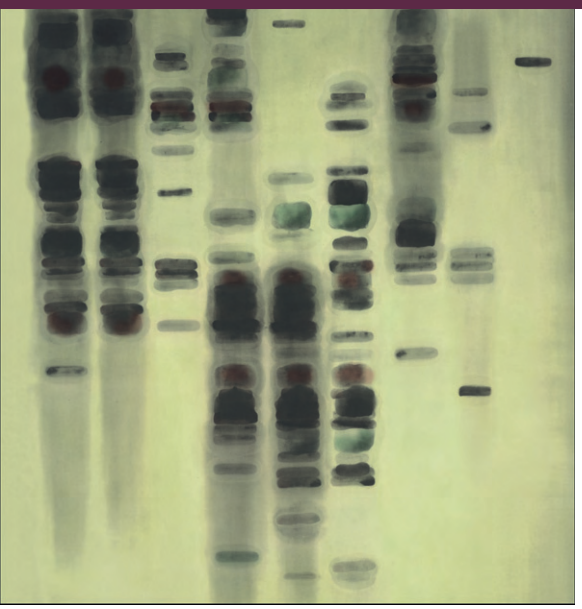
Supplement to PNAS

January 26, 2010 | vol. 107 | suppl. 1 | 1691–1808

Proceedings of the National Academy of Sciences of the United States of America www.pnas.org



Cover image: Pictured is "Bio-gel," an artistic rendering by Dennis Ashbaugh. Evolutionary medicine represents the intersection where evolutionary insights bring a vast and rich body of information to the medical profession and where medical research offers insights, questions, and research opportunities for evolutionary biology. The Sackler Colloquium meeting focused on specific research advances across a wide landscape of medicine and on resolving disjunctions between evolutionary biology and medical science. Image courtesy of Dennis Ashbaugh, freelance artist.

Supplement to the *Proceedings of the National Academy of Sciences of the United States of America*, which includes articles from the Arthur M. Sackler Colloquium of the National Academy of Sciences *Evolution in Health and Medicine*. The complete program is available on the NAS Web site at www.nasonline.org/Sackler_Evolution_Health_Medicine.

Contents

Evolutionary perspectives on health and medicine

Stephen C. Stearns[a,1], Randolph M. Nesse[b], Diddahally R. Govindaraju[c], and Peter T. Ellison[d]

[a]*Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520;* [b]*Departments of Psychiatry and Psychology, University of Michigan, Ann Arbor, MI 48104;* [c]*Department of Neurology, Boston University School of Medicine, Boston, MA 02118; and* [d]*Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138*

Evolution and medicine started an immature romance in the late 19th century that broke up amid violent recriminations in the early 20th century. Thereafter, the relationship remained distant until the partners were reintroduced on a more mature basis by Nesse and Williams' book, *Why We Get Sick: The New Science of Darwinian Medicine* (1). (See ref. 2 for a detailed history.) That book stimulated a symposium in Switzerland in 1996, out of which came a book edited by Stearns (3) that, together with another edited by Trevathan et al. (4), raised interest, connected to the existing body of basic research, and provided materials for the courses that were starting to be offered.

Momentum was further built by several review papers (5, 6), second editions of the two edited books (7, 8), an editorial in *Science* (9), a new textbook (10), and many symposia (Berlin, Rotterdam, York, Copenhagen, New York, Washington, Philadelphia, San Diego, Tucson, and New Haven, among others). Of those symposia, the one held at the National Evolutionary Synthesis Center in 2007 was particularly significant, for it raised medical issues on the home ground of evolutionary biology and brought together the organizers of this Sackler Colloquium. This *PNAS* Supplement marks a significant milestone in the maturation of the field. The range of topics has been expanded, the connections to basic research have been strengthened, the medical community has been more strongly represented, at a higher level, than it had been previously, and the issue of how best to educate future physicians in evolutionary thinking has been developed significantly.

The Interface of Evolution and Medicine

Evolutionary biology and medicine each cover immense scientific landscapes, subsuming many approaches to diverse issues. Evolutionary medicine is not a new specialty or method of practice or critique of medicine. Instead, it consists of the intersections where evolutionary insights bring something new and useful to the medical profession, and where medical research offers new insights, questions, and research opportunities for evolutionary biology. The opportunities are large in the clinic, the research laboratory, and the classroom (3–10). Progress at the interface of evolutionary biology and medicine has given rise to four general messages, three classical themes, and three particularly surprising unique insights.

The four general messages are fundamental but often neglected. First, the view of organisms as machines whose design has been optimized by engineers is as misleading as it is deeply entrenched. Organisms are, instead, bundles of compromises shaped by natural selection to maximize reproduction, not health. They are thus full of unavoidable tradeoffs and constraints (1, 11). Second, because biological evolution is much slower than cultural change, much disease arises from the mismatch of our bodies to modern environments. Third, pathogens evolve much faster than we do, so infection is unavoidable. Fourth, the idea that common heritable diseases are caused by a few defective genes is usually incorrect. An evolutionary view suggests that many genetic variants interact with environments and other genes during development to influence disease phenotypes. Far from suggesting quick new cures, these four general messages help to explain why disease is so prevalent and difficult to prevent.

Three themes at the intersection of evolution and medicine are so well developed they can be considered classic. First, pathogens rapidly evolve resistance to antibiotics just as cancers rapidly evolve resistance to chemotherapy. Second, pathogens evolve strategies to circumvent host defenses, and virulence levels are shaped by natural selection to maximize transmission. Third, human genetic variations that increase disease resistance often have costs, and some variations that increase vulnerability can have benefits. All three classic themes are discussed in articles presented here.

Three previously unexplored insights are particularly surprising. First, humans coevolved with a normal community of symbiotic bacteria and parasitic worms; when they are eliminated by either hygiene or antibiotics, our immune systems can react to this unnatural situation by producing allergies, asthma, and autoimmune disease (12, 13), including very serious ones like Crohn's disease, which can be treated by ingesting eggs of parasitic worms (13). Second, the widespread use of imperfect vaccines, vaccines that do not completely and permanently eliminate the pathogen from the body of the person vaccinated, could lead to an increase in the virulence of the pathogen (14); this is of particular concern in the case of malaria vaccines (15). Third, disruptions of the equilibria achieved in evolutionary conflicts of interest among relatives may be the basis of some mental diseases, particularly autism and schizophrenia, a possibility presented at this meeting, placed in context later in this introduction, and discussed in detail in Crespi et al. (16). All three insights illustrate how evolutionary thinking on medical issues can sometimes illuminate features quite unexpected by nonevolutionary approaches.

The articles in this supplement provide an excellent representation of the topics covered in the Colloquium; however, they cannot, of course, convey the spontaneity or give-and-take that helped to energize the event. All of the presentations and some of the discussion are available for viewing at the National Academy of Sciences Web site (http://www.nasonline.org/site/PageServer?pagename=Sackler_Evolution_Health_Medicine_program). The following overview of the articles helps to situate their contributions both to the Colloquium and as part of a larger effort.

Themes and Articles

The conflict between public good and private interests is at the heart of public health policy. For example, the herd immunity provided by comprehensive vaccination is a public good, but some individuals suffer adverse effects from vaccination. Antibiotic use benefits individuals, but causes the substantial public costs of antibiotic resistance. Althouse, Bergstrom, and Bergstrom develop a quantitative approach to allocation decisions given such externalities and illustrate it with examples of vaccination campaigns and antibiotic management strategies (17). Omenn then provides a comprehensive overview of the public health issues that are impacted by our evolutionary history and current dynamics, and those of our

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences,"Evolution in Health and Medicine" held April 2–3, 2009, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS web site at www.nasonline.org/Sackler_Evolution_Health_Medicine.

Author contributions: S.C.S., R.M.N., D.R.G., and P.T.E. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: stephen.stearns@yale.edu.

pathogens, and argues forcefully that we especially need evolutionary insights when dealing with infectious disease surveillance, gene-by-environment interactions, and global health disparities (18).

Perhaps the clearest basic insight provided by evolution to medicine is the explanation of why we must age (19). Aging is not an adaptation: it is a byproduct of selection for reproductive performance earlier in life. This has been abundantly confirmed by experimental evolution and comparative studies over the last three decades. It is here extended significantly in two articles that get at the mechanisms that mediate the compromises. In the first, Atzmon et al. demonstrate that Ashkenazi centenarians have unusual ability to maintain the length of their telomeres, the caps on the chromosomes made up of repeat DNA sequences that are shortened in each cell division (20). They show that the maintenance of longer telomeres is associated with protection against cognitive deterioration and diseases of aging. In the second, Finch argues that we have achieved a doubling of our lifespan since our last common ancestor with chimpanzees, in part because of evolutionary changes in genes that mediate infection, inflammation, and nutrition (21). Finch focuses in particular on the compromises implicit in the complex effects of apolipoprotein E alleles, which affect immunity, cardiovascular disease, Alzheimer's disease, and brain development, a striking example of the basic insight that our bodies are bundles of evolutionary compromises, not perfect machines designed by engineers (1, 11, 22).

Humans have more cancers than other species for at least three reasons: We now have an extended postreproductive lifespan relatively invisible to natural selection; we are not adapted to the new risk factors generated by civilization, including tobacco, alcohol, a high-fat diet, and contraceptives; and some of our reproductive cancers may be a byproduct of our unique sexuality: continuous cycling, receptivity, and sexual activity. Every cancer evolves within the individual through the multiplication of clones of cells that have accumulated mutations that allow them to escape cell-cycle control. Cancer is virtually inevitable in multicellular organisms that rely on stem cells for tissue maintenance. Frank argues that mutations occurring in cell lineages during development lead to the cell mosaicism that is a precondition for both cancer and certain types of neurodegeneration. He calls for using new technology to measure the dynamics of such genetic cell mosaics to track the evolution of these complex diseases within individual bodies (23).

New medical insights have arisen from the recognition of evolutionary conflicts among relatives. The story begins with Hamilton's work on kin selection in the early 1960s (24). He showed that what matters in evolution is the increase in the numbers of copies of genes in the next generations, no matter through which bodies they are transmitted. Thus, it benefits an organism to sacrifice its own reproductive performance to improve that of a relative if the benefit to the relative, weighted by its degree of relationship, exceeds the cost to the focal organism. A consequence—that asymmetries in inheritance cause evolutionary conflicts among relatives—was developed by Trivers in the 1970s in his theory of parent-offspring conflict (25). In a diploid sexual species a mother is 50% related to all of her offspring, but a focal offspring is 100% related to itself, 50% related to full siblings, and 25% related to half siblings. A mother therefore should divide her investment equally among all offspring, but a focal offspring should try to manipulate her to increase her investment in itself and decrease her investment in its siblings so as to maximize its inclusive fitness. Hamilton's and Trivers's insights have been abundantly confirmed and recognized with major prizes.

Haig took the next step in the early 1990s (26). He saw two things. First, mother and father are also in an evolutionary conflict over investment in offspring whenever the father can have children by more than one female. Then, because the male is 50% related to his own offspring by a female, but 0% related to any offspring she has by another male, he should try to manipulate her to invest in his offspring at the expense of offspring unrelated to him. Second, he noted that there are genes that are differentially imprinted in the germ line, some genes being imprinted – or silenced – in sperm and others in eggs. Some of these genes are expressed in the placenta. They regulate fetal growth and the communications of the fetus with its mother. When the patterns of imprinting are disrupted in genetically engineered mice to express maternal interest without paternal inhibition, the offspring are 10% lighter. When paternal interest is expressed without maternal inhibition, the offspring are 10% heavier. This suggests strongly that the imprinting patterns indeed mediate a parental conflict of interest over investment in offspring, one that may be at the root of pre-eclampsia (dangerously high maternal blood pressure) and gestational diabetes. Here, Haig extends those ideas to show how some of the conflict between the mother and her offspring after birth is being caused by paternal interests, and is mediated by patterns of suckling and rates of maturation (27).

Not all differentially imprinted genes are expressed in the placenta; some are expressed in the brain. That led Crespi and Badcock to postulate, with Haig, that the conflict between maternal and paternal genetic interests over investment is continued after birth and is mediated by infant behavior (28). Such an effect can only be detected when the normal situation—a balance of interests in an evolutionary tug-of-war—is disrupted by a mutation or a developmental event that results in a pathological phenotype. The insight was sparked by the differing effects of deletion or duplication of a single imprinted gene on chromosome 15. When the gene is expressed without the normal paternal inhibition (Prader-Willi syndrome), the mother's interests are expressed without restraint and the child is somnolent, feeds poorly, is easy to care for, and is at high risk (30–70%) of psychosis as an adult. When the opposite pattern occurs (Angelman syndrome), the child is demanding, sleeps poorly, wants to suckle frequently, is difficult to care for, and is at high risk (40–80%) of autism as an adult. Thus, disrupting the equilibrium of an evolutionary conflict of interest appeared to contribute to mental disease.

Here Crespi, Stead, and Elliot extend such analysis of autism and schizophrenia to the impacts of copy number variants (deletions and duplications), further single-gene associations, growth signaling pathways, and brain growth (16). They make a plausible case that the risk of autism is increased by disruption of maternal interests and the uninhibited expression of paternal interests, and that the risk of schizophrenia is increased by the disruption of paternal interests and the uninhibited expression of maternal interests. This is an unconventional but creative approach to serious mental diseases. If it is correct, it will be one of the least expected and most surprising connections in the history of human evolutionary biology. Time will tell.

The processes underlying the origin and emergence of infectious diseases are a key issue in evolutionary medicine. Pathogens with high mutation rates—like RNA viruses—generate enormous genetic diversity and constitute a moving target with which vertebrate immune systems struggle to keep pace. Those high mutation rates also make possible very detailed analysis of their relationships and history, allowing us, for example, to accurately infer the origins of HIV/AIDS (29). Here, Holmes applies his comprehensive knowledge of the evolution of RNA viruses (30) to make two points: lethal mutagenesis may be an underexploited method of viral control, and lack of surveillance of pathogenic/virulent strains circulating in swine impeded

our ability to predict the emergence of H1N1 influenza (31).

Thus far we have mostly reviewed medical consequences of specific evolutionary insights. Another important branch of evolutionary medicine consists of studies that deepen our understanding of basic, general, evolutionary processes. Evolutionary geneticists do much of this work, documenting, for example, evidence for past selection in the genome (32). Meanwhile, specialists on phenotypic evolution are making increasingly important contributions that respond to two facts: selection acts on phenotypes, not on genes, and patients are phenotypes. Both approaches are necessary, and both are represented here: first the genetic, then the phenotypic.

If we reduce evolution to its molecular elements, then the process is initiated by single nucleotide mutations and the consequent substitution, in some cases, of changed single amino acids in proteins. Proteins are composed of hundreds of amino acids, and getting from one functional state to another may be a journey of many steps across a fitness landscape whose topography has until recently been unknown. Carneiro and Hartl present an exquisitely detailed analysis of the fitness landscapes encountered by mutations to three enzymes (33). They conclude that actual proteins display much more additivity and less epistasis than randomly simulated proteins. This finding is important because it means that real biological systems are more likely to be able to attain fitness maxima that had previously been thought inaccessible; they can get across rougher topography in the fitness landscape than we had thought.

The sequencing of the human genome opened the possibility for examining differences among individuals nucleotide by nucleotide. The human genome can now be examined for differences at individual nucleotides, called single nucleotide polymorphisms, at millions of sites in the genome. This finding spurred the hope that by examining such variation across the entire genome, we would be able to discover a majority of the genetic variants involved in any complex human diseases or traits. Some have been discovered; however, the amount of total genetic variation explained by the already discovered genetic variants has been much smaller than had been hoped and promised. In an article that carefully applies basic ideas in evolutionary genetics, Eyre-Walker shows that when we consider how selection acts on the sources of genetic variance in a trait, we find that most of the genetic variance of a trait—most of its heritability—is contributed by mutations at low frequency in the population, and that the effects of rare mutations tend to be much larger than those of common mutation (34). The resulting paradoxical situation has frustrated recent genome-wide association studies: mutations that have strong effects on fitness are likely to be rare in populations, and hence difficult to detect; and mutations that are easy to detect have small effects on disease. This is the most parsimonious evolutionary reason why most genome-wide association studies fail to explain more than a few percent of the variation in a trait.

Recently, interest in epigenetics has increased strikingly (35–37). Epigenetics focuses on developmental changes occurring within a single genome that do not involve changes in DNA sequence. One important class of epigenetic change is mediated by methylation of genes; inheritance of methylation patterns within cell lineages contributes to the stabilization of the differentiated state in different tissues. Feinberg and Irizarry explore an evolutionary consequence of variation across individuals in methylation state: genes that increase such variation among individuals can have higher fitness in a varying environment when the epigenetic variation is realized at the level of the whole organism as phenotypic plasticity, resulting in performance better matched to each state in the varying environment (38). This unique idea has potential to resolve several outstanding puzzles in quite different areas of biology.

Another area in which interest has also recently increased is the structural variation (inversion, deletion, and duplications) in the genome of which copy-number variation is the most abundant form. The classic view of the genome architecture was that each of us had the same number of copies of each of the genomic regions. Once extensive sequence data became available, it became clear that the classical view was false. For example, Sebat et al. (39) examined 20 individuals and found that they differed on average by 11 copy-number polymorphisms, each of which represented on average a sequence of 465 kilobases. Within those sequence intervals, they found copy-number variation in 70 different genomic locations, which involved genes influencing neurological function, regulation of cell growth, regulation of metabolism, and known to be associated with disease. Those early results have been abundantly confirmed. Here Carvalho, Zhang, and Lupski provide a comprehensive review of copy number and other structural variations in the human genome that has allowed them to develop the concepts of genomic instabilities that both cause disease and contribute to adaptation (40). One puts down their article with a sense that structural variation in the genome, and its consequences for health and disease, will be a rich source of research results for a long time to come.

Human populations are usually thought to be poor candidates for studies of basic questions about the evolution and maintenance of fitness traits; the effects of culture are profound, and environments are variable and far different from those the species evolved in. Sometimes, however, special cultural conditions offer something like a natural situation. Kosova, Abney, Ober report data from a Hutterite population where birth control is not used and social stratification is minimized by cultural constraints (41). Armed with an extraordinary database of demographic information over three generations, they ask about the correlations among and heritability of reproductive variables closely correlated with fitness. They find completed family size is influenced by birth rate and even more by age at last reproduction, but that age at last reproduction is little influenced by birth rate. For these traits, heritability estimates for women range from 0.23 to 0.28; for men the heritability is higher, up to 0.68 for completed family size. These data cannot address the basic question of how so much variation persists in heritable traits that correlate highly with fitness. However, they illustrate the potential for continuing evolution of traits in modern societies and how evolutionary thinking can spur creative analysis of a remarkable dataset.

Mutations happen and disease results, but the vast majority of harmful mutations are recessive and subject to selection only when an individual has two copies. Phenotypes with disease from recessive homozygotes are at low frequency because selection has shaped mechanisms in many species, including humans, to foster outbreeding. However, there is substantial cultural variation. Across the globe, 10.4% of spouses are second cousins or closer, but the proportion varies dramatically from <1% to over 50%. In some cultures, first or second cousins are preferred marriage partners because of the social benefits. For instance, Charles Darwin and Emma Wedgewood were first cousins, and Darwin was concerned this might have accounted for health problems in his children. Estimates of the effects of such inbreeding are important not only for practical reasons: they also offer clues to the prevalence of genes that affect fitness, often without any associated identifiable disease. Using data from 69 societies, Bittles and Black report an improved estimate of excess mortality rates in offspring from first cousin marriages of about 3.5% (42). This finding is consistent with many deleterious recessive alleles with usually small effects. They also note a strong trend for decreasing consanguineous marriages in technological societies, with reduced social advantages of marrying relatives. It is interesting to

contemplate the consequences of increased outbreeding for the public health of future populations.

The prevalence of the notion that natural selection has ended for humans illustrates the degree of common misunderstandings about evolution. Individuals with some heritable phenotypes are having more offspring than others, so natural selection continues to shape our species. Major changes take thousands of years, but can we identify any traits associated with variations of reproductive success? Byars, Ewbank, Govindaraju, and Stearns address the question with one of the more remarkable databases in medicine, that from the Framingham Heart Study (43). Using data on lifetime reproductive success, they apply standard evolutionary methods to estimate the selection gradients arising from measured variables, including weight and age at first birth. Sure enough, the role of these factors in selection is observed, and they are even capable of assessing the effects in different decades, to conclude that the most consistently important trait influencing reproductive success is age at first birth, which is predicted to change slowly over successive generations.

As noted already, selection is recorded in genotypes and genomic regions, but natural selection acts on phenotypes. Houle notes that new genomic methods have left our knowledge grossly unbalanced: "the depth of our knowledge of genomes is approaching completeness, whereas our knowledge of phenotypes remains, by comparison, minimal." Most common disease phenotypes are influenced by thousands of genes with millions of variants. If these variants were common and had large effects, progress would be fast, but they are not. In fact, for most common diseases no specific common genes have major effects. We need a new approach. The solution, according to Houle, is phenomics, the large-scale study of high-dimensional phenotypes, and "the natural and inevitable complement to genomics" (44). He advocates developing detailed phenotype-genotype contour maps reminiscent of Sewall Wright's adaptive landscape (45). The same mathematical tools used to describe changes arising from natural selection can be applied to the task of describing the relationships of phenotypes to disease states. Large-scale efforts at phenotyping have not occurred to date, largely because they are expensive, but there are good evolutionary reasons for thinking the payoffs of such a program would be worth the effort.

Filling the Education Gap

The above articles in this special supplement illustrate the value of evolutionary approaches for diverse problems in medicine and public health; however, they also illustrate the opportunities not yet grasped because of the wide gap between evolutionary biology and medicine. Few medical schools have evolutionary biologists on their faculties and none teach evolutionary biology as a basic medical science. Some physicians and medical researchers learn something about evolution before medical school, but few have anywhere near the level of knowledge we demand for other basic sciences. The articles in this supplement illustrate the opportunities in research, but general applications of evolution in medicine may be equally valuable. An evolutionary view corrects mistaken notions of the body as a designed machine, and it gives physicians a feeling for the organism and a sense for what disease is (22). What is needed to fill the gap? Nesse et al. argue that substantially improved evolution education before medical school is needed, and specific renovations of the medical curriculum are also essential (46). Progress in evolution education at specific schools is coming quickly, but new national policies are needed if we are to educate physicians who can make full use of evolution as a crucial basic science for medicine. In addition to changes in medical school curricula, changes in premedical education can have a particularly powerful effect. Requiring competency in evolutionary biology on the Medical College Admissions Test (MCAT) will probably improve understanding of evolutionary issues among clinicians more than any other single measure. In addition to changes in the MCAT itself, every undergraduate institution should offer courses in evolutionary medicine as part of its premedical curriculum.

Will increased investments such as Nesse et al. suggest be worth it? The question is legitimate. Competent medical practice already presupposes long training in many complex subjects, some of which are quite distant from everyday medical practice. Adding another competency to an already packed curriculum requires strong justification. Participants at the Colloquium concluded, as we do, that such justification exists: evolutionary insights are already saving lives, reducing suffering, and can help us to avoid major unpleasant scientific surprises. Ignorance among physicians about fundamental evolutionary principles must be ended. The payoffs of evolutionary thinking are clearest in designing better programs to manage the evolution of antibiotic resistance in pathogens and drug resistance in cancer. Many people can be kept alive longer, in better condition, if we more wisely manage antibiotic treatments and chemotherapy. The potential for anticipating and avoiding unpleasant surprises is greatest where we seek to understand the consequences of large-scale campaigns with vaccines that permit some pathogens to escape: their virulence could increase (15). In addition, evolutionary insights shed light on the reasons for multiple spontaneous abortions (47, 48), preeclampsia, and pregnancy-related diabetes (26); on the potential to treat auto-immune diseases by managing our symbiotic fauna of bacteria and worms (12); on the emergence of new infectious diseases and subsequent changes in their transmissibility and virulence (49); and much more.

Conclusions

The Colloquium that gave a forum for these articles was the culmination of at least a score of smaller meetings; however, it should by no means be viewed as the conclusion. This meeting focused strongly on specific research advances across a wide landscape of medicine. It only touched the surface of public health. It said little about behavioral factors that influence disease. And, the coverage of education and policy recommendations was necessarily brief. We hope that this meeting and the articles in this supplement will inspire many to arrange additional communication ventures, some more focused, some more broad, and many, hopefully, organized by and for practicing physicians.

The general conclusion, looking over the entire supplement, is that existing bridges between medicine and the basic science of evolutionary biology are getting increased traffic, and new ones are being constructed, but significant gulfs remain to be spanned. In particular, current funding mechanisms reinforce a disjunction between evolutionary biology and medical science and make the development of research programs at their intersection problematic. The National Science Foundation and the National Institutes of Health each currently see this area as outside their respective domains, even while advocating increased interdisciplinary research. To move forward, these major federal funding agencies must negotiate a way to close this gap and support innovative science that does not fit within existing funding structures. Science like that represented in this Supplement is too exciting to neglect. It is as though a lost isthmus between two continents has been discovered, one that opens remarkable new frontiers and paths toward powerful strategies for prevention and cure.

1. Nesse RM, Williams GC (1994) *Why We Get Sick: The New Science of Darwinian Medicine* (Vintage Books, New York).
2. Zampieri F (2009) Medicine, evolution and natural selection: An historical overview. *Q Rev Biol* 84: 333–356.
3. Stearns SC ed (1998) *Evolution in Health and Disease* (Oxford University Press, Oxford).
4. Trevathan WR, McKenna JJ, Smith EO ed (1999) *Evolutionary Medicine* (Oxford University Press, New York).
5. Stearns SC, Ebert D (2001) Evolution in health and disease: work in progress. *Q Rev Biol* 76:417–432.
6. Nesse RM, Stearns SC (2008) The great opportunity: Evolutionary applications to medicine and public health. *Evol Appl* 1:28–48.
7. Trevathan WR, McKenna JJ, Smith EO eds (2007) *Evolutionary Medicine, Second Edition* (Oxford University Press, New York).
8. Stearns SC, Koella JK eds (2008) *Evolution in Health and Disease, Second Edition* (Oxford University Press, Oxford).
9. Nesse RM, Stearns SC, Omenn GS (2006) Medicine needs evolution. *Science* 311:1071.
10. Gluckman P, Beedle A, Hanson M (2009) *Principles of Evolutionary Medicine* (Oxford University Press, Oxford, UK).
11. Held LI (2009) *Quirks of Human Anatomy: An Evo-Devo Look at the Human Body* (Cambridge University Press, Cambridge, New York).
12. Zaccone P, Fehervari Z, Phillips JM, Dunne DW, Cooke A (2006) Parasitic worms and inflammatory diseases. *Parasite Immunol* 28:515–523.
13. Rook G ed (2009) *The Hygiene Hypothesis and Darwinian Medicine* (Birkhauser Basel-Boston-Berlin, Boston, MA).
14. Gandon S, Mackinnon MJ, Nee S, Read AF (2001) Imperfect vaccines and the evolution of pathogen virulence. *Nature* 414:751–756.
15. Mackinnon MJ, Read AF (2004) Virulence in malaria: an evolutionary viewpoint. *Philos Trans R Soc Lond B Biol Sci* 359:965–986.
16. Crespi B, Stead P, Elliot M (2009) Comparative genomics of autism and schizophrenia. *Proc Natl Assoc Sci USA* 107(Suppl):1702–1709.
17. Althouse BM, Bergstrom TC, Bergstrom CT (2010) A public choice framework for controlling transmissible and evolving diseases. *Proc Natl Assoc Sci USA* 107 (Suppl):1696–1701.
18. Omenn GS (2009) Evolution and public health. *Proc Natl Assoc Sci USA* 107(Suppl):1702–1709.
19. Williams GC (1957) Pleiotropy, natural selection, and the evolution of senescence. *Evolution* 11:398–411.
20. Atzmon G, et al. (2009) Genetic variation in human telomerase is associated with telomere length in Ashkenazi centenarians. *Proc Natl Assoc Sci USA* 107 (Suppl):1710–1717.
21. Finch CE (2009) Evolution of the human lifespan and diseases of aging: Roles of infection, inflammation, and nutrition. *Proc Natl Assoc Sci USA* 107(Suppl): 1718–1724.
22. Childs B, Wiener C, Valle D (2005) A science of the individual: implications for a medical school curriculum. *Annu Rev Genomics Hum Genet* 6:313–330.
23. Frank SA (2009) Somatic evolutionary genomics: Mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration. *Proc Natl Assoc Sci USA* 107(Suppl):1725–1730.
24. Hamilton WD (1964) The genetical evolution of social behaviour. I. *J Theor Biol* 7:1–16.
25. Trivers RL (1974) Parent-offspring conflict. *Am Zool* 14: 249–264.
26. Haig D (1993) Genetic conflicts in human pregnancy. *Q Rev Biol* 68:495–532.
27. Haig D (2009) Transfers and transitions: Parent-offspring conflict, genomic imprinting, and the evolution of human life history. *Proc Natl Assoc Sci USA* 107(Suppl):1731–1735.
28. Crespi B, Badcock C (2008) Psychosis and autism as diametrical disorders of the social brain. *Behav Brain Sci*, 31:241–261, discussion 261–320.
29. Keele BF, et al. (2006) Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* 313:523–526.
30. Holmes EC (2009) *The Evolution and Emergence of RNA Viruses* (Oxford University Press, Oxford).
31. Holmes EC (2009) The comparative genomics of viral emergence. *Proc Natl Assoc Sci USA* 107(Suppl): 1742–1746.
32. McVicker G, Gordon D, Davis C, Green P (2009) Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet* 5:e1000471.
33. Carneiro MC, Hartl DL (2009) Adaptive landscapes and protein evolution. *Proc Natl Assoc Sci USA* 107(Suppl): 1747–1751.
34. Eyre-Walker A (2009) The genetic architecture of a complex trait and its implications for fitness and genome wide association studies. *Proc Natl Assoc Sci USA* 107(Suppl):1752–1756.
35. West-Eberhard MJ (2003) *Developmental Plasticity and Evolution* (Oxford University Press, Oxford, New York).
36. Jablonka E, Lamb MJ (2005) *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life* (MIT Press, Cambridge, Mass.).
37. Gilbert SF, Epel D (2009) *Ecological Developmental Biology: Integrating Epigenetics, Medicine, and Evolution* (Sinauer Associates, Sunderland, Mass.).
38. Feinberg AP, Irizarry RA (2009) Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Assoc Sci USA* 107 (Suppl):1757–1764.
39. Sebat J, et al. (2004) Large-scale copy number polymorphism in the human genome. *Science* 305:525–528.
40. Carvalho CMB, Zhang FZ, Lupski JR (2010) Genomic disorders—A window into human gene and genome evolution. *Proc Natl Assoc Sci USA* 107(Suppl): 1765–1771.
41. Kosova G, Abney M, Ober C (2009) Heritability of reproductive fitness traits in a human population. *Proc Natl Assoc Sci USA* 107(Suppl):1772–1778.
42. Bittles AH, Black ML (2009) Consanguinity, human evolution, and complex diseases. *Proc Natl Assoc Sci USA* 107(Suppl):1779–1786.
43. Byars SG, Ewbank D, Govindaraju DR, Stearns SC (2009) Natural selection in a contemporary human population. *Proc Natl Assoc Sci USA* 107(Suppl):1787–1792.
44. Houle D (2009) Numbering the hairs on our heads: The shared challenge and promise of phenomics. *Proc Natl Assoc Sci USA* 107(Suppl):1793–1799.
45. Wright S (1988) Surfaces of selective value revisited. *Am Nat* 131:115–123.
46. Nesse RM, et al. (2009) Making evolutionary biology a basic science for medicine. *Proc Natl Assoc Sci USA* 107 (Suppl):1800–1807.
47. Ober C (1992) The maternal-fetal relationship in human pregnancy: an immunogenetic perspective. *Exp Clin Immunogenet* 9:1–14.
48. Ober C, et al. (1997) HLA and mate choice in humans. *Am J Hum Genet* 61:497–504.
49. Ebert D, Bull JJ (2003) Challenging the trade-off model for the evolution of virulence: is virulence management feasible? *Trends Microbiol* 11:15–20.

A public choice framework for controlling transmissible and evolving diseases

Benjamin M. Althouse[a], Theodore C. Bergstrom[b], and Carl T. Bergstrom[a,1]

[a]Department of Biology, University of Washington, Seattle, WA 98195-1800; and [b]Department of Economics, University of California, Santa Barbara, CA 93106

Edited by Peter T. Ellison, Harvard University, Cambridge, MA, and approved November 17, 2009 (received for review July 23, 2009)

Control measures used to limit the spread of infectious disease often generate externalities. Vaccination for transmissible diseases can reduce the incidence of disease even among the unvaccinated, whereas antimicrobial chemotherapy can lead to the evolution of antimicrobial resistance and thereby limit its own effectiveness over time. We integrate the economic theory of public choice with mathematical models of infectious disease to provide a quantitative framework for making allocation decisions in the presence of these externalities. To illustrate, we present a series of examples: vaccination for tetanus, vaccination for measles, antibiotic treatment of otitis media, and antiviral treatment of pandemic influenza.

antibiotic resistance | externalities | vaccination | emerging infectious disease | pandemic influenza

Control measures used to limit the spread of infectious disease often generate both direct effects on the targeted individuals and indirect effects on other parties. As we shall see, these side consequences, or *externalities*, may be either beneficial or detrimental, depending on the biology of the disease and the nature of the interventions.

An effective framework for making health policy decisions must account for both direct effects and externalities. This requirement situates the control of infectious disease as a problem in public choice economics. Several economic studies have considered the role of externalities in public health. Brito et al. (1) first posed the problem, addressing the positive externalities associated with vaccination programs. Subsequent analyses (2, 3) use this approach to illustrate the difficulty of entirely eradicating a disease by vaccination. Gersovitz and Hammer further expand the economic models and extend the analysis to public goods associated with vector control (4–6). Cook et al. (7) propose a graphical approach similar to that presented here and work out a detailed case study for cholera vaccination. However, none of these analyses address the negative externalities associated with the evolutionary process, namely the evolution and spread of antimicrobial resistance. Within the field of infectious disease epidemiology, there is a growing realization that game-theoretic problems arise around vaccine uptake (8–14) and antibiotic resistance evolution (15, 16). However, epidemiologists have yet to adopt a general framework for handling the economic externalities associated with disease control measures.

Here we fuse these two lines of research to sketch a general framework for considering the consequences of control interventions on evolving infectious diseases. To illustrate the fundamental issues that arise, we will step through a sequence of four infectious diseases. We begin by describing tetanus, for which vaccination is a pure private good and no externalities arise. Next we use measles to illustrate a case in which the externalities, namely reduced infectious pressure due to herd immunity (17), are entirely positive. We then consider otitis media, a case in which the externalities, namely selection for antibiotic resistance, are due to evolutionary processes and are purely negative. We conclude with a model of pandemic influenza, in which antiviral treatment simultaneously generates both positive and negative externalities. Each of our examples is intended to be illustrative rather than predictive, and thus we have put a high premium on simplicity; in each case, we use the most basic models possible with simple approximate parameters. Faced with an actual policy decision for a specific disease, one would do well to sacrifice the simplicity here for the realism conferred by more complex disease models. Extension is straightforward to any type of disease model, deterministic or stochastic, analytical or simulation-driven, well-mixed, age-structured, explicitly spatial, or network-based. As far as the economic modeling is concerned, the disease model is simply a "black box" that generates the marginal public benefit curves by specifying the hazard of infection and its derivatives as a function of how broadly and to whom interventions are allocated. Even when these quantities cannot be obtained analytically, numerical methods can be used for all of the calculations described herein.

Economic Method

The economic theory of public choice provides a mathematical framework for making decisions in the presence of externalities (18, 19). In this paper, we are concerned with the allocation of public health interventions: vaccinations, antibiotics, and antivirals. Due to externalities that arise with these goods, the privately obtained levels are not always socially efficient. Some sort of government intervention such as subsidies or taxes may be needed to achieve efficiency.

We begin by describing the basic economic formulation, using vaccination as an example. As we will show, this framework can also be adapted to deal with antimicrobial therapy. Consider a large population N in which a fraction q of the population is vaccinated. The annual hazard of infection is $h(1, q)$ for those who are vaccinated and $h(0, q)$ for those who are not. Let individual i have a utility function $U_i(h_i, x_i)$, where h_i is the hazard rate and x_i is the amount of goods that i consumes in time period t. If i has initial wealth w_i and pays p to be vaccinated, then $x_i = w_i - p$. Individual i's private benefit from vaccination is a function $v_i(q)$, which is defined as the most money that i would be willing to pay to be vaccinated. Thus $v_i(q)$ is implicitly defined by the equation

$$U_i[h(1,q), w_i - v_i(q)] = U_i[h(0,q), w_i]. \quad [1]$$

Although the method generalizes easily to heterogeneous populations, as we will see in the otitis media example, for now consider the special case where consumers have identical utility functions of the form

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Evolution in Health and Medicine" held April 2–3, 2009, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS web site at www.nasonline.org/Sackler_Evolution_Health_Medicine.

Author contributions: B.M.A., T.C.B., and C.T.B. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: cbergst@u.washington.edu.

$$U(h,x) = x - kh. \quad [2]$$

When utility is of the form in Eq. **2**, it follows from Eq. **1** that

$$v(q) = k[h(0,q) - h(1,q)]. \quad [3]$$

If individuals must pay a price c for vaccination, they will prefer to be vaccinated if and only if the private benefits $v(q)$ of vaccination exceed the private cost c.

Assume that the total cost of vaccinating a fraction q of the population is $C(q\,N)$. In this case, the efficient level of q is that which maximizes the sum of individual utilities subject to the constraint that total consumption of goods equals the sum of individual wealths minus total resource costs of vaccinations. This sum is equal to the sum of the utilities of all of the vaccinated individuals plus the sum of the utilities of the unvaccinated, which is

$$\begin{aligned}\sum_i U_i(h,w_i) &= \sum_i w_i - C(qN) - N\,q\,k\,h(1,q) \\ &\quad - N(1-q)kh(0,q) \\ &= \sum_i w_i - C(qN) + N\,q\,v(q) - N\,k\,h(0,q).\end{aligned}$$

It follows that the efficient level of vaccination q_e is that which maximizes

$$N\,q\,v(q) - C(q\,N) - N\,k\,h(0,q). \quad [4]$$

If we differentiate expression **4** with respect to q and divide by N, we have the following first-order necessary condition for efficiency:

$$v(q) - C'(q\,N) - k\left[q\frac{\partial h(1,q)}{\partial q} + (1-q)\frac{\partial h(0,q)}{\partial q}\right] = 0. \quad [5]$$

Let us define

$$X(q) = -k\left[q\frac{\partial h(1,q)}{\partial q} + (1-q)\frac{\partial h(0,q)}{\partial q}\right]. \quad [6]$$

Thus, $X(q)$ measures the marginal value to the population of the indirect effects of vaccination. This is the externality generated by vaccinating a single individual. Note that it is composed of two terms: the $\partial h(1, q)/\partial q$ term reflects the effect of treating one additional individual on the other treated individuals, and the $\partial h(0, q)/\partial q$ term reflects the effect of treating one additional individual on the untreated individuals. In the remainder of this paper, we look at how these terms operate for four illustrative diseases. For tetanus, both terms are zero and no externalities are generated. For measles vaccination, which generates a positive externality for the unvaccinated (but not the vaccinated), the $\partial h(1, q)/\partial q$ is zero and the $\partial h(0, q)/\partial q$ is negative. For antibiotic treatment of otitis media, which generates a negative externality for the other treated individuals (but not the untreated), the $\partial h(1, q)/\partial q$ is positive and the $\partial h(0, q)/\partial q$ is zero. For antiviral treatment of pandemic flu, which generates both positive and negative externalities, both terms are nonzero.

Suppose that individuals are required to pay the marginal cost of their vaccination. Because the private benefit of vaccination is $v(q)$, the equilibrium level of vaccination will be q_p such that $v(q_p) = C'(N\,q_p)$. To induce an efficient outcome, the government can provide a so-called Pigouvian subsidy, subsidizing each vaccination by an amount $X(q_e)$ equal to the marginal value of the indirect effects at the efficient point.

Tetanus

With our economic framework in place, we turn to a series of examples. We begin with one of the few cases in which disease prevention is a purely private good. Tetanus is the severe prolonged contraction of skeletal muscle fibers caused by the neurotoxin tetanospasmin produced by the bacterium *Clostridium tetani*. Whereas tetanus kills hundreds of thousands of individuals worldwide every year, it is virtually nonexistent in industrialized nations, with a reported incidence of 50–100 cases per year in the United States over the past 30 years. This low incidence is due directly to vaccination use (20, 21).

Tetanus is not contagious, making it the only vaccine-preventable disease that is infectious but not transmissible human-to-human (20). As a result, there is no herd immunity from vaccination. In terms of our model, the hazard rate for an unvaccinated individual to become infected is simply a constant, independent of q, and thus $\partial h(0, q)/\partial q = 0$. Furthermore, the vaccine targets the tetanospasmin toxin rather than the bacterium that produces it, thereby minimizing the selective consequences of vaccination on the bacterium itself and minimizing the risk of resistance evolution, even relative to other vaccines (22). Hence, the hazard rate for a vaccinated individual to become infected is also independent of q and so $\partial h(1, q)/\partial q = 0$ as well. Thus, for tetanus, $X(q) = 0$, and there are no externalities associated with vaccination. Tetanus toxoid vaccine is a purely private good; an individual will choose to vaccinate if his or her benefit is larger than the cost, irrespective of what others are doing—and this decision has no side effects on other individuals' risks of disease.

Measles

Next we consider a disease for which control generates positive externalities from reduced transmission. Measles is caused by a paramyxovirus and kills an estimated 242,000 people globally per year, despite an effective and readily available vaccine (23). Vaccination is the primary form of control, and generates a positive externality because high vaccination levels induce "herd immunity" (17) that reduces the risk of infection to nonvaccinated individuals.

To investigate the dependence of hazard rates on vaccination, we consider the stationary equilibrium of a basic susceptible-infectious-recovered (SIR) model of vaccination (see ref. 24). Expected lifespan is T years and population size is constant, with birth rate $\mu = 1/T$ equal to the death rate. Individuals who are vaccinated will be vaccinated at birth. We assume that the vaccine is perfectly protective, so that $h(1, q) = 0$ for all q. At equilibrium, those who are not vaccinated will face a constant hazard rate of contracting measles, but after recovering will not be susceptible. Let γ be the recovery rate and β the infection transmission rate. Let $S(t)$ be the fraction of the population that is susceptible to infection, $I(t)$ be the fraction of the population that is infectious, and $R(t)$ be the fraction that is recovered from the disease and no longer infectious. Where $\dot{S}$, $\dot{I}$, and $\dot{R}$ indicate derivatives with respect to time, the governing differential equations are

$$\begin{aligned}\dot{S} &= \mu(1-q) - \beta\,S\,I - \mu\,S \\ \dot{I} &= \beta\,S\,I - (\gamma+\mu)I \\ \dot{R} &= \gamma\,I + \mu\,q - \mu\,R.\end{aligned} \quad [7]$$

We define *basic reproductive number* $R_0 = \beta/(\gamma + \mu)$. To find the equilibrium level of infection $I^*(q)$ when a fraction q is vaccinated, we take $\dot{S} = \dot{I} = \dot{R} = 0$. We find that

$$I^*(q) = \begin{cases} \frac{\mu}{\beta}(R_0(1-q)-1) & \text{if } R_0(1-q) \geq 1 \\ 0 & \text{if } R_0(1-q) < 1. \end{cases} \quad [8]$$

We see that the equilibrium fraction of infected individuals decreases linearly as the vaccinated proportion of the population

increases, so long as $R_0(1-q) > 1$. If $R_0(1-q) \leq 1$, the disease will be eradicated. The annual hazard rate for an unvaccinated, susceptible individual is given by

$$h(0,q) = \beta I^*(q) = \mu[R_0(1-q)-1]. \quad [9]$$

We assume that vaccination at the beginning of life results in lifetime immunity. Suppose that individuals who are vaccinated are asked to pay the marginal cost of their own vaccination in equal installments over their lifetimes. Let c be the marginal cost of vaccination; then the annual payment would be $c/T = \mu\, c$. The annual utility gain from vaccination is $v(q) = k\ h(0, q)$. At an interior equilibrium, individuals would be indifferent about being vaccinated or not. Thus the privately supported equilibrium proportion of vaccinated individuals would be q_p, such that

$$v(q_p) = k\, h(0,q_p) = k\ \mu[R_0(1-q_p)-1] = \mu\, c. \quad [10]$$

This implies that at equilibrium the fraction of the population that is unvaccinated is

$$1-q_p = \frac{c+k}{kR_0}. \quad [11]$$

The efficient level of vaccination q_e is that which maximizes the sum of individual utilities subject to the constraint that total consumption of goods equals the sum of total wealths minus total resource costs of vaccination. The total annual cost of vaccinations is $C(q\ \mu\ N)$ and the total annual cost of infection is

$$(1-q)N\, k\, h(0,q) = (1-q)N\, k\mu[R_0(1-q_p)-1]. \quad [12]$$

An efficient outcome, therefore, is one that minimizes

$$C(q\ \mu\ N) + (1-q)N\, k\ \mu[R_0(1-q)-1]. \quad [13]$$

If the marginal cost of an additional vaccination is c, the first-order condition for an efficient rate of vaccination q_e is

$$\mu N\, c = \mu N\, k[2R_0(1-q_e)-1]. \quad [14]$$

This implies that

$$1-q_e = \frac{c+k}{2\, kR_0}. \quad [15]$$

We see from Eqs. **11** and **15** that in an efficient outcome, the fraction of the population that is left unvaccinated is just half of the equilibrium unvaccinated fraction if individuals pay the marginal cost of their vaccinations.

In this model, the annual externality from a vaccination is $X(q) = k\ \mu\ R_0(1-q)$. Thus, an annual subsidy of $k\ \mu\ R_0(1-q_e) = \mu(c + k)/2$ for those who choose vaccination would be sufficient to induce an efficient outcome. Fig. 1 illustrates.

As seen in previous analyses, complete eradication can be difficult even with government subsidy, due to the so-called prevalence elasticity in private demand of vaccine: Prevalence of a disease declines through increased vaccination use; the willingness to pay for vaccination decreases as well (2, 25). However, in the framework defined above, increasing vaccine prevalence through government intervention still provides a positive externality in reducing the prevalence of measles (4).

Otitis Media

Otitis media is an inflammation of the duct known as the eustachian tube, which connects the middle ear to the nasopharynx.

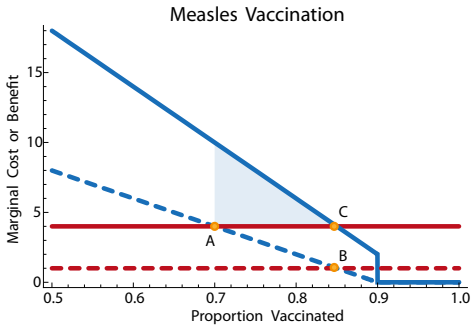


Fig. 1. Measles vaccination with a cost subsidy. The annual marginal cost or benefit is plotted. Solid lines: marginal public cost (red) and marginal public benefit (blue) of vaccination. Once 90% of the population is vaccinated, the disease is eradicated and no further benefit accrues to vaccination. Dashed lines: net private cost (red) is the cost minus the subsidy; net private benefit (blue) of vaccination. Point A: private optimum without subsidy. Point B: private level of vaccination with subsidy. Point C: efficient level of vaccination. An annual subsidy brings the level of vaccination at the private optimum into line with the efficient level. The shaded region illustrates the net welfare gain due to the subsidy. The parameters in this example, with time measured in years and costs in dollars, are as follows. The recovery rate is $\gamma = 100$. So that $R_0 \approx 10$, the transmission parameter is $\gamma = 1000$. The annual valuation of reduced risk is $k = 100$; the total cost of vaccination (financial cost plus perceived risk of being vaccinated) is $c = 200$. Given a death rate of $\mu = 0.02$, the annual cost is $c\ \mu = 4$ and the annual subsidy is $\mu(c + k)/2 = 3$.

Typically caused by ineffective clearing of *Streptococcus pneumoniae* or *Haemophilus influenzae* bacteria from the duct, otitis media is the leading cause of antibiotic prescription in children and adolescents in the United States (26). Due to the position of infection, human-to-human transmission of the infection-causing bacteria is unlikely. With asymptomatic carriage in the nasopharynx extremely common and transmission much more likely from this site than from the eustachian tube, treatment of clinical infection has little or no effect in reducing transmission.

The economics of otitis media treatment differ from measles vaccination, as there exists a negative externality of antibiotic resistance caused by overprescription of antibiotics (27–29). Although most cases of otitis media are caused by bacterial infection, some are viral in etiology and often the two are difficult to distinguish. This commonly leads to unnecessary antibiotic treatment. Moreover, most cases resolve spontaneously without treatment. The American Academy of Pediatrics now recommends that no antibiotic treatment be given for most cases in children over the age of 2 in the absence of severe illness (30).

We adjust our interpretation of the model parameters to fit the specifics of otitis media. We now interpret N as the total number of clinical cases and q as the fraction of cases that receive antibiotic therapy, and we replace the hazard rate h with a function π that gives the chance of developing complications that require further intervention beyond any initial antibiotic therapy.

There is considerable variation across individuals in their need for antibiotic treatment. To account for the individual variation in need for antibiotic treatment, we extend our economic model to consider the case in which individual utility functions differ. Suppose that there is a continuum of consumers such that the utility function of consumer t is

$$U(\pi,x) = x - k(t)\pi,$$

where $t \in [0, 1]$ and k is a nondecreasing function of t.

In an efficient allocation, all of those who are treated must have at least as high a value of k as all of those who are not. If a fraction q of the cases are treated and the persons treated

are chosen efficiently, those who are treated must have $k(t) \geq k(1-q)$ and those who are not treated must have $k(t) \leq k(1-q)$. The total cost of treatment is $C(N\,q)$, and the integral of consumers' utilities is

$$W - C(Nq) - N\,\pi(1,q)\int_{1-q}^{1} k(t)dt - N\,\pi(0,q)\int_{0}^{1-q} k(t)dt, \quad \textbf{[16]}$$

with $W = \sum_i w_i$. Let us define $K^+(q) = \int_{1-q}^{1} k(t)dt$ and $K^-(q) = \int_{0}^{1-q} k(t)dt$. Then expression **16** can be written as

$$W - C(Nq) - N\pi(1,q)K^+(q) - N\pi(0,q)K^-(q). \quad \textbf{[17]}$$

Taking the derivative of expression **17** with respect to q, and dividing by N, the first-order condition for efficiency is

$$C'(q) = k(q)[\pi(0,q) - \pi(1,q)] - K^+(q)\frac{\partial\pi(1,q)}{\partial q} - K^-(q)\frac{\partial\pi(0,q)}{\partial q}$$

If the fraction q of cases are treated, the private benefits to an individual to whom the cost of infection is $k(t)$ will be

$$v(t,q) = k(t)[\pi(0,q) - \pi(1,q)].$$

If marginal cost is a constant c, and individuals purchase their own treatment at a price equal to the marginal cost, an equilibrium will be an outcome in which $k(q)(\pi(0,q) - \pi(1,q)) = c$. Thus, the marginal effect of a change in q on the population, $X(q)$, is much as before but with $K^+(q)$ and $K^-(q)$ replacing the k q and $k(1-q)$ terms:

$$X(q) = -K^+(q)\frac{\partial\pi(1,q)}{\partial q} - K^-(q)\frac{\partial\pi(0,q)}{\partial q}. \quad \textbf{[18]}$$

Antibiotic use typically selects for resistance; as usage increases, we expect the prevalence of antibiotic-resistant strains to increase and, with it, the probability of treatment failure (31). In our model, this means that the probability of complications despite treatment $\pi(1, q)$ will be an increasing function of q. To determine the form of this function, we again can turn to mathematical models in disease epidemiology.

Here we use a simple susceptible-infected-susceptible (SIS) model of antibiotic use in the community to treat a human-commensal bacterium, following ref. 32. Only a small fraction α of individuals carrying the bacterium experience symptoms and are candidates for treatment; the treatment fraction q now refers to the fraction of these symptomatic cases treated. For a population of size N, we let S be the proportion of individuals not carrying the bacterial species of interest, and let I_w and I_r be the proportion of people carrying wild-type bacteria or resistant bacteria, respectively. Let γ_w and γ_r be the rates of spontaneous clearance for wild-type and resistant bacterial carriage, and let γ_t be the clearance rate due to treatment for wild-type carriage. Let β be the force of infection and let σ be the rate at which treated wild-type individuals develop de novo resistance. The governing differential equations are

$$\begin{aligned}
\dot{S} &= -\beta\,S(I_w + I_r) + (\gamma_w + \alpha\,q\,\gamma_t)I_w + \gamma_r\,I_r \\
\dot{I}_w &= \beta\,S\,I_w - (\gamma_w + \alpha\,q\,\gamma_t + \alpha\,q\,\sigma)I_w \\
\dot{I}_r &= \beta\,S\,I_r - \gamma_r\,I_r + \alpha\,q\,\sigma I_w.
\end{aligned} \quad \textbf{[19]}$$

At steady state, sensitive and resistant strains will coexist provided that $\beta > \gamma_w + \alpha\,q\,\gamma_t + \alpha\,q\,\sigma$ and $\kappa > \alpha\,q(\gamma_t + \sigma)$, where $\kappa = \gamma_r - \gamma_w$ is the differential clearance rate. In this case, the equilibrium frequency of resistant bacteria [i.e., $I_r/(I_w + I_r)$ at equilibrium] is $p(q) = \alpha\,q\,\sigma/(\kappa - \alpha\,q\,\gamma_t)$ and the derivative with respect to q is $p'(q) = \alpha\,\kappa\,\sigma/(\kappa - \alpha\,q\,\gamma_t)^2$.

We can now compute $\partial\pi(1,q)/\partial q$ and $\partial\pi(0,q)/\partial q$. The latter is straightforward: Assuming that there is no difference in the virulence of resistant and sensitive strains, the frequency of resistance does not directly impact untreated individuals and thus $\partial\pi(0, q)/\partial q = 0$. The former term depends on the frequency of treatment failure, which in turn is proportional to the frequency of resistant bacteria. Let ρ be the chance that treatment failure leads to complications. Then $\partial\pi(1, q)/\partial q = p'(q)\rho$ and $X(q)$ is negative. The efficient level of treatment q_e lies below the private equilibrium, and some government intervention is required to discourage usage if we are to reach the efficient outcome. In this case, a tax of $X(q_e)$ will shift the private equilibrium to the efficient point. This optimal taxation level is illustrated in Fig. 2.

Pandemic Influenza

Novel influenza A (H1N1) was first reported in Mexico City in late April 2009, and within 6 weeks (June 11) was declared a full pandemic by the World Health Organization. The most widely used antiviral agents, neuraminidase inhibitors oseltamivir and zanamivir, have demonstrated beneficial effects on H1N1 infection and are the WHO-recommended first-line and chemoprophylactic treatment for these viruses (33, 34). However, resistance to antiviral agents is a concern for public health planners (35). Resistance can evolve readily with minimal fitness costs (36), and the use of neuraminidase inhibitor antivirals has the potential to decrease the long-term effectiveness of these drugs.

In this section, we develop a simple illustrative model of pandemic influenza. Our analysis is based on ref. 37, with the simplification that we model only antiviral treatment and not antiviral prophylaxis. Our state variables are as follows: X is the fraction of susceptible individuals in the population; Y_{su} is the fraction of individuals infected with sensitive virus but untreated with the antiviral; Y_{st} is the fraction of individuals infected with sensitive virus and treated with the antiviral; Y_r is the fraction of individuals infected with resistant virus; and Z is the fraction of removed (recovered or dead) individuals in the population. Let γ be the recovery rate and let σ be the rate at which sensitive

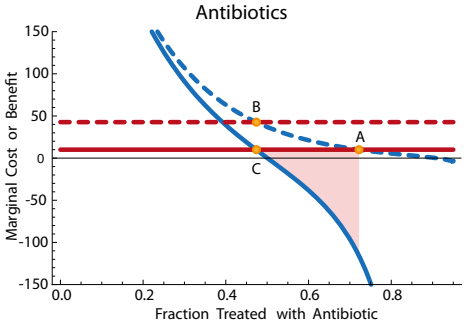


Fig. 2. Antibiotic treatment for otitis media. Solid lines: marginal public cost (red) and marginal public benefit (blue) of antibiotic therapy. Dashed lines: net private cost with tax (red) and net private benefit (blue) of antibiotic therapy. Point A: private optimum without tax. Point B: private level of antibiotic therapy with tax. Point C: efficient level of antibiotic use. The tax brings the level of treatment at the private optimum into line with the efficient level. The shaded region illustrates the net welfare gain due to the tax. Parameters with time in days and costs in dollars are as follows: $\gamma_w = 0.05$ and fitness cost of resistance is a 10% increase in clearance rate so that $\gamma_r = 0.055$. We have $\gamma_t = 0.5$, $\sigma = 0.05$, $\alpha = 0.01$, $\beta = 1$, $\rho = 0.1$. The cost of antibiotics is $c = 10$. The function $k(t)$ specifying the consumer values of treatment is an exponential curve: $k(t) = 5000\,e^{-5(1-t)}$.

infections evolve de novo resistance. Let β_{su}, β_{st}, and β_r be the transmission parameters for untreated sensitive, treated sensitive, and resistant virus, respectively. As in the model of ref. 37, treatment reduces the infectiousness but not the duration of infection. Let q be the fraction of drug-sensitive infections that receive treatment. The governing differential equations are then

$$\begin{aligned}
\dot{X} &= -(\beta_{su}Y_{su} + \beta_{st}Y_{st})X - \beta_r Y_r X \\
\dot{Y}_{su} &= (\beta_{su}Y_{su} + \beta_{st}Y_{st})(1-q)X - \gamma Y_{su} \\
\dot{Y}_{st} &= (\beta_{su}Y_{su} + \beta_{st}Y_{st})q(1-\sigma)X - \gamma Y_{st} \\
\dot{Y}_r &= (\beta_{su}Y_{su} + \beta_{st}Y_{st})q\sigma X + \beta_r Y_r X - \gamma Y_r \\
\dot{Z} &= \gamma(Y_{su} + Y_{st} + Y_r).
\end{aligned} \qquad [20]$$

In the previous examples, we looked at how the treatment fraction influenced the steady-state prevalence of the disease. Such an approach is not appropriate for pandemic influenza, which sweeps through a population and then is eradicated, rather than reaching a steady endemic level. Thus, we will focus in this example on the time course of infection, tracking the number of individuals infected over time.

Fig. 3 shows how the fractions of individuals infected with resistant and sensitive strains, over the course of an epidemic, depend on the fraction of cases treated with antivirals. As seen in previous analyses, intermediate levels of antiviral therapy minimize the total number of cases (37) by reducing the degree to which the epidemic overshoots its eradication threshold (38).

In our economic analysis, individuals have incentive to purchase antivirals at cost c because treatment ameliorates symptoms of influenza, such that k_t and k_u represent the dollar costs of a case of treated and untreated influenza, respectively. Under these assumptions, utility functions are of the form $U(\pi_t, \pi_u, x) = x - k_t\pi_t - k_u\pi_u$, where x is the amount spent on other goods, π_t is the probability that the individual contracts a sensitive case and treats it, and π_r is the probability that an individual contracts a case and does not treat it, either because he or she had chosen not to purchase the antiviral or because the case is resistant.

We consider two different ways in which antivirals might be dispensed. In the *flu kits* scenario (39), the consumer has an advance option to purchase one course of antiviral therapy to be used in the event that he or she is infected in a pandemic. For simplicity, in this scenario, we assume that no further courses of antiviral therapy will be available during the epidemic to those who did not choose to purchase them initially. In the *pharmacy distribution* scenario, no advance purchase is made. Instead, when an individual is infected, he or she has the option to purchase a course of antivirals for immediate use from the pharmacy, but this decision is made without knowing whether the infection is sensitive or resistant.

The π_u and π_t values can be computed directly from the fraction of the population that has been infected by each strain at the end of the epidemic, conditioned on the treatment choices of the individual. Let $\pi_r(q)$ be the fraction of the population that has been infected by resistant strains at the end of the epidemic and let $\pi_s(q)$ be the fraction of the population infected by sensitive strains. In the flu kit scenario, the purchase is made in advance of infection. Ignoring resale and exchange of antivirals, the fraction of sensitive infections treated, q, will be equal to the fraction of the population who chose to purchase the flu kit. In this case, the private benefit is $v(q) = (k_u - k_t)\pi_s(q)$. In the pharmacy distribution scenario, the purchase is deferred until the individual is known to be infected. The fraction of sensitive infections treated will be equal to the fraction of individuals who choose to purchase treatment once infected. In this case, the private benefit is conditioned on infection by some strain of flu:

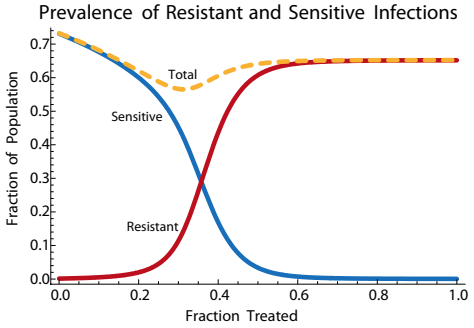


Fig. 3. Fraction of resistant (red) and sensitive (blue) infections for pandemic flu, as a function of the fraction of infections treated. Total fraction infected is indicated by the dashed yellow curve. Parameters follow Lipsitch et al. (37) as follows: the basic reproductive ratio R_0 is 1.8 for sensitive virus. Resistant virus suffers a 10% fitness cost in the form of reduced transmission. Treatment reduces transmission of sensitive virus by 67%. The infectious period is 3.3 days, and each treated case has a 1/500 chance of developing de novo resistance. These values correspond to γ = 1/3.3, σ = 0.002, β_{su} = 1/1.8, $\beta_r = 0.9\beta_{su}$, $\beta_{st} = 0.33\beta_{su}$.

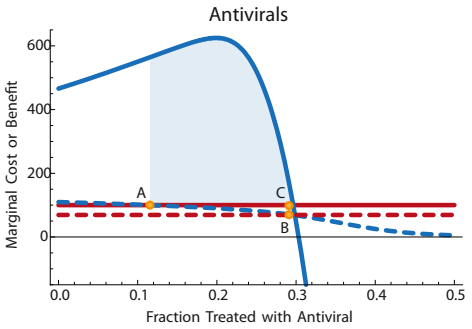


Fig. 4. Antiviral treatment for pandemic flu, flu kit scenario. Solid lines: marginal public cost (red) and marginal public benefit (blue) of antiviral therapy. Dashed lines: net private cost with subsidy (red) and net private benefit (blue) of antiviral therapy. Point A: private optimum without subsidy. Point B: private level of antiviral therapy with subsidy. Point C: efficient level of antiviral use. The subsidy brings the level of treatment at the private optimum into line with the efficient level. The shaded region illustrates the net welfare gain due to the subsidy. Disease parameters are as in Fig. 3, with time in days and costs in dollars. Cost of antivirals is c = 100, and disease valuation parameters are k_t = 850, k_u = 1000.

The π_u and π_t values can be computed directly from the fraction of the population that has been infected by each strain at the end of the epidemic, conditioned on the treatment choices of the individual. Let $\pi_r(q)$ be the fraction of the population that has been infected by resistant strains at the end of the epidemic and let $\pi_s(q)$ be the fraction of the population infected by sensitive strains. In the flu kit scenario, the purchase is made in advance of infection. Ignoring resale and exchange of antivirals, the fraction of sensitive infections treated, q, will be equal to the fraction of the population who chose to purchase the flu kit. In this case, the private benefit is $v(q) = (k_u - k_t)\pi_s(q)$. In the pharmacy distribution scenario, the purchase is deferred until the individual is known to be infected. The fraction of sensitive infections treated will be equal to the fraction of individuals who choose to purchase treatment once infected. In this case, the private benefit is conditioned on infection by some strain of flu:

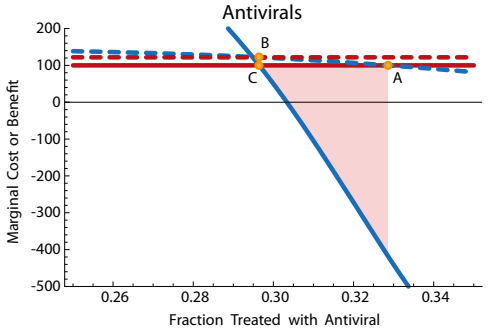


Fig. 5. Antiviral treatment for pandemic flu, pharmacy distribution scenario. Parameters are as in Fig. 4. Here the private market overuses antivirals and a tax is required to bring the level of treatment at the private optimum into line with the efficient level. The shaded region illustrates the net welfare gain due to the tax.

$v(q) = (k_u - k_t)\pi_s(q)/[\pi_s(q) + \pi_r(q)]$. In each case, we can solve numerically for the private equilibrium fraction q of individuals who choose to purchase antiviral therapy. As in our previous examples, the social optimum is the point that maximizes the sum of the utilities. This can also be computed numerically. Fig. 4 and Fig. 5 illustrate.

This model illustrates an interesting tradeoff. When antiviral usage is low, increasing use generates a positive externality in the form of reduced total cases. When antiviral use gets higher, this positive externality is reduced and, moreover, a negative externality arises in the form of reduced antiviral effectiveness. Depending on the distribution technology, the private market may over- or underuse antivirals relative to the efficient level. For the particular parameter values here, too few people stockpile antivirals if required to purchase them in advance of the pandemic, whereas too many use antivirals if allowed to purchase them at the time of infection. Corresponding subsidies or taxes can bring the private equilibrium into line with the efficient point, as illustrated in Fig. 4 and Fig. 5.

Summary

When coupled with mathematical models of disease, the economic theory of public choice provides a robust framework for assessing the economic impact of public health interventions. As we have illustrated, this framework allows decision makers to account for the positive and negative externalities associated with control measures such as vaccination or antimicrobial chemotherapy. Depending on the biology of the disease and the nature of the intervention, the private market may under- or overallocate. Taxes or subsidies can guide the private market to a more efficient level of intervention. Although the models presented here have been deliberately simple, the framework illustrated can be fruitfully applied to more sophisticated quantitative models of infectious disease. Doing so should help public health planners more efficiently control evolving transmissible diseases.

ACKNOWLEDGMENTS. The authors thank Marc Lipsitch for helpful discussions. This work was supported by the National Institute of General Medical Sciences Models of Infectious Disease Agent Study Program cooperative agreement 5U01GM07649 and the MIDAS Center for Communicable Disease Dynamics 1U54GM088588 at Harvard University.

1. Brito DL, Sheshinski E, Intriligator MD (1991) Externalities and compulsory vaccinations. *J Public Econ* 45:69–90.
2. Geoffard P-Y, Philipson T (1997) Disease eradication: Private versus public vaccination. *Am Econ Rev* 87:222–230.
3. Francis PJ (1997) Dynamic epidemiology and the market for vaccinations. *J Public Econ* 63:383–406.
4. Gersovitz M, Hammer JS (2003) Infectious diseases, public policy, and the marriage of economics and epidemiology. *World Bank Res Obs* 18:129–157.
5. Gersovitz M, Hammer JS (2004) The economic control of infectious diseases. *Econ J* 114:1–27.
6. Gersovitz M, Hammer JS (2004) Tax/subsidy policies toward vector-borne infectious diseases. *J Public Econ* 89:647–674.
7. Cook J, et al. (2009) Using private demand studies to calculate socially optimal vaccine subsidies in developing countries. *J Policy Anal Manage* 28:6–28.
8. Bauch CT, Galvani AP, Earn DJD (2003) Group interest versus self-interest in smallpox vaccination policy. *Proc Natl Acad Sci USA* 100:10564–10567.
9. Bauch CT, Earn DJD (2004) Vaccination and the theory of games. *Proc Natl Acad Sci USA* 101:13391–13394.
10. Reluga TC, Bauch CT, Galvani AP (2006) Evolving public perceptions and stability in vaccine uptake. *Math Biosci* 204:185–198.
11. Galvani AP, Reluga TC, Chapman GB (2007) Long-standing influenza vaccination policy is in accord with individual self-interest but not with the utilitarian optimum. *Proc Natl Acad Sci USA* 104:5692–5697.
12. Vardavas R, Breban R, Blower S (2007) Can influenza epidemics be prevented by voluntary vaccination? *PLoS Comput Biol* 3:e85.
13. van Boven M, Klinkenberg D, Pen I, Weissing FJ, Heesterbeek H (2008) Self-interest versus group-interest in antiviral control. *PLoS One* 3:e1558.
14. Medlock J, Galvani AP (2009) Optimizing influenza vaccine distribution. *Science* 325: 1705–1708.
15. Smith DL, Levin SA, Laxminarayan R (2005) Strategic interactions in multi-institutional epidemics of antibiotic resistance. *Proc Natl Acad Sci USA* 102:3153–3158.
16. Foster KR, Grundmann H (2006) Do we need to put society first? The potential for tragedy in antimicrobial resistance. *PLoS Med* 3:e29.
17. Fine PEM, Clarkson JA (1986) Individual versus public priorities in the determination of optimal vaccination policies. *Am J Epidemiol* 124:1012–1020.
18. Samuelson PA (1954) The pure theory of public expenditure. *Rev Econ Stat* 36: 387–389.
19. Samuelson PA (1955) Diagrammatic exposition of a theory of public expenditure. *Rev Econ Stat* 37:350–356.
20. Centers for Disease Control (2008) *Epidemiology and Prevention of Vaccine-Preventable Diseases (The Pink Book)* (Public Health Foundation, Washington, DC), 10th Ed.
21. Centers for Disease Control (1991) Diphtheria, tetanus, and pertussis: Recommendations for vaccine use and other preventive measures: Recommendations of the Immunization Practices Advisory Committee (ACIP). *MMWR Morb Mortal Wkly Rep* 40 (RR-10):1–28.
22. Read AF, Mackinnon MJ (2008) Pathogen evolution in a vaccinated world. *Evolution in Health and Disease*, eds Stearns SC, Koella JC (Oxford University Press, Oxford), 2nd Ed.
23. World Health Organization (2007) *Measles Fact Sheet* (World Health Organization, Geneva).
24. Keeling MJ, Rohani P (2008) *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, Princeton, NJ).
25. Philipson T (1996) Private vaccination and public health: An empirical examination for U.S. measles. *J Hum Resour* 31:611–630.
26. McCaig LF, Besser RE, Hughes JM (2002) Trends in antimicrobial prescribing rates for children and adolescents. *JAMA* 287:3096–3102.
27. Goossens H, Ferech M, Vander Stichele R, Elseviers M; ESAC Project Group (2005) Outpatient antibiotic use in Europe and association with resistance: A cross-national database study. *Lancet* 365:579–587.
28. Lipsitch M (2001) Measuring and interpreting associations between antibiotic use and penicillin resistance in *Streptococcus pneumoniae*. *Clin Infect Dis* 32:1044–1054.
29. Lipsitch M, Samore MH (2002) Antimicrobial use and antimicrobial resistance: A population perspective. *Emerg Infect Dis* 8:347–354.
30. American Academy of Pediatrics Subcommittee on Management of Acute Otitis Media (2004) Diagnosis and management of acute otitis media. *Pediatrics* 113: 1451–1465.
31. Bergstrom CT, Feldgarden M (2008) The ecology and evolution of antibiotic-resistant bacteria. *Evolution in Health and Disease*, eds Stearns SC, Koella JC (Oxford University Press, Oxford), 2nd Ed.
32. Bonhoeffer S, Lipsitch M, Levin BR (1997) Evaluating treatment protocols to prevent antibiotic resistance. *Proc Natl Acad Sci USA* 94:12106–12111.
33. Centers for Disease Control (2009) *Interim Guidance on Antiviral Recommendations for Patients with Novel Influenza A (H1N1) Virus Infection and Their Close Contacts* (Centers Dis Control, Atlanta).
34. World Health Organization (2006) *WHO Rapid Advice Guidelines on Pharmacological Management of Humans Infected with Avian Influenza A (H5N1) Virus* (World Health Organization, Geneva).
35. Centers for Disease Control (2009) *Interim Recommendations for the Use of Influenza Antiviral Medications in the Setting of Oseltamivir Resistance Among Circulating Influenza A (H1N1) Viruses, 2008–09 Influenza Season* (Centers Dis Control, Atlanta).
36. Weinstock DM, Zuccotti G (2009) The evolution of influenza resistance and treatment. *JAMA* 301:1066–1069.
37. Lipsitch M, Cohen T, Murray M, Levin BR (2007) Antiviral resistance and the control of pandemic influenza. *PLoS Med* 4:e15.
38. Handel A, Longini IM (2007) Jr, Antia R (2007) What is the best control strategy for multiple infectious disease outbreaks? *Proc Biol Sci* 274:833–837.
39. Traynor K (2008) Antiinfluenza medication kits need work, FDA advisers conclude. *Am J Health Syst Pharm* 65:2314–2316.

Evolution and public health

Gilbert S. Omenn[1]

Center for Computational Medicine and Bioinformatics, Departments of Internal Medicine and Human Genetics, Medical School and School of Public Health, University of Michigan, Ann Arbor, MI 48109–2218

Edited by Peter T. Ellison, Harvard University, Cambridge, MA, and approved October 23, 2009 (received for review September 11, 2009)

Evolution and its elements of natural selection, population migration, genetic drift, and founder effects have shaped the world in which we practice public health. Human cultures and technologies have modified life on this planet and have coevolved with myriad other species, including microorganisms; plant and animal sources of food; invertebrate vectors of disease; and intermediate hosts among birds, mammals, and nonhuman primates. Molecular mechanisms of differential resistance or susceptibility to infectious agents or diets have evolved and are being discovered with modern methods. Some of these evolutionary relations require a perspective of tens of thousands of years, whereas other changes are observable in real time. The implications and applications of evolutionary understanding are important to our current programs and policies for infectious disease surveillance, gene–environment interactions, and health disparities globally.

cultural evolution | ecogenetics | genome mapping | susceptibility to infection | Western diet

Public health practice and public health research focus on protecting, enhancing, and understanding the health of communities and populations. The scientific disciplines of epidemiology, environmental and occupational health, and health behavior address causes and risk factors of disease over time and space. The substrate for the study of evolution in public health includes international patterns of incidence and prevalence of disease, influences of human and animal behavior, dramatic changes in diet, environmental sources of exposures to infectious agents and chemicals, diverse causes of migration of populations, and climate change. Advances in population genetics and evolutionary biology now facilitate in-depth analysis of gene–environment interactions in human populations and in other species whose life cycles are intimately linked with our own.

This Perspectives article addresses principles and examples of the roles of infectious diseases, cultural/social factors, and diet and metabolism in evolution and public health. It emphasizes implications for gene–environment interactions, global health, health disparities, and health policy.

Principles of Evolutionary Influences in Public Health

The accumulated and ongoing genomic and behavior variation in human populations makes us differentially susceptible to a broad range of disease agents, ranging from infections to obesity.

The interactions of disease agent, intermediate hosts or risk factors, and human host reflect variation and evolution over very different time scales—with microbes the most rapid by far, including microbes in our own microbiome.

Humans—through cultural, behavioral, and technological changes—have become the most disruptive and significant agents of change for the rest of life on the planet.

Susceptibility and Resistance to Infectious Diseases

Throughout human history, infectious diseases have been among the most important causes of mortality and morbidity for humans, including plague, smallpox, tuberculosis (TB), measles, and diarrheal infections (1). Studies of the origins and distribution of infectious diseases examine the geographic distribution, life stage, and evolution of the infectious agent [malaria parasites, TB mycobacteria, cholera bacteria, influenza, severe acute respiratory syndrome (SARS), and HIV]; the geographic distribution and life cycle of intermediate hosts (arthropod vectors for many diseases, birds for avian flu, bats for SARS, and deer and ticks for Lyme disease spirochetes); the geographic distribution of diseases they cause in humans and other species; and the key clues that some population subgroups are strikingly more or less susceptible than others. Infectious agents are also important factors in major "noninfectious" inflammatory diseases, like certain cancers, atherosclerosis, and arthritis (2).

Malaria. The protozoan parasite *Plasmodium falciparum* causes the most severe form of malaria. It causes more than 1 million deaths annually. It occurs over a wide geographic distribution of Africa, the Mediterranean, and south Asia. Altitude is associated with dramatic differences in rates of malaria infection, correlated with the distribution of the mosquitoes. The mosquitoes multiply in stagnant pools of water, a situation probably driven by agricultural practices involving deforestation both long ago and currently. *Plasmodium* species are excellent examples of infectious agents that have an obligatory intermediate host such that the status of humans is closely tied to the geographic distribution and activities of that species, which is the *Anopheles* mosquito in the case of malaria. Malaria particularly attacks children and young adults, providing the substrate for natural selection when genetic or behavioral factors provide differential resistance to infection or propagation of the parasite in humans. The most obvious means of avoiding infection are migration away from geographic areas with high prevalence of *Anopheles* and *Plasmodium* and elimination of the *Anopheles* host with antimalarial chemicals such as dichlorodiphenyltrichloroethane, which was very effective globally before its ban because of adverse effects on bird populations.

In addition, there are dramatic differences in susceptibility of individuals embedded in genomic variation. Children and adults with sickle cell trait (HbS) have red blood cells less hospitable to the life stage of the malaria parasite that infects and propagates in the blood than the red blood cells of individuals with normal HbA. Individuals with HbS are more likely than "normals" to survive infection with *P. falciparum*. In 1954, Allison (3) deduced that malaria was the selective factor that maintained the *HbS* gene in certain population subgroups in the face of high mortality from sickle cell anemia in individuals with a double dose of the *HbS* gene. Understanding that the life stage in red blood cells is critical, he and many other researchers examined the potential role of other genetic abnormalities of red blood cells, with dramatic findings. β-Thalassemias, other hemoglobinopathies (e.g., HbC, HbE), and glucose-6-phosphate dehydrogenase (G6PD)

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Evolution in Health and Medicine" held April 2–3, 2009, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS web site at www.nasonline.org/Sackler_Evolution_Health_Medicine.

Author contributions: G.S.O. designed research, performed research, and wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed at: University of Michigan/CCMB, Room 2017F Palmer Commons, 100 Washtenaw Avenue, Ann Arbor, MI 48109–2218. E-mail: gomenn@umich.edu.

deficiency fit this same pattern of enhanced survival of heterozygotes (gene carriers) in the face of malaria as a negative selective factor (4,5). Sickle cell anemia (*HbSS*), sickle cell trait (*HbS*/*HbA*), and normal hemoglobin (*HbAA*) represent a balanced polymorphism. The incidence of sickle cell anemia in parts of equatorial Africa is as high as 1/25 of the population, compared with 1/400 among African Americans. People ill with malaria have reduced fertility. However, red blood cells of heterozygotes are more readily removed from the circulation than are normal (HbA/A) red blood cells parasitized with *P. falciparum*. A 20% increase in fitness for individuals with the trait could balance an 85% decrease in fitness of homozygous *HbSS* individuals (6).

An entirely different model for evolution of resistance and susceptibility emerged from international studies of blood group antigens on red blood cells. The Fy-allele of the Duffy blood group system on red blood cells is ubiquitous among Black Africans but is very rare or absent in Asian and white populations. Individuals who are Fy^{-}/Fy^{-} have complete resistance against infection with *Plasmodium vivax*, the parasite responsible for a different form of malaria (1). The molecular mechanism of this clinical and public health association is of general importance: the Fy blood group is the receptor through which the *P. vivax* parasite enters erythrocytes. This biochemical polymorphism had sufficient survival advantage in West Africa that nearly the entire population became Duffy-negative. Combination with other infections and poor nutrition has been postulated to make it life-threatening, and hence selective (1). However, proof of causal relations after thousands of years is challenging; thus, another view is that the Duffy-negative allele might have become prevalent for reasons not observable now and acted to prevent this relatively mild form of malaria from becoming pandemic in West Africa.

HIV/AIDS. An analogous discovery of the defective-receptor mechanism of resistance explains the epidemiological observation that some men very highly exposed to the HIV/AIDS virus did not become infected. The most striking specific mechanism involves a mutant CCR5 receptor on lymphocytes with a 32-aa deletion. CCR5 is an essential component of the entry mechanism for HIV. If there is no entry, there is no infection and no transmission risk. There is no explanation yet for what natural selection force led to CCR5 mutations accumulating in the human population. Not all resistant individuals have this mutation; thus, there must be other explanations that could reveal additional important features of HIV infection and targets for prevention or therapy. Of course, the selective factor might have been some other agent altogether. From the pursuit of this line of research, we now know 20 polymorphisms of receptors, coreceptors, cytokine ligands, and HLA genes that influence susceptibility to HIV infection, replication, or relevant innate or adaptive immunity; the detailed modes of action reveal features of evolutionary selection (7). The presence of the CCR5 receptor seems to protect against West Nile virus (8); thus, public health use of CCR5 inhibitors to try to reduce risk for HIV/AIDS could lead to increased risk for West Nile Virus encephalitis. Viruses have a long history of coevolution with molecules of the immune system. Specific HLA-B alleles influence both the rate of progression to AIDS in HIV-infected individuals and the adaptation of viral sequences within the host and at large. There are homologies of human nonprogressors to chimpanzees that tolerate a strain of simian immunodeficiency virus (SIVcpz) without immunopathology. Chimps may have survived a selective sweep after a viral epidemic in the distant past.

Since the emergence of HIV/AIDS in the early 1980s, there has been intense interest in public health and lay circles about the origins of the virus (both HIV-1 and HIV-2). Long-frozen serum samples from central Africa (from the malaria studies) were shown by Nahmias et al. (9) to harbor HIV in at least one case from 1959. Evidence now suggests that these viruses were introduced to humans only in the 20th century in central Africa from lentiviruses in nonhuman primates who suffer no pathology from the infection. Molecular phylogeny studies have compared and classified these viruses. HIV-1 evolved from a strain of SVIcpz in a subspecies of chimpanzees on at least three occasions, whereas HIV-2 originated in SIVsm of sooty mangabeys numerous times. Many other SIVs have not gained a foothold in humans to date (7).

The observation that 8% of the human genome consists of "endogenous retroviral sequences" suggests strongly that our species has a long history of infection with, responses to, and coevolution and coexistence with retroviruses in what is thought to be a dormant state but could include some continuing pathology (10).

Influenza and SARS. With the recent and current international threats of influenza pandemics (H5N1 "avian" and H1N1/2009 "swine"), the public and policymakers realize again that the influenza viruses are highly mutable and capable of adapting rapidly to selective factors in their environments. The H5N1 flu seems to have originated via reassortment among avian flu strains in eastern Asia. The H1N1 strain(s) may have complex origins. Flu strains have variable potential to infect highly exposed humans from their reservoirs in other species and highly variable risk for human-to-human transmission. As reflected in the uncertainties annually about the morbidity and mortality risks from seasonal flu (200,000 hospitalizations and 36,000 deaths in an average year) and from a pandemic each generation or so, we know too little about the variation in susceptibility of humans to influenza viruses other than direct immunity to previously experienced strains. One major barrier limiting cross-transmission of avian influenza into humans (and vice versa) is the evolution of differences in sialic acid linkage binding specificity. The human and avian virus hemagglutinins prefer binding α-2–6- and α-2–3-linked sialic acids, respectively, on epithelial cells in target tissues. In addition, chimpanzees and other great apes do not express the human upper airway epithelial α-2–6-linked sialic acid targets for human influenza viruses (11). Current research utilizes reconstituted influenza strains and reverse genetics to discover the specific genes and gene combinations that may drive virulence and host range. Also, it is feasible to model the effects of vaccines and drugs on the evolution and dynamics of flu strains.

Another remarkable cross-species transmission, from mammals to humans handling infected animals and then to other humans, occurred with the coronavirus SARS in 2002–2003. Fortunately, modern genetic epidemiological methods led to rapid identification and control of this virus after outbreaks and economic disruption in Hong Kong and Toronto, linked by an air traveler. A compelling surveillance strategy was launched by Wolfe et al. (12) to set up stations in remote areas of the world where unusual infectious agents may exist among animals and might get their foothold in humans through infection of highly exposed animal handlers. In general, further mutations and selection would be necessary to make such viruses or other microbes capable of human-to-human transmission.

Microbiome. Our intestinal tract and every surface and orifice are rich sources of microbes in complex communities. There are many more microbial cells than human cells in our bodies. They perform critical functions in digestion and host defenses. We and our microbiota have coevolved; we provide unique habitats that have restricted colonization to a relatively small number of phyla (13). Our changing diet, hygiene practices, medical therapies, chemical exposures, and public health programs continue to lead to changes in the microbiome. Widespread use of broad-spectrum antibiotics has opened habitats to unique organisms. The National Institutes of Health launched a major initiative focused on genomics, ecology, informatics, and clinical implications of the microbiome (http://nihroadmap.nih.gov/hmp/workshop0407/index.asp).

Helminths (Worms). Worms in the intestinal tract used to be "normal" before sanitation. There is some evidence that lack of worm loads has become associated with increased rates of autoimmune disorders, diabetes, and childhood leukemias (14). Public health programs to treat worm infestations have been associated with increased asthma and Crohn disease rates. Cross-reactivity between worm antigens and dust mites may contribute to high rates of asthma among African Americans (15). Conversely, genes that are associated with greater risk for asthma may be protective against worms. An evolutionary bioinformatics approach to worms has been employed by Divergence, Inc. using the worm genome sequences published by the Washington University Genome Center and comparative genomics to identify drug targets in worms that, because of divergent evolution, do not exist in the crops, livestock, or humans they infect (16).

Antibiotic Resistance: An Arms Race Between Species—Evolution in Action

Within the microbial world, there is remarkable interspecies competition and cooperation. Microbes exchange genetic material, even with different genera. They compete for space and food sources, adapting to selective pressures. Fungi have been particularly adept at producing antimicrobial chemicals that protect them against bacteria. Starting with Fleming's use of the extract of *Penicillium* to kill Gram-positive bacteria, patients have benefited from these antibiotics from nature (17). These chemicals may be isolated and used directly, or they may serve as lead compounds for drug development. However, microbes are not passive agents. They respond promptly to negative natural selection in the form of antibiotics by developing genetically transmitted resistance to the action of individual antibiotics or sets of antibiotics. If these microbes are pathogenic to humans, our response is to create generations of antibiotics; hence, the "arms race."

Multiple-Drug-Resistant TB. One of the most threatening situations in public health during the past 20 years was the emergence of multiple-drug-resistant (MDR) TB mycobacteria, especially in patients with HIV/AIDS (18). Health care workers in New York City, New York State, and elsewhere were infected during care of patients with such TB and were at risk for untreatable illnesses. Fortunately, the public health community mobilized aggressively to identify and isolate such patients and provide them with whatever anti-TB therapy was still effective for their organisms in a setting of directly observed administration of the drug. Ensuring full dosage and full course of treatment is essential to avoid selecting additional resistance genotypes. The original outbreaks were contained, but MDR-TB remains a threat worldwide. TB was also an early application of genotyping methods to enhance epidemiological surveillance and discern patterns of transmission, which was a breakthrough for this organism that is so difficult to culture in the clinical laboratory (18).

Multiply Resistant *Staphylococcus aureus*. Among nosocomial or health care-associated infectious threats, multiply resistant *Staphylococcus aureus* (MRSA) is a prime example. The majority of cases of invasive infections in the United States now are acquired outside the hospital but mostly reflect recent hospitalization or surgery; community-acquired and hospital-acquired infections tend to be attributable to quite distinct strains monitored by the Centers for Disease Control and Prevention (CDC) Emerging Infections Program (19). These strains are highly adapted to the human host and are poised to invade wounds and the bloodstream. Infection control requires judicious use of our current arsenal of antimicrobials, excellent sanitary practices, and continued development of drugs. Parallel evolution of MRSA has been observed in different hospitals. Shifts to different antibiotics in the hospital formulary have stopped some hospital epidemics. More complex is the question of how to reduce the risk for and severity of these infections over long periods. Evolutionary biology and ecological theory were used to test the concept of alternating two or more classes of antibiotics over months or years. Bergstrom et al. (20) reported a mathematical model of cycling programs suited to *S. aureus*, *Enterococcus*, and other microbes with single-drug resistance. They concluded that cycling is unlikely to reduce either the evolution or the spread of antibiotic resistance. They proposed an alternative drug use plan called mixing, in which each treated patient receives one of several drug classes used simultaneously in the hospital. At the scale relevant to bacterial populations, mixing imposes greater heterogeneity than cycling does.

***Acinetobacter*.** Evolution of microorganisms can proceed very rapidly in the ambient environment and not just in the laboratory or hospital. For example, Gram-negative *Acinetobacter* bacteria are prevalent in soils and water with only occasional infection of humans. However, in the 1980s, one species, *Acinetobacter baumannii*, emerged as a multidrug-resistant strain that contaminated field hospitals in Iraq during the first Gulf War and was introduced to U.S. hospitals by wounded U.S. Army personnel (21). This strain is highly resistant to drying and disinfectants, making decontamination difficult. Some lineages have acquired additional resistance mechanisms (22).

Effects of Immune Suppression. Another selective feature of modern society is the increasing prevalence of immunocompromised individuals as a result of HIV infection, steroid therapies, cancer chemotherapy, and various genetic immune-deficiency conditions. These individuals are highly vulnerable when hospitalized. Very little investigated is the substrate of previous chemical exposures, especially occupational exposures, that impair immune defenses and lead to pathogenic emergence of otherwise innocuous microbial agents. An example is pneumonia attributable to ordinarily saprophytic organisms in the setting of silicosis of the lung. Among genetic disorders, cystic fibrosis patients are especially susceptible to infection with *Pseudomonas* species in the lung. Cystic fibrosis is a favorite subject for speculation about what selective factors could have led to its high prevalence, including much higher prevalence in white than African-American populations. In an experimental model, mice lacking cystic fibrosis transmembrane conductance regulator (CFTR) protein did not secrete fluid in response to cholera toxin, although heterozygotes experienced 50% less fluid loss from cholera toxin than the normal mouse (23). However, the responsible chloride channel apparently is not the rate-limiting step for fluid loss in humans (15); thus, the selection-by-cholera hypothesis remains quite speculative.

Vaccines Selective for Desired Microbial Characteristics. There are many examples of pathogens increasing in virulence in response to public health interventions (24), but treating infectious diseases or preventing pathogen spread need not result in an arms race. Treatments and vaccines can be designed that select for less rather than more virulence or for more desirable characteristics. The diphtheria toxoid vaccine selects against toxin production, which is what causes disease, rather than other features of *Corynebacterium*. Thus, diphtheria infections and clinical isolations still occur, but the extant strains lack toxin production (25). Vaccination using the seven-conjugate vaccine against *Streptococcus pneumoniae* has reduced carriage of penicillin-resistant serotypes (26) but not invasive isolates (27). A better understanding of the transmission patterns of invasive isolates could enhance vaccination strategies that already select against penicillin-resistant strains.

Immunization is the most important intervention to prevent infectious diseases and improve public health. For vaccine

research and clinical usage, He et al. (28) have created a community-based vaccine ontology to standardize vaccine annotation, integrate information about vaccine types, and support computer-assisted reasoning (www.violinet.org/vaccineontology). Its literature-mining function can assist in capturing information about the evolution of the microbe and its responses to immunization and therapies.

Vector Control. Mosquitoes transmit numerous infectious diseases to humans, including dengue, yellow fever, and malaria. Disease can be prevented by immunizing or treating humans and by protecting humans from bites by infected mosquitoes with insecticide-treated bed-nets or spraying. Not surprisingly, mosquitoes have evolved resistance to insecticides. Read et al. (29) used mathematical modeling to propose a strategy to "evolution-proof" insecticides by targeting older mosquitoes, which, if infected, are more likely to have mature malaria parasites in their salivary glands ready for transmission to humans; this scheme would control disease spread and generate only weak selection for survival and reproduction by resistant mosquitoes.

Cultural Evolution—from Our Origins as Hunters and Gatherers to Contemporary Societies

Throughout 5–7 million years of human evolution, biological evolution and social evolution have been intertwined. Cultural conditions and technologies that affect our lives have been, and will be, a major driving force for biological changes in our species. One of the most remarkable examples was the beginnings of animal husbandry and agriculture 7,000–10,000 years ago; progressive domestication of sheep, goats, and cattle; and introduction of milk from animals as part of the human diet about 6,000 years ago.

Persistence of Intestinal Lactase Activity vs. Lactose Intolerance. The prominent biochemical features of milk are casein protein, calcium salts, water, and lactose (galactose-glucose disaccharide sugar). The ability to digest lactose declines rapidly in most humans after weaning because of a normal decline in the activity of the intestinal enzyme lactase. Before drinking of milk, there was no further need for this enzyme. Populations with a long history of cattle domestication and milk drinking selected for the "persistence of lactase" trait. The prevalence is >90% among northern Europeans (Swedes), ~50% in Spanish and Arab populations, 5–20% among African populations, and 1% among Chinese and Native Americans, and this is a source of health disparities in public health nutrition programs.

There are many interesting evolutionary questions about lactase persistence. How many times has a mutation occurred that was then selected positively to reach high prevalence today? How did the mutation or mutations spread globally? What is the mechanism for what seems to be a regulatory on/off mutation?

The inheritance is as an autosomal dominant gene *LCT* on chromosome 2q21, regulated in Europeans by *cis*-acting elements identified by SNPs just upstream of the *LCT* within introns of the adjacent minichromosome maintenance 6 (*MCM6*) gene (30). The SNP variant T-13910 in *MCM6* appears to be the causal variant for lactase persistence in Europeans. It most likely arose in the Middle East and spread to Northern Europe (31). Itan et al. (32) used a simulation model incorporating genetic and archaeological data to conclude that this allele arose and was selected for among dairying farmers in the central Balkans and central Europe. In Africa, high prevalences do occur in pastoralist groups like the Tutsi (90%) and Fulani (50%), but that variant is absent in nearly all other African groups studied. Working with rural population subgroups in East Africa, Tishkoff et al. (30) reported three previously undescribed variants that account for 20% of phenotypic variation, leaving ample room to discover additional variants, especially with resequencing analyses. The chromosomes with these SNP variants show strong genetic signatures of natural selection. The search for variants was enhanced by choosing to genotype individuals with extremes of plasma glucose increase after ingestion of lactose; outliers are often clues to important mechanisms and risk factors. We must always consider that a trait of interest may be adaptive for more than one reason, and may therefore be selected for some other or additional benefit to reproduction and survival. In addition to protein, calcium, and sugar, milk provides water, which is especially important in arid regions, whereas lactose intolerance leads to water loss via diarrhea.

The social and public policy context of "nutrigenomics" can be illustrated with lactose intolerance. The dairy industry has had a long-running successful campaign with "Got Milk?" advertisements. Originally, the tag line was "Everybody needs milk." The Najavo Indian Nation painted their adobes with federal surplus powdered milk; the unkind comments of outsiders reflected ignorance of the unpleasant gastrointestinal symptoms the milk caused in these people, of whom >95% were lactose-intolerant. In response to objections on behalf of nonwhite U.S. populations, the tagline was changed to "Milk has something for everybody." This case also stimulates us to realize that "the normal state" depends on time and place and environmental conditions. Factors other than the primary gene variant contribute to variation in severity of symptoms, making population testing to identify susceptible individuals before they are symptomatic much more complex. Many cases of irritable bowel syndrome might be attributable to this condition in various populations. Finally, it is interesting that certain European cat breeds have a mutation similar to that in humans and that Asian breeds are particularly intolerant of lactose—apparently reflecting coevolution with humans.

Origins and Evolution of the Western Diet: Basis for the Epidemiology of Chronic Diseases. Many of the diseases associated with contemporary Western populations, and spreading across the globe, have arisen through discordance between our ancient genetically influenced biology and the dietary, cultural, and physical activity patterns of modern societies. There is a lively literature with titles like "Stone agers in the fast lane" (33) and "When the Eskimo comes to town" (34). Food staples and food-processing procedures that were introduced during the Neolithic Period have altered fundamentally seven critical nutritional characteristics of ancient hominin diets: glycemic load, fatty acid composition, macronutrient composition, micronutrient density, acid-base balance, sodium/potassium ratio, and fiber content (35). Most of the food types that dominate present diets were introduced quite recently: dairy products, cereal grains (especially refined grains that lack germ and bran); refined sugars (especially sucrose and fructose); refined vegetable oils (with low ω-3 and high ω-6 fatty acids); alcoholic beverages; salt; and ω-6, saturated, fatty acid-rich mammalian meats. These foods have displaced the wild plant and animal foods of our predecessors. What Cordain et al. (35) call "the evolutionary collision of our ancient genome with the nutritional qualities of recently introduced foods" has contributed mightily to many chronic diseases of Western civilization: obesity, diabetes, cardiovascular disease, high blood pressure, dyslipidemias, osteoporosis, bowel disorders, inflammatory and autoimmune diseases, and several cancers. Several of these conditions are associated with insulin resistance; all remain rare among contemporary hunter-gatherer populations. Modern foods are also net acid generators, compared with net base-producing preagricultural diets. The latter are protective against osteoporosis, muscle wasting, calcium kidney stones, high blood pressure, and exercise-induced asthma. Inversion of the potassium/sodium and base/chloride ratios may cause growth retardation in children and accelerate aging (36). Modern diets are very low in potassium/sodium ratio, which

exacerbates high blood pressure, kidney stones, osteoporosis, asthma, stroke, and other conditions. Finally, modern diets are strikingly fiber-depleted, leading to many gastrointestinal disorders. With these many major changes in the diet, including the diets of children, we may expect selection to be acting on numerous gene variants, most of which may have individual small effects. For example, a variant of the gene *FTO* is associated with increased body mass index and the complex phenotype of obesity (37). An engaging account of the nature and range of modern diets is *The Omnivore's Dilemma* (38).

Genetic variants may be selected positively or negatively to maintain traits that are the optimal average for a population with a stable environment or to move the average genome directionally to match permanently altered aspects of the environment. Changes that began even 10,000 years ago may be too recent to have reached an equilibrium of adaptation; the discordance emerges as diseases.

Evolution of the Thrifty Genotype in Relation to Diabetes. The human behavior of eating regular meals is itself a significant evolutionary change that contributes to our increased consumption of calories. Our "obesogenic environment" produces a mismatch between our evolutionary health status as a hunter-gatherer and present-day life, with many obesity-related diseases (39). Adipose tissue has emerged as an endocrine organ, secreting many hormones and peptides that control eating, metabolism, and storage of excess fat. The phrase "thrifty genotype" was introduced by human genetics pioneer James Neel (40) to describe the benefit of a sustained hyperglycemic response after an occasional hefty meal by hunter-gatherers. That sustained hyperglycemic response is associated now with peripheral resistance to insulin action and the development of diabetes and its complications in the kidney, nerves, arteries, and retina. Native-American populations vary notably in the prevalence and severity of diabetes and diabetic complications, a fertile subject for evolution-based clinical/translational research.

Diabetes also makes people more vulnerable to infections, partly through accumulation of reactive oxygen species, which require effective immune and inflammatory responses and antioxidants to overcome their effects. As Nesse and Stearns (15) have emphasized, our evolutionary legacy is a broad array of symptoms, defense mechanisms, and molecules that may have both protective and damaging features. A striking example is bilirubin, the end product of heme metabolism, which is neurotoxic at high concentrations, especially in infancy. Why, they ask, does the body make such a difficult-to-excrete toxin? It turns out that bilirubin is an effective antioxidant, which may help to delay atherosclerosis and aging. Lipophilic bilirubin and water-soluble glutathione have complementary antioxidant and cytoprotective roles (41). Bilirubin functions and is consumed at a concentration of 10 nM. Evolution has provided a steady source of intracellular bilirubin through the biliverdin reductase cycle, which amplifies bilirubin levels 10,000-fold. A clue comes from the benign genetic disorder Gilbert syndrome, a conjugation enzyme deficiency characterized by increased bilirubin levels, with 6-fold lower rates of heart disease and a 3-fold lower risk for carotid plaques. Sedlak et al. (41) claim that elevated bilirubin is a better index of disease protection than HDL-cholesterol. β-Carotene is another antioxidant that was proposed as a cancer chemopreventive agent, but it turned out to be carcinogenic (42).

Origins of Racial Differences in Human Populations

Racial and ethnic differences have evolved through natural selection in adaptation to different environmental conditions, combined with reproductive isolation. During the most recent glacial period about 100,000 years ago, much of the earth's expanse was covered by ice, providing conditions for separate evolution of whites in the west, mongoloid populations in the east, and blacks in the south (1). Much migration has occurred since then, with admixture of genes. The most conspicuous racial difference is skin pigmentation. An obvious question is why are whites and Asians so lightly pigmented? A plausible hypothesis involves the adaptation in their latitudes to low levels of the UV radiation necessary for conversion of provitamin D to vitamin D in skin. Activated vitamin D is essential for proper calcification of the bones and avoidance of rickets in childhood. Furthermore, rickets in women impairs childbirth through pelvic deformation, leading to death of mothers and infants under primitive medical conditions, a strong selective pressure. An interesting experimental test of this explanation was performed with saddle pigs, which are darkly pigmented in midbody and little pigmented in the dorsal and caudal regions; vitamin D formation after UV irradiation was shown to be greater in the unpigmented areas of skin. Exceptions may be instructive, too, specifically Eskimos and African Pygmies. They experience little UV irradiation in arctic regions and under the tropical rain forest canopy, respectively. Eskimos get activated vitamin D from fish and seal liver, whereas pygmies may get theirs from insect larvae in their diet (1). A gene–environment interaction involving the polymorphic β-2 serum protein Gc may be explained similarly because Gc2 is a more effective carrier protein for vitamin D than Gc1. Vitamin D deficiency is of high epidemiological interest because of increased colon cancer and heart disease risks; recently, the American Association for Clinical Chemistry reported a large increase in testing for vitamin D and its active metabolites.

In a HapMap analysis (43), some of the strongest signals of recent selection appear in five unlinked genes involved in skin pigmentation in Europeans (OCA2, MYO5A, DTNBP1, TYRP1, and SLC24A5), consistent with separate selective events. Embedded in SNP, haplotype, and sequencing study results are ample markers to assess population origins so that subjectively identified race need no longer be a confounding variable in the analysis.

Emerging Topics with Evolutionary Implications

Revealing Natural Selection Potentially Important to Public Health Through Genome Mapping Studies. Genotyping and high-throughput genome sequencing are rapidly producing huge files of data that can guide studies of ongoing evolution relevant to public health. Voight et al. (43) published an analytical method for genome-wide scanning for SNPs that may be signals of recent selection. Their goal was to identify loci in which strong selection has driven mutant alleles to intermediate prevalence—on their way to fixation or to a balanced polymorphism. The key signal of strong directional selection is that the favored allele tends to sit on an unusually long haplotype of low diversity/high homozygosity attributable to a relatively fast increase in prevalence. Windows of consecutive SNPs that contain multiple extreme scores represent clusters attributable to "selection sweeps"; selection coefficients of 0.01–0.04 are sufficient to produce major regional population differences since the separation of African and Eurasian populations about 6,600 years ago (260 generations). The lactase region on chromosome 2 (in Europeans) and the alcohol dehydrogenase (ADH) cluster on chromosome 4 (in East Asians) were confirmed as highly selected by this method. In their application of tag-SNPs to the three-continent HapMap dataset, most signals, but not all, were specific to a geographic region subpopulation, consistent with emergence since the separation of these populations. Because genetic variants have different fitness, they should be loci (or should be in linkage disequilibrium with loci) that contribute significant phenotypic variation, possibly for complex traits and diseases. Among genes showing evidence of sweeps, enrichment was found for the gene ontology categories chemosensory perception, olfaction, gametogenesis, spermatogenesis, fertilization, carbohydrate/lipid/steroid/phosphate metabolism, electron transport, chromatin packaging/remodeling, MHC-1-mediated immunity, peroxisome transport, and vitamin transport (table 2 in ref. 43). This approach is a departure from the candidate gene approach reflected in the CCR5/HIV, HbS/malaria, and lactase persistence/

lactose tolerance examples presented above. Behaviorally, modern human populations have experienced tremendous shifts in habitats, food sources, population densities, and pathogen exposures for long enough to produce selection of certain genes associated with specific complex trait phenotypes. Good examples are CYP3A5 and salt-sensitive high blood pressure, ADH and alcoholism susceptibility, and 17q21 inversion and fertility (43).

It should be useful to combine these findings with genome-wide association studies (GWASs). There is a huge GWAS literature from the past few years, with numerous genomic variants associated with common traits and complex diseases. However, few of the variants are in protein-coding genes with identifiable functional consequences. Moreover, few studies have any assessment of exposures, which is essential for discovery of gene–environment interactions and identification of modifiable risk factors (44). Interactions of individual SNPs with environmental exposures have been reported. *N*-acetyl-transferase 2 (NAT2) genotypes are associated with differential detoxification of arylamines in dye industry occupational exposures and in tobacco smoke, leading to differential risk for cancer of the urinary bladder (45). Interestingly, *O*-acetylation by the same enzyme activates heterocyclic amines that lead to colorectal cancer; only weak main effects of well-done meat consumption (a source of heterocyclic amines), the genes *CYP1A2* and *NAT2* that are involved in their metabolism, or tobacco smoking (which can induce *CYP1A2*) were found for colorectal cancer, but a very high odds ratio of 8.8 was found for those who were both exposed and genetically susceptible, with no significant lower order interactions (46). A conceptual model for the role of genes involved in DNA damage response pathways for double-strand breaks caused by ionizing radiation is the basis for the Women's Environment, Cancer and Radiation Epidemiology (WECARE) study of second breast cancers after radiotherapy of primary breast cancers (47). A special symposium of the 2010 *Annual Review of Public Health* is devoted to genomics and public health, including articles on statistical (48) and epidemiological (49) methods to enhance GWASs.

Two central challenges in evolutionary biology are to understand the genetic and ecological mechanisms that drive adaptation and to recognize the effects of natural selection on a dynamic background of neutral processes of population history, bottlenecks, migration, mutation, recombination, and drift. Coop et al. (50) examined the role of geography and population history in the spread of selectively favored alleles using the Human Genome Diversity Panel of the Centre d'Étude du Polymorphisme Humain (Paris) (CEPH) and the Phase II HapMap. Over the past 50,000–100,000 years, humans have spread out from Africa to colonize essentially the entire planet, thereby experiencing a vast range of climates, diets, and environments as likely selective factors, together with sexual competition, viability selection, and resistance to evolving pathogens on an ongoing basis. Strong evidence of relatively recent adaptation by selection has emerged from haplotype sweep patterns of clusters of SNPs, homozygosity for extended distances, and selection coefficients >1% sustained for long periods for genes involved in resistance to malaria (G6PD and Duffy antigen), lighter skin pigmentation in non-Africans (SLC24A5, SLC45A2, KITLG, and EDAR), and diet and metabolism (lactase and salivary amylase). However, for most high-frequency SNPs that show extreme differentiation between pairs of the three Eurasian, East Asian, or African populations, geographic associations and neutral processes of ancestral relationships and migration still may be largely responsible for the local differences within regions (50).

The *P. falciparum* and *P. vivax* malaria examples (above) show how population genetics helps to reveal the evolution of parasite-host relationships. Population genetics is advancing remarkably through the elucidation of the human genome sequence and the development of high-throughput methods for SNPs, haplotypes, next-generation sequencing of exons, and, soon, the whole genome. A fascinating discovery with phylogenetic analysis has just appeared that radically revises our thinking about the origin of *P. falciparum* (11). For at least 15 years, the evidence pointed to cospeciation of *P. falciparum* in humans and *Plasmodium reichenowi* in chimpanzees, evolved separately from a presumed common ancestor over 5–7 million years in parallel to divergence of their hosts. That was based on one isolate of *P. reichenowi*. With eight previously undescribed isolates, Rich et al. (11) showed that the global totality of *P. falciparum* strains is fully included within the much more diverse *P. reichenowi* variation. All extant *P. falciparum* populations seem to have originated from the parasite infecting chimpanzees by a single-host transfer, possibly as recently as 10,000 years ago. Furthermore, two critical genetic mutations have been elucidated. Inactivation of the gene *CMAH* in the human lineage blocked conversion of sialic acid neuraminidase 5Ac to Neu5Gc, making humans resistant to *P. reichenowi*. In addition, mutations in the dominant invasion receptor EBA 175 made *P. falciparum* prefer the overabundant Neu5Ac precursor. This combination may explain the extreme pathogenicity of *P. falciparum* in humans.

Global Climate Change. Global average surface temperatures have increased 0.8°C (1.4°F) over the past century, mostly in the past 30 years, with a "commitment" to further increases from carbon dioxide already accumulated in the atmosphere (51). Looking forward, we can expect an increasing focus on modeling and predicting coevolution of humans and many relevant plant, microbial, invertebrate, and vertebrate species under the selective forces of global climate change and our attempts to mitigate and adapt to climate change. It is feasible to model and project geographic shifts with temperature and humidity for agriculture and for vector-borne diseases. The Arctic is particularly susceptible to climate change, with warming occurring at a rate twice that of moderate zones, leading to striking changes in the forests and viability of crops, appearance of unfamiliar insects and microbes, and thinning and breakup of the arctic ice. We can anticipate progressive major changes in temperature, humidity, habitats, vectors, and transmission for a host of infectious agents (52). A possible example is the appearance of the fungal pathogen *Cryptococcus gattii* in 1999 in the Pacific Northwest, with 200 human and 400 domestic animal cases now reported in normal individuals; previously, human cases occurred only rarely, mostly among immunocompromised patients.

Seasonality is another important climate variable for infectious diseases, recognized since the time of Hippocrates. Seasonality produces alternating periods of high transmission and population bottlenecks that limit strain diversity and cause rapid genetic shifts. Models show that small seasonal changes in host–pathogen dynamics, including host social behavior and contact rates (53), may be sufficient to create large seasonal surges in disease incidence, with exacerbations likely to arise from climate change (54).

The term *prevention* in public health is analogous to *adaptation* in the climate change literature, ranging from reduction of greenhouse gas emissions to redesign of cities to minimize heat islands and heat waves, to surveillance for diseases like tick-borne Lyme disease, and to mitigation of health disparities in human impacts of rising sea and river levels. Public health professionals will be critical to adaptation strategies, hopefully informed by an evolutionary perspective about the interrelations between living things and their environments.

Human Behavioral Phenotypes. Subjects involving a broad range of normal behaviors and mental illnesses are important in community public health and for initiatives to stimulate healthier choices in personal behavior. For example, genetic variation and evolutionary psychology may help to reveal underlying neural and social determinants of personality traits (55). Darwin was well aware of genetic influences on behavior, as reflected in his

writing on the domestication of animals and comments on the distinctive mental qualities of dogs, horses, and other animals. The dog genome is under study, in part, because of the dramatic differences in behavioral traits among breeds. Humans and other mammals share basic emotion–motivation phenotypes of anger, fear, nurturance, curiosity, and sex-related behaviors shaped during human evolution by social adaptations.

A particularly interesting evolutionary perspective has been applied to uses and effects of psychoactive drugs (56). The use of pure psychoactive chemical agents as drugs and the i.v. and nasal routes of administration are specific evolutionary features of the contemporary human environment. They are "inherently pathogenic" because they bypass adaptive information processing systems and act directly on brain mechanisms that control emotion and behaviors. Drugs of abuse create signals in the brain that indicate falsely the arrival of a fitness benefit such that drug-seeking behavior displaces adaptive behaviors. Video game-playing and snacks high in fat, salt, and sugar were described in similar terms (56). Drugs that block anxiety, low mood, and other negative emotions can be analyzed by analogy to drugs that alter pain, cough, fever, diarrhea, vomiting, and related physical defense mechanisms.

Closing Comment

As illustrated in this article, important examples of practical applications of evolutionary understanding in modern public health include obesity, influenza, and appropriate uses of antibiotics. As documented by the Surgeon General and the CDC, the prevalence of obesity and overweight has increased sharply in the past 30 years, with huge consequences for the burden of chronic diseases and health care costs. The global epidemic of obesity represents a combination of rapidly changing, culture-based, behavior changes and discordant genomic predispositions that cannot be ignored. Meanwhile, we face simultaneous influenza epidemics from seasonal H1N1 strains, with enormous annual variation in impact, from H1N1/2009 swine flu strains with very differential susceptibilities in the human population, and therefore quite different high-risk target populations for prevention, vaccination, and therapy, and, lest we forget, from lingering H5N1 avian flu strains. Finally, we remain in an arms race with bacteria whose environments inside us and around us we are constantly changing, stimulating their own rapid evolution.

Evolution, natural selection, and population dynamics act over very long periods of time. We have learned in recent years, as highlighted by *Science* magazine's "Breakthrough of the Year 2005," that we can actually observe "evolution in action"—in the Galapagos, the Arctic and Antarctic, hospitals, rural and urban waste streams, and many other settings, with impacts on public health and implications for our public health research agenda. We can be confident that evolutionary perspectives will provide a useful foundation for research and communication in public health (57) as well as in medical care (58).

ACKNOWLEDGMENTS. I am grateful to my colleagues Professors Betsy Foxman, Randy Nesse, and Noah Rosenberg for specific inputs for this manuscript.

1. Vogel F, Motulsky AG (1997) *Human Genetics: Problems and Approaches* (Springer, Berlin), 3rd Ed.
2. Ewald PW (2004) Evolution of virulence. *Infect Dis Clin North Am* 18:1–15.
3. Allison AC (1954) Protection afforded by sickle-cell trait against subtertian malareal infection. *Br Med J* 1:290–294.
4. Haldane J (1949) The rate of mutations of human genes. *Hereditas* 35(Suppl): 267–273.
5. Motulsky AG (1964) Hereditary red cell traits and malaria. *Am J Trop Med Hyg* 13 (Suppl):147–158.
6. Gelehrter T, Collins F, Ginsburg D (1998) *Principles of Medical Genetics* (Williams & Wilkins, Baltimore), 2nd Ed, p 51.
7. Heeney JL, Dalgleish AG, Weiss RA (2006) Origins of HIV and the evolution of resistance to AIDS. *Science* 313:462–466.
8. Lim JK, Glass WG, McDermott DH, Murphy PM (2006) CCR5: No longer a "good for nothing" gene—Chemokine control of West Nile virus infection. *Trends Immunol* 27: 308–312.
9. Nahmias AJ, et al. (1986) Evidence for human infection with an HTLV III/LAV-like virus in central Africa, 1959. *Lancet* 1:1279–1280.
10. Contreras-Galindo R, et al. (2008) Human endogenous retrovirus K (HML-2) elements in the plasma of people with lymphoma and breast cancer. *J Virol* 82:9329–9336.
11. Rich SM, et al. (2009) The origin of malignant malaria. *Proc Natl Acad Sci USA* 106: 14902–14907.
12. Wolfe ND, et al. (1998) Wild primate populations in emerging infectious disease research: The missing link? *Emerg Infect Dis* 4:149–158.
13. Turnbaugh PJ, et al. (2007) The human microbiome project. *Nature* 449:804–810.
14. Elliott DE, Summers RW, Weinstock JV (2007) Helminths as governors of immune-mediated inflammation. *Int J Parasitol* 37:457–464.
15. Nesse RM, Stearns SC (2008) The great opportunity: Evolutionary applications to medicine and public health. *Evolutionary Applications* 1:28–48.
16. McCarter JP (2008) Nematology: Terra incognita no more. *Nat Biotechnol* 26:882–884.
17. Lax E (2004) *The Mold in Dr. Florey's Coat: The Story of the Penicillin Miracle* (Henry Hold, New York).
18. Pearson ML, et al. (1992) Nosocomial transmission of multidrug-resistant Mycobacterium tuberculosis. A risk to patients and health care workers. *Ann Intern Med* 117:191–196.
19. Klevens RM, et al; Active Bacterial Core surveillance (ABCs) MRSA Investigators (2007) Invasive methicillin-resistant Staphylococcus aureus infections in the United States. *JAMA* 298:1763–1771.
20. Bergstrom CT, Lo M, Lipsitch M (2004) Ecological theory suggests that antimicrobial cycling will not reduce antimicrobial resistance in hospitals. *Proc Natl Acad Sci USA* 101:13285–13290.
21. Scott P, et al. (2007) An outbreak of multidrug-resistant Acinetobacter baumannii-calcoaceticus complex infection in the US military health care system associated with military operations in Iraq. *Clin Infect Dis* 44:1577–1584.
22. Dijkshoorn L, Nemec A, Seifert H (2007) An increasing threat in hospitals: Multidrug-resistant Acinetobacter baumannii. *Nat Rev Microbiol* 5:939–951.
23. Gabriel SE, Brigman KN, Koller BH, Boucher RC, Stutts MJ (1994) Cystic fibrosis heterozygote resistance to cholera toxin in the cystic fibrosis mouse model. *Science* 266:107–109.
24. Dronamraju K, ed (2004) *Infectious Disease and Host-Pathogen Evolution* (Cambridge Univ Press, Cambridge, UK).
25. Soubeyrand B, Plotkin SA (2002) Microbial evolution: Antitoxin vaccines and pathogen virulence. *Nature* 417:609–610, discussion 610.
26. Dagan R (2009) Impact of pneumococcal conjugate vaccine on infections caused by antibiotic-resistant Streptococcus pneumoniae. *Clin Microbiol Infect* 15(Suppl 3): 16–20.
27. Karnezis TT, Smith A, Whittier S, Haddad J, Jr, Saiman L (2009) Antimicrobial resistance among isolates causing invasive pneumococcal disease before and after licensure of heptavalent conjugate pneumococcal vaccine. *PLoS One* 4:e5965.
28. He Y, et al. (2009) VO: Vaccine ontology. *The First International Conference on Biomedical Ontology (ICBO 2009),* ed Smith B (National Center for Ontological Research, Buffalo, NY), p 172.
29. Read AF, Lynch PA, Thomas MB (2009) How to make evolution-proof insecticides for malaria control. *PLoS Biol* 7:e1000058.
30. Tishkoff SA, et al. (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* 39:31–40.
31. Hollox EJ, Swallow DM (2002) In *The Genetic Basis of Common Diseases*, eds King RA, Rotter JI, Motulsky AG (Oxford Univ Press, Oxford), pp 250–265.
32. Itan Y, Powell A, Beaumont MA, Burger J, Thomas MG (2009) The origins of lactase persistence in Europe. *PLOS Comput Biol* 5:e1000491.
33. Eaton SB, Konner M, Shostak M (1988) Stone agers in the fast lane: Chronic degenerative diseases in evolutionary perspective. *Am J Med* 84:739–749.
34. Schaefer O (1971) When the Eskimo comes to town. *Nutr Today* 6 (Nov–Dec):8–16.
35. Cordain L, et al. (2005) Origins and evolution of the Western diet: Health implications for the 21st century. *Am J Clin Nutr* 81:341–354.
36. Frassetto L, Morris RC, Jr, Sellmeyer DE, Todd K, Sebastian A (2001) Diet, evolution and aging—The pathophysiologic effects of the post-agricultural inversion of the potassium-to-sodium and base-to-chloride ratios in the human diet. *Eur J Nutr* 40: 200–213.
37. Cecil JE, Tavendale R, Watt P, Hetherington MM, Palmer CNA (2008) An obesity-associated FTO gene variant and increased energy intake in children. *N Engl J Med* 359:2558–2566.
38. Pollan M (2006) *The Omnivore's Dilemma: A Natural History of Four Meals* (Penguin Press, Penguin Books Ltd., New York).
39. Power ML, Schulkin J (2009) *The Evolution of Obesity* (Johns Hopkins Univ Press, Baltimore).
40. Neel JV, Weder AB, Julius S (1998) Type II diabetes, essential hypertension, and obesity as "syndromes of impaired genetic homeostasis": The "thrifty genotype" hypothesis enters the 21st century. *Perspect Biol Med* 42:44–74.
41. Sedlak TW, et al. (2009) Bilirubin and glutathione have complementary antioxidant and cytoprotective roles. *Proc Natl Acad Sci USA* 106:5171–5176.
42. Omenn GS, et al. (1996) Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. *N Engl J Med* 334:1150–1155.
43. Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. *PLoS Biol* 4:e72.
44. Omenn GS (2009) From human genome research to personalized health care. *Issues Sci Technol* 25:51–56.

45. Vineis P, et al. (2001) Current smoking, occupation, N-acetyltransferase-2 and bladder cancer: A pooled analysis of genotype-based studies. *Cancer Epidemiol Biomarkers Prev* 10:1249–1252.
46. Le Marchand L, et al. (2001) Combined effects of well-done red meat, smoking, and rapid N-acetyltransferase 2 and CYP1A2 phenotypes in increasing colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev* 10:1259–1266.
47. Bernstein JL, et al. (2004) Study design: Evaluating gene-environment interactions in the etiology of breast cancer—The WECARE study. *Breast Cancer Res* 6 (Suppl): R199–R214.
48. Thomas D (2010) In *Annual Review of Public Health*, ed Fielding JE (Annual Reviews, Palo Alto), Vol 31.
49. Witte JS (2010) In *Annual Review of Public Health*, ed Fielding JE (Annual Reviews, Palo Alto), Vol 31.
50. Coop G, et al. (2009) The role of geography in human adaptation. *PLoS Genet* 5: e1000500.
51. Ebi KL, Semenza JC (2008) Community-based adaptation to the health impacts of climate change. *Am J Prev Med* 35:501–507.
52. Last JM (1993) Global change: Ozone depletion, greenhouse warming, and public health. *Annu Rev Public Health* 14:115–136.
53. Altizer S, et al. (2006) Seasonality and the dynamics of infectious diseases. *Ecol Lett* 9: 467–484.
54. Fisman DN (2007) Seasonality of infectious diseases. *Annu Rev Public Health* 28: 127–143.
55. Bouchard TJ, Jr, Loehlin JC (2001) Genes, evolution, and personality. *Behav Genet* 31: 243–273.
56. Nesse RM, Berridge KC (1997) Psychoactive drug use in evolutionary perspective. *Science* 278:63–66.
57. Nesse RM, Stearns SC, Omenn GS (2006) Medicine needs evolution. *Science* 311:1071.
58. Nesse RM, et al. Making evolutionary biology a basic science for medicine. *Proc Natl Acad Sci USA* 107(Suppl):1800–1807.

Genetic variation in human telomerase is associated with telomere length in Ashkenazi centenarians

Gil Atzmon[a,b,1,2], Miook Cho[a,1], Richard M. Cawthon[c], Temuri Budagov[b], Micol Katz[b], Xiaoman Yang[b], Glenn Siegel[b], Aviv Bergman[d], Derek M. Huffman[a,b], Clyde B. Schechter[e], Woodring E. Wright[f], Jerry W. Shay[f], Nir Barzilai[a,b], Diddahally R. Govindaraju[g], and Yousin Suh[a,b,2]

[a]Departments of Medicine and Genetics, Albert Einstein College of Medicine, Bronx, NY 10461; [b]Institute for Aging Research, Diabetes Research and Training Center, Albert Einstein College of Medicine, Bronx, NY 10461; [c]Department of Human Genetics, University of Utah, Salt Lake City, UT 84112; [d]Department of Systems and Computational Biology, Albert Einstein College of Medicine, Bronx, NY 10461; [e]Department of Family and Social Medicine, Albert Einstein College of Medicine, Bronx, NY 10461; [f]Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390; and [g]Department of Neurology, Boston University School of Medicine, Boston, MA 02118

Edited by Stephen Curtis Stearns, Yale University, New Haven, CT, and accepted by the Editorial Board October 7, 2009 (received for review July 20, 2009)

Telomere length in humans is emerging as a biomarker of aging because its shortening is associated with aging-related diseases and early mortality. However, genetic mechanisms responsible for these associations are not known. Here, in a cohort of Ashkenazi Jewish centenarians, their offspring, and offspring-matched controls, we studied the inheritance and maintenance of telomere length and variations in two major genes associated with telomerase enzyme activity, *hTERT* and *hTERC*. We demonstrated that centenarians and their offspring maintain longer telomeres compared with controls with advancing age and that longer telomeres are associated with protection from age-related diseases, better cognitive function, and lipid profiles of healthy aging. Sequence analysis of *hTERT* and *hTERC* showed overrepresentation of synonymous and intronic mutations among centenarians relative to controls. Moreover, we identified a common *hTERT* haplotype that is associated with both exceptional longevity and longer telomere length. Thus, variations in human telomerase gene that are associated with better maintenance of telomere length may confer healthy aging and exceptional longevity in humans.

longevity | heritability | aging | biomarker

Telomeres consist of the TTAGGG tandem repeats at the ends of chromosomes and are known to protect these regions from degradation and DNA repair activities (1). Telomeres progressively shorten with each cell division in cultured primary human cells (2) until a critically shortened length is achieved, upon which the cells enter replicative senescence (3). Although the relevance of replicative senescence to in vivo aging remains poorly understood, numerous reports suggest that telomere shortening may be associated with organismal aging, with concomitant metabolic decline and increased risk for disease and death (4, 5). For example, several cross-sectional studies in humans have shown that telomere length in white blood cells is inversely related to the age of the cell donor (6–9). Likewise, shorter telomere length has been shown to be associated with age-related disease including coronary heart disease, hypertension, and dementia, as well as general risk factors for disease such as insulin resistance and obesity (10). Furthermore, oxidative stress and inflammation, two major postulated causal factors of aging, are known to accelerate telomere shortening, suggesting that telomere length may be an important biomarker of aging because it reflects the cumulative burden of oxidative stress and inflammation (4, 11). In addition, most (5, 12, 13) but not all studies (14, 15) have shown a positive association between telomere length and overall survival in humans. These results indicate that telomere shortening could be used as a biomarker of disease risk and progression as well as early mortality. However, biological mechanisms responsible for these associations are not known.

Telomere length varies among individuals and families and follows the polygenic mode of inheritance pattern typical of most quantitative traits (16). Heritability estimates for telomere length vary from 35 to 80% (9, 13, 17). Although several candidate genes have been identified as potential modulators of telomere length in humans (17, 18), none of these genes seem to play a direct role in maintenance of telomere length (19). Recently, one of the most obvious candidate genes of telomere maintenance, telomerase, has been shown to play a direct role in the maintenance of telomere length in humans (20). Telomerase is a specialized ribonucleoprotein enzyme complex that adds telomere repeats to the ends of chromosomes and has two essential components: a catalytic component encoded by the human telomerase reverse transcriptase (hTERT) and a human telomerase RNA component (hTERC). The latter component provides the template for nucleotide addition by hTERT. Heterozygote mutations in the *hTERT* and *hTERC* genes lead to short telomeres and are the major risk factors for rare hematopoietic disorders of bone marrow failure, including aplastic anemia and dyskeratosis congenital. These results indicate that the levels of functional telomerase are critical for telomere maintenance (21). Telomerase is expressed at high levels in specific germline cells, proliferating stem-like cells, and many cancers, whereas in normal adult cell types, it is either not expressed or is expressed at very low levels that are not sufficient to maintain telomere length. However, telomerase can be unregulated in these cells under certain conditions to maintain telomere length (22). This suggests that efficient regulation of telomerase gene expression in response to stresses that are known to reduce telomere length such as oxidative damage or inflammation would lead to better telomere maintenance.

We have previously shown that individuals in Ashkenazi families with exceptional longevity have generally been spared major age-related diseases such as cardiovascular disease and diabetes mellitus, which are largely responsible for mortality in the elderly, and that these features are heritable (23). Because studies on individuals with a normal life span suggest that

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Evolution in Health and Medicine" held April 2–3, 2009, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS web site at www.nasonline.org/Sackler_Evolution_Health_Medicine.

Author contributions: G.A., R.M.C., N.B., and Y.S. designed research; G.A., M.C., R.M.C., T.B., X.Y., G.S., D.M.H., and Y.S. performed research; G.A., M.C., M.K., A.B., D.M.H., C.B.S., W.E.W., J.W.S., and Y.S. analyzed data; and G.A., M.C., M.K., A.B., J.W.S., N.B., D.R.G., and Y.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. S.C.S. is a guest editor invited by the Editorial Board.

[1]G.A. and M.C. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: gil.atzmon@einstein.yu.edu or yousin.suh@einstein.yu.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0906191106/DCSupplemental.

common age-related diseases and a shorter life span are associated with shorter telomeres (24–28), we assessed the telomere length in blood leukocytes among subjects with exceptional longevity (centenarians) to investigate if centenarians survived with long telomeres, an indicator of better telomere length maintenance, or short telomeres reflecting chronologically old age. In addition, in view of a strong family history of exceptional longevity and genetic basis of telomere length maintenance, we tested if telomere length as a trait would be inherited by the offspring of centenarians and associated with the status of aging-related diseases. Finally, to study the genetic basis of telomere length maintenance, we took a candidate gene approach to discover all possible variations in two main genes associated with telomerase enzyme activity, *hTERT* and *hTERC*, in centenarians and younger controls for genetic association analysis with human exceptional longevity and telomere length. Our overall hypothesis is that maintenance of telomere length may indicate generalized genomic integrity, which, in turn, may have a profound influence on health and aging in humans.

Results

Telomere Length, Heritability, and Longevity. Phenotypes of centenarians, their offspring (approximate age of 70 years), and offspring-matched controls without a family history of unusual longevity, all of Ashkenazi Jewish descent, are presented in Table 1. The control and offspring groups have similar characteristics of lipid profiles, except that controls have a slightly smaller percentage of large LDL particle size, but these differences are not significant (Table 1). The centenarian group displays a lower mean body mass index and higher HDL levels as compared with controls (Table 1).

When controls and centenarians are combined to represent cross-sectional age groups of unrelated individuals from middle age to oldest old (range: 43–105 years), telomere length declines with age until the age of 85 years (Fig. 1). However, individuals older than 86 years of age as well as the centenarians show longer telomere lengths compared with unrelated individuals younger than 85 years of age, with a mean difference of 0.17 ($P = 0.04$) in the linear regression model adjusting for age and gender. The telomere length of offspring exceeds that of controls in a hierarchical linear model, on average by 0.082 [95% confidence interval (CI): 0.017–0.148; $P = 0.014$]. Because centenarians and their offspring are appreciable singleton families, we used the regression approach for nonindependencies of the observed subjects. To account for parametric model bias as well as for unknown

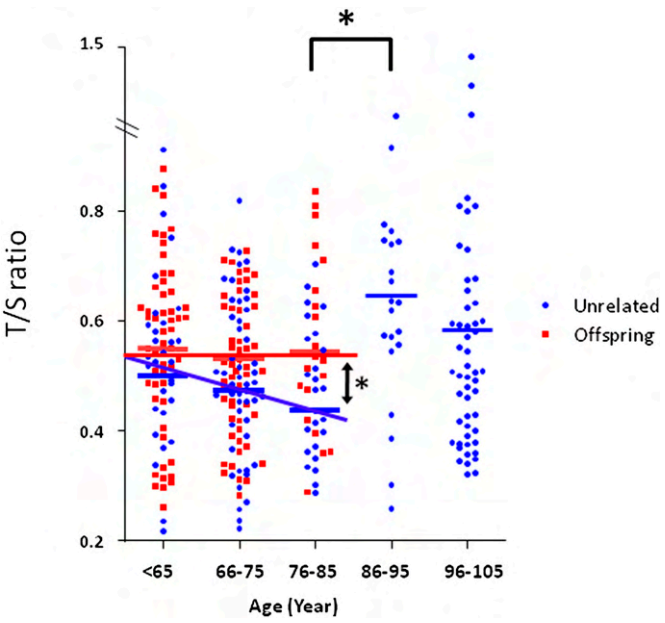


Fig. 1. Comparison of telomere length among Ashkenazi Jewish centenarians ($n = 86$), their offspring ($n = 175$), and controls ($n = 93$). Values are adjusted for age at recruitment and gender in the offspring and control groups and for gender alone in the centenarian group. *$P < 0.05$.

intrafamily correlation, we used a sandwich variance estimator equation and a generalized estimating equation (GEE), respectively. Both analyses repeated our first estimation supporting the observation of significant telomere length differences between offspring and controls, on average, by 0.081 (95% CI: 0.022–0.139; $P = 0.007$) for the sandwich estimation and by 0.082 (95% CI: 0.017–0.147; $P = 0.013$) for the GEE. Moreover, unlike controls, offspring of centenarians do not show an appreciable decline in telomere length with age as compared with controls (Fig. 1). However, the difference in age-related decline did not differ significantly between offspring and controls (adjusted mean difference = 0.002, 95% CI: −0.007 to +0.010; $P = 0.682$) by the same hierarchical linear model described earlier. High heritability (86%; $P = 0.005$) for telomere length between centenarians and their offspring was observed, suggesting that telomere length has a strong genetic component in families with exceptional longevity

Table 1. Phenotypic characteristics of study subjects for telomere length measurements

				P values	
	Controls ($n = 93$)	Offspring ($n = 175$)	Centenarians ($n = 86$)	Offspring vs. controls	Centenarians vs. controls
Age range, years	43–94 (71.8)	44–85 (67.8)	95–105 (97.4)	0.0001	0.0001
Female, %	55	36	63	0.001	0.24
Body mass index, kg/m^2	25 (0.4)	26 (0.3)	22 (0.4)	0.28	0.0001
Waist circumference, inch	34 (0.4)	35 (0.3)	35 (0.5)	0.22	0.57
Cholesterol, mg/dL	197 (3.8)	202 (2.9)	191 (4)	0.28	0.26
Triglyceride, mg/dL	154 (8.9)	146 (6.7)	151 (9.6)	0.52	0.81
LDL, mg/dL	107 (3.5)	112 (2.6)	106 (3.6)	0.22	0.76
HDL, mg/dL	58 (1.7)	61 (1.3)	51 (1.8)	0.21	0.003
LDL particle size, nm	20.9 (0.12)	21.1 (0.08)	21.2 (0.1)	0.29	0.06
HDL particle size, nm	9.23 (0.07)	9.15 (0.05)	9.36 (0.06)	0.37	0.17
Large LDL particle size, % of total	52 (4.4)	62 (3.0)	62 (3.9)	0.07	0.1
Large HDL particle size, % of total	53 (2.3)	50 (1.6)	55 (2.1)	0.26	0.42

The phenotypic characteristics were adjusted for age and gender when controls were compared with offspring and for gender when controls were compared with centenarians. All values are mean (SE).

and that these families have better telomere length maintenance than controls.

Telomere Length, Age-Associated Diseases, and Lipid Profiles of Healthy Aging. Association analysis between telomere length and major age-related diseases among centenarians, their offspring, and controls indicated that significantly shorter telomere lengths (adjusted for age, gender, and group) are present in subjects with hypertension ($P = 0.006$), the metabolic syndrome ($P = 0.03$), or diabetes ($P = 0.03$) compared with subjects without these disorders (Fig. 2). Moreover, when we tested the relation between telomere length and cognitive function among centenarians, centenarians with impaired cognitive function [Mini-Mental State Examination (MMSE) score ≤25] have significantly shorter telomeres as compared with centenarians with normal cognitive function after adjustment for age and gender (Fig. 2; $P = 0.02$). We then tested whether telomere length correlated with the lipid profiles, which are known predictors of aging-related diseases such as coronary artery disease and metabolic syndromes (29). These analyses revealed that telomere length is associated with the lipid profiles of healthy aging, showing a positive association with increased particle sizes of LDL ($P = 0.0001$) and HDL ($P = 0.03$), percentage of large particle sizes of LDL ($P = 0.0002$) and HDL ($P = 0.038$), and levels of Apo-A1 ($P = 0.005$) and HDL ($P = 0.04$) (Table 2), whereas there is a negative association with very LDL levels ($P = 0.008$). Total cholesterol, triglycerides, LDL, and Apo-B levels show no significant correlation (Table 2). These results suggest that longer telomeres are positively correlated with the lipid profiles of healthy aging.

Variation in the Telomerase Genes, Longevity, and Telomere Length. We performed a comprehensive sequence analysis of *hTERT* and *hTERC* genes to detect all possible genetic variants, including the rare variants that may be enriched in centenarians, throughout the coding exons and exon–intron flanking regions in centenarians ($n = 100$) and controls ($n = 80$) using 2D gene scanning and DNA sequencing (Fig. 3). We found a total of 15 sequence variants in the *hTERT* gene, among which 5 were previously unknown unique variants that were not reported in various SNP databases (Table 3). The locations of these variants are shown in Fig. 4. Nine variants in the *hTERT* gene are in the coding region, including two nonsynonymous variants, 893 G > A (Ala-279 Thr) and 3242 G > A (Ala-1062 Thr), which are rare and found only in controls (Table 3). In silico analysis using SIFT (sorts intolerant from tolerant substitutions) (30) and PolyPhen (31), which predict the effects of amino acid changes on protein function, indicates that these changes are likely to be benign. Interestingly, rare synonymous or intronic variants in *hTERT* are enriched in centenarians ($n = 19$) compared with controls ($n = 3$) (Table 2; $P = 0.041$). In addition, three previously undescribed intronic variants in *hTERC* were also found, two of which were rare and found only in centenarians (Table 3).

We genotyped the four common *hTERT* variants, with a minor allele frequency (MAF) of >5% [IVS1–187 T > C, 973 G > A (Ala-305 Ala), 3097 C > T (His-1013 His), IVS16+99 C > T], in 73 centenarians and 49 controls. Information on telomere length was available for all these individuals. We did not find an association of the IVS1–187 T > C, 973 G > A (Ala-305 Ala), and IVS16+99 C > T with longevity. However, significant associations were detected with the 3097 C > T (His-1013 His) variant: the T allele is significantly enriched in centenarians (21.8%) compared with controls (8.2%; $P = 0.026$) (Table 3), and both CT and TT genotypes are significantly enriched in centenarians compared with controls (26.8% and 8.5% vs. 12.2% and 2%, respectively; $P = 0.035$) (Table 4). In contrast, no significant association between the *hTERC* common variant (IVS+63 T > C) and longevity was found (Table 4). Haplotype analysis [see SI Fig. S1 for a linkage disequilibrium (LD) plot derived from the four common variants] revealed that of the four most common haplotypes, Hap 1 is significantly depleted in centenarians compared with controls (34.9% vs. 53.1%; $P = 0.0056$), whereas Hap 3 is significantly enriched in centenarians compared with controls (13.7% vs. 8.2%; $P = 0.007$) when adjusted for age and gender (Table 5). Interestingly, a rare haplotype in controls, Hap 6, is found significantly more frequently in centenarians (1% vs. 6.8%; $P = 0.024$) (Table 5). Association analysis

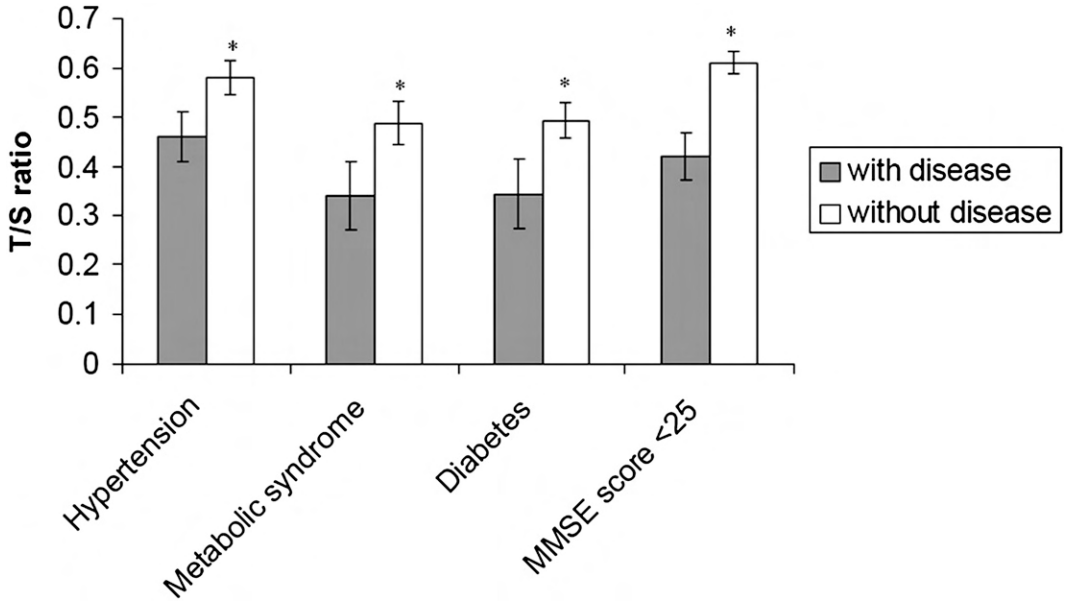


Fig. 2. Telomere length (adjusted for age, gender, and group) according to the absence (gray bars) or presence (white bars) of various age-related diseases in the study subjects. Number of individuals analyzed in each group: hypertension ($n = 125$), without hypertension ($n = 142$), with the metabolic syndrome ($n = 59$), without the metabolic syndrome ($n = 286$), diabetes ($n = 34$), without diabetes ($n = 276$). The MMSE was performed in centenarians only: MMSE score ≤25 (gray bars, $n = 46$) or >25 (white bars, $n = 32$). *$P < 0.05$.

Table 2. Telomere length and lipid profiles

	r	*P* value
Cholesterol, mg/dL	0.06	0.22
Triglyceride, mg/dL	−0.02	0.64
HDL, mg/dL	0.11	0.04
Large HDL, %	0.15	0.038
HDL size, nm	0.15	0.03
Apo-A, mg/dL	0.16	0.005
LDL, mg/dL	0.026	0.62
Large LDL, %	0.26	0.0002
LDL size, nm	0.29	0.0001
Apo-B, mg/dL	−0.003	0.95
Very LDL, mg/dL	−0.19	0.008

Telomere length was adjusted for group, gender, and age at recruitment.

of common *hTERT* or *hTERC* variants with telomere length indicates no significant association (Table S1). However, of the two haplotypes that are enriched in centenarians compared with controls, Hap 3 shows a positive association with telomere length after adjusting for age and gender ($P = 0.007$), whereas Hap 6 does not reach a significant threshold ($P = 0.59$) because of its low frequencies in the tested population (Table 6). In contrast, a centenarian-depleted haplotype, Hap 1, shows no association with telomere length (Table 5).

Discussion

Several studies have demonstrated a close relation between telomere length and life span in humans, including long-lived humans (4). However, interpretations of results from these studies are often confounded because of a lack of adequate controls. We employed a unique study design to overcome this shortcoming in a cohort of Ashkenazi Jewish individuals with exceptional longevity (centenarians), their offspring (approximate age of 70 years), and age- and gender-matched controls without a family history of unusual longevity. The use of offspring of individuals with exceptional longevity and their matched controls provides a powerful approach to identify genetically controlled longevity traits. This approach has previously led to the identification of longevity phenotypes such as lipoprotein sizes and the subsequent discovery of corresponding longevity genotypes (32–34). In the present investigation, we employed the same study design to delineate the relation between telomere lengths, longevity, and diseases of aging as well as to gain insights into the potential role(s) of genetic variations in the *hTERT* and *hTERC* genes on these phenotypes.

This study demonstrates that centenarians and their offspring have significantly longer telomeres than unrelated controls and that this trait is strongly heritable. Moreover, we have demonstrated that offspring of centenarians do not show an appreciable age-related decline in telomere length as is observed in our unrelated control population as well as in other cross-sectional studies (5). Because telomere length in younger persons (<75 years of age) is not significantly different between offspring of centenarians and unrelated controls, these results suggest that families with exceptional longevity have superior telomere length maintenance. Interestingly, unrelated controls older than 86 years of age exhibit longer telomeres than younger individuals (controls), suggesting that the rate of telomere attrition may be an important determinant of overall survival in the general population. The relatively older group may be affected by survival bias that selected out individuals with aging-related disease or those who would have died before reaching the age of 85 years, leaving the survivors with relatively longer telomeres.

Because shorter telomere length is associated with diseases of aging, including hypertension, the metabolic syndrome, and dementia (10), we studied the association between telomere length, major age-related diseases, and lipid profiles in centenarians, their offspring, and controls. Lipid profiles are known predictors of age-related diseases, and we have previously demonstrated that centenarians and their offspring have significantly larger HDL and LDL particle sizes and that these are heritable phenotypes of healthy aging associated with a lower prevalence of morbidity. We found that longer telomeres are indeed associated with lower prevalence of hypertension, the metabolic syndrome, type 2 diabetes, and better cognitive function as well as with healthier lipid profiles.

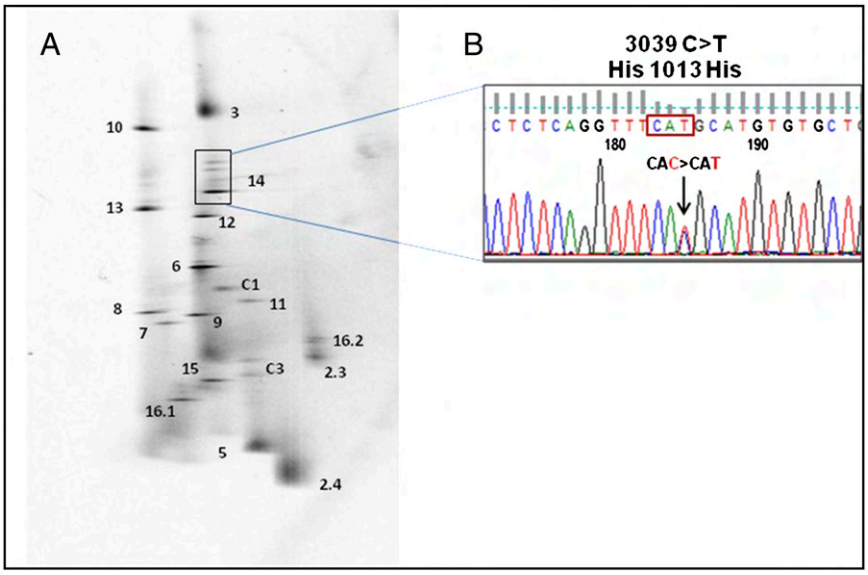


Fig. 3. 2D gene scanning of human telomerase gene (*hTERT* and *hTERC*). The coding regions and exon–intron junctions of the *hTERT* and *hTERC* were amplified by 2-step PCR. Eighteen PCR fragments were displayed in a 2D gel according to their size and melting temperature. (*A*) 2D gene scanning pattern from a centenarian subject with the fragment identification number and a heteroduplex band in exon 14 of the *hTERT*. (*B*) Common genetic variation in exon 14 was identified as 3097 C > T (His-1013 His) by nucleotide sequencing.

Table 3. Telomerase gene variants in centenarians and controls

	Nucleotide change	Protein change	No. of Heterozygotes: Centenarians (n = 100)	No. of Heterozygotes: Controls (n = 80)
hTERT				
Non-synonymous	835 G > A	Ala 279 Thr	0	1
	3148 G > A	Ala 1050 Thr	0	1
Synonymous	915 G > A	Ala 305 Ala	43	27
	1812 A > G	Ala 604 Ala	1	1
	1849 C > T	Leu 618 Leu	1	1
	2355 C > G	**Ser 785 Ser**	1	0
	2481 G > A	**Thr 827 Thr**	3	0
	2739 C > T	**His 913 His**	4	1
	3097 C > T	His 1013 His	28	11
Intronic	**IVS1-211G > A**		2	0
	IVS1-187 T > C		52	36
	IVS4+10 C > T		4	0
	IVS4+21 C > T		1	0
	IVS16+64 C > T		2	0
	IVS16+99 C > T		30	12
hTERC				
Intronic	**IVS-99 C > G**		1	0
	IVS+12 A > G		1	0
	IVS+63 T > C		30	27

Previously unknown unique variants are indicated in bold.

Telomere length, as a quantitative trait, is fairly well studied (19), but genetic factors that influence it are not well understood. Recent studies suggest that heterozygote mutation in *hTERT* and *hTERC* genes, which are the essential components of telomerase, show defective phenotypes in several diseases, indicating that half the usual dose of telomerase is inadequate for maintenance of telomeres with normal length (35, 36). Because living to 100 years of age is a rare phenotype in humans, with a prevalence of 1 in 10,000 individuals in the general population (37), we hypothesized that centenarians may harbor rare gain-of-function mutations in the telomerase genes that may also influence the length of telomeres. Sequence analysis of *hTERT* and *hTERC* genes revealed that rare synonymous or intronic variants in *hTERT* are enriched in centenarians (n = 19) compared with controls (n = 3) (Table 2; P = 0.041). In contrast, centenarians are completely devoid of nonsynonymous variants, whereas there are two control individuals who carry heterozygote nonsynonymous variants, 893 G > A (Ala-279 Thr) and 3148 G > A (Ala-1050 Thr). Although these nonsynonymous changes are predicted to be functionally benign, they are also found in idiopathic pulmonary fibrosis (IPF) patients with a MAF >5% (38), raising the possibility that these variants may have negative

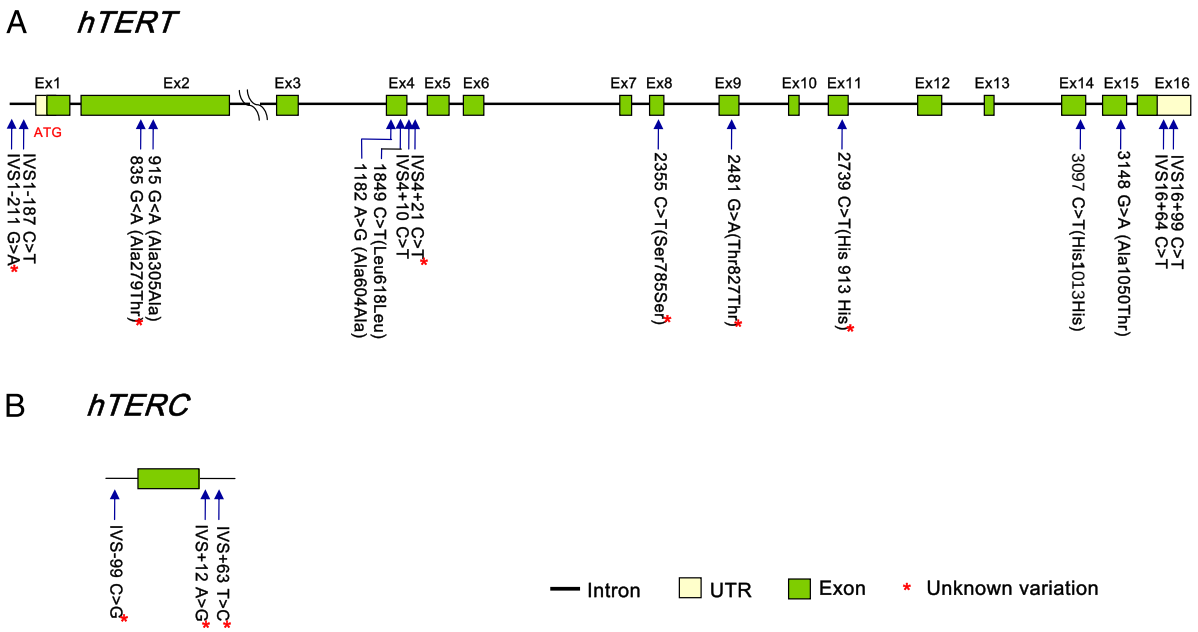


Fig. 4. Schematic representation of exons, introns, UTRs, and variants discovered. (*A*) *hTERT*. (*B*) *hTERC*. Green blocks, coding exons; white blocks, 5′ and 3′ UTRs; *previously unknown variants.

Table 4. Association of hTERT common alleles/genotypes with exceptional longevity

Nature of variation				Controls ($n = 49$)	Centenarians ($n = 73$)	P value[†]
				Allele, %		
TERT	Exon 1	IVS1–187 T > C	C	36.7	42.5	0.3717
	Exon 2	973 G > A (A 305 A)	A	30.6	34.9	0.4834
	Exon 14	3097 C > T (H 1013 H)	T	8.1	21.9	0.0049
	Exon 16	IVS16+99 C > T	T	11.2	16.4	0.2556
TERC		IVS+63 T > C	C	18.5	18.8	0.9402
				Genotype, %		
TERT	Exon 1	IVS1–187 T > C	TT	40.8	31.5	0.4494
			CT	44.9	52.1	
			CC	14.3	16.4	
	Exon 2	973 G > A (A 305 A)	GG	53.1	43.8	0.5313
			GA	32.7	42.5	
			AA	14.3	13.7	
	Exon 14	3097 C > T (H 1013 H)	CC	85.7	63.9	0.035
			CT	12.2	26.8	
			TT	2	8.5	
	Exon 16	IVS16+99 C > T	CC	81.6	68.5	0.0972
			CT	14.3	30.1	
			TT	4.1	1.4	
TERC		IVS+63 T > C	TT	64.4	66.2	0.5653
			CT	34.2	29.9	
			CC	1.4	3.9	

[†]Adjusted for age and gender.

functional effects. In addition, although we did not find a single coding variation in the *hTERC* gene, we found two rare intronic variants only in centenarians. The enrichment of both synonymous and intronic variants of *hTERT* and *hTERC* genes in centenarians is intriguing because these variants are known to play a functional role in the regulation of gene expression through modulation of mRNA stability, mRNA secondary structure, alternative splicing, or translational efficiency (39, 40).

The expression level of hTERT is a major determinant of telomerase activity (41). Recent studies indicate that telomerase activation can be regulated by environmental interventions such as lifestyle changes or stress management (42). There are functional hTERT promoter variants known to influence telomerase expression and telomere length (43). However, because there is very little LD in the *hTERT* gene, it is not likely that the centenarian-enriched rare variants are proxies of functional variants in the *hTERT* promoter. It is tempting to speculate that the rare variants might influence expression of *hTERT* in response to environmental stresses such as inflammation or oxidative stress, which then promotes better maintenance of telomere length. Indeed, mean telomere length is much greater among the carriers of the rare variant compared with the noncarriers after adjusting for age and gender [telomere repeat copy number to single-copy gene copy number (T/S) ratio of 0.62:0.41, crude difference = 0.21; $P = 0.02$ by Student t test, $P = 0.037$ by Mann-Whitney U test], raising the possibility that these rare variants may have a positive functional impact on telomere maintenance. Our study also revealed that common genetic variations in hTERT may influence the telomere length maintenance associated with longer telomeres in families with exceptional longevity.

Haplotypes of the *hTERT* gene derived from four common variants showed associations with longevity and/or telomere length. These are two intronic variants, IVS1–187 T > C and IVS16+99 C > T, and two synonymous variants, 973 G > A (Ala-305 Ala) and 3097 C > T (His-1013 His). These common variants may contribute to the regulation of *hTERT* gene expression. A common intronic *hTERT* variant shown to be associated with susceptibility to IPF may also affect the expression levels of hTERT (44). Taken together, our study demonstrates that centenarians may harbor individually rare but collectively more common genetic variations in genes involved in the telomere maintenance pathway, implicating a role of regulatory variants in

Table 5. hTERT haplotypes and association with exceptional longevity

Hap	IVS1–187 T > C	973 G > A (A305A)	3097 C > T (H1013H)	IVS16+99 C > T	Haplotype frequency, %		
					Centenarians ($n = 74$)	Controls ($n = 49$)	P value*
1	T	G	C	C	34.9	53.1	0.008
2	C	A	C	C	21.9	23.5	0.322
3	T	G	C	T	13.7	9.2	0.007
4	C	A	T	C	10.3	5.1	0.392
5	C	G	C	C	6.2	5.1	0.749
6	T	G	T	C	6.8	1.0	0.024
7–9			Others		6.2	3.1	0.203

*Adjusted for age and gender.

Table 6. hTERT haplotypes and association with telomere length

Haplotype	Hap	Increase (+) or decrease (–) in telomere length (T/S ratio)	*P* value
T G C C	1	−0.0307	0.742
C A C C	2	−0.0759	0.447
T G C T	3	0.4040	0.007
C A T C	4	−0.0222	0.899
C G C C	5	−0.1280	0.648
T G T C	6	0.1235	0.591
Rare	7–9	−0.0837	0.340

the telomerase gene in human longevity and maintenance of telomere length.

In summary, we provide strong correlative evidence that individuals in Ashkenazi Jewish families with exceptional longevity have better maintenance of telomere length and that the telomerase genes may function as important genetic determinants of both human longevity and telomere length. Additionally, our data suggest that both telomere length and variants of telomerase genes may have a cumulative influence on lower disease prevalence and a favorable lipid profile in centenarians and their offspring. Additional comprehensive studies on genetic and genomic variation of centenarians and their offspring comprising candidate genes, especially those known to play a role in telomere maintenance in model organisms, may reveal previously undescribed genomic regions and molecules that are operative in human health and longevity.

Materials and Methods

Study Design and Subjects. A cohort of Ashkenazi Jews with exceptional longevity was recruited as previously described (34). Probands were required to be living independently at 95 years of age as a reflection of good health, although at the time of recruitment, they could be at any level of dependency. In addition, probands had to have a first-degree offspring who was willing to participate in the study. Birth certificates or dates of birth as stated on passports were used to verify the participants' ages. The Ashkenazi population is descended from a founder population (estimated to be several thousands) originating in the 15th century. This "founder effect" resulted in a population both culturally and genetically homogeneous, from which several disease-related genes have been successfully identified (45). A majority of these individuals were born in the United States or moved there before World War II. Telomere measurements (see below) were made on 86 Ashkenazi probands (aged 97.4 ± 0.3 years) defined as having exceptional longevity, 175 offspring of parents with exceptional longevity (aged 67.8 ± 0.6 years), and 93 controls who were offspring of parents with usual survival (aged 71.8 ± 1 years). By definition, parents of controls survived to the age of 85 years or less. The control group consisted largely of spouses of the offspring group. Informed written consent was obtained in accordance with the policy of the Committee on Clinical Investigations of the Albert Einstein College of Medicine.

For each subject, a detailed medical history, physical examination, and blood sample collection were performed as previously described (23, 32, 34). The presence of hypertension, the metabolic syndrome, and diabetes was determined using National Cholesterol Education Program (NCEP), Adult Treatment Panel III (ATP III) criteria (46). Scores on the MMSE (47) ≤25 were considered to indicate impaired cognitive function. Triglyceride, LDL-cholesterol (LDL), and HDL-cholesterol (HDL) levels and average particle sizes were determined by NMR spectroscopy at LipoScience, Inc., as previously described (34). Large lipoprotein particle sizes were defined as >8.9 nm for HDL and >21.3 nm for LDL.

Measuring/Estimating Length by Quantitative Real-Time PCR. Genomic DNA was extracted from blood samples and stored at −80°C. The average (of triplicate) telomere length in leukocytes was estimated using a quantitative PCR (qPCR) method (48). This approach provides relative average telomere lengths in genomic DNA by determining the T/S ratio in experimental samples relative to a reference sample. Briefly, standard curves were generated for telomere lengths and for the single-gene copy amplification reactions from a reference DNA sample that was serially diluted with double-distilled water by 1.68-fold per dilution to produce 5 concentrations of DNA ranging from 7.7 to 61 ng/μL. For analysis of subjects' DNA, triplicate PCR reactions using 2 μL of each DNA dilution were carried out in a 20-μL volume using the LightCycler FastStart DNA Master SYBR Green kit (Roche Diagnostics), with $MgCl_2$ added to a final concentration of 3 mM. Primers for telomeres and the single-copy gene HGB (β-globin gene) were added to final concentrations of 0.1 μM per 0.9 μM and 0.3 μM per 0.7 μM, respectively. The primer sequences are TEL 1b, 5′-CGGTTTGTTTGGGTTTGGGTTTGGGTT-TGGGTTT-GGGTT-3′; TEL 2b, 5′-GGCTTGCCTTACCCTTACCCTTACCCTTACCCTT-ACCCT-3′; HGB1, 5′-GCTTCTGACACAACTGTGTTCACTAGC-3′; and HGB2, 5′-CACCAACTT-CATCCACGTTC-ACC-3′ (48). The enzyme was activated at 95°C for 10 min, followed by 30 cycles at 95°C for 5 s, 58°C for 10 s, and 72°C for 40 s for the HGB reaction or for 20 cycles at 95°C for 5 s, 56°C for 10 s, and 72°C for 60 s for the telomere reaction. All transition rates were set to 20°C/s with the exception of the annealing transition rate in the telomere reaction, which was set to 4°C/s. Forty samples (mixed samples from probands, offspring, and controls) were analyzed by qPCR, both in our laboratory and by one of the authors (R.M.C.) for validation purposes, yielding a correlation of 0.89 ($P < 0.001$). In addition, to ensure reproducibility, when measuring previously unanalyzed samples, and to account for assay variability between batches, we used an anchor sample that was previously measured. The reproducibility between the two measurements was high ($r = 0.94$, $P < 0.001$). Results obtained using this qPCR were highly correlated with those obtained by the traditional terminal restriction fragment length approaches (48). One T/S ratio unit is equivalent to a mean telomere length of 4,270 bp in the leukocytes. The qPCR telomere length determination does not include subtelomeric lengths.

Mutation Screening and Genotyping of hTERT and hTERC. We sequenced genomic DNA isolated from blood of centenarians and controls to discover all possible genetic variations in the exonic regions and intron–exon boundaries of the telomerase genes (*hTERT* and *hTERC*) by a 2D gene scanning (TDGS) method (49). To increase specificity, PCR amplification was designed in two steps. To begin with, triple-multiplex long-distance PCR coamplifies eight large fragments encompassing the coding and splicing junction regions of *hTERT* and *hTERC*. These products serve as a template for quadruple-multiplex short PCRs to amplify target regions in 18 fragments; 16 fragments for hTERT and two fragments for hTERC. Primers are listed in Table S2. The mixture of amplicons after heteroduplexing was separated on the basis of size and melting temperature in a 2D gel. Then, DNA fragments containing heterozygous sequence variation were reamplified from the genomic DNA and sequenced to identify the position and nature of sequence variation. For *hTERT*, exon 1 was analyzed by nucleotide sequencing and exons 2, 4, and 5 were separately analyzed by denaturing gradient gel electrophoresis (DGGE) because of ambiguous melting patterns on a TDGS gel. Genotypes of the common variants in the telomerase gene were determined by DGGE using the same primers designed for the TDGS assay.

Statistical Analyses. Two comparisons have been established: (*i*) between probands and controls for whom crude data were adjusted for gender and (*ii*) between offspring and controls for whom crude data were adjusted for age and gender. Individual phenotypes that were not normally distributed were log natural-transformed for analysis and back-transformed for the current presentation. Associations of telomere length with other variables were estimated using regression on an indicator variable for group, age, and gender. In defining strata for age, cutoff points of 65, 75, 85, 95, and 105 years were used (Fig. 1). Telomere length distribution in probands appears to be nonnormally distributed; thus, to test for statistical significance, we applied nonparametric statistics analysis, and pairwise crude comparisons of telomeres between the study groups were carried out using the Wilcoxon signed-rank test. Parameters were estimated by maximum likelihood methods. The significance of the parameter (association) was evaluated using the likelihood ratio test by comparing the model in which the

parameter is set to 0 with the model in which the parameter is estimated. In addition, we compared telomere length between offspring and controls using a hierarchical linear regression model to account for clustering within families. The regression equation adjusted for age and gender included an indicator variable distinguishing offspring from controls and an offspring vs. controls × age interaction term. We tested the hypotheses that offspring have higher mean telomere length than controls and that the age-related decline in telomere length is less in offspring than in controls. To determine the effect of lipid profile on telomere length, a general linear model adjusted for group, gender, and age at recruitment was used. Correlations of telomere length and lipid profile adjusted for group, gender, and age at recruitment were evaluated using Spearman rank correlation coefficients. Heritability of telomere length was assessed in two ways: First, it was computed in the ASSOC module of SAGE (v6.0.1) (http://darwin.cwru.edu/sage/). Second, narrow sense heritability was estimated from the slope of the linear regression of the traits of each parent on the mean value of the offspring group (16). Data are expressed as mean (SE) as appropriate. Statistical analyses were performed using SAS 9.1 (SAS Institute).

Comparison of the observed numbers of each genotype with those expected under Hardy-Weinberg equilibrium was tested using χ^2 tests for each of the three groups: centenarians, their offspring, and controls. Haplotypes were constructed using the PHASE algorithm (50). The haplotype trend regression analysis for logistical (case-control), linear (telomere length), and proportional hazard with telomere length was performed using JMP genomics 4 (SAS Inc., Cary, NC).

ACKNOWLEDGMENTS. We are indebted to all participants and their families for their dedication and enthusiasm to enroll in the longevity study. We are also grateful to the Hebrew Home for the Aging (Riverdale, NY), Kittay House (Bronx, NY), Hebrew Home Hospital (West Hartford, CT), and Jewish Home for the Aged (New Haven, CT), all under the aegis of the association for the Jewish Aging Services (Washington, DC). This work was supported by grants from the Paul Beeson Physician Faculty Scholar in Aging Award, Ellison Medical Foundation Senior Scholar Award, Glenn Award for Research in Biological Mechanisms of Aging, and theResnick Gerontology Center, and by National Institutes of Health Grants RO1 AG-18728-01A1, RO1 AG024391, PO1 AG027734, PO1 AG17242 and RO1 AG7992, General Clinical Research Center Grant MO1-RR12248, and Diabetes Research and Training Center Grant DK 20541 at the Albert Einstein College of Medicine. Some of the results of this paper were obtained by using the program package S.A.G.E., which is supported by a U.S. Public Health Service Resource Grant (RR03655) from the National Center for Research Resources.

1. de Lange T (2005) Shelterin: The protein complex that shapes and safeguards human telomeres. *Genes Dev* 19:2100–2110.
2. Wong JM, Collins K (2003) Telomere maintenance and disease. *Lancet* 362:983–988.
3. Campisi J, d'Adda di Fagagna F (2007) Cellular senescence: When bad things happen to good cells. *Nat Rev Mol Cell* 8:729–740.
4. Finch CE (2007) *The Biology of Aging* (Academic Press, Burlington, MA).
5. Njajou OT, *et al.* (2007) Telomere length is paternally inherited and is associated with parental lifespan. *Proc Natl Acad Sci USA* 104:12135–12139.
6. Benetos A, *et al.* (2001) Telomere length as an indicator of biological aging: The gender effect and relation with pulse pressure and pulse wave velocity. *Hypertension* 37:381–385.
7. Lindsey J, McGill NI, Lindsey LA, Green DK, Cooke HJ (1991) In vivo loss of telomeric repeats with age in humans. *Mutat Res* 256:45–48.
8. Nawrot TS, Staessen JA, Gardner JP, Aviv A (2004) Telomere length and possible link to X chromosome. *Lancet* 363:507–510.
9. Slagboom PE, Droog S, Boomsma DI (1994) Genetic determination of telomere size in humans: a twin study of three age groups. *Am J Hum Genet* 55:876–882.
10. Aviv A (2006) Telomeres and human somatic fitness. *J Gerontol A Biol Sci Med Sci* 61: 871–873.
11. von Zglinicki T, Martin-Ruiz CM (2005) Telomeres as biomarkers for ageing and age-related diseases. *Curr Mol Med* 5:197–203.
12. Cawthon RM, Smith KR, O'Brien E, Sivatchenko A, Kerber RA (2003) Association between telomere length in blood and mortality in people aged 60 years or older. *Lancet* 361:393–395.
13. Bischoff C, *et al.* (2005) The heritability of telomere length among the elderly and oldest-old. *Twin Res Hum Genet* 8:433–439.
14. Bischoff C, *et al.* (2006) No association between telomere length and survival among the elderly and oldest old. *Epidemiology* 17:190–194.
15. Martin-Ruiz CM, Gussekloo J, van Heemst D, von Zglinicki T, Westendorp RG (2005) Telomere length in white blood cells is not associated with morbidity or mortality in the oldest old: A population-based study. *Aging Cell* 4:287–290.
16. Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics* (Addison Wesley Longman, Essex, UK).
17. Vasa-Nicotera M, *et al.* (2005) Mapping of a major locus that determines telomere length in humans. *Am J Hum Genet* 76:147–151.
18. Andrew T, *et al.* (2006) Mapping genetic loci that determine leukocyte telomere length in a large sample of unselected female sibling pairs. *Am J Hum Genet* 78: 480–486.
19. Blasco MA (2007) Telomere length, stem cells and aging. *Nat Chem Biol* 3:640–649.
20. Bodnar AG, *et al.* (1998) Extension of life-span by introduction of telomerase into normal human cells. *Science* 279:349–352.
21. Garcia CK, Wright WE, Shay JW (2007) Human diseases of telomerase dysfunction: Insights into tissue aging. *Nucleic Acids Res* 35:7406–7416.
22. Shay JW, Wright WE (2007) Hallmarks of telomeres in ageing research. *J Pathol* 211: 114–123.
23. Atzmon G, *et al.* (2004) Clinical phenotype of families with longevity. *J Am Geriatr Soc* 52:274–277.
24. Benetos A, *et al.* (2004) Short telomeres are associated with increased carotid atherosclerosis in hypertensive subjects. *Hypertension* 43:182–185.
25. Brouilette SW, *et al.*; West of Scotland Coronary Prevention Study Group (2007) Telomere length, risk of coronary heart disease, and statin treatment in the West of Scotland Primary Prevention Study: A nested case-control study. *Lancet* 369:107–114.
26. Panossian LA, *et al.* (2003) Telomere shortening in T cells correlates with Alzheimer's disease status. *Neurobiol Aging* 24:77–84.
27. Samani NJ, Boultby R, Butler R, Thompson JR, Goodall AH (2001) Telomere shortening in atherosclerosis. *Lancet* 358:472–473.
28. von Zglinicki T, *et al.* (2000) Short telomeres in patients with vascular dementia: An indicator of low antioxidative capacity and a possible risk factor? *Lab Invest* 80: 1739–1747.
29. Govindaraju DR, *et al.* (2008) Genetics of the Framingham Heart Study population. *Adv Genet* 62:33–65.
30. Ng PC, Henikoff S (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31:3812–3814.
31. Sunyaev S, *et al.* (2001) Prediction of deleterious human alleles. *Hum Mol Genet* 10: 591–597.
32. Atzmon G, *et al.* (2002) Plasma HDL levels highly correlate with cognitive function in exceptional longevity. *J Gerontol A Biol Sci Med Sci* 57:M712–M715.
33. Barzilai N, Atzmon G, Derby CA, Bauman JM, Lipton RB (2006) A genotype of exceptional longevity is associated with preservation of cognitive function. *Neurology* 67:2170–2175.
34. Barzilai N, *et al.* (2003) Unique lipoprotein phenotype and genotype associated with exceptional longevity. *J Am Med Assoc* 290:2030–2040.
35. Artandi SE (2006) Telomeres, telomerase, and human disease. *N Engl J Med* 355: 1195–1197.
36. Verma S, Slutsky AS (2007) Idiopathic pulmonary fibrosis—New insights. *N Engl J Med* 356:1370–1372.
37. Perls TT, Bochen K, Freeman M, Alpert L, Silver MH (1999) Validity of reported age and centenarian prevalence in New England. *Age Ageing* 28:193–197.
38. Tsakiri KD, *et al.* (2007) Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proc Natl Acad Sci USA* 104:7552–7557.
39. Kimchi-Sarfaty C, *et al.* (2007) A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* 315:525–528.
40. Qu HQ, Lawrence SG, Guo F, Majewski J, Polychronakos C (2006) Strand bias in complementary single-nucleotide polymorphisms of transcribed human sequences: Evidence for functional effects of synonymous polymorphisms. *BMC Genomics* 7:213.
41. Ducrest AL, Szutorisz H, Lingner J, Nabholz M (2002) Regulation of the human telomerase reverse transcriptase gene. *Oncogene* 21:541–552.
42. Ornish D, *et al.* (2008) Increased telomerase activity and comprehensive lifestyle changes: A pilot study. *Lancet Oncol* 9:1048–1057.
43. Matsubara Y, *et al.* (2006) Telomere length of normal leukocytes is affected by a functional polymorphism of hTERT. *Biochem Biophys Res Commun* 341:128–131.
44. Mushiroda T, *et al.*; Pirfenidone Clinical Study Group (2008) A genome-wide association study identifies an association of a common variant in TERT with susceptibility to idiopathic pulmonary fibrosis. *J Med Genet* 45:654–656.
45. Lancaster JM, Carney ME, Futreal PA (1997) BRCA 1 and 2—A genetic link to familial breast and ovarian cancer. *Medscape Womens Health* 2:7.
46. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (2001) Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *J Am Med Assoc* 285: 2486–2497.
47. Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 12:189–198.
48. Cawthon RM (2002) Telomere measurement by quantitative PCR. *Nucleic Acids Res* 30:e47.
49. Vijg J, Suh Y (2005) in *Molecular Diagnostics*, eds Patrinos GP, Ansorge W (Elsevier, Amsterdam), pp 95–105.
50. Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 73:1162–1169.

Evolution of the human lifespan and diseases of aging: Roles of infection, inflammation, and nutrition

Caleb E. Finch[1]

Davis School of Gerontology and the University of Southern California, Los Angeles, CA 90089

Edited by Stephen Curtis Stearns, Yale University, New Haven, CT, and accepted by the Editorial Board October 12, 2009 (received for review August 25, 2009)

Humans have evolved much longer lifespans than the great apes, which rarely exceed 50 years. Since 1800, lifespans have doubled again, largely due to improvements in environment, food, and medicine that minimized mortality at earlier ages. Infections cause most mortality in wild chimpanzees and in traditional forager-farmers with limited access to modern medicine. Although we know little of the diseases of aging under premodern conditions, in captivity, chimpanzees present a lower incidence of cancer, ischemic heart disease, and neurodegeneration than current human populations. These major differences in pathology of aging are discussed in terms of genes that mediate infection, inflammation, and nutrition. Apolipoprotein E alleles are proposed as a prototype of pleiotropic genes, which influence immune responses, arterial and Alzheimer's disease, and brain development.

chimpanzee | pathology

Humans have the longest life spans of any primate. Even under the conditions of high mortality experienced by hunter-foragers, the human life expectancy at birth (LE_0) is twice that of wild chimpanzees. This inquiry considers the demographics and pathology of aging in humans and great apes as an approach to understanding how aging processes evolved with longer lifespans. I argue that immune functions and nutrition have been of major importance in the evolution of aging and longevity.

Evolving Demographics of Aging

The human LE_0 has doubled during over an evolutionary span of about 300,000 generations from a great ape ancestor shared with chimpanzees (1, 2). Then during the last 200 years during industrialization and in <10 generations, the LE_0 has doubled again (3, 4), allowing major increases in older ages. The lifespans of intermediate species during human evolution cannot be known, because the spotty skeletal evidence at hand allows only general estimates of age classes. According to tooth wear, early modern *H. sapiens* and *H. neanderthalensis* had a larger proportion of older adults than prior *Homo* species and *Australopithecus* (5).

Turning from the huge gap before historical times, we may model earlier *H. sapiens* demographics by preindustrial populations for which there is good demographic data: Sweden from 1751 (3, 4) and 20th-century hunter-foragers (6–8). Both lived under unhygienic conditions with high burdens of infection and limited access to effective medicine. Their high mortality at early ages of 10%–30% restricted the LE_0 to 30–40 years. Despite low survival, half of those reaching age 20 reached 60 (LE_{20} of 40 years). Thus, most hunter-gatherers survive beyond menopause, unlike wild chimpanzees (7–9). The greater survival to later ages allowed the evolution of stable multigenerational support of the young, a uniquely human trait among primates (7–9).

Mortality across the lifespan forms a J-shaped curve in most mammalian populations: the high early age mortality declines to a minimum (q_{min}) at the approach of adulthood, followed at midlife by exponential accelerations of mortality in association with increased chronic degenerative disease and dysfunctions that collectively define senescence (2, 10). Humans differ from wild chimpanzees by their lower mortality in juvenile and adult ages, and by the later onset of mortality rate acceleration (6, 8, 11) (Fig. 1). In healthy populations of humans and lab animals, the acceleration of mortality is preceded by increasing morbidity from chronic degenerative disease (2, 10). For wild chimpanzees, typical early mortality rates are 20% per year in infancy, within the range of hunter-gatherers, then decreasing to a q_{min} of about 3.5% per year in preadult ages. The chimpanzee life expectancy at birth (LE_0) is about 13 years, whereas those reaching adulthood (age 15) have about 15 years of further life expectancy (6, 11) (Table 1). Very few have survived beyond age 50, even in captivity with modern veterinary care (13). In contrast, human mortality after the early years is much greater, with >2-fold longer LE_0 and >3-fold lower q_{min}, even with limited access to medicine (Table 1). Since 1800, the LE_0 in developed nations rose progressively to >70 years. Only recently has survival to >90 been well documented; currently, centenarians are about 0.01%–0.02% in developed nations (14). Two key factors in human life expectancy are the delayed mortality rate acceleration and lower q_{min} (Fig. 1). The q_{min} merits attention in human evolution (10): even in populations with high infectious burdens and neonatal mortality, the human q_{min} is >50% lower than wild chimpanzees (Table 1). As discussed later, this apparent species difference may be due to stronger immune responses. Since 1800, the industrialized countries have further lowered q_{min} by 25-fold (12).

Causes of Mortality. There is frustratingly little information on diseases of aging in wild chimpanzees and in hunter-foragers for comparison with modern populations. The following summary necessarily includes individual observations as well as larger studies.

Infections. The main cause of mortality throughout human evolution until the 20th century must have been infections, as observed in wild chimpanzees and 20th-century hunter-foragers. Longitudinal studies of the Gombe chimpanzees (Tanzania) since 1960 by Goodall and colleagues identified infections in the majority of deaths (67%) for all ages (Table 2) (15). The oldest individuals frequently had prolonged diarrhea (16, p 104). Infected wounds from accidents or fighting were also a common secondary cause of death (see note to Table 2). The accelerating mortality rates of chimpanzees soon after age 20 (Fig. 1) implies decreasing resistance to infections with aging, as well as synergies of infections with other myocardial damage, discussed below.

The Gombe chimpanzees cannot be considered a pristine population because of their exposure to pathogens from local humans and domestic animals (e.g., mange, polio, and tuberculosis). A recent SIVcpz infection (chimpanzee-derived simian immunodeficiency virus) has been transmitted vertically and horizontally, with >10-fold higher mortality in carriers and lower fertility and

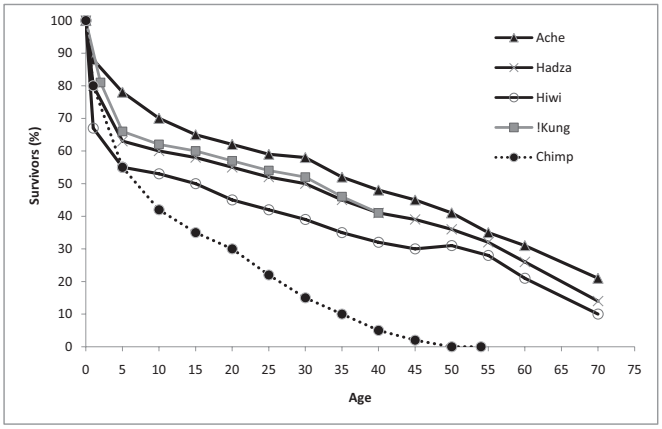
This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Evolution in Health and Medicine" held April 2–3, 2009, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS web site at www.nasonline.org/Sackler_Evolution_Health_Medicine.

Author contributions: C.E.F. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. S.C.S. is a guest editor invited by the Editorial Board.

[1]To whom correspondence should be addressed. E-mail: cefinch@usc.edu.



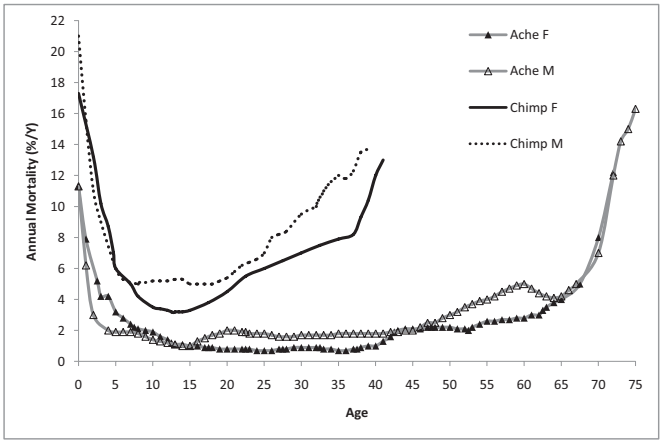


Fig. 1. Demographic comparisons of wild chimpanzees with human populations living under poor hygiene and with little access to medicine. [Reproduced with permission from ref. 6 (Copyright 2000, John Wiley & Sons).] (*A*) Survival curves. (*B*) Age-specific mortality. At all ages after infancy, chimpanzees have higher mortality than the Ache and show acceleration of mortality at least 20 years earlier.

infant survival (17). End stages had depletion of $CD4^+$ T cells and secondary infections, like human HIV. There may be no remaining truly isolated chimpanzee population in which to evaluate pathogen loads and mortality causes, because of increased commercial activity and warfare. There is no detailed profile of native infectious agents in any wild primate population (18)

Human forager-farmers traditionalists with limited access to modern medicine also show infections as a main cause of death (72%) (Table 1) (7, 19). These human populations, although relatively isolated, also had tuberculosis and other pathogenic infections (7) that are unlikely to have been indigenous (20). Notably unlike chimpanzees, a definitive proportion of elderly forager-farmers age 60 or older died from nonspecific senescent causes.

Before the 18th century, there are no national or regional statistical data on mortality rates by age group or causes of death. In the ancient Greco-Roman world, demographic reconstructions agree on short LE_0 ranging 20–35 years (21, 22). These calculations are based on tombstone epitaphs and graveyard samples, which are notoriously unrepresentative (21–23). It may be concluded that few in this era lived longer than 90 years, which is the upper age limit validated in hunter-gatherer-foragers (6, 7). Contagious infections and septic wounds are likely to have been the major causes of death in ancient populations living under unsanitary conditions (2, 21, 22).

The high incidence of infectious causes of death among 20th-century hunter-gathers resembles pre-20th-century populations, where infections, directly or indirectly, were major causes of adult deaths. In national data for England and Wales of 1861, for example, infections caused 25% of female deaths before age 40 (24). Almost all deaths before age 5 were due to infections (9, 24, 25). The much lower q_{min} for humans than chimpanzees (Table 1) suggests corresponding differences of immunological functions, as described later. Data on cause of death for juvenile ages are needed to evaluate the contribution to mortality from transmissible infections, septic wounds from in-group aggression and accidents, and from predation to which subadults are more vulnerable by smaller size and lack of experience.

Exposure to chronic infections and inflammation has major ramifications for aging processes through 2 main fronts: immunosenescence and synergies with chronic diseases that have inflammatory components. In brief, immunosenescence involves depletion of the limited pool of naïve T cells acquired during maturation. During antigenic exposure across life, the pool of memory T cell ($CD8^+$ $CD28^-$ T cells) increases progressively, at least in part from antigenic stimulation by common infections e.g., CMV, HSV, influenza (2, 26–28). A subgroup of elderly with the "immune risk" phenotype for higher mortality have relative depletion of $CD28^+$ T cells and memory T cells with telomere erosion, increased cytokine expression, and other markers of cellular immunosenescence. Because HIV is associated with accelerated memory T-cell accumulation and frailty (29), it is predicted that immunosenescence will be accelerated in the hunter-gatherers with high infectious loads (Table 2).

The associations of high early mortality and shorter life expectancy in historical populations give important clues to early human evolution in highly infectious environments. Crimmins and I examined Sweden and several other 19th-century European populations that had high early age mortality from infections. The correlations of mortality before age 5 with mortality at age 70 were much stronger for birth cohorts than for the periods (5, 25). We proposed a "cohort morbidity hypothesis" in which survivors of early infections carried higher inflammatory loads, which promote chronic diseases with inflammatory components, such as cardiovascular disease. Atherosclerosis, for example, begins before birth, with accumulating lipids, monocytes, and local oxidative damage; "fetal programming" from maternal diet, cholesterol, and stress can influence the later progression of arterial degeneration (2, 30, 31). Higher mortality of elderly to infections could also be involved in cohort effects, e.g., cytomegalovirus (CMV) infections are associated with immunosenescence and cardiovascular disease (2, 32). The progressive reduction of mortality at later ages in birth cohorts with better early survival is likely to involve complex interactions of atherosclerosis and immunity (2).

To further evaluate relationships between infectious exposure and accelerated aging, we examined the 1918 U.S. influenza pandemic for birth cohorts exposed pre- and postnatally (31). Most deaths were secondary to bacterial infections that caused severe pneumonia. This population was considered well-nourished, unlike earlier European cohorts. Specific prenatal influences were found on later aging: the cohort exposed prenatally to the peak pandemic in 1918 had 25% excess ischemic heart disease 60–82 years later, relative to flanking birth quarters. Moreover, the 1919 birth cohort had lower educational achievement and was slightly shorter at WWII enlistment. Because influenza rarely invades the placenta or fetus, these effects may involve stress effects on the fetus with elevations of maternal cortisol and IL-6, and imprinting of the fetal genome (reviewed in ref. 31). Apparently, even brief maternal infections without malnutrition impair postnatal growth and accelerate cardiovascular aging. These findings also extend the Barker theory of developmental origins of adult diseases of aging to effects of stress on the fetus from maternal infections.

Arterial disease. We cannot know the incidence of arterial disease or cancer in pre-20th-century populations because there are no population-based clinical data. Nonetheless, there are indications of arterial diseases in early historical human populations. For

Table 1. Comparative demographics of chimpanzees and humans

	LE_0, years[b]	Infant mortality, %	LE_{adult}, years[b]	q_{min}, % per year[c]	Age, years	Max. lifespan
Chimpanzee (feral) (6)	13	0.20	15	0.035	15	<50
Preindustrial populations						
Forager-horticulturalists (6–8)[a]	33	0.42	40	0.012	11-20	<100
Sweden 1751	35	0.21	40	0.0080	11-20	?[d]
Industrial populations with good nutrition, hygiene, and medicine (12)						
Sweden 1931	64	0.06	49	0.0008	11-20	>100
Sweden 1978	72	0.03	56	0.0003	11-20	
Sweden 2007	79	0.028	61			

[a]Average of 5 nonacculturated groups (6, see table 2); values are rounded, sources in parentheses.

[b]LE_{adult} used age 15 y for chimpanzee and 20 y for human (7, 8). The alleged 75+ age of the Hollywood chimpanzee *Cheeta* was recently discredited (13).

[c]q_{min}, minimum mortality across the lifespan (see Fig. 1).

[d]Rigorous analysis has disproven most claims of longevity >100 years before the later 19th century, when birth records gave more certain identification (14). Sweden in 1751–1760 recorded 24 centenarians per million, which is considered erroneous because of poor records, although it is less than 20th-century norms of 100 centenarians per million; the number declined to 1 per million by 1851, which is still considered uncertain, despite improving records. The accepted record lifespan is 122 of Jeanne Calment, 1875–1997, although there are still skeptics.

arterial disease, the oldest case is the Tyrolean iceman from 5,300 years ago, who died accidentally at about age 45; CT imaging showed calcification of both carotid arteries and portions of aorta and iliac artery (33). Arterial disease was also described for Egyptian mummies from 3,500 years ago (18th Dynasty; $n = 24$) (34, 35): 67% of large arteries were atherosclerotic; of these, 50% were calcified. In modern populations, arterial calcification is a high risk marker for vascular fatal events, with 4-fold more mortality in the following decade (36, 37). Coronary atherosclerosis was also found in mummies from dynastic Egypt (34, 38), China (1150 BCE), and Alaskan Inuit (430 CE) (39). Though these scattered samples cannot inform about the prevalence of atherosclerosis in historical populations or its contribution to mortality, they suggest that advanced atherosclerosis is not a modern condition.

For chimpanzees, the only histopathological data are from captives, which in earlier decades were exposed to varying conditions of husbandry and diet, including dairy products, which are not normal for wild chimpanzees. Up thru 1980, arterial fatty degeneration and sudden death from heart attack or stroke were widely noted as comparable to humans, as represented in these examples from a scattered literature (1). In one U.S. colony, on Yerkes' natural diet, all adults had cerebral arterial lesions, but no coronary lesions (40), whereas the majority of another colony had coronary lesions (41). Premature sudden death from myocardial infarcts was observed elsewhere in 2 young females, an 8-year-old on an unspecified diet (42), and a 10-year-old female with extreme hypercholesterolemia (≥600 mg/dL serum) from a fatty diet (43). More generally, on typical primate diets before 1980, 80% of chimpanzees had elevated cholesterol (*ca*. 200–300 mg/dL serum) (1, table 3*A*). These levels would be considered high risk for cardiovascular events in humans.

Table 2. Cause of death in feral chimpanzees and hunter-gatherers

	Chimpanzees, %[a]	Traditional humans, %[b]
Infections	67 (TB, polio, mange)	73
Violence/accidents	32	17
Senescence	1	10

[a]Feral chimpanzees (Kasekela community of Gombe, Western Tanzania), studied by Jane Goodall and colleagues, 1960–2006, with 73 deaths across all ages (15, 16). This table excludes deaths from poaching and predation; deaths of dependent offspring from maternal death or disability; and from unknown causes. "Illness" represents the largest cause of death and includes polio, mange, and wasting, and respiratory conditions (epidemic and nonepidemic of 48%). Wasting is described as "a conglomerate of enteric diseases, parasitic infections, or perhaps cancer or AIDS-like disease"; some were positive for streptococci and nematode parasites. The life history of the oldest individuals is known in detail (15, 16). Two elderly males aged 41 (Evered and Huxley) died of infected wounds, which I represented as infections. The oldest death (Flo, aged 43) is described as "wasting ... likely secondary to senescence" (15).

[b]Hunter-gatherers and forager-farmers with limited access to modern medicine (8); 7 groups in the 20th century. Senescent deaths were scored for those >60 y, which may have included infections.

Subsequent well-maintained colonies on more standardized diets are puzzlingly divergent for blood lipids: chimpanzees at Yerkes were hypercholesterolemic (44), whereas those at Phoenix had normal cholesterol (45). The Phoenix colony also reported changes in LDL and HDL subfractions that were offsetting in risk by clinical criteria. Other cardiovascular risk indicators included elevated fibrinogen, insulin, and Lp(a); the latter is a species difference, due to increased transcription of the *Lp(a)* gene (46). Markers of oxidation in blood-cell DNA and lipids were higher, though some antioxidants were lower relative to healthy young men. Despite these risk indicators, ischemic coronary artery disease has *not* been identified as the main cause of death in 3 other modern colonies, where most sudden deaths were attributed to congestive heart failure from fibrillation in association with myocardial fibrosis: Yerkes (44), Almagordo (47), and Southwest Foundation (48). Ischemic arterial disease was considered minor in most sudden deaths in these well-maintained colonies, in contrast to the earlier reports. However, myocardial fibrosis was also common in early colonies (40, 41).

Cancer. Chimpanzees and other primates in captivity appear to develop much less neoplasia than humans, as noted in earlier reviews (49–51) and supported by recent studies. To a first approximation, neoplasia was detected in <3% of adults up through older ages. Female chimpanzees from Yerkes and Southwest Foundation had more neoplasia than males, with notable prevalence of uterine leiomyomas (52). The leiomyomas and most other tumors were benign and arose after age 25. Remarkably, no spontaneous mammary carcinoma has been reported in the great apes. In males 25 years and older, benign prostatic hyperplasia is common, and associated with clinical-grade blood prostate-specific

antigen and urinary retention (53). Though prostate neoplasia was not reported (52, 54), later ages need study.

Other primate species also present a low incidence of neoplasia. Adult baboons and monkeys had <3% prevalence at necropsy of >10,000 animals from 3 colonies that included older ages (55–59). Prosimian neoplasia is similar, 1%–3% of adults (60). In view of the absence of mammary carcinoma in chimpanzees, the documentation of mammary carcinoma in prosimians and monkeys (references in ref. 60) give a mandate for continuing surveillance of aging chimpanzees. Provisionally, primate colonies have lower prevalence of malignancy than most modern human populations, e.g., the U.S. lifetime cancer risk was about 40% in 2007 (61). However, the above studies did not present data on the population at risk by age, needed for comparison with human populations. The surviving aging great apes are a vanishing resource because active breeding has been stopped in U.S. colonies. The absence of pregnancy also eliminates a protective factor for breast cancer in humans. There may be no way to obtain autopsy data on wild populations without supporting the bushmeat trade.

The paleopathology of neoplasia may only be approached in bone tumors, which persist in graveyard and fossil skeletons (62). In a large sample of adult bone (>3,500) specimens from pre-Roman Egypt and medieval Germany, about 0.5% of individuals at both sites had macroscopic tumors (>3,500 specimens), similar to that of England in 1900 (63). For comparison, 4,000 baboon autopsies yielded 1 osteoma and 1 osteosarcoma, suggesting a prevalence of <0.1% (55), again consistent with a lower incidence of other types of malignancies in primates than humans. I have not found reports on pathologically confirmed bone tumors in prehistoric human fossils.

Neurodegeneration. The neuropathology of aging in great apes is also surprising. Detailed studies of brains from chimpanzee, gorilla, and orangutan of 40 years or older concur on the rarity of Alzheimer-like neurodegenerative changes of neuronal loss, neuritic plaques (dense amyloid plaques with neuritic degeneration), and neurofibrillary degeneration with tau immunoreactivity (1, 64, 65). In contrast to the great apes, aging monkeys and a prosimian have shown more neurodegenerative changes with varying degrees of neurocytoskeletal abnormalities and amyloid deposits (1, 64–66) and cerebral atrophy (67, 68). Nonetheless, it was recent reported that a 41-year-old chimpanzee died after a stroke with the classic tau-positive neurofibrillary tangles with paired helical filaments (69). This individual also had obesity and chronic hypercholesterolemia. Despite the neurofibrillary tangles, other brain changes were mild: the diffuse amyloid deposits and the absence of major neuronal loss and neuritic plaques do not meet neuropathological criteria for Alzheimer's disease.

Possibly, hypercholesterolemia may promote a subset of Alzheimer-like changes in chimpanzees under some circumstances. In humans, the epidemiological and clinical links of obesity and blood cholesterol to Alzheimer's disease are complex and controversial. Variations of trace elements could be a factor. In rodent and rabbit models of Alzheimer's disease on cholesterol-rich diets, trace iron intake may be a critical variable (70). Lead can also promote later formation of amyloid deposits in monkeys (71). However, none of the rodent or primate models has shown the extensive neuronal loss characteristic of human Alzheimer's disease by early clinical stages. Thus, Alzheimer's disease may be a uniquely human neurodegenerative pathway of aging.

Other age-related changes. Wild chimpanzees of 25 years have increasingly frequent decrepit appearances from bone fractures, skin wounds, tooth loss, weight loss, and difficulty climbing (16, p 104). Degenerative osteoarthitic changes are indicated in some samples. In adult skeletons from Kibale, 75% had some degenerative joint disease, most severe in older females; 65% showed traumatic bone injury from fractures and bite punctures (72). Similarly, an early 20th-century sample from West Africa had prevalent erosive osteoarthritis (73). A Gombe sample, however, had minimal spinal osteoarthritis (74). The uncertain ages and small samples preclude comparisons with humans.

Female reproductive senescence with follicular depletion (menopause) occurs by 50 in chimpanzees in natural populations and in captivity (9, 75–78). Nonetheless, wild females are fertile up through at least 42 years (16). Thus, few if any female chimpanzees survive to reproductive senescence in natural populations. By contrast, most hunter-gatherer females reaching adulthood survive beyond menopause (6, 8, 9). The extended postmenopausal phase also uniquely exposes humans to osteoporotic fractures from low estrogen that are not reported for great apes.

Male reproductive aging is undefined: besides benign prostatic hyperplasia (53), there is no report on how male age influences spermatogenesis or sperm quality. The social hierarchies that determine access to females are dominated by prime-age adult males typically in the late teens to late twenties (16, fig. 15.2); the upper ages overlap the onset of benign prostatic hyperplasia in captive males (53).

Summary on Aging in Chimpanzees. The indications of faster aging in chimpanzees than in humans by the earlier acceleration of mortality require corroboration by age-specific changes in pathology and organ function. Because menopause occurs at about the same age, 50, reproductive declines may be relatively delayed in female chimpanzees. The emerging profile of pathology in aging captive chimpanzees suggests the importance of environmental and husbandry variables for myocardial and brain aging, in which blood cholesterol and trace metals could be important. The low prevalence of ischemic heart disease in modern colonies may represent improvements of husbandry, but the scattered data from earlier colonies do not allow firm conclusions. However, for cancer and myocardial pathology, age-specific rate data are needed for comparison with human aging. Measures of cardiopulmonary function and immunosenescence in captive colonies will also be informative. Ongoing studies of the relict hunter-foragers with limited access to modern medicine (6–8) may be our best basis for comparison with wild chimpanzees. The rarity of malignancy and myocardial infarction, and the absence of Alzheimer's disease, in chimpanzees may prove to be real species differences. Conversely, it is important to know whether the diffuse interstitial fibrosis of aging chimpanzees also occurs in some human populations.

Diet. During human evolution, the diet has shifted to increased consumption of animal tissues, although plant-based foods have always been important (1, 2, 79). The advantages of meat-rich diets include higher density caloric content (reducing efforts in foraging and digestion), and concentrated micronutrients (trace metals and polyunsaturated fatty acids required for optimum development of the musculature and nervous system). However, increased trace metals and fat ingestion could also interact with pathogensis, as noted previously. The greater meat consumption of longer-lived humans than great ape ancestors presents a paradox because in many animal models of human disease and longevity, greater fat and caloric intake is associated with accelerated pathogenesis and shortened lifespan (1, 2). For example, caloric restriction of Alzheimer's transgenic mice attenuated the deposition of brain amyloid and glial reactions (78). Similarly, caloric restriction attenuates atherosclerosis, diabetes, and neoplasia in animal models (2). Moreover, in rodents, caloric restriction slows most aging changes and extends lifespan in proportion to lower intake, over a range of 10%–40%. Conversely, higher fat intake can exacerbate disease in models of atherogenesis, Alzheimer's disease, and neoplasia.

Changes in diet also increased exposure to pathogens and toxins. Uncooked meat, particularly from scavenged old carcasses, would have increased exposure to infectious pathogens. Though cooking can kill most pathogens and increases the digestibility of meat and fibrous plant material (79), cooking also

accelerates nonenzymatic glyco-oxidation to form advanced glycation endproducts (AGEs) that are diabetogenic and proatherosclerotic in animal models and in clinical studies (80, 81). How did humans evolve increased longevity despite the greater fat intake and exposure to pathogens? Finch and Stanford (1) proposed that the diet and longevity shifts during the evolution were supported by meat-adaptive genes, with tradeoffs of mortality and for ingestion of fat and toxins, and pathogen exposure.

Genetic Changes. Before considering specific genes, it is notable that a small part of the DNA difference between humans and chimpanzees shows evidence of positive selection. Though there is 4% DNA sequence divergence, most (*ca.* 3%), represents insertion-deletions (90 Mb difference between species) (82–84). The genome-wide single nucleotide (nt) differences are 1.23%, of which approximately 18% is within-species polymorphisms; thus, the fixed divergence at the species level for proteins is about 1% (82). Genes undergoing positive selection based on the ratio of nonsynonymous:synonymous mutations are overrepresented for immunity and host defense, diet, and brain (85). Moreover, genes associated with immunity and brain have variation clusters of highly localized groups of changes in coding regions (86).

de Magalhães and Church (87) examined human and chimpanzee genomes for longevity gene orthologs from short-lived animal models. Surprisingly, the aging-associated genes had less variation than the average, implying slower evolutionary change in the human lineage, e.g., of *IGF1* and its receptor *IGFR1*, in which loss-of-function mutations increase mouse lifespan (2). Greater coding sequence divergence was observed in *WRN* (Werner's progeroid syndrome), but not other progeria genes, and genes associated with responses to pest/pathogens/parasites (Gene Ontology database accession no. GO:0009613).

The high incidence of neoplasia in humans is not explained so far by DNA sequences. Of 333 cancer-associated genes, the majority are almost identical in chimpanzees (88, 89). The human *BRCA1* has an unusual number of *Alu* repeats that cause gene instability, whereas the chimpanzee *BRCA1* has an 8-kb deletion that truncates the coregulated *NBR2* gene. The breast cancer oncogenes *BRCA2* and *ERBB2* have multiple alleles, whereas chimpanzees have only the lower-cancer-risk human alleles. *BRCA1* also shows evidence for positive selection at the coding level and Hardy-Weinburg disequilibrium in human populations. Influences of *BRCA1* and *-2* alleles on early growth imply tradeoffs for growth and DNA repair relevant to the uniquely human pattern of early breast development with antagonist pleiotropy of later neoplasia (89).

Host defense system genes show evidence for positive selection, as noted. The most details may be available for the major histocompatibility complex (MHC) and sialic acid-binding Ig-like lectins (Siglecs). The MHC system is fundamental to innate and adaptive immunity: its >1,000 genes on CH6 are the most polymorphic of any gene system, particularly for variations in the peptide binding sites that determine host resistance. A major species difference is the loss of polymorphisms in class I*A* and *B* genes. Because the remaining MHC classes had equivalent variety, this class-specific loss of variation suggests a selective sweep (90). An MHC class I peptide that presents a SIV gag peptide to cytotoxic T lymphocytes in macaques (91) could be a target of selection in the ongoing SIVcpz infections noted earlier. There is no easy test of the adaptiveness of the numerous allele differences.

Differences in the Siglec lectin family of proteins (Ig superfamily) cell-surface glycoproteins have specific implications for host-defense evolution in studies from Varki and coworkers (92, 93). Siglecs bind the sialic acids on cell surfaces of macrophages and other immune-related cells. Siglec genes appear to have evolved very rapidly, because there is a much smaller divergence between mice and rats, which had a more distant common ancestor. Human-specific changes arose in at least 10 genes of the 50+ genes involved in sialobiology. Human $CD4^+$ T cells have low expression of Siglecs relative to chimpanzees (93). Siglec-5 manipulation switched the species-type response to T-cell receptor (TCR) stimulation. This species difference may be a factor in T-cell-mediated diseases, including the much milder chimpanzee disease from HIV-1 and hepatitis B or C (94), and the apparent lack of spontaneous rheumatoid arthritis and bronchial asthma. These differences in immunoreactivity could involve the weaker expression of CD33rSiglecs of humans, relative to great apes (93). Siglecs also modulate *Streptococcus* invasiveness (GBS) (95). Direct species comparisons are needed of immune cell responses to specific pathogens and of transcriptomes and kinomes.

Humans also differ by the absence of *N*-glycolylneuraminic acid (Neu5Gc), a major sialic acid of chimpanzees and other great apes (92, 95). A mutation that occurred early in the genus *Homo*, at least before 0.5 M, inactivated the CMAH enzyme that produces Neu5Gc from its precursor Neu5Ac, with implications of Neu5Ac targeting human-specific pathogens. For example, the chimpanzee malarial parasite has a protein that binds preferentially to Neu5Gc during erythrocyte invasion, whereas that of the human parasite *P. falciparum* binds Neu5Ac. The extant *P. falciparum* is likely to have arisen from a single chimpanzee-to-human transfer event (96). The evolving human diet could also have had a role in these complex immunological scenarios, because normal tissues have traces of Neu5Gc, which may be acquired from ingestion of red meat and milk; this could stimulate chronic inflammation induced by anti-Neu5Gc antibodies and also facilitate metastasis (97).

Lastly, I consider the apolipoprotein E (*ApoE*) alleles, which modulate chronic inflammation and many aspects of aging in brain and arteries and which Sapolsky, Stanford, and I (1, 2, 98) have proposed as a meat-adaptive candidate gene in the increases of the human lifespan. Blood apoE mediates the clearance of triglyceride-rich lipoprotein components, and brain apoE transports cholesterol to neurons (100). *ApoE4*, the minor allele in all human populations (<1%–45%), is considered ancestral in the genus *Homo* (99, 101). The uniquely human *apoE3* allele spread about 0.226 million years ago (MYA), range 0.18–0.58 MYA (101). These dates precede the emigration of modern *H. sapiens* from Africa and overlap with the increased organized hunting of large animals and the use of fire (102).

In general, the *apoE4* allele shortens lifespan by several years and accelerates degenerative changes in arteries and brain (2, 99, 100, 103, 104). *ApoE4* carriers have modestly higher total blood cholesterol, more oxidized blood lipids, and greater risk of coronary heart disease (*ca.* 40%) and Alzheimer's disease (depending on the population, *E4*/*E4* homozygotes have >10-fold excess risk).

Table 3. Apolipoprotein E polymorphisms in humans and species differences

ApoE residue (mature peptide)	61[a]	112	158
Human apoE3	R	C	R
ApoE4	R	R	R
Chimpanzee	T	R	R
Gorilla	T	R	R
Orangutan	T	R	R
Mouse	T	R	R

[a]Residue 61 determines apoE protein structure by domain interactions that influence lipid binding by the C terminus (1, 100, 108). Though chimpanzees, other primates, and many mammals have the R112 and R158 that define apoE4, these species differ from human apoE at residue 61. Genetically engineering the mouse apoE with R61T changed lipid-binding affinity to resemble human apoE4 (108). Thus, chimpanzee apoE is predicted to have lipid binding like apoE3 (1). Nonetheless, other amino acid differences from the chimpanzee may be important, because 4 of the 8 residues that showed evidence of positive selection in the human lineage are seated in the lipid-binding C terminus (109).

ApoE4 carriers also have worse outcomes in traumatic brain injury and some neurological conditions. One mechanism may involve heightened inflammatory responses. In transgenic mice with targeted gene (TR) replacement of human apoE alleles, the TR-*apoE4* mice have monocyte reactivity (IL-6, IFN, NO, TNFα) and greater bystander damage to neurons (105). On a fatty diet, TR-*apoE4* mice had larger adipocytes and impaired glucose tolerance (106); however, obesity and diabetes have not shown consistent *apoE* allele associations. Subcellularly, apoE4 causes more lysosomal leakage than apoE3, due to greater membrane disruption from peptide chain unfolding at lysosomal pH ("molten globule") (107); this biophysical feature of apoE4 is unique to humans and is implicated in the greater neurotoxicity of β-amyloid in *apoE4* transgenic models of Alzheimer's disease (100).

Though the chimpanzee apoE has 2 amino acids like apoE4, it is predicted to function like the human apoE3 isoform because of a further coding difference that influences peptide folding (108, 109) (Table 3). The putative apoE3-like function could contribute to the low levels of Alzheimer's and ischemic heart disease in chimpanzees noted previously. Although chimpanzee apoE has not shown allelic variation in small samples (101), serum cholesterol had considerable heritability in a former breeding colony (110).

Besides influencing brain aging, *apoE* alleles also affect brain development. Cortical neurons of young TR-*apoE4* mice have less dendritic complexity (111), which may be a factor in their impaired spatial memory (112). ApoE alleles are increasingly included in studies of human development. In MRI studies of healthy juveniles, the *apoE4* carriers had a thinner entorhinal cortex (113). This regional growth difference is relevant to Alzheimer's disease, which causes early damage in the entorhinal cortex. In sum, the *apoE* alleles are remarkable for the range of pleiotropies on blood cholesterol, immune responses, brain development, and arterial and Alzheimer's disease. As a hedge against overinterpretations of these broad effects, it may be reassuring that apoE alleles have not shown consistent associations with fertility or neoplasia (2).

Given these adverse effects of *apoE4*, at least in modern environments, the persistence of the allele has been proposed as the result of balancing selection, as in malarial protection by heterozygotes of hemoglobinopathies (1, 98). Two examples are under discussion, for which the evidence must be considered as preliminary. In hepatitis C infections, *apoE4* carriers incurred less fibrotic damage by allele dose (114, 115), whereas Brazilian slum children carrying *apoE4* showed less diarrhea and associated impairments of cognitive development (116, 117).

The hyperreactivity of human T cells noted previously, and the inflammatory responses in apoE4 carriers, may be part of an evolved group of heightened immune defenses relative to great apes that decreased baseline mortality represented in the q_{min}, as discussed earlier. However, the heightened immune responses could then have delayed adverse effects in cardiovascular disease and other chronic conditions of aging that involve inflammation (2) and that became more prevalent in the 20th century. This suggestion extends the antagonistic pleiotropy theory of aging in which genes selected for early advantages can have delayed adverse effects that are under weaker selection. The unique human social system of multigenerational support in child nurture has been argued as a key factor in the selection for delayed disability and increased life expectancy at later ages (1, 2, 7–9, 98).

ACKNOWLEDGMENTS. The author is grateful for valuable comments from Kurt Benirshcke and Agit and Nissi Varki. Support was provided by the National Institute on Aging and the Ellison Medical Foundation.

1. Finch CE, Stanford CB (2004) Meat-adaptive genes and the evolution of slower aging in humans. *Q Rev Biol* 79:3–50.
2. Finch CE (2007) *The Biology of Human Longevity. Inflammation, Nutrition, and Aging in the Evolution of Lifespans* (Academic, San Diego).
3. Oeppen J, Vaupel JW (2002) Demography. Broken limits to life expectancy. *Science* 296:1029–1031.
4. Finch CE, Crimmins EM (2004) Inflammatory exposure and historical changes in human life-spans. *Science* 305:1736–1739.
5. Caspari R, Lee SH (2006) Is human longevity a consequence of cultural change or modern biology? *Am J Phys Anthropol* 129:512–517.
6. Kaplan HS, Hill K, Lancaster JB, Hurtado AM (2000) A theory of human life history evolution: Diet, intelligence, and longevity. *Evol Anthropol* 9:156–183.
7. Gurven M, Kaplan H (2007) Hunter-gatherer longevity: Cross-cultural perspectives. *Popul Dev Rev* 33:321–365.
8. Hawkes K, Smith KR, Robson SL (2009) Mortality and fertility rates in humans and chimpanzees: How within-species variation complicates cross-species comparisons. *Am J Hum Biol* 21:578–586.
9. Hawkes K (2004) Human longevity: The grandmother effect. *Nature* 428:128–129.
10. Finch CE, Pike MC, Witten M (1990) Slow mortality rate accelerations during aging in some animals approximate that of humans. *Science* 249:902–905.
11. Hill K, et al. (2001) Mortality rates among wild chimpanzees. *J Hum Evol* 40:437–450.
12. Crimmins EM, Drevenstedt G, Finch CE (2007) Evolution of the mortality curve: Changes in the age of minimum mortality. Paper presented at the Annual Meeting of the Population Association of America, March 27, 2007. Abstract available at http://paa2007.princeton.edu/download.aspx?submissionId=71579.
13. Rosen RD (December 7, 2008) The lie of the jungle. *Washington Post,* p W14.
14. Jeune B, et al. (2009) Jeanne Calment's and her successors. Biographical notes on the longest living humans. Supercentenarians, ed Maier H (Springer, Berlin).
15. Williams JM, et al. (2008) Causes of death in the Kasekela chimpanzees of Gombe National Park, Tanzania. *Am J Primatol* 70:766–777.
16. Goodall J (1986) *The Chimpanzees of Gombe: Patterns of Behavior* (Harvard Univ Press, Boston).
17. Keele BF, et al. (2009) Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. *Nature* 460:515–519.
18. Hopkins ME, Nunn CL (2007) A global gap analysis of infectious agents in wild primates. *Divers Distrib* 13:561–572.
19. Gurven M, Kaplan H, Supa AZ (2007) Mortality experience of Tsimane Amerindians of Bolivia: Regional variation and temporal trends. *Am J Hum Biol* 19:376–398.
20. Black FL (1975) Infectious diseases in primitive societies. *Science* 187:515–518.
21. Parkin TG (1992) *Demography in Roman Society* (Johns Hopkins Press, Baltimore).
22. Scheidel W (2007) Demography. *The Cambridge History of the Greco-Roman World*, Scheidel W, Morris I, Saller R (Cambridge Univ Press, Cambridge, UK), pp 38–86.
23. Hoppa RD, Vaupel JW (2002) The Rostock manifesto for paleodemography. *Paleodemography: Age Distributions from Skeletal Samples* (Cambridge Studies in Biological and Evolutionary Anthropology; 31), Hoppa RD, Vaupel JW, eds (Cambridge Univ Press, Cambridge, UK).
24. Preston SH (1976) *Mortality Patterns in National Populations, with Special Reference to Recorded Causes of Death* (Academic, San Diego).
25. Crimmins EM, Finch CE (2006) Infection, inflammation, height, and longevity. *Proc Natl Acad Sci USA* 103:498–503.
26. Hadrup SR, et al. (2006) Longitudinal studies of clonally expanded CD8 T cells reveal a repertoire shrinkage predicting mortality and an increased number of dysfunctional cytomegalovirus-specific T cells in the very elderly. *J Immunol* 176:2645–2653.
27. Parish ST, Wu JE, Effros RB (2009) Modulation of T lymphocyte replicative senescence via TNF-α inhibition: Role of caspase-3. *J Immunol* 182:4237–4243.
28. Pita-Lopez ML, et al. (2009) Effect of ageing on CMV-specific CD8 T cells from CMV seropositive healthy donors. *Immun Ageing* 6:11.
29. Desquilbet L, et al.; Multicenter AIDS Cohort Study (2009) Relationship between a frailty-related phenotype and progressive deterioration of the immune system in HIV-infected men. *J Acquir Immune Defic Syndr* 50:299–306.
30. Palinski W, Nicolaides E, Liguori A, Napoli C (2009) Influence of maternal dysmetabolic conditions during pregnancy on cardiovascular disease. *J Cardiovasc Transl Res* 2:277–285.
31. Mazumder B, Almond D, Park K, Crimmins EM, Finch CE (2009) Lingering prenatal effects of the 1918 Influenza Pandemic on cardiovascular disease. *J Devel Origin Health Disease* 1:1–9.
32. Spyridopolous I, et al. (2009) Accelerated telomere shortening in leukocyte subpopulations of patients With coronary heart disease. Role of cytomegalovirus seropositivity. *Circulation* 120(14):1364–1372.
33. Murphy WA, Jr, et al. (2003) The iceman: Discovery and imaging. *Radiology* 226: 614–629.
34. Ruffer M (1911) On arterial lesions found in Egyptian mummies. *J Pathol Bacteriol* 15:453–462.
35. Bruetsch WL (1959) The earliest record of sudden death possibly due to atherosclerotic coronary occlusion. *Circulation* 20:438–441.
36. Blaha M, et al. (2009) Absence of coronary artery calcification and all-cause mortality. *JACC Cardiovasc Imaging* 2:692–700.
37. Rennenberg RJ, et al. (2009) Vascular calcifications as a marker of increased cardiovascular risk: A meta-analysis. *Vasc Health Risk Manag* 5:185–197.
38. Cockburn A, Barraco RA, Reyman TA, Peck WH (1975) Autopsy of an Egyptian mummy. *Science* 187:1155–1160.
39. Magee R (1998) Arterial disease in antiquity. *Med J Aust* 169:663–666.
40. Andrus SB, Portman OW, Riopelle AJ (1968) Comparative studies of spontaneous and experimental atherosclerosis in primates. II lesions in chimapnzees, including myocardial infarction and cerebral aneurysms. *Prog Biochem Pharmacol* 4:393–419.
41. Ratcliffe HL (1965) Age and the environment as factors in the nature and frequency of cardiovascular lesions in mammals and birds in the Philadelphia Zoological Gardens. *Ann N Y Acad Sci* 7:715–735.

42. Manning GW (1942) Coronary disease in the ape. *Am Heart J* 23:719–724.
43. Blaton V, et al. (1974) Dietary induced hyperbetalipoproteinemia in chimpanzees: Comparison to the human hyperlipoproteinemia. *Exp Mol Pathol* 20:132–146.
44. Varki N, et al. (2009) Heart disease is common in humans and chimpanzees, but is caused by different pathological processes. *Evol Appl* 2:101–112.
45. Videan EN, et al. (2009) Comparison of biomarkers of oxidative stress and cardiovascular disease in humans and chimpanzees (Pan troglodytes). *Comp Med* 59:287–296.
46. Huby T, et al. (2001) Functional analysis of the chimpanzee and human apo(a) promoter sequences: Identification of sequence variations responsible for elevated transcriptional activity in chimpanzee. *J Biol Chem* 276:22209–22214.
47. Lammey ML, et al. (2008) Interstitial myocardial fibrosis in a captive chimpanzee (Pan troglodytes) population. *Comp Med* 58:389–394.
48. Seiler BM, et al. (2009) Spontaneous heart disease in the adult chimpanzee (Pan troglodytes). *J Med Primatol* 38:51–58.
49. Varki AA (2000) A chimpanzee genome project is a biomedical imperative. *Genome Res* 10:1065–1070.
50. Beniashvili DS (1989) An overview of the world literature on spontaneous tumors in nonhuman primates. *J Med Primatol* 18:423–437.
51. Lapin BA (1982) Use of nonhuman primates in cancer research. *J Med Primatol* 11:327–341.
52. Brown SL, et al. (2009) Neoplasia in the chimpanzee (Pan spp.). *J Med Primatol* 38:137–144.
53. Steiner MS, Couch RC, Raghow S, Stauffer D (1999) The chimpanzee as a model of human benign prostatic hyperplasia. *J Urol* 162:1454–1461.
54. Waters DJ, et al. (1998) Workgroup 4: Spontaneous prostate carcinoma in dogs and nonhuman primates. *Prostate* 36:64–67.
55. Cianciolo RE, Hubbard GB (2005) A review of spontaneous neoplasia in baboons (Papio spp.). *J Med Primatol* 34:51–66.
56. Cianciolo RE, et al. (2007) Spontaneous neoplasia in the baboon (Papio spp.). *J Med Primatol* 36:61–79.
57. Baskerville A, Cook RW, Dennis MJ, Cranage MP, Greenaway PJ (1992) Pathological changes in the reproductive tract of male rhesus monkeys associated with age and simian AIDS. *J Comp Pathol* 107:49–57.
58. Cline JM (2004) Neoplasms of the reproductive tract: The role of hormone exposure. *ILAR J* 45:179–188.
59. Kaspareit J, Friderichs-Gromoll S, Buse E, Habermann G (2007) Spontaneous neoplasms observed in cynomolgus monkeys (Macaca fascicularis) during a 15-year period. *Exp Toxicol Pathol* 9:163–169.
60. Remick AK, Van Wettere AJ, Williams CV (2009) Neoplasia in prosimians: Case series from a captive prosimian population and literature review. *Vet Pathol* 46:746–772.
61. American Cancer Society, 2007.
62. Waldron T (1996) What was the prevalence of malignant disease in the past? *Int J Osteoarcheol* 6:463–470.
63. Nerlich AG, Bachmeier BE (2007) Paleopathology of malignant tumours supports the concept of human vulnerability to cancer. *Nat Rev Cancer* 7:563.
64. Schultz C, Hubbard GB, Rüb U, Braak E, Braak H (2000) Age-related progression of tau pathology in brains of baboons. *Neurobiol Aging* 21:905–912.
65. Erwin JM, Nimchinsky EA, Gannon PJ, Perl DP, Hof PR (2000) The study of brain aging in great apes. *Functional Neurobiology of Aging*, Hof PR, Mobbs CV, eds (Academic, San Diego), pp 447–456.
66. Gearing M, Rebeck GW, Hyman BT, Tigges J, Mirra SS (1994) Neuropathology and apolipoprotein E profile of aged chimpanzees: Implications for Alzheimer disease. *Proc Natl Acad Sci USA* 91:9382–9386.
67. Kraska A, et al. (2009) Age-associated cerebral atrophy in mouse lemur primates. *Neurobiol Aging*, in press.
68. Colman RJ, et al. (2009) Caloric restriction delays disease onset and mortality in rhesus monkeys. *Science* 325:201–204.
69. Rosen RF, et al. (2008) Tauopathy with paired helical filaments in an aged chimpanzee. *J Comp Neurol* 509:259–270.
70. Sparks DL, et al. (2006) Trace copper levels in the drinking water, but not zinc or aluminum influence CNS Alzheimer-like pathology. *J Nutr Health Aging* 10:247–254.
71. Wu J, et al. (2008) Alzheimer's disease (AD)-like pathology in aged monkeys after infantile exposure to environmental metal lead (Pb): Evidence for a developmental origin and environmental link for AD. *J Neurosci* 28:3–9.
72. Carter ML, Pontzer H, Wrangham RW, Peterhans JK (2008) Skeletal pathology in Pan troglodytes schweinfurthii in Kibale National Park, Uganda. *Am J Phys Anthropol* 135:389–403.
73. Rothschild BM, Woods RJ (1991) Reactive erosive arthritis in chimpanzees. *Am J Primatol* 25:49–56.
74. Jurmain R (2000) Degenerative joint disease in African great apes: An evolutionary perspective. *J Hum Evol* 39:185–203.
75. Lacreuse A, et al. (2008) Menstrual cycles continue into advanced old age in the common chimpanzee (Pan troglodytes). *Biol Reprod* 79:407–412.
76. Emery Thompson M, et al. (2007) Aging and fertility patterns in wild chimpanzees provide insights into the evolution of menopause. *Curr Biol* 17:2150–2156.
77. Atsalis S, Videan E (2009) Reproductive aging in captive and wild common chimpanzees: Factors influencing the rate of follicular depletion. *Am J Primatol* 71:271–282.
78. Patel NV, et al. (2005) Caloric restriction attenuates Abeta-deposition in Alzheimer transgenic models. *Neurobiol Aging* 26:995–1000.
79. Wobber V, Hare B, Wrangham R (2008) Great apes prefer cooked food. *J Hum Evol* 55:340–348.
80. Zhao Z, et al. (2009) Advanced glycation end products inhibit glucose-stimulated insulin secretion through nitric oxide-dependent inhibition of cytochrome c oxidase and adenosine triphosphate synthesis. *Endocrinology* 150:2569–2576.
81. Cai W, et al. (2008) Oral glycotoxins determine the effects of calorie restriction on oxidant stress, age-related diseases, and lifespan. *Am J Pathol* 173:327–336.
82. Chimpanzee Sequencing and Analysis Consortium (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437:69–87.
83. Kehrer-Sawatzki H, Cooper DN (2007) Understanding the recent evolution of the human genome: Insights from human-chimpanzee genome comparisons. *Hum Mutat* 28:99–130.
84. Varki A, Geschwind DH, Eichler EE (2008) Explaining human uniqueness: Genome interactions with environment, behaviour and culture. *Nat Rev Genet* 9:749–763.
85. Arbiza L, Dopazo J, Dopazo H (2006) Positive selection, relaxation, and acceleration in the evolution of the human and chimp genome. *PLOS Comput Biol* 2:e38.
86. Wagner A (2007) Rapid detection of positive selection in genes and genomes through variation clusters. *Genetics* 176:2451–2463.
87. de Magalhães JP, Church GM (2007) Analyses of human-chimpanzee orthologous gene pairs to explore evolutionary hypotheses of aging. *Mech Ageing Dev* 128:355–364.
88. Puente XS, et al. (2006) Comparative analysis of cancer genes in the human and chimpanzee genomes. *BMC Genomics* 7:15.
89. Crespi BJ, Summers K (2006) Positive selection in the evolution of cancer. *Biol Rev Camb Philos Soc* 81:407–424.
90. de Groot NG, et al. (2008) Pinpointing a selective sweep to the chimpanzee MHC class I region by comparative genomics. *Mol Ecol* 17:2074–2088.
91. Andrade MC, Leite JP, Cabello PH (2009) Frequency of the major histocompatibility complex Mamu-A*01 allele in a closed breeding colony of rhesus monkey (Macaca mulatta) from Brazil. *J Med Primatol* 38:39–41.
92. Varki A (2009) Multiple changes in sialic acid biology during human evolution. *Glycoconj J* 26:231–245.
93. Nguyen DH, Hurtado-Ziola N, Gagneux P, Varki A (2006) Loss of Siglec expression on T lymphocytes during human evolution. *Proc Natl Acad Sci USA* 103:7765–7770.
94. Bibollet-Ruche F, et al. (2008) The quality of chimpanzee T-cell activation and simian immunodeficiency virus/human immunodeficiency virus susceptibility achieved via antibody-mediated T-cell receptor/CD3 stimulation is a function of the anti-CD3 antibody isotype. *J Virol* 82:10271–10278.
95. Carlin AF, et al. (2009) Molecular mimicry of host sialylated glycans allows a bacterial pathogen to engage neutrophil Siglec-9 and dampen the innate immune response. *Blood* 113:3333–3336.
96. Rich SM, et al. (2009) The origin of malignant malaria. *Proc Natl Acad Sci USA* 106: 14902–14907.
97. Hedlund M, Padler-Karavani V, Varki NM, Varki A (2008) Evidence for a human-specific mechanism for diet and antibody-mediated inflammation in carcinoma progression. *Proc Natl Acad Sci USA* 105:18936–18941.
98. Finch CE, Sapolsky RM (1999) The evolution of Alzheimer disease, the reproductive schedule, and apoE isoforms. *Neurobiol Aging* 20:407–428.
99. Finch CE, Morgan TE (2007) Systemic inflammation, infection, ApoE alleles, and Alzheimer disease: A position paper. *Curr Alzheimer Res* 4:185–189.
100. Mahley RW, Weisgraber KH, Huang Y (2009) Apolipoprotein E: Structure determines function, from atherosclerosis to Alzheimer's disease to AIDS. *J Lipid Res* 50(Suppl): S183–S188.
101. Fullerton SM, et al. (2000) Apolipoprotein E variation at the sequence haplotype level: Implications for the origin and maintenance of a major human polymorphism. *Am J Hum Genet* 67:881–900.
102. Goren-Inbar N, et al. (2004) Evidence of hominin control of fire at Gesher Benot Ya'aqov, Israel. *Science* 304:725–727.
103. Little DM, et al. (2009) Mortality, dementia, and apolipoprotein E genotype in elderly white women in the United States. *J Am Geriatr Soc* 57:231–236.
104. Stephens JW, Bain SC, Humphries SE (2008) Gene-environment interaction and oxidative stress in cardiovascular disease. *Atherosclerosis* 200:229–238.
105. Vitek MP, Brown CM, Colton CA (2009) APOE genotype-specific differences in the innate immune response. *Neurobiol Aging* 30:1350–1360.
106. Arbones-Mainar JM, Johnson LA, Altenburg MK, Maeda N (2008) Differential modulation of diet-induced obesity and adipocyte functionality by human apolipoprotein E3 and E4 in mice. *Int J Obes (Lond)* 32:1595–1605.
107. Ji ZS, et al. (2006) Reactivity of apolipoprotein E4 and amyloid beta peptide: Lysosomal stability and neurodegeneration. *J Biol Chem* 281:2683–2692.
108. Raffai RL, Dong LM, Farese RV, Jr, Weisgraber KH (2001) Introduction of human apolipoprotein E4 "domain interaction" into mouse apolipoprotein E. *Proc Natl Acad Sci USA* 98:11587–11591.
109. Vamathevan JJ, et al. (2008) The role of positive selection in determining the molecular cause of species differences in disease. *BMC Evol Biol* 8:273.
110. Williams-Blangero S, Butler T, Brasky K, Murthy KK (1994) Heritabilities of clinical chemical traits in chimpanzees. *Lab Anim Sci* 44:141–143.
111. Wang C, et al. (2005) Human apoE4-targeted replacement mice display synaptic deficits in the absence of neuropathology. *Neurobiol Dis* 18:390–398.
112. Grootendorst J, et al. (2005) Human apoE targeted replacement mouse lines: h-apoE4 and h-apoE3 mice differ on spatial memory performance and avoidance behavior. *Behav Brain Res* 159:1–14.
113. Shaw P, et al. (2007) Cortical morphology in children and adolescents with different apolipoprotein E gene polymorphisms: An observational study. *Lancet Neurol* 6:494–500.
114. Wozniak MA, et al.; Trent HCV Study Group (2002) Apolipoprotein E-epsilon 4 protects against severe liver disease caused by hepatitis C virus. *Hepatology* 36:456–463.
115. Fabris C, et al. (2005) Low fibrosis progression of recurrent hepatitis C in apolipoprotein E epsilon4 carriers: Relationship with the blood lipid profile. *Liver Int* 25:1128–1135.
116. Oriá RB, et al. (2005) APOE4 protects the cognitive development in children with heavy diarrhea burdens in Northeast Brazil. *Pediatr Res* 57:310–316.
117. Oriá RB, Costa CM, Lima AA, Patrick PD, Guerrant RL (2009) Semantic fluency: A sensitive marker for cognitive impairment in children with heavy diarrhea burdens? *Med Hypotheses* 73:682–686.

Somatic evolutionary genomics: Mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration

Steven A. Frank[1]

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697-2525

Edited by Stephen Curtis Stearns, Yale University, New Haven, CT, and approved August 19, 2009 (received for review August 17, 2009)

Somatic mutations must happen often during development because of the large number of cell divisions to expand from a single-cell zygote to a full organism. A mutation in development carries forward to all descendant cells, causing genetic mosaicism. Widespread genetic mosaicism may influence diseases that derive from a few genetically altered cells, such as cancer. I show how to predict the expected amount of mosaicism and the variation in mosaicism between individuals. I then calculate the predicted risk of cancer derived from developmental mutations. The calculations show that a significant fraction of cancer in later life likely arises from developmental mutations in early life. In addition, much of the variation in the risk of cancer between individuals may arise from variation in the degree of genetic mosaicism set in early life. I also suggest that certain types of neurodegeneration, such as amyotrophic lateral sclerosis (ALS), may derive from a small focus of genetically altered cells. If so, then the risk of ALS would be influenced by developmental mutations and the consequent variation in genetic mosaicism. New technologies promise the ability to measure genetic mosaicism by sampling a large number of cellular genomes within an individual. The sampling of many genomes within an individual will eventually allow one to reconstruct the cell lineage history of genetic change in a single body. Somatic evolutionary genomics will follow from this technology, providing new insight into the origin and progression of disease with increasing age.

amyotrophic lateral sclerosis | Luria–Delbruck fluctuation analysis | somatic phylogenetics

A human develops from a single cell. From that single cell, an individual grows to 10^{13} to 10^{14} cells. That growth requires many cell divisions and, consequently, much somatic mutation must occur during development. Those developmental mutations likely have significant consequences for genetic mosaicism in the body and for the risk of cancer that arises from those mutations (1, 2).

In this article, I will show how to predict the amount of somatic mutation and the amount of genetic variability in the body. I will connect those calculations to the risk of cancer. I also propose that certain neurodegenerative diseases that occur later in life may often derive from early-life somatic mutations that occur during development.

Before turning to the details, let us consider in a general way the magnitude of somatic mutation during development in relation to the number of cells in the body. During development the single-cell zygote expands to $N = 10^{13}$ to 10^{14} cells. How many cell divisions occur during that expansion? Each time a cell divides, the number of cells in the body increases by one, assuming no cell death. So, to start with one cell and expand to N cells requires at least $N - 1$ cell divisions.

How much somatic mutation occurs during development? We do not have good measurements, but we can make some rough calculations. The minimum number of cells divisions is $N - 1 \approx N = 10^{13}$ to 10^{14}. Define the mutation rate per gene per cell division as u. No truly reliable estimates of somatic mutation rates exist, but typically assumed values are of the order $u = 10^{-7}$ to 10^{-6} (3). The total number of mutational events per gene during development is the mutation rate per cell division multiplied by the number of cell divisions, $uN = 10^6$ to 10^8. Thus, every gene in the genome mutates many times.

The value of uN measures the number of mutational events that occur in each gene. But most often, we will be interested in the number of cells that carry a mutation in a particular gene. For example, if a mutation occurs early in development, then that single mutation will carry forward to many descendant cells. By contrast, relatively few cells will carry a mutation that happens late in development. To understand the relation between the number of mutations that occur and the number of cells that carry a mutation, we must place somatic mutations in the context of cell lineage history. In other words, we must think of the body in relation to the lineage history descending from the single ancestral zygote and how mutations accumulate in that lineage history.

The accumulation of change within the lineage history of the body is somatic evolutionary genomics. With $\approx 10^{13}$ to 10^{14} cells in a body, and probably $>10^{16}$ cells produced over a lifetime, the lineage history within a single individual is much greater than for all of the hominids that have ever lived, perhaps as great as for all of the primates that have ever lived.

The tremendous evolutionary history within each human body has, until recently, been hidden by the difficulty of measuring genetic changes in cells. New high-throughput genomic technologies are just opening up the possibility of directly measuring somatic variability and evolution (4). To understand the evolutionary history of the individual and the consequences for disease, we must place somatic genomics within the context of the rate and pattern of evolutionary change in cellular lineages.

Genetic Mosaicism

In this section, I explain in more detail how mutations accumulate in cell lineages. I emphasize that the shape of lineage history differs at different times of life and in different tissues, affecting the patterns of somatic evolution. I also show how to predict the amount of genetic mosaicism in an individual and the variation in mosaicism between individuals. The following sections connect the amount and variation in genetic mosaicism to the risk of diseases such as cancer and neurodegeneration.

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Evolution in Health and Medicine" held April 2–3, 2009, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS web site at www.nasonline.org/Sackler_Evolution_Health_Medicine.

Parts of this article were modified from earlier publications (1, 2).

Author contributions: S.A.F. performed research and wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission. S.C.S. is a guest editor invited by the Editorial Board.

[1]E-mail: safrank@uci.edu

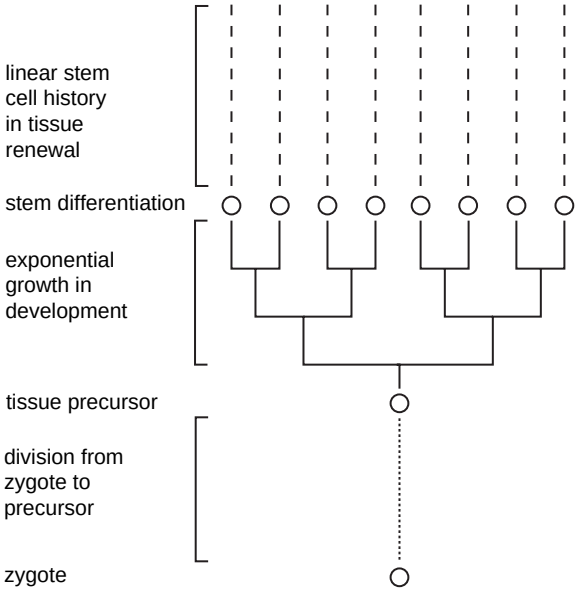


Fig. 1. Lineage history of cells in renewing tissues. All cells trace their ancestry back to the zygote. Each tissue, or subset of tissue, derives from a precursor cell; n_p rounds of cell division separate the precursor cell from the zygote. From a precursor cell, n_e rounds of cell division lead to exponential clonal expansion until the descendants differentiate into the tissue-specific stem cells that seed the developing tissue. In a compartmental tissue, such as the intestine, lineage history of the renewing tissue follows an essentially linear path, in which each cellular history traces back through the same sequence of stem cell divisions (2, 21). At any point in time, a cell traces its history back through n_s stem cell divisions to the ancestral stem cell in the tissue, and $n = n_p + n_e + n_s$ divisions back to the zygote. (Modified from figure 13.1 in ref. 2, based on the original in ref. 1.)

Shape of Cell Lineages. Fig. 1 shows that renewing tissues typically have two distinct phases in the history of their cellular lineages. Early in life, cellular lineages expand exponentially to form the tissue. For the remainder of life, stem cells renew the tissue by dividing to form a nearly linear cellular history

Mutations accumulate differently in the exponential and linear phases of cellular division (1). During the exponential phase of development, a mutation carries forward to many descendant cells. During the renewal phase, a mutation transmits only to the localized line of descent in that tissue compartment: one mutational event has limited consequences (Fig. 2).

Amount and Variation of Genetic Mosaicism: Theory. Consider a renewable tissue, such as the colon epithelium or the hematopoietic system. Those tissues renew throughout life from a set of stem cells. A human colon has approximately $N = 10^8$ stem cells, with probably at least that many stem cells in the hematopoietic system. At the end of development, what fraction of those stem cells carries a somatically derived mutation? To answer this question, we must analyze how mutations during development translate into the number of initially mutated stem cells at the end of development.

Mutations occur stochastically in the small number of cells present early in development. The number of mutant stem cells and the degree of genetic mosaicism will therefore vary greatly between individuals according to a probability distribution called the Luria–Delbruck distribution (5). That distribution describes the number of mutant cells, M, in a population that grows exponentially from one cell to N cells (6, 7).

Suppose, for example, that $N = 10^8$ stem cells must be produced during development to seed a tissue. Exponential growth of one cell into N cells requires, in the absence of cell death, a total of $N - 1$ cellular divisions arranged into approximately $\ln(N)$ cellular generations. In this case, $\ln(10^8) \approx 18$. If the mutation rate per locus per cell division during exponential growth is u_e, then the probability, x, that any final stem cell carries a mutation at a particular locus is approximately the mutation rate per cell division multiplied by the number of cell divisions from that particular cell back to the ancestral progenitor cell. In this case, $x = u_e \ln(N)$. This probability is usually small: for example, if $u_e = 10^{-6}$, then x is of the order of 10^{-5}.

The frequency of initially mutated stem cells may be small, but the number may be significant. The average number of mutated cells at a particular locus is the number of cells, N, multiplied by the probability of mutation per cell, x. In this example, $Nx \approx 10^3$, or ≈1,000.

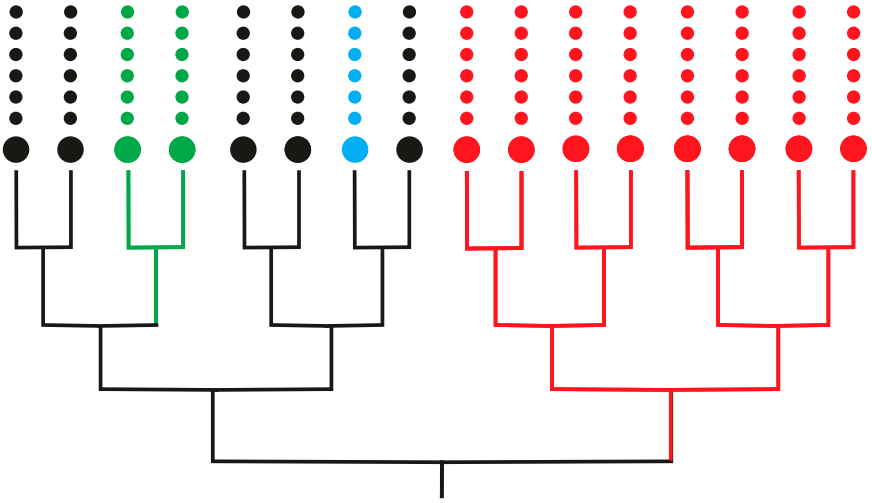


Fig. 2. Mutational events in development occur at different stages in the exponential, branching phase of cellular expansion. In this example, the red mutation happens early, causing a significant fraction of somatic cells to carry the same mutation by descent. By contrast, the green mutation happens late in development, causing only a small fraction of somatic cells to carry the same mutation by descent. Many mutations will arise within the stem cells, each stem cell renewing only a very small fraction of all somatic cells. For example, the blue mutation is private to a single stem cell and will be confined to the small subset of somatic cells derived from that stem cell. Branching lines represent the developmental phase of cellular expansion, the large cells are stem cells, and the small cells in each line form a clone derived from their stem cell ancestor.

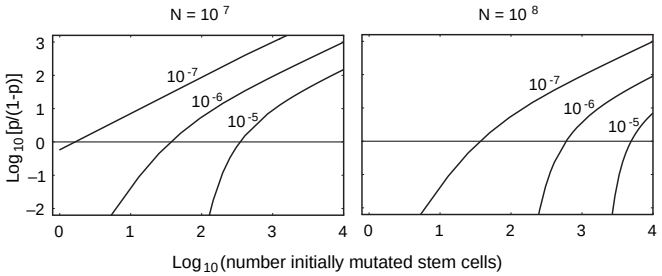


Fig. 3. Number of initially mutated stem cells at the end of development. The N initial stem cells derive by exponential growth from a single precursor cell. Each plot shows the cumulative probability, p, for the number of mutated initial stem cells. By plotting $\log_{10}(p/(1-p))$, the zero line gives the median of the distribution. The number above each line is u_e, the mutation probability per cell added to the population during exponential growth. (I used an actual value of $10^{-5.2}$ rather than 10^{-5} because of computational limitations.) For a single gene, the mutation probability per gene per cell division, u_g, is probably $>10^{-7}$. If there are at least $L = 100$ genes for which initial mutations can influence the progression to cancer, then $u_e = Lu_g \geq 10^{-5}$. Initial mutations may, for example, occur in DNA repair genes, causing an elevated rate of mutation at other loci. Calculations were made with algorithms in Zheng (38). (Modified from figure 13.3 in ref. 2, based on the original in ref. 1.)

I have focused on mutations at a single locus. Mutations at many different loci may predispose to genetically influenced diseases such as cancer. Suppose mutations at L different loci can contribute to predisposition. We can get a rough idea of how multiple loci affect the process by simply adjusting the mutation rate per cell division to be a genome-wide rate of predisposing mutations, equal to $u_e L$. The number of loci that may affect predisposition may reasonably be around $L \approx 10^2$ and perhaps higher. Following the calculation in the previous paragraph, with $L \geq 10^2$, the number of initial stem cells carrying a predisposing mutation would on average be at least 10^5. Some individuals might have two predisposing mutations in a single initial stem cell.

The average number of initially mutated cells tells only part of the story, because the distribution for the number of mutants is highly skewed. A few rare individuals have a great excess; in those individuals, a mutation arises early in development, and most of the stem cells would carry the mutation. Those individuals would have nearly the same risk as one who inherited the mutation.

Fig. 3 shows the distribution for the number of initially mutated stem cells at the end of development. For example, in Fig. 3 *Right*, with a mutation probability per cell division of 10^{-6}, a y-axis value of 2 means that $\approx 10^{-2}$, or 1%, of the population has $>10^4$ initially mutated stem cells at a particular locus ($L = 1$). Similarly, with a mutation probability per cell division of 10^{-7}, a y-axis value of 3 means that $\approx 10^{-3}$, or 0.1%, of the population has $>10^4$ initially mutated stem cells at a particular locus.

Amount and Variation of Genetic Mosaicism: Observations. My simple calculations show that, in a typical individual, every gene mutates somatically many times. Similarly, most cells in the body probably carry at least one somatic mutation. When we place this widespread somatic mosaicism into the context of cell lineage history, as in Fig. 2, we see that the body likely comprises variable-sized patches of somatic mosaicism throughout the genome.

Simple theory tells us that this widespread mosaicism must be present. But few measurements of mosaicism have been accomplished. The lack of measurement occurs because it is not easy to analyze genetic variation between individual cells in the huge population of cells that comprise a single body. I mentioned earlier that the population of cells in a single body greatly exceeds that of all hominid individuals that have ever existed. A comprehensive study of the somatic evolutionary genomics of a single body would be as challenging as an evolutionary analysis of genetic variability for all humans and their hominid ancestors throughout hominid history.

A few analyses of mosaicism do exist (8). Mutation of a skin pigmentation gene in development causes skin cells to be marked by the mutation. The marked cells trace the tips of the somatic evolutionary lineage tree on the body surface. Interestingly, the patterned skin markings vary considerably. For example, several visible skin diseases follow the lines of Blaschko, which trace out what seem like contour lines or whorls over the skin surface (9–12). Other distinct patterns also occur in skin diseases (12). Speckled lentiginous naevus and Becker's naevus follow a mosaic checkerboard pattern; mosaic trisomy of chromosome 13 causes scattered leaf-like shapes of hypopigmentation.

Although somatic evolutionary genomics is difficult at present, genomic technology is advancing rapidly. The recent cancer genome project shows the potential for screening genetic changes in the somatic cells of an individual (4). Wallace's (13) work and future vision for somatic mitochondrial genetics emphasize the potential for analyzing the diseases of increasing age in the context of accumulating somatic mutations in cellular lineages.

The hematopoietic system provides a particularly promising somatic component for future study. Blood cells derived from diverse stem cell populations can easily be sampled and followed over time. Greaves (14) reviewed several lines of evidence demonstrating that developmental mutations in the hematopoietic system cause widespread somatic mosaicism. In many cases, those developmental mutations appear to be the primary cause of childhood leukemia (15–17).

Are the rare childhood forms of leukemia isolated examples, or is somatic mosaicism derived from developmental mutation a hidden risk factor in many cancers? At present, no direct evidence answers this question. The following section considers some theory by which we can predict the risk from developmental mutation. The theory also provides a framework for analyzing the data on somatic genomics that will become available in the future.

Risk of Cancer

Developmental mutations inevitably cause genetic mosaicism. Those cells carrying somatic mutations from development may be predisposed to cancer. In this section, I analyze the increased risk of cancer attributable to somatic mutations in early development.

Frank and Nowak's Model. Fig. 3 shows the probability distribution for the number of stem cells that have mutations. The number of mutated stem cells is $M = Nz$, where N is the total number of stem cells, and z is the frequency of those stem cells that carry a mutation that predisposes to disease. In this section, I focus on the average frequency of mutated stem cells, where x is the average of z, which leads to the average number of mutated stem cells, Nx. In the following section, I discuss the wide variation between individuals in the frequency of mutated stem cells. That variation in mutated stem cells may explain a significant fraction of the variation in cancer risk between individuals.

Mutations during the exponential phase of cellular growth in development cause the average frequency of stem cells with mutations to be $x \approx u_e \ln(N)$ (6), where u_e is the mutation rate during exponential cellular growth in development. Although the frequency of stem cells that start with a mutation may be small, those mutations can contribute substantially to the total risk of cancer. Suppose, by the multistage model of cancer progression, that k rate-limiting mutations are needed to cause

cancer (18–20). To obtain a rough estimate of the total risk, let $R_T = N(1-x)R_k + NxR_{k-1}$, where x is the frequency of stem cells that start with one mutation, and R_k is the risk that a particular stem cell lineage acquires k mutations during the phase of linear division and tissue renewal. The risk, R_k, can be approximated by the γ distribution, which gives the probability for the occurrence of the kth event over a particular time interval. From the γ distribution, $R_k \approx (u_s\tau)^k/k!$, where u_s is the mutation rate per stem cell division, and τ is the total number of stem cell divisions.

The relative odds that cancer derives from stem cells that begin with one mutation acquired in development compared with those stems cells that lack a developmental mutation can be calculated as $F = xR_{k-1}/(1-x)R_k \approx u_e \ln(N)k/u_s\tau$. The next step is to fill in approximate magnitudes for these quantities. We can take $k \approx 5$ for the number of rate-limiting mutations to cause epithelial cancer in humans (18–20), $\ln(N) \approx 20$, and τ from the range 100 to 1,000. This gives the relative odds of cancer deriving from a developmental mutation, F, ranging from $\approx u_e/10\ u_s$ to u_e/u_s.

If mutations accumulate with the same probability per cell division during exponential growth and linear stem cell division, $u_e = u_s$, then mutations arising in development increase risk by 10–100%. If, as Cairns (21) has argued, stem cell mutation rates are much less than mutation rates during exponential growth, $u_s \ll u_e$, then almost all cancer arises from predisposed stem cell lineages that were mutated during development.

Meza et al.'s Model. Meza et al. (22) extended Frank and Nowak's (1) model in three ways. First, they used an explicit quantitative model of colorectal cancer progression to apply direct parameter estimates from data. Second, they analyzed the relative proportion of cancer derived from developmental mutations at each age; they showed that earlier onset of cancer more often derives from developmental mutations than later onset of cancer. Third, they calculated the amount of variation in cancer risk between individuals caused by the stochastic nature of somatic mutations in early development.

Meza et al. (22) began with a model of colorectal cancer progression and incidence that they had previously studied (23). In that model, carcinogenesis progresses through four stages: two initial transitions, followed by a third transition that triggers clonal expansion, and then a final transition to the malignant stage.

In their study, Meza et al. (22) began with the same four-stage model. They then added a Luria–Delbruck process to obtain the probability distribution for the number of stem cells mutated at the end of development. The stochasticity in the Luria–Delbruck process causes wide variation between individuals in the number of mutated stem cells. Meza et al. first calculated the probability that an individual carries Nz initially mutated stem cells at the end of development. To obtain overall population incidence, they summed the probability for each Nz multiplied by the incidence for individuals with Nz mutations.

Meza et al. (22) summed incidence in their four-stage model over the number of initially mutated stem cells to fit the model's predicted incidence curve to the observed incidence of colorectal cancer in the United States. From their fitted model, they then estimated the proportion of cancers attributable to mutations that arise during development. Fig. 4 shows that a high proportion of cancers may arise from mutations during the earliest stage of life.

Cancers at unusually young ages are often attributed to inheritance. However, Fig. 4 suggests that early-onset cancers may often arise from developmental mutations. Developmental mutations act similarly to inherited mutations: if the developmental mutation happens in one of the first rounds of postzygotic cell division, then many stem cells start life with the mutation.

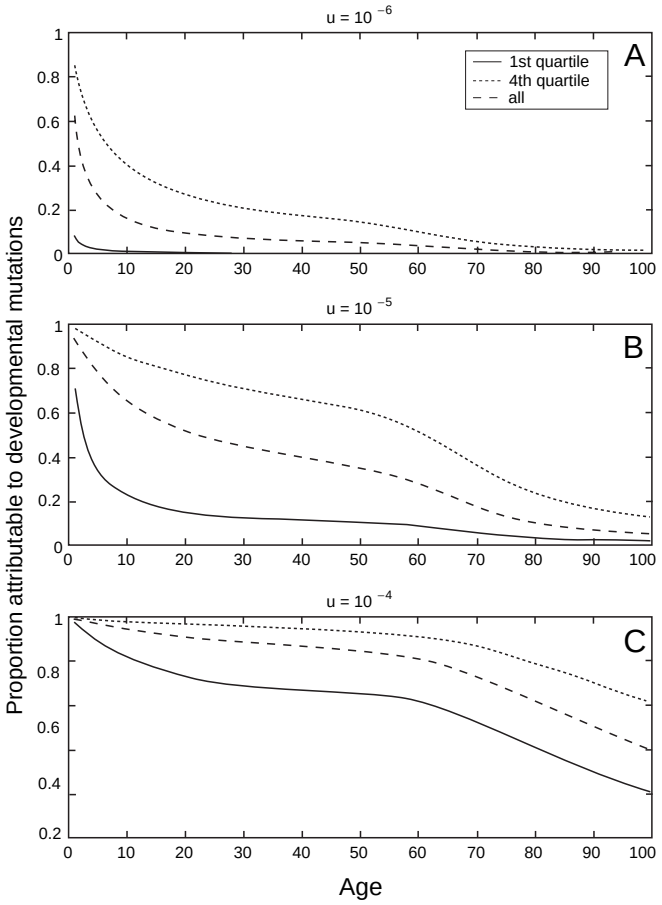


Fig. 4. The proportion of cancers that arise from cells mutated during development. These plots show calculations based on a specific four-stage model of colorectal cancer progression (22). The parameters of the progression model were estimated from incidence data. The values of u above each plot show the mutation rate per year in stem cells. Stem cells likely divide between 10 and 100 times per year, thus a mutation rate per year of at least 10^{-5} per locus seems reasonable. In each plot, the three curves sketch the heterogeneity between individuals in risk attributable to developmental mutations. The first quartile shows the proportion of cancers at each age for those individuals whose risk is in the lowest 25% of the population, in particular, those individuals who by chance have the fewest stem cells mutated during development. Similarly, the fourth quartile shows the risk for the highest 25% of the population with regard to developmental mutations. [Reproduced with permission from ref. 22 (Copyright 2005, Elsevier).]

Inheritance is, in effect, a mutation that happened before the first zygotic division.

Risk of Neurodegenerative Disease

Developmental mutations alter a fraction of cells, leading to genetic mosaicism. Any disease that derives from a small number of mutated cells may be influenced by developmental mutations. For example, cancer progression ultimately encompasses many cells of a tissue, but the initial change in one or a few cells starts the process. Thus, developmental mutations, by seeding the process in many stem cells, enhance the risk that one of those predisposed cells will lead to cancer.

Neurodegenerative diseases such as amyotrophic lateral sclerosis (ALS) appear, at first glance, very different from cancer. Individuals typically seem to be without any symptoms through middle or late life. Then, within a few years, they progress from initial motor problems to widespread motor neuron degeneration and death (24). The widely distributed motor neuron degeneration seems to argue against the disease initiating in a

one or a few cells. However, Armon (25) suggested that the disease does start in a focal area and then spreads to neighboring motor neurons.

In ALS, a particular limb may show the first signs of motor neuron deterioration. The next symptoms frequently occur in the contralateral limb at the same spinal level and at contiguous spinal levels (26). Alternatively, spread may occur up and down the motor neuron system through the primary motor neurons directly connected to muscles and those neurons that feed into the primary motor system (27). Focal initiation followed by spread means that the disease may be initiated in one or a few cells, in the same way as cancer (25).

Developmental mutations and somatic mosaicism would play a role in ALS risk only if the initiating events in the focal cells derived from genetic (or epigenetic) changes in those cells. No direct evidence exists for a role of somatic mutation in initiating ALS. However, inherited germ-line mutations do strongly predispose to ALS, so it is possible that genetics plays a key role in disease (27–32). But the direct role of somatic mutation, and of what follows, remains a speculative hypothesis at present.

The most common genetic mutations associated with ALS lead to misfolded proteins (33–35). Individuals who inherit a predisposing mutation typically appear normal until disease onset usually in the age range of 25–65 years (30). As in inherited predisposition to cancer, other factors must be involved in transforming predisposed cells into the initiating focus of a disease that subsequently spreads.

Most cases seem to arise sporadically, without any evidence of inherited predisposition. Sporadic cases typically have a later age of onset in the range 40–80 years (27). The sporadic disease is mostly similar to the familial form, although some variations in the inherited form have been noted (27). This similarity between sporadic and inherited cases suggests that the sporadic cases may possibly begin from an initiating focus of cells with genetic mutations, but, again, there is no direct evidence for this.

If both inherited and sporadic cases do initiate disease from a small focus of cells that carry a genetic mutation, then, as in cancer, we can predict a continuum of risk. Those who inherit a predisposing mutation have the highest risk and earliest age of onset, with all cells carrying the mutation. Those who suffer a somatic mutation early in development have only slightly lower risk and later age of onset relative to inherited cases, with many cells carrying the mutation. Those who suffer a developmental mutation at some intermediate time in neuronal development have less risk and higher age of onset, with fewer cells, but still a significant number, carrying the predisposing mutation. Those who suffer a late developmental mutation have relatively low risk and late age of onset, with relatively few cells carrying the mutation. Finally, those with very few developmentally mutated cells may form the largest class with lowest risk and latest onset; those with very few developmental mutations may in fact have such low risk that nearly all cases can be ascribed to those who carry a significant number of developmental mutations.

If most cases derive from either inherited or developmental mutations, then the risk of ALS and perhaps other neurodegenerative diseases may be set very early in life. If sporadic cases often derive from developmental mutations, then somatic genomics will reveal an association between age of onset and the fraction of cells that carry a predisposing somatic mutation. The inherited cases would simply be the extreme of the risk continuum, in which all cells carry the predisposing mutation.

What sort of evidence would weigh in favor or against this speculative hypothesis? On the positive side: new somatic genomic measures that show a correlation between early-onset ALS and a somatic mutation widely distributed in neural tissue by descent from a single somatic mutation in a common ancestral cell; a mechanistic pathway that links the somatic mutation to the onset and progression of ALS; and further evidence that disease spreads from a small focus of cells that have been transformed. On the negative side: no correlation between early-onset ALS and widely distributed somatic mutations known to predispose to ALS when those mutations arise in the germ line; and evidence that disease arises over a short period in many widely distributed locations rather than spreading from a small number of foci.

Potential Difficulties

The risks conferred by abundant low-penetrance genes affecting carcinogenesis may be of the same order as the risk from developmental mutations. How can we pick out the proportion of cancers that arise from developmental mutations if they are "drowned out" by population heterogeneity because of a number of weakly penetrant germ-line susceptibility genes?

We can distinguish between the risk caused by developmental mutations and weakly penetrant germ-line mutations by direct evidence from somatic genomics. Evidence favors developmental mutations when: one directly observes that somatic mutations dispersed widely in a tissue arose by normal cell lineage expansion in the development of that tissue; the dispersed somatic mutation derived from a single mutational event in an ancestral cell; and there is an association between those somatic mutations from normal developmental processes and the probability and age of cancer onset. These sorts of measurements are not easy at present, but the technology is developing very rapidly.

Another potential difficulty concerns the simple multistage model of carcinogenesis that I used in making calculations about the consequences of developmental mutations. I used the classical and most commonly cited form of the multistage model presented by Armitage and Doll (18) in 1954. As early as 1957, Armitage and Doll (36) pointed out that the age-specific incidence of certain cancers did not fit the simplest assumptions of their original multistage theory. There has followed a long history of model variants that fit particular assumptions to the age-specific incidence of particular cancers (2).

Recently, Meza et al. (37) made an important contribution to this long history by fitting explicit models of carcinogenesis to the age-specific incidence of colorectal and pancreatic cancers. They show that the classical model of Armitage and Doll (18) is not sufficient, and they develop a specific alternative that accounts for certain aspects of the biology and provides a much better fit to the data.

Given the limitations of the classical Armitage and Doll model that I used in the early sections of this article, how might my conclusions about developmental mutations be affected by alternative and more realistic models of carcinogenesis? The particular quantitative values derived from the model and shown in the figures would be changed. However, the model and the figures were meant only to illustrate the basic qualitative idea that developmental mutation potentially contributes to the risk of cancer and potentially alters the age of onset in those who suffer most strongly from widely dispersed developmental mutations. Those qualitative conclusions hold under any realistic model of carcinogenesis in which somatic mutations (or any heritable changes to cells) play a role.

My recent book (2) went into great detail about the various quantitative models of carcinogenesis in relation to the available data on age-specific incidence. Much can be learned by close study of those issues, and Meza et al. (37) have made a very significant contribution to that topic. But with regard to developmental mutations and somatic genomics, the real problems turn on the coming technological advances that will allow direct observation of how somatic mutations disperse through tissues in relation to cell lineage history and how those somatic mutations influence progression to cancer. Confidence in particular mathematical models must await those advances.

Conclusions

Skin diseases and childhood leukemia show that somatic mosaicism does occur. I have argued that those examples are just hints of the hidden and widespread mosaicism that arises from developmental mutations. Commonly accepted assumptions about cell division, cell lineage history, and somatic mutation lead inevitably to the prediction that mosaicism is common in all genes throughout the genome.

Importantly, the stochastic nature of mutation and the small number of cells in early development predict that the degree of mosaicism varies greatly between individuals. For any gene, a small fraction of individuals in the population will carry a somatic mutation in many cells, the mutation having occurred in early development. Those individuals will be at risk for diseases such as cancer and neurodegeneration that, later in life, can spread from a small focus of genetically predisposed cells.

Other individuals will carry relatively few predisposing somatic mutations derived from development. Those individuals have less risk later in life for diseases such as cancer and neurodegeneration. If this argument is correct, then a significant fraction of risk for diseases later in life may derive from mutational events before birth.

Until recently, genetic technology has not allowed widespread sampling of somatic genomes within an individual. Thus, there is currently no direct evidence for or against this theory of widespread mosaicism and its association with the risk of disease. New technologies promise the ability to sample large numbers of genomes. The sampling of many genomes within an individual will eventually allow one to reconstruct the cell lineage history of genetic change in a single body. Somatic evolutionary genomics will follow from this technology, providing new insight into the origin and progression of disease with increasing age.

ACKNOWLEDGMENTS. I thank Susan Fitzpatrick for introducing me to the genetic basis of neurodegenerative disease and the potential analogy with cancer and Sharon Murphy for providing helpful comments and discussion, particularly with regard to childhood leukemia. My research is supported by National Science Foundation Grant EF-0822399, National Institute of General Medical Sciences MIDAS Program Grant U01-GM-76499, and a grant from the James S. McDonnell Foundation.

1. Frank SA, Nowak MA (2003) Developmental predisposition to cancer. *Nature* 422:494.
2. Frank SA (2007) *Dynamics of Cancer: Incidence, Inheritance, and Evolution* (Princeton Univ Press, Princeton).
3. Araten DJ, et al. (2005) A quantitative measurement of the human somatic mutation rate. *Cancer Res* 65:8111–8117.
4. Stratton MR, Campbell PJ, Futreal PA (2009) The cancer genome. *Nature* 458:719–724.
5. Luria SE, Delbruck M (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 28:491–511.
6. Zheng Q (1999) Progress of a half-century in the study of the Luria–Delbruck distribution. *Math Biosci* 162:1–32.
7. Frank SA (2003) Somatic mosaicism and cancer: Inference based on a conditional Luria–Delbruck distribution. *J Theor Biol* 223:405–412.
8. Gottlieb B, Beitel LK, Trifiro MA (2001) Somatic mosaicism and variable expressivity. *Trends Genet* 17:79–82.
9. Happle R (1993) Mosaicism in human skin. Understanding the patterns and mechanisms. *Arch Dermatol* 129:1460–1470.
10. Siegel DH, Sybert VP (2006) Mosaicism in genetic skin disorders. *Pediatr Dermatol* 23:87–92.
11. Taibjee SM, Bennett DC, Moss C (2004) Abnormal pigmentation in hypomelanosis of Ito and pigmentary mosaicism: The role of pigmentary genes. *Br J Dermatol* 151:269–282.
12. Chuong CM, et al. (2006) What is the biological basis of pattern formation of skin lesions? *Exp Dermatol* 15:547–549.
13. Wallace DC (2005) A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: A dawn for evolutionary medicine. *Annu Rev Genet* 39:359–407.
14. Greaves M (2005) In utero origins of childhood leukemia. *Early Hum Dev* 81:123–129.
15. Greaves MF, Maia AT, Wiemels JL, Ford AM (2003) Leukemia in twins: Lessons in natural history. *Blood* 102:2321–2333.
16. Greaves MF, Wiemels J (2003) Origins of chromosome translocations in childhood leukemia. *Nat Rev Cancer* 3:639–649.
17. Mori H, et al. (2002) Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proc Natl Acad Sci USA* 99:8242–8247.
18. Armitage P, Doll R (1954) The age distribution of cancer and a multistage theory of carcinogenesis. *Br J Cancer* 8:1–12.
19. Hanahan D, Weinberg RA (2000) The hallmarks of cancer. *Cell* 100:57–70.
20. Fearon ER, Vogelstein B (1990) A genetic model for colorectal tumorigenesis. *Cell* 61:759–767.
21. Cairns J (1975) Mutation selection and the natural history of cancer. *Nature* 255:197–200.
22. Meza R, Luebeck EG, Moolgavkar SH (2005) Gestational mutations and carcinogenesis. *Math Biosci* 197:188–210.
23. Luebeck EG, Moolgavkar SH (2002) Multistage carcinogenesis and the incidence of colorectal cancer. *Proc Natl Acad Sci USA* 99:15095–15100.
24. Mitchell JD, Borasio GD (2007) Amyotrophic lateral sclerosis. *Lancet* 369:2031–2041.
25. Armon C (2005) Acquired nucleic acid changes may trigger sporadic amyotrophic lateral sclerosis. *Muscle Nerve* 32:373–377.
26. Caroscio JT, Mulvihill MN, Sterling R, Abrams B (1987) Amyotrophic lateral sclerosis. Its natural history. *Neurol Clin* 5:1–8.
27. Armon C (2003) Epidemiology of ALS/MND. *Motor Neuron Disorders*, eds Shaw P, Strong M (Elsevier, Amsterdam), pp 167–206.
28. Kabashi E, et al. (2008) TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat Genet* 40:572–574.
29. Sreedharan J, et al. (2008) TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science* 319:1668–1672.
30. Strong MJ, Hudson AJ, Alvord WG (1991) Familial amyotrophic lateral sclerosis, 1850–1989: A statistical analysis of the world literature. *Can J Neurol Sci* 18:45–58.
31. Valdmanis PN, Rouleau GA (2008) Genetics of familial amyotrophic lateral sclerosis. *Neurology* 70:144–152.
32. Van Deerlin VM, et al. (2008) TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: A genetic and histopathological analysis. *Lancet Neurol* 7:409–416.
33. Kabashi E, Durham HD (2006) Failure of protein quality control in amyotrophic lateral sclerosis. *Biochim Biophys Acta* 1762:1038–1050.
34. Meiering EM (2008) The threat of instability: Neurodegeneration predicted by protein destabilization and aggregation propensity. *PLoS Biol* 6:e193.
35. Wang Q, Johnson JL, Agar NY, Agar JN (2008) Protein aggregation and protein instability govern familial amyotrophic lateral sclerosis patient survival. *PLoS Biol* 6:e170.
36. Armitage P, Doll R (1957) A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. *Br J Cancer* 11:161–169.
37. Meza R, Jeon J, Moolgavkar SH, Luebeck EG (2008) Age-specific incidence of cancer: Phases, transitions, and biological implications. *Proc Natl Acad Sci USA* 105:16284–16289.
38. Zheng Q (2005) New algorithms for Luria–Delbruck fluctuation analysis. *Math Biosci* 196:198–214.

Transfers and transitions: Parent–offspring conflict, genomic imprinting, and the evolution of human life history

David Haig[1]

Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138

Edited by Stephen Curtis Stearns, Yale University, New Haven, CT, and accepted by the Editorial Board June 11, 2009 (received for review April 16, 2009)

Human offspring are weaned earlier than the offspring of other great apes but take longer to reach nutritional independence. An analysis of human disorders of imprinted genes suggests genes of paternal origin, expressed in infants, have been selected to favor more intense suckling than genes of maternal origin. The same analysis suggests that genes of maternal origin may favor slower childhood growth but earlier sexual maturation. These observations are consistent with a hypothesis in which slow maturation was an adaptation of offspring that reduced maternal fitness, whereas early weaning was an adaptation of mothers that reduced the fitness of individual offspring.

Beckwith–Wiedemann syndrome | genomic imprinting | Prader–Willi syndrome | weaning

Ethnographic data suggest our ancestors consumed more food than they gathered until early adulthood and gathered more food than they consumed thereafter (1). Thus, hominin life history involved a transfer of resources from older producers to younger consumers. Lee modeled the consequences of these transfers for the evolution of age-specific mortality (2, 3). He found that transfers from older to younger individuals mitigate the force of selection against early deaths, because the death of a dependent youngster frees food for other group members, but intensify selection against late-life mortality, because the death of a productive elder reduces food for survivors.

Lee's model assumed consumers and producers were genetically identical, except for new mutations (2). It was as if older producers could provision their younger selves. In sexual life cycles, however, resources are transferred between individuals who may share some, but not all, of their genes (4). If multiple donors transfer resources to multiple recipients, then each donor favors the distribution of resources that maximizes her inclusive fitness, but each recipient favors the distribution that maximizes his inclusive fitness. Individual consumers are predicted to take a larger share of production (if given the opportunity) than the quantity favored by donors. The paradigm of such conflict is the allocation of maternal investment among offspring. If a mother distributes resources in a manner that maximizes her fitness, then each offspring will favor a reallocation from sibs to itself (5). Genes that are expressed differently when inherited via ova than via sperm are predicted to mediate kin conflicts (6, 7). Therefore, the phenotypic effects of such imprinted genes will provide important clues about how transfers among kin have shaped human life history.

Modeling Transfers Among Kin

Patterns of resource transfers within groups and gene transfers between groups are variable among modern human populations and were presumably variable among ancestral populations. Simplification of this complexity is necessary to gain theoretical insight into how resource and gene transfers interact. I will consider a simple gene-transfer model in which females move to new groups when they switch from being net consumers to net producers whereas males remain in their natal group, and an equally simple resource-transfer model in which females put food into household pots, from which their own offspring feed, but men put food into a communal pot, from which all offspring feed. These models will be combined with a model of childhood consumption that identifies conflicts between genes of maternal and paternal origin. These verbal models are deliberately abstract because it is my belief that progress in understanding the action of natural selection in complex human social groups will be advanced by first understanding how selection acts in simpler systems.

Recent human populations exhibit a flexibility of social structures that is not captured in these models: males move to live with their wives' families; both sexes have multiple sexual partners; relationships dissolve; groups split or fuse; related individuals emigrate together; and resources, and genes, are exchanged between groups. Moreover, genes repeatedly leave and reenter local groups as a consequence of regular intermarriage between clans, creating larger regional groups bound together by interlocking kinship, and patterns of relatedness within groups are influenced by demographic stochasticity. Thus, my models assume patterns of gene and resource transfers that are undoubtedly wrong in details, and perhaps wrong in fundamentals. Nevertheless, I hope these models will identify key issues, both empirical and theoretical, that must be addressed by more-realistic analyses, and will provide a baseline against which the effects of departures from my idealized assumptions can be assessed.

One assumption of the models is particularly contentious. I assume mobile females and sedentary males. Patrilocality has been claimed to be the predominant mode of social organization among recent hunter-gatherers (8) but this claim has been strongly disputed (9, 10). My models are not intended to resolve this debate, but rather to explore the theoretical consequences of female-biased dispersal, in part, because I considered male-biased dispersal in ref. 11. Ethnographic data show that both forms of dispersal occur among recent humans and that females often maintain social ties with their natal group after dispersal. I believe that the relative mobility of the sexes in the past is a key unresolved issue for understanding the evolution of human life history.

A challenge for future models will be incorporating variability among social groups. Natural selection deals with variability in 2 contrasting ways that necessitate different modeling approaches. The first is to average across circumstances: strategies evolve that are adaptive on average even though they are maladaptive in some situations. The second is to evolve contingent strategies that enable individuals to respond in different ways to different circumstances. The first process produces a

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Evolution in Health and Medicine" held April 2–3, 2009, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS web site at www.nasonline.org/Sackler_Evolution_Health_Medicine.

Author contributions: D.H. wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission. S.C.S. is a guest editor invited by the Editorial Board.

[1]E-mail: dhaig@oeb.harvard.edu.

Table 1. Human imprinted disorders discussed in text

Human syndrome	Some (epi)genetic causes	Ref(s).
Beckwith–Wiedemann syndrome (BWS)	Excess expression of paternally-expressed *IGF2* Inactivation of maternally-expressed *CDKN1C*	23,57
Silver–Russell syndrome (SRS)	Maternal methylation pattern of paternal 11p15.5 Maternal uniparental disomy 7	23,58
Prader–Willi syndrome (PWS)	Deletion of paternal 15q11–13 Maternal uniparental disomy 15	35,59
Temple syndrome (TS)	Deletion of paternal 14q32 Maternal uniparental disomy 14	49

general-purpose phenotype and the second a repertoire of context-specific adaptations.

Gene-Transfer Model. A young woman enters a group where she lacks relatives and accumulates kin by reproduction. At first her kin are restricted to her own offspring but later include grandoffspring, as her sons mature and reproduce with younger unrelated women. By contrast, a young man remains within his natal group and obtains one or more wives from neighboring groups. The dependent kin of his group consist of his own offspring, his younger sibs, and a mixture of patrilineal nieces, nephews, and cousins. As he ages, the composition of reproductive males in his group gradually shifts from uncles and brothers to sons and nephews. At the same time, the population of dependent young shifts to grandchildren and progressively more distant patrilineal kin.

In this scenario, the productive older individuals of a group consist of unrelated females and related males. Thus, the group's genetic cohesion is maintained via patrilineal descent, including descent from shared paternal grandmothers, while the genetic fissures separate matrilines. Genetic cohesiveness declines as new females enter the group and establish new matrilines. Eventually, the group may become unstable and split, possibly into separate matrilines of older females.

Resource-Transfer Model. In the gene-transfer model, a young wife's only kin among the younger consumers of her group are her own offspring. From her genetic perspective, these are the rightful focus of her (and her husband's) provisioning. By contrast, her husband is one among a group of related males and has a genetic interest in their offspring as well as his own. Therefore, he gains more inclusive fitness from contributions to a common pot than does his wife, and has less incentive to hoard food for his own household.

For purposes of abstraction, I will assume men put food into a communal pot from which all juveniles feed whereas women put food into a household pot from which their own offspring feed. Each additional male hunter enhances the amount of resources put into the communal pot (and spreads the risks from hunting; 12) whereas each additional female gatherer competes for local resources with other households. Mothers increase their fecundity if they can shift some of the burden of childrearing from the household to the communal pot. To mix metaphors, the communal pot is a commons on which the offspring of different households graze and, if unregulated, is expected to be overstocked (3, 13). Fathers favor longer interbirth intervals (lower stocking rates) than mothers because of their greater relatedness to offspring of other households. (This prediction is sensitive to relaxation of the assumptions of stable monogamy and patrilocal residence).

Consumption Model. In the above model, offspring have 3 sources of food: items collected by their mother and contributed to the household pot, items contributed to the communal pot by adult males, and items collected by the child itself. Food taken from the household pot reduces food available to sibs whereas food taken from the communal pot reduces food available to offspring of all households. Food that an offspring forages for itself reduces demands on the household and/or the communal pot. Given a choice between taking an equivalent item from the household pot or communal pot, a child would generally prefer the communal pot because the personal benefit is the same but the cost is spread across a larger group of less-related kin.

Conflicts over resource consumption can exist within the genomes of children between genes of maternal and paternal origin. Suppose that faster feeding from a pot results in the guzzler obtaining more food (individual benefit) at the cost of greater spillage (group cost). The consumers from each household pot comprise a smaller group of maternally-derived alleles than paternally-derived alleles (because some consumers are maternal half-sibs with different fathers) whereas the consumers from the communal pot comprise a smaller group of paternally-derived alleles than maternally-derived alleles (because consumers are the progeny of related fathers but unrelated mothers). Therefore, paternally-expressed genes are predicted to promote faster eating and greater spillage from the household pot, whereas maternally-expressed genes are predicted to promote faster eating and greater spillage from the communal pot (14). If younger children are fed from the household pot but older children are fed from the communal pot, then maternally-derived alleles of children would favor graduation to the communal pot at a younger age than the age favored by paternally-derived alleles.

As a consequence of asymmetries of patrilineal and matrilineal relatedness within social groups, imprinted genes are predicted to influence how much a child consumes and at whose expense, and to accelerate or retard transitions between life-history stages.

Ontogeny of Resource Transfers

Human development is associated with a series of transitions that influence the pattern of resource transfers. After ovulation, the early embryo subsists on tubal and uterine secretions, and reserves deposited in the oocyte. After implantation, the offspring gains direct access to maternal blood via a hemochorial placenta. Parturition is marked by the abrupt loss of the placental haustorium and its replacement by suckling at the breast. Weaning is a more or less gradual process by which milk is first supplemented, and then replaced, by other foods. Adrenarche is defined by increased secretion of adrenal androgens but coincides approximately with the eruption of the first permanent molars and the behavioral and cognitive changes known as the 5- to 7-year shift (15, 16). Gonadarche marks the start of the transition from life as a non-reproductive consumer to life as a reproductive provider. Sexual maturation may be associated with dispersal from the natal group.

The question whether to invest in a child comes logically before the question how much to invest. The next section considers "decisions" to terminate investment and redirect resources to other fitness-enhancing activities. Subsequent sections discuss genetic conflicts over amounts transferred and the timing of ontogenetic transitions. These sections will consider 4 disorders of imprinted gene expression (Table 1): Beckwith–Wiedemann syndrome

(BWS) is associated with excess expression of paternal alleles or deficient expression of maternal alleles, and is expected to exaggerate phenotypic effects that enhance a child's patrilineal inclusive fitness at a cost to its matrilineal inclusive fitness. Silver–Russell syndrome (SRS), Prader–Willi syndrome (PWS), and Temple syndrome (TS) are associated with excess expression of maternal alleles, or deficient expression of paternal alleles, and are expected to exaggerate phenotypic effects that enhance a child's matrilineal inclusive fitness at a cost to its patrilineal inclusive fitness.

Selective Abortion and Infanticide. The death of a child (or embryo) is associated with a fitness cost for its parents and an inclusive fitness cost for other kin, but this loss of fitness via the child can sometimes be compensated by fitness gains via other individuals. Each individual will favor termination of investment in a child if resources can be redeployed to other uses that yield a higher return of inclusive fitness. The costs and benefits of termination may be weighed differently by different group members, depending on their relative relatedness to the discarded child and to other individuals who would benefit from its abandonment. The child is the group member most closely related to itself and, therefore, the least likely to favor its own elimination.

If a child were to die at 7 years, it were better that an infant die at 7 months; if an infant were to die at 7 months, it were better that a babe die at 7 days; and if a babe were to die at 7 days post partum, it were better still that an embryo die at 7 days post conception. It takes 1 death to eliminate 1 copy of a deleterious dominant allele (or 2 copies of a deleterious recessive allele), and the death will have less effect on parental fitness the earlier it occurs. Reproductive compensation for early deaths thus favors the evolution of screening processes, operating before major commitment of resources, that convert small differences of expected fitness into lethal differences (17–19).

Early embryos are easily replaced. Therefore, mothers are expected to be "fastidious" about which embryos implant, and to abandon embryos much more readily than they would abandon a child. The genetic diseases we see at birth are disorders that were compatible with prenatal survival and that evaded detection in utero. Two considerations probably contribute to the inefficiency of prenatal screening. The first is the difficulty of testing many aspects of postnatal gene function in embryos. The second is parent–offspring conflict: embryos have less stringent criteria for continuation of pregnancy than mothers (19).

Mothers probably have effective control over whether, and how much, to invest in offspring during the earliest stages of pregnancy. Implantation, however, marks a major shift in power from mother to offspring. Once an embryo is securely ensconced within the uterus, the offspring has greater control over the delivery of maternal investment than it has at any postnatal stage. Beyond the early stages of pregnancy, a fetus probably has an effective veto on termination of maternal investment and effective control over when to be delivered.

Power to control maternal investment shifts decisively back toward the mother at birth, when a nipple replaces the placenta as the conduit for nutrient transfer. The immediate postnatal period may be the first opportunity to abandon an offspring since the earliest stages of pregnancy (20, 21). Decisions to abandon infants may be influenced by group members with genetic interests distinct from either mother or infant.

Till Birth Do Us Part. Genes of paternal origin are predicted to promote increased demands on mothers during pregnancy whereas genes of maternal origin are predicted to promote reduced demands (6, 7). Strong support for these predictions comes from BWS and SRS, the former associated with fetal overgrowth and the latter with intrauterine growth retardation (22, 23).

The notoriously tight fit between the size of the fetal head and the width of the maternal pelvis suggests that fetuses "choose" to remain inside their mother until the last practical moment. Offspring must have been substantially safer within the uterus than at the breast if the benefits of prolonged gestation were to have outweighed the increased risks of birth complications. I conjecture that longer gestation enhanced the average fitness of offspring but reduced the average fitness of their mothers. There is limited evidence for effects of imprinted genes on gestation length. Gestation is shortened by 2–3 weeks in SRS (22) and preterm and postterm delivery are both increased in PWS (24).

Infancy and Early Childhood. The costs of lactation are borne directly by mothers, although maternal costs may be subsidized by other group members. Supplemental foods are typically introduced early, usually by 6 months (25). At first, these foods must be specially prepared because the infant's gut and dentition are immature. In natural fertility populations, cessation of suckling usually occurs at some time before a child's third birthday with the onset of the mother's next pregnancy (26). More intense suckling and later introduction of solid food prolong lactational amenorrhea (27). Conversely, less intense suckling would shorten interbirth intervals and cause earlier weaning.

Shorter interbirth intervals are associated with increased maternal fecundity but reduced offspring survival (28). Therefore, maternally-derived alleles of infants are predicted to favor lower intensity suckling, greater appetite for supplemental foods, and earlier weaning than paternally-derived alleles. Poor suck is characteristic of infants with SRS, PWS, and TS (24, 29–32). Moreover, infants with SRS show little interest in food and require small frequent feeds (29–31). The large tongues of infants with BWS (33) suggest a role for paternally-expressed genes in the development of the infant's suction pump.

Postnatal feeding is severely perturbed in PWS. This syndrome has been classically described as having 2 phases: poor suck and failure to thrive in infancy, followed by hyperphagia and onset of obesity in early childhood. Recent reviews suggest a more complex 3-phase history with the onset of obesity (18–24 months) occurring before the onset of hyperphagia (5–13 years) (34, 35), but other reviews continue to describe hyperphagia as commencing before excessive weight gain (24).

Two articles have interpreted the change in appetite observed in PWS as a reflection of evolutionary conflict between maternal and paternal alleles over food transfers from parents. Haig and Wharton (36) proposed that paternally-expressed transcripts promote suckling during early infancy but inhibit appetite for supplemental foods at the time of weaning. In the absence of these transcripts, infants with PWS have poor suck during the period of exclusive lactation but develop an insatiable appetite at the time of weaning. The effect of these transcripts was to increase reproductive costs to mothers by extending lactational amenorrhea for the benefit of the offspring (36). Úbeda (37) argued that fathers pay a greater proportion of provisioning costs after weaning and that maternally-expressed transcripts from the PWS region increase demands on fathers for the benefit of weaned offspring. Both articles were based on a 2-phase model of PWS with onset of hyperphagia at the time of weaning. If the onset of hyperphagia is delayed to 5 years or later, then these models may need to be revised because this age correlates more with adrenarche than weaning.

Adrenarche and Preadolescence. Adrenarche occurs at about the age (5–7 years) that the child's immediately younger sib is being weaned because the child's mother is pregnant with the next younger sib. This is an age of significant behavioral and cognitive changes. Children are given more responsibilities, interact more with peers, and develop social norms of reciprocity (15). The sharing decisions of 3- to 4-year-olds are mainly self-centered, whereas 7- to 8-year-olds will share food equitably within their

social group (38). Whether these changes are influenced by adrenal androgens is currently unknown.

I conjecture that adrenarche occurs at an age when the child's sustenance was shifting from predominant reliance on the household pot to greater reliance on the communal pot and self-provisioning. Therefore, genes of maternal origin might be expected to favor earlier adrenarche than genes of paternal origin. Consistent with this hypothesis, premature adrenarche is common in PWS (39), but I know of no data that allow comparison of the relative timing of adrenarche and onset of hyperphagia in PWS.

Children older than 6 years are often expected to work and, cross-culturally, women rather than men are the principal beneficiaries of children's labor (40). Anthropologists have observed marked variation in children's contribution to their own upkeep among hunter-gatherer societies (41, 42). The hyperphagic phase of PWS is associated with behaviors that have been variously described as "foraging" and "food-stealing" (43). It is possible that these terms encompass functionally distinct behaviors. Foraging could reduce withdrawals from the household pot, whereas stealing could increase withdrawals from the communal pot, with both behaviors benefiting genes of maternal origin, but this is mere speculation until the behaviors are better characterized.

Postnatal growth in BWS is characterized by marked acceleration of bone age in infancy and early childhood. Final height is on average 2.5 SD above the mean (44). By contrast, bone age is often delayed in young children with SRS and final height is on average 3.6 SD below the mean (45). This reduction in final height results from slow growth in utero and during the first postnatal months with absence of subsequent catch-up growth (22, 46). In both BWS and SRS, spontaneous puberty occurs at the normal age (44, 45). Length in PWS is within the normal range through the first postnatal year, but declines to the third centile by 3 years, with a further loss in relative height due to absence of the pubertal growth spurt (46).

Fetal and early infant growth are severely perturbed in BWS and SRS. Therefore, genes from chromosome 11p15.5 appear to be major regulators of growth during this period. By contrast, birthweight is within the normal range in PWS, but growth subsequently falters. Therefore, paternally-expressed genes from chromosome 15q11-q13 appear to promote postnatal growth, at least in part, via increased secretion of growth hormone (GH) because GH therapy restores normal adult height in individuals with PWS (47). These observations suggest 2, partially dissociable, phases of growth during early childhood (48).

Childhood is a period of prolonged slow growth. Clinical data from imprinting disorders suggest paternally-expressed genes promote, and maternally-expressed genes inhibit, childhood growth. This implies that larger size benefited offspring at a cost to their mothers' residual reproductive value, although the nature of this tradeoff is not altogether clear. Slow growth, with delayed puberty, would have reduced the rate at which food needed to be supplied to offspring but could have increased the total transfers needed to raise a child to independence.

Puberty. Premature puberty is characteristic of TS (49). Precocious early signs of puberty are also common in PWS (50), but pubertal progression is incomplete, with a weak or absent growth spurt (24, 51). Thus, imprinted genes influence the timing of gonadarche and the pubertal growth spurt, but a clear pattern is absent, perhaps because of the complexity of the underlying selective forces.

In my simple models, gonadarche occurs in the natal group of both sexes. But, when a young couple reproduces, their offspring consume resources in his natal group but her dispersal group. The selective forces associated with the onset of gonadarche are complex because they depend on the relative timing of the transition from being a net consumer to a net producer, of the dispersal of females, and of first reproduction in both sexes.

Offspring presumably benefited from each additional year of prereproductive development by accumulating social, and other, experience that allowed them to function as more effective adults, but this learning period was subsidized by withdrawals from the household and communal pots. From the perspective of other group members, ecological conditions could shift the balance between the benefits of adding an extra adult-sized contributor to the communal pot (favoring earlier male entry to adulthood) and the costs of increased competition from adding an extra household (favoring delayed male entry to adulthood). In contrast, if a daughter's ended sojourn in her natal group increases local resource competition, then one might predict that her paternally-derived genes would favor earlier maturation and dispersal.

Age at puberty is variable among and within human groups and shows strong secular trends (52). Moreover, pubertal timing appears sensitive to both nutritional and social cues. For example, children born in the developing world, but adopted by European families, have high rates of precocious puberty (53) and father absence is associated with early menarche (54). Whether this variability reflects evolved responses to cues of local relatedness and resource transfers remains an open question.

Evolution of Human Life History

Humans take longer to reach nutritional independence than other great apes but have shorter interbirth intervals (25, 55, 56). As a consequence of these two derived features of human life history, mothers often care for "litters" of different-aged offspring. Roughly speaking, a human mother can produce 2 offspring in the time it takes a chimpanzee mother to produce one. Weaning, adrenarche, and first molar eruption are approximately contemporaneous in chimpanzees, but a human mother is weaning her second offspring by the time her first offspring is undergoing adrenarche and cutting its first molar. As a further contrast, a chimpanzee weanling is responsible for feeding itself but a human weanling is fed by others for many years.

I conjecture that prolonged maturation was an adaptation of human offspring that enhanced their individual fitness at a cost to their mothers' fecundity whereas early weaning was an adaptation of mothers that enhanced their fecundity at the expense of offspring survival. This hypothesis is based on substantial evidence that paternally-expressed genes favor more intense suckling, and suggestive evidence that maternally-expressed genes favor earlier sexual maturation. The observation that imprinted genes influence resource transfers and ontogenetic transitions suggests that our distinctive life history has been shaped by conflicting interests of different sets of genes distributed among the individuals of social groups. The resulting life history may be an evolutionary compromise with substantial inefficiencies because of conflict costs.

One of the most promising avenues for testing these ideas will be detailed longitudinal studies of feeding behavior, adrenarche, and pubertal progression in children with various imprinting disorders. Such studies would not only be of evolutionary interest but also of clinical value. For example, many of the health problems in individuals with PWS are associated with obesity and hyperphagia, but the ages of onset of excessive weight gain and excessive feeding appear to be highly variable among individuals. It would be useful to know whether this variability is "random noise" or is systematically associated with differences in family dynamics and how food is presented.

ACKNOWLEDGMENTS. I thank Barry Bogin, Bernie Crespi, Sarah Hrdy, Chris Kuzawa, Ronald Lee, Frank Marlowe, Charles Nunn, Stephen Stearns, Francisco Úbeda and the Fundamental Interconnectedness of All Things Discussion Group for their comments.

1. Kaplan H (1994) Evolutionary and wealth flows theories of fertility: Empirical tests and new models. *Pop Devel Rev* 20:753–791.
2. Lee RD (2003) Rethinking the evolutionary theory of aging: Transfers, not births, shape senescence in social species. *Proc Natl Acad Sci USA* 100:9637–9642.
3. Lee R (2008) Sociality, selection, and survival: Simulated evolution of mortality with intergenerational transfers and food sharing. *Proc Natl Acad Sci USA* 105:7124–7128.
4. Bourke AFG (2007) Kin selection and the evolutionary theory of aging. *Annu Rev Ecol Evol Syst* 38:103–128.
5. Trivers RL (1974) Parent-offspring conflict. *Amer Zool* 14:249–264.
6. Haig D (2000) The kinship theory of genomic imprinting. *Annu Rev Ecol Syst* 31:9–32.
7. Haig D (2004) Genomic imprinting and kinship: How good is the evidence? *Ann Rev Genet* 38:553–585.
8. Ember CR (1978) Myths about hunter-gatherers. *Ethnology* 17:439–448.
9. Alvarez HP (2004) In *Kinship and Behavior in Primates*, eds Chapais B, Berman CM (Oxford Univ Press, New York). pp 420–442.
10. Marlowe FW (2004) Marital residence among foragers. *Curr Anthropol* 45:277–284.
11. Haig D (2000) Genomic imprinting, sex-biased dispersal, and social behavior. *Ann NY Acad Sci* 907:149–163.
12. Smith EA (1988) In *Hunters and gatherers*, eds Ingold T, Riches D, Woodburn J (Berg, Oxford), Vol 1, pp 222–251.
13. Hardin G (1968) The tragedy of the commons. *Science* 162:1243–1248.
14. Haig D, Wilkins JF (2000) Genomic imprinting, sibling solidarity, and the logic of collective action. *Phil Trans R Soc London Ser B* 355:1593–1597.
15. Campbell B (2006) Adrenarche and the evolution of human life history. *Am J Hum Biol* 18:569–589.
16. Del Giudice M, Angeleri R, Manera V (2009) The juvenile transition: A developmental switch point in human life history. *Devel Rev* 29:1–31.
17. Kozlowski J, Stearns SC (1989) Hypotheses for the production of excess zygotes: Models of bet-hedging and selective abortion. *Evolution* 43:1369–1377.
18. Haig D (1990) Brood reduction and optimal parental investment when offspring differ in quality. *Amer Nat* 136:550–556.
19. Haig D (2009) Fertile soil or no man's land: Cooperation and conflict in the placental bed. In *Human Placental Bed Vascular Failure*, eds Pijnenborg R., Brosens I, Romero R (Cambridge Univ Press, Cambridge), in press.
20. Langer WL (1974) Infanticide: A historical survey. *Hist Childh Q* 1 (3):353–365.
21. Overpeck MD, Bremer RA, Trumble AC, Trifiletti LB, Berendes HW (1998) Risk factors for infant homicide in the United States. *N Engl J Med* 339:1211–1216.
22. Kotzot D (2007) Growth parameters in maternal uniparental disomy 7 and 14. *Eur J Ped* 166:1143–1149.
23. Eggermann T, Eggermann K, Schönherr N (2008) Growth retardation versus overgrowth: Silver–Russell syndrome is genetically opposite to Beckwith–Wiedemann syndrome. *Trends Genet* 24:195–204.
24. Butler MG, Hanchett JM, Thompson T (2006) *Management of Prader–Willi Syndrome*, eds Butler MG, Lee PDK, Whitman BY (Springer, New York), 3rd Ed, pp 3–48.
25. Sellen DW (2007) Evolution of infant and young child feeding: Implications for contemporary public health. *Annu Rev Nutr* 27:123–148.
26. Sellen DW, Smay DB (2001) Relationship between subsistence and age at weaning in "preindustrial" societies. *Hum Nat* 12:47–87.
27. Taylor HW, Vázquez-Geffroy M, Samuels SJ, Taylor DM (1999) Continuously recorded suckling behaviour and its effect on lactational amenorrhoea. *J Biosoc Sci* 31:289–310.
28. Koenig MA, Phillips JF, Campbell OM, D'Souza S (1990) Birth intervals and childhood mortality in rural Bangladesh. *Demography* 27:251–265.
29. Wollmann H, Kirchner T, Enders H, Preece MA, Ranke MB (1995) Growth and symptoms in Silver–Russell syndrome: Review on the basis of 386 patients. *Eur J Ped* 154:958–968.
30. Price SM, Stanhope R, Garrett C, Preece MA, Trembath RC (1999) The spectrum of Silver–Russell syndrome. *J Med Genet* 36:837–842.
31. Hannula K, Kere J, Pirinen S, Holmberg C, Lipsanen-Nyman M (2001) Do patients with maternal uniparental disomy for chromosome 7 have a distinct mild Silver–Russell phenotype? *J Med Genet* 38:273–278.
32. Hordijk R, et al. (1999) Maternal uniparental disomy for chromosome 14 in a boy with normal karyotype. *J Med Genet* 36:782–785.
33. Pettenati MJ, et al. (1986) Wiedemann-Beckwith syndrome: Presentation of clinical and cytogenetic data on 22 new cases and review of the literature. *Hum Genet* 74:143–154.
34. McCune H, Driscoll D (2005) In *Pediatric Nutrition in Chronic Diseases and Developmental Disorders*, eds Ekvall SW, Ekvall VK (Oxford Univ Press, New York), 2nd Ed, pp 128–132.
35. Goldstone AP, Holland AJ, Hauffa BP, Hokken-Koelega AC, Tauber M (2008) Recommendations for the diagnosis and management of Prader–Willi syndrome. *J Clin Endocrinol Metab* 93:4183–4197.
36. Haig D, Wharton R (2003) Prader–Willi syndrome and the evolution of human childhood. *Am J Hum Biol* 15:320–329.
37. Úbeda F (2008) Evolution of genomic imprinting with biparental care: Implications for Prader–Willi and Angelman syndromes. *PLoS Biol* 6:e208.
38. Fehr E, Bernhard H, Rockenbach B (2008) Egalitarianism in young children. *Nature* 454:1079–1083.
39. Unanue N, et al. (2007) Adrenarche in Prader–Willi syndrome appears not related to insulin sensitivity and serum adiponectin. *Horm Res* 67:152–158.
40. Bradley C (1993) Women's power, children's labor. *Cross-Cult Res* 27:70–96.
41. Draper P (1976) In *Kalahari Hunter-Gatherers*, eds Lee RB, DeVore I (Harvard Univ Press, Cambridge MA), pp 199–217.
42. Blurton Jones NG, Hawkes K, O'Connell JF (1997) In *Uniting Psychology and Biology*, eds Segal N, Weisfeld G, Weisfeld C (American Psychological Association, Washington, DC) pp 279–313.
43. Page TJ, Finney JW, Parrish JM, Iwata BA (1983) Assessment and reduction of food stealing in Prader–Willi children. *Appl Res Mental Retard* 4:219–228.
44. Sippell WG, Partsch CJ, Wiedemann HR (1989) Growth, bone maturation and pubertal development in children with the EMG syndrome. *Clin Genet* 35:20–28.
45. Davies PSW, Valley R, Preece MA (1988) Adolescent growth and pubertal progression in the Silver–Russell syndrome. *Arch Dis Child* 63:130–135.
46. Wollmann HA, Schultz U, Grauer ML, Ranke MB (1998) Reference values for height and weight in Prader–Willi syndrome based on 315 patients. *Eur J Ped* 157:634–642.
47. Angulo MA, et al. (2007) Final adult height in children with Prader–Willi syndrome with and without human growth hormone treatment. *Am J Med Genet* 143A:1456–1461.
48. Hochberg Z, Albertsson-Wikland K (2008) Evo-devo of infantile and childhood growth. *Ped Res* 64:2–7.
49. Buiting K, et al. (2008) Clinical features of maternal uniparental disomy 14 in patients with an epimutation and deletion of the imprinted *DLK1/GTL2* gene cluster. *Hum Mut* 29:1141–1146.
50. Tauber M, et al. (2000) Auxological and endocrine evolution of 28 children with Prader–Willi syndrome. *Horm Res* 53:279–287.
51. Burman P, Ritzén EM, Lindgren AC (2001) Endocrine dysfunction in Prader–Willi syndrome. *Endocrine Rev* 22:787–799.
52. Parent AS, et al. (2003) The timing of normal puberty and the age limits of sexual precocity. *Endocrine Rev* 24:668–693.
53. Teilmann G, et al. (2007) Early pituitary-gonadal activation before clinical signs of puberty in 5- to 8-year-old adopted girls: A study of 99 foreign adopted girls and 93 controls. *J Clin Endocrinol Metab* 92:2538–2544.
54. Matchock RL, Susman EIJ (2006) Family composition and menarcheal age: Anti-inbreeding strategies. *Am J Hum Biol* 18:481–491.
55. Bogin B (1997) Evolutionary hypotheses for human childhood. *Yearb Phys Anthropol* 40:63–89.
56. Kaplan HS, Robson AJ (2002) The emergence of humans: The coevolution of intelligence and longevity with intergenerational transfers. *Proc Natl Acad Sci USA* 99:10221–10226.
57. Cooper WN, et al. (2005) Molecular subtypes and phenotypic expression of Beckwith–Wiedemann syndrome. *Eur J Hum Genet* 13:1025–1032.
58. Bruce S, Hannula-Jouppi K, Peltonen J, Kere J, Lipsanen-Nyman M (2009) Clinically distinct epigenetic subgroups in Silver–Russell syndrome. *J Clin Endocrinol Metab* 94:579–587.
59. Cassidy SB, Driscoll DJ (2009) Prader–Willi syndrome. *Eur J Hum Genet* 17:3–13.

Comparative genomics of autism and schizophrenia

Bernard Crespi[1], Philip Stead, and Michael Elliot

Department of Biological Sciences, Simon Fraser University, Burnaby, BC, Canada V5A 1S6

Edited by Stephen Curtis Stearns, Yale University, New Haven, CT, and accepted by the Editorial Board September 29, 2009 (received for review June 30, 2009)

We used data from studies of copy-number variants (CNVs), single-gene associations, growth-signaling pathways, and intermediate phenotypes associated with brain growth to evaluate four alternative hypotheses for the genomic and developmental relationships between autism and schizophrenia: (*i*) autism subsumed in schizophrenia, (*ii*) independence, (*iii*) diametric, and (*iv*) partial overlap. Data from CNVs provides statistical support for the hypothesis that autism and schizophrenia are associated with reciprocal variants, such that at four loci, deletions predispose to one disorder, whereas duplications predispose to the other. Data from single-gene studies are inconsistent with a hypothesis based on independence, in that autism and schizophrenia share associated genes more often than expected by chance. However, differentiation between the partial overlap and diametric hypotheses using these data is precluded by limited overlap in the specific genetic markers analyzed in both autism and schizophrenia. Evidence from the effects of risk variants on growth-signaling pathways shows that autism-spectrum conditions tend to be associated with up-regulation of pathways due to loss of function mutations in negative regulators, whereas schizophrenia is associated with reduced pathway activation. Finally, data from studies of head and brain size phenotypes indicate that autism is commonly associated with developmentally-enhanced brain growth, whereas schizophrenia is characterized, on average, by reduced brain growth. These convergent lines of evidence appear most compatible with the hypothesis that autism and schizophrenia represent diametric conditions with regard to their genomic underpinnings, neurodevelopmental bases, and phenotypic manifestations as reflecting under-development versus dysregulated over-development of the human social brain.

genetics | evolution | psychiatry

The Swiss psychiatrist Eugen Bleuler coined the terms "schizophrenia", for the splitting of psychic functions, and "autism", for withdrawal from external reality in patients with schizophrenia, almost exactly a century ago (1). Ever since 1943, when Leo Kanner (2) coopted autism to refer to a new condition involving "disturbance of affective contact" manifested in children, the relationship between schizophrenia and Kanner's autism has remained unclear (3). Kanner originally conceived autism as an early, distinct subtype of schizophrenia (model 1A) (Fig. 1*A*), a view he later renounced in favor of a model, which was also supported by Rutter (4) with the conditions as distinct, separate, and unrelated (model 1B) (Fig. 1*B*). Under each of these two hypotheses, autism and schizophrenia may each grade more or less smoothly, and independently, into so-called normality. Schizophrenia and autism have also been considered as diametric, or opposite sets of conditions (model 1C) (Fig. 1*C*) along a spectrum of social-brain phenotypes from hypodevelopment in autism, to normality, to hyperdevelopment in schizophrenia (5). By a fourth model, autism overlaps broadly yet partially with schizophrenia, sharing some risk factors and phenotypes but not others (model 1D) (Fig. 1*D*). This latter model has been motivated by recent genetic evidence of shared loci and pathways, mediating both autism and schizophrenia (6–8), and by work describing social deficits as central to both Kanner's autism and the "autistic" symptoms of Blueler's schizophrenia (9,10).

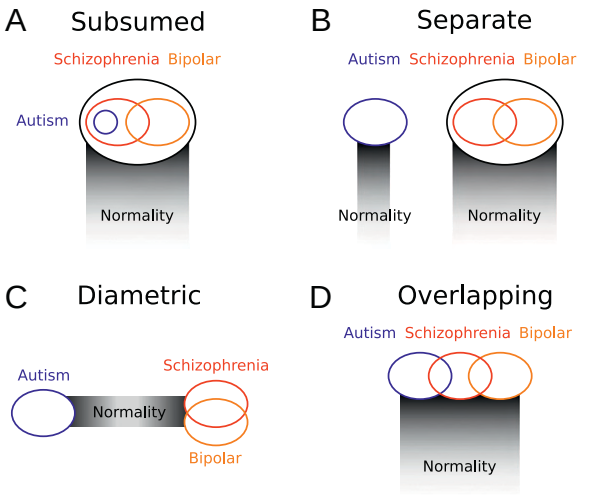


Fig. 1. Alternative models for the genomic and etiological relationships of autism with schizophrenia and bipolar disorder. (*A*) Subsumed. (*B*) Separate. (*C*) Diametric. (*D*) Overlapping.

Differentiating alternative models for the relationships between major human psychiatric conditions has important implications for diagnoses, pharmacological and psychological treatment, and strategies for dissection of the etiology of these disorders at all levels from genes to cognition. In a series of recent papers, Neil Craddock and others (11–13) have demonstrated how genetic data can be deployed to evaluate alternative hypotheses for the relationship of schizophrenia with bipolar disorder, originally described as a dichotomy in work that followed from the pioneering studies of Emil Kraepelin, but now increasingly viewed in terms of intergradation and some form of partial overlap, based on the presence of shared genetic risk factors mediating shared phenotypes (Fig. 1).

In this article, we evaluate alternative hypotheses for the relationship of autistic spectrum conditions with schizophrenia spectrum conditions, by using recent genomic and genetic data to test the predictions that differentiate between models. Thus, under model 1A, alleles or haplotypes affecting autism risk should represent some subset of a larger pool of schizophrenia risk alleles or haplotypes; under model 1B, the two sets of conditions should exhibit independent risk factors; under model 1C, the conditions should be mediated by alternative risk factors

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Evolution in Health and Medicine" held April 2–3, 2009, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS web site at www.nasonline.org/Sackler_Evolution_Health_Medicine.

Author contributions: B.C. designed research; B.C., P.S., and M.E. performed research; B.C. and M.E. analyzed data; and B.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. S.C.S. is a guest editor invited by the Editorial Board.

[1]To whom correspondence should be addressed. E-mail: crespi@sfu.ca

This article contains supporting information online at www.pnas.org/cgi/content/full/0906080106/DCSupplemental.

Table 1. Contingency table analyses of the relative frequencies of copy-number deletions and duplications associated with autism or schizophrenia

CNV locus	Condition	Deletion cases (refs.)	Duplication cases (refs.)	*P*-value (Fisher's exact test)
1q21.1	Autism	**2** (14, 15)	**10** (14–17)	0.001
	Schizophrenia	**15** (18–21)	4 (18, 22)	
15q13.3	Autism	3 (19, 23)	2 (24)	0.849
	Schizophrenia	**10** (18, 19)	4 (18)	
16p11.2	Autism	**14** (16, 25–30)	5 (16, 28, 30)	0.00013
	Schizophrenia	5 (16, 18, 19, 31)	**24** (18, 20, 31)	
16p13.1	Autism	0	**3** (32)	0.434
	Schizophrenia	8 (18, 21, 33)	**23** (18, 22, 33)	
17p12	Autism	4 (14, 16, 26)	1 (14)	0.385
	Schizophrenia	**8** (18, 19, 22)	0	
22q11.21	Autism	1 (14)	**8** (14, 26, 28, 30)	0.000049
	Schizophrenia	**16** (18, 21, 22, 34)	1 (21)	
22q13.3	Autism	**5** (25, 28, 30)	0	0.0079
	Schizophrenia	0	**4** (18)	

Entries in bold indicate that the CNV is statistically documented as a risk factor for the condition specified, from case-control comparisons (Tables S3–S9).

with diametric effects on development; and under model 1D, autism and schizophrenia are expected to overlap for some set of genetic or genomic factors underlying liability to shared psychiatric phenotypes, as do bipolar disorder and schizophrenia. Model 1A and model 1D also predict that overlapping risk factors should influence developmental and phenotypic similarities between autism and schizophrenia. By contrast, model 1C predicts that these two sets of conditions should exhibit phenotypes that deviate in opposite directions from normality, at least for traits (such as aspects of growth or synaptic function) that are underlain by loci subject to variation or perturbation that can cause deviations from typical neurodevelopment or function in opposite directions.

Results

Copy Number Data. Rare copy-number variants (CNVs) at seven loci, 1q21.1, 15q13.3, 16p11.2, 16p13.1, 17p12, 22q11.21, and 22q13.3 (Tables S1 and S2), have been independently ascertained and associated with autism and schizophrenia in a sufficient number of microarray-based comparative genomic hybridization (aCGH) and SNP-based studies to allow statistical analysis of the frequencies of deletions versus duplications in these two conditions (Table 1, Tables S3–S9). For five of the loci (1q21.1, 16p11.2, 16p13.1, 22q11.21, and 22q13.3), specific risk variants have been statistically supported for both autism and schizophrenia using case-control comparisons, which allows direct evaluation of the alternative hypotheses in Fig. 1. One locus (16p13.1) supports a model of overlap, and four loci support the reciprocal model, such that deletions are associated with increased risk of autism and duplications with increased risk of schizophrenia (16p11.2, 22q13.3), or deletions are associated with increased risk of schizophrenia and duplications with increased risk of autism (1q21.1, 22q11.21). For 1q21.1 and 22q11.21, contingency table analyses also indicate highly-significant differences in the frequencies of deletions compared with duplications for the two disorders, such that schizophrenia is differentially associated with deletions and autism with duplications. By contrast, for 16p11.2 and 22q13.3 such analyses show that autism is differentially associated with deletions and schizophrenia with duplications.

Genetic Association Data. Of 45 genes evaluated for association with both autism and schizophrenia and with replicated positive associations, 20 genes exhibit one or more positive associations with both conditions, compared with 2 genes positive for autism but not schizophrenia and 11 genes positive for schizophrenia but not autism; 12 genes showed negative results for both conditions (Table 2). Under a hypothesis in which associations with schizophrenia and autism are distributed independently of each other (corresponding to one interpretation of the "separate" model) (Fig. 1*B*), 20 or more genes associated with both conditions would be observed with a probability of 0.002 (Table S10). We interpret these results as inconsistent with a separate and independent relationship of autism and schizophrenia. The existence of genes associated with autism but not schizophrenia is inconsistent with a strict interpretation of Kanner's original "subsumed" model (Fig. 1*A*) in which autism represents a subtype of schizophrenia; however, the presence of only two genes in this category mitigates against robust rejection of this model, and the relative lack of genetic-association studies of autism compared with schizophrenia may also partially explain this result.

Models 1C (diametric) and 1D (overlapping) both predict broad overlap in risk genes between autism and schizophrenia, and do not necessarily predict an absence or paucity of genes affecting one condition but not the other. In theory, these models can be differentiated by using data on specific risk alleles for specific loci (such as single-nucleotide polymorphisms, haplotypes, or genotypes), which should be partially shared under the overlapping model but different under the diametric model. For the genes *DAO*, *DISC1*, *GRIK2*, *GSTM1*, and *MTHFR*, the same allele, genotype, or haplotype was associated with both autism and schizophrenia, and for the genes *AHI1*, *APOE*, *DRD1*, *FOXP2*, *HLA-DRB1*, and *SHANK3*, alternative alleles, genotypes, or haplotypes at the same loci appear to mediate risk of these two conditions (*SI Text*). For the other genes that have been associated with both conditions, heterogeneity in the genetic markers used, heterogeneity among results from multiple studies of the same genes, and the general lack of functional information preclude interpretation in terms of shared or alternative risk factors.

Discussion

Delineation and classification of psychiatric conditions evolves through an iterative process, driven by societal pressures (35) and advances in our understanding of symptoms and etiology

Table 2. Patterns in overlap between autism-associated genes and alleles and schizophrenia-associated genes and alleles

Autism and schizophrenia						
Shared alleles	Different alleles	Different markers	Complex overlap	Autism, not schiz.	Schiz., not autism	Not autism, not schiz.
DAO†	*AHI1*†	*BDNF*	*CNTNAP2*†	*NRCAM*†	*CTLA4*	*CYP21A2*
DISC1†	***APOE***	*DRD3*†	***COMT***†	*SLC25A12*	*DAOA*	*DBH*
GRIK2	***DRD1***	*EGF*	*NRXN1*†		***DRD2***	*DDC*
GSTM1†	*FOXP2*†	*NTNG1*†	*SLC6A4*†		***DRD4***	*GABRA5*
MTHFR†	*HLA-DRB1*	***RELN***†			***HTR2A***	*GABRG2*
	SHANK3†				*HTR7*	*GABRP*
					NOTCH4	*GSTP1*
					NRG1	*PENK*
					SLC6A3	*RYR3*
					TH	*SYNGAP1*
					TPH1	*TYR*
						YWHAB

Bold genes are included in the list of top 30 SZgene candidates (http://www.schizophreniaforum.org/res/sczgene/default.asp), as determined by effect size.
†At least one of the studies of autism showing positive associations used diagnostic criteria of "autism spectrum," "autistic spectrum," PDD-NOS, or "broad spectrum" autism, as compared to just "autism."

(36). The last several years have been characterized by remarkable advances in our understanding of the genetic and genomic underpinnings of autism, schizophrenia, and bipolar disorder, and evaluation of explicit alternative hypotheses (37) has led to progress well beyond a simple, century-old Kraepelian dichotomy of bipolar disorder distinguished from schizophrenia (12). In this article, we have evaluated alternative hypotheses for the relationship between schizophrenia and autism, using evidence from CNVs and genetic-association studies.

Models of autism as a subset of schizophrenia (Fig. 1*A*), and autism and schizophrenia as independent or separate (model 1B), can be rejected with some degree of confidence, but models involving diametric etiology (model 1C) or partial overlap (model 1D) cannot be clearly rejected. Taken together, most of the data and analyses described here appear to support the hypothesis of autism and schizophrenia as diametric conditions, based primarily on the findings that reciprocal variants at 1q21.1, 16p11.2, 22q11.21, and 22q13.3 represent statistically-supported, highly-penetrant risk factors for the two conditions (Table 1), and that for a number of genes, alternative alleles or haplotypes appear to mediate risk of autism versus schizophrenia.

Additional lines of evidence supporting the diametric hypothesis, from previous studies of autism and schizophrenia, include:

1. Data showing notable rarity of familial coaggregation of autism with schizophrenia (38), in contrast, for example, to strong patterns of co-occurance within pedigrees of schizophrenia, schizoaffective disorder, and bipolar disorder (39).

2. Psychiatric contrasts of Smith-Magenis syndrome with Potocki-Lupski syndrome (due to the reciprocal duplication at the Smith-Magenis locus), Williams syndrome with cases of Williams-syndrome region duplication, and Klinefelter syndrome with Turner syndrome, each of which tends to involve psychotic-affective spectrum phenotypes in the former syndrome, and autistic spectrum conditions in the latter (5, 40).

3. Effects of autism and schizophrenia risk alleles on common growth-signaling pathways, such that autism has been associated with loss of function in genes, such as *FMR1*, *NF1*, *PTEN*, *TSC1*, and *TSC2* that act as negative regulators of the PI3K, Akt, mTOR, or other growth-signaling pathways (41–45), whereas schizophrenia tends to be associated with reduced function or activity of genes that up-regulate the PI3K, Akt, and other growth-related pathways (46–49).

4. Increased average head size, childhood brain volume, or cortical thickness in individuals with: (*i*) idiopathic autism (50–53), (*ii*) the autism-associated duplications at 1q21.1 (17) and 16p13.1 (32) and the autism-associated deletions at 16p11.2 (31), and (*iii*) autism due to loss of function (or haploinsufficiency) of *FMR1* (54), *NF1* (55), *PTEN* (56) and *RNF135* (57). By contrast, reduced average values for brain size and cortical thickness, due to some combination of reduced growth and accelerated gray matter loss, have been demonstrated with notable consistency across studies of schizophrenia (58–62), and such reduced head or brain size has also been associated with the schizophrenia-linked CNVs at 1q21.1 and 22q11.21 (17, 63, 64), and with deletions of 16p13.1 (65).

Despite a convergence of evidence consistent with a model of autism and schizophrenia as diametric conditions, the presence of genetically-based risk factors common to autism and schizophrenia, including deletions, duplications, or specific alleles shared between the conditions (Tables 1 and 2), deletions within the *NRXN1* gene (66), down-regulated *RELN* signaling (67), reduced *GAD1* expression (68), high levels of oxidative stress mediated in part by the *GSTM1* deletion allele (69, 70), and altered folate metabolism mediated by the TT genotype of the *MTHFR* C677T locus (71) support a model of partially-overlapping etiology or indicate that some genetic variants may increase liability to both conditions in otherwise highly-vulnerable individuals.

Epidemiological reports of comorbidity between autism and schizophrenia (72, 73) may be generated by several processes: (*i*) true etiological overlap, (*ii*) diagnostic perspectives, common at least through the 1980s, that conflate autism with "childhood schizophrenia", and (*iii*) false-positive diagnoses of children with premorbidity to schizophrenia as subject to autistic spectrum conditions, especially Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS) (74–77). Similarly, evidence for risk loci, rather than specific risk alleles or haplotypes, differentially shared between autism and schizophrenia (6, 7, 78), supports exclusion of model 1B (Fig. 1) but cannot differentiate between models 1A, 1C, and 1D without data on association of specific alleles with these conditions. Further analyses of alternative models for the relationship of autism with schizophrenia may usefully test for genetic associations of specific alleles, haplotypes, or CNVs at relatively well-established schizophrenia risk loci with autism, using autistic patient populations strictly diagnosed to exclude children with conditions such as PDD-NOS or "autism spectrum", some of whom may actually be premorbid for schizophrenia.

Interpretation of autistic-spectrum conditions and schizophrenia-spectrum conditions as diametric for psychological and psychiatric traits is predicated on a hypothesized axis of neurodevelopment and cognitive-affective functioning that reflects the degree to which the human social brain may be under-

developed versus over-developed to diverse forms of dysfunction (5). Under-development of social phenotypes such as theory of mind, language, sense of self in relation to others, and reciprocal social interaction represent well-recognized manifestations of autism (79–81), which can in some models be linked with neurological phenotypes such as imbalance toward excitatory glutamatergic cortical neurotransmission (82–84) or increased local versus global processing of information (85, 86). By contrast, such psychotic traits as auditory hallucination and thought disorder, paranoia, megalomania, and ascription of causal purpose to inanimate objects may be interpretable in terms of dysregulated hyperdevelopment of language, theory of mind and sense of self, all traits that are highly derived and elaborated in the human lineage (5).

Differentiating between alternative models for the relationship of autism with schizophrenia has far-reaching implications for diagnosis, treatment, and causal analyses of both sets of conditions. Current diagnostic categorization places the most common childhood condition likely to involve premorbidity to schizophrenia, PDD-NOS, formally within the autism spectrum; this system thus implictly presumes either a general lack of premorbidity to schizophrenia sufficiently severe to result in autism spectrum diagnoses, or that autism and schizophrenia overlap in etiology and symptoms. The tendency for males to exhibit worse premorbidity to schizophrenia than females (87, 88), and for earlier-onset schizophrenia to exhibit a higher male bias and a stronger tendency to be mediated by CNVs rather than other factors (89, 90) suggests a notable risk for false-positive diagnoses of autistic spectrum conditions (75–77, 91–93). Apparent direct evidence of such risk comes from tendencies to diagnose autism spectrum conditions in children with deletions at 15q11.2, 15q13.3, and 22q11.21, and duplications of 16p11.2, CNVs for which high risk of schizophrenia has been established from studies of adults (16, 23, 31, 94–97). To the degree that autism and schizophrenia exhibit diametric genetically-based risk factors, inclusion of children premorbid for schizophrenia-spectrum conditions in studies of the genetic bases of autism will substantially dilute and confound the detection of significant results. Until the relationship between autism and schizophrenia is better understood, such risks could be minimized through explicit differential diagnosis of autism versus multiple complex developmental disorder (which is indicative of schizophrenia premorbidity, see ref. 98), and by subsetting genetic analysis of autism by infantile autism versus more broadly-conceived autistic spectrum categories. Finally, to the extent that autism and schizophrenia involve diametric alterations to aspects of neurological function such as hyperglutamatergic, over-excitatory cortical signaling in autism (82, 99) compared with schizophrenia symptoms mediated by hypoglutamatergic cortical states (100, 101), conceptual frameworks for developing pharmaceutical treatment of one set of conditions may prove useful for understanding the other. Fragile X syndrome provides an apparent example, as this autistic condition can be rescued in mice using antagonists of mGlur5 signaling (102), whereas agonists of this receptor represent among the more promising treatments for schizophrenia (103).

The Kraepelian dichotomy between schizophrenia and bipolar disorder has usefully guided psychiatric research programs for many years, but may now impede progress in the development of genetically-based models for the causes of psychotic and affective phenotypes in schizophrenia, bipolar disorder, and major depression (11, 12). Kanner's genius was to recognize a new syndrome among the apparent chaos of human childhood psychopathology, but his adoption of the term autism, already established by Bleuler in the context of social withdrawal in schizophrenia, has led to >65 years of conflating the two sets of conditions, the autistic spectrum and the schizophrenia spectrum, that share social deficits as central phenotypes, but whose causes may differ substantially. To the extent that the autism spectrum and the schizophrenia spectrum represent diametric disorders of the social brain, as suggested by some of the analyses described here, a predictive framework based in evolutionary theory can be developed to guide research into the etiologies of both sets of conditions.

Materials and Methods

Copy Number. CNVs for analysis were determined exclusively from publicly-available data in the aCGH and SNP-based studies listed in Table S2. The seven loci analyzed here, 1q21.1, 15q13.3, 16p11.2, 16p13.1, 17p12, 22q11.21, and 22q13.3 were selected for study strictly on the basis of two criteria: (*i*) statistical support from at least one study, or from analyses conducted here, for the loci as harboring risk variants for schizophrenia and/or autism enriched over controls, and (*ii*) the presence of at least six cases of autism (per se) or schizophrenia combined, with at least one case of autism and one case of schizophrenia associated with CNVs at the locus. Overlap in cases among studies, mainly involving AGRE, Stefansson et al. (19) and ISC (18) samples, was accounted for in ascertainment of cases for inclusion. Data from ref. 18 include information available at http://pngu.mgh.harvard.edu/isc/, and data from ref. 19 include only the authors' phase 1 sample, for which complete data are available for deletions and duplications in regions other than 1q21.1, 15q11.2, and 15q13.3 (19).

Deletions at 15q11.2 that include four genes, *TUBGCP5*, *CYFIP1*, *NIPA2*, and *NIPA1*, are a statistically-documented risk factor for schizophrenia (8, 19), but deletions or duplications of this specific locus have not been reported in studies of autism. Duplications of a much larger region from 15q11.2 to 15q13 (and sometimes including these four genes) are a well-documented risk factor for autism, with seven such duplications reported in the autism studies considered here (16, 25, 26, 28). These data and work implicating overexpression of *CYFIP1* in autism due to duplications of 15q11.2-q13 and Fragile X syndrome (104) suggest that *CYFIP1*, which interacts with the protein product (FMRP) of the *FMR1* gene, may mediate associations of this locus to autism and schizophrenia. However, some cases of autism associated with duplications of 15q11.2-q13 do not include the *CYFIP1* gene, which is inconsistent with any simple causative role for this gene.

Genes affected by CNVs that appear to be associated with both autism and schizophrenia, but have not been reported in a sufficient number of cases for inclusion here, include *APBA2* [2 duplications in schizophrenia (8, 21), one duplication in autism (26)], *CDH8* [4 deletions in autism (14), 1 duplication in schizophrenia (18)], and *ZNF804A* [3 duplications in autism (14), 1 deletion in schizophrenia (18)]. Deletions involving *PARK2* have been reported in association with autism (30), and two deletions and three duplications involving this gene have been reported in cases of schizophrenia (18), but the CNVs exhibit complex patterns of partial overlap that are difficult to interpret without further information.

Genetic Association. Genes in Table 2 were ascertained from the Szgene and AutDB databases and PubMed searches using the terms "autism or schizophrenia" and "gene or genetic". Genes in AutDB in the "functional" category, and in the "syndromic" category, where associations with autism per se have not been reported (e.g., ALDH5A1), were not included. Positive results for schizophrenia also include genes that are positively-associated from meta-analyses in Szgene. PubMed searches are to date as of June 15, 2009. To minimize the possible effects of false-positive results, only genes that exhibited replicated positive associations in one or both conditions or positive associations in meta-analysis for schizophrenia were included in the sets of genes showing positive associations in one or both conditions.

Under a null hypothesis, associations with schizophrenia and autism are distributed among genes independently. Let *s* be the number of genes associated with schizophrenia, *a* be the number of genes associated with autism, and *n* be the total number of genes. The total number of possible combinations of genes with the specified number of associations is given by:

$$ {}_n\mathrm{C}_s \cdot {}_n\mathrm{C}_a $$

where

$$ {}_n\mathrm{C}_x = \frac{n!}{x!(n-x)!} $$

Let *w* be the number of genes associated with both schizophrenia and autism, *x* the number associated only with autism, *y* the number associated only with schizophrenia, and *z* the number associated with neither condition. The total number of possible combinations of genes with the specified sets of associations is given by:

$$ {}_{n}C_{w} \cdot {}_{(n-w)}C_{x} \cdot {}_{(n-w-x)}C_{y} \cdot {}_{(n-w-x-y)}C_{z} $$

Because the latter term is equal to one, the probability of obtaining some configuration of associations under the null hypothesis is given by:

$$ \frac{{}_{n}C_{w} \cdot {}_{(n-w)}C_{x} \cdot {}_{(n-w-x)}C_{y}}{{}_{n}C_{s} \cdot {}_{n}C_{a}} $$

The possible combinations of associations and their probabilities are provided in Table S10.

ACKNOWLEDGMENTS. We thank C. Badcock, J. Friedman, D. Geschwind, R. Glessner, H. Hakonarson, R. Holt, C. Marshall, S. Scherer, J. Sebat, D. St. Clair, H. Stefansson, J. Stone, R. Ullmann, and members of the Simon Fraser University Fab-lab for helpful comments, discussion, or provision of information to us. We thank R. Nesse and S. Stearns for inviting B.C. to the Sackler Colloquium, and NSERC for financial support.

1. Bleuler E (1911) *Dementia praecox or the group of schizophrenias*, trans Zinkin j (1950) (International Universities Press, New York) (German).
2. Kanner L (1943) Autistic disturbances of affective contact. *Nervous Child* 2:217–250.
3. Petty LK, Ornitz EM, Michelman JD, Zimmerman EG (1984) Autistic children who become schizophrenic. *Arch Gen Psychiatry* 41:129–135.
4. Rutter M (1972) Childhood schizophrenia reconsidered. *Journal of Autism and Childhood Schizophrenia* 2:315–337.
5. Crespi B, Badcock C (2008) Psychosis and autism as diametrical disorders of the social brain. *Behav Brain Sci* 31:241–261, and discussion (2008) 31:261–320.
6. Burbach JP, van der Zwaag B (2009) Contact in the genetics of autism and schizophrenia. *Trends Neurosci* 32:69–72.
7. Iossifov I, Zheng T, Baron M, Gilliam TC, Rzhetsky A (2008) Genetic-linkage mapping of complex hereditary disorders to a whole-genome molecular-interaction network. *Genome Res* 18:1150–1162.
8. Kirov G, et al. (2008) Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum Mol Genet* 17:458–465.
9. Frith CD, Frith U (1991) Elective affinities in schizophrenia and childhood autism. *Social Psychiatry: Theory, Methodology and Practice*, ed Bebbington P (Transactions Press, New Brunswick, NJ), pp 65–88.
10. Tordjman S (2008) Reunifying autism disorder and early-onset schizophrenia in terms of social communication disorders. *Behav Brain Sci* 31:278–281.
11. Craddock N, Owen MJ (2005) The beginning of the end for the Kraepelinian dichotomy. *Br J Psychiatry* 186:364–366.
12. Craddock N, O'Donovan MC, Owen MJ (2009) Psychosis genetics: Modeling the relationship between schizophrenia, bipolar disorder, and mixed (or "schizoaffective") psychoses. *Schizophr Bull* 35:482–490.
13. Moskvina V, et al. (2009) Gene-wide analyses of genome-wide association data sets: Evidence for multiple common risk alleles for schizophrenia and bipolar disorder and for overlap in genetic risk. *Mol Psychiatr* 14:252–260.
14. Autism Genome Project Consortium, et al. (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 39:319–328.
15. Mefford HC, et al. (2008) Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med* 359:1685–1699.
16. Weiss LA, et al. (2008) Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 358:667–675.
17. Brunetti-Pierri N, et al. (2008) Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet* 40:1466–1471.
18. International Schizophrenia Consortium (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455:237–241.
19. Stefansson H, et al. (2008) Large recurrent microdeletions associated with schizophrenia. *Nature* 455:232–236.
20. Walsh T, et al. (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 320:539–543.
21. Need AC, et al. (2009) A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet* 5:e1000373.
22. Kirov G, et al. (2009) Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum Mol Genet* 18:1497–1503.
23. Ben-Shachar S, et al. (2009) Microdeletion 15q13.3: A locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders. *J Med Genet* 46:382–388.
24. Miller DT, et al. (2009) Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. *J Med Genet* 46:242–248.
25. Sebat J, et al. (2007) Strong association of de novo copy number mutations with autism. *Science* 316:445–449.
26. Christian SL, et al. (2008) Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. *Biol Psychiatry* 63:1111–1117.
27. Kumar RA, et al. (2008) Recurrent 16p11.2 microdeletions in autism. *Hum Mol Genet* 17:628–638.
28. Marshall CR, et al. (2008) Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* 82:477–488.
29. Bijlsma EK et al. (2009) Extending the phenotype of recurrent rearrangements of 16p11.2: Deletions in mentally retarded patients without autism and in normal individuals. *European Journal of Medical Genetics* 52:77–87.
30. Glessner JT, et al. (2009) Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 459:569–573.
31. McCarthy S, et al. (2009) Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet*, in press.
32. Ullmann R, et al. (2007) Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Hum Mutat* 28:674–682.
33. Ingason A, et al. (2009) Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Mol Psychiatry*, in press.
34. Xu B, et al. (2008) Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* 4:880–885.
35. Mayes R, Horwitz AV (2003) DSM-III and the revolution in the classification of mental illness. *Journal of the History of the Behavioral Sciences* 41:249–267.
36. Kendler KS (April 16, 2009) An historical framework for psychiatric nosology. *Psychol Med*, 10.1017/S0033291709005753.
37. Cannon TD (2009) What is the role of theories in the study of schizophrenia? *Schizophr Bull* 35:563–567.
38. Rutter M (1968) Concepts of autism: A review of research. *J Child Psychol Psyc* 9:1–25.
39. Craddock N, O'Donovan MC, Owen MJ (2005) The genetics of schizophrenia and bipolar disorder: Dissecting psychosis. *J Med Genet* 42:193–204.
40. Crespi B, Summers K, Dorus S (2009) Genomic sister-disorders of neurodevelopment: An evolutionary approach. *Evolutionary Applications* 2:81–100.
41. Belmonte MK, Bourgeron T (2006) Fragile X syndrome and autism at the intersection of genetic and neural networks. *Nat Neurosci* 9:1221–1225.
42. Kwon CH, et al. (2006) Pten regulates neuronal arborization and social interaction in mice. *Neuron* 50:377–388.
43. Hoeffer CA, et al. (2008) Removal of FKBP12 enhances mTOR-Raptor interactions, LTP, memory, and perseverative/repetitive behavior. *Neuron* 60:832–845.
44. Kelleher RJ, 3rd, Bear MF (2008) The autistic neuron: Troubled translation? *Cell* 135:401–437.
45. Cuscó I, et al. (2009) Autism-specific copy number variants further implicate the phosphatidylinositol signaling pathway and the glutamatergic synapse in the etiology of the disorder. *Hum Mol Genet* 18:1795–1804.
46. Emamian ES, Hall D, Birnbaum MJ, Karayiorgou M, Gogos JA (2004) Convergent evidence for impaired AKT1-GSK3beta signaling in schizophrenia. *Nat Genet* 36:131–137.
47. Stopkova P, et al. (2004) Identification of PIK3C3 promoter variant associated with bipolar disorder and schizophrenia. *Biol Psychiatry* 55:981–988.
48. Kalkman HO (2006) The role of the phosphatidylinositide 3-kinase-protein kinase B pathway in schizophrenia. *Pharmacol Ther* 110:117–134.
49. Krivosheya D, et al. (2008) ErbB4-neuregulin signaling modulates synapse development and dendritic arborization through distinct mechanisms. *J Biol Chem* 283:32944–32956.
50. Hardan AY, Muddasani S, Vemulapalli M, Keshavan MS, Minshew NJ (2006) An MRI study of increased cortical thickness in autism. *Am J Psychiatry* 163:1290–1292.
51. Bethea TC, Sikich L (2007) Early pharmacological treatment of autism: A rationale for developmental treatment. *Biol Psychiatry* 61:521–537.
52. Elder LM, Dawson G, Toth K, Fein D, Munson J (2008) Head circumference as an early predictor of autism symptoms in younger siblings of children with autism spectrum disorder. *J Autism Dev Disord* 38:1104–1111.
53. Stanfield AC, et al. (2008) Towards a neuroanatomy of autism: A systematic review and meta-analysis of structural magnetic resonance imaging studies. *Eur Psychiat* 23:289–299.
54. Chiu S, et al. (2007) Early acceleration of head circumference in children with fragile x syndrome and autism. *J Dev Behav Pediatr* 28:31–35.
55. Szudek J, Evans DG, Friedman JM (2003) Patterns of associations of clinical features in neurofibromatosis 1 (NF1). *Hum Genet* 112:289–297.
56. Butler MG, et al. (2005) Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. *J Med Genet* 42:318–321.
57. Douglas J, et al. (2007) Mutations in RNF135, a gene within the NF1 microdeletion region, cause phenotypic abnormalities including overgrowth. *Nat Genet* 39:963–965.
58. Goghari VM, Rehm K, Carter CS, MacDonald AW, 3rd (2007) Regionally specific cortical thinning and gray matter abnormalities in the healthy relatives of schizophrenia patients. *Cereb Cortex* 17:415–424.
59. Gur RE, Keshavan MS, Lawrie SM (2007) Deconstructing psychosis with human brain imaging. *Schizophr Bull* 33:921–931.
60. Bose SK, et al. (2009) The effect of ageing on grey and white matter reductions in schizophrenia. *Schizophr Res* 112:7–13.
61. Goldman AL, et al. (2009) Widespread reductions of cortical thickness in schizophrenia and spectrum disorders and evidence of heritability. *Arch Gen Psychiatry* 66:467–477.
62. Haukvik UK, et al. (2009) Cerebral cortical thickness and a history of obstetric complications in schizophrenia. *J Psychiat Res* 526.
63. Eliez S, Schmitt JE, White CD, Reiss AL (2000) Children and adolescents with velocardiofacial syndrome: A volumetric MRI study. *Am J Psychiatry* 157:409–415.
64. Chow EW, Zipursky RB, Mikulis DJ, Bassett AS (2002) Structural brain abnormalities in patients with schizophrenia and 22q11 deletion syndrome. *Biol Psychiatry* 51:208–215.
65. Hannes FD, et al. (2008) Recurrent reciprocal deletions and duplications of 16p13.11: The deletion is a risk factor for MR/MCA while the duplication may be a rare benign variant. *J Med Genet* 46:223–232.
66. Rujescu D, et al. (2009) Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet* 18:988–996.

67. Persico AM, Levitt P, Pimenta AF (2006) Polymorphic GGC repeat differentially regulates human reelin gene expression levels. *J Neural Transm* 113:1373–1382.
68. Akbarian S, Huang HS (2006) Molecular and cellular mechanisms of altered GAD1/GAD67 expression in schizophrenia and related disorders. *Brain Res Rev* 52:293–304.
69. Harada S, Tachikawa H, Kawanishi Y (2001) Glutathione S-transferase M1 gene deletion may be associated with susceptibility to certain forms of schizophrenia. *Biochem Biophys Res Commun* 281:267–271.
70. Buyske S, et al. (2006) Analysis of case-parent trios at a locus with a deletion allele: Association of GSTM1 with autism. *BMC Genetics* 7:8.
71. Zogel C, et al. (2006) Identification of cis- and trans-acting factors possibly modifying the risk of epimutations on chromosome 15. *Eur J Hum Genet* 14:752–758.
72. Mouridsen SE, Rich B, Isager T (2008) Psychiatric disorders in adults diagnosed as children with atypical autism. A case control study. *J Neural Transm* 115:135–138.
73. Rzhetsky A, Wajngurt D, Park N, Zheng T (2007) Probing genetic overlap among complex human phenotypes. *Proc Natl Acad Sci USA* 104:11694–11699.
74. Sporn AL, et al. (2004) Pervasive developmental disorder and childhood-onset schizophrenia: Comorbid disorder or a phenotypic variant of a very early onset illness? *Biol Psychiatry* 55:989–994.
75. Feinstein C, Singh S (2007) Social phenotypes in neurogenetic syndromes. *Child Adol Psych Cl* 16:631–647.
76. Reaven JA, Hepburn SL, Ross RG (2008) Use of the ADOS and ADI-R in children with psychosis: Importance of clinical judgment. *Clinical Child Psychiatry and Psychology* 13: 81–94.
77. Starling J, Dossetor D (2009) Pervasive developmental disorders and psychosis. *Current Psychiatry Reports* 11:190–196.
78. Konneker T, et al. (2008) A searchable database of genetic evidence for psychiatric disorders. *Am J Med Genet Part B* 147B:671–675.
79. Hill EL, Frith U (2003) Understanding autism: Insights from mind and brain. *Philos Trans R Soc London Ser B* 358:281–289.
80. Williams JH (2008) Self-other relations in social development and autism: Multiple roles for mirror neurons and other brain bases. *Autism Research* 1:73–90.
81. Baron-Cohen S (2009) Autism: The empathizing-systemizing (E-S) theory. *Ann NY Acad Sci* 1156:68–80.
82. Rubenstein JL, Merzenich MM (2003) Model of autism: Increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav* 2:255–267.
83. Fatemi SH (2007) The hyperglutamatergic hypothesis of autism. *Prog Neuropsychopharmacol Biol Psychiatry*, 32:911 author reply (2007) 32:912–913.
84. Markram H, Rinaldi T, Markram K (2007) The intense world syndrome - an alternative hypothesis for autism. *Front Neurosci* 1:77–96.
85. Happé F, Frith U (2006) The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *J Autism Dev Disord* 36:5–25.
86. Lewis JD, Elman JL (2008) Growth-related neural reorganization and the autism phenotype: A test of the hypothesis that altered brain growth leads to altered connectivity. *Dev Sci* 11:135–155.
87. Sobin C, et al. (2001) Early, non-psychotic deviant behavior in schizophrenia: A possible endophenotypic marker for genetic studies. *Psychiatry Res* 101:101–113.
88. Tandon R, Nasrallah HA, Keshavan MS (2009) Schizophrenia, "just the facts" 4. Clinical features and conceptualization. *Schizophr Res* 110:1–23.
89. Remschmidt HE, Schulz E, Martin M, Warnke A, Trott GE (1994) Childhood-onset schizophrenia: History of the concept and recent studies. *Schizophr Bull* 20:727–745.
90. Rapoport J, Chavez A, Greenstein D, Addington A, Gogtay N (2009) Autism spectrum disorders and childhood-onset schizophrenia: Clinical and biological contributions to a relation revisited. *J Am Acad Child Adolesc Psychiatry* 48:10–18.
91. Eliez S (2007) Autism in children with 22q11.2 deletion syndrome. *J Am Acad Child Adolesc Psychiatry*, 46:433–434 author reply (2007) 46:434.
92. Solomon M, Ozonoff S, Carter C, Caplan R (2008) Formal thought disorder and the autism spectrum: Relationship with symptoms, executive control, and anxiety. *J Autism Dev Disord* 38:1474–1484.
93. Sugihara G, Tsuchiya KJ, Takei N (2008) Distinguishing broad autism phenotype from schizophrenia-spectrum disorders. *J Autism Dev Disord*, 38:1998–1999 author reply (2008) 38:2000–2001.
94. Antshel KM, et al. (2007) Autistic spectrum disorders in velo-cardio facial syndrome (22q11.2 deletion). *J Autism Dev Disord* 37:1776–1786.
95. Niklasson L, Rasmussen P, Oskarsdóttir S, Gillberg C (2008) Autism, ADHD, mental retardation and behavior problems in 100 individuals with 22q11 deletion syndrome. *Res Dev Disabil* 30:763–773.
96. Doornbos M, et al. (2009) Nine patients with a microdeletion 15q11.2 between breakpoints 1 and 2 of the Prader-Willi critical region, possibly associated with behavioural disturbances. *Eur J Med Genet* 52:108–115.
97. Jolin EM, Weller RA, Weller EB (2009) Psychosis in children with velocardiofacial syndrome (22q11.2 deletion syndrome). *Curr Psychiatry Rep* 11:99–105.
98. Sprong M, et al. (2008) Pathways to psychosis: A comparison of the pervasive developmental disorder subtype Multiple Complex Developmental Disorder and the "At Risk Mental State". *Schizophr Res* 99:38–44.
99. Spence SJ, Schneider MT (2009) The role of epilepsy and epileptiform EEGs in autism spectrum disorders. *Pediatr Res* 65:599–606.
100. Coyle JT (2004) The GABA-glutamate connection in schizophrenia: Which is the proximate cause? *Biochem Pharmacol* 68:1507–1514.
101. Kehrer C, Maziashvili N, Dugladze T, Gloveli T (2008) Altered excitatory-inhibitory balance in the NMDA-hypofunction model of schizophrenia. *Front Mol Neurosci* 1:6.
102. Dölen G, Bear MF (2008) Role for metabotropic glutamate receptor 5 (mGluR5) in the pathogenesis of fragile X syndrome. *J Physiol* 586:1503–1508.
103. Conn PJ, Lindsley CW, Jones CK (2009) Activation of metabotropic glutamate receptors as a novel approach for the treatment of schizophrenia. *Trends Pharmacol Sci* 30:25–31.
104. Nishimura Y, et al. (2007) Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways. *Hum Mol Genet* 16:1682–1698.

The comparative genomics of viral emergence

Edward C. Holmes[a,b,1]

[a]Center for Infectious Disease Dynamics, Department of Biology, Pennsylvania State University, University Park, PA 16802; and [b]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892

Edited by Stephen Curtis Stearns, Yale University, New Haven, CT, and accepted by the Editorial Board September 28, 2009 (received for review July 8, 2009)

RNA viruses are the main agents of emerging and re-emerging diseases. It is therefore important to reveal the evolutionary processes that underpin their ability to jump species boundaries and establish themselves in new hosts. Here, I discuss how comparative genomics can contribute to this endeavor. Arguably the most important evolutionary process in RNA virus evolution, abundant mutation, may even open up avenues for their control through "lethal mutagenesis." Despite this remarkable mutational power, adaptation to diverse host species remains a major adaptive challenge, such that the most common outcome of host jumps are short-term "spillover" infections. A powerful case study of the utility of genomic approaches to studies of viral evolution and emergence is provided by influenza virus and brought into sharp focus by the ongoing epidemic of swine-origin H1N1 influenza A virus (A/H1N1pdm). Research here reveals a marked lack of surveillance of influenza viruses in pigs, coupled with the possibility of cryptic transmission before the first reported human cases, such that the exact genesis of A/H1N1pdm (where, when, how) is uncertain.

evolution | influenza | RNA virus | lethal mutagenesis | mutation rate

The recent appearance of swine-origin H1N1 influenza A virus (A/H1N1pdm) in humans serves as a pointed reminder of the global burden of morbidity and mortality caused by influenza viruses. More generally, A/H1N1pdm highlights the ability of RNA viruses to jump species barriers and emerge in new hosts, in this case transferring from pigs to humans. Although the H1N1 subtype of influenza A virus is a familiar one, most famously as the cause of the devastating influenza pandemic of 1918–1919 (1), the current A/H1N1pdm epidemic is notable in that it is caused by a viral lineage that is phylogenetically distinct from the other swine influenza viruses sampled over the last 20 years (2, 3). Hence, despite the considerable effort that has gone into the surveillance and characterization of influenza viruses, and especially in wild birds since the appearance of highly pathogenic avian A/H5N1 viruses (4–6), there has been a marked gap in our surveillance of these viruses in pigs. This represents a serious oversight as swine viruses are already adapted for transmission in mammalian populations and spillover infections from pigs to humans are relatively commonplace (7).

Another important aspect of the A/H1N1pdm epidemic is how quickly genome sequence data for this virus was generated and placed into the public domain. Indeed, so rapid was the generation of sequence data that analysis was effectively undertaken in real time and laudably often with full public access (see http://tree.bio.ed.ac.uk/groups/influenza for an excellent example). More generally, the rapid generation of genome sequence data represents a very powerful way to determine cause of diseases of unknown etiology. This is illustrated by the case of colony collapse disorder (CCD), in which honey bees leave the hive and seemingly die, such that the hive community eventually collapses. Here, comparative genome sequencing was quickly able to identify the RNA virus Israeli acute paralysis virus (IAPV) as the most likely agent for CCD (8), although this is yet to satisfy Koch's postulates. As an added bonus, because all of the DNA present in infected hives was sequenced, rather than just that of IAPV, this metagenomics approach to pathogen discovery allowed characterization of much of the microbial flora carried by honey bees, encompassing viruses, bacteria, fungi, and others. In short, the generation and analysis of complete genome sequence data are close to becoming the default way of characterizing new viral pathogens (9).

The aim of this article is to demonstrate how new genomic-scale approaches are able to provide unique insights into the processes that govern the emergence and evolution of RNA viruses. In doing so I make general statements about the nature of RNA virus evolution and highlight some of the key evolutionary lessons learned from the ongoing A/H1N1pdm pandemic in particular. As a sidebar, this work illustrates the increasingly important role played by evolutionary biology in the study of infectious disease.

The Evolutionary Genetics of Viral Emergence

Even allowing for their relative abundance, RNA viruses seem particularly prone to causing emerging diseases in humans and other animals (10). Although these infectious agents have defining characteristics, perhaps the most important from the perspective of their evolution is their capacity for mutation. The vast majority of estimates of mutation rates in RNA viruses are in the range of 0.1 to 1.0 mutations per genome, per replication (11), several orders of magnitude higher than those in most DNA-based organisms (Fig. 1). Such remarkable error rates are evidently a function of replication with a low-fidelity RNA-dependent RNA polymerase, the only protein shared by all RNA viruses and which lacks any of the proof-reading abilities associated with the higher-fidelity DNA polymerases. Although there is an ongoing debate as to what selective forces (if any) are responsible for such high error rates (11, 12), it is likely that RNA viruses survive this enormous mutational burden by an equally remarkable reproductive power, manifest as many progeny in each infected cell, infected cells, and infected hosts (13).

High mutation rates, coupled with rapid replication, are also the basis for the high rates of nucleotide substitution (fixation) recorded in RNA viruses. Mean substitution rates in this case are usually in the realm of 10^{-3} to 10^{-4} nucleotide substitutions per site per year (subs/site/year) and hence some six orders of magnitude higher than those seen in eukaryotes (11, 14). Clearly, evolutionary rates of this magnitude are a major reason clinically important traits, such as drug resistance, escape from vaccine coverage, and host range expansion, appear so readily in some RNA viruses. In the case of A/H1 N1pdm, estimates of genome-wide substitution rates, at $\approx 5 \times 10^{-3}$ subs/site/year (15), are broadly similar to those seen in other human influenza viruses (16), and hence at the upper end of those seen in RNA viruses generally (14). Whether this means that A/H1N1pdm will gen-

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Evolution in Health and Medicine" held April 2–3, 2009, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS web site at www.nasonline.org/Sackler_Evolution_Health_Medicine.

Author contributions: E.C.H. designed research, performed research, analyzed data, and wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission. S.C.S. is a guest editor invited by the Editorial Board.

[1]To whom correspondence should be addressed. E-mail: ech15@psu.edu.

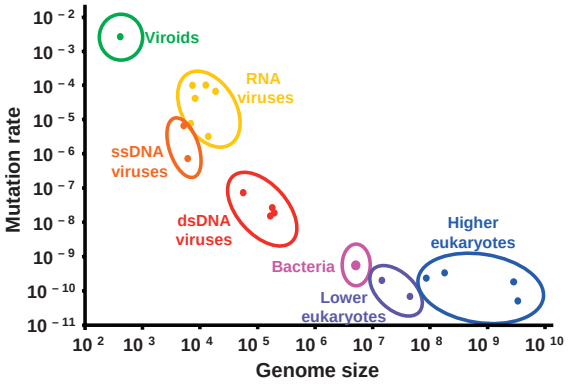


Fig. 1. Relationship between mutation rate per nucleotide site and genome size for different genomic systems including viruses. [Reproduced with permission from ref. 19 (Copyright 2009, AAAS).]

erate similar levels of antigenic variation to that seen in other human influenza A viruses will require a longer sampling period.

The high evolutionary rates associated with RNA viruses have several important implications. First, it is likely that high mutation rates limit the genome size of RNA viruses to a median value of ≈10 Kb (and the largest RNA viruses are the coronaviruses at only 29–32 Kb). Specifically, if mutation rates are constant per nucleotide, then the error-prone replication of longer RNA molecules will lead to the accumulation of excessive numbers of deleterious mutations and hence major fitness losses (17). Although a number of theories have been proposed to explain the constrained genome sizes of RNA viruses, such as an inability to package excessively large genomes, the power of the mutation rate hypothesis is that it can be extended to other microbial organisms. In particular, ssDNA viruses, such as the vertebrate parvoviruses, exhibit mutation rates that fall closer to those seen in RNA viruses than dsDNA viruses, and all have genomes that are <12 Kb in length (Fig. 1). Similarly, rates of nucleotide substitution in ssDNA viruses fall within the range seen in RNA viruses (11). Although the reasons that underpin the high evolutionary rates of ssDNA viruses are unclear, one possibility is that mutations in ssDNA viruses are largely the result of postreplicatory processes, such as deamination (18). At the other extreme, recent work has revealed that viroids, small (< 500 nt), highly structured RNA elements that cause a variety of diseases in plants, experience mutation rates that exceed even those seen in RNA viruses (ref. 19 and Fig. 1).

Another important implication of the high mutation rates of RNA viruses, and one that at first glance seems rather paradoxical, is that they may represent an Achilles' heel for their treatment. The background to this idea is that RNA viruses sit close to what can be notionally thought of as an "error threshold," beyond which major fitness losses are likely (17). Although the complexity of fitness landscapes means that there is unlikely to be an absolute threshold value *per se* (20), the general association between viral mutation rates and genome sizes strongly suggests that overly large genomes are subject to severe fitness costs. Given this upper bound on mutation rates (and genome sizes), it then follows that artificially increasing error rates through exposure to mutagens such as ribavirin or 5′-fluorouracil would also lead to major fitness losses (21). Although there is some debate over the precise mechanistic basis to this "lethal mutagenesis" (22), the available experimental data offer strong support to the applicability of this exciting form of antiviral therapy, particularly when the mutagens are combined with more standard replication inhibitors (23). In addition, if there is indeed a fundamental relationship between genome size and error rate then the methods of lethal mutagenesis should equally apply to ssDNA viruses and viroids. The complication, of course, is the evolution of resistance. Although there have been claims that lethal mutagenesis will not be subject to all of the mechanisms of resistance that plague other forms of microbial control (24), in reality there is nothing in the biology of lethal mutagenesis that suggests it is "evolution proof," particularly as single point mutations can result in improved polymerase fidelity (25).

Although RNA viruses have an enhanced capacity for mutation, which must in part underpin their ability to jump species boundaries and successfully emerge in new hosts, rapid mutation is not the only evolutionary process that needs to be discussed when considering their emergence. Of obvious importance is the fitness distribution of new mutations (26). As with most systems, the majority of mutations that arise in RNA viruses are deleterious and simply act to reduce fitness, particularly in the absence of high rates of recombination and heterozygosity (13). Measures of mutational fitness in single-cell assays suggest that as many as 40% of all mutations in vesicular stomatitis virus may be lethal, with many of the remainder falling into the slightly deleterious class (27). When this effect is magnified across all of the cell types a virus may infect, and considering the complexities of the viral life cycle during which natural selection can act at a variety of times and levels (28), it is evident that the vast majority of mutations will reduce fitness. As a case in point, only ≈1% of poliovirus virions released from a cell are able to complete a full replication cycle (28). Hence, most mutations that will ultimately aid adaptation to a new host species are likely to be strongly deleterious in the donor host (29).

As deleterious mutations are expected to be young (30), the preponderance of deleterious and slightly deleterious mutations is also manifest in phylogenetic analyses as an excess of nonsynonymous mutations on the tips of trees (31). It is this phenomenon that explains the increasingly common observation that rates of nucleotide substitution are higher in the short term, such as among sequences sampled from within single hosts or from individuals directly connected by transmission, than in the long term, such as between epidemics (32, 33). This may also be true of A/H1N1pdm; rate estimates based on the analyses of A/H1N1pdm sequences alone, collected over very short time scales (days to weeks), are greater than those estimated over longer time scales when A/H1N1pdm sequences are combined with other (swine) H1N1 lineages separated by years of evolution (3).

RNA viruses therefore seem to occupy a region of evolutionary parameter space that is acutely different to that where higher eukaryotes reside (13). That genome sizes are constrained by a high mutation rate means that RNA viruses may be less adaptable then is often envisaged, because there will be strong selection against evolutionary processes that act to increase genome size, such as gene duplication and lateral gene transfer (13). Such constraints are also reflected in the fact that the majority of cross-species transmissions of RNA viruses result in transient spillover infections rather than fully endemic pathogens (34). Hence, even though there is frequent exposure, the majority of RNA viruses are unable to fully adapt to new host species (35). A/H1N1pdm is again an important exemplar: between 2005 and 2009 11 human patients in the United States experienced swine influenza infections, yet only A/H1N1pdm has resulted in epidemic spread (7). Understanding the factors that determine whether a new infection will simply spill-over or spread on a larger scale is critical to predicting the future of any new emerging disease.

There are two important generalities about the nature of viral emergence that shed some light on the evolutionary mechanisms at play. First, vector-borne RNA viruses are subject to stronger

selective constraints than those viruses transmitted by other routes (36) and correspondingly are less able to establish productive infections in new host species (10). This effect is likely because vector-borne viruses are subject to strong antagonistic pleiotropy, such that mutations favored in one host type are injurious in another, and which greatly limits adaptability after host jumps (37, 38), although different levels of diversity may be generated in the vertebrate and invertebrate components of the transmission cycle (39, 40). Second, a simple (but often broken) rule of thumb is that the more closely donor and recipient host species are related, the easier it will be for any virus (and likely any other pathogen) to jump between them and establish a productive infection (13, 41). Such a tendency arises because the cell types in these host species, their receptors, and likely other key components of the virus–host interaction will diverge along with their hosts to eventually reach a point where they become unrecognizably different for any RNA virus. For example, dengue viruses from nonhuman primates seem to able to replicate in human cells without any additional mutations (42), and only a single mutation appears responsible for the successful transfer of Venezuelan equine encephalitis virus from rodents to horses (43). In contrast, 13 mutations may be required for avian influenza viruses to establish productive infections in humans (44). This point also emphasizes how swine influenza viruses are in some sense "preadapted" to replicate in humans because they already contain the suite of key mutations required for productive replication in mammals. Hence, a simple rule of emergence is that viruses that have achieved this feat once have an inherent capacity to do it again.

The Evolutionary Genomics of Influenza Virus

One virus where genome sequence data has already had a profound impact on evolutionary studies is influenza. The key event in the genomics revolution for influenza A virus was the instigation of the Influenza Genome Sequencing Project (IGSP) in 2005 (ref. 45; see www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html). Although influenza viruses have long been the under the gaze of evolutionary biologists, there were surprisingly few complete genome sequences available for analysis before the start of the IGSP. Today, however, >4,000 complete genome sequences of influenza viruses have been generated from a diverse array of avian and mammalian hosts.

One of the most important observations stemming from the data generated under the IGSP is that intrasubtype (i.e., within the A/H3N2 or A/H1N1 human subtypes) reassortment occurs very frequently. As a corollary, this also means that the mixed infection of individual hosts with multiple viral strains is also commonplace, which in turn raises questions about the extent of cross-protective immunity (46). Although the importance of reassortment for the cross-species transmission of influenza virus has a long history (47), complete genome sequence data provide the only clear insight into the frequency and determinants of this process (48). In addition, complete genome sequence analysis reveals how by placing gene segments in new genomic configurations, reassortment can sometimes generate isolates with altered antigenic properties, which in turn may lead to vaccine failure (49, 50). In short, reassortment seems to be a more important process in the day-to-day evolution of influenza A virus than previously realized, and attempts to predict future antigenic evolution without a consideration of reassortment are unlikely to be successful.

A second key insight stemming from the IGSP data is that specific human populations, such as that of New York State, where sampling has been particularly widespread in time and space, are characterized by the circulation of multiple viral lineages during any single season (16, 48–51). Not only does the cocirculation of lineages ensure that there is abundant raw material for reassortment, but it means that viral lineages must be continually imported during the time course of the influenza season (16, 49, 50). It is therefore not the case that a single viral lineage enters a population at the start of the influenza season (winter in the Northern Hemisphere), gradually diffuses through the population over the subsequent 6 months, before dying out the next summer. The same can be expected of A/H1N1pdm in the years that follow. The genetic diversity within a single population is also extremely well mixed spatially. For example, across the United States as a whole, phylogeographic analysis reveals that even relatively geographically isolated communities harbor similar amounts of viral diversity as major cities with more expansive travel networks (51), highlighting how rapidly this virus is able to spread through populations. Influenza virus is clearly readily able to exploit human contact networks, so that the coinfection that fuels reassortment could occur in any number of locations.

The expanse of genomic information on influenza also sheds new light on the genesis of drug-resistant viruses. One of the most important, and unexpected, outcomes of these studies is that direct drug-selection pressure is not always responsible for drug resistance. This is clearly the case with the adamantanes (amantadine and rimantadine), a group of antivirals to which subtype A/H3N2 viruses have shown a global rise in resistance in recent years (52). The most common cause of adamantane resistance is a single amino acid change (Ser31Asn) in the M2 protein. What is most striking in this case is that the Ser31Asn mutation has increased abruptly in frequency in populations where adamantanes are rarely used, such as the United States. Therefore, rather than being caused by direct selection pressure, it is more likely that the Ser31Asn mutation has become fortuitously linked to an antibody escape mutation located on another genome segment (53, 54). That this might be a more general phenomenon is suggested by the fact that the same hitch-hiking process may now be taking place with the neuraminidase inhibitor oseltamivir. In this case, there has been a dramatic rise in oseltamivir resistance in A/H1N1 viruses in many locations, including the United States where nearly all viruses are oseltamivir resistant (55, 56). Because these drugs are not widely used in many populations, linkage to another beneficial mutation again seems the most probable explanation for the rise of oseltamivir resistance, although there is no evidence for reassortment in this case. Although swine-origin A/H1N1 viruses are currently generally sensitive to oseltamivir (but resistant to adamantanes) it is possible that future reassortment among cocirculating human and swine-origin H1N1 viruses will change this picture.

Despite the genomic revolution, aspects of the evolution and epidemiology of influenza A virus remain opaque. In particular, although there is some evolutionary evidence for interaction between the A/H1N1 and A/H3N2 influenza viruses, which experience distinctive out-of-phase dynamics (15), and between influenza viruses A and B (57), the exact cause of these interactions remains elusive. For example, does this competition involve some form of nonspecific cross-immunity or ecological interference? The appearance of A/H1N1pdm adds a new urgency to revealing these interactions: it is obviously important to determine whether this newly emerged virus will outcompete the other influenza viruses circulating in human populations, particularly as estimates of its basic reproductive number (R_0) indicate that it has the capacity to spread widely (58). Similarly, it is unclear why the "seasonal" A/H1N1 and A/H3N2 viruses that currently cocirculate in human populations have such different epidemiological dynamics. Seasonal A/H1N1 viruses are characterized by relatively slow antigenic evolution (such that relatively few mutations accumulate at antigenic sites), yet greater circulating genetic diversity, whereas A/H3N2 viruses are characterized by lower levels of circulating genetic diversity, but more rapid antigenic change (50), manifest as the fact that the

A/H3N2 vaccine component needs to be updated on a regular basis (59). Finally, what epidemiological and evolutionary processes determine the phylodynamics of the influenza HA protein, manifest as "ladder-like" phylogenies and regular changes in antigenic type, is still the source of considerable debate (60–62).

As noted above, a key goal for the future must be to track the potential reassortment of the cocirculating influenza viruses, including the sporadic cases of highly pathogenic A/H5N1 avian influenza virus that have appeared in humans since 2003. Particular attention should be paid to viral surveillance in East and Southeast Asia, which seems to act as the global source population for influenza viruses (63), and may eventually prove to be the case for A/H1N1pdm. Indeed, the recent emergence of A/H1N1pdm means that three distinct lineages of influenza A virus are currently circulating in human populations, an event that is unprecedented in modern human history. Similarly, it will also be of fundamental importance to determine the spatial dynamics of A/H1N1pdm in human populations, and particularly whether they follow the same general pathways as identified for other influenza A viruses (63).

Those evolutionary analyses of A/H1N1pdm undertaken to date have shed light on a number of key issues. Phylogenetic analyses of complete genome sequence data have revealed the series of reassortment events responsible for the origin of A/H1N1pdm and how this virus has spread rapidly in both time and space (3). For example, since its first appearance A/H1N1pdm has spread to >160 countries (ref. 64; www.who.int/csr/don/2009_09_11/en/index.html), including multiple introductions into both Asia and Europe from the Americas (Fig. 2 and ref. 15). As this spatial diffusion continues it will also be essential to track the antigenic evolution of A/H1N1pdm and determine whether it bears more resemblance to the slow antigenic drift of A/H1N1 or the speedier evolution of A/H3N2. More generally, it will be important to understand why the A/H1N1pdm lineage was able to successfully emerge in human populations, and the determinants of this process, when most of the other lineages of swine influenza virus that periodically spill over into human populations fail to become established as endemic pathogens. In addition, although A/H1N1pdm was first reported in Mexico, whether the reassortment events that generated this virus occurred in that country (or continent) is less clear.

Finally, molecular clock estimates of the time of origin of H1N1pdm date it to a period spanning the end of 2008 through the first 2 months of 2009 (15), even allowing for a change of rate as the virus spreads in a new host (3). Although these estimates clearly depend on the sample of viruses used in the analysis, such that inclusion of earlier isolates from Mexico (or elsewhere) may push back times of ancestry to some extent, they are compatible with a period of "cryptic" viral transmission during which A/H1N1pdm went unnoticed by health authorities for several months. Interestingly, the identification of periods of cryptic viral transmission appears to be a common observation when using molecular clocks to date the onset of viral epidemics.

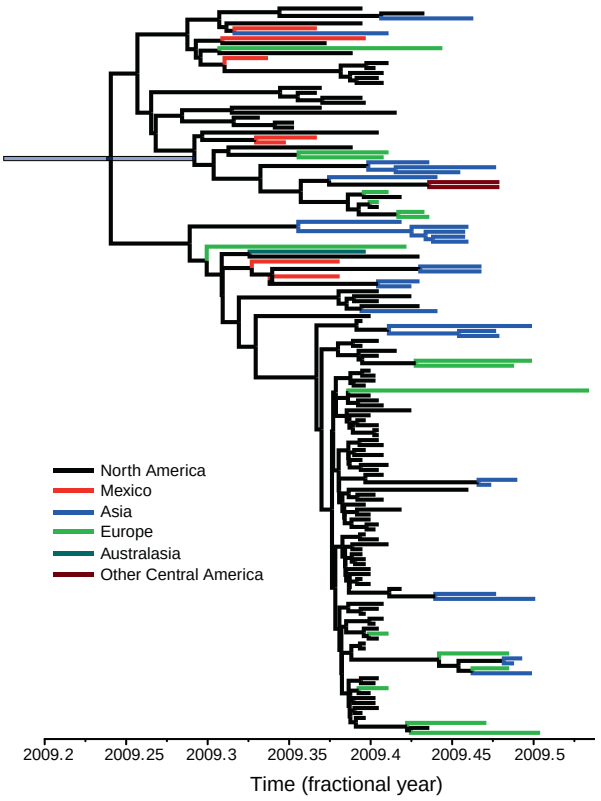


Fig. 2. Maximum clade credibility tree of a sample of 147 complete genomes (13,130 nt) of A/H1N1pdm showing the spatial diffusion of this virus. The tree was estimated by using the Bayesian Markov Chain Monte Carlo method available in the BEAST package (67). The data were analyzed by assuming a relaxed (uncorrelated lognormal) molecular clock under the HKY85 model of nucleotide substitution with a different substitution rate for each codon position and a Bayesian skyline coalescent prior. The chain was run for 200 million generations (with a 10% burn-in). Branches are color-coded by place of origin. The bar at the root node represents the 95% highest probability density for the age of that node (*x* axis). In all cases tip times reflect the time of sampling. Although there is a heavy sampling bias toward American strains, there have clearly been multiple exportation events to localities such as Asia and Europe. Data kindly provided by Andrew Rambaut (University of Edinburgh, Edinburgh, Scotland, UK).

Noteworthy examples include the genotype C rhinoviruses, first described in 2007 although molecular clock estimates place their ancestry at >250 years ago (65), and HIV in the Americas, where the virus was not identified until the early 1980s even though molecular clocks place its time of emergence in the Americas to the late 1960s or early 1970s, long before the earliest AIDS cases (66). As such, molecular evolutionary analyses offer a way to explore the early stages of emergence characterized by hidden viral transmission.

1. Taubenberger JK, Reid AH, Frafft AE, Bijwaard KE, Fanning TG (1997) Initial genetic characterization of the 1918 "Spanish" influenza virus. *Science* 275:1793–1796.
2. Garten RJ, et al. (2009) Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science* 325:197–201.
3. Smith GJ, et al. (2009) Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 459:1122–1125.
4. Dugan VG, et al. (2008) The evolutionary genetics and emergence of avian influenza viruses in wild birds. *PLoS Pathog* 4:e1000076.
5. Obenauer JC, et al. (2006) Large-scale sequence analysis of avian influenza isolates. *Science* 311:1576–1580.
6. Olsen B, et al. (2006) Global patterns of influenza a virus in wild birds. *Science* 312:384–388.
7. Shinde V, et al. (2009) Triple-reassortant swine influenza A (H1) in humans in the United States, 2005–2009. *N Engl J Med* 360:2616–2625.
8. Cox-Foster DL, et al. (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* 318:283–287.
9. Lipkin WI (2009) Microbe hunting in the 21st century. *Proc Natl Acad Sci USA* 106:6–7.
10. Woolhouse MEJ (2002) Population biology of emerging and re-emerging pathogens. *Trends Microbiol* 10:S3–S7.
11. Duffy S, Shackelton LA, Holmes EC (2008) Rates of evolutionary change in viruses: Patterns and determinants. *Nat Rev Genet* 9:267–276.
12. Elena SF, Sanjuán R (2005) Adaptive value of high mutation rates of RNA viruses: Separating causes from consequences. *J Virol* 79:11555–11558.
13. Holmes EC (2009) *The Evolution and Emergence of RNA Viruses*, eds Harvey PH, May RM (Oxford Univ Press, Oxford, UK), Oxford Series in Ecology and Evolution pp 6–8.
14. Jenkins GM, Rambaut A, Pybus OG, Holmes EC (2002) Rates of molecular evolution in RNA viruses: A quantitative phylogenetic analysis. *J Mol Evol* 54:152–161.

15. Rambaut A, Holmes EC (2009) The early molecular epidemiology of the swine-origin A/H1N1 human influenza pandemic. *PLoS Currents Influenza* Aug 18:RRN1003.
16. Rambaut A, et al. (2008) The genomic and epidemiological dynamics of human influenza A virus. *Nature* 453:615–619.
17. Eigen M (1992) *Steps Toward Life* (Oxford Univ Press, New York).
18. Duffy S, Holmes EC (2009) Validation of high rates of nucleotide substitution in geminiviruses: Phylogenetic evidence from East African cassava mosaic viruses. *J Gen Virol* 90:1539–1547.
19. Gago S, Elena SF, Flores R, Sanjuán R (2009) Extremely high mutation rate of a hammerhead viroid. *Science* 323:1308.
20. Wiehe T (1997) Model dependency of error thresholds: The role of fitness functions and contrasts between the finite and infinite sites models. *Genet Res* 69:127–136.
21. Anderson JP, Daifuku R, Loeb LA (2004) Viral error catastrophe by mutagenic nucleosides. *Annu Rev Microbiol* 58:183–205.
22. Bull JJ, Sanjuán R, Wilke CO (2007) Theory of lethal mutagenesis for viruses. *J Virol* 81:2930–2939.
23. Pariente N, Sierra S, Lowenstein PR, Domingo E (2001) Efficient virus extinction by combinations of a mutagen and antiviral inhibitors. *J Virol* 75:9723–9730.
24. Martin V, Grande-Perez A, Domingo E (2008) No evidence of selection for mutational robustness during lethal mutagenesis of lymphocytic choriomeningitis virus. *Virology* 378:185–192.
25. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R (2005) Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 439:344–348.
26. Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. *Nat Rev Genet* 8:610–618.
27. Sanjuán R, Moya A, Elena SF (2004) The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc Natl Acad Sci USA* 101:8396–8401.
28. Krakauer DC, Komarova NL (2003) Levels of selection in positive-strand virus dynamics. *J Evol Biol* 16:64–73.
29. Duffy S, Turner PE, Burch CL (2006) Pleiotropic costs of niche expansion in the RNA bacteriophage ϕ6. *Genetics* 172:751–757.
30. Nielsen R, Weinrich DM (1999) The age of nonsynonymous and synonymous mutations in animal mtDNA and implications for the mildly deleterious theory. *Genetics* 153:497–506.
31. Pybus OG, et al. (2007) Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. *Mol Biol Evol* 24:845–852.
32. Ramsden C, et al. (2008) High rates of molecular evolution in hantaviruses. *Mol Biol Evol* 25:1488–1492.
33. Holmes EC (2003) Patterns of intra- and inter-host nonsynonymous variation reveal strong purifying selection in dengue virus. *J Virol* 77:11296–11298.
34. Wolfe ND, Dunavan CP, Diamond J (2007) Origins of major human infectious diseases. *Nature* 447:279–283.
35. Parrish CR, et al. (2008) Cross-species viral transmission and the emergence of new epidemic diseases. *Micro Mol Biol Rev* 72:457–470.
36. Woelk CH, Holmes EC (2002) Reduced positive selection in vector-borne RNA viruses. *Mol Biol Evol* 19:2333–2336.
37. Coffey LL, et al. (2008) Arbovirus evolution in vivo is constrained by host alternation. *Proc Natl Acad Sci USA* 105:6970–6975.
38. Greene IP, et al. (2005) Effect of alternating passage on adaptation of sindbis virus to vertebrate and invertebrate cells. *J Virol* 79:14253–14260.
39. Jerzak GV, Bernard K, Kramer LD, Shi PY, Ebel GD (2007) The West Nile virus mutant spectrum is host-dependent and a determinant of mortality in mice. *Virology* 360:469–476.
40. Vasilakis N, et al. (2009) Mosquitoes put the brake on arbovirus evolution: Experimental evolution reveals slower mutation accumulation in mosquito than vertebrate cells. *PLoS Pathog* 5:e1000467.
41. Davies TJ, Pedersen AB (2008) Phylogeny and geography predict pathogen community similarity in wild primates and humans. *Proc Biol Sci* 275:1695–1701.
42. Vasilakis N, et al. (2007) Potential of ancestral sylvatic dengue-2 viruses to re-emerge. *Virology* 358:402–412.
43. Anishchenko M, et al. (2006) Venezuelan encephalitis emergence mediated by a phylogenetically predicted viral mutation. *Proc Natl Acad Sci USA* 103:4994–4999.
44. Finkelstein DB, et al. (2007) Persistent host markers in pandemic and H5N1 influenza viruses. *J Virol* 81:10292–10299.
45. Ghedin E, et al. (2005) Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* 437:1162–1166.
46. Ghedin E, et al. (2009) Mixed infection and the genesis of influenza diversity. *J Virol* 83:8832–8841.
47. Bean WJ, Jr, Cox NJ, Kendal AP (1980) Recombination of human influenza A viruses in nature. *Nature* 284:638–640.
48. Holmes EC, et al. (2005) Whole genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *PLoS Biol* 3:e300.
49. Nelson MI, et al. (2006) Stochastic processes are key determinants of the short-term evolution of influenza A virus. *PLoS Pathog* 2:e125.
50. Nelson MI, Simonsen L, Viboud C, Miller MA, Holmes EC (2007) Phylogenetic analysis reveals the global migration of seasonal influenza A viruses. *PLoS Pathog* 3:e131.
51. Nelson MI, et al. (2008) Molecular epidemiology of A/H3N2 and A/H1N1 influenza virus during a single epidemic season in the United States. *PLoS Pathog* 4:e1000133.
52. Bright RA, Shay DK, Shu B, Cox NJ, Klimov AI (2006) Adamantane resistance among influenza A viruses isolated early during the 2005–2006 influenza season in the United States. *J Am Med Assoc* 295:891–894.
53. Simonsen L, et al. (2007) The rapid global spread of reassortant human influenza A/H3N2 viruses conferring adamantane resistance. *Mol Biol Evol* 24:1811–1820.
54. Nelson MI, Simonsen L, Miller MA, Viboud C, Holmes EC (2009) The origin and global emergence of adamantane-resistant A/H3N2 influenza viruses. *Virology* 388:270–278.
55. Rameix-Welti MA, Enouf V, Cuvelier F, Jeannin P, van der Werf S (2008) Enzymatic properties of the neuraminidase of seasonal H1N1 influenza viruses provide insights for the emergence of natural resistance to oseltamivir. *PLoS Pathog* 4:e1000103.
56. Weinstock DM, Zuccotti G (2009) The evolution of influenza resistance and treatment. *J Am Med Assoc* 301:1066–1069.
57. Chen R, Holmes EC (2008) The evolutionary dynamics of human influenza B virus. *J Mol Evol* 66:655–663.
58. Fraser C, et al. (2009) Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science* 324:1557–1561.
59. Smith DJ, et al. (2004) Mapping the antigenic and genetic evolution of influenza virus. *Science* 305:371–376.
60. Ferguson NM, Galvani AP, Bush RM (2003) Ecological and immunological determinants of influenza evolution. *Nature* 422:428–433.
61. Koelle K, Cobey S, Grenfell B, Pascual M (2006) Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science* 314:1898–1903.
62. Recker M, Pybus OG, Nee S, Gupta S (2007) The generation of influenza outbreaks by a network of host immune responses against a limited set of antigenic types. *Proc Natl Acad Sci USA* 104:7711–7716.
63. Russell CA, et al. (2008) The global circulation of seasonal influenza A (H3N2) viruses. *Science* 320:340–346.
64. WHO (2009) *Pandemic (H1N1) 2009: Update 65* (WHO, Geneva).
65. Briese T, et al. (2008) Global distribution of novel rhinovirus genotype. *Emerg Infect Dis* 14:944–947.
66. Gilbert MT, et al. (2007) The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci USA* 104:18566–18570.
67. Drummond AJ, Rambaut (2007) A BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.

Adaptive landscapes and protein evolution

Maurício Carneiro and Daniel L. Hartl[1]

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138

Edited by Diddahally R. Govindaraju, Boston University School of Medicine, and accepted by the Editorial Board August 31, 2009 (received for review July 10, 2009)

The principles governing protein evolution under strong selection are important because of the recent history of evolved resistance to insecticides, antibiotics, and vaccines. One experimental approach focuses on studies of mutant proteins and all combinations of mutant sites that could possibly be intermediates in the evolutionary pathway to resistance. In organisms carrying each of the engineered proteins, a measure of protein function or a proxy for fitness is estimated. The correspondence between protein sequence and fitness is widely known as a fitness landscape or adaptive landscape. Here, we examine some empirical fitness landscapes and compare them with simulated landscapes in which the fitnesses are randomly assigned. We find that mutant sites in real proteins show significantly more additivity than those obtained from random simulations. The high degree of additivity is reflected in a summary statistic for adaptive landscapes known as the "roughness," which for the actual proteins so far examined lies in the smallest 0.5% tail of random landscapes.

antibiotic resistance | fitness landscape | molecular evolution

Attempts to control agents of infectious disease or their vectors have been frustrated time and again by the evolution of resistance in the targeted proteins. How proteins evolve under strong selection is therefore an important line of inquiry, particularly in regard to whether evolutionary pathways can be reproduced or predicted.

The modern concept of protein evolution as a kind of walk in sequence space seems to have originated with John Maynard Smith (1). Responding to a criticism of the theory of natural selection that the number of possible polypeptide sequences is so large that no functional protein could conceivably have arisen by random mutation, Maynard Smith emphasized that favorable mutations are incorporated into a protein sequentially, not simultaneously. He argued by analogy with a word game called change-one-letter, in which the object at each turn is to change one letter in a word to yield a meaningful different word. His example was sequential changes from WORD to GENE as follows: WORD → WORE → GORE → GONE → GENE. His rationale was that, in Darwinian evolution, each change in a protein sequence should be better (or at least no worse) than the present sequence. The basis of these assumptions, he argued, was "that no sensible alternatives have been suggested and that no evidence exists at the moment to invalidate them." And so it is today, despite intelligent design and other creationist critiques.

One limitation of the analogy to the change-one-letter game is that it is usually unknown whether altering a particular amino acid in a protein results in a change in fitness that is beneficial, neutral, or deleterious, hence it is unclear which amino acid replacements are allowed. By means of studying a protein whose sequence can be changed experimentally, and choosing a proxy measure of fitness (such as catalytic activity, protein stability, or drug resistance), the change-one-letter analogy can be converted into an experimental program for studying the pathways of protein evolution (2–7). In most such studies, the number of amino sites allowed to change is deliberately chosen to be relatively small to keep the number of possible combinations of changed sites within the realm of what current technology allows.

Here, we summarize results from several studies that have followed this experimental program (2, 5, 7) and compare the results with expectations based on computer simulations in which the fitness of each combination of mutant sites is assigned at random. We find that, in each case, mutant combinations in actual proteins show significantly more additive effects than would be expected by chance. These results are discussed in the wider context of fitness landscapes in protein space.

The Roads Not Taken

For every realized evolutionary path in sequence space there are other roads not taken. General discussions of evolutionary pathways began ≈80 years ago in the work of Haldane (8) and Wright (9). Wright's article is far better known than Haldane's, probably because Wright's had been written in response to a specific request for a short piece describing his mathematical evolutionary theories for an audience of nonspecialists (10). The general idea is that points in a multidimensional space consisting of gene combinations (appropriate for individuals) or allele frequencies (appropriate for populations) is projected onto two dimensions, and a third dimension representing the fitness of each genotype (or the average fitness of each population) is added. Because the simplest models of natural selection result in increasing fitness (11), evolution can be thought of as a sort of walk on a fitness landscape, which may be smooth with one highest fitness peak or rough with multiple submaximal fitness peaks separated by valleys of lower fitness.

Wright's diagram (9) showed a surface with two local fitness maxima. It illustrated how he envisioned evolution to take place under increased mutation or relaxed selection, decreased mutation or intensified selection, weak or strong inbreeding, a change of environment, or in a subdivided population. The diagram was a great success and was picked up and republished in numerous other papers and books (10). The diagram prominently highlighted Wright's shifting balance theory of evolution (12), in which random genetic drift plays a key role in enabling a population to explore its adaptive landscape notwithstanding peaks and valleys.

The problem that the shifting balance theory was supposed to solve is depicted in the context of protein evolution in Fig. 1. The height of each cube in Fig. 1 is proportional to the fitness of a haploid organism (or that of a homogeneous population of haploid individuals) whose genome encodes a protein with any of four possible combinations of amino acids at two distinct sites. For simplicity, only two possible amino acids at each site are considered, hence the choices are binary and the combinations can be designated as 00, 10, 01, and 11. The model of protein evolution is essentially that of Maynard Smith (1), which has become known as the strong-selection, weak-mutation model

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Evolution in Health and Medicine" held April 2–3, 2009, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS web site at www.nasonline.org/Sackler_Evolution_Health_Medicine.

Author contributions: D.L.H. designed research; M.C. performed research; M.C. and D.L.H. analyzed data; and M.C. and D.L.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. D.R.G. is a guest editor invited by the Editorial Board.

[1]To whom correspondence should be addressed. E-mail: dhartl@oeb.harvard.edu

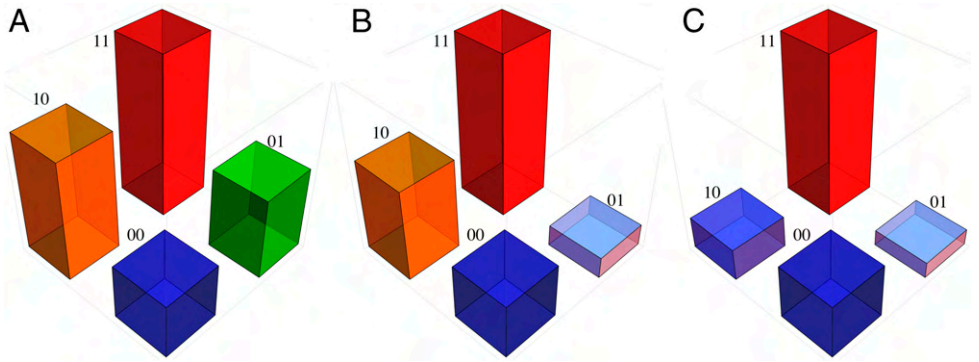


Fig. 1. Examples of gene interaction (epistasis) in fitness landscapes. Each cube's height is proportional to the fitness of organisms having mutant proteins with combinations of two variant amino acids, arbitrarily designated 00, 10, 01, and 11. (*A*) Magnitude epistasis: One highest fitness peak, two allowed paths to 11. (*B*) Sign epistasis: One highest fitness peak, one allowed path to 11. (*C*) Reciprocal sign epistasis: One highest peak (combination 11) and one submaximal peak (00); no paths from 00 to 11.

(13). Evolution on the landscape occurs through random mutation, one site at a time, with the probability of fixation of any beneficial amino acid replacement proportional to its selective advantage (14). The genetically heterogeneous populations that exist during the transitions between states are not depicted, on the grounds that, under strong selection and weak mutation, the time to fixation is short relative to the waiting time between favorable mutations.

Suppose the initial population in Fig. 1*A* is fixed for the all-0 amino acid sequence 00. Mutations to either 10 or 01 are likely to become fixed, and either of these states can mutate to the still more favorable state 11. In Fig. 1*B*, the evolutionary pathway to 11 through 10 is still accessible, but that through 01 is not, owing to the decrease in fitness between 01 and 00. In Fig. 1*C*, all pathways to 11 are blocked by the reduced fitness of the intermediates, and the population becomes stranded on the submaximal fitness peak 00. Random genetic drift can alleviate this situation because, with a small fitness differentials and a small enough population size, a population at 00 could, by chance, evolve into one fixed for either 10 or 01, and from either of these states go to 11, thereby achieving the highest fitness state in the landscape. In principle, the shifting balance theory would work in this manner, but there are many difficulties in practice (15). There is a convenient terminology for the types of fitness landscapes shown in Fig. 1: The pattern depicted in *A* exemplifies magnitude epistasis, that in *B* exemplifies sign epistasis, and that in *C* exemplifies reciprocal sign epistasis (16). Except when interpreted as a Wright-type metaphor (9), fitness landscapes with a greater dimensionality than that shown in Fig. 1 cannot be depicted in two dimensions.

Random Fitness Landscapes of Low Dimensionality

A rich literature deals with fitness landscapes in which the fitnesses of genotypes are assigned at random, either with statistical independence or specified patterns and strengths of correlation (17–21). Much but not all of this literature focuses on landscapes of high dimensionality, and it deals with issues such as the fitness ultimately achieved (22), the role of mutation bias (23), noisy fitness mappings (24), genetic robustness (25), and whether the likelihood of becoming stranded at a submaximal fitness peak is reduced at high dimensionality (26). Our present focus is on fitness landscapes of low dimensionality, because these are the types of landscapes presently amenable to experimental investigation.

Fig. 2 shows some results of simulated fitness landscapes whose dimensionality is in the range amenable to experimental study using current techniques. At each site the choices are binary (either 0 or 1). The combination of all zeroes is assigned a fitness of 0. We use malthusian parameters for fitness, which means that the growth rate of a homogeneous population consisting of organisms with a fitness of 0 is Exp[0] = 1 (11). The combination of all ones is assigned a fitness of 1. Every other mutant combination is assigned a fitness at random and independently with a uniform distribution on [0, 1]. This model is similar to the so-called NK model with $K = N - 1$ (17); however, it differs in that the fitnesses of the all-0 and all-1 states are not random variables. For each of 10,000 randomly assigned fitness

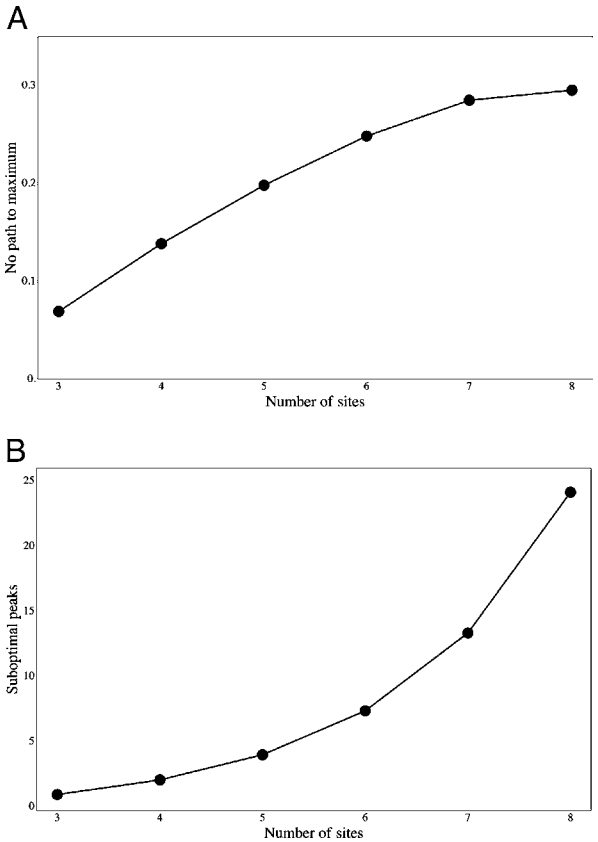


Fig. 2. Some features of random fitness landscapes. (*A*) Average proportion of random fitness landscapes with no allowable paths to the maximum. (*B*) Average number of accessible submaximal fitness peaks among random fitness landscapes. Results of 10,000 simulations of random fitness landscapes are shown. Fitness of the all-0 combination was assigned a value of 0 (malthusian fitness), that of all-1 combination was assigned a value of 1.0, and those of all other combinations were taken from a random uniform distribution on [0, 1].

landscapes, we assumed an initial population consisting of individuals of the all-0 genotype and let mutations occur to the alternative sites at random, one at a time. If a mutation decreases fitness it is discarded, but if the mutation increases fitness, it is regarded as defining an allowed step in an evolutionary pathway, and a transition to the mutant state takes place. The mutation-selection process was repeated until we had mapped all paths from the all-0 state to any state in which no single-step mutation could increase fitness further. Each allowed path was also assigned a probability of occurrence according to the rule that the probability of fixation of a favorable mutation is proportional to its selective advantage (14).

Fig. 2*A* shows the average proportion of random landscapes that have no allowable evolutionary path (an allowable path increases fitness at each step) from the all-0 state to the all-1 state, as a function of the number of amino acid sites. The minimum is at three sites, and the number increases almost linearly at first, but then seems to level off at ≈30%. The values for 9–13 sites are similar to those for 8 sites. At the same time, as the number of sites increases (Fig. 2*B*), the number of submaximal fitness peaks increases, from near 1 at $n = 3$ sites to ≈25 at $n = 8$, and the exponential increase continues for 9–13 sites. These are, we must emphasize, submaximal fitness peaks that are accessible through a sequence of single steps of mutation and selection, each step of which increases fitness. In our modification of the NK model, it can be shown from results in ref. 17 that the number of submaximal fitness peaks with n sites is given by $2^n/(n + 1)$, but some of these submaximal fitness peaks may not be accessible. To revert to the landscape analogy, these submaximal fitness peaks are inaccessible because they are surrounded by a fitness "moat."

Although the majority of random fitness landscapes of low dimensionality include one or more paths to the maximum (Fig. 2*A*), the chance of any population reaching the maximum is actually quite bleak. Weighing the probability of each successive fixation by the fitness advantage of the new mutant, the overall probability of reaching the maximum on a fitness landscape with three sites is 0.53 ± 0.38. This average is somewhat misleading because the distribution of probabilities is strongly bimodal: starting from the all-0 state, ≈1/3 of the landscapes have a probability of reaching the maximum of 1.0, and the remaining have an average probability of reaching the maximum of ≈0.30. For four binary sites, the probability of reaching the maximum averages 0.18 ± 0.25, and for five binary sites it is 0.04 ± 0.10. Each of the latter distributions is strongly skewed toward 0. If it was Wright's intuition that complex interactions between genes result in fitness landscapes that make it difficult for any evolving population to attain the maximum fitness, then his intuition is validated, at least for random landscapes of low dimensionality.

Roughness

As might be expected, random fitness landscapes show considerable variation, and hence it is unclear how one might compare one landscape to the next or any set of landscapes with data from actual proteins. One feature of fitness landscapes that does admit of comparison is the "roughness," defined as the root mean sum of squares of the residual variation after removing the main additive effects of each amino acid site (3). The main additive effects are obtained by least squares. For two amino acid sites, to take a concrete example, the main additive effects of sites 1 and 2 (ε_1 and ε_2) are obtained by minimizing

$$Q = (f_{11} - 1)^2 + (f_{10} - 1 - \varepsilon_1)^2 + (f_{01} - 1 - \varepsilon_2)^2 + (f_{00} - 1 - \varepsilon_1 - \varepsilon_2)^2,$$

where f_{ij} is the fitness of an organism whose genome encodes a protein with the amino acids i and j ($i, j = 0, 1$) at the two sites, and $f_{11} = 1$. Hence $\varepsilon_1 = (1/3)(f_{00} + 2f_{10} - f_{01} - 2)$ and $\varepsilon_2 = (1/3)(f_{00} - f_{10} - 2f_{01} - 2)$. The roughness of a landscape is defined as roughness = $\sqrt{Q/4}$. For a generalization to any number of alternative amino acids at any number of sites, see ref. 3. For a fitness landscape in which the main effects of the amino acid replacements are completely additive, the roughness equals 0. For example, if the fitnesses corresponding to the cubes in Fig. 1 are assigned values of 0.25, 0.50, 0.75, and 1.0 according to their height, then the roughness of the landscape in Fig. 1*A* is 0, that of Fig. 1*B* is 0.1443, and that of Fig. 1*C* is 0.2886.

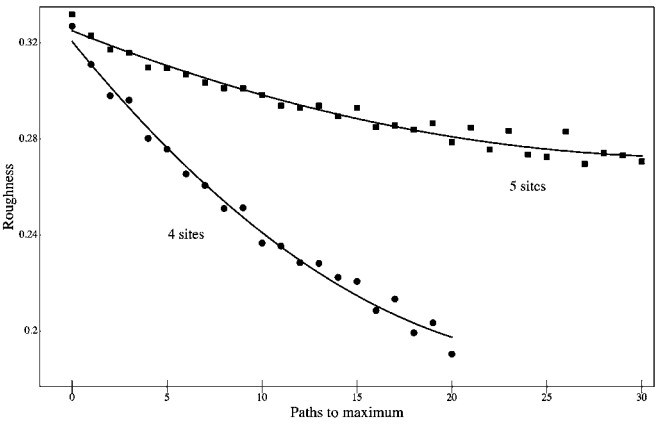


Fig. 3. Roughness of random fitness landscapes with four variant sites (●) or five variant sites (■), as a function of number of paths to the maximum, among the random landscapes described in the Fig. 2 legend. Curves are quadratic fit by least squares.

Roughness serves as one convenient metric by which fitness landscapes can be compared. Fig. 3 shows the relation between roughness and number of accessible paths to the maximum for landscapes with four or five binary sites. As might have been expected on intuitive grounds, the average roughness decreases as the number of paths to the maximum increase. Less intuitive are the patterns in Fig. 4, which show the relation between number of accessible submaximal fitness peaks and roughness. For landscapes with more than two such submaximal peaks, there is little or no relation to roughness. Virtually the same patterns emerge from an analysis of 100,000 random landscapes as those shown here for 10,000 landscapes.

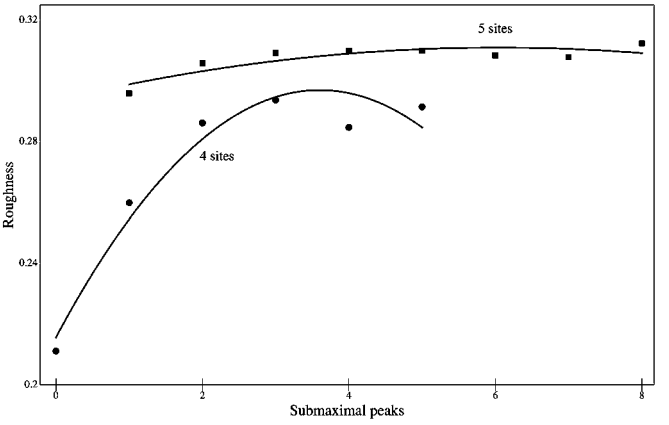


Fig. 4. Roughness of random fitness landscapes with four variant sites (●) or five variant sites (■), as a function of number of accessible submaximal fitness peaks, among the random landscapes described in the Fig. 2 legend. Curves are quadratic fit by least squares.

Table 1. Roughness of empirical fitness landscapes

Protein (Ref.)	Sites	Roughness	Mean	SD	*P* value
Lysozyme (2)	T40S, I55V, T91S	0.00388	0.0242	0.00653	<0.001
Dihydrofolate reductase (7)	N51I, C59R, S108N, I164L	0.22449	0.34520	0.04767	≈0.005
TEM β-lactamase (5)	g4205a, A42G, E104K, M182T, G238S	0.27667	0.36567	0.02491	≈0.0002

Mean roughness of simulated landscapes is based on random permutations. SD of simulated landscapes is based on random permutations. Approximate *P* value is for the difference between the observed and simulated mean roughness.

Actual Fitness Landscapes

How do real fitness landscapes compare with those in which fitnesses are randomly assigned? Table 1 shows three examples with a small number of binary mutant sites in which all possible mutant combinations have been created and assayed for some measure of protein function or some proxy for fitness. In the case of lysozyme, the assay of protein function is thermal stability (2), for dihydrofolate reductase the fitness proxy is the concentration of pyrimethamine that decreases growth rate by 50% (7), and for TEM β-lactamase (TEM stands for Temoniera, the name of the patient from whom the enzyme was first isolated) the fitness proxy is minimal inhibitory concentration of cefotaxime (5). Lysozyme illustrates a case with three binary sites, dihydrofolate reductase with four binary sites, and TEM β-lactamase with five binary sites (g4205a is a regulatory site, not an amino acid-coding site).

In each case, we estimated the roughness of the actual fitness landscape and compared it with the distribution of the roughness values of 10,000 simulated landscapes obtained by random permutations of all of the fitness values excluding those of the all-0 and all-1 states (3). Approximate *P* values were estimated based on the deviation between the observed roughness and the simulated mean in units of SD. In all cases the observed roughness is highly significantly less rough than that expected with random permutations. These results are consistent with other studies of empirical fitness landscapes that include more sites (3, 4, 6).

Biologically, the reduced roughness of actual fitness landscapes means that the effects of mutant sites show highly significantly more additivity than those obtained from random simulations. This inference does not diminish the potential importance of interactions among sites (epistasis). Perfect additivity would yield a roughness of 0, whereas the observed value for dihydrofolate reductase is 4.7 SD > 0, and that for TEM β-lactamase is 5.6 SD > 0. The result does, however, suggest that reciprocal sign epitasis, in which individually deleterious mutations become beneficial when combined (6, 16), is not pervasive in the handful of examples that have thus far been examined in detail.

A Tail of Random Landscapes

The significant additivity of actual fitness landscapes prompts another look at the seemingly bleak prospect of an evolving population attaining the highest fitness peak in a random landscape. It suggests that comparison with random landscapes is untenable, and that one should instead examine only the tail end of the roughness distribution of random landscapes in which the sites in the simulated landscapes are more additive than those in the distribution as a whole. Because the largest *P* value in Table 1 is ≈0.5%, we examined only those 500 landscapes comprising the least rough 0.5% of the roughness distribution among 100,000 random and uncorrelated fitness landscapes. The results were quite different from those described earlier. For three, four, and five binary sites, the probability of attaining the maximum was 0.993 ± 0.074, 0.708 ± 0.320, and 0.219 ± 0.225, respectively, and the number of allowable paths to the maximum was 5.96 ± 0.32, 18.6 ± 6.4, and 29.6 ± 21.1, respectively. It therefore appears that the subset of random landscapes showing approximately the levels of additivity as actual molecules would offer a good chance of fixation of the allele with maximum fitness, without the need to invoke random genetic drift, noisy fitnesses, changing environments, or other ad hoc processes. Each of these is an important process in its own right, but may not be essential in exploring fitness landscapes with the levels of additivity actually observed.

Evolutionary Pathways to Higher Fitness

Fitness landscapes with low but nonzero roughness result from sites that show more additivity than expected by chance. They nevertheless show magnitude epistasis, in which the fitness effects of a mutant site in different genetic backgrounds differ in magnitude but not in sign. Many also show sign epistasis, in which a mutant site has opposite effects depending on the genetic background. Although reciprocal sign epistasis, in which individually harmful mutations are favorable in combination, cannot be neglected because it is observed in a few combinations (5, 16), nevertheless it seems not to be pervasive. The major practical implication of landscapes featuring mainly magnitude and sign epistasis is that they constrain the pathways of protein evolution without shutting off pathways to the maximum. In the case of TEM β-lactamase (5), for example, only 18 of 120 theoretically possible evolutionary pathways to highest resistance are allowable (i.e., show increased resistance at each step), and a mere five pathways account for ≈80% of the probability. Likewise for transgenic bacteria carrying the dihydrofolate reductase gene from the malaria parasite (7), in which only 10 of 24 theoretically possible pathways are allowable, and just three pathways account for ≈90% of the probability.

The relatively high probabilities of a small number of pathways means that evolution on low-roughness pathways has a degree of predictability and reproducibility that would not necessarily be expected (27). Experimental studies of fitness landscapes may therefore be informative for processes that have happened, or are happening, in nature. For example, the high-probability evolutionary pathways identified for the evolution of pyrimethamine resistance of the malaria dihydrofolate reductase studied in *Escherichia coli* coincide exactly with the inferred stepwise acquisition of pyrimethamine resistance in the malaria parasite itself, as inferred from amino acid polymorphisms in extant populations and in vitro studies of the mutant enzymes (7). Such good agreement between studies in transgenic organisms (in this case, organisms in different kingdoms) may not be expected in general, but this particular example offers hope that much of importance can be learned from judicious choice of protein, model organism, and experimental protocol.

Should the Fitness Landscape Be Buried?

The landscape metaphor is continuously alluring, "a powerful quantitative concept in biology" (28). However, its acclaim has been mixed. Wright's conflation of the landscape for individual fitness with that for population average fitness has led to confusion and controversy. Among the most severe critics is Wright's biographer (ref. 10, p. 316), who called adaptive landscapes "unintelligible, ... meaningless in any precise

sense." Another thoughtful observer has recommended that it "is time to give up the pictorial metaphor of the landscape entirely" (29). Wright himself seemed momentarily to have misgivings. In a 1986 letter to Provine, he says "The object [in 1932] was to give pictorial representations of elementary evolutionary processes, ... but sources of confusion in the multidimensional nature of the field as a whole, and the contributions of each locus to the combinations, may have made this attempt a mistake." But by 1988, in his last published article, appearing 2 months before his death, Wright seems to have changed his mind. He wrote "I think that [Provine] was looking for something more mathematical than was intended ... It is assumed that the genotypes are packed, side by side, in a two-dimensional space in such a way that each is surrounded by genotypes that differ by only one gene replacement. Correspondence with geographical continuity is a secondary consideration ... It is obvious that this two-dimensional surface of selective values cannot accurately represent relations that are multidimensional both among and within loci. It is useless for mathematical purposes" (30).

Poor Adaptive Landscape

If one may be permitted a metaphor for a metaphor, one could think of the adaptive landscape as a small pack burro that has been loaded with excessive baggage. The mistreated beast has been asked to carry central optimizing principles in population genetics, developmental biology, systems biology, gene regulation, neural dynamics, computer algorithms, protein folding, manufacturing strategy, technology policy, and who knows what else (e.g., refs. 19, 21, and 28). Should the overloaded landscape metaphor, therefore, be abandoned? We think yes and no. The adaptive landscape is a metaphor, nothing more, and like all metaphors and analogies is misleading when pushed too far. Even the change-one-letter game becomes absurd if you start the game with a word such as "syzygy." It is asking too much of the adaptive landscape metaphor to accommodate limit cycles or changing environments. Wright invented it as nothing more than a visual aid for nonmathematical biologists who were attending the 1932 International Congress of Genetics in Ithaca, New York (10). It should be taken in the spirit in which he intended. Fitness landscapes should not be abandoned, but rather studied in less picturesque but more quantitative ways. An approach using summary statistics such as roughness seems promising, but there may be other characterizations of fitness landscapes that are equally or more informative.

ACKNOWLEDGMENTS. We thank Kyle M. Brown for comments on the manuscript and Daniel M. Weinreich for many helpful conversations about fitness landscapes. This work was supported by National Institutes of Health Grant R01GM079536 (to D.L.H.).

1. Maynard Smith J (1970) Natural selection and the concept of a protein space. *Nature* 225:563–564.
2. Malcolm BA, Wilson KP, Matthews BW, Kirsch JF, Wilson AC (1990) Ancestral lysozymes reconstructed, neutrality tested, and thermostability linked to hydrocarbon packing. *Nature* 345:86–89.
3. Aita T, Iwakura M, Hasumi Y (2001) A cross-section of the fitness landscape of dihydrofolate reductase. *Protein Eng* 14:633–638.
4. Lunzer M, Miller SP, Felsheim R, Dean AM (2005) The biochemical architecture of an ancient adaptive landscape. *Science* 310:499–501.
5. Weinreich DM, Delaney NF, DePristo MA, Hartl DL (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312:111–114.
6. Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ (2007) Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* 445:383–386.
7. Lozovsky ER, et al. (2009) Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proc Natl Acad Sci USA* 106:12025–12030.
8. Haldane JBS (1931) A mathematical theory of natural selection, part VIII: Metastable populations. *Proc Cambridge Philos Soc* 27:137–142.
9. Wright S (1932) The roles of mutation, inbreeding, cross-breeding and selection in evolution. *Proc Sixth Int Cong Genet* 1:356–366.
10. Provine WB (1986) *Sewall Wright and Evolutionary Biology* (Univ Chicago Press, Chicago).
11. Hartl DL, Clark AG (2007) *Principles of Population Genetics* (Sinauer, Sunderland, MA).
12. Wright S (1931) Evolution in Mendelian populations. *Genetics* 16:97–159.
13. Gillespie JH (1984) Molecular evolution over the mutational landscape. *Evolution (Lawrence, Kans)* 38:1116–1129.
14. Haldane JBS (1927) A mathematical theory of natural and artificial selection, part V: Selection and mutation. *Proc Cambridge Philos Soc* 28:838–844.
15. Coyne JA, Barton NH, Turelli M (1997) A critique of Sewall Wright's shifting balance theory of evolution. *Evolution (Lawrence, Kans)* 51:643–671.
16. Weinreich DM, Watson RA, Chao L (2005) Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution (Lawrence, Kans)* 59:1165–1174.
17. Kauffman S, Levin S (1987) Toward a general theory of adaptive walks on rugged landscapes. *J Theor Biol* 128:11–45.
18. Kauffman SA (1993) *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford Univ Press, New York).
19. Kauffman SA (1995) *At Home in the Universe: The Search for Laws of Self-Organization and Complexity* (Oxford Univ Press, New York).
20. Gavrilets S (2004) *Fitness Landscapes and the Origin of Species* (Princeton Univ Press, Princeton).
21. Frenken K (2005) *Innovation, Evolution, and Complexity Theory* (Edward Elgar, Cheltenham, UK).
22. Rokyta DR, Beisel CJ, Joyce P (2006) Properties of adaptive walks on uncorrelated landscapes under strong selection and weak mutation. *J Theor Biol* 243:114–120.
23. Stoltzfus A (2006) Mutation-biased adaptation in a protein NK model. *Mol Biol Evol* 23:1852–1862.
24. Levitan B, Kauffman SA (1995) Adaptive walks with noisy fitness measurements. *Mol Diversity* 1:53–68.
25. Kim Y (2007) Rate of adaptive peak shifts with partial genetic robustness. *Evolution (Lawrence, Kans)* 61:1847–1856.
26. Gravner J, Pitman D, Gavrilets S (2007) Percolation on fitness landscapes: Effects of correlation, phenotype, and incompatibilities. *J Theor Biol* 248:627–645.
27. Orr HA (2005) The genetic theory of adaptation: A brief history. *Nat Rev Genet* 6:119–127.
28. Ao P (2009) Global view of bionetwork dynamics: Adaptive landscapes. *J Genet Genomics* 36:63–73.
29. Kaplan J (2008) The end of the adaptive landscape metaphor? *Biol Philos* 23:625–638.
30. Wright S (1988) Surfaces of selective value revisited. *Am Nat* 131:115–123.

Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies

Adam Eyre-Walker

School of Life Sciences, University of Sussex, Brighton BN1 7FR, United Kingdom

Edited by Diddahally R. Govindaraju, Boston University School of Medicine, and accepted by the Editorial Board December 8, 2009 (received for review July 3, 2009)

A model is investigated in which mutations that affect a complex trait (e.g., heart disease) also affect fitness because the trait is a component of fitness or because the mutations have pleiotropic effects on fitness. The model predicts that the genetic variance, and hence the heritability, in the trait is contributed by mutations at low frequency in the population, unless the mean strength of selection of mutations that affect the trait is very small or weakly selected mutations tend to contribute disproportionately to the trait compared with strongly selected mutations. Furthermore, it is shown that each rare mutation tends to contribute more to the variance than each common mutation. These results may explain why most genome-wide association studies have failed to find associations that explain much of the variance. It is also shown that most of the variance in fitness contributed by new nonsynonymous mutations is caused by mutations at very low frequency in the population. This implies that most low-frequency SNPs, which are observed in current resequencing studies of, for example, 100 chromosomes, probably have little impact on the variance in fitness or traits. Finally, it is shown that the variance contributed by a category of mutations (e.g., coding or regulatory) depends largely upon the mean strength of selection; this has implications for understanding which types of mutations are likely to be responsible for the variance in fitness and inherited disease.

disease | evolution

Most biological traits, and many of the characters that are of most interest to humans, are complex, or quantitative, in nature; they are determined by many mutations in multiple loci. These traits include diseases such as heart disease, type II diabetes, and schizophrenia, but also traits of commercial and biological interest such as milk yield and clutch size.

Because of the great health and commercial implications, there has been considerable interest in finding the mutations and genes involved in complex traits, particularly those involved in human disease, both to aid in predicting risk and to further understand the genetic basis of disease. This project has advanced rapidly in humans in the past few years with the help of the HAPMAP project (1, 2) and genome-wide association studies (3). By the end of December 2008, more than 300 associations had been reported in more than 70 common diseases (4). Unfortunately, although these studies have successfully identified many mutations associated with a variety diseases and traits, these associations explain very little of the variance in each trait (5–8). A good example is human height; three separate studies involving approximately 63,000 individuals have identified 54 markers associated with height, but, all told, these explain less than 10% of the variation in human height, a trait that has a heritability of 80% (9–11).

A number of different explanations have been proffered as to why the associations discovered by genome-wide association studies explain so little of the variance; these have included epistasis, epigenetics, incomplete association, and structural variation (6, 8). It has also been suggested that rare variants with large or moderate effects might contribute much of the variance, and that such mutations might be rare because they have deleterious fitness effects (8, 12, 13). As pleiotropy appears to be very common (14), mutations that affect a trait may be subject to selection, either because the trait is a component of fitness or because the mutations have pleiotropic effects on fitness. As a consequence, mutations with large effects on the trait may be kept at low frequency because of their deleterious effects on fitness (8, 12, 13). Here I explore a pleiotropic model in which mutations affect both fitness and a trait (15, 16).

Let us consider a population of diploid organisms under a Fisher-Wright model with stationary population size. We will assume that mutations are semidominant and that mutations affect both fitness and some trait of interest, such as body size; we assume that the trait is not subject to selection directly unless the trait is fitness itself. Let us assume that all mutations are deleterious, although some may be sufficiently weakly selected to be effectively neutral; let the fitnesses of the three genotypes be 1, 1-s and 1–2s and the effects of the mutations on the trait be 0, z, and 2z. Let z and s be related to each other by the following model:

$$z(S, \varepsilon; \delta, t) = \delta S^{\tau}(1 + \varepsilon) \quad [1]$$

where $S = 4N_e s$, ε is normally distributed with a mean of zero and an SD of σ, and δ randomly takes a value of +1 or −1 with equal probability. Examples of the relationship are given in Fig. 1. The δ parameter transforms the distribution of effects such that mutations have equal probabilities of increasing or decreasing the trait, unless the trait is fitness, in which case δ is +1. The strength of association between the effects of mutations on the trait and fitness is dependent upon two parameters, σ and τ. As σ becomes larger, so the dependency decreases. However, this parameter turns out to be unimportant in terms of how the variance in the trait is distributed with respect to allele frequency (as discussed later). In contrast, τ is important; it measures how the SD in the trait, or equivalently the mean absolute effect of a mutation on the trait, increases with the strength of selection. If τ is 1, this increase is linear; so a mutation that has a 10 fold larger effect on fitness will also, on average, have a 10 fold larger effect on the trait. If τ is 0, the effects of a mutation on trait and fitness are independent.

We assume that S is Γ-distributed:

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Evolution in Health and Medicine" held April 2–3, 2009, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS web site at www.nasonline.org/Sackler_Evolution_Health_Medicine.

Author contributions: A.E.-W. designed research, performed research, analyzed data, and wrote the paper.

The author declares no conflicts of interest.

This article is a PNAS Direct Submission. D.R.G. is a guest editor invited by the Editorial Board.

E-mail: a.c.eyre-walker@sussex.ac.uk.

This article contains supporting information online at www.pnas.org/cgi/content/full/0906182107/DCSupplemental.

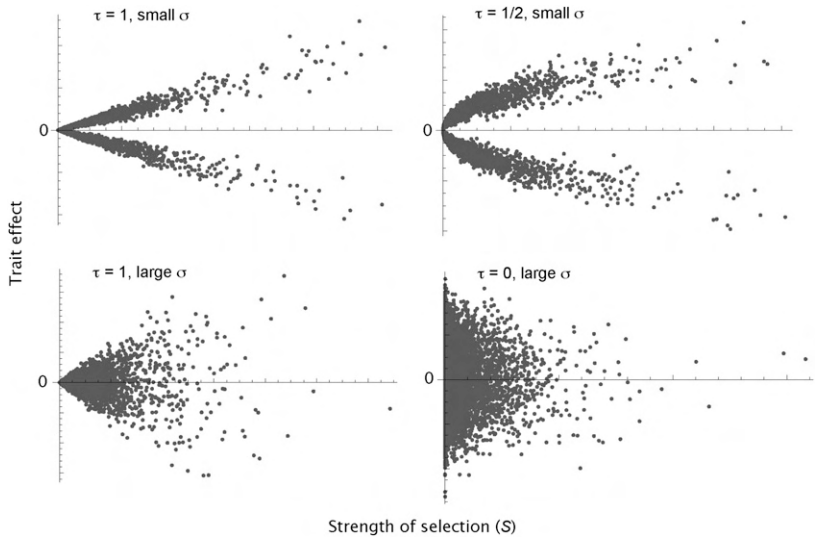


Fig. 1. The relationship between trait and fitness, assuming the distribution of fitness is a Γ-distribution (β = 0.20 and $\bar{S}$ = 3,000) for different values of *t* and σ.

$$D(S;\bar{S},\beta) = \frac{(\beta/\bar{S})^{\beta} S^{\beta-1} e^{-\beta s/\bar{s}}}{\Gamma(\beta)} \quad [2]$$

where β is the shape parameter and $\bar{S}$ is the mean strength of selection.

If we assume free recombination and that the effects of mutations combine additively, then we can write down an expression for the additive genetic variance contributed by mutations as a function of their frequency in the population, *x*, as:

$$V(x) = 2N_e u \int_{-\infty}^{\infty}\int_{0}^{\infty} D(S;\bar{S},\beta)N(\varepsilon;0,\sigma)H(S,x)U(z(S,\tau,\varepsilon),x)dSd\varepsilon \quad [3]$$

where $N(\varepsilon;0,\sigma)$ is the distribution of ε,

$$H(S,x) = 2\left(\frac{1-e^{S(1-x)}}{x(1-x)(1-e^{S})}\right) \quad [4]$$

and

$$U(z,x) = 2x(1-x)z^2 \quad [5]$$

H(S,x) is the time that a new mutation of selective strength *S* spends at a frequency *x* (17) and *U(z,x)* is the variance in the trait contributed by a mutation of effect *z* at frequency *x*. Eq. **3** takes into account mutation, selection, and genetic drift. It can solved to yield the following:

$$V(x) = \frac{\theta}{\Gamma(\beta)}\left(\frac{\beta}{\bar{S}}\right)^{\beta}(1+\sigma^2)\Gamma(2\tau+\beta)\left(Zeta\left(2\tau+\beta,x+\frac{\beta}{\bar{S}}\right) - Zeta\left(2\tau+\beta,\frac{\bar{S}+\beta}{\bar{S}}\right)\right) \quad [6]$$

where *Zeta* is the Hurwitz Zeta function and θ is $4N_e u$. It is useful to know, for plotting Eq. **6**, the density of the variance as a function of $\log_{10}$ of the allele frequency; this is $V'(y) = ve^{vy}V(e^{vy})$, where *y* is $\log_{10}(x)$ and *v* is $\log_e(10)$.

The total additive genetic variance in the trait is:

$$V_T = \int_0^1 V(x)dx$$

which simplifies to:

$$V_T = \frac{\theta}{\Gamma(\beta)}\left(\frac{\beta}{\bar{S}}\right)^{\beta}(1+\sigma^2)\Gamma(2\tau+\beta)\left(\frac{\left(Zeta\left(2\tau+\beta-1,\frac{\beta}{\bar{S}}\right) - Zeta\left(2\tau+\beta-1,\frac{\bar{S}+\beta}{\bar{S}}\right)\right)}{(2\tau+\beta-1)} - Zeta\left(2\tau+\beta,\frac{\bar{S}+\beta}{\bar{S}}\right)\right) \quad [7]$$

It is evident on inspection of Eqs. **6** and **7** that $V(x) / V_T$ is independent of θ and σ. A Mathematica notebook of these equations is available from the author's Web site (*www.lifesci.susx.ac.uk/home/Adam_Eyre-Walker/*).

Results

In our model we assume that mutations have effects on both fitness and a trait of interest, such as the chance of developing diabetes. The effects of the mutation on fitness and the trait are correlated, and this correlation can vary from very strong, when the trait is fitness, to very weak, when the effects are independent of one another. Let us start by considering the case in which the trait is fitness itself (i.e., δ = +1, τ = 1, and σ = 0), and as a starting point let us consider the distribution of fitness effects (DFE) to be that inferred for new amino acid mutations in humans [β = 0.20, $\bar{S}$ = 3000 (18); see also refs. 19 and 20]. It is apparent that, under this distribution, the vast majority of the variance, and hence heritability, in fitness is contributed by mutations that are very rare in the population; 96% of the variance is contributed by mutations

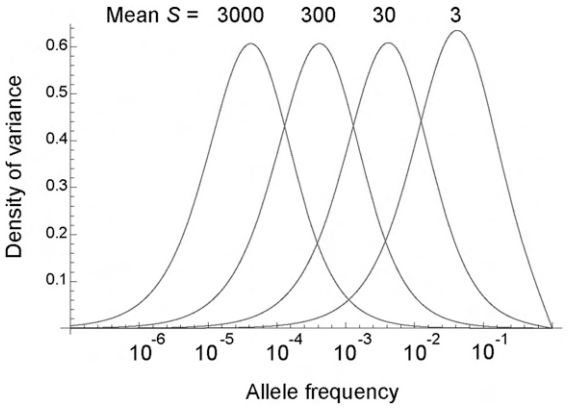


Fig. 2. The density of variance, $V(x) / V_T$, in the trait as a function of allele frequency when the trait is fitness for different mean strengths of selection when β is 0.20.

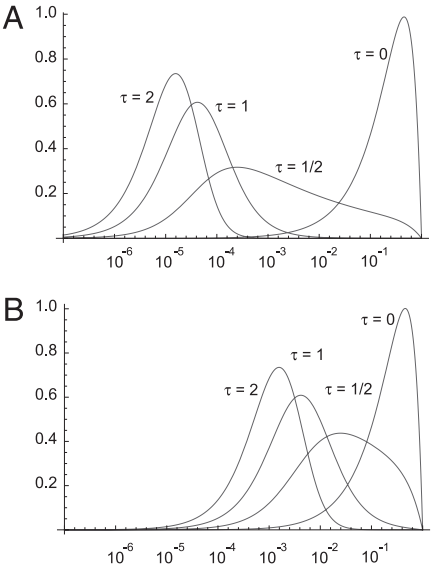


Fig. 3. The density of variance, $V(x) / V_T$, in the trait as a function of allele frequency, for different values of τ and different mean strengths of selection (β = 0.20): (*A*) $\bar{S}$ of 3,000, (*B*) $\bar{S}$ of 30.

that are at a frequency of less than 0.001 (Fig. 2). This pattern depends little on the shape of the distribution (Fig. S1), although more of the variance tends to be contributed by mutations at high frequencies when the distribution is less leptokurtic (i.e., higher values of β). However, the relationship between the variance and allele frequency does depend strongly upon the mean strength of selection; the lower the mean fitness, the more of the variance is contributed by mutations at relatively high frequencies in the population (Fig. 2). Nevertheless, even if the mean $N_e s$ is just 10, we expect 95% of the additive variance to be contributed by mutations at less than 5% in the population and 73% to be contributed by mutations at less than 1%.

It is therefore apparent that the majority of the variance in fitness tends to be contributed by mutations that are rare in the population unless the mean strength of selection is small. This is a simple consequence of two facts: mutations with large effect tend to contribute disproportionately to the variance, and mutations that have large effect are rare because natural selection is effective at minimizing their frequency in the population. Some insight into this can be gained from simple population genetics. The variance contributed by a mutation of selective strength s at a frequency x in the population is $2x(1-x)s^2$. If the mutation is deleterious, it will be rare, so the variance is approximately $2xs^2$. At equilibrium, the mean frequency of such a mutation in the population is approximately u/s, so the variance contributed by such mutations is $2us$. The variance therefore depends upon the mutation rate and the strength of selection; hence categories of mutations that are numerous or are strongly selected tend to contribute most to the variance in fitness.

It is also of interest to determine the variance in fitness contributed by a single mutation at a particular frequency. This can be calculated by dividing the variance contributed by mutations at a frequency by the density of mutations at that frequency. As expected, there are more mutations at lower frequencies, but on a log scale this difference in density is relatively small over a broad range of parameters (Fig. S2); hence the variance contributed by single mutations at a particular frequency is very similar to the proportion of variance contributed by all mutations at that frequency, with common mutations contributing slightly more variance on a mutation-by-mutation basis than they do as a category (Fig. S3).

Let us now consider the case in which trait and fitness are not perfectly correlated, and let us first consider the case in which the mean absolute effect of a mutation on the trait is linearly related to the strength selection on the mutation ($\tau = 1, \sigma > 0$). So if two mutations differ by twofold in fitness, they will, on average, differ twofold in their average absolute effects on the trait, although mutations can have very large or small effects on the trait depending on the magnitude of σ. Surprisingly, the relationship between the variance in the trait and allele frequency under this model is identical to the case when the trait is fitness itself (Fig. 3); so even if σ is very large and the effects on trait and fitness are very poorly correlated, most of the variance is contributed by rare mutations unless the mean strength of selection is low. This can be seen be by considering $V(x) / V_T$; this expression is independent of σ (as described earlier).

In contrast, if the effects of the mutation on fitness and trait are independent ($\tau = 0, \sigma > 0$), all of the variance in the trait is contributed by common mutations (Fig. 3 and Fig. S4); these are neutral mutations segregating at relatively high frequency. In between these extremes we see a shift from one pattern to the other. For example, if τ is 1/2, such that the mean absolute trait value increases as the square root of the mean strength of selection, then more variance is contributed by high-frequency mutations, than when the trait is fitness or when the SD increases linearly with selection (Fig. 3 and Fig. S4). As the strength of selection increases, the proportion of variance explained by high allele frequency decreases; the proportion is also dependent upon the shape parameter of the DFE, but here we see a different pattern to that observed when the trait is fitness. When $\tau << 1$, the relationship between variance and allele frequency becomes more dependent upon the shape parameter, and as the DFE becomes less leptokurtic (i.e., increasing shape parameter), less of the variance in the trait is contributed by high-frequency mutations. However, when τ is relatively large we see the opposite pattern, and less leptokurtic distributions have slightly more variance being contributed by relatively common mutations.

We have so far considered the proportion of the variance in the trait being contributed by mutations at different allele frequencies under a single unimodal DFE. However, in reality, both fitness and trait are likely to be governed by complex multimodal distributions, composed of several different distributions; indeed there is some evidence of this (21). It seems likely that the overall distribution will, at minimum, be a combination of the distribution of nonsynonymous point and small indel mutations, point and small indel mutations in regulatory sequences, and copy number variant mutations. It is thus of interest to investigate the absolute level of variance contributed by a certain distribution and how this depends upon the shape and mean of the distribution and the

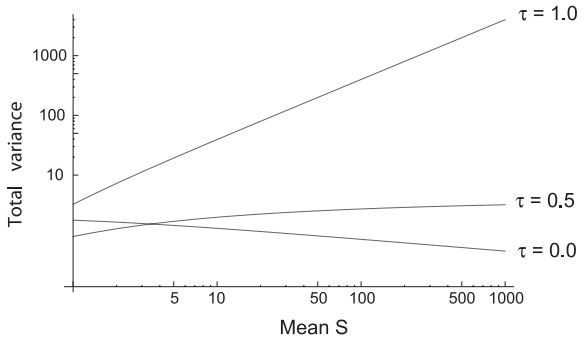


Fig. 4. The total variance, V_T, as a function of the mean strength of selection and the parameter τ. The shape parameter, β, was assumed to be 0.20, and σ was set at 1.

relationship between trait and fitness. A visual inspection of Eq. **7** shows that the total variance is linearly related to the square of σ, but otherwise the relationship between the total variance and any one parameter (e.g., β) depends on the values of the other two parameters (e.g., τ and $\bar{S}$). However, the total variance is largely independent of the shape parameter, β. The variance increases with the mean strength of selection, and the rate at which it increases depends on τ (Fig. 4). When τ is 1, the increase in variance is linear with a slope of 1, such that a 10-fold increase in the mean strength of selection yields a 10-fold increase in the variance. This dependency declines as τ decreases (Fig. 4).

Discussion

We have investigated a model in which mutations that affect a trait of interest also have affects on fitness, even if the effects are very small, either because the trait is a component of fitness or because the mutations have pleiotropic effects on other traits, which are themselves aspects of fitness (15, 16). Under this model, most of the variance in fitness is contributed by mutations of large effect that are very rare in the population, unless the mean strength of selection acting upon the mutations is very low (i.e., $\overline{N_e s} < 10$) or the mean absolute effect of a mutation on the trait (equivalently, the SD of trait effects) increases less than linearly with the strength of selection. Surprisingly, the model is independent of the variation in the trait, σ, so the correlation between the effects of a mutation on the trait and fitness can be very small and still most of the variance in the trait will come from low-frequency mutations if the SD of the trait increases linearly with the strength of selection.

One might argue that the model relating fitness and trait is unrealistic because, if σ is not very large, then the distribution of mutational effects on the trait is bimodal. As an alternative, we investigated a model in which the distribution of mutational effects on the trait is unimodal: $z(S, \varepsilon; \tau) = S^{\tau}\varepsilon$. Examples of this relationship are given in Fig. S5. If we substitute this model into Eq. **3** and simplify, we get an equation that is very similar to Eq. **6**; it differs only in that the $(1 + \sigma^2)$ term vanishes. The model therefore behaves in an almost identical manner in all respects.

The relationship between the variance in a trait and allele frequency has been previously investigated by Pritchard (22), who concluded that slightly deleterious mutations would contribute most to the variance in the trait, and that most of this variance would be from mutations at moderate frequencies in the population. However, the model assumed that the effect of mutations on trait and fitness were uncorrelated, so this is consistent with the results presented here for a τ of 0. There seems little reason to believe that the effects would not be correlated.

A number of assumptions have been made within the model. First, it is assumed that population sizes are stationary. However, population size expansions or contractions can affect the distribution of alleles in a population, and this may influence how the variance is distributed between different allele frequency classes. Furthermore, it has been shown that demography can have implications for the diversity of alleles that cause a disease (23, 24). Second, the model assumes that mutations are being held in a balance among mutation, selection, and genetic drift. However, the maintenance of quantitative genetic variation still remains the subject of debate (25). It is possible that positive selection may increase the frequency of mutations, either directly because some mutations are advantageous, or indirectly through genetic "hitchhiking" (26), and this will have consequences for the variance in the trait. Third, we assume that all traits are equal; however, we are most interested in common diseases, and these may be common simply because some mutations are segregating at high frequency by chance.

The behavior of the model depends critically on the how the effects of a mutation of the trait and fitness are related; if the effects are independent, most of the variance is contributed by mutations segregating at high frequencies in the population, but if the mean absolute effect of a mutation on the trait is linearly related to the strength of selection, most of the variance is contributed by mutations at low frequencies. Unfortunately we know relatively little about how mutations affect fitness and traits, although it is clear from work in Drosophila, mice, and humans that pleiotropy is widespread (14). The one study to specifically address the relationship between fitness and a putatively neutral trait (27) showed that mutations with large effects on two traits, abdominal and sternopleural bristle number, also had negative effects on viability, but the level of resolution was not sufficient to determine whether the absolute effect on the trait increased linearly with the strength of selection (27). Without additional information, there seems little reason to believe that the relationship would not be linear (i.e., τ = 1 in both models). As a consequence it seems likely that most of the variance in the trait will be contributed by mutations segregating at low frequencies unless the mean strength of selection is very low (i.e., $N_e s < 10$).

Unfortunately, we do not know the DFE for all new mutations in humans or any other organism (21). In humans, we have some information about the DFE for amino acid mutations; using the site frequency spectrum and assuming a Γ-distribution for the DFE, it has been estimated that the distribution is highly leptokurtic (i.e., shape parameter of 0.20) with a mean strength of selection $\bar{S}$ of approximately 3,000 (ref. 18; see also refs. 19, 20, 28, and 29). The estimate of the mean strength of election must be treated with some caution because, in analyses of this sort, there is little information about the mean. Nevertheless it is clear that the mean is likely to be greater than 100 as the majority of mutations are inferred to have effects larger than this (18–20, 28, 29) and we therefore expect the vast majority of the variance in fitness to be contributed by mutations below 1/100 (Fig. 1); if the mean is greater than 1,000, most of the variance is contributed by mutations below 1/1,000. This suggests that very little of the variance in fitness is being contributed by nonsynonymous mutations discovered in current resequencing projects, which have sample sizes of approximately 100 chromosomes (30–33). If we want to find the nonsynonymous mutations that might contribute to the variance in fitness or traits, we need to be prepared to sequence at least 1,000, and maybe 10,000, individuals. Furthermore, it may be that we will also need to sequence deeply to see mutations causing disease if they tend to be at all strongly selected.

Although we know something about the DFE of nonsynonymous or amino acid mutations, we know less about the DFE for mutations in noncoding DNA. It has been estimated that approximately 3.6% of the genome is noncoding DNA subject to some level of selective constraint, as opposed to 0.6% of the genome in coding sequences (34). However, several lines of evidence suggest the strength of selection is substantially lower in functional noncoding sequences than in coding DNA. First, a

recent analysis of SNPs estimated the mean strength of selection acting upon new mutations in human conserved nongenic sequences (CNGs) to be such that S is approximately 50 (35), whereas a similar analysis in coding sequences estimated the strength of selection on new nonsynonymous mutations to be approximately 3,000 (18). Second, CNGs show higher levels of divergence and diversity than coding sequences and less skew toward rare polymorphisms (34, 36, 37). And third, CNGs and sequences flanking protein coding sequences are less constrained in hominids than in rodents, which suggests that many mutations in these sequences are weakly selected (36–38).

Although they must be treated with great caution, the best estimates suggest the strength of selection acting upon CNG mutations is nearly 100-fold less than that acting upon nonsynonymous mutations. This would therefore suggest that the vast majority of the variance in fitness is contributed by nonsynonymous mutations, as there are only about 10 times more CNGs than nonsynonymous sites. Furthermore, if the relationship between trait and fitness is the same for regulatory and nonsynonymous mutations, this implies that most of the variance in traits will also be contributed by nonsynonymous mutations. This is because, to a first approximation, the variance contributed by a category of mutations is the total mutation rate for the category multiplied by the mean strength of selection.

We have so far discussed point mutations, but insertions, deletions, and genomic rearrangements are often involved in disease (39). These may contribute substantial variance because, although they are probably less numerous than point mutations, they are also expected to have larger effects on fitness and on traits. Unfortunately, we currently do not know either the rate or effects of indel and genomic rearrangements.

The fact that most of the variance in a trait tends to be contributed by rare mutations, unless the means strength of selection acting upon mutations that affect the trait is very low, has clear implications for genome-wide association studies. These studies have successfully identified many mutations associated with a variety of diseases and traits in humans (4, 7), but disappointingly, these variants explain little of the variance in any of the traits (6, 8). This is perhaps not surprising given the analysis presented here; if mutations have effects on the trait and on fitness, most of the variance in the trait is contributed by mutations that are rare in the population, which genome-wide association studies would never, and may never, be able to detect. Furthermore, current genome-wide association studies may actually be misleading us to some extent. It is possible that the genes that have the largest effect on a trait also tend to be the most strongly selected, so we will never find associations in those genes; instead we may be finding associations in genes in which most mutations are relatively weakly selected, but which also correspondingly have small effects on the trait. Such a bias is consistent with the observation that genes that contain a nonsynonymous SNP associated with disease appear to be subject to weaker natural selection than other genes containing nonsynonymous SNPs, as judged by the rate of nonsynonymous to synonymous substitution (40, 41).

ACKNOWLEDGMENTS. I am very grateful to David Waxman for mathematical help and to Nina Stoletzki, Peter Keightley, David Houle, Mary Clare King, Raju Govindaraju, Peter Visscher, Monty Slatkin, Bill Hill, and two anonymous referees for helpful comments and discussion.

1. The_International_HapMap_Consortium (2005) A haplotype map of the human genome. *Nature* 437:1299–1320.
2. The_International_HapMap_Consortium (2007) A second generationhuman haplotype map of over 3.1 million SNPs. *Nature* 449:851–862.
3. Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 6:95–108.
4. Donnelly P (2008) Progress and challenges in genome-wide association studies in humans. *Nature* 456:728–731.
5. Weiss KM (2008) Tilting at quixotic trait loci (QTL): an evolutionary perspective on genetic causation. *Genetics* 179:1741–1756.
6. Maher B (2008) Personal genomes: The case of the missing heritability. *Nature* 456: 18–21.
7. Hindorff LA, et al. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 106: 9362–9367.
8. Manolio TA, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461:747–753.
9. Gudbjartsson DF, et al. (2008) Many sequence variants affecting diversity of adult human height. *Nat Genet* 40:609–615.
10. Lettre G, et al.; Diabetes Genetics Initiative; FUSION; KORA; Prostate, Lung Colorectal and Ovarian Cancer Screening Trial; Nurses' Health Study; SardiNIA (2008) Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet* 40:584–591.
11. Weedon MN, et al.; Diabetes Genetics Initiative; Wellcome Trust Case Control Consortium; Cambridge GEM Consortium (2008) Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* 40:575–583.
12. Bodmer W, Bonilla C (2008) Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 40:695–701.
13. Goldstein DB (2009) Common genetic variation and human traits. *N Engl J Med* 360: 1696–1698.
14. Flint J, Mackay TF (2009) Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res* 19:723–733.
15. Hill WG, Keightley PD (1988) Interrelations of mutation, population size, artificial and natural selection. Proceedings of the second international conference on quantitative genetics, eds Weir BS, Eisen EJ, Goodman MM, Namkoong G (Sinauer, Sunderland, MA), pp 57–70.
16. Keightley PD, Hill WG (1990) Variation maintained in quantitative traits with mutation-selection balance: pleiotropic side-effects on fitness traits. *Proc R Soc Lond B* 242:95–100.
17. Wright S (1938) The distribution of gene frequencies under irreversible mutation. *Proc Natl Acad Sci USA* 24:253–259.
18. Boyko AR, et al. (2008) Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet* 4:e1000083.
19. Eyre-Walker A, Woolfit M, Phelps T (2006) The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* 173:891–900.
20. Keightley PD, Eyre-Walker A (2007) Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymrophism frequencies. *Genetics* 177:2251–2261.
21. Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. *Nat Rev Genet* 8:610–618.
22. Pritchard JK (2001) Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 69:124–137.
23. Peng B, Kimmel M (2007) Simulations provide support for the common disease-common variant hypothesis. *Genetics* 175:763–776.
24. Reich DE, Lander ES (2001) On the allelic spectrum of human disease. *Trends Genet* 17:502–510.
25. Johnson T, Barton NH (2005) Theoretical models of selection and mutation on quantitative traits. *Phil Trans R Soc B* 360:1411–1425.
26. Williamson SH, et al. (2007) Localizing recent adaptive evolution in the human genome. *PLoS Genet* 3:e90.
27. Lyman RF, Lawrence F, Nuzhdin SV, Mackay TF (1996) Effects of single P-element insertions on bristle number and viability in *Drosophila melanogaster*. *Genetics* 143:277–292.
28. Yampolsky LY, Kondrashov FA, Kondrashov AS (2005) Distribution of the strength of selection against amino acid replacements in human proteins. *Hum Mol Genet* 14: 3191–3201.
29. Kryukov GV, Pennacchio LA, Sunyaev SR (2007) Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet* 80:727–739.
30. Akey JM, et al. (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biol* 2:e286.
31. Bustamante CD, et al. (2005) Natural selection on protein-coding genes in the human genome. *Nature* 437:1153–1157.
32. Hinds DA, et al. (2005) Whole-genome patterns of common DNA variation in three human populations. *Science* 307:1072–1079.
33. Livingston RJ, et al. (2004) Pattern of sequence variation across 213 environmental response genes. *Genome Res* 14 (10A, 10A)1821–1831.
34. Asthana S, et al. (2007) Widely distributed noncoding purifying selection in the human genome. *Proc Natl Acad Sci USA* 104:12410–12415.
35. Torgerson DG, et al. (2009) Evolutionary processes acting on candidate cis-regulatory regions in humans inferred from patterns of polymorphism and divergence. *PLoS Genet* 5:e1000592.
36. Kryukov GV, Schmidt S, Sunyaev S (2005) Small fitness effect of mutations in highly conserved non-coding regions. *Hum Mol Genet* 14:2221–2229.
37. Keightley PD, Kryukov GV, Sunyaev S, Halligan DL, Gaffney DJ (2005) Evolutionary constraints in conserved nongenic sequences of mammals. *Genome Res* 15:1373–1378.
38. Keightley PD, Lercher MJ, Eyre-Walker A (2005) Evidence for widespread degradation of gene control regions in hominid genomes. *PLoS Biol* 3:e42.
39. Lupski JR (2009) Genomic disorders ten years on. *Genome Med* 1:42.
40. Thomas PD, Kejariwal A (2004) Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. *Proc Natl Acad Sci USA* 101:15398–15403.
41. Blekhman R, et al. (2008) Natural selection on genes that underlie human disease susceptibility. *Curr Biol* 18:883–889.

Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease

Andrew P. Feinberg[1] and Rafael A. Irizarry

Center for Epigenetics, Johns Hopkins University, Baltimore, MD 21205

Edited by Peter T. Ellison, Harvard University, Cambridge, MA, and approved November 17, 2009 (received for review September 2, 2009)

Neo-Darwinian evolutionary theory is based on exquisite selection of phenotypes caused by small genetic variations, which is the basis of quantitative trait contribution to phenotype and disease. Epigenetics is the study of nonsequence-based changes, such as DNA methylation, heritable during cell division. Previous attempts to incorporate epigenetics into evolutionary thinking have focused on Lamarckian inheritance, that is, environmentally directed epigenetic changes. Here, we propose a new non-Lamarckian theory for a role of epigenetics in evolution. We suggest that genetic variants that do not change the mean phenotype could change the variability of phenotype; and this could be mediated epigenetically. This inherited stochastic variation model would provide a mechanism to explain an epigenetic role of developmental biology in selectable phenotypic variation, as well as the largely unexplained heritable genetic variation underlying common complex disease. We provide two experimental results as proof of principle. The first result is direct evidence for stochastic epigenetic variation, identifying highly variably DNA-methylated regions in mouse and human liver and mouse brain, associated with development and morphogenesis. The second is a heritable genetic mechanism for variable methylation, namely the loss or gain of CpG dinucleotides over evolutionary time. Finally, we model genetically inherited stochastic variation in evolution, showing that it provides a powerful mechanism for evolutionary adaptation in changing environments that can be mediated epigenetically. These data suggest that genetically inherited propensity to phenotypic variability, even with no change in the mean phenotype, substantially increases fitness while increasing the disease susceptibility of a population with a changing environment.

DNA methylation | epigenetics | evolution | stochastic variation

A key tenet of *Origin of Species* argues that phenotype is the result of many discrete traits that are individually and exquisitely selected, to quote Darwin, "detecting the smallest grain in the balance of fitness," which has been described as Newtonian in its dependence on static forces acting in consistent ways (1). This concept is the basis for quantitative trait loci proposed by R. A. Fisher (2). This concept has led to the modern basis of population genetics that continuous variation exists within a population, yet selection is on individuals, which has led to models of balancing or purifying selection at the extremes of phenotype (1). The classic model also has significant limitations in explaining common human disease; common variants can explain only a small fraction of a given disease phenotype, even the most well understood, such as adult-onset diabetes and height (3).

Epigenetics, the study of nonsequence-based changes in DNA and associated proteins, was first suggested by Jablonka to play a role in evolution through Lamarckian inheritance, that is, direct modification of the genome by the environment, which is then transmitted transgenerationally (4). Two examples are commonly cited: changes in coat color caused by dietary modifications of DNA methylation of the *agouti* gene in mice (5, 6) and methylation of the *axin-fused* allele in kinked tail mice (6, 7). Both of these examples involve methylation of a retrotransposon LTR sequence, and thus fit into various genetic exceptions to classical Darwinian thinking, including anticipation due to trinucleotide repeat expansion and lateral gene transfer in the evolution of influenza strains (8). But they have not been shown to be general mechanisms for either speciation or developmental differences across species, so-called "evo-devo," or for canalization, a term coined by Waddington to refer to a mechanism by which environmental perturbations during development are corrected by the genetic program, leading to a consistent developmental plan (9). Indeed, canalization remains a "black box," as noted by West-Eberhard (8). Others have discussed the potential role for Lamarckian inheritance in disease; for example, Slatkin proposed a model of transgenerational epigenetic Lamarckian inheritance and noted that such modifications must persist for many generations to contribute substantially to average risk (10), which has implications for public health management (11). Although not disputing an important contribution of Lamarkcian inheritance, here we propose an alternative view in which *genetic* modification could provide stochastic phenotypic variation favored by selection in changing environments, and also provide an alternative non-Lamarckian role for epigenetics in evolution.

A New Advance Over Darwinism: Stochastic Variation, Not Lamarckian Inheritance

It has occurred to us that increased variability with a given genotype might itself increase fitness. This could arise by genetic variants that do not change the mean phenotype but do change the variability of phenotype. A natural mechanism to use to consider such a model is epigenetic plasticity during development, for example, varying DNA methylation patterns. This idea differs from Lamarckian inheritance, in that in our model the genetic change is inherited, and this change leads to increased epigenetic variation. It also differs from the likely role of epigenetics in modifying mutation rate, both through C to T transition due to deamination of methylcytosine and through modified rates of chromosomal rearrangement (12, 13). As a proof of principle, we revisited previously generated data sets (14) of genome-scale analysis of DNA methylation in human and mouse tissues and explored them in two new ways. First, we investigated whether there were regions of variable methylation *across individuals* for a given tissue type. Then we explored whether tissue-specific differentially methylated regions (T-DMRs) differed across species and whether the underlying DNA sequence could account for these differences.

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Evolution in Health and Medicine," held April 2–3, 2009, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS web site at www.nasonline.org/Sackler_Evolution_Health_Medicine.

Author contributions: A.P.F. designed research; A.P.F. and R.A.I. performed research; R.A.I. contributed new reagents/analytic tools; A.P.F. and R.A.I. analyzed data; and A.P.F. and R.A.I. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: afeinberg@jhu.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0906183107/DCSupplemental.

Variably Methylated Regions Across Individuals

To assess the degree of intrinsic variability in DNA methylation of a given tissue, we set out to identify the location of the most highly variable regions of DNA methylation in mouse liver from four individuals. We chose this specific tissue because it is relatively homogeneous. We examined newborns in whom polyploidy is minimal, although copy number would not be expected to affect DNA methylation, because our method controls for copy number (15). Environmental effects were minimized by examining inbred mice (indeed, littermates from the same cage). Surprisingly, many loci throughout the genome showed striking variations in DNA methylation, which we term *variably methylated regions* (VMRs). Surprisingly, these VMRs were significantly enriched in the vicinity of genes with Gene Ontogeny (GO) functional categories for development and morphogenesis (Table 1) when using either all genes for comparison or all regions present on the CHARM array, indicating that enrichment is not explained solely by high CpG content, because the array itself is designed to assay high-CpG regions. Examples of developmental genes with VMRs—*Bmp7*, involved in early embryogenic programming and bone induction, *Pou3f2*, involved in neurogenesis and stem cell reprogramming, and *Ntrk3*, involved in body position sensing—are shown in Fig. 1.

Furthermore, the VMRs were associated with a functional property: expression. As shown in Fig. 2, VMRs within 500 bp of a transcriptional start site (TSS) exhibited a stronger association between gene expression variability and methylation variability.

We then examined human liver for the presence of VMRs. Similar to our mouse results, we found significant variability. Where the VMRs were near genes, as in the mouse, there was a strong enrichment in the vicinity of genes with GO functional categories for development and morphogenesis when controlled for the mouse CHARM array (Table 2).

We then performed a similar analysis on mouse brain. The results were even more striking. For example, Fig. 3 shows two examples of VMRs: *Bmpr2*, the receptor for the morphogenetic BMP protein, and *Irs1*, a key mediator of insulin-driven differentiation. Our findings indicate that VMRs are present across tissues and species, are enriched in development-related genes, and are related to phenotype, at least at the level of expression of the proximate gene.

Also note that VMRs often are located near tissue-varying DMRs (T-DMRs), suggesting a mechanism by which they might evolve into each other over time. This is illustrated in Fig. 4 for mouse *Ptp4a1*, a protein tyrosine phosphatase involved in maintaining differentiated epithelial tissues, and for human *FOXD2*, a forkhead transcription factor involved in embryogenesis.

Tissue-Specific Differentially Methylated Regions Across Species

Next, we were interested in whether changes in differential methylation across species (mouse and human) could be traced back to an underlying genetic basis. To address this question, we focused on

Table 1. Enrichment scores of GO categories of genes in the vicinity of VMRs in mouse liver

GOBPID	*P* value	Odds ratio	Expected count	Count	Size	Term
GO:0048699	2.8E-05	2.0	26.9	49	384	Generation of neurons
GO:0009880	8.5E-05	4.9	2.8	11	41	Embryonic pattern specification
GO:0030030	0.00033	2.0	19.1	35	272	Cell projection organization
GO:0021517	0.00034	8.8	1.0	6	15	Ventral spinal cord development
GO:0035107	0.00041	2.9	6.2	16	89	Appendage morphogenesis
GO:0048666	0.00046	2.0	17.2	32	245	Neuron development
GO:0032990	0.00050	2.2	12.3	25	175	Cell part morphogenesis
GO:0009887	0.00052	1.6	35.9	56	512	Organ morphogenesis
GO:0021515	0.00055	6.2	1.5	7	22	Cell differentiation in spinal cord
GO:0048812	0.00065	2.2	11.8	24	168	Neurite morphogenesis
GO:0060173	0.00068	2.7	6.5	16	93	Limb development
GO:0007411	0.00075	2.8	5.9	15	85	Axon guidance
GO:0006270	0.00088	9.5	0.8	5	12	DNA replication initiation
GO:0001708	0.0010	4.6	2.1	8	31	Cell fate specification
GO:0000904	0.0014	2.0	13.2	25	188	Cell morphogenesis involved in differentiation
GO:0048869	0.0017	1.3	86.5	112	1,231	Cellular developmental process
GO:0007420	0.0020	1.9	15.0	27	214	Brain development
GO:0048663	0.0021	3.6	2.9	9	42	Neuron fate commitment
GO:0042415	0.0031	19.9	0.3	3	5	Norepinephrine metabolic process
GO:0009954	0.0033	4.9	1.5	6	22	Proximal/distal pattern formation
GO:0042472	0.0033	3.1	3.7	10	53	Inner ear morphogenesis
GO:0048598	0.0035	1.7	19.4	32	277	Embryonic morphogenesis
GO:0007417	0.0050	2.9	3.9	10	57	Central nervous system development
GO:0021846	0.0053	7.6	0.7	4	11	Cell proliferation in forebrain
GO:0021520	0.0058	13.2	0.4	3	6	Spinal cord motor neuron cell fate specification
GO:0021521	0.0058	13.2	0.4	3	6	Ventral spinal cord interneuron specification
GO:0045773	0.0058	13.2	0.4	3	6	Positive regulation of axon extension
GO:0021536	0.0065	4.2	1.7	6	25	Diencephalon development
GO:0035116	0.0067	5.1	1.2	5	18	Embryonic hindlimb morphogenesis
GO:0007275	0.0076	1.2	124.8	149	1,776	Multicellular organismal development
GO:0007423	0.0076	1.8	13.4	23	191	Sensory organ development
GO:0030326	0.0090	2.6	4.2	10	61	Embryonic limb morphogenesis
GO:0035270	0.0095	2.7	3.6	9	52	Endocrine system development
GO:0006268	0.0097	9.9	0.49	3	7	DNA unwinding during replication
GO:0021546	0.0097	9.9	0.49	3	7	Rhombomere development
GO:0048856	0.0099	1.2	106.1	128	1,538	Anatomical structure development

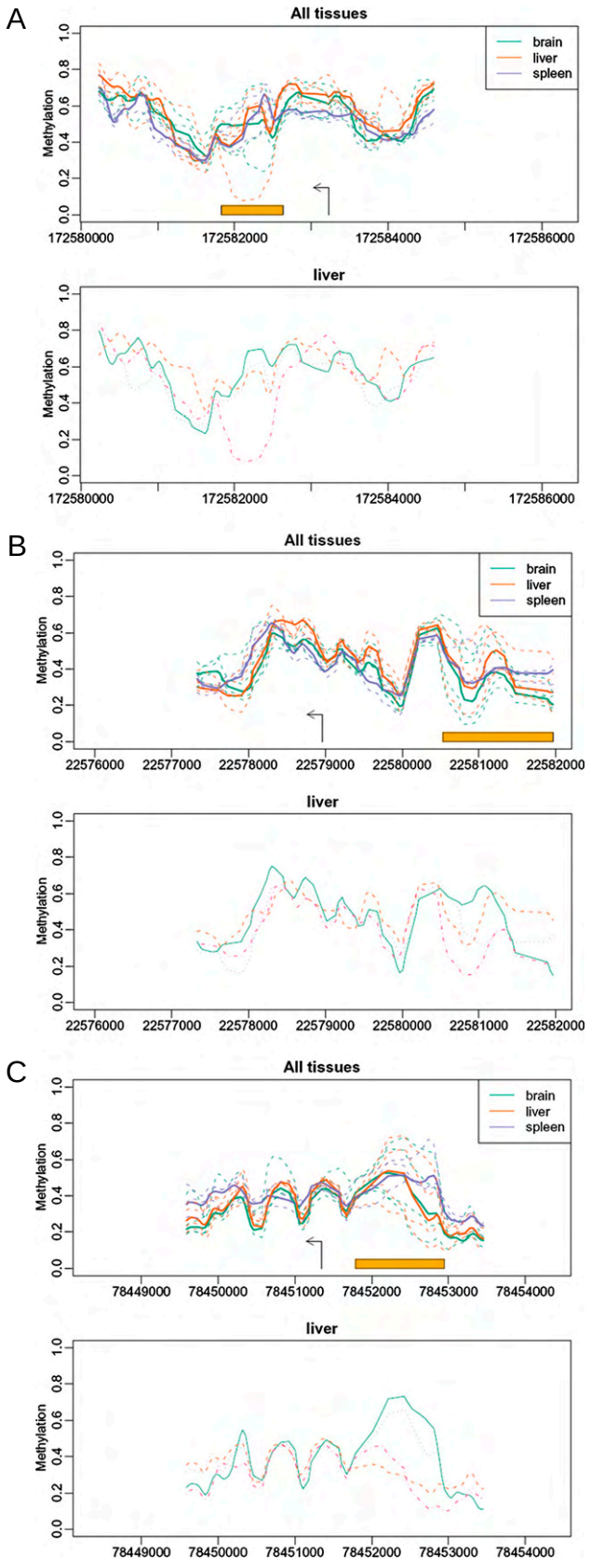


Fig. 1. Examples of developmental genes with VMRs in livers from isogenic mice raised in the same environment. Shown are *Bmp7* (*A*), *Pou3f2* (*B*), and *Ntrk3* (*C*), involved in early embryogenic programming and bone induction, neurogenesis and stem cell reprogramming, and body position sensing, respectively. In each paired plot, the top panel shows estimated methylation levels from various biological replicates from three different tissues: brain, liver, and spleen (dashed lines). The thicker solid lines represent the average curves for each tissue. The orange bar denotes the region in which our statistical method detected a VMR. The bottom panel highlights the liver. Only the four liver curves are shown. The different line types and colors represent the four individual mice.

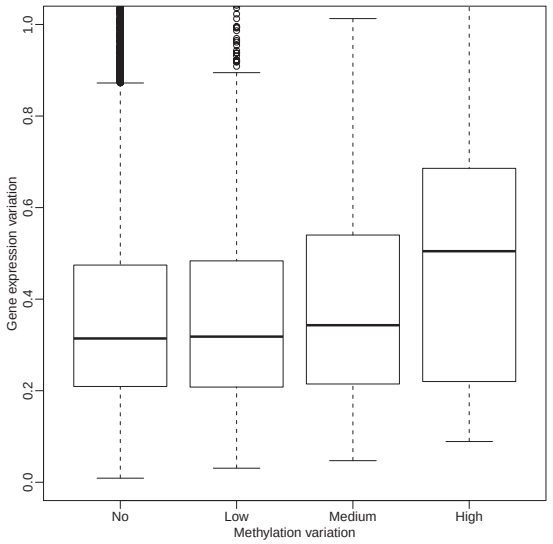


Fig. 2. VMRs are associated with variability in gene expression of nearby genes. The human liver VMRs detected with our statistical algorithm were divided into three types: low variation (lowest 70%), high variation (highest 5%), and medium variation (the remainder). The VMRs within 500 bases from a gene's transcription start site were associated with that gene. The expression measurements were obtained for the same human livers, and the SD across subjects was used to quantify variability. These boxplots show the distribution of this variability stratified by VMR variability. The first boxplot represents genes not associated with a VMR.

T-DMRs, given the wealth of data gathered in previous studies and their relevance to human diseases, such as cancer. Previously we reported that DMRs that distinguish colorectal cancer from normal colonic mucosa (C-DMRs) are enriched for T-DMRs, and this finding was validated in a large independent set of samples. In many cases, the loss of differential methylation in one species was related to an underlying loss of CpGs at the corresponding CpG island or nearby CpG island shore (14). A typical example of an evolutionary change in differential methylation involved *LHX1*, a transcriptional regulator essential for vertebrate head organization and mesoderm organization, (shown in Fig. 5). Note the T-DMR in human that is not in mouse on the left of the TSS. The human has gained CpGs at a CpG island shore (with the island shown in orange tick marks in the bottom panel). In contrast, both species have a moderate CpG count to the right of the TSS, and both have DMRs in this region. This is an example of how a genetic variation (i.e., gain of CpGs) allows for development-relevant tissue-specific differences in a highly conserved gene. Thus, differential methylation that itself differs across species may be due to underlying sequence variation at the site of these DMRs. Additional examples of this are available at rafalab.jhsph.edu/evometh.pdf.

Increased Stochastic Variation Would Increase Fitness in a Varying Environment

To model the role of epigenetic variation in natural selection, we performed three simulations based on a single quantitative phenotype that contributes to fitness, arbitrarily called *Y*. We assumed

Table 2. Enrichment scores of GO categories of genes in the vicinity of VMRs in human liver

GOBPID	*P* value	Odds ratio	ExpCount	Count	Size	Term
GO:0009790	1.8E-05	1.8	43.1	70	320	Embryonic development
GO:0019222	2.3E-05	1.3	319.5	379	2,372	Regulation of metabolic process
GO:0006355	4.0E-05	1.3	239.6	292	1,779	Regulation of transcription, DNA-dependent
GO:0032774	5.0E-05	1.3	246.8	299	1,832	RNA biosynthetic process
GO:0009887	5.3E-05	1.6	54.1	82	402	Organ morphogenesis
GO:0048704	8.4E-05	4.0	5.2	15	39	Embryonic skeletal system morphogenesis
GO:0001501	8.5E-05	1.9	27.8	48	207	Skeletal system development
GO:0051093	8.5E-05	1.7	43.5	68	323	Negative regulation of developmental process
GO:0016339	0.00012	7.2	2.2	9	17	Calcium-dependent cell-cell adhesion
GO:0009952	0.00013	2.5	12.3	26	92	Anterior/posterior pattern formation
GO:0048518	0.00017	1.3	133.2	171	989	Positive regulation of biological process
GO:0019219	0.00025	1.2	269.0	317	1,997	Regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
GO:0007389	0.00028	2.0	22.3	39	166	Pattern specification process
GO:0010468	0.00029	1.2	272.3	320	2,029	Regulation of gene expression
GO:0043009	0.00032	2.1	18.7	34	140	Chordate embryonic development
GO:0031326	0.00037	1.2	279.8	327	2,077	Regulation of cellular biosynthetic process
GO:0006350	0.00038	1.2	267.6	314	1,987	Transcription
GO:0001824	0.00040	4.9	3.0	10	23	Blastocyst development
GO:0010556	0.00048	1.2	271.3	317	2,014	Regulation of macromolecule biosynthetic process
GO:0050678	0.00051	3.6	4.8	13	36	Regulation of epithelial cell proliferation
GO:0048863	0.00064	7.5	1.7	7	13	Stem cell differentiation
GO:0019827	0.00076	9.6	1.3	6	10	Stem cell maintenance
GO:0007399	0.00080	1.4	84.5	112	631	Nervous system development
GO:0000165	0.00089	2.0	16.0	29	119	MAPKKK cascade
GO:0043284	0.0011	1.2	327.0	372	2,428	Biopolymer biosynthetic process
GO:0043583	0.0014	2.7	7.2	16	54	Ear development
GO:0042472	0.0016	3.5	4.1	11	31	Inner ear morphogenesis
GO:0048468	0.0016	1.4	62.6	85	465	Cell development
GO:0007420	0.0017	1.8	21.2	35	158	Brain development
GO:0034645	0.0017	1.2	346.4	390	2,572	Cellular macromolecule biosynthetic process
GO:0001656	0.0018	3.8	3.6	10	27	Metanephros development
GO:0035239	0.0018	2.6	7.4	16	55	Tube morphogenesis
GO:0043066	0.0019	1.7	26.9	42	200	Negative regulation of apoptosis
GO:0045747	0.002	Inf	0.4	3	3	Positive regulation of Notch signaling pathway
GO:0045597	0.0027	1.9	15.6	27	116	Positive regulation of cell differentiation
GO:0043067	0.0030	1.4	58.7	79	436	Regulation of programmed cell death
GO:0032501	0.0037	1.2	297.8	336	2,211	Multicellular organismal process
GO:0007156	0.0039	1.9	13.7	24	102	Homophilic cell adhesion
GO:0021546	0.0039	12.8	0.8	4	6	Rhombomere development
GO:0065007	0.0040	1.1	633.7	677	4,704	Biological regulation
GO:0045884	0.0043	5.5	1.7	6	13	Regulation of survival gene product expression
GO:0048523	0.0043	1.2	129.7	157	963	Negative regulation of cellular process
GO:0021915	0.0044	3.2	4.0	10	30	Neural tube development
GO:0001525	0.0046	1.9	14.6	25	109	Angiogenesis
GO:0048856	0.0048	1.2	202.8	235	1,525	Anatomical structure development
GO:0048646	0.0049	2.2	8.8	17	66	Anatomical structure formation
GO:0000122	0.0055	1.7	21.1	33	157	Negative regulation of transcription from RNA polymerase II promoter
GO:0045595	0.0055	1.8	16.4	27	123	Regulation of cell differentiation
GO:0007507	0.0063	1.8	16.5	27	123	Heart development
GO:0000070	0.0065	4.1	2.4	7	18	Mitotic sister chromatid segregation
GO:0021545	0.0067	4.8	1.8	6	14	Cranial nerve development
GO:0006366	0.0070	1.3	59.7	78	448	Transcription from RNA polymerase II promoter
GO:0048869	0.0073	1.2	149.1	176	1,107	Cellular developmental process
GO:0008284	0.0076	1.5	28.1	41	209	Positive regulation of cell proliferation
GO:0001708	0.0079	3.4	3.0	8	23	Cell fate specification
GO:0007020	0.0081	8.5	0.9	4	7	Microtubule nucleation
GO:0001655	0.0083	2.2	7.8	15	58	Urogenital system development
GO:0001666	0.0083	2.2	7.8	15	58	Response to hypoxia
GO:0000281	0.0087	19.3	0.5	3	4	Cytokinesis after mitosis
GO:0009058	0.0088	1.1	405.0	442	3,007	Biosynthetic process
GO:0035270	0.0093	2.5	5.7	12	43	Endocrine system development
GO:0001649	0.0094	2.6	5.1	11	38	Osteoblast differentiation
GO:0048699	0.0096	1.4	40.4	55	300	Generation of neurons
GO:0007215	0.0099	4.2	2.0	6	15	Glutamate signaling pathway

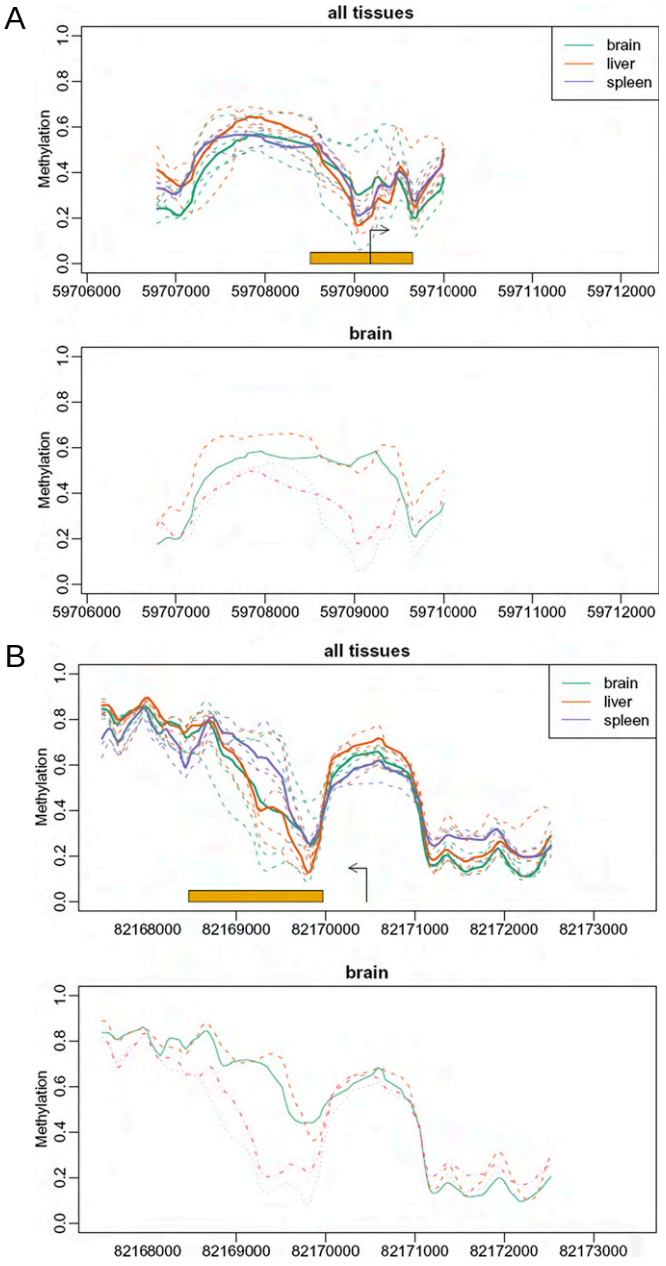


Fig. 3. Examples of developmental genes with VMRs in brains from isogenic mice raised in the same environment. Shown are *Bmpr2*, the receptor for the morphogenetic BMP protein (*A*), and *Irs1*, a key mediator of insulin-driven differentiation (*B*). Labeling is as in Fig. 1.

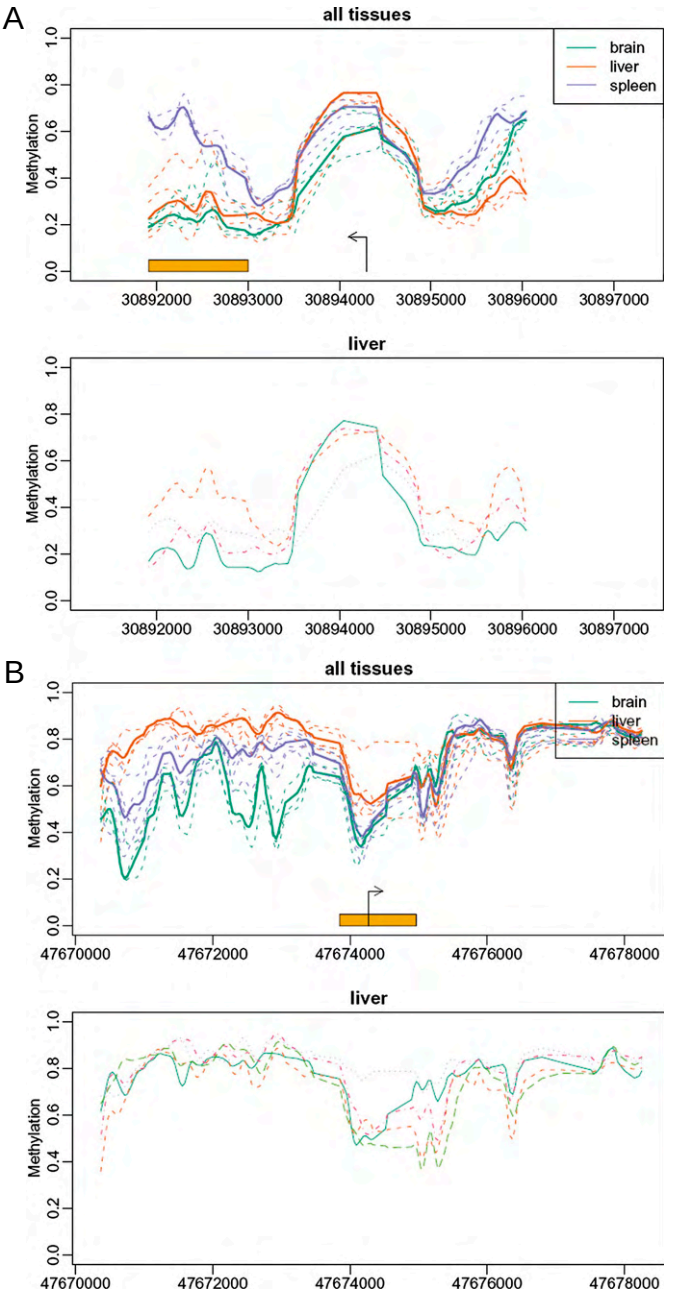


Fig. 4. VMRs are often located near T-DMRs. Shown are mouse *Ptp4a1*, a protein tyrosine phosphatase involved in maintaining differentiated epithelial tissues (*A*), and human *FOXD2*, a forkhead transcription factor involved in embryogenesis (*B*). Labeling is as in Fig. 1. In (*A*), the VMR and T-DMR coincide, whereas in (*B*), they are adjacent.

that mutations of eight genomic locations affected the expected value of Y, with four mutations increasing Y and four decreasing Y. For two of the simulations (simulations 1 and 2), we included a novel stochastic element controlled by eight mutations, four of which increased the *variance* of Y across the population given an identical genotype and four of which decreased this variance. Mathematical details are given in *Materials and Methods*.

In simulation 1, we emulated natural selection in a fixed environment favoring positive Y but including a novel stochastic epigenetic element, such that eight mutations affect the average of Y and eight mutations affect the variance of Y. As expected, this simulation favored the genotype with the largest expected value and the smallest variance (Fig. 6*A*). Simulation 2 was the same as simulation 1, but in this case we allowed a *changing* environment across generations that favor at times large Y and at times small Y. In this simulation, the most highly variable genotype was selected for and dominated by the 1,000th generation (Fig. 6*A*). In simulation 3, we did not permit the variance to change. In this case, 72% of the iterations resulted in extinction before the 1,000th generation. This occurred because the genotype selected in one environment was not fit for the environment change after a dramatic environmental change. In contrast, when variance was allowed to change (simulation 2), extinction never occurred.

In addition, we also emulated genome-wide association studies (GWAS) for Y. The individuals that did not survive were considered diseased, and the survivors were considered controls. An

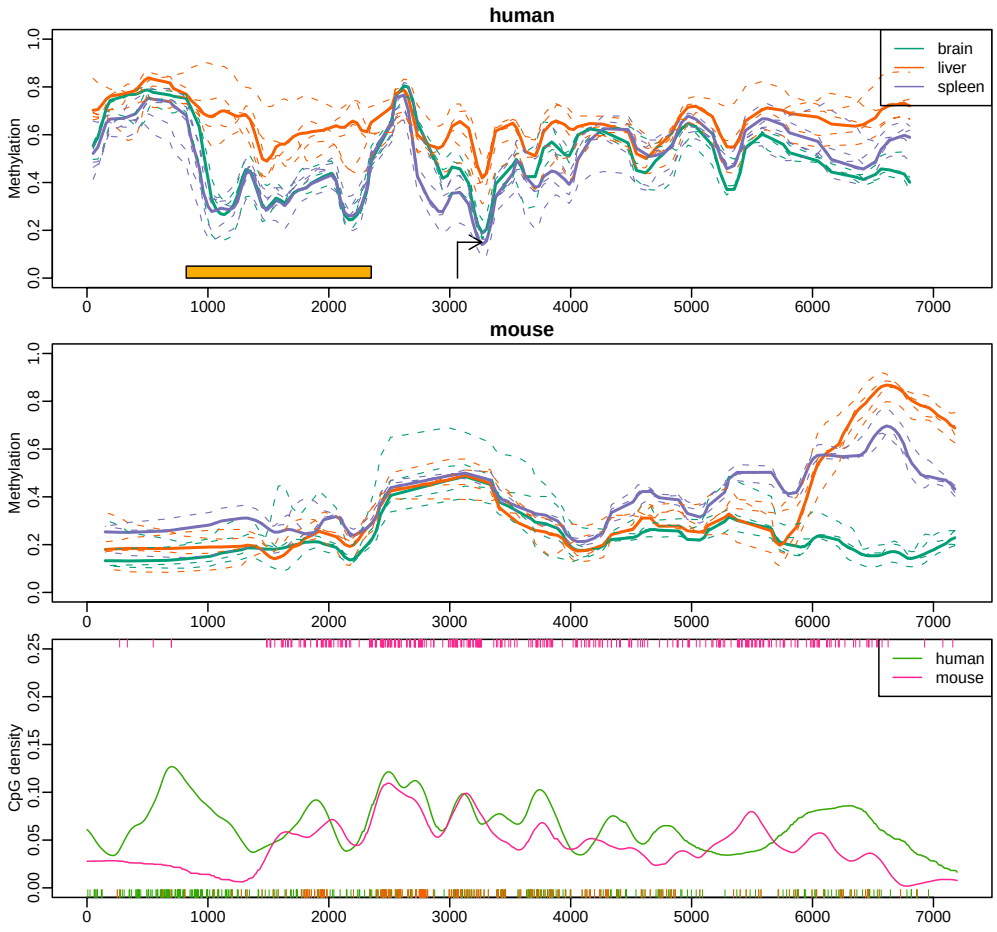


Fig. 5. An underlying genetic basis for species differences in DMRs. A 7,500-bp human region was mapped to the mouse genome. The *x*-axis shows an index so that mapped bases are on top of one another. (*Top*) Methylation profiles for each human sample. As in Fig. 1, the dashed lines represent the individuals, and the solid lines represent the tissue averages. (*Middle*) The same plot for mouse. (*Bottom*) Ticks representing CpG locations for human and mouse. The orange ticks represent CpGs that were conserved. The curves represent CpG counts in a moving window of size 200 bases. Note that the lack of CpGs in the mouse at the beginning of the regions is associated with a difference in methylation patterns between species. Shown is *LHX1*, a transcriptional regulator essential for vertebrate head organization and mesoderm organization. Note the DMR in human that is not in mouse on the left of the TSS. The human has gained CpGs at a CpG island shore (orange tick marks). In contrast, both species have a moderate CpG count to the right of the TSS, and both have DMRs in this region.

interesting finding was that the odds ratios for association between the genes known to affect fitness with disease hovered around 1.10 (Fig. 6*B*). The reason for this is because many of the diseased individuals were unfit only because of the affect of SNPs on variation, not because of the usual SNP-defined genetic change that directly affects function. This is simply a result of the low heritability that results from a large variance. Thus, the results of the epigenetic variation model are in agreement with results from current GWAS studies that explain very little attributable risk of disease.

Discussion

Here we have proposed a model in which increased variability with a given genotype might increase fitness not by changing mean phenotype, but rather by changing the variability of phenotype with a given genotype. We also have provided a possible mechanism by which such enhanced variability could be genetically inherited and lead to increased stochastic epigenetic variation during development. Note that the genomic loci for such variation would be well defined in our model; we have provided examples of these loci. Although these loci do not represent the primary engine of development, they do provide *plasticity* in the developmental program by virtue of the stochastic variation that they impart through the genes in their proximity.

Our model differs from that of a transgenerational epigenetic effect on phenotypic variation and disease risk (16), in that in our model, the genetic variant is inherited and contributes to enhanced phenotypic variation, which can be mediated epigenetically in each generation. It also differs from a hypermutable genetic-switching model, in which the genotype itself changes from generation to generation, increasing phenotypic plasticity (17).

Our model provides a mechanism for developmental plasticity and evolutionary adaptation to a fluctuating environment. Although the model is general and does not necessitate epigenetic variation, we have demonstrated the existence of VMRs that affect phenotype (i.e., gene expression) in isogenic mice raised in an identical environment, and have shown that similar VMRs exist in humans as well. We also have reported a potential genetic mechanism for differences in tissue-specific methylation across species—namely, the gain or loss of a CpG island or the associated shore. The localization near a specific gene would provide specificity of the effect of variation, but the mechanism for variation could entail the relationship to tissue-specific promoters, transcription factor binding sites, population variation in CpG density in these regions, or a combination of such factors. Distinguishing among these possibilities will require further experimentation.

Nonetheless, our model makes a specific prediction: that heritable genetic variation affects stochastic phenotypic variation. Thus, one should be able to identify SNPs that contribute to variance but not mean phenotype. Such SNPs do not *necessitate* an epigenetic mechanism for their influence, but at least some of them would be predicted to be in linkage disequilibrium to VMRs, such as those described above. The VMRs provide a possible mechanism for phenotypic variation in a given genetic background, and we have direct evidence for this at least at the level of expression of the proximate gene. Waddington (9) also proposed that in a given environment, phenotypes eventually become genetically assimilated, and that the sequence differences in CpG islands and shores could provide a mechanism for both gain and loss in evolution of developmental variation mediated by DNA methylation.

Our model and our data differ from Lamarckianism, which argues that the environment modifies the genome. While not

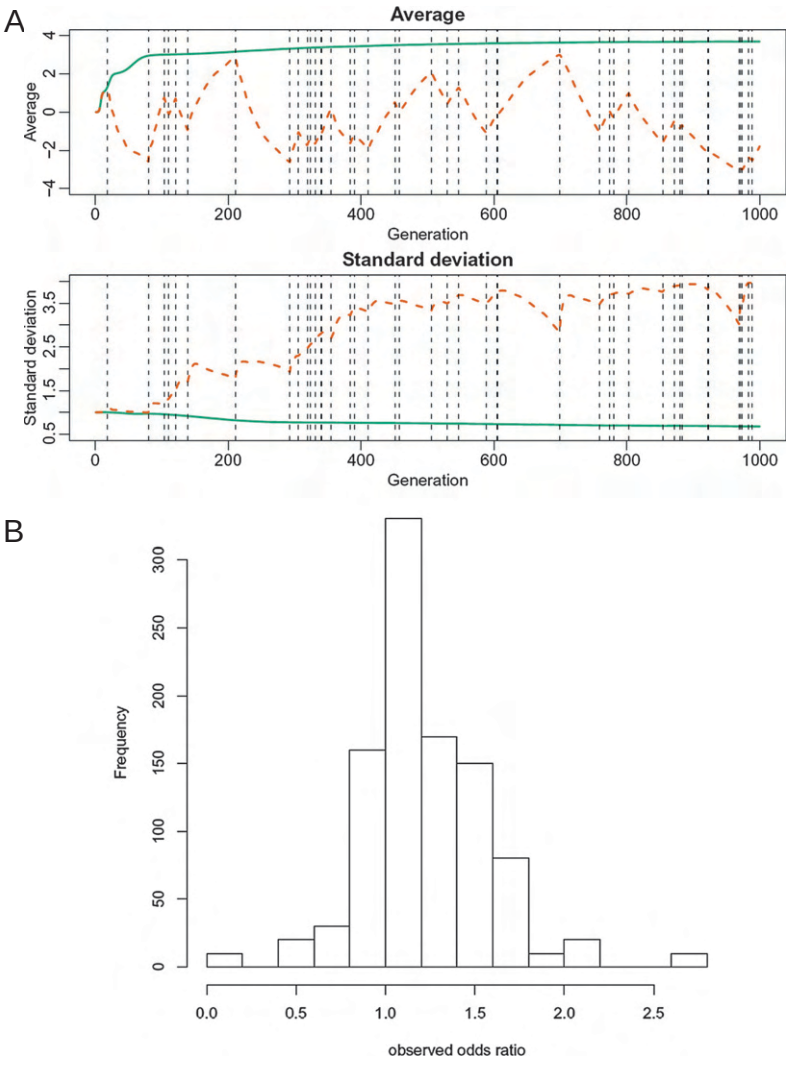


Fig. 6. Results of simulations demonstrating that increased stochastic variation in the epigenome would increase fitness in a varying environment. (*A*) Simulations of natural selection. For each simulation, we computed the population average and SD of the phenotype as a function of generation. Two simulations are shown: simulation 1, natural selection in a fixed environment favoring positive *Y* but including a novel stochastic epigenetic element, such that eight mutations affect average *Y* and eight mutations affect variance of *Y*, and simulation 2, similar to simulation 1 but in this case allowing a changing environment across generations that favor at times positive *Y* and at times negative *Y*. The top panel shows the average (across all iterations) population average of *Y* as a function of generation for simulation 1 (green) and simulation 2 (orange). The dashed vertical lines indicate the generations at which the environment was changed in simulation 2. The bottom panel shows the average (across all iterations) population standard deviation of *Y*. Note that with a changing environment, the average *Y* fluctuates around a common point, but the SD of *Y* increases consistently. (*B*) Emulation of GWAS analysis based on simulation 2 (varying variance of *Y*). Observed odds ratios are for SNPs that change the mean phenotype.

disputing the existence of such inheritance, here we propose a *genetic* mechanism that may underlie this ability to vary epigenetically. We also depart from the neo-Darwinian and classical population genetics principle that heritable quantitative phenotypic variation is due entirely to the additive effect of individual trait loci. Here the heritable component is in part be a propensity to variation itself, adding an element of randomness to the phenotypic outcome. Thus, selection would be determined in part by the ability to vary around a setpoint, rather than by the setpoint itself. This notion is consistent with the idea of "order for free" of Stuart Kaufmann (18). Although Kaufman did not anticipate a role for epigenetics in evolution, inherent epigenetic variation itself will create new possibilities for ordered function—a question that now might be addressable mathematically, given our identification of a possible measurable substrate for this variation, namely DNA methylation. Of course, we do not know how much variation can be tolerated; at some point of increased variation, the individual species "identity" might deteriorate.

Our model also may help explain observations in the evolutionary and epigenetic literature that have seemed paradoxical. In epigenetics, the apparent high degree of instability in the fidelity of epigenetic marks is puzzling. For example, cell lines propagated clonally show a high frequency of random monoallelic expression (19). This epigenetic instability may have been first described while observing individual cancer cells (20), and data show clear epigenetic differences between identical twins (21). In evolutionary biology, social insects show environment-mediated phenotypic differences in social castes, and the distribution of those differences can be selected for (22), leading those authors to speculate that an epigenetic mechanism might be involved (23); the bee would be an outstanding model for testing these ideas. Finally, substantial variations in phenotype of crayfish from an identical genotype have been reported (24). The authors also observed variable global DNA methylation, but as a phenotype, not a mechanism, and found no relationship between methylation and phenotype; they did not examine individual genes (24). We suggest that the mechanism for phenotypic variation is epigenetic, and that increased variation would promote fitness.

Finally, not only variable phenotypes in normal tissue, but also variable disease phenotypes, might be obtained through inherent epigenetic variation. This is because a genetic variant providing a higher variance in phenotype also will increase the tails at both ends of the phenotype; that is, the same variant increasing fitness in one environment will increase the risk of decreasing fitness in a different environment. In support of this idea, we analyzed DMRs that are present in human but not in mouse, and found that many of these genes are associated with human disorders of development as well as common complex diseases, including *TAL1* (leukemia), *FOXD3* (several disorders), *HHEX* (diabetes), *PLCE1* (nephrotic syndrome), *NKX2* (heart trunk malformation), *TLX1* (leukemia), *FEZ1* (esophageal cancer), *ALX4* (forebrain absence), *SHANK3* (brain/immune defect), *NKX2* (heart malformations), and *IGF2*

(colorectal and other cancers). We also note that in cancer the high degree of epigenetic variation (the mechanism of which has proved elusive) would follow directly from our evolutionary model. Thus, rather than arising from a varying environment acting across generations, cancer may arise in part from a repeatedly changing microenvironment due to, for example, repeated exposures to carcinogens, which would select for epigenetic heterogeneity, and thus the ability of cells to grow outside of their normal milieu.

Materials and Methods

Tissue Samples and CHARM. Human tissues were obtained from the Stanley Foundation, and mouse tissues from C57BL/6 wild-type mice were obtained from Jackson Laboratory. Sample preparation and the CHARM DNA methylation analysis from which the data sets were derived are described in more detail elsewhere (14, 15).

VMRs. First, the microarray raw data from CHARM arrays (14) were transformed into estimated methylation percentages for each genomic location represented by a probe. These values were then smoothed (14) to obtain estimated methylation profiles for each sample. Then for each tissue, the SD for each location was computed. A region of locations surpassing a 99.95% percentile of all of the variances was designated a VMR.

Simulations. To create the simulation, we expanded the Fisher-Wright neutral selection model. In the neutral model, we started with N individuals and to create the next generation, we selected N individuals at random with replacement. This implies that the number of children for each individual follows a multinomial distribution, with population size remaining fixed at N. To introduce selection, we permitted each individual to die with probability $1\text{-}p_n$, with the survival probability p_n depending on a phenotype, Y_n. For the next generation, we selected N individuals, with replacement, from those that survived. For the simulation shown here, we quantified this relationship with a simple logistic function, $log\{ p_n /(1\text{-}p_n) \}= a + bY_n$. Note that if b is positive, then *positive Y* individuals are more fit, and if b is negative, then *negative Y* individual are more fit. We then assumed the existence of M SNPs, X_m, $m = 1,\ldots,M$, that affect the phenotype. We assumed two possible polymorphisms, designated 0 and 1, and denoted the expected change on the phenotype by β_j, $j = 1,\ldots,M$. We refer to $(X_1,\ldots,X_M)$ as the genotype. Note that there are 2^M different genotypes.

We followed Fisher's additive model for complex traits and assumed that the phenotype was a random variable with

$$Y_n = \beta_1 X_{n,1} + \beta_2 X_{n,2} + \ldots + \beta_M X_{n,M} + e_n.$$

Here e represents variation not explained by the standard genetic model and assumed to be a Gaussian random quantity with mean 0 and standard deviation s. Note that each genotype will have a different average Y value, determined by the effects β. We then added an epigenetic variation term caused by sequence changes (e.g., the addition of a CpG island that allows the presence of a VMR or T-DMR). We modeled this by incorporating another feature; we assumed the existence of M SNPs that altered the individual's variability (i.e., changed s). This is the epigenetic scenario, in which we are incorporating sequence variation that affects the variability of the phenotype, without altering the mean of the phenotype. This would be analogous to the earlier examples of loss or gain of CpGs that lead to the loss or gain of differentially methylated regions. We denote this epigenetic variation-inducing sequence change by Z and the effects by γ, and assume that

$$log2(s_n) = \gamma_1 Z_{n,1} + \gamma_2 Z_{n,2} + \ldots + \gamma_m Z_{n,m}.$$

Simulation 1. We started this simulation with an isogenic population and permit mutations to occur independently and at random at rate r. We ran this simulation with $n = 10{,}000$, $a = -4$, $b = 4$, $M = 8$ with $(\beta_1,\ldots,\beta_8) = (-1,-1,-1,-1,1,1,1,1)$, $s = 1$, and $r = 10^{-4}$. Note that these values of a and b imply that a average individual ($Y = 0$) has about a 1% chance of surviving. In contrast, an individual with the (0,0,0,0,1,1,1,1) genotype has about a 99% chance of surviving. For the epigenetic part of our model, we used $(\gamma_1,\ldots,\gamma_8)=(-1,-1,-1,-1,1,1,1,1)/2$. This implies that some mutations increase phenotype variance by 50% and others decrease it by 50%. We ran 1,000 generations 250 times.

Simulation 2, environment changing. We repeated simulation 1 except that we imitated dramatic environmental changes that changed the environment and its relationship with phenotype and fitness. The occurrence of these events was assumed to be random at a rate of 1 per 25 generations. Such a change resulted in b changing from 4 to -4. This implies that after the first event, smaller-than-average individuals were more fit than taller-than-average individuals. To check whether the outcome was stable, we considered a more skewed initial condition. Specifically, we reran the original simulation using 12 different sets of initial parameters. We first increased the number of iterations to 5,000. We then varied the environment changing rate to be 1 per 5, 1 per 10, 1 per 25, or 1 per 50 generations. Finally, we varied the number of mutating SNPs to be 2, 8, or 16. The conclusions from these simulations were as expected: Variability increased fitness, particularly in a changing environment (see Fig. S1).

Simulation 3. Simulation 3 was the same as simulation 1, except we did not permit mutations to affect the variance of Y.

ACKNOWLEDGMENTS. We thank Elisabet Pujadas for providing helpful discussions and comments on the manuscript, Simon Tavaré for pointing out evolution papers containing simulations, and Sarah Wheelan for help with BLAST. This work was supported by National Institutes of Health Grants P50 HG003233 and R01 GM083084.

1. Weiss KM (2004) The smallest grain in the balance. *Evol Anthropol* 13:122–126.
2. Barton NH, Briggs DEG, Eisen JA, Goldstein DB, Patel NH (2007) *Evolution* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY).
3. Goldstein DB (2009) Common genetic variation and human traits. *N Engl J Med* 360: 1696–1698.
4. Jablonka E, Lamb MJ (1995) *Epigenetic Inheritance and Evolution: The Lamarckian Dimension* (Oxford University Press, New York).
5. Cooney CA, Dave AA, Wolff GL (2002) Maternal methyl supplements in mice affect epigenetic variation and DNA methylation of offspring. *J Nutr* 132 (Suppl 8) 2393S–2400S.
6. Waterland RA, Jirtle RL (2003) Transposable elements: Targets for early nutritional effects on epigenetic gene regulation. *Mol Cell Biol* 23:5293–5300.
7. Rakyan VK, et al. (2003) Transgenerational inheritance of epigenetic states at the murine *Axin^{Fu}* allele occurs after maternal and paternal transmission. *Proc Natl Acad Sci USA* 100:2538–2543.
8. West-Eberhard MJ (2003) *Developmental Plasticity and Evolution* (Oxford Univ Press, New York).
9. Waddington CH (1935) *How Animals Develop* (Allen & Unwin, London).
10. Slatkin M (2009) Epigenetic inheritance and the missing heritability problem. *Genetics* 182:845–850.
11. Handel AE, Ramagopalan SV (2009) Public health implications of epigenetics. *Genetics* 182:1397–1398.
12. Carbone L, et al. (2009) Evolutionary breakpoints in the gibbon suggest association between cytosine methylation and karyotype evolution. *PLoS Genet* 5:e1000538.
13. Janion C (1982) Influence of methionine on the mutation frequency in *Salmonella typhimurium*. *Mutat Res* 94:331–338.
14. Irizarry RA, et al. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 41: 178–186.
15. Irizarry RA, et al. (2008) Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res* 18:780–790.
16. Nadeau JH (2009) Transgenerational genetic effects on phenotypic variation and disease risk. *Hum Mol Genet* 18 (R2):R202–R210.
17. Salathé M, Van Cleve J, Feldman MW (2009) Evolution of stochastic switching rates in asymmetric fitness landscapes. *Genetics* 182:1159–1164.
18. Kauffman SA (1994) *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford Univ Press, New York).
19. Gimelbrant A, Hutchinson JN, Thompson BR, Chess A (2007) Widespread monoallelic expression on human autosomes. *Science* 318:1136–1140.
20. He L, et al. (1998) Hypervariable allelic expression patterns of the imprinted *IGF2* gene in tumor cells. *Oncogene* 16:113–119.
21. Kaminsky ZA, et al. (2009) DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet* 41:240–245.
22. Page RE, Jr., Scheiner R, Erber J, Amdam GV (2006) 8. The development and evolution of division of labor and foraging specialization in a social insect (*Apis mellifera* L.). *Curr Top Dev Biol* 74:253–286.
23. Omholt SW, Amdam GV (2004) Epigenetic regulation of aging in honeybee workers. *Sci Aging Knowl Environ* 26:pe28.
24. Vogt G, et al. (2008) Production of different phenotypes from the same genotype in the same environment by developmental variation. *J Exp Biol* 211:510–523.

Genomic disorders: A window into human gene and genome evolution

Claudia M. B. Carvalho[a], Feng Zhang[a], and James R. Lupski[a,b,c,1]

[a]Department of Molecular and Human Genetics and [b]Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030; and [c]Texas Children's Hospital, Houston, TX 77030

Edited by Diddahally R. Govindaraju, Boston University School of Medicine, Boston, MA, and accepted by the Editorial Board November 5, 2009 (received for review July 22, 2009)

Gene duplications alter the genetic constitution of organisms and can be a driving force of molecular evolution in humans and the great apes. In this context, the study of genomic disorders has uncovered the essential role played by the genomic architecture, especially low copy repeats (LCRs) or segmental duplications (SDs). In fact, regardless of the mechanism, LCRs can mediate or stimulate rearrangements, inciting genomic instability and generating dynamic and unstable regions prone to rapid molecular evolution. In humans, copy-number variation (CNV) has been implicated in common traits such as neuropathy, hypertension, color blindness, infertility, and behavioral traits including autism and schizophrenia, as well as disease susceptibility to HIV, lupus nephritis, and psoriasis among many other clinical phenotypes. The same mechanisms implicated in the origin of genomic disorders may also play a role in the emergence of segmental duplications and the evolution of new genes by means of genomic and gene duplication and triplication, exon shuffling, exon accretion, and fusion/fission events.

chromosomal rearrangements | low copy repeats | segmental duplications | copy-number variation

Genomic Disorders Result from Copy-Number Variation

One decade ago the concept of genomic disorders was proposed predicated on two major premises: First, the conveyed clinical phenotype does not result from a point mutation, but rather from genomic rearrangements and second, the DNA rearrangement results from instability incited by genome architectural features (1, 2). It was considered that elucidating the rules for the mechanisms of human genomic rearrangements could potentially provide insights into what regions of the human genome are susceptible to instability. Structural variation can produce copy-number variation (CNV) that has been implicated in Mendelian diseases and common traits such as obesity (3, 4), neurobehavioral traits (4–8), and craniofacial features (9, 10), as well as in sporadic diseases (1, 2, 11, 12). The clinical phenotype conferred will vary depending on the genes and the genomic region involved and may result from distinct mechanisms including gene dosage effects, gene disruption, and position effects or by unmasking a recessive allele (13–16).

Mechanistically, the instability and thus mutability of our genome can be facilitated by the ubiquitous presence of repeat sequences, such as low copy repeats (LCRs) or segmental duplications (SDs), as well as by the presence of repetitive sequences such as short interspersed nuclear elements (SINEs) and long interspersed nuclear elements (LINEs). Characterization of many genomic rearrangements causative of human diseases revealed two rearrangement types that could be distinguished at a given locus: recurrent and nonrecurrent rearrangements. Recurrent rearrangements have the same size and fixed breakpoints that cluster in LCRs (17); these LCRs can act as homologous recombination substrates. Nonrecurrent rearrangements have varied sizes and breakpoints for each patient. The mapping and delineation of the smallest region of overlap (SRO) pertaining to nonrecurrent duplications and deletions in a given patient cohort can be used to delineate the genes or regulatory sequences within the dosage-sensitive genomic interval mediating the phenotypic consequences of the genomic change. A subtype of the nonrecurrent rearrangements is characterized by one breakpoint grouping, but not clustering, in a genomic region. The breakpoint grouping can be coincident with genomic intervals laden with genomic sequence elements able to form unusual non-B DNA structures, such as hairpins and cruciforms, potentially stimulating specific mechanisms that drive nonrecurrent rearrangements (17).

The Human Genome Is Enriched in Both Repeated and Repetitive Sequences

LCRs were defined as intrachromosomal duplications ≥10 kb in length and with ≥97% sequence identity that probably arose by duplication of genomic segments resulting in paralogous regions of the human genome (15). SDs were defined as segments of DNA containing ≥90% of sequence identity and ≥1 kb in length (18); both terms are used interchangeably. In contrast, repetitive sequences were defined much earlier (1968) by Britten and Kohne using reassociation kinetics (19), therefore constituting a different class of repeats. LCR/SD became apparent during mechanistic studies of genomic disorders and their genomewide nature was independently revealed during studies of the sequence of the draft haploid human genome (20, 21); these were not revealed by reassociation kinetics. In fact, as much as 5.4% of our genome is duplicated (≥1 kb and ≥90% identity) (22). Also, 52% of the remaining gaps in the reference haploid human genome, refractory regions to all techniques available at the moment, are flanked by LCRs with >90% identity (23). According to the Human Genome Sequence Consortium, "by far, the most difficult regions of the genome were those containing near-exact segmental duplications" (23). The analysis of the (almost) finished human euchromatic genome sequence provided in 2004 enabled the confirmation of several remarkable LCR features already documented by previous studies. LCRs are present across the entire human genome, they can be inter- or intrachromosomal, and they often contain partial or complete gene sequences with intron–exon structure. In addition, LCRs can be classified into three categories: pericentromeric, subtelomeric, or those present in interstitial. Pericentromeric and subtelomeric LCRs are biased toward interchromosomal LCRs and organized as a complex mosaic of duplications; by contrast, interstitial LCRs are enriched for interspersed LCRs (22).

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Evolution in Health and Medicine" held April 2–3, 2009, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS web site at www.nasonline.org/Sackler_Evolution_Health_Medicine.

Author contributions: C.M.B.C., F.Z., and J.R.L. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. D.R.G. is a guest editor invited by the Editorial Board.

[1]To whom correspondence should be addressed. E-mail: jlupski@bcm.tmc.edu.

Proximal 17p as a Model for Human Genomic Disorders and Evolution

Studies of rearrangements involving chromosome 17p11.2p12 showed that the proximal 17p chromosome is marked by several direct and inverted interspersed LCRs (2). The ~7.5-Mb LCR-rich region evolved to a complex genomic architecture that involved serial segmental duplication events during primate evolution (24), often emanating from preferential LCR-containing genomic intervals or cores, some regions with apparent increased mutation rates and others with apparent reduced recombination (12, 25–27), potentially reflecting inversion polymorphisms. It is also the site of the breakpoint for an evolutionary translocation, t(4;19), that occurred in an ancestral gorilla chromosome (28). Four genomic disorders, Charcot–Marie–Tooth disease type 1A [CMT1A (MIM118220)], hereditary neuropathy with liability to pressure palsies [HNPP (MIM 162500)], Smith–Magenis microdeletion syndrome [SMS (MIM 182290)] and Potocki–Lupski microduplication syndrome [PTLS (MIM 610883)], map to this region. CMT1A is a length-dependent distal symmetric polyneuropathy caused by a 1.4-Mb duplication generated by nonallelic homologous recombination (NAHR) between the distal CMT1A-REP and proximal CMT1A-REP (29). HNPP is a milder condition with susceptibility to asymmetric neuropathy; it results from the reciprocal deletion of the same genomic segment (30). SMS is a multiple congenital anomaly mental retardation syndrome with obesity, sleep disturbance, and specific behavioral abnormalities due to a recurrent 3.7 Mb deletion generated by NAHR between two LCRs, the so-called proximal and distal SMS-REPs (27, 31). PTLS is due to the reciprocal duplication and manifests as failure to thrive and neurobehavioral abnormalities, including features of autism (6, 32). NAHR between the same LCRs can produce the microdeletion or the reciprocal duplication disorder (6, 32–34), but the large number of LCRs spanning the region can also result in uncommon recurrent rearrangement, using alternative LCR as homologous recombination (HR) substrates (35). Furthermore, nonrecurrent deletions/duplications are potentially stimulated by other LCRs in the region (12, 32, 36).

The 17p11.2p12 region also undergoes other structural variations observed in the population such as an inversion involving the distal SMS-REP and middle SMS-REPs [Database of Genomic Variants (DGV), http://projects.tcag.ca/variation/]. Furthermore, structural changes therein can occur somatically and be associated with cancer (37–40). The common breakpoint of the i(17q) chromosome, formally idic(17)(p11.2), frequently observed in patients with hematological malignancies associated with poor prognosis, maps to the LCRs REPA and -B located between middle and proximal SMS-REPs (41). The same region is very polymorphic within the population (42). A summary of the characterized proximal 17p evolutionary, constitutional (i.e. germ-line), and somatic rearrangement events is shown in Fig. 1.

The presence of the specific dosage-sensitive gene within the 17p11.2p12 chromosome was demonstrated by the identification of rare patients who had disease-causative point mutations in the dosage-sensitive gene rather than large genomic rearrangements including that gene: for example, patients with HNPP and without a deletion who had loss-of-function *PMP22* point mutations (nonsense/frameshift) and rare CMT1A patients without duplication who instead had gain-of-function *PMP22* point mutations (43–45). Dosage alteration of the *retinoic acid inducible 1* (*RAI1*) gene causes most of the clinical phenotypes observed in patients with SMS, an observation also supported by mouse models (10, 46–48). Furthermore, nonsense and frameshift point mutations within that gene were detected in patients with SMS who did not have a genomic deletion of *RAI1*, implicating haploinsufficiency as a major contributing factor for the disease (46–48). Point mutations leading to gain-of-function are predicted to cause PTLS but such patients have not yet been identified.

Delineation of the NAHR Mechanism Enabled the Prediction of Novel Genomic Disorders

NAHR is a frequent mechanism underlying disease-associated genomic rearrangements. LCRs are the usual substrates for NAHR due to their high degree of sequence identity. Experimental observations have implicated the existence of recombination hotspots for the occurrence of the crossovers within the LCRs (33, 49, 50). Two LCRs involved in a particular NAHR can be interchromosomal, intrachromosomal, or intrachromatidal, and they can be either directly or inversely oriented to each other.

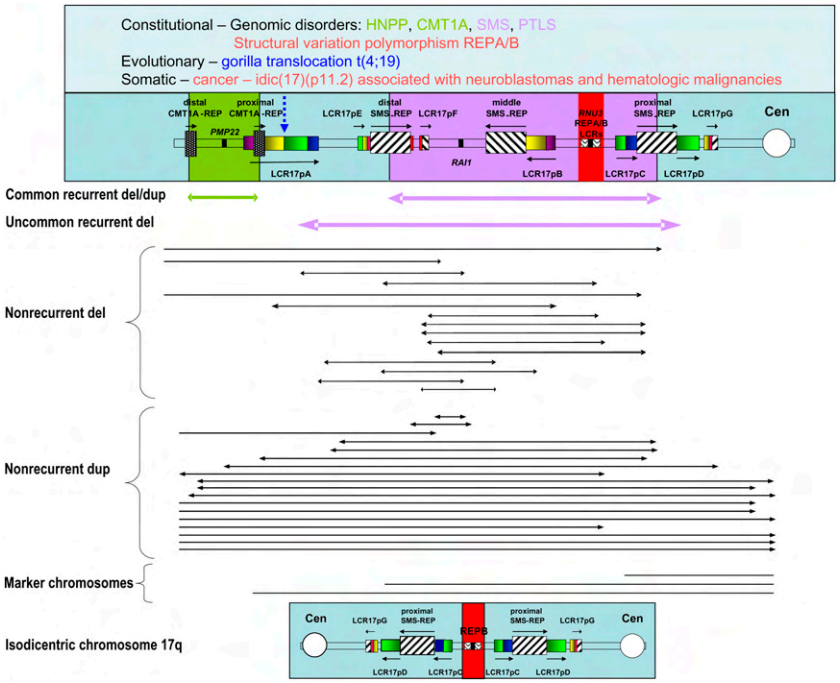


Fig. 1. Schematic representation of the genome architecture susceptible to rearrangements in the proximal chromosome 17p. The low copy repeats are shown in rectangles (color-coded or similar symbols for given repeats), along with the distribution of the rearrangement breakpoints. (*Upper*) Diverse alterations (constitutional, evolutionary, somatic) thus far documented for this region. They are color coded for matching the involved segment on 17p. The green horizontal arrow below represents the recurrent duplication and deletion causative of CMT1A and HNPP, respectively; purple horizontal arrows represent the recurrent deletion and duplication causative of SMS and PTLS (3.7 Mb) and the recurrent but uncommon deletion causative of SMS (~5 Mb). Black arrows below represent the uncommon nonrecurrent deletions and duplications causative of SMS and PTLS, respectively. Solid black line: marker chromosome breakpoints. (*Lower*) Schematic representation of the isodicentric chromosome 17q, formally designated idic (17)(p11.2), generated according to the model proposed by Barbouti et al. (41) and adapted, with permission, from refs. 12 and 113.

The rearrangement will generate different products accordingly, i.e., duplication, deletion, inversion, or translocation (1, 15). Experimentally, deletions occur twice as often as duplications during meiosis in male germ cells (51).

The high frequency of interspersed LCRs in the human genome predicts many regions of genomic instability that could potentially undergo NAHR-mediated rearrangements and be associated with genomic disorders. In a "genome-first" approach, Sharp et al. (7) developed a BAC-based array Comparative Genomic Hybridization (aCGH) designed to interrogate 130 genomic intervals flanked by directly orientated LCRs >10 kb in length, with >95% identity, and within a distance of 50 kb to 5 Mb. Such an approach was used to screen patient cohorts with defined phenotypes, such as mental retardation and congenital anomalies, enabling the detection of five microdeletions (at 17q21.31, 17q12, 15q24, 15q13.3, and 1q21.1) and further description of five novel genomic disorders (7). Therefore, knowledge of the NAHR mechanism has played a pivotal role in uncovering new human syndromes with profound consequences for clinical genetics.

Other Mechanisms Produce Nonrecurrent Rearrangements

Nonrecurrent rearrangements can be generated by NAHR between repetitive sequences such as SINEs and LINEs (36), but other molecular mechanisms are also implicated for their origin, including nonhomologous end joining (NHEJ), fork stalling and template switching (FoSTeS), microhomology-mediated break-induced replication (MMBIR), and retrotransposition (reviewed in refs. 16, 17, 52, 53). NHEJ is one of the repair pathways responsible for double-strand break (DSB) repair in cells. Following detection of DSBs, NHEJ rejoins the broken DNA ends without the requirement for homology; this process requires the preparation of damaged ends using base removal and insertions of new bases, without ensuring sequence restoration around the break (54). FoSTeS is a recently described replication-based mechanism proposed to explain the complex *PLP1* duplications at Xq22, associated with the genomic disorder Pelizaeus–Merzbacher disease [PMD (MIM 312080)] (55). It was proposed that during DNA replication the DNA replication forks could stall, and the 3′ end of the newly synthesized strand could resume DNA synthesis on a different template in a second nearby replication fork. Microhomologies between the switched template and the original fork are used to prime replication. DNA deletion or duplication can be generated depending on whether the template switching occurred to a new replication fork located upstream or downstream. Inversions can also be produced depending on the direction of the fork progression and if the leading or the lagging strands are used on the switched template. The disengaging/resuming replication in a different fork/extension process can occur multiple times, producing complex rearrangements (17, 55). The FoSTeS model has been further generalized and the molecular details are provided in the MMBIR model that appears to be operative in all domains of life (52). In this model, the replication fork stalls, by virtue of the presence of a nick on the template strand resulting in a collapsed fork as the replication fork proceeds through the nick. The collapsed fork generates a one-ended, double-stranded, DNA that is resected to expose the 3′ end, which can mediate a break-induced replication (BIR) using microhomology to prime the template switch.

The presence of complex rearrangements in several genomic disorders has been increasingly detected due to the greater resolution of the advanced genome technologies. Recent examples include *MECP2* duplications (56) [MRXSL (MIM 300260)] and duplications in 17p13.3 involving the *PAFAH1B1* (LIS1) and/or the *YWHAE* (14-3-3e) genes (57). Remarkably, Zhang et al. (58) observed as much as 57% of the nonrecurrent PTLS-associated duplications involving the 17p11.2 region can be complex rearrangements (58). The extent to which FoSTeS/MMBIR is involved in the generation of human structural variation is still undetermined as breakpoint sequences of the complex rearrangements are particularly difficult to obtain.

Genomic Architecture Incites Rearrangements

The presence of LCRs in a specific genomic region increases the probability of occurrences of new rearrangements, such as duplications, deletions, gene conversions, and inversions therein or in the flanking segments (36). Indeed, the association between structural variation and LCRs has been shown in several studies, including those examining the genome of different populations (59–63), those analyzing individual genomic loci (through human-specific disease studies) (1, 9, 12, 32, 55, 56, 64–70), and genomic evolutionary studies (24, 71). In fact, it has been shown that between human and chimpanzee ~70–80% of inversions and ~40% of deletions/duplications map to regions containing LCRs (71). Interestingly, the unique regions flanking segmental duplications are ~10 times more probable to become duplicated compared to other randomly distributed regions, a phenomenon termed "duplication shadowing" (22) that partially explains the nonrandom distribution and the complex mosaic patterns of LCRs. This observation is supported by a recent comparative study in primates where it was shown that LCRs do not arise randomly, but are likely to arise within or adjacent to another LCR already present (24, 72, 73). Therefore, the unique regions flanking LCRs will eventually undergo rearrangements that can either create new LCRs or add new complexities to the previous one; additionally, ectopic homologous recombination and gene conversion can produce homogenization, maintaining the sequence conservation within the LCRs (15).

The role of the LCRs in recurrent rearrangements as substrates for NAHR is well established (1). However, LCRs can also be associated with nonrecurrent rearrangements generated by FoSTeS/MMBIR. Inoue et al. (68) analyzed families with PMD due to deletion of the *PLP1* gene at Xq22 and found the distal breakpoints in two of three cases were embedded in LCRs. This finding was supported by the results of Lee et al. (55, 74) who studied PMD patients carrying *PLP1* duplications. Later a statistically significant association between LCRs and the distal breakpoints of duplications involving the *MECP2* gene in male patients with neurodevelopmental delay was shown (56). Approximately 77% (23/30) of the distal duplication breakpoints map within or nearby one of the LCRs (LCRJ and LCRK) that are located 47 and 201 kb telomeric to the *MECP2* gene. LCRJ is formed by two genes that constitute the Opsin array, *OPN1LW* (long-wave sensitive) and *OPN1MW* (middle-wave sensitive). In vertebrates, the visual pigments are the products of five families of Opsin genes that probably have arisen by multiple gene duplication events at least 540 million years ago (Mya) (reviewed in ref. 75). The ability to absorb three different wavelengths (short, medium, and long) in the retina is not found among many mammals and constitutes a distinctive feature of the primates. Such evolutionary acquisition enabled primates to see three primary colors (blue, green, and red), changing their vision from dichromatic to trichromatic. From a molecular evolutionary standpoint, the event that enabled trichromatic vision in Catarrhines was the duplication of the X chromosome Opsin gene that occurred ~35 Mya followed by gene diversification (75–77). Interestingly, in humans, differences regarding the sensitivity to distinguish red–green colors are very common. As much as 8% of Caucasian males present color-vision defects and polymorphisms resulting from frequent chromosomal rearrangements and gene conversions at the Opsin locus (78). Of note, the evolutionary timing of the switch from dichotamous to trichotamous color vision coincides with the loss of many functional olfactory receptor genes and may reflect increased dependence of higher primates on vision versus olfaction to sense one's environment (79).

In our cohort of patients with *MECP2* duplication, the stronges breakpoint bias was observed in patients with complex rearrangements

(triplications embedded within duplications) (56) in which both duplication and triplication breakpoints map within a low copy repeat pair termed LCRs K (Fig. 2). The LCRK1 and LCRK2 are positioned in an inverted orientation with respect to each other, have 99% sequence identity, and are 11.3 kb in length (56, 80). The region between the LCRs K, which contains the *FLNA* and *EMD* genes, is inverted in 18% of individuals of European descent (81). Nonrecurrent deletions involving one of the LCRs K and the *EMD* gene have been reported to cause X-linked Emery–Dreifuss muscular dystrophy [EDMD (MIM 310300)] (81). Caceres et al. (82) identified the presence of the LCRs K in diverse eutherians, suggesting that they are derived from an ancestral duplication and probably have a single common origin. In addition, inversion events occurred at least 10 independent times along the eutherian evolution (82).

Duplication Rearrangements and the Emergence of Novel Traits

In his seminal work, Ohno (83) proposed that gene duplications coupled with rapid sequence diversification may play a fundamental role in evolution. Increasing evidence from experimental studies in diverse organisms has confirmed his prediction. In fact, duplications may act as a "reservoir" for producing adaptative phenotypes (84), but also they can cause a dramatic increase in the dosage of specific genes, producing an immediate advantageous effect (85).

In primates, LCRs are implicated in lineage-specific gene creation and potentially in speciation as well. A comparative study between human and chimpanzee genomes estimated that 2.7% of euchromatic sequences were differentially duplicated between chimpanzee and human (86). In contrast, single-base pair differences account for 1.2% of the genetic difference (87). Therefore, some of the genes that distinguish human from chimpanzee arose and/or expanded as LCRs. The salivary amylase gene (*AMY1*), which encodes a protein that catalyzes the first step in digestion of dietary starch and glycogen, constitutes an interesting example. It has approximately three times more copies in humans compared to chimpanzees and copy-number differences correlate positively with the higher levels of salivary amylase protein (88). The copy number of *AMY1* shows evidence of positive selection in populations with high-starch consumption, suggesting that its copy-number increase in humans was selectively favored due to the concomitant increase of starch consumption in agricultural societies (88). Also, the human-specific amplification of the aquaporin-7 gene (*AQP7*), coupled to positive selection, provides another example of adaptative traits that emerged after lineage-specific gene duplication; the aquaporin-7 protein is involved in water, glycerol, and urea membrane transport and may have contributed to enabling the human capacity for endurance running (89, 90).

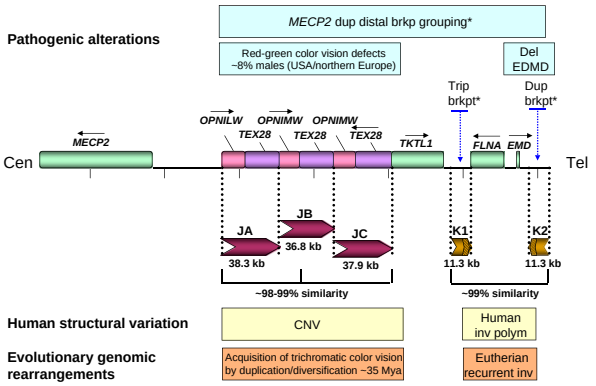


Fig. 2. Schematic representation of the *MECP2* telomeric region. (*Top*) Blue boxes represent the pathogenic rearrangements documented in the literature thus far: distal breakpoint grouping of most of the patients with *MECP2* duplications, deletions and/or gene conversions of the Opsin genes that cause color blindness, and deletions of the *EMD* gene that cause Emery–Dreifuss muscular dystrophy (EDMD). (*Middle*) The genomic context telomeric to *MECP2*. LCRJ spans 114 kb and is formed by three genes and/or pseudogenes that constitute the Opsin array, *OPN1LW*, *OPN1MW*, and *TEX28*. The nearby LCRs, K1 and K2, are positioned in inverted orientation, have 99% sequence identity, and are 11.3 kb in length. Hatched bars within arrows inside the LCRs K represent the small region that is 100% identical between them. Blue arrows show alignment of the join points of the patients carrying complex rearrangements (triplications embedded in duplications). (*Bottom*) Human structural variation (yellow rectangles) includes CNVs and inversions; evolutionary genomic rearrangements (orange rectangles) include the duplication of the Opsin gene and further acquisition of the trichromatic color vision during the primate evolution in addition to a recurrent inversion that has been occurring multiple times in eutherians. *, based on data reported in Carvalho et al. (56).

Using interspecies cDNA CGH in five hominoid species including humans, Fortna et al. (89) identified 140 genes showing human lineage-specific variation in copy number, most of them (134/140) due to amplification. Several genes are implicated in neuronal function, including a neurotransmitter transporter for γ-aminobutyric acid (GABA) (*SLC6A13*) and the gene encoding the neuronal apoptosis inhibitory protein (*NAIP*), which is suspected to have a role in neuronal proliferation and/or brain size in humans (89). Remarkably, they showed that the neuronal-expressed DUF1220 domain, which presents the highest number of copies in humans compared to primates, is apparently under positive selection (91). It is estimated that perhaps as many as 34 human genes encode a DUF1220 domain; these genes map to several genomic sites on chromosome 1 with the majority localized to the 1q21.1 region. Rearrangements of 1q21.1 are associated with congenital anomalies, mental retardation, and neuropsychiatric phenotypes (7, 9, 92). Sikela et al. found a high correlation between DUF1220 domain copy number and human head circumference, suggesting that DUF1220 domains may have a role in shaping human brain size (93). Brunetti-Pierri et al. (9) recently showed that 1q21.1 deletion is associated with microcephaly whereas 1q21.1 duplication is associated with macrocephaly. In fact, rearrangements involving 1q21.1 represent an interesting example of copy-number variation causing developmental and behavioral phenotypes. Noteworthy, 63.6% of the content of the 1q21 chromosome sequence is represented by LCRs. Recent studies showed recurrent deletions and duplications in patients with a broad range of clinical phenotypes including dysmorphic features, congenital anomalies, mental retardation, and neuropsychiatric conditions such as attention deficit hyperactivity disorder (ADHD), autism, anxiety/depression, and antisocial behavior (7, 9, 92). Deletions have also been recently associated with schizophrenia (5, 94). The association of microdeletion with microcephaly and duplication with macrocephaly can be potentially explained by the copy-number alteration of the human-specific paralog of the gene *HYDIN*. In mice, mutations causing premature termination of the *Hydin* gene product were reported to cause hydrocephalus (95). The 1q21.1 paralog *HYDIN* copy results from a 360-kb interchromosomal duplication from the 16q22.2 segment (96) containing the original *HYDIN* gene.

Exon Shuffling and the Emergence of New Genes

Along with gene duplication, exon shuffling is also implicated in the generation of novel genes and proteins and, once more, the LCRs might play a pivotal role underlying that event. In 1978, Walter Gilbert launched the concept of exon shuffling when he proposed that recombination between introns could rearrange exons, creating new transcription units, and consequently new proteins could be formed (97). In the primate lineage, including humans, there is evidence of exon shuffling generating novel genes, e.g., the creation of testis-specific genes (98, 99). Some additional evidence for exon shuffling observed in human and mouse subjects is listed in Table 1.

Table 1. Examples of exon shuffling and their potential mechanisms

Organism	Involved gene	Mechanism	Microhomology at breakpoint	Reference
Mouse	aA-crystallin	Illegitimate recombination*	CCCAT	(123)
	Gnb5, *Myo5a*	Nonhomologous recombination*	GG	(124)
Human	LDL receptor, EGF precursor	NA	NA	(125, 126)
	Multiple genes	L1 retrotransposition	NA	(127)
	Kua, *UEV*	Gene fusion	NA	(128)
	PMCHL1, *MCH*	Retrotransposition	NA	(129)
	ATM	Retrotransposition	NA	(130)
	TRE-2 (*USP6*)	Gene fusion	NA	(99)
	PIPSL, PIP5K1A, PSMD4	L1 retrotransposition	NA	(131)

NA, not available.
*Microhomology was shown at the breakpoint, which can be alternatively interpreted to be caused by the FoSTeS/MMBIR mechanism.

An interesting example of the "birth" of a gene due to duplication and exon shuffling is the proximal CMT1A-REP. This LCR arose by genomic rearrangement whereby exon VI of the *COX10* gene and surrounding 25-kb intronic sequences (i.e., distal CMT1A-REP) were duplicated and inserted 1.4 Mb more proximal on 17p within the human–chimpanzee ancestral chromosome. This one event created proximal CMT1A-REP (Fig. 3) and gave birth to two novel genes through exon accretion and fission, respectively (24, 26). Interestingly, both novel genes, *HREP* and *CDRT1*, are expressed in humans although they have different tissue specificity: *HREP* is expressed in heart and skeletal muscle whereas *CDRT1* is mainly expressed in pancreas (26, 100). The original *COX10* (distal CMT1A-REP) is highly expressed in multiple tissues (101); its protein product farnesylates the heme group incorporated into cytochrome oxidase that is important for mitochondrial function.

Another example is the hominoid testis-specific gene TRE-2 (*USP6*) that emerged during primate evolution resulting from the chimeric fusion of two genes, *USP32* and *TBC1D3* (Table 1) (99). *TBC1D3* itself is derived from a segmental duplication that underwent multiple gene duplications during primate evolution (99). Interestingly, *TBC1D3* underwent mutations with respect to its closest homolog, *USP6NL* and acquired the features of an adaptor molecule involved in the macropinocytic process (102).

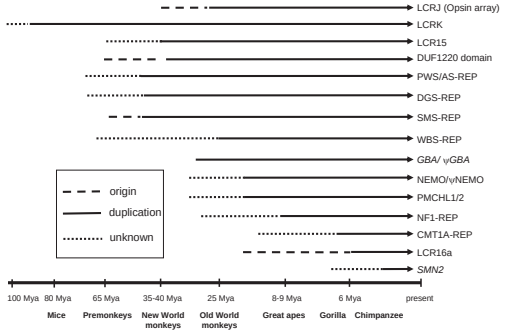


Fig. 3. Duplication of selected LCRs during molecular evolution of the primates (updated from ref. 122). The figure is not to scale. LCRJ, Opsin and *TEX28* array at Xq28; LCR15, LCR highly repeated in chromosome 15q11-q14; LCRK, LCR flanking the genes *FLNA* and *EMD* at Xq28; PWS/AS, Prader–Willi and Angelman syndromes; DGS, DiGeorge syndrome; SMS, Smith–Magenis syndrome; WBS, Williams–Beuren syndrome; *GBA*, glucocerebrosidase gene; *NEMO*, gene mutated in incontinentia pigmenti; PMCHL1/2, chimeric genes derived from the melanin-concentrating hormone gene; NF1, neurofibromatosis 1; CMT1A, Charcot-Marie-Tooth disease type 1A; LCR16a, low copy repeats on chromosome 16; *SMN2*, gene mutated in spinal muscular atrophy. This figure was adapted, with permission, from ref. 122.

Can FoSTeS/MMBIR Account for Exon Shuffling Events?

It has been estimated that at least ~19% of exons in eukaryotic genes were formed by exon shuffling (103). However, the underlying mechanisms are not fully understood. Two mechanisms have been proposed, illegitimate recombination (104) and retrotransposed exon insertion (105); nevertheless, many exon rearrangements are not readily explained by either mechanism. Recently, our group (58) reported complex rearrangements, including triplications, detected at the join points of duplications in patients with PTLS and in patients with *PMP22* duplication and deletions. Importantly, the complex patterns implicating FoSTeS/MMBIR could be detected at different levels of genome resolution from involving megabases of the human genome to small genomic intervals containing a single gene or even only one exon (58). The sequencing of the join points of the deletion involving just one exon of the *PMP22* gene revealed a complex pattern including a small insertion in an inverted orientation. This complex rearrangement of a coding exon caused by FoSTeS/MMBIR suggests that this mechanism may contribute to exon shuffling (58). The replicative FoSTeS/MMBIR mechanism could readily shuffle any given exon by a template switch before and after that exon anywhere within the flanking introns (58).

Birth Defects: Evolution in Real Time

Structural variation in the human genome encompasses a wide range of different alterations, including aneuploidies, heteromorphisms, fragile sites, repetitive elements, micro- and minisatellites, insertions, deletions, inversions, duplications, balanced and unbalanced translocations, and complex genomic and chromosomal rearrangements (60, 62, 106–110). Some of the changes are large enough to be visualized by light microscopy whereas others require special techniques, e.g., submicroscopic alterations can be detected by CGH (62, 107), whereas inversions can be detected by paired-end sequencing techniques (60, 109, 111) or by PCR-based approaches (112).

In the human genome, the de novo locus-specific mutation rates for genomic rearrangements were estimated on the basis of disease prevalence rates as $\sim 10^{-6}$–10^{-4} (113, 114). This range is two to four orders of magnitude greater than the locus-specific rates for base pair changes ($\sim 10^{-8}$). Therefore, CNVs may frequently occur de novo and can be associated with sporadic birth defects. This contention is supported by a recent study on neonates with various birth defects in which a high frequency of de novo pathological CNVs was identified (115). aCGH was used to screen 638 neonates with different birth defects including dysmorphic features, multiple congenital anomalies, congenital heart disease, cleft palate, etc. Pathogenic CNVs were detected in up to 20% of subjects (115). In patients with a clinical indication of suspected chromosomal abnormalities, the rate of de novo CNV detection was as high as 66.7% (115), three times greater than the published rates for chromosome studies.

Can Structural Variation Produce Atavism?

Atavism is a concept proposed by Darwin in 1868 to term the reappearance of ancestral characteristics in individuals of a species in further generations. Evidence for atavistic traits has been found in horses and whales (116). In humans proposed atavistic traits include extra nipples, the ability to move the scalp, natural "earring" holes, and hypertrichosis (117). Hypertrichosis is a rare condition characterized by excessive generalized or localized hairiness (118). Marcias-Flores et al. (119) described a family with a severe form of X-linked hypertrichosis. Figuera et al. (118) mapped the X-linked locus to the chromosome Xq24-q27.1, but autosomal-dominant inheritance patterns with associated clinical signs, e.g., gingival hyperplasia, skeletal abnormalities, mental retardation, and others, have also been described (reviewed in ref. 120). Recently, Sun et al. (121) mapped the congenital generalized hypertrichosis terminalis (CGHT) trait to chromosome 17q24.2-q24.3 by linkage analysis in three Han Chinese families. They identified nonrecurrent microdeletions in three CGHT families and also found one de novo microduplication in a sporadic patient, as causative of the trait. The candidate gene has not yet been identified but they postulated a long-range position effect potentially due to the presence of the *SOX9* gene nearby.

Conclusions

In conclusion, the development of the concept of genomic disorders, and the definition of the mechanisms for formation (e.g., NAHR, FoSTeS/MMBIR) of the rearrangements underlying these conditions, has led to improvement in clinical ascertainment and the discovery of novel syndromes. Such studies revealed a great deal of new information about human genome structure and evolution and delineated the role of the genomic architecture, including repetitive (e.g., SINEs and LINEs) and repeat sequences (LCRs/SDs), as a facilitator of genomic instability that can cause disease. Adaptative traits can be driven by structural variation as exemplified by the amylase (*AMY1*) copy-number variation associated with the change of human eating habits. Moreover, increasing data regarding human CNVs and how they can convey neuropsychiatric phenotypes suggest that CNVs may play a major role in human cognition and other complex traits.

1. Lupski JR (1998) Genomic disorders: Structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet* 14:417–422.
2. Lupski JR (2009) Genomic disorders ten years on. *Genome Med* 1:42.
3. Sha BY, et al. (2009) Genome-wide association study suggested copy number variation may be associated with body mass index in the Chinese population. *J Hum Genet* 54:199–202.
4. Walz K, Paylor R, Yan J, Bi W, Lupski JR (2006) *Rai1* duplication causes physical and behavioral phenotypes in a mouse model of dup(17)(p11.2p11.2). *J Clin Invest* 116:3035–3041.
5. International Schizophrenia Consortium (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455:237–241.
6. Potocki L, et al. (2000) Molecular mechanism for duplication 17p11.2—the homologous recombination reciprocal of the Smith-Magenis microdeletion. *Nat Genet* 24:84–87.
7. Sharp AJ, et al. (2006) Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet* 38:1038–1042.
8. Sharp AJ, et al. (2007) Characterization of a recurrent 15q24 microdeletion syndrome. *Hum Mol Genet* 16:567–572.
9. Brunetti-Pierri N, et al. (2008) Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet* 40:1466–1471.
10. Yan J, et al. (2004) Reduced penetrance of craniofacial anomalies as a function of deletion size and genetic background in a chromosome engineered partial mouse model for Smith-Magenis syndrome. *Hum Mol Genet* 13:2613–2624.
11. Gu W, Lupski JR (2008) CNV and nervous system diseases—What's new? *Cytogenet Genome Res* 123:54–64.
12. Stankiewicz P, et al. (2003) Genomic disorders: Genome architecture results in susceptibility to DNA rearrangements causing common human traits. *Cold Spring Harbor Symp Quant Biol* 68:445–454.
13. Henrichsen CN, Chaignat E, Reymond A (2009) Copy number variants, diseases and gene expression. *Hum Mol Genet* 18:R1–R8.
14. Lupski JR, Stankiewicz P (2005) Genomic disorders: Molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet* 1:e49.
15. Stankiewicz P, Lupski JR (2002) Genome architecture, rearrangements and genomic disorders. *Trends Genet* 18:74–82.
16. Zhang F, Gu W, Hurles ME, Lupski JR (2009) Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet* 10:451–481.
17. Gu W, Zhang F, Lupski JR (2008) Mechanisms for human genomic rearrangements. *Pathogenetics* 1:4.
18. Bailey JA, et al. (2002) Recent segmental duplications in the human genome. *Science* 297:1003–1007.
19. Britten RJ, Kohne DE (1968) Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* 161:529–540.
20. Eichler EE (1998) Masquerading repeats: Paralogous pitfalls of the human genome. *Genome Res* 8:758–762.
21. Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE (2001) Segmental duplications: Organization and impact within the current human genome project assembly. *Genome Res* 11:1005–1017.
22. Bailey JA, Eichler EE (2006) Primate segmental duplications: Crucibles of evolution, diversity and disease. *Nat Rev Genet* 7:552–564.
23. IHGSC (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431:931–945.
24. Stankiewicz P, Shaw CJ, Withers M, Inoue K, Lupski JR (2004) Serial segmental duplications during primate evolution result in complex human genome architecture. *Genome Res* 14:2209–2220.
25. Bi W, et al. (2002) Genes in a refined Smith-Magenis syndrome critical deletion interval on chromosome 17p11.2 and the syntenic region of the mouse. *Genome Res* 12:713–728.
26. Inoue K, et al. (2001) The 1.4-Mb CMT1A duplication/HNPP deletion genomic region reveals unique genome architectural features and provides insights into the recent evolution of new genes. *Genome Res* 11:1018–1033.
27. Park SS, et al. (2002) Structure and evolution of the Smith-Magenis syndrome repeat gene clusters, SMS-REPs. *Genome Res* 12:729–738.
28. Stankiewicz P, Park SS, Inoue K, Lupski JR (2001) The evolutionary chromosome translocation 4;19 in Gorilla gorilla is associated with microduplication of the chromosome fragment syntenic to sequences surrounding the human proximal CMT1A-REP. *Genome Res* 11:1205–1210.
29. Lupski JR, et al. (1991) DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell* 66:219–232.
30. Chance PF, et al. (1993) DNA deletion associated with hereditary neuropathy with liability to pressure palsies. *Cell* 72:143–151.
31. Chen KS, et al. (1997) Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. *Nat Genet* 17:154–163.
32. Potocki L, et al. (2007) Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype. *Am J Hum Genet* 80:633–649.
33. Bi W, et al. (2003) Reciprocal crossovers and a positional preference for strand exchange in recombination events resulting in deletion or duplication of chromosome 17p11.2. *Am J Hum Genet* 73:1302–1315.
34. Shaw CJ, Bi W, Lupski JR (2002) Genetic proof of unequal meiotic crossovers in reciprocal deletion and duplication of 17p11.2. *Am J Hum Genet* 71:1072–1081.
35. Shaw CJ, Withers MA, Lupski JR (2004) Uncommon deletions of the Smith-Magenis syndrome region can be recurrent when alternate low-copy repeats act as homologous recombination substrates. *Am J Hum Genet* 75:75–81.
36. Shaw CJ, Lupski JR (2004) Implications of human genome architecture for rearrangement-based disorders: The genomic basis of disease. *Hum Mol Genet* 13 (Spec No 1):R57–R64.
37. Mendrzyk F, et al. (2006) Isochromosome breakpoints on 17p in medulloblastoma are flanked by different classes of DNA sequence repeats. *Genes Chromosomes Cancer* 45:401–410.
38. Babicka L, et al. (2006) Complex chromosomal rearrangements in patients with chronic myeloid leukemia. *Cancer Genet Cytogenet* 168:22–29.
39. McCabe MG, et al. (2006) High-resolution array-based comparative genomic hybridization of medulloblastomas and supratentorial primitive neuroectodermal tumors. *J Neuropathol Exp Neurol* 65:549–561.
40. Fabris S, et al. (2008) Molecular and transcriptional characterization of 17p loss in B-cell chronic lymphocytic leukemia. *Genes Chromosomes Cancer* 47:781–793.
41. Barbouti A, et al. (2004) The breakpoint region of the most common isochromosome, i (17q), in human neoplasia is characterized by a complex genomic architecture with large, palindromic, low-copy repeats. *Am J Hum Genet* 74:1–10.
42. Carvalho CM, Lupski JR (2008) Copy number variation at the breakpoint region of isochromosome 17q. *Genome Res* 18:1724–1732.
43. Nicholson GA, et al. (1994) A frame shift mutation in the *PMP22* gene in hereditary neuropathy with liability to pressure palsies. *Nat Genet* 6:263–266.
44. Roa BB, et al. (1993) Charcot-Marie-Tooth disease type 1A. Association with a spontaneous point mutation in the *PMP22* gene. *N Engl J Med* 329:96–101.
45. Valentijn LJ, et al. (1992) Identical point mutations of *PMP-22* in Trembler-J mouse and Charcot-Marie-Tooth disease type 1A. *Nat Genet* 2:288–291.
46. Slager RE, Newton TL, Vlangos CN, Finucane B, Elsea SH (2003) Mutations in *RAI1* associated with Smith-Magenis syndrome. *Nat Genet* 33:466–468.
47. Bi W, et al. (2005) Inactivation of *Rai1* in mice recapitulates phenotypes observed in chromosome engineered mouse models for Smith-Magenis syndrome. *Hum Mol Genet* 14: 983–995.

48. Bi W, et al. (2007) *Rai1* deficiency in mice causes learning impairment and motor dysfunction, whereas *Rai1* heterozygous mice display minimal behavioral phenotypes. *Hum Mol Genet* 16:1802–1813.
49. Lupski JR (2004) Hotspots of homologous recombination in the human genome: Not all homologous sequences are equal. *Genome Biol* 5:242.
50. Reiter LT, et al. (1996) A recombination hotspot responsible for two inherited peripheral neuropathies is located near a mariner transposon-like element. *Nat Genet* 12:288–297.
51. Turner DJ, et al. (2008) Germline rates of *de novo* meiotic deletions and duplications causing several genomic disorders. *Nat Genet* 40:90–95.
52. Hastings PJ, Ira G, Lupski JR (2009) A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet* 5:e1000327.
53. Hastings PJ, Lupski JR, Rosenberg SM, Ira G (2009) Mechanisms of change in gene copy number. *Nat Rev Genet* 10:551–564.
54. Weterings E, van Gent DC (2004) The mechanism of non-homologous end-joining: A synopsis of synapsis. *DNA Repair (Amst)* 3:1425–1435.
55. Lee JA, Carvalho CM, Lupski JR (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 131:1235–1247.
56. Carvalho CM, et al. (2009) Complex rearrangements in patients with duplications of *MECP2* can occur by fork stalling and template switching. *Hum Mol Genet* 18:2188–2203.
57. Bi W, et al. (2009) Increased LIS1 expression affects human and mouse brain development. *Nat Genet* 41:168–177.
58. Zhang F, et al. (2009) The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat Genet* 41:849–853.
59. Kidd JM, et al. (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature* 453:56–64.
60. Korbel JO, et al. (2007) Paired-end mapping reveals extensive structural variation in the human genome. *Science* 318:420–426.
61. Mehan MR, Freimer NB, Ophoff RA (2004) A genome-wide survey of segmental duplications that mediate common human genetic variation of chromosomal architecture. *Hum Genomics* 1:335–344.
62. Redon R, et al. (2006) Global variation in copy number in the human genome. *Nature* 444:444–454.
63. Shaw CJ, Lupski JR (2005) Non-recurrent 17p11.2 deletions are generated by homologous and non-homologous mechanisms. *Hum Genet* 116:1–7.
64. Antonacci F, et al. (2009) Characterization of six human disease-associated inversion polymorphisms. *Hum Mol Genet* 18:2555–2566.
65. Ben-Shachar S, et al. (2009) Microdeletion 15q13.3: A locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders. *J Med Genet* 46:382–388.
66. Edelmann L, et al. (1999) A common molecular basis for rearrangement disorders on chromosome 22q11. *Hum Mol Genet* 8:1157–1167.
67. Groth M, et al. (2008) High-resolution mapping of the 8p23.1 beta-defensin cluster reveals strictly concordant copy number variation of all genes. *Hum Mutat* 29:1247–1254.
68. Inoue K, et al. (2002) Genomic rearrangements resulting in *PLP1* deletion occur by nonhomologous end joining and cause different dysmyelinating phenotypes in males and females. *Am J Hum Genet* 71:838–853.
69. Pentao L, Wise CA, Chinault AC, Patel PI, Lupski JR (1992) Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. *Nat Genet* 2:292–300.
70. Stankiewicz P, Lupski JR (2006) The genomic basis of disease, mechanisms and assays for genomic disorders. *Genome Dyn* 1:1–16.
71. Kehrer-Sawatzki H, Cooper DN (2008) Molecular mechanisms of chromosomal rearrangement during primate evolution. *Chromosome Res* 16:41–56.
72. Marques-Bonet T, et al. (2009) A burst of segmental duplications in the genome of the African great ape ancestor. *Nature* 457:877–881.
73. Kolb J, et al. (2009) Cruciform-forming inverted repeats appear to have mediated many of the microinversions that distinguish the human and chimpanzee genomes. *Chromosome Res* 17:469–483.
74. Lee JA, et al. (2006) Role of genomic architecture in *PLP1* duplication causing Pelizaeus-Merzbacher disease. *Hum Mol Genet* 15:2250–2265.
75. Jacobs GH (2008) Primate color vision: A comparative perspective. *Vis Neurosci* 25: 619–633.
76. Hunt DM, et al. (1998) Molecular evolution of trichromacy in primates. *Vision Res* 38: 3299–3306.
77. Nathans J, Piantanida TP, Eddy RL, Shows TB, Hogness DS (1986) Molecular genetics of inherited variation in human color vision. *Science* 232:203–210.
78. Nathans J, Thomas D, Hogness DS (1986) Molecular genetics of human color vision: The genes encoding blue, green, and red pigments. *Science* 232:193–202.
79. Gilad Y, Przeworski M, Lancet D (2004) Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol* 2:E5.
80. del Gaudio D, et al. (2006) Increased *MECP2* gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. *Genet Med* 8:784–792.
81. Small K, Iber J, Warren ST (1997) Emerin deletion reveals a common X-chromosome inversion mediated by inverted repeats. *Nat Genet* 16:96–99.
82. Caceres M, Sullivan RT, Thomas JW (2007) A recurrent inversion on the eutherian X chromosome. *Proc Natl Acad Sci USA* 104:18571–18576.
83. Ohno S (1970) *Evolution by Gene Duplication* (Springer-Verlag, Berlin).
84. Larkin DM, et al. (2009) Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Res* 19:770–777.
85. Innan H (2009) Population genetic models of duplicated genes. *Genetica* 137:19–37.
86. Cheng Z, et al. (2005) A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* 437:88–93.
87. Fujiyama A, et al. (2002) Construction and analysis of a human-chimpanzee comparative clone map. *Science* 295:131–134.
88. Perry GH, et al. (2007) Diet and the evolution of human amylase gene copy number variation. *Nat Genet* 39:1256–1260.
89. Fortna A, et al. (2004) Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS Biol* 2:E207.
90. Lupski JR (2007) An evolution revolution provides further revelation. *BioEssays* 29: 1182–1184.
91. Popesco MC, et al. (2006) Human lineage-specific amplification, selection, and neuronal expression of DUF1220 domains. *Science* 313:1304–1307.
92. Mefford HC, et al. (2008) Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med* 359:1685–1699.
93. Dumas L, Sikela JM (2009) DUF1220 domains, cognitive disease, and human brain evolution. *Cold Spring Harbor Symp Quant Biol*, in press.
94. Stefansson H, et al. (2008) Large recurrent microdeletions associated with schizophrenia. *Nature* 455:232–236.
95. Davy BE, Robinson ML (2003) Congenital hydrocephalus in *hy3* mice is caused by a frameshift mutation in *Hydin*, a large novel gene. *Hum Mol Genet* 12:1163–1170.
96. Doggett NA, et al. (2006) A 360-kb interchromosomal duplication of the human *HYDIN* locus. *Genomics* 88:762–771.
97. Gilbert W (1978) Why genes in pieces? *Nature* 271:501.
98. Babushok DV, Ostertag EM, Kazazian HH, Jr (2007) Current topics in genome evolution: Molecular mechanisms of new gene formation. *Cell Mol Life Sci* 64:542–554.
99. Paulding CA, Ruvolo M, Haber DA (2003) The *Tre2* (*USP6*) oncogene is a hominoid-specific gene. *Proc Natl Acad Sci USA* 100:2507–2511.
100. Inoue K, Lupski JR (2002) Molecular mechanisms for genomic disorders. *Annu Rev Genomics Hum Genet* 3:199–242.
101. Murakami T, Reiter LT, Lupski JR (1997) Genomic structure and expression of the human heme A:farnesyltransferase (*COX10*) gene. *Genomics* 42:161–164.
102. Frittoli E, et al. (2008) The primate-specific protein TBC1D3 is required for optimal macropinocytosis in a novel ARF6-dependent pathway. *Mol Biol Cell* 19:1304–1316.
103. Long M, Betran E, Thornton K, Wang W (2003) The origin of new genes: Glimpses from the young and old. *Nat Rev Genet* 4:865–875.
104. van Rijk A, Bloemendal H (2003) Molecular mechanisms of exon shuffling: Illegitimate recombination. *Genetica* 118:245–249.
105. Esnault C, Maestre J, Heidmann T (2000) Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* 24:363–367.
106. Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. *Nat Rev Genet* 7:85–97.
107. Iafrate AJ, et al. (2004) Detection of large-scale variation in the human genome. *Nat Genet* 36:949–951.
108. Sebat J, et al. (2004) Large-scale copy number polymorphism in the human genome. *Science* 305:525–528.
109. Tuzun E, et al. (2005) Fine-scale structural variation of the human genome. *Nat Genet* 37:727–732.
110. Zhang F, Carvalho CM, Lupski JR (2009) Complex human chromosomal and genomic rearrangements. *Trends Genet* 25:298–307.
111. Bailey JA, Kidd JM, Eichler EE (2008) Human copy number polymorphic genes. *Cytogenet Genome Res* 123:234–243.
112. Flores M, et al. (2007) Recurrent DNA inversion rearrangements in the human genome. *Proc Natl Acad Sci USA* 104:6099–6106.
113. Lupski JR (2007) Genomic rearrangements and sporadic disease. *Nat Genet* 39: S43–S47.
114. van Ommen GJ (2005) Frequency of new copy number variation in humans. *Nat Genet* 37:333–334.
115. Lu XY, et al. (2008) Genomic imbalances in neonates with birth defects: High detection rates by using chromosomal microarray analysis. *Pediatrics* 122:1310–1318.
116. Hall BK (1995) Atavisms and atavistic mutations. *Nat Genet* 10:126–127.
117. Cantu JM, Ruiz C (1985) On atavisms and atavistic genes. *Ann Genet* 28:141–142.
118. Figuera LE, Pandolfo M, Dunne PW, Cantu JM, Patel PI (1995) Mapping of the congenital generalized hypertrichosis locus to chromosome Xq24-q27.1. *Nat Genet* 10:202–207.
119. Macias-Flores MA, et al. (1984) A new form of hypertrichosis inherited as an X-linked dominant trait. *Hum Genet* 66:66–70.
120. Garcia-Cruz D, Figuera LE, Cantu JM (2002) Inherited hypertrichoses. *Clin Genet* 61: 321–329.
121. Sun M, et al. (2009) Copy-number mutations on chromosome 17q24.2-q24.3 in congenital generalized hypertrichosis terminalis with or without gingival hyperplasia. *Am J Hum Genet* 84:807–813.
122. Stankiewicz P, Lupski JR (2002) Molecular-evolutionary mechanisms for genomic disorders. *Curr Opin Genet Dev* 12:312–319.
123. van Rijk AA, de Jong WW, Bloemendal H (1999) Exon shuffling mimicked in cell culture. *Proc Natl Acad Sci USA* 96:8074–8079.
124. Jones JM, et al. (2000) The mouse neurological mutant flailer expresses a novel hybrid gene derived by exon shuffling between *Gnb5* and *Myo5a*. *Hum Mol Genet* 9:821–828.
125. Sudhof TC, et al. (1985) Cassette of eight exons shared by genes for LDL receptor and EGF precursor. *Science* 228:893–895.
126. Sudhof TC, Goldstein JL, Brown MS, Russell DW (1985) The LDL receptor gene: A mosaic of exons shared with different proteins. *Science* 228:815–822.
127. Moran JV, DeBerardinis RJ, Kazazian HH Jr. (1999) Exon shuffling by L1 retrotransposition. *Science* 283:1530–1534.
128. Thomson TM, et al. (2000) Fusion of the human gene for the polyubiquitination coeffector UEV1 with *Kua*, a newly identified gene. *Genome Res* 10:1743–1756.
129. Courseaux A, Nahon JL (2001) Birth of two chimeric genes in the Hominidae lineage. *Science* 291:1293–1297.
130. Ejima Y, Yang L (2003) Trans mobilization of genomic DNA as a mechanism for retrotransposon-mediated exon shuffling. *Hum Mol Genet* 12:1321–1328.
131. Babushok DV, et al. (2007) A novel testis ubiquitin-binding protein gene arose by exon shuffling in hominoids. *Genome Res* 17:1129–1138.

Heritability of reproductive fitness traits in a human population

Gülüm Kosova[a,b], Mark Abney[b], and Carole Ober[a,b,c,1]

[a]Committee on Genetics, Genomics, and Systems Biology, and Departments of [b]Human Genetics and [c]Obstetrics and Gynecology, University of Chicago, Chicago, IL 60637

Edited by Diddahally R. Govindaraju, Boston University School of Medicine, and accepted by the Editorial Board September 10, 2009 (received for review July 6, 2009)

The genetic basis of fitness traits has been studied widely in animals, yet the contribution of genetic variation to these traits in humans is controversial. In particular, it is difficult to disentangle genetic versus environmental effects on fertility, because of within-family correlations of sociocultural, economic, and other nongenetic factors that influence family sizes. In this study, we investigated the genetic architecture of reproductive fitness traits in a fertile human population whose communal lifestyle assures uniform and equal access to resources. Our study revealed significant heritabilities for reproductive traits in both men and women, after accounting for common household effects shared among siblings and demographic changes in reproductive practices. Furthermore, our results indicate that both autosomal and X-linked additive and dominance variances contribute to these traits. We therefore propose that reproductive traits should be amenable to genetic mapping studies, and the results we present here will facilitate the search for the novel genes influencing natural fertility in humans.

life history traits | human fertility

Reproductive fitness reflects the ability of individuals to pass on their genes to subsequent generations. Fitness traits, also referred to as life-history traits, include measures of fertility and mortality and are complex phenotypes that are direct targets of Darwinian selection. Understanding the genetic basis of variation in these traits and inheritance in animals has long been a central theme in evolutionary biology (1). However, partitioning the observed variation into the genetic and environmental sources, and therefore determining the heritability of these traits, remains challenging in humans. As a result, current theories on the evolution and heritability of fitness, and the empirical data, come largely from animal studies (for examples, see refs. 2–8). However, studies of model organisms suggest that hundreds of genes influence fertility in mammals (9). Standing variation in any of those genes could contribute to interindividual differences in fitness in natural populations.

The difficulty in assessing genetic contributions to human fertility is caused in part by the fact that human family sizes are often deliberately limited, with few populations reaching their true reproductive potential, and because the many nongenetic factors that influence human family size are often shared within families. As a result, disentangling the effects of shared genes from shared environment is often impossible. For example, parent–child correlations in family sizes have been reported for a number of human populations (10–16). However, in nearly all of these studies, the investigators concluded that social or cultural transmission, such as patterns of emigration (10), polygyny and higher male mortality (11), education or marital age (12), or differential access to resources and ability to acquire a mate (15), but not genetic factors, accounted for the observed intergenerational correlations in family size. One exception was a recent study in the Hutterites, which attributed parent–offspring correlation in family sizes to genetic causes (16).

The Hutterites are a young founder population of European descent that are particularly amenable to studies of reproductive fitness. The proscription of contraception and desire for large families, resulted in median completed family sizes (CFSs) >10 and interbirth intervals <2 years in the 1960s (17, 18). Moreover, they practice a communal agrarian lifestyle, which ensures that all members are exposed to a relatively similar environment and have equal access to resources, including wealth, education, and medical care. In an earlier study, intergenerational correlations in family size were measured in 161 three-generation completed Hutterite families (16). Significant correlations between the family size of a couple and that of their sons ($r = 0.29$; $P < 10^{-6}$) and their daughters ($r = 0.18$; $P = 0.0041$) were reported. Because socio-cultural factors known to influence family sizes are remarkably uniform between the Hutterite families (19), observed correlations were interpreted as evidence for genetic contributions (i.e., heritability) to this trait (16).

Here, we defined three measures of fertility in the extant population to assess different components of a couple's reproductive fitness (Table 1). These traits include measures of reproductive capacity [(CFS and age at last reproduction (ALR)] and reproductive rate (birth rate). These measures were corrected for age and cohort effects and length of the reproductive period when relevant (see *Methods*). We also modeled the effects of shared household environment for each trait. The genetic variance for these reproductive fitness measures were formally estimated in Hutterite men and women with proven fertility, who are members of a single 13-generation pedigree (20). The results reported here lay the foundation for future studies to identify novel genes that influence natural fertility and contribute toward theoretical considerations on the evolution of fitness traits.

Results

Reproductive Fitness Traits in the Hutterites. The characteristics of the reproductive phenotypes in ≈450 Hutterite couples are shown in Table 1. In general, there is a wide range of variability in each of the traits. CFSs range between 1 and 17 (Fig. 1) and, not surprisingly, were significantly correlated with the number of years from marriage to last birth ($r = 0.87$, $P < 0.0001$; Fig. 2*A*). Wife's age at marriage was a significant predictor of CFS and ALR ($P < 0.0001$), with fewer births and later ALR at later ages at marriage (Fig. 2 *B* and *C*). Changes in the reproductive practices over time were also remarkable, with earlier maternal

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Evolution in Health and Medicine" held April 2–3, 2009, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS web site at www.nasonline.org/Sackler_Evolution_Health_Medicine.

Author contributions: G.K. and C.O. designed research; G.K. performed research; M.A. contributed new reagents/analytic tools; G.K. analyzed data; and G.K., M.A., and C.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. D.R.G. is a guest editor invited by the Editorial Board.

[1]To whom correspondence should be addressed. E-mail: gulum@uchicago.edu

This article contains supporting information online at www.pnas.org/cgi/content/full/0906196106/DCSupplemental.

Table 1 Definitions and characteristics of three reproductive phenotypes

Phenotype	Definition	Sample size	Mean ± SD	Range
CFS*†	Total number of births (liveborn and stillborn births were included)	353	7.14 ± 3.11 births	1 to 17 births
Birth rate†‡	Number of births per year of marriage	459	0.53 ± 0.14 births	0.21 to 1.05 births
ALR*†	Age at which the wife had her last child	353	35.07 ± 5.31 years	22.51 to 47.13 years

For all phenotypes we counted multiple births as one birth. There were 33 multiple births in our sample (32 twins, 1 triplet), 30 in completed families and 3 in incomplete families.
*Wife's age at marriage included as a covariate in analyses (mean 22.45, SD 2.62, range 17.57 to 35.15 years).
†Wife's birth year included as a covariate in analyses (mean 1950, SD 16.12, range 1899 to 1984).
‡Number of years from marriage to last birth included as a covariate in analyses (mean 12.27, SD 5.31, range 0.78 to 25.86 years).

birth years associated with more births, higher rates of reproduction, and later ALRs ($P < 0.0001$; Fig. 2 *D–F*). Inclusion of wife's age at marriage and wife's birth year as covariates in the multivariate model assured that the residuals of the phenotypes are independent of these demographic variables and pedigree depth (see *Methods* and Figs. S1 and S2).

Birth rates are more variable in smaller completed families (≤7 children; Fig. 3*A*). In larger families, however, higher reproductive rates and less variation are observed. Similarly, birth rates increase with increasing years of marriage (longer reproductive period), particularly for larger family sizes (Fig. 3*B*). Neither wife's age at menarche nor birth control use (ever) was a significant predictor of any fertility measure in a subset of 399 and 456 couples, respectively, for whom this information was available, and these covariates were not considered further.

Correlations between reproductive fitness traits are shown in Fig. 4. CFS was correlated with the other two traits ($r > 0.50$; $P < 0.0001$). However, no correlation was observed between birth rate and ALR ($r < 0.10$; $P > 0.05$), suggesting that two distinct components of reproductive fitness are captured by these measures: one by the measures of CFS and ALR, which might reflect the reproductive capacity of a couple, and a second by birth rate, which might reflect reproductive success and/or gamete quality.

Heritability Estimates of Reproductive Fitness Traits. We estimated heritabilities by using a variance-component, maximum-likelihood method, developed for large, inbred pedigrees (20, 21), and evaluated models with autosomal additive, autosomal dominance, and X-linked additive variance components, in addition to an environmental variance component that included a shared household effect (see *Methods*). The full and most parsimonious models for each trait in women and men are shown in Table 2.

CFS. In women, a model including X-linked additive and environmental variance is favored the most [based on the Akaike Information Criterion (AIC) score and likelihood ratio test], but was only borderline significant compared with a model including environmental variance only ($P = 0.056$), yielding narrow and broad heritabilities of 0.22. Inclusion of other variance components in the model (autosomal additive, dominance and shared household) did not contribute significantly to heritability of this trait and was not favored statistically. In men, however, there was significant heritability caused by both autosomal and X-linked additive and dominance variances ($P = 0.0006$), yielding a narrow heritability ($h^2_A + h^2_x$) of 0.20 and broad heritability of 0.68. To address the possibility that the observed dominance effect is caused by the shared household among brothers, we also evaluated a model that includes a shared household variance component in addition to a dominance variance component. The contribution of shared household was not significant in the presence of dominance variance.

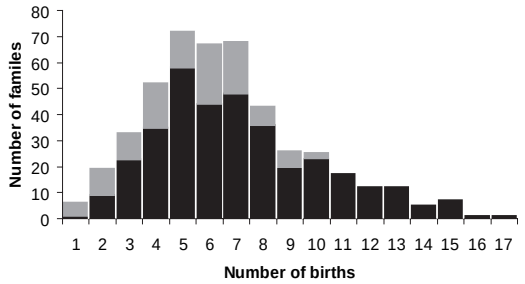


Fig. 1. Family sizes of the Hutterite couples included in this study. Black bars represent the sizes of completed families ($n = 353$; see *Methods* for the definition of completed families); gray bars represent the sizes of incomplete families ($n = 112$).

Birth rate. Similar to the results of CFS, X-linked additive is the main source of genetic variance for birth rate in women, yielding narrow and broad heritabilities of 0.28 ($P = 0.033$). In men, however, a model including all three genetic variance components was favored ($P = 0.021$), giving a total narrow heritability ($h^2_A + h^2_x$) of 0.21 and broad heritability of 0.54. Inclusion of a shared household effect reduced the estimate for the dominance variance slightly; however, this did not result in improvement in the overall heritability model (based on the AIC score and likelihood ratio test).

ALR. In both women and men, autosomal additive variance accounted for all of the genetic variance of this phenotype, with narrow and broad heritabilities of 0.23 in women ($P = 0.019$), and 0.34 in men ($P = 0.0001$).

Discussion

Reproductive fitness traits are complex phenotypes influenced by environmental and genetic factors and have been studied extensively in both animal and human populations (2–8, 22). Despite this fact the extent to which these traits are shaped by genetic forces (or the nature of this action) in humans remains controversial (e.g., refs. 1 and 23), mainly because of the difficulties in disentangling the effects of shared social and environmental factors from shared genes within human families, and in assessing the true reproductive potential of the individuals. We demonstrate here in a human population with equal access to resources, shared cultural practices, and among the highest fertility rates ever recorded (18), that both autosomal and X-linked additive variances and autosomal dominance variance contribute to the genetic architecture of reproductive fitness traits.

All of the traits considered in this study had significant narrow and broad heritabilities in both women and men. In women, birth rate had a significant X-linked additive variance, whereas ALR had a significant autosomal additive component (Table 2), consistent with the fact that ALR and birth rate are not themselves correlated ($r = 0.03$; Fig. 4) and further suggesting that they measure different components of reproductive fitness. CFS in women, however, was only marginally heritable, with possible X-linked additive variance. The smaller heritability estimate for CFS indicates a larger role for nongenetic factors in determining family size in women compared with either the rate

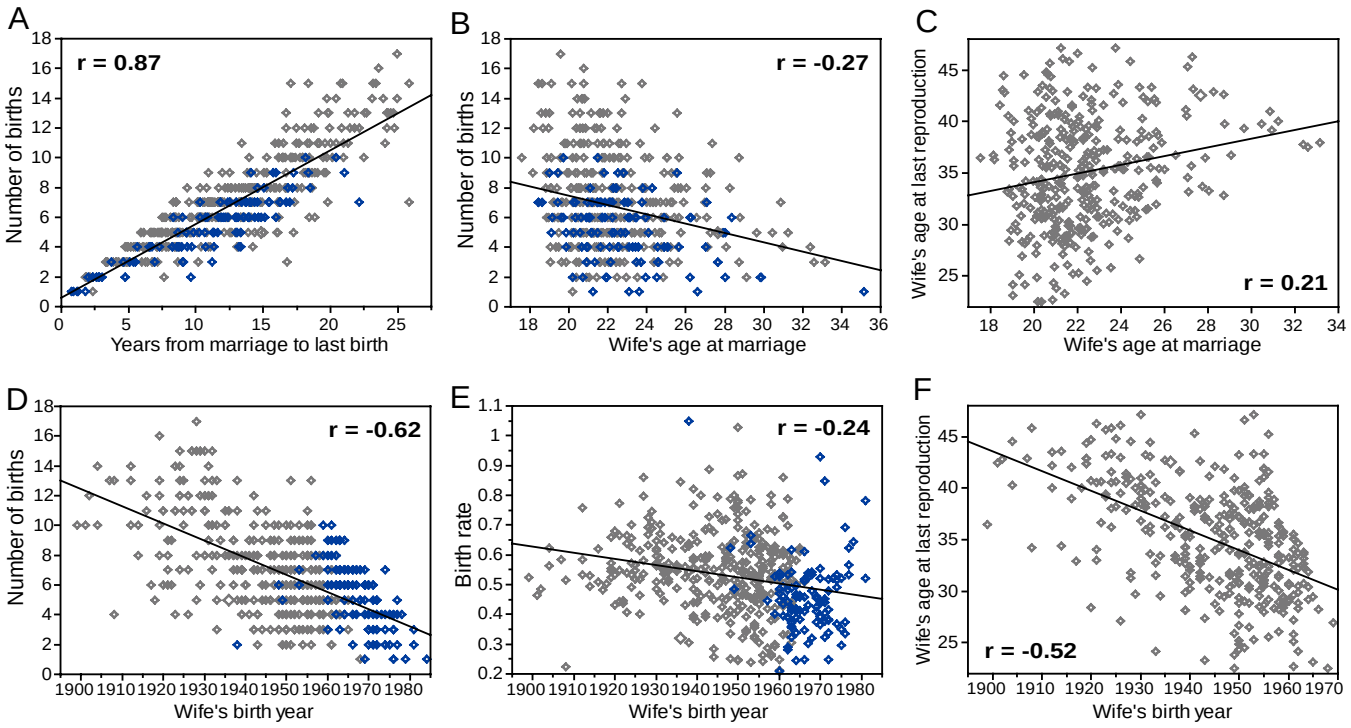


Fig. 2. Correlations between fertility traits and their significant covariates (see Table 1 legend). Gray diamonds represent completed families, blue diamonds represent incomplete families. Only traits for which the covariates were significant are shown. Pearson's correlation coefficient, *r*, is reported for each plot. (*A*) Number of years from marriage until last birth was included as a covariate to correct for length of the reproductive period. (*B* and *C*) Wife's age at marriage was included to correct for maternal age effects. (*D–F*) Wife's birth year was included to correct for cohort effects (48).

of reproduction or the women's ALR. This finding can be interpreted in the context of Price and Schluter's argument (8), which posits that complex phenotypes are composite traits determined by multiple phenotypes with less complex architecture. In this example, birth rate and ALR can be considered components of the composite traits of CFS. The heritability of CFS would therefore include all of the environmental variances affecting their individual components and additional environmental variation, resulting in an overall greater proportion of environmental variance, and thus, smaller heritability, contributing to the composite traits (8). In this respect, birth rate and ALR can be viewed as independent fitness components contributing to family size, and the smaller heritability of CFS might be the result of the cumulative and increased role of environmental variances in this trait in women. The implication of X chromosome genes for birth rate and possibly for CFS is particularly intriguing, given that the X chromosome is enriched for genes associated with sexual development and reproduction (24). In addition, determinants of ovarian function map to the X chromosome, including genes for folliculogenesis, premature ovarian failure, and infertility (for example, see refs. 25 and 26).

However, at least one variance component contributes significantly to all male fitness traits (Table 2). The heritability of ALR was caused exclusively by autosomal additive variance, whereas both autosomal and X-linked additive and dominance variances contribute to genetic architecture of CFS and birth rate. It was surprising to us that the wife's ALR, which is highly correlated with the husband's ALR ($r = 0.98$), was heritable in men. Although declines in sperm quality with age have been reported (27), this decline has usually been attributed to environmental causes, such as prenatal or postnatal exposures to hormones and chemical compounds (refs. 28 and 29 and references therein), and not generally to genetic factors. Our results suggest that genetic factors directly contribute to the age-related decline in sperm quality in men, or they determine sensitivity to environmental exposures that affect sperm quality.

It is challenging to partition the total genetic variance between variance components, because the effects captured by one model might possibly be confounded by another. In particular, one could expect that the power to distinguish the effects of autosomal and X-linked additive genetic variances in females may be low, because of their similar inheritance patterns. Likewise, shared household effects might mimic dominance variance, because dominance effects are largely driven by correlation among siblings. Nonetheless, the unusually large Hutterite pedigree structure, and inclusion of all pairs of individuals who are related to each other through multiple lines of descent, allows us to estimate these effects simultaneously. In particular, the closed nature of the Hutterite population and the small number of founding genomes result in a nonzero probability of any two individuals sharing both of their alleles identical by descent (IBD) (30), a situation that would not be expected in many human populations. For example, among the 107,880 pairwise relationships between the 465 wives included in this study, there are 313 sib pairs and 107,567 nonsib pairs with dominance variance coefficients >0. Likewise, for 465 husbands, there are 493 sib pairs and 106,611 nonsib pairs with dominance variance coefficients >0. The sheer number of nonsib relatives, each contributing a small effect, allows us to differentiate between dominance versus (nongenetic) household effects. Indeed, if these two variance components were estimating the same effect, the variance would be shared between them, and varying the initial values for the maximum-likelihood procedure would alter the distributions of the variance estimates between these components. However, this was not the case; multiple iterations of maximum likelihood with different starting points all converged to the same estimates in our dataset. Therefore, the unique Hutterite pedigree structure and the known relationships be-

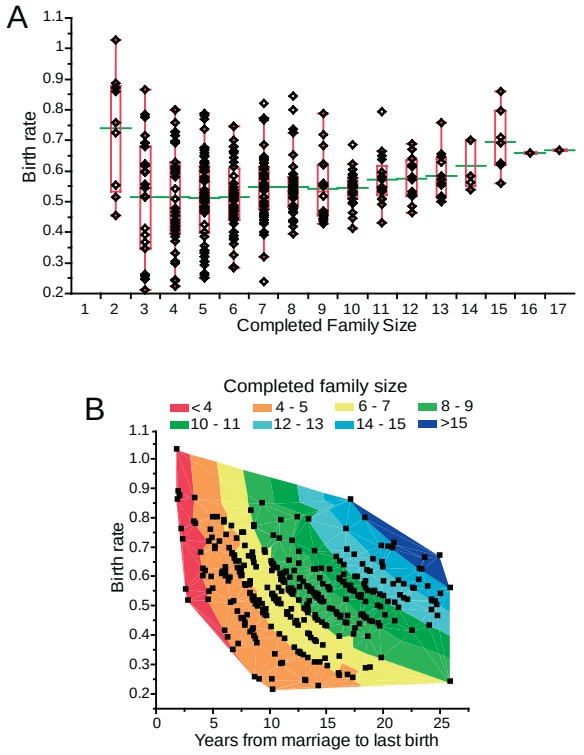


Fig. 3. Changes in birth rate by CFS (*A*) and years from marriage to last birth (*B*) stratified by the CFSs. Green horizontal lines in *A* show the mean value of the trait.

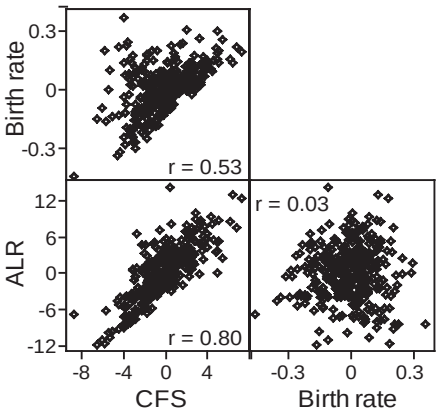


Fig. 4. Correlations between the residuals of different fertility traits. Pearson's correlation coefficient, *r*, is reported for each correlation.

tween all individuals enabled us to separate effects caused by autosomal and X-linked sources (as in female birth rate vs. ALR) and caused by dominance and shared household (as in male CFS and birth rate). However, precise quantification of the relative contributions of all variance components to the overall phenotypic variance of these traits remains challenging, because the increased number of parameters (hence, degrees of freedom) in these tests results in larger estimates of the standard errors and reduces the power to detect significant effects when all of the possible components are included in the model simultaneously.

It is noteworthy that males have overall higher heritability for reproductive fitness traits than females. This finding is also consistent with results of our earlier study in the Hutterites, in which the correlation between family sizes of a couple and their sons was higher and more significant than the correlation between family sizes of a couple and their daughters (16). This result may be caused by the overall larger number of genes involved in male compared with female reproductive processes (31) and, therefore, their combined effects may account for more of the observed phenotypic variation between the males. Alternatively, environmental factors affecting female fitness likely differ from those affecting male fitness, which might result in larger proportions of the total variance attributed to environmental variance in women (32). In addition, the effects of genetic models other than those considered here may contribute to variation in fitness. In particular, our analyses focused on male-specific and female-specific traits. We did not simultaneously consider partner effects, although parental combinations of some genes, such as HLA, are known to influence reproductive outcomes this population (33). It is possible, even likely, that parental combination of other genes also contribute to reproductive fitness in this population.

The results we report here are also consistent with previous studies and evolutionary predictions in several respects. First, the narrow heritabilities we estimated for our reproductive fitness traits compare well with those derived from studies of wild animal populations (e.g., refs. 7 and 32) and preindustrial humans (22). For example, narrow heritabilities for fitness traits estimated in various animal populations are low, ranging between 0 and 0.30 (7, 32). Pettay et al. (22) studied fitness traits in preindustrial Finns and were able to show significant narrow heritabilities in women ($h^2 = 0.18 - 0.76$) for traits similar to those considered in this study, although they did not estimate dominance or X-linked variance components. However, their heritability estimations were higher and more significant for female traits compared with male traits. Such differences, however, are not necessarily unexpected given that the Pettay study was conducted in individuals living in "premodern" conditions; whereas the Hutterites in our study enjoy a modern lifestyle, including access to 20th- and 21st-century health care, and heritability estimates for the same traits are expected to differ in different environments. Nevertheless, both studies suggest the presence of significant heritabilities of reproductive fitness traits. Second, traits closely associated with fitness are expected to exhibit lower narrow heritabilities than morphological and physiological traits [a common interpretation of Fisher's fundamental theorem of natural selection (34)], and this prediction has been supported by numerous comparative studies of heritability estimates in animals (3, 7, 35). Even though it is beyond the scope of this study to compare the heritabilities of fitness versus nonfitness traits, we observed that the narrow heritabilities for the reproductive fitness traits considered here (mean $h^2 = 0.25$, range 0.20 to 0.34) are generally lower than the narrow heritabilities we previously reported for 20 quantitative physiological and anthropometric traits (also with significant genetic variance components) in the same population (mean $h^2 = 0.47$, range 0.16 to 0.81) (21). Last, our results also support theories that argue for significant contributions of nonadditive genetic factors, especially dominance variance, in the overall genetic architecture of fitness traits (e.g., refs. 2 and 23), as observed in male CFS and birth rate. Thus, these results are consistent with those of Crnokrak and Roff (2) that showed the presence of significant dominance components for fitness traits studied in wild animal species. Our study further supports a role for dominance variance in the genetic architecture of fitness traits in humans and suggests that ignoring nonadditive genetic variance components may lead to significant underestimates of the total heritability for fitness traits in natural populations.

The effects of a founder event and inbreeding on shaping the genetic variance are also a subject of debate. Goodnight (36) argues that founder effects may result in conversion of nonadditive (i.e., epistatic) variance into additive variance. We do not

Table 2 Variance component and heritability estimates of the fitness traits in females and males

Trait	Model	Estimate for variance components, SE				Heritability estimates, SE			P value
		Environment	Additive	Dominance	X-linked additive	h^2_A	h^2_X	H^2	
				Females					
CFS	E, A, D, X	4.590 (1.807)	0.000 (1.161)	0.565 (2.211)	1.225 (1.703)	0.00 (0.18)	0.20 (0.23)	0.28 (0.26)	0.2911
	E, X	5.054 (0.510)			0.699 (0.430)		0.22 (0.12)	0.22 (0.12)	0.0559
Birth rate	E, A, D, X	11.665 (3.696)	0.000 (2.446)	0.000 (4.524)	2.218 (1.947)	0.00 (0.14)	0.27 (0.19)	0.27 (0.21)	0.2089
	E, X	11.670 (1.061)			2.211 (0.988)		0.28 (0.10)	0.28 (0.10)	0.0330
ALR	E, A, D, X	15.043 (5.984)	4.230 (4.056)	0.000 (7.347)	0.014 (2.504)	0.19 (0.21)	0.05 (0.24)	0.23 (0.28)	0.1392
	E, A	15.049 (2.147)	4.305 (2.212)			0.23 (0.11)		0.23 (0.11)	0.0192
				Males					
CFS	E, A, D, X	1.822 (1.593)	0.624 (0.849)	2.783 (2.300)	0.534 (0.818)	0.11 (0.14)	0.09 (0.13)	0.68 (0.27)	0.0006
Birth rate	E, A, D, X	6.512 (3.229)	1.841 (1.756)	4.724 (4.595)	1.075 (1.606)	0.13 (0.12)	0.08 (0.11)	0.54 (0.23)	0.0211
ALR	E, A, D, X	12.932 (4.888)	6.407 (3.005)	0.000 (6.703)	0.000 (1.062)	0.34 (0.14)	0.00 (0.10)	0.34 (0.24)	0.0009
	E, A	12.933 (2.044)	6.407 (2.421)			0.34 (0.11)		0.34 (0.11)	0.0001

For each trait, a full model, in which all the variance components were tested simultaneously, are shown first; followed by the most parsimonious model (if different from the full model) that captures the total genetic variance with the fewest variance components. P values are obtained by χ^2 likelihood ratio test against the model that includes environmental variance only. E, environmental; A, autosomal additive; D, dominance; X, X-linked additive variance components; h^2_A, narrow heritability caused autosomal additive effects; h^2_X, narrow heritability caused by X-linked additive effects; H^2, broad heritability.

know the extent to which epistatic interactions affects our phenotypes; however, we think it is unlikely to be the case because estimates of heritabilities for ≥20 physiological and anthropometric traits in the Hutterites are quite similar to estimations for those same traits in other populations (21, 30, 37, 38). If there was an inflation of the additive genetic variance component caused by the Hutterite founding event, it should affect traits broadly and not be limited to reproductive traits. However, one could argue that because inbreeding increases the frequency of homozygosity for recessive alleles and, hence, the variation caused by these genes (39), our estimates of dominance variance in the Hutterites could be inflated. However, heritability estimates of nonreproductive traits in the Hutterites showed significant contributions of additive variance to most, and dominance variance to only a few, of those traits (21, 30, 37, 38). Therefore, we do not think that a systematic bias caused by either a founder effect or inbreeding exists in this sample and that the presence of dominance variance in the Hutterites represents true nonadditive genetic effects on reproductive fitness traits. Last, the Hutterite communal lifestyle results in a remarkably uniform environment, particularly with regard to sociocultural factors that affect family sizes, which likely maximizes the effects of genetic variance on phenotypic variance in this population.

However, how genetic variation that influences human fertility is maintained in a population over generations is an intriguing question. It could be caused by pleiotropic effects of the contributing genes, if an allele, or tightly linked variation that is beneficial for reproduction has detrimental effects on other physiological processes or at different stages of the lifecycle. Such fitness tradeoffs are similar to those proposed in Williams' antagonistic pleiotropy hypothesis (40). Alternatively, there may be many genes influencing fertility, each with small effects and individually contributing very little to overall fitness. In that case, selection acting on individual genes may be too weak to drive these alleles to fixation or elimination.

In summary, the correlations in reproductive fitness traits revealed by this study reflect an underlying genetic architecture of male and female fertility and indicate that these traits should be amenable to genetic mapping studies to identify novel genes influencing natural variation in reproductive fitness. As specific fertility genes are identified, and their functions are elucidated, it may be possible to directly examine how genetic diversity has shaped variation in human fitness traits. Furthermore, we suggest that more severe mutations in genes associated with normal variation in fertility may also account for some proportion of infertility, which is present in ≈10% of the general population (41).

Methods

Subjects. The Hutterites are a young founder population who originated in the South Tyrol in the 16th century (19, 42). In the 1870s, ≈900 Hutterites migrated from Europe to the United States (19, 42), and today their >40,000 descendants live on communal farms (called colonies) in the northern United States and western Canada. The subjects of this study are 525 Hutterite couples living in South Dakota, all of whom can be traced back to 62 ancestors who were born in the early 1700s to 1800s (43). The Hutterites in our studies are related to each other through multiple lines of descent in a 13-generation pedigree consisting of 3,028 individuals (20). The mean pedigree depth was 7.60 ± 0.62 generations for the husbands and 7.68 ± 0.62 generations for the wives in our study.

The mean inbreeding coefficient of these individuals is 0.034 (SD 0.015), approximately equivalent to that of first cousins once removed (1½ cousins). Despite this high level of inbreeding, the Hutterites are among the most fertile human populations with relatively few (≈2%) childless couples and small interbirth intervals (18). Moreover, the Hutterites' communal agrarian lifestyle ensures that all individuals have similar environmental exposures and equal access to resources. In particular, the limited use of contraception and uniform desire for large families result in large sibships (e.g., see Fig. 1). Last, the Hutterites are strictly monogamous, although second (or third) marriages occur after the death of a spouse. These unique features make the Hutterites ideally suited for genetic studies of fertility, because their family sizes and rates of conception may reveal the true human reproductive potential (18).

Sample Composition. We obtained birth, death, and marriage dates from records compiled by the Hutterite ministers. In addition, reproductive history interviews were conducted in person by C.O. with 525 ever married women during field trips to Hutterite colonies between 1982 and 2007 (33, 44, 45). All births were updated to at least 2002 for these women. These interviews elicited information on births, miscarriages (approximate dates and gestational ages), infertility (>1 year inability to conceive or use of infertility treatment), birth control use (type, dates, and duration), ages at menarche and last menses, surgical sterilization, medication use, and maternal illnesses that could affect fertility. Of the 525 interviewed couples, 60 couples were excluded from this study for one of the following reasons: the couple was childless or conceived after treatment for infertility (n = 20), conception before marriage (the first child was born before or within 28 weeks after the marriage; n = 30), medical conditions in the wife could have limited her fertility (Rh incompatibility, severe arthritis, severe depression, ovarian cancer, and ovariectomy; n = 7), or incomplete information for the couple (n = 3). The remaining 465 couples were considered for the analyses of male and female reproductive traits.

The wives represented 267 full sibships and one half sibship (sisters having the same father but different mother); the husbands represented 212 full sibships and one half sibship (brothers having the same father but different mother). Three wives and seven husbands included in this study reported a second marriage. However, in six of these cases (two women and four men),

the subjects had completed their families (see below) in their first marriage and did not have any children with their second spouses. In those cases, only the first marriages were considered. One woman had children in two marriages. Only the data on her first marriage were used, but her family size was considered incomplete. For the remaining three men, reproductive history interviews were available for their second wives only; therefore only the data on their second marriages were used, but these three men were excluded from the analyses of CFS. Because our studies in the Hutterites are population-based and participation within each colony was high (>95%), there are no known ascertainment biases that could affect the interpretation of our results.

Measures of Reproductive Fitness. The definitions of the reproductive fitness measures considered in this study are shown in Table 1. We defined families as "completed" if either the wife was >45 years of age and was not widowed before then, or the couple had not had a child in >6 years ($n = 353$). For birth rate, we first determined total interbirth interval for each couple having two or more children. Birth rate is then calculated as [(the number of births − 1)/(total interbirth intervals)]. Six couples had only one child at the time of the analysis and were therefore excluded from the analyses of birth rate.

For all fitness traits, we fit a multivariate linear regression model and included as covariates the wife's age at marriage to correct the maternal age effects (Fig. 2 *B* and *C*) and wife's birth year to correct for the demographic changes in reproductive behaviors (Fig. 2 *D–F*) and pedigree depth (Fig. S2). Because there is a high correlation between the wife's age and husband's age in our sample ($r = 0.98$), we used only the wife's age in our analyses of both male and female fertility. In addition to these two covariates, number of years from marriage to last birth was also included for the analyses of birth rate to correct for the length of the reproductive period. Residuals of all of the traits were normally distributed.

Estimating the Heritability of the Reproductive Fitness Traits. We considered male and female fertility separately because of the prior expectation that different biological processes influence natural variation in fertility between the sexes (31). Furthermore, because the wife in each couple has different degrees of relatedness to all other Hutterites than the husband of each couple, the heritabilities can be estimated independently for males and females.

Heritabilities were estimated by using the 3,028-person Hutterite pedigree, using a variance component, maximum-likelihood method, as described (20). Briefly, we modeled each phenotype as a multivariate normal, with mean $X\beta$ and covariance Σ, where X is a matrix of covariates and β is a vector of effect sizes. The matrix Σ is given by $\Sigma = 2\Phi_a\sigma_a^2 + \Delta\sigma_d^2 + 2\Phi_x\sigma_x^2 + H\sigma_s^2 + I\sigma_e^2$, where Φ_a and Φ_x are the autosomal and X-linked kinship coefficient matrices, respectively, Δ is the probability of individuals sharing two alleles IBD, and H and I are the shared household and identity matrices, respectively. Environmental and "shared household" variance components were considered for modeling the nongenetic effects; autosomal additive, autosomal dominance, and X-linked additive variance components were considered for modeling the genetic effects. Estimations of variance component parameters for the genetic effects are explained in detail elsewhere (20, 21). To estimate shared household effects, we created a matrix such that full and half-sib pairs were scored as 1, and all other pairs were scored as 0. We considered full and half-sibs equivalent for the purpose of this analysis because full sibs and half sibs are raised in the same household. There was only one family with half sibs in our study.

For each fertility trait, we fit environmental, autosomal additive, X-linked additive, and dominance variance components simultaneously. The traits for which the dominance variance was significant were also evaluated with models that included a shared household variance component. In addition, we evaluated reduced models for each trait, by excluding the nonsignificant variance components from the model, to obtain a more precise estimate of the significant components. Different variance component models were compared based on their AIC and Bayesian Information Criterion scores (46, 47), and the model that captured the most variance with fewest variance components was reported as the most "parsimonious" model (Table 2). Autosomal and X-linked narrow heritabilities (h_A^2, h_x^2) were calculated for all models as $h_A^2 = (1 + f_a)V_A/V_T$ and $h_x^2 = (1 + f_x)V_X/V_T$, where f is the autosomal or X-linked average inbreeding coefficient of the population, and V_A, V_X and V_T are the autosomal additive, X-linked additive and total phenotypic variances, respectively. Last, broad heritabilities (H^2) were calculated as $H^2 = 1 - [(V_E + V_S)/V_T]$, where V_S is the variance caused by shared household effect, when included. P values were calculated by χ^2 likelihood ratio test against the model including environmental variance only. Using this method previously with >20 quantitative traits, we have shown that heritability estimates for anthropometric and physiologic phenotypes in the Hutterites are similar to estimates in other populations (21, 30, 37, 38).

ACKNOWLEDGMENTS. We thank Rebecca Anderson, Jessica Chong, Gaixin Du, Josef Jurek, Lin Pan, and Ying Sun for technical assistance and the Hutterites for their participation. This work was supported by National Institutes of Health Grants HD21244 to C.O. and HG02899 to M.A.

1. Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. *Nature* 452:169–175.
2. Crnokrak P, Roff DA (1995) Dominance variance: Associations with selection and fitness. *Heredity* 75:530–540.
3. Houle D (1992) Comparing evolvability and variability of quantitative traits. *Genetics* 130:195–204.
4. Kruuk LE, et al. (2000) Heritability of fitness in a wild mammal population. *Proc Natl Acad Sci USA* 97:698–703.
5. McCleery RH, et al. (2004) Components of variance underlying fitness in a natural population of the great tit Parus major. *Am Nat* 164:E62–E72.
6. Merila J, Sheldon BC (2000) Lifetime reproductive success and heritability in nature. *Am Nat* 155:301–310.
7. Mousseau TA, Roff DA (1987) Natural selection and the heritability of fitness components. *Heredity* 59:181–197.
8. Price T, Schluter D (1991) On the low heritability of life-history traits. *Evolution (Kans)* 45:853–856.
9. Matzuk MM, Lamb DJ (2002) Genetic dissection of mammalian fertility pathways. *Nat Cell Biol* 4(Suppl):41–49.
10. Austerlitz F, Heyer E (1998) Social transmission of reproductive behavior increases frequency of inherited disorders in a young-expanding population. *Proc Natl Acad Sci USA* 95:15140–15144.
11. Helgason A, et al. (2003) A populationwide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: Evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. *Am J Hum Genet* 72:1370–1388.
12. Imaizumi Y, Nei M, Furusho T (1970) Variability and heritability of human fertility. *Ann Hum Genet* 33:251–259.
13. Neel JV (1970) Lessons from a "primitive" people. *Science* 170:815–822.
14. Pearson KAL, Bramley-Moore L (1899) On the inheritance of fertility in mankind. *Philos Trans R Soc London Ser B* 192:282–330.
15. Madrigal L, Relethford JH, Crawford MH (2003) Heritability and anthropometric influences on human fertility. *Am J Hum Biol* 15:16–22.
16. Pluzhnikov A, et al. (2007) Correlation of intergenerational family sizes suggests a genetic component of reproductive fitness. *Am J Hum Genet* 81:165–169.
17. Mange AP (1964) Growth and inbreeding of a human isolate. *Hum Biol* 36:104–133.
18. Sheps MC (1965) An analysis of reproductive patterns in an American isolate. *Popul Stud* 19:65–80.
19. Hostetler J (1974) *Hutterite Society* (John Hopkins Univ Press, Baltimore).
20. Abney M, McPeek MS, Ober C (2000) Estimation of variance components of quantitative traits in inbred populations. *Am J Hum Genet* 66:629–650.
21. Pan L, Ober C, Abney M (2007) Heritability estimation of sex-specific effects on human quantitative traits. *Genet Epidemiol* 31:338–347.
22. Pettay JE, et al. (2005) Heritability and genetic constraints of life-history trait evolution in preindustrial humans. *Proc Natl Acad Sci USA* 102:2838–2843.
23. Merila J, Sheldon BC (1999) Genetic architecture of fitness and nonfitness traits: Empirical patterns and development of ideas. *Heredity* 83:103–109.
24. Saifi GM, Chandra HS (1999) An apparent excess of sex- and reproduction-related genes on the human X chromosome. *Proc Biol Sci* 266:203–209.
25. Toniolo D, Rizzolio F (2007) X chromosome and ovarian failure. *Semin Reprod Med* 25:264–271.
26. Vaiman D (2002) Fertility, sex determination, and the X chromosome. *Cytogenet Genome Res* 99:224–228.
27. Kidd SA, Eskenazi B, Wyrobek AJ (2001) Effects of male age on semen quality and fertility: A review of the literature. *Fertil Steril* 75:237–248.
28. Carlsen E, et al. (1992) Evidence for decreasing quality of semen during past 50 years. *Br Med J* 305:609–613.
29. Swan SH, Elkin EP, Fenster L (1997) Have sperm densities declined? A reanalysis of global trend data. *Environ Health Perspect* 105:1228–1232.
30. Abney M, McPeek MS, Ober C (2001) Broad and narrow heritabilities of quantitative traits in a founder population. *Am J Hum Genet* 68:1302–1307.
31. Torgerson DG, Whitty BR, Singh RS (2005) Sex-specific functional specialization and the evolutionary rates of essential fertility genes. *J Mol Evol* 61:650–658.
32. Visscher PM, Hill WG, Wray NR (2008) Heritability in the genomics era: Concepts and misconceptions. *Nat Rev Genet* 9:255–266.
33. Ober C, et al. (1998) Human leukocyte antigen matching and fetal loss: Results of a 10-year prospective study. *Hum Reprod* 13:33–38.
34. Fisher RA (1930) *The Genetical Theory of Natural Selection* (Clarendon, Oxford).
35. Gustafsson L (1986) Lifetime reproductive success and heritabilities: Empirical suport for Fisher's fundamental theorem. *Am Nat* 128:761–764.
36. Goodnight CJ (1988) Epistasis and the effect of founder events on the additive genetic variance. *Evolution (Lawrence, Kans)* 42:441–454.
37. Ober C, Abney M, McPeek MS (2001) The genetic dissection of complex traits in a founder population. *Am J Hum Genet* 69:1068–1079.
38. Weiss LA, et al. (2006) The sex-specific genetic architecture of quantitative traits in humans. *Nat Genet* 38:218–222.

39. Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics* (Longman, New York), 4th ed.
40. Williams GC (1957) Pleiotropy, natural selection, and the evolution of senescence. *Evolution (Lawrence, Kans)* 11:398–411.
41. Behrman SJ, Kistner RN (1975) *Progress in Infertility* (Little, Brown & Co., Boston), 2nd Ed.
42. Steinberg AG, et al. (1967) Genetic studies in an inbred human isolate. *Proceedings of the Third International Congress of Human Genetics*, eds Crow JF, Neel JV (Johns Hopkins Univ Press, Baltimore), pp 267–290.
43. Ober C, et al. (1997) HLA and mate choice in humans. *Am J Hum Genet* 61:497–504.
44. Ober C, et al. (2003) Variation in the HLA-G promoter region influences miscarriage rates. *Am J Hum Genet* 72:1425–1435.
45. Ober C, et al. (1992) Decreased fecundability in Hutterite couples sharing HLA-DR. *Am J Hum Genet* 50:6–14.
46. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automatic Control* 19:716–723.
47. Schwartz G (1978) Estimating the dimension of a model. *Ann Stat* 6:461–464.
48. Ober C, Hyslop T, Hauck WW (1999) Inbreeding effects on fertility in humans: Evidence for reproductive compensation. *Am J Hum Genet* 64:225–231.

Consanguinity, human evolution, and complex diseases

A. H. Bittles[a,b,1] and M. L. Black[a]

[a]Centre for Comparative Genomics, Murdoch University, South Street, Perth WA 6150, Australia; and [b]Centre for Human Genetics, Edith Cowan University, Joondalup Drive, Perth WA 6027, Australia

Edited by Diddahally R. Govindaraju, Boston University School of Medicine, Boston, MA, and accepted by the Editorial Board August 27, 2009 (recevied for review june 25, 2009)

There is little information on inbreeding during the critical early years of human existence. However, given the small founding group sizes and restricted mate choices it seems inevitable that intrafamilial reproduction occurred and the resultant levels of inbreeding would have been substantial. Currently, couples related as second cousins or closer ($F \geq 0.0156$) and their progeny account for an estimated 10.4% of the global population. The highest rates of consanguineous marriage occur in north and sub-Saharan Africa, the Middle East, and west, central, and south Asia. In these regions even couples who regard themselves as unrelated may exhibit high levels of homozygosity, because marriage within clan, tribe, caste, or biraderi boundaries has been a long-established tradition. Mortality in first-cousin progeny is ≈3.5% higher than in nonconsanguineous offspring, although demographic, social, and economic factors can significantly influence the outcome. Improving socioeconomic conditions and better access to health care will impact the effects of consanguinity, with a shift from infant and childhood mortality to extended morbidity. At the same time, a range of primarily social factors, including urbanization, improved female education, and smaller family sizes indicate that the global prevalence of consanguineous unions will decline. This shift in marriage patterns will initially result in decreased homozygosity, accompanied by a reduction in the expression of recessive single-gene disorders. Although the roles of common and rare gene variants in the etiology of complex disease remain contentious, it would be expected that declining consanguinity would also be reflected in reduced prevalence of complex diseases, especially in population isolates.

community genetics | inbreeding | reproduction | health | social structure

It is generally accepted that the founding population size of *Homo sapiens* was small, with effective population estimates ranging downward from ≈10,000 to 1,900–2,800 and ≈1,000 to ≈700 (1–4). With such limited total numbers and population dispersal caused by a hunter–gatherer existence, a substantial level of inbreeding would have been inevitable, almost certainly involving multiple loops of kin relationships. Close-kin unions continued during the subsequent slow population growth of human groups living mainly in scattered rural settlements, with bottlenecks caused by periodic epidemics, famines, and warfare. Even in mid 19th-century Europe and North America first-cousin marriage remained both socially accepted and quite widely favored, especially among the more privileged classes (5, 6). Against this background it is puzzling that in recent generations human inbreeding has been subject to widespread negative opinion and prejudice in Western societies.

Ironically, suspicion as to the advisability of first-cousin marriage had been raised by Charles Darwin (7) in the improbabl context of a book on self-fertilization in orchids. In keeping with family tradition Darwin married his first cousin Emma Wedgwood in 1839, with 10 children born over the next 17 years. Although happily married, after the death of three of their children, including his favorite daughter Annie in 1851 probably of tuberculosis, Darwin became concerned that their union may have been a mistake from a biological perspective. However, studies conducted by his son George (8) into the prevalence and basic health outcomes of contemporary first-cousin marriage in Great Britain helped to convince Darwin to the contrary, on the grounds that "the widely different habits of life of men and women in civilized nations, especially among the upper classes, would tend to counterbalance any evil from marriages between healthy and somewhat closely related persons" (9). But by that stage the topic of cousin marriage had become a matter of often acrimonious public debate on both sides of the Atlantic, and by the end of the 19th century legislation banning first-cousin unions had been enacted by 12 state legislatures in the United States (5).

As indicated in the title of this review, a central aim is to consider the influence of consanguinity on complex genetic disorders. As a starting point, the historical background to Western and other world attitudes toward consanguinity will be briefly examined, followed by discussion of the relationship between consanguinity and community endogamy in determining population profiles of genetic disease, the current global prevalence of consanguineous unions, and the overall impact of first-cousin marriage on survival and health.

Civil and Religious Regulation of Consanguineous Marriage

The roots of negative Western attitudes toward consanguinity extend back over 1,500 years. In the Eastern Roman Empire the legality of first-cousin marriage had been confirmed by the Emperor Arcadius in 400 AD (10), possibly in acceptance of the marriage regulations defined in the Old Testament Book of Leviticus 18:7–18. But according to the Venerable Bede writing in the early 8th century (11), in 597 AD Augustine the first Archbishop of Canterbury was advised by Pope Gregory I that first-cousin marriage was banned by sacred law, a somewhat overly enthusiastic interpretation of Leviticus 18:6, "none of you shall approach to any that is near kin to him, to uncover their nakedness." Depending on the translation of Bede consulted, Gregory I further advised that first-cousin unions "do not result in children" (11), an opinion that is factually incorrect (12), or that "the offspring of such marriages cannot thrive" (10), which also is at best an overstatement.

Until 1917 the Roman Catholic Church required dispensation for unions between couples related as first, second, or third cousins (equivalent to a coefficient of inbreeding, $F \geq 0.0039$),

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Evolution in Health and Medicine" held April 2–3, 2009, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS web site at www.nasonline.org/Sackler_Evolution_Health_Medicine.

Author contributions: A.H.B. designed research; A.H.B. and M.L.B. performed research; M.L.B. analyzed data; and A.H.B. and M.L.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. D.R.G. is a guest editor invited by the Editorial Board.

[1]To whom correspondence should be addressed. E-mail: abittles@ccg.murdoch.edu.au.

This article contains supporting information online at www.pnas.org/cgi/content/full/0906079106/DCSupplemental.

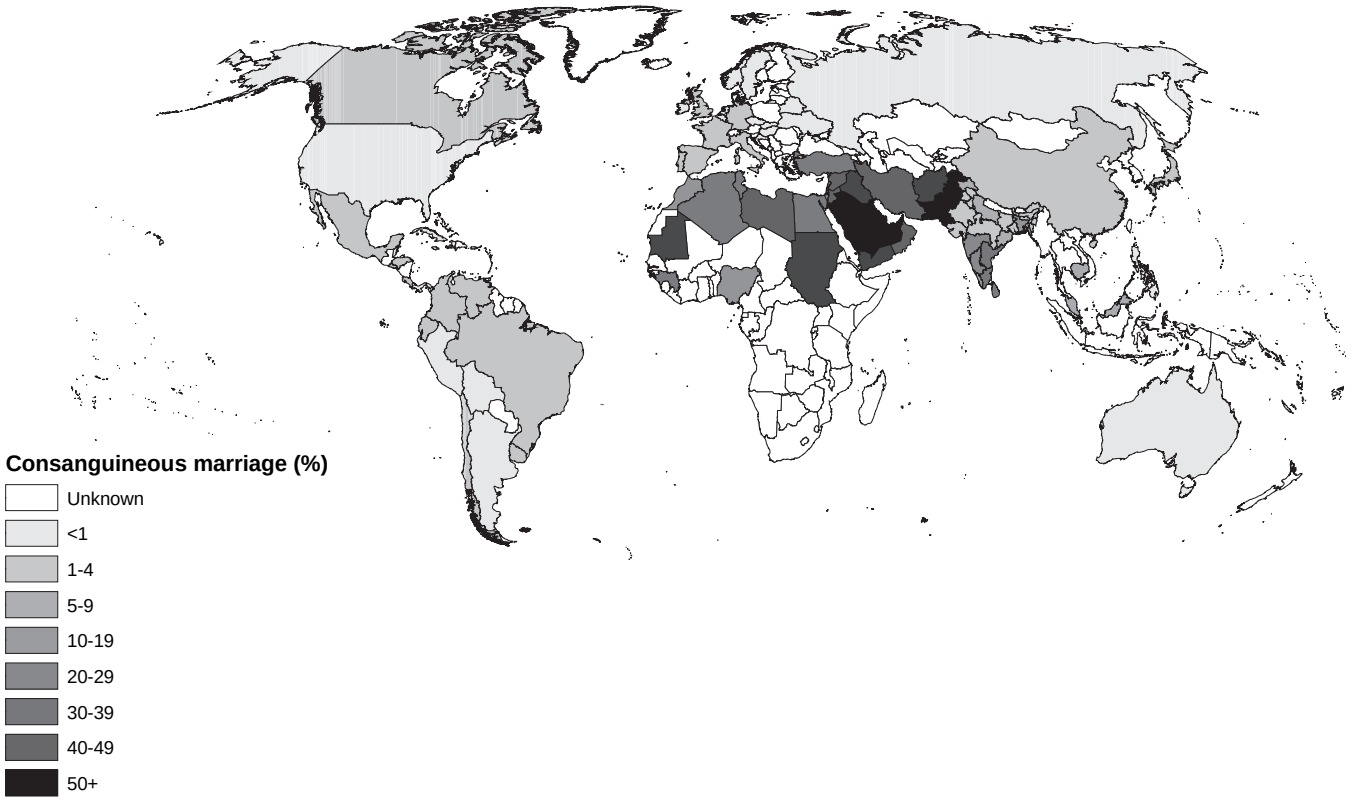


Fig. 1. Global distribution of marriages between couples related as second cousins or closer ($F \geq 0.0156$).

with a wide range of reasons accepted as grounds for consanguinity dispensation, e.g., the small size of the local population, advanced bridal age, or lack of dowry (13). As a result of misunderstanding after the switch from the Roman to the Germanic system for calculating degrees of consanguinity, during the late 11th to the early 13th centuries the requirement for dispensation expanded to include fourth-, fifth-, and sixth-cousin marriages ($F \geq 0.00006$), a level of regulation that rapidly proved impractical at local level (10). Because Luther had attacked the dispensation requirements for consanguineous unions as representing the rules of the church rather than of divine intention, and as a revenue-raising device (10), after the Reformation the Protestant denominations largely accepted the Levitical marriage proscriptions with no restriction on first-cousin unions.

The Levitical guidelines also permit uncle–niece marriage ($F = 0.125$), which along with first-cousin marriages are still practiced in many Sephardi Jewish communities. Marriage regulations in Islam permit first-cousin and double first-cousin ($F = 0.125$) marriages, but uncle–niece unions are prohibited by the Quran. Contrary to common belief there is no encouragement of consanguinity within Islam, and although the Prophet Muhammad married his daughter Fatima to his ward and first cousin Ali, several hadith (sayings of the Prophet) endorse marriage between nonrelatives (14). It therefore seems that the strong preference for first-cousin marriage in most Muslim countries, principally the parallel paternal subtype, i.e., between a man and his father's brother's daughter, reflect both pre-Islamic Arab tradition and the rules introduced in the Quran enabling female inheritance of wealth (15).

First-cousin marriage is generally permitted within Buddhism, but the marriage regulations in Hinduism are more complex. According to the north Indian tradition believed to date back to 200 BC, pedigrees are examined over an average of seven generations on the male side and five generations on the female side to preclude a consanguineous union (16). Whereas in Dravidian south India, cross first-cousin marriage (between a man and his mother's brother's daughter) and more especially uncle–niece marriages are favored across all castes. Because of their customary nature, cross-cousin marriages were recognized by the government of India in the Hindu Marriage Act of 1955 and the legality of uncle–niece marriages was confirmed in the Hindu Code Bill of 1984 (17).

The Current Global Prevalence of Consanguineous Marriage

As illustrated in Fig. 1, based on detailed information accessible at the Global Consanguinity website (www.consang.net), close-kin marriage continues to be preferential in many major populations, with the influence of religion apparent in the major regional differences in consanguinity prevalence across the globe (18). Despite anthropological reports indicating consanguineous marriage throughout sub-Saharan Africa, and in populous Asian countries including Bangladesh and Indonesia, little quantitative information on consanguinity is available from these regions. Nevertheless, current data indicate that some 10.4% of the 6.7 billion global population are related as second cousins or closer ($F \geq 0.0156$). Although the overall prevalence of consanguineous marriage seems to be declining, in some countries the present-day rates of consanguinity exceed those of the preceding generation, possibly reflecting greater overall survival to adulthood that in turn increases the numbers of marriageable biological relatives (19).

Large-scale emigration of people from countries where consanguinity is preferential to North America, Europe, and Oceania was an important demographic feature of the latter half of the 20th century. As previously indicated, first-cousin marriages ($F = 0.0625$) have the potential to cause legal problems for migrants and state law enforcement authorities in the United States because these unions are now either illegal or a criminal

offense in 31 of 50 states (5, 6, 20), despite a unanimous recommendation in 1970 that all such state laws should be rescinded (21). In Western Europe there are at least 10 million resident migrants from regions where consanguinity is preferential, and it is the possibility that the progeny of consanguineous unions are more likely to be affected by recessive genetic disorders that has aroused greater controversy, for example, with calls by some legislators for a ban on first-cousin marriages in the United Kingdom's Pakistani community (19, 22). Although a decline in first-cousin marriage has been observed in the Norwegian Pakistani community (23), no similar trend seems to have occurred in the United Kingdom's Pakistani population (24) or in the Turkish or Moroccan communities in Belgium (25), and a rapid reduction in the preference for consanguineous unions by first- and second-generation migrant families in Europe appears improbable.

The Comparative Roles of Consanguinity and Endogamy in Genetic Studies

Intracommunity marriage is the norm in regions where consanguineous marriage is favored, usually contracted within long-established male lineages, e.g., within the clan (hamula) and tribe in Arab societies, within caste in India, and intrabiraderi in Pakistan. Because gene flow between communities is highly restricted in most traditional societies, adjacent villages or even coresident subcommunities may exhibit very different inherited disease profiles, reflecting local founder mutations and genetic drift (18). These characteristics have been demonstrated in tribe-specific single gene disorders in Saudi Arabia (26–28), the differential origins and expansion patterns of β-globin mutations in an Israeli Arab village (29), and village- and lineage-specific predisposing genes for visceral leishmaniasis in Sudan (30). Under these circumstances and whether or not the parents are known to be consanguineous, a recessive founder or de novo mutation of chronic effect can rapidly increase in frequency within a particular community or subcommunity, resulting in the birth of an affected child. In communities with a high level of consanguineous marriage, the diagnosis of a recessive disorder in one or more members of the same family is generally indicative of a recent mutation, whereas the presence of a rare disorder in several families suggests an older mutational event or previous admixture through marriage with a person from another community (31).

Population substructure, whether caused by ethnic, geographical, religious, or social divisions, often results in variant marker allele frequencies in different subpopulations. The occurrence of type 1 errors, i.e., false positive results, is of major importance in case-control studies, association studies, and clinical trials (32, 33). Conflicting opinions have been expressed as to the impact of population stratification on genomewide studies with, for example, the claim that in the United Kingdom if persons of non-European ancestry are excluded "the extent of population stratification in the British population is generally modest" (34). Conversely, in the more homogenous Icelandic population it was believed that population substructure had to be considered in the sampling strategy, with the implication that it would be of much greater importance in larger populations with more diverse genetic origins (35). Because genomic studies consistently report that a large majority (93–95%) of genetic variation is within-population (36), the latter opinion is unsurprising and highlights the need for vigilance in case-control studies to preclude spurious associations.

As discussed in the following sections, population stratification may also be of critical importance in the investigation of consanguinity-associated morbidity and mortality, with straightforward comparisons drawn between the progeny of first cousins versus unrelated parents of dubious validity unless both sets of parents are known to be members of the same clan, tribe, caste, or biraderi (19). For this reason, in many populations the clan or its hereditary social/occupational equivalent may be the most logical unit for genetic screening and genetic counseling programs, as exemplified by the distribution pattern of β-thalassemia in Oman where >50% of cases were diagnosed in just one of the 185 major tribes and subtribes (37, 38).

Consanguinity and Health

Within genetics, contemporary attention on consanguineous marriage continues to be largely focused on the expression and identification of rare autosomal recessive alleles, a recent example being a comparative study in Norway of progressive encephalopathy in Pakistani migrants and the indigenous population (39). But as indicated in Fig. 2, from an overall health perspective consanguinity is a much wider and more complex topic involving major social, economic, and demographic influences, differential reproductive behavior, and early- and late-onset morbidity and mortality. A thorough appreciation of the salient nongenetic variables is therefore essential in addressing the concerns of individuals, families, and communities with regard to reproductive choices, and in designing genetic education and genetic counseling programs for consanguineous couples.

The highest overall prevalence of consanguineous unions is in poor rural communities, which are typified by low levels of maternal education, early age at marriage and first birth, short birth intervals, and longer reproductive spans (15, 40–42). Each of these factors is independently associated with larger family sizes and higher rates of infant and early childhood mortality, with reproductive compensation for early losses a further complicating issue in assessing the overall health outcomes of consanguinity (12). Comprehensive genetic education and premarital genetic counseling programs can help to lessen the burden of genetic diseases in such communities, as reported in Israeli Arab and Bedouin villages (43–45). While in Middle Eastern countries such as Bahrain educational programs aimed at high school children, and through them their parents and relatives, have had a marked beneficial effect in reducing the incidence of sickle cell disease (46). There are, however, current limitations to the success of these initiatives in many low-income countries, in particular the lack of clinicians, genetic counselors, nurses, and scientific support staff with appropriate specialist training (47). Patients referred for genetic counseling may also expect directive advice as to whether or not to proceed with a pregnancy, with failure to provide an opinion interpreted as a lack of knowledge on the part of the clinician (48), and even when specific rulings have been provided by religious authorities permitting prenatal diagnosis of genetic diseases and selective termination of a pregnancy, this option may remain unacceptable to individual couples (15).

Consanguinity, Mortality, and Morbidity

To investigate the impact of consanguinity on deaths from ≈6 months gestation to an average of 10 years of age, a metaanalysis was conducted directly comparing prereproductive mortality in first-cousin versus nonconsanguineous progeny within specific populations. The study sample comprised 69 populations resident in 15 countries located across four continents, with a total sample size of 2.14 million (Table S1). An unweighted linear regression comparing mean mortality in first-cousin versus nonconsanguineous progeny in each population was plotted according to the standard equation $y = a + bx$. The results are presented in Fig. 3 as a scatter diagram and show a mean excess mortality at first-cousin level of 3.5% ($r^2 = 0.70$; $P < 0.00001$) that is consistent across the range of control mortalities, i.e., the level of excess consanguinity-associated mortality is independent of the basal (nonconsanguineous) death rate in each study population. The estimate of 3.5% excess deaths among first-cousin progeny compares with an earlier global estimate of 4.4%

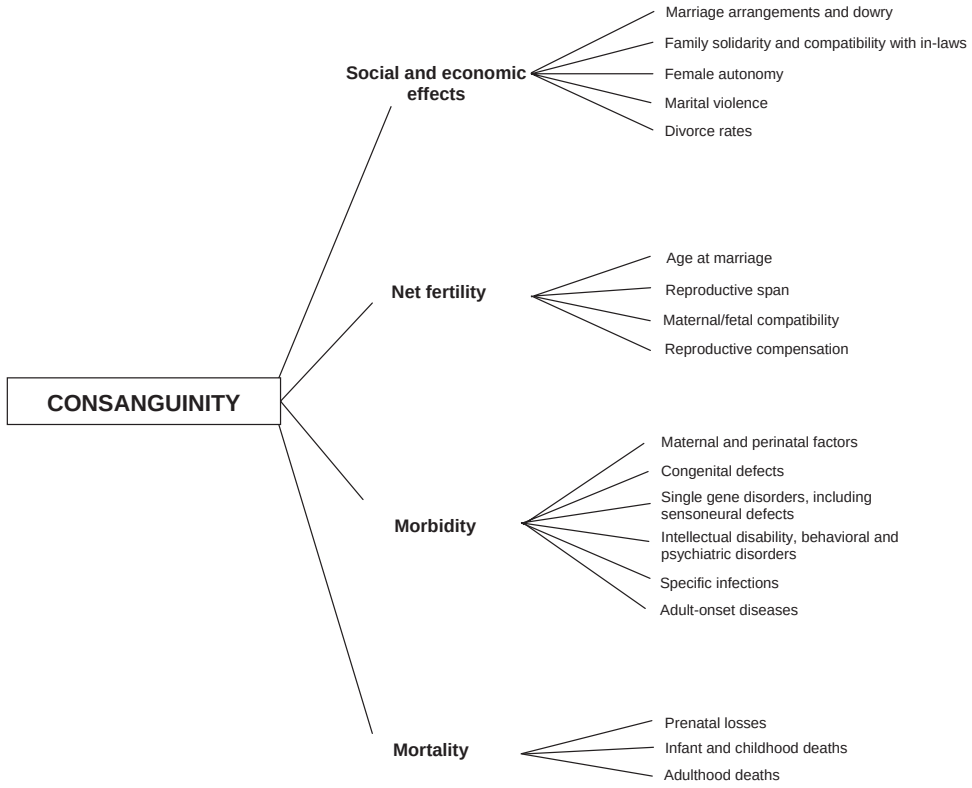


Fig. 2. Influences and outcomes of consanguineous marriage.

excess mortality (49) calculated from 38 studies each of which was included in the present analysis, and it matches the 3.5% excess mortality derived for Italian data of the early to mid 20th century (13).

Initial estimates of the adverse effects of consanguineous marriage, expressed as lethal gene equivalents, had produced significantly higher values for consanguinity-associated mortality, mainly because of lack of control for the negative correlation between consanguinity and socioeconomic status (50). Although control for the effects of nongenetic variables was improved in the present study, the mean value of 3.5% excess mortality at the first-cousin level is an upper-level estimate that may be subject to further downward revision as data from better-designed studies become available.

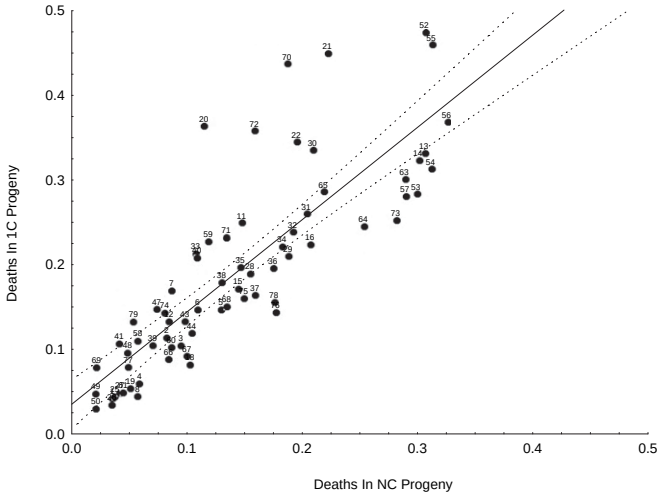


Fig. 3. Comparative mortality in first cousin (1C; $F = 0.0625$: y axis) versus nonconsanguineous progeny (NC; $F = 0$: x axis) in 69 study populations.

The influence of first-cousin marriage on the prevalence of autosomal recessive single-gene disorders was examined as part of an investigation into consanguinity-associated morbidity in a Pakistani community in the United Kingdom (51). From the results of this 5-year prospective study it was calculated that there would be a ≈7/1,000 increase in autosomal recessive disorders per 0.01 increase in the mean coefficient of inbreeding (52). Thus, in a national population such as Pakistan where ≈50% of marriages were between first cousins ($F = 0.0625$) (53) some 22/1,000 extra single-gene disorders would be expected.

Unfortunately, the original study omitted control for population subdivision, which has been shown to be a notable feature of indigenous and migrant Pakistani populations (54–56), and as previously noted is typical of many more traditional populations. Wahlund effect predicts that subdivided populations characteristically exhibit higher than predicted levels of homozygosity. Given the known levels of population substructure associated with biraderi membership in Pakistan and the Pakistani community in the United Kingdom, nonconsanguineous couples are at higher risk of sharing the same recessive disease mutation than counterparts in populations where limited or no substructure exists. The consequent random consanguinity effect on the distribution and expression patterns of recessive disease genes means that in populations with significant subdivision the beneficial health outcomes that have been claimed through simply avoiding consanguineous marriage are almost certainly exaggerated and require reassessment (19, 57).

Consanguinity and Complex Diseases

There has been extended debate on the nature of the genetic contribution to complex diseases, i.e., whether the common disease/common variant or the common disease/rare variant hypothesis is more applicable (58), with the role of copy number

variants also proposed (59, 60). Consanguinity would be expected to exert a greater influence on the etiology of complex diseases if rare autosomal recessive alleles were causally implicated, whereas if disease alleles that are common in the gene pool are involved then intrafamilial marriage would have a proportionately lesser effect. However, because both gene–gene interactions and numerous nongenetic factors in prenatal and postnatal life also contribute to the disease phenotype, a single all-embracing solution to the genetics of complex diseases is highly improbable.

Major genomewide analyses of diseases with onset primarily in childhood and adulthood have identified associations with specific chromosomal regions, e.g., for type 1 and type 2 diabetes (61, 62), although these studies have emphasized the large numbers of genes involved and the small increased risk that appears to be associated with most individual variants. Concern also has been expressed that concentration on the identification of gene variants via patients with the disease under study rather than full genome sequencing of randomly ascertained samples could lead to significantly inflated rates of false positives (63).

Investigations into the effects of consanguinity on congenital defects have produced quite varied results, in large part because of a lack of standardized assessment protocols and the different environmental and socioeconomic circumstances of the study populations. Using nonconsanguineous progeny as controls, estimates of the excess level of congenital defects in first-cousin offspring have ranged from 0.7% to 7.5% (64–68), but the Latin American Collaborative Study of Congenital Malformation based on 34,1902 newborns found a significant association with consanguinity only for hydrocephalus, postaxial polydactyly, and bilateral oral and facial clefts (69).

A different picture emerges from the large literature on congenital heart defects, which are conservatively estimated to have an incidence of 50/1,000 live births (70). Although a consistent positive association between consanguinity and disorders such as ventricular septal defect and atrial septal defect has been demonstrated, indicating the involvement of common variants, both positive and negative associations with patent ductus arteriosus, atrioventricular septal defect, pulmonary atresia, and tetralogy of Fallot have been reported in different populations (71–74), suggestive of community-specific founder mutations. It is, however, also possible that nonstandardized diagnostic protocols may have contributed to the variant findings reported by different study centers.

As yet relationships between consanguinity and complex diseases of adulthood have been significantly underinvestigated, and the few studies published have relied mainly on rudimentary sampling strategies, with simple consanguineous versus nonconsanguineous comparisons in disease prevalence and inadequate attention paid to possible genetic or demographic subdivisions. Accordingly, the results obtained often are contradictory, e.g., with both positive and negative associations reported between consanguinity and breast cancer (75–77), and consanguinity and heart disease (75, 78, 79). Long-term studies conducted on the Dalmatian islands in the Adriatic Sea have indicated a positive association between inbreeding and a very wide range of common adulthood disorders, including hypertension, coronary heart disease, stroke, cancer, uni/bipolar depression, asthma, gout, peptic ulcer, and osteoporosis (80–82). The data thus suggest virtually ubiquitous causal involvement of rare autosomal recessive genes in adult-onset disease in this population, with the more general corollary that increasing genomewide heterozygosity after a decline in consanguineous marriage should lead to a widespread reduction in the burden of common genetic diseases (83).

The Dalmatian studies have the very considerable advantage of demographically well-characterized populations with known ethnic origins, although the actual definitions used in assessing the comparative levels of inbreeding are genetically quite imprecise and principally reflect village endogamy rather than consanguinity *per se*. As previously discussed, until the early 20th century church dispensation would have been required for marriages between spouses related as third cousins or closer ($F \geq 0.0039$) in these devoutly Roman Catholic communities (13). In the absence of church records indicating dispensation for marriages contracted within the prevailing consanguinity regulations, the consanguineous relationships examined may principally have been random rather than preferential in nature and reflected restricted marriage partner choices. The analysis of genealogical data covering four to five generations showed substantial levels of consanguinity in some communities, with mean coefficients of inbreeding ranging from $\alpha = 0.002$ to 0.049 calculated at village level, indicating major variations in local marriage patterns driven by both the history and the geographical location of each settlement (80).

Pedigree-based estimates of consanguinity and the resultant levels of homozygosity have several limitations; in particular, they do not provide information on close-kin marriages that have occurred in distant generations and thus underestimate cumulative inbreeding effects, and with rare exceptions incorrectly ascribed paternity is not recorded. To complement the pedigree-based approaches previously adopted and avoid these difficulties, high-density genome scans were used to estimate individual autozygosity (Froh) from uninterrupted runs of homozygosity (ROH). An appropriate length threshold was empirically derived for ROH and the method was applied to data derived from residents of the Dalmatian islands, the Orkney islands off the north coast of Scotland, mainland Scotland, and the state of Utah (84). Initial comparisons of Froh values ranging from 0.5, i.e., with a minimum length threshold of 0.5 Mb, to 1 (length threshold 1 Mb) and 5 (length threshold 5Mb) with pedigree data from the Orkneys indicated good correlation with pedigree-based mean coefficients of inbreeding and so confirmed the applicability of the method for the direct assessment of autozygosity. The method has been further applied to investigate changes in autozygosity through time in two American study populations. The steady decreases observed in the size and frequency of ROH > 1 Mb in length in these populations were ascribed to expanded marriage pools and larger effective population sizes and interpreted as indicating future ongoing reductions in the frequency of rare recessive disorders (85).

When applied to behavioral disorders genomewide analysis has indicated the potential contribution of thousands of alleles of very small effect in schizophrenia and bipolar disorder, with significant genetic overlap between the two disease states (86, 87). At the same time, homozygosity mapping in autism (88) and a case-control study of bipolar disorder type 1 in consanguineous progeny (89) both implicated the causal expression of rare recessive genes. ROH similarly have been shown to be significantly more common in patients with schizophrenia spectrum disorders, suggesting the involvement of recessive alleles in the etiology of the disorder (90). Reverting to earlier comments on the relationship between endogamy and consanguinity, an association between consanguinity and Alzheimer disease was demonstrated in a genealogical study of the Saguenay region in Québec (91), and multiple loci for Alzheimer disease were identified in a highly endogamous and consanguineous Israeli Arab kindred (92), in both cases indicative of founder mutations. Thus, from a more general perspective these results strengthen the argument that all association studies on complex diseases would benefit from a sound prior knowledge of community demographic and genetic structure.

Discussion

Although consanguinity is a highly complex and multifaceted topic (Fig. 2), the claimed social and cultural advantages, such as

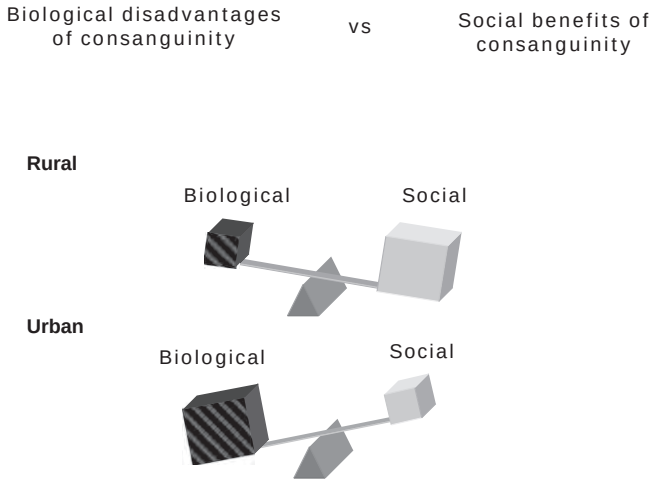


Fig. 4. Contrasting biological and social outcomes of consanguineous marriage in traditional rural and modern urban settings.

ease of marriage arrangements, enhanced female autonomy, more stable marital relationships, greater compatibility with in-laws, lower domestic violence, lower divorce rates, and the economic benefits of reduced dowry and the maintenance of any landholdings (15, 41, 42, 47, 93–95) have received much less attention than studies into adverse genetic outcomes. It therefore is not surprising that the prevailing Western public and medical opinion with regard to consanguinity is largely negative. There is the additional problem that in many societies that favor consanguineous unions marriages are usually arranged by and/or meet with prior parental approval, a practice frequently misrepresented and criticized as "forced marriage" (15).

For families living in impoverished rural areas with limited or no formal education or access to medical services, young age at marriage and first pregnancy, short birth intervals, and high infant and childhood mortality rates primarily caused by infectious and nutritional disorders, the social and economic advantages offered by consanguineous marriage and the strengthening of family relationships often outweigh the biological disadvantages of close-kin marriage for a majority of families (96, 97). The current scenario in urban populations is quite different, especially in developed countries with better living and public health conditions, low levels of infectious disease, and ready access to modern health facilities. Newborns with a genetic disorder that in previous generations may have died in infancy of no known cause are now referred to specialist centers for diagnosis, and they and their families can anticipate a lifespan that will extend at least into adolescence and more probably into mid to late adulthood, usually requiring ongoing medical care.

Unless a de novo mutation has been identified the diagnosis will effectively involve other family members as potential or obligate carriers and so could become a negative factor in all future family marriage arrangements (19, 98). For this reason, in disorders with a very adverse clinical outcome and involving multiple affected family members, such as progressive retinopathy and amelogenesis (99) and severe intellectual disability (100), marriage to a nonrelative may not be a realistic option, resulting either in celibacy or continued intrafamilial marriage. Within the wider community, greater understanding and acceptance of genetic explanations for familial patterns of disease and the unfavorable medical outcomes experienced by some consanguineous families can significantly influence the perceived balance of advantage and disadvantage associated with intrafamilial marriage (Fig. 4). Therefore, in conjunction with increasing difficulty in finding a marriageable cousin of acceptable age because of rapidly declining family sizes, future global reductions in the prevalence of consanguinity appear to be inevitable (19).

What effect will this predicted reduction in consanguinity have in terms of human evolution and on the prevalence of genetic disease? Recent studies have identified the ongoing role of positive natural selection during an extended period when effective population sizes were small and consanguinity would have been high (101–103), and the very rapid increases in global population numbers over the course of the last 150 years would suggest even greater acceleration in the pace of current and future human adaptive evolution (104). Although the mixing of previously separated breeding groups should lead to a marked initial reduction in the global prevalence of rare autosomal recessive disorders (85), the subsequent dispersal of phenotypically normal heterozygotes through newly agglomerated breeding pools will in time result in the "random" mating of nonconsanguineous carriers of recessive mutations. But the rate at which these changes in mating patterns occur will necessarily be more rapid in increasingly panmictic urbanized populations than in endogamous ethnic, religious, geographical, or social isolates.

Whether similar predictions are possible for complex diseases will very much depend on the proportional contribution of recessive genes, and more especially rare recessive genes, to individual diseases in different populations. For the moment the greatest promise in identifying genes of major effect for complex diseases continues to reside in endogamous communities with extensive genealogical records (105). Convincing support for this approach is provided by the high frequencies of autosomal recessive disease genes diagnosed in numerically small, highly endogamous Arab Israeli communities (106). Yet, surprisingly, in these communities and other isolates where consanguinity is much less common, multiple mutations in specific disease genes have been identified where a single founder mutation would more usually have been expected (29, 107). Because limited genetic diversity and restricted allelic heterogeneity are generally expected in isolated founder populations, it also is salutary that a genomewide association analysis of obesity and other metabolic disorders in a Pacific island community, in which reduced haplotype diversity and extended linkage disequilibrium had already been demonstrated, failed to detect major contributory alleles and instead indicated the presence of common variants of small effect (108, 109).

Having largely been ignored for many years, the specific roles of population bottlenecks and consanguinity in influencing variation between and within populations are now receiving due attention, with special focus on homozygosity in identifying recent common ancestry via ROH analysis (110). The potential complexity of the interrelationships between consanguinity and human health and disease was highlighted by the reported association between consanguinity and predisposition to major infectious diseases (111). If these findings are substantiated, by ameliorating the risk of exposure to infectious agents a global decline in consanguinity could also providentially reduce the risk of inflammatory disease and hence the development of coronary disease in middle and old age (112).

Time will tell whether these as yet tenuous epidemiological connections can be sustained. In the interim, it is important to emphasize that in assessing the impact of consanguinity on any aspect of health a clear causal relationship needs to be established, rather than reliance on speculation driven solely by the presence of a close kin union in the family pedigree. At the same time, rigorous control for population stratification should be a prerequisite in the many populations where community subdivisions exist if confused and confusing conclusions are to be avoided.

ACKNOWLEDGMENTS. We thank the referees for constructive comments. A.H.B. is supported by National Science Foundation Grant 0527751.

1. Harpending HC, et al. (1998) Genetic traces of ancient demography. *Proc Natl Acad Sci USA* 95:1961–1967.
2. Tenesa A, et al. (2007) Recent human effective population size estimated from linkage disequilibrium. *Genome Res* 17:520–526.
3. Liu H, Prugnolle F, Manica A, Balloux F (2006) A geographically explicit genetic map of worldwide human-settlement history. *Am J Hum Genet* 79:230–237.
4. Zhivotovsky LA, Rosenberg NA, Feldman MW (2003) Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers. *Am J Hum Genet* 72:1171–1186.
5. Ottenheimer M (1996) *Forbidden Relatives: The American Myth of Cousin Marriage* (Univ Illinois Press, Urbana), pp 19–41.
6. Bittles AH (2003) The bases of Western attitudes to consanguineous marriage. *Dev Med Child Neurol* 45:135–138.
7. Darwin C (1862) *On the Various Contrivances by Which British and Foreign Orchids Are Fertilized by Insects, and on the Good Effects of Intercrossing* (John Murray, London), pp 359–360.
8. Darwin GH (1875) Marriages between first cousins in England and Wales and their effects. *J Stat Soc* 38:153–184.
9. Darwin C (1876) *The Effects of Cross and Self-Fertilization in the Vegetable Kingdom* (John Murray, London), pp 460–461.
10. Goody J (1985) *The Development of the Family and Marriage in Europe* (Cambridge Univ Press, Cambridge, UK), pp 48–60 and 134–146.
11. Bede (1991) *The Ecclesiastical History of the English People* (Penguin Books, London), Revised Ed, pp 79–81.
12. Bittles AH, Grant JC, Sullivan SG, Hussain R (2002) Does inbreeding lead to increased human fertility? *Ann Hum Biol* 29:111–131.
13. Cavalli-Sforza LL, Moroni A, Zei G (2004) *Consanguinity, Inbreeding, and Genetic Drift in Italy* (Princeton Univ Press, Princeton).
14. Hussain R (1999) Community perceptions of reasons for preference for consanguineous marriages in Pakistan. *J Biosoc Sci* 31:449–461.
15. Bittles AH, Hamamy HA (2009) Consanguinity and endogamy in Arab countries. *Genetic Disorders among Arab Populations*, ed Teebi A (Springer, Heidelberg), 2nd Ed, in press.
16. Kapadia KM (1958) *Marriage and the Family in India* (Oxford Univ Press, Calcutta), 2nd Ed, pp 117–137.
17. Bittles AH (2002) Endogamy, consanguinity, and community genetics. *J Genet* 81:91–98.
18. Bittles AH (2009) Consanguinity, genetic drift, and genetic diseases in populations with reduced numbers of founders. *Human Genetics: Principles and Approaches*, eds Vogel F, Motulsky AG, Antonarakis SE, Speicher M (Springer, Heidelberg), 4th Ed, in press.
19. Bittles AH (2008) A community genetics perspective on consanguineous marriage. *Commun Genet* 11:324–330.
20. Paul DB, Spencer HG (2008) "It's OK, we're not cousin by blood": The cousin marriage controversy in historical perspective. *PLoS Biol* 6:e320.
21. National Conference of Commissioners (1970) *Handbook on Uniform State Laws and Proceedings of the Annual Conference Meeting in its Seventy-Ninth Year* (Port City Press, Baltimore).
22. Dyer O (2005) MP is criticized for saying that marriage of first cousins is a health risk. *Br Med J* 331:1292.
23. Grjibovski AM, Magnus P, Stoltenberg C (2009) Decrease in consanguinity among parents of children born in Norway to women of Pakistani origin: A registry-based study. *Scand J Pub Health* 37:232–238.
24. Shaw A (2000) Kinship, cultural preference, and immigration: Consanguineous marriage among British Pakistanis. *J R Anthropol Inst* 7:315–334.
25. Reniers G (1998) *Postmigration Survival of Traditional Marriage Patterns: Consanguineous Marriage Among Turkish and Moroccan Immigrants in Belgium* (Department of Population Studies, University of Gent, Gent, Belgium), Interuniversity Papers in Demography, PPD-1 Working Paper 1998-1.
26. Ozand PT, et al. (1990) Prevalence of different types of lysosomal storage diseases in Saudi Arabia. *J Inherit Metab Dis* 13:849–861.
27. Ozand PT, Devol EB, Gascon GG (1992) Neuro-metabolic diseases at a national referral center: Five years experience at the King Faisal Specialist Hospital and Research Centre. *J Child Neurol* 7 (Suppl): S4–S11.
28. Rashed M, Ozand PT, Al Aqeel A, Gascon GG (1994) Experience of King Faisal Specialist Hospital and Research Center with Saudi organic acid disorders. *Brain Dev* 16 (Suppl): 1–6.
29. Zlotogora J, et al. (2005) The origin and expansion of four different β-globin mutations in a single Arab village. *Am J Hum Biol* 17:659–661.
30. Miller EN, et al. (2007) Y chromosome lineage- and village-specific genes on chromosomes 1p22 and 6q27 control visceral leishmaniasis in Sudan. *PLoS Genet* 3:e71.
31. Zlotogora J, Hujerat Y, Barges S, Shalev SA, Chakravarti A (2006) The fate of 12 recessive mutations in a single village. *Ann Hum Genet* 71:202–208.
32. Heiman GA, Hodge SE, Gorroochurn P, Zhang J, Greenberg DA (2004) Effect of population stratification on case-control association studies. *Hum Hered* 58:30–39.
33. Wang K (2009) Testing for genetic association in the presence of population stratification in genomewide association studies. *Genet Epidemiol*, in press.
34. The Wellcome Trust Case Control Consortium (2007) Genomewide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–678.
35. Helgason A, Yngvadóttir B, Hrafnkelsson B, Gulcher J, Stefánsson K (2005) An Icelandic example of the impact of population structure on association studies. *Nat Genet* 37:90–95.
36. Rosenberg NA, et al. (2002) Genetic structure of human population. *Science* 298:2381–2385.
37. Rajab A, Patton MA (1997) Major factors determining the frequencies of hemoglobinopathies in Oman. *Am J Med Genet* 71:240–242.
38. Rajab A, Patton MA (1999) Analysis of the population structure in Oman. *Commun Genet* 2:23–25.
39. Strømme P, et al. (2009) Parental consanguinity is associated with a 7-fold increased risk of progressive encephalopathy: A cohort study from Oslo, Norway. *Eur J Paed Neurol*, in press.
40. Bittles AH, Mason WH, Greene J, Appaji Rao N (1991) Reproductive behavior and health in consanguineous marriages. *Science* 252:789–794.
41. Bittles AH (1994) The role and significance of consanguinity as a demographic variable. *Pop Dev Rev* 20:561–584.
42. Khlat M (1997) Endogamy in Arab countries. *Genetic Disorders Among Arab Populations*, eds Teebi A, Farag TI (Oxford Univ Press, New York), pp 63–80.
43. Shiloh S, Reznik H, Bat-Miriam-Katznelson M, Goldman B (1995) Premarital genetic counseling to consanguineous couples: Attitudes, beliefs, and decisions among counseled, noncounseled, and unrelated couples in Israel. *Soc Sci Med* 41:1301–1310.
44. Raz AE, Atar M (2004) Cousin marriage and premarital carrier matching in a Bedouin community in Israel: Attitudes, service development, and educational intervention. *J Fam Planning Reprod Health Care* 30:49–51.
45. Zlotogora J, Carmi R, Lev B, Shalev SA (2009) A targeted population carrier screening program for severe and frequent genetic diseases in Israel. *Eur J Hum Genet* 17:591–597.
46. Al Arrayed S (2005) Campaign to control genetic blood diseases in Bahrain. *Commun Genet* 8:52–55.
47. Hamamy H, Bittles AH (2009) Genetic clinics in Arab communities: Meeting individual, family, and community needs. *Pub Health Genomics* 12:30–40.
48. Eldahdah L, et al. (2007) Outcome of chromosomally abnormal pregnancies in Lebanon: Obstetricians' roles during and after prenatal diagnosis. *Prenat Diagn* 27:525–534.
49. Bittles AH, Neel JV (1994) The costs of human inbreeding and their implications for variations at the DNA level. *Nat Genet* 8:117–121.
50. Bittles AH, Makov E (1988) Inbreeding in human populations: An assessment of the costs. *Human Mating Patterns*, eds Mascie-Tyalor CGN, Boyce AJ (Cambridge Univ Press, Cambridge, UK), pp 153–167.
51. Bundey S, Alam H (1993) A 5-year prospective study of the health of children in different ethnic groups, with particular reference to the effect on inbreeding. *Eur J Hum Genet* 1:206–219.
52. Christianson A, Howson CP, Modell B (2006) *Global Report on Birth Defects* (March of Dimes, White Plains, NY), pp 83–84.
53. Hussain R, Bittles AH (1998) The prevalence and demographic characteristics of consanguineous marriages in Pakistan. *J Biosoc Sci* 30:261–275.
54. Wang W, et al. (2000) A genome-based study of consanguinity in three coresident endogamous Pakistan communities. *Ann Hum Genet* 64:41–49.
55. Qamar R, et al. (2002) Y-chromosomal DNA variation in Pakistan. *Am J Hum Genet* 70:1107–1124.
56. Overall ADJ, Ahmad M, Thomas MG, Nichols RA (2003) An analysis of consanguinity and social structure within the U.K. Asian population using microsatellite data. *Ann Hum Genet* 67:525–537.
57. Overall ADJ (2009) The influence of the Wahlund effect on the consanguinity hypothesis: Consequences for recessive disease incidence in a socially structured Pakistani population. *Hum Hered* 67:140–144.
58. Schork NJ, Murray SS, Frazer KA, Topol EJ (2009) Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev* 19:212–219.
59. Estevill X, Armengol L (2007) Copy number variants and common disorders: Filling the gaps and exploring complexity in genomewide association studies. *PLoS Genet* 3:e190.
60. Need AC, et al. (2009) A genomewide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet* 5:e1000373.
61. Todd JA, et al. (2007) Robust associations of four new chromosome regions from genomewide analyses of type 1 diabetes. *Nat Genet* 39:813–815.
62. Lyssenko V, et al. (2008) Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med* 359:2220–2232.
63. Li B, Leal SM (2009) Discovery of rare variants via sequencing: Implications for the design of complex trait association studies. *PLoS Genet* 5:e1000481.
64. Schull WJ (1958) Empirical risks in consanguineous marriages: Sex ratio, malformation, and viability. *Am J Hum Genet* 10:294–343.
65. Jaber L, Merlob P, Bu X, Rotter JI, Shohat M (1992) Marked parental consanguinity as a cause for increased major malformations in an Israeli Arab community. *Am J Med Genet* 44:1–6.
66. Stoltenberg C, Magnus P, Lie TR, Daltveit AK, Irgens LM (1997) Birth defects and parental consanguinity in Norway. *Am J Epidemiol* 145:439–448.
67. Zlotogora J (2002) What is the birth defect risk associated with consanguineous marriage? *Am J Med Genet* 109:70–71.
68. Bromiker R, Glam-Baruch M, Gofin R, Hammerman C, Amitai Y (2004) Association of parental consanguinity with congenital malformations among Arab newborns in Jerusalem. *Clin Genet* 66:63–66.
69. Rittler M, Liascovich R, López-Camelo J, Castilla EE (2001) Parental consanguinity in specific types of congenital anomalies. *Am J Med Genet* 102:36–43.
70. Pierpont ME, et al. (2007) Genetic basis for congenital heart defects: Current knowledge. *Circulation* 115:3015–3038.
71. Gnanalingham MG, Gnanalingham KK, Singh A (1999) Congenital heart disease and parental consanguinity in South India. *Acta Paediatr* 88:473–474.
72. Becker SM, Al Halees Z, Molina C, Paterson RM (2001) Consanguinity and congenital heart disease in Saudi Arabia. *Am J Med Genet* 99:8–13.

73. Nabulsi MM, et al. (2003) Parental consanguinity and congenital heart malformations in a developing country. *Am J Med Genet* 116A:342–347.
74. Khalid Y, et al. (2006) Consanguineous marriage and congenital defects: A case-control study in the neonatal period. *Am J Med Genet* 140:1524–1530.
75. Shami SA, Qaisar R, Bittles AH (1991) Consanguinity and adult morbidity in Pakistan. *Lancet* 338:954–955.
76. Liede A, et al. (2002) Contribution of BRAC1 and BRAC2 mutations of breast and ovarian cancer in Pakistan. *Am J Hum Genet* 71:595–606.
77. Denic S, Bener A (2001) Consanguinity decreases risk of breast cancer: Cervical cancer unaffected. *Br J Cancer* 85:1675–1679.
78. Ismail J, et al. (2004) Risk factors for nonfatal myocardial infarction in young South Asian adults. *Heart* 90:259–263.
79. Jaber L, Shohat T, Rotter JI, Shohat M (1997) Consanguinity and common adult diseases in Israeli Arab communities. *Am J Med Genet* 70:346–348.
80. Rudan I, et al. (2003) Inbreeding and the genetic complexity of human hypertension. *Genetics* 163:1011–1021.
81. Rudan I, et al. (2003) Inbreeding and risk of late-onset complex disease. *J Med Genet* 40:925–932.
82. Rudan I, et al. (2004) Inbreeding and susceptibility to osteoporosis in Croatian island isolates. *Coll Antropol* 28:585–602.
83. Campbell H, et al. (2007) Effects of genomewide heterozygosity on a range of biomedically relevant human quantitative traits. *Hum Mol Genet* 16:233–241.
84. McQuillan R, et al. (2008) Runs of homozygosity in European populations. *Am J Hum Genet* 83:359–372.
85. Nalls MA, et al. (2009) Measures of autozygosity in decline: Globalization, urbanization, and its implications for medical genetics. *PLoS Genet* 5:e1000415.
86. Lichenstein P, et al. (2009) Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: A population-based study. *Lancet* 373:234–239.
87. The International Schizophrenia Consortium (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disease. *Nature* 460:748–752.
88. Morrow EM, et al. (2008) Identifying autism loci and genes by tracing recent shared ancestry. *Science* 321:218–223.
89. Mansour H, et al. (2009) Consanguinity associated with increased risk for bipolar I disorder in Egypt. *Am J Med Genet B Neuropsychiatr Genet* 150:879–885.
90. Lencz T, et al. (2007) Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc Natl Acad Sci USA* 104:19942–19947.
91. Vézina H, et al. (1999) A genealogical study of Alzheimer disease in the Saguenay region of Québec. *Genet Epidemiol* 16:412–425.
92. Farrer LA, et al. (2003) Identification of multiple loci for Alzheimer disease in a consanguineous Israeli–Arab community. *Hum Mol Genet* 12:415–422.
93. Hussien FH (1971) Endogamy in Egyptian Nubia. *J Biosoc Sci* 3:251–257.
94. Assaf S, Khawaja M (2009) Consanguinity trends and correlates in the Palestinian Territories. *J Biosoc Sci* 41:107–124.
95. Clark CJ, Hill A, Jabber K, Silverman JG (2009) Violence during pregnancy in Jordan: Its prevalence and associated risk and protective factors. *Violence Against Women* 15:720–735.
96. Bittles AH (2001) Consanguinity and its relevance to clinical genetics. *Clin Genet* 60:89–98.
97. Bittles AH (2005) Endogamy, consanguinity, and community disease profiles. *Commun Genet* 8:17–20.
98. Shaw A, Hurst JA (2009) "I don't see any point in telling them": Attitudes to sharing genetic information in the family and carrier testing of relatives among British Pakistani adults referred to a genetics clinic. *Ethn Health* 14:205–224.
99. Jalili IK, Smith NJD (1988) A progressive cone-rod dystrophy and amelogenesis imperfecta: A new syndrome. *J Med Genet* 25:738–740.
100. Basel-Vanagaite L, et al. (2007) Genetic screening for autosomal recessive nonsyndromic mental retardation in an isolated population in Israel. *Eur J Hum Genet* 15:250–253.
101. Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. *PLoS Biol* 4:e72.
102. Sabeti PC, et al. (2007) Genomewide detection and characterization of positive selection in human populations. *Nature* 449:913–918.
103. Williamson SH, et al. (2007) Localizing recent adaptive evolution in the human genome. *PLoS Genet* 3:e90.
104. Hawks J, Wang ET, Cochran GM, Harpending HC, Moyzis RK (2007) Recent acceleration of human adaptive evolution. *Proc Natl Acad Sci USA* 104:20753–20758.
105. Varilo T, Peltonen L (2004) Isolates and their potential use in complex gene mapping efforts. *Curr Opin Genet Dev* 14:316–323.
106. Zlotogora J, Shalev S, Habiballah H, Barjes S (2000) Genetic disorders among Palestinian Arabs: 3. Autosomal recessive disorders in a single village. *Am J Med Genet* 92:343–345.
107. Zlotogora J (2007) Multiple mutations responsible for frequent genetic diseases in isolated populations. *Eur J Hum Genet* 15:272–278.
108. Bonnen PE, et al. (2006) Evaluating potential for whole-genome studies in Kosrae, an isolated population in Micronesia. *Nat Genet* 38:214–217.
109. Lowe JK, et al. (2009) Genomewide association studies in an isolated founder population from the Pacific island of Kosrae. *PLoS Genet* 5:e10000365.
110. Auton A, et al. (2009) Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res* 19:795–803.
111. Lyons EJ, Frodsham AJ, Zhang L, Hill AVS, Amos W (2009) Consanguinity and susceptibility to infectious diseases in humans. *Biol Lett* 5:574–576.
112. Crimmins EM, Finch CE (2006) Infection, height, and longevity. *Proc Natl Acad Sci USA* 103:498–503.

Natural selection in a contemporary human population

Sean G. Byars[a], Douglas Ewbank[b], Diddahally R. Govindaraju[c], and Stephen C. Stearns[a,1]

[a]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520-8102; [b]Population Studies Center, University of Pennsylvania, Philadelphia, PA 19104-6299; and [c]Department of Neurology, Boston University School of Medicine, Boston, MA 02118-2526

Edited by Peter T. Ellison, Harvard University, Cambridge, MA, and approved September 16, 2009 (received for review June 25, 2009)

Our aims were to demonstrate that natural selection is operating on contemporary humans, predict future evolutionary change for specific traits with medical significance, and show that for some traits we can make short-term predictions about our future evolution. To do so, we measured the strength of selection, estimated genetic variation and covariation, and predicted the response to selection for women in the Framingham Heart Study, a project of the National Heart, Lung, and Blood Institute and Boston University that began in 1948. We found that natural selection is acting to cause slow, gradual evolutionary change. The descendants of these women are predicted to be on average slightly shorter and stouter, to have lower total cholesterol levels and systolic blood pressure, to have their first child earlier, and to reach menopause later than they would in the absence of evolution. Selection is tending to lengthen the reproductive period at both ends. To better understand and predict such changes, the design of planned large, long-term, multicohort studies should include input from evolutionary biologists.

evolutionary rates | heritability | *Homo sapiens* | medical traits

Are contemporary humans experiencing natural selection and evolving in response to it? The answer to that question depends on whom one asks. A long tradition in the medical community (1) holds that natural selection does not operate on contemporary human populations because medicine keeps "alive many who otherwise would have perished" (2). No evolutionary biologist would now agree with that claim, for natural selection works through differential reproductive success rather than simple differential survival, and individuals in contemporary human populations vary in lifetime reproductive success (LRS). Selection operates on any trait that varies and is correlated with LRS, and traits respond to selection with change across generations if they vary genetically. But what traits is selection operating on? Do they include the traits treated by physicians? Previous work (e.g., ref. 3) has shown that human life history traits, most significantly age at first reproduction, are currently under selection, but evidence for selection operating on traits of medical importance is scarce. Here, we report estimates of natural selection, and the potential genetic response to selection, in the women of the first two generations of the Framingham Heart Study (FHS) population. The traits we analyzed include traits of medical significance: total cholesterol (TC), systolic blood pressure (SBP), diastolic blood pressure (DBP), and blood glucose (GLU). We had three general aims: first, to correct the still widespread misconception that natural selection is not operating on contemporary humans; second, to make quantitative predictions about future evolutionary change for specific traits with medical significance; and third, to register firmly a point of general cultural interest that follows directly from our first two aims: We are still evolving, and for some traits we can make short-term predictions about our future evolution.

The Framingham Heart Study

The FHS was established in 1948 in Framingham, MA, by the National Heart, Lung, and Blood Institute and Boston University to identify factors that contribute to cardiovascular disease. It is the longest running multigenerational study in medical history. The people originally enrolled in the study were of predominantly European ancestry (20% United Kingdom, 40% Ireland, 10% Italy, 10% Quebec). The original cohort (n = 5,209) has been examined every 2 years, a total of 29 times between 1948 and 2008. The offspring cohort (n = 5,124) has been examined approximately every 4 years, a total of eight times between 1971 and 2008 (4). There is also a third generation cohort (n = 4,095) that is not included in this study because many in it have not yet completed reproduction. At each examination many physical and blood chemistry traits are measured and a questionnaire is administered, yielding data on >70 traits. Data are deidentified by the FHS and delivered to the National Institutes of Health dbGaP database, from which we downloaded them for analysis. In this study, we use only the data on individuals who were measured three or more times.

Measuring Selection in a Multicohort Medical Study

Natural selection has been measured many times in natural populations of animals and plants (5) using methods inspired by Robertson (6), developed by Lande and Arnold (7), and refined by Janzen and Stern (8), Hereford et al. (9), and others. To apply those methods to contemporary human populations requires consideration of several special features of data on humans. Some, such as cultural variation related to education, smoking, and medication, we dealt with as covariates. Others could in principle be measured on natural populations of animals and plants but in practice often are not; these include repeated measures on individuals that establish the developmental trajectories of multiple traits with age and long-term observations of populations across several generations that reflect secular demographic trends. Both make the measurement of traits more complex: at what age and in what portion of a secular trend—a change in conditions across time rather than age—should the expression of the trait be measured? The solution we chose was to calculate the response surface of each trait for age and time and to express the measurement of that trait for each individual as an average deviation from that surface (e.g., Fig. 1). Thus, for several traits we asked whether through their adult years individuals tended to have higher or lower values than other individuals of the same age measured in the same year. Because many individuals have been measured repeatedly in the FHS, the response surface can be estimated accurately. And how should

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Evolution in Health and Medicine" held April 2–3, 2009, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS web site at www.nasonline.org/Sackler_Evolution_Health_Medicine.

Author contributions: D.R.G. and S.C.S. designed research; S.G.B. performed research; S.G.B. and D.E. analyzed data; and D.R.G. and S.C.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: stephen.stearns@yale.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0906199106/DCSupplemental.

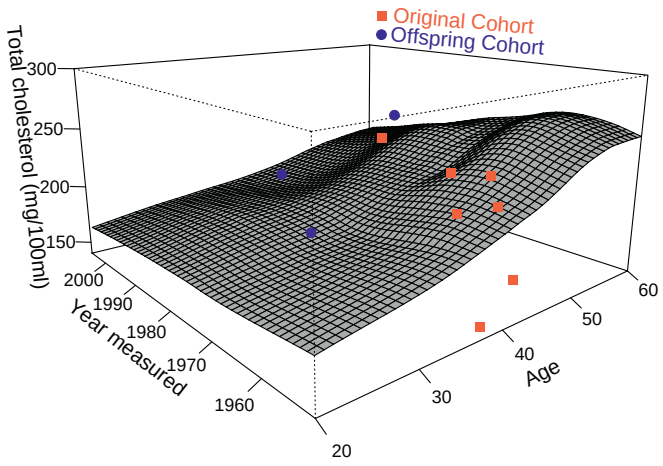


Fig. 1. TC between the ages of 20 and 60, 1955–2003. Because the measurement of TC changes with the age of the individual and the year of measurement, we fitted a surface to all measurements, calculated the residuals of all of the measurements for each individual from this surface, and expressed the trait measurement as a single value: the average of the residuals. We only used individuals who had been measured at least three times (n = 2,227 women measured 16,516 times for TC, with one individual's measures from each cohort added for reference).

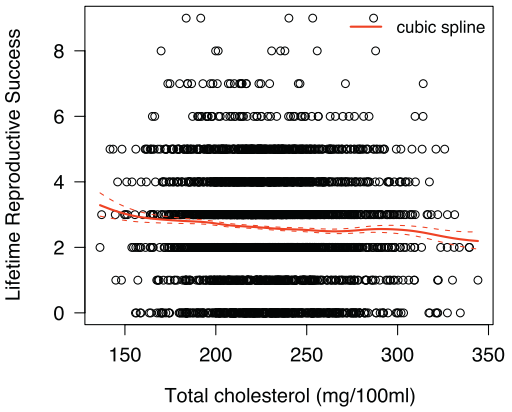


Fig. 3. Plot of average cholesterol values against LRS. Residuals from the surface in Fig. 1 were converted back to the original metric by adding the global mean to each value and then averaged to yield a single average residual estimate for each woman. A cubic spline [using glms40 (25)] was then fitted to these residuals, which have the effects of age and year of measure removed, to give a visual impression of the variation in the data and the strength of selection. The cubic spline is not the linear selection gradient. Dashed lines are ± 1 SE of fitness associations (solid line).

one express relative LRS in a population that goes through a baby boom and a baby bust? We asked whether an individual had relatively high LRS compared with others bearing children during the same time period. To do so we divided the population into six periods of differing average LRS by year of birth and expressed the fitness of women relative to the average for their year of birth (Fig. 2).

Measuring Evolutionary Change

To study evolutionary change, one conceives of the process as occurring in two steps. First, selection acts through differences in LRS. With multitrait phenotypes there are both direct effects on the individual traits and indirect effects mediated by the phenotypic correlations among the traits. Thus, a trait may be favored by selection only because it is correlated with other traits that are directly associated with greater LRS. The direct effects of selection are captured by the linear selection gradients (β) estimated as partial regression coefficients (Table 1); the indirect effects are mediated by the phenotypic variance/covariance matrix P (Table 2).

Second, the multitrait response to selection also depends on direct and indirect effects of inheritance. Selection differentials will have no long-term consequences if the traits are not passed to the next generation. The direct effects on the response are determined by the additive genetic variances of the individual traits; these are the diagonal elements in the genetic variance–covariance matrix G (Table 3). The indirect effects are mediated by the off-diagonal elements of that matrix, the genetic covariances. Inheritance can either be genetic or cultural. We are not able to fully differentiate the effects of genes and culture with these data.

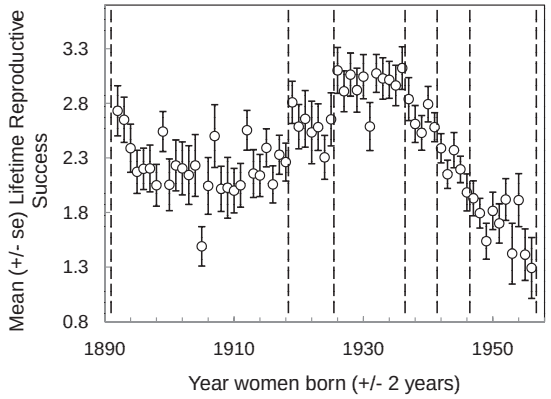


Fig. 2. LRS, by year of birth, for women in the FHS. To deal with secular demographic change, we divided the data into six periods and divided the relative reproductive success of each woman by the mean reproductive success of the women in her group.

Selection Gradients

Selection gradients measure the extent to which individuals with a given value of a quantitative trait tend to have higher or lower fitness (LRS). If individuals with low levels tend to have higher LRS, then the next generation is apt to have a lower average level of that trait (Fig. 3). We measured significant linear selection gradients acting to reduce total cholesterol (TC), height (HT), SBP, and age at first birth and to increase weight (WT) and age at menopause. These results strongly suggest that natural selection is acting on the women of the FHS through differences in the number of children they had during their reproductive ages (Table 4).

In estimating the selection gradients for most of the traits, we controlled for level of education, foreign or native birth, and whether the individual smoked. We also estimated quadratic terms to detect stabilizing selection (Table S1). These terms indicate whether a continuation of past trends will lead a trait to change at a steady or increasing rate or whether it will tend to level off. A total of 5.74% of the variation in LRS was explained by all factors combined (linear, quadratic, and interaction terms). The impact of stabilizing selection is reflected in our projections of evolutionary change (Table 5).

Inheritance of Traits

We estimated the inheritance of traits by taking advantage of the longitudinal, multigeneration, family design for the FHS. It

Table 1. Mean values and direct and total selection gradients

	TC, mg/100 mL	WT, kg	HT, cm	SBP, mmHg	DBP, mmHg	GLU, mg/100 mL	Age at menopause	Age at first birth
Means	2.350	1.811	2.205	2.106	1.895	1.950	1.689	1.418
β	−0.743	0.861	−3.999	−0.963	0.982	−0.848	1.280	−1.267
$P \times \beta$	−2.841	2.581	−0.533	−0.323	0.526	−2.554	1.981	−6.940

Means, − $\log_{10}$. β, linear selection gradient from partial regressions. $P \times \beta$, direct and indirect selection from P matrix, × 1,000.

yielded pedigrees, some quite complex, for hundreds of families. We applied a maximum-likelihood method implemented in the software package SOLAR (24) to extract from these pedigrees estimates of additive genetic variance and covariance. The heritability of a trait, perhaps a more familiar term, is its additive genetic variance divided by its total phenotypic variance, and the genetic correlation between two traits is their additive covariance divided by the square root of the product of their additive variances. Thus, in essence we were measuring the heritabilities of and genetic correlations between all of the traits and expressing them in a form suitable for evolutionary projections. This method cannot completely discriminate between vertical cultural transmission and genetic transmission, but the use of all degrees of relationship, not just those between parents and offspring, does to some degree get around this problem. Our estimates of heritabilities were: HT, 0.84 ± 0.01 (SE); TC, 0.61 ± 0.02; SBP, 0.53 ± 0.02; WT, 0.52 ± 0.02; DBP, 0.49 ± 0.02; age at menopause, 0.47 ± 0.05; GLU, 0.34 ± 0.02; and age at first birth, 0.09 ± 0.02. These are similar to heritabilities found in other studies examining phenotypic (10, 11) and life history traits (12) in humans.

Projecting Evolutionary Change

Table 5 presents the projected changes on the assumption that conditions will continue to mirror the averages encountered by this population over the past 60 years. In the next 10 generations mean TC among women is projected to decline from the average of 224 over the past 60 years to 216 (209.3–222.5) mg/100 mL (95% C.I.; see *Methods*). Because the environment has changed over the past 60 years (Fig. 2) and will continue to change, these results suggest that whatever changes in environment occur evolutionary changes will lead to mean cholesterol levels among women that are ≈0.8 (0.14–1.46) mg/100 mL lower in the next generation than they would be in the absence of evolution.

Similarly, we expect that as a result of evolution, in the next generation mean body WT among women will increase by 0.2 (−0.20 to 0.62) kg then stabilize; HT will decrease a bit, ≈0.2 (0.03–0.39) cm; SBP will decrease by ≈0.25 (−0.05 to 0.53) mmHg; DBP will remain essentially unchanged; blood GLU will decrease slightly by 0.8 (−0.09 to 0.29) mg/100 mL; age at menopause will increase by ≈1.0 (−0.23 to 2.15) months; and age at first birth will decrease by ≈0.5 (−0.6 to 1.7) months. The rates of projected evolution in haldane units (SD per generation) range from 0.032 (HT) to 0.002 (DBP), slower than those estimated for Galapagos finches and Trinidadian guppies but comparable to those estimated for New Zealand chinook salmon and Hawaiian mosquitofish (13). In sum, as a result of evolution future generations of women in this population are predicted to be slightly shorter and stouter, to have lower values for TC and SBP, to have their first child slightly earlier, and to reach menopause slightly later than they would have otherwise. These are small, gradual evolutionary changes in the middle to lower range of those observed in contemporary populations of non-human species.

Secular Changes

To see whether selection intensities were changing over the period of the study, we broke the dataset down into three periods defined by year of birth: 1892–1913 (n = 716), 1914–1935 (n = 842), and 1936–1956 (n = 669). Most of the significance in the selection gradients we estimated in Tables 1–3 was contributed by variation among women in the first period. The only trait consistently under significant selection in all three periods was age at first birth (period 1: $\beta = -1.120$, $P = 0.0037$; period 2: $\beta = -1.107$, $P = 0.0018$; period 3: $\beta = -1.488$, $P = 0.018$).

Would Unrecorded Early Mortality Have Changed Any of Our Conclusions?

LRS should be measured from birth to the end of reproduction. However, we could only study individuals who survived to adulthood and had measurements on adult traits. During most of human history, high mortality at early ages was a major factor driving evolution, but now most children survive well into the reproductive ages.

To determine whether excluding early deaths from LRS could have biased our results, we simulated what would happen if a trait was associated with early death and, therefore, no reproduction. For example, we asked what would happen if higher cholesterol was associated with an increased risk of not surviving to age 20 and, therefore, remaining childless? What would the sample look like if we were to include those people who died? We might see a group of childless women whose adult (never actually measured) cholesterol levels were on average higher than other women. We simulated such effects by adding to the dataset a group of phantom women for two levels of preadult mortality, 0.011 (survival to age 20, $p_{20} = 0.989$, the average for women in the United States in 2002) and 0.06 ($p_{20} = 0.94$, the average for women in the United States in 1939–1940) and for

Table 2. Phenotypic variance/covariance matrix

	TC, mg/100 mL	WT, kg	HT, cm	SBP, mmHg	DBP, mmHg	GLU, mg/100 mL	MEN	BIR
TC	3.860	0.022	−0.089	0.336	0.278	0.300	−0.180	0.112
WT	0.022	5.710	0.426	1.040	1.070	1.030	0.045	0.040
HT	−0.089	0.426	0.266	−0.039	0.009	−0.023	0.022	0.109
SBP	0.336	1.040	−0.039	2.590	1.800	0.688	−0.005	−0.222
DBP	0.278	1.070	0.009	1.800	1.920	0.399	−0.020	−0.183
GLU	0.300	1.030	−0.023	0.688	0.399	3.570	−0.032	0.021
MEN	−0.180	0.045	0.022	−0.005	−0.020	−0.032	1.460	0.094
BIR	0.112	0.040	0.109	−0.222	−0.183	0.021	0.094	5.378

P matrix, $\log_{10}$, all values × 1,000. MEN, age at menopause; BIR, age at first birth.

Table 3. Additive genetic variance/covariance matrix

	TC, mg/100 mL	WT, kg	HT, cm	SBP, mmHg	DBP, mmHg	GLU, mg/100 mL	MEN	BIR
TC	2.371 ± 0.097	−0.126 ± 0.101	−0.070 ± 0.022	0.180 ± 0.068	0.108 ± 0.059	0.071 ± 0.080	0.040 ± 0.081	−0.095 ± 0.089
WT	−0.126 ± 0.101	3.014 ± 0.143	0.376 ± 0.021	0.250 ± 0.079	0.408 ± 0.065	0.329 ± 0.090	0.030 ± 0.100	0.002 ± 0.103
HT	−0.070 ± 0.022	0.376 ± 0.021	0.225 ± 0.005	−0.020 ± 0.017	0.022 ± 0.015	−0.035 ± 0.020	0.027 ± 0.029	0.074 ± 0.024
SBP	0.180 ± 0.068	0.250 ± 0.079	−0.020 ± 0.017	1.386 ± 0.066	0.723 ± 0.018	0.454 ± 0.060	−0.090 ± 0.064	−0.127 ± 0.072
DBP	0.108 ± 0.059	0.408 ± 0.065	0.022 ± 0.015	0.723 ± 0.018	0.950 ± 0.050	0.226 ± 0.055	−0.088 ± 0.056	−0.173 ± 0.065
GLU	0.071 ± 0.080	0.329 ± 0.090	−0.035 ± 0.020	0.454 ± 0.060	0.226 ± 0.055	1.235 ± 0.098	0.040 ± 0.053	−0.121 ± 0.080
MEN	0.040 ± 0.081	0.030 ± 0.100	0.027 ± 0.029	−0.090 ± 0.064	−0.088 ± 0.056	−0.052 ± 0.053	0.695 ± 0.084	−0.054 ± 0.043
BIR	−0.095 ± 0.089	0.002 ± 0.103	0.074 ± 0.024	−0.127 ± 0.072	−0.173 ± 0.065	−0.121 ± 0.080	−0.054 ± 0.043	0.537 ± 0.137

G matrix, $\log_{10}$, values × 1,000, ± 1. MEN, age at menopause. BIR, age at first birth.

each of 10 traits, by giving each phantom individual a value for the trait equal to the population mean plus 2 SD. For 90% of the 20 cases, the unstandardized β values remained in the same direction, positive or negative, and for 86% of the traits P values did not change significance. For example, selection on cholesterol was always negative and always significant (20 of 20 cases); selection on HT was always negative (20 of 20) and almost always significant (19 of 20); selection on WT was almost always positive (18 of 20) and usually significant (16 of 20); selection on age at menopause was almost always positive (19 of 20) and almost always significant (19 of 20); selection on SBP was usually positive (17 of 20) but only significant approximately half the time (11 of 20); selection on GLU was almost always negative (19 of 20) but rarely significant (2 of 20); patterns in the other traits were mixed. Apparently the proportion not surviving to age 20 is so low that if they were in the dataset our major conclusions would not be likely to change.

Conclusions

Natural selection is acting slowly and gradually on traits of medical importance and on life history traits in the FHS population. Selection varied in intensity, becoming generally less intense over time, but not in direction, and it has only operated consistently over the entire period to reduce age at first birth. Predictions for one generation are fairly reliable, but whether selection will be consistent and sustained enough to bring about significant genetic change can only be answered with longer periods of observation of more traits relevant to human health.

These results suggest slow evolutionary change. It is noteworthy, although not surprising, that both age at first birth and age at menopause appear to be changing so as to lengthen the reproductive period, which is consistent with previous findings (3). Because fertility is the driving force behind evolution in modern populations, we might have found larger effects of evolution on the levels of sex hormones and related traits had they been measured. The impact of fertility on selection could prove especially important now that many couples that would otherwise remain childless can produce offspring with medical assistance.

The traits we studied were those available from the FHS, which was focused on heart disease, not reproduction. To better understand and predict evolutionary changes, the design of planned large, long-term, multicohort studies should include input from evolutionary biologists and, in particular, should consider measuring traits that might be closely associated with reproductive success.

Methods

The FHS continues today. Study design and entry criteria for the FHS are detailed in Dawber et al. (14) and Kannel et al. (15). We included subjects from the original and offspring cohorts; they received physical examinations and questionnaires administered by trained interviewers every 2 and 3–4 years, respectively. The Institutional Review Board requirements have been adequately addressed for all of the participants in the FHS, and formal approval for the data used was obtained from the dbGaP (www.ncbi.nlm.nih.gov/sites/entrez?Db = gap).

Relative Fitness. We calculated LRS for women who reported at a postmenopausal age how many live births they had had. Of the 5,372 women in the original and offspring cohorts, 739 were removed because of missing values for menopause (or menopause had not been reached) and 852 were removed because the number of live births was not recorded after a postmenopausal age. Sample size was further reduced to 3,224 after excluding an additional 557 women who reached menopause unnaturally (e.g., ovaries removed, hysterectomy, radiation, chemotherapy) before the age of 45 and to 2,238

Table 4. Selection gradients acting on women in the Framingham population (combined dataset, women born between 1892 and 1956)

Trait	N	μ	SD	β	β_μ	P	% linear	% Quad
TC, mg/100 mL	2,227	223.2	48.2	−0.743	−1.447	0.0011	0.295	
HT, cm	2,227	160.7	6.4	−3.999	−6.672	0.0002	0.689	
WT, kg	2,227	65.7	13.3	0.861	0.985	<0.0001	0.197	0.394
SBP, mmHg	2,227	127.2	22.1	−0.963	−1.545	0.0236	0.019	
DBP, mmHg	2,227	79.1	11.9	0.982	1.317	0.0724	-	
GLU, mg/100 mL	2,227	89.6	22.8	−0.848	−0.548	0.0596	-	0.492
Diabetes	2,227	-	-	0.076	0.008	0.3473	-	
Age at first birth	1,448	26.5	4.6	−1.267	−1.758	<0.0001	*	
Age at menopause	2227	49.2	4.1	1.280	2.839	0.0035	0.689	0.098

Associations with age at first birth were estimated only for women who reported at least one live birth. N, sample size of individual women; μ, mean; β, unstandardized directional selection gradient from a multiple linear regression model that included interaction and quadratic terms; β_μ, mean standardized selection gradient; P, significance estimates from the Poisson model; %, percentage of variation in lifetime reproductive success explained. Presenting the mean and SD for diabetes would be inappropriate because it was recorded as presence/absence.*, The percentage of variation in lifetime reproductive success explained by age at first birth is not strictly comparable to that explained by the other traits because the sample was smaller; in that separate analysis it explained ≈5%. One covariate, level of education, was significant; it explained 0.295% of the variation in lifetime reproductive success. Only one of the 45 two-way interaction terms was significant, estrogen therapy × diastolic blood pressure; it explained 0.197% of the variation in lifetime reproductive success.

Table 5. Projected evolutionary change, untransformed values

	TC, mg/100 mL	WT, kg	HT, cm	SBP,mmHg	DBP, mmHg	GLU, mg/100 mL	Age at menopause	Age at first birth
Generation								
0	223.9	64.7	160.2	127.6	78.5	89.1	48.9	26.18
5	219.8 ± 1.58	65.4 ± 0.58	159.3 ± 0.41	126.3 ± 0.71	78.6 ± 0.34	88.4 ± 0.52	49.3 ± 0.23	25.96 ± 0.23
10	215.9 ± 3.35	65.6 ± 1.33	158.1 ± 0.90	125.2 ± 1.47	78.7 ± 0.73	88.1 ± 1.03	49.7 ± 0.51	25.74 ± 0.48
%	3.6	1.4	1.3	1.9	0.3	1.1	1.6	1.7
Haldanes	0.016	0.007	0.032	0.010	0.002	0.004	0.020	0.010

The averages for generation 0 are those observed for the Framingham original and offspring cohorts between the ages of 20 and 60 over the years 1955–2003. Values for later cohorts are the means that would be expected given the average conditions over the period of observation and evolutionary change. %, percent change in traits from generation 0 to 10. Haldanes: rate of evolution in SD per generation.

after excluding those who had been measured <3 times for any of the continuous traits. Accuracy of LRS estimates was improved by using the FHS pedigree file (and other data) to correct for number of live births >5. Twenty-four women (1.0%) who had five or more live births were recorded as having had five as part of the deidentification of the data. We estimate that ≈7 of those 24 women actually had more than the five births (generally only one more) indicated by the dataset; we recorded all 24 as having had five births.

Main Phenotypes. We used some traits known to be associated with cardiovascular disease. They included TC, WT, HT, SBP, DBP, and serum GLU. TC included both serum (earlier survey rounds) and plasma (later) measures; both were treated as one here. Blood pressure was measured twice in each examination; only the first physician's measure was used here. Methods for the various cholesterol measures throughout the study have been described (16, 17).

For both cohorts, WT, HT, SBP, DBP, and GLU were measured from examination 1 (1948 and 1971 onward for original and offspring cohorts, respectively). Number of measures per trait varied depending on how many exams a woman attended throughout the study.

We found complex nonlinear changes in traits with age and year of measure that may reflect demographic and developmental changes throughout the study period and a woman's life. We removed these effects by taking residuals from a 3D surface of a generalized additive model of each trait by age and year of measure, then converted these back into the original metric by adding the overall trait mean (Fig. 2). We used measures across the ages of 20–60, which span the reproductive years of modern human women, and only included women with three or more measures for each trait, because there is considerable intraindividual trait variation over time (18, 19). Residuals were then averaged to obtain one point per trait per woman and log-transformed to correct for deviations from normality.

Because traits may be affected by number of offspring (e.g., women might gain WT with number of births), thereby potentially confounding the covariation between traits and fitness, we compared traits measured at an age when women had not yet reproduced (according to age at first birth estimates from the FHS pedigree) or at the earliest age a trait was measured (up to the age of 30), to the same traits measured at an age when women in the sample had completed fertility (between the ages of 45 and 55). We found no significant difference in an ANOVA for any of the traits examined, suggesting few, if any, cumulative effects of number of offspring on the traits examined.

Covariates. Demographic covariates included level of education and whether foreign or native born. We assumed that women from the offspring cohort were all born in the United States. Education was coded as number of years of education completed; the few women with missing values here were assumed to have a minimum of 8 years of education.

Other covariates included self-reported use of cholesterol-lowering medication (e.g., statins, fibrates, resins), self-reported use of medication for hypertension or high blood pressure, presence or absence of diabetes mellitus, self-reported smoking status, and estrogen use (hormone therapy or contraceptive use), which can affect cholesterol levels (20), other traits (21, 22), and women's LRS. We only used information on medication use across the ages of 20–60. Smoking status was recorded as whether a woman was or had been a smoker. Of the remaining 2,238 women, only 11 (or 0.49%) were recorded as having taken cholesterol-lowering medication between the ages of 20 to 60; they were removed from the dataset, yielding a final sample size of 2,227. All covariates (except smoking intensity and education level) were coded as binary (present/absent) in the multiple regression analyses.

Projections. We estimated the effect of directional selection on the mean value of each trait in the population in the next generation (the response to selection) with this equation:

$$\Delta z = G\beta.$$

Δz is a vector of predicted changes in the population means of characters between the observed and next generation, G is the additive genetic variance-covariance matrix, and β is a vector of the selection gradients. Predicted changes in trait means should be interpreted cautiously for the direction and intensity of selection and the additive genetic variances and covariances all change over time and across generations.

To estimate the cumulative effect over several generations, we projected the mean change between generations using the equation $\Delta z = G\beta = GP^{-1}s$, where P is the phenotypic variance-covariance matrix and s is the vector of changes in the traits that would occur in one generation as a result of selection alone. The change between generations in the variance–covariance matrix for traits was then estimated as $P^* - P = P\gamma P - ss^T$, where γ is the matrix of interaction terms among the traits calculated from the coefficient estimates from a linear regression (7). We assumed that G, s, and γ remain fixed across generations.

Confidence Intervals for Projections. The confidence intervals for the projection reflect the uncertainty in the β, γ, and G matrixes by using Monte Carlo simulations. Random number generators were used to run 5,000 projections with different estimates for each parameter. The β and γ were handled simultaneously by using a multinomial random number generator and the matrix of variances and covariances among the regression coefficients (including the intercept). The coefficients on the linear and quadratic terms were then used to set up β and γ. Each element of the G matrix was handled separately by using a random number generator and the standard errors produced by SOLAR (24). The projections were all done on the log of traits.

Statistical Analysis. LRS from women measured for TC, WT, HT, SBP, DBP, GLU, age at menopause, age at first birth, age at death, and level of education were used to assess whether these traits influenced women's fitness. Multiple linear regression was used to infer the strength and direction of selection, and multiple Poisson regression was used to test statistical significance.

We adjusted LRS to account for fluctuations in fertility over time (Fig. 2) relative to women in the same birth cohort (birth years 1892–1918, 1919–1925, 1926–1936, 1937–1941, 1942–1946, 1947–1956). Relative fitness was calculated by dividing each woman's number of births by the mean for her birth cohort. The Poisson models included binary markers for each cohort. We first used education level, smoking, and country of birth as controls to remove their potentially confounding effects on LRS. When we then assessed selection on cultural traits, we included education level and smoking as covariates. In all analyses we included quadratic terms for the main traits (TC, WT, HT, SBP, DBP, GLU, age at menopause, age at first birth) and two-way interactions between traits and covariates. Quadratic coefficients were doubled to estimate stabilizing/disruptive selection gradients (23). Coefficients from the multiple linear regressions were standardized by using trait means [see Hereford et al. (9)]. Partial regression coefficients, termed directional selection gradients (β), estimate the magnitude of directional selection on a trait i, with the effects of selection on all of the other measured traits removed.

***G* Matrix Estimates.** We used average residuals (described above) to control for changes in heritability over time and age. Additive genetic variance and covariance estimates were made with a pedigree-based maximum-likelihood

method implemented in SOLAR (24). Significance was determined by likelihood ratio tests.

ACKNOWLEDGMENTS. We thank the National Evolutionary Synthesis Center for support and the other members of our National Evolutionary Synthesis Center working group (Charles Goodnight, David Houle, Martin Larson, Trudy Mackay, Shamil Sunyaev, Anatoli Yashin, and Sebastian Zoellner) for helpful comments. D.E. was supported by National Institutes of Health (National Institute on Aging) Grant P30 AG12836, the Boettner Center for Pensions and Retirement Security at the University of Pennsylvania, and National Institutes of Health (National Institute of Child Health and Development Population Research Infrastructure Program) Grant R24 HD-044964.

1. Tait L (1869) Has the law of natural selection by survival of the fittest failed in the case of man? *Dublin Q J Med Sci* 47:102–113.
2. Bynum WF (1983) Darwin and the doctors: Evolution, diathesis, and germs in 19th century. *Gesnerus* 40:43–53.
3. Kirk KM, et al. (2001) Natural selection and quantitative genetics of life-history traits in Western women: A twin study. *Evolution (Lawrence, Kans)* 55:423–435.
4. Govindaraju DR, et al. (2008) Genetics of the Framingham Heart Study population. *Adv Genet* 62:33–65.
5. Kingsolver JG, et al. (2001) The strength of phenotypic selection in natural populations. *Am Nat* 157:245–261.
6. Robertson A (1966) A mathematical model of culling process in dairy cattle. *Anim Prod* 8:95–108.
7. Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution (Lawrence, Kans)* 37:1210–1226.
8. Janzen FJ, Stern HS (1998) Logistic regression for empirical studies of multivariate selection. *Evolution (Lawrence, Kans)* 52:1564–1571.
9. Hereford J, Hansen TF, Houle D (2004) Comparing strengths of directional selection: How strong is strong? *Evolution (Lawrence, Kans)* 58:2133–2143.
10. Brown WM, et al. (2002) Age-stratified heritability estimation in the Framingham Heart Study families. *BMC Genet* 4(Suppl 1): S32.
11. Mathias RA, et al. (2002) Comparison of year-of-exam and age-matched estimates of heritability in the Framingham Heart Study data. *BMC Genet* 4(Suppl 1): S36.
12. Kohler HP, Rodgers JL, Christensen K (1999) Is fertility behavior in our genes? Findings from a Danish twin study. *Popul Dev Rev* 25:253–288.
13. Hendry AP, Kinnison MT (1999) Perspective–The pace of modern life: Measuring rates of contemporary microevolution. *Evolution (Lawrence, Kans)* 53:1637–1653.
14. Dawber TR, Kannel WB, Lyell LP (1963) An approach to longitudinal studies in a community: Framingham Study. *Ann NY Acad Sci* 107:539–556.
15. Kannel WB, et al. (1979) Investigation of coronary heart disease in families: Framingham offspring study. *Am J Epidemiol* 110:281–290.
16. Abell LL, Levy BB, Brodie BB, Kendall FE (1952) A simplified method for the estimation of total cholesterol in serum and demonstration of its specificity. *J Biol Chem* 195:357–366.
17. Warnick GR, Benderson J, Albers JJ (1982) Dextran sulfate Mg^{2+} precipitation procedure for quantitation of high-density-lipoprotein cholesterol. *Clin Chem* 28:1379–1388.
18. Frishman WH, et al. (1992) Serum lipids and lipoproteins in advanced age: Intraindividual changes. *Ann Epidemiol* 2:43–50.
19. Bookstein L, Gidding SS, Donovan M, Smith FA (1990) Day-to-day variability of serum cholesterol, triglyceride, and high-density-lipoprotein cholesterol levels: Impact in the assessment of risk according to the National Cholesterol Education Program Guidelines. *Arch Intern Med* 150:1653–1657.
20. Demissie S, et al. (2006) Estrogen receptor-α variants are associated with lipoprotein size distribution and particle levels in women: The Framingham Heart Study. *Atherosclerosis* 185:210–218.
21. Fox CS, et al. (2005) Sex-specific association between estrogen receptor-α gene variation and measures of adiposity: The Framingham Heart Study. *J Clin Endocrinol Metab* 90:6257–6262.
22. Peter I, et al. (2005) Variation in estrogen-related genes and cross-sectional and longitudinal blood pressure in the Framingham Heart Study. *J Hypertens* 23:2193–2200.
23. Stinchcombe JR, et al. (2008) Estimating nonlinear selection gradients using quadratic regression coefficients: Double or nothing? *Evolution (Lawrence, Kans)* 62:2435–2440.
24. Blangero J, Almasy L, Dyer T, Peterson C (1999) Sequential oligogenic linkage analysis routines: SOLAR (Solar Software, Westerville, OH), Version 4.2.0. Available at http://solar.sfbrgenetics.org/.
25. Schluter D, Nychka D (1994) Exploring fitness surfaces. *Am Nat* 143:597–616.

Numbering the hairs on our heads: The shared challenge and promise of phenomics

David Houle[1]

Department of Biological Science, Florida State University, Tallahassee, FL 32306-4295

Edited by Diddahally R. Govindaraju, Boston University School of Medicine, Boston, MA, and accepted by the Editorial Board September 21, 2009 (received for review July 22, 2009)

Evolution and medicine share a dependence on the genotype–phenotype map. Although genotypes exist and are inherited in a discrete space convenient for many sorts of analyses, the causation of key phenomena such as natural selection and disease takes place in a continuous phenotype space whose relationship to the genotype space is only dimly grasped. Direct study of genotypes with minimal reference to phenotypes is clearly insufficient to elucidate these phenomena. Phenomics, the comprehensive study of phenotypes, is therefore essential to understanding biology. For all of the advances in knowledge that a genomic approach to biology has brought, awareness is growing that many phenotypes are highly polygenic and susceptible to genetic interactions. Prime examples are common human diseases. Phenomic thinking is starting to take hold and yield results that reveal why it is so critical. The dimensionality of phenotypic data are often extremely high, suggesting that attempts to characterize phenotypes with a few key measurements are unlikely to be completely successful. However, once phenotypic data are obtained, causation can turn out to be unexpectedly simple. Phenotypic data can be informative about the past history of selection and unexpectedly predictive of long-term evolution. Comprehensive efforts to increase the throughput and range of phenotyping are an urgent priority.

disease | genotype–phenotype map | natural selection | G matrix | dimensionality

Medicine emphasizes proximal cause, for example, in the case of infectious disease, exposure to disease-causing microbes, environmental and genetic factors that have shaped the properties of exposure, antimicrobial therapy, and treatment of symptoms. Evolutionary biology approaches these same factors retrospectively in terms of the evolutionary history of the microbe and human host, that is the factors that have shaped the niche of microbes and humans, the evolutionary factors that allow or promote the existence of genetic variants, and prospectively the potential for the microbe to evolve in response to our therapies. The two approaches are reciprocally illuminating.

I want to point out another point of contact between medicine and evolutionary biology that is less appreciated: they both depend on our knowledge of the relationship between genotype and phenotype, the genotype–phenotype (G-P) map. This concept has a long history in evolutionary thinking. An early and influential statement of the importance of the G-P map in evolutionary biology is that of Lewontin (1), whose map is redrawn in Fig. 1*A*. The evolutionary process takes place in two "spaces." The first is the genotype space (G space), which consists of all possible genotypes. Populations move in this space over generations in response to natural selection and genetic processes. Natural selection, however, takes place in continuous phenotype space (P space), the space of all possible phenotypes. The genotype of an individual strongly influences the location in P space through the process of epigenesis, the totality of interactions of genes and environment, including all aspects of development (2). The properties of the phenotype produced influence its probability of survival and success at reproducing its genotype. This process of weighting genotypes by phenotypic success (and potentially epigenetic inheritance) then indirectly changes the mean genotype. Finally, transmission influences the mean of the next generation through the processes of segregation, mutation, and recombination. This process is repeated over many generations.

Medical genetics seeks to understand the genetic causes of variation in human morbidity and mortality. As represented in Fig. 1*B*, doing so involves unraveling the same transformations in and between G and P spaces as does understanding the process of evolution. Epigenesis and transmission are the same in both realms; the determination of disease state from phenotypes is precisely analogous to natural selection, although disease state may or may not influence reproductive success. Proximal causation of disease state takes place in P space and must ultimately be studied there. Current methods for explaining G-P relationships are, however, based almost entirely on determining the positions of subpopulations in G space, bypassing P space except as a classifier. For disease genetics, individuals are rather crudely sorted into diseased and healthy subpopulations so that their genetic differences can be compared. Analogous approaches are commonly used for simple continuous phenotypes, such as human height. The techniques of Mendelian analysis, candidate gene studies, and association studies are in this sense all association studies.

Thanks to genomics, we now have, or can readily obtain, abundant population data on genotypes. In addition, efforts to extend high-throughput techniques to aspects of the epigenetic process relatively close to the genome, such as gene expression, protein interactions, and metabolism, have greatly increased our ability to detect genetic influences of these subcellular phenotypes. If we consider, however, multicellular phenotypes, such as morphology, physiology, and behavior, our capabilities have remained relatively unchanged over the last 20 years. Commercially available gene chips now allow the simultaneous assay of the expression of an entire genome, but the average investigator of variation in whole organism phenotypes is not far removed from previous generations who took out the calipers, made a single measurement, and wrote it down in a notebook with a pencil. As a result, the depth of our knowledge of genomes is approaching completeness, whereas our knowledge of phenotypes remains, by comparison, minimal. Part of the explanation for this strong imbalance is certainly that P space is vastly more vast than G space.

All biologists need the problem of G-P relationships to be solved, or at least thoroughly described, but the need in evolu-

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Evolution in Health and Medicine" held April 2–3, 2009, at the National Academy of Sciences in Washington, DC. The complete program and audio files of most presentations are available on the NAS web site at www.nasonline.org/Sackler_Evolution_Health_Medicine.

Author contributions: D.H. designed research, performed research, analyzed data, and wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission. D.R.G. is a guest editor invited by the Editorial Board.

[1]E-mail: dhoule@bio.fsu.edu.

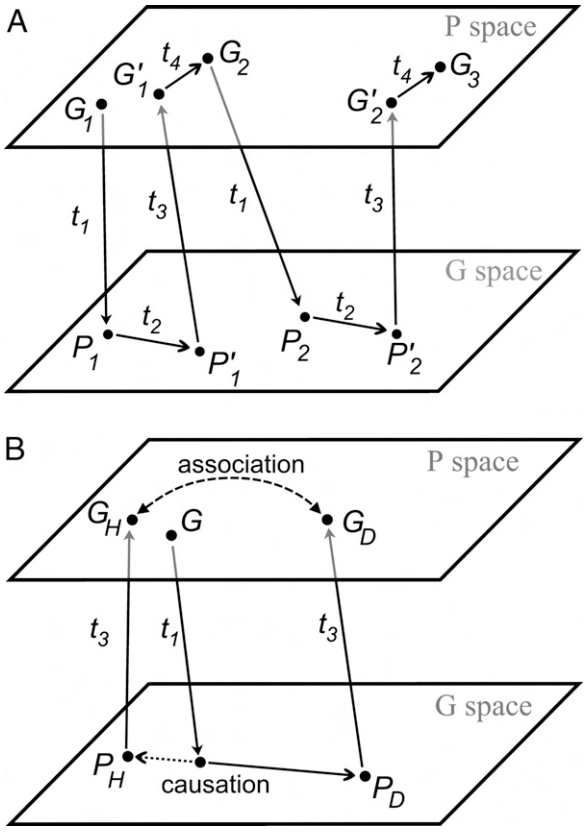


Fig. 1. G-P maps in evolution and medicine. Circles represent population mean genotypes and phenotypes, and arrows indicate the processes by which genotypic and phenotypic means are interconverted. (*A*) In the evolutionary realm, epigenesis (transformation t_1) transforms genomic information into the whole-organism phenotype. Natural selection (transformation t_2) alters the proportions of types within the population of phenotypes, potentially changing the phenotypic mean. This process alters the frequency of genotypes by transformation t_3, the inverse of epigenesis. Finally, reproduction results in transmission (t_4) of genotypes to the next generation, possibly again altering the mean genotype as a result of mutation and recombination. This process is repeated over many generations, moving both the population genotype and phenotype through their spaces. (*B*) The medical realm shares the process of epigenesis (t_1). Any influence of the phenotypes on the likelihood that an individual will be healthy or diseased is reflected in the mean phenotype of healthy (P_H) and diseased (P_D) individuals, in a process precisely analogous to natural selection. This differential sorting of genotypes depending on the phenotype they produce affects the genotypic means of healthy and diseased individuals through t_3. Differences between healthy and diseased subpopulations in G space are detected in association studies. Proximal causation of disease is studied in the P space.

tionary biology is particularly acute, because no predictive science of evolutionary dynamics can emerge without such understanding. The study of natural selection is even more primitive than our knowledge of phenotypes, but only by combining a G-P map with detailed knowledge of natural selection can one predict what aspects of the genome can evolve in response. As of now, however, this effort is pinned to the type of association studies diagrammed in Fig. 1*B*, which rely on crude, simplistic phenotypic measures to categorize individuals, and conduct the remainder of the analysis in G space.

Does this relative lack of phenotypic information matter? There is a growing realization that phenotypically naive association studies are unlikely to explain more than a minority of genetic causation (3, 4). For some phenotypes, even driving an association project to complete description seems likely to give us a list of thousands of genes and perhaps millions of variants, all with individually small effects. For these phenotypes we need alternative approaches to the G-P map.

Making sense of the evolutionary process requires that the phenotype as a whole be approached; understanding the causation of disease in human does as well. The large-scale study of high-dimensional phenotypes is phenomics; phenomics is the natural and inevitable complement to genomics. Implementation of a phenomic approach faces two critical challenges. One is obtaining comprehensive phenotypic data, and the second is learning how to use such data. The title of this article includes a quotation from Luke 12:7, where God is ascribed the power to evaluate the tiniest details of existence, not only to number the hairs of our heads, but to understand their meaning. Can we hope to do as well?

Studies in G Space

The thrust of modern biology and much of medicine is that the most effective way to understand phenotypes, including disease state and mortality, is to understand how genes function. The assumption is that failures in gene function either directly cause failures of organismal function or mimic the effects of failures with other causes. For example, epidemiologists have explicitly turned to the concept of "Mendelian randomization" (5) to test hypotheses about even environmental causes of disease. This approach exploits genetic variants that manipulate the effective factor hypothesized to be responsible for altered risk. Because relatives that share exposure to environmental factors may nevertheless differ in their relevant genotypes, we may get a less confounded picture of causation. Furthermore, we know that virtually all common diseases show evidence of inheritance. As the human genome was being published, optimism that these genetic and genomic approaches would unravel the majority of disease causation ran high (e.g., ref. 6).

This G-space approach to disease has undoubtedly had great successes. As of June 2009, The Online Mendelian Inheritance in Man database (7) listed 2,908 disorders that can be traced to defects at particular human genes. More than 3,700 additional disorders show evidence for inheritance, although they have not yet been traced to a precise genomic location, so these numbers are very likely to continue to increase. If progress along these lines can continue, then maybe we do not need to worry about our relative inability to measure phenotypes.

The available methods for detecting genetic association have important limitations that give reason to doubt how far we can go in understanding phenotypic causation (8). For example, traditional Mendelian analyses of disease state have been supercharged by the availability of abundant markers, but require that the overall probability of disease in the random population be low, a set of candidate loci be in hand, and both the penetrance (probability of developing the disease state when the variant is present) and effect size (detectability of the disease phenotype) of an allele be very high. When these assumptions are met, first-degree relatives, such as siblings, of those with the disease are at a high relative risk of disease (λ), and therefore have a high odds ratio (OR), a ratio of the probability of being affected when an allele is present to the probability of being affected when it is absent, which is typically approximately twice λ. Such analyses, therefore, only effectively uncover causation of rare syndromes that can be caused by single genes. Typical relative risk values for mapped Mendelian variants are well over 10. Some examples of Mendelian genetic diseases are cystic fibrosis ($\lambda = 500$), phenylketonueria ($\lambda = 500$), and sickle cell anemia ($\lambda = 18$) (7).

The relative risks of common, chronic, and late-onset diseases are typically much lower. The two leading causes of death in the United States are heart disease and cancer (9), but the average λ for 28 types of cancer was just 2.2 (10), whereas a typical number for heart disease is 3 (for the presence of coronary artery

disease in sibs of heart attack patients; ref. 11). Finding causation in such cases requires candidate gene or genomewide association (GWA) studies. GWA studies (e.g., ref. 12) are powerful when a causal allele is common enough to be present in multiple individuals in a sample and penetrance is as low as 10%, enabling alleles with small ORs of 1.1 or so to be detected, if sample size is very large. Common alleles, however, almost never have large effects (detected alleles usually have ORs between 1.2 and 1.5; ref. 8), and therefore explain little of the variance in disease. The candidate gene or "rare variant" approach intensively screens for variants in samples with and without the disease. Enrichment of rare variants in the diseased population, followed by verification that the function of the candidate gene is altered, indicates that altered function increases disease. Typical rare variants have ORs between 2 and 10, but individually explain little of the variance in disease susceptibility because of their rarity. The discovery that de novo copy-number variants are commonly associated with disease in a wide variety of syndromes (see, e.g., refs. 13 and 14) provides hope that these readily detected mutations will lead us to large numbers of new candidate genes. The GWA and candidate gene approaches are now being brought together by intensive resequencing of case-control populations (see, e.g., ref. 15).

This combination of available techniques is thus incapable of detecting variants with low penetrance or those at loci that are not yet candidates. These classes of variation appear to be quite common as only a small proportion of the total genetic causation can so far be assigned to a genomic location in most syndromes (3, 4), a result that suggests that low-penetrance alleles explain a substantial proportion of disease susceptibility. One response to this problem is to improve our methods, by explicitly addressing the mechanisms of low penetrance, such as genotype–environment interaction and epistasis (16).

Evidence is strong, however, that these holes in our understanding signal a deeper problem that cannot be fully addressed by better association studies. A particularly revealing set of GWA studies recently discovered multiple new regions with effects on height in human populations (17–19). The total sample size of genotyped individuals over those three studies was ≈85,000; each study accumulated large samples by combining data from many different GWA studies that incidentally recorded height. The studies collectively identified 52 loci that affect height, 40 of which were previously unknown. On one level, this result seems to confirm the usefulness of GWA studies (20), and each article proudly points out clustering of the associations near loci known to influence bone growth. The more important message, however, is the proportion of variation in height explained; the studies explain just 2.9%, 2.0%, and 3.7%, respectively of the variation in human height in populations of European ancestry. These figures suggests the distinct possibility that something approaching the entire genome is capable of influencing height (4), a conclusion supported by the finding that one-third of nonlethal mouse gene knockouts affect body weight (21).

Given these results, the goal of understanding GP relationships can probably advance only partway by association mapping. Many voices of caution have argued for a scaling back of the genomic rhetoric to match diminished expectations (8, 16). Others are ready to counsel that the entire enterprise of association mapping should be abandoned (e.g., refs. 3 and 22). To those focused on the overriding G-P problem the first response is eminently sensible, but unsatisfying, because it leaves no prospect of a solution. The second is defensible only if an alternative is available.

A Phenomic Alternative?

Fig. 1 suggests a natural alternative to G space studies that incorporates some concepts from quantitative genetics and evolutionary biology. These concepts are relatively straightforward if a single trait is involved, whose value can be symbolized p, with population mean P. First consider the study of natural selection during evolution, represented in Fig. 1*A*. Fitness is a function of trait value, f(p). When this function is standardized to a value of 1 at the population mean [i.e., f(P_1) = 1] the derivative at the population mean is the selection gradient, β. The gradient is approximated by the ratio between the covariance between f and p and the population variance in p, $\beta = COV_{f,P}/V_P$ (23), and gives the rate at which relative fitness changes for a unit change in the trait. It is readily estimated as the regression of relative fitness on trait value. The gradient allows the calculation of the change in mean phenotype after one round of natural selection (P_1 and P_1' in Fig. 1*A*) as $P_1' - P_1 = V_p\beta = COV_{f,P}$.

A very similar approach could be used to examine the function expressing the change in disease (or health) probability as a function of phenotype f(p). In this case, the means of healthy and diseased individuals (P_H and P_D; Fig. 1*B*) are readily calculated, and the gradients that transform the population mean to either P_H or P_D can be obtained as e.g., $\beta_H = (P_H - P)/V_P$. This procedure is equivalent to using an indicator of disease state as an analog for fitness. For example, to obtain the health gradient β_H we could use an indicator, x, that has a value of 1 for a healthy individual and 0 in a diseased one. If the proportion of healthy individuals is h, regressing x/h on p gives the disease gradient that expresses the change in relative probability of health for a unit change in trait value. When a large change in disease probability occurs in the range of the data, logistic regression will provide a better estimate of β_H (24).

This thinking is most useful when generalized to multiple traits (25), that is to P space, where the phenotype is a vector **p**. The weighting function, f(**p**) (which can be either a fitness function or a disease function) is a multidimensional surface. The derivative of this function is a gradient vector $\boldsymbol{\beta}$ with elements that summarize the direction in which f(**p**) fitness or disease state probability increases the most rapidly. Quadratic (or even higher-order) terms can also be fit, capturing the curvature of f(**p**) around the population mean. The resulting matrix of quadratic terms is called $\boldsymbol{\gamma}$ (25).

Estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ is by multiple regression and has the same advantages in this context as any other: if the function is well-behaved and the traits that actually cause the dependent variable to vary are in the analysis, the elements of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ will reveal the relative importance of each phenotype in determining the outcome. This could, for example, reveal which of a large number of possible phenotypes are most predictive of disease. If, however, some or all of the causal traits are missing from the analysis, because they are almost certain to be when only one trait is analyzed, the estimated $\boldsymbol{\beta}$ can underestimate or overestimate the importance of each trait, perhaps giving a misleading picture of which phenotypes matter (25, 26). Neither medical researchers nor evolutionary biologists currently have access to anything approaching complete phenotypic data, so full use of this approach awaits widespread implementation of phenomic-scale measurements.

In the evolutionary realm, the results of natural selection are transmitted back to the genotypes passed on to the next generation. The process of epigenesis (t_1 in Fig. 1) is to some degree indeterminate so individuals with the same genotype produce a variety of phenotypes. This variation itself may be partly predictable from the study of interactions between genotype and environment. As a result, some of the change in phenotype brought about by applying function f(**p**) is caused by the deviations of an individual from the average phenotype that its genotype would produce, and only some of the weighting function is transmitted back to G space in transformation t_3. The amount that is transmitted for a single trait is captured by the

additive genetic variance, V_A, which is that part of V_P that causes offspring to resemble their parents. The expected change in mean phenotype in a single trait between generations (neglecting the transformations caused by mutation and recombination, which are usually small) is then $P_2 - P_1 = V_A\beta$, which can be rearranged to give the more familiar form $P_2 - P_1 = V_A/V_P \cdot COV_{f,P} = h^2S$. In the multivariate case, inheritance is captured by a matrix of variances and covariances, **G**, where the diagonal elements are the additive genetic variances for each trait, and the off-diagonal elements are the additive genetic covariances between the traits. The resulting transformation across generations is then $\mathbf{P_2} - \mathbf{P_1} = \mathbf{G}\beta$.

Steps Toward Phenomics

Assessment of variation at a few locations in the genome was not enough to characterize location in G space, so we have turned to genomics. Similarly, the logic of natural selection and disease causation in P space makes clear that studying a few traits cannot be enough. In the last 10 years, calls for enhanced phenotyping have become increasingly common, although the logic behind these arguments has been varied and not always explicit (27–33). These calls have increasingly been taken up and led to concrete increases in our phenotyping ability (e.g., refs. 32 and 34–38) of differing scale and complexity.

Clearly, phenomic measurements must be extensive, covering many different aspects of the phenotype, such as morphology, behavior, physiology, etc. Less obviously, phenomics must also be intensive; that is, it must lead to detailed characterization of each major aspect of the phenotype. For example, the genetic variants that affect function of the human heart are very likely to have pleiotropic effects on other body parts and functions, calling for extensive measurements of other systems. In addition, the heart itself cannot be adequately characterized by a small number of summary parameters like cardiac output, but must be approached in terms of the full complexities of physiological capacities, morphology, etc., calling for intensive measurements of the heart. Phenomic efforts are rising to both challenges. For example, the mouse research community is adopting a standard set of protocols for extensive measurement covering many different aspects of the phenotype (37, 38), and intensive measurements of mouse morphology are being pursued by other groups (e.g., ref. 39). Most important, the mouse community is focused on associating this detailed phenomic data with particular genotypes and their recombinants.

An easy objection to putting resources into phenomics is that most of what we might measure may prove irrelevant. Although the genotype has a finite extent and discrete content and can therefore be measured exhaustively, the phenotype is both continuous in multiple dimensions and infinitely divisible in some dimensions. For example the state of the phenotype can be measured at an infinitely great number of time points. If the goal is exhaustive measurement of the phenotype, it will forever remain beyond our reach. Rather, the goal must be defined in terms of understanding. How intensively we need to measure the phenotype to achieve goals like understanding the proximate causation of natural selection or disease is an open question that must be addressed with respect to a particular goal, such as predicting susceptibility to a particular disease, or response to a particular selection pressure. Both the genotype and especially the phenotype are immensely complex; our hope must be that any particular problem becomes simpler when viewed from a favorable perspective. Buchanan et al. (22) nicely summarized this hope with the metaphor of an hourglass with the full genotype at one end and the full description of the phenotype at the other. In between, we hope, is the waist of the hour glass, where measurement of just a few key aspects of the organism (which could be any combination of genetic, environmental, and phenotypic measurements) are maximally informative about the problem we want to address, for example, fitness or disease susceptibility. This task is to increase the range of data that can be applied to a problem in hopes that, once the key pieces are in hand, we can build simplified, but powerful, models of causation.

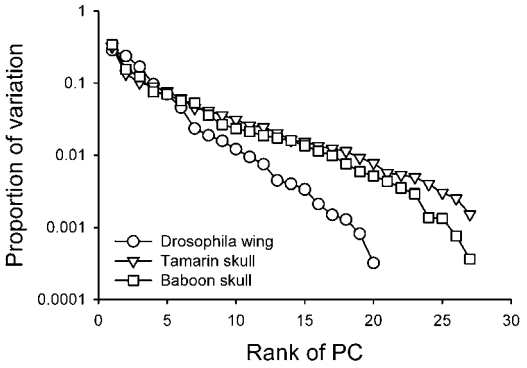


Fig. 2. Proportion of additive genetic variation on each principal component (PC) axis of the G matrix for three sets of morphological data: 30 measurements of baboon skulls (ref. 57 and C. Roseman, personal communication), 39 skull measurements of a combined estimate from two tamarin species, *Saguinus fusicollis* and *Saguinus oedipus* (refs. 58 and 59 and J. M. Cheverud, personal communication), and 12 *Drosophila melanogaster* wing vein intersections (42).

Early Lessons from Phenomic Data

Evolutionary biologists have been increasingly dealing with intensive datasets where a single aspect of the phenotype is subjected to detailed characterization, and the outcomes of these studies can give a taste of what we might learn by expanding such studies to more be more comprehensive. I discuss four areas where we can draw tentative conclusions not accessible from a genomic viewpoint alone.

Phenotypic Datasets Have High Dimensionality. Perhaps the largest class of highly multivariate datasets are obtained from studies of morphological form, assessed by means of the spatial locations of landmark points that are recognizable across a series of specimens or the curves that connect such points. In a relative handful of instances, such data have been subjected to a genetic analysis, in which the variation was partitioned into additive genetic variance and all other sources of variation. This process yields a **G** matrix with as many rows as the number of measurements, minus a few degrees of freedom for estimating the spatial orientation of each specimen relative to the others (40). When each **G** matrix is subjected to a principal components analysis (PCA), we can see just how much variation will be missed in studies that only measure of handful of traits. PCA rotates the **G** matrix to a new set of directions (the eigenvectors). Each direction has an amount of variation associated with it, its eigenvalue. The eigenvectors are chosen to maximize the range of the eigenvalues, so PCA allows us to measure the amount of variation in the least variable combinations of the original traits.

Three eigenvalue distributions are shown in Fig. 2, for measurements of baboon and tamarin and for *Drosophila* wing vein intersections. The analyses of these data remove size, leaving only variation in shape (41). On a log scale, the decrease in the amount of genetic variation in shape explained is approximately linear, suggesting an exponential distribution. What is most remarkable is how slowly the amount of variation falls from the kth most variable direction to the $k + 1$th direction. For the tamarin and baboon skulls, the variation falls by an average of 18% with each dimension, whereas for the flies it falls at the rate of 29% per dimension. No single summary measure can capture even 50% of the variation in the shape of these structures.

A second question is just how many aspects of form must be measured to characterize the genetic variation fully, that is, what is the dimensionality of the genetic variation. The fly-wing study used a particularly large number of families (800) and individuals (17,000), and demonstrated that at least 17 of the 20 possible dimensions had significant genetic variation (42). Further analysis using a restricted maximum-likelihood approach (43) revealed significant variation in all 20 directions. No similar analyses of the primate skull datasets have been done, and the sample size in each of those studies was substantially smaller. Nevertheless, the overall pattern of decrease in variation is quite similar and suggests that the genetic dimensionality in each of these species is also quite large. Full characterization of the genetics of these phenotypes cannot be undertaken from a small sample of measurements.

Studies of Selection Can Reveal That Only Some Combinations of Traits Are Important. Although the dimensionality of genetic variation is high, the selection-gradient analysis described above may well turn up some low-dimensional combination of traits that predicts an important outcome, either fitness or disease. Such a result could potentially indicate the narrow waist of a causal hourglass (22). The ability to attract a mate is an important component of fitness, and attractiveness can be readily assayed by directly observing matings, allowing selection gradient analyses to be performed. Blows and colleagues (44, 45) characterized the relative abundances of nine cuticular hydrocarbons (CHCs) in a population of *Drosophila serrata*, then compared the compositions of males to their success in competitive mating trials. The standardized selection gradient was extremely strong (change in relative fitness of 76% >1 SD change in the relative proportions of different CHCs), suggesting female preferences for higher proportions of several CHCs and antipathies toward others.

This logic of simplification extends to sets of phenotypes near a fitness optimum, where fitness decreases away from the population mean. In such cases, the important parameters are the quadratic coefficients in the γ matrix. These can be manipulated to allow interpretation, even in very high dimensional space (46). Brooks et al. (47) studied sexual selection on the calls of a cricket by synthesizing variation in five aspects of the call so that they could assess attractiveness of phenotypes not actually found in nature. They found that mate choice favored an intermediate optimum phenotype and that females paid strong attention to just two of the possible directions in P space. These cases suggest that a phenomic approach that begins with extensive and intensive measurements can then turn around to indicate some low-dimensional subset of these that is actually important in a particular context. The advantage of passing through a phenomic phase is that which combinations of traits are actually important is not apparent at the start.

Studies of Selection Can Suggest Past History of Trait Evolution. Both of the sexual-selection studies cited above went further and compared the pattern of selection on phenotypes to the pattern of genetic variation in the population studied. In each case, those aspects of the phenotype that were most strongly selected also had little genetic variation (45, 48). This pattern suggests a persistent mismatch between the phenotypes that females prefer and the ability of the males to produce them. The result is that the genetic consequences of female choice are very small; successful males are those that happen differ from normal in the favored direction, perhaps because they have experienced a favorable environment. Fig. 3*A* represents this relationship between population variation and selection schematically. The gray ellipse represents the expected phenotype (averaged over environmental factors) of the genotypes in the population. There is plenty of variation in the genetic basis of the phenotype, but it is oriented orthogonal to the direction of selection.

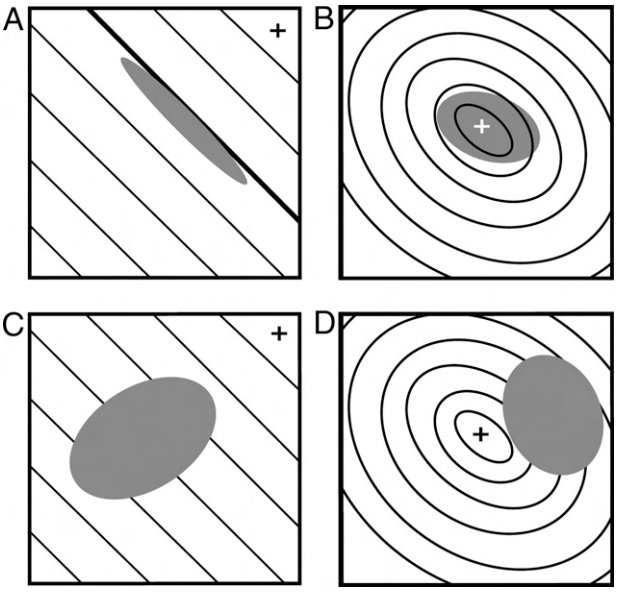


Fig. 3. Relationship between genetic variation in P space and the probability of disease or of fitness. Gray ellipses represent the distribution of genotypic variation in phenotype. + indicates the phenotype with the lowest probability of disease (or highest fitness), and the black lines are iso-lines that mark a particular level of probability of disease. (*A*) Constraints on the possible genotypes. The heavy diagonal line represents the constraint. A genotype above and to the right of this line cannot evolve. (*B*) Population mean is near the optimum, but mutation creates variation around that optimum. (*C*) A population in a novel environment, evolving toward a new optimum. (*D*) An aging population, where deterministic changes in phenotype caused by senescence drive the population away from the optimum.

This finding suggests that the comparison of the pattern of genetic variation for phenotype with the probability of disease could be very informative about the nature of the genetic variation in human disease. If the contours in Fig. 3 are now taken to represent disease probabilities instead of fitnesses (with + indicating a healthy phenotype with low disease probability), several possible scenarios might be found. Fig. 3*A* would represent an outcome with low λ, such as our inevitable demise caused by aging. Mutation-selection balance might produce a distribution like that in Fig. 3*B*, where the population generally matches the optimum state, but individuals with extreme phenotypes have increased probability of disease. Note that this disease may not be the same in each direction. The emerging hypothesis that many psychiatric disorders represent overexpression or underexpression of continuous personality traits provides a possible example, in which deviation in one direction leads to autism and deviation in another leads to schizophrenia (49). Diseases of civilization might lead to a pattern like that in Fig. 3*C*, where there is ample variation that has not yet been removed by a long history of natural selection. In the environment of evolutionary adaptedness, the selective pattern on the same variation might have been like that in Fig. 3 *A* or *B*. The shift from those patterns to that shown in Fig. 3*C* would be caused by genotype–environment interactions that alter the relative consequences of genetic variation. A second kind of alteration in the probability landscape might occur with age. In this case, Fig. 3*B* might represent the probability landscape during the reproductive years, and Fig. 3*D* might represent the landscape in the same population at an advanced age.

Genetic Variation Predicts Long-Term Evolution. There are many reasons to believe that the genetic variation that segregates

within a population might be irrelevant to long-term evolution. For example, most variation could be in the form of unconditionally deleterious mutations destined for quick elimination from the population, and conversely those rare mutations that will lead to major phenotypic changes might not be polymorphic for long. Contrary to this expectation, comparison of standing genetic variation in phenotype with patterns of among-species divergence suggests the relationship can be reasonably strong. Most work along these lines has relied on the relationship between the direction with the most genetic variance in P space, the first eigenvector of **G**, called $\mathbf{g}_{max}$, and the direction of evolutionary change (50). In most cases, the angle between these directions is less than expected under random models of change. Recent work has widened the scope of such comparisons to include all possible directions in multivariate P space, and here again those directions that show evolutionary change tend to have the most variation (51–53). The existence of relationships between variation and evolution suggests that the variation present in populations reflects deep conserved properties of the G-P map in ways that are not fully understood.

Phenomics: What Needs To Be Done

The foremost reason that G space is the favored locale for G-P studies is clear: "Collecting phenotypic data . . . is expensive and time consuming . . . " (54). Fifty years ago few could have imagined how our ability to obtain molecular data would increase; 20 years ago few could have imagined the scale at which we can now collect genomic information; 10 years ago few anticipated that genome-scale data could become as cheap as it now is. A key to this set of transformations was the vision of the Human Genome Project, which brought intellectual, technical, and financial resources to bear on genomes. Now is the time for a phenome project bringing the same kinds of gains in throughput and economic efficiency to the study of the phenotype.

Many biologists share my enthusiasm for the prospects of phenomics. There are increasing numbers of self-described phenome projects that should be wholeheartedly supported. The most useful of these take advantage of species where differentiated genotypes already exist as a scaffold onto which phenotype information can be added (37–38, 55). Inspection of the details of these projects, however, reveals that they are makeshift, shoestring operations compared with the magnitude of the challenges. We are pursuing phenomics as a piecemeal, small-science endeavor.

The need for a bigger-science approach is most apparent in the development of high-throughput approaches to phenotyping. To take one example, imaging is an extremely promising source of phenotypic data. The analysis of images should be generalizable across many different organisms and many different sorts of phenotypes (morphology of course, but also flows, spatial locations of metabolites, etc.). To maximize throughput, one would obviously optimize hardware for rapid, repeatable imaging, but also optimize specimen handling, automate phenotyping in software, and solve database issues to allow the handling of the massive amount of data that would result, among other challenges. The efforts of biologists who exploit imaging for phenotyping always far short of these ideals, however. Biologists use a huge variety of different, often ad hoc techniques for dealing with such data. In many cases, automation is restricted to use of a computer mouse. Sophisticated approaches are often applied to reduce the complexity of the phenotype measured to just a few dimensions, rather than to acquire intensive phenotypic data. My own approach to *Drosophila* wing measurement (34) reduces handling time of a live specimen to about a minute for all operations, but could readily be improved in various ways. For example, the low resolution and depth of field in the images prevents us from characterizing the cells and hairs clearly visible on the wing; the software we depend on was written to recover the locations, but not the thicknesses of veins. Because we already have an immobilized specimen, why not characterize body parts other than the wing?

The fundamental problem for phenomics is that the need for expertise is truly transdisciplinary (33). We need engineers, computer scientists, mathematicians, and statisticians as much as all flavors of expertise in biology (56). The time for the Human Genome Project did not arrive until fast and inexpensive methods were developed. Coordinated large-scale efforts to develop such approaches are what is currently missing from phenome efforts. As in the case of images, general approaches to phenotyping applicable across many organisms are surely possible for groups with the right expertise.

Therefore, although biologists continue valuable piecemeal efforts toward phenomics, we need large-scale efforts with the following aims: (*i*) further development of robust, general high-throughput phenotyping techniques; (*ii*) combined sequencing and phenotyping efforts that expand from the handful of genotypically controlled model systems, such as mice, to encompass natural population variation; and (*iii*) further development of analytical approaches that can use high-dimensional genotypic, endo-phenotypic, and end-phenotypic data to generate well-supported hypotheses for further testing.

Short of the ideal project outlined above, humans are clearly the one outbreeding species where the prospects for informative phenomics are the greatest. We have the peculiar tendency to measure our own species obsessively; the biomedical community is the one best positioned to provide the most complete phenomic data. The ultimate reward is to understand the G-P maps needed to turn biology and medicine from descriptive to predictive sciences.

We did not begin to study genomes because we care about genotypes; we study genomes because we care about phenotypes, the health and well-being of humans and the diversity of life on Earth. Now is the time to begin to take the study of the phenotype as seriously as we take the study of the genotype. We must number, locate, and measure even the hairs of our heads, the details of the phenotype, so that we can understand which of those details matter.

ACKNOWLEDGMENTS. I thank the organizers Randy Nesse and Raju Govindaraju for the invitation to participate in the symposium, Charles Roseman and Jim Cheverud for sharing unpublished data, and Stevan J. Arnold and an anonymous reviewer for detailed comments. This work was supported by National Science Foundation Grants DEB-0344417 and DEB-0129219 and the National Institutes of Health through National Institutes of Health Roadmap for Medical Research Grant U54 RR021813.

1. Lewontin RC (1974) *The Genetic Basis of Evolutionary Change* (Columbia Univ Press, New York).
2. Waddington CH (1942) The epigenotype. *Endeavor* 1:18–20.
3. Weiss KM (2008) Tilting at quixotic trait loci (QTL): An evolutionary perspective on genetic causation. *Genetics* 179:1741–1756.
4. Goldstein DB (2009) Common genetic variation and human traits. *N Engl J Med* 360:1696–1698.
5. Davey Smith G, Ebrahim S (2003) Mendelian randomization: Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 32:1–22.
6. Blangero J (2004) Localization and identification of human quantitative trait loci: King Harvest has surely come. *Curr Opin Genet Dev* 14:233–240.
7. McKusick-Nathans Institute of Genetic Medicine (2009) OMIM: Online Mendelian Inheritance in Man. Available at www.ncbi.nlm.nih.gov/sites/entrez?db=omim. Accessed June 30, 2009.
8. Bodmer W, Bonilla C (2008) Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 40:695–701.
9. Kung H-C, Hoyert DL, Xu J, Murphy SL (2008) Deaths: Final data for 2005. *Nat Vital Stat Rep* 56:1–121.
10. Goldgar DE, Easton DF, Cannonalbright LA, Skolnick MH (1994) Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Nat Cancer Inst* 86:1600–1608.
11. Pohjola-Sintonen S, Rissanen A, Liskola P, Luomanmaki K (1998) Family history as a risk factor of coronary heart disease in patients under 60 years of age. *Eur Heart J* 19:235–239.

12. Wellcome Trust Case Control Consortium (2007) Genomewide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–678.
13. Lupski JR (2007) Genomic rearrangements and sporadic disease. *Nat Genet* 39:S43–S47.
14. Stefansson H, et al. (2008) Large recurrent microdeletions associated with schizophrenia. *Nature* 455:232–236.
15. Tenesa A, Dunlop MG (2009) New insights into the aetiology of colorectal cancer from genomewide association studies. *Nat Rev Genet* 10:353–358.
16. Guerra S, Martinez FD (2008) Asthma genetics: From linear to multifactorial approaches. *Annu Rev Med* 59:327–341.
17. Weedon MN, et al. (2008) Genomewide association analysis identifies 20 loci that influence adult height. *Nat Genet* 40:575–583.
18. Lettre G, et al. (2008) Identification of 10 loci associated with height highlights new biological pathways in human growth. *Nat Genet* 40:584–591.
19. Gudbjartsson DF, et al. (2008) Many sequence variants affecting diversity of adult human height. *Nat Genet* 40:609–615.
20. Hirschhorn JN (2009) Genomewide association studies: Illuminating biologic pathways. *N Engl J Med* 360:1699–1701.
21. Reed D, Lawler M, Tordoff M (2008) Reduced body weight is a common effect of gene knockout in mice. *BMC Genetics* 9:4.
22. Buchanan AV, Weiss KM, Fullerton SM (2006) Dissecting complex disease: The quest for the philosopher's stone? *Int J Epidemiol* 35:562–571.
23. Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics* (Addison Wesley Longman, Essex, UK) 4th Ed.
24. Janzen FJ, Stern HS (1998) Logistic regression for empirical studies of multivariate selection. *Evolution (Lawrence, Kans)* 52:1564–1571.
25. Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution (Lawrence, Kans)* 37:1210–1226.
26. Mitchell-Olds T, Shaw RG (1987) Regression analysis of natural selection: Statistical inference and biological interpretation. *Evolution (Lawrence, Kans)* 41:1149–1161.
27. Weng G, Bhalla US, Iyengar R (1999) Complexity in biological signaling systems. *Science* 284:92–96.
28. Bassingthwaighte JB (2000) Strategies for the physiome project. *Ann Biomed Eng* 28:1043–1058.
29. Paigen K, Eppig JT (2000) A mouse phenome project. *Mamm Genome* 11:715–717.
30. Houle D (2001) in *The Character Concept in Evolutionary Biology*, ed Wagner GP (Academic, New York), pp 109–140.
31. Freimer N, Sabatti C (2003) The human phenome project. *Nat Genet* 34:15–21.
32. Oti M, Huynen MA, Brunner HG (2008) Phenome connections. *Trends Genet* 24:103–106.
33. Bilder RM, et al. (2009) Phenomics: The systematic study of phenotypes on a genome-widescale. *Neuroscience*, in press.
34. Houle D, Mezey J, Galpern P, Carter A (2003) Automated measurement of *Drosophila* wings. *BMC Evol Biol* 3:25.
35. Ohya Y, et al. (2005) High-dimensional and large-scale phenotyping of yeast mutants. *Proc Natl Acad Sc USA* 102:19015–19020.
36. Vizeacoumar FJ, Chong Y, Boone C, Andrews BJ (2009) A picture is worth a thousand words: Genomics to phenomics in the yeast *Saccharomyces cerevisiae*. *FEBS Lett* 583:1656–1661.
37. Beckers J, Wurst W, de Angelis MH (2009) Toward better mouse models: Enhanced genotypes, systemic phenotyping, and envirotype modelling. *Nat Rev Genet* 10:371–380.
38. Grubb SC, Maddatu TP, Bult CJ, Bogue MA (2009) Mouse phenome database. *Nucleic Acids Res* 37:D720–D730.
39. Kristensen E, Parsons TE, Hallgrimsson B, Boyd SK (2008) A novel 3D image-based morphological method for phenotypic analysis. *IEEE Trans Biomed Eng* 55:2826–2831.
40. Zelditch ML, Swiderski DL, Sheets HD, Fink WL (2004) *Geometric Morphometrics for Biologists: A Primer* (Elsevier, Amsterdam).
41. Mosimann JE (1970) Size allometry: Size and shape variables with characterizations of the lognormal and generalized gamma distributions. *J Am Stat Assoc* 65:930–945.
42. Mezey JG, Houle D (2005) The dimensionality of genetic variation for wing shape in *Drosophila melanogaster*. *Evolution (Lawrence, Kans)* 59:1027–1038.
43. Kirkpatrick M, Meyer K (2004) Direct estimation of genetic principal components: Simplified analysis of complex phenotypes. *Genetics* 168:2295–2306.
44. Hine E, Lachish S, Higgie M, Blows MW (2002) Positive genetic correlation between female preference and offspring fitness. *Proc R Soc London Ser B* 269:2215–2219.
45. Blows MW, Chenoweth SF, Hine E (2004) Orientation of the genetic variance–covariance matrix and the fitness surface for multiple male sexually selected traits. *Am Nat* 163:329–340.
46. Phillips PC, Arnold SJ (1989) Visualizing multivariate selection. *Evolution (Lawrence, Kans)* 43:1209–1222.
47. Brooks R, et al. (2005) Experimental evidence for multivariate stabilizing sexual selection. *Evolution (Lawrence, Kans)* 59:871–880.
48. Hunt J, Blows MW, Zajitschek F, Jennionos MD, Brooks R (2007) Reconciling strong stabilizing selection with the maintenance of genetic variation in a natural population of black field crickets (*Teleogryllus commodus*). *Genetics* 177:875–880.
49. Crespi B, Summers K, Dorus S (2009) Genomic sister-disorders of neurodevelopment: An evolutionary approach. *Evol Appl* 2:81–100.
50. Schluter D (1996) Adaptive radiation along genetic lines of least resistance. *Evolution (Lawrence, Kans)* 50:1766–1774.
51. Hansen TF, Armbruster WS, Carlson ML, Pélabon C (2003) Evolvability and genetic constraint in *Dalechampia* blossoms: Genetic correlations and conditional evolvability. *J Exp Zool B Mol Dev Evol* 296:23–39.
52. Hansen TF, Houle D (2008) Measuring and comparing evolvability and constraint in multivariate characters. *J Evol Biol* 21:1201–1219.
53. Hunt G (2007) Evolutionary divergence in directions of high phenotypic variance in the ostracode genus *Poseidonamicus*. *Evolution (Lawrence, Kans)* 61:1560–1576.
54. Hancock AM, Di Rienzo A (2008) Detecting the genetic signature of natural selection in human populations: Models, methods, and data. *Annu Rev Anthropol* 37:197–217.
55. Canine Phenome Project (2009) The Canine Phenome Project. Available at www.caninephenome.org Accessed July 16, 2009.
56. Anonymous (2007) Geneticist seeks engineer; must like flies and worms. *Nat Methods* 4:463.
57. Willmore KE, Roseman CC, Rogers J, Richtsmeier JT, Cheverud JM (2009) Genetic variation in baboon craniofacial sexual dimorphism. *Evolution (Lawrence, Kans)* 63:799–806.
58. Cheverud JM (1995) Morphological integration in the saddle-backed tamarin (*Saguinus fusicollis*) cranium. *Am Nat* 145:63–89.
59. Cheverud JM (1996) Quantitative genetic analysis of cranial morphology in the cotton-top (*Saguinus oedipus*) and saddle-back (*S. fuscicollis*) tamarins. *J Evol Biol* 9:5–42.