Am. J. Hum. Genet. 64:18–23, 1999

MOLECULAR EVOLUTION '99
When Less Is More: Gene Loss as an Engine of Evolutionary Change

Maynard V. Olson

Departments of Medicine (Division of Medical Genetics) and Genetics, University of Washington, Seattle

Evolutionary change results from differences in the reproductive success of individuals with different genotypes. The downside of this process is easy to grasp: selection constantly purges deleterious mutations from the gene pool. However, we know remarkably little about evolution's upside—that is, about the types of mutations that commonly lead to increased fitness. To understand the biology of natural populations—including, most notably, that of the human—we need testable ideas about the types of mutations that evolution is likely to have favored in the recent past. Here I explore one such idea, the proposal that loss of gene function may represent a common evolutionary response of populations undergoing a shift in environment and, consequently, a change in the pattern of selective pressures. If, as I suggest, less is often more, where gene function is concerned, adaptive loss of function may occur frequently and may spread rapidly through small populations.

Because mutations that lead to loss of function are numerous, this class of change (if adaptive) is the most likely outcome when a novel selection acts on a population. Loss-of-function mutations will occur far more often than will a shift in the target specificity of a protein or in the patterns of spatial or temporal regulation of a gene—and certainly will occur more often than a gene will acquire a new regulatory system. Once its function is lost—unless the lesion involves a complete deletion of the gene—the mutated gene will persist in the genome and may be available for reversion if the selective environment shifts once more. Evidence supporting the plausibility of the "less-is-more" hypothesis comes from both mammalian and microbial genetics. Here, on the basis of diverse examples drawn from these organisms, I propose the testable view that gene loss is a major motif of molecular evolution.

Well-known human examples of conditionally advantageous mutations—those that improve fitness in particular environments—include a number of biallelic and multiallelic systems in which heterozygotes enjoy a conditional advantage. In these cases, alleles that are clearly maladaptive when present in homozygous form are nevertheless maintained at high frequency in some populations. For example, enteric disease and iron-deficient diets, respectively, have been proposed as selective pressures that may confer a heterozygote advantage on mutations that, when homozygous, cause cystic fibrosis and hemochromatosis (Gabriel et al. 1994; Crawford et al. 1995). Similarly, hemoglobinopathies are common in many human populations because of a slight heterozygote advantage in high-malaria environments. Still, our knowledge of conditionally beneficial alleles in humans is biased by our reliance on homozygous disease states to bring the slight conditional benefits of heterozygous phenotypes to our attention. Moreover, genetic drift and founder effects are hard to exclude as explanations when mutant alleles are present at high frequencies in particular populations, and hypotheses about specific selective mechanisms are always difficult to prove. Work with model genetic systems in the laboratory circumvents some of these difficulties.

Selection for Year-Round Fecundity in Mice

To make these ideas more concrete, consider some observations about the reproductive behavior of wild and laboratory mice. Wild strains of *Mus musculus,* the species from which laboratory mice were derived, display a seasonal pattern of reproduction that is typical of animals living at nonequatorial latitudes. These wild strains display the same diurnal cycles of melatonin synthesis in the pineal gland that are present in nearly all mammals and that play a central role in the regulation of seasonal reproduction (Tamarkin et al. 1985). Given its evolutionary conservation, this mechanism for monitoring changes in the length of daylight hours and for adjusting reproductive behavior in response to them is likely to be the ancestral state. In contrast, laboratory mice, whose reproduction is uncoupled from seasonal

Received October 23, 1998; accepted for publication November 13, 1998; electronically published December 21, 1998.

Address for correspondence and reprints: Dr. Maynard V. Olson, Departments of Medicine (Division of Medical Genetics) and Genetics, Box 352145, University of Washington, Seattle, WA 98195. E-mail: mvo@u.washington.edu

This article represents the opinion of the author and has not been peer reviewed.

0002-9297/99/6401-0005$02.00

change, have no pineal-melatonin synthesis. The defect is due to the presence of recessive mutations in two genes, which almost certainly encode the two enzymes required for the conversion, in two steps, of seratonin to melatonin (Ebihara et al. 1986; Goto et al. 1994). A likely hypothesis is that these mutations accumulated through inadvertent selection for unregulated breeding, a highly desirable characteristic of domesticated mice. Similar selections are also likely to have occurred frequently in the wild when mice have populated environments where continuous breeding is at least transiently advantageous. Indeed, a wild mouse species that lives at equatorial latitudes and that reproduces continuously throughout the year—the Venezuelan cane mouse—has normal pineal-melatonin synthesis, but this species' reproductive behavior is uncoupled from this metabolic pathway (Bronson and Heideman 1992). Most likely, the cane mouse, under selection for high reproductive rates, has lost other genetic functions required for full operation of the pineal-melatonin system. Perhaps nonreproductive benefits of diurnal melatonin synthesis have preserved the "front end" of the system in cane mice, whereas this system was lost in laboratory mice. The different responses of the two species may relate to the heavy-handed selection applied during domestication or to more subtle biological or environmental issues. These examples illustrate how readily organisms can gain conditional benefits by discarding the functionality of widely conserved genes. Given that the overwhelming preponderance of mutations cause complete or partial loss of gene function, it seems likely that evolution has made frequent positive use of this most common class of genetic change.

Disposable Genes

If one is to argue for broad applicability of the less-is-more principle in evolution, several questions arise. First, how cavalierly can organisms dispense with genetic functions, without suffering seriously adverse effects? Second, if the selection is envisioned as acting on homozygotes rather than on heterozygotes, is it plausible that these genotypes arise often enough to play a major evolutionary role in diploid, sexually reproducing populations? Third, how can evolutionary processes generate novelty as readily as they apparently do, if selection is constantly stripping away hard-won genetic functions? Finally, how do inactivated genes contribute to the subsequent evolution of the organism?

With respect to the dispensability of genetic functions, the yeast *Saccharomyces cereviseae* is the best-studied model. This single-celled eukaryote possesses fully developed mitotic and meiotic cell-division cycles, synthesizes functional mitochondria, and differentiates along several alternative developmental pathways, but it maintains only 6,000 genes in a genome 0.5% the size of a mammalian genome (Broach et al. 1991). Despite this obviously streamlined genetic system, Goebl and Petes (1986) note that 85% of yeast genes can be ablated without an effect on haploid viability. Some viable knockout strains display deleterious phenotypes under particular circumstances, but such phenotypes are often subtle. We have no idea what fraction of the environments to which wild yeast are adapted have been successfully mimicked in the laboratory, but the data show that even an organism with a parsimonious genome can dispense with a surprisingly large fraction of its genetic functions while preserving its complete life cycle. Intriguingly, Goebl and Petes (1986) also mention that some of the strains in which genes had been disrupted grew better than "wild type" on a rich medium. Thus, although there are in nature undoubtedly many circumstances in which genes that are dispensable in the laboratory play critical roles, there may also be specialized ecological niches—analogous to growth on a rich medium—in which selection favors less than a complete repertoire of functional genes.

Propagation of Gene Loss in Human Populations

Because *Saccharomyces* readily forms fully homozygous diploids through self-mating in the wild, homozygosis of conditionally beneficial mutations—including gene loss—may be more efficient than the corresponding process in mammals. For our species, population size and mating structure have a large influence on the probability that an autosomal recessive loss-of-function mutation will ever have the opportunity to expand through selection, even if homozygotes enjoy major conditional benefits. The best estimates of past population sizes make it quite plausible that loss-of-function mutations could have contributed to human evolution even if their conditionally beneficial effects were fully recessive. The level of genetic variation in the current human population suggests that typical population sizes were ~10,000 individuals during most of the past several hundred thousand years, the interval during which population estimates can be made from sequence data on modern humans (Harpending et al. 1998). Since 10^{-4} is a typical frequency for recessive genetic diseases, it is likely that even a random-mating population of 10,000 humans would frequently contain individuals homozygous for loss-of-function mutations in any dispensable human gene (Vogel and Motulsky 1996). Fragmentation of populations into relatively inbred subgroups, such as almost certainly occurred during most of human evolution, further increases the likelihood that a recessive mutation will become homozygous within a few generations.

In the human, the best-studied examples of adaptive

gene loss involve pathogen resistance. One dramatic case involves the relationship between the Duffy-negative blood group and resistance to *Plasmodium vivax* (Tournamille et al. 1995), a malaria parasite that causes less-severe disease than is caused by the better-known *P. falciparum*. A similar example involves resistance to AIDS in individuals homozygous for a null mutation in the *CCR5* gene (Libert et al. 1998; Stephens et al. 1998). Both of these examples involve genetic loss of different chemokine receptors that are essential for entry of the pathogens into target cells. In the Duffy-negative case, the recessive mutation is in a promoter element required for expression of the chemokine receptor DARC in erythroid lineages, whereas in resistance to AIDS a *CCR5* coding-region deletion leads to loss of function in all tissues, even though the functionally important change is apparently specific to macrophages. In neither case has any deleterious phenotype been detected in individuals homozygous for the mutations. The allele frequency of the Duffy-negative mutation is 100% in some regions of western Africa, whereas that for the *CCR5* deletion is >15% in certain areas of northern Europe. Although *vivax* malaria is a plausible source of the selective pressure that led to the prevalence of Duffy-negative genotypes in western Africa, the analogous selective pressure for the *CCR5* deletion allele remains unclear, since AIDS is thought to be a modern human disease.

Inevitable historical uncertainties about past selective pressures should not cloud the basic message of these examples. Both involve conditionally beneficial loss-of-function mutations that have become common in large human populations, almost certainly in response to strong selection. Furthermore, they have done so without the concomitant cost of genetic disease. Presumably, averaged over all human experience, DARC expression in erythrocytes and CCR5 expression in many cell types improve fitness—otherwise, these highly conserved functions would have been lost long ago. However, the conditional benefits of losing the functions in particular environments appear to have been overwhelming, whereas the costs of the losses appear to have been slight. Similar considerations are likely to apply to the better-known O blood group, an instance in which recurrent loss-of-function mutations have expanded in primate populations, albeit in response to unknown selective forces (O'hUigin et al. 1997).

Given the difficulty of understanding the evolutionary significance of common human genotypes, we should be slow to conclude that conditionally beneficial loss-of-function mutations are only important for particular classes of genes, such as pathogen-resistance genes. Other candidates, whose adaptive loss might prove to be examples of the less-is-more model, include genes that mediate our responses to particular foods. *Aldolase B* mutations, for instance, are surprisingly common, despite the fact that homozygous deficiency leads to fructose intolerance (Ali et al. 1998). Similarly, *glutathione transferase* null alleles persist at high frequencies, despite their statistical association with increased levels of some cancers (Board et al. 1990). The evolutionary benefits of these alleles remain speculative, but it must be emphasized that virtually all aspects of the human environment have undergone repeated upheavals during the past 100,000 years, the period during which humans have populated the globe. Change has been particularly dramatic during the past 15,000 years, as our species has adapted to the end of the last ice age—an event that is now thought to have been dramatically sudden. The ensuing spread of agriculture not only revolutionized human diets and pathogen exposures but also profoundly influenced the physiological stresses experienced by humans, as well as their behavioral and social adaptations (Diamond 1997). The possibility that the selective shedding of genetic functions played a major role in reshaping the human gene pool during this period deserves consideration as we seek to understand the functional significance of diversity in many classes of human genes.

Adaptive Reversions

In the long run, evolution must innovate in more fundamental ways than simply by giving up what it had once created. However, if one accepts the concept that, in response to short-term selections, loss-of-function mutations frequently expand rapidly in populations, their possible role in long-term innovation also deserves consideration. Although genes that are never needed—and, in particular, those that are sometimes best done without—will eventually erode by mutation, many genes are likely to have functions that sometimes increase and sometimes decrease reproductive fitness. Particularly when population sizes are appropriate and environmental fluctuations occur over tens to thousands of generations—a situation that may describe many environmental changes that have occurred during human evolution—cycles of loss of function followed by reversion may be important modes of molecular evolution. At one extreme, when environmental variation is sufficiently rapid that individuals typically experience frequent environmental fluctuations within their lifetimes, gene regulation is likely to evolve as a means of accommodating the on-again, off-again benefits of particular genetic functions. At the other extreme, a species confronted by a singular environmental shift that suddenly requires a genetic function that was shed by mutation in the distant past will not have a reservoir of "near-functional" alleles in the population within which reversion events can occur at a meaningful frequency. However, at intermediate time scales, reversion of pre-

viously shed genes may be a common response to the recurrence of selective forces that act intermittently.

Bacteria provide the best-described model for this process. A surprising finding of the early 1980s was the discovery that many bacteria harbor cryptic operons (i.e., operons that have been inactivated by mutation but that can still readily revert to functional status; Hall et al. 1983). For example, most natural isolates of *E. coli* carry cryptic genes encoding enzymes required for the utilization of β-glucoside sugars such as cellobiose. Under selection for growth on these sugars, revertants are readily obtained in the laboratory. As explained by Hall and Xu (1992, p. 688), "The persistence of cryptic genes in the face of mutational pressure is an interesting puzzle for population biologists, and our current model is that they are retained by alternately selecting for loss and regain of function in different environments." There is no reason to imagine that this evolutionary process is limited to microorganisms. Whenever strong selection for the conditional benefits of a loss-of-function mutation leads to rapid population expansion, a large pool of mutant genes becomes available within which revertants can arise. When conditions change, these revertants, which would be expected to have dominant phenotypes when arising in a homozygous-null background, may be favored by selection. Furthermore, cycles of mutation and reversion have the potential for molecular experiments that are bolder than those typical of single-step changes. This potential is most obvious in the case of frameshift mutations, which are common both in cryptic bacterial genes and in mutant genes associated with human genetic diseases. For example, of the 700 distinct mutant alleles that have been reported in the human genes *BRCA1* and *BRCA2,* 42% of the *BRCA1* alleles and 54% of the *BRCA2* alleles are frameshift mutations (Brody and Biesecker 1998). Frameshift mutations often revert by the occurrence of a second, complementary frameshift separated from the first by many codons. In this way, two mutations can change a whole string of adjacent amino acids, leading to rapid evolution of protein sequence.

A spectacular example of the potential for rapid evolution in genes that appear to be alternately under selection for function and loss of function are the genes for bacterial restriction endonucleases. Among the large number of sequenced restriction-endonuclease genes, there are almost never recognizable sequence motifs, even when the encoded enzymes have nearly identical properties (Wilson and Murray 1991; Bickle and Kruger 1993). Indeed, some pairs of endonucleases whose amino acid sequences are too dissimilar to align, such as *Bam*H1 and *Eco*RI, have nearly identical three-dimensional structures (Aggarwal 1995). Although there is a long-standing debate as to whether the lack of sequence similarity among endonuclease genes of similar structures and functions reflects convergent or divergent evolution (Jeltsch et al. 1995), it is plausible that these genes diverge rapidly because the lineage of a particular gene traverses frequent cycles of mutation and reversion.

The rationale for this model is as follows. Genes for a restriction endonuclease and a protective methylase are invariably tightly linked and are thought to spread to new genomic environments largely through horizontal transfer (Jeltsch and Pingoud 1996). Transfer of a restriction-endonuclease gene into a naive host is a precarious transaction, since the cleavage of the new host's unprotected sites risks death of the cell. However, once the methylase gene is expressed, the host genome is protected from cleavage by the endonuclease, and the host cell stands to benefit from the ability of the endonuclease to destroy the unprotected DNA of incoming viruses. Perhaps even more important, reversion of a mutated endonuclease gene benefits the gene itself, because loss of a gene encoding a functional restriction-methylation system is frequently lethal to the host cell (i.e., these genes fulfill the basic criteria for "selfish DNA"; Naito et al. 1995). Although regulatory mechanisms may play a role in controlling the relative phenotypic lags associated with expression of endonuclease and methylase genes after transfer into a new host (Karyagina et al. 1997), it remains plausible that, at the transfer step, a mutated endonuclease gene will often be at a great selective advantage over a functional one. In contrast, once the gene is established in a new host that expresses an active methylase, reversion of the mutation would be favored. This cycle would be expected to lead, as is observed, to rapid divergent evolution of endonuclease genes, whereas methylase genes remain relatively well conserved.

Traces of Lost Genes

The idea that genetic loss may be an important engine of evolutionary change is counterintuitive. We like to think that organisms achieve better fitness by having "better" genes, not broken ones. Over the broad sweep of evolutionary time, this principle must be true, but loss and regain of gene function may be common over shorter stretches of a species' history; if so, this pattern should be evident from genomic comparisons between populations and among related species. Indeed, the less-is-more theory can be tested more readily than many evolutionary hypotheses. It predicts that study populations acquired on the basis of conditionally beneficial phenotypes—in the case of modern urban human beings, low blood pressure on a high-salt diet, failure to gain weight on a rich diet, lack of anxiety under stress, or longevity—will frequently be enriched for individuals homozygous for null alleles in genes relevant to the phenotype. The same logic applies to phenotypes that are

presently deleterious but that conveyed conditional benefits in the past (e.g., an extreme tendency to store excess calories as body fat). Because they present with disease phenotypes, this class of conditionally beneficial genotypes—Neel's (1962) "thrifty genes"—fit more comfortably into traditional views of human genetics than do genotypes that continue to convey conditional benefits in distinctively modern environments. Obviously, in testing the less-is-more hypothesis, we will have the best success with phenotypes whose underlying biochemistry is relatively well understood. However, advances in basic biology—as well as in the technical capability to scan large numbers of genes for mutations—may be expected to expand rapidly the range of phenotypes that are accessible to this type of analysis in the years ahead. At the DNA level, loss-of-function mutations are often recognizable against a high background of neutral mutations, as is the case for nonsense mutations, frameshift mutations, and obvious splicing defects. Expression surveys at the mRNA or protein level offer alternative screening methods. In some instances, the pattern of variation in a gene (e.g., ratios of synonymous to nonsynonymous amino acid substitutions) may provide clues that the gene has been under alternating selection for function and loss thereof.

From a theoretical standpoint, the less-is-more model suggests a reformulation of our concept of "wild type." Although a cherished term in genetics, "wild type" has never had a particularly clear definition. Operationally, most geneticists think of a wild-type allele as one that has nothing obvious wrong with it. By extension, a wild-type organism is one composed entirely of wild-type genes. Perhaps such an organism, in the unlikely event that it could be found or constructed, would be poorly suited to all the environments in which its species normally lives. A contemporary, anthropogenic analogy would be to a "fully loaded" sports-utility vehicle. Such a vehicle typically would not be manufactured and, indeed, would be poorly suited to any particular use. There is, of course, a core set of functions—spark plugs, fuel pumps, and so forth—that are shared by all operable units and that vary little from vehicle to vehicle. Other features, such as seats and tires, are present and functional in all units but vary greatly from one to another. However, there is also a long list of features—four-wheel drive, catalytic converters, trailer attachments, security systems, cruise control—whose outright desirability is highly dependent on the vehicle's proposed use. The more complex the vehicle's design and the larger the variety of uses to which it might be put, the more the divergence between the "fully loaded" abstraction and the configuration of any actual unit.

Particularly in human and agricultural genetics, where intensive molecular analysis is now being brought to bear on diverse natural populations, there are likely to be many opportunities to test the hypothesis that loss-of-function mutations play a major role in evolution. Although statistical-genetic methods offer a more agnostic approach to the correlation of genotypes to complex phenotypes, it is unlikely that these methods alone will lead to a comprehensive view of how the genomes of contemporary organisms have been shaped by past selective pressures. Broad insights into these processes will depend on new ways of thinking about molecular evolution and on the ability to couple evolutionary theory to our growing knowledge of cellular and organismal function. Perhaps the less-is-more principle will provide this coupling for a broad range of evolutionary adaptations.

Acknowledgments

I would like to acknowledge discussions with a number of colleagues—particularly Richard Roberts, Arno Motulsky, and Joseph Takahashi—which were important sources of examples used in this article.

References

Aggarwal AK (1995) Structure and function of restriction endonucleases. Curr Opin Struct Biol 5:11–19

Ali M, Rellos P, Cox TM (1998) Hereditary fructose intolerance. J Med Genet 35:353–365

Bickle TA, Kruger DH (1993) Biology of DNA restriction. Microbiol Rev 57:434–450

Board P, Coggan M, Johnston P, Ross V, Suzuki T, Webb G (1990) Genetic heterogeneity of the human glutathione transferases: a complex of gene families. Pharmacol Ther 48:357–369

Broach JR, Pringle JR, Jones EW (eds) (1991) The molecular and cellular biology of the yeast *Saccharomyces*. Vols 1–3. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Brody LC, Biesecker BB (1998) Breast cancer susceptibility genes BRCA1 and BRCA2. Medicine 77:208–226

Bronson FH, Heideman PD (1992) Lack of reproductive photoresponsiveness and correlative failure to respond to melatonin in a tropical rodent, the cane mouse. Biol Reprod 46: 246–250

Crawford DH, Powell LW, Leggett BA, Francis JS, Fletcher LM, Webb SI, Halliday JW, et al (1995) Evidence that the ancestral haplotype in Australian hemochromatosis patients may be associated with a common mutation in the gene. Am J Hum Genet 57:362–367

Diamond JM (1997) Guns, germs, and steel: the fates of human societies. WW Norton, New York

Ebihara S, Marks T, Hudson DJ, Menaker M (1986) Genetic control of melatonin synthesis in the pineal gland of the mouse. Science 231:491–493

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

Goebl MG, Petes TD (1986) Most of the yeast genomic sequences are not essential for cell growth and division. Cell 26:983–992

Goto M, Oshima I, Hasegawa M, Ebihara S (1994) The locus controlling pineal serotonin N-acetyltransferase activity (Nat-2) is located on mouse chromosome 11. Brain Res Mol Brain Res 21:249–354

Hall BG, Xu L (1992) Nucleotide sequence, function, activation, and evolution of the cryptic asc operon of *Escherichia coli* K12. Mol Biol Evol 9:688–706

Hall BG, Yokoyama S, Calhoun DH (1983) Role of cryptic genes in microbial evolution. Mol Biol Evol 1:109–124

Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, Sherry ST (1998) Genetic traces of ancient demography. Proc Natl Acad Sci USA 95:1961–1967

Jeltsch A, Kroeger M, Pingoud A (1995) Evidence for an evolutionary relationship among type-II restriction endonucleases. Gene 160:7–16

Jeltsch A, Pingoud A (1996) Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. J Mol Evol 42:91–96

Karyagina A, Shilov I, Tashlitskii V, Khodoun M, Vasil'ev S, Lau PCK, Nicholskaya I (1997) Specific binding of SsoII DNA methyltransferase to its promoter region provides the regulation of SsoII restriction-modification gene expression. Nucleic Acids Res 25:2114–2120

Libert F, Cochaux P, Beckman G, Samson M, Aksenova M, Cao A, Czeizel A, et al (1998) The Δccr5 mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in northeastern Europe. Hum Mol Genet 7:399–406

Naito T, Kusano K, Kobayashi I (1995) Selfish behavior of restriction-modification systems. Science 267:897–899

Neel JV (1962) Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? Am J Hum Genet 14: 353–362

O'hUigin C, Sato A, Klein J (1997) Evidence for convergent evolution of A and B blood group antigens in primates. Hum Genet 101:141–148

Stephens JC, Reich DE, Goldstein DE, Shin HD, Smith MW, Carrington M, Winkler C, et al (1998) Dating the origin of the *CCR5-Δ32* AIDS-resistance allele by the coalescence of haplotypes. Am J Hum Genet 62:1507–1515

Tamarkin L, Baird CJ, Almeida OFX (1985) Melatonin: a coordinating signal for mammalian reproduction? Science 227:714–720

Tournamille C, Colin Y, Cartron JP, Le-Van-Kim C (1995) Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. Nat Genet 10:224–228

Vogel F, Motulsky AG (1996) Human genetics: problems and approaches, 3d ed. Springer-Verlag, New York

Wilson GG, Murray NE (1991) Restriction and modification systems. Annu Rev Genet 25:585–627