What tangled web: barriers to rampant horizontal gene transfer

Kurland, Charles

Published in:
BioEssays

DOI:
10.1002/bies.20258

2005

What tangled web: barriers to rampant horizontal gene transfer

Charles G. Kurland

Summary
Dawkins in his *The Selfish Gene*[1] quite aptly applies the term ''selfish'' to parasitic repetitive DNA sequences endemic to eukaryotic genomes, especially vertebrates. Doolittle and Sapienza[2] as well as Orgel and Crick[3] enlivened this notion of selfish DNA with the identification of such repetitive sequences as remnants of mobile elements such as transposons. In addition, Orgel and Crick[3] associated parasitic DNA with a potential to outgrow their host genomes by propagating both vertically via conventional genome replication as well as infectiously by horizontal gene transfer (HGT) to other genomes. Still later, Doolittle[4] speculated that unchecked HGT between unrelated genomes so complicates phylogeny that the conventional representation of a tree of life would have to be replaced by a thicket or a web of life.[4] In contrast, considerable data now show that reconstructions based on whole genome sequences are consistent with the conventional ''tree of life''.[5–10] Here, we identify natural barriers that protect modern genome populations from the inroads of rampant HGT.
***BioEssays* 27:741–747, 2005.**

Introduction

The standard model for the inheritance of genes in modern organisms rests on the Darwinian conception of selection of the fittest phenotype and the linear descent of the corresponding selected sequences in tree-like phylogenetic representations. This is commonly referred to as vertical inheritance. In contrast, both the transfer of sequences to non-canonical sites within a genome or the irregular exchange of sequences between different related or unrelated genomes was referred to as horizontal gene transfer (HGT) by Orgel and Crick.[3] Here, the term HGT is applied likewise to all such transfers to emphasize the common mechanisms of sequence exchange within and between genomes.

Molecular genetics arose from studies of bacterial inheritance that exploited different modes of HGT such as those mediated by viruses and other mobile elements as well as by transforming DNA. Nevertheless, it took time to realize that HGT might also be relevant to genome evolution, particularly among the vertebrates. This line of thought first surfaced with the conjecture that seemingly nonfunctional repetitive sequences common to vertebrate genomes could be viewed as selfish parasites.[1] This important idea is mentioned only in passing in Richard Dawkins' *The Selfish Gene*.[1] However, it was eventually embraced by Doolittle and Sapienza[2] as well as by Orgel and Crick,[3] who recognized the parallels between the highly repetitive sequences in eukaryote genomes, and the terminal repeat sequences of bacterial mobile elements. Thus, repetitive sequences in eukaryote genomes could be viewed as the signatures of parasites such as viruses and transposons.[2,3]

Orgel and Crick[3] noted that such parasitic sequences would have two modes of propagation. One mode is by replication of the transgenic genome and the other is by infective transfer to other genomes as for example through transposons or viruses. Accordingly, these parasitic DNA sequences can in principle propagate at a faster rate than their host genomes. That is to say, they are infectious and, if unopposed, they tend to expand the size of genomes without limit. On the other hand, Orgel and Crick[3] also recognized that such parasitic sequences could be disruptive to the host genome, at the very least by adding to its replication costs. Nevertheless, if their infective propensities exceeded the loss of fitness to the host, the parasitic sequences would be able to spread faster than the host genome population replicates.

Still later, another dimension of HGT was introduced for alien sequences that encode proteins and RNA. Thus, alien HGT was invoked to account for the phylogenetic anomalies detected when the first large-scale sequence data was accessed. Indeed, it was suggested that alien HGT was so frequent that it necessitated a paradigm shift from the standard model to one in which rampant HGT becomes the ''essence of the phylogenetic process''.[4]

In fact, only a relatively small number of phylogenetic anomalies had been forwarded as evidence for rampant HGT.[4] For example, in a more recent discussion, thirty anomalous sequences identified in nearly as many genomes were presented as significant examples of HGT.[11] Since each of the genomes in this cohort contains thousands of coding sequences, it is evident that the anomalous sequences are far too few to validate the rampant HGT paradigm. Still, the idea of

Department of Microbial Ecology, University of Lund, Lund, Sweden.
E-mail: kurland@telia.com
DOI 10.1002/bies.20258
Published online in Wiley InterScience (www.interscience.wiley.com).

the HGT paradigm shift became a seemingly permanent fixture in the literature by the end of the 1990s.

The grip of the rampant HGT paradigm could be loosened by enlarging the search for phylogenetic anomalies from anecdotal samplings to systematic studies of whole genome sequences. It then became apparent that the previously highlighted anomalies were not the tip of an HGT iceberg. Instead, most of the detectable phylogenetic anomalies seem to arise from paralogy, deletion, biased mutation rates, etc and not from HGT.[5–10] Whatever their origins, the phylogenetic anomalies in whole genome sequences are not sufficiently numerous to jumble reconstructions based on these sequences. Rather, there is remarkably good accord between, on the one hand, phylogenetic reconstructions based on ribosomal RNA sequences and on the other, reconstructions based on the pooled protein sequences from as many as 55 fully sequenced genomes.[5–10] If the rampant HGT paradigm were relevant to modern genomes, reconstructions with the pooled protein sequences would not have been resolvable as a single tree that is coherent with reconstructions based on the corresponding ribosomal RNA sequences.

Here, rather than recount the criticisms of the rampant HGT paradigm,[5,12–14,16] the natural barriers that oppose the invasiveness of transferred sequences are discussed briefly. These include the destruction of transferred sequences by random mutations, the loss of fitness by transgenic cells and host modulation of the infectivity of transferred sequences. Such factors provide formidable barriers to the invasion of genomes by infective mobile elements as well as by alien coding sequences. Without such barriers genomes would inevitably expand to extinction.

Selfish behavior

Notions of genetic selfishness have played an important role in the texts that introduced the notion of HGT. We explore the idiosyncratic ways authors have used "selfish" in this context in order to clarify the view developed here of HGT as an invasive mode of sequence propagation.

We begin somewhat arbitrarily with Hamilton's now classic early work on kin selection and inclusive fitness.[17,18] Hamilton demonstrated that apparently altruistic behavior in organisms could be interpreted as the behavioral phenotypes of genetic determinants that favor their own propagation in succeeding progeny. Here, seemingly altruistic behavior patterns are understood as an expression of genes that are "selfish" in that their inclusive fitness supports their selection over other genes or alleles. An important consequence of Hamilton's work is that we no longer speak about group selections such as "selection of the species". Instead, we speak in the behavioral context about "selfish genes".

The import of Hamilton's work is not derived from his use of the term "selfish" in a genetic context. What is important is that Hamilton developed formalisms with which to calculate the genetic consequences of behavior.[17,18] His assessment was carried out with a measure of fitness that depends solely on the expected impact of behavior on the frequencies with which genes for behavior are propagated in populations. The term selfish in this behavioral context is useful then as a denial of genetic altruism. Nevertheless, from the point of view of population genetics, it is business as usual.

The wider currency of "selfish genes" in contemporary texts is in good measure due to a book published more than a decade after Hamilton's classics. *The Selfish Gene* by Richard Dawkins[1] initially asserts that all genes are selfish in the sense used by Hamilton. However, genes in general are not usually thought of as altruistic. So, while the use of selfish in the behavioral context may be liberating, it does not as far as I know solve any technical problems in the wider context of classical genetics. On the other hand, "selfish genes" mark classical genetics as a part of the behavioral ecologist's territory very much as large carnivores mark their territories with arbitrary but distinctive signals.

Caveat

According to Dawkins "if you look at the way that natural selection works, it seems to follow that anything that has evolved by natural selection should be selfish" (ref. 1, p. 4). Thus, his notion of selfish evolution is initially like Hamilton's,[17,18] that of conventional Darwinian selection of the fittest individuals. It turns out to be very important in reading his book to recall this way of using "selfish" because Dawkins uses the term in several different ways that are relevant to our discussion of HGT. Furthermore, there are instances in the text where Dawkins hints, perhaps unintentionally, that selfish genes can explain things that are inexplicable by conventional notions of selection because selfish genes transcend the confines of selection for the fittest phenotype.

For example, Dawkins' offers his view of mutators, genes that increase the copying errors in other genes. He claims, "by definition a copying error is to the disadvantage of the gene which is miscopied. [Nevertheless, if]it is to the advantage of the selfish mutator that induces it, the mutator can spread through the gene pool" (ref. 1, p. 47). The implication here is that, even if the effects of a mutator gene are always destructive for other genes, the mutator might be propagated because of its own selfishness. How could this be?

Dawkins might have clarified such comments on mutators by mentioning that mutations in fact are not always destructive, though they lead to the replacement of alleles. Rather, copying errors may be adaptive and mutant alleles may be selected, particularly, under challenging circumstances. Indeed, mutator genes that enhance the frequencies of adaptive mutations are identifiable in microbial cultures growing under selective conditions. Such mutators are propagated in the surviving progeny by hitchhiking within the genomes that have been selected for their mutant phenotypes.[19] Or, as Dawkins might

have said, mutator genes propagate selfishly because copying errors are not always destructive.

Dawkins motivated his discussion of sex by underscoring the difficulties of understanding the selective forces that have driven its evolution. He offers the notion of selfish genes to resolve these difficulties. Thus, ''if crossing over benefits a gene for crossing over that is sufficient explanation for the existence of crossing over. And, if sexual, as opposed to non-sexual, reproduction benefits a gene for sexual reproduction that is sufficient explanation for the existence of sexual reproduction. Whether or not it benefits all the rest of an individual's genes is comparatively irrelevant. Seen from the selfish gene's point of view, sex is not bizarre at all.'' (ref. 1, p. 47).

To be sure, Dawkins confesses, ''This comes perilously close to being a circular argument. . . . [Nevertheless,] I believe there are ways of escaping from the circularity but this book is not the place to pursue the question.'' (ref. 1, p. 47)

Selfish DNA

The selfish metaphor seems to work well to describe ''surplus DNA'' or ''selfish DNA'' as ''parasitic DNA''.[1] Thus, Dawkins identifies eukaryotic repetitive sequences as quintessentially selfish in the common parlance sense because they seem not to encode any functions for their hosts. Nevertheless, they are massively propagated in the genomes of eukaryotes. In primates they make up a much larger fraction of the genome than do protein-coding sequences.

Note, the sense of ''selfish'' as ''parasitic'' is a second, distinguishable way that the word has been employed by Dawkins.[1] Genome parasites are not ''selfish'' in the sense employed initially by Hamilton.[17,18] From the perspective of the host genome, selfish sequences that are propagated by natural selection of the fittest individuals in a population are simply selected sequences. Parasitic repetitive sequences might be counter-selected or nearly neutral from the perspective of their host genomes.[3] Thus, the insertion of a parasitic sequence into a noncoding region of a genome may have little or no influence on its fitness, but an insertion into a coding sequence would most likely be deleterious. Furthermore, expression of the inserted sequence might have a debilitating effect on the host phenotype. In any case, the inserted parasitic sequence would raise the cost of replication of the transgenic genome.[3]

The term ''selfish'' is further complicated by a third interpretation that was applied by Doolittle and Sapienza[2] in a more extreme sense than that initially employed by Dawkins.[1] Doolittle and Sapienza[2] initially identified eukaryotic repetitive sequences with the mobile elements identified earlier in prokaryotes. However, their interpretation of such eukaryotic repetitive sequences is that, because they do not encode useful information for their hosts, they lack a phenotype. They go on to infer that these sequences must be propagated by what Dawkins in another context refers to as purely ''genic competition''.[1] This they suggest is a departure from the conventional paradigm of selection for the fittest phenotype.[2]

The inspiration for this interpretation seems to come from superficial anomalies such as the mutator genes mentioned by Dawkins,[1] and segregation distorter genes in *Drosophila* as well as sterility factors in a variety of creatures, to which we will return. As Doolittle and Sapienza say, such genes are so selfish ''as to promote their own spread through a population at the ultimate expense of the evolutionary fitness of that population''. In other words they seem to be propagated independently of selection for the fittest phenotype. So, Doolittle and Sapienza[2] propose that these determinants ''arise and are maintained by what [they] call non-phenotypic selection''. Whatever it means, ''non-phenotypic selection'' is presented as a version of Dawkins' genic competition.

The failure to couple the ''non-phenotypic selection'' of repetitive sequences within a host genome to the ''parasitic phenotype'' of mobile elements is puzzling. So too is Doolittle and Sapienza's preoccupation with a shift away from the conventional paradigm of phenotypic selection.[2] As it turns out, some of the genetic determinants that have been identified as distorters of segregation and sex ratios are simply cellular parasites[20] while others are genome parasites.[21,22]

Some informative parasites

Segregation distortion, male sterility, male killing and feminization are common transmissible characters among insects.[20] These determinants often are inherited maternally, transmitted via the cytoplasm of oocytes, and ''curable'' by heat shock or by antibiotics. The cytoplasmic agents responsible for these transmissible characters are microorganisms that may be as disparate as mycoplasma (Spiroplasma), alpha-proteobacteria (Wolbachia) and microsporidia (Amblyospora).[20] And, most important, as parasites are known to do, they are propagated by their propensity to infect prospective hosts, in these cases by the ovarian route. Likewise, transposable elements that distort sex ratios in insects are identifiable as genome parasites that can be duplicated vertically as well as transferred horizontally.[21,22] All of these parasites are of interest here because they can help to demystify the behavior of repetitive DNA sequences.

The seemingly anomalous characteristic of the genetic determinants responsible for distortion of sex ratios or segregation is that they are persistent residents in organisms even though they have a demonstrably negative influence on the fitness of their hosts. One might be tempted therefore to search for some cryptic service that these parasites might perform for their hosts in order to compensate for distortions of the sex ratio and the like. But, if they are infective parasites, their propagation by multiple infections of host cells can far outstrip their negative influences on individual host fitness. In that case, no cryptic service would be required.

Orgel and Crick[3] note that, in addition to normal replication and segregation to progeny, what may be referred to as vertical transfer, selfish DNA also may propagate by transposing to different positions in its host genome or hopping to other related or unrelated genomes. The latter is horizontal transfer, and it is recognizable as an infectious mode of propagation for transposable elements. When the dynamics of propagation are modeled, it is found that, for sufficiently large rates of infectious propagation, there may be compensation for the loss of fitness created by the transposable element. In effect, the deleterious determinants may be propagated as though they were selected because they are capable of multiplying faster than their hosts.[3,16,22] Accordingly, mobile DNA sequences are well described as the ultimate parasites.[3]

Natural barriers

Orgel and Crick[3] discuss at some length the notion that a sequence acquired by HGT might turn out to improve the fitness of a host organism. Although they argue that this would not be a common event, they conclude, ''It would be surprising if the host organism did not occasionally find some use of particular selfish DNA sequences''. Accordingly, transferred sequences may on occasion become selected sequences, but for the reasons outlined below this is not a common expectation.

In fact, an important objective of subsequent population genetic studies of HGT has been to understand how host populations accommodate the infectious feature of genomic parasites.[22–24] Thus, genomic parasites have the potential of multiplying faster than their hosts do. If uncontrolled, this tendency could lead to the extinction of the host. What are the barriers to the spread of parasitic DNA?

Here, the notion of a barrier to HGT is used to describe molecular mechanisms that reduce the probability that a transferred sequence will be propagated invasively within a population of genomes. A barrier to the spread of a transferred sequence might result from disabling the transfer mechanism by mutation or from the loss of fitness to the host genome carrying a transferred sequence. In either case, the infective propensity of the transferred sequence is reduced.

Two inescapable costs to fitness are associated with the transposition of parasitic DNA.[3] One cost is the mutational load created by the insertion of transposable sequences into coding sequences. The other is the metabolic cost of replicating noncoding DNA, which for the aggregate of many such sequences might be substantial. Significant amounts of noncoding DNA might also lengthen the time required to replicate the genome, and increase cell size necessary to accommodate this extra DNA. This may account for the positive correlations observed between total genome size, cell size and cellular division times.[22–24] Insertion of transposable elements into chromosomes also may create destructive rearrangements and deletions. Likewise, the expression of the coding sequences that are carried by some transposable elements may cause serious harm to host cells by the induction of DNA strand breaks.[22] Finally, there are the distortions of chromosome segregation and biased sex ratios mentioned above.

Accordingly, the accumulation of significant quantities of transposable elements would tend to lower the fitness of the infected organisms compared to that of uninfected competitors. So, on the one hand, even infective genome parasites are subject to the constraints of selection for fitness. On the other hand, such counterselection comes as an added negative load on the infected host population, which suggests that there is also a premium on mechanisms that constrain the spread if not completely eliminate parasitic DNA from genomes. It is tempting to speculate that the rapid growth rates and small cell sizes characteristic of prokaryotes with their relatively low incidence of parasitic DNA are dependent on stringent containment of transposable elements. Indeed, calculations have shown that deletion rates in *Escherichia coli* are sufficient to contain the influx of inserted sequences.[16]

The prokaryotic tendency to purge nearly all transposable elements may be shunned by most eukaryotes because of their much larger coding capacities. Thus, for larger genomes cure by deletion may be worse than stoically bearing a larger steady-state load of parasitic DNA. Thus, the larger the aggregate coding capacity of a genome, the more likely it is that the genome can be lethally damaged by any given rate of random deletion. So, while augmented random deletion rates in eukaryotes might in principle eliminate parasitic sequences, the attendant lethal load on the population might be so high as to preclude the survival of the host genomes (Kurland, unpublished results).

There are exceptions to such parasite-tolerant genomes among vertebrates. Indeed, the sanitized genomes of smooth skinned puffer fish and their relatives[25,26] are very suggestive. Here, the cost to the population created by purging parasitic sequences[26] might be unusually tolerable because the burst of progeny from a single pair of puffer fish is so large that the demise of most of the zygotes before sexual maturity might be inconsequential. The existence of these minimal genomes in puffer fish suggests that eukaryotes potentially do have efficient mechanisms for eliminating parasitic DNA when their life style permits extreme measures.

There is evidence from both prokaryotic and eukaryotic genomes that short deletions are more prevalent than short insertions.[27–31] However, on the one hand, calculations suggest that short deletions might not be sufficiently frequent to limit the spread of transposable elements in eukaryotes.[30,31] On the other hand, the best estimates of deletion rates are biased towards very short deletions because of the reporter systems used to detect them.[31] A heuristic guess then is that very large deletions of eukaryotic genomes may contribute to the containment or elimination of transposable elements.[31]

Finally, cellular pathways involving methylation of DNA and RNA as well as the modification of histones have begun to emerge as silencing mechanisms for transposable elements.[32] The notion that silencing mechanisms that function in cellular differentiation programs may also provide barriers to the spread of infective DNA seems very attaractive.

We are still far from a detailed understanding of how the sizes of genomes evolve, or how the load of parasitic DNA is stabilized in eukaryotes.[31–34] However, given the menu of deleterious effects generated by infective sequences, it is a safe bet that their propagation is always constrained in genome populations that survive their initial infection. The very fact that different eukaryotic genomes carry characteristic loads of repetitive[3,25,31–34] sequences suggests that these are maintained in some sort of dynamic balance between their propensity to multiply and the contingent capacities of their hosts to eliminate or to contain them.

Invasive coding sequences

The provocative fact is that alien gene transfer between prokaryote genome populations is fairly frequent. Nevertheless, there is a wealth of data showing that alien coding sequences though well represented in bacterial genomes are often transients and that their residence is usually restricted to small subpopulations within a host global population.[5–16] Since alien sequences do not persist very long nor do they seem to spread very far within the host genome population, the frequencies with which they occur are not reflected in the phylogeny of global populations. Simply stated, transferred coding sequences are rarely stable residents of an entire global population of prokaryotes. In order to understand why this is so, we need to understand the barriers to the migration of alien sequences through large populations of small genomes.

First, in natural populations, the transfer of a particular alien DNA sequence is likely to be seeded in only a relatively small number of host cells. The efficiency with which it can invade the global population is determined, as above, by the combination of its infectivity and its selective impact on the host cell. If the combination of these two factors is positive, there will be a net tendency for the alien sequence to be invasive.[16] Opposing this invasive tendency are destructive random mutations and the size of the global population.[16] For the very large populations of bacteria, which may be of the order of $10,^{21}$ strong selection is required to fix the alien sequence within the global population.[16] When the alien sequences are weakly selected, neutral or counter selected, the lifetime of the alien gene within the acceptor population will be cut short by random destructive mutations.[14,16]

If the alien sequence is infective as for a transposon, the infectivity will function like selection.[16] That is to say repeated insertions can replace infective sequences that are purged by random mutation. However, it is worth emphasizing that only sequences that support the infectivity per se will be invasive. Coding sequences that hitchhike on a mobile element will be susceptible to mutational degradation unless they are themselves selected.[16] So, the bottom line is that only strong continuous selection or repeated infection will allow transferred sequences to run the gauntlet of random mutation to establish residence in a large transgenic global population.

Second, genes do not evolve in atomistic, self-contained modes. Rather, gene products evolve coordinately and cooperatively so that mutually compatible molecular structures that interact physically and that are regulated in coherent, dependent ways are selected in modern organisms, as they may not be in the most-primitive cellular collectives characteristic of earliest evolution.[41,42] Accordingly, HGT, like most mutational change tends to be deleterious to the workings of modern cellular collectives because it tends to introduce molecular incompatibilities. Thus, an alien gene product that evolves in a molecular environment different from that provided by a new host cell will not be structurally tuned to those host components with which it must interact. For this reason, alien transfers are often toxic and transgenic cells will tend to be less robust than native cells in the absence of specific selective agents.[14,35,43–48] The functional incompatibilities of alien transfers with the molecular environment of a host cell will retard its spread through the host population and enhance the probabilities of it being purged by random mutations.[14,16]

A third general barrier to the propagation of transgenic organisms is the patchiness of the environment. Even when a selective condition is at work driving the propagation of a transgenic phenotype, that selective condition will not in general be uniformly distributed over the entire range of the organism. Where the selective condition is absent, the transgenic will tend to be excluded or lost. The bottom line is that strongly selected sequences will persist only as long as specific selection for their phenotypes persists. In bacteria, for example, plasmids carrying antibiotic resistance may be perpetuated where there is antibiotic, but elsewhere that plasmid is often debilitating to its host.[14,16,35,47,48] Even simple mutations such as nucleotide substitutions that lead to antibiotic resistance are often associated with growth defects in the absence of antibiotic.[46–48] This is the reason that hospitals are such dangerous places for encounters with antibiotic-resistant bugs.

These constraints on the invasiveness of alien coding sequences are expressed in the dynamics of HGT in microbial populations.[16,36–40] Thus, alien sequences, either intact genes or degraded pseudogenes, in one patch of a bacterial taxa are likely to be missing in other patches of the same taxa. Furthermore, comparisons of the distributions of genome sequences in different bacterial patches of the same taxa or in diverse taxa provide strong evidence that these sequences are turning over at a quite appreciable rate.[16,28,36–40] It is estimated that, in *Escherichia coli*, remnants of a neutral alien

gene sequence would circulate in small patches of organisms for roughly one million years before being purged from the genome.[16] Other calculations suggest that any particular alien neutral sequence may be present in roughly 0.1 percent of the *E. coli* global population.[16] So, when not driven by strong selection, HGT will appear in a heterogeneous patchwork of transient sequences diffusing through the global population in various states of mutational disarray. Since alien gene transfers tend to be transient residents in minute patches, they have little impact on the phylogeny or the global adaptations of their microbial hosts.

Conclusion

Genomes that continuously expand due to the uninhibited acquisition of horizontally transferred sequences are genomes that are earmarked for extinction. At the very least, an unchecked increase in replication costs and replication time associated with the most-benign sequence insertions must ultimately exact its toll in a loss of fitness compared to genomes that are less tolerant of HGT. Although the investment in repetitive sequences in modern animal genomes may seem excessive, it is evident that selective pressures temper the steady-state levels of transposable elements in animal cells.[21–23] Though there are mechanisms to modulate the infectivity of horizontally transferable sequences, once inserted, a transferred sequence will tend to be faithfully propagated in succeeding generations unless it is purged from the genome. It is therefore not surprising that, in all organisms that have been studied, deletion rates exceed insertion rates during replication.(27–31)

We have explored two extreme modes of HGT. One based on the invasiveness of transgenic bacteria in very large populations of single cells with relatively small genomes. The other is based on infectious mobile elements propagating in relatively small populations of multicellular organisms with relatively large genomes. In both, an adequate combination of deletion rates as well as mechanisms to modulate the infectivity of mobile sequences is paramount to the survival of the host genomes. Moreover, the expression of transferred sequences is likely to have a negative impact on the fitness of transgenic cells due to the incompatibilities of gene products that have evolved under different cellular constraints.[12,13,16] Counter-selective factors such as these that oppose the invasiveness and stability of inserted sequences in modern organisms constitute the natural barriers to rampant HGT.

Doolittle's essay[4] describing an extreme view of the impact of alien gene transfer on genome evolution may be read as a return to earlier thinking about selfish genes that escape the normal constraints of selection for fitness.[1,2] More recent data show that Doolittle's essay greatly overestimated the incidence of alien gene transfers as well as their impact on phylogeny.[5–10] It is arguable that such overestimates of HGT seemed reasonable at the time because a fundamental dimension of genome evolution was neglected, namely, that size matters.

Acknowledgments

I thank the Kungliga Fysiografiska Sällskapet, Lund, Sweden and the Alan Wilson Center, Palmerston North, New Zealand for support to visit the Alan Wilson Centre at Massey University, Palmerston North, New Zealand, where versions of this paper were written. I also am very grateful to Otto Berg, Jeffrey Kurland, Peter Lockhart, David Penny, Adam Wilkins and Irmgard Winkler for much useful criticism and advice.

References

1. Dawkins R. 1976. *The Selfish Gene*. New York: Oxford University Press.
2. Doolittle WF, Sapienza C. 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601–603.
3. Orgel LE, Crick FHC. 1980. Selfish DNA: the ultimate parasite. Nature 284:604–607.
4. Doolittle WF. 1999. Phylogenetic classification and the universal tree. Science 284:2124–2129.
5. Snel B, Bork P, Huynen M. 1999. Genome phylogeny based on gene content. Nat Genet 21:108–110.
6. Snel B, Bork P, Huynen M. 2002. Genomes in flux: the evolution of Archaeal and Proteobacterial gene content. Genome Res. 12:17–25.
7. Fitz-Gibbon ST, House CH. 1999. Whole genome-based phylogenetic analysis of free-living microorganisms. Nucleic Acids Res. 27:4218–4222.
8. Tekaia F, Lazcano A, Dujon B. 1999. The genomic tree as revealed from whole pProteome comparisons. Genome Res. 9:550–557.
9. Brown JR, Douady CJ, Italia MJ, Marshall WE, Stanhope MJ. 2001. Universal trees based on large combined protein sequence data sets. Nat Genet 28:281–285.
10. Korbel JO, Snel B, Huynen MA, Bork P. 2002. SHOT: A Web Server for the construction of genome phylogenies. Trends Genet. 18:158–162.
11. Gogarten JP, Doolittle WF, Lawrence JG. 2002. Prokaryotic evolution in light of gene transfer. Mol Biol Evol 19:2226–2238.
12. Kurland CG, Canback B, Berg OG. 2003. Horizontal gene transfer: a critical view. Proc Natl Acad Sci USA 100:9658–9662.
13. Kurland CG. 2004. Paradigm Lost. In: Microbial Evolution: Concepts and Controversies, Ed. Jan Sapp. New York: Oxford University Press.
14. Kurland CG. 2000. Something for everyone. EMBO Reports 1:92–95.
15. Canback B, Andersson SGE, Kurland CG. 2002. The global phylogeny of glycolytic enzymes. Proc Natl Acad Sci USA 99:6097–6102.
16. Berg O, Kurland CG. 2002. Evolution of microbial genomes: Sequence acquisition and loss. Mol Biol Evol 19:2265–2276.
17. Hamilton WD. 1964. The genetical evolution of social behavior, I. J Theoretical Biology 7:1–16.
18. Hamilton WD. 1964. The genetical evolution of social behavior, II. J Theoretical Biology 7:17–52.
19. Taddei F, Radman M, Maynard-Smith J, Touponce B, Gouyon PH, Godelle B. 1997. Role of mutator alleles in adaptive evolution. Nature (London) 387:700–702.
20. Ishikawa H. 2003. Insect Symbiosis: an Introduction. In: Bourtzis K, Miller TA, editors. Insect Symbiosis. Boca Raton: CRC Press LLC. p 1–21.
21. Charlesworth BP, Sniegowski PD, Stephan W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215–220.
22. Charlesworth B, Langley CH. 1989. The population genetics of Drosophila transposable elements. Ann Rev Genet 23:251–287.
23. Green MM. 1980. Transposable elements in Drosophila and other insects. Ann Rev Genet 14:109–120.
24. Gregory TR. 2002. Genome size and developmental complexity. Genetica 115:131–146.
25. Brenner S, Elgar G, Sanford R, Macrae A, Venkatesh B, Aparacio S. 1993. Characterization of the pufferfish (Fugu) as a compact model vertebrate genome. Nature 366:265–268.
26. Neafsey DE, Palumbi SR. 2003. Genome size evolution in pufferfish: a comparative analysis of diodontid and tetradontid puffer fish. Genome Research 13:821–830.

27. Andersson SGE, Kurland CG. 1998. Reductive evolution of resident genomes. Trends Microbiol 6:263–268.
28. Mira A, Ochman H, Moran NA. 2001. Deletional bias and the evolution of bacterial genomes. Trends Genet 17:589–596.
29. Petrov DA, Hartl DL. 1998. High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species group. Mol Biol Evol 15: 293–302.
30. Petrov DA, Sangster TA, Johnston JS, Hartl DL, Shaw KL. 2000. Evidence for DNA loss as a determinant of genome size. Science 287:1060–1062.
31. Gregory TR. 2004. Insertion deletion biases and the evolution of genome size. Gene 324:15–34.
32. Matzke MA, Mettte MF, Aufsatz W, Jakowitsch J, Matzke AJM. 1999. Host defenses to parasitic sequences and the evolution of epigenetic control mechanisms. Genetica 107:271–287.
33. Hartl DL. 2000. Molecular melodies in high and low C. Nature Rev Genet 1:145–149.
34. Nuzhdin SV. 1999. Sure facts, speculations and open questions about the evolution of transposable copy number. Genetica 107:129–137.
35. Björkman J, Nagaev I, Berg OG, Hughes D, Andersson DI. 2000. Effects of environment on compensatory mutations to ameliorate the costs of antibiotic resistance. Science 287:1479–1482.
36. Lawrence JG, Ochman H. 1997. Amelioration of bacterial genomes: rates of change and exchange. J Mol Evol 44:383–397.
37. Lawrence JG, Ochman H. 1998. Molecular archaeology of the *Escherichia coli*. Genome Proc Natl Acad Sci USA 95:9413–9417.
38. Hooper SD, Berg OG. 2002. Gene import or deletion—a study of the difference genes in *Escherichia coli* Strains K12 and O157:H7. J Mol Evol 54:734–744.
39. Ochman H, Jones IB. 2000. Evolutionary dynamics of full genome content in *Escherichia coli*. EMBO J 19:6637–6643.
40. Perna NT, Plunkett G, III, Burland V, et al. 2000. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature (London) 409: 529–533.
41. Woese CR. 1998. The universal ancestor. Proc Natl Acad Sci USA 95: 6854–6859.
42. Woese CR. 2000. Interpreting the universal phylogenetic tree. Proc Natl Acad Sci USA 97:8392–8396.
43. Asai T, Zaporojets D, Squires C, Squires CL. 1999. An *Escherichia coli* strain with all chromosomal rRNA operons inactivated: complete exchange of rRNA genes between bacteria. Proc Natl Acad Sci USA 96: 1971–1976.
44. Rao AR, Varshney U. 2001. Specific interaction between the ribosome recycling factor and the elongation factor G from *Mycobacterium tuberculosis* mediates peptidyl-tRNA release and ribosome recycling in *Escherichia coli*. EMBO J 20:2977–2986.
45. Ohnishi M, Janosi L, Shuda M, Matsumoto H, Terawaki T, Kaji A. 1999. A. molecular cloning, sequencing, purification and characterization of *Pseudomonas aeruginosa* ribosome recycling factor. J Bacteriol 181: 1281–1291.
46. Kurland CG. 1992. Translational accuracy and the fitness of bacteria. Annu Rev Genet 26:29–50.
47. Andersson DI, Levin BR. 1999. The biological cost of antibiotic resistance. Curr Op Microbiol 2:487–491.
48. Bergstrom CT, Lipsitch M, Levin BR. 2000. Natural selection, infectious transfer and the existence conditions for bacterial plasmids. Genetics 155:1505–1519.