When natural selection gives gene function the cold shoulder

Asher D. Cutter and Richard Jovelin*

It is tempting to invoke organismal selection as perpetually optimizing the function of any given gene. However, natural selection can drive genic functional change without improvement of biochemical activity, even to the extinction of gene activity. Detrimental mutations can creep in owing to linkage with other selectively favored loci. Selection can promote functional degradation, irrespective of genetic drift, when adaptation occurs by loss of gene function. Even stabilizing selection on a trait can lead to divergence of the underlying molecular constituents. Selfish genetic elements can also proliferate independent of any functional benefits to the host genome. Here we review the logic and evidence for these diverse processes acting in genome evolution. This collection of distinct evolutionary phenomena – while operating through easily understandable mechanisms – all contribute to the seemingly counterintuitive notion that maintenance or improvement of a gene's biochemical function sometimes do not determine its evolutionary fate.

Keywords:
gene function; genome evolution; natural selection

Introduction

Modern evolutionary theory lays out clearly the familiar textbook determinants of heritable change over time as recorded in the DNA of genomes. It is easy to be lured into the view that evolution is strictly progressive, that selection perpetually favors the functional improvement of each individual gene in the genome. But this is not the case [1]. It should come as no shock that mutation and genetic drift can cause fluctuations in gene frequency in ways that undermine organismal fitness, provided that mutation rates are high enough or population sizes small enough. That is, over time, chance in reproduction can allow a detrimental mutation – even one that knocks out a gene's functional activity – to rise in abundance and become the predominant allele in the species. Classic experimental evolution studies deftly illustrate the truth of this principle [2]. Narrow expression breadth can make a gene especially susceptible to the influence of mutation and drift despite selection for the maintenance or improvement of gene activity, as for gene expression restricted to post-reproductive individuals that are subject to weaker selection (e.g. senescence). Genetic drift can also do more than just randomly change allele frequencies, potentially setting the stage for profound evolutionary changes at the heart of basic cellular processes and genomic features (e.g. introns) [3]. But under which other circumstances can evolutionary change occur independent of the specific benefits or costs of gene functional activity? When does functional activity of a gene not guarantee its persistence? Which roles can selection play in promoting this process? It turns out that the diverse answers to these superficially counterintuitive questions are straightforward (Fig. 1), if not always standing together in the limelight of evolutionary thinking.

DOI 10.1002/bies.201500083

Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada

***Corresponding author:**
Asher D. Cutter
E-mail: asher.cutter@utoronto.ca

Gene function can degrade as a byproduct of selection interacting with genetic linkage

Natural selection is a powerful agent of gene frequency change, capable of taking an allele from rarity to near fixation in short order. Specifically, a beneficial allele will sweep through a population in about $\ln(2N_e s)/s$ generations [4]. For reasonably strong selection (s) in a population with effective population size (N_e) less than a million, this corresponds to hundreds of generations or less. In such a selective sweep, the favorable allele also will drag with it a stretch of the genome to make a long allelic haplotype that far exceeds the size of the beneficial mutation itself. Again, we can predict the genomic span of such genetic hitchhiking: $s/[r \ln (N_e s)]$ basepairs of DNA sequence [5, 6]. The lower the recombination rate (r) and the stronger the selection and the smaller the population, the wider will be the

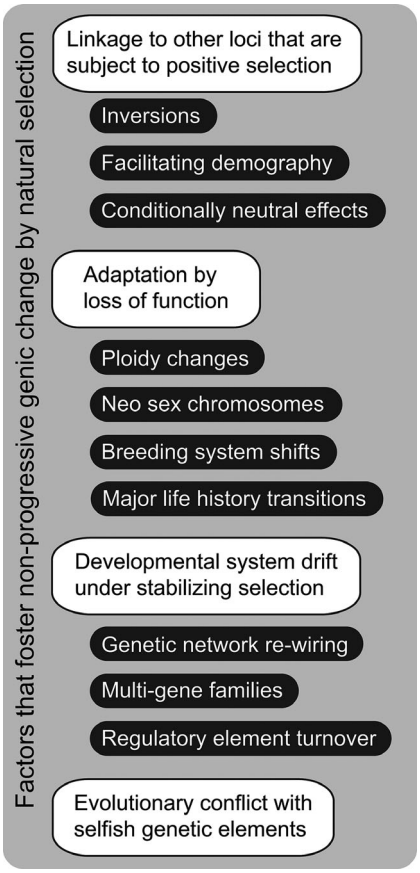

Figure 1. Diverse sources of change, mediated by natural selection, can lead to the evolution of genic degeneration or non-progressive genic change with respect to biochemical function and activity.

expanse of the genome captured within the swept haplotype. This means that selection at one gene influences the evolution of other genes, simply due to the happenstance of physical proximity in the genome. Alleles at other genes within the genomic region that are linked to the selectively favored site can thus rise in frequency and become fixed in the population, irrespective of any effects on function, so long as any negative fitness effects are sufficiently weak compared to the favored linked allele. Conversely, mutations increasing organismal fitness may be kept at low frequency in the population, or may be lost altogether if they are linked to a strong deleterious allele because of "background selection" [7]. In practice, selective interference between any such linked alleles with selection coefficients of opposing sign will narrow the width of the hitchhiking haplotype [6], but nevertheless it is not unexpected that the reach of genetic hitchhiking will span many, many genes.

Which factors can facilitate how natural selection on one locus can drive incidental divergence affecting gene function at linked loci? Here, we will highlight three features that may be especially potent. First, DNA inversions can capture whole swathes of genes in the genomic neighborhood of a beneficial allele. Such "inverted haplotypes" cannot recombine with standard DNA segments, and so will be passed along intact, the alleles all linked to each other and transmitted as a single unit [8]. Any inversion that captures a beneficial allele could easily also contain function-altering alleles at other loci that would not be able to recombine off of the inverted haplotype. Second, demographic factors that broaden the width of the hitchhiking region include small population size and self-fertilization. Populations in which individuals fertilize themselves will have an overabundance of homozygous genotypes, and recombination is impotent to create new gene combinations between homozygous genotypes. Consequently, hitchhiking haplotypes will be especially long, and capture more genes [5]. Self-fertilizing *Caenorhabditis* nematodes provide an exceptional example of how selfing plays into this phenomenon, made even more extreme because the parts of the genome with higher densities of genes also have lower natural rates of recombination [9–11]. Third, a transient eco-environmental shift that provides the evolutionary arena for the favorable, sweeping allele might also provide circumstances for conditional neutrality of function-altering alleles at linked loci. That is, those linked alleles might exert conflicting fitness effects in the old environment, but be selectively neutral in the new environment and so not generate selective interference during the sweep. While perhaps this sounds suspiciously convenient, it turns out that conditional neutrality as a form of genotype-by-environment interaction is a common feature of the genetic architecture of trait variation in nature [12]. All three of these evolutionary scenarios will facilitate the evolution of function-altering changes through indirect means owing to their incidental genetic linkage to alleles at other loci that experience positive selection.

Adaptation by loss of function can cause gene degradation

We commonly conceive of adaptation as the acquisition of novel traits, perhaps less consciously presuming that novel genic gain of function changes are responsible or, at least, that gene functional activity improves during adaptation. However, under many circumstances, adaptation actually operates by loss of gene function through the acquisition of null mutations, gene deletions, and regulatory disruptions. Thus, adaptation by loss of function exemplifies another way that gene function need not matter in evolution [13], in the sense that improved biochemical activity of a protein is not favored by natural selection. There are many examples of this phenomenon in nature (e.g. flower coloration and timing [14–16], HIV resistance [17]), in domestication traits (e.g. myostatin "double-muscling" of cattle [18], crop height [19]), and in laboratory experimental evolution [20, 21]. Notably, adaptation through gene loss can lead to rewiring of metabolic networks to exploit the new conditions without modifying individual protein function at all [22]. More rapid sampling of distant portions of the fitness landscape might thus occur through beneficial null mutations than through small-effect gain-of-function mutations. Null mutations can enjoy strong selection favoring their spread, and yet genome integrity imposes an inevitable upper limit on how pervasive this mechanism of adaptation can be in the long term. Many loss-of-function mutations, however, are not necessarily permanent. For example, reversion point mutations to a pre-mature stop codon could reconstitute gene function or, more likely mechanisms, secondary indel mutation can restore reading frame to a frame shift-induced null allele or retrotransposons can excise themselves from regulatory regions [23]. For such reversible mutations, permanent loss of function requires killing the gene two or more times over with mutation.

Major evolutionary transitions can set the stage for extensive adaptation by gene loss. For example, the genesis of new sex chromosomes can produce selection for chromosome degradation by gene loss [24]. Drastic shifts in life history also make opportunity for this process to run wild, as in the evolution of self-fertilization and asexuality [25]. A convergent evolutionary feature of plants and animals that evolve self-compatibility is the degeneration of traits that would otherwise facilitate outbreeding, yielding a suite of characteristic traits known as the "selfing syndrome" [26, 27]. Oftentimes, the origins of self-fertilization are too recent for degeneration by drift to provide a plausible explanation for the rapid evolution of diminutive reproductive characters, implicating adaptation to the selfing lifestyle [28]. That is, while relaxed selection on genes encoding male and outcrossing functions heightens the relative role of drift in their molecular evolution, it appears that selection also must often favor their elimination from the genome. In brassicaceous plants and *Caenorhabditis* nematodes, this manifests as genomic shrinkage and deletion of male-biased genes [29–31].

Do similar fates await the genes in species that experience other sharp changes in selective regime? An open question is whether all or most major adaptive shifts to novel conditions, of which shifts in ploidy, sex linkage, and breeding system represent powerful examples, will commonly involve extensive adaptation by loss of gene function. Moreover, many genes perform multiple functions, owing to changes in the spatio-temporal context of their expression within an organism. From the empiricist's perspective, aware of only a subset of a gene's biological roles, it may appear that function has been lost, despite selection having dominated its evolution through fitness effects in other biological contexts. The permanence of gene loss will depend on sufficient time for fixation of null alleles and, in the case of reversible null alleles, sufficient time for subsequent mutations to reinforce a null as a true pseudogene. Consequently, the novel ecological conditions must persist for long enough for this to happen. Null mutations might be predisposed to contribute to such adaptation, in the following way: the myriad ways that mutation can induce loss of gene function provides a steady rain of null alleles into the genomes of a population. Loss of function alleles tend to be recessive [32], allowing them to accumulate as part of the standing genetic variability within populations owing to the trouble that selection has in completely weeding out recessive alleles. Strong selection upon an eco-environmental shift will favor those alleles in the standing variability with greatest fitness effects, irrespective of dominance [33]. Consequently, the potentially large fitness effect of null alleles could predispose them in the adaptive response to such altered selection regimes [34].

Stabilizing selection can lead to evolutionary changes

Whole genome duplication is common across the tree of life and an important foundation for diversification. Most gene duplicates are deleted and lost over evolutionary time [35]. However, the initial functional redundancy of new paralogs can permit their preservation through the accumulation of complementary mutations in their regulatory control elements [36]. Stated differently, mutations to distinct regulatory elements in each gene duplication act to partition the activity of the single ancestral copy between the descendant paralogs. Such sub-functionalization evolving independently in separate lineages can then cause orthologs to act at different times or places in an organism, in other words, to perform different functions, despite maintaining overall functional capacity [37]. Expanded gene families can experience rapid turnover of their closely-related members [38], implying selective exchangeability in the individual functional effects of members. Perhaps counterintuitively, these processes suggest that it can be selection to maintain an unchanging organismal phenotype that drives the presence or absence of a given gene in the genome, or the divergence of gene function among gene family members [38]. We are not aware of an explicit model for this kind of stabilizing selection scenario on the phenotypes associated with large multigene families comprised of similar and partially redundant members, though it is reminiscent of Developmental System Drift [39] and the "dosage balance hypothesis" for ploidy changes [35]. However, we suspect that extensions to polygenic selection models might predict extensive population polymorphism in gene presence-absence (i.e. copy number variation, CNV) and in premature stop alleles in such multigene families [40, 41]. Similarly, radical turnover in the sequences and locations of non-coding regulatory motifs can occur despite stabilizing selection on gene expression [42], meaning that selection on the activity of the gene as a whole dominates over the specific function of any individual regulatory motif. More generally, gene regulatory networks diverge constantly because of functional redundancy, compensatory mutations, epistatic interactions, and gene co-option that maintain developmental programs of traits under stabilizing selection [43–45]. This pattern, called Developmental System Drift, describes the often observed disconnect between homology at the phenotypic and molecular levels [39].

Functional degradation can be caused by selfish genomic elements

From the perspective of the organism, selfish genetic elements, such as the transposable elements (TEs) that reside in their genomes, are usually devoid of function. And yet, TEs nevertheless proliferate to pepper the genomes of virtually all organisms. Of course, the influence of TEs is more nuanced than this blanket statement might suggest, as attested, for example, by their role in mediating adaptations [46, 47] and, perhaps, in spawning the origins of molecular mechanisms of genomic surveillance [48, 49]. While TEs have the potential to play adaptive roles in evolution, most often, organismal fitness suffers from TE activity, owing to the detrimental consequences of their mutational insertion. Moreover, many TEs do not even encode the enzymes

necessary for their own transposition, instead exploiting those coding sequences lodged and expressed by other classes of TE [50]. Thus, TEs persist as an agent of evolutionary change in spite of both the detrimental influence of TEs on the functioning of host genes and the largely harmful function of TEs as genes, owing to their ability to replicate themselves independently of cellular DNA replication. This kind of logic applies to many systems of selfish elements [47].

Conclusion and outlook

Deciphering gene function presents a long-standing puzzle to biologists of diverse stripes, from reductionists intrigued by the biochemical activities of proteins to those seeking maps of genotype to organismal phenotype. In most cases, we should anticipate evolution to depend intricately on the specific functional attributes of individual genes. But not always. Here we have summarized a variety of the ways in which evolutionary change can occur independently of, or in spite of, a given gene's functional activity. Most of these evolutionary processes are individually familiar, such as adaptation by loss of function and the interaction of selection and linkage, which we have integrated under a common theme. But how often is it the case that selection favors biochemical improvement of gene function versus degradation in adaptive responses? How pervasive is linkage in restraining per-gene responses to selection? How critical is the balance between gene duplication and gene degradation in genome evolution? And, how important is selection compared to genetic drift in leading to degradation of gene activity? Elucidating answers to these questions will help integrate the diverse ways in which selection shapes genomes, from favoring mutations that create novel functional attributes to mutations that degrade gene function.

Acknowledgments

We thank the constructive comments of S. Wright and C. Tsai on a previous version of the article and two anonymous reviewers for insightful suggestions. A.D. C. is supported by funds from the Natural Sciences and Engineering Research Council of Canada, a Canada Research Chair, and the National Institutes of Health (R01-GM096008).

The authors have declared no conflicts of interest.

References

1. **Jacob F.** 1977. Evolution and tinkering. *Science* **196**: 1161–6.
2. **Rich SS**, **Bell AE**, **Wilson SP.** 1979. Genetic drift in small populations of Tribolium. *Evolution* **33**: 579–84.
3. **Lynch M**, **Conery JS.** 2003. The origins of genome complexity. *Science* **302**: 1401–4.
4. **Barton NH.** 2010. Genetic linkage and natural selection. *Phil Trans Roy Soc Lond Series B: Biol Sci* **365**: 2559–69.
5. **Cutter AD**, **Payseur BA.** 2013. Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat Rev Genet* **14**: 262–74.
6. **Neher RA.** 2013. Genetic draft, selective interference, and population genetics of rapid adaptation. *Annu Rev Ecol Evol Syst* **44**: 195–215.
7. **Charlesworth B.** 2012. The effects of deleterious mutations on evolution at linked sites. *Genetics* **190**: 5–22.
8. **Kirkpatrick M**, **Barton N.** 2006. Chromosome inversions, local adaptation and speciation. *Genetics* **173**: 419–34.
9. **Andersen EC**, **Gerke JP**, **Shapiro JA**, **Crissman JR**, et al. 2012. Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nat Genet* **44**: 285–90.
10. **Rockman MV**, **Skrovanek SS**, **Kruglyak L.** 2010. Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. *Science* **330**: 372–6.
11. **Thomas CG**, **Wang W**, **Jovelin R**, **Ghosh R**, et al. 2015. Full-genome evolutionary histories of selfing, splitting and selection in Caenorhabditis. *Genome Res* **25**: 667–78.
12. **Anderson JT**, **Willis JH**, **Mitchell-Olds T.** 2011. Evolutionary genetics of plant adaptation. *Trends Genet* **27**: 258–66.
13. **Olson MV.** 1999. When less is more: gene loss as an engine of evolutionary change. *Am J Hum Genet* **64**: 18–23.
14. **Johanson U**, **West J**, **Lister C**, **Michaels S**, et al. 2000. Molecular analysis of FRIGIDA, a major determinant of natural fariation in Arabidopsis flowering time. *Science* **290**: 344–7.
15. **Hoekstra HE**, **Coyne JA.** 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61**: 995–1016.
16. **Smith SD**, **Rausher MD.** 2011. Gene loss and parallel evolution contribute to species difference in flower color. *Mol Biol Evol* **28**: 2799–810.
17. **Novembre J**, **Galvani AP**, **Slatkin M.** 2005. The geographic spread of the CCR5 Δ32 HIV-resistance allele. *PLoS Biol* **3**: e339.
18. **Grobet L**, **Royo Martin LJ**, **Poncelet D**, **Pirottin D**, et al. 1997. A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. *Nat Genet* **17**: 71–4.
19. **Doebley JF**, **Gaut BS**, **Smith BD.** 2006. The molecular genetics of crop domestication. *Cell* **127**: 1309–21.
20. **Cooper VS**, **Lenski RE.** 2000. The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* **407**: 736–9.
21. **Lang GI**, **Rice DP**, **Hickman MJ**, **Sodergren E**, et al. 2013. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* **500**: 571–4.
22. **Hottes AK**, **Freddolino PL**, **Khare A**, **Donnell ZN**, et al. 2013. Bacterial adaptation through loss of function. *PLoS Genet* **9**: e1003617.
23. **Moxon R**, **Bayliss C**, **Hood D.** 2006. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu Rev Genet* **40**: 307–33.
24. **Bachtrog D.** 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet* **14**: 113–24.
25. **van der Kooi CJ**, **Schwander T.** 2014. On the fate of sexual traits under asexuality. *Biol Rev* **89**: 805–19.
26. **Cutter AD.** 2008. Reproductive evolution: symptom of a selfing syndrome. *Curr Biol* **18**: R1056–R8.
27. **Ornduff R.** 1969. Reproductive biology in relation to systematics. *Taxon* **18**: 121–33.
28. **Slotte T**, **Hazzouri KM**, **Stern D**, **Andolfatto P**, et al. 2012. Genetic architecture and adaptive significance of the selfing syndrom in Capsella. *Evolution* **66**: 1360–74.
29. **Wright SI**, **Ness RW**, **Foxe JP**, **Barrett SCH.** 2008. Genomic consequences of outcrossing and selfing in plants. *Intl J Plant Sci* **169**: 105–18.
30. **Thomas CG**, **Li RH**, **Smith HE**, **Woodruff GC**, et al. 2012. Simplification and desexualization of gene expression in self-fertile nematodes. *Curr Biol* **22**: 2167–72.
31. **Fierst JL**, **Willis JH**, **Thomas CG**, **Wang W**, et al. 2015. Reproductive mode and the evolution of genome size and structure in Caenorhabditis nematodes. *PLoS Genet* **11**: e1005323.
32. **Agrawal AF**, **Whitlock MC.** 2011. Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics* **187**: 553–66.
33. **Orr HA**, **Betancourt AJ.** 2001. Haldane's sieve and adaptation from the standing genetic variation. *Genetics* **157**: 875–84.
34. **Orr HA.** 2005. The genetic theory of adaptation: a brief history. *Nat Rev Genet* **6**: 119–27.
35. **Conant GC**, **Birchler JA**, **Pires JC.** 2014. Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Curr Opin Plant Biol* **19**: 91–8.
36. **Force A**, **Lynch M**, **Pickett FB**, **Amores A**, et al. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531–45.
37. **Jovelin R**, **He X**, **Amores A**, **Yan YL**, et al. 2007. Duplication and divergence of fgf8 functions in teleost development and evolution. *J Exp Zool B Mol Dev Evol* **308**: 730–43.
38. **Demuth JP**, **Hahn MW.** 2009. The life and death of gene families. *BioEssays* **31**: 29–39.
39. **True JR**, **Haag ES.** 2001. Developmental system drift and flexibility in evolutionary trajectories. *Evol Devel* **3**: 109–19.

40. **Menashe I**, **Man O**, **Lancet D**, **Gilad Y.** 2003. Different noses for different people. *Nat Genet* **34**: 143–4.
41. **Maydan J**, **Lorch A**, **Edgley M**, **Flibotte S**, et al. 2010. Copy number variation in the genomes of twelve natural isolates of *Caenorhabditis elegans*. *BMC Genomics* **11**: 62.
42. **Wray GA**, **Hahn MW**, **Abouheif E**, **Balhoff JP**, et al. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* **20**: 1377–419.
43. **Verster AJ**, **Ramani AK**, **McKay SJ**, **Fraser AG.** 2014. Comparative RNAi screens in *C. elegans* and *C. briggsae* reveal the impact of developmental system drift on gene function. *PLoS Genet* **10**: e1004077.
44. **Barriere A**, **Gordon KL**, **Ruvinsky I.** 2012. Coevolution within and between regulatory loci can preserve promoter function despite evolutionary rate acceleration. *PLoS Genet* **8**: e1002961.
45. **Baker CR**, **Tuch BB**, **Johnson AD.** 2011. Extensive DNA-binding specificity divergence of a conserved transcription regulator. *Proc Natl Acad Sci USA* **108**: 7493–8.
46. **Schlenke TA**, **Begun DJ.** 2004. Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. *Proc Natl Acad Sci USA* **101**: 1626–31.
47. **Werren JH.** 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc Natl Acad Sci USA* **108**: 10863–70.
48. **Obbard DJ**, **Gordon KHJ**, **Buck AH**, **Jiggins FM.** 2009. The evolution of RNAi as a defence against viruses and transposable elements. *Phil Trans Roy Soc Lond Series B: Biol Sci* **364**: 99–115.
49. **Song J**, **Liu J**, **Schnakenberg SL**, **Ha H**, et al. 2014. Variation in piRNA and transposable element content in strains of *Drosophila melanogaster*. *Genome Biol Evol* **6**: 2786–98.
50. **Bennetzen JL**, **Wang H.** 2014. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu Rev Plant Biol* **65**: 505–30.