Positive and negative selection on the mitochondrial genome

Colin D. Meiklejohn, Kristi L. Montooth and David M. Rand

Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912, USA

Several recent studies have confirmed that mitochondrial DNA variation and evolution are not consistent with the neutral theory of molecular evolution and might be inappropriate for estimating effective population sizes. Evidence for the action of both positive and negative selection on mitochondrial genes has been put forward, and the complex genetics of mitochondrial DNA adds to the challenge of resolving this debate. The solution could lie in distinguishing genetic drift from ‘genetic draft’ and in dissecting the physiology of mitochondrial fitness.

Introduction

The neutral theory of molecular evolution predicts that the amount of genetic variation within a population will be positively correlated with the size of the population [1,2]. Large populations experience relatively less genetic drift and, as a result, retain more segregating alleles than small populations where new mutations are easily lost. In a recent study, Bazin et al. [3] compiled published animal mitochondrial (mt)DNA sequences and observed that the average amount of mitochondrial nucleotide polymorphism was remarkably invariant across taxa spanning a wide range of inferred population sizes. By contrast, a smaller set of nuclear-encoded sequences showed the expected positive relationship between average genetic diversity and population size. Bazin et al. [3] concluded that mtDNA might, in fact, be a poor indicator of population size, and suggested that frequent episodes of natural selection are the cause of this discrepancy [3]. Specifically, they invoke an evolutionary process that has been termed ‘genetic draft’ [4] (Box 1 and Glossary) to explain their observation. According to the draft hypothesis, the reduction in variation caused by recurrent selective sweeps balances the expected greater levels of neutral diversity in larger populations, making variation independent of population size [4]. The conclusions of Bazin et al. [3] challenge the utility of mtDNA variation as a reliable indicator of population size and, consequently, this suggestion has generated some debate in the literature [5–8].

Positive selection on mitochondrial genes?

If genetic draft is responsible for the observed independence of mtDNA polymorphism from population size, this should leave the signature of positive selection in the mtDNA of species with larger population sizes. Bazin et al. [3] tested this prediction by considering the ratio of nonsynonymous to synonymous fixations between species ($d_{S}/d_{S}$) in whole mitochondrial genomes of taxa with small (vertebrates) versus large (invertebrates) population sizes. Although all $d_{S}/d_{S}$ ratios were well below one, indicating that purifying selection is the predominant force in mtDNA evolution, the $d_{S}/d_{S}$ values were on average higher in invertebrates, consistent with the idea that more amino acid fixations have been adaptive in these species. However, this is also consistent with a hypothesis of more relaxed selective constraint on amino acid substitutions in invertebrates than in vertebrates.

Bazin et al. [3] also combined mtDNA polymorphism and divergence data in McDonald–Kreitman (MK) tests [9]. The MK test compares the ratio of amino acid replacement variation to synonymous variation within species with the ratio of replacement to synonymous divergence between species. Under neutrality these two ratios should be equal. The direction and degree of departures from this neutral prediction can be quantified with the neutrality index, $N_{I}$ [10]. Assuming that synonymous sites are evolving neutrally, $N_{I} < 1$ indicates an excess of amino acid divergence, whereas $N_{I} > 1$ indicates an excess of amino acid polymorphism. The median $N_{I}$ was significantly lower in invertebrate species than in vertebrates, again suggesting that species with larger population sizes are more prone to positive selection on mtDNA. The authors cite this as evidence that mitochondrial genomes in species with large population sizes are subject to selective sweeps, which are required to balance the expected greater levels of neutral diversity in larger populations, making variation independent of population size [4].

Glossary

- $d_{S}/d_{S}$: the ratio of nonsynonymous (amino acid-altering) changes per nonsynonymous site to synonymous changes per synonymous site. This ratio is used to infer the type of evolutionary pressure acting on a protein: when $d_{S}/d_{S} > 1$ positive Darwinian selection is inferred; when $d_{S}/d_{S} < 1$, purifying selection is inferred.
- Genetic draft: the reduction in nucleotide diversity in a genomic region linked to a locus that experiences recurrent fixation of beneficial mutations that sweep through the population. Draft has the counterintuitive effect that, as population size increases, genetic diversity decreases and deleterious fixation events increase relative to the neutral expectation.
- Genetic hitchhiking: the effect of evolutionary processes at one locus on a linked locus.
- McDonald–Kreitman (MK) test: a test of the neutral theory of molecular evolution for protein coding regions. This test compares the ratio of nonsynonymous to synonymous changes within species with the ratio of nonsynonymous to synonymous changes between species.
- Purifying selection: natural selection that removes deleterious variation from populations.
- Selective sweep: natural selection that increases the frequency of beneficial mutations in populations; most clearly demonstrated by a $d_{S}/d_{S}$ ratio > 1.

Corresponding author: Rand, D.M. (David_Rand@brown.edu).
Available online 5 April 2007.
frequent adaptation, with the consequence that polymorphism is reduced relative to the neutral expectation through genetic draft.

These results contrast with several previous studies that reported neutrality indices for mtDNA indicative of an excess of amino acid polymorphism ($NI > 1.0$), which is consistent with the hypothesis that most segregating functional variants are deleterious and rarely contribute to divergence [11–13]. Furthermore, a recent study compared complete mtDNA sequences in sexual and asexual lineages of *Daphnia* [14] and found higher $d_S/d_S$ in the asexual lineages. Asexual reproduction decreases the genetic effective population size, which reduces the efficacy of natural selection [15]. If relaxed selection associated with reduced effective population size increases amino acid substitution rates, this suggests that most of these substitutions are deleterious.

Why did Bazin *et al.* [3] find evidence for positive selection when previous studies consistently support purifying selection on mtDNA? The original studies that documented purifying selection had more vertebrate samples than invertebrates, [11,12] and those cases with $NI < 1$ were often insects or other invertebrates [13], so recognizing this pattern might have required Bazin’s large meta-analysis [3]. However, the results of Ref. [3] might also be influenced by peculiarities of the neutrality index and by biases in the genes chosen by researchers for mtDNA sequence analysis in vertebrates versus invertebrates.

In a note on Bazin’s study [3], Wares *et al.* [5] argue that $NI$ becomes increasingly biased below 1.0 when distant outgroups are used in MK tests, as such outgroups can result in an underestimation of synonymous fixed differences. However, Bazin *et al.* [6] point out that the

![Figure 1](https://www.sciencedirect.com)

**Figure 1.** Gene- and taxon-dependent heterogeneity in mitochondrial evolution. Box plots of Neutrality index ($NI$) scores from McDonald–Kreitman tests of mitochondrial genes in vertebrates (Vert) and invertebrates (Inv) reveal gene-specific patterns. Data are from the supplementary data of Bazin *et al.* [3] and are presented for (a,d) all genes and for (b,c) cytochrome oxidase subunits (COX), (e,f) the cytochrome $b$ gene (CYTB) and (c,d) NADH dehydrogenase subunits (ND). $NI$ values $< 1$ suggest positive selection; $NI > 1$ suggests negative selection. (a–d) All data, including species for which $NI = 0$; (e–h) the results excluding cases with $NI = 0$. The pattern of consistent positive selection (a–d) disappears when cases of $NI = 0$ are omitted (e–h). As recent positive selection and strong functional constraint are both likely causes of $NI = 0$, the distributions of $NI$ that reflect the true selective forces acting on mitochondrial genes must lie somewhere in between. The persistence of the median $NI < 1$ among invertebrate ND datasets in the lower panel is also strong support for positive selection. Thick line, median; box, 25–75% quartiles; error bars, range; the width of the box is proportional to sample size. $NI$ values $> 20$ were forced to 20. $P$-values are from Wilcoxon rank sum tests for differences in $NI$ between vertebrates and invertebrates.
average interspecific divergence was almost identical for invertebrates and vertebrates, and they consider it unlikely to be the cause of the differences in median NI between these groups.

If no amino acid polymorphisms are observed in a sample, NI is zero; and if no amino acid differences between species are observed, NI is undefined. Although a recent selective sweep could rob a population of all its replacement polymorphisms, small sample sizes or strong purifying selection can also produce this result. In Bazin’s study [3], more than a third of the datasets produced NI = 0 (40% of invertebrate datasets and 24% of vertebrate datasets). The higher incidence of NI = 0 in invertebrate datasets might result in the appearance of stronger positive selection in these taxa, if these zeros are caused by something other than selective sweeps. If datasets with NI = 0 are excluded, the distributions of NI among both vertebrates and invertebrates broadly overlap neutrality, with the median NI in vertebrates now >1 (Figure 1e–h).

Previous studies have demonstrated significant heterogeneity in rates of sequence evolution across the different protein complexes encoded by the mtDNA [16] (Montooth et al. unpublished). In vertebrates and invertebrates, cytochrome oxidase subunit (COX) genes in general and COXI specifically seem to have lower rates of divergence than the cytochrome B (cytB) gene, which suggests that there is greater functional constraint on COX proteins. Stronger functional constraint might result in more datasets with no replacement polymorphisms, which would produce NI = 0 in the absence of positive selection. Consistent with this, 35% of COXI datasets versus 20% of cytB datasets show NI = 0, and excluding such datasets has the greatest impact on the distribution of NI values from COX in invertebrates (see Figure 1 and the supplementary material online). In Bazin’s study [3], 3% and 74% of vertebrate sequences came from the COXI and cytB genes, respectively, whereas for invertebrates taxa these numbers are 64% and 8%. This differential use of mitochondrial genes to study invertebrate and vertebrate taxa might contribute to the observation of low NI values among invertebrates.

However, even excluding cases of NI = 0, the assemblage of studies using the ND genes (which encode subunits of electron transport chain complex I, NADH dehydrogenase) shows evidence for positive selection in invertebrates (NI < 1; Figure 1) and neutrality in vertebrates. As some datasets might have NI = 0 due to a recent sweep, excluding these cases produces a bias against detecting positive

**Box 1. Genetic draft**

Genetic draft [4] is a result of the hitchhiking process associated with positive selection acting on beneficial mutations. Because positive selection moves favored alleles through a population much more rapidly than genetic drift, neutral variants linked to a beneficial allele hitchhike to high frequency, resulting in a decrease in polymorphism in the region of the genome surrounding the favorable mutation [26] (Figure 1). The number of advantageous mutations that appear in a population is a product of the beneficial mutation rate and the population size. If the fixation rate of beneficial mutations is limited by their appearance, then larger populations will fix more of them. As a consequence, loci in organisms with large population sizes will, on average, have experienced a sweep more recently than those in small populations, and should harbor less genetic variation relative to the neutral expectation. However, our understanding of the effects of selection on linked neutral variation depends critically on precisely modeling the dynamic processes of selection, drift and recombination. For example, the behavior of a new beneficial allele is governed by both selection and drift until it reaches a high enough frequency for selection to take over; this initial phase can have important consequences for the structure of linked variation following the sweep [27,28]. These and other elaborations [29] on the original descriptions [30–32] of selective sweeps will undoubtedly influence the specifics and significance of genetic draft as we learn more about these population genetic processes.

**Figure 1**. Genetic drift and genetic draft have different effects on levels of neutral variation in large versus small populations. (a) Neutral mutations that are destined for fixation (black lines) enter a population and drift to fixation at a rate inversely proportional to population size. Because of this, expected levels of genetic diversity (dashed lines) are higher in larger populations. (b) Under genetic draft, beneficial alleles (red) sweep to fixation and purge neutral variation at linked sites (blue). On average, this reduces nucleotide variation below the neutral expectation. This reduction, indicated by the arrow and the solid black horizontal line, is greater in larger populations as a result of more frequent sweeps, leading to levels of neutral variation that are independent of population size.
selection, which makes this inference of adaptation in the ND complex in invertebrates even more compelling. These considerations indicate that there is support for the major finding of Bazin et al. [3] concerning mitochondrial adaptation, but they also highlight the difficulties of explaining the independence of polymorphism from population size using the neutrality index alone. Moreover, the different patterns of selection in the COX and ND genes suggest that the physiological connection between mtDNA sequence and fitness differs between protein complexes, possibly in taxon-specific ways.

Caught in the draft: slightly deleterious mutations and selective sweeps

There are several ways in which evolution without frequent positive selection on mtDNA variants could nonetheless produce a constant level of mitochondrial polymorphism. First, recent theoretical work has demonstrated that certain population models can generate a decoupling of variation and population size in the absence of selection [17]. It remains to be seen whether such processes are good models for mtDNA evolution. Second, the excess amino acid divergence between species could be the result of genetic draft fixing segregating deleterious mutations. Under draft, the rate of fixation of deleterious variants is an increasing function of population size [18], which could account for the elevated $d_S/d_S$ seen in invertebrates. This would require periodic selective sweeps to generate the genetic draft; if replacement variants are deleterious, what produces the positive selection coefficient that drives the sweep? Bazin et al. [3] suggest that perhaps only a fraction of replacement substitutions have been directly selected, and the rest have hitchhiked to fixation. However, the most intriguing possibility they raise is that mitochondrial sweeps result from cytoplasmic factors, such as maternally transmitted symbionts. Wolbachia and other intracellular parasites are known to cause cytoplasmic sweeps in nature [19]; such endosymbionts are also prevalent in the two groups Bazin et al. [3] identify as having the highest mtDNA $d_S/d_S$, insects and nematodes [20], and are not known to infect vertebrates. Indeed, one noteworthy study [21] contrasted a species of Drosophila that harbors Wolbachia with a closely related uninfected species, and found that the mtDNA of the infected species showed much lower diversity as well as an elevated $d_S/d_S$.

Concluding remarks

There is clearly a great deal to be learned about how selection shapes mitochondrial genome evolution. Mitochondrially encoded proteins evolve much more slowly than predicted by the high mutation rate in the mitochondrial genome [22], so purifying selection on nonsynonymous mutations is clearly an important component of mitochondrial evolution. However, Bazin et al. [3] force a careful evaluation of the role of positive selection. If selective sweeps are so common in mtDNA, why does it consistantly have higher levels of nucleotide polymorphism than nuclear DNA [3,23], in which a reduction in polymorphism as a result of genetic draft can be seen in regions of low recombination [24,25]? In cases in which mtDNA variation is low, cytoplasmic agents are probable causes of the reduced polymorphism [19], again consistent with genetic draft resulting from sources other than positive selection on mitochondrial variants directly. Finally, mtDNA variation in mammals does scale with population size [7], so the effect of genetic draft might be restricted to species with very large effective population sizes.

Acknowledgements

This work was supported by NIH grants GM072399 to C.D.M., GM076812 to K.L.M., and GM067862 to D.M.R. and NSF grant DEB 0108500 to D.M.R. We thank four anonymous reviewers whose comments greatly improved this manuscript.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2007.03.008.

References

21 Shoemaker, D.D. et al. (2004) Decreased divergence but increased substitution rate in host mtDNA as a consequence of Wolbachia endosymbiont infection. Genetics 168, 2049–2058
More is not always better: the genetic constraints of polyploidy

Peter H. Thorpe, Sergio González-Barrera and Rodney Rothstein

Department of Genetics & Development, Columbia University Medical Center, 701 West 168th Street, New York, NY 10032-2704, USA

Polyploidy

Most prokaryotes and some unicellular eukaryotes contain a single haploid genome. By contrast, many eukaryotes are diploid, which is the standard for sexually reproducing species. At the extreme, certain organisms are 'gluttons' for genetic information, having adapted to maintain multiple copies of their genome. Examples of these polyploid organisms include many plant species; however, polyploid vertebrates are rare. Aside from the established benefits of sexual reproduction, duplicate genomes can confer an evolutionary advantage over haploids because redundant genes can mutate without detriment to the organism. Whole-genome duplication has had an important role in the evolution of diverse species such as bacteria, plants and primates [1–3]. Indeed, there is evidence that such an event occurred recently in one species of mammal [4], although this remains controversial [5].

Human tetraploid cells

In humans, tetraploid cells, which contain four sets of chromosomes, are found in both germline and somatic cells. Germline tetraploids are rare and are believed to arise either by mitotic failure early in embryogenesis or through cell fusion, and such tetraploidy is typically lethal [6]. In addition, triploidy, which occurs in 1 in 10 000 live births, is the result of either fertilization of an egg with two sperm (dispermy) or fusion of the ovum with a polar body and subsequent fertilization by a haploid sperm [7]. In contrast to these infrequent germline events, somatic genome duplication is common in humans. Tissues with a high regenerative capacity, such as the liver, muscle, brain, placenta and certain bone marrow cells, have a higher incidence of polyploidy [8–12]. Perhaps genome multiplication is an adaptive response in normally non-proliferating tissues (such as the heart), as a buffer against oxidative stress and genotoxic damage. However, under different circumstances, polyploid cells seem to be at a disadvantage; they show increased chromosome instability and DNA repair defects [13,14]. Furthermore, cancer cells are commonly aneuploid, and there is evidence that this karyotype can arise through a polyploid intermediate [15]. Indeed, polyploidy might drive cellular transformation [16]. In addition, aging cells have a higher incidence of polyploidy, and the higher incidence of polyploidy in aging men correlates with their increased mortality compared with women [17].

A genetic screen for genes essential in tetraploids

Recently, Storchová et al. [18] undertook a systematic approach to identify genes that are essential for the viability of yeast polyploid cells. They used diploid strains created from the Saccharomyces cerevisiae gene disruption library, containing defined single-gene deletions, to mate together to form tetraploids (Box 1). Approximately 1% of the genes screened (39 in total) were sensitive to tetraploidy, and these fall into three functional groups. First are genes that encode the mitotic spindle, including components of the spindle pole body (SPB, the yeast microtubule organizing center; Figure 1a). The second group contains genes involved in chromosome cohesion.