

# Genetic hitchhiking, mitonuclear coadaptation, and the origins of mt DNA barcode gaps

Geoffrey Hill<sup>1</sup>

<sup>1</sup>Auburn University

July 2, 2020

## Abstract

DNA barcoding based on mitochondrial (mt) nucleotide sequences is an enigma. Neutral models of mt evolution predict DNA barcoding cannot work for recently diverged taxa, and yet, mt DNA barcoding accurately delimits species for many bilaterian animals. Meanwhile, mt DNA barcoding often fails for plants and fungi. I propose that because mt gene products must cofunction with nuclear gene products, the evolution of mt genomes is best understood with full consideration of the two environments that impose selective pressure on mt genes: the external environment and the internal genomic environment. Moreover, it is critical to fully consider the potential for adaptive evolution of not just protein products of mt genes but also of mt transfer RNAs and mt ribosomal RNAs. The tight linkage of genes on mt genomes that do not engage in recombination could facilitate selective sweeps whenever there is positive selection on any element in the mt genome, leading to the purging of mt genetic diversity within a population and to the rapid fixation of novel mt DNA sequences. Accordingly, the most important factor determining whether or not mt DNA sequences diagnose species boundaries may be the extent to which the mt chromosomes engage in recombination.

## Introduction

Bilaterian animals carry mitochondrial (mt) genotypes that typically show little variation among individuals within a species but with substantial variation between species (Dasmahapatra and Mallet 2006; Lane 2009b; Bucklin et al. 2011; Zahiri et al. 2014). There is some debate regarding how reliably mt DNA genotype diagnoses species, particularly closely related species (Dupuis et al. 2012; Ratnasingham and Hebert 2013), but for birds (Tavares and Baker 2008), mammals (Clare et al. 2007), turtles (Naro-Maciel et al. 2010), boney fish (Ward and Holmes 2007), amphibians (Vences et al. 2005), spiders (Coddington et al. 2016), butterflies (Janzen et al. 2009), ants (Smith et al. 2005), parasitoid wasps (Smith et al. 2008), mayflies, stoneflies, and caddisflies (Morinière et al. 2017), among other bilaterian taxa, mt genotypes change abruptly at the great majority of species boundaries. This pattern of differentiation in mt genes among species has led to the use of mt nucleotide sequences as a diagnostic tool in species identification, the so called “DNA barcode”(Hebert et al. 2003a). The substantial divergence in mt DNA sequence observed between most sister pairs of animal taxa is termed the “barcode gap” (Hebert et al. 2003a).

Three mechanisms have been proposed to explain this pattern of diversity of mt genotypes: (1) variations in mt DNA nucleotide sequences are neutral and are fixed via drift in isolated populations (Moritz et al. 1987; Hickerson et al. 2006; Lynch et al. 2006; Zink and Barrowclough 2008; Smith 2016), (2) there have been repeated episodes of extreme population bottlenecks involving the majority of bilaterian taxa (Stoeckle and Thaler 2014) or (3) directional selection on mitochondrial genotypes leads to rapid divergence when gene flow between populations is disrupted (Gershoni et al. 2009; Chou and Leu 2010; Hill 2016; James et al. 2016). In this essay, I focus on the necessity of coadaptation with the nuclear (N) genome throughout the evolution of the mt genome as a foundation for explaining the population structure of mt genomes. I propose that, far from being an unexpected or inexplicable pattern, the tight congruence between mt genotype and

species boundaries may be an inevitable consequence of the need for mt and N gene products to cofunction to enable aerobic respiration, especially when the mt chromosome does not engage in recombination. In reviewing previous efforts to explain DNA barcode gaps, I consider the nearly exclusive focus on amino acid substitutions and the protein-coding genes of the mt genome as potential targets of selection, which has discounted the potential key role played by selection on genes coding for tRNA and rRNA as well as selection on origin of replication regions of mt DNA (Ruiz-Pesini and Wallace 2006; Ellison and Burton 2008a; Barreto and Burton 2013; Adrion et al. 2016; Barreto et al. 2018). I also consider a potential pivotal role played by recombination of mitochondrial chromosomes in the generation of mt DNA barcode gaps. I propose that a better understanding of the evolutionary mechanism that generates the genetic structure of mt DNA across eukaryotes is critical not only with regards to assessing the value of DNA barcodes as a tool in taxonomy (Rubinoff et al. 2006; Baker et al. 2009) but also for a better understanding of the process of speciation (Hill 2016; Sunnucks et al. 2017; Tobler et al. 2019).

## Neutral Models

It was long assumed that the great majority of the evolution of mt genomes was neutral and hence that genetic structure of mt DNA within and among populations was necessarily the product of drift (Ballard and Kreitman 1995; Avise 2004; Lynch et al. 2006). The assumption of neutrality in changes to mt genotypes emerges from the recognition that all protein-coding genes in the animal mt genome code for subunits of the electron transport system and therefore that the protein products of the mt genome are among the most system-critical proteins in the entire animal genome (Lane 2011; Bar-Yaacov et al. 2012). Functional changes to such mission-critical genes was proposed to be so rare as to be realistically ignored, leaving the assumption that observed evolutionary changes in the mt genome will be neutral (Saccone et al. 2000). The rapid coalescence of mt genotype compared to N genotype in populations of eukaryotes was proposed to arise as a simple consequence of the small effective population size of the mt genome in relation to the N genome—a result of the mt genome being haploid and maternally transmitted (Palumbi et al. 2001; Hickerson et al. 2006; Zink and Barrowclough 2008).

Arguments for using mt DNA as a neutral marker of evolution rested on the assumption that essentially all selection on mitochondrial genotypes would be in the form of purifying selection to maintain the current forms of mt-encoded proteins with no functional change in gene products and with no functional variation between groups (Rand et al. 1994; Stewart et al. 2008). Synonymous changes to the nucleotide sequence, which are defined as changes that do not affect the amino acid sequence of a protein, were predicted to evolve via genetic drift and thus to accumulate across evolutionary time at a rate proportional to population size and mutation rate (Wilson et al. 1985; Lynch et al. 2006; Stoeckle and Thaler 2014). However, fundamental predictions of the neutral hypothesis for mitochondrial evolution have not been supported. Neutral theory predicts that genetic variation within a population should be proportional to the size of that population. Contrary to this prediction, there is no consistent relationship between population size and variation in mt DNA sequence (Bazin et al. 2006; Nabholz et al. 2009; Stoeckle and Thaler 2014). Moreover, the fixation of distinct mt genotypes between populations of at least some vertebrates (for which the rates of mutation of mt DNA are fairly well characterized) seems to occur much faster than predicted by neutral theory (Ballard and Whitlock 2004; Hickerson et al. 2006). And finally, in contradiction to neutral theory, isolation by distance is unreliable for mt DNA (Teske et al. 2018). All things considered, neutral theory does not seem like the place to begin in an investigation of the evolution of mt DNA and the origins of the mt DNA barcode gap (Kern and Hahn 2018).

## Demographic Bottlenecks

In a recent essay, Stoeckle and Thaler (2014) posed the question: “A universal selection-driven mtDNA clock implies all organisms are evolving at about the same rate . . . . What could cause similar rates of change for diverse organisms in diverse environments?”. Stoeckle and Thaler (2014) proposed that the external environment of organisms could cause periodic extreme reductions in the population sizes of essentially all organisms, perhaps particularly at the point of divergence of incipient sister species. Severe demographic bottlenecks in population size would purge populations of genetic diversity in mt genotype and fix differences

between species, potentially creating the pattern of mt DNA barcode gaps observed in bilaterian animals (Stoeckle and Thaler 2014). No explanation is given for why such bottlenecks would reduce variation in mt genotypes but not N genotypes. Moreover, this hypothesis requires that, at regular intervals that average a few hundred thousand years, essentially every species is subjected to an extreme population bottleneck (Stoeckle and Thaler 2014). These authors speculated that complex “food web, predator-prey, and parasite-host interactions” might sum to a common selective pressure on animal mitochondrial genomes “with long-term planetary climate cycles as the ultimate driver of evolution” (Stoeckle and Thaler 2014). I know of no evidence for such periodic synchronized collapse of all populations of all bilaterian organisms to create the pattern of mt DNA barcode gaps observed in bilaterian animals.

### Change in mt DNA Genotype via Selection

An alternative hypothesis to both neutral drift and demographic bottlenecks for the generation of mt DNA barcode gaps between species is directional selection on mt genotypes. Natural selection has the potential to shape the mt genome in response to two distinct environments: the external environment (both biotic and abiotic) and the internal genomic environment created by the N genome (Rand et al. 2004; Zhu et al. 2014; Barreto et al. 2018; Sloan et al. 2018; Hill 2019a). There is now overwhelming evidence that the mt genome of at least some animal lineages—and likely all animal lineages—are subject to periods of directional selection as adaptive responses to the external environment (Dowling et al. 2008; Ballard and Pichaud 2014; Kazancioğlu and Arnqvist 2014). In particular, thermal and chemical environments, oxygen pressure, diet, salinity, and UV exposure can all exert natural selection on the mt genome and lead to adaptive changes in protein-coding genes (Ballard and Pichaud 2014; Hill 2019a). The adaptive evolution of mt genomes in response to external environments is now a major research topic in evolutionary biology (Sunnucks et al. 2017; Hill et al. 2019), and such changes to the nucleotide sequence of mitochondria in response to directional selection pose a serious challenge to core arguments for why mt DNA sequences will often fail as a tool for diagnosing species (Hickerson et al. 2006). Adaptive divergence of mt genotype in response to external environment is a key reason why mt DNA is predicted to rapidly diverge between allopatric populations (Gershoni et al. 2009; Tobler et al. 2019).

Perhaps even more important, and certainly more pervasive, than changes to mt DNA gene sequence in response to external environment is the potential for perpetual evolutionary change in the mt DNA in response to changes in the internal genomic environment (Chou and Leu 2010; Burton and Barreto 2012; Barreto et al. 2018; Sloan et al. 2018; Hill 2020). The coadaptation of gene complexes is a foundational concept in evolutionary biology (Dobzhansky 1937; Wright 1942). In a discussion of the evolution of mt genomes, however, it is essential to grasp that there are unique features to the co-evolution and coadaptation between mt gene products and the products of a small list of N genes that code for products that function in intimate interaction with mt gene products (N-mt genes) (Hill 2019a; Shtolz and Mishmar 2019). First, the system that depends on coadaptation of mt and N-mt genes—the electron transport system—is the most critical biochemical system in the bodies of eukaryotes that depend on energy from aerobic respiration (Wallace 2010; Lane 2014). Second, because of the complexity of the ETS in controlling the flow of electrons and pumping of protons, very small changes to interacting components can have huge fitness effects (Lane 2011; Sloan et al. 2018; Hill 2019a; Hill et al. 2019). Third, mitonuclear coadaptation involves two genomes that can potentially undergo independent evolution (Rand et al. 2004; Gershoni et al. 2014; Wolff et al. 2014). Fourth and finally, the mt genome of animals does not generally engage in recombination (Barr et al. 2005) and so mitochondrial genes form one linkage group such that selection on one mt gene can affect the frequencies of other mt genes (Meiklejohn et al. 2007; Oliveira et al. 2008). Functional divergence in mt DNA will be particularly effective in creating Dobzhansky-Muller incompatibilities in hybrid offspring and hence in establishing barriers to gene flow because the mt DNA must maintain tight coadaptation with the N genome (Burton and Barreto 2012; Hill 2017).

If changes in mt genotype between species were entirely neutral, then matching the N genes of one species with the mt genes of a closely related species—either through hybridization or in cell culture by directly manipulating genomes—should result in no change in mitochondrial function in the resulting cells or orga-

nisms. Indeed, this logical extension of the neutral theory of mitochondrial evolution led to a failed research program to propagate endangered species by pairing mitochondria of donor species to the N genome of the species to be saved (Lanza et al. 2000). Observations from cybrid and hybrid studies, however, clearly established that, once sets of mt and N-mt genes diverge in nucleotide sequences to the extent seen in sister species, incompatibilities in non-coadapted gene sets cause a reduction in mitochondrial function when they are forced to work together (reviewed in Hill (2019a)). Mitonuclear incompatibilities in cybrid cells and hybrid organisms is strong evidence that the evolution of mt genotypes is not neutral with respect to the genomic environment (Barrientos et al. 1998; Ellison and Burton 2008b; Lee et al. 2008; Garvin et al. 2011; Latorre-Pellicer et al. 2016).

The evolution of uniquely coadapted mt and N-mt genotypes is a critical concept because it potentially explains both how the mt genotypes of sister species rapidly diverge and why there is so little introgression of divergent mt genotypes between species within most clades of bilaterian animals (Burton and Barreto 2012; Hill 2016). The evolution of a clean mt DNA barcode gap requires that the propagation of population-specific mitochondrial genotypes are constrained to remain within species boundaries across generations (Hebert et al. 2003b). Even a small amount of introgressive flow of mitochondrial genotypes, which would be inevitable under neutral models of mitochondrial evolution if species lived in sympatry, would add unacceptable ambiguity into barcoding efforts (Papadopoulou et al. 2008). In the rare cases in which mitochondria do introgress across species boundaries, the introgression tends to be rampant, with complete replacement of one mitochondrial genotype by another (Hill 2019b). All of these patterns are consistent with a process whereby coevolution of mt and N-mt genotypes leads to loss of fitness (at the level of the individual organism) when mt genotypes are paired to N-mt genes to which they are not coadapted. The barcode gap is more than an arbitrary marker of species boundaries—it is the functional boundary that reinforces the uniqueness of a species' mitonuclear genotype (Lane 2009a; Chou and Leu 2010; Burton and Barreto 2012; Hill 2016, 2017).

### **The Gene Content of a Barcode Gap**

The pattern of little variation within a species but substantial variation between species is the reason that DNA barcoding is proposed as a useful tool for taxonomists (Meyer and Paulay 2005). But what, specifically, are the fixed differences in nucleotide sequences that create barcode gaps? In vertebrates, including birds (Kerr 2011), mammals (Tobe et al. 2010), and fish (Ward and Holmes 2007), variation in amino acid sequence is rare in the barcoding region of the cytochrome c oxidase subunit one (COX1) gene. Thus, the barcode gap that is commonly observed using the conventional COX1 barcode gene (Kerr et al. 2007; Tavares and Baker 2008) is comprised almost entirely of synonymous nucleotide changes, and there is evidence for strong purifying selection on the nonsynonymous nucleotide positions within the COX1 barcode gene (Stewart et al. 2008; Kerr 2011; Popadin et al. 2013). In contradiction to the prediction that adaptive evolution of the COX1 gene might underlie the evolution of DNA barcode gaps (Hill 2016), there is too little variation in the amino acid sequence of the product of the COX1 gene between sister taxa for this prediction to be correct (Kerr 2011). The paradox of the COX1 barcode gene is that, despite departure from expectations of neutral theory, there seems to be little opportunity for adaptive divergence creating the differences among species in the nucleotide sequence of the COX1 barcode gene (Kwong et al. 2012). Certainly there are a handful of very well documented cases of COX1 adaptively diverging between sister taxa in response to changes in the oxygen pressure (Scott et al. 2011; Luo et al. 2013; Tomasco and Lessa 2014) or hydrogen sulfide exposure (Pfenninger et al. 2014) in the external environment. Such adaptive divergences in COX1 genotype, however, cannot account for the barcode gap that has been documented between thousands of sister taxa.

A paucity of non-synonymous changes in the barcode region of the COX1 gene is not difficult to explain. COX1 is the least changeable gene in the entire mitochondrial genome (da Fonseca et al. 2008; Kerr 2011). The conserved nature of COX1 is a major reason that it was chosen as the barcode gene: primer sets developed for model species tend to work for non-model species (Hebert et al. 2003a). COX1 is one of thirteen protein subunits of Complex IV of the ETS, which is the rate-controlling enzyme in the OXPHOS system (Pacelli et al. 2011; Arnold 2012), and COX1 holds the key catalytic position of that crucial enzyme

(Wang and Pollock 2007). Thus, Complex IV is a particularly critical enzyme in animal systems that depend on aerobic respiration, and the barcode gene, COX1, is the most critical subunit of this most critical enzyme (Pierron et al. 2012). I propose that the resolution of this paradox of species-specific variation in the COX1 barcode sequence without functional changes in the COX1 gene lies in the tight linkage of genes on the mitochondrial chromosome, and genetic hitchhiking of neutral substitution in the COX1 barcoding region with adaptive changes in other regions of the mt genome (Meiklejohn et al. 2007).

### **Synonymous/nonsynonymous vs functional/nonfunctional**

For most pairs of sister species that recently evolved from a common ancestor and now have a DNA barcode gap, there is no difference in the amino acid sequence of the portion of the COX1 gene serving as the barcode gene. However, this does not necessarily mean that there are no functional changes to other protein-coding genes that include seven subunits of Complex I, one subunit of Complex III, two (additional) subunits of Complex IV, and two subunits of Complex V. Indeed, the seven mitochondrially encoded protein subunits of Complex I are much more frequently implicated in adaptive divergences between sister taxa than Complex IV subunits (da Fonseca et al. 2008; Garvin et al. 2014). At least some sister taxa also carry fixed differences in amino acid sequence for subunits of Complexes III and V (reviewed in Hill (2019a)). A hypothesis that is worthy of testing is that the pattern of little variation within species but substantial differences between species in mt DNA sequence arises entirely as a consequence of strong selection on adaptive amino acid substitutions in mt-encoded proteins (da Fonseca et al. 2008). Given available data, however, I do not think that an adaptive protein evolution hypothesis will be the primary solution to the paradox of the mt DNA barcode gap, because purifying selection is, indisputably, the dominant force in the evolution of all mt protein-coding genes (Stewart et al. 2008; Kerr 2011).

I propose that the key to explaining the evolution of the mt DNA barcode gap lies in giving full consideration to the fact that all of the genes encoded by the animal mitochondrial genome will evolve via natural selection primarily in response to the internal genomic environment (Sunnucks et al. 2017; Sloan et al. 2018; Hill et al. 2019). Most of the genes in the mitochondrial genome code for products other than proteins; in bilaterian animals, 24 out of 37 mitochondrial genes (65%) code for transfer RNA (tRNA) or ribosomal RNA (rRNA) (Rand et al. 2004; Burton and Barreto 2012). Every one of these genes maintains coadaptation with the N-encoded genes through coevolution; in other words, there is a prediction of perpetual directional adaptive evolution of all of the products of the mt genome in response to the internal genomic environment (Kivisild et al. 2006; Hill 2019a; Wei et al. 2019; Zaidi and Makova 2019). The selective driver of this process of adaptive evolution of mt genes is compensatory coevolution, whereby mt genes evolve so as to compensate for deleterious N genotypes and vice versa (Rand et al. 2004; Dowling et al. 2008; Osada and Akashi 2012; Barreto et al. 2018; Hill 2020). Even the non-coding region of the animal mitochondrial RNA, which serves as the origin of replication site for transcription and replication, coevolves with N genes (Gaspari et al. 2004; Ellison and Burton 2008a, 2010). The expression of mt and N genes that code for co-functioning units must also be co-regulated, another important level of mitonuclear coadaptation (Barshad et al. 2018; Calvo et al. 2019).

There is a large and rapidly growing literature showing that single nucleotide substitutions in each of the non-protein-coding genes of the animal mitochondrion has important fitness consequences (reviewed in Hill 2019). Many of these fitness effects play out in relation to the external environment of the organism (Hoekstra et al. 2013), but the source of the hypothesized perpetual evolutionary change of all of the products of the mt genome would be selection to maintain coadaptation with products of the N genome to enable cellular respiration (Meiklejohn et al. 2013; Barreto et al. 2018; Hill 2020). Because it is dependent on random mutations in both the N and mt genomes, coevolution of co-functioning mt and N genes to maintain mitochondrial function will be idiosyncratic, unpredictable, and not repeatable (Blount et al. 2018). Directional selection on both the mt and N genomes to maintain mitonuclear coadaptation will create the sort of divergence in mt genotypes between species that give rise to a DNA barcode gap (Burton and Barreto 2012; Hill 2016). The key missing element is: how would divergence in a tRNA, rRNA, or the control region affect the barcode region of the COX1 gene or other synonymous substitution in protein coding genes?

## Genetic Hitchhiking

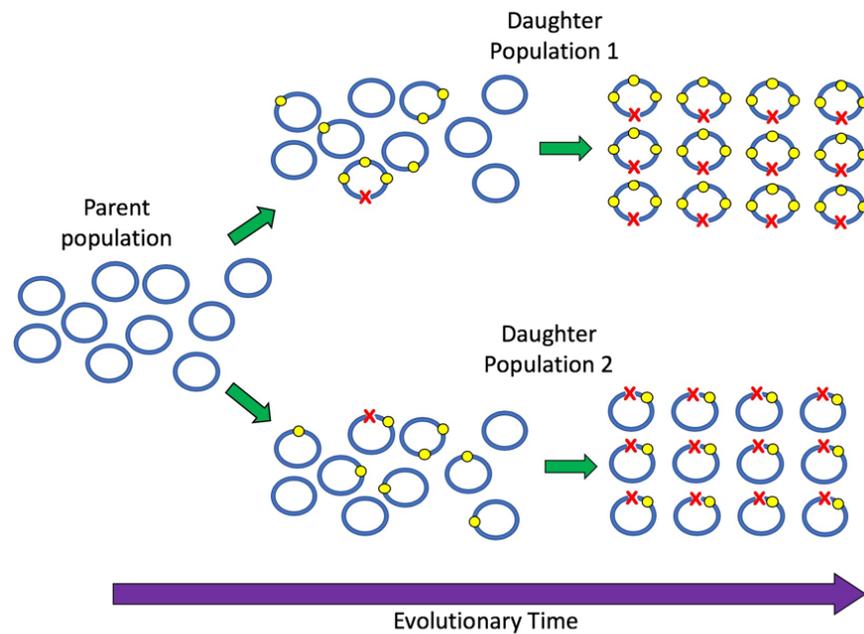


Figure 1. An illustration of rapid evolution of divergent mitochondrial genotypes between allopatric daughter populations via genetic hitchhiking. Blue rings represent the mitochondrial genome of an individual in a population. Yellow dots represent functionally neutral changes to the nucleotide sequence. Red Xs represent functional changes that bestow a benefit and that are subject to positive selection. Selection on the beneficial alleles leads to the fixation of those beneficial alleles along with any neutral traits that happen to be linked to them and purges the population of genetic variants.

The animal mitochondrial genome is single, effectively non-recombining chromosome, and the genes on this chromosome form one linkage group (Gray 1999). Under such circumstances, genetic hitchhiking is inevitable (Maynard Smith and Haigh 1974). Genetic hitchhiking results when strong positive selection on one genetic element causes an increase in the frequency of not only the element under selection but also of all of the genetic elements to which it is linked (Gillespie 2000; Meiklejohn et al. 2007)(Figure 1). The implications of genetic hitchhiking for the creation of a mt DNA barcode gap are inescapable (Costa and Carvalho 2010). If a favorable mutation occurs in any part of the mitochondrial genome—if for instance there is a nucleotide substitution in a mt-tRNA that improves the speed and accuracy of translation of mRNA (Adrion et al. 2016)—then positive selection for that mutation would cause an increase in the frequency of *the entire mitochondrial genotype that held that mutation*. If the mt chromosome that carried that favorable allele happened to also carry a unique, neutral mutation in the barcoding region of the COX1 gene, then that COX1 mutation would rise in frequency along with the mt-tRNA gene. Selection for the favorable allele could lead to rapid fixation of the new genotype, purging all diversity in mitochondrial genotypes within that population (Fig. 1). This process of genetic hitchhiking would essentially pull the mitochondrial genotype through a series of bottlenecks that would simultaneously purge standing variation within a population and fix differences in mt genotype between populations, creating the pattern of barcode gaps that typify the genomic structure of animals (Maynard Smith and Haigh 1974; Barton 2000; Meiklejohn et al. 2007). Because the mt and N genomes are inherited independently and N genes engage in recombination each generation, N genes could escape the bottleneck events affecting gene frequency in the mt genome.

The power of this explanation is that the proposed process would be ubiquitous among animals. Across most bilaterian animals, the genes that contribute to the function of the electron transport system are rigidly conserved—the same N genes cofunction with the same mt genes in a fruit fly and a chimpanzee

(Boore 1999; Gissi et al. 2008). A common set of interacting genes that are subject to the same functional constraints is exactly the circumstance that would give rise to a universal, selection-driven mt biological clock that runs faster than predicted by neutral theory (Hickerson et al. 2006). Adaptations to the external environment would only add noise to the dominant mode of evolution driven by mitonuclear coevolution.

Selective sweeps arising from the rapid fixation of mt variants under positive selection is a process already under discussion regarding the pattern of variation in mitochondrial genotypes within and among populations (Meiklejohn et al. 2007; Kerr 2011). By adding a need to consider both the protein coding and non-coding genes of the mt genome to the list of genes likely to be subject to at least periodic positive selection, a much greater opportunity for frequent selection sweeps is recognized. The majority of gene products of the mt genome is tRNAs, and the rate of mutation and evolutionary change of tRNA is much greater than the rate of amino acid substitutions in protein coding genes (Thornlow et al. 2018). Moreover, changes to mitochondrial tRNAs can have large effects on function and fitness. Numerous maternally inherited mitochondrial diseases are caused by nucleotide substitution on genes coding for mt tRNAs (Suzuki et al. 2011) and effects in non-human animals have also been documented (Meiklejohn et al. 2013). Given that function of mt tRNAs is dependent on the genotype of N-encoded Amino-acyl tRNA synthetase and N-encoded post-transcriptional processing proteins, we would predict positive selection for better performing variants as well as negative selection for dysfunctional variants (Pett and Lavrov 2015; Adrion et al. 2016). The same arguments for the importance of functional evolution of mitochondrial tRNAs also apply to mitochondrial-encoded rRNA (Scheel and Hausdorf 2014). Mitochondrial rRNA evolves at a rate that is an order of magnitude faster than the N-encoded ribosomal proteins (Barreto and Burton 2013) and these changes have functional consequences: as with tRNA, human inherited diseases are linked to nucleotide changes in mt rRNA (Scheper et al. 2007). Changes to the nucleotide sequence of the control region also can have functional consequences in terms of human disease (Chinnery et al. 2002) and functional divergence of the control region among sister taxa of animals can play a role in postzygotic isolation of populations (Ellison and Burton 2010). Positive selection on any of these non-protein coding genes should lead to selective sweeps that would fix neutral changes across the mitochondrial genome, including in DNA barcode regions, and this process would be perpetual and inevitable because of the necessity of coadaptation of the mitochondrial and N genomes.

Eyre-Walker (2006) pointed out that there is an interesting interaction between population size, genetic diversity, and genetic hitchhiking. As the size of a population increases, the amount of genetic diversity contained within that population, both in the mt and N genomes, will increase. This increased within-population diversity of mt genomes would work against the effectiveness of mt DNA barcodes for large populations. However, larger populations offer greater potential for the appearance of adaptive variants of mt genes, and hence a greater opportunity for genetic hitchhiking and selective sweep. He suggested that these two opposing forces might tend to negate each other, leaving genetic diversity of mt (but not N) genotypes largely independent of population size.

### **Selection sweeps of the W chromosome or mt DNA?**

The W chromosomes of birds have very low rates of variation suggesting that the genes on this chromosome have been subjected to selective sweeps (Berlin and Ellegren 2004; Ellegren 2013; Smeds et al. 2015). This low rate of variation in the genes on the avian W chromosome led Berlin et al. (2007) to hypothesize that selective sweeps on genes in the W chromosome would also result in selective sweeps on the mt genome, due to strict maternal linkage (perfect co-transmission) of mt DNA and the W chromosome. Following this logic, Berlin et al. (2007) proposed that evidence of selective sweeps of the mt genome would be evidence for positive selection on W genes. However, it could also work the other way: selective sweeps of the mt genome would cause selective sweeps and loss of variation in the W chromosome (Marais 2007; Lane 2008). The fact that the W chromosome of all birds investigated show signs of loss of genetic diversity via selective sweeps is thus, potentially, further support for the idea that the mt DNA barcode pattern is a consequence of selective sweeps. Birds are not the only taxa with ZW sex determination and co-transmission of mt and W chromosomes but to date there is much more sequencing data available for the W chromosome of birds than for any other ZW taxa. A broader survey of genetic diversity of W-linked genes might make it possible to

distinguish whether selection on the W-linked genes or mt-linked genes is responsible for observed patterns of low genetic diversity.

### Compensation-Draft Feedback

The co-evolution of mt and N genes has been proposed to lead to rapid serial fixation of alleles if a positive feedback loop arises as a consequence of changes and counter-changes between coevolving mt and N-mt genes. This idea is called the compensation-draft feedback hypothesis (Oliveira et al. 2008). Compensatory coevolution describes a situation whereby co-functioning sets of mt and N-mt genes are each under strong selection to improve aspects of performance that arise from the products of the other genome (Hill 2020). For instance, it was experimentally demonstrated that in a lab population of nematodes that the mt genome rapidly evolved a novel genotype to compensate for an OXPHOS dysfunction created by a N-mt allele (Christy et al. 2017). This example involves the interaction of protein-coding genes in an experimental lab setting, but for the reasons stated above, in most natural populations, the interacting mt and N gene products may be involved in transcription, translation, and replication of mt genes. The rapid fixation of mt genomes that carried this single adaptive nucleotide change might also have led to fixation of slightly deleterious alleles that happened to be associated with that allele. By this process of compensation-draft feedback, selective sweeps fix one problem while creating future problems that can be fixed through further selective sweeps when solutions happen to evolve. Such a series of selective sweeps would perpetually suppress within-population variation in mt genotypes while rapidly generating unique mt nucleotide sequences among populations thereby giving rise to mt DNA barcode gaps.

### An explanation for why mitochondrial barcoding fails

DNA barcoding using sequences from mt-encoded proteins or rRNA works very well for bilaterian animals, but it is much less effective in delimiting species boundaries of some other eukaryotic taxa, particularly plants (Chase et al. 2005; Kress et al. 2005) and fungi (Xu 2016) but also Porifera (sponges) and Anthozoa (corals and sea anemones) (Huang et al. 2008). Recombination of mt genomes, which is rare or non-existent in bilaterian animals, slows down or stops selective sweeps because beneficial alleles can be fixed in a mt genotype independent of the frequencies of other genes on the mt chromosome (Charlesworth et al. 1993; Rokas et al. 2003; White et al. 2008). The hypothesis for the evolution of barcode gaps that I outline in this paper, therefore, provides testable hypotheses for why mt DNA barcoding might fail for some taxa. If the efficacy of barcoding is dependent on selective sweeps, which in turn is dependent on lack of recombination of mt genomes, then it follows that taxa with recombination of mt genes will have a poor mt DNA barcode signal. Intriguingly, the mt genomes of Porifera and Anthozoa, for which mt DNA barcoding also works poorly, include introns, have very low mutation rates, and likely engage in recombination (Gissi et al. 2008; Huang et al. 2008; Brockman and McFadden 2012). Recombination of mitochondrial DNA has been documented in some plants and fungi (Barr et al. 2005), but the scope of recombination across these eukaryotic groups remains poorly known. For plants, the extent of recombination and the potential for selection sweeps is likely irrelevant to a failure of an effective mt DNA barcode—the rates of nucleotide substitution in plants (with some exceptions) is far lower than in other eukaryotic taxa, leaving little opportunity for the evolution of species-specific mt genotypes (Cowan et al. 2006). A broad-scale comparison of the efficacy of mt DNA barcoding in relation to rates of recombination and nucleotide substitution of mt DNA could be very illuminating.

Rampant introgression of mt genomes, wherein the mitochondrial genotype of one species replaces the mt genotype of another species with little change to N genotypes, will also erase a barcode signal (Toews and Brelsford 2012; Hill 2019b). Such mt introgression is hypothesized to occur when (1) the fitness gain from a better adapted heterospecific mitochondrion compensate for fitness losses from mitonuclear incompatibilities, (2) escape from mutational erosion and loss of mt function compensate for loss of mitonuclear incompatibilities, or (3) when a maternally transmitted parasite like *Wolbachia* infects a new host species and, because it is co-transmitted with mitochondria, causes the spread the mt genotype of the original host species in the new host species (Sloan et al. 2017; Hill 2019b). The effects of endosymbionts may be particularly problematic for the persistence of a mt DNA barcode gap because endosymbionts can drag mitochondria

across a species boundary and could be an explanation for why phenotypically distinct populations of animals like blowflies (Diptera: Calliphoridae) which have high rates of infection by endosymbionts often share a mitochondrial genotype (Whitworth et al. 2007). Loss of a uniquely coadapted mitonuclear genotype could be viewed as loss of species identity such that a lack of a DNA barcode gap in cases of rampant mt introgression is correctly failing to diagnose a collapsed species (Vonlanthen et al. 2012). Such an argument carries a risk of circularity, but the congruence between mt DNA barcode gaps and both conventional species boundaries (Hebert et al. 2003b; Tavares and Baker 2008) and the transitions in ornamentation used during mate choice for species recognition (Hill 2018) establishes a clear link between transitions in mitochondrial genotype and transitions between populations that taxonomists have recognized as species. The cases of rampant introgression of mt genomes then become rare exceptions that can be explained within the context of the mitonuclear compatibility species concept (Hill 2019b).

### Diagnosing species with mitochondrial genes

There is a widely held view that the use of mt DNA to diagnose species is a certain-to-be-flawed approximation that evolutionary biologists must endure until advances in sequencing technology allow us to do the job correctly with N genes. Because the mitochondrial genome is a single linkage group (at least for bilaterian animals), it is proposed that sequences from multiple N genes will reveal species boundaries with fundamentally better accuracy than the mt genes (Pazhenkova and Lukhtanov 2019). For instance, Chase et al. (2005) wrote that we will advance from mt gene sequences to “more sophisticated barcoding tools, which would be multiple, low-copy nuclear markers with sufficient genetic variability and PCR-reliability” to “identify the ‘genetic gaps’ that are useful in assessing species limits”. Along the same lines, Edwards et al. (2005) commented that ‘in our view, maternally inherited mtDNA can never capture enough of a species’ history to delimit species on its own’ and that ‘mtDNA should not have priority over N genes in avian species delimitation.’ Furthermore, it is sometimes stated that mt genomes introgress across species boundaries more readily than N alleles (Bonnet et al. 2017). The success of DNA barcoding across the majority bilaterian animals is conspicuous evidence that introgression of mt genomes across species boundaries is a rare rather than a common event. As evolutionary biologists compare N genes and mt genes between closely related species of bilaterian animals, the typical pattern that emerges is that the boundaries revealed by N genes are fuzzy while the boundaries between mitochondrial genotypes are discrete (Barrowclough and Zink 2009; Petit and Excoffier 2009; Toews et al. 2016; Hill 2019a). This pattern, of course, is why mt genes are used as DNA barcode genes. If mitonuclear interactions underlie the process of speciation, species limits are best defined by coadapted sets of co-functioning mt and N-mt genes, and a close proxy to this true species diagnosis is simply mt genotype (Hill 2017).

### Summary

In many taxa of bilaterian animals, there is little diversity in mitochondrial genotype within a species but substantial variation between species. This pattern is the basis for mt DNA barcoding as a means for identifying species. Despite the failure of neutral theory to explain this pattern of mt DNA sequence variation, most of the variation in the nucleotide sequence of barcode genes is neutral with respect to function. In other words, changes to the nucleotide sequence of mt DNA are evolving in a non-neutral manner despite the fact that they have no functional consequences. A solution to this paradox is that directional selection on any gene in the mitochondrial genome, including genes that code for rRNA and tRNA, can lead to selective sweeps that eliminate genetic diversity and fix neutral or slightly deleterious alleles in other parts of the mt genome. It is proposed that genetic hitchhiking by neutral elements in the DNA barcoding region explains how the DNA barcode gap evolves. This hypothesis proposes that mt DNA barcodes will only be effective when there is little or no recombination of mt genes, potentially explaining why DNA barcoding fails for some groups of eukaryotes.

### FIGURE LEGEND

Figure 1. An illustration of rapid evolution of divergent mitochondrial genotypes between allopatric daughter populations via genetic hitchhiking. Blue rings represent the mitochondrial genome of an individual in a

population. Yellow dots represent functionally neutral changes to the nucleotide sequence. Green bars represent functional changes that bestow a benefit and that are subject to positive selection. Selection on the beneficial alleles leads to the fixation of those beneficial alleles along with any neutral traits that happen to be linked to them while at the same time purging the population of genetic variants.

## Funding

This essay was prepared while the author was supported in part by NSF grant NSF-IOS-1754152.

## Acknowledgements

I thank Wendy Hood and the Hill/Hood lab groups for comments on a draft of this paper.

## Data Accessibility

No original data were used in writing this theoretical paper.

## Literature Cited

- Adrion, J. R., P. S. White, and K. L. Montooth. 2016. The roles of compensatory evolution and constraint in aminoacyl tRNA synthetase evolution. *Mol. Biol. Evol.* 33:152–161.
- Arnold, S. 2012. The power of life—cytochrome c oxidase takes center stage in metabolic control, cell signalling and survival. *Mitochondrion* 12:46–56.
- Avise, J. C. 2004. *Molecular Markers, Natural History, and Evolution*.
- Baker, A. J., E. S. Tavares, and R. F. Elbourne. 2009. Countering criticisms of single mitochondrial DNA gene barcoding in birds. *Mol. Ecol. Resour.* 9:257–268.
- Ballard, J. W. O., and M. Kreitman. 1995. Is mitochondrial DNA a strictly neutral marker? *Trends Ecol. Evol.* 10:485–488.
- Ballard, J. W. O., and N. Pichaud. 2014. Mitochondrial DNA: More than an evolutionary bystander. *Funct. Ecol.* 28:218–231.
- Ballard, J. W. O., and M. C. Whitlock. 2004. The incomplete natural history of mitochondria. *Mol. Ecol.* 13:729–744.
- Bar-Yaacov, D., A. Blumberg, and D. Mishmar. 2012. Mitochondrial-nuclear co-evolution and its effects on OXPHOS activity and regulation. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1819:1107–1111.
- Barr, C. M., M. Neiman, and D. R. Taylor. 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytol.* 168:39–50.
- Barreto, F. S., and R. S. Burton. 2013. Evidence for compensatory evolution of ribosomal proteins in response to rapid divergence of mitochondrial rRNA. *Mol. Biol. Evol.* 30:310–314.
- Barreto, F. S., E. T. Watson, T. G. Lima, C. S. Willett, S. Edmands, W. Li, and R. S. Burton. 2018. Genomic signatures of mitonuclear coevolution across populations of *Tigriopus californicus*. *Nat. Ecol. Evol.* 2:1250–1257.
- Barrientos, A., L. Kenyon, and C. T. Moraes. 1998. Human xenomitochondrial cybrids. *J. Biol. Chem.* 273:14210.
- Barrowclough, G. F., and R. M. Zink. 2009. Funds enough, and time: mtDNA, nuDNA and the discovery of divergence. *Mol. Ecol.* 18:2934–2936.
- Barshad, G., A. Blumberg, T. Cohen, and D. Mishmar. 2018. Human primitive brain displays negative mitochondrial-nuclear expression correlation of respiratory genes. *Genome Res.* 28:952–967.
- Barton, N. H. 2000. Genetic hitchhiking. *Philos. Trans. R. Soc. B Biol. Sci.* 355:1553–1562.

- Bazin, E., S. Glémin, and N. Galtier. 2006. Population Size Does Not Influence Mitochondrial Genetic Diversity in Animals. *Science* (80-. ). 312:570–572.
- Berlin, S., and H. Ellegren. 2004. Chicken W: A genetically uniform chromosome in a highly variable genome. *Proc. Natl. Acad. Sci. U. S. A.* 101:15967–15969.
- Berlin, S., D. Tomaras, and B. Charlesworth. 2007. Low mitochondrial variability in birds may indicate Hill-Robertson effects on the W chromosome. *Heredity* (Edinb). 99:389–396.
- Blount, Z. D., R. E. Lenski, and J. B. Losos. 2018. Contingency and determinism in evolution: Replaying life’s tape. *Science* 362.
- Bonnet, T., R. Leblois, F. Rousset, and P.-A. Crochet. 2017. A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. *Evolution* (N. Y). 71:2140–2158.
- Boore, J. L. 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* 27:1767–1780.
- Brockman, S. A., and C. S. McFadden. 2012. The mitochondrial genome of *Paraminabea aldersladei* (Cnidaria: Anthozoa: Octocorallia) supports intramolecular recombination as the primary mechanism of gene rearrangement in octocoral mitochondrial genomes. *Genome Biol. Evol.* 4:994–1006.
- Bucklin, A., D. Steinke, and L. Blanco-Bercial. 2011. DNA Barcoding of Marine Metazoa. *Ann. Rev. Mar. Sci.* 3:471–508.
- Burton, R. S., and F. S. Barreto. 2012. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Mol. Ecol.* 21:4942–4957.
- Calvo, E., S. Cogliati, P. Hernansanz-Agustin, M. Loureiro-Lopez, A. Guaras, R. A. Casuso, F. Garcia-Marques, R. Acin-Perez, Y. Marti-Mateos, J. C. Silla-Castro, M. Carro-Alvarellos, J. R. Huertas, J. Vazquez, and J. A. Enriquez. 2019. Functional role of respiratory supercomplexes in mice: segmentation of the Qpool and SCAF1. *bioRxiv* 826115.
- Charlesworth, D., M. T. Morgan, and B. Charlesworth. 1993. Mutation accumulation in finite outbreeding and inbreeding populations. *Genet. Res.* 61:39–56.
- Chase, M. W., N. Salamin, M. Wilkinson, J. M. Dunwell, R. P. Kesanakurthi, N. Haidar, and V. Savolainen. 2005. Land plants and DNA barcodes: Short-term and long-term goals. *Philos. Trans. R. Soc. B Biol. Sci.* 360:1889–1895.
- Chinnery, P. F., G. A. Taylor, N. Howell, D. T. Brown, T. J. Parsons, and D. M. Turnbull. 2002. Point Mutations of the mtDNA Control Region in Normal and Neurodegenerative Human Brains. *Am. J. Hum. Genet.* 68:529–532.
- Chou, J. Y., and J. Y. Leu. 2010. Speciation through cytonuclear incompatibility: Insights from yeast and implications for higher eukaryotes. *BioEssays* 32:401–411.
- Christy, S. F., R. I. Wernick, M. J. Lue, G. Velasco, D. K. Howe, D. R. Denver, and S. Estes. 2017. Adaptive Evolution under Extreme Genetic Drift in Oxidatively Stressed *Caenorhabditis elegans*. *Genome Biol. Evol.* 9:3008–3022.
- Clare, E. L., B. K. Lim, M. D. Engstrom, J. L. Eger, and P. D. N. Hebert. 2007. DNA barcoding of Neotropical bats: species identification and discovery within Guyana. *Mol. Ecol. Notes* 7:184–190. Wiley Online Library.
- Coddington, J. A., I. Agnarsson, R. C. Cheng, K. Čandek, A. Driskell, H. Frick, M. Gregorič, R. Kostanjšek, C. Kropf, M. Kweskin, T. Lokovšek, M. Pipan, N. Vidergar, and M. Kuntner. 2016. DNA barcode data accurately assign higher spider taxa. *PeerJ* 2016:1–25.
- Costa, F. O., and G. R. Carvalho. 2010. New insights into molecular evolution: Prospects from the barcode of life initiative (BOLI). *Theory Biosci.* 129:149–157.

- Cowan, R. S., M. W. Chase, W. J. Kress, and V. Savolainen. 2006. 300,000 Species to identify: Problems, progress, and prospects in DNA barcoding of land plants. *Taxon* 55:611–616.
- da Fonseca, R. R., W. E. Johnson, S. J. O'Brien, M. J. Ramos, and A. Antunes. 2008. The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics* 9:119.
- Dasmahapatra, K. K., and J. Mallet. 2006. Taxonomy: DNA barcodes: recent successes and future prospects. *Heredity (Edinb)*. 97:254–255.
- Dobzhansky, T. G. 1937. *Genetics and the Origin of Species*. Columbia University Press, New York.
- Dowling, D. K., U. Friberg, and J. Lindell. 2008. Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends Ecol. Evol.* 23:546–554.
- Dupuis, J. R., A. D. Roe, and F. A. H. Sperling. 2012. Multi-locus species delimitation in closely related animals and fungi: One marker is not enough.
- Edwards, S. V., S. B. Kingan, J. D. Calkins, C. N. Balakrishnan, W. B. Jennings, W. J. Swanson, and M. D. Sorenson. 2005. Speciation in birds: Genes, geography, and sexual selection. *Proc. Natl. Acad. Sci.* 102:6550–6557.
- Ellegren, H. 2013. The Evolutionary Genomics of Birds. *Annu. Rev. Ecol. Evol. Syst.* 44:239–259.
- Ellison, C. K., and R. S. Burton. 2010. Cytonuclear conflict in interpopulation hybrids: The role of RNA polymerase in mtDNA transcription and replication. *J. Evol. Biol.* 23:528–538.
- Ellison, C. K., and R. S. Burton. 2008a. Genotype-dependent variation of mitochondrial transcriptional profiles in interpopulation hybrids. *Proc. Natl. Acad. Sci.* 105:15831–15836.
- Ellison, C. K., and R. S. Burton. 2008b. Interpopulation hybrid breakdown maps to the mitochondrial genome. *Evolution (N. Y)*. 62:631–638.
- Eyre-Walker, A. 2006. Size Does Not Matter for Mitochondrial DNA. *Science (80- )*. 312:537–538.
- Garvin, M. R., J. P. Bielawski, and A. J. Gharrett. 2011. Positive darwinian selection in the piston that powers proton pumps in Complex I of the mitochondria of Pacific salmon. *PLoS One* 6.
- Garvin, M. R., J. P. Bielawski, L. A. Sazanov, and A. J. Gharrett. 2014. Review and meta-analysis of natural selection in mitochondrial complex I in metazoans. *J. Zool. Syst. Evol. Res.* 53:1–17.
- Gaspari, M., M. Falkenberg, N.-G. Larsson, and C. M. Gustafsson. 2004. The mitochondrial RNA polymerase contributes critically to promoter specificity in mammalian cells. *EMBO J.* 23:4606–4614.
- Gershoni, M., L. Levin, O. Ovadia, Y. Toiw, N. Shani, S. Dadon, N. Barzilai, A. Bergman, G. Atzmon, J. Wainstein, A. Tsur, L. Nijtmans, B. Glaser, and D. Mishmar. 2014. Disrupting mitochondrial-nuclear coevolution affects OXPHOS complex i integrity and impacts human health. *Genome Biol. Evol.* 6:2665–2680.
- Gershoni, M., A. R. Templeton, and D. Mishmar. 2009. Mitochondrial bioenergetics as a major motive force of speciation. *Bioessays* 31:642–650.
- Gillespie, J. H. 2000. Genetic drift in infinite populations. *Genetics* 155:909–919.
- Gissi, C., F. Iannelli, and G. Pesole. 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity (Edinb)*. 101:301–320.
- Gray, M. W. 1999. Evolution of organellar genomes. *Curr. Opin. Genet. Dev.* 9:678–687.
- Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. deWaard. 2003a. Biological identifications through DNA barcodes. *Proc. R. Soc. B Biol. Sci.* 270:313–321.

- Hebert, P. D. N., S. Ratnasingham, and J. Waard. 2003b. Barcoding animal life : cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. B* 270:S96–S99.
- Hickerson, M. J., C. P. Meyer, and C. Moritz. 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. *Syst. Biol.* 55:729–739.
- Hill, G. E. 2016. Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. *Ecol. Evol.* 6:5831–5842.
- Hill, G. E. 2020. Mitonuclear Compensatory Coevolution. *Trends Genet.*
- Hill, G. E. 2019a. *Mitonuclear Ecology*. Oxford University Press, Oxford, UK.
- Hill, G. E. 2018. Mitonuclear Mate Choice: A Missing Component of Sexual Selection Theory? *BioEssays* 40:1700191.
- Hill, G. E. 2019b. Reconciling the Mitonuclear Compatibility Species Concept with Rampant Mitochondrial Introgression. *Integr. Comp. Biol.* 59:912–924.
- Hill, G. E. 2017. The mitonuclear compatibility species concept. *Auk* 134:393–409.
- Hill, G. E., J. C. Havird, D. B. Sloan, R. S. Burton, C. Greening, and D. K. Dowling. 2019. Assessing the fitness consequences of mitonuclear interactions in natural populations. *Biol. Rev.* 94:1089–1104.
- Hoekstra, L. A., M. A. Siddiq, and K. L. Montooth. 2013. Pleiotropic effects of a mitochondrial-nuclear incompatibility depend upon the accelerating effect of temperature in *Drosophila*. *Genetics* 195:1129–1139.
- Huang, D., R. Meier, P. A. Todd, and L. M. Chou. 2008. Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding. *J. Mol. Evol.* 66:167–174.
- James, J. E., G. Piganeau, and A. Eyre-Walker. 2016. The rate of adaptive evolution in animal mitochondria. *Mol. Ecol.* 25:67–78.
- Janzen, D. H., W. Hallwachs, P. Blandin, J. M. Burns, J. M. Cadiou, I. Chacon, T. Dapkey, A. R. Deans, M. E. Epstein, B. Espinoza, J. G. Franclemont, W. A. Haber, M. Hajibabaei, J. P. W. Hall, P. D. N. Hebert, I. D. Gauld, D. J. Harvey, A. Hausmann, I. J. Kitching, D. Lafontaine, J. F. ?ois Landry, C. Lemaire, J. Y. Miller, J. S. Miller, L. Miller, S. E. Miller, J. Montero, E. Munroe, S. R. Green, S. Ratnasingham, J. E. Rawlins, R. K. Robbins, J. J. Rodriguez, R. Rougerie, M. J. Sharkey, M. A. Smith, M. A. Solis, J. B. Sullivan, P. Thiaucourt, D. B. Wahl, S. J. Weller, J. B. Whitfield, K. R. Willmott, D. M. Wood, N. E. Woodley, and J. J. Wilson. 2009. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Mol. Ecol. Resour.* 9:1–26.
- Kazancıoğlu, E., and G. Arnqvist. 2014. The maintenance of mitochondrial genetic variation by negative frequency-dependent selection. *Ecol. Lett.* 17:22–27.
- Kern, A. D., and M. W. Hahn. 2018. The neutral theory in light of natural selection. *Mol. Biol. Evol.* 1–6.
- Kerr, K. C. R. 2011. Searching for evidence of selection in avian DNA barcodes. *Mol. Ecol. Resour.* 11:1045–1055.
- Kerr, K. C. R., M. Y. Stoeckle, C. J. Dove, L. A. Weigt, C. M. Francis, and P. D. N. Hebert. 2007. Comprehensive DNA barcode coverage of North American birds. *Mol. Ecol. Notes* 7:535–543.
- Kivisild, T., P. Shen, D. P. Wall, B. Do, R. Sung, K. Davis, G. Passarino, P. A. Underhill, C. Scharfe, A. Torroni, R. Scozzari, D. Modiano, A. Coppa, P. De Knijff, M. Feldman, L. L. Cavalli-Sforza, and P. J. Oefner. 2006. The role of selection in the evolution of human mitochondrial genomes. *Genetics* 172:373–387.
- Kress, W. J., K. J. Wurdack, E. A. Zimmer, L. A. Weigt, and D. H. Janzen. 2005. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci.* 102:8369–8374.

- Kwong, S., A. Srivathsan, G. Vaidya, and R. Meier. 2012. Is the COI barcoding gene involved in speciation through intergenomic conflict? *Mol. Phylogenet. Evol.* 62:1009–1012.
- Lane, N. 2009a. Biodiversity: On the origin of bar codes. *Nature* 462:272–274.
- Lane, N. 2014. Bioenergetic constraints on the evolution of complex life. *Cold Spring Harb. Perspect. Biol.* 6.
- Lane, N. 2008. Mitochondria and the W chromosome: Low variability on the W chromosome in birds is more likely to indicate selection on mitochondrial genes. *Heredity (Edinb.)* 100:444–445.
- Lane, N. 2011. Mitonuclear match: Optimizing fitness and fertility over generations drives ageing within generations. *Bioessays* 33:860–869.
- Lane, N. 2009b. On the origin of bar codes. *Nature* 462:272–274.
- Lanza, R. P., B. L. Dresser, and P. Damiani. 2000. Cloning Noah’s Ark. *Sci. Am.* 283:84–89.
- Latorre-Pellicer, A., R. Moreno-Loshuertos, A. V. Lechuga-Vieco, F. Sánchez-Cabo, C. Torroja, R. Acín-Pérez, E. Calvo, E. Aix, A. González-Guerra, A. Logan, M. L. Bernad-Miana, E. Romanos, R. Cruz, S. Cogliati, B. Sobrino, Á. Carracedo, A. Pérez-Martos, P. Fernández-Silva, J. Ruíz-Cabello, M. P. Murphy, I. Flores, J. Vázquez, and J. A. Enríquez. 2016. Mitochondrial and nuclear DNA matching shapes metabolism and healthy ageing. *Nature* 535:561–565.
- Lee, H. Y., J. Y. Chou, L. Cheong, N. H. Chang, S. Y. Yang, and J. Y. Leu. 2008. Incompatibility of Nuclear and Mitochondrial Genomes Causes Hybrid Sterility between Two Yeast Species. *Cell* 135:1065–1073. Elsevier Ltd.
- Luo, Y., X. Yang, and Y. Gao. 2013. Mitochondrial DNA response to high altitude: A new perspective on high-altitude adaptation. *Mitochondrial DNA* 24:313–319.
- Lynch, M., B. Koskella, and S. Schaack. 2006. Mutation Pressure and the Evolution of Organelle Genomic Architecture. *Science (80- )*. 311:1727–1730.
- Marais, G. A. B. 2007. Sex chromosomes and mitochondrial DNA polymorphism in birds: The Hill-Robertson effects extend from nucleus to mitochondria. *Heredity (Edinb.)* 99:357–358.
- Maynard Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. *Genet. Res.* 23:23–35.
- Meiklejohn, C. D., M. A. Holmbeck, M. A. Siddiq, D. N. Abt, D. M. Rand, and K. L. Montooth. 2013. An Incompatibility between a Mitochondrial tRNA and Its Nuclear-Encoded tRNA Synthetase Compromises Development and Fitness in *Drosophila*. *PLoS Genet.* 9:e1003238.
- Meiklejohn, C. D., K. L. Montooth, and D. M. Rand. 2007. Positive and negative selection on the mitochondrial genome. *Trends Genet* 23:259–263.
- Meyer, C. P., and G. Paulay. 2005. DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biol.* 3:1–10.
- Morinière, J., L. Hendrich, M. Balke, A. J. Beermann, T. König, M. Hess, S. Koch, R. Müller, F. Leese, P. D. N. Hebert, A. Hausmann, C. D. Schubart, and G. Haszprunar. 2017. A DNA barcode library for Germany’s mayflies, stoneflies and caddisflies (Ephemeroptera, Plecoptera and Trichoptera). *Mol. Ecol. Resour.*, doi: 10.1111/1755-0998.12683.
- Moritz, C., T. E. Dowling, and W. M. Brown. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu. Rev. Ecol. Syst.* 18:269–292. Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA.
- Nabholz, B., S. Glemin, and N. Galtier. 2009. The erratic mitochondrial clock: variations of mutation rate, not population size, affect  $\{\{\text{mtDNA}\}\}$  diversity across birds and mammals. *Bmc Evol. Biol.* 9:13.

- Naro-Maciel, E., B. Reid, N. N. Fitzsimmons, M. Le, R. O. B. Desalle, and G. Amato. 2010. DNA barcodes for globally threatened marine turtles: a registry approach to documenting biodiversity. *Mol. Ecol. Resour.* 10:252–263. Wiley Online Library.
- Oliveira, D. C. S. G., R. Raychoudhury, D. V. Lavrov, and J. H. Werren. 2008. Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp *Nasonia* (Hymenoptera: Pteromalidae). *Mol. Biol. Evol.* 25:2167–2180.
- Osada, N., and H. Akashi. 2012. Mitochondrial-nuclear interactions and accelerated compensatory evolution: Evidence from the primate cytochrome c oxidase complex. *Mol. Biol. Evol.* 29:337–346.
- Pacelli, C., D. Latorre, T. Cocco, F. Capuano, C. Kukat, P. Seibel, and G. Villani. 2011. Tight control of mitochondrial membrane potential by cytochrome c oxidase. *Mitochondrion* 11:334–341.
- Palumbi, S. R., F. Cipriano, and M. P. Hare. 2001. Predicting Nuclear Gene Coalescence From Mitochondrial Data: the Three-Times Rule. *Evolution* (N. Y). 55:859.
- Papadopoulou, A., J. Bergsten, T. Fujisawa, M. T. Monaghan, T. G. Barraclough, and A. P. Vogler. 2008. Speciation and DNA barcodes: testing the effects of dispersal on the formation of discrete sequence clusters. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 363:2987–2996.
- Pazhenkova, E. A., and V. A. Lukhtanov. 2019. Nuclear genes (but not mitochondrial DNA barcodes) reveal real species: Evidence from the *Brenthis* fritillary butterflies (Lepidoptera, Nymphalidae). *J. Zool. Syst. Evol. Res.* 57:298–313.
- Petit, R. J., and L. Excoffier. 2009. Gene flow and species delimitation. *Trends Ecol. Evol.* 24:386–393.
- Pett, W., and D. Lavrov. 2015. Mito-nuclear interactions in the evolution of animal mitochondrial tRNA metabolism. *Genome Biol. Evol.* 7:2089–2101.
- Pfenninger, M., H. Lerp, M. Tobler, C. Passow, J. L. Kelley, E. Funke, B. Greshake, U. K. Erkoç, T. Berberich, and M. Plath. 2014. Parallel evolution of *cox* genes in H<sub>2</sub>S-tolerant fish as key adaptation to a toxic environment. *Nat. Commun.* 5:3873.
- Pierron, D., D. E. Wildman, M. Huttemann, G. C. Markondapatnaikuni, S. Aras, and L. I. Grossman. 2012. Cytochrome c oxidase: Evolution of control via nuclear subunit addition. *Biochim. Biophys. Acta - Bioenerg.* 1817:590–597. Elsevier B.V.
- Popadin, K. Y., S. I. Nikolaev, T. Junier, M. Baranova, and S. E. Antonarakis. 2013. Purifying selection in mammalian mitochondrial protein-coding genes is highly effective and congruent with evolution of nuclear genes. *Mol. Biol. Evol.* 30:347–355.
- Rand, D. M., M. Dorfsman, and L. M. Kann. 1994. Neutral and non-neutral evolution of *Drosophila* mitochondrial DNA. *Genetics* 138:741–756.
- Rand, D. M., R. A. Haney, and A. J. Fry. 2004. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol. Evol.* 19:645–653.
- Ratnasingham, S., and P. D. N. Hebert. 2013. A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. *PLoS One*, doi: 10.1371/journal.pone.0066213.
- Rokas, A., E. Ladoukakis, and E. Zouros. 2003. Animal mitochondrial DNA recombination revisited. *Trends Ecol. Evol.* 18:411–417.
- Rubinoff, D., S. Cameron, and K. Will. 2006. A genomic perspective on the shortcomings of mitochondrial DNA for “barcoding” identification. *J. Hered.* 97:581–594.
- Ruiz-Pesini, E., and D. C. Wallace. 2006. Evidence for adaptive selection acting on the tRNA and rRNA genes of human mitochondrial DNA. *Hum. Mutat.* 27:1072–1081. Wiley Online Library.

- Saccone, C., C. Gissi, C. Lanave, A. Larizza, G. Pesole, and A. Reyes. 2000. Evolution of the mitochondrial genetic system: An overview. *Gene* 261:153–159.
- Scheel, B. M., and B. Hausdorf. 2014. Dynamic evolution of mitochondrial ribosomal proteins in Holozoa. *Mol. Phylogenet. Evol.* 76:67–74. Elsevier Inc.
- Scheper, G. C., M. S. van der Knaap, and C. G. Proud. 2007. Translation matters: protein synthesis defects in inherited disease. *Nat. Rev. Genet.* 8:711–723.
- Scott, G. R., P. M. Schulte, S. Egginton, A. L. M. Scott, J. G. Richards, and W. K. Milsom. 2011. Molecular evolution of cytochrome c oxidase underlies high-altitude adaptation in the bar-headed goose. *Mol. Biol. Evol.* 28:351–363.
- Shtolz, N., and D. Mishmar. 2019. The Mitochondrial Genome—on Selective Constraints and Signatures at the Organism, Cell, and Single Mitochondrion Levels. *Frontiers Media S.A.*
- Sloan, D. B., J. C. Havird, and J. Sharbrough. 2017. The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Mol. Ecol.* 26:2212–2236.
- Sloan, D. B., J. M. Warren, A. M. Williams, Z. Wu, S. E. Abdel-Ghany, A. J. Chicco, and J. C. Havird. 2018. Cytonuclear integration and co-evolution. *Nat. Rev. Genet.* 19:635–648.
- Smeds, L., V. Warmuth, P. Bolivar, S. Uebbing, R. Burri, A. Suh, A. Nater, S. Bureš, L. Z. Garamszegi, S. Hogner, J. Moreno, A. Qvarnström, M. Ruzic, S. A. Sæther, G. P. Sætre, J. Török, and H. Ellegren. 2015. Evolutionary analysis of the female-specific avian W chromosome. *Nat. Commun.* 6.
- Smith, D. R. 2016. The mutational hazard hypothesis of organelle genome evolution: 10 years on. *Mol. Ecol.* 25:3769–3775.
- Smith, M. A., B. L. Fisher, and P. D. N. Hebert. 2005. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: The ants of Madagascar. *Philos. Trans. R. Soc. B Biol. Sci.*, doi: 10.1098/rstb.2005.1714.
- Smith, M. A., J. J. Rodriguez, J. B. Whitfield, A. R. Deans, D. H. Janzen, W. Hallwachs, and P. D. N. Hebert. 2008. Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. *Proc. Natl. Acad. Sci.* 105:12359 LP – 12364.
- Stewart, J. B., C. Freyer, J. L. Elson, and N.-G. Larsson. 2008. Purifying selection of mtDNA and its implications for understanding evolution and mitochondrial disease. *Nat Rev Genet* 9:657–662.
- Stoeckle, M. Y., and D. S. Thaler. 2014. DNA barcoding works in practice but not in neutral theory. *PLoS One* 9:e100755.
- Sunnucks, P., H. E. Morales, A. M. Lamb, A. Pavlova, and C. Greening. 2017. Integrative approaches for studying mitochondrial and nuclear genome co-evolution in oxidative phosphorylation. *Front. Genet.* 8:25.
- Suzuki, T., A. Nagao, and T. Suzuki. 2011. Human Mitochondrial tRNAs: Biogenesis, Function, Structural Aspects, and Diseases. *Annu. Rev. Genet.* 45:299–329.
- Tavares, E. S., and A. J. Baker. 2008. Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. *Bmc Evol. Biol.* 8:14.
- Teske, P. R., T. R. Golla, J. Sandoval-Castillo, A. Emami-Khoyi, C. D. Van Der Lingen, S. Von Der Heyden, B. Chiazzari, B. Jansen Van Vuuren, and L. B. Beheregaray. 2018. Mitochondrial DNA is unsuitable to test for isolation by distance. *Sci. Rep.* 8:1–9. Springer US.
- Thornlow, B. P., J. Hough, J. M. Roger, H. Gong, T. M. Lowe, and R. B. Corbett-Detig. 2018. Transfer RNA genes experience exceptionally elevated mutation rates. *Proc. Natl. Acad. Sci.* 115:8996–9001.

- Tobe, S. S., A. C. Kitchener, and A. M. T. Linacre. 2010. Reconstructing mammalian phylogenies: a detailed comparison of the cytochrome b and cytochrome oxidase subunit I mitochondrial genes. *PLoS One* 5:e14156.
- Tobler, M., N. Barts, and R. Greenway. 2019. Mitochondria and the origin of species: bridging genetic and ecological perspectives on speciation processes. *Integr. Comp. Biol.*
- Toews, D. P. L., and A. Brelsford. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Mol. Ecol.* 21:3907–3930.
- Toews, D. P. L., S. A. Taylor, R. Vallender, A. Brelsford, B. G. Butcher, P. W. Messer, and I. J. Lovette. 2016. Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers. *Curr. Biol.* 26:2313–2318. Elsevier Ltd.
- Tomasco, I. H., and E. P. Lessa. 2014. Two mitochondrial genes under episodic positive selection in subterranean octodontoid rodents. *Gene* 534:371–378. Elsevier B.V.
- Vences, M., M. Thomas, R. M. Bonett, and D. R. Vieites. 2005. Deciphering amphibian diversity through DNA barcoding: chances and challenges. *Philos. Trans. R. Soc. London B Biol. Sci.* 360:1859–1868. The Royal Society.
- Vonlanthen, P., D. Bittner, A. G. Hudson, K. A. Young, R. Müller, B. Lundsgaard-Hansen, D. Roy, S. Di Piazza, C. R. Largiadèr, and O. Seehausen. 2012. Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature* 482:357. Nature Publishing Group.
- Wallace, D. C. 2010. Bioenergetics, the origins of complexity, and the ascent of man. *Proc. Natl. Acad. Sci.* 107:8947–8953.
- Wang, Z. O., and D. D. Pollock. 2007. Coevolutionary patterns in cytochrome c oxidase subunit I depend on structural and functional context. *J. Mol. Evol.* 65:485–495.
- Ward, R. D., and B. H. Holmes. 2007. An analysis of nucleotide and amino acid variability in the barcode region of cytochrome c oxidase I (cox1) in fishes. *Mol. Ecol. Notes* 7:899–907. Wiley Online Library.
- Wei, W., S. Tuna, M. J. Keogh, K. R. Smith, T. J. Aitman, P. L. Beales, D. L. Bennett, D. P. Gale, M. A. K. Bitner-Glindzicz, and G. C. Black. 2019. Germline selection shapes human mitochondrial DNA diversity. *Science* (80-. ). 364. American Association for the Advancement of Science.
- White, D. J., J. N. Wolff, M. Pierson, and N. J. Gemmell. 2008. Revealing the hidden complexities of mtDNA inheritance. *Mol. Ecol.* 17:4925–4942.
- Whitworth, T. L., R. D. Dawson, H. Magalon, and E. Baudry. 2007. DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). *Proc. R. Soc. B Biol. Sci.* 274:1731–1739.
- Wilson, A. C., R. L. Cann, S. M. Carr, M. George, U. B. Gyllensten, K. M. Helm-Bychowski, R. G. Higuchi, S. R. Palumbi, E. M. Prager, R. D. Sage, and M. Stoneking. 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol. J. Linn. Soc.* 26:375–400.
- Wolff, J. N., E. D. Ladoukakis, J. A. Enriquez, and D. K. Dowling. 2014. Mitonuclear interactions: evolutionary consequences over multiple biological scales. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130443.
- Wright, S. 1942. Genetics and the origin of species. *J. Hered.* 33:283–284.
- Xu, J. 2016. Fungal DNA barcoding. *Genome* 59:913–932.
- Zahiri, R., J. D. Lafontaine, B. C. Schmidt, J. R. DeWaard, E. V Zakharov, and P. D. N. Hebert. 2014. A transcontinental challenge - A test of DNA barcode performance for 1,541 species of Canadian Noctuoidea (Lepidoptera). *PLoS One* 9:1–12.

Zaidi, A. A., and K. D. Makova. 2019. Investigating mitonuclear interactions in human admixed populations. *Nat. Ecol. Evol.* 3:213–222. Nature Publishing Group.

Zhu, C. T., P. Ingelmo, and D. M. Rand. 2014. GxGxE for Lifespan in *Drosophila*: Mitochondrial, Nuclear, and Dietary Interactions that Modify Longevity. *PLoS Genet.* 10:e1004354.

Zink, R. M., and G. F. Barrowclough. 2008. Mitochondrial DNA under siege in avian phylogeography. *Mol. Ecol.* 17:2107–2121.

### Hosted file

Fig 1 barcode draft\_2.pptx available at <https://authorea.com/users/338217/articles/464276-genetic-hitchhiking-mitonuclear-coadaptation-and-the-origins-of-mt-dna-barcode-gaps>