Why should mitochondria define species?

Category: Article

Authors: Mark Y. Stoeckle[1]*

David S. Thaler[2]*

*Corresponding Authors

The authors contributed equally to the manuscript and both are corresponding authors

Affiliations: 1. Program for the Human Environment
The Rockefeller University
1230 York AVE
New York, NY 10065
USA
email mark.stoeckle@rockefeller.edu
phone: 212 327-7917
fax 212 327-7519

2. Biozentrum, University of Basel
Klingelbergstrasse 50/70
CH - 4056 Basel
Switzerland
david.thaler@unibas.ch davidsthaler@gmail.com

Abstract

More than a decade of DNA barcoding encompassing about five million specimens covering 100,000 animal species supports the generalization that mitochondrial DNA clusters largely overlap with species as defined by domain experts. Most barcode clustering reflects synonymous substitutions. What evolutionary mechanisms account for synonymous clusters being largely coincident with species? The answer depends on whether variants are phenotypically neutral. To the degree that variants are selectable, purifying selection limits variation within species and neighboring species have distinct adaptive peaks. Phenotypically neutral variants are only subject to demographic processes—drift, lineage sorting, genetic hitchhiking, and bottlenecks. The evolution of modern humans has been studied from several disciplines with detail unique among animal species. Mitochondrial barcodes provide a commensurable way to compare modern humans to other animal species. Barcode variation in the modern human population is quantitatively similar to that within other animal species. Several convergent lines of evidence show that mitochondrial diversity in modern humans follows from sequence uniformity followed by the accumulation of largely neutral diversity during a population expansion that began approximately 100,000 years ago. A straightforward hypothesis is that the extant populations of almost all animal species have arrived at a similar result consequent to a similar process of expansion from mitochondrial uniformity within the last one to several hundred thousand years.

Precis

1. Mitochondrial Cytochrome Oxidase Subunit I DNA barcodes (COI barcodes, often shortened to "DNA barcodes" or "barcodes" in this article) began as an aid to animal species identification and made no claims of contributing to evolutionary theory. Five million DNA barcodes later the consistent and commensurable pattern they present throughout the animal kingdom is one of the most general in biology. In well-studied groups the majority of DNA barcode clusters agree with domain experts' judgment of distinct species.
2. The tight clustering of barcodes within species and unfilled sequence space among them are key facts of animal life that evolutionary theory must explain. Many aspects of speciation are complex. Barcodes are unique in being quantifiably commensurable across all animal species and almost always yielding the same single simple answer [1].
3. Either of two evolutionary mechanisms might account for the facts: a) species-specific selection, or b) demographic processes acting independently of phenotype.
4. Most barcode variation consists of synonymous codon changes. Since the assumption of neutrality of mitochondrial synonymous codons was asserted, many exceptions in nuclear genes and prokaryotic systems have been found.
5. New arguments are presented that synonymous codon changes in mitochondrial genes are neutral to a greater extent than nuclear genes.
6. Extensive data on modern humans make our species a valuable model system for animal evolution as a whole. The mitochondrial variation within the modern human population is about average when compared to the extant populations of most animal species.
7. Similar neutral variation of humans and other animals implies that the extant populations of most animal species have, like modern humans, recently passed though mitochondrial uniformity.

History of COI barcoding. DNA barcoding was first proposed as a tool for practical taxonomy and to democratize actionable biological knowledge [2, 3]. At its origin DNA barcoding made no claim of contributing to evolutionary theory. Previous work bode well for mitochondrial genomes being reliably similar within animal species yet in many cases distinct among neighbor species [4, 5]. The particular mitochondrial sequence that has become the most widely used, the 648 base pair (bp) segment of the gene encoding mitochondrial cytochrome *c* oxidase subunit I (COI), reached a tipping point because widely applicable reliable primers and methods useful for both vertebrates and invertebrates were adopted by a critical mass of a community [6, 7]. Skeptics of COI barcoding [8] raised a number of objections about its power and/or generality of as a single simple metric applicable to the entire animal kingdom, including 1) the small fraction of the genome (about 5% of the mitochondrial genome and less than one millionth of the total organism's genome) might not be sensitive or representative [9, 10]; 2) since animal mitochondria are inherited maternally the apparent pattern of speciation from mitochondria is vulnerable to distortion when females and males roam differently [11]; 3) the mitochondrial chromosome is subject to types of selection not experienced by the nuclear genome [12]: replicon competition within each organelle [13], among organelles inside each cell [14-16], including differential segregation of organelles at cell division [17]; and 4) mitochondria in some groups are sensitive to agents such as Wolbachia that are not known to affect nuclear genes [18]. Mitochondrial pseudogenes in the nucleus sometimes confused analysis[19]. Anecdotally, some

domain experts felt that only specialists can reliably recognize species in each group and that "DNA taxonomy" was felt as necessarily inferior or a threat.

The current field of COI barcodes is no longer fragile but neither is it complete. As of late 2016 there were close to five million COI barcodes between the Genbank and BOLD databases. Objections can now be seen in the cumulative light of these data and more than a decade's experience. There is no longer any doubt that DNA barcodes are useful and practical (Figs. 1,2). The agreement with specialists encompasses most cases in several important animal domains. Many cases where DNA barcodes and domain specialists do not agree reflect geographic splits within species or hybridization between species. Others upon further investigation been attributed to mislabeling or sequence error [20]. Some may represent *bona fide* exceptions to the rule that mitochondrial sequence clusters coincide with species defined by other means. In the great majority of cases COI barcodes yield a close approximation of what specialists come up with after a lot of study. Birds are one of the best characterized of all animal groups and COI barcode clusters have been tabulated as agreeing with expert taxonomy for 94% of species [21].

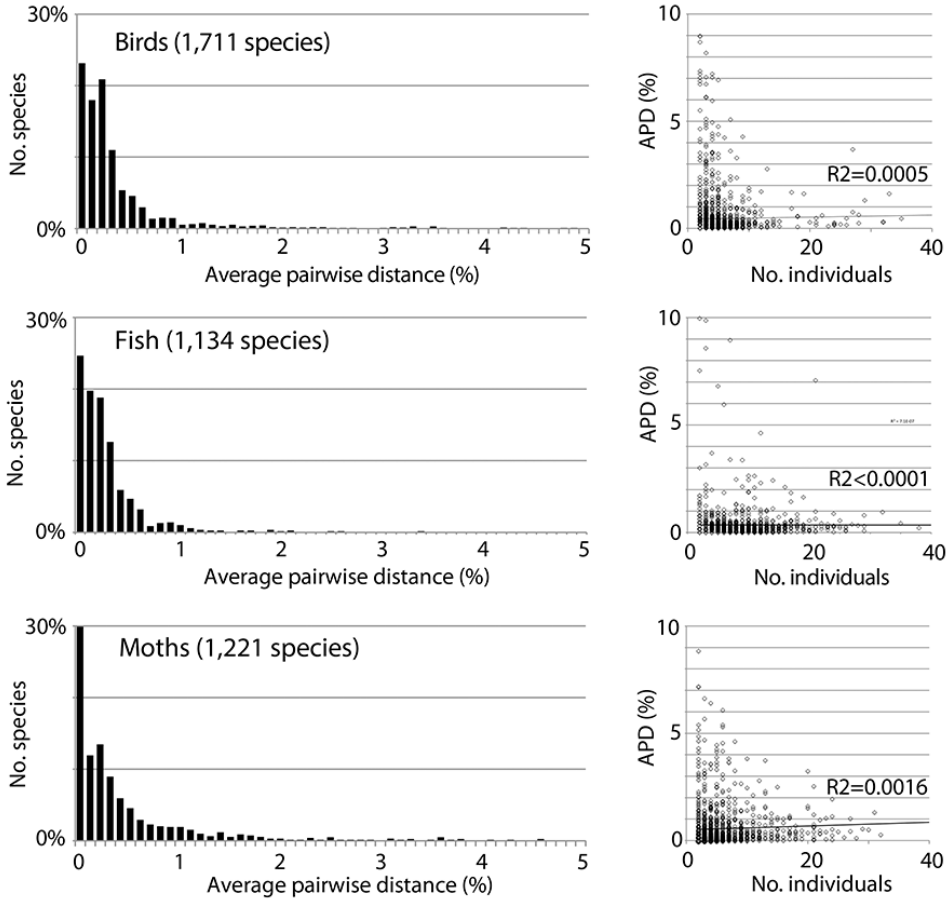

Fig. 1. Low intraspecific COI barcode variation is the norm in animals, not an artifact of handpicking examples or small sample size. Variation is expressed as average pairwise difference (APD) between individuals.

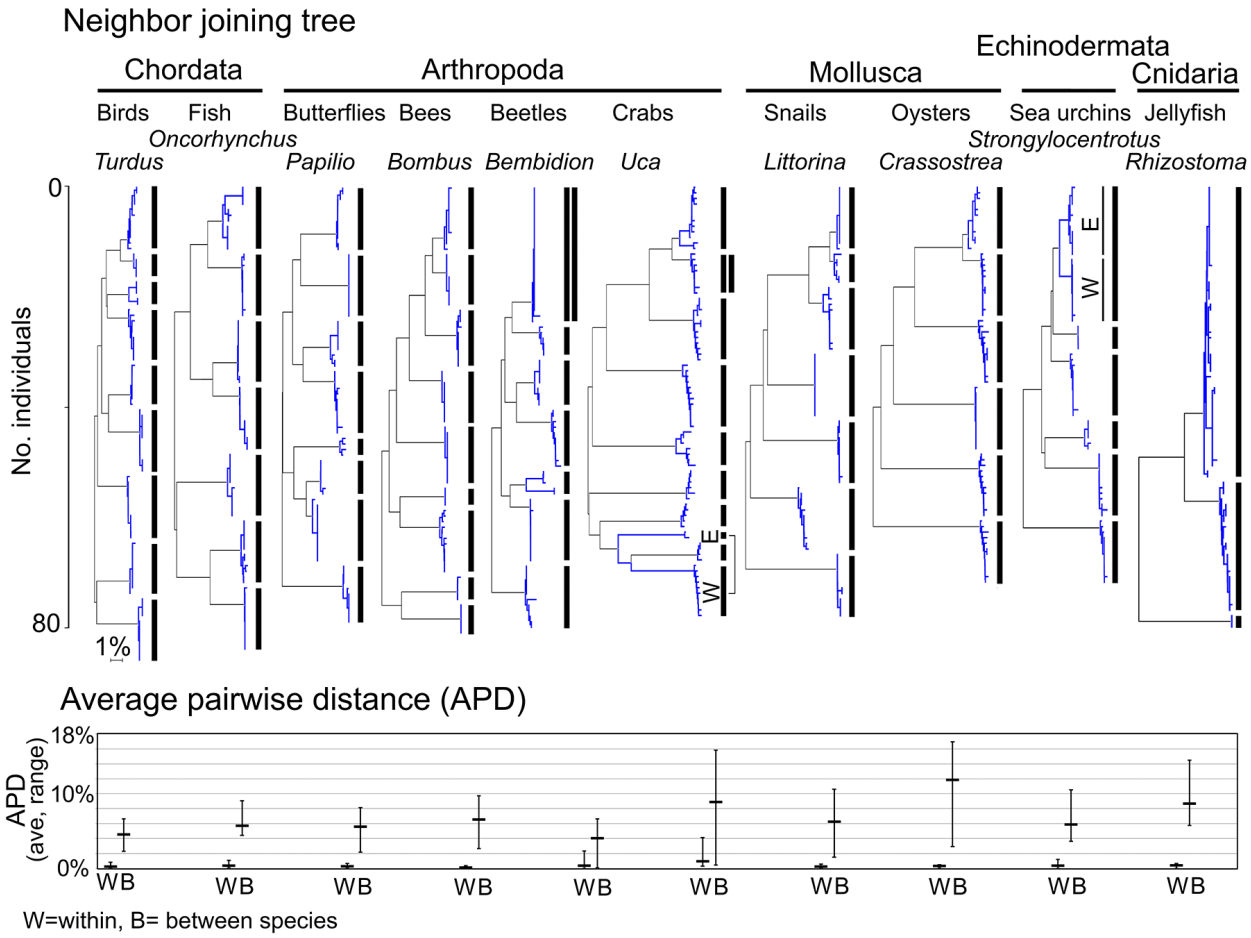

Fig. 2. Relatively large interspecific differences, as compared to uniformly small intraspecific differences, are the norm in animals. Together these yield the familiar clustering pattern that enables DNA barcode species identification**.** Shown are neighbor-joining (NJ) trees and average pairwise distance (APD) within and between sets of closely-related congeneric species. At top, NJ trees. At bottom, APDs for the same congeneric sets, with average (horizontal bar) and range (vertical bar) of intraspecific and interspecific APDs shown. Exceptions to the one species/one cluster rule include cases with multiple clusters within species, corresponding to geographically isolated populations [marked as (W)estern and (E)astern], and cases with clusters shared between species, marked by double vertical lines.

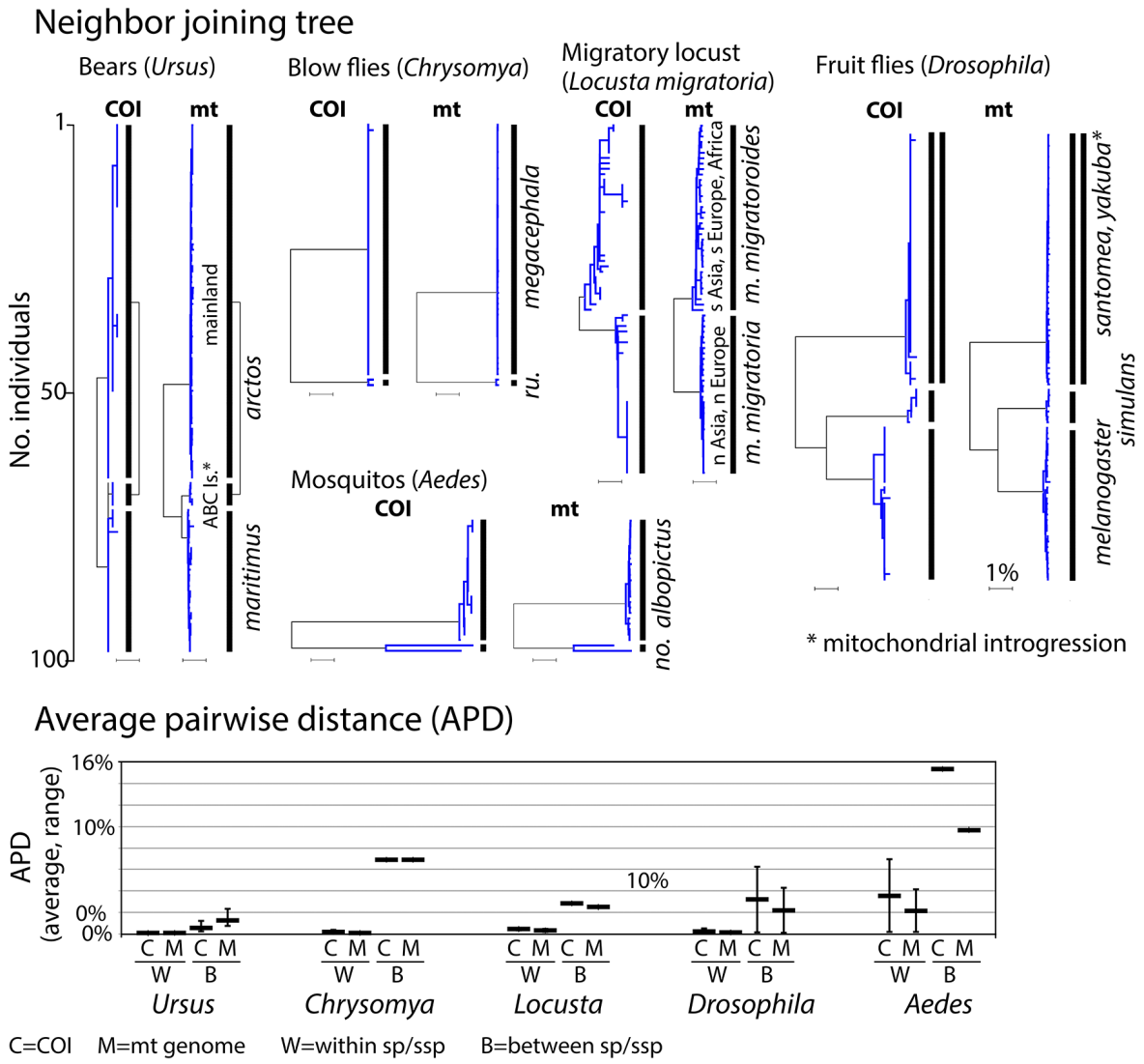

Fig. 3. Clustering of 0.6 kbp COI barcode segments accurately represents the complete 12 kbp coding mitogenome. At top, COI and mt genome NJ trees exhibit similar clustering patterns. At bottom, average pairwise differences within and between species in each set are about the same whether calculated from COI barcodes or coding mitogenomes. As in Fig. 2 legend, apparent exceptions with phylogeographic divisions (locusts) or shared or overlapping clusters (bears, fruit flies) are noted.

Exceptions to the rule that each species is a single unique cluster. Most exceptions to COI-clusters-represent-species are also exceptions to the general rule that species are single interbreeding populations. These include cases with phylogeographic divisions within species and those with shared or overlapping barcode clusters (Figs. 2,3).

In most well-studied cases of shared or overlapping barcodes, nuclear genome analysis demonstrates these anomalies are due to hybridization resulting in mitochondrial introgression from one species into the other. If recent, and complete across the whole population, introgression erases mitochondrial differences between species. Introgression events in the more distant past and those involving only part of a species produce more complex patterns, as illustrated by *Ursus* bears (Fig 3). Based on nuclear and mitochondrial genome analysis, polar bears (*U. maritimus*) hybridized with "ABC island" brown bears (*U. arctos*) about 50,000 years ago, with introgressive replacement of ABC *arctos* mitogenomes by *maritimus* mitogenomes. The mitochondrial lineages subsequently diverged, but ABC island brown bear mtDNA remains closer to polar bear than to mainland brown bears. Nuclear genomic analysis supports taxonomic classification of ABC island and mainland populations as subspecies of brown bear.

Incomplete lineage sorting with retention of ancestral polymorphisms is a plausible mechanism for shared or overlapping mitogenomes, also called paraphyly. However, in all cases we know of, when analyzed for nuclear and mitochondrial differences, ongoing or historical hybridization is considered the likely cause.

General across the animal kingdom. DNA barcodes based on mitochondrial sequences might have failed to be sensitive, general, practical, or to agree with the judgment of experts in each domain. Five million DNA barcodes later some exceptions have been found, however, the power, generality, and validity of the COI barcode approach for identifying animal species is no longer in question, at minimum, for several major groups. A general observation is that barcode clusters correspond best to species in well-studied animal groups, where taxonomists have mostly decided and agreed upon what species are. Thus there is good support in several major phyla, including Chordata, Arthropoda, Mollusca, Echinodermata. We note that these phyla are estimated to contain about ¾ of named animal species.

Incompletely studied groups. In the remaining 23 animal phyla, there are examples where clusters match species, but the overall picture is muddier. Many are small animals, difficult to distinguish morphologically, and have attracted relatively little taxonomic or DNA barcode study. Major incompletely studied groups include Annelida, Nematoda, Platyhelminthes, Porifera, and Rotifera. We expect that with further study these phyla will fit a pattern similar to that in more established groups. However, at this stage it takes cherry-picking to find examples that match the better-studied phyla and one cannot make a data-based case for the general validity of DNA barcoding in these phyla.

Beyond using the DNA barcode as an aid to taxonomy, the enormity of data now available make it appropriate to extend the applications of the "broad but not deep" vista that COI barcode uniquely provides [20, 22, 23]. In a founding document of phylogeography Avise and colleagues noted the long-standing divide in biology between the intellectual lineages of Linnaeus for whom species are discrete entities and those of Darwin who emphasize incremental change within species and leading to new species [4]. They presciently proposed that mitochondrial analysis would provide a way to bridge the intellectual gap. DNA barcoding now provides the most comprehensive database allowing a kingdom-wide and quantitative realization of that vision.
Differing definitions of species. There are approximately 30 different definitions of species in the biological, philosophical, and taxonomical literatures [24]. Almost all of them share the idea

that species are distinct entities in biology and the corollary idea that there are discontinuities among species [25]. In their clarifying and valuable analyses Mayr [25] and de Queiroz [11] point out that all definitions of species involve separate monophyletic evolutionary lineages (with important exceptions where symbiosis or horizontal gene transfer are key [26-29]. Different distinguishing factors such as mating incompatibility, ecological specialization, and morphological distinctiveness evolve, in various cases, in a different temporal sequence. During the process, as species diverge and emerge some of these characteristics will be fulfilled while others are not. Disagreement is inevitable when different properties are considered necessary and sufficient to fit one or another definition of "species".

There are two important observations regarding how COI barcodes fit into the differing definitions of species. First, the cluster structure of the animal world found in COI barcode analysis is independent of any definition(s) of species. Second, domain experts' judgments of species tend to agree with barcode clusters and many apparent deviations turn out to be "exceptions that prove the rule". Controversy around the edges, e.g. disagreements about whether or not borderline cases constitute species or subspecies [30, 31] should not obscure visualizing the overall structure of animal biodiversity. It is unavoidable that some cases will be considered as species by one definition and not another. Controversial cases can illuminate in the context of William Bateson's adage to "treasure your exceptions" [32] but they should not obscure the agreement for most cases and an appreciation of the overall structure within the animal kingdom. This pattern of life, close clustering within individual species with spaces around clusters, can be visualized and demonstrated in different ways and with different statistics (e.g., Figs. 1-4). It qualifies as an empirically-determined evolutionary law [33]. Barcode distribution is arrived at independently but consistent with a view of biology as composed of discrete entities that on different levels include organisms [34] and species [35].

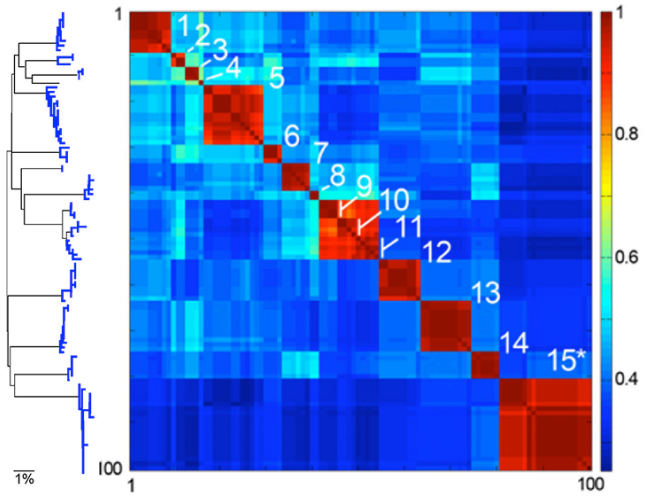

Fig. 4. Species are islands in sequence space. COI barcode NJ tree and Klee diagram of American Robin (*Turdus migratorius*) and closely related *Turdus* species. To generate dataset, a single American robin COI barcode was used to search GenBank using BLAST, and the top 100 matches were downloaded. In Klee diagram, numbers indicate species, asterisk marks *T. migratorius* sequences, and indicator vector correlation scale is at right, with 1 representing 100% sequence identity.

The pattern of DNA barcode variance is the central fact of animal life that needs to be explained by evolutionary theory. In 'The Structure of Scientific Revolutions' Thomas Kuhn makes the point that every scientific model takes certain facts of nature or experimental results as the key ones it has to explain [36]. We take the clustering structure of COI barcodes—small variance within species and often but not always sequence gaps among nearest neighbor species—as the primary fact that a model of evolution and speciation must explain. The pattern of life seen in barcodes is a commensurable whole made from thousands of individual studies that together yield a generalization. The clustering of barcodes has two equally important features: 1) the variance within clusters is low, and 2) the sequence gap among clusters is empty, i.e., intermediates are not found. Beyond the qualitative descriptor "low" for the variance within species there is a quantitative statement. The average pairwise difference among individuals (APD; equivalent to population genetics parameter π) within animal species is between 0.0% and 0.5%. The most data are available for modern humans, who have an APD of 0.1% calculated in the same way as for other animals [37].
The agreement of barcodes and domain experts implies that explaining the origin of the pattern of DNA barcodes would be in large part explaining the origin of species. Understanding the mechanism by which the near-universal pattern of DNA barcodes comes about would be tantamount to understanding the mechanism of speciation.

The clustering pattern of life was elegantly articulated by Dobzhansky in his 1937 book Genetics and the Origin of Species [38] from which an extensive quote is merited. Only through DNA barcodes can the same metric be used so that the "feeling that it must be right" can now be given a single quantitative meaning across the entire animal kingdom:

If we assemble as many individuals living at a given time as we can, we notice that the observed variation does not form a single probability distribution or any other kind of continuous distribution. Instead, a multitude of separate, discrete, distributions are found. In other words, the living world is not a single array of individuals in which any two variants are connected by unbroken series of intergrades, but an array of more or less distinctly separate arrays, intermediates between which are absent or at least rare. Each array is a cluster of individuals, usually possessing some common characteristics and gravitating to a definite modal point in their variation.… Therefore the biological classification is simultaneously a man-made system of pigeonholes devised for the pragmatic purpose of recording observations in a convenient manner and an acknowledgement of the fact of organic discontinuity.

Two models have the potential to explain the structure of COI barcodes in the extant animal kingdom: 1) COI barcode clusters represent species-specific adaptations, 2) extant populations have recently passed through diversity-reducing regimes whose consequences for sequence diversity are indistinguishable from clonal bottlenecks. This way of framing the problem is similar to that raised by the analysis of isozymes by electrophoresis more than half a century previous [39]. The key difference being that the COI barcode data are vast and commensurable across the animal kingdom. "Commensurable" means using the same measurement and being directly comparable. "The "awesome power of" [40] mitochondrial COI barcodes allows the same metric to be used animal kingdom-wide. Without DNA barcodes,

generalizations to all animals have to be based on putting together data sets gathered and analyzed by different methods.

Most barcode variation among neighboring species—and also within species—consists of synonymous codon changes (Fig. 5). The question that determines which of the two mechanisms is most plausible is whether or not synonymous codons in the mitochondrial genome are selectively neutral. If purifying selection does not act on synonymous codons in the mitochondrial genome then population processes must be acting to suppress neutral variance.

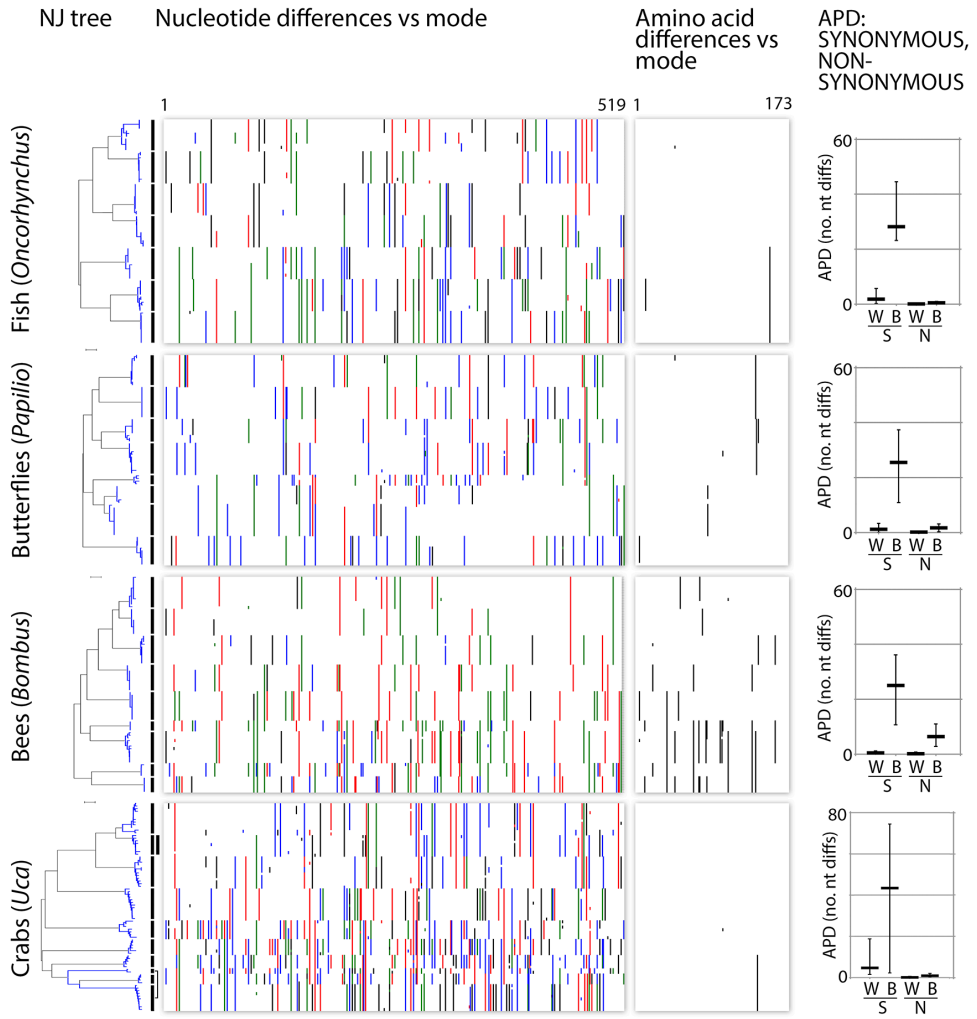

Fig. 5. mtDNA clusters reflect synonymous substitutions. Charts depict nucleotide and amino acid differences from the mode for congeneric COI barcode sets in Fig. 3. Nucleotide differences are colorized (A=green; C=blue; G=black; T=red). To minimize contribution of sequence errors and missing data, the 648 bp barcode region is trimmed by 10% at either end, leaving 519 nt/173 amino acids. At right, synonymous (S) and non-synonymous (N) average pairwise distances within (W) and between (B) species. Horizontal bar indicates mean and vertical line indicates maximum and minimum.

Are synonymous codons in mitochondria neutral? Comparative rates along phylogenies have been used to argue that amino acid changes in the mitochondrial genome are subject to purifying selection but synonymous substitutions are not [23, 41, 42]. Across the animal kingdom the preponderance of SNP variation in mitochondrial sequences consists of synonymous codon changes. Are these synonymous codons targets for purifying and/or adaptive selection strong enough to be responsible for the low variance within species and/or the different consensus sequence among neighboring species? Codon bias in the mitochondrial genome has been shown at the phylogenetic level of order but there is no evidence for different codon bias among neighboring species [43, 44]. Furthermore, the number of synonymous codons relevant to the discussion of DNA barcodes (0.0%-0.5% within species, 0.0%-5.0% for neighboring species) is not enough to alter codon bias. Nearby species do not differ in overall codon bias or GC ratios. If synonymous codons are differentially selected in DNA barcodes, this selection must be acting at the level of the placement of individual codons rather than their cumulative average effects on base composition.

An assumption of neutrality for synonymous codons is no longer a "slam dunk" [45], i.e., is not a certain conclusion. Avise et al in 1987 made an absolutist statement on the irrelevance of synonymous codons to selective fitness:

First, in a mechanistic sense, we already "know" that most of the particular mtDNA genotypic variants segregating in populations probably have, by themselves, absolutely no differential effect on organismal fitness. These include, for example, base substitutions in silent positions of protein-coding genes, and some substitutions and small addition/ deletions in the nontranscribed D-Ioop region. These changes are disproportionately common in mtDNA (25) and are ones for which only the most ardent selectionist would argue a direct link to organismal fitness.

This statement merits critical evaluation in light of three intervening decades of molecular genetics. (Spoiler alert: in this case re-examination strengthens the original assertion for reasons that the authors themselves could not have anticipated.)

Since 1987 numerous examples have emerged where even very few synonymous codon changes make important and selectable differences in organismal, cellular, or viral physiology [46-52]. Synonymous codons may also modulate protein folding or membrane insertion concomitant with translation [53], as suggested for a synonymous codon that modulates the severity of cystic fibrosis [54]. The cited cases involve cytoplasmic translation of bacteria, bacteriophage, and drosophila nuclear genes. However, they require one to critically re-examine the assertion of absolute and universal neutrality with regard to mitochondria.

New justification for an old assertion. The proven cases where synonymous codons have phenotypes are all consequent to their differential rate of translation. Codon-specific translation rates are in turn attributed to different concentrations of codon-specific tRNAs, known as "isoacceptors." This has been proven directly in an experimental system where overexpression of the cognate tRNA changes a low-expression codon to a high one [55, 56]. Isoacceptor concentrations differ between species and even within individuals in a tissue-specific manner [57, 58]. The frequencies of synonymous codons and the concentration of their isoacceptor tRNAs with complementary anticodons coevolve [59]. Codon-specific isoacceptor tRNAs are tightly and dynamically regulated; they play important roles in the differential regulation of gene

expression [60]. The human nuclear genome encodes tRNAs with 51 distinct anticodons for the 20 amino acids [61]. In addition to isoacceptors there are dozens to hundreds of "isodecoder" genes in the nuclear genome. Isodecoders are tRNAs that share an anticodon sequence but differ elsewhere. The different sequences of isodecoders are often also associated with different post-transcriptional modifications [61]. It is likely that isodecoders add a further important layer to differential gene expression depending on a codon's sequence and tissue context but this remains to be proven.

In striking contrast to the multitude of different nuclear tRNA genes and cytoplasmic tRNAs animal mitochondria have only 22 different tRNA types to translate the 20 amino acids [62, 63]. With two exceptions, isoacceptor tRNAs are not available inside animal mitochondria. Leucine and serine each have two mitochondrial tRNAs with different anticodons; the remaining 18 amino acids are each translated by a single tRNA that covers all cognate codons. The best documented mechanism for altering the efficiency of translation is when the changed codon(s) are near to the first-translated end of the gene [55]. The approximately even distribution of synonymous variation among mitochondrial genes in modern humans [37] is most compatible with neutrality [64].

Speculatively, synonymous codon changes could affect gene expression by mechanisms independent of tRNAs. These include changes in mRNA secondary structure and stability and the binding of specific factors, protein or miRNA. Modification of splicing is a candidate and possibly important mechanism by which synonymous codons alter protein structure and function. However, in contrast to most nuclear genes in most animals, animal mitochondrial protein-encoding genes do not have introns. In animal mitochondria there are no alternative spliced forms whose ratios could be modulated by synonymous codons near splice sites.

Kimura's insight that a preponderance of synonymous substitutions is evidence for neutral evolution [65] now appears to be more universally valid for the mitochondrial than for the nuclear genome. For nuclear genomes one finds a growing number of cases and mechanisms where synonymous codons have phenotypes and are subject to selection. In contrast, for the mitochondrial genomes of animals there is not a single example of any synonymous codon having a phenotype. Furthermore, the known mechanisms that allow synonymous codons to alter the phenotypes of nuclear genes are impossible in mitochondria. Mitochondrial sequences yield straightforward and uncomplicated phylogenetic analysis and species-level identification for reasons beyond those known by those who originally proposed them. Sometimes you get lucky.

The case of the missing G's. Codons that end in G are underrepresented by a factor of about ten in animal mitochondria. Previously we interpreted the lack of third position G's in mitochondrial coding sequences is evidence of a role for extreme purifying selection in determining the COI DNA barcoding gap [20] but we now find this argument flawed. On the one hand there appears to be purifying selection against codons that end in G but apparent selection is similar neighboring species. With selection against G-ending codons in all animal species it could not be a source of species-specific adaptive peaks. Further insight into the lack of G in the third position follows from a focused review of the wobble hypothesis in the context of mitochondria. Francis Crick set forth a set of stereochemical models in which a single tRNA anticodon pairs with multiple codons for the same amino acid. Crick called his idea the "Wobble hypothesis"

because it postulated flexible pairing between the 3' base of the codon with the 5' base of the anticodon [66]. The Wobble hypothesis is brilliantly insightful, however, details of the pairing scheme have changed with knowledge of extensive post-transcriptional chemical modifications of tRNAs. More than 150 different chemical modifications of RNAs are now characterized [67]; the greatest concentration of RNA modification is found on anticodons. Only certain modifications at the U at the 5' position in the anticodon allow efficient Wobble-pairing with G [68-71]. Wobble G is rare in animal mitochondrial codons[20] consistent with the fact that Wobble G recognition-specific modifications have not been found in animal mitochondrial tRNAs [72, 73].

Several human pathologies are correlated by Genome Wide Association Studies (GWAS) to SNPs that change synonymous codon and it is expected that more will be found[74]. Two cautions apply when considering Genome Wide Association Studies linking synonymous codons with human mitopathologies: 1) So far as we are aware there are no inferences of pathologies based on synonymous substitutions in the mitogenome [75], 2) GWAS are subject to artifacts of inference that encourage erroneous confidence [76, 77]. GWAS are hypothesis generators, not proof. Anecdotally, workers in mitochondrial pathologies are well aware of synonymous codons, and so the absence of evidence for any human pathology due to a synonymous codon change in the mitochondrial genome is not due to a lack of looking. In contrast, missense mutations in mitochondrial tRNA genes are "hotspots" for human pathologies [78]; they lead to large scale insertion of inappropriate amino acids.

Selectionist arguments. Thomas Kuhn points out that science stalls when different camps that study the same aspects of nature use different vocabularies, cite only within their own discipline and ignore or disparage each other. There appears to be an unfortunate isolation in the literature between camps that advocate mitochondrial selection and those that rely on demographic reasoning. Here we argue that selectionist perspectives are valid in some mitochondrial systems and plausible in others. However, exceptions abound and as a whole we find that a selectionist perspective is not robust enough to account for the animal-kingdom wide facts of the barcode gap.

Even within a single species different external environments may select for particular alleles of mitochondrially-encoded enzymes [79-82]. In modern humans there are two different amino acid alleles of the mitochondrially-encoded ATPase. This subunit partitions the proton gradient of mitochondria in two ways: it can use the gradient to form the covalent to join inorganic phosphate to ADP in order to make ATP. Alternatively, if the proton gradient runs down without storing energy in the synthesis of ATP heat is immediately released. The allele predisposed to ATP synthesis is more frequent in human populations who inhabit tropical regions. Conversely, the allele biased toward instantaneous heat generation is more frequent in colder regions [83]. The argument for environmentally-driven selection for this allele is logical and inspires interesting experiments [84]. The plausible but unproven possibility of selection for a single allelic case of amino acid substitution is a small pebble in the scale when compared to the evidence for the apparent neutrality of most mitochondrial variation. Most human mitochondrial variation, similar to that of other animal species, consists of synonymous codon changes. However, in principle, the linkage of a single selected amino acid could drive a species-wide sweep of the entire linked mitochondrial genome.

The model of an optimum sequence for each species has two subcategories: a) optimal for the external environment, and b) optimal integration with other genes of the organism [85]. These two mechanisms can work together and various permutations have been suggested with more or less emphasis on selection for external conditions such as environmental temperature or internal compatibility with nuclear genes [81, 86-88]. Compatibility among the thousand or so nuclear genes whose products enter the mitochondria and the 13 gene products coded for by the mitochondrial genome can lead to reproductive isolation and incompatibility [21, 87, 89]. Incompatibility of mitochondrial and nuclear genes can cause reproductive isolation either immediately or via decreased fitness of progeny [88, 90-93]. Mitochondrial introgression in some cases has been proposed to favor the co-introgression of compatible nuclear alleles that form subunits of mitochondrial complexes [94, 95]. This is a potentially important perspective but its generality is unclear.

We see arguments against stepping from specific examples of nuclear-mitochondrial incompatibilities to a general theory of this effect as the major driver of animal speciation. 1) There are many examples of fertile and fit interspecies hybridizations including cases with mitochondrial sequences that diverge by 4% or more [96-103]. 2) Some well-studied cases of mating incompatibility have nothing to do with nuclear-mitochondrial incompatibility. Examples include chromosomal inversion [104, 105] and the activation of endogenous retroviruses-transposons [106]. 3) Nuclear-mitochondrial incompatibility is a subset of physiological incompatibility. Other definitions of species, e.g. behavioral, geographical, ecological, do not require physiological mating incompatibility in any form [11]. COI barcode clustering is more widely a fact than can be accounted for by mating incompatibility in general and nuclear-mitochondrial incompatibility in particular. 5) Finally, there is no example in which mating incompatibility or weakness of inter-species hybrids is attributable to the synonymous codons that constitute the major fraction of barcode gaps.

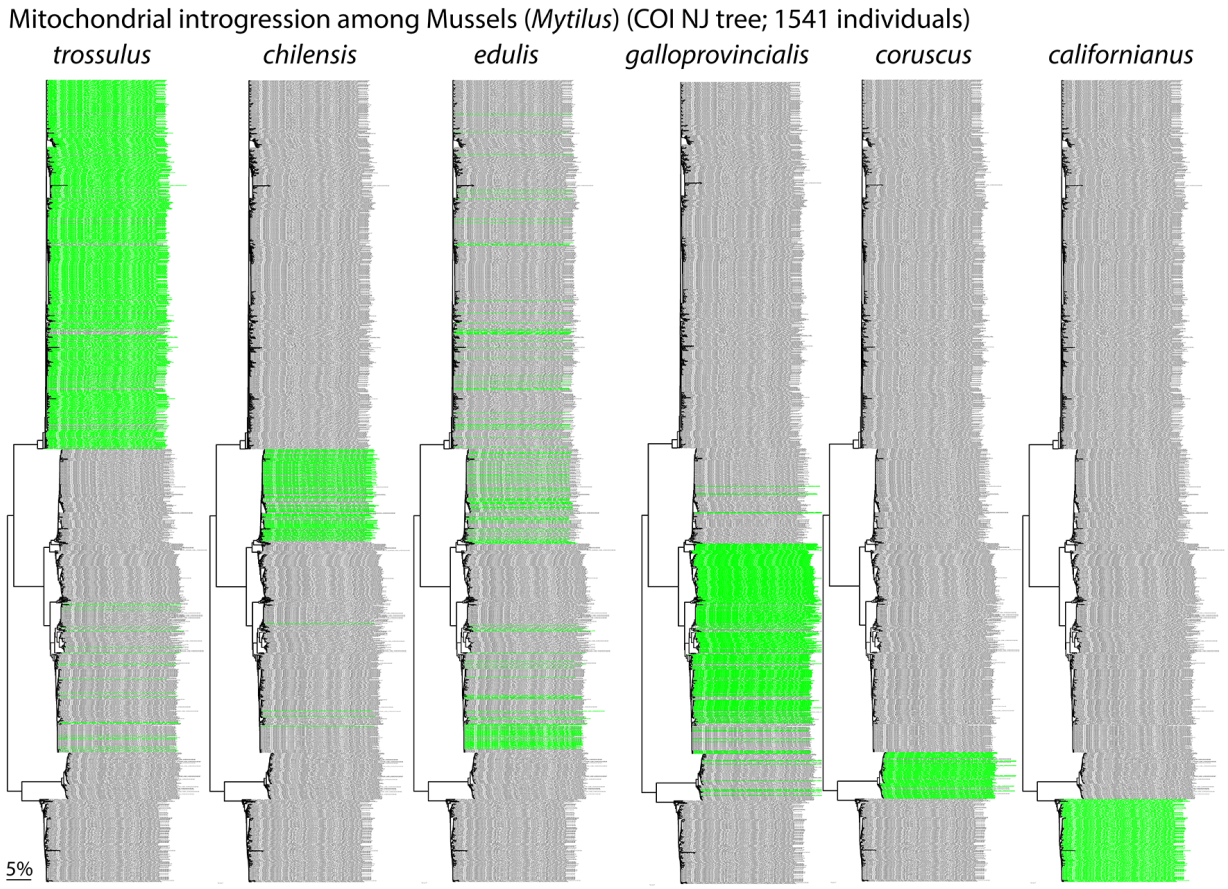

Fig. 6. Fertile hybrids. *Mytilus* mussels exhibit complex patterns of mitochondrial and nuclear introgression, reflecting multiple historical and recent hybridization events, some following introduction of non-native species for aquaculture. F1 hybrids are fertile even though parental species differ by 10-20% in COI nucleotide sequence. This supports view that mtDNA clustering is not due to species-specific adaptations.

The average pairwise difference of the COI barcode in modern humans is 0.1%, i.e., about average for the animal kingdom. However, the most extreme differences between individual humans approach 1%. This difference is as great as many distinctions among neighboring species. Modern humans are a single population. Darwin made this point with respect to visible phenotypes and it applies even more strongly when neutral variants are considered:

Hereafter, we shall be compelled to acknowledge that the only distinction between species and well-marked varieties is, that the latter are known, or believed, to be connected at the present day by intermediate gradations, whereas species were formerly thus connected [107].

The possibility of preferred combinations of nuclear and mitochondrial alleles within a species is intriguing and there is one example of experimental support. An inbred strain of mouse was shown to have non-optimal physiology when the mitochondrial genome from a different inbred line was crossed in (10 backcrosses to the nuclear line all using the female descendent from the first mitochondrial donor) [108]. This finding has been extrapolated as justification to urge

studies of nuclear-mitochondrial compatibility in human three-parent IVF (*in vitro* fertilization) [109]. On the other hand, human mitochondrial transfer experiments have found no analogous effect [110], arguably, owing to the different genetic structure of our species when compared to inbred mouse strains [111]. The differences in the two mouse mitochondrial genomes at issue include missense in the coding region, tRNA alterations and ori-region changes as well as synonymous codon changes. There are no data to pinpoint which sequences make a difference, in particular no evidence for a phenotype of synonymous codon changes, which the authors mark as "silent" (Extended Data Table 1 in [108]).

This line of work and controversy adds evidence that in some cases mitochondrial-nuclear incompatibility may interfere with mating or health of offspring. However, the work does not show any effect of synonymous codon changes. No matter which mechanism for speciation is responsible in any specific case, the 0.0%-0.5% accumulation of synonymous variance independent of population size or apparent species age is a biological fact. The variable distance between the most closely related living species presumably reflects differing numbers of extinct intermediate sequences.

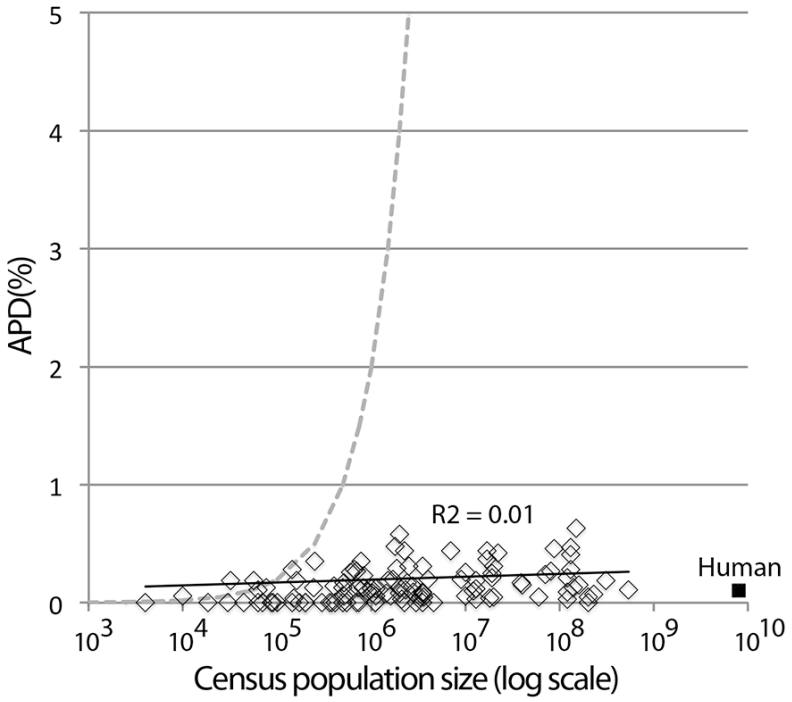

Fig. 7. Kimura's equilibrium model alone is insufficient to account for usual levels of intraspecific variation in animal species. APD and census population size for 112 bird species without phylogeographic clusters are shown. Dashed line is expected APD limit due to lineage sorting (AVP = 2 N μ, where N = population size, and μ = mutation rate, using 10^-8 substitutions/site/generation, or 1% per My, assuming generation time is 1 y). Average effective population size in the birds shown is 70 thousand (range 0-300 thousand); average census population size is 30 million (range 5 thousand to 500 million). Human mitochondrial variation (population 7.5 billion, APD 0.1%) is typical of that in other animal species.

Conditions that favor clonal uniformity are frequent in biology. Bottlenecks, founder effects, lineage sorting, and gene sweeps decrease genetic diversity in a population [112]. The question is how widespread these effects are in the context of defining animal species and if it is possible distinguish them in other than a rhetorical manner. Here we emphasize the overlap—in fact the near congruence—in the conditions that favor each of these mechanisms.

Based on contemporary mitochondrial sequence data alone it is impossible to distinguish an organismal bottleneck from mitochondrial and Y chromosome specific lineage sorting since both mechanisms make the same prediction of a uniform mitochondrial sequence in the past [113].

A positively selected allele has the potential to sweep through a population and by hitchhiking [114, 115] or genetic draft [116] carry the entirety of the linked genome along thereby resetting mitochondrial variation to zero. This scenario requires a single maternal lineage replace all others [114]. It is reasonable to hypothesize that somewhere on the entire mitochondrial genome there arises a positively selected amino acid substitution leading to the replacement of the entire linked genome in the entire population. One should not mince words about what a mitochondrial genome sweep requires: the entire population's mitochondrial genome must re-originate from a single mother.

These three pathways toward sequence uniformity should not be thought of as enemies because they converge in both cause and effect. Lineage sorting is most efficient when the population is small, when the number of different mitochondrial genotypes is small, and when the population is either stable or shrinking [117]. An extreme diminishment of population size followed by population expansion is the definition of a bottleneck. Lineage sorting is diminished during periods of population growth and does not occur at all during exponential growth when all neutral lineages leave progeny [118, 119]. The same conditions that favor lineage sorting also favor gene sweeps, which in the context of a totally-linked genome means one mitochondrial genome. The concept of "gene sweeps" emphasizes positive selection whereas "lineage sorting" emphasizes neutral events. Bottlenecks are extreme forms of the same conditions.

Bottlenecks followed by expansion are the dominant mechanism for evolution in the microbial majority of life and it might seem odd to think animals should be exceptional [120]. Ever since Koch, microbiologists have streaked out their bacteria to begin experiments with pure, i.e., clonal cultures [121]. The first experiments showing evolution of new mutants from clonal starting populations were the classical cases of proof that bacterial genes follow the patterns expected from random mutation that grow indistinguishably from sibs when unselected ([118, 122-124]). Clonal outgrowth and replacement of the inoculating population was inferred from the earliest chemostat [125] as well as later serial transfer experiments [126]. Epidemiology shows that repeated bottlenecks play dominant roles in the natural evolution of microbial pathogens including protists, bacteria and viruses [127-133]. A visually impressive demonstration of successive clonal selection and population outgrowth is seen in time lapse studies of bacteria serially mutating to new heights of antibiotic resistance [134]. On the host side of the equation, the clonal selection theory of immune system development was controversial when first proposed but its logic and experimental support proved compelling [135, 136].

Mayr made a specific proposal for the role of extreme bottlenecks in speciation that followed from a founder effect (originally 1942, here quoted from a reprise based on interviews in 2004 [137]):

The reduced variability of small populations is not always due to accidental gene loss, but sometimes to the fact that the entire population was started by a single pair or by a single fertilized female. These "founders" of the population carried with them only a very small proportion of the variability of the parent population. This "founder" principle sometimes explains even the uniformity of rather large populations, particularly if they are well isolated and near the borders of the range of the species.

Eldredge and Gould used this idea of allopatric speciation in small isolated populations that then rapidly expanded to account for the abrupt transitions seen in the broad range of the fossil record [138].

Models of allopatric or peripatetic speciation invoke a bottleneck with an additional feature: What emerges from the bottleneck looks or acts differently, i.e., it is a bona fide new species. It may be more frequent that what emerges from a bottleneck looks and acts like a middling representative of what went in.

If mitochondria are considered "honorary prokaryotes" then their dominant mode of frequent processes that lead to clonal selection and outgrowth are not counterintuitive. Whether sequence uniformity occurred via differential growth of 100%-linked organelle genomes, linked genomes within organelles, organismal bottlenecks, lineage sorting through many generations of an organismal population the result is the same: The mitochondrial sequence frequently becomes uniform.

Purifying selection in linked genomes slows but does not stop the accumulation of neutral variation. Drift and lineage sorting during population stasis or shrinkage decrease variation. The efficiency of decrease depends on the number of haplotypes in the population, as well as the numbers and distributions of female offspring among parents with different haplotypes[10]. A key prediction of naïve neutral theory that does not hold up against extensive barcode data from across the animal kingdom is that larger populations or older species should harbor more neutral variation [20, 139, 140]. The key incompatibility of naïve neutral theory with biological fact is that the theory considers populations at equilibrium in the sense that the population be at stable numbers for approximately as many generations as the mutation rate per generation. The evolution of modern humans offers a specific solution to the animal-kingdom-wide dilemma of missing neutral mutations.

Modern humans. More approaches have been brought to bear on the emergence and outgrowth of *Homo sapiens sapiens* (i.e. the modern human population) than any other species including full genome sequence analysis of thousands of individuals and tens of thousands of mitochondria, paleontology, anthropology, history and linguistics [62, 141-143]. The congruence of these fields supports the view that modern human mitochondria and Y chromosome originated from conditions that imposed a single sequence on these genetic elements between 100,000 and 200,000 years ago [144-146]. Contemporary sequence data cannot tell whether mitochondrial

and Y chromosomes clonality occurred at the same time, i.e. consistent with the extreme bottleneck of a founding pair, or via sorting within a founding population of thousands that was stable for tens of thousands of years [117]. As Kuhn points out unresolvable arguments tend toward rhetoric.

Summary and conclusion. Science greedily seizes simplicity among complexity. Speciation occurs through via alternative pathways distinct in terms of the number of genes involved and the abruptness of transitions [147]. Nuclear variance in modern humans varies by loci in part due to unequal selection [148] and the linkage of neutral sites to those that undergo differential selection. Complexity is the norm when dealing with variance of the nuclear ensemble [149-153]. It is remarkable that despite the diversity of speciation mechanisms and pathways the mitochondrial sequence variance in almost all extant animal species should be constrained within narrow parameters.

Mostly synonymous and apparently neutral variation in mitochondria within species shows a similar quantitative pattern across the entire animal kingdom. The pattern is that that most—over 90% in the best characterized groups—of the approximately five million barcode sequences cluster into groups with between 0.0% and 0.5% variance as measured by APD, with an average APD of 0.2%.

Modern humans are a low-average animal species in terms of the APD. The molecular clock as a heuristic marks 1% sequence divergence per million years which is consistent with evidence for a clonal stage of human mitochondria between 100,000- 200,000 years ago and the 0.1% APD found in the modern human population [37, 154, 155]. A conjunction of factors could bring about the same result. However, one should not as a first impulse seek a complex and multifaceted explanation for one of the clearest, most data rich and general facts in all of evolution. The simple hypothesis is that the same explanation offered for the sequence variation found among modern humans applies equally to the modern populations of essentially all other animal species. Namely that the extant population, no matter what its current size or similarity to fossils of any age, has expanded from mitochondrial uniformity within the past 200,000 years. Nonhuman animals, as well as bacteria and yeast, are often considered “model systems” whose results can be extrapolated to humans. The direction of inference is reversible. Fossil evidence for mammalian evolution in Africa implies that most species started with small founding populations and later expanded [156] and sequence analysis has been interpreted to suggest that the last ice created widespread conditions for a subsequent expansion [157]. The characteristics of contemporary mitochondrial variance may represent a rare snapshot of animal life evolving during a special period. Alternatively, the similarity in variance within species could be a sign or a consequence of coevolution [158].

Mitochondria drive many important processes of life[159] [160] [161]. There is irony but also grandeur in this view that, precisely because they have no phenotype, synonymous codon variations in mitochondria reveal the structure of species and the mechanism of speciation. This vista of evolution is best seen from the passenger seat.

Acknowledgements. Bruce Levin suggested the phrase “honorary prokaryote” in reference to mitochondria. Others have used the phrase in reference yeast or to phage introns. Thanks to Glen Bjork, Manny Goldman, Ken Zahn for discussions on tRNAs, Jesse Ausubel, Frank Stahl for encouragement and comments and the Sloan Foundation for support.

References

1. Avise, J.C. and K. Wollenberg, *Phylogenetics and the origin of species.* Proc Natl Acad Sci U S A, 1997. **94**(15): p. 7748-55.
2. Hebert, P.D., et al., *Biological identifications through DNA barcodes.* Proc Biol Sci, 2003. **270**(1512): p. 313-21.
3. Stoeckle, M., *Taxonomy, DNA, and the Bar Code of Life.* BioScience, 2003. **53**: p. 2-3.
4. Avise, J.C., et al., *Intraspecific phylogeography: the mitochondrial bridge between population genetics and systematics.* Ann. Rev. Ecol. Syst., 1987. **18**: p. 489–522.
5. Moore, W.S., *Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees.* Evolution, 1995. **49**: p. 718–726.
6. Gladwell, M., *The Tipping Point: How Little Things Can Make a Big Difference*. 2000: Little Brown. 279.
7. Hebert, P.D., et al., *Identification of Birds through DNA Barcodes.* PLoS Biol, 2004. **2**(10): p. e312.
8. Galtier, N., et al., *Mitochondrial DNA as a marker of molecular diversity: a reappraisal.* Molec. Ecol., 2009. **18**: p. 4541–4550.
9. Hickerson, M.J., C.P. Meyer, and C. Moritz, *DNA barcoding will often fail to discover new animal species over broad parameter space.* Syst Biol, 2006. **55**(5): p. 729-39.
10. Avise, J.C., B.W. Bowen, and F.J. Ayala, *In the light of evolution X: Comparative phylogeography.* Proc Natl Acad Sci U S A, 2016. **113**(29): p. 7957-61.
11. de Queiroz, K., *Species Concepts and Species Delimitation.* Syst. Biol., 2007. **56**: p. 879–886.
12. Albert, B., et al., *Dynamics of plant mitochondrial genome: model of a three-level selection process.* Genetics, 1996. **144**(1): p. 369-82.
13. Taylor, D.R., C. Zeyl, and E. Cooke, *Conflicting levels of selection in the accumulation of mitochondrial defects in Saccharomyces cerevisiae.* Proc Natl Acad Sci U S A, 2002. **99**(6): p. 3690-4.
14. Birky, C.W., Jr., *On the origin of mitochondrial mutants: evidence for intracellular selection of mitochondria in the origin of antibiotic-resistant cells in yeast.* Genetics, 1973. **74**(3): p. 421-32.
15. Otten, A.B., et al., *Differences in Strength and Timing of the mtDNA Bottleneck between Zebrafish Germline and Non-germline Cells.* Cell Rep, 2016. **16**(3): p. 622-30.
16. Rebolledo-Jaramillo, B., et al., *Maternal age effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA.* Proc Natl Acad Sci U S A, 2014. **111**(43): p. 15474-9.
17. Vevea, J.D., et al., *Inheritance of the fittest mitochondria in yeast.* Trends Cell Biol, 2014. **24**(1): p. 53-60.
18. Ballard, J.W., B. Chernoff, and A.C. James, *Divergence of mitochondrial dna is not corroborated by nuclear dna, morphology, or behavior in Drosophila simulans.* Evolution, 2002. **56**(3): p. 527-45.
19. Song, H., et al., *Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified.* Proc Natl Acad Sci U S A, 2008. **105**(36): p. 13486-91.
20. Stoeckle, M.Y. and D.S. Thaler, *DNA barcoding works in practice but not in (neutral) theory.* PLoS One, 2014. **9**(7): p. e100755.

21. Hill, G.E., *Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap.* Ecol Evol, 2016. **6**(16): p. 5831-42.
22. Hebert, P.D., P.M. Hollingsworth, and M. Hajibabaei, *From writing to reading the encyclopedia of life.* Philos Trans R Soc Lond B Biol Sci, 2016. **371**(1702).
23. Pentinsaari, M., et al., *Molecular evolution of a widely-adopted taxonomic marker (COI) across the animal tree of life.* Sci Rep, 2016. **6**: p. 35275.
24. Zachos, F.E., *Species Concepts in Biology: Historical Development, Theoretical Foundations and Practical Relevance*. 2016, Switzerland: Springer International. 220.
25. Mayr, E. *The species problem: A symposium presented at the Atlanta meeting of the American Association for the Advancement of Science, December 28–29, 1955*. 1957. Washington, DC: American Association for the Advancement of Science.
26. Margulis, L., *Symbiosis in Cell Evolution*. 1981: WH Freeman. 419.
27. Sapp, J., *Saltational symbiosis.* Theory Biosci, 2010. **129**(2-3): p. 125-33.
28. Gilbert, S.F., J. Sapp, and A.I. Tauber, *A symbiotic view of life: we have never been individuals.* Q Rev Biol, 2012. **87**(4): p. 325-41.
29. Thaler, D.S., *The cytoplasmic structure hypothesis for ribosome assembly, vertical inheritance, and phylogeny.* Bioessays, 2009. **31**(7): p. 774-83.
30. Groves, C., *Species concept in primates.* Am J Primatol, 2012. **74**(8): p. 687-91.
31. Zachos, F.E., *Taxonomy: Species splitting puts conservation at risk.* Nature, 2013. **494**(7435): p. 35.
32. Bateson, W., *The Methods and Scope of Genetics*. 1908: Cambridge University Press.
33. Van Valen, L., *A new evolutionary law.* Evolutionary Theory, 1973. **1**: p. 1-30.
34. Buss, L.W., *The Evolution of Individuality*. 1987: Princeton University Press. 201.
35. Gould, S., *The Structure of Evolutionary Theory*. 2002: Belknap Press. 1464.
36. Kuhn, T., *The Structure of Scientific Revolutions* 1962: University of Chicago Press. 226.
37. Thaler, D.S. and M.Y. Stoeckle, *Bridging two scholarly islands enriches both: COI DNA barcodes for species identification versus human mitochondrial variation for the study of migrations and pathologies.* Ecology and Evolution, 2016. **6**: p. 6824–6835.
38. Dobzhansky, T., *Genetics and the origin of species 3rd ed*. 1937, New York: Columbia University Press.
39. Lewontin, R.C., *Twenty-five years ago in Genetics: electrophoresis in the development of evolutionary genetics: milestone or millstone?* Genetics, 1991. **128**(4): p. 657-62.
40. Stahl, F. and C. Gross, *Ira Herskowitz (1946-2003).* Cell, 2003. **114**: p. 9-10.
41. Castellana, S., S. Vicario, and C. Saccone, *Evolutionary patterns of the mitochondrial genome in Metazoa: exploring the role of mutation and selection in mitochondrial protein coding genes.* Genome Biol Evol, 2011.
42. Kerr, K.C., *Searching for evidence of selection in avian DNA barcodes.* Mol Ecol Resour, 2011. **11**(6): p. 1045-55.
43. Selva Kumar, C., et al., *Influence of certain forces on evolution of synonymous codon usage bias in certain species of three basal orders of aquatic insects.* Mitochondrial DNA, 2012. **23**(6): p. 447-60.
44. Wei, L., et al., *Analysis of codon usage bias of mitochondrial genome in Bombyx mori and its relation to evolution.* BMC Evol Biol, 2014. **14**: p. 262.
45. *Scientific language is becoming more informal.* Nature, 2016. **539**(7628): p. 140.
46. Saikia, M., et al., *Codon optimality controls differential mRNA translation during amino acid starvation.* RNA, 2016.

47. Komar, A.A., *The Yin and Yang of codon usage.* Hum Mol Genet, 2016. **25**(R2): p. R77-R85.
48. Supek, F., *The Code of Silence: Widespread Associations Between Synonymous Codon Biases and Gene Function.* J Mol Evol, 2016. **82**(1): p. 65-73.
49. Bull, J.J., *Evolutionary reversion of live viral vaccines: Can genetic engineering subdue it?* Virus Evol, 2015. **1**(1).
50. Li, M., et al., *Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11.* Nature, 2012. **491**(7422): p. 125-8.
51. Napolitano, M.G., et al., *Emergent rules for codon choice elucidated by editing rare arginine codons in Escherichia coli.* Proc Natl Acad Sci U S A, 2016. **113**(38): p. E5588-97.
52. Lajoie, M.J., D. Soll, and G.M. Church, *Overcoming Challenges in Engineering the Genetic Code.* J Mol Biol, 2016. **428**(5 Pt B): p. 1004-21.
53. Quax, T.E., et al., *Codon Bias as a Means to Fine-Tune Gene Expression.* Mol Cell, 2015. **59**(2): p. 149-61.
54. Bali, V., et al., *A synonymous codon change alters the drug sensitivity of DeltaF508 cystic fibrosis transmembrane conductance regulator.* FASEB J, 2016. **30**(1): p. 201-13.
55. Zahn, K. and A. Landy, *Modulation of lambda integrase synthesis by rare arginine tRNA.* Mol Microbiol, 1996. **21**(1): p. 69-76.
56. Powell, J.R. and K. Dion, *Effects of codon usage on gene expression: empirical studies on Drosophila.* J Mol Evol, 2015. **80**(3-4): p. 219-26.
57. Bulmer, M., *The selection-mutation-drift theory of synonymous codon usage.* Genetics, 1991. **129**(3): p. 897-907.
58. Bulmer, M., *Coevolution of codon usage and transfer RNA abundance.* Nature, 1987. **325**(6106): p. 728-30.
59. Krakauer, D.C. and V.A. Jansen, *Red queen dynamics of protein translation.* J Theor Biol, 2002. **218**(1): p. 97-109.
60. Wilusz, J.E., *Controlling translation via modulation of tRNA levels.* Wiley Interdiscip Rev RNA, 2015. **6**(4): p. 453-70.
61. Parisien, M., X. Wang, and T. Pan, *Diversity of human tRNA genes from the 1000-genomes project.* RNA Biol, 2013. **10**(12): p. 1853-67.
62. Lott, M.T., et al., *mtDNA Variation and Analysis Using Mitomap and Mitomaster.* Curr Protoc Bioinformatics, 2013. **44**: p. 1 23 1-26.
63. Salinas-Giege, T., R. Giege, and P. Giege, *tRNA biology in mitochondria.* Int J Mol Sci, 2015. **16**(3): p. 4518-59.
64. Kimura, M., *The neutral theory of molecular evolution and the world view of the neutralists.* Genome, 1989. **31**(1): p. 24-31.
65. Kimura, M., *Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution.* Nature, 1977. **267**(5608): p. 275-6.
66. Crick, F.H., *Codon--anticodon pairing: the wobble hypothesis.* J Mol Biol, 1966. **19**(2): p. 548-55.
67. Machnicka, M.A., et al., *MODOMICS: a database of RNA modification pathways--2013 update.* Nucleic Acids Res, 2013. **41**(Database issue): p. D262-7.
68. Yokoyama, S., et al., *Molecular mechanism of codon recognition by tRNA species with modified uridine in the first position of the anticodon.* Proc Natl Acad Sci U S A, 1985. **82**(15): p. 4905-9.

69. Osawa, S., et al., *Recent evidence for evolution of the genetic code.* Microbiol Rev, 1992. **56**(1): p. 229-64.
70. Nasvall, S.J., P. Chen, and G.R. Bjork, *The wobble hypothesis revisited: uridine-5-oxyacetic acid is critical for reading of G-ending codons.* RNA, 2007. **13**(12): p. 2151-64.
71. Rozov, A., et al., *Novel base-pairing interactions at the tRNA wobble position crucial for accurate reading of the genetic code.* Nat Commun, 2016. **7**: p. 10457.
72. Watanabe, K., *Unique features of animal mitochondrial translation systems. The non-universal genetic code, unusual features of the translational apparatus and their relevance to human mitochondrial diseases.* Proc Jpn Acad Ser B Phys Biol Sci, 2010. **86**(1): p. 11-39.
73. Suzuki, T. and T. Suzuki, *A complete landscape of post-transcriptional modifications in mammalian mitochondrial tRNAs.* Nucleic Acids Res, 2014. **42**(11): p. 7346-57.
74. Bali, V. and Z. Bebok, *Decoding mechanisms by which silent codon changes influence protein biogenesis and function.* Int J Biochem Cell Biol, 2015. **64**: p. 58-74.
75. Alston, C.L., et al., *The genetics and pathology of mitochondrial disease.* J Pathol, 2016.
76. Wood, A.R., et al., *Another explanation for apparent epistasis.* Nature, 2014. **514**(7520): p. E3-5.
77. Fish, A.E., J.A. Capra, and W.S. Bush, *Are Interactions between cis-Regulatory Variants Evidence for Biological Epistasis or Statistical Artifacts?* Am J Hum Genet, 2016. **99**(4): p. 817-830.
78. Abbott, J.A., C.S. Francklyn, and S.M. Robey-Bond, *Transfer RNA and human disease.* Front Genet, 2014. **5**: p. 158.
79. Ruiz-Pesini, E., et al., *Effects of purifying and adaptive selection on regional variation in human mtDNA.* Science, 2004. **303**(5655): p. 223-6.
80. Wallace, D.C., *The mitochondrial genome in human adaptive radiation and disease: on the road to therapeutics and performance enhancement.* Gene, 2005. **354**: p. 169-80.
81. Wallace, D.C., *A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine.* Annu Rev Genet, 2005. **39**: p. 359-407.
82. Dingley, S.D., et al., *Mitochondrial DNA variant in COX1 subunit significantly alters energy metabolism of geographically divergent wild isolates in Caenorhabditis elegans.* J Mol Biol, 2014. **426**(11): p. 2199-216.
83. Wallace, D.C., *Mitochondria as chi.* Genetics, 2008. **179**(2): p. 727-35.
84. Kenney, M.C., et al., *Mitochondrial DNA variants mediate energy production and expression levels for CFH, C3 and EFEMP1 genes: implications for age-related macular degeneration.* PLoS One, 2013. **8**(1): p. e54339.
85. Orr, H.A., *Dobzhansky, Bateson, and the genetics of speciation.* Genetics, 1996. **144**(4): p. 1331-5.
86. Lane, N., *Biodiversity: On the origin of bar codes.* Nature, 2009. **462**(7271): p. 272-4.
87. Bar-Yaacov, D., et al., *Mitochondrial Involvement in Vertebrate Speciation? The Case of Mito-nuclear Genetic Divergence in Chameleons.* Genome Biol Evol, 2015. **7**(12): p. 3322-36.
88. Gibson, J.D., et al., *Genetic and developmental basis of F2 hybrid breakdown in Nasonia parasitoid wasps.* Evolution, 2013. **67**(7): p. 2124-32.
89. Gershoni, M., A.R. Templeton, and D. Mishmar, *Mitochondrial bioenergetics as a major motive force of speciation.* Bioessays, 2009. **31**(6): p. 642-50.

90. Burton, R.S. and F.S. Barreto, *A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities?* Mol Ecol, 2012. **21**(20): p. 4942-57.
91. Willett, C.S. and R.S. Burton, *Evolution of interacting proteins in the mitochondrial electron transport system in a marine copepod.* Mol Biol Evol, 2004. **21**(3): p. 443-53.
92. Gebiola, M., et al., *"Darwin's corollary" and cytoplasmic incompatibility induced by Cardinium may contribute to speciation in Encarsia wasps (Hymenoptera: Aphelinidae).* Evolution, 2016.
93. Chou, J.Y. and J.Y. Leu, *Speciation through cytonuclear incompatibility: insights from yeast and implications for higher eukaryotes.* Bioessays, 2010. **32**(5): p. 401-11.
94. Beck, E.A., et al., *Gene flow between Drosophila yakuba and Drosophila santomea in subunit V of cytochrome c oxidase: A potential case of cytonuclear cointrogression.* Evolution, 2015. **69**(8): p. 1973-86.
95. Cong, Q., et al., *When COI barcodes deceive: complete genomes reveal introgression in hairstreaks.* Proc Biol Sci, 2017. **284**(1848).
96. Weckstein, J., et al., *Anomalous Variation in Mitochondrial Genomes of White-crowned (Zonotrichia leucophrys) and Golden-crowned (Z. atricapilla) Sparrows: Pseudogenes, Hybridization, or Incomplete Lineage Sorting?* The Auk 2001. **118**: p. 231–236.
97. Toews, D.P., et al., *Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers.* Curr Biol, 2016. **26**(17): p. 2313-8.
98. Dabrowski, A., et al., *Geographic variability in mitochondrial introgression among hybridizing populations of Golden-winged (Vermivora chrysoptera) and Blue-winged (V. pinus) Warblers.* Conservation Genetics, 2005. **6**: p. 843–853.
99. Brelsford, A., B. MILÁ, and D. IRWIN, *Hybrid origin of Audubon's warbler.* Molec. Ecol., 2011. **20**: p. 2380–2389.
100. McGuire, J.A., et al., *Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotaphytid lizards.* Evolution, 2007. **61**(12): p. 2879-97.
101. Llopart, A., et al., *Sequential adaptive introgression of the mitochondrial genome in Drosophila yakuba and Drosophila santomea.* Mol Ecol, 2014. **23**(5): p. 1124-36.
102. von Beeren, C., et al., *Interbreeding among deeply divergent mitochondrial lineages in the American cockroach (Periplaneta americana).* Sci Rep, 2015. **5**: p. 8297.
103. Hogner, S., et al., *Deep sympatric mitochondrial divergence without reproductive isolation in the common redstart Phoenicurus phoenicurus.* Ecol Evol, 2012. **2**(12): p. 2974-88.
104. Noor, M.A., et al., *Chromosomal inversions and the reproductive isolation of species.* Proc Natl Acad Sci U S A, 2001. **98**(21): p. 12084-8.
105. Navarro, A. and R. Faria, *Pool and conquer: new tricks for (c)old problems.* Mol Ecol, 2014. **23**(7): p. 1653-5.
106. Hill, T., C. Schlotterer, and A.J. Betancourt, *Hybrid Dysgenesis in Drosophila simulans Associated with a Rapid Invasion of the P-Element.* PLoS Genet, 2016. **12**(3): p. e1005920.
107. Darwin, C.A., *The Origin of Species*. 2cnd ed. 1860 (reprinted 1958), New York: New American Library. 448.
108. Latorre-Pellicer, A., et al., *Mitochondrial and nuclear DNA matching shapes metabolism and healthy ageing.* Nature, 2016. **535**(7613): p. 561-5.

109. Wallace, D.C., *Genetics: Mitochondrial DNA in evolution and disease.* Nature, 2016. **535**(7613): p. 498-500.
110. Kang, E., et al., *Mitochondrial replacement in human oocytes carrying pathogenic mitochondrial DNA mutations.* Nature, 2016.
111. Shoubridge, E., *Biomedicine: Replacing the cell's power plants.* Nature, 2016.
112. Templeton, A.R., *The reality and importance of founder speciation in evolution.* Bioessays, 2008. **30**(5): p. 470-9.
113. Mende, M.B. and A.K. Hundsdoerfer, *Mitochondrial lineage sorting in action--historical biogeography of the Hyles euphorbiae complex (Sphingidae, Lepidoptera) in Italy.* BMC Evol Biol, 2013. **13**: p. 83.
114. Smith, J.M. and J. Haigh, *The hitch-hiking effect of a favourable gene.* Genet Res, 1974. **23**(1): p. 23-35.
115. Strobeck, C., J.M. Smith, and B. Charlesworth, *The effects of hitchhiking on a gene for recombination.* Genetics, 1976. **82**(3): p. 547-58.
116. Oliveira, D.C., et al., *Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp nasonia (hymenoptera: pteromalidae).* Mol Biol Evol, 2008. **25**(10): p. 2167-80.
117. Avise, J.C., J.E. Neigel, and J. Arnold, *Demographic influences on mitochondrial DNA lineage survivorship in animal populations.* J Mol Evol, 1984. **20**(2): p. 99-105.
118. Luria, S.E. and M. Delbruck, *Mutations of bacteria from virus sensitivity to virus resistance.* Genetics, 1943. **28**: p. 490-510.
119. Stewart, F., D.M. Gordon, and B.R. Levin, *Fluctuation analysis: the probability distribution of the number of mutants under different conditions.* Genetics, 1990. **124**: p. 175-185.
120. Tibayrenc, M., J.C. Avise, and F.J. Ayala, *In the light of evolution IX: Clonal reproduction: Alternatives to sex.* Proc Natl Acad Sci U S A, 2015. **112**(29): p. 8824-6.
121. Koch, R., *Classics in infectious diseases. The etiology of tuberculosis: Robert Koch. Berlin, Germany 1882.* Rev Infect Dis, 1982. **4**(6): p. 1270-4.
122. Levin, B.R., D.M. Gordon, and F.M. Stewart, *The origin of mutations: mundane processes can account for departure from Luria-Delbruk expectations.* 1991.
123. Newcombe, H.B., *Origin of bacterial variants.* Nature, 1949. **164**(4160): p. 150.
124. Lederberg, J. and E.M. Lederberg, *Replica plating and indirect selection of bacterial mutants.* J. Bact., 1952. **63**: p. 399-406.
125. Novick, A. and L. Szilard, *Experiments with the Chemostat on spontaneous mutations of bacteria.* Proc Natl Acad Sci U S A, 1950. **36**(12): p. 708-19.
126. Maddamsetti, R., R.E. Lenski, and J.E. Barrick, *Adaptation, Clonal Interference, and Frequency-Dependent Interactions in a Long-Term Evolution Experiment with Escherichia coli.* Genetics, 2015. **200**(2): p. 619-31.
127. Grenfell, B.T., et al., *Unifying the epidemiological and evolutionary dynamics of pathogens.* Science, 2004. **303**(5656): p. 327-32.
128. Weir, W., et al., *Population genomics reveals the origin and asexual evolution of human infective trypanosomes.* Elife, 2016. **5**: p. e11473.
129. Tibayrenc, M. and F.J. Ayala, *Reproductive clonality in protozoan pathogens--truth or artifact? A comment on Ramirez and Llewellyn.* Mol Ecol, 2015. **24**(23): p. 5778-81.

130. Tomasini, N., et al., *How often do they have sex? A comparative analysis of the population structure of seven eukaryotic microbial pathogens.* PLoS One, 2014. **9**(7): p. e103131.
131. Tibayrenc, M. and F.J. Ayala, *How clonal are Neisseria species? The epidemic clonality model revisited.* Proc Natl Acad Sci U S A, 2015. **112**(29): p. 8909-13.
132. Diaz Caballero, J., et al., *Selective Sweeps and Parallel Pathoadaptation Drive Pseudomonas aeruginosa Evolution in the Cystic Fibrosis Lung.* MBio, 2015. **6**(5): p. e00981-15.
133. Strelkowa, N. and M. Lassig, *Clonal interference in the evolution of influenza.* Genetics, 2012. **192**(2): p. 671-82.
134. Baym, M., et al., *Spatiotemporal microbial evolution on antibiotic landscapes.* Science, 2016. **353**(6304): p. 1147-51.
135. Burnet, F.M., *Immunological recognition of self.* Science, 1961. **133**(3449): p. 307-11.
136. Lederberg, J., *Genes and antibodies.* Science, 1959. **129**(3364): p. 1649-53.
137. Provine, W.B., *Ernst Mayr: Genetics and speciation.* Genetics, 2004. **167**(3): p. 1041-6.
138. Eldredge, N. and S.J. Gould, *Punctuated equilibria: an alternative to phyletic gradualism*, in *Models in Paleobiology*, T.J.M. Schopf, Editor. 1972, Freeman, Cooper and Co.: San Francisco. p. 82-115.
139. Bazin, E., S. Glemin, and N. Galtier, *Population size does not influence mitochondrial genetic diversity in animals.* Science, 2006. **312**(5773): p. 570-2.
140. Goodall-Copestake, W., et al., *Swarms of diversity at the gene cox1 in Antarctic krill.* Heredity (Edinb), 2010. **104**: p. 513-518.
141. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.
142. Qiu, J., *How China is rewriting the book on human origins.* Nature, 2016. **535**: p. 22-25.
143. Smith, F. and J. Ahern, eds. *The Origins of Modern Humans: Biology Reconsidered*. 2013, Wiley: New Jersey 454.
144. Cann, R.L., M. Stoneking, and A.C. Wilson, *Mitochondrial DNA and human evolution.* Nature, 1987. **325**(6099): p. 31-6.
145. Cann, R., *All about mitochondrial eve: an interview with Rebecca Cann. Interview by Jane Gitschier.* PLoS Genet, 2010. **6**(5): p. e1000959.
146. Gibbons, A., *Y chromosome shows that Adam was African.* Science, 1997. **278**(5339): p. 804-5.
147. Nosil, P., et al., *Tipping points in the dynamics of speciation*
. nature Ecology & Evolution, 2017. **1**: p. 1-8.
148. Sudmant, P.H., et al., *An integrated map of structural variation in 2,504 human genomes.* Nature, 2015. **526**(7571): p. 75-81.
149. Templeton, A.R., *The theory of speciation via the founder principle.* Genetics, 1980. **94**(4): p. 1011-38.
150. Nei, M., *Bottlenecks, genetic polymorphism and speciation.* Genetics, 2005. **170**(1): p. 1-4.
151. Johnson, T., *Beneficial mutations, hitchhiking and the evolution of mutation rates in sexual populations.* Genetics, 1999. **151**(4): p. 1621-31.
152. Hernandez, R.D., et al., *Classic selective sweeps were rare in recent human evolution.* Science, 2011. **331**(6019): p. 920-4.

153. Campos, J.L., et al., *The relation between recombination rate and patterns of molecular evolution and variation in Drosophila melanogaster.* Mol Biol Evol, 2014. **31**(4): p. 1010-28.
154. Zuckerkandl, E. and L. Pauling, *Molecules as documents of evolutionary history.* J Theor Biol, 1965. **8**(2): p. 357-66.
155. Lee, M.S. and S.Y. Ho, *Molecular clocks.* Curr Biol, 2016. **26**(10): p. R399-402.
156. Vrba, E.S. and D. DeGusta, *Do species populations really start small? New perspectives from the Late Neogene fossil record of African mammals.* Philos Trans R Soc Lond B Biol Sci, 2004. **359**(1442): p. 285-92; discussion 292-3.
157. Grant, W.S., *Problems and Cautions With Sequence Mismatch Analysis and Bayesian Skyline Plots to Infer Historical Demography.* J Hered, 2015. **106**(4): p. 333-46.
158. Thompson, J., *Relentless Evolution* 2013: University of Chicago Press.
159. Horvitz, H.R., *Nobel lecture. Worms, life and death.* Biosci Rep, 2003. **23**(5-6): p. 239-303.
160. Mitchell, P., *Chemiosmotic coupling in oxidative and photosynthetic phosphorylation. 1966.* Biochim Biophys Acta, 2011. **1807**(12): p. 1507-38.
161. Pagliarini, D.J. and J. Rutter, *Hallmarks of a new era in mitochondrial biochemistry.* Genes Dev, 2013. **27**(24): p. 2615-27.