MtDNA: The small workhorse of evolutionary studies

Rob DeSalle¹, Bernd Schierwater¹,²,³, Heike Hadrys¹,²,³

¹Sackler Institute for Comparative Genomics, American Museum of Natural History, 79th Street at Central Park West, New York, NY 10024. ²Institut für Tierökologie und Zellbiologie, Stiftung Tierärztliche Hochschule Hannover, Hannover, Germany. ³EEB, Yale University, New Haven, Connecticut 06511, USA

TABLE OF CONTENTS

1. Abstract
2. Introduction
3. Entrenched uses of mtDNA
   3.1. Animal population genetics and phylogeography
   3.2. Animal systematics using mtDNA sequences
4. Emerging approaches using mtDNA in evolutionary studies
   4.1. Using molecular morphology to infer phylogeny
   4.2. Paleo Human mtDNA studies
   4.3. Taxonomy and museum collections
   4.4. DNA barcoding
   4.5. Mitochondrial DNA next generation sequencing (NGS) and evolutionary studies
   4.6. Evolutionary dynamics of mtDNA: natural selection
5. The future of mtDNA in evolutionary studies
6. References

1. ABSTRACT

The double-stranded, circular mitochondrial DNA (mtDNA), which is present in all eukaryotic life forms, was initially discovered and characterized in the last century and has been widely used in evolutionary studies. Since then, a large number of studies have taken advantage of the genetic information encoded in this genome. Because of its small size in animals (in general), the technical ease of manipulating mitochondrial genome and the dynamics of its evolutionary change, this genome has been the workhorse of evolutionary studies over the past three decades. However, the ease with which nuclear DNA can be manipulated due to next generation sequencing (NGS) methods, has recently caused an expected dip in the use of mtDNA in evolutionary studies. This review examines the future of mitochondrial DNA as a useful tool in studies centered around evolution.

2. INTRODUCTION

A plethora of reviews about the utility of mitochondrial DNA in evolutionary studies has been produced in the more than three decades that this molecule has been used as a tool for evolutionary studies. The sheer volume of the reviews and the breadth of subject matter they span (see timeline in Figure 1) points to an incredibly important position for mitochondrial DNA in evolutionary studies. The unique characteristics of this small molecule make it an important tool in evolutionary biology, as summarized in Table 1. The major role played by mtDNA as a marker in evolutionary studies has been in two subdisciplines of evolutionary biology—population genetics and systematics. Specifically, mtDNA has played a central role in understanding human population genetics and the movement of our species across the planet. mtDNA has also been used as an evolutionary focus in studies of natural selection. This review examines the important role of mtDNA in the development of modern evolutionary biology and attempts to address the future utility of this tiny workhorse of evolutionary studies.

3. INGRAINED USES OF mtDNA

One of the major reasons mtDNA has been so useful in population genetics is the rapidity with which it evolves. The haploid nature of mtDNA means that coalescence of neutral genes will be positively correlated with the effective population size of the species (1, 2). This means that from population genetics theory mtDNA should evolve four times faster than the average nuclear gene. Hence, mtDNA can be used to follow divergence in very closely related taxa and even within species. The long tradition of using this marker in population and species boundary studies in animals was started by John Avise who authored several seminal papers in the early 1980’s and popularized the term “phylogeography”. Since then literally thousands of papers focused on thousands of different species have been published (Figure 1).
Figure 1. Timeline for major developments of animal mtDNA analysis. The curve was established by searching Google Scholar for the terms “mitochondrial DNA” AND “evolution” and counting the hits at two year intervals starting in 1980 and ending in 2016. The Y axis shows the number of hits. A similar search for “evolution” was linear in the same time period. The arrows point to specific years where seminal events occurred in the history of mitochondrial use as an evolutionary marker. The colors on the graph represent where different technological developments occurred. The red section can in general be considered Pre-Polymerase Chain Reaction, the green section is post-PCR but pre genomics, the blue region represents first generation genomics sequencing and the yellow section represents the period where Next Generation Sequencing (NGS) techniques became available. Seminal events shown in the figure are coded as follows: 1. Brown et al., 1979 (9) Rapid evolution of mtDNA; Avise et al., 1979 (79) mtDNA RFLP population structure; 2. Brown, 1980 (32). Human RFLP mtDNA; 3. Anderson et al., 1981 (80) First human mtDNA – genome; 4. Clary et al., 1982 (81) Drosophila mtDNA genome; 5. Aquadro et al., 1983 (33) Seven human mtDNA gene study; 6. Higuchi et al., 1984 (82) – Extinct quagga mtDNA; 7. Cann et al., 1987 (34) mitochondrial Eve; Avise et al., 1987 (83) Phylogeography; 8. Kocher et al., 1989 (84) Universal mtDNA primers; 9. Birky, 1991 (2) mtDNA popgen; 10. Excoffier et al., 1992 (85) mtDNA molecular variance; 11. Tamura, 1983 (86) Human dloop Poppen; Torroni et al., 1993 (87) Human Haplogroups named; 12. Baker and Palumbi, 1994 (48) – Whale mtDNA ID; Avise, 1994 (88) Molecular Markers; 13. Krings et al., (89) Neanderthal mtDNA; 14. Kogelnik et al., 1998 (90) MITOMAP; 15. Hebert et al., 2003 (47) DNA barcoding; 16. 2004, CBoL (47) established; 17. Briggs et al., 2009 (39) Five Neanderthal mtDNA genomes; 2009 Mitochondrial DNA Journal launched; van Oven and Kayser 1992 (81); Haplogroups refined; 18. Dabney et al., 2013 (92-94) ultrashort mtDNA sequencing; 19. Picardi et al., 2012 (95) 1000 genomes mtDNA; 20. Meyer et al., 2016 (96) 450,000 year old mtDNA from Sima de los Huesos hominins. We have started this “history” in 1979 when the modern DNA analysis era kicks in. However, the following are significant events prior to this date. 1949 - Slonimski and Ephrussi isolated yeast mutants which were defective for cell respiration and hypothesized presence of some non-Mendelian genetic characters; 1960 - Chevremont demonstrated that mitochondria incorporated tritiated thymidine, a marker nucleoside: nucleic acid metabolism in mitochondria; 1962 - Nass and Nass demonstrated by morphological studies that mitochondria contained DNA; 1965 - Saccone and colleagues showed that isolated mitochondria were able to synthesize RNA; 1965 - Kroon demonstrated that intact mitochondria or fragments could incorporate amino acids, signaling presence of a protein translation system in organelle; 1967 - Clayton and Vinogra isolated circular dimer and concatenate forms of mtDNA in human cancer cell lines; 1974 - Bogenhagen and Clayton revealed multicopy state of mtDNA in human and mouse cells; 1974 – Berk and Clayton clarified several features of mtDNA replication in mouse cells, including its asymmetry in time and space; 1975 - First complete mitochondrial genomes cloned by Chang and colleagues.

Table 1. Advantages and disadvantages of mtDNA in evolutionary studies

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Size (in most cases &lt;20,000 bp)</td>
<td>• Ease of annotation</td>
<td></td>
</tr>
<tr>
<td>Ratio of copies to nuclear genome is high</td>
<td>• Ease of isolation and manipulation; Also allows for isolation from long dead tissues</td>
<td></td>
</tr>
<tr>
<td>Protein genes + ribosomal genes + AT rich region</td>
<td>• Increased range of evolutionary rates</td>
<td></td>
</tr>
<tr>
<td>Maternally Inherited</td>
<td>• Clonal and hence genetics simple</td>
<td></td>
</tr>
<tr>
<td>Ne ½ of nuclear genes</td>
<td>• Goes to fixation faster</td>
<td></td>
</tr>
<tr>
<td>Gene order easily determined</td>
<td>• Gene order phylogenetics possible¹</td>
<td></td>
</tr>
<tr>
<td>Non-recombining</td>
<td>• Clonal and hence genetics simple</td>
<td></td>
</tr>
<tr>
<td>High mutation rate</td>
<td>• Rapid change accommodates examining closely related organisms</td>
<td></td>
</tr>
<tr>
<td>Sometimes heteroplasmic</td>
<td>• Destroys advantage of clonality</td>
<td></td>
</tr>
<tr>
<td>Insertions in nuclear genome</td>
<td>• nuMts cause horizontal evolution problems</td>
<td></td>
</tr>
<tr>
<td>Different regions of genome often show</td>
<td>• Because mtDNA is clonal this causes</td>
<td></td>
</tr>
<tr>
<td>Phylogenetic incongruence with each other</td>
<td>• Interpretation problems</td>
<td></td>
</tr>
</tbody>
</table>

¹See Figures 2 and 3 for information on taxonomic groups and limitations of this advantage
3.1. Animal population genetics and phylogeography

Animal mtDNA population studies have focused on phylogeographic patterns using tree building methods. In this approach, individuals from populations are analyzed for polymorphisms in their mtDNA. These polymorphisms are then used as characters to construct a branching diagram of the relationships of the individuals in the populations. The distribution of individuals in the tree can often times be interpreted in a population genetics context. mtDNA is also used to examine demography using techniques like the mismatch distribution approach (3, 4) and the standard analysis of variance population genetics parameters (5). In addition methods, like Templeton’s (6) nested clade analysis have been used mostly with mtDNA data. With the advent of NGS technology, one might think that mtDNA would be replaced by nuclear markers. True to form, a recent survey of the phylogeographic literature Garrick et al. (7) showed that nuclear DNA Single Nucleotide Polymorphisms (SNPs) have increased steadily. However, their survey also showed that mtDNA “continues to represent an important component of phylogeographic data”, in combination with or in comparison to nuclear markers. This trend is encouraging for the future use of mtDNA as a marker in population genetics.

3.2. Animal systematica using mtDNA sequences

Early on in the development of molecular systematics, sequences from mtDNA and the nuclear small subunit ribosomal RNA were used to generate animal phylogenies. The focus on these markers was the result of the limits of the technology available in the 1980’s and early 1990’s. As with population genetics and phylogeography, literally thousands of publications have resulted from the use of mtDNA sequences in systematics. While it would be impossible to address all of these studies in this review, several aspects of mtDNA evolution in the context of systematics became clear from the many phylogenetic studies that were generated.

The fourfold greater rate of evolution of mtDNA is an advantage and a disadvantage at the same time for systematics. The high rate of change means that any mtDNA tree will have a very high probability of resolving correctly short internodes in a phylogenetic tree compared to most nuclear marker (8). While this might be preferable for population genetics, and for studying phylogeny near the species boundary, it produces problems for systematics with deep evolutionary history and accelerated evolutionary rates (e.g. orders of insects; deep mammalian relationships etc.). Because of its high rate of change mtDNA can back-mutate rapidly and cause convergence in systematic studies. This phenomenon was clearly demonstrated in one of the first papers to use mtDNA as a source of characters for phylogenetics as shown in Figure 2 in Brown et al. (9) and is still demonstrable in one of the latest, most comprehensive analyses of complete mitogenomes in insects (10). The back-mutations cause saturation of nucleotide sites, and this saturation results in convergence of characters and long branches which are then susceptible to long branch attraction producing messy phylogenetic signal. There are several approaches that researchers have taken to deal with this problem. The most used is the development of models that take into account the saturation. While models can correct for some of the convergence, such models are only corrective at shallow phylogenetic levels. On the other hand, Simon and Hadrys (10) showed, that in insects depending on the group, complete mitogenomes can recover intraordinal relationships in agreement with morphological and nuclear molecular data sets. In contrast they demonstrate that the limits caused by convergence still exist in inferring deep hexapod (interordinal) relationships using the existing models.

The question with regard to using mtDNA in phyllogenetics then becomes one of how to treat the data in the analysis and much work has been accomplished in this context. Due to the strong knowledge of how mtDNA sequences evolve, researchers have formulated models and weighting schemes for use in phylogenetics. The most extreme model is to give a probability of 1.0 to changes that are transversions and a 0.0 probability to transitions. This effectively removes transitions from the analysis and is called “transversion parsimony” (11). Another extreme model is to weight the probability of change such that the third position is removed from the analysis. This is because most change in mtDNA occurs in the third position of coding regions of the molecule and would then be susceptible to saturation. Other models are easily incorporated into phylogenetic analysis especially when using maximum likelihood and Bayesian analysis. And since the amino acid sequence of mtDNA genes evolve relatively slowly (again third positions are mostly impacted by saturation) some researchers have simply fallen back on using the amino acid sequences of mtDNA genes in phyllogenetic studies and have implemented likelihood models based on empirical amino acid transition patterns (12). Following the protocol of testing for the best model using likelihood ratio tests is advised (13, 14) whether using mtDNA nucleotide sequences or amino acid sequences in likelihood analysis.

It has become very clear though that mtDNA sequences may be most informative in combination with other nuclear markers or gene sequences. The first molecular systematics studies in the 1980’s used mtDNA alone to generate trees. When these first phylogenetic hypotheses from mtDNA appeared it was evident that substantial incongruence of mtDNA based trees and morphological trees existed (15). Later in the 1990’s when it became easier to obtain information from nuclear genes, researchers started to realize that the information
mtDNA and evolution

from mtDNA often times conflicted with evidence from nuclear DNA (8, 16-18).

Data combination or concatenation methods and the capacity to analyze the relative support of different data sets (19) showed that while mtDNA trees are frequently incongruent with trees from other sources of data, there is considerable hidden support in all data sets for a concatenated analysis. A more recent example of this can be found in Song et al. (20) who used mt-genome data and the inclusion of rRNA genes to generate a phylogenetic hypothesis of Holometabola (insects that develop through larval stages), one of the most species rich and controversial branches of the tree of life. Based on the largest taxon sampling of Holometabola to date, Song et al. (20) tested mtDNA nucleotide and amino acid data sets using several tree building algorithms and models. With densely sampled mitochondrial genomes they outlined a practical approach to recovering reasonable hypotheses of Holometabola phylogeny, as corroborated by nuclear and morphological data. The inclusion of rRNA genes with mtDNA sequences and removal of fast-evolving sites under a site heterogeneous model correctly recovered most of the deep branches in the Holometabola. We should also point out, however, that there is some controversy concerning concatenation methods in phylogenetics (see for instance (21)).

While mtDNA has been used quite successfully in animal phylogenetics several problems have arisen that warn against its indiscriminate use. One problem that has been particularly difficult to detect in the past is the presence of nuclear mitochondrial pseudogenes (Numts). Numts result from the translocation of mitochondrial sequences from the mitochondrial genome into the nuclear genome and, once integrated, these non-functional sequences accumulate mutations freely. The potential for Numt amplification in addition to, or even instead of, the authentic target mtDNA sequence can seriously confound population genetic and phylogenetic analyses. In a study on gorillas (22) the prevalence of Numts, for example, obscured the presence of two genetically divergent clades and affected the understanding of their evolutionary history as well as future conservation and management plans. Numts also will impact inferences at the population genetics level. As an example, the presence of Numts in *Aedes aegypti* (the tiger mosquito) might have seriously confounded the interpretation of this mosquito’s demographic distribution in the wild. In a combined experimental and bioinformatics approach utilizing the recent genome sequence of *Ae. aegypti* the authors (23) showed that Numts are indeed prevalent in *Ae. aegypti* and that they more than likely affect demographic studies of this species using mtDNA.

Another caveat concerning mtDNA usage in phylogenetics is to always keep in mind that it follows maternal lineages of animals. This basic aspect of mtDNA inheritance becomes relevant for the interpretation of phylogenetic patterns generated from mtDNA in animals that do not reproduce with random mates. For instance, if the mating system of a group of animals is matrilineal, mtDNA sequences will bias systematic patterns toward the matrilineal pattern. If the evolutionary history is not congruent with the matrilineal history then the inferred organismal phylogeny from mtDNA will be biased. However, if unraveling matrilineal patterns of history is a goal of research the mtDNA genome can be exploited efficiently to test hypotheses about this interesting demographic behavior (24-26).
4. EMERGING APPROACHES USING mtDNA IN EVOLUTIONARY STUDIES

In the following section we discuss several more recent applications of the use of mtDNA in evolutionary biology. Detailed examination of the structural aspects of mtDNA gene products have been used in phylogenetic studies. Recent capacity to isolate DNA from Paleolithic samples of human bones has opened the way for analyses of human mtDNA variation at a time frame never before accomplished. DNA barcoding and the use of mtDNA in taxonomy are also emerging approaches in modern mtDNA research in evolution. And finally the application of next generation sequencing technology has been suggested as the new wave of mtDNA research in evolutionary biology.

4.1. Using molecular morphology to infer phylogeny

It has been argued that mtDNA can be useful in systematic studies by converting the information from mtDNA genomes into structural information. The overall structure of mitochondrial DNA genomes has been used to decipher relationships in Cnidaria. Bridge et al., (27) used the chromosome structure the Cnidarian classes to hypothesize that Anthozoa are the ancestral class (having a circular and hence ancestral position) relative to hydrozoans, scyphozoans and cubozoans (all three having linear genomes). Ender and Schierwater (28) examined the predicted stem and loop folding structures of 16S rDNA molecules to score several “morphological” characters relevant to lower Metazoan phylogenetics.

More frequently though the gene order of mitochondrial genomes has been used to examine phylogeny in many animal systems (29-32), much in the same way that banding patterns of polytene chromosomes of fruit flies were used to determine Drosophila phylogenies. At this point in time it could be argued that the spectrum of structural information inherent in mt genomes has been insufficiently explored for use in phylogenetics. Also, possible complications arising from having a wide spectrum of mitochondrial genome sizes, genetic inventories, and gene arrangements or even linearization of mtDNA genomes in different animal taxa have not been adequately addressed yet.

The overview in Figure 2 shows that the largest animal mtDNA genomes are circular and more than 40Kb in size, while the smallest are less than 10Kb in size. mtDNA genome linearization and fragmentations have occurred as a synapomorphy in the Medusozoa but are also found as an exception in some Porifera (Figure 2). With respect to the gene inventory the most complete mtDNA genomes are found in placozoans, while the derived Ctenophora show the most incomplete genomes. Within the Bilateria similar secondary fragmentations of the mtDNA genome are observed in some arthropods, but with the structure remains circular. Substantial secondary expansions due to duplication events of whole mtDNA genome regions are found in some molluscs. Overall, it looks like mtDNA genomes in diploblastic animals are in many respects overly messy for comparative studies at higher taxonomic levels.

In sharp contrast, the typical bilaterian mtDNA genome is quite uniform. It typically shows a standard size of 16Kb, harbors all respiratory chain genes, a complete set of tRNA genes to encode the standard set of 20 amino acids, lacks any introns and is always circular. Thus within Bilateria and especially for groups within the Bilateria, comparative studies of structure are not problematic. As an example, in the insects, the gene order within the mtDNA genome varies between only between orders (Figure 3A). Within an insect order gene order is for the most part perfectly conserved. In Figure 3B an example is shown for the Odonata, a group that has radiated more than 300 Mio. years ago. The only structural differences seen here relate to the length of the control region and are not evident in gene order at all.

4.2. Paleo human mtDNA studies

mtDNA has been the focus of an immense body of work in human evolution studies. Since the first publications using mtDNA to examine human variation were generated in 1980 (33) to the subsequent sequencing of mitochondrial genes in human population genetics (34) and to the mitochondrial “Eve” hypothesis (35) mtDNA has had an integral role in clarifying human population variation. Current genome sequencing projects have the extra added benefit of yielding mtDNA genomes from the subjects in such studies (36) and this coupled with a very sophisticated view of mtDNA haplotype biology (37, 38) ensures that human mtDNA studies will continue to be useful in understanding human population genetics.

In addition, human mtDNA has also been a relatively important source of information for paleo-DNA studies (39). To demonstrate the power of this paleo DNA approach several studies deserve mention. Several H. neanderthalensis individuals have had their mtDNA genomes sequenced (40) and three non- H. neanderthalensis specimens from the Denisova cave have had mtDNA isolated from remains (41). The oldest human paleo mtDNA isolated so far is from the Sima de los huesos cave in Sierra de Atapuerca in Spain dated at 430,000 years old. Two tour de force studies of paleo-mtDNA are by Llamas et al., (42) and Fu et al., (43). The Llamas et al., (42) study sequenced the whole mtDNA genomes of 92 pre-Columbian South American skeletons dating from 9000 to 500 years ago. This large study showed that the diversity of pre-Columbian H. sapiens was large and further that “European colonization caused a substantial loss of pre-Columbian lineages” (42). The Fu et al., (43) study examined the paleo DNA of 51 modern H. sapiens subfossils dating as far back as
35000 years ago. This latter study hypothesizes that a single founder population for *H. sapiens* in Europe during the Ice Age and also documents a great deal of migration and population turnover as major themes in the evolution of European *H. sapiens*. A study by Ozga *et al.* (44) explored the potential of archaeological human dental calculus as a source for human ancestry information. Using an in-solution and enrichment technique with subsequent high-throughput NGS, complete mt-genomes were reconstructed from dental calculus of "pre-Contact" native North American skeletons dating back 700 years ago. Although the study itself is focused on the material
mtDNA and evolution

...and technique, it opens up some new opportunities and sources to study human past and health.

The number of paleo-mtDNA studies is too large to mention here, but Haber et al., (45) have recently reviewed the field and have summarized the role of paleo-DNA and mtDNA in particular in our understanding of human population genetics and movement. They point out that three major areas of interest have been addressed using paleo mtDNA. These include studies that address how modern humans expanded across the globe and focuses on potential admixture of H. sapiens with archaic species. Second, paleo mtDNA has been used to pin down the patterns with which modern Europe have been colonized through the Neolithic Transition. Third, paleo mtDNA has been extremely useful in determining the patterns with which H. sapiens populated the western hemisphere. Given the ability to isolate paleo-mtDNA and the extensive background work on mtDNA haplotypes of extant human populations, this molecule will no doubt continue for some time to be an important tool in studying human variation and population genetics even despite the emergence of spectacular technology that can sequence paleo nuclear genomes.

4.3. Taxonomy and museum collections

The field of taxonomy is changing rapidly, because of the “increase in taxonomic breadth of mtDNA databases” as Timmermans et al., (46) suggest. MtDNA sequence data are already widely used in species identification and classification, however, increasingly without formal species identification based on Linnaean taxonomy. The danger of this DNA-based taxonomy lays in its non-association to existing biological information linked to the “Linnaean” nomenclature. Consequently, the historical knowledge repositories in the literature and natural history collections could as Timmermans et al., (47) suggest “become a relic of the past and specimen to species links become untraceable”.

Beyond the prominent studies of human evolution the improved applicability of paleo-mtDNA to museum collections, other taxonomic groups are being focused on using museum collections. This approach has added a new and incredibly useful aspect to taxonomy by directly linking type material, tissue collections or vouchers to its individual (original) genetic “barcode”. What this means is that classical type specimens which are often times quite old are now being linked to actual DNA sequences. This molecular information can then be compared with material from outside sources and other biological studies of the organisms in focus. While conventional PCR-based approaches are susceptible to contamination, various new techniques based on shotgun NGS and assembly of full mt-genomes with improved bioinformatics can overcome these limits.

Complete mitogenome “taxonomies”of pooled, large scale museum-samples are now underway to link museum and outside sources. While these approaches still have to overcome technical limitations and problems associated with the endeavor (among-sample variation, very short sequences achievable with preserved specimens, critical assembly of mitochondrial genomes of closely related species in a museum pool, the fact that specimens and corresponding DNA are in very different conditions) numerous researchers are making progress towards its realization. As Timmermans et al. (46) state “mitochondrial sequences are particularly accessible to bulk sequencing because of the potential full assembly of organelle genomes present in high copy number from mixtures of specimens”. As a proof of concept they tested a pool of 35 British butterflies from a natural history collection and successfully extracted the standard mt-DNA barcode sequences as well as mitochondrial genomes from a large number of the species in the pool. While these approaches still have to overcome some limitations and problems associated with sequencing technology (among-sample variation due to DNA degradation and very short sequences achievable with preserved specimens to name two) critical assembly of mitochondrial genomes of closely related species in a pool, specimens and corresponding DNA are in very different conditions etc. numerous researchers are making progress towards its realization (e.g. 47).

4.4. DNA barcoding

In 2003 Hebert et al., (48) made the suggestion that the mitochondrial DNA COI gene could be used as a universal tool for identification of animal species. The idea of using DNA sequences for identifying species was not new (see (49, 50) for early examples), however, the utility of Hebert et al., (48) suggestion was to systematically use the sequences of the mitochondrial COI gene as a sort of a DNA barcode. They pointed out that even in a 15 base pair stretch of that gene (or alternatively if there were 15 polymorphic sites in short regions of the COI gene) that over 1 billion different combinations of sequence potentially could be found. This large number of potential nucleotide combinations led them to suggest that a single small region of the COI gene would be adequate for DNA barcoding of animals. Since the initial 2003 paper thousands of publications have appeared either using or discussing the use of DNA barcodes in animal identification.

The approach has not been without controversy. Most of the controversy arises from imprecise definitions of what DNA barcoding can be used for in evolutionary and biodiversity studies. It is useful to discriminate between using DNA barcodes to “identify” species that already have strong taxonomic work done on them, and using DNA barcoding to “discover” or “discriminate” and systematize new and unknown species (51-55). Classical taxonomists argue that an integrative approach

879 © 1996-2017
using DNA and morphology should be used to identify new species and to do taxonomy (56, 51). However once new species have been identified there is no reason to shy away from using DNA sequences to identify already characterized species.

It should be noted that the most popular approach to analysis of DNA barcode data can be found on the Barcode of Life Database (BoLD) website (http://www.boldsystems.org; (57)). In addition, several informatics approaches have been developed to utilize DNA barcode data for identification of species (algorithms reviewed in 58). According to Little and Stevenson (58) these approaches can be divided into four major categories – clustering methods, similarity methods, combination methods and diagnostic methods. The similarity and clustering methods (and by default the combination methods) contrary to most taxonomic approaches utilize distances as a criterion for identification of specimens in the database. While it is beyond the scope of this review to delve into the mechanisms of these other approaches, we point out that an important, yet under-represented bioinformatics modification is to use a character-based approach, the so-called CAOS approach (59-61) to concatenate multiple "layers" of either sequences of the mitochondrial COI gene, other mt-gene fragments or nuclear gene fragments (62). By identifying taxon-specific characters within a traditional or new barcode region many problems of the traditional DNA barcoding can be overcome and different operational taxonomic units, from population to genera, can be identified as this has been shown for example in insects (63, 64). At the same time the flagging of new or previously unknown species is possible and character based barcodes can readily be incorporated into matrices including organismal characters (e.g. morphological, developmental, ecological and behavioral data).

Despite the controversies, DNA barcoding has been an extremely active area of research for evolutionary and biodiversity focused scientists. The research is summarized on the Barcode of Life Database (BoLD) website. As of May, 2016 nearly 5 million DNA barcode sequences had been deposited into the BoLD database. This number of barcodes encompasses over 160,000 animal species. More recently, Ratnasingham and Hebert, (65) have suggested that DNA barcode index numbers can be given to species with DNA barcode information calling such numbers the BIN system. Such an approach they hope will "aid revisionary taxonomy by flagging possible cases of synonymy, and by collating geographical information, descriptive metadata, and images for specimens that are likely to belong to the same species, even if it is undescribed." While DNA barcoding has progressed from a lofty idea in 2003 to a full fledged and active movement amongst organismal biologists, Coissac et al., (66) make the suggestion that in addition to accelerating the standard DNA barcode approach, that researchers should also augment that approach with NGS methods that can skim genomes for extended DNA barcodes outside of the mtDNA COI gene.

4.5. Mitochondrial DNA next generation sequencing (NGS) and evolutionary studies

Most of the mtDNA studies at the population and systematic level are done using standard Sanger DNA sequencing approaches. Recent interest in expanding the role of mtDNA not only to meta-barcoding approaches of single mt-gene-fragments but to the genome level in evolutionary studies have attempted to frame future work using Next generation sequencing approaches. Part of the problem with utilizing NGS approaches is adapting the massively parallel approaches of NGS to single species. Crampton-Platt et al., (67) have outlined adaptation of NGS approaches that they call mitochondrial metagenomics (MMG; also called mito-metagenomics (68) and mitogenome skimming (69) for the study of multiple mitochondrial genomes of animals (see above). These approaches allow for the bulk processing of mixtures of animals (usually smallish insects) in single metagenomic sequencing run. Such approaches rely on the recovery of full mitochondrial genomes using informatic approaches. A possible role for mtDNA has also been suggested for detection of eukaryotes in environmental DNA (eDNA) studies. Some eDNA studies use a mitochondrial gene as a marker because their high copy number enhances the likelihood of DNA detection in environmental samples (summarized in (70)). There is even a suggestion that such approaches might be useful in estimating biomass of species in ecological assemblages (67). These developments would raise the recent meta-barcoding and high-throughput eDNA approaches to a new level of assessing biodiversity patterns.

4.6. Evolutionary dynamics of mtDNA: natural selection

While mtDNA variation is often times assumed to be neutral, several studies have examined the potential of natural selection to act upon it (71-73). Several studies have examined the role of natural selection in mtDNA in specific animal systems, but two recent meta-analyses are illuminating as to overall patterns of natural selection on mtDNA. Garvin et al., (73) performed a meta-analysis of over 200 animal species by examining natural selection in mtDNA genomes. Their analysis showed that "the ND5 subunit of complex I is a repeated target of positive Darwinian selection in diverse taxa" (74). In another meta-analysis of over 500 animal species, James et al., (74) conducted McDonald Kreitman tests for selection and concluded that the majority of mtDNA mutations are slightly deleterious. However, they also detected a significant proportion (26%) of mutations that were non-synonymous and hence potentially being influenced by natural selection. Interestingly, they also showed a correlation of the rate of adaptive evolution of mtDNA mutations with synonymous diversity, which
they interpret as “that at least some adaptive evolution is limited by the supply of mutations” ((75); p. 67).

Natural selection at the level of the entire mtDNA molecule has also been examined (76). Because mitochondria are found in large numbers per cell, the strength of natural selection on individual molecules of mtDNA should be very low. On the other hand, mtDNA molecules that promote their own replication no matter what cell function might be operating should be strongly selected for. Haig (76) suggests that this evolutionary paradox has been solved by the transfer of critical genes from the mitochondrion to the nucleus. A few very essential genes still exist in the mitochondrial genome and according to Haig (76) these are managed in an evolutionary sense by homogenization through bottlenecking in female germlines and elimination of “low quality” mitochondrial genomes with deleterious mutations. Another study that has addressed the shrinkage of the mitochondrial genome since animals diverged looked at the full genomes of over 2000 animals and modeled all possible mtDNA genome gene combinations. Using this approach Johnston and Williams (77) were able to show that there are three universals in determining whether or not a gene in the mtDNA genome will be lost. First if the gene is involved as a central player in a biochemical process then it will be retained. Second, if the gene codes for a hydrophobic protein it will be retained and finally if the gene has a high proportion of G’s and C’s in its sequence it will be retained. This in silico study points to the importance of meta-datasets and of using models to unravel an interesting evolutionary problem.

Another broad evolutionary question that can be asked about the mtDNA genome concerns its origin and maintenance in animal cells. It is widely accepted that the mitochondrion arose as an endosymbiotic capture event by an ancestral eukaryotic cell of a alpha-proteobacterial donor (77). What has not been settled is whether the endosymbiotic capture event was early or late. Pittis and Gabaldón (78) recently report that the event was an evolutionarily late one because many of the bacterial proteins that appeared before the endosymbiont capture are active in the intracellular membrane system. This result has been interpreted to mean that the ancestral eukaryotic cell that captured the mitochondrial ancestor already had a high degree of complexity. On the other hand, Esposti (79) suggests an alternative to this late hypothesis. However this controversy is settled it will remain of interest for some time.

5. THE FUTURE OF mtDNA IN EVOLUTIONARY STUDIES

It is not too bold of a statement to suggest that mtDNA will continue to be a workhorse of modern evolutionary biology despite some suggestions that it has run its course as an important tool. Certainly, the molecule will continue to be used as a marker in phylogeography studies. There is no reason to exclude mtDNA as a population genetics tool, as long as its biology and limitations are recognized a priori in such studies. In fact, mtDNA phylogeography may be the first thing a researcher may want to establish when doing population studies or when doing full scale phylogeographic studies. Its role in paleo-systematics and human paleo research will still be an important one as a result of the technical limitations this field faces. With the potential of new techniques when working with museum material the utility of mtDNA in studies in this arena are both promising and exciting. The “next generation barcoding techniques” for museum voucher specimens is also an important endeavor that collection based scientists will need to ponder in the near future too.

The use of mtDNA in DNA barcoding studies is a reflection of its utility at the level of understanding species boundaries and species identification. Its utility at this level is very obvious and the continuous drive to complete a comprehensive DNA barcoding database for the species of animals on this planet will also continue to be an important endeavor for research in the future. However, its utility might easily be supplanted by more genomic based approaches (66). Indeed, mtDNA is still at the forefront of integrative taxonomy, with its improved potential to recover and “barcode” mito-genomes of museum samples and its continuous exploration of species boundaries as a marker for environmental metabarcoding approaches. Mito-metagenomics also promises to be a useful approach for the understanding of closely related species that are a part of ecological systems and at the same time allow for more sophisticated, reliable and fast methods to detect invasive species.

While the dynamics of mitochondrial DNA evolution (i.e., natural selection and origin of mtDNA genomes) are fairly well worked out, novel aspects of mtDNA genome evolution will more than likely be discovered when researchers start to examine in detail the dynamics of change in specific animal groups. Here the fast accumulation of comparative information at the mtDNA genome and individual gene level will help to examine the evolutionary forces involved in adaption and speciation processes. In our estimation, while a dip in the actual use of mtDNA as a tool in evolutionary studies is both demonstrated and expected (Figure 1), the use of mtDNA will continue in evolutionary studies in many areas of comparative and evolutionary biology. As Neupert, (80) has pointed out, “The pathways involved in generating mitochondrial diversity are still almost completely unknown … (and) very little is known about the intimate relationship between mitochondrial architecture and function. Here regulation of mitochondrial gene expression will be a wide open field of future research on mitochondrial evolution.” Being able to decipher the steps involved in mediating these regulatory pathways would indeed be an important advance in our understanding of
the evolutionary process. mtDNA, the small workhorse of evolutionary biology, has a lot of work left to do.

6. REFERENCES

4. S. Schneider, L. Excoffier: Estimation of Past Demographic Parameters from the Distribution of Pairwise Differences when the Mutation Rates Vary Among Sites: Application to Human Mitochondrial DNA. Genetics 152.3., 1079-1089 (1999)
mtDNA and evolution


36. M.A. Diroma, C. Calabrese, D. Simone, M.


55. P.Z. Goldstein, R. DeSalle: Integrating DNA Barcode Data and Taxonomic Practice:
Determination, Discovery, and Description. Bioessays 33, 135-147 (2011) DOI: 10.1002/bies.201000036
PMid:2184470

PMid:16243769


60. T. Bergmann, H. Hadrys, G. Breves, B. Schierwater: Character-Based DNA Barcoding: A Superior Tool for Species Classification. Berliner und Münchener Tierärztliche Wochenschrift 122.1.1/12, 446-450 (2009)


PMid:25712507

PMid:17999953 PMCid: PMC2212734


PMid:27006764 PMCid: PMC4802855

PMid:25979752 PMCid: PMC4494052

PMid:26479867


PMid:15012752

72. J. Ballard, O. William, D.M. Rand: The Population Biology of Mitochondrial DNA and


78. A.A. Pittis, T. Gabaldón: Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. Nature 531, 101-104 (2016) DOI: 10.1038/nature16941


91. M. Krings, A. Stone, R.W. Schmitz, H. Krainitzki, M. Stoneking, S. Pääbo: Neandertal
DOI: 10.1016/S0092-8674(00)80310-4

DOI: 10.1093/nar/26.1.112 
PMid:9399813 PMCid: PMC147233


DOI: 10.1073/pnas.1314445110 
PMid:24019490 PMCid: PMC3785785


DOI: 10.1038/nature17405 
PMid:26976447


Key Words: Paleo DNA, DNA Barcoding, Molecular Morphology, Natural Selection, Next Generation Sequencing, Review

Send correspondence to: Rob DeSalle, Sackler Institute for Comparative Genomics, American Museum of Natural History, 79th Street at Central Park West, New York, NY 10024, Tel: 646-226-965, E-mail: desalle@amnh.org