AS I SEE IT

Misconceptions regarding nuclear mitochondrial pseudogenes (Numts) may obscure detection of mitochondrial evolutionary novelties

Nikolaos V. Schizas*

Department of Marine Sciences, University of Puerto Rico at Mayagüez, Isla Magueyes Laboratories, Call Box 9000, Mayagüez, Puerto Rico 00681, USA

ABSTRACT: The proliferation of DNA ‘barcoding’ as a way to catalogue all terrestrial and marine species has attracted much-needed attention to the diversity of life. The reliance of barcoding on cytochrome c oxidase subunit I (COI) for species identification has brought into the spotlight the use and interpretation of mitochondrial data. An increasing number of mitochondrial-like sequences have been discovered, which are generally regarded as nuclear mitochondrial pseudogenes or Numts. The across-the-board approach to categorize unusual mitochondrial DNA (mtDNA) sequences as Numts may obscure the detection of evolutionary novelties in mtDNA. Alternative scenarios are presented where unusual mtDNA sequences are not Numts.

KEY WORDS: Nuclear mitochondrial pseudogenes · Numts · Mitochondrial DNA · mtDNA · Cytochrome c oxidase subunit I · COI · Census of Marine Life

mtDNA and Barcoding of Life

The idea of barcoding life, which entails the sequencing of ~650 nucleotides of the mitochondrial gene cytochrome c oxidase subunit I (COI), has been promoted as a useful tool to catalogue all living species, showcased by the Barcode of Life (BoL) project (Hebert et al. 2003, 2010). The presence of universal primers (Folmer et al. 1994), the matrilineal transmission of mitochondrial DNA (mtDNA) in most species, the perceived absence of recombination (Elson & Lightowlers 2006), and, on average, the higher mutation rate compared to nuclear coding genes (Gissi et al. 2008; but see for further discussion Hudson & Turelli 2003, Ballard & Whitlock 2004) render this COI region suitable to measure biodiversity from a molecular point of view. Barcoding has been extended to marine metazoans under the auspices of the Census of Marine Life (Bucklin et al. 2011), now more than a decade-long effort. The application of this technique has led to an appreciation of standing genetic diversity and the discovery of new genetic lineages in species, subspecies, and potential cryptic species in the marine environment (Barber & Boyce 2006, Gómez et al. 2007, Bucklin et al. 2011). However, despite the haploid nature of mtDNA, non-identical mtDNA-like sequences may exist in one individual, and oftentimes they amplify with or instead of the target mtDNA. These sequences have been labeled nuclear mitochondrial pseudogenes (Numts) and are copies of mtDNA genes or almost-complete mitochondrial genomes that have been translocated to the nuclear genome (Lopez et al. 1994, Kim et al. 2006). Initially, the mitochondrial genes and the newly translocated nuclear copies of mitochondrial genes are identical, but over time, if there is reduced
selection pressure on the nuclear copy, nucleotide substitutions and indels may introduce stop codons and shifts in the reading frame of protein-coding genes, resulting in non-functional mtDNA-like sequences. Detection of Numts derived from mitochondrial ribosomal RNA (rRNA) or transfer RNA (tRNA) genes is more challenging, since Numts may be identified only by changes in the inferred secondary structure of the transcribed RNA (Zhang & Hewitt 1996). However, conservation of the secondary structure does not eliminate the possibility of ribosomal Numts (Olson & Yoder 2002).

These unusual mtDNA-like sequences have been found in protists, plants, fungi, and animals (Hazkanic-Covo et al. 2010). More relevant to marine biologists, Numts seem to be especially common in crustaceans (Williams & Knowlton 2001, Buhay 2009, Schubart 2009), sea urchins (Jacobs & Grimes 1986), tunicates (Richly & Leister 2004), and fishes (Antunes & Ramos 2005), and have been found more recently in sponges (Covo et al. 2010). More relevant to marine biologists, Numts will be found most likely in all marine organisms.

The proliferation of questionable mitochondrial sequences in GenBank and the use of these sequences in studies is problematic (Buhay 2009). The most well-known example of an inadvertent inclusion of Numts is by Woodward et al. (1994), where contaminant human Numts were interpreted as dinosaur DNA (Zischler et al. 1995). Buhay (2009, her Table 1) identified in GenBank at least 14 different data sets that contain Numts; however, the problem is surely more widespread, as the search was restricted to crustacean COI data sets. The inclusion of Numts in species identification and phylogeographic/phylogenetic studies could yield an over-inflated number of species and misleading patterns of population subdivision/phylogenies (van der Kuyl et al. 1995, Song et al. 2008, Buhay 2009, Schubart 2009). The percent divergence between Numts and mtDNA sequence varies, but segments of extremely high similarity are present (Woischink & Moraes 2002). Since population-level studies are usually comprised of highly similar sequences, the inclusion of non-(or slightly) differentiated Numts will surely go undetected. Therefore, countermeasures such as performing PCR on diluted DNA templates, cDNA amplification, and long-range amplifications should be taken to preferably amplify orthologous mtDNA sequences (Calvignac et al. 2011). Attention also should be placed on a more critical use and curation of mitochondrial sequences (Buhay 2009), especially nowadays with the increasing number of BoL studies. Inclusion of non-orthologous mtDNA sequences in BoL will lead to an overestimation of the number of species (Song et al. 2008), but advocates of BoL are not considering this potential pitfall as a problem, because of the low incidence of non-orthologous sequences in their studies (Hebert et al. 2004).

**Discrimination between Numts and authentic mtDNA sequences**

The general labeling of all mtDNA-like sequences as Numts deserves further discussion. There are several possibilities where the unusual sequences could represent real mitochondrial sequences and not Numts, such as male and female mitochondrial lineages in a species, genomes of aging mitochondria, and/or amplifications of damaged DNA templates.

There are well-known instances where both male and female lineages of mitochondrial genomes exist in single individuals. Several marine and freshwater bivalve species carry both maternally and paternally inherited mtDNAs and transmit the copy specific to their sex to their progeny through a mechanism known as doubly uniparental inheritance (DUI) (Zouros et al. 1994, Doucet-Beaupré et al. 2010). The extent of this phenomenon in other molluscs or other phyla is unknown. The average sequence divergence between the male and the female mtDNA lineages in 3 *Mytilus* species is ~8.3% (Rawson & Hilbish 1995), and can be >20% in other species (Breton et al. 2007). These divergence estimates would alarm a ‘Numt-aware’ biologist who is expecting to recover a single mitochondrial sequence from a homoplasmic specimen. Both mitochondrial lineages are fully functional (Dalziel & Stewart 2002, Obata et al. 2011), yet the amplification of 2 divergent copies of the same gene (one male and one female copy) may lead investigators to wrongly conclude that they have amplified Numts and exclude one of the sequences from further analysis.

There is a second mechanism by which true but highly unusual mitochondrial sequences may be amplified. According to the mitochondrial free radical theory (Harman 1992), as organisms age, reactive oxygen species produced during respiration damage mitochondrial proteins, lipids, and mtDNA (Kujoth et al. 2007). Oxidatively damaged mtDNA can be repaired by excision repair enzymes, but the activity of repair enzymes declines with increasing age (de Souza-Pinto et al. 2008). Unrepaired mutations caused by both damaged mtDNA and DNA replica-
tion errors accumulate over time, leading to the decline of cellular functions and eventual aging (Kujoth et al. 2007, Holt 2010). In an aging cell, the population of mitochondria is increasingly dominated by mutant and by partially deleted mtDNA, since the shorter, mutant mtDNA can replicate faster than wild-type mtDNA (Diaz et al. 2002). Rearrangements and deletions of mtDNA have been found in *Caenorhabditis elegans* (Melov et al. 1995) and *Drosophila melanogaster* (Yui et al. 2003) in aging studies.

Similar to DNA templates from aging tissues, amplification and sequencing of damaged or ancient templates may yield sequences with erroneous base substitutions or chimeric amplicons (Pääbo et al. 1990). Even though fresh tissues are sought for genetic comparisons, museum specimens may be the only source of rare species. Sometimes, these specimens are decades or even over a century old, and the method of fixation may be inadequate for DNA work (e.g. use of formalin, or low-quality ethanol, or poor preservation techniques). DNA extractions and amplifications from these templates may produce unusual results, yet still the resulting sequences could be of mitochondrial origin. The incidence of *in vitro* recombination of authentic and Numt sequences during PCR, which may also produce anomalous mtDNA-like sequences, has been attributed to damaged templates (Thalmann et al. 2004).

**Perspectives on mtDNA evolutionary novelties**

True Numts are not hard to find. Whole-genome analyses have shown unequivocally that Numts exist in eukaryotic genomes (Richly & Leister 2004, Kim et al. 2006, Hlaing et al. 2009, Hazkani-Covo et al. 2010). However, the use of the term ‘Numts’ for every unusual sequence is a misnomer because there are other alternatives that may explain the genomic origin and functionality of these sequences. The term ‘mtDNA-like’ better describes these sequences, unless proven otherwise. How important is this distinction? Perhaps for the marine molecular systematist, population geneticist, or barcoding investigator, it is not critical, because these unusual DNA sequences could be detected and excluded from the analysis. A flow chart of suggested actions to reduce the risk of including non-homologous mtDNA sequences has been laid out by Song et al. (2008). However, unusual sequences should not be called Numts, not only because it may be a misnomer, but because researchers may be passing up the opportunity to explore evolutionary novelties in mtDNA. In addition to gender-specific lineages and age-related deterioration of mtDNA, other exemplar mitochondrial novelties include recombination, heteroplasm in introgression (Rokas et al. 2003), gene rearrangements, and unconventional architectures of mtDNA genomes.

One of the long-held analytical advantages of using mtDNA markers for genetic studies was the absence of recombination. Ever since the highly debated paper by Awadalla et al. (1999) brought mitochondrial recombination back into the spotlight, it has been generally accepted that there is both intra- (Lunt & Hyman 1997) and inter-mitochondrial recombination (Ladoukakis & Zouros 2001), regardless of how difficult it is to detect statistically (Rokas et al. 2003). Mitochondrial recombination has been proposed as one of the possible mechanisms to explain the evolution of tandem repeats (Hoelzel et al. 1993, Campbell & Barker 1999), a common feature in mtDNA. Tandem duplication followed by deletion likely causes gene rearrangements (Moritz et al. 1987, Lavrov et al. 2002). Usually the mitochondrial gene arrangements are highly conserved, which is why they have been used extensively in deep metazoan phylogenetic studies (Boore et al. 1995, Lavrov et al. 2002). However, the degree of conservation of gene arrangements can greatly vary from taxon to taxon (Miya et al. 2001, Rawlings et al. 2001, Cunha et al. 2009). During phylogenetic studies, investigators usually use a set of primers to amplify each desired gene for all species. If the DNA primers are designed in the flanking regions of the target gene, and gene rearrangements of the target or the flanking regions have occurred in some species, the PCR reactions will either fail or result in an unusually long or short sequence. Such results may indicate a novel rearrangement of metazoan mtDNA. However, an investigator may disregard these sequences as non-orthologous mtDNA sequences or Numts.

During interspecific crosses (i.e. introgression), the mechanism eliminating paternal mitochondria from the zygote can break down, allowing mitochondria from both species to co-exist in the hybrid F1, resulting in heteroplasm. Recombination between the non-homologous mitochondria will result in haplotypes that may persist in the population through backcrosses with either parental species (Rokas et al. 2003). Depending on the spread of these recombinant haplotypes into the population, amplicons from such specimens with heteroplasmic mtDNA may result in unusual sequences with heterozygote positions, unreadable sequences (if the sequences differ by an indel or more), or chimeric sequences if the marker transcends the area of recombination. Het-
eroplasm through introgression has been observed in crossings between blue mussel species (Kijewski et al. 2006). Heteroplasm has also been reported in cetaceans (Vollmer et al. 2011), anchovies (Magoulas & Zouros 1993), and flounders (Hoarau et al. 2002), although in these cases, the unusual sequences are most likely caused by paternal leakage of mtDNA. If researchers are unaware that introgression may be taking place in the species of study, such sequences can be easily disregarded as Numts or as other contaminants. Since introgression is rather common in the marine environment (see review by Arnold & Fogarty 2009), opportunities may be missed to document instances of hybridization.

The typical metazoan mtDNA genome is a circular molecule, about 16 kb long, consisting of 13 protein-coding genes, 22 tRNA- and 2 rRNA-coding genes, and an AT-rich control region. The genetic information is tightly packed and characterized by stasis in gene content, compared to other eukaryotes (Adams & Palmer 2003). A glimpse of ancestral features of the metazoan mitochondrial genome is offered by the largest metazoan mtDNA genome in the basal phylum Placozoa (Signorovitch et al. 2007), introns within protein-coding regions in both Cnidaria and Placozoa (Boore 1999, Signorovitch et al. 2007), and the large intergenic regions in Porifera (Erpenbeck et al. 2009).

We are now discovering that among the higher metazoans, mtDNA genomes also vary in size, gene content, gene order, and rates of sequence evolution. Exemplar non-conventional architectures of metazoan mtDNA genomes are the unusually large mtDNA genomes of the isopod Armadillidium vulgare (Raymond et al. 1999), the frequent re-arrangements in demosponges (Wang & Lavrov 2008) and vermetid gastropods (Rawlings et al. 2010), and the 2 circular mitochondrial chromosomes in the freshwater rotifer Brachionus plicatilis (Suga et al. 2008). The marine realm offers a unique evolutionary novelty in the mitochondrial genomes of cnidarian octocorals, which possess a coding gene for a mismatch repair protein (MSH) (Pont-Kingdon et al. 1995) recently inferred to be the first case of horizontal gene transfer into a metazoan mitochondrial genome (Bilewitch & Degnan 2011), a hypothesis that challenges the way we thought metazoan mitochondrial genomes evolve.

**Conclusions**

As we continue to analyze the mtDNA from an ever-increasing number of species, we are becoming more aware of the complexities of the sequenced data. We should avoid the dichotomy of choices (mtDNA or Numts), because there are many surprises left to discover in mtDNA evolution. With the advent of next-generation sequencing, more genome projects will yield more Numts (Hazkani-Covo et al. 2010) and will not only expose the depth of true Numts in the nuclear genomes, but also help in understanding the evolutionary history of these mitochondrial-derived sequences and novelties of the mitochondrial genomes.

**Acknowledgments.** The manuscript was improved significantly by comments made by 3 anonymous reviewers.

**LITERATURE CITED**

- Bilewitch JP, Degnan SM (2011) A unique horizontal gene transfer event has provided the octocoral mitochondrial genome with an active mismatch repair gene that has potential for an unusual self-contained function. BMC Evol Biol 11:228


Pääbo S, Irwin DM, Wilson AC (1990) DNA damage promotes jumping between templates during enzymatic


Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proc Natl Acad Sci USA 105: 13486–13491


Editorial responsibility: Matthias Seaman, Oldendorf/Luhe, Germany

Submitted: April 13, 2011; Accepted: September 21, 2012

Proofs received from author(s): October 10, 2012