Myrmecological News	12	217-218	Vienna, September 2009

Focus

Numts and mitochondrial pseudogenes

Prof. Dr. Andrew T. Beckenbach, Department of Biological Sciences, Simon Fraser University, Burnaby, BC V5A 1S6 Canada. E-mail: beckenba@sfu.ca

Myrmecol. News 12: 217-218 (online 29 June 2009)
ISSN 1994-4136 (print), ISSN 1997-3500 (online)
Received 22 May 2009; revision received 2 June 2009; accepted 2 June 2009

Mitochondrial gene sequences are widely used for population and phylogenetic analyses (SIMON & al. 2006) and form the basis for current DNA barcoding initiatives (HEBERT & al. 2003). A major advantage is the unambiguous homology of the functional mitochondrial genes across the animal kingdom. This "homology" feature extends only to the functional copy of each gene. Duplicates of mitochondrial genes often arise, both as duplications within the mitochondrial genome, and as **nu**clear insertions of **mit**ochondrial sequences (**numts**). These duplicates can greatly complicate the use of mitochondrial sequences for phylogenetic reconstruction, while at the same time provide unique opportunities to study the evolution of DNA sequences (BENSASSON & al. 2001). The purpose of this Focus article is to encourage the analysis and reporting of numts and mitochondrial pseudogenes when they are encountered.

Methodology

Prior to the introduction of PCR, analysis of mitochondrial sequences required shotgun cloning from total genomic DNA and screening of those clones with a DNA probe derived from known mitochondrial sequence. The screening process finds sequences having a general, overall similarity to part or all of the probe sequence. This process often revealed false positives, having high similarity to at least part of the probe, but not corresponding to the functional genes. The majority of these non-functional copies have been shown to be numts.

PCR combines the screening and purification processes into a single step, and eliminates the need for cloning when a single product is obtained. Selectivity is determined by exact, or nearly exact, match of two 20 to 25 nucleotide sequences corresponding to the primers used in the PCR amplification. It is likely that each primer binds at multiple sites within the nuclear and mitochondrial genomes, but amplification occurs only where primer binding sites are located an appropriate distance apart, on opposite strands and in the proper orientation. In practice, this requirement usually turns out to be more selective than probing of a genomic library, where a partial match to any part of the probe generates a positive.

Recent complete nuclear genome sequencing efforts have allowed more exact searches for numts in certain model organisms (RICHLY & LEISTER 2004, PAMILO & al. 2007). In contrast to the molecular (wet lab) methods, this approach specifically targets non-functional mitochondrial-like sequences in the nuclear genome. Much like the pre-PCR methodology, the genome database is probed with part or all of the mitochondrial DNA sequence for that species, using a search algorithm such as BLAST. This technique finds all mitochondrial-like sequences in the nuclear genome, locates the sequences precisely within the genome and provides the sequence of each numt. Thus a complete survey of numts can be made for that species.

Properties of numts and mitochondrial pseudogenes

Scans of complete nuclear genome sequences have revealed considerable information about numts. They are apparently absent from the genome of the mosquito *Anopheles gambiae*, present in only five sites (total of 777 bp) in *Drosophila melanogaster*, but are present in thousands of sites in other organisms such as *Homo sapiens* and the honeybee, *Apis mellifera* (see RICHLY & LEISTER 2004, BEHURA 2007, PAMILO & al. 2007). At present, there is little pattern evident in the taxon distribution of numts. Though they are absent from one *Anopheles* mosquito, they are present at about 100 sites in another mosquito, *Aedes aegypti* (HLAING & al. 2009).

In most organisms, the majority of numts are small, with about 90% smaller than 500 bp in length (RICHLY & LEISTER 2004). Small numts should not interfere with standard PCR surveys of mtDNA, but in most species with large numbers of numts at least a few copies include significant portions of the mitochondrial genome, exceeding 1000 bp (RICHLY & LEISTER 2004, PAMILO & al. 2007, HLAING & al. 2009).

Much less is known about mitochondrially encoded pseudogenes. Duplications within the mitochondrial genome occur frequently and probably account for much of the observed size variation encountered in mitochondrial genome sequences (BOORE 1999).

Fate of numts and mitochondrial pseudogenes

Numts likely arise when part or all of the mtDNA is directly integrated into sites in the nuclear genome, or mitochondrial transcripts are reverse transcribed and integrated. They are non-functional at the time of integration (BENSASSON & al. 2001). Once integrated, the evolution of numts slows down dramatically despite the apparent absence of selective constraints. Experimental evidence from vertebrates and *Drosophila* indicates a 10× reduction in mutation rate of numts compared to the functional mitochondrial sequences. As a result, numts appear frozen in time, and insertions occurring 40 - 60 million years ago are easily recognizable (BENSASSON & al. 2003).

The fate of duplications within the mitochondrial genome is quite different. Such duplications should experience the same mutation rate as the functional mitochondrial genes, but in the absence of purifying selection will accumulate mutations much more rapidly. Mitochondrial duplications appear to quickly lose their identity, perhaps accounting for the rarity of pseudogenes in completely sequenced mitochondrial genomes.

Practical considerations

It has been noted that inadvertent inclusion of pseudogene sequences in phylogenetic surveys can lead to misleading phylogenies (SORENSEN & QUINN 1998). A taxon represented by an ancient numt may appear as the sister to the branch harboring the numt. A taxon represented by a mitochondrial pseudogene may show an exaggerated branch length. How serious a problem do numts and mitochondrial pseudogenes present?

An empirical study of PCR selectivity for functional mitochondrial sequences in human genomic extracts showed virtually no risk of co-amplification of numts using standard techniques, despite the high frequency of numts within the genome (GOIOS & al. 2008). Part of the reason is that most numts are small and therefore lack one or both primer binding sites. The few numts large enough to include both binding sites are present in only two copies per diploid cell. As the mitochondrial genome may be present in thousands of copies per cell, the numt amplification represents only minor contaminants.

On the other hand, several studies have shown numts to pose significant problems with phylogenetic analyses (ZHANG & HEWITT 1996, WILLIAMS & KNOWLTON 2001). One property of numts may make them more likely to amplify than the mitochondrial sequence from which they were derived: the slow rate of evolution of numts. Mutations of the mitochondrial sequences may reduce binding ability of conserved primers, while numts are not affected.

Mitochondrial pseudogenes may also pose a problem, as they are present in each mitochondrial genome. This property would allow them to co-amplify along with the functional copy, and could be particularly problematic for recent duplications.

Numts and mitochondrial pseudogenes, when encountered in phylogenetic surveys, inevitably require additional effort to sort out the problem. BENSASSON & al. (2001) provide suggestions for identifying numts. Distinguishing pseudogenes from the functional gene sequences is not always straightforward. For protein coding genes, we expect the functional gene to be an open reading frame of the appropriate length. Unfortunately, numts because of their slow rate of evolution, may not acquire obvious defects for many millions of years. Worse, the recent discovery of frameshift mutations in functional mitochondrial protein coding genes in many animal groups including ants (BECKENBACH & al. 2005) shows that even this criterion may result in misidentifying functional sequences as non-functional pseudogenes.

Numts and mitochondrial pseudogenes in ants

There appear to be few published studies of numts in ants. MARTINS & al. (2007) characterize two numts derived from the mitochondrial *cox1* gene in ants of the genus *Atta*. KRONAUER & al. (2007) identified several probable numts in a survey of *Dorylus* ants while BEIBL & al. (2007) found no evidence of numt contamination in two genera of ants. In addition, several unpublished mitochondrial-like pseudogenes have been deposited in GenBank. Numts are almost certainly under-reported in ants, as in many other organisms. The amount of effort necessary to analyze taxa having pseudogene contamination may seem excessive if the primary goal is to obtain a molecular phylogeny. Nonetheless, numts and mitochondrial pseudogenes are of considerable interest from the standpoint of molecular evolution.

References

BECKENBACH, A.T., ROBSON, S.K.A. & CROZIER, R.H. 2005: Single nucleotide +1 frameshifts in an apparently functional mitochondrial cytochrome b gene in ants of the genus *Polyrhachis*. – Journal of Molecular Evolution 60: 141-152.

BEHURA, S.K. 2007: Analysis of nuclear copies of mitochondrial sequences in honeybee (*Apis mellifera*) genome. – Molecular Biology and Evolution 24: 1492-1505.

BEIBL, J., BUSCHINGER, A., FOITZIK, S. & HEINZE, J. 2007: Phylogeny and phylogeography of the Mediterranean species of the parasitic ant genus *Chalepoxenus* and its *Temnothorax* hosts. – Insectes Sociaux 54: 189-199.

BENSASSON, D., FELDMAN, M.W. & PETROV, D.A. 2003: Rates of DNA duplication and mitochondrial DNA insertion in the human genome. – Journal of Molecular Evolution 57: 343-354.

BENSASSON, D., ZHANG, D.-X., HARTL, D.L. & HEWITT, G.M. 2001: Mitochondrial pseudogenes: evolution's misplaced witnesses. – Trends in Ecology & Evolution 16: 314-321.

BOORE, J.L. 1999: Animal mitochondrial genomes. – Nucleic Acids Research 27: 1767-1780.

GOIOS, A., PRIETO, L., AMORIM, A. & PEREIRA, L. 2008: Specificity of mtDNA-directed PCR–Influence of NUclear MTDNA (NUMT) contamination in routine samples and techniques. – International Journal of Legal Medicine 122: 341-345.

HEBERT, P.D.N., CYWINSKA, A., BALL, S.L. & DEWAARD, J.R. 2003: Biological identifications through DNA barcodes. – Proceedings of the Royal Society of London Series B 270: 313-322.

HLAING, T., TUN-LIN, W., SOMBOON, P., SOCHEAT, D., SETHA, T., MIN, S., CHANG, M.S. & WALTON, C. 2009: Mitochondrial pseudogenes in the nuclear genome of *Aedes aegypti* mosquitoes: implications for past and future population genetic studies. – BioMed Central Genetics 10: 11.

KRONAUER, D.J.C., SCHÖNING, C., VILHELMSEN, L.B. & BOOMSMA, J.J. 2007: A molecular phylogeny of *Dorylus* army ants provides evidence for multiple evolutionary transitions in foraging niche. – BioMed Central Evolutionary Biology 7: 56.

MARTINS Jr., J., SOLOMON, S.E., MIKHEYEV, A.S., MUELLER, U.G., ORTIZ, A. & BACCI Jr., M. 2007: Nuclear mitochondrial-like sequences in ants: evidence from *Atta cephalotes* (Formicidae: Attini). – Insect Molecular Biology 16: 777-784.

PAMILO, P., VILJAKAINEN, L. & VILHAVAINEN, A. 2007: Exceptionally high density of NUMTs in the honeybee genome. – Molecular Biology and Evolution 24: 1340-1346.

RICHLY, E. & LEISTER, D. 2004: NUMTs in sequenced eukaryotic genomes. – Molecular Biology and Evolution 21: 1081-1084.

SIMON, C., BUCKLEY, T.R., FRATI, F., STEWART, J.B. & BECKENBACH, A.T. 2006: Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. – Annual Review of Ecology, Evolution and Systematics 37: 545-579.

SORENSON, M.D. & QUINN, T.W. 1998: Numts: a challenge for avian systematics and population biology. – The Auk 115: 214-221.

WILLIAMS, S.T. & KNOWLTON, K. 2001: Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus *Alpheus*. – Molecular Biology and Evolution 18: 1484-1493.

ZHANG, D.-X. & HEWITT, G.M. 1996: Highly conserved nuclear copies of the mitochondrial control region in the desert locust, *Schistocerca gregaria*: some implications for population studies. – Molecular Ecology 5: 295-300.