



# Mitochondrial DNA Part A

## DNA Mapping, Sequencing, and Analysis

ISSN: 2470-1394 (Print) 2470-1408 (Online) Journal homepage: <http://www.tandfonline.com/loi/imdn21>

## No homology means there can be no analyses; a comment on Jose & Harikrishnan

Timothy J. Page & Dirk Steinke

To cite this article: Timothy J. Page & Dirk Steinke (2016): No homology means there can be no analyses; a comment on Jose & Harikrishnan, Mitochondrial DNA Part A, DOI: [10.1080/24701394.2016.1267159](https://doi.org/10.1080/24701394.2016.1267159)

To link to this article: <http://dx.doi.org/10.1080/24701394.2016.1267159>



Published online: 26 Dec 2016.



Submit your article to this journal [↗](#)



Article views: 3



View related articles [↗](#)



View Crossmark data [↗](#)

Full Terms & Conditions of access and use can be found at  
<http://www.tandfonline.com/action/journalInformation?journalCode=imdn21>

LETTER TO THE EDITOR

## No homology means there can be no analyses; a comment on Jose & Harikrishnan

Sir,

A recent paper (Jose & Harikrishnan 2016) considers some issues surrounding the use of mitochondrial (mt) cytochrome *c* oxidase subunit I (*COI* or *COX1*) sequences as DNA barcodes. They provide a particular example of how the presence of two commonly used fragments of the gene, that are largely non-homologous, could lead to confusion for taxonomy, in this case for decapod crustaceans. Jose and Harikrishnan (2016) show the divergence between two *COI* fragments (termed 'Folmer' and 'Palumbi') for some freshwater shrimp taxa (Family Palaemonidae, Genus *Macrobrachium*) through a number of analyses, including neighbor-joining phylograms, sequence distance calculations and pairwise  $F_{ST}$ .

Firstly, we do not think there really should be any confusion, as the 'Folmer' portion of *COI* (a.k.a. 5' *COI*; LCO-HCO) is commonly accepted as the 'DNA barcode' (Hebert et al. 2003; Raupach & Radulovici 2015). While there are plenty of arguments for and against DNA barcoding (Collins & Cruickshank 2013), and it is true that other fragments could serve just as well as a barcode (Page & Hughes 2010) (and are used as such for protists, plants, fungi), it is the very standardization of a precise fragment and term that makes DNA barcoding particularly useful. The presence of the other *COI* fragment ('Palumbi', a.k.a. 3' *COI*; Jerry-Pat) is well-known, since the primers date from the early 1990s (Palumbi et al. 1991). The relationship between the two fragments is no secret and has been explored specifically within crustaceans (Lefébure et al. 2006; Roe & Sperling 2007). The effect of relying on a precise mitochondrial gene fragment for species identification has also been previously considered for this same genus of shrimps (Page & Hughes 2011).

Secondly, we see no problem or potential confusion with the presence of multiple DNA fragments, as these are very useful for phylogenetic analyses. It is basic scientific procedure to ensure that you know precisely which loci you are employing, as this represents simple data-quality checking. It is generally not hard to work out where a mitochondrial sequence belongs in a genome, as is the case for the example taxon, *Macrobrachium rosenbergii*, for which there exists an annotated mitochondrial genome (Miller et al. 2005). A simple GenBank BLASTN (blast.ncbi.nlm.nih.gov) or BOLD search (www.barcodinglife.org) should also tell you whether your fragment has been sequenced previously for the taxon in question and closely related taxa. However, the results need to be scrutinized carefully to see how close the resulting sequences really are.

Thirdly and most importantly, it should not really be possible to do any meaningful analyses or derive any results

from the datasets in Jose and Harikrishnan (2016) since they are not homologous. When sequences are aligned (as they have been in Jose & Harikrishnan 2016, somehow), it is a hypothesis and a statement of homology (basically, comparing like with like). It is saying that we believe position X in our alignment has a shared evolutionary history for all of our sequences, and thus can be compared. It is the same basic precept for a comparative morphological analysis, where the relevant character being compared must represent the same structure in evolutionary terms, or else no meaningful analyses can be undertaken. It is a classic scenario of ensuring that apples are compared solely with other apples.

Jose and Harikrishnan (2016) align the two different fragments (of roughly 600–650 bases each). This is theoretically possible since there is a 30 base pair overlap between both, however, a quick look at the results shows the sequences were probably not aligned thus. We tried to replicate the analyses by aligning the *M. rosenbergii* sequences from Jose and Harikrishnan (2016) in MEGA version 6 (Tamura et al. 2013). We also included the full *COI* sequence from the *M. rosenbergii* mt genome (GenBank Reference Sequence: NC\_006880.1) as a scaffold from which to hang the two fragments, and so we were able to successfully align the sequences accurately. The first 627 bases of the Folmer fragment align only with the genome sequence. The remaining final 30 bases of the Folmer sequence also align with the first 30 bases of the Palumbi fragment, which then go on to continue for another 573 bases, aligning only with the genome. While fairly meaningless, as there is only a 30 base overlap in the full 1230 base pair dataset, it is indeed possible to create a neighbor-joining phylogram and to derive Kimura two-parameter model (K2P) distance measurements between the fragments, in line with Jose and Harikrishnan (2016) and traditional DNA barcoding practice (Costa et al. 2007). Our K2P distance here is 0.07, which reflects two bases that differ between the two fragments within the small overlapping region.

However, this does not reflect the results from Jose and Harikrishnan (2016), which find K2P distances of greater than 1.00 ( $F_{ST}=1$ ), which is essentially 100% different. The only way we could recreate these results was if the two completely different fragments were forced together into a sort of chimera, with no concern for homology, and then analyzed. It is not valid to force non-homologous regions of DNA into a phylogeny as this is essentially meaningless. You might as well measure the wing length 'distance' in per cent between an albatross and a hippo, or listen to one hand clapping; all are essentially philosophical exercises.

Jose and Harikrishnan (2016) say that the inclusion of non-homologous loci could lead to 'severe problems in

establishing ancestral species, since the phylogenetic tree will possess inconsistent topology', but that is not really the case. Basic preliminary data checking should establish that the included sequences represent different fragments, and will not (and cannot) be fully aligned in the first place, thus there will be no topology and no subsequent confusion.

As Jose and Harikrishnan (2016) point out, there are many legitimate questions surrounding DNA barcoding still needing exploration. These include whether the locus labels or taxonomic names associated with sequences on public databases are accurate (Steinke & Hanner 2011) and whether those sequences themselves are of a high enough quality (Buhay 2009). Buhay (2009) makes a pertinent (and hilarious) point that unrecognized dubious data, such as nuclear copies of mitochondrial DNA segments (numts) (but in her example, the text of her favorite pumpkin soup recipe), can be aligned with DNA sequences and then used to produce a meaningless phylogram with statistical outputs. Ironically, there is at least one legitimate use for non-directly homologous regions in a phylogeny. This occurs when a taxon has no close relatives, and so a numt can be used as an outgroup (Hay et al. 2004), although ultimately the mitochondrial gene and its nuclear copy do have a common root. Homology itself is at the root of all of these types of evolution-based analyses, for DNA barcoding or general phylogenetics, and we forget this at the expense of meaningful results.

## Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

## References

- Buhay JE. 2009. "COI-Like" sequences are becoming problematic in molecular systematic and DNA barcoding studies. *J Crust Biol.* 29:96–110.
- Collins RA, Cruickshank RH. 2013. The seven deadly sins of DNA barcoding. *Mol Ecol Resour.* 13:969–975.
- Costa FO, deWaard JR, Boutillier J, Ratnasingham S, Dooh RT, Hajibabaei M, Hebert PDN. 2007. Biological identifications through DNA barcodes: the case of Crustacea. *Can J Fish Aquat Sci.* 64:272–295.
- Hay JM, Sarre SD, Daugherty CH. 2004. Nuclear mitochondrial pseudogenes as molecular outgroups for phylogenetically isolated taxa: a case study in *Sphenodon*. *Heredity.* 93:468–475.
- Hebert PDN, Cywinska A, Ball SL, DeWaard JR. 2003. Biological identifications through DNA barcodes. *Proc Biol Sci.* 270:313–321.
- Jose D, Harikrishnan M. 2016. Non-homologous COI barcode regions: a serious concern in decapod molecular taxonomy. *Mitochondrial DNA Part A.* doi: 10.3109/19401736.2015.1137902.
- Lefebvre T, Douady CJ, Gibert J. 2006. Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. *Mol Phylogenet Evol.* 40:435–447.
- Miller AD, Murphy NP, Burrige CP, Austin CM. 2005. Complete mitochondrial DNA sequences of the decapod crustaceans *Pseudocarcinus gigas* (Menippidae) and *Macrobrachium rosenbergii* (Palaemonidae). *Mar Biotechnol.* 7:339–349.
- Page TJ, Hughes JM. 2010. Comparing the performance of multiple mitochondrial genes in the analysis of Australian freshwater fishes. *J Fish Biol.* 77:2093–2122.
- Page TJ, Hughes JM. 2011. Neither molecular nor morphological data have all the answers; with an example from *Macrobrachium* (Decapoda: Palaemonidae) from Australia. *Zootaxa.* 2874:65–68.
- Palumbi SR, Martin A, Romano S, McMillan WO, Stice L, Grabowski G. 1991. A simple fool's guide to PCR Honolulu. Honolulu, HI: University of Hawaii Press.
- Raupach MJ, Radulovici AE. 2015. Looking back on a decade of barcoding crustaceans. *Zookeys.* 539:53–81.
- Roe AD, Sperling FAH. 2007. Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. *Mol Phylogenet Evol.* 44:325–345.
- Steinke D, Hanner R. 2011. The FISH-BOL collaborators' protocol. *Mitochondrial DNA.* 22:10–14.
- Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 30:2725–2729.

Timothy J. Page

*Australian Rivers Institute, Griffith University, Nathan, QLD, Australia*

*Water Planning Ecology, Queensland Dept. of Science, Information Technology and Innovation, Dutton Park, QLD, Australia*

 penguintim@hotmail.com

Dirk Steinke

*Biodiversity Institute of Ontario, University of Guelph, Guelph, ON, Canada*

Received 8 September 2016; accepted 27 November 2016

© 2016 Informa UK Limited, trading as Taylor & Francis Group