

## FULL LENGTH RESEARCH PAPER

# Thrice better than once: quality control guidelines to validate new mitogenomes

Fidel Botero-Castro, Frédéric Delsuc, and Emmanuel J. P. Douzery

Université Montpellier 2, CNRS &amp; IRD, Institut des Sciences de l'Évolution (UMR 5554), Montpellier, France

**Abstract**

Mitogenomic data are increasingly used in evolutionary biology and ecology, stressing the importance for double checking the authenticity of DNA sequences. For example, Szcześniak et al. (2013) recently published the mitochondrial genome of a bat, the Leschenault's rousette (*Rousettus leschenaultii*). Here we show using straightforward phylogenetic analyses of available chiropteran sequence data that the taxonomic attribution of the reported mitogenome is erroneous. The purportedly-new complete mitochondrial genome likely belongs to the Egyptian fruit bat (*R. aegyptiacus*) for which a reference sequence already exists. We propose that future articles reporting complete mitochondrial genome sequences should mandatorily include maximum likelihood trees inferred from (i) the standard barcoding marker for the taxon under focus, which would benefit from the massive data available in public databases, and (ii) the available mitogenomes of closely related species. We also strongly advise these trees be presented as phylograms so that all pertinent phylogenetic information is displayed in the form of a topology and its associated branch lengths. Along with compulsory information on the geographical location and origin of the specimen, these new standards should help avoiding the publication of taxonomically misidentified mitogenomes that might end up as reference sequences in public databases and re-used in subsequent meta-analyses.

**Keywords**

Barcoding, bats, phylogeny

**History**

Received 21 January 2014  
Revised 25 February 2014  
Accepted 01 March 2014  
Published online 4 April 2014

**Introduction**

Since the report of the first complete mitochondrial genomes of the PCR era (e.g. Arnason et al., 1991), the number of publications of new complete mitogenomes has steadily increased, thanks to next-generation sequencing technologies (NGS) and a renewed interest in mitogenomic data. Complete mitochondrial genomes have, for example, proven useful for resolving intra-familial relationships in mammals including Elephantidae (Rohland et al., 2007), Ursidae (Krause et al., 2008), Mustelidae (Yu et al., 2011a), Delphinidae (Vilstrup et al., 2011), Bovidae and Cervidae (Hassanin et al., 2012; Wada et al., 2010), and Phyllostomidae (Botero-Castro et al., 2013). Mitogenomes are now routinely produced in large numbers to attain comprehensive taxon sampling in phylogenetic studies of diverse groups such as Primates (Finstermeier et al., 2013; Guschanski et al., 2013), frogs (Zhang et al., 2013), and particularly in teleost fish, in which the complete mitochondrial genome has been adopted as the marker of choice early on (e.g. Miya et al., 2013). Finally, mitogenomes will also probably quickly become the norm in paleogenetic (Enk et al., 2011), phylogeographic (Hirata et al., 2013) and barcoding (Dettai et al., 2012; Nelson et al., 2012) studies of metazoan taxa.

With such a high number of mitochondrial genomes quickly populating public databases (e.g. MitoZoa: D'Onorio de Meo et al., 2012) and scientific journals like *Mitochondrial DNA* with its devoted section "Mitogenome Announcements", there is an increasing need for upstream quality controls (Shen et al., 2013). Indeed, several difficulties can occur at different stages of the process of sequencing and assembling whole mitochondrial genomes. These potential pitfalls can strongly affect the use of mitogenomes in subsequent evolutionary analyses.

A first non-negligible issue comes from the misidentification of biological samples used for sequencing. Mostly due to errors in the taxonomic assignation of the source specimen, this can lead to attributing complete mitochondrial genomes to the incorrect species. This can have even more profound implications if the misidentified sequence ends up as the reference mitochondrial genome for the species in public curated databases such as the RefSeq section of GenBank (Bidartondo et al., 2008). This can also potentially magnify the number of sequences whose identification will be based solely on comparisons with the reference sequence and can thus blur phylogenetic studies.

A second issue resides in the occurrence of nuclear copies of mitochondrial genes (NUMTs). These copies can be PCR co-amplified with or amplified instead of the actual mitochondrial genes, which can result in chimerical sequences after assembly. This has been, for example, the case of the reference mitogenome of the common goat (*Capra hircus* Parma et al., 2003) for which it has been shown that several undetected NUMTs have been assembled instead of the genuine mitochondrial genes (Hassanin et al., 2010).

Correspondence: Emmanuel J. P. Douzery, Université Montpellier 2, CNRS & IRD, Institut des Sciences de l'Évolution (UMR 5554), CC064 – Place Eugène Bataillon, 34 095 Montpellier Cedex 05, France. E-mail: emmanuel.douzery@univ-montp2.fr

Finally, the sequencing process is never immune of potential contaminations from the environment or from crossed-contaminations with other samples extracted or sequenced simultaneously. This risk is increased in the construction of NGS libraries where a small exogenous DNA fragment can be amplified millions of times and, thus, sequenced instead of the target sample.

In a recent publication, Szcześniak et al. (2013) reported the sequence of the complete mitochondrial genome of the Leschenault's rousette (*Rousettus leschenaultii*; KC702803.1), providing the second mitogenome for this genus of pteropodid bats in addition to the available Egyptian fruit bat (*R. aegyptiacus*; NC\_007393.1). According to the paper, the sequence obtained from Illumina sequencing of a total genomic DNA extract has a 3200x coverage and no reference was made to the occurrence of NUMTs. However, this paper contains a number of intriguing points that are urgent to clarify, as this sequence might become the mitogenomic reference for the species. By re-analyzing the new mitogenome in the light of available mitochondrial sequences, we show that this sequence does not belong to *R. leschenaultii* and is most probably a second mitochondrial genome for *R. aegyptiacus*. We take opportunity of this particular case to suggest quality control guidelines that will hopefully help the future publication of accurate new mitochondrial genome sequences.

## Materials and methods

First, a complete mitogenome dataset including the 18 taxa of Szcześniak et al. (2013) was built in order to assess the phylogeny and level of divergence among bats. Accession numbers of all mitogenomes are provided in Supplementary Table 1.

Second, we extracted two barcoding markers from the complete mitogenomes: the protein-coding genes *cox1* and *cytb*. Barcoding markers have been shown to contribute to the detection of sequencing error, contamination, taxonomy inaccuracy, and species misidentification (Shen et al., 2013). As of August 2013, there were 63 *cox1* sequences for *R. aegyptiacus* and 21 for *R. leschenaultii* publicly available from the online identification tool of BOLD systems (Barcode Of Life Database: <http://www.boldsystems.org> – “Identification” section) (Hebert et al., 2003). Also, 281 *cytb* sequences were available from public databases for the genus *Rousettus*. We therefore analyzed the two datasets corresponding to *cox1* and *cytb* sequences available for the genus *Rousettus*. Sequences of the pteropodid *Pteropus dasymallus* (NC\_002612.1) were used as outgroup to root the phylogenetic trees.

All three datasets were aligned using Muscle (Edgar, 2004) and further adjusted manually within Geneious (Kearse et al., 2012). The control region was excluded from the mitogenomic dataset due to alignment ambiguities. Besides, the mitogenomic alignment was cleaned with trimAl (Capella-Gutiérrez et al., 2009) using the *automated1* option in order to remove ambiguously aligned positions. Final alignments included 15,330 sites for the mitogenomic dataset, 657 sites for *cox1*, and 1140 sites for *cytb*.

For each alignment, a maximum likelihood (ML) tree was inferred under the GTR+GAMMA model as implemented in RAXML v7.4.2 (Stamatakis, 2006). Node robustness was evaluated by performing 100 thorough bootstrap replicates (called with the -b option in RAXML). The resulting topologies and branch lengths were used to evaluate the phylogenetic position and divergence of the recently released mitogenome (Szcześniak et al., 2013) with respect to other *R. leschenaultii* and *R. aegyptiacus* available sequences.

## Results and discussion

### Phylogenetic evidence for misidentification of a new *R. leschenaultii* mitogenome

When a new mitogenome is reported, its amount of genetic divergence with respect to closely related taxa can be assessed. Szcześniak et al. (2013) stated that when compared to the available mitochondrial genome of *R. aegyptiacus*, their new mitogenome has a similarity of 99.7% (i.e. a genetic distance of only 0.3%). This value alone is a warning about a potential identification problem since it appears surprisingly low for two complete mitochondrial genomes belonging to different bat species. Indeed, in mammals, interspecific genetic distances for sister taxa have been shown to reach more than 5% based on *cytb* (Nabholz et al., 2008). Moreover, this value is often used as a threshold to identify cryptic species, leading to the description of species that could not be distinguished on the basis of morphological characters solely, according to the genetic species concept (Baker & Bradley, 2006). For example, within the phyllostomid bat genus *Carollia*, the species *C. benkeithi* (Solari & Baker, 2006) and *C. sowelli* (Baker et al., 2002) have been distinguished thanks to molecular data.

The quasi-identity of the new mitochondrial sequence with the one of *R. aegyptiacus* is also evidenced in the ML mitogenomic tree. Unfortunately, Szcześniak et al. (2013) chose to represent their ML tree as a cladogram (Figure 1, right). This kind of representation actually erases half of the phylogenetic information since it only displays a topology with arbitrary branch lengths, thus masking the evolutionary divergence among taxa. The suspicious similarity between the two sequences is best viewed in the ML phylogram (Figure 1, left) in which there is virtually no difference between their terminal branch lengths reflecting an extremely low number of substitutions. The inferred phylogram also shows that the divergence between the two mitogenomes is considerably lower if compared to other pairs of congeneric bat species such as within the genera *Artibeus*, *Rhinolophus*, and *Pteropus*.

### A barcoding perspective on the analysis of new mitogenomes

Following from the previous observations, we can reasonably hypothesize that the newly provided sequence would in fact be a second mitochondrial genome for *R. aegyptiacus* instead of a new one for *R. leschenaultii*. This can be evaluated by comparing the newly reported sequences using mitochondrial markers such as *cox1* and *cytb* for which a diversity of barcoding sequences for bats in general, and for *Rousettus*, in particular, is available in public databases. If the new mitogenome actually belongs to *R. leschenaultii*, and assuming that there has not been introgression or hybridization between the two species, *R. leschenaultii* sequences should appear monophyletic in the two gene trees.

Our results show with high confidence that the new mitochondrial genome is actually an additional sequence for *R. aegyptiacus* and not one of *R. leschenaultii*. The barcoding fragment of the *cox1* gene shows a similarity ranging from 97.4 to 100% with respect to *R. aegyptiacus* sequences over 657 sites. In contrast, similarities with respect to *R. leschenaultii* vary between 93.1 and 92.8%, i.e. at least 7% of the interspecific divergence in the *cox1* gene. This result is confirmed by the phylogenetic analyses of *cox1* – and also *cytb* – data sets, both of which consistently place the new sequences nested within *R. aegyptiacus* instead of *R. leschenaultii*, whereas the two species otherwise form strongly supported mitochondrial clades (Figure 2). In terms of sequence divergence, the clades of *R. aegyptiacus* and *R. leschenaultii* differ by much more than the

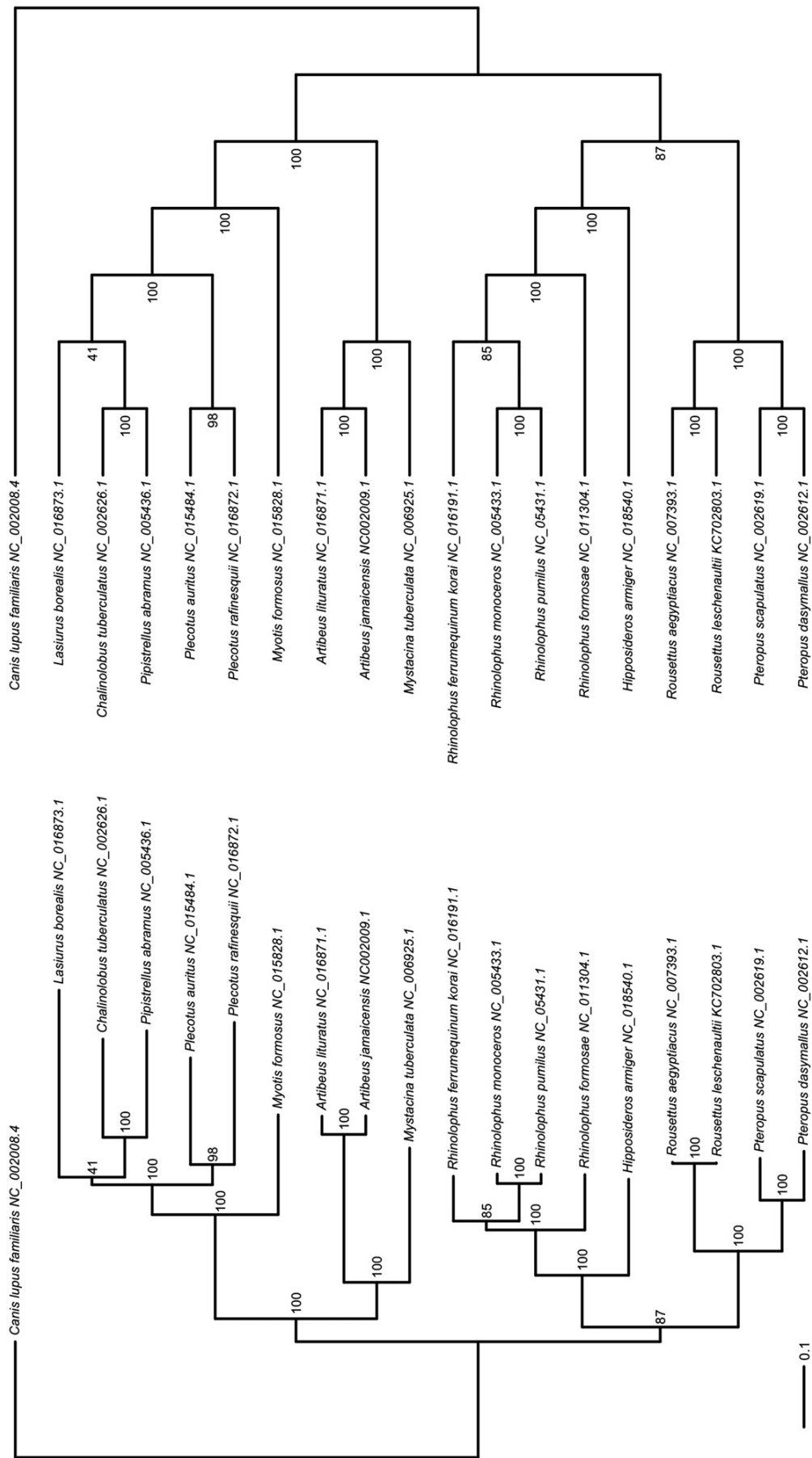


Figure 1. Comparison between the maximum likelihood tree inferred from complete mitochondrial genomes of chiropterans drawn as a phylogram (left) or as a cladogram (right). The topology is the same, but the phylogram representation incorporates branch lengths (scale: 0.1 expected substitution per site). Bootstrap support is given for each node.

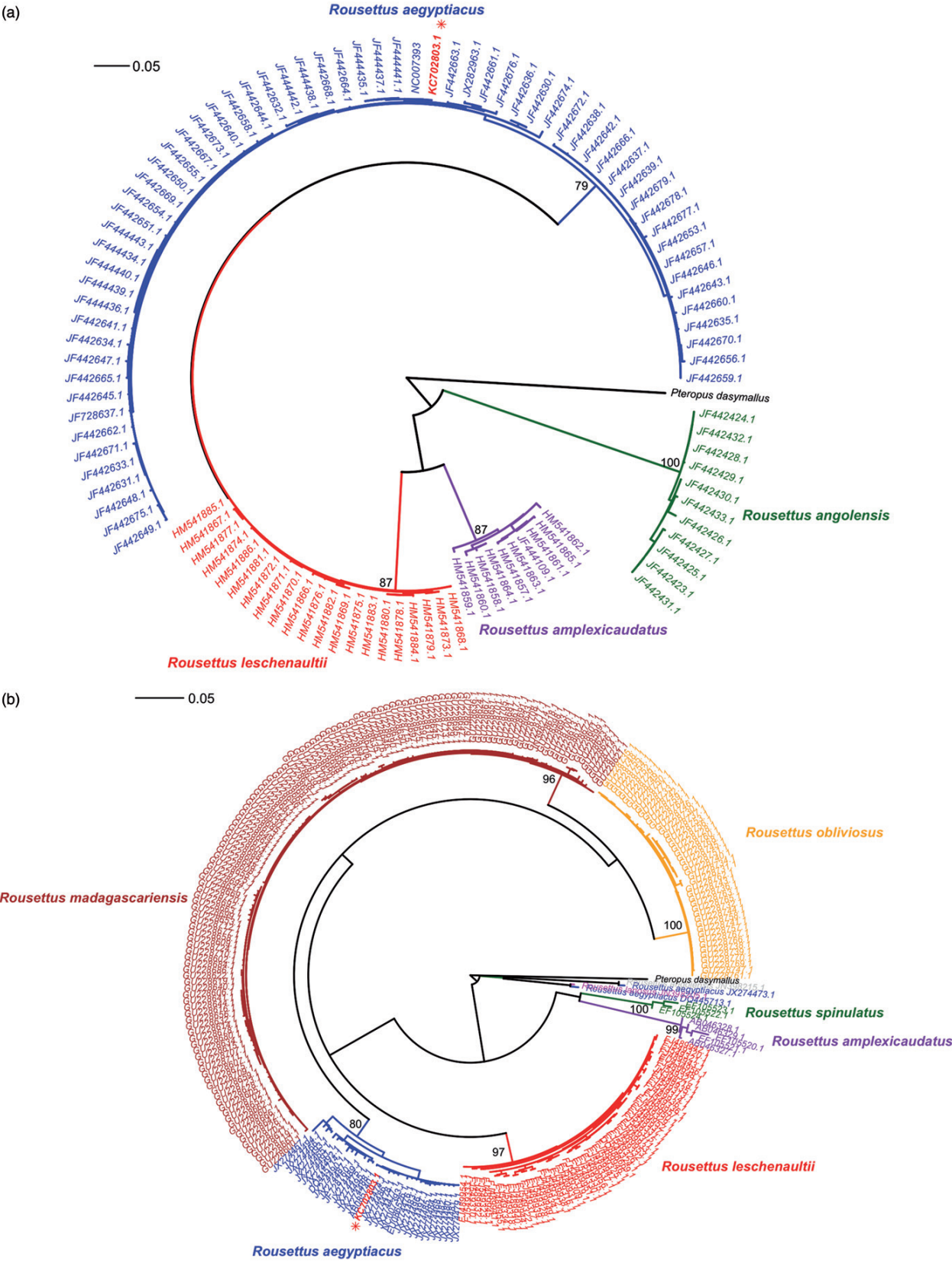


Figure 2. Maximum likelihood phylogenies inferred from mitochondrial *cox1* (a) and *cytb* (b) genes sequences available for the genus *Roussettus*. Asterisks show the position of the corresponding sequences extracted from the mitochondrial genome published by Szcześniak et al. (2013).



reported value of 0.3%. Actually, the divergence between these two taxa ranges from 6% to 10% for *cytb* and *cox1*, respectively.

Although our results identify the newly reported mitogenome as likely belonging to *R. aegyptiacus* rather than to *R. leschenaultii*, and the obtained trees do not show any obvious signs of introgression between the two species, we cannot exclude the possibility that the specimen sequenced by Szcześniak et al. (2013) actually corresponds to a hybrid or introgressed individual carrying the mitochondrial genome of *R. aegyptiacus*. In the absence of any information about the geographic origin of the specimen used for sequencing the mitogenome, it is impossible to evaluate such a hypothesis, which would need further nuclear information to be properly tested (Nesi et al., 2011). However, given that these two species have non-overlapping geographic ranges in Africa / West Asia (*R. aegyptiacus*) and South Asia (*R. leschenaultii*), the occurrence of hybrids appears unlikely. This example highlights the common difficulties encountered for identifying bats based on morphology alone. Among chiropterans, morphological convergence is a recurrent phenomenon (see for example Clare et al. (2007), Mayer & von Helversen (2001), and Murray et al. (2012)) and the order appears to be rich in cryptic species. This pleads for the generalized use of molecular data to obtain a more accurate depiction of taxonomy and biodiversity of bats.

### Proposed guidelines for accurately reporting new mitogenomes

In response to the increase in publication of new mitochondrial genomes, *Mitochondrial DNA* created a dedicated “Mitogenome Announcement” section. However, as illustrated by the case of Szcześniak et al. (2013), some quality controls are needed to validate the authenticity of the published mitogenomes. In line with the arguments exposed above, we would like to call for attention on the importance of both accurately identifying samples and verifying the authenticity of the final sequences before announcing new mitogenomes. We suggest that “Mitogenome Announcement” papers should include, at least, the following three mandatory quality controls.

- (1) *Provide detailed information on the origin of the sample used for mitogenome sequencing.* Ideally, the sample should be attached to a specimen voucher deposited in a recognized museum and accessible through multi-institution, multi-collection museum databases like ARCTOS (<http://arctos.database.museum>). Of note, the exact procedures and rules for registering new material to a holding institution may depend upon the country, institution, and curator policies. At the very end, minimal information on a sample should include the country of origin, a geographical locality of sampling with GPS coordinates if available, the sampling date, the name of the collector or the donating institution, and the source of the biological material used for DNA extraction (e.g. the catalog number assigned by the final institution to the corresponding specimen/voucher). The more accurate the information, the easier the taxonomic identification and any feedback on it.
- (2) *Conduct a phylogenetic analysis of the new mitogenome in the context of closely related species.* This will generate a phylogenetic tree that we strongly advise to be presented as a ML phylogram with node bootstrap support values, i.e. providing information on both the topology and its reliability, and the associated branch lengths (Figure 1). An important issue is the taxonomic sampling as the availability of mitogenome sequences greatly varies from one group to

another. We therefore suggest using the 20 phylogenetically closest taxa that should allow for a clear depiction of both the evolutionary affinities of the new mitogenome and the degree of divergence as compared to its closest relatives. Of course, this might be adjusted with respect to the available sequences for a given taxonomic level. For instance, if no congeneric species is available, sampling should ideally include members of the closest genera, and/or the closest families, and so on. For densely sampled groups, including between one and three representatives of each major lineages should provide an informative and easily readable tree.

- (3) *Provide a barcoding identification assessment of the sample thanks to a ML tree based on the closest available sequences.* Identifying a sample is facilitated by the wealth of information contained in public databases like BOLD, which includes 1,790,548 *cox1* sequences for more than 120,000 animal species as of August 2013. This barcoding database relies on a portion of the *cox1* gene, which has proven useful for identifying and describing new species in several groups of animals including bats (Clare et al., 2007, 2011). Similarly, the GenBank database of the National Center of Biotechnology Information (NCBI) includes a massive collection of both nuclear and mitochondrial markers. Furthermore, the strength of these databases relies on the detection of misidentified sequences (Shen et al., 2013), provided that sequences are available for the same marker for different individuals and populations of a given taxon. For bats, there are currently 18,671 public *cox1* sequences in BOLD and more than 7500 *cytb* sequences available in GenBank. A complementary use of these databases should allow trustfully identifying most samples. For instance, a similarity search against these databases coupled to a phylogenetic analysis would have shown that the specimen sequenced by Szcześniak et al. (2013) was likely misidentified. Of course, the barcode approach is also valuable for new mitogenomes for which there is already a reference deposited in public databases, from another individual of either the same population or from a different locality. In this particular case, a phylogenetic tree will cross-validate all sequences, and this accurate sample identification will be of capital importance when studying intra-specific or geographical genetic variation.

In this study, we illustrated the advantages of evaluating sequences in a phylogenetic context. This approach complements the already available tools for sequence comparison and provides a more accurate identification of the sequences. It can shed some light on the occurrence of contaminations or chimerical assemblies. It will also provide useful information about divergence of the new mitogenome with respect to other available individuals/taxa.

We hope that following these quality control guidelines will reinforce the quality standards of the papers published in *Mitochondrial DNA* and contribute to maintain and improve the reliability of the data stored in public databases.

### Acknowledgements

We would like to thank the Editor in Chief Rob DeSalle for helpful comments on an early version of the manuscript. Phylogenetic analyses benefited from the Montpellier Bioinformatics Biodiversity platform. This publication is contribution No. 2014-019 of the Institut des Sciences de l'Evolution de Montpellier (UMR 5554 – UM2 + CNRS + IRD).

### Declaration of interest

The authors declare no conflict of interest.

## References

- Arnason U, Gullberg A, Widegren B. (1991). The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. *J Mol Evol* 33:556–8.
- Baker RJ, Bradley RD. (2006). Speciation in mammals and the genetic species concept. *J Mammal* 87:643–62.
- Baker RJ, Solari S, Hoffmann F. (2002). A new Central American species from the *Carollia brevicauda* complex. Occasional Papers, Museum of Texas Tech University, 217:i, 1–12.
- Bidartondo MI and 255 co-authors. (2008). Preserving accuracy in GenBank. *Science* 319:1616.
- Botero-Castro F, Tilak MK, Justy F, Catzeflis F, Delsuc F, Douzery EJ. (2013). Next-generation sequencing and phylogenetic signal of complete mitochondrial genomes for resolving the evolutionary history of leaf-nosed bats (Phyllostomidae). *Mol Phylogenet Evol* 69:728–39.
- Capella-Gutiérrez S, Silla-Martínez J, Gabaldón T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–3.
- Clare EL, Lim BK, Engstrom MD, Eger JL, Hebert PDN. (2007). DNA barcoding of Neotropical bats: Species identification and discovery within Guyana. *Mol Ecol Notes* 7:184–90.
- Clare EL, Lim BK, Fenton MB, Hebert PDN. (2011). Neotropical bats: Estimating species diversity with DNA barcodes. *PLoS One* 6:e22648.
- D'Onorio De Meo P, D'Antonio M, Griggio F, Lupi R, Borsani M, Pavesi G, Castrignanò T, et al. (2012). MitoZoa 2.0: A database resource and search tools for comparative and evolutionary analyses of mitochondrial genomes in Metazoa. *Nucleic Acids Res* 40:D1168–72.
- Dettai A, Gallut C, Brouillet S, Pothier J, Lecointre G, Debruyne R. (2012). Conveniently pre-tagged and pre-packaged: Extended molecular identification and metagenomics using complete metazoan mitochondrial genomes. *PLoS One* 7:e51263.
- Edgar RC. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
- Enk J, Devault A, Debruyne R, King CE, Treangen T, O'Rourke D, Salzberg SL, et al. (2011). Complete Columbian mammoth mitogenome suggests interbreeding with woolly mammoths. *Genome Biol* 12: R51.
- Finstermeier K, Zinner D, Brameier M, Meyer M, Kreuz E, Hofreiter M, Roos C. (2013). A mitogenomic phylogeny of living primates. *PLoS One* 8:e69504.
- Guschanski K, Krause J, Sawyer S, Valente L, Bailey S, Finstermeier K, Sabin R, et al. (2013). Next-generation museomics disentangles one of the largest primate radiations. *Syst Biol* 62:539–54.
- Hassanin A, Bonillo C, Nguyen BX, Cruaud C. (2010). Comparisons between mitochondrial genomes of domestic goat (*Capra hircus*) reveal the presence of numts and multiple sequencing errors. *Mitochondrial DNA* 21:68–76.
- Hassanin A, Delsuc F, Ropiquet A, Hammer C, Jansen Van Vuuren B, Matthee C, Ruiz-Garcia M, et al. (2012). Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *CR Biologies* 335:32–50.
- Hebert PDN, Cywinska A, Ball SL, Dewaard JR. (2003). Biological identifications through DNA barcodes. *Proc R Soc Lond B* 270: 313–22.
- Hirata D, Mano T, Abramov A, Baryshnikov G, Kosintsev P, Vorobiev A, Raichev E, et al. (2013). Molecular phylogeography of the brown bear (*Ursus arctos*) in Northeastern Asia based on analyses of complete mitochondrial DNA sequences. *Mol Biol Evol* 30:1644–52.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, et al. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–9.
- Krause J, Unger T, NoçOn A, Malaspina A-S, Kolokotronis S-O, Stiller M, Soibelzon L, et al. (2008). Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evol Biol* 8:220.
- Mayer F, Von Helversen O. (2001). Cryptic diversity in European bats. *Proc R Soc Lond Ser B, Biol Sci* 268:1825–32.
- Miya M, Friedman M, Satoh TP, Takeshima H, Sado T, Iwasaki W, Yamanoue Y, et al. (2013). Evolutionary origin of the Scombridae (tunas and mackerels): Members of a paleogene adaptive radiation with 14 other pelagic fish families. *PLoS One* 8:e73535.
- Murray SW, Campbell P, Kingston T, Zubaid A, Francis CM, Kunz TH. (2012). Molecular phylogeny of hipposiderid bats from Southeast Asia and evidence of cryptic diversity. *Mol Phylogenet Evol* 62:597–611.
- Nabholz B, Glemis N, Galtier N. (2008). Strong variations of mitochondrial mutation rate across mammals – the longevity hypothesis. *Mol Biol Evol* 25:120–30.
- Nelson L, Lambkin C, Batterham P, Wallman J, Dowton M, Whiting M, Yeates D, Cameron S. (2012). Beyond barcoding: A mitochondrial genomics approach to molecular phylogenetics and diagnostics of blowflies (Diptera: Calliphoridae). *Gene* 511:131–42.
- Nesi N, Nakouné E, Cruaud C, Hassanin A. (2011). DNA barcoding of African fruit bats (Mammalia, Pteropodidae). The mitochondrial genome does not provide a reliable discrimination between *Epomophorus gambianus* and *Micropteropus pusillus*. *C R Biologies* 334:544–54.
- Parma P, Feligini M, Greppi G, Enne G. (2003). The complete nucleotide sequence of goat (*Capra hircus*) mitochondrial genome. *Mitochondrial DNA* 14:199–203.
- Rohland N, Malaspina AS, Pollack JL, Slatkin M, Matheus P, Hofreiter M. (2007). Proboscidean mitogenomics: Chronology and mode of elephant evolution using mastodon as outgroup. *PLoS Biol* 5:e207.
- Shen Y-Y, Chen X, Murphy RW. (2013). Assessing DNA barcoding as a tool for species identification and data quality control. *PLoS ONE* 8: e57125.
- Solari S, Baker RJ. (2006). Mitochondrial DNA sequence, karyotypic, and morphological variation in the *Carollia castanea* species complex (Chiroptera: Phyllostomidae) with description of a new species. Occasional Papers, Museum of Texas Tech University, 254:i, 1–16.
- Stamatakis A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–90.
- Szczeniak M, Yoneda M, Sato H, Makiłowska I, Kyuwa S, Sugano S, Suzuki Y, et al. (2013). Characterization of the mitochondrial genome of *Rousettus leschenaulti*. *Mitochondrial DNA*. [Epub ahead of print]. doi: 10.3109/19401736.2013.809451.
- Vilstrup JT, Ho SY, Foote AD, Morin PA, Krieb D, Krützen M, Parra GJ, et al. (2011). Mitogenomic phylogenetic analyses of the Delphinidae with an emphasis on the Globicephalinae. *BMC Evol Biol* 11:65.
- Wada K, Okumura K, Nishibori M, Kikkawa Y, Yokohama M. (2010). The complete mitochondrial genome of the domestic red deer (*Cervus elaphus*) of New Zealand and its phylogenetic position within the family Cervidae. *Anim Sci J* 81:551–7.
- Yu L, Peng D, Liu J, Luan P, Liang L, Lee H, Lee M, et al. (2011a). On the phylogeny of Mustelidae subfamilies: Analysis of seventeen nuclear non-coding loci and mitochondrial complete genomes. *BMC Evol Biol* 11:92.
- Zhang P, Liang D, Mao R, Hillis D, Wake D, Cannatella D. (2013). Efficient sequencing of Anuran mtDNAs and a mitogenomic exploration of the phylogeny and evolution of frogs. *Mol Biol Evol* 30: 1899–915.

Supplementary material available online

## Supplementary Tables