C. R. Biologies 334 (2011) 544–554

Contents lists available at ScienceDirect

Comptes Rendus Biologies

www.sciencedirect.com

Evolution/Évolution

DNA barcoding of African fruit bats (Mammalia, Pteropodidae). The mitochondrial genome does not provide a reliable discrimination between *Epomophorus gambianus* and *Micropteropus pusillus*

« Codes-barres ADN » des chauves-souris frugivores africaines (Mammalia, Pteropodidae). Le génome mitochondrial ne permet pas une discrimination fiable entre Epomophorus gambianus *et* Micropteropus pusillus

Nicolas Nesi [a,b], Emmanuel Nakouné [c], Corinne Cruaud [d], Alexandre Hassanin [a,b,*]

[a] *UMR 7205, origine, structure et évolution de la biodiversité, département systématique et évolution, Muséum national d'histoire naturelle, case postale n° 51, 55, rue Buffon, 75005 Paris, France*
[b] *Service de systématique moléculaire, département systématique et évolution, Muséum national d'histoire naturelle, 43, rue Cuvier, 75005 Paris, France*
[c] *Laboratoire des arbovirus, virus des fièvres hémorragiques, virus de la grippe et rage, institut Pasteur de Bangui, BP 923 Bangui, Central African Republic*
[d] *Genoscope, Centre national de séquençage, 2, rue Gaston-Crémieux, CP5706, 91057 Evry cedex, France*

ARTICLE INFO

Article history:
Received 14 January 2011
Accepted after revision 18 May 2011
Available online 24 June 2011

Keywords:
Taxonomy
Phylogeny
Megachiroptera
Mitochondrial genome
Nuclear DNA
Topological incongruence
Interspecific hybridization
Ancestral polymorphism

ABSTRACT

Sequences of the mitochondrial cytochrome c oxidase subunit I (*COI*) gene have been shown to be useful for species identification in various groups of animals. However, the DNA barcoding approach has never been tested on African fruit bats of the family Pteropodidae (Mammalia, Chiroptera). In this study, the *COI* gene was sequenced from 120 bats collected in the Central African Republic and belonging to either *Epomophorus gambianus* or *Micropteropus pusillus*, two species easily diagnosed on the basis of morphological characters, such as body size, skull shape and palatal ridges. Two additional molecular markers were used for comparisons: the complete mitochondrial cytochrome *b* gene and the intron 7 of the nuclear β-fibrinogen (*FGB*) gene. Our results reveal an unexpected discordance between mitochondrial and nuclear genes. The nuclear *FGB* signal agrees with our morphological identifications, as the three alleles detected for *E. gambianus* are divergent from the fourteen alleles found for *M. pusillus*. By contrast, this taxonomic distinction is not recovered with the analyses of mitochondrial genes, which support rather a polyphyletic pattern for both species. The conflict between molecular markers is explained by multiple mtDNA introgression events from *M. pusillus* into *E. gambianus* or, alternatively, by incomplete lineage sorting of mtDNA haplotypes associated with positive selection on *FGB* alleles of *M. pusillus*. Our work shows the failure of DNA barcoding to discriminate between two morphologically distinct fruit bat species and highlights the importance of using both mitochondrial and nuclear markers for taxonomic identification.

Mots clés :
Taxinomie
Phylogénie

RÉSUMÉ

Les séquences du gène mitochondrial de la première sous-unité de la cytochrome c oxydase (COI) sont de plus en plus utilisées comme « codes-barres ADN » pour identifier les espèces animales. Toutefois, cette approche n'a jamais été testée sur les chauves-souris

* Corresponding author.
E-mail address: hassanin@mnhn.fr (A. Hassanin).

1631-0691/$ – see front matter
doi:10.1016/j.crvi.2011.05.003

Megachiroptera;
Génome mitochondrial
ADN nucléaire
Incohérence topologique
Hybridation interspécifique
Polymorphisme ancestral.

frugivores d'Afrique. Lors de cette étude, le gène *COI* a été séquencé à partir de 120 chauves-souris collectées en République centrafricaine et appartenant à *Epomophorus gambianus* et *Micropteropus pusillus*, deux espèces faciles à différencier sur la base de caractères morphologiques, tels que la taille corporelle, la forme du crâne ou les plis palataux. Deux autres marqueurs moléculaires ont été utilisés : le gène mitochondrial du cytochrome *b* et l'intron 7 du gène nucléaire de la chaîne bêta du fibrinogène (*FGB*). Nos résultats révèlent une discordance inattendue entre les gènes mitochondriaux et le gène nucléaire. Les données nucléaires confirment les identifications morphologiques puisque les trois allèles détectés chez *E. gambianus* sont divergents des quatorze allèles découverts chez *M. pusillus*. En revanche, les analyses reposant sur les gènes mitochondriaux ne soutiennent pas cette distinction taxonomique, puisque les deux espèces apparaissent polyphylétiques. Un tel conflit peut être expliqué par de multiple évènements d'introgression du génome mitochondrial de *M. pusillus* vers *E. gambianus*, ou alternativement par la persistance d'haplotypes mitochondriaux ancestraux chez les deux espèces, associée à une sélection positive sur les allèles du gène *FGB* chez *M. pusillus*. Notre cas d'étude montre que l'approche des « codes-barres ADN » ne permet pas de distinguer deux espèces de chauves-souris morphologiquement très différentes. Ainsi, nous préconisons de réaliser les identifications moléculaires à l'aide d'une approche combinant le gène mitochondrial *COI* avec un ou plusieurs marqueurs nucléaires.

1. Introduction

The family Pteropodidae (Old World fruit bats) represents one of the most diversified taxa of the order Chiroptera with 189 species placed in 43 genera [1–3]. Its geographic distribution covers most tropical and subtropical zones of the Old World, from Africa (38 species) to Australia and Pacific islands (68 species), through India (13 species), East Asia (14 species) and Southeast Asia (84 species).

Many African fruit bat species have been recognised as potential reservoirs for viruses, including *Eidolon helvum*, *Epomophorus gambianus*, *Epomophorus wahlbergi*, *Epomops buettikoferi, Micropteropus pusillus* and *Rousettus aegyptiacus* for the Lagos bat virus [4,5], *Eidolon helvum*, *Epomops franqueti, Hypsignathus monstrosus* and *Myonycteris torquata* for the Ebola virus [6,7] and *Rousettus aegyptiacus* for the Marburg virus [8]. Whereas Lagos bat virus infection affects only cats and dogs, Ebola and Marburg viruses cause hemorrhagic fever, a severe, often-fatal disease in both human and great ape populations. Fruit bats appear to play a major role in disseminating zoonoses. To assess future threats posed by bat-associated zoonoses to humans, there is a need for accurate knowledge of the factors underlying the emergence of disease, as well as the specific relationships between the host and virus. One fundamental step for these studies is correct and accurate identification of the suspected reservoir species.

Species have traditionally been identified using a taxonomic key based on morphological characters. In the case of African fruit bats, genera can be distinguished by their skull shape, dental formula and unique palatal ridge pattern [9]. At the specific level, only a few characters can be used and most of them consist of adult body and skull dimensions. This lack of information makes identification tricky, notably when the specimens are juvenile and subadult, or when body measurements overlap greatly between closely related species. To deal with these difficulties, the DNA Barcode project was conceived as a standard system for identifying animal species. DNA barcoding is based on the subunit I of the cytochrome c oxidase (*COI*) gene from the mitochondrial DNA (mtDNA) genome, which acts as a 'barcode' to identify animal species [10]. Although the efficiency of DNA Barcoding has gained increasing validation, *COI* has not been used extensively on mammals and it has never been applied to African fruit bats.

In this study, the DNA Barcoding approach was tested on African fruit bats collected in the Central African Republic (CAR) in 2008. Our first *COI* data revealed very low genetic distance (0.14%) between two specimens belonging to two different species, *Epomophorus gambianus* Ogilby, 1835 and *Micropteropus pusillus* Peters, 1867. This result was highly unexpected, given that they are two morphologically distinct species. *E. gambianus* is a large fruit bat (mean weight of 105.5 g), whereas *M. pusillus* is a small-sized bat (mean weight of 27.5 g) (Fig. 1A). The skull of *E. gambianus* is long and narrow, whereas that of *M. pusillus* is short and broad (Fig. 1B). Finally, their six palatal ridges show distinct patterns: in *E. gambianus*, all are undivided, except for the last two postdental ridges, which are notched or narrowly divided, whereas, in *M. pusillus*, all are divided by a deep, continuous median groove, except for the first ridge, which is prominent and undivided (Fig. 1B).

M. pusillus occurs in West and Central Africa (Fig. 2) in a variety of habitats, like savannah woodlands and vegetation belts on the edge of the forest [11]. *E. gambianus* is also a savannah and woodland species. It is sympatric with *M. pusillus* from Senegal to southern Ethiopia, but is also found in South-East Africa (Fig. 2).

Here we conduct a specific study on *E. gambianus* and *M. pusillus,* to understand better why *COI* sequences of these two species show such a low genetic distance. A total of 120 specimens were sequenced for two mitochondrial genes, i.e., *COI* (685 nt) and the complete cytochrome *b* gene (*Cytb*; 1140 nt), and the nuclear β-fibrinogen intron 7 (*FGB*; 700 nt). Our analyses reveal that mitochondrial markers do not provide a reliable signal for identifying *E. gambianus* and *M. pusillus*, whereas the nuclear marker supports the taxonomic distinction. Two hypotheses are

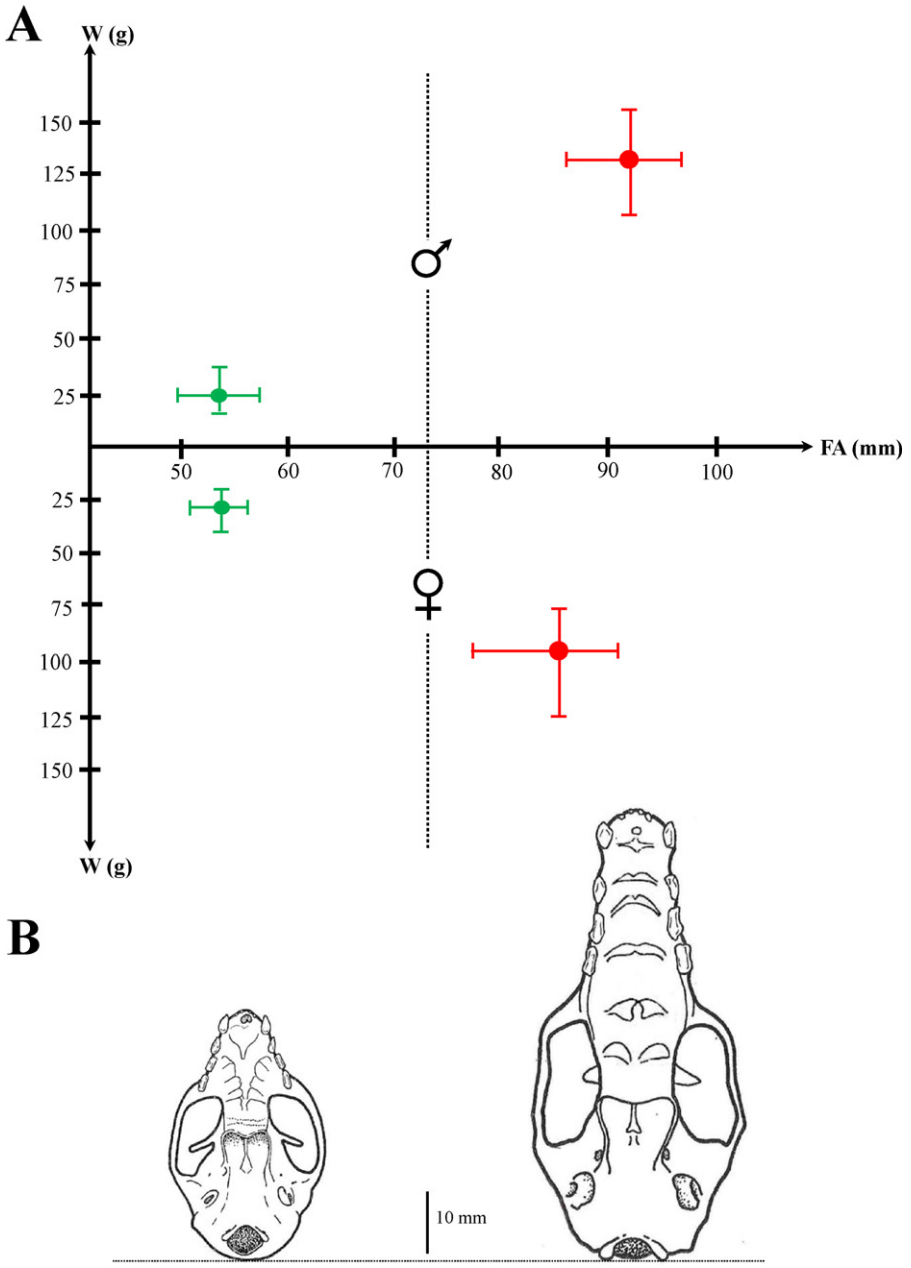

Fig. 1. Morphological differences between *Micropteropus pusillus* (on the left) and *Epomophorus gambianus* (on the right). A. Forearm length (FA, mm) and body weight (W, g) of males (above the abscissa) and females (below the abscissa). Data were extracted from Bergmans [39,47]. B. Skull, ventral view (modified after [54,55]).

retained to explain the misleading signal of mitochondrial genes: mtDNA introgression and incomplete lineage sorting of mtDNA haplotypes.

2. Materials and methods

2.1. Taxonomic sample

All samples were collected in CAR in 2008. Fruit bats were determined using the species identification key published by Bergmans [9]. A total of 23 individuals of *E. gambianus* from four localities and 97 individuals of *M. pusillus* from five geographic localities were included in this study. The localities are depicted in Fig. 2 and detailed in Table S1 (see Supplementary Material). In agreement with previous phylogenetic results [12], three outgroup species were used to root the tree: *Epomops franqueti*, which is another species of the tribe Epomophorini; *Myonycteris torquata*, which is a member of the tribe Myonycterini; and *Rousettus aegyptiacus*, which is the most distant outgroup, as it belongs to Rousettinae, a different subfamily from Epomophorinae.

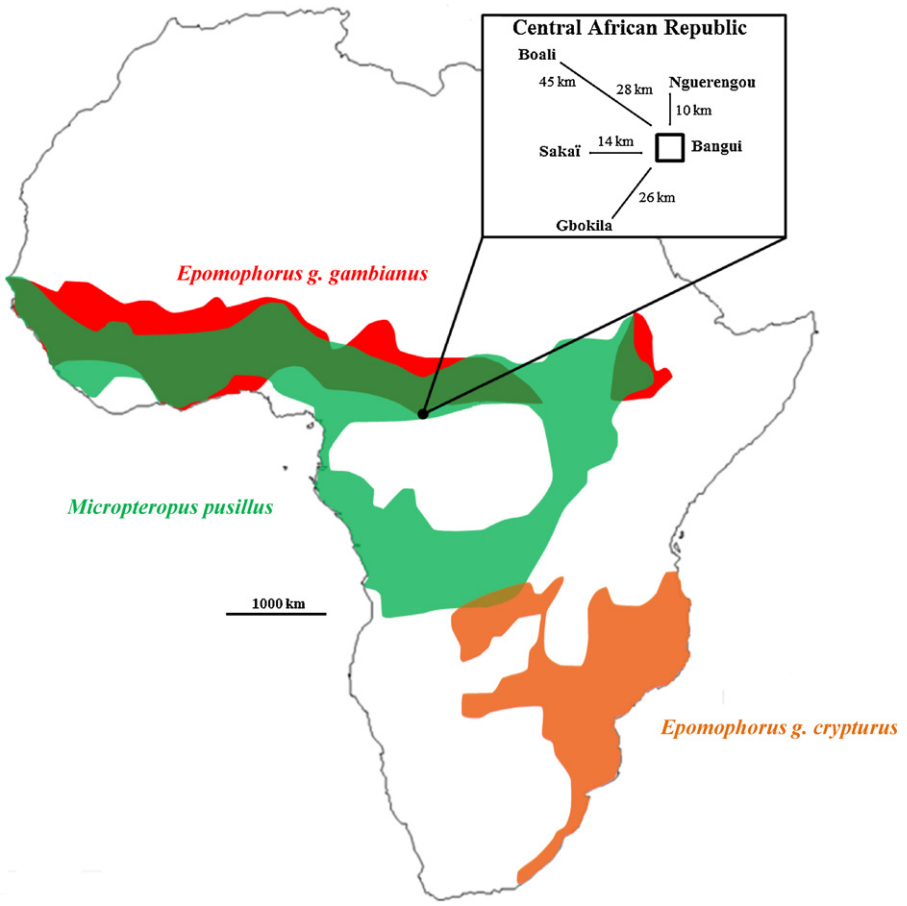

Fig. 2. Distribution map of *Epomophorus gambianus* (red) and *Micropteropus pusillus* (green) (adapted from Bergmans [39,47] and [2]). The five localities sampled for this study are detailed in the box. Two subspecies have been described for *E. gambianus*: *E. g. gambianus* in the two northern areas (red) and *E. g. crypturus* in South-East Africa (orange) [39]. The latter subspecies has been elevated to species rank, i.e. *Epomophorus crypturus*, in several recent classifications [1,2].

2.2. DNA extraction, amplification and sequencing

Total genomic DNA was extracted from muscle or patagium samples using QIAGEN DNeasy Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol.

Three genes were sequenced for this study: two belonging to the mitochondrial genome (*COI* and *Cytb*) and one from the nuclear genome (a non-coding fragment: *FGB*). Primers used for polymerase chain reaction (PCR) were: UTyrLA and C1L705 for *COI*, UGluMA and LThrCH for *Cytb*, and 5′-CCA-CAA-CRG-CAT-GTT-CTT-CAG-CAC-3′ and 5′-GTA-TCT-GCC-ATT-TGG-ATT-GGC-TGC-3′ for *FGB* [13,14]. Amplifications were done in 20 μl using 2 μl of Buffer 10X with $MgCl_2$, 0.8 μl of dNTP (6.6 mM), 0.12 μl of Taq DNA polymerase (2.5 U, Qiagen, Hilden, Germany) and 0.32 μl of the two primers at 10 μM. The standard PCR conditions were as follows: 4 min at 94 °C; 35 cycles of denaturation/annealing/extension with 30 s at 94 °C for denaturation, 45 s at 50 °C for annealing and 90 s at 72 °C for extension; and 10 min at 72 °C.

PCR products were purified using ExoSAP Kit (GE Healthcare, Buckinghamshire, UK) and then sequenced in both directions using an automated DNA Sequencer (Applied Biosystems 3100). These two last steps were performed by GENOSCOPE (Ivry-sur-Seine, France). Nuclear fragments that contained multiple heterozygous single-nucleotide polymorphisms (characterized by double peaks in the chromatogram) were cloned using a QIAGEN PCR Cloning^{plus} Kit in order to isolate the exact sequence of each allele in heterozygous individuals. Sequences were edited and assembled using Sequencher 4.7 (Gene Codes Corporation). Sequences generated for this study were deposited in the EMBL/DDBJ/Genbank database (accession numbers JF728368-JF728760; Table S1 for details).

2.3. Sequence data analysis

Sequences were aligned by eye using BioEdit version 7.0.9 [15]. No gaps were included in the alignments of *COI* and *Cytb*. For *FGB*, two unambiguous gaps were inferred to align the most divergent outgroup species, *R. aegyptiacus*, with other taxa.

Phylogenetic analyses were performed using Neighbor Joining (NJ) and Bayesian methods. The three genes (*COI*: 685 nt; *Cytb*: 1140 nt; *FGB*: 700 nt) were analyzed

separately to evaluate their own signal and to detect any incongruence. The mitochondrial markers were then combined (*COI* + *Cytb*). MODELTEST version 3.7 [16] was run to determine the best-fitting evolutionary model. According to AIC criterion, the selected model was HKY + G for *COI* and *Cytb*, TVM + G for *COI* + *Cytb* and HKY for *FGB*. PAUP* 4b10 [17] was used for the NJ analyses with the selected model. Nodal support was assessed by the bootstrap analysis (1000 replicates). Bayesian inferences were conducted with MrBayes v3.1.2 [18]. Posterior probabilities (PP) were calculated using four independent Markov Chains run for 10,000,000 Metropolis-coupled MCMC generations with tree sampling every 1000 generations. TRACER [19] was used to calculate the effective sample size statistic to make sure that a sufficient number of generations were run and to decide on the length of the burn-in period. Analyses were run twice independently to check for convergence of the results.

Heterozygosity of nuclear genes complicates phylogenetic analyses because the two alleles from a heterozygous individual are not necessarily most closely related to each other. In practice, for many phylogenies based on diploid gene sequences, allelic sequences are not separated and heterozygous sites are treated as polymorphic. This procedure can make phylogenetic inference ambiguous, especially for closely related taxa [20]. However, the pairing of different alleles can provide information on interbreeding and may be useful in delimitating species. Therefore, the nuclear sequences of *FGB* were analyzed using the median-joining algorithm available in Network version 4.5.1 [21].

The pairwise genetic distances were calculated with PAUP* 4b10 [17] using the Kimura 2-parameter (K2P) correction. For each species, we calculated the number of polymorphic sites, number of haplotypes, nucleotide diversity (π) and haplotype diversity (h) [22,23] using DnaSP version 5.10.01 [24].

An "Isolation with Migration" (IM) model was used to test introgression rate between the two fruit bat species. We used a Bayesian MCMC method in the IMa version 2.0 coalescent program [25], which attempts to fit the data to a null model where an ancestral population bifurcates into two allopatric populations 1 and 2. This MCMC simulation estimates the joint posterior probability of six demographic parameters: population size of species 1 and 2 (N_{e1}, N_{e2}) and of the ancestral species (N_{eA}), migration (m_1, m_2) and divergence time (t). In our study, subscript 1 refers to the *M. pusillus* species and subscript 2 refers to the *E. gambianus* species. Although migration typically refers to movements of individuals between populations, the m parameter is better interpreted as the rate of mtDNA introgression for the purposes of this study. In this way, the m_1 parameter refers to introgression from *E. gambianus* into *M. pusillus*, whereas the m_2 parameter refers to introgression from *M. pusillus* into *E. gambianus*. As recommended in Hey & Nielsen [26], the inheritance scalars were set to 1 for the nuclear locus (*FGB*) and 0.25 for the mitochondrial sequences (*COI* + *Cytb*) and the HKY model was used. To assess convergence, we checked effective sample sizes throughout the run and compared results between three independent runs. Following the approach used by Won and Hey [27], we first ran IMa using large, flat priors for each parameter. Based on those runs, we defined narrower upper bounds that encompassed the full posterior distribution of each parameter. In the final runs, priors were set to 5 for t, 2 for m, 1000 for N_{eA} and 100 for N_{e1} and N_{e2}. The input files were executed in IMa using 10 Metropolis-coupled chains with 100,000 steps for burn-in followed by 10 million steps for parameter estimation.

3. Results

3.1. Analyses of mitochondrial genes

The 5′ region of *COI* (685 nt) and the complete *Cytb* gene (1140 nt) were sequenced for all samples available for *E. gambianus* (23 individuals) and *M. pusillus* (97 individuals). Mitochondrial genes showed high nucleotide and haplotype diversities for both species (Table 1), suggesting large effective size and stable populations (i.e., no recent bottleneck). Among *COI* sequences, a total of 60 haplotypes were identified: seven specific to *E. gambianus*, 49 specific to *M. pusillus* and four shared between the two species (Table S1). Among *Cytb* sequences, thirteen haplotypes were found for *E. gambianus*, 71 for *M. pusillus* and none were shared between the two taxa (Table S1). Using the alignment combining *COI* and *Cytb* sequences, we identified thirteen haplotypes for *E. gambianus*, 83 for *M. pusillus* and none were shared between the two taxa (Table S1).

The Bayesian tree reconstructed from the alignment combining *COI* and *Cytb* genes (123 taxa, 1825 nt) is shown in Fig. 3. All individuals of *E. gambianus* and *M. pusillus* fall

Table 1
Genetic variability of molecular markers sequenced for *Epomophorus gambianus* and *Micropteropus pusillus*.

Locus	Species	N	S	Hap	$h \pm$ SD	$\pi \pm$ SD	Het
Cytb	*M. pusillus*	97	134	71	0.984 ± 0.007	0.00859 ± 0.00057	–
	E. gambianus	23	48	13	0.933 ± 0.001	0.00987 ± 0.00062	–
COI	*M. pusillus*	97	70	53	0.956 ± 0.014	0.00935 ± 0.00068	–
	E. gambianus	23	24	11	0.909 ± 0.034	0.01042 ± 0.0007	–
COI + Cytb	*M. pusillus*	97	204	83	0.995 ± 0.003	0.00888 ± 0.00056	–
	E. gambianus	23	72	13	0.933 ± 0.033	0.01008 ± 0.00056	–
FGB	*M. pusillus*	97	16	14	0.333 ± 0.058	0.00118 ± 0.00031	17
	E. gambianus	23	2	3	0.360 ± 0.010	0.00055 ± 0.00017	6

N: number of specimens; S: number of polymorphic sites; Hap: number of haplotypes; h: haplotype diversity; π: nucleotide diversity; SD: standard deviation; Het: number of heterozygotes.

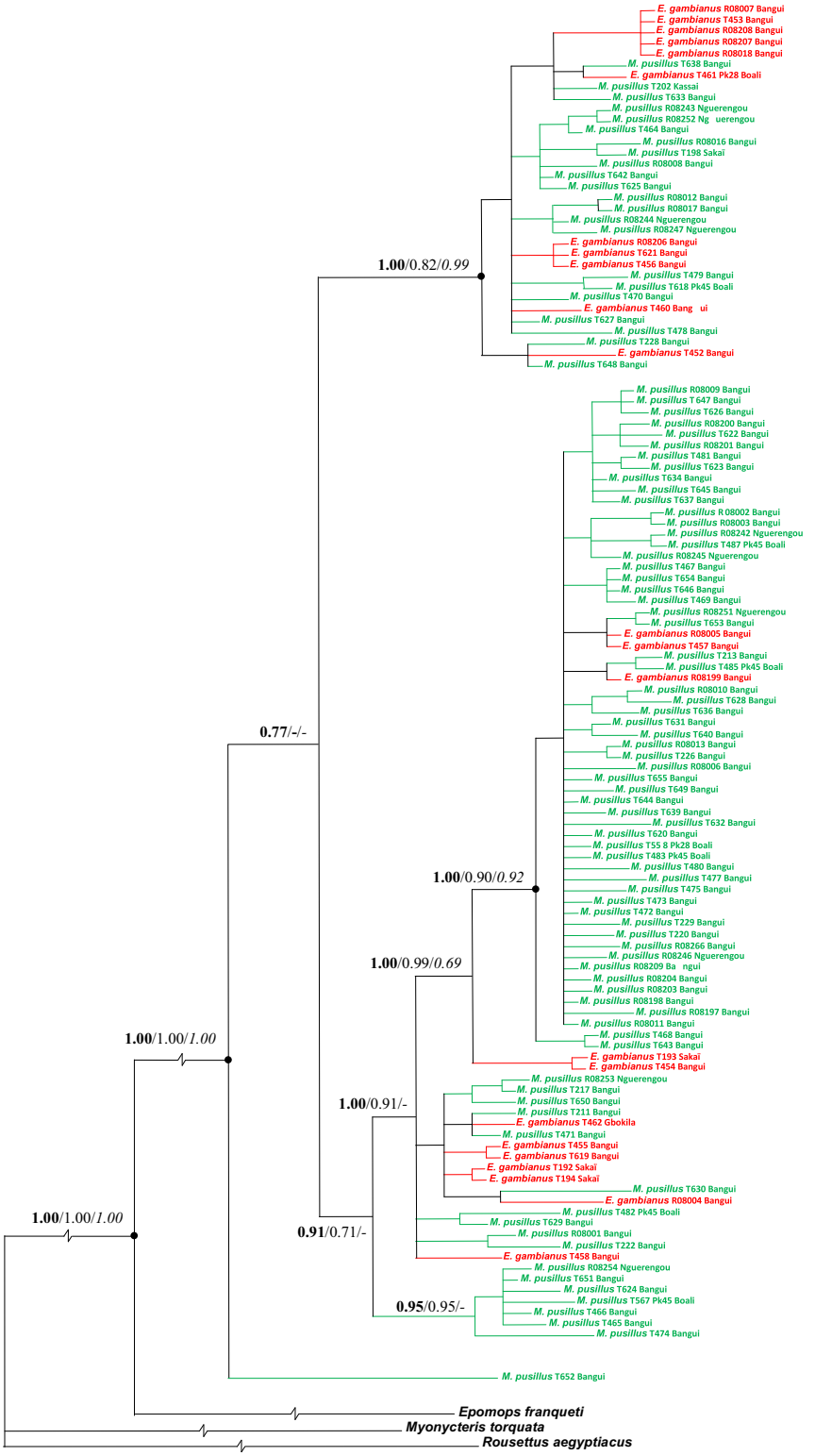

Fig. 3. Bayesian tree obtained from the analysis of mitochondrial genes. The values on the branches indicate posterior probabilities (> 0.75) calculated from left to right with three alignments: the combination of *COI* and *Cytb* genes, *Cytb* and *COI*.

into a robust clade (PP = 1), but neither species was found to be monophyletic. The trees obtained from separate analyses of *COI* and *Cytb* genes (Figs. S3 and S1) show very similar relationships, but the *COI* tree is less resolved than the *Cytb* tree, because *COI* has much less phylogenetic information content (165 and 89 parsimony informative sites for *Cytb* and *COI*, respectively). Moreover, the analysis including one complete *Cytb* sequence of *Epomophorus wahlbergi* (accession numbers DQ445706) revealed that the genus *Epomophorus* is paraphyletic, as *E. wahlbergi* occupies a more basal position with respect to the clade uniting *E. gambianus* and *M. pusillus* sequences (Fig. S2).

In Fig. 4, the comparisons between pairwise *COI* genetic distances show that DNA barcodes cannot be used for discriminating between *E. gambianus* and *M. pusillus*. Indeed, interspecific distances, which are comprised between 0% and 2.5%, are not higher than intraspecific distances found for *E. gambianus* (0–2%) and *M. pusillus* (0–2.5%). As a consequence, a query sequence cannot be correctly assigned to either *E. gambianus* or *M. pusillus*.

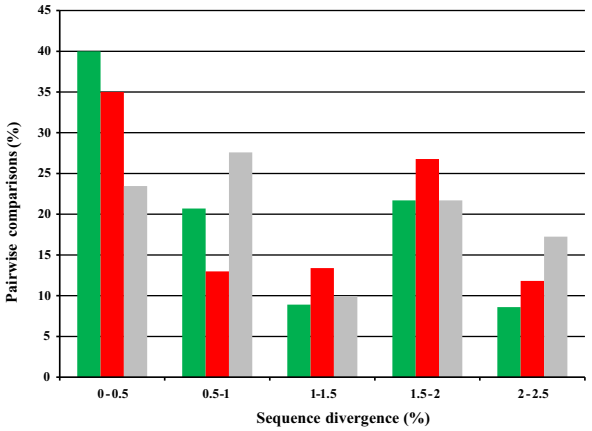

Fig. 4. Intra- and interspecific *COI* distances. Pairwise distances were calculated with the Kimura 2-parameter correction for comparisons within *Micropteropus pusillus* (green histograms), within *Epomophorus gambianus* (red histograms) and for interspecific comparisons (grey histograms).

3.2. Analyses of nuclear alleles

The nuclear *FGB* gene (700 nt) was also sequenced for all samples available for *E. gambianus* and *M. Pusillus*. The nuclear gene showed low nucleotide and haplotype diversities for both species (Table 1). We found three alleles for *E. gambianus* and fourteen alleles for *M. pusillus* (Table S1). None of them was shared between the two species. We identified six heterozygous individuals for *E. gambianus* (26%) and seventeen for *M. pusillus* (17.5%) (Table 1). In contrast with mtDNA data, the comparisons of *FGB* sequences indicate a higher haplotype diversity for *E. gambianus* and a higher nucleotide diversity for *M. pusillus* (Table 1).

The allelic network of *FGB* is presented in Fig. 5. The three alleles of *E. gambianus* are closely related and separated by only one mutation. The fourteen alleles of *M. pusillus* can be divided into two linked groups: the first one forms a star-like topology with a main central allele (Mp1), from which ten alleles are derived by one mutational step (Mp2 to Mp9; Mp12; Mp13) and two alleles by two mutational steps (Mp10 and Mp11); the

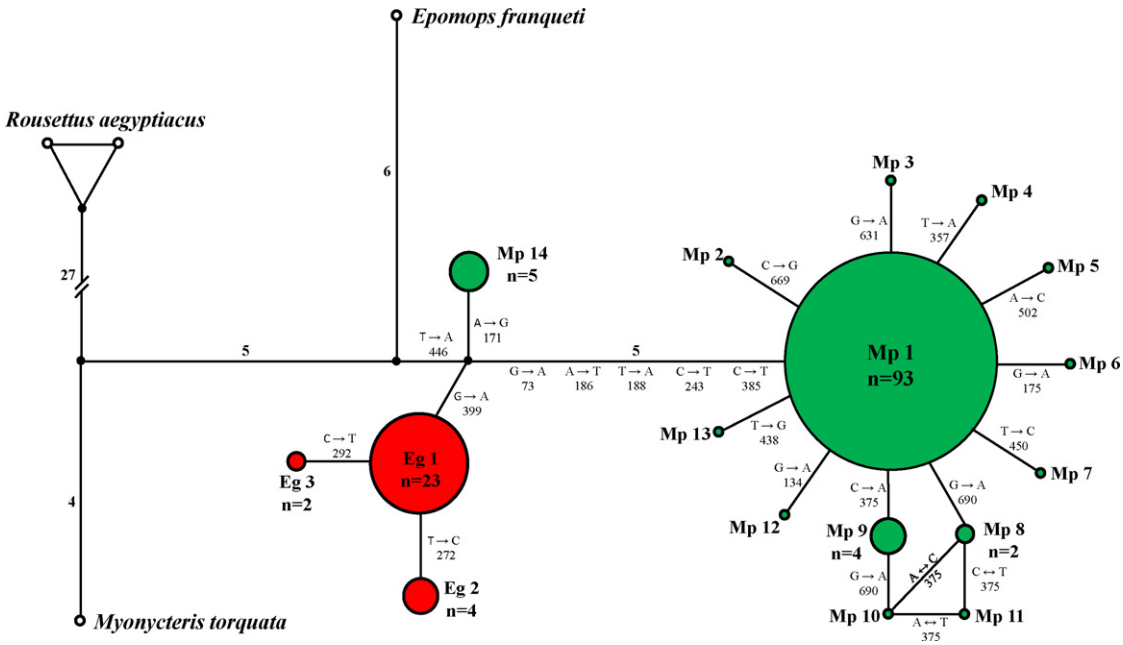

Fig. 5. Median-joining network of *FGB* alleles. The three alleles found for *Epomophorus gambianus* are indicated by red circles. The fourteen alleles found for *Micropteropus pusillus* are indicated by green circles. The circle size is proportional to the number of alleles found in the populations with the exception of alleles Mp1 and Eg1. The number of substitutions (greater than one) between alleles are indicated on the branches. We detailed the nature and position of all substitutions separating the alleles of *E. gambianus* and *M. pusillus*.

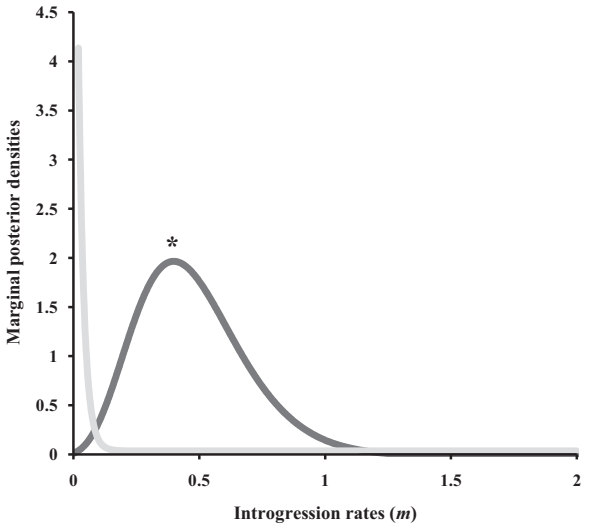

Fig. 6. The marginal posterior probability distributions of the introgression parameter, as estimated from the isolation with migration (IMa) program. The grey curve corresponds to the hypothesis of mtDNA introgression from *E. gambianus* into *M. pusillus*. The black curve corresponds to the hypothesis of mtDNA introgression from *M. pusillus* into *E. gambianus*. Statistical significance as assessed by the likelihood ratio test is indicated by * (< 0.001).

second group corresponds to allele Mp14. Interestingly, the allele Mp14 is found in only five heterozygous individuals of *M. pusillus* and it is always associated with the most frequent allele, Mp1. In addition, Mp14 is six mutational steps distant from other alleles of *M. pusillus*, but it differs by only two mutations from Eg1, the most common allele of *E. gambianus*.

3.3. Introgression rates

The introgression (or migration) parameter curves obtained from the isolation with migration program (IMa version 2.0) are presented in Fig. 6. In the case of introgression of *E. gambianus* into *M. pusillus*, the curve has a peak near zero and the probability of the peak is very near to what the probability is at zero. In this case, we cannot reject an introgression rate of zero. However, in the opposite direction, i.e., introgression from *M. pusillus* into *E. gambianus*, the peak is far from zero and the estimated probability of zero introgression is zero; we can therefore reject a model of no introgression from *M. pusillus* into *E. gambianus*.

4. Discussion

4.1. Discordance between mitochondrial and nuclear data sets

According to Bergmans [9], *Epomophorus* and *Micropteropus* can be ranged in the tribe Epomophorini with the three other genera *Epomops*, *Hypsignathus* and *Nanonycteris*. By analyzing mtDNA sequences of *Cytb* and 16S rRNA genes, Juste et al. [12] have concluded on a sister-group relationship between *Epomophorus wahlbergi* and *Micropteropus pusillus*, whereas *Epomops franqueti* was found to be more divergent. Other species of Epomophorini were however not included in the molecular analyses. Although our mtDNA results confirm that *Epomophorus* and *Micropteropus* are closely-related genera, it was highly unexpected to find a paraphyletic pattern for *Epomophorus* (Fig. S2) and a polyphyletic pattern for both *E. gambianus* and *M. pusillus* species (Fig. 3). Indeed, *M. pusillus* can be easily distinguished from *E. gambianus* and *E. wahlbergi* on the basis of morphological characteristics, such as body size (Fig. 1; weight: 20–35 g *versus* 56–155 g for *E. gambianus* and 54–125 g for *E. wahlbergi*; forearm length: 48–56 mm *versus* 75–100 mm for *E. gambianus* and 68–95 mm for *E. wahlbergi*) and the pattern of palatal ridges (Fig. 1; only five bridges for *E. wahlbergi*, see in Bergmans [9]). Since our analyses of the nuclear gene *FGB* agree with morphology, as both *M. pusillus* and *E. gambianus* possess distinct alleles (Fig. 5), we can conclude that mtDNA data are problematic. Three hypotheses can be advanced to explain the incongruence of mtDNA with nuclear and morphological data: the amplification of Numts, mtDNA introgression and incomplete lineage sorting of mtDNA haplotypes.

4.2. Can Numts explain the mitonuclear discordance?

Nuclear mitochondrial pseudogenes or 'Numts' are segments of mtDNA translocated to the nuclear genome. These paralogous sequences are commonly found in animal genomes and can be accidentally amplified instead of the targeted mitochondrial genes [28]. The undetected presence of Numts in the analyses can result in erroneous interpretations of phylogenetic relationships, because Numts evolve under constraints that are different than those of mtDNA [29–31]. The nuclear and mtDNA genomes have different rates and patterns of mutations and, because Numts are not functional, they do not evolve under purifying selection, explaining why they readily accumulate stop codon and frameshift mutations [32].

Four major arguments suggest that the *COI* and *Cytb* sequences produced in the present study cannot be considered as being Numts: (1) as the primers used here for PCR and sequencing were specially designed to amplify mtDNA genes, all our chromatograms were perfectly readable, without double peaks; (2) there is a high identity (99.8%) between our *Cytb* sequences of *M. pusillus* and the partial sequence previously obtained by Juste et al. [12] using a different set of primers; (3) there is no stop codon or indel (insertion or deletion); and (4) the tree reconstructed from *COI* sequences is very similar to that obtained from *Cytb* sequences (Figs. S3 and S1), with several identical nodes robustly supported in both analyses (Fig. 3). Therefore, there is no doubt about the mitochondrial origin of our *COI* and *Cytb* sequences.

4.3. Rampant mitochondrial introgression into *Epomophorus gambianus*?

The occasional mating between distinct species can result in genetic introgression of the mtDNA genome or nuclear alleles of one species into the gene pool of another

species. This is a cause of gene/species tree discordance. Mitochondrial introgression following hybridization has been widely inferred, especially when dealing with differences in genetic signal between mtDNA and nuclear markers [13,33–36], but documented cases in bats are very rare [37]. The mtDNA from one species may completely replace that of another species, without leaving any trace of nuclear introgression or morphological change [38]. In such cases, the mtDNA tree generally shows a typical topology, in which one of the two species is found to be monophyletic and the other paraphyletic. In our case, the mtDNA tree shows a more complex pattern, in which both species are polyphyletic. Assuming the hypothesis of mtDNA introgression implies therefore that interspecific breeding occurred frequently and at different periods of time. In addition, IMa analyses suggest that mtDNA introgression events occurred from *M. pusillus* into *E. gambianus*. According to this scenario, females of *M. pusillus* have been integrated several times into *E. gambianus* populations and all thirteen mtDNA haplotypes detected here for *E. gambianus* were transferred from *M. pusillus*.

Three main arguments concerning DNA variation, ecology and sexual behaviour render this scenario plausible.

Firstly, mtDNA variations between *E. gambianus* and *M. pusillus* are between 0 and 2.5% for *COI* and between 0 and 2.4% for *Cytb*. These low levels of variation are comparable with intraspecific divergence calculated for *COI* and *Cytb* sequences in several species of the family Pteropodidae (Tables S2 and S3). By contrast, the complete *Cytb* sequence available for *E. wahlbergi* is divergent from those of *E. gambianus* and *M. pusillus* (4.8–5.7%). These comparisons support therefore the fact that all mtDNA sequences produced here for *E. gambianus* and *M. pusillus* originated from only one species, namely *M. pusillus*.

Secondly, both species are found in sympatry from Senegal to CAR (Fig. 2), where they occupy the same type of habitat, i.e. the forest-savannah mosaic [9,39], which is a transition zone, or ecotone, between the tropical moist broadleaf forests and the drier savannahs. As a consequence, they were often collected in the same mist nets (personal observations) and were observed roosting in the same tree [40]. The geographic distribution of *M. pusillus* (Fig. 2) suggests, however, that it is more dependent on rainforests than *E. gambianus*. Hybrid zones often coincide with ecotones, as previously shown in African elephants [33] or mouse lemurs [41]. Astonishingly, the situation in African fruit bats is comparable with that of African elephants, as multiple events of mtDNA introgression have been inferred from the smallest forest species (*M. pusillus*/*L. cyclotis*) into the largest savannah species (*E. gambianus*/*L. africana*).

Thirdly, introgressive hybridization is known to occur more frequently between sister-species [42], for which prezygotic isolation (behavioural and mechanical) and postzygotic isolation (zygote mortality and hybrid inviability and sterility) have not been fully accomplished. Here, both species share very similar reproductive patterns. Females have two parturition periods during rainy seasons around April and September/October in West Africa. Furthermore, adult males exhibit similar sexual characteristics (white tufts of erectile hair grow within circular pocket-like folds of skin on the shoulders) and have the same typical courtship behaviour (with courting calls, the display of the white shoulder patches and wing flapping) [40,43].

However, three arguments dispute an introgression from *M. pusillus* into *E. gambianus*. Firstly, the male courting call is at a different frequency: 1750 and 2800 Hz for *Epomophorus* and *M. pusillus*, respectively [44]. Secondly, the two species have different male karyotypes: *M. pusillus* is XY_1Y_2, with $2n = 35$ and a fundamental number (FN) of 64 [45], whereas *E. gambianus* is XO, with $2n = 35$ and a fundamental number (FN) of 68 [46]. The karyotype available for *E. gambianus* was established from two specimens from Kenya and Zimbabwe, i.e. from two populations of *E. gambianus crypturus*. Since this subspecies has been elevated to full species status in recent classifications, i.e. *Epomophorus crypturus* [1,2], we can expect a different karyotype for specimens from CAR. Thirdly, and probably most importantly, their body size is very different, as adult males of *E. gambianus* are four times heavier than females of *M. pusillus* (Fig. 1; [39,47]). Such morphological differences may preclude copulation and cause mortality in pregnant *M. pusillus* females carrying an overly large hybrid foetus.

4.4. Incomplete lineage sorting of mtDNA haplotypes and positive selection in the FGB gene?

As an alternative to the hypothesis of mtDNA introgression, the phenomenon of incomplete lineage sorting, due to the retention of ancient polymorphisms, can be invoked to explain the polyphyletic pattern observed here for mtDNA haplotypes of *E. gambianus* and *M. pusillus*. If the mtDNA genome has not reached coalescence at the species level, some lineages from one species may be more closely related to a lineage in the other species than to other lineages within the same species. Incomplete lineage sorting can take place between sister-species, when the recent speciation event occurs before sorting is completed [48,49]. A recent speciation event between *E. gambianus* and *M. pusillus* is here supported by the fact that both mtDNA and nuDNA markers show very low interspecies distances ($< 2.5\%$ for mtDNA and $< 1\%$ for nuDNA).

The shorter the coalescence time, the less likely the gene will suffer from the 'incomplete lineage sorting' problem. In this regard, mtDNA has an advantage, because it coalesces in one-fourth the time of nuclear genes [48,49]. Therefore, if incomplete lineage sorting of mtDNA haplotypes is possibly the cause of the polyphyly, we should assume a faster time of coalescence for the *FGB* nuclear gene than for the mitochondrial genome.

A strong positive selection in the *FGB* gene may have led to selecting one or several particular allelic forms rapidly in one or both species before the coalescence of mtDNA. Actually, positive selection can be advanced for *M. pusillus*, because most *FGB* alleles detected in this species are highly derived with respect to those found in *E. gambianus*. Indeed, Mp1 to Mp13 alleles differ by five mutations from

the ancestral allelic form shared by *E. gambianus* and *M. pusillus* (Fig. 5). In accordance with the hypothesis of positive selection, the primitive allele Mp14 is very rare in *M. pusillus* (2.6%) and it is always found in heterozygosity with a derived Mp1 allele, suggesting that at least one derived allele is necessary to ensure the function of *FGB* in *M. pusillus*.

The *FGB* gene codes for the beta polypeptide chains of the fibrinogen, a glycoprotein involved in blood coagulation. In humans, several polymorphisms in the *FGB* gene were found to affect the concentration of clotting factor and therefore the plasma viscosity [50,51] or to protect against myocardial infarction [52]. In addition, previous physiological studies on fruit bats have concluded that there is a linear correlation between body mass and heart pulse rate [53]. In other words, the smallest bats, which have the highest wing beat frequency, have the highest heart rate. Therefore, we can expect to have a higher selective pressure for *FGB* alleles in *M. pusillus* than in *E. gambianus*.

5. Conclusion

Our study has revealed that mtDNA barcodes are not useful for assigning an unknown specimen to either *E. gambianus* or *M. pusillus*. By chance, these two species can be easily identified in the field. Within *Epomophorus*, several sympatric species are however difficult to distinguish using a morphological taxonomic key. For example, the distinction between *E. anselli*, *E. gambianus, E. labiatus* and *E. minimus*, may be tricky in East Africa, as diagnosing morphological characters are only measurements, which show important overlapping variations between species and which cannot be used on juvenile and subadult specimens.

Additional nuclear markers and other individuals from allopatric populations of *E. gambianus* (i.e., from South-East Africa) and *M. pusillus* (i.e., from Equatorial Africa), as well as from other species of *Epomophorus* and *Micropteropus,* need to be included in further molecular studies in order to decipher between introgression and incomplete lineage sorting and to provide taxonomic conclusions at genus and species levels.

Disclosure of interest

The authors declare that they have no conflicts of interest concerning this article.

Acknowledgements

We are very grateful to Carine Ngoagouni who helped in collecting tissue samples. We thank Pr. Alain Le Faou, from the *Institut Pasteur* of Bangui, for permission and support to conduct research in CAR. We thank Céline Bonillo for help in DNA cloning. This work was supported by the *Muséum National d'Histoire Naturelle* (MNHN), CNRS, *Institut Pasteur* of Bangui, PPF "*État et structure phylogénétique de la biodiversité actuelle et fossile*," and "*Consortium National de Recherche en Génomique*." It is part of agreement No. 2005/67 between the Genoscope and the MNHN on the project "Macrophylogeny of Life".

Appendix A. Supplementary data

There is supplementary material associated to the electronic version of this article available at doi:10.1016/j.crvi.2011.05.003.

References

[1] N.B. Simmons, Order Chiroptera, in: D.E. Wilson, D.M. Reeder (Eds.), 3rd Ed., Mammals species of the world: a taxonomic and geographic reference, 1, Johns Hopkins University Press, 2005, pp. 312–529.
[2] IUCN Red List of Threatened Species. Version 2010.4. www.iucnredlist.org.
[3] K.M. Helgen, L.E. Helgen, D.E. Wilson, Pacific flying foxes (Mammalia: Chiroptera): two new species of Pteropus from Samoa, probably extinct, Am Museum Novitates 3646 (2009) 1–37.
[4] D.T.S. Hayman, A.R. Fooks, D. Horton, R. Suu-Ire, A.C. Breed, A.A. Cunningham, J.L.N. Wood, Antibodies against Lagos bat virus in Megachiroptera from West Africa, Emerg. Infect. Dis. 14 (2008) 926–928.
[5] W. Markotter, I. Kuzmin, C.E. Rupprecht, L.H. Nel, Phylogeny of Lagos bat virus: challenges for lyssavirus taxonomy, Virus Res. 135 (2008) 10–21.
[6] E.M. Leroy, B. Kumulungui, X. Pourrut, P. Rouquet, A. Hassanin, P. Yaba, A. Delicat, J.T. Paweska, J.P. Gonzalez, R. Swanepoel, Fruit bats as reservoirs of Ebola virus, Nature 438 (2005) 575–576.
[7] D.T.S. Hayman, P. Emmerich, L.F. Yu, L.F. Wang, R. Suu-Ire, A.R. Fooks, A.A. Cunningham, J.L.N. Wood, Long-term survival of an urban fruit bat seropositive for Ebola and Lagos viruses, Plos One 5 (2010) e11978.
[8] J.S. Towner, X. Pourrut, C.G. Albarino, C.N. Nkogue, B.H. Bird, G. Grard, T.G. Ksiazek, J.P. Gonzalez, S.T. Nichol, E.M. Leroy, Marburg virus infection detected in a common African bat, Plos One 2 (8) (2007) e764.
[9] W. Bergmans, Taxonomy and biogeography of African fruit bats (Mammalia, Megachiroptera). 5. The genera Lissonycteris Andersen, 1912, Myonycteris Matschie, 1899 and Megaloglossus Pagenstecher, 1885; general remarks and conclusions; annex: key to all species, Beaufortia 47 (1997) 11–90.
[10] P.D.N. Hebert, A. Cywinska, S.L. Ball, J.R. DeWaard, Biological identification through DNA barcodes, Proc R Soc Lond B 270 (2003) 313–321.
[11] R.W. Hayman, J.E. Hill, Order Chiroptera, in: The Mammals of Africa. An identification manual, Part 2, Smithsonian Institution Press, City of Washington, 1971.
[12] J. Juste, Y. Alvarez, E. Tabares, A. Garrido-Pertierra, C. Ibanez, J.M. Bautista, Phylogeography of African fruitbats (Megachiroptera), Mol. Phylogenet. Evol. 13 (1999) 596–604.
[13] A. Hassanin, A. Ropiquet, Resolving a zoological mystery: the Kouprey is a real species, Proc. R. Soc. Lond. B. 274 (2007) 2849–2855.
[14] A. Hassanin, A. Ropiquet, F. Delsuc, C. Hammer, B. Jansen van Vuuren, C. Matthee, M. Ruiz-Garcia, F. Catzeflis, V. Areskoug, T.T. Nguyen, A. Couloux. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes, submitted.
[15] T.A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucleic Acids Symposium Series 41 (1999) 95–98.
[16] D. Posada, K.A. Crandall, MODELTEST: testing the model of DNA substitution, Bioinformatics 14 (1998) 817–818.
[17] D.L. Swofford, PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. 0b10, Sinauer Associates, Sunderland, MA, 2003.
[18] J.P. Huelsenbeck, F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees, Bioinformatics 17 (2001) 754–755.
[19] A. Rambaut, A.J. Drummond, Tracer Version 1.4 software available at http://tree.bio.ed.ac.uk/software/tracer.
[20] T. Sota, A.P. Vogler, Reconstructing species phylogeny of the carabid beetles *Ohompterus* using multiple nuclear DNA sequences: heterogeneous information content and the performance of simultaneous analyses, Mol. Phylogenet. Evol 26 (2003) 139–154.
[21] H.J. Bandelt, P. Forster, A. Rohl, Median-joining networks for inferring intraspecific phylogenies, Mol. Biol. Evol. 16 (1999) 37–48.
[22] M. Nei, Molecular evolutionary genetics, Colombia University Press, 1987.

[23] F. Tajima, Measurement of DNA polymorphism, in: Mechanisms of molecular evolution, Sinauer Associates, 1993 p. 37–59.
[24] P. Librado, J. Rozas, DnaSP v5: a software for comprehensive analysis of DNA polymorphism data, Bioinformatics 25 (2009) 1451–1452.
[25] J. Hey, Isolation with migration models for more than two populations, Mol. Biol. Evol. 27 (2010) 905–920.
[26] J. Hey, R. Nielsen, Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. Persimilis*, Genetics 167 (2004) 747–760.
[27] Y.J. Won, J. Hey, Divergence population genetics of chimpanzees, Mol. Biol. Evol. 22 (2005) 297–307.
[28] D. Bensasson, D.X. Zhang, D.L. Hartl, G.M. Hewitt, Mitochondrial pseudogenes: evolution's misplaced witnesses, Trends Ecol. Evol. 16 (2001) 314–321.
[29] D.A. Triant, J.A. Dewoody, The occurrence, detection, and avoidance of mitochondrial DNA translocations in mammalian systematic and phylogeography, J. Mammalogy 88 (2007) 908–920.
[30] H. Song, J.E. Buhay, M.F. Whiting, K.A. Crandall, Many species in one: DNA barcoding overestimates in the number of species when nuclear mitochondrial pseudogenes are coamplified, PNAS 105 (2008) 13486–13491.
[31] A. Hassanin, A. Ropiquet, A. Couloux, C. Cruaud, Evolution of the Mitochondrial Genome in mammals living at high altitude: new insights from a study of the Tribe Caprini (Bovidae, Antilopinae), J. Mol. Evol. 68 (2009) 293–310.
[32] A. Hassanin, C. Bonillo, X.N. Bui, C. Cruaud, Comparisons between mitochondrial genomes of domestic goat (*Capra hircus*) reveal the presence of Numts and multiple sequencing errors, Mitochondrial DNA 21 (2010) 68–76.
[33] A.L. Roca, N. Georgiadis, S.J. O'Brien, Cytonuclear genomic dissociation in African elephant species, Nature Genetics 37 (2005) 96–100.
[34] A. Ropiquet, A. Hassanin, Hybrid origin of the Pliocene ancestor of wild goats, Mol. Phylogenet. Evol. 41 (2006) 395–404.
[35] P.C. Alves, J. Melo-Ferreira, H. Freitas, P. Boursot, The ubiquitous mountain hare mitochondria: multiple introgressive hybridization in hares, genus *Lepus*, Philo, Transactions of the Royal Society B: Biological Sciences 363 (2008) 2831–2839.
[36] R.J. Petit, L. Excoffier, Gene flow and species delimitation, Trends Ecol. Evol. 24 (2009) 386–393.
[37] P. Berthier, L. Excoffier, M. Ruedi, Recurrent replacement of mtDNA and cryptic hybridization between two sibling bat species *Myotis myotis* and *Myotis blythii*, Pro. R. Soc. Lond. B. 273 (2006) 3101–3109.
[38] J.W.O. Ballard, M.C. Whitlock, The incomplete natural history of mitochondria, Mol. Ecol. 13 (2004) 729–744.
[39] W. Bergmans, Taxonomy and biogeography of African fruit bats (Mammalia, Megachiroptera). 2. The genera *Micropteropus* Matschie 1899 *Epomops* Gray 1870 *Hypsignathus* H. Allen 1861 *Nanonycteris* Matschie 1899 and *Plerotes* Andersen 1910, Beaufortia 39 (1989) 89–153.
[40] M.C. Boulay, C.B. Robbins, *Epomophorus gambianus* in mammalian species, Mammalian Species, 344, American Society of Mammalogists, 1989 p. 1–5.
[41] M. Gligor, J.U. Ganzhorn, D. Rakotondravony, O.R. Ramilijaona, E. Razafimahatratra, H. Zischler, A. Hapke, Hybridization between mouse lemurs in an ecological transition zone in southern Madagascar, Mol. Ecol. 18 (2009) 520–533.
[42] L. Gay, G. Neubauer, M. Zagalska-Neubauer, C. Debain, J.M. Pons, P. David, P.A. Crochet, Molecular and morphological patterns of introgression between two large white-headed gull species in a zone of recent secondary contact, Mol. Ecol. 16 (2007) 3215–3227.
[43] N.T. Owen-Ashley, D.E. Wilson, *Micropteropus pusillus*, Mammalian Species, 577, American Society of Mammalogists, 1998 p. 1–5.
[44] J. Haft, Beobachtungen zu Balzverhalten und Jugentwicklung beim Zwerg-Epualettenflughund, *Micropteropus pusillus* Peters, 1868 (Mammalia: Megachiroptera: Epomophorini), Zoologische Garten 72 (2002) 61–67.
[45] M.W. Haiduk, R.J. Baker, L.W. Robbins, D.A. Schlitter, Chromosomal evolution in African megachiroptera: G- and C-band assessment of the magnitude of change in similar standard karyotypes, Cytogenet. Cell Genet. 29 (1981) 221–232.
[46] R.L. Peterson, D.W. Nagorsen, Chromosomes of fifteen species of bats (Chiroptera) from Kenya and Rhodesia, Life Sci. Occas. Pap. 27 (1975) 1–16.
[47] W. Bergmans W, Taxonomy and biogeography of African fruit bats (Mammalia, Megachiroptera). 1. General introduction Material and Methods Results the genus *Epomophorus* Bennett 1836, Beaufortia 38 (1988) 75–146.
[48] J.C. Avise (Ed.), Phylogeography: the history and formation of species, Harvard Univ. Press, Cambridge, 2000.
[49] D.J. Funk, K.E. Omland, Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA, Ann. Rev. Ecol. Evol. Syst. 34 (2003) 397–423.
[50] G.D.O. Lowe, F.G.R. Fowkes, J. Dawes, P.T. Donnan, S.E. Lennie, E. Housley, Blood-viscosity, fibrinogen, and activation of coagulation and leucocytes in peripheral arterial-disease and the normal population in the Edinburg artery study, Circulation 87 (1993) 1915–1920.
[51] F.M. Van't Hooft, S.J.F. von Bahr, A. Silveira, A. Iliadou, P. Eriksson, A. Hamsten, Two common, functional polymorphisms in the promoter region of the beta-fibrinogen gene contribute to regulation of plasma fibrinogen concentration, Arterioscl. Thromb. Vasc. Biol. 19 (1999) 3063–3070.
[52] X.F Lu, H.J. Yu, X.Y. Zhou, L.Y. Wang, J.F. Huang, D.F. Gu, Influence of fibrinogen beta-chain gene variations on risk of myocardial infarction in a Chinese Han population, Chinese Med. J. 121 (2008) 1549–1553.
[53] R.E. Carpenter, Flight physiology of intermediate sized fruit-bats (Family: Pteropodidae), J. Exp. Biol. 120 (1986) 619–747.
[54] K. Andersen, Catalogue of the Chiroptera in the Collection of the British Museum, Megachiroptera, 1, Trustees British Museum (Natural History), London, 1912.
[55] D.R. Rosevear, The bats of West Africa, British Museum (Natural History), London, 1965.