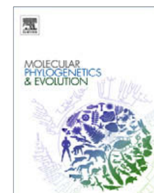
Molecular Phylogenetics and Evolution 50 (2009) 485–495

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev



# High mitochondrial diversity in geographically widespread butterflies of Madagascar: A test of the DNA barcoding approach

Marjorie C. Linares [a], Iván D. Soto-Calderón [a,b], David C. Lees [c], Nicola M. Anthony [a,*]

[a] Department of Biological Sciences, Computer Resources Center Building, Room 200, University of New Orleans, New Orleans, LA 70148, USA
[b] Biology Institute, University of Antioquia, Medellín AA1226, Colombia
[c] Department of Entomology, Natural History Museum, Cromwell Rd., London SW7 5BD, UK

ARTICLE INFO

*Article history:*
Received 16 May 2008
Revised 6 November 2008
Accepted 11 November 2008
Available online 21 November 2008

*Keywords:*
*Heteropsis*
*Wolbachia*
Lepidoptera
Barcoding
Comparative phylogeography

ABSTRACT

The standardized use of mitochondrial cytochrome *c* oxidase subunit I (COI) gene sequences as DNA barcodes has been widely promoted as a high-throughput method for species identification and discovery. Species delimitation has been based on the following criteria: (1) monophyletic association and less frequently (2) a minimum 10× greater divergence between than within species. Divergence estimates, however, can be inflated if sister species pairs are not included and the geographic extent of variation within any given taxon is not sampled comprehensively. This paper addresses both potential biases in DNA divergence estimation by sampling range-wide variation in several morphologically distinct, endemic butterfly species in the genus *Heteropsis*, some of which are sister taxa. We also explored the extent to which mitochondrial DNA from the barcode region can be used to assess the effects of historical rainforest fragmentation by comparing genetic variation across *Heteropsis* populations with an unrelated forest-associated taxon *Saribia tepahi*. Unexpectedly, generalized primers led to the inadvertent amplification of the endosymbiont *Wolbachia*, undermining the use of universal primers and necessitating the design of genus-specific COI primers alongside a *Wolbachia*-specific PCR assay. Regardless of the high intra-specific genetic variation observed, most species satisfy DNA barcoding criteria and can be differentiated in the nuclear phylogeny. Nevertheless, two morphologically distinguishable candidate species fail to satisfy the barcoding 10× genetic distance criterion, underlining the difficulties of applying a standard distance threshold to species delimitation. Phylogeographic analysis of COI data suggests that forest fragmentation may have played an important role in the recent evolutionary diversification of these butterflies. Further work on other Malagasy taxa using both mitochondrial and nuclear data will provide better insight into the role of historical habitat fragmentation in species diversification and may potentially contribute to the identification of priority areas for conservation.

© 2008 Elsevier Inc. All rights reserved.

## 1. Introduction

DNA barcoding (henceforth "barcoding") has been heralded as a major new tool in species description and discovery (Hebert et al., 2003). Proponents of barcoding initiatives (such as the Consortium for the Barcoding of Life) argue that it will revolutionize our current understanding of biological diversity (Ratnasingham and Hebert, 2007), especially in hyper-diverse or cryptic taxa such as tropical lepidoptera (Hajibabaei et al., 2006; Hebert et al., 2004a). Although mitochondrial DNA (mtDNA) has been widely promoted as a vehicle for barcoding studies, mtDNA markers such as COI have long been considered the work-horse of intra-specific phylogeography (Avise 2004) and invertebrate phylogenetic systematics (Caterino et al., 2000). Barcoding has also attracted considerable controversy and several important criticisms of this approach have been raised. These include the undue emphasis it places on molecular data (Will et al., 2005; Rubinoff, 2006), its over-reliance on a single marker system (Elias et al., 2007; Hickerson et al., 2006; Neigel et al., 2007; Roe and Sperling, 2007) and the costs and limitations imposed by the quality of the reference database (Cameron et al., 2006; Ekrem et al., 2007). Moreover, interpretation of mitochondrial genetic diversity may be hampered by inadvertent amplification of paralogous nuclear copies (Numts) (Bensasson et al., 2001) and heritable endosymbionts such as *Wolbachia* (Hurst and Jiggins, 2005; Whitworth et al., 2007). Despite these shortcomings, barcoding may prove to be an efficient tool for rapid assessment of taxonomic diversity, especially in species groups that are otherwise difficult to study (e.g. Evans et al., 2007).

Within the mitochondrial barcoding framework, species delimitation is based on one of two criteria: (1) reciprocal monophyly (Wiens and Penkrot, 2002) and/or less frequently (2) a genetic dis-

* Corresponding author. Fax: +1 504 280 6121.
*E-mail address:* nanthony@uno.edu (N.M. Anthony).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.11.008

tance cutoff such as the proposed 10× average intra-specific distance (Hebert et al., 2004b). Despite their intuitive appeal, these criteria may be difficult to satisfy in geographically widespread or closely related taxa (Elias et al., 2007; Moritz and Cicero, 2004). Over-estimation of species number could readily arise from (i) under-estimation of intra-specific variation through poor geographic sampling and (ii) over-estimation of inter-specific distances through the omission of sister species (Moritz and Cicero, 2004; Meyer and Paulay, 2005). Moreover, substantial overlap between inter and intra-specific variability may exist, complicating the adoption of a fixed distance threshold for species delimitation (Meyer and Paulay, 2005; Meier et al., 2006).

In order to address these two important limitations to barcoding analyses, a geographically comprehensive study of several widespread taxa within the hyper-diverse butterfly genus *Heteropsis* was initiated using the lepidopteran-specific cytochrome oxidase I (COI) primers published in Hebert et al. (2004a). Sequences from two nuclear genes (elongation factor [EF-$1\alpha$] and *wingless*) were also obtained from representative samples of each species in order to provide independent verification of species identifications made through mitochondrial barcoding data.

*Heteropsis* belongs to a group of Old World grass-feeding genera in the subfamily Satyrinae, traditionally treated as the subtribe Mycalesina in the tribe Elymniini (Lees et al., 2003), but now

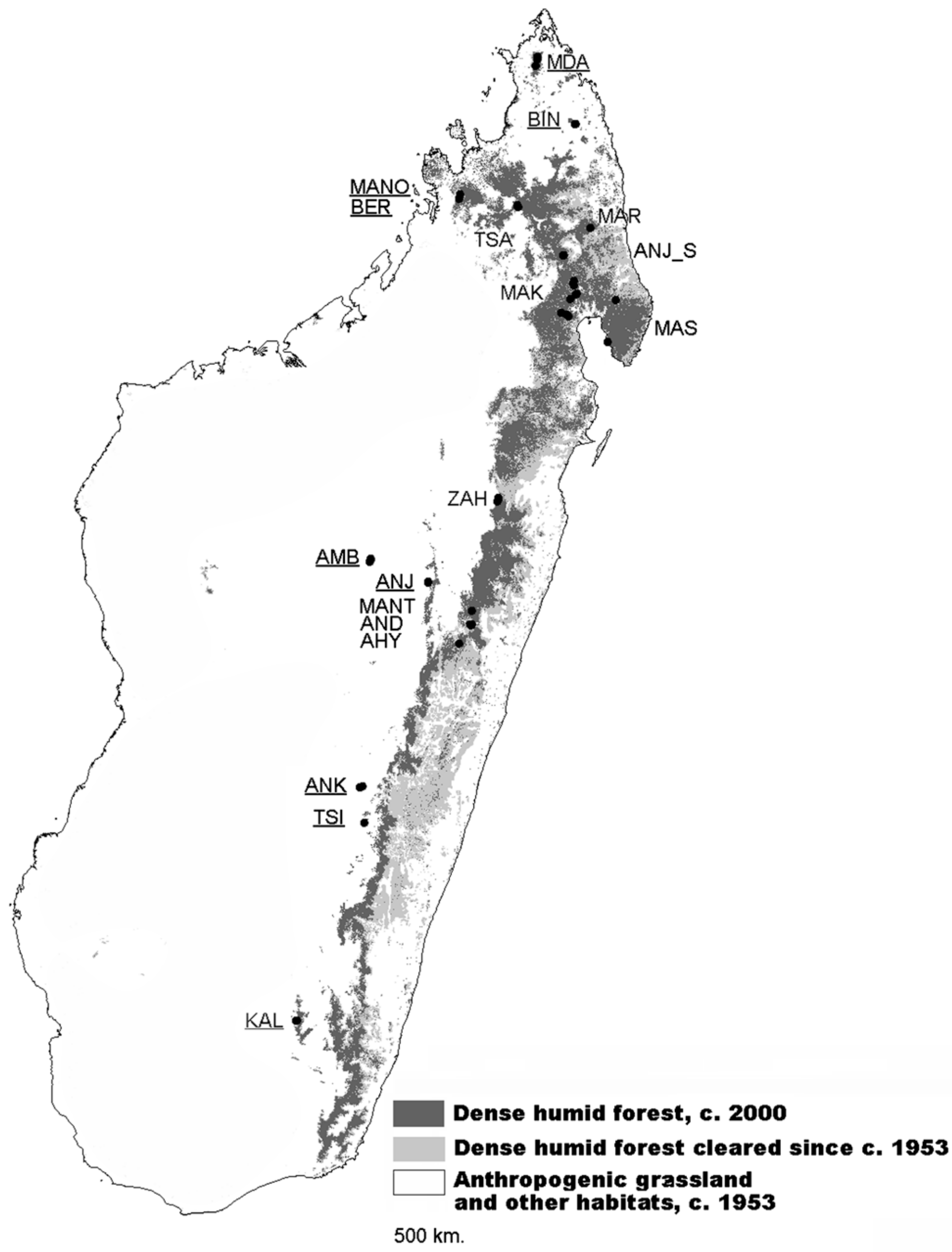



**Fig. 1.** Location and fragmentation history of sites sampled in this analysis. The letter codes correspond to geographic locations listed in Supplementary Table 1. Sites that are underlined represent "old" forest fragments lacking a clearly documented connection history. Sites that are not underlined form part of an essentially contiguous main rainforest block or are known to be recently fragmented (AND). Dark areas represent the near current extent of forest cover. Lighter grey areas represent forest that has been cleared since the ~1950s.

**Table 1**
Genbank *wsp* sequences used in phylogenetic analysis of *Wolbachia* variation.

| Host | Genbank ID | Strain or isolate | Super-group | Reference |
|---|---|---|---|---|
| *Drosophila melanogaster* | DQ235407 | *w*Mel | A | Paraskevopoulos et al., 2007 |
| *Drosophila prosaltans* | DQ118779 | *w*Pro | A | Miller and Riegler, 2006 |
| *Drosophila simulans* | AF020067 | *w*Cof | A | Braig et al., 1998 |
| *Drosophila simulans* | DQ235409 | *w*Au | A | Paraskevopoulos et al., 2007 |
| *Drosophila simulans* | DQ235408 | *w*Ha | A | Unpublished |
| *Drosophila quadraria* | DQ412096 | *w*Ri | A | Mateos et al., 2006 |
| *Drosophila willistoni* | AY620229 | *w*Wil | A | Miller and Riegler, 2006 |
| *Rhagoletis cerasi* | AF418557 | *w*Cer2 | A | Riegler and Stauffer, 2002 |
| *Haematopinus suis* | AY596785 | *w*Pol1 | A | Perotti et al., 2004 |
| *Hypolimnas bolina* | EF025179 | *w*Bol2 | A | Charlat et al., 2006 |
| *Supella longipalpa* | EF193198 | 3A | A | Vaishampayan et al., 2007 |
| *Blatella* sp. | EF193197 | 2E | A? | Vaishampayan et al., 2007 |
| *Drosophila mauritiana* | DQ412107 | *w*No | B | Mateos et al., 2006 |
| *Drosophila sechellia* | AF468036 | *w*Sn | B | Charlat et al., 2002 |
| *Drosophila simulans* | AF020069 | *w*Ma | B | Braig et al., 1998 |
| *Acraea alcinoe* | AJ271196 | | B | Jiggins, unpublished |
| *Pseudacteon nudicornis* | AY878109 | *w*Pnud | B | Dedeine et al., 2005 |
| *Cotesia aeamiae* | DQ241818 | | B | Rincon et al., unpublished |
| *Protocalliphora sialia* | DQ842482 | 21 | B | Baldo et al., 2006 |
| *Ostrinia scapularis* | DQ842481 | 32 | B | Baldo et al., 2006 |
| *Encarsia formosa* | DQ842471 | 33 | B | Baldo et al., 2006 |
| *Leptopilina victoriae* | DQ380527 | *w*Lvic | B | Gavotte et al., 2007 |
| *Cimex lectularis* | DQ842459 | 36 | F | Baldo et al., 2006 |

placed in the tribe Satyrini (Peña and Wahlberg, 2008). This genus was selected because it contains several examples of very recently diverged species (Torres et al., 2001) that are morphologically distinct. In addition, the taxa selected for study show varying degrees of site-specific phenotypic variability (Lees, 1997), are all endemic to Madagascar and are distributed widely across entire biomes or different forest habitats (Lees et al., 1999). Thus, their broad geographic distribution and high morphological variation makes them particularly good candidates for testing the accuracy of barcoding methods. Morphological characters used to distinguish *Heteropsis* species include male androconial organs and semiochemicals (Lees, 1997; Kubbinga, 2006), male inflated hind wing veins and genitalic characters, wing ocellus configuration and wing coloration (Lees, 1997).

As some phenotypically distinct populations within several *Heteropsis* species are confined to one or more distinct forest fragments, this study also provided an ideal opportunity to test the ability of mtDNA to detect the effects of historical fragmentation. In keeping with the principle of comparative phylogeography (Bermingham and Moritz, 1998), the rainforest interior species *Saribia tepahi* (Boisduval, 1833), within the family Riodinidae, subfamily Nemeobiinae, tribe Abisarini (Hall and Harvey, 2002), was also included as an additional test of the role of historical fragmentation in evolutionary diversification. Our primary hypothesis was that if historical fragmentation plays an important role in diversification, species of both genera should exhibit congruent patterns of population genetic structure. As indicated by recent satellite and aerial photography imagery represented in Fig. 1, many forest sample sites belong to a once contiguous eastern rainforest block, fragmented by slash and burn activity since ~1950s (Green and Sussman, 1990). In contrast, other fragments found on the central Plateau (underlined in Fig. 1) were isolated from the main forest block prior to 1950 (Harper et al., 2007) and possibly before 1900 (see Cowan, 1882). For example, whereas Ankazomivady forest (Fisher and Robertson, 2002) appears to be clearly isolated on a map from between 1949 and 1957 (Humbert and Cours-Darne, 1965; Harper et al., 2007), other rainforest fragments (like Binara and Montagne d'Ambre) may have been unconnected to a similar forest type for several thousand years, given that they are on isolated mountains surrounded by dry deciduous forests.

As our preliminary work based on the barcoding COI primers of Hebert et al. (2004a) also amplified the COI homologue in the endosymbiont *Wolbachia* (see also AY800177; Deans et al., 2006), primers were redesigned to better match the lepidopteran sequence. A PCR diagnostic was then used to survey butterfly samples for *Wolbachia* in order to evaluate whether infection could have influenced patterns of mitochondrial diversity (Dean et al., 2003; Jiggins, 2003; Shoemaker et al., 2004) and therefore the interpretation of barcoding data (Hurst and Jiggins, 2005). Lastly, we also examined sequence data for the presence of nuclear translocations of mtDNA which, if left undetected, may also represent another potential source of error (Song et al., 2008).

## 2. Materials and methods

### 2.1. Study sites, taxonomic and geographic sampling

Our study material includes a number of closely related taxa with subtle morphological differences that could either be new cryptic species, or geographically divergent subpopulations. With about 70 species within the genus, a full taxonomic sampling of *Heteropsis* was not feasible, so we examined three clades as defined by Lees (1997) and Torres et al. (2001), each of which contain widespread species that occur in a range of recent and historical forest fragments.

Butterflies were hand netted or captured in fruit-bait traps. Each specimen was placed in a glassine envelope and one or more legs were removed and placed in ethanol. Alternatively, a portion of the abdomen from the dried specimen was removed for DNA extraction. Table S1 lists all of the butterfly specimens examined, their specimen number, geographic site of origin and corresponding site code. Butterflies were placed to genus and species by one of the co-authors (DL) and by morphological comparison with type specimens. In most cases this could be done simply by wing shape and pattern and by male androconial morphology.

Specimens of the mycalesine species in this study, numbered according to their respective clades, are as follows: (1) the *exocellata* clade represented by *Heteropsis exocellata* (Mabille, 1880), (2) the *ankova* clade comprising *H. turbata* (Butler, 1880) and *H. pallida* (Oberthür, 1916) and (3) the *H. subsimilis* clade comprising

*H. pauper* (Oberthür, 1916), *H. subsimilis* (Butler, 1979) and two undescribed *Heteropsis* taxa (*H.* 23 and *H.* 25). Lastly, the riodinid *S. tepahi* was included as a taxonomic and ecological contrast to the focal group of interest. Like most other taxa in this study, this species is restricted to rainforest interior habitats but is capable of more powerful flight (Lees, unpubl.).

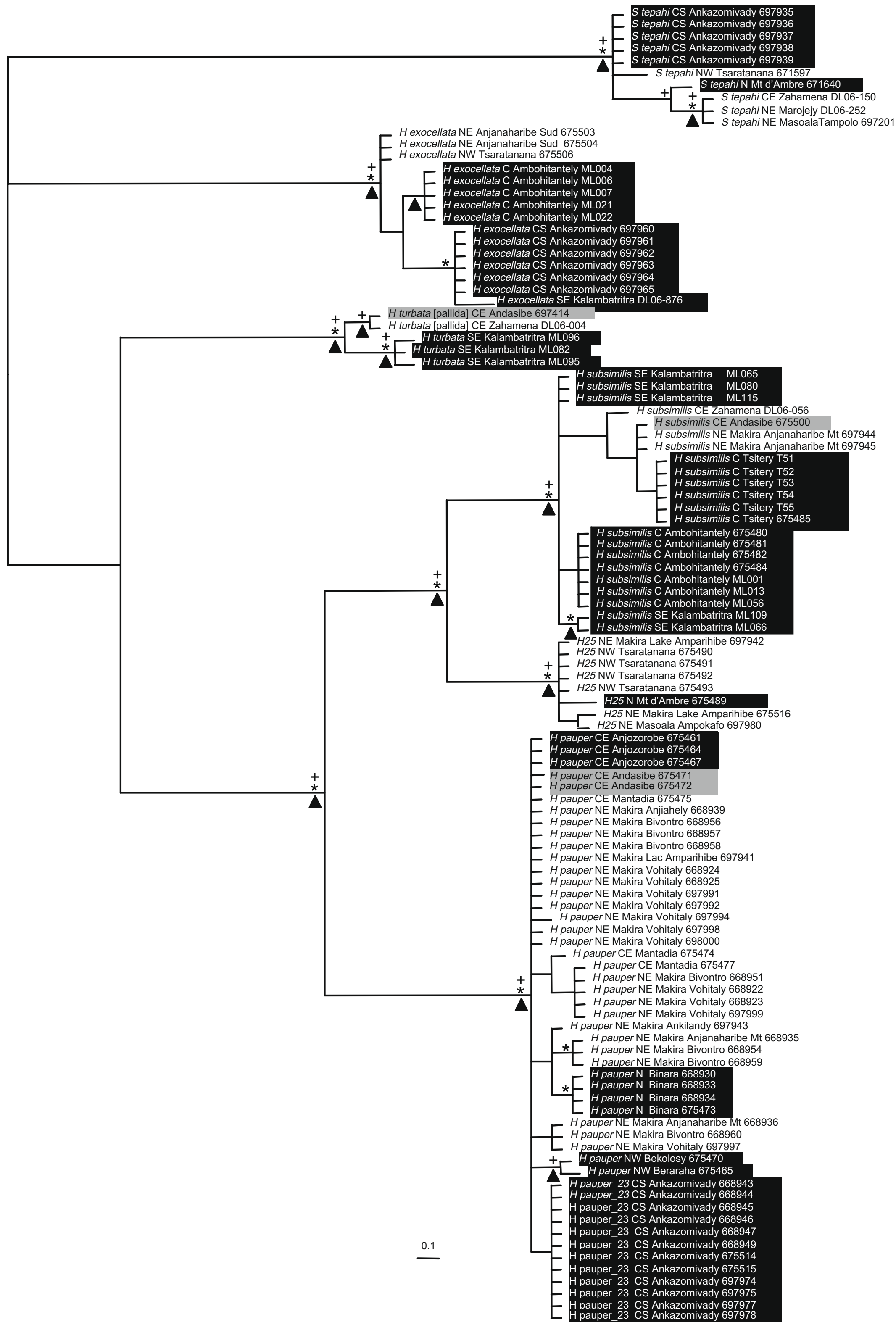



**Fig. 2.** Bayesian phylogeny of butterfly COI sequences. Support for individual nodes is indicated by phylogenetic method where (+) denotes 75% or more bootstrap support in either Maximum Parsimony, or (▲) Maximum Likelihood, and (∗) indicates 95% or greater posterior support in Bayesian analysis. Sequences highlighted in black are from older fragments whereas those highlighted in grey are from a recent forest fragments (Andasibe). Sites that are not highlighted at all are from the contiguous eastern rainforest block.

Butterfly specimens were collected from rainforest sites across Madagascar (Fig. 1). Species exhibit differences in both ecological specialization and elevational zonation (Kremen, 1994; Lees et al., 1999). *Heteropsis* species feed on grasses or bamboos (Poaceae) as larvae and usually specialize on over-ripe fruit as adults (Lees, 1997). Both *H. turbata* and *H. pallida* are unusual in that they fly just outside the edge of the forest margin in open marshy grassland and are therefore less likely to have populations isolated by forest fragmentation. In contrast, the riodinid butterfly species *S. tepahi* specializes on the shrub genus *Oncostemon* (Myrsinaceae). *H. subsimilis*, *H. pauper*, and *S. tepahi* all occur from sea level to at least 1650 m. above sea level in a variety of rainforest habitats throughout Madagascar (Kremen, 1994; Lees et al., 2003). *H. turbata* is widespread from 440 to 2000 m. (with *H. pallida* known between 910–970 m.) and *H. exocellata* is also widespread in rainforest from 670 to 1850 m. However, *H.* 25 is restricted to more northerly rainforests between 700 and 1650 m. (Lees, unpubl.).

Taxa also show intraspecific morphological differences. Butterflies in the first clade (*H. exocellata*) are phenotypically variable such that various local forms have been described as different species. For example, "*Henotesia*" *benedicta* was originally described from the Ambositra region (adjacent to Ankazomivady forest) whereas "*Henotesia*" *aberrans* was described from Mt. Tsaratanana, based on differences in the male genital valves (Paulian, 1951). However, these two species were later synonymised with *H. exocellata* by Lees et al. (2003), so it is of interest that we here resampled the populations from which these types were probably drawn. In the second clade, we also resampled *H. pallida*, which has dorsal (space Cu1a) wing ocellus rings which are yellowish rather than orange to reddish in *H. turbata*, and was originally described from the "Antsianaka" region located around Zahamena (Fig. 1). *H. pallida* has been variously regarded as a 'race' of *H. turbata* (d'Abrera, 1997) or as a distinct species (Lees 1997; Lees et al. 2003). In the *subsimilis* clade, *H.* 23 and *H.* 25 have been recognized as morphologically distinct taxa (Lees 1997; Torres et al. 2001). *H.* 23 is distinguished by a smoother hindwing margin and lighter color than is typical of *H. pauper*. This taxon, only known from Ankazomivady forest (Lees, 1997) was not reciprocally monophyletic to populations of *H. pauper* sampled from Masoala and Ranomafana (Torres et al., 2001) and is thus referred to hereafter as *H. pauper*_23). *H.* 25 is an undescribed taxon in the *subsimilis* clade that displays a relatively more uniform yellowish ventral wing coloration pattern and differently inflated male hindwing veins (presumably androconial in function), compared to its inferred sister species *H. subsimilis* (Lees, 1997; Torres et al., 2001). Finally, *S. tepahi* is made up of several distinct geographically localized but as yet undocumented morphotypes (Lees, unpubl. data) and has no potential synonyms.

### 2.2. DNA amplification and sequencing

Genomic DNA extractions of butterfly specimens were carried out using standard methods for animal tissue (Sambrook and Russell, 2001) or using the DNeasy kit (Qiagen), according to manufacturer's instructions. Initial amplification of a 648 bp 5′ fragment of the *Heteropsis* mitochondrial COI gene was carried out using the Lep-F1 and Lep-R1 primer combination and thermo-cycling conditions outlined by Hebert et al. (2004a). PCR reactions were carried out in a total volume of 50 μL containing 4.0 mM $MgCl_2$, 0.2 μM of each primer, 0.2 mM dNTPs, 1× Buffer and 1.25 U of *Taq* DNA polymerase (Invitrogen). Thermo-cycling conditions were as follows: one initial cycle of 1 min at 94 °C followed by six cycles of 94 °C for 1 min, 45 °C for 1 min 30 s, 72 °C for 1 min 15 s, then 36 cycles of 94 °C for 1 min, 51 °C for 1 min 30 s and 72 °C for 1 min 15 s, with a final step of 72 °C for 5 min. After the discovery that the Lep-F1/R1 primer combination co-amplified *Wolbachia*, new butterfly COI primers were redesigned from existing lepidopteran sequences: Lep-F4 5′ CGGTACTTCTTTAAGTTTAA 3′ and Lep-R4b 5′

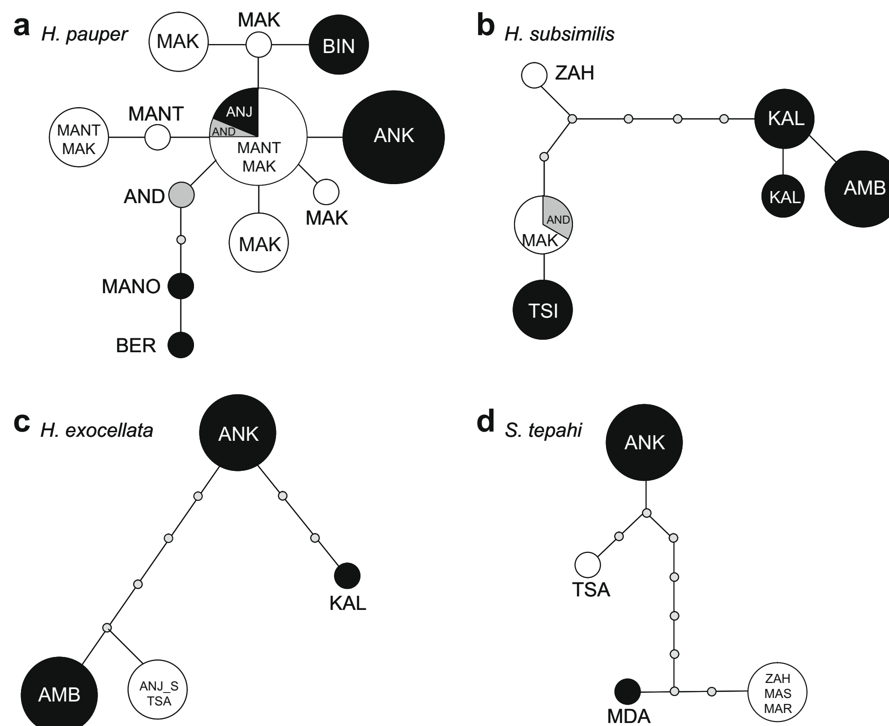



**Fig. 3.** Haplotype network of COI sequences obtained from: (a) *H. exocellata*, (b) *H. pauper*, (c) *H. subsimilis* and (d) *Saribia tepahi*. Dark haplotypes (black) are from older forest fragments whereas, grey haplotypes are from the recent forest fragment (Andasibe). All other haplotypes (white) are from contiguous rainforest. The smallest circles are inferred haplotypes.

AAATATAAACTTCTGGATGT 3′. These COI primers were also modified to amplify *S. tepahi*: Lep-F5s 5′ TGGAACATCTTTAAGTTTAT 3′ and Lep-R5s 5′ AAATATAAACTTCTGGATG 3′. The same PCR protocol was used for reactions using the F4/R4b and F5s/R5s primer combinations, except that the annealing temperature for the first six cycles was changed to 42 °C and for the remaining 36 cycles to 45 °C.

For the nuclear DNA phylogenetic analysis, one to five individuals were selected from each species. Two different regions were amplified: (1) a 1139 bp fragment of the EF-*1α* gene amplified in three overlapping segments using the primers Starsky, Luke, Cho, Verdi (Peña et al., 2006), EF51.9 and EFrcM4 (Monteiro and Pierce, 2001); (2) a 320 bp fragment of the *wingless* gene using primers LepWG1 and LepWG2 (Brower and DeSalle, 1998). *Bicyclus anyana* was also included as an outgroup in the phylogeny (AY218258, AY218276).

In order to assess the host range and geographic distribution of *Wolbachia* infections in the taxa included in this study, ~600 bp fragment of the *wsp* gene was PCR amplified from a subset of *Heteropsis* spp. ($n = 76$) and *S. tepahi* ($n = 6$) samples using primers published in Jeyaprakesh and Hoy (2000). A PCR positive and negative controls were included using *Wolbachia* infected and uninfected *Drosophila* strains respectively, purchased from the *Drosophila* stock center (http://flystocks.bio.indiana.edu/). PCR reactions were carried out in a 50 μl reaction volume containing 1x buffer, 1.75 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 μM of each primer and 1.5 U of *Taq* DNA polymerase (Invitrogen) using cycling reaction conditions of Jeyaprakesh and Hoy (2000). From samples that tested positive, we subsequently sequenced representatives of *Wolbachia wsp* gene sequences from all infected host species. Sequencing for all genetic markers was carried out using the BigDye version 1.1 (ABI) and reactions were run on an ABI 3100 automatic sequencer.

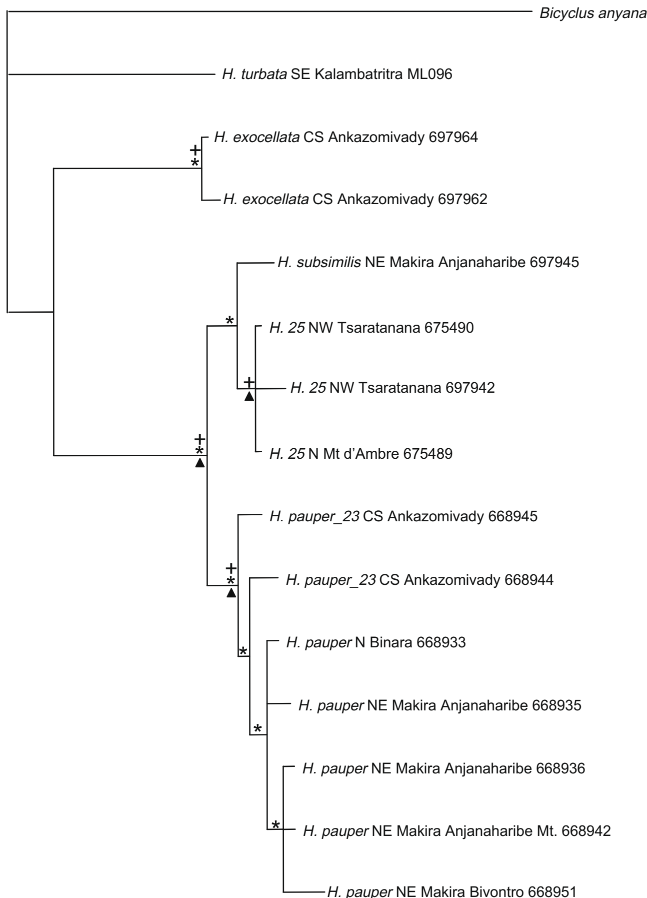



**Fig. 4.** Bayesian phylogeny of nuclear *EF-1α* and *wingless* DNA sequences obtained from the genus *Heteropsis*. Support for individual nodes is indicated by phylogenetic method where (+) denotes 75% or more bootstrap support in either Maximum Parsimony, or (▲) Maximum Likelihood, and (*) indicates 95% or greater posterior support in Bayesian analysis.

### 2.3. Data analysis

Alignment of butterfly sequences was carried out using the Clustal X algorithm (Thompson et al., 1997), except in the case of *Wolbachia* where some sequences were highly divergent and therefore difficult to align. To overcome this problem, *WSP* sequences were translated into amino acid sequences using the bacterial genetic code, aligned and then reverse translated back into nucleotide sequences using the program RevTrans (Wernersson and Pedersen, 2003). To check that there were no reading frameshifts and stop codons in the mitochondrial COI data, we translated the COI sequences using the invertebrate genetic code in Bioedit 7.0 (Hall, 1999), modified to code AGG as Lysine in Lepidoptera (Abascal et al., 2006).

Phylogenetic analyses of butterfly mitochondrial sequences ($n = 109$), nuclear DNA sequences ($n = 14$) and representative *wsp* sequences from A, B and F *Wolbachia* strains ($n = 43$) were carried out using maximum parsimony (MP), maximum likelihood (ML) and Bayesian methods. Compatibility between nuclear gene datasets was assessed using the partition homogeneity test implemented in PAUP 4.0b10 (Swofford 2000). *Wolbachia wsp* sequences other than those obtained from this study were taken from previous studies and from data available in Genbank (Table 1). MP and ML analyses were carried out using PAUP 4.0b10 and PhyML (Guindon and Gascuel, 2003) respectively. Bayesian analyses were carried out using the Monte Carlo Markov Chain method (MCMC) implemented in MrBayes (Ronquist and Huelsenbeck 2003).

In parsimony analyses, a starting tree was obtained using the stepwise addition option and heuristic searches were conducted using the tree-bisection-reconnection (TBR) heuristic algorithm. All character changes were considered unordered and unweighted. ML analysis was carried out under a general time-reversible (GTR) model using a neighbor-joining starting tree, as implemented in PhyML. Base frequencies, substitution rate categories and among site rate heterogeneity were estimated from the data. The strength of support for individual nodes in both MP and ML analyses was assessed by 1000 bootstrap replicates of the data. For Bayesian analyses, a GTR model was also adopted allowing for among site rate variation and invariant sites. Priors for model parameters were not defined *a priori* and were left at their default settings. In order to ensure that the MCMC did not get trapped in local optima (Leaché and Reeder, 2002), output was compared from two separate analyses, each made up of three heated chains and a cold chain, using the program TRACER (Rambaut and Drummond, 2007). The proportion of samples to be discarded as "burn in" was assessed by looking at the output from the *sump* command in MrBayes and by examining the MCMC trace files. In each case, runs were only accepted if the effective sample size (ESS) was greater than 500 for all model parameters. Convergence across analyses was assessed by verifying whether different runs attained the same stationary distribution and average log likelihood values. In general, chains were run for 10,000,000 iterations and trees were sampled every 1000 generations. For analysis of the nuclear data, chains were run for 1,000,000 iterations with the same sampling frequency. Support for a specific node was accepted if the relevant bootstrap value was ⩾75% and a posterior probability of ⩾0.95.

Inter versus intra-specific pair wise Kimura 2-parameter genetic distances were estimated for each species using the program MEGA v3.1 (Kumar et al., 2004). A haplotype tree was constructed

**Table 2**
*Wolbachia* distribution by site, species and presence (+) or apparent absence (–).

| Host butterfly | AMB | AND | ANK | ANJ | BER | BIN | MAK | MANO | MANT | TSA | TSI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Heteropsis pauper_23* | | | – | | | | | | | | |
| *Heteropsis pauper* | | + | + | + | + | + | + | – | – | | |
| *Heteropsis* 25 | | | | | | | – | | | + | |
| *Heteropsis subsimilis* | + | | | | | | + | | | | + |
| *Heteropsis exocellata* | | | + | | | | | | | | |
| *Saribia tepahi* | | | – | | | | + | | | | |

for the four species with 10 or more mtDNA sequences (*H. pauper*, *H. subsimilis*, *H. exocellata* and *S. tepahi*) using the program TCS version 1.21 (Clement et al., 2000). An analysis of molecular variance of the mitochondrial COI data was also carried out on the same taxonomic data sets using the program ARLEQUIN (Excoffier et al., 2005). In order to test for the effects of fragmentation history on genetic structure, populations within contiguous eastern rainforest or within forest fragments created since the 1950s were grouped into one region and sites representing older forest blocks (Fig. 1) were treated as additional individual groups. Alternatively, a recent forest fragment (Andasibe) formed since c.1900 was also treated as an additional separate group.

## 3. Results

### 3.1. Butterfly mitochondrial and nuclear DNA

Translation of COI sequences did not reveal evidence of frameshifts or stop codons, suggesting that these sequences are likely to be mitochondrial in origin. Tree-based analyses of DNA barcoding data provided strong support for the monophyly of *S. tepahi* and *Heteropsis* species (*H. exocellata, H. pauper, H. subsimilis*, and *H.* 25 and *H. turbata* when including *H. pallida* ['*H. turbata sensu lato*']) but not the novel morph '*H. pauper*_23' (=*Henotesia* sp. 23 [male] and 30 [female] of Torres et al. 2001; Fig. 2). Phylogenetic support for individual species nodes was high across all three methods, despite minor differences in overall topology. In keeping with barcoding criteria, all species were monophyletic. The average pair wise Kimura 2-parameter differences across all *Heteropsis* taxa were almost 20-fold greater (0.0862 ± 0.0325) than the mean pair-wise differences within taxa (0.0043 ± 0.0029). All pair-wise comparisons between individual candidate species pairs exceeded the suggested 10×-threshold of Hebert et al. (2004b) with the exception of *H.* 25 and *H. subsimilis* (8.44-fold), and *H. pauper*_23 and *H. pauper* (2.67-fold).

Phylogenetic analysis also indicated that populations inhabiting older fragments are frequently differentiated from nearby contiguous forest populations, although branch support for these cases is weak (Fig. 2). Network analyses show that within *Heteropsis* spp., older fragments are frequently fixed for a single haplotype that is shallowly derived from other sequences in contiguous forest sites (Fig. 3). *S. tepahi* shows a similar pattern but with deeper divergences between haplotypes from contiguous forest sites. Haplotype sharing between older fragments and contiguous forest was only observed for a single site: Anjozorobe (in the case of *H. pauper*). In contrast, haplotypes sampled from contiguous forest were sometimes shared across distant localities, indicating that few if any barriers to gene flow exist.

Analysis of molecular variance supported the regional structure evident in the phylogeny and network analysis. The among-regional component of the total genetic variance within *H. pauper* (55.84%) and *H. subsimilis* (65.87%) is high and reflects the importance of fragmentation history. Most of the remaining variance within *H. pauper* is found within populations (33.34%), whereas within *H. subsimilis* it is mostly among populations (31.33%). Similarly, *S. tepahi* exhibits a marked population genetic structure with a moderate regional genetic component of 30.21% and a high among population component of 69.79%. When the recent forest fragment (Andasibe) was treated as a separate group, the among-regional component dropped very slightly for *H. pauper* (54.32%) but substantially (42.59%) for *H. subsimilis*.

The partition homogeneity test indicates no evidence for incongruence between nuclear datasets ($p = 0.37$). Phylogenetic analysis of the EF-*1α* and *wingless* nuclear data provided good support for the species resolution observed in the mitochondrial data (Fig. 4). With the exception of *H. pauper*_23, all species where more than one individual was sequenced are monophyletic.

### 3.2. Wolbachia wsp variation

*Wolbachia wsp* gene fragments were detected at multiple sites in all *Heteropsis* and *Saribia* species tested in this study (Table 2).

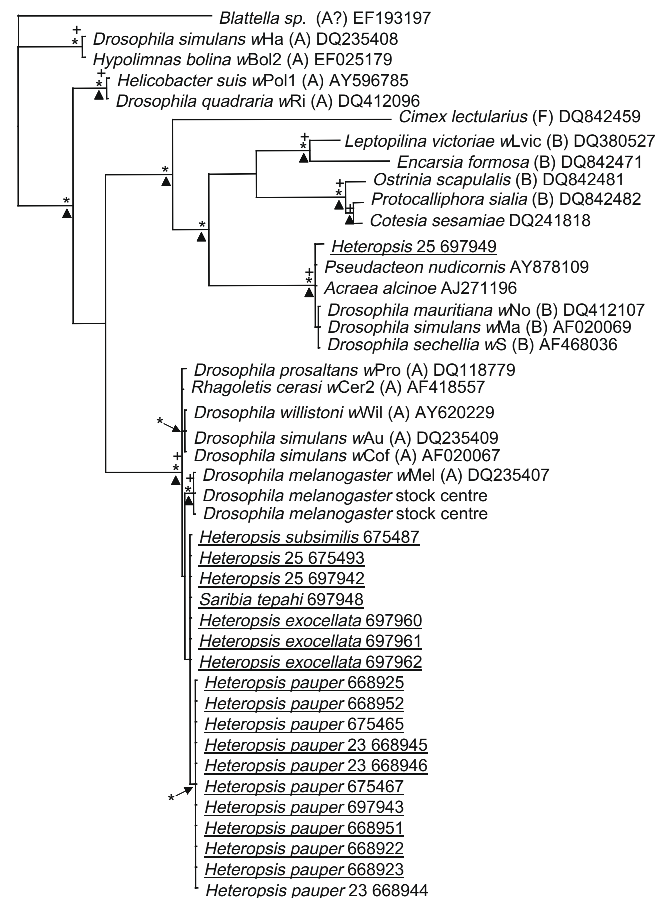



**Fig. 5.** Bayesian phylogeny of *Wolbachia wsp* sequences sampled in this study (underlined) and their relationship to supergroup sequences obtained from Genbank. Support for individual nodes is indicated by phylogenetic method where (+) denotes 75% or more bootstrap support in either Maximum Parsimony, or (▲) Maximum Likelihood, and (*) indicates 95% or greater posterior support in Bayesian analysis.

Phylogenetic analysis revealed evidence of two independent infections (Fig. 5). The first belongs to supergroup A and is found in all species tested in this study. When compared to available Genbank sequence data using the BLAST tool, this sequence type displays 99% identity to the *w*Au and *w*Wil strains from *Drosophila* (DQ235409, AY620229). A second sequence type belonging to supergroup B was also diagnosed in *H.* 25 and shares a high identity (99%) to *w*PNud strain identified in the parasitoid fly *Pseudoacteon nudicornis* and nymphalid butterfly *Acraea alcinoe* and a slightly lower (98%) identity to the *w*Ma, *w*No and *w*S strains from *Drosophila simulans*, *D. mauritania* and *D. sechellia*, respectively.

## 4. Discussion

### 4.1. Universal primers

One important technical challenge to the success of a barcoding approach is the availability of universal primers (Ekrem et al., 2007). As demonstrated from this study, the uncritical application of lepidopteran-specific COI primers (Hebert et al., 2004a) led to inadvertent amplification of the homologous COI domain in the *Wolbachia* genome and necessitated the redesign of suitable primers from butterfly COI sequence data. Close examination of Hebert's COI primers revealed that they match the COI domain of the *Wolbachia* genome (AE017196) at nearly every nucleotide position. Hebert's COI primers were originally modified from the LCO1490 and HCO2198 COI primers of Folmer et al. (1994). These primers were designed for 11 invertebrate phyla and have also been shown to amplify the *Wolbachia* genome (Deans et al., 2006). This result underscores a major problem with the widespread application of universal primers for DNA barcoding i.e. non-specific species amplification. In this case, the extreme conservation of the COI barcode primer region inadvertently led to the amplification of homologous sequence across very distantly related taxa, despite the fact that these primers were designed to be "lepidopteran-specific" (Hebert et al., 2004a,b). In light of this, we strongly recommend that whenever feasible, tissues containing high *Wolbachia* copy number should be avoided (Dobson et al., 1999) and that COI barcoding primers be designed from the taxonomic group of interest. Designing specific primers for every focal group of interest would nevertheless present significant technical challenges to the global barcoding initiative, particularly in groups where the taxonomy is poorly understood.

Adopting a universal primer approach may also lead to the unintended amplification of Numts (see Bensasson et al., 2001), and lead to species misidentification and/or over-estimation of species number (Song et al., 2008). Conversely, over-estimation of intra-specific genetic distances due to the inadvertent incorporation of Numts might also under-estimate species number. Although we did not find evidence for Numt contamination in this dataset, as evidenced by the absence of stop codons and/or frame-shifts and congruence between mitochondrial and nuclear data, it is difficult to rule out the possibility that recent translocations may still be present. To safeguard against Numt contamination, we propose adopting the precautions proposed by Song et al. (2008) for barcoding studies, including the validation of species identification with complementary data.

### 4.2. DNA barcoding

Findings from this study provide support for the taxonomic recognition of several geographically widespread morphologically based species of *Heteropsis*: namely *H. pauper, H. subsimilis, H. exocellata* and *H. turbata.* However, we would like to draw attention to several caveats in the interpretation of our data. Firstly, although all the original candidate species satisfy barcoding criteria in being monophyletic, not all have strong statistical support. Branch support is a critical first step in species identification and should not be overlooked in barcoding efforts (Brower, 2006), as has been the case in the past (e.g. Hebert et al., 2004a). Secondly, not all candidate taxa exhibit 10× or greater among versus within species genetic distances, leading us to reject some species designations based on distance alone. The so-called 10× threshold criterion has not been widely applied, perhaps partly due to the difficulty of establishing a consistent threshold for delimiting species. Both theoretical (e.g. Hickerson et al., 2006) and empirical (e.g. Meyer and Paulay, 2005; Meier et al., 2006) studies argue that a fixed cut-off could lead to substantial error in species identification, especially where taxonomic sampling is incomplete and/or species are only recently derived, leading us to question the universality of such an approach. It is well known that gene trees are not necessarily congruent with species trees, especially in recently diverged taxa, due to retention of shared ancestral polymorphisms (Funk and Omland, 2003; Pamilo and Nei, 1988), gene flow (Slatkin and Maddison, 1989) and poor taxonomy (Meyer and Paulay, 2005). Therefore, it is unclear how barcoding can be used to resolve the taxonomic status of recently diverged species without the addition of supplementary morphological, behavioral or ecological data. Lastly, results here show that uncritical application of universal primers can lead to inadvertent amplification of the endosymbiont *Wolbachia*. In order to avoid the risk of co-amplification of non-target species, it might therefore be necessary to a redesign primers appropriately. Recent studies also indicate that lateral gene transfer events from the *Wolbachia* genome into multicellular eukaryotic hosts can also occur (Hotopp et al., 2007) and may lead to identical barcodes in morphologically distinct species (Whitworth et al., 2007). These events have been observed in established public databases and may also confound routine COI analysis if their presence is overlooked.

### 4.3. Comparative phylogeography

Analyses of COI variation also support the role of historical forest fragmentation in shaping recent differentiation at the population level. However, there is little if any statistical support for historically fragmented populations as distinct clades, consistent with their recent isolation. In four out of seven of the older forest fragments where more than one individual/species was sampled (Ankazomivady, Ambohitantely, Binara and Tsitery), populations were fixed for a unique haplotype. The small size of these fragments (0.28–41 $km^2$) suggests that drift has likely driven these populations to fixation. In contrast, *H. subsimilis* and *H. turbata* populations sampled from the forest fragment of Kalambatritra, each harbor two unique haplotypes, consistent with the current larger area of this site (~141 $km^2$). Haplotype sharing between older fragments and contiguous forest sites was observed in only one instance, the Anjozorobe sample of *H. pauper*. Further sampling may help to resolve whether this site, which constitutes part of a major arm of the eastern forest block that runs along the Angavo massif, was more recently fragmented than supposed.

Comparative phylogeographic analyses of geographically widespread taxa can be a powerful tool for disentangling potential mechanisms of evolutionary diversification. In Madagascar, mtDNA analyses of mouse lemurs (Heckman et al., 2007; Yoder et al., 2000; Yoder and Heckman, 2006), frogs (Vences et al., 2004; Vieites et al., 2006), reptiles (Boumans et al., 2007) and tenrecs (Olson et al., 2004) suggest that a phylogeographic division exists between northern and southern populations. This pattern might be related to the topography of the island (Vences et al., accepted for publication; Wollenberg et al., 2008), and contrasts sharply with the observed east-west bioclimatic division thought

to divide many wide-ranging species across the island (e.g. Andreone et al., 2000; Stothard et al. 2001). Rivers are also thought to have played an important role as barriers to gene flow in lemurs (Pastorini et al., 2003) and this may be true for other species such as the poison frog *Mantella bernhardi* (Vieites et al., 2006) and several reptiles (Boumans et al. 2007). In contrast, other studies suggest a role of both isolation by distance and past historical refugia in explaining patterns of genetic variation in a rainforest tree (Andrianoelina et al., 2006). Elevational gradients might also be important in structuring diversity (e.g. Hall, 2005; Lees et al., 1999) but remain little tested at the molecular level in Malagasy fauna (but see Olson et al., 2004). Surprisingly, findings from the best sampled taxon in this study (*H. pauper*) suggest neither a pronounced north-south phylogeographic division nor an effect of riverine barriers, although sampling of more southerly contiguous forest sites is required to refute or support such hypotheses. Similarly, results suggest that within the eastern/northern contiguous forest, butterfly populations are relatively unstructured. However, localized differentiation is evident in several species sampled from isolated central Plateau fragments. In order to gain a better understanding of the potential mechanisms underlying genetic diversity, further work should be directed towards broader taxonomic comparisons of Malagasy butterflies with contrasting ecological associations and dispersal capabilities.

### 4.4. Wolbachia infection

Findings from this study also demonstrate widespread infection of butterfly species with the endosymbiont *Wolbachia*. Butterflies in almost all cases are infected with the supergroup A strain although in one case an unrelated supergroup B strain was identified. Interestingly this sequence is almost identical to that identified in the fly parasitoid *Pseudacteon* spp., suggesting a potential mechanism of inter-species transfer (Dedeine et al., 2005). Although we have used existing supergroup affiliations to classify *Wolbachia* strains in this study (Zhou et al., 1998), widespread recombination within the *wsp* gene precludes their definitive identification (Baldo et al., 2005; Baldo and Werren, 2007). We note however that sequences obtained from *Heteropsis* and *Saribia* share 98–99% identity with other well established *Wolbachia* reference strains from *Drosophila* and other insects. Interestingly, previous studies have shown that *Wolbachia* infection may bias patterns of mitochondrial diversity (Narita et al., 2006; Shoemaker et al., 2004; Whitworth et al., 2007) and potentially mislead phylogeographic inference (Hurst and Jiggins, 2005). We see no obvious association between host lineages and *Wolbachia* infections but future work should compare the prevalence and type of infection in tandem with ongoing phylogeographic analyses to rigorously assess the extent to which endosymbiont infection biases mitochondrial diversity.

### 4.5. Conservation implications

The COI data provide novel insights into the possible role of historical forest fragmentation in evolutionary diversification of butterflies in Madagascar. In addition to climate-induced historical fragmentation, Madagascar has experienced a massive (40.4%) deforestation due to anthropogenic causes. This loss in forest cover has occurred over the past 50 years and leads to an estimates 9.1% commitment to extinction of a representative set of 2243 species that included 297 species of butterflies (Allnutt et al. 2008). The new distributional data here have already been integrated into the national biodiversity database REBIOMA which was recently used to help prioritize the ongoing expansion of the Malagasy national park system, now nearing completion (Kremen et al., 2008). More specifically, we highlight here that certain forest fragments, such as Ankazomivady, have thus far been largely ignored on the basis of reserve planning, perhaps because they were thought to be peripheral parts of the eastern forest block. These fragments not only hold unique phenotypic but also genetic diversity, whose preservation could hold clues to the effects of past climate change on evolutionary diversification.

### 4.6. Conclusions

In summary, mtDNA sequence data reinforced some but not all species identified from morphological data. Nuclear data also provide additional support for those taxa identified by DNA barcoding criteria and morphology. However, for recently diverged taxa such as *H. pauper*_23 and *H*. 25, straightforward application of barcoding rules may prove problematic. Generalized primers also led to the inadvertent amplification of non-target *Wolbachia* sequences. This subsequently alerted us to the widespread presence of this bacterium across the host species tested in this study and suggest that future arthropod barcoding surveys should include *Wolbachia* detection assays alongside mitochondrial and nuclear sequencing efforts. Finally, COI data also provide novel insights into the possible role of historical forest fragmentation and may prove useful in the identification of priority areas for conservation in the expansion of the Malagasy national park system (Kremen et al., 2008).

## Acknowledgments

We would like to thank the Wildlife Conservation Society, numerous local guides and the following individuals for assistance with field logistics and collecting specimens in Madagascar: Claire Kremen, Louise Holloway, Tiana Raharitsimba, Ranaivosolo Ravomiarana. Madame Emilienne Razafimatatra and Madame Olga Ramilijaona assisted with obtaining permits. Tom Allnutt kindly supplied the forest shape coverages. Laboratory work was funded through the NSF DEB 0405263 award to the University of New Orleans Undergraduate Mentoring in Environmental Biology (UMEB) program. Fieldwork was funded by: BBSRC (including BBS/B/04358), Leverhulme Trust F/696/I and National Geographical Society #7332-02.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2008.11.008.

## References

Abascal, F., Posada, D., Knight, R.D., Zardoya, R., 2006. Parallel evolution of the genetic code in arthropod mitochondrial genomes. PLoS Biol. 4, e127.

Allnutt, T.F., Ferrier, S., Manion, G., Powell, G.V.M., Ricketts, T.H., Fisher, B.L., Harper, G.J., Irwin, M.E., Kremen, C., Labat, J.-N., Lees, D.C., Pearce, T.A., Rakotondrainibe, F., 2008. A method for quantifying biodiversity loss and its application to a 50-year record of deforestation across Madagascar. Conserv. Lett. 1, 173–181.

Andreone, F., Vences, M., Guarino, F.M., Aprea, G., 2000. Natural history and larval morphology of *Boophis occidentalis* (Anura: Mantellidae: Boophinae) provide new insights into the phylogeny and adaptive radiation of endemic Malagasy frogs. J. Zool. 257, 425–438.

Andrianoelina, O., Rakotondraoelina, H., Ramamonjisoa, L., Maley, J., Danthu, P., Bouvet, J.-M., 2006. Genetic diversity of *Dalbergia monticola* (Fabaceae) an endangered tree species in the fragmented oriental forest of Madagascar. Biodivers. Conserv. 15, 1109–1128.

Avise, J.C., 2004. Molecular Markers, Natural History, and Evolution, second ed. Sinauer Associates, Sunderland, Massachusetts. p. 684.

Baldo, L., Dunning Hotopp, J.C., Jolley, K.A., Bordenstein, S.R., Biber, S.A., Choudhury, R.R., Hayashi, C., Maiden, M.C.J., Tettelin, H., Werren, J.H., 2006. Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. Appl. Environ. Microbiol. 72, 7098–7110.

Baldo, L., Werren, J.H., 2007. Revisiting *Wolbachia* supergroup typing based on wsp: spurious lineages and discordance with MLST. Curr. Microbiol. 55, 81–87.

Bensasson, D., Zhang, D.-X., Hartl, D.L., Hewitt, G.M., 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol. Evol. 16, 314–321.

Bermingham, E., Moritz, C., 1998. Comparative phylogeography: concepts and applications. Mol. Ecol. 7, 367–369.

Boumans, L., Vieites, D.R., Glaw, F., Vences, M., 2007. Geographical patterns of deep mitochondrial differentiation in widespread Malagasy reptiles. Mol. Phylogen. Evol. 45, 822–839.

Braig, H.R., Zhou, W., Dobson, S.L., O'Neill, S.L., 1998. Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. J. Bacteriol. 180, 2373–2378.

Brower, A.V.Z., DeSalle, R., 1998. Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: the utility of wingless as a source of characters for phylogenetic inference. Insect Mol. Biol. 7, 73–82.

Brower, A.V.Z., 2006. Problems with DNA barcodes for species delimitation: 'ten species' of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae). Syst. Biodiv., 127–132.

Caterino, M.S., Cho, S., Sperling, F.A.H., 2000. The current state of insect molecular systematics: a thriving tower of Babel. Ann. Rev. Entomol. 45, 1–54.

Cameron, S., Rubinoff, D., Will, K., 2006. Who will actually use DNA Barcoding and what will it cost? Syst. Biol. 55, 844–847.

Charlat, S., Nirgianaki, A., Bourtzis, K., Mercot, H., 2002. Evolution of *Wolbachia*-induced cytoplasmic incompatibility in *Drosophila simulans* and *D. sechellia*.. Evolution 56, 1735–1742.

Charlat, S., Engelstadter, J., Dyson, E.A., Hornett, E.A., Duplouy, A., Tortosa, P., Davies, N., Roderick, G.K., Wedell, N., Hurst, G.D.D., 2006. Competing selfish genetic elements in the butterfly *Hypolimnas bolina*. Curr. Biol. 16, 2453–2458.

Clement, M., Posada, D., Crandall, K., 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9, 1657–1660.

Cowan, W.D., 1882. Geographical excursions in the Betsileo, Tanala and Bara countries. Proc. R. Geogr. Soc. 5, 21–37.

d'Abrera, B.L., 1997. Butterflies of the Afrotropical Region, Part 1. Papilionidae, Pieridae, Nymphalidae, Satyridae, second ed. Hill House, London.

Dean, M.D., Ballard, K.J., Glass, A., Ballard, J.W.O., 2003. Influence of two *Wolbachia* strains on population structure of east African *Drosophila simulans*. Genetics 165, 1959–1969.

Deans, A.R., Gillespie, J.J., Yoder, M.J., 2006. An evaluation of ensign wasp classification (Hymenoptera: Evaniidae) based on molecular data and insights from ribosomal RNA secondary structure. Syst. Entomol. 31, 517–528.

Dedeine, F., Ahrens, M., Calcaterra, L., Shoemaker, D.D., 2005. Social parasitism in fire ants (*Solenopsis* spp.): a potential mechanism for interspecies transfer of *Wolbachia*. Mol. Ecol. 14, 1543–1548.

Dobson, S.L., Bourtzis, K., Braig, H.R., Jones, B.F., Zhou, W., Rousset, F., O'Neill, S.L., 1999. *Wolbachia* infections are distributed throughout insect somatic and germ line tissues. Insect Biochem. Mol. Biol. 29, 153–160.

Ekrem, T., Willassen, E., Stur, E., 2007. A comprehensive DNA sequence library is essential for identification with DNA barcodes. Mol. Phylogen. Evol. 43, 530–542.

Elias, M., Hill, R.I., Willmott, K.R., Dasmahapatra, K.K., Brower, A.V., Mallet, J., Jiggins, C.D., 2007. Limited performance of DNA barcoding in a diverse community of tropical butterflies. Proc. R. Soc. Lond. Ser. B: Biol. Sci. 274, 2881–2889.

Evans, K.M., Wortley, A.H., Mann, D.G., 2007. An assessment of potential diatom "barcode" genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in *Sellaphora* (Bacillariophyta). Protist 158, 349–364.

Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evol. Bioinform. Online 1, 47–50.

Fisher, .L., Robertson, H.G., 2002. Comparison and origin of forest and grassland ant assemblages in the high plateau of Madagascar (Hymenoptera: Formicidae). Biotropica 34, 155–167.

Folmer, O., Black, M., Hoen, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of the mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Sel. Mol. Mar. Biol. Biotech. 3, 294–299.

Funk, D.J., Omland, K.E., 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu. Rev. Ecol. Syst. 34, 97–423.

Gavotte, L., Henri, H., Southamer, R., Charif, D., Charlat, S., Boulétreau, M., Vavre, F., 2007. A survey of the bacteriophage WO in the endosymbiotic bacteria *Wolbachia*. Mol. Biol. Evol. 24, 427–435.

Green, G., Sussman, R., 1990. Deforestation history of the eastern rain forests of Madagascar from satellite images. Science 248, 212–215.

Guindon, S., Gascuel, O., 2003. PhyML—A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwachs, W., Hebert, P.D.N., 2006. DNA barcoding distinguishes species of tropical Lepidoptera. Proc. Natl. Acad. Sci. USA 103, 971–978.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41, 95–98.

Hall, J.P.W., Harvey, D.J., 2002. A survey of androconial organs in the Riodinidae (Lepidoptera). Zool. J. Linn. Soc. 136, 171–197.

Hall, J.P.W., 2005. Montane speciation patterns in *Ithomiola* butterflies (Lepidoptera: Riodinidae): are they consistently moving up in the world? Proc. R. Soc. Lond. Ser. B: Biol. Sci. 272, 2457–2466.

Harper, G.J., Steininger, M.K., Tucker, C.J., Juhn, D., Hawkins, F., 2007. Fifty years of deforestation and forest fragmentation in Madagascar. Environ. Conserv. 34, 325–333.

Hebert, P.D.N., Cywinska, A., Ball, S.L., DeWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. Ser. B: Biol. Sci. 270, 313–321.

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., Hallwachs, W., 2004a. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc. Natl. Acad. Sci. USA 101, 14812–14817.

Hebert, P.D.N., Stoeckle, M.Y., Zemlack, T.S., Francis, C.M., 2004b. Identification of birds through DNA barcodes. PLoS Biol. 2, 1657–1663.

Heckman, K.L., Mariani, C.L., Rasoloarison, R., Yoder, A.D., 2007. Multiple nuclear loci reveal patterns of incomplete lineage sorting and complex species history within western mouse lemurs (*Microcebus*). Mol. Phylogen. Evol. 43, 353–367.

Hickerson, M.J., Meyer, C.P., Moritz, C., 2006. DNA Barcoding will often fail to discover new animal species over broad parameter space. Syst. Biol. 55, 729–739.

Hotopp, J.C., Clark, M.E., Oliveira, D.C.S.G., Foster, J.M., Fischer, P., Muñoz Torres, M.C., Giebel, J.D., Kumar, N., Ishmael, N., Wang, S., Ingram, J., Nene, R.V., Shepard, J., Tomkins, J., Richards, S., Spiro, D.J., Ghedin, E., Slatko, B.E., Tettelin, H., Werren, J.H., 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science 317, 1753–1756.

Humbert, H., Cours-Darne, G., 1965. Carte internationale du tapis vegetal et des conditions ecologiques. 3 coupures au 1/1,000,000 de Madagascar. Trav. Sect. Sci. Tech. Inst. Franç. Pondichéry.

Hurst, G.D.D., Jiggins, F.M., 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proc. R. Soc. Lond. Ser. B: Biol. Sci. 272, 1525–1534.

Jeyaprakesh, A., Hoy, M.A., 2000. Long PCR improves *Wolbachia* DNA amplification: *wsp* sequences found in 76% if sixty-three arthropod species. Insect Mol. Biol. 9, 393–405.

Jiggins, F.M., 2003. Male-killing *Wolbachia* and mitochondrial DNA: selective sweeps, hybrid introgression and parasite population dynamics. Genetics 164, 5–12.

Kremen, C., 1994. Biological inventory using target taxa: a case study of the butterflies of Madagascar. Ecol. Appl. 4, 407–422.

Kremen, C., Cameron, A., Moilanen, A., Phillips, S., Thomas, C.D., Beentje, H., Dransfield, J., Fisher, B.L., Glaw, F., Good, T.C., Harper, G.J., Hijmans, R.J., Lees, D.C., Louis, E., Nussbaum, R.A., Raxworthy, C.J., Razafimpahanana, A., Schatz, G.E., Vences, M., Vieites, D.R., Wright, P.C., Zjhra, M.L., 2008. Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. Science 320, 222–225.

Kubbinga, B., 2006. Evo-devo of sexually selected traits in *Heteropsis iboina* (Lepidoptera, Satyrinae, Mycalesina). Comparative study of eyespot development and description of candidate male sex pheromones. M.Sc. thesis, University of Leiden, Leiden.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief. Bioinform. 5, 150–163.

Leaché, A.D., Reeder, T.W., 2002. Molecular systematics of the Eastern Fence Lizard (*Sceloporus undulatus*): a comparison of parsimony, likelihood, and Bayesian approaches. Syst. Biol. 51, 44–68.

Lees, D.C., 1997. Systematics and biogeography of Madagascan mycalesine butterflies (Lepidoptera: Satyrinae). Ph.D. thesis, University of London, London.

Lees, D.C., Kremen, C., Andriamampianina, L., 1999. A null model for species richness gradients: bounded range overlap of butterflies and other rainforest animals in Madagascar. Biol. J. Linn. Soc. 67, 529–584.

Lees, D.C., Kremen, C., Raharitsimba, H., 2003. Classification, diversity and endemism of the butterflies (Papilionoidea and Hesperioidea): a revised species checklist. In: Goodman, S.M., Benstead, J.P. (Eds.), The Natural History of Madagascar. University of Chicago Press, Chicago, pp. 762–793.

Mateos, M., Caastrezana, S.J., Nankivell, B.J., Estes, A.M., Markow, T.A., Moran, N.A., 2006. Heritable endosymbionts of *Drosophila*. Genetics 174, 363–376.

Meier, R., Kwong, S., Vaidya, G., Ng, P.K.L., 2006. DNA Barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Syst. Biol. 55, 715–728.

Meyer, C.P., Paulay, G., 2005. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol. 3, 2229–2238.

Miller, W.J., Riegler, M., 2006. Evolutionary dynamics of wAu-like *Wolbachia* variants in neotropical Drosophila spp. Appl. Environ. Microbiol. 72, 826–835.

Monteiro, A., Pierce, N., 2001. Molecular phylogeny of *Bicyclus* butterflies (Satyridae) using COI, COII and EF1α. Mol. Phylogen. Evol. 18, 264–281.

Moritz, C., Cicero, C., 2004. DNA Barcoding: Promise and Pitfalls. PLoS Biol. 2, e354.

Narita, S., Nombra, M., Kato, Y., Fukatsu, T., 2006. Genetic structure of sibling butterfly species affected by *Wolbachia* infection sweep: evolutionary and biogeographical implications. Mol. Ecol. 15, 1095–1108.

Neigel, J.E., Domingo, A., Stake, J., 2007. DNA barcoding as a tool for coral reef conservation. Coral Reefs 26, 487–499.

Olson, L.E., Goodman, S., Yoder, A.D., 2004. Illumination of cryptic species boundaries in long-tailed shrew tenrecs (Mammalia: Tenrecidae; Microgale) provides insights into geographic variation and distributional constraints. Biol. J. Linn. Soc. 83, 1–22.

Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5, 568–583.

Paraskevopoulos, C., Bordenstein, S.R., Wernegreen, J.J., Werren, J.H., Bourtzis, K., 2007. Toward a *Wolbachia* multilocus sequence typing system: Discrimination of *Wolbachia* strains present in *Drosophila* species. Curr. Microbiol. 53, 388–395.

Pastorini, J., Thalmann, U., Martin, R.D., 2003. A molecular approach to comparative phylogeography of extant Malagasy lemurs. Proc. Natl. Acad. Sci. USA 100, 5879–5884.

Paulian, R., 1951. Études sur les lépidoptères malagaches. Nouveaux satyrides (1). Mém. Inst. Scient. Madagascar (A) 6, 387–394.

Peña, C., Wahlberg, N., Weingartner, E., Kodandaramaiah, U., Nylin, S., Freitas, A., Brower, A.V.Z., 2006. Higher level phylogeny of Satyrinae butterflies (Lepidoptera: Nymphalidae) based on DNA sequence data. Mol. Phylogen. Evol. 40, 29–49.

Peña, C., Wahlberg, N., 2008. Prehistorical climate change increased diversification of a group of butterflies. Biol. Lett. 4, 274–278.

Perotti, M.A., Catala, S.S., Ormeno, A.V., Zelazowska, M., Bilinski, S.M., Braig, H.R., 2004. The sex ratio distortion in the human head louse is conserved over time. BMC Genet. 5, 10.

Rambaut, A., Drummond, A.J., 2007. Tracer v1.4. Available from: <http://beast.bio.ed.ac.uk/Tracer>.

Ratnasingham, S., Hebert, P.D.N., 2007. BOLD: the barcode of life data system. Mol. Ecol. Notes 7, 355–364.

Riegler, M., Stauffer, C., 2002. *Wolbachia* infections and superinfections in cytoplasmically incompatible populations of the European cherry fruit fly *Rhagoletis cerasi* (Diptera, Tephritidae). Mol. Ecol. 11, 2425–2434.

Roe, A.D., Sperling, F.A.H., 2007. Patterns of evolution of mitochondrial cytochrome *c* oxidase I and II DNA and implications for DNA barcoding. Mol. Phylogen. Evol. 44, 325–345.

Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Rubinoff, D., 2006. DNA barcoding evolves into the familiar. Conserv. Biol. 20, 1548–1549.

Sambrook, J., Russell, D.W., 2001. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Shoemaker, D.D., Dyer, K.A., Ahrens, M., McAbee, K., Jaenike, J., 2004. Decreased diversity but increased substitution rate in host mtDNA as a consequence of *Wolbachia* endosymbiont infection. Genetics 168, 2049–2058.

Slatkin, M., Maddison, W.P., 1989. A cladistic measure of gene flow inferred from the phylogeny of alleles. Genetics 123, 603–661.

Stothard, J.R., Brémond, P., Andriamaro, A., Sellin, B., Sellin, E., Rollinson, D., 2001. *Bulinus* species on Madagascar: molecular evolution, genetic markers and compatibility with *Schistosoma haematobium*. Parasitology 123, 261–275.

Song, H., Buhay, J.E., Whiting, M.F., Crandall, K.A., 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proc. Natl. Acad. Sci. USA 105, 13486–13491.

Swofford, D., 2000. PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods), Version 4.0b2. Sinauer Associates, Sunderland, MA.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Torres, E., Lees, D.C., Vane-Wright, R.I., Kremen, C., Leonard, J.A., Wayne, R.K., 2001. Examining monophyly of a large radiation of Madagascan butterflies (Lepidoptera: Satyrinae: Mycalesina) based on mitochondrial DNA sequence data. Mol. Phylogen. Evol. 20, 460–473.

Vaishampayan, P.A., Dhotre, D.P., Rakeshkumar, P.G., Lalwani, P., Ghate, H., Patole, M.S., Shouche, Y.S., 2007. Molecular evidence and phylogenetic affiliations of *Wolbachia* in cockroaches. Mol. Phylogen. Evol. 44, 1346–1351.

Vences, M., Chiari, Y., Raharivololoniaina, L., Meyer, A., 2004. High mitochondrial diversity within and among populations of Malagasy poison frogs. Mol. Phylogen. Evol. 30, 295–307.

Vences, M., Wollenberg, K.C., Lees, D.C., Vieites, D.R. Accepted for publication. Madagascar as a model region of species diversification. Trends Ecol. Evol..

Vieites, D.R., Chiari, Y., Vences, M., Andreone, F., Rabemananjara, F., Bora, P., Nieto-Román, S., Meyer, A., 2006. Mitochondrial evidence for distinct phylogeographic units in the endangered Malagasy poison frog *Mantella bernhardi*. Mol. Ecol. 15, 1617–1625.

Wernersson, R., Pedersen, A.G., 2003. RevTrans—Constructing alignments of coding DNA from aligned amino acid sequences. Nucleic Acids Res. 31, 3537–3539.

Whitworth, T.L., Dawson, R.D., Magalon, H., Baudry, E., 2007. DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). Proc. R. Soc. Lond. Ser. B: Biol. Sci. 274, 1731–1739.

Wiens, J.J., Penkrot, T.L., 2002. Delimiting species based on DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). Syst. Biol. 51, 69–91.

Will, K.W., Mishler, B.D., Wheeler, Q.D., 2005. The perils of DNA barcoding and the need for integrative taxonomy. Syst. Biol. 54, 844–851.

Wollenberg, K.C., Vieites, D.R., van der Meijden, A., Glaw, F., Cannatella, D.C., Vences, M., 2008. Patterns of endemism and species richness in Malagasy cophyline frogs support a key role of mountainous areas for speciation. Evolution 62, 1890–1907.

Yoder, A.D., Heckman, K., 2006. Mouse lemur phylogeography revises a model of ecogeographic constraint in Madagascar. In: Fleagle, J., Lehman, S.M. (Eds.), Primate Biogeography: Progress and Prospects. Springer Press, New York, pp. 255–268.

Yoder, A.D., Rasoloarison, R., Goodman, S., Ganzhorn, J.S., Irwin, J.A., Atsalis, S., Ravosa, M.J., Ganzhorn, J.U., 2000. Remarkable species diversity in Malagasy mouse lemurs (Primates, Microcebus). Proc. Natl. Acad. Sci. USA 97, 11325–11330.

Zhou, W., Rousset, F., O'Neill, S., 1998. Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. Proc. R. Soc. Lond. Ser. B: Biol. Sci. 265, 509–515.