MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2010) **10**, 60–68 doi: 10.1111/j.1755-0998.2009.02724.x

DNA BARCODING

Tissue segregation of mitochondrial haplotypes in heteroplasmic Hawaiian bees: implications for DNA barcoding

KARL N. MAGNACCA* and MARK J. F. BROWN*†
*Department of Zoology, School of Natural Sciences, Trinity College Dublin, Dublin 2, Ireland, †School of Biological Sciences, Royal Holloway University of London, Egham, Surrey TW20 0EX, UK

Abstract

The issue of mitochondrial heteroplasmy has been cited as a theoretical problem for DNA barcoding but is only beginning to be examined in natural systems. We sequenced multiple DNA extractions from 20 individuals of four Hawaiian *Hylaeus* bee species known to be heteroplasmic. All species showed strong differences at polymorphic sites between abdominal and muscle tissue in most individuals, and only two individuals had no obvious segregation. Two specimens produced completely clean sequences from abdominal DNA. The fact that these differences are clearly visible by direct sequencing indicates that substantial intra-individual mtDNA diversity may be overlooked when DNA is taken from small tissue fragments. At the same time, differences in haplotype distribution among individuals may result in incorrect recognition of cryptic species. Because DNA barcoding studies typically use only a small fragment of an organism, they are particularly vulnerable to sequencing bias where heteroplasmy and haplotype segregation are present. It is important to anticipate this possibility prior to undertaking large-scale barcoding projects to reduce the likelihood of haplotype segregation confounding the results.

Keywords: Apoidea, DNA barcoding, heteroplasmy, *Hylaeus*, tissue segregation

Received 9 February 2009; revision received 10 April 2009; accepted 17 April 2009

Introduction

The key assumption behind the use of DNA sequences for identification is that only a single sequence occurs in an individual. Indeed, this is promoted as one of the advantages of using DNA over morphology, as it allows one to associate morphologically different life stages and identify fragmentary material that would otherwise be impossible to recognize (Barber & Boyce 2006; Webb *et al.* 2006; Costa *et al.* 2007). In the debate over the utility and flaws of DNA barcoding, most of the attention given to problems of multiple sequences has been focused on nuclear copies of mitochondrial sequences (numts; Song *et al.* 2008). These occur when fragments of mtDNA – which can range from small segments of genes to significant portions of the entire genome – are duplicated and inserted into the nuclear genome as non-functional units. They can be numerous and widespread (Zhang & Hewitt 1996; Bensasson *et al.* 2000; Pamilo *et al.* 2007), but are usually recognizable as pseudogenes because of the presence of nonsense mutations such as stop codons, insertions and deletions in coding sequences. Heteroplasmy – the coexistence of multiple mitochondrial haplotypes in a single individual – is often mentioned as a potential problem (Rubinoff *et al.* 2006; Song *et al.* 2008), but is infrequently documented. If present, however, it can cause serious problems for identification.

In humans, heteroplasmy is well-known and intensively studied because of its link with mitochondrial disease (Chinnery & Turnbull 2000; Ballana *et al.* 2008). In other metazoans, particularly insects, length polymorphism in the control region is documented in a variety of groups (Boyce *et al.* 1989; Zhang & Hewitt 1997; Nardi *et al.* 2001). Aside from unusual cases such as the *Mytilus* mussels, which have a modified inheritance mechanism

Correspondence: Karl N. Magnacca, Fax: +353 (0)1 6778094; E-mail: knm956@gmail.com

(Zouros *et al.* 1994), sequence polymorphism has been reported from a wide variety of non-human metazoans but relatively few specific examples (Satta *et al.* 1988; Walton & Butlin 1997; García *et al.* 2003; Frey & Frey 2004; Magnacca & Danforth 2006). As fixation or elimination of mitochondrial changes has been documented to occur rapidly within individuals (Koehler *et al.* 1991), heteroplasmy in wild populations has been considered to be a rare, transient condition (Satta *et al.* 1988; Song *et al.* 2008) or restricted to highly-variable sites such as the control region where spontaneous mutants arise frequently (Boyce *et al.* 1989; Zhang & Hewitt 1997). In both insects (Volz-Lingenhöhl *et al.* 1992) and nematodes (Tsang & Lemire 2002), laboratory strains heteroplasmic for haplotypes with multiple protein-coding genes deleted are able to thrive, although they may suffer from subtle phenotypic effects that would lead to their elimination in the wild (Liau *et al.* 2007). However, the persistence of even these deleterious haplotypes in heteroplasmy over long periods of time indicates that multiple lineages of mitochondria are passed on as a normal part of mitochondrial inheritance, and the 'mitochondrial bottleneck' (Ashley *et al.* 1989) will not inevitably lead to homoplasy in a few generations. Barring recombination, mutations will accumulate independently in separate lineages, leading to increasing sequence divergence unless homoplasmy is achieved through random drift or selection.

The distribution of haplotypes in heteroplasmic individuals is often not uniform, but varies across different tissues (Lécher *et al.* 1994; Chinnery 2002; Liau *et al.* 2007). The causative mechanism for haplotype distribution is complex and uncertain (Battersby & Shoubridge 2001), but the consistency of segregation indicates that certain haplotypes have a selective advantage in some tissues and a disadvantage in others (Jenuth *et al.* 1997). Consequently, the proportion of haplotypes in a given tissue can vary greatly across the life of an individual, even a relatively short-lived one (Jenuth *et al.* 1997; Battersby & Shoubridge 2001). Such segregation may pose a major additional problem for DNA barcoding, particularly with respect to associating different life stages and partial material.

To date, most of the study of heteroplasmy in animals has been conducted on experimentally derived laboratory cultures (Volz-Lingenhöhl *et al.* 1992; Jenuth *et al.* 1997; de Stordeur 1997; Tsang & Lemire 2002), or in relation to relatively infrequent human mitochondrial pathologies (Chinnery 2002). Here, we examine haplotype segregation in heteroplasmic species of endemic Hawaiian bees, *Hylaeus* (*Nesoprosopis*) (Colletidae). Nearly half of the species examined have been found to be heteroplasmic; although most are too rare to be sequenced on a large scale, of 49 species (266 individuals), 18 appear to be consistently heteroplasmic, 28 are consistently homoplasmic and only three species are variable (K. N. Magnacca & M. J. F. Brown, unpublished data). The conclusion of heteroplasmy rather than pseudogene amplification is based on the high proportion of third-position, synonymous polymorphisms (comparable to that found in intraspecific, between-island comparisons in homoplasmic species); the absence of stop codons, indels, and unusual amino acid changes and the clustering of dominant and subdominant haplotypes of each species into a monophyletic group. Heteroplasmy is concentrated on the 'wood-nesting clade' (Magnacca & Danforth 2006), where it occurs in 20 of 27 species. Many species are threatened and rare (U.S. Fish and Wildlife Service 1994; Magnacca 2007) and females pose significant identification problems (Daly & Magnacca 2003), making them a prime target for DNA identification.

Materials and methods

Total DNA was extracted from 10 specimens of *Hylaeus pubescens*, three *Hylaeus andrenoides*, three *Hylaeus muranus*, and four *Hylaeus setosifrons* using the Qiagen DNeasy Blood & Tissue extraction kit, following the manufacturer's protocol, with the exception of three (those with codes beginning with 'C') which were performed using the phenol–chloroform–isoamyl alcohol method (Doyle & Doyle 1990). All specimens were collected by KNM and stored in 95% ethanol. Collection details are given in Table 1. For *H. pubescens*, three extractions were taken from each specimen: thoracic flight muscle, mid and hind right legs, and abdomen. The last was composed primarily of the reproductive organs, sometimes including portions of the gut. DNA was extracted from the legs and abdomen of the remaining species, and in one specimen of each, the gut and reproductive organs were extracted separately (Table 1).

The 'standard' barcoding fragment of cytochrome oxidase I was amplified using a modified version of the commonly-used primer LCO (5'-TATCAACCAATCAT AAAGATATTGG-3') (Folmer *et al.* 1994), with a shorter version of 'Nancy' (Simon *et al.* 1994) as the reverse (5'-CCCGGTAAAATTAAAATATAAAC-3'). PCR was run using standard Taq (Invitrogen Corp.) with the following program: a starting denaturation at 94 °C for 180 s, followed by 35 cycles of 94 °C for 30 s, 48 °C for 45 s, and 72 °C for 60 s, concluding with a final extension at 72 °C for 240 s. PCR products were sequenced using an ABI 3130xl capillary automated sequencer (Applied Biosystems, Inc.), either at the School of Natural Sciences, Trinity College (extractions T155–T195, including all *H. pubescens* sequences) or by Macrogen (Seoul, South Korea). Several extractions were sequenced on both; these showed no significant differences between the two. The PCR primers were used for sequencing, and all were

Table 1 Specimens and extractions for sequencing. Extraction codes beginning with a C were done using the PCI protocol (see Methods)

Specimens					Extractions				
Species	Island	Locality	Date	Ind.	Abdomen	Legs	Thorax	Reprod.	Gut
Hylaeus (*Nesoprosopis*) *andrenoides*	Kauai	Alakai Swamp Trail, 3600 ft	21 August 1999	1	C79	T139			
Hylaeus (*Nesoprosopis*) *andrenoides*	Kauai	Paaiki Valley, 2700 ft	2 July 2000	2		T57		T187	T188
Hylaeus (*Nesoprosopis*) *andrenoides*	Kauai	Mohihi Trail, 4000 ft	8 July 2000	3	T189	T58			
Hylaeus (*Nesoprosopis*) *muranus*	Hawaii	Old Jap. Sch., Volcano, 3750 ft	5 August 2000	1	C175	T140			
Hylaeus (*Nesoprosopis*) *muranus*	Hawaii	Old Jap. Sch., Volcano, 3750 ft	6 August 2000	2		T86		T190	T191
Hylaeus (*Nesoprosopis*) *muranus*	Hawaii	Old Jap. Sch., Volcano, 3750 ft	29 July 2002	3	T192	T87			
Hylaeus (Nesoprosopis) pubescens	Hawaii	Kilauea Field Station, 3800 ft	1 January 1999	1	T156	T155	T157		
Hylaeus (*Nesoprosopis*) *pubescens*	Hawaii	Tree Planting Rd., 4200 ft	5 January 1999	2	T159	T158	T160		
Hylaeus (*Nesoprosopis*) *pubescens*	Hawaii	Hilina Pali Rd., 3300 ft	11 January 1999	3	T162	T161	T163		
Hylaeus (*Nesoprosopis*) *pubescens*	Hawaii	Kilauea Iki, 3900 ft	5 June 1999	4	T165	T164	T166		
Hylaeus (*Nesoprosopis*) *pubescens*	Hawaii	Tree Moulds, 4000 ft	9 August 1999	5	T168	T167	T169		
Hylaeus (*Nesoprosopis*) *pubescens*	Hawaii	Olaa Small Tract, 3800 ft	15 June 2000	6	T171	T170	T172		
Hylaeus (*Nesoprosopis*) *pubescens*	Hawaii	Kilauea Caldera, 3600 ft	29 July 2000	7	T174	T173	T175		
Hylaeus (*Nesoprosopis*) *pubescens*	Hawaii	Olaa, Wright Rd., 3800 ft	30 July 2000	8	T177	T176	T178		
Hylaeus (*Nesoprosopis*) *pubescens*	Hawaii	Kona Forest Unit camp, 5300 ft	2 August 2000	9	T180	T179	T181		
Hylaeus (*Nesoprosopis*) *pubescens*	Hawaii	Earthquake Trail, 4000 ft	30 June 2002	10	T183	T182	T184		
Hylaeus (*Nesoprosopis*) *setosifrons*	Hawaii	Tree Moulds, 4000 ft	9 August 1999	1	C80	T150			
Hylaeus (*Nesoprosopis*) *setosifrons*	Hawaii	Earthquake Trail, 4000 ft	9 August 1999	2	T193	T128			
Hylaeus (*Nesoprosopis*) *setosifrons*	Hawaii	Old Jap. Sch., Volcano, 3750 ft	5 August 2000	3	T185	T129	T186		
Hylaeus (*Nesoprosopis*) *setosifrons*	Hawaii	Kipuka Puaulu, 4000 ft	31 July 2002	4		T130		T194	T195

Ind. = individual, abdomen = mixed reproductive and gut tissue, reprod. = reproductive organs.

sequenced from both directions. The final length of the fragment was 654 bases. Chromatograms were edited using FinchTV (Geospiza Inc.). Dominant sequences were generated using the highest peak at each site; if the peaks were equal in both the forward and reverse chromatograms, the polymorphism was retained. Sequences were submitted to GenBank (accession number(s) FJ411518–20, FJ411696–98, FJ411728, FJ411736–39, FJ656055–97); see electronic supplemental material for details and sequences, trace files and specimen data submitted to BOLD (http://www.barcodinglife.org) as part of the project '*Hylaeus* haplotype segregation'.

Results

All four species showed clear segregation of mitochondrial haplotypes. In *H. pubescens*, the best-sampled species, segregation was strong (i.e., with chromatogram peak heights indicating a high degree of difference in intra-individual haplotype proportions between extractions) in 60% of individuals and moderate in 20%. Among the other species, segregation was strong in 80% of individuals and moderate in 20%. Two distinct categories of relative haplotype abundance were observed for each species, generally corresponding to the abdominal and muscle tissues. These are designated by letters in Table 2 and some of the characteristic differences between them are shown in Fig. 1. Intermediates were also seen in some cases, but overall the extreme forms predominated. Only two extractions – C79 (*H. andrenoides*) and T185 (*H. setosifrons*) – produced sequences that appeared clean (i.e., no secondary peaks above the low background noise level). Both of these were abdominal extractions and the corresponding muscle sequences from these individuals were highly polymorphic, similar to other specimens of the species. Sites that displayed a single peak in one extraction and two codominant peaks in the other from the same individual, or weakly polymorphic sites where the dominant base differed between extractions, were extremely common (Fig. 1).

Sequences from reproductive and gut tissue were identical and muscle sequences from the legs and thorax were generally identical. The exceptions to the latter were three specimens of *H. pubescens*, which all carried amino acid polymorphisms at the same site (serine-glycine in specimen 7, and serine-aspartic acid in specimens 5 and 8). It is noteworthy that these three were also the only *H. pubescens* in which other polymorphisms typical of form 'A' appear in muscle tissue, suggesting that amino acid changes altered the selective balance among haplotypes. Fixed synonymous differences are present between individuals of *H. andrenoides* and *H. setosifrons* in the abdominal extractions (indicated by subtypes 1 and 2 in Table 2) that did not affect haplotype segregation in those species.

Table 2 Distribution of haplotype segregation across tissue types and specimens

Species	Individual	Sex	Abdomen	Legs	Thorax
Hylaeus andrenoides	1	m	C1	D	
H. andrenoides	2	m	C2*	D	
H. andrenoides	3	f	C1	CD	
Hylaeus muranus	1	f	E	F	
H. muranus	2	f	EF*	F	
H. muranus	3	f	E	F	
Hylaeus pubescens	1	f	B	B	B
H. pubescens	3	f	B	B	B
H. pubescens	2	m	AB	B	B
H. pubescens	4	m	A	B	B
H. pubescens	6	f	A	B	B
H. pubescens	9	f	A	B	B
H. pubescens	10	m	A	B	B
H. pubescens	5	m	A	AB	B
H. pubescens	8	f	A	AB	B
H. pubescens	7	f	AB	A	AB
Hylaeus setosifrons	1	m	G1	H	
H. setosifrons	2	m	G1	H	
H. setosifrons	3	f	G1	H	H
H. setosifrons	4	m	G2*	H	

Letters indicate distinct haplotype patterns; those with two letters (e.g. 'AB') are intermediates. Asterisks indicate where reproductive and gut tissue were extracted and sequenced separately (results identical in all cases)

In *H. pubescens*, abdominal samples exhibited a greater degree of polymorphism – more sites were clearly polymorphic and in general, peak heights at polymorphic sites were more balanced. In the other species, the situation was reversed. Identical patterns of heteroplasmy was observed in reproductive tissues of both males and females of all species, ruling out contamination in females by sperm or fertilized eggs containing paternal mtDNA (Meusel & Moritz 1993).

The dominant haplotypes differed significantly between tissue types. Uncorrected genetic distance varied from 1.22% to 4.44% depending on the species, with up to four amino acid differences (Table 3). The closest interspecific distance between dominant haplotypes here was 4.74% with three amino acid differences, between *H. muranus* and *H. pubescens* (both extractions taken from the legs). In comparison, across the Hawaiian *Hylaeus* radiation as a whole, the range of non-heteroplasmic interspecific variation is 2.45–12.71%, while that of intraspecific variation is 0–5.35% (K. N. Magnacca & M. J. F. Brown, unpublished data; one species, *H. anthracinus*, reaches 9.17% intraspecific divergence, but may be composed of three cryptic species). The latter range is as a result of the isolation of island populations, as most intra-island populations are nearly homosequential.

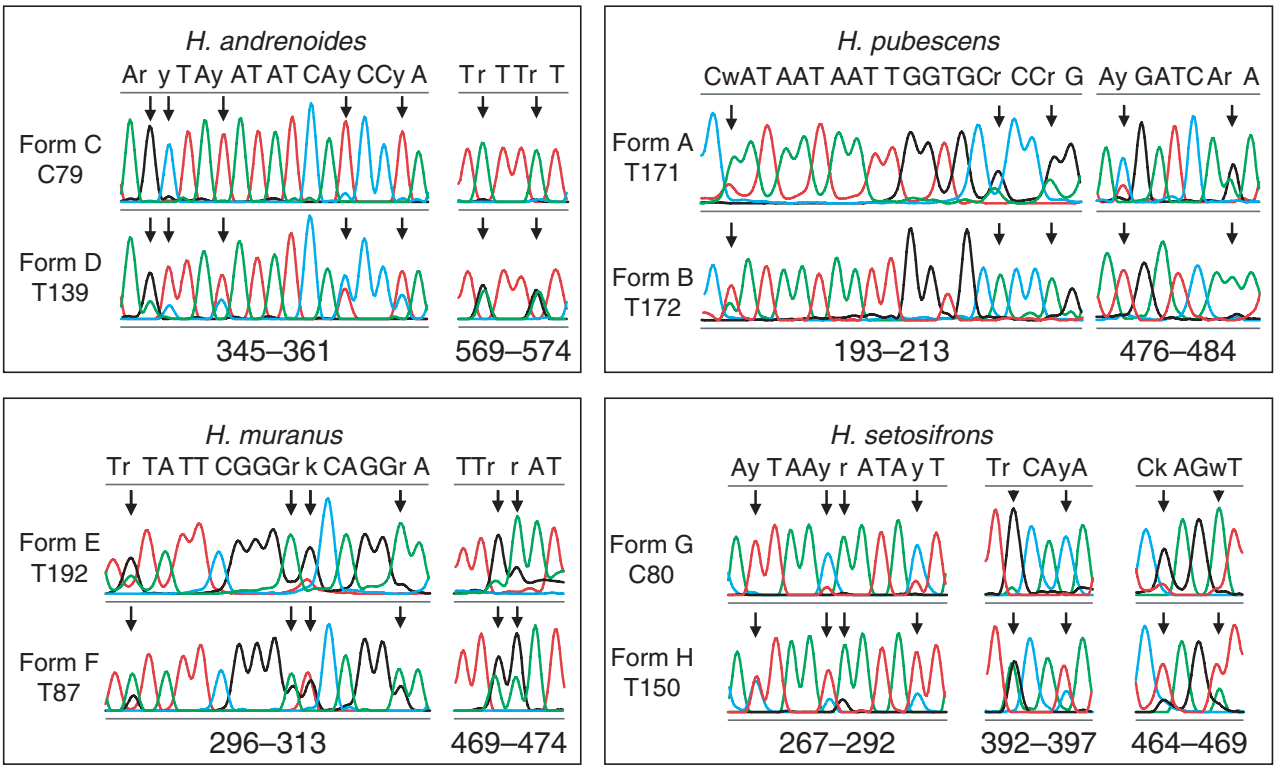

Fig. 1 Chromatogram traces showing tissue differences between abdominal (above) and muscle (below) extractions from single individuals of each species. Arrows denote sites varying between tissues. Number ranges refer to the position of the segment in the 654 bp sequence.

Table 3 Uncorrected (p) distances between dominant haplotypes of the four species examined

Species	Abdomen	Muscle	Amino acid differences	Distance, %
Hylaeus andrenoides	C79	T139	1	3.06
Hylaeus muranus	T192	T87	3	1.22
Hylaeus pubescens	T171	T172	1	3.06
Hylaeus setosifrons	T193	T128	4	4.44

The dominant haplotypes of *H. muranus* are much more similar than those of the other species, yet three of the eight sites where they differ cause amino acid changes. This suggests that the two predominant haplotypes are recently diverged and are segregating based on amino acid differences, while a number of other haplotypes – possibly older and more or less neutral, because they differ less between tissue types – exist at lower proportions in the background.

Discussion

The discovery of strong tissue segregation in a group with species-wide heteroplasmy has serious implications for DNA barcoding. It has generally been assumed that individuals are typically uniform in mtDNA because of uniparental inheritance and the 'mitochondrial bottleneck' (Hebert *et al.* 2003), and any exceptions are evolutionary peculiarities (Zouros *et al.* 1994) or transient states (Satta *et al.* 1988). When heteroplasmy is discussed at all in the context of barcoding, it is implicit that sequencing any portion of an individual obtains a complete picture of its haplotype profile (Rubinoff *et al.* 2006). However, as shown here, this is not necessarily the case. Tissue segregation is a particular problem with regard to 'cryptic species' uncovered by sequencing. None of the specimens here produced two different, clean sequences from different tissues, but it is entirely possible that such a scenario could occur. Two of the present extractions did result in unambiguous sequences; in other groups where tissue-specific selection is stronger, multiple sequences could easily come out completely clean. In Hawaiian *Hylaeus*, the level of heteroplasmic polymorphism overlaps that of interspecific divergence (K. N. Magnacca & M. J. F. Brown, unpublished data) and without further integrative taxonomic analysis, these could incorrectly 'identify' cryptic species.

In all species in the present study, the dominant haplotypes differed by at least one amino acid, suggesting that functional selection is a potential cause for the

varying proportions in different tissues. As the entire mitochondrial genome is generally inherited as a unit without recombination (but see Tsaousis *et al.* 2005), there is no way to be certain which gene(s) is (are) responsible for tissue bias and which is (are) simply accumulating functionally neutral changes. Furthermore, once heteroplasmy is established, different mitochondrial haplotypes may evolve in competition with one another for tissue-specific dominance. The example of *H. pubescens* specimens 5, 7 and 8 may show the start of this, with an amino acid change causing muscle tissue to switch from strictly B to the A or AB haplotype pattern. Clearly, the specimen sample size and gene sequencing coverage are too small at present to say that this particular change is the cause of the shift, but it is nevertheless an intriguing case. It should be noted that mtDNA is highly A/T biased in bees (93.6% at third positions for this fragment across all Hawaiian *Hylaeus*; K. N. Magnacca & M. J. F. Brown, unpublished data) and intra-individual selection for translational efficiency could theoretically occur without or aside from amino acid changes. However, in the fragment analysed here, there is no obvious difference in bias between dominant haplotypes, even though most polymorphisms are transitions. Another factor in haplotype distribution is the nuclear genome: individuals carrying different nuclear alleles are known to exhibit different patterns of haplotype distribution in tissues (Dunbar *et al.* 1995; Battersby *et al.* 2003). It is notable that in the Hawaiian *Hylaeus*, distinct segregation appears to be limited to the *pubescens* species group sampled here. Although multiple extractions from single individuals were not performed on them, a comparison of muscle and abdominal extractions from heteroplasmic species of the *dumetorum* and *connectens* species groups showed no clear evidence of segregation on the level observed in the *pubescens* group.

Haplotype selection is also related to the age of individuals, which is usually uncertain unless specimens are reared or immature. In mice (Jenuth *et al.* 1997) and humans (Chinnery & Turnbull 2000), heteroplasmic haplotype proportions have been demonstrated to change dramatically over a scale of months to years, or even more rapidly when cultured *in vitro* (Battersby & Shoubridge 2001). The ability of mitochondrial haplotypes to shift from low or undetectable levels to dominant or codominant under more favourable conditions may explain the persistence of clear heteroplasmy in *Hylaeus* despite the typically lower levels of polymorphism observed in the reproductive organs. To our knowledge, no evaluation has been made of the segregation time of haplotypes in insects, but there is no reason to believe that it would be shorter than in mice. Many arthropods have relatively long lifespans, up to or exceeding one year, which is more than sufficient for major haplotype shifts to occur. Temporal dynamics pose a second problem for DNA barcoding, as one of its most significant proposed uses is the association of morphologically monotonous or divergent larvae with their more distinctive or familiar adult forms (Barber & Boyce 2006; Webb *et al.* 2006). Although such cases have so far been successful (Greenstone *et al.* 2005), the possibility of changes in haplotype proportions during metamorphosis between larvae and adults – which may involve major morphological reorganization (Snodgrass 1956; Chapman 1998) and/or occur over a period of several years (Barber & Boyce 2006) – adds to the likelihood of overestimation of diversity in heteroplasmic taxa. As the basis of segregation is not understood and is likely to be taxon-specific (or even haplotype-specific), a full accounting of its effects would require more intensive sampling than is currently practical. In addition, where sampling is biased towards older individuals, heteroplasmy may be missed because of greater tissue segregation. The weak segregation exhibited by some specimens in this study and the contrasting clean sequences in others of the same species may be indicative of young and old individuals respectively. It may be possible to approximately age some bees by the wear of pubescence, mandibles, and wings, or by the time of year they are collected in temperate areas. However, the Hawaiian *Hylaeus* exhibit relatively little outward signs of ageing, are not strictly seasonal, and their lifespan is unknown.

The persistence of heteroplasmy may also be related to tissue segregation and even actively selected for. Such a system has been documented in laboratory cultures of the nematode *Caenorhabditis elegans*, where an experimentally-induced heteroplasmic condition is maintained because of forces that cause individuals with biased haplotype distributions to produce offspring with a more balanced ratio (Tsang & Lemire 2002). This is in spite of the fact that the mutant haplotype contains a deletion of four protein-coding and seven tRNA genes; while heteroplasmic worms do not exhibit outward defects, they have lower reproductive fitness that would lead to rapid elimination in nature (Liau *et al.* 2007). Likewise, adaptation of mitochondrial haplotypes to different tissues in *Hylaeus* may be reproductively advantageous to the genomes, but detrimental to the organism as a whole. It is noteworthy that three of the groups in which apparently persistent sequence heteroplasmy is documented in wild taxa – Hawaiian *Hylaeus*, Indonesian *Chitaura* grasshoppers (Walton & Butlin 1997) and Mauritian *Drosophila* (Satta *et al.* 1988) – are island endemic species. In situations where rapid niche partitioning has taken place and competition from other taxa is low, there is a potential to accumulate subtly deleterious mutations (affecting characters such as growth rate,

flight activity in cold weather, etc.) that only become evident when conditions change, as with the arrival of competition from new invasive species. This is compatible with the hypothesis that high rates of amino acid substitution observed in island endemic species are because of relaxed selection (Johnson & Seger 2001; Woolfit & Bromham 2005), and is an interesting subject for future study.

Much of this discussion is hypothetical, because of the limited documentation of heteroplasmy in animals. As shown elsewhere (K. N. Magnacca & M. J. F. Brown, unpublished data), varying levels of polymorphism do not appear to impact seriously the accuracy of identification in Hawaiian *Hylaeus* outside the *fuscipennis* complex and none of the specimens sequenced in this study produced two different, clean sequences. So far, none of the studies on cryptic species in bees has found significant heteroplasmy in the species involved (Danforth *et al.* 1998; Kuhlmann *et al.* 2007; Murray *et al.* 2008; Gibbs 2009). However, as DNA barcoding increases in popularity, it is inevitable that heteroplasmy will be discovered in more and more taxa (White *et al.* 2008). In barcoding studies, DNA is typically extracted from small portions of a single tissue (such as an insect leg or vertebrate blood or skin) to leave the original specimen as complete as possible. Moreover, the tissue used for DNA extraction is often not specified for individual specimens (Ball *et al.* 2005; Cywinska *et al.* 2006) or not mentioned at all (Barrett & Hebert 2005; Monaghan *et al.* 2005; Smith *et al.* 2005; Kaila & Stahls 2006). While leaving intact voucher specimens is valuable for the purposes of integrative taxonomy (Dayrat 2005), it also means that the intra-individual genetic content is poorly sampled. It is important to recognize heteroplasmy and accompanying haplotype segregation as an issue prior to the launch of large-scale barcoding initiatives to establish protocols to control for potential pitfalls down the road. At a minimum, this should include reporting of the body part(s) and/or tissue type used for DNA extraction so that later workers can have some basis for comparison of results. Preferably, barcode reference databases should include sequences from extractions of different tissues, and check multiple extractions from single individuals if discrepancies are found. These precautions will increase the long-term utility of barcode data sets and reduce the possibility of overestimation of species diversity.

Acknowledgements

We wish to thank the Hawaii Division of Forestry and Wildlife and the U.S. National Park Service for permission to collect specimens. This project was funded by Science Foundation Ireland grant EEEOBF131 to MJFB.

References

Ashley MV, Laipis PJ, Hauswirth WW (1989) Rapid segregation of heteroplasmic bovine mitochondria. *Nucleic Acids Research*, **17**, 7325–7331.

Ball SL, Hebert PDN, Burian SK, Webb JM (2005) Biological identifications of mayflies (Ephemeroptera) using DNA barcodes. *Journal of the North American Benthological Society*, **24**, 508–524.

Ballana E, Govea N, de Cid R *et al.* (2008) Detection of unrecognized low-level mtDNA heteroplasmy may explain the variable phenotypic expressivity of apparently homoplasmic mtDNA mutations. *Human Mutation*, **29**, 248–257.

Barber P, Boyce SL (2006) Estimating diversity of Indo-Pacific coral reef stomatopods through DNA barcoding of stomatopod larvae. *Proceedings of the Royal Society B-Biological Sciences*, **273**, 2053–2061.

Barrett RDH, Hebert PDN (2005) Identifying spiders through DNA barcodes. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **83**, 481–491.

Battersby BJ, Shoubridge EA (2001) Selection of a mtDNA sequence variant in hepatocytes of heteroplasmic mice is not due to differences in respiratory chain function or efficiency of replication. *Human Molecular Genetics*, **10**, 2469–2479.

Battersby BJ, Loredo-Osti JC, Shoubridge EA (2003) Nuclear genetic control of mitochondrial DNA segregation. *Nature Genetics*, **33**, 183–186.

Bensasson D, Zhang D-X, Hewitt GM (2000) Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. *Molecular Biology and Evolution*, **17**, 406–415.

Boyce TM, Zwick ME, Aquadro CF (1989) Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. *Genetics*, **123**, 825–836.

Chapman RF (1998) *The Insects: Structure and Function*, 4th edn. Cambridge University Press, Cambridge, UK.

Chinnery PF (2002) Modulating heteroplasmy. *Trends in Genetics*, **18**, 173–176.

Chinnery PF, Turnbull DM (2000) Mitochondrial DNA mutations in the pathogenesis of human disease. *Molecular Medicine Today*, **6**, 425–432.

Costa FO, deWaard JR, Boutillier J *et al.* (2007) Biological identifications through DNA barcodes: the case of the Crustacea. *Canadian Journal of Fisheries and Aquatic Sciences*, **64**, 272–295.

Cywinska A, Hunter FF, Hebert PDN (2006) Identifying Canadian mosquito species through DNA barcodes. *Medical and Veterinary Entomology*, **20**, 413–424.

Daly HV, Magnacca KN (2003) *Insects of Hawaii vol. 17. Hawaiian Hylaeus (Nesoprosopis) Bees (Hymenoptera: Apoidea)*. University of Hawaii Press, Honolulu.

Danforth BN, Mitchell PL, Packer L (1998) Mitochondrial DNA differentiation between two cryptic *Halictus* (Hymenoptera: Halictidae) species. *Annals of the Entomological Society of America*, **91**, 387–391.

Dayrat B (2005) Towards integrative taxonomy. *Biological Journal of the Linnaean Society*, **85**, 407–415.

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus*, **12**, 13–15.

Dunbar DR, Moonie PA, Jacobs HT, Holt IJ (1995) Different cellular backgrounds confer a marked advantage to either

mutant or wild-type mitochondrial genomes. *Proceedings of the National Academy of Sciences of the USA*, **92**, 6562–6566.

Folmer O, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome C oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Frey JE, Frey B (2004) Origin of intra-individual variation in PCR-amplified mitochondrial cytochrome oxidase I of *Thrips tabaci* (Thysanoptera: Thripidae): mitochondrial heteroplasmy or nuclear integration? *Hereditas*, **140**, 92–98.

García BA, Mandfredi C, Fichera L, Segura EL (2003) Variation in mitochondrial 12S and 16S ribosomal DNA sequences in natural populations of *Triatoma infestans* (Hemiptera: Reduviidae). *American Journal of Tropical Medicine and Hygiene*, **68**, 692–694.

Gibbs J (2009) Integrative taxonomy identifies new (and old) species in the *Lasioglossum (Dialictus) tegulare* (Robertson) species group (Hymenoptera, Halictidae). *Zootaxa*, **2032**, 1–38.

Greenstone MH, Rowley DL, Heimbach U, Lundgren JG, Pfannenstiel RS, Rehner SA (2005) Barcoding generalist predators by polymerase chain reaction: carabids and spiders. *Molecular Ecology*, **14**, 3247–3266.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B-Biological Sciences*, **270**, 313–321.

Jenuth JP, Peterson AC, Shoubridge EA (1997) Tissue-specific selection for different mtDNA genotypes in heteroplasmic mice. *Nature Genetics*, **16**, 93–95.

Johnson KP, Seger J (2001) Elevated rates of nonsynonymous substitution in island birds. *Molecular Biology and Evolution*, **18**, 874–881.

Kaila L, Stahls G (2006) DNA barcodes: evaluating the potential of COI to differentiate closely related species of Elachista (Lepidoptera : Gelechioidea : Elachistidae) from Australia. *Zootaxa* **1170**, 1–26.

Koehler CM, Lindberg GL, Brown DR *et al.* (1991) Replacement of bovine mitochondrial DNA by a sequence variant within one generation. *Genetics*, **129**, 247–255.

Kuhlmann M, Else GR, Dawson A, Quicke DLJ (2007) Molecular, biogeographical and phenological evidence for the existence of three western European sibling species in the *Colletes succinctus* group (Hymenoptera: Apidae). *Organisms, Diversity & Evolution*, **7**, 155–165.

Lécher P, Béziat F, Alziari S (1994) Tissular distribution of heteroplasmy and ultrastructural studies of mitochondria from a *Drosophila subobscura* mitochondrial deletion mutant. *Biology of the Cell*, **80**, 25–33.

Liau WS, Gonzalez-Serricchio AS, Deshommes C, Chin K, LaMunyon CW (2007) A persistent mitochondrial deletion reduces fitness and sperm performance in heteroplasmic populations of *C. elegans*. *BMC Genetics*, **8**, 8.

Magnacca KN (2007) Conservation status of the native bees of Hawaii, *Hylaeus* (*Nesoprosopis*) (Hymenoptera: Apoidea). *Pacific Science*, **61**, 173–190.

Magnacca KN, Danforth BN (2006) Evolution and biogeography of native Hawaiian *Hylaeus* bees (Hymenoptera: Colletidae). *Cladistics*, **22**, 393–411.

Meusel MS, Moritz RFA (1993) Transfer of paternal mitochondrial DNA during fertilization of honeybee (*Apis mellifera* L.) eggs. *Current Genetics*, **24**, 539–543.

Monaghan MT, Balke M, Gregory TR, Vogler AP (2005) DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **360**, 1925–1933.

Murray TE, Fitzpatrick U, Brown MJF, Paxton RJ (2008) Cryptic species diversity in a widespread bumble bee complex revealed using mitochondrial DNA RFLPs. *Conservation Genetics*, **9**, 653–666.

Nardi F, Caparelli A, Fanciulli PP, Dallai R, Frati F (2001) The complete mitochondrial DNA sequence of the basal hexapod *Tetrodontophora bielanensis*: evidence for heteroplasmy and tRNA translocations. *Molecular Biology and Evolution*, **18**, 1293–1304.

Pamilo P, Viljakainen L, Viljakainen A (2007) Exceptionally high density of NUMTs in the honeybee genome. *Molecular Biology and Evolution*, **24**, 1340–1346.

Rubinoff D, Cameron S, Will K (2006) A genomic perspective on the shortcomings of mitochondrial DNA for ''barcoding'' identification. *Journal of Heredity*, **97**, 581–594.

Satta Y, Toyohara N, Ohtaka C *et al.* (1988) Dubious maternal inheritance of mitochondrial DNA in *D. simulans* and evolution of *D. mauritiana*. *Genetical Research*, **52**, 1–6.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.

Smith MA, Fisher BL, Hebert PDN (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **360**, 1825–1834.

Snodgrass RE (1956) Crustacean metamorphoses. *Smithsonian Miscellaneous Collections*, **131**, 1–78.

Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 13486–13491.

de Stordeur E (1997) Nonrandom partition of mitochondria in heteroplasmic Drosophila. *Heredity*, **79**, 615–623.

Tsang WY, Lemire BD (2002) Stable heteroplasmy but differential inheritance of a large mitochondrial DNA deletion in nematodes. *Biochemistry and Cell Biology*, **80**, 645–654.

Tsaousis AD, Martin DP, Ladoukakis ED, Posada D, Zouros E (2005) Widespread recombination in published animal mtDNA sequences. *Molecular Biology and Evolution*, **22**, 925–933.

U.S. Fish and Wildlife Service (1994) Endangered and threatened wildlife and plants; animal candidate review for listing as endangered or threatened species; proposed rule. *Federal Register* **59**, 58982–59028.

Volz-Lingenhöhl A, Solignac M, Sperlich D (1992) Stable heteroplasmy for a large-scale deletion in the coding region of *Drosophila subobscura* mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 11528–11532.

Walton C, Butlin RK (1997) A phylogeny for grasshoppers of the genus *Chitaura* (Orthoptera: Acrididae) from Sulawesi, Indonesia, based on mitochondrial DNA sequence data. *Biological Journal of the Linnaean Society*, **62**, 365–382.

Webb KE, Barnes DKA, Clark MS, Bowden DA (2006) DNA barcoding: a molecular tool to identify Antarctic marine larvae. *Deep-Sea Research Part II-Topical Studies in Oceanography*, **53**, 1053–1060.

White DJ, Wolff JN, Pierson M, Gemmell NJ (2008) Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology*, **17**, 4925–4942.

Woolfit M, Bromham L (2005) Population size and molecular evolution on islands. *Proceedings of the Royal Society B-Biological Sciences*, **272**, 2277–2282.

Zhang D-X, Hewitt GM (1996) Nuclear integrations: challenge for mitochondrial DNA markers. *Trends in Ecology & Evolution*, **11**, 247–251.

Zhang D-X, Hewitt GM (1997) Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematics and Ecology*, **25**, 99–120.

Zouros E, Ball AO, Saavedra C, Freeman KR (1994) An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 7463–7467.

Supporting Information

Additional supporting information may be found in the online version of this article.

Appendix S1 Extraction details and GenBank codes for sequences used in this study.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.