Molecular Ecology Resources (2009) **9** (Suppl. 1), 257–268 doi: 10.1111/j.1755-0998.2009.02650.x

BARCODING VERTEBRATES

# Countering criticisms of single mitochondrial DNA gene barcoding in birds

ALLAN J. BAKER, ERIKA SENDRA TAVARES and REBECCA F. ELBOURNE
*Department of Natural History, Royal Ontario Museum, 100 Queen's Park, Toronto, ON, Canada M5S 2C6, and Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada M5S 3B2*

**Abstract**

**General criticisms of a single mtDNA gene barcodes include failure to identify newly evolved species, use of species-delimitation thresholds, effects of selective sweeps and chance occurrence of reciprocal monophyly within species, inability to deal with hybridization and incomplete lineage sorting, and superiority of multiple genes in species identification. We address these criticisms in birds because most species are known and thus provide an ideal test data set, and we argue with selected examples that with the exception of thresholds these criticisms are not problematic for avian taxonomy. Even closely related sister species of birds have distinctive COI barcodes, but it is not possible to universally apply distance thresholds based on ratios of within-species and among-species variation. Instead, more rigorous methods of species delimitation should be favoured using coalescent-based techniques that include tests of chance reciprocal monophyly, and times of lineage separation and sequence divergence. Incomplete lineage sorting is also easily detected with DNA barcodes, and usually at a younger time frame than a more slowly evolving nuclear gene. Where DNA barcodes detect divergent reciprocally monophyletic lineages, the COI sequences can be combined with multiple nuclear genes to distinguish between speciation or population subdivision arising from high female philopatry or regional selective sweeps. Although selective sweeps are increasingly invoked to explain patterns of shallow within-species coalescences in COI gene trees, caution is warranted in this conjecture because of limited sampling of individuals and the reduced power to detect additional mtDNA haplotypes with one gene.**

*Keywords*: criticisms of COI barcodes, efficacy in birds, single versus multiple genes, species identification

*Received 17 November 2008; revision received 6 January 2009; accepted 25 January 2009*

## Introduction

Because a huge proportion of the biodiversity on the planet has yet to be identified and named, DNA barcoding with a short standardized sequence of the mitochondrial DNA gene cytochrome oxidase I (COI) was proposed as a rapid and cost-effective method to help document as many species as possible before many go extinct (Hebert *et al*. 2003a, b). The proposal initially met with severe criticism either because it was perceived to be a form of DNA taxonomy, and appeared to fly in the face of current methods of phylogenetic systematics based on a large number of characters from multiple genes, morphology, and other potentially informative traits, or because of the potential problems with a single-locus threshold to identify species (Seberg *et al*. 2003; Lee 2004; Moritz & Cicero 2004; Will & Rubinoff 2004; Will *et al*. 2005; Dasmahapatra & Mallet 2006; Meier *et al*. 2006; Shearer & Coffroth 2008).

COI barcode sequences of about 650 bp are now known to harbour sufficient numbers of variable sites to be widely effective in identifying species of invertebrates and vertebrates (Hebert *et al*. 2004a, b; Smith *et al*. 2005, 2007, 2008; Ward *et al*. 2005; Clare *et al*. 2007; Tavares & Baker 2008). Furthermore, deep genealogical splits within species flag lineages that warrant additional study as potential cryptic species or as historical phylogeographical subdivisions with species. However, slower rates of evolution in mtDNA of flowering plants have required the use of plastid genes to identify species (Chase *et al*. 2007; Lahaye *et al*. 2008),

Correspondence: Erika Sendra Tavares, Fax: 1-416-586-5553; E-mail: erika.st@gmail.com

and the presence of introns in COI in some fungi present potential problems for the single gene approach (Seifert *et al*. 2007). Even in taxonomically mature groups such as birds where COI barcodes have been able to identify species with > 94% success (Hebert *et al*. 2004b; Kerr *et al*. 2007; Tavares & Baker 2008), the use of a single gene threshold in a maternally inherited genome has been strongly criticized. In particular, problems might be expected in distinguishing species in closely related sister lineages, in species that have undergone introgressive hybridization, in clades where interspecific and intraspecific variation overlap, in taxa with strongly diverged reciprocally monophyletic lineages that could arise by chance or selective sweeps rather than by speciation, or in emergent species that share ancestral polymorphisms in mtDNA (e.g. Moritz & Cicero 2004; Meyer & Paulay 2005; Rubinoff *et al*. 2006; Trewick 2008). We address each of these concerns with some examples from our own work or from the literature, and argue that these problems in birds are easily identified with COI barcodes, and can be resolved with subsequent studies using appropriate analyses.

## Materials and methods

### *DNA barcoding*

Total genomic DNA was extracted from tissues using phenol, Chelex or membrane purification with glass fibre filtration plates [Acroprep 96 Filter Plate-1.0 μm Glass, PALL Corporation (Ivanova *et al*. 2006)]. Polymerase chain reaction (PCR) amplification with bird primers and sequencing of the standardized DNA barcode from 5′ end of the *COI* gene was performed following the same protocols specified in Tavares & Baker (2008). Sequences were obtained using an ABI 3100 (Applied Biosystems), and were checked for ambiguities in Sequencher 4.1.2 (GeneCodes Corp.). Multiple alignments of DNA barcodes were performed in MacClade 4 (Maddison & Maddison 2005). *COI* sequences produced for this study (10 species and 128 sequences) and details of the specimens analysed are deposited in the project 'Royal Ontario Museum-Birds 2' in the Completed Projects selection of the Barcode of Life Data System (BOLD, GenBank Accession nos FJ582508–FJ582635). Additional sequences used are available in the Completed Projects selection of the BOLD in the 'Birds of North America' project (GenBank Accession nos DQ432694–DQ433261, DQ433274–DQ433846, DQ434243–DQ434805), and published work (Milá *et al*. 2007; Nyári 2007), and are detailed in Table S1, Supporting information.

### *Comparisons with other genes*

To test the resolving power of COI in detecting within-species divergence with small samples normally analysed by investigators, we compared DNA barcodes with short control region sequences obtained previously from purple martins (*Progne subis*), as reported in Baker *et al*. (2008), and from little blue penguins (*Eudyptula minor*). For the latter we used primers ND6LR (5′-GCCCCCCGAGAYAACCCCCG-3′, Oliver Haddrath) and LBPCRH (5′-AATGAGAAT-GTYMYTGTATACGGRTAT-3′, Kristen Choffe) to amplify a 350-bp fragment from the hypervariable 5′ end of the control region using the same amplification and sequencing protocols as for purple martins (Baker *et al*. 2008, GenBank Accession nos FJ589573–FJ589632). Additionally, to test tree topologies obtained with COI barcodes vs. multiple genes, we amplified and sequenced the mtDNA genes ATPase 6, cytb, 12SrDNA (GenBank Accession nos FJ603651–FJ603668) and included COI from samples of snipes (*Gallinago gallinago* and *G. delicata*) using the same protocols for shorebirds outlined in Baker *et al*. (2007). Bayesian analysis was performed in MrBayes 3.2 (Ronquist & Huelsenbeck 2003) using a partitioned likelihood approach, considering each gene as a partition. The best-fit models of nucleotide evolution for each gene was selected with the Akaike criterion in MrModelTest 2 (Nylander 2004): Hasegawa–Kishino–Yano (HKY) for COI, cytb and 12S, and HKY with a proportion of invariable sites for ATPase 6. Two Markov chains were run for 10 million generations, sampling trees each 100 generations. The burn-in was determined by the convergence of likelihood scores (100 000 sample trees of which 10 000 initial trees were discarded). The same comparison was also made with samples of yellow-rumped warblers using Bayesian analysis of COI, ATPase and cytb published by Milá *et al*. (2007), and the thrush-like Schiffornis (*Schiffornis turdina*) using the Bayesian analysis of COI, ND2 and cytb published by Nyári (2007).

### *Tests of selective neutrality*

To test whether the inference that shallow coalescences in species cluster of COI sequences are the imprint of selective sweeps on mtDNA genomes, we computed the statistics *Fs*, *F** and *D** of Fu & Li (1993) and Fu (1997). These statistics are based on an excess of young mutations (rare alleles), genetic hitchhiking on selected genes, or for background selection against deleterious alleles. To make these tests more sensitive we used the faster evolving control region sequences as they are linked to COI in the mtDNA genome, and the large sample available for the New Zealand little blue penguins, with the advantage that the Australian samples (which appear to be a sibling species) could be employed as an outgroup.

## Results

### *DNA barcodes in closely related species*

DNA barcoding of sister-species pairs defined with rigorous multigene phylogenies showed that even closely related

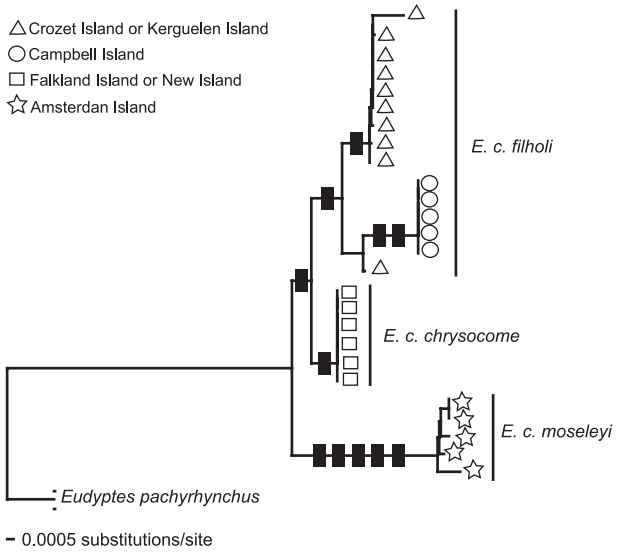


**Fig. 1** Neighbour-joining tree of COI barcodes of three putative subspecies of rockhopper penguins. Substitutional differences distinguishing lineages are mapped on the tree as solid bars.

species of birds from diverse clades could be identified successfully (Tavares & Baker 2008). However, a more stringent test is whether newly emergent species can be identified, so we selected two examples at the interface between within-species differentiation and speciation from the literature. The first involves rockhopper penguins, which on the basis of morphological differentiation and allopatric breeding populations in the southern oceans are thought to represent either three subspecies, or alternatively two separate species. The southern rockhopper (*Eudyptes chrysocome chrysocome*) breeds on the Falklands Islands and islands near Cape Horn, the eastern rockhopper (*Eudyptes chrysocome filholi*) breeds in the southeastern Indian Ocean, the subantarctic islands of New Zealand and Macquarie Island, and the larger northern rockhopper (*Eudyptes chrysocome moseleyi*) breeds on islands in the South Atlantic (Tristan da Cunha and Gough Islands) and the mid-southern Indian Ocean (Amsterdam and St. Paul Islands). The alternative taxonomic arrangement is that the northern rockhopper constitutes a separate species from the southern + eastern rockhoppers (Jouventin 1982; Cooper *et al*. 1990).

A phylogeny based on three mtDNA genes (12S, COI, cytb) recovered three reciprocally monophyletic clades (Fig. 1), corresponding to the three subspecies (Banks *et al*. 2006). Despite obvious morphological differences that suggest only two species should be recognized, the authors argued that genetic distances among the three subspecies were equivalent to those among penguin species reported in Baker *et al*. (2006), and therefore suggested that three species might be recognized. The rockhoppers are also structured into three reciprocally monophyletic groups by their distinctive COI barcodes, so we applied a coalescent test (Rosenberg 2007) to determine if subdivision is likely to have arisen by random branching within one species by chance. The test rejected this hypothesis (*E. c. filholi* vs. *E. c. chrysocome*, $P = 1.8 \times 10^{-6}$; *E. c. chrysocome* vs. *E. c. moseleyi*, $P = 1.2 \times 10^{-6}$), and taxonomic distinctiveness seems appropriate. Ultimately, the decision to recognize two or three species depends not only on genetic data but also on independent corroborating data adduced from other characters such as morphology, behaviour and life-history traits. In this example, DNA barcoding with COI was able to mirror the results achieved with a multigene phylogeny, thus indicating the efficacy of a single mtDNA gene approach.

A second example purports to show the opposite — that COI barcodes could not distinguish between two Mesoamerican sedentary forms of the yellow-rumped warbler complex (*Dendroica coronata*), whereas multiple mtDNA genes (COI, ATPase, control region) could separate them (Milá *et al*. 2007). The complex consists of four subspecies within a single species, comprised of two migratory forms in North America (myrtle warbler *Dendroica coronata coronata* and Audubon's warbler *Dendroica coronata auduboni*), and two largely sedentary forms in Mexico (*Dendroica coronata nigrifrons*) and Guatemala (*Dendroica coronata goldmani*). COI barcode sequences detected the large mutational gap in minimum-spanning networks between the North American and Mesoamerican forms placing them in reciprocally monophyletic groups (Fig. 2). Given the > 3% sequence divergence and phenotypic differences between these two groups, it might be reasonable to recognize them as separate species irrespective of whether COI barcodes or multigene evidence was used in support of this proposition. However, because COI on its own did not also recover reciprocal monophyly of the two Mesoamerican forms, Milá *et al*. (2007) cautioned that it might not always be a good choice for barcoding species of birds. The genetic distance between these two forms is very low, and when corrected for within-species variation it is approximately 0.04%. Thus in this example, the failure to recover very weakly differentiated monophyletic groups in Mesoamerican subspecies has been conflated with the failure of DNA barcoding as a species identification tool. Furthermore, the authors were loathe to commit to reclassifying any of these forms as separate species. This is really an excellent example of a study that might follow a DNA barcoding survey in which the complex would have been flagged as being composed of two potential species.

### *Species-delimitation thresholds*

The use of thresholds, such as the 10 times rule for the ratio of between-species and among-species variation in DNA barcodes or some empirically derived amount of sequence divergence (e.g. 2.7%), has been shown to be effective in

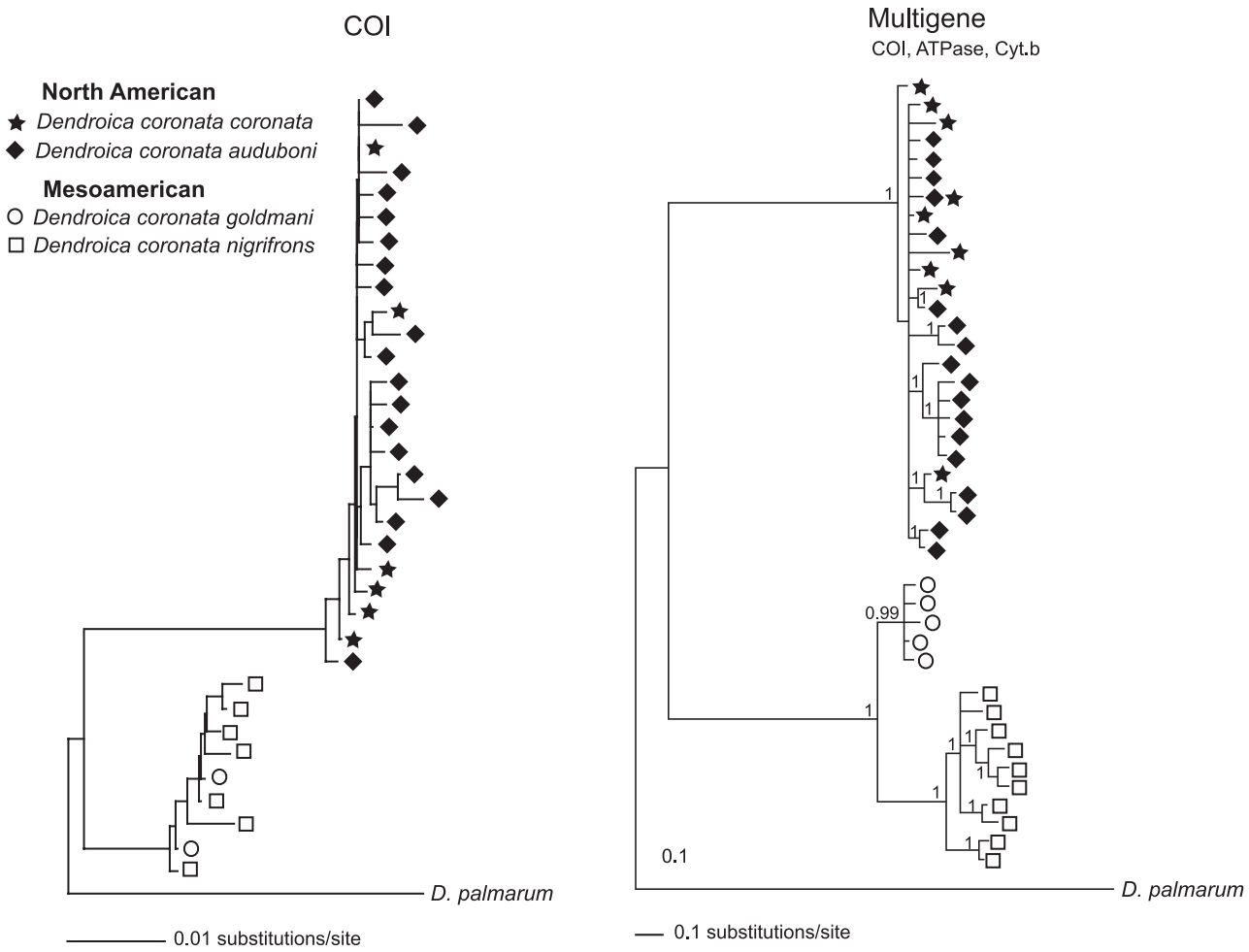


**Fig. 2** Neighbour-joining tree of COI barcodes vs. Bayesian topology recovered with three mtDNA genes (COI, ATPase, cyt b) of the yellow-rumped warbler *Dendroica coronata complex*. Support at the nodes in the Bayesian tree is given by posterior clade probabilities (Milá *et al*. 2007).

birds (Hebert *et al*. 2003a). However, there is some overlap between barcodes that identify species vs. within-species variation, so the limiting conditions to test the efficacy of species-delimitation thresholds is at the boundary between the two. We present two contrasting examples to demonstrate that although thresholds are useful to flag population splits for further investigation, they are not universally applicable. The first example is from a phylogeographical study of the purple martin (*Progne subis*) using the fast-evolving mtDNA control region sequences of 214 birds sampled on either side of the Rocky Mountains in the USA and Canada (Baker *et al*. 2008). Two haplogroups on either side of the Rocky Mountains were revealed that differed by 14 mutations (3.5%). The study also surprisingly revealed that a few chicks sampled from nest boxes in the British Columbia and Washington populations had divergent eastern haplotypes rather than the common western haplotypes, indicating that the populations, presently classified as different subspecies, are interfertile. COI barcodes of the two subspecies differ by 2.0–3.5% and nine nucleotide substitutions (Fig. 3) at the margin suggested by one threshold, but given their morphological similarity and successful interbreeding it would be unwise to suggest they are different species.

Another example is the thrush-like schiffornis (*Schiffornis turdina*) from the hyper-diverse Neotropical avifauna. A range-wide molecular study by Nyári (2007) using three mtDNA genes (ND2, COI, cytb) recovered seven well differentiated phylogroups (3.4–9.6% sequence divergence) that corresponded to different ecogeographical regions (Fig. 4). The author suggested that six species should be recognized, which curiously included a polyphyletic group apparently justified on independent evidence from vocalizations. He also concluded that COI barcodes alone did not recover the monophyly of *S. turdina* or details in the phylogroup, and that the 10 times rule would not apply as average within-species differentiation was 6.7%. If the six phylogroups were accorded species status as he suggested, the highest within-species divergence would be 1.7%, so

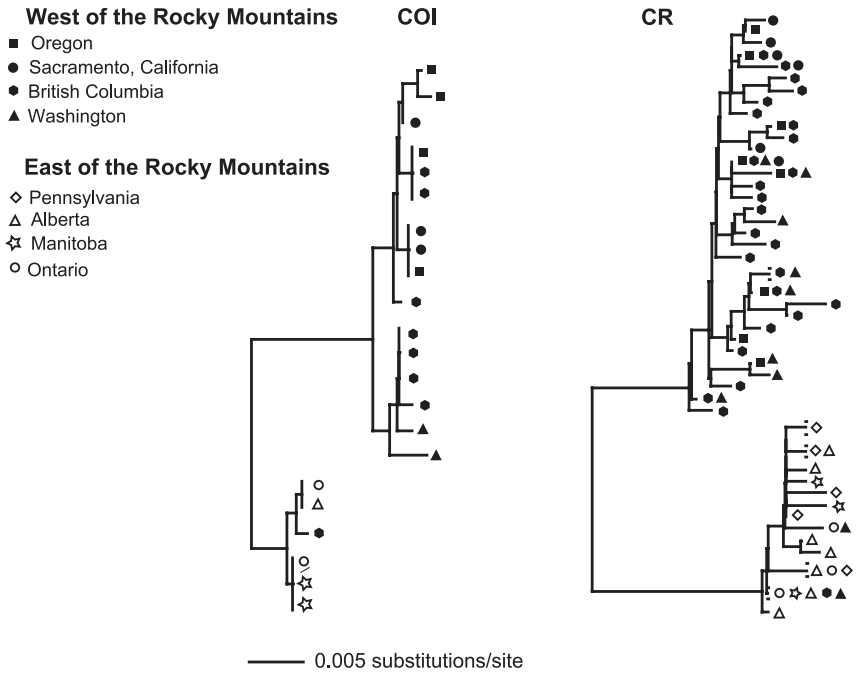


**Fig. 3** Neighbour-joining trees of COI barcodes vs. the hypervariable control region of purple martin populations samples west and east of the Rocky Mountains. These populations are currently classified respectively as separate subspecies *Progne subis arboricola* and *P. s. subis*.

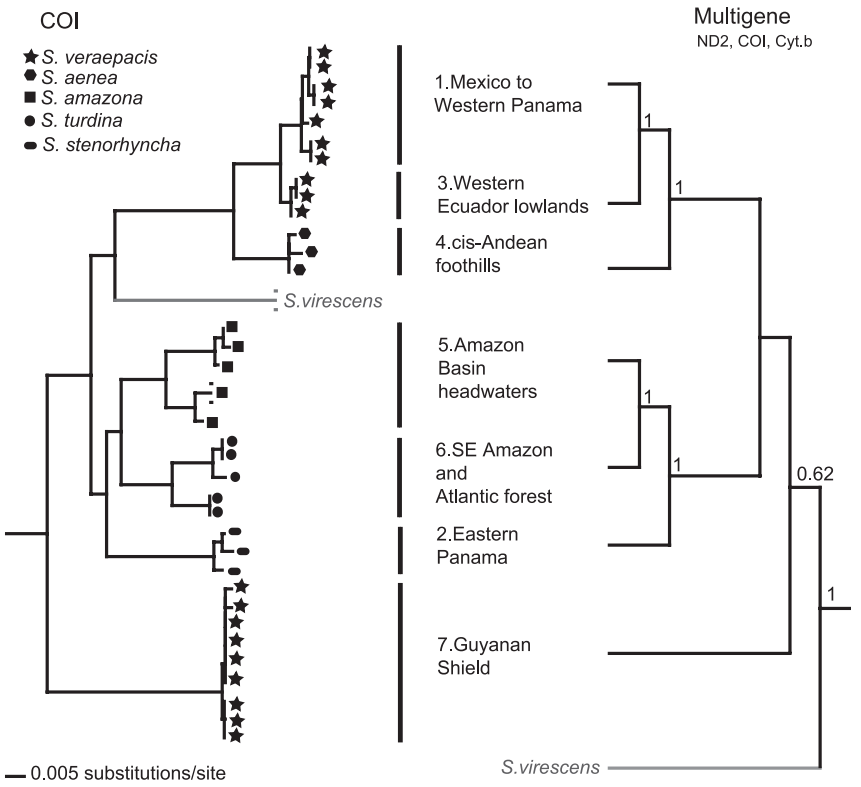


**Fig. 4** Neighbour-joining trees of COI barcodes vs. three mtDNA genes (ND2, COI, Cytb) for Neotropical thrush-like schiffornis samples. Symbols at the branch tips indicate suggested species limits following Nyári (2007).

even then the 10 times rule would be too stringent and fail to identify these new species. Furthermore, COI sequences do recover the seven species phylogroups he detected with the three-gene analysis, and whether or not they form a monophyletic group is not germane to the efficacy of barcoding in species identification. A straightforward COI barcoding analysis likely would have detected most of the well-differentiated phylogroups with initial sampling, and would provide the justification for expanding the study to seek genetic boundaries within the species complex.

**Table 1** Presence or absence of reciprocal monophyly relative to mtDNA divergence and associated divergence time (Johnson & Cicero 2004) between closely related sister species of birds

| Sister-species pair | Estimated divergence time (years ago) | Percentage of mtDNA divergence | Reciprocally monophyletic (COI) |
|---|---|---|---|
| *Zonotrichia atricapilla*, *Z. leucophrys* | 50 000 | — | No |
| *Carduelis flammea*, *C. hornemanni* | 100 000 | — | No |
| *Plectrophenax nivalis*, *P. hyperboreus* | 100 000 | — | No |
| *Anas cyanoptera*, *A. discors* | 150 000 | — | No |
| *Baeolophus atricristatus*, *B. bicolor* | 200 000 | 0.4 | Yes |
| *Sphyrapicus nuchalis*, *S. ruber* | 250 000 | — | No |
| *Empidonax difficilis*, *E. occidentalis* | 350 000 | 0.7 | Yes |
| *Centrocercus minimus*, *C. urophasianus* | 850 000 | 1.8 | Yes |
| *Catharus minimus*, *C. bicknelli* | 900 000 | 1.8 | Yes |

### *Selective sweeps and chance occurrence of reciprocal monophyly*

On the basis of simulations conducted using the neutral coalescent and the simple two-locus/single-incompatibility Bateson-Dobzhanksy-Muller model of speciation by geographical isolation, Hickerson *et al*. (2006) concluded that neither the thresholds of reciprocal monophyly or the 10 times rule would prevent large error rates in detecting recently diverged species with single mtDNA barcodes. Under the assumptions of their model, they found that the thresholds had error rates < 10% only when populations had been isolated for > 4 million generations. While we do not dispute that the 10 times rule is far too conservative for identifying recently diverged lineages, empirical evidence from barcoding of closely related sister-species pairs differing by 0.6–0.9% (Table 1) confirms that reproductively isolated species can be detected with COI sequences in 100 000–150 000 generations (assuming a generation time of 2 years in small passerines), rather than > one million generations as found in the simulations. This result is consistent with the faster neutral coalescent time of mtDNA genes relative to nuclear DNA (Zink & Barrowclough 2008).

If mtDNA is affected frequently by selective sweeps mediated by cytonuclear interactions, then COI barcodes in different species should accumulate fixed differences quite rapidly under moderate to strong selection. This had led to the suggestion that the distinctive COI barcodes in > 94% of bird species are attributable to recurrent selective sweeps (e.g. Hebert *et al*. 2004b). Although selective sweeps could well occur occasionally in the history of matrilineal lineages, this interpretation is not reflected in DNA barcodes of closely related sister species of birds diverging by less than 0.6%; instead their barcodes share incompletely sorted ancestral polymorphisms or are mixed by subsequent hybridization (Table 1). Furthermore, tests based on detection of an excess of young mutations (rare alleles), genetic hitchhiking on selected genes, or for background selection against deleterious alleles (Fu & Li 1993; Fu 1997) usually do not reject selective neutrality of mtDNA variation in species of birds (e.g. Zink *et al*. 2006; Baker *et al*. 2008).

While it is tempting to impute selective sweeps to explain reciprocally monophyletic groups in different species in the COI gene tree, these shallow coalescences are due at least in part to the mutation rate in the barcode region that limits fine-scale resolution of haplotypes, and also to insufficient sampling across the species ranges. Two examples illustrate this point, both comparing haplotype variation detected with COI vs. the faster mutating control region when sample sizes are large enough to depict intraspecific variation in more detail. In the cackling goose (*Branta hutchinsii*), limited sampling of five individuals normally employed in DNA barcoding would have a high likelihood of sampling the common haplotype, as shown in Fig. 5, thus appearing to represent the signature of a selective sweep. More extensive sampling of 24 individuals, however, recovered additional haplotypes in a pattern consistent with neutrality. Similarly, in the little blue penguin the increased resolution afforded by the control region sequences not only supports the deep split between New Zealand and Australian populations (suggesting they are sibling species) but also suggests that this mtDNA division is likely not due to a selective sweep (Fig. 6). Using the Australian sibling species as an outgroup, the CR sequences in the New Zealand species do not deviate significantly from expectations of selective neutrality (Fu & Li test with an outgroup, $D = 0.471$, $P > 0.10$; $F = -0.581$, $P > 0.10$).

### *Hybridization and incomplete lineage sorting*

A common problem for avian taxonomists is to identify hybrids which not only occur between congeneric species but also more rarely between different genera. This is because reproductive isolation can take a long time to evolve in birds, and complete loss of hybrid fertility and viability has been estimated to take as long as 7 million and 11.5 million

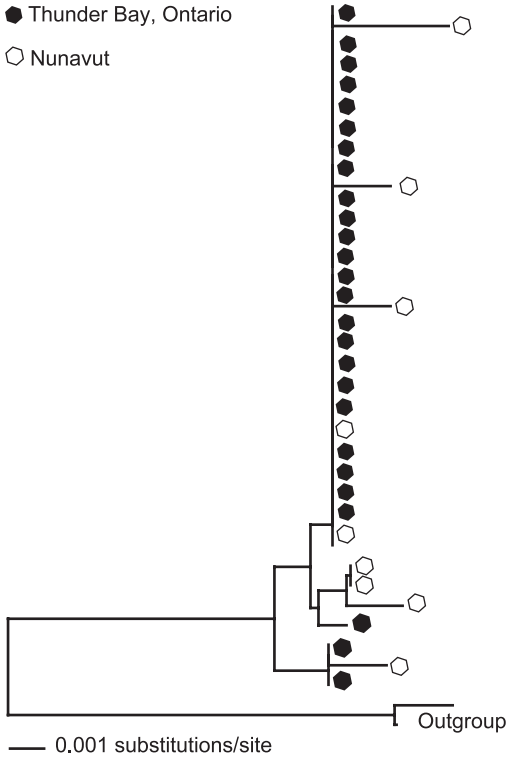


**Fig. 5** Neighbour-joining tree of COI barcodes of the cackling goose showing increased detection of rarer haplotypes with larger sample size.

years, respectively (Price 2007). For example, mallards and northern pintails are estimated to have speciated about four million years ago, but they still produce fully fertile hybrids. Against this backdrop of a potentially prolonged period of introgressive hybridization it has been suggested that a single matrilineally transmitted COI barcode would fail to delimit hybridizing taxa, and thus multiple nuclear markers would be required. Whereas this would be a much more severe problem with taxonomically poorly studied organisms with many undescribed taxa, in birds it is unlikely to cause much difficulty. In the event that introgressive hybridization has transferred mtDNA genomes between species then DNA barcoding will obviously not be able to identify individuals to species, and in that sense it fails. However, the advantage of using barcodes is that it will immediately flag the sharing of haplotypes between phenotypically distinct species, and invite further studies using additional genes to discriminate between alternative hypotheses of hybridization and the retention of shared ancestral polymorphisms. For example, the surprising finding in a mtDNA phylogeographical study of the *Dendroica coronata* complex was that two phenotypically differentiated forms in North America, the myrtle warbler *D. c. coronata* and Audubon's warbler *D. c. auduboni*, share haplotypes across their ranges (Milá *et al*. 2007). There is a relatively small hybrid zone where they come into contact in British Columbia and Alberta, and they were previously thought to be separate species (Hubbard 1969). Extensive sampling revealed that only haplotypes found in both forms were at higher frequency and were central in haplotype networks, and thus they concluded that ancestral polymorphism rather that an ancient bout of introgressive hybridization was the most likely explanation for this phylogeographical pattern. Analyses of mtDNA sequences alone were sufficient to reach this conclusion, though it would be instructive to test it with fast evolving nuclear genes such as microsatellites given that the two forms have only occupied their ranges since ice sheets retreated in the last 10 000 years.

Macaroni penguins (*Eudyptes chrysolophus*) are known to hybridize with other congeneric species such as royal (*E. schlegeli*) and rockhopper (*E. chrysocome*) penguins (White & Clausen 2002). Although the similarity of the barcodes in this tree (Fig. 7) could also be due to shared ancestral haplotypes, this is extremely unlikely because the divergence time of the two parental species has been estimated at about 1.5 million years ago (Baker *et al*. 2006), which is ample time for sorting into reciprocally monophyletic clades. Putative hybrids are readily detected with DNA barcodes (Fig. 7), but this does not pose any problem in taxonomy because their phenotypes are intermediate to the parental species. Because hybridization between the species is rare, introgression is not the problem it would be if significant amounts of backcrossing were occurring. However, even in species where introgression is extensive, both nuclear and mtDNA genes will be mixed and thus will identify the individuals as recombinants of two known species. Hybridization is not likely to be a major problem in birds because of their mature taxonomy, phenotypic differences between species, and numerous examples of well-studied hybrid zones. A possible exception is the very recently evolved white-headed gulls in the northern hemisphere (Hebert *et al*. 2004b), where hybridization and ancestral polymorphism are confounded, but species limits in this complex remain as problematic for detailed genetic analyses (Liebers *et al*. 2004; Pons *et al*. 2005) as they are for DNA barcodes.

Another instructive example of the ability of COI barcode sequences to test putative species limits in very recently diverged lineages is provided by the common snipe complex (*Gallinago gallinago*). Wilson's snipe was considered to be a subspecies (*Gallinago gallinago delicata*) of the common snipe because of their overall morphological similarities, but in 2002 the AOU checklist committee raised it to separate species status as *G. delicata* (Banks *et al*. 2002). This decision was made on differences in winnowing display sounds generated by differences in the number and size of its outer tail feathers (Thönen 1969; Tuck 1972; Miller 1996), similar to differences between other closely related species in the

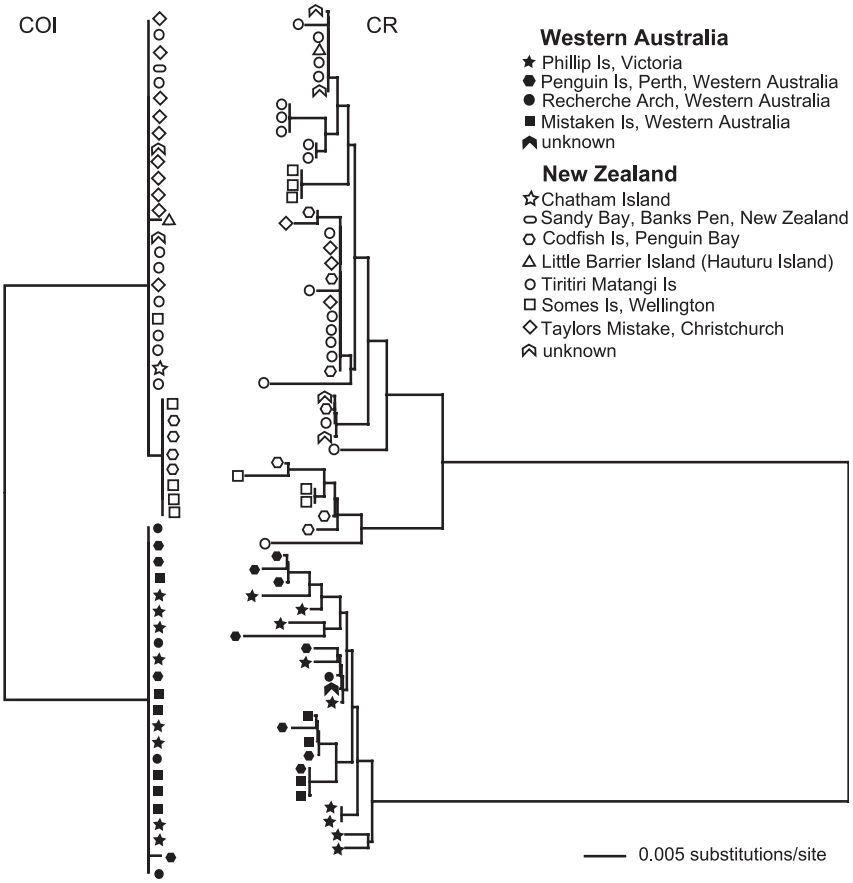


**Fig. 6** Neighbour-joining trees showing increased resolution of haplotypes of little blue penguins from Australia and New Zealand using a hypervariable part of the mtDNA control region relative to COI barcodes.

genus. However, DNA barcoding was unable to distinguish either species, suggesting that they share a common ancestral COI haplotype. A multigene phylogeny constructed with sequences from three additional mtDNA genes (ATPase6, cytb,12S) placed *delicata* within the *G. gallinago* clade, suggesting that it is at best a subspecies of the latter (Fig. 8). More sampling of individuals is required to verify this placement, but if it is correct then barcoding has proved its effectiveness in refuting species status in this case, rather than the reverse.

### *Single mtDNA barcode vs. multiple nuclear genes*

DNA barcoding of species with a single maternally inherited gene like COI has been criticized because it would fail to detect a significant number of closely related species, as moderate to strong divergent selection on genes causing reproductive isolation or species recognition could evolve rapidly. The choice of COI was based on the generally faster rate of mutation in the mtDNA genome than the nuclear genome, and its low cost and applicability to a wide range of taxa via PCR amplification and sequencing protocols that are amenable to high throughput approaches. Impressive progress has been made in barcoding diverse taxa including a range of vertebrates and invertebrates, but slower rates of evolution in flowering plants, for example, have necessitated a switch to multiple and more informative plastid genes. However, it is clearly advisable to utilize a single barcode to inventory biodiversity where it is shown to be efficacious and accurate, as biodiversity around the world is at risk and extinction rates are likely to increase in the near future.

Nevertheless, exciting analytical developments in the field of statistical phylogeography have the potential to detect divergence of lineages very early in the process of speciation, and have the potential to distinguish species and infer phylogenies even when multiple gene trees are not sorted into reciprocally monophyletic groups (Knowles 2000, 2001; Maddison & Knowles 2006). Somewhat ironically, an earlier criticism of DNA barcoding was that it would fail when the gene tree was unsorted with respect to closely related lineages, but COI has now been analysed along with multiple independent nuclear genes to reconstruct recent speciation events in montane grasshoppers in the genus *Melanoplus*. So we are coming full circle from the million years or so estimate of time for the evolution of reproductive isolation in birds (Price 2007) to the late Pleistocene (< 200 000 years) to recognize the early stages of lineage diversification that separates species. This development puts species identification more in line with the

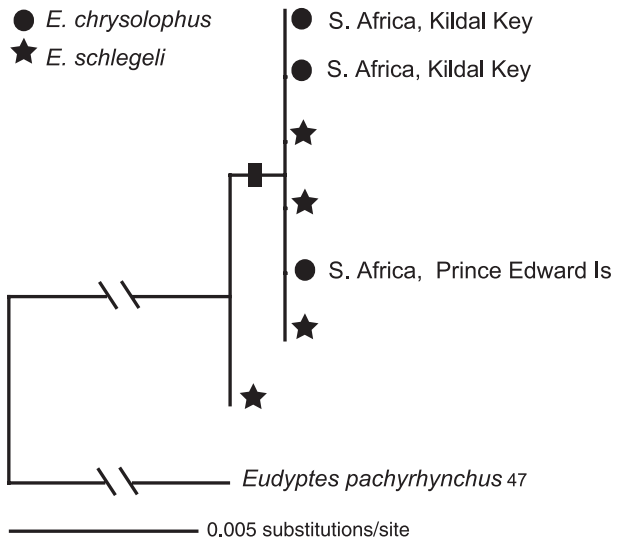


**Fig. 7** Neighbour-joining tree of COI barcodes of royal (*Eudyptes schlegeli*) and macaroni (*E. chrysolophus*) penguins showing shared haplotypes probably due to hybridization.

concept of phylogenetic species, but with the addition of a reasonably long time period since populations split from a common ancestral population. Most ironically of all, the famous textbook example of speciation in action is provided by Darwin's finches, yet the species of *Geospiza* which have some amazing variations in bill and body size are not reciprocally monophyletic in their mtDNA (Sato *et al*. 1999), they have a very short divergence time of 16 000–63 000 years ago estimated with mtDNA (Zink 2002), and they hybridize without fitness consequences (Grant & Grant 1994). By none of these criteria are they reproductively isolated or even phylogenetic species, as any DNA barcode study would reveal. Avian taxonomists are therefore placing heavy weight on morphological and ecological differences between lineages in continuing to recognize them as separate species, albeit in the very early stages of speciation. Introgressive hybridization and selection were shown to contribute to significant convergent evolution of *Geospiza fortis* and *Geospiza scandens* (Grant *et al*. 2004), and under favourable ecological conditions these lineages could possibly fuse into a hybrid species.

## Discussion

Despite earlier criticisms of species identification with a single mtDNA gene, DNA barcoding is gaining worldwide acceptance as a rapid way to document as much of the biodiversity of the planet in the face of the accelerating rate of loss of species expected in this century. The upcoming International Barcode of Life (*i*BOL) project involving an international consortium of agencies is the outcome of proof-of-concept research done on a range of plants, fungi, animals and protists. Although difficulties remain that may require different markers from a range of genes, as in land plants and some fungi, the overall goal still remains to use a universal barcode as far as this is possible. In taxonomically mature groups, such as birds and mammals, COI barcodes effectively classify the great majority of

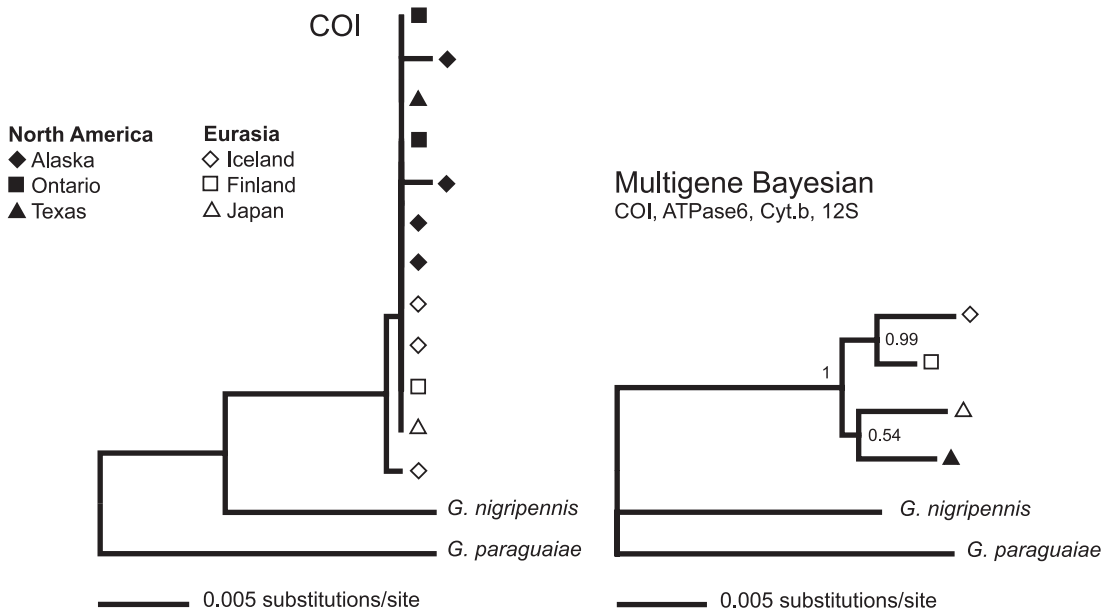


**Fig. 8** Neighbour-joining trees of COI barcodes and Bayesian tree constructed from four mtDNA genes (COI, Cytb, ATPase, 12S) showing that *Gallinago delicata* haplotypes (North America) are embedded within the *Gallinago gallinago* cluster. Support at the nodes in the Bayesian tree is given by posterior probabilities.

species and identify divergent lineages that warrant further investigation (Hebert *et al*. 2004b; Clare *et al*. 2007; Tavares & Baker 2008).

We have shown that closely related sister species pairs of birds defined previously on independent data sets have species-specific single gene barcodes, but there is overlap of within-species and among-species variation (Tavares & Baker 2008). Thus arbitrary thresholds for species delimitation based on a genetic distance of say 2.7% sequence divergence or a ratio of 10 times the among to within species variation (the 10 times rule) are too conservative and will miss species, as critics had maintained. The solution to this problem can be found in coalescent techniques that combine not only branching pattern tests for chance occurrence of reciprocal monophyly (Rosenberg 2007), but also the amount of genetic differentiation as a proxy for the time since lineages last exchanged genes. In practice, taxonomists rely on these criteria whether or not they accept the biological species or phylogenetic species concepts, as the acid tests of interbreeding and hybrid fitness are untestable in allopatric lineages.

The suggestion that single mtDNA gene barcodes will fail to recognize recently diverged biological species because divergent selection on nuclear loci can rapidly isolate lineages (Hickerson *et al*. 2006) is not so far borne out in birds. The evolution of pre-mating isolation between lineages is instead estimated to take at least a million years, which is long enough to allow the accumulation of about 1–2% sequence divergence among sister species (Pereira & Baker 2006). Complete loss of hybrid fertility and viability in birds is estimated to take an order of magnitude longer (Price 2007). However, this is not to say that speciation is some species pairs could be more rapid, but neither gene sequences nor phenotypic changes are likely to provide convincing evidence that the lineages have speciated in less than 100 000 years. The reverse is observed in birds and other vertebrates, as stabilizing or convergent selection on morphologically cryptic sibling or hybridizing species have been detected primarily with DNA sequences (e.g. Grant *et al*. 2004; Johnson & Cicero 2004). Introgressive hybridization and unsorted ancestral polymorphism in sister species can be identified with DNA barcodes, though telling them apart relies on evidence synthesized from genetics, phenotypic differences between lineages and possibly the geographical spread of haplotypes beyond contact zones (Milá *et al*. 2007).

Recent developments of coalescent applications in phylogeography have demonstrated the worth of using multigene approaches, which not only can employ both mitochondrial and nuclear genes, but also can resolve species trees even when individual genes are not sorted into reciprocally monophyletic lineages (e.g. Jennings & Edwards 2005; Maddison & Knowles 2006; Knowles & Carstens 2007). COI sequences generated from DNA barcoding projects thus provide a species-rich library for use along with nuclear genes for resolving divergent monophyletic lineages that might represent unrecognized species. In this case, the alternative hypotheses of population subdivision via restricted female dispersal and male-biased gene flow, or regional selective sweeps acting on mtDNA, will need to be tested. Selective sweeps in different lineages will increase the efficacy of COI barcodes in identifying species, but whether or not within-species haplotype diversity is consistent with this interpretation and occurs commonly in birds must await higher resolution analyses of larger samples.

## Acknowledgements

We thank Kristen Choffe and Sergio Luiz Pereira for providing some sequences used in this study. We are grateful to Janet Hinshaw of the University of Michigan Museum of Zoology for a tissue loan of a common snipe. This work was supported by funding through the Canadian Barcode of Life Network from Genome Canada through the Ontario Genomics Institute, NSERC, and other sponsors listed at www.BOLNET.ca, and the ROM Governors' Fund.

## Conflict of interest statement

The authors have no conflict of interest to declare and note that the funders of this research had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## References

Baker AJ, Pereira SL, Haddrath OP, Edge KA (2006) Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 11–17 (doi: 10.1098/rspb.2005.3260).

Baker AJ, Pereira SL, Paton TA (2007) Phylogenetic relationships and divergence times of Charadriiformes genera: multigene evidence for the Cretaceous origin of at least 14 clades of shorebirds. *Biology Letters*, **3**, 205–209 (doi: 10.1098/rsbl.2006.0606).

Baker AJ, Greenslade AD, Darling L, Finlay JC (2008) High genetic diversity in the Blue-listed British Columbia population of the Purple Martin maintained by multiple sources of immigrants. *Conservation Genetics*, **9**, 495–505.

Banks J, Van Buren A, Cherel Y, Whitfield JB (2006) Genetic evidence for three species of rockhopper penguins, *Eudyptes chrysocome*. *Polar Biology*, **30**, 61–67.

Banks RC, Cicero C, Dunn JL *et al*. (2002) Forty-third supplement to the American Ornithologists' Union Check-list of North American Birds. *Auk*, **119**, 897–906.

Chase MW, Cowan RS, Hollingsworth PM, van den Berg C *et al*. (2007) A proposal for a standardised protocol to barcode all land plants. *Taxon*, **56**, 295–299.

Clare EL, Kim BK, Engstrom MD, Eger JL, Hebert PDN (2007) DNA barcoding of Neotropical bats: species identification and discovery within Guyana. *Molecular Ecology Notes*, **7**, 184–190.

Cooper J, Brown CR, Gales RP *et al*. (1990) Diets and dietary segregation of crested penguins (*Eudyptes*). In: *Penguin Biology* (eds Davis LS, Darby JT), pp. 131–156. Academic Press, San Diego, California.

Dasmahapatra KK, Mallet J (2006) DNA barcodes: recent successes and future prospects. *Heredity*, **97**, 254–255.

Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Fu Y-X, Li W-H (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.

Grant PR, Grant BR (1994) Phenotypic and genetic effects of hybridization in Darwin's finches. *Evolution*, **48**, 297–316.

Grant PR, Grant BR, Markert JA, Keller LF, Petren K (2004) Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. *Evolution*, **58**, 1588–1599.

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

Hebert PDN, Ratnasingham S, deWaard JR (2003b) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences*, **270**, S96–S99.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004a) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences, USA*, **101**, 14812–14817.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004b) Identification of birds through DNA barcodes. *Public Library of Science Biology*, **2**, 1657–1663.

Hickerson MJ, Meyer CP, Moritz C (2006) DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology*, **55**, 729–739.

Hubbard JP (1969) The relationships and evolution of the *Dendroica coronata* complex. *Auk*, **86**, 393–432.

Ivanova NV, DeWaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes*, **6**, 998–1002.

Jennings WB, Edwards SV (2005) Speciational history of Australian grass finches (*Poephila*) inferred from 30 gene trees. *Evolution*, **59**, 2033–2047.

Johnson NK, Cicero C (2004) New mtDNA data affirm the importance of Pleistocene speciation in North American birds. *Evolution*, **58**, 1122–1130.

Jouventin P (1982) *Visual and Vocal Signals in Penguins, Their Evolution and Adaptive Characters*. Verlag Paul Parey, Berlin, Germany.

Kerr KC, Stoeckle MY, Dove CJ *et al*. (2007) Comprehensive DNA barcode coverage of North American birds. *Molecular Ecology Notes*, **7**, 535–543.

Knowles LL (2000) Tests of Pleistocene speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of western North America. *Evolution*, **54**, 1337–1348.

Knowles LL (2001) Genealogical portraits of speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of the Rocky Mountains. *Proceedings of the Royal Society B: Biological Sciences*, **268**, 319–324.

Knowles LL, Carstens BC (2007) Delimiting species without monophyletic gene trees. *Systematic Biology*, **56**, 887–895.

Lahaye R, van der Bank M, Bogarin D *et al*. (2008) DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences, USA*, **105**, 2923–2928.

Lee MSY (2004) The molecularisation of taxonomy. *Invertebrate Systematics*, **18**, 1–6.

Liebers D, de Knijff P, Helbig AJ (2004) The herring gull complex is not a ring species. *Proceedings of the Royal Society of London B: Biological Sciences*, **273**, 11–17.

Maddison W, Maddison D (2005) *MacClade 4: Analysis of Phylogeny and Character Evolution*, Version 4.08. Sinauer & Associates, Sunderland, Massachusetts.

Maddison WP, Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, **55**, 21–30.

Meier R, Shiyang K, Vaidya G, Ng P (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, **55**, 715–728.

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *Public Library of Science Biology*, **3**, e422.

Milá B, Smith TB, Wayne RK (2007) Speciation and rapid phenotypic differentiation in the yellow-rumped warbler *Dendroica coronata* complex. *Molecular Ecology*, **16**, 159–173.

Miller EH (1996) Acoustic differentiation and speciation in shorebirds. In: *Ecology and Evolution of Acoustic Communication in Birds* (eds Kroodsma DE, Miller EH), pp. 59–88. Comstock/Cornell University Press, Ithaca, New York.

Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. *Public Library of Science Biology*, **2**, 1529–1531.

Nyári AS (2007) Phylogeographic patterns, molecular and vocal differentiation, and species limits in *Schiffornis turdina* (Aves). *Molecular Phylogenetics and Evolution*, **44**, 154–164.

Nylander JAA (2004) *MrModelTest v2*. Program distributed by the author. Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden.

Pereira SL, Baker AJ (2006) A mitogenomics timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. *Molecular Biology and Evolution*, **23**, 1731–1740.

Pons J-M, Hassanin A, Crochet P-A (2005) Phylogenetic relationships within the Laridae (Charadriiformes: Aves) inferred from mitochodrial markers. *Molecular Phylogenetics and Evolution*, **37**, 686–699.

Price T (2007) *Speciation in Birds*. Roberts and Company Publishers, Greenwood Village, Colorado.

Ronquist FR, Huelsenbeck JPH (2003) MrBayes: Bayesian inference of phylogeny. *Bioinformatics*, **19**, 1572–1574.

Rosenberg NA (2007) Statistical tests for taxonomic distinctiveness from observations of monophyly. *Evolution*, **61**, 317–323.

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.

Sato A, O'Huigin C, Figueroa F *et al*. (1999) Phylogeny of Darwin's finches as revealed by mtDNA sequences. *Proceedings of the National Academy of Sciences, USA*, **96**, 5101–5106.

Seberg O, Humphries CJ, Knapp (2003) Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends in Ecology & Evolution*, **18**, 63–65.

Seifert KA, Sampson RA, deWaard JR *et al*. (2007) Prospects for fungus identification using CO1 DNA barcodes, with *Pencillium* as a test case. *Proceedings of the National Academy of Sciences, USA*, **104**, 3901–3906.

Shearer TL, Coffroth MA (2008) Barcoding corals: limited by interspecific divergence, not intraspecific variation. *Molecular Ecology Resources*, **8**, 247–255.

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 of London. Series B, Biological Sciences*, **360**, 1825–1834.

Smith MA, Wood DM, Janzen DH, Hallwachs W, Hebert PDN (2007) DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. *Proceedings of the National Academy of Sciences, USA*, **104**, 4967–4972.

Smith MA, Poyarkov NA, Hebert PDN (2008) CO1 DNA barcoding amphibians: take the chance, meet the challenge. *Molecular Ecology Resources*, **8**, 235–246.

Tavares ES, Baker AJ (2008) Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. *BMC Evolutionary Biology*, **8**, 81 (doi: 10.1186/1471-2148-8-81).

Thönen W (1969) Auffallender Unterschied zwischen den instrumentalen Balzlauten der europäischen und nordamerikanischen Bekassine Gallinago gallinago. *Ornithologische Beobachter*, **66**, 6–13.

Trewick SA (2008) DNA Barcoding is not enough: mismatch of taxonomy and genealogy in New Zealand grasshoppers (Orthoptera: Acrididae). *Cladistics*, **24**, 240–254.

Tuck LM (1972) *The Snipes: A Study of the Genus Capella*. Canadian Wildlife Service Monograph Series, no. 5., Ontario, Canada.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London. Series B, Biological Science*, **360**, 1847–1857.

White RW, Clausen AP (2002) Rockhopper *Eudyptes chrysocome chrysocome* x Macaroni *E. chrysolophus* penguin hybrids apparently breeding in the Falkland Islands. *Marine Ornithology*, **30**, 40–42.

Will KW, Rubinoff D (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics*, **20**, 47–55.

Will KW, Mishler B, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, **54**, 844–851.

Zink RM (2002) A new perspective on the evolutionary history of Darwin's finches. *Auk*, **119**, 864–871.

Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, **17**, 2107–2121.

Zink RM, Drovetski SV, Rohwer S (2006) Selective neutrality of mitochondrial ND2 sequences, phylogeography and species limits in *Sitta europaea*. *Molecular Phylogenetics and Evolution*, **40**, 679–686.

## Supporting Information

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

**Table S1** Detailed information about specimens used in this study: collection localities, museum collection identification, Bold identification and Genbank accession numbers for each gene.

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