Coding sequence polymorphism in avian mitochondrial genomes reflects population histories

AUSTIN L. HUGHES and MARY ANN K. HUGHES*
Department of Biological Sciences, University of South Carolina, Coker Life Sciences Bldg., 700 Sumter Street, Columbia, SC 29208, USA

Abstract

Nucleotide sequence diversity at mitochondrial protein-coding loci from 72 species of birds from different geographical regions was analysed in order to test the hypothesis that temperate zone species show population genetic effects of past glaciation. Temperate zone species showed reduced nucleotide diversity in comparison to tropical mainland species, suggesting that the latter have long-term effective population sizes due to population bottleneck effects during the most recent glaciation. This hypothesis was further supported by evidence of an unusually high estimated rate of population growth in species breeding in North America and wintering in the New World tropics (Nearctic migrants), consistent with population expansion after a bottleneck. Nearctic migrants also showed evidence of an abundance of rare nonsynonymous (amino acid-altering) polymorphisms, a pattern suggesting that slightly deleterious polymorphisms drifted to high frequencies during a bottleneck and are now being eliminated by selection. Because the shape of the North American land mass limited the area available for refugia during glaciation, the bottleneck effects are predicted to have been particularly strong in Nearctic migrants, and this prediction was supported. The reduced genetic diversity of Nearctic migrants provides an additional basis for concern for the survival of these species, which are threatened by loss of habitat in the winter range and by introduced disease.

Keywords: Nearctic migrant birds, nearly neutral theory, population bottleneck, population effects of glaciation, slightly deleterious mutations

Received 24 September 2006; revision accepted 13 November 2006

Introduction

Conservation biologists have argued that maintaining genetic diversity in natural populations can be an important factor in assuring their continued survival, both because polymorphisms at certain key loci (such as immune system loci) may be essential for survival and because genome-wide polymorphism underlies the additive genetic variance enabling short-term adaptive responses to environmental change (O'Brien 1994; Templeton 1994). Thus, an understanding of global patterns of genetic polymorphism and their causes can play an important role in devising comprehensive conservation strategies. The level of polymorphism present in a species in turn largely reflects the species' demographic history.

*Deceased
Correspondence: Austin L. Hughes, PhD, Fax: 1-803-777-4002;
E-mail: austin@biol.sc.edu

For animal and plant species inhabiting the Earth's temperate zones, a major factor in that history was the most recent glaciation between 60 000 and 10 000 years ago, which may have caused population bottlenecks and a consequent loss of genetic diversity in numerous temperate zone species (Avise *et al.* 1988; Bucklin & Wiebe 1998; Leonard *et al.* 2000; McCusker *et al.* 2000; Lessa *et al.* 2003). In birds (Aves), species which breed in the temperate zones but winter in the tropics are particularly likely to have undergone bottlenecks during glaciation, as a result of the restriction of available breeding habitat (Steadman 2005; Williams & Webb 1996). Because of their inability to survive low winter temperatures, migrants are expected to have experienced more severe range reduction due to glaciation than nonmigrants (Williams & Webb 1996). Furthermore, because both the distance and direction of migration are known to have a genetic basis (Berthold 1991; Berthold & Pulido 1994; Pulido *et al.* 1996, 2001), climate change is likely to have caused substantial mortality due to misdirected migration.

Most nucleotide sequence polymorphism is believed to result from genetic drift affecting selectively neutral or nearly neutral variants (Kimura 1983; Nei 1987). Since the rate of fixation of neutral or nearly neutral variants is inversely related to effective population size, the extent of polymorphism provides an index of a species' long-term effective population size (Fuerst *et al.* 1977). Positive natural selection, including both balancing selection and directional selection, also affects the frequency of certain sequence variants. Balancing selection can maintain polymorphisms, potentially for a much longer time than neutral polymorphisms typically last (Takata & Nei 1990), while directional selection can lead to fixation of selectively advantageous variants, often with an accompanying reduction in polymorphism in genomic regions closely linked to the site or sites subject to selection (Maynard Smith & Haigh 1974; Charlesworth 1992).

However, there is abundant evidence that by far the most prevalent form of natural selection in nature is not positive selection but purifying selection; that is, selection acting to eliminate selectively deleterious alleles (Kimura & Ohta 1974; Ruiz-Pesini *et al.* 2004). The prevalence of purifying selection is supported by the observation that, in comparisons of ornologous protein-coding genes, the number of synonymous nucleotide substitutions per synonymous site generally exceeds the number of non-synonymous (amino acid-altering) nucleotide substitutions per nonsynonymous site (Kimura 1977; Nei 1987). Assuming that mutation affects synonymous and nonsynonymous sites equally, the higher rate of substitution at the former implies that the nonsynonymous mutations are eliminated prior to fixation to a greater extent than are synonymous mutations.

The pattern of purifying selection also provides information regarding population history. In populations of small effective population size (N_e), selection is expected to be inefficient at removing slightly deleterious mutations, which can then drift to high frequency or even become fixed (Ohta 1976, 2002). On the other hand, if a species experiences population growth after a severe bottleneck, there will be a number of slightly deleterious alleles that drifted to high frequencies when population size was small but are subject to purifying selection when N_e becomes larger. There is evidence of such a pattern in the human population, where single-nucleotide polymorphisms predicted to have a deleterious effect on protein structure or gene expression occur at lower frequencies than synonymous polymorphisms in the same genes (Hughes *et al.* 2003, 2005).

Bazin *et al.* (2006) compared patterns of nucleotide substitution in nuclear and mitochondrial genes of animals and argued that polymorphism in mitochondrial genomes does not reflect population history but rather shows evidence of repeated selective sweeps (Maynard Smith & Haigh 1974). Their study relied on these two assumptions:

(i) that patterns of neutral sequence polymorphism should be explainable by differences between vertebrates and invertebrates with respect to long-term N_e; and (ii) that N_e values of vertebrates are lower than those of invertebrates. However, these assumptions may be questioned, given the complexity of life-history differences between vertebrates and invertebrates and the numerous differences in population biology between mitochondrial and nuclear genomes (Fay & Wu 1999; Weinreich & Rand 2000). A more appropriate test for an association between long-term N_e and mitochondrial polymorphism would compare more closely related taxa whose population histories are expected to have resulted in different N_e.

Here we analyse an extensive database of protein-coding gene sequences from the mitochondrial genomes of a worldwide sample of avian species in order to test for correlations between geographical distribution and the pattern of genetic polymorphism. In these analyses, we assume that biogeographical categories are correlated with population history. On average, we assume that tropical species mainland species, should have larger long-term N_e than temperate-zone mainland species because of the longer lasting climatic stability of the tropics in comparison to the temperate zones over the past 700 000 years, during which seven well-defined glaciations occurred (Barron 1984; Kastner & Goñi 2003; Lessa *et al.* 2003), and we predict a particularly strong impact of glaciation on migrants. Note this represents an appropriate timescale for analysis of polymorphism within vertebrate species, since most polymorphisms in species with long-term effective populations on the order of those seen in vertebrates (= 10^5) will not be older than a few hundred thousand years, with the exception of a few balanced polymorphisms such at those of the major histocompatibility complex (MHC, Takata & Nei 1990). Within the tropics, mainland species are predicted on average to have greater N_e than island species because of the larger average range sizes of mainland species (Stattersfield *et al.* 1998) and because founder effects may reduce effective population size of island species (Estoup & Clegg 2003; Miller & Lambert 2004).

Methods

Sequence data

Our analyses used 103 data sets, each of which consisted of a set of four or more aligned allelic partial or complete sequences for one of five mitochondrial protein-coding genes (COI, *ND2*, *ND3*, cyt*b*, and *ATP6*). Sequences were aligned by the CLUSTAL x program (Thompson *et al.* 1994). These data sets included a total of 2377 individual sequences and represented 72 species; 19 of these species were represented by two or more data sets (see Table S1, Supplementary material). A total of 3237 sites were

polymorphic within species. Species were placed in four biogeographical categories: (i) Nearctic migrant, including species breeding in the Nearctic region and wintering in the Neotropics; (ii) other temperate zone, including both year-round residents of the Nearctic or other temperate regions (e.g. Australian) and temperate-to-tropic migrants of the Old World; (iii) tropical island, including species whose life cycle is confined to one or more small (< 40 000 km²) oceanic islands; and (iv) tropical mainland, including species resident on tropical continental areas or larger tropical islands (e.g. Madagascar). Nearctic migrants were analysed separately from other temperate zone species because of the availability of data from a substantial number of these species and because of the expectation that the effects of glaciation might have been particularly acute in Nearctic migrants. Because of the shape of the North American land-mass, we predicted that there would have been a relative paucity of refugia in North America, in comparison to the Palearctic (Hewitt 2004). We followed the species-level taxonomy of the authors who obtained the sequences and the higher-level taxonomy of Sibley & Monroe (1990).

Statistical analyses

The number of synonymous substitutions per synonymous site and the number of nonsynonymous substitutions per nonsynonymous site were estimated by Li's (1993) method, using the MEGA 2 software (Kumar *et al*. 2001). This method was used because it takes into account transitional bias; and, as is typical in vertebrate mitochondrial genomes, there was a strong transitional bias in the present data, with the transition:transversion ratio (R) in all data sets estimated at 7.2:1. Within each data set, the mean for all pairwise comparisons of the number of synonymous substitutions per synonymous site provided an estimate of nucleotide diversity at synonymous sites (π_S); and the mean for all pairwise comparisons of the number of nonsynonymous substitutions per nonsynonymous site provided an estimate of nucleotide diversity at synonymous sites (π_N) (Nei & Kumar 2000). Weighted averages of π_S and π_N for the 19 species represented by more than one data set were obtained by weighting with the numbers of synonymous and nonsynonymous sites, respectively, estimated by the modified Nei–Gojobori method (Zhang *et al*. 1998), assuming an R of 7.2.

Gene diversity (Nei 1987; p. 177) was estimated separately at each polymorphic site, as in Hughes *et al*. (2003); where x_i is the frequency of the *i*th allele (nucleotide) at a given locus (site), the gene diversity is $1 - \Sigma x_i^2$. Polymorphic sites were classified as synonymous or nonsynonymous, based on the coding effect of the nucleotide change. There were 29 sites (0.9%) that could not be so classified; in these cases, because of multiple polymorphic sites within a single codon, the coding effect of a given substitution depended on the pathway taken by evolution. These 29 sites were therefore excluded from analyses of gene diversity at individual polymorphic sites.

In order to examine the relative frequency of rare alleles at synonymous and nonsynonymous sites, we compared the average number of nucleotide differences and the number of segregating sites (Tajima 1989) separately for synonymous and nonsynonymous sites (Hughes 2005). We computed separately for synonymous and nonsynonymous polymorphisms, the difference $k-S/a_1$, where k is the mean number of nucleotide differences for all pairwise comparisons among n allelic sequences, S is the number of segregating sites, and a_1 is the sum from $i = 1$ to $n - 1$ of $1/i$, which provides an adjustment for sample size (Tajima 1989). We then computed the ratio of this difference to the absolute value of the minimum possible value of the difference, which would occur if all polymorphisms were singletons (Schaeffer 2002). We designate this ratio Q_{syn} in the case of synonymous polymorphisms and Q_{non} in the case of nonsynonymous polymorphisms. Comparing Q_{syn} and Q_{non} provides an index of the relative abundance of rare alleles at synonymous and nonsynonymous sites, with a strongly negative value indicating an abundance of rare alleles (Hughes 2005). We applied this method to all data sets ($N = 82$) that included both synonymous and nonsynonymous polymorphisms. We used these statistics instead of Tajima's D (Tajima 1989) because the latter is dependent on sample size and thus not directly comparable among data sets. Note that in the ratio defined above, the standard error term (which is responsible for the sample size dependence of D) cancels out.

We used the FLUCTUATE program (Kuhner *et al*. 1998) with default settings except that the observed R (7.2) was input, in order to estimate the rate of population growth per mutation rate per generation (g) for each individual data set, assuming exponential population growth or decline. This program uses a coalescent-based approach, which assumes strict neutrality of sequence polymorphisms (Kuhner *et al*. 1998). Even if population growth or decline has occurred in a given species, this growth or decline may not have followed an exponential model. Moreover, the assumption of neutrality was not valid for at least some nonsynonymous polymorphisms in the present data (see below). Because of these violations of its underlying assumptions, results obtained by this method should be considered approximate. In cases where there were multiple data sets for a given species, we averaged the results for the different data sets in order to obtain an estimate for g for the species.

A variety of statistical analyses were conducted using as the unit of analysis the individual polymorphic site, the data set, or the species. Most polymorphisms within a species are likely to have arisen since speciation, the most

obvious exception being balanced polymorphisms (Takata & Nei 1990), which are unlikely in mitochondrial genomes (Nielsen & Weinreich 1999). Since we analysed polymorphism within species, the data for each species is thus phylogenetically and statistically independent of data for other species (Felsenstein 1985; Hughes *et al.* 2006). Consistent with this expectation, in preliminary analyses, there were no statistically detectable effects of order, superfamily, family, or genus (data not shown). Likewise, although data sets differed with respect to the numbers of sequences available for analyses, there were no statistically detectable effects of number of sequences (data not shown). Both parametric analyses based on means and nonparametric analysis based on medians yielded similar results, and only the former are reported here.

Results

Mean π_S and π_N at mitochondrial protein-coding loci within each of 72 avian species (see Table S1, Supplementary material) were compared by paired-sample *t*-test (Fig. 1A). Overall mean π_S (0.0385 ± 0.0058 SE.) was over an order of magnitude greater than overall mean π_N (0.0027 ± 0.0004); and the difference was highly significant ($P < 0.001$; Fig. 1A). Moreover, in the case of 58 species for which there were data on both synonymous and nonsynonymous polymorphisms, the mean gene diversity at synonymous polymorphic sites was significantly greater than mean gene diversity at nonsynonymous polymorphic sites (paired *t*-test; $P = 0.001$; not shown). When mean gene diversities at polymorphic sites were computed for the 72 species, mean gene diversity at synonymous polymorphic sites was significantly greater than that at polymorphic nonsynonymous sites (Fig. 1B).

There was a significant difference with respect to mean π_S values among the four biogeographical categories of species ($P = 0.014$; Fig. 2A). The lowest mean π_S was that for Nearctic migrants (0.023 ± 0.008), while the highest was that for tropical mainland species (0.065 ± 0.015). Individual comparisons with a family error rate of 5% revealed a significant difference between mean π_S for Nearctic migrants and that for tropical mainland species and a significant difference between mean π_S for other temperate zone species and that for tropical mainland species (Fig. 2A). Likewise, there was a significant difference with respect to mean π_N values among the four biogeographical categories ($P < 0.001$; Fig. 2B). Mean π_N values for all other categories were significantly different with a family error rate of 1% from that for tropical mainland species (Fig. 2B).

Using a general linear models procedure, we applied a factorial analysis of variance to gene diversity at individual polymorphic sites (see Table S2, Supplementary material) in order to test effects of the category of site (synonymous
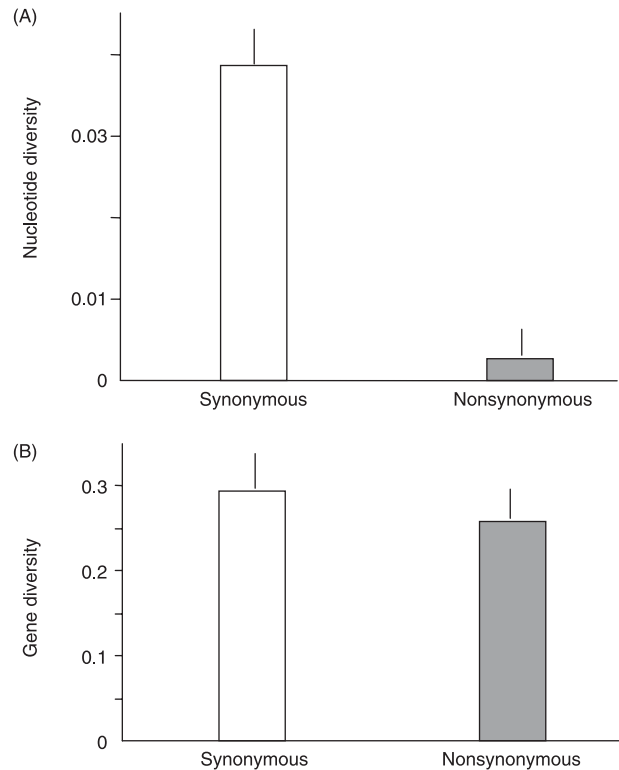

Fig. 1 (A) Mean nucleotide diversity at synonymous (π_S) and nonsynonymous (π_N) sites in mitochondrial protein-coding genes for 72 species of birds ($P < 0.001$; paired *t*-test; two-tailed). (B) Mean gene diversity at polymorphic synonymous and nonsynonymous sites in mitochondrial protein-coding genes for 72 species of birds ($P = 0.001$; paired *t*-test; two-tailed). Vertical lines indicate standard errors.

or nonsynonymous) and the biogeographical category of the species (Fig. 3A). There was a highly significant effect of site category ($P < 0.001$), with mean gene diversity being consistently lower at nonsynonymous than at synonymous sites (Fig. 3A). Likewise, there was a highly significant effect of biogeographical category ($P < 0.001$), with the lowest mean gene diversities for both synonymous and nonsynonymous sites being found in Nearctic migrants (Fig. 3A). However, there was no significant interaction between site category and biogeographical category. Thus, this analysis did not reveal a significant difference among biogeographical categories with respect to the relative magnitude of mean gene diversity at synonymous and nonsynonymous polymorphic sites.

We used the Q statistic, computed separately at synonymous and nonsynonymous sites, as a measure of the relative abundance of rare alleles in the 82 data sets that included both synonymous and nonsynonymous polymorphisms. There was no significant difference among biogeographical categories with respect to Q_{syn} (based on synonymous sites) (Fig. 3B). By contrast, there was a highly significant difference among biogeographical categories with respect
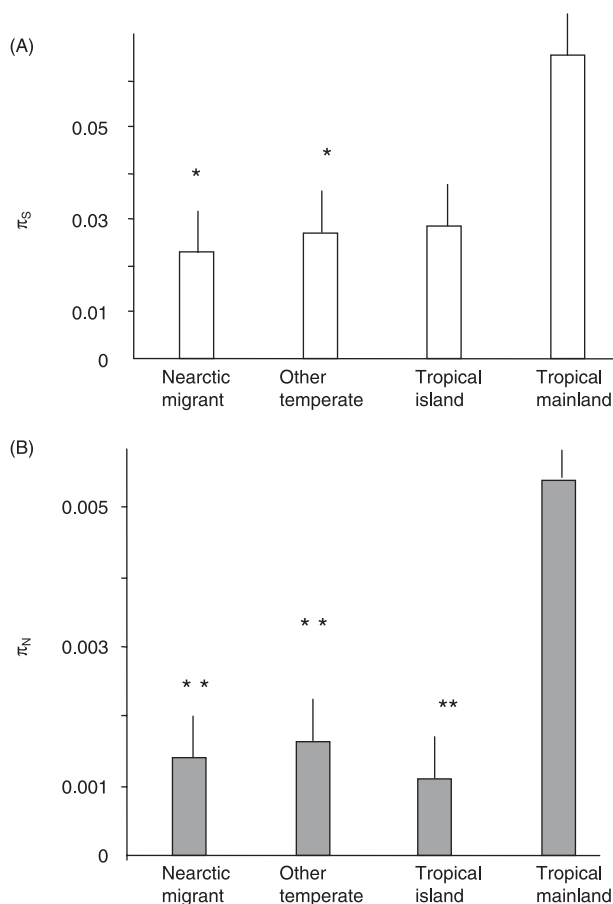
Fig. 2 Mean nucleotide diversity at (A) synonymous (π_S) and (B) nonsynonymous (π_N) sites in mitochondrial protein-coding genes for 72 species of birds classified by biogeographical category. One-way ANOVA testing for differences among categories with respect to π_S: $F_{3,68} = 3.78$; $P = 0.014$; and π_N ($F_{3,68} = 8.57$; $P < 0.001$. Results of Dunnett's post-hoc comparisons (with family error rate) of category means with the mean value for tropical mainland species: *$P < 0.05$; **$P < 0.01$. Vertical lines indicate standard errors.
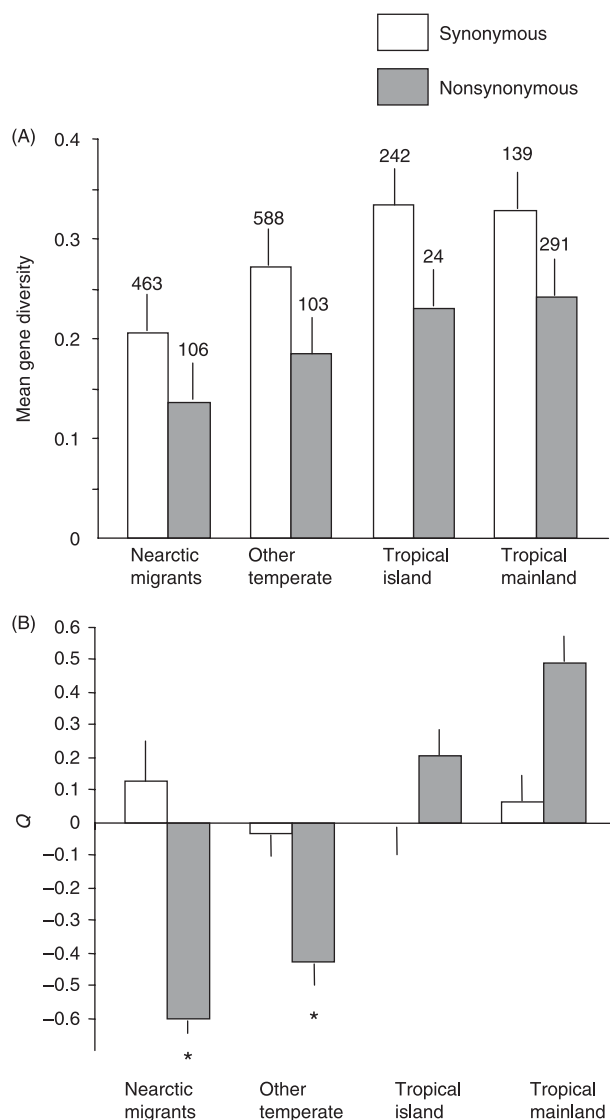

Fig. 3 (A) Mean gene diversity at polymorphic synonymous and nonsynonymous sites categorized by biogeographical category. There were significant effects of site category (synonymous or nonsynonymous; $F_{1,3200} = 63.02$; $P < 0.001$); and of biogeographical category ($F_{3,3200} = 45.59$; $P < 0.001$); but no significant interaction between site category and biogeographical category ($F_{3, 3200} = 0.35$; n.s). Numbers of polymorphic sites in each category are shown above means. (B) Mean Q at synonymous (Q_{syn}) and nonsynonymous (Q_{non}) sites in 103 data sets categorized by biogeographical category. One-way ANOVA for Q_{syn}, $F_{3,78} = 0.09$; n.s.; for Q_{non}, $F_{3,78} = 4.28$; $P = 0.007$. Results of Dunnett's post-hoc comparisons (with family error rate) of category means with the mean value for tropical mainland species: * $P < 0.05$. Vertical lines indicate standard errors.

to Q_{non} (based on nonsynonymous sites; $P = 0.007$; Fig. 3B). By individual comparisons with a 5% family error rate, both Nearctic migrants and other temperate species showed significant differences from tropical mainland species with respect to mean Q_{non} (Fig. 3B).

Estimates of the population growth parameter (g) differed significantly among categories ($F_{3,68} = 3.70$; $P = 0.016$; Fig. 4). The highest mean g was that of Nearctic migrants (3827 ± 918), while the lowest was that of tropical mainland species (968 ± 456), about a quarter that for Nearctic migrants. These two means differed significantly by individual comparisons with a 5% family error rate (Fig. 4).

Discussion

Analysis of polymorphism at mitochondrial protein-coding loci from 72 avian species revealed a substantial impact

of purifying selection. The fact that the mean nucleotide diversity at synonymous sites greatly exceeded that at nonsynonymous sites (Fig. 1A) is evidence that purifying selection has acted to eliminate a substantial fraction of nonsynonymous mutations occurring at these loci.
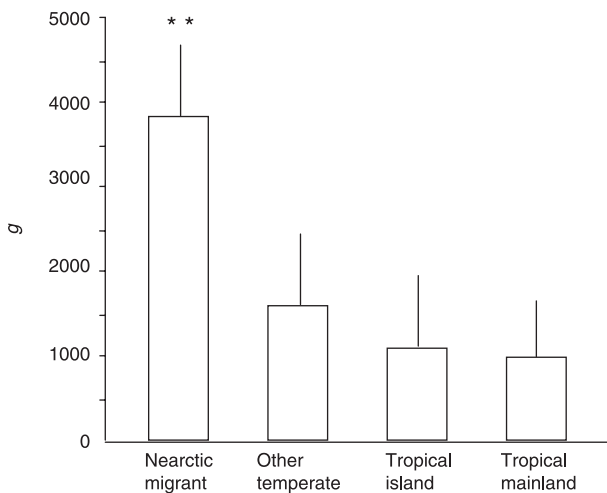
Fig. 4 Mean estimates of the population growth parameter (g) for 72 avian species; one-way ANOVA among categories, $F_{3,68} = 3.70$; $P = 0.016$. Results of Dunnett's post-hoc comparisons (with family error rate) of category means with the mean value for tropical mainland species: **$*P < 0.01$. Vertical lines indicate standard errors.

Furthermore, reduced genetic diversity at nonsynonymous polymorphic sites (Fig. 1B) is evidence that many non-synonymous polymorphisms are subject to ongoing purifying selection, acting to reduce population frequency of slightly deleterious alleles (Hughes *et al.* 2003, 2005). These results are not consistent with the hypothesis of Bazin *et al.* (2006) that positive selection rather than purifying selection predominates in the case of mitochondrial genomes.

Biogeographical categories differed with respect to nucleotide diversity (Fig. 2); gene diversity at both synonymous and nonsynonymous sites (Fig. 3A); the abundance of rare nonsynonymous variants, as measured by Q_{non} (Fig. 3B); and estimated population growth parameter (g) (Fig. 4). These striking differences evidently reflect different population histories. Tropical mainland species showed the highest nucleotide diversities (Fig. 2), supporting the hypothesis that these species have had relatively large long-term effective population sizes in comparison to the other categories. Several previous studies have shown increased within-species genetic diversity in the tropics in a variety of organisms (Alvarez-Buylla *et al.* 1996; Hewitt 2004; Martin & McKay 2004). Increased genetic diversity in the tropics is frequently associated with population subdivision (Mart & McKay 2004). Such genetic differentiation among subpopulations requires long-term stability and a high effective population size for the species as a whole (Wakeley 2000).

Tropical island species showed lower gene diversities than tropical mainland species, although the difference was significant only in the case of nonsynonymous sites (Fig. 2). This result is consistent with the hypothesis that

nucleotide diversity should reflect long-term effective population size, which is expected to be lower in island species because of their more restricted geographical distribution. Note that most nonsynonymous polymorphisms in tropical species are expected to be strictly neutral (rather than 'nearly neutral') because these species will not have gone through the recent bottleneck due to glaciation.

In comparison to tropical mainland species, temperate zone species showed significantly reduced gene diversities at both synonymous and nonsynonymous sites, a result consistent with the hypothesis that these species have reduced long-term effective population sizes as a result of population bottlenecks caused by glaciation. Although estimation of the growth parameter was complicated by the presence of purifying selection on certain nonsynonymous sites, high estimates of the growth parameter in temperate zone species (Fig. 4) were consistent with a rebound after a bottleneck during glaciation. Nearctic migrants showed the highest mean estimated growth parameter, suggesting that the bottleneck-and-rebound effect may have been most pronounced in this group. The loss of breeding habitat may have been particularly acute for species breeding in the Nearctic region and wintering in the Neotropics (here designated Nearctic migrants). As a consequence of the funnel-like shape of the North American land mass, the availability of refugia from glaciation was much reduced in Nearctic migrants in comparison, for example, to Palearctic migrants (Hewitt 2004). Thus, a particularly pronounced effect of glaciation on genetic diversity is not surprising in Nearctic-to-Neotropical migrant birds.

Nearctic migrants also showed strongly negative values of Q_{non}, significantly different from those of tropical mainland species, whereas there were no significant differences among biogeographical categories with respect to Q_{syn}. Strongly negative Q_{non} indicates that there are abundant cases of rare nonsynonymous polymorphisms. Under a population bottleneck, the 'nearly neutral theory' predicts that slightly deleterious mutations can drift to high frequencies because purifying selection cannot eliminate them effectively in a small population (Ohta 1976, 2002). When a bottlenecked population subsequently increases in size, purifying selection is predicted to act to remove such slightly deleterious alleles, leading to a decrease in gene diversity at such sites in comparison to linked neutral sites. Consistent with this prediction, in Nearctic migrants we observed a significant abundance of rare polymorphisms at nonsynonymous sites, where deleterious mutations are likely to occur, but not at synonymous sites. Note that, in mitochondrial genomes, the elimination of slightly deleterious mutations is slowed by the lack of recombination (Weinreich & Rand 2000; Lowe 2006), although the latter can be compensated to some extent by back-mutation due to the higher mutation rate and strong transitional bias (Kumar 1996) in mitochondrial genes. Thus, the overall results are consistent

with the hypothesis that mitochondrial protein-coding genes show a strong signature of past population history, contrary to the conclusion of Bazin *et al.* (2006). When sufficient data on nuclear genes of birds become available, it will be important to test whether the same signature of past evolutionary history is detectable in nuclear genes (see Ballard & Whitlock 2004).

It might be hypothesized that the reduced genetic diversity in the temperate zones and in Nearctic migrants in particular can be attributed to some factor other than glaciation, such as human activities. However, other than a few well-known cases of extinction, temperate zone birds have not experiences unusually severe population reduction. Species with restricted ranges have increased incidence of extinction and endangerment (Hughes 2004), and restricted range species occur much more frequently in the tropics than in the temperate zones (Stattersfield *et al.* 1998). Moreover, human impacts are mostly too recent to have had a substantial impact on genetic diversities of avian species, since, in order to have a significant impact, a bottleneck must last for many generations (Nei *et al.* 1975).

Under the hypothesis that temperate zone species will show genetic effects of past glaciation, species that breed in North America and winter in the tropics might be expected to show strong bottleneck effects because the shape of the North American land mass severely restricted available breeding habitat during the last glaciation. Consistent with this prediction, the overall strongest evidence of genetic bottleneck effects was observed in Nearctic migrants. These species have been subjects of conservation concern because of recent population declines, some but not all of which may be attributed to wintering habitat loss due to clearing of Neotropical forests (Askins *et al.* 1990; Pimm & Askins 1995). That these species have relatively reduced genetic diversity and, potentially, an elevated frequency of slightly deleterious mutations adds further urgency to conservation efforts, since these genetic characteristics may make them particularly vulnerable to recently introduced infectious agents such as the West Nile virus (Van der Meulen *et al.* 2005; McLean 2006).

Acknowledgement

References

Alvarez-Buylla ER, García-Burrios C, Lara-Moreno Martínez-Ramos M (1996) Demographic and genetic models in conservation biology: applications and perspectives for tropical rain forest tree species. *Annual Review of Ecology and Systematics*, **27**, 387–421.

Askins RA, Lynch F, Greenberg R (1990) Population declines in migratory birds in eastern North America. *Current Ornithology*, **7**, 1–57.

Avise JC, Ball RM, Arnold J (1988) Current versus historical population sizes in vertebrate species with high gene flow: a comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations. *Molecular Biology and Evolution*, **5**, 331–344.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Barron EJ (1984) Ancient climates: investigation with climate models. *Reports of Progress in Physics*, **47**, 1563–1599.

Bazin E, Glémin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science*, **312**, 570–572.

Berthold P (1991) Orientation in birds: spatiotemporal programmes and genetics of orientation. *EXS*, **60**, 86–105.

Berthold P, Pulido F (1994) Heritability of migratory activity in a natural bird population. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **257**, 311–315.

Bucklin A, Wiebe PH (1998) Low mitochondrial diversity and small effective population sizes of the copepods *Calanus finmarchicus* and *Nannocalanus minor*: possible impact of climatic variation during recent glaciation. *Journal of Heredity*, **8**, 383–392.

Charlesworth B (1992) New genes sweep clean. *Nature*, **356**, 475–476.

Estoup A, Clegg M (2003) Bayesian inferences on the recent island colonization history by the bird *Zosterops lateralis*. *Molecular Ecology*, **12**, 657–674.

Fay JC, Wu C-I (1999) A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. *Molecular Biology and Evolution*, **16**, 1003–1005.

Felsenstein J (1985) Phylogenies and the comparative method. *American Naturalist*, **125**, 1–15.

Fuerst PA, Chakraborty R, Nei M (1977) Statistical studies on protein polymorphism in natural populations. I. Distribution of single-locus heterozygosity. *Genetics*, **86**, 455–483.

Hewitt GM (2004) The structure of biodiversity — insights from molecular phylogeography. *Frontiers in Zoology*, **2004**, I: 4.

Hughes AL (2004) A statistical analysis of factors associated with historical extinction and current endangerment of non-passerine birds. *Wilson Bulletin*, **116**, 330–336.

Hughes AL (2005) Evidence for abundant slightly deleterious polymorphisms in bacterial populations. *Genetics*, **169**, 533–538.

Hughes AL, Packer B, Welch R, Bergen AW, Chanock SJ, Yeager M (2003) Widespread purifying selection at polymorphic sites in human protein-coding loci. *Proceedings of the National Academy of Sciences, USA*, **100**, 15754–15757.

Hughes AL, Packer B, Welch R, Bergen AW, Chanock SJ, Yeager M (2005) Effects of natural selection on inter-population divergence at polymorphic sites in human protein-coding loci. *Genetics*, **170**, 1181–1187.

Hughes AL, Hughes MA, Friedman R (2006) Variable intensity of purifying selection on cytotoxic T-lymphocyte epitopes in hepatitis C virus. *Virus Research*, (in press).

Kastner TP, Goñi MA (2003) Constancy in the vegetation of the Amazon Basin during the Late Pleistocene: evidence from the organic matter composition of Amazon deep sea fan sediments. *Geology*, **31**, 291–294.

Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature*, **267**, 275–276.

Kimura M (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.

Kimura M, Ohta T (1974) On some principles governing molecular evolution. *Proceedings of the National Academy of Sciences, USA*, **71**, 2848–2852.

Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics*, **149**, 429–434.

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA 2: molecular evolutionary genetics analysis software. *Bioinformatics*, **17**, 1244–1245.

Leonard JA, Wayne RK, Cooper A (2000) Population genetics of Ice Age brown bears. *Proceedings of the National Academy of Sciences, USA*, **97**, 1651–1654.

Lessa EP, Cook JA, Patton JL (2003) Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. *Proceedings of the National Academy of Sciences, USA*, **100**, 10331–10334.

Li W-H (1993) Unbiased estimates of the rates of synonymous and nonsynonymous substitution. *Journal of Molecular Evolution*, **36**, 96–99.

Lowe L (2006) Quantifying the genomic decay paradox due to Muller's ratchet in human mitochondrial DNA. *Genetical Research*, **87**, 133–159.

Martin PR, McKay JK (2004) Latitudinal variation in genetic divergence of populations and the potential for future speciation. *Evolution*, **58**, 938–945.

Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favorable gene. *Genetical Research*, **23**, 23–35.

McCusker MR, Parkinson E, Taylor EB (2000) Mitochondrial DNA variation in rainbow trout (*Onchorhynchus mykiss*) across its native range: testing biogeographical hypotheses and their relevance to conservation. *Molecular Ecology*, **9**, 2089–2108.

McLean RG (2006) West Nile virus in North American birds. *Ornithological Monographs*, **60**, 44–64.

Miller HC, Lambert DM (2004) Genetic drift outweighs balancing selection in shaping post-bottleneck major histocompatibility complex variation in New Zealand robins (Petroicidae). *Molecular Ecology*, **13**, 3709–3721.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei M, Kumar S (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1–10.

Nielsen R, Weinreich DM (1999) The age of nonsynonymous and synonymous mutations in animal mtDNA and implications for the mildly-deleterious theory. *Genetics*, **153**, 497–506.

O'Brien S (1994) A role for molecular genetics in biological conservation. *Proceedings of the National Academy of Sciences, USA*, **91**, 5748–5755.

Ohta T (1976) Role of very slightly deleterious mutations in molecular evolution and polymorphism. *Theoretical Population Biology*, **10**, 254–275.

Ohta T (2002) Near-neutrality in evolution of genes and gene regulation. *Proceedings of the National Academy of Sciences, USA*, **99**, 6134–16137.

Pimm SL, Askins RA (1995) Forest losses predict bird extinctions in eastern North America. *Proceedings of the National Academy of Sciences, USA*, **92**, 9343–9347.

Pulido F, Berthold P, Mohr G, Querner U (2001) Heritability of the timing of autumn migration in a natural bird population. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **268**, 953–959.

Pulido F, Berthold P, van Noordwijk AJ (1996) Frequency of migrants and migrant activity are genetically correlated in a bird population: evolutionary implications. *Proceedings of the National Academy of Sciences, USA*, **93**, 14642–14647.

Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC (2004) Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science*, **303**, 223–226.

Schaeffer SW (2002) Molecular population genetics of sequence length diversity in the Adh region of *Drosophila pseudoobscura*. *Genetical Research*, **80**, 163–175.

Sibley CG, Monroe BL (1990) *Distribution and Taxonomy of Birds of the World*. Yale University Press, New Haven, Connecticut.

Stattersfield AJ, Crosby MJ, Long AJ, Wege DC (1998) *Endemic Bird Areas of the World: Priorities for Biodiversity Conservation*. BirdLife International, Cambridge, UK.

Steadman DW (2005) The paleoecology and fossil history of migratory landbirds. In: *Birds of Two Worlds: the Ecology and Evolution of Migration* (eds Greenberg R, Marra PP), pp. 5–17. Johns Hopkins University Press, Baltimore, Maryland.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **125**, 585–595.

Takata N, Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics*, **124**, 967–978.

Templeton AR (1994) Biodiversity at the molecular genetic level: experiences from disparate macroorganisms. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **345**, 59–64.

Thompson JD, Higgins DG, Gibson T (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Van der Meulen KM, Pensaert MB, Nauuwynck HJ (2005) West Nile virus in the vertebrate world. *Archives of Virology*, **150**, 637–657.

Wakeley J (2000) The effects of subdivision on the genetic divergence of populations and species. *Evolution*, **54**, 1092–1101.

Weinreich DM, Rand DM (2000) Contrasting patterns of non-neutral evolution in proteins encoded in nuclear and mitochondrial genomes. *Genetics*, **156**, 385–399.

Williams TC, Webb TH III (1996) Neotropical bird migration during the Ice Ages: orientation and ecology. *Auk*, **113**, 105–118.

Zhang J, Rosenberg HF, Nei M (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proceedings of the National Academy of Sciences, USA*, **98**, 3708–3713.

Supplementary material

The following supplementary material is available for this article:

Table S1 Summary of species and genes used in analyses.

Table S2 Data on individual polymorphic sites.