Published in final edited form as:

Genetica. 2010 December ; 138(11-12): 1271–1276. doi:10.1007/s10709-010-9527-z.

Reduced microsatellite heterozygosity in island endemics supports the role of long-term effective population size in avian microsatellite diversity

Austin L. Hughes

Department of Biological Sciences, University of South, Carolina, Coker Life Sciences Bldg., 700 Sumter St., Columbia, SC 29208, USA

Austin L. Hughes: austin@biol.sc.edu

Abstract

Because of the high mutation rate of microsatellites, polymorphism at microsatellite loci might be predicted to reflect the effective population size over a time span of about 10,000 years and thus to be associated with biogeographic factors impacting species on that time frame. This prediction was tested by comparing heterozygosity at microsatellite loci from 294 bird species, including 58 species endemic to oceanic islands. Controlling statistically for phylogenetic effects, mean heterozygosity was significantly reduced in oceanic island endemics compared to other species. There was also an effect of current endangerment, statistically independent of the effect of island endemism. These results support the hypothesis that long-term effective population size can be an important causative factor behind differences among species with respect to microsatellite heterozygosity.

Keywords

Aves; Effective population size; Endangered species; Heterozygosity; Microsatellite

The effective population size (N_e) is a fundamental parameter of biological populations for a variety of reasons, including the fact that the efficiency of natural selection is an increasing function of N_e . N_e is thus of particular interest to conservation biology, because when N_e is very small, it may not be possible for selection to remove deleterious variants, some of which may then drift to fixation (Lande 1998; Kohn et al. 2006). Furthermore, because N_e determines the rate at which neutral polymorphisms are lost due to drift, N_e is correlated with the average gene diversity (heterozygosity) at selectively neutral loci (Nei 1987). As a consequence, examination of gene diversity at neutral loci can provide an important indicator of N_e and thus of a population's prognosis for long-term genetic health (Kohn et al. 2006). Note that in general, N_e is not the same as the current census size (N) of a population (Nei 1987). Although N_e and N may be roughly correlated across species, it is expected that there will be some exceptions to such a trend; for example, species with high N but low N_e as a result of recent expansion after a severe bottleneck (Nei 1987).

An important fact about N_e is that it can be measured over different time scales (Lynch 2007). For example, the long-term N_e of a species reflects changes in population size over a

Correspondence to: Austin L. Hughes, austin@biol.sc.edu.

Electronic supplementary material The online version of this article (doi:10.1007/s10709-010-9527-z) contains supplementary material, which is available to authorized users.

very long time frame—hundreds of thousands of years in the case of eukaryotes. Because of the relatively low mutation rate of nucleotide sequences, polymorphism in nuclear or mitochondrial genomes of animals reflects long-term population sizes over periods of several hundred thousand years. Thus, the effects of ancient bottlenecks—for example those caused by glaciation events (Hughes and Hughes 2007)—are likely to be reflected in current polymorphism. By contrast, although the mutation rates vary considerably among microsatellite arrays, it is generally agreed that microsatellite mutation rates are several orders of magnitude higher than the rates of point mutation in DNA sequences (Ellegren 2004; Schlötterer 2000). As a consequence, it might be expected that microsatellite polymorphism should reflect N_e over a shorter time-frame, perhaps the last 1,000–10,000 years.

Evans and Sheldon (2008) presented evidence that microsatellite heterozygosity in birds is negatively correlated with the degree of current extinction risk as categorized by the World Conservation Union (IUCN). These authors did not directly address the causal factors behind this relationship, although they did mention the impacts of very recent anthropogenic impacts on population size as a possible factor in reducing heterozygosity of endangered species (Evans and Sheldon 2008). On the other hand, population genetics theory predicts that even an extreme bottleneck will not necessarily have an immediate impact on heterozygosity, unless there is a founder effect; rather the reduction of heterozygosity is expected to be more pronounced when the bottleneck is prolonged (Nei et al. 1975). Since anthropogenic effects generally are recent (since the European expansion of the past 500 years; Hughes 1999), anthropogenically caused bottlenecks may not always have occurred over a sufficient time span to cause a substantial loss of heterozygosity.

Thus, it might be expected that biogeographic factors acting over a longer time scale than recent human activity may impact avian effective population size and thus heterozygosity, independently of any effects of current endangerment. One example of such a factor might be oceanic island endemism, because an insular distribution imposes a restricted range (Purvis et al. 2000; Hughes and Hughes 2007). Here I test the hypothesis that microsatellite heterozygosity in island endemic species of birds is reduced in comparison to that of species with a continental distribution, independently of recent anthropogenic impacts.

Even though events (such as population bottlenecks) occurring during or after speciation are likely to have a major impact on the effective population size of many species (Hughes and Hughes 2007), in at least some cases effective population size may be correlated with phylogenetic history as well (Fisher and Owens 2004). For example, ecological factors associated with relatively reduced population size (such as occupying a top-predator niche; Fa and Purvis 1997) may be correlated with anciently evolved adaptations of phylogenetic lineages. In addition, it is possible that different clades may have different mutation rates, related to factors such as generation time (Nabholz et al. 2009). Therefore, I use a statistical approach that tests for major phylogenetic effects and allows testing for other effects controlling for the effect of phylogeny.

Methods

Estimates of mean heterozygosity (H) at polymorphic microsatellite loci were gathered from the published literature for 294 bird species belonging to 86 families (Supplementary Table S1). 161 values were taken from Evans and Sheldon (2008); the remainder were from other published sources (Supplementary Table S1). When available, expected heterozygosity (“gene diversity,” Nei 1987) was used rather than observed heterozygosity, because of the lower variance of the former (Nei 1987); but preliminary analyses indicated that, when the authors provided both expected and observed heterozygosities, these values were closely

correlated. Only species for which data were available for at least four polymorphic loci were included. The mean number of polymorphic loci was 11.5 (median = 9.0; range = 4–106). The mean number of individuals analyzed was 96.0 (median = 44.0; range = 4–1,057). Neither the number of loci ($r = -0.046$; n.s.) nor the number of individuals ($r = 0.062$; n.s.) was significantly correlated with H .

Only loci polymorphic in the species under study were included in the analyses. In fact, most papers reporting heterozygosity at microsatellite loci report data only for polymorphic loci because the authors' interest is in providing tools for future ecological or behavioral work. If the authors did report data on invariant loci, those loci were not included in the computation of the mean value for the species in the present study. Primmer et al. (2005) found that cross-species microsatellite amplification and polymorphism are inversely related to evolutionary distance between avian species. In most cases analyzed here, the primers for the microsatellites were developed specifically for the species to which they were applied. In the few cases where data from cross-species primers were included, the primers were applied to closely related, congeneric species. Because only polymorphic loci were used, any primer set that did not amplify or did not show polymorphism was excluded.

H was used as dependent variable in analyses using a fully nested general linear models (GLM) approach. The super-ordinal relationships among avian orders were incorporated in GLM models (Grafen 1989) using the phylogeny of Livezey and Zusi (2007). This phylogeny was based on morphological data, permitting rooting of the tree with non-avian theropods; and it is consistent in most respects with published molecular phylogenies (Livezey and Zusi 2007). Applied to the orders of birds for which data were available in the present data set, this phylogeny comprised 7 nested levels (Fig. 1). Families, nested within orders, provided an eighth level. The model consisting of the 8 nested classificatory variables based on the phylogeny is referred to below as the “phylogeny model.”

Two additional categorical variables were used:

1. *IS* By this variable species were categorized as insular (having a breeding range restricted to one or more oceanic islands having a total land area less than or equal to that of Madagascar ($8 \times 10^5 \text{ km}^2$) or continental). In preliminary analyses, 6 species inhabiting large islands (greater than $1 \times 10^5 \text{ km}^2$ in total land area) were excluded from the analysis; but the results (not shown) were essentially the same as with the larger data set. Therefore only the latter results are reported below.
2. *END* By this variable species were categorized as *non-endangered*, *vulnerable*, or *endangered* according to information from IUCN website (<http://www.redlist.org>). The IUCN uses five categories to classify species, which I combined into three categories because some categories included only a small number of species in the data set; however, in preliminary analyses, the results were similar when categories were not combined or were combined in somewhat different ways from that reported below. My *non-endangered* category (219 species) corresponded to IUCN's “least concern.” In the category *vulnerable* (41 species) I combined IUCN's “near-threatened” (10 species) and “vulnerable” (31 species) categories. Likewise, in the category *endangered* (28 species), I combined IUCN's “endangered” (23 species) and “critically endangered” (5 species) categories.

In the GLM analysis, I used a model in which *IS* was nested within family, which was nested within the higher phylogenetic categories, and *END* was nested within *IS*. I refer to this below as the “full model.” Because this model was fully hierarchical, effects at any level of interest could be explored by examining residuals from the next higher order model; that is, a model including only the levels higher than the level of interest. A model that included only the phylogenetic categories is referred as the “phylogeny model”; the residuals from the

phylogeny model are equivalent to deviations from the family means. All means are presented plus or minus standard errors. Deleted-t residuals (equivalent to a test of the improvement in the fit of the model by removing an individual observation) were used to identify outliers from the full model.

Results

A fully hierarchical model based on phylogenetic variables, insular versus continental breeding distribution (*IS*), and endangerment (*END*) was used to predict micro satellite heterozygosity (*H*) in 294 species of birds, including 58 island endemic species (Supplementary Table S1). There were four significant phylogenetic effects.

1. There was a significant effect ($F_{1,160} = 4.90$; $P = 0.028$) at phylogenetic level 7 (as numbered in Fig. 1); that is, a difference in mean *H* between the two Cohorts of Neornithes (Livezey and Zusi 2007): Paleognathae (mean = 0.736 ± 0.069 SE; $N = 3$ species) and Neognathae (mean = 0.639 ± 0.009 SE; $N = 291$ species).
2. There was a significant effect ($F_{1,160} = 5.93$; $P = 0.016$) at phylogenetic level 3 (Fig. 1). Examination of residuals showed that this effect was due to a difference between the two sections within the subdivision Dendromithes (Livezey and Zusi 2007): Raptores (mean residual from the next higher order model = -0.098 ± 0.024 ; $N = 28$) and Anomalogonates (mean residual from the higher order model = 0.016 ± 0.011 ; $N = 169$). Raptores includes Falconiformes and Strigiformes, while Anomalogonates includes Cuculiformes, Psittaciformes, Coraciiformes, Piciformes, and Passeriformes (Fig. 1).
3. There was a significant effect ($F_{8,160} = 2.74$; $P = 0.007$) corresponding to a difference among orders within phylogenetic level 1 (Fig. 1). Examination of residuals showed that this effect was due to a difference between the two orders in the Galloanserae: Anseriformes (mean residual from the next higher order model = -0.075 ± 0.033 ; $N = 21$) and Galliformes (mean residual from the higher order model = 0.083 ± 0.031 ; $N = 19$).
4. There was a significant effect of families within orders ($F_{63,161} = 1.40$; $P = 0.049$); many differences among families contributed to this effect.

In the full model, there was a highly significant effect of *IS* within families ($F_{17,161} = 3.61$; $P < 0.001$). This effect was due to lower mean *H* in island species than in continental species (Fig. 2a). The mean residual from the phylogeny model (equivalent to the deviation from the family mean) was 0.015 ± 0.007 in the case of continental species, but -0.061 ± 0.018 in the case of insular species (Fig. 2a). In addition, there was a significant effect of *END* within *IS* ($F_{30,161} = 2.08$; $P = 0.002$) in the full model. The mean residual from a model including the phylogeny plus *IS* (the phylogeny + *IS* model) was 0.008 ± 0.006 for non-endangered species; 0.001 ± 0.018 for vulnerable species; and -0.059 ± 0.025 for endangered species (Fig. 2b).

In the full model, all phylogenetic effects accounted for 42.9% of the total variance; *IS* accounted for 13.4% of the total variance; and *END* accounted for 12.7% of the total variance. The entire model thus accounted for 67.2% of the variance in *H*. Significant negative deleted-t residuals identified 6 species with substantially lower *H* than predicted by the model (Table 1). Two of the 6 species were endangered, the Galápagos Penguin and the Crested Ibis; and the former species is also an island endemic (Table 1).

Discussion

Analysis of microsatellite heterozygosity (H) in 294 species of birds showed significantly reduced H in island endemics in comparison to mainland species. Because the long-term effective population sizes of island endemics are likely to be relatively small (Hughes and Hughes 2007), the results support the hypothesis that long-term effective population size can be an important causative factor behind differences among species with respect to microsatellite H . There was an independent association of current endangerment with reduced H , which accounted for a slightly smaller percentage of the overall variance in H than did island endemism. This finding supported the hypothesis that recent anthropogenic impacts have been a factor reducing genetic diversity of certain endangered species (Evans and Sheldon 2008). However, the effect was only seen in the species categorized as endangered, not those characterized as vulnerable.

In 89 avian species for which an estimate of the current population size was available, Evans and Sheldon (2008) found a positive correlation between microsatellite H and population size. In population genetics theory, heterozygosity is expected to be correlated with effective population size, not current census number (Nei 1987); but in general effective population size is probably roughly correlated with current census number (Nei and Graur 1984). Thus, this result of Evans and Sheldon (2008) is consistent with the hypothesis that microsatellite H is a function of effective population size. Only in species that have undergone substantial expansion in the past 10,000 years or so, would we expect effective population size, on the time scale relevant for microsatellite polymorphism, to be substantially lower than census number. The data of Evans and Sheldon (2008) show certain outliers with much lower microsatellite heterozygosity than expected by census number; and these species are likely to be those that have undergone a recent expansion.

Besides effective population size, the major other factor expected to contribute to microsatellite polymorphism is mutation rate. Mutation rate certainly differs among microsatellite loci (Brinkmann et al. 1998); but, since the present analyses were applied to the mean across several microsatellite loci, such differences may contribute relatively little to differences between avian lineages. In the case of mtDNA sequences, Nabholz et al. (2009) presented evidence that mutation rate differences among lineages, rather than differences in effective population size may account for differences among species with respect to diversity. For example, Nabholz et al. (2009) reported a lower mtDNA substitution rate in Falconiformes than in Piciformes or Passeriformes, and a lower mtDNA substitution rate in Anseriformes than in Galliformes. Both of the latter trends parallel significant phylogenetic effects on microsatellite H observed in the present analyses, suggesting that microsatellite mutation rates might be correlated with those of mtDNA.

On the other hand, ecological adaptations of major lineages may give rise to differences in long-term effective population size, and such ecological differences, rather than mutation rate differences, may account for the observed phylogenetic effects. For example, population sizes of top predators are generally low relative to prey species (Fa and Purvis 1997); and the lower H in Falconiformes and Strigiformes (diurnal raptors and owls) in comparison with the other orders of Dendroornithes (including parrots, rollers, woodpeckers, and perching birds) is consistent with lower long-term effective population sizes in the former. Similarly, the higher mean H in Galliformes (pheasants, quail, and their relatives) than Anseriformes (ducks, geese, and their relatives) may be due to the ecological difference between these orders. Because of their aquatic adaptation, Anseriformes may have restricted available habitat in comparison to the terrestrial Galliformes, resulting in reduced average effective sizes, an effect that may be magnified by a positive occupancy–abundance relationship (Holt et al. 2002). Thus, at present, it is not possible to determine whether the

observed phylogenetic effects are due to mutation rate differences, to differences in long-term effective population size resulting from distinct ecological adaptations, or to some combination of both factors.

The six species, two of them endangered, with significant negative deleted-t residuals provided examples of species with unexpectedly low H that was not well explained by the model (Table 1). One of the two endangered species, the Galápagos Penguin has probably always had a small population, restricted to two small oceanic islands, and its current population numbers in the thousands (Boersma 1998). Unlike most other severely endangered species in the data set (Supplementary Table S1), the Crested Ibis is not an island endemic; but there are only some 80 remaining members of this species, including those in captivity (Wingfield et al. 2000). The low H values of the four non-endangered species are less easily explained. The King Eider is a widespread species of the North American arctic; the fact that this species shows little spatial genetic structure (Pearce et al. 2004) suggests that it may have expanded since the last glaciation from a small ancestral population and thus may have a low long-term effective population size. The Long-tailed Manakin is a lek-forming species (Trainer and McDonald 1995), while both the Superb Starling (Rubenstein 2005) and the Southern Emu-wren (Maguire and Mulder 2004) are cooperative breeders. Thus all three of these species have unusual mating systems that are likely to reduce effective population size, because of unequal contributions of adults to breeding (Nei 1987).

The conclusion that island endemicity is associated with differences among species with respect to microsatellite H is consistent with the hypothesis that microsatellite heterozygosity is determined to a substantial extent by genetic drift, as expected under population genetics theory if microsatellite polymorphism is largely selectively neutral or nearly so (Nei 1987). Moreover, even though island endemics are particularly likely to be endangered (Hughes 2004; Purvis et al. 2000), the present results suggest that the impact of long-term effective population size on H at microsatellite loci is independent of the reductions in population size that have occurred in certain species as a result of recent anthropogenic endangerment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by grant GM43940 from the National Institutes of Health.

References

- Boersma PB. Population trends of the Galápagos Penguin: impacts of El Niño and La Niña. *Condor*. 1998; 100:245–253.
- Brinkmann B, Klitsch M, Neuhuber F, Hühne J, Rolf B. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet*. 1998; 62:1408–1415. [PubMed: 9585597]
- Ellegren H. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet*. 2004; 5:435–445. [PubMed: 15153996]
- Evans SR, Sheldon BC. Interspecific patterns of genetic diversity in birds: correlations with extinction risk. *Conserv Biol*. 2008; 22:1016–1025. [PubMed: 18616741]
- Fa JE, Purvis A. Body size, diet and population density in Afro tropical forest mammals: a comparison with neotropical species. *J Anim Ecol*. 1997; 66:98–112.

- Fisher DO, Owens IP. The comparative method in conservation biology. *Trends Ecol Evol.* 2004; 19:391–398. [PubMed: 16701291]
- Grafen A. The phylogenetic regression. *Phil Trans R Soc Lond B.* 1989; 326:119–157. [PubMed: 2575770]
- Holt AR, Gaston KJ, He F. Occupancy-abundance relationships and spatial distribution: a review. *Basic Appl Ecol.* 2002; 3:1–13.
- Hughes AL. Differential human impact on the survival of genetically distinct avian lineages. *Bird Conserv Int.* 1999; 9:147–154.
- Hughes AL. A statistical analysis of factors associated with historical extinction and current endangerment of non-passerine birds. *Wilson Bull.* 2004; 116:330–336.
- Hughes AL, Hughes MA. Coding sequence polymorphism in avian mitochondrial genomes reflects population histories. *Mol Ecol.* 2007; 16:1369–1376. [PubMed: 17391262]
- Kohn MH, Murphy WJ, Ostrander EA, Wayne RK. Genomics and conservation genetics. *Trends Ecol Evol.* 2006; 21:629–637. [PubMed: 16908089]
- Lande R. Risk of population extinction from fixation of deleterious and reverse mutations. *Genetica.* 1998; 102/103:21–27. [PubMed: 9766960]
- Livezey BC, Zusi RL. Higher-order phylogeny of modern birds (Teropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. *Zool J Linn Soc.* 2007; 149:1–95. [PubMed: 18784798]
- Lynch, M. The origins of genome architecture. Sinauer; Sunderland: 2007.
- Maguire GS, Mulder RA. Breeding biology and demography of the southern emu-wren (*Stipiturus malachurus*). *Aust J Zool.* 2004; 52:583–604.
- Nabholz B, Glémin S, Galtier N. The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *BMC Evol Biol.* 2009; 9:54. [PubMed: 19284537]
- Nei, M. Molecular evolutionary genetics. Columbia University Press; New York: 1987.
- Nei M, Graur D. Extent of protein polymorphism and the neutral mutation theory. *Evol Biol.* 1984; 17:73–118.
- Nei M, Maruyama T, Chakabarty R. The bottleneck effect and genetic variability in populations. *Evolution.* 1975; 29:1–10.
- Pearce JM, Talbot SL, Pierson BJ, Petersen MR, Scribner KT, Dickson DL, Mosbacj A. Lack of spatial genetic structure among nesting and wintering King Eiders. *Condor.* 2004; 106:229–240.
- Primmer CR, Painter JN, Koskinen MT, Palo J, Merilä J, Piironen J. Factors affecting avian cross-species microsatellite amplification. *J Avian Biol.* 2005; 36:348–360.
- Purvis A, Gittlemann JL, Cowlishaw G, Mace GM. Predicting extinction risk in declining species. *Proc R Soc Lond B.* 2000; 267:1947–1952.
- Rubenstein DR. Isolation and characterization of polymorphic microsatellite loci in the plural cooperatively breeding superb starling, *Lamprotornis superbus*. *Mol Ecol Notes.* 2005; 5:739–744.
- Schlötterer C. Evolutionary dynamics of microsatellite DNA. *Chromosoma.* 2000; 109:365–371. [PubMed: 11072791]
- Trainer JM, McDonald DB. Singing performance, frequency matching and courtship success of long-tailed manakins (*Chiroxiphia linearis*). *Behav Ecol Sociobiol.* 1995; 37:249–254.
- Wingfield JC, Ishii S, Kikuchi M, Wakabayashi S, Sakai H, Yamaguchi N, Wada M, Chikatsuji K. Biology of a critically endangered species, the Toki (Japanese Crested Ibis) *Nipponia nippon*. *Ibis.* 2000; 142:1–11.

Fig. 1. Relationships of avian orders used in data analysis, based on the phylogeny of Livezey and Zusi (2007). *Numbers* indicate hierarchical levels used in general linear model

Fig. 2.

a Box-and-whiskers plots showing mean (*solid circle*) and median (*horizontal line*) residual from the phylogeny model for species classified as continental or insular. **b** Box-and-whiskers plots showing mean (*solid circle*) and median (*horizontal line*) residual from the phylogeny + *IS* model for species classified as non-endangered, vulnerable, or endangered. In each case, numbers of species are indicated. The *box* indicates the first quartile (Q1) and third quartile (Q3), and the whiskers the range from $Q1 - 1.5(Q3 - Q1)$ to $Q3 + 1.5(Q3 - Q1)$

Table 1
Species with significant negative deleted-t residuals from full model

Species	<i>H</i>	IUCN category	Deleted-t residual
Galápagos Penguin <i>Spheniscus mendiculus</i>	0.063	Endangered	-2.02 [*]
Crested Ibis <i>Nipponia nippon</i>	0.121	Endangered	-3.62 ^{***}
King Eider <i>Somateria spectabilis</i>	0.310	Least concern	-2.19 [*]
Long-tailed Manakin <i>Chiroxiphia linearis</i>	0.402	Least concern	-2.50 [*]
Southern Emuwren <i>Stipiturus malachurus</i>	0.558	Least concern	-2.14 [*]
Superb Starling <i>Lamprolornis superbus</i>	0.421	Least concern	-2.40 [*]

Significance of deleted-t residual:

* $P < 0.05$;

*** $P < 0.001$