PROCEEDINGS OF THE ROYAL SOCIETY B

Proc. R. Soc. B (2006) 273, 1049–1053
doi:10.1098/rspb.2005.3425

Are subspecies useful in evolutionary and conservation biology?

Albert B. Phillimore* **and Ian P. F. Owens**

Division of Biology and NERC Centre for Population Biology, Imperial College London, Silwood Park, Ascot, Berkshire SL5 7PY, UK

The taxonomic rank of subspecies remains highly contentious, largely because traditional subspecies boundaries have sometimes been contradicted by molecular phylogenetic data. The most complete meta-analysis to date, for instance, found that only 3% of traditional avian subspecies represented distinct phylogenetic lineages. However, the global generality of this phenomenon remains unclear due to this previous study's narrow geographic focus on continental Nearctic and Palearctic subspecies. Here, we present a new global analysis of avian subspecies and show that 36% of avian subspecies are, in fact, phylogenetically distinct. Among biogeographic realms we find significant differences in the proportion of subspecies that are phylogenetically distinct, with Nearctic/Palearctic subspecies showing significantly reduced levels of differentiation. Additionally, there are differences between island and continental subspecies, with continental subspecies significantly less likely to be genetically distinct. These results indicate that the overall level of congruence between taxonomic subspecies and molecular phylogenetic data is greater than previously thought. We suggest that the widespread impression that avian subspecies are not real arises from a predominance of studies focusing on continental subspecies in North America and Eurasia, regions which show unusually low levels of genetic differentiation. The broader picture is that avian subspecies often provide an effective short-cut for estimating patterns of intraspecific genetic diversity, thereby providing a useful tool for the study of evolutionary divergence and conservation.

Keywords: subspecies; birds; phylogenetics; monophyly; islands

1. INTRODUCTION

The taxonomic rank of subspecies has been the subject of long-running controversy (Mayr 1982), primarily because of the failure of some molecular studies to identify traditional subspecies as phylogenetically distinct (Barrowclough 1980; Mayr & Ashlock 1991; O'Brien & Mayr 1991; Ball & Avise 1992; Burbrink *et al.* 2000). Traditionally, subspecies have been recognized on the basis of discontinuities in the geographical distribution of phenotypic traits (Mayr & Ashlock 1991). However, several high-profile studies have revealed a mismatch between such phenotypically defined subspecies and phylogenetic clusters identified using modern molecular methods (Barrowclough 1980; Ball & Avise 1992; Zink 1996; Burbrink *et al.* 2000). A particularly influential case of this concerns the seaside sparrow (*Ammodramus maritimus*) of North America. This species is an excellent example of morphological variation within a single species, with traditional taxonomy recognizing a large number of morphological subspecies, including the, now extinct, dusky seaside sparrow (*A. m. nigrescens*) (Clements 2000). However, molecular phylogenetic analysis has suggested that *A. m. nigrescens* did not represent a distinct evolutionary lineage (Avise & Nelson 1989). Thus, the conservation efforts directed at the dusky seaside sparrow constitute a vivid illustration that traditional subspecies nomenclature can provide a misleading impression of the true geographical pattern of intraspecific differentiation and can arguably misdirect conservation effort (Zink 2004). Indeed, the only previous meta-analysis of the extent of congruence between traditional subspecies boundaries and molecular phylogenies found that only 3% of traditional avian subspecies were distinct phylogenetic units and concluded that the subspecies rank 'continue[s] to hinder progress in taxonomy, evolutionary studies and especially conservation' (Zink 2004).

But how general is the finding that only 3% (7 out of 230) of avian subspecies are evolutionarily distinct lineages? At present this is not clear, because the previous meta-analysis was almost entirely restricted to subspecies from just two biogeographic realms, the Nearctic and Palearctic, with over 95% of case studies coming from these regions (220 out of 230 subspecies, across 41 species) (Zink 2004). Furthermore, the previous meta-analysis focused almost exclusively on continental subspecies, whereas we know that a sizeable proportion of the world's avian subspecies are island-dwelling (Clements 2000). From the evidence presented to date, therefore, it is impossible to know whether the very low level of genetic distinctiveness reported for avian subspecies in the previous meta-analysis is a general phenomenon, or an idiosyncrasy of the continental birds of North America and Eurasia.

The overall aim of this study was to re-address the question of the extent of congruence between traditional avian subspecies and molecular phylogenetics (derived

---

* Author for correspondence (albert.phillimore@imperial.ac.uk).

The electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2005.3425 or via http://www.journals.royalsoc.ac.uk.

Received 25 October 2005
Accepted 24 November 2005

from mitochondrial sequence data) on a global scale. We therefore collated a new database on avian phylogenetic studies that incorporated taxa from diverse biogeographic realms and included both continental and island-dwelling subspecies. We then used this database to test (i) the overall proportion of traditional subspecies that are phylogenetically distinct, (ii) whether there are consistent differences among biogeographic realms, or between continental and island-dwelling taxa, in the proportion of subspecies that are phylogenetically distinct, and (iii) whether methodological factors may explain the apparent differences among studies in the proportion of subspecies that are phylogenetically distinct. We used birds as a model system for this study because of the relative maturity of the taxonomy for this class, the large number of phylogenetic studies available in the literature, and the prominence of this group in the long-running debate regarding the utility of the subspecies rank (Barrowclough 1980; Mayr & Ashlock 1991; Ball & Avise 1992; Burbrink *et al.* 2000; Zink 2004).

2. MATERIAL AND METHODS

(a) *Congruence*

Following the previous meta-analysis of birds (Zink 2004), we tested for congruence between traditional taxonomic subspecies and molecular phylogenetic studies by examining whether traditional avian subspecies were monophyletic. Phylogenetic monophyly of a subspecies was defined as; the situation when all individuals of a particular subspecies were more similar to one another in terms of mitochondrial DNA haplotype than they were to any other sampled individuals of other subspecies (Moritz 1994). The minimum criteria for including a particular subspecies in the test of monophyly were, first, that more than one subspecies within a single species was sampled in the phylogeny, and second, that more than one individual of the particular subspecies was sampled in the phylogeny.

(b) *Data collection*

Our nomenclature for avian subspecies followed a traditional ornithological taxonomy (Clements 2000). Information on molecular phylogenetic hypotheses of the relationships among subspecies were obtained from the literature (see electronic supplementary material). To minimize potential sources of bias, we adopted several criteria for data collection. First, we only considered phylogenies that were reconstructed from sequence data (subsequently referred to as 'molecular' phylogenies) and were published, or in press, from 2001 onwards. This cut-off date was chosen to ensure that the taxonomy used here (Clements 2000), for identifying traditional avian subspecies, pre-dated the phylogenetic information. Second, where the subspecific boundaries adopted by the author(s) of a molecular phylogeny deviated from those recognized by our chosen taxonomy, we adopted those of the latter to ensure consistency across studies in our database. Third, polyphyletic species were excluded from the database, except in cases of paraphyly. In cases of paraphyly only the larger monophyletic portion of the species phylogeny was included, unless, however, the embedded non-conspecific represented a peripheral isolate, such as an island species; in which case we considered all of the subspecies within the paraphyletic species. Finally, where multiple molecular phylogenies were presented in a publication, we selected the phylogeny to be used in our analyses based on two criteria. First, we selected the tree(s) that included most individuals, and second, we selected trees according to the method of phylogenetic tree reconstruction (following the preference order: Bayesian > maximum likelihood > maximum parsimony > neighbour joining).

To test whether the proportion of subspecies that were monophyletic differed between realms we recorded the biogegraphical realm(s) (Olson *et al.* 2001), in which each included subspecies was found. If an included subspecies' geographical range overlapped with the nearctic and palearctic realms, the parent species was classified as being in the nearctic/palearctic category, even when other daughter subspecies also occurred in other realms. Similarly, to test for differences in the incidence of monophyly on continents versus islands we collected data on the number of the sampled subspecies within a species that were restricted to islands (defined according to Dahl 1991).

Finally, to test whether other methodological differences between studies may influence the proportion of subspecies that were monophyletic, we also recorded the average number of individuals sampled for each included subspecies within a species, and the average number of localities sampled per subspecies within a species. Sampling localities were considered separate if they were 100 km or more apart, or if they were separated from other sampling localities by a water barrier. This definition of sampling localities was intended to facilitate a meaningful count of the number of populations sampled, while minimizing inflation caused by dense local sampling.

(c) *Univariate analyses*

We used a χ^2 contingency tables test with Yates's correction to test whether the proportion of subspecies that were monophyletic differed, first, between the nearctic/palearctic realms and the other biogeographic realms, and second, between continental and island subspecies.

(d) *Multivariate analyses*

The univariate approach treated subspecies as independent data points, whereas they actually represent pseudo-replicates in terms of both phylogenetic non-independence and phylogenetic methodology. To ensure that our results were not an artefact of such phylogenetic non-independence we subsequently adopted a multivariate approach that explicitly addressed the problem of intraspecific pseudo-replication by modelling the presence and absence of monophyly within a species as a binomial response. We used generalized linear models (GLMs) to test the ability of biogeographic realms, island-dwelling, taxon sampling and geographical sampling intensity to explain variation in the proportion of subspecies that were monophyletic. As the dependent variable was a proportion and was over-dispersed, we employed a quasi-binomial error structure. We built multivariable models and tested the significance of terms via model comparisons using ANOVA tables with *F*-tests (Crawley 2002). All analyses were conducted in the R environment (R Development Core Team 2004).

Initially, we built a maximal multivariate model that included all terms as covariates. Subsequently, however, we used a stepwise procedure to reduce this to a minimal adequate model. Model simplification was conducted via sequential deletion of the terms that were found to be least significant using ANOVA tables and *F*-tests (Crawley 2002).

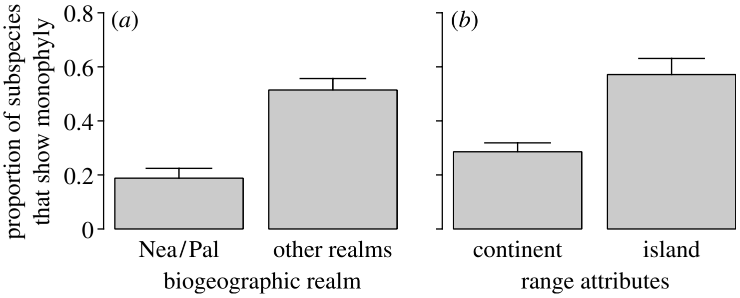

Figure 1. (*a*) Barplot showing the proportion of subspecies that are monophyletic in the nearctic/palearctic realms versus the rest of the world. (*b*) Barplot showing the proportions of island and continental subspecies that are monophyletic. Error bars indicate standard errors. (Nea/Pal refers to the nearctic and palearctic realms.)

Terms that had a *p*-value less than 0.1 were retained in the minimum adequate model.

Model fit was measured in the form of a pseudo-r^2, according to equation (2.1) (Maddalla 1983), wherein dev_{res} and dev_{null} are the residual and null deviance of the model, respectively, and n is the sample size

$$r^2 = \frac{(1-\exp((\text{dev}_{\text{res}}-\text{dev}_{\text{null}})/n))}{(1-\exp(-\text{dev}_{\text{null}}/n))}. \tag{2.1}$$

3. RESULTS

(a) *Congruence*

In total we collated molecular phylogenetic data on 259 avian subspecies belonging to 67 species (see electronic supplementary material). Across all biogeographic realms, and including both continental and island-dwelling taxa, a total of 94 (36%) of these subspecies exhibited monophyly.

(b) *Univariate analyses*

When we tested for differences among biogeographic realms (figure 1*a*), we found that in the nearctic/palearctic realms only 19% of subspecies (22 out of 117) were monophyletic, whereas in the rest of the world this value increased to 51% (72 out of 140 subspecies). This disparity between realms, in terms of the proportion of subspecies that were found to be monophyletic, was highly significant ($\chi_1^2=27.9$, $p<0.001$). Similarly, when we tested for differences between continental and island-dwelling subspecies (figure 1*b*), monophyly was higher among island subspecies (57% of subspecies, 40 out of 70) than among continental subspecies (29% subspecies, 54 out of 189), which again proved significant ($\chi_1^2=16.8$, $p<0.001$).

(c) *Multivariate analyses*

The maximal GLM, which included all potential explanatory variables, explained 56% of the variation among studies in the proportion of subspecies that were monophyletic. Both biogeographic realm and island-dwelling were significant terms in this model (table 1). However, there was no robust evidence that density of taxon sampling or density of spatial sampling predicted the extent to which subspecies were found to be monophyletic.

When we used a stepwise deletion procedure to construct a minimum adequate model, we found that biogeographic realm and island-dwelling were the only terms retained in this model (table 2). Again both terms were highly significant, with those subspecies found on islands outside the nearctic and palearctic realms exhibiting the highest levels of monophyly (figure 2).

Table 1. Maximal multivariate model to explain differences among species in the proportion of subspecies that are monophyletic. ($n=61$, Pseudo-$r^2=0.56$ Coefficients and standard errors are reported on a logit scale.)

independent variable	coefficient	s.e.	*t*-value	*p*-value
intercept	−0.89	0.99	−0.89	0.38
taxon sampling	−0.96	0.86	−1.12	0.27
spatial sampling	0.15	0.52	0.29	0.77
island-dwelling	1.42	0.57	2.46	<0.05
region	1.25	0.49	2.52	<0.05

Table 2. Minimum adequate model to explain differences among species in the proportion of subspecies that are monophyletic. ($n=61$, pseudo-$r^2=0.56$ Coefficients and standard errors are reported on a logit scale.)

independent variable	coefficient	s.e.	*t*-value	*p*-value
intercept	−1.94	0.39	−4.94	<0.001
island-dwelling	1.48	0.55	2.67	<0.05
region	1.45	0.45	3.20	<0.01

4. DISCUSSION

Our results show that, across all biogeographic realms and including both continental and island-dwelling taxa, approximately 36% of traditional avian subspecies are phylogenetically distinct. This global estimate is noteworthy because it is an order of magnitude higher than the estimate obtained by the previous meta-analysis (Zink 2004). The major explanation for the discrepancy between our results and those obtained in the previous meta-analysis appears to be that the previous study was almost entirely restricted to continental subspecies from two biogeographic realms, the nearctic and palearctic. Our analyses show that continental subspecies sampled from the nearctic and palearctic realms are significantly less likely to be monophyletic than subspecies found on islands and/or outside these realms. Both the difference between realms and the positive correlation with island-dwelling are significant in all models, and these patterns are not attributable to either intraspecific phylogenetic non-independence or to differences between studies in the intensity of taxon or geographical sampling.

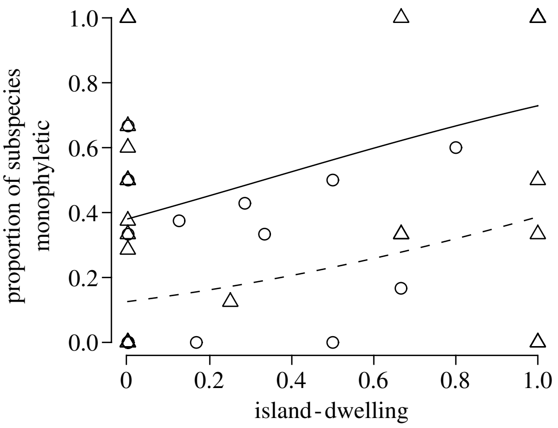

Figure 2. Relationship between island-dwelling and the proportion of subspecies that are monophyletic, shown separately for different biogeographic realms. Island-dwelling is measured as the proportion of the subspecies within a species that are found solely on islands. Data points derived from the nearctic/palearctic realm are denoted with circles, while those from the rest of the world are denoted with triangles. The dotted and solid lines represent the predicted slopes (back-transformed from logits) for the nearctic/palearctic realms and other realms respectively.

These results pose the question, why is the level of genetic distinctness among subspecies higher on islands, and outside the nearctic and palearctic realms? The answer to the first element of this question, regarding islands, is relatively straightforward. It has long been established that populations on islands encounter a physical impediment to gene-flow between populations, and it is therefore expected that such populations may diverge in isolation (Mayr 1963). Furthermore, island populations tend to be smaller than those on continents, meaning that the fixation of neutral genes is likely to take place with greater rapidity on islands. Continental subspecies, on the other hand, will often have geographical ranges that directly abut, or even overlap, those of conspecific subspecies, and thus any phenotypic adaptation to local environments will need to take place in the face of gene flow. Theory predicts that, in the absence of pre- or post-zygotic isolation, even a small amount of on-going gene flow may reduce the rate of divergence among populations so, all other things being equal, continental subspecies should diverge at a slower rate than their island-dwelling counterparts (Hastings & Gavrilets 1999; Gavrilets 2004). The reasons for the reduced levels of monophyly among palearctic and nearctic subspecies are less obvious. One possibility, of course, is that the implementation of subspecies status differs between regions. However, an alternative biological explanation may be that subspecies in the nearctic and palearctic realms are relatively young due to post-glacial re-colonization (Pianka 1966). We therefore propose that subspecies from the nearctic and palearctic tend to show a relatively low level of phylogenetic differentiation because, compared to other biogeographic realms, insufficient time has elapsed for coalescence to have taken place.

Although this study has revealed much higher levels of phylogenetic differentiation than expected from previous studies, there are reasons to think that even these could be underestimates. This is because we have followed previous studies in using monophyly as our index of the phylogenetic distinctiveness of subspecies (Zink 2004), which may be a rather conservative index of population differentiation for several reasons. Mitochondrial DNA requires $1.8\times Ne$ generations (Ne=effective population size) for there to be a 95% probability that an isolated population is identified as being monophyletic (Hudson & Coyne 2002). Processes such as historical-introgression and incomplete lineage-sorting may generate incongruence between the actual divergence of populations and the gene tree recovered (Funk & Omland 2003). Under such conditions the likelihood of finding reciprocal monophyly in recently divergent populations will be reduced. Where a divergent island population is derived from a small part of a widespread continental species it is possible that the island subspecies may have achieved monophyly, while the continental source subspecies may not have. This arises because individuals from the source population may be more genetically similar to the now separate island population than they are to other continental populations (Hudson & Coyne 2002). Moreover, from a population genetic perspective, rather than a phylogenetic perspective, populations can undergo genetic divergence despite the ameliorating influence of gene flow (Gavrilets *et al.* 2000). Indeed, instances of populations diverging in the face of gene flow are starting to emerge from recent empirical studies (Garant *et al.* 2005; Postma & Van Noordwijk 2005). It will therefore be interesting to see whether future studies based on population genetic measures of population differentiation detect even higher rates of divergence among traditional subspecies.

Whether our analyses of bird subspecies are indicative of the general pattern in other taxonomic groups remains to be seen. Nevertheless, we conclude that, in the case of birds, the low degree of phylogenetic distinctness among continental subspecies from North America and Eurasia should not obscure the general picture that, in other regions, more than 50% of traditional subspecies are phylogenetically distinct. Our results suggest that in many circumstances subspecies are likely to prove useful in estimating the historical patterns of divergence among populations (O'Brien & Mayr 1991). Finally, therefore, we propose that subspecies may, in fact, be of considerable conservation utility, as proxies for the sub-structure found within species. We suggest that the conservation utility of subspecies is likely to be greatest in situations where molecular data is absent, a scenario that is likely to be most frequently encountered in the Equatorial and Southern Hemisphere regions.

We thank T. Barraclough, N. Collar, C. Godfray, J. Hadfield, S. Meiri, D. Orme, A. Purvis, T. Price, D. Quicke, G. Thomas, R. Zink and three anonymous reviewers for discussion or comments. This work was supported by the Natural Environment Research Council (UK).

REFERENCES

Avise, J. C. & Nelson, W. S. 1989 Molecular genetic relationships of the extinct dusky seaside sparrow. *Science* **243**, 646–649.

Ball, R. M. & Avise, J. C. 1992 Mitochondrial DNA phylogeographic differentiation among avian populations and the evolutionary significance of subspecies. *Auk* **109**, 626–636.

Barrowclough, G. F. 1980 Genetic and phenotypic differentiation in a wood warbler (genus *Dendroica*) hybrid zone. *Auk* **97**, 655–668.

Burbrink, F. T., Lawson, R. & Slowinski, J. B. 2000 Mitochondrial DNA Phylogeography of the polytypic North American rat snake (*Elaphe obsoleta*): a critique of the subspecies concept. *Evolution* **54**, 2107–2118.

Clements, J. F. 2000 *Birds of the World: a checklist*. Vista, CA: Ibis.

Crawley, M. J. 2002 *Statistical computing: an introduction to data analysis using S-plus*. Chichester, UK: Wiley.

Dahl, A. L. 1991 *Island directory. UNEP regional seas directories and bibliographies*, No. 35, pp.573. Nairobi, Kenya: UNEP

Funk, D. J. & Omland, K. E. 2003 Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal DNA. *Annu. Rev. Ecol. Evol. Syst.* **34**, 397–423. (doi:10.1146/annurev.ecolsys.34.011802.132421)

Garant, D., Kruuk, L. E. B., Wilkin, T. A., McCleery, R. H. & Sheldon, B. C. 2005 Evolution driven by differential dispersal within a wild bird population. *Nature* **433**, 60–65. (doi:10.1038/nature03051)

Gavrilets, S. 2004 *Fitness landscapes and the origin of a species*. Monographs in population biology. Princeton, NJ: Princeton University Press.

Gavrilets, S., Li, H. & Vose, M. D. 2000 Patterns of parapatric speciation. *Evolution* **54**, 1126–1134.

Hastings, A. & Gavrilets, S. 1999 Global dispersal reduces local diversity. *Proc. R. Soc. B* **266**, 2067–2070. (doi:10.1098/rspb.1999.0888)

Hudson, R. R. & Coyne, J. A. 2002 Mathematical consequences of the genealogical species concept. *Evolution* **56**, 1557–1565.

Maddalla, G. S. 1983 *Limited-dependent and qualitative variables in econometrics*. Cambridge, UK: Cambridge University Press.

Mayr, E. 1963 *Animal species and evolution*. Cambridge, MA: Harvard University Press.

Mayr, E. 1982 Of what use are subspecies? *Auk* **99**, 593–595.

Mayr, E. & Ashlock, P. D. 1991 *Principles of systematic biology*. New York: McGraw-Hill.

Moritz, C. 1994 Defining evolutionary significant units for conservation. *Trends Ecol. Evol.* **9**, 373–375. (doi:10.1016/0169-5347(94)90057-4)

O'Brien, S. J. & Mayr, E. 1991 Bureaucratic mischief: recognizing endangered species and subspecies. *Science* **251**, 1187–1188.

Olson, D. M. *et al.* 2001 Terrestrial ecoregions of the worlds: a new map of life on Earth. *Bioscience* **51**, 933–938.

Pianka, E. R. 1966 Latitudinal gradients in species diversity: a review of concepts. *Am. Nat.* **100**, 65–75. (doi:10.1086/282398)

Postma, E. & Van Noordwijk, A. J. 2005 Gene flow maintains a large genetic difference in clutch size at a small spatial scale. *Nature* **433**, 65–68. (doi:10.1038/nature03083)

R Development Core Team 2004 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Zink, R. M. 1996 Comparative phyloeography in North American birds. *Evolution* **50**, 308–317.

Zink, R. M. 2004 The role of subspecies in obscuring avian biological diversity and misleading conservation policy. *Proc. R. Soc. B* **271**, 561–564. (doi:10.1098/rspb.2003.2617)