

REVIEW: Report by R.R. Ramey on *Zapus* subspecies

General problems:

- The overall tone of the manuscript lacks objectivity. Conclusions would be more convincing if data and results were presented from a less biased perspective.
- The report lacked context beyond the *Z.h.preblei* problem – the issues dealt with in this report (i.e., recent diversification of lineages) are complex and there is considerable literature available on the topic. Yet none of this was discussed.
- The molecular data are quite limited (only 355 base pairs of sequence) and these provide insufficient resolution. Thus, results are inconclusive.
- Criteria used for exclusion of particular specimens are rather unclear and seemingly subjective.
- Presentation of data is confusing and lacks sufficient and necessary detail. There are numerous typos that speak of haste in preparation.
- Manuscript is rather ambiguous with regard to various descriptions.
- Authors equate their results (a gene tree at best) with a species tree.
- Comments below often refer to (page/paragraph/line number) of the report.

1. Techniques, phylogenetic evaluation. Appropriate methods and markers?

Yes and no. The overall approach seems appropriate. However, there are a number of issues that are not addressed in the manuscript, some of which may substantially impact results. Details that support results and conclusions are lacking. Additional and appropriate analyses could have been performed for both molecules and morphology.

Molecular work:

DNA extraction was appropriate. DNAeasy Kit from Qiagen is known to produce clean DNA from difficult samples. PCR amplifications are appropriate, although 200-300 ng of DNA in a 25 ul reaction volume seems large, but might be necessitated by low quality (i.e., low molecular weight) of extracted DNA. It is also appropriate to sequence the target region in both directions, obtaining sequences for both strands.

Museum specimens are indeed a valuable resource in that specimens from a large geographic area can be made available, and a study can thus be executed in relatively short time. However, the quality of DNA extracted from museum specimens is often inferior. It is most often fragmented and consists of small pieces in low quantity that

makes it difficult to reliably amplify target sequences (Pääbo 1989). Thus, primers should be designed from sequences obtained from fresh tissues, and must be selected to produce short products approximately 100—200 base pairs in length (as per Drew et al. 2003). The problem is that this procedure requires up to 5x more PCR amplifications than normal, which in turn increases cost and reduces sample sizes. However, it does sample a large number of overall base pairs, which is important. This apparently was not done in the present study.

Authors allude to some of these problems (page 6) and in fact developed internal primers to amplify difficult samples using nested PCR. However, cross-contamination is an issue with such “ancient DNA” samples. While DNA can be amplified from minute amounts of tissue using forensic techniques, contamination of such templates with high quality DNA from other samples is a major concern. Another aspect is that PCR can incorporate the wrong base during replication; if such a mistake is incorporated early in PCR cycles, it will be reproduced in all subsequent cycles. Again, this is a major issue for ancient DNA samples where little template DNA is available to start the reaction. One way to address this issue would be to generate independent replicate amplifications/sequences of samples so as to calculate genotyping error. This is particularly important if haplotypes differ by only a single base, as is seemingly the case with the present data set.

One must assume that authors took all necessary precautions to avoid contamination of their samples, but it would certainly be more convincing if indeed they explicitly stated in their lab procedures the manner in which they dealt with these issues.

Descriptions are often vague. For example, authors state (6/2/1-2) that some DNA extracts did not amplify well, but there is no information on how many? It is also not clear how they could get amplified DNA when the initial PCR did not amplify at all (6/2/2)?

Another more substantial problem encountered when working with ancient DNA is the size of fragments (i.e., numbers of base pairs) that can be reliably amplified and sequenced (as noted above). The current analysis is based on 355 base pairs (bp) of sequence data - this is a marginal data set for population-level analyses judging from today's standards. As a general rule, at least 1,000 bp should be evaluated to substantiate findings and make results conclusive. An analysis of several independent molecular markers that corroborate findings would also make the study more convincing.

The control-region (or D-loop) is generally a good marker to examine recently diverged taxa because it has a high rate of evolution. Presumably *Z.h.preblei* became isolated post-pleistocene (6/2/9), yet a time span of 10,000 years is about the limit for mtDNA resolution. Taxa that are more recent diverged would be difficult to detect via mtDNA analysis. The control-region does not code for a protein which explains its fast rate of evolution, but this is also a drawback in that it limits the types of analyses that can be done. For example, those that rely on codon position cannot be utilized.

In addition, it is not at all clear which section of the control region was sequenced. Here, site positions should be provided relative to a standard sequence available from GenBank. The control-region consists of rather variable segments at the 5' and 3' end, and a relatively conserved middle region. Since nested primers are designed internally from the flanking regions, it is likely that the region sequenced in this study straddles the conserved middle segment, and thus encapsulates only a moderate amount of genetic diversity. This is an issue in that the analyzed fragment does not provide sufficient resolution to determine interrelationships of the taxa under study (see below). The question then becomes, is the lack of variation due to similarities among OTUs, or is it instead a function of the conservative (and limited) nature of the molecular marker?

Molecular data analyses

Sample sizes are appropriate (Crandall et al. 2000). Phylogenetic analyses seem appropriate. Standard procedures were used to generate phylogenies, and data were first examined via ModelTest so as to determine the model of sequence evolution. However, details are lacking for the AMOVA. Were alternative genetic structures tested for significance?

In addition, the text is confusing and it is not at all clear how many samples were indeed used for analyses. On page 6 (6/1/1/), authors state that 151 sequences were aligned, whereas the heading for Fig. 2 indicated that analyses are based on 176 samples. The sum of all specimens listed per haplotype for the ingroup (i.e., *Z.hudsonius*) is 151.

Also, the basis for exclusion of specimens that showed haplotype characteristic of other subspecies is rather vague (page 8). Why is it reasonable to assume that those *Z.h.campestris* with haplotype L/Pa/C-2 were misidentified and can thus be excluded from analyses, whereas those *Z.h.campestris* showing haplotype C/P-1 through 4 are not? Details on collection data are also lacking. How reliable are the locality definitions? Further, why wasn't the identification of these specimens confirmed by re-examination? Authors state that this is a strong suit of voucher specimens.

Why is a Neighbor joining phylogram presented, instead of an MP, ML or BA tree? Authors state that other analysis produced "similar" trees, but phylograms of these should be provided so that tree topology and nodal support can be examined.

A haplotype network or minimum spanning tree of haplotypes would also be informative. The shallow terminal branches of the phylogram suggest that haplotypes differ by single base pairs. Further, haplotype diversity statistics and an appendix showing haplotypes/variable sites should be provided (see also comment under Point 3).

It is not clear which samples were used to calculate nucleotide diversity. High nucleotide diversity in *Z.h.campestris* and *Z.h.preblei* could be due to divergent and "mis-identified"

individuals, as suggested by the high standard deviation (8/5). Again, it's not clear whether these samples were included in the calculation of nucleotide diversity.

Morphometric analyses

There are several perceived difficulties with the morphometric analyses.

(10/2/3) – “In several cases, fewer measurements were taken because of breakage or not taken because of previous breakage.” These should be enumerated in the report, along with the museum numbers of the specimens excluded. If this is not done, then how will another scientist be able to replicate the study?

Discriminant analysis is an inappropriate multivariate procedure for this study in that it requires that specimens be *a priori* allocated to group. If indeed one is testing for group membership, then pre-allocation to group biases the study. Given the ambiguity of specimen assignment in the molecular analyses, a more effective means of evaluation would have been a principal components analysis of morphometric data based on the variance-covariance matrix. Data should also be first tested for normality.

(10/3/10—11) – “Males and females were pooled in the analyses because of a lack of cranial sexual dimorphism in *Z.princeps* and *Z.hudsonius* (Connor and Shenk 2003).” However, Connor and Shenk evaluated *Z.h.preblei* and *Z.p.princeps* whereas the current report evaluated *Z.h.preblei* and *Z.h.campestris*. If indeed the object is to test the sexes of two subspecies for potential differences in morphology, then one should not apply as the test those results previously generated for different subspecies. Additionally, if the object is to evaluate group membership (in this case, sex) using morphological criteria, then pre-allocation to group would again bias the analysis (as per caveats regarding discriminant analysis above).

It is also somewhat confusing that other subspecies of *Z.hudsonius* were not examined morphologically as well. And since the researchers went to all the trouble to measure their specimens, why did they not take other (additional) standard morphometric measures?

(11/1/8) – “...only larger for one measurement...and it was only marginally significant (P=0.037).” Again, a value judgement that undermines the objectivity of the study.

2. Are conclusions about taxonomic validity of *Z.h.preblei* logical and defensible?

I personally cannot follow the logic. If *Z.h.preblei* and *Z.h.campestris* should be synonymized based on shared haplotypes, then other *Z.hudsonius* subspecies must be synonymized as well. This logic could even be extended to *Z.hudsonius* and *Z.p.princeps*, since haplotypes of the latter were found within *Z.hudsonius*. This suggests either a very complex taxonomic problem confounded by quite recent (i.e.,

post-Pleistocene) diversification, or a problem with the resolution of the molecular marker (as above).

The limitations of the data affect resolution of analyses and thus render results inconclusive. Relationships among haplotypes are not (or only poorly) resolved in the neighbor joining tree (Fig. 2). Additional sequence data from fast evolving, independent markers are needed (as recommended by Haig 1998). Data based on a different marker might still remain incongruent, but that in itself reveals important aspects of the phylogenetic history of a species (Hey et al. 2003).

Not clear why *Z.p.princeps* was selected as outgroup. The work by J. Cook is unpublished and thus unavailable for evaluation. Monophyly does not render one taxa as outgroup for another.

It is not clear why hybridization between *Z.h.preblei* and *Z.p.princeps* should invalidate the taxonomic status of *Z.h.preblei* (3/4/5-7)? The biological species concept (BCS) uses reproductive isolation as a criterion of demarcation, but it is generally recognized that the ability to hybridize is a pleisomorphic (i.e., ancestral) trait that offers little with regard to recent diversification of species.

Museum collection data (e.g., date of collection, precise collection locality etc.) should be provided in an Appendix. This would clarify the validity of original identifications and also provide further information about DNA quality in that reviewers could judge the ages of various samples.

Conclusions based on AMOVA are not justified. High percentage of within vs among subspecies diversity is influenced by resolution of the marker and demographics of the population (e.g., bottlenecks, population fluctuations, effective population size, etc.).

The criterion of “greater genetic diversity among putative taxa than within” (8/5/1—4) is a flawed concept. Genetic diversity is dependent on population size and population history. Paetkau (1999) emphasized that population demographics do influence retention of genetic diversity, including ancestral haplotypes and time to complete lineage sorting (or reciprocal monophyly).

3. Alternative interpretations for genetic data?

Identical haplotypes in *Z.h.preblei* and *Z.h.campestris* could be explained by:

- Retention of ancestral polymorphism and incomplete lineage sorting
- Homoplasy (similar character state but independent evolutionary origin)
- Genotyping error

Pattern could also be explained by retention of ancestral polymorphism in *Z.h.preblei* and *Z.h.campestris*. In other words, the detected variation stems from mutations that occurred prior to their divergence, and which is still retained in both subspecies. The marker does not provide enough resolution to differentiate the two lineages and the data suggest incomplete lineage sorting at this level of resolution. A more extensive data set based on markers with appropriate evolutionary rate might reveal additional mutations that are not shared between the two subspecies. It is important to note that there is a clear frequency difference regarding the identities of individuals contained in the four *Z.h.preblei* haplotypes (C/P1—4). Only one or two specimens of *Z.h.campestris* are found within these haplotypes, with the majority of individuals (90%, 89%, 86% and 82%) being *Z.h.preblei*. If indeed *Z.h.preblei* simply represents a recent range extension of *Z.h.campestris*, then one would expect to find *Z.h.preblei* haplotypes scattered throughout the *Z.h.campestris* clade. The fact that *Z.h.preblei* haplotypes cluster together suggests they diverged from one another and thus underscores the argument that *Z.h.preblei* might be on its own evolutionary trajectory.

Paetkau (1999) raised concerns about applying purely genetic identification criterion for an ESU and pointed out that population demographics do indeed influence time to complete lineage sorting. Even low frequency haplotypes are retained in large populations over sustained time, whereas small populations lose genetic diversity randomly through drift.

An alternative although less likely hypothesis is that the shared haplotypes arose independently in both subspecies thus representing homoplasy rather than homology.

As alluded to above, genotyping error is a concern with museum-based molecular studies. Independent replicates (including DNA extraction, amplification and sequencing) would corroborate findings or reveal genotyping errors. Again, more information on haplotypes would help to minimize this problem.

There are various analyses that could be performed differently or in addition to the ones presented. For example, AMOVA could be used to test hypotheses of alternative genetic structure, and data could be examined for an isolation-by-distance effect. Most importantly, a Nested Clade Analysis would enable separation of historic and demographic events (as suggested by Crandall et al. 2000). This is particularly important to verify the hypothesis of “founder effects and range expansion” (9/2/8-9). Such analyses helped to clarify genetic structure in fragmented populations of another small mammal with a controversial conservation history (Swei et al. 2003).

The assumption that microsatellite DNA loci will provide less resolution than mitochondrial DNA sequences (5/1/1-4) is completely erroneous. It is also unclear why *Z.h.preblei*, if shown to be distinct based on mtDNA sequence data, must then be tested for hybridization, and furthermore, why microsatellite DNA loci should be used for this task (nuclear markers yes, but preferably not microsatellite DNAs).

4. Additional or divergent taxonomic conclusions based on these data?

Alternative conclusions are most certainly possible. However, limitations of the genetic data hamper any conclusions, and render as speculative any taxonomic interpretations.

Shared haplotypes between *Z.h.preblei* and *Z.h.campestris* could simply be due to a lack of resolution for the limited data set. Further, incomplete lineage sorting is to be expected given the presumed time frame of divergence. This is particularly so if a population is large or expanding (*Z.h.campestris*) and if ancestral haplotypes are retained (Paetkau 1999). However, the presence of different haplotype frequencies suggests that *Z.h.preblei* is on its own trajectory and could warrant DPS (“distinct population status”) if indeed corroborated with a more comprehensive genetic evaluation.

It also seems as if the report confuses a “gene tree” with a “species tree.” As a means of explanation, a phylogeny of a species represents a multitude of nested component trees, each reflecting the history of populations as determined by single characters. When a component tree is derived from DNA information (e.g., haplotype sequences) it is referred to as a “gene tree” (Avice 2000). On the other hand, a “species tree” can be viewed (Avice 1994:126) as “...a single pedigree that extends [historically] as an unbroken chain of parent-offspring genetic transmission...” Hence, species trees are histories of organisms (i.e., pedigrees) whereas gene trees are histories of single traits (Avice 2000). The distinction is important.

In phylogenetic reconstruction, one must understand that a gene tree does not necessarily reflect a species tree. Not only can gene trees differ in their topology one from another, but also from species trees. Differences are due to a variety of biological factors (e.g., stochastic lineage sorting, introgressive hybridization, horizontal transfer, etc.; Avice 1994, 2000). Yet a common practice is to use gene trees as an estimate of the species tree (as herein). This is a tenuous association at best, because there are ample reasons for gene trees and species trees to be discordant (as above). This is particularly true in the present report because the gene tree is based on very few base pairs of data. Thus, the original hypothesis carried the caveat “distinguishable....using mitochondrial DNA sequence data” ((4/2/6), yet this caveat was somehow dropped in the conclusions and the gene tree then became instead the species tree for the taxon.

5. Interpretations about possible mechanisms of reduced gene flow between *Z.h.preblei* and other subspecies of *Z.hudsonius*?

I also cannot follow the logic of the argument that “founder effect and range expansion” contradict evidence of restricted genetic exchange. I would agree that low haplotype diversity in *Z.h.preblei* suggests a population bottleneck. But if it was due to a recent founder event, one would expect the haplotypes in *Z.h.preblei* to represent various *Z.h.campestris* haplotypes and not just those that form a distinct cluster.

The existence of four very similar haplotypes suggests a bottleneck within *Z.h.preblei* populations followed by subsequent expansion with low effective population size. It also indicates a lack of (or at least reduced levels of) gene flow between *Z.h.preblei* and other *Z.hudsonius* subspecies. Hey *et al.* (2003) argued that bottlenecks will obscure genetic divergence among populations, even within such ancient lineages as Tuataras. Based on the data in this study, it appears that the Front Range population is of “recent” (i.e., post-Pleistocene) origin. Again, the molecular data are limited in their number and capabilities, and thus do not provide sufficient resolution from which to draw conclusions. Additional genetic data are needed.

Thus, genetic divergence originating through range expansion and subsequent reduced gene flow is indeed a potential mechanism for speciation. For example, Abbott and Double (2003a,b) used mitochondrial and microsatellite DNA to discover that shy albatross arose from white albatross through range expansion.

6. Crandall et al. (2000) definition of ESU?

The definition for ESUs as provided in the report is rather vague. For example, Moritz (1994) included significant differences in nuclear alleles as an additional criterion for designation of an ESU.

Moritz (1994) also emphasized the distinction between the biological definition of ESU (and MU) and the genetic criteria. He argues that “the term ‘significant’ in ESU should be seen as recognition that the set of populations has been historically isolated and accordingly, is likely to have a distinct potential” (Moritz 1994:373). Further, his genetic criterion for recognizing an ESU includes mitochondrial and nuclear loci: “ESUs should be reciprocally monophyletic for mtDNA alleles and show significant divergence of allele frequencies at nuclear loci.” (Moritz 1994:373).

Ecological exchangeability is one criterion proposed by Crandall et al. (2000) and this suggestion has merit. However, its application is hampered by lack of ecological data for most rare species. Pacific salmon are probably a notable exception. However this report does not provide data (or even references) concerning the ecological characteristics of the subspecies under study. Did I miss something here?

Further, Crandall et al. stressed that “failure to reject the null hypothesis does not imply that the null hypothesis is true, but could simply be a result of the lack of relevant data” (Crandall et al. 2000:293). I would argue this is indeed the case here.

7. Clear ecological distinctions between *Z.h.preblei* and closely related taxa?

Authors state that they “examined literature for evidence of ecological differences between subspecies” (5/2/6-7), yet they do not provide a single reference. Also, additional morphological data might provide some insight?

References Cited:

- Abbott, C.L., and M.C. Double. 2003a. Phylogeography of shy and white-capped albatrosses inferred from mitochondrial DNA sequences: implications for population history and taxonomy. *Molecular Ecology* 12:2747—2758.
- Abbott, C.L., and M.C. Double. 2003b. Genetic structure, conservation genetics and evidence of speciation by range expansion in shy and white-capped albatrosses. *Molecular Ecology* 12:2953—2962.
- Avise, J. C. 1994. *Molecular Markers, Natural History and Evolution*. Chapman & Hall, New York.
- Avise, J. C. 2000. *Phylogeography: the History and Formation of Species*. Harvard University Press, Cambridge.
- Drew, R. E., J. G. Hallett, K. B. Aubrey, K. W. Cullings, S. M. Koepf, and W. J. Zielinski. 2003. Conservation genetics of the fisher (*Martes pennanti*) based on mitochondrial DNA sequencing. *Molecular Ecology* 12:51—62.
- Haig, S.M. 1998. Molecular contributions to conservation. *Ecology* 79(2):413—425.
- Hey, J.M., C.H. Daugherty, A. Cree, and L.R. Maxson. 2003. Low genetic divergence obscures phylogeny among populations of *Sphenodon*, remnant of an ancient reptile lineage. *Molecular Phylogenetics and Evolution* 29:1-19.
- Mortiz, C.1994. Defining 'Evolutionary Significant Units' for conservation. *Trends in Ecology and Evolution* 9(10):373—375.
- Pääbo, S. 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Science, USA* 86:1939—1943.
- Paetkau, D. 1999. Using genetics to identify intraspecific conservation units: a critiques of current methods. *Conservation Biology* 13(6):1507—1509.
- Swei, A. P.V. Brylski, W.D. Spencer, S.C. Dodd, and J.L. Patton. 2003. Hierarchical genetic structure in fragmented populations of the Little Pocket Mouse (*Perognathus longimembris*) in Southern California. *Conservation Genetics* 4(4):501—514.