EXHIBIT C

Is the Preble's meadow jumping mouse an evolutionarily distinct subspecies? Comments on the report by King et al. (2006)

Rob Roy Ramey II*, John D. Wehausen[§], Hsiu-Ping Liu[†], Clinton W. Epps[#], and Lance M. Carpenter**

*Department of Conservation Biology, Denver Zoological Foundation, Steele St., Denver, Colorado 80205
§ University of California, White Mountain Research Station, 3000 East Line Street, Bishop, CA 93514
†University of Denver, Department of Biological Sciences, 2190 E. Iliff, Denver, CO 80208
University of California, Department of ESPM, 137 Mulford Hall, Berkeley, CA 94720-3114
** Department of Zoology, Denver Museum of Nature & Science, 2001 Colorado Blvd., Denver, CO 80205

18 May 2006

Introduction

King et al. (2006, hereafter King et al.) produced a second independent set of genetic data on the prebles jumping mouse (*Z. h. preblei*) and their analyses thereof led to the opposite conclusions of Ramey et al. (2005) regarding the uniqueness of that subspecies. King et al. argue that their different conclusions result from sampling design, tissues used (museum specimens versus fresh ear punches), amount of molecular genetic data used (longer and more mtDNA sequences, more microsatellite loci), and analytical methods used, and criteria for uniqueness). We take issue with their interpretation of differences between the studies. Their portrayal of our work was inaccurate and the difference between conclusions is largely a function of basic conceptual differences. While the King et al. sampling represents a notable effort, it falls short in both sampling design and strength of inference that they attempt to present so forcefully.

Morphometric Considerations

King et al. ignored several key sources of information on morphological variation and adaptive differences that did not support their conclusions. These included a literature review conducted by Cryan, one of the coauthors of King et al. (Cryan 2004), who could not find evidence of adaptive differences among putative subspecies, or ecological differentiation that would lead to adaptive differences. Similarly, King et al. ignored the results of Jones (1981) who had found no support for any subspecies of *Z. hudsonius*, despite measuring over 9,000 specimens of *Zapus*.

We also retested the original morphological basis of Krutzsch's (1954) description of *Z. h. preblei* and found no support for his results. That alone would be sufficient basis to reject the taxonomic separation of *Z. h. preblei*. Rather than acknowledging that finding, King et al. dismissed the use of morphology in general because it might not reflect genetic differences, an argument they attempted to support by selective use of references. However, by their argument, these mice should never have been listed under the ESA, a listing that King et al. now strive to defend.

Conceptual Basis and Thresholds for Uniqueness

We based our analyses on a definition of subspecies provided by Ball and Avise (1992) to avoid the long history of taxonomic subspecies decisions having no definitional basis. Ball and Avise (1992) proposed that subspecies represent major subdivisions in the gene pool diversity. By that definition, subspecies are similar to evolutionary significant units (ESUs) as discussed by Moritz (1994a) in requiring deeper historic phylogenetic separation – an important criterion of Crandall et al. (2000). Our uniqueness criteria were set in advance in an attempt to focus on the need for subspecies to represent virtual discontinuities in the gene pool diversity.

King et al. cite Avise and Ball (1990) and Moritz (1994b) as their basis for subspecies and state that they tested the null hypothesis that purported subspecies of *Zapus husdonius* comprise a single homogeneous unit. They claim to accept as evidence of subspecific distinctiveness the conditions previously defined by Avise and Ball (1990) and Moritz (1994b) "as significant phylogeographic separation of mtDNA alleles between subspecies (or populations) combined with congruent phylogeograhic structure for nuclear loci". By their analyses, the definition of subspecies they use differs greatly from ours and appears to be equivalent to what Moritz (1994a) defined as management units. Their different conclusions are substantially a function of these fundamental conceptual differences

In their criticisms of Ramey et al. (2005), King et al. did not acknowledge the ways in which the molecular results of Ramey et al. (2005) were similar to theirs. These include: 1) shallow levels of evolutionary divergence found among putative subspecies for mtDNA and microsatellites; 2) support in mtDNA analyses for a *Z. h. pallidus/luteus* clade and a *Z. h. preblei/campestris/intermedius* clade; 3) putative subspecies were not reciprocally monophyletic or even close to being so; and 4) few unique alleles found in *Z. h. preblei* despite a larger sample size for this putative subspecies. These similarities are important because of their bearing on how different conceptual approaches to subspecies affected differences in conclusions.

Statistical Significance Versus Biological Significance.

In their null hypothesis testing of genetic homogeneity, King et al. equate statistical significance with biological significance – an analytical approach that deviates from ours. It is known that with a large enough sample size it is possible to find statistical significance in almost any comparison, especially when intervening variation is ignored. As pointed out by Hedrick (2001): *"the statistical power for determining differentiation between groups is closely related to the number of independent alleles, so, that for even for a few highly variable microsatellite loci, there can be high statistical power. When there is such high statistical power, extremely small molecular genetic differences between groups become statistically significant."*

King et al. screened a large numbers of individuals from few populations for a large number of microsatellites (27 total). Although King et al. found a high level of statistical significance in their comparisons (similar to our results), the degree of differentiation among *Z. h. preblei*, *Z. h. campestris* and *Z. h. intermedius* were the lowest of any of the pairwise comparisons for mtDNA (King et al. Tables 8 and 9) and microsatellites (King et al. Table 6). This low degree of differentiation is illustrated by the fact that only 4 unique alleles were reported in *Z. h. preblei* (out of 279 total), the lowest number of unique alleles for any subspecies sampled. It appears that King et al. are reporting statistical significance in tests that are of little biological relevance.

King et al. reported high levels of correct assignment to subspecies using the program STRUCTURE. These authors attribute this to *Z. h. preblei* having "considerable evolutionary differentiation" from other subspecies and to shortcomings of the Ramey et al. (2005) study. However, King et al. failed to acknowledge that this high level of correct assignment could be an artifact of sampling design (as discussed below). Moreover, the higher levels of correct assignment also are very likely to result from the larger number of microsatellite loci surveyed. King et al.'s findings raise a valid critique of both studies - use of assignment tests such as STRUCTURE may not be an appropriate tool for evaluating taxonomic separation, because of the sensitivity of these tests to the number of loci employed.

The probability of correct assignment of individuals to populations can increase substantially with the number of microsatellite loci used. For example, in a study that examined the distribution of variation within and among human populations, Rosenburg et al. (2005) found that increasing the number of microsatellite loci increased the statistical significance of assigning individual humans to geographic populations. These authors, who employed 993 microsatellite loci in their study, reported that: *"human genetic diversity consists not only of clines, but also of clusters, which STRUCTURE observes to be repeatable and robust."* In light of these observations it is worth asking: If a high level of correct assignment can be detected using microsatellite loci, should populations automatically be considered as subspecies or distinct population segments (DPSs)? If King et al.'s approach were applied to human populations, how many subspecies of *Homo sapiens* would King et al. recognize? Future efforts employing these types of analyses may need to establish threshold assignment probabilities for a set

number of loci with a given amount of variation per locus to allow comparability between studies.

Sampling Questions

Sampling Distribution

King et al. sampled many individuals from few local subpopulations (within a 6km radius) within each putative subspecies, whereas Ramey et al. (2005) sampled many populations, but few individuals per population, across the range of each putative subspecies. An ideal study design would incorporate both approaches, thereby sampling many individuals across the range of each putative subspecies. However, this is not often practical because of logistical and funding constraints. Given the choice, which strategy provides the most objective test of subspecies uniqueness?

King et al. claim that their sampling strategy allows more appropriate statistical testing, but they do not provide any supporting evidence. Instead, the sampling approach of King et al. has created artificial gaps in the distribution of genetic variation, making it less appropriate for subspecies comparisons. Most importantly, their sampling approach ignores the variation found across the range of the putative subspecies (or hypothetical DPSs), especially near zones of current or recent contact.

It is notable that King et al. did not make use of available specimens from southeastern Wyoming. These are the areas where shared variation among putative subspecies is most likely to be found (because of current or recent genetic exchange between *Z. h. preblei* and *Z. h. campestris*). Because King et al. sampled only one subpopulation of *Z. h. luteus* and two subpopulations each of *Z. h. campestris, Z. h. intermedius,* and *Z. h. pallidus,* artificial gaps have been inserted between all of the subspecies. Therefore, the sampling strategy of King et al. predisposed the results to exaggeration of genetic distances among putative subspecies. This same effect was recently reported by Rosenburg et al. (2005) in comparisons of microsatellite data from human populations.

The sampling scheme of King et al. also created an artificial gap within the range of *Z. h. preblei.* This resulted in the recognition of two potentially listable DPSs under the ESA. As noted in Ramey et al. (2005) the Denver metropolitan area creates a manmade gap in the range of *Z. h. preblei.* King et al. artificially increased the size of that gap (by ~70km) by excluding a large number of samples from Boulder County, Colorado. Furthermore, by not taking into account that growth of the Denver Metropolitan could affect the distribution of genetic variation of mice in this area (e.g. local extirpation and bottlenecks), it appears that the results of King et al. could lead to the identification of DPSs solely based on manmade separation of ranges and recent anthropogenetic changes to the environment.

Despite its limitations in inferring degree of within-population variation, the sampling scheme used by Ramey et al. is (2005) is supported both theoretically and empirically by

4

other authors for the broader scale comparisons across species (Lynch and Crease 1986; Presa et al. 2002; Rosenburg et al. 2005).

Sources of Material

King criticized Ramey et al. (2005) for using museum specimens, claiming a wide variety of problems associated with such tissue, based mostly on literature for ancient DNA samples, not museum specimens collected within the past 45 years. In truth, Ramey et al. (2005) used a mixture of museum specimens and frozen tissue. However, the specimens used by King et al. are subject to the same issues they raise regarding museum specimens.

Ear punch samples were the primary material used by King et al. and were collected without measures to prevent cross-contamination of specimens. Ear punch samples are small (~1mm in diameter) and collected with a hand-held ear punch tool. Cross-contamination occurs when the DNA of one or more samples is mixed. When the polymerase chain reaction (PCR) is used to amplify copies of DNA from the sample, contaminated DNA is amplified as well, potentially leading to genotyping errors and erroneous results. Because mouse ear punches are minute in size and contain small amounts of DNA these small samples are susceptible to the effects of cross-contamination. As a result, stringent contamination control procedures are needed for sample collection and handling (see Ramey et al. 2000 for procedures used for extracting DNA from individual scabies mites). In the case of King et al., flaming or bleaching of the ear punch tool and forceps would have prevented the carry over of blood or tissue remnants to other ear punch samples. Also, a new pair of laboratory gloves should have been worn when obtaining each specimen to prevent cross-contamination of ear punches as well as the collection tubes, forceps, and sources of ethanol used to fill collection tubes.

Cross contamination in all of the ear punch samples in King et al. is a strong possibility because *the same ear punch tool was used repeatedly to sample multiple mice in all of the ear punch collections.* The samples collected by Cryan and Ellison (2005) were not collected with any of the cross-contamination control procedures discussed above. The ear punch samples collected in Colorado by Schoor and Shenk were only wiped with alcohol (R. Taylor, pers. comm.; Riggs et al. 1997), which cannot be expected to remove contaminating DNA from the inside of the punch tube or reliably remove all contaminating DNA from the outside of the tube. In all cases, there is no documentation that laboratory gloves were ever worn.

Cross-contamination or carryover of genetic material between ear punches is analogous to the spread of infectious diseases through the reuse of contaminated needles. A small amount of contaminating material is all that is needed to produce problems, which is why experienced geneticists go to great lengths to avoid cross-contamination in the collection and handling of specimens.

A difference between King et al. and Ramey et al. (2005) is that the latter relied on vouchered museum specimens. These are accessioned into the collection of an accredited museum with the skin, skull, and detailed collections information to accompany each sample. This means that the specimens used are publicly available and additional follow-up work can be conducted on any sample (e.g. additional genetic analyses, morphometric, and pelage comparisons) whereas the same cannot be said of the ear punch specimens that were used in King et al.

Historic samples, such as vouchered museum specimens, can help the understanding of patterns of genetic variation that occurred in natural populations prior to human induced bottlenecks and local extirpations. If only recently collected samples are used, as advocated by King et al., we limit our ability to investigate the extent to which current patterns of variation are due to natural or recent anthropogenic processes. Ramey et al. (2005) we suggested that *"criteria for genetic uniqueness need to adequately identify natural discontinuities in gene pool variation and distinguish these from recent (e.g. last 100 years) differences that may be due to genetic drift or human-induced bottlenecks or isolation (Hedrick et al. 2001; Brown et al. 2004)."* The use of museum specimens is a part of such an investigative strategy. In contrast, King et al. propose: *"any new study of this type should focus on newly trapped individuals...rather than museum skins."*

King et al. assert that the shared mtDNA haplotypes found by Ramey et al. (2005) between *Z. h. preblei* and *Z. h. campestris* must have been the result of contamination of museum specimens. Their support for this assertion comes from the claim that they collected a large number of samples from the same locations as the specimens used by Ramey et al. (2005) but did not find any shared haplotypes between *Z. h. campestris* and *Z. h. preblei*. While the potential contamination of museum specimens is a point well taken, there are several reasons to take issue with King et al.'s assertion. They did not consider alternative explanations, did not accurately report the basis of their assertion, and they ignored the fact that Ramey et al. (2005) employed strict contamination control procedures to minimize the chances of cross-contamination or PCR carryover occurring (see Ramey et al. 2000). Additionally, at least two negative controls were used per nested PCR amplification and these did not contain detectable PCR-amplified DNA.

King et al. did not consider the alternative explanations. The first is that the shared mtDNA haplotypes that we observed were actual shared variation rather than contamination. Second, King et al. might not have found these same haplotypes because they have been lost from *Z. h. campestris* due to genetic drift, extirpation and recolonization, or a selective sweep. A similar shift of mtDNA haplotypes, from common to rare, has been previously reported in a study comparing contemporary specimens with museum specimens of *Peromyscus* that were collected one hundred years apart (Pergams et al. 2003). Third, King et al. incorrectly reported that they sampled individuals from the same locations as our museum specimens, but they relied on samples collected over very short distances, typically less than 5km (Cryan and Ellison 2005). At this scale, mice can be expected to be very closely related to each other and not representative of the variation in the population as a whole. Thus they may have sampled different populations of *Z. h. campestris* than Ramey et al. (2005). Fourth, this could be

the result of previously undetected mtDNA heteroplasmy (although we did not see other evidence of this in our study).

Standardization Between Studies

King et al. misrepresented that they had not obtained necessary samples from the Denver Museum of Nature and Science that could have been used to standardize microsatellite alleles between the two studies. It appears that King had obtained *all* of the samples used in Ramey et al. (2005) but either did not run these specimens or did not release the results.

Analysis of Molecular Variance

We concur with King et al. that the AMOVA criterion that we proposed for mtDNA may not be an ideal measure with which to test the uniqueness of subspecies or distinct populations. As found by King et al., if there are slight differences among mtDNA haplotypes, but those haplotypes are fixed or nearly fixed in populations, it will have a substantial effect on the value of Φ_{ST}. That could lead to the erroneous designation of weakly differentiated populations as subspecies or DPS.

We take issue with the frequent use by King et al. of qualitative assessments for F_{ST} and other divergence parameters as being high, moderate, or low. These are subjective assessments that detract from the goal of providing objective tests of genetic uniqueness.

Conclusions

As noted in Ramey at al. (2005) and by at least one peer reviewer of our earlier work (Sites), consistent thresholds for defining conservation units below the level of subspecies have been lacking. However, in proposing thresholds, we also acknowledged that: *"Any such threshold can be seen as arbitrary; however, we hope to establish reasonable threshold levels for these sorts of tests where they have often been absent. Appropriate thresholds can be debated and revised, but we feel that the first step in establishing standards and objectively applying them is to state them explicitly."* It is legitimate to debate thresholds, but the need for them is obvious -- there are many endangered taxa and not enough resources to conserve them. If conservation effort is allocated to non-distinct or weakly differentiated populations, other more unique taxa lose out. Hypothesis testing relative to these thresholds can provide objective assessments of degree of uniqueness. As legislated in the 1982 amendments to the US-ESA and repeated by numerous authors (cited in Ramey et al. 2005) conservation of endangered taxa would be best served if the allocation of conservation effort were prioritized based on degree of genetic uniqueness.

Literature Cited:

Avise, J. C., and R. M. Ball Jr. (1990) Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford Survey of Evolutionary Biology* 7: 45–67.

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

Crandall.K.A., Bininda-Emonds O.R.P, Mace, G.M., & Wayne R.K. (2000) Considering evolutionary processes in conservation biology. *Trends Ecol. Evol.* **15**: 290-295.

Cryan, P. M. (2004) Synthesis of existing information on meadow jumping mice (*Zapus hudsonius*) in the northern Great Plains. An information review compiled for the U.S. Fish and Wildlife Service, Region 6, Denver, Colorado. USGS Fort Collins Science Center, 2150 Centre Ave., Building C, Fort Collins, CO 80526. 27pp.

Cryan, P. M. & Ellison, L. E. (2005) Distributional survey of the meadow jumping mouse (*Zapus hudsonius*) in the northern Great Plains: Trapping Report, Summer 2005. Progress report compiled for U.S. Fish and Wildlife Service, Region 6, Denver, Colorado. United States Geological Survey, Fort Collins Science Center, Fort Collins, CO.

King, T. L., Switzer, J. F., Morrison, C. L., Eackles, M. S., Young, C. C., Lubinski, B., & Cryan, P. (2006) Comprehensive analysis of molecular phylogeographic structure among meadow jumping mice (*Zapus hudsonius*) reveals evolutionarily distinct subspecies. A report submitted to U.S. Fish and Wildlife Service. January 27, 2006.

Jones, G.S. (1981) The systematics and biology of the genus *Zapus* (Mammalia, Rodentia, Zapodidae) unpublished Ph.D. dissertation. Indiana State University, Terre Haute.

Hedrick, P. W. (2001) Conservation genetics: where are we now? *Trends Ecol. Evol.* 16:629-636.

Krutzsch P. H. (1954) North American jumping mice (genus *Zapus*) *Univ. Kansas Publ. Mus. Nat. Hist.* 4: 349-472.

Lynch, M. & Crease, T.J. (1986) The analysis of population survey data on human variation. *Mol. Bio. Evol.* 7: 377-394.

Moritz C (1994a) Defining 'evolutionarily significant units' for conservation. *Trends Ecol. Evol.* **9**: 373-375.

Moritz C (1994b) Applications of mitochondrial DNA analysis in conservation: a critical review. *Molec. Ecol.* **3**: 401-411.

Oyler-McCance, S. J., Kahn, N. W., Burnham, K. P., Braun, C. E., and Quinn, T. W. (1999) A population genetic comparison of large and small-bodied sage grouse in Colorado using microsatellite and mitochondrial DNA markers. *Molecular Ecology* **8**: 1457-1465.

Pergams, O. R. W., Barnes, W. M. & Nyberg, D. (2003) Mammalian microevolution: Rapid change in mouse mitochondrial DNA. *Nature* **423**: 397.

Presa, P., Pardo, B. G., Martínez, P., & Bernatchez, L. (2002) Phylogeographic Congruence Between mtDNA and rDNA ITS Markers in Brown Trout. *Mol. Bio. Evol.* **19**: 2161-2175.

Ramey, R.R., W.M. Boyce, B. Farrell, and S. Kelley. (2000). Phylogeny and host specificity of psoroptic mange mites as indicated by ITS sequence data. *Journal of Medical Entomology* **37**: 791-796.

Ramey R. R. II, Liu, H-P, Epps C. W., Carpenter L M, & Wehausen J. D. 2005. Genetic relatedness of the Preble's meadow jumping (*Zapus hudsonius preblei*) to nearby subspecies of *Z. hudsonius* as inferred from variation in cranial morphology, mitochondrial DNA and microsatellite DNA: implications for taxonomy and conservation. *Animal Cons.* **8**: 329-346.

Riggs, L. A., Dempcy, J. M., & Orrego, C. (1997) Evaluating distinctiveness and evolutionary significance of Preble's meadow jumping mouse: phylogeography of mitochondrial DNA non-coding region variation. Report to the Colorado Division of Wildlife, Denver, Colorado.

Rosenberg, N. A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J. K., & Feldman, M. W. (2005) Clines, clusters, and the effect of study design on the inference of human population structure. *PLOS Genetics* **1**: 660-671.