

## PROVIDING RELIABLE AND ACCURATE GENETIC CAPTURE–MARK–RECAPTURE ESTIMATES IN A COST-EFFECTIVE WAY

KEVIN S. MCKELVEY,<sup>1</sup> U.S. Forest Service, Rocky Mountain Research Station, 800 East Beckwith, Missoula, MT 59801, USA  
MICHAEL K. SCHWARTZ, U.S. Forest Service, Rocky Mountain Research Station, 800 East Beckwith, Missoula, MT 59801, USA

**Abstract:** Capture–mark–recapture (CMR) estimates assume no misidentification of individuals captured and are extremely sensitive to identification errors. A large body of published literature has demonstrated that non-invasively derived genetic tags are error-prone, and the potential biases associated with these errors are large. We provided methods to reduce and evaluate these errors. Paetkau (2004, this issue), in his comments concerning our paper, argues that no formal, statistical error testing is necessary and that good laboratory practices are sufficient to remove all error. However, he provides only anecdotal evidence that this is the case. Given the presence of a variety of errors in genetic tags and the potential for large biases associated with these errors, we argue that scientific norms require formal tests to demonstrate the absence of errors. The primary purpose of our study was to provide such tests in a manner that is not cost-prohibitive.

*JOURNAL OF WILDLIFE MANAGEMENT* 68(3):453–456

**Key words:** allele, binomial, capture, DNA, dropout, estimation, genetic, genotyping, mark, microsatellite, molecular tagging, non-invasive, population.

DNA-based estimates of animal distribution and abundance are excellent, cost-effective techniques. Here, the debate is how to ensure that samples are error-free; that is, a perfect correspondence exists between genetic tags and unique individuals. Two general categories of errors occur: insufficient numbers of loci amplified leading to common tags among multiple individuals (“Shadow effect;” Mills et al. 2000), and genotyping/recording errors leading to multiple tags associated with the same individual. In McKelvey and Schwartz (2004, this issue) we concentrated on the latter, but our “Difference in Capture History” (DCH) test works for detecting both sources of error.

Based on simulations, we know that small amounts of genotyping error in the dataset can have large, unacceptable consequences for CMR estimates (Waits and Leberg 2000, McKelvey and Schwartz 2004). Further, a large body of literature indicates that genotyping errors are pervasive in non-invasively derived DNA samples (Taberlet et al. 1996, Gagneux et al. 1997, Morin et al. 2001, Eggert et al. 2003, Fernando et al. 2003). Because the effects of genotyping errors are potentially large, scientific norms require the researcher to both control error and demonstrate this control. The only real question is: how? Several approaches to remove these errors have been suggested (Taberlet et al. 1996, Morin et al. 2001, Miller et al. 2002, Creel et al. 2003); the latest (McKelvey and Schwartz 2004) is criticized by Paetkau (2004).

First, we note that Paetkau’s (2004) criticisms largely deal with his views that the prevailing non-invasive sampling literature deals unnecessarily with genotyping and other errors. He doesn’t state that our tests are bad, but in his opinion these measures are unnecessary and costly. He argues that having diligent, experienced workers and a strict protocol is sufficient to remove error. Similarly, he views repeatedly running all samples (multi-tubing) to expose and eliminate errors as unnecessary, even though this is the accepted methodology in the published literature (Taberlet et al. 1996, Goossens et al. 1998, Constable et al. 2001, Morin et al. 2001, Creel et al. 2003). We find his antagonism to our approach particularly mystifying because our primary goals were to formalize the good—but ad hoc—methodologies that he has published (Paetkau 2003) and to provide a formal test for error removal that would require a minimum number of additional assays.

Second, we disagree with Paetkau’s (2004) assertion that “the suggestions of McKelvey and Schwartz (2004) only serve to widen [the] gap, offering no demonstrable benefit over established methods at a cost that would threaten the practical utility of the technique.” The “established method” in the field of non-invasive sampling is to run each homozygote sample at least 7 times (Taberlet et al. 1996). In fact, Paetkau (2003) even subscribes to this approach when he has samples in which allelic dropout is suspected: “In situations where allelic dropout is suspected, but where the reanalysis confirms the original data, I now *repeat the reanalysis up to 7 times...*”

<sup>1</sup> E-mail: kmckelvey@fs.fed.us

(emphasis ours). Thus, conventionally accepted approaches would require running 7 markers 7 times each, or 49 runs per individual. Both Paetkau's (2003) and our methods attempt to minimize the overall number of times each sample is run to preserve DNA. The difference is that our approach statistically validates the claim of no error, whereas his does not.

Third, our method is not as onerous as Paetkau (2004) portrays. While we ran our simulations with 15 loci, the necessary number of loci is largely based on the tag size necessary to reduce shadow effects to insignificant levels ( $L_{base}$ , in McKelvey and Schwartz 2004). Amplifying enough loci to largely eliminate shadow effects is a necessary prerequisite prior to using CMR tests in any case, and a formal error test therefore requires that at least 1 additional locus be amplified—a very minimal additional cost considering the gain of assuring that the sample has low error rates. Of course, a variety of benefits are associated with amplifying more loci; specifically, the resulting data can be used for more precise gene flow, relatedness, or paternity questions.

Fourth, Paetkau (2004) questions whether our "Examining Bimodality" (EB) test is useful, stating that the issue is one of efficiency rather than efficacy, and he argues that a tag size-independent approach in which all organisms differing at  $\leq 3$  loci are multi-tubed is more efficient. The EB test is indeed about efficiency and is among the many things that one might do in the lab to try to control and identify errors; yet, this test alone does not guarantee that errors are reduced to insignificant levels. Our second test, the DCH test, is the formal test for error removal. It is a very generic test, testing the hypothesis that capture history is not a function of tag size and composition. As such, the DCH test checks for both genotyping and shadow errors but does not differentiate between the 2. Our test therefore has 2 parts: first, isolate those samples that are likely to contain errors; and second, enlarge and change tag composition to test whether the capture history, and therefore the resulting CMR estimate, is sensitive to these changes.

Paetkau (2004) primarily discusses the EB test, especially as it relates to tag length. Proper tag length will vary with population, based on heterozygosity, allelic diversity, and a variety of subtle issues such as the degree to which the captures of family members are correlated. As an example, the population presented in Paetkau (2004: Fig. 1) is extremely heterozygous at all amplified loci

(many individuals differ at all loci), and error appears to be largely limited to 1 error per sample. This combination allows Paetkau (2004) to apply a bimodality test while using a short tag. However, we argue that in general, choosing a longer tag will provide additional information and lead to both more efficient error identification and more robust error-checking. If, for instance, in our simulated population (McKelvey and Schwartz 2004) we chose a tag size of 6 loci, most individuals in the population would be separated by  $\leq 3$  loci (Fig. 1A; data free from genotyping errors). While nothing is wrong with multi-tubing most of the samples following Paetkau's rule, no evidence suggests that this process will target the errors or that the process is necessary. Figure 1B shows a sample from the same population and using the same markers, but with a 5% per-locus allelic dropout rate. Compared to the distribution in Fig. 1A, the number of individuals differing at only 1 locus is elevated, yet we cannot use this distributional information to separate the samples containing errors. Increasing the tag size to 8 loci, however, allows much better error identification (Fig. 1C). If Fig. 1C failed to show any trace of bimodality, we would have no reason to reanalyze half the samples prior to running the DCH test.

Interestingly, the DCH test applied to the sample in Fig. 1A indicates that, with a 6-locus tag, capture history is related to tag size and composition; a 6-locus tag is not sufficient to entirely remove shadow errors, despite the probability of identity (PI) =  $5 \times 10^{-6}$ . Our simulated population has high heterozygosity compared with many wild populations and does not contain the large numbers of highly related individuals that characterize some non-invasive samples. Thus, while populations may exist where a 6-locus tag is sufficient both to separate error and to eliminate shadow, we feel that this situation is uncommon.

Fifth, we question Paetkau's (2004) calculation of error rates. In truth, he did "identify 210 samples that had 1 error in their initial genotype, and 16 samples with 2-locus errors" (Paetkau 2004), but errors, both genotyping and shadow, were likely missed. Lacking a formal test, he has no way of knowing. Therefore, he is forced to make equivocal statements concerning the number of individuals in the Yukon sample. He cannot make a precise, scientific statement that the 335 samples equal 58 individuals. As we mention in McKelvey and Schwartz (2004), our EB test is simply a formalization of the lab methods that Paetkau

(2003) applies. However, our DCH test, which he has not run, potentially provides strong evidence that error—in all of its manifestations—was reduced to insignificant levels. As Paetkau (2004) strongly believes that he has removed all error, we are baffled by his hostility toward a test that, with minimal extra effort, would provide validation.

Sixth, Paetkau (2004) argues that concentrating DNA “allows more DNA to be used in each individual reaction, ... lowering the probability that this result will be inaccurate.” While concentrating the sample to increase DNA is obviously desirable, real limits exist to the degree to which one can get away with concentrating all the DNA into a single assay. For instance, Paetkau (2004) does not know a priori which samples will exhibit dropout errors. He therefore needs to hold back sufficient DNA to multi-tube any given sample if problems arise. Further, the efficacy of concentrating DNA, given Paetkau’s (2004) lack of a robust error check, is unknown. Morin et al. (2001) suggest that with <25 pg of DNA per reaction, allelic dropout is nearly certain (Morin et al. 2001:Table 3). After DNA concentration, Paetkau may have more or less than 25 pg per reaction; concentrating DNA does not guarantee samples containing >25 pg. Without quantitative methods (Morin et al. 2001), we have no way to evaluate the efficacy of concentrating DNA. As with reanalyzing samples that only differ at 1 or 2 loci, concentration is a good idea, but a direct test to its efficacy is required.

**DISCUSSION**

In general, scientific data consist of numerical estimates with associated error rates. Only when numbers are produced using methods that the scientific community agrees decrease error to trivial levels do we accept numbers without measures of quality. In DNA tagging, errors can be exposed by changing the tag size or using different genetic markers to build the tag. If the tag is large enough to eliminate shadow effects and the amplified loci are free from genotyping errors, then changes in either tag size or composition will have no effect on the derived capture history. Any changes in capture history associated with changing the tag indicate errors, and the degree to which the capture history changes indicates the magnitude of error. Selective multi-tubing of those samples and loci responsible for capture history changes will expose the nature of the errors. These ideas are the basis of our tests. While we believe that the methods we propose will prove effective and efficient, we know that

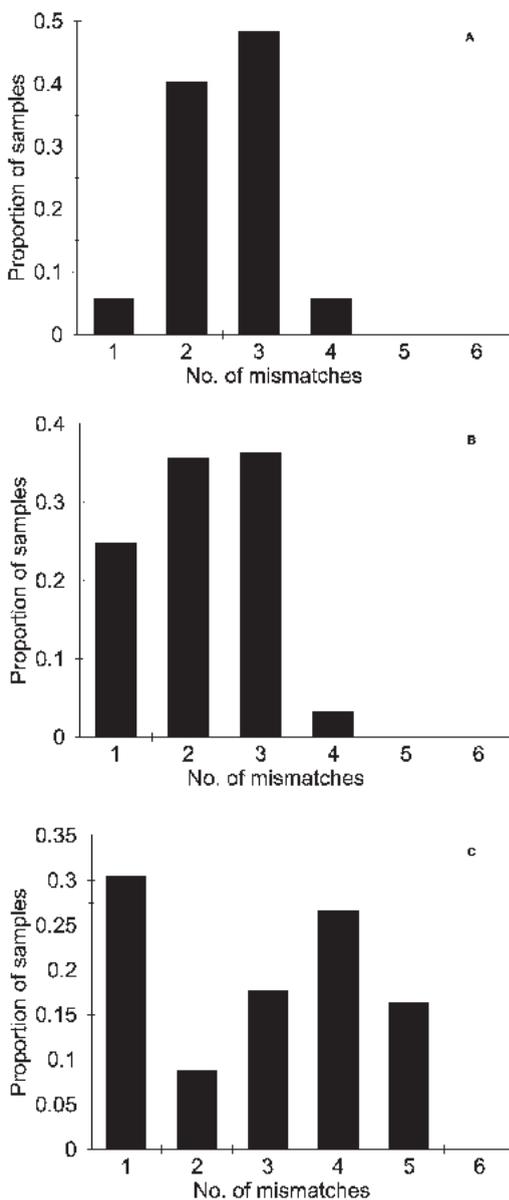


Fig. 1. Distributions of number of mismatches between samples based on tags of 6 (A, B) and 8 (C) loci. Data are from a simulated population (heterozygosity = 0.78) described in McKelvey and Schwartz (2004). The simulated sample analyzed in Fig. 1A is error free; samples analyzed in Figs. 1B and 1C contain 5% allelic dropout error per locus. Using a larger tag (1C) allows error (indicated by a bimodal distribution) to be identified and isolated.

other approaches and other variants of the approaches we outline are available. We cannot, however, accept Paetkau’s (2004) suggestion that the judgment of even the most diligent, experi-

enced technicians be substituted for formal quality checks. Capture–mark–recapture estimators produce population estimates that are assumed to be unbiased with confidence intervals that are assumed to be valid (Otis et al. 1978). The data we feed into CMR estimators therefore need to be of known quality. While Paetkau (2004) is concerned that managers will reject non-invasive surveys because of an emphasis on their potential pitfalls, we believe that the real risks to the utilization of these approaches lie in mistakes due to inadequate error removal or the failure to withstand scrutiny because error rates have not been quantified. Overall, non-invasive surveys likely will not be accepted by managers in the near future unless we can quantify the error rate and describe how these errors may affect population projections. Our tests provide a cost-effective way to efficiently use non-invasive DNA samples to test whether errors are present, and ultimately describe the errors.

#### LITERATURE CITED

- CONSTABLE, J. L., M. V. ASHLEY, J. GOODALL, AND A. E. PUSEY. 2001. Noninvasive paternity assignment in Gombe chimpanzees. *Molecular Ecology* 10:1279–1300.
- CREEL, S., G. SPONG, J. L. SANDS, J. ROTELLA, J. ZEIGLE, L. JOE, K. M. MURPHY, AND D. SMITH. 2003. Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. *Molecular Ecology* 12:2003–2009.
- EGGERT, L. S., J. A. EGGERT, AND D. S. WOODRUFF. 2003. Estimating population sizes for elusive animals: the forest elephants of Kakum National Park, Ghana. *Molecular Ecology* 12:1389–1402.
- FERNANDO, P., T. N. C. VIDYA, C. RAJAPAKSE, A. DANGOLLA, AND D. J. MELNICK. 2003. Reliable noninvasive genotyping: fantasy or reality? *Journal of Heredity* 94:115–123.
- GAGNEUX, P., C. BOESCH, AND D. S. WOODRUFF. 1997. Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology* 6:861–868.
- GOOSSENS B., L. P. WAITS, AND P. TABERLET. 1998. Plucked hair samples as a source of DNA: reliability of dinucleotide microsatellite genotyping. *Molecular Ecology* 7:1237–1241.
- MCKELVEY, K. S., AND M. K. SCHWARTZ. 2004. Genetic errors associated with population estimation using non-invasive molecular tagging: problems and new solutions. *Journal of Wildlife Management* 68:439–448.
- MILLER, C. R., P. JOYCE, AND L. P. WAITS. 2002. Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics* 160:357–366.
- MILLS, L. S., J. J. CITTA, K. P. LAIR, M. K. SCHWARTZ, AND D. A. TALMON. 2000. Estimating animal abundance using noninvasive DNA sampling: promise and pitfalls. *Ecological Applications* 10:283–294.
- MORIN, P. A., K. E. CHAMBERS, C. BOESCH, AND L. VIGILANT. 2001. Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology* 10:1835–1844.
- OTIS, D. L., K. P. BURNHAM, G. C. WHITE, AND D. R. ANDERSON. 1978. Statistical inference from capture data on closed animal populations. *Wildlife Monographs* 62.
- PAETKAU, D. 2003. An empirical exploration of data quality in DNA-based population inventories. *Molecular Ecology* 12:1375–1387.
- . 2004. The optimal number of markers in genetic capture–mark–recapture studies. *Journal of Wildlife Management* 68:449–452.
- TABERLET, P., S. GRIFFIN, B. GOOSSENS, S. QUESTIAU, V. MANCEAU, N. ESCARAVAGE, L. P. WAITS, AND J. BOUVET. 1996. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* 26:3189–3194.
- WAITS, J. L., AND P. L. LEBERG. 2000. Biases associated with population estimation using molecular tagging. *Animal Conservation* 3:191–199.