Molecular Ecology (2003) **12**, 2011–2017 doi: 10.1046/j.1365-294X.2003.01874.x

SHORT COMMUNICATION

The use of nuclear and mitochondrial single nucleotide polymorphisms to identify cryptic species

N. M. BELFIORE,*‡ F. G. HOFFMAN,† R. J. BAKER† and J. A. DEWOODY*
*Department of Forestry & Natural Resources, Purdue University, 195 Marstellar St, West Lafayette, IN 47907-2033, USA,
†Department of Biological Sciences, Texas Tech University, Lubbock, USA

Abstract

There is growing interest in the use of single nucleotide polymorphisms for evolutionary and population genetics. We tested the efficacy of one of the available single nucleotide polymorphism techniques, single-base extension, in distinguishing four cryptic species of *Microtus*. Sequence data were available for these species at nuclear and mitochondrial loci and their identity could be independently confirmed using karyotypes. We found that the development and optimization of single nucleotide polymorphisms required extensive effort, and that the method accurately identified the correct nucleotide at single nucleotide polymorphism sites ≈ 90% of the time at the conserved nuclear locus. Correct identification rates were much lower at the highly variable mitochondrial locus.

Keywords: cryptic species, diagnostic DNA methods, *Microtus*, single base extension, single nucleotide polymorphisms (SNP)

Received 1 December 2002; revision received 27 March 2003; accepted 27 March 2003

Introduction

Single-nucleotide polymorphisms (SNPs) have enormous potential for studies of evolutionary and population genetics, including, for example, the characterization of recombination rates, population origins, and gene flow. Recent analyses of the human genome estimate that it contains between 1.42 and 2.3 million SNPs scattered across 3 billion base pairs suggesting that the overall frequency of SNPs is 0.0005–0.0008 (The International SNP, Working Group 2001; Lee 2002). This suggests that large numbers of SNPs are available for use as neutral or nearly neutral genotyping markers. Nonhuman genomes are poorly characterized, but preliminary surveys suggest that SNP frequencies may be an order of magnitude higher in mice and passerine birds (Lindblad-Toh *et al*. 2000; Primmer *et al*. 2002). This observed 10-fold difference between humans and nonhumans most likely reflects the genomic regions that were surveyed in mice and birds, that is, regions known to show genetic variation by other means. Whatever their absolute numbers, it is apparent that more than a million SNPs occur in vertebrate genomes and that they are potentially powerful markers for evolutionary and population geneticists.

Spurred by the increase in high-throughput genotyping techniques and the drive to map complex diseases, many laboratories are in pursuit of reliable, relatively simple approaches to analysing SNPs. However, only a few investigators have exploited SNPs in studies of nonmodel organisms (e.g. Bensch *et al*. 2002; Primmer *et al*. 2002). One major reason for the paucity of SNP studies is the lack of candidate sequence data available for most species. In addition, the accuracy of novel SNP assay techniques must be confirmed by independent methods, and this is often difficult to achieve in studies of wild populations.

There are four primary approaches to SNP genotyping: allele-specific hybridization, allele-specific nucleotide incorporation, allele-specific oligonucleotide ligation, and allele-specific invasive cleavage (Kwok 2001). Of these four, the allele-specific nucleotide incorporation approach is: (i) conceptually simple and uses equipment common to many genetics laboratories, (ii) relatively economical, and (iii) potentially robust in the laboratory. Thus, it is an approach likely to bring SNP genotyping to evolutionary and population genetics.

Correspondence: N. M. Belfiore. ‡Present address: Museum of Vertebrate Zoology, 3101 VLSB # 3160, University of California, Berkeley, CA 94720, USA; Fax: + 1 510 643 8238; E-mail: nmb@uclink.berkeley.edu

Here we present one of the first studies to use SNP diagnostic methods in an ecological context, the identification of morphologically cryptic species. Cryptic species are usually defined as morphologically indistinct lineages separated by species-level genetic differences. Their evolution is poorly understood (Schneider & Moritz 1999; Martin & Bermingham 2000), but they are occasionally identified during broad geographical surveys of genetic variation within a purported species (e.g. Ryan & Bloomer 1999; Martin & Bermingham 2000; Witt & Hebert 2000; Kreiser 2001). Our SNP assays were developed and tested in four Eurasian vole species, *Microtus agrestis*, *M. arvalis*, *M. oeconomus* and *M. rossiaemeridionalis*, whose ranges overlap and which exhibit morphological crypsis to varying degrees (Wilson & Reeder 1993; Baker *et al*. 1996a). Of these four species, *M. arvalis* and *M. rossiaemeridionalis* are morphologically indistinguishable. Adult *M. agrestis* and *M. oeconomus* can be distinguished from each of the other species, but juveniles are easily confused with one another or with *M. arvalis*/*M. rossiaemeridionalis*. To date, there is no evidence of hybridization among these four *Microtus* species. Despite their similar morphologies, each species has a distinct karyotype (Baker *et al*. 1996a; Baker unpublished data) and is readily identified by nuclear (DeWoody 1999) or mitochondrial DNA (mtDNA) sequences (DeWoody *et al*. 1999). Our confidence that these species are genetically distinct, combined with the available diagnostic sequence data, made this group an ideal test of SNP assays. Potentially, SNPs could provide a viable alternative to existing molecular methods of species diagnosis (e.g. Nekrutenko *et al*. 2000; Meksem *et al*. 2001; Greig *et al*. 2002).

We sought to develop both nuclear and cytoplasmic SNP markers that could be used to differentiate among these morphologically similar species. We used previously published sequence data from each species to develop diagnostic SNP tests, and we confirmed our assays by independent means (i.e. DNA sequencing and karyotyping).

Materials and methods

Sampling

Thirty-eight voles were sampled near Chernobyl, Ukraine (see Baker *et al*. 1996a; RJ Baker unpublished data). Each individual vole in our study was karyotyped using established methods (Baker *et al*. 1996a; RJ Baker unpublished data). Tissue or DNA from these karyotyped individuals was sent to Purdue University for the SNP assays. Twenty-seven of the individuals were blind-coded, and 11 samples (controls) were labelled to species according to the karyotype data. Genomic DNA was extracted from tissue samples using standard techniques (Sambrook & Russell 2001).

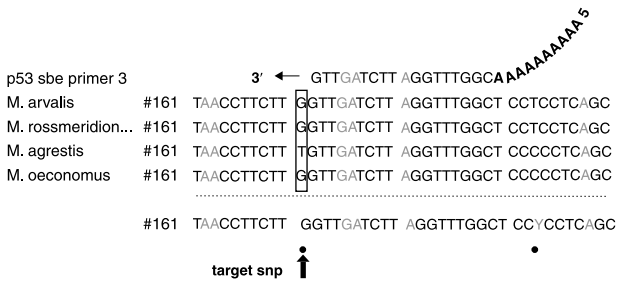


Fig. 1 The single base extension approach. The SBE primer anneals adjacent to the target SNP. Labelled ddNTPs in the reaction mix extend a single base, thus terminating the reaction and labelling the product, which is the size of the primer (in this example, 27 bp) plus one nucleotide.

Locus and SNP selection

DNA sequence variation at mitochondrial and nuclear loci has been studied in these species as part of a larger project to understand and characterize the effects of ionizing radiation exposure on wildlife (Baker *et al*. 1996b; DeWoody 1999) and studies investigating the systematics and molecular evolution of these species (e.g. DeWoody *et al*. 1999; Conroy & Cook 2000). We used the mitochondrial cytochrome *b* and nuclear *p53* sequences from those studies, and from others obtained via GenBank (Accession nos: AF01418–47, AF119271, AF159402–3, AF163902, U54472–95; Conroy & Cook 1999; Martin *et al*. 2000) to develop SNP assays.

A schematic diagram of the single-base extension (SBE) method for SNP detection is shown in Fig. 1 (Syvanen 1999). Briefly, a primer is designed such that its 3′ nucleotide is one nucleotide 5′ of a target SNP. During the SBE reaction, only fluorescent-labelled dideoxynucleotides are available to extend the 3′-end of a primer in a primer–template molecule. Thus, primer extension proceeds for only a single nucleotide before polymerization is halted, and the resulting fragment is fluorescent-labelled for detection on an automated DNA sequencer. In order to multiplex the SBE reaction, primers must have common annealing requirements and differ in length by at least 3–4 bp to avoid overlapping peaks on the chromatogram. To achieve this size disparity, different-length strings of nonannealing bases are added to the 5′-end of the primers; we found the simplest approach of adding poly(A)s was effective (Lindblad-Toh *et al*. 2000; JJ Agresti, pers. commun.). The SBE occurs in a reaction mixture with purified polymerase chain reaction (PCR) product containing the target SNP region (i.e. the SBE template).

We aligned all available sequences for each species using SEQUENCHER (GeneCodes) and located interspecific polymorphisms. We selected among all 'interspecies candidate SNPs' (Primmer *et al*. 2002) by identifying only SNPs that did not co-occur with intraspecific polymorphisms, and

those that were flanked by stretches of inter- and intraspecific monomorphic sequence of sufficient length for a primer. Most within-species SNPs are assumed to be biallelic (e.g. Kruglyak 1997; Wang *et al*. 1998; Glaubitz *et al*. 2003); consequently, closely related species rarely have more than two alleles at interspecific SNP loci (a pattern that was evident when comparing the four *Microtus* species here).

In order to diagnose an unknown DNA sample to species, SNP genotypes were assessed at multiple sites. SBE combinations were selected to minimize the number of template and SBE amplifications for a full diagnosis. SBE templates were amplified using primers for mammalian cytochrome *b* (Irwin *et al*. 1991) and *p53* (DeWoody 1999) that produced the smallest product containing the target SNPs.

SNP detection

We amplified two portions of the cytochrome *b* gene using primers L14724 and H15149 to amplify the first 557 bp (cyt*b*2), and L15775 and H15915 to amplify the last 153 bp (cyt*b*1) (Irwin *et al*. 1991). For cyt*b*1, genomic DNA was mixed with 0.2 mM dNTP, 1.5 mM $MgCl_2$, 2.0 U *Taq* DNA polymerase (Promega), 0.35 μM each primer and 1× Promega buffer. For cyt*b*2, primer concentrations were increased to 0.5 μM and the $MgCl_2$ concentration was increased to 2.0 mM. Both cytochrome *b* loci were amplified individually in a total reaction volume of 25 μL with the following thermal profile: 94 °C for 5 min, 35 cycles of 94 °C for 30 s, 40 °C for 1 min, 72 °C for 1 min, then 72 °C for 5 min.

We amplified a 479-bp region of the *p53* gene with p53C and p53H (DeWoody 1999), within which several combinations of three SNP loci would potentially differentiate species. The template amplification included 1.5 mM $MgCl_2$, 0.2 mM dNTP, 2.0 U *Taq* DNA polymerase (Promega), 1× Promega buffer, 0.25 μM each primer and genomic DNA in a total reaction volume of 25 μL. Thermal conditions were 94 °C for 5 min, 35 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min, then 72 °C for 2 min. All PCR amplifications were conducted in MJ Research thermal cyclers.

PCR products (i.e. SBE templates) were purified with QIAquick PCR Purification Kits (Qiagen). SBE template DNA was gel quantified and diluted. Although the SBE reaction does not require a kit (Lindblad-Toh *et al*. 2000), we used the ABI PRISM® SNaPshot™ kit (Applied Biosystems). We followed the manufacturer's recommendations for relative concentrations, including 0.2 μM primer, 2.5 μL SNaPshot Multiplex Ready Reaction Mix and 0.02–0.05 pmol SBE template in a final volume of 5 μL. The thermal profile consisted of 25 rapid-ramping cycles of 96 °C for 10 s, 58 °C for 5 s and 60 °C for 30 s. Post-extension products were dephosphorylated and then mixed 1:2 with formamide loading dye before loading 0.75 μL in 4.75% polyacrylamide gels (Long Ranger™, BMA) and running for 45′ on an ABI 377.

We used 11 control DNA samples from known species to optimize amplification and running conditions to minimize nonspecific amplification and noise in the electropherogram. Once reaction conditions were optimized, we genotyped the unknown samples at both the nuclear and mitochondrial genes. These data were then tabulated and each individual was identified to species based only on the SNP data. The species identifications based on the SNP data were then compared with those from the karyotypes in order to evaluate the efficacy of the SNP assays.

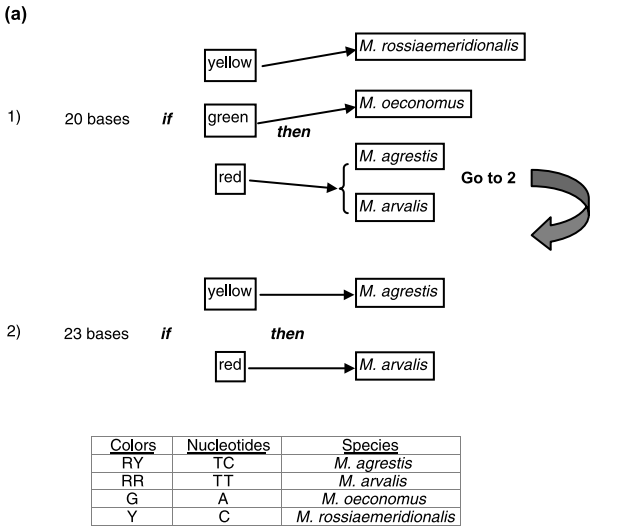
Post-hoc sequence confirmation

We sequenced the SBE template of some individuals that were either unscorable at either locus or were incorrectly assigned to a species different from that assigned by their karyotype. Sequencing was carried out in both directions (*p53*C and *p53*H, H15149 and L14724, H15915 and L15775). For difficult *p53* samples, we reamplified the template using *p53*C plus *p53*D primers (DeWoody 1999) in order to sequence from the *p53*D direction. Sequencing products were precipitated using a standard sodium acetate/ethanol protocol, resuspended in water, and visualized via an ABI 3700.

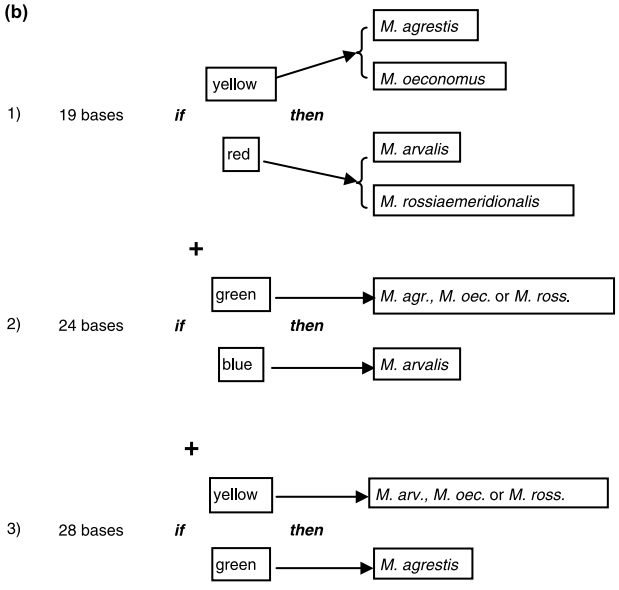
Results

The average pairwise difference among these four species at the mitochondrial cytochrome *b* gene is ≈ 11% (*Microtus agrestis* $n = 2$, *M. arvalis* $n = 16$, *M. oeconomus* $n = 1$, *M. rossiaemeridionalis* $n = 9$). Within the 1143 bp of this gene, there were few invariant locations of sufficient length for SBE primer design. Nevertheless, our first combination of two SBE primers (cyt*b*1 target base 1089: 5′-GCCTACTTCGCTATCATCGT-3′; cyt*b*2 target base 204: 5′-AAAGCATTCTCATCAGTAGCCCA-3′) worked well in singleplex reactions. Multiple adjustments in template concentration, primer concentrations and annealing conditions failed to result in successful multiplex reactions (Fig. 2a).

In contrast, the average pairwise difference among species is only 2% for the relevant portion of the *p53* gene (788 bp; DeWoody 1999) (*M. agrestis* $n = 1$, *M. arvalis* $n = 17$, *M. oeconomus* $n = 3$, *M. rossiaemeridionalis* $n = 10$). We designed and tested multiple primers before the SBE reaction identified the correct nucleotide in the controls. However, once we designed robust primers (p53-1 target base 348: 5′-AGGAAGGTTGCTGCAAGGT-3′; p53-2 target base 436: 5′-AAAAACCTGAGGTTTACTTGAACA-3′;



Colors	Nucleotides	Species
RY	TC	*M. agrestis*
RR	TT	*M. arvalis*
G	A	*M. oeconomus*
Y	C	*M. rossiaemeridionalis*



Colors	Nucleotides	Species
YGG	CAA	*M. agrestis*
RBY	TGC	*M. arvalis*
YGY	CAC	*M. oeconomus*
RGY	TAC	*M. rossiaemeridionalis*

Fig. 2 (a) Flow diagram of the two-steps of SBE reactions performed at the cytochrome *b* loci to diagnose unknown DNA samples to species. (b) Flow diagram of the multiplex SBE reactions performed at three *p53* loci to diagnose unknown DNA samples to species. In this case, we labelled: A = green, C = yellow, G = blue and T = red.

p53-3 target base 171: 5′-AAAAAAAAAAGCCAAAC CTAAGATCAAC-3′), little optimization was necessary for multiplex reactions (Fig. 2b).

Peaks at the *p53* loci were consistently cleaner and more intense than those at the cytochrome *b* loci. Ten of the eleven (91%) controls were correctly typed every time using the *p53* loci. Nine of the eleven (82%) controls were correctly typed every time using the combination of cytochrome *b* loci. Of the 27 unknown samples, we correctly assigned 20 (74%) to species using the *p53* loci, and 4 (19%) to species using the cytochrome *b* loci. However, only 23 of the 27 (85%) unknowns were unambiguously scorable using the *p53* loci. Of these, 20 of the 23 (87%) of the species calls with *p53* were correct. Few of the cytochrome *b* SBE amplifications were entirely clean and unequivocal. We nevertheless scored 21 of the 27 unknowns, after scoring rules were established based on the electropherogram profiles of the controls. However, only 4 of these 21 (19%) were correct.

Sequence data were obtained for six of the *p53*-miscalled unknown samples and the one control that was ambiguously scorable. The ambiguous control sample was a heterozygote for a 9 bp indel, which explains why the SBE calls were unreliable.

In the miscalled unknown individuals, the sequences at all three *p53* SBE priming sites matched the SBE primers exactly except in one instance. In that case, one mismatching base was observed 7 bp 5′ of the 3′-end of the p53-1 priming site in an individual known karyotypically to be *M. oeconomus*. However, in the same individual, the nucleotide at the target SNP corresponding to the mismatch-containing SBE priming site was that expected for *M. rossiaemeridionalis*, not *M. oeconomus*. Thus, the SBE reaction did accurately identify the nucleotides at all three SNPs for this *M. oeconomus* individual, but the presence of a previously unobserved interspecific polymorphism at this one site resulted in species misidentification. Two of the other five misclassified unknowns also showed polymorphisms at the same SNP site. As the SBE reactions correctly identified the nucleotides at those sites, it was the presence of unknown polymorphisms at the SNP sites that accounts for these misclassifications. The rest (three) of the six unknowns sequenced showed no mismatches at the *p53* SBE priming sites or unexpected nucleotides at the target SNP sites and thus are considered true miscalls.

At the cytochrome *b* regions that included the cyt*b*1 and cyt*b*2 SBE sites, sequence data were generated for 11 of the miscalled unknowns and 1 of the miscalled controls. Of these 12 sequences, 9 contained one or two mismatches at the priming sites, representing previously unknown polymorphisms at those sites. One of these sequences also revealed a new polymorphism at the actual SNP site. The remaining three sequences matched the SBE priming sites exactly, though one also revealed a new polymorphism at the SNP site.

Discussion

In principle, SNPs should be useful in studies of many biological phenomena, including recombination rates,

introgression and species identification. In practice, we had a final success rate for SNP genotyping with the SBE method of 89% after accounting for the heterozygote and the three cases of unknown polymorphisms at the *p53* locus. However, the success rate at the mitochondrial locus was less than that expected by chance alone.

Our SBE amplifications failed in two unknown samples at all of the *p53* loci, even after multiple tries, although it succeeded at the cytochrome *b* loci for those same individuals. In the other 31 ambiguous cases (which include every SBE amplification event, i.e. single or multiplexed, *p53* or cytochrome *b*), scoring difficulties were generally attributed to multiple bands on the gel at or near the expected size range. These extra peaks most likely resulted from nonspecific binding of the SBE primers at other locations on the SBE template. Other possibilities include the single base extension of residual PCR primers in the SBE template mixture from incomplete PCR purification or single base extensions of template molecules. However, the size of either of these products is often predictable (i.e. primer or template size plus one nucleotide) and thus they can at times be excluded from scoring. Another source of error is single or multiple base extensions with residual, unlabelled dNTPs from the original PCRs that were not completely removed during PCR purification. The SBE is terminated and visualized by the addition of labelled ddNTPs, thus incomplete removal of dNTPs used in PCR should result in a series of peaks (one nucleotide apart) that are similar in appearance to a short sequencing reaction.

As with PCR primer design, nonspecific SBE primer binding could not always be predicted by evaluating sequence homology a priori. Therefore, SBE primer optimization must proceed empirically. We present a list of troubleshooting suggestions in Table 1.

It is clear that the *p53* assays produced more repeatable and accurate SBE results than the cytochrome *b* assays. We considered further optimization or redesigning of the SBE primers at the cytochrome *b* locus, but there were few candidate diagnostic SNPs (i.e. species-specific SNPs in otherwise monomorphic regions) in the gene. The incidence of mitochondrial heteroplasmy in these voles is generally low (Baker *et al*. 1999), but we considered the possibility that it is still sufficiently high to produce variable PCR products. However, the sequencing results for a subset of the unscorable and miscalled individuals at the *p53* and the cytochrome *b* loci helped to explain the discrepancy between the loci. After accounting for unknown polymorphisms, the accuracy rate of the SBE reactions in the conserved *p53* gene (2% interspecies sequence divergence) was quite high (89%). It is not possible to calculate an accuracy rate at the cytochrome *b* locus even after completing the sequencing, because most of the SBE reaction products were unscorable. However, the post-hoc sequencing results indicate that in a highly polymorphic region such as

Table 1 Troubleshooting suggestions for SBE optimization

Problem	Source	Solution
Too many peaks near target peak	Nonspecific SBE primer annealing	Restrict SBE template to ≤ 150 bp Minimize the number of SBE templates in a multiplex reaction Increase annealing temperature in SBE reaction Decrease template or primer concentration in SBE reaction Design longer SBE primers Design SBE primers to anneal to the opposite strand Design SBE primers targeted at different SNPs
	Extension of other molecules (leftover PCR primers, truncated SBE template or SBE primers)	Use SAP, or other noncolumn clean-up of PCR product before SBE reaction Purchase PAGE purified SBE primers
Series of nonoverlapping, adjacent peaks near target size	Excess dNTPs from PCR permitting short sequencing reaction	Use SAP, or other noncolumn clean-up of PCR product before SBE reaction

the cytochrome *b* gene (11% interspecies sequence divergence), hidden polymorphisms are likely to undermine SBE scoring by: (i) altering the nucleotides at the actual target SNPs, (ii) altering priming conditions at the SBE sites such that the primers cannot anneal reliably, and (iii) creating alternative primer annealing sites elsewhere on the SBE template. Others have successfully accommodated priming site polymorphisms using degenerate SBE primers (G-P Saetre, pers. commun.). Of course, such polymorphisms must be identified via DNA sequencing prior to SNP genotyping.

At any locus, many sequences are required in order to detect most of the extant polymorphisms at the target SBE sites. In our case, we selected candidate interspecies diagnostic SNPs and designed the SBE primers based on between 1 and 17 sequences per species, and this was (in hindsight) inadequate. Specifically at the *p53* locus, the three cases of previously unobserved polymorphism at the SNP sites were in species for which we had few (1–3) sequences from which to design our assays. In the case of cytochrome *b* (and in other variable genes), it may not be practical to identify all possible intra- and interspecific sequence variation a priori.

The design of our SNP genotyping assays likely contributed to their mixed success. For example, although we followed the purification and clean-up instructions suggested by the ABI PRISM® SNaPshot™ kit, detailed inquiry and experience leads us to speculate that more rigorous procedures should be followed (Table 1). Assay success also was predicated on the number and variability of the sequences in GenBank. For instance, had more sequences been available for review at cytochrome *b*, then we would have identified a greater proportion of the polymorphic sites. The a priori knowledge of those polymorphisms would have allowed the design of more robust assays.

We set out to determine whether SNP diagnostic techniques are practical for use with nonmodel species. Primmer *et al*. (2002) examined three methods for obtaining adequate sequence data: (i) designing PCR primers based on comparative sequence data to amplify and then sequence regions in the target species, (ii) sequencing target species' microsatellite alleles to find SNPs in flanking sequences, and (iii) sequencing random clones from libraries created from target species DNA. They showed that these are all viable approaches to developing SNP analysis techniques for wild species. In this experiment, we tested the efficacy of the SBE method for genotyping SNPs that were identified by comparative DNA sequencing.

We discovered that the SBE method requires extensive time and labour to optimize, and even then may not be 100% accurate. Our stated accuracy rate of ≈ 90% at the conserved nuclear locus refers to the actual accuracy of the assays, which were known only after post-hoc sequencing was completed. In practice, we accurately identified an unknown's correct species 74% of the time with *p53* and 19% of the time using cytochrome *b*. We have shown that many individuals must be sequenced before designing SNP surveys in order to detect as much underlying polymorphism as possible a priori. In addition, the use of positive controls of known sequence during the optimization phase is absolutely critical to be able to assess the accuracy of the SNP diagnosis. Based on the results of this study, several informative candidate SNPs may be required to develop a diagnostic SNP toolkit that is robust. Large-scale studies may benefit from the time and effort invested in SNP genotyping optimization, as SNP genotyping has the potential to be quicker and easier than other genotyping methods. It also has potential in applications such as characterizing the sequence variation in microsatellite flanking regions (e.g. Makova *et al*. 2000; Blankenship *et al*. 2002). With currently available methods and technology, SBE assays of SNPs are akin to the development of novel microsatellite markers: they require a priori sequence data and extensive optimization. Unfortunately for some applications, 100 independently segregating SNPs are no more informative than 16–20 polymorphic microsatellite loci (Glaubitz *et al*. 2003; see also Wang *et al*. 1998). However, SNPs are more amenable to large-scale automation (e.g. DNA chips) and thus, many laboratories are currently pursuing methods to apply SNPs to genotyping applications for nonmodel organisms.

Acknowledgements

We thank J. Glaubitz, Olin E. Rhodes and members of the DeWoody laboratory. This is ARP publication #16964 from Purdue University.

References

Baker RJ, DeWoody JA, Wright AJ, Chesser RK (1999) On the utility of heteroplasmy in genotoxicity studies: an example from Chornobyl. *Ecotoxicology*, **8**, 301–309.

Baker RJ, Hamilton MJ, Van Den Bussche RA *et al*. (1996a) Small mammals from the most radioactive sites near the Chernobyl nuclear power plant. *Journal of Mammology*, **77**, 115–170.

Baker RJ, Van Den Bussch RA, Wright AJ *et al*. (1996b) High levels of genetic change in rodents of Chernobyl. *Nature*, **380**, 707–708.

Bensch S, Åkesson S, Irwin DE (2002) The use of AFLP to find an informative SNP: genetic differences across a migratory divide in willow warblers. *Molecular Ecology*, **11**, 2359–2366.

Blankenship SM, May B, Hedgecock D (2002) Evolution of a perfect simple-sequence-repeat locus in the context of its flanking sequence. *Molecular Biology and Evolution*, **19**, 1943–1951.

Conroy CJ, Cook JA (1999) MtDNA evidence for repeated pulses of speciation within arvicoline and murid rodents. *Journal of Mammology Evolution*, **6**, 221–245.

Conroy CJ, Cook JA (2000) Molecular systematics of a holarctic rodent (*Microtus*: Muridae). *Journal of Mammology*, **81**, 344–359.

DeWoody JA (1999) Nucleotide variation in the *p53* tumor-suppressor gene of voles from Chernobyl, Ukraine. *Mutation Research*, **439**, 25–36.

DeWoody JA, Chesser RK, Baker RJ (1999) A translocated mitochondrial cytochrome *b* pseudogene in voles (Rodentia: *Microtus*). *Journal of Molecular Evolution*, **48**, 380–382.

Glaubitz JC, Rhodes OE, DeWoody JA (2003) Prospects for inferring pairwise relationships with single nucleotide polymorphisms (SNPs). *Molecular Ecology*, **20**, 1039–1047.

Greig C, Robertson JM, Banks MA (2002) Rapid PCR-based species tests for threatened sympatric salmonids. *Conservation Genetics*, **3**, 83–86.

Irwin DM, Kocher TD, Wilson AC (1991) Evolution of the cytochrome *b* gene of mammals. *Journal of Molecular Evolution*, **32**, 128–144.

Kreiser BR (2001) Mitochondrial cytochrome *b* sequences support recognition of two cryptic species of plains killifish, *Fundulus zebrinus* and *Fundulus kansae*. *American Midland Naturalist*, **146**, 199–209.

Kruglyak L (1997) The use of a genetic map of biallelic markers in linkage studies. *Nature Genetics*, **17**, 21–24.

Kwok P (2001) Methods for genotyping single nucleotide polymorphisms. *Annual Review of Genomics and Human Genetics*, **2**, 235–258.

Lee C (2002) Irresistible force meets immovable object: SNP mapping of complex diseases. *Trends in Genetics*, **18**, 67–69.

Lindblad-Toh K, Winchester E, Daly MJ *et al.* (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nature Genetics*, **24**, 381–386.

Makova KD, Nekrutenko A, Baker RJ (2000) Evolution of microsatellite alleles in four species of mice (Genus *Apodemus*). *Journal of Molecular Evolution*, **51**, 166–172.

Martin AP, Bermingham E (2000) Regional endemism and cryptic species revealed by molecular and morphological analysis of a widespread species of neotropical catfish. *Proceedings of the Royal Society of London B*, **267**, 1135–1141.

Martin Y, Gerlach G, Schlotterer C, Meyer A (2000) Molecular phylogeny of European muroid rodents based on complete cytochrome *b* sequences. *Molecular Phylogenetics and Evolution*, **16**, 37–47.

Meksem K, Ruben E, Hyten D, Triwitayakorn K, Lightfoot DA (2001) Conversion of AFLP bands into high-throughput DNA markers. *Molecular Genetics and Genomics*, **265**, 207–214.

Nekrutenko A, Makova KD, Baker RJ (2000) Isolation of binary species-specific PCR-based markers and their value for diagnostic applications. *Gene*, **249**, 47–51.

Primmer CR, Borge T, Lindell J, Saetre G (2002) Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. *Molecular Ecology*, **11**, 603–612.

Ryan PG, Bloomer P (1999) The long-billed lark complex: a species mosaic in southwestern Africa. *Auk*, **116**, 194–208.

Sambrook J, Russell DW (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schneider C, Moritz C (1999) Rainforest refugia and evolution in Australia's wet tropics. *Proceedings of the Royal Society of London B*, **266**, 191–196.

Syvanen AC (1999) From gels to chips: 'minisequencing' primer extension analysis of point mutations and single nucleotide polymorphisms. *Human Mutation*, **13**, 1–10.

The International SNP Working Group (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, **409**, 928–933.

Wang DG, Fan J-B, Siao C-J *et al.* (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, **280**, 1077–1082.

Wilson DE, Reeder DM, eds. (1993) *Mammal Species of the World: A Taxonomic and Geographic Reference*, 2nd edn. Smithsonian Institution Press, Washington, DC.

Witt JDS, Hebert PDN (2000) Cryptic species diversity and evolution in the amphipod genus *Hyalella* within central glaciated North America: a molecular phylogenetic approach. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 687–698.

N. M. Belfiore is a population geneticist and ecologist who uses genetic methods to address complex questions in ecology and evolution. F. G. Hoffman and R. J. Baker are collaborators with long-standing interests in mammalian evolution. J. A. DeWoody's laboratory applies molecular markers and computer simulations to problems in conservation and population genetics.
