Molecular Ecology 1997, 6, 575–580

SHORT COMMUNICATION

Detection and isolation of nuclear haplotypes by PCR-SSCP

G. ORTÍ, M. P. HARE* and J. C. AVISE
Department of Genetics, University of Georgia, Athens, GA 30602, USA

Abstract

SSCP (single-strand conformational polymorphism) is used widely in the field of human biomedicine, but its potential as a population genetics tool for the recovery of nuclear gene genealogies remains to be realized. We describe and illustrate a use for SSCP in the physical isolation of nuclear haplotypes that circumvents several difficulties associated with more conventional cloning procedures. The DNA sequence can be determined directly from the isolated haplotypes and used for phylogenetic inference. SSCP provides a convenient first step toward generating nuclear genealogies for population studies.

Keywords: SSCP, allelic variation, nuclear haplotypes, gene genealogies, *Limulus*, *Crassostrea*

Received 10 September 1996; revision accepted 14 December 1996

Introduction

The concept of gene trees – the phylogeny of alleles or haplotypes for any specified stretch of DNA – has prompted phylogenetic thought and historical perspectives in population biology (e.g. Wilson *et al.* 1985; Avise *et al.* 1987; Avise 1989). However, due to both biological and technical complications, few empirical studies have assessed nuclear gene genealogies in natural populations (e.g. Hey & Klinman 1993; Bernardi *et al.* 1993; Palumbi & Baker 1994; Vogler & DeSalle 1994). The primary biological snag is historical intragenic recombination among homologous haplotypes, a phenomenon that requires statistical methods for detection (Sawyer 1989; Hein 1990; Maynard Smith 1992; Templeton *et al.* 1992) and if common can obscure the otherwise lineal histories of allelic descent. The primary technical difficulty, and the one addressed in the current paper, is the physical isolation of individual haplotypes from diploid tissue sources.

Here we describe and exemplify a general approach for isolating single-locus haplotypes from diploid organisms. The approach involves 'single-strand conformational polymorphism' or SSCP (Orita *et al.* 1989a,b), a popular technique that allows rapid detection and isolation of nucleotide polymorphisms in nondenaturing gels. Since its inception, over 1100 publications have reported the use of SSCP, mainly in biomedical research and as a clinical diagnostic tool for human disease mutations (BIOSYS, Philadelphia, PA). Only a small fraction of these studies (33 publications, < 3%) were indexed as 'population genetics', but most of the latter (30) were surveys of human populations, with none involving a genealogical perspective. The potential of SSCP for evolutionary biology is not yet fully appreciated (Lessa & Appelbaum 1993; Potts 1996).

Correspondence: Guillermo Ortí. Fax: + 1-706-542-3910. E-mail: gorti@bscr.uga.edu

*Present address: Department of Organismic and Evolutionary Biology, Harvard University Biolabs, 16 Divinity Ave., Cambridge, MA 02138, USA.

Materials and Methods

DNA samples from American oysters *Crassostrea virginica* were procured for previous studies, and PCR primers for anonymous single-copy nuclear (scn) DNA loci were developed (Karl & Avise 1992; Hare *et al.* 1996). For the purposes of an example, the isolation of haplotypes from one such locus (CV-32) is presented here. To assess the efficiency of SSCP for detection and isolation of haplotypes, only individuals known to be heterozygous for a restriction site at this locus (Hare *et al.* 1996) were tested. Amplification by PCR was performed with 0.25 μM of each primer, CV-32.3LS (5'- CTGTACTTGTTTTCCCGTT-3') and CV-32.7R (5'-CATTCATTTTCCTCCAAGTTTTGTT-3'), 200 μM of each dNTP, 1 mM $MgCl_2$, 1 × reaction buffer, and 1 unit *Taq* polymerase (Promega), in a final volume of 50 μL. Cycling conditions were 94 °C for 1 min, 51 °C for 45 s, 72 °C for 1 min, for 25 cycles. For verification prior to SSCP analysis, PCR products of the expected size (447 bp) were visualized in ethidium bromide-stained 2% agarose gels.

From the horseshoe crab *Limulus polyphemus* (originally collected by Saunders *et al.* 1986), scnDNA loci were obtained by a procedure similar to that described by Sültmann *et al.* (1995). Briefly, randomly amplified fragments generated by the PCR with a non-specific primer (5′-CCTTATATAAAGCAAACCAGACGGCAGCA-3′) were size-selected and cloned using the pGem-T vector system (Promega). Cloned inserts were sequenced from various individuals, aligned and searched for polymorphisms. Specific PCR primers were designed based on the alignment. Several scnDNA loci were obtained by this procedure (manuscript in preparation), but for purposes of example only the results from locus LP-1 are shown here. Specific primers for this locus (LP1s: 5′-CTTTAAGCTGCTACCCTAAC-3′; LP1a: 5′-CAGAGATAGAGACGACAGTTG-3′) amplify a fragment about 775 bp long. Amplification by PCR was performed with 0.4 μM of each primer, 1 mM of each dNTP, 2 mM $MgCl_2$, 1 × PCR buffer, and 1 unit of *Taq* polymerase, in a final volume of 50 μL. Cycling conditions were 94 °C for 1 min, 51 °C for 1 min, 72 °C for 1 min, for 29 cycles. The presence of a single clear band was verified in 2% agarose gels prior to SSCP analysis.

For both the oysters and horseshoe crabs, nonisotopic SSCP analysis was conducted on the PCR-amplified fragments following the protocol of Hongyo *et al.* (1993), with some modifications. Briefly, 10–20 μL of unpurified PCR product (roughly 0.5–1.6 μg of DNA) were mixed with 5 μL of denaturing/loading buffer containing 0.4 μL of 1 M methylmercury hydroxide (Matthey Electronics, Inc., War Hill, MA), 1 μL of 15% Ficoll loading buffer (with 0.25% bromophenol blue and 0.25% xylene cyanol), and 3.6 μL of 1 × TBE buffer (90 mM Tris, 92 mM boric acid, 2.5 mM EDTA). This mixture was denatured for 4 min at 85 °C and immediately chilled on ice before loading to the SSCP gel.

Six to 16% polyacrylamide (39 : 1 acrylamide to bisacrylamide) TBE gels (16 cm × 14 cm × 1.5 mm) were run with 1 × TBE buffer on a vertical electrophoresis system (Fisher Biotech model VE16-1). Refrigerated water from a thermostatically controlled circulator (Brinkmann RC 20B, Lauda, Germany) was passed through a vertical cooling chamber to maintain the upper buffer and gel at a constant temperature. In a modification of the gel rig, air bubbles were introduced into the upper buffer to eliminate thermal gradients between the cooling chamber and gel. Gels were run at constant power (8–12 W) for 8–20 h, and with the upper buffer temperature set to 3–15 °C. Temperature, power, acrylamide concentration and fragment size affected the running time. Acrylamide concentration and temperature were optimized for each fragment to obtain maximum separation of allelic variants.

Gels were stained for at least 20 min with a 2 μg/mL ethidium bromide solution, and destained for 5 min with distilled water. Bands were visualized and photographed under UV light, and a small fraction of each band was excised from the gel with the tip of a 200-μL glass micropipette. These acrylamide plugs were placed individually in tubes with 50 μL of distilled water and stored at –20 °C. Prior to reamplification, the gel samples were heated to 80 °C for 10 min.

Genetic differences among the haplotypes isolated by SSCP were verified by direct sequencing, either from single stranded template generated by asymmetric PCR (Gyllensten & Erlich 1988), or from reamplified double-stranded products (fmol cycle-sequencing, Promega), and the results will be presented elsewhere in the context of phylogenetic analyses.

Results

The ability of SSCP to detect and isolate nuclear haplotypes is summarized in Table 1. Oyster samples (n = 47) from Massachusetts to Louisiana revealed 21 distinct alleles at the CV-32 locus. Figure 1 shows the effect of

	CV-32	LP-1
No. individuals assayed by SSCP	47	44
No. individuals showing band separation in SSCP gels*	37	26
No. unique alleles isolated	21	11
Minimum difference between alleles showing band separation	2 bp subst. + 1 bp-indel (lane 7, Fig. 1)	1 bp subst. (lane 7, Fig. 2a)
Maximum difference between alleles showing band separation	8 bp subst. + 1 bp-indel (not shown)	6 bp subst. + 39 bp-indel (lane 6, Fig. 2a)

* Three- or four-banded pattern, allowing excision of individual haplotypes from the gel.

Table 1 Results of SSCP assays for oysters (locus CV-32) and horseshoe crabs (locus LP-1). About 79% of the oyster heterozygotes were detected by SSCP, since all 47 individuals tested were known to be heterozygous for a restriction site (Hare *et al.* 1996). Genotypes of the horseshoe crabs were unknown at the outset, except for eight individuals (among the 26 that showed band separation in SSCP gels) that were heterozygous for a 39-bp deletion (Fig. 1)

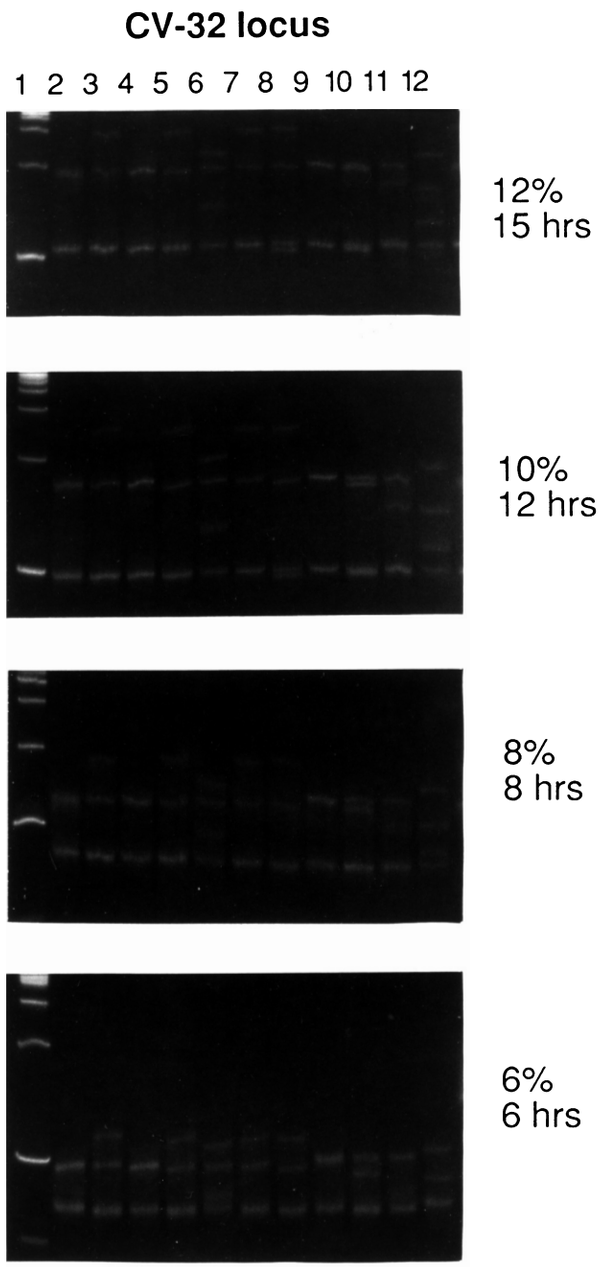
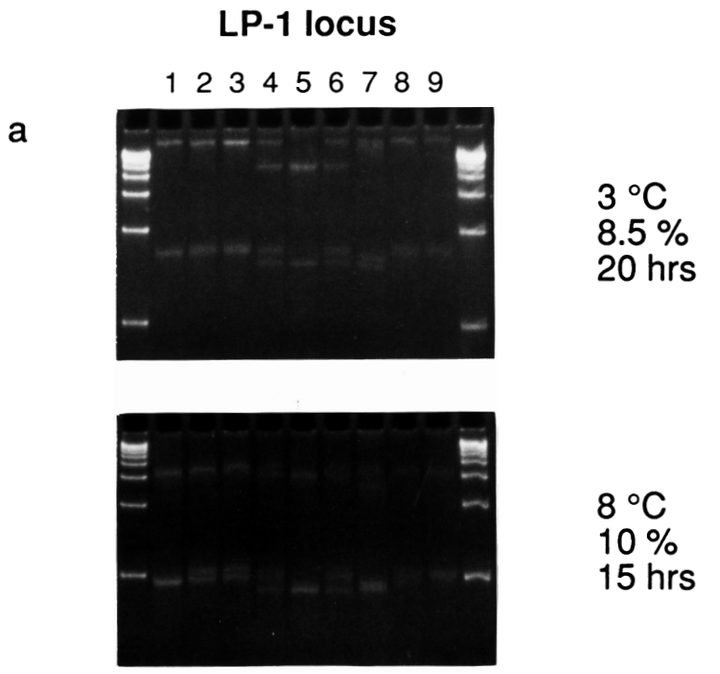


Fig. 1 Separation of allelic variants of the CV-32 locus (447 bp) by SSCP analysis at 3 °C. Polyacrylamide gels (6–12%, bottom to top) were run at 3 °C (upper buffer temperature) for 6–15 h (bottom to top) at constant power (9 W). Each lane of the gel was loaded with PCR products from single individuals, whose genotypes were as follows (upper/lower case indicates absence/presence of a restriction site). Lane: (1) 1-kb ladder; (2) n_1n_1; (3) n_1N_1; (4) n_2N_2; (5) n_3N_1; (6) n_4N_3; (7) n_4N_4; (8) n_5N_1; (9) n_9N_5; (10) n_6N_5; (11) n_7N_5; (12) n_8N_6. Variation among alleles (determined by sequencing) included 1–9 bp substitutions and 1–3 indels with length variation from 1 to 4 bp.

acrylamide concentration on band separation. Homozygous individuals (e.g. lane 2), or heterozygotes whose haplotypes fail to separate (e.g. lane 9) typically show a two-banded pattern, corresponding to the pyrimidine-rich and purine-rich DNA strands. Successful isolation of haplotypes from heterozygotes is evidenced when the upper and/or lower strands separate in the gel (resulting in three- or four-banded patterns, e.g. lanes 5 and 6, respectively). Note that the lower bands of heterozygote n_5N_1 in lane 8 separated only at gel concentrations ≥ 10%. By sequencing, these alleles proved to differ by six nucleotide substitutions. Conversely, good separation of the upper bands in heterozygote n_6N_5 in lane 10 (four nucleotide substitutions) was obtained in 6–10% acrylamide gels but not in 12% gels. In general, optimal acrylamide gel concentrations for the CV-32 locus was determined empirically (Fig. 1) to be about 8–10%, in agreement with previous reports (Glavac & Dean 1993).

For horseshoe crabs, the primer-pair LP1s-LP1a successfully amplified a single 775-bp fragment in most individuals tested. Samples from New Hampshire to the Gulf of Mexico in Florida (n = 44) revealed 11 distinct alleles at the LP-1 locus (Table 1). Optimal conditions for the separation of alleles were obtained at 7–8 °C with 9–10% polyacrylamide gels (Fig. 2a). In lanes 4 and 6, the four-banded patterns correspond to genotypes L_1S_1 and L_6S_1 that carry a 39-bp deletion (Fig. 2b) in one of the alleles. In lanes 2 and 3, genotypes L_1L_4 and L_1L_5 (the alleles proved to differ by three nucleotide substitutions) separated at their lower strands at 8 °C, but not at 3 °C. In some cases (e.g. the heterozgyote in lane 7), the lower strands separated at one temperature (3 °C) and the upper strands separated at a different temperature (8 °C). This last observation demonstrates that no single optimum set of conditions for allelic separation need exist among multiple heterozygous genotypes. Three randomly picked samples (out of 18) that did not show band separation at the LP-1 locus were confirmed to be homozygotes upon sequencing, suggesting a relatively high efficiency of this technique (Table 1).

In summary, numerous variant haplotypes at scnDNA loci from oysters and horseshoe crabs were isolated successfully by the PCR-SSCP approach. Most prior applications of SSCP report high efficiency for the detection of mutations in DNA fragments ≤ 200 bp-long, but fragments as large as 775 bp were isolated and sequenced in this study. Other authors (Hayashi 1992; Hayashi & Yandell 1993; Glavac & Dean 1993) have concluded that the efficiency of SSCP is highly sequence dependent and, thus, may vary significantly among fragments, depending on running conditions optimized for each locus. Alterations in single-strand DNA mobility in nondenaturing electrophoretic gels may be predicted by DNA folding computer software (Nielsen *et al.* 1995), but for most



b

```
CTTTAAGCTGCTACCCTAAC->TCGGGCTTGAGTAAAAATAGTGTTAACTTTGTTTATAA
AAGATCATATTTTGTTGTGGAGCAAATCAGAAACCAGTTGTTAAAACTGTATTTGTTGAT
TGGGTTATCCAGGTTTTATTGTATGATATATGTGATTTCACGTTTAATGAAATTAAATTA
TCTTTTAGGAATATTTATTTAGTTGGATACAAATAAGGAAAAAATAGGGAACATACAATG
CCAAGAACTACTGCACCAAAATTATGTAGAAAATTTTAATACAAGACT
  (1)   CAAACAAGTATGTCCTCCTGTTTTAGCTTGAATAAAAAACA
  (2)   CAAACAATTATATCCTCCTGTTTTAGTTTGATTAAAAGACA
  (3)   CAAACAAGCATACCCT--GGAGTTTGTCAGAATACAAGACA
  (4)   CAAACAAGTACACACT--ATTTTTAGTTTGAATACAAGACA
  (5)   CAAACAAGCATACCTTGATGTTTTAGTCTGAATACAAGACA
  (6)   CAAATAAGCATACCCTGATGTTTTAGTTTGAATATAAGACA
AAAACAAGCATACTGTTACGTCTCTGGTCGTTAGTTAATTTTCCGTAAAAATTCTTTCCG
AAGAAAATTAGCGATACAAGTCCTTTCTTAAATAAAACTAGTTTATTCTTTCTCAATACA
ACTTTGCAATACACTTAATTTTCATAGTCACGCTATAACTTAAACAAACGCCATAATTCA
CAACGAACGAACGATTCACgcgcttgatt<-CAACTGTCGTCTCTATCTCTG
```

Fig. 2 (a) Separation of allelic variants at the LP-1 locus by SSCP analysis under different conditions of temperature, acrylamide concentration, and running time. Both gels were run at constant power (9 W). Each lane was loaded with PCR products amplified from individual horseshoe crabs, with genotypes as follows. Lane: (1) L_1L_1; (2) L_1L_4; (3) L_1L_5; (4) L_1S_1; (5) S_1S_1; (6) L_6S_1; (7) L_8L_9; (8) L_2L_7; (9) L_2L_2. Left- and right-most lanes are a 1-kb ladder. (b) DNA sequence (775 bp) of the LP-1 locus (primers are shown at both ends, marked by arrows). Six repeats of an imperfectly conserved 41-bp minisatellite are numbered. The third repeat is absent in the 'short' alleles (designated 'S'). The other polymorphic sites (shown bold and underlined) vary among the S and the L ('long') alleles.

applications empirical assessments are necessary. Our data support the idea that temperature and gel concentration are critical parameters affecting mobility shifts that allow separation by SSCP.

Discussion

Several experimental approaches and specialized genetic systems have been suggested for isolating individual haplotypes from nuclear loci (review by Avise 1994: 134). These include: (i) extraction of identical-by-descent chromosomes via controlled crosses (e.g. in *Drosophila*, Aquadro *et al.* 1986); (ii) use of haploid tissues, or of haploid phases of a haplo-diploid life cycle (e.g. Guttman & Dykhuizen 1994); (iii) use of sex-linked genetic markers (e.g. Bishop *et al.* 1985); and (iv) cloning of PCR products (Scharf *et al.* 1986). This latter approach clearly has the broadest taxonomic applicability. However, because cloning occurs through a single molecule, the possibility of nucleotide misincorporation by *Taq* polymerase must be considered (e.g. Keohavong & Thilly 1989; Palumbi & Baker 1994). In practice, possible *Taq* misincorporations either have been ignored for phylogenetic analyses (Palumbi & Baker 1994; Vogler & DeSalle 1994), or several separate clones from each individual were sequenced to distinguish allelic variation present in the population from PCR-cloning artefacts (Bernardi *et al.* 1993). PCR-SSCP offers a simpler and less laborious alternative to the direct cloning of PCR products because it bypasses the problem of *Taq* error and the need for multiple sequencing of each haplotype.

Another advantage conferred by SSCP separation (compared with the cloning of PCR products) is the ability

to detect haplotype variants with low amplification efficiencies. When PCR amplification from a heterozygote yields a low copy number of one allele, many clones must be sequenced to isolate both haplotypes from an individual. In contrast, a mutant haplotype in SSCP gels can be detected when it comprises as little as 3% of a mixture of PCR products (Hongyo *et al.* 1993; Law *et al.* 1996). An alternative broadly applicable method for allele-specific sequencing using restriction enzymes and biotinylation has been described recently (Zhang & Hewitt 1996). Shortcomings of this method are that interallelic sequence differences must be known a priori (a special difficulty if length variation exists among alleles), and that 'restriction cuttable' sites for the polymorphic DNA sequences in question must be identified.

An essential characteristic sought for target loci in population genealogical analysis is the presence of substantial polymorphism in a relatively short stretch of DNA. In general, shorter fragments are easier to separate in SSCP gels and also should be less likely to have undergone intragenic recombination in recent history. Conversely, larger fragments should tend to accumulate more polymorphic sites, but SSCP separation becomes less efficient and recombination perhaps more likely. In conclusion, several advantages exist for SSCP in the detection and isolation of nuclear haplotypes for gene-genealogical studies at the population level. With physical methods for haplotype isolation now available, it will next be important to determine the extent to which rapidly evolving nuclear genes can be identified that also are relatively free of homoplasy and intragenic recombination.

Acknowledgements

D. Guttman has brought to our attention the potential of SSCP for this particular application. Research was supported by a grant from the NSF to J.C.A.

References

Aquadro CF, Desse SF, Bland MM, Langley CH, Laurie-Ahlberg CC (1986) Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaste*r. *Genetics*, **114**, 1165–1190.

Avise JC (1989) Gene trees and organismal histories: A phylogenetic approach to population biology. *Evolution*, **43**, 1192–1208.

Avise JC (1994) *Molecular Markers, Natural History and Evolution*. Chapman & Hall, New York.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, **18**, 489–522.

Bernardi G, Sordino P, Powers DA (1993) Concordant mitochondrial and nuclear DNA phylogenies for populations of the teleost fish *Fundulus heteroclitus*. *Proceedings of the National Academy of Science USA*, **90**, 9271–9274.

Bishop CE, Boursot P, Baron B, Bonhomme F, Hatat D (1985) Most classical *Mus musculus domesticus* laboratory mouse strains carry *Mus musculus musculus* Y chromosomes. *Nature*, **315**, 70–72.

Glavac D, Dean M (1993) Optimization of the single-strand conformation polymorphism (SSCP) technique for detection of point mutations. *Human Mutation*, **2**, 404–414.

Guttman DS, Dykhuizen DE (1994) Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science*, **266**, 1380–1383.

Gyllensten UB, Erlich HA (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQa locus. *Proceedings of the National Academy of Science USA*, **85**, 7652–7655.

Hare MP, Karl SA, Avise JC (1996) Anonymous nuclear DNA markers in the American oyster and their implications for the heterozygote deficiency phenomenon in marine bivalves. *Molecular Biology and Evolution*, **13**, 334–345.

Hayashi K (1992) PCR-SSCP: A method for detection of mutations. *GATA*, **9**, 73–79.

Hayashi K, Yandell DW (1993) How sensitive is PCR-SSCP? *Human Mutation*, **2**, 338–346.

Hein J (1990) Reconstructing evolution of sequences subject to recombination using parsimony. *Mathematical Biosciences*, **98**, 185–200.

Hey J, Klinman RM (1993) Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* complex. *Molecular Biology and Evolution*, **10**, 804–822.

Hongyo T, Buzard GS, Calvert RJ, Weghorst CM (1993) 'Cold SSCP': a simple, rapid and non-radioactive method for optimized single-strand conformation polymorphism analyses. *Nucleic Acids Research*, **21**, 3637–3642.

Karl SA, Avise JC (1992) Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. *Science*, **256**, 100–102.

Keohavong P, Thilly WG (1989) Fidelity of DNA polymerases in DNA amplification. *Proceedings of the National Academy of Science USA*, **86**, 9253–9257.

Law JC, Facher EA, Deka A (1996) Nonradioactive single-strand conformation polymorphism analysis with application for mutation detection in a mixed population of cells. *Analytical Biochemistry*, **236**, 373–375.

Lessa EP, Appelbaum G (1993) Screening techniques for detecting allelic variation in DNA sequences. *Molecular Ecology*, **2**, 119–129.

Maynard Smith J (1992) Analyzing the mosaic structure of genes. *Journal of Molecular Evolution*, **34**, 126–129.

Nielsen DA, Novoradovsky A, Goldman D (1995) SSCP primer design based on single-strand DNA structure predicted by a DNA folding program. *Nucleic Acids Research*, **23**, 2287–2291.

Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T (1989a) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proceedings of the National Academy of Science USA*, **86**, 2766–2770.

Orita M, Suzuki Y, Sekiya T, Hayashi K (1989b) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics*, **5**, 874–879.

Palumbi SR, Baker CS (1994) Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Molecular Biology and Evolution*, **11**, 426–435.

Potts WK (1996) PCR-based cloning across large taxonomic distances and polymorphism detection: MHC as a case study. In: *Molecular Zoology* (eds. Ferraris JD, Palumbi SR), pp. 181–194. Wiley-Liss, Inc., New York.

Saunders NC, Kessler LG, Avise JC (1986) Genetic variation and geographic differentiation in mitochondrial DNA of the horseshoe crab, *Limulus polyphemus*. *Genetics*, **112**, 613–627.

Sawyer S (1989) Statistical tests for detecting gene conversion. *Molecular Biology and Evolution*, **6**, 526–538.

Scharf SJ, Horn GT, Erlich HA (1986) Direct cloning and sequence analysis of enzymatically amplified genomic sequences. *Science*, **233**, 1076–1078.

Sültmann H, Mayer WE, Figueroa F, Tichy H, Klein J (1995) Phylogenetic analysis of cichlid fishes using nuclear DNA markers. *Molecular Biology and Evolution*, **12**, 1033–1047.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**, 619–633.

Vogler AP, DeSalle R (1994) Evolution and phylogenetic information content of the ITS-1 region in the tiger beetle *Cincindella dorsalis*. *Molecular Biology and Evolution*, **11**, 393–405.

Wilson AC, Cann RL, Carr SM *et al.*(1985) Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society*, **26**, 375–400.

Zhang D.X., Hewitt GM (1996) An effective method for allele-specific sequencing using restriction enzyme and biotinylation (ASSURE B). *Molecular Ecology*, **5**, 591–594.

Guillermo Ortí is a postdoctoral associate and Matt Hare a recent PhD graduate from the laboratory of John Avise at the University of Georgia. This work is part of a broader research program concerned with the principles and processes governing the geographic distributions of genealogical lineages (phylogeography). M. Hare is now a postdoctoral associate with S. Palumbi at Harvard University.
