UC Irvine
UC Irvine Previously Published Works

Title

PCR-based assays of mendelian polymorphisms from anonymous single-copy nuclear DNA: techniques and applications for population genetics.

Permalink
https://escholarship.org/uc/item/5c9414pc

Journal
Molecular Biology and Evolution, 10(2)

ISSN
0737-4038

Authors
Karl, SA
Avise, JC

Publication Date
1993-03-01

DOI
10.1093/oxfordjournals.molbev.a040002

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, availalbe at
https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

PCR-based Assays of Mendelian Polymorphisms from Anonymous Single-Copy Nuclear DNA: Techniques and Applications for Population Genetics[1]

Stephen A. Karl[2] *and John C. Avise*
Department of Genetics, University of Georgia

This paper outlines a PCR-based approach for population genetics that offers several advantages over conventional Southern blotting methods for revealing restriction-fragment-length polymorphisms (RFLPs) in nuclear DNA. Primers are constructed from clones isolated from a nuclear DNA library, and these primers subsequently are employed in in vitro syntheses of homologous regions. Amplified products are then screened directly for RFLPs by using gel-staining procedures. Population applications for this PCR-based approach, including potential strengths and weaknesses, are exemplified by two RFLP data sets generated to estimate (*a*) male-mediated gene flow in the green turtle (*Chelonia mydas*) and (*b*) geographic population genetic structure in the American oyster (*Crassostrea virginica*). Restriction assays of amplified products from 14 or 15 independent primer pairs in each species revealed polymorphisms at several loci that proved highly informative in the population genetic analyses. In general, the Mendelian polymorphisms produced by this PCR-based approach will provide useful genetic markers for population studies, particularly in situations where simpler and less expensive allozyme methods have failed, for whatever reason, to provide adequate information.

Introduction

Estimates of genetic parameters in natural populations often require molecular polymorphisms (preferably neutral) from multiple independent loci. Following the introduction of allozyme methods to population genetics (Harris 1966; Lewontin and Hubby 1966), thousands of studies have utilized protein polymorphisms to address such issues as population variability, genetic parentage, gene flow, and species phylogeny. Nonetheless, concerns regarding (*a*) possible effects of non-neutrality on some allozyme variants and (*b*) the degree of bias associated with the choice of loci known to code for important metabolic proteins have prompted searches for additional classes of molecular markers from numerous unlinked genes.

In recent years, variation at the level of DNA has been assayed more directly as restriction-fragment-length polymorphisms (RFLPs) in Southern blots, using cloned single-copy nuclear (scn) DNAs as probes (Southern 1975). These approaches alleviate some of the concerns associated with allozyme studies: (*a*) many of the nucleotide changes underlying scnRFLPs probably reflect silent substitutions or other genetic

1. Key words: polymerase chain reaction, anonymous nuclear loci, restriction-fragment-length polymorphism, *Chelonia mydas, Crassostrea virginica,* gene flow.

2. Present address: Center for Theoretical and Applied Genetics, Cook College, Rutgers University, P.O. Box 231, New Brunswick, New Jersey 08903-0231.

Address for correspondence and reprints: Stephen A. Karl, Center for Theoretical and Applied Genetics, Cook College, Rutgers University, P.O. Box 231, New Brunswick, New Jersey 08903-0231.

Mol. Biol. Evol. 10(2):342–361. 1993.
© 1993 by The University of Chicago. All rights reserved.
0737-4038/93/1002-0007$02.00

changes likely to be selectively neutral; and (*b*) use of a large (in principle, nearly unlimited) number of scnDNA probes should provide a broader representation of the nuclear genome. Southern-blotting approaches suffer, however, from the considerable effort needed to produce and isolate clones that reveal informative scnRFLPs. There are also expenses and difficulties incurred by the requirement of microgram quantities of nuclear (n) DNA, large amounts of radioactive nucleotide, and long exposure times for development of autoradiographs. These drawbacks have limited severely the number of studies that have successfully employed scnDNA markers in a population context (Quinn and White 1987).

Here we describe and evaluate an approach, based on the polymerase chain reaction (PCR) (Saiki et al. 1985, 1988; Mullis and Faloona 1987) for screening scnRFLPs that in many ways is superior to similar analyses by Southern blotting. Although the approach does not at present match the speed and cost-effectiveness of allozyme assays, it may prove extremely useful in situations where protein-electrophoretic methods have failed (for whatever reason) to resolve population questions. We will first outline the laboratory methods. Then, on the basis of our empirical experience with these scnDNA techniques in marine turtles (Karl et al. 1992) and oysters (Karl et al. 1992), we will provide an overview of applications as well as limitations of this PCR-based approach for uncovering Mendelian genotypes at multiple unlinked loci.

Material and Methods

Initial steps in the PCR-RFLP protocol are similar to those of Southern blotting. An nDNA library is constructed, and clones are isolated. Beyond this point, PCR-RFLP analyses depart from Southern-blotting procedures by (*a*) eliminating the need for large quantities of genomic DNA from each individual, (*b*) screening for restriction-site variation directly in PCR-amplified DNA homologous to the cloned insert, and (*c*) visualizing RFLPs by ethidium bromide (EtBr) staining of the agarose gels. Once suitable PCR primers have been developed, hundreds of specimens can be screened with only moderate effort.

Primer Construction

Genomic DNA libraries can be constructed for each species by using nDNA extracted from fresh or frozen tissue. In our laboratory, the DNA is banded on an EtBr/CsCl equilibrium density gradient by a procedure modified from that of Herrmann and Frischauf (1987). The nDNA band is recovered from the gradient, and CsCl and EtBr are extracted. Nuclear DNA is digested with a restriction endonuclease that empirically has been determined to produce a majority of fragments in the 500–5,000-bp range. The entire digestion product is then ligated directly into a phagemid vector.

Recombinant phagemids are recovered by transforming *Escherichia coli* cells and plating them under selective conditions. Each isolated recombinant is labeled pdXX (p = plasmid; d = cloning of DNA; and XX = two-letter genus and species code) and numbered. Putative recombinants are screened for insert size by digestion of mini-DNA preparations (Maniatis et al. 1982, pp. 368–369), and genomic copy number is determined by dot blot analysis using labeled total cell DNA as probe. Since the amount of radioactive label incorporated into a particular probe sequence is proportional to its abundance in the genome, the signal intensity of each clone dot provides an indication of genomic copy number (fig. 1). Clones are scored tentatively

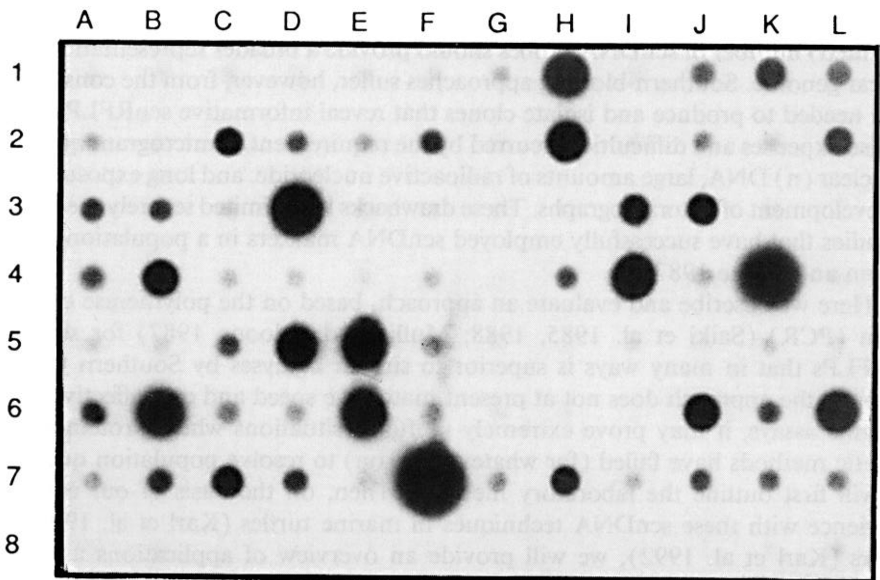


FIG. 1.—Dot blot of the green turtle nDNA library. DNA was isolated from each of several clones, bound to a nylon filter, and hybridized overnight with radioactively labeled nDNA. The blot was washed under stringent conditions and exposed to X-ray film for 24 h. The dot in sector 1A is a negative control containing plasmid DNA only. Positive controls for hybridization efficiency consisting of dot sectors containing 150 ng of total cell DNA (data not shown) were also included in the hybridization. Examples of tentative scoring for copy number: clone in dot 1B, single copy; 1H, high repetitive; 1K, moderate repetitive; and IJ, low repetitive.

as single-copy and low-, moderate-, or high-repetitive DNA (corresponding to absent, weak, moderate, or strong autoradiogram signals, respectively). Single- and low-copy clones with inserts of 500–2,000 bp then are chosen, and sequences of the first 100–150 nucleotides from both ends are determined by the dideoxy chain-termination method (Sanger et al. 1977) using the Sequenase T_7 sequencing kit (U.S. Biochemicals).

PCR primers are selected from these sequences by using a computer program (such as NAR; Rychlik and Rhoads 1989). Appropriate primers are ~20–25 nucleotides long, have little or no secondary structure, and lack regions of significant inter-primer complementarity. These sequences also are analyzed for putative open reading frames (by using the MAP routine in the GCG sequence-analysis software package; Devereux et al. 1984). If an open reading frame is found, the 3′ end of the primer is situated at the second nucleotide position of a putative codon.

Screening for Polymorphisms

DNA may be isolated from the population samples by a procedure modified from that of Herrmann and Frischauf (1987). A small sample of tissue is ground in 50 mM Tris-HCl pH 8.0, 100 mM NaCl, and 100 mM EDTA at room temperature. The homogenate is lysed by the addition of sodium dodecyl sulfate, and the lysate is extracted with phenol and chloroform. In some species (such as oysters), excess RNA can interfere with the subsequent PCR amplifications. In these cases, lysates are incubated for 3 h at 37°C in the presence of 10 μg RNase A, after which the samples again are treated with phenol and chloroform. In either case, DNA is precipitated by

the addition of sodium acetate and ethanol. All DNAs can be stored indefinitely in TE buffer (10 mM Tris-HCl pH 8.0 and 1 mM EDTA) at 4°C.

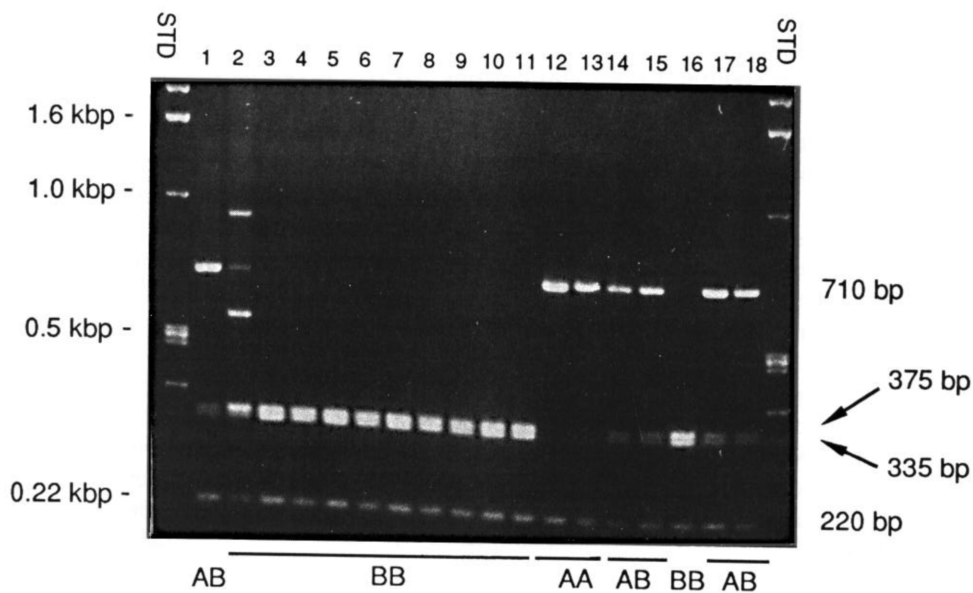
To identify polymorphic restriction sites, DNAs from population samples are amplified with the anonymous primers and *Taq* DNA polymerase under the manufacturer's (e.g., Promega) standard reaction conditions, with the addition of $MgCl_2$ and bovine serum albumin to final concentrations of 2.5 mM and 0.1 μg/μl, respectively. Exact thermal cycling parameters are determined empirically and are varied, depending on the primers used. For our samples, conditions generally were 35–40 cycles of 1 min at 94°C, 1 min annealing, and 1 min at 72°C. The only condition commonly changed was the annealing temperature, from 55°C to 65°C. Cycling times and the concentrations of $MgCl_2$, primer, template, and dNTPs can also be varied in attempts to optimize the amplifications. However, in our experience the most important factors are $MgCl_2$ concentration and the addition of bovine serum albumin.

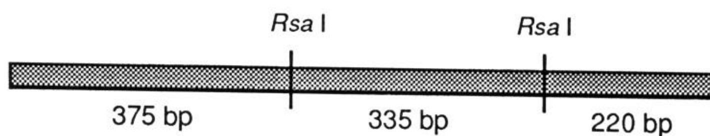
To check the amount and fidelity of amplification, we electrophoresed an aliquot of each undigested sample, through a 2.0% agarose gel stained with EtBr. Successfully amplified DNA then was used directly in restriction digestions, without further purification. Restriction-fragment profiles for several enzymes were generated, and polymorphic sites were deduced by the patterns of fragment gains and losses (fig. 2A). Because the locus amplified is bounded by primers and therefore is of defined size, site polymorphisms can be interpreted readily (fig. 2B and C). In addition, profile changes due to insertions or deletions can be identified either by size differences in the undigested amplified DNA or by consistent fragment size changes across digestions with different enzymes (fig. 3).

Genotype Assignment

When restriction-site polymorphisms are considered individually (or when primer regions contain a single polymorphic restriction site), haplotype designations at a locus are evident directly from the diploid genotypes ("++," "+−," and "−−," where + and − indicate restriction-site presence and absence, respectively). However, for loci with two or more polymorphic enzyme sites, the *cis* versus *trans* phase of multiple heterozygotes is not apparent immediately. Furthermore, the phase of double heterozygotes cannot be deduced from gel profiles alone when the variant sites are separated by an invariant site(s). In this case, both variable sites are flanked by nonvariable sites or the primers, so no unique fragments are produced that permit haplotype phases to be distinguished (fig. 4; Quinn and White 1987).

Nonetheless, this problem may be circumvented partially by additional linkage considerations. From the diploid genotypes of all individuals except multiple heterozygotes, the phase of each haplotype can be determined directly (e.g., a ++/+− diploid must consist of one ++ and one +− haplotype). For example, suppose that in a two-site polymorphic system the scoring of all such unambiguous haplotypes reveals only two of the four possible haplotype classes (e.g., "+−" and "−+"). This would indicate that the presence versus absence states are in strong or complete disequilibrium, most likely because of limited recombination among physically linked markers. Therefore, the double heterozygotes likely represent a combination of the "+−" and "−+" haplotypes and provisionally can be scored accordingly. If the polymorphic sites are adjacent (i.e., not separated by an invariant site), then appropriate digests of DNA from double heterozygotes further can confirm these haplotype assignments, because the enzymes produce different restriction patterns for the *cis* versus *trans* configurations of sites (fig. 4).


A
STD
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
STD
1.6 kbp -
1.0 kbp -
0.5 kbp -
0.22 kbp -
710 bp
375 bp
335 bp
220 bp
AB
BB
AA
AB
BB
AB


B
Rsa I
Rsa I
375 bp
335 bp
220 bp

C
Genotype:
AA
AB
BB
710 bp
375 bp
335 bp
220 bp

Results

We successfully have employed this PCR-based approach to analyze both male-mediated gene flow in the green turtle (Karl et al. 1992) and geographic population genetic structure in the American oyster (Karl and Avise 1992). Biological aspects of these studies have been published elsewhere and will be reiterated here only briefly. Rather, we will focus on previously unpublished aspects of these genetic surveys, to illustrate methodological nuances that to a considerable degree determine strengths and weaknesses of the PCR methodology for generating scnRFLPs.

Anonymous Loci

Green Turtle (Chelonia mydas) *Survey*

Ninety-seven genomic clones from a *Hin*dIII nDNA library were screened for insert size, and the inserts showed an average ± SD length of 1,479 ± 657 bp. Seventy of these clones were assayed for copy number by dot blot hybridization: 29 (41%) appeared to be single-copy, 17 (24%) low-repetitive, 13 (19%) moderate-repetitive, and 11 (16%) high-repetitive genomic elements. Independence of the clones was not assessed. From 15 mostly single- or low-copy clones (table 1), DNA was isolated by CsCl gradient purification (several "quick prep" methods for the production of plasmid DNA suitable for DNA sequencing proved unreliable), and nucleotide sequences from both ends of the inserts were determined. Sequence data then were employed to synthesize primers (by Oligos Etc., Gilford, Conn.). The sequence of these primer pairs is provided in Appendix A.

Table 1 summarizes some of the results for these 15 primer sets. Three primer pairs (involving one single- and two low-copy loci) failed completely to amplify nDNA; five (involving four low- and one moderate-repetitive loci) amplified nDNA but produced several different-size amplification products (both larger and smaller fragments generally persisted even under more stringent amplification conditions); and the remaining seven primer pairs produced a single amplification product of the expected length. The latter subsequently were screened for polymorphic restriction sites in samples from various green turtle populations.

American Oyster (Crassostrea virginica) *Survey*

Initially, 29 clones from a *Hin*dIII nDNA library were screened for insert size, and the inserts showed an average ± SD length of 1,755 ± 675 bp. All clones were assayed for copy number by dot blot hybridization: 10 (35%) appeared to be single-copy, 7 (24%) low-repetitive, 9 (31%) moderate-repetitive, and 3 (10%) high-repetitive genomic elements. Observed frequencies of these categories thus are quite similar to those for the green turtle. Independence of clones was not assessed. In this first phase of analysis, 11 clones (5 single-copy and 3 low-, 2 moderate-, and 1 high-repetitive)

FIG. 2.—Example of RFLPs from single-copy nDNA loci in green turtles. Primers were constructed and nDNA was amplified by using procedures described in the text. A, Agarose gel showing 18 individuals screened at the CM-14 locus by *Rsa*I. Diploid genotypes are indicated below each lane (the extra bands in lane 2 are partial digestion products). B, Relative map positions of the variable (*left*) and nonvariable (*right*) *Rsa*I sites, as determined by partial restriction digests. C, Restriction-fragment profiles predicted for individuals homozygous for the absence (i.e., −/−) or presence (i.e., +/+) of the variable site, and of heterozygotes (i.e., −/+).

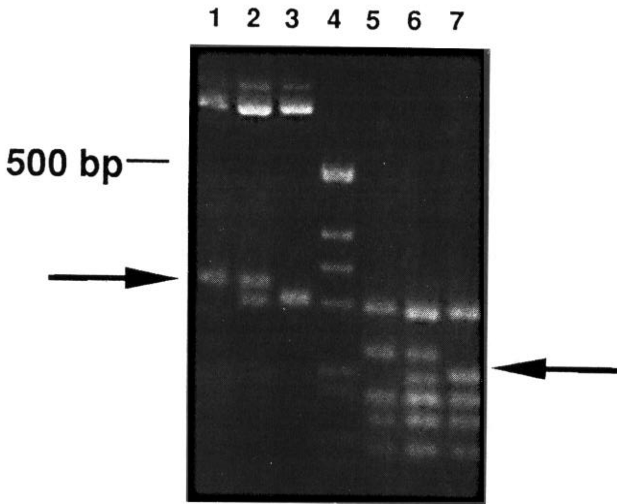


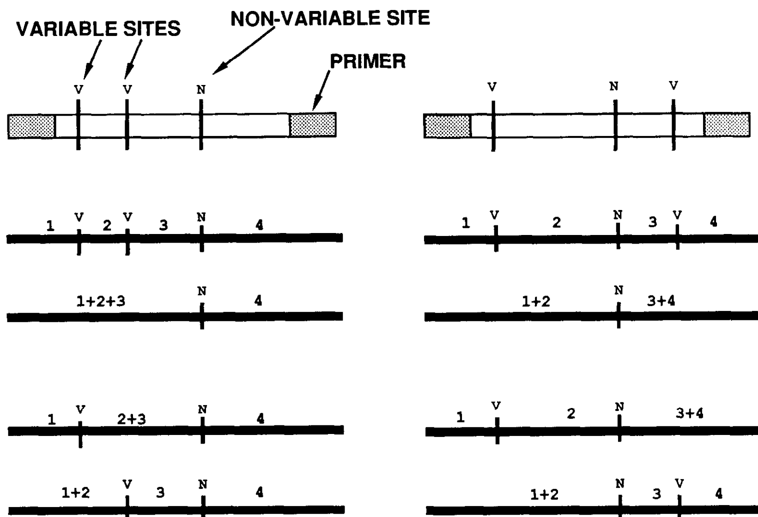
FIG. 3.—Restriction-fragment patterns for PCR-amplified DNA (locus CV-32) from American oysters, digested with *Sau*3A and *Taq*I. All individuals are homozygous for both the *Sau*3A and the *Taq*I recognition sites. Lane 4, 1-kb size standard. Lanes 1 and 5, Restriction-fragment patterns of individuals homozygous for a large (1,000-bp) allele. Lanes 3 and 7, Homozygotes for a small (965-bp) allele. Lanes 2 and 6, Heterozygotes for the size polymorphism. This 35-bp size difference was observed with several different restriction enzymes. Variable size fragments are indicated by the arrows.

were sequenced, and primers were synthesized as above (experiment 1, table 2). The sequence of these primer pairs is provided in Appendix B. One of these primer pairs failed to amplify genomic DNA, and another did so only unreliably. Three primer pairs (two single-copy and one moderate-repetitive clone) produced more than a single size amplification product that persisted even under stringent conditions.

The remaining six primer pairs appeared to amplify a single size product. However, restriction digests for two of these (CV-05 and CV-20) resulted in fragment profiles apparently too complex to be attributable to a single, diploid locus (fig. 5). These complex patterns where highly variable between individuals but were consistent under various amplification conditions within an individual. For two of the loci whose primer pairs proved unsuitable for PCR-RFLP analysis (CV-03, which gave multiple amplification products, and CV-05, which gave a complex fragment profile), new primer pairs were made by searching the sequence data for alternative priming sites. Neither of these new primer sets (CV-03.2 and CV-05.2) alleviated the problems of its predecessors. A second-generation primer was also made from another oyster clone, but for different reasons. One of the initial primer pairs (CV-32) was capable of amplifying only certain individuals. After one of the primers (chosen arbitrarily) was repositioned at a location 14 nucleotides 5′ of the old one, DNA from all individuals amplified. The exact reason for this change was not determined.

In an attempt to circumvent some of the above problems, we constructed a second nDNA library from oysters (experiment 2, table 2). Oyster nDNA again was digested to completion, this time by using *Nde*II. This enzyme produced a large number of fragments of length <1,000 bp. The digested DNA was ligated directly into a phagemid vector cut with *Bam*HI and then was used to transform *Escherichia coli.* One hundred eleven clones were surveyed, as before, for insert size, and the inserts showed an average ± SD length of 789 ± 296 bp. Ninety-two of these clones were

Adjacent Variable Sites | **Non-Adjacent Variable Sites**



Cis Heterozygote

Trans Heterozygote

Fragments Produced

Cis: 1, 2*, 3, 4, 4, 1+2+3* | 1, 2, 3, 4, 1+2, 3+4

Trans: 1, 3, 1+2*, 2+3*, 4, 4 | 1, 2, 3, 4, 1+2, 3+4

FIG. 4.—Diagram illustrating *cis* and *trans* configuration for adjacent and nonadjacent polymorphic enzyme sites. Restriction fragments produced after restriction endonuclease digestion are numbered. There are fragments diagnostic for the *cis* or *trans* configuration produced if the variable sites are separated by nonvariable site(s).

screened for genomic copy number. Clones that appeared on first examination to be single-copy were rescreened, and 13 of these proved repetitive, resulting in a final total of 43 (47%) single-copy, 26 (28%) low-repetitive, 18 (20%) moderate-repetitive, and 5 (5%) high-repetitive clones. Three primer pairs were synthesized, as before, from single-copy clones. The sequence of these primer pairs is provided in Appendix B. One of these primer pairs produced multiple amplification products, whereas the other two (CV-195 and CV-233) produced only a single fragment (table 2).

Genetic Variation

Green Turtle

Forty restriction endonucleases were employed to assay the seven green turtle scn loci that had been amplified successfully by the PCR, revealing a total of 206 restriction sites (table 3). However, because recognition sequences for some enzymes are subsets of those for others [e.g., the sequence for *Taq*I (TCGA) is contained within that for *Sal*I (GTCGAC)], the number of distinct sites reduced to 166, representing a total of 758 bp (or 9.0% of the entire 8,415 bp of sequence amplified). Table 4

Table 1
Summary of All Primers Constructed from Green Turtle nDNA Clones

Locus	Fragment Size (bp)	Genomic Copy No.	Product[a]	No. of Polymorphic Enzymes
CM-01 ...	1,380	Single	S	0
CM-04 ...	3,500	Moderate	M	...
CM-07 ...	1,600	Single	NO	...
CM-08 ...	1,050	Low	NO	...
CM-12 ...	1,195	Single	S	2
CM-14 ...	930	Low	S	2
CM-23 ...	1,250	Low	M	...
CM-28 ...	1,400	Single	S	0
CM-35 ...	1,500	Low	NO	...
CM-39 ...	1,350	Single	S	1
CM-40 ...	1,600	Low	M	...
CM-45 ...	1,000	Single	S	1
CM-46 ...	1,000	Low	M	...
CM-67 ...	1,160	Low	S	3
CM-70 ...	1,400	Low	M	...

[a] S = single amplification product of expected size; M = multiple amplification products; and NO = no amplification.

summarizes this information, as well as additional data on polymorphisms observed across the seven scn loci. Two loci (CM-01 and CM-28) contained no site polymorphisms. Other loci exhibited either one (CM-39 and CM-45), two (CM-12 and CM-14), or three (CM-67) polymorphic sites, but no single endonuclease produced more than one polymorphism within a locus. In addition, two loci (CM-39 and CM-45) contained size polymorphisms (data not shown). Overall, ~1.2% of base pairs proved polymorphic (table 4).

The five polymorphic loci formed the basis for the published estimates of interpopulational gene flow and other population genetic parameters from the green turtle data (Karl et al. 1992). Most loci from most populations were in agreement with Hardy-Weinberg expectations (only 3 of 75 locus × population comparisons deviated significantly). Furthermore, strong linkage disequilibrium among polymorphic sites within a locus allowed assignments of multisite haplotype phase (as described above).

American Oyster

Forty restriction endonucleases were employed to assay the six oyster scn loci that had been amplified successfully by the PCR (table 5), and these revealed a total of 107 restriction sites (89 different sites, after correction for overlaps in recognition sequence, representing 7.1% of the entire 5,955 bp of sequence amplified). Table 6 summarizes this information, as well as additional data on polymorphisms observed across the six scn loci. Two loci contained no (CV-36) or very rare (CV-233) site polymorphisms. Other loci exhibited one to four polymorphic sites, and one enzyme (*Dde*I) produced two polymorphic sites within a locus (CV-195). In addition, size variation was observed at all four polymorphic loci (data not shown). Overall, ~2.6% of base pairs proved polymorphic (table 6), a value approximately double that for green turtles (1.2%).

One site polymorphism chosen from each of the four strongly variable scn loci provided the basis for the published estimates of geographic population structure of

Table 2
Summary of All Primers Constructed from American Oyster nDNA Clones

Locus	Fragment Size (bp)	Genomic Copy No.	Product[a]	No. of Enzymes Polymorphic[b]
Experiment 1 (*Hin*dIII library):				
CV-01	2,200	Single	M	...
CV-03	1,600	Single	M	...
CV-05	1,600	Single	S	C
CV-06	1,000	Low	NO	...
CV-07	1,500	Low	S	2
CV-15	4,000	Moderate	U	...
CV-19	1,500	Single	S	4
CV-20	1,600	Low	S	C
CV-32	1,000	Single	S	1
CV-35	1,100	Moderate	M	...
CV-36	1,050	High	S	0
Experiment 2 (*Nde*II library):				
CV-195	770	Single	S	2
CV-233	830	Single	S	1
CV-244	1,400	Single	M	...

[a] M = multiple amplification products; S = single amplification product of expected size; NO = no amplification; and U = unreliable amplification.
[b] C = complex restriction profile.

the American oyster (Karl and Avise 1992). In contrast to the green turtle data, several scn loci in oyster populations showed significant deficits of heterozygotes, relative to Hardy-Weinberg expectations (unpublished data), a phenomenon noted previously with regard to allozyme genotypes in oysters and other bivalves (Singh and Green 1984; Singh and Zouros 1987). Whether these deficiencies are due to biological factors or technical difficulties is currently uncertain. Because only one site polymorphism was scored per locus in the population surveys, no conclusions can be reached regarding the degree of linkage disequilibrium within the assayed oyster DNA sequences.

Discussion

Strengths and Limitations

The PCR-based approach for generating scnRFLPs shares with Southern blotting procedures the necessity for construction and screening of a nuclear library for suitable clones. It has a significant advantage over Southern blotting procedures in that, once polymorphisms have been identified, large numbers of individuals can be assayed quickly and with minimal expense. Thus, much smaller starting quantities of nDNA are required (extending the number of enzymes that can be used to assay an individual), and the time required from amplification to gel scoring is shortened considerably. Furthermore, because the regions amplified are bounded by primers, the distinction between DNA length polymorphisms and restriction-site polymorphisms within a locus usually is apparent immediately. This contrasts with Southern blotting, where probes reveal regions of varying length, depending on positions of the flanking restriction sites (and additional mapping procedures may be required to distinguish site heterogeneity from length heterogeneity).

In both the turtle and oyster, library construction was straightforward and pro-

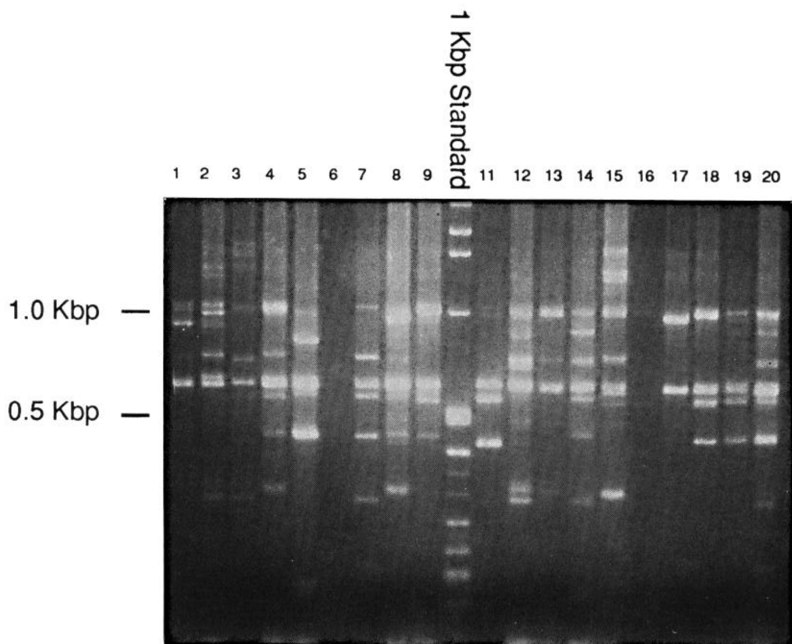


FIG. 5.—Complex restriction profile from the American oyster (locus CV-20). DNA was amplified and then digested with *Hin*dII. The DNA in lane 11 is the amplified cloned insert from clone pdCV-20, also digested with *Hin*dII. Because of its clonal origin, the latter represents a single allele from the population.

duced ample usable clones. Although we confined attention to single-copy or low-repetitive clones (39%–47% of the total), high-repetitive clones also could have been employed. Presumably, repetitive sequences are flanked by single-copy sequences from which locus-specific primers could be designed. Some of these loci might show variation similar to micro- or minisatellite sequences described by other researchers (Jeffreys et al. 1988, 1991; Horn et al. 1989).

Several enzymes failed to recognize sites at all screened loci (tables 3 and 5), whereas others (e.g., *Hin*dII, *Hin*fI, and *Rsa*I) consistently cut and commonly were polymorphic. This information can be useful in designing an efficient screening strategy—the screen should commence with enzyme sites likely to be polymorphic and should exclude those that typically fail to find recognition sequences.

One potential complication in the scoring of RFLP genotypes arises if mutational differences exist, at a primer site, among alleles in the population (especially near the 3′ end); then it is possible that not all alleles will amplify successfully under PCR. Such nonamplified alleles could produce errors in genotype assignment (such as the scoring of some true heterozygotes as homozygotes). If nonamplified ("null") alleles were common in the population, they should occur also as homozygotes and be evidenced as such on gels. However, DNA from some individuals might also fail to amplify, for reasons (such as poor-quality DNA as substrate) unrelated to mutations in the primer regions.

Why some primer pairs completely failed to amplify nDNA is unknown. Perhaps the sequences of the primers were incorrectly determined or had unfavorable properties, such as low GC content, cryptic primer intercomplementarity, or otherwise unstable

Table 3
Numbers of Restriction Sites for the 40 Restriction Enzymes Used to Screen Seven scnDNA Loci in Green Turtles

	No. of Restriction Sites for Locus[a]						
Enzyme[b]	CM-01 (1,380)	CM-12 (1,195)	CM-14 (930)	CM-28 (1,400)	CM-39 (1,350)[c]	CM-45 (1,000)[c]	CM-67 (1,160)
*Alu*I	1	0/1	2	4	5	3/4	4/5
*Ava*I	0	1	0	0	0	0	0
*Ava*II	2	3	1	1	1	0	0
*Bam*HI	0	1	0	0	0	0	1
*Bcl*I	0	0	0	0	1	0	0
*Bgl*II	1	1	0	0	2	0	0
*Bst*NI	6	3	1	1	0	0	1
*Cfo*I	1	0	1	2	1	0	0/1
*Cla*I	1	0	0	0	0	0	0
*Dde*I	3	5	4	7	2	1	2/3
*Dra*I	2	2	1	1	2	0	0
*Dra*II	1	2	2	0	0	0	0
*Eco*RI	1	0	0	1	1	1	0
*Fok*I	3	2	1	1	0	2	2
*Hae*III	2	1	2	2	0	0	3
*Hin*dII	0	2/3	0	1	1	0	0
*Hin*fI	4	3	1/2	2	2/3	1/2	3
*Kpn*I	0	0	0	0	0	1	0
*Msp*I	0	1	0	1	0	0	0
*Mvn*I	...[d]	0	0	0	2	0	0
*Nci*I	0	1	0	1	0	0	0
*Nde*I	0	0	0	2	0	0	1
*Nde*II	2	3	1	4	4	3	4
*Nsi*I	0	1	0	0	0	1	1
*Pst*I	0	2	1	1	0	0	0
*Pvu*II	0	0	1	2	0	1	1
*Rsa*I	2	1	1/2	2	1	2	0
*Sal*I	0	0	0	1	0	0	0
*Sma*I	0	1	0	0	0	0	0
*Spe*I	0	0	0	1	0	0	0
*Stu*I	1	0	0	0	0	0	0
*Taq*I	...[d]	0	1	1	0	0	0
*Xba*I	1	0	0	2	0	0	0
Total	34	38	23	41	26	18	26

[a] Nos. in parentheses are the sizes (in bp) of each amplified locus.
[b] *Bgl*I, *Bst*EII, *Eco*RV, *Ksp*I, *Pvu*I, *Sac*I, and *Xho*I failed to identify recognition sequences.
[c] Size is polymorphic.
[d] Not assayed.

binding properties. In the oysters, simply moving the primer pair for locus CV-32 did indeed solve the problem of nonamplification. However, other second-generation primers (CV-03.2 and CV-05.2) failed to correct the multiple amplification products, suggesting that the specific sequences of these primers were not at fault. Rather, the larger region (i.e., the locus) from which these primers were made may have been present as multiple copies in the genome. In general, some of the expense of producing

Table 4
Summary of Preliminary Screen for RFLPs at Seven scnDNA Loci in Green Turtles

Primer	Total No. of Sites Observed	Total No. of Base Pairs Surveyed[a]	% of Sequence Surveyed	% of Base Pairs Polymorphic[b]	No. of Sites Polymorphic
CM-01	26	115	8.3%	0%	0
CM-12	29	140	11.7%	1.4%	2
CM-14	19	79	8.5%	2.5%	2
CM-28	33	157	11.2%	0%	0
CM-39	23	106	7.9%	0.9%	1
CM-45	14	64	6.4%	1.6%	1
CM-67	22	97	8.4%	3.1%	3
Overall ...	166	758	9.0%[c]	1.4%[c]	9

[a] Calculated as the number of observed restriction sites for an enzyme times the number of base pairs in the recognition sequence (Nei 1987, pp. 101–102), summed across all enzymes.

[b] Calculated as the number of polymorphic sites divided by the total number of base pairs surveyed, with the assumption that each polymorphic site represents a single nucleotide difference.

[c] Average across all primer pairs.

failed primers might be recovered by using them in other PCR-based procedures, such as random amplified polymorphic DNA (RAPD) analysis (Williams et al. 1991) or gene walking (Parker et al. 1991).

Other PCR-based techniques also are available and have been employed to generate population genetic data. For example, several studies using RAPDs have provided information on strain identification and population structure, as well as on genomic mapping (review in Hadrys et al. 1992). Since the primers employed in RAPD assays are short, random sequences, much less effort is required for their production, and large numbers of primers can be screened readily. However, unlike the case for RFLP data, RAPD alleles typically behave in a dominant/recessive fashion (presence of a band indicates successful amplification, and absence indicates unsuccessful amplification). Thus an amplification from even one of a pair of homologous chromosomes from a diploid individual produces a DNA fragment, and it is not possible, without additional effort, to distinguish all homozygote from heterozygote genotypic classes at each RAPD locus. Furthermore, because RAPD alleles are identified solely by size, there is the potential for incorrect scoring due to comigration of nonhomologous fragments.

For generation of Mendelian polymorphisms, another available PCR-based approach that is of potential use to population genetics involves assay of particular mini- or microsatellite VNTR (variable number of tandem repeat) loci via use of primers from conserved single-copy flanking regions (Jeffreys et al. 1988; Horn et al. 1989). Such approaches have been employed, for example, in genetic analyses of human populations (Balazs et al. 1992). One potential complication in this approach involves proper characterization of the large number of alleles (often with similar DNA fragment mobilities) sometimes present at a locus (e.g., see Budowle et al. 1991). The overall magnitude of effort and expense involved in these techniques probably is comparable to that required for the scnDNA methods described in the current study.

Table 5
Numbers of Restriction Sites for the 40 Restriction Enzymes Used to Screen Six scnDNA Loci in American Oysters

	No. of Restriction Sites for Locus[a]					
Enzyme[b]	CV-07 (1,390)[c]	CV-19 (1,440)[c]	CV-32 (1,000)[c]	CV-36 (930)	CV-195 (620)[c]	CV-233 (575)
*Alu*I	2	0	7	1	...[d]	0
*Bcl*I	0	1	0	0	...[e]	...[e]
*Bgl*II	0	0	0	1	...[e]	...[e]
*Bst*NI	0	1/2	0	0	0	0
*Cfo*I	0	0	2	0	...[d]	0
*Cla*I	2	1	1	0	...[d]	0
*Dde*I	1	4	1	1	0/1/2	1
*Dra*I	2	3	1	3	1	0/1
*Dra*II	0	1	0	1	1	0
*Eco*RI	0	0	0	0	0	1
*Eco*RV	0	0	0	1	0	0
*Hae*III	0	1	1	1	0	1
*Hind*II	0	2	0	0	...[d]	0
*Hin*fI	2/3	1	2	0	...[d]	0
*Kpn*I	1	0	0	0	...[d]	0
*Msp*I	1	0	1	2	...[d]	0
*Mvn*I	...[d]	...[d]	2	...[d]	...[d]	0
*Nci*I	0	0	1	0	...[d]	0
*Nde*II	...[d]	3	1	...[d]	...[e]	...[e]
*Nsi*I	0	4	0/1	0	...[d]	1
*Pst*I	1	0	0	0	...[d]	0
*Pvu*II	0	0	4	0	...[d]	0
*Rsa*I	2/3	1/2	1	2	1/2	2
*Spe*I	0	1	0	0	...[d]	0
*Stu*I	0	0/1	0	0	...[d]	0
*Taq*I	2	2	4	2	...[d]	0
*Xba*I	1	0/1	0	0	...[d]	0
Total	19	30	30	15	6	7

[a] Nos. in parentheses are the sizes (in bp) of each amplified locus.
[b] *Ava*I, *Ava*II, *Bam*HI, *Bgl*I, *Bst*EII, *Hind*III, *Ksp*I, *Nde*I, *Pvu*I, *Sac*I, *Sal*I, *Sma*I, and *Xho*I failed to identify recognition sequences (many of these were not surveyed in CV-195).
[c] Size is polymorphic.
[d] Not assayed.
[e] Site cut by enzyme used to make library.

Nature of the Data

In some respects, RFLP data from scnDNA loci are intermediate to those produced by protein electrophoresis versus RFLP analyses of mtDNA. Allozyme alleles represent unordered multistate characters revealed at each of several unlinked loci, whereas RFLPs from mtDNA reflect single-locus haplotypes whose phylogenetic order may be inferred from the patterns of linked site changes. Barring intragenic recombination or gene conversion, numerous RFLPs from each scnDNA locus could, in theory, provide the same type of gene-genealogical information as that for the 17-kb mtDNA molecule. However, both the relatively small size of the PCR-amplified loci (typically ~1 kb in our assays) and the presumed slower rate of evolution in scnDNA (relative to mtDNA) mean that few linked polymorphisms will normally be found

Table 6
Summary of Preliminary Screen for RFLPs at Six scnDNA Loci in American Oysters

Primer	Total No. of Sites Observed	Total No. of Base Pairs Surveyed[a]	% of Sequence Surveyed	% of Base Pairs Polymorphic[b]	No. of Sites Polymorphic
CM-07	16	78	5.6%	2.5%	2
CV-19	24	121	8.4%	3.3%	4
CV-32	24	110	11.0%	0.9%	1
CV-36	13	60	6.5%	0%	0
CV-195	5	22	3.5%	13.6%	3
CV-233	7	34	5.9%	2.9%	1
Overall ...	89	425	6.8%[c]	3.9%[c]	11

NOTE.—Nine individuals (18 haplotypes) from three geographically remote regions were surveyed with up to 40 enzymes, to detect polymorphic restriction sites.

[a] Calculated as the number of observed restriction sites for an enzyme times the number of base pairs in the recognition sequence (Nei 1987, pp. 101–102), summed across all enzymes.

[b] Calculated as the number of polymorphic sites divided by the total number of base pairs surveyed, with the assumption that each polymorphic site represents a single nucleotide difference.

[c] Average across all primer pairs.

(although direct sequencing of the amplified products might increase this number). In our experience with green turtles and oysters, we never observed more than four polymorphisms within a locus (despite use of up to 40 restriction enzymes). Furthermore, assumptions of complete linkage disequilibrium occasionally may be violated, resulting in problems of inferring haplotypes from RFLP patterns in diploid individuals. For all of these reasons, opportunities for gene-genealogical assessment normally may be limited.

Applications

In comparison with conventional allozyme methods, the elucidation of Mendelian polymorphisms by the PCR-based approach is more expensive and labor intensive. For the turtle and oyster surveys described above, many months of effort devoted to library construction and screening, primer generation, and restriction with numerous enzymes produced totals of only five and four strong restriction-site polymorphisms that proved useful in the population genetic surveys for the two respective species. Under what circumstances, then, might the PCR-based method be advocated as an appropriate methodology for population genetics? At least two general classes of situation can be exemplified by the published surveys on *Chelonia mydas* and *Crassostrea virginica*.

A. Insufficient Allozyme Variation

In some species, conventional allozyme surveys fail to reveal adequate numbers of polymorphic markers for reliable estimation of population genetic parameters. For example, in an electrophoretic survey of 23 allozyme loci in several green turtle rookeries from around the world, only two low-frequency polymorphisms were reported, and both occurred at a single locale (Bonhomme et al. 1987). The authors attributed the lack of observable differentiation among rookeries to extensive gene exchange, but this conclusion should be viewed with caution, because of the paucity of variation suitable for estimating gene-flow parameters. Our global survey of *Chelonia mydas*,

based on the five strongly polymorphic scn loci, revealed significant population genetic structure: standardized interlocality allele frequency variances (F_{ST} values) among loci were 0.05–0.21, producing estimates of interrookery gene flow (*Nm* values) in the range of ~1.0–4.8 (Karl et al. 1992).

B. Apparent Conflicts of Allozyme Data with Other Evidence

In some situations, population genetic parameters inferred from allozyme data may appear to clash with those derived from other lines of information, and additional genetic markers are required to resolve the dilemma. A case in point involves geographic population structure observed in *Crassostrea virginica.* An allozyme survey of 11 polymorphic loci revealed remarkable uniformity in allele frequencies through most of the species' range (Buroker 1983). This result contrasts strikingly with the nearly fixed differences in mitochondrial (mt) DNA haplotype frequency, which can distinguish most Atlantic populations from Gulf of Mexico populations (Reeb and Avise 1990). Our survey of American oysters, based on four strongly polymorphic scn loci, revealed a nuclear genetic "break" roughly coincident, in geographic position, with that registered by mtDNA (Karl and Avise 1992). If it is assumed that the Atlantic-Gulf genetic dichotomy reflects historical population separations coupled with severe contemporary restrictions on gene flow, then the allozyme data for oysters are positively misleading indicators of population structure. One likely possibility is that balancing selection on protein electrophoretic characters in oysters stabilizes allozyme frequencies, even in the face of severe constraints to interregional genetic exchange.

In summary, the PCR-based approach to revealing scnRFLPs can add another powerful weapon to the molecular evolutionists' arsenal. Although the approach as presently conducted cannot challenge the ease and low cost of standard protein electrophoresis for revealing Mendelian polymorphisms, in certain circumstances, such as those exemplified by the studies of the marine turtles and oysters, the scnDNA approach may be both useful and necessary.

Acknowledgments

We would like to thank Chris and Jane Bowen for expert field assistance. This work was supported by an NIH predoctoral training grant to S.A.K. and by NSF grant BSR-9005940.

APPENDIX A

Sequences of the turtle primers synthesized in this study. All sequences are 5′ to 3′ and are arbitrarily designated as right (R) and left (L).

CM-01:
R: GGAAGCTTTC AAGCTACACT GT
L: GGAAGCTTCA TAACACAAAA CC

CM-04:
R: GAGAAGGCCA GTGTATCAAG AG
L: CTCTCATAGG TTCTCAAGCT G

CM-07:
R: AAGCTTGGAT CCTAAATCTTTGTCTG
L: AAGCTTCTGA AGTTGTTGGC

CM-08:
R: AAATCCAAAG GAGAAGAAAG
L: ATCAAGAATA CTGCGGGAGA

CM-12:
R: AGCTGAAGCC AATGAAGAAG AA
L: GCTCAGGTTT AGCTCGAAGG T

CM-14:
R: TAAGCATTAT ACGTCACGGA
L: AGTATTTGGG CAGAACAGAA

CM-23:
R: ATAAATCTGT TTGCTTGTCA CG
L: ATCCAACCTA AACCTCCCAC A

CM-28:
R: TAAATGCCAG GTATGTAACTC
L: GATTGCTGGT CTCTGGAAGG CT

CM-35:
R: TTTGCGAATA CAGACTAACA CG
L: CTGCTCCGAA AACAAACGCT

CM-39:
R: TGCTAGTTTT GTTAGTTCTG GT
L: ATAGTGGATT GGAGAAGTTG TT

CM-40:
R: AATCCAAGCC TGAGCCTAAA
L: ACAAGCGTAT TTTCCTGGCA

CM-45:
R: CTGAAAGTGT TGTTGAATCC AT
L: CCGCAAGCAA AACATTCTCT

CM-46:
R: AAGCACGATA AAGGTTCTGG
L: CTCATAAGAA AGCCCTCCCA

CM-67:
R: GAATATAAGA TTTCATACCC CA
L: TTTAATTCTG AAAACTGCTC TT

CM-70:
R: ATGTACAGGG CCTAGCACAA
L: GTGGCTAGTT GTTCAAAACC AT

APPENDIX B

Sequences of the oyster primers synthesized in this study. All sequences are 5′ to 3′ and are arbitrarily designated as right (R) and left (L).

CV-01:
R: GAAGCATCAG GAAAGGACCG
L: GCGGGATTTC TGTTCGTTTA

CV-03:
R: AGGGTGATTG ATGCTTTTCT
L: TTTATGTTAC GAGGCAATGT CA

CV-03.2:
R: CACTGATCTC TCTTGTTGCC TTCCT
L: AGGCTGAAAC CCAGATTGTG CATCT

CV-05:
R: AAGCTTTTGT TTGTCTGTGG
L: GAATTTGCAT GTAATTTTCA C

CV-05.2:
R: GGGGAATATT TCAGGAATGT GTGGTT
L: TGGTGCTTAT TCAAAACGAA TTTCA

CV-06:
R: AATTTCACCT AAGAATGATT TA
L: TAGCAGATTT CAGAGAGCCT

CV-07:
R: AAGCTTTAGC CTTCAACTCA GACAA
L: AAGCTTTAAG GTAGAAGCAA ATTA

CV-15:
R: GGAAGCTTTG CTATAGGTTA GT
L: GGAAGCTTGT TTTGTAGAAT CGGG

CV-20:
R: GGAAGCTTCT CGGAGACAAG CT
L: GGAAGCTTGT CTCTTCCCTC TG

CV-32:
R: GGAAGCTTCA GTCTTTAGAT TATACATG
L: GGAAGCTTAC AAAACAAGCT CGGCTA

CV-32.2:
R: GGAAGCTTTA TTATCTAACA GTCT

CV-35:
R: AGCTCGAAAA CGGTAAGAGC TGGAC
L: CTTCCGGTTT TGTCACTTCC GGTTC

CV-36:
R: GCTTCCGGGT TCGTCATTTC CGGTC
L: ATCAGTCTAT TCATGTGTTC TGCTC

CV-195:
R: GGATCAGAAG GAAAGCAACA GCAC
L: AACGTTTGAT GGAACAAGGG AAACT

CV-244:
R: CATATTGGCG GTTTCATATT CAAAG
L: TTATCTATGG ACAAAGCGTG ACCCT

LITERATURE CITED

BALAZS, I., J. NEUWEILER, P. GUNN, J. KIDD, K. K. KIDD, J. KUHL, and L. MINGJUN. 1992. Human population genetic studies using hypervariable loci. I. Analyses of Assamese, Australian, Cambodian, Caucasian, Chinese and Melanesian populations. Genetics **131**:191–198.

BONHOMME, F., S. SALVIDIO, A. LEBEAU, and G. PASTEUR. 1987. Comparaison génétique des tortues vertes (*Chelonia mydas*) des Oceans Atlantique, Indien et Pacifique: une illustration

apparente de la théorie mullerienne classique de la structure génétique des populations? Genetica **74**:89–94.

BUDOWLE, B., A. M. GIUSTI, J. S. WAYE, F. S. BAECHTEL, R. M. FOURNEY, D. E. ADAMS, L. A. PRESLEY, H. A. DEADMAN, and K. L. MONSON. 1991. Fixed-bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci, for use in forensic comparisons. Am. J. Hum. Genet. **48**:841–855.

BUROKER, N. E. 1983. Population genetics of the American oyster *Crassostrea virginica* along the Atlantic coast and the Gulf of Mexico. Mar. Biol. **75**:99–112.

DEVEREUX, J., P. HAEBERLI, and O. SMITHIES. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.

HADRYS, H., M. BALICK, and B. SCHIERWATER. 1992. Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. Mol. Ecol. **1**:55–63.

HARRIS, H. 1966. Enzyme polymorphisms in man. Proc. R. Soc. Lond. [Biol.] **164**:298–310.

HERRMANN, B., and A. M. FRISCHAUF. 1987. Isolation of genomic DNA. Methods Enzymol. **152**:180–183.

HORN, G. T., B. RICHARDS, and K. W. KLINGER. 1989. Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. Nucleic Acids Res. **17**:2140.

JEFFREYS, A. J., A. MACLEOD, K. TAMAKI, D. L. NEIL, and D. G. MONCKTON. 1991. Minisatellite repeat coding as a digital approach to DNA typing. Nature **354**:204–209.

JEFFREYS, A., V. WILSON, R. NEUMANN, and J. KEYTE. 1988. Amplification of human minisatellites by the polymerase chain reaction: towards DNA typing of single cells. Nucleic Acids Res. **16**:10953–10971.

KARL, S. A., and J. C. AVISE. 1992. Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. Science **256**:100–102.

KARL, S. A., B. W. BOWEN, and J. C. AVISE. 1992. Global population genetic structure and male-mediated gene flow in the green turtle (*Chelonia mydas*): RFLP analyses of anonymous nuclear loci. Genetics **131**:163–173.

LEWONTIN, R. C., and J. L. HUBBY. 1966. A molecular approach to the study of genetic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura.* Genetics **54**:595–609.

MANIATIS, T., E. F. FRITSCH, and J. SAMBROOK. 1982. Molecular cloning: a laboratory manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MULLIS, K. B., and F. A. FALOONA. 1987. Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. Methods Enzymol. **155**:335–350.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

PARKER, J. D., P. S. RABINOVITCH, and G. C. BURMER. 1991. Targeted gene walking polymerase chain reaction. Nucleic Acids Res. **19**:3055–3060.

QUINN, T. W., and B. N. WHITE. 1987. Identification of restriction-fragment-length polymorphisms in genomic DNA of the lesser snow goose (*Anser caerulescens caerulescens*). Mol. Biol. Evol. **4**:126–143.

REEB, C. A., and J. C. AVISE. 1990. A genetic discontinuity in a continuously distributed species: mitochondrial DNA in the American oyster, *Crassostrea virginica.* Genetics **124:** 397–407.

RYCHLIK, W., and R. E. RHOADS. 1989. A computer program for choosing optimal oligonucleotides for filter hybridizations, sequencing and *in vitro* amplification of DNA. Nucleic Acids Res. **17**:8543–8551.

SAIKI, R. K., D. H. GELFAND, S. STOFFEL, S. J. SCHARF, R. HIGUCHI, G. T. HORN, K. B. MULLIS, and H. A. ERLICH. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**:487–491.

SAIKI, R. K., S. SCHARF, F. FALOONA, K. B. MULLIS, G. T. HORU, H. A. ERLICH, and N. ARNHEIM. 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230**:1350–1354.

SANGER, F., S. NICKLEN, and A. R. COULSON. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

SINGH, S. M., and R. H. GREEN. 1984. Excess of allozyme homozygosity in marine molluscs and its possible biological significance. Malacologia **25**:569–581.

SINGH, S. M., and E. ZOUROS. 1987. Genetic variation associated with growth rate in the American oyster (*Crassostrea virginica*). Evolution **32**:342–353.

SOUTHERN, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.

WILLIAMS, J. G. K., A. R. KUBELLK, K. J. LIVAK, J. A. RAFALSKI, and S. V. TINGEY. 1991. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. **18**:6531–6535.

MASATOSHI NEI, reviewing editor

Received June 11, 1992; revision received September 15, 1992

Accepted September 28, 1992