Population Structure in the American Oyster as Inferred by Nuclear Gene Genealogies

Matthew P. Hare[1] *and John C. Avise*

Department of Genetics, University of Georgia

Multiple haplotypes from each of three nuclear loci were isolated and sequenced from geographic populations of the American oyster, *Crassostrea virginica.* In tests of alternative phylogeographic hypotheses for this species, nuclear gene genealogies constructed for these haplotypes were compared to one another, to a mitochondrial gene tree, and to patterns of allele frequency variation in nuclear restriction site polymorphisms (RFLPs) and allozymes. Oyster populations from the Atlantic versus the Gulf of Mexico are not reciprocally monophyletic in any of the nuclear gene trees, despite considerable genetic variation and despite large allele frequency differences previously reported in several other genetic assays. If these populations were separated vicariantly in the past, either insufficient time has elapsed for neutral lineage sorting to have achieved monophyly at most nuclear loci, or balancing selection may have inhibited lineage extinction, or secondary gene flow may have moved haplotypes between regions. These and other possibilities are examined in light of available genetic evidence, and it is concluded that no simple explanation can account for the great variety of population genetic patterns across loci displayed by American oysters. Regardless of the source of this heterogeneity, this study provides an empirical demonstration that different sequences of DNA within the same organismal pedigree can have quite different phylogeographic histories.

Introduction

Phylogeography is a discipline concerned with the demographic, behavioral, and historical factors influencing the geographic distributions of gene lineages (Avise et al. 1987). In theory, historical dispersal and demography consort to shape the phylogeographic architectures of neutral loci within and among closely related species. Thus, phylogenetic methods applied to gene lineages and their geographic distributions provide an opportunity to examine evolutionary processes underlying contemporary population genetic patterns (Crandall and Templeton 1993; Templeton and Georgiadis 1996). By focusing on relationships of alleles as well as allele frequencies, phylogeographic approaches have helped to bridge the formerly distinct fields of systematic biology and population genetics (Avise 1989).

Because of its rapid sequence evolution and maternal, nonrecombining mode of inheritance in higher animals, mitochondrial (mt) DNA provided both the original stimulus for phylogeographic perspectives and the vast majority of empirical applications in the field (Avise 1997). These features have made mtDNA, in principle as well as in practice, arguably the most informative molecule available for phylogeographic studies (Wilson et al. 1985; Avise 1995; Moore 1995). Nonetheless, there are dangers in relying on a single locus for evolutionary inferences about populations or species. The matrilineal pathway of descent is but a minuscule fraction of the total hereditary history of a population (Maddison 1995; Avise and Wollenberg 1997). Whether an mtDNA genealogy is representative of the myriad genealogies for nuclear loci depends on such factors as the gender-specific historical demography of the assayed species, the effective population size of mtDNA relative to nuclear (n)DNA, and the nature of selection operating on nuclear and cytoplasmic genes.

Genealogical perspectives have made explicit the distinction between gene trees and population trees or species trees (Neigel and Avise 1986; Pamilo and Nei 1988; Takahata 1989; Ball, Neigel, and Avise 1990; Wu 1991) and have prompted development of genealogical concordance principles as a conceptual guide to relate significant partitions in gene trees to significant historical partitions at the level of organismal phylogeny (Avise and Ball 1990; Baum and Shaw 1995; Avise and Wollenberg 1997). In principle, phylogenetic subdivisions in gene trees securely register phylogenetic subdivisions in organismal phylogenies only when various of the following distinct categories of genealogical concordance can be demonstrated (Avise 1996): (1) phylogeographic concordance across multiple variable characters within a surveyed gene, (2) concordance of significant genealogical partitions across multiple independent (unlinked and nonepistatic) genes within a species, (3) concordance in the geographic position of significant gene tree partitions across multiple codistributed species, and (4) concordance of gene tree partitions with traditional (nonmolecular) biogeographic evidence.

Here, we explicitly address category 2 of genealogical concordance by examining gene genealogies for each of three nuclear loci in populations of the American oyster, *Crassostrea virginica.* We take advantage of recently developed methods of single-stranded conformation polymorphism analysis (PCR-SSCP; Orita et al. 1989; Orú, Hare, and Avise 1997) to isolate and then sequence individual haplotypes from diploid oyster tissues. Results are compared to a previous report of mtDNA phylogeography for this species (Reeb and Avise 1990). Except for phylogeographic studies on humans (e.g., Hammer et al. 1997 and references therein) and a few other model organisms such as *Drosophila*

[1] Present address: Department of Organismic and Evolutionary Biology, Harvard University.

Key words: phylogeography, gene flow, lineage sorting, gene tree, ancestral polymorphism, polyphyly, oyster.

Address for correspondence and reprints: Matthew P. Hare, Department of Organismic and Evolutionary Biology, Harvard University Biolabs, 16 Divinity Avenue, Cambridge, Massachusetts 02138. E-mail: mhare@oeb.harvard.edu.

Mol. Biol. Evol. 15(2):119–128. 1998

(e.g., DeSalle and Giddings 1986), few reports have attempted comparisons of nDNA genealogies with mtDNA genealogies within and among closely related taxa (Bernardi, Sordino, and Powers 1993; Burton and Lee 1994), and even fewer published studies have compared multiple nuclear gene trees at the intraspecific level.

Background on the American Oyster

Crassostrea virginica is a broadcast spawning, eurythermal, hermaphroditic bivalve that inhabits estuarine waters from New Brunswick, Canada, to the Yucatan Peninsula, Mexico (Galtsoff 1964). Prior reports based on mtDNA RFLPs (Reeb and Avise 1990) and RFLP allele frequencies at four anonymous nuclear loci (Karl and Avise 1992) indicated a pronounced genetic distinction between most oyster populations along the Atlantic coast versus those in the Gulf of Mexico. The zone of genetic transition between the two forms occurs along the east coast of Florida and localizes to the Cape Canaveral area, where allele frequency shifts of 50%–75% occur over a distance of 20 km (Hare, Karl, and Avise 1996). Yet, mating within local populations in the step-cline region appears to be nearly random, indicating an absence of pronounced intrinsic reproductive barriers between Gulf and Atlantic oysters (Hare and Avise 1996).

At least three categories of genealogical concordance mentioned above (categories 1, 3, and 4) suggest that the Atlantic and Gulf populations may have had a relatively deep history of vicariant population separation: multiple RFLP variants distinguish the two populations in the mtDNA gene tree; conspecific populations of several other coastal species similarly assayed show pronounced genetic differences between the Atlantic and Gulf coasts (reviews in Avise 1992, 1996); and the boundary between the Atlantic and Gulf forms often localizes to the east coast of Florida in a region long recognized by marine biogeographers as a demarcation zone (evidenced by concentrations of species' distributional limits) between the temperate Carolinian and the subtropical Caribbean zoogeographic provinces (Briggs 1974). The evidence for genetic differentiation between Atlantic and Gulf oyster populations also includes dramatic allele frequency differences at several anonymous nuclear genes (Karl and Avise 1992; Hare and Avise 1996) and at one surveyed allozyme locus (*GOT-2*; W. W. Anderson, personal communication).

On the other hand, a lack of separation for the Atlantic and Gulf populations of oysters is suggested by an approximate uniformity of allele frequencies at most allozyme (Buroker 1983) and some nDNA loci (McDonald, Verrelli, and Geyer 1996) throughout relevant portions of the range of the species; by the retention of reproductive compatibility between the Atlantic and Gulf populations (Hare and Avise 1996); and, in general, by the potential for long-distance dispersal and high gene flow conferred by the planktonic larvae of this otherwise sessile bivalve (a factor originally invoked to explain the allele frequency uniformity in the allozyme assays; Buroker 1983).

Allele frequency differences at several anonymous nuclear genes (Karl and Avise 1992; Hare and Avise 1996) are in general agreement with the mtDNA phylogeographic pattern and suggest that among-locus genealogical concordance (category 2) also might prove to characterize the Atlantic and Gulf oyster populations. However, these nuclear data consist solely of allele frequencies from RFLP assays and include no information on genealogical relationships of the alleles. Here, we analyze nucleotide sequences at each of three nuclear loci to infer genealogical patterns of allelic descent. For one of these loci in particular, the phylogenetic analyses should help to decide among three competing hypotheses previously advanced (Karl and Avise 1992) to account for the nonfixed (but substantial) differences in nuclear RFLP allele frequencies between Atlantic and Gulf oysters: (1) historical gene flow in secondary contact zones that may have transferred alleles between regional populations formerly separated, (2) evolutionary convergence (homoplasy) of genealogically distinct Atlantic and Gulf alleles to identical RFLP phenotypes, and (3) incomplete lineage sorting of allelic polymorphisms that predate any Atlantic/Gulf population separation.

Materials and Methods

Samples

Adult oysters were collected in 1989 from Massachusetts (MA), South Carolina (SC), Panacea in west Florida (PN), and Louisiana (LA) (Karl and Avise 1992). Haplotypes examined in this study were taken preferentially from the distant geographic sites (MA and LA) and exclusively from heterozygotes carrying alleles that could be electrophoretically separated (see below). One of the loci (CV-32) was examined previously at the RFLP level (Karl and Avise 1992; Hare and Avise 1996) and showed large allele frequency differences (as gauged by *Nsi* I digests) between Atlantic and Gulf oyster populations. To take advantage of existing data on the CV-32 polymorphism and to test hypothesized convergence in RFLP patterns, *Nsi* I-present and *Nsi* I-absent alleles were examined in equal numbers from *Nsi* I heterozygotes at each of the four localities, regardless of the *Nsi* I population frequency. In contrast, amplification products from two newly examined loci (CV-23 and CV-myc) were isolated and sequenced from electrophoretic heterozygotes chosen at random.

For two of the loci, sequences from a closely related (Littlewood 1994) species, *C. rhizophorae*, were generated and employed to root the respective gene trees. However, our assays failed to amplify a usable CV-23 sequence from *C. rhizophorae*, and hence the CV-23 gene tree for *C. virginica* is presented as unrooted.

Laboratory Techniques

Nuclear loci were examined by PCR amplification from genomic DNA, isolation of haplotypes from heterozygotes by SSCP (single-strand conformational polymorphism), reamplification, and direct sequencing. Apparent SSCP homozygotes were not analyzed further. In

Table 1
Locus Attributes Including PCR Product Size, Length of Analyzed DNA Sequence, PCR Primers Used, and GenBank Accession Number for *Crassostrea virginica* and *C. rhizophorae* Sequences

Locus	PCR Product (bp)	Sequence Length (bp)		Primer Sequences	GenBank Accession No.
CV-23.....	966[a]	534	23.0L,	5′-ACAATCAAAGACCACCGTCGCTAATT-3′	*C.v.*—AF025657
			23.1R,	5′-TTGTATCTTCA CATTGCGTCCACTT-3′	
			23.2Rs,	5′-AAAATCATTGGGGTATCTT-3′	
CV-myc	559	508	mycL,	5′-CTGTCGGGGAGTGTATGAGCCTTGG-3′	*C.v.*—AF024523
			mycR,	5′-TGGCTTGTGGGTTGTTGGTGTTCAT-3′	*C.r.*—AF029312
			myc2L,	5′-TCATTAGCTTCCTGACTAGAATTTC-3′	
CV-32.....	446	386	32.3Ls,	5′-CTGTACTTGTTTTCCCGTT-3′	*C.v.*—AF025658
			32.7R,	5′-CATTCATTTTCCTCCAAGTTTTGTT-3′	*C.r.*—AF025659

[a] Variable size due to VNTR length variation.

general, PCR included 6–12 pmol of each primer per reaction, 200 μM dNTPs, 2–4 mM $MgCl_2$, 0.1 μg/μl bovine serum albumin, and 0.5 μl genomic DNA in 25- or 50-μl reaction volumes. PCR typically included 30 cycles at 94°C for 30 sec, 55°C for 30 sec, and 72°C for 60 sec.

Table 1 presents the PCR primers and the lengths of sequences amplified for each of the three nuclear loci. The nucleotide sequences from which the primers were derived are deposited in GenBank.

The CV-32 locus was isolated as an anonymous genomic clone and has been studied with respect to a *Nsi* I RFLP polymorphism (Karl and Avise 1992, 1993; Hare, Karl, and Avise 1996). For this study, DNA sequence surrounding the *Nsi* I site was collected using CV-32 primer sequences and PCR conditions described in Orţ, Hare, and Avise (1997).

The CV-23 locus was isolated from a size-selected (1–3 kb) genomic library by screening with multiplexed tetranucleotide probes ($(GACA)_4$, $(GATA)_4$, $(GGAT)_4$). Primers CV-23.0L and CV-23.1R (table 1) were designed to amplify a complex VNTR motif plus 0.5 kb of flanking sequence. After the isolation of haplotypes (see below), primers CV-23.0L and CV-23.2Rs were used for the nested reamplification and sequencing of non-VNTR DNA.

For the CV-myc locus, primers were designed from the published cDNA sequence of a putative *C. virginica* homolog of deuterostome c-*myc* (Marsh and Chen 1995). In vertebrates, c-*myc* is a proto-oncogene whose protein product appears to function in the regulation of developmental differentiation and apoptosis. Two intron/exon boundaries were proposed for the *C. virginica* gene by Marsh and Chen (1995) based on sequence similarities with human c-*myc*. Primers CV-mycL and CV-mycR (table 1) were designed to amplify 500 bp of coding sequence containing the putative exon 2–3 boundary. The resulting 1,500-bp PCR product from *C. virginica* genomic DNA had the expected coding sequence adjacent to unique (presumably intron) sequence. Primers CV-mycR and CV-myc2L (table 1) were subsequently used to amplify and sequence primarily intron DNA.

Methods for SSCP separation, isolation, and sequencing of haplotypes are detailed elsewhere (Orţ, Hare, and Avise 1997) with specific reference to CV-32. For CV-myc, allelic separation of PCR products by SSCP was best in 8% polyacrylamide gels run at 13°C for 11 h at a constant 8 W, or 9% gels at 3°C for 17–20 h at 10 W. Because the CV-23.0L and CV-23.1R primers flank a VNTR, allelic separations were achieved with heterozygous PCR products run slowly in 1.5% agarose. From each CV-23 heterozygote, EtBr-stained haplotype bands were stabbed with 20-gauge needles and rinsed in 50 μl H_20. This tempate DNA (1 μl) was used in nested reamplifications to generate sequencing template.

For sequencing, 50-μl reamplification reactions were column-purified (Wizard, Promega), eluted with H_20, and cycle-sequenced either manually (fmol, Promega) or with an automated ABI sequencer. Both DNA strands were sequenced using the original PCR primers and internal sequencing primers.

Data Analyses

Sequences were aligned by eye to minimize gaps and nucleotide substitutions. Measures of genetic diversity and recombination were calculated with SITES (Hey and Wakeley 1997). Parsimony (PAUP 3.1; Swofford 1991) was used to reconstruct phylogenetic relationships among alleles and to calculate sequence differences. Maximum-likelihood phylogeny reconstructions were done using fastDNAml 1.0 (Olsen et al. 1994).

Results

Haplotypes ranging in size from ca. 380 to 530 bp were successfully amplified from the three target loci and, after electrophoretic separation from heterozygotes, reamplified and sequenced directly (table 1). Haplotype isolation from heterozygotes was facilitated by heterozygosities ≥81% in nonsequencing assays of all three loci. Sequence variation suggests that all three loci are predominantly noncoding DNA. For CV-32 and CV-23, this evidence includes frameshift indels and multiple stop codons in all three reading frames, and these sequences showed no greater than 70% similarity to any sequences in BLAST searches of existing databases. These features also were characteristic of the CV-myc sequences, except for 81 bp at the 3′ end and 36 bp near the 5′ end, where near-identity with published coding sequence was found (Marsh and Chen 1995).

Table 2
Measures of Polymorphism and Selection for DNA Sequences at Three Nuclear Loci in Gulf and Atlantic Populations of American Oysters

	Base Pairs Noncoding	*N*	*k*	*s*	No. of indels	Haplotype Diversity	Maximum Sequence Difference	4Nc per bp[a]	π per bp[a] (×100)	Ts/Tv
CV-23	534	15	14	48	9	0.96	0.038	0.52	1.99	1.19
Atlantic		11	10	11		0.93	0.036		1.79	
Gulf		4	4	4		0.86	0.028		2.53	
CV-myc	391	18	12	10	5	0.93	0.012	0.07	0.59	1.13
Atlantic		11	7	6		0.87	0.010		0.35	
Gulf		7	5	5		0.83	0.010		0.53	
CV-32	386	48	21	32	3	0.92	0.026	0.10	0.96	1.59
Atlantic		24	12	21		0.83	0.023		0.81	
Gulf		24	9	13		0.86	0.023		0.96	
mtDNA	ca. 390[b]	212	82	NA	NA	0.85	0.03	NA	NA	NA
Atlantic		103	31	NA	NA	0.57	0.01	NA	NA	NA
Gulf		109	51	NA	NA	0.80	0.01	NA	NA	NA

NOTE.—*N* = number of haplotypes examined; *k* = number of different alleles observed; *s* = number of variable nucleotide sites; 4Nc = recombination parameter (Hudson 1987); π = mean pairwise sequence divergence; Ts/Tv = average transition/transversion ratio from all pairwise comparisons.

[a] Values calculated for noncoding sequence only.

[b] mtDNA assays involved restriction fragment comparisons (Reeb and Avise 1990), so this value is the approximate number of nucleotides in the recognition sites of the enzymes employed, and several of the statistical parameters are not applicable (NA).

Sequence Variation

Information on genetic variation at each locus is summarized in table 2. From 12 to 21 different alleles per locus were detected in our samples, involving from 10 to 48 variable nucleotide positions. Three to nine indels per locus also were observed.

The sampling of alleles from heterozygotes, performed here for the convenience of haplotype isolation, is expected to underestimate the population frequency of common alleles and overestimate the frequency of relatively rare alleles. The magnitude of the bias can be extreme if there is an allele at high frequency in the population (maximum allele frequency among heterozygotes is 0.5), but is much lower or absent at loci with an even dispersion of allele frequencies. Diversity measures such as haplotype diversity and nucleotide diversity (π) that are sensitive to allele frequency will therefore be biased upward by heterozygote sampling when population allele frequencies are uneven. This bias is partly responsible for haplotype diversity values that were ≥ 0.92 across the three nuclear loci (table 2). Nucleotide heterozygosity measured as the mean number of pairwise sequence differences (π, table 2), was highest at CV-23 and lowest at CV-myc.

The number of polymorphic sites within aligned sequences (*s*, table 2) is not sensitive to allele frequency, but an estimate based on sequences from heterozygotes will converge toward the true population value faster than one based on a random sample of the same size. Nucleotide diversity by this measure was greatest at CV-23 and lowest at CV-myc.

Phylogenetic Inference

Multiple equally short parsimony trees were found for each locus. Strict-consensus maximum-parsimony trees are presented for CV-23, CV-myc, and CV-32 in figures 1, 2, and 3, respectively. None of the three gene trees displayed a pattern of reciprocal monophyly between Atlantic and Gulf oyster populations. On the contrary, at each locus, the unrooted gene tree topology showed a polyphyletic relationship among Atlantic and Gulf alleles. Rooting of the CV-myc and CV-32 trees with an outgroup sequence had little effect on their topology. Polyphyletic topologies were also found using maximum-likelihood and neighbor-joining tree reconstruction methods (not shown).

The CV-23 data had the largest number of informative sites coupled with the highest level of homoplasy. Details of the CV-23 topology depended on the alignment of variable-length mononucleotide strings, but in every case, strict-consensus clades contained both Atlantic and Gulf alleles. Both the CV-myc and CV-32 trees had few informative sites and were poorly resolved. Nonetheless, for all three loci, the most parsimonious trees (MPTs) were shorter by several (3–6) steps than the shortest constrained trees in which the Atlantic and Gulf alleles were reciprocally monophyletic. To test whether these tree length differences are greater than those expected from random variation among the informative characters, a comparison was made between the character step changes in an MPT and the minimum changes required in a tree constrained to be reciprocally monophyletic (Templeton 1983). Two-tailed probability values for this nonparametric Wilcoxon signed-ranks test ranged from 0.07 to 0.11 for the three loci.

Discussion

As described in the *Introduction,* several previous lines of genetic evidence had suggested a history of vicariant differentiation and secondary contact for Atlantic and Gulf of Mexico populations of the American oyster (Avise 1996). However, in this study, we find for three

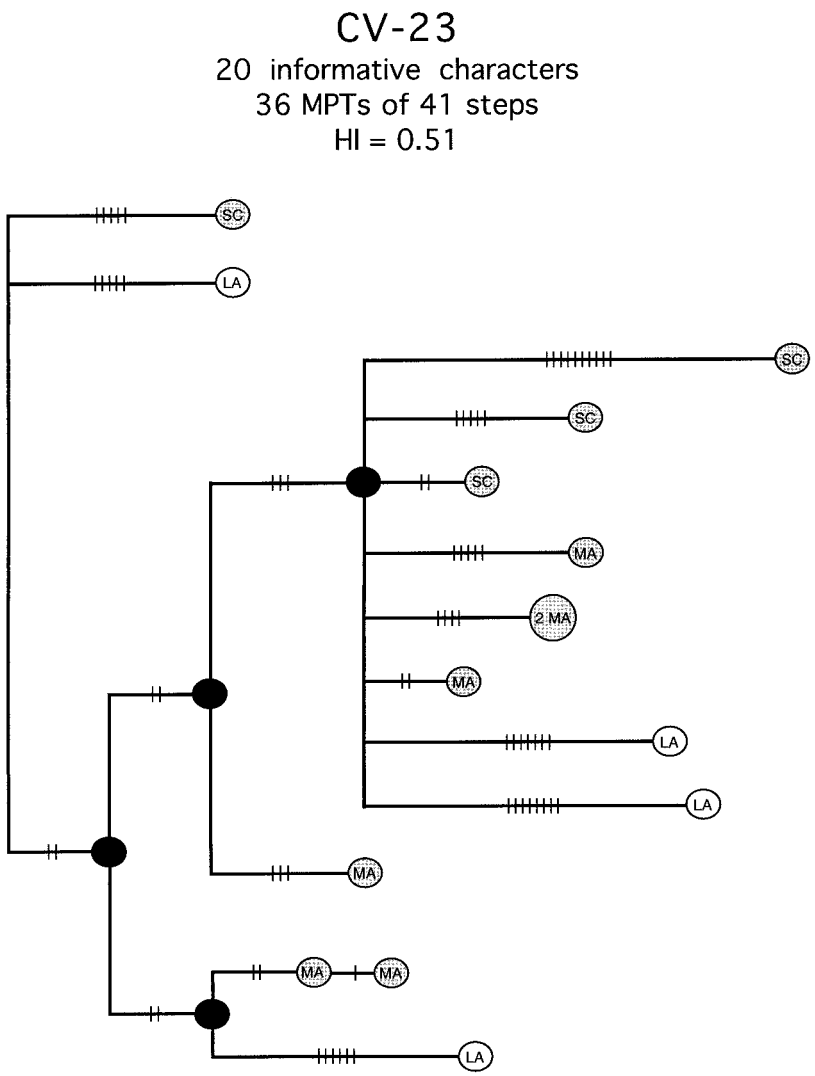

FIG. 1.—Unrooted maximum-parsimony tree for CV-23. The topology is a strict consensus of the shortest trees found using a branch-and-bound search with informative characters only. The number of informative sites, the number of shortest trees (MPTs), the tree length (steps), and the homoplasy index (HI) are shown. Alleles observed more than once are represented by larger circles. Circles are shaded for alleles that originated in the Atlantic; black circles at interior nodes represent unobserved hypothetical ancestors. Branch lengths are proportional to the approximate number of step changes required for all variable characters (also shown by dashes across each branch).

nuclear gene genealogies that oyster haplotypes are not aligned phylogenetically into recognizably distinct Atlantic and Gulf lineages. Allelic sampling error cannot explain these results, because at each locus, no conceivable haplotype could now be sampled that would convert the gene trees to reciprocal monophyly. This genealogical conclusion is also insensitive to biased allele frequencies resulting from the heterozygote sampling design. Thus, the current data in this case fail to confirm category 2 of genealogical concordance, wherein a substantial agreement with mtDNA might have been anticipated. Several explanations are possible for the poor concordance between the mtDNA and nuclear gene genealogies with regard to the Atlantic/Gulf dichotomy.

Vicariance and Stochastic Lineage Sorting

Monophyly, Paraphyly, and Polyphyly

When an ancestral population is sundered by geographic or other barriers to gene flow, a process of ''lineage sorting'' is initiated that proceeds rather inexorably toward an eventual status of reciprocal monophyly for gene trees that include both daughter populations (Neigel and Avise 1986; Pamilo and Nei 1988). Intermediate stages typically involve gene-tree polyphyly and then paraphyly. Under neutral models, genetic drift governs this process such that smaller effective population sizes (N_e) lead to higher rates of random lineage extinction and fixation (i.e., of lineage sorting) on average.

In a random-mating dioecious species, mtDNA is expected to have a fourfold lower N_e than autosomal nuclear loci (all else being equal), because it is effectively haploid and maternally transmitted (Birky, Fuerst, and Maruyama 1989). Under protandrous hermaphroditism characteristic of oysters, this difference in N_e reduces to about twofold, assuming that most individuals survive to breed as females. (Note: These are only general expectations. Additional demographic factors of potential importance to rates of lineage sorting include the

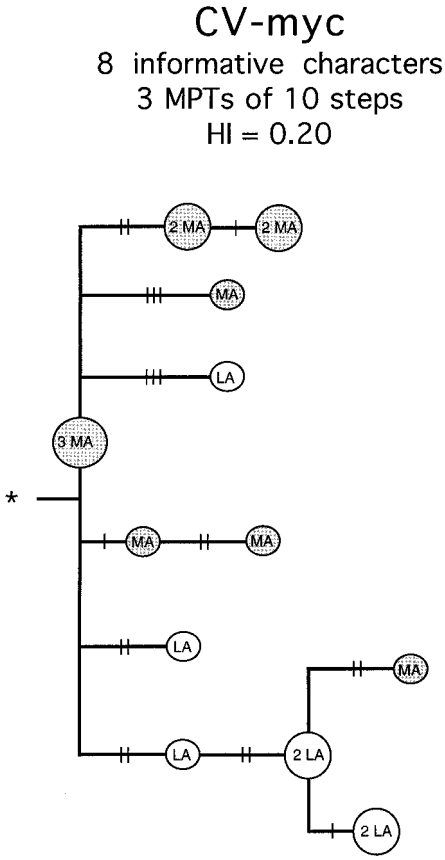

FIG. 2.—Maximum-parsimony tree for CV-myc (see legend to fig. 1). An asterisk marks the observed rooting of the tree with a single *C. rhizophorae* sequence as outgroup.

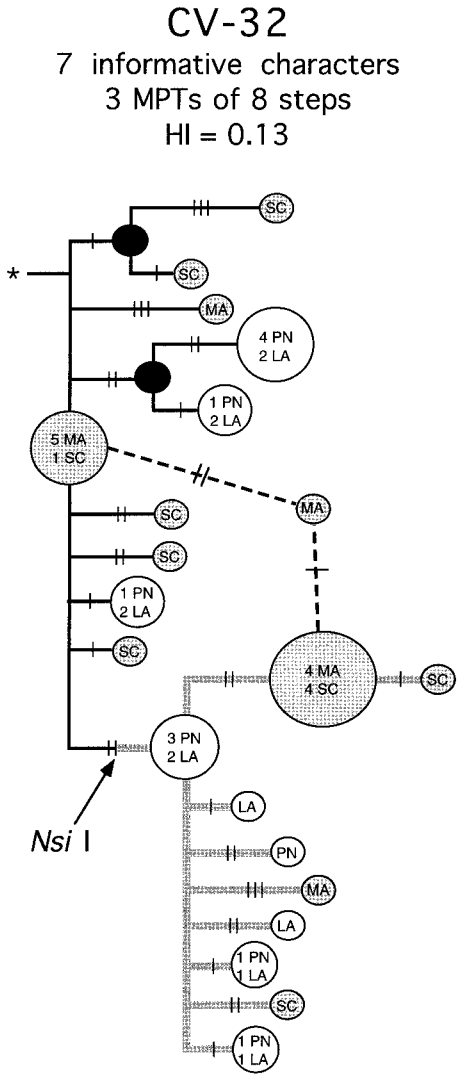

FIG. 3.—Maximum-parsimony tree for CV-32 (see legend to fig. 1). An asterisk marks the observed rooting of the tree with a single *C. rhizophorae* sequence as outgroup. Wide gray branches on the tree connect haplotypes with a *Nsi* I restriction site. Dashed lines connect a hypothesized recombinant haplotype (see text) to the tree.

nature of the mating system, relative variances in reproductive success between genders, and the magnitudes of population structure. Under some combinations of these parameters, N_e for mtDNA can actually exceed that for nuclear genes [Birky, Fuerst, and Maruyama 1989; Hoelzer 1997].) In any event, one possibility is that the mitochondrial alleles have had sufficient time to sort to reciprocal monophyly in vicariantly separated Atlantic and Gulf populations of oysters, whereas alleles at most nuclear loci have not.

However, expected coalescence times have a large stochastic variance (Nei 1987, pp. 394–396). This variance across loci in lineage-sorting outcomes can be expanded further by natural selection. Balancing selection at a locus can inhibit allelic extinction and thereby greatly extend the time to reciprocal monophyly (Takahata 1990). Conversely, differential directional selection in two daughter taxa can accelerate the progression to reciprocal monophyly (Stephan and Mitchell 1992). The target of selection need not be within the assayed segment of DNA: ''genetic hitchhiking'' can extend the effect of selection across mechanistically nonfunctional sequences within a larger nonrecombining region (Maynard Smith and Haigh 1974).

Alternatively, the difference in coalescence patterns between mitochondrial and nuclear loci in this study might be explained by directional selection and adaptive fixation of mtDNA variants in the Atlantic and the Gulf (e.g., Rand, Dorfsman, and Kann 1994). Two independent observations suggest that this is not the case. First, the prevalence of Atlantic–Gulf mtDNA differences in several unrelated maritime species (e.g., horseshoe crab, seaside sparrow, tiger beetle) suggests common elements of biogeographic history (and not merely contemporary selection factors alone) affecting coastal species in the region. Second, considerable mtDNA polymorphism within both Atlantic and Gulf oyster populations (Reeb and Avise 1990) is incompatible with recent selective sweeps (unless there is an exceptionally high mutation rate regenerating allelic diversity).

Intragenic Recombination

Another characteristic distinguishing nDNA from mtDNA is that meiotic recombination permits even physically linked nucleotides on a nuclear chromosome

to occasionally evolve quasi-independently along different genealogical paths through an organismal pedigree. Might intragenic recombination alone produce apparent poly- or paraphyly for nuclear haplotypes between vicariantly separated populations? After a status of reciprocal monophyly through lineage sorting has been achieved in a gene genealogy for two separate populations or taxa, the field for recombination is within each taxon and not between taxa (Baum and Shaw 1995). With respect to genealogical distinctions between such taxa, any empirical complications of intragenic recombination should then diminish with the further evolution of shared derived character states. Thus, intragenic recombination by itself should not be a major factor in regenerating the appearance of a poly- or paraphyletic condition in a gene genealogy for long-isolated populations.

At a practical level, intragenic recombination is merely a subset of processes that could produce homoplasy in a gene genealogy. For example, a pair of polymorphic sites (with two states each) in a sequence alignment can have all four possible haplotypes represented in the sample only if recombination has occurred between the two sites or there have been parallel or back mutations. Using this ''four-gamete test'' devised by Hudson and Kaplan (1985), the minimum numbers of recombination events were 6, 1, and 0 for CV-23, CV-32, and CV-myc, respectively. Hudson (1987) has defined a recombination parameter 4Nc that, when calculated for our data under assumptions of an infinite-sites model, similarly suggests that there is a five- to sevenfold difference in recombination rate across the three loci. The inferred recombinant allele at the CV-32 locus was also evident from direct visual inspection: the 5′ and 3′ ends of the haplotype were identical, respectively, to those of two separate and divergent haplotypes also present in the same Massachusetts collection. This allele was not included in phylogenetic analyses, but instead was added to the already-completed tree in figure 3 as a recombinant bridging the two haplotypes to which it presumably is allied. If intragenic recombination has truly occurred with moderate frequency at CV-23, then for that locus one of the basic assumptions underlying the phylogenetic analysis has, of course, been violated.

Inadequate Resolution

Several measures of allelic diversity (table 2) suggest that nucleotide substitution rates vary across the three nuclear loci. It is important to separate the effects of mutation rate on the process of lineage sorting from the empirical ability to detect this process. The theoretical dynamics of neutral lineage sorting to reciprocal monophyly are independent of mutation and substitution rate (Neigel and Avise 1986; Hudson 1990). In other words, lineage sorting and coalescent processes proceed regardless of whether unique derived mutations mark (for our eyes) the allelic lineages after population separation. A higher rate of mutation and substitution (without homoplasy) simply makes the underlying process more visible empirically.

Our nuclear gene tree reconstructions are limited by low numbers of cladistically informative sites (figs. 1–3). For this reason, maximum-parsimony trees constrained to display reciprocal monophyly between the Atlantic and the Gulf were not significantly worse than the shortest unconstrained trees observed. However, there were no Atlantic or Gulf synapomorphies observed at any locus. Furthermore, to alter the para- or polyphyletic structure in each data set to one of reciprocal monophyly by adding characters (instead of making existing characters more homoplasious in a constrained tree) would require the addition of multiple synapomorphs uniting all Atlantic haplotypes to the exclusion of those in the Gulf, and vice versa. By these latter lines of reasoning, the absence of reciprocal monophyly in the nuclear gene trees is probably not merely an artifact of limited phylogenetic resolution.

Lineage Sorting at the CV-32 Locus

If the retention of ancestral polymorphisms explains the CV-32 gene tree polyphyly in the Atlantic and Gulf populations of *C. virginica*, then an intriguing view of the lineage sorting processes emerges from a combined examination of the locus genealogy (fig. 3) and the CV-32 allele frequencies (Hare and Avise 1996). The low frequency of *Nsi* I-absent alleles in the Gulf suggests that few lineages need disappear before an Atlantic *Nsi* I-absent clade would become monophyletic (this can be appreciated from fig. 3 only if it is remembered that Gulf alleles are disproportionately represented on the *Nsi* I-absent side of the genealogy due to our nonrandom sampling design). The same reasoning applies to *Nsi* I-present alleles that are at low frequency in the Atlantic. Thus, although the phylogeny of haplotypes at the CV-32 locus does not distinguish Atlantic from Gulf oyster populations, the allele frequencies at this locus nearly do, and relatively few lineage extinctions would be required to produce a gene tree displaying reciprocal monophyly.

In accounting for occurrence of Gulf-predominant *Nsi* I RFLP alleles in the Atlantic, and vice versa (Karl and Avise 1992; Hare and Avise 1996), evolutionary convergence of the *Nsi* I phenotype could not be eliminated as a possible contributing factor. The localization of the *Nsi* I-present and *Nsi* I-absent states to different branches of the CV-32 gene genealogy (fig. 3) implies that convergent evolution at the *Nsi* I site has not occurred in the haplotypes we sampled.

Vicariance with Secondary Gene Flow

Secondary gene flow between populations is another possible source of apparent para- or polyphyly in gene trees, because it can produce genealogical patterns similar to those expected from differential lineage sorting in populations separated for evolutionary times that are short relative to effective population size (Avise, Ankney, and Nelson 1990; Wakeley 1996). Several studies have found evidence that mtDNA introgresses across species boundaries more readily than do nuclear loci (Avise 1994, pp. 287–289), but the conundrum to be explained in oysters is just the reverse. If secondary

gene flow following vicariant population separation accounts for gene lineages shared between Atlantic and Gulf oysters, then this exchange appears to have been more pronounced for nDNA than for mtDNA.

As an explanation for the shared allelic lineages between populations, contemporary gene flow does generate one phylogenetic prediction distinct from that of incomplete lineage sorting in long-separated populations. Under the former hypothesis, identical alleles should, in this case, be shared by Atlantic and Gulf oyster populations. Under the latter hypothesis, the alleles producing the para- or polyphyletic gene trees would not be identical (unless artifactually so because of limited empirical resolving power), but instead would merely represent longer-separated allelic lineages still shared between regions. Of course, for purposes of data analysis, any under-sampling of extant variation within either population also would mimic the effects of lineage sorting and might leave some true contemporary gene flow undetected.

The sequence data from all three nuclear loci examined here are compatible with the idea that barriers to gene flow currently exist between the Atlantic and Gulf oyster populations. In no case were identical alleles observed in the Atlantic and the Gulf (fig. 3). Most telling is the situation for CV-32: Despite the fact that long-distance gene flow within regions was indicated by several identical haplotypes shared by both Atlantic locales (South Carolina and Massachusetts), and by both Gulf locales (western Florida and Louisiana), never was an identical allele observed in both the Atlantic and the Gulf (fig. 3). Furthermore, the polymorphic *Nsi* I site at CV-32 was previously shown to display a sharp step-cline in allele frequency, suggesting strong barriers to gene flow at Cape Canaveral in eastern Florida (Hare and Avise 1996).

Restrictions on Gene Flow, No Vicariance

Using computer-simulated populations, Neigel and Avise (1986) showed that pronounced but stochastic phylogeographic ''breaks'' in a neutral-gene genealogy can arise in continuously distributed low-gene-flow populations along a linear habitat (such as a coastline), even in the complete absence of historical vicariance. Furthermore, such patterns, at face value, can generally match the mtDNA phylogeographic patterns reported, for example, in horseshoe crabs (Saunders, Kessler, and Avise 1986). Might such population models account for the joint nuclear/mitochondrial data now available for the American oyster? Several lines of reasoning suggest that the answer is ''no.''

First, such stochastic models alone do not predict categories 3 or 4 of genealogical concordance, which apply to the southeastern U.S. coastal faunas considered collectively. Thus, the models would not have predicted concordant Atlantic/Gulf distinctions for multiple taxa, or any concordance between the gene tree partitions and zoogeographic provincial boundaries. Second, evidence from allozymes (Buroker 1983) as well as from the wide geographic distributions of identical nuclear haplotypes (current study) indicates that the American oyster is not an exceptionally low-gene-flow species within either the Atlantic or Gulf coast regions.

Combinations of Factors

Considered collectively, the population genetic data available for the American oyster admit no simple explanation for the diverse phylogeographic patterns observed across loci. The mtDNA data, the frequencies of alleles at several anonymous nuclear loci, and the agreement of mtDNA phylogeographic patterns in oysters with those of several other coastal species in the southeastern U.S. all suggest a history of vicariant separation between Atlantic and Gulf oyster populations. They also suggest that pronounced contemporary barriers to gene flow exist, at least along the east coast of Florida. On the other hand, the nuclear gene genealogies presented here suggest that if historical vicariance has played any role in oyster population genetics, the relevant parameters of time since isolation and effective population size must have been such as to inhibit as yet the attainment of reciprocal monophyly through lineage sorting.

Various forms of historical and contemporary natural selection may have complicated matters further. For example, the possibility remains that balancing selection on most allozyme loci has played a role in maintaining similar electromorph frequencies in all populations. Alternatively, if allozymes are neutral, divergent directional selection in the face of moderate gene flow conceivably might account for the striking interregional differences in mtDNA and nuclear RFLP allele frequencies. Yet another complicating demographic factor might come into play: ''chaotic patchiness'' in local population-genetic constitution due to huge variances in reproductive success across families in this highly fecund species (Hedgecock, Tracey, and Nelson 1982). Hedgecock (1994) has argued from genetic evidence that effective population sizes in oysters are vastly smaller than might be supposed given their enormous census numbers. If so, idiosyncratic effects of genetic drift, both historical and contemporary, might influence local and regional population genetic outcomes in ways that are both somewhat unpredictable in space and time and difficult to capture adequately in finite samples from natural populations.

Conclusions

Because so few intraspecific studies that attempt comparisons among multiple gene genealogies are as yet available, we do not know whether the heterogeneity of phylogeographic patterns across loci in oysters will prove to be a common or a rare outcome. In gene-genealogical studies of humans, various autosomal loci, Y-linked genes, and mtDNA have painted richly varied phylogeographic pictures that suggest far greater complexity in human origins than was formerly supposed (see Gibbons 1997; Hammer et al. 1997; Hey 1997; and references therein). At this point, perhaps the most important realization is that for a variety of reasons involving demography, history, and selection, independent DNA sequences within the same organismal pedigree

sometimes can present quite different phylogeographic patterns.

Acknowledgments

Work was supported by an NIH training grant to M.P.H. and by an NSF grant and University of Georgia funds to J.C.A. We wish to thank members of the Avise laboratory, M. Goodisman, S. Palumbi, K. Shaw, and two anonymous reviewers for helpful suggestions and comments.

LITERATURE CITED

AVISE, J. C. 1989. Gene trees and organismal histories: a phylogenetic approach to population biology. Evolution **43**: 1192–1208.

———. 1992. Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. Oikos **63**:62–76.

———. 1994. Molecular markers, natural history and evolution. Chapman & Hall, New York.

———. 1995. Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. Conserv. Biol. **7**:45–67.

———. 1996. Toward a regional conservation genetics perspective: phylogeography of faunas in the southeastern United States. Pp. 431–470 *in* J. C. AVISE and J. L. HAMRICK, eds. Conservation genetics: case histories from nature. Chapman & Hall, New York.

———. 1998. The history and purview of phylogeography: a personal reflection. Mol. Ecol. (in press).

AVISE, J. C., C. D. ANKNEY, and W. S. NELSON. 1990. Mitochondrial gene trees and the evolutionary relationship of mallard and black ducks. Evolution **44**:1109–1119.

AVISE, J. C., J. ARNOLD, R. M. BALL JR., E. BERMINGHAM, T. LAMB, J. E. NEIGEL, C. A. REEB, and N. C. SAUNDERS. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Syst. **18**:489–522.

AVISE, J. C., and R. M. BALL. 1990. Principles of genealogical concordance in species concepts and biological taxonomy. Oxf. Surv. Evol. Biol. **7**:45–67.

AVISE, J. C., and K. WOLLENBERG. 1997. Phylogenetics and the origin of species. Proc. Natl. Acad. Sci. USA **94**:7748–7755.

BALL, R. M., J. E. NEIGEL, and J. C. AVISE. 1990. Gene genealogies within the organismal pedigrees of random-mating populations. Evolution **44**:360–370.

BAUM, D. A., and K. L. SHAW. 1995. Genealogical perspectives on the species problem. Pp. 289–303 *in* P. C. HOCH and A. G. STEPHANSON, eds. Experimental and molecular approaches to plant biosystematics. Monogr. Syst. Bot. Mo. Bot. Gard. **53**.

BERNARDI, G., P. SORDINO, and D. A. POWERS. 1993. Concordant mitochondrial and nuclear DNA phylogenies for populations of the teleost fish *Fundulus heteroclitus.* Proc. Natl. Acad. Sci. USA **90**:9271–9274.

BIRKY, C. W. JR., P. FUERST, and T. MARUYAMA. 1989. Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. Genetics **121**:613–627.

BRIGGS, J. C. 1974. Marine zoogeography. McGraw-Hill, New York.

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.

BURTON, R. S., and B.-N. LEE. 1994. Nuclear and mitochondrial gene genealogies and allozyme polymorphism across a major phylogeographic break in the copepod *Tigriopus californicus.* Proc. Natl. Acad. Sci. USA **91**:5197–5201.

CRANDALL, K. A., and A. R. TEMPLETON. 1993. Empirical tests of some predictions from coalescent theory with some applications to intraspecific phylogeny. Genetics **134**:959–969.

DESALLE, R., and L. V. GIDDINGS. 1986. Discordance of nuclear and mitochondrial DNA phylogenies in Hawaiian *Drosophila.* Proc. Natl. Acad. Sci. USA **83**:6902–6906.

GALTSOFF, P. S. 1964. The American oyster. Fishery bulletin no. 64, U.S. Government Printing Office, Washington, D.C.

GIBBONS, A. 1997. Ideas on human origins evolve at anthropology gathering. Science **276**:535–536.

HAMMER, M. F., A. B. SPURDLE, T. KARAFET et al. (11 coauthors). 1997. The geographic distribution of human *Y* chromosome variation. Genetics **145**:787–805.

HARE, M. P., and J. C. AVISE. 1996. Molecular genetic analysis of a stepped multilocus cline in the American oyster (*Crassostrea virginica*). Evolution **50**:2305–2315.

HARE, M. P., S. A. KARL, and J. C. AVISE. 1996. Anonymous nuclear DNA markers in the American oyster and their implications for the heterozygote deficiency phenomenon in marine bivalves. Mol. Biol. Evol. **13**:334–345.

HEDGECOCK, D. 1994. Does variance in reproductive success limit effective population sizes of marine organisms? Pp. 122–134 *in* A. R. BEAUMONT, ed. Genetics and evolution of aquatic organisms. Chapman & Hall, London.

HEDGECOCK, D., M. L. TRACEY, and K. NELSON. 1982. Genetics. Pp. 297–403 *in* L. G. ABELE, ed. The biology of Crustacea. Vol. 2. Academic Press, New York.

HEY, J. 1997. Mitochondrial and nuclear genes present conflicting portraits of human origins. Mol. Biol. Evol. **14**:166–172.

HEY, J., and J. WAKELEY. 1997. A coalescent estimator of the population recombination rate. Genetics **145**:833–846.

HOELZER, G. A. 1997. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees revisited. Evolution **51**:622–626.

HUDSON, R. R. 1987. Estimating the recombination parameter of a finite population model without selection. Genet. Res. Camb. **50**:245–250.

———. 1990. Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. **7**:1–44.

HUDSON, R. R., and N. L. KAPLAN. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111**:147–164.

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

———. 1993. PCR-based assays of Mendelian polymorphisms from anonymous single-copy nuclear DNA: techniques and applications for population genetics. Mol. Biol. Evol. **10**: 342–361.

LITTLEWOOD, D. T. J. 1994. Molecular phylogenetics of cupped oysters based on partial 28S rRNA gene sequences. Mol. Phylogenet. Evol. **3**:221–229.

MCDONALD, J. H., B. C. VERRELLI, and L. B. GEYER. 1996. Lack of geographic variation in anonymous nuclear polymorphisms in the American oyster, *Crassostrea virginica.* Mol. Biol. Evol. **13**:1114–1118.

MADDISON, W. 1995. Phylogenetic histories within and among species. Pp. 273–287 *in* P. C. HOCH and A. G. STEPHANSON,

eds. Experimental and molecular approaches to plant biosystematics. Monogr. Syst. Bot. Mo. Bot. Gard. **53**.

MARSH, A. G., and T. T. CHEN. 1995. A divergent cDNA homologue of the c-*myc* proto-oncogene in the eastern oyster *Crassostrea virginica*: implications for Myc evolution. Mol. Mar. Biol. Biotechnol. **4**:185–192.

MAYNARD SMITH, J., and J. HAIGH. 1974. The hitchhiking effect of a favorable gene. Genet. Res. **23**:23–35.

MOORE, W. S. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution **49**:718–726.

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

NEIGEL, J. E., and J. C. AVISE. 1986. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. Pp. 515–534 *in* E. NEVO and S. KARLIN, eds. Evolutionary processes and theory. Academic Press, New York.

OLSEN, G. J., H. MATSUDA, R. HAGSTROM, and R. OVERBECK. 1994. FastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput. Appl. Biosci. **10**:41–48.

ORITA, M., H. IWAHANA, H. KANAZAWA, K. HAYASHI, and T. SEKIYA. 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc. Natl. Acad. Sci. USA **86**:2766–2770.

ORTÍ, G., M. P. HARE, and J. C. AVISE. 1997. Detection and isolation of nuclear haplotypes by PCR-SSCP. Mol. Ecol. **6**:575–580.

PAMILO, P., and M. NEI. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. **5**:568–583.

RAND, D. M., M. L. DORFSMAN, and L. M. KANN. 1994. Neutral and non-neutral evolution of *Drosophila* mitochondrial DNA. Genetics **138**:741–756.

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**: 397–406.

SAUNDERS, N. C., L. G. KESSLER, and J. C. AVISE. 1986. Genetic variation and geographic differentiation in mitochondrial DNA of the horseshoe crab, *Limulus polyphemus*. Genetics **112**:613–627.

STEPHAN, W., and S. J. MITCHELL. 1992. Reduced levels of DNA polymorphism and fixed between-population differences in the centromeric region of *Drosophila ananassae*. Genetics **132**:1039–1045.

SWOFFORD, D. L. 1991. PAUP: phylogenetic analysis using parsimony. Version 3.1. Illinois Natural History Survey, Champaign.

TAJIMA, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**: 585–595.

TAKAHATA, N. 1989. Gene genealogy in three related populations: consistency probability between gene and population trees. Genetics **122**:957–966.

———. 1990. A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. Proc. Natl. Acad. Sci. USA **87**:2419–2423.

TEMPLETON, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution **37**:221–244.

TEMPLETON, A. R., and N. J. GEORGIADIS. 1996. A landscape approach to conservation genetics: conserving evolutionary processes in the African Bovidae. Pp. 398–430 *in* J. C. AVISE and J. L. HAMRICK, eds. Conservation genetics: case histories from nature. Chapman & Hall, New York.

WAKELEY, J. 1996. Distinguishing migration from isolation using the variance of pairwise differences. Theor. Popul. Biol. **49**:369–386.

WILSON, A. C., R. L. CANN, S. M. CARR et al. (11 co-authors). 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linn. Soc. **26**:375–400.

WU, C.-I. 1991. Inferences of species phylogeny in relation to segregation of ancient polymorphisms. Genetics **127**:429–435.

MARCY K. UYENOYAMA, reviewing editor

Accepted October 24, 1997