Contrasting Population Structure from Nuclear Intron Sequences and mtDNA of Humpback Whales

Stephen R. Palumbi and C. Scott Baker†*

*Kewalo Marine Lab and †Department of Zoology, University of Hawaii

Powerful analyses of population structure require information from multiple genetic loci. To help develop a molecular toolbox for obtaining this information, we have designed universal oligonucleotide primers that span conserved intron-exon junctions in a wide variety of animal phyla. We test the utility of exon-primed, intron-crossing amplifications by analyzing the variability of actin intron sequences from humpback, blue, and bowhead whales and comparing the results with mitochondrial DNA (mtDNA) haplotype data. Humpback actin introns fall into two major clades that exist in different frequencies in different oceanic populations. It is surprising that Hawaii and California populations, which are very distinct in mtDNAs, are similar in actin intron alleles. This discrepancy between mtDNA and nuclear DNA results may be due either to differences in genetic drift in mitochondrial and nuclear genes or to preferential movement of males, which do not transmit mtDNA to offspring, between separate breeding grounds. Opposing mtDNA and nuclear DNA results can help clarify otherwise hidden patterns of structure in natural populations.

Introduction

Recent studies of population structure and genetic diversity have employed new analytical and technical tools that provide high-resolution genetic information. These tools, including polymerase chain reaction (PCR) amplification and sequencing of DNA fragments (Saiki et al. 1988; Hillis and Moritz 1990; Vigilant et al. 1991; Baker et al. 1993*b*) and the phylogenetic analysis of allelic variants (Slatkin and Maddison 1989, 1990), combine to provide unprecedented power to detect and interpret genetic variation in natural populations. Many of the data for these analyses have come from sequences of mitochondrial genes because (1) mitochondrial DNA (mtDNA) has a fast rate of sequence evolution (Brown et al. 1982), (2) gene content is conserved across broad taxonomic boundaries (Brown 1985), and (3) universal primers are available to provide access to the mitochondrial genome in a wide variety of different taxa (Kocher et al. 1989; Palumbi et al. 1991).

Despite its advantages, the analysis of mtDNA sequence variability is widely recognized as suffering from two major weaknesses. First, the mitochondrial genome is but a single genetic locus. Lack of recombination links the 37 different genes of the animal mitochondrial genome into a single genetic entity (Wilson et al. 1985), transmitted in its entirety to progeny (Attardi 1985). Reliance on a single genetic locus greatly diminishes power to detect significant spatial or temporal structure. This is because genetic drift in populations involves random change in gene frequencies, and these changes will not occur in exactly the same way in independent loci. Thus, an accurate investigation of spatial or temporal genetic structure should consider the variety of patterns seen in different loci (Slatkin and Maddison 1990; Slatkin and Hudson 1991). In addition, the phylogenetic tree derived from a single locus may not accurately reflect the genetic history of a population, species, or genus (Ball et al. 1990). Avise and others (e.g., Karl and Avise 1992; Karl et al. 1992) have argued that it is the concordance of phylogenetic patterns across several loci that provides the most insightful views of population structure.

A second problem with exclusive reliance on mtDNA data is that such analyses allow only the reconstruction of maternal lineages (Wilson et al. 1985; Avise et al. 1987). In species unlikely to show gender-biased dispersal (such as many plants and marine animals), this distinction may be unimportant. However, for species in which complex social behaviors exist (e.g., many vertebrates [Greenwood 1983]), differences in the dispersal of males versus females may mean that population structure of maternally inherited mtDNA may differ from that of biparentally

Key words: mtDNA, introns, PCR, population structure, whales.

Address for correspondence and reprints: Stephen R. Palumbi, Kewalo Marine Lab, 41 Ahui Street, Honolulu, Hawaii 96813.

Mol. Biol. Evol. 11(3):426–435. 1994.

0737-4038/94/1103-0010$02.00

inherited nuclear DNA (Karl and Avise 1992; Karl et al. 1992).

Our understanding of systematic relationships and population structure will benefit from high-resolution analysis of nuclear DNA alleles to complement ongoing studies of mtDNA sequence variation (Lessa 1992; Slade et al., 1993). Parallel analysis of nuclear DNA and mtDNA variation allows the partitioning of maternal and biparental components of dispersal and population structure (e.g., see Karl et al. 1992). Study of sequence-based nuclear DNA variation can also take advantage of current theoretical developments in phylogenetic reconstruction (Felsenstein 1988), the cladistic measurement of gene flow (Slatkin and Maddison 1990), and the coalescent approach to population genetic models (Takahata 1988).

Universal PCR primers anneal to exons of highly conserved nuclear genes (such as actin) and are of broad taxonomic utility. Because the PCR products span non-coding introns (which often occur in conserved places within genes), the amplified DNA segment has potentially high rates of sequence evolution. Sequence analysis of the introns provides high-resolution allelic phylogenies comparable to those from mtDNA studies. Alternatively, restriction-fragment-length polymorphism (RFLP) analysis of these products can provide a rapid means of identifying major allelic variants in natural populations (Lessa and Applebaum 1993; Slade et al., in press). The broad applicability, potential high rates of variability, and well-understood genetic background of this nuclear allelic system make it attractive for molecular studies of population structure and genetic diversity (Lessa 1992; Slade et al., 1993).

Here, we use exon-primed, intron-crossing (EPIC) PCR amplification to assay variation in nuclear loci of humpback whales (*Megaptera novaengliae*). High-resolution mtDNA patterns are available for this species, on the basis of sequences of mitochondrial D-loops (Baker et al. 1993*b*), allowing a comparison of nuclear DNA and mtDNA sequence variation.

Methods

To date we have developed a suite of four primer pairs that amplify introns in conserved genes (Appendix). Many more such primers can be developed from the extensive genetic databases available (e.g., see the primers in the work of Lessa and Applebaum [1993] and Slade et al. [1993]). To illustrate the use of these primers, we focused on the first intron in the gene for the muscle and cytoskeletal protein actin. Genes in this small family are known to be highly conserved across animal phyla (e.g., see Foran et al. 1985), and codon bias limits heterogeneity at fourfold and twofold sites. In addition, some intron positions are also conserved. For example, the first intron occurs near amino acid position 41 in most species examined (e.g., see Kowbel and Smith 1989), allowing exon primers to be developed both upstream and downstream from the intron.

Nucleotide sequences surrounding amino acid position 41 were aligned, and PCR primers were designed that are 85%–95% similar between species in different phyla (fig. 1). Using these primers, we performed PCR amplifications in 25 μl total volume, using buffers and reagents described by Palumbi et al. (1991) and using *Taq* polymerase (Cetus Perkin Elmer). The thermal profile for initial amplifications with EPIC primers was as follows: 3 cycles at 94°C for 30 s, 45°C for 30 s, and 72°C for 90 s, followed by 35 cycles at 94° for 30 s, 55°C for 30 s, and 72°C for 90 s.

Amplification of nuclear introns produces copies of both alleles in diploid organisms. If these alleles have very different sequences, direct sequencing of the PCR product often yields confusing results. This is especially true if the alleles differ by insertions or deletions: in this case the sequencing ladders appear to be superimpositions of two different ladders that are slightly out of phase with one another. Because of the importance of separating different allele sequences during direct sequencing of PCR products, amplified introns were cloned in a phagemid vector (pBluescript; Stratagene). Each cloned copy of the intron represents a single allele, and unambiguous sequences of both alleles in an individual can be obtained by sequencing multiple clones. Cloning of PCR products is often difficult, but in this case it was enhanced by the TA cloning system described by Marchuk et al. (1991). Plasmid DNA was restriction-digested with *Eco*RV, and overhanging T nucleotides were added (Marchuk et al. 1991). PCR products (100 ng) were ligated with 100 ng vector for 1.5 h at room temperature

```
                --- primer ACT 1--------->   G   R   P   R   H   Q
STARFISH C   GCT GTC TTC CCG TCC ATC GTC GGG AGA CCC CGC CAC CAG
PURP CY2A    ... ... ... ..A ... ... ..T ..C ... ... ..T ... ...
MOUSE        ... ... ... ... ... ... ..G ..T C.C ..A ... ... ...
DROS         ..C ... ... ..A ..G ..T ..G ..A C.T ... ..T ... ...
YEAST        ... ... ... ..A ..T ... ... ..T ... ..A A.A ... ..A
STARFISH M   ... ..G ... ..A ... ..T ..A ... ... ..A ..T ... ...
XENOPUS      ... ..G ... ..T ... ..T ..G ..A C.T ... ..T ... ...
C.ELEGANS    ..C ..G ... ..A ... ..T ... ..A ... ..A ..T ..T ..A
CHICKEN      ..A ..T ... ..T ... ... ..G ..T C.T ... A.G ... ...
ARTEMIA      ..C ..T ... ..C ... ..T ... ..C C.T ..T A.A ... ...

              G   V  <------ primer ACT 2-----------
MOUSE        GGT GTC ATG GTG GGT ATG GGA CAG AAG GAC TCC TAC GTA
CHICKEN      ... ... ... ... ... ... ..T ... ... ... ... ... ...
CARP         ... ... ... ..T ... ... ... ... ... ... AG. ... ..T
STARFISH C   ... ..G ... ..T ... ... ..T ... ... ... AG. ... ..C
STARFISH M   ... ..G ... ..A ... ... ..T ... ... ... AG. ... ...
PURP CY2A    ... ... ... ..T ... ... ... ..A ... ... AG. ... ..T
C.ELEGANS    ... ..G ... ..C ... ... ... ... ... ... ... ... ..C
```

FIG. 1.—Alignment of actin-coding sequences in different phyla. Top and bottom panels are sequences 5′ and 3′ of codon 41, respectively, where a conserved intron/exon junction occurs (see Kowbel and Smith 1989). Overbars mark the extent of the two universal actin primers used in this study. Amino acids coded by conserved codons flanking the intron boundary are marked. Sequences are from GenBank.

and were transformed into competent cells (Sambrook et al. 1989).

Screening of these colonies was done by PCR: inserts were amplified from bacterial colonies by using 25-μl PCR reactions (standard reagent concentrations; Palumbi et al. 1991) and two oligonucleotide primers that annealed to the pBluescript vector on either side of the insert (M13 and M13 reverse primers as provided by Stratagene). Each colony to be screened was touched lightly with a flame-sterilized bacterial transfer loop. Each loop was then twirled in a separate 25-μl PCR cocktail (complete with all reagents, including enzyme) to release bacteria into the amplification buffer. After overlaying one drop of oil, reactions were amplified for 35 cycles at 94°C for 30 s, 55°C for 20 s, and 72°C for 60 s (short inserts may allow shorter extension times). PCR products were electrophoresed in 2% agarose/Tris-borate ethylenediaminetetraacetate (Sambrook et al. 1989) to identify the sizes of the inserts in each colony.

For colonies that showed inserts of the appropriate sizes, single-strand template was produced by helper-phage infection as described by the vector vendor. Single-strand DNA was precipitated with polyethylene glycol (Sambrook et al. 1989). Complete sequences were obtained using oligonucleotide primers spaced 250 bp apart along the intron (Operon).

The combination of T-vector cloning and PCR screening greatly speeds the analysis of many alleles. However, a disadvantage of this approach for nuclear introns is that *Taq* polymerase errors are incorporated into cloned products, and these errors become part of the allele sequence. We estimated the frequency of these errors by sequencing multiple clones in heterozygotes and comparing several individual sequences from the same allele. In this way, we estimated the error rate at ~1 transition substitution/1,000 bases.

DNA sequences were aligned by eye and were analyzed phylogenetically by using PAUP (Swofford 1991). Interspecific analyses using other phylogenetic methods (e.g., maximum parsimony and neighbor joining) showed no topological differences from PAUP trees (data not shown).

The cetacean DNA samples analyzed here were from free-ranging whales that were sampled by biopsy as described elsewhere (Baker et al. 1993*b*). Spooled genomic DNA was extracted from skin by standard proteinase K/phenol/chloroform methods (see Sambrook et al. 1989; Baker et al. 1993*b*). Introns that were sequenced were derived from individuals sampled either in wintering grounds in Hawaii or on feeding grounds in California. RFLP analysis of introns was performed on animals from all oceanic populations, including the North Atlantic, the Antarctic Ocean, Hawaii, Alaska, California, and Mexico (for details, see Baker et al. 1993*b*).

Results

Initial amplifications with act I and act II generally revealed several products. These products are derived either from multiple loci of this small gene family, from actin pseudogenes, or from random sections of the genome that have accidental homology to one or more primers. To identify the nature of these multiple bands, they were excised from agarose gels, reamplified, and cloned in pBluescript. To ascertain which of the amplified products were from the target intron, the sequences were compared with those expected on the basis of the intron/exon junctions of published actin genomic DNA sequences (fig. 1). Clones that did not show the predicted boundary amino acids were not analyzed further. Sequences from confirmed introns were used to design taxon-specific primers that produced good amplifications at high stringency.

With these procedures, a 1,409-bp actin intron was sequenced, in its entirety, from blue and humpback whales. Although other actin introns exist and are amplified by the universal primers, we chose the largest confirmed intron, in order to increase the chance of finding significant intraspecific variation. This large intron also was apparent in amplifications of most species of Mysticete whales, allowing us to use these intron data to compare species. After whale-specific primers were constructed, a smaller, 1,200-bp subsection of the intron was amplified and sequenced for bowhead whales and for multiple individual humpback whales sampled from two regional subpopulations known to be very different in mitochondrial haplotype (fig. 2).

Sequence variation between species is substantial (fig. 2) but is not noticeably clustered along the intron length. In particular, there are no obvious "hypervariable" regions as commonly seen in mtDNA D-loop sequences (e.g., see Cann et al. 1987). Transitions outnumber transversions by approximately threefold when blue whale sequences are compared with humpback whale sequences, whereas there are approximately equal numbers of the two types of substitutions when humpback whales are compared with bowhead whales. Three length changes are apparent, including a 19-bp deletion in the humpback sequence (at about position 776). Overall, the humpback sequence is ~3% different from the blue whale sequence and ~4% different from the bowhead sequence. These low percentages are consistent with other reports of slow rates of molecular evolution in cetaceans (Hoezel et al. 1991; Schlotterer et al. 1991; Baker et al. 1993*b*).

Despite these low rates of change between species, the intron sequence is long enough to show a substantial number

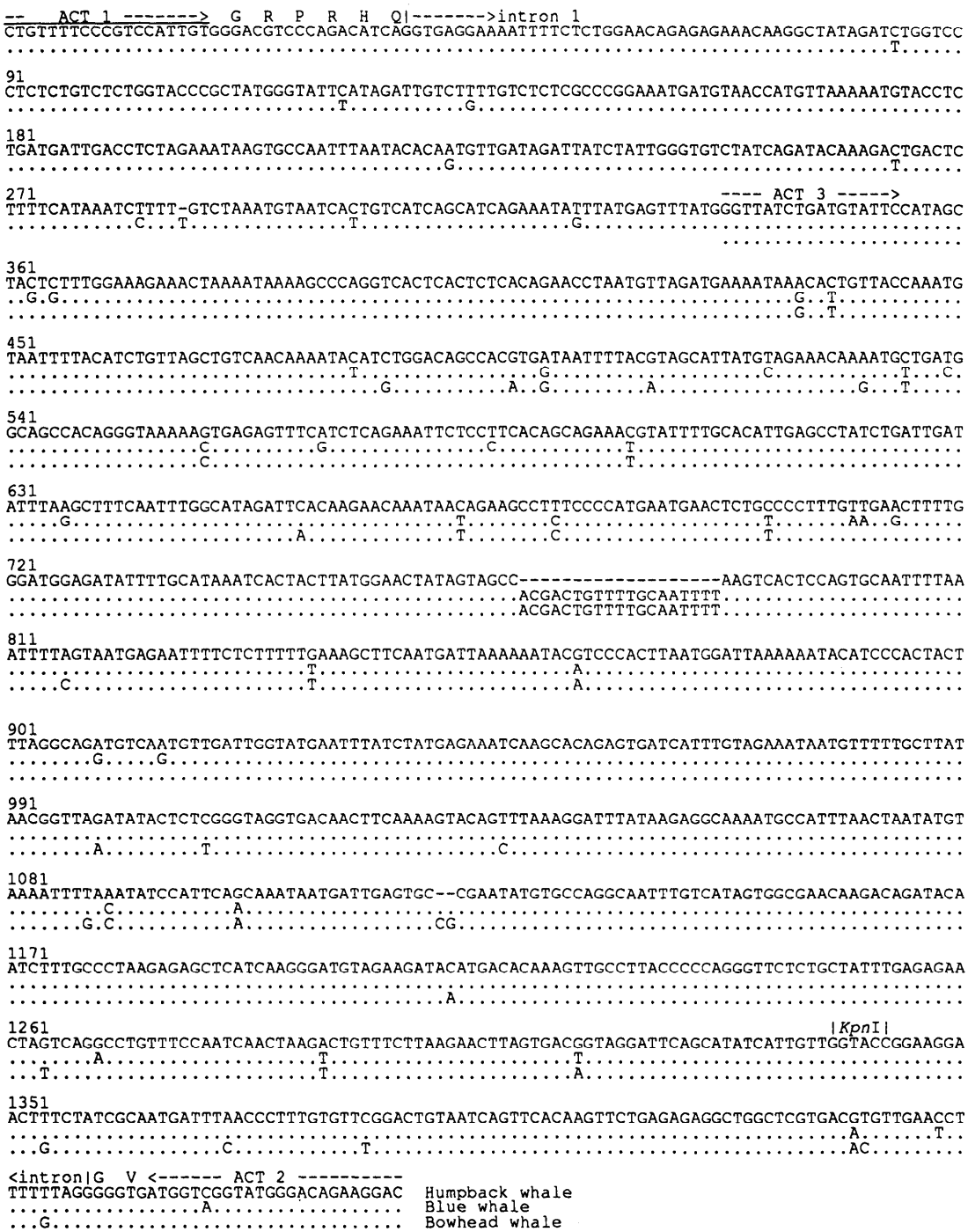

FIG. 2.—Complete actin intron sequences for representative blue and humpback whales. The original primers, the flanking codons, and the intron/exon boundaries are shown above the sequence. A dash (-) represents a deletion.

of intraspecific polymorphisms. Within the 10 humpback whale sequences, there are 17 polymorphic nucleotide positions, 8 of which are phylogenetically informative (i.e., occur in more than one allele). These phylogenetically informative positions are particularly important because it is unlikely that random *Taq* polymerase errors will occur at the same nucleotide positions. Alleles differ from one another by ≤12 base substitutions. There are only two transversion polymorphisms in this sample, and there are no intraspecific length polymorphisms.

Phylogenetic analysis of the 10 humpback alleles sequenced (fig. 3) yields a single most parsimonious tree

```
                                1111111111
                      34446680011122333
POSITION:             92390462404969145
                      29610356135801543
-----------------------------------
Humpback HI 1a   CAAACC?TAGGAAAGCC
Humpback HI 1b   CAGATCGTCAAGTGATT
Humpback HI 2a   CAGATCATCAGGAGATT
Humpback HI 2b   CAGACCGTAGGGAAGCC
Humpback HI 3    CAGACCGTAGGGAAGCC
Humpback CA 8a   CAAGCCGCAGGGAAGCC
Humpback CA 8b   CAAACCGTAGGGAAGCC
Humpback CA 11a TAGATTGTAAGGAAGCC
Humpback CA 11b CGGACCGTAGGGAAGCC
Humpback CA 54   CAAACCGTAGGGAAGCC
```

FIG. 3.—Variable positions in actin introns from humpback whales. Numbers correspond to the positions in fig. 2. Humpback sequences are from the Farallon Islands of California (CA) and the island of Hawaii (HI). Two sequences from the same individual (e.g., HI 2) represent two alleles in that individual and are distinguished by lowercase letters "a" or "b."

(by branch-and-bound search using PAUP), which has the same topology as trees generated by the neighbor-joining or maximum-likelihood method (as implemented in PHYLIP 3.5). The phylogenetic tree shows that alleles can be divided into two broad clades (clade A and clade B in fig. 4) differing by an average of 10 substitutions. Individuals from Hawaii and California are distributed throughout the tree (fig. 4), and there is no obvious geographic structure to the cladogram.

One of the nucleotide substitutions that distinguishes the two clades alters a restriction site for the endonuclease *Kpn*I. To describe the frequency of these clades in a larger sample, oligonucléotide primers that

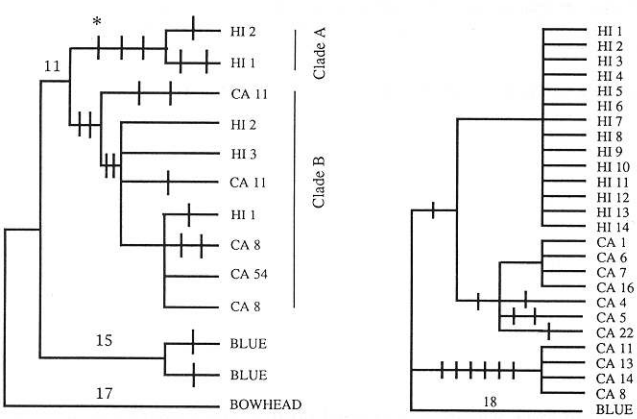

FIG. 4.—*Left,* Best phylogenetic tree of actin intron alleles, based on parsimony analysis (PAUP, branch-and-bound search). Vertical lines mark single nucleotide substitutions. Numbers above branches represent the substitutions between species. The asterisk (*) denotes the nucleotide substitution at position 1344, which alters a *Kpn*I site. Designations of individual whales are as in fig. 3. *Right,* Parallel phylogenetic tree, showing the relationships of mtDNA D-loop sequences for humpback whales from Californian and Hawaiian populations. Note that all 14 Hawaiian individuals showed identical sequences, distinguishable from the sequences of the Californian mtDNAs. For sequences, see the work of Baker et al. (1993*b*).

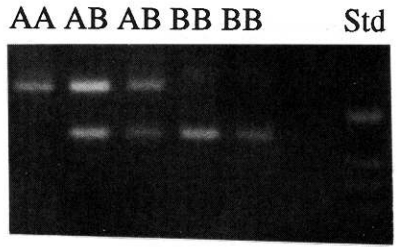

FIG. 5.—Major actin allele types in humpback whales, for allele clades A and B, distinguished by agarose electrophoresis after *Kpn*I digestion (overnight at 37°C in PCR cocktail). Allele designations have been confirmed by sequencing. Std = size standard.

flank this polymorphic site were used to amplify intron alleles from 128 individual humpback whales. The amplified products were digested with *Kpn*I and were visualized on agarose gels (fig. 5). Although there is much allelic variation within these clades that is not detected by this method, this simple test allows examination of the frequencies of these allele clades in a large number of individuals. The proportions of A and B alleles in all samples are in Hardy-Weinberg equilibrium (χ^2 test; $P>0.50$) and are similar in the Hawaiian and California samples (randomized χ^2; Roff and Bentzen 1989; $P>0.5$). However, the North Atlantic and South Pacific populations differ from those in the Northern Pacific (randomized χ^2; $P<0.01$; table 1).

Discussion

EPIC amplifications can be a powerful way to assess highly variable nuclear DNA sequences in taxa for which no other previous DNA sequence data are available. In our analysis, we demonstrate species differences in actin intron sequences and show significant population-level variation for these sequences from humpback whales. Parallel analyses using the same act I and act II primers have produced similar results for sea urchins, including substantial within- and between-species sequence variation (Palumbi, in press). Slade et al. (1993) and Lessa

Table 1
Frequencies of Alleles in Clade A and Clade B, for 128 Humpback Whales from Four Populations

	AA	AB	BB	*P*(B)
Hawaii (n=32)	8	16	8	0.50
California (n=30)	5	10	15	0.66
North Atlantic (n=42)	2	15	25	0.77
Southern oceans (n=24)	2	7	15	0.77

NOTE.—Populations that are indistinct are joined by vertical lines (randomized χ^2 and contingency-table analyses; $P>0.05$). Hawaii and California includes individuals from wintering grounds in Alaska and Mexico, respectively.

and Applebaum (1993) show that several other primer pairs, particularly those for aldolase and the *DQA* locus of the major histocompatibility system, also provide good amplifications from vertebrates. In the Appendix, we list three other primer pairs that tend to give consistent amplifications with a variety of taxa.

Intron primers can be used to examine nuclear loci in a large number of individuals or species. Lessa (1992) used globin primers to amplify and analyze variability between introns in populations of pocket gophers. He showed how denaturing gel gradient electrophoresis could be combined with intron amplifications to allow rapid analysis of alleles without sequencing. Slade et al. (1993) used a similar approach in a preliminary analysis of seal populations. In their study, variable restriction sites within introns were used to distinguish alleles, and they explain several complementary strategies for investigating genetic differences among EPIC products.

Technical Difficulties of Using Highly Conserved Primers

Like virtually all other PCR methods for sampling the nuclear genome, EPIC amplifications involve several technical and conceptual problems that need to be addressed in every new study. Foremost is the fact that primers for a conserved gene region that work on different species tend also to be conserved between different loci. For example, the actin primers described here show a good match to both cytoplasmic and muscle loci of actin (fig. 1), and initial amplifications usually yield several products corresponding to the different loci amplified. Most nuclear genes occur in small gene families or have closely related pseudogenes, so that this is likely to be a general result of using EPIC primers (Slade et al. 1993).

An advantage of these multiple products is that several, presumably independent loci can be identified using the initial primers (e.g., the aldolase primers in Lessa and Applebaum 1993), and these can be examined separately by using primers redesigned to be specific to only a single locus. A disadvantage is that it is necessary both to identify homologous loci in related species and to distinguish between allelic variation at a single locus and variation between loci.

In principle there are several ways to solve these problems in any new system. In our experience (also see Slade et al., in press), these different loci often have introns of different size, and these can usually be separated on agarose gels. In the current work, we selected the largest introns that were apparent after initial amplifications. Yet, how can we know whether these are from homologous loci? First, different loci generally arise from gene duplications far in the past, and loci tend to be more different from one another than are alleles at the same locus. Thus, a cladogram of sequences from closely related species should show that the sequence differences between presumed alleles within species are no greater than the sequence differences among species. Figure 4 shows that this is the case for the actin intron that we investigated here.

Furthermore, a cladogram of homologous sequences from a suite of species can be tested for concordance with other data. The topological relationships of actin introns from blue, humpback, and bowhead whales are the same as those derived from mtDNA, nuclear DNA, and morphological and paleontological analysis (see Barnes et al. 1985; Arnason et al. 1993; Baker et al. 1993*b*). If the sequences that we obtained were from different loci, the resulting phylogenetic trees would reflect the duplication of the loci, not the divergence of whale species. Concordance of our intron cladogram with other phylogenetic analyses of whales suggests that these sequences are from homologous loci.

An additional, independent way to investigate whether sequence variants represent different alleles at a locus or different loci is to test whether the distribution of sequence difference is as expected from Hardy-Weinberg equilibrium. For example, in our humpback whale data, we have shown the existence of two intron-sequence clades (clade A and clade B in fig. 4). Do these clades represent different loci? This is not likely, since the clades are most closely related to each other, and since blue and bowhead whale sequences fall outside both clades in a parsimony tree (fig. 4). Yet, a gene duplication *could* have occurred within the humpback lineage (which has existed for ~5 Myr; Barnes et al. 1985) and could have given rise to two loci that we now see as clades A and B. If this were the case, then each individual would possess both A- and B-type sequences, and all individuals would show two bands after restriction digestion (corresponding to the locus with a *Kpn*I site and to the locus without one). Instead, individuals often showed only one band on agarose gels (fig. 5), and the presumed allelic clades A and B always occurred in Hardy-Weinberg proportions within populations. These results show that the sequence variants that we present here are due to allelic variation at a single locus.

Strong demonstration of the allelic nature of sequence variants—by using either cladograms among closely related species or distribution of variants within species—is important in every new use of EPIC primers. The complex nature of the nuclear genome makes using these primers more difficult than using comparable mtDNA primers (e.g., see Kocher et al. 1989), but the potential for examining many different independent loci makes these primers attractive choices in population genetic surveys.

Interpreting Sequence Differences

A second potential difficulty with the techniques that we describe is the identification of allelic differences from sequence data. The sequences that we obtain from cloned PCR products include 1–2 polymerase errors/1,000 bases, so a fraction of sequence variants in the data set is due to amplification errors, not to true genetic variation. However, because *Taq* errors occur at random, it is unlikely that sequence differences that occur in a number of different alleles are artifactual. Thus, phylogenetically informative positions are most likely to represent true genetic variation. For instance, actin intron polymorphisms in humpback whales (fig. 3) include eight phylogenetically informative substitutions that cluster alleles into distinct clades (fig. 4). These are unlikely to be cloning or amplification artifacts.

By contrast, it is more difficult to be certain that a polymorphism seen in only a single sequence is not the result of a PCR error, rather than the result of true genetic variability. The eight transition substitutions that were seen in only one actin intron allele (fig. 3) fall into this category. Yet, these "singletons" do not substantially affect analyses of phylogeny or population structure, because, by definition, they only occur in a single sequence. In the phylogenetic analysis of alleles (fig. 4), these substitutions fall along branch tips leading to individual sequences. Although these potentially false substitutions have little effect on population comparisons, they will have two minor effects on analysis of EPIC amplifications: (1) they will increase apparent allelic diversity, and (2) they will result in a slight increase in average sequence heterogeneity.

If single sequence variants are impossible to verify, how can both alleles possessed by an individual be identified? If two sequences from an individual differ at phylogenetically informative positions, then it can be assumed that these represent different alleles. For example, sequences from Hawaii humpback individual 1 (HI 1) fell into both allele clades A and B (figs. 3 and 4). Because these clades are defined by numerous phylogenetically informative substitutions, these sequence differences are clearly not due to *Taq* errors. Restriction-endonuclease digestion with *Kpn*I verified that this individual was an AB heterozygote. Likewise, sequence differences between clones from HI 2 and CA 11 include differences at phylogenetically informative positions.

By contrast, if sequences from a single individual differ at one or two non–phylogenetically informative positions, then these differences might be due to polymerase errors. Sequencing a third clone will reveal which of these differences is due to random errors and which represents true allelic differences. To illustrate this approach we sequenced a third clone from HI 2 (HI 2c). This sequence was identical to the one listed first in figure 3 (HI 2a), except that it contained an A→G transition substitution at the 5′ end of the intron. All other sequences in the analysis show an "A" at this position, including the other allele of HI 2 (HI 2b, which differs from the first allele at eight positions). Thus, HI 2c probably represents a duplicate of HI 2a, with a *Taq* error introducing a single transition substitution.

Analysis of EPIC amplifications by restriction enzymes avoids the problems of *Taq* error. This is because the analysis is performed on the bulk amplification, not on a single clone of a single amplified DNA strand. Thus, even if *Taq* errors create a few DNA strands that do not possess a particular restriction site, these will not be visible in a stained agarose gel. However, restriction analysis presents its own set of problems, including the difficulty of assigning polymorphic restriction sites to the correct allele (see Slade et al. 1993). Using sequence data to identify phylogenetically informative restriction polymorphisms is probably the best mix of sequencing and RFLP approaches to intron analysis.

Population Genetics of Humpback Whales

Examination of patterns of geographic variation in actin introns among humpback whales can be compared with previous results from mtDNA analysis. For actin introns, geographic patterns between oceans or hemispheres parallel our results for mtDNA. In both analyses, the North Atlantic and South Pacific populations are genetically distinct from the northern Pacific populations (Baker et al. 1993*b*). In this comparison, the mtDNA and nuclear DNA analyses give similar estimates of gene flow. Baker et al. (1993*b*) estimate worldwide immigration rates ($N_e m_f$) for humpback whales to be 0.5–1.0 females per generation. An $N_e m_e$ value based on a single-locus F_{ST} calculated from the data in table 1 is ~3.0. Because this value represents gene flow from males as well as females, it should be twice as high as the estimate from mtDNA. This low-to-moderate amount of gene flow has allowed the buildup of some significant differences in gene frequencies, especially among mtDNA haplotypes.

Over smaller spatial scales (e.g., within the northern Pacific) mtDNA and nuclear DNA results show contrasting patterns. mtDNA sequences and RFLPs (Baker et al. 1990, 1993*b*) have shown that Hawaiian and Californian populations are genetically distinct. Not only are the mtDNA haplotypes different in the two regions, but Hawaii shows a uniquely low level of mtDNA diversity—lower than that in any other tested population of humpback whales (fig. 4; Baker et al. 1993*b*). Yet, the actin intron alleles show no major distinction between Hawaii and California individuals, and there is no evidence for decreased heterozygosity in Hawaii

compared with California. An F_{ST} calculated from Californian and Hawaiian samples indicates high gene flow ($N_e m_e = 10$), far higher than the value calculated from mtDNA data ($N_e m_f < 0.1$). Thus, between the two adjacent subpopulations (California and Hawaii) in our study, mitochondrial and nuclear alleles present opposing pictures of the population structure.

For humpback whales, this discrepancy may be the result of gender-biased migration between breeding grounds. If males tend to move between breeding grounds more often and to breed more successfully than do females, then nuclear alleles may be mixed more thoroughly than are the maternally inherited mitochondrial sequences. It has long been known that the "song of the humpback whale," a presumed male mating signal, is very similar in both Hawaii and the breeding grounds of the Californian whales in Mexico (Payne and Guinee 1983). Moreover, the song changes from year to year in a parallel fashion in these two separate breeding populations. These observations, plus observations of migratory movements by naturally marked individuals (Baker et al. 1986) and DNA fingerprinting similarities between Hawaiian and California populations (Baker et al. 1993*a*), have long led to speculation about exchange of males between these breeding grounds.

The genetic results presented here are consistent with such exchange. However, alternative explanations are also possible. The higher effective population size for nuclear genes than for mtDNA will slow genetic drift of intron alleles relative to mtDNA haplotypes (Wilson et al. 1985). Thus, low mtDNA diversity in Hawaii could be due to faster genetic drift for haploid, maternally inherited genes than for diploid, biparentally inherited loci.

In addition, genetic drift is a stochastic process and, simply by chance alone, may not have acted strongly at the actin locus that we examined. In many studies of protein polymorphisms, population structure is indicated by only a subset of loci. Clearly, data from additional intron studies are necessary to fully characterize the gene flow patterns within humpback whales and to provide a more robust estimate of migration between subpopulations. Use of several different loci will also allow comparisons of estimates of gene flow that are not confounded by the inherent differences between maternally and biparentally inherited genes.

Acknowledgments

We thank W. O. McMillan, E. Metz, A. Perry, D. Powers, and J. Quattro for discussion, criticism, or technical help, and we thank B. Abernathy, J. Bannister, J. Calambodkidis, J. Lien, O. Vasquez, R. Lambertsen, and M. Weinrich for assistance in sample collection. Many of these data were collected while SRP was on sabbatical leave at the Hopkins Marine Station (HMS) of Stanford University. We thank Dennis Powers for laboratory space, and we thank the staff and students of HMS for help and lively discussions. This work was supported by grants from the NSF, IWC, and National Geographic Society.

APPENDIX

Loci Useful for Exon-primed, Intron-crossing Amplifications

Actin

For details about this set of primers, see text and figure 1.

Creatine Kinase

Creatine kinase supplies muscles with energy by transferring a phosphate group from creatine phosphate to ADP. This enzyme is only known in vertebrates and echinoderms, but a very similar enzyme occurs in the protostome phyla (e.g., the Arthropods), where it is called "arginine kinase." Creatine kinase has three unlinked loci in most vertebrates known and occurs in a tandem triple repeat in sea urchins. Of the following primers shown in figure A1, CK6 and CK7 work well on most vertebrates tested. CK6 also works well with AK7 when amplifying the arginine kinase locus of arthropod DNAs, which appears to be a single-copy gene in *Drosophila* (Collier 1990).

Cytochrome c

Cytochrome c is a small protein vital to oxidative phosphorylation during electron transport. The amino-terminus and carboxy-terminus of the protein are very highly conserved, making it possible to develop primers that are appropriate for taxa as diverse as plants, animals, and yeast (Kemmerer et al. 1991). There are generally

```
CK6-5'          GAC CAC CTC CGA GTC ATC TCZ ATG
MOUSE           ... ... ... ..C ..G ... ... ...
CHICKEN         ... ..T ACC A.. ..G ... ..C ...
FISH            ... ... ... ... ... ... ..G ...
URCHIN          ..T ... ACT ..G ..T ..T ..C ...


CK7-3'          CAG GTG CTC GTT CCA CAT GAA
MOUSE           ... ... ... ... ... ... ...
CHICKEN         G.. .C. T.. A.. ... ... ...
FISH            ... ... ... ... ... ... ...
URCHIN          A.. A.. T.T ... ... ... ...


AK7-3'           GT GCC AAG GTT GGT DGG GCA
LOBSTER          .. ... ... ... ... G.. ...
URCHIN CK        .. A.. ... ... ..A T.. A..
HUMAN CK         .. ... C.. ... ..A.T.. ...
```

FIG. A1.—EPIC primers for creatine kinase and arginase kinase genes. References for full sequences are as follows: urchin, Wothe et al. (1990); chicken, Wirz et al. (1990); and lobster, Dumas and Camonis (1993) (GenBank accession HVARGK).

```
cytC-C-5'       AAG TGT GCY CAR TGC CAC AC
insect          CGC ..C ..C ..G ... ... ..
yeast           .GA ... CTA ..A ... ... ..
rice            ... ..C ..C ..G ... ... ..

cytC-B-3'       CAT CTT GGT GCC GGG GAT GTA TTT CTT
insect          ... ... ... ... ... ... ... C.. ...
yeast           ... ... ... T.. T.. T.. T.. A.. ...
rice            ... ... ... A.. T.. ... ... ... ...
```

FIG. A2.—EPIC primers for cytochrome c genes. Sequences are from Kemmerer et al. (1991).

one or two introns in cytochrome c. Vertebrates (including chickens, rodents, and humans) have a single intron at amino acid 56. Rice has one intron at position 29 and one at position 73, although other plants have slightly different intron positions (Kemmerer et al. 1991). The two insects studied (*Drosophila* and *Manduca*) have no introns. Although the plant introns seem to be hefty and might be useful in population or systematic studies, the vertebrate introns that we have found so far are disappointingly small (see fig. A2).

Proto-oncogene *int*

The proto-oncogene *int* was first identified as a gene in mice into which a tumor virus was inserted. The gene was subsequently found to be homologous to the product of the *wingless* locus in *Drosophila*. Amino acid conservation is high in this gene, and studies of genomic clones have revealed several introns in conserved positions.

The primers shown in figure A3 amplify the third intron in *int*. In mice this intron is ~600 bp long. In whales it is only ~300 bp. These primers work well in most vertebrates and have also been successful with urchins and nemerteans. In some taxa, *int* is obviously a small multigene family and gives a number of amplification products. Note that at position 12 there is a wide variety of different bases in different taxa, and we have included an inosine at this position. Inosines do not bind well with anything, but they do not disrupt the annealing of adjacent bases. They are a sort of molecular "filler" in primers.

```
INT A-5'        AAC CTT CAC AAC AAY GAG GC
HUMAN           ... ... ... ... ... ... ..
MOUSE           ... ... ... ... ... ... ..
XENOPUS         ..T ..G ... ... ..C C.. ..
DROSOPHILA      ..T ..G ... ... ..C ... ..

INT B-3'        TT GCA CTC TTG ICG CAT YTC
HUMAN           .. ... ... C.. T.. ... ...
MOUSE           .. ... ... ... G.. ... ...
XENOPUS         .. ... T.. C.. A.. ... ...
DROSOPHILA      .. ... ... C.. T.. ... ...
```

FIG. A3.—EPIC primers for proto-oncogene *int*. References for full sequences are as follows: human and mouse, van Ooyen et al. (1985), *Xenopus*, Noordermeer et al. (1989); and *Drosophila*, Rijsewijk et al. (1987).

LITERATURE CITED

ÁRNASON, Ú., A. GULLBERG, and B. WIDEGREN. 1993. Cetacean mitochondrial DNA control region: sequences of all extant baleen whales and two sperm whale species. Mol. Biol. Evol. **10**:960–970.

ATTARDI, G. 1985. Animal mitochondrial DNA: An extreme example of genetic economy. Int. Rev. Cytol. **93**:93–145.

AVISE, J. C., J. ARNOLD, R. MARTIN BALL, 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.

BAKER, C. S., D. GILBERT, M. WEINRICH, R. LAMBERTSEN, J. CALAMBOKIDIS, B. MCARDLE, G. CHAMBERS, and S. J. O' BRIEN. 1993*a*. Population characteristics of DNA fingerprints in humpback whales (Megaptera novaeangliae). J. Hered. **84**:281–290.

BAKER, C. S., L. M. HERMAN, A. PERRY, W. S. LAWTON, J. M. STRATEY, A. A. WOLMAN, G. D. KAUFMAN, H. E. WINN, J. D. HALL, J. M. REINKE, and J. OSTMAN. 1986. Migratory movement and population structure of humpback whales (*Megaptera novaeangliae*) in the central and eastern North Pacific. Mar. Ecol. Prog. Ser. **31**:105–119.

BAKER, C. S., A. PERRY, J. BANNISTER, M. WEINRICH, R. ABERNETHY, J. CALAMBOLCIDIS, J. LIEN, R. LAMBERTSEN, J. URBÁN RAMÍREZ, O. VASQUEZ, P. CLAPHAM, A. ALLING, S. O' BRIEN, and S. R. PALUMBI. 1993*b*. Abundant mitochondrial DNA variation and world-wide population structure in humpback whales. Proc. Natl. Acad. Sci. USA **90**:8239–8243.

BAKER, S., S. R. PALUMBI, R. H. LAMBERTSEN, M. T. WEINRICH, J. CALAMBOKIDIS, and S. J. O'BRIEN. 1990. The influence of seasonal migration on geographic distribution of mitochondrial DNA haplotypes in humpback whales. Nature **344**:238–240.

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.

BARNES, L. G., D. P. DOMMING, and C. F. RAY. 1985. Status of studies on fossil marine mammals. Mar. Mammal Sci. **1**:15–53.

BROWN, W. M. 1985. The mitochondrial genome of animals. Pp. 95–130. *in* R. J. MacIntyre, ed. Evolutionary genetics. Plenum, New York.

BROWN, W. M., E. M. PRAGER, A. WANG, and A. C. WILSON. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18**:225–239.

CANN, R. C., M. STONEKING, and A. C. WILSON. 1987. Mitochondrial DNA and human evolution. Nature **325**:31–36.

COLLIER, G. E. 1990. Evolution of arginine kinase within the genus *Drosophila*. J. Hered. **81**:177–182.

DUMAS, C., and J. CAMONIS. 1993. Cloning and sequence analysis of the cDNA for arginine kinase of lobster muscle. J. Biol. Chem. **268**:21599.

FELSENSTEIN, J. 1988. Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. **22**:521–565.

FORAN, D. R., P. J. JOHNSON, and G. P. MOORE. 1985. Evolution of two actin genes in the sea urchin *Strongylocentrotus fransiscanus*. J. Mol. Evol. **22**:108–116.

GREENWOOD, P. J. 1983. Mating systems and the evolutionary consequences of dispersal. Pp. 116–131 *in* I. R. Swingland and P. J. Greenwood, eds. The ecology of animal movement. Clarendon, Oxford.

HILLIS, D., and C. MORITZ. 1990. Molecular systematics. Sinauer, Sunderland, Mass.

HOEZEL, R. A., J. M. HANCOCK, and G. DOVER. 1991. Evolution of the cetacean mitochondrial D-loop region. Mol. Biol. Evol. **8**:475–493.

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

KEMMERER, E. C., M. LEI, and R. WU. 1991. Isolation and molecular evolutionary analysis of a cytochrome c gene from *Oryza sativa* (rice). Mol. Biol. Evol. **8**:212–226.

KOCHER, T. D., W. K. THOMAS, A. MEYER, S. V. EDWARDS, S. PÄÄBO, F. X. VILLABLANCA, and A. C. WILSON. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA **86**:6196–6200.

KOWBEL, D. J., and M. J. SMITH. 1989. The genomic nucleotide sequences of two differentially expressed actin-coding genes from the sea star Pisaster ochraceus. Gene **77**:297–308.

LESSA, E. P. 1992. Rapid surveying of DNA sequence variation in natural populations. Mol. Biol. Evol. **9**:323–330.

LESSA, E. P., and G. APPLEBAUM. 1993. Screening techniques for detecting allelic variation in DNA sequences. Mol. Ecol. **2**:119–129.

MARCHUK, D., M. DRUMM, A. SAULINO, and F. COLLINS. 1991. The construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res. **19**:1154.

NOORDERMEER, J., F. MEIJLINK, P. VERRIJZER, F. RIJSEWIJK, and O. DESTREE. 1989. Isolation of the *Xenopus* homolog of *int*-1/*wingless* and expression during neurla stages of early development. Nucleic Acids Res. **17**:11–18.

PALUMBI, S. R. PCR and molecular systematics. *In* D. Hillis and C. Moritz, eds. Molecular systematics, 2d. ed. Sinauer, Sunderland, Mass. (in press).

PALUMBI, S. R., A. P. MARTIN, S. ROMANO, W. O. MCMILLAN, L. STICE, and G. GRABOWSKI. 1991. The simple fool's guide to PCR. Spec. Publ. Department of Zoology, University of Hawaii, Honolulu.

PAYNE, R., and L. N. GUINEE. 1983. Humpback whale (*Megaptera novangliae*) songs as an indicator of 'stocks.' Pp. 333–358 *in* R. Payne, ed. Communication and behavior of whales. Westview, Boulder.

RIJSEWIJK, F., M. SCHUERMANN, E. WAGENAAR, P. PARREN, D. WEIGEL, and R. NUSSE. 1987. The *Drosophila* homolog of the mouse mammary oncogene *int*-1 is identical to the segment polarity gene *wingless*. Cell **50**:649–657.

ROFF, D. A., and P. BENTZEN. 1989. The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small samples. Mol. Biol. Evol. **6**:539–545.

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

SAMBROOK, J., E. F. FRITSCH, and T. MANIATIS. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SCHLOTTERER, C., C. B. AMOS, and D. TAUTZ. 1991. Conservation of polymorphic simple sequence loci in cetacean species. Nature **354**:63–65.

SLADE, R. W., C. MORITZ, A. HEIDEMAN, and P. T. HALE 1993. Rapid assessment of single copy nuclear DNA variation in diverse species. Mol. Ecol. **2**:359–373.

SLATKIN, M., and R. R. HUDSON. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics **129**:555–562.

SLATKIN, M., and W. P. MADDISON. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics **123**:603–613.

———. 1990. Detecting isolation by distance using phylogenies of genes. Genetics **126**:249–260.

SWOFFORD, D. L. 1991. PAUP: phylogenetic analysis with parsimony, version 3.0. Illinois. Natural History Survey, Champaign.

TAKAHATA, N. 1988. The coalescent in two partially isolated diffusion populations. Genet. Res. **52**:213–222.

VAN, OOYEN, A., V. KWEE, and R. NUSSE. 1985. The nucleotide sequence of the human *int*-1 mammary oncogene: evolutionary conservation of coding and non-coding sequences. EMBO J. **4**:2905–2909.

VIGILANT, L., M. STONEKING, H. HARPENDING, K. HAWKES, and A. C. WILSON. 1991. African populations and the evolution of human mitochondrial DNA. Science **253**:1503–1507.

WILSON, A. C., R. L. CANN, S. M. CARR, M. GEORGE, U. B. GYLLENSTEN, K. HELM-BYCHOWSKI, R. G. HIGUCHI, S. R. PALUMBI, E. M. PRAGER, R. D. SAGE, and M. STONEKING. 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linnean Soc. **26**:375–400.

WIRZ, T., U. BRANDLE, T. SOLDATI, J. P. HOSSLE, and J.-C. PERRIARD. 1990. A unique chicken B-creatine kinase gene gives rise to two B-creatine kinase isoproteins with distinct N-termini by alternative splicing. J. Biol. Chem. **265**:11656–11666.

WOTHE, D. D., H. CHARBONNEAU, and B. M. SHAPIRO. 1990. The phosphocreatine shuttle of sea urchin sperm: flagellar creatine kinase resulted from a gene triplication. Proc. Natl. Acad. Sci. USA **87**:5203–5207.

JULIAN ADAMS, reviewing editor

Received June 17, 1993

Accepted December 30, 1993