

PREDICTING NUCLEAR GENE COALESCENCE FROM MITOCHONDRIAL DATA: THE THREE-TIMES RULE

STEPHEN R. PALUMBI,^{1,2} FRANK CIPRIANO,^{1,3} AND MATTHEW P. HARE^{1,4}

¹Department of Organismic and Evolutionary Biology, Harvard University, Cambridge Massachusetts 02138

²E-mail: spalumbi@oeb.harvard.edu

Abstract.—Coalescence theory predicts when genetic drift at nuclear loci will result in fixation of sequence differences to produce monophyletic gene trees. However, the theory is difficult to apply to particular taxa because it hinges on genetically effective population size, which is generally unknown. Neutral theory also predicts that evolution of monophyly will be four times slower in nuclear than in mitochondrial genes primarily because genetic drift is slower at nuclear loci. Variation in mitochondrial DNA (mtDNA) within and between species has been studied extensively, but can these mtDNA data be used to predict coalescence in nuclear loci? Comparison of neutral theories of coalescence of mitochondrial and nuclear loci suggests a simple rule of thumb. The “three-times rule” states that, on average, most nuclear loci will be monophyletic when the branch length leading to the mtDNA sequences of a species is three times longer than the average mtDNA sequence diversity observed within that species.

A test using mitochondrial and nuclear intron data from seven species of whales and dolphins suggests general agreement with predictions of the three-times rule. We define the coalescence ratio as the mitochondrial branch length for a species divided by intraspecific mtDNA diversity. We show that species with high coalescence ratios show nuclear monophyly, whereas species with low ratios have polyphyletic nuclear gene trees. As expected, species with intermediate coalescence ratios show a variety of patterns. Especially at very high or low coalescence ratios, the three-times rule predicts nuclear gene patterns that can help detect the action of selection. The three-times rule may be useful as an empirical benchmark for evaluating evolutionary processes occurring at multiple loci.

Key words.—Forensics, introns, molecular systematics, monophyly, whales.

Received May 23, 2000. Accepted December 22, 2000.

The application of molecular sequence data to systematics, population biology, and forensic identification of species often depends on whether monophyly has evolved at particular loci between taxa (Hudson 1992; Moritz 1994; Moore 1995; Palumbi and Cipriano 1998). Monophyly of genes is basic to some views of the species concept (Baum and Shaw 1995) and has been used to define evolutionarily significant units for management of threatened species (Moritz 1994). A population that has been separate long enough to develop gene monophyly has probably developed novel combinations of alleles at many genetic loci and may possess unique evolutionary features including adaptation to local environment that should be considered by resource managers (Moritz 1994). The shape of intra- and interspecific gene trees has also been used to infer the demographic history of species (Takahata 1995) and speciation patterns (M. P. Hare, F. Cipriano, and S. Palumbi, unpubl. ms.). Monophyly is also important if molecular tools are used in the identification of managed species in commercial markets (whales: Baker et al. 1993; Cipriano and Palumbi 1999; seals: Malik et al. 1997; canned fish: Rehbein et al. 1997; cannabis: Jagadish et al. 1996; caviar: Cohen 1997). This is because species identification must be based on unique derived features to limit identification errors. Because such features only occur at loci that are monophyletic, understanding the evolution of sequence monophyly is an important practical and conservation goal.

Despite its importance, there is a gap between our theoretical understanding of the evolution of monophyly and

practical application of theory to empirical data. For example, under what circumstances can monophyly of nuclear alleles be expected in a particular species, and under what circumstances should few loci be expected to be monophyletic? Coalescence theory provides a powerful framework for understanding the evolution of monophyly at neutral loci (Neigel and Avise 1986; Avise and Ball 1990; Hudson 1990, 1992), showing in general that monophyly of alleles develops more quickly in species with low genetically effective population size (Neigel and Avise 1986; Hudson 1990). However, application of theory to particular populations and species requires two types of data generally not available: (1) species origination times; and (2) long-term genetically effective population sizes (Hudson 1990).

Another simple result from coalescence theory is that monophyly of alleles is expected to evolve more quickly in mitochondrial or chloroplast DNA than in nuclear loci. This is because organelle DNAs tend to be effectively haploid and inherited uniparentally, and thus have a genetically effective population size that is approximately four times smaller than that for most nuclear loci (Birky et al. 1989). This relationship allows a quantitative comparison between coalescence of organelle and nuclear genes and suggests a simple rule of thumb that may be useful for interpreting evolutionary processes affecting coalescence. In this paper, we outline the theoretical justification for the relationship between organelle and nuclear monophyly and show how this leads to a quantitative prediction of the probability of coalescence at a neutral nuclear locus. We test these predictions using sequence data from mitochondrial and nuclear loci in whales and dolphins, point out circumstances in which the predictions might fail in interesting ways, and show how this approach can provide insight about molecular evolutionary patterns between species.

³ Present address: Department of Biology, San Francisco State University, San Francisco, California 94132.

⁴ Present address: Department of Zoology, University of Maryland, College Park, Maryland 20742.

Coalescence Theory and the Evolution of Monophyly

In the absence of hybridization, lineage sorting at a locus will inevitably result in all alleles tracing back to a single lineage, one of many that may have existed in the ancestor (for different but complementary definitions of monophyly, see Tajima 1983; Nei 1987; Li 1997). Eventually, all loci evolving neutrally within a species will become monophyletic (Nei 1987), but the rate at which this occurs is difficult to predict without independent estimates of long-term effective population size. The mean time to coalescence is $4N_e$ generations for nuclear genes, where N_e represents the diploid effective population size (Kimura and Ohta 1969; Tajima 1983; Maynard Smith 1987). Selection, recombination, and linkage to other genes also affects coalescence at particular loci (Hudson 1990).

Neigel and Avise (1986) used computer simulations to monitor the approach to monophyly in mitochondrial lineages; in their simulations, monophyly develops more quickly than for nuclear loci. This is because mean time to monophyly is about $2N_f$ generations for mitochondrial mutations, where N_f is the genetically effective number of females in the population (Birky 1991). For species with a 1:1 sex ratio and equal variance in reproductive success for males and females, $2N_f \approx N_e$, and so mitochondrial monophyly develops four times faster than for the average nuclear locus. Birky (1991, Pp. 121–122) lists several important exceptions to this generalization.

These analyses show that when a mitochondrial locus reaches monophyly, the average neutral nuclear locus in that species is still polyphyletic or paraphyletic. Conversely, when the average nuclear locus is monophyletic within a species, the mitochondrial locus has been monophyletic for a long time interval, on average $3N_e$ generations (Palumbi and Cipriano 1998). Moreover, mitochondrial diversity within such species is likely to have been accumulating for about $2N_f \approx N_e$ generations since the most recent mitochondrial fixation event (Tajima 1983; Hudson 1990), and the average mitochondrial genetic distance within the species is about $2N_f\mu$ (where μ is the average mutation rate per lineage per generation; Tajima 1983; Kuhner et al. 1995). Thus, for a species that has been distinct for $4N_e$ generations, mitochondrial DNA (mtDNA) diversity averages $2N_f\mu \approx N_e\mu$ and the branch leading to the mtDNA cluster has an average length of about $3N_e\mu$ (Fig. 1).

Thus, a simple rule of thumb is that species showing mtDNA branches much more than three times longer than the average mtDNA diversity within that species are likely to show nuclear alleles that are also monophyletic (Palumbi and Cipriano 1998). In contrast, species with mtDNA branches much less than three times as long as the average intraspecific diversity are less likely to show a majority of nuclear loci that are monophyletic (Fig. 1). This “three-times rule” is a simple outgrowth of expected differences in drift between nuclear and organelle genes. It may be useful as a qualitative benchmark similar to the one long used by population geneticists to determine the impact of gene flow on evolutionary dynamics (e.g., Nm values much less than 1.0 result in independent evolutionary trajectories; Slatkin 1987).

It is possible to make the three-times rule more quantitative

by estimating the probability of genetic coalescence of nuclear loci in terms of the shape of the mitochondrial tree. For a large sample of alleles, Tavaré (1984, eq. 6.4) gives the probability of coalescence of a randomly chosen nuclear locus by time t_n in the past as:

$$g_1(t) = 1 - \sum_{k=2}^{\infty} \exp\{-k(k-1)t_n/2\}[(2k-1)(-1)^k], \quad (1)$$

where t_n is the number of generations since species divergence, measured in multiples of $2N_e$. To use this equation requires that we estimate t_n from mitochondrial coalescence patterns.

Imagine we have mitochondrial data from two species, A and B (Fig. 2A). To estimate t_n for species A in terms of mitochondrial coalescence patterns, we first divide the history of this lineage into two parts: the number of generations from the present back to coalescence of the mitochondrial sequences in species A ($2N_f$ generations assuming neutral coalescence patterns; Nei 1987) and the rest of the time (L generations) back to the ancestral node: $t = L + 2N_f$. We define the coalescence ratio (C_R) from the mitochondrial gene tree as the length of the branch L ($L\mu$) divided by the average intraspecific nucleotide diversity ($2N_f\mu$): $C_R = L/2N_f$.

We can rewrite Tavaré’s equation in terms of coalescence ratios if we can define t_n in terms of C_R . From the above,

$$t = 2C_R N_f + 2N_f, \quad (2a)$$

$$t = 2N_f(C_R + 1). \quad (2b)$$

Recalling that t_n is expressed in units of $2N_e$ generations, $t_n = t/2N_e \approx t/4N_f$ and so $t_n = (C_R + 1)/2$.

As a result, Tavaré’s equation (6.4) can be expressed as

$$g_1(t) = 1 - \sum_{k=2}^{\infty} \exp\{-k(k-1)(C_R + 1)/4\} \times [(2k-1)(-1)^k]. \quad (3)$$

Note that this equation assumes a large sample size, and that k would be indexed up to K instead of infinity for a sample of K alleles. Using this equation provides a prediction of the fraction of neutral nuclear loci, evolving independently, that are expected to show coalescence at various mitochondrial coalescence ratios (Fig. 2B). Although this conceptualization fails to take into account differences in coalescence rate in ancestral and derived populations, it provides a general starting point for understanding the quantitative behavior of the three-times rule under simplified conditions. One major stumbling block is that the above assumes that speciation time and mitochondrial divergence happen together (t generations in the past; see Fig. 2A). However, mitochondrial divergence is expected to precede speciation by an average of N_f generations (if ancestral and derived populations are the same size). In this case, $t = L + N_f$ and $t_n = (C_R + 1/2)/2$. At high coalescence ratios, this distinction will be minor. Further theoretical work might refine the relationship between mitochondrial coalescence ratios and nuclear gene coalescence under diverse circumstances.

The three-times rule, whether it is used in a qualitative or quantitative fashion, makes a large number of assumptions that may not apply in particular cases, including complete

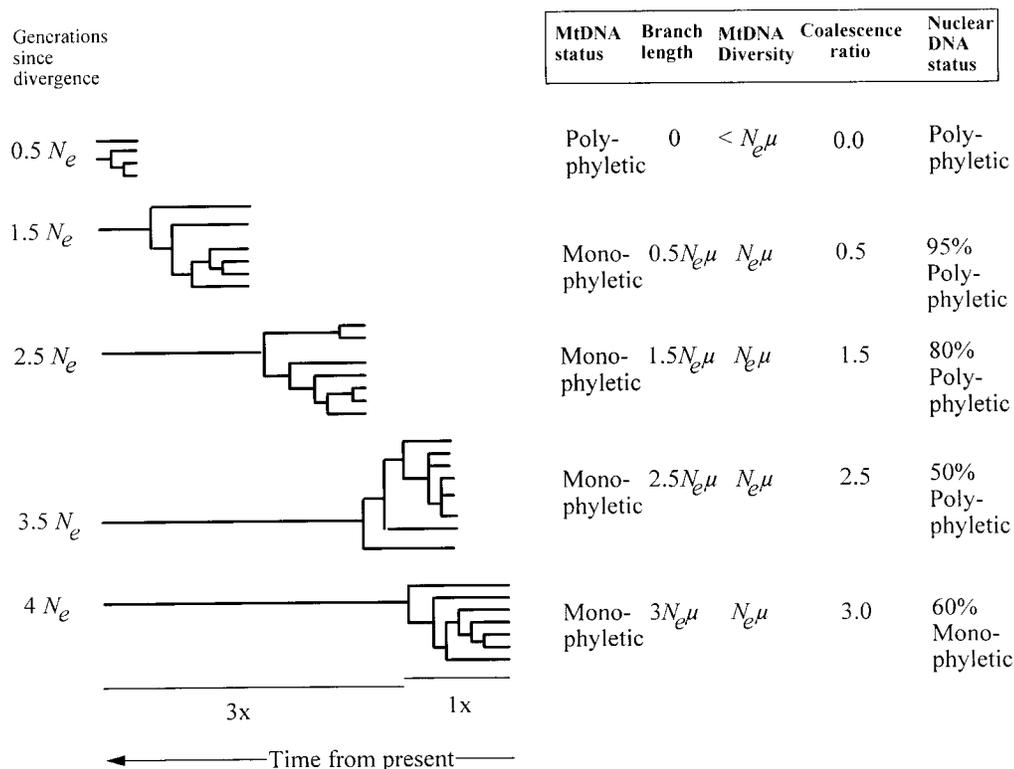
MtDNA trees for species that diverged up to $4N_e$ generations ago

FIG. 1. Evolution of mtDNA tree shape after species divergence. Five views are shown. At the top is a view of a species as it might appear $0.5N_e$ generations after divergence. Also shown are hypothetical, successive views of the same species after 1.5, 2.5, 3.5, and $4N_e$ generations, assuming stable population sizes and strictly neutral genetic variation. At first, mtDNA variation within the species is polyphyletic. After monophyly develops, mitochondrial data include a cluster of variable haplotypes connected by a branch to the rest of an interspecific phylogeny. Neutral theory suggests that nucleotide diversity among haplotypes averages $N_e\mu$, where N_e is the effective population size (assumed to be twice the number of breeding females in this case) and μ is the substitution rate per generation. Over time, the average mtDNA diversity within the population stays approximately the same (if population size and μ remain constant), but the branch leading to the haplotype cluster lengthens. A neutral model suggests the length of the branch leading to the haplotype cluster will average $(D - N_e)\mu$, where D is the number of generations since divergence. Average nuclear loci will be monophyletic when the length of the branch leading to the mtDNA clade defining a species is three times greater than the average mtDNA variation within the species, because the average nuclear locus becomes monophyletic after $D = 4N_e$ generations. Nuclear gene status is based on equation (3) for a sample size of 10 alleles.

neutrality, constant population size, equal sex ratios, equal variance in reproductive success, no recombination, and no hybridization. In addition, in the absence of selection, lineage sorting occurs randomly and so there is wide variance in fixation times among neutral loci (Fig. 2B; also see Kimura 1983; Nei 1987; Hudson 1992). Coalescence times may also be affected by linkage. Loci linked to genes under directional selection will show rapid coalescence (Hudson and Kaplan 1994). In contrast, at nucleotide positions near an amino acid polymorphism in the *Adh* gene of *Drosophila melanogaster* under balancing selection, intraspecific diversity is extremely high and coalescence has slowed (Kreitman 1991). Loci in regions of the genome with lower recombination may be affected more often by genetic hitchhiking (Begun and Aquadro 1992).

Keeping these limitations in mind, the three-times rule can nevertheless provide a yardstick with which we can use mtDNA data to make predictions about the genealogies of nuclear loci. Moreover, this simple rule can be employed to

gauge the accuracy of the neutral assumptions underlying many applications of coalescence theory. In fact, failure of the three-times rule to correctly predict nuclear monophyly may be more interesting than its success because failure of its predictions can lead to further understanding of the nature of selection acting across nuclear, mitochondrial, and chloroplast genomes. Here, we provide a preliminary test of the three-times rule using mitochondrial and nuclear data from congeneric species of baleen whales, dolphins, and data gleaned from the literature for other species. Once verified, the three-times rule may prove useful in predicting patterns of nuclear gene coalescence from the extensive mtDNA and cpDNA data available in a variety of diverse systems.

METHODS

Mitochondrial Data

Mitochondrial sequences from the 5' end of the control region were obtained as described in Baker et al. (1993). For

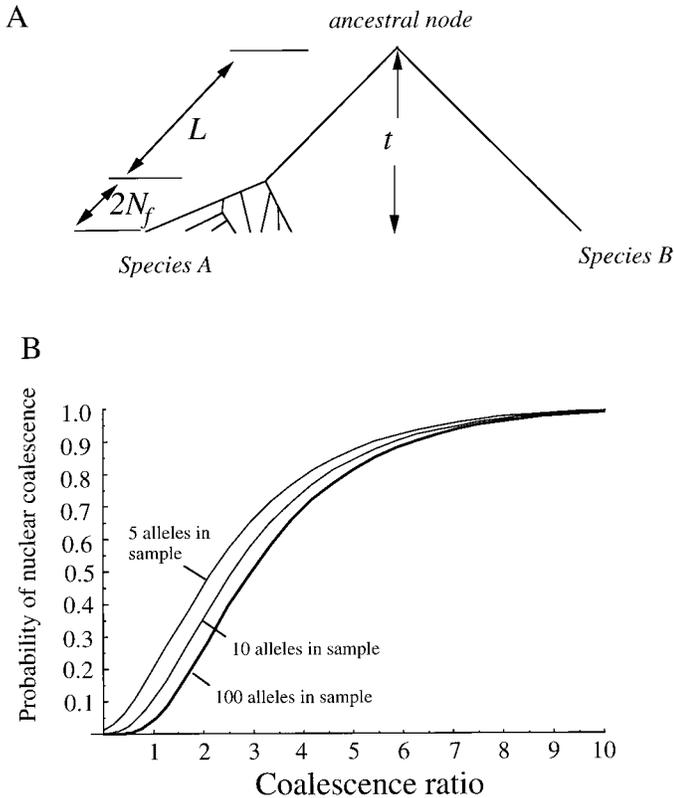


FIG. 2. (A) Conceptual framework for understanding the relationship between coalescence ratio ($C_R = L/2N_f$) and time since divergence. (B) Relationship between the probability of nuclear gene coalescence and the mitochondrial coalescence ratio. The bold line represents the case of a large sample of alleles using equation (3). Probability of monophyly is higher for smaller samples and has been evaluated using equation 6.4 in Tavaré (1984) for samples sizes of 10 and five alleles.

humpback whales, we used control region haplotypes ($n = 63$) found in a worldwide survey (Baker et al. 1993). For blue, fin, and minke whales, we used type sequences ($n = 9$, $n = 21$, and $n = 84$ respectively) from our forensic evaluations of the Asian whale meat market (Baker et al. 1996, Table 2). The outgroup used for whale comparisons was the bowhead whale (*Baleana mysticetus*). Cytochrome *b* data for dolphins are from Cipriano (1997) and M. P. Hare, F. Cipriano, and S. Palumbi (unpubl. ms.). The porpoise *Phocoena phocoena* was used as an outgroup for dolphin comparisons (for systematic details see Cipriano 1997).

Nuclear Loci

To determine whether particular nuclear loci are monophyletic within species, we focused on the distribution of fixed and polymorphic nucleotide positions within intron DNA from two conserved nuclear genes: the muscle protein actin and butyrophilin, a glycoprotein specific to fat globules produced by mammary glands (Davey et al. 1997). Such introns can be amplified for a wide variety of taxa using exon-priming intron-crossing polymerase chain reaction (PCR; Palumbi and Baker 1994) and be used to distinguish species or (if heterozygosity is high) individuals (Lessa and Applebaum

1993; Slade et al. 1993; Villablanca et al. 1998; Cipriano and Palumbi 1999).

To test the three-times rule for baleen whales, we sequenced partial actin introns from blue, fin, humpback, and minke whales and examined the resulting alleles for patterns of within and between species variation. Actin introns were amplified from baleen whales using PCR primers designed from sequences of full-length actin introns from humpback, blue, and bowhead whales (Palumbi and Baker 1994; Palumbi and Cipriano 1998). For dolphins, longer sequences were required for phylogenetic resolution, and we amplified a 1906-bp intron between exons 3 and 4 of butyrophilin (BTM) using primers designed in exon 3 (5'-GTCTCTGATGATGGGGAGTA) and exon 4 (5'-ATCTCTCCACTCTCTTGMAC) from a cow/human/mouse alignment of Genbank sequences. PCR, cloning, and sequencing were performed as previously described (Palumbi et al. 1991; Cipriano and Palumbi 1999; M. P. Hare, F. Cipriano, and S. Palumbi, unpubl. ms.).

Data Analysis

Mitochondrial and intron sequence analyses were performed using PAUP version 4.0b4a (Swofford 1993) for parsimony-based and distance-based phylogenetic analyses. For distance analyses, we used the Kimura two-parameter model to correct for multiple substitutions, although we also compared these results to results from other methods (Tamura-Nei, Jukes-Cantor, and gamma corrected distances, all implemented in PAUP 4.0b4a).

To test predictions of the three-times rule, we measured branch lengths leading to each species from a neighbor-joining tree constructed with mitochondrial sequences. These branch lengths provide an estimate of $L\mu$. Population surveys of mitochondrial sequences can be used in two ways to estimate $2N_f$. Because mitochondrial sequences coalesce at a point $2N_f$ generations in the past under neutral assumptions (Neigel and Avise 1986; Nei 1987), the maximum genetic pairwise distance in a random sample should be $4N_f\mu$. As a result, half the maximum pairwise distance can be taken as an estimate of $2N_f\mu$ (Moore 1995). Alternatively, the expectation for the average pairwise nucleotide diversity among mitochondrial sequences is also $2N_f\mu$, and this value can be used to estimate the coalescence ratio. In this study, we used the latter method to estimate $2N_f\mu$, from mitochondrial sequences compared pairwise because estimates of mitochondrial diversity are less sensitive to vagaries of sample size (when $n > 5$ or so) than are estimates of maximum mtDNA divergence (Avise 1994; Moore 1995).

To calculate the coalescence ratio, we divide the branch length leading to the most recent common ancestor of haplotype diversity within the species, by our estimate of average intraspecific nucleotide diversity ($C_R = L\mu/2N_f\mu$).

RESULTS

Molecular Tests of the Three-Times Rule

Whales

Bootstrap analysis of mtDNA control region sequences from blue, fin, humpback, and minke whales shows that all four species are monophyletic (Fig. 3). Estimates of branch

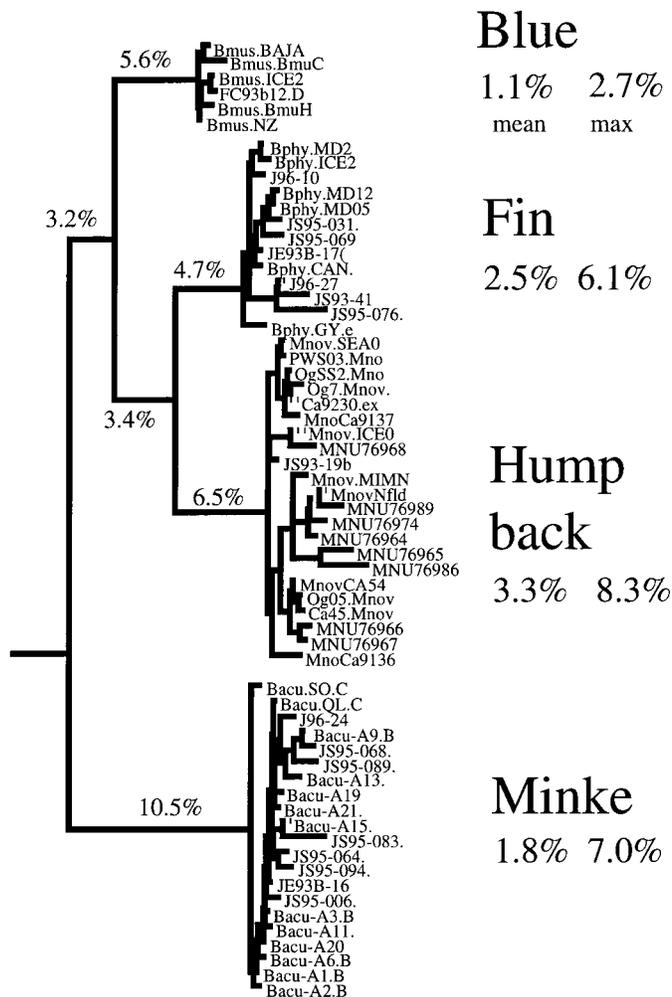


FIG. 3. Neighbor joining tree of Kimura two-parameter genetic distances for control region sequences between blue, fin, and humpback whales using sequences from bowhead whales (genus *Balaena*) as an outgroup. Genetic distance is shown on each interspecific branch, which all have bootstrap values > 0.99. For minke and humpback whales, we used all available sequences to calculate branch lengths and intraspecific diversity, but only a subset are shown here for clarity. Mean intraspecific nucleotide diversity and maximum sequence difference are shown for each species. Sample codes correspond to names of individual sequences in genbank.

lengths leading to individual species indicates minke and blue whales meet the three-times rule criterion, but for humpback and fin whales, the rule is not strongly met (Table 1).

Branch lengths are not very dependent on the type of analysis performed, with variation of less than 0.2% for interspecific branches calculated from Kimura two-parameter, Tamura-Nei, or Jukes-Cantor corrections for multiple substitutions. Maximum-likelihood trees could not be constructed with the large dataset used here, but comparisons of distances calculated from subsamples of the dataset show good correspondence with other methods. One exception is that branch lengths are 50–100% longer when a gamma-corrected genetic distance is calculated, probably because the different nucleotide positions in the d-loop region sequenced here have variable evolutionary rates (Hoelzel and Dover 1991).

Calculations of average mtDNA heterogeneity within spe-

TABLE 1. Evaluation of the three-times rule for whales species.

	Mitochondrial branch length ¹	mtDNA diversity ²	Coalescence ratio	Predicted % monophyletic ³	Nuclear monophyly observed?
Blue	5.6	1.1	5.1	85%	yes
Fin	4.7	2.5	1.9	45%	yes
Humpback	6.5	3.3	2.0	47%	yes
Minke	10.5	1.8	5.8	90%	yes
<i>Lagenorhynchus acutus</i> (N. Atlantic)	4.2	0.7	6.0	90%	yes
<i>L. obliquidens</i> (N. Pacific)	0.5	0.3	1.8	45%	no
<i>L. obscurus</i> (S. Pacific)	0.4	0.8	0.5	1%	no

¹ Number of differences per 100 bp.

² Average pairwise percent sequence difference.

³ Estimates based on equation (6.4) of Tavaré (1984) as modified in the text and shown in Figure 2B for samples of five alleles.

cies are also largely insensitive to assumptions of the analysis, with variation in mean nucleotide variability of 0.1% or less for Kimura two-parameter multiple-hit corrections with or without variation in rate among sites. Fin whales and minke showed unimodal plots of pairwise difference, whereas others (e.g., humpback whales) were distinctly bimodal, the latter probably reflecting the population structure of humpback whale populations around the world (Baker et al. 1993). Maximum genetic distance among sequences is two to three times the average distance among sequences compared pairwise (Fig. 3), slightly higher than the value predicted by strict neutral theory (Nei 1987).

The coalescence ratio (the ratio of branch length to intraspecific diversity) ranges between 1.9 and 6.0 (Table 1). At these ratios, and with a sample size of five alleles, we expect 45–85% of neutral nuclear loci to be monophyletic (Table 1). Minke whales have the highest ratio, showing over five times more divergences along the branch leading to this species than the average intraspecific diversity. Fin and humpback whales show the lowest coalescence ratios: The length of the mtDNA branch for these species with distinct global population structure is only twice as high as intraspecific diversity.

Sequences were obtained from a 310-bp portion of the first actin intron from six alleles of blue, fin, and humpback whales, as well as eight alleles from minke whales and two from the outgroup bowhead whale. Alleles were variable within and between species (Fig. 4), showing 35 variable positions and 16 sites that were informative in parsimony reconstructions. There was one deletion shared among all humpback sequences.

All whale species were monophyletic at this nuclear locus, with between one and four fixed nucleotide substitutions defining branches. Fin and blue whales are monophyletic with high bootstrap support (97% and 88%, respectively), although for humpback and minke whales, support for monophyly was more moderate (75% and 60%, respectively). All humpback sequences share a unique derived substitution at position 703 (numbering relative to Palumbi and Baker 1994; marked by an asterisk in Fig. 4) and a unique 19-bp deletion. However, there is also a reversion substitution at 855 in sam-

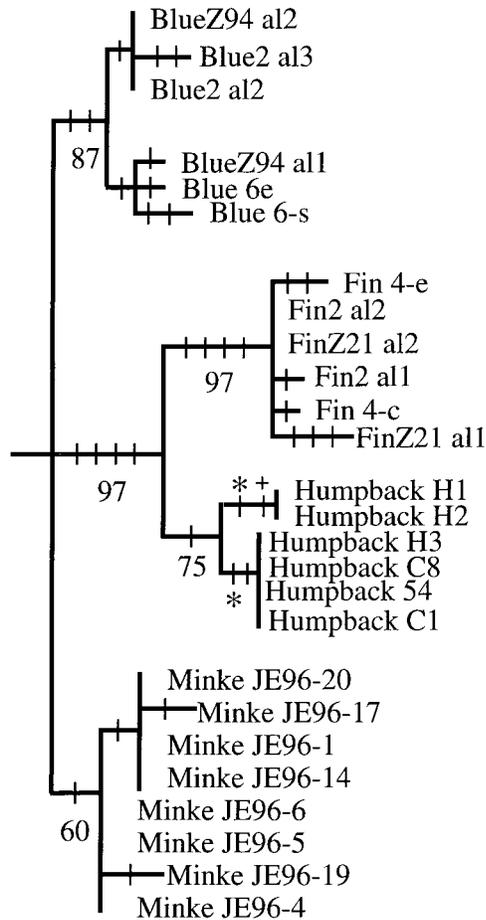


FIG. 4. Neighbor-joining phylogenetic tree of partial actin intron sequences from blue, fin, humpback, and minke whales using sequences from bowhead whales as an outgroup. Humpback sequences are from Palumbi and Baker (1994), fin and blue whale sequences are from Cipriano and Palumbi (1999), and minke whale sequences are deposited in Genbank accession nos. AF372134–AF372159). Tic marks on branches represent the number of unique substitutions placed by parsimony analysis along those branches. Values beneath branches represent percent bootstrap support. The asterisk denotes a nucleotide position that is hypothesized to have changed in parallel in the two humpback lineages shown. Alternatively, an equally parsimonious reconstruction involves a reversion at the position marked by a plus sign to a nucleotide state shared with the outgroup, bowhead whales.

ples H1 and H2 (marked by a plus sign) that causes maximum-parsimony procedures to occasionally group these alleles outside of the main humpback cluster, thus reducing support for monophyly.

Dolphins

Cytochrome *b* data, based on 496 bp from 14 individuals, show that the Atlantic species *Lagenorhynchus acutus* is monophyletic with respect to outgroups in other genera and to two Pacific species (bootstrap 100%), but that the two Pacific species display only a small degree of genetic divergence from one another (Fig. 5). Sequences from the Southern Hemisphere species *L. obscurus* are clustered by three nucleotide substitutions found nowhere else in the dataset, and bootstrap analyses based on parsimony or distance criteria

group these sequences together in a monophyletic cluster 80% of the time. Sequences of the northern Pacific white-sided dolphin, *L. obliquidens*, display only a single shared derived substitution, clustering together in bootstrap analyses about 60% of the time.

The cytochrome *b* branch leading to the Atlantic species is long, showing approximately 4.2% sequence change from outgroups that include porpoises and other delphinid dolphins, whereas branches leading from the common ancestor of the two Pacific species are both short, representing 0.4% and 0.5% sequence evolution, respectively (Fig. 5). Analyses based on Kimura two-parameter, Tamura-Nei, and gamma-corrected distance analysis gave similar results, except that the gamma-corrected analysis gave similar results, except that the gamma-corrected analysis was 50% longer. Sequences of d-loops from a limited number of these individuals showed similar phylogenetic and distance patterns (Cipriano 1997). Coalescence ratios vary from 0.5 to 6.0 in these three species (Table 1), and the predicted proportion of monophyletic loci varies from 1% to 90%.

The result from 1904 bp alleles of butyrophilin intron I is quite different. Two sequences from the North Atlantic species differ at 24 positions from the Pacific species (Fig. 5), but alleles from the two Pacific species are clearly polyphyletic. Two major clusters of alleles are apparent in phylogenetic reconstructions, and sequences from both species appear in both clusters (Fig. 5). The shortest tree relating sequences of the Pacific species is 11 substitutions in length (Fig. 5), resolving the sequences into two species-specific monophyletic clusters adds four more steps: two reversals and two parallel substitutions.

DISCUSSION

The Three-Times Rule

For neutral loci and a panmictic population, nuclear coalescence is expected when the mitochondrial branch leading to a species clade is three times longer than the average mitochondrial diversity within the species. Although this expectation is a simple consequence of the transmission genetics of organelle and nuclear genes, estimating the probability of nuclear coalescence based on mitochondrial or chloroplast data has never been formalized. By modifying equations that predict the proportion of nuclear loci that coalesce over time, we show how the coalescence ratio can be used to estimate the probability of nuclear monophyly in a species (Fig. 2). As proposed, the three-times rule is merely a heuristic guide to general coalescence expectations, much like the well-known rule that gene flow of more than one individual per generation prevents evolutionary divergence (Slatkin 1987). Like the gene flow rule, the strongest nuclear gene tree expectations will derive from mitochondrial coalescence ratios much larger than three or much smaller than three. Under these conditions, nuclear monophyly is predicted to be frequent or rare, respectively. At intermediate coalescence ratios, nuclear monophyly will be seen at some but not all neutral loci, because coalescence occurs stochastically and some nuclear loci coalesce faster than others by chance.

Data from seven species of whales and dolphins confirm the general predictions of this approach. We observed recip-

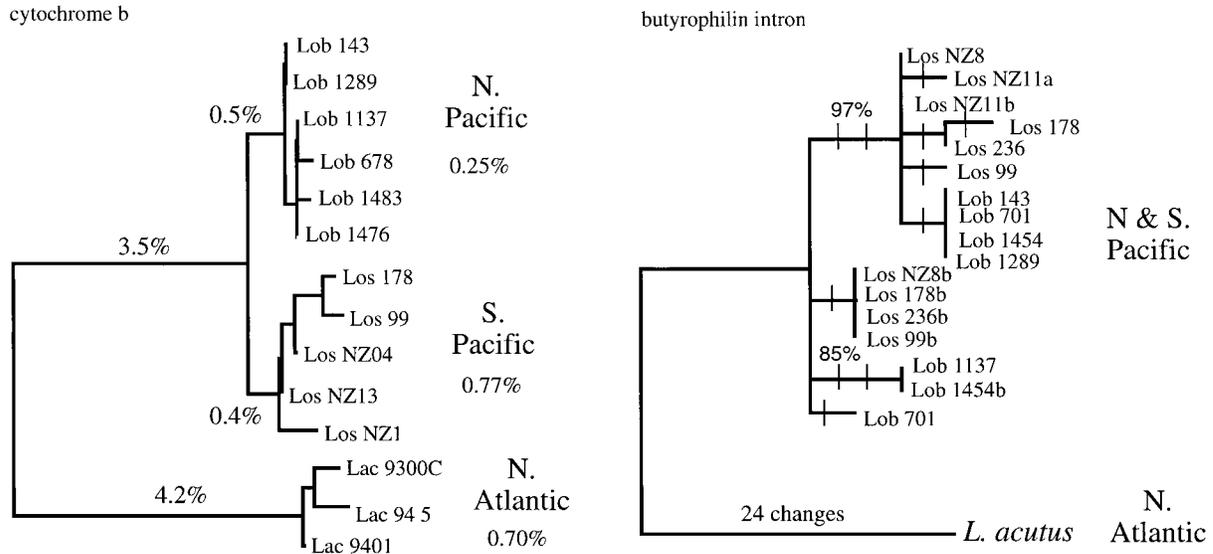


FIG. 5. Mitochondrial and nuclear phylogenetic patterns for three species of dolphins. Mitochondrial cytochrome *b* data (left) show strong monophyly of the North Atlantic species *Lagenorhynchus acutus* (Lac), but only small net divergence of the Pacific species *L. obscurus* (Los, S. Pacific) and *L. obliquidens* (Lob, N. Pacific). Branch lengths represent genetic distances calculated with Kimura two-parameter corrections used in a neighbor-joining tree. Intron data from the nuclear locus butyrophilin shows alleles from *L. acutus* clustering together ($n = 2$), but alleles from the other two species in a polyphyletic assemblage.

rocal monophyly at nuclear loci in species with high coalescence ratios, whereas species with low ratios tended to show polyphyly. In three species with coalescence ratios of 5.0–6.0, nuclear introns (actin or butyrophilin) were monophyletic. At the other extreme, in one species with an exceptionally small ratio of 0.5, the nuclear data showed polyphyly. Moderate coalescence ratios, as expected, gave variable results. Among three comparisons with coalescence ratios of about 2.0, nuclear introns were polyphyletic in one case and monophyletic in two cases (Table 1).

A Spectrum of Coalescence Ratios

A full test of the three-times rule should involve coalescence comparisons of multiple loci for single species (M. P. Hare, F. Cipriano, and S. Palumbi, unpubl. ms.), but the results reported here suggest that species may fall along a spectrum of coalescence ratios allowing strong predictions in some cases. When species have high coalescence ratios (above 10), then 95% of nuclear loci are predicted to be monophyletic. For such cases, observation of polyphyletic loci could signal the action of balancing selection or other evolutionary processes that act to slow coalescence. A well-known case is the transspecies phylogenetic pattern of major histocompatibility complex (MHC) alleles in primates (Takahata 1990). Even though humans have low mitochondrial diversity and a coalescence ratio of about 9.0, MHC alleles are polyphyletic with respect to chimpanzees and gorillas. This pattern is thought to be maintained by selection for MHC heterozygosity (Hughes and Nei 1988).

A second example occurs in the marine birds known as murrelets. Three species, *Brachyramphus brevirostris*, *B. marmoratus marmoratus*, and *B. marmoratus perdix*, are monophyletic with respect to each other on the basis of cytochrome *b* sequences (Friesen et al. 1996; Kidd and Friesen

1998). Estimates of genetic distances within and between species show coalescence ratios > 10 for both *B. m. perdix* and *B. m. marmoratus*, but only 1.77 for *B. brevirostris*. From these results, the three-times rule correctly predicts nuclear monophyly within *B. m. perdix* and polyphyly in *B. brevirostris*. However, *B. m. marmoratus* is not monophyletic at any of four nuclear loci examined. In this case, failure of the three-times rule for *B. m. marmoratus* could reflect past introgression of nuclear genes through hybridization (Friesen et al. 1997).

At the other end of the spectrum, low coalescence ratios predict nuclear polyphyly across the nuclear genome. In such instances, observation of strong monophyly at a locus may signal the action of directional selection. For example, three closely related tropical sea urchins in the genus *Echinometra* have mitochondrial coalescence ratios of only 0.0–1.0, and mtDNAs from some species are not monophyletic with respect to others (Palumbi et al. 1997). As predicted, most nuclear loci are polyphyletic in these species (S. R. Palumbi, unpubl. data). Nevertheless, these species are reciprocally monophyletic at the nuclear locus for the gamete-recognition protein *bindin* (Metz and Palumbi 1996). These data suggest that selection on the *bindin* locus has accelerated lineage sorting within species. Such selection is evident in the enhanced replacement substitutions relative to silent substitutions in *bindin* (Metz and Palumbi 1996), which may be an important part of the mechanisms of reproductive isolation in this group. Identification of positive selection at loci is sometimes problematic (Clark and Civetta 2000), and coalescence patterns, because they integrate over a long time scale and are sensitive to slight differences in selection (Tajima 1983; Takahata 1990), may provide a more general method for detecting selection between closely related species (Ting et al. 2000).

Strong predictions of polyphyly in closely related species may be a powerful tool in the identification of other loci related to species differentiation. For example, few nuclear loci should be monophyletic by chance in species of *Echinometra*, and thus any loci discovered that are monophyletic represent good candidates for the action of directional selection distinguishing species. This approach may provide a search criterion for candidate loci involved in speciation in closely related taxa for which, unlike *Drosophila*, defined introgressed lines cannot easily be created (Zeng and Singh 1995).

Another example derives from recent work on multilocus differentiation in *Drosophila*, showing how strong three-times rule predictions may clarify conflicting patterns of nuclear coalescence. For example, among *Drosophila* nuclear loci show conflicting patterns of separation between the closely related species *D. pseudoobscura pseudoobscura* and *D. pseudoobscura bogotana*. At the *Adh* locus there are no fixed differences and many shared polymorphisms, whereas at the *period* locus there are six fixed differences and one shared polymorphism (Wang et al. 1997). At moderate coalescence ratios, this variation in nuclear phylogenetic patterns would be expected and would not require any explanation except different rates of lineage sorting at different loci. If the coalescence ratios between these two species were very high, then unexpected *Adh* polyphyly might indicate introgression (Wang et al. 1997). However, mitochondrial data suggest *D. p. bogotana* is very closely related to *D. p. pseudoobscura*, and mtDNA sequences are either paraphyletic or polyphyletic (M. Noor, pers. comm.). At such low coalescence ratios, it is the *period* monophyly that is unexpected, suggesting selection at this locus is operating between species to accelerate monophyly. A similar result has recently been described for the *Odysseus* gene in *D. simulans* and *D. mauritiana* (Ting et al. 2000).

In the middle of the coalescence spectrum, fewer strong conclusions can be drawn because individual loci may show just about any phylogenetic pattern. For example, when coalescence ratios vary between 0.75 and 6.00, then 10–90% of nuclear loci are predicted to be monophyletic. Thus, our observation of actin monophyly in fin and humpback whales is consistent with the neutral expectation that 25–50% of loci should be monophyletic for the sample sizes we report here, but our power to detect any substantial departure from neutral expectations is very low. In such cases, analysis of a large number of loci is needed to precisely test the predictions of the three-times rule.

Failure of Three-Times Rule Assumptions

Three-times rule predictions are only strong if the relationship between coalescence ratio and nuclear monophyly generally follows neutral predictions. Other than selection at nuclear loci, possible reasons for failure of the three-times rule to explain nuclear patterns include selection on mtDNA (Nachman et al. 1996; Templeton 1996), differential introgression of nuclear versus mitochondrial genes between species (Avice 1994), and chimeric recombination between nuclear loci (Cheung et al. 1999). In addition, if ancestral populations were much larger and more genetically diverse than

the ones studied now, a significant fraction of the divergence between mtDNA clades might have arisen before species formation (Wakeley and Hey 1997), and this will deflate the coalescence ratio estimate. Males and females can have different effective population sizes (Birky 1991), altering the expected relationship between nuclear and mitochondrial coalescence (Hoelzer 1997). Finally, the observed mitochondrial tree for a given species reflects organelle genome coalescence in the past. Estimates of this process and of the shape of the mitochondrial tree are subject to sampling error: Intraspecific diversity is likely to be underestimated and branch length overestimated by small mitochondrial sample sizes.

Sampling of nuclear alleles also impacts coalescence predictions. If nuclear alleles are sampled from relatively few distinct populations, then nuclear monophyly may be observed more often than predicted. For example, in the fin whale mitochondrial dataset, three individuals from Japanese markets (J96–27, JS93–41, JS95–076) stand out in a divergent clade, 4–6% different from the other sequences. We do not have actin data from these samples, but if they represent a distinct subpopulation, they may possess actin introns distinct from other fin whales. A possibility that requires theoretical treatment is whether comparisons of mitochondrial and nuclear coalescence patterns should take into account the population structure apparent at mitochondrial versus nuclear loci.

Further examination of the three-times rule are needed to test whether it provides a general backdrop for understanding nuclear monophyly. Datasets including multiple, polymorphic nuclear loci will be particularly crucial. A major focus could be how often mitochondrial-nuclear comparisons show patterns predicted by simple neutral expectations and under what conditions complicating factors like hybridization and selection arise. Our knowledge of lineage sorting, ancestral population sizes, genomic selection, and the evolution of monophyly will be incomplete until we make appropriate comparisons across independent loci and empirically describe the extent of genealogical discordance within species trees. Toward this end, the three-times rule allows the large databases already available on mtDNA and cpDNA variation to be used to provide first-order predictions about the behavior of nuclear loci in these species.

ACKNOWLEDGMENTS

We thank N. Davies, S. France, S. Lavery, and F. Villablanca for discussion and criticism of an earlier draft. J. Wakeley led us to Tavaré's equation, helped us develop quantitative predictions, and pointed out the importance of ancestral polymorphism. A. Palumbi produced the curve in Fig. 2 from equation (1). We also thank C. S. Baker, R. Brownell, A. Dizon, A. Helden, G. Lento, L. Medrano, J. Reyes, and L. Siemann for tissue samples or unpublished reference sequences. Monitoring of whale meat markets was supported by Earthtrust and the Pew Charitable Trusts. Development of nuclear markers and the theory of species identifications was supported by the National Science Foundation.

LITERATURE CITED

- Avise, J. C. 1994. Molecular markers, natural history, and evolution. Chapman and Hall, New York.
- Avise, J. C., and R. M. Ball, Jr. 1990. Principles of genealogical concordance in species concepts and biological taxonomy. *Oxf. Surv. Evol. Biol.* 7:45–67.
- Baker, C. S., and S. R. Palumbi. 1994. Which whales are hunted? A molecular genetic approach to monitoring whaling. *Science* 265:1538–1539.
- Baker, C. S., A. Perry, J. L. Bannister, M. T. Weinrich, R. B. Abernethy, J. Calambokidis, J. Lien, R. H. Lambertsen, J. Urban-Ramirez, O. Vasquez, P. J. Clapham, A. Alling, S. J. O'Brien, and S. R. Palumbi. 1993. Abundant mitochondrial DNA variation and world-wide population structure in humpback whales. *Proc. Natl. Acad. Sci. USA* 90:8239–8243.
- Baker, C. S., F. Cipriano, and S. R. Palumbi. 1996. Molecular genetic identification of whale and dolphin products from commercial markets in Korea and Japan. *Mol. Ecol.* 5:671–685.
- Baum, D. A., and K. L. Shaw. 1995. Genealogical perspectives on the species problem. Pp. 289–303 in P. C. Hoch, and A. G. Stephenson, eds. *Experimental and molecular approaches to plant biosystematics*. Missouri Botanical Garden, St. Louis, MO.
- Begun, D. J., and C. F. Aquadro. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *Drosophila melanogaster*. *Nature* 356:519–520.
- Birky, C. W. J. 1991. Evolution and population genetics of organelle genes: mechanisms and models. Pp. 112–134 in R. K. Selander, A. G. Clark, and T. S. Whittam, eds. *Evolution at the molecular level*. Sinauer Associates, Sunderland, MA.
- Birky, C. W., 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.
- Carson, H. L., and A. R. Templeton. 1984. Genetic revolutions in relation to species phenomena: the founding of new populations. *Annu. Rev. Ecol. Syst.* 15:97–131.
- Cheung, B., R. S. Holmes, S. Esté, and I. R. Beacham. 1999. Evolution of Class I alcohol dehydrogenase genes in Catarrhine primates: gene conversion, substitution rates, and gene regulation. *Mol. Biol. Evol.* 16:23–36.
- Cipriano, F. 1997. Antitropical distributions and speciation in dolphins of the genus *Lagenorhynchus*: a preliminary analysis. Pp. 305–316 in A. E. Dizon, S. J. Chivers, and W. F. Perrin, eds. *Molecular genetics of marine mammals*. Society for Marine Mammalogy, New York.
- Cipriano, F., and S. R. Palumbi. 1999. Genetic tracking of a protected whale. *Nature* 397:307–308.
- Clark, A. G., and A. Civetta. 2000. Protamine wars. *Nature* 403:261–263.
- Cohen, A. 1997. Sturgeon poaching and black market caviar: a case study. *Environ. Biol. Fishes* 48:423–426.
- Davey, H. W., S. L. Ogg, Y. Husaini, R. G. Snell, I. V. Korobko, I. H. Mather, and R. Wilkins. 1997. Structure and sequence of the bovine butyrophilin gene. *Gene* 199:57–62.
- Friesen, V. L., J. F. Piatt, and A. J. Baker. 1996. Evidence from cytochrome b sequences and allozymes for a “new” species of alcid: the long-billed murrelet (*Brachyramphus perdix*). *Condor* 98:681–690.
- Friesen, V. L., B. C. Congdon, H. E. Walsh, and T. P. Birt. 1997. Intron variation in marbled murrelets detected using analyses of single-stranded conformational polymorphisms. *Mol. Ecol.* 6:1047–1058.
- Hoelzel, A. R., and G. A. Dover. 1991. Evolution of the cetacean mitochondrial D-loop region. *Mol. Biol. Evol.* 8:475–493.
- Hoelzer, G. A. 1997. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees revisited. *Evolution* 51:622–626.
- Hudson, R. R. 1990. Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* 7:1–44.
- . 1992. Gene trees, species trees, and the segregation of ancestral alleles. *Genetics* 131:509–512.
- Hudson, R. R., and N. Kaplan. 1994. Gene trees with background selection. Pp. 140–153 in B. Golding, ed. *Non-neutral Evolution: Theories and molecular data*. Chapman and Hall New York.
- Hughes, A. L., and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex loci reveals over dominant selection. *Nature* 335:167–170.
- Jagdish, V., J. Robertson, and A. Gibbs. 1996. RAPD analysis distinguishes *Cannabis sativa* samples from different sources. *Forensic Science International* 79:113–121.
- Kidd, M. G., and V. L. Friesen. 1998. Sequence variation in the Guillemot (Alcidae: *Cephus*) mitochondrial control region and its nuclear homologue. *Mol. Biol. Evol.* 15:61–70.
- Kimura, M. 1983. *The neutral theory of molecular evolution*. Cambridge Univ. Press, Cambridge, U.K.
- Kimura, M., and T. Ohta. 1969. The average number of generations until fixation of a mutant gene in a finite population. *Genetics* 61:763–771.
- Kreitman, M. 1991. Detecting selection at the level of DNA. Pp. 204–221 in R. K. Selander, A. G. Clark, and T. S. Whittam, eds. *Evolution at the molecular level*. Sinauer Associates, Sunderland, MA.
- Kuhner, M., J. Yamato, and J. Felsenstein. 1995. Estimation of effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics* 140:1421–1430.
- Lessa, E., and G. Applebaum. 1993. Screening techniques for detecting allelic variation in DNA sequences. *Mol. Ecol.* 2:119–129.
- Li, W.-H. 1997. *Molecular evolution*. Sinauer Associates, Sunderland, MA.
- Malik, S., P. J. Wilson, R. J. Smith, D. M. Lavigne, and B. N. White. 1997. Pinniped penises in trade: a molecular-genetic investigation. *Conserv. Biol.* 11:1365–1374.
- Maynard Smith, J. 1987. On the equality of origin and fixation times in genetics. *J. Theor. Biol.* 128:247–252.
- Metz, E. C., and S. R. Palumbi. 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Mol. Biol. Evol.* 13:391–406.
- Moore, W. S. 1995. Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 49:718–726.
- Moritz, C. 1994. Defining “evolutionary significant units” for conservation. *Trends Ecol. Evol.* 9:373–375.
- Nachman, M. W., W. M. Brown, M. Stoneking, and C. F. Aquadro. 1996. Nonneutral mitochondrial DNA variation in humans and chimpanzees. *Genetics* 142:953–963.
- Nei, M. 1987. *Molecular evolutionary genetics*. Columbia Univ. 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.
- Palumbi, S. R. 1996. What can molecular genetics contribute to marine biogeography? An urchin’s tale. *J. Exp. Mar. Biol. Ecol.* 203:75–92.
- Palumbi, S. R., and C. S. Baker. 1994. Contrasting populations structure from nuclear intron sequences and mtDNA or humpback whales. *Mol. Biol. Evol.* 11:426–435.
- . 1996. Molecular genetic approaches in conservation. Pp. 25–37 in T. B. Smith, and R. K. Wayne, eds. *Molecular genetic approaches in conservation*. Oxford Univ. Press, New York.
- Palumbi, S. R., and F. Cipriano. 1998. Species identification using genetic tools: the value of nuclear and mitochondrial gene sequences in whale conservation. *J. Hered.* 89:459–464.
- Palumbi, S. R., G. Grabowsky, T. Duda, N. Tachino, and L. Geyer. 1997. Speciation and the evolution of population structure in tropical Pacific sea urchins. *Evolution* 51:1506–1517.
- Palumbi, S. R., A. P. Martin, S. Romano, W. O. McMillan, L. Stice, and G. Grabowski. 1991. *The simple fool’s guide to PCR*. Department of Zoology Special Publication, University of Hawaii, Honolulu, HI.
- Rehbein, H., G. Kress, and T. Schmidt. 1997. Application of PCR-SSCP to species identification of fishery products. *J. Sci. Food Agric.* 74:35–41.
- 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. 1987. Gene flow and the geographic structure of natural populations. *Science* 236:787–792.
- Swofford, D. L. 1993. *Phylogenetic Analysis Using Parsimony (PAUP)*, vers. 3.1.1. Illinois Natural History Survey, Champaign, IL.
- Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–460.
- Takahata, N. 1990. A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proc. Natl. Acad. Sci.* 87:2419–2423.
- . 1995. A genetic perspective on the origin and history of humans. *Annu. Rev. Ecol. Syst.* 26:343–372.
- Tavaré, S. 1984. Line-of-descent and genealogical processes and their applications in populations genetic models. *Theor. Popul. Biol.* 26:199–164.
- Templeton, A. R. 1996. Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the hominoid primates. *Genetics* 144:1263–1270.
- Ting, C.-T., S.-C. Tsaur, and C.-I. Wu. 2000. The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. *Proc. Natl. Acad. Sci. USA* 97: 5313–5316.
- Villablanca, F. X., G. Roderick, and S. R. Palumbi. 1998. Invasion genetics of the Mediterranean fruit fly: variation in multiple nuclear introns. *Mol. Ecol.* 7:547–560.
- Wakeley, J., and J. Hey. 1997. Estimating ancestral population parameters. *Genetics* 145:847–855.
- Wang, R. L., J. Wakeley, and J. Hey. 1997. Gene flow and natural selection in the origin of *Drosophila psweudoobscura* and close relatives. *Genetics* 147:1091–1106.
- Zeng, L.-W., and R. A. Singh. 1995. A general method for identifying major hybrid male sterility genes in *Drosophila*. *Heredity* 75:331–341.

Corresponding Editor: J. Neigel