

# Principles of genealogical concordance in species concepts and biological taxonomy

JOHN C. AVISE and  
R. MARTIN BALL, JR.

## 1. INTRODUCTION

For more than 50 years, the 'biological species concept' (BSC) has been a major theoretical framework orienting research on the origins of evolutionary diversity (Dobzhansky 1937). Under the BSC, species are 'groups of interbreeding natural populations that are reproductively isolated from other such groups' (Mayr 1970, p. 12). Numerous authors have expressed sentiments on the BSC similar to those of Ayala (1976, p. 18):

among cladogenetic processes, the most decisive one is speciation – the process by which one species splits into two or more. . . Species are, therefore, independent evolutionary units. Adaptive changes occurring in an individual or population may be extended to all members of the species by natural selection; they cannot, however, be passed on to different species.

Thus under the BSC, species are perceived as biological and evolutionary entities that are more meaningful and less arbitrary than other taxonomic categories such as subspecies or genera (Dobzhansky 1970). The BSC has served to focus attention on questions concerning the evolution of intrinsic reproductive barriers (RBs), including: What genetic changes produced RBs, and hence new species? What morphological, developmental or behavioral traits are involved? What ecological, demographic or evolutionary conditions favor RB evolution?

As judged by its continued widespread employment in textbooks and as a guide to research strategy (eg. Coyne and Orr 1989), the BSC appears to have survived a variety of criticisms (both philosophical and operational) leveled against it over the last 30 years (Ehrlich 1961; Ehrlich and Raven 1969; Levin 1979; Raven 1976; Sokal and Crovello 1970; Sokal 1973; Wiley 1978). Recently, another serious challenge has come from some systematically-oriented evolutionists who argue that the BSC lacks a sufficient phylogenetic perspective, and hence provides an inappropriate guide to the origins and products of evolutionary diversification (Cracraft 1983; de Queiroz and Donoghue 1988; Donoghue 1985; Eldredge and

Cracraft 1980; McKittrick and Zink 1988; Mishler and Donoghue 1982; Nelson and Platnick 1981; Rosen 1979). Many of these critics of the BSC argue that 'reproductive isolation should not be a part of species concepts' (McKittrick and Zinc 1988, p. 3). This has led to another call for abandonment of the BSC, and its replacement by the 'phylogenetic species concept' (PSC).

We believe there is much of value in the PSC, but that some of its proponents go too far in suggesting a total abandonment of the BSC. The purpose of this chapter is to introduce another conceptual approach for taxonomic and species recognition based on principles of genealogical concordance defined below. These principles derive most easily from theories and observations in molecular evolution, but can also be applied to hereditary morphological, behavioral and other phenotypic attributes traditionally studied by systematists. Concepts of genealogical concordance in taxonomic recognition combine what we perceive to be the better elements of the PSC and BSC.

## 2. SYNOPSIS AND CRITIQUE OF THE PHYLOGENETIC SPECIES CONCEPT

Under the PSC, a species has been defined as a monophyletic group composed of 'the smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent' (Cracraft 1983, p. 170). Cracraft intentionally avoids explicit reference to reproductive disjunctions, and instead focuses directly on the distributions of diagnostic, heritable trait(s). As emphasized by Cracraft (1983, p. 170), assemblages constituting a phylogenetic species 'simply must be diagnosable from all other species'. Each successful diagnosis is taken to indicate a phylogenetic separation – the key concept underlying the phylogenetic species. The PSC has several suggested advantages over the BSC, including the following (Cracraft 1983): (a) assemblages diagnosed under the PSC are phylogenetic units, and as such provide more informative subjects for evolutionary and ecological study; (b) the PSC eliminates the need for direct concern with reproductive compatibilities (that in nature are often difficult to observe directly, and normally cannot be assessed among allopatric forms); and (c) attention is focused on the geographic and genealogical histories of populations.

We are sympathetic to the general goal of the PSC of calling for greater attention to historical relationships of populations (Avise 1989a; Avise *et al.* 1987a). However, current formulations of the PSC (e.g. Cracraft 1983; McKittrick and Zink 1988) have limitations as a replacement for the BSC:

1. *The number of species recognized under the PSC depends on the*

*resolving power of the analytical tools available.* The first clause of the PSC definition, 'the smallest diagnosable cluster of individual organisms', indicates that any diagnostic trait uniting an array of individuals to the exclusion of others is sufficient to define a phylogenetic species (provided also that the trait is inherited and monophyletic). Some authors (e.g. Cracraft 1983) suggest that the diagnostic trait may be either primitive or derived, whereas others (McKittrick and Zink 1988) restrict consideration to derived traits only. However, there is considerable sentiment that a single diagnostic trait, no matter how 'trivial', is sufficient for clade definition (Wiley 1981), and hence for species recognition under the PSC. For example, McKittrick and Zink (1988, pp. 9–10) contemplate the prospect of finding a group of birds with one extra hooklet on the barb of the seventh primary feather, and conclude 'There is no theory to suggest that a trait must be of a certain quality or magnitude to provide historical information or to delimit species. . . . Thus, no character is potentially more or less useful as a tool to reconstruct patterns of speciation (*sensu* the PSC).'

What would be the consequences of recognizing a distinct species for each diagnostic trait? Cracraft (1983, p. 173) suspects that within ornithology 'the phylogenetic species concept should not increase the number of taxa already recognized' (although many subspecies would be elevated to species status). McKittrick and Zink (1988, p. 9) suspect that closer scrutiny of biochemical, morphological and other characters on a microgeographic scale will likely reveal the existence of many more phylogenetic avian taxa, but that 'The notion that there should be an upper limit to the number of species described does not appear to have any value, heuristic or otherwise.'

We believe that such statements concerning the PSC overlook the huge size and extreme variability of eukaryotic genomes. Evidence from molecular biology demonstrates enormous genetic polymorphism within most taxa. For example, nucleotide diversity (a measure of heterozygosity at the nucleotide level) ranges from about 0.002 to 0.019 for various loci (Nei 1987, p. 267). Because a typical gene consists of several thousand nucleotide pairs, randomly drawn haplotypes from conspecific individuals will normally differ in nucleotide sequence. Even cursory assays of restriction-fragment-length-polymorphisms (RFLPs) of particular genes such as mitochondrial DNA (mtDNA) or minisatellite nuclear DNA have revealed extensive intraspecific genetic heterogeneity. Most recognized biological species are already divisible into large numbers of diagnosable subunits (often geographically subdivided: Avise *et al.* 1987a), and in some taxa nearly every organism can be distinguished with the limited genetic information already at hand (Avise *et al.* 1989; Burke 1989). The data from direct nucleotide sequencing, and from multiple loci, will make such levels of genetic distinction commonplace (e.g. Jeffreys *et al.* 1985;

Kocher *et al.* 1989; Lander 1989). If each individual organism is genetically unique at a high level of resolution, then the grouping of individuals requires that we ignore distinctions that occur below some arbitrary threshold. The evolutionary significance of any such threshold must surely be questionable.

2. *Unless persistent extrinsic (geographic) or intrinsic RBs are present, different gene genealogies will usually disagree in the boundaries for 'species' under the PSC.* A strict application of the PSC definition is also difficult to apply given the vast numbers of gene genealogies and their expected idiosyncratic distributions within and among population pedigrees. The nuclear genome of most birds and mammals, for example, consists of about 2–3 billion nucleotide pairs. Under a conventional mutation rate of  $\mu = 10^{-9}$  per nucleotide site per generation (Nei 1987), a typical individual is likely to carry 2–3 newly arisen mutations, and a species composed of even a few million animals would be expected to carry several million new mutations every generation. Under a reasonable population demography (Poisson distribution of surviving progeny with mean 2 per family), nearly two-thirds of new neutral mutants are expected to survive for at least one generation, and nearly 2 per cent will likely survive beyond 100 generations (Spiess 1977, pp. 376–7). Each new mutation that survives (a derived trait) will have its own particular geographic and population distribution, depending on such factors as its place of origin and age, the fitness conferred on its bearers, and the historical gene flow regime of the species (Fig. 1). Except in the special evolutionary circumstances discussed beyond (which involve intrinsic or extrinsic RBs), little or no concordance should exist among the assemblages of individuals diagnosable with independently-derived mutations or their derivatives: non-overlaps, partial overlaps, or nested arrangements in group membership should typically characterize various organismal assemblages diagnosed by independent genetic traits (Fig. 1). Due to inherently stochastic aspects of the hereditary process – mutational origins, allelic segregation during meiosis and the vagaries of population demography – alleles at each genetic locus that have ‘trickled’ through an organismal pedigree will exhibit a unique phylogenetic tracing (Ball *et al.* 1990).

3. *Shared ancestry in sexually reproducing organisms implies historical membership in a reproductive community.* The PSC could clearly be applied to the identification of ‘species’ in asexually reproducing organisms. However, the second clause of Cracraft’s (1983) PSC definition – parental pattern of ancestry and descent – was intended to extend the PSC to sexually reproducing organisms. Species then constitute a phylogenetic ‘lineage’ within which matings and successful reproduction have taken place (McKittrick and Zink 1988). Thus when the PSC is applied to

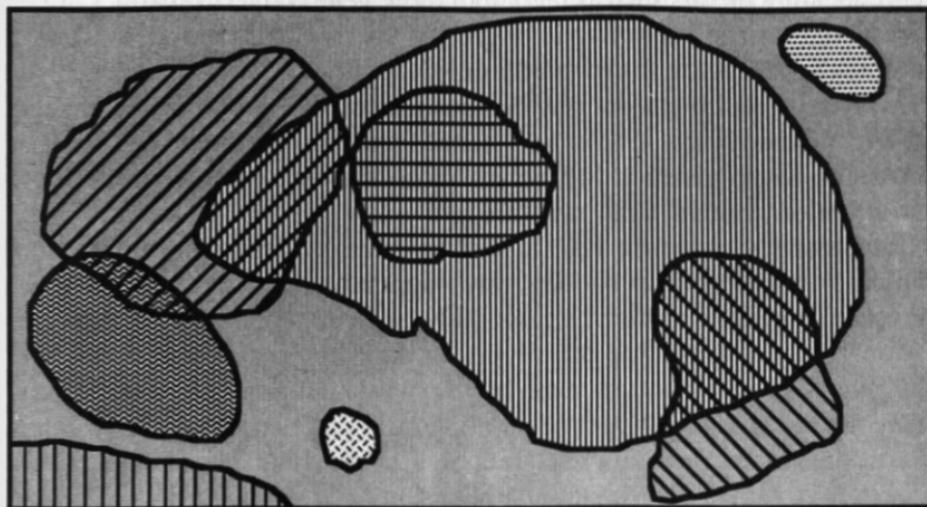


Fig. 1. Examples of possible geographic distributions of various derived mutations in a continuously distributed species with limited dispersal capability (compared to the geographic range occupied).

sexually reproducing forms, historical reproductive communication and continuity implicitly accounts for shared possessions of heritable diagnostic traits. In other words, 'species' under the PSC are recognizable precisely because their members derive from reproductive communities of individuals (a general philosophy that the BSC has always emphasized).

Thus overall, if a broader framework of the PSC is to contribute to a significant advance in systematic practice (and we believe it can), a shift from issues of diagnosability to those of magnitudes and patterns of phylogenetic differentiation (and of the historical and reproductive reasons for such patterns) will be required. A powerful approach to these issues should involve reference to the following principles of genealogical concordance.

### 3. GENEALOGICAL CONCORDANCE

The extant haplotypes (DNA sequences) present in any 'species' represent the gene lineages that have survived through an organismal pedigree. Within any pedigree, such lineage tracings (gene phylogenies) can differ greatly from locus to locus (Ball *et al.* 1990), due to the vagaries of meiotic segregation, mating pattern and the reproductive success of individuals through which the alleles were transmitted. Such differences among gene phylogenies within an organismal pedigree arise inevitably, even when all loci experience nucleotide substitutions at the same rate, and when

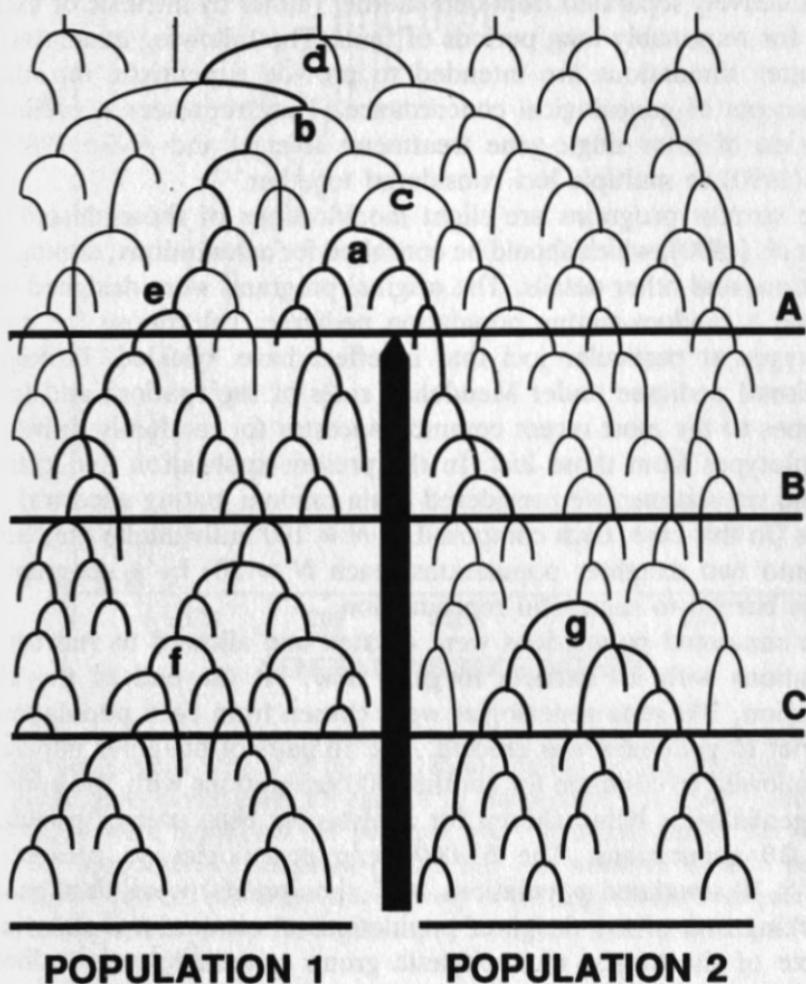
complicating factors such as recombination, gene conversion and 'sampling error' due to the idiosyncratic origins of the particular mutations assayed are neglected. A clear distinction must therefore be drawn between gene phylogenies and organism phylogenies (Avise 1989a; Nei 1987; Takahata 1989; Wilson *et al.* 1985), and hence between phylogenetic diagnoses based on single genetic traits versus those based on broader trends in the information content from multiple loci.

Suppose, as shown in Fig. 2, that an ancestral, random mating population is sundered into two daughter populations, through either a geographic or genetic barrier to gene flow. Immediately following this separation, at any locus some haplotypes in daughter population 1 will likely be genealogically closer to some haplotypes in daughter population 2 than they are to other haplotypes in population 1 (and vice versa). Phylogenetic partitions based on traits encoded by a single gene would therefore be discordant with the population subdivision. The status of particular gene phylogenies in these daughter populations changes through time due to demographically based processes of lineage sorting, until eventually all remaining haplotypes in population 1 are genealogically closer to others in population 1 than to any haplotypes in population 2, and vice versa (Neigel and Avise 1986; Pamilo and Nei 1988; Tajima 1983). The rate of the process is demography-dependent, but commonly takes about  $2N_e - 4N_e$  generations, where  $N_e$  is the effective size of the daughter populations. In other words, in terms of any gene phylogeny, two isolated daughter populations are expected to evolve from conditions of initial polyphyly, through paraphyly, and eventually to a state of reciprocal monophyly (Neigel and Avise 1986), at which point the major phylogenetic subdivisions in the gene genealogy become coincident with the major population-level subdivisions (as defined by the intrinsic or extrinsic barriers to reproduction). Secondary admixture and introgressive hybridization between the two populations could, of course, blur the evidence for this historical separation.

Thus in taxonomic recognition, the operational challenge involves an assessment of when the arrays of individuals grouped by particular genetic trait(s) coincide with the historical, organismal-level partitions that are seldom observable directly. Under what biological or evolutionary conditions should the biotic partitions registered by various genetic traits faithfully mirror the phylogenetic separations of the taxa that we might wish to recognize formally?

### 3.1 Gene-gene phylogenetic concordances

One important consideration must be whether many independent gene phylogenies (those from unlinked and non-epistatic loci) provide *concordant* support for the organismal assemblages identified. As shown below,



*Fig. 2.* Schematic presentation of the distribution of haplotype lineages (at a single gene) through an ancestral population subdivided (at time A) by a geographic or other barrier to reproduction. With respect to this particular gene genealogy, between levels A and B the daughter populations 1 and 2 are polyphyletic, i.e. at any point in that time interval, some individuals in population 1 are genealogically closer to some individuals in population 2 than to other individuals in population 1, and vice versa (see nodes a, b, c and d). Between levels B and C, the populations exhibit a paraphyletic relationship in the gene tree, i.e. at any point in that time interval, all individuals in population 1 form a monophyletic subset (tracing to node e) within the more ancient gene tree of population 2, some of whose extant members diverged at nodes c and d. Below time level C, populations 1 and 2 are reciprocally monophyletic in the gene tree, tracing to nodes f and g, respectively.

such concordances are likely to arise only when populations have been reproductively separated from one another (either by intrinsic or extrinsic RBs) for reasonably long periods of time. The following examples from computer simulations are intended to provide a heuristic introduction to concepts of genealogical concordance. They represent a preliminary extension of prior single-gene treatments (Neigel and Avise 1986; Ball *et al.* 1990) to multiple loci considered together.

The current programs are slight modifications of those described by Ball *et al.* (1990), which should be consulted for assumptions, demographic conditions and other details. The original programs were designed to: (a) produce a random-mating population pedigree; (b) choose for analysis haplotypes at particular loci that in effect have 'trickled' through the organismal pedigree under Mendelian rules of segregation; and (c) find the times to the most recent common ancestor for randomly drawn pairs of haplotypes from those loci. In the present application and extension of these simulations, we considered again random mating ancestral populations (in this case, each composed of  $N \approx 100$  individuals) that become split into two daughter populations (each  $N \approx 50$ ) by a geographic or genetic barrier to successful reproduction.

Ten simulated populations were created and allowed to run for 1000 generations with no barriers to gene flow. At the end of the 1000th generation, 100 gene genealogies were chosen from each population and a barrier to gene flow was erected. The 10 pairs of daughter populations were allowed to continue for another 500 generations with 100 additional gene genealogies being chosen for analysis for each pair of populations every 10 generations. The 51 000 gene genealogies so created (100 genes  $\times$  10 simulated populations  $\times$  51 time points) were then analyzed by picking one of the daughter populations at random and determining the size of the largest monophyletic group containing only individuals from that daughter population. By this process, we hoped to show the development of concordant gene patterns through time. The results are shown in Fig. 3.

The results from these preliminary simulations support our intuitive expectations that populations isolated from one another for increasing lengths of time should be genealogically differentiable by increasing numbers of loci. Thus, if the phylogenetic histories of many independent genes in an empirical survey were to separate individuals into concordant arrays, such a finding would be consistent with long-term reproductive separation of those arrays. Furthermore, concordance among the gene arrays is highly unlikely if there were no reproductive barrier. Therefore, we suggest that such population subdivisions concordantly identified by multiple independent genetic traits should constitute the population units worthy of recognition as phylogenetic taxa (see below).

Our simulations involve monitoring actual times to common ancestry

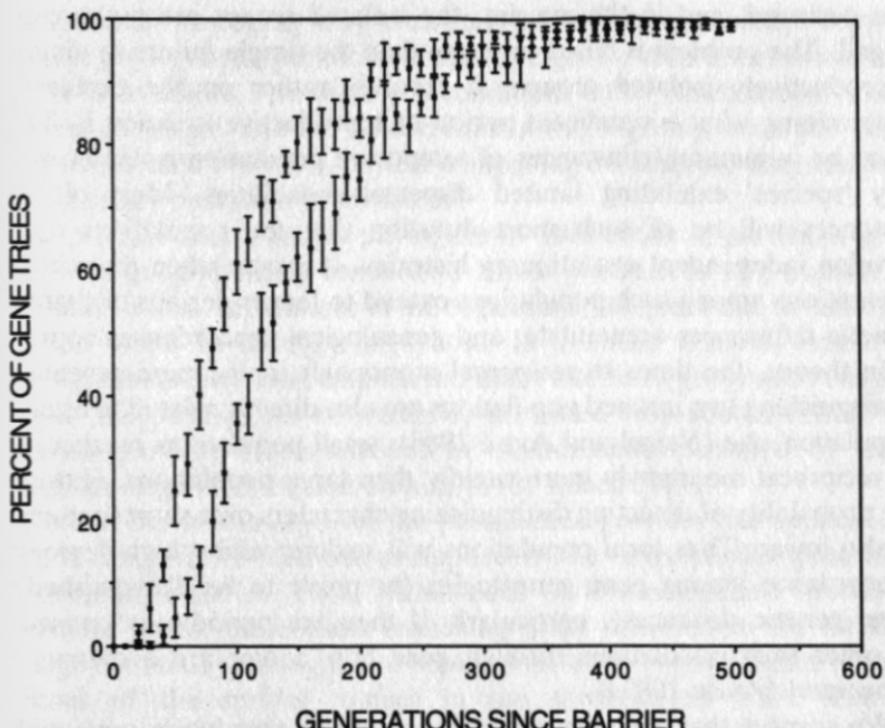


Fig. 3. The time-course of the approach to monophyly in the gene genealogies of daughter populations separated by an absolute barrier to reproduction at time zero. In computer simulations, for each of 10 independent pedigrees, genealogies of 100 genes were monitored per pedigree every 10th generation. The vertical axis is the proportion of outcomes ( $\pm 0.5$  S.D.) in which at least 75 per cent (upper curve) (or 100 per cent, lower curve) of the individuals were part of the largest monophyletic group (in a gene tree) in the daughter population (see the text for additional details).

in the pedigree (and thus ignore stochastic errors in the gene tree arising from the mutational process). In reality, many loci will evolve too slowly to provide markers for identifying recently separated populations and, in addition, the sensitivities of assay methods employed may be inadequate to detect all genetic differences that exist. Thus in practice, it is unrealistic to suppose that most or all loci should contribute to the concordance distinguishing historically separated population units. None the less, unless multiple genetic traits distinguish arrays of organisms, those arrays cannot necessarily be assumed to reflect significant phylogenetic population subdivisions.

It might be argued that the requirement of multiple gene concordance for taxonomic recognition is overly restrictive. After all, genealogical concordance will not develop until some time after reproductive isolation

has occurred, and in the interim, the isolated groups are not acknowledged. The problem is based on more than the simple failure to diagnose reproductively isolated groups. It revolves rather on the problem of determining what a significant period of reproductive isolation is. There must be innumerable instances of temporary population isolation within any 'species' exhibiting limited dispersal capabilities. Most of these instances will be of such short duration that the populations cannot develop independent evolutionary histories. It is only when reproductive separations among such populations extend to longer periods of time that genetic differences accumulate, and genealogical concordances appear.

In theory, the times to reciprocal monophyly in the gene genealogies distinguishing two isolated populations are also directly related to effective population size (Neigel and Avise 1986): small populations reach a state of reciprocal monophyly more rapidly than large populations. However, the probability of detecting distinguishing characters over short time-scales is also lower. Thus local populations will seldom exhibit high degrees of concordance among gene genealogies (or prove to be distinguished by large genetic distances), particularly if they are periodically connected to other such populations through gene flow and/or are evolutionarily ephemeral (Avise 1989a).

We suspect that the number of phylogenetic population units within most currently recognized 'biological species', as identified by genealogical concordance, will be relatively low (certainly far less than the number of 'local populations' or family units, though often greater than 1 – see below). Such phylogenetic units, supported by concordant distributions of multiple, independent traits (which can be biochemical, morphological, behavioral, etc., provided they have independent genetic bases), should represent the population assemblages that have been isolated from one another by long-term impediments to interbreeding.

A distinction should also be made between the use of multiple genes (or characters) in a discriminatory versus a concordance sense. For example, discriminant function analysis (Sneath and Sokal 1973) is a multivariate statistical approach designed to maximally separate populations based on the accumulated information from many characters, each of which may overlap in distribution between the populations. Although concordance principles are similar in that they also apply to multiple characters, the concern is not whether the populations can be distinguished, but rather with the level of concordant support for such distinctions.

### 3.2 Gene-geography concordances among taxa

The branches in the phylogenetic trees for particular loci often show strong geographic clustering (Fig. 1; Avise *et al.* 1987a). While we have argued that justification for the recognition of the major subdivisions in

an organismal phylogeny may ultimately necessitate concordant support from the phylogenetic partitions of multiple genes, such information may seldom be available. Are there any conditions under which concordances between geography and the subdivisions in single-gene genealogies (such as those provided by mtDNA) yield compelling evidence for longstanding, phylogenetic population subdivisions?

We propose that if major phylogenetic distinctions in particular gene trees were geographically concordant across populations of a number of independent taxa, separations in the organismal pedigrees due to historical isolation would be strongly implicated. In a similar fashion, vicariance biogeographers have long emphasized that patterns of geographic congruence in the phylogenies of multiple, unrelated taxa should reflect the historical patterns of disjunctions in environments occupied by those organisms (e.g. Platnick and Nelson 1978; Rosen 1978).

Work in our laboratory over the past decade provides two applications of such comparative methods as applied to the 'intraspecific' gene trees registered in mtDNA. First, within each of five recognized species of freshwater fishes, the earliest branching point observed in the mtDNA phylogenies readily distinguished populations in the eastern versus western portions of the species' ranges in the southeastern USA (Fig. 4; Bermingham and Avise 1986). These concordant patterns in the mtDNA gene trees therefore suggest two major areas of endemism for southeastern fish populations, a result further supported by a conventional biogeographic reconstruction involving concentrations of species' distributional limits in the region (Swift *et al.* 1985). For one of the assayed species (*Lepomis macrochirus*), allozyme data were also available, and they provided strong, independent genetic support (gene–gene concordance) for the phylogenetic units identified by mtDNA (Fig. 4; Avise *et al.* 1984).

A second example of geographic concordance in mtDNA phylogenies across a number of taxa involves coastal marine species in the southeastern USA (Avise *et al.* 1987b; Bowen and Avise, unpublished; Lamb and Avise, unpublished; Reeb and Avise, 1990; Saunders *et al.* 1986). Within each of six taxonomically recognized species or species groups, ranging from oysters (*Crassostrea virginica*) and horseshoe crabs (*Limulus polyphemus*) to toadfishes (*Opsanus beta* and *O. tau*), diamondback terrapins (*Malaclemys terrapin*) and seaside sparrows (*Ammodramus maritimus*), the earliest (and most strongly supported) separations in the reconstructed mtDNA phylogenies involved distinction of most Atlantic coast populations from those in the Gulf of Mexico and southeastern Florida (Fig. 5). These assayed species are confined to coastal margins, and are typically associated with saltmarsh and estuarine conditions. Reeb and Avise (1990) discuss the paleoclimatic and geologic evidence for Pliocene/Pleistocene disjunctions in suitable habitats that likely initiated the phylogenetic population separations, as well as the ecologic and

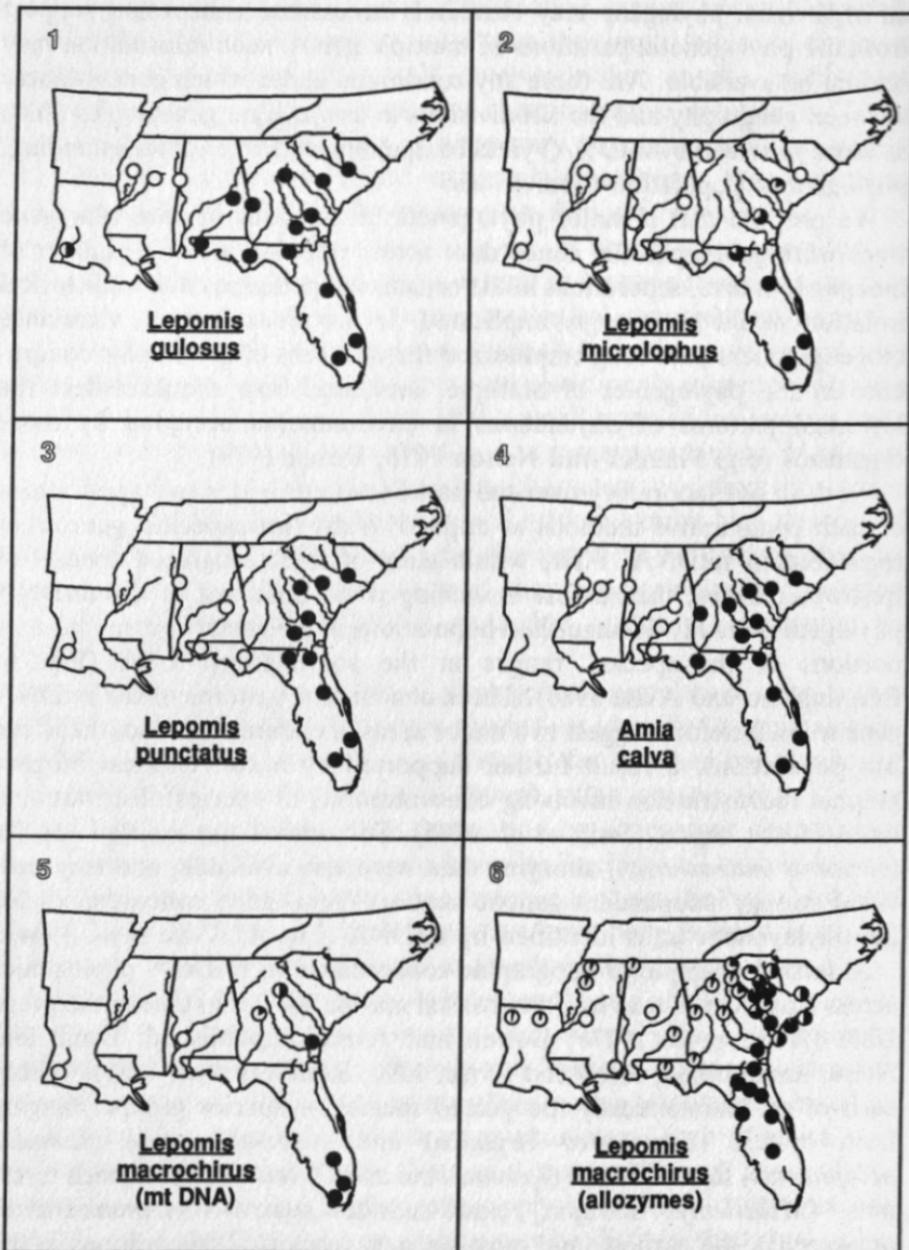


Fig. 4. Pie diagrams showing the geographic distributions of the two major mtDNA phylogenetic branches observed within each of five species of freshwater fishes: 1, warmouth sunfish; 2, redear sunfish; 3, spotted sunfish; 4, bowfin (Birmingham and Avise 1986); 5, bluegill sunfish (Avise *et al.* 1984). Also shown (panel 6) are frequencies in the bluegill sunfish of the two electromorphs at the allozyme locus *Got-2* (which are also very similar to observed geographic distributions at alleles at another nuclear gene, *Es-3*: (Avise and Smith 1974).

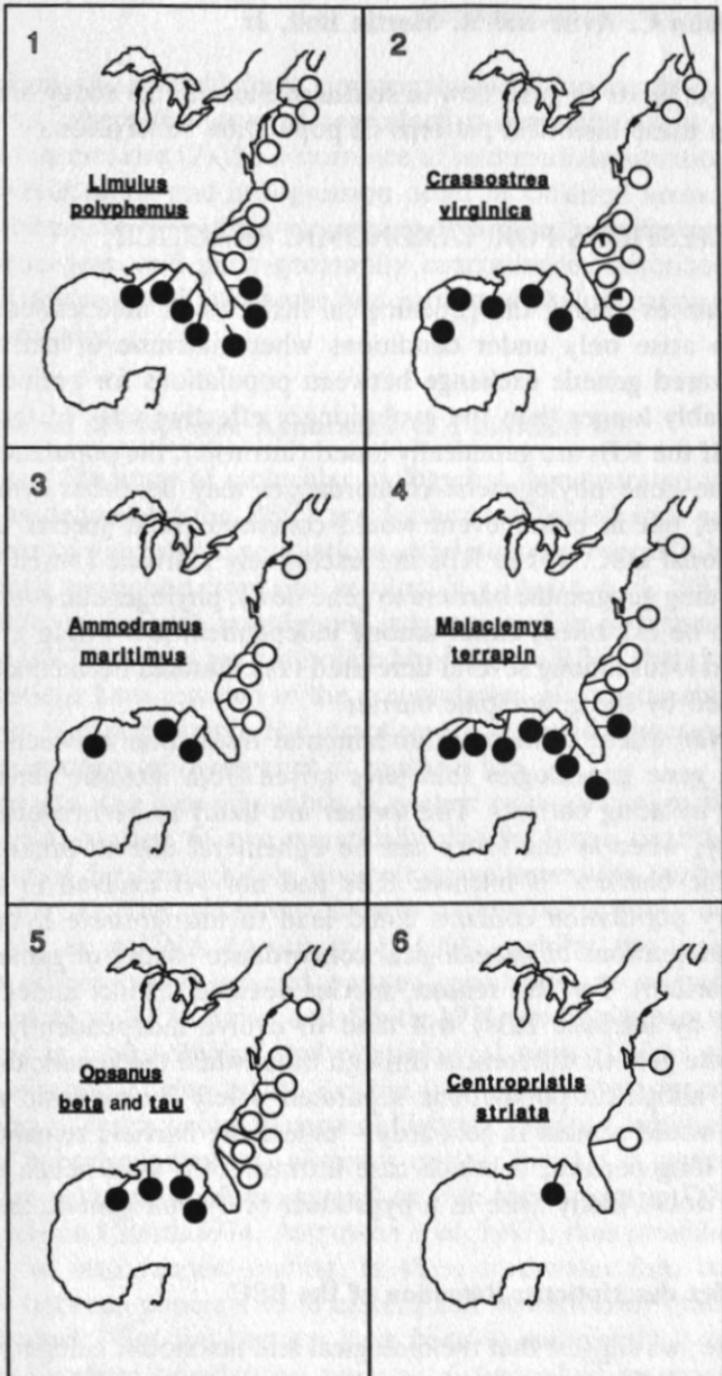


Fig. 5. Pie diagrams showing the geographic distributions of the two major mtDNA phylogenetic branches observed within each of six coastal-restricted marine species: 1, horseshoe crab (Saunders *et al.* 1986); 2, American oyster (Reeb and Avise 1990); 3, seaside sparrow (Avise and Nelson 1989); 4, diamondback terrapin (Lamb and Avise, in prep.); 5, toadfish (Avise *et al.* 1987b); 6, black sea bass (Bowen and Avise, in prep.).

hydrologic limits to gene flow in southern Florida that today may help to maintain these historical patterns of population subdivision.

#### 4. SUGGESTIONS FOR TAXONOMIC PRACTICE

Concordances among the genealogical histories of independent loci are likely to arise only under conditions where intrinsic or extrinsic RBs have severed genetic exchange between populations for periods of time considerably longer than the evolutionary effective sizes of those populations. If the RBs are genetically based (intrinsic), the populations exhibiting gene-gene phylogenetic concordances may be either sympatric or allopatric, but in either event would constitute valid species under the conventional BSC. If the RBs are exclusively extrinsic (based solely on longstanding geographic barriers to gene flow), phylogenetic concordances may also be exhibited, either among independent loci within a taxon, or at a given locus among several unrelated taxa that had been concomitantly subdivided by the geographic barrier.

However, there remains a fundamental distinction between concordances in gene genealogies that have arisen from intrinsic versus purely extrinsic isolating barriers. The former are likely to be irreversible evolutionarily, whereas the latter can be ephemeral due to breakdowns of geographic barriers. If intrinsic RBs had not yet evolved in allopatry, secondary population contacts could lead to introgressive hybridization and disintegrations of genealogical concordance (decay of gametic-phase disequilibrium). For this reason, species deemed distinct under the BSC (isolated by intrinsic RBs) will tend to evolve independently and will accumulate genetic differences through time, while the genetic differences between allopatric populations separated solely by extrinsic gene-flow barriers would remain in jeopardy – unless the barriers remained intact for very long periods, in which case intrinsic RBs (and hence biological species) would likely arise as a byproduct of overall genetic divergence.

##### 4.1 Species descriptions: Retention of the BSC

Therefore, we suggest that the biological and taxonomic category ‘species’ continue to refer to groups of actually or potentially interbreeding populations isolated by *intrinsic* RBs from other such groups. In other words, we favor a retention of the philosophical framework of the BSC. However, the *evidence* for evolutionarily significant RBs (whether intrinsic or extrinsic) will normally be indirect, and will involve concordant genetic differences among the populations involved. We recognize (as have many others) that there are operational difficulties with a strict application of this BSC, particularly as applied to allopatric forms, due to: (1) elements

of uncertainty and subjectivity in applying the definition to certain disjunct populations, where the issue of reproductive compatibility in nature is difficult to assess; and (2) the occurrence of intermediate situations where limited hybridization and introgression occur in localized areas. Despite these problems, by objectively examining organisms from the perspective of the gene-gene and gene-geography concordances described in this chapter, assessments of the degree and pattern of phylogenetic separation will be facilitated.

#### 4.2 Subspecies descriptions: Application of a modified PSC

The growing literature of molecular biology has demonstrated that many species (as defined by the BSC) are further subdivided into genetically distinct sets of geographic populations exhibiting considerable historical, phylogenetic separation from one another (e.g. Avise *et al.* 1987; Wilson *et al.* 1985). While such populations may not qualify as distinct species under the BSC (i.e. they are not isolated by intrinsic RBs), their longstanding separations have resulted in the accumulation of considerable genetic differences that in principle and practice are reflected in geographically congruent patterns of divergence at multiple loci.

To illustrate, the bluegill sunfish (*Lepomis macrochirus*) in the southeastern USA consists of two genetically distinct forms distributed east versus west of the Apalachicola River drainage separating Alabama from Georgia (Fig. 4). The two forms differ in nucleotide sequence divergence by  $\cong 8.5\%$  in mtDNA (Avise *et al.* 1984), exhibit two nearly fixed allozyme differences at assayed nuclear genes (and an overall genetic distance of  $D \cong 0.15$ : Avise and Smith 1974) and evidence additional differences in morphological and physiological traits (Hubbs and Allen 1944; Hubbs and Lagler 1958). Yet the two forms hybridize extensively in a secondary contact zone in parts of Georgia and the Carolinas. Within particular hybrid populations, alleles at nuclear loci are in gametic phase equilibrium with one another, as well as with the distinct mtDNA genotypes (Avise and Smith 1974; Asmussen *et al.* 1987), thus providing strong evidence for near random-mating. In these freshwater fish, barriers to gene flow between populations in eastern and western river drainages are purely extrinsic. But the barriers have been of sufficiently longstanding duration that these populations are now distinguished by many genetic characteristics. We propose that such sets of populations, concordantly recognizable by many independent genetic differences, reflect the major phylogenetic partitions within a biological species that are worthy of formal subspecies recognition. (The two bluegill forms has indeed been assigned the Latin trinomials *L.m. macrochirus* and *L.m. purpurescens*).

We therefore suggest the following definitional guidelines for the taxonomic category 'subspecies' in sexually reproducing organisms: *subspecies*

are groups of actually or potentially interbreeding populations phylogenetically distinguishable from, but reproductively compatible with, other such groups. Importantly, the evidence for phylogenetic distinction must normally come from the concordant distributions of multiple, independent, genetically based traits.

Because the populations constituting distinct subspecies are reproductively compatible, subspecies will normally be allopatric (though some of their populations may meet in secondary hybrid zones), and any significant phylogenetic partitions will be registered by multiple loci exhibiting congruent geographic distributions. The longer the geographic isolation, the greater the opportunity for genetic divergence and also for the accumulation of concordant phylogenetic distinctions at multiple loci. It remains to be seen how often genealogical concordance will be observed among geographic populations within biological species, such that subspecies designations are warranted, but we suspect that many biological species (such as the bluegill sunfish) will prove to have rather fundamental phylogenetic subdivisions resulting from historical population separations.

## 5. ADVANTAGES AND DIFFICULTIES OF CONCORDANCE PRINCIPLES IN TAXONOMY

### 5.1 Advantages of concordance principles

1. *The category 'subspecies' will rest on a firmer empirical foundation.* Conventionally, subspecies descriptions have been based on one or a few traits (such as pelage color or size) that allowed distinction of a high percentage of individuals in a geographic region from those in other areas (e.g. Mayr 1969). As noted and criticized by Wilson and Brown (1953, p. 104), 'The tendency in this method has been to delimit races on the basis of one or several of the most obvious characters. . . ; the remainder of the geographically variable characters are then ignored, or if they are considered at all, they are analyzed only in terms of the subspecific units previously defined.' We agree with Wilson and Brown's (1953, p. 110) contention that 'geographical variation should be analyzed first in terms of genetically independent characters, which would then be employed synthetically to search for possible racial groupings'.

2. *Major phylogenetic units within biological species will be afforded formal taxonomic recognition.* Sets of long-isolated populations, reflecting historical disjunctions in suitable habitat, probably occur within the boundaries of many biological species (as delimited by intrinsic RBs). Such populations would be afforded subspecies status under concordance principles. The resulting taxonomies can be of great use in reconstructing

historical biogeography (Avise *et al.* 1987), as well as in summarizing the apportionment of intraspecific genetic diversity that should aid the management and preservation efforts of conservation biology (Avise 1989b). With the partitions in intraspecific phylogeny properly recognized (by concordance principles), the interpretation of characters at variance with the primary pattern should also be facilitated. For example, characters likely to be under intense selection pressures related to ecological circumstances (such as pelage color or body size) may often be geographically discordant with the phylogenetic subdivisions. Thus under concordance principles (as under the PSC, but not the BSC), explicit attention is focused on phylogenetic histories of populations.

3. *The category 'species' will remain similar to that currently employed, and intrinsic RBs retain primacy as a conceptual guide to species distinctions.* Unlike the PSC, which would require a drastic revision of taxonomic designations at the species level, application of concordance principles should have little effect on current species-level taxonomies. The conventional procedure for distinguishing biological species already rests implicitly on concordance principles – intrinsic reproductive barriers are deduced from the indirect evidence of differences (sometimes requiring close scrutiny) in numerous morphological, behavioral and other assayable traits. Because reproductive assessments seldom can be made directly (see Sokal and Crovello 1970), such character-state surrogates of reproductive unions and disjunctions are likely to remain of prime utility in species descriptions.

To the extent that genetically-based RBs are irreversible evolutionarily, they constitute irrevocable partitions of biotic diversity. The genes directly responsible for RBs, by cementing biotic subdivisions, will have a correlated effect on the phylogenetic partitions of many other loci in the genome (in the presence of RBs, all neutral loci will eventually evolve to a status of concordant reciprocal monophyly). Thus under concordance principles (as under BSC, but not the PSC), intrinsic RBs properly occupy a position of fundamental evolutionary significance.

## 5.2 Potential difficulties of concordance principles

1. *Some species may be overlooked.* Occasionally, populations may have evolved intrinsic barriers to reproduction so recently that phylogenetic separation and concordance are not yet evident in genes other than those directly responsible for the RBs themselves. While it seems unlikely that RB-producing genetic traits alone would normally distinguish species (except perhaps in very young polyploid assemblages, or in other situations where chromosomal or other genetic changes rapidly and recently precipitated intrinsic reproductive isolation), some 'good' biological species could

be missed under concordance principles (they would likely be overlooked under the PSC and BSC also, unless the critical genetic trait conferring reproductive isolation were examined).

Some events that give species under the BSC may not be immediately recognizable under an implementation of concordance principles. In the special case where a single character change results in the intrinsic reproductive isolation of two groups, there will be no immediate concordance between that character and others, although such concordance will inevitably develop through time. Overlooking valid biological species that were recently isolated is of special concern because such taxa should afford the chance to study the initial stages of speciation. Thus intrinsic RBs should retain primacy as a conceptual basis for species recognition (as under the BSC).

2. *Subjective taxonomic judgments may be required for intermediate levels of genetic congruence and divergence.* Admittedly, there are several gray areas in this heuristic construct for taxonomic recognition under concordance principles. For example, how many gene phylogenies must support the subspecies distinctions? And, how much concordance must characterize the geographic distributions of the gene phylogenies? In principle, levels of genealogical concordance can vary along a continuum (see Fig. 3), so any specific suggestions for implementing concordance methods will necessarily involve an element of arbitrariness. However, all taxonomic schemes that divide the elements of the continuous pattern of evolutionary differentiation into discrete categories (such as subspecies and species) face similar difficulties with intermediate situations.

3. *Obtaining phylogenies for independent genes may be difficult.* Ideally, complete information on the histories of allelic relationships at each of many loci would be most desirable in the search for genealogical concordance. However, apart from the large number of mtDNA phylogenies presented in recent years (Avise *et al.* 1987a; Wilson *et al.* 1985), very few gene trees have as yet been generated for any taxa at the microevolutionary scale (Aquadro *et al.* 1986; Avise 1989a; Langley *et al.* 1988), despite the increased availability of laboratory methods for restriction site mapping and nucleotide sequencing. In the assay of nuclear genomes, one technical complication involves the isolation and assay of haplotypes from diploid organisms; but a more serious problem involves the likelihood that intragenic recombination or gene conversion will have shuffled nucleotide sequences at a locus, thus confounding reconstruction of gene genealogies (Aguadé *et al.* 1989; Hudson and Kaplan 1988; Stephens 1985; Templeton *et al.* 1987). It remains to be seen whether significant disequilibria involving intragenic restriction sites or nucleotide

sequences (of use to phylogenetic reconstruction) will commonly be found among geographically separated populations.

As complete gene genealogies may remain difficult to obtain for most nuclear loci, frequencies of phylogenetically unordered alleles [such as those provided by allozyme methods, or restriction fragment length polymorphism (RFLP) analyses: Avise 1989a] will continue to provide an important source of molecular character states normally used in the search for geographically concordant population subdivisions. Differences in morphological, behavioral and other phenotypic attributes, provided they are genetically based and independent, will of course continue to be important characters for survey in the search for patterns of geographic and genetic concordance.

## 6. CONCLUSIONS

One important root of the PSC probably traces to Simpson's (1951) paleontological perspective on taxa, summarized in his definition of an evolutionary species as 'a lineage (ancestral-descendant sequence of populations) evolving separately from others and with its own unitary evolutionary role and tendencies' (Simpson 1961, p. 153). Current versions of the PSC, apparently motivated by a perceived lack within the BSC of an adequate emphasis on history and phylogeny, have led some PSC proponents to reject the BSC's emphasis on reproductive isolation. Principles of genealogical concordance provide a compromise or composite stance between the BSC and PSC. Concepts of concordance are far from new in systematics (see, for example, Wilson and Brown 1953) and numerous statements can be found in support of the desirability of concordant information prior to taxonomic recognition. For example, Mayr (1969, p. 192) notes that 'geographic variation in the salamander *Plethodon jordani* is too discordant to justify the recognition of formal subspecies, even though the variation of each individual character shows a definite geographic trend'. Yet such sentiments too seldom have been followed, and many taxa continue to be recognized on the basis of one or a few diagnostic traits. The new generation of systematists may avoid a repeat of such errors by requiring concordance among several independent characters before advocating formal taxonomic recognition of putative population disjunctions. By focusing on the phylogenetic consequences of intrinsic RBs, and by emphasizing that important historical partitions can also be present within biological species because of extrinsic RBs, concordance principles should provide a useful set of philosophical and operational guidelines for the recognition of biotic and taxonomic diversity.

## ACKNOWLEDGMENTS

This work was supported by NSF grants to J.C.A., and by a NIH Genetics Training grant to R.M.B.

## REFERENCES

- Aguadé, M., Miyashita, N. and Langley, C. H. (1989). Restriction-map variation at the *Zeste-tko* region in natural populations of *Drosophila melanogaster*. *Mol. Biol. Evol.* **6**, 123-30.
- Aquadro, C.F., Desse, S. F., Bland, M. M., Langley, C. H. and Laurie-Ahlberg, C. C. (1986). Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. *Genetics* **114**, 1165-90.
- Asmussen, M. A., Arnold, J. and Avise, J. C. (1987). Definition and properties of disequilibrium statistics for associations between nuclear and cytoplasmic genotypes. *Genetics* **115**, 755-68.
- Avise, J. C. (1989a). Gene trees and organismal histories: A phylogenetic approach to population biology. *Evolution* **43**, 1192-1208.
- (1989b). A role for molecular genetics in the recognition and conservation of endangered species. *Trends Ecol. Evol.* **4**, 279-81.
- and Nelson, W. S. (1989). Molecular genetic relationships of the extinct Dusky seaside sparrow. *Science* **243**, 646-8.
- and Smith, M. H. (1974). Biochemical genetics of sunfish. I. Geographic variation and subspecific intergradation in the bluegill, *Lepomis macrochirus*. *Evolution* **28**, 42-56.
- , Bermingham, E., Kessler, L. G. and Saunders, N. C. (1984). Characterization of mitochondrial DNA variability in a hybrid swarm between subspecies of bluegill sunfish (*Lepomis macrochirus*). *Evolution* **38**, 931-41.
- , Arnold, J., Ball, R. M. Jr, Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A. and Saunders, N. C. (1987a). Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Ann. Rev. Ecol. Syst.* **18**, 489-522.
- , Reeb, C. A. and Saunders, N. C. (1987b). Geographic population structure and species differences in mitochondrial DNA of mouthbrooding marine catfishes (Ariidae) and demersal spawning toadfishes (Batrachoididae). *Evolution* **41**, 991-1002.
- , Bowen, B. W. and Lamb, T. (1989). DNA fingerprints from hypervariable mitochondrial genotypes. *Mol. Biol. Evol.* **6**, 258-69.
- Ayala, F. J. (1976). Molecular genetics and evolution. In *Molecular evolution* (ed. F. J. Ayala), pp. 1-20. Sinauer, Sunderland, Mass.
- Ball, R. M., Jr, Neigel, J. E. and Avise, J. C. (1990). Gene genealogies within the organismal pedigrees of random mating populations. *Evolution* **44**, 1109-19.
- Bermingham, E. and Avise, J. C. (1986). Molecular zoogeography of freshwater fishes in the southeastern United States. *Genetics* **113**, 939-65.
- Burke, T. (1989). DNA fingerprinting and other methods for the study of mating success. *Trends Ecol. Evol.* **4**, 139-44.

- Coyne, J. A. and Orr, H. A. (1989). Patterns of speciation in *Drosophila*. *Evolution* **43**, 362–381.
- Cracraft, J. (1983). Species concepts and speciation analysis. In *Current ornithology* (ed. R. F. Johnston), Vol. 1, pp. 159–87. Plenum Press, New York.
- Dobzhansky, T. (1937). *Genetics and the origin of species*. Columbia University Press, New York.
- (1970). *Genetics of the evolutionary process*. Columbia University Press, New York.
- Donoghue, M. J. (1985). A critique of the biological species concept and recommendations for a phylogenetic alternative. *Bryologist* **88**, 172–81.
- Ehrlich, P. R. (1961). Has the biological species concept outlived its usefulness? *Syst. Zool.* **10**, 167–76.
- and Raven, P. H. (1969). Differentiation of populations. *Science* **165**, 1228–32.
- Eldredge, N. and Cracraft, J. (1980). *Phylogenetic patterns and the evolutionary process*. Columbia University Press, New York.
- Hubbs, C. L. and Allen, E. R. (1944). Fishes of Silver Springs, Florida. *Proc. Fla. Acad. Sci.* **6**, 110–30.
- and Lagler, K. F. (1958). *Fishes of the Great Lakes region*. University of Michigan Press, Ann Arbor.
- Hudson, R. R. and Kaplan, N. L. (1988). The coalescent process in models with selection and recombination. *Genetics* **120**, 831–40.
- Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985). Individual-specific 'fingerprints' of human DNA. *Nature* **316**, 76–9.
- Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Paabo, S., Villablanca, F. X. and Wilson, A. C. (1989). Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proc. Natl Acad. Sci. USA* **86**, 6196–200.
- Lander, E. S. (1989). DNA fingerprinting on trial. *Nature* **339**, 501–505.
- Langley, C. H., Shrimpton, A. E., Yamazaki, T., Miyashita, N., Matsuo, Y. and Aquadro, C. F. (1988). Naturally occurring variation in the restriction map of the Amy region of *Drosophila melanogaster*. *Genetics* **119**, 619–29.
- Levin, D. A. (1979). The nature of plant species. *Science* **204**, 381–4.
- Mayr, E. (1969). *Principles of systematic zoology*. McGraw-Hill, New York.
- (1970). *Populations, species, and evolution*. Belknap Press, Harvard University, Cambridge, Mass.
- McKittrick, M. C. and Zink, R. M. (1988). Species concepts in ornithology. *The Condor* **90**, 1–14.
- Mishler, B. D. and Donoghue, M. J. (1982). Species concepts: A case for pluralism. *Syst. Zool.* **31**, 491–503.
- Nei, M. (1987). *Molecular evolutionary genetics*. Columbia University Press, New York.
- Neigel, J. E. and Avise, J. C. (1986). Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. In *Evolutionary processes and theory* (ed. E. Nevo and S. Karlin), pp. 515–34. Academic Press, London.
- Nelson, G. J. and Platnick, N. I. (1981). *Systematics and biogeography: Cladistics and vicariance*. Columbia University Press, New York.

- Pamilo, P. and Nei, M. (1988). Relationships between gene trees and species trees. *Mol. Biol. Evol.* **3**, 254–9.
- Platnick, N. I. and Nelson, G. J. (1978). A model of analysis for historical biogeography. *Syst. Zool.* **27**, 1–16.
- de Queiroz, K. and Donoghue, M. J. (1988). Phylogenetic systematics and the species problem. *Cladistics* **4**, 317–38.
- Raven, P. H. (1976). Systematics and plant population biology. *Syst. Botan.* **1**, 284–316.
- Reeb, C. A. and Avise, J. C. (1990). A genetic discontinuity in a continuously distributed species: Mitochondrial DNA in the American oyster, *Crassostrea virginica*. *Genetics* **124**, 397–406.
- Rosen, D. E. (1978). Vicariant patterns and historical explanation in biogeography. *Syst. Zool.* **27**, 159–88.
- (1979). Fishes from the uplands and intermontane basins of Guatemala: Revisionary studies and comparative geography. *Bull. Amer. Mus. Nat. Hist.* **162**, 267–376.
- Saunders, N. C., Kessler, L. G. and Avise, J. C. (1986). Genetic variation and geographic differentiation in mitochondrial DNA of the horseshoe crab, *Limulus polyphemus*. *Genetics* **112**, 613–27.
- Simpson, G. G. (1951). The species concept. *Evolution* **5**, 285–98.
- (1961). *Principles of animal taxonomy*. Columbia University Press, New York.
- Sneath, P. H. A. and Sokal, R. R. (1973). *Numerical taxonomy*. W. H. Freeman, San Francisco.
- Sokal, R. R. (1973). The species problem reconsidered. *Syst. Zool.* **22**, 360–74.
- and Crovello, T. J. (1970). The biological species concept: A critical evaluation. *Amer. Nat.* **104**, 127–53.
- Spiess, E. B. (1977). *Genes in populations*. John Wiley, New York.
- Stephens, J. C. (1985). Statistical methods of DNA sequence analysis: Detection of intragenic recombination or gene conversion. *Mol. Biol. Evol.* **2**, 539–56.
- Swift, C. C., Gilbert, C. R., Bortone, S. A., Burgess, G. H. and Yerger, R. W. (1985). Zoogeography of the freshwater fishes of the southeastern United States: Savannah River to Lake Ponchartrain. In *Zoogeography of North American freshwater fishes* (ed. C. H. Hocutt and E. O. Wiley), pp. 213–65. John Wiley, New York.
- Tajima, F. (1983). Evolutionary relationships of DNA sequences in finite populations. *Genetics* **105**, 437–60.
- Takahata, N. (1989). Gene genealogy in three related populations: Consistency probability between gene and population trees. *Genetics* **122**, 957–66.
- Templeton, A. R., Boerwinkle, E. and Sing, C. F. (1987). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics* **117**, 343–51.
- Wiley, E. O. (1978). The evolutionary species concept reconsidered. *Syst. Zool.* **27**, 17–26.
- (1981). *Phylogenetics: The theory and practice of phylogenetic systematics*. John Wiley, New York.
- Wilson, A. C., Cann, R. L., Carr, S. M., George, M. Jr, Gyllensten, U. B.,

- Helm-Bychowski, K. M., Higuchi, R. G., Palumbi, S. R., Prager, E. M., Sage, R. D. and Stoneking, M. (1985). Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol. J. Linn. Soc.* **26**, 375-400.
- Wilson, E. O. and Brown, W. L. Jr (1953). The subspecies concept and its taxonomic application. *Syst. Zool.* **2**, 97-111.