

GENEALOGICAL PERSPECTIVES ON THE SPECIES PROBLEM¹

David A. Baum² and Kerry L. Shaw³

ABSTRACT. The species category marks the boundary between systematics and population biology, hence a clear understanding of species is critical for both these fields and their intercommunication. There are several distinct entities in nature that have been subsumed under the single term "species" (e.g., genealogical units, interbreeding groups, ecological units, and morphological clusters). We focus on one species-entity of particular concern to phylogenetic systematists, the "genealogical species." We start from the assumption that genealogical species are basal taxa (taxa containing no less inclusive taxa) and that they reside at the boundary between reticulate and divergent genealogy. These two assumptions are met if we define species as basal, exclusive groups of organisms, where exclusive groups are ones whose members are all more closely related to each other than to any organisms outside the group. We explore two genealogical measures of relatedness to see how they affect the implementation of this species definition. If relatedness is measured in terms of recency of common organismic descent, then species delimitation becomes problematic. We argue that this is due to measuring relatedness in terms of organismal descent, and suggest that we should instead focus on the genealogical histories of genes. This can be accomplished with insights from coalescent theory and in so doing a tractable, theoretically consistent definition of the genealogical species can be developed.

The species is probably the most important concept in organismic biology, yet it is also one of the most problematic and confusing. Despite many attempts to clarify the definition of the term "species" the problem still attracts much attention and intense controversy. The dominant approach to the problem is to view the species as a single entity with many distinct biological properties (e.g., Mayr, 1942; Wiley, 1978; Paterson, 1985; Templeton, 1989). In this framework the "species problem" is the search for the single property or combination of properties that define the species. However, a number of biologists reject the assumption that only one entity is being referred to by the term "species." Instead, they emphasize that the term has been used to refer to several, equally meaningful but distinct entities (e.g., Holsinger, 1984; Kitcher, 1984; de Queiroz and Donoghue, 1988; Wilkinson, 1990; Ereshefsky, 1992). Under this pluralistic perspective the primary aim is to conceptually separate the species-entities that have been heretofore conflated, characterize each of them rigorously, and develop a terminology that avoids confusion (Ereshefsky, 1992).

We subscribe to the pluralistic conception of species and in this paper focus on a species-entity of particular concern to phylogenetic systematists, the "genealogical species." Starting from some minimal assumptions, we proceed to develop a rigorous definition of this

¹For useful discussion and comments on earlier drafts of this paper we would like to thank: John Beatty, Joel Cracraft, Keith Crandall, Jerrold Davis, Kevin de Queiroz, Michael Donoghue, Jeff Doyle, Marc Ereshefsky, Anne Gerber, William Hahn, Richard Harrison, Robert Kowal, Melissa Luckow, Wayne Maddison, and Elliott Sober. We are grateful to Michael Donoghue for his efforts in arranging the symposium. David Baum thanks Ken Sytsma and the members of his lab for their support during the writing of this paper. Kerry Shaw thanks the Templeton lab for many useful discussions on the subject of species.

²Botany Department, University of Wisconsin, Madison, WI 53706, U.S.A. Current address: Harvard University Herbaria, 22 Divinity Ave., Cambridge, MA 02138, U.S.A.

³Biology Department, Washington University, St. Louis, MO 63130, U.S.A. Current address: Section of Ecology and Systematics, Corson Hall, Cornell University, Ithaca, NY 14853-2701, U.S.A.

entity using the terminology and assumptions of phylogenetic systematics. Our starting assumptions are (1) genealogical species are basal taxa (*sensu* Cracraft, 1989), that is, taxa that contain no included taxa, and (2) genealogical species reside at the boundary between reticulate and divergent genealogy. Before defining the genealogical species we will discuss these two premises and their most salient implications.

SPECIES ARE BASAL TAXA

If species are basal taxa then whatever criteria are used to define taxa must also apply to species. So what are taxa? Within phylogenetic systematics⁴, taxa are groups of organisms formally recognized because they share a common genealogical history (de Queiroz, 1992; de Queiroz and Gauthier, 1992). In defining taxa, one therefore only considers the evolutionary history of groups of organisms, ignoring both their current properties and their future fate (Lidén and Oxelman, 1989; Kluge, 1990; Lidén, 1992). While we do not deny that the current properties and future fate of organisms are important, they are irrelevant to the definition of phylogenetic taxa and the genealogical species described here. We are interested in the species as a product of evolution, not a player in evolution. Thus, while some *biologically important species-entities may be defined on the basis of either current properties (e.g., interbreeding, morphological similarity) or future tendencies (e.g., persistence as an independent evolutionary lineage)*, these species-entities are not conceptually equivalent to the genealogical species discussed here.

In phylogenetic systematics, monophyly is considered the most important historical attribute of taxa. However, despite its importance in systematic theory, a monophyletic criterion cannot be used to define taxa if species are to be basal taxa (Nixon and Wheeler, 1990). This is true whether one uses a narrow or broad definition of the term monophyly. The narrow concept (e.g., Hennig, 1966; Nixon and Wheeler, 1990; Wheeler and Nixon, 1990) defines a monophyletic group as a group of species descended from a common ancestral species. Given the use of the term "species" in this definition, it is clear that species cannot be considered monophyletic (McKittrick and Zink, 1988). Thus, if taxa are defined as monophyletic groups in the narrow sense, species cannot be taxa.

The alternative position defines a monophyletic group more broadly as a group descended from a common ancestor, where ancestors are any entities that are capable of leaving descendants (e.g., taxa, populations, single organisms, breeding pairs, cells; de Queiroz and Donoghue, 1988). Under this definition, monophyly can logically apply to species, but species would not be basal taxa as monophyletic groups of organisms occur at all levels from higher taxa to family groups. Although monophyly, defined in the broad sense (the sense we will use for the remainder of this paper), may be a necessary property of some taxa, it is not a sufficient criterion for taxa since many monophyletic groups are not taxa. Some other historical property is therefore needed to define taxa.

We suggest that taxa should be defined as "exclusive" groups of organisms, where an exclusive group is one whose members are all more closely related to each other than to any organisms outside the group (de Queiroz and Donoghue, 1990b; Baum, 1992). The most important characteristic of an exclusive group is that although one exclusive group may be nested within another, two exclusive groups cannot overlap. This means that exclusive groups can only be delimited when relationships are hierarchic. Hence, if taxa are defined as exclusive groups of organisms they will always be hierarchically related to each other.

⁴"Phylogenetic systematics" is used here to refer to the school of systematics that stresses the primacy of evolutionary history (exemplified by de Queiroz and Gauthier, 1990). It should not be confused with "cladistics" which emphasizes the hierarchical distribution of characters (exemplified by Nelson, 1989). This distinction is discussed in more detail by de Queiroz and Donoghue (1990b).

Since we start from the premise that genealogical species are taxa, then they too are exclusive groups. Furthermore, since species are basal taxa, species cannot contain nested within them any other exclusive groups. Thus, we define genealogical species as basal, exclusive groups of organisms.

SPECIES RESIDE AT THE RETICULATE/DIVERGENT BOUNDARY

Our second starting premise is that genealogical species reside at the boundary between reticulate and divergent genealogy. It is therefore necessary to examine why we think such a genealogical discontinuity exists and what form it takes. Below we describe the two basic genealogical patterns, divergence and reticulation, by contrasting the genealogies of sexual and asexual organisms.

Asexual organisms have purely divergent genealogies. Each individual has only one parent in each previous generation, which means that all the organisms existing at any one moment in time can be arranged into a strictly hierarchical pattern of relationships. As long as reproduction is asexual, this hierarchical structure will exist across all levels of biological organization, stretching from the highest taxonomic levels to the individual organisms.

Sexual organisms have a fundamentally different genealogy due to the fact that each individual has two or more parents in all previous generations. The genealogy is "reticulate," with ancestor-descendant lineages that anastomose through interbreeding. Therefore, organisms form overlapping patterns of relationship. For example, each of us is potentially a member of two groups of first cousins, one on our father's side and one on our mother's side. However, one's maternal and paternal first cousins are not each other's first cousins. Hierarchies (as defined by Gregg, 1954) cannot contain overlapping sets and, thus, this example demonstrates that the pattern of relationships arising from a reticulate genealogy is fundamentally non-hierarchic.

In contrast to the relationships within sexually reproducing populations, the higher taxa formed by biparental organisms can be arranged in a non-overlapping nested array. The divergent genealogy that is inferred to have generated this hierarchical structure is termed a "phylogeny." Biparental organisms thus have a distinctive genealogical pattern. The organisms within a population are related by a reticulate pattern of descent whereas the higher taxa to which these organisms belong are related by a divergent pattern of descent (the phylogeny). Clearly there must be some boundary (however fuzzy) at which the reticulate genealogies of sexual organisms become transmuted into the divergent relations of higher taxa. There must be a taxon whose relationships with other taxa are predominantly divergent but whose parts (the organisms ascribable to the taxon) are related to each other by a predominantly reticulate genealogy. We refer to this entity as the "genealogical species⁵."

As discussed above, exclusive groups of organisms are hierarchically related to each other and, thus, the existence of exclusive groups indicates that a genealogy is divergent. Similarly, when genealogies are reticulate, exclusive groups cannot arise. Therefore, the lowest level at which exclusive groups emerge should approximate to the reticulate/divergent boundary and, hence, defining species as basal, exclusive groups of organisms satisfies both our starting assumptions.

⁵This concept is a phylogenetic species concept insofar as it is defined primarily on the basis of historical relatedness rather than current properties or future fate. However, it will become clear that it has several novel features that distinguish it from other phylogenetic species concepts (e.g., Cracraft, 1983, 1989; Donoghue, 1985; Mishler and Brandon, 1987; de Queiroz and Donoghue, 1988, 1990a, 1990b; Nixon and Wheeler, 1990; Wheeler and Nixon, 1990; Davis and Manos, 1991; Davis and Nixon, 1992; reviewed by McKittrick and Zink, 1988; Avise and Ball, 1990; Frost and Hillis, 1990; Baum, 1992). We use the term "genealogical species" to avoid confusion with these other concepts.

EXCLUSIVITY AND THE EVALUATION OF RELATEDNESS

Central to the definition of exclusivity is degree of relatedness, and criteria are needed whereby relatedness can be evaluated. In accordance with our emphasis on evolutionary history we discuss two genealogical measures: recency of common organismal descent and recency of genetic coalescence. The former approach focuses on the genealogy of organisms, whereas the latter focuses on the genealogies of genes. Our main claim is that, despite its simplicity, recency of common organismal descent leads to severe problems when used to define genealogical species. We show this to be the case whether one views taxa as comprising both ancestors and descendants (the “time-extended” approach) or just contemporaneous organisms (the “time-limited” approach). We argue that the problems with evaluating relatedness through organismic ancestry arise because reticulation is discernible only at the gene-level and not at the organismal level. Thus, only an approach that looks at genetic coalescence will successfully place genealogical species at the reticulate/divergent boundary.

Recency of Common Ancestry. In phylogenetic systematics, degree of relatedness is traditionally determined by recency of common descent. Organisms are considered closely related if they share a common ancestor in the recent past, distantly related if their most recent common ancestor (MRCA) is ancient. Using this measure of relatedness, an exclusive group of organisms (a taxon) is one whose MRCA existed more recently than the common ancestor of any member of the group and any organism outside the group.

One obvious advantage of this definition of exclusivity is that, in a divergent genealogy, monophyletic groups and only monophyletic groups are exclusive. As Hennig (1966) realized, monophyletic groups can be identified by the possession of synapomorphies. This means that when a genealogy is divergent, cladistic methods can identify exclusive groups of organisms (i.e., taxa). However, the symmetry of monophyly and exclusivity does not apply when genealogies are reticulate (de Queiroz and Donoghue, 1990b; Baum, 1992). In a reticulate genealogy, monophyletic groups of organisms occur (although in practice they are difficult to detect because synapomorphy ceases to be evidence of monophyly; de Queiroz and Donoghue, 1988; Nixon and Wheeler, 1990), but they will not be exclusive. This means that monophyly is irrelevant to the definition or practical recognition of genealogical species.

There are two alternative approaches to delimiting exclusive groups when recency of common descent is used to evaluate relationships, and these lead to quite different outcomes. The “time-extended” approach considers a group of organisms (living or not) exclusive if they share a more recent common ancestor with each other than with any organisms (living or not) outside the group. The “time-limited” approach judges a group of contemporaneous organisms exclusive if they share a more recent common ancestor with each other than with any contemporaneous organisms outside the group. We will discuss the time-extended and time-limited approaches in turn showing that, when considering organismic measures of common ancestry, neither of them places the genealogical species at the reticulate/divergent boundary.

The time-extended approach requires quite exceptional biological circumstances for exclusive groups to be formed. Consider the hypothetical phylogeny shown in Figure 1. An ancestral population has divided into two at time t , after which time the two descendant “lineages” have not interbred. Superficially this diagram implies that the two descendant “lineages” are non-overlapping and, hence, each shows exclusivity of descent. However, close scrutiny of the events at the time of the split reveals that in fact neither of the two descendant “lineages” are exclusive. For example, the MRCA of labeled individuals **A** and **B** is organism **C** living at time $t-3$. However, the common ancestor of **A** and **D** (a member of the other descendant lineage) is organism **E** living at time $t-1$. Under the “time-extended”

approach, "lineage 1" is therefore not exclusive (the same can be shown for lineage 2). No matter how long the two "lineages" persist without interbreeding, they will never become exclusive.

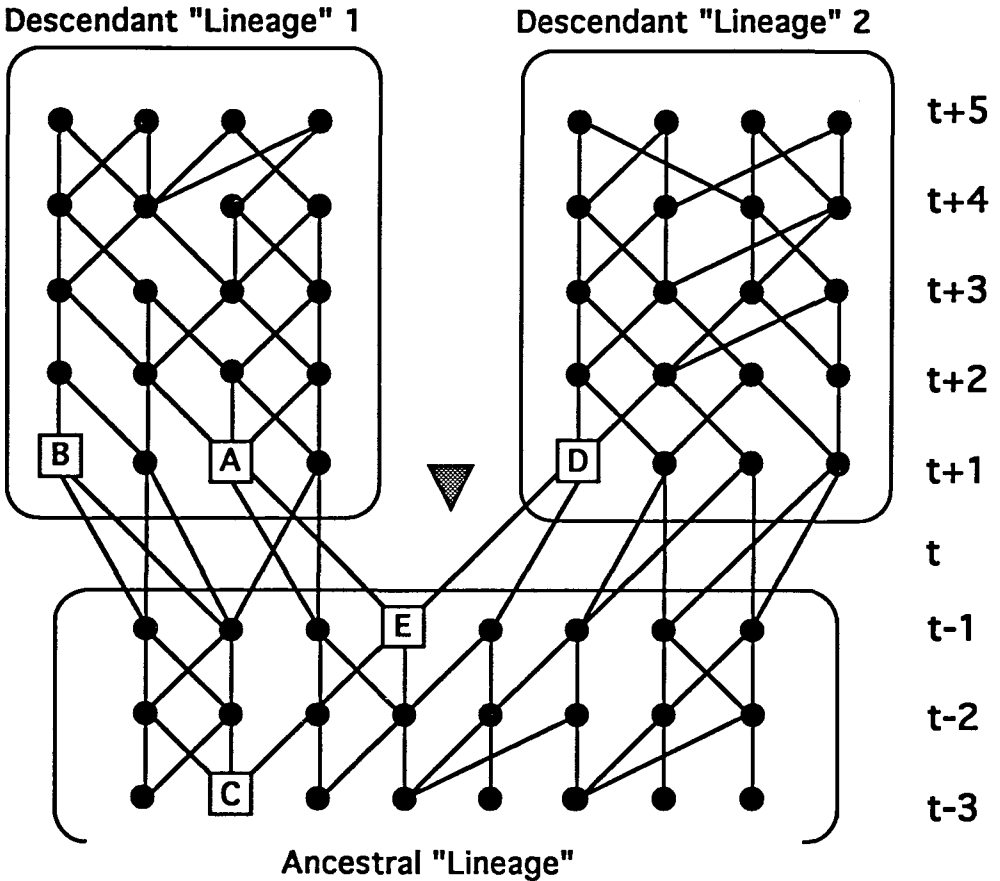


Figure 1. A hypothetical organismic genealogy of biparental, hermaphroditic annuals. An ancestral "lineage" divides at time t into two descendant "lineages." The MRCA of labeled organisms A and B is organism C and that of organisms A and D is organism E.

The only way that exclusive groups can form is when a single individual or breeding pair gives rise to a "lineage" whose descendants never interbreed with other "lineages." Such groups may form (e.g., when a founder population is derived from a single breeding pair) but only under very rare conditions in nature. Furthermore, evolutionary theory predicts that divergence and the formation of hierarchical structure occurs without such extreme "bottlenecks." Thus, it is clear that measuring relatedness using recency of common ancestry under a time-extended view of taxa results in genealogical species emerging well above the reticulate/divergent boundary.

Unlike the time-extended approach, the time-limited definition of exclusivity succeeds in separating the two descendant "lineages" in Figure 1. Looking at the most recent generation only ($t+5$), the four members of each hypothetical "lineage" share a more recent

ancestor with each other than with any of the organisms in the other "lineage." They are therefore exclusive and, assuming they contain within them no nested groups, they would be considered genealogical species. This is encouraging because it means that widely accepted modes of speciation (e.g., allopatric or parapatric) can generate exclusive taxa without passing through the extreme bottlenecks required under the time-extended view.

The problem with the time-limited definition of exclusivity is that exclusive taxa (species) will also emerge through chance patterns of mating within panmictic populations. Figure 2 illustrates this situation using a hypothetical reticulate genealogy of a population of hermaphroditic annuals. The organisms living in the current generation would be assigned to five exclusive groups (A, B, C, D, and E). For each group one MRCA is labeled, and it can be seen that this MRCA lived closer to the present than the ancestor shared with any contemporaneous organisms outside the group. Of these exclusive groups, A, B, and D are basal, exclusive groups in that they contain no nested exclusive groups. Thus, if exclusivity is defined using recency of common descent under the time-limited view, then A, B, and D would have to be considered genealogical species. However, to elevate such small and closely related groups of organisms to the rank of species clearly would be of little practical value. Further, we maintain that recognizing this level of genealogical structure results in placing the species taxon well below the reticulate/divergent boundary.

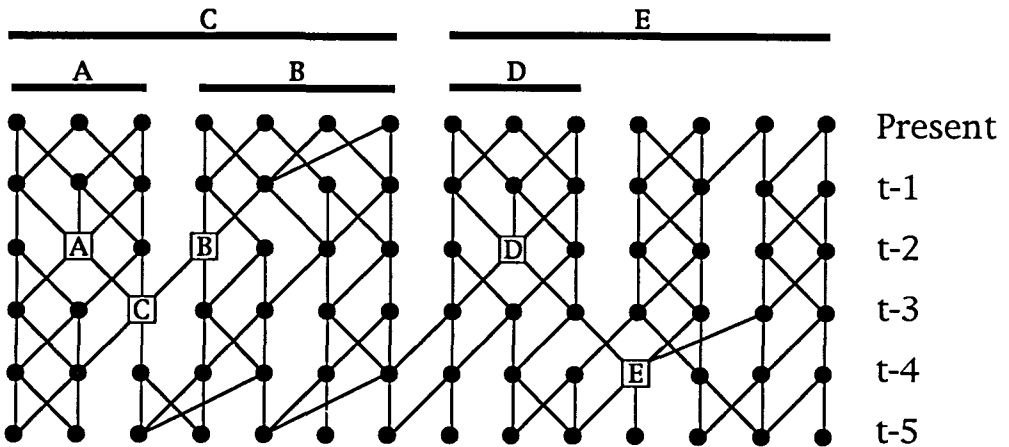


Figure 2. A hypothetical organismic genealogy for a population of biparental, hermaphroditic annuals. Time-limited exclusive groups (see text) are marked with a bar. One of the MRCAs of each exclusive group is labeled with the same letter as the group.

How can we retain the exclusivity criterion without having to recognize species at such low taxonomic levels? In other phylogenetic species concepts this problem (or analogous problems) was solved by invoking current or future interbreeding (as opposed to past interbreeding, i.e., reticulate genealogy) as a secondary criterion of species delimitation. Cra-craft (1983, 1989), Nixon and Wheeler (1990), Davis and Manos (1991), and Davis and Nixon (1992) use a reproductive cohesion criterion (or some variant thereof) to individuate the units within which diagnostic character variation is indicative of a "parental pattern of ancestry and descent." Mishler and Brandon (1987) employ current interbreeding as one of several criteria that permit the assignment of monophyletic groups to the rank of species, whereas de Queiroz and Donoghue (1988) suggest that biological populations be terminal units in the search for monophyletic species.

Despite a long-standing tradition within evolutionary biology of incorporating interbreeding criteria into species concepts, there are good reasons for rejecting it here. As discussed earlier, we are interested in a species taxon that is a product of evolution and therefore this taxon must only be defined on the basis of history (Lidén and Oxelman, 1989; Kluge, 1990; Lidén, 1992). Thus incorporating information on the inferred fate of groups of organisms into a species definition would undermine our most basic assumptions about the nature of taxa.

The need for an *ad hoc* criterion (such as interbreeding) to elevate the level at which species emerge is an inevitable consequence of defining exclusivity in terms of recency of common organismic descent. We do not suggest rejecting the exclusivity criterion as it is an important and powerful concept in phylogenetic theory. Instead, the problems encountered can be overcome by retaining exclusivity but using a more appropriate measure of genealogical relatedness.

Recency of common organismic descent as a measure of relatedness focuses on a level of biological organization that does not manifest reticulation and divergence. This, we believe, is why it causes genealogical species to emerge above (time-extended view) or below (time-limited view) the reticulate/divergent boundary. An organismic measure of relatedness implicitly ignores the fact that many of the genes found in a group of organisms do not come from their MRCA. Many ancestors other than the most recent contribute to a group's genetic make-up and should be considered in the evaluation of relatedness. This mosaic pattern arises because sexual reproduction involves recombination: the reshuffling of gene combinations from generation to generation. Indeed, it is recombination, not simply biparental reproduction, that is the cause of reticulation. By focusing on organismal MRCAs, genetic reshuffling is undetectable and therefore reticulation is obscured.

These problems can be overcome if we change our focus from organismal to genic genealogy. Whereas the previous focus on organismic pedigrees conceals the mosaic history of genomes, an approach based on gene-history permits us to detect and more accurately explore the boundary between reticulation and divergence. Until recently we lacked a terminological and conceptual framework in which individual gene-histories could be divorced from the pedigrees of the organisms containing them. However, the recently developed branch of population genetics known as coalescent theory offers a framework in which individual and collective gene history can be discussed.

Exclusivity of Coalescence. Coalescent theory is an actively developing branch of population genetics concerned with tracing the genealogical history of a present-day sample of genes (see, Hudson [1990] and Maddison [1995] for useful introductions). A "gene" is defined in coalescent theory as a non-recombining genetic element (for example, the chloroplast genome could be a gene). Since genes show no recombination, by definition, gene-genealogies will always be divergent even if the genes are sampled from within a sexually reproducing population. Thus, if you take a sample of genes and trace their lineages backward through time there are successively fewer ancestral genes in previous generations. Eventually if you go far enough back in time, the genic lineages leading to the sampled genes will "coalesce" to a single lineage.

Coalescent theory suggests an alternative means to evaluate relatedness. If an homologous (orthologous) gene is sampled from two organisms one can ask: How far back in time do these genic lineages coalesce? The depth of this coalescence is a measure of relationship between the sampled genes. The further back in time coalescence occurs, the more distantly related they are. Before incorporating this measure of relatedness into our definitions of taxa and genealogical species, it is important to consider how the coalescent histories of different genes from the same set of organisms are related in both divergent and reticulate genealogies.

In a divergent organismic genealogy, such as that generated by asexual reproduction, all parts of the genome have a genealogical history that exactly matches the organismal pedigree (i.e., coalescence will occur in the MRCA). Due to the lack of recombination the whole genome can be considered a single gene (as defined above) and therefore all parts of the genome will have concordant patterns of coalescence. In contrast, the coalescent histories of genes sampled from within a sexually reproducing population will not match the organismal pedigree. Furthermore, although each gene has a divergent genealogy, different genes (alleles) within a locus and genes from different, unlinked loci may have discordant genealogies. Consequently, different genes may give conflicting estimates of the degree of relatedness between the organisms from which they are sampled. For example, a pair of siblings will be judged very closely related if at the locus considered they each received the same alleles from their parents. However, if the locus sampled is one for which they received different alleles, the siblings might be judged very distantly related.

These examples illustrate the general fact that genes will have concordant genealogical histories when the organismal genealogy is divergent, but discordant histories when organismal genealogy is reticulate. As discussed by Avise and Ball (1990), genetic concordance thus provides an objective, measurable criterion for determining whether a genealogy is reticulate or divergent. These insights allow us to redefine exclusivity using recency of coalescence: A group of organisms is exclusive if their genes coalesce more recently within the group than between any member of the group and any organisms outside the group.

To see the significance of this definition of exclusivity consider how exclusive groups (i.e., taxa) emerge. Imagine a single panmictic population. The genealogical relationships of organisms are reticulate (as in Figure 2), such that each individual organism receives a mixture of genes from its two parents. Because each gene has its own history, the genes at different loci sampled from the same pair of organisms can suggest quite different estimates of these organism's relatedness. Now imagine that an extrinsic barrier to reproduction arises which splits this population into two. Initially, a gene sampled from an organism in one descendant population is as likely to show more recent coalescence with genes in the other population as with genes in its own population. After a time, however, dependent primarily on the effective population size, some genic lineages will go extinct in one or the other (or both) descendant populations. Eventually, genes sampled from within the populations will be found to coalesce after the ancestral population split. If the two populations are isolated long enough all gene-genealogies will be concordant (as pointed out by Avise and Ball [1990]) and will indicate that the populations are exclusive. These populations then would satisfy the criteria for being basal exclusive taxa and would constitute genealogical species. However, it should be emphasized that genes at different loci will coalesce in different ancestral organisms and not all of these organisms need have been living at the same time.

Depending on the mode of inheritance and the ploidy of the genes under consideration, different genes within a genome will have different effective sizes (Templeton, 1987) and consequently different expected rates of coalescence. For example, X chromosome specific genes (in an XY or XO sex determining system) will coalesce more rapidly than autosomal genes in diploid organisms, because their effective population size is $3/4$ that of autosomal genes. Likewise, genes favored under directional selection will coalesce more rapidly than neutral genes, whereas genes under frequency dependent or balancing selection will coalesce more slowly. We expect the depth of the reticulate/divergent border to depend on mean times to coalescence. However, even when a vast majority of the genome is concordant in supporting the exclusivity of a particular group of organisms we expect some genes (especially those under balancing or frequency-dependent selection) to show discordant patterns. Furthermore, even for neutral genes there will be variation in actual times to coales-

cence due to stochastic factors. Thus we predict that genealogical species will have “fuzzy” boundaries. The theoretical implications of fuzziness are discussed in more detail below.

Our definition of the genealogical species resembles the definition of subspecies proposed by Avise and Ball (1990). However, they endorsed a monistic rather than pluralistic view of the species problem, and subscribed to the Biological Species Concept, which groups organisms on the basis of reproductive compatibility (Mayr, 1942). As discussed previously, reproductive criteria are not employed in the delimitation of genealogical species. Contrary to Avise and Ball (1990), we do not assume that basal exclusive groups will always be nested within reproductively compatible groups. It seems likely that under certain demographic situations, genealogical species will contain reproductively incompatible sub-groups which are not themselves exclusive. Hence, a reproductive grouping criterion may delimit either more or less inclusive groups than the genealogical criterion presented here. Furthermore, a group delimited on the basis of reproductive compatibility could show partial overlap with a group defined on the basis of exclusivity of coalescence. How reproductive and genealogical groups compare and the degree to which they are coextensive is currently an unexplored question and an exciting subject for future research.

PRACTICAL IMPLEMENTATION

To make the coalescent approach more concrete, Figure 3 gives a hypothetical example of how genealogical species can be studied in nature. Four rapidly evolving, unlinked loci have been sampled (e.g., mitochondrial DNA, chloroplast DNA, and two independently assorting nuclear genes) and one allele has been sequenced from each of eleven individuals in two different geographical populations (a–e from one, f–k from the other). The DNA sequences are used to generate four separate gene trees, each assumed to represent the coalescent histories of one of the genes.

In order to delimit exclusive groups of organisms based on these separate gene trees it is necessary to look for points of concordance among the gene trees. This can be achieved using a strict consensus⁶ of the gene trees, treating the genes from the same organism as equivalent terminal taxa (Figure 3). Resolved nodes in the consensus tree indicate genetic concordance, whereas unresolved nodes indicate discordance. In the worked example, the consensus tree suggests that organisms f–k are an exclusive, coalescent group. These data would thus support the hypothesis that the population comprising organisms f–k constitutes a genealogical species. In contrast, the data suggest that the population containing a–e does not constitute a genealogical species. Indeed, organisms a–e cannot be assigned to any species as the smallest exclusive group containing them (a–k) is not basal (it contains the exclusive taxon f–k). Such a finding would suggest that a–e constitutes a “metaspecies” (see Donoghue, 1985; Mishler and Brandon, 1987; de Queiroz and Donoghue, 1988, 1990a; Kluge, 1989), which we would define as: a non-exclusive assemblage of organisms that can only be assigned to a taxon that contains within it at least one genealogical species. Unlike some authors, we employ the term metaspecies to describe an ontological situation (organisms that are not members of any species) rather than an epistemological one (groups that cannot be assigned to recognized species due to a lack of evidence). Accordingly, if organisms a–e are members of a metaspecies then even with complete knowledge of all organismic and genic relationships, these organisms could not be assigned to any genealogical species. A more extensive discussion of metaphyly is beyond the scope of this paper. However, we argue that

⁶Further work is needed to determine what methods apart from the strict consensus might be appropriate for inferring the boundary between genetic concordance and discordance. This is a particularly acute problem when more than one allele per locus from a single organism is included in which case the strict consensus method in its traditional form cannot be applied.

some pieces of the genealogy of life will inevitably be left over when we carve out genealogical species, and that these pieces must be terminologically distinguished so as to avoid them being ascribed the properties of exclusive taxa.

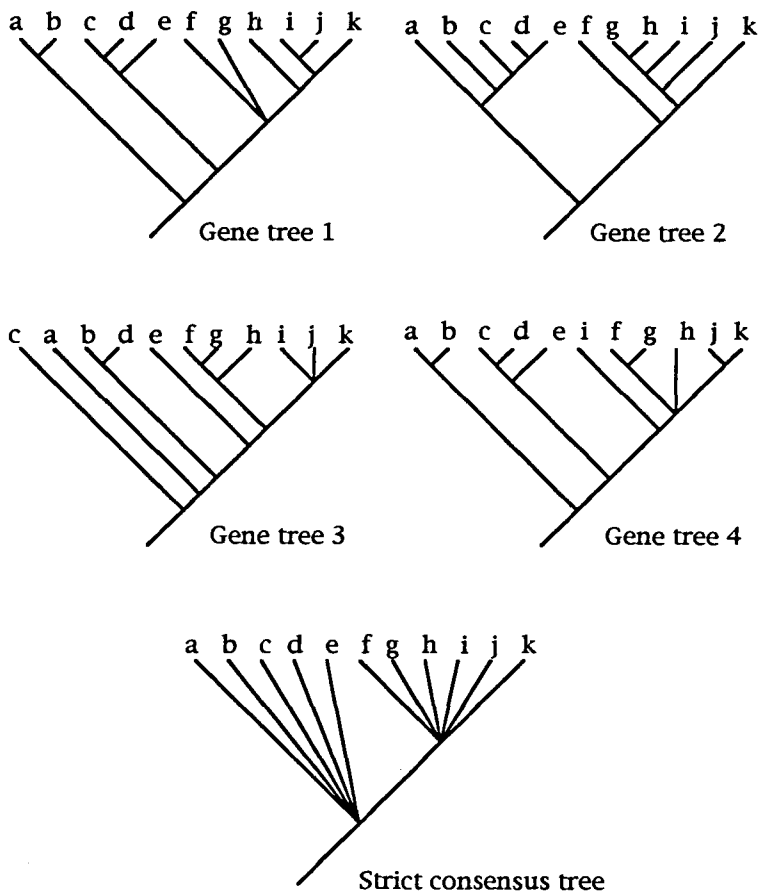


Figure 3. A worked example of the implementation of the genealogical species concept. Eleven organisms (a–k) are sampled from two geographic populations (a–e; f–k). From each individual, four unlinked genes are sequenced and gene trees constructed using standard phylogenetic procedures. Points of genetic concordance are identified using a strict consensus of these trees (see text).

An analysis such as that described in the worked example should be seen as a first step in a process of reciprocal illumination. An increased sample of organisms or of genes may contradict a group's exclusivity. As with existing species concepts, the decision as to when a sufficiently large sample has been studied will, in practice, rest with the individual systematist and the systematic community as a whole. Likewise, the extent to which rare non-concordant gene-genealogies (such as those under balancing selection) cause the rejection of a putative species will be determined by convention rather than any objective criteria.

While this hypothetical example illustrates how the genealogical species concept can be applied in practice, it will be clear to most systematists that the degree of study involved greatly exceeds the time and resources available for most taxonomic studies. However, we argue that this is not a reason for rejecting our genealogical perspective. As discussed by Hennig (1966), Simpson (1961), and others, the taxonomic naming of a species should be

viewed as a hypothesis. In our view, this means that the role of a species concept is not to provide a convenient recipe for “discovering” species, but rather to provide a conceptual criterion against which species-hypotheses can be judged. We consider it important to have theoretically consistent criteria for testing taxon status even if these criteria cannot always be applied due to practical limitations.

In practice, systematists are unlikely to be able to undertake full-scale coalescent analysis. However, empirical research on a few model systems could be used to determine what type of phenotypic characters are good predictors of species status. For example, the presence of fixed differences between reproductively isolated populations suggests that they have been genetically isolated long enough to allow for the fixation of unique variation. However, we emphasize that although the presence of such “diagnostic” variation can be used to hypothesize that a group of organisms constitutes a genealogical species, it cannot by itself validate the group’s species status (in contrast to the “phylogenetic species concept” of Cracraft, 1983, 1989; Nixon and Wheeler, 1990; Davis and Nixon, 1992). There are many reasons why a non-exclusive group could have fixed differences from close relatives (e.g., strong directional selection on the diagnostic locus, lineage-sorting from an ancestrally polymorphic stock or the retention of a plesiomorphic character that has been superseded in close relatives). Similarly, since coalescence occurs with or without mutation, genealogical species can potentially exist without any observable diagnostic variation (though the neutral mutation rate is probably sufficiently high that genealogical species will have at least some detectable variation at the molecular level).

Another criterion that may have utility in hypothesizing that a group is a genealogical species is the presence of barriers to reproduction or different specific mate-recognition systems (*sensu* Paterson, 1985). However, the same limitations apply to these reproductive criteria as to other diagnostic variation. Reproductively isolated groups need not be distinct genealogical species (if the isolating mechanisms are under strong selection and, hence, coalesce more rapidly than the rest of the genome). Likewise, there is no reason to think that genealogical species will necessarily evolve isolating mechanisms. The difference between reproductive and non-reproductive diagnostic variation is that the former is a better predictor that the group will become exclusive in the future, even if it is not so now. This difference in future tendency is, however, irrelevant to the delimitation of genealogical species, though it is an important consideration for microevolutionary theory and for some other species concepts.

The treatment of hybridizing groups and asexual organisms are common problems for all species concepts. Hybrids cause difficulties for the genealogical species concept as well, because hybridization between species can lead to a loss of exclusivity (and hence taxon status) not just for the species involved but also for the higher level taxa in which they are nested. However, rather than sweeping hybridization under the carpet, the genealogical species concept highlights the fact that hybridization is merely a form of reticulate genealogy, and this can be analyzed from a coalescent perspective. It is therefore an empirical question whether or not genealogical species can be delimited within groups of introgressing “species” (e.g., within *Quercus* and *Helianthus*).

Asexual taxa have the same ontological status as sexual taxa in that both are exclusive groups of organisms. However in asexual organisms the absence of reticulation means they do not form genealogical species as defined here (some workers, e.g., Vrana and Wheeler [1992] have failed to recognize this important difference between sexual and asexual organisms). Although basal taxa may exist in asexual organisms (usually sib-groups) these are not comparable to genealogical species (except in so far as they are taxa). In the terminology of Donoghue (1985), asexual taxa are defined by the same “grouping criterion” as

sexual taxa (i.e., exclusivity) but lack the "ranking criteria" developed here for sexual taxa (i.e., proximity to the reticulate/divergent boundary). It may be desirable to develop a subjective ranking criterion (e.g., phenetic disjunction) to identify some asexual taxa as "species," but these would not be genealogical species.

Once recognized, genealogical species will be suitable terminal taxa for cladistic analyses. This concern has played a critical role in motivating the development of the "phylogenetic species concept" (Cracraft, 1983, 1989; Nixon and Wheeler, 1990; Wheeler and Nixon, 1990). The primary issue is that the methods of cladistics will only elucidate evolutionary history when the operational taxonomic units show non-reticulate relationships (de Queiroz and Donoghue, 1988; Nixon and Wheeler, 1990). Thus, if species are to be terminal taxa their relations must be strictly divergent. Since the genealogical species described here are related to each other by a divergent genealogy they are appropriate terminal taxa for higher-level cladistic analyses.

THEORETICAL IMPLICATIONS

Does the coalescent approach satisfy our starting premises? Our first assumption, that species are taxa, is met because all taxa including species can be viewed as exclusive groups of organisms. Our second assumption, that species reside at the boundary between reticulation and divergence, is likewise met. As our discussion has shown, only divergent genealogies generate coalescent, exclusive groups of organisms. Thus, making species basal exclusive taxa places them at the boundary between reticulation and divergence.

An important implication of the gene-genealogy approach is that exclusive groups of organisms are only meaningfully delimited among organisms living at one point in time. Thus, when relatedness is evaluated using recency of genetic coalescence, basal exclusive groups are time-limited entities (defined above). There are, nonetheless, two alternative ways of relating these basal coalescent-exclusive groups to genealogical species, differing in whether species themselves are or are not time-limited.

If genealogical species are considered to be basal coalescent-exclusive groups (as we have assumed up until now) then, clearly, they too must be viewed as time-limited entities. This approach makes species analogous to the instantaneous morphologies (semaphoronts) that make up the development pathway of organisms (Hennig, 1966). It has the advantage that species can be delimited without knowledge of the future and the only past information needed is the genealogical history that gave rise to the extant organisms. Nonetheless, viewing species (and higher taxa) as time-limited poses theoretical problems for paleontology because fossil organisms cannot meaningfully be ascribed to modern groups (see Griffiths, 1974a, 1974b). Within earlier slices of time taxa can be delimited, but in an ontological sense the taxa so delimited would not be the same as taxa existing in other slices of time (though they may comprise cross-sections of the same lineages).

If, instead of viewing genealogical species as basal exclusive groups one views them as a series of successive, basal exclusive groups connected by direct descent, genealogical species (and taxa in general) would be time-extended entities. To avoid all of life being subsumed under a single species, it is however necessary to identify criteria to mark the beginning ("birth") and end ("death") of species. Following Hennig (1966) and Ridley (1987), species can be seen as beginning at an event of cladogenesis (the point in time when a single ancestral group first breaks up into two or more descendant, basal, exclusive groups) and ending at extinction or subsequent cladogenesis. This time-extended view of species and taxa is compatible with current thinking in phylogenetic systematics (e.g., de Queiroz and Gauthier, 1990) and permits the placement of fossil organisms in extant taxa. However, it has the disadvantage that complete delimitation of species requires knowledge of genealogical rela-

tionships in previous slices of time. In addition, acceptance of a time-extended view of taxa requires major modifications to the existing taxonomic system (see Griffiths, 1974b, 1976; de Queiroz and Gauthier, 1992).

Since the ramifications of the time-extended and time-limited view for systematic theory and practice remain unclear, we do not wish to advocate strongly either one. This uncertainty does not, in our opinion, undermine the utility of the genealogical species concept. Indeed, we hope that the genealogical species concept will lead to renewed interest in temporal-extent as a factor in systematics, thereby improving our understanding of the ontological status of species and other taxa.

An implication of the genealogical species concept that conflicts with traditional systematic thinking is that species have inherently fuzzy boundaries. This fuzziness is not the result of ignorance, but is due to the different rates of coalescence of different parts of the genome (caused by a combination of stochastic and selective processes). Consequently, there is no single point in time at which the boundary between reticulation and divergence is crossed. It is important to stress that this fuzziness is not an attribute of the concept itself, which like any theory or model is an idealized abstraction, but arises in the application of the concept to real organisms. Few (if any) real objects have absolute boundaries if they are observed carefully enough. Nonetheless, provided the boundary around an entity corresponds to some (however gradual) discontinuity, a conceptual description of the entity can have great heuristic value (Simpson [1961] makes a similar argument with regard to his evolutionary species concept). For example, the biological population is an important concept in microevolutionary theory despite lacking a sharp boundary (in this case expressed as probability of gene exchange). Likewise, the fuzzy boundary of the genealogical species may clash with systematic practice but it does not threaten the concept's theoretical value in macroevolutionary theory.

The genealogical species concept provides a point of contact between the fields of population biology and systematics. In defining this entity we have employed models of gene-genealogy developed by coalescent theory, a branch of population genetics, to describe and define taxa, the basic units of the nested hierarchy studied in systematics. The approach therefore provides a conceptual link between the process of genetic isolation and the hierarchical structure of nature. This can be seen clearly by the important empirical questions that the genealogical species concept poses. Just as the biological species concept (Mayr, 1942) focused attention on reproductive isolating mechanisms, the genealogical species concept focuses attention on the mechanisms by which divergent (phylogenetic) patterns of relationship emerge out of reticulating patterns. The combined insights of coalescent theory and modern phylogenetic systematics allow the problem to be clearly stated and, we believe, permit the development of an empirical research program bridging both population genetics and systematics.

It is beyond the scope of this paper to contrast the genealogical species with the many published "solutions" to the species problem. Our opinion is that even if this genealogical species is accepted as being a meaningful entity, this would not make other species concepts obsolete. There are many criteria by which one can structure knowledge of biodiversity. Many of these criteria are important and, hence, the entities delimited are necessary theoretical concepts (a point stressed by de Queiroz and Donoghue, 1988). In the case of the term "species," several entities of biological importance reside at a similar biological level (Ereshefsky, 1992), the most obvious being genealogical units (genealogical species), interbreeding units (i.e., biological populations), phenetic units (i.e., morphological clusters), and ecological units (i.e., filled niches). Although we maintain that within phylogenetic systematics the genealogical species is the entity we should be concerned with, other species-

entities may be relevant in different theoretical contexts. It is therefore important to dismantle the multiple expectations biologists have of "species" and characterize each distinct species-entity independently of the others. We can then evaluate empirically whether these different entities are coextensive in nature. If they match entirely then we can return to a single, universal "species" concept. If, as is much more likely, their boundaries match under only a limited set of situations, we will be forced to decide which entities are relevant in which theoretical contexts and which is the species-entity that should be recognized in our systems of classification.

LITERATURE CITED

- Avice, J. C. and R. M. Ball. 1990. Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford Surv. Evol. Biol.* 7: 45–67.
- Baum, D. A. 1992. Phylogenetic species concepts. *Trends Ecol. Evol.* 7: 1–2.
- Cracraft, J. 1983. Species concepts and speciation analysis. *Curr. Ornith.* 1: 159–187.
- _____. 1989. Speciation and its ontology: The empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. Pp. 28–59 in D. Otte and J. A. Endler (eds.), *Speciation and its Consequences*. Sinauer Press, Sunderland, MA.
- Davis, J. I. and P. S. Manos. 1991. Isozyme variation and species delimitation in *Puccinellia nuttalliana* complex (Poaceae): An application of the phylogenetic species concept. *Syst. Bot.* 16: 431–445.
- _____. and K. C. Nixon. 1992. Populations, genetic variation, and the delimitation of phylogenetic species. *Syst. Biol.* 41: 421–435.
- Donoghue, M. J. 1985. A critique of the biological species concept and recommendations for a phylogenetic alternative. *Bryologist* 88: 172–181.
- Ereshefsky, M. 1992. Eliminative pluralism. *Philos. Sci.* 59: 671–690.
- Frost, D. R. and D. M. Hillis, 1990. Species in concept and practice: Herpetological applications. *Herpetologist* 46: 87–104.
- Gregg, J. R. 1954. *The Language of Taxonomy; an Application of Symbolic Logic to the Study of Classificatory Systems*. Columbia Univ. Press, New York.
- Griffiths, G. C. D. 1974a. On the foundations of biological systematics. *Acta Biotheor.* 23: 85–131.
- _____. 1974b. Some fundamental problems in biological classification. *Syst. Zool.* 22: 338–343.
- _____. 1976. The future of Linnaean nomenclature. *Syst. Zool.* 25: 168–173.
- Hennig, W. 1966. *Phylogenetic Systematics*. Univ. Illinois Press, Urbana, IL.
- Holsinger, K. E. 1984. The nature of biological species. *Philos. Sci.* 51: 293–307.
- Hudson, R. R. 1990. Gene genealogies and the coalescent process. *Oxford Surv. Evol. Biol.* 7: 1–44.
- Kitcher, P. 1984. Species. *Philos. Sci.* 51: 308–333.
- Kluge, A. G. 1989. Metacladistics. *Cladistics* 5: 291–294.
- _____. 1990. Species as historical individuals. *Biol. & Philos.* 5: 417–431.
- Lidén, M. 1992. Species—where's the problem? *Taxon* 41: 315–317.
- _____. and B. Oxelman. 1989. Species—pattern or process? *Taxon* 38: 228–232.
- Maddison, W. 1995. Phylogenetic histories within and among species. Pp. xx–xx in P. C. Hoch and A. G. Stevenson (eds.), *Experimental and Molecular Approaches to Plant Biosystematics*. Monogr. Syst. Bot. Missouri Bot. Gard. 53.
- Mayr, E. 1942. *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. Columbia Univ. Press, New York.

- McKittrick, M. C. and R. M. Zink. 1988. Species concepts in ornithology. *Condor* 90: 1–4.
- Mishler, B. D. and R. N. Brandon. 1987. Individuality, pluralism, and the phylogenetic species concept. *Biol. & Philos.* 2: 397–414.
- Nelson, G. 1989. Cladistics and evolutionary models. *Cladistics* 5: 275–289.
- Nixon, K. C. and Q. D. Wheeler. 1990. An amplification of the phylogenetic species concept. *Cladistics* 6: 211–223.
- Paterson, H. E. H. 1985. The recognition concept of species. Pp. 21–29 in E. S. Vrba (ed.), *Species and Speciation*. Transvaal Museum, Pretoria.
- de Queiroz, K. 1992. Phylogenetic definitions and taxonomic philosophy. *Biol. & Philos.* 7: 295–313.
- _____ and M. J. Donoghue. 1988. Phylogenetic systematics and the species problem. *Cladistics* 4: 317–338.
- _____ and _____. 1990a. Phylogenetic systematics and species revisited. *Cladistics* 6: 83–90.
- _____ and _____. 1990b. Phylogenetic systematics or Nelson's version of cladistics? *Cladistics* 6: 61–75.
- _____ and J. Gauthier. 1990. Phylogeny as a central principle in taxonomy: Phylogenetic definitions of taxon names. *Syst. Zool.* 39: 307–322.
- _____ and _____. 1992. Phylogenetic taxonomy. *Annual Rev. Syst. Evol.* 23: 449–480.
- Ridley, M. 1987. The cladistic solution to the species problem. *Biol. & Philos.* 4: 1–16.
- Simpson, G. G. 1961. *Principles of Animal Taxonomy*. Columbia Univ. Press, New York.
- Templeton, A. R. 1987. Genetic systems and evolutionary rates. Pp. 218–234 in K. S. W. Campbell and M. F. Day (eds.), *Rates of Evolution*. Allen and Unwin, Boston.
- _____. 1989. The meaning of species and speciation: A genetic perspective. Pp. 3–27 in D. Otte and J. A. Endler (eds.), *Speciation and its Consequences*. Sinauer Press, Sunderland, MA.
- Vrana, P. and W. Wheeler. 1992. Individual organisms as terminal entities: Laying the species problem to rest. *Cladistics* 8: 67–72.
- Wheeler, Q. D. and K. C. Nixon. 1990. Another way of *looking* at the species problem: A reply to de Queiroz and Donoghue. *Cladistics* 6: 77–81.
- Wiley, E. O. 1978. The evolutionary species concept reconsidered. *Syst. Zool.* 27: 17–26.
- Wilkinson, M. 1990. A commentary on Ridley's cladistic solution to the species problem. *Biol. & Philos.* 5: 433–446.