Journal of Classification 17:165-173 (2000)
DOI: 10.1007/s003570000015

*Special Section on Reticulate Evolution*

# Reticulation inside the Species Boundary

Peter E. Smouse

Rutgers University

**Abstract:** Subspecific evolution is thought to be governed by the same radiating processes that govern supraspecific evolution, but evolution within a species generally entails substantial amounts of genetic exchange (and reticulation) within the taxon. Subspecific trees based on a strictly radiating model can be a serious distortion of the evolutionary process. The traditional insistence on using tree methods is based more on algorithm availability than on process reality. We deploy a number of devices to avoid dealing with reticulation, one of which is to distinguish between *gene trees* and *species trees*. We construct different lineage histories for different genes, in spite of the fact that intragenic recombination ensures that building a *gene tree* can become an exercise in averaging over disparate (and reticulating) segmental phylogenies. Combining data across disparate *gene trees* leads to an average *species tree*, but whether that represents anything real is dubious. Another ploy is to study mitochondrial and/or chloroplast genomes, confidently asserted to be inherited in strictly lineal fashion, without recombination. Evidence is mounting, however, that even these organellar elements have recombination and that their phylogenies are reticulate. Given the generally reticulate process of evolution at the subspecific level, we should model the collection of relationships more as a redundant and multiply connected network than as a strictly radiating phylogeny.

**Keywords:** Biogeography; Dispersal; Homoplasy; Host-parasite relationships; Reticulogram.

---

Author's Address: Department of Ecology, Evolution & Natural Resources, Cook College, Rutgers University, New Brunswick, New Jersey 08901-8551, USA; e-mail: smouse@aesop.rutgers.edu

## 1. Reticulation within a Species

The attraction of viewing subspecific evolution in strictly phylogenetic terms is a natural outgrowth of the view that "evolution is evolution", at whatever level. The same processes of genetic radiation that govern the separate evolution of two or more supraspecific taxa also govern the process by which they have arrived at the state of being "separate species". The difficulty is that while the recognizable similarity that defines the taxa could (in principle) be the sole result of evolutionary radiation (phylesis), this similarity also generally entails some level of ancestral genetic exchange within each taxon, either in the form of meiosis, syngamy, and recombination (in sexually reproducing species), or in the form of episodic genetic transfer of parts of the genome, as in many bacteria and viruses (Sneath 2000, in this Special Section). The mechanisms of genetic transfer are not the issue; the evolutionary consequences are. Subspecific evolution is only partly tree-like.

The lure of phylogenetic practice has turned the examination of subspecific evolution into a cottage industry, based on the construction of strictly radiating trees. The resulting trees sometimes reflect the process of diversification closely, but in other cases, they are almost surely a serious distortion of the actual evolutionary process. Arguments about the trees we construct all too often degenerate into a discussion of the choice of reconstruction algorithms or the data (characters) employed in tree construction, but the question of whether we should be using a tree to represent a highly reticulating process seems to be largely out of bounds. We use trees to describe subspecific evolution, not because the process is necessarily tree-like, but because we have available algorithms to model the process as tree-like. We are all comfortable with the fact that our trees are estimates, and that they almost surely contain estimation error, as long as the algorithm is working properly. Where subspecific evolution is highly reticulate, our strictly radiating trees are suspect (however, see Wang, Wakeley, and Hey 1997). Forcing an evolutionary reticulogram into a strictly radiating tree form involves some distortion; that distortion can occasionally be severe and our inferred trees positively misleading.

## 2. Lineage Sorting

One of the devices we use to finesse the reticulation problem is to make a distinction between *gene trees* and *species* trees (Avise 1989). Consider a single gene, and imagine an ancestral allele (character state) that

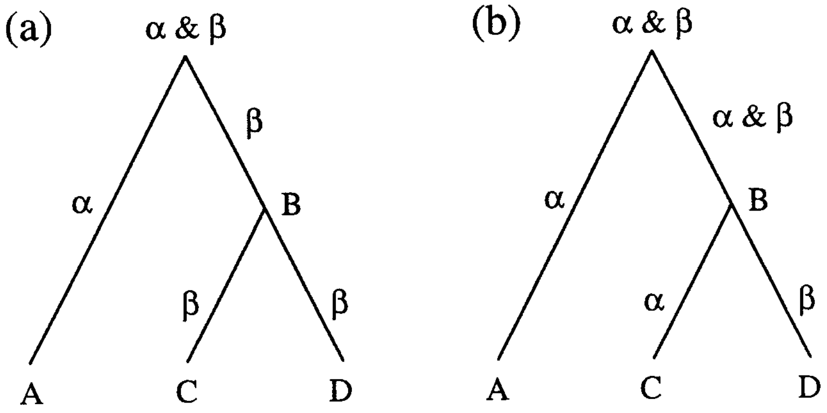

Figure 1. Consequences of lineage sorting: (a) one-to-one assortment of polymorphic allelic states (α & β) with divergent lineages; (b) lingering polymorphism within one of the lineages; reconstruction in (a) will be correct, but that in (b) will be erroneous, if based on the locus in question.

is the same in all individuals. A mutation occurs, and the population then contains two alleles/states (α and β); if both persist, the population becomes polymorphic for that gene. Eventually, perhaps even after the species has split into independently evolving lineages (**A** and **B**), one allele (α) becomes fixed (monomorphic) in species **A**, the other (β) in species **B**. The gene is said to have experienced phyletic radiation, and all is well with subsequent phylogenetic analysis, because subsequent radiation within either or both daughter species will create no inferential difficulties (see Figure 1a).

Suppose, however, that **A** becomes fixed for α, while B remains polymorphic (α and β). Now suppose that **B** subsequently splits, with fixation of the α allele in derivative species **C** and the β allele in derivative species **D** (see Figure 1b). This process of "lineage sorting" yields a data set whose subsequent phylogenetic analysis groups species **A** and **C**. The usual view is that evolution has misled us, but at least the algorithm has done what it was supposed to do, so that there is no deeper philosophical problem. Of course, this process is proceeding for thousands of genes simultaneously, and some genes (lineages) sort one way, others another. Gene by gene analysis yields contradictory trees for different genes. That situation is inferentially awkward for construction of a *species* tree, but at least the algorithm is behaving as advertised, one gene at a time.

The species tree, constructed from the pooled set of characters (genes), each with a different lineage, is inevitably an average of disparate outcomes. The question of how (indeed whether) to combine these disparate *gene trees* into a single *species tree* is still the subject of discussion. One could take the view that with enough genes, the average tendency is the *species tree*. Alternatively, one could view that average answer as representing nothing real, with the truth to be found in the collection of *gene trees*. Either way, the fact that the individual gene lineages are reticulation-free is thought to have removed us from the horns of an embarrassing algorithmic dilemma. We acknowledge the *gene tree*/*species tree* dichotomy, and let it go at that.

Most diverging species sets show incomplete reproductive isolation, particularly early in the process of phyletic radiation. Even where fixation of alternate alleles (character states) has already taken place, subsequent exchange can scramble the phyletic signal among a set of closely related taxa. Recombination among different genes within the collection of hybrid derivatives can thoroughly scramble the pattern of diversification. Viewed from a later evolutionary vantage point, the inconsistencies among a set of resulting *gene trees* resemble the results of lineage sorting. In view of the frequency of genetic exchange, early in the radiation process, the long-term evolutionary consequences of lineage sorting and reticulation are confounded. With the passage of evolutionary time, any useful distinction between them will quickly be lost in stochastic noise, which increases as we move backward in time, as also pointed out by Sneath (2000) in this Special Section.

## 3. Intragenic Recombination

The implications of recombination, either as a result of normal meiotic segregation within radiating taxa or as a result of the genetic shuffling that follows sister-taxon hybridization, have to be taken a step further. We know that recombination occurs within single genes, and that while the "real time" rate of intragenic recombination is low, the "evolutionary rate" is at least as high as that of the point mutations that provide the substrate variation for subsequent radiation. Imagine an ancestral allele ($\alpha$), of length say 100,000 nucleotides, and imagine that a pair of point mutations occur over time, yielding a trio of alleles ($\alpha$, $\beta$, and $\gamma$). These alleles are shown in Figure 2, where the locations of the relevant point mutations are shown.

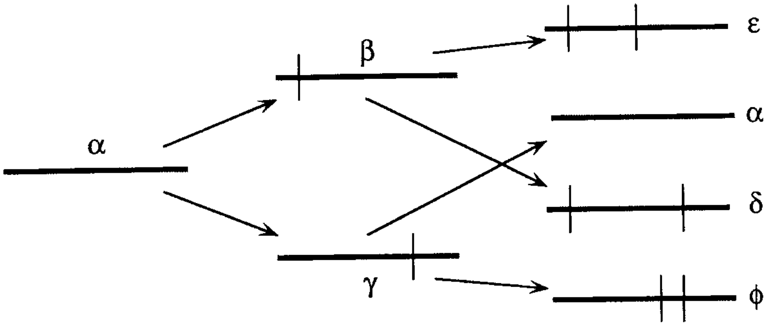

Figure 2. Evolution of allelic states under both point mutation (β, γ, ε, ϕ) and intragenic recombination (δ); α is the initial state. The crossing arrows represent intragenic recombination between the altered nucleotides at the indicated positions of the β and γ alleles.

The point is that one can generate new character states (allelic variants) solely by recombining (reticulating) the existing allelic states. The more extant states there are, the more (re)combinatorial possibilities for novelty there are. Strobeck and Morgan (1978) and Morgan and Strobeck (1979) have shown that intragenic recombination can be a more important generator of allelic (character state) novelty than point mutation itself. There is ample empirical support for the importance of such variation in the evolution of single genes (e.g., Long, Chakravarti, Boehm, Antonarakis, and Kazazian 1990). If we insist on using strictly phylogenetic approaches to evolutionary reconstruction, we are going to have to analyze small (sub-gene) sectors as units of analysis. Even building a strictly radiating *gene tree* for a whole gene is (by default) an exercise in averaging over (sometimes divergent) segmental phylogenies, and the use of a strictly branching algorithm ignores the fact that the process is instead partially (or substantially) reticulate.

## 4. Uniparental Inheritance

Population geneticists interested in subspecific evolution have gone to considerable lengths to avoid the complications of recombinational shuffling. Myriad studies have been conducted with animal mitochondrial DNA, and a smaller number with plant chloroplast DNA, both confidently asserted to show strictly uniparental inheritance. That is, to avoid recombination, we study the evolution of an organellar genome, not that of the species that carries it. We pay lip service to the distinction, but we routinely ignore it in practice. We still have to allow for lineage sorting and/or introgressive transfer of organellar genomes across species, because nuclear and organellar genes may come from different lineages, but we comfort ourselves with the thought that we can discount the possibility of recombination within the organellar genomes themselves (Smouse 1998).

It is now clear that organellar inheritance is not rigidly uniparental, and that there is enough slippage of transmission to place disparate organellar genomes in the same individual (e.g., Kondo, Satta, Matsuura, Ishiwa, Takahata, and Chigusa 1990; Zouros, Oberrhauser, Saavedra, and Freeman 1994; Ankel-Simon and Cummins 1996), where they can occasionally recombine by irregular mechanisms. The character data upon which organellar phylogenies are based include large numbers of homoplasies, usually thought to represent recurrent mutation (multiple occurrences of the same mutational change that are not monophyletic), but recent studies have shown that any reasonable rate of recurrent mutation could not lead to the observed levels of character homoplasy (see Eyre-Walker, Smith, and Maynard Smith 1999). Moreover, there is spatial autocorrelation along the organellar genome, representing adjacent sets of character-state changes that only make sense if there have been occasional recombination/rearrangement events (Awadalla, Eyre-Walker, and Maynard Smith 1999). Such events must be rare in "real time", but common enough in "evolutionary time" to provide substrate variation for evolution.

Whether the excessive homoplasy represents either mutational saturation or reticulate processes (recombination of disparate genomes), phylogenetic reconstruction is problematic. It can always be done, but whether we have anything credible at the end is none too clear (Smouse 1998), and even our most confident subspecific trees are subject to challenge.

## 5. Whither Hence?

The real question, however, is how to build reticulate reconstructions that reflect the actual evolutionary history. In this Special Section, Lapointe (2000) has much to say on this subject, but suffice it that an ideal treatment of reticulate evolution is still ahead of us. Most of our phylogenetic trees are designed to place the objects (or Operational Taxonomic Units, OTUs) at the branch tips. For supraspecific taxa, where the ancestral intermediates are extinct, the radiating tree form conveys the essential pattern of relationships

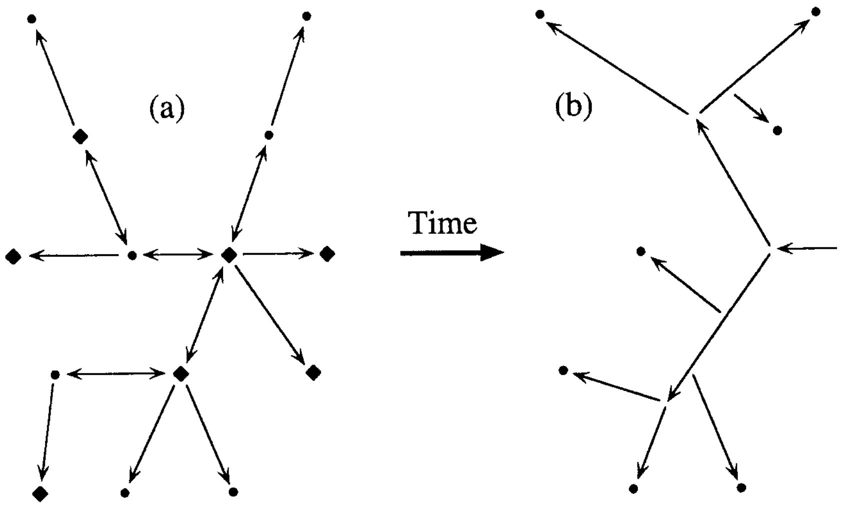

Figure 3. Development of an subspecific pattern of diversification into an interspecific phylogeny, as a consequence of extinction of the ancestral intermediates: (a) "strawberry plant" relationships of the polymorphic variants within a single species, with diamonds indicating the lineages that will become extinct; (b) radiating phylogeny, constructed from the lineages that survive evolutionary time.

among extant taxa, provided we allow for the more overt reticulation events. For subspecific phylogenetic studies, where many of the intermediate types that connect the branch-tip objects are still present within the sample, we are better served by spanning trees (Kruskal 1956; Prim 1957), where the extant objects can be either branch tips or internal nodes. Mutational or recombinational homoplasy within a species implies alternative connections, described by a *median network* (Bandelt, Forster, Sykes, and Richards 1995), with closed loops used for unresolvable homoplasic sets of connections. A median network is phylogenetically less traditional than a strictly radiating network, but it is probably a closer representation of what actually happened.

Barring closed loops, a spanning tree is like a strawberry plant (see Figure 3a). For the inner connections, directionality may be difficult to infer from the extant objects, but for the outer connections, directionality may be more obvious. Over the course of evolutionary time and phyletic radiation, the strawberry plant loses many of its nodes, particularly the innermost nodes, rendering most of the initial lineage sorting and hybridization, artifacts moot and simplifying the supraspecific structure (see Figure 3b).

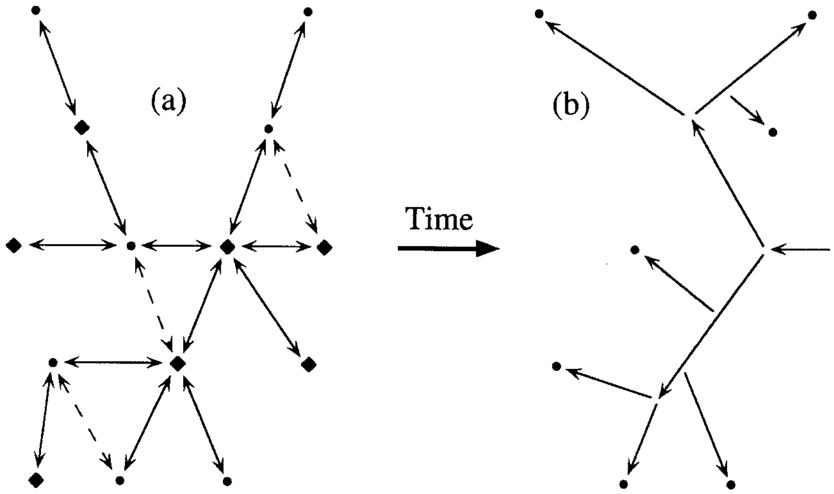

Figure 4. Development of a subspecific reticulogram into an interspecific phylogeny, as a consequence of the ancestral intermediates: (a) subspecific reticulogram, produced by recurrent mutation or reticulation, with alternate connections between adjacent internal objects indicated by solid and dashed lines; (b) radiating phylogeny, constructed from the lineages that survive evolutionary time.

Species formation may have at least as much to do with extinction of the intermediates as it does with the development of strictly radiating lineages.

Now, add to that strictly radiating network a series of closed loops, apparent homoplasies caused either by mutation, recombination, or lateral transfer of character states, and we have something more like the result shown in Figure 4a, where there are multiple connections between internal objects. As the lineages diverge into related species, the analytically awkward intermediates become extinct. Eventually, we reach a point where it is convenient to connect the surviving objects with a strictly radiating tree (see Figure 4b). There is nothing much to be gained by worrying about the precise evolutionary pathway by which an extant object has navigated the numerous changes that have occurred along any one branch; suffice it that *n* changes have occurred. Within a species, however, extinction has often not had time to erase the awkward intermediates, and we cannot realistically (nor should we attempt to) ignore them. Subspecific evolution is highly homoplasic (and frequently reticulate), and we should treat it accordingly. Standard practice has not (yet) caught up with reality.

## References

ANKEL-SIMON, F., and CUMMINS, J. M. (1996), "Misconceptions about Mitochondria and Mammalian Fertilization: Implications for Theories of Human Evolution," *Proceedings of the National Academy of Sciences USA, 93,* 13859-13863.

AVISE, J. C. (1989), "Gene Trees and Organismal Histories: A Phylogenetic Approach to Population Biology," *Evolution, 43,* 1192-1208.

AWADALLA, A., EYRE-WALKER, A., and MAYNARD SMITH, J. (1999), "Linkage Disequilibrium and Recombination in Hominid Mitochondrial DNA," *Science, 286,* 2524-2525.

BANDELT, H.-J., FORSTER, P., SYKES, B. C., and RICHARDS, M. B. (1995), "Mitochondrial Portraits of Human Populations Using Median Networks," *Genetics, 141,* 743-755.

EYRE-WALKER, A, SMITH, N. H., and MAYNARD SMITH, J. (1999), "Reply to Macaulay et al. (1999): Mitochondrial DNA Recombination – Reasons to Panic," *Proceedings of the Royal Society of London, Series B, 266,* 2041-2042.

KONDO, R., SATTA, Y., MATSUURA, E. T., ISHIWA, H., TAKAHATA, N., and CHIGUSA, S. I. (1990), "Incomplete Maternal Transmission of Mitochondrial DNA in *Drosophila*," *Genetics, 126,* 657-663.

KRUSKAL, J. B. (1956) "On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem," *Proceedings of the American Mathematical Society, 7,* 48-50.

LAPOINTE, F.-J. (2000), "How to Account for Reticulation Events in Phylogenetic Analysis: A Comparison of Distance-Based Methods," *Journal of Classification, 17,* 175-184.

LONG, J. C., CHAKRAVARTI, A, BOEHM., C. D., ANTONARAKIS, S., and KAZAZIAN, H. H. (1990), "Phylogeny of Human β-Globin Haplotypes and its Implications for Recent Human Evolution," *American Journal of Physical Anthropology, 81,* 113-130.

MORGAN, K., and STROBECK, C. (1979), "Is Intragenic Recombination a Factor in the Maintenance of Genetic Variation in Natural Populations?," *Nature, 277,* 383-384.

PRIM, R. C. (1957), "Shortest Connection Networks and Some Generalizations," *Bell System Technical Journal, 36,* 1389-1401.

SMOUSE, P. E. (1998), "To Tree or Not to Tree," *Molecular Ecology, 7,* 399-412.

SNEATH, P. H. A. (2000), "Reticulate Evolution in Bacteria and Other Organisms: How Can We Study It?," *Journal of Classification, 17,* 159-163.

STROBECK, C., and MORGAN, K. (1978), "The Effect of Intragenic Recombination on the Number of Alleles in a Finite Population," *Genetics, 88,* 829-844.

WANG, R. L., WAKELEY, J., and HEY, J. (1997), "Gene Flow and Natural Selection in the Origin of *Drosophila pseudoobscura* and Close Relatives," *Genetics, 47,* 1091-1106.

ZOUROS, E., OBERHAUSER, B. A., SAAVEDRA, C., AND FREEMAN, K. R. (1994), "An Unusual Type of Mitochondrial DNA Inheritance in the Blue Mussel *Mytilus*," *Proceedings of the National Academy of Sciences USA, 91,* 7463-7467.