

Journal of Classification 17: 153-195 (2000)

Special Section on Reticulate Evolution

Pierre Legendre, Guest Editor

Pierre Legendre

Reticulate evolution: from bacteria to philosopher

Journal of Classification 17: 153-157

Peter H. A. Sneath

Reticulate evolution in bacteria and other organisms: how can we study it?

Journal of Classification 17: 159-163

Peter E. Smouse

Reticulation inside the species boundary

Journal of Classification 17: 165-173

François-Joseph Lapointe

How to account for reticulation events in phylogenetic analysis: a comparison of distance-based methods

Journal of Classification 17: 175-184

F. James Rohlf

Phylogenetic models and reticulations

Journal of Classification 17: 185-189

Pierre Legendre

Biological applications of reticulation analysis

Journal of Classification 17: 191-195

Special Section on Reticulate Evolution

Foreword

Reticulate Evolution: From Bacteria to Philosopher

Pierre Legendre

Université de Montréal

Reticulate evolution refers to evolutionary processes that cannot fully be represented by the classical (bi)furcating tree model. In a provocative paper in *Science*, Doolittle (1999) stressed the importance of reticulate evolution in the form of lateral gene transfer (defined below) for the evolution of bacteria and higher groups of organisms.

During my years of formation as an evolutionary biologist, I came across such phenomena as hybridization and allopolyploidy, which did not fit and could not be represented using the classical evolutionary tree model. At the time, these phenomena were considered to be exceptions in the evolutionary process rather than the rule. In the mid-1970s, two important contributions appeared which demonstrated the generality of reticulation as an evolutionary process: Sneath (1975) summarized the biological evidence from various fields and showed for the first time how reticulate evolution could be represented using modified cladograms, while Sonea and Panisset (1976, 1981) showed that lateral gene transfer was a very general method of

Author's Address: Département de sciences biologiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec H3C 3J7, Canada; e-mail: Pierre.Legendre@umontreal.ca

[**Editor's Note:** I am grateful to Professor Pierre Legendre for accepting the invitation to initiate and coordinate the project that led to this Special Section. — Phipps Arabie]

evolution among bacteria. Sonea and Mathieu (2000) have reviewed the more recent evidence from the world of prokaryotes.

The biological reticulate evolution concepts developed during the 1970's and 1980's awaited the development of mathematical methods for the reconstruction of reticulograms; a reticulogram is a network capable of representing a reticulate evolutionary structure. Without them, the concepts could not be implemented using, in particular, the molecular data that are becoming increasingly available. Such methods have started to appear in the numerical classification literature; they include statistical methods for reconstructing reticulate patterns as well as methods for testing the goodness-of-fit of reticulate models. In June 1993, during the joint meeting of the Classification Society of North America and the Numerical Taxonomy Group held at the University of Pittsburgh, reticulate models for evolution were presented by Philippe Casgrain, John A. Hartigan, and Arthur R. Lee. The papers included in this Special Section discuss the biological concepts that form the foundation of this type of analysis as well as the methods presently available for the reconstruction of reticulograms.

Reticulate patterns of relationships are found in nature in the following phylogenetic situations:

- (1) In bacterial evolution, lateral gene transfer (LGT) is the mechanism by which bacteria can exchange genes across "species" through a variety of mechanisms which are described in a contribution to this Special Section written by Sneath (2000). Lateral gene transfer can be studied either in the deep phylogeny, as in the classical work of Margulis (1981) which summarized the endosymbiont hypothesis (Figure 1; see also Doolittle 1999, Figure 3), or in presently evolving groups.
- (2) Reticulations appear as the result of allopolyploidy in plants, which leads to the instantaneous appearance of a new species possessing the chromosome complement of its two parent species.
- (3) Reticulate evolution also appears during micro-evolution within species, in sexually reproducing eukaryotes, after genetic differentiation of allopatric populations followed by gene exchange through migration. This point is developed in the contribution of Smouse (2000) to this Special Section.
- (4) Homoplasy, which is the portion of phylogenetic similarity resulting from evolutionary convergence (i.e., parallel evolution and reversals), can be represented by reticulations added to a phylogenetic tree.

Reticulate patterns are also found in such non-phylogenetic problems as:

- (5) Host-parasite relationships involving host transfer.
- (6) Vicariance and dispersal biogeography.

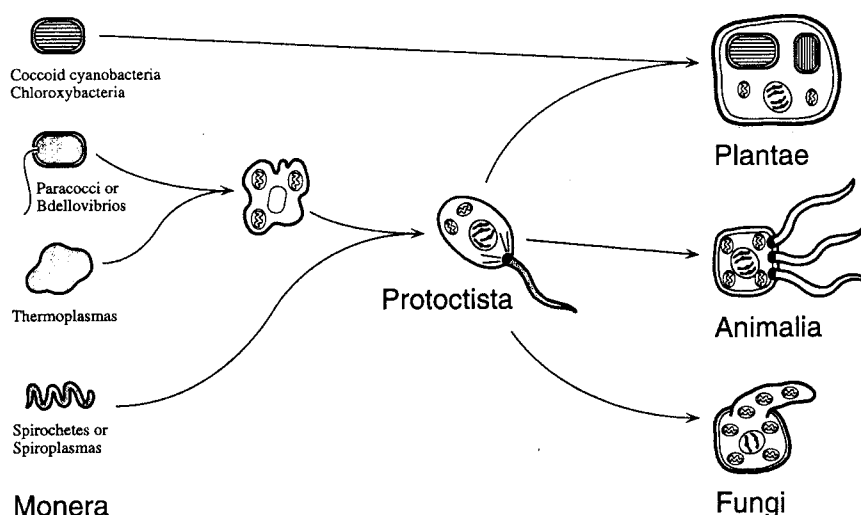


Figure 1. The endosymbiosis hypothesis describing the origin of the eukaryotic cells (i.e., having well-defined nuclei), found in the Protoctista, plants, animals, and fungi, by symbiosis. The five kingdoms of living organisms are in large print. The Monera, or Prokaryotes, have no true nucleus; they comprise the bacteria, blue-green algae, spirochaetes, etc. The algae, protozoa, slime molds, etc., are members of the Protoctista. Modified from Margulis (1981, Figure 1-1).

A reticulogram is a type of graph capable of representing relationships among organisms that may have more than one path from an organism to another; such a structure, which contains cycles, cannot be represented by a phylogenetic tree, which is acyclic by definition. Actually, phylogenetic trees are particular cases of reticulograms and include the extra property that the path from the root to any object is unique. Various numerical methods have been proposed in the literature to represent reticulated phylogenies and other non-tree-like relationships. Lapointe (2000) describes some of these methods in this Special Section, showing and contrasting for biologists and statisticians how these methods variously model and represent the data.

The need for reticulate evolution studies is discussed from the biological point of view in Rohlf's (2000) contribution to this Special Section. Reticulations can be regarded from a strictly mathematical perspective: they represent a way of modifying a tree model in such a way as to obtain a better approximation of the data. Reticulograms may reveal contradictory information that emerges from the data when they are forced

into a tree topology. In some problems, these two points of view are the main and eventually the only interpretations that can be given to reticulations added to a tree. On the other hand, in problems where one can reasonably assume that the process that has led to the generation of the data is tree-like (e.g., evolution), reticulations can be interpreted as representing either actual cases of gene exchange among non-related lineages (or branches) of the tree, or the result of homoplasy, which is the portion of phylogenetic similarity resulting from convergence. A third category of problems exists (e.g., ecological biogeography) where a tree is a good model for the main data generation process, but other processes can be assumed to have acted concomitantly; in the end, the model sought for the data may contain a tree-like portion plus additional edges. Some of these ideas are developed in the contribution of Legendre (2000) to this Special Section.

Studies of reticulate evolution have been limited, until recently, by (a) the lack of interest of biologists in the algorithms developed by methodologists to reconstruct reticulograms, (b) the lack of reasonable evolutionary models for reticulation analysis, and (c) the lack of goodness-of-fit criteria indicating whether or not a reticulate model is better, in some sense, than a non-reticulated phylogenetic tree nested within it. As long as such criteria are not available, researchers are not likely to consider the reticulate evolutionary model, even in situations where reticulate gene exchange mechanisms are known to exist.

Reticulate patterns of evolution pose a new challenge to evolutionary biologists who have been trained, after Darwin, to believe that the evolution of life could conveniently be summarized and modeled by a branching structure. The existence of reticulated patterns confronts this belief, with two consequences: on the one hand, evolutionary biologists hesitate to study the reticulated facet of evolution because they are reluctant to abandon the paradigm in which they have been trained; on the other hand, those who would like to do so lack an alternative set of tools to represent this new facet of life. In 1998, Smouse questioned the use of trees in studies of the genetic divergence of populations of single species. He expands upon his concerns in a contribution to this Special Section. In a recent paper, Doolittle (1999) pinpointed some important philosophical aspects of the problem, reminding us that LGT cannot be dismissed as trivial and that if the molecular phylogeneticists have failed to find the true tree of the history of life, it is "not because their methods are inadequate or because they have chosen the wrong genes, but because the history of life cannot properly be represented as a tree." The reconciliation of our paradigms and models with known phylogenetic processes should be seen as a challenge to modelers, including classification methodologists.

The objective of this Special Section is to stimulate research about reticulation methodology. Reticulation analysis is a new challenge for biologists, classification scientists, and statisticians. We need to generalize and extend the standard techniques of tree reconstruction to adapt them to reticulate structures.

In the following pages, we will describe the biological bases of reticulate patterns in phylogenies and related biological problems. We will also review the extant methods for inferring reticulate evolution and link them to the phylogenetic, biogeographic, or ecological problems that gave rise to them. By doing so, we hope that the turn of the century will be the opportunity for many scientists to participate in the effort to create the methodology needed for better understanding and representing evolutionary processes.

References

- DOOLITTLE, W. F. (1999), "Phylogenetic Classification and the Universal Tree," *Science*, 284, 2124-2128.
- 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.
- LEGENDRE, P. (2000), "Biological Applications of Reticulation Analysis," *Journal of Classification*, 17, 191-195.
- MARGULIS, L. (1981), *Symbiosis in Cell Evolution*, San Francisco, CA: Freeman.
- ROHLF, F. J. (2000), "Phylogenetic Models and Reticulations," *Journal of Classification*, 17, 185-189.
- SMOUSE, P. E. (1998), "To Tree or Not To Tree," *Molecular Ecology*, 7, 399-412.
- SMOUSE, P. E. (2000), "Reticulation Inside the Species Boundary," *Journal of Classification*, 17, 165-173.
- SNEATH, P. H. A. (1975), "Cladistic Representation of Reticulate Evolution," *Systematic Zoology*, 24, 360-368.
- SNEATH, P. H. A. (2000), "Reticulate Evolution in Bacteria and Other Organisms: How Can We Study It?," *Journal of Classification*, 17, 159-163.
- SONEA, S., and MATHIEU, L. G. (2000), *Prokaryotology — A Coherent View*, Montréal: Presses de l'Université de Montréal.
- SONEA, S., and PANISSET, M. (1976), "Pour une nouvelle bactériologie," *Revue Canadienne de Biologie*, 35, 103-167.
- SONEA, S., and PANISSET, M. (1981), *Introduction à la nouvelle bactériologie*, Montréal: Presses de l'Université de Montréal and Paris: Masson.

Special Section on Reticulate Evolution

Reticulate Evolution in Bacteria and Other Organisms: How Can We Study It?

Peter H. A. Sneath

University of Leicester

Abstract: The mechanisms of lateral gene transfer in bacteria and eukaryotes are briefly described. The methods of reconstruction of branching phylogenies (which can be represented as trees) can be modified to accommodate reticulate evolutionary patterns. These methods are often applied to molecular data. The elucidation of reticulate evolution requires data of very high quality and quantity. There is also a need for new concepts, terminology, and methods for this field.

Keywords: Bacteria; Lateral gene transfer; Molecular data; Phylogenetic tree; Reticulate evolution.

1. Lateral Transfer in Bacteria

Much is known about the lateral gene transfer (LGT) in bacteria. This knowledge is important because of the increasing spread of antibiotic resistance. More and more bacteria of medical significance are becoming resistant to antibiotics. The genes for resistance are usually situated on plasmids, which behave as accessory bacterial chromosomes. Plasmids are most commonly transferred by bacterial conjugation. In this process, two bacterial cells come into contact, and a tubular connection is established between them. Plasmids can then pass from one bacterium to the other.

Author's Address: Department of Microbiology and Immunology, University of Leicester, P.O. Box 138, University Road, Leicester LE1 9HN, England.

Antibiotic in the environment (in the patient under treatment, or else in natural conditions, as in the soil) produces strong selective pressure. Very rare recombinational events can be selected, leading to new forms of bacteria that are resistant to an antibiotic. Such recombination can occur between distantly related bacteria. Less commonly, DNA molecules can escape from one bacterium and penetrate another (bacterial transformation). Another mechanism is bacterial transduction, where genes are included in the genome of bacterial viruses (bacteriophages) and are transmitted when these viruses infect other bacteria. Many of these mechanisms have long been known; see Sneath (1975) and Sonea and Panisset (1976) for reviews.

All these mechanisms depend on selection for some genetic trait that is advantageous to the recipient bacterium. Further, the transferred genes can subsequently become incorporated into the main bacterial chromosome, and thus become permanent in the new lineage. These mechanisms are very widespread in bacteria, so that one can understand how easily single genes can be transferred laterally to a wide range of organisms. Complexes of several genes are less easily transferred, because the gene complex must first become assembled in the same plasmid, virus, or DNA fragment, but examples of such occurrences are known. Much useful information can be found in Baumberg, Young, Wellington, and Saunders (1995).

Similar mechanisms are believed to explain the much rarer transfer of genes in eukaryotes. Bacterial DNA can enter eukaryotic cells (for example in the case of crown galls in plants caused by infection by agrobacteria). Viruses can incorporate genes from their hosts, and some viruses, the retroviruses (of which the AIDS virus HIV is an example), can themselves integrate into the chromosomes of eukaryotes, and it is believed that integrated retroviruses may sometimes separate from the chromosome. Thus, viruses can, in principle, carry genes to new organisms.

Viruses may also perhaps cause transfer indirectly. Thus, plant viruses often destroy the green chloroplasts and cause chlorotic (i.e., white) shoots. One can then envisage that a chloroplast from some other plant species that is resistant to the virus could be transferred mechanically, perhaps by insects, or even by rare pollen grains that contain functional chloroplasts. In this way, one could have lateral transfer of a chloroplast.

2. The Study of Branching and Reticulate Phenomena

The advent of abundant molecular sequences has revolutionized the practice of evolutionary reconstruction. The logical problems of reconstructing a branching phylogenetic tree are not too difficult if one

accepts certain basic assumptions, such as evolutionary parsimony. On this assumption, the preferred pattern is the one for which the amount of evolution is smallest. The advent of abundant molecular sequences has revolutionized this field. The details, however, may be difficult to establish. One example of such problems is the distorting effect of rapid bursts of evolution on methods to recover the topology of the tree (Carmean and Crespi 1995). Fortunately, there are some parts of the genome that seem to reflect very well the main branching pattern and agree with fossil evidence. Such regions often consist of “housekeeping genes”, those that are essential to the functioning of the cell and which would malfunction if replaced by genes from a distantly related organism. Such genes are unlikely to be successfully transmitted laterally in reticulate evolution.

One set of such genes are those of the ribosomes, which are involved in protein synthesis. One of these, known as the small subunit, 16S or 18S gene, has been widely used for reconstructing branched trees, because of the pioneering work of Woese and his colleagues (Woese and Fox 1977; Woese 1987). Its sequence consists of about 1600 nucleotides, which is sufficient to reconstruct the major branching pattern of all living organisms, though not the finer details at the level of genera and species. There are, however, difficulties in recognizing such genes (Doolittle 1996). We still have little information on the complete genomes of organisms that could help to solve this problem, but as the knowledge grows this task could be a fertile field for study. By contrast, genes that do not show this evolutionary stability would be candidates for exploring reticulate evolution.

The elucidation of reticulate evolution requires data of very high quality and quantity, even when suitable sequences have been identified. As soon as fusions are permitted in trees, the number of alternative patterns becomes extremely large. Similar phenomena are well known in other fields; if one adds a time dimension to the spatial dimensions, the analyses become very complicated. Furthermore, there are often gene duplications which superficially look much the same as lateral transfers (Delwicke and Palmer 1996). Yet when suitable data are available, the development of new algorithmic methods will be an exciting challenge for the future.

3. Methods for Studying Reticulate Evolution

One form of reticulate evolution is crossing over within genes, whereby part of a gene from one parent exchanges its position in the offspring with the other part of the gene from the other parent; see also Smouse (2000) in this Special Section. Such exchanges are usually found

only within one species of organism; they are very frequent and constitute the commonest form of reticulate evolution below the genus level. Methods for detecting cross-overs, or similar phenomena, have been developed in recent years, and many are based on comparing two sequences of the same gene from different individuals. If evolution has been limited to branching, the differences will usually be scattered more-or-less randomly along the sequence. But if a cross-over has occurred in the immediate ancestry of the individuals, the differences will be concentrated in the region that has been involved in the crossing over. Useful references are Maynard Smith (1992) and Sawyer (1989). Sneath (1998) has discussed the influence of codon positions on the statistics. Also useful are papers of Sang (1995) and Estabrook, Sil-Ad, and Reznicek (1996).

Another approach is to reconstruct trees from various parts of the genome and to compare these trees instead of pairs of sequences. Templeton, Routman, and Phillips (1995) have extended these concepts to formal cladistic analysis of the genome. Such studies are, in principle, able to show the direction of transfer of genetic information, but there has been little study of this capability. In practice, it may be difficult to determine directionality (Thorpe 1982).

In a branching tree, the amount of genetic divergence between organisms is approximately proportional to the time to their most recent common ancestor. When reticulation is present, no such simple representation can be made (Sneath 1975). There is therefore a need for new concepts and terminology for this field. Another point is that evolutionary diagrams should have some indication of the reliability in different areas. The usual way is to give "bootstrap" values for selected phylogenetic groups. The bootstrap values often do not stand out clearly, so that a graphic method may be more informative. One such method (Sneath and Radbourne 1991) can readily be adapted to reticulate patterns where bootstrapping is less easy to apply.

It should also be remembered that evidence of past events fades over time. This fading should be considered as approximately logarithmic. Thus, the evidence of a past event can be represented as the "signal" of the event, and the remaining information can be considered "noise". If the mean signal is 50% after a thousand years, one must expect that after two thousand years the mean signal will be only 25%, and 12.5% after three thousand years. The variance will also increase similarly. How best to incorporate this principle into methodology remains to be studied, but it is important to give more evidence to recent events than those in the distant past.

References

- BAUMBERG, S., YOUNG, J. P. W., WELLINGTON, M. H., and SAUNDERS, J. R. [eds.] (1995), *Population Genetics of Bacteria*, Cambridge, U.K.: Cambridge University Press.
- CARMEAN, D., and CRESPI, B. J. (1995), "Do Long Branches Attract Flies?," *Nature*, 373, 661.
- DELWICKE, C. F., and PALMER, J. D. (1996), "Rampant Horizontal Transfer and Duplication of Rubisco Genes in Eubacteria and Plasmids," *Molecular Biology and Evolution*, 13, 873-882.
- DOOLITTLE, W. F. (1996), "Some Aspects of the Biology of Cells and their Possible Evolutionary Significance," *Symposium of the Society for General Microbiology*, 54, 1-21.
- ESTABROOK, G. F., SIL-AD, N. L., and REZNICEK, A. A. (1996), "Hypothesizing Hybrids and Parents Using Character Intermediacy, Parental Distances, and Equality," *Taxon*, 45, 647-662.
- MAYNARD SMITH, J. (1992), "Analysing the Mosaic Structure of Genes," *Journal of Molecular Evolution*, 24, 126-129.
- SANG, T. (1995), "New Measurements of the Distribution of Homoplasy and Reliability of Parsimonious Cladograms," *Taxon*, 44, 77-82.
- SAWYER, S. (1989), "Statistical Tests for Detecting Gene Conversion," *Molecular Biology and Evolution*, 6, 526-538.
- SMOUSE, P. E. (2000), "Reticulation Inside the Species Boundary," *Journal of Classification*, 17, 165-173.
- SNEATH, P. H. A. (1975), "Cladistic Representation of Reticulate Evolution," *Systematic Zoology*, 24, 360-368.
- SNEATH, P. H. A. (1998), "The Effect of Evenly Spaced Constant Sites on the Distribution of the Random Division of a Molecular Sequence," *Bioinformatics*, 14, 608-616.
- SNEATH, P. H. A., and RADBOURNE, J. C. (1991), "DOTDND: a FORTRAN-77 Program for Showing Graphically the Confidence or Uncertainty in Phylogenetic Trees," *Computers & Geosciences*, 17, 689-718.
- SONEA, S., and PANISSET, M. (1976), "Pour une nouvelle bactériologie," *Revue Canadienne de Biologie*, 35, 103-167.
- TEMPLETON, A. R., ROUTMAN, E., and PHILLIPS, C. A. (1995), "Separating Population Structure from Population History: A Cladistic Analysis of the Geographical Distribution of Mitochondrial DNA Haplotypes in the Tiger Salamander, *Ambystoma tigrinum*," *Genetics*, 140, 767-782.
- THORPE, R. S. (1982), "Reticulate Evolution and Cladism: Tests for the Direction of Evolution," *Experientia*, 38, 1242-1244.
- WOESE, C. R. (1987), "Bacterial Evolution," *Microbiological Review*, 51, 221-271.
- WOESE, C. R., and FOX, G. E. (1977), "Phylogenetic Structure of the Prokaryotic Domain: The Primary Kingdoms," *Proceedings of the National Academy of Sciences of the USA*, 74, 5088-5090.

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 (phyletic), 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* (Avice 1989). Consider a single gene, and imagine an ancestral allele (character state) that

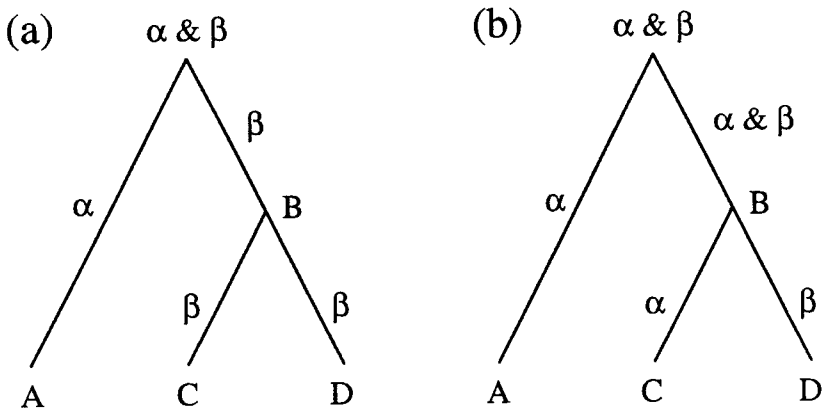


Figure 1. Consequences of lineage sorting: (a) one-to-one assortment of polymorphic allelic states (α & β) with divergent lineages; (b) lingering polymorphism within one of the lineages; reconstruction in (a) will be correct, but that in (b) will be erroneous, if based on the locus in question.

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 (α), of length say 100,000 nucleotides, and imagine that a pair of point mutations occur over time, yielding a trio of alleles (α , β , and γ). These alleles are shown in Figure 2, where the locations of the relevant point mutations are shown.

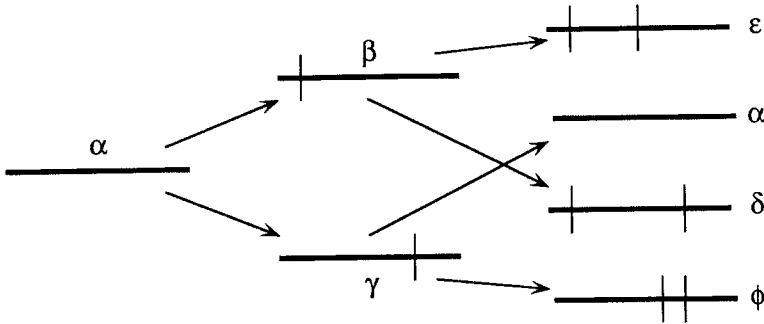


Figure 2. Evolution of allelic states under both point mutation (β , γ , ϵ , ϕ) and intragenic recombination (δ); α is the initial state. The crossing arrows represent intragenic recombination between the altered nucleotides at the indicated positions of the β and γ alleles.

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, Oberhauser, 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

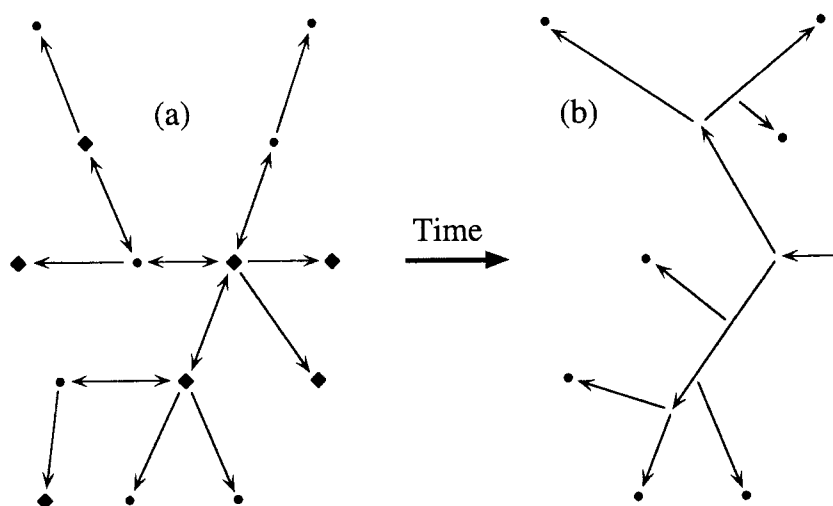


Figure 3. Development of an subspecific pattern of diversification into an interspecific phylogeny, as a consequence of extinction of the ancestral intermediates: (a) “strawberry plant” relationships of the polymorphic variants within a single species, with diamonds indicating the lineages that will 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 homoplastic 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).

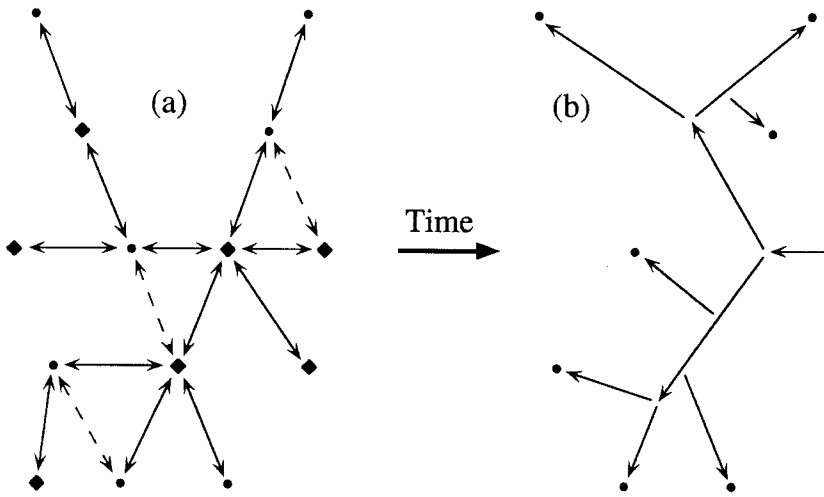


Figure 4. Development of a subspecific reticulogram into an interspecific phylogeny, as a consequence of the ancestral intermediates: (a) subspecific reticulogram, produced by recurrent mutation or reticulation, with alternate connections between adjacent internal objects indicated by solid and dashed 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 homoplastic (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.

Special Section on Reticulate Evolution

**How to Account for Reticulation Events in Phylogenetic Analysis:
A Comparison of Distance-Based Methods**

François-Joseph Lapointe

Université de Montréal

Abstract: This paper presents a review of mathematical techniques capable of representing reticulate events in phylogenetics. Two families of methods are identified; they relax either the ultrametric inequality defining dendrograms or the four-point condition defining additive trees. Pyramids and weak hierarchies are techniques developed to fit dendrograms with overlapping clusters. Splitsgraphs and reticulograms are extensions of the additive tree model; they allow one to fit a dissimilarity matrix using a graph containing reticulations. The four methods are applied to a data set; the results are compared and discussed in a phylogenetic setting.

Keywords: Additive tree; Dendrogram; General graph; Pyramids; Reticulogram; Splitsgraph; Weak hierarchy.

I am grateful to Guy Cucumel, Vladimir Makarenkov, David Bryant, and Andreas Dress for helping me understand some of the mathematical aspects of the methods presented in this paper, to Pierre Legendre for inviting me to participate in this Special Section, to Jim Rohlf for encouraging me to write this paper, and to Peter Sneath and Peter Smouse for constructive comments on a previous version. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche du Québec.

Author's Address: François-Joseph Lapointe, Département de sciences biologiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec H3C 3J7, Canada; e-mail: francois-joseph.lapointe@umontreal.ca

1. Introduction

The basic graph-theoretic model used in phylogenetic analysis is that of a tree. Sometimes, trees are rooted to represent ancestor-descendent relationships among their nodes, and sometimes, their branches are weighted to represent the amount of evolutionary change along those branches. Two different types of weighted trees are commonly used to depict evolutionary relationships among species. *Dendrograms* are used to represent rooted weighted trees in which all terminal nodes are equidistant from the root, whereas *additive trees* are used to represent unrooted weighted trees; additive trees can also be rooted by selecting one of the nodes to form the root of the tree. The distinction is important in evolutionary biology because dendrograms represent trees that satisfy the molecular clock hypothesis stating that all lineages evolved at the same rate (Figure 1a). This assumption is not always made for additive trees (Figure 1d). For the purpose of the present paper, it suffices to say that dendrograms satisfy the well-known ultrametric inequality (Hartigan 1967) and that additive trees satisfy the more general four-point condition (Buneman 1974); a more detailed presentation of dendrograms and additive trees is found in Lapointe and Legendre (1991).

A tree is not always a suitable graphical representation of the evolutionary relationships among species. In fact, it is not uncommon for species to exchange genetic material laterally instead of vertically (along the branches of the tree). These so-called reticulation events violate the branching evolutionary model by introducing cycles in a graph and causing conflicting signals in the data. Other representations must be used to depict such evolutionary phenomena, which cannot adequately be represented in the form of trees. Four such *reticulistic* techniques are described below and are applied to the same data set for comparison.

2. Pyramids

Pyramids, introduced by Diday and Bertrand (1986), are a generalization of the hierarchical clustering framework. Whereas a dendrogram can be defined as a nested set of nonoverlapping clusters (Figure 1a), pyramids represent a set of clusters that may overlap without necessarily being nested (Figure 1b). For any given pair of clusters *C* and *D* in a dendrogram *H* that have a nonempty intersection, either *C* is contained in *D*, or *D* is contained in *C*. In Figure 1a, for example, the cluster {*Pan paniscus*, *Pan troglodytes*} is contained in the cluster {*Homo sapiens*, *Pan paniscus*, *Pan troglodytes*} of *H*. In the case of a pyramidal graph *P*, the

intersection of two clusters C and D that have a nonempty intersection is always a cluster of \mathbf{P} . In Figure 1b, for example, the intersection of the clusters $\{Pan\ paniscus, Pan\ troglodytes\}$ and $\{Homo\ sapiens, Pan\ troglodytes\}$ is the singleton $\{Pan\ troglodytes\}$ of \mathbf{P} .

A dissimilarity matrix \mathbf{D} is said to be pyramidal iff \mathbf{D} is also a Robinsonian matrix (Table 1c, upper triangular). This property means that, for any triplet i, j, k , from an *ordered* set of species, the dissimilarity value d_{ik} must be larger than or equal to the maximum of d_{ij} and d_{jk} . Interestingly, an ultrametric matrix \mathbf{U} can always be permuted to form a Robinsonian matrix, so that a dendrogram actually represents a special type of pyramids with at most $n-1$ different clusters. Just like dendrograms, pyramids can be obtained by agglomerative algorithms. In Figure 1b, the pyramidal representation of the dissimilarities presented in Table 1a was obtained from the equivalent of the complete linkage algorithm: two clusters are joined at a given height if they satisfy the clustering rule and have not been aggregated *twice* before; in the case of dendrograms, two clusters are joined if they satisfy the clustering rule and have not been aggregated *once* before. By allowing species to be included in overlapping clusters, pyramids can thus be used to depict reticulation events in a set of species that can be ordered in a Robinsonian matrix. A program to compute pyramids is available at the following WWW address: <<http://genome.genetique.uvsq.fr/Pyramids/>>.

3. Weak Hierarchies

Weak hierarchies have been proposed by Bandelt and Dress (1989) to fit dendrograms with a few additional nonnested clusters (i.e., reticulations). In short, the method proceeds by creating weak clusters of species, as opposed to the so-called strong clusters found in dendrograms. From a similarity matrix \mathbf{S} , a weak cluster C is formed if any two species i and j that belong to C are more similar to each other than any other species k outside of C is similar to at least one of i and j (Bandelt and Dress 1989); that is, s_{ij} must be larger than the minimum of s_{ik} and s_{kj} for every species k which is not a member of C (for strong clusters, s_{ij} must be larger than the maximum of s_{ik} and s_{kj}). Using a set-theoretic point of view, a weak hierarchy \mathbf{W} is obtained if the intersection of any three (strong or weak) clusters C, D , and E of \mathbf{W} is equal to one of the binary intersections $C \cap D$, $C \cap E$, or $D \cap E$. For example, one can check that the similarity s_{ij} between *Homo sapiens* and *Pan troglodytes* in \mathbf{S} (where $s_{ij} = 1 - d_{ij}$; Table 1a) is larger than the lesser of the similarities s_{ik} and s_{kj} between any other species k and either *H. sapiens* or *P. troglodytes*; as a consequence, the pair $\{Homo\ sapiens, Pan\ troglodytes\}$ represents a weak cluster of \mathbf{W} . Then, because

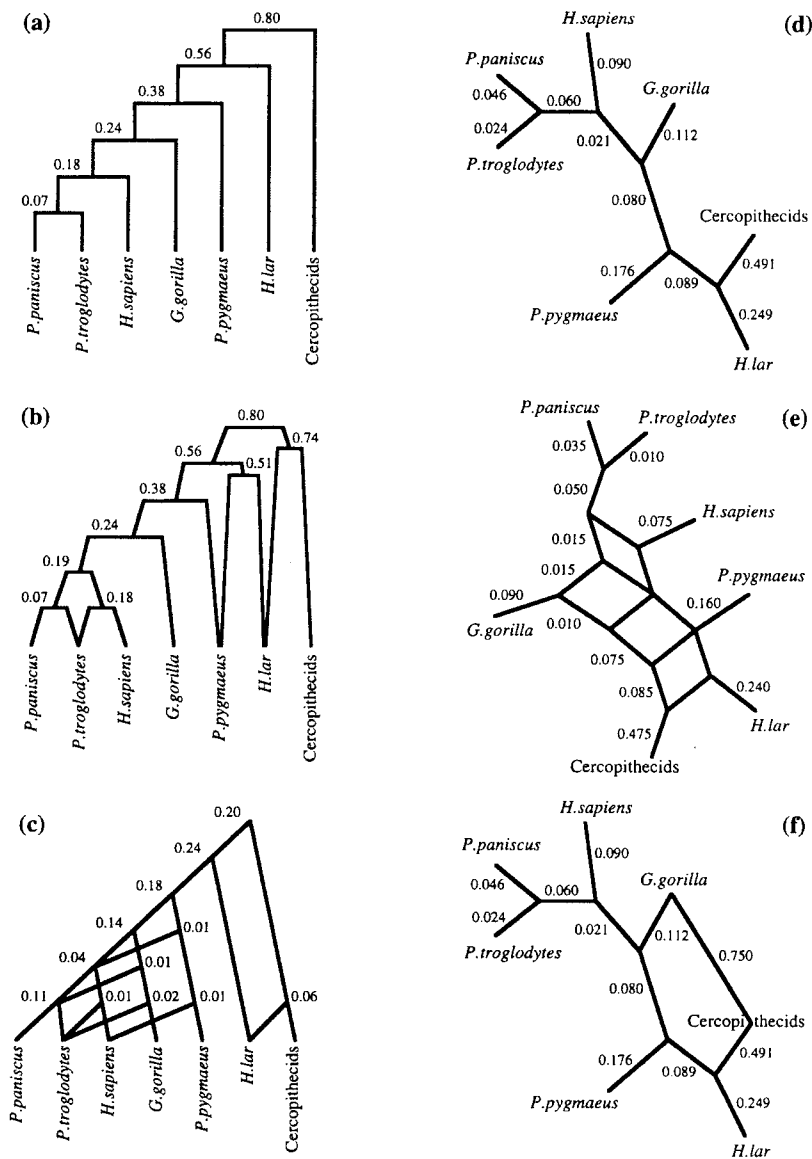


Figure 1. Different representations of the dissimilarity matrix of Table 1a. (a) A complete linkage dendrogram with clustering levels. The distance between two species is given by the height of the lowest cluster that includes these species. The corresponding ultrametric distances are presented in Table 1b (upper triangular). (b) Complete linkage pyramids with clustering levels. The distance between two species is given by the height of the lowest cluster that includes these species. The corresponding (Robinsonian) pyramidal distances are presented in Table 1c (upper triangular). (c) A weak hierarchy (*caption continues on next page*)

the intersection of the clusters $\{H. sapiens, P. paniscus, P. troglodytes, G. gorilla\}$, $\{H. sapiens, P. paniscus, P. troglodytes\}$ and $\{H. sapiens, P. troglodytes\}$ in the weak hierarchy derived from Table 1a is equal to at least one of the three binary intersections, these four species form a weak hierarchy (see Figure 1c).

A weak hierarchy is an extension of a dendrogram and represents all weak and strong clusters. Consequently, any dendrogram is a weak hierarchy, whereas pyramids are nothing but weak hierarchies with the additional property that a linear order of the species can be defined such that every cluster is an interval relative to that order. Using the clusters of a weak hierarchy, one can compose a similarity matrix additively (Table 1c, lower triangular; where $d_{ij} = 1 - s_{ij}$) by attaching a weight to each cluster and letting the similarity of a pair of species i and j be the sum of the weights of all the clusters (weak or strong) containing the pair $\{i, j\}$; see the algorithm in Bandelt and Dress (1989). Furthermore, given the weighted weak hierarchy, one can reconstruct all of its clusters as well as their respective weights from the associated similarity matrix. A complete linkage type of algorithm has been developed by Bandelt and Dress (1989) to approximate a similarity matrix \mathbf{S} by a weak hierarchy (see Figure 1c). A computer program to compute weak hierarchies can be obtained by writing to Professor H.-J. Bandelt: Mathematisches Seminar, Universität Hamburg, Bundesstrasse 55, D-20146 Hamburg, Germany.

4. Splitsgraph

As in the case of dendrograms, reticulations are not allowed in additive trees (see Figure 1d). To produce unrooted phylogenies in which

Figure 1 (*continued*) obtained by a complete linkage-type method applied to the matrix presented in Table 1a, with corresponding weights attached to the clusters. The similarity between two species is computed as the sum of the weights of all the clusters that include these species. The corresponding distances are presented in Table 1c (lower triangular), where $d_{ij} = 1 - s_{ij}$. (d) An additive tree, with edge lengths, obtained by a least-squares algorithm. The distance between two species is computed as the sum of the edge lengths along the path connecting these species. The corresponding path-length distances are presented in Table 1b (lower triangular). (e) A splitsgraph representation with edge lengths; all parallel edges have equal lengths. The distance between two species is computed as the shortest path-length distance between these species over all possible paths. The corresponding path-length distances are presented in Table 1d (upper triangular). (f) A reticulogram, with edge lengths, obtained by adding reticulations onto the additive tree presented in Figure 1d. The distance between two species is computed as the shortest path-length distance between these species over all possible paths. The corresponding path-length distances are presented in Table 1d (lower triangular). For clarity, edge lengths in the figure are not represented proportional to their actual lengths.

Table 1

a: Initial dissimilarity matrix (modified from Bandelt and Dress 1989)

	1	2	3	4	5	6	7
1. <i>H. sapiens</i>	0.0000	0.1900	0.1800	0.2400	0.3600	0.5200	0.7700
2. <i>P. paniscus</i>	0.1900	0.0000	0.0700	0.2300	0.3700	0.5600	0.8000
3. <i>P. troglodytes</i>	0.1800	0.0700	0.0000	0.2100	0.3700	0.5100	0.7700
4. <i>G. gorilla</i>	0.2400	0.2300	0.2100	0.0000	0.3800	0.5400	0.7500
5. <i>P. pygmaeus</i>	0.3600	0.3700	0.3700	0.3800	0.0000	0.5100	0.7600
6. <i>H. lar</i>	0.5200	0.5600	0.5100	0.5400	0.5100	0.0000	0.7400
7. Cercopithecids	0.7700	0.8000	0.7700	0.7500	0.7600	0.7400	0.0000

b: Distances corresponding to the dendrogram of Figure 1a (upper triangular matrix) and the additive tree of Figure 1d (lower triangular)

	1	2	3	4	5	6	7
1. <i>H. sapiens</i>	0.0000	0.1800	0.1800	0.2400	0.3800	0.5600	0.8000
2. <i>P. paniscus</i>	0.1960	0.0000	0.0700	0.2400	0.3800	0.5600	0.8000
3. <i>P. troglodytes</i>	0.1739	0.0701	0.0000	0.2400	0.3800	0.5600	0.8000
4. <i>G. gorilla</i>	0.2233	0.2393	0.2173	0.0000	0.3800	0.5600	0.8000
5. <i>P. pygmaeus</i>	0.3672	0.3832	0.3612	0.3683	0.0000	0.5600	0.8000
6. <i>H. lar</i>	0.5287	0.5447	0.5227	0.5298	0.5140	0.0000	0.8000
7. Cercopithecids	0.7707	0.7867	0.7647	0.7719	0.7560	0.7400	0.0000

c: Distances corresponding to the pyramids of Figure 1b (upper triangular matrix¹) and the weak hierarchy of Figure 1c (lower triangular²)

	1	2	3	4	5	6	7
1. <i>H. sapiens</i>	0.0000	0.1900	0.1800	0.2400	0.3800	0.5600	0.8000
2. <i>P. paniscus</i>	0.1900	0.0000	0.0700	0.2400	0.3800	0.5600	0.8000
3. <i>P. troglodytes</i>	0.1800	0.0700	0.0000	0.2400	0.3800	0.5600	0.8000
4. <i>G. gorilla</i>	0.2400	0.2300	0.2100	0.0000	0.3800	0.5600	0.8000
5. <i>P. pygmaeus</i>	0.3600	0.3700	0.3700	0.3800	0.0000	0.5100	0.8000
6. <i>H. lar</i>	0.5600	0.5600	0.5600	0.5600	0.5600	0.0000	0.7400
7. Cercopithecids	0.8000	0.8000	0.8000	0.8000	0.8000	0.7400	0.0000

¹ The matrix is Robinsonian if the species are ordered as in the pyramids {2, 3, 1, 4, 5, 6, 7}.

² The distances were obtained by subtracting the similarity values from one: $d_{ij} = 1 - s_{ij}$.

d: Distances corresponding to the splitsgraph of Figure 1e (upper triangular matrix) and the reticulogram of Figure 1f (lower triangular)

	1	2	3	4	5	6	7
1. <i>H. sapiens</i>	0.0000	0.1700	0.1450	0.2050	0.3250	0.4900	0.7400
2. <i>P. paniscus</i>	0.1960	0.0000	0.0450	0.2050	0.3450	0.5100	0.7600
3. <i>P. troglodytes</i>	0.1739	0.0701	0.0000	0.1800	0.3200	0.4850	0.7350
4. <i>G. gorilla</i>	0.2233	0.2393	0.2173	0.0000	0.3500	0.5150	0.7350
5. <i>P. pygmaeus</i>	0.3672	0.3832	0.3612	0.3683	0.0000	0.4850	0.7350
6. <i>H. lar</i>	0.5287	0.5447	0.5227	0.5298	0.5140	0.0000	0.7300
7. Cercopithecids	0.7707	0.7867	0.7647	0.7500	0.7560	0.7400	0.0000

the species manifest reticulation, the splitsgraph method of Bandelt and Dress (1992) can be used. This technique relies on split decomposition, a procedure for decomposing distances canonically into a sum of simpler metrics. For each quadruplet of species $\{i, j, k, l\}$, the algorithm looks at the three possible tree topologies that may be used to split the four species in two groups of two (i.e., ij/kl , ik/jl , and il/jk), with their corresponding sum of distances (i.e., $ij + kl$, $ik + jl$, and $il + jk$). Instead of *selecting the most probable* topology (i.e., the one with the smallest distance sum) as the estimate of the relationships, Bandelt and Dress's (1992) method *excludes the most improbable* of the three topologies at each step. The global splits (computed over all possible quadruplets) which never realize the most improbable topologies are accepted and depicted as a "splitsgraph" (Dress, Huson, and Moulton 1996).

In contrast to additive trees in which any edge splits the tree into two connected subtrees, incompatible splits cannot always be depicted by a single edge but will give rise to a series of parallel edges of equal lengths; the length of these parallel edges represents the isolation index of a given split. Therefore, a splitsgraph is a representation, composed of parallelograms plus individual edges, providing a visual representation of the support for contradictory patterns in the data (see Figure 1e). Unlike additive trees, in which the path-length distance between two species i and j is computed by adding the edge lengths along the path between these species, path-lengths in a splitsgraph are the shortest lengths of all paths from species i to j (also corresponding to the sum of all weighted splits separating two species, Table 1d, upper triangular). For example, the path length between *Pan paniscus* and *Gorilla gorilla* in Figure 1e is 0.2050, instead of 0.2393 in the additive tree of Figure 1d. A splittability index can be used to indicate the fit of the weighted system of splits, depicted as a splitsgraph, to the original dissimilarities in **D**. A computer program to compute splitsgraphs is available at the following WWW address: <http://bibiserv.techfak.uni-bielefeld.de/splits/>.

5. Additive Tree with Reticulations

Recently, Makarenkov and Legendre (submitted) have proposed an algorithm to add reticulations onto an additive tree so as to maximize the fit between the data and a reticulogram, which is an evolutionary graph in which the data may be related nonuniquely to a common ancestor (Makarenkov and Legendre 2000). This graph is computed by gradually improving the approximation of the dissimilarities as extra edges are added to the graph. Contrary to the other methods, this technique uses an

optimality criterion to determine the minimum number of reticulations required to reach a maximum fit to the data; a least-squares loss function computed as the sum of the squared differences between the original dissimilarities in **D** and the path-length distances **P** on the reticulogram is minimized. Because there is more than one way to compute the path lengths between two species i and j , the minimum path-length distance over all possible paths from i to j is recorded in **P** (see Table 1d, lower triangular). For instance, the path-length distance between *Gorilla gorilla* and the Cercopithecids in Figure 1f is the length of the reticulate edge connecting these two species (0.7500) rather than the sum of the edge lengths along the original and unique path found in the additive tree (0.7719 in Figure 1d).

Makarenkov and Legendre (submitted) described three stopping rules to determine the number of reticulations to be added to an additive tree. Criterion Q1 takes into account the value of the loss function as well as the number of degrees of freedom of the reticulogram under construction; two other criteria, Q2 and AIC, have also been proposed by those authors. A statistical procedure could possibly be implemented to assess the significance of individual reticulations, using the Q1 statistic, for a graph bearing n edges compared to one with $n-1$ edges. A program to compute additive trees and reticulograms is available at the WWW address <http://www.fas.umontreal.ca/biol/legendre/>.

6. Discussion

To produce reticulograms, it is difficult to select a single best method among the four described in this paper. Pyramids allow for overlapping clusters and can perfectly fit a dissimilarity matrix if there exists a permutation order of the species such that the dissimilarities are Robinsonian. In the case of weak hierarchies, an optimal collection of weighted weak clusters is sought to reconstruct a similarity measure that approximates the original similarities. Both methods should therefore be able to fit the dis/similarities better than a dendrogram without reticulations. When reticulograms based on extended additive trees are sought, the splitsgraph method, which detects incompatible splits in the data, can be used to obtain a graphical representation of a dissimilarity matrix. Allowing cycles in a graph produces in turn a better fit of the model to the data. One can also use the method proposed by Makarenkov and Legendre (2000) which seeks to improve the representation of a dissimilarity matrix by adding reticulations to a previously estimated additive tree.

In the example used throughout this paper, a cophenetic correlation of 0.99749 was obtained between the ultrametric matrix (Table 1b, upper

triangular) and the input dissimilarity matrix (Table 1a), indicating a good fit of the data by a dendrogram. A slightly larger correlation of 0.99897 was found between the input data and the path-length distances associated with the additive tree (Table 1b, lower triangular). As expected, the correlations for all reticulistic methods were even larger. For the pyramidal distances (Table 1c, upper triangular) and the distances associated with the weak hierarchy (Table 1c, lower triangular), the correlations were respectively 0.99769 and 0.99754. Similarly, correlations of 0.99914 and 0.99922 were obtained for the splitsgraph (Table 1d, upper triangular) and the additive tree with one extra edge (Table 1d, lower triangular).

Interestingly, the various methods produced somewhat different results; the biological meaning of these representations is of great importance. Whereas overlap is only allowed among contiguous clusters of species in pyramids, weak hierarchies can be used to represent reticulations between distant species or clusters (see Bandelt and Dress 1989). Similarly, the extra edges fitted on a tree when using the Makarenkov and Legendre algorithm tend to join distant species, as shown by the various examples presented by these authors (Makarenkov and Legendre 2000, and submitted). In such cases, reticulations may simply represent incompatibilities in the data resulting from convergent evolution. Another option, allowing the detection of a larger number of incompatibilities, is the splitsgraph. However, since they create a series of multiple parallel edges, splitsgraphs may quickly be saturated with extra vertices and edges, making it difficult to display them as planar graphs (Dress, Huson, and Moulton 1996).

7. Conclusion

This paper presented four different but somewhat related approaches to account for reticulation events in phylogenetic analysis. This list is not exhaustive; other techniques are currently available and being developed to produce reticulograms from gene frequencies (Xu 2000), binary data (Smouse 1998), or multistate characters using median graphs (Bandelt, Forster, and Rohl 1999). There are also clustering methods that can produce overlapping clusters. It is worth mentioning that the split decomposition method (Bandelt and Dress 1992) can be applied in other contexts than with distance data. To produce a splitsgraph all one needs is a phylogenetic method (parsimony or maximum likelihood) to decide, for each quadruplet, which of the three topologies is the most inappropriate. Likewise, an evolutionary parsimony criterion could be used to modify the Makarenkov and Legendre approach. Instead of searching for a reticulogram minimizing

a least-squares criterion, extra edges could be added to parsimonious trees obtained from standard algorithms so as to minimize the number of character-state changes on those trees. Whichever approach is selected, one should be aware that in spite of interesting mathematical properties, the different reticulistic methods will not necessarily produce biologically meaningful results. Model-based techniques should be developed to serve that purpose. On the other hand, simulation studies are badly needed to evaluate the relative performances of the extant competing methods. In addition, more comparative studies are required to determine the success rate of the different algorithms to recover known phylogenies that include species of reticulate origins like hybrids or allopolyploids (e.g., McDade 1997).

References

- BANDELT, H. J., and DRESS, A. W. (1989), "Weak Hierarchies Associated With Similarity Measures: An Additive Clustering Technique," *Bulletin of Mathematical Biology*, 51, 133-166.
- BANDELT, H. J., and DRESS, A. W. (1992), "Split Decomposition: A New and Useful Approach to Phylogenetic Analysis of Distance Data," *Molecular Phylogenetics and Evolution*, 1, 242-252.
- BANDELT, H. J., FORSTER, P., and ROHL, A. (1999), "Median-Joining Networks for Inferring Intraspecific Phylogenies," *Molecular Biology and Evolution*, 16, 37-48.
- BUNEMAN, P. (1974), "A Note on the Metric Properties of Trees," *Journal of Combinatorial Theory (B)*, 17, 48-50.
- DIDAY, E., and BERTRAND, P. (1986), "An Extension of Hierarchical Clustering: The Pyramidal Representation," in *Pattern Recognition in Practice*, Eds., E. S. Gelsema and L. N. Kanal, Amsterdam: North-Holland, 411-424.
- DRESS, A., HUSON, D., and MOULTON, V. (1996), "Analyzing and Visualizing Sequence and Distance Data Using SPLITSTREE," *Discrete Applied Mathematics*, 71, 95-109.
- HARTIGAN, J. A. (1967), "Representation of Similarity Matrices by Trees," *Journal of the American Statistical Association*, 62, 1140-1158.
- LAPOINTE, F.-J., and LEGENDRE, P. (1991), "The Generation of Random Ultrametric Matrices Representing Dendrograms," *Journal of Classification*, 8, 177-200.
- MAKARENKOV, V., and LEGENDRE, P. (2000), "Improving the Additive Tree Representation of a Dissimilarity Matrix Using Reticulations," in *Data Analysis, Classification, and Related Methods*, Eds., H. A. L. Kiers, J.-P. Rasson, P. J. F. Groenen, and M. Schader, Berlin: Springer, 35-46.
- MAKARENKOV, V., and LEGENDRE, P. "General Network Representation of a Dissimilarity Matrix: Adding Reticulations to an Additive Tree," (submitted).
- MCDADDE, L. A. (1997), "Hybrids and Phylogenetic Systematics III. Comparison With Distance Methods," *Systematic Botany*, 22, 669-683.
- SMOUSE, P. E. (1998), "To Tree or Not To Tree," *Molecular Ecology*, 7, 399-412.
- XU, S. (2000), "Phylogenetic Analysis Under Reticulate Evolution," *Molecular Biology and Evolution*, 17, 897-907.

Special Section on Reticulate Evolution

Phylogenetic Models and Reticulations

F. James Rohlf

State University of New York, Stony Brook

Abstract: The relative importance of genetic exchanges other than simple lineal descent (e.g., hybridization and lateral gene exchanges) on methods to represent evolutionary relationships is discussed. The importance of additional theoretical work in this field is emphasized.

Keywords: Evolutionary relationship; Gene tree; Lateral gene transfer; Multidimensional scaling; Parsimony.

1. Introduction

Recent developments have made it clear that more attention needs to be paid to other ways of representing evolutionary relationships among organisms than to the traditional phylogenetic tree. Such branching diagrams show ancestor-descendant relationships as a minimally connected graph with directed edges and no loops or cycles. While each internal vertex (corresponding to a hypothetical ancestor) can have links to two or more descendant vertices, internal vertices can have only one edge connecting them to their ancestral vertex. It has long been recognized that hybrids between living species (corresponding to the terminal vertices of the tree)

Author's Address: Department of Ecology and Evolution, State University of New York, Stony Brook, NY 11794-5245, USA; e-mail: rohlf@life.bio.sunysb.edu

cannot be represented correctly on such diagrams. There have been suggestions for *ad hoc* adjustments, such as showing the parental species and the hybrid as a trifurcation (i.e., connected by edges to the same ancestral vertex, which assumes that the parent species have the same immediate common ancestor) or leaving the hybrid out of the analysis (and thus not placed directly on the tree) and then showing the hybrid's relationship to its two parental species by dotted lines.

Of course, hybridization has also happened in the past, and such hybrids may by now have produced large groups of species as their descendants. However, this observation does not seem likely to pose an important practical problem, at least in multicellular plants and animals. If a pair of species is close enough to hybridize, then determining which parent or which hybrid eventually gave rise to a major taxonomic group is likely to remain below the resolving power of phylogenetic studies.

Figure 3 of Doolittle's (1999) provocative paper shows an impressively complicated hypothesis for the evolutionary history of the kingdoms of living organisms. The evolutionary relationships are shown as a directed graph in which *many* of the internal vertices have more than one edge connecting them to an ancestral vertex. In this diagram, most of the kingdoms of living organisms are linked together in a complicated reticulation – especially near the root of the graph. This diagram is based on evidence of lateral gene transfers (LGT) between endosymbionts and their hosts for some of the earliest and most primitive single-cell organisms. While LGT is common in bacteria, it is rare in eukaryotes (see Sneath 2000, in this Special Section) and thus not expected to create very important complications within lineages of multicellular plants and animals. Of course, the artificial transfer of genes through genetic engineering could affect the future evolution even of eukaryotes.

Whether diagrams such as Figure 3 of Doolittle's (1999) provide a reasonable representation of the tree of life involves some philosophical issues. If a single gene is transferred by some mechanism from species *A* to a distantly related species *B*, is it reasonable to consider species *A* as one of species *B*'s ancestors? If a human gene is transferred by genetic engineering into *E. coli*, should man then be considered an ancestor of this species of bacteria? Of course, this process could take place repeatedly in nature, with multiple genes coming to species *B* from species *A* as well as from a variety of other sources. Standards (probably inherently arbitrary) would have to be developed to decide when species *A* has contributed enough genetic material for it to be considered as one of the ancestors of species *B*. To take these complications into account, fundamental changes are needed in the ways genetic relationships among organisms are usually represented.

2. Applications

A particular phylogenetic study may be based on the analysis of a single gene or a single molecule. In such cases, the possibility of a reticulate pattern must be taken into account. A number of approaches are possible that might be useful for the analysis of such data. De Soete and Carroll's (1996) review includes methods that could be used to represent relationships in data by graphs that could even include reticulations. The most obvious approach is to extend existing parsimony and maximum likelihood methods for estimating phylogenetic trees by allowing additional edges and possibly additional internal vertices in the graph. They would be added when their presence would significantly decrease the length of the graph or increase its likelihood under some statistical model for evolution.

An important property of the usual phylogenetic trees is the nesting of discrete character states for those characters that are compatible with the relationships indicated by a tree. The concept of compatibility needs to be generalized to allow for the possibility of reticulations in a phylogenetic tree. An alternate approach is to fit a graph (possibly including reticulations) to a matrix of distances among the terminal vertices. However, even when using the same criterion (e.g., minimal length), discrete-character-based and distance-based methods are not equivalent and can lead to different results, because fitting character states rather than distances to a tree constrains the possible solutions when a Manhattan (i.e., "city-block") distance is used as the metric. This distinction is important to many evolutionary biologists, who give much greater weight to character-based solutions. Distance-based solutions seem more appropriate for data sets made of continuous characters.

3. Prospects

The computational complexity of distance-based methods to estimate conventional phylogenetic trees is NP hard (Day 1987, 1996), even without the added complication of the possibility of reticulations. Fitting graphs with possible reticulations is likely to increase the amount of computational effort greatly for any method that attempts to find an optimal solution. There is clearly a need for the development of new heuristic approaches that will make it practical to apply these new methods. One such proposal is that of Makarenkov and Legendre (2000, and submitted).

There are many other areas of biology for which methods to estimate reticulated patterns of relationships would be useful. The problem of estimating genetic relationships among geographically separated popula-

tions is an obvious area of application. Because of the different underlying models, correspondingly different methods are appropriate here. If the gene flow among populations results from a continuous spatial diffusion process, then characters are not expected to exhibit the nesting property described above. Thus, distance-based methods are more likely to work well for such applications. In a two- or three-dimensional physical space with continuous environmental gradients, the expected graph connecting local populations could resemble a two- or three-dimensional lattice. In such cases, it is possible that multidimensional scaling methods that locate populations in a two- or three-dimensional space may provide the most useful summary of the relationships among the populations. On the other hand, gene flow may be spatially constrained. If, for example, the movement of individuals is constrained to follow along a river and its tributaries, then the genetic relationships among populations may resemble an evolutionary tree (with reticulations if long distance dispersal is also possible occasionally). Thus, biologists must be prepared to consider a very wide variety of models when working near or below the species level.

Of course, new species may often arise from populations that show such complicated patterns of genetic affinities, stemming from complicated geographical, temporal, and environmental constraints on gene flow among the populations of the parent species and its possible hybrids with other species. There are also constraints on the combinations of phenotypic traits that are possible. Perhaps only relatively few of the theoretically possible combinations of traits represent viable organisms that are able to survive. If so, then the trajectories of the evolving clades will be greatly constrained in the paths they can take through the multivariate character space. This observation implies that a simple branching diagram (even with reticulations) represents a very simplistic representation of evolution and of the information that is likely to be present in a comprehensive set of data. Clearly, there are many opportunities for mathematicians and statisticians to contribute.

References

- DAY, W. H. E. (1987), "Computational Complexity of Inferring Phylogenies from Dissimilarity Matrices," *Bulletin of Mathematical Biology*, 49, 461-467.
- DAY, W. H. E. (1996), "Complexity Theory: An Introduction for Practitioners of Classification" in *Clustering and Classification*, Eds., P. Arabie, L. J. Hubert and G. De Soete, River Edge, NJ: World Scientific, 199-233.
- DE SOETE, G., and CARROLL, J. D. (1996), "Tree and Other Network Models for Representing Data," in *Clustering and Classification*, Eds., P. Arabie, L. J. Hubert and G. De Soete, River Edge, NJ: World Scientific, 157-197.

- DOOLITTLE, W. F. (1999), "Phylogenetic Classification and the Universal Tree," *Science*, 284, 2124-2128.
- MAKARENKOV, V., and LEGENDRE, P. (2000), "Improving the Additive Tree Representation of a Dissimilarity Matrix Using Reticulations," in *Data Analysis, Classification, and Related Methods*, Eds., H. A. L. Kiers, J.-P. Rasson, P. J. F. Groenen, and M. Schader, Berlin: Springer, 35-46.
- MAKARENKOV, V., and LEGENDRE, P. "General Network Representation of a Dissimilarity Matrix: Adding Reticulations to an Additive Tree," (submitted).
- SNEATH, P. H. A. (2000), "Reticulate Evolution in Bacteria and Other Organisms: How Can We Study It?," *Journal of Classification*, 17, 159-163.

Special Section on Reticulate Evolution

Biological Applications of Reticulation Analysis

Pierre Legendre

Université de Montréal

Abstract: This paper describes phylogenetic and ecological problems where mechanisms other than lateral gene transfer create, in the data, complexities that can be resolved by reticulation analysis. One such mechanism is homoplasy, which is the portion of phylogenetic similarity resulting from convergence. In the study of host-parasite relationships, reticulations provide a way of representing the putative historical events that may have led to a lack of fit between host and parasite trees. In historical biogeography (vicariance or dispersal), different processes may act concomitantly, producing data that can partly be represented by a tree and partly by additional edges (reticulations).

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

1. Introduction

Reticulations can be contemplated from a strictly statistical perspective: they represent a way of modifying a tree model in such a way as to obtain a better-fitting approximation for a distance matrix that does not satisfy the four-point condition which defines an additive tree (Buneman 1974). This paper describes some phylogenetic and ecological problems where mechanisms other than lateral gene transfer create, in the data,

Author's Address: Département de sciences biologiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec H3C 3J7, Canada; e-mail: Pierre.Legendre@umontreal.ca

complexities that can be resolved by reticulation analysis, and where the explanation of the reticulations identified in the course of the analysis goes beyond a strictly statistical point of view.

2. Homoplasy

There are phylogenetic problems where reticulation analysis can prove useful, irrespective of the absence of lateral gene transfer between taxa. The objective in this case is not to model actual reticulation events in evolution, as in Sneath (2000) and Smouse (2000) in this Special Section, but to use a diagram containing reticulations to describe more accurately the patterns found in the data. This situation arises with homoplasy which is the portion of phylogenetic similarity resulting from evolutionary convergence, i.e., parallel evolution and reversals. Adding reticulations to a phylogenetic tree can represent homoplasy, which represents an "excess of similarity" in the data. The inclusion of reticulations into the model, according to some goodness-of-fit criterion, indicates that the distance between two taxa is smaller, in the observed data, than the "patristic distance" which is the distance along the path of edges of the estimated tree. In such cases, addition of reticulations to a tree produces a graph (reticulogram) that better fits the data. For reticulations involving distant taxa on the tree, where no lateral gene transfer can be invoked, the result represents the homoplasy, which could not be adequately represented by the tree alone. An example of reticulate analysis of the phylogeny of primates, based upon mitochondrial DNA, is presented by Makarenkov and Legendre (2000).

3. Host-Parasite Relationships

The study of host-parasite relationships is another field where reticulation analysis may point out interesting phenomena. Page and Charleston (1998) emphasize the similarities between the problems posed by genes tracking organisms (i.e., being associated with them through time), parasites tracking hosts, and organisms tracking geological and geographical changes (see next section). The co-evolution of hosts and parasites is one of the few situations in phylogenetic analysis where one can perform a strong test of hypothesis under the primary model that the evolution of parasites, if they are host-specific, may mimic the evolution of their hosts. It is useful to falsify this basic model, because any departure points to some interesting mechanism, such as multiple hosts (which may intervene in different parts of the life cycle), host switching, or migration. See Brooks, Thorson, and

Mayes (1981) and Page (1994) for reviews of the methods of reconstruction of the history of host-parasite relationships. Reticulation analysis provides for representing the putative historical events that may have led to the lack of fit between host and parasite trees: the evolutionary tree of the hosts is applied to the parasite resemblance matrix, then reticulations are added to the tree to increase the goodness-of-fit of the model to the parasite data.

4. Biogeography

Historical biogeography is interested in the simultaneous evolution of taxa, species assemblages (ecological communities), and geographic areas. The major division of the field is between *dispersal* and *vicariance* biogeography. The dispersal approach follows the movements of organisms through geographic areas, while the vicariance approach studies how the creation of barriers, or other geologic processes, may have split up ("vicariated") biota, leading to speciation. *Vicariance* is the property of closely related taxa that occupy similar biota located in distinct and often widely separated geographic areas, e.g., on opposite sides of a sea or mountain chain whose formation was the vicariating event. Dispersal and vicariance are constantly alternating in nature. It seems appropriate to call upon dispersal when analyzing intermediate time and space processes, like postglacial reinvasion of a territory by organisms, whereas the vicariance paradigm is pertinent when studying problems involving long temporal and broad spatial scales, like continental drift.

Important texts on vicariance biogeography are Nelson and Platnick (1981) and Humphries, Ladiges, Roos, and Zandee (1988). Different explanations for the occurrence of the same, closely related, or weakly related species of animals and plants in different geographic regions can be hypothesized. It may not be easy to decide among them; Sneath (1982) gives examples of the evidence needed for this type of analysis. Reticulation analysis may be helpful to vicariance biogeography in enabling the analysis to depart from a strict tree-like set of relationships among geographic areas and suggesting that the resulting faunas or floras are the result of events that may have followed several geographic paths.

Dispersal biogeography has been studied for centuries, leading to such important concepts as the theory of island biogeography (MacArthur and Wilson 1967). Some researchers conduct multivariate analyses of the changes that have taken place in communities (species composition data) across space to delineate patterns and seek explanatory hypotheses. Even when the generating process can be assumed to be mostly tree-like in its

geography, reticulograms may provide better explanations of the data than do classical additive trees. A few examples, where reticulograms were used to describe lateral (i.e., geographic) exchanges of information between populations or communities of biological organisms, are given by Makarenkov and Legendre (submitted), who used a spatially-constrained form of their reticulation analysis method to revisit two data sets that had previously been studied under an additive-tree paradigm. The first example concerned community composition after postglacial (Pleistocene) reinvasion of a territory by freshwater fish (Legendre and Legendre 1984); addition of reticulations to the tree-like structure suggested meaningful improvements to the hypothesis describing dispersal routes. The second data set described the morphological similarity among local populations of muskrats along a river network (Le Boulengé, Legendre, de le Court, Le Boulengé-Nguyen, and Languy 1996); again, the reticulations added to the tree-like structure fitted the morphological distance matrix better than did the tree alone and suggested meaningful geographic routes for gene flow. The goodness-of-fit criteria used for selecting the most appropriate number of reticulations to be added to the tree in these two examples incorporated the number of degrees of freedom, and thus took into account the number of parameters of the models. These two examples are analogous to the lateral gene transfer among taxa which is studied by phylogeneticists using reticulation analysis.

References

- BROOKS, D. R., THORSON, T. B., and MAYES, M. A. (1981), "Freshwater Stingrays (Potamotrygonidae) and their Helminth Parasites; Testing Hypotheses of Evolution and Coevolution," in *Advances in Cladistics*, Eds., V. A. Funk and D. R. Brooks, New York: Columbia University Press, 147-175.
- BUNEMAN, P. (1974), "A Note on the Metric Properties of Trees," *Journal of Combinatorial Theory (B)*, 17, 48-50.
- HUMPHRIES, C. J., LADIGES, P. Y., ROOS, M., and ZANDEE, M. (1988), "Cladistic Biogeography," in *Analytical Biogeography – An Integrated Approach to the Study of Animal and Plant Distributions*, Eds., A. A. Myers and P. S. Giller, London: Chapman and Hall, 371-404.
- LE BOULENGÉ, É., LEGENDRE, P., DE LE COURT, C., LE BOULENGÉ-NGUYEN, P., and LANGUY, M. (1996), "Microgeographic Morphological Differentiation in Muskrats," *Journal of Mammalogy*, 77, 684-701.
- LEGENDRE, P., and LEGENDRE, V. (1984), "Postglacial Dispersal of Freshwater Fishes in the Québec Peninsula," *Canadian Journal of Fisheries and Aquatic Sciences*, 41, 1781-1802.
- MacARTHUR, R. H., and WILSON, E. O. (1967), *The Theory of Island Biogeography*, Princeton, NJ: Princeton University Press.

- MAKARENKOV, V., and LEGENDRE, P. (2000), "Improving the Additive Tree Representation of a Dissimilarity Matrix Using Reticulations," in *Data Analysis, Classification, and Related Methods*, Eds., H. A. L. Kiers, J.-P. Rasson, P. J. F. Groenen, and M. Schader, Berlin: Springer, 35-46.
- MAKARENKOV, V., and LEGENDRE, P. "General Network Representation of a Dissimilarity Matrix: Adding Reticulations to an Additive Tree," (submitted).
- NELSON, G., and PLATNICK, N. I. (1981), *Systematics and Biogeography: Cladistics and Vicariance*. New York: Columbia University Press.
- PAGE, R. D. M. (1994), "Parallel Phylogenies: Reconstructing the History of Host-Parasite Assemblages," *Cladistics*, 10, 155-173.
- PAGE, R. D. M., and CHARLESTON, M. A. (1998), "Trees Within Trees: Phylogeny and Historical Associations," *Trends in Ecology and Evolution*, 13, 356-359.
- SMOUSE, P. E. (2000), "Reticulation Inside the Species Boundary," *Journal of Classification*, 17, 165-173.
- SNEATH, P. H. A. (1982), "Book Review: *Systematics and Biogeography: Cladistics and Vicariance*, by Gareth Nelson and Normal Platnick," *Systematic Zoology*, 31, 208-217.
- SNEATH, P. H. A. (2000), "Reticulate Evolution in Bacteria and Other Organisms: How Can We Study It?," *Journal of Classification*, 17, 159-163.