

RECONSTRUCTING PATTERNS OF RETICULATE EVOLUTION IN PLANTS¹

C. RANDAL LINDER^{2,4} AND LOREN H. RIESEBERG³

²Section of Integrative Biology and the Center for Computational Biology and Bioinformatics, University of Texas–Austin,
1 University Station—A6700, Austin, Texas 78712 USA; and ³Department of Biology, Jordan Hall, Indiana University,
Bloomington, Indiana 47405 USA

Until recently, rigorously reconstructing the many hybrid speciation events in plants has not been practical because of the limited number of molecular markers available for plant phylogenetic reconstruction and the lack of good, biologically based methods for inferring reticulation (network) events. This situation should change rapidly with the development of multiple nuclear markers for phylogenetic reconstruction and new methods for reconstructing reticulate evolution. These developments will necessitate a much greater incorporation of population genetics into phylogenetic reconstruction than has been common. Population genetic events such as gene duplication coupled with lineage sorting and meiotic and sexual recombination have always had the potential to affect phylogenetic inference. For tree reconstruction, these problems are usually minimized by using uniparental markers and nuclear markers that undergo rapid concerted evolution. Because reconstruction of reticulate speciation events will require nuclear markers that lack these characteristics, effects of population genetics on phylogenetic inference will need to be addressed directly. Current models and methods that allow hybrid speciation to be detected and reconstructed are discussed, with a focus on how lineage sorting and meiotic and sexual recombination affect network reconstruction. Approaches that would allow inference of phylogenetic networks in their presence are suggested.

Key words: gene tree/species tree; hybrid speciation; phylogenetics; polyploidy; population genetics; recombination.

Phylogenetic trees are the main tool for representing evolutionary relationships among biological entities at the level of species and above. Biologists, mathematicians, statisticians, and computer scientists have developed a variety of methods for reconstructing these events, with the usual model being a phylogenetic tree. Over the last 30 years, biologists have come to embrace reconstruction of phylogenetic trees as a major research goal (Hillis, 1997; Huelsenbeck et al., 1997; Felsenstein, 2001) with the ultimate aim of inferring the evolutionary relationships of all of the extant and, whenever possible, fossil species on the earth (Soltis and Soltis, 2001; Bininda-Emonds et al., 2002; Watanabe, 2002).

Phylogenetics, because it reflects the history of transmission of life's genetic information, has unique power to organize our knowledge of diverse organisms, genomes, and molecules beyond merely providing the order and timing of speciation events. A reconstructed phylogeny helps guide interpretation of the evolution of organismal characteristics, providing hypotheses about the lineages in which traits arose and under what circumstances, thus playing a vital role in studies of adaptation and evolutionary constraints (e.g., Felsenstein, 1985; Maddison, 1990; Martins, 1995; Liberles et al., 2001; Merritt and Quattro, 2001). Phylogenetic trees also help elucidate patterns and dynamics of speciation and, to some extent, extinction when fossil data are available (Futuyma, 1998; Carroll et al., 2001).

In the second half of the twentieth century, trees were inferred primarily from morphological characters, but in the last decade or so, DNA sequences have become the primary data for phylogenetic inference. DNA sequences have a number of advantages in phylogenetic reconstruction, but they are not

without their problems. Points of strength include presence in nearly all organisms, a near perfect guarantee that sequence information is heritable, an abundant set of characters for reconstruction, sequences that evolve at different rates, and good models of sequence evolution for use in reconstruction. On the negative side are potential problems with paralogous sequences, aligning sequences so that positional homology of individual nucleotides is maintained, and the limited number of character states for nucleotides (Hillis et al., 1996; Moritz and Hillis, 1996). Usually these problems can be dealt with, mostly by careful selection of molecules that evolve at appropriate rates and that are either uniparentally inherited or that are known or assumed to undergo rapid concerted evolution. Nonetheless, the green-plant clade of the tree of life has some special characteristics relative to most of the animal and fungal clades that bring some of these problems to the fore and that demand our attention if we are to correctly infer relationships among plants. In particular, the evolutionary history of plants is not really a tree at all for some taxa. Rather it is a network, in which there have been a large number of reticulate evolutionary events, especially hybrid speciation, both polyploid and diploid (Stebbins, 1950; Grant, 1981; Arnold, 1997; Otto and Whitton, 2000). As Ford Doolittle (1999, p. 2124) wrote, "Molecular phylogeneticists will [fail] to find the 'true tree', 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."

Routine reconstruction of hybrid speciation in the manner of phylogenetic trees—for example, (1) searches of alternative reconstructions using optimality criteria and algorithms or heuristics with explicit evolutionary models, (2) extensive testing of methods on large sets of simulated phylogenies, and (3) parametric and nonparametric methods for assessing support for particular solutions—requires special methods that are, as yet, largely unavailable. Moreover, unlike tree reconstruction, numerous independently inherited sequences are required for

¹ Manuscript received 26 December 2003; revision accepted 22 June 2004.

The authors thank Lucinda McDade, Jeff Palmer, and one anonymous reviewer for their constructive comments on this paper and the NSF (CRL, LHR) and NIH (LHR) for funding to study hybrid speciation.

⁴ E-mail: rlinder@mail.utexas.edu.

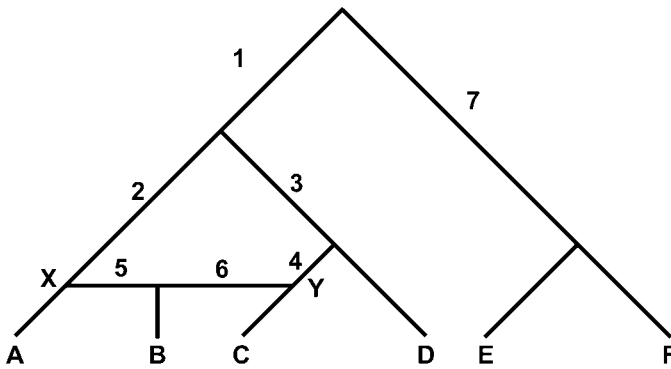


Fig. 1. Example of a phylogenetic network with a single hybrid species (B). Internal branches are numbered to allow the tripartition metric to be illustrated. See Table 1 for the tripartitions induced by each internal branch.

confident reconstruction of networks, and these kinds of data sets are currently rare. Finally, although phylogenetic reconstruction using methods that only recover trees requires some accounting for a number of population genetic processes—especially when using biparentally inherited markers—reconstruction of a network of relationships requires explicit incorporation of the effects of population genetic processes because they can mimic network patterns and, therefore, interfere with obtaining an accurate estimate of the network. In this article, we will (1) discuss some of the special needs for network detection and reconstruction, including methods developed to date, (2) explain how population genetic processes can affect our ability to accurately infer phylogenetic relationships in trees and networks, and (3) suggest some research directions for addressing these issues so the network of plant life can be accurately inferred. Our focus will be on network reconstruction using DNA sequence data.

The nature of hybrid speciation—In hybrid speciation, two otherwise independent lineages recombine sexually to create a new species (Fig. 1, species X, Y, and B). Hybrid speciation occurs in at least two ways: allopolyploid speciation and diploid (homoploid) hybrid speciation. Allopolyploidy is hybrid speciation between two species resulting in a new species that has the complete diploid chromosome complement of both its parents. The parents need not have the same base chromosome number. Allopolyploidy generally results in instantaneous speciation because any backcrossing to the diploid parents produces a high proportion of unviable or sterile triploid offspring. Diploid hybrid speciation results from a normal sexual event in which each gamete has a haploid complement of the nuclear chromosomes from its parent, but gametes that form the zygote come from different species. Because hybrids must have partial fertility or viability for hybrid speciation to be successful, backcrossing to the parents is often possible. Therefore, it is thought that speciation also requires hybrids to be isolated from parental species by selection for life in a novel environment, as seen in the few cases of demonstrated diploid hybrid speciation (Rieseberg and Carney, 1998). Not surprisingly, the number of identified diploid hybrid species is much lower than the number of allopolyploid species. Autopolyploidy occurs when the normal genome of a single species is duplicated in its entirety to produce a triploid or tetraploid offspring. It is sometimes treated as a form of hybrid speciation, but when autopolyploid lineages are postzygotically iso-

lated from their parent, they are more properly considered a specialized form of normal (bifurcating) speciation because only a single parental species is involved in their production.

Patterns indicative of hybrid speciation—To understand how hybrid speciation might be detected and reconstructed using DNA sequence data, consider how a single nucleotide site evolves down a simple network (Fig. 1). Assume the non-hybrid taxa are normal diploid organisms, in which each chromosome consists of a pair of homologs. In a diploid hybridization event, the hybrid (e.g., species B) inherits one of two homologs from each chromosome from each of its two parents (X and Y). Because homologs assort at random into gametes, each has an equal probability of ending up in the hybrid. In polyploid hybridization, both homologs from both parents are contributed to the hybrid. Prior to the hybridization event, each nucleotide site on each homolog has evolved in a treelike fashion at the species level, even though meiotic and sexual recombination will have caused strings of nucleotides on a homolog to have different histories from other strings. Because each nucleotide site on each homolog has come from one of the two parents, at the level of individual nucleotides on a homolog, each nucleotide has evolved on one tree contained inside the species-level network representing the hybridization event. For example, in Fig. 1, a nucleotide inherited from the X parent of hybrid B will be part of the subtree in which species A and B are sister taxa, and a nucleotide inherited from the Y parent of the same hybrid will be part of the subtree in which species B and C are sister.

The critical insight is that even when species relationships are properly represented as a *network*, each nucleotide site evolves down one of the *trees* contained inside the network. In other words, at the lowest possible level of evolutionary change, the correct representation is a tree. Because sets of tightly linked nucleotides that have not been recombined will share a common evolutionary history, each parent of the hybridization event can potentially be inferred.

Three lines of evidence might be employed to detect and reconstruct hybrid speciation. First, in the absence of other processes that might produce topologically incongruent trees, detection of hybrid speciation could be as simple as looking for sets of incongruent trees from separate data analyses on independent data sets, each representing a different parent of the hybridization (Maddison, 1997; Nakhleh et al., 2004). In theory, reconstruction of each hybrid speciation event could be accomplished accurately with just one marker or a small set of biparentally inherited markers that evolve at the appropriate rate. In reality, the number of biparentally inherited markers will have to be larger to distinguish incongruence due to hybrid speciation from incongruence due to population genetic and stochastic processes, which we discuss later in this article. The second way to detect hybridization would be to combine DNA sequences from multiple independent loci into a single analysis and look for phylogenetic signals that indicate a set of two or more histories, for example, by doing splits decomposition (Bandelt and Dress, 1992; Huson, 1998; Bryant and Moulton, 2002). As with the incongruence approach, this could work well in the absence of confounding processes. A third approach would involve searching for associations among genetically linked markers, i.e., linkage disequilibrium. The expectation is that tightly linked markers in a hybrid species are significantly more likely to come from the same parent and therefore to display linkage disequilibrium. Linkage dis-

equilibrium is often employed to detect contemporary hybridization events, but it also has provided perhaps the most convincing evidence for ancient hybridization events as well. For example, Doebley et al. (1984) found that an individual of *Zea diploperennis* had two allozymes that were common in maize. Because the two allozyme loci were tightly linked on chromosome six, their presence most likely was the result of introgression from maize rather than lineage sorting. Likewise, Rieseberg et al. (1996, 2003) showed that the genomes of hybrid sunflower species, which originated more than 63 000 years ago, contain blocks of linked markers (i.e., chromosomal segments) from both parental species. Hybrid speciation is the only plausible explanation for this pattern. Clearly, the linkage disequilibrium approach would be most powerful if employed in combination with phylogenetic incongruence. Under the assumption of hybrid (recombinational) speciation (Müntzing, 1930), separate phylogenetic reconstructions of individual DNA regions or loci that are part of a tightly linked set of loci should have the topology of only one side of the hybridization. These reconstructions would be topologically incongruent with reconstructions based upon clusters of regions or loci from the other parent of the hybridization.

Early phylogenetic studies of hybrid speciation—Although the problem of hybridization was mostly ignored in early phylogenetic studies, several approaches were suggested for the treatment of hybrids. Most frequently, it was proposed that hybrids be detected by other biosystematic tools and then excluded from phylogenetic study (e.g., Wagner, 1983). The other common suggestion was for inclusion of all taxa in initial phylogenetic analyses, followed by searches for phylogenetic signatures of hybridization such as character conflict and polytomies (e.g., Funk, 1985). Unfortunately, analyses of the placement of known hybrids in phylogenetic trees failed to reveal predictable hybrid phylogenetic patterns, at least for morphological features, leading McDade (1992) to predict that phylogenetic approaches were unlikely to be an effective tool for detecting hybrids.

On the other hand, early molecular phylogenetic studies were more successful at detecting the footprints of hybridization. The first studies comparing biparental nuclear and uniparental plastid phylogenies revealed discrepancies that were interpreted to result from hybridization (Palmer et al., 1983, 1985), and just a few years later, Rieseberg and Soltis (1991) were able to compile 36 such examples. Although these early studies were perhaps too quick to attribute patterns of phylogenetic incongruence to hybridization, it was clear that phylogenetic incongruence offered a powerful means for detecting past hybridization. More recent reviews have updated the list of known examples of phylogenetic incongruence (Rieseberg, 1996; Arnold, 1997) and hybrid speciation (Rieseberg, 1997). Others have discussed population genetic processes that could produce similar patterns (e.g., Wendel and Doyle, 1998) or offered simple computer programs for detecting hybrids in phylogenetic trees (Rieseberg and Morefield, 1995). Most of this work focused on detecting introgressive hybridization or diploid hybrid speciation because detecting hybrid speciation was considered trivial when ploidy changed (Rieseberg, 1997). However, because autopolyploids also undergo changes of ploidy the mere presence of polyploidy is insufficient for inferring hybrid speciation. In addition, if a clade includes multiple polyploid species with the same or similar numbers of chromosomes, looking for changes in ploidy cannot determine

whether there has been only one hybrid speciation event followed by bifurcating speciation of the initial polyploid or several independent polyploidization events.

Mathematical models of hybrid speciation—Mathematicians refer to the network depicted in Fig. 1 as a directed acyclic graph (DAG). It is directed because the tree is rooted, and so time (and information) flows through it in a directed way; it is acyclic because the flow of time and information never turns back on itself to trace through any node more than once. Hence, even though the graphical representation of the hybrid speciation event might appear to be a cycle, it technically is not. Strimmer et al. (2001) developed a model for applying maximum likelihood to directed splits graphs; however, splits graphs are representations of possible incompatibilities in sequence data sets and not phylogenetic networks. Hallett and Lagergren (2001) used a set of simplifying assumptions to create DAGs that were more biologically realistic than splits graphs and created a method for inferring lateral gene transfer events when one is attempting to reconcile gene trees and species trees. Linder et al. (2003) proposed a model of phylogenetic networks that is based on DAGs to describe the topology of phylogenetic networks, adding a set of (mostly simpler) conditions to ensure that resulting DAGs reflect the properties of biological reticulation.

For Linder et al. (2003), a phylogenetic network is a rooted DAG in which the internal nodes are partitioned into tree nodes and network nodes. A tree node has one ancestral branch and two or more descendant branches (allowing for polytomies). A network node has two ancestral branches and only one descendant branch. Similarly, branches are partitioned into tree branches and network branches. A tree branch has a tree node at its younger end, and a network branch has a network node at its younger end. Tree branches are directed from the root of the network towards the tips, and the network branches are directed from their tree-node endpoint towards their network-node endpoint. Visually, in Fig. 1, tree branches are angled or vertical, and network branches are horizontal. DNA sequences are assumed to evolve only on the tree branches, although a small amount of change could theoretically occur on the network branches (i.e., a mutation could occur during the evolutionarily instantaneous time it takes for an interspecific sexual event to occur). Because hybrid speciation requires a pair of species to sexually recombine, network branches must occur at the same instant in time and originate from concurrent tree branches.

As with phylogenetic tree inference, the design and analysis of methods for detection and reconstruction of phylogenetic networks have several components: (1) software for simulation studies that can generate model networks and evolve DNA sequences down the networks (so inferred networks using detection and reconstruction methods can be compared to model networks for accuracy), (2) algorithms and software for reconstructing phylogenetic networks, and (3) methods for assessing support for a particular reconstruction. Whereas the phylogenetics community has produced many tree simulation tools and reconstruction and support methods—many of which are good—much still needs to be done with respect to network evolution.

Software tools for generating random phylogenetic networks and simulating sequence evolution down phylogenetic networks have been developed for hybrid speciation (Nakhleh et al., 2003). These tools are adaptations and extensions of

TABLE 1. The set of tripartitions induced by each internal branch in the example network in Fig. 1.

Branch	Tripartition
1	(({A,B,C,D}), {}, {E,F})
2	(({A}), {B}, {C,D,E,F})
3	(({C,D}), {B}, {A,E,F})
4	(({C}), {B}, {A,D,E,F})
5	({ }, {B}, {A,C,D,E,F})
6	({ }, {B}, {A,C,D,E,F})
7	(({E,F}), { }, {A,B,C,D})

those used for the simulation of tree evolution (Rambaut and Grassly, 1997). When hybrid speciation events occur in the simulator, parents of the event are determined by the set of species that have the appropriate level(s) of ploidy and a probability function determined by the genetic distances among the possible parents available at the time of the hybrid speciation event. The choice of genetic distance as the determinant of the probability of hybrid speciation was chosen because it is generally true that more genetically distant species are less likely to successfully hybridize. However, not enough is currently known about the genetics of hybridization to include more detailed options for what determines the probability of successful hybridization.

Performance studies that assess network reconstruction methods need to be able to measure the error (distance) between the phylogeny of a group and the estimate of it. For such a measure to be a metric, it must be symmetric (count the same number of false positives—branches in the reconstruction that are not in the model—and false negatives—branches that are in the model but not the reconstruction) and be zero only when the phylogeny and its estimate are the same. Ideally, a network metric would reduce to an appropriate tree measure for cases in which there is no hybrid speciation, i.e., the metric should handle trees as a degenerate form of network, not a separate class of graphs that require independent measures. Error metrics are commonplace for trees, with the most common being the Robinson–Foulds (R-F) distance (Robinson and Foulds, 1981). The R-F measure tallies the number of bipartitions (the pair of sets of species produced by removing an internal branch on a tree) that appear in the true tree but not the reconstructed tree (false negatives) and the number of bipartitions that appear in the reconstructed tree but not the model tree (false positives). These numbers are then standardized according to the number of internal branches in the tree so that the metric varies between 0 and 1. The full set of bipartitions is produced by systematically removing each of the internal branches in turn and comparing the taxa that appear in each bipartition produced by branch removal. Identical model and reconstructed trees have an R-F measure of 0.

Linder et al. (2003) developed an extension of the Robinson–Foulds measure that meets the criteria of a metric and that reduces to the standard R-F distance when the reconstruction is a tree. Whereas the R-F metric breaks model and reconstructed trees into their full sets of bipartitions, the network metric is based on a tripartition (Fig. 1, Table 1). When an internal branch is removed from either the model or reconstructed network, the taxa are partitioned according to the following rules. Taxa below the removed branch, i.e., that are later in time than the younger node of the removed branch and that can only be reached by that branch, go in the first partition. Taxa below the removed branch that can be reached via

that branch but also by another branch that is not below the removed branch go in the second partition. Finally, any taxa that are not below the removed branch go in the third partition. For example, removal of branch 2 in Fig. 1 causes species A to go in the first partition because it can only be reached below branch 2. (It is important to remember that information only flows in one direction on the network, so it is not possible to reach A via branches 5 and 6.) Species B goes into the second partition because it is below branch 2 but is also reachable via branch 6, and the remaining species are not below branch 2. In general, taxa that can only be reached by a single path, no matter which branch is removed, evolved on a tree within the network and will only appear in the first and third partitions. They form the standard bipartition sets that would be formed under R-F. This characteristic is what causes the tripartition metric to be equivalent to R-F when the network is a tree and allows alternative tree and network reconstructions to be directly compared on the same scale. Any taxa that appear in the second partition are hybrids and will only appear when there are network events. Model and reconstructed networks that are identical will have measures of 0, just like R-F.

Methods for detecting and reconstructing phylogenetic networks—Of the three possibilities for detecting and reconstructing hybrid speciation, only the incongruence and the combined data approaches have been developed into formal methods, but both are at early stages of development. Because none of these methods has been well studied in simulation studies, we do not yet know how well they perform or the degree to which any of these approaches will generally infer networks. Our discussion of the current network reconstruction methods is, therefore, brief.

A small number of methods attempt to both detect and reconstruct hybrid speciation events using combined data (Satath and Tversky, 1977; Huson, 1998; Bandelt et al., 1999; Xu, 2000; Bryant and Moulton, 2002), i.e., data from multiple, independent genes or DNA regions, but none are entirely satisfactory, especially at reconstruction. In general, the methods produce an unacceptable number of false positives. The problems most likely arise because combined data are used and because the methods lack sufficient biological rationale.

Within combined data approaches, three general methods have been proposed. The first approach builds a tree and then adds network branches to turn it into a network, using a greedy approach to optimize some cost criteria (Clement et al., 2000; Makarenkov, 2001; Addario-Berry et al., 2003; Makarenkov and Legendre, 2004). The second approach builds many trees (sometimes using different subsets of the data) and attempts to reconcile them. If reconciliation fails, conflict might be explained by a reticulation event. This is the basic idea behind median networks (Bandelt et al., 1995, 1999, 2000), as well as the molecular-variance parsimony approach (Excoffier et al., 1992). Finally, incompatibilities in the data are characterized in advance of any reconstruction (for example, by looking for non-additivity in a distance matrix) to provide a collection of the possible resolutions through reticulation. The researcher is left to choose which resolution is preferable. This approach is used in the splits-based methods (Bandelt and Dress, 1992; Huson, 1998; Huber et al., 2001; Bryant and Moulton, 2002). Splits-based methods do not build or even propose a specific network, but present all consistent choices, a potential problem when the number of choices is large.

Reconstruction methods based on phylogenetic incongru-

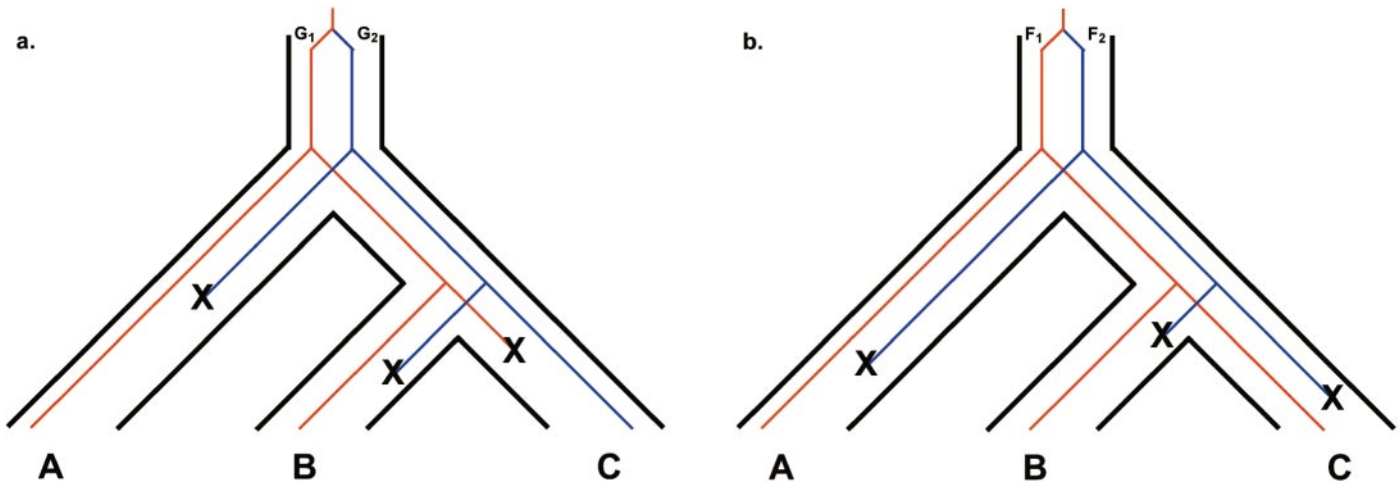


Fig. 2. An example of the gene tree/species tree problem. The species phylogeny is represented by black lines. The gene trees are represented by colored lines. (a) Prior to the root of the ABC clade, a gene (G_1 , in red) is either duplicated or mutates to produce a new allele (G_2 , in blue). Both versions of G are inherited at the two speciation events, but G_2 is lost in the lineages leading to species A and B, and G_1 is lost in the lineage leading to species C. The tree that would be reconstructed from the paralogous versions of G would incorrectly indicate that species A is the sister species of B. (b) A second gene (F) from the ABC clade where only F_2 is lost. Using the F_1 orthologs produces the correct set of species relationships. Note that the two genes produce incongruent trees, which indicates the possibility that B is a hybrid of A and C.

ence are only in the earliest stages of development, but they appear promising. Nakhleh et al. (2004) have developed an algorithm (SpNet) that is efficient at detecting and reconstructing hybrid speciation events under the special condition that the network is “galled,” that is, when each hybrid speciation event is evolutionarily independent from all the other hybrid speciation events in the network. In addition, simulation studies have shown that, in the presence of the sort of stochastic noise that is expected in DNA sequences, SpNet has a significantly lower false positive rate than NeighborNet (Bryant and Moulton, 2002), a combined data approach. It remains for incongruence approaches to be expanded to phylogenetic networks that include hybrids that are themselves parents in later hybrid speciation events.

Confounding population genetic processes—Were it not for population genetic events and systematic and stochastic variation in the evolutionary rates of DNA sequences, distinguishing between tree and network reconstructions would be computationally expensive, but nonetheless achievable. With long enough DNA sequences, reasonably short inferred branches, and sufficient computational power, networks would be detectable and in some cases readily reconstructable. Unfortunately, evolutionary histories are reticulate at levels below species and can give the appearance of being reticulate at the level of species even when they are not. Reticulation often occurs at the levels of chromosomes and genomes as well as species, which can mislead inference of hybrid speciation in both separate and combined data analyses. These other levels of reticulation can mimic patterns expected under hybrid speciation even when the underlying phylogeny is a tree. In addition, lineage sorting—the stochastic sorting of alleles following divergence from a polymorphic ancestor—as well as independent gene duplication and random loss in multiple genes can produce incongruent tree reconstructions that could be interpreted as hybrid speciation. (See Rokas et al., 2003 for a discussion of these issues.)

Multiple alleles and gene duplication—For recently diverged species, coalescence of alleles at a single locus may predate speciation. This is particularly common for nuclear genes, for which effective population sizes are double (for hermaphrodites) or quadruple (for species with separate sexes) that of organellar genes. As a consequence, relationships among allelic lineages in a set of species (i.e., the gene tree) may reflect stochastic sorting processes rather than species relationships. This produces the classic gene tree/species tree problem—whether the gene tree accurately reflects the species tree, which is the object of phylogenetic reconstruction. If alleles for different genes assort differently during speciation (which is likely), then incongruent trees will be reconstructed, which is exactly the same pattern used to identify hybrid speciation events.

The possibility for misinterpretation increases if the genes being analyzed are duplicated because researchers must distinguish between orthologous and paralogous sequences as well as lineage sorting among alleles for each gene. Orthologous sequences are those that have evolved from a single most recent common ancestor (MRCA) at the root of a clade, whereas paralogous sequences result from gene duplications that evolved prior to the MRCA of the clade (or any subclades within the clade that is to be reconstructed) (Fig. 2a). Because duplicated genes are subject to random loss in different species—via random production of pseudogenes—duplicated genes are subject to the gene tree/species tree problem in much the same manner as lineage sorting of alleles at a locus. Gene trees that are accurately reconstructed from the same alleles in a single ortholog will be identical to the species trees as long as coalescence times postdate speciation (Fig. 2b), but it is not always possible to be certain that all of the gene sequences used for phylogenetic reconstruction are orthologous. When paralogs are mistakenly used for reconstructing the gene tree, the “species tree” inferred will usually be incorrect. The one case in which paralogs will not affect species tree inference is when the duplication events are within the terminal branches.

When the origin of paralogous copies is within an internal branch, lineage sorting, inadequate sampling of the alleles of a gene, or confusing which gene duplicate is used for reconstruction can produce incorrect phylogenetic inferences. If all of the orthologs are present in the extant taxa from which the DNA sequences are taken, then the use of paralogs in tree reconstruction can be ameliorated by more extensive sampling of the species. However, a number of population genetic processes can cause orthologs to be randomly or systematically lost in some species: genetic drift and population bottlenecks (random) and natural selection (systematic). Thus, when a species lacks a particular ortholog, it is possible to use a paralog without being aware of it. Under these circumstances, an incorrect phylogenetic inference can be strongly supported by the data (high nonparametric bootstrap values under parsimony, distance, or ML methods or high posterior probabilities under Bayesian methods). Separate reconstructions that use two or more genes with different lineage sorting events can give the appearance of well-supported incongruent phylogenetic hypotheses and possibly lead to incorrect inference of reticulation events. Determining whether DNA sequences are orthologous in distantly related species is a current topic of research. Many papers discuss and provide algorithms for the gene tree/species tree problem, as well as some of its related problems, such as distinguishing orthologs from paralogs (see Maddison, 1997; Page and Charleston, 1997a, b; Eulenstein et al., 1998; Ma et al., 1998; Pamilo and Nei, 1988; Stege, 1999; Arvestad et al., 2003; Rokas et al., 2003).

Recombination—Meiotic recombination occurs in every sexual generation, causing individual nuclear chromosomes to contain two or more evolutionary histories. Over many generations, meiotic recombination may lead to chromatids that are complex mosaics of many evolutionary histories. Contiguous strings of nucleotides on a chromosome that share a single evolutionary history are referred to as haplotype blocks (Wang et al., 2002).

Sexual recombination commonly acts at the population level and recombines the evolutionary histories of genomes. Each parent contributes half of its original nuclear genome—one sister chromatid from each chromosome—and each of these chromosomes have themselves undergone meiotic recombination during the process of producing gametes. Because different parts of each parent's contribution to the genome of the next generation may have a different evolutionary history from that of the other parent's contribution, sexual recombination is a form of population-level reticulation. Organellar genomes (mitochondria and plastids) are haploid and usually inherited uniparentally, so they do not usually undergo sexual or meiotic recombination.

Sexual and meiotic recombination can cause at least two types of problems for detecting and reconstructing hybrid speciation. First, recombination coupled with drift and selection can cause different lineages to inherit different alleles at particular loci. The net effect of this is the same as lineage sorting, leading to incongruence among reconstructions of different loci. Second, errors in reconstruction could be generated by running analyses under the assumption that individual sequences represent a single evolutionary history when, in fact, they are (re)combinations of multiple histories.

Detecting recombination is a major topic of study in population genetics, with a commensurate number of publications. Studies of specific systems abound—any literature search us-

ing the keyword “recombination” will immediately bring up hundreds of references. Mostly, such recombinations are meiotic in nature. In phylogenetic work, detecting recombination (from a variety of sources) is at the heart of many approaches to the reconstruction of ancestral genomes or lines of descent (Hein, 1990, 1993; Griffiths and Marjoram, 1996; Smith and Smith, 1998; Holmes et al., 1999; McGuire et al., 2000; Strimmer et al., 2001; Wiuf et al., 2001; Worobey, 2001; McVean et al., 2002). Posada and Crandall (2001) have studied the accuracy of methods for detecting recombination from a collection of DNA sequences; their papers contain a wealth of references.

Detecting the presence of recombination is only the first step in assessing the evolutionary history of a DNA region. Characterizing recombinations that did take place is the goal. An intermediate goal along this path is to determine which recombination events might have taken place, as is done in many studies and implemented in several programs (Huson, 1998; Makarenkov, 2001; Bryant and Moulton, 2002; Zhang et al., 2002; Addario-Berry et al., 2003; Wall and Pritchard, 2003; Zhang and Jin, 2003). Some of these programs also attempt to determine the number of recombination events. Overall, the goal is to produce one or more recombination networks that optimize some criterion (perhaps a generalization of a criterion used in tree reconstruction, such as minimum evolution, parsimony, or maximum likelihood). None of the existing programs yet achieve this final goal, and none attempt to analyze meiotic recombination and hybrid speciation simultaneously.

Suggestions for future work—Phylogenetic network detection and reconstruction methods are at an early stage of development. Nonetheless, certain recommendations can be made for how to distinguish true hybrid speciation events from population genetic “noise.”

Distinguishing incongruent trees produced by population genetic processes from true hybrid speciation can be approached on the principle that all of the population genetic forces should usually produce random sets of incongruent trees, whereas hybrid speciation events should produce sets of incongruent trees that occur more often than would be expected by chance. At the species level, lineage sorting and recombination should both create gains and losses of gene lineages in extant taxa that have no particular relationship to the species network. The predictions become even more powerful if the linkage relationships among the sequenced genes are also considered (Huynen and Bork, 1998). With hybrid speciation, topological congruence should be greater among tightly linked than unlinked genes, but no association between linkage and topology is expected under divergent models of evolution. These approaches to network reconstruction will require both computational advances and practical advances in available markers.

Computationally, network generation tools will need to be extended to explicitly include lineage sorting and recombination, singly and in combination. This will allow researchers to simulate different levels (rates) of these population-level processes and then systematically assess their effects on the ability of current and future methods to correctly infer hybrid speciation events. It will be important to determine how these processes affect reconstruction when (1) the number of hybrid speciation events varies, (2) the number of taxa in the network varies, (3) the depth of the hybrid speciation events vary, (4) the complexity of the network varies, that is, when the types

of ploidy are more or less constrained and the hybrid speciation events are more or less independent from one another, (5) the number of independent DNA sequences used for reconstruction varies, and (6) the linkage relationships among markers varies. For example, it is to be expected that as the number of hybrid speciation events increases, a larger number of independent DNA regions will be needed to reliably detect and reconstruct hybrid speciation. However, at this point, nothing is known about the rate at which the number of regions needed will increase under different population genetic conditions.

Empirically, a significant effort is needed to develop a relatively large set of DNA regions that can be used for network reconstruction. Because mitochondria and plastids are primarily nonrecombining and uniparentally inherited, they cannot be used for multiple independent regions. Some labs have begun to use multiple single copy nuclear regions in phylogenetic reconstruction (Cronn et al., 2002; Mathews et al., 2002), but there has not been a concerted effort to develop “universal” or nearly universal single copy nuclear regions for green plants. A much larger number of nuclear regions needs to be developed. Ideal regions will be single copy (to increase the chance that orthology will be preserved) and will span a wide range of evolutionary rates so that different levels of the network can be reconstructed. There may, however, be a limit below which it will be virtually impossible to produce accurate network reconstruction because recombination will have so thoroughly mixed the evolutionary history of nuclear chromosomes that the size of haplotype blocks will be too short to provide enough informative variation. It may be possible to get around this problem to some extent by choosing regions that have low rates of recombination—centromeric and telomeric regions, for example—for deeper levels of reconstruction and reserve areas with higher rates of recombination for shallower levels. Studies need to be conducted to determine how many DNA regions are needed to make the distinctions at different levels of statistical confidence and at different levels in the network.

Developing a set of DNA regions for routine sequencing is a large undertaking, but one that is technically feasible by using at least two approaches. The first approach takes advantage of complete plant genome sequences to discover single copy regions, highly conserved regions, and linkage relationships. For example, one could compare the rice and *Arabidopsis* genomes to find regions that are single copy and tightly linked in both and that are sufficiently conserved to serve as PCR primers. Clearly, this approach is not without its problems. It is computationally difficult, and, biologically, there is no guarantee that what is single copy, well conserved, and linked between rice and *Arabidopsis* will be true throughout all plants (Lynch, 2002). However, as more plant genomes become available, it will be possible to more reliably assess whether a region or gene has desirable characteristics.

An alternative approach to whole genome comparisons would be to use data from the many expressed sequence tag (EST) projects for plants to find conserved regions and primers or compare whole genome sequences with EST libraries (Fulton et al., 2002). These approaches would provide a much larger set of species from which primer conservation could be ascertained, but they would not always readily lend themselves to determination of other important parameters: (1) whether conserved genes are broadly single copy, (2) the physical distance between primer pairs, (3) whether primers span an intronic region that would be useful for lower level reconstruc-

tion, and (4) linkage relationships among ESTs. Nonetheless, with sufficient effort, the best possible set of regions for network reconstruction will eventually emerge.

Summary—Because of their high level of hybrid speciation, plants present novel problems in phylogenetic reconstruction. Although biologically based and validated methods for network reconstruction are under development, only a limited set of reticulations can be correctly inferred at this time. In addition, the population genetic processes of meiotic and sexual recombination as well as lineage sorting can masquerade as hybrid speciation when only a small number of DNA regions are used to attempt reconstruction of hybrid speciation events. We have suggested that one of the most fruitful ways to reliably distinguish them is by using multiple independent DNA regions, particularly if linkage relationships are known. Parametric and nonparametric bootstrap methods need to be extended to network reconstruction to provide confidence assessments for different resolutions of data sets. Work also needs to be undertaken to provide a much larger set of DNA regions for network reconstruction. We conjecture that successful approaches in phylogenetic networks will combine population genetics and phylogenetics and will lead to interesting questions in many technical areas, including statistical inference, molecular phylogenetics, and computer science.

LITERATURE CITED

- ADDARIO-BERRY, L., M. T. HALLETT, AND J. LAGERGREN. 2003. Towards identifying lateral gene transfer events. *Proceedings of the Eighth Pacific Symposium on Biocomputing* (PSB03): 279–290.
- ARNOLD, M. L. 1997. Natural hybridization and evolution. Oxford University Press, New York, New York, USA.
- ARVESTAD, L., A. C. BERGLUND, J. LAGERGREN, AND B. SENNBLOD. 2003. Bayesian gene/species tree reconciliation and orthology analysis using MCMC. *Bioinformatics* 19: i7–i15.
- BANDEL, H. J., AND A. W. M. DRESS. 1992. Split decomposition: a new and useful approach to phylogenetic analysis of distance data. *Molecular Phylogenetics and Evolution* 1: 242–252.
- BANDEL, H. J., P. FORSTER, AND A. ROEHL. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16: 37–48.
- BANDEL, H. J., P. FORSTER, B. C. SYKES, AND M. B. RICHARDS. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* 141: 743–753.
- BANDEL, H. J., V. MACAULAY, AND M. RICHARDS. 2000. Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Molecular Phylogenetics and Evolution* 16: 8–28.
- BININDA-EMONDS, O. R. P., J. L. GITTLEMAN, AND M. A. STEEL. 2002. The (super) tree of life: procedures, problems, and prospects. *Annual Review of Ecology and Systematics* 33: 265–289.
- BRYANT, D., AND V. MOULTON. 2002. NeighborNet: an agglomerative method for the construction of planar phylogenetic networks. In R. Guigó and D. Gusfield [eds.], *Algorithms in bioinformatics, Second International Workshop, WABI, Rome, Italy, 2002. Lecture Notes in Computer Science* 2452: 375–391.
- CARROLL, S. B., J. K. GRENIER, AND S. D. WEATHERBEE. 2001. From DNA to diversity. Blackwell Science, Oxford, UK.
- CLEMENT, M., D. POSADA, AND K. CRANDALL. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657–1660.
- CRONN, R. C., R. L. SMALL, T. HASELKORN, AND J. F. WENDEL. 2002. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *American Journal of Botany* 89: 707–725.
- DOEBLEY, J. F., M. M. GOODMAN, AND C. W. STUBER. 1984. Isoenzymatic variation in *Zea* (Gramineae). *Systematic Botany* 9: 203–218.
- DOOLITTLE, W. F. 1999. Phylogenetic classification and the universal tree. *Science* 284: 2124–2129.

- EULENSTEIN, O., B. MIRKIN, AND M. VINGRON. 1998. Duplication-based measures of difference between gene and species trees. *Journal of Computational Biology* 5: 135–148.
- EXCOFFIER, L., P. E. SMOUSE, AND J. M. QUATTRO. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.
- FELSENSTEIN, J. 1985. Phylogenies and the comparative method. *American Naturalist* 125: 1–15.
- FELSENSTEIN, J. 2001. The troubled growth of statistical phylogenetics. *Systematic Biology* 50: 465–467.
- FULTON, T. M., R. VAN DER HOEVEN, N. T. EANNETTA, AND S. D. TANKSLEY. 2002. Identification, analysis, and utilization of conserved orthog set markers for comparative genomics in higher plants. *The Plant Cell* 14: 1457–1467.
- FUNK, V. A. 1985. Phylogenetic patterns and hybridization. *Annals of the Missouri Botanical Garden* 72: 681–715.
- FUTUYMA, D. J. 1998. Evolutionary biology. Sinauer Associates, Sunderland, Massachusetts, USA.
- GRANT, V. 1981. Plant speciation. Columbia University Press, New York, New York, USA.
- GRIFFITHS, R. C., AND P. MARJORAM. 1996. Ancestral inference from samples of DNA sequences with combination. *Journal of Computational Biology* 3: 479–502.
- HALLETT, M. T., AND J. LAGERGREN. 2001. Efficient algorithms for lateral gene transfer problems. In Proceedings of the Fifth Annual International Conference on Computational Biology (RECOMB01), Montreal, Quebec, Canada, 2001, 149–156.
- HEIN, J. 1990. Reconstructing evolution of sequences subject to recombination using parsimony. *Mathematical Biosciences* 98: 185–200.
- HEIN, J. 1993. A heuristic method to reconstruct the history of sequences subject to combination. *Journal of Molecular Evolution* 36: 396–405.
- HILLIS, D. M. 1997. Primer: phylogenetic analysis. *Current Biology* 7: R129–R131.
- HILLIS, D. M., B. K. MABLE, A. LARSON, S. K. DAVIS, AND E. A. ZIMMER. 1996. Nucleic acids IV: sequencing and cloning. In D. M. Hillis, C. Moritz, and B. K. Mable [eds.], *Molecular systematics*, 321–384. Sinauer Associates, Sunderland, Massachusetts, USA.
- HOLMES, E. C., M. WOROBAY, AND A. RAMBAUT. 1999. Phylogenetic evidence for recombination in dengue virus. *Molecular Biology and Evolution* 16: 405–409.
- HUBER, K. T., E. E. WATSON, AND M. D. HENDY. 2001. An algorithm for constructing local regions in a phylogenetic network. *Molecular Phylogenetics and Evolution* 19: 1–8.
- HUELSENBECK, J. P., B. RANNALA, AND Z. YANG. 1997. Statistical tests of host-parasite cospeciation. *Evolution* 51: 410–419.
- HUSON, D. H. 1998. SplitsTree: a program for analyzing and visualizing evolutionary data. *Bioinformatics* 14: 68–73.
- HUYNEN, M. A., AND P. BORK. 1998. Measuring genome evolution. *Proceedings of the National Academy of Science, USA* 95: 5849–5856.
- LIBERLES, D. A., D. R. SCHREIBER, S. GOVINDARAJAN, S. G. CHAMBERLIN, AND S. A. BENNER. 2001. The adaptive evolution database (TAED). *Genome Biology* 2: 1–6.
- LINDER, C. R., B. M. E. MORET, L. NAKHLEH, A. PADOLINA, J. SUN, A. THOLSE, TIMME, AND T. WARNOW. 2003. An error metric for phylogenetic networks. Technical Report TR–CS-2003–2026, University of New Mexico, Albuquerque, New Mexico, USA.
- LYNCH, M. 2002. Gene duplication and evolution. *Science* 297: 945–947.
- MA, B., M. LI, AND L. ZHANG. 1998. On reconstructing species trees from gene trees in terms of duplications and losses. In Proceedings of the second annual international conference on computational molecular biology (RECOMB98), New York, New York, USA, 1998, 182–191.
- MADDISON, W. 1990. A method for testing the correlated evolution of two binary characters: are gains or losses concentrated on certain branches of a phylogenetic tree? *Evolution* 44: 304–314.
- MADDISON, W. 1997. Gene trees in species trees. *Systematic Biology* 46: 523–536.
- MAKARENKOV, V. 2001. T-REX: reconstructing and visualizing phylogenetic trees and tication networks. *Bioinformatics* 17: 664–668.
- MAKARENKOV, V., AND P. LEGENDRE. 2004. From a phylogenetic tree to a reticulated network. *Journal of Computational Biology* 11: 195–212.
- MARTINS, E. P. 1995. Phylogenies and comparative data, a microevolutionary perspective. *Philosophical Transactions of the Royal Society of London, B* 349: 85–91.
- MATHEWS, S., R. E. SPANGLER, R. J. MASON-GAMER, AND E. A. KELLOGG. 2002. Phylogeny of Andropogoneae inferred from phytochrome B, GBSSI, and *ndhF*. *International Journal of Plant Sciences* 163: 441–450.
- MCDADE, L. A. 1992. Hybrids and phylogenetic systematics II: the impact of hybrids on cladistic analysis. *Evolution* 46: 1329–1346.
- MCGUIRE, G. F., F. WRIGHT, AND M. J. PRENTICE. 2000. A Bayesian model for detecting past recombination events and DNA multiple alignments. *Journal of Computational Biology* 7: 159–170.
- MCVEAN, G., P. AWADALLA, AND P. FEARNHEAD. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160: 1231–1241.
- MERRITT, T. J., AND J. M. QUATTRO. 2001. Evidence for a period of directional selection following gene duplication in a neurally expressed locus of triosephosphate isomerase. *Genetics* 159: 689–697.
- MORITZ, C., AND D. M. HILLIS. 1996. Molecular systematics: context and controversies. In D. M. Hillis, C. Moritz, and B. K. Mable [eds.], *Molecular systematics*, 1–16. Sinauer Associates, Sunderland, Massachusetts, USA.
- MÜNTZING, A. 1930. Outlines to a genetic monograph of the genus *Galeopsis*. *Hereditas* 13: 185–341.
- NAKHLEH, L., J. SUN, T. WARNOW, R. LINDER, B. M. E. MORET, AND A. THOLSE. 2003. Towards the development of computational tools for evaluating phylogenetic network reconstruction methods. In Proceedings of the Eighth Pacific Symposium on Biocomputing (PSB03): 315–326.
- NAKHLEH, L., T. WARNOW, AND C. R. LINDER. 2004. Reconstructing reticulate evolution in species: theory and practice. In Proceedings of the Eighth Annual International Conference on Research in Computational Molecular Biology, San Diego, California, USA, 2004, 337–346.
- OTTO, S. P., AND J. WHITTON. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* 24: 401–437.
- PAGE, R., AND M. A. CHARLESTON. 1997a. From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem. *Molecular Phylogenetics and Evolution* 7: 231–240.
- PAGE, R., AND M. A. CHARLESTON. 1997b. Reconciled trees and incongruent gene and species trees. In B. Mirkin, F. R. McMorris, F. S. Roberts, and A. Rzehtsky [eds.], *Mathematical hierarchies in biology*, 57–70. American Mathematical Society, Providence, Rhode Island, USA.
- PALMER, J. D., R. A. JORGENSEN, AND W. F. THOMPSON. 1985. Chloroplast DNA variation and evolution in *Pisum* patterns of change and phylogenetic analysis. *Genetics* 109: 195–214.
- PALMER, J. D., C. R. SHIELDS, D. B. COHEN, AND T. J. ORTON. 1983. Chloroplast DNA evolution and the origin of amphidiploid *Brassica* species. *Theoretical & Applied Genetics* 65: 181–189.
- PAMILO, P., AND M. NEI. 1988. Relationship between gene trees and species trees. *Molecular Biology and Evolution* 5: 568–583.
- POSADA, D., AND K. A. CRANDALL. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proceedings of the National Academy of Science, USA* 98: 13757–13762.
- RAMBAUT, A., AND N. C. GRASSLY. 1997. Seq-gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Computational and Applied Bioscience* 13: 235–238.
- RIESEBERG, L. H. 1996. Distribution of spontaneous plant hybrids. *Proceedings of the National Academy of Science, USA* 93: 5090–5093.
- RIESEBERG, L. H. 1997. Hybrid origins of plant species. *Annual Review in Ecology and Systematics* 28: 359–389.
- RIESEBERG, L. H., AND S. E. CARNEY. 1998. Plant hybridization. *New Phytologist* 140: 599–624.
- RIESEBERG, L. H., AND J. D. MOREFIELD. 1995. Character expression, phylogenetic reconstruction, and the detection of reticulate evolution. In P. C. Hoch and A. G. Stephenson [eds.], *Experimental and molecular approaches to plant biosystematics*, 333–353. Missouri Botanical Garden, St. Louis, Missouri, USA.
- RIESEBERG, L. H., O. RAYMOND, D. M. ROSENTHAL, Z. LAI, K. LIVINGSTONE, T. NAKAZATO, J. L. DURPHY, A. E. SCHWARZBACH, L. A. DONOVAN, AND C. LEXER. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301: 1211–1216.
- RIESEBERG, L. H., B. SINERVO, C. R. LINDER, M. C. UNGERER, AND D. M. ARIAS. 1996. Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. *Science* 272: 741–745.
- RIESEBERG, L. H., AND D. E. SOLTIS. 1991. Phylogenetic consequences of

- cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* 5: 65–83.
- ROBINSON, D. R., AND L. R. FOULDS. 1981. Comparison of phylogenetic trees. *Mathematical Biosciences* 53: 131–147.
- ROKAS, A., B. L. WILLIAMS, N. KING, AND S. B. CARROLL. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425: 798–804.
- SATTATH, S., AND A. TVERSKY. 1977. Additive similarity trees. *Psychometrika* 42: 319–345.
- SMITH, J. M., AND N. H. SMITH. 1998. Detecting recombination from gene trees. *Molecular Biology and Evolution* 15: 590–599.
- SOLTIS, P. S., AND D. E. SOLTIS. 2001. Molecular systematics: assembling and using the tree of life. *Taxon* 50: 663–677.
- STEBBINS, G. L. 1950. Variation and evolution in plants. Columbia University Press, New York, New York, USA.
- STEGE, U. 1999. Gene trees and species trees: the gene-duplication problem is fixed-parameter tractable. In Algorithms and data structures. Sixth International Workshop, WADS'99, Vancouver, Canada, *Lecture Notes in Computer Science* 1663: 288–293.
- STRIMMER, K., C. WIUF, AND V. MOULTON. 2001. Recombination analysis using directed graphical models. *Molecular Biology and Evolution* 18: 97–99.
- WAGNER, W. H., JR. 1983. Reticulistics: the recognition of hybrids and their role in cladistics and classification. In N. I. Platnick and V. Funk [eds.], *Advances in cladistics*, proceedings of the second meeting of the Willi Hennig Society. Columbia University Press, New York, USA.
- WALL, J. D., AND J. K. PRITCHARD. 2003. Assessing the performance of the haplotype block model of linkage disequilibrium. *American Journal of Human Genetics* 73: 502–515.
- WANG, N., J. M. AKEY, K. ZHANG, R. CHAKRABORTY, AND L. JIN. 2002. Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. *American Journal of Human Genetics* 71: 1227–1234.
- WATANABE, M. 2002. Describing the “Tree of Life”: attainable goal or stuff of dreams? *Bioscience* 52: 875–880.
- WENDEL, J. F., AND J. J. DOYLE. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. In D. E. Soltis, P. S. Soltis, and J. J. Doyle [eds.], *Molecular systematics of plants II: DNA sequencing*, 256–296. Kluwer Academic Publishers, Boston, Massachusetts, USA.
- WIUF, C., T. CHRISTENSEN, AND J. HEIN. 2001. A simulation study of the reliability of recombination detection methods. *Molecular Biology and Evolution* 18: 1929–1939.
- WOROBAY, M. 2001. A novel approach to detecting and measuring recombination: new sights into evolution in viruses, bacteria, and mitochondria. *Molecular Biology and Evolution* 18: 1425–1434.
- XU, S. Z. 2000. Phylogenetic analysis under reticulate evolution. *Molecular Biology and Evolution* 17: 897–907.
- ZHANG, K., AND L. JIN. 2003. HaploBlockFinder: haplotype block analyses. *Bioinformatics* 19: 1300–1301.
- ZHANG, J., W. L. ROWE, J. P. STRUEWING, AND K. H. BUETOW. 2002. HapScope: a software system for automated and visual analysis of functionally annotated haplotypes. *Nucleic Acids Research* 30: 5213–5221.