10

Phylogenetic Incongruence: Window into Genome History and Molecular Evolution

Jonathan F. Wendel and Jeff J. Doyle

The field of systematic biology has been revitalized and transformed during the last few decades by the confluence of phylogenetic thinking with ready access to the tools of molecular biology. Indeed, the title of this volume and the fact that it is already in its second edition offers ample testimony to the impact that molecular approaches have had on efforts to reconstruct the phylogenetic history of plants. Concomitant with the proliferation of molecular tools has been a growing awareness that reliance on a single data set may often result in insufficient phylogenetic resolution or misleading inferences. Accordingly, it is an increasingly widespread practice to apply multiple data sets to a common group of taxa. One of the consequences of analyzing multiple data sets is that the phylogenies inferred may differ from each other in one or more details. This phylogenetic incongruence is not rare; to the contrary, it is almost the rule rather than the exception, being evident to varying degrees.

Given the prevalence of phylogenetic incongruence, the question naturally arises as to whether two or more independent data sets should only be analyzed separately or whether they should be combined into a global analysis. This question stems in part from the recognition that different character sets may have different underlying evolutionary histories and are therefore expected, in many instances, to lead to different reconstructions of the sampled taxa. Despite considerable past and present discussion regarding optimal treatment of multiple data sets, no clear consensus has emerged as to the most appropriate course of action (Miyamoto, 1985; Hillis, 1987; Kluge, 1989; Barrett et al., 1991; Doyle, 1992; Bull et al., 1993; Eernisse and Kluge, 1993; de Queiroz et al., 1995; Miyamoto and Fitch, 1995). The reader is referred to these sources for explication of arguments on both sides of the issue.

Complicating this issue is the realization that not all incongruence is equal in magnitude and that topological differences between competing phylogenies may have a number of different causes, including some that are artifactual. Consequently, there have been efforts to measure incongruence and evaluate whether it is "real" and hence potentially biologically significant, or spurious, due to insufficient character evidence, excessive homoplasy, or some other

We thank our many graduate students and colleagues with whom we have shared innumerable illuminating conversations. We also thank R. Cronn, R. Small, and L. Clark for comments on the manuscript and T. Seelanan for assistance with the figures. Much of the authors' research has been sponsored by the National Science Foundation, whose support we gratefully acknowledge.

cause (Mickevich and Farris, 1981; Templeton, 1983; Faith, 1991; Swofford, 1991; Rodrigo et al., 1993; Lutzoni and Vilgalys, 1994; Farris et al., 1995; Mason-Gamer and Kellogg, 1996; Kellogg et al., 1996; Sites et al., 1996; Lyons-Weiler et al., 1996, 1997; Graham et al., 1997; Seelanan et al., 1997). The question of whether different data sets should be combined into a single global analysis thus becomes intertwined with the issue of assessing the reality of topological differences. Because this latter topic is discussed in detail in Chapter 11, it is not expanded upon here.

Surrounding much of the discussion of the treatment of multiple data sets has been an implicit or explicit assumption that phylogenetic incongruence is inherently undesirable. That is, incongruence of topologies derived from different data sets is often viewed as an unfortunate but unavoidable side effect of phylogenetic analysis, and as such it represents an impediment to achieving phylogenetic understanding. Against this backdrop, perhaps it is not surprising that a considerable portion of the literature on the treatment of multiple data sets has focused on *reconciliation* of alternative estimates of phylogeny or on assessing which of two competing resolutions is better supported by the available evidence.

In this chapter, we present a different perspective on phylogenetic incongruence. The unifying theme of the chapter is that rather than representing an undesirable outcome or a problem that requires a solution, phylogenetic incongruence is touted as an important observation that often reflects something interesting in the biology of the group under study, and accordingly, its appearance may alert us to one or more evolutionary processes that would not have been suspected in the absence of incongruence. To this extent, phylogenetic incongruence may be *desirable*, as it often illuminates previously poorly understood evolutionary phenomena. In this chapter we enumerate and discuss the various processes that underlie phylogenetic discord and attempt to assess their relative importance as causative agents. In expanding upon previous treatments of this subject (Sytsma, 1990; Doyle, 1992; Kadereit, 1994; de Queiroz et al., 1995; Brower et al., 1996), our intention is to provide an introduction to the relevant issues and to the causes of phylogenetic incongruence, as well as to offer a single-source entry point to the recent literature. We focus on plants, although literature from other groups is cited where appropriate or particularly illustrative.

CAUSES OF PHYLOGENETIC INCONGRUENCE

Before delving into the causes of phylogenetic incongruence, it is necessary to address the question of whether the evidence for it, in any particular case, is substantial enough to warrant a conclusion that it actually exists. Phrased another way, it is important to evaluate whether the conflict is "significant" and therefore possibly reflective of different evolutionary histories for two or more data sources, or "insignificant," meaning that it does not hold up to inspection and standard measures of statistical evaluation (Mickevich and Farris, 1981; Templeton, 1983; Faith, 1991; Swofford, 1991; Rodrigo et al., 1993; Farris et al., 1995; Mason-Gamer and Kellogg, 1996; Huelsenbeck et al., 1996; Kellogg et al., 1996; Sites et al., 1996; Seelanan et al., 1997; see Chapter 11). If the conflict is judged to be insignificant, using appropriate criteria, the possibility remains that the phylogenetic incongruence *may* reflect one or more underlying biological processes that are differentially affecting distinct data sets, but it is also possible that the incongruence has a more mundane genesis. Examples of the latter would include the many cases where particular clades are weakly supported and where alternative resolutions are only slightly less parsimonious. Indeed, this type of "soft incongruence" (Seelanan et al., 1997) is actually *expected* to occur as long as weakly supported nodes exist in trees arising from different data sources (e.g., Kim and Jansen, 1994; Olmstead and Sweere, 1994; Hoot et al., 1995; Kellogg et al., 1996; Seelanan et al., 1997). In *Krigia*, for example, the expected and cladistically supported (by morphology and cpDNA) sister-species relationship between *Krigia biflora* and *K. montana* was not obtained in a tree based on internal transcribed spacer (ITS) data; this incongruence, however, disappeared in trees one step longer than the shortest.

Similarly, in an analysis of the Solanaceae based on three different molecular data sets, Olmstead and Sweere (1994) only observed disagreement in the Solanoideae, where character support was minimal, suggesting that the lack of complete congruence among gene trees reflects an absence of sufficient signal rather than fundamentally different evolutionary histories.

The foregoing discussion and examples are intended to underscore an important notion, namely, that *not all incongruence is created equal*. Minor points of topological disparity, often reflecting no special biological process or differential history, are commonly observed, and in fact are expected in many situations. In the context of the present chapter, this type of conflict is not especially interesting in that it cannot be assumed to result from an evolutionary process that differentially affects data sets. More interesting are examples of "hard incongruence" (Seelanan et al., 1997), where alternative resolutions in two or more trees derived from distinct data sets are statistically supported as incongruent. In these instances one might justifiably suspect that the discord was generated by an evolutionary process that differentially affected the various sources of data, so it becomes important to ask what this process or processes are.

This form of phylogenetic incongruence provides the basis for the remainder of this chapter. As shown in Table 10.1, the evolutionary processes potentially responsible are many and varied. For purposes of presentation, these have been divided into three broad categories, reflecting technical causes, processes that operate at the whole organism level, and gene or genome-level processes. To a certain extent, this classification is subjective, in that several of the listed phenomena might reasonably be placed elsewhere in the table. Phylogenetic sorting of ancestral polymorphisms (lineage sorting), for example, might be conceived of as either an organism-level or genome-level process, depending on context and one's perspective. Notwithstanding this arbitrary aspect of the listing, each category of phenomena and its subdivisions are discussed in turn.

A final comment concerns the "phenotype" in question, namely, differing phylogenetic resolutions of two or more taxa based on two or more data sets. In some if not most cases, more than one of the myriad phenomena listed in Table 10.1 might conceivably have generated the discord of interest. This fact alone serves to underscore an important cautionary note about phylogenetic incongruence, specifically, that even when the discord is striking and strongly supported, its genesis will probably not be evident from phylogenetic analysis alone, and may not be evident even when all available evidence is considered. Thus, inferring the cause is often problematic.

Table 10.1. Phenomena that may lead to conflicting phylogenetic hypotheses.

Technical causes	Insufficient data
	Gene choice
	Sequencing error
	Taxon sampling
Organism-level processes	Convergent or rapid morphological evolution
	Rapid diversification
	Hybridization/Introgression
	Lineage sorting
	Horizontal transfer
Gene and genome-level processes	Intragenic recombination
	Orthology/paralogy conflation
	Interlocus interactions and concerted evolution
	Rate heterogeneity among taxa
	Rate heterogeneity among sites
	Base compositional bias
	RNA editing
	Nonindependence of sites

Technical Causes

Phylogenetic resolutions based on two or more data sets may differ due to experimental or technical reasons rather than differing underlying histories. The examples mentioned above, where weakly supported clades differ from alternative resolutions by only one to several steps, are a case in point, with the technical cause being "insufficient data." Other experimental variables that may play a role in generating discordant trees include appropriateness of the molecular tool, quality of the data, and density of taxon sampling.

Gene Choice One of the fundamental principles of molecular systematics is that the rate of molecular evolution for a particular sequence

should be optimized, to the extent possible, to the scale of divergence in the taxa whose phylogenetic relationships are being explored. If the gene evolves too slowly, signal will be insufficient and hence little phylogenetic resolution may be expected. Alternatively, if the rate of evolution is too high relative to the scale of taxon divergence, phylogenetic signal may be obscured by homoplasy. Less than optimal choices at the outset of a study may lead to incongruence among the resulting gene trees, either from lack of sufficient signal in one or more of the data sets (hence, a spurious resolution) or from excessive sequence evolution, evidenced as long branches and the associated problem of homoplasy masquerading as synapomorphy ("long-branch attraction"; Felsenstein, 1978). These issues of gene choice and its relationship to taxonomic rank (or more precisely, scale of divergence, as not all families and genera, for example, are equivalently aged or evolve at the same rate) are discussed in greater depth in Chapter 1. The essential point here is that topological incongruence among gene trees may have an experimental basis.

A related technical or experimental issue concerns the length of DNA sequences employed for the construction of gene trees. Just as substitution rates need to be sufficient to provide phylogenetic signal, enough nucleotides need to be sampled to overcome the sampling error inherent in data representing single stretches of DNA (Cummings et al., 1995; de Quieroz et al., 1995; Hillis, 1996). Thus, sequence lengths need to be considered in addition to evolutionary rates in the design of phylogenetic projects. The total length of sequence needed is difficult to predict a priori; it depends on many factors, including taxon sampling density, the amount of and level of heterogeneity in molecular evolutionary rates, the degree of divergence among taxa within the ingroup, and the amount and distribution of homoplasy. The simulation by Hillis (1996) is instructive in this regard. Using a model phylogeny of 228 angiosperms based on 18S ribosomal DNA sequences (generated by D. Soltis, P. Soltis, D. Nickrent, and colleagues; see Soltis, Soltis, Nickrent et al., 1997), Hillis explored the relationship between accuracy of phylogenetic inference and number of nucleotides in the data set. He found that over 90% of the model tree was accurately reconstructed using sequences approximately the length of *rbcL*, with an asymptotic approach to over 99% correct using 5,000 base pairs.

Sequencing Error In addition to gene choice and sequence length, the accuracy of the sequence data collected may influence the topologies obtained (Clark and Whittam, 1992), and therefore possibly lead to discordant gene trees. Sequencing error is probably of most concern in studies where sequence divergence, and hence the number of potentially informative sites, is very small, although it is conceivable that it also influences resolution among closely related terminals in studies involving more divergent taxa. Although most investigators are alert to this potential problem, it is often impractical to verify the accuracy of each and every apomorphy scored during an initial round of sequencing. Moreover, in studies of recently diverged taxa, where sequencing errors are likely to have their biggest impact, a common solution to the problem of insufficient sequence evolution is to employ nongenic sequences such as introns or spacers. These regions, while often providing the desired higher rates of sequence evolution, may also lead to higher error rates in scoring, because of the absence of the continual alignment checks present in genes (e.g., codons), and because of their inherently higher indel frequencies (Golenberg et al., 1993; Morton and Clegg, 1993; Gielly and Taberlet, 1994; Johnson and Soltis, 1994, 1995; van Ham et al., 1994; Baldwin et al., 1995; Kelchner and Wendel, 1996). Indels may complicate sequence alignments and hence homology assessments (see Chapter 4), potentially leading to alignment errors. Sequencing error may also result from sampling nuclear loci that are heterozygous, particularly when sequences are generated from PCR pools rather than cloned products. Finally, sequencing error may arise from amplification of contaminating DNAs, as a number of recent studies have documented (e.g., Olmstead and Palmer, 1994; Liston et al., 1996; Zhang et al., 1997).

Taxon Sampling A final and related technical issue concerns taxon sampling, which, if in-

sufficiently dense, has long been recognized as a cause of long-branch attraction (Felsenstein, 1978) and hence, potentially, phylogenetic incongruence. This problem may or may not be avoidable, depending on the study group in question and its historical pattern of speciation and extinction. In some cases increasing taxon sampling density may lead to improved phylogenetic accuracy (Wheeler, 1992), although this is not assured (Kim, 1996).

In addition to the issue of taxon sampling *density*, taxon *identification* is also relevant, in that misidentified samples in one study may lead to an erroneous phylogenetic inference and hence incongruence with a second study or data set. In many molecular phylogenetic studies, one or more taxa may be obtained as DNA extracts, or samples for DNA extraction may be secured as fresh or dried leaves from collectors, colleagues, or botanical gardens. The potential for mislabeling or misidentification underscores the necessity of using appropriately vouchered materials in molecular phylogenetic work.

Organism-level Processes

Perhaps more interesting than cases of incongruence resulting from technical causes are those where the discord reflects different underlying evolutionary histories for the data sets in question. As listed in Table 10.1 and discussed subsequently, a variety of processes may be involved. Most of these entail some variant on a common theme, whereby one or more genes has experienced a perturbation in its history relative to other genes. Because the genes will have experienced different histories, it is expected that this will be reflected in reconstructions of these histories.

A simple example illustrates an important point. Imagine a single allele that somehow is transferred across a species barrier, perhaps through hybridization and subsequent introgression, from a donor taxon to a recipient species, where it becomes fixed. If this is the sole introgressed allele, a gene tree based on this particular gene will likely be incongruent with the "species tree" and may also conflict with trees based on other, nonintrogressant genes. Notice that in the latter case, that involving a comparison of gene trees, *neither* gene tree is more correct than the other, given the important caveats that sufficient data have been accurately gathered and that there have been no other perturbing forces. Phrased alternatively, phylogenetic hypotheses based on sequences from one DNA segment may be in conflict with hypotheses generated from a different DNA segment *even when both gene trees are correct in all details.* This, in fact, is the expectation in many cases, particularly at the species level (Davis and Nixon, 1992; Doyle, 1992, 1995; Baum and Shaw, 1995; Maddison, 1995, 1996). In the example given, each gene tree faithfully reproduces the history of diversification of the alleles sampled in the various taxa studied. The fact that one allele was interspecifically introgressed across a species barrier does not invalidate its gene tree. A pitfall, however, arises from a tempting interpretive extrapolation, namely, that a gene tree faithfully reproduces the history of organismal divergence. This may or may not be true, as this example underscores.

So gene trees may be incongruent with other gene trees and/or with species trees. The challenge then, is not only to infer the phylogeny of the organisms under study, but also to interpret the evolutionary history of the data sources used. These twin objectives often involve reciprocal illumination, whereby phylogenetic analysis leads to the recognition of incongruence, which informs subsequent experiments or facilitates revised interpretations. The latter, in turn, provide a more complete portrait of organismal history.

Convergent or Rapid Morphological Evolution The relative merits of morphological and molecular data in phylogeny reconstruction have been discussed at length (e.g., Hillis, 1987; Donoghue and Sanderson, 1992; Patterson et al., 1993; Kadereit, 1994). A point on which there is near universal agreement is that morphological evolution may differ fundamentally from molecular evolution in that single genetic changes often underlie dramatic morphological transformations (Gottlieb, 1984; Doebley, 1993). Sytsma (1990) and Kadereit (1994) highlight these transformations as a root cause of phylogenetic incongruence, suggesting that they are

typically unaccompanied by similar levels of molecular divergence. Essentially this is a form of incongruence arising from rate differences, where the rates in question are morphological on the one hand and molecular on the other. To the extent that this phenomenon causes incongruence, it may serve to identify cases where genes with major influences on morphogenesis have had profound evolutionary effects.

In addition to evolutionary rate differences and timing of morphological transformations, evolutionary convergence in morphology may cause incongruence between trees inferred from molecular and morphological data, as noted by Sytsma (1990). In some cases this convergence may be striking, as with the repeated evolution of carnivory in plants (Albert et al., 1992).

Issues concerning convergent morphological evolution and the evolutionarily sporadic nature of morphological change are already widely appreciated and need not be elaborated here. Perhaps it is worthwhile, though, to draw a parallel with molecular evolution, which is often viewed as being more stochastically regular and free from the type of convergence envisioned for morphological traits. Thus, there is no evidence that major convergence events have taken place in *rbcL* evolution, for example, at least to the extent that sequences from disparate lineages would have become phylogenetically linked via this process. It seems likely that this freedom from wholesale convergence will generally be true for molecular data, or at least for DNA sequence data, although there may be exceptions.

On the other hand, it is equally clear that molecular tools are not completely free from convergence. There are now many examples, at a variety of taxonomic levels, of convergent molecular changes in characters that at one time might have been presupposed to be unique. These include the repeated losses of cpDNA genes and introns (Downie et al., 1991, 1994; Doyle et al., 1995), convergent inversions of minute as well as large pieces of cpDNA (Downie and Palmer, 1994; Hoot and Palmer, 1994; Kelchner and Wendel, 1996), and the remarkable discovery of "intron homing" in rDNA of fungi (Hibbett, 1996), where a particular group I intron is reported to have been precisely inserted at an identical nucleotide position in disparate groups of homobasidiomycetes. These and other examples bear witness to the potential for molecular convergence, and to the extent that these labile characters influence gene tree topologies, they may also be a source of phylogenetic incongruence. Given this potential lability, individual structural mutations may prove most informative when evaluated in the context of phylogenies inferred using other data sources.

Rapid Diversification If organismal divergence events are temporally compressed relative to the scale of molecular evolution, phylogenetically inferred internodes on gene trees may be short and difficult to resolve with confidence. In these cases, the relevant clades tend to be weakly supported, often decaying in trees only one to several steps longer than the most parsimonious trees. These same clades are often not recovered in subsequent analyses using additional molecular markers. This short internode or short interior branch phenomenon may be a common cause of misleading phylogenetic inference as well as phylogenetic incongruence.

The short internode problem is a relative concept, meaning that it is dependent on the scale of divergence and the data type being used. It is, however, applicable to all levels of organismal divergence, and examples of phylogenetic incongruence attributable to this cause abound at a variety of taxonomic levels in both plants and animals (e.g., Olmstead and Sweere, 1994; Fehrer, 1996; Lara et al., 1996; Baldwin, 1997). At the family level, for example, several short branches are apparent near the base of the Poaceae in trees based on cpDNA characters, regardless of the data set, and alternative molecular data sets often yield conflicting resolutions of these major clades. Clark et al. (1995) and Mathews and Sharrock (1996) found a BOP clade (Bambusoid, Oryzoid, Pooid) using *ndhF* and phytochrome data, respectively, whereas alternative resolutions were obtained using *rbcL* (Barker et al., 1995; Duvall and Morton, 1996) and cpDNA restriction site (Davis and Soreng, 1993) data. Similarly, some cladistic relationships in the Asteraceae (Kim et al., 1992; Kim and Jansen, 1995), Lardizabalaceae (Hoot et al., 1995), and Saxifragaceae (Morgan and Soltis,

1993; Soltis et al., 1993, 1996; Johnson and Soltis, 1994, 1995; Soltis et al., 1996) are unstable vis-à-vis source of molecular data. At the tribal and generic levels in the Gossypieae (Malvaceae) and in *Gossypium* itself, Seelanan et al. (1997) attributed varying resolutions, as inferred from ITS, *ndhF*, and cpDNA restriction site data, to the short interior branch phenomenon.

In these and many other examples, the absolute meaning of "short interior branch" is dependent on the scale of divergence. Thus, a short interior branch in, for example, a phylogenetic study of families or orders using a slowly evolving molecule such as *rbcL* will have a rather different temporal interpretation than a similar branch in a phylogeny of closely related species using a fast-evolving sequence such as a nuclear intron. In both cases, short refers to the relative paucity, and hence presumable instability, of support, but in the former case the actual temporal spacing of divergence events may be an order of magnitude or more greater than in the latter case. Both types of phenomena are commonly referred to in the literature as "rapid radiations," so the expression is a relative one. The pattern underlying the interpretation, though, namely a star phylogeny, is the same regardless of scale.

Regardless of the issue of absolute temporal scale, the many examples of unstable short interior branches highlight an evolutionary phenomenon that may be fairly common. In many cases, it may be that this form of phylogenetic incongruence is revealing with respect to the history of the group under study. Thus, rather than focusing exclusively on resolution or reconciliation of the conflict between data sets, it might be worthwhile to acknowledge the potential significance of the incongruence and seek its genesis. Star phylogenies might, for example, alert us to a history of fragmentation of a formerly widespread population system into a series of geographically isolated populations, or to a relatively rapid diversification following a novel ecological or morphological adaptation (see fig. 16, p. 176, Soltis and Soltis, 1995), such as radiation of allopolyploids following their formation (e.g., Wendel, 1989). In practice, alternative resolutions of short interior nodes represent a form of soft incongruence, as defined in the introduction to this chapter, and in these cases the conflict might disappear with additional data. Nevertheless, the perspective offered here is that the incongruence is important in its own right, in that it focuses our attention on the possibility of a history of rapid radiation, whatever that means ecologically or temporally for the group in question.

Hybridization and Introgression One of the most significant and conspicuous insights to emerge from molecular systematic investigations in the last decade is that hybridization and introgression are even more widespread in plant populations than earlier suspected on morphological grounds (Anderson, 1949; Heiser, 1973; Rieseberg and Wendel, 1993; Rieseberg, 1995). In nearly all cases, the initial evidence for the phenomenon has consisted of an unexpected phylogenetic result, usually between cpDNA restriction sites and morphology but also among various other combinations of organellar and nuclear markers. Many of these cases apparently represent ancient introgression episodes, events that were wholly unsuspected prior to their serendipitous discovery using molecular markers. A testimonial to the prevalence of cryptic hybridization and introgression is that examples of the process have accumulated quickly in the literature (e.g., Palmer et al., 1983, 1985; Doebley, 1989; Smith and Sytsma, 1990; Rieseberg et al., 1990; Rieseberg, 1991; Wendel et al., 1991). In their 1991 review, Rieseberg and Soltis listed 37 examples of cpDNA introgression, and within the span of just 5 years this number had grown to over 100 (Rieseberg et al., 1996). Mechanisms that underlie interspecific introgression of organellar genomes have been reviewed (Rieseberg et al., 1996), as has the potential evolutionary significance of the process (Rieseberg and Soltis, 1991; Arnold, 1992; Rieseberg and Wendel, 1993; Rieseberg, 1995; Rieseberg et al., 1996). Irrespective of these important questions of mechanism and significance, cpDNA introgression evidently is a common process, and hence it is necessary to consider its phylogenetic implications.

cpDNA introgression is diagrammatically illustrated in Fig. 10.1. If plastome transfer is unaccompanied by nuclear introgression, an inferred cpDNA tree will likely differ from gene

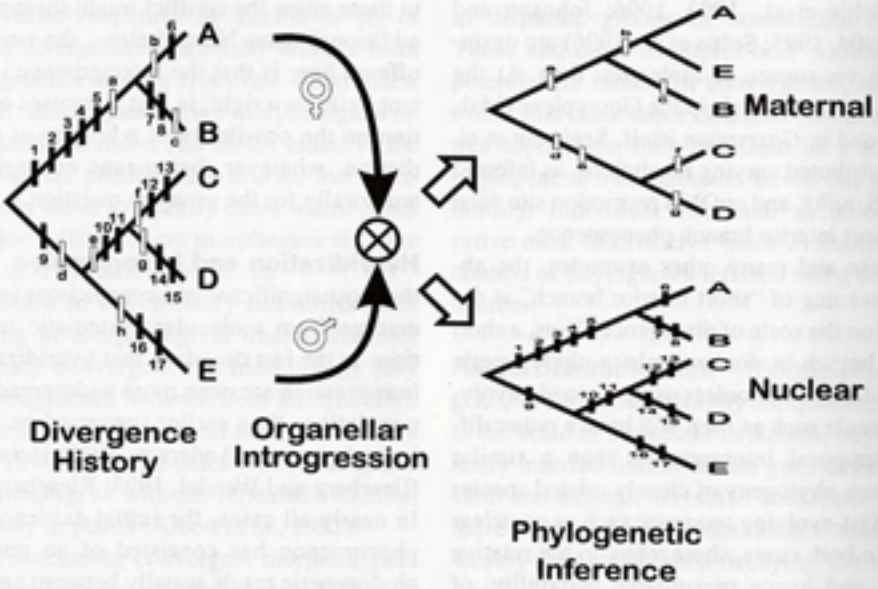

Figure 10.1. Introgression as a cause of phylogenetic incongruence. A divergence history for five taxa (A through E) is illustrated on the left. During divergence, mutations occur in both nuclear (solid bars) and organellar (open bars) genes. Hybridization between Taxon A, as female, and Taxon E, as the pollen parent, followed by organellar introgression into the paternal lineage, leads to cytoplasmic replacement in Taxon E, which acquires an A organellar complement. Phylogenies based on maternal (top right) and nuclear (bottom right) genes will therefore be incongruent. For simplicity, there is no homoplasy in either data set and no molecular evolution subsequent to hybridization.

trees based on nuclear sequences and from phylogenies inferred from morphological data. Hence, phylogenetic incongruence is the expected result. Given the apparently high frequency of cytoplasmic introgression, the phenomenon is likely to be among the most common causes of phylogenetic discord. An important point about cpDNA introgression is that because the plastome is nonrecombinant, it is inherited as a unit, and so all cpDNA sequences and/or restriction sites are expected to be transmitted together as a single haplotype, regardless of how many nucleotides or restriction sites are scored (Doyle, 1992). Consequently, phylogenetic incongruence is not expected among gene trees based on two or more different cpDNA data sets.

Among the several generalizations to emerge with respect to introgression is that cytoplasmic gene flow occurs at an apparently higher frequency than nuclear introgression. Moreover, patterns of organellar and nuclear introgression are typically asymmetrical. That is, cytoplasmic gene flow is frequently observed without evidence of nuclear introgression, whereas nuclear introgression without concomitant cytoplasmic introgression has rarely been demonstrated. Rieseberg et al. (1996) present a useful analysis of why patterns of nuclear and cytoplasmic gene flow may differ in plant populations and include a discussion of some of the potentially important ecological and genetic factors. Especially uncommon have been examples of introgression of nuclear genes between plant populations for which there is no ongoing evidence of hybridization (Talbert et al., 1990; Wendel et al., 1995b). Presumably, this scarcity of evidence does not reflect the actual absence of ancient nuclear introgression, *nor necessarily* a quantitative difference in its frequency relative to cytoplasmic introgression. Instead, it probably reflects both methodological and conceptual factors, including the difficulty of detecting introgressant loci in large, complex plant genomes, and problems in demonstrating that candidate markers are truly introgressant, as op-

posed to phenotypically convergent, symplesiomorphic, or a consequence of sorting of ancestral polymorphisms (see Lineage Sorting, below). As nuclear molecular tools become more widely applied, examples of nuclear introgression will continue to accumulate (Rieseberg et al., 1991, 1996; Brubaker et al., 1993; Wendel et al., 1995b).

From a phylogenetic reconstruction standpoint, nuclear introgression may differ from cytoplasmic introgression in significant ways, although both may lead to phylogenetic incongruence. For a single, nonrecombining introgressant nuclear gene, the effects will be similar to those of plastome introgression, in that gene trees based on this sequence will reflect the evolutionary history of that particular gene rather than the taxa from which they arose. If, however, a reconstruction is based on a number of independent nuclear markers, such as restriction fragment length polymorphism (RFLP) or allozyme loci for which individual gene trees are not first reconstructed, the resolution observed should depend on many factors, including the proportion of loci that are introgressed, their linkage relationships in the genome, and the antiquity of the introgression event. First generation hybrids are expected to resolve in cladistically basal positions (McDade, 1990, 1992), and it is likely that introgressants will be found to resolve at every level in a hierarchy, depending on the factors listed previously as well as the breeding history of the introgressed population.

Given these considerations, it should be apparent why interpretation of introgression may not be straightforward, except in those cases where evidence for cytoplasmic introgression is convincing. That is, the phenotype of incongruence may be self-evident, but its genesis may be elusive. Further complicating the analysis is the issue of antiquity of the putative reticulation event, both in an absolute sense and relative to the scale of cladogenesis and extinction. The more ancient the reticulation, the more difficult will be its detection, particularly for nuclear markers, which tend to be more polymorphic than cytoplasmic makers and which recombine. On the other hand, there are continuing reports of reasonably ancient reticulations first evidenced by "aberrant" plastomes (Doebley, 1989; Wendel et al., 1991; Soltis and Kuzoff, 1995), demonstrating that the effects of reticulation may remain phylogenetically detectable for perhaps millions of years.

One important implication of these examples, and others, is that hybridization and introgression may influence interpretations of relationships not only at the very tips of phylogenetic trees, which is where interspecific progeny are most likely possible, but also at higher levels of divergence among species that no longer are able to form fertile hybrids. In this respect it is noteworthy that many of the reports of cytoplasmic introgression involve ancient hybrids among lineages that presently cannot be crossed, but apparently once did (Palmer et al., 1983; Doebley, 1989; Wendel et al., 1991; Soltis and Kuzoff, 1995). A particularly striking example of this phenomenon is offered by *Gossypium gossypioides,* a species from Oaxaca, Mexico (Wendel et al., 1995b). Evidence indicates that the nuclear genome of this species is heavily introgressed with sequences from taxa that presently exist in Africa and Asia. This is an example of introgression between taxa whose modern derivatives are not only incompatible, but have geographic ranges in different hemispheres, and hence, have no opportunity for sexual contact.

Perhaps it is in this arena of hybridization and introgression that phylogenetic incongruence has had its greatest impact in systematic biology. Despite the many uncertainties of interpretation given any single set of conflicting cladograms, the incongruence itself has inspired a search for explanations, which has led to a renewed interest in the possibility that hybridization is a potent force in evolution (Arnold, 1992; Rieseberg, 1995).

Lineage Sorting Genetic polymorphism is a conspicuous characteristic of plant populations. This polymorphism has been extensively documented by allozyme analysis (Hamrick and Godt, 1989; Soltis and Soltis, 1989), and more recently has been confirmed and quantified by DNA sequencing (e.g., Gaut and Clegg, 1993b; Hanson et al., 1996; Innan et al., 1996). Thus, it comes as no surprise that individuals may be heterozygous at a particular nuclear gene used

for phylogeny reconstruction, or that variation exists for a locus within a population, or that multiple alleles exist within a particular species. These multiple alleles are related to each other by a complex history involving myriad factors, such as mutational processes, interaction with other alleles via recombination or gene conversion, and a whole series of population level features that may have influenced genetic transmission. Genes within species therefore have a history, which may be reconstructed by phylogenetic analysis to yield an intraspecific gene tree (Doyle, 1992, 1995, 1996; Maddison, 1995, 1996).

It is of interest to consider the nature of gene trees in the context of naturally occurring polymorphism levels and to extend the gene tree concept to interspecific comparisons. Consider, for example, the case where ancestral alleles are polymorphic and a particular polymorphism is maintained through one or more speciation events. The daughter lineages arising from this speciation event each have two alleles, and these alleles are shared across the species boundary. Accordingly, each of the two alleles may be referred to as being older than the species to which they belong. Extending this process of organismal divergence through time, and introducing allele extinction through normal stochastic or selective means and allele genesis through mutation it is possible to envision how gene trees might fail to reflect the phylogeny of the species that house them.

Lineage sorting, or phylogenetic sorting, refers to a special subset of gene tree—species tree relationships (Neigel and Avise, 1986; Pamilo and Nei, 1988). Specifically, if a polymorphism transcends one or more organismal divergence events, and if the polymorphism fails to survive later speciation events, the polymorphism is "sorted" into its component alleles. A simple example of this process is illustrated in Fig. 10.2. Notice that the sine qua non of lineage sorting is the existence of an ancestral polymorphism, which may exist for any nuclear or organellar sequence, or, for that matter, for a morphological trait (Roth, 1991). As long as there is polymorphism, the possibility exists that it may survive speciation and thereby contribute to a future lineage sorting event.

With respect to the subject of the present chapter, the significance of lineage sorting is

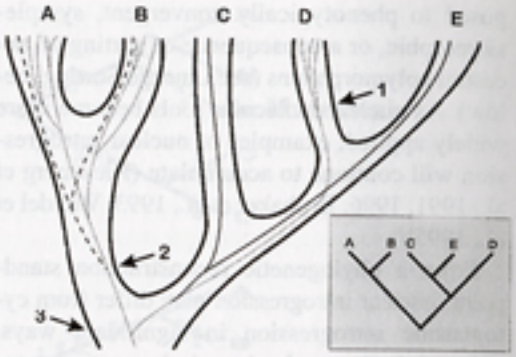

Figure 10.2. Lineage sorting, coalescence, and phylogenetic incongruence. Illustrated are the historical relationships among alleles at a single locus in five hypothetical species, A through E. Alleles are continually created through mutational processes and lost via selection and drift, but for simplicity, only alleles that have survived to the present are depicted. Alleles within a species may be monophyletic, as exemplified in this example by the three alleles in species D; these alleles share a most recent common ancestor, or *coalesce*, at the point marked by arrow number 1. In other cases, alleles within a species are not monophyletic, as shown by species A and B, each of which possess both dotted and dashed alleles; coalescence of these alleles occurs at the point marked by arrow number 2, antedating the divergence of species A and B. Coalescence of all alleles in the figure (including solid alleles) occurs at arrow number 3. Differential survival of alleles in alternative lineages will lead to instances where historical relationships among alleles (= gene tree) are not the same as those of the species that house them (= species tree). In the example shown, dotted alleles survive to the present in species D but solid alleles do not, whereas the opposite allelic survivorship pattern occurs in species C and E. Hence, sampling of a single allele from each taxon will lead to recovery of the gene tree shown in the inset, which is incongruent with the history of organismal divergence. The process by which the allelic polymorphism present in the common ancestor of species C, D, and E was lost is termed *lineage sorting*.

that it may be a cause of phylogenetic incongruence. This is explained most easily by referring to Figs. 10.2 and 10.3, which show how lineage sorting may lead to discordant gene trees and species trees, and by extension, between two or more gene trees. One important manifestation of lineage sorting is that alleles *from different* species may be more closely related to each other than are alleles *within the same* species (Figs. 10.2, 10.3). Consequently, and because the process by definition involves alleles that are older than their species, alleles at any given locus within a particular species are not necessar-

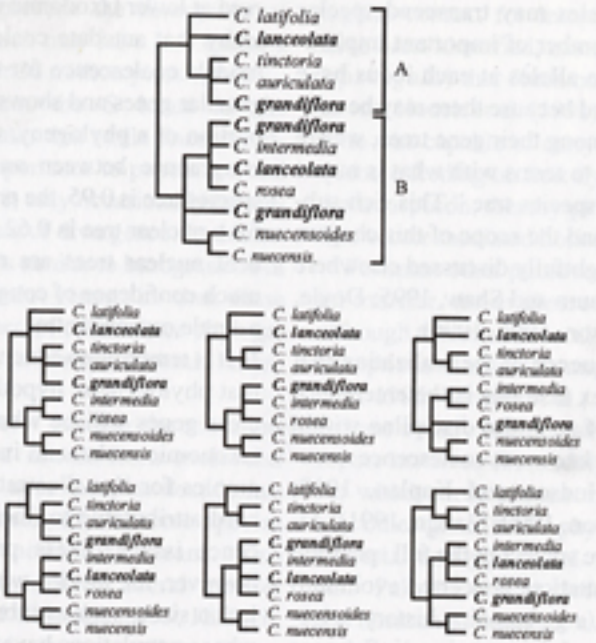

Figure 10.3. The effect of sampling or lineage sorting on a real data set. *Coreopsis lanceolata* and *C. grandiflora* are each polymorphic for cpDNA haplotypes, whose relationships to one another and to cpDNA haplotypes from other *Coreopsis* species are shown in the top cladogram (after Mason-Gamer et al., 1995). Haplotypes fall into two primary clades, designated A and B. If only one individual was sampled from each of the two species *C. lanceolata* and *C. grandiflora*, or if there had been additional haplotype lineage extinction, any one of the six topologically different hypotheses of species relationships (below) might have been obtained.

ily expected to be monophyletic. To this extent, phylogenetic incongruence among gene trees is an *expected* outcome of lineage sorting at lower taxonomic ranks, the divergence levels for which lineage sorting processes are usually thought to operate.

Lineage sorting may affect both organellar and nuclear genes, but because of the generally slow rate of cpDNA evolution and population genetic considerations, there are relatively few examples of cpDNA polymorphisms that transcend species boundaries. One example is offered by *Coreopsis*, in which two plastome classes (A and B) found in a wide survey of *Coreopsis grandiflora* may have been phylogenetically sorted into various populations of this species and other *Coreopsis* taxa (Mason-Gamer et al., 1995). Additional examples include wild potatoes (Castillo and Spooner, 1997) and populations of two species of *Phacelia* (Levy et al., 1996).

More common are empirical examples of nuclear genes where alleles transcend species boundaries, and this number is expected to grow rapidly in the coming years as additional surveys are undertaken of multiple alleles within species (see, for example, Cronn et al., 1996 for 5S rDNA polymorphisms shared among closely related *Gossypium* species). At present, some of the best examples in plants come from maize and its relatives, where alleles that are more ancient than their species have been described for *Adh* (Gaut and Clegg, 1993a; Goloubinoff et al., 1993), *c1* (Hanson et al., 1996), and the internal transcribed spacer region of the major rDNA repeat (Buckler and Holtsford, 1996a). In each of these examples, gene tree topologies are consistent with the operation of lineage sorting, although the authors correctly take pains to discuss the alternative possibility of interspecific gene flow. This uncertainty as to cause deserves emphasis here: The incongruence footprint left by lineage sorting is often identical to that expected from other processes, such as introgression (see Assessing the Cause).

The fact that alleles may transcend species boundaries has a number of important implications. First, because alleles at each locus have their own history, and because there may be considerable discord among their gene trees, we are challenged to come to terms with what is meant by the expression "species tree." This rich subject is perhaps beyond the scope of this chapter, but it has been thoughtfully discussed elsewhere (see in particular Baum and Shaw, 1995; Doyle, 1995, 1996; Maddison, 1995, 1996).

A second consequence of the realization that alleles have complex histories is the emergence and development of an entire discipline within population genetics known as coalescence or coalescent theory (Hudson and Kaplan, 1988; Ewens, 1990; Hudson, 1990; Slatkin, 1991). Alleles at any locus are subject to the full spectrum of internal (e.g., mutation, concerted evolution) and external forces (e.g., breeding history, selection, drift). Effects of these various influences are expected to differ for alleles within a species as well as for alleles from different species. At any locus, a given pair of alleles may share a relatively recent common ancestor, in which case coalescence is said to be relatively fast, or alternatively, the most recent common ancestor may antedate several speciation events, in which case coalescence is relatively ancient.

One result from coalescent theory that is particularly relevant to phylogenetic issues is that effective population size has a large effect on coalescence times: Alleles in small populations reach fixation or are lost more rapidly than alleles from large populations, and hence the time to coalescence is less in small than in large populations. Accordingly, coalescence of organellar genes may be faster than nuclear genes (Moore, 1995), all else being equal (see, however Hoelzer, 1997), because organellar genes often have lower effective population sizes than nuclear genes (Birky et al., 1983; see, however, Chesser and Baker, 1996). This population genetical process, combined with the generality that infraspecific polymorphism levels are typically low for plastome genes, accounts for the relative rarity of lineage sorting for cpDNA-based gene trees (see, however, Mason-Gamer et al., 1995). It also accounts for the expectation that in general, lineage sorting is only expected to be a potential source of discord at lower taxonomic ranks, levels in the hierarchy that antedate coalescence. Moore (1995) models coalescence for neutral nuclear and organellar genes and shows that for a three-species portion of a phylogeny, when the probability of congruence between an organellar tree and a species tree is 0.95, the probability of congruence for a nuclear tree is 0.62. Moreover, 16 independent nuclear trees are required to generate as much confidence of congruence as obtained with a single organellar tree.

It is tempting to conclude from this discussion that phylogenetic hypotheses derived from nuclear genes will be wholly unreliable at lower taxonomic ranks, and indeed, the foregoing examples for *Zea* illustrate the potential for discord attributable to lineage sorting and coalescence issues. These problems may dissolve, however, for species with small effective population sizes, such as inbreeding species or those whose populations have few individuals. Unfortunately, this information rarely exists in the detail necessary to assess accurately the likelihood that lineage sorting is a factor in allele topologies (see Doyle, 1995 for additional discussion). At present, we must simply be alert to the possibility that it is a factor, especially in phylogenetic analyses at the generic level and below.

A final comment concerns the possibility that lineage sorting may affect tree topologies at higher levels of divergence. This prospect emerges from polymorphisms that are maintained by natural selection through many speciation events (e.g., Gaur et al., 1992). Polymorphisms maintained by balancing selection will have longer coalescence times than neutral loci, and thus may contribute to phylogenetic incongruence at higher taxonomic ranks. Perhaps the most compelling example of this process in plants concerns the self-incompatibility (S) locus in Solanaceae, where high molecular diversity is ensured by enforced outcrossing (Rivers et al., 1993). Phylogenetic and phenetic analyses of alleles at this locus have demonstrated the existence of alleles from different genera that are grouped together in clades (Ioerger et al., 1990; Clark and Kao, 1991; Richman et al., 1996). Some of these allelic polymorphisms have survived what must be an enormous number of speciation events, as coalescence times for alleles shared between *Petunia* and

Solanum appear to precede the divergence of these two genera, which may have last shared a common ancestor over 30 million years ago.

This example from the S locus in the Solanaceae demonstrates that while lineage sorting is generally of importance to population and species-level studies, it may occasionally be significant at higher levels of divergence. In this respect, lineage sorting is similar to introgression of nuclear or cytoplasmic genes, in that these processes also are expected to influence phylogenetic inference mostly at lower taxonomic ranks.

Horizontal Transfer Among the more unexpected conclusions drawn from molecular phylogenetic studies over the past decade has been that genes occasionally traverse species boundaries through nonsexual means. This process, termed horizontal or lateral transfer (Amábile-Cuevas and Chicurel, 1993; Kidwell, 1993; Syvanen, 1994; "xenology" or "paraxenology" of Patterson, 1988), has long been suspected to play a role in the evolution of bacteria (Lan and Reeves, 1996), but in recent years an increasing number of putative examples have been reported from eukaryotes (Houck et al., 1991; Clark et al., 1994; Robertson and Lampe, 1995; Delwiche and Palmer, 1996; Hibbett, 1996; Moens et al., 1996). From a phylogenetic standpoint, horizontal transfer generates discordant trees exactly like the processes of hybridization and introgression, although perhaps only for a single gene and potentially across great phylogenetic distances.

One of the more compelling and well-known examples involves "P element" transposons in *Drosophila* (Houck et al., 1991; Clark et al., 1994). A wealth of phylogenetic and phenetic evidence supports the proposal that these elements first made their way into the *D. melanogaster* lineage via horizontal transfer from a taxon in the *D. willistoni* clade. This evidence includes the near sequence identity of elements extracted from *D. melanogaster* populations that possess the element, the high sequence similarity between elements from the *D. melanogaster* and *D. willistoni* lineages, and peculiarities of their distribution in *D. melanogaster* populations and laboratory stocks. A remarkable and unique feature of this example is that both a likely vector and an ecological context for interspecific DNA transfer have been identified, namely, a mite (*Proctolaelaps regalis*) that feeds on *Drosophila* eggs and larvae. Houck et al. (1991) demonstrated that the mite was capable of carrying P elements subsequent to feeding on fruit fly strains that possessed the transposon, thereby providing a rare glimpse into vector-mediated horizontal transfer.

For most other putative examples of horizontal transfer, the evidence is not as compelling, although there are several recent and interesting exceptions (Delwiche and Palmer, 1996; Hibbett, 1996; Moens et al., 1996). Delwiche and Palmer's study is especially noteworthy, as their report involves the widely utilized gene *rbcL*, for which more sequences exist in plants than any other gene. From phylogenetic analyses of bacterial and plant *rbcL* sequences, Delwiche and Palmer (1996) inferred approximately six horizontal transfer events involving *rbcL* and various combinations of proteobacteria, cyanobacteria, and plastids (see Chapter 13).

These and other examples demonstrate that this non-Mendelian process almost certainly occurs in nature, although its extent and evolutionary significance remain unclear. In addition to the process being relatively rare and difficult to catch in the act, the initial evidence for it in any particular study usually consists of the simple observation of a noteworthy discord between a gene tree and the expected organismal relationships. As is clear from the present chapter, many other evolutionary processes may result in the same observation of phylogenetic incongruence, and these alternatives are not always readily excluded as formal possibilities (Cummings, 1994). Thus, the patterns reported by Delwiche and Palmer (1996) may have resulted, as the authors note, from a complex history of gene duplication and loss, with differential survival in various lineages resulting in "apparent" horizontal transfer. Similarly, the putative horizontal transfer of group I introns into the rDNA of mushroom-forming fungi may also have resulted from a highly homoplasious pattern of intron loss in clades lacking the putative insertion (Hibbett, 1996). Inferences of lateral transfer of members of replicative families, such as P elements and retrotransposons (VanderWiel et al., 1993) are additionally complicated by their

inherent history of duplication and differential survival of orthologous genes (see also section Orthology/Paralogy Conflation).

At present, it seems clear that horizontal transfer occurs, although we do not yet understand its frequency or significance. Given the many other causes of phylogenetic incongruence, it seems unwise, however, to invoke it on this basis alone, without due consideration of the other possibilities and without additional supporting evidence (Cummings, 1994). This evidence may be difficult to garner, particularly if the putative transfer event is ancient, but some of the alternatives may still be addressed using arguments based on parsimony (cf. Delwiche and Palmer, 1996; Hibbett, 1996).

Gene and Genome-level Processes

Most of the evolutionary processes included under this heading cause phylogenetic incongruence because of some manifestation of nonindependence. This applies to both intra- and intergenic recombination, various forms of concerted evolution, evolutionary rate heterogeneity among sites or taxa, and most obviously, when nucleotide sites in a molecule evolve in a directly dependent fashion, such as stemmed bases in ribosomal RNAs. Despite the diversity of underlying molecular mechanisms encompassed by these phenomena, they exhibit the common thread of perturbing relationships among terminals in gene trees through the effects of interaction (among alleles or genes) or nonindependence (among sites), with the consequence that the resulting gene tree is incongruent with other gene trees. Our understanding of these processes is still incomplete, and at present empirical demonstrations of their effects on phylogenetic inference remain reasonably small in number. As a result, the relative significance of each phenomenon to systematists is not known.

Intragenic Recombination An evolutionary history that involves reticulation is not accurately depicted by a strictly bifurcating tree. Trees are, of course, the products of cladistic analysis, and hence the reality of hybridization poses problems for the ideal of phylogeny reconstruction. This problem applies equally to whole organisms, using morphological data sets (Funk, 1985; McDade, 1990, 1992, 1995) and to genes, using sequence data from a sampling of alleles. In the latter case, intragenic (or interallelic) recombination (or gene conversion) leads to the evolution of composite molecules that possess characteristics of both parental alleles. When this occurs, alleles will no longer have arisen exclusively from normal mutational processes, and their relationships may no longer be depicted in a strictly hierarchical fashion. Forcing a treelike topology onto a sampling of alleles that originated from both nonrecombinant and recombinant processes will result in a reconstruction that is, at least in part, erroneous. As a consequence, the resulting gene tree may be incongruent with phylogenetic estimates obtained using other sources of data.

That interallelic recombination occurs is already evident, despite the fact that there are still relatively few studies where multiple alleles of a locus have been sequenced from individual plant taxa. Perhaps the best example concerns alcohol dehydrogenase in maize (Gaut and Clegg, 1993a; Hanson et al., 1996), where recombination was documented among alleles at *Adh1* (nine recombinations among six alleles sampled) and *Adh2* (five recombinants in 12 alleles sampled). Hanson et al. (1996) also suggest that there have been three recombination events among 27 sequences at the anthocyanin regulatory locus, *c1*. Innan et al. (1996) described the evolutionary history of 17 sequences of *Arabidopsis Adh* as having included four intragenic recombination events.

Detecting recombination in a sampling of alleles is not always straightforward, and the reliability of an inference is likely to depend on several factors, including the level of confidence that the alleles involved are from the same locus as opposed to a paralogous locus, the relative antiquity of the recombination event, the amount of sequence divergence between the two alleles involved, and the distribution of differences along the length of the gene. Thus, ancient recombination between two similar alleles is likely to remain undetected, whereas a single recent gene conversion or reciprocal recombination between two highly dissimilar alleles will probably be readily apparent. In the latter case, this may be manifested as a "chimeric" allele, where the 5′

and 3′ portions of the allele (or smaller regions) appear to have different phylogenetic histories and phenetic relationships with other alleles sampled (e.g., Innan et al., 1996). Despite these potential difficulties in detection, several tests for recombination have been developed (Sawyer, 1989; Fitch and Goodman, 1991; Hein, 1993; Templeton and Sing, 1993; Jakobsen and Easteal, 1996) and computer programs that assist in the analysis are available (Hein, 1993; Jakobsen and Easteal, 1996; Hey and Wakeley, 1997).

In some cases, recombination may be so extensive that it will remain undetected as such except under extraordinary circumstances. A case in point concerns the rDNA ITS from *Gossypium gossypioides,* which available evidence suggests was derived from recombination between two highly divergent ITS sequences that were brought together in a common nucleus as a result of interspecific hybridization (Wendel et al., 1995b). The ITS sequence in question is not a simple chimeric product, but instead consists of a mosaic of nucleotides contributed by both parental clades distributed in an apparently random fashion along the length of the sequence. In a follow-up study, phylogenetic data were gathered that suggest that the 5S rDNA locus in *G. gossypioides* was similarly recombined (Cronn et al., 1996).

Recombination, like hybridization, causes homoplasy when included in a phylogenetic analysis. Consequently it is not surprising that recombinant alleles behave cladistically like morphological hybrids (Funk, 1985; McDade, 1990, 1992, 1995) in that they resolve in relatively basal positions in clades occupied by one parent or the other. This prediction is fulfilled by the recombinant rDNA alleles from *Gossypium gossypioides,* as shown in Fig. 10.4. The three

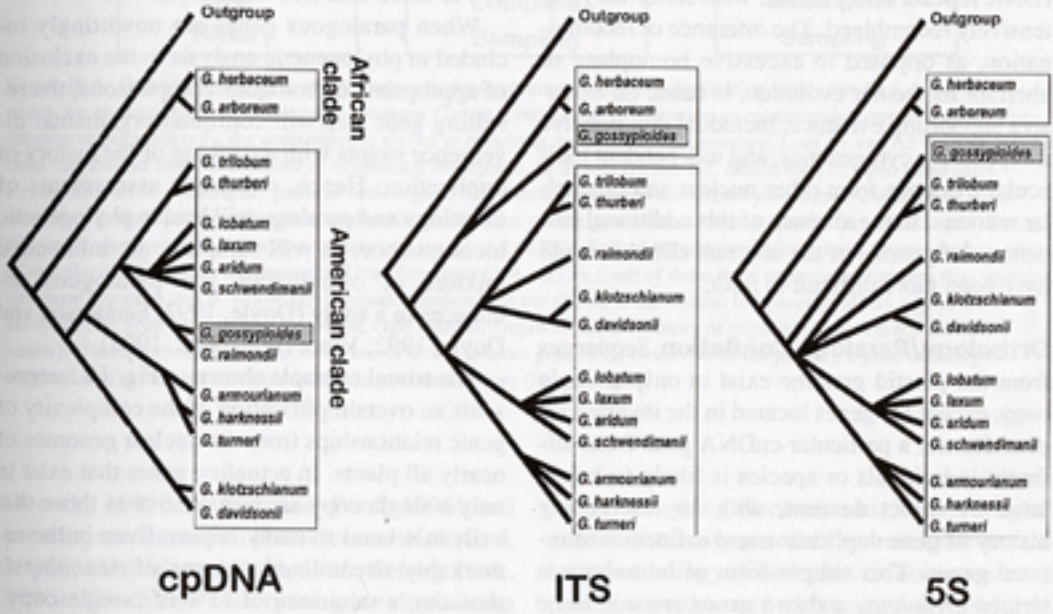

Figure 10.4. Effect of recombination on phylogenetic resolution. Shown on the left are relationships among American and African diploid species of *Gossypium,* as inferred from cpDNA restriction site variation (Wendel and Albert, 1992; Seelanan et al., 1997). The primary division is into strongly supported clades consisting of diploids from different hemispheres. Within the American clade, *G. gossypioides* appears as the sister taxon of *G. raimondii,* which is the conventional placement supported by most other genetic and cytogenetic data (Wendel et al., 1995b). Evidence suggests, though, that the history of *G. gossypioides* includes an episode of hybridization and introgression with a genome like that presently found in African diploid species. This hypotheses was initially invoked to account for phylogenetic trees that are incongruent with the conventional tree shown on the left. Specifically, in trees based on both ITS (center) and 5S (right) ribosomal DNA sequences (redrawn from Wendel et al., 1995b and Cronn et al., 1996, respectively) *G. gossypioides* occupies phylogenetically basal positions, although in different primary clades in each case. This basal resolution is consistent with the highly recombined nature of the ribosomal repeats in *G. gossypioides,* and illustrates that the phylogenetic effects of molecular recombination are similar to those of morphological hybridization (Funk, 1985; McDade, 1990, 1992, 1995).

panels illustrate phylogenetic observations from a nonrecombinant molecule (cpDNA, left), which cladistically resolves in a terminal "nonrecombinant" position, and from the two recombinant rDNA genes (ITS, center; 5S, right), which resolve basally, although in different parental clades in each case.

Recombination may be difficult to distinguish from other causes of phylogenetic discord, such as undetected and unexpected interspecific introgression. In some cases, an inference of recombination may be justified by the pattern of molecular variation, such as when an obvious chimeric allele is observed. In other cases, the source of the conflict may not be so readily apparent. With respect to the peculiar rDNA sequences of *Gossypium gossypioides, both* processes appear to have been involved. Specifically, an organismal history of hybridization and introgression is suggested to have brought "native" and "alien" rDNA repeats into contact, whereafter they extensively recombined. The inference of recombination, as opposed to excessive homoplasy or aberrant molecular evolution, is based on extensive supporting evidence, including comparative morphology, cytogenetics, and independent molecular evidence from other nuclear and organellar markers. In the absence of this additional evidence, the genesis of the aberrant rDNAs would have been more difficult to infer.

Orthology/Paralogy Conflation Sequences from the plastid genome exist in only a single copy, except for genes located in the inverted repeat. Hence, a particular cpDNA gene from different individuals or species is likely to be related by direct descent, with no intervening history of gene duplication and extinction of related genes. This simple form of homology is termed *orthology,* and two genes are said to be *orthologous* if their relationship originated from organismal cladogenesis (see Chapter 4). Orthologous sequences are appropriate for phylogenetic analysis, in that their history may reveal organismal divergence events, assuming the absence of the other potentially confounding influences described in this chapter.

Genic relationships may take forms other than orthology, however. Consider the case, for example, where a gene becomes duplicated (yielding locus A and locus B) and each copy of the gene is transmitted to two daughter lineages (1 and 2) following an organismal divergence event. In this case there are two sets of orthologous relationships, one involving locus A from both lineages (A1 vs. A2) and the other involving locus B from both lineages (B1 vs. B2). In addition, however, there are relationships among nonorthologous genes from different species (A1 vs. B2 and A2 vs. B1) as well as between the duplicated copies within each species (A1 vs. B1, A2 vs. B2). In this example the duplicated sequences within and between lineages exhibit *paralogy,* and the genic relationships are termed *paralogous* (Fig. 10.5). Paralogy is simpler to define than orthology, in that a rigorous criterion exists: If two loci occur in the same individual genome, then they must be paralogous. Orthology, however, describes a genetic relationship that is inextricably tied to a phylogenetic concept involving two or more taxa (see Chapter 4).

When paralogous genes are unwittingly included in phylogenetic analysis to the exclusion of appropriate orthologous comparisons, the resulting gene tree will confound organismal divergence events with a tracking of the history of duplication. Hence, erroneous assessments of orthology and paralogy will lead to phylogenetic incongruence, as will sampling an unintended mixture of orthologous and paralogous sequences in a study (Doyle, 1992; Sanderson and Doyle, 1992; VanderWiel et al., 1993).

The trivial example shown in Fig. 10.5 represents an oversimplification of the complexity of genic relationships from the nuclear genomes of nearly all plants. In actuality, genes that exist in only a single copy are rare relative to those that exist in several to many copies. Even in the remarkably streamlined genome of *Arabidopsis thaliana,* a minimum of 15% of "single-copy" nuclear sequences are duplicated (Kowalski et al., 1994; McGrath et al., 1993). Most nuclear genes from most plants exist in several to many copies, these reflecting a complicated history of sequence duplication, through polyploidization (Masterson, 1994), retrotransposition (VanderWiel et al., 1993; Wessler et al., 1995; Kumar, 1996), or other mechanisms, as well as processes of sequence loss, via genic and/or chromosomal deletion (e.g., Dubcovsky and

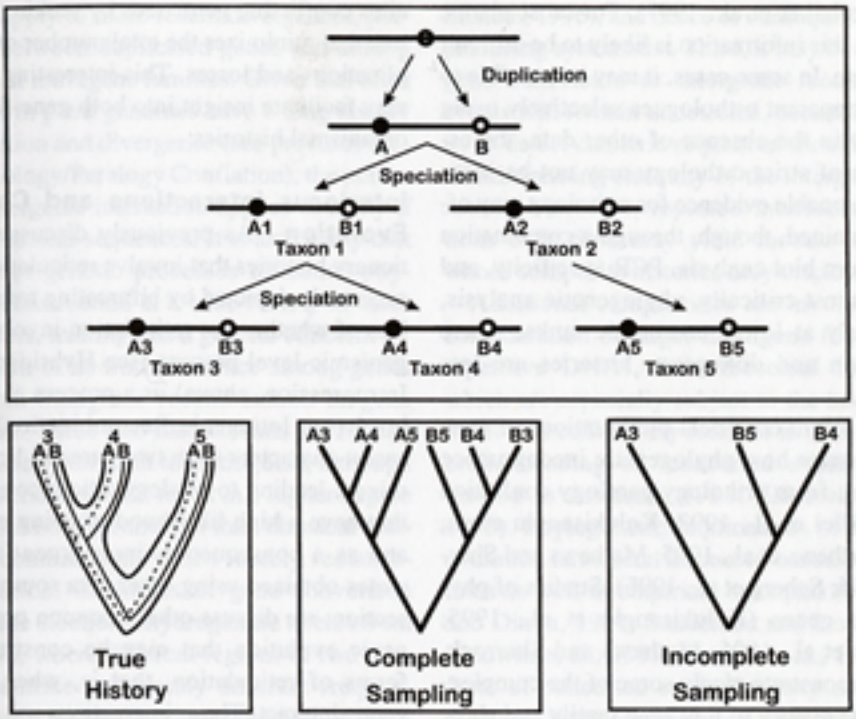

Figure 10.5. Orthology, paralogy, and phylogenetic incongruence. Top: Gene duplication yields two loci, A and B, which are each transmitted into daughter lineages (Taxon 1 and Taxon 2) following an organismal divergence event. For illustrative purposes only, the duplication is shown as occurring on the same chromosome. Loci A1 and A2 become distinct from one another as a consequence of organismal divergence, as do loci B1 and B2; hence, each of these locus pairs represents a set of *orthologous* genes or *orthologues*. Nonorthologous genes are termed *paralogues*. *Paralogous* genes arise from gene duplication; the term applies to relationships among genes from different species (A1 vs. B2 and A2 vs. B1) and to duplicated copies within species (A1 vs. B1; A2 vs. B2). Following additional divergence events, the two sets of orthologues (A and B) are transmitted into three taxa (3, 4, and 5). Each of these three organismal lineages thus contains two gene copies, A and B, whose phylogenetic relationships are shown in the bottom left panel. If all six copies are sampled, two gene trees will be recovered, each faithfully reproducing the history of organismal divergence (bottom center). If sampling is incomplete, however, or if one or more genes fails to survive to the present, a gene tree may be obtained that is incongruent with the true organismal history (bottom right).

Dvorák, 1995; Gottlieb and Ford, 1996) and pseudogene formation (Buckler and Holtsford, 1996b; Gottlieb and Ford, 1996; Seberg et al., 1996). Given this history of gene duplication and loss, homology for any given gene system from two or more species *may* include orthologous comparisons, but will almost certainly include a diverse array of paralogous relationships. Duplications responsible for the latter may have been ancient or more recent. Hence paralogy is a relative concept (VanderWiel et al., 1993), as two genes may be more or less paralogous than two other genes, depending on the relative recency of their shared common ancestry.

The foregoing discussion highlights why sequence similarity per se is an unreliable indicator of orthology. Given that orthologous comparisons are the only ones potentially useful for phylogeny reconstruction, it is essential that orthology be established prior to a phylogenetic study. Orthologous relationships can be hypothesized with varying degrees of confidence from using criteria such as tissue specificity (Doyle, 1991), expression patterns (Doyle, 1994a), Southern blot analysis (Matthews and Sharrock, 1996), and perhaps most convincingly, comparative mapping (Bonierbale et al., 1988; Whitkus et al., 1992; Kowalski et al., 1994; Pereira et al.,

1994; Brubaker et al., 1997). For most plants, however, this information is likely to be difficult to generate. In some cases, it may be possible to amplify apparent orthologues selectively using PCR, but in the absence of other data, the assumption of strict orthology may not be justified. Reasonable evidence for orthology may often be obtained, though, through a combination of Southern blot analysis, PCR specificity, and perhaps most critically, phylogenetic analysis, particularly at lower taxonomic ranks, where duplication and divergence histories are less complex.

Many recent empirical presentations of gene trees illustrate how phylogenetic incongruence may result from orthology/paralogy conflation (VanderWiel et al., 1993; Kolukisaoglu et al., 1995; Mathews et al., 1995; Mathews and Sharrock, 1996; Seberg et al., 1996). Studies of phytochrome genes (Kolukisaoglu et al., 1995; Mathews et al., 1995; Mathews and Sharrock, 1996) demonstrate nicely some of the complexities in the history of this gene family and show that different orthology and paralogy issues apply to various levels in the taxonomic hierarchy. VanderWiel et al. (1993) show how incongruence with organismal relationships is the *expected* outcome for phylogenetic analyses of multigene families such as retrotransposable elements, due to the complex history of duplicative transposition combined with incomplete sampling.

A somewhat different type of confounded orthology/paralogy relationship may arise in cases involving organellar genes where there has been duplicative gene transfer to the nucleus during organellar evolution (Nugent and Palmer, 1991; Gantt et al., 1991). If nuclear copies of the duplicated gene are sampled from only a portion of the taxa included in a study, gene trees may show unexpected relationships. To the extent that the underlying phenomenon may be diagnosed, the inferred phylogeny may therefore provide critical information on the timing of the gene transfer event.

A final comment concerns the intriguing possibility that a "species tree" may be estimated from gene trees even in the presence of confounded orthology/paralogy relationships. Guigó et al. (1996) present a method that reconciles conflicting gene trees by seeking a species tree that minimizes the total number of gene duplications and losses. This interesting approach may facilitate insight into both gene-family and organismal histories.

Interlocus Interactions and Concerted Evolution As previously discussed, evolutionary histories that involve reticulation are not accurately depicted by bifurcating trees, regardless of whether the reticulation involves an organismic-level process (see Hybridization and Introgression, above) or a process at the genic level (see Intragenic Recombination). In both cases, characters from two parental lineages are mixed, leading to phylogenetic reconstructions that have a high likelihood of being erroneous, and as a consequence, incongruent with estimates obtained using other data sources. In this section, we discuss other common processes in genic evolution that may be construed to be forms of reticulation, that is, when different genes interact. These interactions are multifarious and complex, but they all lead to nonindependent character evolution, through the recombining, blending, or altering of different DNA sequences.

Intergenic recombination may influence gene tree interpretation at any level in the taxonomic hierarchy. In a phylogenetic reconstruction of serine protease genes from humans, for example, different trees were obtained from different domains of the proteins; these were interpreted as reflecting ancient mosaicism, or "exon shuffling" between formerly independent gene regions (Ikeo et al., 1995). This example is only one of many of exon shuffling, mostly from vertebrates (Stone and Schwartz, 1990; Patthy, 1991; Doolittle, 1995) but including at least a couple of examples from higher plants (Domon and Steinmetz, 1994; Schena and Davis, 1994). To the degree that proteins have been assembled by this process, their gene sequences will be reticulate, which has obvious ramifications for phylogenetic analysis.

The extent to which exon shuffling needs to be considered by plant systematists is not known, but it is likely that the phenomenon applies primarily to the deepest levels in plant phylogeny. Other, better-known forms of genic in-

teraction apply to more recent divergences, particularly between duplicated genes and among members of multigene families. Given that most sequences in plant genomes have a long history of duplication and divergence (see previous section, Orthology/Paralogy Conflation), the potential for intergenic interaction applies to many if not most nuclear sequences. It is also likely that the relevant genetic processes influence phylogeny reconstruction at a wide variety of taxonomic levels, making this a general concern.

One form of nonindependence among genes arises from intergenic recombination or gene conversion. These two mechanisms are related and are usually difficult to distinguish, although gene conversion should result in a higher degree of sequence homogenization than classical reciprocal recombination, which merely redistributes variation. Examples of gene conversion may include alcohol dehydrogenase alleles from *Gossypium*, where different regions of two alleles demonstrate remarkably different frequencies of indels and nucleotide substitutions (Millar and Dennis, 1996). In *Pisum*, duplicated copies of a glutamine synthetase gene are over 99% identical at the nucleotide level, except in the middle portion of the genes. Walker et al. (1995) attribute high sequence similarity in the distal regions to the homogenizing influence of gene conversion, and interpret the reduced similarity in the middle part of the genes as representing unconverted portions of sequence. Similarly, paralogous copies of *PgiC* in some species of *Clarkia* may have experienced gene conversion (Gottlieb and Ford, 1996).

These examples have several implications. First, in each case, the chimeric or recombinant nature of the genes involved was instrumental in revealing an important aspect of their evolutionary history, namely, interaction with other, similar sequences. Thus, the reticulation that confounds phylogeny reconstruction yields insights into molecular evolutionary process. Second, there are still relatively few molecular systematic investigations using nuclear genes (other than rDNA, but see below), yet given even this small number, phylogenetically significant gene conversion and recombination events have been detected, suggesting that the process may be common, as has been shown in yeast (Goldman and Lichten, 1996), and hence of consequence for the practicing systematist. Third, it may be that most gene conversion or intergenic recombination events will remain undetected, because the ability to detect chimeric sequences should decrease with increasing antiquity of the intergenic interaction. Moreover, repeated interlocus interactions over time may yield mosaic sequences whose composite histories defy disclosure.

Additional complexities are involved in the consideration of larger multigene families and repetitive DNAs, like ribosomal RNA loci, which are especially subject to the homogenizing forces collectively referred to under the umbrella heading of "concerted evolution" (reviewed in Arnheim, 1983; Elder and Turner, 1995). Phylogenetic implications of concerted evolution have been discussed based on simulations as well as empirical examples (e.g., Hillis and Dixon, 1991; Sanderson and Doyle, 1992; Baldwin et al., 1995; Wendel et al., 1995a). Effects of concerted evolution may range from complete homogenization of multigene families to occasional gene conversion events that affect only a small portion of the members of a given multigene family. Meagher et al. (1989), for example, showed that gene conversion has played a prominent role in the evolution of RUBISCO small subunit genes (*rbcS*) in plants, whereby sequences from single lineages are more similar to each other than would be expected under a scenario of independent evolution. Concerted evolution is incomplete, however, as evidenced by the considerable *rbcS* diversity that is retained within species. A similar process of partial and intermittent homogenization has been invoked for actin evolution in angiosperms (Moniz de Sá and Drouin, 1996). This latter study demonstrates that gene conversion events have occurred at various times during evolution of the actin gene family, and that these may be detected at phylogenetic depths ranging from the rank of class (Liliopsida) to genus (*Zea*).

The mechanisms by which these occasional gene conversion events take place, and the forces responsible for their sporadic nature, are incompletely understood. From a phylogenetic perspective, though, these examples of partial homogenization pose real problems for interpretation of both organismal and gene-family

history (Sanderson and Doyle, 1992). A reasonably accurate gene tree is a minimum requirement for inferring a species tree, and this may be obtained both in cases where concerted evolutionary forces are absent and when they are strong enough to maintain relative infraspecific homogeneity. Partial homogenization, however, generates recombinant sequences that are neither orthologous nor paralogous. Hence, no history of bifurcation among genes or organisms may be recovered. Consequently, nuclear genes that are subject to *incomplete* concerted evolution are expected to be the least useful for reconstruction of organismal relationships.

A variant on this theme of incomplete concerted evolution involves cases where the degree of homogenization varies among taxa. This variation has been shown in legumes, for example, where both leghemoglobin (Doyle, 1994b) and 7S seed storage protein genes (Doyle et al., 1992) appear to evolve concertedly in some but not all legume taxa. This process may lead to cases where orthologues of particular genes in a taxon experiencing weak concerted evolutionary forces may simply no longer exist in related taxa where the genes have been homogenized by stronger concerted evolutionary processes.

Concerted evolutionary forces are often stronger for tandemly repeated sequences, such as ribosomal genes, than they are for dispersed members of multigene families spread across several or more chromosomal loci, at least in part because one major mechanism, unequal crossing over, is only possible for tandemly arranged families. Ribosomal DNA genes are organized into arrays consisting of hundreds to thousands of identical to nearly identical repeats, indicative of their concerted mode of evolution. This feature of rDNA, in addition to their high copy number, has been responsible for the rapid growth and utilization of these sequences in phylogeny reconstruction (Hillis and Dixon, 1991; Baldwin et al., 1995; see Chapter 7).

Numerous phylogenetic studies in plants have been conducted using the 18S–26S rDNA repeat, allowing several insights into the various possibilities for molecular evolution of repeated sequences; each of these has a differing relevance to phylogenetic inference and phylogenetic incongruence. One possibility is that homogenization of repeats may be incomplete, and indeed, considerable sequence variation may exist among repeats within species. Buckler and Holtsford (1996b), for example, in an analysis of ITS sequences, demonstrate both intra-array sequence diversity and the presence of rDNA pseudogenes. Because of incomplete homogenization, interspecific orthology is thus not assured. Similarly, inter-array homogenization may not occur, even in circumstances where rDNA repeats within arrays are reasonably homogeneous. This results in the long-term maintenance of more than one rDNA repeat type (e.g., Suh et al., 1993; Kim and Jansen, 1994; Sang et al., 1995; Waters and Schaal, 1996), which may in some cases actually preserve paralogous relationships. An additional evolutionary outcome, especially in cases where divergent repeats are united in a common nucleus as a result of hybridization or polyploidy, is homogenization of repeats within (intralocus concerted evolution) and among (interlocus concerted evolution) arrays (Wendel et al., 1995a; Brochmann et al., 1996). Wendel et al. (1995a) showed that this process might occur bidirectionally in allopolyploids derived from a common progenitor bearing divergent rDNA arrays, such that different derivative taxa are fixed for alternative repeat types. This situation is formally equivalent to the random loss of different paralogous copies of duplicated genes in different lineages, a process that also has been demonstrated for rRNA genes (Dubcovsky and Dvorák, 1995; Danna et al., 1996).

These examples bear witness to the many possibilities for intergenic interactions in the evolution of multigene families, and accordingly, illustrate the complexities that need to be considered if they are to be used in phylogeny reconstruction, especially at the level of genus and below. There is irony in the fact that much of what has been learned about the evolutionary behavior of these multigene families was first revealed by the discordant gene trees they produce, once again underscoring the enhanced understanding of process that emerged from an initial observation of phylogenetic incongruence.

Rate Heterogeneity among Taxa A common observation in molecular systematic stud-

ies is evolutionary rate heterogeneity among the sequences sampled (Wu and Li, 1985; Britten, 1986). This phenomenon is widely acknowledged (Bousquet et al., 1992; Gaut et al., 1992, 1993), and is evidenced either intuitively by the observation of dramatic branch length variation in phylogenetic reconstructions, or by statistical tests of rate homogeneity (Wu and Li, 1985; Tajima, 1993; Takezaki et al., 1995; Gaut et al., 1996, 1997; see Chapter 9). Substitution rate heterogeneity may have a number of different underlying causes, but in some cases it is partially accounted for by variation in generation times, with which rates are often inversely correlated (Wu and Li, 1985; Li et al., 1987; Ohta, 1993; Gaut et al., 1996, 1997).

Regardless of the causative mechanism, rate heterogeneity is of interest from a phylogenetic perspective. In particular, when some branches are long and others are short, the long branches may experience long-branch attraction (Felsenstein, 1978), due to parallel substitutions in two lineages becoming mistakenly identified as actual synapomorphy. This phenomenon has been much discussed and is appreciated as a cause of misleading phylogenetic inference, and hence as a cause of phylogenetic incongruence.

A striking example both of rate heterogeneity and its influence on phylogeny reconstruction is offered by 18S rDNA sequence variation in parasitic and nonparasitic angiosperms (Nickrent and Starr, 1994; see Chapter 8). Relative rate tests showed that substitution rates were greatly accelerated in holoparasitic taxa: When compared with *Glycine*, the mean number of substitutions per site (K) for five autotrophic angiosperms was 0.036, whereas for four holoparasites the value was 3.5 times higher ($K = 0.126$). Nickrent and Starr (1994) further discuss how the accelerated substitution rates in the parasitic angiosperms confound attempts to deduce their phylogenetic placement, presumably due to long-branch attraction.

It is of interest to note that long-branch attraction has at least two fundamentally different causes. One, rate heterogeneity, is a molecular evolutionary phenomenon and thus it might be modeled in such a way that it is accommodated in phylogenetic reconstruction methods. The other cause relates to taxon sampling density, as discussed earlier under Taxon Sampling. In this case it may be possible to "break up" long branches by a strategic increase in sampling (Wheeler, 1992; see, however, Kim, 1996), although the long branches may also arise from natural divergence and extinction patterns that are beyond investigator control. Distinguishing whether long branches reflect molecular evolutionary phenomena or sampling phenomena may not always be straightforward, although in most cases tests of rate heterogeneity (Wu and Li, 1985; Tajima, 1993; Takezaki et al., 1995; Gaut et al., 1996, 1997; see Chapter 9) should prove diagnostic.

Rate Heterogeneity among Sites One of the more obvious aspects of DNA sequence data for protein-encoding genes is that the amount of variation detected varies by codon position. This is for the most part a straightforward consequence of the degeneracy of the genetic code by codon position in conjunction with the operation of selective contraints. Given that 70% of the possible nucleotide changes in the third position are silent, whereas nearly all substitutions in the first and second codon positions result in amino acid replacements, it is not surprising that in most molecular systematic studies greater variation is detected at the third position than in the other two codon positions. This is an example of among-site rate heterogeneity, but it is not the only one. Substitution rates among sites may vary systematically in non-protein-encoding sequences just as they do among different regions or domains of protein-encoding genes (see Chapter 6).

When among-site rate variation exists, evolutionary change may be concentrated in a minority of nucleotide positions, while many other sites experience few to no substitutions. In many cases, this circumstance poses few problems vis-à-vis phylogeny estimation, particularly when substitution rates at the variable sites are low enough that evolutionary history is not obscured by multiple hits. This characterization may apply, for example, to most published studies at the family level using *rbcL*, where a preponderance of changes are in third codon positions. In other instances, third positions may be too homoplasious, while other sites provide insufficient information. Consequently, sequences

that are characterized by severe rate variation tend to be less informative on a per nucleotide basis than those with a more even distribution of substitutions among sites.

A good example of this rate distributional phenomenon is provided by Steele and Vilgalys (1994) in their comparison of *matK* and *rbcL* sequence evolution. They showed that the former sequence evolves at twice the rate of the latter, but in addition, that substitutions are more evenly distributed among codon positions. In a comparison of tobacco and rice *rbcL* genes, for example, 79% of differences are at third positions, whereas for *matK*, this value is much lower (50%). Thus, the likelihood of multiple substitutions that obscure phylogenetic information is greater in the case of the more rate-heterogeneous (*rbcL*) gene. Similar results have been reported for other taxonomic groups; in a comparison of 25 saxifragaceous taxa, for example, substitutions are far more evenly distributed among codon positions for *matK* than for *rbcL* (Johnson and Soltis, 1995).

Considerable attention has recently been given to among-site rate variation and its potential phylogenetic consequences (Kuhner and Felsenstein, 1994; Wakeley, 1994; Sullivan et al., 1995; Sullivan, 1996; Yang, 1996; see Chapter 6). These consequences include phylogenetic incongruence, as exemplified by the study of Sullivan et al. (1995) on murid rodents. In this case, cytochrome *b* gene sequences, but not 12S rRNA gene sequences, yielded a phylogeny that is well corroborated by morphology and other molecular data sets. Sullivan et al. (1995) attributed the misleading signal in the latter gene to extreme among-site rate variation.

Because of the potential for among-site rate variation to yield misleading phylogenetic signal and hence phylogenetic discord, approaches that accommodate among-site rate variation in phylogenetic analysis have been developed. These often involve likelihood and distance methods, as recently reviewed by Yang (1996; see Chapter 5).

Base Compositional Biases In most DNA sequences the four different nucleotide bases do not occur in equal proportions. This may be true for sequences as a whole, for different partitions such as stemmed and looped portions of rRNA molecules, and, in the case of protein-encoding genes, for different codon positions (e.g., Vawter and Brown, 1993; Martin, 1995; Hershkovitz and Zimmer, 1996). These compositional biases may be associated with asymmetries in nucleotide transformational probabilities (Albert and Mishler, 1992; Albert et al., 1993; Vawter and Brown, 1993; Collins et al., 1994b), which may systematically distort patterns of character-state reconstruction. In response to recognition of these biases, a number of corrective measures have been developed for inclusion in phylogenetic reconstruction methods (Albert and Mishler, 1992; Albert et al., 1993; Knight and Mindell, 1993; Collins et al., 1994a; Lockhart et al., 1994; Gaultier and Gouy, 1995).

Compositional bias, and the related phenomenon of transformational bias, may lead to misleading phylogenetic inferences, as noted in several studies (Lockhart et al., 1994; Martin, 1995; Steel et al., 1995; see Chapter 5). In particular, taxa typically become spuriously linked by virtue of sharing a compositional bias, such as a high G+C content. A case in point is the study by Lockhart et al. (1994), who showed that the apparent support for a bird/mammal clade in vertebrate 18S rRNA sequences is due to independently acquired compositional similarities, and that this support evaporates when compositional biases are accounted for. Once the data are transformed (using LogDet transformation) there is actually strong support for a bird/crocodilian relationship using the same data. Similarly, inferred phylogenetic relationships in sharks are distorted by compositional variation in cytochrome *b* sequences (Martin, 1995). Because compositional biases may lead to erroneous reconstructions, as these examples show, they may also be a source of phylogenetic incongruence with other gene trees or trees derived from morphological data sets.

RNA Editing In this ubiquitous process, mRNA sequences, primarily from organellar genes, are modified by the editing of particular bases, usually by conversion of C to U, but also by U to C substitutions, at least in plant mitochondria (Gray and Covello, 1993; Araya et al., 1994; Hiesel et al., 1994; Malek et al., 1996;

Yoshinaga et al., 1996). As a result of these nucleotide changes, the genetic information in the transcript may differ from that of the gene, leading to a translation product with one or more amino acids that would not have been predicted by the DNA sequence. Explanations for the existence of RNA editing include preservation of highly conserved amino acids (e.g., Hirose et al., 1994), restoration of function to DNA sequences that have suffered mutations that create premature stop codons (Yoshinaga et al., 1996), and compensation for T to C transitional drift (Malek et al., 1996).

That this process may be a potential source of phylogenetic incongruence was recently highlighted in phylogenetic analyses of plant mitochondrial *cox* genes (Bowe and dePamphilis, 1996; C. dePamphilis, pers. comm.). As discussed by Bowe and dePamphilis (1996), genomic sequences that undergo RNA editing are appropriate to include in phylogenetic analysis, because the process operates at the transcriptional level and thus should not affect historical information stored in the DNA sequences. Inclusion in an analysis of both DNA sequences and mRNA sequences that have been edited, however, may lead to conflict, as mRNA sequences will be spuriously linked by apparent synapomorphies of edited bases at positions affected by RNA editing. Because mRNAs (via cDNA synthesis) are sometimes incorporated into phylogenetic studies, especially when sequences are down-loaded from databases, this may be a practical consideration. RNA editing is also a concern, however, due to a peculiar form of orthology and paralogy conflation. Specifically, Bowe and dePamphilis (1996) show that phylogenetic incongruence may arise from inclusion of a mixture of unprocessed DNA sequences with processed genes ("processed paralogs" in their terminology) resulting from genomic integration of reverse-transcribed, previously edited RNAs. This result reinforces the necessity, invoked above, of documenting orthology during phylogenetic analysis. In addition, this example underscores the insights into molecular evolutionary history that may be revealed through phylogenetic discord. In this respect RNA editing may be distinguished from other potential sources of paralogy by careful analyses of potentially edited sites (Nugent and Palmer, 1991; Bowe and dePamphilis, 1996).

Nonindependence of Sites One of the assumptions of phylogenetic analysis of nucleotide sequences is that each position is independent of other positions, at least in models where characters (= positions) are equally weighted (see, however, Albert et al., 1993, for an example of character weighting). While the condition of complete independence among sites is probably never met in a strict sense, in many instances it may not be too seriously violated. For some types of sequences, however, structural considerations alone predict that the assumption of independence is unwarranted. Perhaps best known in this respect are genes that encode ribosomal RNAs, which are single-stranded but have a secondary structure that includes stemmed regions containing Watson-Crick base pairs. Because rRNA is divided into domains where bases are either paired or unpaired, it is likely that different evolutionary constraints operate in each case.

Stemmed bases are subject to selection for compensatory mutations so that base-pairing is maintained, although phylogenetic evidence indicates that an unpaired base may persist for some time following a mutation, even if it does ultimately become compensated (Gatesy et al., 1994). Because mutations typically become compensated, some thought has been given to differential weighting of paired and unpaired nucleotide positions for purposes of phylogenetic analysis (see Chapter 7). Dixon and Hillis (1993) studied 28S RNA genes from selected vertebrates and recommend reducing the weight accorded stem characters by 20% relative to loop characters, whereas Springer et al. (1995), in a study of 12S rRNA gene sequences from mammals, suggest a more extreme weighting of stemmed positions (circa 40% downweighting). Though these weighting recommendations differ quantitatively, they underscore the possibility of nonindependence of nucleotide positions in sequence data. Importantly, Dixon and Hillis (1993) further show how alternative weighting can modify the phylogenetic resolution obtained (see, however, Chapter 7), raising the possibility that failure to account for nonindependence

among sites may lead to discordance with other phylogenetic estimates.

ASSESSING THE CAUSE

In this chapter we have drawn attention repeatedly to the enhanced understanding of evolutionary process that has derived from observations of discordant phylogenies. Given that many phenomena may lead to the same phylogenetic "phenotype," that is, incongruence, the question naturally arises as to how to discriminate among the many possible causes when one is faced with a particular unexpected or conflicting resolution. Often, there may be insufficient information to render a confident diagnosis, and the cause of incongruence simply remains unknown. Rodman et al. (1993), for example, report discordant resolutions for the Limnanthaceae in a comparison of trees based on morphological and molecular data, and note that the explanation for incongruence is obscure. Unexplained phylogenetic discord, as in this example, is prevalent, and it is also disquieting in that it hampers robust phylogenetic conclusions for the taxa involved, without an apparent rationalization that might yield insight into process.

This difficulty of inferring the underlying phenomenon when different causes result in similar phylogenetic observations is a common and vexing problem. Take lineage sorting and introgression, for example, which often leave the same incongruence footprint. Mason-Gamer et al. (1995) note the difficulty of distinguishing these causes in their analysis of the distribution of cpDNA haplotypes in *Coreopsis*. Hanson et al. (1996) and Buckler and Holtsford (1996a) note a similar difficulty in distinguishing between the two alternatives for shared alleles at nuclear loci (*c1* and ITS, respectively) in *Zea*.

Given an observation of incongruence and in the absence of other information, it is unlikely that its cause will be confidently diagnosed in many cases. An escape from the impasse is often provided by taking advantage of ancillary information and other data sources. Geographic distribution data, for example, may be instrumental in favoring introgression over lineage sorting by virtue of putative introgressant alleles occurring only in areas of sympatry (Rieseberg and Wendel, 1993). Alternatively, it may be possible to distinguish between lineage sorting and introgression on the basis of the distribution of alleles at other loci and by the occurrence patterns of alleles among individuals. Although both introgression and lineage sorting may affect loci widely distributed in the genome, in some cases these two processes may be distinguished, particularly if most putatively introgressed alleles co-occur in the same individuals while most other individuals appear pure (Brubaker and Wendel, 1994; Brubaker et al., 1993). Under a scenario of lineage sorting, the marker alleles in question might be expected to be more randomly distributed among individuals. Quantitative arguments may also facilitate diagnosis of cause, as for example in many of the cases of plastome introgression (Rieseberg and Soltis, 1991; Reiseberg and Wendel, 1993; Rieseberg, 1995; Rieseberg et al., 1996). To attribute the alien cpDNA to ancestral retention may invoke the long-term maintenance of a polymorphism level that is deemed improbable, and hence the alternative of hybridization and introgression may be preferred. A clever, complementary approach was recently suggested by Hanson et al. (1996), who, in noting the presence of *c1* alleles shared among *Zea* taxa, used postulated divergence times and mutation rates in simulations of sequence evolution, with the aim of addressing the likelihood that alleles could have survived since speciation without accumulating differences. The authors take pains to point out that their method does not preclude the possibility of introgression, merely that introgression need not be invoked to account for alleles shared among taxa.

Additional quantitative arguments may derive from parsimony itself. When confronted with a possible case of trans-specific, nonsexual gene flow (horizontal transfer), it is necessary to consider the formal alternative of differential duplications and losses of genes, leading to a mixed sampling of orthologous and paralogous sequences (Fig. 10.5; see also fig. 4 in VanderWiel et al., 1993). Resolution often relies on parsimony criteria, where, for example, the number of homoplasious gene gains and/or losses that would need to be invoked is deemed too high (Delwiche and Palmer, 1996; Hibbett, 1996).

These examples, and many others cited in this chapter, indicate that a holistic approach often will prove helpful in assessing the cause of phylogenetic incongruence. It is difficult to generalize about recommended approaches or guidelines, however, because each instance of incongruence has peculiarities unique to the taxa involved and their histories. When a diversity of evidence is harnessed, though, the discord often dissolves into understanding.

CONCLUSIONS

Our intention in preparing this chapter was to draw attention to the many biological processes that lead to discordant phylogenetic hypotheses by providing brief explanations of the various phenomena and providing pointers to relevant literature. In addition, though, we hoped to emphasize the many insights into evolutionary process that were stimulated by initial observations of incongruence. These insights have been many and varied, and have led to an enhanced appreciation of phenomena that operate at the organismal as well as molecular levels. A partial list would include issues surrounding the terms *gene tree* and *species tree;* the phenomena of lineage sorting and introgression; the potential for reticulation and/or concerted evolution among alleles, genes, and among organisms; the possibility of myriad forms of hybrid or recombinational speciation as well as rapid radiation; and the potential significance of horizontal gene transfer in evolution. Each of these processes confounds phylogeny reconstruction while offering insight into some aspect of the biology of the organisms under study. This is an impressive list, and it includes phenomena about which little was known prior to molecular systematic investigations that led to an alerting and informing incongruence. Accordingly, one might justifiably argue that an enhanced understanding of the evolutionary process is the most important achievement of molecular systematics.

Perhaps it is of more than just historical interest to note the interplay between the tools used to infer phylogeny and our appreciation of the appropriate use of those tools. In our effort to infer phylogeny, incongruence has emphasized the need to understand the evolutionary history of the data sources used. These interconnected objectives have often involved reciprocal illumination, whereby phylogenetic analysis using a particular molecular tool has led to incongruence, which has led to increased knowledge of the evolutionary forces that impact the use of that particular tool in phylogeny reconstruction. A case in point would be ribosomal DNA, where phylogenetic analysis led to recognition of the possibility of interlocus concerted evolution, which has obvious relevance to orthology and paralogy concerns. Similarly, phylogenetic use of rDNA sequences has enhanced our understanding of character and character-state weighting issues, which led to refinements in phylogenetic methodology.

So we return to the premise of the chapter: Phylogenetic incongruence is not merely an evil to be reconciled, but is instead a glimmer of light. When it shines with sufficient brightness, we are offered a window through which evolutionary process may be seen more clearly.

LITERATURE CITED

Albert, V. A., and B. D. Mishler. 1992. On the rationale and utility of weighting nucleotide sequence data. Cladistics **8:**73–83.

Albert, V. A., S. E. Williams, and M. W. Chase. 1992. Carnivorous plants: phylogeny and structural evolution. Science **257:**1491–1495.

Albert, V. A., M. W. Chase, and B. D. Mishler. 1993. Character-state weighting for cladistic analysis of protein-coding DNA sequences. Annals of the Missouri Botanical Garden **80:**752–766.

Amábile-Cuevas, C. F., and M. E. Chicurel. 1993. Horizontal gene transfer: gene flow from parent to child is the basis of heredity. American Scientist **81:**332–341.

Anderson, E. 1949. Introgressive Hybridization. Wiley, New York.

Araya, A., D. Begu, and S. Litvak. 1994. RNA editing in plants. Physiologia Plantarum **91:**543–550.

Arnheim, N. D. 1983. Concerted evolution of multigene families. **In** Evolution of Genes and Proteins, eds. M. Nei. and R. K. Koehn, pp. 38–61. Sinauer Associates, Sunderland, Massachusetts.

Arnold, M. L. 1992. Natural hybridization as an evolutionary process. Annual Review of Ecology and Systematics **23:**237–261.

Baldwin, B. G. 1997. Adaptive radiation of the Hawaiian silversword alliance: congruence and conflict of phylogenetic evidence from molecular and non-molecular investigations. **In** Molecular Evolution and Adaptive Radiation, eds. T. J. Givnish and K. J. Systma, pp. 103–128. Cambridge University Press, Cambridge.

Baldwin, B. G., C. S. Campbell, J. M. Porter, M. J. Sanderson, M. F. Wojciechowski, and M. J. Donoghue. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Annals of the Missouri Botanical Garden **82:**247–277.

Barker, N. P., H. P. Linder, and E. H. Harley. 1995. Polyphyly of Arundinoideae (Poaceae): evidence from *rbcL* sequence data. Systematic Botany **20:**423–435.

Barrett, M., M. J. Donoghue, and E. Sober. 1991. Against consensus. Systematic Zoology **40:**486–493.

Baum, D. A., and K. L. Shaw. 1995. Genealogical perspectives on the species problem. **In** Experimental and Molecular Approaches to Plant Biosystematics, Monographs in Systematic Biology No. 53, eds. P. C. Hoch and A. G. Stevenson, pp. 289–303. Missouri Botanical Garden, St. Louis.

Birky, C. W., T. Maruyama, and P. Fuerst. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts and some results. Genetics **103:**513–527.

Bonierbale, M. W., R. L. Plaisted, and S. D. Tanksley. 1988. RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics **120:**1095–1103.

Bousquet, J., S. H. Strauss, A. H. Doerksen, and R. A. Price. 1992. Extensive variation in evolutionary rate of *rbcL* gene sequences among seed plants. Proceedings of the National Academy of Sciences U.S.A. **89:**7844–7848.

Bowe, L. M., and C. W. dePamphilis. 1996. Effects of RNA editing and gene processing on phylogenetic reconstruction. Molecular Biology and Evolution **13:**1159–1166.

Britten, R. J. 1986. Rates of DNA sequence evolution differ between taxonomic groups. Science **231:**1393–1398.

Brochmann, C., T. Nilsson, and T. M. Gabrielsen. 1996. A classic example of postglacial allopolyploid speciation re-examined using RAPD markers and nucleotide sequences: *Saxifraga osloensis* (Saxifragaceae). Symbolae Botanicae Upsaliensis **31:**75–89.

Brower, A. V. Z., R. DeSalle, and A. Vogler. 1996. Gene trees, species trees, and systematics: a cladistic perspective. Annual Review of Ecology and Systematics **27:**423–450.

Brubaker, C. L., and J. F. Wendel. 1994. Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*: Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). American Journal of Botany **81:**1309–1326.

Brubaker, C. L., J. A. Koontz, and J. F. Wendel. 1993. Bidirectional cytoplasmic and nuclear introgression in the New World cottons, *Gossypium barbadense* and *G. hirsutum*. American Journal of Botany **80:**1203–1208.

Brubaker, C. L., A. H. Paterson, and J. F. Wendel. 1997. Chromosome structural evolution in allotetraploid cotton and its diploid progenitors. Genetics, in press.

Buckler, E. S., and T. P. Holtsford. 1996a. *Zea* systematics: ribosomal ITS evidence. Molecular Biology and Evolution **13:**612–622.

Buckler, E. S., and T. P. Holtsford. 1996b. *Zea* ribosomal repeat evolution and substitution patterns. Molecular Biology and Evolution **13:**623–632.

Bull, J. J., J. P. Huelsenbeck, C. W. Cunningham, D. L. Swofford, and P. J. Waddell. 1993. Partitioning and combining data in phylogenetic analysis. Systematic Biology **42:**384–397.

Castillo, R. O., and D. M. Spooner. 1997. Phylogenetic relationships of wild potatoes, *Solanum* Series *Conicibaccata* (Sect. *Petota*). Systematic Botany **22:**45–83.

Chesser, R. K., and R. J. Baker. 1996. Effective sizes and dynamics of uniparentally and diparentally inherited genes. Genetics **144:**1225–1235.

Clark, A. G., and T.-H. Kao. 1991. Excess nonsynonymous substitution at shared polymorphic sites among self-incompatibility alleles of Solanaceae. Proceedings of the National Academy of Sciences U.S.A. **88:**9823–9827.

Clark, A. G., and T. S. Whittam. 1992. Sequencing errors and molecular evolutionary analysis. Molecular Biology and Evolution **9:**744–752.

Clark, J. B., W. P. Maddison, and M. G. Kidwell. 1994. Phylogenetic analysis supports horizontal transfer of *P* transposable elements. Molecular Biology and Evolution **11:**40–50.

Clark, L. G., W. Zhang, and J. F. Wendel. 1995. A phylogeny of the grass family based on *ndhF* sequence data. Systematic Botany **20:**436–460.

Collins, T. M., F. Kraus, and G. Estabrook. 1994a. Compositional effects and weighting of nucleotide sequences for phylogenetic analysis. Systematic Biology **43:** 449–459.

Collins, T. M., P. H. Wimberger, and G. J. P. Naylor. 1994b. Compositional bias, character-state bias and character-state reconstruction using parsimony. Systematic Biology **43:**482–496.

Cronn, R. C., X. Zhao, A. H. Paterson, and J. F. Wendel. 1996. Polymorphism and concerted evolution in a multigene family: 5S ribosomal DNA in diploid and allopolyploid cottons. Journal of Molecular Evolution **42:**685–705.

Cummings, M. P. 1994. Transmission patterns of eukaryotic transposable elements: arguments for and against horizontal transfer. Trends in Ecology and Evolution **9:**141–145.

Cummings, M. P., S. P. Otto, and J. Wakeley. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. Molecular Biology and Evolution **12:**814–822.

Danna, K. J., R. Workman, V. Coryell, and P. Keim. 1996. 5S rRNA genes in tribe Phaseoleae: array size, number, and dynamics. Genome **39:**445–455.

Davis, J. I., and K. C. Nixon. 1992. Populations, genetic variation, and the delimitation of phylogenetic species. Systematic Biology **41:**421–435.

Davis, J. I., and R. J. Soreng. 1993. Phylogenetic structure of the grass family (Poaceae) as inferred from chloroplast DNA restriction site variation. American Journal of Botany **80:**1444–1454.

Delwiche, C. F., and J. D. Palmer. 1996. Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. Molecular Biology and Evolution **13:**873–882.

de Queiroz, A., M. J. Donoghue, and J. Kim. 1995. Separate versus combined analysis of phylogenetic evidence.

Annual Review of Ecology and Systematics **26:** 657–681.

Dixon, M. T., and D. M. Hillis. 1993. Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. Molecular Biology and Evolution **10:**256–267.

Doebley, J. 1989. Molecular evidence for a missing wild relative of maize and the introgression of its chloroplast genome into *Zea perennis*. Evolution **43:**1555–1559.

Doebley, J. 1993. Genetics, development and plant evolution. Current Opinion in Genetics and Development **3:**865–872.

Domon, C., and A. Steinmetz. 1994. Exon shuffling in anther-specific genes from sunflower. Molecular and General Genetics **244:**312–317.

Donoghue, M. J., and M. J. Sanderson. 1992. The suitability of molecular and morphological evidence in reconstructing plant phylogeny. **In** Molecular Systematics of Plants, eds. P. S. Soltis, D. E. Soltis, and J. J. Doyle, pp. 340–368. Chapman & Hall, New York.

Doolittle, R. F. 1995. The multiplicity of domains in proteins. Annual Review of Biochemistry **64:**287–314.

Downie, S. R., and J. D. Palmer. 1994. A chloroplast DNA phylogeny of the Caryophyllates based on structural and inverted repeat restriction site variation. Systematic Botany **19:**236–252.

Downie, S. R., R. G. Olmstead, G. Zurawski, D. E. Soltis, P. S. Soltis, J. C. Watson, and J. D. Palmer. 1991. Six independent losses of the chloroplast DNA *rpl2* intron in dicotyledons: molecular and phylogenetic implications. Evolution **45:**1245–1259.

Downie, S. R., D. S. Katz-Downie, K. H. Wolfe, P. J. Calie, and J. D. Palmer. 1994. Structure and evolution of the largest chloroplast gene (ORF2280): internal plasticity and multiple gene losses during angiosperm evolution. Current Genetics **25:**367–378.

Doyle, J. J. 1991. Evolution of higher-plant glutamine synthetase genes: tissue specificity as a criterion for predicting orthology. Molecular Biology and Evolution **8:**366–377.

Doyle, J. J. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. Systematic Botany **17:**144–163.

Doyle, J. J. 1994a. Evolution of a plant homeotic multigene family: toward connecting molecular systematics and molecular developmental genetics. Systematic Biology **43:**307–328.

Doyle, J. J. 1994b. Phylogeny of the legume family: an approach to understanding the origins of nodulation. Annual Review of Ecology and Systematics **25:**325–349.

Doyle, J. J. 1995. The irrelevance of allele tree topologies for species delimitation and a non-topological alternative. Systematic Botany **20:**574–588.

Doyle, J. J. 1996. Homoplasy connections and disconnections: genes and species, molecules and morphology. **In** Homoplasy and the Evolutionary Process, eds. M. J. Sanderson and L. Hufford, pp. 37–66. Academic Press, New York.

Doyle, J. J., M. Lavin, and A. Bruneau. 1992. Contributions of molecular data to Papilionoid legume systematics. **In** Molecular Systematics of Plants, eds. P. S. Soltis, D. E. Soltis and J. J. Doyle, pp. 223–251. Chapman & Hall, New York.

Doyle, J. J., J. L. Doyle, and J. D. Palmer. 1995. Multiple independent losses of two genes and one intron from legume chloroplast genomes. Systematic Botany **20:**272–294.

Dubcovsky, J., and J. Dvorák. 1995. Ribosomal RNA multigene loci: nomads of the Triticeae genomes. Genetics **140:**1367–1377.

Duvall, M. R., and B. R. Morton. 1996. Molecular phylogenetics of Poaceae: an expanded analysis of *rbcL* sequence data. Molecular Phylogenetics and Evolution **5:**352–358.

Eernisse, D. J., and A. G. Kluge. 1993. Taxonomic congruence versus total evidence and amniote phylogeny inferred from fossils, molecules and morphology. Molecular Biology and Evolution **10:**1170–1195.

Elder J. F., and B. J. Turner. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. Quarterly Review of Biology **70:**297–320.

Ewens, W. 1990. Population genetics theory—the past and the future. **In** Mathematical and Statistical Developments of Evolutionary Theory, ed. S. Lessard, pp. 177–227. Kluwer, New York.

Faith, D. P. 1991. Cladistic permutation tests for monophyly and nonmonophyly. Systematic Zoology **40:**366–375.

Farris, J. S., M. Källersjö, A. G. Kluge, and C. Bult. 1995. Testing significance of incongruence. Cladistics **10:**315–319.

Fehrer, J. 1996. Conflicting character distribution within different data sets on cardueline finches: artifact or history? Molecular Biology and Evolution **13:**7–20.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Systematic Zoology **27:**401–410.

Fitch, D. H. A., and M. Goodman. 1991. Phylogenetic scanning: a computer-assisted algorithm for mapping gene conversions and other recombinational events. Computer Applications in the Biosciences **7:**207–216.

Funk, V. A. 1985. Phylogenetic patterns and hybridization. Annals of the Missouri Botanical Garden **72:**681–715.

Gantt, J. S., S. L. Baldauf, P. J. Calie, N. F. Weeden, and J. D. Palmer. 1991. Transfer of *rpl22* to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. EMBO Journal **10:**3073–3078.

Gatesy, J., C. Hayashi, R. Desalle, and E. Vrba. 1994. Rate limits for mispairing and compensatory change: the mitochondrial ribosomal DNA of antelopes. Evolution **48:**188–196.

Galtier, N., and M. Gouy. 1995. Inferring phylogenies from DNA sequences of unequal base compositions. Proceedings of the National Academy of Sciences U.S.A. **92:**11317–11321.

Gaur, L. K., Hughes, E. R. Heise, and J. Gutknecht. 1992. Maintenance of *DQB1* polymorphisms in primates. Molecular Biology and Evolution **9:**599–609.

Gaut, B. S., and M. T. Clegg. 1993a. Molecular evolution of the *Adh1* locus in the genus *Zea*. Proceedings of the National Academy of Sciences U.S.A. **90:**5095–5099.

Gaut, B. S., and M. T. Clegg. 1993b. Nucleotide polymorphism in the *Adh1* locus of pearl millet (*Pennisetum glaucum*) (Poaceae). Genetics **135:**1091–1097.

Gaut, B. S., S. V. Muse, W. D. Clark, and M. T. Clegg. 1992. Relative rates of nucleotide substitution at the *rbcL* locus of monocotyledonous plants. Journal of Molecular Evolution **35:**292–303.

Gaut, B. S., S. V. Muse, and M. T. Clegg. 1993. Relative rates of nucleotide substitution in the chloroplast genome. Molecular Phylogenetics and Evolution **2:**89–96.

Gaut, B. S., B. R. Morton, B. C. McCaig, and M. T. Clegg. 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proceedings of the National Academy of Sciences U.S.A. **93:**10274–10279.

Gaut, B. S., L. G. Clark, J. F. Wendel, and S. V. Muse. 1997. Comparisons of the molecular evolutionary process at *rbcL* and *ndhF* in the grass family (Poaceae). Molecular Biology and Evolution **14:**769–777.

Gielly, L., and P. Taberlet. 1994. The use of chloroplast DNA to resolve plant phylogenies: non-coding versus *rbcL* sequences. Molecular Biology and Evolution **11:**769–777.

Goldman, A. S. H., and M. Lichten. 1996. The efficiency of meiotic recombination between dispersed sequences in *Saccharomyces cerevisiae* depends upon their chromosomal location. Genetics **144:**43–55.

Golenberg, E. M., M. T. Clegg, M. L. Durbin, J. F. Doebley, and D. P. Ma. 1993. Evolution of a non-coding region of the chloroplast genome. Molecular Phylogenetics and Evolution **2:**52–64.

Goloubinoff, P., S. Pääbo, and A. C. Wilson. 1993. Evolution of maize inferred from sequence diversity of an *Adh2* gene segment from archaeological specimens. Proceedings of the National Academy of Sciences U.S.A. **90:**1997–2001.

Gottlieb, L. D. 1984. Genetics and morphological evolution in plants. American Naturalist **123:**681–709.

Gottlieb, L. D., and V. S. Ford. 1996. Phylogenetic relationships among the sections of *Clarkia* (Onagraceae) inferred from the nucleotide sequence of *PgiC*. Systematic Botany **21:**45–62.

Graham, S. W., J. R. Kohn, B. R. Morton, J. E. Eckenwalder, and S. C. H. Barrett. 1997. Phylogenetic congruence and discordance among one morphological and three molecular data-sets from Pontederiaceae. Systematic Biology, in press.

Gray, M. W., and P. S. Covello. 1993. RNA editing in plant mitochondria and chloroplasts. FASEB Journal **7:**64–71.

Guigó, R., I. Muchnik, and T. F. Smith. 1996. Reconstruction of ancient molecular phylogeny. Molecular Phylogenetics and Evolution **6:**189–213.

Hamrick, J. L., and M. J. W. Godt. 1989. Allozyme diversity in plant species. **In** Plant Population Genetics, Breeding and Genetic Resources, eds. A. H. D. Brown, M. T. Clegg, A. L. Kahler, and B. S. Weir, pp. 43–63. Sinauer Associates, Sunderland, Massachusetts.

Hanson, M. A., B. S. Gaut, A. O. Stec, S. I. Fuerstenberg, M. M. Goodman, E. H. Coe, and J. F. Doebley. 1996. Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. Genetics **143:**1395–1407.

Hein, J. 1993. A heuristic method to reconstruct the history of sequences subject to recombination. Journal of Molecular Evolution **36:**396–405.

Heiser, C. B. 1973. Introgression re-examined. Botanical Review **39:**347–366.

Hershkovitz, M. A., and E. A. Zimmer. 1996. Conservation patterns in angiosperm rDNA ITS2 sequences. Nucleic Acids Research **24:**2857–2867.

Hey, J., and J. Wakeley. 1997. A coalescent estimator of the population recombination. Genetics **145:**833–846.

Hibbett, D. S. 1996. Phylogenetic evidence for horizontal transmission of group I introns in the nuclear ribosomal DNA of mushroom-forming fungi. Molecular Biology and Evolution **13:**903–917.

Hiesel, R., B. Combettes, and A. Brennicke. 1994. Evidence for RNA editing in mitochondria of all major groups of land plants except the Bryophyta. Proceedings of the National Academy of Sciences U.S.A. **91:**629–633.

Hillis, D. M. 1987. Molecular versus morphological approaches to systematics. Annual Review of Ecology and Systematics **18:**23–42.

Hillis, D. M. 1996. Inferring complex phylogenies. Nature **383:**130–131.

Hillis, D. M., and M. T. Dixon. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Quarterly Review of Biology **66:**411–453.

Hirose, T., T. Wakasugi, M. Sugiura, and H. Kössel. 1994. RNA editing of tobacco *petB* mRNAs occurs both in chloroplasts and non-photosynthetic proplastids. Plant Molecular Biology **26:**509–513.

Hoelzer, G. A. 1997. Inferring phylogenies from mtDNA variation: mitochondrial gene-trees versus nuclear gene-trees revisited. Evolution **51:**622–626.

Hoot S. B., and J. D. Palmer. 1994. Structural rearrangements, including parallel inversions, within the chloroplast genome of *Anemone* and related genera. Journal of Molecular Evolution **38:**274–281.

Hoot, S. B., A. Culham, and P. R. Crane. 1995. The utility of *atpB* gene sequences in resolving phylogenetic relationships: comparison with *rbcL* and 18S ribosomal DNA sequences in the Lardizabalaceae. Annals of the Missouri Botanical Garden **82:**194–207.

Houck, M. A., J. B. Clark, K. R. Peterson, and M. G. Kidwell. 1991. Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. Science **253:**1125–1129.

Hudson, R. R. 1990. Gene genealogies and the coalescent process. Oxford Surveys in Evolutionary Biology **7:**1–41.

Hudson, R. R., and N. L. Kaplan. 1988. The coalescent process in models with selection and recombination. Genetics **120:**831–840.

Huelsenbeck, J. P., J. J. Bull, and C. W. Cunningham. 1996. Combining data in phylogenetic analysis. Trends in Ecology and Evolution **11:**152–158.

Ikeo, K., K. Takahashi, and T. Gojobori. 1995. Different evolutionary histories of kringle and protease domains in serine proteases: a typical example of domain evolution. Journal of Molecular Evolution **40:**331–336.

Innan, H., F. Tajima, R. Terauchi, and N. T. Miyashita. 1996. Intragenic recombination in the *Adh* locus of the wild plant *Arabidopsis thaliana.* Genetics **143:**1761–1770.

Ioerger, T. R., A. G. Clark, and T.-H. Kao. 1990. Polymorphism at the self-incompatibility locus in Solanaceae predates specification. Proceedings of the National Academy of Sciences U.S.A. **87:**9732–9735.

Jakobsen, I. B., and S. Easteal. 1996. A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences. Computer Applications in the Biosciences **12:**291–295.

Johnson, L. A., and D. E. Soltis. 1994. *matK* DNA sequences and phylogenetic reconstruction in Saxifragaceae *s. str.* Systematic Botany **19:**143–156.

Johnson, L. A., and D. E. Soltis. 1995. Phylogenetic inference in Saxifragaceae *sensu stricto* and *Gilia* (Polemoniaceae) using *matK* sequences. Annals of the Missouri Botanical Garden **82:**149–175.

Kadereit, J. W. 1994. Molecules and morphology, phylogenetics and genetics. Botanica Acta **107:**369–373.

Kelchner, S. A., and J. F. Wendel. 1996. Hairpins create minute inversions in non-coding regions of chloroplast DNA. Current Genetics **30:**259–262.

Kellogg, E. A., R. Appels, and R. Mason-Gamer. 1996. When genes tell different stories: incongruent gene trees for diploid genera of Triticeae (Gramineae). Systematic Botany **21:**321–347.

Kidwell, M. G. 1993. Lateral transfer in natural populations of eukaryotes. Annual Review of Genetics **27:**235–256.

Kim, J. 1996. General inconsistency conditions for maximum parsimony: effects of branch lengths and increasing numbers of taxa. Systematic Biology **45:**363–374.

Kim, K.-J., and R. K. Jansen. 1994. Comparisons of phylogenetic hypotheses among different data sets in dwarf dandelions (*Krigia,* Asteraceae): additional information from internal transcribed spacer sequences of nuclear ribosomal DNA. Plant Systematics and Evolution **190:**157–185.

Kim, K.-J., and R. K. Jansen. 1995. *ndhF* sequence evolution and the major clades in the sunflower family. Proceedings of the National Academy of Sciences U.S.A. **92:**10379–10383.

Kim, K.-J., R. K. Jansen, R. S. Wallace, H. J. Michaels, and J. D. Palmer. 1992. Phylogenetic implications of *rbcL* sequence variation in the Asteraceae. Annals of the Missouri Botanical Garden **79:**428–445.

Kluge, A. G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidea, Serpentes). Systematic Zoology **38:**7–25.

Knight, A., and D. P. Mindell. 1993. Substitution bias, weighting of DNA sequence evolution and the phylogenetic position of Fea's viper. Systematic Biology **42:**18–31.

Kolukisaoglu, H. Ü., S. Marx, C. Wiegmann, S. Hanelt, and H. A. W. Schneider-Poetsch. 1995. Divergence of the phytochrome gene family predates angiosperm evolution and suggests that *Selaginella* and *Equisetum* arose prior to *Psilotum.* Journal of Molecular Evolution **41:**329–337.

Kowalski, S. P., T.-H. Lan, K. A. Feldmann, and A. H. Paterson. 1994. Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. Genetics **138:** 499–510.

Kuhner, M. K., and J. Felsenstein. 1994. A simulation of phylogeny algorithms under equal and unequal evolutionary rates. Molecular Biology and Evolution **11:**459–468.

Kumar, A. 1996. The adventures of the Tyl-*copia* group of retrotransposons in plants. Trends in Genetics **12:** 41–43.

Lan, R., and P. R. Reeves. 1996. Gene transfer is a major factor in bacterial evolution. Molecular Biology and Evolution **13:**47–55.

Lara, M. C., J. L. Patton, and M. N. F. da Silva. 1996. The simultaneous diversification of South American echimyid rodents (Hystricognathi) based on complete cytochrome b sequences. Molecular Phylogenetics and Evolution **5:**403–413.

Levy, F., J. Antonovics, J. E. Boynton, and N. W. Gillham. 1996. A population genetic analysis of chloroplast DNA in *Phacelia.* Heredity **76:**143–155.

Li, W.-H., M. Tanimura, and P. M. Sharp. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. Journal of Molecular Evolution **25:**330–342.

Liston, A., W. A. Robinson, J. M. Oliphant, and E. R. Alvarez-Buylla. 1996. Length variation in the nuclear ribosomal DNA internal transcribed spacer region of nonflowering seed plants. Systematic Botany **21:**109–120.

Lockhart, P. J., M. A. Steel, M. D. Hendy, and D. Penny. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Molecular Biology and Evolution **11:**605–612.

Lutzoni, F., and R. Vilgalys. 1994. Integration of morphological and molecular data sets in estimating fungal phylogenies. Canadian Journal of Botany **73:** S649–S659.

Lyons-Weiler, J., G. A. Hoelzer, and R. J. Tausch. 1996. Relative apparent synapomorphy analysis (RASA) I: the statistical measurement of phylogenetic signal. Molecular Biology and Evolution **13:**749–757.

Lyons-Weiler, J., G. A. Hoelzer, and R. J. Tausch. 1997. Relative apparent synapomorphy analysis (RASA) II: optimal outgroup analysis. Systematic Biology, in press.

Maddison, W. P. 1995. Phylogenetic histories within and among species. In Experimental and Molecular Approaches to Plant Biosystematics, Monographs in Systematic Biology No. 53, eds. P. C. Hoch and A. G. Stevenson, pp. 273–287. Missouri Botanical Garden, St. Louis.

Maddison, W. P. 1996. Molecular approaches and the growth of phylogenetic biology. **In** Molecular Zoology: Approaches, Strategies and Protocols, eds. J. D. Ferraris and S. R. Palumbi, pp. 47–63. Wiley, New York.

Malek, O., K. Lättig, R. Hiesel, A. Brennicke, and V. Knoop. 1996. RNA editing in bryophytes and a molecular phylogeny of land plants. EMBO Journal **15:**1403–1411.

Martin, A. P. 1995. Mitochondrial DNA sequence evolution in sharks: rates, patterns and phylogenetic inferences. Molecular Biology and Evolution **12:**1114–1123.

Mason-Gamer, R. J., and E. A. Kellogg. 1996. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). Systematic Biology **45:**524–545.

Mason-Gamer, R. J., K. E. Holsinger, and R. K. Jansen. 1995. Chloroplast DNA haplotype variation within and among populations of *Coreopsis grandiflora* (Asteraceae). Molecular Biology and Evolution **12:**371–381.

Masterson, J. 1994. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science **264:**421–424.

Mathews, S., and R. A. Sharrock. 1996. The phytochrome gene family in grasses (Poaceae): a phylogeny and evidence that grasses have a subset of the loci found in dicot angiosperms. Molecular Biology and Evolution **13:**1141–1150.

Mathews, S., M. Lavin, and R. A. Sharrock. 1995. Evolution of the phytochrome gene family and its utility for phylogenetic analyses of angiosperms. Annals of the Missouri Botanical Garden **82:**296–321.

McDade, L. A. 1990. Hybrids and phylogenetic systematics I. Patterns of character expression in hybrids and their implications for cladistic analysis. Evolution **44:** 1685–1700.

McDade, L. A. 1992. Hybrids and phylogenetic systematics II. The impact of hybrids on cladistic analysis. Evolution **46:**1329–1346.

McDade, L. A. 1995. Hybridization and phylogenetics. **In** Experimental and Molecular Approaches to Plant Biosystematics, Monographs in Systematic Botany No. 53, eds. P. C. Hoch and A. G. Stevenson, pp. 305–331. Missouri Botanical Garden, St. Louis.

McGrath, J. M., M. M. Jansco, and E. Pichersky. 1993. Duplicate sequences with a similarity to expressed genes in the genome of *Arabidopsis thaliana*. Theoretical and Applied Genetics **86:**880–888.

Meagher, R. B., S. Berry-Lowe, and K. Rice. 1989. Molecular evolution of the small subunit of ribulose bisphosphate carboxylase: nucleotide substitution and gene conversion. Genetics **123:**845–863.

Mickevich, M. F., and J. S. Farris. 1981. The implications of congruence in *Menidia*. Systematic Zoology **30:** 351–370.

Millar, A. A., and E. S. Dennis. 1996. The alcohol dehydrogenase genes of cotton. Plant Molecular Biology **31:**897–904.

Miyamoto, M. M., 1985. Consensus cladograms and general classifications. Cladistics **1:**186–189.

Miyamoto, M. M., and W. M. Fitch. 1995. Testing species phylogenies and phylogenetic methods with congruence. Systematic Biology **44:**64–76.

Moens, L., J. Vanfleteren, Y. Van de Peer, K. Peeters, O. Kapp, J. Czeluzniak, M. Goodman, M. Blaxter, and S. Vinogradov. 1996. Globins in nonvertebrate species: dispersal by horizontal gene transfer and evolution of the structure-function relationships. Molecular Biology and Evolution **13:**324–333.

Moniz de Sá, M., and G. Drouin. 1996. Phylogeny and substitution rates of angiosperm actin genes. Molecular Biology and Evolution **13:**1198–1212.

Moore, W. S. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution **49:**718–726.

Morgan, D. R., and D. E. Soltis. 1993. Phylogenetic relationships among members of the Saxifragaceae s.l. based on *rbcL* sequence data. Annals of the Missouri Botanical Garden **80:**631–660.

Morton, B. R., and M. T. Clegg. 1993. A chloroplast DNA mutational hotspot and gene conversion in a non-coding region near *rbcL* in the grass family (Poaceae). Current Genetics **24:**357–365.

Neigel, J. E., and J. C. Avise. 1986. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. **In** Evolutionary Processes and Theory, eds. E. Nevo and S. Karlin, pp. 515–534. Academic Press, New York.

Nickrent, D. L., and E. M. Starr. 1994. High rates of nucleotide substitution in nuclear small-subunit (18S) rDNA from holoparasitic flowering plants. Journal of Molecular Evolution **39:**62–70.

Nugent, J. M., and J. D. Palmer. 1991. RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. Cell **66:**473–481.

Ohta, T. 1993. An examination of the generation-time effect on molecular evolution. Proceedings of the National Academy of Sciences U.S.A. **90:**10676–10680.

Olmstead, R. G., and J. D. Palmer. 1994. Chloroplast DNA systematics: a review of methods and data analysis. American Journal of Botany **81:**1205–1224.

Olmstead, R. G., and J. A. Sweere. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. Systematic Biology **43:**467–481.

Palmer, J. D., C. R. Shields, D. B. Cohen, and T. J. Orten. 1983. Chloroplast DNA evolution and the origin of amphidiploid *Brassica* species. Theoretical and Applied Genetics **65:**181–189.

Palmer, J. D., R. A. Jorgenson, and W. F. Thompson. 1985. Chloroplast DNA variation and evolution in *Pisum*: patterns of change and phylogenetic analysis. Genetics **109:**195–213.

Pamilo, P., and M. Nei. 1988. Relationships between gene trees and species trees. Molecular Biology and Evolution **5:**568–583.

Patterson, C. 1988. Homology in classical and molecular biology. Molecular Biology and Evolution **5:**603–625.

Patterson, C., D. M. Williams, and C. J. Humphries. 1993. Congruence between molecular and morphological phylogenies. Annual Review of Ecology and Systematics **24:**153–188.

Patthy, L. 1991. Modular exchange principles in proteins. Current Opinion in Structural Biology **1:**351–361.

Pereira, M. G., M. Lee, P. Bramel-Cox, W. Woodman, J. Doebley, and R. Whitkus. 1994. Construction of an

RFLP map in sorghum and comparative mapping in maize. Genome **37**:236–243.

Richman, A. D., M. K. Uyenoyama, and J. R. Kohn. 1996. Allelic diversity and gene genealogy at the self-incompatibility locus in the Solanaceae. Science **273**: 1212–1216.

Rieseberg, L. H. 1991. Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. American Journal of Botany **78**:1218–1237.

Rieseberg, L. H. 1995. The role of hybridization in evolution: old wine in new skins. American Journal of Botany **82**:944–953.

Rieseberg, L. H., and D. E. Soltis. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evolutionary Trends in Plants **5**:65–84.

Rieseberg, L. H., and J. F. Wendel. 1993. Introgression and its consequences. **In** Hybrid Zones and the Evolutionary Process, ed. R. Harrison, pp. 70–109. Oxford University Press, New York.

Rieseberg, L. H., S. Beckstrom-Sternberg, and K. Doan. 1990. *Helianthus annuus* ssp. *texanus* has chloroplast DNA and nuclear ribosomal RNA genes of *Helianthus debiolis ssp. cucumerifolius*. Proceedings of the National Academy of Sciences U.S.A. **87**:593–597.

Rieseberg, L. H., H. C. Choi, and D. Ham. 1991. Differential cytoplasmic versus nuclear introgression in *Helianthus*. Journal of Heredity **82**:489–493.

Rieseberg, L. H., J. Whitton, and C. R. Linder. 1996. Molecular marker incongruence in plant hybrid zones and phylogenetic trees. Acta Botanica Neerlandica **45**:243–262.

Rivers, B. A., R. Bernatzky, S. J. Robinson, and W. Jahnen-Dechent. 1993. Molecular diversity at the self-incompatibility locus is a salient feature in natural populations of wild tomato (*Lycopersicon peruvianum*). Molecular and General Genetics **238**:419–427.

Robertson, H. M., and D. J. Lampe. 1995. Recent horizontal transfer of a *mariner* transposable element among and between Diptera and Neuroptera. Molecular Biology and Evolution **12**:850–862.

Rodman, J., R. A. Price, K. Karol, E. Conti, K. J. Systma, and J. D. Palmer. 1993. Nucleotide sequences of the *rbcL* gene indicate monophyly of mustard oil plants. Annals of the Missouri Botanical Garden **80**:686–699.

Rodrigo, A. G., M. Kelley-Borges, P. R. Bergquist, and P. L. Bergquist. 1993. A randomization test of the null hypothesis that two cladograms are sample estimates of a parametric phylogenetic tree. New Zealand Journal of Botany **31**:257–268.

Roth, V. L. 1991. Homology and hierarchies: problems solved and unresolved. Journal of Evolutionary Biology **4**:167–194.

Sanderson, M. J., and Doyle, J. J. 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. Systematic Biology **41**:4–17.

Sang, T., D. J. Crawford, and T. F. Stuessy. 1995. Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: implications for biogeography and concerted evolution. Proceedings of the National Academy of Sciences U.S.A. **92**:6813–6817.

Sawyer, S. 1989. Statistical tests for detecting gene conversion. Molecular Biology and Evolution **6**:526–538.

Schena, M., and R. W. Davis. 1994. Structure of homeobox-leucine zipper genes suggests a model for the evolution of gene families. Proceedings of the National Academy of Sciences U.S.A. **91**:8393–8397.

Seberg, O., G. Petersen, and C. Baden. 1996. The phylogeny of *Psathyrostachys* Nevski (Triticeae, Poaceae)—are we able to see the wood for the trees? **In** Proceedings of the 2nd International Triticeae Symposium, eds. R. R.-C. Wang, K. B. Jensen, and C. Jaussi, pp. 247–253. Utah State University, Logan.

Seelanan, T., A. Schnabel, and J. F. Wendel. 1997. Congruence and consensus in the cotton tribe. Systematic Botany **22**:259–290.

Sites, J. W., S. K. Davis, T. Guerra, J. B. Iverson, and H. L. Snell. 1996. Character congruence and phylogenetic signal in molecular and morphological data sets: a case study in the living iguanas (Squamata, Iguanidae). Molecular Biology and Evolution **13**:1087–1105.

Slatkin, M. 1991. Inbreeding coefficients and coalescence times. Genetical Research of Cambridge **58**:167–175.

Smith, R. L., and K. J. Sytsma. 1990. Evolution of *Populus nigra* (sect. *Aegiros*): introgressive hybridization and the chloroplast contribution of *Populus alba* (sect. *Populus*). American Journal of Botany **77**:1176–1187.

Soltis, D. E., and R. K. Kuzoff. 1995. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). Evolution **49**:727–742.

Soltis, D. E., and P. S. Soltis (eds.). 1989. Isozymes in Plant Biology. Dioscorides Press, Portland Oregon.

Soltis, D. R. Morgan, A. Grable, P. S. Soltis, and R. Kuzoff. 1993. Molecular systematics of Saxifragaceae *sensu stricto*. American Journal of Botany **80**:1056–1081.

Soltis, D. E., L. A. Johnson, and C. Looney. 1996. Discordance between ITS and chloroplast topologies in the *Boykinia* group (Saxifragaceae). Systematic Botany **21**:169–185.

Soltis, D. E., P. S. Soltis, D. L. Nickrent, L. A. Johnson, W. J. Hahn, S. B. Hoot, J. A. Sweere, R. K. Kuzoff, K. A. Kron, M. W. Chase, S. M. Swensen, E. A. Zimmer, S.-M. Chaw, L. J. Gillespie, W. J. Kress, and K. J. Sytsma. 1997. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. Annals of the Missouri Botanical Garden **84**:1–49.

Soltis, P. S., and D. E. Soltis. 1995. Plant molecular systematics: inferences of phylogeny and evolutionary processes. Evolutionary Biology **28**:139–194.

Springer, M., L. J. Hollar, and A. Burk. 1995. Compensatory substitutions and the evolution of the mitochondrial 12S rRNA gene in mammals. Molecular Biology and Evolution **12**:1138–1150.

Steel, M., P. J. Lockhart, and D. Penny. 1995. A frequency-dependent significance test for parsimony. Molecular Phylogenetics and Evolution **4**:64–71.

Steele, K. P., and R. Vilgalys. 1994. Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene *matK*. Systematic Botany **19**:126–142.

Stone, E. M., and R. J. Schwartz (eds.). 1990. Intervening Sequences in Evolution and Development. Oxford University Press, New York.

Suh, Y., L. B. Thien, H. E. Reeve, and E. A. Zimmer. 1993. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of ribosomal DNA in Winteraceae. American Journal of Botany **80:**1042–1055.

Sullivan, J. 1996. Combining data with different distributions of among-site rate variation. Systematic Biology **45:**375–380.

Sullivan, J., K. E. Holsinger, and C. Simon. 1995. Among-site rate variation and phylogenetic analysis of 12S rRNA in Sigmodontine rodents. Molecular Biology and Evolution **12:**988–1001.

Swofford, D. L. 1991. When are phylogeny estimates from molecular and morphological data incongruent? **In** Phylogenetic Analysis of DNA Sequences, eds. M. M. Miyamoto and J. Cracraft, pp. 295–333. Oxford University Press, New York.

Sytsma, K. J. 1990. DNA and morphology: inference of plant phylogeny. Trends in Ecology and Evolution **5:**104–110.

Syvanen, M. 1994. Horizontal gene transfer: evidence and possible consequences. Annual Review of Genetics **28:**237–261.

Tajima, F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics **135:**599–607.

Takezaki, N., A. Rzhetsky, and M. Nei. 1995. Phylogenetic test of the molecular clock and linearized trees. Molecular Biology and Evolution **12:**823–833.

Talbert, L. E., J. F. Doebley, S. Larson, and V. L. Chandler. 1990. *Tripsacum andersonii* is a natural hybrid involving *Zea* and *Tripsacum:* molecular evidence. American Journal of Botany **77:**722–726.

Templeton, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution **37:**221–244.

Templeton, A. R., and C. F. Sing. 1993. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analysis with cladogram uncertainty and recombination. Genetics **134:**659–669.

VanderWiel, P. S., D. F. Voytas, and J. F. Wendel. 1993. *Copia*-like retrotransposable element evolution in diploid and polyploid cotton (**Gossypium** L.). Journal of Molecular Evolution **36:**429–447.

van Ham, R. C. H. J., H. T'Hart, T. H. M. Mes, and J. M. Sandbrink. 1994. Molecular evolution of non-coding regions of the chloroplast genome in the Crassulaceae and related species. Current Genetics **25:**558–566.

Vawter, L., and W. M. Brown. 1993. Rates and patterns of base change in the small subunit ribosomal RNA gene. Genetics **134:**597–608.

Wakeley, J. 1994. Substitution-rate variation among sites and the estimation of transition bias. Molecular Biology and Evolution **11:**426–442.

Walker, E. L., N. F., Weeden, C. B. Taylor, P. Green, and G. M. Coruzzi. 1995. Molecular evolution of duplicate copies of genes encoding cytosolic glutamine synthetase in *Pisum sativum.* Plant Molecular Biology **29:**1111–1125.

Waters, E. R., and B. A. Schaal. 1996. Biased gene conversion is not occurring among rDNA repeats in the *Brassica* triangle. Genome **39:**150–154.

Wendel, J. F. 1989. New World tetraploid cottons contain Old World cytoplasm. Proceedings of the National Academy of Sciences U.S.A. **86:**4132–4136.

Wendel, J. F., and V. A. Albert. 1992. Phylogenetics of the cotton genus (*Gossypium* L.): character-state weighted parsimony analysis of chloroplast DNA restriction site data and its systematic and biogeographic implications. Systematic Botany **17:**115–143.

Wendel, J. F., J. McD. Stewart, and J. H. Rettig., 1991. Molecular evidence for homoploid reticulate evolution in Australian species of *Gossypium.* Evolution **45:** 694–711.

Wendel, J. F., A. Schnabel, and T. Seelanan. 1995a. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proceedings of the National Academy of Sciences U.S.A. **92:**280–284.

Wendel, J. F., A. Schnabel, and T. Seelanan. 1995b. An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. Molecular Phylogenetics and Evolution **4:**298–313.

Wessler, S. R., T. E. Bureau, and S. E. White. 1995. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Current Opinion in Genetics and Development **5:**814–821.

Wheeler, W. C. 1992. Extinction, sampling, and molecular phylogenetics. **In** Extinction and Phylogeny, eds. M. J. Novacek and Q. D. Wheeler, pp. 205–215. Columbia University Press, New York.

Whitkus, R., J. Doebley, and M. Lee. 1992. Comparative genome mapping of sorghum and maize. Genetics **132:**1119–1130.

Wu, C.-I., and W-H. Li. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proceedings of the National Academy of Sciences U.S.A. **82:**1741–1745.

Yang, Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends in Ecology and Evolution **11:**367–372.

Yoshinaga, K., H. Iinuma, T. Masuzawa, and K. Ueda. 1996. Extensive RNA editing of U to C in addition to C to U substitution in the *rbcL* transcripts of hornwort chloroplasts and the origin of RNA editing in green plants. Nucleic Acids Research **24:**1008–1014.

Zhang, W., J. F. Wendel, and L. G. Clark. 1997. Bamboozled again! Inadvertent isolation of fungal rDNA sequences from bamboos (Poaceae: Bambusoideae). Molecular Phylogenetics and Evolution **8:** 205–217.