The Role of Phylogenetics in Comparative Genetics[1]

Douglas E. Soltis* and Pamela S. Soltis

Department of Botany and the Genetics Institute, University of Florida, Gainesville, Florida 32611; and Florida Museum of Natural History and the Genetics Institute, University of Florida, Gainesville, Florida 32611

WHY PHYLOGENY MATTERS

Many biologists agree that a phylogenetic tree of relationships should be the central underpinning of research in many areas of biology. Comparisons of plant species or gene sequences in a phylogenetic context can provide the most meaningful insights into biology. This important realization is now apparent to researchers in diverse fields, including ecology, molecular biology, and physiology (see recent papers in *Plant Physiology*, e.g. Hall et al., 2002a; Doyle et al., 2003). Examples of the importance of a phylogenetic framework to diverse areas of plant research abound (for review, see Soltis and Soltis, 2000; Daly et al., 2001). One obvious example is the value of placing model organisms in the appropriate phylogenetic context to obtain a better understanding of both patterns and processes of evolution. The fact that tomato (*Lycopersicon esculentum*) and other species of this small genus actually are embedded within a well-marked subclade *Solanum* (and, hence, are more appropriately referred to as species of *Solanum*; tomato has been renamed as of *Solanum lycopersicon*; e.g. Spooner et al., 1993; Olmstead et al., 1999) is a powerful statement that is important to geneticists, molecular biologists, and plant breeders in that it points to a few close relatives of *S. lycopersicon* (out of a genus of several hundred species) as focal points for comparative genetic/genomic research and for crop improvement. Snapdragon (*Antirrhinum majus*) was historically part of a broadly defined Scrophulariaceae, a family that is now known to be grossly polyphyletic (i.e. not a single clade). Phylogenetic studies indicate that Scrophulariaceae should be broken up into several families (Olmstead et al., 2001), and snapdragon and its closest relatives are part of a clade recognized as the family Plantaginaceae.

A phylogenetic framework has revealed the patterns of evolution of many morphological and chemical characters, including complex pathways such as nitrogen-fixing symbioses, mustard oil production, and chemical defense mechanisms (for review, see Soltis and Soltis, 2000; Daly et al., 2001). However, the importance of phylogeny reconstruction applies not only to the organisms that house genes but also to the evolutionary history of the genes themselves. For example, are the genes under investigation the members of a single well-defined clade, all members of which appear to descend from a recent common ancestor as a direct result of speciation (orthologous genes), or do the sequences represent one or more ancient duplications (paralogous genes; see also Doyle and Gaut, 2000)? Gene families are, of course, the norm in studies of nuclear genes, but investigators are often bewildered by the diversity of genes encountered in a survey of a family of genes from a diverse array of plants. Phylogenetic methodology offers several solutions by permitting inferences of putative orthology among a set of sequences.

Examples of the phylogenetic analysis of gene families abound (e.g. genes encoding: heat shock proteins, Waters and Vierling, 1999; phytochrome, Kolukisaoglu et al., 1995; Mathews and Sharrock, 1997; and actin, McDowell et al., 1996). A noteworthy recent example involves MADS box genes, which encode transcription factors that control diverse developmental processes in plants. Some of the best known examples of MADS box genes include the A, B, and C class floral genes that control the identity of floral organs (for review, see Ma and dePamphilis, 2000). Phylogenetic analyses indicate that a minimum of seven different MADS box gene lineages were already present in the common ancestor of extant seed plants approximately 300 million years ago (mya; Becker et al., 2000). Thus, a diverse tool kit of MADS box genes was available before the origin of the angiosperms.

A phylogenetic perspective also provides the basis for comparative genomics (e.g. Soltis and Soltis, 2000; Walbot, 2000; Daly et al., 2001; Kellogg, 2001; Hall et al., 2002a; Mitchell-Olds and Clauss, 2002; Pryer et al., 2002; Doyle and Luckow, 2003). However, obtaining the appropriate phylogenetic perspective may be difficult: What phylogenetic hypotheses are already available for the group of interest? Are phylogenetic studies underway on a particular group, and is it possible to obtain unpublished trees? Is the phylogenetic underpinning for a lineage of interest sound enough for use in comparative genetic/genomic analyses? Not all phylogenetic trees are of equal quality, and the most fruitful phylogenomic compar-

[1] This work was supported in part by the National Science Foundation (Deep Time Research Coordination Network and the Floral Genome project grants).

* Corresponding author; e-mail dsoltis@botany.ufl.edu; fax 352–846–2154.

http://www.plantphysiol.org/cgi/doi/10.1104/pp.103.022509.

isons will be those based on the strongest phylogenetic inferences.

We cannot address all of the crucial issues relating to the importance of phylogeny in a comprehensive fashion and, therefore, will focus on a few main topics. We provide: (a) phylogenetic summaries and references for major clades of land plants, with an emphasis on angiosperm model systems; (b) a "primer" of phylogenetic methods, including evaluation of parsimony, distance, maximum likelihood (ML), and Bayesian methods, the importance of measures of internal support in phylogenetic inference, and methods of analysis of large data sets; and (c) use of molecular data to estimate divergence times of genes or organisms. A major goal is to foster increased interaction and communication between phylogeneticists and physiologists/molecular geneticists by providing contacts and references for those requiring a phylogenetic backbone for analyses.

SELECTION OF TAXA AND PHYLOGENETIC TREES IN COMPARATIVE STUDIES. A SUMMARY OF LAND PLANT PHYLOGENY

One question that systematists are frequently asked is: Where would I find the most recent phylogenetic tree for group (fill in the blank)? We provide a brief summary of relevant trees below, with a focus on land plants. In addition, selected trees for angiosperms can be found at http:/www.mobot.org/MOBOT/research/APweb//, http://www.flmnh.ufl.edu/deeptime/and http://plantsystematics.org/). Researchers can also consult Tree of Life (http://tolweb.org/tree/phylogeny.html) and TreeBASE (http://www.treebase.org/treebase). Phylogenetic questions can also be posed directly to experts working on various groups of plants; a partial list of phylogenetic consultants is provided in Table I (for a larger list, see also http://www.flmnh.ufl.edu/deeptime/).

Land Plants. Origin and Relationships

Understanding patterns of gene and genome evolution across land plants requires an understanding of the phylogeny of land plants, or embryophytes. Molecular data indicate that the sister group (i.e. the closest relative; two sister groups share a common ancestor not shared with any other group) of land plants is Charales (stoneworts) from the charophycean lineage of green algae (Karol et al., 2001; Fig. 1; see also http://www.flmnh.ufl.edu/deeptime/).

Table I. *Partial list of phylogenetic experts for various clades of land plants.*
For a larger list of experts, see the Deep Time Web site (http://www.flmnh.ufl.edu/deeptime/).

Clade(s)	Contact Person	E-Mail Address
Mosses, liverworts	Jonathan Shaw, Department of Biology, Duke University, Durham, NC 27708	shaw@duke.edu
Ferns	Kathleen Pryer, Department of Biology, Duke University, Durham, NC 27708	pryer@duke.edu
Basal angiosperms	Douglas Soltis, Department of Botany, University of Florida, Gainesville, FL 32611	dsoltis@botany.ufl.edu
	Pamela Soltis, Florida Museum of Natural History, University of Florida, Gainesville, FL 32611	psoltis@flmnh.ufl.edu
Monocots	Walter Judd, Department of Botany, University of Florida, Gainesville, FL 32611	wjudd@botany.ufl.edu
	Mark Chase, Molecular Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, TW9 3DS UK	M.Chase@rbgkew.org.uk
Poaceae	Elizabeth Kellogg, Department of Biology, University of Missouri, St. Louis 8001 Natural Bridge Rd, St. Louis, MO 63121	kellogg@msx.umsl.edu
Rosids	Walter Judd, Department of Botany, University of Florida, Gainesville, FL 32611	wjudd@botany.ufl.edu
	Mark Chase, Molecular Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, TW9 3DS UK	M.Chase@rbgkew.org.uk
	Douglas Soltis, Department of Botany, University of Florida, Gainesville, FL 32611	dsoltis@botany.ufl.edu
Fabaceae	Jeff Doyle, Department of Plant Science, Cornell University, Ithaca, NY 14853	jjd5@postoffice.mail.cornell.edu
	Matt Lavin, Department of Plant Sciences, Montana State University, Bozeman, MT 59717	mlavin@montana.edu
Brassicaceae	Ishan Al-Shehbaz, Missouri Botanical Garden, P.O. Box 299, St. Louis, MO 63166	Ishan.Al-Shehbaz@mobot.org
Asterids	Walter Judd, Department of Botany, University of Florida, Gainesville, FL 32611	wjudd@botany.ufl.edu
	Richard Olmstead, Department of Botany, University of Washington, Seattle, WA 98195	olmstead@u.washington.edu
Caryophyllales	Mark Chase, Molecular Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, TW9 3DS UK	M.Chase@rbgkew.org.uk

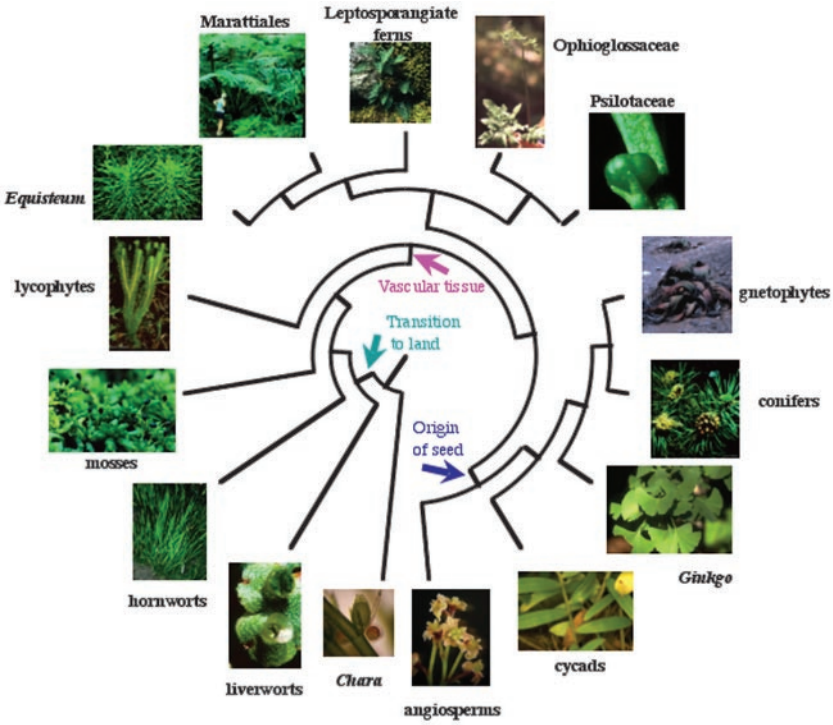

Figure 1. Summary of phylogenetic relationships among major lineages of embryophytes (land plants). Charales are the sister group of the embryophytes. Within the embryophytes, liverworts, hornworts, and mosses are the basal most lineages; however, their precise branching order is uncertain. One of the best supported topologies is depicted with liverworts, hornworts, and mosses as successive sisters to the tracheophytes (vascular plants). Within tracheophytes, there are two clades: monilophytes and spermatophytes (seed plants). Data from Karol et al. (2001), Pryer et al. (2001), and Soltis et al. (2002). Photograph of *Chara* courtesy of R. McCourt; photographs of *Welwitschia* and *Ophioglossum* courtesy of H. Wilson; photographs of *Ginkgo* sp. and *Zamia* courtesy of J. Manhart; photograph of *Polypodium* courtesy of J. Reveal; *Anthoceros* taken from CalPhotos (http://elib.cs.berkeley.edu); other photographs from the online teaching collection of the Botanical Society of America (http://www.botany.org/).

Plants colonized the land approximately 450 mya. Within the land plants, the three lineages long known as the "bryophytes" (liverworts, hornworts, and mosses) do not form a single clade in most analyses but instead form a grade that subtends the tracheophytes (Fig. 1). Furthermore, the precise branching order of the three "bryophyte" lineages remains ambiguous, with different topologies suggested by various data sets. A branching order of liverworts, hornworts, and mosses has emerged as one favored arrangement (e.g. Karol et al., 2001); other data suggest that hornworts, followed by a clade of mosses + liverworts, are the basal branches of the embryophytes (Renzaglia et al., 2000).

Tracheophytes

Vascular plants (tracheophytes) constitute a large and well-defined clade of land plants comprising the lycophytes (e.g. *Lycopodium*, *Selaginella*, and *Isoetes*) as sister to two well-marked clades—monilophytes and seed plants (Pryer et al., 2001; Fig. 1).

Monilophytes (or Moniliforms)

Both molecular and morphological analyses of tracheophytes have recognized a clade of *Equisetum*, Marattiaceae, Psilotaceae, Ophioglossaceae, and leptosporangiate ferns (Kenrick and Crane, 1997; Pryer et al., 2001). Kenrick and Crane (1997) first suggested the presence of this clade (based on one morphological character) and designated these plants Moniliformopses or "moniliforms"; they are now referred to more commonly as monilophytes (Judd et al., 2002). This monilophyte clade unites ancient lineages not previously considered closely related and is sister to a clade of all remaining tracheophytes—the seed plants (Fig. 1).

Seed Plants

Despite repeated efforts, it has been difficult to resolve phylogenetic relationships among extant seed plants, that is, angiosperms and the four lineages of living gymnosperms: cycads, *Ginkgo biloba*, conifers, and Gnetales (for review, see Donoghue and Doyle, 2000; Soltis et al., 2002). Analyses of morphological data generally concur in suggesting that angiosperms and Gnetales are sister groups (the "anthophyte" hypothesis), with extant gymnosperms paraphyletic (that is, not forming a clade but rather a grade; Donoghue and Doyle, 2000).

However, the sister group relationship of Gnetales and angiosperms has not been supported by most molecular analyses. Analyses of combined data sets of multiple genes representing all three plant genomes (plastid, mitochondrion, and nucleus) have found strong support for a clade of extant gymnosperms (Fig. 1; e.g. Bowe et al., 2000; Chaw et al., 2000; Pryer et al., 2001; Soltis et al., 2002). However, some extinct gymnosperms (e.g. Caytoniales and Bennettitales) may be more closely related to angiosperms than to any lineage of living gymnosperm (Donoghue and Doyle, 2000). Cycads and *Ginkgo biloba* are sisters to the remaining living gymnosperms. The relationship between cycads and *Ginkgo biloba*

is unclear; in some analyses, cycads and *Ginkgo biloba* are successive sisters to a clade of conifers and Gnetales, whereas in others, *Ginkgo biloba* and cycads form a clade that is sister to other extant gymnosperms. Some molecular analyses support a surprising placement of Gnetales within conifers as sister to Pinaceae (Bowe et al., 2000; the "gne-pine" hypothesis of Chaw et al., 2000).

The placement of Gnetales within conifers is an excellent example of a molecular phylogenetic result that must be viewed with caution, for several reasons. First, the placement of Gnetales within conifers is supported largely by mitochondrial genes; genes from other genomes do not place Gnetales within conifers. Furthermore, there is conflict between first and second versus third codon positions of cpDNA genes, with different positions supporting different placements of Gnetales. In addition, because most analyses of seed plants have involved small numbers of taxa, the gne-pine hypothesis may be an artifact of inadequate taxon sampling in some analyses. Our current interpretation of relationships among extant seed plants, showing Gnetales as sister to all conifers, is depicted in Figure 1. Analysis of extant gymnosperms exemplifies the complexities inherent in phylogenetic analysis of ancient lineages that have undergone significant extinction.

Angiosperms

The impact of molecular phylogenetic analyses on the angiosperms (flowering plants) has been particularly profound (e.g. Qiu et al., 1999; Graham and Olmstead, 2000; Soltis et al., 2000; Bremer et al., 2002; see below). Because of the wealth of molecular phylogenetic data, angiosperms became the first major group of organisms to be reclassified based largely on molecular data (Angiosperm Phylogeny Group [APG], 1998); data have accumulated so rapidly that this classification was recently revised (APG II, 2003). Readers will find that some family circumscriptions and ordinal groups have changed considerably from traditional classifications (e.g. Cronquist, 1981). Comprehensive trees depicting family level relationships for nearly all of the 300+ angiosperm families (e.g. Soltis et al., 2000; Zanis et al., 2002) and the APG II classification are available at http://www.flmnh.ufl.edu/deeptime/. Although recent classifications (e.g. Cronquist, 1981) may still provide some useful family descriptions, these classifications do not depict current concepts of phylogeny. For interpretations of data in a phylogenetic context and for consistency, authors are urged to follow the APG II (2003) classification.

Think "Eudicots." Abandon "Dicots"

The angiosperms, a clade of 260,000+ species (Takhtajan, 1997), first appeared in the fossil record, conservatively, approximately 130 mya (Hughes, 1994). Standard classifications divided the angiosperms into two large groups, typically recognized at the Linnean rank of class: Magnoliopsida (dicots) and Liliopsida (monocots). Thus, standard comparative studies of physiological pathways and genetic/genomic data have spanned this "monocot-dicot split." However, even preliminary morphology-based studies of angiosperms suggested that this "monocot-dicot split" did not accurately portray relationships. Molecular phylogenetic analyses clearly indicate that the traditional "dicots" are paraphyletic, with the monocots (a clade of $\pm$65,000 species) emerging from among the basal branches of angiosperms (Fig. 2). Following this basal grade of monocots and traditional "primitive dicots" (e.g. Amborellaceae, Nymphaeaceae, Austrobaileyales, and magnoliid clade) is a well-supported clade, the eudicots (Fig. 2). The eudicot clade contains 75% of all angiosperm species, united by the shared feature of triaperturate pollen (pollen with three grooves). The term "monocots" is still useful in that it designates a clade. In contrast, the term "dicots" should be abandoned because it does not correspond to a clade. This change in concept and terminology has already been accepted by many entry level biology and botany textbooks. Comparisons of genes or characters should be based on sister groups, if possible or, minimally, on other monophyletic groups. For example, because the sister group of the monocots remains uncertain, monocots could be compared with members of eudicots or magnoliids. Most of the published molecular comparisons of monocots and dicots have used eudicots as placeholders (e.g. Arabidopsis, *Brassica* spp., and *Antirrhinum* spp.) for the dicots. Thus, many such comparisons are still valid, even if the terminology used ("dicot") was incorrect.

NO SUBCLASSES

Perhaps the best known classification of angiosperms is that of Cronquist (1981), who recognized six subclasses of dicots, Magnoliidae, Hamamelidae, Rosidae, Dilleniidae, Caryophyllidae, and Asteridae, and five subclasses of monocots, although these were followed less frequently. Molecular phylogenies indicate that these subclasses, like the classes Magnoliopsida and Liliopsida, should also be abandoned. The Magnoliidae are paraphyletic, and both the Hamamelidae and Dilleniidae are grossly polyphyletic, with constituent members appearing throughout much of the angiosperm tree. Thus, "Magnoliidae," "Hamamelidae," and "Dilleniidae" do not refer to monophyletic groups, and these names are no longer valid. Cronquist's concepts of Rosidae, Asteridae, and Caryophyllidae must be expanded and revised to correspond to monophyletic groups; these clades are the rosids, asterids, and Caryophyllales sensu

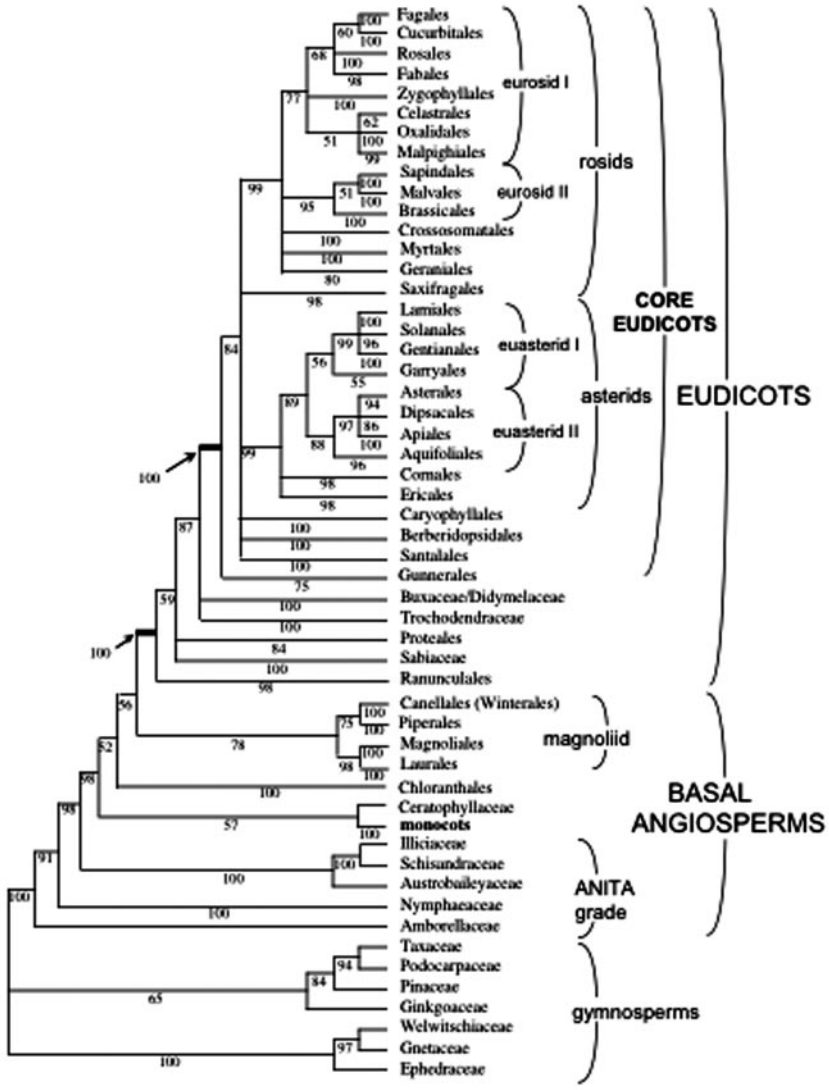

Figure 2. Summary tree of angiosperm relationships based on Soltis et al. (2000, 2003), with basal angiosperm relationships modified following Zanis et al. (2002). Numbers are jackknife values. Values for basal angiosperms are from Zanis et al. (2002); the value for the placement of Gunnerales is from Soltis et al. (2003); other eudicot values are from Soltis et al. (2000).

(APG II, 2003). Although Caryophyllales are recognized at the ordinal level (see APG II, 2003), both rosids and asterids are supraordinal groups that are not assigned a Linnean rank in the APG II classification.

It is important to note that deep-level angiosperm phylogeny is not yet resolved. Relationships among the major clades of eudicots (e.g. rosids, asterids, Caryophyllales, Saxifragales, Santalales, and a few smaller clades) are unresolved (Fig. 2), presenting a limitation for many areas of comparative biology, including comparative genomics.

MODEL GROUPS. OPPORTUNITIES FOR COMPARATIVE GENETICS AND GENOMICS IN THE ANGIOSPERMS

The phylogenetic trees available for many families of angiosperms facilitate interpretation of the evolution of diverse characters (molecular, physiological, and genetic). These trees also aid in the appropriate choice of representative taxa for comparative studies (see also Daly et al., 2001; Hall et al., 2002a); it is often useful to choose representative taxa from across the breadth of a clade and not simply one or two taxa from only a small part of the diversity of that clade.

Because trees depicting organismal phylogenies have accumulated so rapidly, it is often difficult for the nonexpert to know how to obtain a tree for a group of interest. Unfortunately, there is no single source that serves as a compendium of all intrafamilial phylogenetic trees. Judd et al. (2002) provide trees and relevant references for many families of tracheophytes. However, because it is an entry level textbook, many families are not covered. Therefore, we provide a short list of experts (Table I) who can assist with phylogenetic questions for major groups of embryophytes. A larger list is available on the Deep Time website.

Monocots

Molecular analyses have clarified many (but far from all) relationships within monocots (Chase et al., 2000; Soltis et al., 2000), and further analyses are underway (M. Chase and J. Davis, personal communication). The sister group of the monocots remains unclear, but the most comprehensive analyses suggest Ceratophyllaceae (Zanis et al., 2002; Fig. 2).

Poaceae

The Poaceae, or grass family, are an ideal focal point for comparative genetic/genomic research (Kellogg, 2001). The Grass Phylogeny Working Group (2001) has provided the most comprehensive and best supported tree for the grass family. Complete sequencing of the rice (*Oryza sativa*) genome and of entire cpDNA genomes for some genera, as well as extensive genetic/genomic data for crops including wheat (*Triticum aestivum*), sorghum (*Sorghum bicolor*), and maize (*Zea mays*), make tribe Triticeae of particular interest; a firm phylogenetic framework is available not only for the tribe (Kellogg, 2001) but also for individual genera, such as *Hordeum* (Petersen and Seberg, 2003).

Antirrhinum Spp. (Snapdragon and Relatives)

Snapdragon (Plantaginaceae and Lamiales) is one of the best model systems for the study of floral developmental genetics and offers numerous opportunities for comparative genetic and genomic research. Although *Antirrhinum* spp. have long been placed in the family Scrophulariaceae, molecular phylogenetic studies indicate that the traditionally recognized Scrophulariaceae are not a single clade but actually represent a number of distinct clades: Scrophulariaceae in the strict sense; Plantaginaceae, which includes *Antirrhinum*, *Plantago*, and *Veronica*; Orobanchaceae, which contains all of the parasitic taxa formerly placed in either Orobanchaceae or Scrophulariaceae; the new family Calceolariaceae; an expanded Stilbaceae; and an expanded Phyrmaceae (Olmstead et al., 2001).

Solanaceae

Solanaceae contain a number of model organisms, including tomato and potato (*Solanum tuberosum*), tobacco (*Nicotiana tabacum*), peppers (*Capsicum annuum*), and petunia (*Petunia hybrida*). The family has also served as a model for studies of reproductive incompatibility and organization of the nuclear genome. A molecular phylogenetic framework and a provisional reclassification are now available for the family (Olmstead et al., 1999). Molecular studies have also confirmed that Convolvulaceae represent the sister group of Solanaceae (Soltis et al., 2000). As noted, tomato (formerly *Lycopersicon*) is clearly embedded within the large genus *Solanum*, which also includes potatoes. Thus, potato and tomato share very similar linkage maps (e.g. Tanksley et al., 1988; Doganlar et al., 2002) because they share a recent common ancestor.

Legumes (Fabaceae)

The closest relative of the Fabaceae has long been considered a mystery. Phylogenetic analyses have recently shown the closest relatives of Fabaceae to be Surianaceae and Polygalaceae (Soltis et al., 2000). Considerable progress has been made in recent years in clarifying relationships across the family as a whole and also within subclades within the family (Doyle and Luckow, 2003). Recent analyses have also identified the closest relatives of several important crop genera, including *Medicago*, *Gycine*, and *Pisum* (e.g. Kajita et al., 2001; Hu et al., 2002; for review, see Doyle and Luckow, 2003).

Brassicaceae

Brassicaceae offer important opportunities in comparative genomics by extending out from the complete genome sequence of Arabidopsis (e.g. Hall et al., 2002a; Mitchell-Olds and Clauss, 2002). Initial molecular phylogenetic analyses indicated the presence of a broadly defined Brassicaceae (Brassicaceae sensu lato) that also include Capparaceae. More recently, Hall et al. (2002b) found evidence for three well-supported clades within Brassicaceae sensu lato—Capparaceae subfamily Capparoideae, Capparaceae subfamily Cleomoideae, and Brassicaceae sensu stricto—with the latter two clades as sister groups. Rather than a single broadly defined family Brassicaceae, it may be more appropriate to recognize three families: Capparaceae, Cleomaceae, and Brassicaceae (Hall et al., 2002b). The model plants *Brassica* sp. and Arabidopsis are in Brassicaceae. It may be informative to include members of Capparaceae (e.g. *Capparis* spp.) and Cleomaceae (*Cleome* spp.) in comparative genetic and genomic analyses.

Recent phylogenetic studies of Arabidopsis and relatives (Koch et al., 1999, 2001, 2003; Koch, 2003; O'Kane and Al-Shehbaz, 2003) have provided an initial tree for Brassicaceae sensu stricto and identified an Arabidopsis clade that contains the closest relatives of Arabidopsis. However, a more comprehensive analysis of the family is required and is well underway (M. Beilstein, E. Kellogg, and I. Al-Shehbaz, personal communication).

Brassicales

Brassicaceae are part of a well-supported Brassicales (i.e. "glucosinolate clade"; e.g. Rodman et al., 1998; Soltis et al., 2000), a clade of 15 families that were not considered closely related in recent classifications (e.g. Cronquist, 1981). The order offers the

opportunity to investigate the evolution of a host of features considered characteristic of Brassicaceae. Some aspects of genomic and genic diversification will be better understood by extending out from Brassicaceae to relatives in Brassicales.

PHYLOGENY RECONSTRUCTION. A PRIMER

Alignment ("Garbage in; Garbage out")

Alignment of nucleotide and amino acid sequences is a major consideration, particularly in studies of genes from divergent taxa (e.g. rice and Arabidopsis). It seems obvious to state that the phylogenetic analysis of sequences begins with the appropriate alignment of the data themselves, yet alignment remains one of the most difficult and poorly understood facets of molecular data analysis. Detailed coverage of the topic is beyond the scope of this Update, but excellent overviews are provided by Doyle and Gaut (2000) and Simmons and Ochoterena (2000). We will simply restate, as Doyle and Gaut (2000) stress, that researchers should not accept alignments produced with the default settings of any computer algorithm without a critical evaluation by eye. Furthermore, there may be multiple "good" alignments, and all of these should be subjected to phylogenetic analysis.

Life after Neighbor Joining (NJ)

Inferences of orthology require phylogenetic analysis. Although expression patterns and knowledge of function may provide clues to orthology relationships, orthology, by definition, requires historical analysis to disentangle the products of gene duplication and speciation (for useful review of orthology and paralogy, see Doyle and Gaut, 2000; Jensen, 2001; Koonin, 2001). Thus, molecular biologists and geneticists suddenly need to become phylogeneticists. Although molecular phylogeny reconstruction is a relatively young discipline, it nonetheless has a rich and sometimes contentious background, encompassing diverse philosophies and methodologies that are not necessarily apparent to users of most available computer packages. Several approaches can be used in phylogeny reconstruction of molecular sequences: maximum parsimony (MP), maximum likelihood (ML), distance-based methods such as NJ, and Bayesian inference (BI), a new method of phylogenetic inference (Huelsenbeck et al., 2002). All of these methods have strengths and weaknesses (e.g. Swofford et al., 1996; Lewis, 1998; Doyle and Gaut, 2000; Huelsenbeck et al., 2002; Nei and Kumar, 2000), some of which are summarized in Table II.

Although there is a desire among many investigators for rapid phylogeny reconstruction and "instant tree," it may be prudent to explore several methods (e.g. Swofford et al., 1996; Doyle and Gaut, 2000; Nei and Kumar, 2000). There remains a tendency to place more trust in phylogenetic results supported by multiple approaches (Doyle and Gaut, 2000). Regardless of method of phylogenetic inference, however, some measure of internal support (e.g. bootstrap, jackknife, and posterior probabilities; see below) is essential.

Many non-systematists employ NJ to the exclusion of other methods (Nei and Kumar, 2002). The distance measures used in NJ and other distance methods are typically based on models of nucleotide substitution. The NJ algorithm is fast and readily available in software packages such as MEGA (http://www.megasoftware.net/) and PAUP*. However, it also has important weaknesses. For example, NJ provides only a single tree, precluding comparison with other topologies. In reality, many optimal trees may be found in MP and ML analyses, depending on the data set, and these methods allow all optimal or near-optimal trees to be compared. Furthermore, different trees can be obtained with NJ depending on the entry order of the taxa (Farris et al., 1996; see Table II). One solution is to run multiple NJ analyses with different random entry orders of the taxa, accompanied by bootstrap or jackknife analysis (see below). Finally, because sequence differences are summarized as distance values, it is impossible to identify the specific character changes that support a branch. Although proponents of NJ, Nei and Kumar (2000) nonetheless argue for a pluralistic approach. Other methods of phylogenetic inference should be explored in addition to NJ.

MP is preferred by many phylogeneticists because of its theoretical basis and the diagnosable units it produces. The advantages of parsimony over NJ are several (Table II), an important one being that parsimony seeks to recover all shortest trees. Depending on the data set, a parsimony search may yield one (or a few) to hundreds or thousands of equally short trees. These shortest trees can be summarized in a strict consensus tree, which depicts only the nodes present in all equally short trees. In addition, MP analysis provides diagnoses (i.e. specific sets of characters) for each clade and branch lengths in terms of the number of steps (or changes) on each branch of a tree.

Statistical methods of phylogeny reconstruction, incorporating models of nucleotide (or amino acid) substitution, are preferred by many molecular phylogeneticists (see Lewis, 1998). Both ML and BI rely on such models to reconstruct both topology and branch lengths and, thus, are computationally intensive. ML analysis finds the likelihood of the data, given a tree and a model of molecular evolution. Like ML, BI has had a long tenure in statistics. However, it has only recently been introduced into phylogenetics (see Huelsenbeck et al., 2001, 2002). Although BI uses the same models of evolution as some other methods of phylogenetic analyses (e.g. ML and NJ), it represents a powerful tool and perhaps the wave of the future in phylogenetic inference. BI is based on a quantity referred to as the posterior probability of a

Table II. *Comparison of methods of phylogeny reconstruction*

Method	General	Advantages	Disadvantages
Parsimony	Simplest explanation is the best (Ockham's razor)	By minimizing no. of steps, it also minimizes the no. of additional hypothesis (parallel or reversal nucleotide substitutions)	Different results may be obtained based on the entry order of sequences (therefore, perform multiple searches)
	Select the tree or trees that minimize the amount of change (no. of steps)	Searches identify numerous equally parsimonious (shortest) trees; treats multiple hits as an inevitable source of false similarity (homoplasy)	Relatively slow (compared with NJ) with large data sets
		Basic method can be modified by weighting schemes to compensate for multiple hits	Highly unequal rates of base substitution may cause difficulties (e.g. long branch attraction)
		Readily implemented in PAUP*	
		Can identify individual characters that are informative or problematic	
		Can infer ancestral states	
NJ	Involves estimation of pair-wise distances between nucleotide sequences	Fast	Different results may be obtained based on the entry order of sequences
	Pair-wise distances compensate for multiple hits by transforming observed percent differences into an estimate of the no. of nucleotide substitutions using one of several models of molecular evolution	Provides branch lengths	Only a single tree produced; cannot evaluate other trees
	Minimum evolution is a common distance criterion for picking an optional tree (sum of all branch lengths is the smallest)	Uses molecular evolution model	Branch lengths presented as distances rather than as discrete characters (steps)
	NJ algorithm provides a good approximation of the minimum evolution tree	Readily implemented in PAUP* and MEGA	Cannot identify characters that are either informative or problematic
			Cannot infer ancestral states
Maximium Likelihood	Involves estimating the likelihood of observing a set of aligned sequences given a model of nucleotide substitution and a tree	A statistical test (the likelihood ratio test) can be used to evaluate properties of trees	Computationally very intensive (much slower than other methods
		Nucleotide substitution models are used directly in the estimation process, rather than indirectly (as in parsimony)	Practical with only small nos. (fewer than 50) of sequences
		Flexible, models that can incorporate parameters of base frequencies, substitution rates, and variation in substitution rates and, therefore, are "general"; Jukes-Cantor sets a single substitution rate and is more "restricitive"	
		Easily implemented in PAUP*	
		Uses all of the data (invariable sites and unique mutations are still informative, unlike parsimony analysis)	
Bayesian	Uses a likelihood function and an efficient search strategy	Based on the likelihood function, from which it inherits many of its favorable statistical properties	Very large memory demands
	Based on a quality called the posterior probability of a tree	Uses models as in ML	
	Researcher may specify belief in a prior hypothesis prior to analysis	Can be used to analyze relatively large data sets	
		Provides support values	Posterior probabilities (measure of internal support) can be overestimates

tree, a value that can be interpreted as the probability that a tree is correct, given the data. BI uses a likelihood function to compute the posterior probability. Although BI allows the researcher to specify a prior belief in relationships (Table II; Huelsenbeck et al., 2001, 2002), this option has not been explored extensively to date, and Bayesian analyses typically assign equal prior probability values to all possible trees. Whereas ML is not feasible for large data sets (more than perhaps 50 taxa), BI (as implemented in MrBayes; see Huelsenbeck et al., 2001) incorporates a faster search strategy (using Markov chains) and can be used on data sets of several hundred taxa to find tree, branch lengths, and support (but see Suzuki et al., 2002).

Certainly a frustrating aspect of phylogenetic analysis to those outside of the field is the number of inference methods available. NJ is widely used, in part, because of its speed and ready availability in computer packages such as MEGA. It also is part of alignment packages such as MegAlign (http://www.dnastar.com/cgi-bin/php.cgi?r10.php). However, parsimony can be readily implemented using PAUP* (Swofford, 1998; NJ and ML are also part of the PAUP package). PAUP* is often not employed by molecular biologists, however, because the user friendly version with pull-down menus is made for Macintosh, not Windows, operating systems.

Internal Support for Clades

Some measure of internal support for clades should be provided on all phylogenetic trees. Resampling approaches, such as the bootstrap and the jackknife, are easily computed using PAUP* for parsimony, NJ, and ML analyses, and parsimony jackknifing is performed by *Jac* (Farris et al., 1996). The pros and cons of the jackknife versus bootstrap have been discussed (e.g. Farris et al., 1996; Soltis and Soltis, 2003). A reasonable number of replications should be employed, but "reasonable" varies with the size of the data set, the specifications of the analysis, and the patience of the investigator. It has been argued (Farris et al., 1996) that resampling methods should maximize the number of replicates at the expense of detailed searches in each replicate. Thus, with "fast" methods that conduct little or no branch swapping per replicate, 1,000 or more replicates are quickly obtained. A smaller number of replicates (e.g. 100) may be suitable for bootstrap and jackknife analyses that include detailed searches per replicate.

Interpretations of bootstrap and jackknife values vary (for review, see Soltis and Soltis, 2003), although few view these values in a strict statistical sense. Bootstrap values are conservative, but biased, measures of phylogenetic accuracy (Hillis and Bull, 1993), with values of 70% or greater corresponding to "true" clades in experimental phylogenies (Hillis and Bull, 1993). Thus, some consider values of 70% or more as indicators of strong support, whereas others reserve "strong support" for values of 90 or 95% and above. Although different phylogenetic methods may yield different optimal topologies, the differences generally involve poorly supported clades. Those clades that are strongly supported generally appear in topologies regardless of the method of phylogenetic inference. Additional measures of support include the decay index or Bremer support (Bremer, 1994) for parsimony analyses and the posterior probabilities generated in BI.

Measures of internal support indicate those relationships in which we should, and should not, have confidence. A recently identified clade of MADS-box genes appears as the sister group to the well-known B class floral genes that specify the identity of petals and stamens in Arabidopsis and snapdragon. Becker et al. (2002) termed this new clade B_{sister} and determined that these genes are present in diverse seed plants. Although the monophyly of the B_{sister} clade received 92% bootstrap support, the placement of the B_{sister} clade as sister to the clade of B class genes received only 77% bootstrap support. With this level of support, it is reasonable to question whether the B_{sister} clade is really the sister group of the clade of B class genes. Increased sampling of B_{sister} genes from additional taxa and more rigorous analyses are needed to establish with certainty the placement of the B_{sister} clade within the MADS box genes of plants.

MOLECULAR CLOCKS. RATES AND DATES OF GENE DIVERSIFICATION

Many efforts to date evolutionary divergences using a molecular clock have yielded age estimates that are grossly inconsistent with the fossil record, regardless of method of tree construction. For example, molecular-based estimates of divergence times in plants reveal a vast range of dates. Using molecular data, the age of the angiosperms has been estimated as 350 to 420 mya, greater than 319, 200, to 140 to 190 mya (for review, see Sanderson and Doyle, 2001). However, the oldest unequivocal angiosperm fossils are 125 to 135 mya (for review, see Soltis et al., 2002).

Many sources of error and bias can affect molecular-based estimates of divergence times (see Sanderson and Doyle, 2001; Soltis et al., 2002). Obviously, an incorrect topology will yield erroneous estimates, with the magnitude of the problem depending on the extent of the topological error (Sanderson and Doyle, 2001). Inaccurate calibration will bias the resulting estimates. Also problematic are heterogeneous rates of evolution among lineages (see Sanderson and Doyle, 2001; Soltis et al., 2002). Inadequate taxon sampling can compound the problem. Estimates of divergence times can also vary among genes or other data partitions (e.g. among codon positions). Another potential source of error is the method used to estimate divergence dates. Sanderson and Doyle

(2001) used molecular data to examine angiosperm divergences and found that the age of crown group angiosperms ranges from 68 to 281 mya, depending on data, tree, and assumptions, with most estimates falling between 140 and 190 mya.

Given that rate heterogeneity among lineages is common in most molecular-based trees, can we reliably use molecular data to estimate divergence times? Simple clock-based approaches to estimating divergence times are not likely to yield meaningful estimates. However, several approaches have been proposed when the assumption of rate constancy is violated: linearized trees (Takezaki et al., 1995), nonparametric rate smoothing (Sanderson, 1997, 1998), penalized likelihood (Sanderson, 2002), Bayesian approaches (e.g. Huelsenbeck et al., 2002; Thorne and Kishino, 2002), and "PATH" (Britton et al., 2002; for review of methods and instructions for implementing nonparametric rate smoothing, see http://www.flmnh.ufl.edu/deeptime/). Although methods to accommodate deviations from a steady molecular clock are still under development, it is nonetheless possible to estimate dates of divergence, given: (a) a reliable calibration point or points, (b) adequate sampling of taxa and characters, and (c) a method that is robust to rate heterogeneity. Confidence intervals for the estimated dates and consistency with the fossil record provide means for assessing the reliability of age estimates. Despite attempts to accommodate deviations from constant evolutionary rates, however, confidence intervals are typically large, and divergence times should be interpreted carefully.

SUMMARY AND FUTURE PROSPECTS

An exciting recent development is the merging of phylogenetics and genomics. Phylogenetic hypotheses have become the framework for the choice of organisms in genomic analyses, and more and more molecular biologists are using phylogenetic trees to guide their sampling of taxa for comparative research. This trend will continue. Systematics is moving rapidly; therefore, molecular biologists are encouraged to contact systematics "experts" for help in obtaining the best supported trees for a given clade of interest. We stress the importance of a rigorous phylogenetic analysis of data. It is ironic, for example, that researchers may spend years gathering gene sequence data, but then want an immediate phylogenetic "answer" within seconds or minutes. A thorough phylogenetic analysis, evaluating alternative alignments, exon versus intron boundaries, using different phylogenetic methods, and obtaining estimates of internal support, may take several weeks or more, and this should not be considered an unreasonable investment of time. Our review of issues relating to phylogeny reconstruction also illustrates the need for more "quick courses" in phylogeny reconstruction for molecular biologists interested in constructing gene trees.

ACKNOWLEDGMENTS

We thank Jeff Doyle, Bernie Hauser, Alice Harmon, and two anonymous reviewers for helpful comments on earlier drafts of this paper.

Received February 27, 2003; returned for revision March 30, 2003; accepted May 12, 2003.

LITERATURE CITED

Angiosperm Phylogeny Group (1998) An ordinal classification for the families of flowering plants. Ann Missouri Bot Gard **85:** 531–553

Angiosperm Phylogeny Group II (2003) An updated classification of the angiosperms. Bot J Linn Soc **141:** 399–436

Becker A, Kaufmann K, Freialdenhoven A, Vincent C, Li MA, Saedler H, Theissen G (2002) A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes. Mol Genet Genomics **266:** 942–950

Becker A, Winter K-U, Meyer B, Saedler H, Theissen G (2000) MADS-box gene diversity in seed plants 300 million years ago. Mol Biol Evol **17:** 1425–1434

Bowe LM, Coat G, DePamphilis CW (2000) Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. Proc Natl Acad Sci USA **97:** 4092–4097

Bremer K (1994) Branch support and tree stability. Cladistics **10:** 295–304

Bremer B, Bremer K, Heidari N, Erixon P, Olmstead RG, Källersjö M, Anderberg A, Barkhordarian E (2002) Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. Mol Phylogenet Evol **24:** 274–301

Britton T, Oxelman B, Vinnersten A, Bremer K (2002) Phylogenetic dating with confidence intervals using mean path lengths Mol Phylogenet Evol **24:** 58–65

Chase MW, Soltis DE, Soltis PS, Rudall PJ, Fay MF, Hahn WJ, Sullivan S, Joseph J, Molvray M, Kores PJ et al. (2000) Higher-level systematics of the monocutyledons: an assessment of current knowledge and a new classification. *In* KL Wilson, DA Morrison, eds, Monacots: Systematics and Evolution. CSIRO Publishing, Victoria, Australia, pp 3–16

Chaw SM, Parkinson CL, Cheng Y, Vincent TM, Palmer JD (2000) Seed plant phylogeny inferred from all three plant genomes: monophyly of extant gymnosperms and origin of Gnetales from conifers. Proc Natl Acad Sci USA **97:** 4086–4091

Cronquist A (1981) An Integrated System of Classification of Flowering Plants. Columbia University Press, New York

Daly DC, Cameron KM, Stevenson DW (2001) Plant systematics in the age of genomics. Plant Physiol **127:** 1328–1333

Doganlar S, Frary A, Daunay MC, Lester RN, Tanksley SD (2002) A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the Solanaceae. Genetics **161:** 1697–16711

Donoghue MJ, Doyle JA (2000) Seed plant phylogeny: demise of the anthophyte hypothesis? Curr Biol **10:** R106–R109

Doyle JJ, Gaut B (2000) Evolution of genes and taxa: a primer. Plant Mol Biol **42:** 1–23

Doyle JJ, Luckow MS (2003) The rest of the iceberg: legume diversity and evolution in a phylogenetic context. Plant Physiol (in press)

Farris JS, Albert VA, Källersjö M, Lipscomb D, Kluge AG (1996) Parsimony jackknifing outperforms neighbor-joining. Cladistics **12:** 99–124

Graham SW, Olmstead RG (2000) Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. Am J Bot **87:** 1712–1730

Grass Phylogeny Working Group (2001) Phylogeny and subfamilial classification of the grasses (Poaceae). Ann Missouri Bot Gard **88:** 373–457

Hall AE, Fiebig A, Preuss D (2002a) Beyond the *Arabidopsis* genome: opportunities for comparative genomics. Plant Physiol **129:** 1439–1447

Hall JC, Sytsma KJ, Iltis HH (2002b) Phylogeny of Capparaceae and Brassicaceae based on chloroplast sequence data. Am J Bot **89:** 1826–1842

Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst Biol **42:** 182–192

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. Science **294:** 2310–2314

Huelsenbeck JP, Larget B, Miller RE, Ronquist F (2002) Potential applications and pitfalls of Bayesian inference of phylogeny. Syst Biol **51:** 673–688

Hu J-M, Lavin M, Wojciechowski M, Sanderson MJ (2002) Phylogenetic analysis of nuclear ribosomal ITS/5.8S sequences in the tribe Millettieae (Fabaceae): *Poecilanthe-Cyclolobium*, the core Millettieae, and the *Callerya* group. Syst Bot **27:** 722–733

Hughes NF (1994) The Enigma of Angiosperm Origins. Cambridge University Press, Cambridge, UK

Jensen RA (2001) Orthologs and paralogs: we need to get it right. Genome Biol **2:** 1002.1–1002.3

Judd WS, Campbell CS, Kellogg EA, Stevens PF, Donoghue MJ (2002) Plant Systematics: A Phylogenetic Approach. Sinauer Associates, Inc., Sunderland, MA

Kajita T, Ohashi H, Tateishi Y, Bailey CD, Doyle JJ (2001) *rbcL* and legume phylogeny with particular reference to Phaseoleae, Millettieae, and allies. Syst Bot **26:** 515–536

Karol KG, McCourt RM, Cimino MT, Delwiche CF (2001) The closest living relatives of land plants. Science **294:** 2351–2352

Kellogg EA (2001) Evolutionary history of the grasses. Plant Physiol **125:** 1198–1205

Kenrick P, Crane PR (1997) The Origin and Early Evolution of Land Plants. Smithsonian Institution Press, Washington, DC

Koch M (2003): Molecular Phylogenetics, Evolution and Population Biology in the Brassicaceae. *In* AK Sharma, A Sharma A, eds, Plant Genome: Biodiversity and Evolution, Vol 1: Phanerogams. Science Publishers, Inc., Enfield, NH (in press)

Koch M, Bishop J, Mitchell-Olds T (1999) Molecular systematics and evolution of *Arabidopsis* and *Arabis*. Plant Biol **1:** 529–537

Koch M, Haubold B, Mitchell-Olds T (2001) Molecular systematics of the Brassicaceae: evidence from coding plastid *matK* and nuclear *Chs* sequences. Am J Bot **88:** 534–544

Koch M, Mummenhoff K, Al-Shehbaz IA (2003): Molecular systematics, evolution, and population biology in the mustard family (Brassicaceae): a review of a decade of studies. Ann Missouri Bot Gard (in press)

Kolukisaoglu HM, Marx MS, Weigmann C, Hanelt S, Schneider-Portsch AW (1995) Divergence of the phytochrome gene family predates angiosperm evolution and suggests that *Selaginella* and *Equistem* arose prior to *Psilotum*. J Mol Evol **41:** 329–337

Koonin EV (2001) An apology for orthologs: or brave new memes. Genome Biol **2:** 1005.1–1005.2

Lewis P (1998) Maximum likelihood as an alternative to parsimony for inferring phylogeny using nucleotide sequence data. *In* DE Soltis, PS Soltis, JJ Doyle, eds, Molecular Systematics of Plants II: DNA Sequencing. Kluwer, Boston, pp 132–187

Ma H, dePamphilis C (2000) The ABCs of floral evolution. Cell **101:** 5–8

Mathews S, Sharrock RA (1997) Phytochrome gene diversity. Plant Cell Environ **20:** 666–671

McDowell JM, Huang S, McKinney EC, An Y-Q, Meacher RB (1996) Structure and evolution of the actin gene family in *Arabidopsis thaliana*. Genetics **142:** 587–602

Mitchell-Olds T, Clauss MJ (2002) Plant evolutionary genomics. Curr Opin Plant Biol **5:** 74–79

Nei M, Kumar S (2000) Molecular Evolution and Phylogenetics. Oxford University Press, Oxford

O'Kane SL, Al-Shehbaz IA (2003) Phylogenetic position and generic limits of *Arabidopsis* (Brassicaceae) based on sequences of nuclear ribosomal DNA. Ann Missouri Bot Gard (in press)

Olmstead RG, dePamphilis CW, Wolfe AD, Young ND, Elisons WJ, Reeves A (2001) Disintegration of the Scrophulariaceae. Am J Bot **88:** 348–361

Olmstead RG, Sweere JA, Spangler RE, Bohs L, Palmer J (1999) Phylogeny and provisional classification of the Solanaceae based on chloroplast DNA. *In* M Nee, DE Symon, RN Lester, JP Jessop, eds, Solanaceae IV. Royal Botanic Gardens, Kew, UK, pp 111–117

Petersen G, Seberg O (2003) Phylogenetic analyses of the diploid species of *Hordeum* (Poaceae) and a revised classification of the genus. Syst Bot **28:** 293–306

Pryer KM, Schneider H, Smith AR, Cranfill R, Wolf P, Hunt JS, Sipes SD (2001) Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. Nature **409:** 618–622

Pryer KM, Schneider H, Zimmer EA, Banks JA (2002) Deciding among green plants for whole genome studies. Trends Plant Sci **7:** 550–554

Qiu YL, Lee J, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M, Zimmer EA, Chen Z, Savolainen V, Chase MW (1999) The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. Nature **402:** 404–407

Renzaglia KS, Duff RJ, Nickrent DL, Garbary DJ (2000) Vegetative and reproductive innovations if early land plants: implications for a unified phylogeny. Philos Trans R Soc Lond B **355:** 769–793

Rodman JE, Soltis PE, Sytsma KJ, Karol KG (1998) Parallel evolution of glucosinolate biosynthesis inferred from congruent nuclear and plastid gene phylogenies. Am J Bot **85:** 997–1006

Sanderson MJ (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. Molec Biol Evol **14:** 1218–1231

Sanderson MJ (1998) Estimating rate and time in molecular phylogenies: beyond the molecular clock. *In* DE Soltis, PS Soltis, JJ Doyle, eds, Molecular Systematics of Plants II. Kluwer, Boston, pp 242–264

Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol **19:** 101–109

Sanderson MJ, Doyle JA (2001) Sources of error and confidence intervals in estimating the age of angiosperms from *rbcL* and 18S rDNA data. Am J Bot **88:** 1499–1516

Simmons MP, Ochoterena H (2000) Gaps as characters in sequence-based phylogenetic analyses. Syst Biol **49:** 369–381

Soltis DE, Soltis PS (2000) Contributions of plant molecular systematics to studies of molecular evolution. Plant Mol Biol **42:** 45–75

Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF et al. (2000) Angiosperm phylogeny inferred from a combined data set of 18S rDNA, rbcL and atpB sequences. Bot J Linn Soc **133:** 381–461

Soltis PS, Soltis DE (2003) The bootstrap in phylogeny reconstruction. Am Stat (in press)

Soltis PS, Soltis DE, Savolainen V, Crane PR, Barraclough T (2002) Rate heterogeneity among lineages of land plants: integration of molecular and fossil data and evidence for molecular living fossils. Proc Natl Acad Sci USA **99:** 4430–4435

Spooner DM, Anderson GJ, Jansen RK (1993) Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes and pepinos (Solanaceae). Am J Bot **80:** 676–688

Susuki Y, Glazko GV, Nei M (2002) Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc Natl Acad Sci USA **99:** 16138–16143

Swofford DL (1998) PAUP* 4.0: Phylogenetic Analysis Using Parsimony (and Other Methods). Sinauer Associates, Sunderland, MA

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996) Phylogenetic inference. *In* DM Hillis, C Moritz, BK Mable, eds, Molecular Systematics. Sinauer Associates, Sunderland, MA, pp 407–514

Takezaki N, Rzhetsky A, Nei M (1995) Phylogenetic test of the molecular clock and linearized trees. Mol Biol Evol **12:** 823–833

Takhtajan A (1997) Diversity and classification of flowering plants. Columbia University Press, New York

Tanksley SD, Bernatsky R, Lapitan NL, Prince JP (1988) Conservation of gene repertoire but not gene order in pepper and tomato. Proc Natl Acad Sci USA **84:** 6419–6423

Thorne JL, Kishino H (2002) Divergence time and evolutionary rate estimation with multilocus data. Syst Biol **51:** 689–702

Walbot V (2000) A green chapter in the book of life. Nature **408:** 794–795

Waters ER, Vierling E (1999) The diversification of plant cytosolic small heat shock proteins preceded the divergence of mosses. Mol Biol Evol **16:** 127–139

Zanis MJ, Soltis DE, Soltis PS, Mathews S, Donoghue MJ (2002) The root of the angiosperms revisited. Proc Natl Acad Sci USA **99:** 6848–6853