

Supertree Analysis of the Plant Family Fabaceae

Tiffany J. Morris
Computational Biosciences Program
Arizona State University

Martin F. Wojciechowski
Internship Advisor
School of Life Sciences
Arizona State University

Internship:
June 2004-December 2004

Report Number: 04-21

Table of Contents

Abstract.....	page 3
Introduction.....	page 4
Materials and Methods.....	page 6
Results.....	page 10
Discussion.....	page 11
Future Directions of the Project.....	page 12
Appendix.....	page 14
Table 1.....	page 14
Table 2.....	page 14
Table 3.....	page 15
Figure 1.....	page 18
Captions for Figures 2-4.....	page 19
Figure 2.....	page 20
Figure 3.....	page 23
Figure 4.....	page 26
References.....	page 41
Studies Referenced for Analysis.....	page 42

Abstract

The “Tree-of-Life” is a national and international project to collect information about the origin, evolution, and diversity of organisms, with the goal of producing a tree of all life on Earth (Pennisi, 2003). The obstacles to achieving this goal are many. From questions related to the kinds and number of data to be used, to building that phylogeny, to the methodological and computational resources required to analyze the massive amounts of data expected to be necessary to bring this to fruition. The development of efficient methods for phylogenetic analyses of thousands (hundreds of thousands) of taxa has become increasingly important to this endeavor, as well as somewhat controversial. Supertree construction is one of these methods proposed for reconstructing phylogenies on a large scale (Bininda-Emonds, 2004). This method combines phylogenetic trees (topologies), which overlap taxonomically, rather than the primary data, to produce a larger tree using standard methods of analyses such as parsimony. This method also has the advantage of being able to incorporate trees derived from many different kinds of data into tree estimation. For the study presented here, a comprehensive literature search identified 185 phylogenetic trees published since 1984 on the plant family Fabaceae (the “legumes”). Of these, 68 trees were chosen for supertree analysis and were divided into three major groups, two of which correspond to strongly-supported monophyletic groups identified by recent studies that sampled extensively across the family and are based on standard phylogenetic analyses of single data sets (e.g. gene sequences). Supertrees of taxa that correspond to the three subfamilies Mimosoideae, Caesalpinioideae, and Papilionoideae will be presented here.

Introduction

Phylogenetic analysis allows us to statistically evaluate the differences between species or comparative analyses between large groups of species. The result of these analyses produce phylogenies (evolutionary trees), which allow us to visualize the relationships between the species analyzed (Bininda-Emonds, 2004). The “Tree-of-Life” project is the ultimate phylogenetic analysis. The project described in this paper is expected to contribute to the tree of life by producing a large-scale phylogeny of the flowering plant family Fabaceae (or Leguminosae, the legumes). The Fabaceae family was chosen for this analysis because it is one of the largest families of flowering plants with 750 genera and approximately 18,000 species (Wojciechowski et al., 2004), and can serve as a model for such an approach. Fabaceae is the third largest family of flowering plants, it is widely distributed on every continent and in a wide variety of habitats, and many species are agriculturally and economically important. The higher-level phylogenetic analysis of the family has been the subject of on-going investigation by the project’s advisor. Approximately 150-200 smaller-scale phylogenetic studies, based on morphological and molecular evidences, have been published in the primary literature on this family since the late 1980’s, but an in-depth analysis comparing all the available data has not been done.

There are two basic ways to approach this problem of how to construct larger phylogenies, with hundreds or thousands of taxa. The “total evidence” approach requires the data that is being combined to be compatible (at least two taxa in common) while the “taxonomic congruence” approach requires the data to have the same set of taxa. The concatenation of multiple sequences from a maximal number of taxa from sequence databases is an option, but given the current sparse and often incomplete nature of the amount of data in databases like Genbank (URL, <http://www.ncbi.nlm.nih.gov/>) on a per taxon basis this

approach leaves many “gaps” in coverage (Sanderson et. al., 2003). One approach to this problem is the construction of a “supertree”.

Supertrees are estimates of phylogeny assembled from sets of smaller estimates (source trees), which share some but not necessarily all their taxa in common (Sanderson et al., 1998). Informally supertrees have been a part of the methods for systematics for a long time. Small, nested trees were simply pasted together; this was the only method available to obtain a bigger, more comprehensive representation of the phylogeny. Now more formal methods are available that estimate phylogenies more objectively using the widest selection of information (Bininda-Emonds, 2004). To reconstruct a supertree of legumes, tree topologies rather than primary character data from published phylogenetic studies can be combined using several of these formal methods.

Parsimony methods are commonly used in phylogenetic analysis and were the methods of choice utilized for most analyses included in this study. The basic idea of parsimony is to find a tree that minimizes the number of changes of the state of the characters being analyzed (Felsenstein, 2004). The algorithm we used for this study is the matrix representation using parsimony (MRP). This method can combine incompatible source trees using existing phylogenetic software (Bininda-Emonds, 2004). In MRP, the nodes of each source tree are recoded as a binary character and scored for each taxon. If the taxon is descended from the focal node it is scored as a “1”; if it is not descended from the focal node but is present in the same source tree it is scored as a “0” (zero); all other taxa are considered missing and are scored as “?” in the new matrix (Bininda-Emonds et al., 2002).

Materials and Methods

A thorough literature search was conducted to identify all phylogenetic studies that have been conducted on the Fabaceae family, essentially since the late 1980s. To store the data for each of the studies I designed a database in Microsoft Access (see Appendix Figure 1), and extracted the following information from each study to store in the database: complete literature citation, the main taxon the study investigated (e.g. the genus *Medicago*), the total number of taxa used in the phylogenetic analysis, the outgroup(s) utilized for the phylogenetic analysis, the kind of characters used to compare taxa (molecular sequences, morphological, etc), the phylogenetic method used (parsimony, maximum likelihood, neighbor joining, etc.), the type of statistical support value used in the analyses, if any (bootstrap, jackknife, etc.), whether data from the study is stored in Genbank or TreeBASE, the types of trees presented in the paper (strict consensus, single representative tree derived from parsimony, maximum likelihood, etc.), and whether the data in the study was “independent” or if it had been used in an analysis that resulted in a tree presented in an earlier paper. We examined each of the studies to determine which ones would be used for our supertree analysis. We eliminated those studies in which less than two of the taxa in the study overlapped with taxa in any other studies; this is a requirement of formal supertree methods (Bininda-Emonds et al., 2002). We also eliminated those studies with data we considered non-independent, meaning there was overlap in the primary data used for previous analyses. This is considered an important issue in supertree construction (Bininda-Emonds et al., 2004) and as more studies are published and the data is reused for subsequent analyses, it may become more difficult to control for. For example, many authors publish a study on taxon X and then add more of the same (e.g., sampling more taxa) or new data (e.g., another gene sequence) and reanalyze the original data as part of a larger dataset containing the new data. Thus, these two studies overlap in the primary data and if used in the supertree construction the shared data sets will effectively be

weighted and may have more influence on the supertree results (Bininda-Emonds et al., 2004). It was important to examine studies by the same authors and determine if datasets had been reused in this fashion and if so, to eliminate the older study from our analysis. Studies published on hypothetical phylogenetic trees and other non-relevant studies (i.e., not based on analyses of primary data) were also eliminated from consideration for inclusion in our supertree analysis.

The next step was to select a representative tree (the “best tree”) from each study and to obtain the tree descriptions for these trees in Newick format (the standard used in most phylogenetic analysis/presentation software such as PAUP and PHYLIP, e.g., Maddison et al., 1997). We obtained these tree descriptions either from TreeBASE (an online phylogenetic tree database: URL, www.treebase.org), directly from the authors of the studies, or by recreating the tree topology exactly as it appears in the published figure(s) from the study using MacClade (Maddison and Maddison, 2003), a program that allows tree topologies to manually be entered resulting in a tree description.

Once we had collected the tree descriptions the next step was to edit them. Many of the trees descriptions we obtained contained spelling errors, name standardization errors, and formatting errors, requiring that they be analyzed word by word in addition to locating missing commas and additional extraneous punctuation. In addition, many trees contained multiple accessions of the same species, or included other non-necessary information in the taxon name (e.g., GenBank accession number). In these situations, the accession numbers were removed and the duplicate taxa were also removed if their relationship was resolved with respect to the other accessions of the same taxon.

Once we obtained the tree descriptions and had begun the editing process we began to consider what programs would be used to construct the actual supertrees. Two online supertree

programs seemed promising. Rod Page of the University of Glasgow has a supertree website (<http://darwin.zoology.gla.ac.uk/cgi-bin/supertree.pl>) as does Iowa State University (http://genome.cs.iastate.edu/supertree/userdata_analysis/userdata_analysis.html). Generation of supertree matrices using these sites proved to have limitations and considering the large number of taxa we were using they were not appropriate for our analysis. We opted to use different programs and complete the analyses ourselves.

As we began to decide how to approach the analysis, we realized it made sense to break the analysis up into smaller more manageable pieces. We divided the studies based on which subfamily the taxa in the study most represented. Trees files (a file containing tree descriptions of all the input or source trees) in NEXUS format (Maddison et al., 1997) were created for each subfamily.

To create the supermatrix from the source trees file, we utilized a program called “r8s” (Sanderson, 2003). r8s is a program designed for estimating absolute rates of molecular evolution, but it also has options for phylogenetic analysis. r8s used the source trees in the trees file and the MRP algorithm to create a supermatrix for subsequent analysis by standard phylogenetic methods such as parsimony.

At this point, we also decided to add additional characters (equivalent to adding “topological constraints”) to our analysis. Legumes as a family, as well as the subfamilies Mimosoideae and Papilionoideae have been highly supported (by nonparametric bootstrap and Bayesian probabilities) as monophyletic groups in many studies. To give the support for these groups additional “weight” in our analysis we added extra character(s) in each supermatrix. For example, in each of the supermatrices for the three subfamilies an additional binary character was added for all taxa in the legume family. For this character, all outgroups were scored as 0 and all taxa in the group were scored as 1.

To minimize the topological problems caused by the various studies using different taxa as outgroups, we removed taxa in each subfamily supermatrix that represented the other two subfamilies. For example, in the supermatrix constructed from source trees containing Mimosoideae taxa, all taxa in the Papilionoideae subfamily were removed and an additional character was added giving weight only to Mimosoideae. In the Caesalpinioideae supermatrix, all taxa in the Mimosoideae and Papilionoideae subfamilies were removed and additional characters were added for monophyletic groups that were highly supported (100% bootstrap values) in the source trees. This procedure was repeated in constructing the supermatrix containing the source trees corresponding to the papilionoids. (This method of matrix representation is described in Bininda-Emonds, 2002).

Once a supermatrix was constructed, edited accordingly, and the appropriate topological constraints had been added, the matrix was analyzed using the software program PAUP* (Phylogenetic Analysis Using Parsimony, and other methods*; Swofford, 2002). For each supermatrix analysis, we conducted three separate heuristic searches using different addition sequence procedures: simple, closest, and random. For each heuristic search we saved a maximum of 5000 trees (MAXTREES = 5000), holding five trees at each step, and branch swapping using the TBR (tree bisection-reconnection) branch-swapping algorithm. Once the searches were completed, the 5000 stored trees were combined using consensus methods into a final 90% majority rule consensus tree, and included all other compatible groupings. The consensus tree was rooted with an outgroup. This consensus tree represents the supertree for each of the subfamilies.

Results

The literature search resulted in 185 studies published since 1984. These studies range in size from 216 to 1797 taxa and use a variety of different characters in their analyses (gene sequences, non-coding DNA sequences, morphology, binary characters – e.g., loss of one copy of the chloroplast DNA inverted repeat, or presence/absence of a large inversion). After eliminating the non-relevant studies, studies without the minimum two overlapping taxa, and those with data non-independence, the total number of candidate studies remaining for the analysis was 68.

We began with the tree descriptions for 16 trees that had been utilized in a smaller, previously published legume supertree on the taxon Hologalegina (Wojciechowski et al., 2000). Fourteen of the remaining trees were available in TreeBASE, but other tree descriptions, particularly the larger ones, were requested directly from the authors of the studies. Authors were generally helpful and contributed 9 tree descriptions to the project. The majority of the remaining tree descriptions (18 trees) were created using MacClade, a program that allows tree topologies to be reconstructed manually and then saved in the usual format as a tree description. Of the 68 trees of interest, we were unable to obtain 11 as of November 15. Hopefully these trees can be obtained at a later time and incorporated into the final family supertree.

The 57 studies that were obtained and used for my analysis were divided into groups according to subfamilies and are listed in the Appendix, Tables 1-3. The number of taxa included in each study is shown to indicate the size of the original author's analysis. One study, Wojciechowski et al. (2004) was used as a “backbone” or framework study to provide a family wide tree for all three subfamilies and is therefore listed in each table.

A supermatrix was constructed for each subfamily using r8s and edited as described above. Parsimony analyses of the supermatrix for the mimosoid group, containing 216 taxa and 429 characters, resulted in the consensus supertree which is shown in Figure 2. Parsimony

analyses of the supermatrix for the caesalpinoid group, containing 343 taxa and 605 characters, resulted in the consensus supertree, which is shown in Figure 3. Finally, parsimony analyses of the supermatrix for the papilionoid group, containing 1797 taxa and 2045 characters, resulted in the consensus supertree, which is shown in Figure 4.

Discussion

The division of a final supertree for the family into three separate, smaller, supertrees, made sense because the three subfamilies (papilionoids, mimosoids, and caesalpinoids) are typically studied independently. Mimosoideae and Papilionoideae are both strongly supported as monophyletic groups nested in the basal grade that contains all taxa of the Caesalpinioideae. Our plan was to obtain a supertree for each of the subfamilies and then combine them into a super-supertree, which will be completed at a later time by this project's advisor.

The supertree method has been applied to studies involving several plant, animal, and bacterial groups including grasses, Primates, Carnivora, Marsupialia (Bininda-Emonds, 2004). However, the largest of these previous studies was done on Chiroptera (bats) and contained 916 taxa. Our study currently includes over twice that many taxa and when it is finally completed will likely include even more. Our study will be useful in evaluating the supertree method as an option for solving such computationally complex problems. Inferring an optimal phylogenetic tree becomes NP-complete with 1000's of taxa causing the possible tree space to grow exponentially (Page, 2004). The supertree method is one way to make this problem more approachable.

An additional issue that was not dealt with in this study is the problem of synonymy, that is, many organisms have more than one accepted scientific name. There are software tools

and comprehensive taxonomic lists available to deal with this, but it is difficult to deal with and as the number and sources of data increases the problem becomes more complex (Page, 2004).

Future directions of the project

Our goal is to complete the project by combining the supertrees for each of the subfamilies into one supertree for the entire legume family. This project is being done in collaboration with Drs. M. J. Sanderson and M. McMahon at UC Davis who are now compiling a maximal concatenated dataset (“supermatrix”) of the DNA and protein sequences from legumes that are currently available in GenBank, using the approach described in Sanderson et al. (2003). The final supertree obtained at the conclusion of this study and the one obtained from the supermatrix constructed at UC Davis will be compared to that derived from a recently completed tree of the family based on standard phylogenetic analyses of a single chloroplast gene sequence dataset (Wojciechowski et al., 2004). Completion of a supertree will also enable detailed comparisons of the efficiency and accuracy of tree reconstruction methods of large clades of the tree of life.

Appendix

Table 1: Mimosoideae Studies. All the studies used for the construction of the Mimosoid Supertree and the number of taxa in each one.

	Year Published	Lead Author	Number of Taxa
1	2004	Wojciechowski M.F.	34/330 taxa*
2	2003	Hughes C.E	72 taxa
3	2003	Miller J.T	60 taxa
4	2000	Clarke H.D	26 taxa

*Number of mimosoid taxa/total number of taxa

Table 2: Caesalpinioideae Studies. All the studies used for the construction of the Caesalpinioideae Supertree and the number of taxa in each one.

	Year Published	Lead Author	Number of Taxa
1	2004	Wojciechowski M.F.	33/330 taxa*
2	2003	Haston E.M.	28 taxa
3	2003	Herendeen P.S.	220 taxa
4	2003	Schnabel A.	13 taxa
5	2003	Simpson B.B	81 taxa
6	2002	Davis C.C	7 taxa
7	2001	Brouat C.	13 taxa
8	1998	Schnabel A.	13 taxa

*Number of Caesalpinioideae taxa/total number of taxa

Table 3: Papilionoid Studies. All the studies used for the construction of the Papilionoid Supertree and the number of taxa in each one.

	Year Published	Lead Author	Number of Taxa
1	2004	Wojciechowski M.F	262/330 taxa*
2	2004	Allan G.J	52 taxa
3	2004	McMahon M.	240 taxa
4	2004	Pardo C.	78 taxa
5	2004	Ree R.	15 taxa
6	2003	Ainoche A.	34 taxa
7	2003	Allan G.J	55 taxa
8	2003	Crisp M.D.	66 taxa
9	2003	Dong T.X.X	10 taxa
10	2003	Kang Y.	56 taxa
11	2003	Lavin M.	12 taxa
12	2003	Schrire B.D.	109 taxa
13	2003	Steele K.P.	84 taxa
14	2002	Badr A.	37 taxa
15	2002	Cubas P.	57 taxa
16	2002	Doi K.	23 taxa
17	2002	Hu J-M	42 taxa
18	2002	Mayer	12 taxa
19	2002	Percy D.M.	50 taxa
20	2001	Bena G.	77 taxa
21	2001	Chandler G.T.	57 taxa
22	2001	Lavin M.	61 taxa
23	2001	Lavin M.	95 taxa
24	2001	Pennington R.T.	122 taxa
25	2000	Allan G.J.	42 taxa
26	2000	Crisp M.D.	99 taxa
27	2000	Murphy D.J.	19 taxa
28	1999	Ainoche A-K	49 taxa
29	1999	Delgado-Salinas A.	132 taxa
30	1999	Wagstaff S.J.	39 taxa
31	1999	Wojciechowski M.F.	115 taxa
32	1998	Asmussen C.B.	42 taxa
33	1998	Bena G.	13 taxa
34	1998	Downie S.R.	62 taxa
35	1998	Fennel S.R.	10 taxa
36	1998	Lavin M.	34 taxa
37	1997	van Oss H.	8 taxa
38	1996	Sanderson M.J.	41 taxa
39	1995	Pennington R.T	27 taxa
40	1994	Liston A.	51 taxa
41	1993	Bruneau A.	66 taxa
42	1993	Doyle J.J.	53 taxa

43	1993	Sanderson M.J.	33 taxa
44	1992	Liston A.	64 taxa

*Number of Papilionoideae taxa/total number of taxa

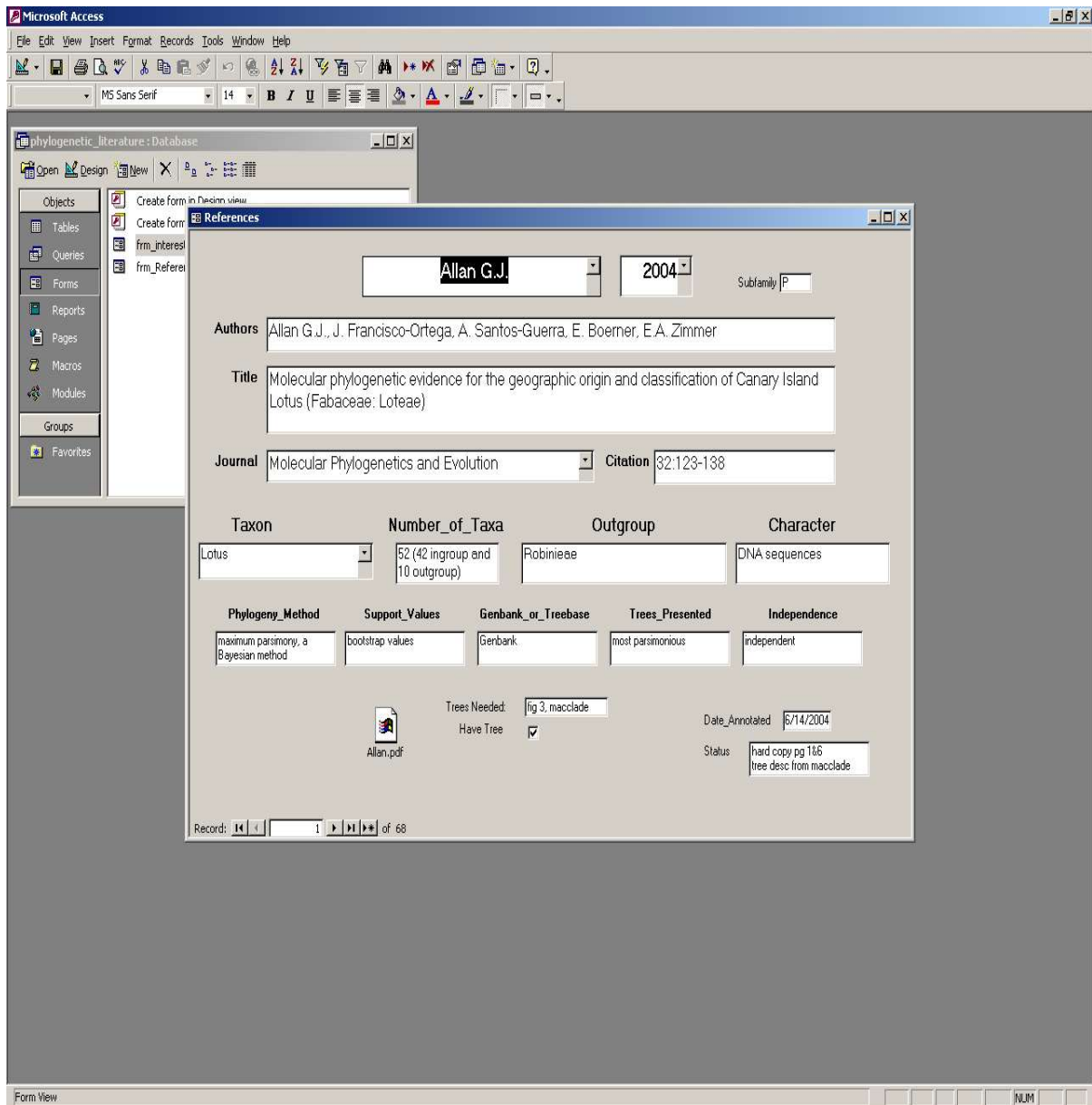


Figure 1. Screen shot of Access Database used to store information collected from phylogenetic legume studies published in the primary literature.

Figure Captions

Figure 2.

Supertree of Mimosoideae subfamily based on MRP analysis of supermatrix of 4 source trees. Supermatrix contained 216 taxa and 429 characters. Tree shown is a 90% majority rule consensus of 5000 equally parsimonious trees derived from a heuristic search of the supermatrix using the closest addition sequence procedure and TBR branch swapping algorithm holding 5 trees at each step.

Figure 3.

Supertree of Caesalpinioideae subfamily based on MRP analysis of supermatrix of 8 source trees. Supermatrix contained 343 taxa and 605 characters. Tree shown is a 90% majority rule consensus of 5000 equally parsimonious trees derived from a heuristic search of the supermatrix using the closest addition sequence procedure and TBR branch swapping algorithm holding 5 trees at each step.

Figure 4.

Supertree of Papilionoideae subfamily based on MRP analysis of supermatrix of 44 source trees. Supermatrix contained 1797 taxa and 2045 characters. Tree shown is a 90% majority rule consensus of 5000 equally parsimonious trees derived from a heuristic search of the supermatrix using the closest addition sequence procedure and TBR branch swapping algorithm holding 5 trees at each step.

Figure 2.



Figure 2.



Figure 3.



Figure 3.



Figure 4.



Figure 4.



Figure 4.



Figure 4.



Figure 4.



Figure 4.



Figure 4.



Figure 4.



Figure 4.



Figure 4.



Figure 4.



Figure 4.



Figure 4.



Figure 4.



References

- Bininda-Emonds, O. R. P. 2004. The evolution of supertrees. *Trends in Ecology and Evolution* 19:315-322.
- Bininda-Emonds, O. R.P., K.E. Jones, S.A. Price, M. Cardillo, R.Grenyer, and A. Purvis. 2004. Garbage in, garbage out: data issues in supertree construction. Chapter 12 in *Phylogenetic Supertrees: Combining Information to Reveal the Tree of Life. Computational Biology* 3:267-280.
- Bininda-Emonds, O. R.P. 2002. The (Super) Tree of Life: Procedures, Problems, and Prospects. *Annual Review of Ecological Systems* 33:265-289.
- Daniel, P., and C. Semple. 2004. Supertree Algorithms for Nested Taxa. Chapter 7 in *Phylogenetic Supertrees: Combining Information to Reveal the Tree of Life. Computational Biology* 3:151-171.
- Felsenstein, J. 2004. *Inferring Phylogenies*. Sinauer Associates, Sunderland, Massachusetts.
- Liu, F-G-R. et al. 2001. Molecular and morphological supertrees for eutherian (placental) mammals. *Science* 291: 1786-1789.
- Maddison, D. R., D. L. Swofford, and W. P. Maddison. 1997. NEXUS: an extensible file format for systematic information. *Systematic Biology* 46: 590-621.
- Page, Roderic D.M. 1998. *Molecular Evolution: A Phylogenetic Approach*. Blackwell Science, Malden, MA.
- Page, Roderic D.M. Taxonomy, Supertrees, and the Tree of Life. 2004. Chapter 11 in *Phylogenetic Supertrees: Combining Information to Reveal the Tree of Life. Computational Biology* 3: 247-265.
- Pennisi, E. 2003. Modernizing the Tree of Life. *Science* 300: 1692-1697.
- Salamin, N., et al. 2002. Building Supertrees: An Empirical Assessment Using the Grass Family (Poaceae). *Systematic Biology*. 51:136-150.
- Sanderson, M. J. 2003. r8s, version 1.6 (User's Manual, April 2003). Distributed by the author (<http://ginger.ucdavis.edu/r8s/>). University of California, Davis.
- Sanderson, M. J., A. Purvis, and C. Henze. 1998. Phylogenetic supertrees: assembling the trees of life. *Trends in Ecology and Evolution* 13:105-109.
- Sanderson, M. J., A. C. Driskell, R. H. Ree, O. Eulenstein, and S. Langlely. 2003. Obtaining maximal concatenation. *Figure 4. Continued from figure 2 and 3* *Journal of Molecular Evolution* 57: 1-11.
- Wojciechowski, M. F. 2003. The phylogeny of the ' *Journal of Molecular Evolution* 57: 1-11.

approach. In *Advances in Legume Systematics*, part 9, pp. 277-298. P. Herendeen and A. Bruneau, eds. Royal Botanic Garden, Kew.

References for Studies included in Supertree Analysis

- Ainoche A-K and R.J. Bayer. 1999. Phylogenetic relationships in *Lupinus* (Fabaceae: Papilionoideae) based on internal transcribed spacer sequences (ITS) of nuclear ribosomal DNA. *American Journal of Botany*. 86:590-607.
- Ainoche A., et al. 2003. Phylogenetic Relationships Within Tribe Genisteae (Papilionoideae) with Special Reference to Genus *Ulex*. *Advances in Legume Systematics*. B.B. Klitgaard and A. Bruneau eds., Part 10: Higher Level Systematics. 239-252. The Royal Botanic Gardens, Kew.
- Allan G.J. and J.M. Porter. 2000. Tribal delimitation and phylogenetic relationships of Loteae and Coronilleae (Faboideae: Fabaceae) with special reference to *Lotus*: evidence from nuclear ribosomal ITS sequences. *American Journal of Botany*. 87:1871-1881.
- Allan G.J., et al. 2003. Molecular Phylogenetic Analyses of Tribe Loteae (Leguminosae): Implications for Classification and Biogeography. *Advances in Legume Systematics*. B.B. Klitgaard and A. Bruneau eds., Part 10: Higher Level Systematics. 371-393. The Royal Botanic Gardens, Kew.
- Allan G. J., et al. 2004. Molecular phylogenetic evidence for the geographic origin and classification of Canary Island *Lotus* (Fabaceae: Loteae). *Molecular Phylogenetics and Evolution*. 32:123-138.
- Asmussen C.B. and A. Liston. 1998. Chloroplast DNA characters, phylogeny, and classification of *Lathyrus* (Fabaceae). *American Journal of Botany*. 85:387-401.
- Badr A., et al. 2002. Systematic relationships in Lathyruseae. *Lathyrus* (Fabaceae) based on amplified fragment length polymorphism (AFLP) data. *Canadian Journal of Botany*. 80:962-969.
- Barker N.P., B.D. Schrire, and J-H Kim. 2000. Generic Relationships in the Tribe Indigofereae (Leguminosae: Papilionoideae) Based on Sequence Data and Morphology. *Advances in Legume Systematics*. Herendeen P.S. and A. Bruneau ed., Part 9. 311-337. The Royal Botanic Gardens, Kew.
- Bena G., et al. 1998. Ribosomal External and Internal Transcribed Spacers: Combined Use in the Phylogenetic Analysis of *Medicago* (Leguminosae). *Journal of Molecular Evolution*. 46:299-306.
- Bena G. 2001. Molecular phylogeny supports the morphologically based taxonomic transfer of the "medicagoid" *Trigonella* species to the genus *Medicago* L. *Plant Systematics and Evolution*. 229:217-236.

- Brouat C., L. Gielly, and D. McKey. 2001. Phylogenetic relationships in the genus *Leonardoxa* (Leguminosae: Caesalpinioideae) inferred from chloroplast trnL intron and trnL-trnF intergenic spacer sequences. *American Journal of Botany*. 88:143-149.
- Bruneau A. 1996. Phylogenetic and biogeographical patterns in *Erythrina* (Leguminosae: Phaseoleae) as inferred from morphological and chloroplast DNA characters. *Systematic Botany*. 21:587-605.
- Bruneau A., et al. 2000. Phylogenetic relationships in tribes Macrolobieae and Detarieae as inferred from chloroplast trnL intron sequences. *Advances in Legume Systematics*. P. Herendeen and A. Bruneau eds., Part 9. 121-149, The Royal Botanic Gardens, Kew.
- Bruneau A. and J.J. Doyle. 1993. Cladistic-Analysis of Chloroplast DNA Restriction Site characters in *Erythrina* (Leguminosae, Phaseoleae). *Systematic Botany*. 18:229-247.
- Bruneau A., J.L. Doyle, and J.J. Doyle. 1995. Phylogenetic evidence in Phaseoleae: evidence from chloroplast restriction site character. *Advances in Legume Systematics*. M. D. Crisp, J.J. Doyle eds., Part 7: Phylogeny. 309-330, The Royal Botanic Gardens, Kew.
- Chandler G.T., R.J. Bayer, and M.D. Crisp. 2001. A molecular phylogeny of the endemic Australian genus *Gastrolobium* (Fabaceae: Mirbelieae) and allied genera using chloroplast and nuclear markers. *American Journal of Botany*. 88:1675-1687.
- Chappill J.A. 1995. Cladistic analysis of the Leguminosae: the development of an explicit phylogenetic hypothesis. *Advances in Legume Systematics*. M.D. Crisp and J.J. Doyle eds., Part 7: Phylogeny. 1-9, The Royal Botanic Gardens, Kew.
- Clarke H.D., S.R. Downie, and D.S. Seigler. 2000. Implications of chloroplast DNA restriction site variation for systematics of *Acacia* (Fabaceae: Mimosoideae). *Systematic Botany*. 25:618-632.
- Crisp M.D. and L.G. Cook. 2003. Phylogeny and Embryo Sac Evolution in the Endemic Australasian Papilionoid Tribes Mirbelieae and Bossiaeeae. *Advances in Legume Systematics*. B.B. Klitgaard and A. Bruneau eds., Part 10: Higher Level Systematics. 253-268. The Royal Botanic Gardens, Kew.
- Crisp M.D., S. Gilmore, and B-E. van Wyk. 2000. Molecular phylogeny of the genistoid tribes of papilionoid legumes. *Advances in Legume Systematics*. P. Herendeen and A. Bruneau eds., Part 9. 249-276, The Royal Botanic Gardens, Kew.
- Cubas P., C. Pardo, and H. Tahiri. 2002. Molecular approach to the phylogeny and systematics of *Cytisus* (Leguminosae) and related genera based on nucleotide sequences of nrDNA (ITS region) and cpDNA (trnL-trnF intergenic spacer). *Plant Systematics and Evolution*. 233:223-242.
- Davis C.C., et al. 2002. Phylogeny and Biogeography of *Cercis* (Fabaceae): Evidence from Nuclear Ribosomal ITS and Chloroplast ndhF Sequence Data. *Systematic Botany*. 27:289-302.

- Delgado-Salinas A., et al. 1999. Phylogenetic analysis of cultivated and wild species of Phaseolus (Fabaceae). *Systematic Botany*. 24:438-460.
- Doi K., et al. 2002. Molecular phylogeny of genus Vigna subgenus Ceratotropis based on rDNA ITS and atpB-rbcL intergenic spacer of cpDNA sequences. *Genetica*. 114:129-145.
- Dong T.X.X., et al. 2003. Phylogeny of Astragalus in China: Molecular Evidence from the DNA Sequences of 5S rRNA Spacer, ITS, and 18S rRNA. *Journal of Agricultural and Food Chemistry*. 51:6709-6714.
- Downie S.R., et al. 1998. Multiple independent losses of the plastid rpoC1 intron in Medicago (Fabaceae) as inferred from phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacer sequences. *Canadian Journal of Botany*. 76:791-803.
- Doyle J.J. and J.L. Doyle. 1993. Chloroplast DNA phylogeny of the papilionoid legume tribe Phaseoleae. *Systematic Botany*. 18:309-327.
- Fennel S.R., et al. 1998. Phylogenetic relationships between Vicia faba (Fabaceae) and related species inferred from chloroplast trnL sequences. *Plant Systematics and Evolution*. 212:247-259.
- Gervais G.Y.F. and A. Bruneau. 2002. Phylogenetic analysis of a polyphyletic African genus of Caesalpinioideae (Leguminosae): Monopetalanthus Harms. *Plant Systematics and Evolution*. 235:19-34.
- Haston E.M., G.P. Lewis, and J.A. Hawkins. 2003. A Phylogenetic Investigation of the Peltophorum Group (Caesalpinieae: Leguminosae). *Advances in Legume Systematics*. B.B. Klitgaard and A. Bruneau eds., Part 10: Higher Level Systematics. 149-159. The Royal Botanic Gardens, Kew.
- Herendeen P.S., et al. 2003. Phylogenetic Relationships in Caesalpinoid Legumes: A Preliminary Analysis Based on Morphological and Molecular Data. *Advances in Legume Systematics*. B.B. Klitgaard and A. Bruneau eds., Part 10: Higher Level Systematics. 37-62. The Royal Botanic Gardens, Kew.
- Hu J-M, et al. 2002. Phylogenetic analysis of nuclear ribosomal ITS/5.8 S sequences in the tribe Millettieae (Fabaceae): Poecilanthe-Cyclolobium, the core Millettieae, and the Callerya group. *Systematic Botany*. 27:722-733.
- Hughes C.E., et al. 2003. Relationships Among Genera of the Informal Dichrostachys and Leucaena Groups (Mimosoideae) Inferred from Nuclear Ribosomal ITS Sequences. *Advances in Legume Systematics*. B.B. Klitgaard and A. Bruneau eds., Part 10: Higher Level Systematics. 221-238. The Royal Botanic Gardens, Kew.
- Kajita T., H. Ohashi, Y. Tateishi, C.D. Bailey, and J.J. Doyle. 2001. rbcL and legume phylogeny, with particular reference to Phaseoleae, Millettieae, and Allies. *Systematic Botany*. 26:515-536.

- Kang Y. et al. 2003. A Preliminary Phylogenetic Study of the Subgenus Pogonophace (Astragalus) in China Based on ITS Sequence Data. *Acta Botanica Sinica*. 45:140- 145.
- Lavin M. et al. 2001. Identifying Tertiary radiations of Fabaceae in the Greater Antilles: alternatives to cladistic vicariance analysis. *International Journal of Plant Sciences*. 162:S53-S76.
- Lavin M. et al. 2001. The dalbergioid legumes (Fabaceae): delimitation of a pantropical monophyletic clade. *American Journal of Botany*. 88:412-434.
- Lavin M. et al. 2003. Phylogeny of Robinoid Legumes (Fabaceae) Revisited: Coursetia and Gliricidia Recircumscribed, and a Biogeographical Appraisal of the Caribbean Endemics. *Systematics Botany*. 28:387-409.
- Lavin M. E. Eshbaugh, J-M Hu, S. Mathews, and R.A. Sharrock. 1998. Monophyletic subgroups of the tribe Millettieae (Leguminosae) as revealed by phytochrome nucleotide sequence data. *American Journal of Botany*. 85:412-433.
- Liston A. 1992. Variation in the chloroplast gene rpoC1 and rpoC2 of the genus *Astragalus* (Fabaceae): evidence from restriction site mapping of a PCR-amplified fragment. *American Journal of Botany*. 79:953-961.
- Liston A. and A.J. Wheeler. 1994. The phylogenetic position of the genus *Astragalus* (Fabaceae): Evidence from the chloroplast genes rpoC1 and rpoC2. *Biochemical Systematics and Ecology*. 22:377-388.
- Luckow M. 1995. A Phylogenetic Analysis of the Dichrostachys Group (Mimosoideae: Mimoseae). *Advances in Legume Systematics*. M.D. Crisp and J. J. Doyle eds., Part 7: Phylogeny. 63-76, The Royal Botanic Gardens, Kew.
- Luckow M., et al. 2003. A Phylogenetic Analysis of the Mimosoideae (Leguminosae) Based on Chloroplast DNA Sequence Data. *Advances in Legume Systematics*. B.B. Klitgaard and A. Bruneau eds., Part 10: Higher Level Systematics. 197-220. The Royal Botanic Gardens, Kew.
- Luckow M. and F.C.H. Hopkins. 1995. A cladistic analysis of *Parkia* (Leguminosae: Mimosoideae). *American Journal of Botany*. 82:1300-1320.
- Mayer M.S., and S.K. Bagga. 2002. The phylogeny of *Lens* (Leguminosae): new insight from ITS sequence analysis. *Plant Systematics and Evolution*. 232:145-154.
- McMahon M. and L. Hufford. 2004. Phylogeny of Amorpheae (Fabaceae: Papilionoideae). *American Journal of Botany*. 9:1219-1230.
- Miller J.T., et al. 2003. A Phylogenetic Analysis of the Acacieae and Ingeae (Mimosoideae: Fabaceae) based on trnK, matK, psbA-trnH, and trnL/trnF Sequence Data. *Systematic Botany*. 28:558-566.

- Murphy D.J., F. Udovicic, and P.Y. Ladiges. 2000. Phylogenetic analysis of Australian Acacia (Leguminosae : Mimosoideae) by using sequence variations of an intron and two intergenic spacers of chloroplast DNA. *Australian Systematic Botany*. 13:745-754.
- Osaloo S.K., A.A. Maassoumi, and N. Murakami. 2003. Molecular Systematics of the genus Astragalus L. (Fabaceae): Phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacers and chloroplast gene ndhF sequences. *Plant Systematics and Evolution*. 242:1-32.
- Pardo C. et al. 2004. Molecular Phylogeny and systematics of Genista (Leguminosae) and related genera based on nucleotide sequences of nrDNA (ITS region) and cpDNA (trnL-trnF intergenic spacer). *Plant Systematics and Evolution*. 244:93-119.
- Pennington R.T. 1995. Cladistic-Analysis of Chloroplast DNA Restriction Site characters in Andira (Leguminosae, Dalbergieae). *American Journal of Botany*. 82: 526-534.
- Pennington R.T., et al. 2001. Phylogenetic relationships of basal papilionoid legumes based upon squences of the chloroplast trnL intron. *Systematic Botany*. 26:537-556.
- Percy D.M., et al. 2002. Different Fates of Island Brooms: Contrasting Evolution in Adenocarpus, Genista, and Teline (Genisteae, Fabaceae) in the Canary Islands and Madeira. *American Journal of Botany*. 89:854-864.
- Ree R. et al. 2004. Heterogeneous Selection on LEGCYC Paralogs in Relation to Flower Morphology and the Phylogeny of Lupinus (Leguminosae). *Molecular Biology and Evolution*. 21:321-331.
- Richardson J.E., et al. 2001. Rapid Diversification of a Species-Rich Genus of Neotropical Rain Forest Trees. *Science*. 293:2242-2245.
- Sanderson M.J. and J.J. Doyle. 1993. Phylogenetic relationships in North American Astragalus (Fabaceae) based on chloroplast DNA restriction site variation. *Systematic Botany*. 18:395-408.
- Sanderson M.J. and M.F. Wojciechowski. 1996. Diversification rates in a temperate legume clade: Are there "so many species" of Astragalus (Fabaceae)? *American Journal of Botany*. 83:1488-1502.
- Schnabel A., et al. 2003. Phylogenetic relationships in Gleditsia (Leguminosae) based on ITS sequences. *American Journal of Botany*. 90:310-320.
- Schnabel A. and J.F. Wendell. 1998. Cladistic biogeography of Gleditsia (Leguminosae) based on ndhF and rpl16 chloroplast gene sequences. *American Journal of Botany*. 85:1753-1765.
- Schrire B.D., et al. 2003. Towards a Phylogeny of Indigofera (Leguminosae-Papilionoideae): Identification of Major Clades and Relative Ages. *Advances in Legume Systematics*. B.B.

- Klitgaard and A. Bruneau eds., Part 10: Higher Level Systematics. 269-302. The Royal Botanic Gardens, Kew.
- Simpson B.B., et al. 2003. Progress Towards Resolving the Relationships of the Caesalpinia Group (Caesalpinieae: Caesalpinioideae: Leguminosae). *Advances in Legume Systematics*. B.B. Klitgaard and A. Bruneau eds., Part 10: Higher Level Systematics. 123-148. The Royal Botanic Gardens, Kew.
- Steele K.P. and M.F. Wojciechowski. 2003. Phylogenetic Analyses of Tribes Trifolieae and Vicieae, Based on Sequences of the Plastid Gene matK (Papilionoideae: Leguminosae). *Advances in Legume Systematics*. B.B. Klitgaard and A. Bruneau eds., Part 10. 355-370. The Royal Botanic Gardens, Kew.
- van der Bank M., et al. 2002. Systematics of the tribe Podalyrieae (Fabaceae) based on DNA, morphological and chemical data. *Botanical Society of the Linnean Society*. 139:159-170.
- van Oss H., Y. Aron, and G. Ladizinsky. 1997. Chloroplast DNA variation and evolution in the genus *Lens* Mill. *Theoretical and Applied Genetics*. 94:452-457.
- Wagstaff S.J., P.B. Heenan, and M.J. Sanderson. 1999. Classification, origins, and patterns of diversification in New Zealand Carmichaelinae (Fabaceae). *American Journal of Botany*. 86:1346-1356.
- van Wyk B.E. and A.L. Schutte. 1995. Phylogenetic Relationships in the Tribes Podalyrieae, Liparieae and Crotalarieae. *Advances in Legume Systematics*. M.D. Crisp and J. J. Doyle eds., Part 7: Phylogeny. 283-308, The Royal Botanic Gardens, Kew.
- Wojciechowski, M. F., M. Lavin, and M. J. Sanderson. 2004. A phylogeny of legumes (Leguminosae) based on analysis of the plastid matK gene resolves many well supported subclades within the family. *American J. Botany*, 91:1846-1862.
- Wojciechowski M.F., M.J. Sanderson, and J-M. Hu. 1999. Evidence of the monophyly of *Astragalus* (Fabaceae) and its major subgroups based on nuclear ribosomal DNA ITS and chloroplast DNA trnL intron data. *Systematic Botany*. 24:409-437.