

**ASSEMBLING THE TREE OF LIFE TO ENABLE THE PLANT SCIENCES (iPTOL):  
A PROPOSAL FOR AN iPLANT GRAND CHALLENGE WORKSHOP**

*Organizers:*

**Michael J. Donoghue**, Department of Ecology and Evolutionary Biology and Peabody Museum of Natural History, Yale University; 203-432-2074 (phone), 203-432-7907 (fax), [michael.donoghue@yale.edu](mailto:michael.donoghue@yale.edu) (primary contact for this proposal): Donoghue's research focuses on the uses of phylogenies in understanding morphological and molecular evolution, ecology, and biogeography. As Director of the Peabody Museum he has managed numerous K-12 outreach programs, including a museum exhibition entitled *Travels in the Great Tree of Life*.

**Michael J. Sanderson**, Department of Ecology and Evolutionary Biology, University of Arizona; 520-626-6848 (phone), 520-621-9190 (fax), [sanderm@email.arizona.edu](mailto:sanderm@email.arizona.edu): Sanderson's research emphasizes method and algorithm development for data mining and large scale phylogenetic tree construction.

**Douglas E. Soltis**, Department of Botany, University of Florida; 352-273-1963 (phone), 352-846-2154 (fax), [dsoltis@botany.ufl.edu](mailto:dsoltis@botany.ufl.edu): Douglas Soltis's research focuses on angiosperm phylogeny, especially the Saxifragales, and the genetic and genomic consequences of polyploidy.

**Pamela S. Soltis**, Florida Museum of Natural History, University of Florida; 352-273-1964 (phone), 352-846-2154 (fax), [psoltis@flmnh.ufl.edu](mailto:psoltis@flmnh.ufl.edu): Pamela Soltis's research interests are in angiosperm phylogeny, the origin and evolution of the floral genetic program, and the evolution by polyploidy.

**Val Tannen**, Department of Computer and Information Science, University of Pennsylvania; 215-898-2665 (phone), 215-898-0587 (fax), [val@cis.upenn.edu](mailto:val@cis.upenn.edu): Tannen's research is on foundations of data management with applications to phyloinformatics; he leads the NSF AToL project "Processing Phylodata" (pPOD) as well as the database focus for the NSF CIPRES project.

**Todd J. Vision**, Department of Biology, University of North Carolina at Chapel Hill; 919-843-4507 (phone), 919-962-1625 (fax), [tjv@bio.unc.edu](mailto:tjv@bio.unc.edu): Vision's research focuses on computational and informatic applications to evolutionary genetics and genome evolution, with an emphasis on the macroevolution of gene order and gene content in the angiosperms.

*Potential participants in this GCW (\* = have indicated their interest in participating)*

R. Beaman\* (FL, biodiversity informatics), L. Bohs (UT, Solanaceae NSF PBI), G. Burleigh\* (NC, large trees) S. Cannon\* (IA, genome diversification), L. Comai\* (CA, polyploidy), P. Crane (IL, paleobotany), J. Doyle\* (CA, paleobotany), W. Friedman\* (CO, morphology), T. Givnish\* (WI, ecology, monocot AToL), S. Graham (Canada, barcoding), D. Hibbett\* (MA, fungal informatics), K. Huber (Sweden, reticulate phylogenies), R. Jansen\* (TX, cpDNA), W. Judd\* (FL, morphological systematics), E. Kramer (MA, evo-devo), J. Leebens-Mack\* (GA, genome pipelines, monocot AToL), P. Lockhart (New Zealand, plant phylogenetics), D. Maddison (AZ, TOLweb, MacClade, Mesquite), S. Magallon\* (Mexico, paleobotany), S. Mathews\* (MA, gymnosperms), B. Mishler\* (CA, mosses), B. Moret\* (Switzerland, CIPRES), E. Mossel\* (CA, comp. sci., large trees), T. Munzner (Canada, visualization), R. Olmstead\* (WA, asterids), J. Palmer\* (IN, mol. biol.), W. Piel\* (CT, TreeBASE), K. Pryer\* (NC, monilophytes), Y. Qiu (MI, large mtDNA trees), S. Rhee (CA, genome informatics), L. Salter-Kubatko\* (OH, comp. sci., tree building), V. Savolainen\* (UK, comparative studies), J. Shaw\* (NC, liverworts), S. Smith\* (NC, data mining, large trees), M. Steel\* (New Zealand, computer science), P. Stevens\* (MO, APWeb), D. Swofford (NC, phylogeny software), D. Tank\* (ID, large trees), G. Theissen (Germany, MADS, evo-devo), C. Webb (MA, Phylomatic, Phylocom, ecology), G. Weiblen\* (MN, coevolution), R. Zechman\* (CA, green algae). *Summary: 21 US States, 7 Countries, 9 women.*

## SUMMARY OF THE PROPOSED WORKSHOP

Knowledge of the phylogeny of green plants has expanded very dramatically over the past two decades, yet the task in front of us remains enormous. In the meantime, knowledge of evolutionary relationships is yielding new insights across the plant sciences, from comparative genomics, to adaptation, speciation, community assembly, and ecosystem function. It has become clear, for example, that phylogenetic trees provide a rigorous quantitative framework for testing comparative hypotheses in structural, functional, and developmental biology. And, beyond elucidating species relationships, phylogenetic tools are proving to be crucial in unraveling population dynamics and the history of gene families and protein evolution, to name only a few productive extensions.

The rate of accumulation of the data relevant to phylogenetic biology has far exceeded all expectations, and we now find ourselves awash in unconsolidated information that could shed light on the history of plants. We are, therefore, confronted with a major irony – just as we are learning to confidently infer phylogenetic relationships and to use trees to orient comparative studies, we are confronting major barriers in managing and synthesizing the relevant data, in visualization, and in the development of tools to put this exploding knowledge-base to work. This conundrum presents a Grand Challenge whose solution clearly requires dramatic expansion of a developing intersection between phylogenetic biology and the computer sciences, broadly conceived, with an eye toward integrating and synthesizing data and facilitating research and education across the plant sciences.

We propose to bring a group ca. 35 relevant plant and computer scientists together to conceive a Grand Challenge Project focused on assembling all knowledge of the phylogeny of green plants, to make this readily accessible, and to integrate it throughout the botanical sciences. More specifically, we will envision "discovery environments" that would allow the mining and synthesis of all of the relevant literature, the underlying molecular and morphological data, and all associated metadata, and within which this information could be visualized, broadly disseminated, and utilized in brand new ways. Key objectives will be the development of new tools to streamline the assembly of datasets to carry out sophisticated, and in many cases, massive analyses, and the integration of phylogenetic knowledge into studies involving data from other plant science domains. We also will explore strategies for insuring the maintenance, expansion, and broadest use of the cyberinfrastructure that is developed.

This is just the right time for this effort. Beyond a burgeoning print literature, most of the hard-won information on phylogeny is not readily accessible to the research and education communities now, and the problem will get much worse unless we address it soon and succeed in developing the right cyberinfrastructure. The necessary developments are so broad and cross-disciplinary that they are highly unlikely to materialize through normal NSF funding. A concerted iPlant effort in this area is not only necessary, but is also highly likely to yield successes, as the plant phylogenetics community has an established history of broad-scale collaboration and has consistently been willing to adopt new cyber-solutions. This area is also rich in problems of direct interest to computer scientists and bioinformaticians, whose solutions would be of very general use. Furthermore, a variety of these problems have begun to be addressed through other major efforts, including the NSF's CIPRES and ATOL programs. Most importantly, early success in addressing the plant phylogeny problem would enable developments throughout the plant sciences, and would be especially useful in connection with other Grand Challenge Projects supported through the iPlant Collaborative. In the process, a phylogeny-focused project would promote the integration of plant evolutionary biologists throughout the iPlant community, and, in doing so, would insure the breadth and broad impact of the project as a whole. Finally, the Plant Tree of Life provides exciting opportunities for education and outreach at all levels.

## BIOGRAPHICAL SKETCHES

### MICHAEL J. DONOGHUE

Department of Ecology and Evolutionary Biology  
& Peabody Museum of Natural History  
Yale University, New Haven, CT 06520-8105  
Phone: 203-432-2074; michael.donoghue@yale.edu

#### PROFESSIONAL PREPARATION:

Michigan State University, East Lansing, MI	Botany	B.S. (honors), 1976
Harvard University, Cambridge, MA	Biology	Ph.D., 1982

#### APPOINTMENTS:

**2000-present** G. E. Hutchinson Professor of Ecology and Evolutionary Biology, with appointments in the Department of Geology and Geophysics, in the School of Forestry and Environmental Studies, and as Curator of Botany in the Peabody Museum. Chair, Department of Ecology and Evolutionary Biology (2001-02); Director, Peabody Museum of Natural History (2003-2008).

**1993-2000** Professor of Organismic and Evolutionary Biology, Harvard University, and Director of the Harvard University Herbaria (1995-99).

**1998-1999** Visiting Professor, Department of Biology, Stanford University.

**1985-1993** Assistant Professor (1985-88), Associate Professor (1988-90), and Professor (1990-92) of Ecology and Evolutionary Biology, University of Arizona.

**1982-1985** Assistant Professor of Biology, San Diego State University.

#### FIVE RELEVANT PUBLICATIONS (of 183):

Donoghue, M. J. 2004. Immeasurable Progress on the Tree of Life. Pp. 548-552 in Cracraft, J. and M. J. Donoghue (eds.), *Assembling the Tree of Life*. Oxford University Press, New York.

Donoghue, M. J. 2005. Key innovations, convergence, and success: macroevolutionary lessons from plant phylogeny. *Paleobiology* 31(2): 77-93.

Moore, B. R., S. A. Smith, and M. J. Donoghue. 2006. Increasing data transparency and estimating phylogenetic uncertainty in supertrees: approaches using nonparametric bootstrapping. *Syst. Biol.* 55: 662-676.

Howarth, D. G. and M. J. Donoghue. 2006. Phylogenetic analyses of the "ECE" (CYC/TB1) clade reveal duplications that predate the core eudicots. *Proc. Nat. Acad. Sci. USA* 103: 9101-9106.

Smith, S. A. and M. J. Donoghue. Mega-phylogenies link rates of molecular evolution to life history in flowering plants. *Science* (in review).

#### FIVE ADDITIONAL RELEVANT PUBLICATIONS:

Piel, W. H., M. J. Sanderson, and M. J. Donoghue. 2003. The small-world dynamics of tree networks and data mining in phyloinformatics. *Bioinformatics* 19: 1162-1168.

Webb, C. O. and M. J. Donoghue. 2004. Phylomatic: tree assembly for applied phylogenetics. *Molecular Ecology Notes* 5: 181-183

Moore, B. R. and M. J. Donoghue. 2007. Correlates of diversification in the plant clade Dipsacales: Geographic movement and evolutionary innovations. *Amer. Nat.* S28-S55.

Cantino, P. D., J. A. Doyle, S. W. Graham, W. S. Judd, R. G. Olmstead, P. S. Soltis, D. E. Soltis, and M. J. Donoghue. 2007. Towards a phylogenetic nomenclature of *Tracheophyta*. *Taxon* 56: 822-846.

Edwards, E. J., C. J. Still, and M. J. Donoghue. 2007. The relevance of phylogeny to studies of global change. *Trends Ecol. Evol.* 22: 243-249.

## RECENT SYNERGISTIC ACTIVITIES:

**Selected Service, 1995-03:** President, Society of Systematic Biologists (1994-95); Steering Committee, "Evolution, Science, and Society" (1995-97); DIVERSITAS Steering Committee (2001- ), Vice Chair (2002- ), Chair, bioGENESIS (2006); NAS Committee, Int. Union Biol. Sci. (1999-04); NCEAS Working Groups: Biogeography (2000-02), Adaptive Radiations (2001-02); Phylogeny and Ecology (2002); Executive Committee, Discovering Life in America (2001- ); Visiting Committees: Princeton Univ. (1995, 2002), SUNY Stony Brook (1997), UC Berkeley (1998, 2001), Cornell (1999), Arnold Arboretum (2001, 2003), Brown Univ. (2002); Board of Directors, Natural Science Collections Alliance (2004- ).

**Selected Honors, 1997-02:** Fellow, AAAS (1997); Glaser Distinguished Visiting Prof., Florida Int. Univ. (1998); Eminent Biologist Lecturer, Carnegie Museum (1999), Perspectives in Biology Lecturer, Wake Forest Univ. (1999); Distinguished Alumni Award, Michigan State Univ. (2005); Elected, US National Academy of Sciences (2005); Elected, American Academy of Arts and Sciences (2008)

**Selected Invited Presentation, 2000-05:** "Frontiers in phylogenetic biology," Bot. Soc. America (2000); "Transference of function," FASEB, Vermont (2000, with D. Baum); "Phylogenies and global change," Amsterdam (2001); "Historical biogeography," European Soc. Evol. Biol., Denmark (2001); "Plant evolution," Cold Spring Harbor (2001); "Dipsacales flower evolution," Zurich (2002); "Evolution of Biomes," Royal Society, (2004); "Homology," Botanical Society of America (2004); "Biodiversity," Paris (2005).

**Additional Activities:** *Databases* -- TreeBASE: a database of phylogenetic knowledge (with W. Piel, M. Sanderson,); <http://www.treebase.org>. *Selected Symposia/Meetings Organized* -- "Northern Hemisphere Biogeography" (with J. Wen), International Botanical Congress, Vienna, Austria (2005); "Phytogeography of the Northern Hemisphere" (with P. Manos), NESCent Working Group (2006-08); "Phylogenies and Biodiversity Science," DIVERSITAS Open Science Conference, Oaxaca, Mexico (2005)

## COLLABORATORS AND OTHER AFFILIATIONS:

**Co-authors and co-editors (past 5 years, excluding students and postdocs – see below):** G. Burleigh (NESCent), N. Arens (Smith Coll.), C. Campbell (U. Maine), P. Cantino (Ohio U.), K. Chan (U. British Columbia), L. Clark (Iowa State), J. Cracraft (Am. Mus. Nat. Hist.), C. David (Harvard U.), C. Delwiche (U. Maryland), J. Doyle (UC Davis), T. Eriksson (Stockholm) T. Field (U. Tennessee), J. Gauthier (Yale), G. Gilbert (UC Santa Cruz), S. Graham (U. British Columbia), N. Havill (Yale U.), B. Jacobs (Leiden), C. Jaramillo (STRI), W. Judd (U. Florida), E. Kellogg (U. Missouri), B. LePage (U. Pennsylvania), M. Loreau (McGill U.), J. Lundberg (Stockholm), P. Manos (Duke U.), D. Miranker (U. Texas), S. Mathews (Arnold Arb.), L. Nakhleh (U. Texas), R. Olmstead (U. Washington), E. Smets (Leiden), M. Smith (Paris), D. Soltis (U. Florida), P. Soltis (U. Florida), P. Stevens (U. Missouri), N. Theis (U. Massachusetts), J. Wiens (Stony Brook), K. Wurdack (Smithsonian Inst.), T. Yahara (Kyushu U.), A. Yoder (Duke U.).

**Graduate Students (21 total):** J. Dice (CalTrans), M. Sanderson (U. Arizona), J. Malusa (Discov. Mag.), L. Abbott, J. M. Porter (Rancho Santa Ana Gard.), G. Bharathan (SUNY Stony Brook), D. Ferguson (Louisiana State U.), C. Maley (Wistar Inst.), G. Weiblen (U. Minnesota), R. Spangler (U. Minnesota), R. Ree (Field Mus.), C. Davis (Harvard U.), C. Bell (U. New Orleans), E. Edwards (Brown), S-T. Kim (Tubingen, Germany), B. Moore (UC Berkeley). Current: R. Novick, S. Smith, S. Carlson, J. Beaulieu, A. Greenberg

**Recent Postdoctoral Associates (24 total):** L. Schulthies (Cal. State, Hayward), C. Webb (Arnold Arboretum), R. Winkworth (Fiji), D. Howarth (St. Johns U.). Current: M. Evans, D. Tank

**Ph.D. advisor:** Carroll E. Wood, Jr., Harvard University

## MICHAEL J. SANDERSON

Professor, Department of Ecology and Evolutionary Biology  
University of Arizona, Tucson, AZ 85721  
E-mail: sanderm@email.arizona.edu

### Professional Preparation

- B. S., 1982, (cum laude, Physics) University of Arizona.  
Ph. D., 1989, (Ecology and Evolutionary Biology) University of Arizona.

### Appointments

- 2006 – Professor, Dept. of Ecology and Evolutionary Biology, University of Arizona  
2001 – 2006 Professor, Section of Evolution and Ecology, University of California, Davis  
1997 – 2001 Associate Professor, Section of Evolution and Ecology, University of California, Davis.  
1995 – 1997 Assistant Professor, Section of Evolution and Ecology, University of California, Davis.  
1992 – 1995 Assistant Professor, Department of Biology, University of Nevada, Reno.  
1992 – 1992 (one semester) Adjunct Assistant Professor, Department of Ecology and Evolutionary Biology, University of Arizona.  
1989 – 1991 Alfred P. Sloan Foundation Postdoctoral Fellow, L. H. Bailey Hortorium, Cornell University.

### Five Publications Relevant to Proposed Research (of 93)

- Sanderson, M. J. 2008. Phylogenetic signal in the eukaryotic tree of life. *Science*, in press.  
Sanderson, M. J., D. Boss, D. Chen, K. A. Cranston, and A. Wehe. 2008. The PhyLoTA Browser: processing GenBank for molecular phylogenetics research. *Syst. Biol.* In press.  
Sanderson, M. J., and M. M. McMahon. 2007. Inferring angiosperm phylogeny from EST data with widespread gene duplication. *BMC Evol. Biol.*, 7 (suppl. 1): S3.  
McMahon, M. M., and M. J. Sanderson. 2006. Phylogenetic Supermatrix Analysis of GenBank Sequences from 2228 Papilionoid Legumes. *Syst. Biol.*, 55:818-836.  
Driskell, A. C., C. Ané, J. G. Burleigh, M. M. McMahon, B. O'Meara, and M. J. Sanderson. 2004. Prospects for building the tree of life from large sequence databases. *Science*, 306:1172-1174.

### Five Other Publications

- Sanderson, M. J. Construction and annotation of large phylogenetic trees. 2007. *Australian Syst. Bot.* 4:287-301.  
Sanderson, M. J. 2006. Paloverde: an OpenGL 3-D phylogeny browser. *Bioinformatics*, 22:1004-1006.  
Burleigh, J. G., A. C. Driskell, and M. J. Sanderson. 2006. Supertree bootstrapping methods for assessing phylogenetic variation among genes in genome-scale data sets. *Syst. Biol.* 55:426-440.  
Ané, C., J. G. Burleigh, M. M. McMahon, and M. J. Sanderson. 2005. Covarion structure in plastid genome evolution: a new statistical test. *Mol. Biol. Evol.*, 22:914-924.  
Scotland, R., and M. J. Sanderson. 2004. The significance of few versus many in the tree of life. *Science*, 303:643.

### Synergistic Activities

- PhyLoTA Browser database: <http://loco.biosci.arizona.edu/cgi-bin/pb.cgi>.  
'r8s', vers. 1.71. Software for the analysis of molecular rates of evolution and the reconstruction of divergence times: <http://loco.biosci.arizona.edu/r8s>  
TreeBASE, an electronic database of phylogenetic knowledge: <http://www.treebase.org>.  
"Workshop in Applied Phylogenetics". Bodega Marine Lab, Bodega Bay California, Taught every year since 2000, and in 2004 with an appended workshop on supertree methodology.  
Organized an NSF-sponsored workshop: "Workshop on Phyloinformatics; October 20-21, 2000.

### **Collaborators and other Affiliations**

Graduate advisor; Michael J. Donoghue, Yale University

Postdoctoral advisor: Jeff J. Doyle, Cornell University

Graduate advisees and current addresses (7):

Michael Plotkin, UC Davis

Jer-Ming Hu, Assistant Professor, National Taiwan University

Shelah Morita, Postdoc, NC State University

Rosita Scherson, Postdoc, Universidad de Chile

Brian O'Meara, Postdoc, Duke U (NESCent)

Travis Wheeler, U of Arizona

Ed Gilbert, U of Arizona

Postdoctoral Advisees and current addresses (12):

Chris Henze, NASA Advanced Supercomputer Division, Ames Research Lab

Olaf Bininda-Emonds, Technical University of Munich

Susana Magallon, Ass't Prof., UNAM, Mexico City

Charles Nunn, Max Planck Institute, Leipzig

Amy Driskell, Smithsonian Institution

Mike Alfaro, Ass't Prof., Washington State

Rick Ree, Ass't Curator., Field Museum of Natural History

Gordon Burleigh, NESCENT Fellow, Duke

Campbell Webb, Arnold Arboretum, Harvard

Cecile Ane, Ass't Prof., University of Wisconsin

Justen Whittall, Ass't Prof., Santa Clara University

Karen Cranston, U of Arizona

David Hearn, U of Arizona

Brad Boyle, U of Arizona

Collaborators not listed above or in publications (no co-editors in last 24 months):

Boss, Darren, U of Arizona, Chen, D., Iowa State U, Eulenstein, O., Iowa State U, Fernandez-Baca, D., Iowa State U, Goloboff, P., INSUE, CONICET, Instituto Miguel Lillo, Argentina, Kim, J., U Pennsylvania, Wehe, A., Iowa State U, Beaman, R., Yale U, grant co-PI, Cellinese, N., Yale U., grant co-PI, Davis, C., Harvard U, grant co-PI, Hilu, K., Virginia Tech, grant co-PI, Judd, W., U Florida, grant co-PI, Manchester, S. U Florida, grant co-PI, Olmstead, R. U Washington, grant co-PI, Qiu, Y.-L., U Michigan, grant co-PI, Sierwald, P., Field Museum, grant co-PI, Soltis, D.S., U Florida, grant co-PI, Soltis, P.S., U Florida, grant co-PI, Sytsma, K., U Wisconsin, grant co-PI, Wurdack, K., Smithsonian Institution, grant co-PI

## DOUGLAS E. SOLTIS

Department of Botany & the Genetics Institute  
University of Florida, Gainesville, FL 32611  
Phone: 352-273-1963; dsoltis@botany.ufl.edu

### Professional Preparation:

B.S. (Biology) College of William and Mary, 1975  
M.A. (Biology) Indiana University, 1977  
Ph.D. (Biology) Indiana University, 1980  
Postdoctoral experience, University of British Columbia, summer 1981

### Appointments:

William R. Ogg Fellowship, Indiana University 1979-1980  
Associate Instructor, Indiana University, 1975-1979  
Postdoctoral experience, University of British Columbia, summer 1981  
Assistant Professor, The University of North Carolina at Greensboro, 1980-1983  
Assistant Professor, Washington State University, 1983-1986  
Associate Professor, Washington State University, 1986-1990  
Acting Director of the Ownbey Herbarium, 1990-91  
Professor, Washington State University, 1990-2000  
Professor, University of Florida, 2000-present  
Chair, Department of Botany, University of Florida, 2006-present

### Five Relevant Publications:

Moore, M. M., C. D. Bell, P. S. Soltis, and D. E. Soltis. 2007. Using plastid genomic-scale data to resolve enigmatic relationships among basal angiosperms. *Proc. Natl. Acad. Sci., USA* 104: 19363–19368.  
Jian, S., P. S. Soltis, M. Gitzendanner, M. Moore, R. Li, T. Hendry, Y. Qiu, A. Dhingra, C. Bell, and D. E. Soltis. 2007. Resolving an ancient, rapid radiation in Saxifragales. *Systematic Biology* 57: 1-20.  
Soltis, D. E., M. Gitzendanner, and P. S. Soltis. 2007. A Bayesian analysis of the three-gene, 567-taxon data set for angiosperms. *International Journal of Plant Sciences* 168: 137-157.  
Soltis, D. E. 2007. Reconstructing the tree of life: taxonomy and systematics of species rich taxa. *Evolution* 61: 3007–3011.  
Soltis, D. E., P. S. Soltis, M. W. Chase, and P. Endress. 2005. *Phylogeny and Evolution of Angiosperms*. Sinauer Associates, Sunderland, MA.

### Five Additional Publications:

Soltis, D. E., H. Ma, M. Frohlich, P. Soltis, V. Albert, D. Oppenheimer, N. Altman, C. dePamphilis, and J. Leebens-Mack. 2007. The floral genome: an evolutionary history of polyploidy and shifting patterns of gene expression. *Trends Plant Science* 12: 358-367.  
Soltis, D. E., V. A. Albert, J. Leebens-Mack, J. Palmer, R. Wing, C. dePamphilis, H. Ma, J. E. Carlson, N. Altman, S. Kim, K. Wall, A. Zuccolo, P. S. Soltis. 2007. The *Amborella* Genome Initiative: A genome for understanding the evolution of angiosperms. *Genome Biology*, in press  
Soltis, D. E., A. Chanderbali, S. Kim, M. Buzgo, and P. S. Soltis. 2007. The ABC model and its applicability to basal angiosperms. *Annals of Botany* 100: 155–163.  
Soltis, D. E., P. S. Soltis, D. W. Schemske, J. F. Hancock, J. N. Thompson, B. C. Husband, and W. S. Judd. 2007. Autopolyploidy in angiosperms: have we grossly underestimated the number of species? *Taxon*: 56:13-30.  
Bell, C. D., D. E. Soltis, and P. S. Soltis. 2005. The age of the angiosperms: A molecular time-scale without a clock. *Evolution* 59: 1245-1258.

**Synergistic Activities (recent and representative):**

Planning Committee, International Organization of Plant Biosystematists, Slovakia, 2008  
Associate Editor, *American Journal of Botany*, 2006-present  
Head, Advisory Committee, Polyploidy—Plant Genome Grant, 2001-2005  
Deep Green, Executive Committee, 2000-present  
Advisory board, *Kew Bulletin* (UK), 2004-2007  
Co-Organizer of PAG polyploidy symposia, 2005-2008  
Co-Organizer, NSF-DFG Biodiversity Conference, November, 2005  
Polyploidy International Conference co-organizer, 2003  
Co-Organizer of multiple Deep Time symposia and workshops, 2002-2007

**Conflicts of Interest:****(i) Collaborators (last 48 months):**

D. Albach, U. Mainz; V. Albert, U. Oslo; G. A. Allen, U. Victoria; T. J. Barraclough, Imperial College, Silwood; M. Bennett, Royal Botanic Gardens, Kew; B. Bremer, Royal Swedish Acad. Sci.; K. Bremer, Stockholm U.; M. W. Chase, Royal Botanic Gardens, Kew; P. Crane, U. Chicago; C. dePamphilis, Penn State Univ.; M. J. Donoghue, Yale U.; C-Z. Fan, North Carolina State U.; K. Hilu, Virginia Polytechnic Inst.; L. Hufford, Washington State U.; J. Leebens-Mack, U. Georgia; A. Leitch, U. London; I. Leitch, Royal Botanic Gardens, Kew; H. Ma, Penn State U.; P. Manos, Duke U.; S. Mathews, Harvard; M. Moody, U. Connecticut; D. Nickrent, Southern Illinois U.; R. G. Olmstead, U. Washington; Y-L. Qiu, U. Michigan; J. L. Reveal, U. Maryland; V. Savolainen, Imperial College, Silwood; P. F. Stevens, U. Missouri-St. Louis; E. A. Zimmer, Smithsonian Institution

**(ii) Advisors:** M.A./Ph.D. Advisor: G. J. Gastony; Postdoctoral Advisor: B. A. Bohm

**(iii) Major Thesis Advisor:** F. A. Bryan (1983); L. H. Rieseberg (1987; U. Brit. Col.); B. Ness (1989, Anguin College); R. D. Noyes (1989; Univ. Colorado); S. J. Brunsfeld (1990, U. Idaho, deceased); G. M. Plunkett (1994; Virginia Commonwealth); Q-Y. Xiang (1995, North Carolina State U.); L. A. Johnson (1997; Brigham Young U.), D. Strenge (1998, Battelle Research), A. Rabe (1999, Nature Conservancy), R. K. Kuzoff (1998; U. Wisc.-Stevens Pt.), M. E. Mort (1999; U. Kansas), M. Zanis (2002; Purdue U.); Stacie A. Kageyama (2001; U. Oregon); C. Notis (2004; U. Michigan—Flint); C. Edwards (2001-07, U. Wyoming); M. Arakaki (2002-); M-J. Yoo (2002-); S. Brockington (2002-); J. Clayton (2003-); J. Koh (2005-); S. Servick (2006-); N. Miles (2007-); L. Majure (2007-); M. Latvis (2007-).

**(iv) Postdoctoral-Scholar Sponsor:** T. Ranker (1987-88; U. Colorado); E. Conti (1994-95; U. of Zurich); M. Fishbein (1998-2000; Portland State U.); J. C. Pires (2000-2001; U. Missouri); A. Scheen (2001-2002; U. Oslo); E. Melendez-Ackerman (2002-2003; U. Puerto Rico); S. Kim (2001-07; Korean Inst. Biol. Res.); M. Buzgo (2002-07; Louisiana State U.-Shreveport); H. Wang (2002-03; U. Florida); J. Tate (2002-2006; Massey U.); C. Bell (2003-04; New Orleans U.); V. Symonds (2005-06; Massey U.); A. Powell (2005-06; U. Evansville); M. Moore (2005-07; Oberlin); S. Jian (2005-06; Chinese Acad. Sci.); H. Wang (2006-07; Chinese Acad. Sci.); A. Doust (2006-07; Oklahoma State U.); E. Mavrodiev (2002-); A. Chanderbali (2003-); R. Buggs (2007-).

**PAMELA S. SOLTIS**

Florida Museum of Natural History and the Genetics Institute  
University of Florida, Gainesville, FL 32611  
Phone; 352-273-1964; [psoltis@flmnh.ufl.edu](mailto:psoltis@flmnh.ufl.edu)

**Professional Preparation:**

B.A. (Biology), Central College, Pella, IA, 1980, Summa Cum Laude  
M. Phil. (Botany), University of Kansas, 1984, with Honors  
Ph.D. (Botany), University of Kansas, 1986

**Appointments:**

Curator, Florida Museum of Natural History, and Professor, Genetics Institute, University of Florida,  
October, 2000 – present  
Distinguished Professor, University of Florida, 2007 – present  
Fulbright Distinguished Scholar, Royal Botanic Gardens, Kew, England, and Imperial College, Silwood  
Park, England, 2000 – 2001  
Assistant, Associate, and Full Professor, Department of Botany, Washington State University, 1986 –  
2000  
Mellon Senior Fellow, Smithsonian Institution, 1994-95

**Five Relevant Publications:**

Soltis, D. E., P. S. Soltis, P. K. Endress, and M. W. Chase. 2005. *Angiosperm Phylogeny and Evolution*.  
Sinauer Associates, Sunderland, MA.  
Soltis, P. S. and D. E. Soltis. 2004. The origin and diversification of angiosperms. *American Journal of  
Botany* 91: 1614-1626.  
Jian, S., P. S. Soltis, M. Gitzendanner, M. Moore, R. Li, T. Hendry, Y. Qiu, A. Dhingra, C. Bell, and D.  
E. Soltis. 2008. Resolving an ancient, rapid radiation in Saxifragales. *Systematic Biology* 57: 1-20.  
Bell, C. D., D. E. Soltis, and P. S. Soltis. 2005. The age of the angiosperms: A molecular time-scale  
without a clock. *Evolution* 59: 1245-1258.  
Davies, T. J., T. G. Barraclough, M. W. Chase, P. S. Soltis, D. E. Soltis, and V. Savolainen. 2004.  
Darwin's abominable mystery: Insights from a supertree of the angiosperms. *Proceedings of the  
National Academy of Sciences, USA* 101: 1904-1909.

**Five Additional Relevant Publications:**

Soltis, P. S., D. E. Soltis, S. Kim, A. Chanderbali, and M. Buzgo. 2006. Expression of floral regulators in  
basal angiosperms and the origin and evolution of the ABC model. In D. E. Soltis, J. Leebens-Mack,  
and P. S. Soltis (eds.), *Developmental Genetics of the Flower. Advances in Botanical Research*  
44:483-506.  
Cui, L., P. K. Wall, J. Leebens-Mack, B. G. Lindsay, D. Soltis, J. J. Doyle, P. Soltis, J. Carlson, A.  
Arumuganathan, A. Barakat, V. Albert, H. Ma, and C. W. DePamphilis. 2006. Widespread genome  
duplications throughout the history of flowering plants. *Genome Research* 16: 738-749.  
Soltis, P. S., D. E. Soltis, V. Savolainen, P. R. Crane, and T. Barraclough. 2002. Rate heterogeneity  
among lineages of land plants: integration of molecular and fossil data and evidence for molecular  
living fossils. *Proceedings of the National Academy of Sciences, USA* 99: 4430-4435.  
Zanis, M. J., D. E. Soltis, P. S. Soltis, S. Mathews, and M. J. Donoghue. 2002. The root of the  
angiosperms revisited. *Proceedings of the National Academy of Sciences, USA* 99: 6848-6853.

Soltis, P. S., D. E. Soltis, and M. W. Chase. 1999. Angiosperm phylogeny inferred from multiple genes: A research tool for comparative biology. *Nature* 402: 402-404.

**Synergistic Activities (recent representative activities):**

*Society Service:* President, Society of Systematic Biologists, 2004-06  
President, Botanical Society of America, 2006-09

*Editorial Service:* Associate Editor, *Systematic Biology*, 2001-07  
Associate Editor, *Evolution*, 2003-07

*NSF Workshops:* The Future of AToL, Washington, DC, 2008  
Species Diversity on Earth, Washington, DC, 2005 (co-organizer)  
Databases in Plant Systematics, Gainesville, FL, 2003 (co-organizer)

*Research Collaborations:* International Polyploidy Conference, London, 2003 (co-organizer)  
Deep Time Research Coordination Network, 2001-present  
Floral Genome Project, 2001-present  
Ancestral Angiosperm Genome Project, 2006-present

*Advisory Board Chair:* NSF Plant Genome Polyploidy Project (L. Comai, PI)

**Conflicts of Interest:**

**(i) Collaborators & Co-authors:** V. Albert (U. Oslo), N. Altman (Penn State U.), M. Bennett (Royal Bot. Gard., Kew), T. Borsch (U. Bonn), P. Cantino (Ohio U.), J. Carlson (Penn State U.), M. Chase (Royal Bot. Gard., Kew), J. Chen (U. Texas), L. Cui (Penn State U.), C. Davis (Harvard), C. dePamphilis (Penn State U.), M. Donoghue (Yale), J. Doyle (Cornell), P. Endress (U. Zurich), S. Farris (Stockholm U.), M. Frohlich (Natural History Museum, London), P. Herendeen (George Washington U.), K. Hilu (Virginia Tech U.), R. Huck (U. Florida), L. Hufford (Washington State U.), H. Kong (Chinese Acad. Sci.), A. Kovarik (Czech Acad. Sci.), J. Leebens-Mack (U. Georgia), A. Leitch (U. London), I. Leitch (Royal Bot. Gard., Kew), Y. Lim (U. London), H. Ma (Penn State U.), P. Manos (Duke), S. Mathews (Harvard), R. Matyasek (Czech Acad. Sci.), M. Moody (U. Connecticut), L. Mueller (Cornell), L. Oliveira (U. Vicosa), R. Olmstead (U. Washington), Y.-L. Qiu (U. Michigan), L. Ronse DeCraene (Royal Bot. Gard., Edinburgh), M. Sanderson (U. Arizona), V. Savolainen (Imperial College, Silwood), S. Schlarbaum (U. of Tennessee), K. Sytsma (U. Wisconsin), S. Tanksley (Cornell), G. Theissen (U. Jena), L. Zahn (Penn State U.)

**(ii) Graduate Advisor:** W. L. Bloom

**(iii) Thesis Advisor/Postgrad Sponsor: Students:** P. G. Wolf (PhD 1990; Utah State U.); R. E. B. Kirkpatrick (MS 1988; UC-Berkeley); T. S. Richter (MS 1990); M. S. Mayer (PhD 1993; U. San Diego); J. L. Schultz (PhD 1996; Lewis-Clark State College); T. M. Hardig (PhD 1998; U. Montevallo); L. M. Cook (PhD 1998; Washington State U.); D. Albach (MS 1998; U. Mainz); J. A. Koontz (PhD 2000; Augustana College); M. A. Gitzendanner (PhD 2000; U. Florida); P. Speranza (PhD 2005; U. de la República); A. Morris (PhD 2006; U. S. Alabama); H. Loring (MS 2006); C. Edwards (PhD 2007; U. Wyoming); R. Vergara (2002 -); S. Brockington (2002 -); **Post-docs:** T. Ranker (U. Colorado); S. Novak (Boise State U.); J. C. Pires (U. Missouri); S. Kim (Korean Inst. Biol. Res.); M. Buzgo (Louisiana State U.-Shreveport); H. Wang (U. Florida); J. Tate (Massey U.); C. Bell (U. New Orleans); A. Powell (U. Evansville); V. Symonds (Massey U.); M. Moore (Oberlin); S. Jian (Chinese Acad. Sci.); H. Wang (Chinese Acad. Sci.); A. Doust (Oklahoma State U.); E. Mavrodiev; A. Chanderbali; R. Buggs

## VAL TANNEN

Computer and Information Science Department  
University of Pennsylvania  
Levine Hall, 3330 Walnut St., Philadelphia, PA 19104  
val@cis.upenn.edu, (215)898-2665, fax -0587

### PROFESSIONAL PREPARATION

Polytechnic Institute of Bucharest, Computer Engineering, BSE/MSE 1977.  
Massachusetts Institute of Technology, Applied Mathematics/Computer Science, PhD, 1987.  
MIT Laboratory for Computer Science, Postdoc in Computer Science, 1987.

### APPOINTMENTS

1999–present, Professor, Department of Computer and Information Science, University of Pennsylvania.  
2002–2003, Visiting Professor, MPLA, University of Athens, and Visiting Researcher, FORTH, Crete, Greece (sabbatical leave)  
1993–1999, Associate Professor, Department of CIS, UPenn.  
1994–1995, Visiting Professor, Université de Paris XI, Orsay, and Visiting Researcher, INRIA, Rocquencourt, France (sabbatical leave).  
1987–1993, Assistant Professor, Department of CIS, UPenn.  
1985, 1986, Researcher, IBM Thomas J. Watson Research Center.

### SELECTED RECENT PUBLICATIONS

“Annotated XML: Queries and Provenance” J.N. Foster, T.J. Green, and V. Tannen. Proceedings PODS 2008.

“Update Exchange with Mappings and Provenance” T.J. Green, G. Karvounarakis, Z. Ives, and V. Tannen. Proceedings VLDB 2007: 675-686

“ORCHESTRA: Facilitating Collaborative Data Sharing” T.J. Green, G. Karvounarakis, N. Taylor, O. Biton, Z. Ives, and V. Tannen. Demo, Proceedings SIGMOD 2007: 1131-1133

“Provenance semirings” T.J. Green, G. Karvounarakis, and V. Tannen. Proceedings PODS 2007, pp.31-40.

“Data Integration in the Life Sciences” S. Cohen Boulakia and V. Tannen (eds.) Proceedings 4th International Workshop, DILS 2007, Philadelphia, Springer LNCS 4544 (LNBI), 2007

“Models for Incomplete and Probabilistic Information” T.J. Green and V. Tannen, Proceedings EDBT 2006, IIDB Workshop, also in Data Engineering Bull., Vol. 29, Mar.2006, pp.17-24.

“Query Reformulation with Constraints” A. Deutsch, L. Popa, and V. Tannen. SIGMOD Record 35(1): 65-73 (2006)

“XML Queries and Constraints, Containment and Reformulation”, A. Deutsch and V. Tannen. Theoretical Computer Science, Vol.336, May 2005, pp.57-87.

“Semantic Web and Databases” C. Bussler, I. Fundulaki, and V. Tannen (eds.) Revised Selected Papers from SWDB, Toronto, 2004. Springer LNCS 3372, 2005.

“MARS: A System for Publishing XML from Mixed and Redundant Storage”, A. Deutsch and V. Tannen. Proceedings VLDB 2003.

“Viewing the Semantic Web through RVL Lenses.” A. Magkanaraki, V. Tannen, V. Christophides, and D. Plexousakis. Proceedings of 2nd Int’l Semantic Web Conference (best paper award) 2003. Full version in J. Web Sem. 1(4): 359-375 (2004)

### **SYNERGISTIC ACTIVITIES**

NSF Presidential Young Investigator 1990–1995, Editor, “Electronic Journal of Discrete Mathematics and Theoretical Comp. Sci.”, Area Editor, “Encyclopedia of Databases”, Springer-Verlag 2008.

**Program Committees (recent):** 2009 Mendelzon Wksh’p on Foundations of Data Management (AMW), 2009 Int’l Conf. on Extended Database Technology (EDBT), 2009 Int’l Conf. of Data Engineering (ICDE), 2008 Principles of Database Systems Symp. (PODS), 2008 Int’l Conf. on Extended Database Technology (EDBT), 2008 Int’l Conf. on Very Large Databases (Demos) (VLDB), Co-Chair, 2007 Data Integration in Life Sciences (DILS), Vice Chair, 2007 Semantic Web Conf. (ISWC), 2007 Int’l Conf. on Scalable Uncertainty Management (SUM), Area Chair, Scientific and Biological Databases and Bioinformatics for 2006 Int’l Conf. of Data Engineering (ICDE), 2006 Int’l Conf. on Extended Database Technology (EDBT), 2006 Semantic Web Conf. (ISWC), 2005 Principles of Database Systems Symp. (PODS), Co-Chair, 2004 Semantic Web and Databases Wksh’p (SWDB), 2004 Web and Databases Wksh’p (WebDB), 2004 XML Database Symp. (XSym), 2004 Semantic Web Conf. (ISWC), 2003 Int’l Wksh’p on Database Prog. Lang. (DBPL), 2003 Int’l Conf. on Very Large Databases (VLDB), 2002 Int’l World Wide Web Conf. (WWW11), 2001 Int’l World Wide Web Conf. (WWW10), 2001 Int’l Conf. on Database Theory (ICDT),

**Keynote and Conference Invited Talks (since 1997):** Int’l Provenance and Annotation Wksh’p, Salt Lake City, June 2008; Symposium on Provenance in Databases, e-Science Institute Edinburgh, May 2008; Workshop on Principles of Provenance, Edinburgh, Nov.2007; Mendelzon Workshop on Foundation of Databases and the Web, Chile, Nov.2006; Int’l Workshop on Exchange and Integration of Data, Brixen-Bressanone, Italy, June 2006; University of Washington / Microsoft Research Summer Institute on Infrastructure for Managing Imprecise Information in Relational Database Systems”, Salish Lodge, Aug.2005, Grids and Applied Language Theory: NeSC Workshop, Edinburgh, Oct.2003; Foundations of XML: Dagstuhl Workshop, Sept.2001.

**Recent Invited Lectures** at SUNY Buffalo, UC San Diego, Microsoft Research, TU Vienna, University of Washington, Edinburgh University, Indiana University, National Technical University of Athens, INRIA Rocquencourt, Bell Labs, ETH Zurich, AT&T Labs, Ecole Polytechnique Paris, TU Eindhoven, University of Antwerpen, University of Torino, M.I.T., Ecole Normale Supérieure Paris, Imperial College London, University of Southern California, University of Paris VII, Carnegie Mellon University, Hewlett-Packard Laboratories, Stanford University, University of Pisa.

**SOME FORMER PhD STUDENTS** Atsushi Ohori, Tohoku University, Dan Suciu Univ. of Washington, Lucian Popa, IBM Almaden, Alin Deutsch, UC San Diego, Arnaud Sahuguet, Google.

**Collaborators (not PhD students) in the last 4 years:** J.N. Foster, N. Taylor, UPenn, O. Biton, S. Cohen-Boulakia, Univ. Paris 11, W. Piel, Yale, Z. Ives, UPenn.

**PhD Students and PostDocs in the last 5 years:** T. J. Green, G. Karvounarakis. Total number of PhD students: 9, Total number of postdocs: 3.

**PhD Thesis and PostDoc Supervisor:** A. Meyer, MIT.

## TODD J. VISION

Department of Biology, University of North Carolina at Chapel Hill  
Campus Box 3280, Chapel Hill, NC 27599  
Phone: 919.843.4507, Fax: 919.962.1625, email: [tjv@bio.unc.edu](mailto:tjv@bio.unc.edu)

### Professional Preparation

<u>Institution</u>	<u>Area</u>	<u>Degree/Position</u>	<u>Year(s)</u>
University of Chicago	Biological Sciences	B.A.	1992
Princeton University	Evolutionary Genetics	M.S.	1995
Princeton University	Evolutionary Genetics	Ph.D.	1998

### Appointments

2007-present	Associate Professor, Department of Biology, UNC Chapel Hill Affiliations: Curriculum in Bioinformatics and Computational Biology, Curriculum in Genetics and Molecular Biology, Carolina Genome Sciences Center
2006-present	Associate Director for Informatics, National Evolutionary Synthesis Center (NESCent), Durham, NC
2001-2007	Assistant Professor, Department of Biology, UNC Chapel Hill
1999-2001	Postdoctoral Associate, Center for Agricultural Bioinformatics, USDA-ARS, Ithaca, NY
1998-1999	Postdoctoral Associate, Dept. of Plant Breeding, Cornell University, Ithaca NY

### Five Most Relevant Recent Publications (out of 26 total)

- Hartmann S, Vision TJ (2008) Using ESTs for phylogenomics: can one accurately infer a phylogenetic tree from a gappy alignment? *BMC Evolutionary Biology* 8, 95.
- Lapp H, and 24 others (2007) The 2006 NESCent Phyloinformatics Hackathon: a Field Report. *Evolutionary Bioinformatics* 3, 357–366.
- Leebens-Mack J, and 26 others (2006) Taking the first steps towards a standard for reporting in phylogenies: Minimal Information About a Phylogenetic Analysis (MIAPA). *OMICS* 10, 231-237.
- Hartmann S, Lu D, Phillips J, Vision TJ (2006) Phytome: a platform for plant comparative genomics. *Nucleic Acids Research* 34, D724-730.
- Remington DL, Vision TJ, Guilfoyle TJ, Reed JW (2004) Contrasting modes of diversification in the Aux/IAA and ARF gene families. *Plant Physiology* 135, 1738-1752.

### Other relevant and significant publications

- Bouck A, Vision TJ (2007) The molecular ecologist's guide to expressed sequence tags. *Molecular Ecology* 16, 907-924.
- Crayton ME, Powell BC, Vision TJ, Giddings MC (2006) The evolution of alternatively spliced exons within the DSCAM gene family. *BMC Evolutionary Biology* 6, 16.
- Ganko EW, Meyers BC, Vision TJ (2007) Divergence in expression between duplicated genes in Arabidopsis. *Molecular Biology and Evolution* 24, 2298-2309.
- Knies JL, Dank KK, Vision TJ, Hoffman N, Swanstrom RI, Burch CL (in press) Compensatory evolution in RNA secondary structures increases substitution rate variation among sites. *Molecular Biology and Evolution* doi:10.1093/molbev/msn130.
- Vision TJ, Brown DG and Tanksley SD. 2000. The origins of genomic duplications in Arabidopsis. *Science* 290, 2114-7.

## Synergistic Activities

- Development and support of software programs and databases for genomics (MapPop, FISH, Phytome, mimulusevolution.org).
- Development of “The Power Within” curriculum, on phylogenomics and the origin of the eukaryotic cell, for the Destiny Science Bus, which delivers inquiry-based science learning activities to underserved minority K-12 teachers and students in NC.
- Co-organizer of NESCent Phyloinformatics Summer Course (2007, 2008); NESCent Phyloinformatics Summer of Code (2007, 2008); NESCent Phyloinformatics Hackathon (2007, 2008).

## Collaborators and other Affiliations

**Coauthors on papers & grants (last 5 yrs, excluding students, postdocs & UNC faculty):** C. Ané (U Wisconsin), T. Bradshaw (U Washington), J. Bowers (U Georgia), E. Brenner (NYBG), G. Burleigh (U. Florida), S. Cannon (USDA), M. Clement (Brigham Young), J. Comstock (unaff.), M. Crayton (Xavier U), C. Cunningham (Duke U), C. dePamphilis (Penn. State), R. deSalle (AMNH), J. Doyle (Cornell), J. Eisen (UC Berkeley), Oliver Eulenstein (Iowa State), D. Fernandez-Baca (U. Iowa), X. Gu (Iowa State), J. Harshman (unaff), B. Hemminger, M. Holder (U Kansas), R. Holland (EMBL), I. Holmes (UC Berkeley), R. Jansen (UT Austin), T. Katayama (Osaka U), L. Katz (Smith College), E. Kellogg (U Missouri), S. Knapp (U. Georgia), J. Knies (Brown), E. Koonin (NCBI), H. Lapp (NESCent), J. Leebens-Mack (U Georgia), P. Lewis (U Connecticut), P. Mabee (U South Dakota), A. Mackey (GSK), B. Martin (Oklahoma State), B. Meyers (U of Delaware), W. Michener (U New Mexico), B. Mishler (UC Berkeley), B. Osborne (BioTeam), H. Philippe (U Montreal), W. Piel (Yale), S. Kosakovsky Pond (UC San Diego), W-G. Qiu (CUNY), D. Schemske (Michigan State), C. Pires (U Missouri), Y-L. Qiu (U Michigan), S. Rhee (Stanford U), K Sjölander (UC Berkeley), D. Soltis (U Florida), P. Soltis (U Florida), J. Stajich (UC Berkeley), L. Stein (CSHL), D. Stevenson (NYBG), A. Stoltzfus (NIST), T. Thierer (Biomatters), A. Vilella (EMBL), L. Szekely (U South Carolina), K. Smith (NESCent), J. Tang (U South Carolina), C. Tauer (Oklahoma State U), J. Tomkins (Clemson U), R. Vos (U British Columbia) T. Warnow (UT Austin), M. Westerfield (U Oregon), J. Willis (Duke U), C. Zmasek (Burnham Inst.)

**Graduate students:** 2 current: Toby Clarke, Suja Thomas; 2 former: Sugata Chakravarty (MS 2004, UNC Charlotte), Ruchir Shah (PhD 2006, Constella Group).

**Postdocs:** 1 current: Eric Ganko; 7 former: Amy Bouck (Pioneer Hi-Bred), María Chacón (U. Nacional de Colombia), Jixin Deng (Baylor U.), Stefanie Hartmann (U. Potsdam), Maria Tsompana (SUNY Buffalo), Rajkumar Rathinavelu (Indian Tobacco Co.), Zongli Xu (NIEHS)

**Ph.D. advisors:** H. Hollocher (U Notre Dame), D. Stratton (U. Vermont)

**Postdoc advisors:** S. Cartinhour (USDA), S. Tanksley (Cornell U).

## GRAND CHALLENGE WORKSHOP DESCRIPTION

### Statement of the Scientific Problem

Amazing progress has been made over the past few decades in resolving the Tree of Life. Plant scientists have been leaders in this global effort, and are widely regarded as being in the vanguard both in organizing large collaborative efforts and in pushing the limits in the analysis, synthesis, and use of phylogenetic information. While there is much to be proud of, the magnitude of the task that still lies ahead is daunting, and we find ourselves struggling to keep up with the massive datasets that are accumulating through our own efforts and those of our colleagues throughout the plant sciences. Data-mining pipelines are underdeveloped, as are tools for the assembly, analysis, and visualization of larger and larger trees. We still do not have proper tools to integrate and display data from multiple sources (e.g., molecular sequences, genomic data, expression data, morphological and developmental data, fossil evidence), and the results of most phylogenetic studies have not been assembled and made readily accessible, and, therefore, remain effectively unavailable to the wide variety of potential user communities in research and in education.

All of this emphasizes the clear need to greatly expand the intersection between phylogenetic biology and computer science, computational biology, and bioinformatics. Some important attempts have been made recently to bridge this gap, such as the NSF-funded CIPRES project. This effort has highlighted the enormous needs that exist, but has also led to the realization that solutions to these problems require efforts that focus on the needs of a particular research community. The iPlant Collaborative presents a unique opportunity to connect plant phylogenetic biologists with colleagues in the relevant computer sciences, to craft real world solutions, not only to deliver better phylogenetic inferences, but to render this information truly useful and to enable research and education across the plant sciences. In turn, this effort would provide a model that could extend well beyond the plant sciences.

The purpose of the Grand Challenge Workshop that we propose is to begin this process; that is, to bring together a set of relevant plant phylogeneticists with computer scientists and informatics experts to visualize a cyberinfrastructure that would propel us all forward. Importantly, the beneficiaries of such a development would not just be phylogenetic biology and the plant sciences, but the computer sciences as well, as many of the problems in this realm are of general theoretical and practical interest. Likewise, our ultimate aim is to enable new, synthetic research and education across the entire scientific community. Approaches and tools developed through the iPlant Collaborative will be useful to those studying other branches of the Tree of Life and other problems – plant structure, function, and gene family evolution, to name just a few – that require phylogenetic analysis of large datasets.

It is important to stress that the likelihood of success is especially high in the plant sciences, where the plant Tree of Life community has already demonstrated a willingness to collaborate on a grand scale. Beginning in the early 1990's, a series of major efforts have been made to compile and analyze large datasets of gene sequences from all three plant genomes (nuclear, plastid, and mitochondrial). These team efforts involved many of the investigators associated with the present proposal, who have helped to build and maintain an international network of expertise and data sources. This has spawned several NSF Research Coordination Networks (RCN's), including the Deep Green, Deep Time, and Deep Gene projects. As a result of these efforts, angiosperms became the first major group of organisms to be classified based on molecular phylogenetic analyses (the so-called APG system), and recently we have provided the outline of a truly phylogenetic classification of all vascular plants. These advances mark a turning point not only in terms of our knowledge, but in the way that we operate as a community.

Currently, six NSF-funded Assembling the Tree of Life (AToL) projects focus on green plants, five of which address specific clades (basal green plants, liverworts, seed plants, angiosperms, and monocots) and coordinate the activities of a community of plant scientists, involving morphologists, paleobotanists, and molecular phylogeneticists, with direct connections to ongoing genomics and evo-devo efforts. A sixth project focuses on the use of whole plastid genome sequences for phylogenetics. Excellent communication has been established among these efforts, which is especially evident in their development of common informatics solutions (e.g., TOLKIN).

As an example of a specific effort, the angiosperm AToL group has resolved 12 of the most problematical deep-level nodes, providing a firm foundation for future studies. We have explored challenges in the analysis of large datasets, including supermatrix approaches with large amounts of missing data. In campanulids, we have constructed a supertree of ca. 5,000 (of ~30,000) species from >200 phylogenetic studies. We are currently building a 16-gene dataset from all three plant genomes for hundreds of species to provide a new, comprehensive phylogenetic backbone (replacing our widely used 3-gene, 567-taxon tree). We have also assembled matrices of over 100 complete plastid genome sequences. Finally, it is noteworthy that plant phylogeneticists were among the first to promote global databases of phylogenetic data matrices and trees (e.g., TreeBASE), and have been leaders in the development of key data-mining and analytical tools (e.g., PhyLoTa, Phytome), as well as resources to increase the accessibility of phylogenetic information (e.g., APWeb, Phylomatic).

### **Why the problem requires cutting-edge computer science, bioinformatics, and computational biology tools, rather than off-the-shelf solutions**

Building an effective cyberinfrastructure for phylogenetic biology will require solutions to a host of practical and theoretical problems in computer science, bioinformatics and phylogenetics itself. Some are well characterized but computationally daunting, such as developing good heuristic solutions for several NP-complete problems for which very large data inputs are now at hand (multiple sequence alignment; construction of phylogenetic trees; assembly of synthetic "supertrees"). Others are not so well characterized but are emerging as problems unanticipated before the availability of such large quantities of sequences and other data (e.g., phylogenetic incongruence between different regions of the genome; the complexities of gene family diversification). Still other problems arise because of the breadth and heterogeneous nature of data that can be brought to bear on building phylogenetic trees: morphology, development, gene expression and other postgenomic data, etc. The canonical informatics problem of data integration is exemplified in our domain by the vast array of diverse datasets that retain footprints of evolutionary history.

Few of these problems currently have off-the-shelf solutions. There is, however, substantial prior experience in the phylogenetics community for building a next-generation phylogenetics cyberinfrastructure. On the one hand, the diversity of mathematical and computational methodologies for solving specific problems in phylogenetics has never been higher (e.g., recent work on building reticulate histories in hybridizing plants). This has been enabled in large part by substantial buy-in from the math and computer science communities. On the informatics side, the NSF-funded Cyberinfrastructure for Phylogenetic Research (CIPRES) project, which will soon be ending, tackled several key issues, including an upgrade of TreeBASE, the community repository of phylogenetic trees. Through its AToL and other programs, the NSF has supported several smaller infrastructure-related projects, including TOLWeb, TOLKIN, PhyLoTA, pPod, PhyloWidget, and others. Moreover, many bioinformatics resources have added phylogenetic components in recent years (e.g., tree functions in PFAM and GenBank's BLAST server). However, an overarching infrastructure fostering true high-level integration is still largely missing. These efforts need to be focused and coordinated, to sculpt, reorganize, synthesize, and integrate data and render them more broadly useful for plant biology by orders of magnitude.

### **Description of the datasets currently available or that will soon be available**

Phylogenetic biology in many respects is not data limited. GenBank alone contains at least one DNA sequence for about 10% of all species known to science: this includes some 26 million sequences across 67,000 species of green plants. Phenotypic data have also been databased for protein structure (PDB), gene expression patterns (NCBI GEO), pathway networks (KEGG), and external morphology (MorphBank, etc.), to name just a few. These tend to be smaller repositories, but their information content is high owing to the inherent complexity of their traits compared to simple nucleotide sequences. Other kinds of databases are also relevant: paleontological databases archive important data on fossil taxa (Paleobiology Database); biodiversity collections databases (e.g., GBIF) can tie species to georeferenced collection sites, bringing geography, ecology and climate into the mix; and nomenclatural-taxonomic

databases, which help to solve important problems in disambiguating the identities of species. In the next few years, the greatest explosion of growth in data for phylogenetics will no doubt arise from next-generation sequencing technologies that promise to deliver whole genome sequences for a substantially larger sample of the plant tree of life than we have seen up to now.

Thus we need not worry about stimulating the growth of the relevant databases. If anything, the problem is largely the opposite - the phylogenetics community is wallowing in more data than can be effectively navigated, summarized, rendered into usable representations, or importantly, synthesized in "grand" and innovative ways.

### **Goals and outcomes of this GCW**

The primary goal of the workshop is to bring experts on plant phylogeny and several related botanical disciplines together with computer scientists, computational biologists, and bioinformaticians, to envision a cyberinfrastructure to enable phylogenetic research and to make the results of such research maximally useful across the plant sciences. Among those present will be leaders of a number of phyloinformatics projects that are actively trying to bridge this gap. The meeting will be designed to accommodate the exchange of information and the identification of current barriers, and then to map these barriers to specific cyber-solutions.

We expect this GCW to lead to the development of a Grand Challenge Project within the iPlant Collaborative, and that in this context "discovery environments" would be developed to advance the cause of integration, synthesis, and training. In the meantime, a white paper will be produced from the GCW to summarize the current state of affairs and to project a vision for the future. The GCW will also serve the immediate purpose of stimulating connections among the participants and the many perspectives and projects that they represent.

Special attention will be paid to the needs of the wide variety of communities who might benefit by the use of phylogenetic knowledge. This includes research communities spanning the plant sciences, from genomics to ecosystem ecology. We also will address challenges and opportunities in education and outreach, including the K-12 and college levels, informal science education, and coordination with ongoing efforts, for example at the National Evolutionary Synthesis Center (NESCent).

### **What other groups are working in this or related areas? Which people from these groups are likely to be involved in the GCW?**

A variety of ongoing projects bear on the aims of this GCW, and we intend to bring representatives of these efforts together, along with other plant phylogeneticists and computer scientists. These include the NSF-funded CIPRES project, various AToL projects, and a variety of databasing efforts and web resources. The leaders of most of these relevant projects will be invited to participate and have already expressed interest in joining us. Several CIPRES PI's are on our list, including B. Moret (original lead PI on the project), M. Donoghue, B. Piel, R. Jansen, B. Mishler, E. Mossel, D. Swofford, V. Tannen, and D. Maddison, as well as one CIPRES-funded graduate student (S. Smith).

There are now six AToL projects focused on green plants (see above). PI's of all of these are represented among our prospective participants: B. Mishler, M. Donoghue, and R. Olmstead (green plants), J. Shaw (liverworts), S. Mathews (seed plants), D. and P. Soltis, M. Donoghue, R. Olmstead, W. Judd, and Y.-L. Qiu (angiosperms), T. Givnish and J. Leebens-Mack (monocots), and plastid genome analysis (R. Jansen). Two other AToL projects focused on the development of tools for phylogenetics are represented by PI's M. Sanderson (PhyLoTa) and V. Tannen (pPod). Of the AToL projects focused on other groups of organisms, the most relevant work has been the development of a data-mining and analysis pipeline by the Fungal Tree of Life Project; this project will be represented by PI D. Hibbett.

The primary databasing efforts and internet resources will be represented by T. Vision, (Phytome), B. Piel (TreeBASE), D. Maddison (Tree of Life Web Project), R. Beaman (TOLKIN), C. Webb (Phylomatic), and P. Stevens (Angiosperm Phylogeny Website). Other prospective participants have been closely associated with key paleontological databases (P. Crane, J. Doyle, and S. Magallon), with morphological projects such as MorphBank (R. Beaman), with various nomenclature databases (B. Piel,

M. Donoghue), with the international “barcoding” effort (S. Graham), and with the emerging Encyclopedia of Life project (B. Piel, D. Maddison). Also among our invitees will be the developers of key phylogenetic software packages, such as PAUP\* (D. Swofford), MacClade and Mesquite (D. Maddison), Phylocom (C. Webb), and Phyutility (S. Smith). L. Bohs and R. Beaman will represent the NSF’s Planetary Biodiversity Projects (PBI) on Solanaceae and *Euphorbia*, respectively. Finally, developers and users of various pipelines and tools for management of genomic data will be represented; for example, by J. Leebens-Mack and D. and P. Soltis for the Floral Genome and Ancestral Angiosperm Genome Projects.

### **Format of the meeting and outline of the agenda**

We envision a three-day meeting of ca. 35-40 participants, including representatives from the iPlant Collaborative and the Board of Directors. The meeting would be held during the winter months in Oracle, AZ. Owing to various planning and timing constraints that we have already identified, the meeting could take place in late November, 2008, or in late January, 2009. We imagine the workshop taking place on a Thursday-Saturday. Participants might arrive early on Wednesday for a hike in nearby Catalina State Park, and for a reception that evening at the Oracle Conference Center. The first day of the meeting would be devoted to short presentations by participants to highlight progress and barriers in plant phylogeny and to review a variety of relevant computational tools. During the second day a variety of break-out groups will circumscribe the key barriers and their possible cyber-solutions, identifying common themes, overlaps, and possible synergies. The morning of the third day would center on education and outreach opportunities, on mobilizing the relevant communities, and on a timetable for future activities. The final afternoon would focus on outlining a Grand Challenge Project and on specific writing tasks and other commitments. On the Sunday following the meeting, the organizers and selected others will remain in Oracle to begin to write up the results of the GCW, and to plan the next steps.

### ***Day 1 (Thursday)***

#### **AM – Session 1, 8:00 am – 12:30 pm, D. Soltis, moderator**

8:00 Welcome, introductions, and opening comments by GCW PI Michael Donoghue and iPlant Representatives (Rich Jorgensen, others) – Overview of the iPlant Collaborative; “Grand Challenge Projects”; goals for this workshop, i.e., to identify cyberinfrastructure needs for the assembly and use of the Plant Tree of Life; development of the core of a proposal for an iPlant Grand Challenge Project

*8:30-12:30 Phylogenetic knowledge across green plants: progress, issues, and barriers*

8:30 Green Algae, Rick Zechman

9:00 “Bryophytes,” Brent Mishler and Jon Shaw

9:30 Monilophytes, Kathleen Pryer

10:00 Coffee

10:30 Seed Plants, Sarah Mathews and Jim Doyle

11:00 Angiosperms, Doug and Pam Soltis

11:30 Monocots, Tom Givnish and Vincent Savolainen

12:00 Asterids, Dick Olmstead and David Tank

#### **PM – Session 2, 1:15 – 5:30 pm, P. Soltis, moderator**

1:15 Morphological perspectives, Walter Judd and Ned Friedman

1:45 Paleo perspectives, Peter Crane and Susana Magallon

*2:15-6:00 Analytical and informatics challenges*

2:15 Challenges in inferring large trees, Mike Sanderson and Elchanan Mossel

2:45 GenBank, PhyLoTa, Phyutility, Gordon Burleigh and Stephen Smith

3:15 Coffee

3:30 Synteny data and Phytome, Jim Leebens-Mack and Todd Vision

4:00 Tree visualization, Tamara Munzner and Mike Sanderson

4:30 TreeBASE, PhyloWidget, TOLWeb, and Mesquite, Bill Piel and David Maddison

- 5:00 TOLKIN and pPod, Reed Beaman and Val Tannan  
5:30 PhyloMatic and APWeb, Cam Webb and Peter Stevens

**Day 2 (Friday)**

**AM – Session 3, 8:00 am – 12:30 pm, M. Sanderson, moderator**

- 8:00 Take stock of the presentations made on Day 1, and summarize major conclusions. With this background, return to the issue of a “Grand Challenge Problem.” What are the major issues and major needs? What issues require special attention from the computer science, cyberinfrastructure, and informatics communities?

9:45 Coffee

10:00 Break-out groups to connect needs to possible cyber-solutions (5 groups of 7-8 people)

11:30 Summaries and identification of topics for further discussion

**PM – Session 4, 1:30 – 5:30 pm, V. Tannen, moderator**

- 1:30 Break-out groups on topics identified in the AM sessions, each to focus on refining a problem and its possible solutions (the exact make-up and sizes of the break-out groups will depend on the issues identified)

3:00 Coffee

3:30 Group presentations and general discussion, with the aim of identifying common themes and possible synergies

**Day 3 (Saturday)**

**AM – Session 5, 8:00 am – 12:30 pm, T. Vision, moderator**

Organizing the community and forming our iPlant team.

- 8:00 Break-out groups to discuss organization and governance, sociological barriers, mechanisms for communication, integration across sub-teams working on major sub-problems, communication with other iPlant initiatives, and possible education, training, and outreach opportunities.

10:00 Coffee

10:30 Reconvene to share suggestions and develop consensus

11:00 Development of an outline for a Grand Challenge Project proposal

**PM – Session 6, 1:00 – 5:30 pm, M. Donoghue, moderator**

1:00 Break-out groups to draft key sections of the proposal

3:00 Coffee

3:30 Continue drafting

4:00 Reconvene to discuss progress, remaining tasks, coordinate writing and other assignments

**How can the iPlant Collaborative be useful to this GCW effort?**

We trust that the iPlant Collaborative team in Tucson will provide logistical support for the meeting at Biosphere2, especially for transportation issues, lodging, and meals, and will also coordinate with Conference Center staff on details of the meeting itself, including a reception, coffee breaks, and meals. Beyond this base-line support, we look to the Collaborative especially for advice and assistance in coordinating our GCW efforts with other such iPlant efforts. We would be delighted to entertain suggestions about plant and computer scientists whom we could include from other projects and other areas of expertise. Specifically, although a number of excellent computer scientists are among our prospective participants, we would welcome additional suggestions, and we would like to include iPlant computer staff who are already in place. Similarly, although a variety of plant science disciplines are represented by our prospective participants, we would benefit from additional suggestions, especially to connect with model-organism genome projects and their associated informatics projects. Likewise, members of our group may be appropriate in forming links to other iPlant projects.

## SELECTED RECENT REFERENCES AND WEB RESOURCES

### *Plant Phylogeny*

- Bell, C. D., D. E. Soltis, and P. S. Soltis. 2005. The age of the angiosperms: A molecular time-scale without a clock. *Evolution* 59: 1245-1258.
- Cantino, P. D., J. A. Doyle, S. W. Graham, W. S. Judd, R. G. Olmstead, P. S. Soltis, D. E. Soltis, and M. J. Donoghue. 2007. Towards a phylogenetic nomenclature of *Tracheophyta*. *Taxon* 56: 822-846.
- Cui, L., P. K. Wall, J. Leebens-Mack, B. G. Lindsay, D. Soltis, J. J. Doyle, P. Soltis, J. Carlson, A. Arumuganathan, A. Barakat, V. Albert, H. Ma, and C. W. DePamphilis. 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Research* 16: 738-749.
- Davies, T. J., T. G. Barraclough, M. W. Chase, P. S. Soltis, D. E. Soltis, and V. Savolainen. 2004. Darwin's abominable mystery: Insights from a supertree of the angiosperms. *Proceedings of the National Academy of Sciences, USA* 101: 1904-1909.
- Frohlich, M. W. and M. W. Chase. 2007. After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature* 450: 1184-1189.
- Howarth, D. G. and M. J. Donoghue. 2006. Phylogenetic analyses of the "ECE" (CYC/TB1) clade reveal duplications that predate the core eudicots. *Proc. Nat. Acad. Sci. USA* 103: 9101-9106.
- Jansen, R. K., et al. 2007. Analysis of 81 genes from 64 chloroplast genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences, USA* 104: 19369–19374 .
- McMahon, M. M., and M. J. Sanderson. 2006. Phylogenetic supermatrix analysis of GenBank sequences from 2228 papilionoid legumes. *Syst. Biol.*, 55:818-836.
- Moore, M. J., C. D. Bell, P. S. Soltis, and D. E. Soltis. 2007. Using plastid genomic-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences USA* 104: 19363–19368.
- Qiu, Y-L., et al. 2006. The deepest divergences in land plants inferred from phylogenomic evidence. *Proceedings of the National Academy of Sciences, USA* 103: 15511-15516.
- Soltis, D. E., P. S. Soltis, M. W. Chase, and P. Endress. 2005. *Phylogeny and Evolution of Angiosperms*. Sinauer Associates, Sunderland, MA.
- Soltis, D. E., M. Gitzendanner, and P. S. Soltis. 2007. A Bayesian analysis of the three-gene, 567-taxon data set for angiosperms. *International Journal of Plant Sciences* 168: 137-157.

### *Phyloinformatics*

- Driskell, A. C., C. Ané, J. G. Burleigh, M. M. McMahon, B. O'Meara, and M. J. Sanderson. 2004. Prospects for building the tree of life from large sequence databases. *Science*, 306:1172-1174.
- Dutilh, B.E., V. Van Noort, R. Van Der Heijden, T. Boekhout, B. Snel, and M. A. Huynen. 2007. Assessment of phylogenomic and orthology approaches for phylogenetic inference. *Bioinformatics* 23: 815-824.
- Hartmann S, Lu D, Phillips J, Vision TJ (2006) Phytome: a platform for plant comparative genomics. *Nucleic Acids Research* 34, D724-730.
- Hartmann S, Vision TJ (2008) Using ESTs for phylogenomics: can one accurately infer a phylogenetic tree from a gappy alignment? *BMC Evolutionary Biology* 8, 95.
- Hibbett, D. S., R. H. Nilsson, M. Snyder, M. Fonseca, J. Costanzo, and M. Shonfeld. 2005. Automated phylogenetic taxonomy: An example in the homobasidiomycetes (mushroom-forming fungi). *Systematic Biology* 54: 660-668.
- Huson, D. H., D. C. Richter, C. Rausch, T. DeZulian, M. Franz, and R. Rupp. 2007. Dendroscope: An interactive viewer for large phylogenetic trees. *BioMed Central Bioinformatics* 8(1): 460.
- Jordan, G. E. and W. H. Piel. 2008. PhyloWidget: web-based visualizations for the tree of life. *Bioinformatics*, doi:10.1093/bioinformatics/btn235.
- Kauff, F., C. J. Cox, and F. Lutzoni. 2007. WASABI: An Automated Sequence Processing System for Multigene Phylogenies. *Systematic Biology* 56: 523-531.

- Lahaye, R., M. van der Bank, D. Bogarin, J. Warner, F. Pupulin, G. Gigot, O. Maurin, S. Duthoit, T. G. Barraclough, and V. Savolainen. 2008. DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences, USA* 105: 2923-2928.
- Leebens-Mack J, and 26 others (2006) Taking the first steps towards a standard for reporting in phylogenies: Minimal Information About a Phylogenetic Analysis (MIAPA). *OMICS* 10, 231-237.
- Munzner, T., F. Guimbretiere, S. Tasiran, L. Zhang, and Y. Zhou. 2003. Treejuxtaposer: Scalable tree comparison using focus+context with guaranteed visibility. *ACM Transactions on Graphics* 22(3): 453-462.
- Nakhleh, L., D. Miranker, F. Barbancon, W. Piel, and M. Donoghue. 2003. Requirements of phylogenetic databases. In Proc. 3<sup>rd</sup> IEEE Symp. on Bioinformatics and Bioengineering (BIBE'03): 141-148.
- Piel, W. H., M. J. Sanderson, and M. J. Donoghue. 2003. The small-world dynamics of tree networks and data mining in phyloinformatics. *Bioinformatics* 19: 1162-1168.
- Sanderson, M. J. 2006. Paloverde: an OpenGL 3D phylogeny browser. *Bioinformatics* 22: 1004-1006.
- Sanderson, M. J. Construction and annotation of large phylogenetic trees. 2007. *Australian Syst. Bot.* 4:287-301.
- Sanderson, M. J., D. Boss, D. Chen, K. A. Cranston, and A. Wehe. 2008. The PhyLoTA Browser: processing GenBank for molecular phylogenetics research. *Syst. Biol.* In press.
- Stamatakis, A., T. Ludwig, and H. Meier. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21: 456-463.
- Webb, C. O. and M. J. Donoghue. 2004. Phylomatic: tree assembly for applied phylogenetics. *Molecular Ecology Notes* 5: 181-183.
- Webb, C. O., D. D. Ackerly, and S. W. Kembel. 2007. Phylocom: software for the analysis of community phylogenetic structure and trait evolution. Version 3.41. <http://www.phylodiversity.net/phylocom/>.
- Zwickl, D. J. 2006. GARLI Genetic algorithm for rapid likelihood inference. Software available at <http://www.bio.utexas.edu/grad/zwickl/web/garli.html>.

#### *Selected Phyloinformatic Web Resources*

- TreeBASE; a database of phylogenetic knowledge: <http://www.treebase.org>
- PhyLoTA; tools to build the Tree of Life from sequence databases: <http://loco.biosci.arizona.edu/phyloata/>
- TOLKIN; a phylodiversity information management web application: <http://www.tolkin.org/>
- CIPRES; Cberinfrastructure for Phylogenetic Research: <http://www.phylo.org/>
- AToL; Assmebling the Tree of Life: <http://atol.sdsc.edu/>
- APWeb; angiosperm phylogeny website: <http://www.mobot.org/MOBOT/research/APweb/>
- Mesquite; software for phylogenetic research: <http://mesquiteproject.org/>
- TOLWeb; a tree of life web project: <http://www.tolweb.org/>
- MorphBank: <http://www.morphbank.net>
- Paleobiology DataBase: <http://www.paleodb.org>
- Phytome: <http://www.phytome.org>
- Phyutility; tools for phylogenetic editing and analysis: <http://code.google.com/p/phyutility/>
- Barcoding: <http://www.barcodinglife.org>
- Phylomatic; tree assembly for applied phylogenetics: <http://www.phylodiversity.net/phyloomatic/>
- pPod; core technologies to enable the integration of AToL information: <http://www.phylodata.org>
- PaloVerde; a 3-D tree visualization tool: <http://loco.biosci.arizona.edu/paloverde/paloverde.html>
- Dendroscope: <http://www-ab.informatik.uni-tuebingen.de/software/dendroscope/>
- PhyloWidget; large tree visualization and editing: <http://www.phylowidget.org/>

## POTENTIAL REVIEWERS

**Michael Clegg**, Ecology and Evolutionary Biology, School of Biological Sciences, University of California, Irvine. Phone: (949) 824-4490. Email: [mclegg@uci.edu](mailto:mclegg@uci.edu)

**Peter Endress**, Institute of Systematic Botany, University of Zurich, Switzerland. Email: [pendress@systbot.uzh.ch](mailto:pendress@systbot.uzh.ch)

**David Hillis**, Section of Integrative Biology, University of Texas, Austin. Phone: (512) 71-5792. E-Mail: [dhillis@mail.utexas.edu](mailto:dhillis@mail.utexas.edu)

**Ralph Quatrano**, Department of Biology, Washington University, St. Louis. Phone: (314) 935-6850. Email: [rsq@wustl.edu](mailto:rsq@wustl.edu)

**Peter Raven**, Missouri Botanical Garden, St. Louis, MO. Phone: (314) 577-5111. E-Mail: [peter.raven@mobot.org](mailto:peter.raven@mobot.org)

**Barbara Schaal**, Department of Biology, Washington University, St. Louis. Phone: (314) 935-6822. Email: [schaal@biology.wustl.edu](mailto:schaal@biology.wustl.edu)

**Tod Stuessy**, University of Vienna, Austria. E-Mail: [tod.stuessy@univie.ac.at](mailto:tod.stuessy@univie.ac.at)

**Virginia Walbot**, Department of Biology, Stanford University. Phone: (650) 723-2007. Email: [walbot@stanford.edu](mailto:walbot@stanford.edu)

*Note: owing to long-standing conflicts with several of the PI's, if possible please avoid sending this for review to K. Nixon and others associated with the Bailey Hortorium of Cornell University.*