OMICS A Journal of Integrative Biology
Volume 10, Number 2, 2006

Taking the First Steps towards a Standard for Reporting on Phylogenies: Minimum Information about a Phylogenetic Analysis (MIAPA)

JIM LEEBENS-MACK,[1,*] TODD VISION,[2] ERIC BRENNER,[3] JOHN E. BOWERS,[4] STEVEN CANNON,[5] MARK J. CLEMENT,[6] CLIFFORD W. CUNNINGHAM,[7] CLAUDE dePAMPHILIS,[1] ROB deSALLE,[8] JEFF J. DOYLE,[9] JONATHAN A. EISEN,[10] XUN GU,[11] JOHN HARSHMAN,[12] ROBERT K. JANSEN,[13] ELIZABETH A. KELLOGG,[14] EUGENE V. KOONIN,[15] BRENT D. MISHLER,[16] HERVÉ PHILIPPE,[17] J. CHRIS PIRES,[18] YIN-LONG QIU,[19] SEUNG Y. RHEE,[20] KIMMEN SJÖLANDER,[21] DOUGLAS E. SOLTIS,[22] PAMELA S. SOLTIS,[23] DENNIS W. STEVENSON,[3] KERR WALL,[1] TANDY WARNOW,[24] and CHRISTIAN ZMASEK[25]

[1]Department of Biology, Institute of Molecular Evolutionary Genetics, and Huck Institutes of Life Sciences, Pennsylvania State University, University Park, Pennsylvania.
[2]Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.
[3]New York Botanical Garden, Bronx, New York.
[4]Applied Genetic Technology Center, Departments of Crop and Soil Science, Botany, and Genetics, University of Georgia, Athens, Georgia.
[5]USDA-ARS, and Department of Agronomy, Iowa State University, Ames, Iowa.
[6]Networked Computing Laboratory, Computer Science Department, Brigham Young University, Provo, Utah.
[7]Department of Biology, Duke University, Durham, North Carolina.
[8]Division of Invertebrate Zoology, American Museum of Natural History, New York, New York.
[9]L.H. Bailey Hortorium, Cornell University, Ithaca, New York.
[10]UC Davis Genome Center, Section of Evolution and Ecology, and the Department of Medical Microbiology and Immunology, University of California, Davis, California.
[11]Department of Genetics, Development and Cell Biology, Iowa State University, Ames, Iowa.
[12]Pepperwood Way, San Jose, California.
[13]Section of Integrative Biology and Institute of Cellular and Molecular Biology, University of Texas, Austin, Texas.
[14]Department of Biology, University of Missouri, Saint Louis, Missouri.
[15]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland.
[16]University Herbarium, Jepson Herbarium, and Department of Integrative Biology, University of California, Berkeley, California.
[17]Canadian Institute for Advanced Research, Centre Robert Cedergren, Departement de Biochimie, Universite de Montreal, Succursale Centre-Ville, Montreal, Canada.
[18]Division of Biological Sciences, University of Missouri–Columbia, Columbia, Missouri.
[19]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan.
[20]Department of Plant Biology, Carnegie Institution, Stanford, California.
[21]Department of Bioengineering, University of California, Berkeley, California.
[22]Department of Botany and the Genetics Institute, University of Florida, Gainesville, Florida.
[23]Florida Museum of Natural History and the Genetics Institute, University of Florida, Gainesville, Florida.
[24]Department of Computer Science, University of Texas at Austin, Austin, Texas.
[25]Burnham Institute for Medical Research, La Jolla, California.
[*]*New address:* Department of Plant Biology, University of Georgia, Athens, Georgia 30602.

ABSTRACT

In the eight years since phylogenomics was introduced as the intersection of genomics and phylogenetics, the field has provided fundamental insights into gene function, genome history and organismal relationships. The utility of phylogenomics is growing with the increase in the number and diversity of taxa for which whole genome and large transcriptome sequence sets are being generated. We assert that the synergy between genomic and phylogenetic perspectives in comparative biology would be enhanced by the development and refinement of minimal reporting standards for phylogenetic analyses. Encouraged by the development of the Minimum Information About a Microarray Experiment (MIAME) standard, we propose a similar roadmap for the development of a Minimal Information About a Phylogenetic Analysis (MIAPA) standard. Key in the successful development and implementation of such a standard will be broad participation by developers of phylogenetic analysis software, phylogenetic database developers, practitioners of phylogenomics, and journal editors.

This paper is part of the special issue of OMICS on data standards.

INTRODUCTION

PHYLOGENIES HAVE PROVIDED a historical framework for interpreting the evolution of form and function since Darwin (1859) and Haeckel (1866) published their iconic tree figures some 150 years ago. In recent years, phylogenetics has come to play a multifaceted role in genomic analyses and interpretation of genomics data. Phylogenetic analyses are now being performed on a genomic scale in order to address issues ranging from the prediction of gene and protein function (Eisen, 1998; Sjölander, 2004; Engelhardt et al., 2005) to organismal relationships (Philippe et al., 2005; Delsuc et al., 2005), to the influences of polyploidy (Bowers et al., 2003; Byrne and Wolfe, 2005) and horizontal gene transfer (Ge et al., 2005; Simonson et al., 2005) on genome content and structure (Wolf et al., 2002), to the reconstruction of ancestral genome characteristics (Blanchette et al., 2004).

The fundamental nature of inferences drawn from all of these applications underscores the growing importance of genomic and sub-genomic investigations of species covering the spectrum of organismal diversity. Phylogenomic analyses, defined here broadly as the integration of phylogenetic and genomic analysis (Eisen and Fraser, 2003), place genome sequence, gene expression (Gu and Gu, 2003; Gu 2004; Gu et al., 2005; Duarte et al., 2006) and functional data in a historical context and thereby help to elucidate those processes shaping the structure and function of genes, genetic systems and whole organisms. The development and refinement of searchable phylogeny databases such as TreeBase (Piel et al., 2003) or gene tree databases (Duret et al., 1994; Sjölander, 2004; Roth et al., 2005; Hartmann et al., 2006; Li et al., 2006) is an important step in the advancement of phylogenomics, but only a miniscule fraction of published phylogenies are currently deposited in a database. What is worse, the alignments for many published phylogenies are not easily accessible, and methods of analysis are not adequately described. These are serious impediments to those wanting to test the robustness of published phylogenies, conduct cross-study comparisons of phylogenetic inferences, or draw new inferences from meta-analyses.

Accurate phylogenetic trees provide a valuable historical context for a variety of comparative analyses, and can be applied to a host of biological questions unforeseen by the original authors. This is particularly true in phylogenomics, where many applications require the investigation of phylogenetic trees for a large number of independent gene/protein families. If inadequately documented, however, even the most carefully constructed phylogenetic analysis will languish in the pages of a journal. Thus, a key step in the continued ability of phylogenomics to take full advantage of the rapidly expanding volume of sequence data will be the development of reporting standards for phylogenetic analyses, along with databases from which these metadata can easily be retrieved. In this paper, we propose a roadmap to develop a set of reporting standards

for phylogenetic analyses. Using the MIAME standard (Brazma et al., 2001) as a model, we call for a community-wide effort to develop a Minimal Information About a Phylogenetic Analysis (MIAPA) standard.

CONSIDERATIONS FOR DEVELOPING STANDARDS FOR REPORTING PHYLOGENETIC ANALYSES

The papers in this special issue constitute a series of case studies on the importance of standard practices for reporting the results of various types of experiments in a way that facilitates the ability of scientists to use these data in subsequent studies (Field and Sansone, *this issue*). The motivating question behind the MIAME standard for microarray experiments was as follows: "What is the minimum information necessary for an independent scientist to carry out an independent analysis of the data?" (Quackenbush, 2005).

The motivation for the MIAPA standard is the same, as is the challenge: minimizing the reporting requirements while maximizing the information available to those interpreting the results of a study (Brazma, 2001; Brazma et al., 2001; Ball and Brazma, *this issue*). The phylogenetics community is coming together to develop this standard with careful consideration of the types of future analyses that are likely to be performed and the data required. For example, systematists may combine pre-existing phylogenies into supertree analyses (Davies et al., 2004; Page, 2005), while genomicists may combine them to investigate the timing of genome duplication events (Chapman et al., 2004). At the same time, investigators may require access to the alignments and component sequences used to build the selected phylogenies in order to perform independent phylogenetic analyses on single or combined datasets. Thus, just as the MIAME standard was designed to accommodate the nested organization of gene expression levels derived from signal quantification matrices derived in turn from raw image data (Brazma et al., 2001), the MIAPA standard would need to accommodate phylogenies derived from analysis of alignments derived in turn from raw sequence data.

A decision that was integral to development and success of the MIAME standards was that they should be applicable to a wide variety of microarray technologies and no one platform or hybridization protocol was prescribed. Similarly, we suggest that the MIAPA standard should be agnostic concerning methods of alignment and phylogenetic reconstruction. The diversity of methods of phylogenetic inference is perhaps even greater than the diversity of applications to which phylogenies may be applied (Swofford et al., 1996; Felsenstein, 2004; Delsuc et al., 2005) and novel methods are likely to be developed in the future. Parsimony, likelihood, Bayesian and distance-based approaches have all been adapted for analyses of the various data types relevant to phylogenomics, including aligned nucleotide and protein sequences, gene structure (insertions and deletions), gene content, motif frequencies (Qi et al., 2004) and gene order (Moret et al., 2001). Multiple sequence alignment has its own diverse set of methodologies, and, in some approaches, a multiple sequence alignment and phylogenetic tree are constructed simultaneously (Gladstein and Wheeler, 1997; Edgar and Sjölander, 2003; Lunter et al., 2005; Fleissner et al., 2005). The relative performance of these different methods is an area of active research, but it is clear that no single method is optimal for all data sets (Swofford et al., 2001; Spencer et al., 2005). Benchmark datasets have been compiled for comparing the performance of alignment algorithms (van Walle et al., 2004; Thompson et al., 2005) but there are few comparable benchmarks for phylogenetic algorithms, and so comparisons have relied largely on analyses of simulated or contrived data sets (Huelsenbeck 1995; Swofford et al., 2001; Spencer et al., 2005, but see Hillis et al., 1992; Cunningham et al., 1997). Thus, for a variety of reasons, methodological diversity in phylogenetics is likely to be the state of affairs for the foreseeable future. No matter how phylogenies are constructed, however, a comprehensive description of how a set of sequences was aligned, and how phylogenetic trees were derived from an alignment would allow researchers to evaluate their confidence in a phylogeny and run their own analyses if they see fit.

The six required components of the MIAME standards proposed in 2001 (Brazma et al., 2001) included descriptions of (1) the experimental design for a complete study; (2) the design of each array and the identity of each spot on the arrays used in the study; (3) the biological sample extraction preparations and labelling procedures used for each hybridization; (4) the hybridization protocols; (5) the measurements, including imaging and signal quantification parameters; and (6) the normalization and control information. At this stage, it would be premature to specify the details of the MIAPA standard, but Figure 1 offers a start-

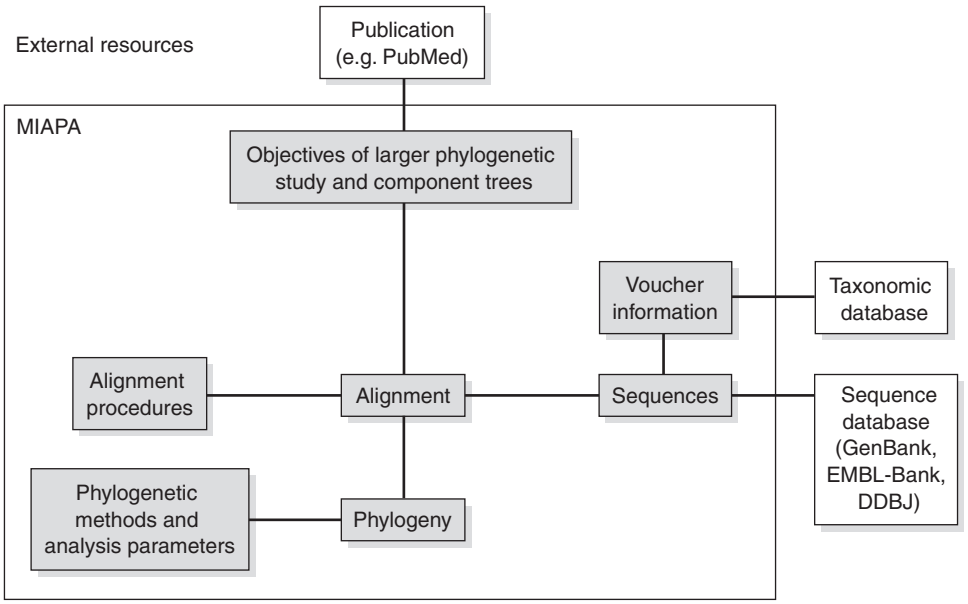

FIG. 1. A schematic diagram showing the components of a phylogenetic analysis that could be included in a minimal reporting standard.

ing point for considering MIAPA's essential components that it might include. By analogy with the MIAME standards, minimum reporting standards for phylogenetic analyses are likely to include (1) a description of the objectives of the phylogenetic analysis and the component trees included in a study (many phylogenetic studies produce multiple trees based on different data sets or analytical methods); (2) the raw sequences or character descriptions; (3) sample voucher information; (4) a description of procedures for establishing character homology (e.g., sequence alignment); (5) the sequence alignment or some other character matrix; (6) detailed description of the phylogenetic analysis, including search strategies and parameter values (specific commands for the analysis program would be optimal); and (7) the phylogenies including branch lengths and support values (e.g., bootstrap). The schematic shown in Figure 1 is likely to be incomplete. For example, it is not clear whether or how to report measures of node support, such as bootstrap values, and phylogenetic analyses are often performed on data matrices other than nucleotide and protein sequence alignments. If the reporting standard were focused on sequence data, referencing an external database for the unaligned and unmasked sequences would require that all sequence identifiers in a database such as GenBank would be stable over the long term. If the standard were to extend to phylogenetic analyses of morphological characters, character descriptions and data matrices could be deposited in MorphBank (⟨www.morphbank.com⟩) or MorphoBank (⟨www.morphobank.org⟩). Following the MIAME model (Brazma et al., 2001), the scheme in Figure 1 is reliant on an external database or other such databases (e.g., the taxonomy database at NCBI) for information about the taxonomic placement of the studied organisms. However, it might be better to require the full taxonomy of the studied organisms to be reported in order to allow a full search of the taxonomic hierarchy (Page, 2005). We suggest that sample voucher information be included in the reporting standard in order to properly synthesize future combined data matrices or build supertrees. The phylogenetics community will have to grapple with these issues and more as we formalize the reporting standard, and we reiterate that Figure 1 is presented simply as a starting point for deeper consideration.

DEFINING A ROADMAP FOR THE DEVELOPMENT OF MIAPA STANDARDS

The nearly universal adoption of the MIAME standards and their impact on all aspects of microarray-based expression profiling was driven by necessity. The deliberate process by which they were constructed

started with an international meeting of what became the Microarray Gene Expression Data Society (⟨www.mged.org⟩) in 1999 and culminated in an open letter first published in Nature Genetics in 2001. There has been continued refinement of the standards at annual meetings (Ball and Brazma, *this issue*). Much of the success of MIAME must also be attributed to the fact that MGED engaged commercial interests and database managers. The compliance of microarray databases (Parkinson et al., 2005; Barrett et al., 2005) was facilitated by the development of formal protocols for data exchange, namely the microarray gene expression object model (MAGE-OM) implemented in XML (MAGE-ML) (Spellman et al., 2002). User-friendly systems for submission of expression data and metadata that built upon these protocols (Mukherjee et al., 2005) have further promoted widespread compliance with MIAME guidelines among investigators.

Development of the MIAPA standard must also involve developers of phylogenetic analysis software (Felsenstein, 2005; Goloboff et al., 2004; Kumar et al., 2004; Ronquist and Huelsenbeck, 2003; Swofford, 2001; Roshan et al., 2004; ⟨www.phylo.org⟩), existing public databases for organismal (Piel et al., 2003) and gene family phylogenies (Duret et al., 1994; Sjölander, 2004; Roth et al., 2005; Hartmann et al., 2006; Li et al., 2006), as well as editors of the journals in which phylogenetic analyses are published. A well-defined protocol for saving and transferring phylogenetic metadata should be considered, one that would complement existing formats such as New Hampshire (or Newick) and PhyloXML (⟨www.phyloxml.org⟩).

Following the example of MGED, development of the MIAPA standard could be advanced through an international conference of representative stakeholders in conjunction with open discussions across the phylogenetics community. We will be soliciting involvement in an organizational conference at scientific meetings this coming summer and publishing proposals for the MIAPA standard in the journals most read by the phylogenetics community. We anticipate these efforts will culminate in an open letter to the editors of all journals publishing phylogenies in which MIAPA will be described in detail. In addition, the standard would be most viable if accompanied by software and database tools that would facilitate utility and widespread compliance.

CONCLUSION

These are ambitious objectives, but the time is ripe for the development and implementation of minimal reporting standards for phylogenetic analyses. Widespread recognition of the importance of phylogenetics to genome biology comes at a time when recent advances have increased the rate of sequence generation by orders of magnitude (Margulies et al., 2005). Increases in sequencing capacity and concomitant cost decreases are spurring a rapid expansion in the availability of whole genome sequences (Liolios et al., 2006) or subgenomic sequence data (Lee et al., 2005). Beyond doubt, this flood of sequence data will spur a corresponding flood of comparative analyses in which phylogenetic trees play a central role. Indeed, many computational and statistical methods for functional genomic analysis are being developed, which are, more or less, phylogeny-based. When reporting of phylogenetic analyses is brought more fully into the informatics age, it will have manifold beneficial effects on the utility and impact of phylogenomics.

ACKNOWLEDGMENTS

We thank Dawn Field and Peter Sterk for organizing the "Cataloguing our Current Genome Collection" workshop at EBI, where some of these ideas were first proposed. We also thank Dawn Field, Susanna Sansone, and Eugene Kolker for their roles in putting together this special issue.

REFERENCES

BALL, C.A., and BRAZMA, A. (2006). MGED standards: work in progress OMICS (*this issue*).

BARRETT, T., SUZEK, T.O., TROUP, D.B., et al. (2005). NCBI GEO: mining millions of expression profiles—database and tools. Nucleic Acids Res **33,** D562–D566.

BLANCHETTE, M., GREEN, E.D., MILLER, W., et al. (2004). Reconstructing large regions of an ancestral mammalian genome *in silico*. Genome Res **14,** 2412–2423.

BOWERS, J.E., CHAPMAN, B.A., RONG, J., et al. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **422,** 433–438.

BRAZMA, A. (2001). On the importance of standardisation in life sciences. Bioinformatics **17,** 113–114.

BRAZMA, A., HINGAMP, P., QUACKENBUSH, J., et al. (2001). Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. Nat Genet **29,** 365–371.

BYRNE, K.P., and WOLFE, K.H. (2005). The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res **15,** 1456–1461.

CHAPMAN, B.A., BOWERS, J.E., SCHULZE, S.R., et al. (2004). A comparative phylogenetic approach for dating whole genome duplication events. Bioinformatics **20,** 180–185.

CUNNINGHAM, C.W., JENG, K., HUSTI, J., et al. (1997). Parallel molecular evolution of deletions and nonsense mutations in bacteriophage T7. Mol Biol Evol **14,** 113–116.

DARWIN, C. (1859). *The Origin of Species by Means of Natural Selection* (John Murray, London).

DAVIES, T.J., BARRACLOUGH, T.G., CHASE, M.W., et al. (2004). Darwin's abdominable mystery: insights from a supertree of the angiosperms. Proc Natl Acad Sci USA **101,** 1904–1909.

DELSUC, F., BRINKMANN, H., and PHILIPPE, H. (2005). Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet **6,** 361–375.

DUARTE, J.M., CUI, L., WALL, P.K., et al. (2006). Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of Arabidopsis. Mol Biol Evol **23,** 469–478.

EDGAR, R.C., and SJOLANDER, K. (2003). SATCHMO: sequence alignment and tree construction using hidden Markov models. Bioinformatics **19,** 1404–1411.

EISEN, J.A. (1998). Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. Genome Res **8,** 163–167.

EISEN, J.A., and FRASER, C.M. (2003). Phylogenomics: intersection of evolution and genomics. Science **300,** 1706–1707.

ENGELHARDT, B.E., JORDAN, M.I., MURATORE, K.E., et al. (2005). Protein molecular function prediction by Bayesian phylogenomics. PLoS Comput Biol **1,** e45.

FIELD, D., and SANSONE, S.-A. (2006). A special issue on data standards. OMICS (*this issue*).

FELSENSTEIN, J. (2004). *Inferring Phylogenies* (Sinauer Associates, Sunderland, MA).

FELSENSTEIN, J. (2005). *PHYLIP* (*Phylogeny Inference Package*), version 3.6 (distributed by author, Department of Genome Sciences, University of Washington, Seattle, Washington).

FLEISSNER, R., METZLER, D., and VON HAESELER, A. (2005). Simultaneous statistical multiple alignment and phylogeny reconstruction. Syst Biol **54,** 548–561.

GE, F., WANG, L.S., and KIM, J. (2005). The cobweb of life revealed by genome-scale estimates of horizontal gene transfer. PLoS Biol **3,** e316.

GLADSTEIN, D.S., and WHEELER, W.C. (1997). POY: the optimization of alignment characters. Available at: ⟨ftp.amnh.org/pub/molecular⟩.

GOLOBOFF, P.A., FARRIS, J.S., and NIXON, K.C. (2004). *TNT Tree Analysis Using New Technology*, version 1.0. Available at: ⟨www.cladistics.com⟩.

GU, J., and GU, X. (2003). Induced gene expression in human brain after the split from chimpanzee. Trend Genet. **19,** 63–65.

GU, X. (2004). Statistical framework for phylogenetic analysis of expression profiles. Genetics **167,** 531–542.

GU, X., ZHANG, Z., and HUANG, W. (2005). Rapid evolution of expression and regulatory network after yeast gene/genome duplications. Proc Natl Acad Sci USA **102,** 707–712.

HAEKEL, F. (1866). *Generelle Morphologie der Organismen* (G. Reimer, Berlin).

HARTMANN, S., LU, D., PHILLIPS, J., et al. (2006). Phytome: a platform for plant comparative genomics. Nucleic Acids Res **34,** D724–D730.

HILLIS, D.M., BULL, J.J., WHITE, M.E., et al. (1992). Experimental phylogenetics: generation of a known phylogeny. Science **255,** 589–592.

HUELSENBECK, J.P. (1995). The robustness of two phylogenetic methods: four-taxon simulations reveal a slight superiority of maximum likelihood over neighbor joining. Mol Biol Evol **12,** 843–849.

KUMAR, S., TAMURA, K., and NEI, M. (2004). MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform **5,** 150–163.

LEE, Y., TSAI, J., SUNKARA, S., et al. (2005). The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. Nucleic Acids Res **33,** D71–D74.

LI, H., COGHLAN, A., RUAN, J., et al. (2006). TreeFam: a curated database of phylogenetic trees of animal gene families. Nucleic Acids Res **34,** D572–D580.

LIOLIOS, K., TAVERNARAKIS, N., HUGENHOLTZ, P., et al. (2006). The Genomes On Line Database (GOLD) v.2: a monitor of genome projects worldwide. Nucleic Acids Res **34,** D332–D334.

LUNTER, G., MIKLOS, I., DRUMMOND, A., et al. (2005). Bayesian coestimation of phylogeny and sequence alignment. BMC Bioinform **6,** 83.

MARGULIES, M., EGHOLM, M., ALTMAN, W.E., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature **437,** 376–380.

MORET, B.M., WANG, L.S., WARNOW, T., et al. (2001). New approaches for reconstructing phylogenies from gene order data. Bioinformatics **17,** S165–S173.

MUKHERJEE, G., ABEYGUNAWARDENA, N., PARKINSON, H., et al. (2005). Plant-based microarray data at the European Bioinformatics Institute. Introducing AtMIAMExpress, a submission tool for *Arabidopsis* gene expression data to ArrayExpress. Plant Physiol **139,** 632–636.

PAGE, R.D.M. (2005). Towards a taxonomically intelligent phylogenetic database. Technical Reports in Taxonomy 04-01, presented at DBiBD at Edinburgh.

PARKINSON, H., SARKANS, U., SHOJATALAB, M., et al. (2005). ArrayExpress—a public repository for microarray gene expression data at the EBI. Nucleic Acids Res **33,** D553–D555.

PHILIPPE, H., SNELL, E.A., BAPTESTE, E., et al. (2004). Phylogenomics of eukaryotes: impact of missing data on large alignments. Mol Biol Evol **21,** 1740–1752.

PIEL, W.H., SANDERSON, M.J., and DONOGHUE, M.J. (2003). The small-world dynamics of tree networks and data mining in phyloinformatics. Bioinformatics **19,** 1162–1168.

QI, J., WANG, B., and HAO, B.I. (2004). Whole proteome prokaryote phylogeny without sequence alignment: a K-string composition approach. J Mol Evol **58,** 1–11.

QUACKENBUSH, J. (2005). Extracting meaning from functional genomics experiments. Toxicol Appl Pharmacol **207,** 195–199.

ROSHAN, U.W., MORET, B.M., WARNOW, T., et al. (2004). Rec-I-DCM3: a fast algorithmic technique for reconstructing large phylogenetic trees. Proc IEEE Comput Syst Bioinform Conf 98–109.

ROTH, C., BETTS, M.J., STEFFANSSON, P., et al. (2005). The Adaptive Evolution Database (TAED): a phylogeny based tool for comparative genomics. Nucleic Acids Res **33,** D495–D497.

SIMONSON, A.B., SERVIN, J.A., SKOPHAMMER, R.G., et al. (2005). Decoding the genomic tree of life. Proc Natl Acad Sci USA **102,** 6608–6613.

SJOLANDER, K. (2004). Phylogenomic inference of protein molecular function: advances and challenges. Bioinformatics **20,** 170–179.

SPELLMAN, P.T., MILLER, M., STEWART, J., et al. (2002). Design and implementation of microarray gene expression markup language (MAGE-ML). Genome Biol **3,** RESEARCH0046.

SPENCER, M., SUSKO, E., and ROGER, A.J. (2005). Likelihood, parsimony, and heterogeneous evolution. Mol Biol Evol **22,** 1161–1164.

SWOFFORD, D.L. (2003). *PAUP*: Phylogeneic Analyses Using Parsimony* (** and Other Methods*) (Sinauer Associates, Sunderland, MA).

SWOFFORD, D.L., OLSEN, G.J., WADDELL, P.J., et al. (1996). Phylogenetic inference. In *Molecular Systematics*, 2nd ed. D.M. Hillis, C. Moritz, and B.K. Mable, eds. (Sinauer Associates, Sunderland, MA), pp. 407–514.

SWOFFORD, D.L., WADDELL, P.J., HUELSENBECK, J.P., et al. (2001). Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. Syst Biol **50,** 525–539.

THOMPSON, J.D., KOEHL, P., RIPP, R., et al. (2005). BAliBASE 3.0: latest developments of the multiple sequence alignment benchmark. Proteins **61,** 127–136.

VAN WALLE, I., LASTERS, I., and WYNS, L. (2005). SABmark—a benchmark for sequence alignment that covers the entire known fold space. Bioinformatics **21,** 1267–1268.

WOLF, Y.I., ROGOZIN, I.B., GRISHIN, N.V., et al. (2002). Genome trees and the tree of life. Trends Genet **18,** 472–479.

Address reprint requests to:
Dr. Jim Leebens-Mack
Department of Plant Biology
University of Georgia
Athens, GA 30602

E-mail: jleebensmack@plantbio.uga.edu