SCHOOL OF BIOLOGICAL SCIENCES
THE UNIVERSITY OF TEXAS AT AUSTIN

A6500 • Austin, Texas 78712 • (512) 232-3691 • FAX (512) 232-3699

November 22, 2000

Dear Tree of Life Workshop Participants:

Thank you for making time in your busy schedules to attend the *Tree of Life Workshop* in Austin, Texas. We anticipate a very full, productive, and hopefully enjoyable two days here in the heart of Texas.

By this time, you should have received itinerary information and/or confirmation numbers for your air travel arrangements and hotel accommodations. If you have not received your air & hotel confirmation numbers, please contact me as soon as possible.

Enclosed you will find a workshop itinerary, a participant list, and some updated information regarding computer access and entertainment plans. Please take a moment to read these materials. We suggest that you bring this packet of information with you as you travel as it contains information that could prove helpful.

Please plan on congregating in the hotel lobby by 7:55a.m. on Sunday, December 3rd. We will have two 15-passenger vans ready to shuttle you and the rest of our guests to the University of Texas at Austin campus where the workshop will take place. We hope to have the first session begin by 8:30 that morning.

If you have any questions about the workshop or your travel arrangements, please do not hesitate to contact me. On behalf of everyone here at UT, we truly look forward to hosting you at the Tree of Life Workshop.

Regards,

Amy Y. Kim

Research Needs in Phyloinformatics

Final report from a workshop held at the
University of California, Davis, October 20 and 21, 2000

Sponsored by the National Science Foundation

SUMMARY

Phylogenetic trees provide a rigorous framework for all aspects of basic and applied comparative biology. They are also having a dramatic impact on biomedical research and other problems of direct relevance to human health and well-being, such as managing threats from invasive species. Recognition of the important scientific and societal benefits arising from phylogenetic knowledge is generating interest in reconstructing a comprehensive phylogenetic Tree of Life (**TOL**), a structure that might well include a million species or more. The unprecedented scope of such an enterprise prompted us to organize a workshop to address the informatics needs of such an effort. A group of 26 phylogenetic biologists and computer scientists met at UC Davis on October 20-21, 2000 to make specific recommendations about how to proceed. Discussion led to broad agreement on the mission and goals of phyloinformatics, the elements needed in a Phyloinformatics Infrastructure (**PII**), and recommendations about implementation strategies.

The primary mission of the PII should be the archiving of phylogenetic knowledge and syntheses of alternative views of the Tree of Life. This should permit interactive browsing of the TOL by a diverse audience of users with different needs and levels of expertise. For research scientists, sophisticated capabilities for phylogenetically driven queries should be included to permit prediction and inference in comparative studies. For educators and others involved in public outreach, innovative and intuitive visualizations of the TOL combined with rich multimedia content should be developed. Finally, ongoing assessment of the state of archived phylogenetic knowledge should help optimize the efficient reconstruction of the TOL.

Two main elements should comprise the PII. The first is a centralized primary database containing individual phylogenetic trees and the data and metadata that generated them. The second is a set of secondary databases and/or toolkits charged with the task of constructing synthetic views of the TOL based on the data archived in the primary database. The main point of contact of the primary database with the broader network of biological databases in the outside world will be through existing and future species name databases. However, an additional component of the primary database should be a database of higher taxon names (such as "Mammalia" or "mammal"), because these are directly associated with phylogenetic relationships.

Recommendations for implementation strategies that would facilitate the establishment of the PII included:

- Establishment of a PII Center charged with setting up and curating the primary database, and providing supporting staff
- Support of basic research in several areas that will require significant innovation. These include (but are not limited to) the problem of visualization of large trees and collections of trees, development of a phylogenetic query language, resolution and display of phylogenetic ambiguity and conflict, and development of strategies to handle synonymy and resolve conflicts among names for species and higher taxa
- Sponsorship of graduate and postdoctoral training programs in phyloinformatics, modeled along the lines of the NSF IGERT program

I. Background: The Tree of Life

Biology, biomedicine and bioinformatics are relying increasingly on phylogenetic trees to provide a common framework for comparative studies across a tremendous diversity of organisms. In July 2000 phylogenetic biologists met at Yale University[1] for the first of a series of workshops on the prospects for building a comprehensive Tree of Life (**TOL**) in the next 10-15 years—with perhaps 75% of living taxa in it (~1-1.5 million species). In reflecting on this ambitious goal, participants took note of the fact that the number of phylogenetic studies is increasing rapidly in response to the demand for phylogenies. The number of trees published is doubling every 5 years, and the number of sequences in GenBank that might be used to build trees is doubling even faster, roughly every year. The size and scope of individual trees are also increasing rapidly, as recent publications of trees with hundreds to thousands of species demonstrate. Both the trees and their underlying data comprise a vast and growing resource that, in total, represents our current understanding of the Tree of Life.

Unfortunately the informatics infrastructure necessary to support this burgeoning database has not kept pace with the rate of data accumulation. Only a small fraction of published trees have been archived, while the remaining trees and their data languish all but lost for many purposes. Necessary software tools have not been developed to take full advantage of existing data, and to permit integration with existing biological databases. Significant investment and innovation is crucial merely to keep up with current demand and prevent the continued loss of expensive and unique data. If existing data sets and trees are to contribute to the much larger project of building a comprehensive Tree of Life, they must be archived in such a way that they are accessible and maximally useful to both producers and consumers of phylogenetic trees. To accomplish this, a set of essential informatics problems must be solved.

Compared to other major scientific disciplines, the phylogenetics community has invested relatively little effort into database development. The scale of work discussed at the first Tree of Life workshop is vastly beyond anything that existing small-scale efforts

---

[1] Final report available at http://research.amnh.org/biodiversity/acrobat/tol_workshop_report.pdf

were designed to handle. For example, in the last 10 years the TreeBASE[2] database has accessioned about 1300 trees and their supporting data, covering 15,000 species. Data for aligned small subunit ribosomal DNA sequences for about 20,000 taxa are stored in the Ribosomal Database Project (RDP II)[3]. The Tree of Life Project[4] contains over 1700 pages of phylogenies and more than 11,000 taxa. While each of these projects fills a particular informatics niche, these efforts span barely 1% of known biodiversity, and it is unclear which, if any, of them might serve as the best model for a dramatically scaled-up database that would underlie a truly comprehensive Tree of Life. Following upon the recommendations of the first workshop, we organized a second meeting aimed at identifying the elements necessary for an appropriately broad informatics infrastructure.

II. Description of Workshop and Participants

We held a two day meeting at the University of California, Davis, on October 20-21, 2000. The purpose of the Davis workshop was to bring together phylogenetic biologists, computer scientists, and database experts to discuss the unique informatics challenges posed by any attempt to reconstruct the Tree of Life on a truly comprehensive scale.

The meeting consisted of a mixture of breakout sessions, brief reports, and plenary discussions. The first morning included reports from the Yale workshop, group discussion of the vision behind the Tree of Life initiative and its specific goals, and presentations by developers of existing database projects (TreeBASE, Tree of Life Project, Ribosomal Database Project [RDP II]), and databases with which phylogenetic databases might likely be integrated (National Biological Information Infrastructure[5] [NBII], and Global Biodiversity Information Facility[6] [GBIF]). An agenda for three 2-hour breakout sessions with six specific topics occupied the remainder of the meeting. Each topic was covered by two independent groups of 4-6 members, and all groups reported back to the entire workshop with brief oral and written summaries of their findings.

The following 26 people from disparate fields participated, along with 3 NSF observers (James Rodman, Grace Wyngaard, and Terry Yates).

Organizers

Michael J. Sanderson, University of California, Davis. Interests: phylogenetic theory

Meredith Lane, Academy of Natural Sciences, Philadelphia. Interests: plant systematics, bioinformatics

---

[2] http://herbaria.harvard.edu/treebase

[3] http://www.cme.msu.edu/RDP/html/index.html

[4] http://www.treeoflifeproject.org

[5] http://www.nbii.gov

[6] http://www.gbif.org

Systematists

Joel Cracraft (co-organizer of Yale Workshop), American Museum of Natural History. Interests: avian systematics, large data sets
Michael Donoghue (co-organizer of Yale workshop), Yale University. Interests: plant systematics, large data sets
Brad Shaffer, UC Davis. Interests: vertebrate systematics, conservation biology
Gavin Naylor, Iowa State University. Interests: vertebrate systematics, molecular evolution
Peter Cranston, UC Davis. Interests: insect systematics, biodiversity
Chuck Delwiche, University of Maryland. Interests: basal eukaryote systematics, genome evolution
Tom Bruns, UC Berkeley. Interests: fungal systematics
Mary McKitrick, Smith College. Interests: bird systematics
Dirk Redecker, UC Berkeley and UC Davis. Interests: fungal systematics

Phyloinformaticists and authors of phylogenetic software

Roderick D. M. Page, University of Glasgow. Interests: phylogenetic software (Component, TreeMap, and others) and theory, host-associate evolution
David R. Maddison, University of Arizona. Interests: phylogenetic databases and analysis (Tree of Life Project, MacClade), beetle systematics
Emilia Martins, University of Oregon. Interests: phylogenetic comparative methods and software (COMPARE), behavior
Bill Piel, University of Leiden. Interests: invertebrate systematics, phylogenetic databases (TreeBASE)
Tim Lilburn, Michigan State University. Interests: phylogenetic databases (Ribosomal Database Project II)

Database and data visualization experts

Michael Freeston, Kings College, Aberdeen. Interests: databases
Jim Gannon, Parabon Computation. Interests: distributed computing platforms
Susanne Chambers, Parabon Computation. Interests: databases
Hasan Jamil, Mississippi State University. Interests: expert systems, artifical intelligence
Peter Karp, SRI International. Interests: microbial bioinformatics
Anne Frondorf, USGS. Interests: biological databases
Gary Waggoner, USGS. Interests: biological databases
Carol Bult, Jackson Labs. Interests: genome databases
Chris Henze, NASA Ames Research Labs. Interests: 3D visualization, large databases, algorithms in computational biology
Tamara Munzner, Stanford University. Interests: visualization of large graphs and networks

III. Conclusions and Recommendations of the Workshop

A. Mission and Goals of a Phyloinformatics Infrastructure (PII)

Significant scientific and societal benefits come from organizing biological knowledge according to phylogenetic relationships. Specific examples include reconstructing the history of functional changes in gene and protein sequences linked to disease, identifying the place of origin of emerging infectious diseases and their vectors (e.g., hantaviruses, West Nile Virus) and tracing individual contact histories, identifying invasive species and reconstructing their geographic origins, and providing a comparative framework for bioinformatic databases such as GenBank.

As a framework for organizing basic information about all biological diversity, the potential user community for the PII is extremely broad, including scientists, educators, students in K-12, the university, and the general public. Considerable discussion at the workshop focussed on enumerating the user communities for the Tree of Life and identifying their specific needs. Below we summarize these in a series of mission statements.

- Archiving of phylogenetic knowledge and syntheses. The basic mission of the PII is archiving of phylogenetic trees and the raw data, methods and algorithms used to construct them. Every effort should be made to obtain all data already published in the literature, and all new data submitted for publication. This fundamental knowledge-base should be complemented by updated syntheses of the comprehensive TOL, derived from expert syntheses, explicit algorithms that continually analyze the data archive, or both. These latter projects will produce a continually expanding, constantly-updated TOL as one of its final products

- Browsing the Tree of Life. The most visible product of the PII will be a set of tools for interactively browsing the TOL. Interfaces will have to be constructed to provide users with vastly different levels of expertise a sense of the information contained in the TOL. Interactive and visually creative systems will be needed to permit users to navigate across a tree structure that will ultimately encompass a million nodes or more. Because the demands of an elementary school user are so different from those of a phylogenetics researcher, we anticipate a need for several final TOL's and browsing tools, or several ways of visualizing a single, comprehensive tree.

- Phylogenetically driven data mining. Phylogenetic trees allow predictions about poorly known species by virtue of their relationship to better known species. Extraction of data about biological diversity should be facilitated by use of queries to the TOL. Development of novel means to query phylogenetically organized data is essential. The TOL will be the source of phylogenetically driven queries that can be distributed to other biological databases or across the internet as a whole.

- Scientific prediction and inference. The PII should provide a simple application programming interface (API) to researchers in the scientific community who develop phylogenetic tools for inference and prediction. Examples include tools for estimating the age and place of origin of viral epidemics, or for predicting functions of genes or structures of proteins from information about related genes. Developing the PII in a flexible, open format to facilitate developers of novel software applications is essential.

- Knowledge assessment. As this archive of phylogenetic knowledge catches up with our current knowledge, it will be possible to use the database to identify weak links in the TOL that need to be strengthened. Poorly sampled groups will be highlighted, serving as an essential call for additional research efforts in these groups. This knowledge assessment should also identify pivotal points on the tree that require particularly intensive research efforts due to their critical information content, their lack of phylogenetic resolution, or both.

- Breadth of access. Delivering the Tree of Life in its most informative and inspiring form to the most people is a primary goal of this project. This should include but not be limited to standard WWW access to the data and trees. The project should strive to provide visually stimulating and interactive tools in a variety of venues, including public museums and school classrooms.

B. Components of a Phyloinformatics Infrastructure (Fig. 1)

Participants agreed on the broad outline of components needed for an informatics infrastructure to support the TOL. Below we describe these components; in the next section we discuss their implementation.

Primary database. A single **primary** database should form the core of the PII. Its responsibility would be archiving of the data associated with published phylogenetic studies. For each study this would include the raw data (organized in matrices of taxon names by characters) and the phylogenetic trees stemming from analyses of these data. We anticipate that ultimately tens- to hundreds of thousands of partially overlapping trees will reside in this database, providing the raw material for construction of alternative views of the Tree of Life. These individual trees are analogous to the individual sequences that must be spliced together to form large "contigs" in the human genome project. A subset of the necessary features of this database can be seen in the existing TreeBASE archive.

Various other data associated with these studies would also be included as available. Some involve further description of the organisms: images of the whole organisms or their features; textual descriptions of the characters or taxa, and so on. Other potentially useful data pertain to the trees, such as estimated values of support for various parts of the tree(s), lengths of branches, and estimates of the ages of nodes in the tree.

Data acquisition would emphasize direct electronic submission by authors of studies, but also a significant effort to archive a now extensive backlog of published studies. All participants agreed that submission must be a requirement of publication in major journals, as is currently the case for publication of DNA sequence data in GenBank or EMBL.

Secondary synthetic database. One or more **secondary** databases would be charged with the responsibility of synthesizing and reporting the Tree(s) of Life to various user communities. Since the primary database consists of a large number of relatively small, unconnected trees, the synthesis of composite trees is a logically separate effort. We considered two fundamental ways that such synthetic TOL's could be constructed. One involves explicit algorithms (so-called supertree methods) to construct a continuously-updated tree based solely on the input trees from the primary database. The other combines the input data and expert opinion on specific taxa to create a TOL. In much the same way as GenBank provides the raw data for more derivative databases like the HIV sequence database, or the RDP, which have more specific goals, our secondary databases would be designed according to competing strategies for building a TOL, and targeted to different end user communities. For example, the design of the existing "Tree of Life Project", which is rich in multimedia content, serves a particular niche of users who may want a polished expert perspective about the TOL. Other users will want trees complete with all of the details of their construction, statistical support, redundancy between trees, quality of the original data, and a host of other types of data and metadata.

Data submission gateway. The ultimate success of an archive of phylogenetic trees will depend on the commitment of the scientific community to direct electronic submission of data, as is the case for the sequence databases. Tools for electronic submission of sequence data have improved dramatically in recent years, and a necessary element of the PII must be the development of a simple submission procedure and associated software. This is a context in which human curation will be essential (see below).

Interoperability with other biological databases. The primary and secondary databases must interoperate with several important classes of existing biological databases (Fig. 1). This will necessitate the invention and refinement of metadata standards to permit integration of different types of data. Most important is interoperability with species name databases. Species names are the immediate link between the PII and virtually all other forms of biological knowledge in the "outside world." These same species name databases are also the resource that allows interoperability of other biological databases with each other, emphasizing the critical nature of this seemingly simple "list of names". Strong support was expressed for developing partnerships with existing and new databases that archive species names and authorities, as opposed to establishing an expensive, redundant species list database within the PII.

Connections to other databases would range from relatively simple hypertext links to more sophisticated integration via metadata standards. Integration with collections databases would provide voucher specimen information. Links to sequence databases would provide information about molecular annotation, and links to geographic information systems and other databases would add tremendous value to phylogenetically driven searches. Finally, internet agents could be designed to scour other databases or the World Wide Web as a whole for general information on sets of taxon names generated by phylogenetic queries (Fig. 1).

Higher taxon name database and synonymy. Extensive discussion at the workshop revolved around the problem of synonymy—two or more names associated with the same taxon. Synonymy occurs in names of species and also in names of "higher taxa", or groups of species. Synonymy at the species level is likely to be resolved by external databases currently under development. However, higher taxon names are directly connected to phylogenetic concepts of relationships (e.g., the name "Mammalia" is a statement about the closeness of relationship of all animal species with mammalian features), and there was overwhelming support for including a database of higher taxon names within the PII. This database is critical because many users' initial entry into the database will be via a higher taxon keyword rather than by a species name such as "Homo sapiens". Many formal higher taxon names have vernacular equivalents, such as "mammal" for "Mammalia," which will have to be included if the database is to be broadly accessible. Creating this database will be a major undertaking, and issues involved in defining the content of higher taxonomic groups and their synonym with other names were all debated by participants. Whatever the final strategy, all agreed that it makes practical sense to have a component database for these names directly linked to the PII database.

C. Implementation: Specific Recommendations

The overarching conclusion from the workshop was the need to support a diversity of activities to foster the development of a PII, ranging from basic research to targeted investment in specific physical and human resources.

C.1. Physical Resources

Recommendation. A PII Center should be established. The center should be charged with constructing, maintaining and curating the primary database.

Rationale. Broad support for the establishment of a center for the primary database emerged from the workshop. Participants argued that the scope of the primary database was far beyond the capabilities of current efforts in the community, and yet sufficiently well-defined that it would benefit from being housed in a single center where support staff could provide maintenance and curation. Centralization would avoid duplication of effort and lessen interoperability problems that fall within the purview of the primary archive of data and trees. A single facility would centralize decision-making on technical issues, which would facilitate integration with other databases. Intellectual synergy would be encouraged by co-locating research scientists, visiting scholars, and support staff in one place. Investment in such a center would lead to its establishment as a global resource and encourage international cooperation with ongoing biodiversity and bioinformatics efforts such as GBIF, the Global Biodiversity Information Facility.

Implementation. The PII Center should be associated with an institution that can support it on several levels. Institutions with strong libraries, extensive natural history collections, and expertise in information technology would be prime candidates. Natural history museums or large research universities were viewed as the most likely possibilities. The establishment of this center should receive high priority, given that many other aspects of the PII are contingent upon it. *A strong recommendation was for a call for proposals to establish the PII Center.*

Design of the primary database. The workshop identified several key issues related to the design of the primary database. Some of these are fundamental, such as reaching consensus on the database schema ("ontology"), including the determination of what information should be included in the database. For example, although most molecular sequence data can be described by a well-characterized alphabet of symbols, the same is not true for morphological data. As discussed in the Yale workshop, a morphological database might range from digitized images to high-dimensionality shape data, to biochemical structure descriptions. Design of a database that handles such diverse data types is a challenge that must be met by the PII.

Decisions about the types of trees to be accessioned must also be addressed. A significant fraction of published phylogenies represent gene trees rather than species trees. Gene trees may support or conflict with each other or with their corresponding species tree. Reticulation and lateral gene transfer, as seen in bacteria and other taxa, provides a different kind of conflict in the sense that a single, bifurcating Tree of Life is not an accurate description of relationships. It is essential that decisions on the database design reflect the likelihood that both bifurcating and reticulate trees of species and characters can all play an important role in the TOL.

Slightly more technical issues also must be resolved, including the establishment of metadata standards and the design of a simple application programming interface (API) to foster rapid development of tools that take advantage of the primary database.

It is not clear what strategy NSF should take to ensure that these issues are resolved in a timely fashion. The overriding concern is that there be significant and specific input from the phylogenetics and bioinformatics community on database design to construct an optimal database environment. This effort could itself form the basis of a targeted call for research proposals on database design prior to a competition for establishing a PII Center. Because so much of the PII hinges on deployment of the primary database, these would have to have a quick turnaround time (perhaps one year). Another strategy would be to require the Center proposals to include specific plans for the establishment of collaborative working groups charged with making technical recommendations on database design.

C. 2. Human Resources

Short and long term maintenance. Staffing of the PII Center must be sufficient to provide for the initial construction of the primary database, the initial population of the database with data (ported from existing databases like TreeBASE), initiation of literature-based data acquisition, and curation of incoming data.

In the long term, as with other biological databases, support must be found to guarantee the continued growth and curation of the databases, mainly in terms of support for staff. Support from specific user communities should be sought early in the process. For example, the widely acknowledged utility of phylogenies of viruses in human epidemiology may help convince funding agencies or private foundations that support biomedical research of the need for maintenance of the PII. Support in perpetuity would also be one of the criteria on which the home institution for the PII Center would be based.

Curation. Based on extensive experience with existing small-scale phylogenetic database projects (RDP, TreeBASE, and Tree of Life Project), a clear recommendation was that human curation will be an absolute necessity during data acquisition and later annotation. Many specific problems with data submission can only be resolved by communication between authors and PII Center staff. Another problem requiring human intervention is the enforcement of metadata standards necessary for integration of data between the PII and external databases. Individuals with experience in the aforementioned databases, as well as parallel projects like GenBank and EMBL, should be part of a team of experts to help guide staffing recommendations.

Training. The ultimate success of the phyloinformatics component of building the TOL will require participation by individuals well versed in biology, phylogenetic methods, and computer science. Although there is growing interest in and commitment to multidisciplinary training programs in bioinformatics, many such efforts are focussed on applications in molecular biology and genomics rather than phylogenetics. To insure that experts are trained in skills necessary to handle the data involved in the TOL we suggest the following. **Recommendation. NSF should sponsor a competition for establishing a graduate and postdoctoral training program in phyloinformatics, modeled along the lines of the IGERT program.** This should be inaugurated immediately so that trained professionals will be able to come into the system as early as possible in the process. Another vehicle for training might be the support of postdocs and visiting fellows at the PII Center, who could be involved in development and enhancements of the databases and associated tools.

Outreach. As the TOL comes to fruition, it will be increasingly important to insure that user communities are educated about its existence and utility. Strategies for dissemination of results beyond the research community should be developed. Involvement of natural history museums and outreach to K-12 educators seems especially important. NSF might well consider sponsoring teacher training workshops, K-12 field trips and computer access, and other strategies to ensure that the general public has full access to this "ultimate" tool for bringing biodiversity home to the most people.

Collaboration with scientific journals. Archiving of phylogenetic data in existing databases is now recommended by several journals. Journals and funding agencies must be encouraged to make data submission a **requirement**. This will maximize the quality and quantity of data added to the PII. Any reluctance on the part of journals will presumably be mitigated by the existence of a PII Center and the perception of permanence associated with it.

C.3. Basic Research

Recommendation. Several components of the PII require solutions to substantial technical or scientific problems detailed below. These will require the kind of innovation fostered by competitive funding programs from NSF.

Secondary Databases. Several alternative schemes for the construction of synthetic and comprehensive TOL's were discussed at this and the Yale workshop. One model is "direct construction" of very large trees based on piecing together the underlying data matrices themselves and applying conventional phylogenetic algorithms. Another is "supertree construction", in which large trees are constructed from sets of smaller trees (rather than the original data used to construct those trees) using explicit algorithms that are now under development. Yet a third approach is "human expert synthesis," which begins with trees in the primary database but relies for synthesis on external knowledge supplied by experts in the relevant taxa. The Tree of Life Project has pursued this third strategy using trees from the literature. Other strategies are obviously also possible.

Because much theoretical and empirical work remains to be done in these areas, workshop participants agreed that it was far too early to recommend specific strategies for the design, content or implementation of the secondary database(s) responsible for handling the TOL itself. Moreover, it is unlikely that any one strategy would serve all user communities equally well, and therefore a compelling need for multiple secondary databases could quickly emerge as the PII develops. NSF-sponsored competition would foster innovation in this area and the development of a cogent strategy that guarantees success.

Visualization. Effective display of what will ultimately be an extremely large Tree of Life (and/or a large collection of smaller trees) is an enormous challenge. However, visualization of the results of the growing TOL is also likely to be one of the most exciting products to emerge from this effort. Visualization of extremely large graphs is an area of active research in computer science today (partly motivated by an interest in the complex network of connections in the internet). Fostering such research and its application to large phylogenies should be a high priority.

Among the important elements of a system for visualization of the Tree of Life are (1) interactive mechanisms for browsing very large trees (>1,000,000 taxa); (2) highlighting attributes of taxa and branches, including estimates of statistical support, indications of ambiguity, character states or geographic ranges; (3) extraction of subtrees; (4) comparisons of sets of trees ("forests") from the primary database and their relation to synthesized trees; (5) abstract visualization of arbitrary relationships among trees in the primary database including topological similarity, or relationships between host and parasite trees, to name just a few possibilities.

To make these visualizations stimulating and accessible to both scientists and the general public, exploration of 3D, virtual reality, immersive, and even more exotic media should be encouraged, as well as alternative representations of trees as Venn diagrams, cityscapes, etc. In other words, creativity should be encouraged, perhaps by enlisting the assistance of artists and graphic designers.

Phylogenetic Query Language (PQL). Queries that might be submitted to the PII databases are expected to range from simple ones such as "list all trees that contain species of whales" to more complex ones such as "list all flowering plants closely related to species that synthesize caffeine". Even more complex queries might include seemingly simple requests, such as "find all 'living fossils'", which encapsulates a series of related steps—perhaps a living fossil is defined as groups that have only one species but whose nearest relative diverged at least 100 million years ago. No existing query language is designed to permit searches tailored to the structure of phylogenetic trees. Considerable research is needed in this area, both to determine the needs of the user community and to develop the tools to fulfill these needs.

Phylogenetic Ambiguity. Significant attention should be paid to representing various types of ambiguity in the TOL. Because of the likelihood that conclusions reached by users would be tempered by uncertainty in the TOL itself, there was broad agreement that the strength of hypotheses about relationships and the existence of alternatives should be conveyed to users in creative and informative ways. Navigation of alternative trees and visualization of strengths of support were seen as desirable features of any research-oriented TOL. Although alternative specific solutions were discussed, it was clear that there is a strong need for innovation in this area. Considerable potential for interaction with computer scientists and mathematicians exists in the difficult problem of describing ambiguity in the sense of differences between trees and graphs.

Common names and natural language processing. Most users outside the phylogenetic systematics community will naturally be inclined to use common names for taxonomic groups when working with the Tree of Life. Experience with prototypes like TreeBASE indicates that users often begin with common group names like "mammal", "plant" or "mushroom". Some of these names have precise correspondences to names in the Tree of Life; others do not. Suggested solutions to the problem included software that would point the user back to a portion of the tree, to allow the user to define the precise clade of interest. The translation of these words into non-English languages of the international audience of users also represents a daunting task. To realize the PII mission of widest possible access to the TOL, significant investment should be made in research exploring interfaces that will be truly usable by the nonspecialist and the general public. Handling the ambiguity of common names and reducing the obscurity of conventional queries by use of natural language should be emphasized.

IV. Conclusion: Sense of the Meeting

Participants in the workshop expressed strong support for building an information infrastructure for managing the data associated with a large-scale effort to *reconstruct the Tree of Life (a Phyloinformatics Infrastructure, or PII). Enthusiasm on the part of participating biologists was matched by confidence on the part of computer scientists that such an enterprise was feasible. Both research groups agreed that interesting and challenging problems at the intersection of biology and computer science would take center stage in a phyloinformatics initiative. Indeed, an interdisciplinary approach was viewed as essential to the success of such a project.*

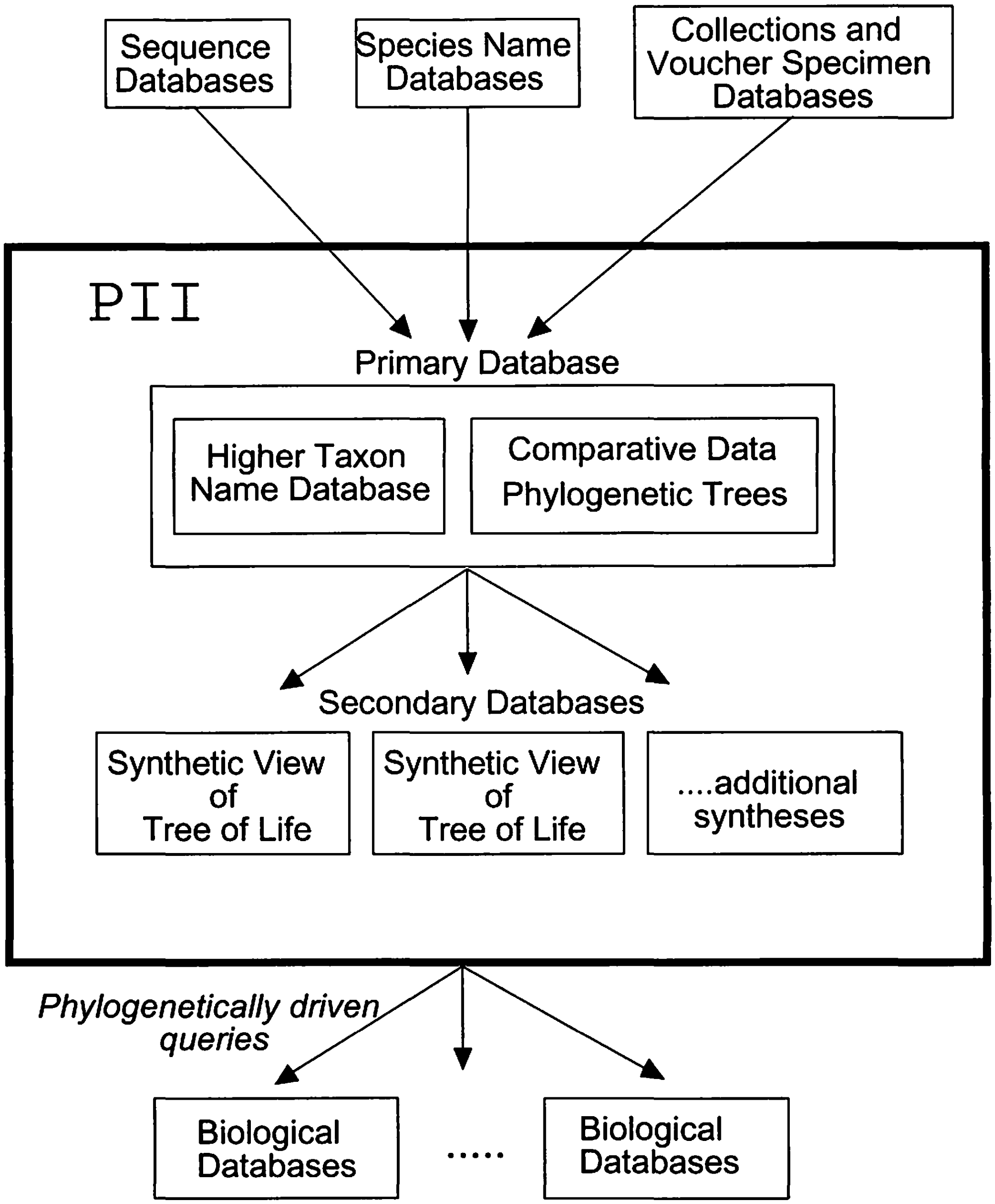


Fig. 1. Simplified diagram of the main elements of a Phyloinformatics Infrastructure (PII) and its relationship to databases outside of it.

Sabre virtually there

NAVIGANT
1-512-459-8203
WE APPRECIATE YOUR BUSINESS!

HOME | MY VT | DESTINATIONS | TRAVEL TOOLS

View: FLIGHTS Language: English Detailed View

Trip Plans for: **JONATHAN EISEN** **Sabre Reservation Code:** **UCOAOS**

Travel Arranger Comments:

Important Notes: **

** ELECTRONIC TICKETLESS RESERVATIONS RECEIPT **
** PRESENT 1 GOVT ISSUED PHOTO ID/DRIVERS LIC **
** WITH YOUR CONFIRMATION NUMBER AT CHECK-IN **
** YOUR TICKETLESS CONFIRMATION NUMBER IS **
** *WSGRLO
** IF TRIP IS CANCELLED OR TICKET IS NOT USED **
** PLEASE ADVISE THE TRAVEL OFFICE TO PROCESS **
** REFUNDS, EXCHANGES,OR VOID TICKETS **
** TOTAL AMOUNT INVOICED USD 335.60
****

FLIGHTS

Sat, Dec 2 **UNITED AIRLINES, UA 1975** Check Gate Info

From: WASHINGTON DULLES, VA (IAD)
To: AUSTIN, TX (AUS)
Class: Economy
Status: Confirmed
Meal: Dinner
Aircraft: Boeing 737-300
Flight Time: 3 hours and 39 minutes

Departs: 5:45p
Arrives: 8:24p
Seat: 03C
Confirmation: WSGRLO
Smoking: No
Mileage: 1303

Tue, Dec 5 **UNITED AIRLINES, UA 1422** Check Gate Info

From: AUSTIN, TX (AUS)
To: WASHINGTON DULLES, VA (IAD)
Class: Economy
Status: Confirmed
Meal: Lunch
Aircraft: Boeing 737-500 Jet
Flight Time: 3 hours and 05 minutes

Departs: 12:40p
Arrives: 4:45p
Seat: 04C
Confirmation: WSGRLO
Smoking: No
Mileage: 1303

Back To Top

CAR RENTAL

Back To Top

HOTEL & LODGING

OM OMNI AUSTIN HOTEL

Address: 700 SAN JACINTO
AUSTIN TEXAS 78701
Phone: 512-476-3700
Room type: NONSMOKING 1 KING
Rate: 85.00N / night
Status: Confirmed
Confirmation: 256228
Notes: HOTEL FAX 512-397-4888
CANCEL BY 4PM 30NOV

Check in: Dec 2
Check out: Dec 5

Back To Top

OTHER

Sat, Feb 3

City: WASHINGTON DULLES, VA (IAD)
Status: Confirmed
Information: THANK YOU FOR CHOOSING NAVIGANT

Back To Top

Tree of Life Workshop

TENTATIVE ITINERARY, DEC. 3-DEC.4

SATURDAY, DECEMBER 2, 2000

Recommended arrival into Austin, Texas.
Hotel check-in at the Omni in Downtown Austin begins at 3:00 p.m.

SUNDAY, DECEMBER 3, 2000

7:55a.m.	Convene in hotel's front lobby to await van pick up to the UT Austin campus (please be prompt)
8:00a.m.	Vans load & depart. Each van seats 15 passengers and will be driven by volunteer UT students and/or lecturers.
8:15a.m.	Arrive at the workshop facility (ACES building, Electronics Seminar Room 2.402 on the ground floor of the ACES building). Check-in at the registration table to receive your name badge and to sign travel & reimbursement papers. Breakfast of pastries, bagels, fresh fruit/coffee, juice.
8:30a.m.	Background on the Tree of Life Initiative & introductions
8:45a.m.	Reports from the previous workshops at Yale University and U.C. Davis.
9:15a.m.	General discussion of structure of TOL Centers, TOL Networks
10:15a.m.	Coffee break
10:30a.m.	General discussion of new technologies needed; development of human resources
11:30a.m.	Formation of break-out groups, discussion of objectives
12:00p.m.	Boxed lunches provided; discussions continue over lunch
1:00p.m.	Break-out groups convene in Rooms 2.404B and 2.444
3:30p.m.	Coffee Break and discussion in your small groups
3:45p.m.	Reconvene in large group in Room 2.402. Reports from initial break-out groups; target priorities for Day 2
6:15p.m.	Cocktail reception at the Texas Memorial Museum at The University of Texas at Austin
7:00p.m.	Dinner at the Texas Memorial Museum (menu: mixed green salad, sauteed chicken saltimbocca with mixed vegetables and roasted new potatoes, chocolate gauche cake)
9:00p.m.	Vans pick up & return guests to the Omni.

MONDAY, DECEMBER 4, 2000

7:55a.m.	Convene in hotel's front lobby to await van pick up to the UT Austin campus (please be prompt)
8:00a.m.	Vans load & depart for UT campus
8:15a.m.	Breakfast buffet is served in Room 2.402 (breakfast tacos, fresh fruit/coffee, juice)
8:30a.m.	Large group meeting—set goals & priorities for today's break-out groups
8:45a.m.	Form break-out groups (may be different from Day 1). Groups report to rooms 2.404B or 2.444 at this time
10:15a.m.	Coffee Break
10:30a.m.	Continue break-out group session until noon
12:00p.m.	Leave break-out group and convene in Room 2.402 for boxed lunches in large group setting
1:00p.m.	Convene in Room 6.304 (6th floor) for reports from break-out groups
3:15p.m.	Coffee Break and discussion in Rm 6.304
3:30p.m.	Back to break-out group rooms 2.404B and 2.444 to write group reports
5:45p.m.	Workshop formally ends. Vans pick up guests and return them to the Omni to prepare/get refreshed for tonight's dinner cruise. The riverboat is heated and enclosed.
6:30p.m.	Vans reload to take guest to the riverboat boarding dock—Please be Prompt!
7:15p.m.	Boat boarding begins
7:30p.m.	The Flagship Texas Riverboat departs. Dinner and cocktail reception aboard the boat

Tree of Life Workshop Itinerary (continued)

(CONTINUED FROM MONDAY, DECEMBER 4, 2000)

9:30p.m.	The Flagship Texas riverboat returns to shore. Vans take guests back to hotel.
10:00p.m.	Expected return to the hotel.

TUESDAY, DECEMBER 5, 2000

Recommended date of departure from Austin
Hotel check-out is scheduled at or before 12 noon.

-Fin-

TREE OF LIFE WORKSHOP

December 3-December 4, 2000

Workshop Site:	The University of Texas at Austin
Organized By:	David M. Hillis, The University of Texas at Austin Wayne Maddison, The University of Arizona
Sponsored By:	The National Science Foundation

Attendees

Participants were chosen to represent a wide diversity of taxonomic groups, research expertise in data collection and analysis, research approaches, technology development, and experience in the organization and administration of national centers. See attached page for list of attendees.

Hotel Accommodations

The Omni in Downtown Austin is the official hotel of the Tree of Life Workshop. The Omni is located at 700 San Jacinto at 8th Street, Austin, Texas 78701. The hotel can be reached by phone at (512) 476-3700 or (800) THE-OMNI or by facsimile at (512) 320-5882. Relatives, friends, or colleagues can call the toll-free Omni number to reach your hotel room (please instruct them to ask for the Downtown Austin location).

Workshop Location

All of the workshop sessions will be held in the Applied Computational Engineering Sciences Building (A.C.E.S.) at The University of Texas at Austin. The street address is 201 East 24th Street, Austin, Texas 78712. Shuttles will take workshop guests to and from the workshop each day.

Computer Access

The School of Biological Sciences will have laptop computers (Macintosh & PC) available for use during both workshop days. These laptop computers are for shared use. You are encouraged to bring your own computer if you would like accessibility during non-workshop hours. The Omni Hotel has modems available in each guest room and high-speed, wireless access in the hotel lobby. Each seminar room in the ACES building at UT Austin is equipped with at least four high-speed ethernet ports. They are configured for DHCP. Please set your laptop network settings accordingly.

Meals

All meals will be provided for you during the workshop dates of Sunday, December 3 and Monday, December 4. If you have special dietary needs or would like to request a vegetarian/vegan meal, please contact Amy Kim at (512) 232-7073 or via email at amykim@mail.utexas.edu.

At the close of the workshop on Monday evening, a special dinner cruise will be held for Tree of Life Workshop guests as well as some University of Texas faculty and students. Attendees will board the Flagship Texas, a luxury charter yacht, for a two hour excursion of Lake Travis which is located west of Austin on the scenic Colorado River. A catered dinner and bar service will be offered for your enjoyment. In the case of inclement weather, the Flagship Texas riverboat has two heated, fully covered state rooms. The dinner cruise will be two hours in duration from 7:30p.m. to 9:30p.m.

Ground Transportation

To & From the Airport: Taxi's and shuttle busses are widely available at the Austin-Bergstrom International Airport. A taxi from the airport to downtown will cost approximately $25.00. "Supershuttle of Austin" is a widely used shuttle bus service. A one-way rate from the airport to downtown is $9.00, with a round-trip rate of $16.00. Supershuttle reservations can be made by calling (512) 929-3900 or online at http://supershuttle.com. Please retain transportation receipts for reimbursement.

Tree of Life Workshop

PARTICIPANT LIST

ORGANIZERS

David M. Hillis, University of Texas, Austin. Interests: Phylogenetic methods for large data sets; viral evolution; phylogenetic applications; metazoan systematics

Wayne Maddison, University of Arizona. Interests: Computational phylogenetics; phylogenetic software; spider systematics

PARTICIPANTS

Judy Blake, The Jackson Laboratory. Interests: Genome informatics

Jeffrey Boore, DOE Joint Genome Institute. Interests: Comparative Genomics.

Stuart Brand, Co-chair, All Species Initiative. Interests: Applications of information technology to large biological problems

David C. Cannatella, University of Texas, Austin. Interests: Comparative analysis, vertebrate systematics and morphology

Joel Cracraft (co-organizer of Yale Workshop), American Museum of Natural History. Interests: avian systematics, large data sets

Michael Donoghue (co-organizer of Yale workshop), Yale University. Interests: plant systematics, large data sets

Jonathan Eisen, The Institute for Genomic Research. Interests: Bacterial phylogeny, genomics

John Huelsenbeck, University of Rochester. Interests: Statistical analysis of phylogenies; phylogenetic methods

Robert Jansen, University of Texas, Austin. Interests: Plant systematics; using rearrangements of genomes for phylogenetic analysis

Kevin Kelly, Editor, Wired Magazine; Co-chair, All Species Initiative. Interests: Applications of information technology to large biological problems

Junhyong Kim, Yale University. Interests: Phyloinformatics, biostatistics, and developmental evolution

Leonard Krishtalka, University of Kansas. Interests: Systematics, museum databases

Francois Lutzoni, Field Museum. Interests: Systematics of lichens, co-evolution

David M. Maddison, University of Arizona. Interests: Phylogenetic software and theory, insect systematics

Maureen O'Leary, State University of New York, Stony Brook. Interests: Paleontological approaches to phylogenetics

Jim Reichman, University of Santa Barbara. Director, National Center for Ecological Analysis and Synthesis

Timothy B. Rowe, The University of Texas at Austin. Interests: The evolution and anatomical structure of reptiles, birds, and mammals; multimedia and X-ray scanning technology

Chris Simon, University of Connecticut. Interests: Insect systematics

David Swofford, Smithsonian Institution, Laboratory of Molecular Systematics. Interests: Phylogenetic analysis (including software development)

Tandy Warnow, University of Texas, Austin. Interests: Computational phylogenetics, computation and visualization of large phylogenetic trees

Greg Wray, Duke University. Interests: Developmental evolution, echinoderm evolution

Anne Yoder, Northwestern University Medical School. Interests: Primate systematics, molecular evolution

Elizabeth Zimmer, Smithsonian Institution, Laboratory of Molecular Systematics. Interests: Plant molecular systematics

NATIONAL SCIENCE FOUNDATION PARTICIPANTS

Matthew Kane
James E. Rodman
Joann Roskoski
Grace A. Wyngaard
Terry Yates

FACTORIES

What?

- high throughput
- centralized

outside Box

Why?

- cheaper
- faster
- better?
- access to more people
- ability to change technology rapidly
- storage and archiving

WHAT ARE THE STEPS WHERE NEEDED?

- rate limiting
- technology expensive or big

STEPS IN PROCESSES

OLECULAR (1)organism → ② tissues → ③ DNA → ④ CLONES → SEQUENCE OR OTHER

ORPHOLOGICAL (1)organism → (2) store/transport → (3) imaging/measure

OSSIL ~~(1) organis~~
(1) fossil → (2) store/transpat → (3)

RFPs for factories/high throughput

- could be NSF facility
- could be associated with ceter, museum, university

(1) Morphometric - high throughput digital museum
- Also could be useful to have remote/portable technologies

(2) Tissue/cells to DNA/libraries

(3) Sequencing/molecular characters

(4) Fossil isolation

RFPs for new technology

① archiving 1° + 2° materials

② new technologies to speed up/cheapen/make accessible gathering of phylogenetically informative characters in any stage of process, in particular rate limiting or expensive steps.

③ archiving information/databases.

④ quality control

⑤ species identifaction A$$istance

Other issues related to factories/high throughput

① interaction across steps. Storage of samples needs to be linked to methods of acquiring characters.

② some could be local, some in TOL centers, some in centralized facilities. Not all need to be done in centers.

③ could be private but can't count on it

④ credit for data/research needs to be addressed.

⑤ coordination with other projects very important. For example whole genomes will be useful. TOL goals might help select genomes.

Computing the Tree of Life

PROJECT LEADERS
David Hillis
University of Texas, Austin
Paul Lewis
University of Connecticut

A fundamental objective of biology is to produce an accurate tree of life for the world's 1.7 million known species. Before this tree can be completed, biologists face three computational challenges. They must develop the tools to analyze the large data sets and determine relationships for all living organisms, train people to collect the data and use the computational tools once they are developed, and engineer an infrastructure that can support the data collection and computational systems. Along with a team of scientists at the University of Texas, Austin, David Hillis is working to overcome these obstacles with the help of NPACI supercomputers.

Through the study of evolutionary relationships—a field known as *phylogenetics*—researchers can now use computers to reconstruct the history of a virus such as influenza or HIV, determine its origin, and use this information to predict the course of the epidemic or to develop potential vaccines. With the help of supercomputers, scientists like Hillis will someday take phylogenetics a step further and complete the tree of life.

PARALLEL ALGORITHMS FOR GENETICS

Hillis, director of Texas' School of Biological Sciences, is working on approaches to reconstruct the history of thousands of species and, with this history, to interpret biological information on the behavior, physiology, or molecular evolution of the species. Hillis' primary limitation is the inability to handle the large, complex data sets that are becoming more and more common within phylogenetic analysis.

"In practice, a modern phylogenetic analysis consists of finding the pattern of relationships that best accounts for the DNA or RNA sequence data, given a particular model of evolution," Hillis said. "Evaluating a potential solution to a phylogenetic problem is computationally very intensive, and the number of possible solutions to a phylogenetic problem increases precipitously as more organisms are added. This puts the phylogeny problem into the class of computational problems that are NP-complete; the solution time goes up exponentially as a function of the number of sequences in the analysis."

The rapid increase of the size of phylogenetic data sets is largely due to discoveries in molecular biology and genetics that are producing gene sequences of more and more organisms. These large sets of sequences create phylogenetic problems that are very difficult to solve—even with the world's most powerful supercomputers.

Because of the difficulty of the problem, there is no provably optimal algorithm for solving it; instead, there are several heuristic approaches. Hillis' project is looking at the performance of an approach based on genetic algorithms, which can be easily scaled to take advantage of large pools of computer resources. They are intrinsically parallel algorithms, and the project has used the Cray T3E at the Texas Advanced Computation Center to refine and test the approach.

COMMUNITY-WIDE TOOLS

Through collaboration with Paul Lewis, professor of ecology and evolutionary biology at the University of Connecticut, Hillis' group is testing a genetic-algorithm-based software package for parallel supercomputers that helps scientists depict the relationships among large sets of organisms. Lewis' program, GAML, has recently been applied to estimate the evolutionary relationships among the more than 100,000 species of flowering plants. Due to its large size, the flowering plant problem is such that conventional approaches cannot be used to address it effectively.

"GAML has been shown to have great potential to work across large, distributed networks of workstations, as well as on supercomputers," Hillis said. "Future large-scale phylogenetic analyses might therefore take advantage of unused computer resources or, as has been done most successfully by the Search for Extraterrestrial Intelligence, of the computer resources of interested amateurs."

Beyond his own phylogenetics studies, Hillis is working to make such computational tools widely available for determining and manipulating large sets of complex data, such as the entire database of tens of thousands of HIV sequences. Once Hillis and Lewis perfect GAML, they will share it with the biological community. And as a partner in NPACI's Bioinformatics Infrastructure alpha project, Hillis is contributing phylogenetics algorithms to the suite of large-scale analyses that the infrastructure will make possible.

"With GAML, we are hopeful that we are one step closer to reconstructing the complete tree of life," Hillis said. "Reconstructing the tree of life is a major objective of evolutionary biology and will require a serious, coordinated international effort to complete. However, once we complete this feat, scientists will have the framework in place to incorporate newly discovered species into the tree. We will then have a more lucid understanding of the patterns of biodiversity and an extremely powerful tool for interpreting all aspects of biological variation." —*KMB* ▼

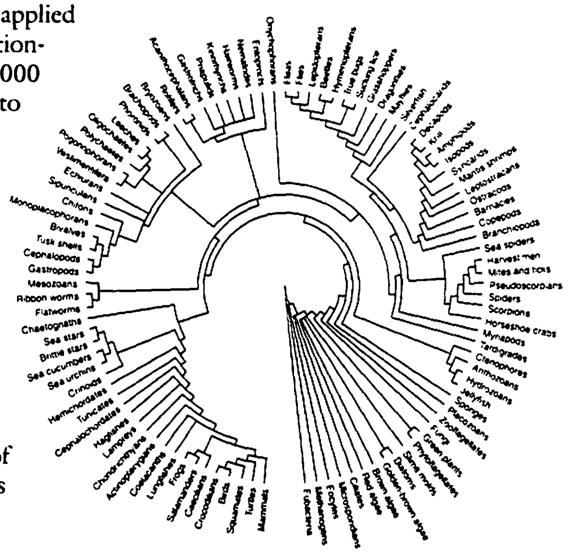
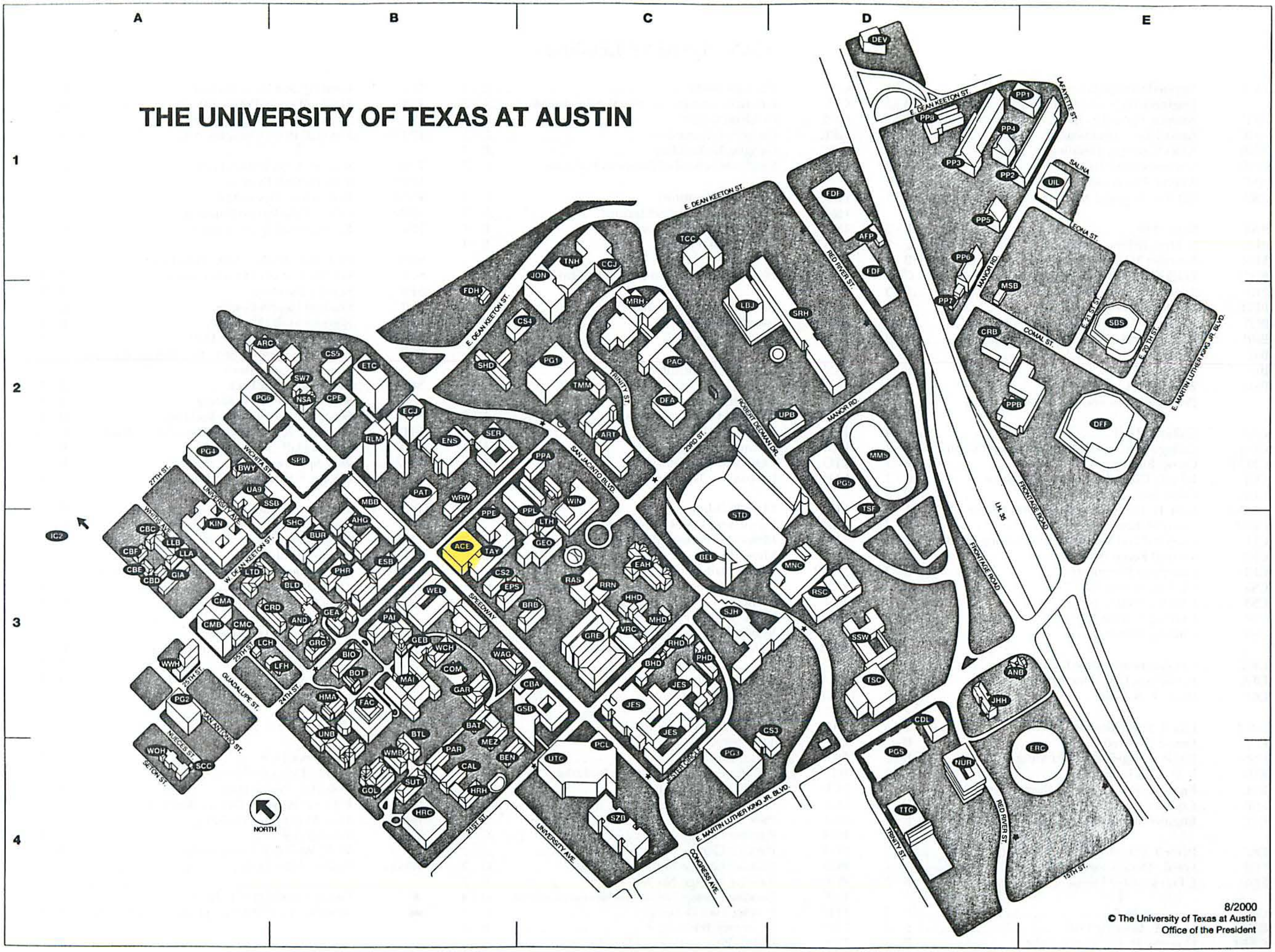


Figure 1. A Branch from the Tree of Life

A phylogenetic estimate from a small sample of the tree of life created by David Hillis. With current computational tools, phylogenetic analyses for 1,000 species is possible with adequate computer resources. However, it is currently impossible to reach a reasonable solution for 500,000 species, even with months of computation. Hillis, Paul Lewis, and their collaborators are studying how to make analysis of such large data sets possible.

THE UNIVERSITY OF TEXAS AT AUSTIN



8/2000

Office of the President

Tree of Life Workshop

ITINERARY, DEC. 3-DEC.4

SUNDAY, DECEMBER 3, 2000

7:55a.m.	Out-of-town guests convene in hotel's front lobby to await van pick up to the UT Austin campus. Vans will depart at 8:00am.
8:15a.m.	Arrive at the workshop facility (ACES building, Electronics Seminar Room 2.402 on the ground floor). Check-in at the registration table to receive your name badge and to sign travel & reimbursement papers. Breakfast of pastries, bagels, fresh fruit/coffee, juice.
8:30a.m.	Group meeting begins in 2.402 (introduction to the TOL initiative, background reports from previous workshops)
10:15a.m.	Coffee break
10:30a.m.	Group meeting continues (discussion of new technologies; development of human resources, formation of break-out groups, discussion of objectives)
12:00p.m.	Boxed lunches provided
12:30p.m.	Group tour of the CT Lab in the Geological Sciences Building (host: Tim Rowe)
1:00p.m.	Small groups convene in Rooms 2.404B , 2.444, or 3.116 (if 3rd group necessary)
3:30p.m.	Coffee Break and discussion in your small groups
3:45p.m.	Reconvene in large group in Room 2.402. Reports from initial break-out groups; target priorities for Day 2
6:15p.m.	Cocktail reception at the Texas Memorial Museum at The University of Texas at Austin
7:00p.m.	Dinner at the Texas Memorial Museum
9:00p.m.	Vans pick up & return guests to the Omni.

MONDAY, DECEMBER 4, 2000

7:55a.m.	Out-of-town guests convene in hotel's front lobby to await van pick up to the UT Austin campus. Vans depart promptly at 8:00a.m.
8:15a.m.	Breakfast buffet is served in Room 2.402 (breakfast tacos, fresh fruit/coffee, juice)
8:30a.m.	Large group meeting in Rm. 2.402 convenes briefly to set goals & priorities for today's break-out groups
8:45a.m.	Form break-out groups (may be different from Day 1). Groups report to rooms 2.404B , 2.444, or 3.116.
10:15a.m.	Coffee Break in small group rooms
10:30a.m.	Continue small group sessions until noon
12:00p.m.	Lunch is served in Room 2.402 (large group)
12:30p.m.	Tour of the Visualization Lab in the ACES Building (host: Tim Rowe)
1:00p.m.	Convene in Room 6.304 (6th floor) for reports from small groups
3:15p.m.	Coffee Break and discussion in Rm 6.304
3:30p.m.	Back to small group rooms 2.404B , 2.444, or 3.116 to write group reports
5:45p.m.	Workshop formally ends. Vans pick up guests and return them to the Omni to prepare/get refreshed for tonight's dinner cruise. The riverboat is heated and enclosed.
6:30p.m.	Vans reload to take guest to the riverboat boarding dock—Please be Prompt!
7:00p.m.	Boat boarding begins
7:30p.m.	The Flagship Texas Riverboat departs. Dinner and cocktail reception.
9:30p.m.	The Flagship Texas riverboat returns to shore. Vans take guests back to hotel.

~fin~

TREE OF LIFE WORKSHOP

Helpful Numbers/Addresses

TRANSPORTATION

American Yellow Checker Cab Co. / Tel: (512) 452-9999
Austin Cab / Tel: (512) 478-2222
Supershuttle of Austin / Tel: (512) 929-3900

HOTEL

The Omni Hotel
700 San Jacinto at 8th Street
Tel: (512) 476-3700 or (800) THE-OMNI
Facsimile (512) 320-5882

UT TRAVEL AGENCY

Navigant International (Agent: Pamela Folds)
TEL: (800) 798-0292

WORKSHOP RELATED VENUES

- A.C.E.S. Building
  201 East 24th Street (between 24th & Speedway)
  The University of Texas, Austin
  Austin, Texas 78712
  Telephone (512) 232-9090
  Facsimile (512) 232-9084

- Texas Memorial Museum
  2400 Trinity Street
  Austin, Texas 78705
  Tel: (512) 471-1604

- The Flagship Texas Riverboat
  10 Stoney Creek Cove
  Austin, TX 78734
  Office Tel: (512) 261-6484
  Ship's Cellular Phone: (512) 461-4774

LOCAL WORKSHOP HOST

David M. Hillis
Work Tel: (512) 232-3690
Home Tel: (512) 472-8614
Mobile Tel: (512) 970-5762

LOCAL WORKSHOP COORDINATOR

Amy Kim
Work Tel: (512) 232-7073
Home Tel: (512) 891-6381
Mobile Tel: (512) 217-3218

GENERAL AUSTIN INFORMATION

Austin Bergstrom International Airport / Tel: (512) 530-ABIA
Austin Visitor's & Conventions Bureau / Tel: (800) 299-1700 or (512) 463-8586
Austin local time & temperature readings / Tel: (512) 973-3555

ACES BUILDING - GROUND FLOOR