Databases and ontologies

TimeTree: a public knowledge-base of divergence times among organisms

S. Blair Hedges[1,*], Joel Dudley[2] and Sudhir Kumar[2]

[1]Department of Biology, Pennsylvania State University, University Park, PA 16802-5301, USA and
[2]Center for Evolutionary Functional Genomics, The Biodesign Institute and the School of Life Sciences, Arizona State University, Tempe, AZ 85287-5301, USA

ABSTRACT

Summary: Biologists and other scientists routinely need to know times of divergence between species and to construct phylogenies calibrated to time (timetrees). Published studies reporting time estimates from molecular data have been increasing rapidly, but the data have been largely inaccessible to the greater community of scientists because of their complexity. TimeTree brings these data together in a consistent format and uses a hierarchical structure, corresponding to the tree of life, to maximize their utility. Results are presented and summarized, allowing users to quickly determine the range and robustness of time estimates and the degree of consensus from the published literature.

Availability: TimeTree is available at http://www.timetree.net

Contact: sbh1@psu.edu

The evolutionary history of life includes two primary components: phylogeny and timescale. Phylogeny refers to the branching order (relationships) of species or other taxa within a group, crucial for understanding the inheritance of traits and for erecting classifications. However, a timescale is equally important because it provides a way to compare the phylogeny directly with the evolution of other organisms and with planetary history, such as geology, climate, impacts and other features (Fig. 1). Temporal information can come from either the fossil record (Benton, 1993) or molecular clocks, which are often calibrated using the fossil record (reviewed in Hedges and Kumar, 2003; Kumar, 2005; Wray, 2001).

Publications involving molecular clocks, now numbering in the hundreds, are appearing more frequently (Kumar, 2005). In large part, this has been driven by the rapid increase in sequence data that have become available in recent years, as well as the development of methods that accommodate rate variation in estimating divergence times and their confidence intervals (Drummond *et al.*, 2006; Kumar *et al.*, 2005; Sanderson, 1997; Thorne *et al.*, 1998). Eventually, this effort will lead to a phylogenetic tree of life that is calibrated to time (i.e. a timetree of life). This is especially important for the many groups that have little or no fossil record. Important biological questions have been addressed with these methods, including the origin of life, rise in oxygen on Earth, origin of eukaryotes, and colonization of land by plants and fungi, and radiation of animal phyla (reviewed in Hedges, 2002). While each study addresses a common theme or question, the global result is that there are now

thousands of individual published time estimates among organisms and species.

We developed the TimeTree database so that these data would be readily accessible to the scientific community. Scientists often need to know times of divergence to calculate rates of change, to search for connections with Earth history and to construct timetrees, among other reasons. They also need to know the range of time estimates and whether or not there is a consensus among studies. Most time estimates are not included in abstracts that are searchable on the web, but instead are usually buried in the text in a variety of formats (including trees) that are often difficult to assimilate and compare among studies.

The home page (Fig. 2) briefly describes the database and presents a choice of two user-query forms: finding the divergence of two taxa, and finding all time estimates from a particular author. For the taxon-query, the user inputs two names (not order specific) such as 'cat' (suborder Feliformia) and 'dog' (Caniformia). TimeTree first uses a synonym list (from NCBI) to translate from many common names, or phonetically encodes the query string. Then, the two most-inclusive groups containing those taxa are identified, and all taxa in one group are compared with all taxa in the other group to find all published times of divergence for the evolutionary split. The guide tree used is based on NCBI's Taxonomy Browser, which currently contains approximately 225 000 taxa. The advantage of this hierarchical approach is that any query between any two species in the two most-inclusive groups defining a node will yield results if there are any data bearing on the node. For example, all current time estimates of the Feliformia-Caniformia divergence are derived from cat-dog sequence comparisons, but a query of mongoose (Feliformia) and raccoon (Caniformia) will yield exactly the same results because their divergence pertains to the same node. The guide tree will be updated as NCBI's Taxonomy Browser is revised and as new developments occur in the field.

The results of a taxon-query (Fig. 3) are presented in tabular form and list all published studies and time estimates bearing on that species divergence. These are grouped by study, and the user is given the option to show or hide the details of each entry. Those details include divergence time (millions of years ago), number of genes, gene type (e.g. nuclear and organellar), data type (e.g. nucleotide and amino acid), specific taxa compared (which could be different from queried taxa, as noted above), and the source of the information within the article (e.g. the specific table or figure) (Fig. 3a). Confidence limits and standard errors are shown when
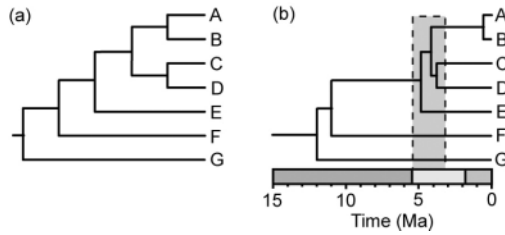
---

*To whom correspondence should be addressed.

Fig. 1. Relationship between a phylogenetic tree (**a**) showing only the branching order of seven species, and the corresponding timetree (**b**) showing both branching order and times of divergence. Colored horizontal panels represent geologic periods, and the gray vertical bar represents a climatic or geologic event.

Fig. 2. TimeTree home page showing taxa-based (e.g. 'cat' versus 'dog') and author-based query options. The hierarchical nature of the database is illustrated at right.

Fig. 3. Results page showing individual studies reporting time estimates for a query for 'cat' versus 'dog' (**a**) and the summary results (**b**). Details for some studies were hidden from view by checking the box at left, but all were included in the summary.

the mouse pointer hovers over the divergence times. Taxon names are linked to taxonomy entries (including sequence information) in NCBI's Taxonomy Browser, and author's names are linked to Pubmed records in NCBI showing abstracts. Time estimates are summarized on the Results page, with the means and weighted means shown for all data and partitions (Fig. 3b).

This first version (v1.0) of TimeTree is limited in taxonomic scope to tetrapods (amphibians, reptiles, mammals and birds). It includes 70 studies and 1102-divergence times. Additional tetrapod data are being added by our group using the publications submitted by the community to us and those we have identified by searching the published literature. Future versions will be expanded in taxonomic scope to include all of life. They will also include enhanced tree-based graphics (e.g. timetrees and timelines) and fossil times, and will permit user-submission of data and community-based quality control. The latter will be facilitated by a TimeTree Consortium of experts, already in place, who will provide advice and review new submissions of data. User assistance is included through a list of Frequently Asked Questions.

With its simplicity of use and ability to handle common names of organisms, educators will find timetree useful as a classroom tool. In summary, TimeTree is a web-based resource for researchers and educators from all fields of science to assist in the discovery of patterns and the formulation of hypotheses, or simply to learn about the timetree of life.

REFERENCES

Benton,M.J. (1993) *The Fossil Record 2*. Chapman and Hall, London.

Drummond,A.J. *et al.* (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol.*, **4**, 699–710.

Hedges,S.B. (2002) The origin and evolution of model organisms. *Nat. Rev. Genet.*, **3**, 838–849.

Hedges,S.B. and Kumar,S. (2003) Genomic clocks and evolutionary timescales. *Trends Genet.*, **19**, 200–206.

Kumar,S. (2005) Molecular clocks: four decades of evolution. *Nat. Rev. Genet.*, **6**, 654–662.

Kumar,S. *et al.* (2005) Placing confidence limits on the molecular age of the human-chimpanzee divergence. *Proc. Natl Acad. Sci. USA*, **102**, 18842–18847.

Sanderson,M.J. (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.*, **14**, 1218–1231.

Thorne,J.L. *et al.* (1998) Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.*, **15**, 1647–1657.

Wray,G.A. (2001) Dating branches on the tree of life using DNA. *Genome Biol.*, **3**, 1–7.