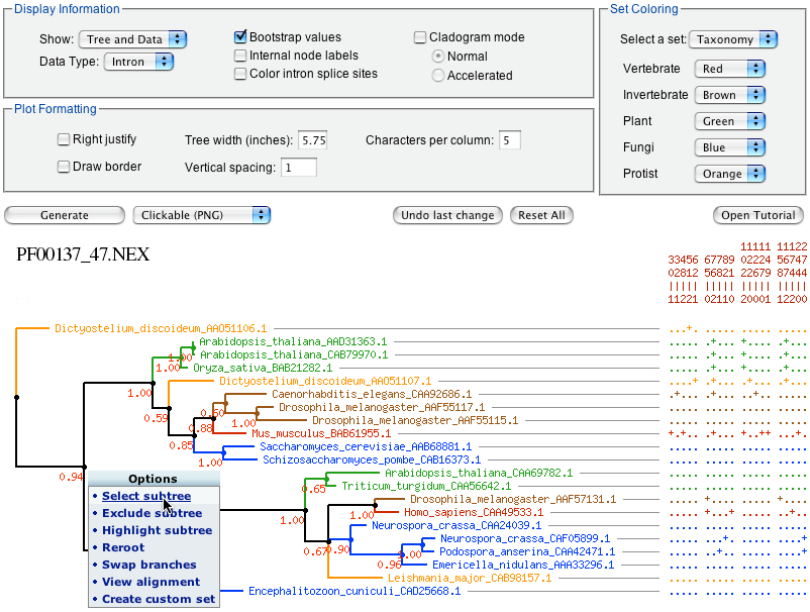
Nexplorer: Phylogenetic Exploration of Sequence Family Data

Genome sequence data are an enormous resource. The practical ability to tap this resource does not depend on direct analysis of individual sequences, but rather on comparative methods that use data from different species. Genome annotation, drug target discovery, biomolecule engineering, and medical genetics rely on sequence comparisons. However, most such efforts are still a long way from the ideal approach, i.e., a probabilistic model of evolutionary change using a phylogenetic tree. The success of more robust models will depend on software tools (bioinformatics) that facilitate the storage, exchange, processing, and visualization of genomic data together with phylogenetic trees.

A Stoltzfus (Div. 831)
V. Gopalan, W.-G. Qiu, and M. Chen (UMBI/CARB)

To facilitate the phylogeny-based analysis of comparative data, we have developed *Nexplorer*, a web-based program that combines a useful set of features for viewing and manipulating data with the capacity to generate publication-quality graphics of a tree with a sequence alignment or other data. The foundation for the functionality of *Nexplorer* is our earlier work implementing NEXPL, an applications programming interface for the NEXUS file format used by evolutionary biologists. The use of NEXPL and NEXUS provide a foundation for building software tools that utilize trees and data together. *Nexplorer* solves the technical problem of providing an unimposing interface to this functionality.

This year the *Nexplorer* server came online at http://www.molevol.org/nexplorer, providing images that allow the user to view and modify data sets, and to save data sets and images for later use.



A view of the Nexplorer server interface showing a sample set of data (intron data for the ATP Synthase Subunit C family), with control panel (above), data matrix (right), and gene phylogeny (left). The pop-up menu shows node-specific operations. Taxonomic coloring (key, upper right) reveals important patterns such as sub-families.

To make this server more useful, we also provided thousands of pre-computed data sets corresponding to families of protein-coding genes.

We anticipate that *Nexplorer* will be used by the medical and drug discovery communities to discover patterns in genomic data, and more generally, that it will stimulate an appreciation for the use of comparative data for many other applications.

Our plans for *Nexplorer* include updates to data sets, as well as expansions in the functionality of the server to allow greater flexibility in sub-selecting and modifying data sets.

Publication
Gopalan, V., Qiu, W. G., Chen, M. Z., and Stoltzfus, A. ***Nexplorer*: *Phylogeny-based exploration of sequence family data.*** *Bioinformatic*s, 2006 Jan 1;22(1):120-1. Epub 2005 Nov 2.