# Chapter 10
# Phyloinformatics: Toward a Phylogenetic Database

**Roderic D. M. Page**

**Summary**

Much of the interest in the "tree of life" is motivated by the notion that we can make much more meaningful use of biological information if we query the information in a phylogenetic framework. Assembling the tree of life raises numerous computational and data management issues. Biologists are generating large numbers of evolutionary trees (phylogenies). In contrast to sequence data, very few phylogenies (and the data from which they were derived) are stored in publicly accessible databases. Part of the reason is the need to develop new methods for storing, querying, and visualizing trees. This chapter explores some of these issues; it discusses some prototypes with a view to determining how far phylogenetics is toward its goal of a phylogenetic database.

## 10.1 Introduction

Cracraft [87] defined *phyloinformatics* as "an information system that is queried using the hierarchical relationships of life." Much of the interest in the "tree of life" [391] is motivated by the notion that we can make much more meaningful use of biological information if we query the information in a phylogenetic framework. Rather than being limited to queries on single species or arbitrarily defined sets of species, phyloinformatics aims to query data using sets of evolutionarily related taxa (Figure 10.1).

Implementing such a system raises a number of issues, several of which have been discussed at various workshops.[1] My aim in this chapter is

[1]Examples include the tree of life workshops held at Yale and the Universities of California at Davis and Texas at Austin (reports available from http://taxonomy.zoology.gla.ac.uk/rod/docs/tol/) and the tree of life workshop at DIMACS (http://dimacs.rutgers.edu/Workshops/Tree/).

to explore some of the database and data visualization issues posed by phylogenetic databases. In particular I discuss taxonomic names, supertrees, and navigating phylogenies. I review some recent work in this area and discuss some prototypes with a view to determining how far phylogenetics is toward its goal of a phylogenetic database.

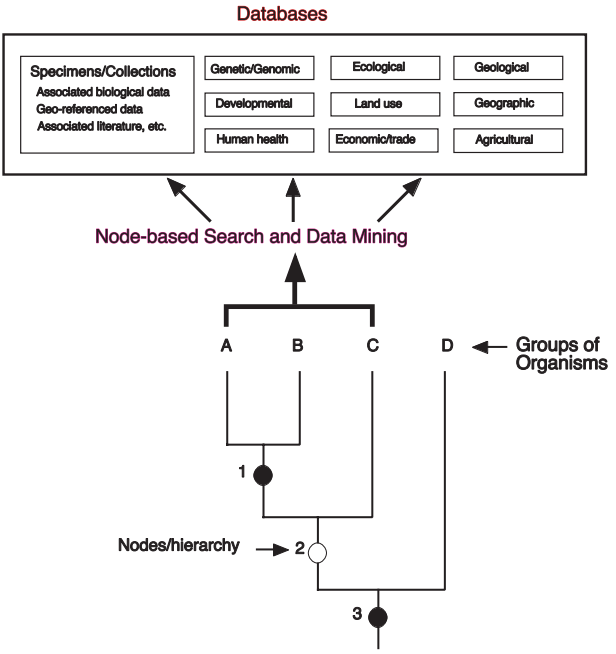


**Fig. 10.1.** Diagram illustrating a "phyloinformatic search strategy." Instead of undertaking searches on a single taxa at a time, queries would use sets of related taxa, such as the taxa A–C in the subtree rooted at node 2. From reference [87].

### 10.1.1 Kinds of Trees

It is useful to distinguish between at least two different kinds of trees—classifications and phylogenies. Figure 10.2 shows a classification and a phylogeny for the plant order Nymphaeales (waterlilies).

**Classification.** A Linnaean classification can be represented as a rooted tree with all nodes labeled. Each node has a "rank," such as order, family, genus, or species (see Figure 10.2). Although the relative position of a rank in the taxonomic hierarchy is fixed, ranks are essentially arbitrary in that they are rarely comparable across different taxonomic groups. For example,