Education of Computational Biology

Dean L. Zeller
Kent State University

The main view of evolution is that all living species are derived from a common ancestor, and new species are created by a population splitting into two or more smaller populations with different characteristics. At some point, these populations are unable to cross-breed, as they are no longer genetically compatible. As such, a tree-structure has been used to model the evolution of species over time. While this structure is true in a vast majority of cases, there are circumstances in which a tree model is insufficient to simulate all aspects of evolution. Modeling the evolutionary process has taken on a different form to allow for the genetic possibilities that give a tree model difficulty. Phylogenetic distance graphs (k-leaf roots) are now used to generate a more accurate model of evolution. Converting a k-leaf root into a phylogeny is a problem that still unsolved for values of $k \geq 5$, severely limiting scientists in their phylogenetic analysis.

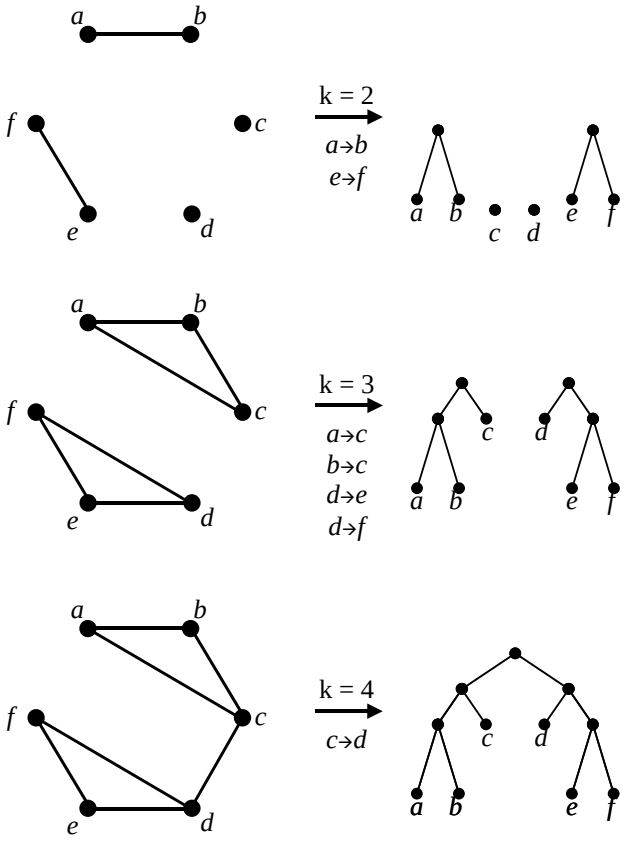
This presentation outlines a series of assignments in computational biology. The lessons require little prior knowledge and the subject matter is appropriate for any mathematics, computer science, or biology class. Two assignments on evolution trees and phylogenetic distance graphs have been tested on an introductory computer science class at Kent State University, with excellent results. The assignments focus on the algorithmic viewpoint of phylogeny reconstruction. The figures below illustrate two examples in reconstructing a phylogeny from a table of distance data among hypothetical species.

Bioinformatics is a technology still in its infancy. It is ripe with discoveries yet to be made. There are very few computational biologists when compared to the number of scientists of other fields, such as chemistry, physics, or geology. In order to create competent computational biologists in the next generation, lessons in bioinformatics must be published. Graph theory is a concept that can be learned at any age if taught correctly.

This presentation represents a needed shift of thinking in academics regarding topics in computational biology. Difficult topics such as computer science and genetics are commonly thought of as too complex for high school, and are usually saved until college. In order to encourage and properly train the computational biologists of tomorrow, the material should be covered at an early age. Assignments with this factor in mind should be created and tested for the educational community.

Data Table 1

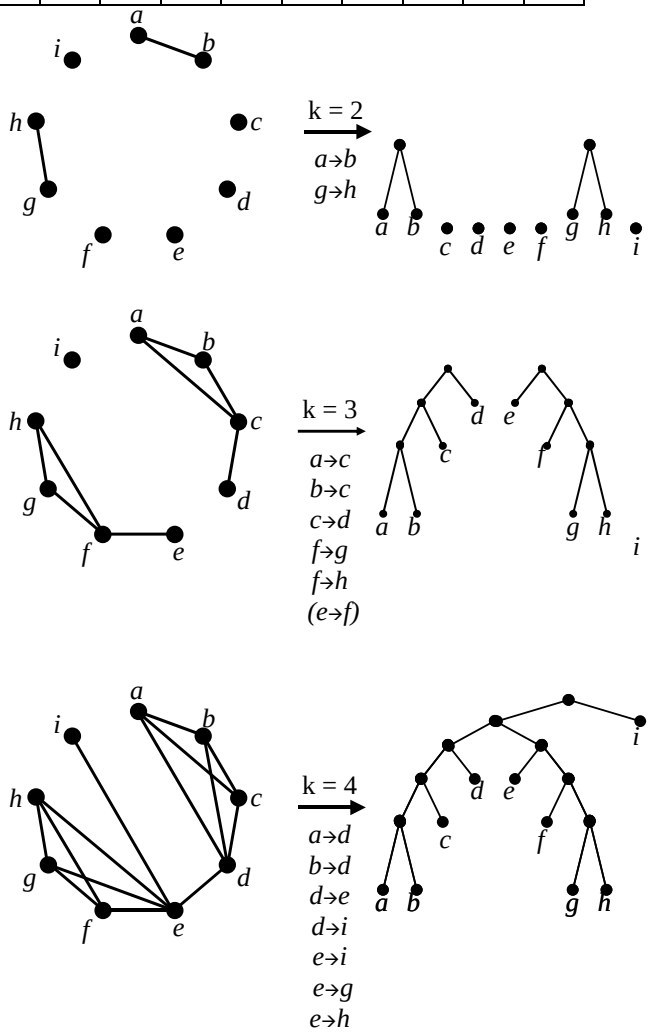
	a	*b*	*c*	*d*	*e*	*f*
a	0	2	3	5	6	6
b		0	3	5	6	6
c			0	4	5	5
d				0	3	3
e					0	2
f						0



Phylogeny reconstruction complete at k=4.

Data Table 2

	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*
a	0	2	3	4	6	7	8	8	6
b		0	3	4	6	7	8	8	6
c			0	3	5	6	7	7	5
d				0	4	5	6	6	4
e					0	3	4	4	4
f						0	3	3	5
g							0	2	6
h								0	6
i									0



Phylogeny reconstruction complete at k=4.