**Polyploidapalooza**

**Module 4: Polyploid phylogenetics**

Kaitlin E. Allen & David C. Blackburn

**Abstract**

Polyploidy plays an important role in the evolution of life but is often overlooked in undergraduate classrooms. In this series of modules, we explore how polyploid species are formed, how they evolve and adapt, and their economic and conservation importance. Module 4 explores how to read and understand polyploid phylogenetic trees and how the genomic structure of polyploid organisms impacts phylogenetic analysis.

**Objectives**

* Conduct an overview of how to read and understand basic phylogenies
* Explore polyploid phylogenies in the form of multi-labeled trees - including the difference between allopolyploid and autopolyploid topographies
* Explore allopolyploid phylogenies in the form of species networks

**Logistics**

Department: Biology

 Level: Upper undergraduate

Prep time: 0 hours

 Class time: 2 hours

**Pre-module preparation**

List of terms - come back to these terms if you need to while reading the text.

 **Phylogenetic tree**—diagram showing the bifurcating evolutionary relationships

between species

**Phylogenetic network**—diagram showing evolutionary relationships between species

that also includes species formed through hybridization

**Lineages**—representation of a species on a phylogenetic tree, the length can

correspond

to the amount of evolutionary divergence between species

**Node**—a representation of the most recent common ancestor on a tree, the point where two lineages come together.

**Clade**—two or more lineages that share a node form a clade

 Sister species—species connected by a node, the closest relative of a species or

clade

 **Bifurcating**—Splitting into two lineages from a node

**Introduction**

 Phylogenetics is the study of how living organisms are related to each other. You can think of it as similar to a family tree, where each species or population in the tree is descended from a common ancestor. These relationships are often determined using similarities in morphology or genetics. Typically, these relationships are represented as **bifurcating**, meaning that when new species are formed the ancestral species (represented as a **node)** splits into two **lineages**. In the first activity of this module, we will explore how to read and interpret a bifurcating **phylogenetic tree.**

 Polyploid species exhibit more complicated phylogenetic relationships than diploid species because they have more than one genome, each undergoing its own evolutionary processes. When the evolutionary relationships of genes from each genome are plotted in a bifurcating phylogenetic tree, the two categories of polyploids (autopolyploids and allopolyploids) show different patterns. Autopolyploids, which are polyploids with duplicated genomes of the same species, show branching patterns where the two genomes are closely related to each other. Meanwhile, allopolyploids, which are polyploids formed by hybridization between two distinct species, will show branching patterns where the two genomes are found separately in the tree and are most closely related to the species that hybridized to form the polyploid. Because polyploid species are represented multiple times in the phylogeny, due to their multiple genomes, these trees are called multi-labeled gene trees. This will be covered in the second activity of this module. In allopolyploids, these trees can be redrawn as reticulating **Phylogenetic networks** to represent the hybridization that formed the species (activity 3).

**Activity 1: How to read and interpret phylogenetic trees**

Before we get into how to work with polyploid multi-labeled trees and networks, we will quickly review the basics of how to read and interpret phylogenies in general. Below is a tree with five species (labeled A, B, C, D, & E). Each ancestral node splits into two lineages, meaning that this tree is bifurcating. The species that are most closely related to each other (called **sister species**) are connected by a node (labeled 1–4). Species like C are said to be sister to the lineage containing both species A and B. Together these three species share a common ancestor at node 2. Species C is equally related to A and B, just as your cousin would be equally related to you and your sibling. Groups of species that are closely related to each other by a common node are called **clades**, so A and B are a clade joined by node 3; A, B, and C are a clade joined by node 2, and A, B, C, D, and E are a clade joined by node 1. Species that are not joined by a common node do not form clades, so A and D do not form a clade unless all of the other species that are connected by that node are included.



Using the phylogeny above, answer the five (5) questions below to check your comprehension of phylogenetic relationships before moving forward to apply these concepts to polyploid phylogenies.

1. Using the terminology above, what is D's relationship to E?

**Sister species**

1. What node joins species D and E?

**Node 4**

1. What node joins D and E with A, B and C?

**Node 1**

1. Do C and E form a clade? Why or why not?

**C and E do not form a clade by themselves because they are not joined by a common ancestor/node. They could be considered as part of a larger clade joined by node 1**

1. Do D and E form a clade? Why or why not?

**Yes because they are joined by a common ancestor/node 4**

**Activity 2: Multi-labeled trees**

Now that we have reviewed the basics of reading phylogenies, we will apply these concepts to polyploid phylogenies. There are two types of polyploids that we will cover in this activity: autopolyploids and allopolyploids. These are formed in different ways and show different phylogenetic patterns. Autopolyploids are formed through genome duplication within a single species. In a phylogenetic tree showing the evolutionary relationship between the genomes of autopolyploids and other closely related species, their duplicated genomes will appear as each other's closest relatives, and each will be labeled with the same species name since they come from the same species. In the first phylogeny below, species A is an example of an autopolyploid with two sub-genomes (A1 and A2). Allopolyploid species are formed through hybridization between two distinct parental species. In a bifurcating phylogenetic tree showing the evolutionary relationships between the genomes of different closely related species, the genomes of the allopolyploid species will each appear as more closely related to the species that hybridized, rather than to each other, as in an autopolyploid. Like autopolyploids, each lineage associated with the allopolyploid will be labeled as the same species since they come from the same species. In the second phylogeny below, species B is an allopolyploid with sub-genomes B1 and B2.



**Tree 1: autopolyploid multi-labeled phylogenetic tree**



**Tree 2: allopolyploid multi-labeled phylogenetic tree**

For this exercise use the phylogenies above to answer the following four (4) questions.

1. List two species in the above trees that are autopolyploid based on their phylogenetic relationships

**A and E in tree 1**

1. List one species in the above trees that is allopolyploid based on its phylogenetic relationship

**B in tree 2**

1. How can you tell that tree 1 represents autopolyploids while tree 2 represents allopolyploids?

**Tree 1 has autopolyploids because duplicate genomes are sister to each other. Tree 2 has allopolyploids because duplicate genomes are sister to other species**

1. What is the biological difference between auto and allopolyploids that causes the differences in phylogenetic patterns?

**Autopolyploids formed through duplication in a single species, allopolyploids formed by duplication through hybridization between different species**

**Activity 3: Allopolyploid species networks**

 As we reviewed above, allopolyploids are formed through hybridization and their genomes appear in the phylogenetic tree closest to the hybridizing species. Because of this, the hybridization pattern that created the allopolyploid can be deciphered from the relationships in the phylogenetic tree. This tree can then be redrawn as a phylogenetic “network” where lineages can be shown coming back together into the polyploid species, which, even though it shares the genomes of other species in the tree, is now on its own evolutionary path. Below you can see the same tree with the allopolyploid species B as determined in activity 2.



From this phylogeny representing parts of the genome, you can see that one sub-genome from species B is most closely related to species A and the other sub-genome is more closely related to species C. This means that B is an allopolyploid through an hybridization event between these species A and C. A network representing relationships among species can be drawn to show this hybridization pattern, as below.



Next, take a look at the phylogenetic tree below and answer the following three (3) questions, including one which will have you practice drawing a phylogenetic network on your own.



1. Which species is allopolyploid in this phylogeny?

**Species D**

1. Which two species are the allopolyploid species most closely related to?

**Species C and E**

1. Draw the phylogenetic network based on this phylogeny in a style similar to that above.



**Acknowledgments**

We would like to thank Makenzie Mabry, Michelle Gaynor, Libby Ellwood, and Alva Mihalik for their thoughtful feedback on these modules.

**References**

Phylogenetic tree figures created by Kaitlin Allen