Apr 27th, 3:00 PM - 4:00 PM

Branching Out: Generating an Evolutionary Tree of Southeast Asian Plants with Computational Tools

Conner J. Copeland
University of Montana, Missoula, conner.copeland@umconnect.umt.edu

Travis Wheeler
University of Montana, Missoula

Jedediah Brodie
University of Montana, Missoula

Copeland, Conner J.; Wheeler, Travis; and Brodie, Jedediah, "Branching Out: Generating an Evolutionary Tree of Southeast Asian Plants with Computational Tools" (2018). *University of Montana Conference on Undergraduate Research (UMCUR)*. 1.
https://scholarworks.umt.edu/umcur/2018/pmposters/1

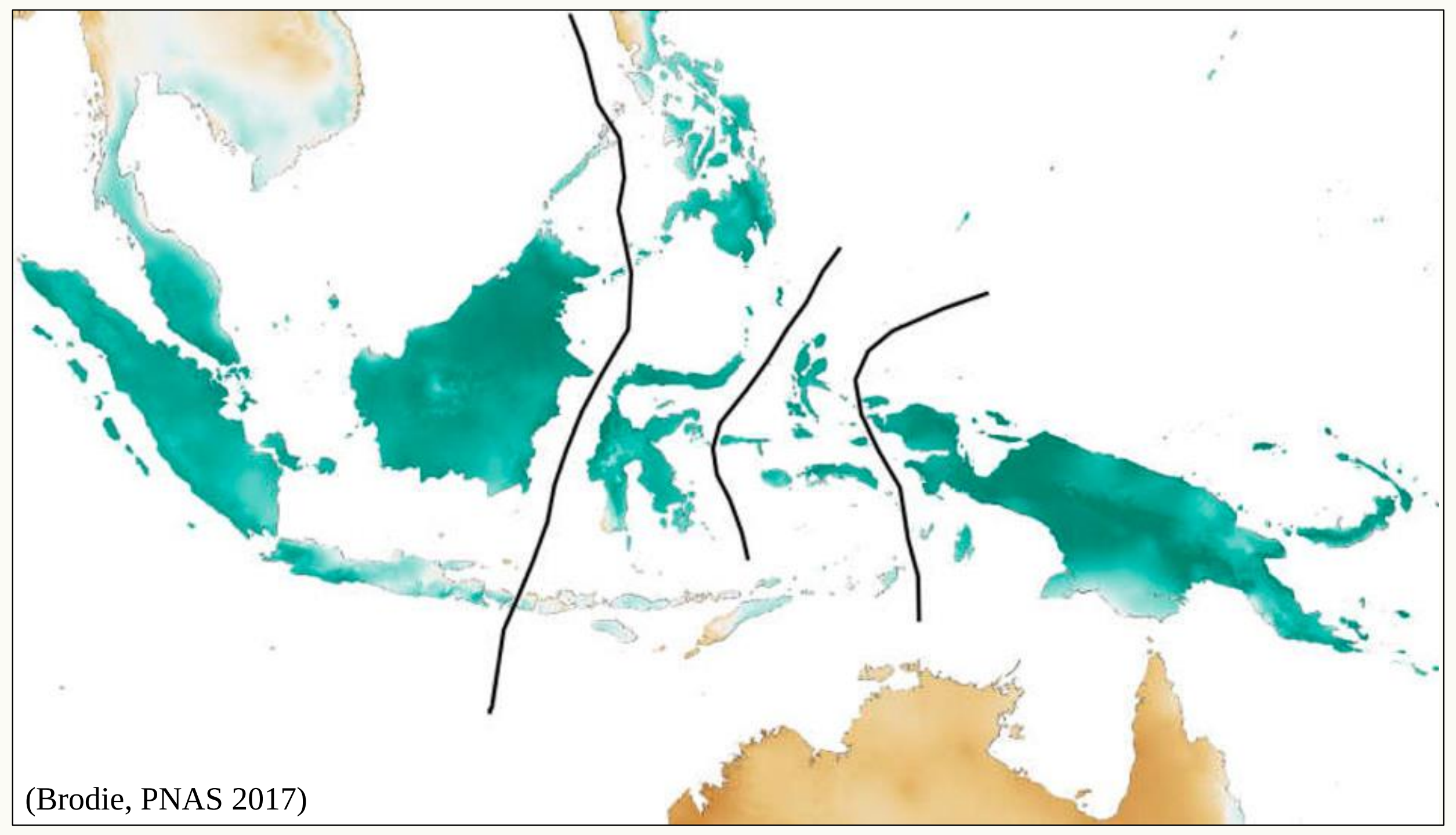
Branching Out: Generating an Evolutionary Tree of Southeast Asian Plants with Computational Tools

Conner Copeland[1], Jedediah Brodie[2], Travis Wheeler[1]

[1]Dept. of Computer Science, University of Montana, Missoula, MT, USA; [2]Dept. of Biological Sciences, University of Montana, Missoula, MT, USA

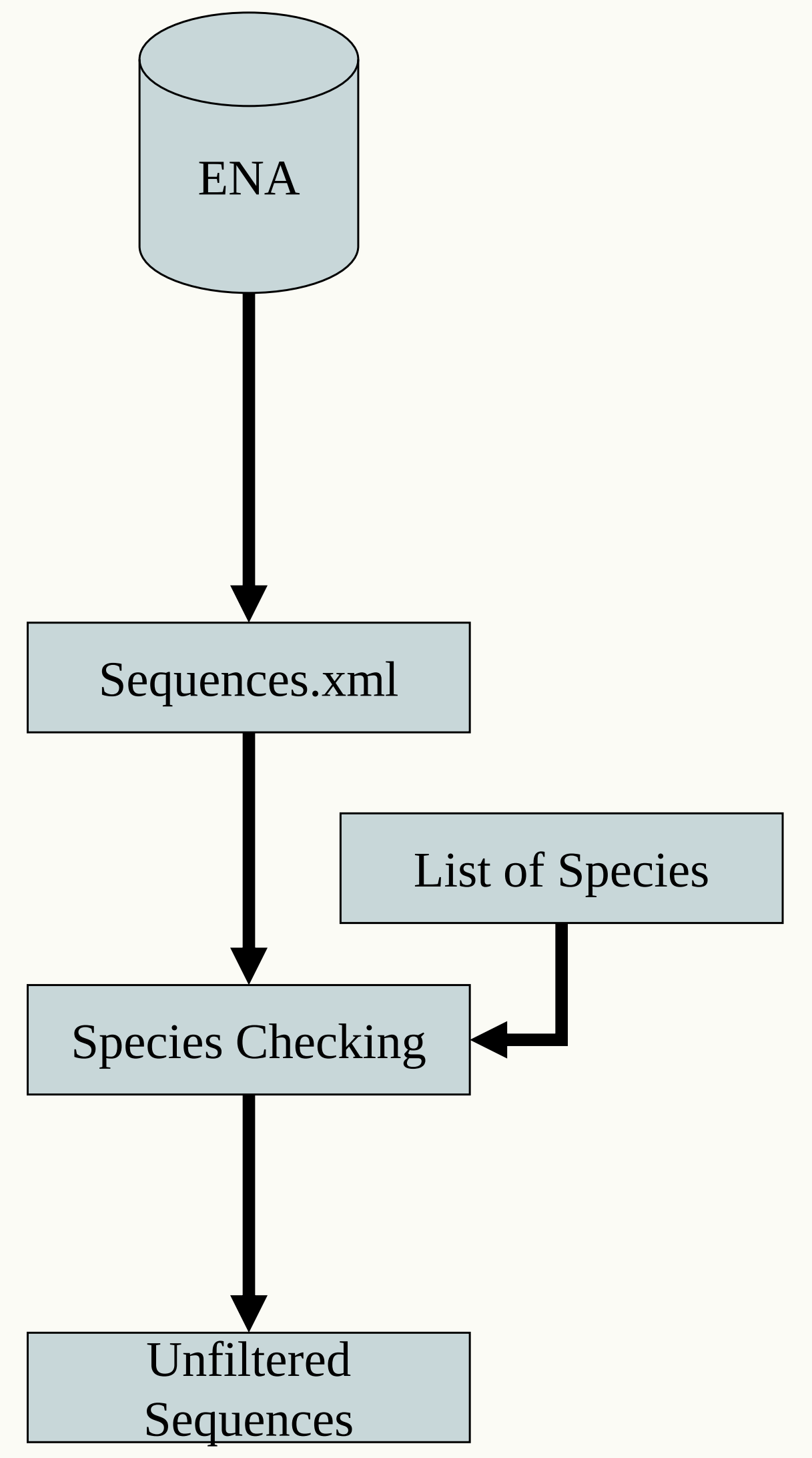
Marsupials, Mammals, and Angiosperms: An Untold Tale of Evolutionary History

The islands of the Southeast Asian archipelago are home to a distinctive gradient of animal species: Islands closer to mainland Asia are home to mammals, while those closer to Australia host marsupials. This is illustrated by the figure below, which demonstrates how the region can be divided into sections based on animal populations. However, the islands are home to plant species from the same genera, presenting a unique opportunity to investigate how interactions with different animals have shaped the evolution of the islands' plants. Efforts to research this phenomenon by the Biology Department's Dr. Jedediah Brodie have been hampered by a lack of a phylogeny, or evolutionary tree, of these plant species. This project seeks to provide a phylogeny for use in his research, and to fill an important hole in our knowledge of evolutionary history.



(Brodie, PNAS 2017)

Sequence Capture



ENA

Sequences.xml

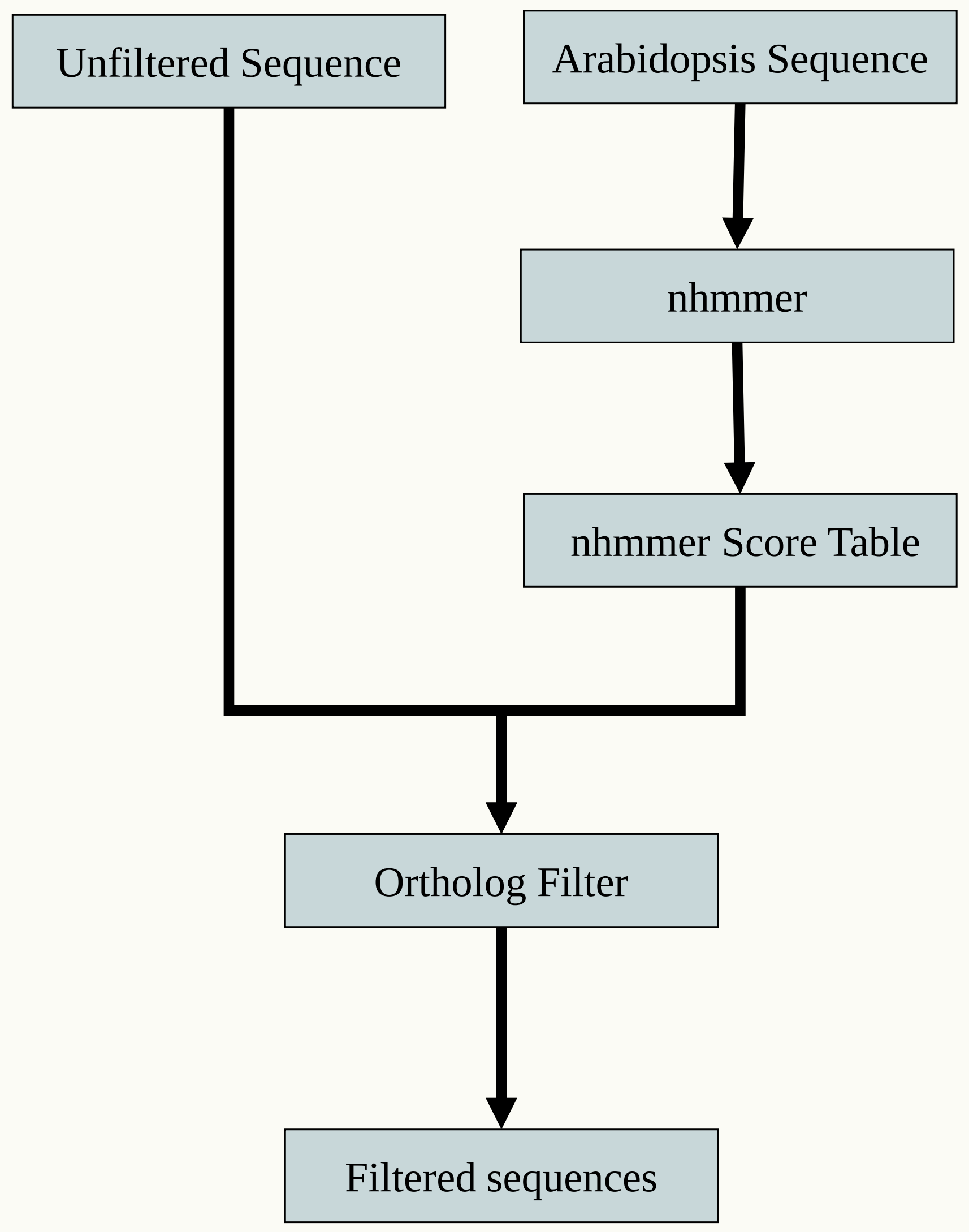
List of Species

Species Checking

Unfiltered Sequences

Using a scripted interface to the European Nucleotide Archive, we downloaded DNA sequence data for three genomic regions corresponding to a list of Southeast Asian plant species, saving entries belonging to any species on the list. Sequences for over 1200 species were captured.

Ortholog Filtering



Unfiltered Sequence

Arabidopsis Sequence

nhmmer

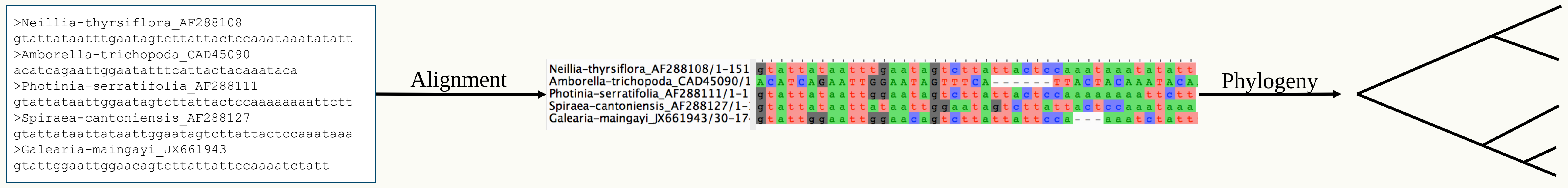
nhmmer Score Table

Ortholog Filter

Filtered sequences

Many of the species in this group were represented by multiple homologous copies of the searched genes. We restricted this list in an attempt to capture orthologous copies of the genes. This was achieved by comparing each sequence to a single instance of the gene from Arabidopsis Thaliana (using HMMER), and picking the highest-scoring match.

Generalized Phylogenetic Inference Pipeline



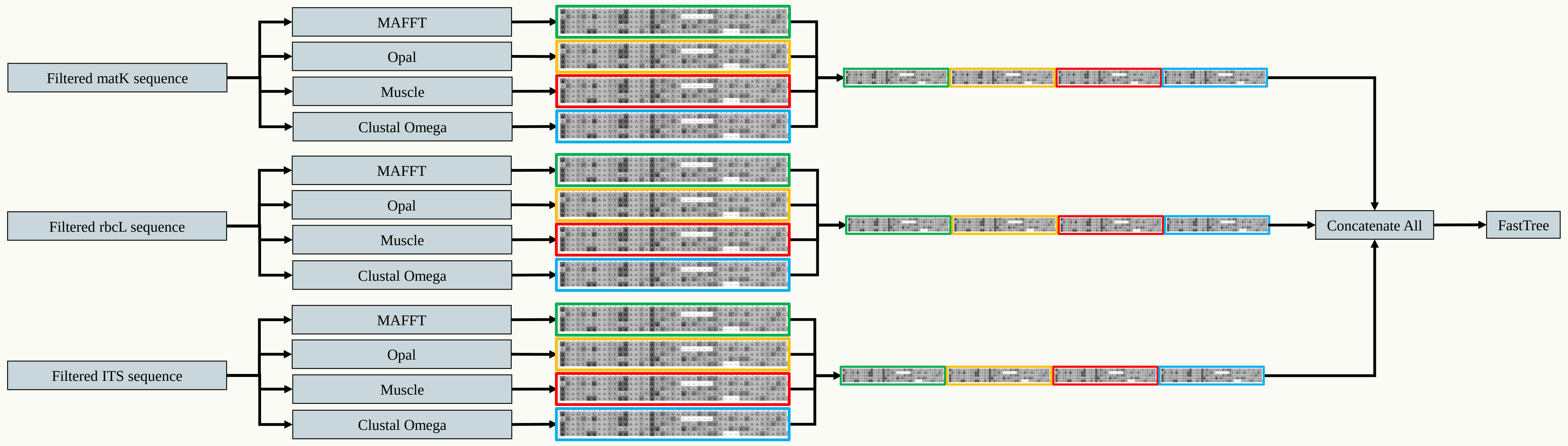
```
>Neillia-thyrsiflora_AF288108
gtattataatttgaatagtcttattactccaaataaatatatt
>Amborella-trichopoda_CAD45090
acatcagaattggaatatttcattactacaaataca
>Photinia-serratifolia_AF288111
gtattataattggaatagtcttattactccaaaaaaaattctt
>Spiraea-cantoniensis_AF288127
gtattataattataattggaatagtcttattactccaaataaa
>Galearia-maingayi_JX661943
gtattggaattggaacagtcttattattccaaaatctatt
```

Alignment

Phylogeny

Generally, phylogenetic inference begins with DNA sequences being run through Multiple Sequence Alignment (MSA) tools, which line up sections of DNA that are likely related and place gaps where the sequences don't match. Phylogenetic inference tools then use the alignments to create likely evolutionary trees.

Our Pipeline



Filtered matK sequence

Filtered rbcL sequence

Filtered ITS sequence

MAFFT

Opal

Muscle

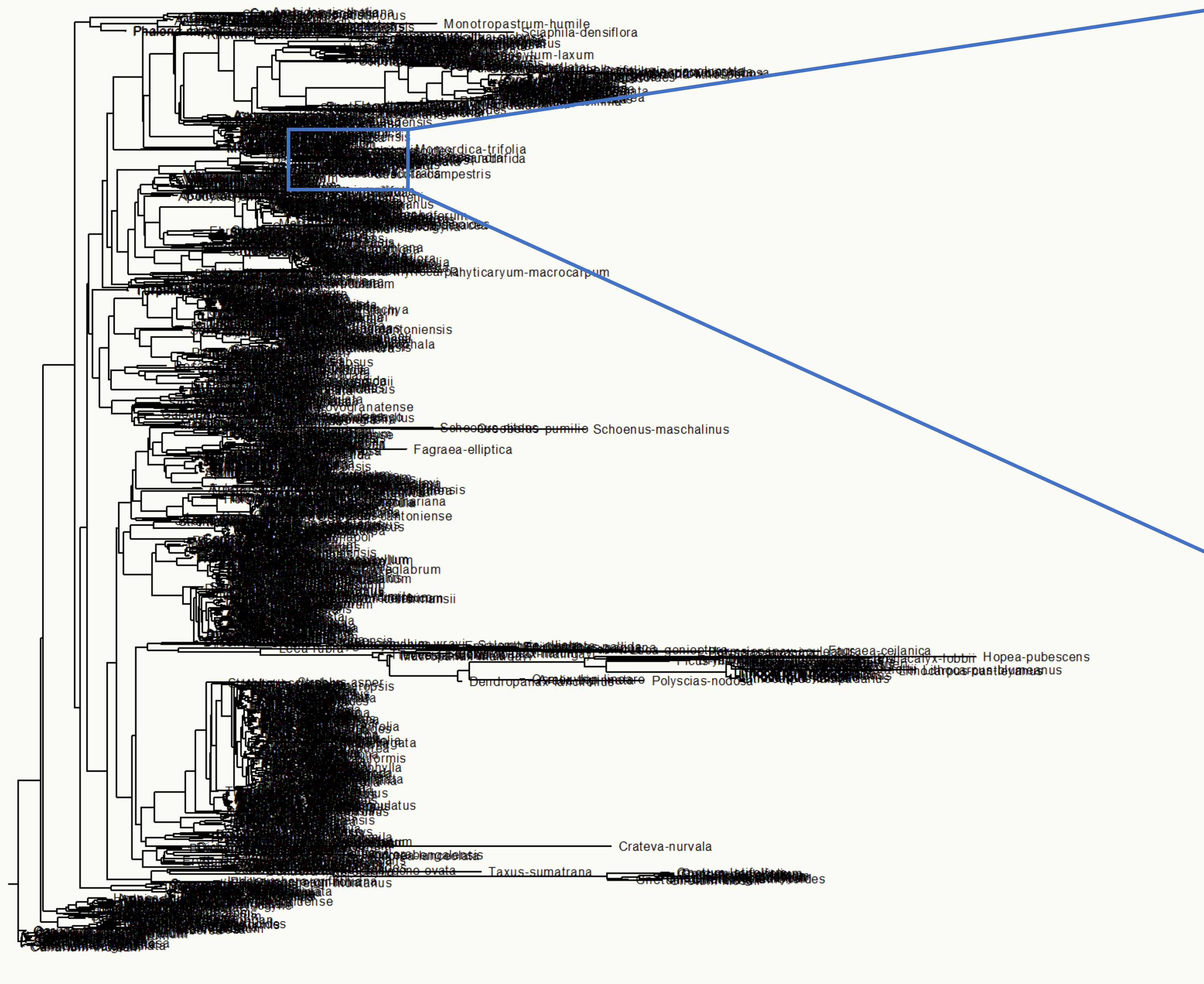
Clustal Omega

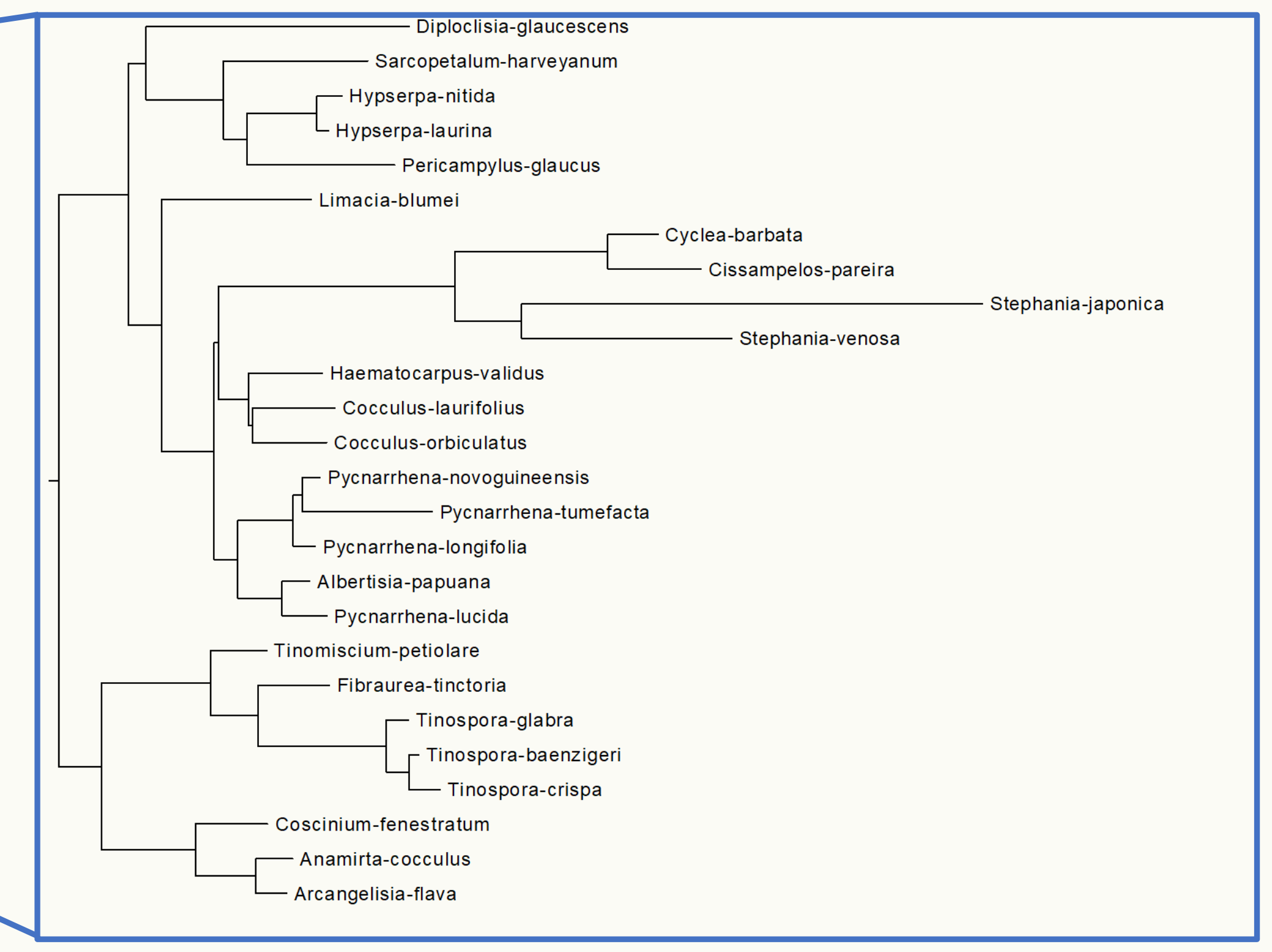
Concatenate All

FastTree

Our pipeline works similarly to the general method, using 3 sequences (matK, rbcL, and Internal Transcribed Spacers 1 and 2) and 4 MSA-building tools (Opal, MAFFT, Muscle, and Clustal Omega). Multiple tools were used to limit the impact that software bias may have on the alignment input to the phylogeny inference software. In this way, highly supported regions are likely to be found in all alignments, while uncertain regions will differ from tool to tool, receiving less support. The tool FastTree was used to build an approximate Maximum Likelihood phylogeny from the resulting concatenated MSA.

Current Results





The figure to the left represents the phylogeny constructed using the pipeline above, and includes over 1200 species. In the zoomed region shown above, we demonstrate that typical genus-level clusters are largely represented in the tree. Development of this phylogeny will support future analyses of phylogeographic and species-coevolutionary processes.

Thanks to the members of Wheeler Lab for their help with script design, bugfixing, and Regex!