Phylogenies derived from somatic mutations agree with physical topologies in *Eucalyptus*

Adam J Orr [1,2] Robert Lanfear [3,4] Reed Cartwright [1,2]

[1]School of Life Sciences, Arizona State University
[2]Biodesign Institute, Arizona State University
[3]Department of Biological Sciences, Macquarie University
[4]College of Medicine, Biology and Environment, Australian National University

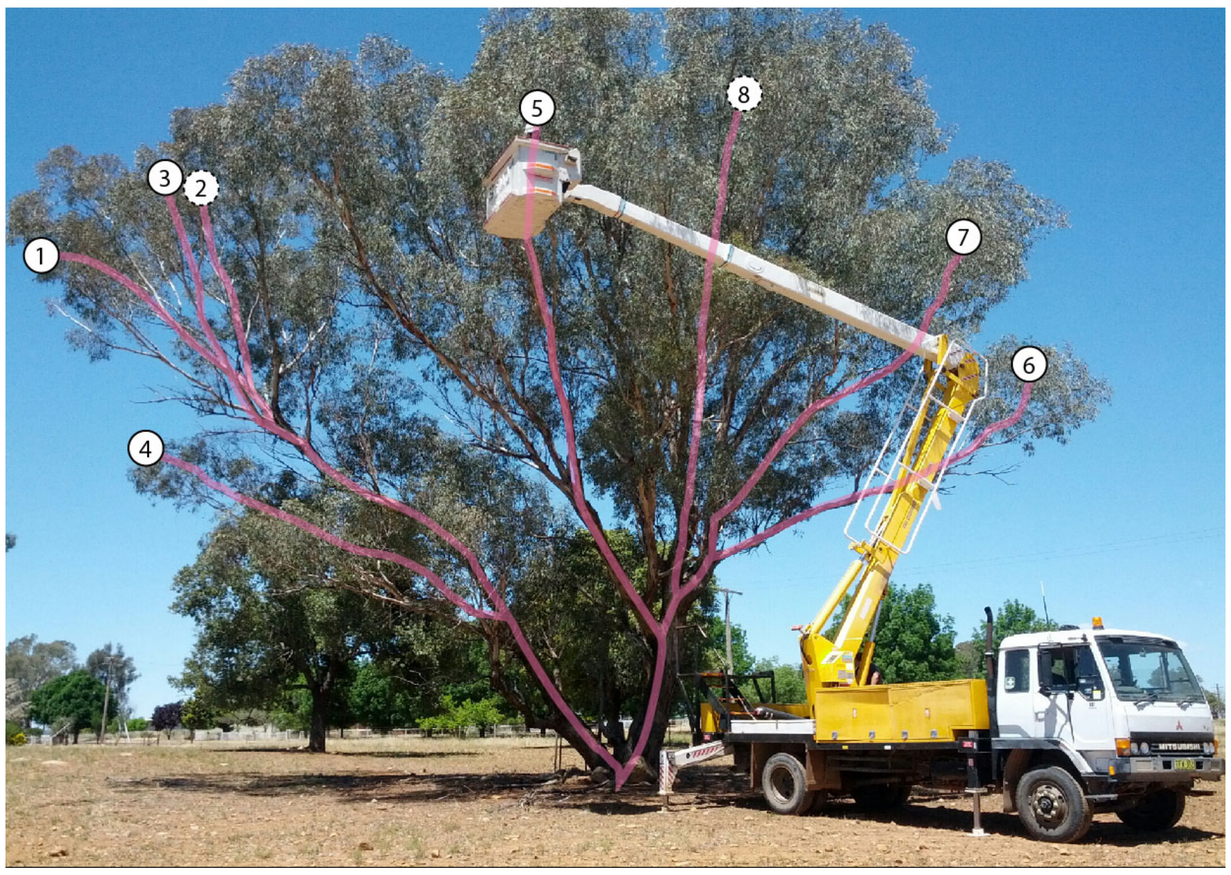
Abstract

Eucalyptus melliodora, a tree native to eastern Australia, has strong timber and a high nectar load, making it economically important and a vital source of food for nectar-consuming species. In 1993, an individual of the species was discovered that harbors a somatic change conferring herbivore resistance to a section of branches on the tree via differential terpenoid production. Though transcriptomic analysis was inconclusive in determining the genetic source of this variation, we attempt to do so using ultra-deep whole-genome sequencing of 8 samples in triplicate. We call variants using a reference-free De-Bruijn variant caller *DiscoSNP++* and by the GATK best practices using a reference from a closely-related species. We find that the phylogeny of the variants identified by both methods reflects the branching pattern of the tree, though the phylogeny is affected by short interior nodes. While we have yet to validate the source of the herbivore resistance, this data presents an opportunity for further study of how somatic mutations are produced and spread in plants and how to properly resolve short interior nodes.

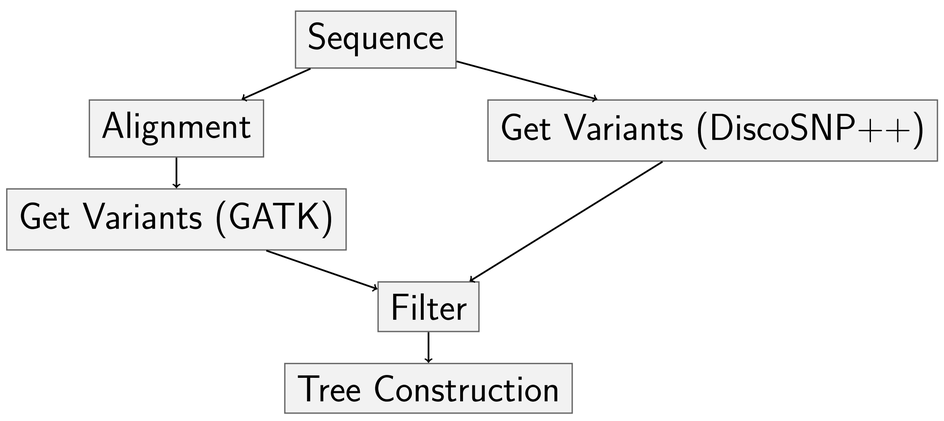
Introduction

- Somatic mutations are rare but sequencing errors are common, making somatic mutations difficult to detect.
- Little is known about the spread of somatic mutations, despite the key role they play in cancer development.
- If the pattern of mutations in the phylogeny matches the branching pattern of the plant, then plants can be used to easily validate somatic mutations.

Methods: Variant Detection



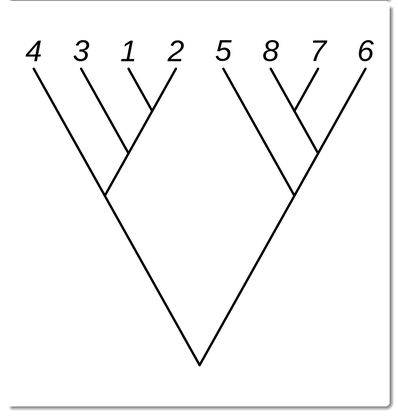
- Samples 1-8 were collected in triplicate
- Each replicate was Illumina sequenced
- Each sequence was aligned to genome of *Eucalyptus grandis* using bwa mem.
- Variants were called using GATK's UnifiedGenotyper or DiscoSNP++, a reference-free variant caller
- Nonvariable sites and gaps were removed
- Variants were removed if the genotypes of all replicates of a sample were not identical
- A maximum likelihood tree was constructed with RAxML



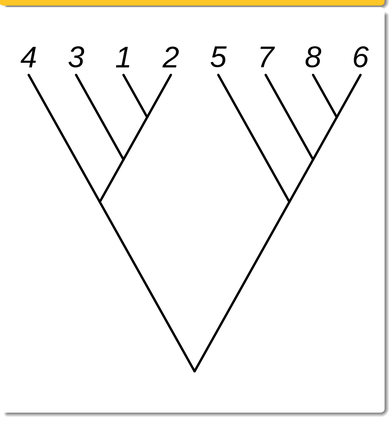
Results: Variant Detection

- The toplogy of the tree built from variants called by reference-free method DiscoSNP++ closely resembles the branching pattern of the plant.
- Its performance matches that of GATK used with a reference from the closely-related species *Eucalyptus grandis*.
- A short internode causes incorrect placement of branches 6, 7, and 8 for both methods.

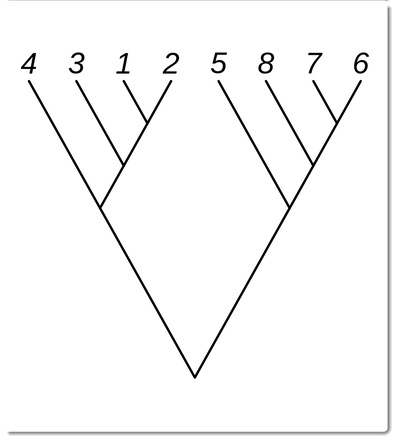
Tree Created with Variants From GATK



Tree Created with Variants From DiscoSNP++



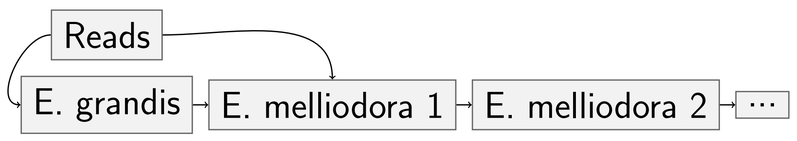
True Topology



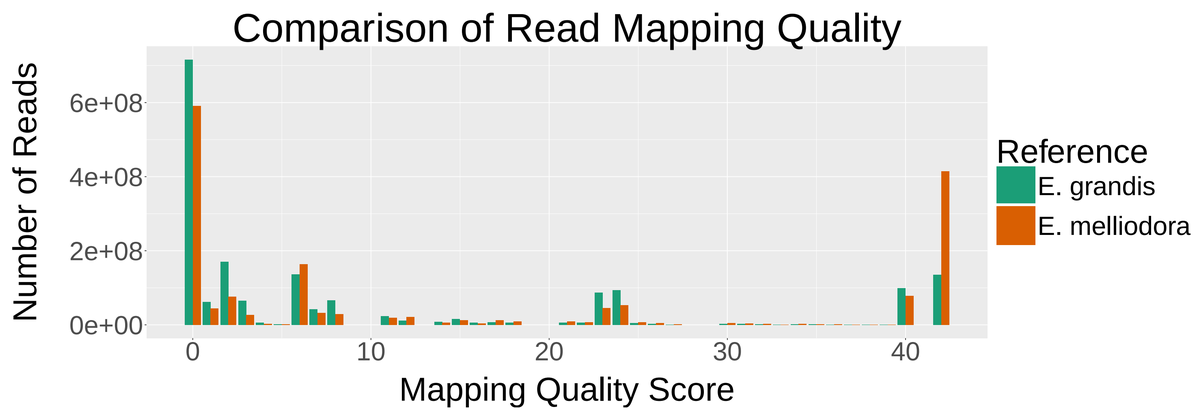
Next steps: Reference Improvement

To improve resolution of short internodes, we attempt to modify the *E. grandis* genome to make it more suitable for *E. melliodora* by:

- Aligning the reads to the *E. grandis* genome
- Creating a consensus sequence from this alignment
- Aligning the reads to this consensus to create a draft *E. melliodora* genome



Results: Reference Improvement



Alignment to the consensus sequence produces an alignment with higher overall quality scores than alignment to the *Eucalyptus grandis* reference.

Conclusions

- Phylogenies of somatic mutations within a *Eucalyptus* tree match the branching patterns of the tree using both a reference-based and a reference-free variant caller.
- Aligning reads to a close relative, obtaining a consensus sequence, then realigning to that consensus seems to improve alignment quality.

Acknowledgements

This work is supported by grant NIH R01-HG007178.