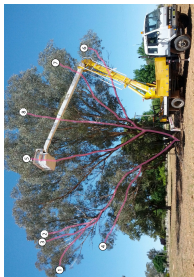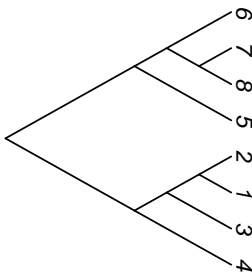# Methods for Detecting Somatic Mutation in Plants
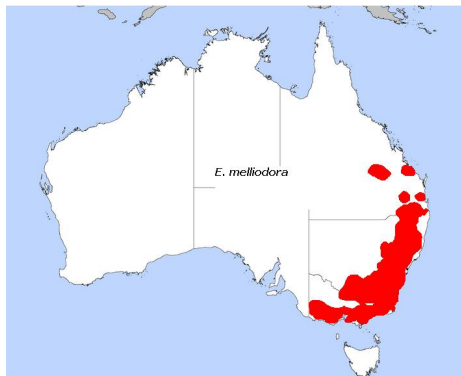
Adam Orr

5/02/17

The Yellow Box Tree: *Eucalyptus melliodora*

- Produces **5 times** more nectar than smaller trees.
- Food source for bees
- Strong wood used for bridges

E. melliodora

1

---

[1]https://commons.wikimedia.org/wiki/File:E._melliodora.JPG

A Genetic Mosaic

- Edwards identified as mosaic in 1993[2]
- Sheep pen in Yeoval, New South Wales
- Differential oil production gives protection from Christmas beetles
- Is this mutation a controlled process?

[2] *Edwards PB, Wanjura WJ, Brown WV. Oecologia 1993, 95:551557.*

Somatic Mutations are Commercially Interesting

Definition

A **somatic mutation** is a mutation that occurs in non-germline cells

- Nectarines arose from a somatic mutation on a peach tree
- In botany, this is called a **sport**
- Limited understanding of how plants grow

[3]

---

[3] https://commons.wikimedia.org/wiki/File:
White_nectarine_and_cross_section02_edit.jpg

Broad Implications

- How do somatic mutations spread? How can we study this?
- Cancers and other somatic diseases.
- A tree as a system for studying somatic mutation.
- The tree has a built-in control

What mutation is causing the herbivore resistance phenotype?
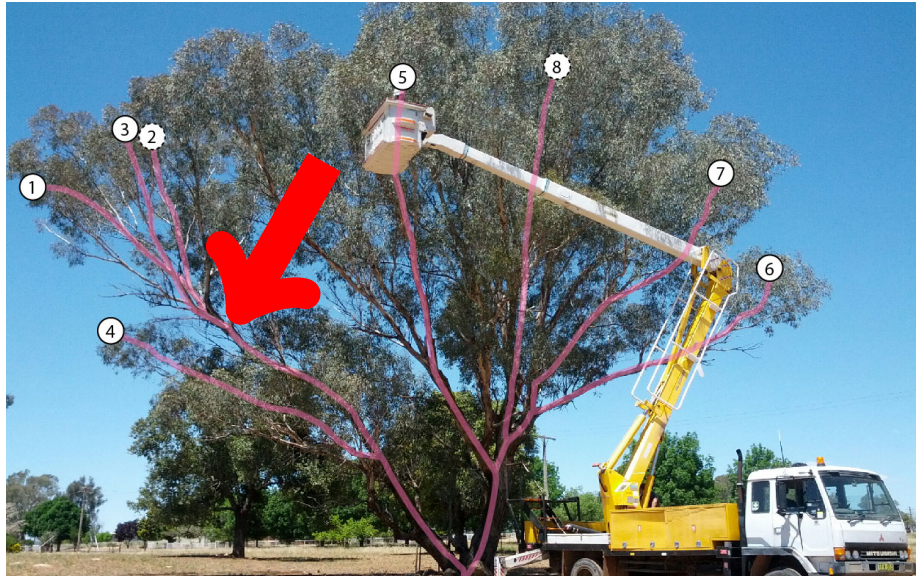
Mutations are very rare, but sequencing errors are very common.

Somatic mutations are hard to find

- Errors accumulate during PCR prior to sequencing - then propagate
- Errors accumulate in amplification steps during sequencing
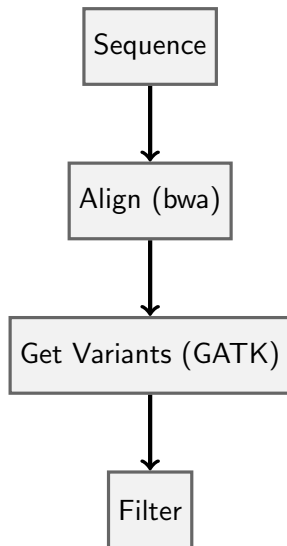- Technical error from sequencer

Sequencing error alone is $\sim 10^{-2}$ while mutation rate after error-checking is $\sim 10^{-10}$
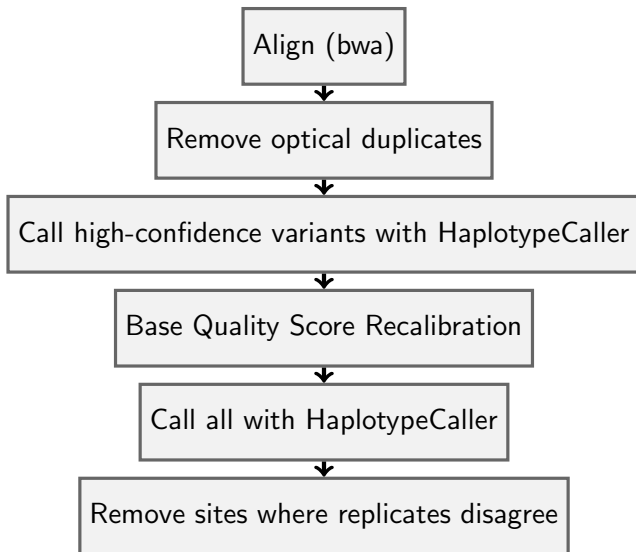
Study Methodology

Definition

Coverage: Average number of times a single base is sequenced.

- Sequence 8 samples in triplicate
- Ultra-deep coverage for each replicate ($\sim$30X)
- Align sequence to genome of *Eucalyptus grandis*
- Use replicates to remove false positives

```
Sequence
   |
   v
Align (bwa)
   |
   v
Get Variants (GATK)
   |
   v
Filter
```

GATK Best Practices

```
        ┌─────────────────────────────────────────┐
        │              Align (bwa)                 │
        └─────────────────────────────────────────┘
                          ↓
        ┌─────────────────────────────────────────┐
        │        Remove optical duplicates         │
        └─────────────────────────────────────────┘
                          ↓
┌─────────────────────────────────────────────────────────┐
│  Call high-confidence variants with HaplotypeCaller      │
└─────────────────────────────────────────────────────────┘
                          ↓
        ┌─────────────────────────────────────────┐
        │       Base Quality Score Recalibration   │
        └─────────────────────────────────────────┘
                          ↓
        ┌─────────────────────────────────────────┐
        │        Call all with HaplotypeCaller     │
        └─────────────────────────────────────────┘
                          ↓
        ┌─────────────────────────────────────────┐
        │    Remove sites where replicates disagree│
        └─────────────────────────────────────────┘
```
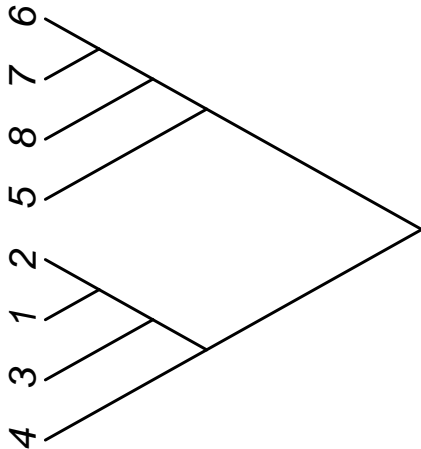
Mutation Pattern Approximately Matches Tree Structure
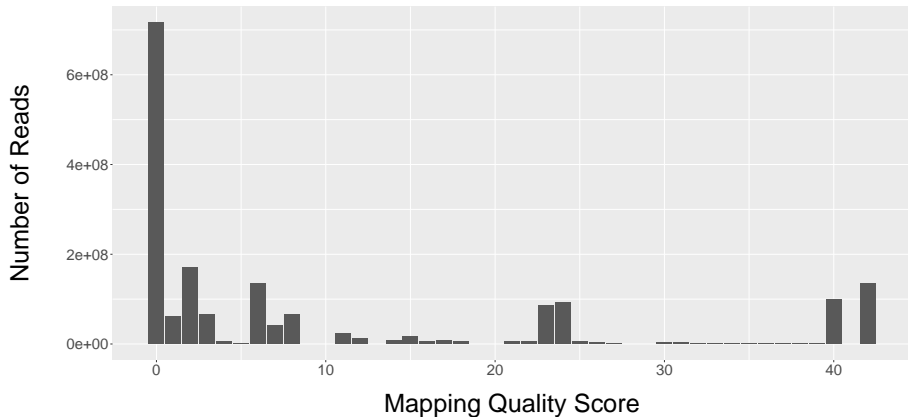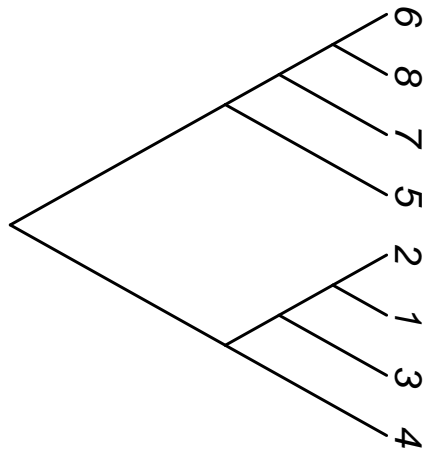

GATK Best Practices Tree
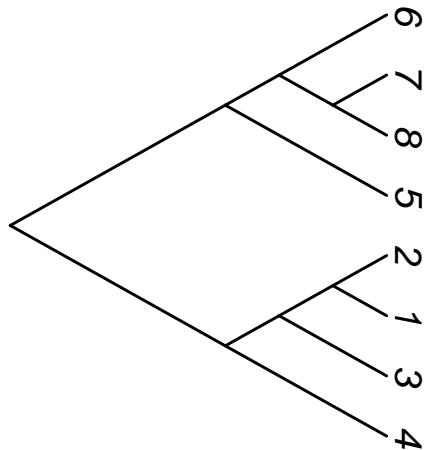
True Tree

Most Reads Are Not Mapped to the *E. grandis* Reference
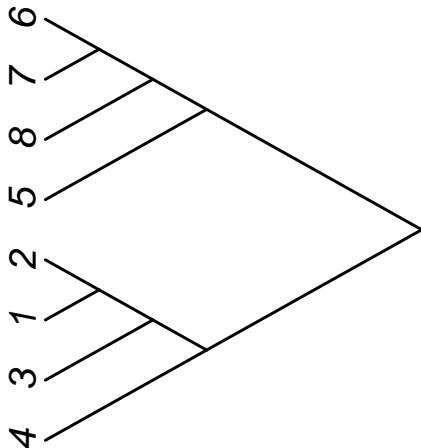

Quality of Reads Mapped
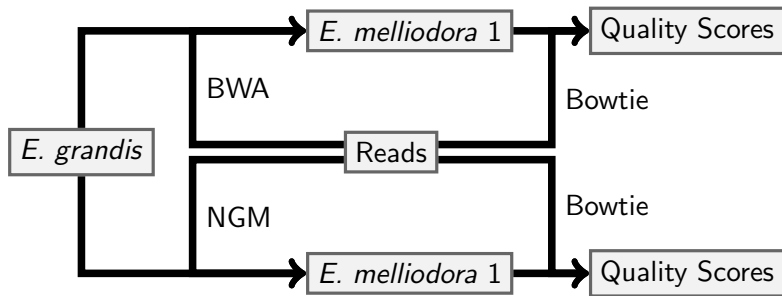
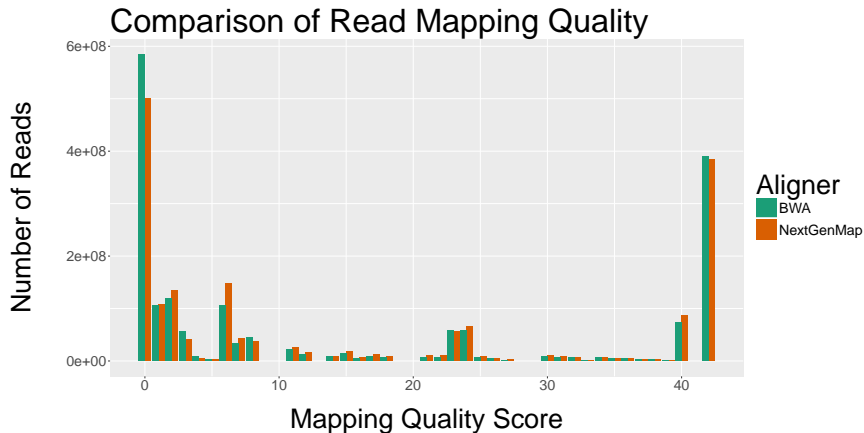# A Reference-Free Method Performs Similarly

DiscoSNP++ Tree
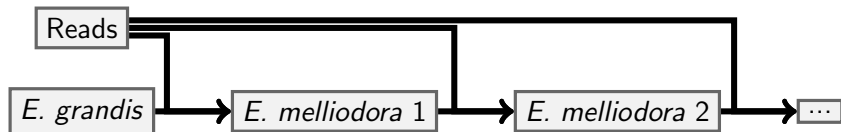
True Tree

Improving Alignment

The Choice of Aligner Impacts Mapping Quality

Comparison of Read Mapping Quality

If you want something done right...

Use *E. melliodora* genome as a starting place, then generate a new reference and map to that reference.

Our New Reference Improves Mapping Quality

Comparison of Read Mapping Quality

Next Steps

- Iterate to improve the reference we've created.
- Filter out repetitive elements that make mapping difficult
- Error correct reads

Dissertation Ideas

- Create variant caller able to utilize error models from diverse sequencing platforms
- GATK Base-Quality Score Recalibration-like process without need for alignment and truth-set variant calls
- Study error correction impact on alignment and variant calling