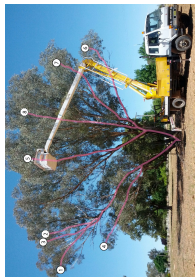
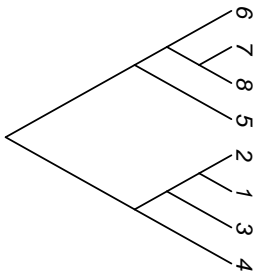


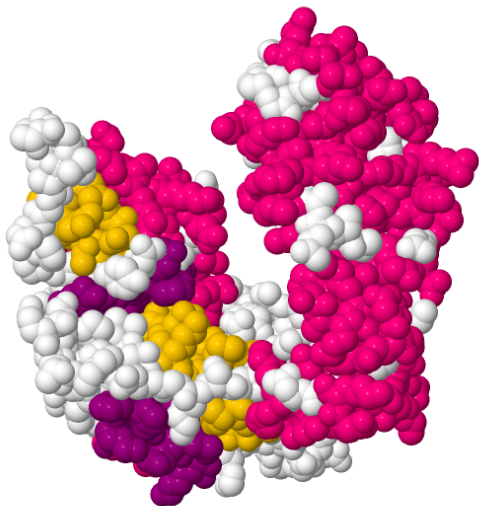
Calculating a non-model *Eucalyptus* Individual's Somatic Mutation Rate

Adam Orr  @AdamJOrr

2/16/18



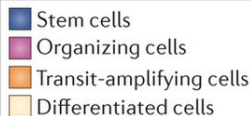
Somatic Mutations Occur During Replication Even Without Exposure to Mutagens



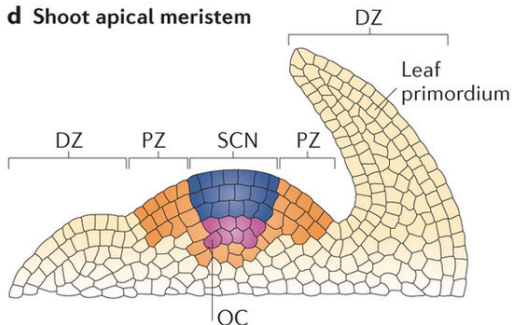
- DNA Polymerase β
- Mutation rate
 $\sim 10^{-9}$

PDB: 7ICG

Plants Grow Directionally



d Shoot apical meristem



- The genetic structure of the plant *should* mirror its physical structure.

⁰Heidstra & Sabatini (2014) Plant and animal stem cells: similar yet different.

Do Plants Evolve Differently?

NATURE VOL. 320 27 MARCH 1986

NEWS AND VIEWS

305

Somatic mutation

Do plants evolve differently?

from William J. Sutherland and Andrew R. Watkinson

It has recently been stressed that single ornamental plant with variegated leaves plants may consist of a mosaic of genetic clones. It is essential to remove branches

trees. This model also shows that somatic mutation may be an important source of variation within trees and tree populations. It seems likely that this mechanism is unimportant for species with relatively few meristems (such as peas and maize) but more important for long-lived and extensive species. Thus, individual clones of the aspen (*Populus tremuloides*) can cover

The relative importance of somatic and gametic mutations in plants cannot be assessed until the necessary measurements are made. But it is clear that somatic mutation could be important in many plant species.

Plants are weird.

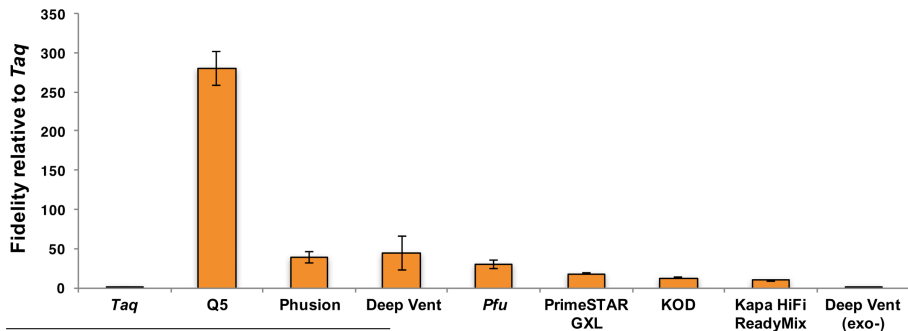
- Segregated germline?
- Dedifferentiation
- Selfing
- Mutational stress response?

Why are somatic mutations difficult to detect?

Mutations are very rare, but sequencing errors are very common.

Sequencing error alone is $\sim 10^{-2}$ while mutation rate after error-checking is $\sim 10^{-9}$

- Errors accumulate during PCR prior to sequencing - then propagate.
- *Taq* $\sim 10^{-4}$
- Technical error from sequencer



⁰ Potapov V, Ong JL (2017) Examining Sources of Error in PCR by Single-Molecule Sequencing

Why Care About Somatic Mutations?

Disease

- Cancer

Development

- Understanding the relationship between tissues

Agriculture

- Looking for interesting phenotypes in clonally reproducing species

Evolution

- Determining the relationship between somatic and germline mutation rate



⁰ https://commons.wikimedia.org/wiki/File:White_nectarine_and_cross_section02_edit.jpg

Project Goals

Is it possible?

Can we detect mutations with sufficient accuracy to reconstruct the physical structure of a tree?

The Relevant Measurements

What's the mutation rate like? Is there evidence of hypermutation?

A Genetic Mosaic

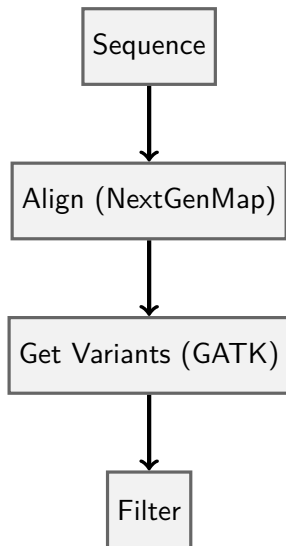


- Edwards identified as mosaic in 1993¹
- Sheep pen in Yeoval, New South Wales
- Differential oil production gives protection from Christmas beetles

¹Edwards PB, Wanjura WJ, Brown WV. *Oecologia* 1993, 95:551–557.

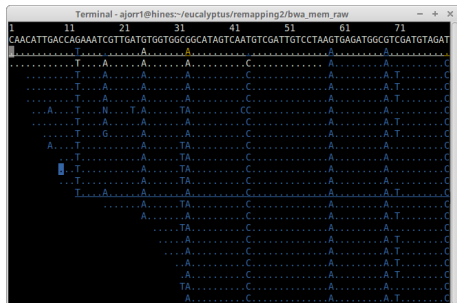
Study Methodology

- Sequence 8 samples in triplicate
- ~10X coverage for each replicate
- Align sequence to genome of *Eucalyptus grandis*
- Use replicates to remove false positives



Study Methodology

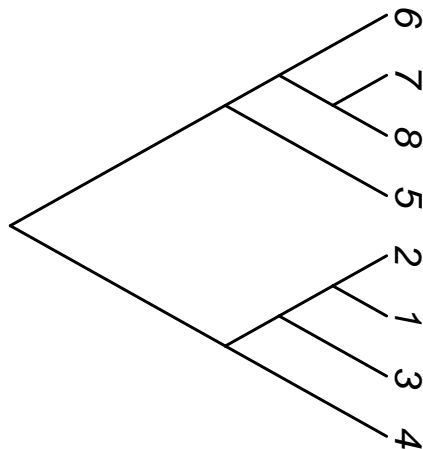
- Sequence 8 samples in triplicate
- ~10X coverage for each replicate
- Align sequence to genome of *Eucalyptus grandis*
- Use replicates to remove false positives



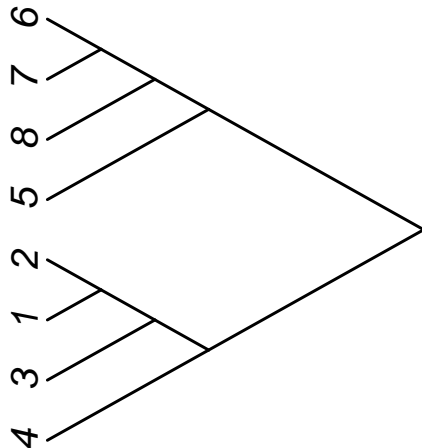
A terminal window titled "Terminal - ajorrt@hines:~/eucalyptus/remapping2/bwa_mem_raw" displays sequence alignment results. The top line shows a reference sequence: "1 11 21 31 41 51 61 71 CAACATTGACCAGAAATCGTTGATGGTGGCGGCATAGTCAATGTCGATTGTCTTAAGTGAGATGGCGTCGATGTAGAT". Below this, multiple lines of alignment are shown, with columns of dots representing gaps and letters (A, C, G, T) representing matches or mismatches. The alignment is visualized with a grid of dots and letters, indicating the mapping of individual sequencing reads to the reference genome.

Mutation Pattern Approximately Matches Tree Structure

GATK Best Practices Tree

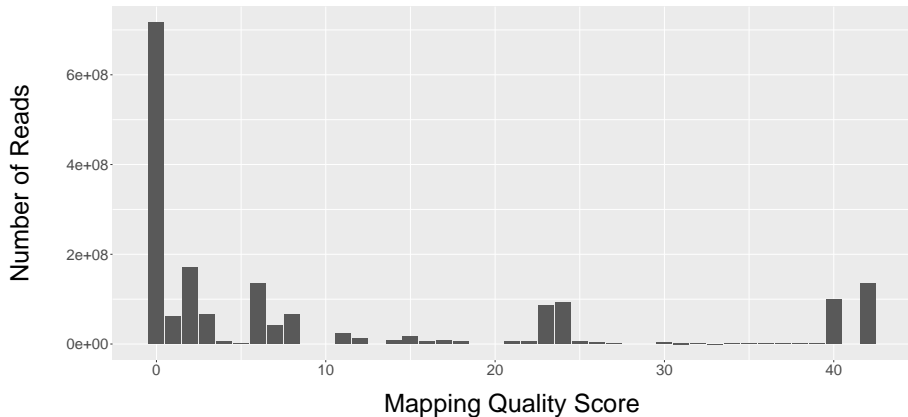


True Tree



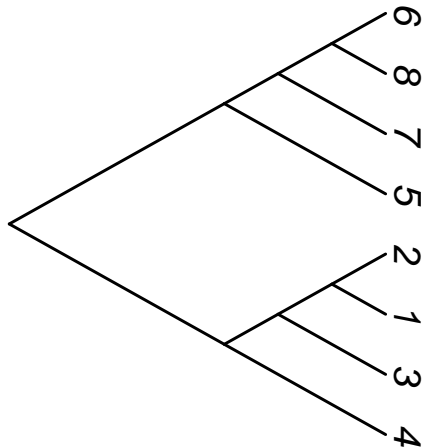
Most Reads Are Not Mapped to the *E. grandis* Reference

Quality of Reads Mapped

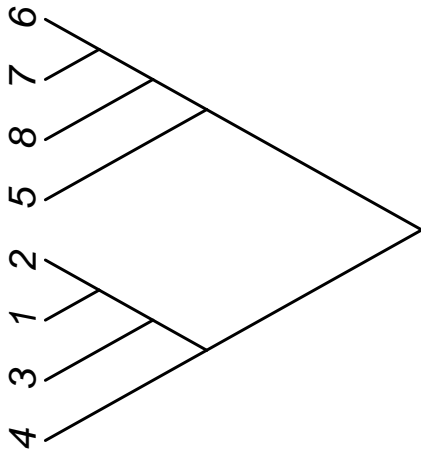


A Reference-Free Method Performs Similarly

DiscoSNP++ Tree

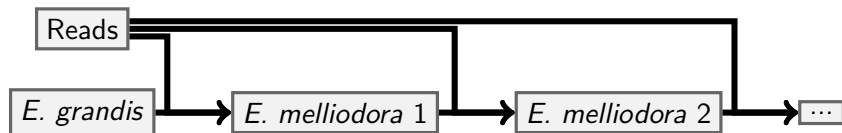


True Tree

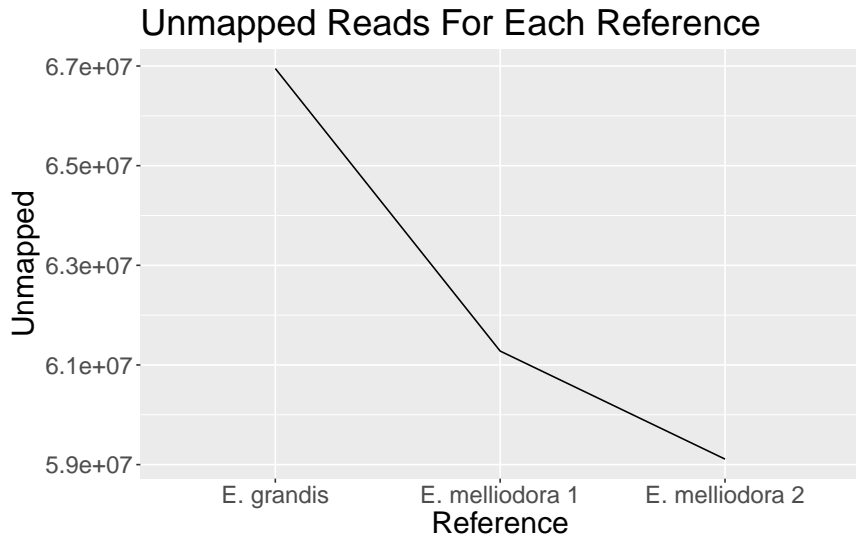


Approximating a Genome

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



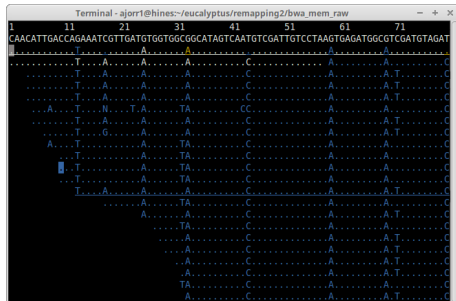
Our New Reference Has Fewer Unmapped Reads



Filtering Variants

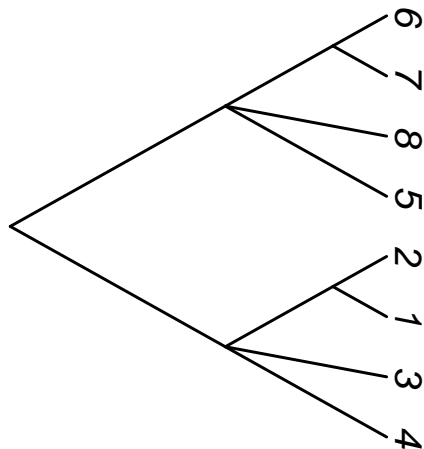
Remove variants likely from alignment errors:

- at sites with excessive depth (>500).
- with excessive levels of heterozygosity.
- within 50 bases of an indel.
- in repeat regions

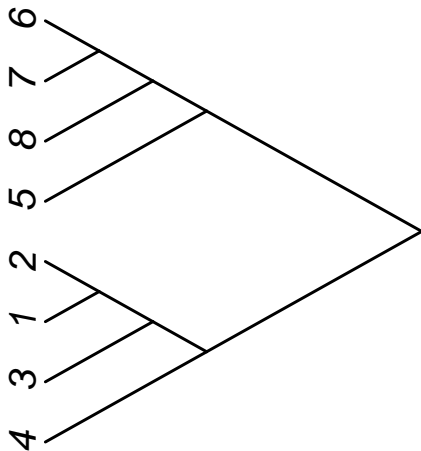


Removing Variants in Repeat Regions Improves Tree Topology

Predicted Variants

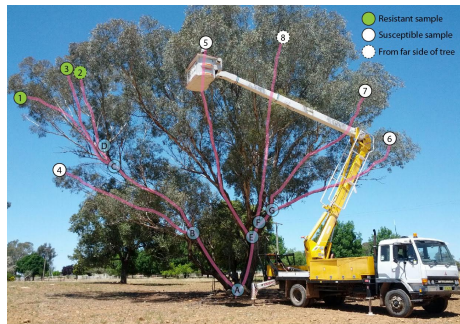


True Tree



Mutation Rates


- Detected ~ 143 mutations.
- Estimated true-positive rate of $\sim 22\%$.
- $143 \times \frac{1}{.22} = 640$ mutations.
- ~ 7.1 mutations per meter
- Somatic mutations account for 116 mutations in gametes.
- 2.34×10^{-7} mutations per site per generation



Acknowledgements

- Advisor: Reed Cartwright  @MinionLab
- Robert Lanfear, Australian National University  @RobLanfear

Pipeline:  <https://github.com/adamjorr/somatic-variation>

Talk:  <https://github.com/adamjorr/talks>



This work is supported by grants NIH R01-HG007178 and NSF DBI-1356548.