

# Patterns of somatic mutation correspond to physical topology in *Eucalyptus*

Adam J Orr<sup>1,2</sup> Robert Lanfear<sup>3</sup> Reed Cartwright<sup>1,2</sup>

<sup>1</sup>School of Life Sciences, Arizona State University

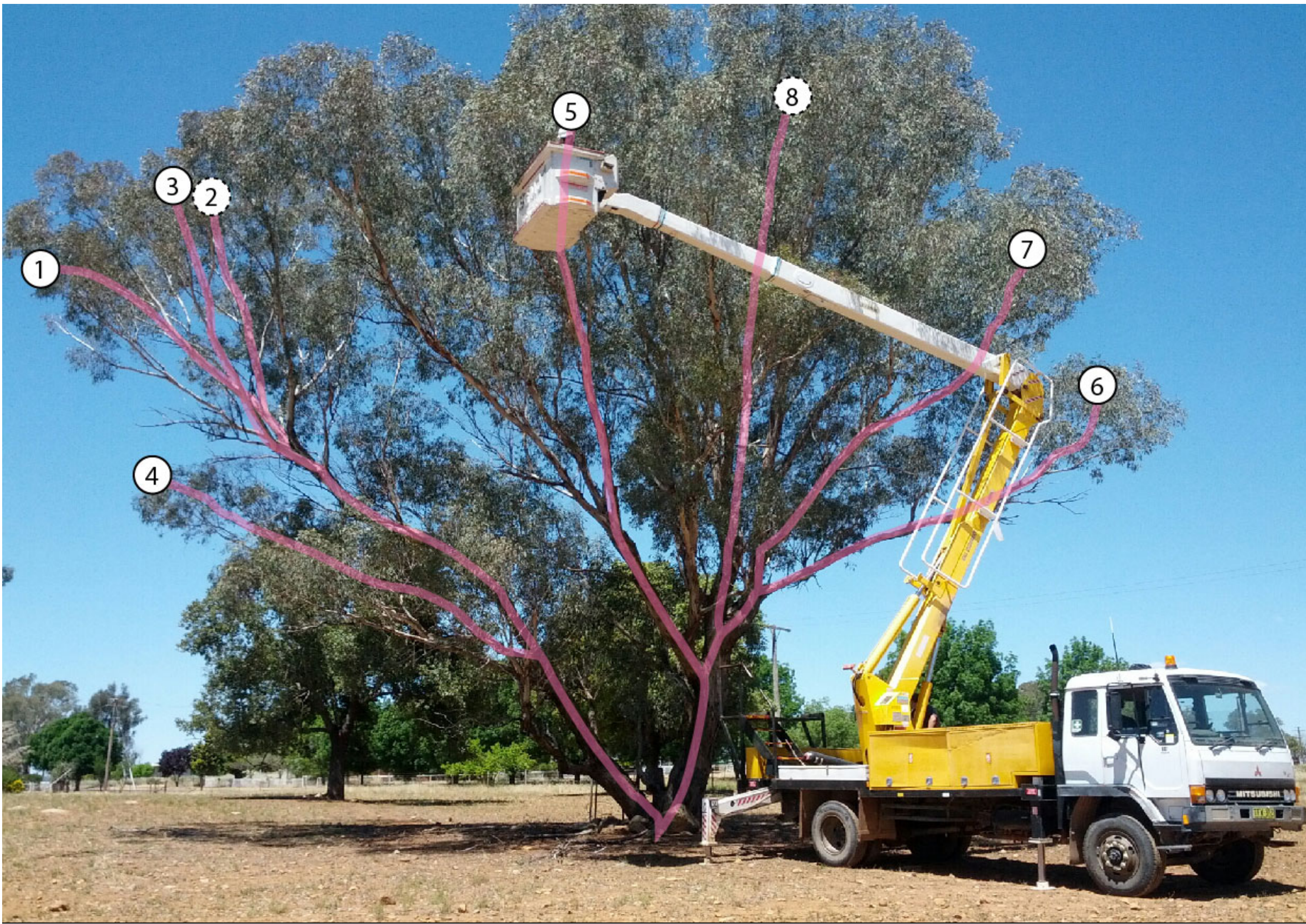
<sup>2</sup>Biodesign Institute, Arizona State University

<sup>3</sup>College of Medicine, Biology and Environment, Australian National University

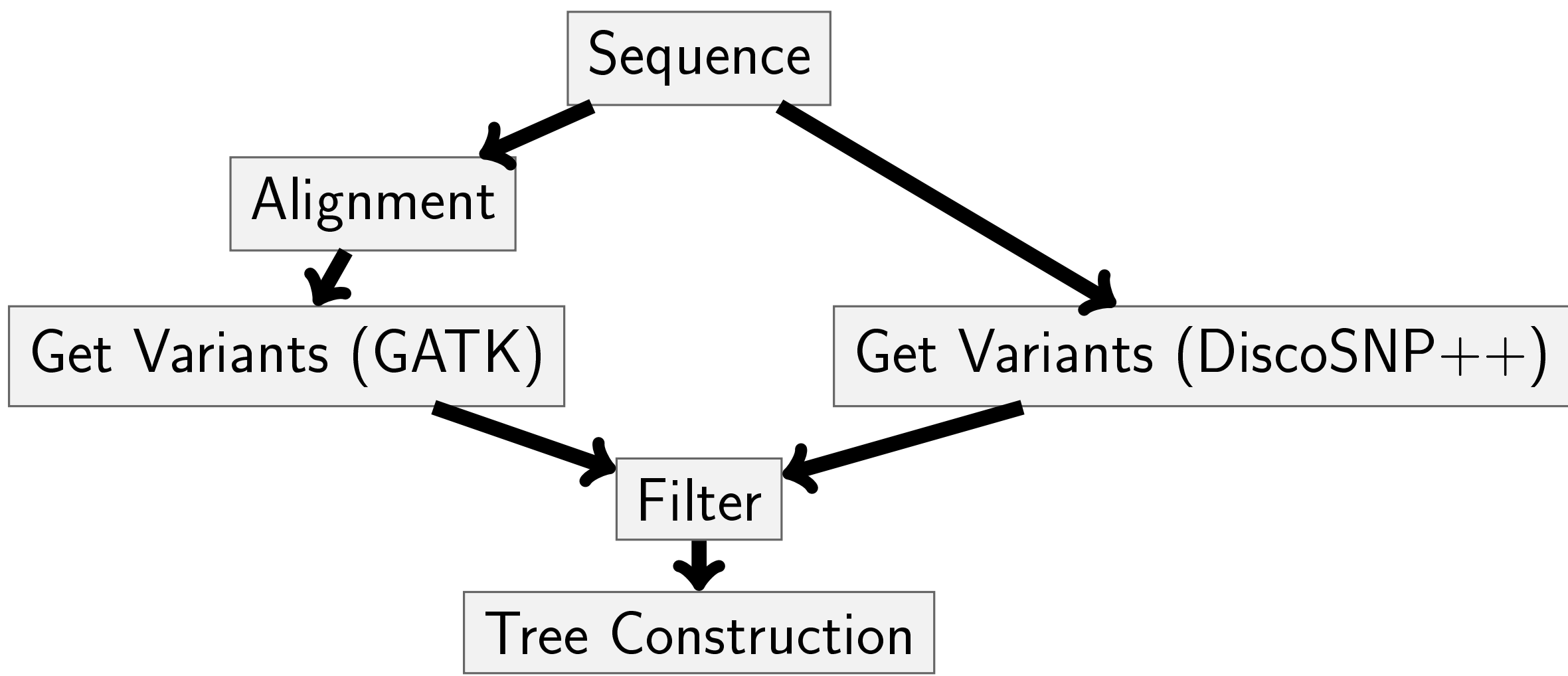
### Introduction

- ▶ 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.
- ▶ The branching pattern of trees can be used to validate methods for detecting somatic mutations.

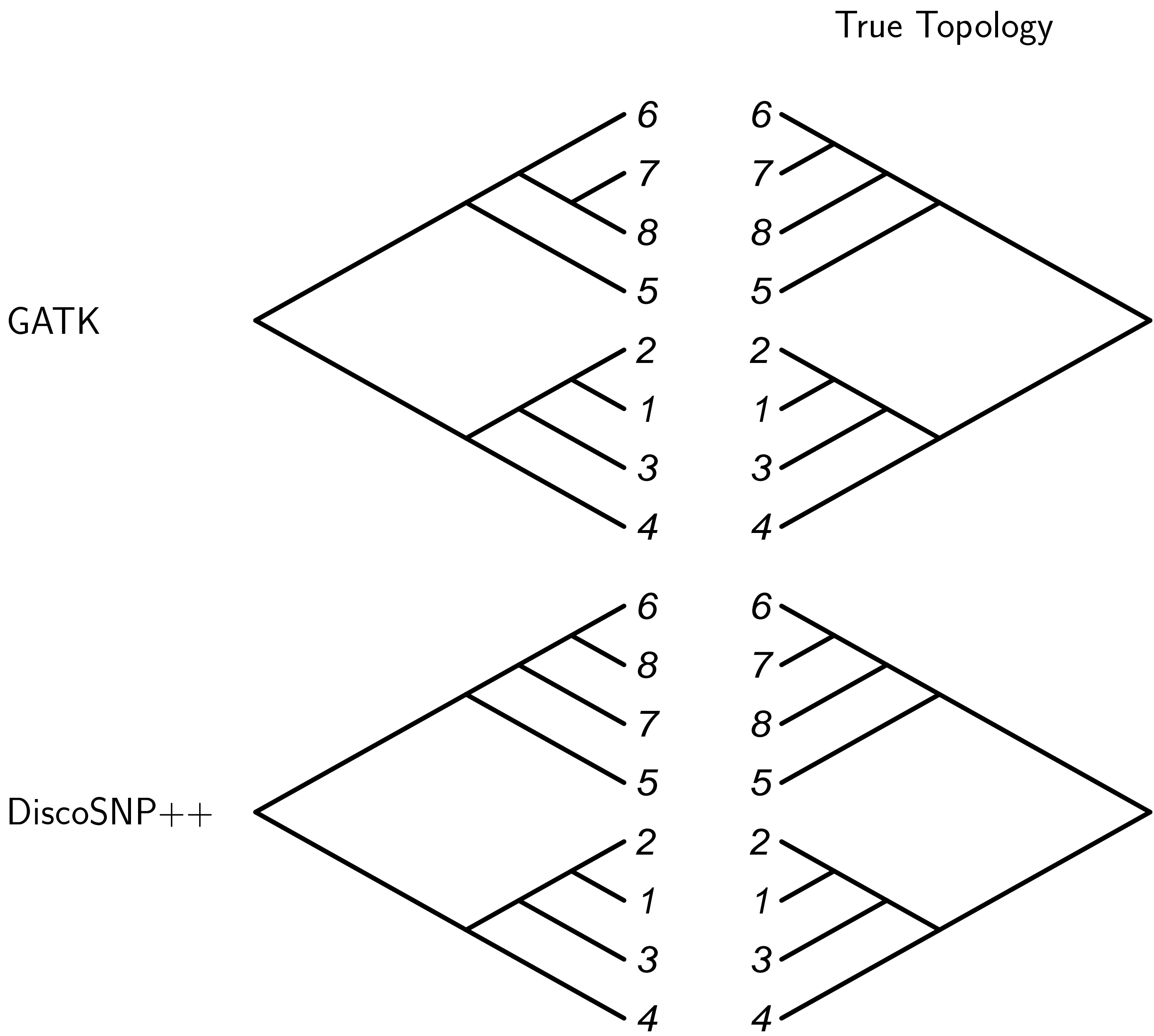
### Methods: Variant Detection



- ▶ 8 samples collected in triplicate
- ▶ Variants were removed if the genotypes of all replicates of a sample were not identical

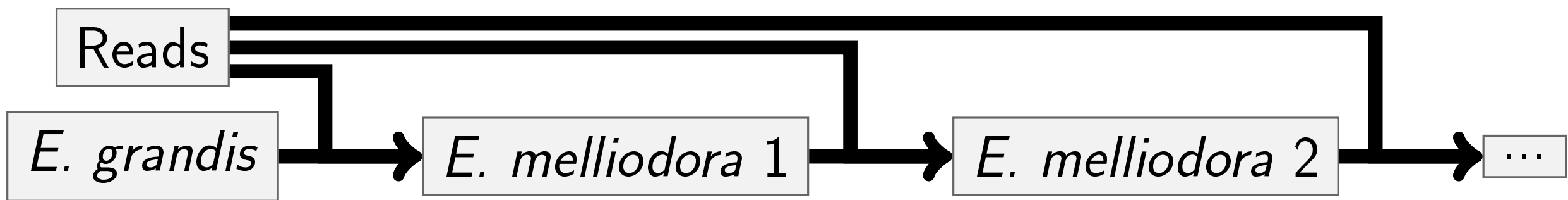


### Results: Variant Detection

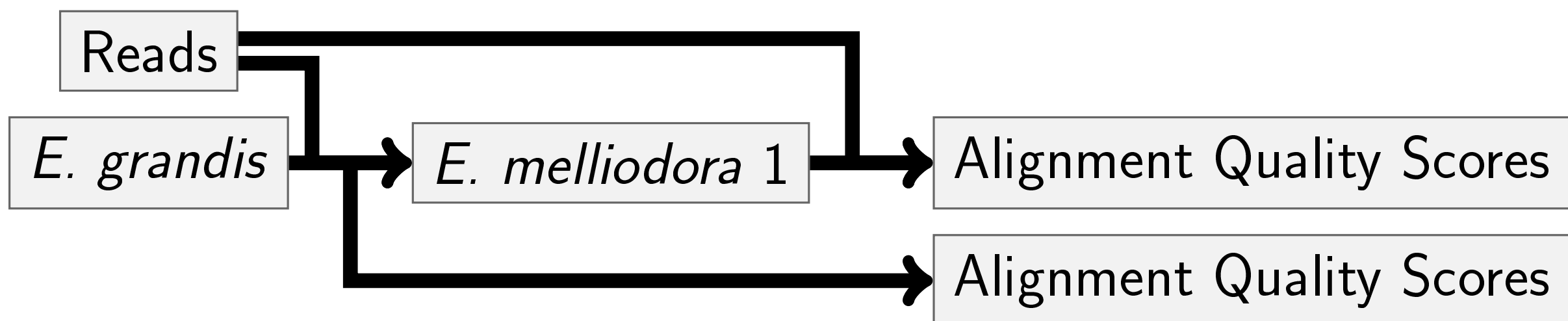


### Next steps: Reference Improvement

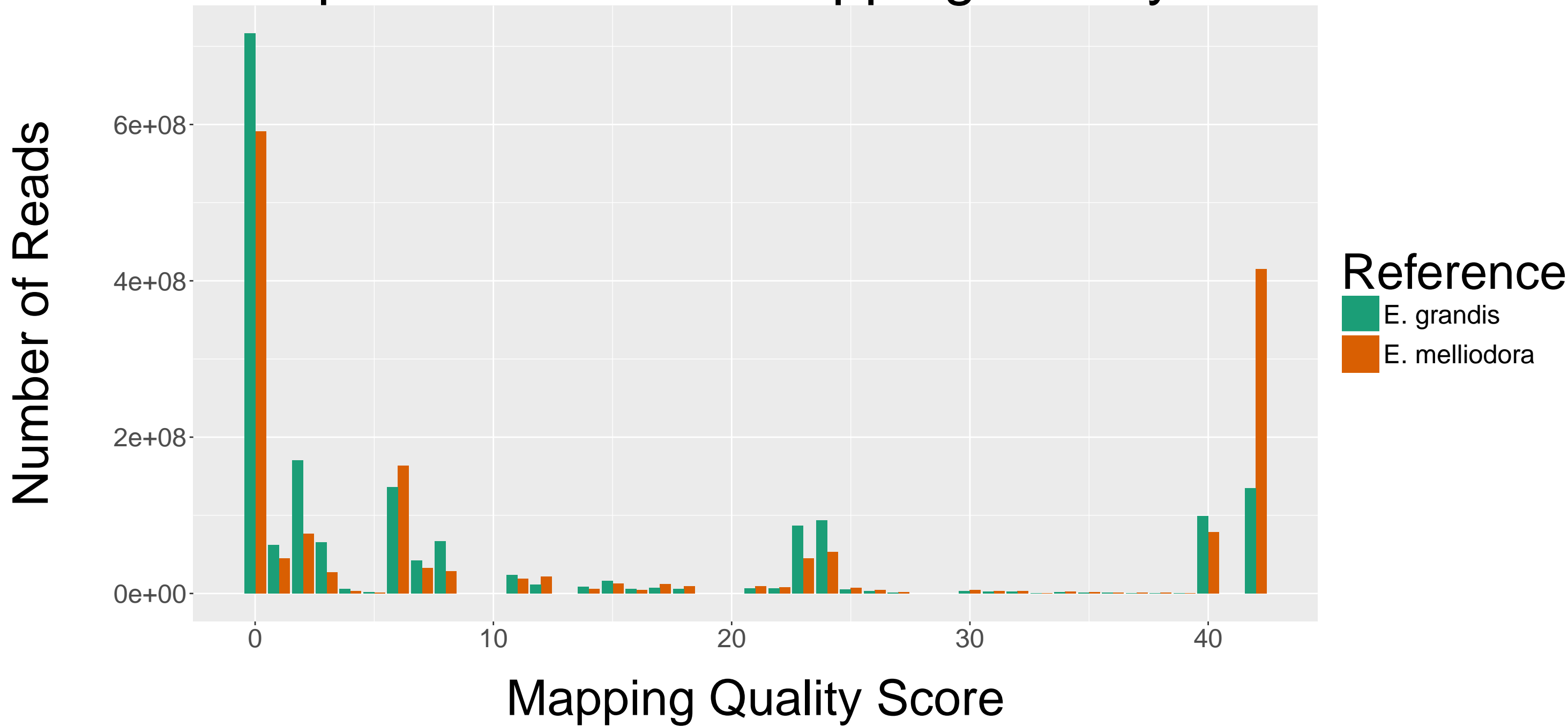
We improve our reference by aligning to the *E. grandis* genome, then creating a consensus from that alignment. This can be repeated until convergence.



### Results: Iterative Mapping Improves Reference

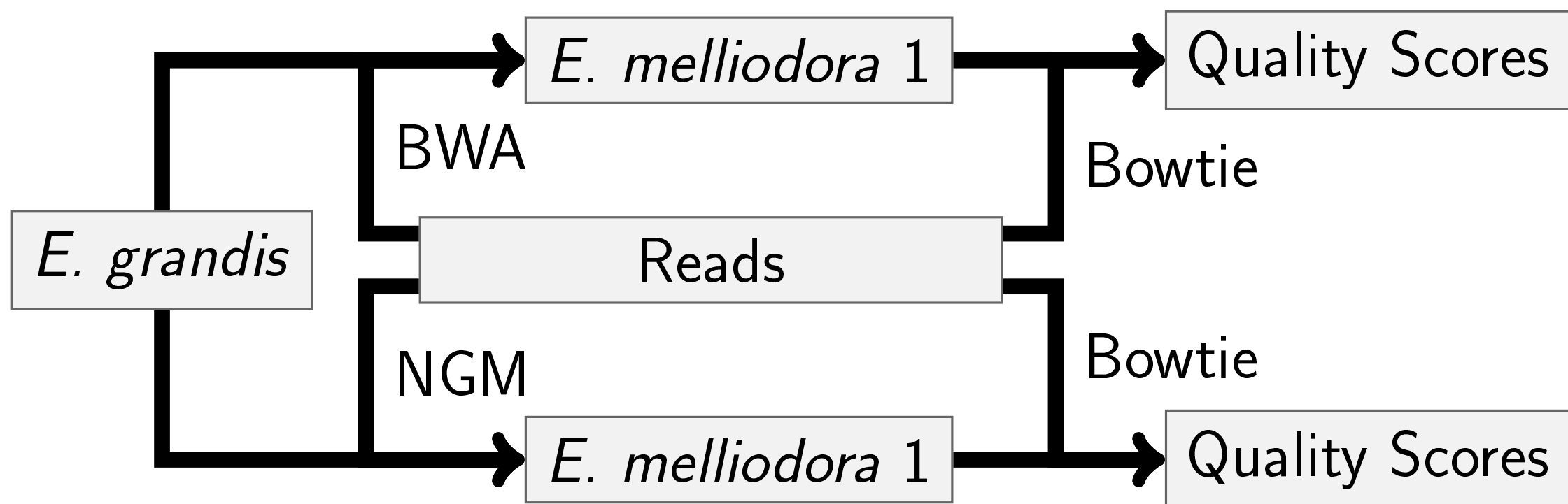


#### Comparison of Read Mapping Quality

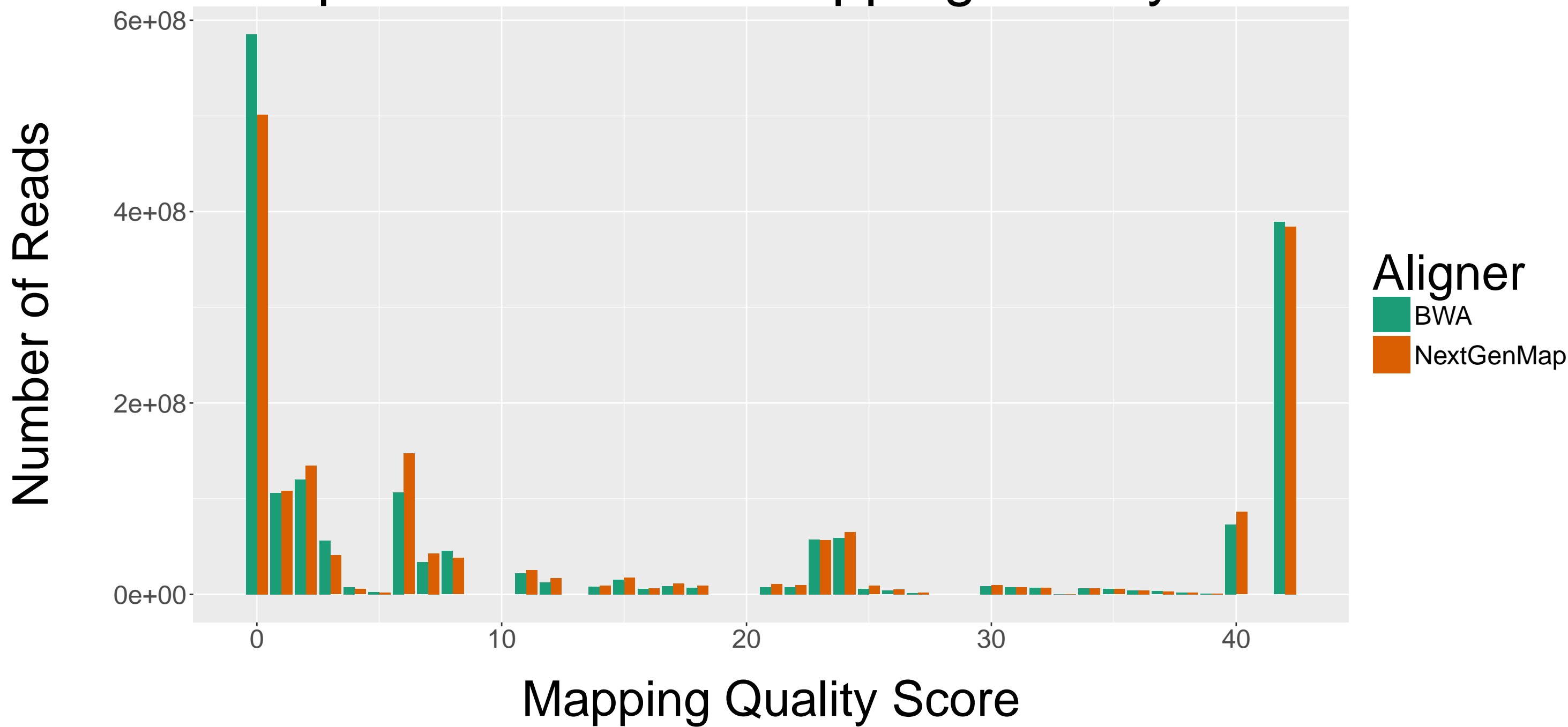


Alignment to the consensus sequence produces an alignment with higher overall quality scores than alignment to the *Eucalyptus grandis* reference.

### Results: Aligner Choice Affects Generated Reference Quality



#### Comparison of Read Mapping Quality



### 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 grants NIH R01-HG007178 and NSF DBI-1356548.