Quaint: An R Package for detecting introgression across a phylogeny using discordant gene tree topologies.

Ethan Baldwin
Department of Plant Biology, University of Georgia

Background

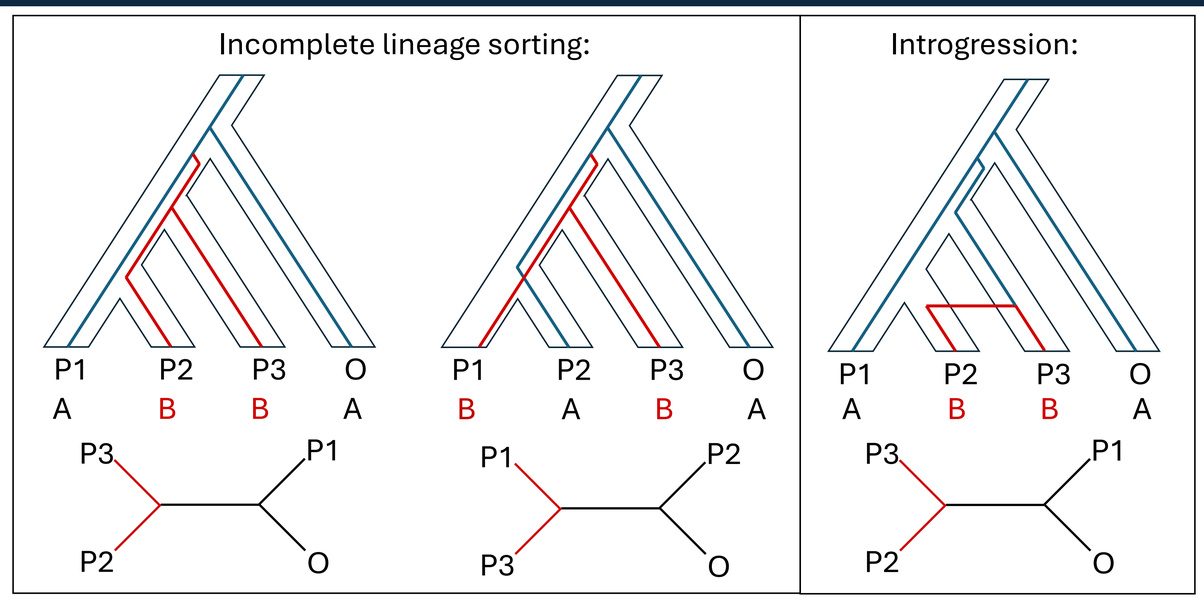
- While many methods for detecting introgression using allele frequencies and/or nucleotide site patterns, there are few methods that utilize gene tree topologies to infer introgression.
- The use of gene trees allows for introgression to be assessed between species that where SNP calling is infeasible due to divergence or the lack of a reference genome.
- Quaint is a **qu**artet-based **a**nalysis of **int**rogression implemented in R. It takes as input a species tree and a set of gene trees.
- Using a similar framework as the ABBA-BABA test, Quaint infers introgression by examining imbalances in quartet topology frequencies found in the gene trees.

Quartet-based ABBA-BABA test

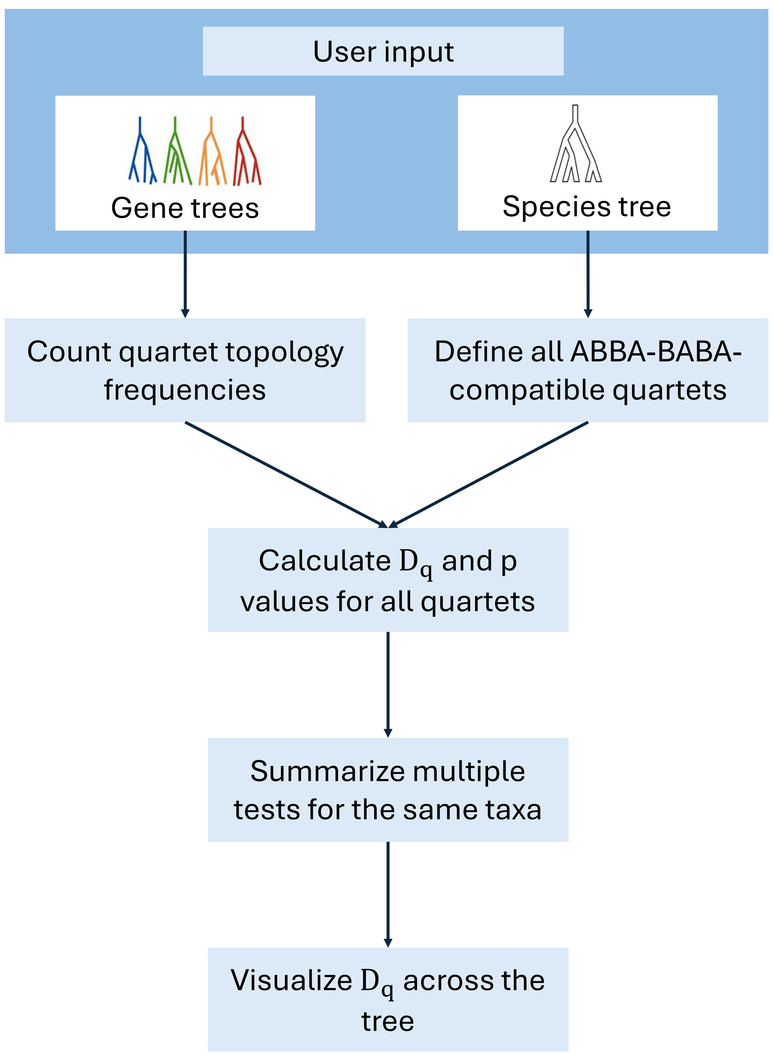
- Incomplete lineage sorting produces ABBA and BABA quartet topologies in equal frequency.
- Gene flow between P2 and P3 produces only ABBA quartet topologies.
- Quaint calculates a D-like statistic ($\mathrm{D_q}$) that quantifies how skewed the ABBA quartet topologies are:

$$D_q(P1, P2, P3, O) = \frac{ABBA - BABA}{ABBA + BABA}$$

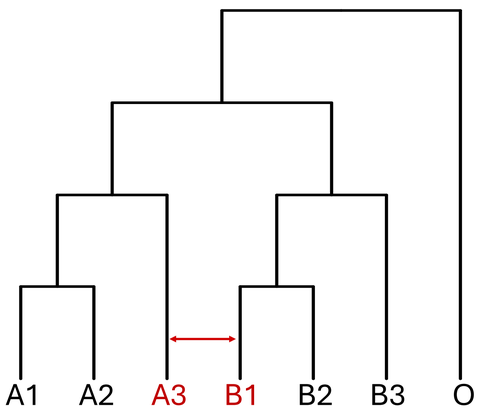
- Quaint uses a chi-square test to test the hypothesis that ABBA > BABA.



Quaint workflow



Summarizing multiple tests



Multiple quartets can be used to test for gene flow between A3 and B1:

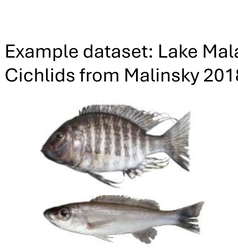
$$Q_{A3,B1} = \left\{ \begin{array}{l} (A2, A3, B1, O) \\ (A1, A3, B1, O) \\ (B2, B1, A3, O) \\ (B3, B1, A3, O) \end{array} \right\}$$

$\mathrm{D_q}$ is summarized across all of these quartets by taking the mean across them:

$$\overline{D_q}(A3, B1) = \mathrm{mean}(D_q(\mathrm{q}) \mid \mathrm{q} \in Q_{A3,B1})$$

Visualization of results

Example dataset: Lake Malawi Cichlids from Malinsky 2018[2].



Availability

github.com/ethan-baldwin/quaint
Contact: ethan.baldwin@uga.edu

1. Green, R. E. et al. A draft sequence of the Neandertal genome. Science 328, 710–722 (2010).
2. Malinsky, M., Svardal, H., Tyers, A.M. et al. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nat Ecol Evol 2, 1940–1955 (2018).