# Guessing RCCX modules from a collapsed pangenome

Jean Monlong

Pangenome with collapsed RCCX region

- A pangenome was built from the HPRC haplotypes using PGGB.
- The global structure of the pangenome looks good
  - The modules were collapsed nicely
  - Haplotypes loop back, traversing the collapsed RCCX part of the pangenome multiple times.
  - The large transposable element polymorphism at the beginning of the RCCX module is visible too.
  - No other major rearrangement.
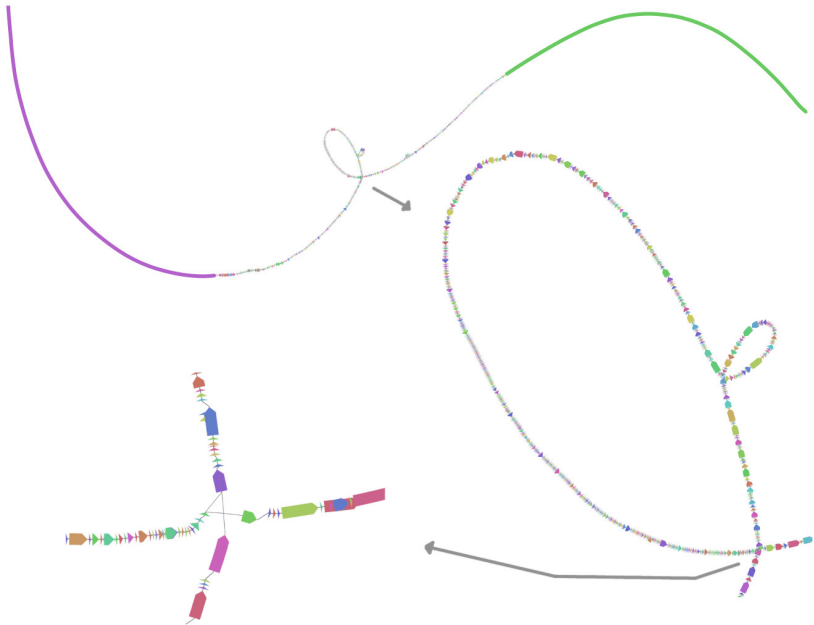
Pangenome with collapsed RCCX region

Can we guess the location of module 1 (with CY21A1P) and module 2 (with CYP21A2) in the haplotypes based on how they traverse the pangenome?

- ▶ Each module 1 should tend to traverse similar nodes across different haplotypes.
- ▶ A PCA-like might analysis might pick up the two types of modules.
- ▶ If we can guess the modules, we won't have to rely on annotation (and filter out haplotypes with suspicious annotations).

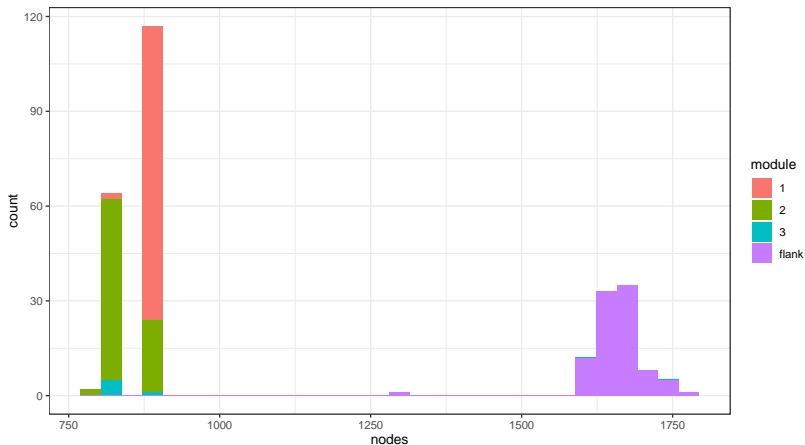# Find the looping edge from the reference path

As the edge where the node IDs differ the most

path	pos	node	strand	next.node	jump
ref	1846	2689	+	1336	1353

Split each haplotype at the looping edges
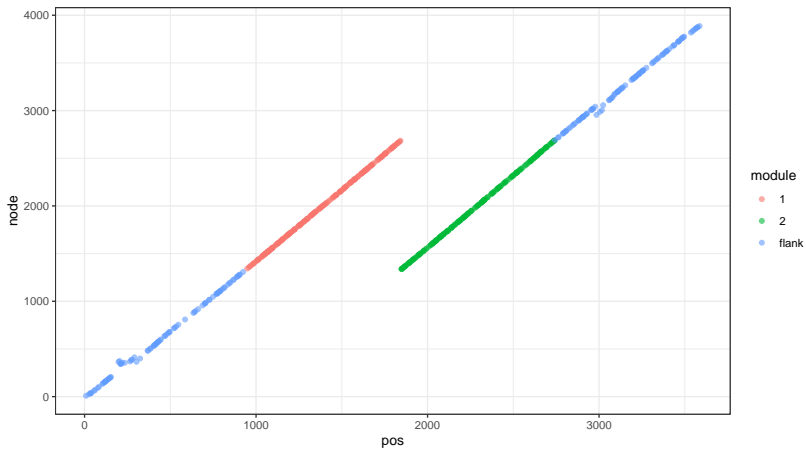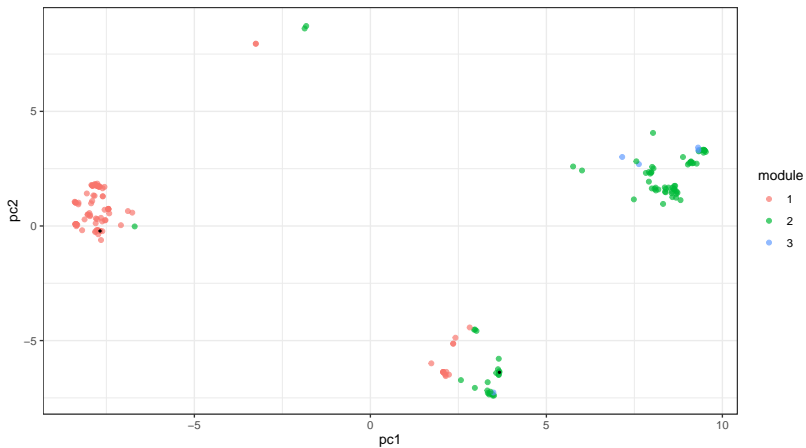
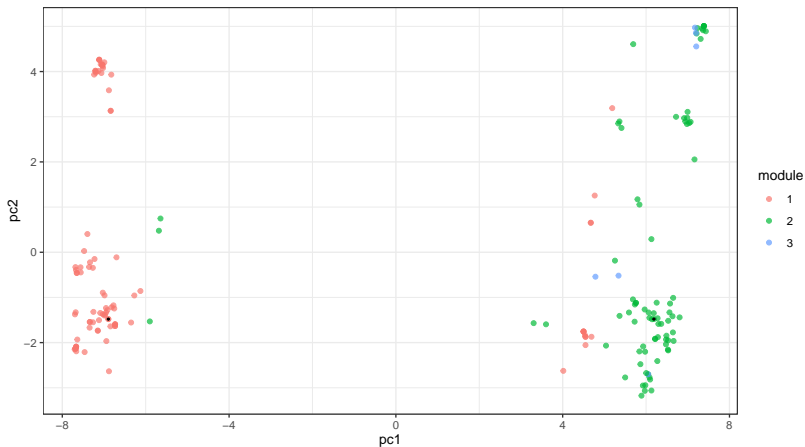# Any small sub-path to remove?

No, they seem to have consistent size.
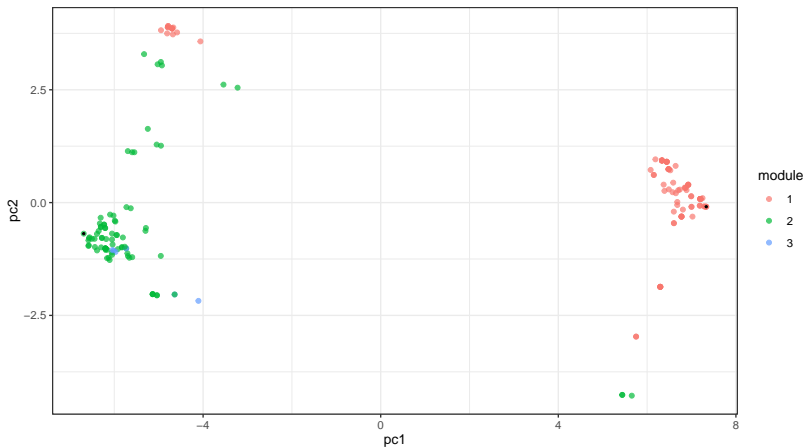
PCA with all the module nodes

Black points highlights modules in the reference haplotype.
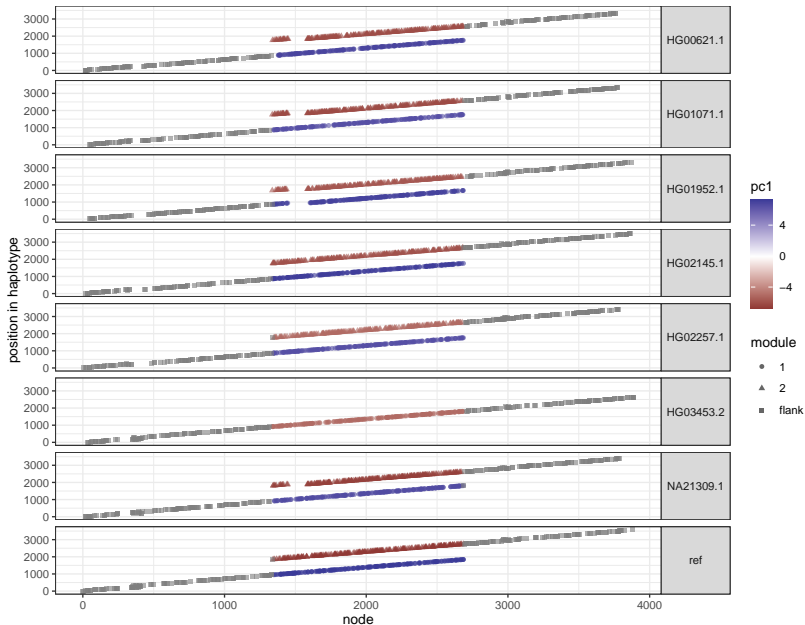
Removing the TE polymorphism

Black points highlights modules in the reference haplotype.
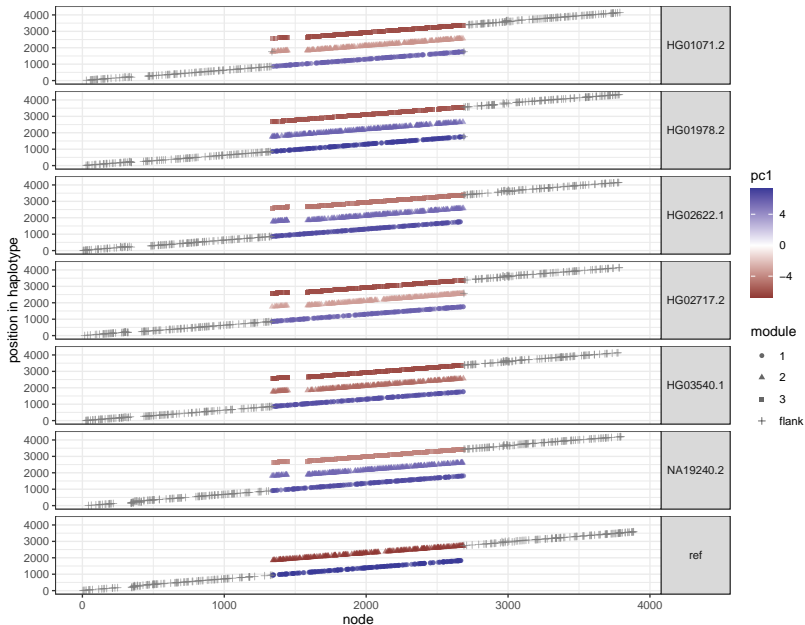
Using module-spec nodes in the reference haplotype

Black points highlights modules in the reference haplotype.

Random haplotypes colored by PC1

Tri-modular haplotypes

Uni-modular haplotypes