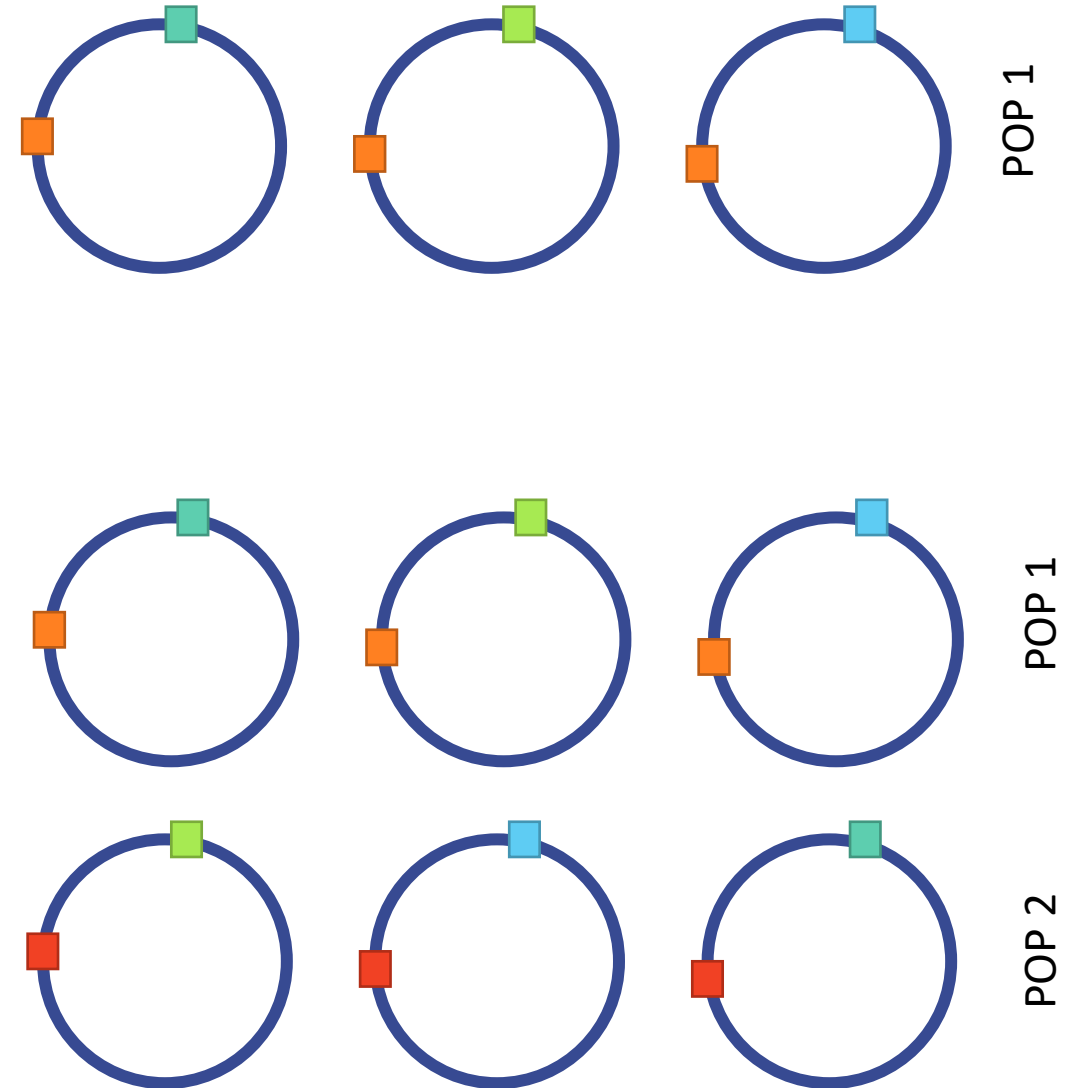# Outlier Analysis Workshop INTRODUCTION

Pilot, 2022, UNSW

| What is outlier analysis?

In outlier locus detection, loci that demonstrate significantly higher or lower within- or among-population genetic differentiation than expected under neutrality are identified as outliers.
~covariates?

| Common outlier analysis programs

- PCAdapt
- localPCA (?)

- Lositan
- Arlequin (DFdist)
- outFLANK
- Pairwise Fst
- Bayescan
- Baypass

- RDA
- Samβada
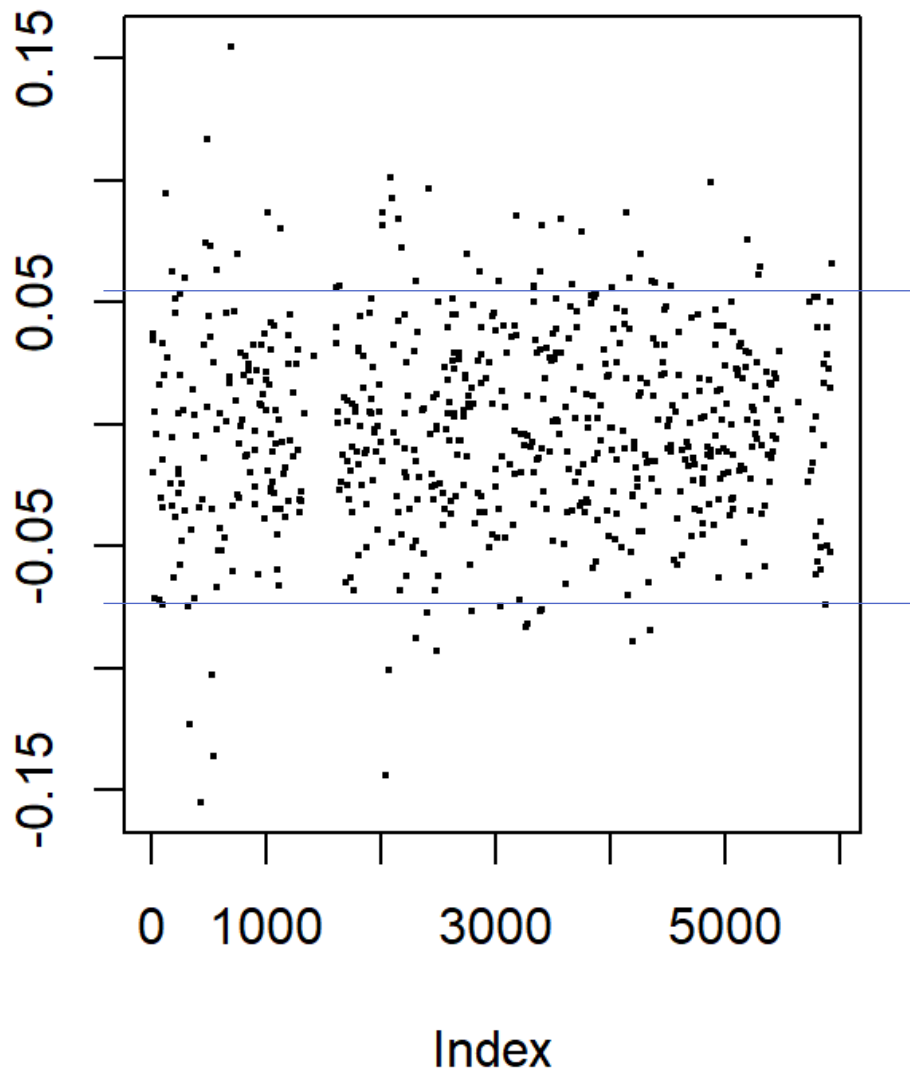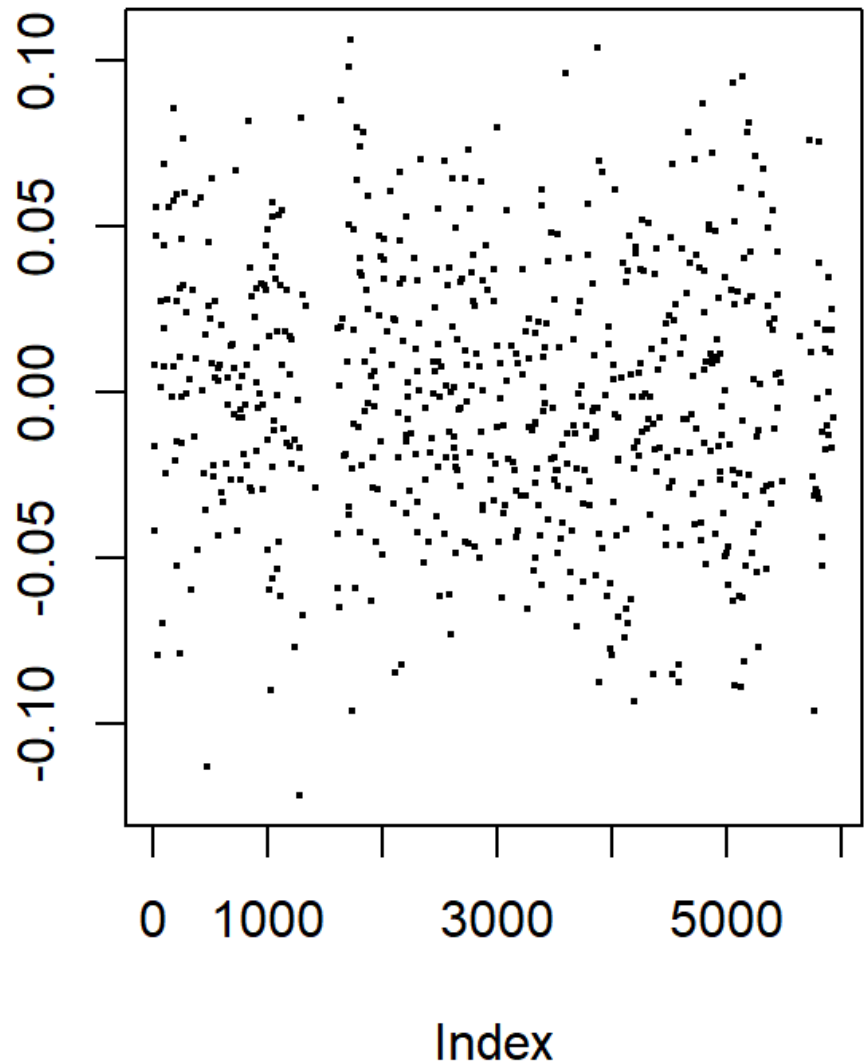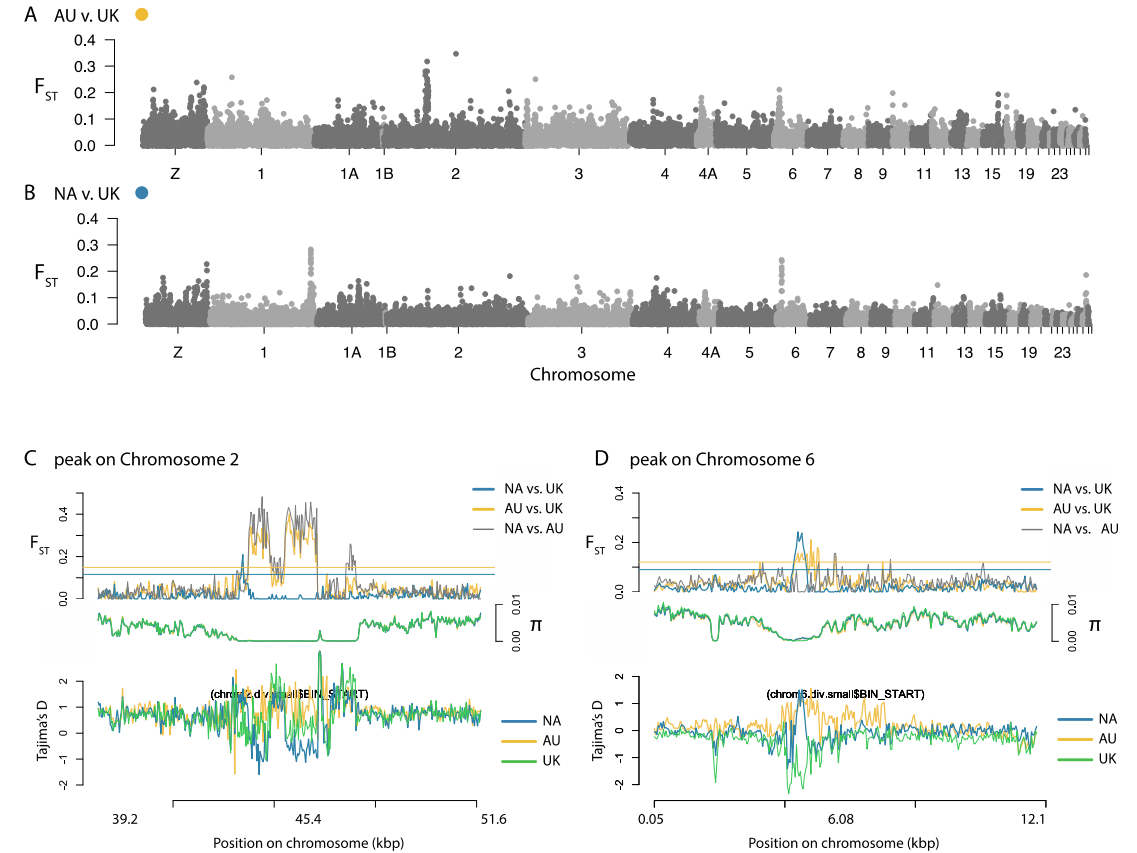- Bayenv
- Baypass

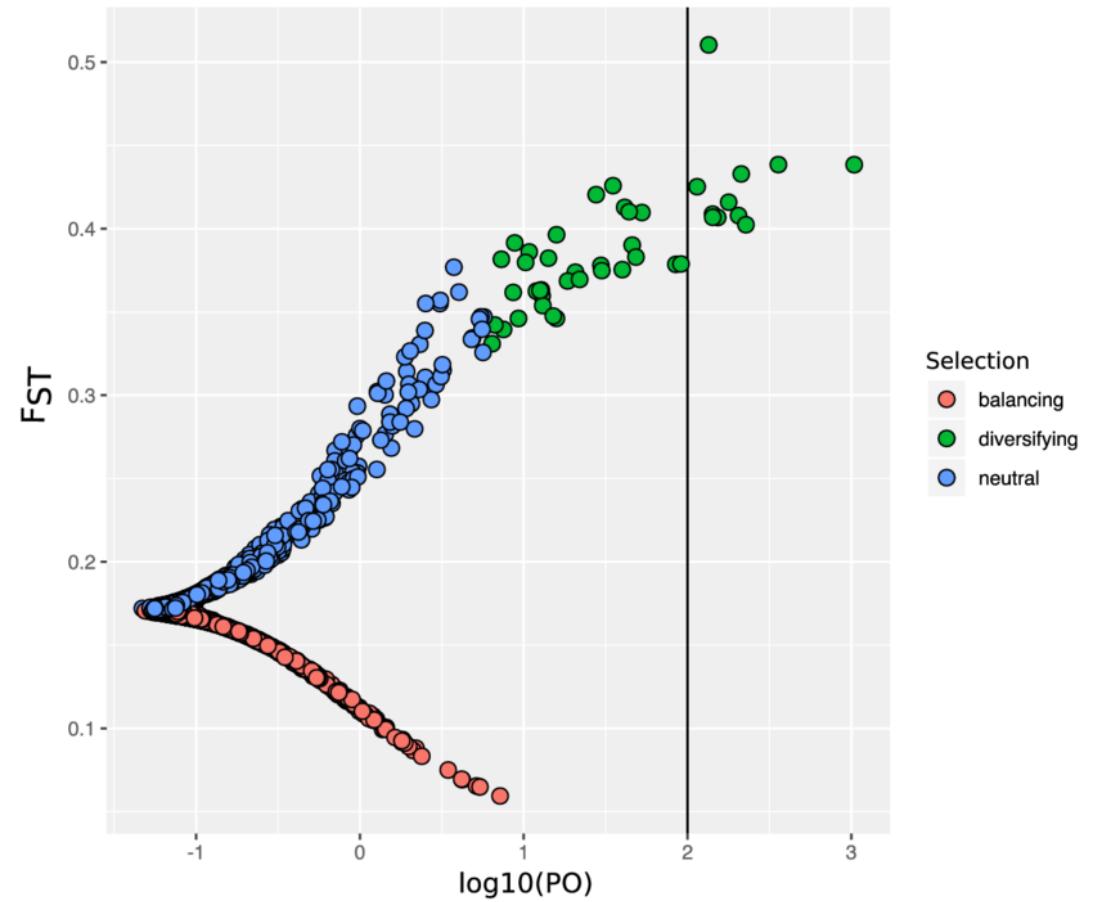# TOOLS

PC

- D
- L
  b
  o

- O

VCFtools Fst

- Requires group designations

- Looks for genetic variants that are statistically divergent between pairwise groupings

- …Fst based

Bayescan

- Requires group designations.

- Looks for genetic variants that are statistically divergent between any number of groupings

- Bayesian

Baypass (bayenv2 updated)

- Requires group designations.

- Looks for genetic variants that are statistically divergent betweenany number of groupings. Accounts for population structure.

- Bayesian

- Can process covariates (e.g. morphology or environmental)
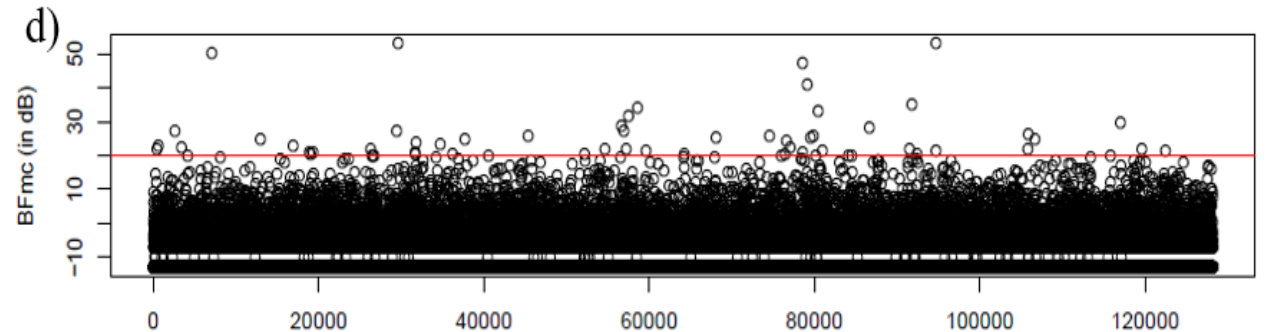
CAVEATS & PITFALLS

Population structure = trouble

Strong pop structure may confound:

- Create false signals

- Hide real ones

References Salloum *et al.* 2020; Stuart *et al.* 2021

Divergence =/= Selection

Allele frequency shifts may be due to:

- Founder effects
- Allele surfing/range edge effects
- Drift
- Non adaptive processes e.g. Spatial sorting

Divergence also =/= adaptation

References Salloum *et al.* 2020; Stuart *et al.* 2021

Selection =/= Novel Range Selection

Divergence between populations is all relative

Determining which is the ancestoral state and which is undergoing 'selection' requires extra data or analysis

References Stuart *et al.* 2022

EXTRA THOUGHTS

Multiple Testing

Using multiple outlier tests is not unusual (max 5)

Combine results by union or overlap union (visualized using Venn or Upset plots)

More interesting methods for combining lists exist, e.g.

- Geometric means (Salloum *et al*. 2020)

- Alternate means (Sherwin *et al.* 2022)

Downstream Analysis

List of outliers is not the final step.

Biologically driven questions around e.g.:

Balancing selection using Betascan (requires WGS)

GO using Gorilla (requires ref annotation or variant annotation)

Protein alteration using VEP (requires variant annotation)

Location using bedtools (requires rough chr lvl assembly)