

Outlier Analysis Workshop

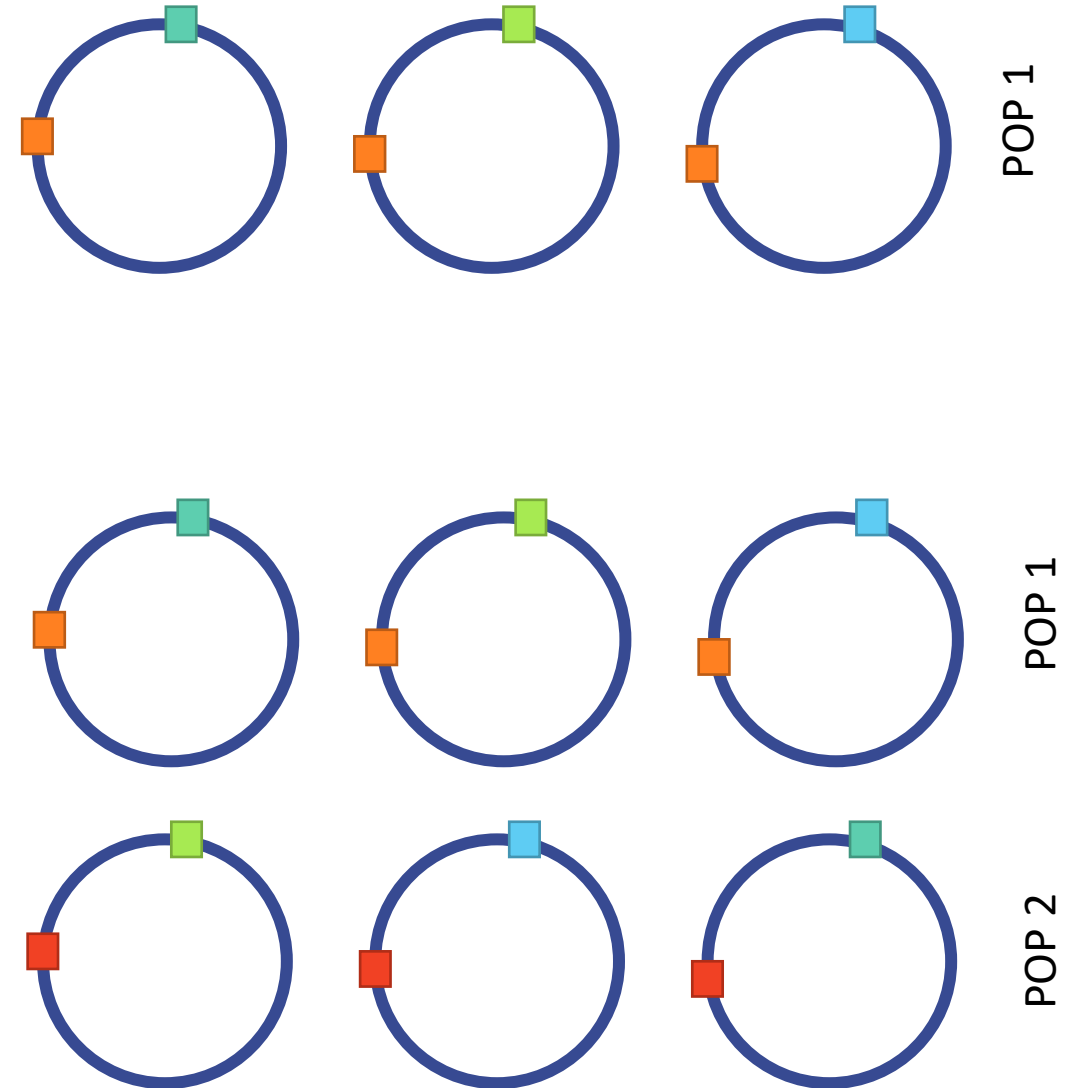
INTRODUCTION

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?



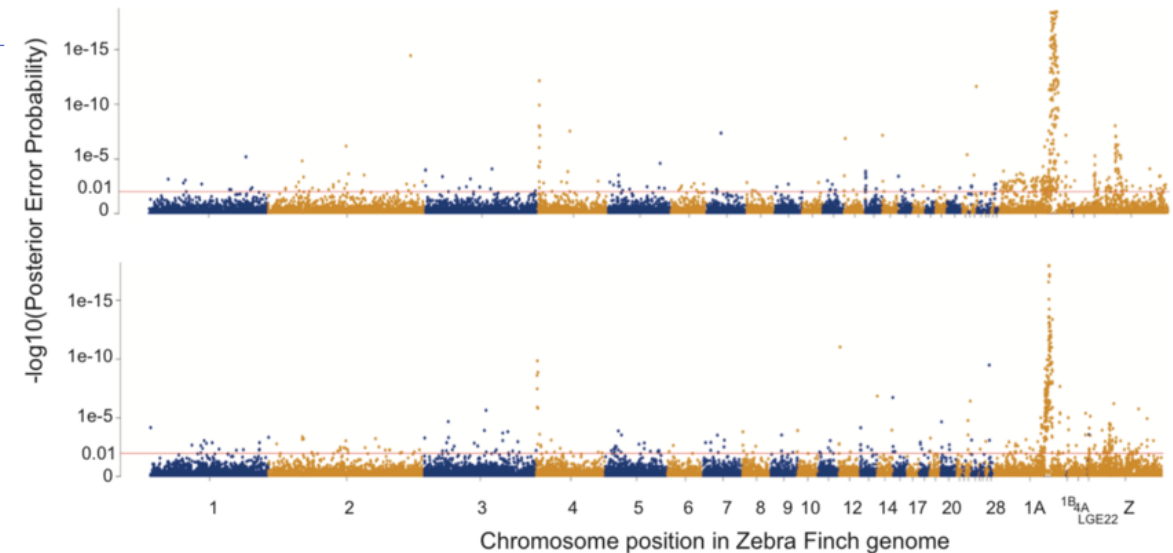
| Common outlier analysis programs

- PCAdapt
- localPCA (?)
- Lositan
- Arlequin (DFdist)
- outFLANK
- Pairwise Fst
- Bayescan
- Baypass
- RDA
- Samβada
- Bayenv
- Baypass

TOOLS

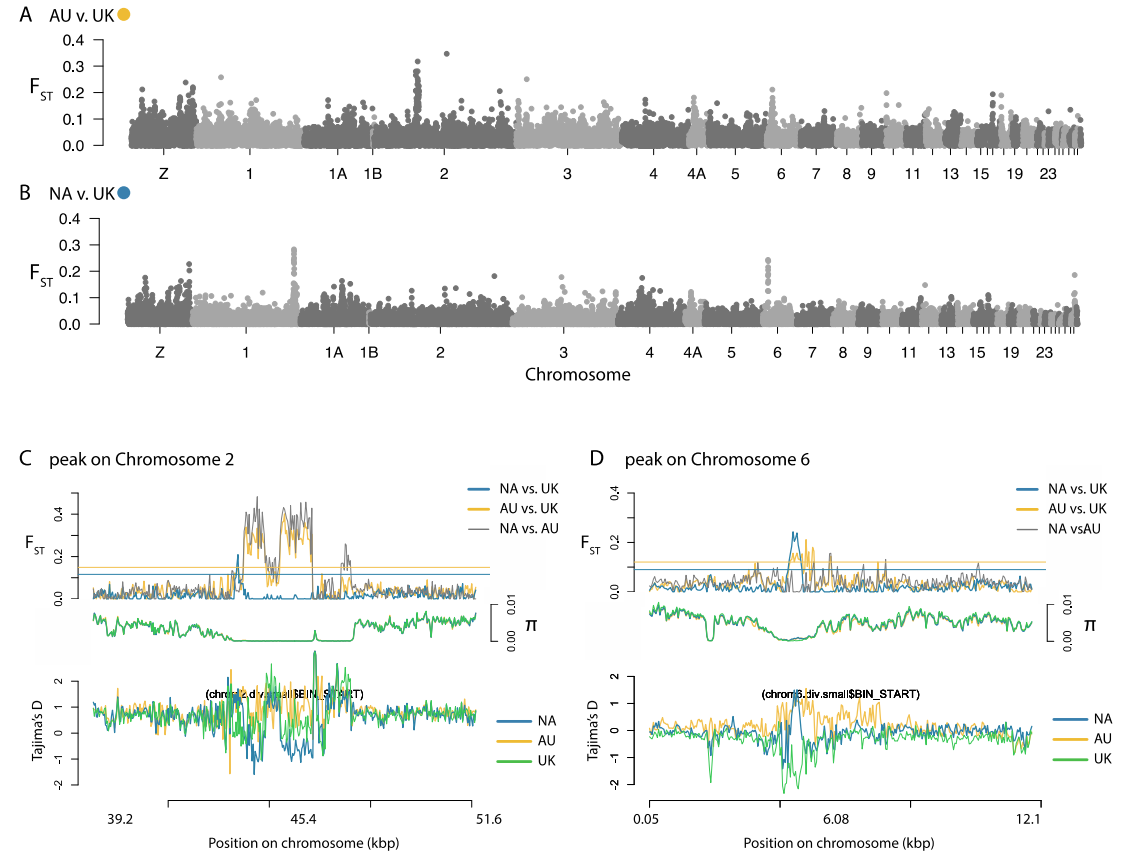
PCAdapt

- Doesn't require group designations
- Looks for genetic variants that behave different relative to the rest of the genome
- Ordination based



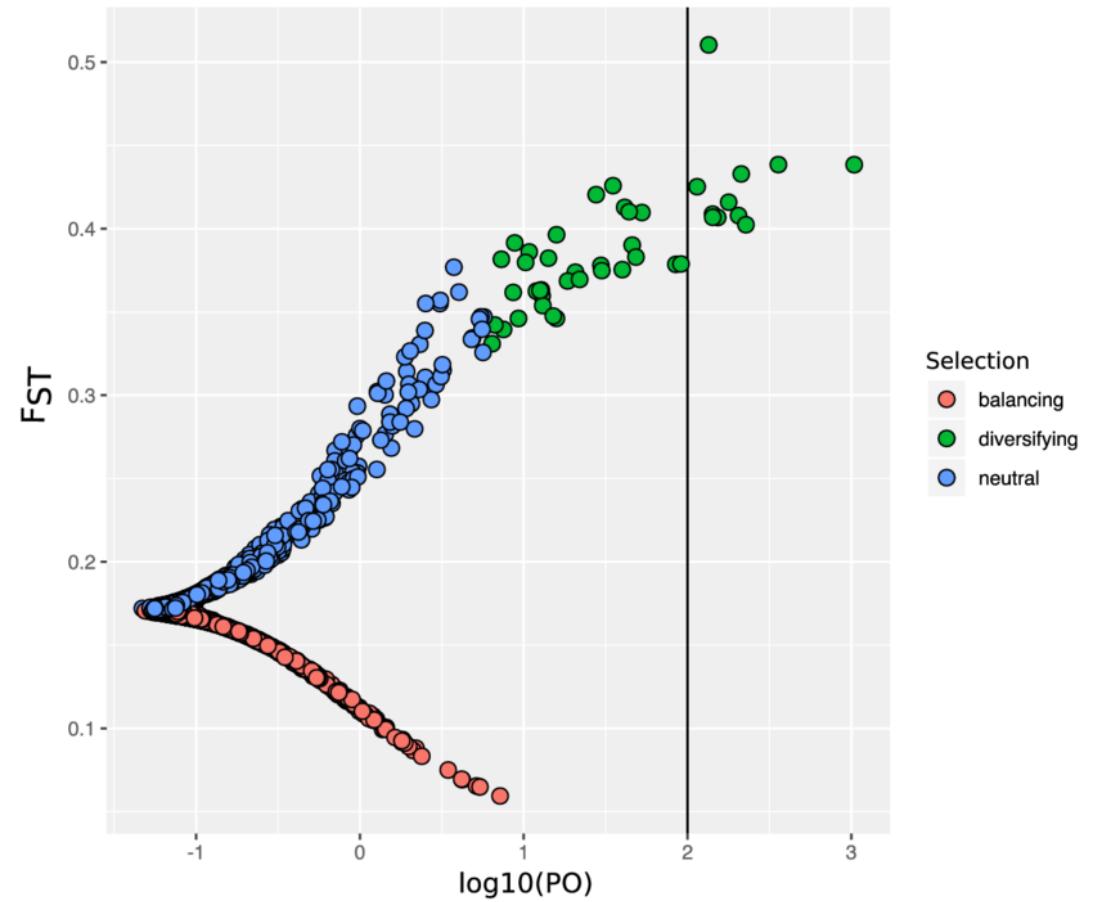
VCFTools Fst

- Requires group designations
- Looks for genetic variants that are statistically divergent between pairwise groupings
- ...Fst based



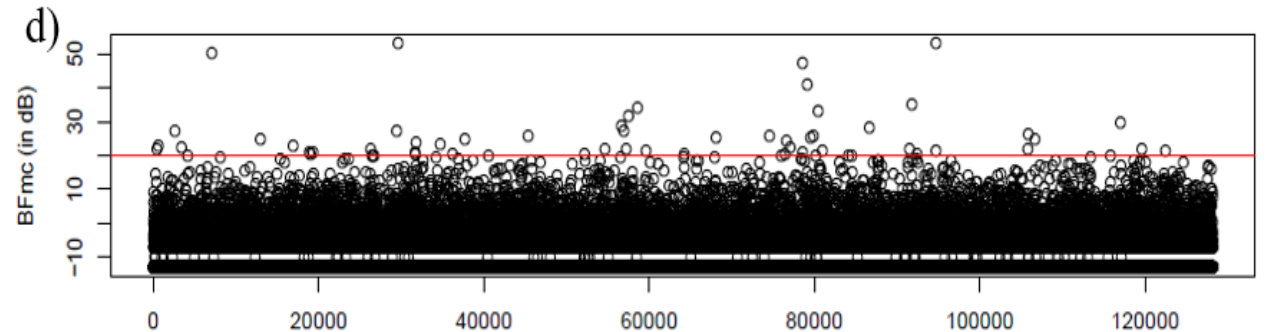
Bayescan

- 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 between any 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

Salloum, Priscila M., et al. "Finding the adaptive needles in a population-structured haystack: A case study in a New Zealand mollusc." *Journal of Animal Ecology* 91.6 (2022): 1209-1221.

Stuart, Katarina C., et al. "Signatures of selection in a recent invasion reveal adaptive divergence in a highly vagile invasive species." *Molecular Ecology* 30.6 (2021): 1419-1434.

Divergence \neq 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 \neq adaptation

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

Selection \neq Novel Range Selection

Divergence between populations is all relative

Determining which is the ancestral 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)

Sherwin, William B. "Bray-Curtis (AFD) differentiation in molecular ecology: Forecasting, an adjustment (AA), and comparative performance in selection detection." *Ecology and Evolution* 12.9 (2022): e9176.

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)

Some important papers

- <https://onlinelibrary.wiley.com/doi/10.1111/mec.14549>
- <https://pubmed.ncbi.nlm.nih.gov/29486732/>
- <https://onlinelibrary.wiley.com/doi/full/10.1002/ece3.9176>