https://github.com/clairemerot/physalia_adaptation_course

| Theme | Software | Purpose | Relevant Dataset |
|---|---|---|---|
| Variant Calling | STACKS | Loci assembly, Preprocess data, variant calling & population genomics | RADseq (short-read) for assembly and variant calling<br><br>Whole-genome data can be used for population-level analysis ($D_{xy}$, $F_{ST}$, $\pi$ etc.) |
| Population structure | Custom R scripts, basic R packages (dplyr, magrittr, tibble, ggplot2, reshape2), vcftools (to convert to geno format) | Detect systematic patterns in measures of mean relatedness between individuals (PCA) | All genomic datasets |
| | fastStructure | Detect population structure and visualize admixed populations using a Bayesian framework | Large genomic datasets (only handles bi-allelic loci) |
| | DAPC (using R package adegenet) | Detect population structure, summarize and visualize variation among groups | All genomic datasets with multiple populations (in different localities for example) |
| Detecting adaptive loci | StAMPP (R Package) | Compute pairwise $F_{ST}$ as a measure of population differentiation | Genomic data from different populations |
| | Various R packages - see Github (Isolation by distance) | To investigate whether differentiation is due to geographic distance | Genomic data from different populations |
| | Plink | LD Pruning (due to inversion or sex) | All population genetics |

| | | | |
|---|---|---|---|
| | OutFLANK | To investigate outliers between populations that drive differentiation | Genomic data from different populations |
| | Baypass | To investigate outliers between populations that drive differentiation alongside environmental associations | Genomic data from different populations and environmental data |
| | RDA | To investigate polygenic sources of adaptation using whole-genome variation alongside alongside geographic or ecological variation | Genomic and environmental (geographic/ecological) data |
| Exploring heterogeneity in the genome (haploblocks) | Lostruct | Sliding PCA | All genomic data |
| | Plink and other scripts (see Github) | Linkage disequilibrium and detecting inversions | |
| | Vcftools, Various R packages | Divergence ($F_{ST}$) and differentiating haplotypes | |
| | Vcftools, Various R packages, windowscanr | Study heterozygosity | |
| Copy Number Variants (CNVs) | Custom R script, edgeR | Detecting CNVs from SNP data | All genomic data |
| | RDA | CNV-environment association | Genomic and environmental (geographic/ecological) data |
| Genome annotation and making biological sense of the adaptive loci | SNPeff | Annotating the genome | De novo genomes |
| | Fisher's exact test | Enrichment of one category of SVs | All genomic data |
| | Bedtools, Goseq | Find intersection between SNPs and genes, GO enrichment | |
| | RepeatModeler2, RepeatMasker, Bedtools | Annotate TEs and find intersection between SNPs and TEs | |