What else can we do with lcWGS data?

- Genotype-phenotype association
- Genotype-environment associations
- Introgression, local ancestry or gene flow analyses
- Selection scans
- Relatedness analyses
- ….

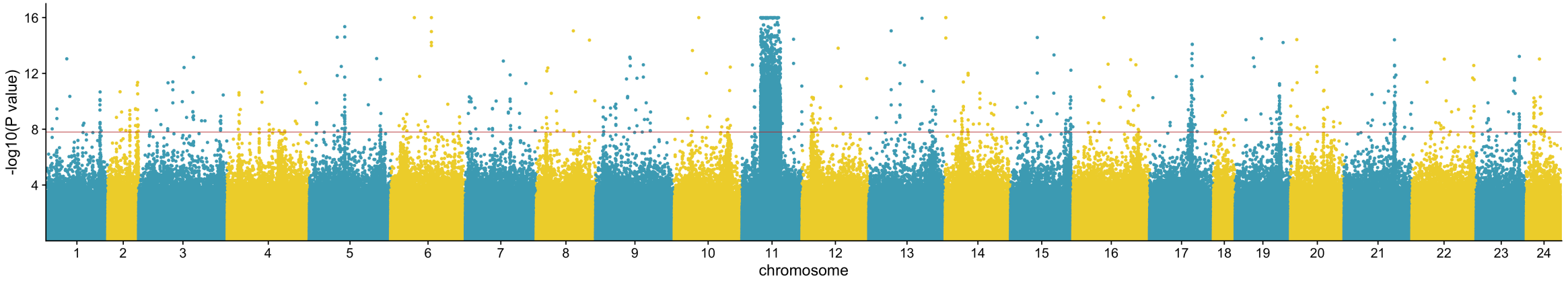
Genotype-phenotype/environment association

GWAS in ANGSD:

```
./angsd -doAsso
abcAsso.cpp:
        -doAsso 0
        1: Frequency Test (Known Major and Minor)
        2: Score Test
        4: Latent genotype model
        5: Score Test with latent genotype model - hybrid test
        6: Dosage regression
        7: Latent genotype model (wald test) - NOT PROPERLY TESTED YET!
  Frequency Test Options:
        -yBin           (null)  (File containing disease status)

  Score, Latent, Hybrid and Dosage Test Options:
        -yBin           (null)  (File containing disease status)
        -yCount         (null)  (File containing count phenotypes)
        -yQuant         (null)  (File containing phenotypes)
        -cov            (null)  (File containing additional covariates)
        -sampleFile             (null)  (.sample File containing phenotypes and covariates)
        -whichPhe       (null)  Select which phenotypes to analyse, write phenos comma seperated ('phe1,phe2,...'), only works with a .sample
file
        -whichCov       (null)  Select which covariates to include, write covs comma seperated ('cov1,cov2,...'), only works with a .sample
file
        -model  1
        1: Additive/Log-Additive (Default)
        2: Dominant
        3: Recessive

        -minHigh        10      (Require atleast minHigh number of high credible genotypes)
        -minCount       10      (Require this number of minor alleles, estimated from MAF)
        -assoThres      0.000001        Threshold for logistic regression
        -assoIter       100     Number of iterations for logistic regression
        -emThres        0.000100        Threshold for convergence of EM algorithm in doAsso 4 and 5
        -emIter 40      Number of max iterations for EM algorithm in doAsso 4 and 5

        -doPriming      1       Prime EM algorithm with dosage derived coefficients (0: no, 1: yes - default)

        -Pvalue 0       Prints a P-value instead of a likelihood ratio (0: no - default, 1: yes)

  Hybrid Test Options:
        -hybridThres            0.050000        (p-value value threshold for when to perform latent genotype model)
```

Genotype-phenotype/environment association

GWAS in Atlantic silversides for the extent of temperature-dependent sex determination:



Population-level GWAS/GEA in BayPass

Allele counts as input

```
--- file begins here ---
81 19 86 14 2 98 8 92 32 68 23 77
89 11 81 19 9 91 1 99 27 73 27 73
89 11 91 9 0 0 15 85 77 23 80 20

[...97 more lines...]

--- file ends here ---
```

Population-specific allele counts can be estimated from population-specific allele frequencies (.mafs.gz files) produced by angsd

```
#Estimate minor allele count (x) and major allele count (y) for each population
JIGA.maf$x=round((JIGA.maf$knownEM)*(2*(JIGA.maf$nInd)), digits = 0)
JIGA.maf$y=(2*(JIGA.maf$nInd))-JIGA.maf$x
```

Population-level GWAS/GEA in BayPass

Allele counts as input

Pop1 Pop2

```
      file begins here ---
snp1  81 19 86 14 2 98 8 92 32 68 23 77
snp2  89 11 81 19 9 91 1 99 27 73 27 73
snp3  89 11 91 9 0 0 15 85 77 23 80 20

      [...97 more lines...]

      --- file ends here ---
```

Population-level GWAS/GEA in BayPass

Allele counts as input

```
--- file begins here ---
81 19 86 14 2 98 8 92 32 68 23 77
89 11 81 19 9 91 1 99 27 73 27 73
89 11 91 9 0 0 15 85 77 23 80 20

[...97 more lines...]

--- file ends here ---
```

Population-specific allele counts can be estimated from population-specific allele frequencies (.mafs.gz files) produced by angsd

```
#Estimate minor allele count (x) and major allele count (y) for each population
JIGA.maf$x=round((JIGA.maf$knownEM)*(2*(JIGA.maf$nInd)), digits = 0)
JIGA.maf$y=(2*(JIGA.maf$nInd))-JIGA.maf$x
```

Demographic inference

Option 1: Use the 1D and 2D (or 3D …) SFS produced in angsd or winSFS with demographic inference software (e.g. ∂a∂i, GADMA, fastsimcoal2, stairwayplot2)

Important: SFS can be problematic with very low coverage data (< 2x), which might lead to mis-inferences. But new approaches such as winSFS might help with that.
NOTE: Don’t use a minor allele frequency filter!

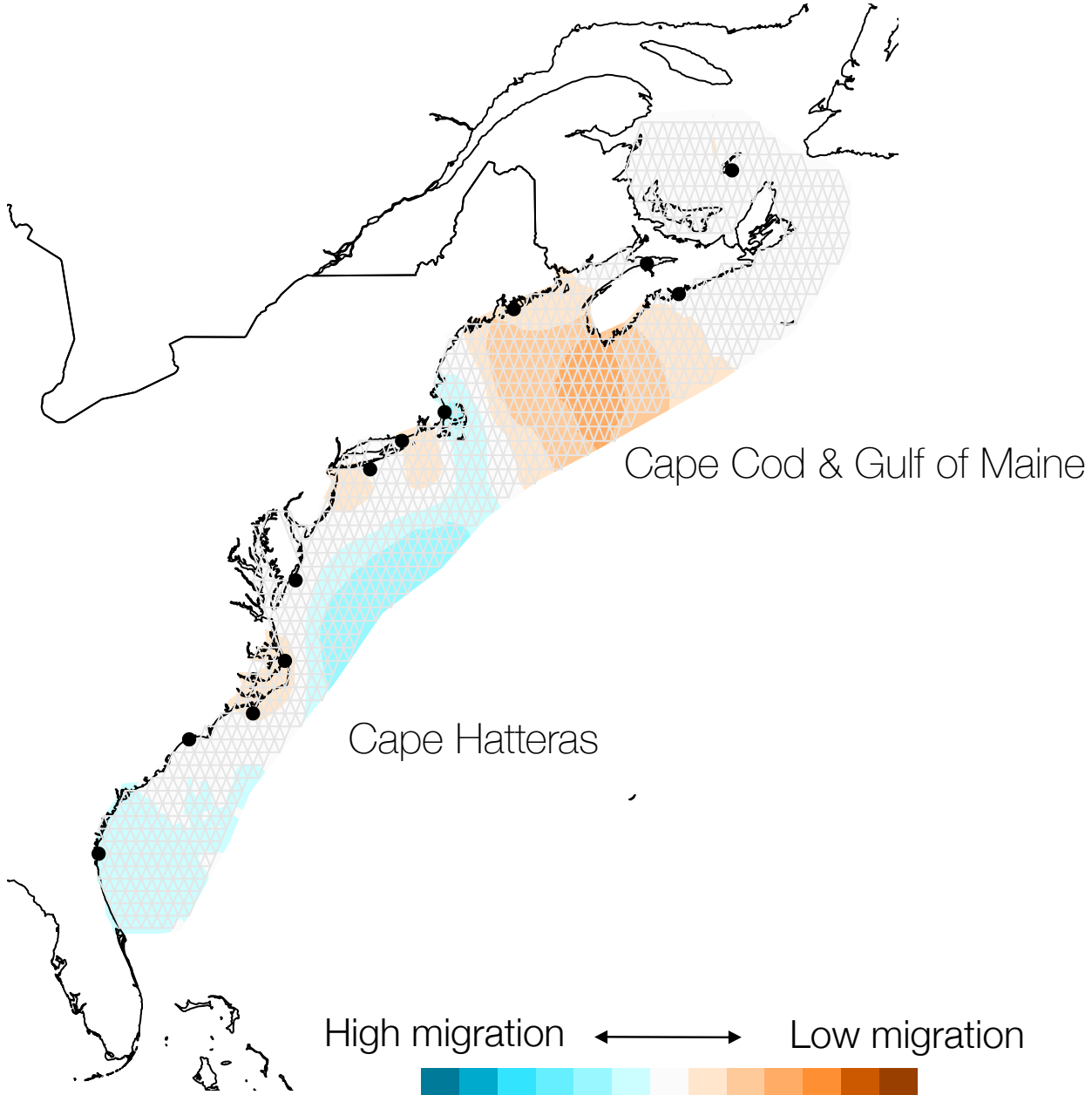
Option 2: Calculate summary statistics from lcWGS data and use them for demographic inference using ABC.

Disclaimer: Have not tested this and one would need to implement custom approaches.

Effective migration surfaces

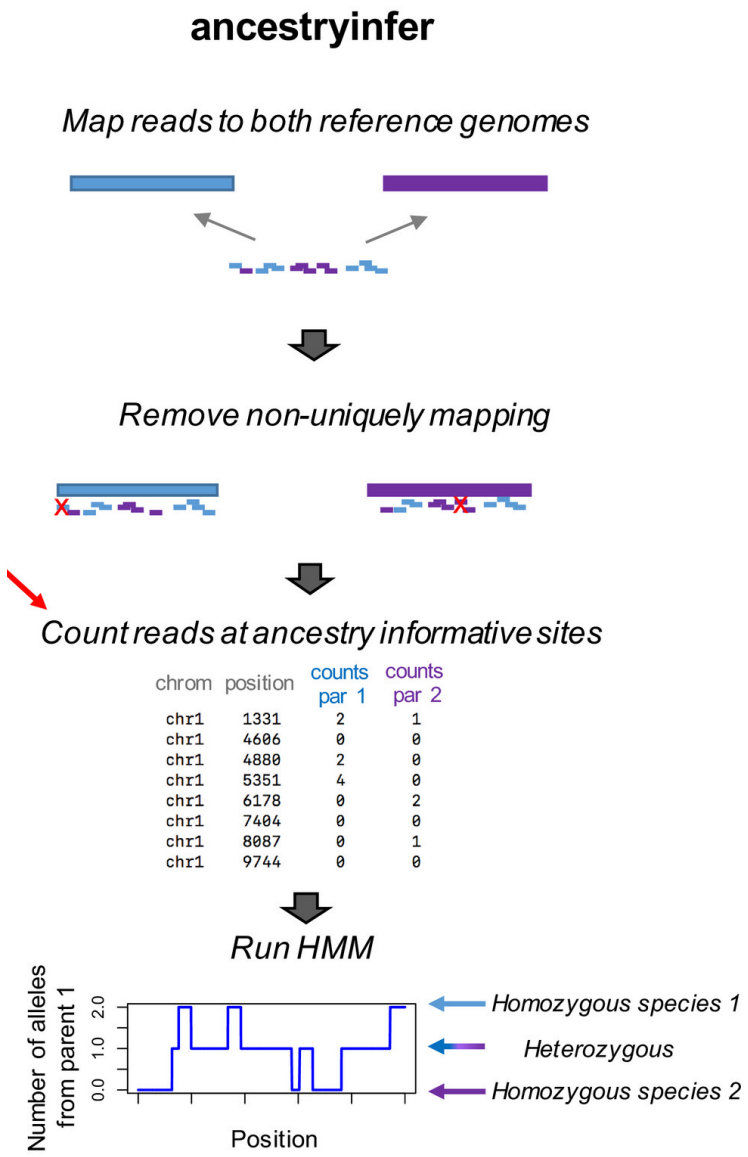
EEMS, Petkova et al. 2016

Input: Pairwise Fst matrix



Local ancestry

ancestryinfer, Schumer et al. 2020 MER



Input: References genomes and fastq files for individuals.

Caveats: One needs ancestry informative SNPs.

Selective sweep analysis

Sweepfinder2

position	x	n	folded
460000	9	100	0
460010	100	100	0
460210	30	78	1
463000	0	94	0
…	…	…	…

x = allele count
n = number of samples
folded = polarized or not? (1 = not polarized)

1. Estimate SFS for the entire genome

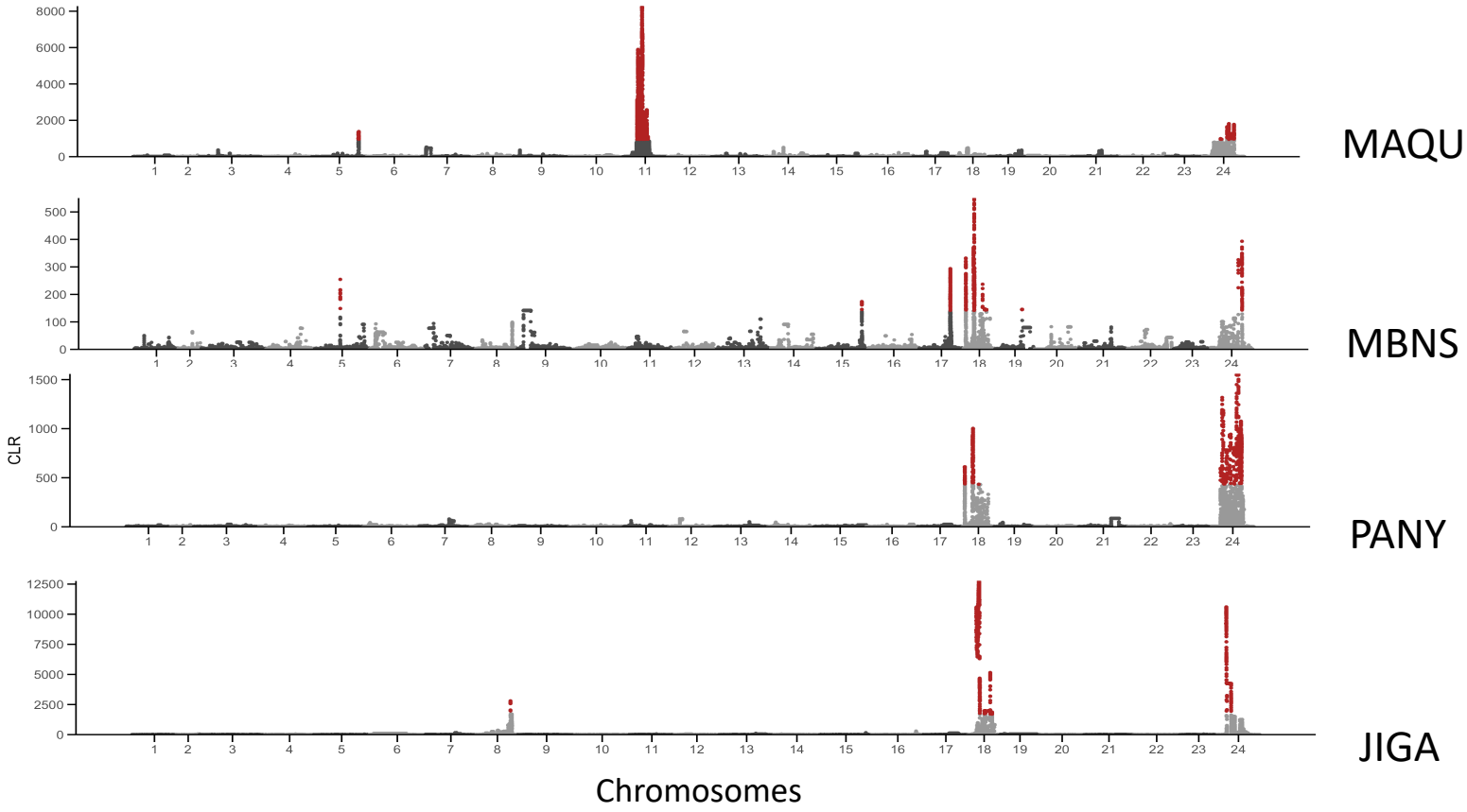
2. Run Sweepfinder2 by chromosome for the entire genome

Selective sweep analysis (by population)

Sweepfinder2

position	x	n	folded
460000	9	100	0
460010	100	100	0
460210	30	78	1
463000	0	94	0
…	…	…	…

x = allele count
n = number of samples
folded = polarized or not? (1 = not polarized)



1. Estimate SFS for the entire genome

2. Run Sweepfinder2 by chromosome for the entire genome

Selective sweep analysis

Ohana

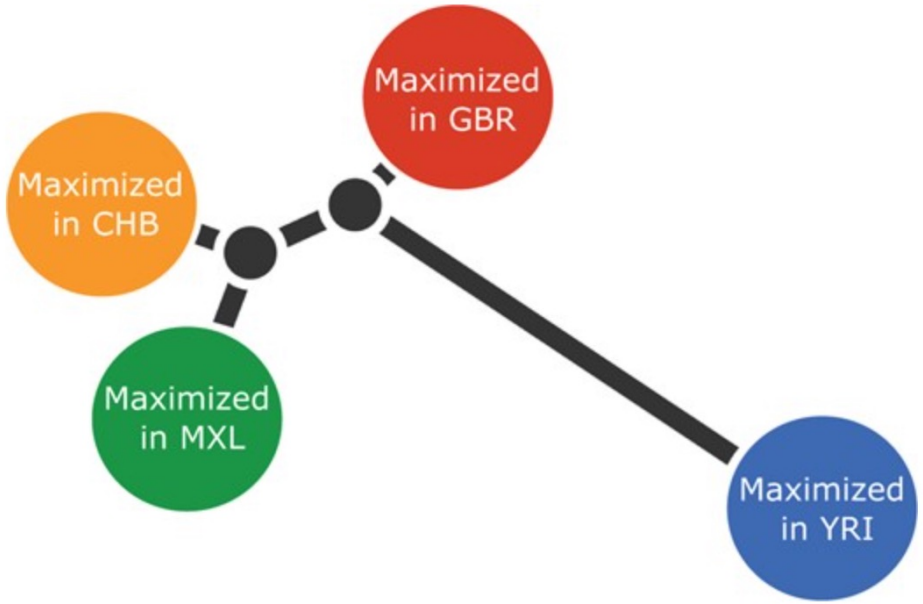
JOURNAL ARTICLE

Detecting Selection in Multiple Populations by Modeling Ancestral Admixture Components

Jade Yu Cheng ✉, Aaron J Stern, Fernando Racimo, Rasmus Nielsen

Molecular Biology and Evolution, Volume 39, Issue 1, January 2022, msab294,
https://doi.org/10.1093/molbev/msab294
Published: 09 October 2021

Fig. 4.

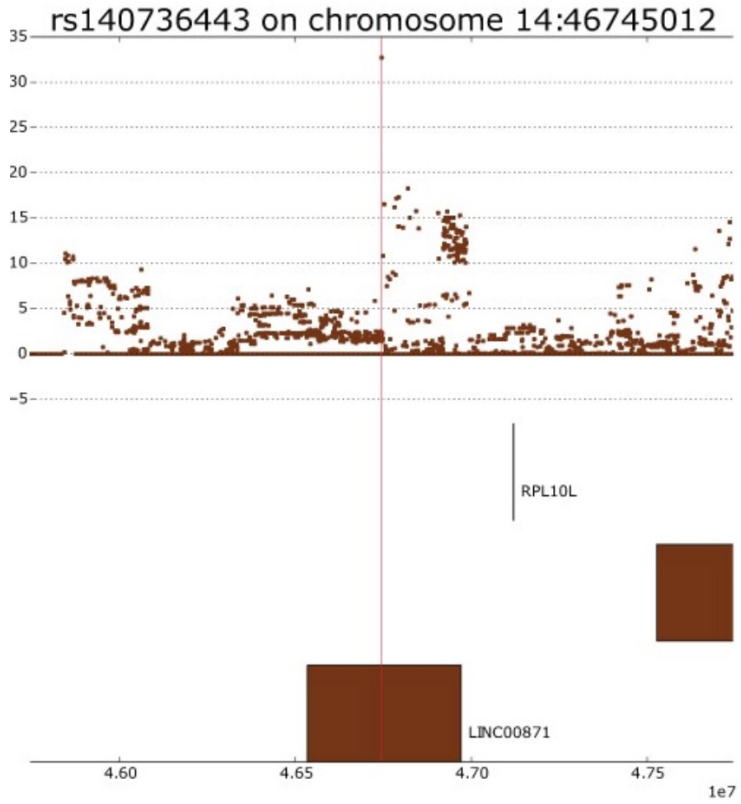


Input: Beagle genotype likelihood file

1. Reconstruct population structure
2. Correct allele frequencies using inferred structure
3. Scan for signatures of selection in specific ancestry clusters

Selective sweep analysis

Ohana



JOURNAL ARTICLE

Detecting Selection in Multiple Populations by Modeling Ancestral Admixture Components

Jade Yu Cheng ✉, Aaron J Stern, Fernando Racimo, Rasmus Nielsen

Molecular Biology and Evolution, Volume 39, Issue 1, January 2022, msab294,
https://doi.org/10.1093/molbev/msab294
Published: 09 October 2021

Input: Beagle genotype likelihood file

1. Reconstruct population structure
2. Correct allele frequencies using inferred structure
3. Scan for signatures of selection in specific ancestry clusters

Other approaches:

Ancestry relationships/Gene flow:

- D-stats/ABBA-BABA using angsd: -doAbbababa
- TreeMix*
- Effective migration surfaces (EEMS)

Relatedness:

- ngsRelate
- LowKi
- NGSremix

Selection:

- PCAngsd (e.g. pcadapt)
- Sweepfinder2*
- Ohana
- BayPass*
- Parallel allele frequency changes (Afvaper)
- Allele frequency differentiation

Linkage Mapping:

- LepMap3

* Based on allele counts

Other approaches:

Ancestry relationships/Gene flow:

- D-stats/ABBA-BABA using angsd: -doAbbababa
- TreeMix*
- Effective migration surfaces (EEMS)

Relatedness:

- ngsRelate
- LowKi
- NGSremix

Selection:

- PCAngsd (e.g. pcadapt)
- Sweepfinder2*
- Ohana
- BayPass*
- Parallel allele frequency changes (Afvaper)
- Allele frequency differentiation

Linkage Mapping:

- LepMap3

Code for running some of the additional analyses:

https://github.com/therkildsen-lab/genomic-data-analysis/tree/master

Other approaches:

Ancestry relationships/Gene flow:
- D-stats/ABBA-BABA using angsd: -doAbbababa
- TreeMix*
- Effective migration surfaces (EEMS)
- Local ancestry inference (ancestryinfer)

Relatedness:
- ngsRelate
- LowKi
- NGSremix

Selection:
- PCAngsd (e.g. pcadapt)
- Sweepfinder2*
- Ohana
- BayPass*
- Parallel allele frequency changes (Afvaper)
- Allele frequency differentiation

Linkage Mapping:
- LepMap3

More detailed information in our paper: