

# Exercises in Marine Ecological Genetics

## 09. Selection and mutation

- Filter SNPs by linkage
- Simulate selection and mutation
- Apply genome-wide scans and identify outlier loci ( $F_{ST}$  / selection)

Martin Helmkamp

# Download course materials using git

Go to project directory

```
cd dir          # e.g. Documents/meg23_exercises  
ls -l           # view directory contents, long format
```

Update course repository

```
cd meg23_repo  
git pull
```



In case of an error message

```
cd ..                # go back to project directory  
rm -rf meg23_repo    # delete old repository  
git clone https://github.com/mhelmkampf/meg23\_repo.git
```

# Avoiding version conflict

Please do not save over files in the course repository. Instead, save your own scripts to the local subdirectory (including copies of course scripts you would like to edit), e.g with

```
cp code/09_sel.R ../local/09_sel_lc.R          # cp [source] [destination]
```

# Filter SNPs by linkage

## Exercise 1

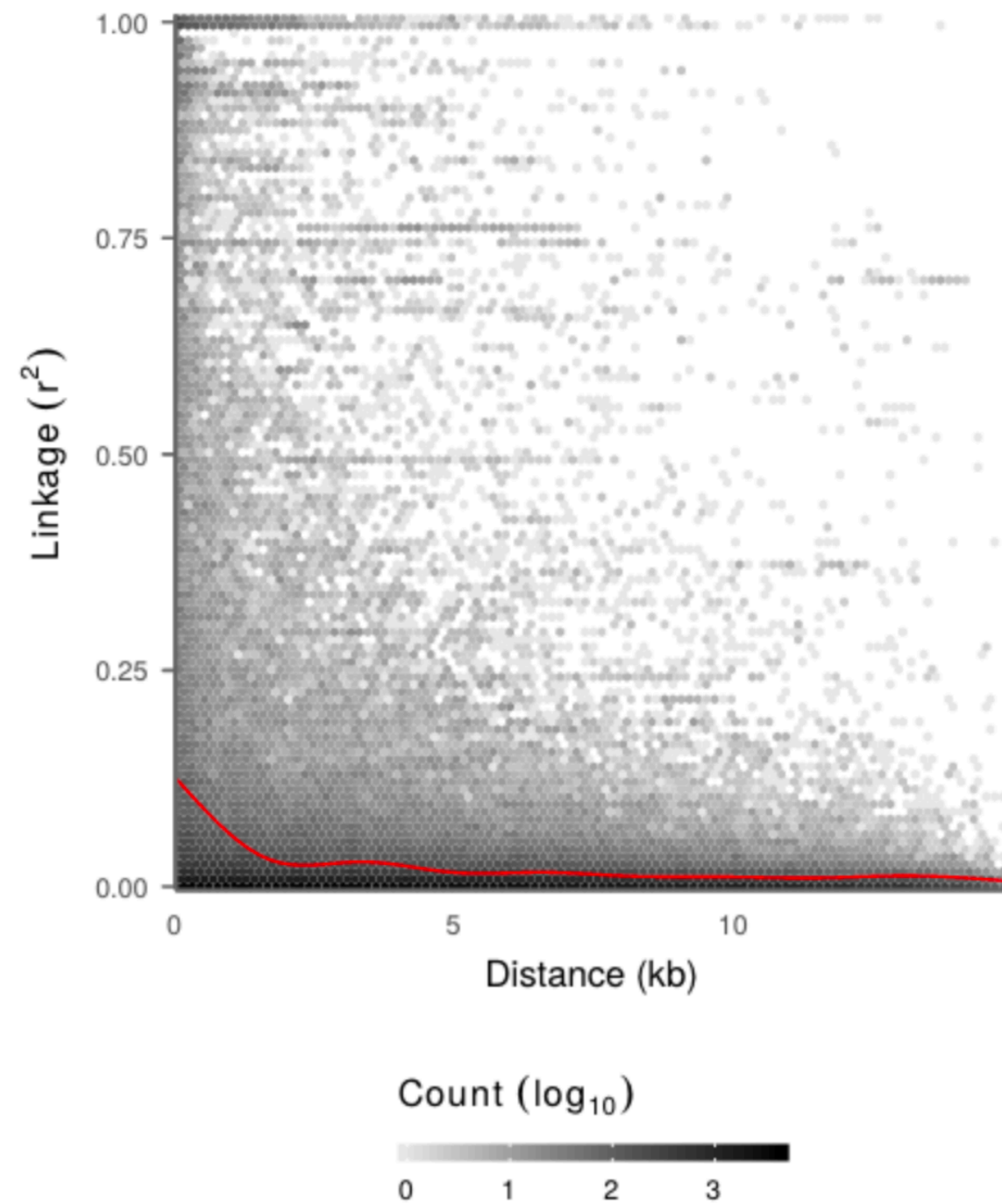
```
### Prune by LD (r2) with BCFtools
bcftools +prune snps_hamlets_filtered.vcf.gz \
  -m 0.4 \
  -Oz -o snps_hamlets_04r.vcf.gz

### Calculate r2 and D statistics with VCFtools (first 50 kb only)
vcftools --gzvcf snps_hamlets_04r.vcf.gz \
  --chr LG12 \
  --from-bp 1 \
  --to-bp 50000 \
  --hap-r2 \
  --stdout > LD_snps_hamlets_04r.tsv

### Download from cluster to "local"
scp <account>@carl.hpc.uni-oldenburg.de:/user/<account>/local/LD* .
```

# Decay of linkage with physical distance

## Exercise 1



Hench et al. 2019 (Nat Ecol Evol)

# LD statistics

## Recap

$$D_{AB} = p_{AB} - p_A p_B$$

Product of allele frequencies  
Haplotype frequency

Coefficient of linkage disequilibrium between two alleles  
0 to  $\pm 1$ , but constrained by allele frequencies

$$D' = D / D_{\max}$$

Max value given allele frequencies

$D$  normalized with respect to allele frequencies  
0 to  $\pm 1$ , full range (0: no association,  $\pm 1$ : perfect LD)

$$r^2 = \frac{D^2}{p_A (1 - p_A) p_B (1 - p_B)}$$

Correlation coefficient of linkage disequilibrium  
0 to 1, but constrained by allele frequencies

a.k.a.  $\rho$  (rho)

# Simulate selection and mutation

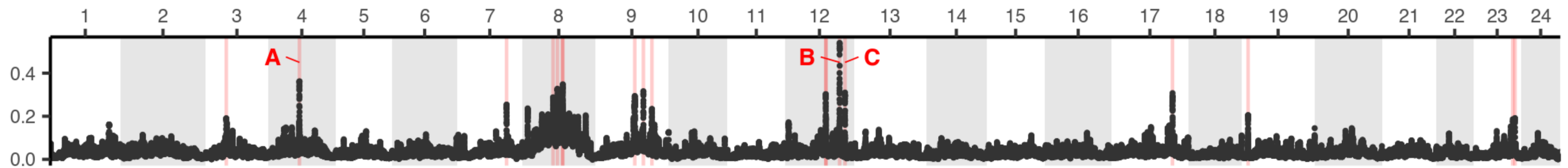
## Exercise 2

<http://evolutiongenetics.georgetown.edu/simulations/driftselection/>

Category	Genotype-specific fitness		
	$W_{AA}$	$W_{Aa}$	$W_{aa}$
Selection against a recessive phenotype	1	1	$1 - s$
Selection against a dominant phenotype	$1 - s$	$1 - s$	1
General dominance (dominance coefficient $0 \leq h \leq 1$ )	1	$1 - hs$	$1 - s$
Heterozygote disadvantage (underdominance for fitness)	1	$1 - s$	1
Heterozygote advantage (overdominance for fitness)	$1 - s$	1	$1 - t$

# Genome-wide scans using windows

## Exercise 3

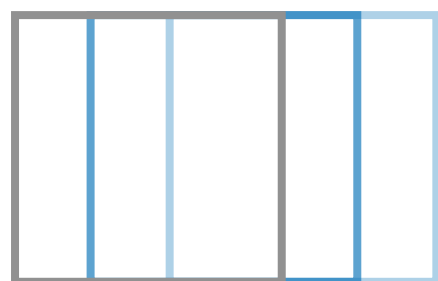


Hench et al. 2022, PNAS



...

Non-overlapping windows



...

Overlapping (sliding) windows



Window size

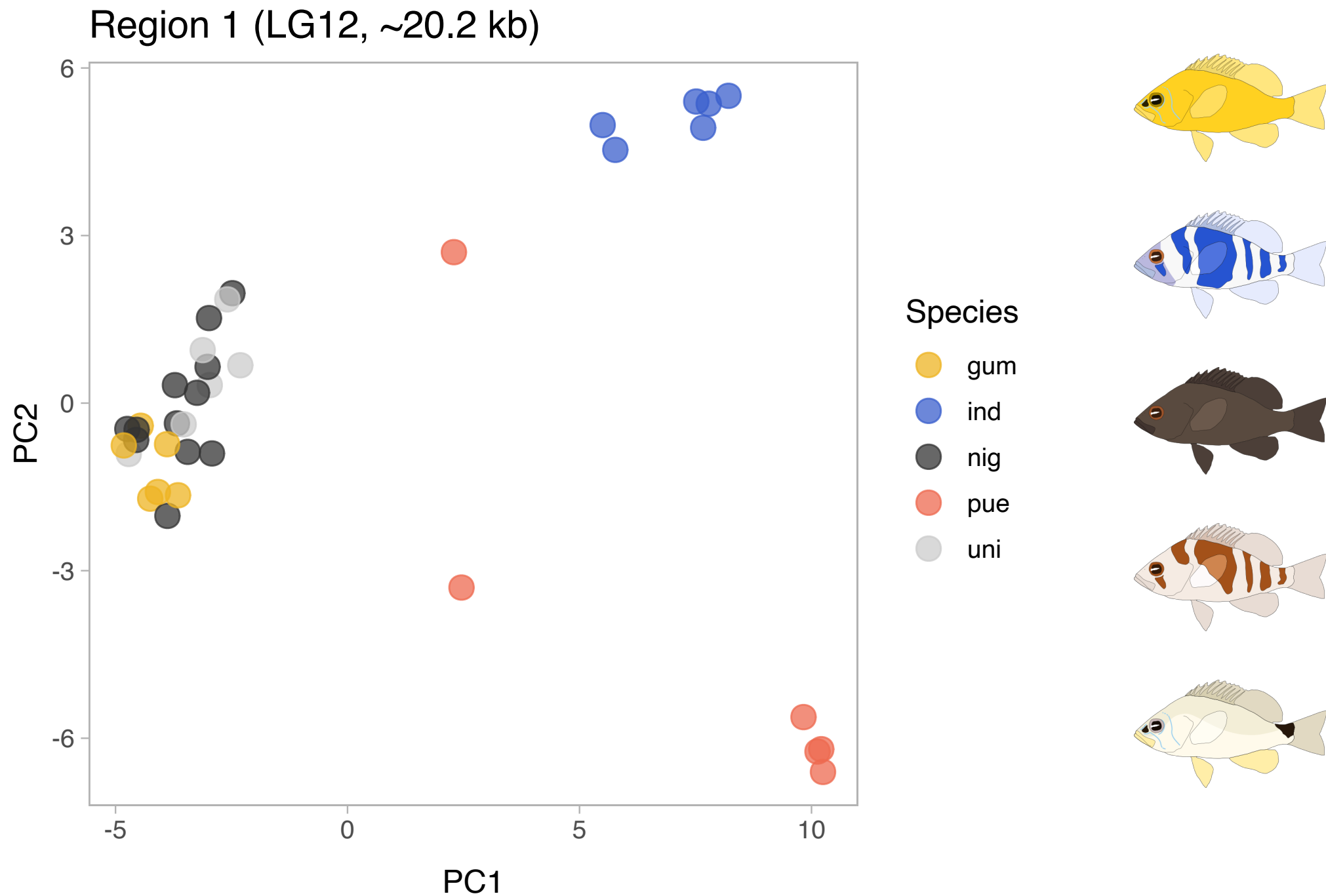


Increment / step size



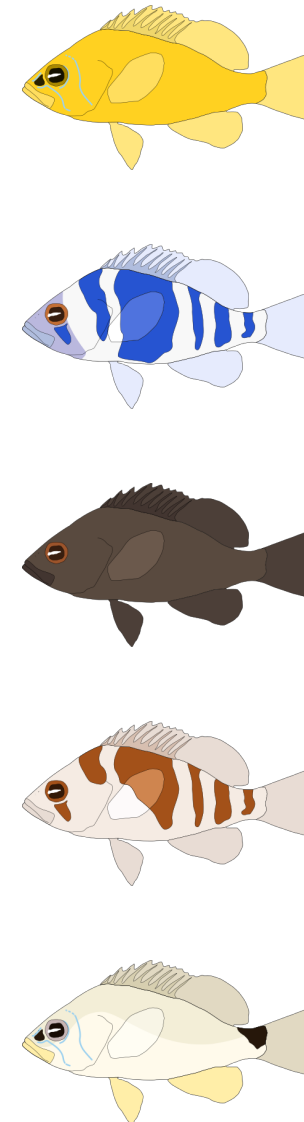
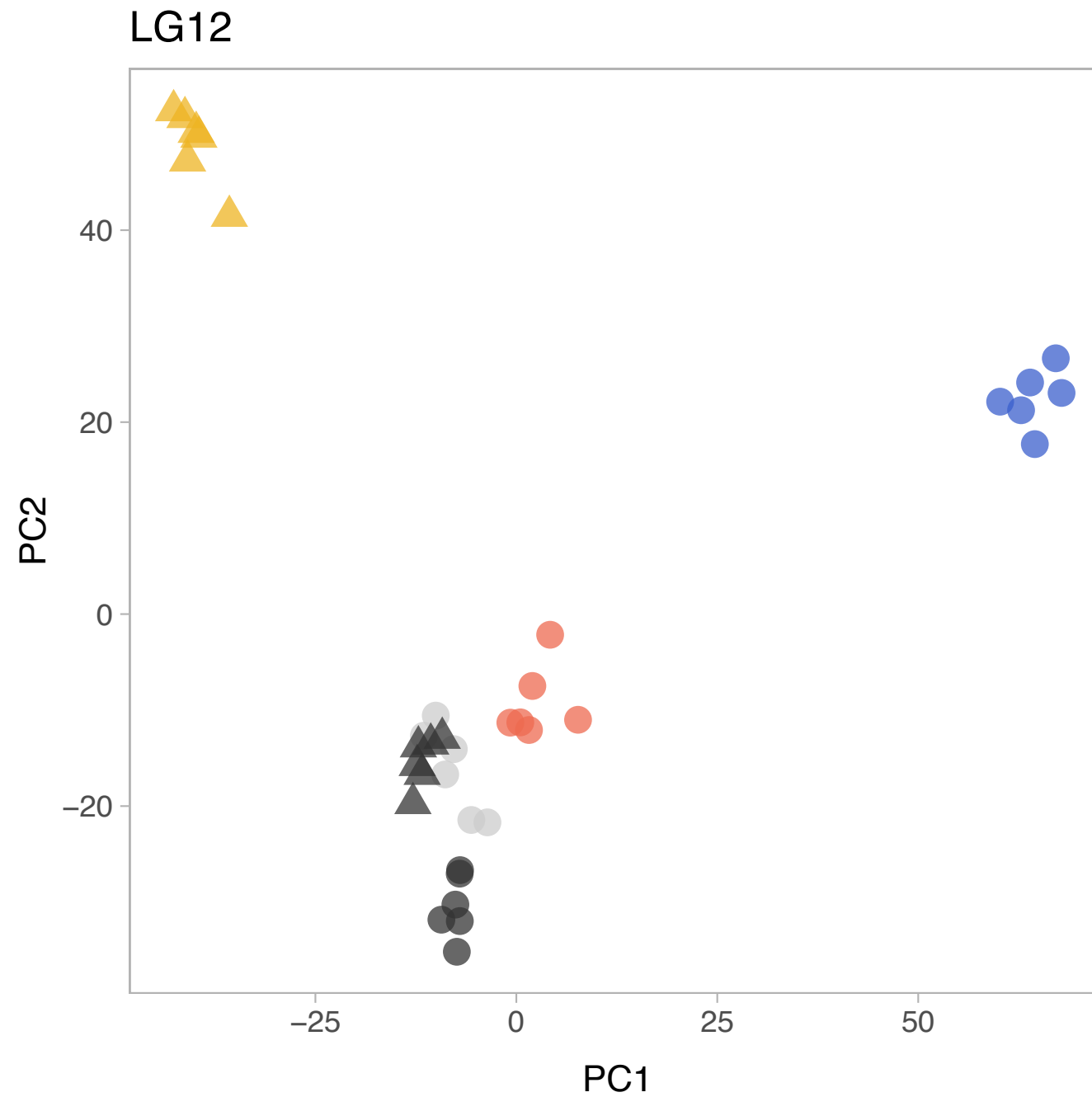
# $F_{ST}$ -based PCA of outlier region 1 on LG12

## Exercise 3



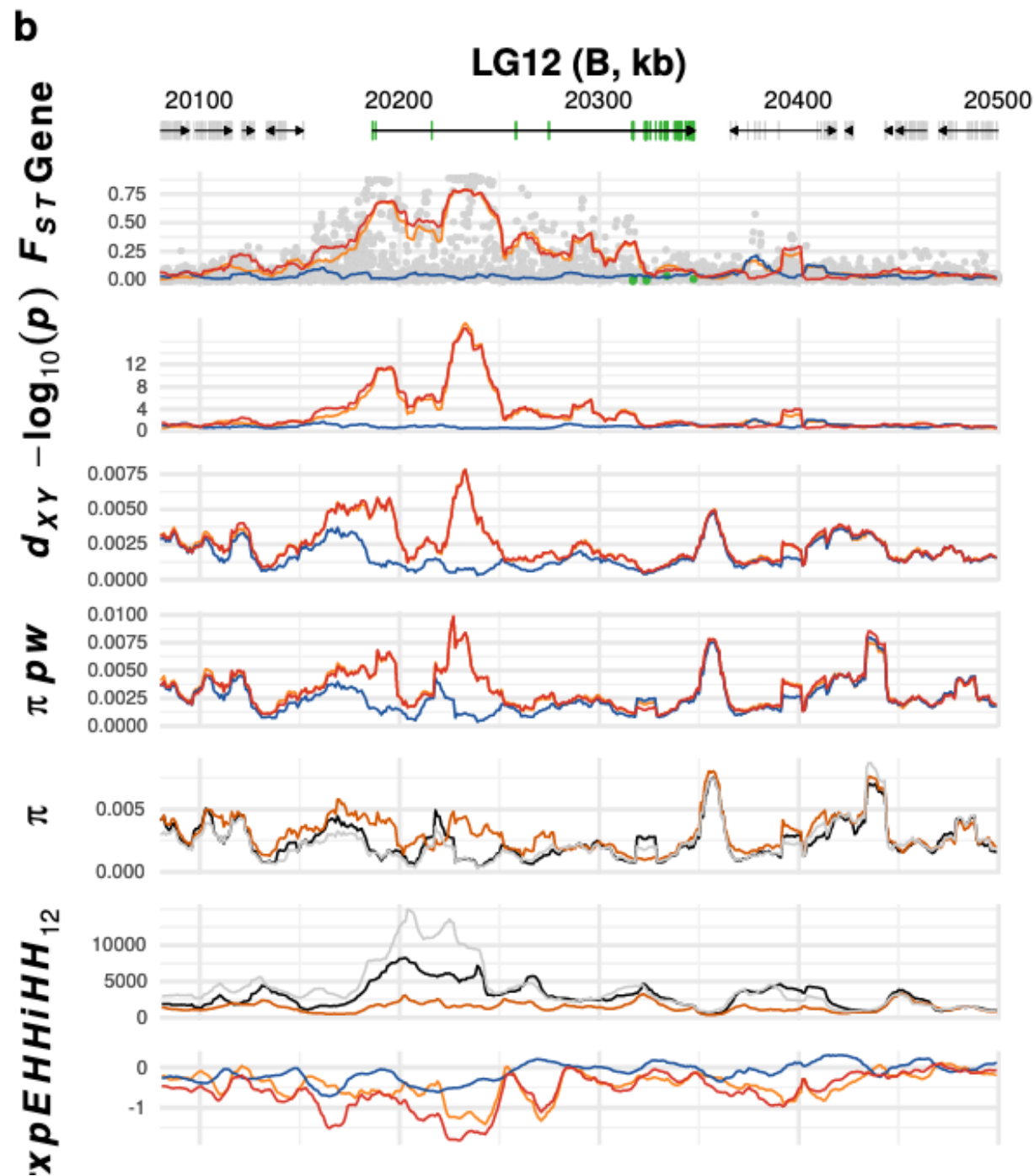
# $F_{ST}$ -based PCA, whole chromosome

## Exercise 3



# Genome-wide scans centered on region 1

## Exercise 3



### Genomewide scan for selection

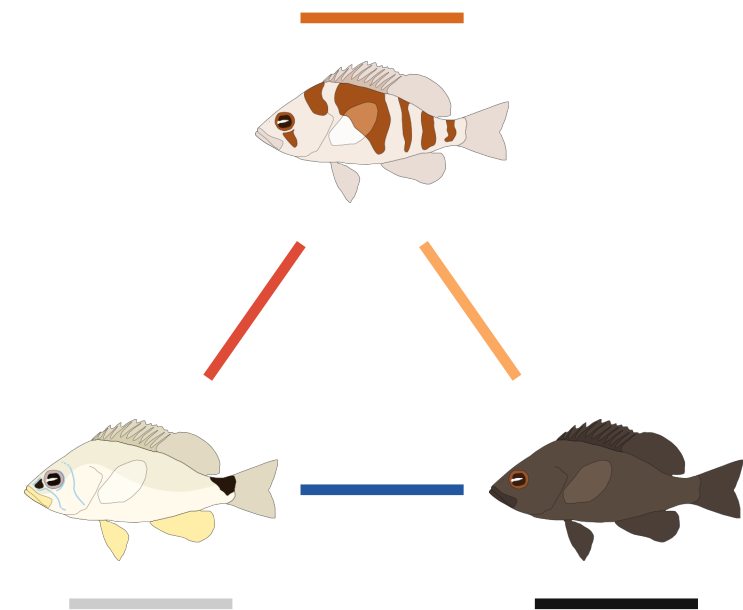
selscan

--ihh12 \ # or --xpehh

--vcf \

--map \

--out



Hench et al. 2019 (Nat Ecol Evol)

# Course outline

May be subject to change

Class	Date	Topics	Script
01	Apr 14	Introduction, software installation	01_intro.R
02	Apr 21	Hardy-Weinberg equilibrium	02_hwe.R
03	Apr 28	Genetic drift and effective population size	03_drift.R
04	May 05	Population structure and gene flow	04_structure.R
05	May 12	Isolation by distance (lecture online, exercises in person)	05_ibd.R
–	May 19	<a href="#">Himmelfahrt break</a>	–
06	May 26	Genome sequencing and assembly	06_genseq.sh
07	Jun 02	Genotyping, SNPs and population genomics	07_snps.sh
08	Jun 09	Recombination and linkage disequilibrium	08_recomb.R
–	Jun 16	<a href="#">Student presentations</a>	–
09	Jun 23	Selection and mutation	09_sel.R
10	Jun 30	DNA barcoding	10_barcode.sh
11	Jul 07	Metabarcoding	11_meta.sh
–	Jul 14	To be determined	–