

# Exercises in Marine Ecological Genetics

## 08. Detecting selection

- Detect selection by identifying  $F_{ST}$  outlier loci
- Perform genome scans using overlapping windows
- Learn about haplotype-based tests
- Investigate candidate genes under selection



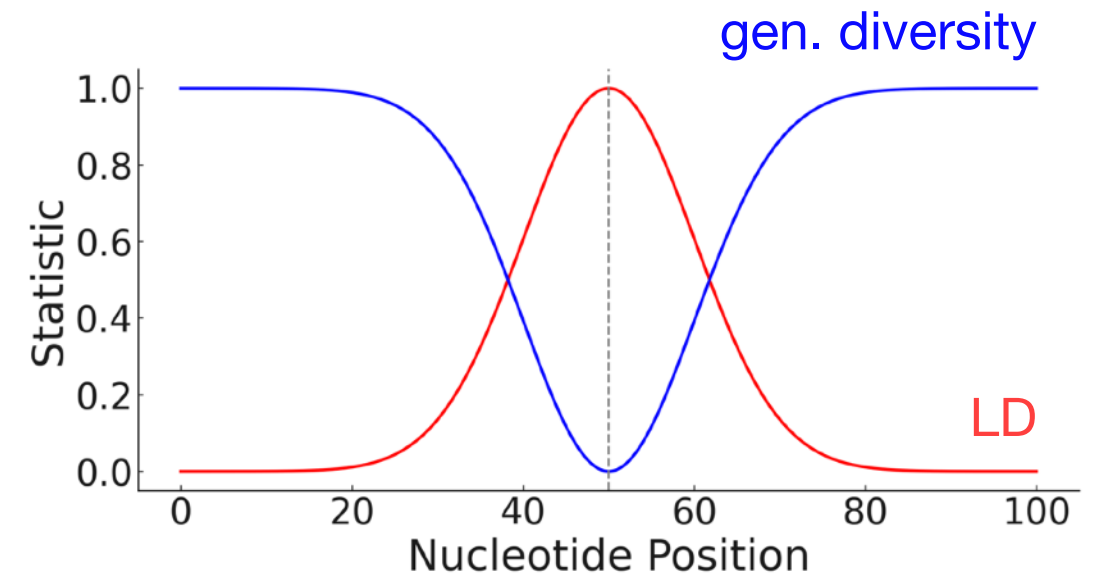
Martin Helmkamp

<https://github.com/mhelmkampf/meg25>

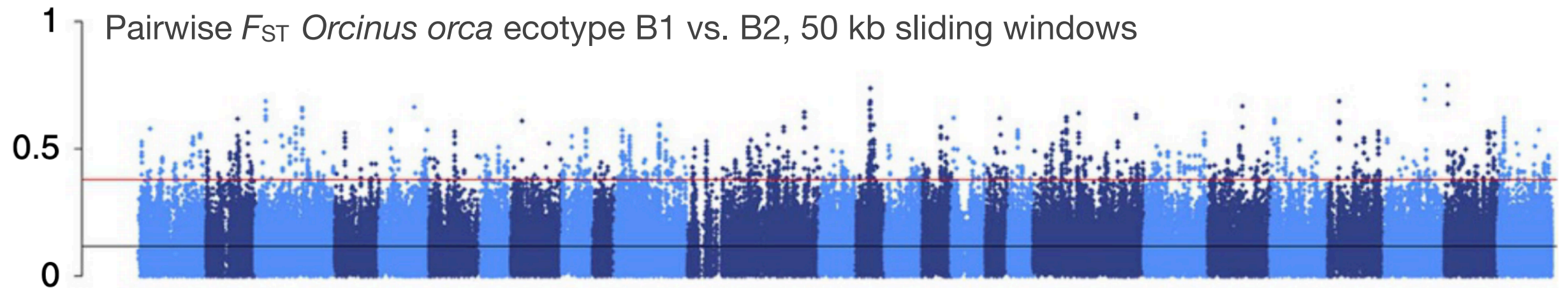
# Detecting selection

Positive selection influences genetic variation by:

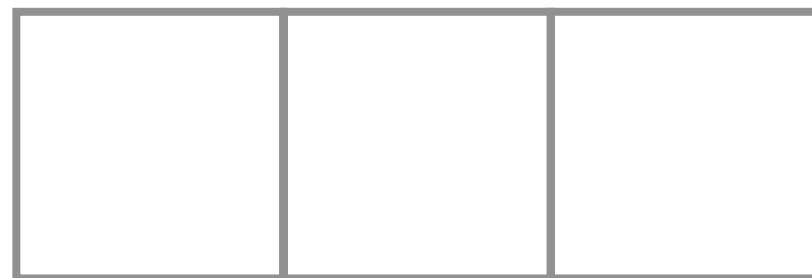
- Reducing genetic diversity (e.g.  $\pi$ )
- Increasing linkage disequilibrium
- Skewing allele frequencies / abundance of rare alleles (Tajima's D)
- Increasing population differentiation ( $F_{ST}$ )
- Extending homozygous haplotypes (iHS, EHH)



# Genome-wide scans using windows

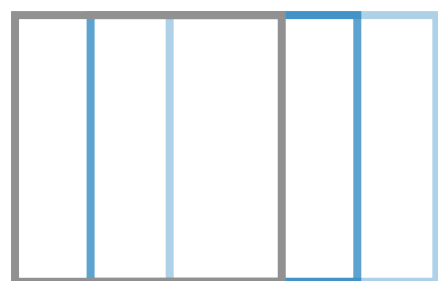


Footo et al. 2016, *Nature Communications*



...

Non-overlapping windows



...

Overlapping / sliding windows



Window size

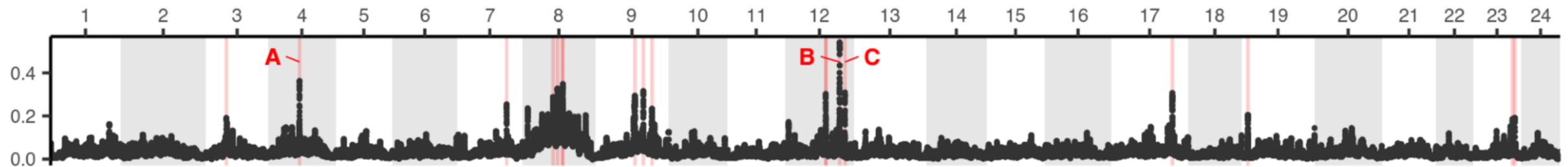


Increment / step size

# $F_{ST}$ scan across whole genome

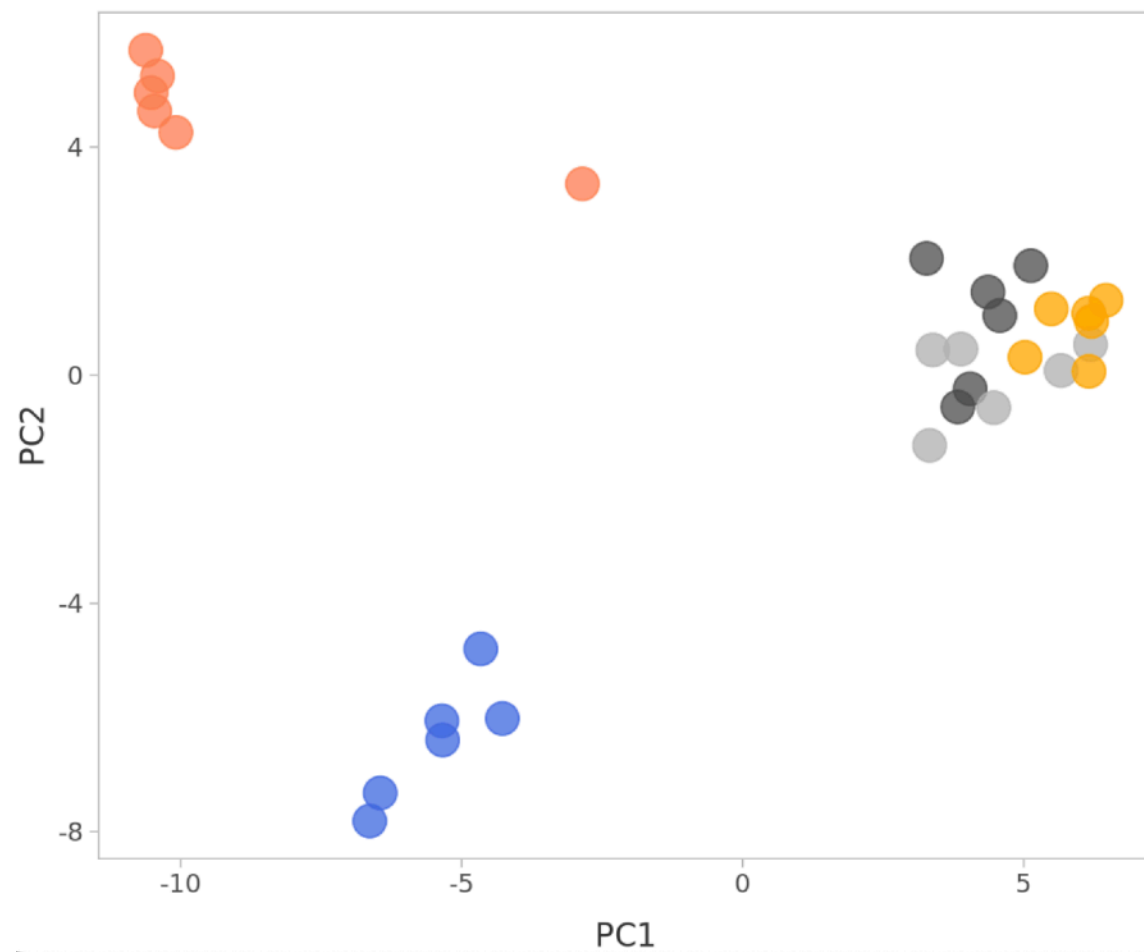
VCFtools, 50 kb sliding windows, joint  $F_{ST}$  of 14 Caribbean hamlet populations

Red = 99.8<sup>th</sup> percentile

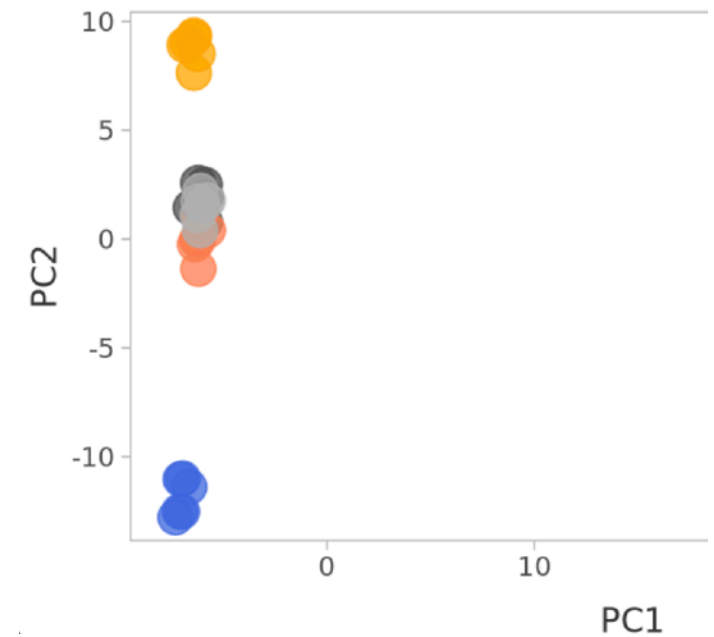
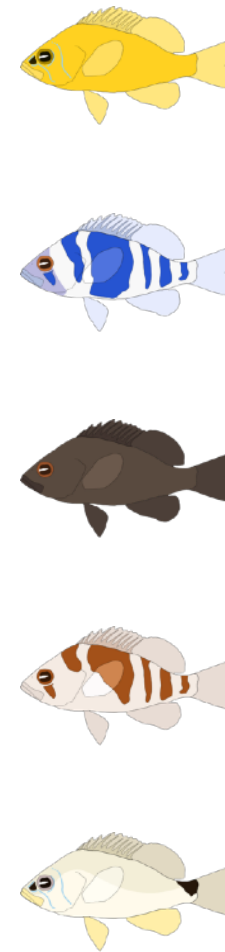


Hench et al. 2022, PNAS

# Outlier region and whole chromosome compared



PCA high  $F_{ST}$  region



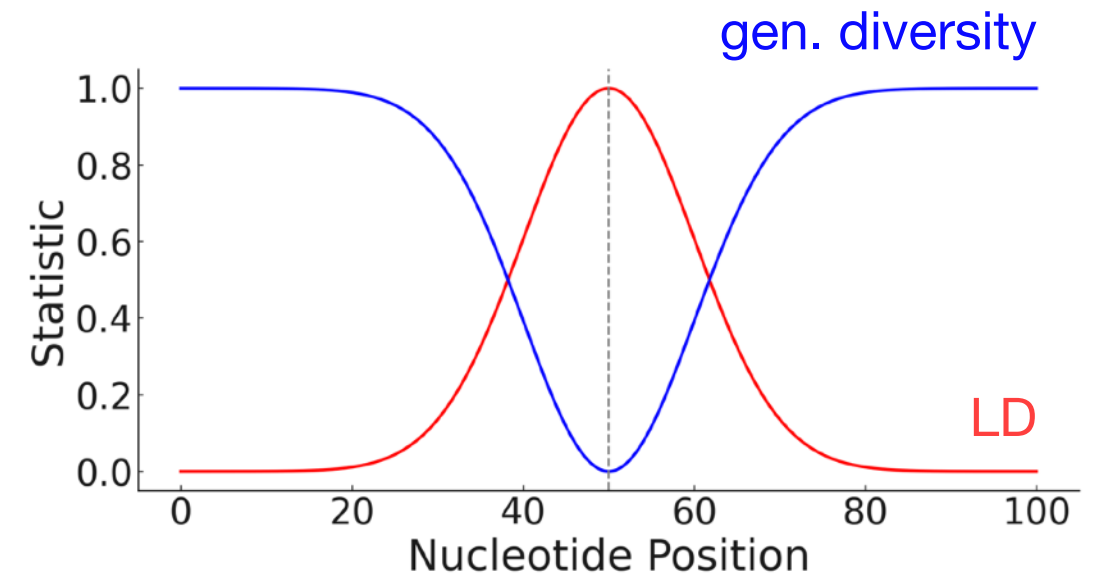
PCA whole chromosome 12

# General Feature Format (GFF)

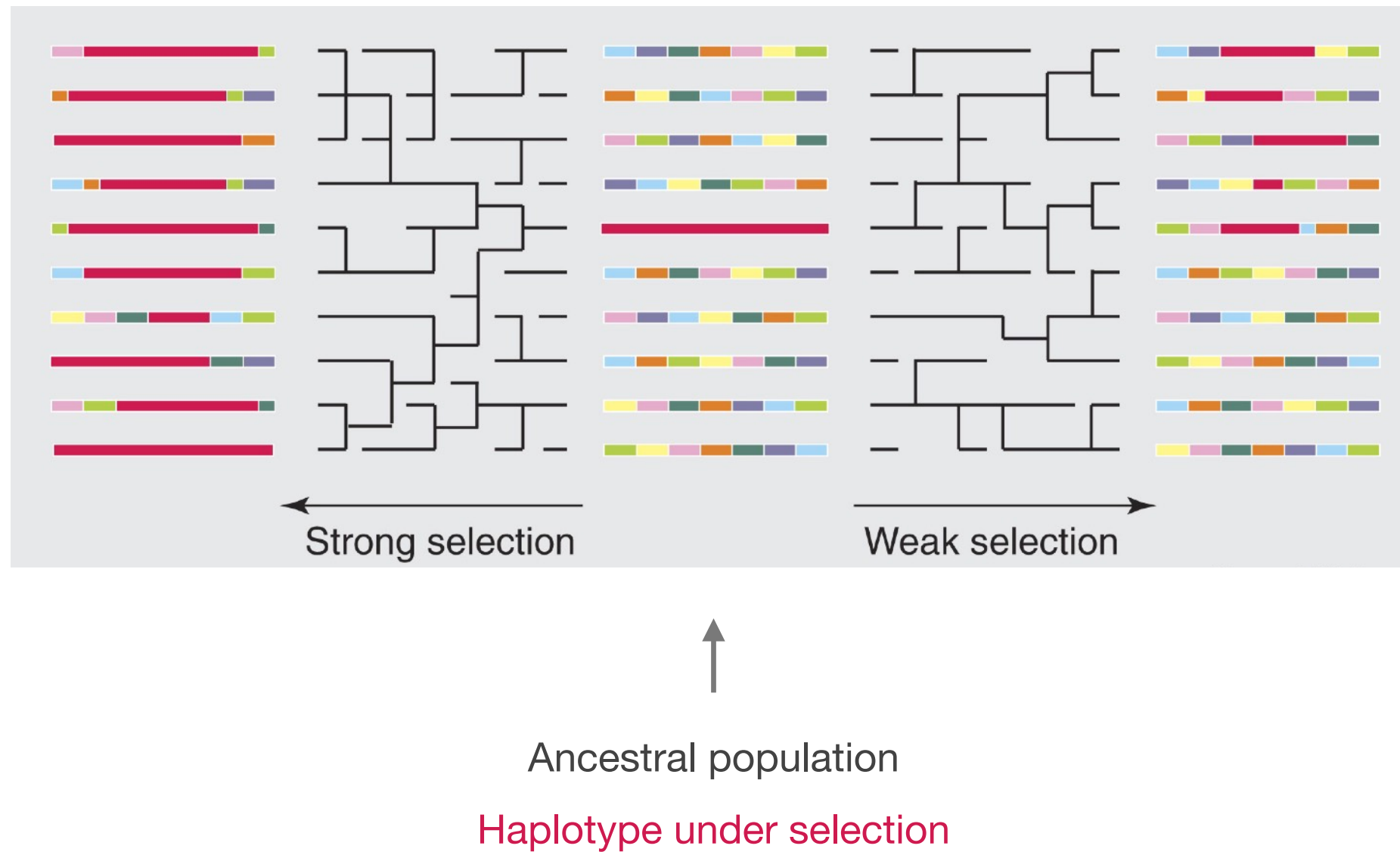
```
##gff-version 3;Parentgene=;Parentgenename=  
Contig49      .          gene          406          8055          .          +  
      .          ID=HPv1G000000000001;...;Parentgenename=ZNF484  
Contig49      .          mRNA          406          8055          .          +  
      .          ID=HPv1M000000000001;...;Parentgenename=ZNF484  
Contig49      .          exon          406          700          .          +  
      .          ID=HPv1E000000000001;...;Parentgenename=ZNF484  
Contig49      .          exon          5740         5843          .          +  
      .          ID=HPv1E000000000002;...;Parentgenename=ZNF484  
Contig49      .          exon          5993         6110          .          +  
      .          ID=HPv1E000000000003;...;Parentgenename=ZNF484  
Contig49      .          exon          7271         8055          .          +  
      .          ID=HPv1E000000000004;...;Parentgenename=ZNF484  
... 
```

# Haplotype-based methods

- Positive selection increases the frequency of a beneficial mutation in the population
- Nearby loci “hitchhike” due to linkage disequilibrium, faster than recombination can break up the association (**selective sweep**)
- As the haplotype rises in frequency, more individuals carry the same extended haplotype around the selected site (**extended haplotype homozygosity**)
- Haplotype-based methods (e.g. iHS, XP-EHH) detect this pattern of long, high-frequency homozygous haplotypes as a signature of recent positive selection
- Longer homozygous haplotypes = stronger evidence for selection



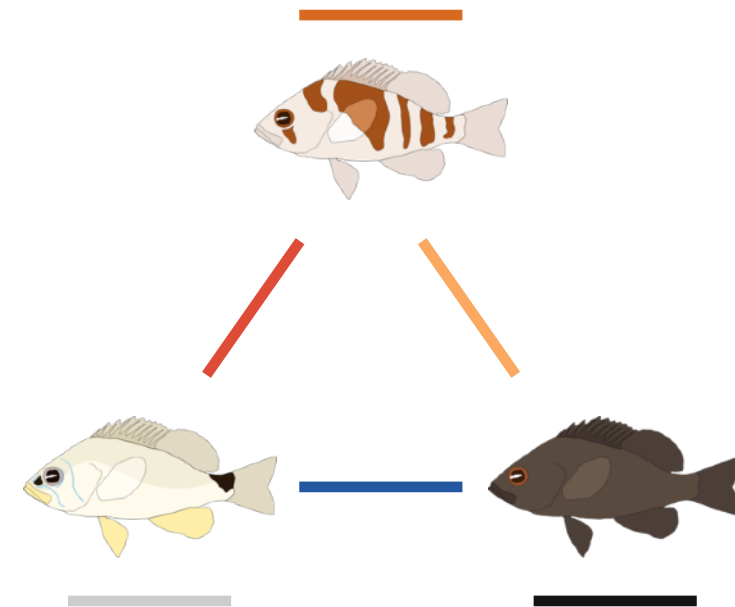
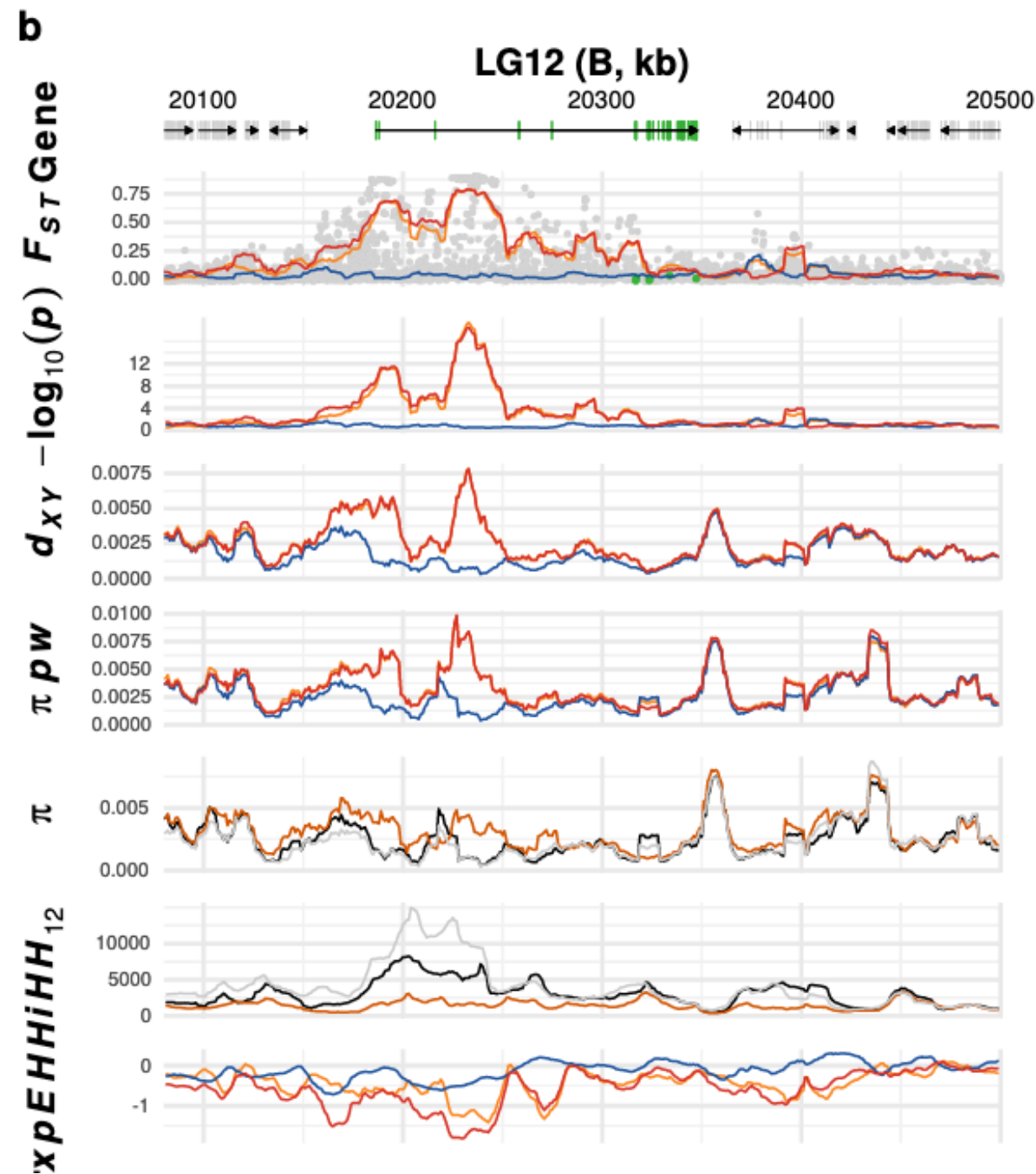
# Haplotype-based methods



Gibson 2007, *Current Biology*



# Genome-wide scans centered on *casz1* region



Hench et al. 2019, *Nat Ecol Evol*

# Phylogeny of *casz1* region

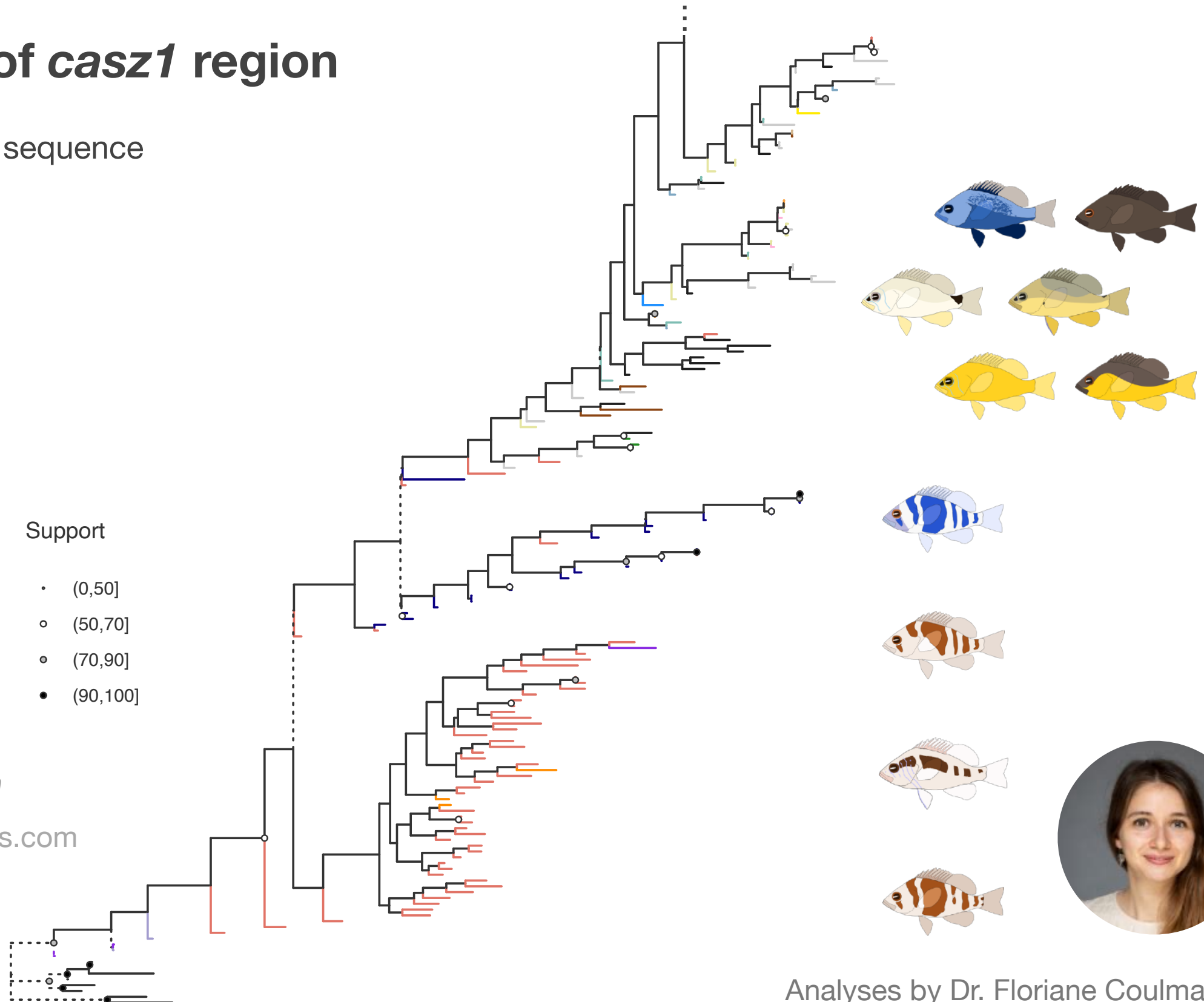
RAxML | ~ 200 kb sequence

Support

- (0,50]
- (50,70]
- (70,90]
- (90,100]

*Serranus tortugarum*

Image by whitecorals.com



Analyses by Dr. Floriane Coulmance

```
vcftools ... --weir-fst-pop --fst-window-step 5000 --fst-window-size 50000
```

```
scan_hh()          ### Scan for EHH (Extended Haplotype Homology), rehh package
```

```
manhattanplot()    ### Create Manhattan plot, rehh package
```

- Loci with unusually high  $F_{ST}$  may be under divergent selection between populations
- Haplotype-based methods detect recent selection by identifying long, unbroken stretches of shared DNA (extended homozygous haplotypes)
- Sliding-window scans smooth out noise and help detect regional signals of selection along the genome