

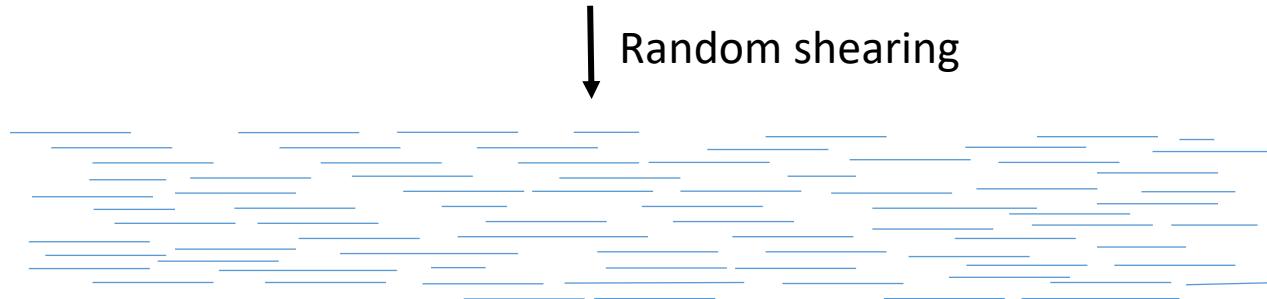
Summary of the analysis
pipeline until now

&

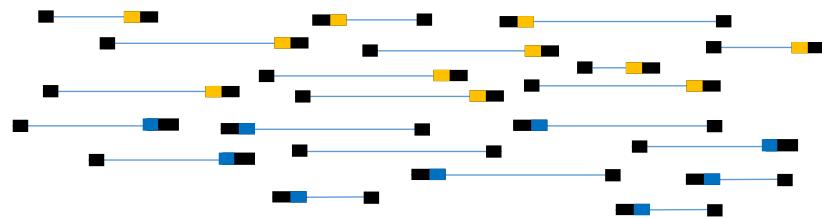
Introduction of
Pundamilia cichlids

Whole-genome short-read sequencing

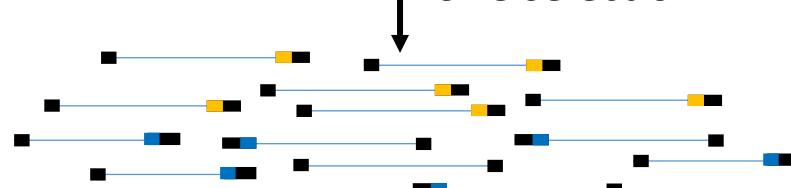
DNA



Illumina adapter ligation
incl. individual index

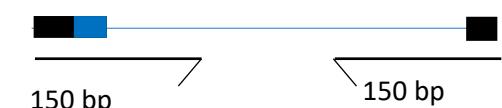


Size selection

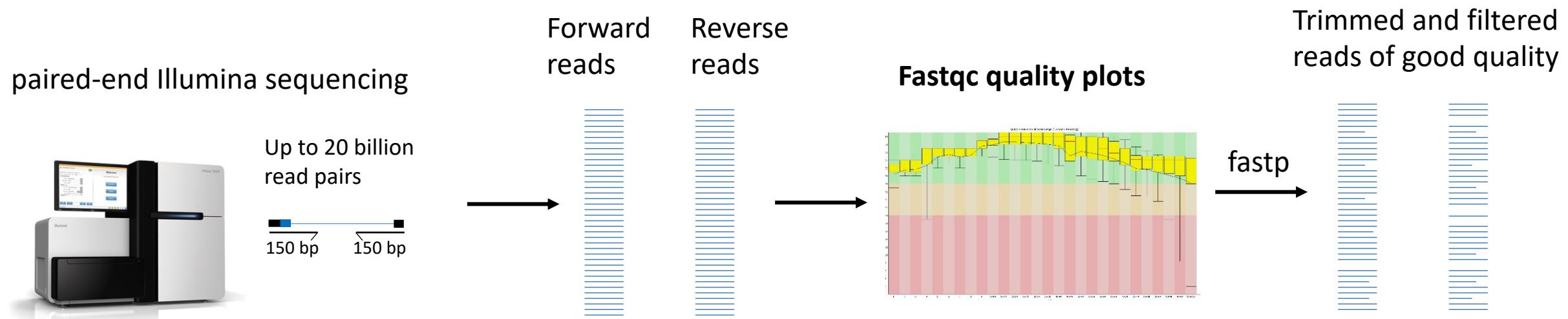


paired-end sequencing

Up to 20 billion read pairs



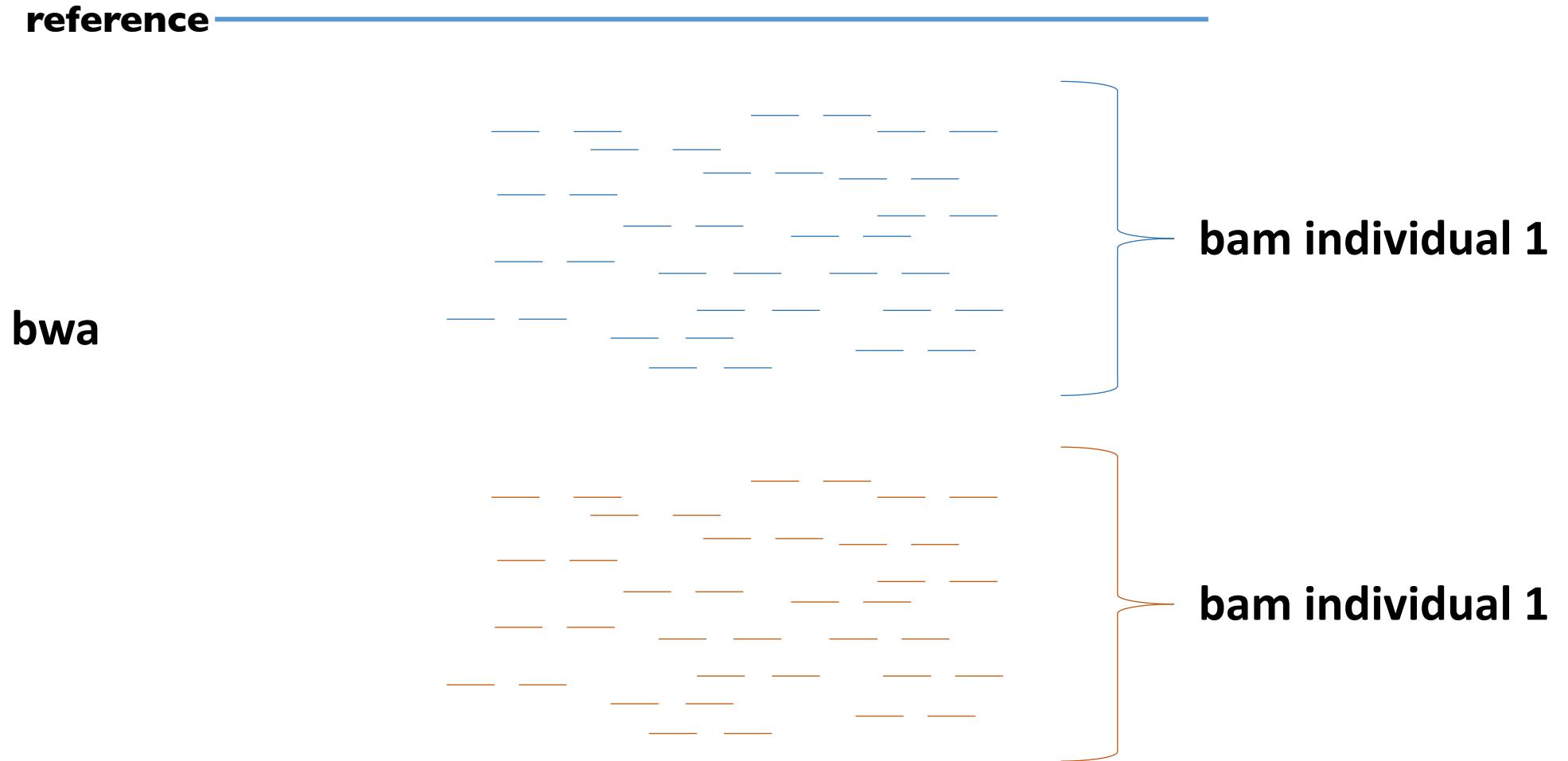
1. Quality check and trimming raw reads



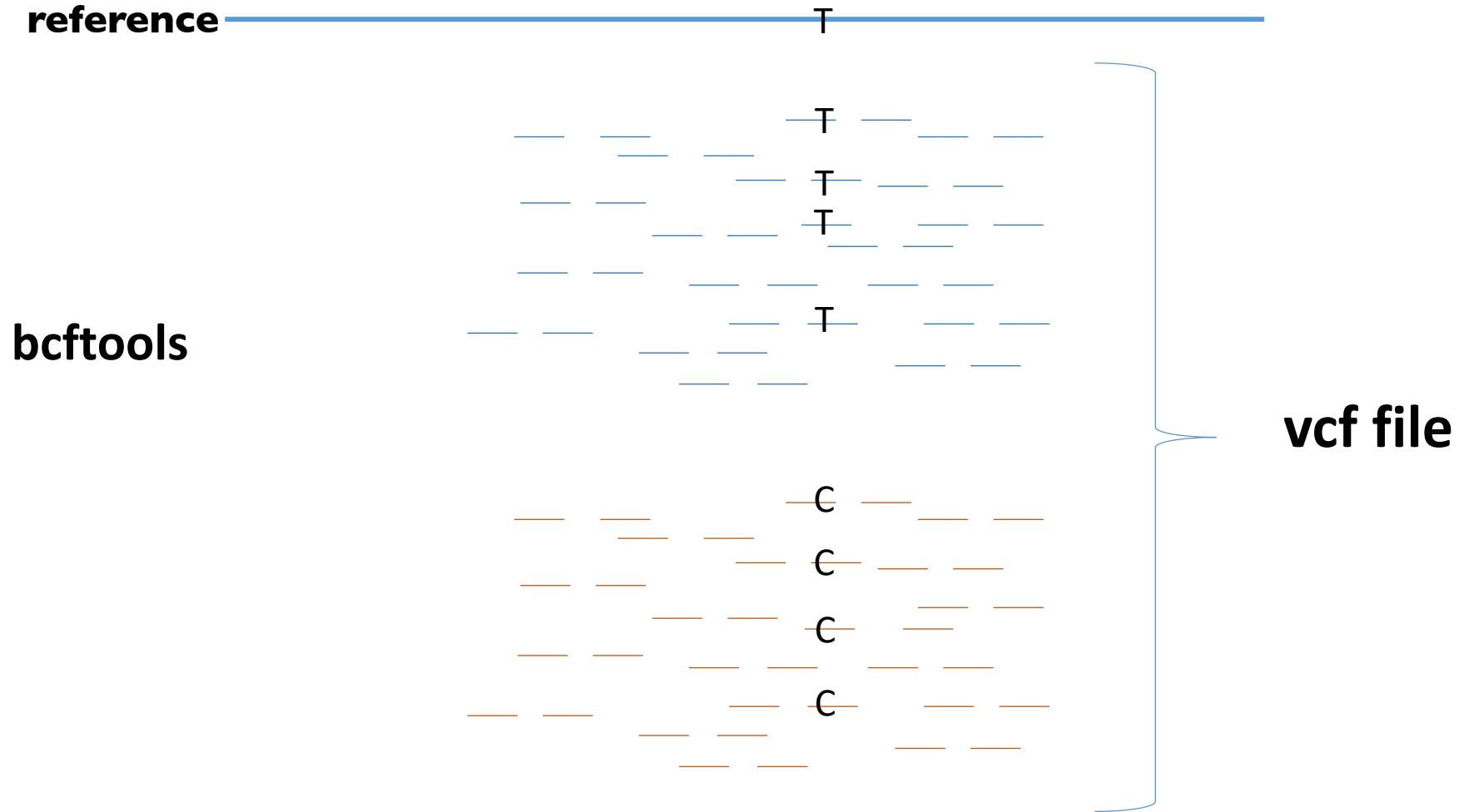
Fastq file of reads

@HWUSI-EAS611:34:6669YAXX:1:1:5069:1159 1:N:0:
TCGATAATACCGTTTTCCGTTGATGTTGATACCATT
+
II

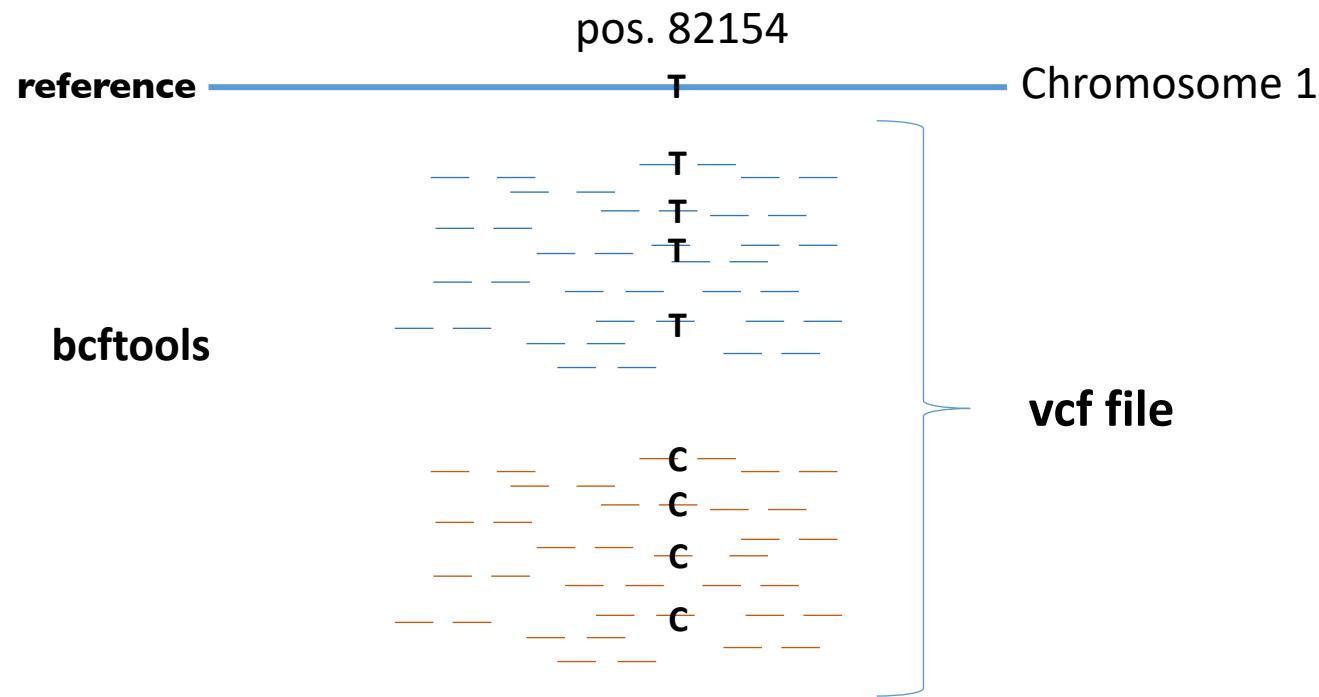
2. Alignment to the reference genome with bwa



3. Variant and genotype calling with bcftools



3. Variant and genotype calling with bcftools



vcf file: Genotypes for each individual at genomic sites

##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype"										
#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	GEN	
chr1	82154	.	T	C	.	GT	0/0 1/1	0/0 1/1 0/0 1/1	0/0 1/1 0/0 1/1	1/1
chr1	752566	.	T	.	.	GT	0/0 0/0	0/0 0/0 0/0 0/0	0/0 0/0 0/0 0/0	0/0
chr1	752721	.	T	C	.	GT	1/1 1/1	1/1 1/1 1/1 1/1	1/1 1/1 1/1 1/1	1/1
chr1	752721	.	A	.	.	GT	./. 0/0	0/0 0/0 0/0 0/0	0/0 0/0 0/0 0/0	0/0

Fastqc

Check read quality

fastp

Quality trim

Remove adapter

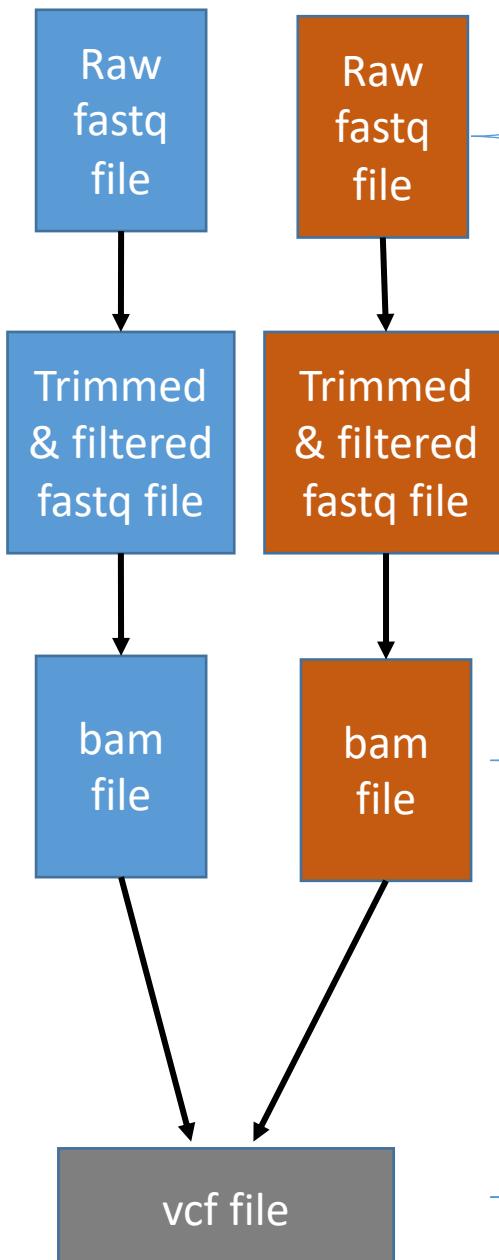
Remove polyG-tails

bwa

Align to reference

bcftools

Call variants
and genotypes



Fastq: raw reads with sequencing quality information

```

@HWUSI-EAS611:34:6669YAXXX:1:1:5069:1159 1:N:0:
TCGATAATACCGTTTTTCCGTTGATGTTGATACCATT
+
IIHIIHIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
  
```

sam file: reads mapped to the reference genome -> binary version: bam file

```

HWI-ST1145:74:C101DACXX:7:1114:2759:41961      16     chr20    193953  50    100M   *     0     0
        TGCTGGATCATCTGGTTAGTGGCTTCTGACTCAGAGGACCTTCGCCCCCTGGGCAGTGGACCTCCAGTGATTCCCCTGACATAAGGGGCATGGACGA
        G     DCDDDDDEDDDDDDCDDDDDDCCCDDDCDDDDDEEC>DFFEJJJJJIGJJJIHGBHHGJIJJJJJJGJJJIJJJJJIHJJJJJJHHHHHFFFFCCC
        AS:i:-16    XM:i:3 X0:i:0 XG:i:0 MD:Z:60G16T18T3 NM:i:3 NH:i:1
HWI-ST1145:74:C101DACXX:7:1204:14760:4030      16     chr20    270877  50    100M   *     0     0
        GGCTTATTGGTAAAAAAGGAATAGCAGATTAAATCAGAAATTCCCACCTGGCCCAAGCAGCACCAACCAGAAAGAAGGGAAAGAACAGGAAAAACCA
        C     DDDDDDDDDCDDDDDDDDDEEEEEEFFFEFFEGHHHHFGDJJIHJJJIJJJJIIIGGFJJJIHIIIIJJJJJIGHHFAHGFHJHFGGHFFFDD@BB
        AS:i:-11    XM:i:2 X0:i:0 XG:i:0 MD:Z:0A85G13 NM:i:2 NH:i:1
HWI-ST1145:74:C101DACXX:7:1210:11167:8699      0      chr20    271218  50    50M4700N50M   *     0
        0      GTGGCTTCCACAGGAATGTTGAGGATGACATCCATGTCTGGGTGCACTTGGTCTCCGAAGCAGAACATCCTCAAATATGACCTCTG
  
```

vcf file: Genotypes for each individual at genomic sites

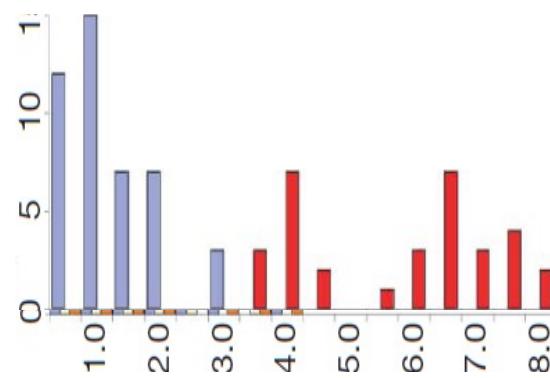
```

##FORMAT=<ID=GT,Number=1>Type=String>Description="Genotype"
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT GEN
chr1 82154 . T C . GT 0/0 1/1 0/0 1/1 0/0 1/1
chr1 752566 . T . . GT 0/0 0/0 0/0 0/0 0/0 0/0
chr1 752721 . T C . GT 1/1 1/1 1/1 1/1 1/1 1/1
chr1 752721 . A . . GT ./ 0/0 0/0 0/0 0/0 0/0
  
```

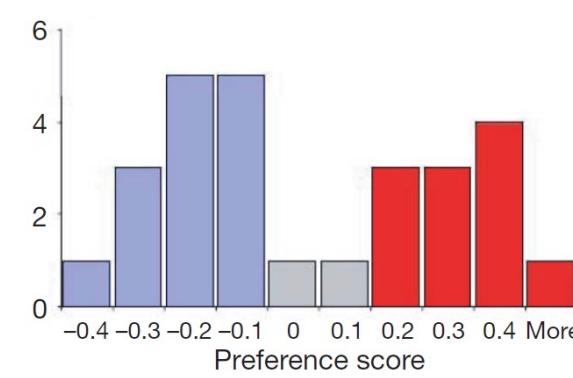
Pundamilia species complex



Water depth distribution



Female preference

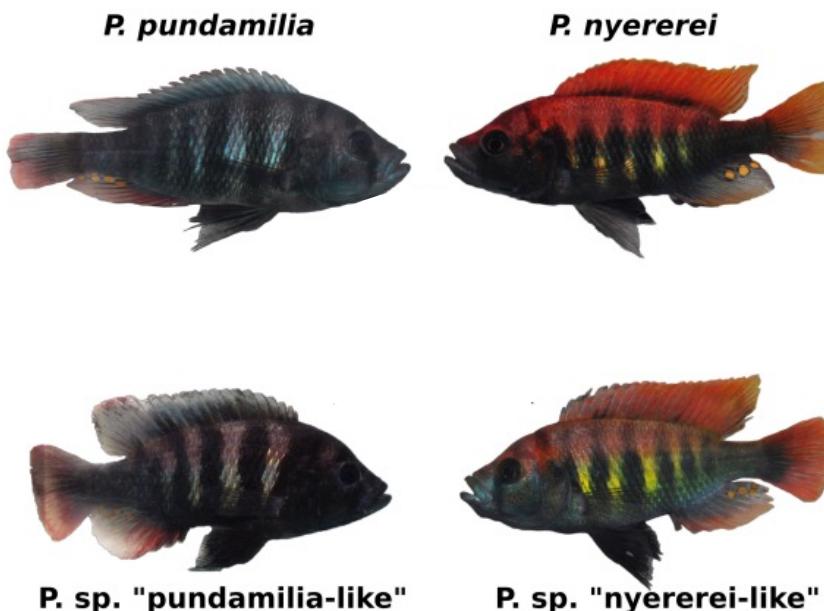


LWS opsin alleles

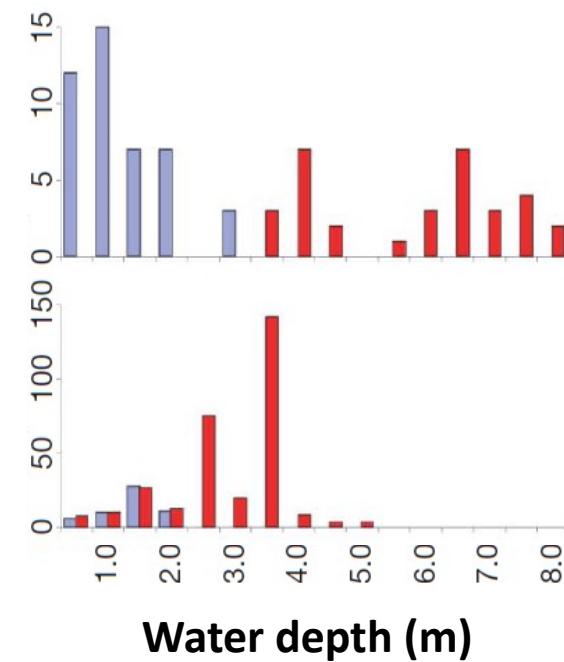


Parallel differences in phenotype, water depth, color vision

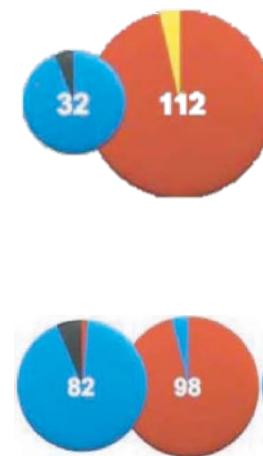
Male nuptial coloration



Water depth distribution



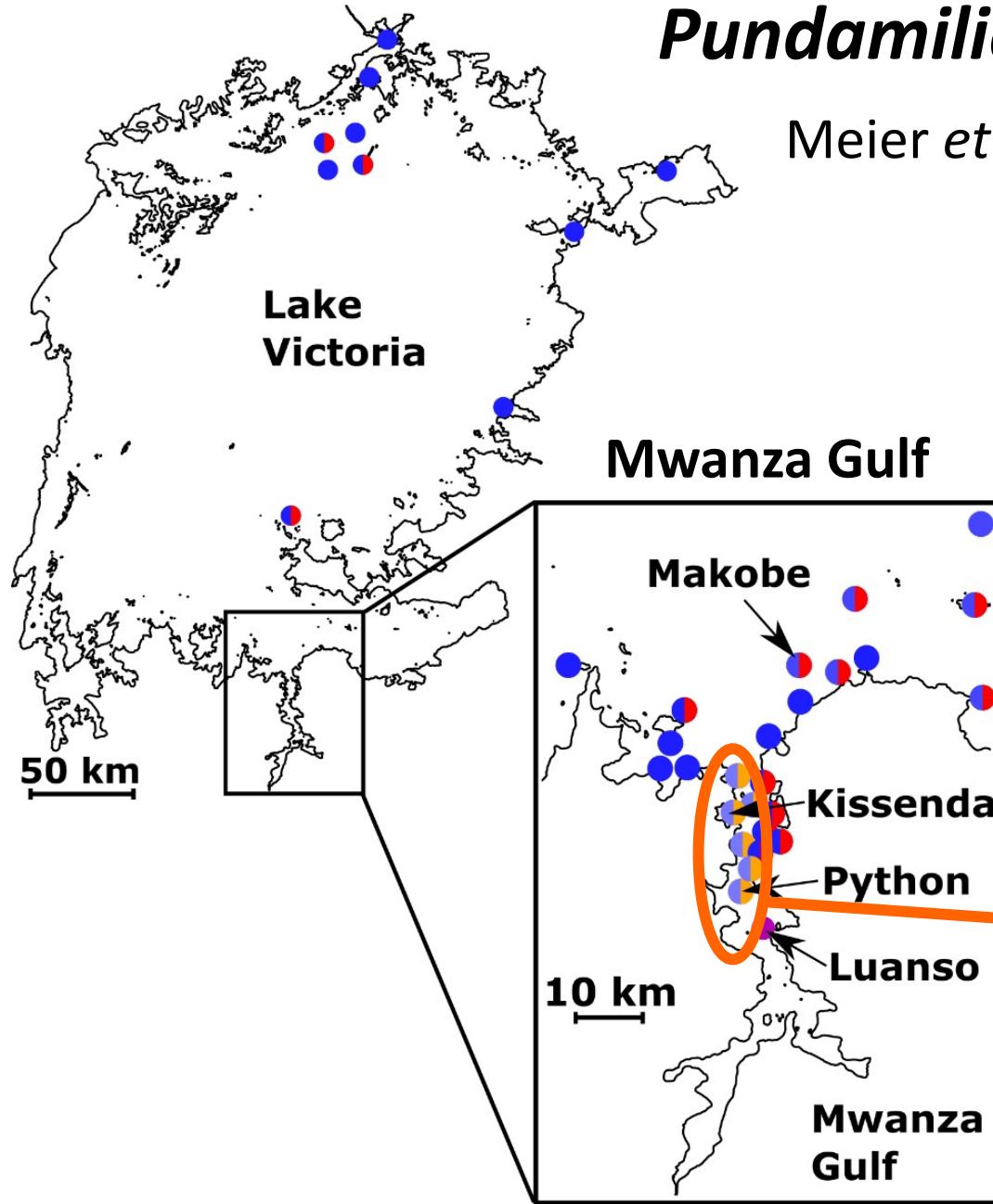
LWS opsin haplotypes



Seehausen *et al.*, 2008
Seehausen, 2009

Pundamilia species complex

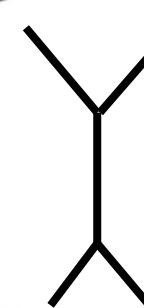
Meier et al. (2017) MolEcol



P. pundamilia



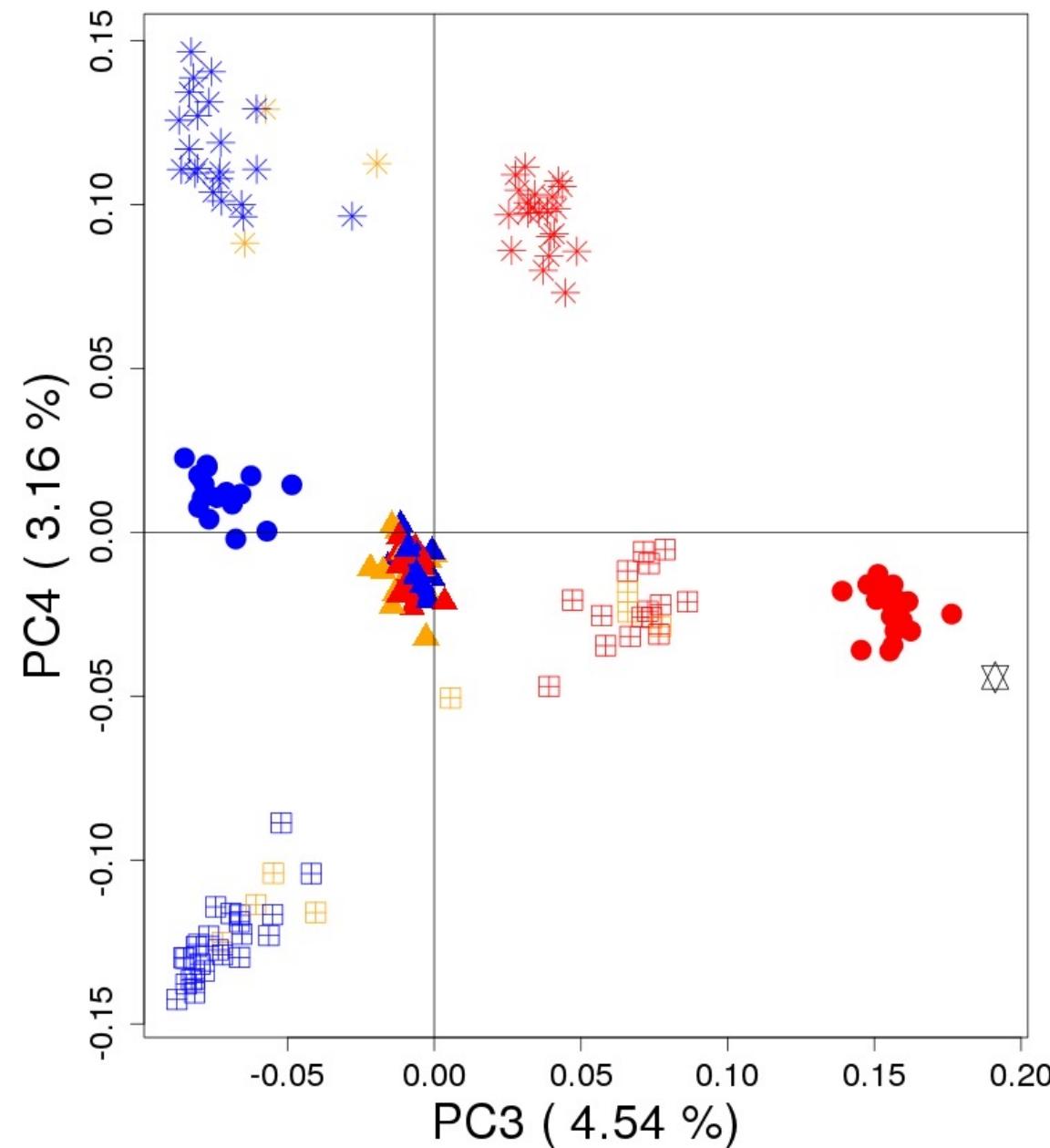
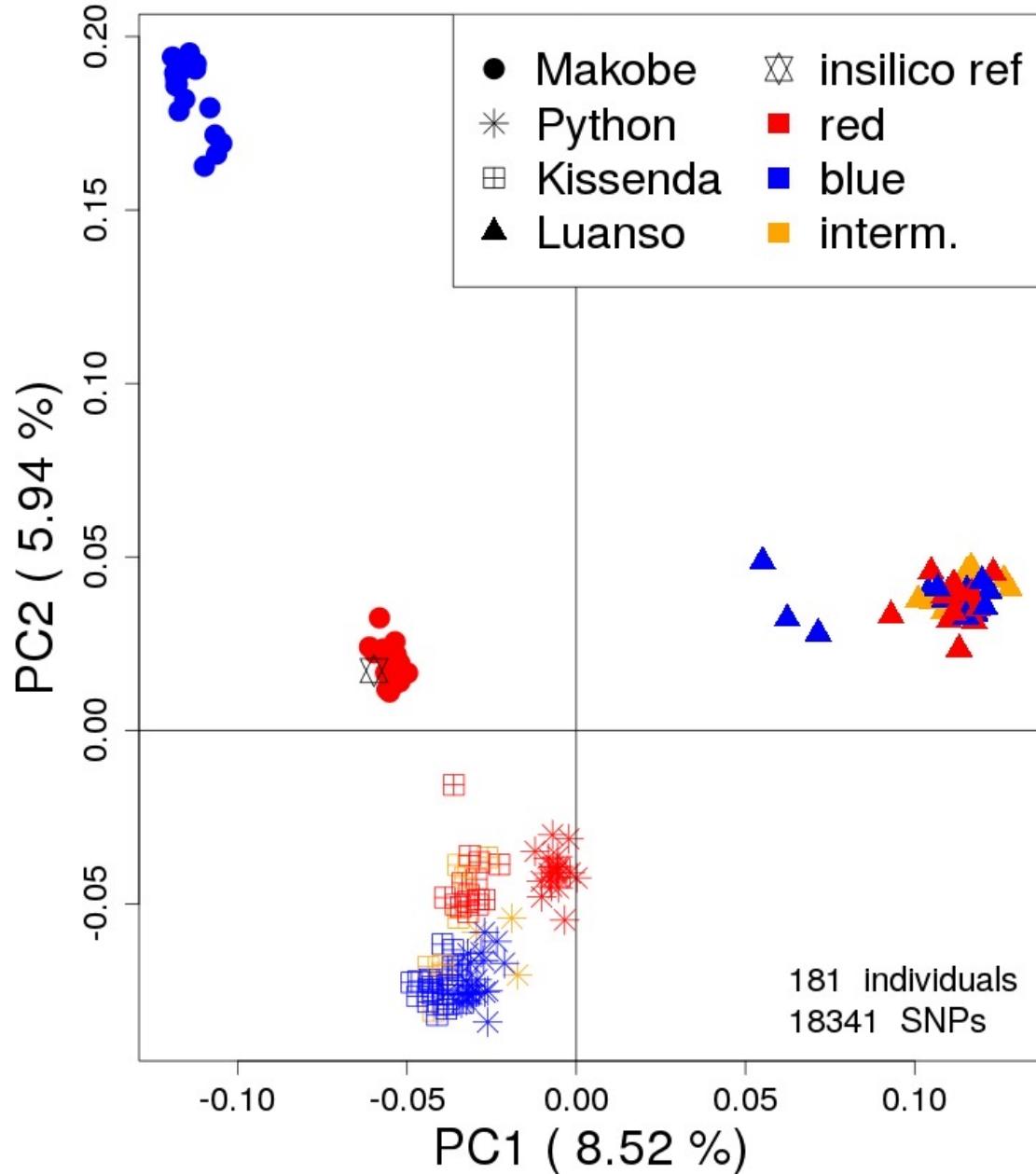
P. nyererei



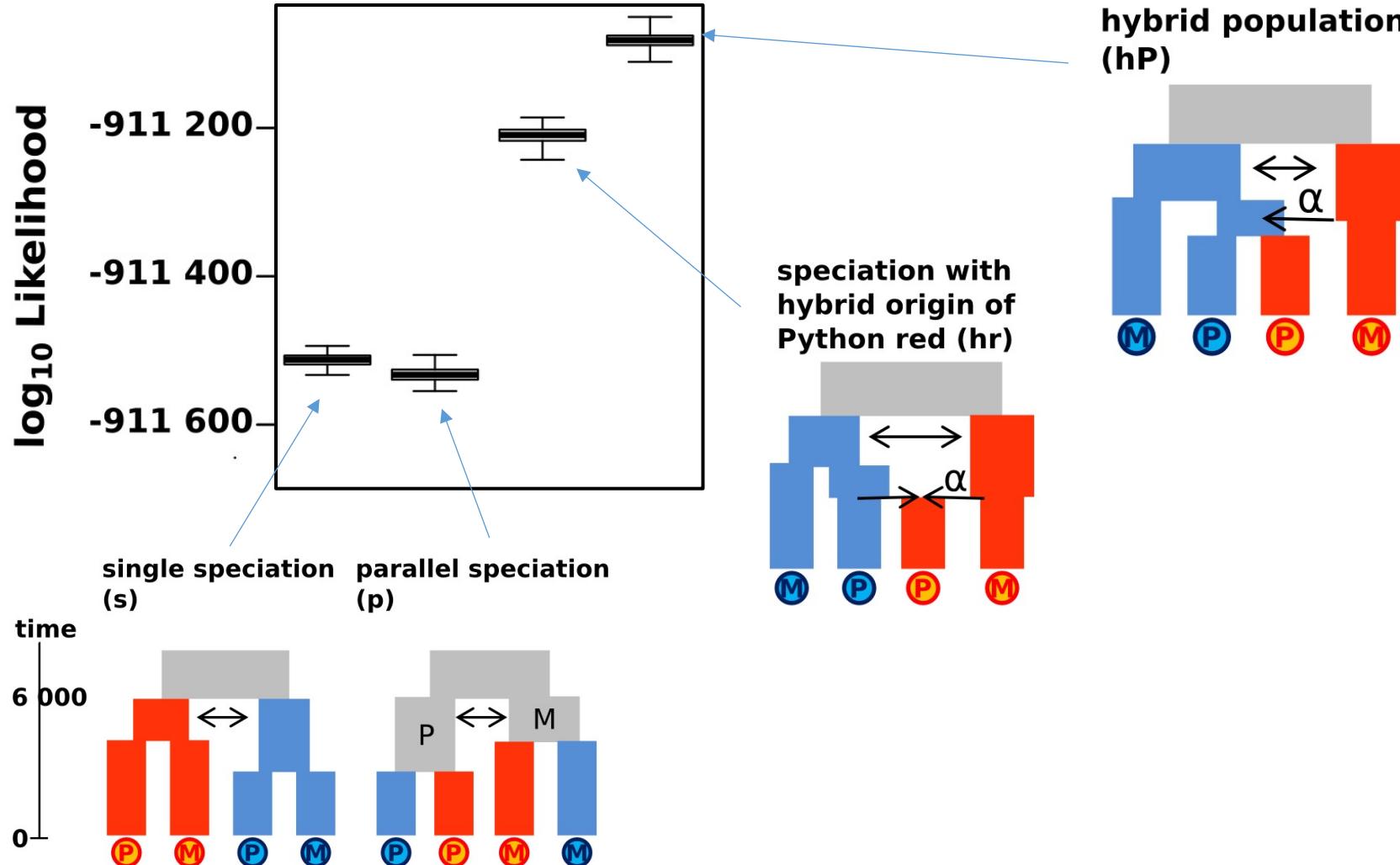
P. sp. "pundamilia-like"



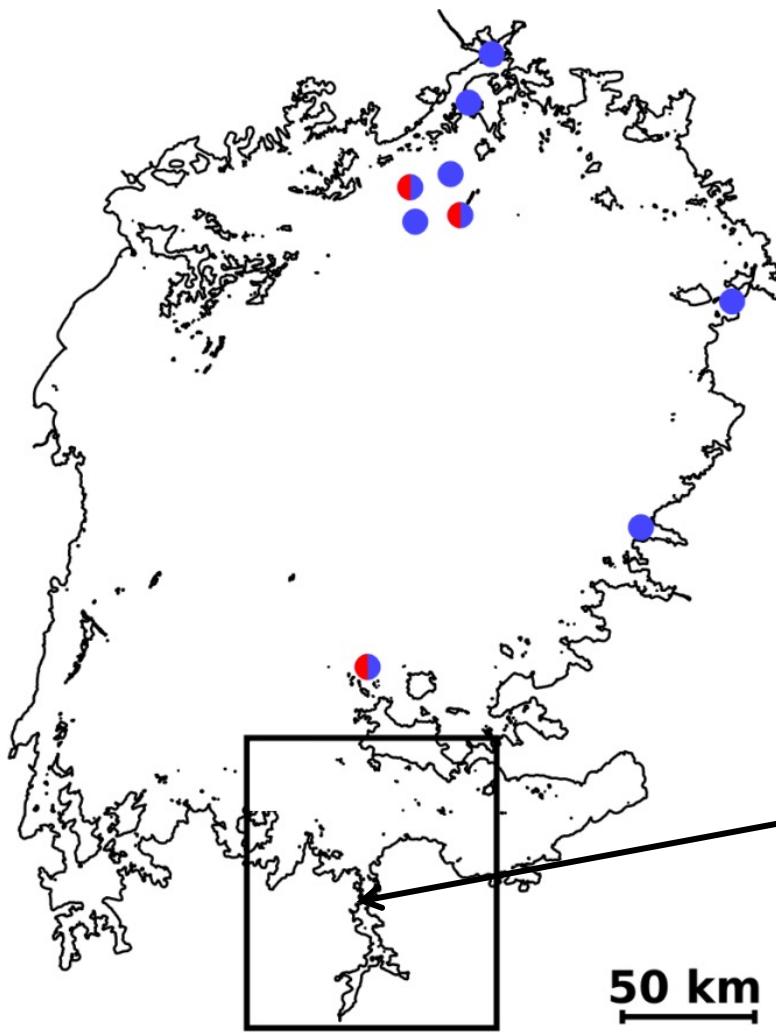
P. sp. "nyererei-like"



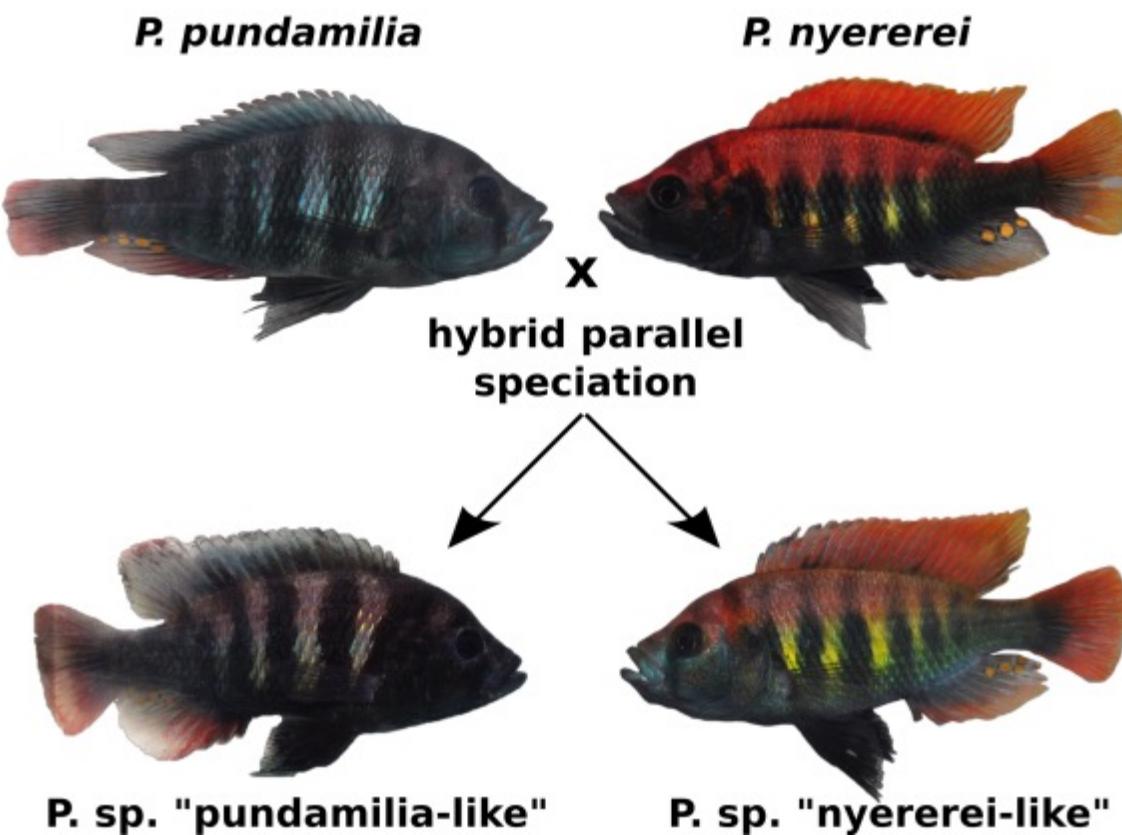
Demographic modeling of whole genomes reveals speciation from a hybrid population in the Mwanza Gulf species pair



Recent hybrid parallel speciation



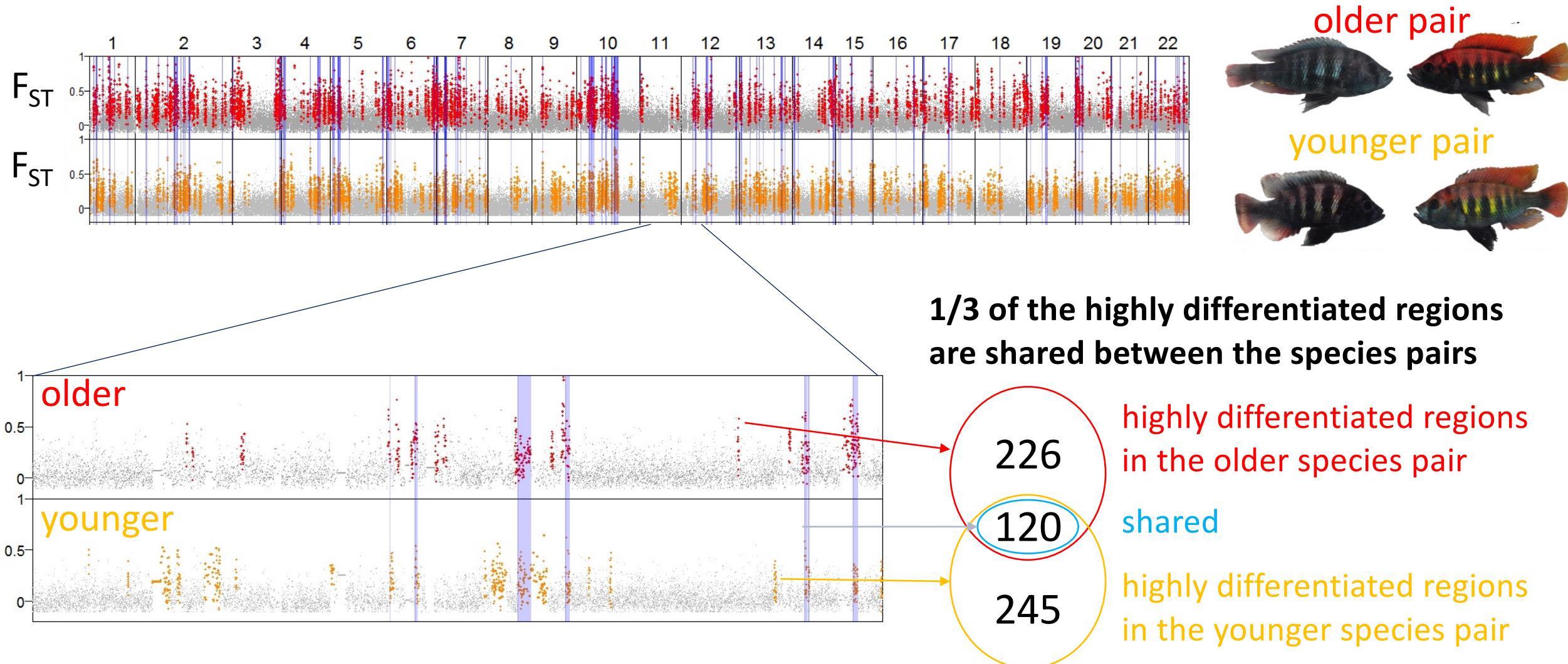
Mwanza Gulf



older original
species pair

younger
species pair of
hybrid origin

Many regions are highly differentiated between the species

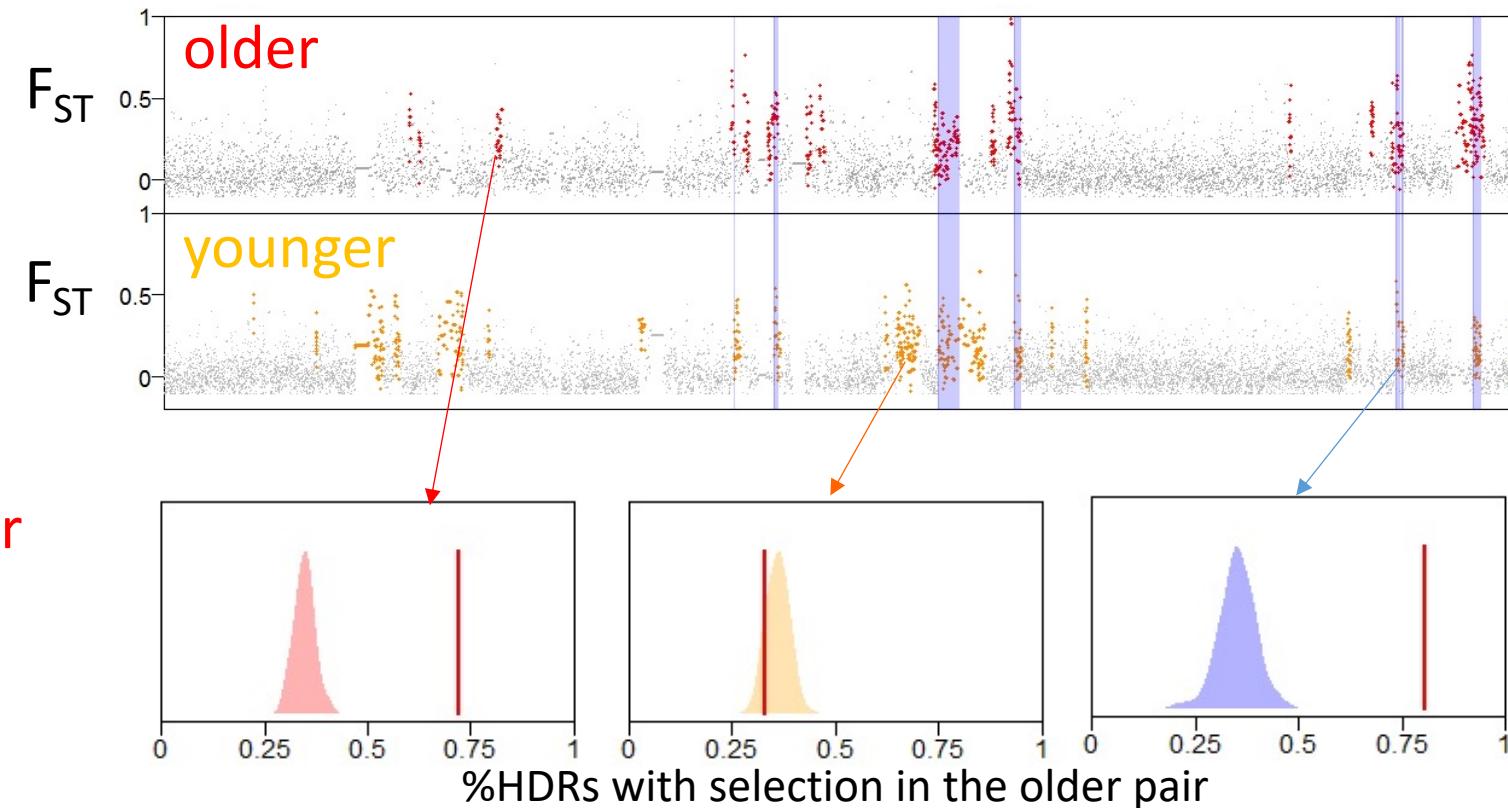


Enrichment of selection statistics support the action of selection

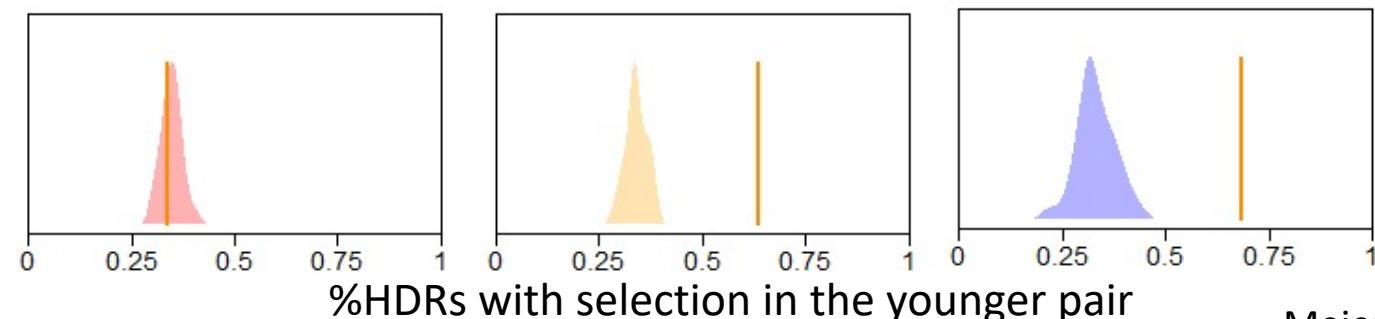
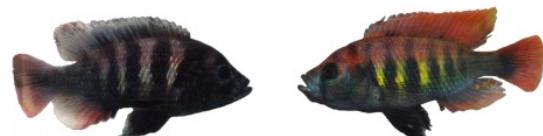
Selection statistics:

d_{xy}
Tajima's D
 $\Delta\pi$
XP-EHH
iHS

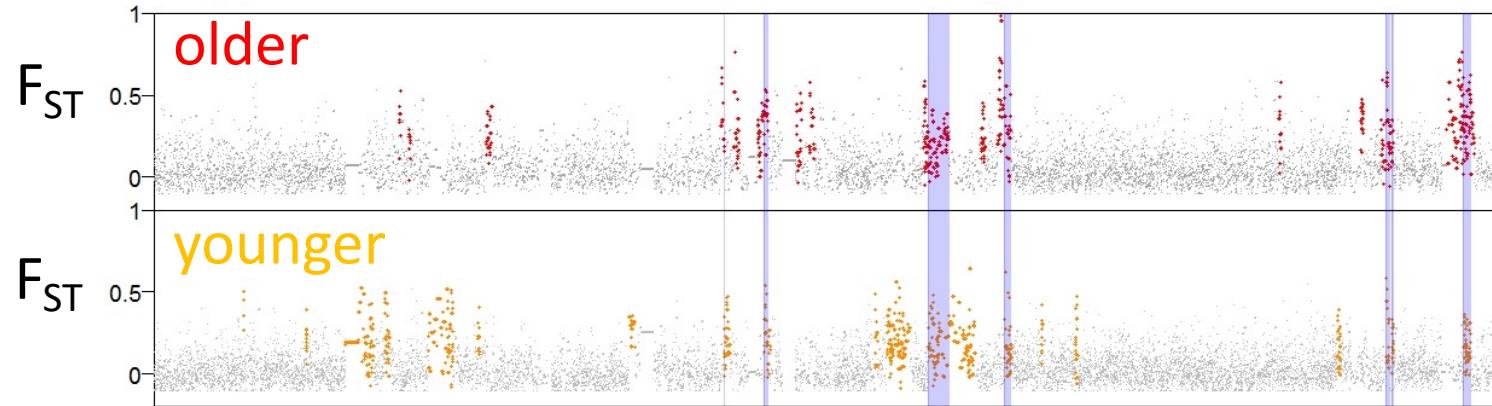
Selection in the older pair



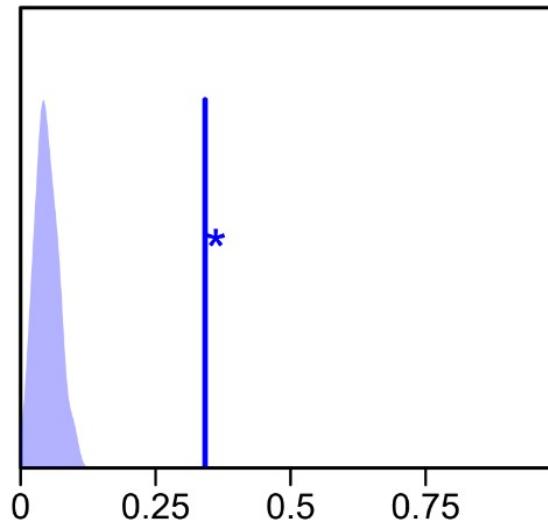
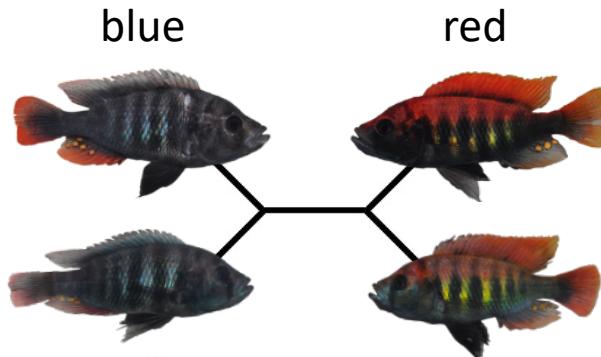
Selection in the younger pair



Highly differentiated regions shared by both species pairs show parallel allele frequency differences



Species group by color



%shared HDRs with color topology

Sorting of admixture variation under parallel selection pressures

