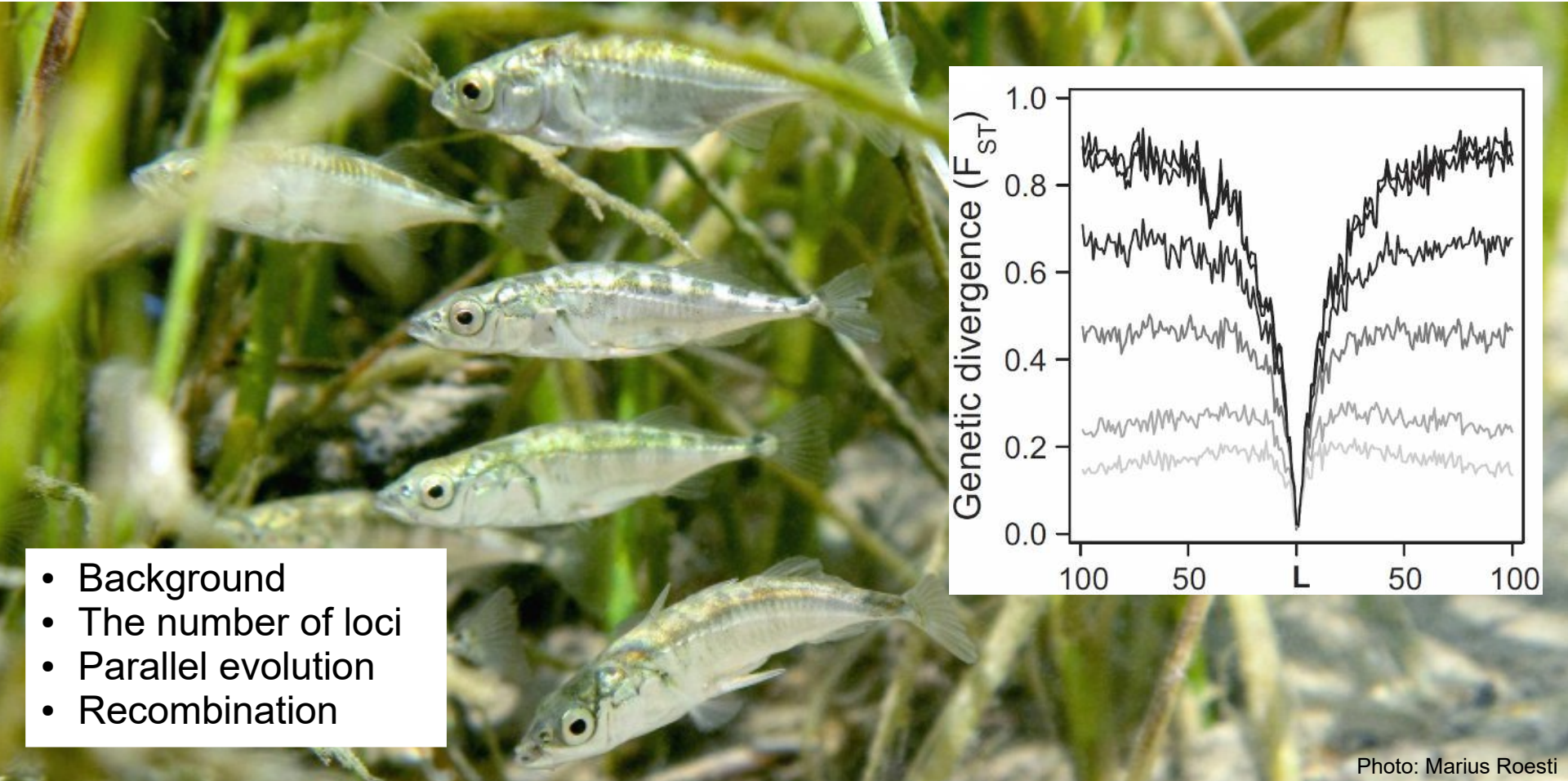


Evolutionary population genomics



- Background
- The number of loci
- Parallel evolution
- Recombination

Photo: Marius Roesti



Daniel Berner, University of Basel
daniel.berner@unibas.ch
http://www.salzburgerlab.org/team/daniel_berner

Cesky Krumlov
2 Feb 2016

Collaborators

Marius Roesti



Dario Moser

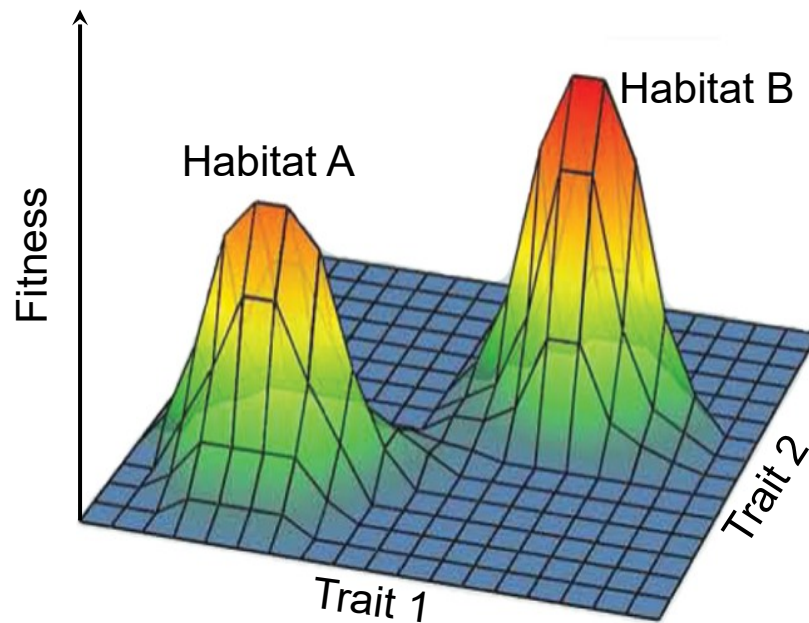


Andrew Hendry
McGill, Montreal

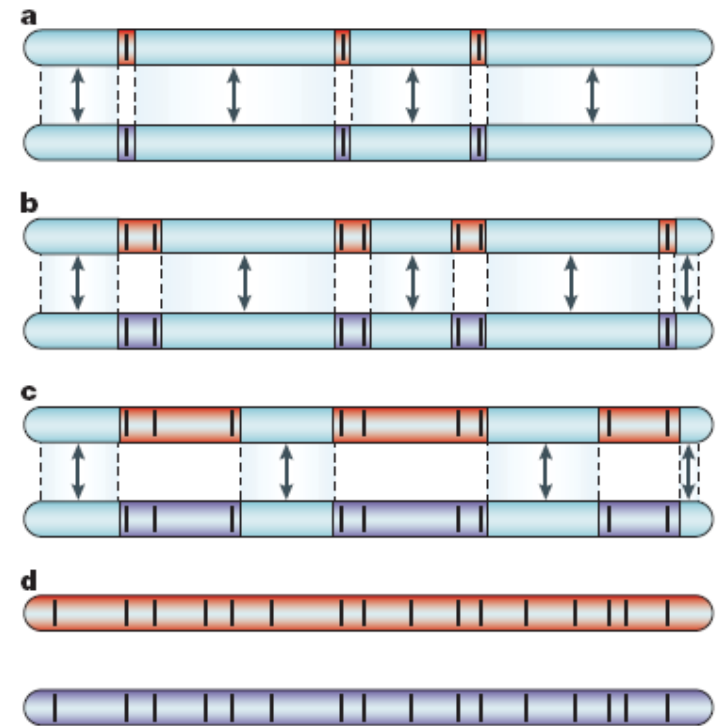
Walter Salzburger
Uni Basel



Focus and questions



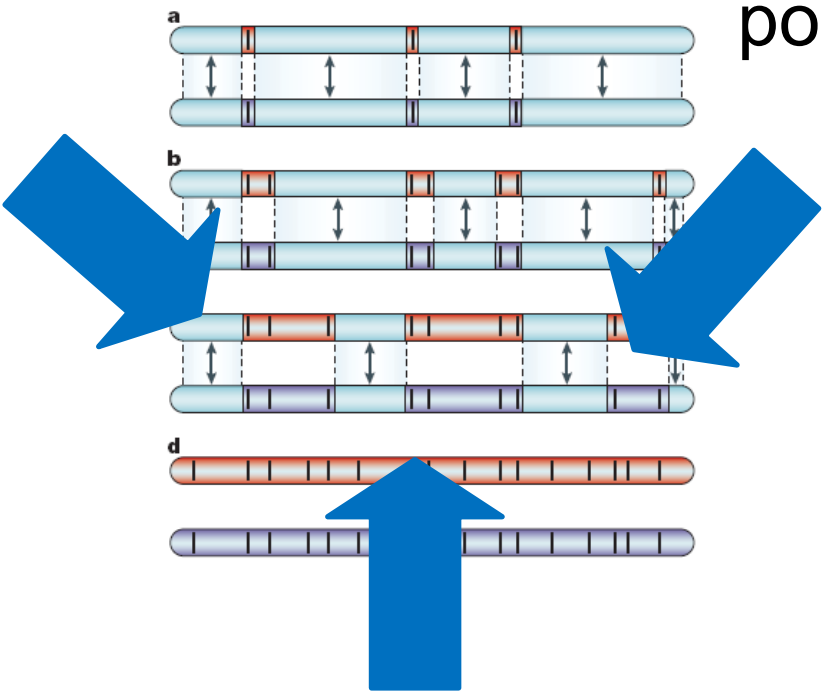
- How many genes underlie adaptive divergence?
- How are they arranged in the genome?
- Where do the alleles come from?
- What is their function?
- What is their effect size?
- To what extent are alleles reused in multiple populations?



Wu & Ting 2004 Nat Rev Genet

Phenotypic
information
from natural
populations
and laboratory
crosses

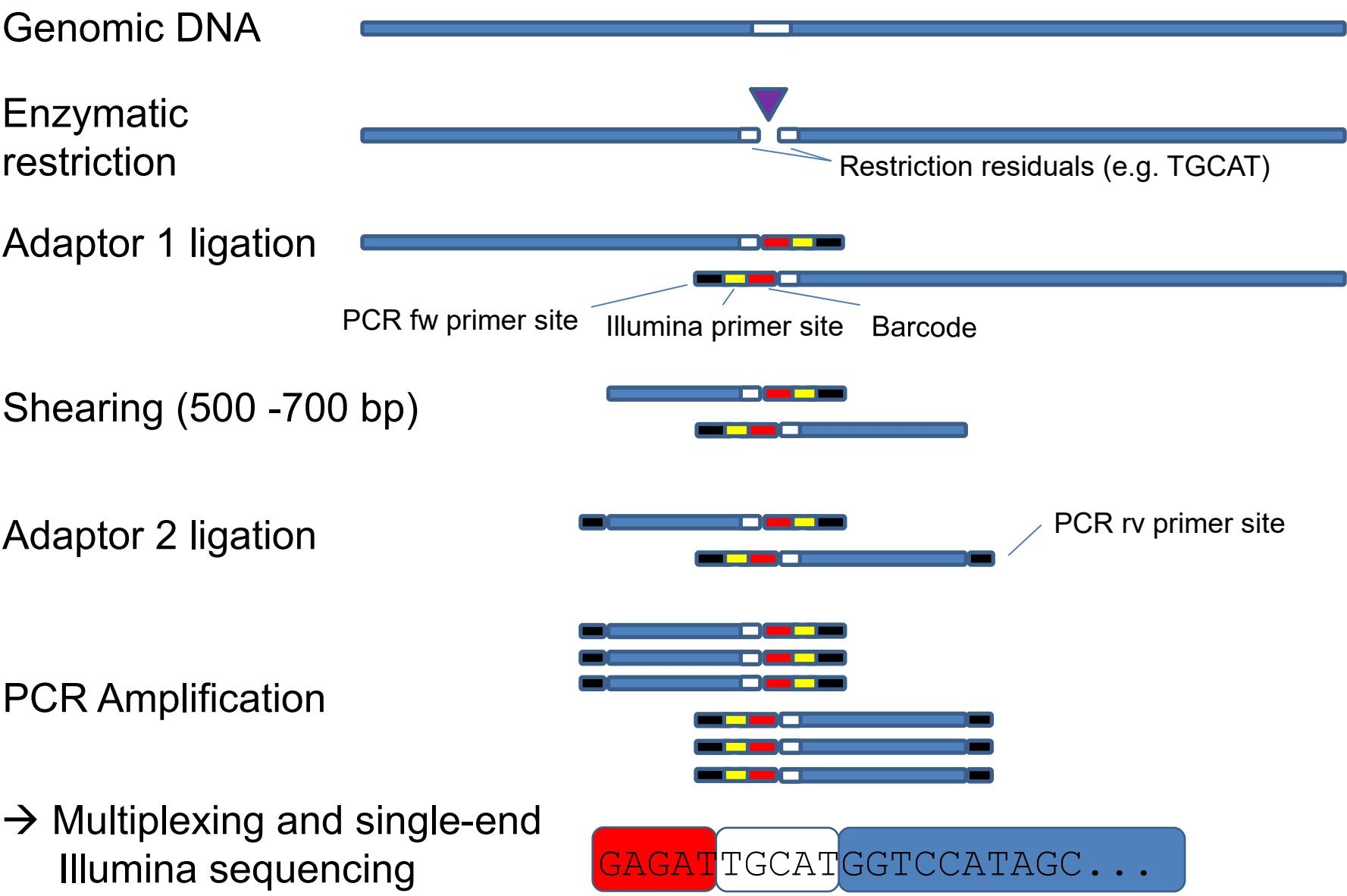
Genome-wide
polymorphism
data



Complexity
reduction
using *in silico*
experiments

RADseq as a tool for *de novo* SNP discovery

RAD = Restriction site-associated DNA





From reads to SNPs

Demultiplexing,
alignment to
reference (typical
coverage: $\geq 40\times$)

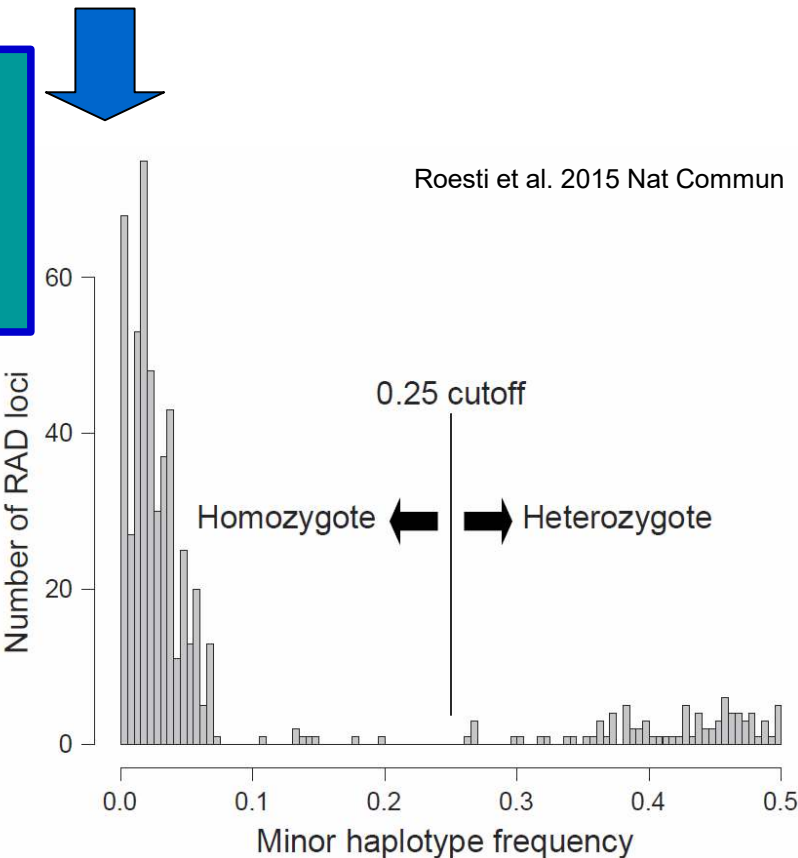
```
TGGCTGCTGCTTTTCTTGCTTTTAGATGGCATAACATTGATACACATTTCTTTTAGATTTTGTG
TGGCTGCTGCTTTTCTTGCTTTTAGATGGCATAACATTGATACACATTTCTTTTAGATTTTGTG
TGGCTGCTGCTTTTCTTGCTTTTAGATGGCATAACATTGATACACATTTCTTTTAGATTTTGTG
TGGCTGCTGCTTTTCTTGCTTTTAGATGGCATAACATTGATACACATTTCTTTTAGATTTTGTG
TGGCTGCTGCTTTTCTTGCTTTTAGATGGCATAACATTGATACACATTTCTTTTAGATTTTGTG
```

Consensus
genotyping for
each
individual



Keep
It
Simple and
Stupid

Robust RAD genotyping based
on the distribution of whole
haplotypes



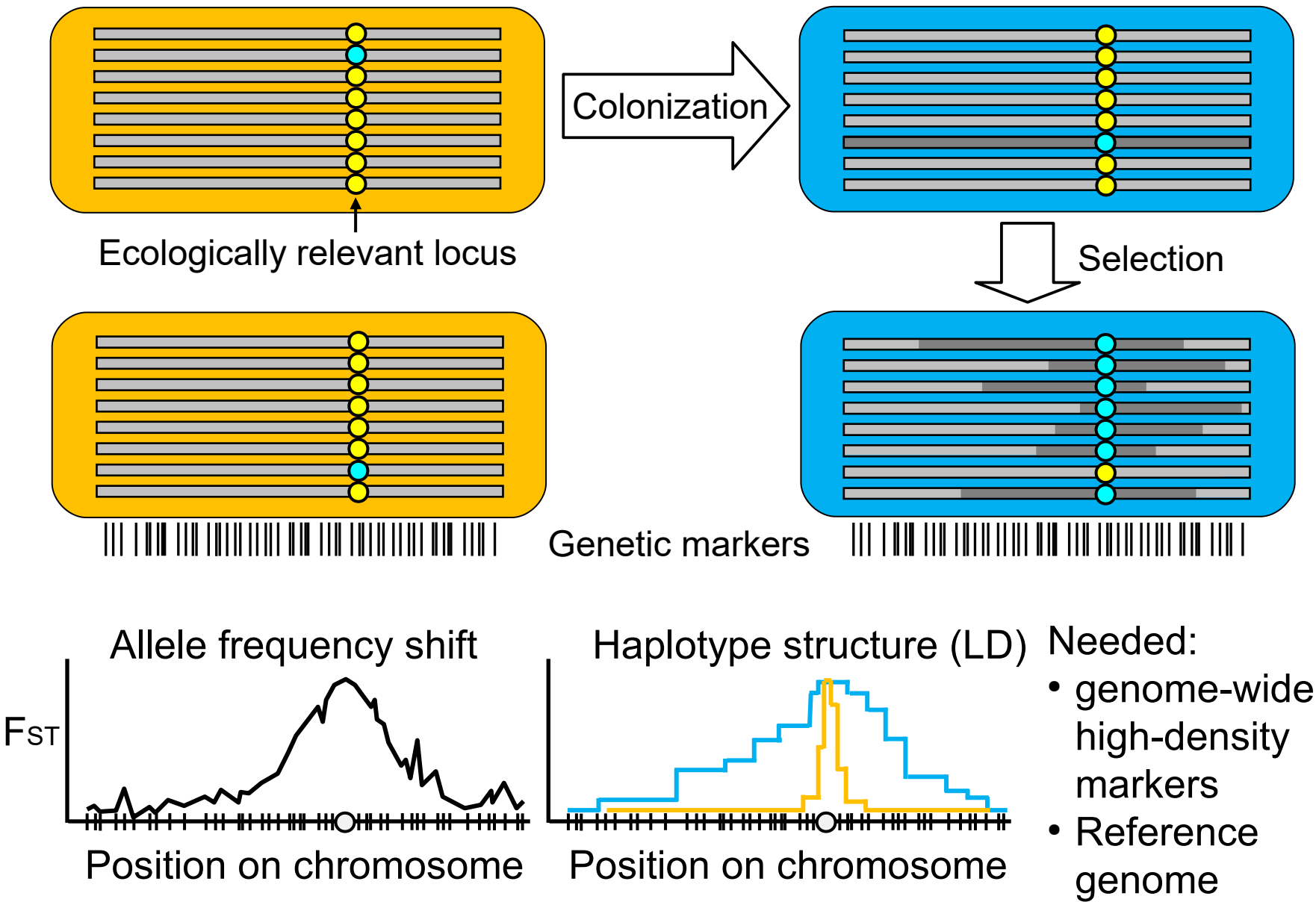
SNP calling

Pop1

Pop2



Basic logic of SNP-based genome scans for selection



How many genes underlie adaptive divergence?



Threespine stickleback

- Ancestrally marine
- Fossil record back to 13 My
- Colonization of freshwater, mainly postglacial
- Repeated and predictable divergence from marine ancestor



Truckee formation, Nevada (USA), 10my

Lake-stream divergence in stickleback

- Parallel evolution - lake-stream divergence is replicated many times



Berner et al. 2010
Mol Ecol

- Predictable niche differentiation (pelagic vs. benthic prey)



Lake



Stream

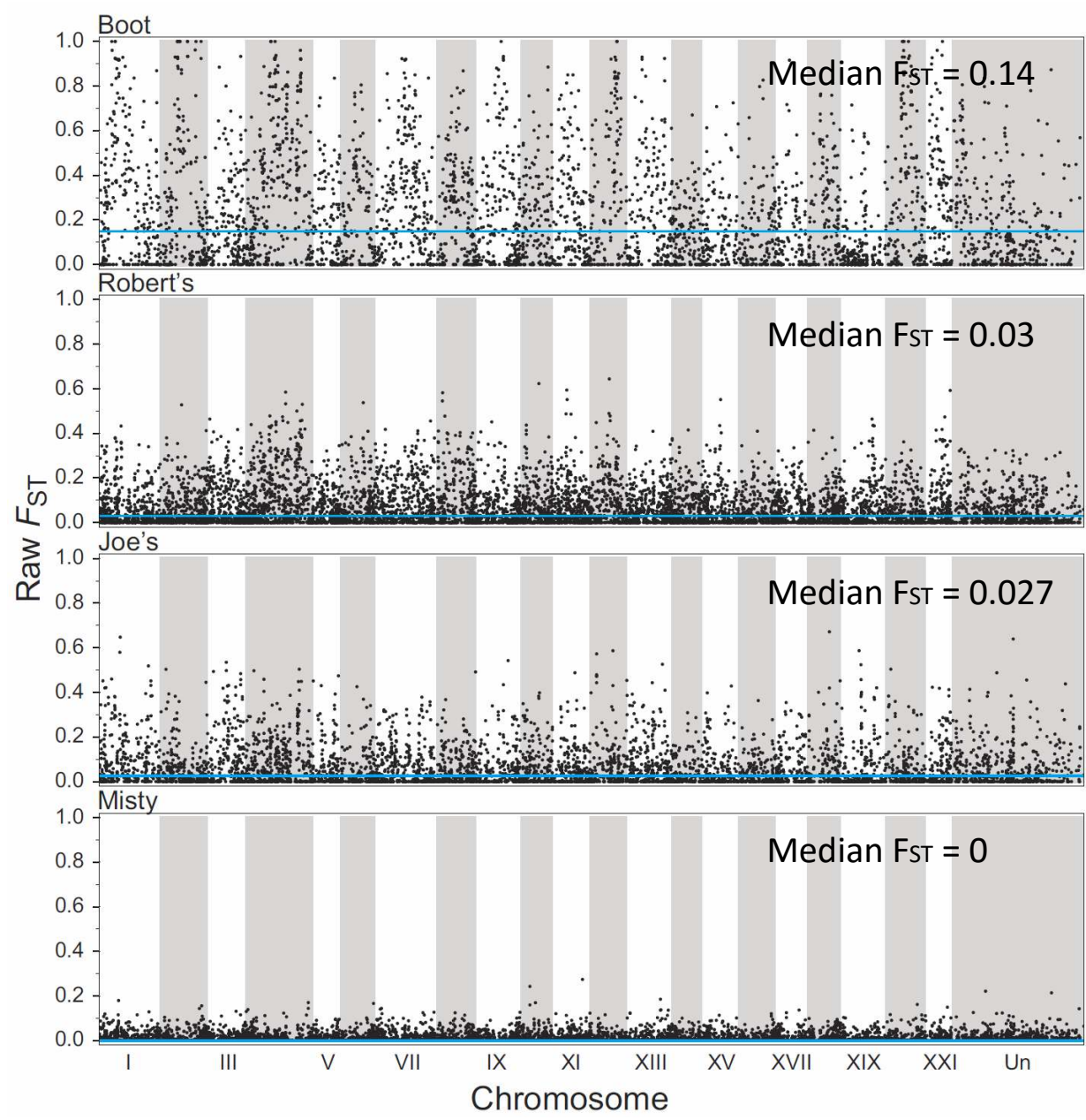


Early divergence mapping in lake-stream stickleback Roesti et al. 2012 Mol Ecol

- 4 lake-stream population pairs from Canada, 27 individuals from each habitat
- SNPs generated by RADseq (Sbf1 cutter, c. 20k restriction sites across the genome)
- Calculation of F_{ST} at all SNPs

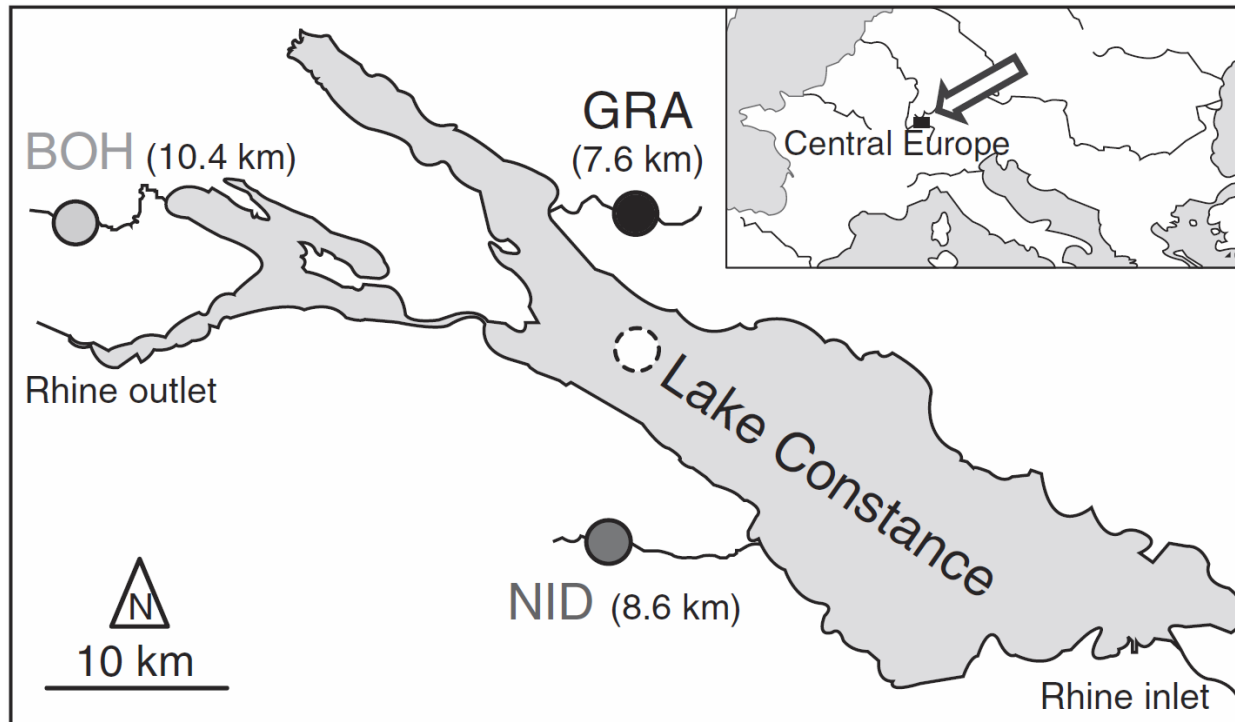


Chapter 1



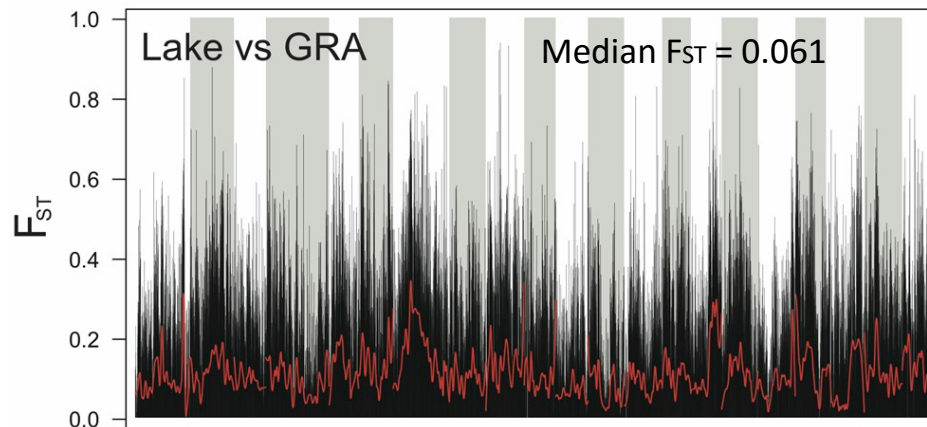
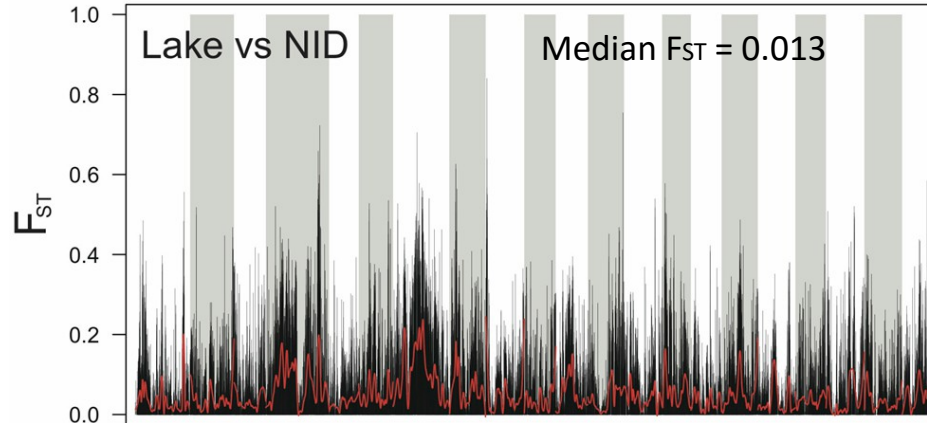
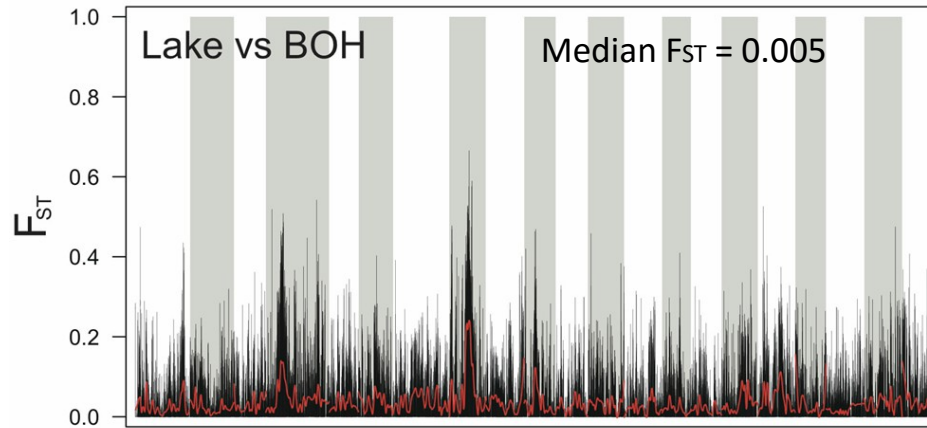
- Variable magnitude of baseline genomic differentiation
- Divergence is heterogeneous across the genome
- Very many loci seem to be under divergent selection

Higher-resolution divergence mapping Roesti et al. 2015 Nat Commun



- Lake Constance and 3 stream populations, 22-25 individuals from each habitat
- SNPs generated by RADseq (NsiI cutter, c. 140k restriction sites across the genome)
- Calculation of F_{ST} at all SNPs

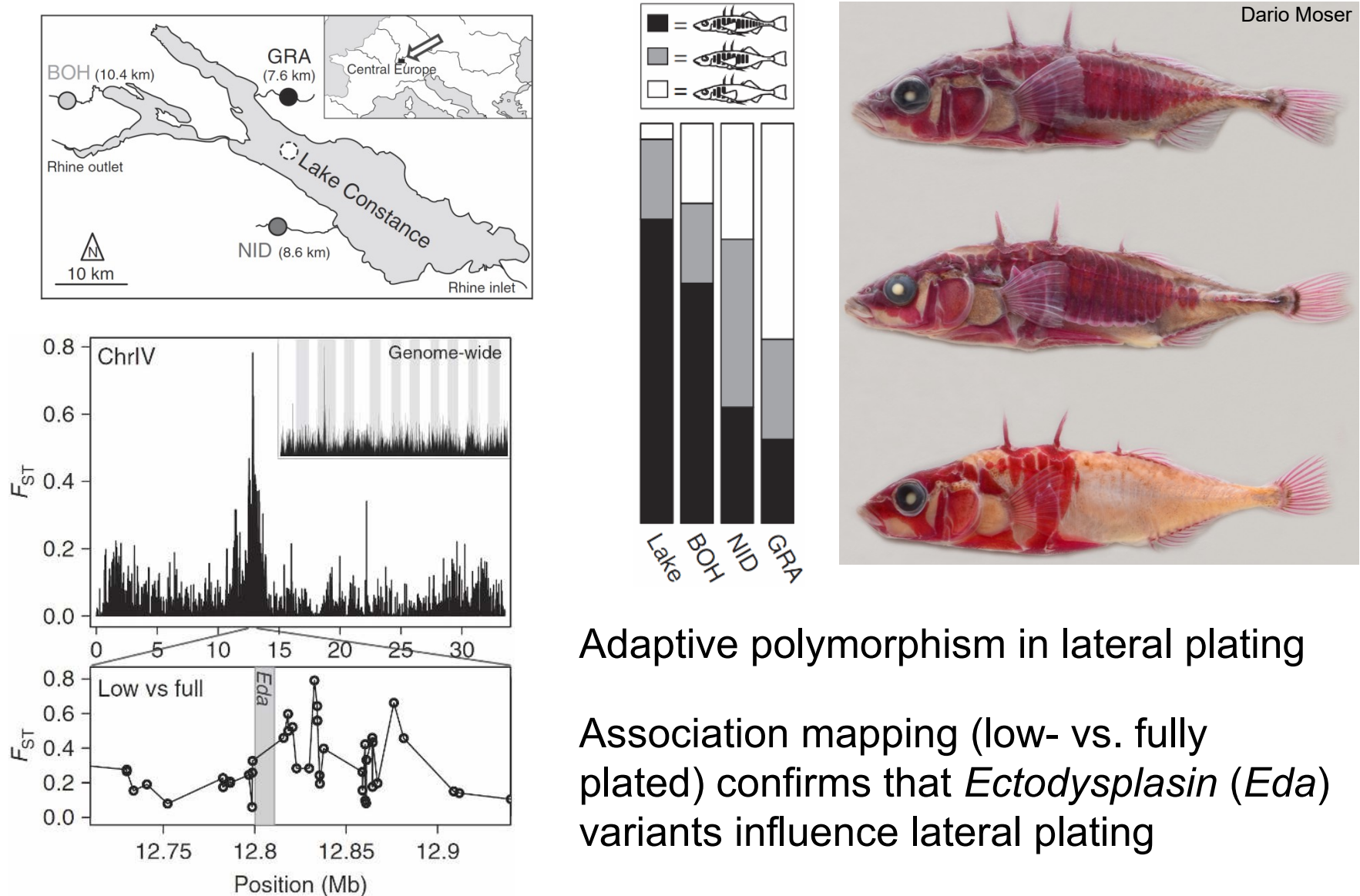
Chapter 1



Similar findings:
variable magnitude of baseline
genomic differentiation; highly
polygenic, heterogeneous
divergence

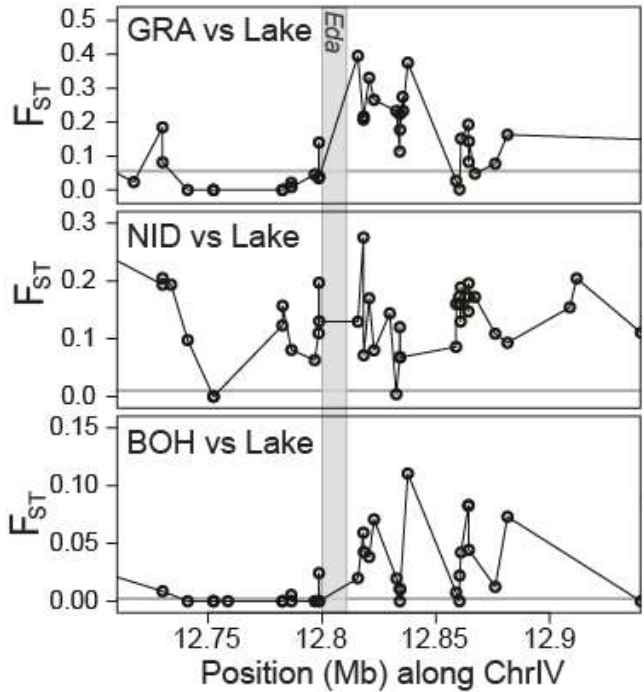
Support for highly polygenic selection offered by the *Eda* locus

Roesti et al. 2015 Nat Commun



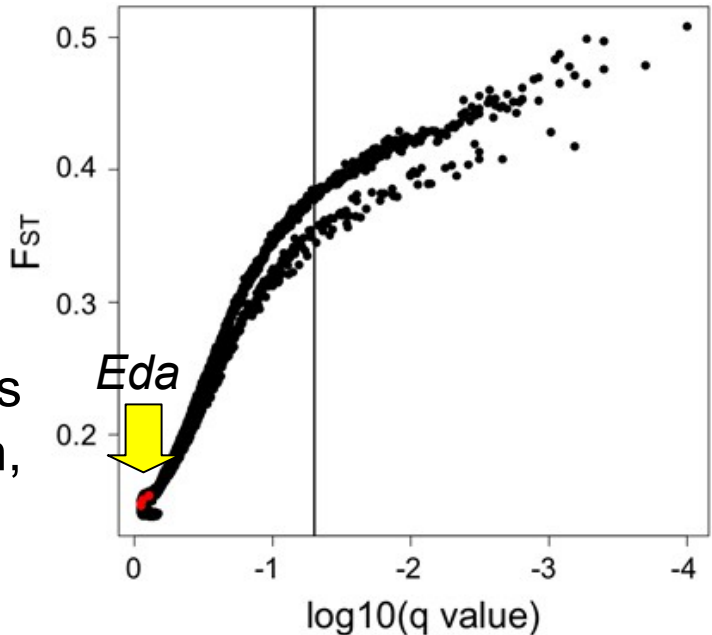
Adaptive polymorphism in lateral plating

Association mapping (low- vs. fully plated) confirms that *Ectodysplasin* (*Eda*) variants influence lateral plating

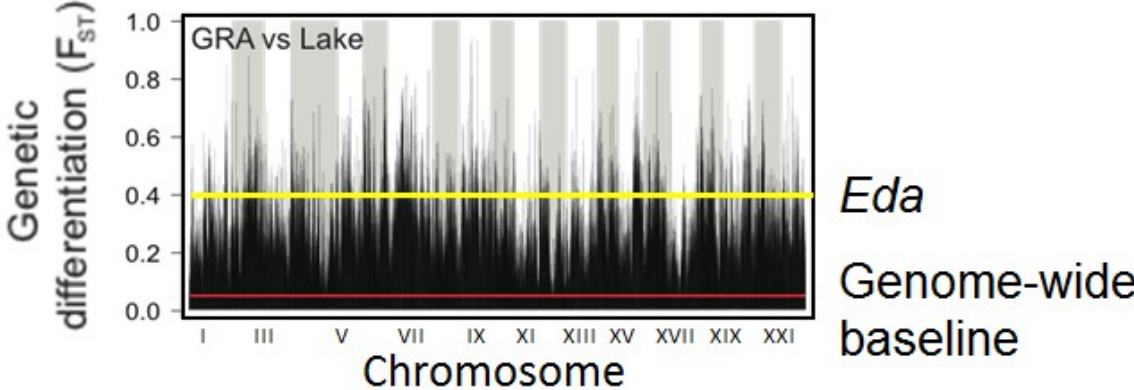


F_{ST} profiles around *Eda*, based on the populations

‘Outlier’ analysis using BayeScan, GRA vs. lake



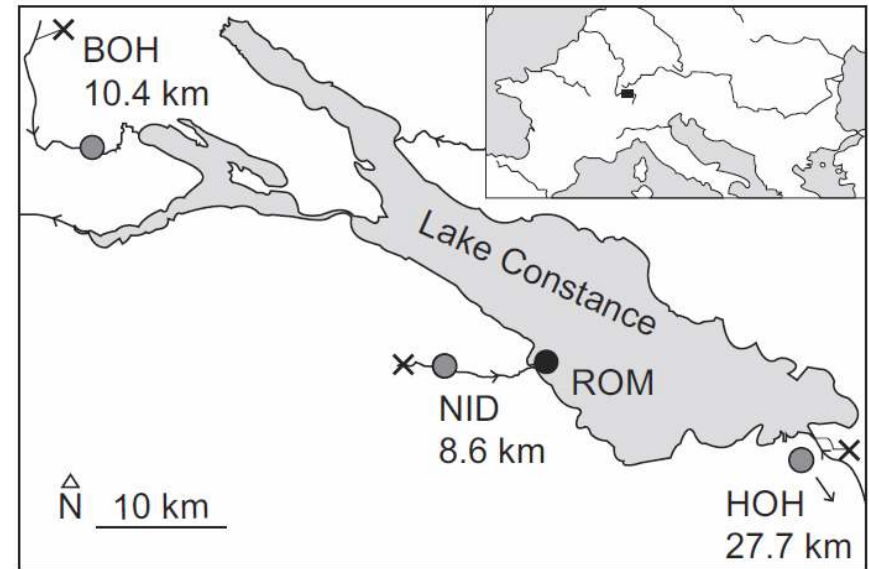
F_{ST} genome-wide, GRA vs Lake

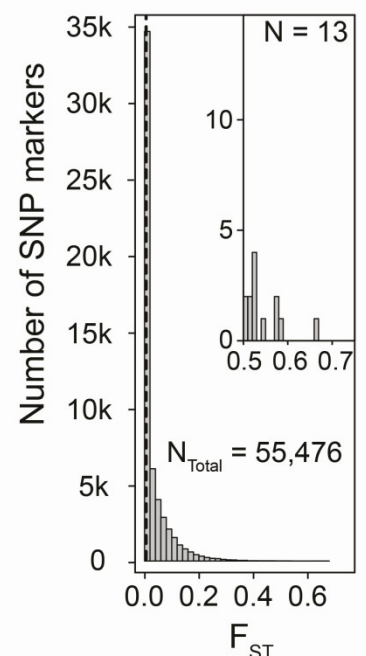
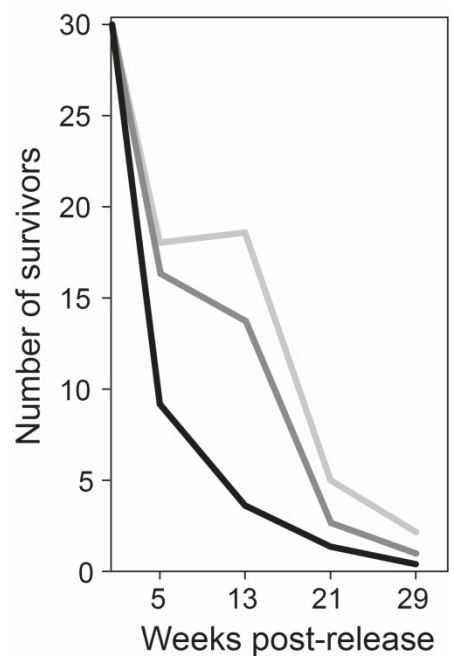
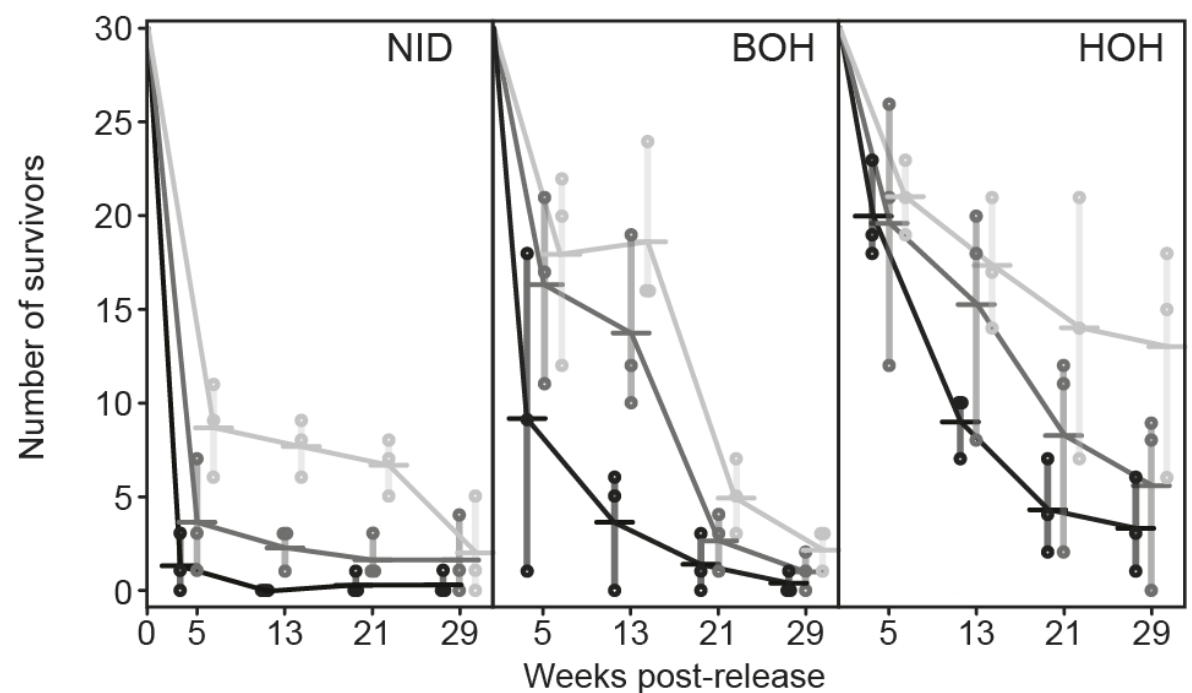


- Hundreds of loci are involved in lake-stream divergence
- Detection issue...

What is the fitness correlate of genomic divergence? Moser et al. 2016 J Evol Biol

- Field enclosure experiment (3x3)
- Stocking with 90 lab-reared, marked individuals (lake, stream, F1 hybrid)
- Tracking of survival by recapture over 29 weeks

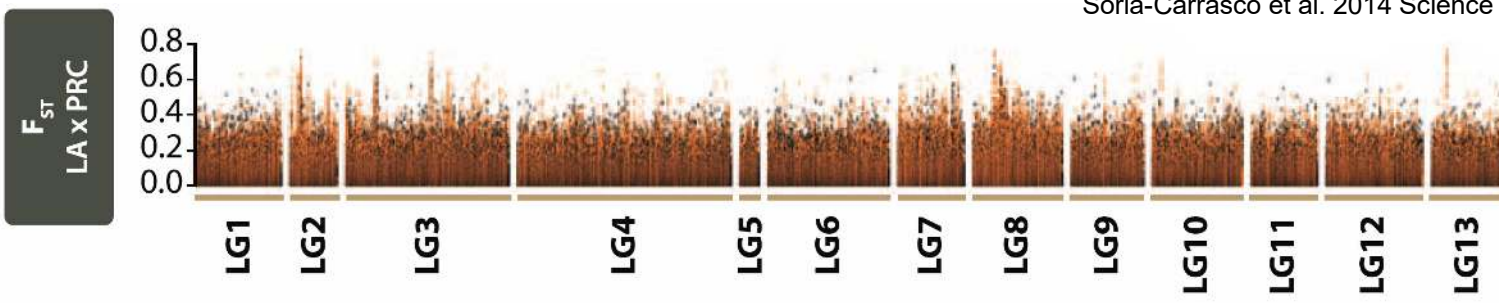
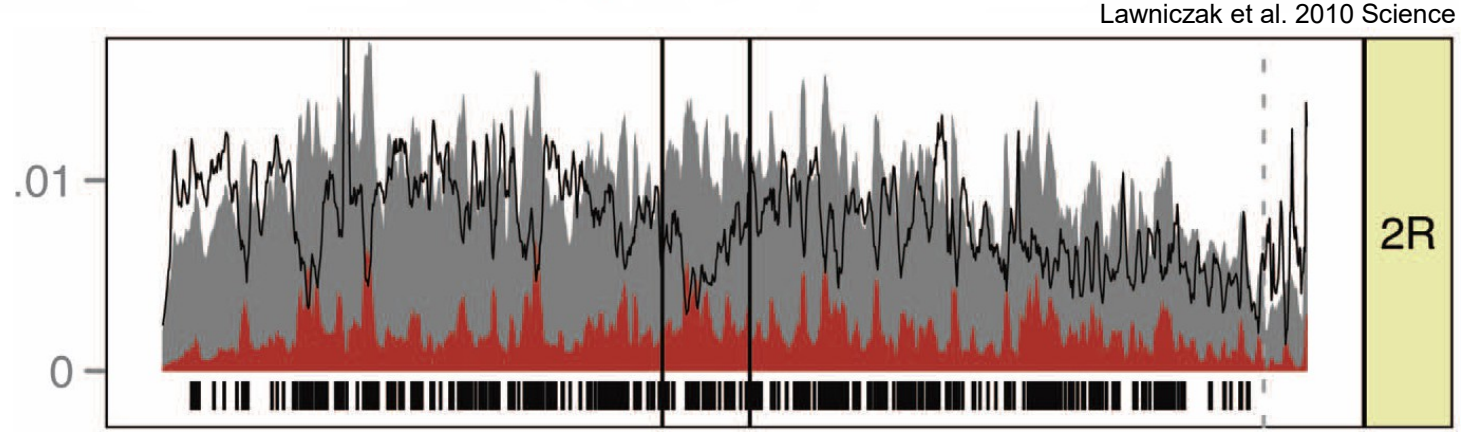
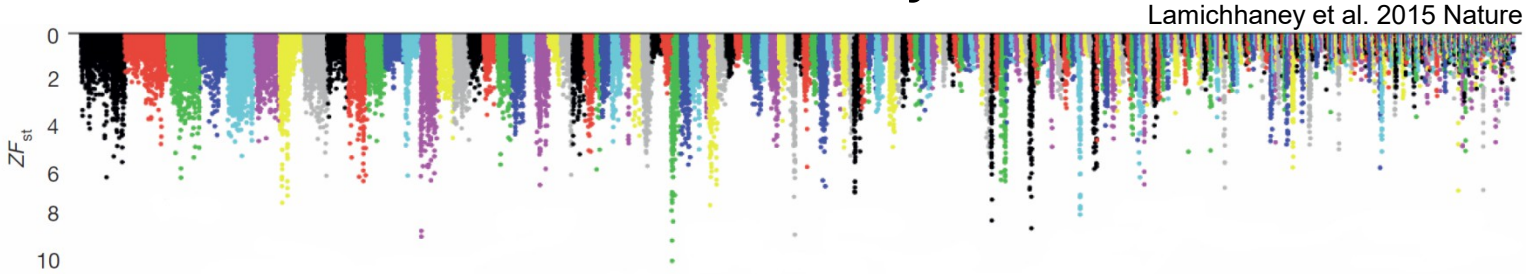




Strong fitness differences despite weak baseline genomic differentiation

Modest allele frequency differences at many ecologically important loci allow strong adaptive divergence

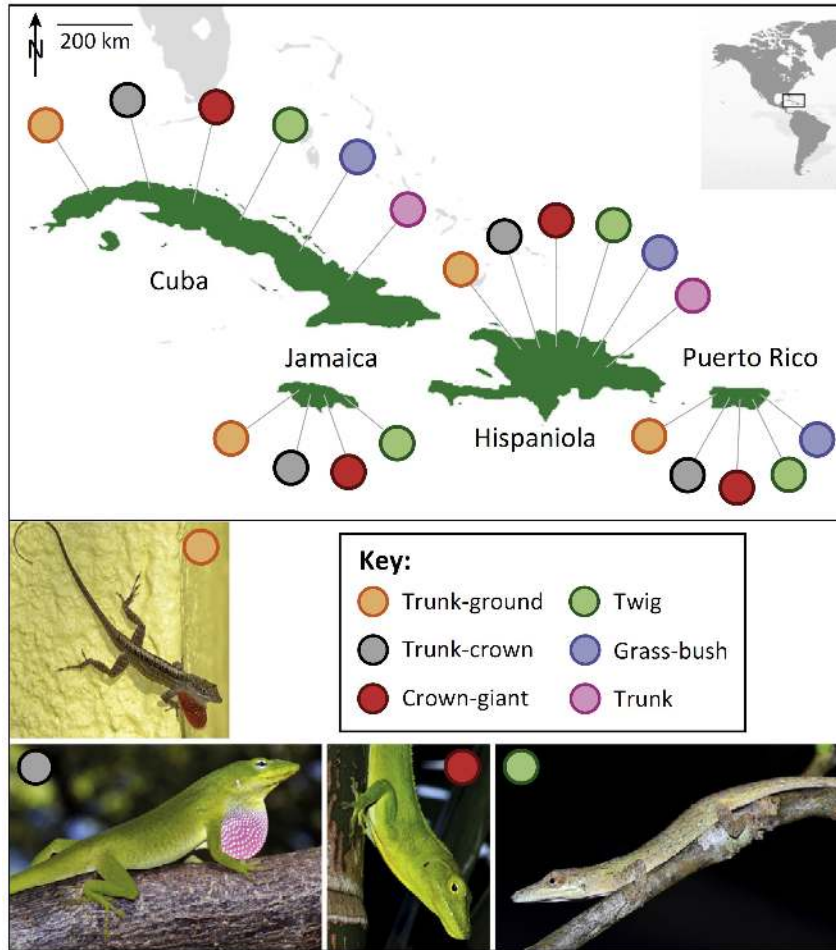
Evidence on the number of loci from other systems



Conclusions

- Evolutionary population genomics benefits from integrative research approaches
- It does not always have to be full-genome sequencing
- A robust genome assembly is the key to informative genomic investigations
- Divergence into different ecological niches generally involves a great number of allele frequency shifts
- We should abandon the idea that we can count or characterize these shifts comprehensively

Parallel evolution



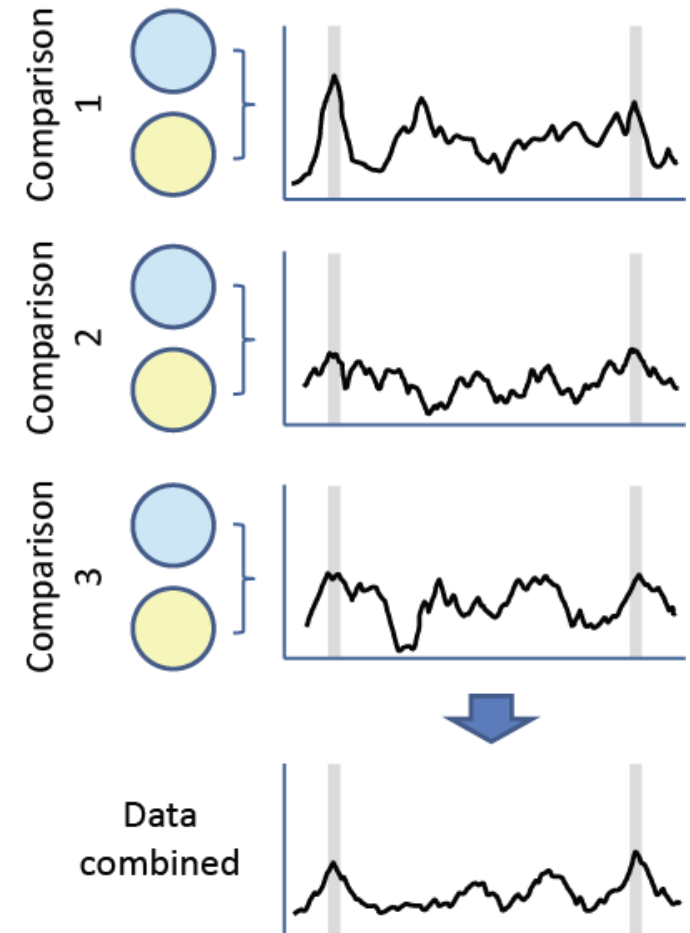
Berner & Salzburger 2015 Trends Genet

- Recurrent evolution of habitat-associated phenotypes
- Deterministic outcome of natural selection
- Focus initially phenotypic, now often genetic

1. Searching for adaptation loci using parallel evolution in Canadian lake-stream stickleback pairs

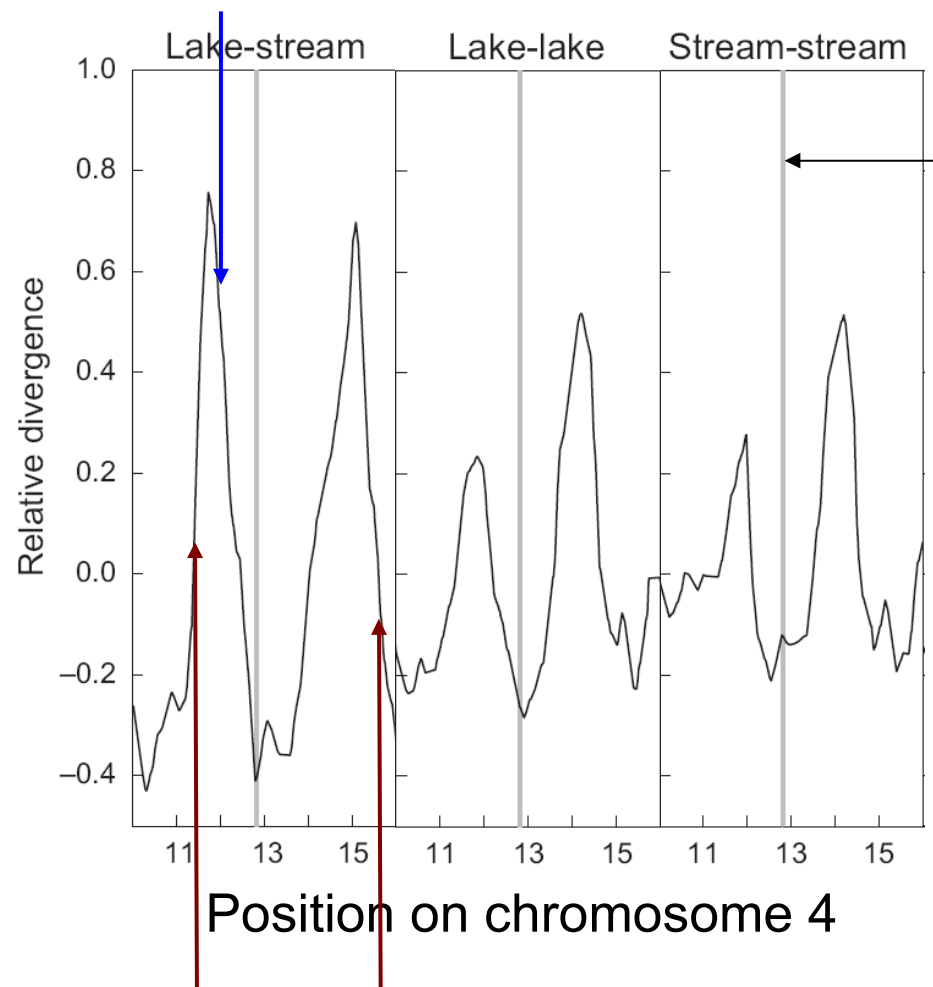
Roesti et al. 2012, 2014 Mol Ecol

Logic: screening for high-differentiation genome regions shared among the replicate lake-stream pairs



Other lake-stream genome scan

Deagle et al. 2012 Proc R Soc B

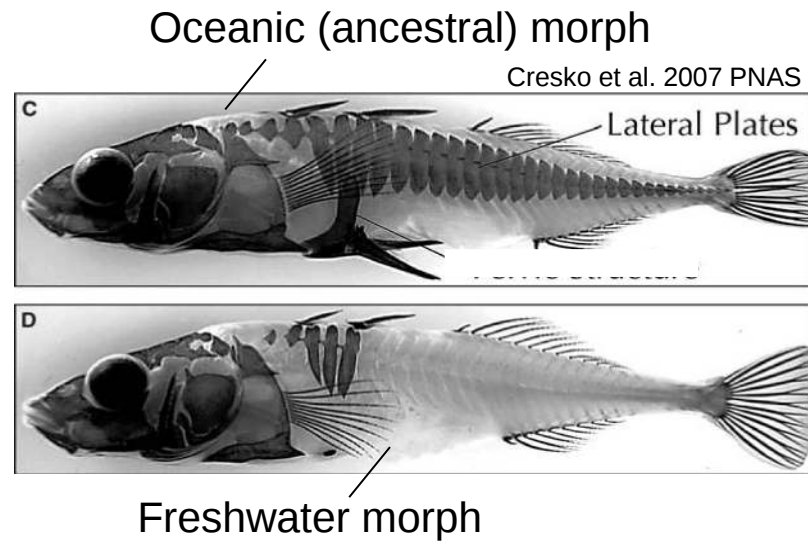


Benthic-pelagic genome scan

Jones et al. 2012 Curr Biol

Can we be certain that F_{ST} peaks reflect loci under parallel divergent selection?

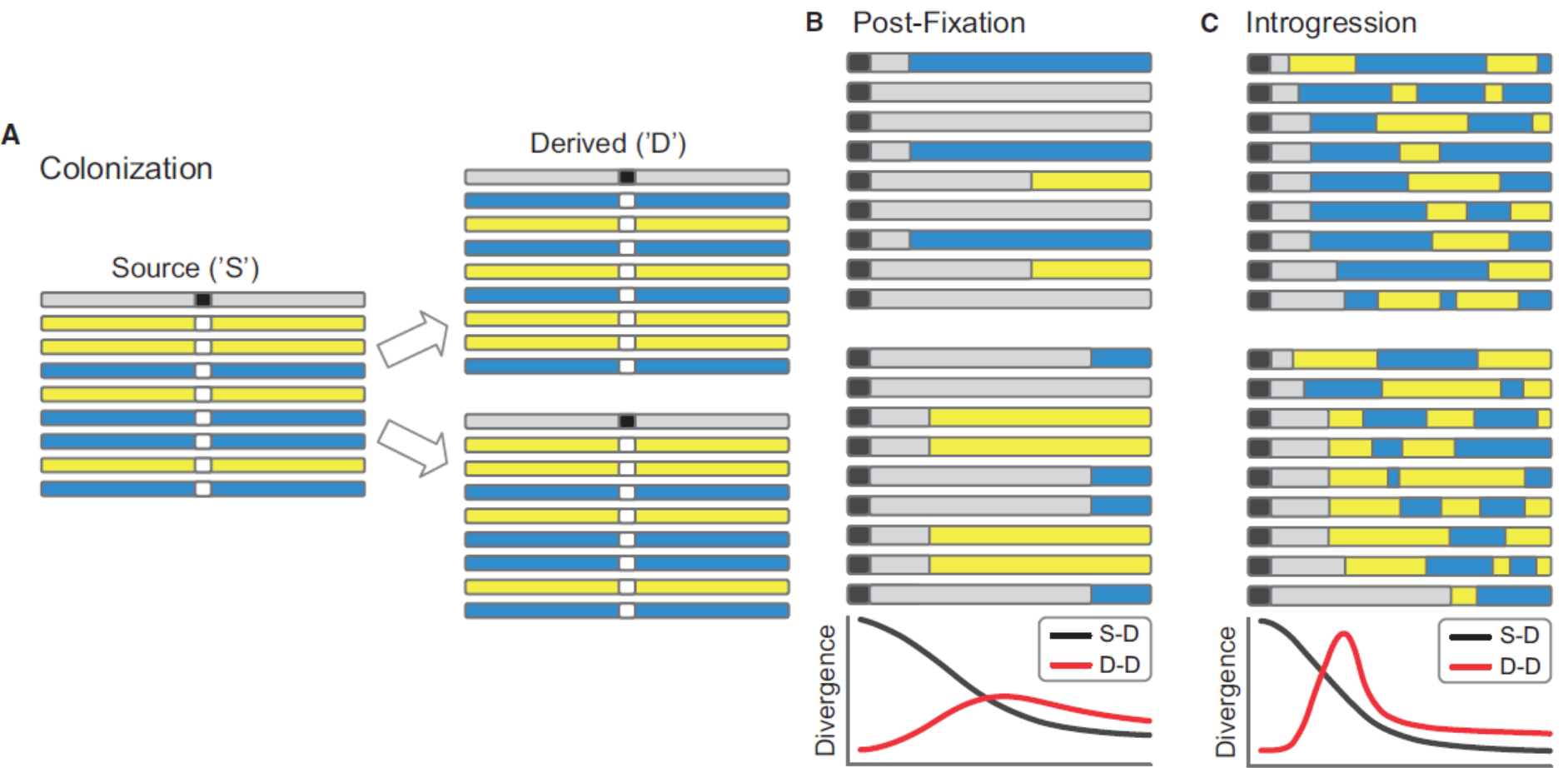
Ectodysplasin (*Eda*) gene



Freshwater populations share the low-plated phenotype, driven by *Eda* variants

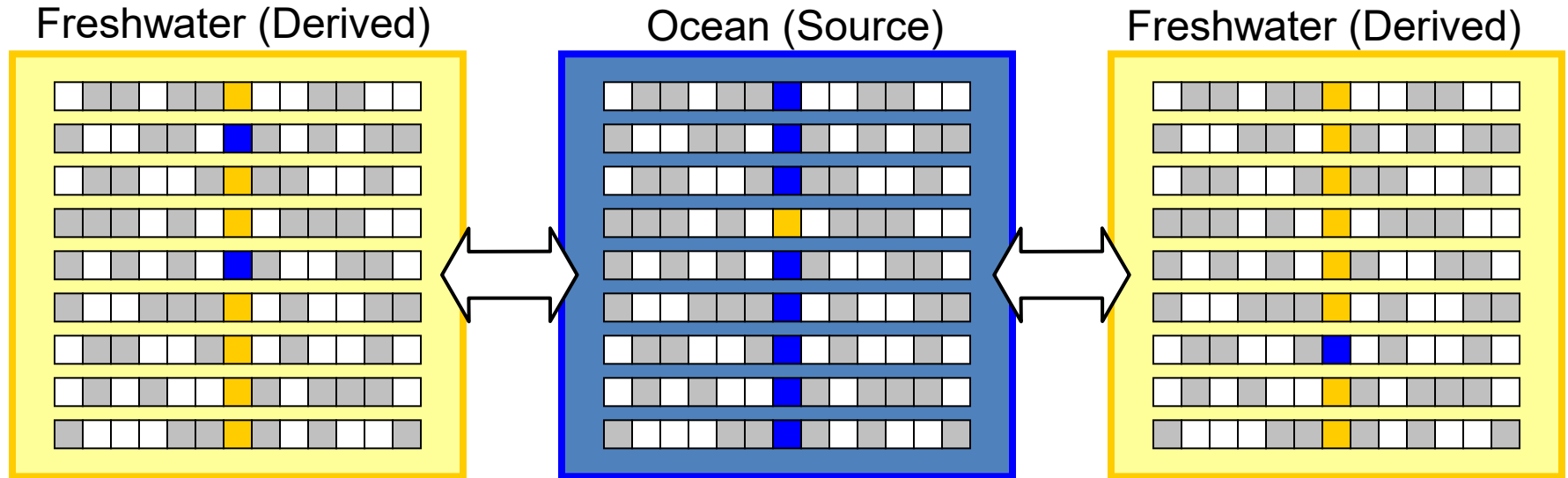
Hypothesis

Peak-valley-peak pattern among derived populations is driven by repeated adaptation from shared genetic variation at a single locus

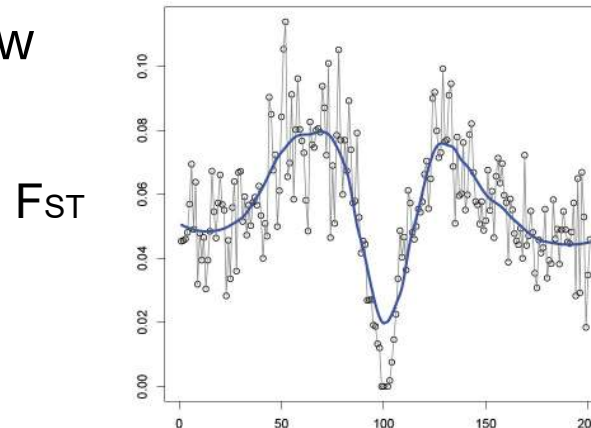


Simulation study

- Multiple independent colonization of new habitats from ancestral source with standing variation at selected locus



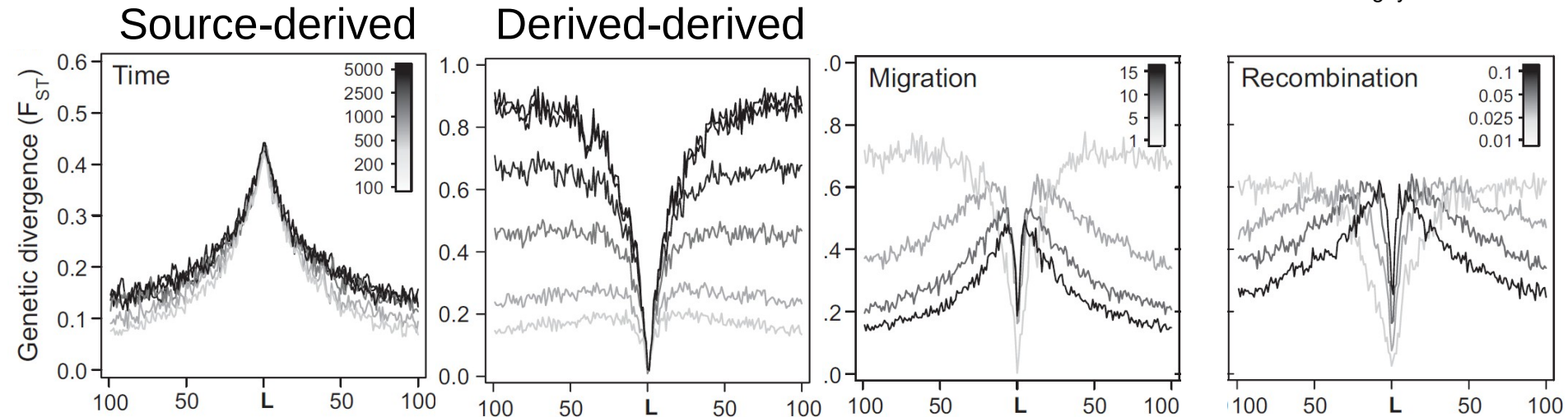
- Adaptation with gene flow



- Parameters explored: time, N colonizers, carrying capacity, migration rate, selection strength, recombination rate



Sergey Gavrillets

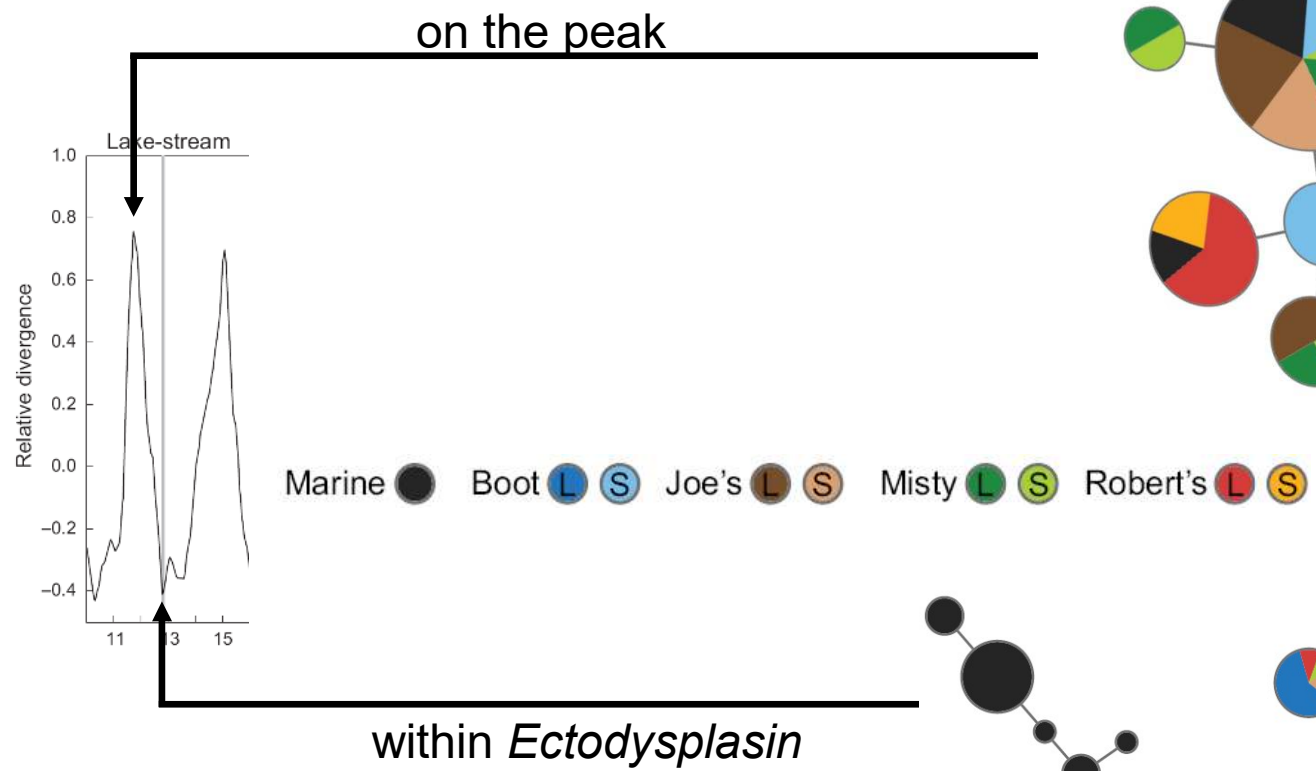


The selected locus acts as a localized barrier to gene flow

Source-derived contrasts reveal the classical signature of divergence with gene flow

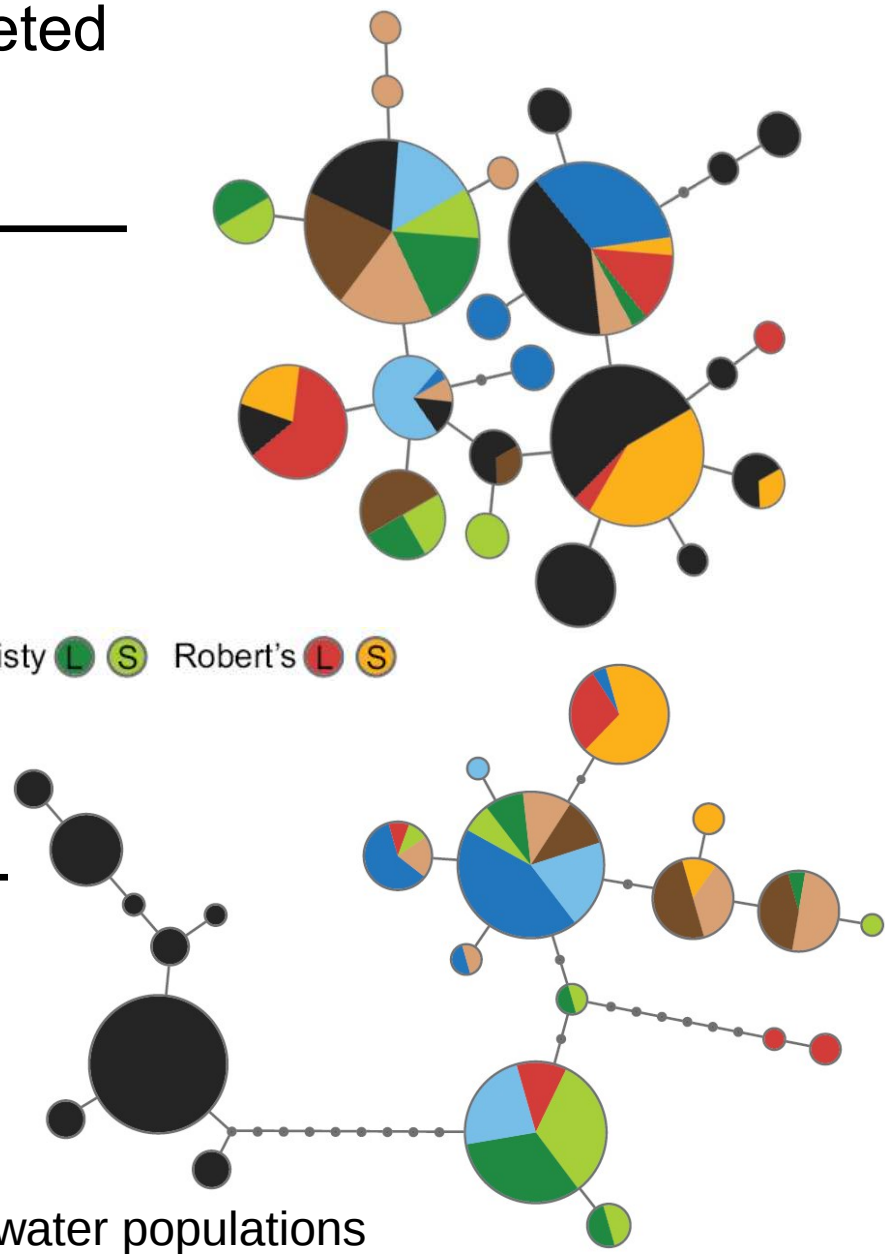
Derived-derived contrasts: repeated adaptation from shared variation drives a peak-valley-peak pattern

Confirming the model through targeted sequencing

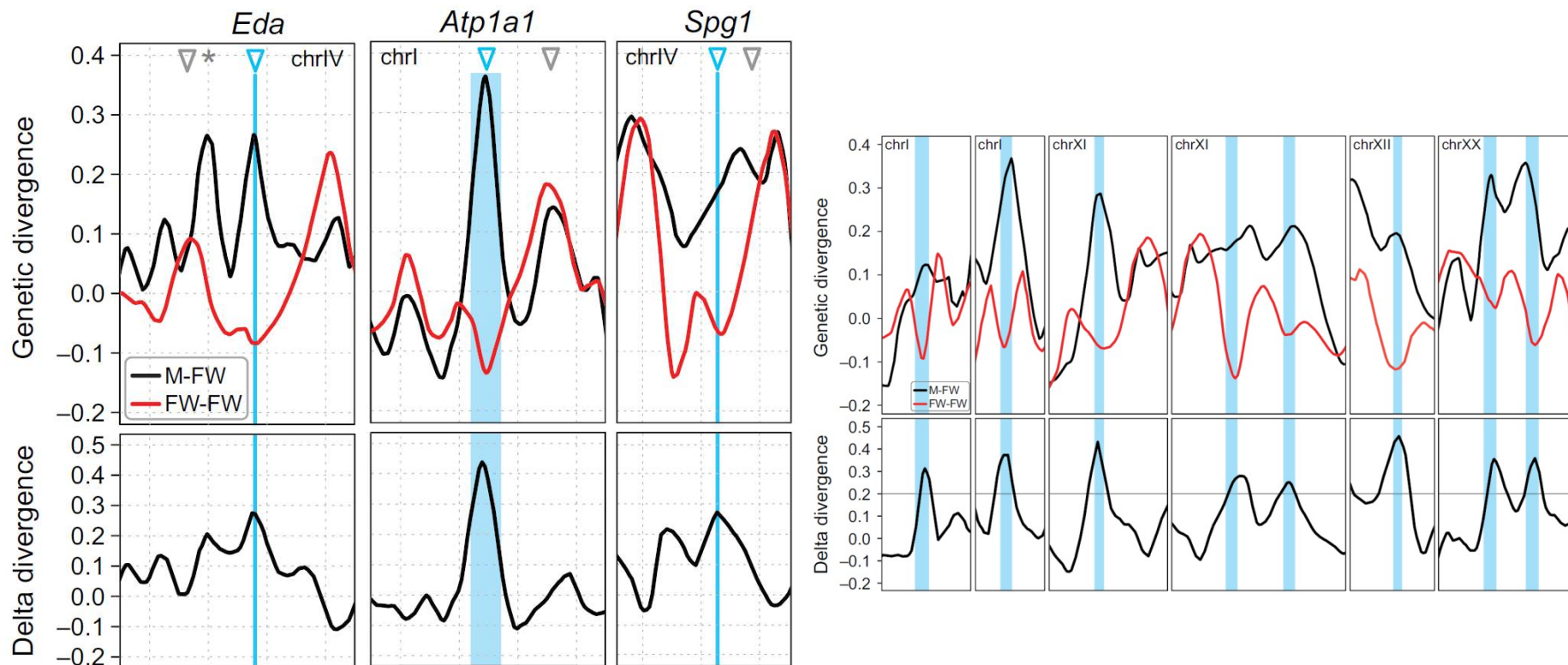


Complete marine-freshwater isolation at *Eda* but not away from it – localized barrier to gene flow

Extensive haplotype sharing among the freshwater populations – adaptation from shared genetic variation



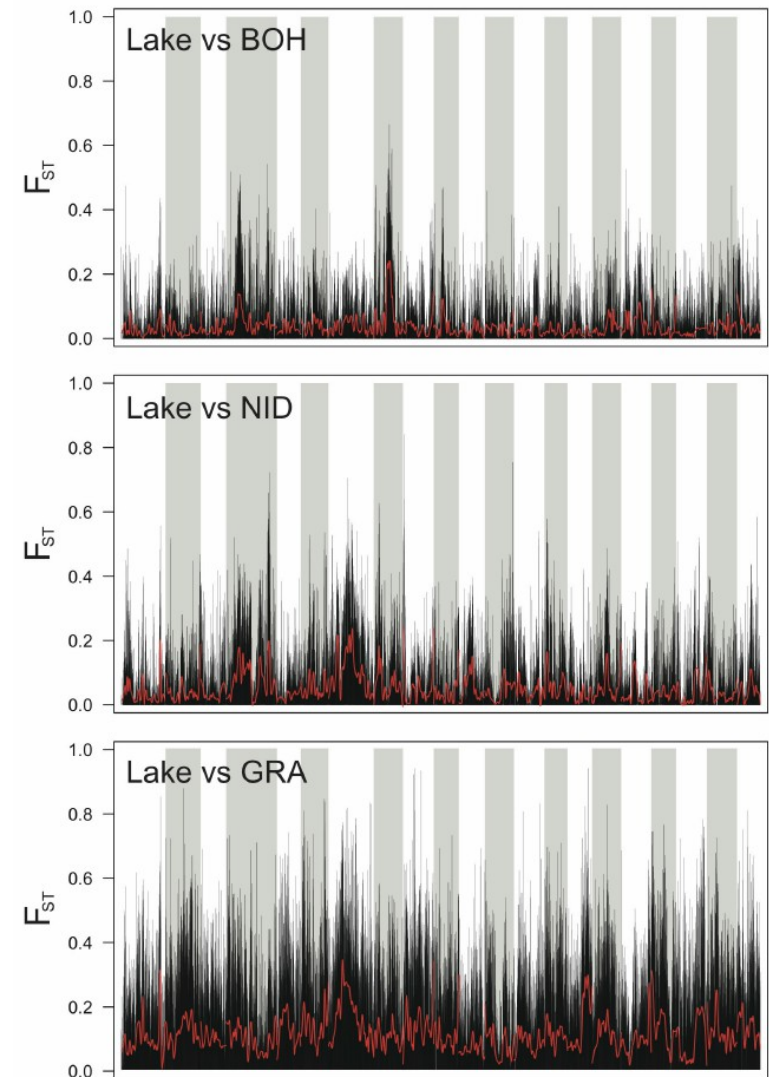
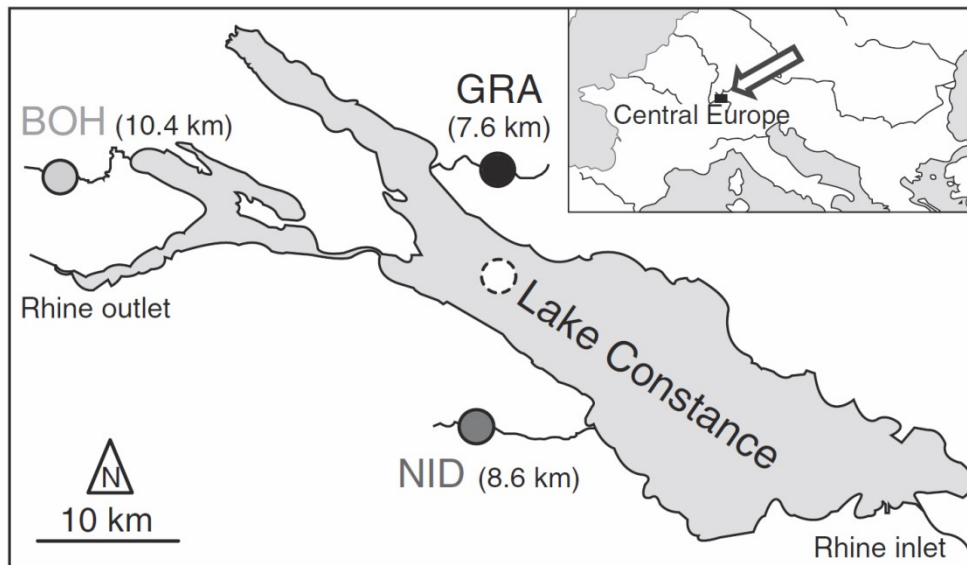
The peak-valley-peak signature as a discovery tool



- Marine-freshwater divergence is highly polygenic and involves extensive allele recycling
- Derived freshwater populations still carry the footprints of the ancestral marine-freshwater divergence
- Parallel evolution is useful for the detection of adaptation loci – but one must look for the right signature at the right ecological level

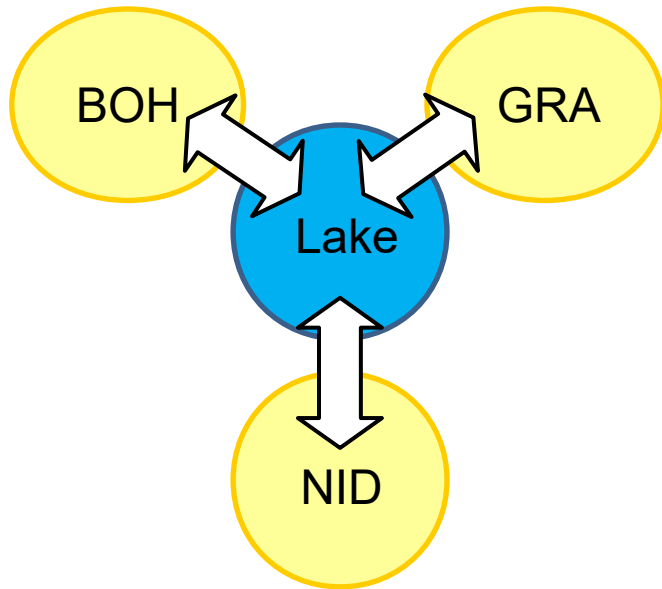
2. Searching for adaptation loci using parallel evolution in lake-stream stickleback from the Lake Constance region Roesti et al. 2015 Nat Commun

Logic: screening for high-differentiation genome regions shared across multiple lake-stream contrasts



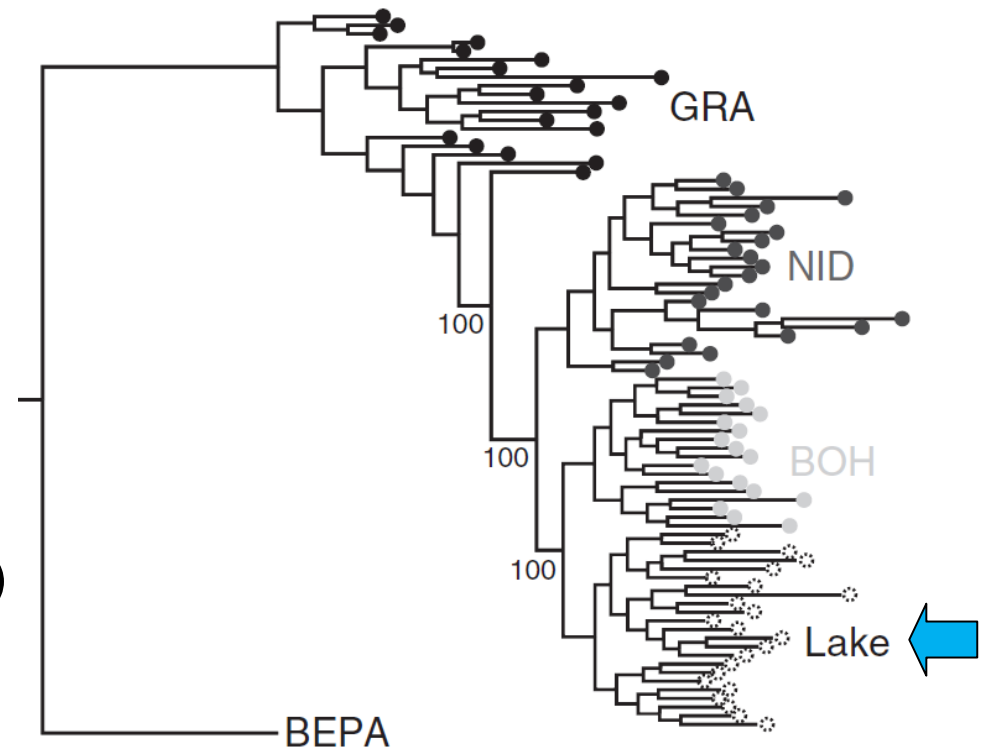
Assumptions ...

- The lake population is the ancestor
- Stream adaptation has occurred three times in parallel

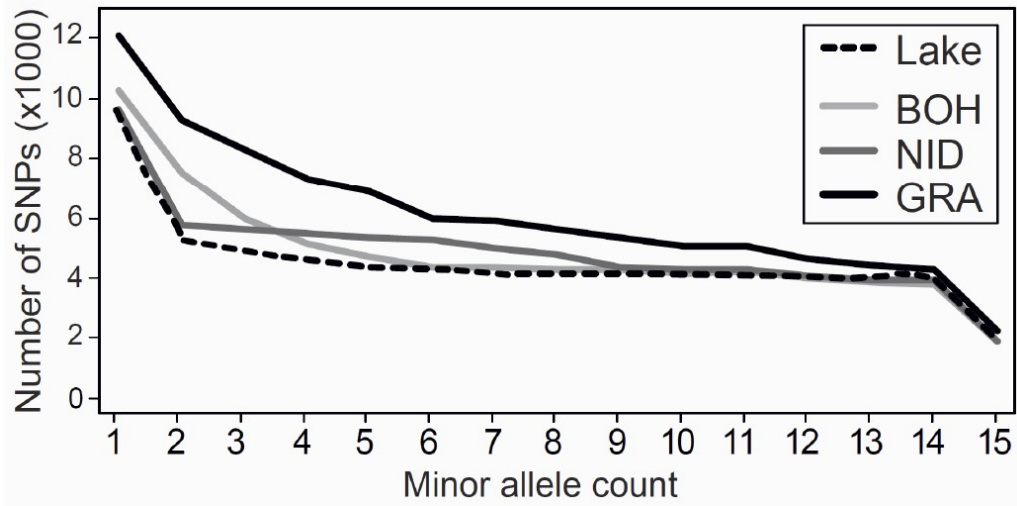


Phylogeny
(51 k SNPs)

... and reality



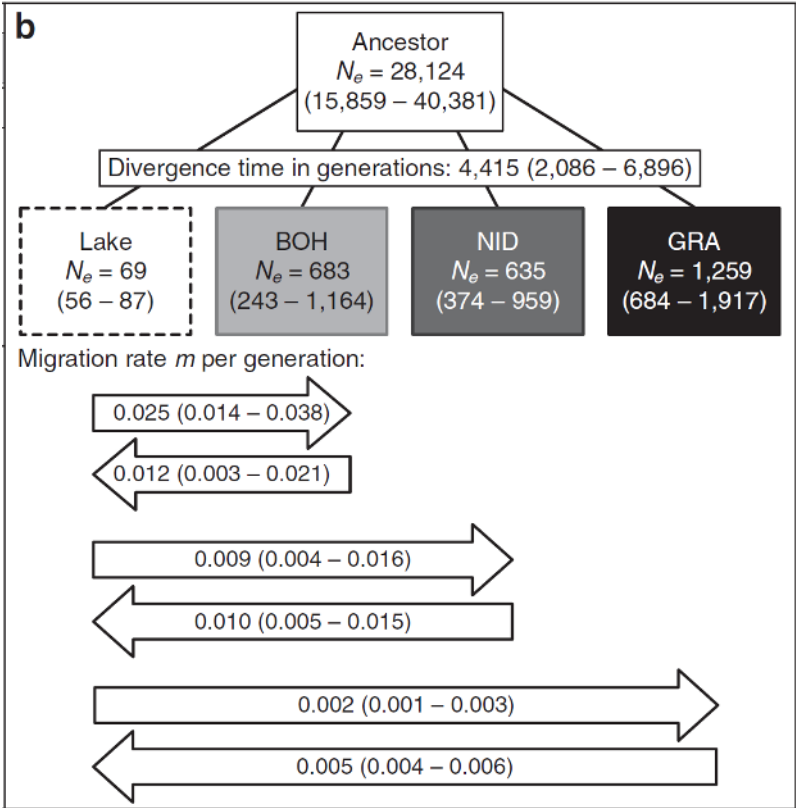
The lake population seems derived from stream fish



Site frequency spectra from the four populations

Demography with fastsimcoal, 14.8 million positions

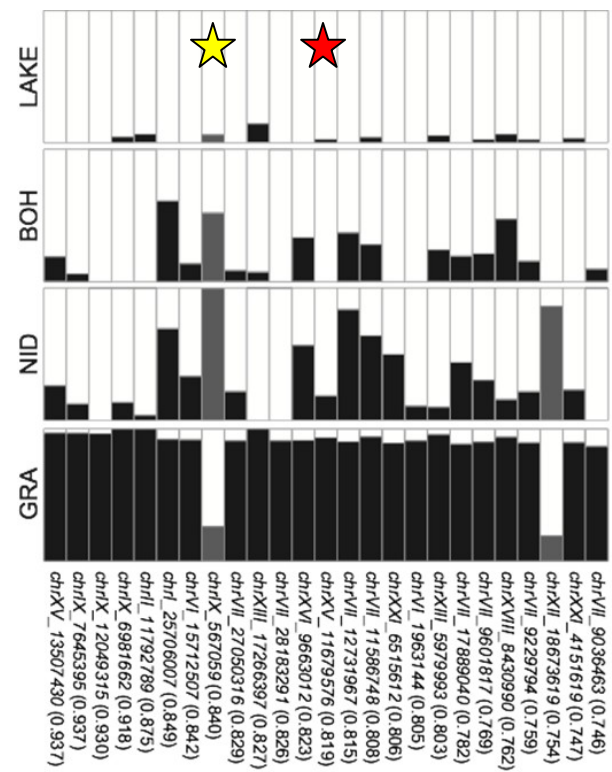
The lake population has the lowest genetic diversity



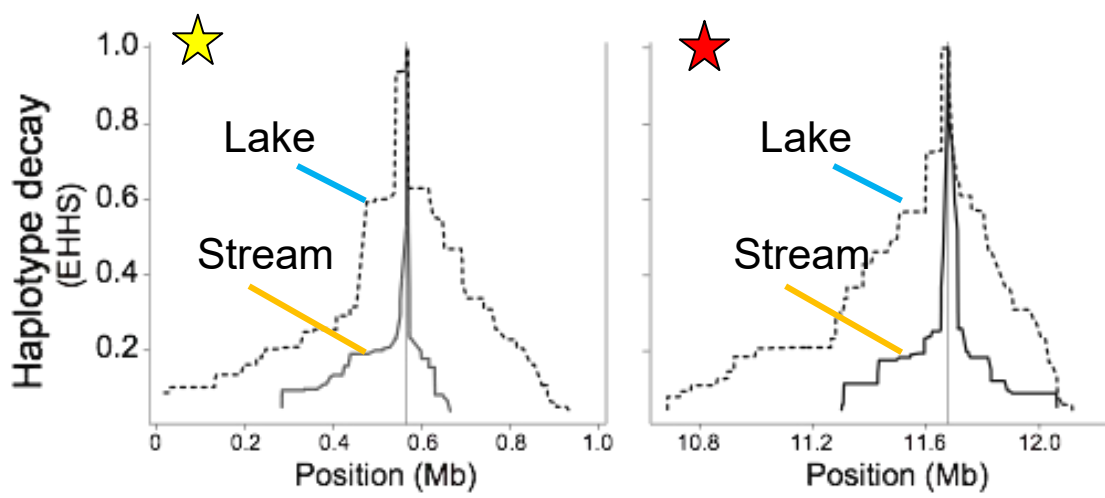
Hypothesis: The lake population has diverged from an ancestral stream form

Evidence from selected regions

F_{ST} extremes from all lake-stream comparisons



Haplotype tract length around F_{ST} extremes

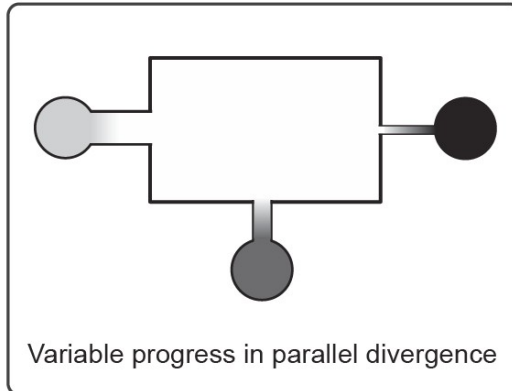
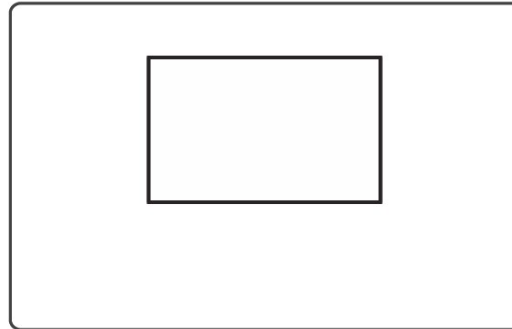
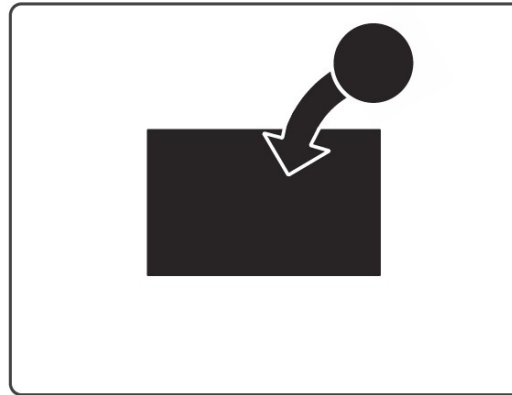


Selection has occurred mainly in the lake
Selective sweeps have eliminated genetic variation

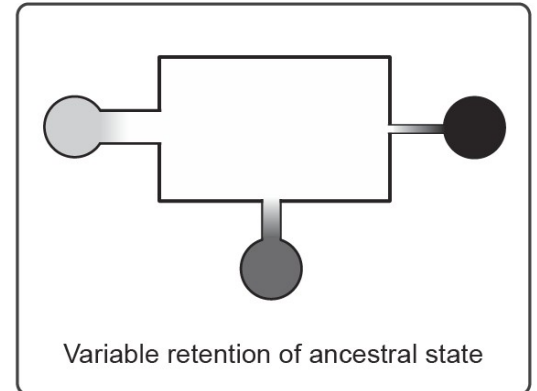
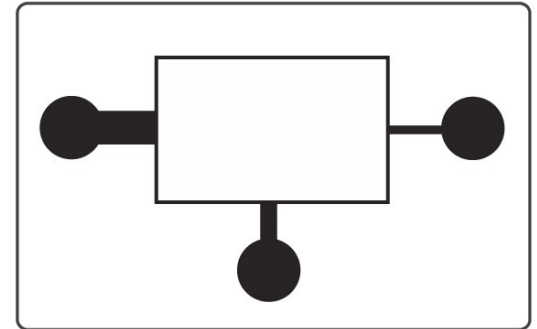
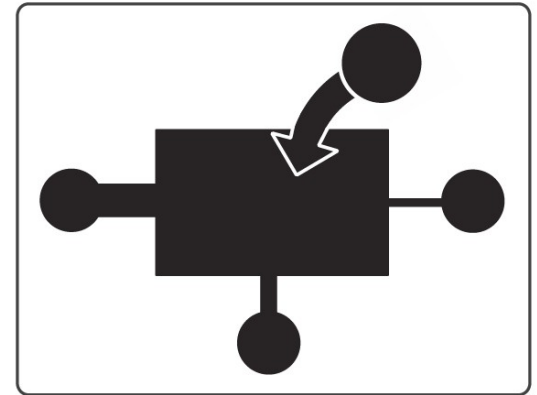
Resolving the history of adaptive divergence in Lake Constance stickleback

- Two alternative processes can result in similar patterns of population differentiation
- The population structure in the Lake Constance region is consistent with ecological vicariance
- Adaptation has occurred only once; conclusions about the determinism of natural selection are flawed

Parallel divergence



Ecological vicariance

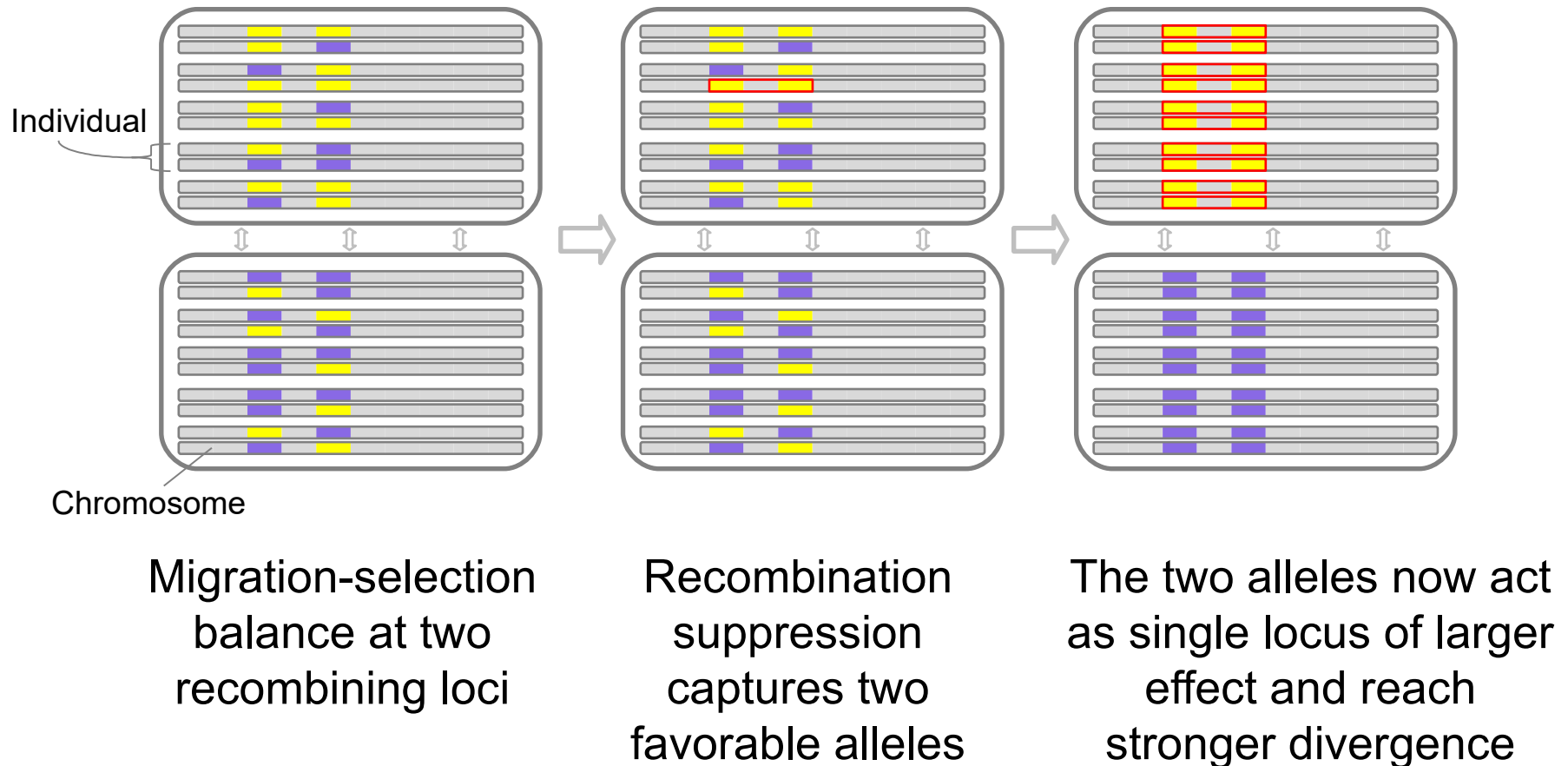
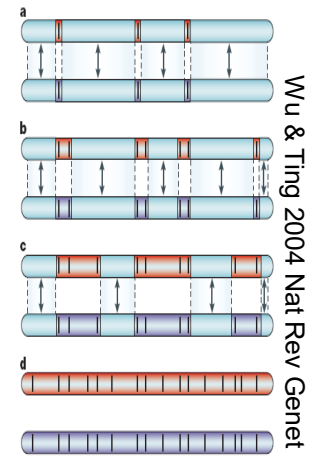


Conclusions

- Studying parallel evolution can be tricky: signatures in the genome might reflect ancient selection unrelated to the focal population contrast
- Parallel evolution can be mimicked by alternative evolutionary histories
- Robust ecological and demographic information, and sequence data from loci under selection, are valuable when inferring parallel evolution

Recombination and genomics

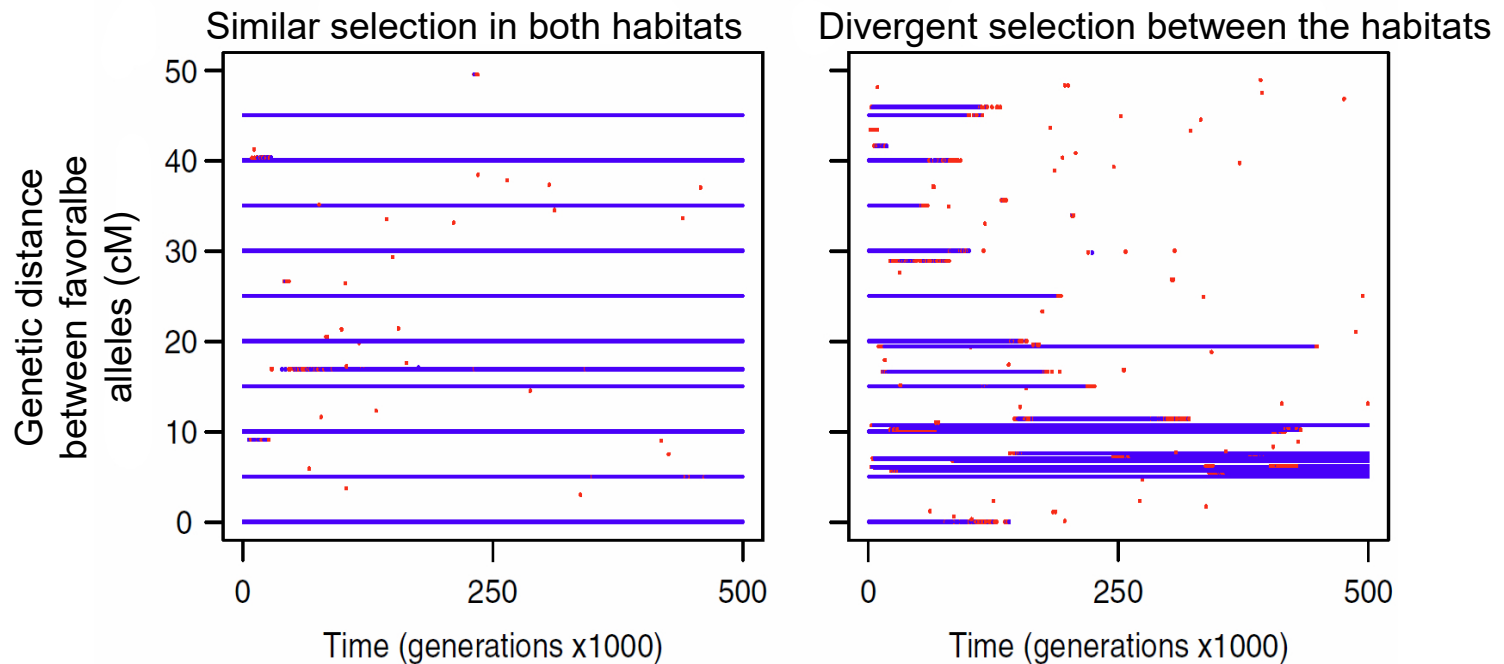
- Divergence with gene flow – a migration-selection antagonism
- The recombination rate influences this antagonism



Theoretical evidence of the adaptive benefit of reduced recombination

Yeaman 2013 PNAS

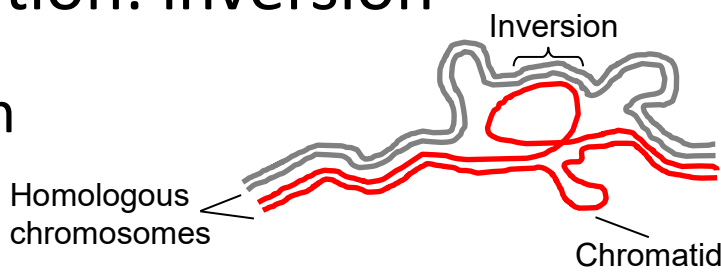
- Two habitats, 10 loci under migration-selection balance along a single chromosome
- Mutation: cut-paste positional relocation of the different alleles



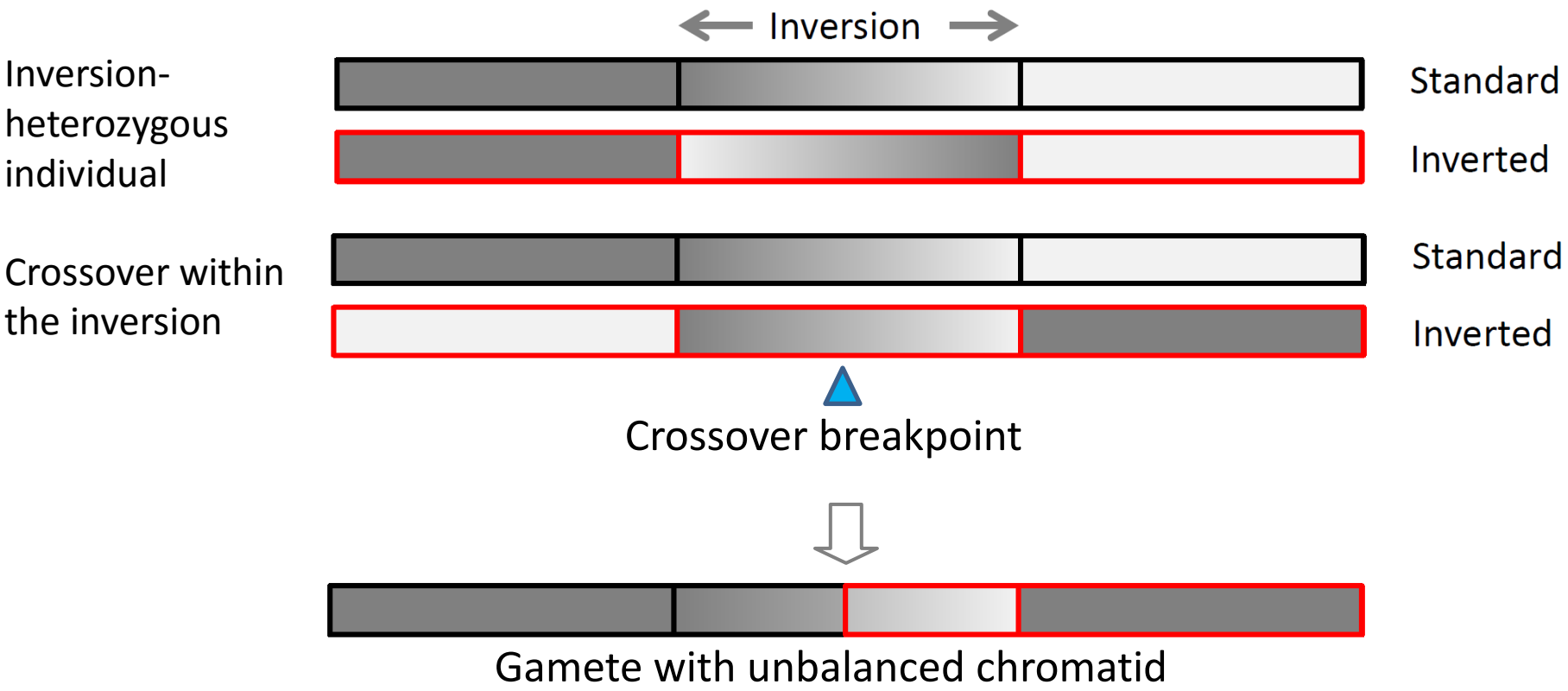
Selection favors a more compact genomic architecture (physical proximity = lower recombination frequency)

An easy way to suppress recombination: inversion

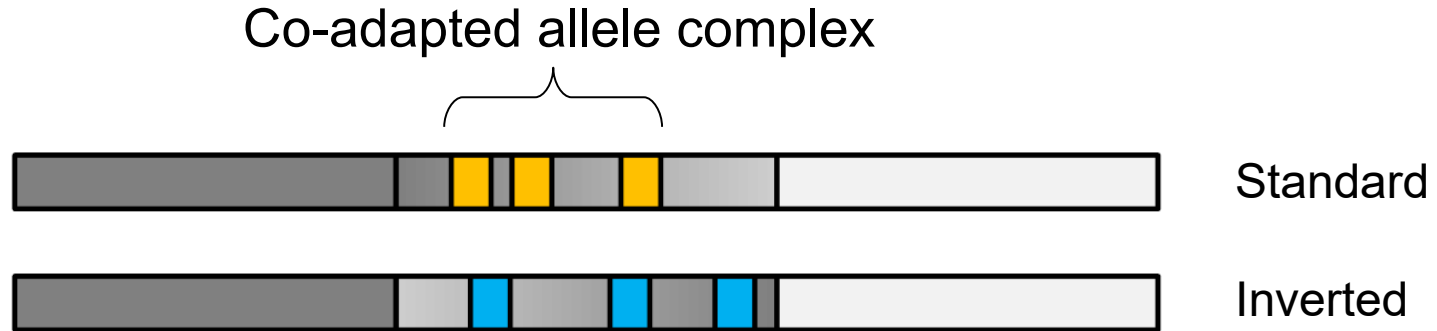
The inversion loop hinders recombination



When recombination still occurs within the inversion, the recombined gametes display gene duplications/deletions



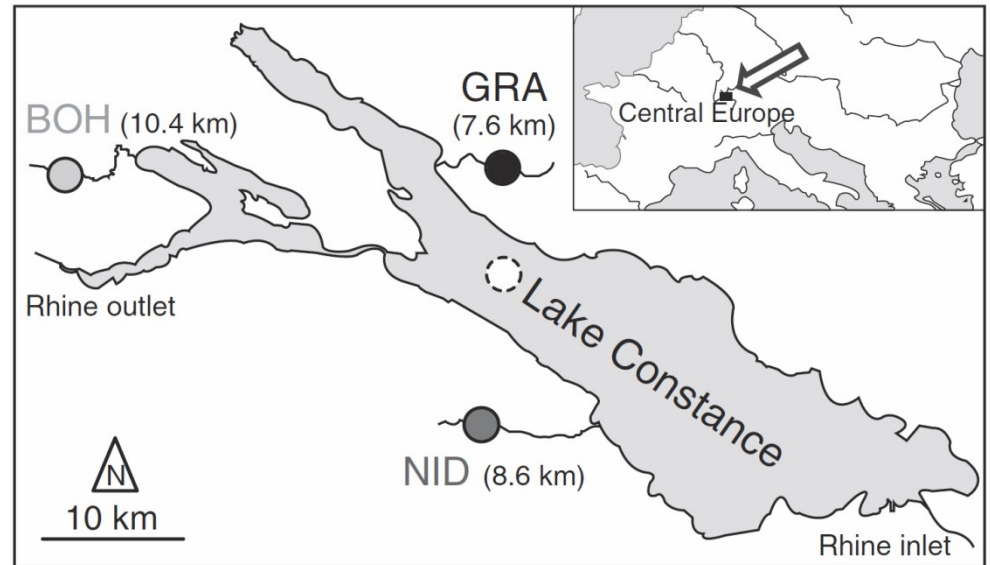
Prediction: inversions coupling adaptive alleles should be important to adaptive divergence



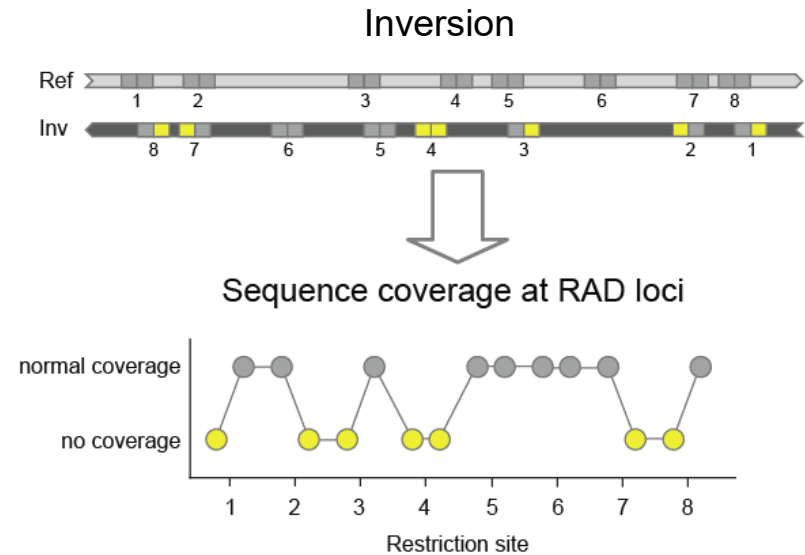
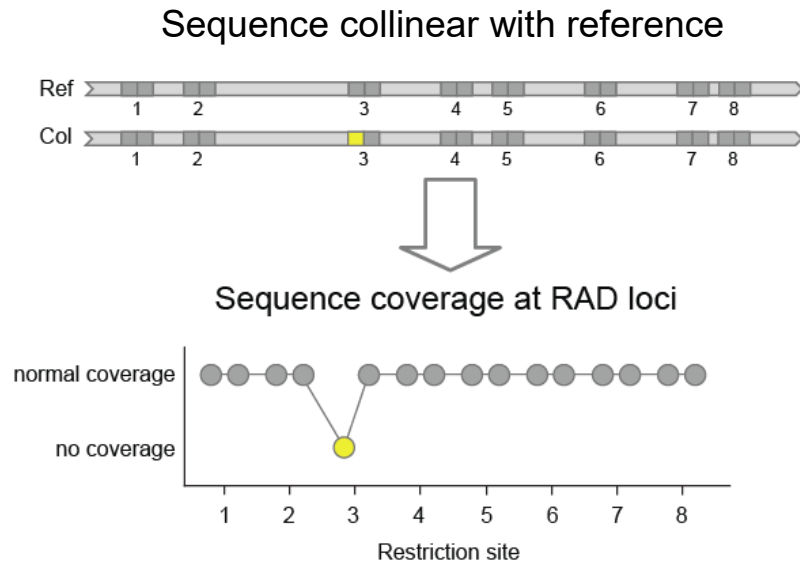
Exploring the significance of inversions in lake and stream stickleback

Roesti et al. 2015 Nat

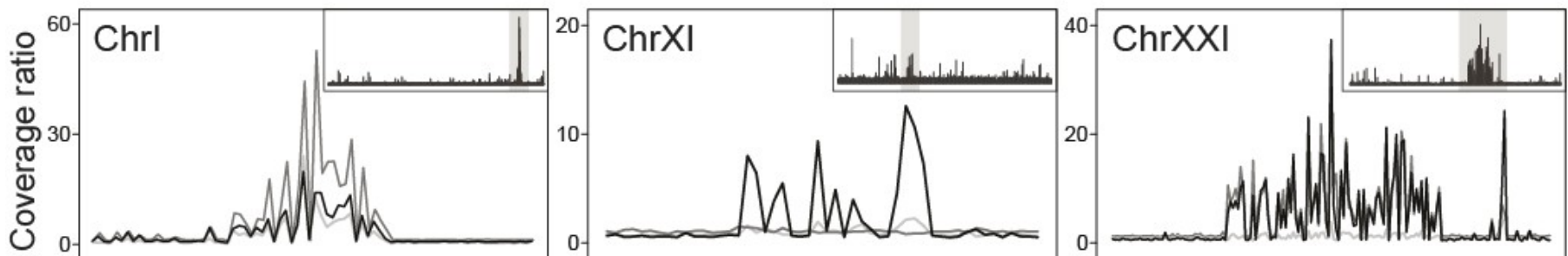
Commun



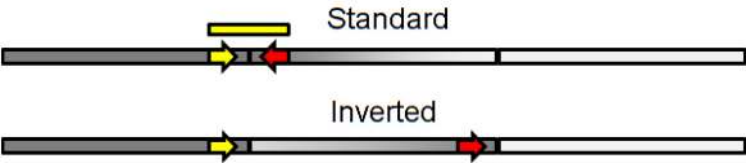
- Screen for inversions in Lake Constance stickleback
- Method: distortion in sequence coverage among RAD loci



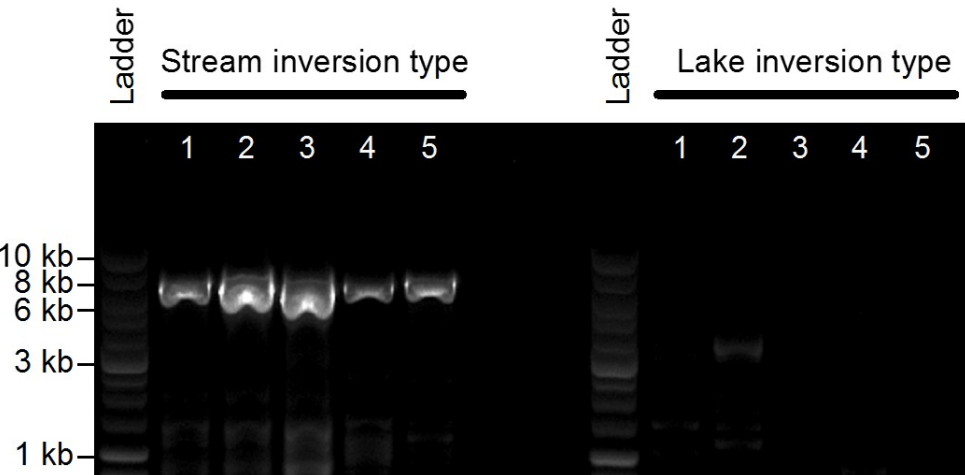
Coverage scans in lake-stream stickleback



Confirming the inversions using PCR primers across breakpoints



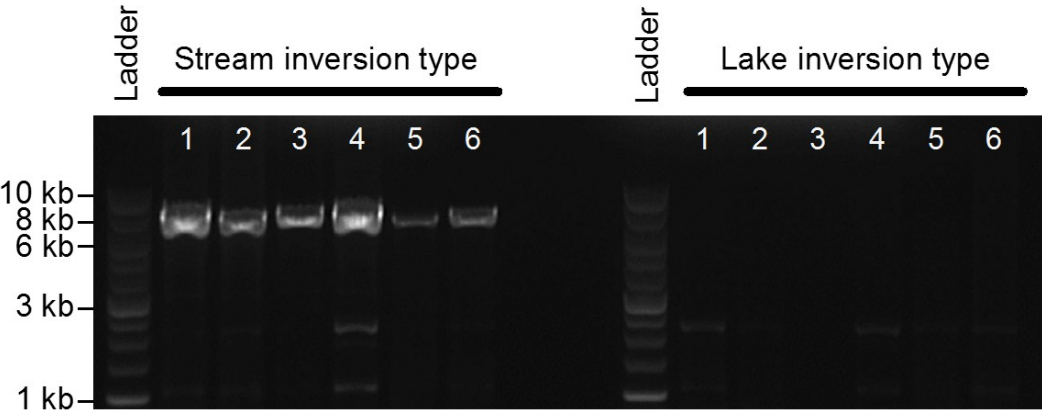
ChrI inversion



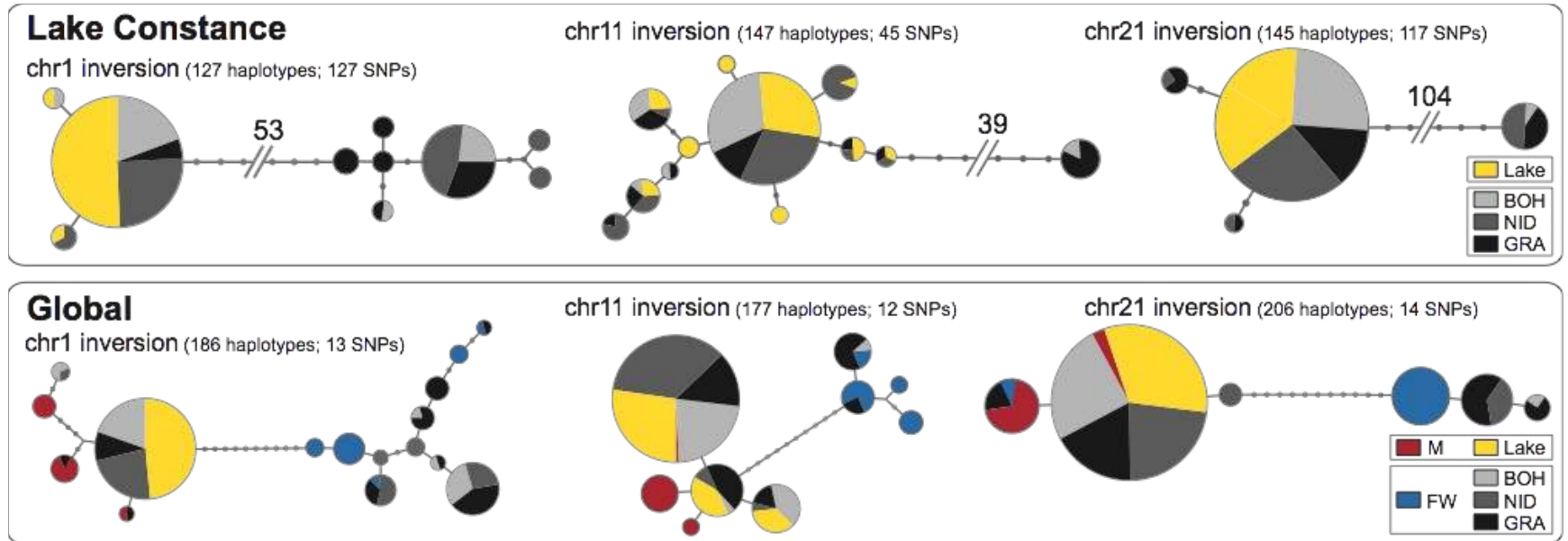
Three major inversions segregate in the Lake Constance basin

Detected previously in marine-freshwater stickleback Jones et al. 2012 Nature

ChrXI inversion



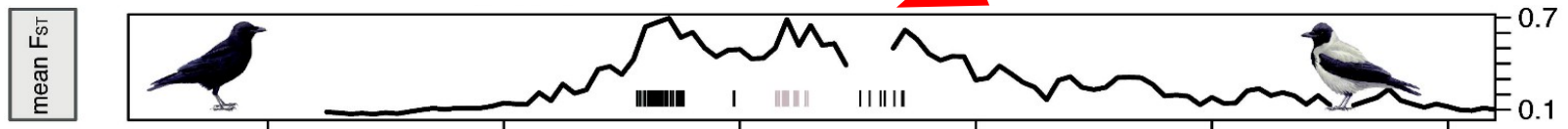
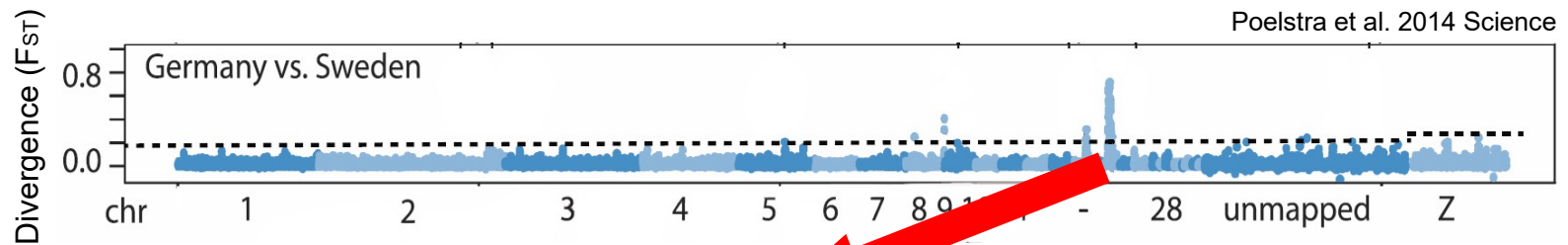
Exploring the ecology of the inversions based on inversion-specific haplotype genealogies



Jones et al. 2012 Nature

- Ancient inversions have been re-cycled for lake-stream divergence
- Likely ecological factor: pelagic vs. benthic life style
- Parallel evolution at a global scale

Evidence from other systems for a role of inversions in adaptive divergence



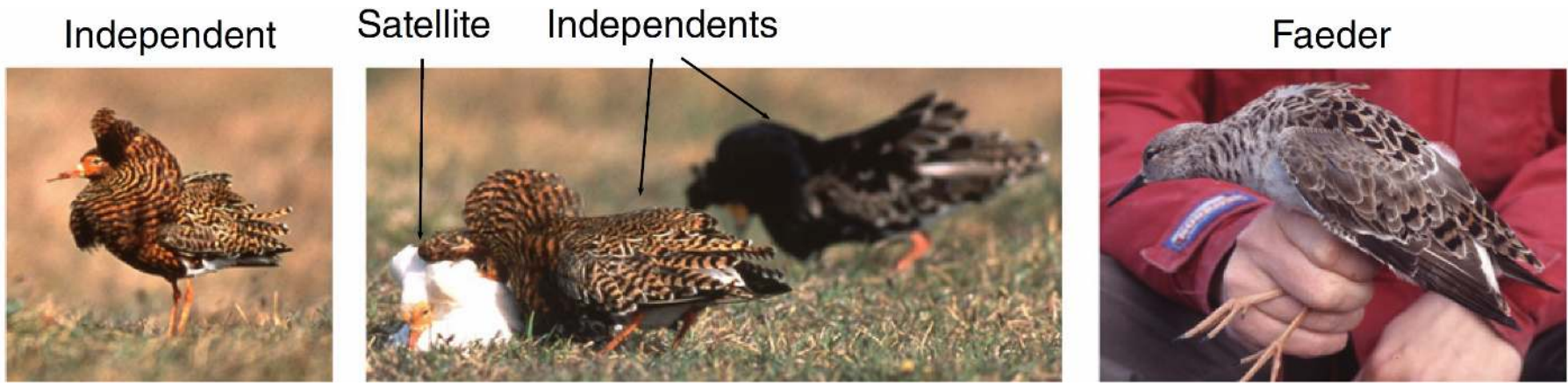
1,883,205

2,459,610

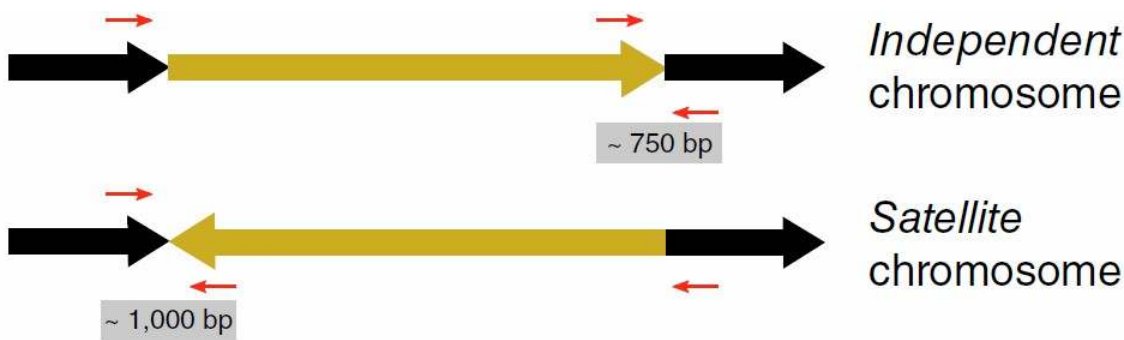
2,494,964



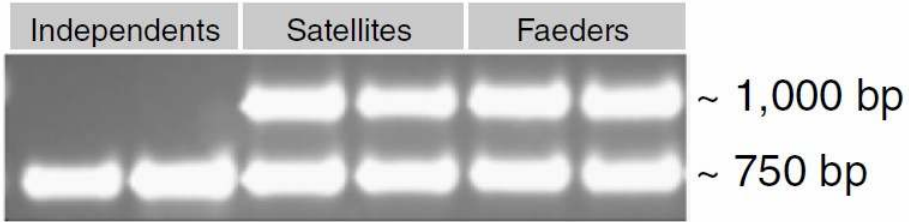
Hypothesis: inversion coupling plumage color and perception alleles



Lamichhaney et al. 2015 Nat Genet



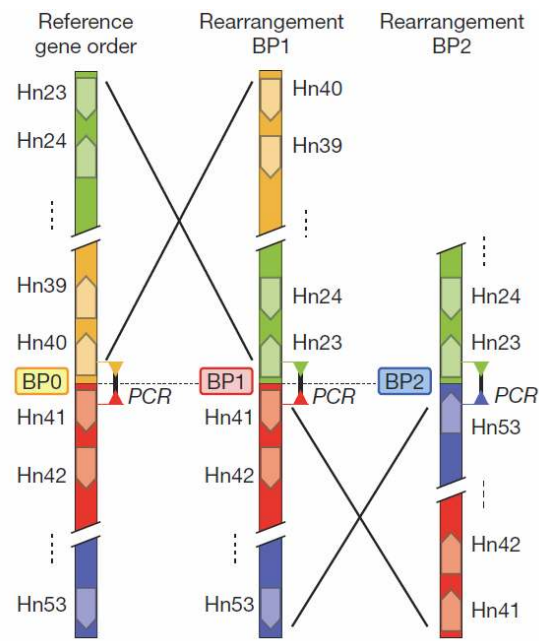
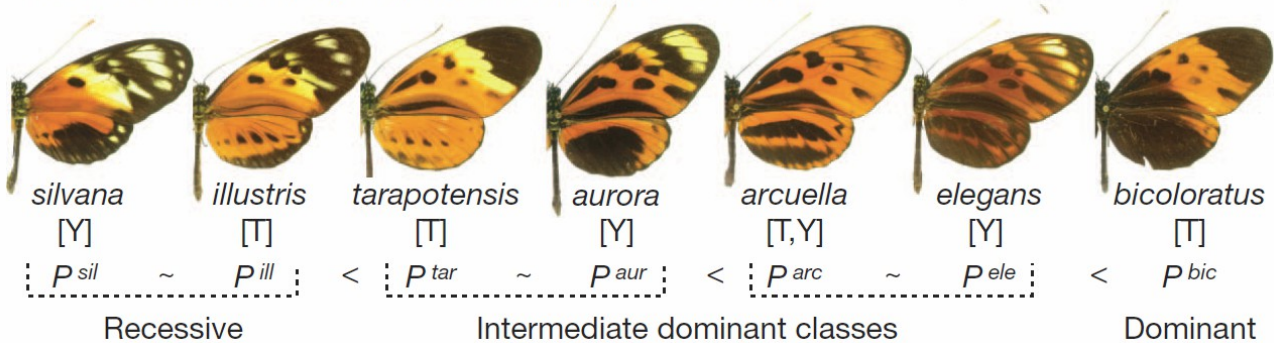
Inversion coupling
behavior and plumage
color alleles



Melinaea : models



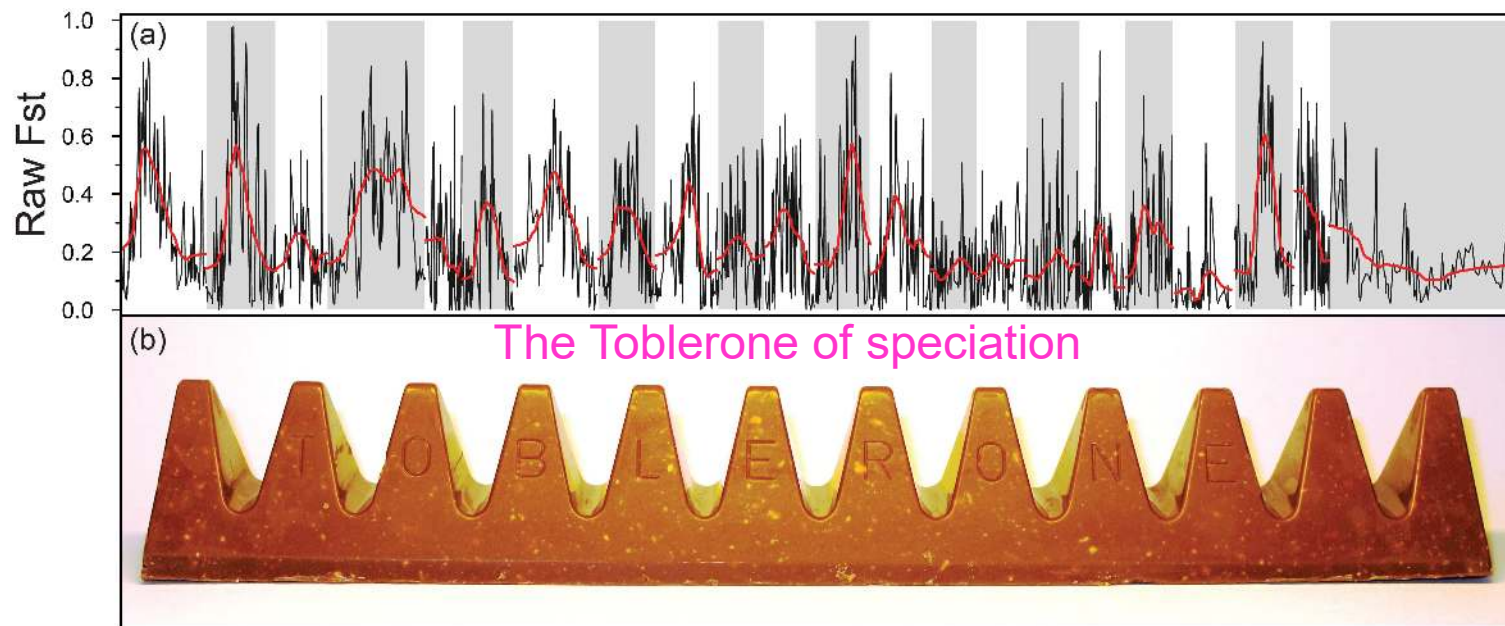
Heliconius numata : polymorphic mimic, sympatric morphs



Wing color elements
coupled in adjacent
inversions

Aren't we overlooking something rather important?

Genomic divergence between lake and stream stickleback from Vancouver Island is biased toward chromosome centers Roesti et al. 2012 Mol Ecol

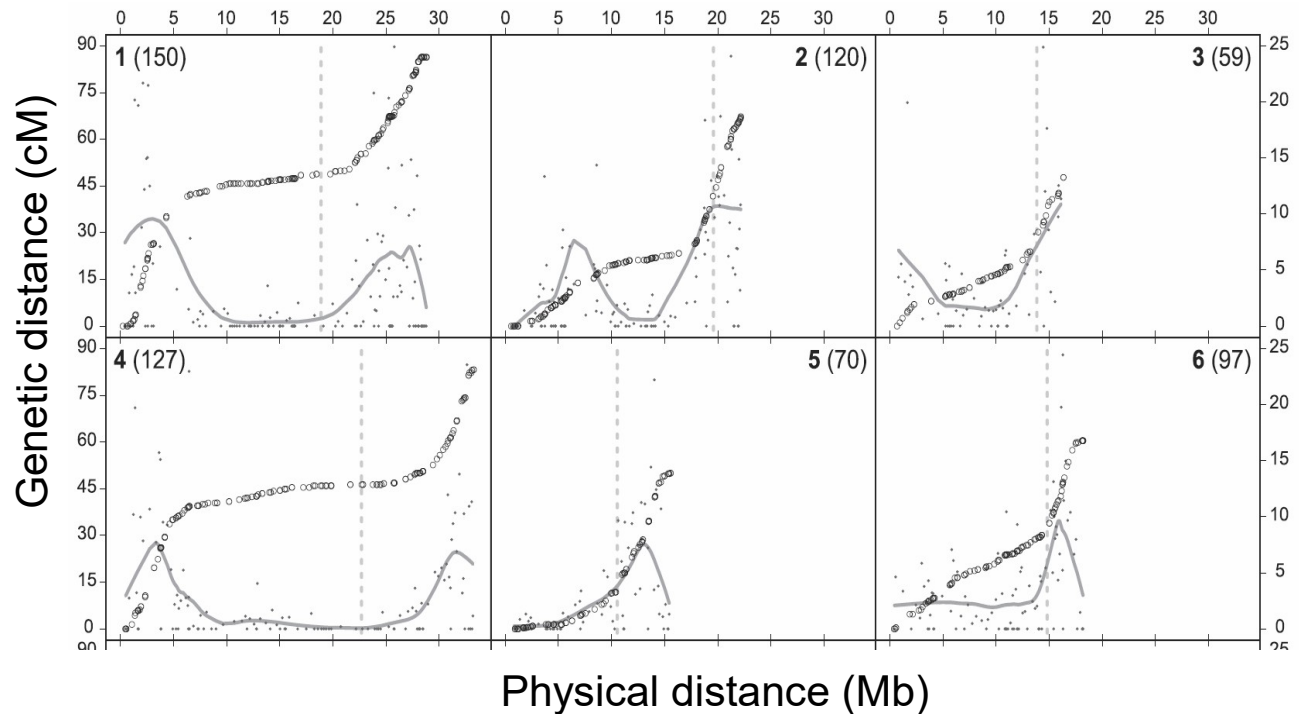


Let's call this CCBD for **C**hromosome **C**enter-**B**iased **D**ivergence

A recombination-based explanation for CCBD

Roesti et al. 2012, 2013 Mol Ecol

Linkage mapping (F2 intercross, N = 280) reveals reduced recombination rate in stickback chromosome centers

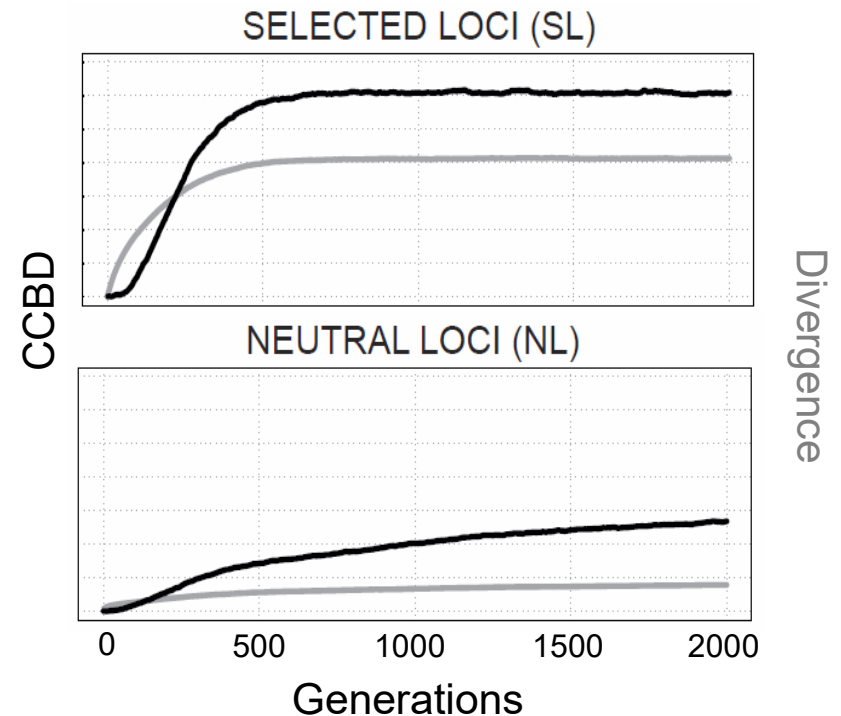
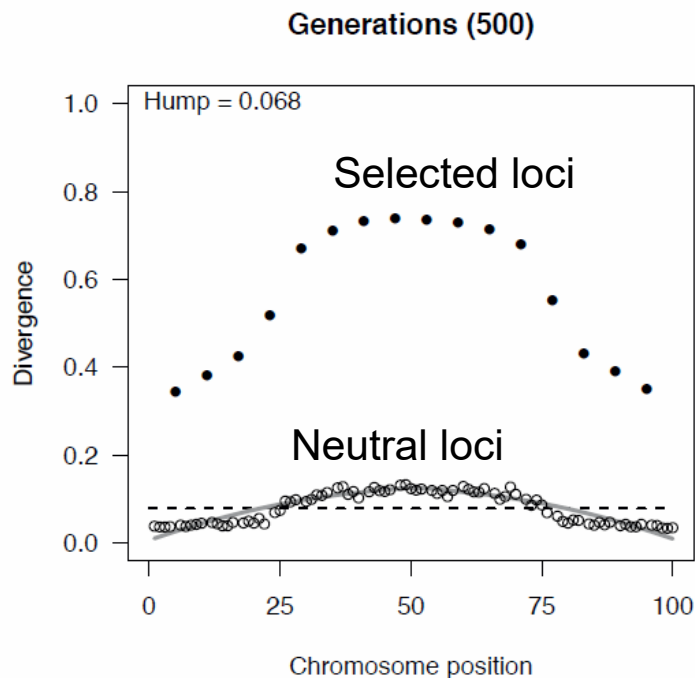


Genomic divergence in lake-stream stickleback is highly polygenic

Hypothesis: The coupling of alleles in chromosome centers makes the introgression of DNA in chromosome centers more difficult than in the peripheries

Validation using adaptive divergence *in silico* Berner & Roesti, in progress

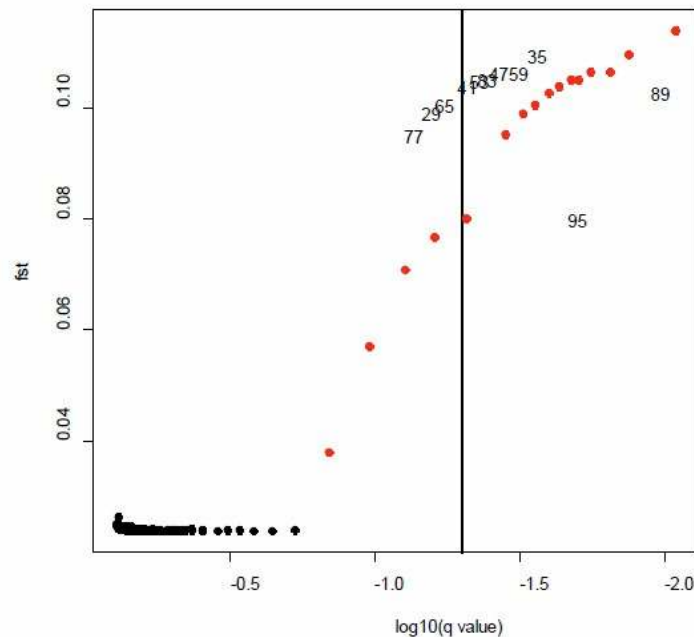
- Adaptive divergence with gene flow between two populations, using standing variation at many loci under selection
- Single chromosome; neutral loci interspersed between the selected ones
- Recombination is biased to the peripheries
- Parameters: Pop size, time, selection strength, migration rate, recombination bias, etc



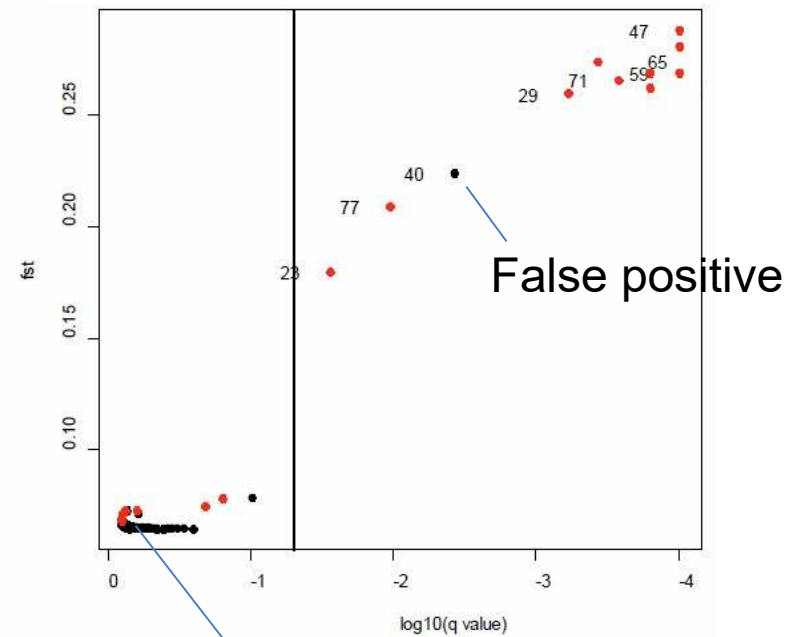
One of many consequences of CCBD

Biased outlier detection: standard F_{ST} outlier scan using BayeScan

Recombination rate
homogeneous along
chromosome



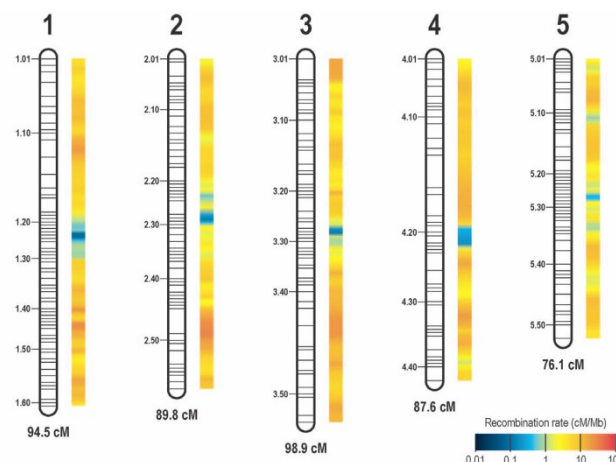
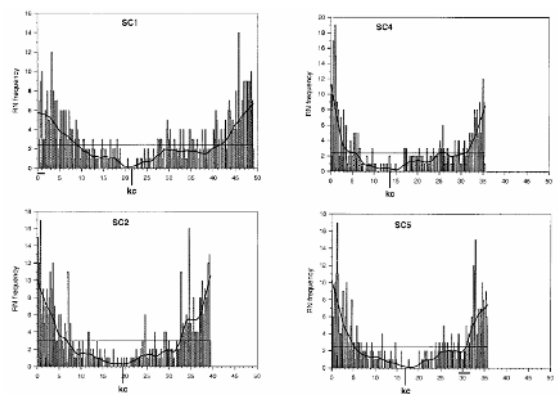
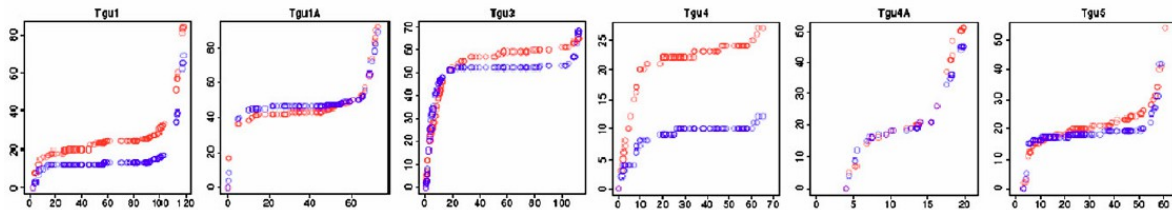
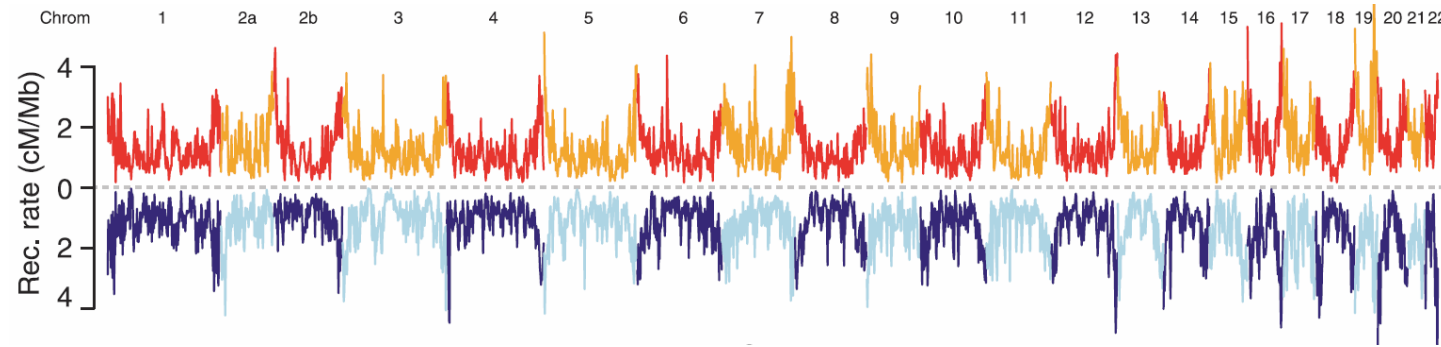
Recombination rate biased
toward chromosome center



False negatives

We ARE overlooking something rather important!

Elevated recombination in chromosome peripheries seems taxonomically universal



Conclusions

- Recombination suppression via inversions clearly contributes to adaptive divergence
- Our perspective on the general adaptive significance of inversions, however, is biased by simple large-effect systems
- The consequences broad-scale heterogeneity in recombination rate along chromosomes is largely ignored in current genomics (theory coming up!)

Thanks

- Marius Roesti, Dario Moser, Benjamin Kueng, Anja Frey, Walter Salzburger, Andrew Hendry, Sergey Gavrilets
- Christian Beisel, Ina Niessen (Department of Biosystems Science and Engineering, D-BSSE Basel)

-
- SNF
 - University of Basel

