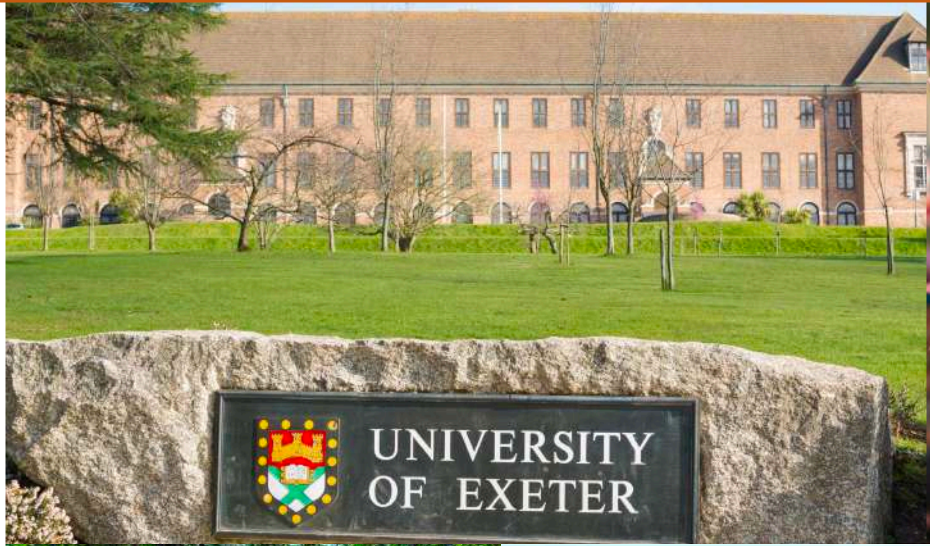
Population Genomics

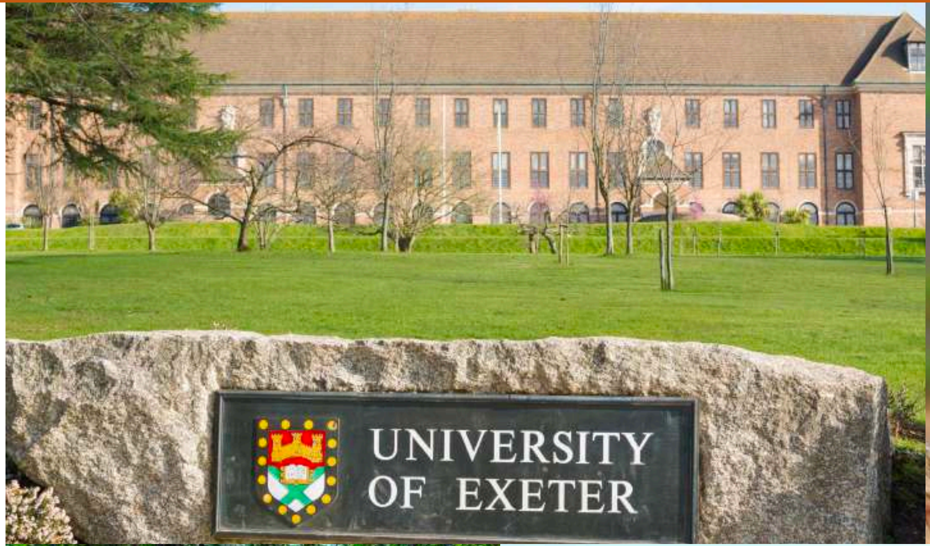
A bit about me…PhD

A bit about me…PhD

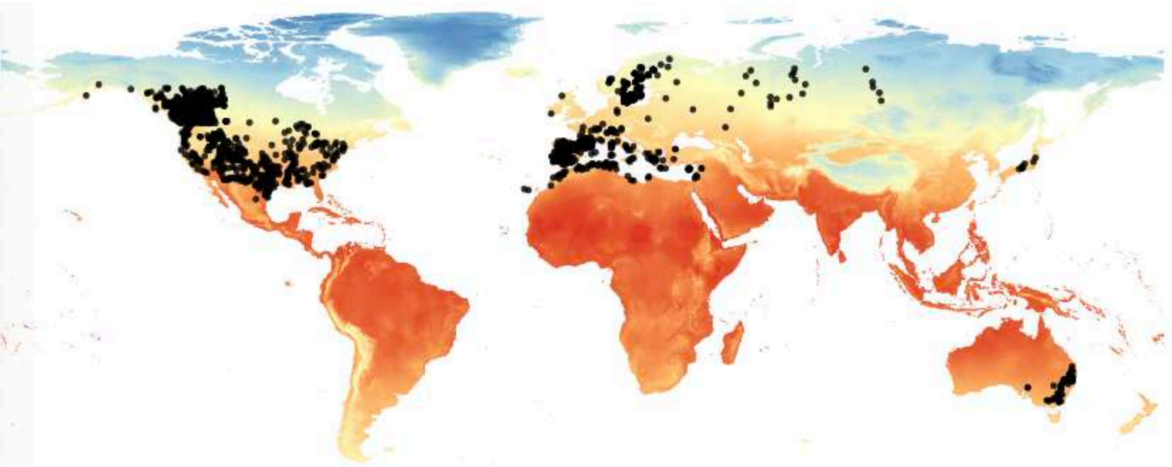
A bit about me…Postdoc 1



A bit about me…Postdoc 1



A bit about me…Postdoc 2



A bit about me…Selling Out

SNP:	S1	S2	S3	S4
Ind 1	0	1	0	1
Ind 2	1	1	1	0
Ind 3	0	1	0	1
Ind 4	1	0	0	1

A bit about me…Selling Out

Species:	S1	S2	S3	S4
Site 1	0	1	0	1
Site 2	1	1	1	0
Site 3	0	1	0	1
Site 4	1	0	0	1

Population Genetics – What the?!

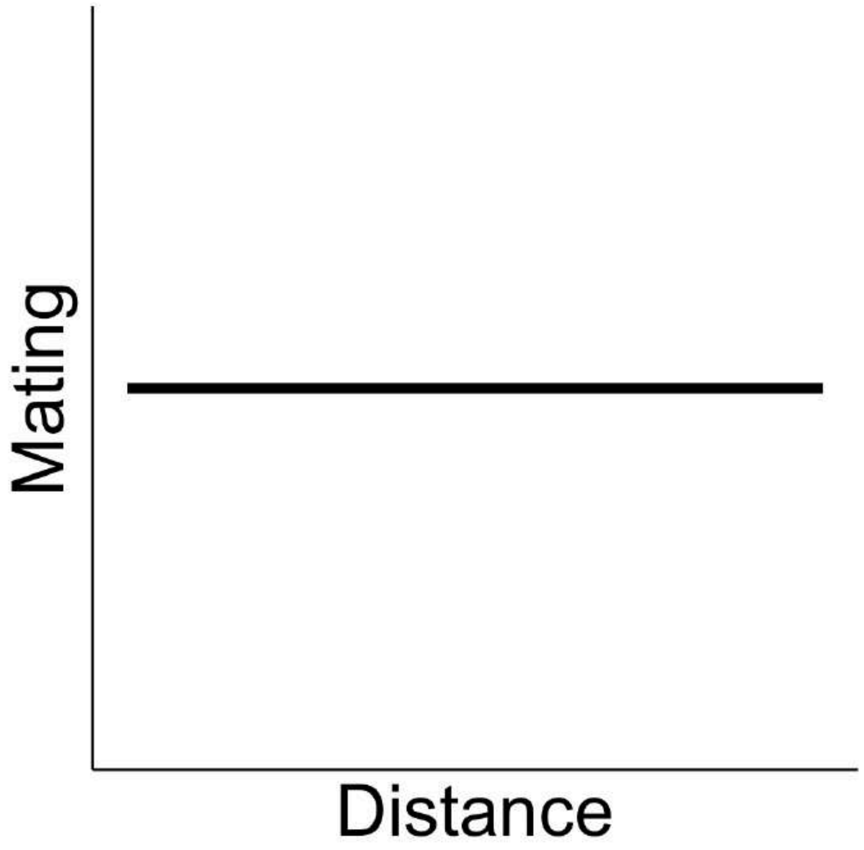
How is genetic variation distributed across a species’ range?

How is genetic variation distributed across the genome?

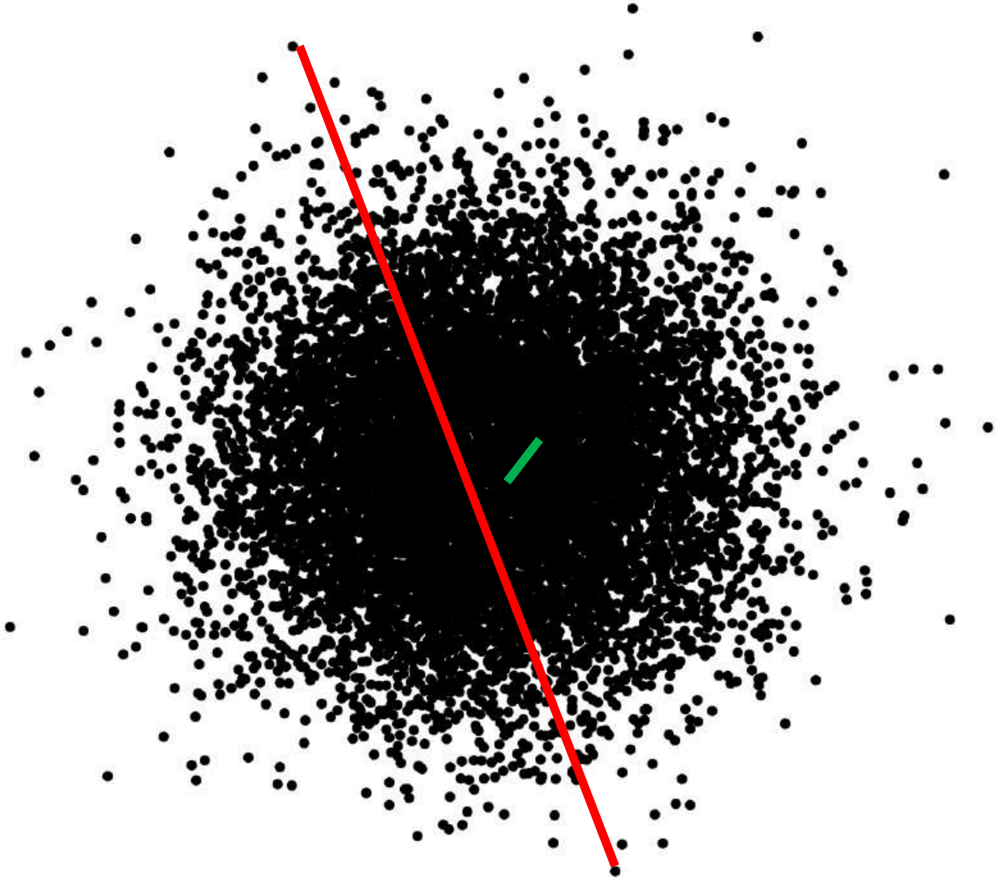
Where does genetic variation come from? Why does it persist?

What do we mean by a ‘population’

Panmixia

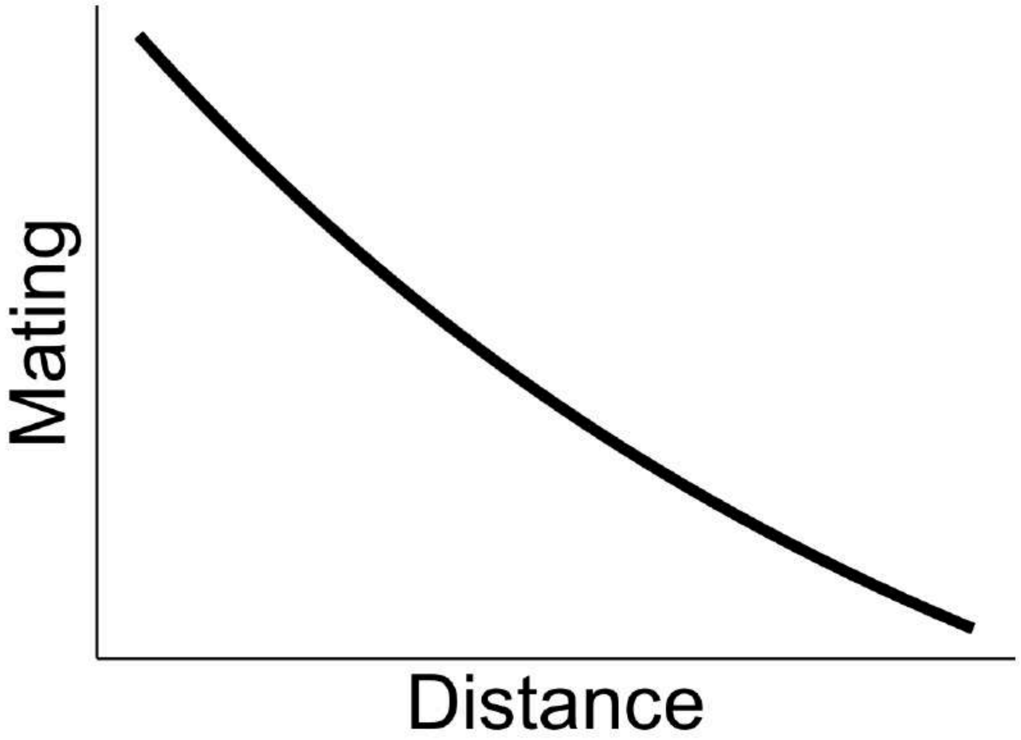


M = M

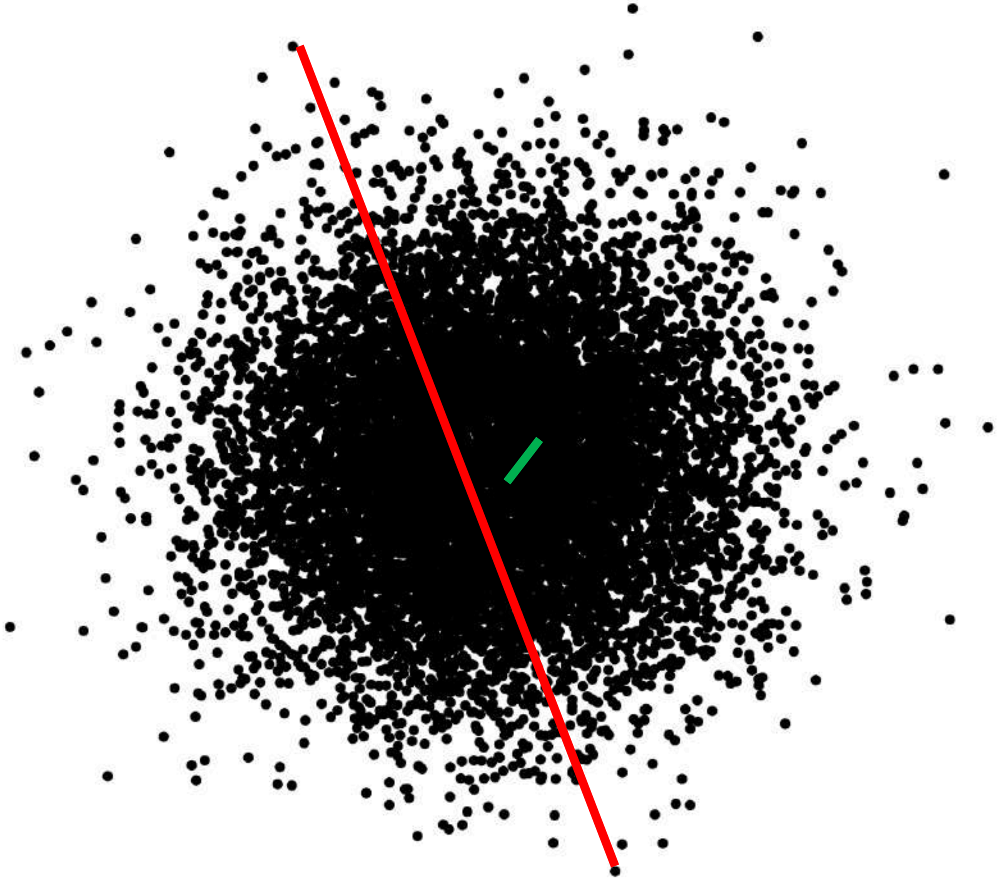


What do we mean by a ‘population’

Continuous spatial populations

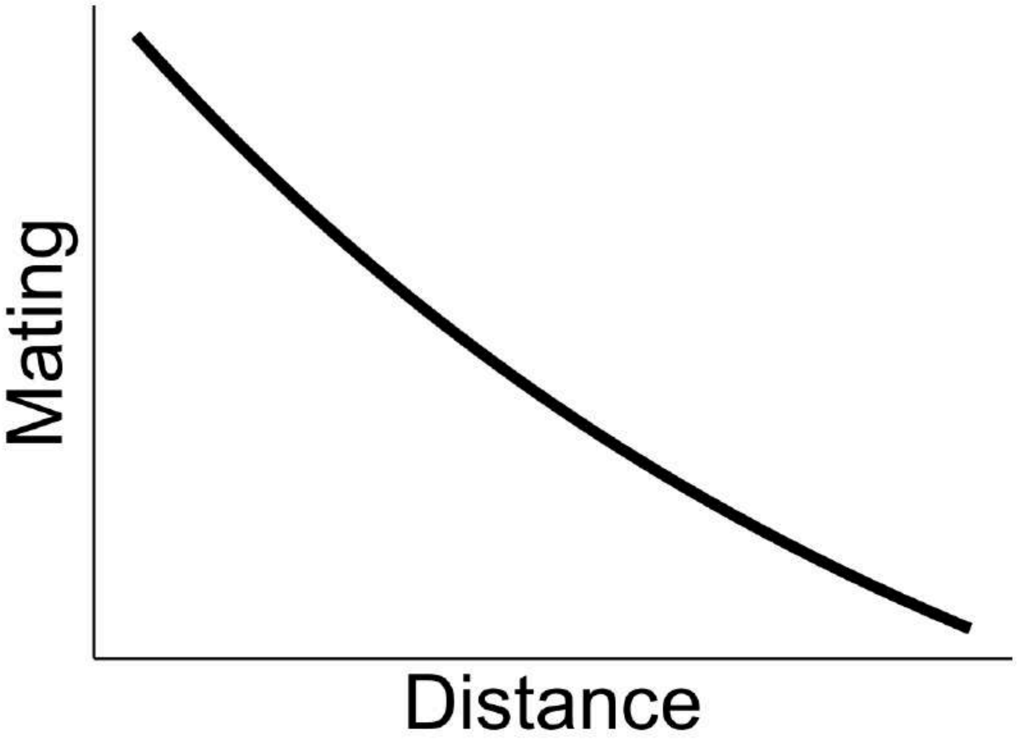


M < M

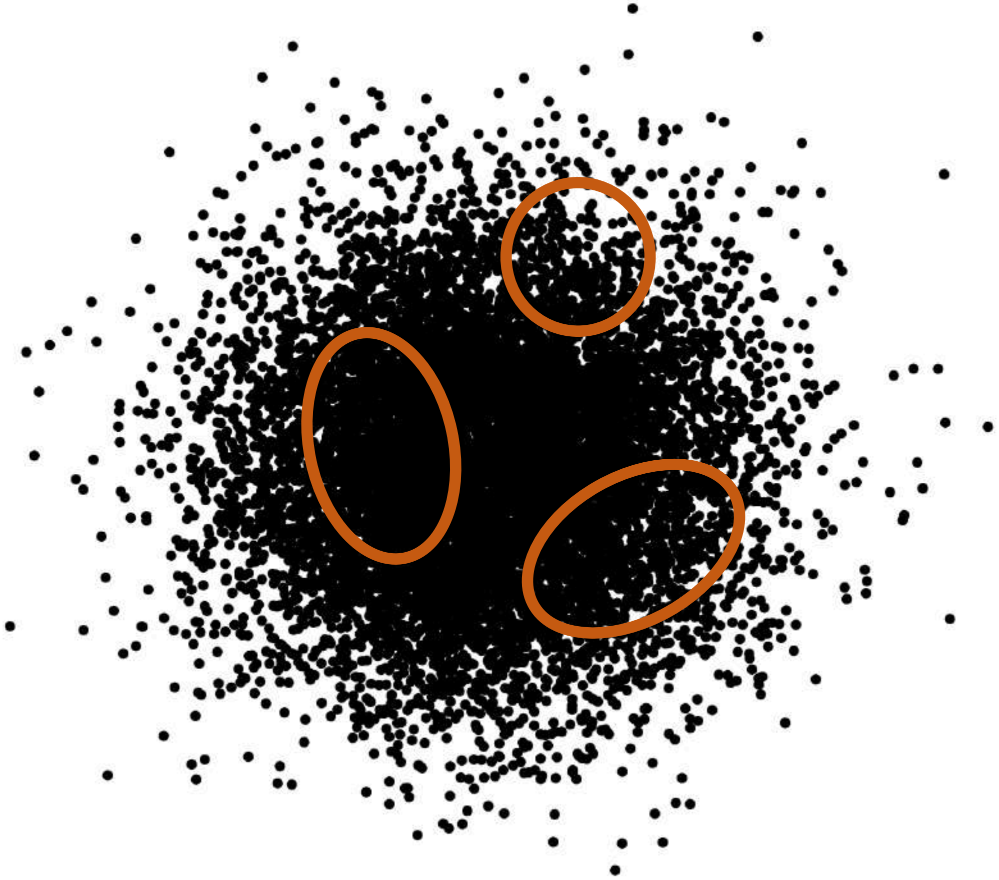


What do we mean by a ‘population’

Continuous spatial populations

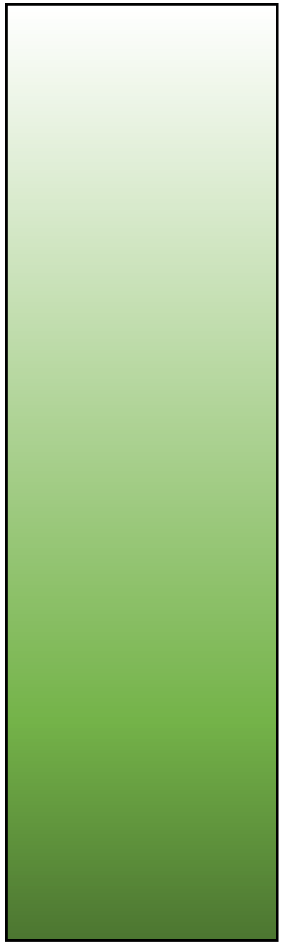


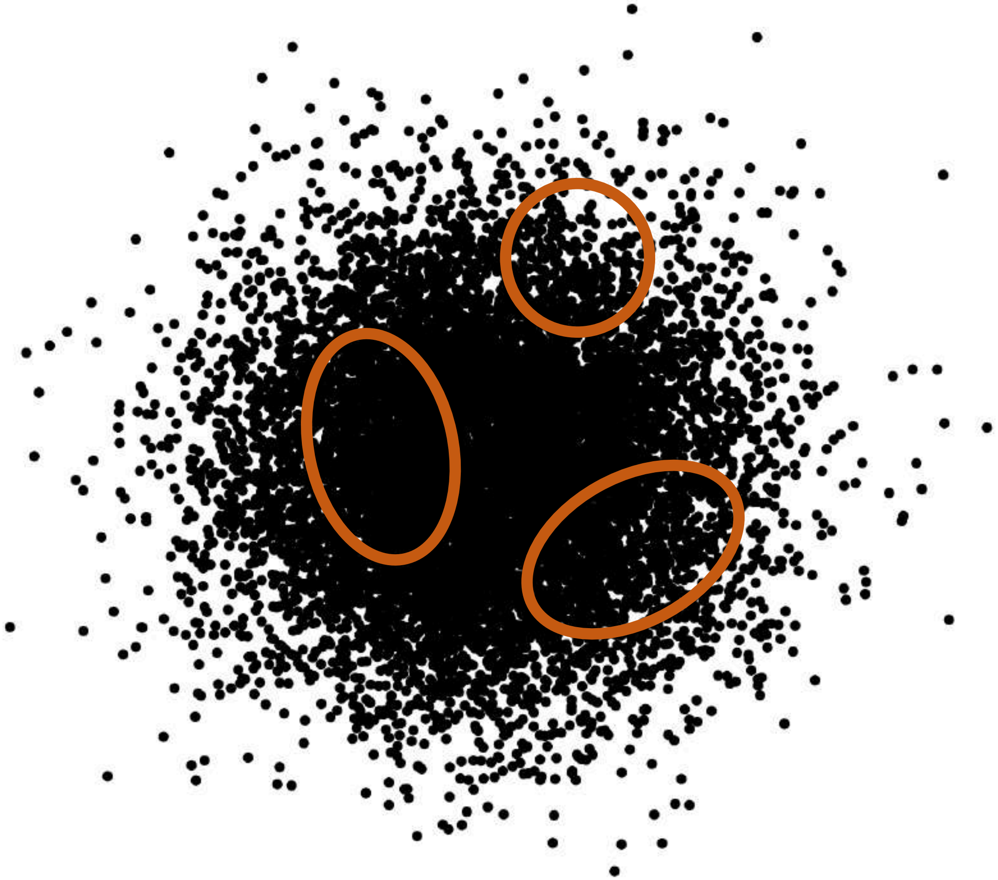
M < M



What do we mean by a ‘population’

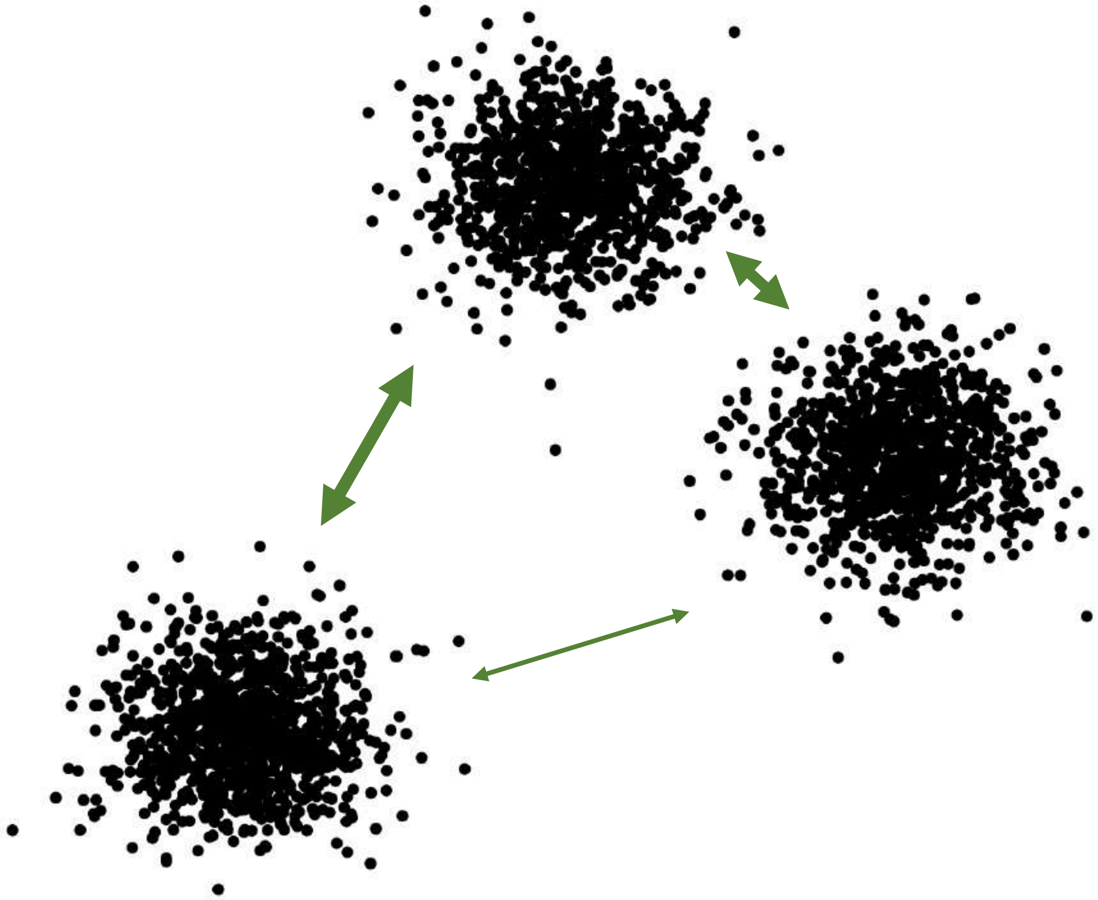
Continuous spatial populations





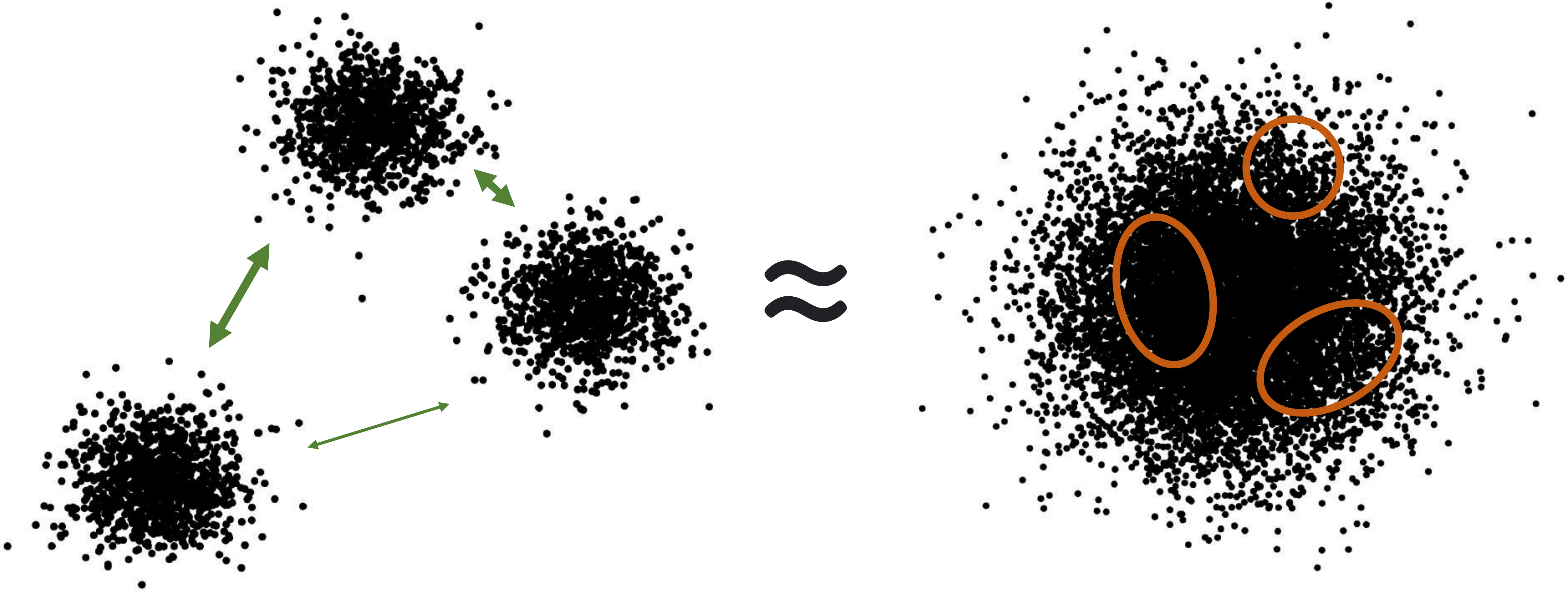
What do we mean by a ‘population’

Discrete spatial populations



What do we mean by a ‘population’

Discrete spatial populations



Population Genetics – What the?!

Recombination

Gene Flow

Selection

Drift

Mutation

Mutation

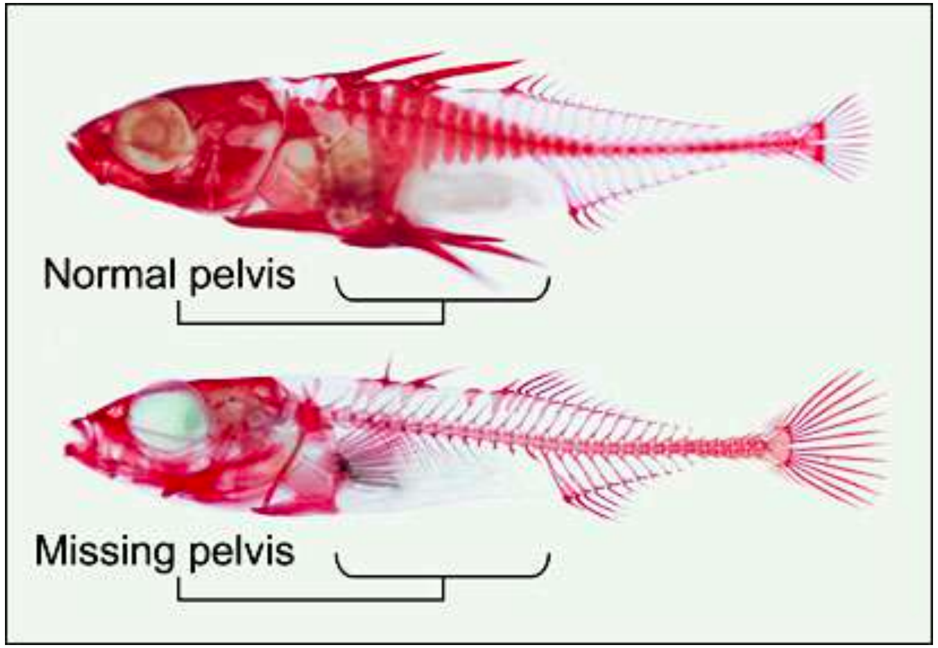
Mutation is **RANDOM (with respect to fitness)**

Mutation is **NON-RANDOM (with respect to location)**

The ultimate source of novel variation

Often we are interested in :
Pop-scaled mutation rate $= 4N\mu$
Distribution of fitness effects = DFE

Mutation



EVOLUTION

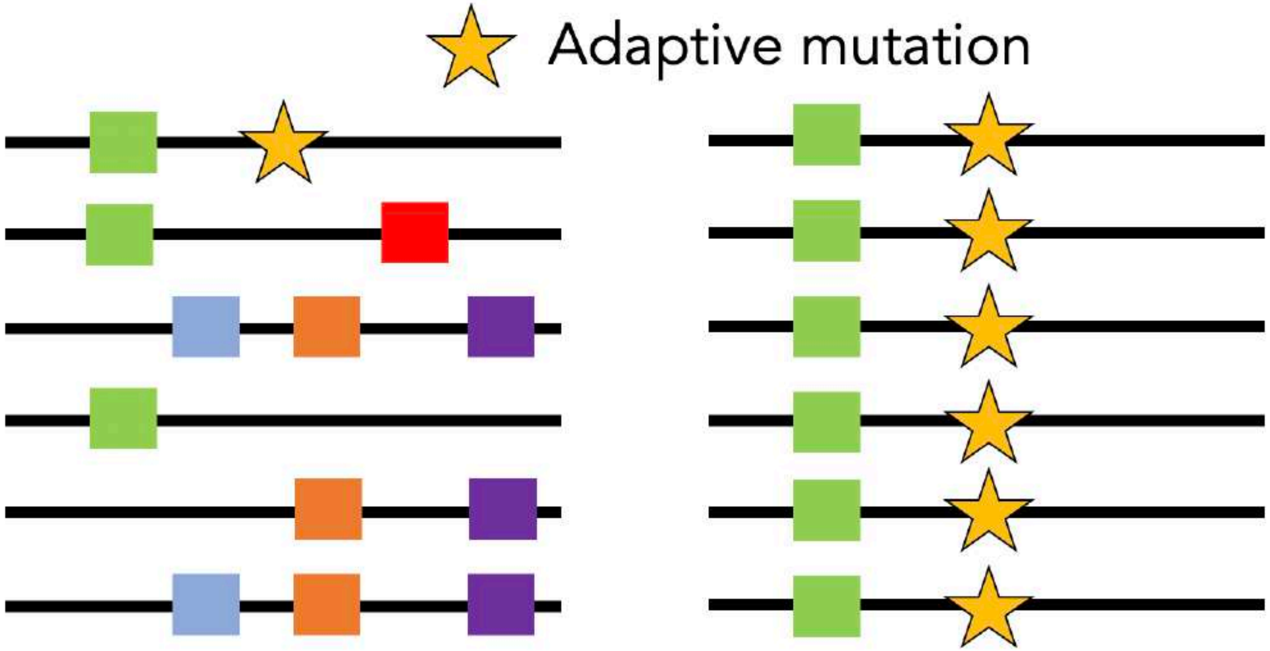
DNA fragility in the parallel evolution of pelvic reduction in stickleback fish

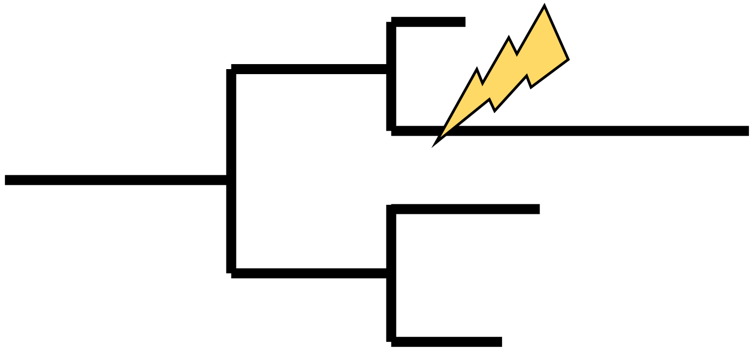
Kathleen T. Xie[1,2,3], Guliang Wang[4], Abbey C. Thompson[1,5], Julia I. Wucherpfennig[1], Thomas E. Reimchen[6], Andrew D. C. MacColl[7], Dolph Schluter[8], Michael A. Bell[9]*, Karen M. Vasquez[4], David M. Kingsley[1,2]†

Selection - Positive

Reduces variation within populations (linked selection)

Increases variation among populations (unless ‘global’)





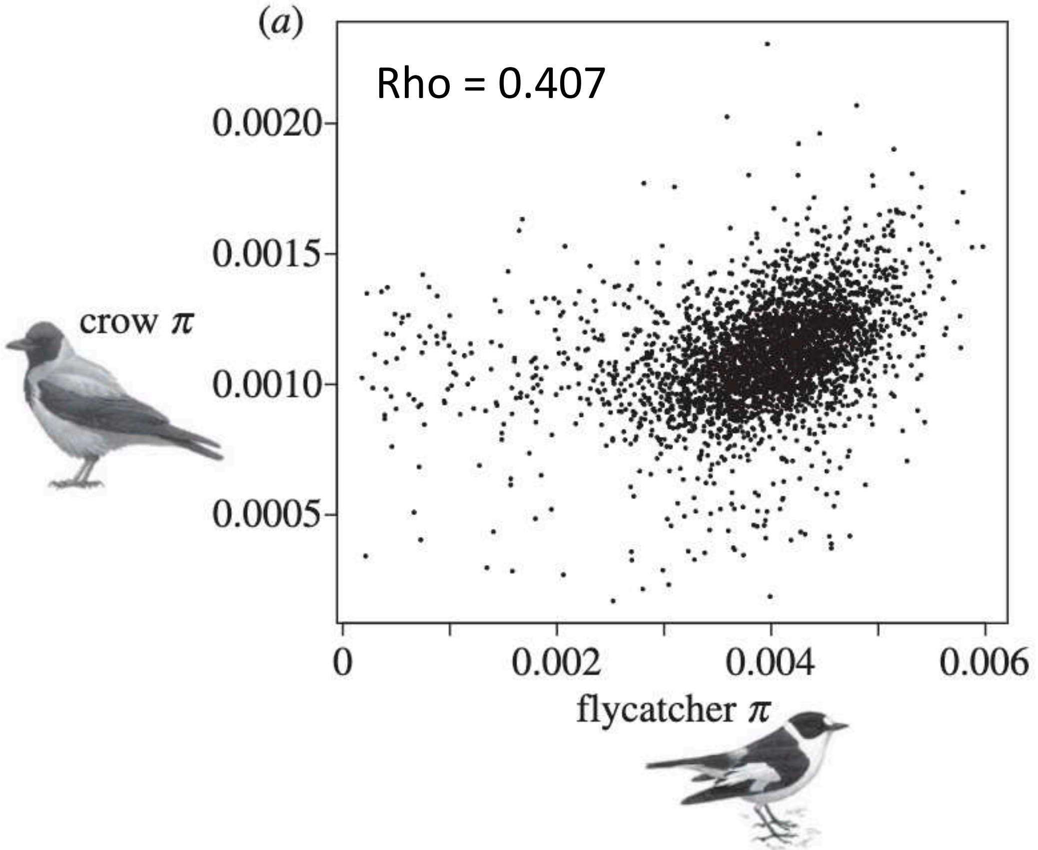
Selection - Background

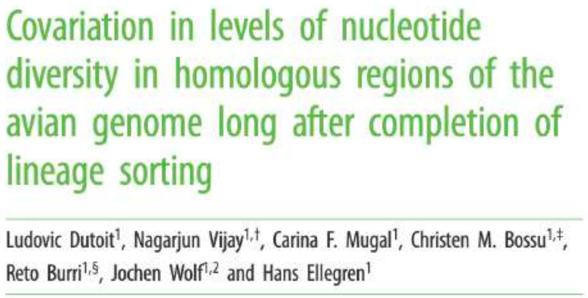
Purifying selection removes deleterious variants

Deleterious mutations take linked neutral variation with them

Gene density increases deleterious likelihood

Low recombination/tight linkage removes more neutral variation




Covariation in levels of nucleotide diversity in homologous regions of the avian genome long after completion of lineage sorting

Ludovic Dutoit[1], Nagarjun Vijay[1,†], Carina F. Mugal[1], Christen M. Bossu[1,‡], Reto Burri[1,§], Jochen Wolf[1,2] and Hans Ellegren[1]

Selection - Balancing

Hotspots and maintenance of genetic variation in the genome

Can be driven by different forms of selection, e.g. Negative Frequency Dependent Selection or Heterozygote Advantage

Commonly associated with immune genes e.g. MHC in response to pathogens

Selection - Balancing

🔒 | RESEARCH ARTICLE | SEXUAL SELECTION

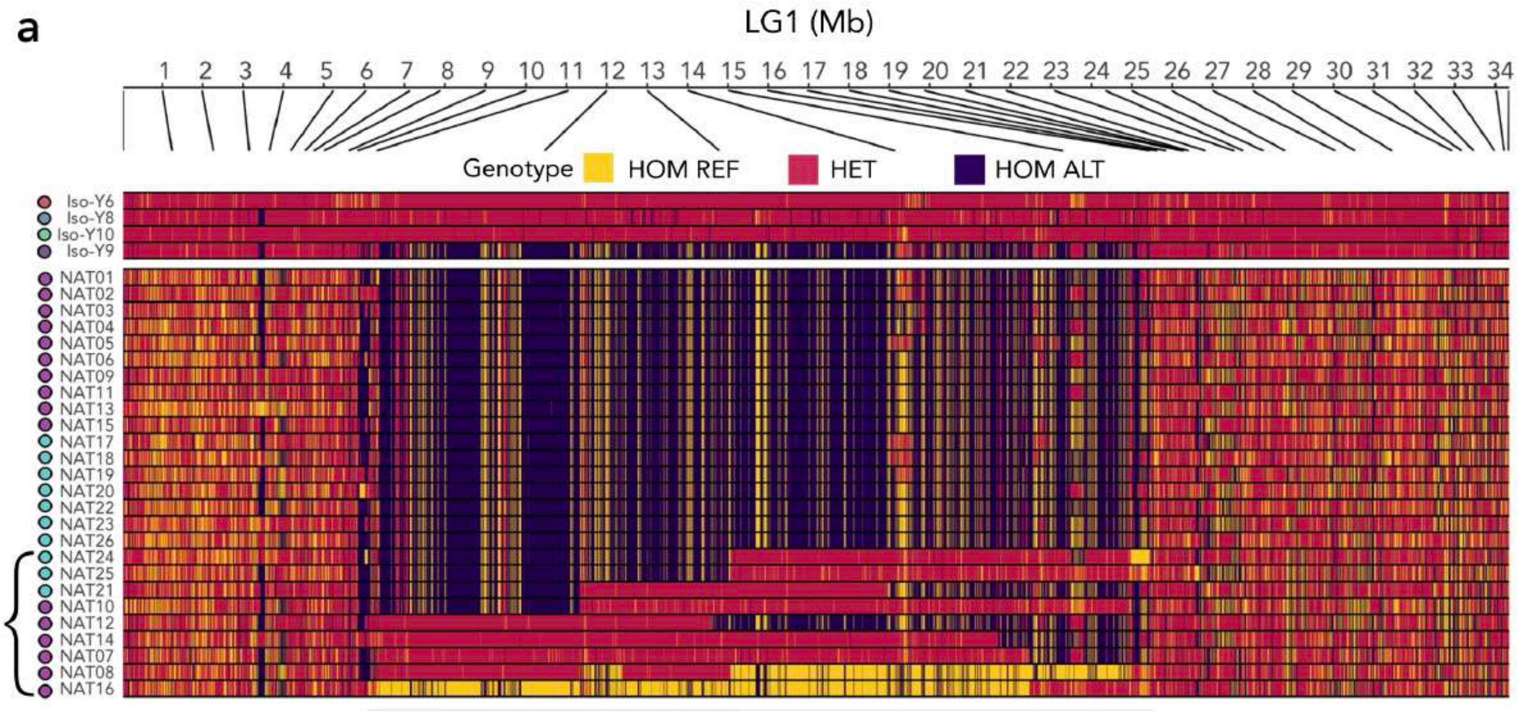
Female preference for rare males is maintained by indirect selection in Trinidadian guppies

TOMOS POTTER, JEFF ARENDT, RONALD D. BASSAR, BETH WATSON, PAUL BENTZEN, JOSEPH TRAVIS, AND DAVID N. REZNICK Authors

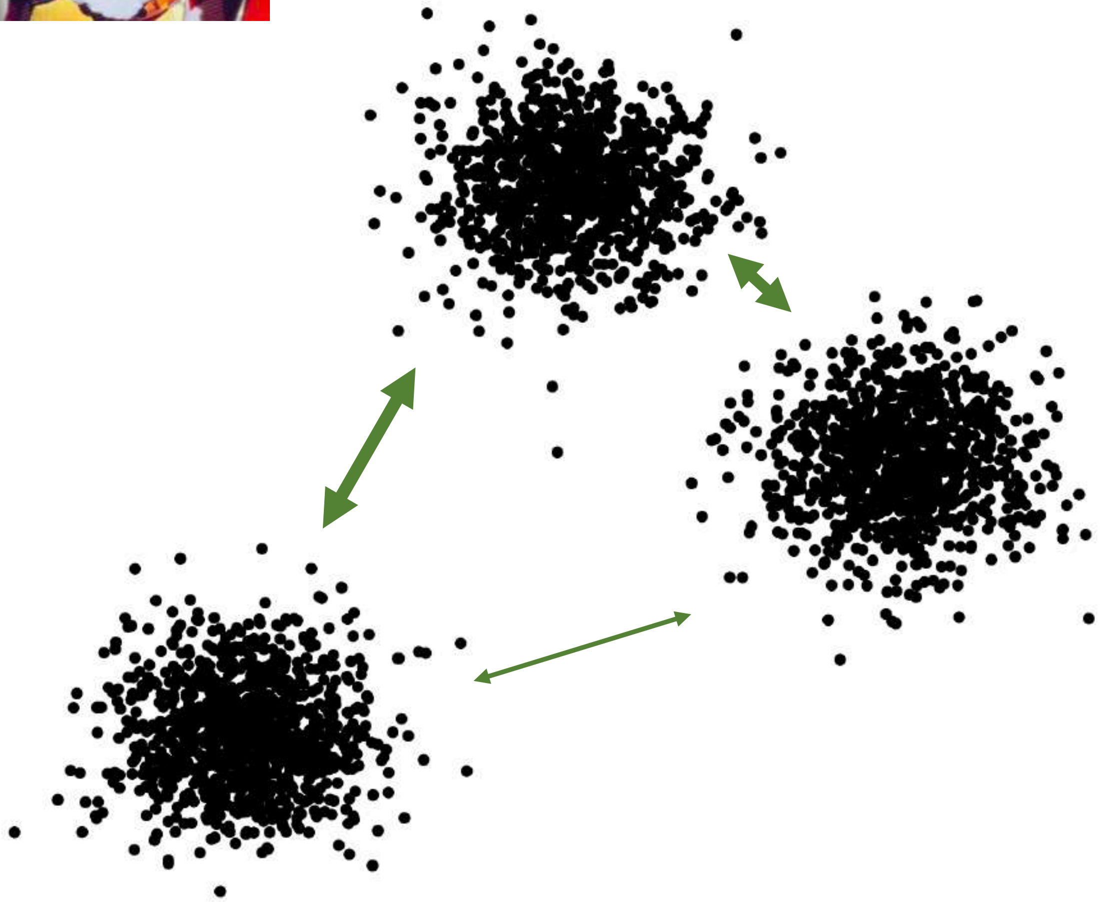
Article | Open Access | Published: 09 March 2022

A large and diverse autosomal haplotype is associated with sex-linked colour polymorphism in the guppy

Josephine R. Paris ✉, James R. Whiting, Mitchel J. Daniel, Joan Ferrer Obiol, Paul J. Parsons, Mijke J. van der Zee, Christopher W. Wheat, Kimberly A. Hughes & Bonnie A. Fraser



Gene Flow



Transfer of alleles among populations.

Can be asymmetric, unidirectional and episodic.

Works against population sub-division.

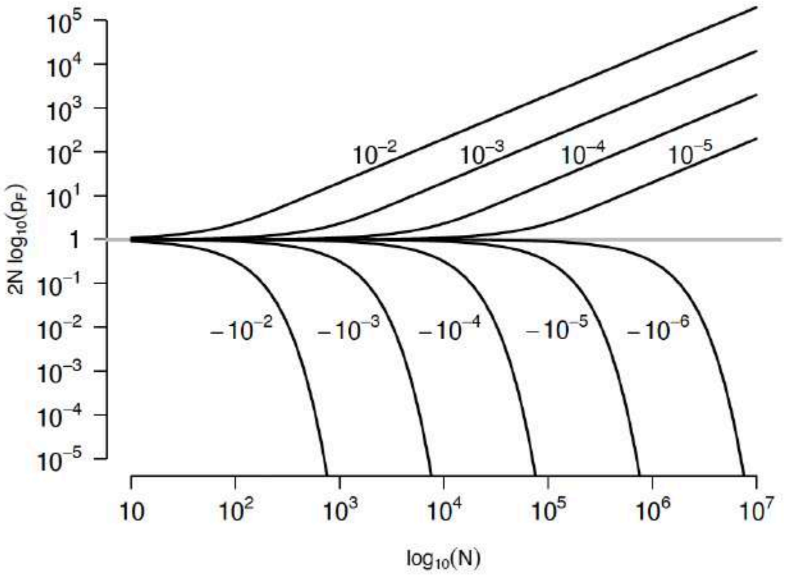
Inflates diversity by increasing effective population size (novel diversity)

Genetic Drift

Stochastic fluctuations in allele frequency

Probabilistic, determined by population size and selection

Drives population subdivision and neutral divergence



Neutral fixation = 1/2N

Recombination

Variation in Recombination Rate and Its Genetic Determinism in Sheep Populations

Morgane Petit,* Jean-Michel Astruc,† Julien Sarry,* Laurence Drouilhet,* Stéphane Fabre,* Carole R. Moreno,* and Bertrand Servin*,1

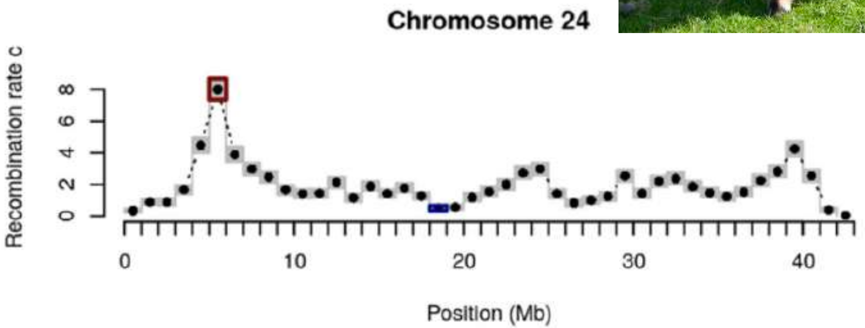
~1 CO per chromosome = more recombination on smaller chromosomes

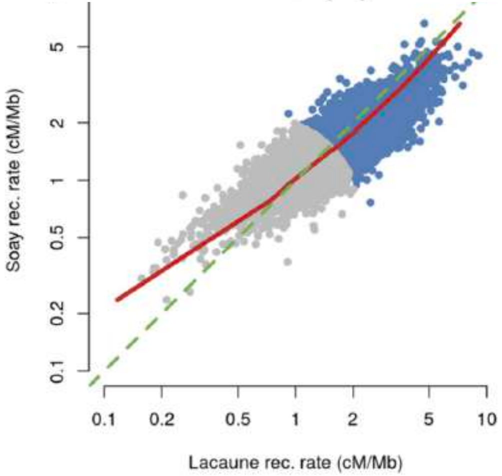
Moderates linkage across the genome

Low recombination resists introgression

Recombination generates novel haplotypes

Conserved landscapes, but variation as well





What determines genetic diversity?

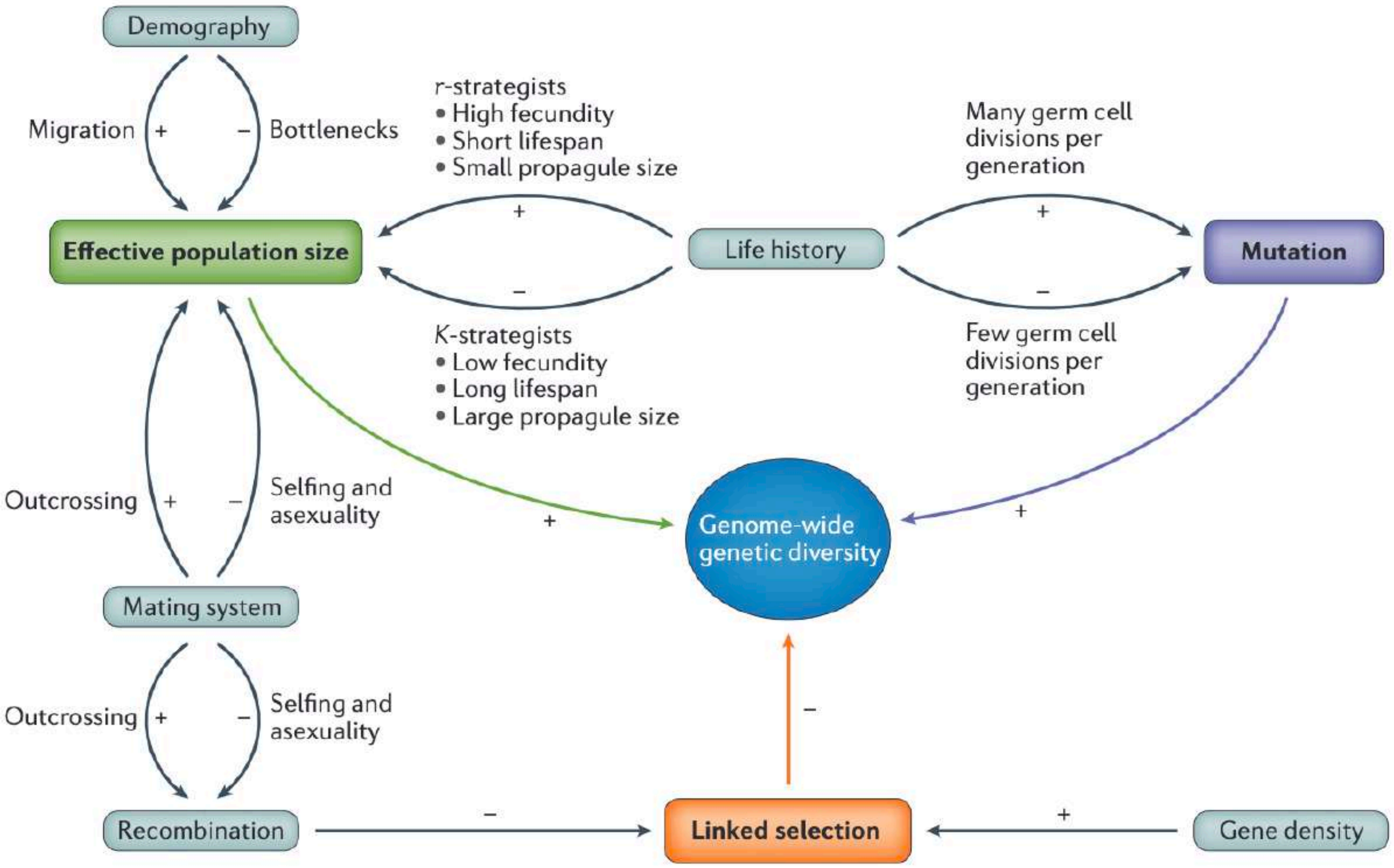


Figure 2 | **Overview of determinants of genetic diversity.** Effective population size, mutation rate and linked selection are the main factors affecting diversity. These factors are in turn governed by several other parameters. The direction of correlation is indicated by the + and – symbols. Selfing, self-fertilization.

Ellegren and Galtier 2017

Applications for population genomics

Conservation genomics
- Understanding connectivity and structure among populations
- Quantifying temporal changes in population diversity and extinction risk

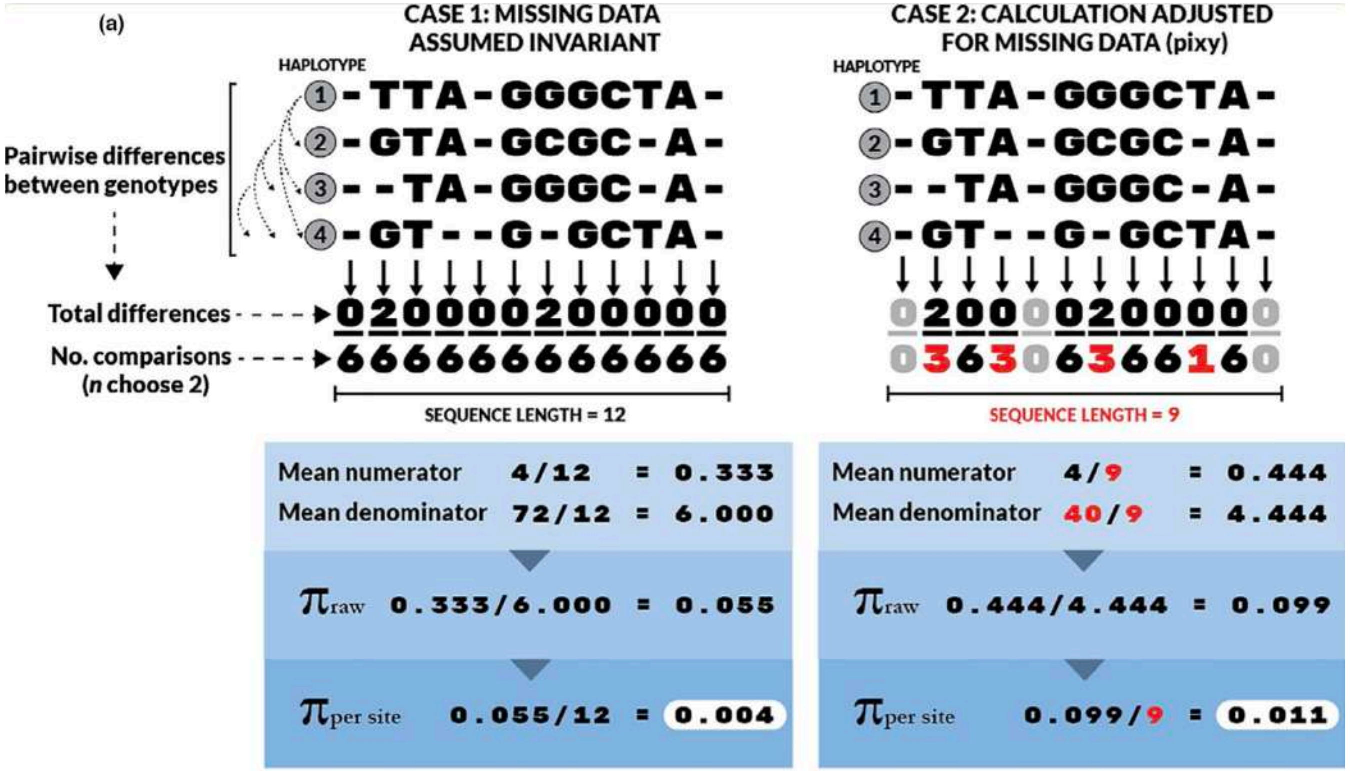
Agricultural genomics
- Identifying novel variation within wild crop relatives
- Genomic selection and guided breeding

Medical genomics
- Tracking pathogen evolution (e.g. Covid!)
- Understanding disease risk of individuals and populations

Evolutionary biology
- Identifying candidate loci under selection and adaptive genes
- Identifying barriers to gene flow and understanding speciation
- Understanding the evolutionary history of species

PopGen Toolkit – Summary Statistics

Pi (nucleotide diversity)



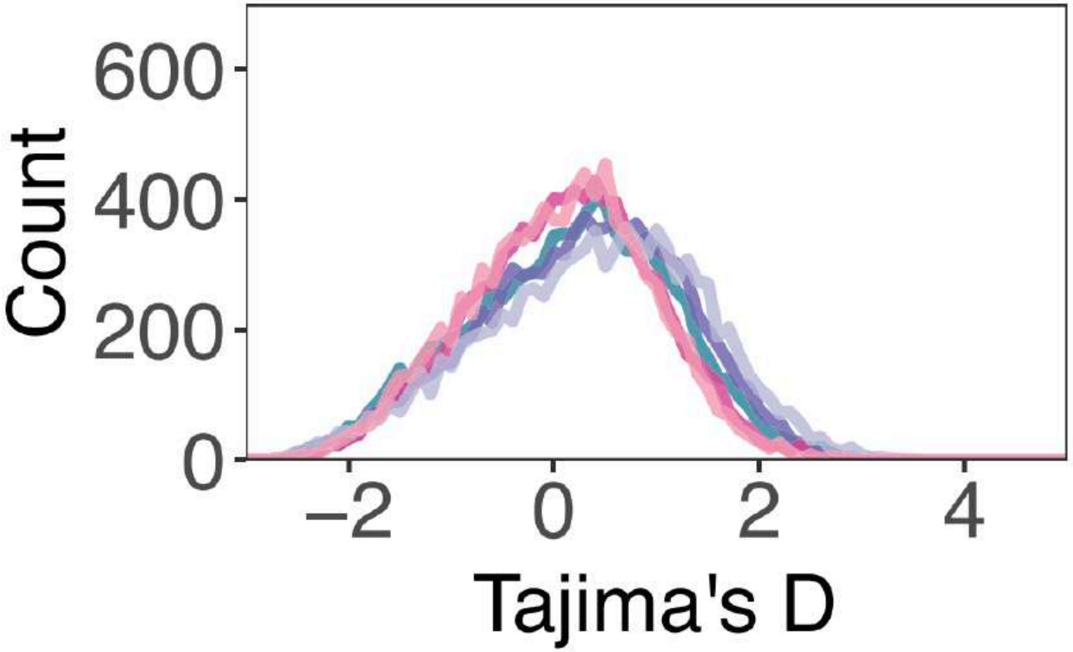
PIXY: Korunes and Samuk 2021

PopGen Toolkit – Summary Statistics

Pi (nucleotide diversity)

Tajima's D

$$D = \frac{\pi - S/a_1}{\sqrt{V}}$$



van der Zee et al. 2022

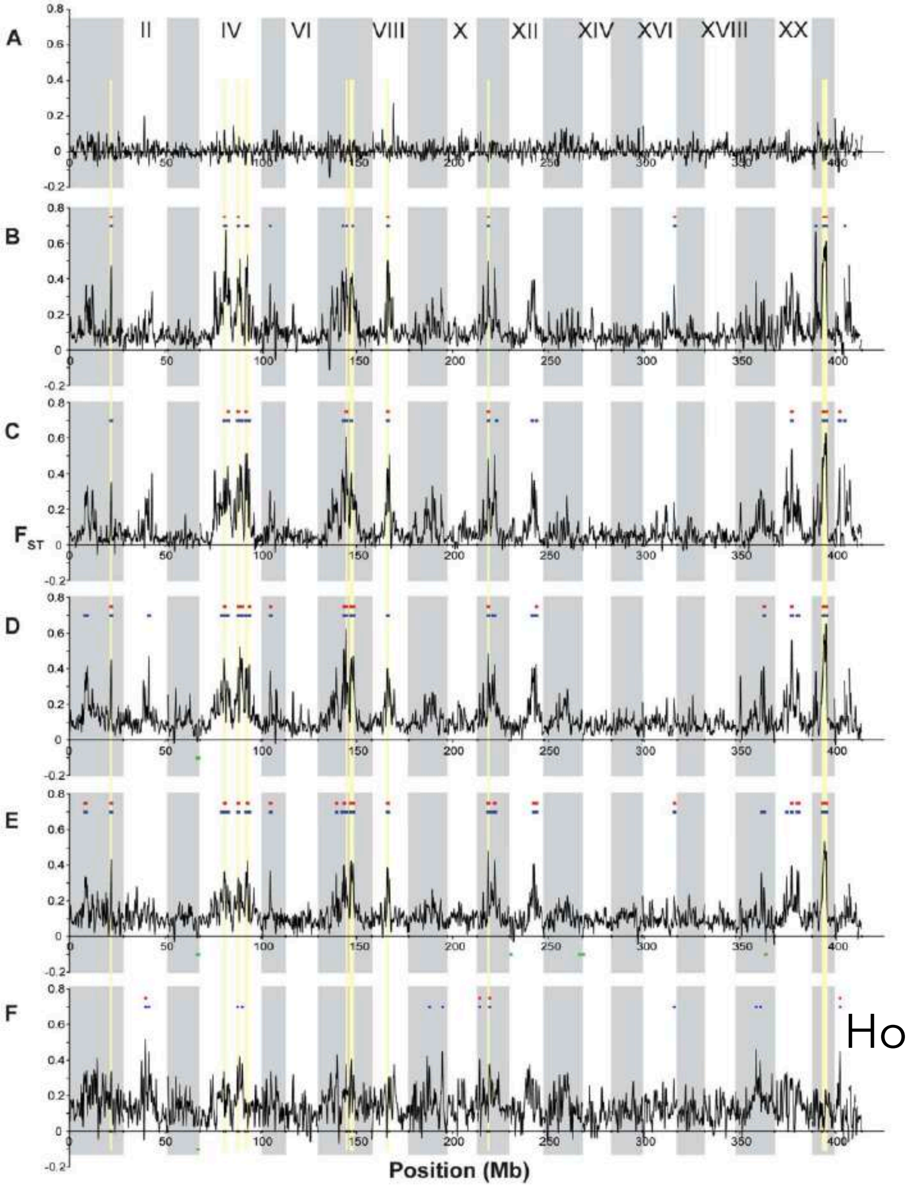
PopGen Toolkit – Summary Statistics

Pi (nucleotide diversity)

Tajima's D

F_{ST}

$$\frac{\pi_{\text{Between}} - \pi_{\text{Within}}}{\pi_{\text{Between}}}$$



Hohenlohe et al. 2010

PopGen Toolkit – Summary Statistics

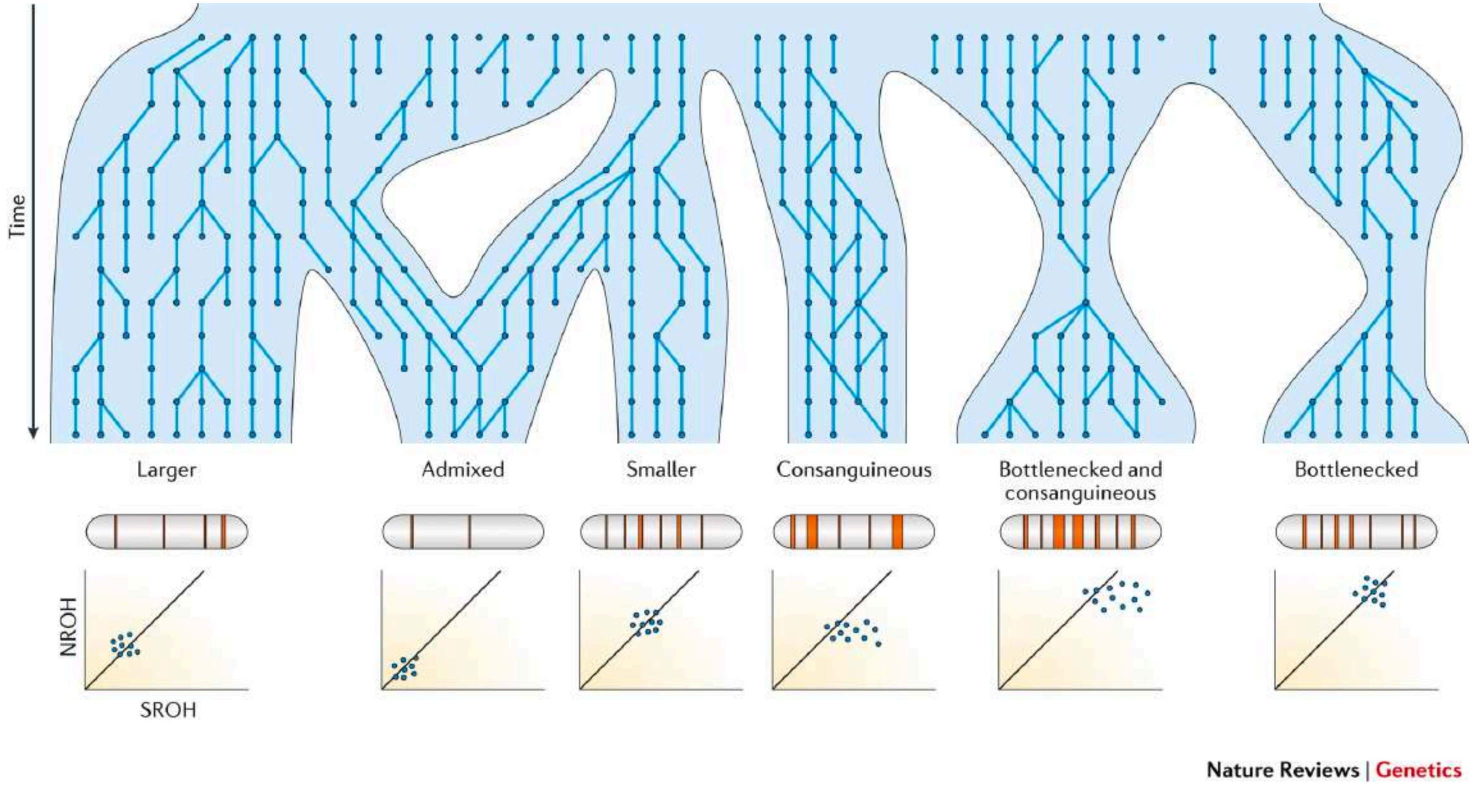
Pi (nucleotide diversity)

Tajima's D

F_{ST}

ROH

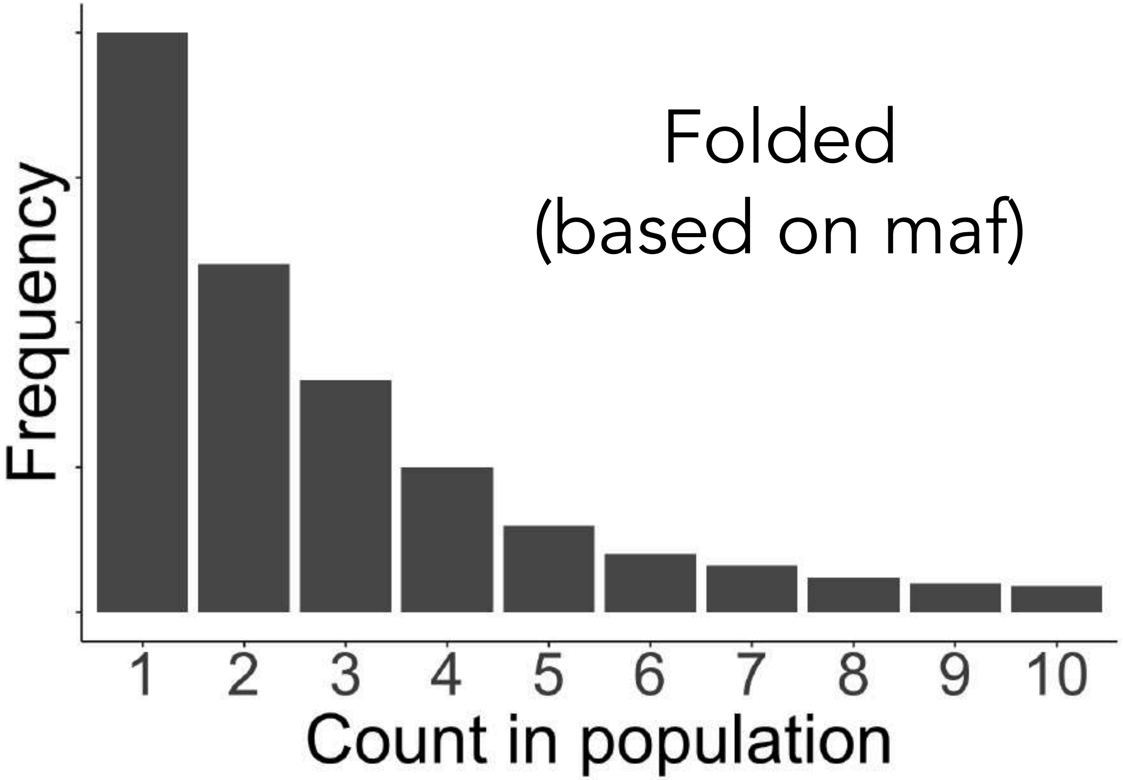
(Runs of Homozygosity)

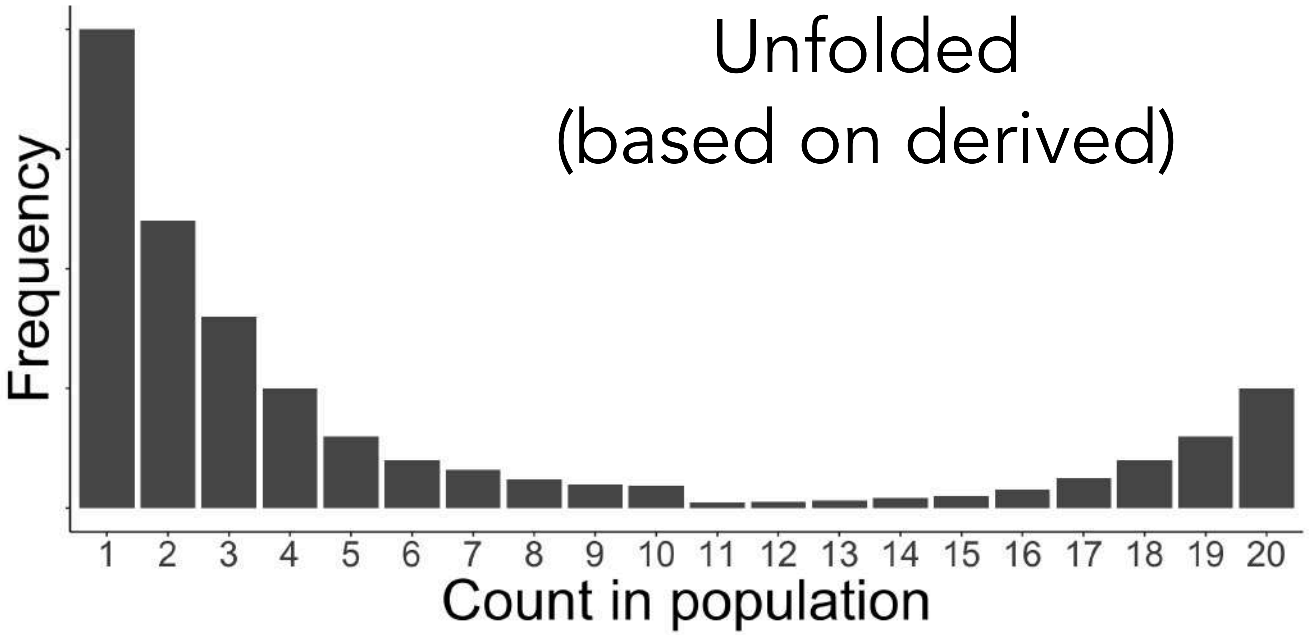


Ceballos et al. 2018

PopGen Toolkit – Site Frequency Spectrum

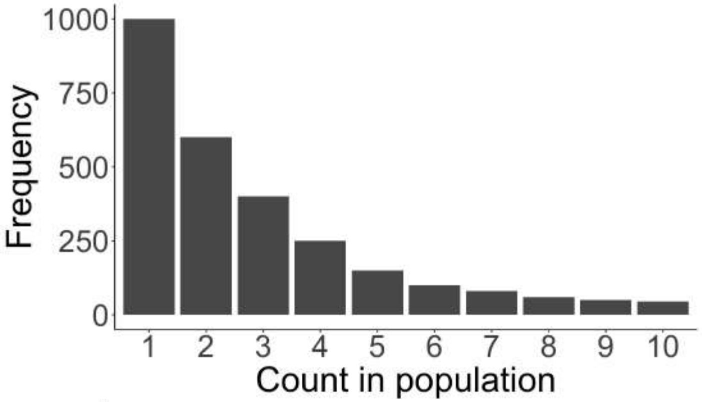
A summary of the frequency of allele counts within a population


Folded
(based on maf)

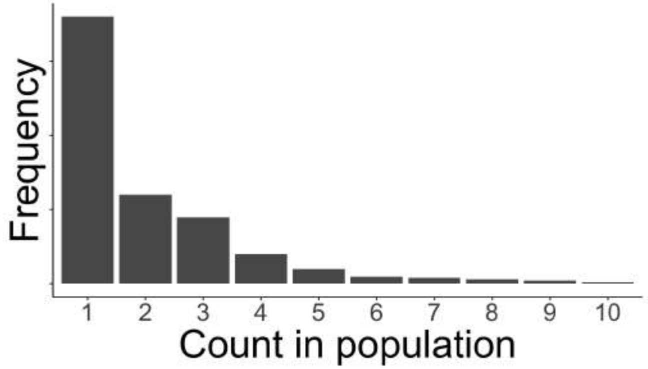

Unfolded
(based on derived)

Do we know ancestral/derived allele?

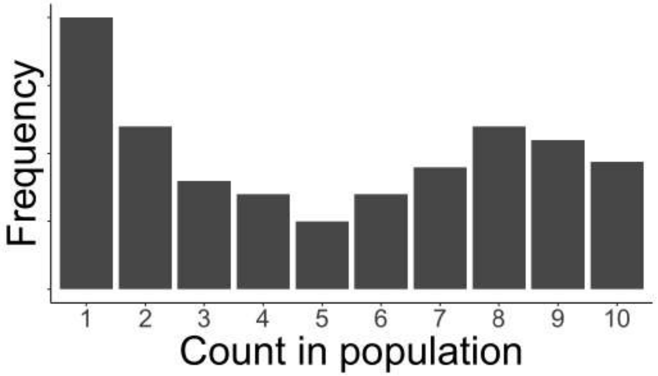
PopGen Toolkit – Site Frequency Spectrum



Folded - Neutral

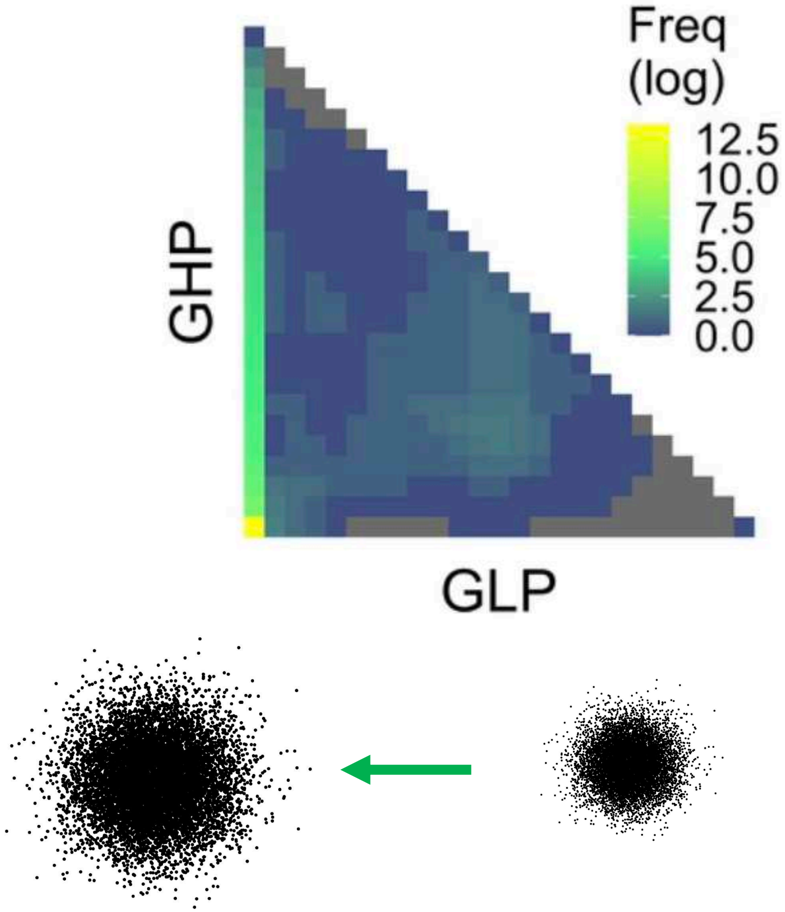


Folded – Selection

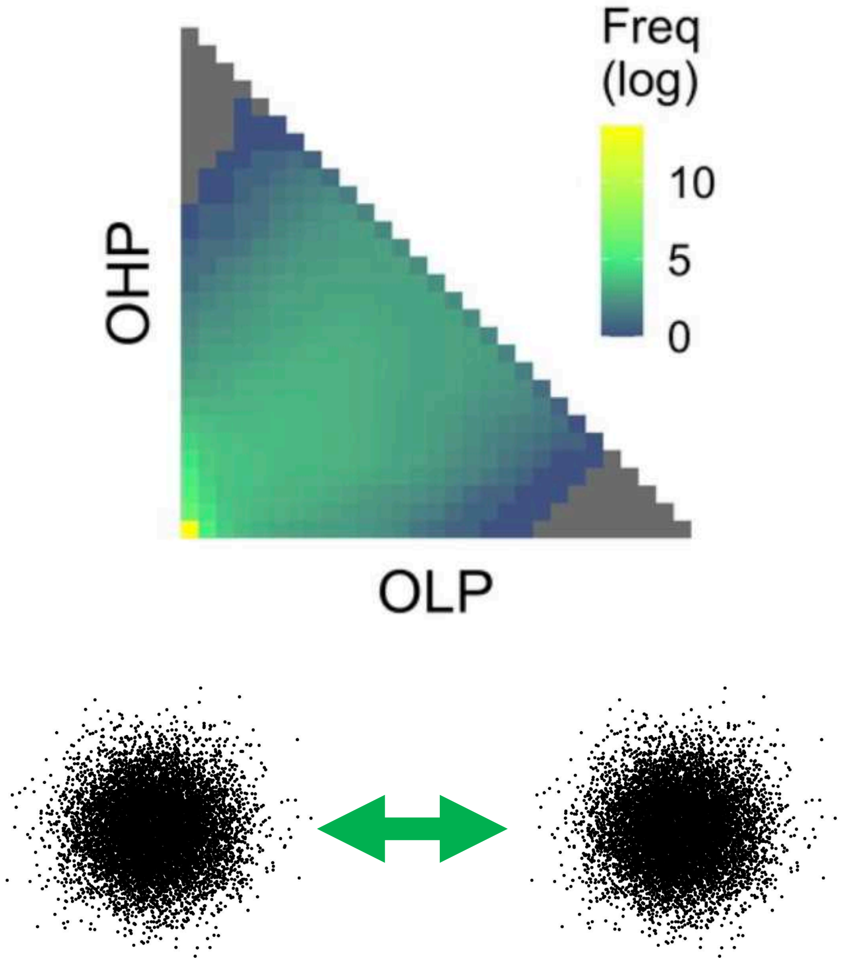


Folded – Balancing

PopGen Toolkit – Site Frequency Spectrum



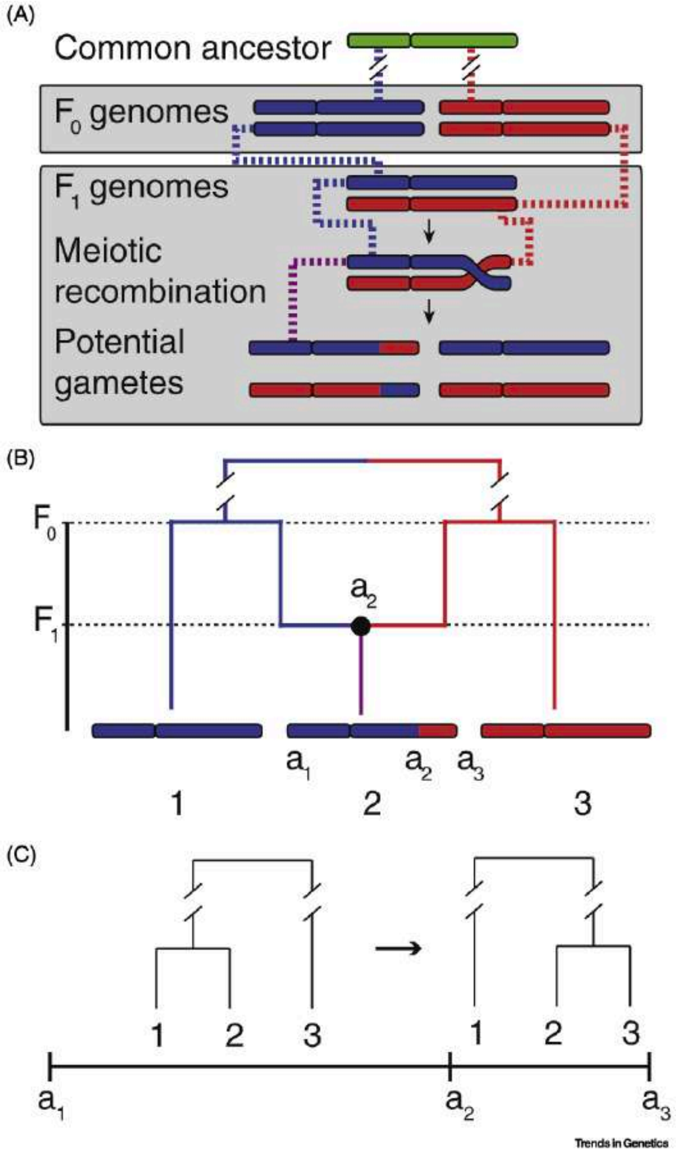
2D/Joint



PopGen Toolkit – The future… (is now)

Ancestral recombination graphs

- ARGweaver
- RELATE
- tsinfer



Hejase et al 2020

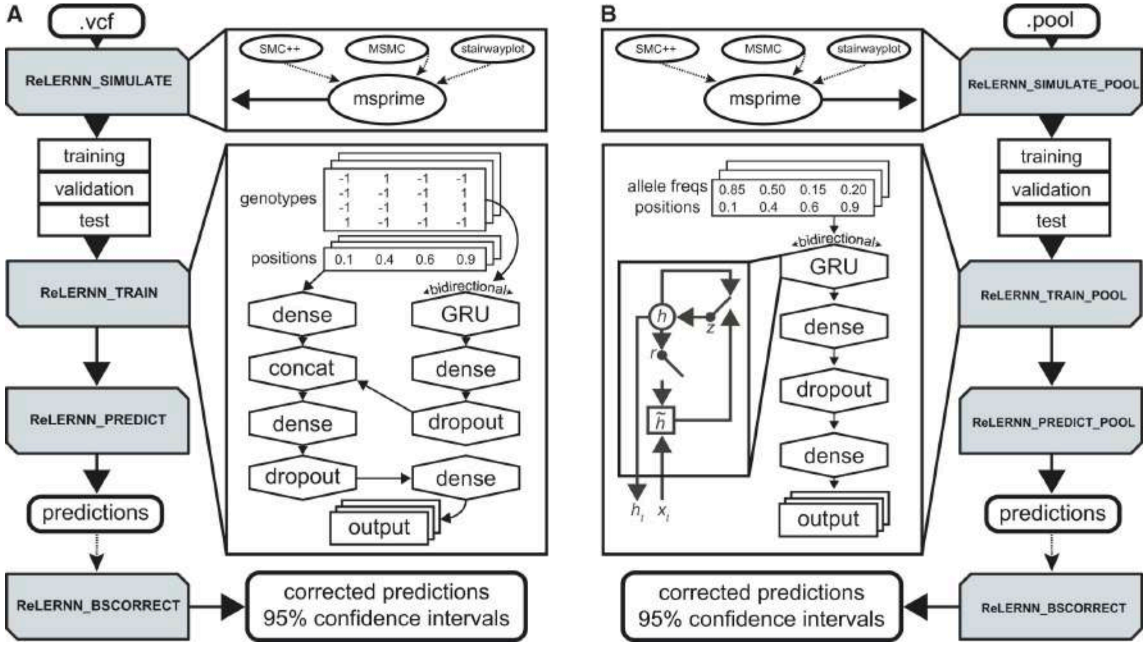
PopGen Toolkit – The future… (is now)

Ancestral recombination graphs

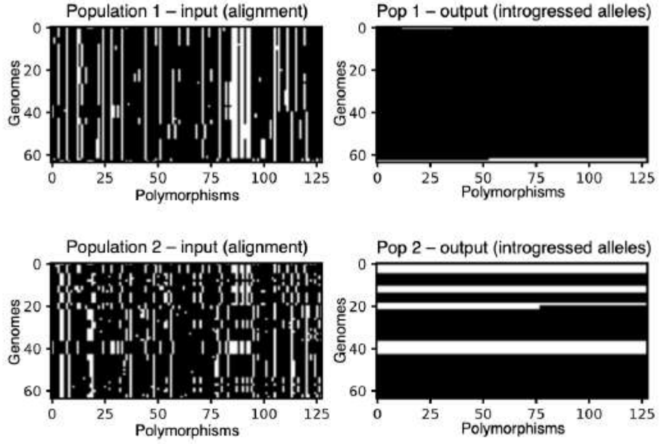
- ARGweaver
- RELATE
- tsinfer

Machine-learning

- Classification algorithms (e.g. diploS/HIC)
- Neural networks (e.g. RELERNN)
- Image processing (e.g. IntroUNET)



Adrion et al 2020



Ray et al 2023

PopGen Toolkit – The future… (is now)

Ancestral recombination graphs

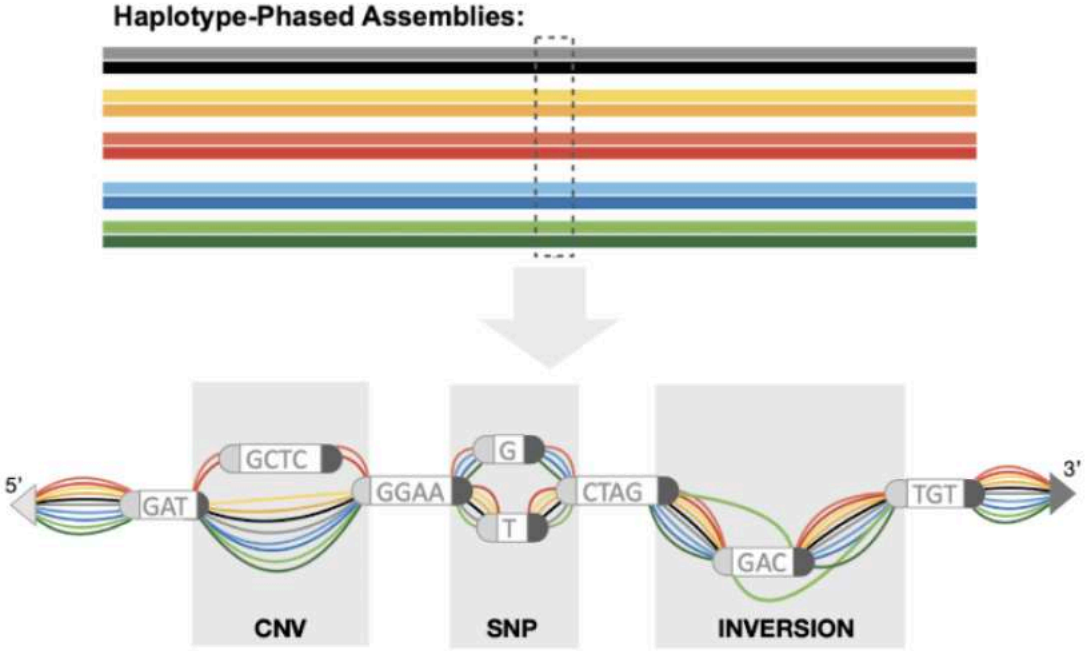
- ARGweaver
- RELATE
- tsinfer

Machine-learning

- Classification algorithms
- Neural networks (e.g. RELERNN)
- Image processing (e.g. IntroUNET)

Pangenomics/SV (beyond the SNP…)

- How applicable are our models?
- What is a population/species in a pangenome sense?



Human Pangenome Reference Consortium

Practical Session

Understanding the (pretend) evolutionary history of Scottish wild cat re-introduction to England

Using a simulated VCF, we'll work through a simple popgen workflow covering:

- Describing population structure
  - PCA
  - DAPC
- Quantifying population diversity
  - Nucleotide diversity
- Identifying selective sweeps
  - SweepFinder2
- Testing adaptive introgression
  - D-Suite (ABBA-BABA / D-statistics)