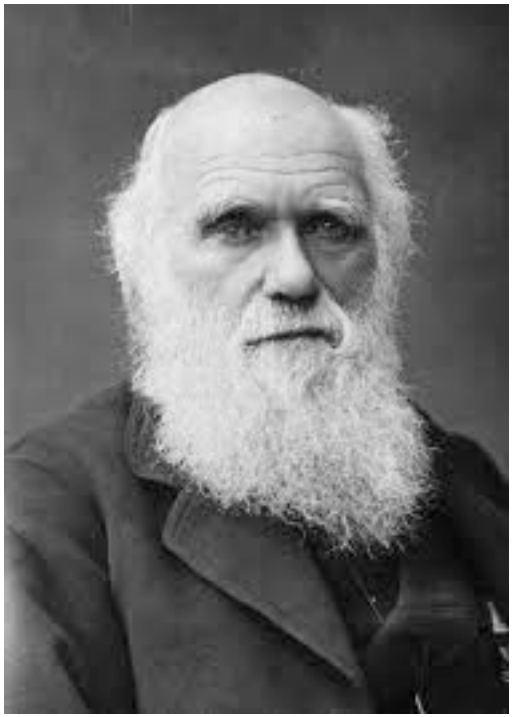
Lecture 3: Experimental Evolution

Evolution, Fast and Slow

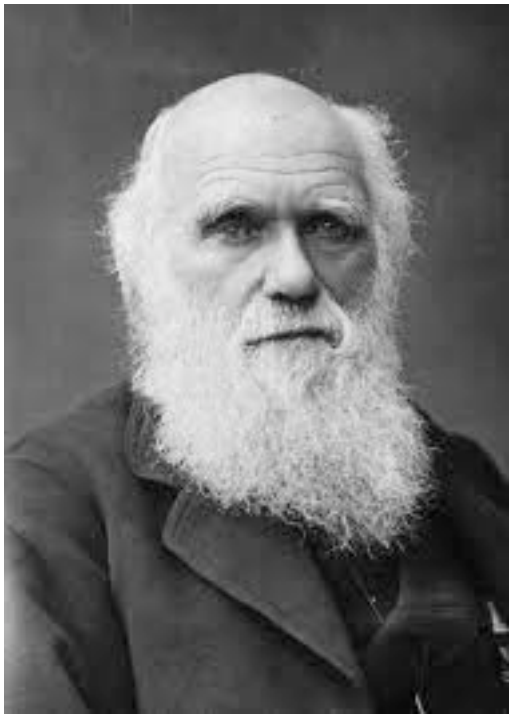


Natural selection is slow

"Natural selection will always act very slowly, often only at long intervals of time, and generally on only a few inhabitants of the same region at the same time"

The Origin of Species, 1859

Evolution, Fast and Slow



Artificial selection can produce substantial changes

"Slow though the process of selection may be, if feeble man can do much by his powers of artificial selection, I can see no limit to the amount of change …which may be effected in the long course of time by nature's power of selection"

The Origin of Species, 1859

Evolution, Fast and Slow

Animal and plant breeding have brought about substantial phenotypic changes across the course of human history

E. A. Ostrander, NEJM, 2012

Evolution, Fast and Slow

Animal and plant breeding have brought about substantial phenotypic changes across the course of human history

Teosinte

↓

Modern maize

Evolution, Fast and Slow

Trait-selection experiments have shown substantial phenotypic changes on a time-scale of decades

The Illinois Long-term Evolution Experiment

Began in 1896:

Aim to improve the characteristics of maize

Protein content : Animal feed

Oil content : Corn oil

Initially 163 ears of corn. Analyse for chemical content.

Use top 24 ears to seed a high protein/oil crop

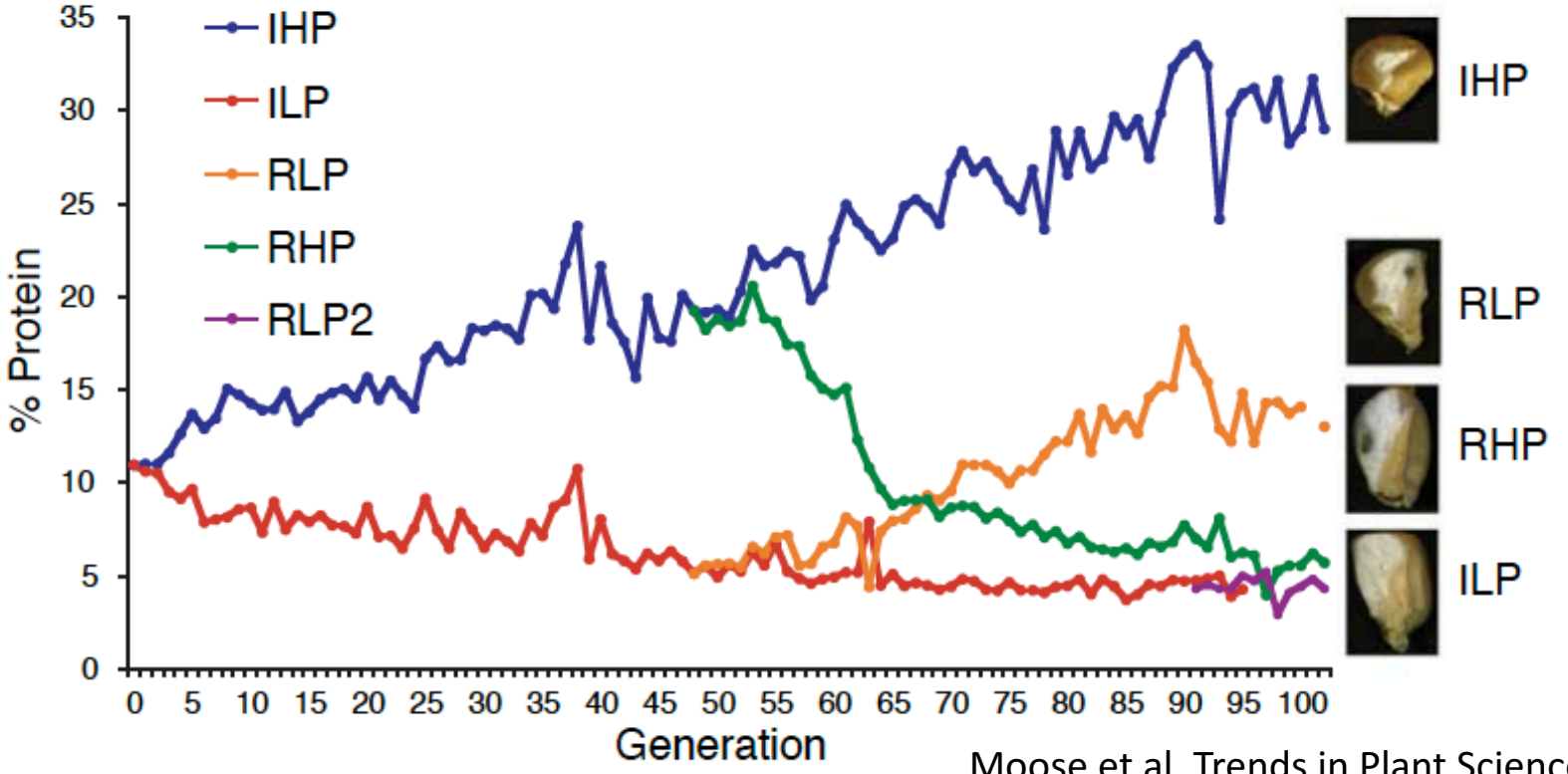
Use bottom 12 ears to seed a low protein/oil crop

Repeat each year

Evolution, Fast and Slow

The Illinois Long-Term Selection Experiment

Select for high and low protein content

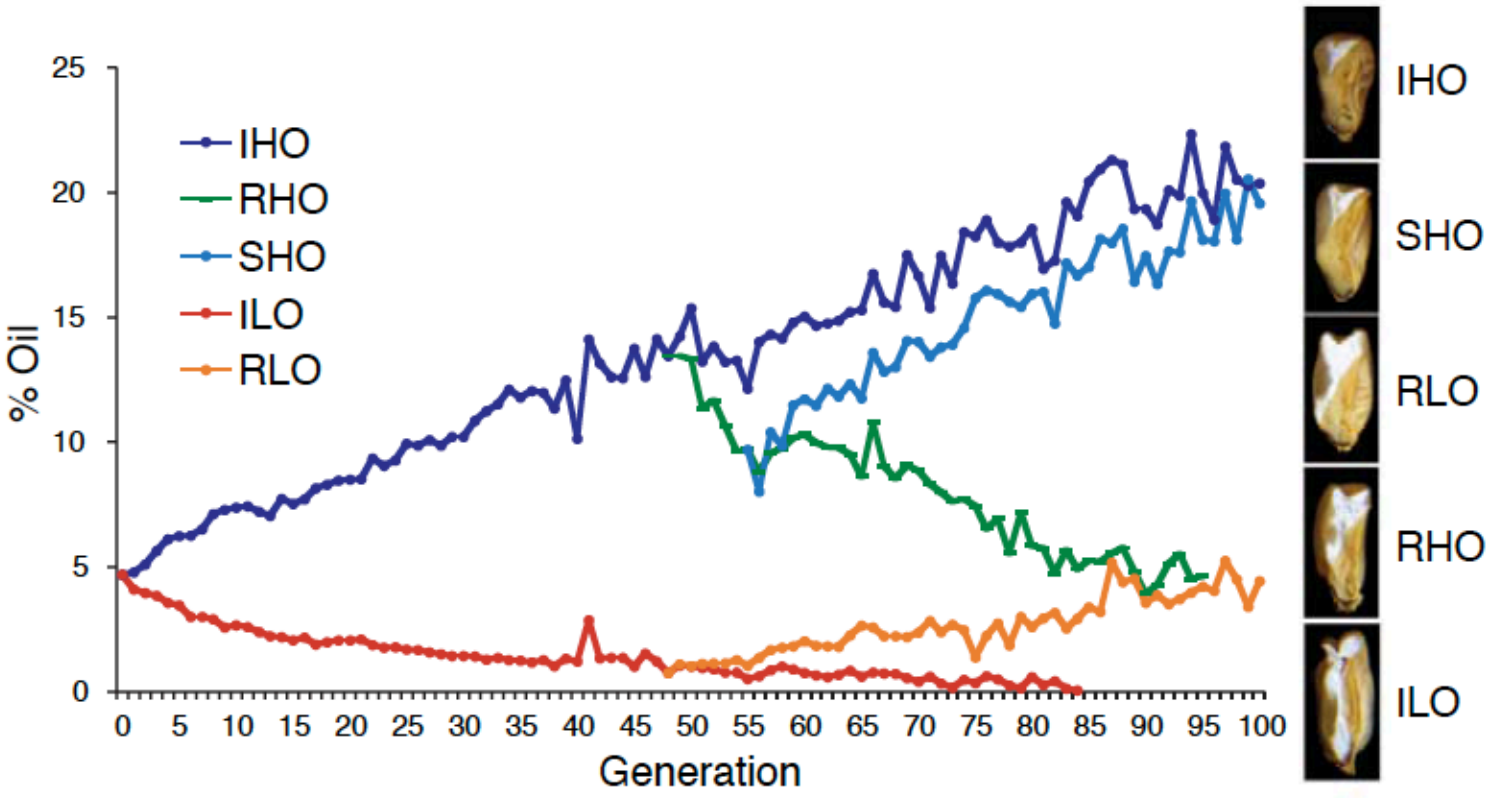


Moose et al, Trends in Plant Science, 2004

Evolution, Fast and Slow

The Illinois Long-Term Selection Experiment

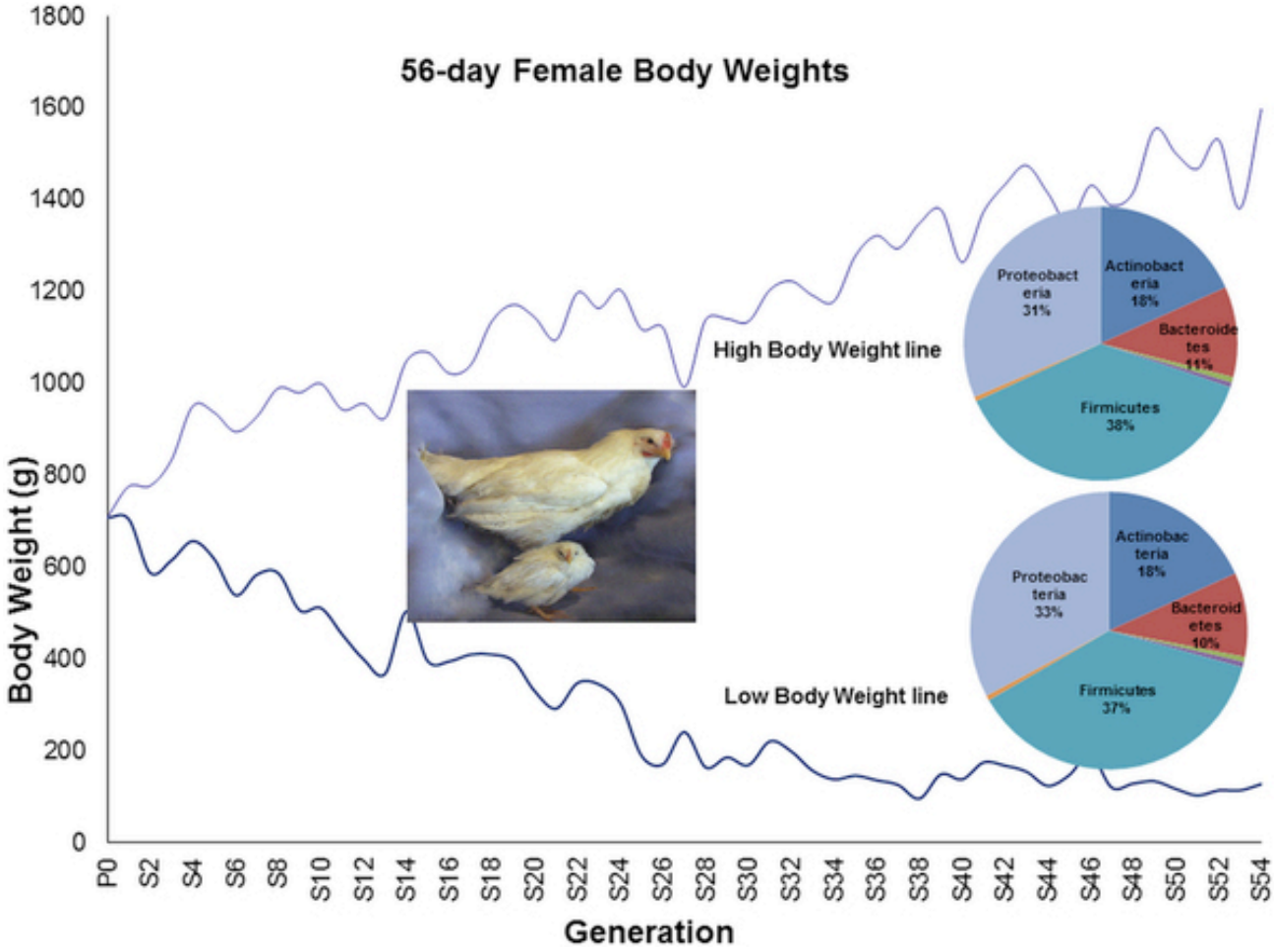
Select for high and low oil content



Moose et al, Trends in Plant Science, 2004

Evolution, Fast and Slow

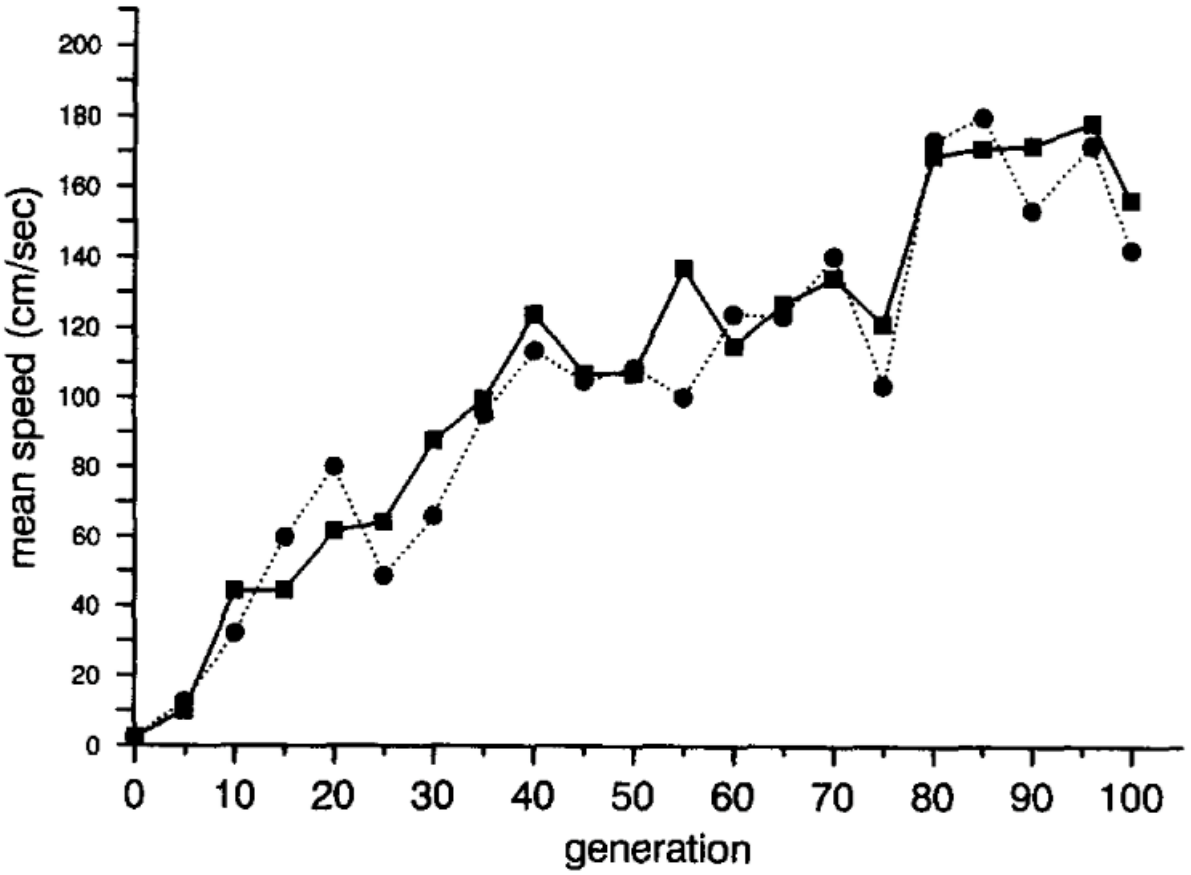
Selection for body weight in chickens



Zhao et al, Nature Scientific Reports, 2012

Evolution, Fast and Slow

Selection for flight speed in *Drosophila melanogaster*



Weber, Genetics, 1996

Evolutionary experiments

Definition

The study of evolutionary changes occurring in experimental populations as a consequence of conditions (environmental, demographic, genetic, social, and so forth) imposed by the experimenter

Kawecki et al, Trends in Ecology and Evolution, 2012

Evolutionary experiments

Experiments can be conducted in either wild or laboratory conditions

Pros of lab evolution:

Environment is known and can be controlled

Experiments can be replicated

Cons of lab evolution:

Lab conditions are unnatural (artificial light, food, climate)

-Adaptation to lab conditions may occur

Phenotypes which would not be viable in the wild may occur

-Lack of predators etc.

What happens in the real world cannot always be replicated in the lab

-Drosophila response to climatic variation

Evolutionary experiments

Example: Laboratory evolution of Escherichia coli

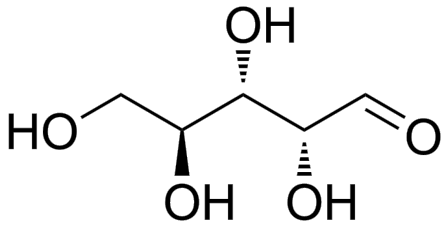
Evolutionary experiments

Example: Lab evolution of Escherichia coli

Begun in 1988. Twelve near-identical populations of E. coli.

Differ in ability to metabolise arabinose

6 Ara^+, 6 Ara^-

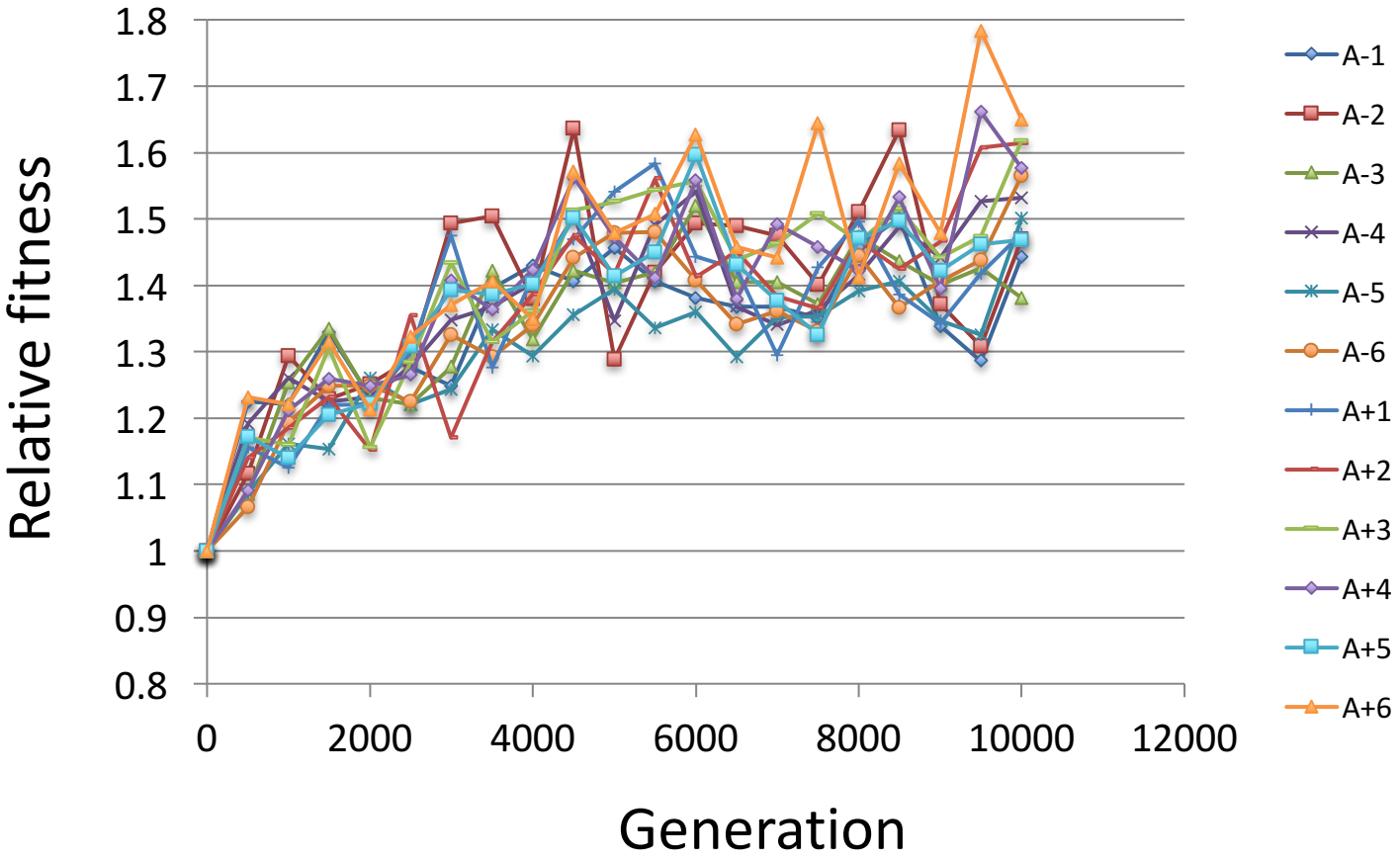


Populations grow over time: each day transfer 1% of each population to a new flask.

Evolutionary experiments

Example: Lab evolution of Escherichia coli

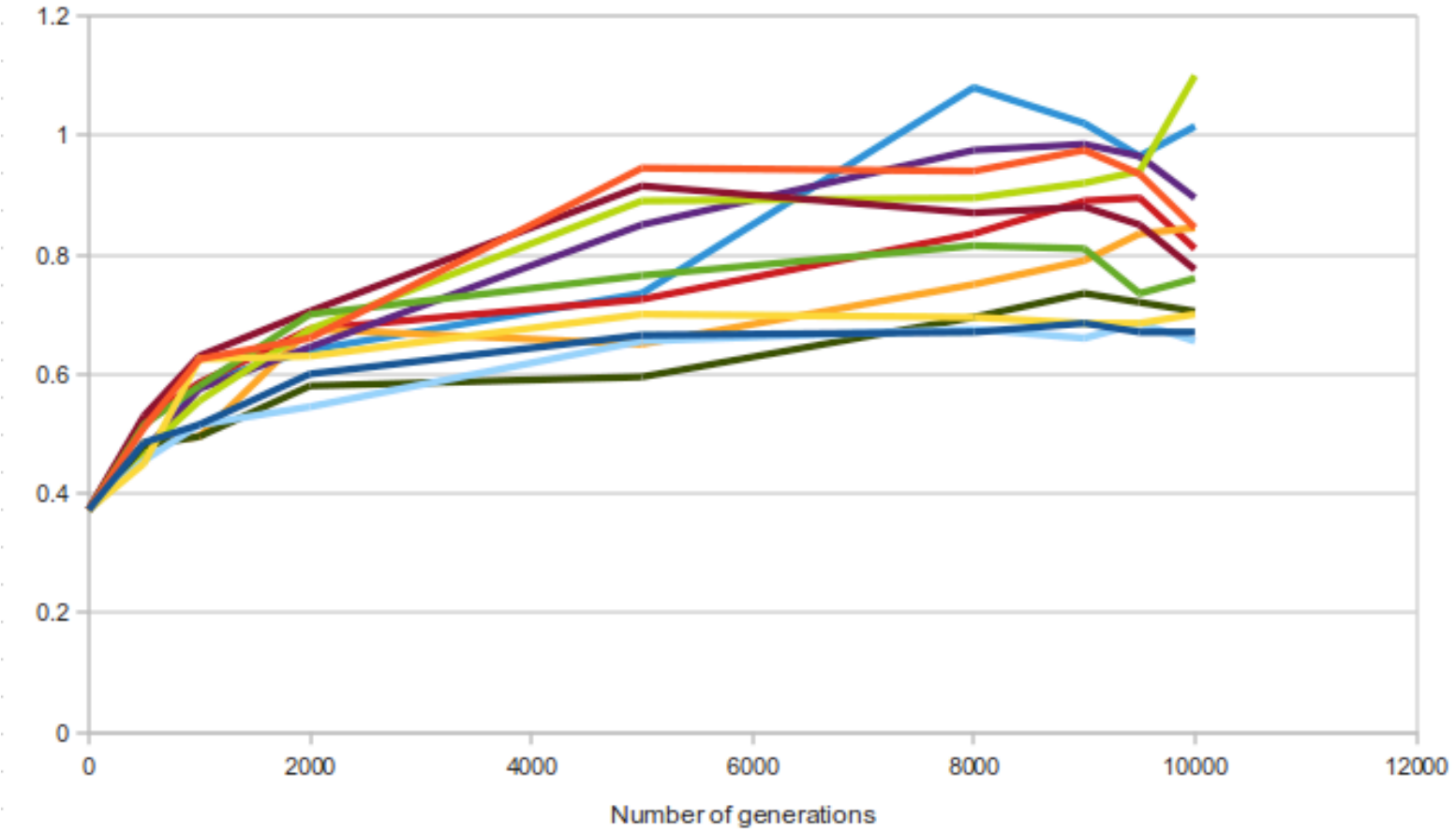
Change in fitness over first 10,000 generations



Evolutionary experiments

Example: Lab evolution of Escherichia coli

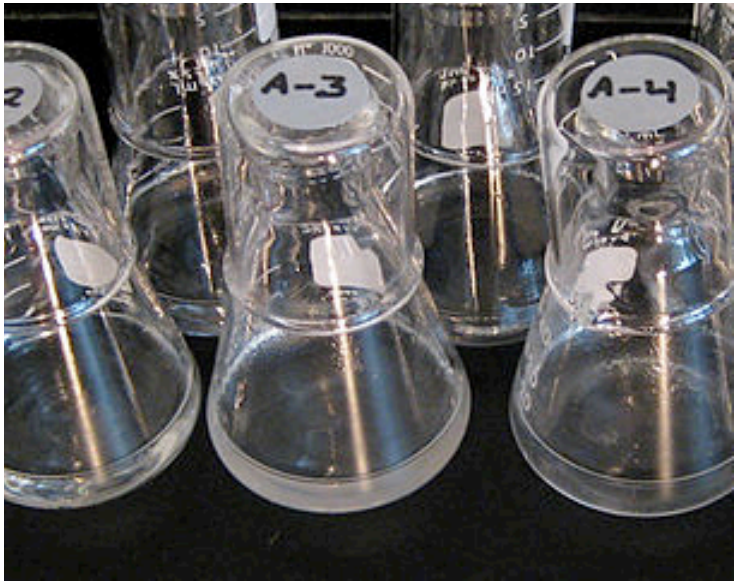
Change in cell size over first 10,000 generations



Evolutionary experiments

Evolution of citrate metabolism

One population has evolved the ability to digest citrate



More available nutrient: leads to an increase in the population size

Genome sequencing shows the underlying mechanisms of evolution

Photo credit: *Brian Baer and Neerja Hajela*

Evolutionary experiments

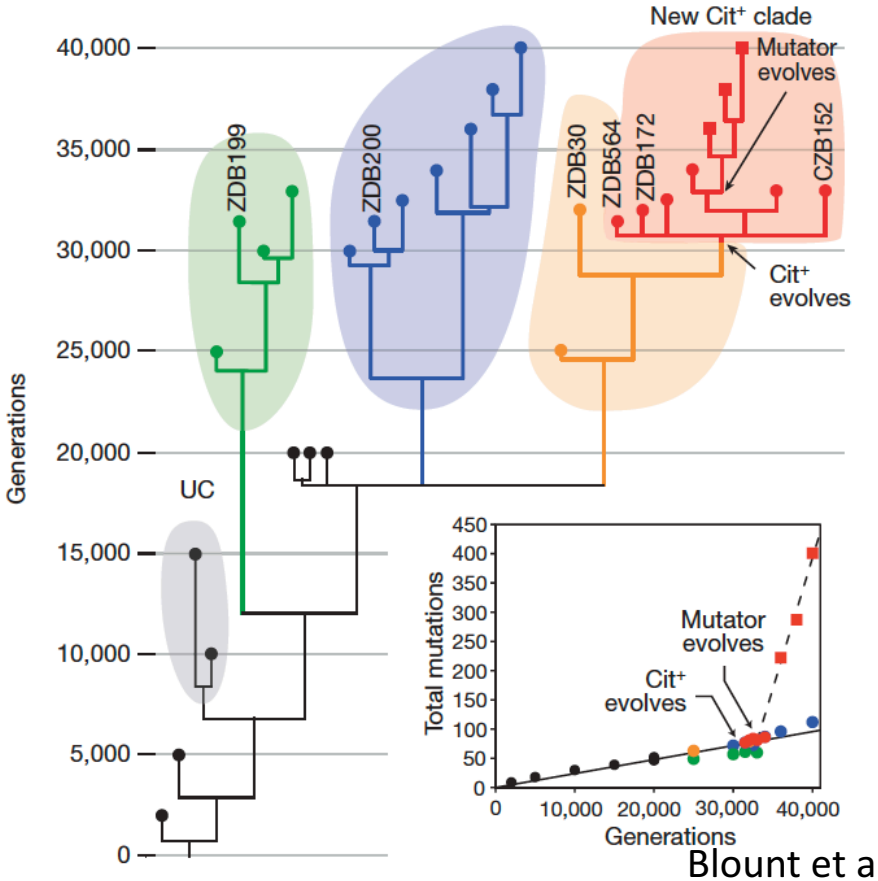
Evolution of citrate metabolism

Phylogenetics shows a structured population evolving over time

Identify three consistent clades persisting over time.

The ability to metabolise citrate occurred in one clade.

Following this, a mutator phenotype emerged.



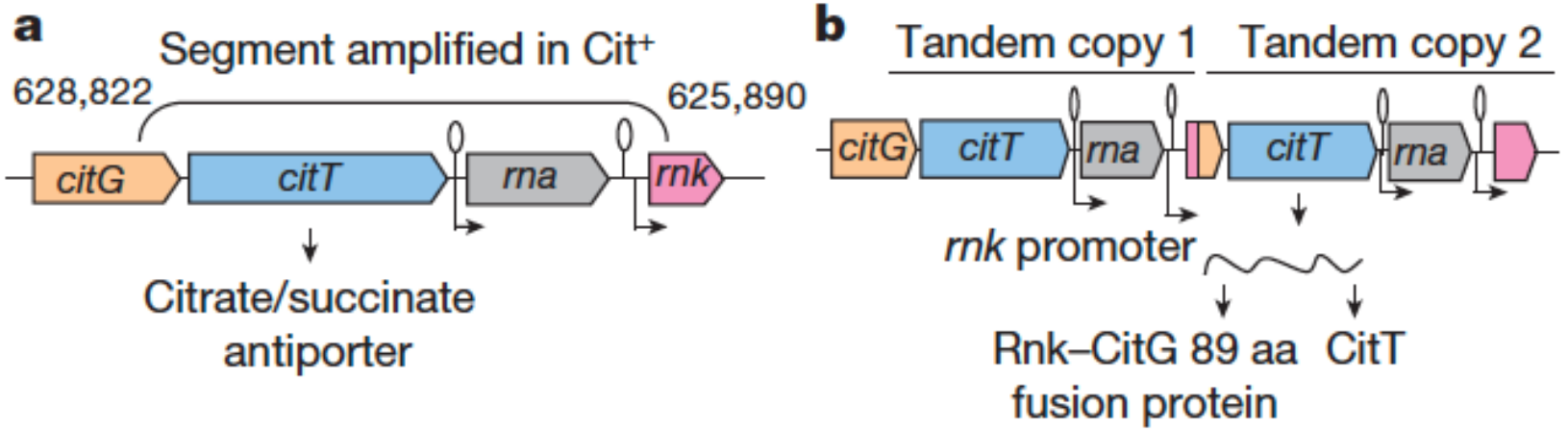
Blount et al, Nature, 2012

Evolutionary experiments

Evolution of citrate metabolism

Ability to metabolise citrate emerged from a duplication event

Duplication of part of a promoter leads to the expression of citT gene



Amplification produces limited metabolism of citrate
Conveys slight fitness advantage

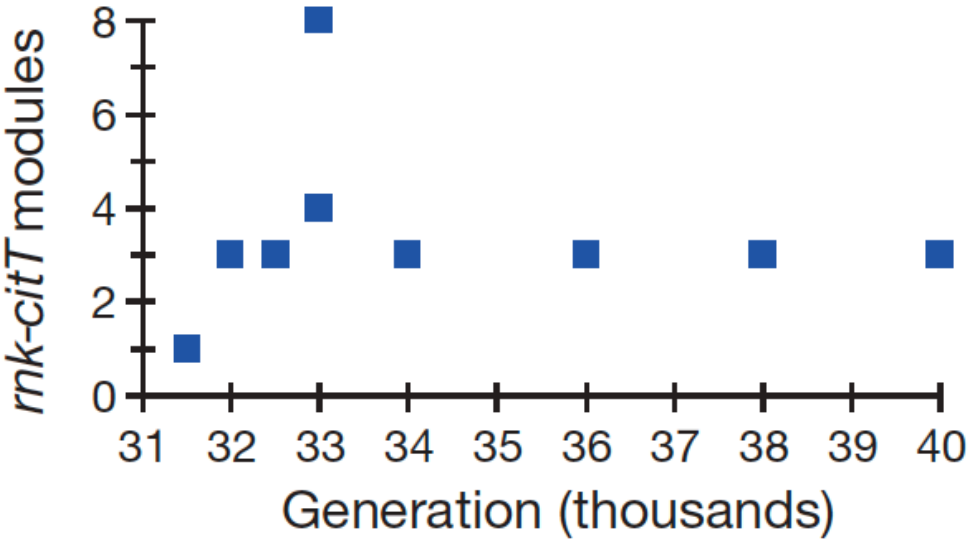
Blount et al, Nature, 2012

Evolutionary experiments

Evolution of citrate metabolism

Further duplication events are seen

Number of copies of the repeated module is unstable



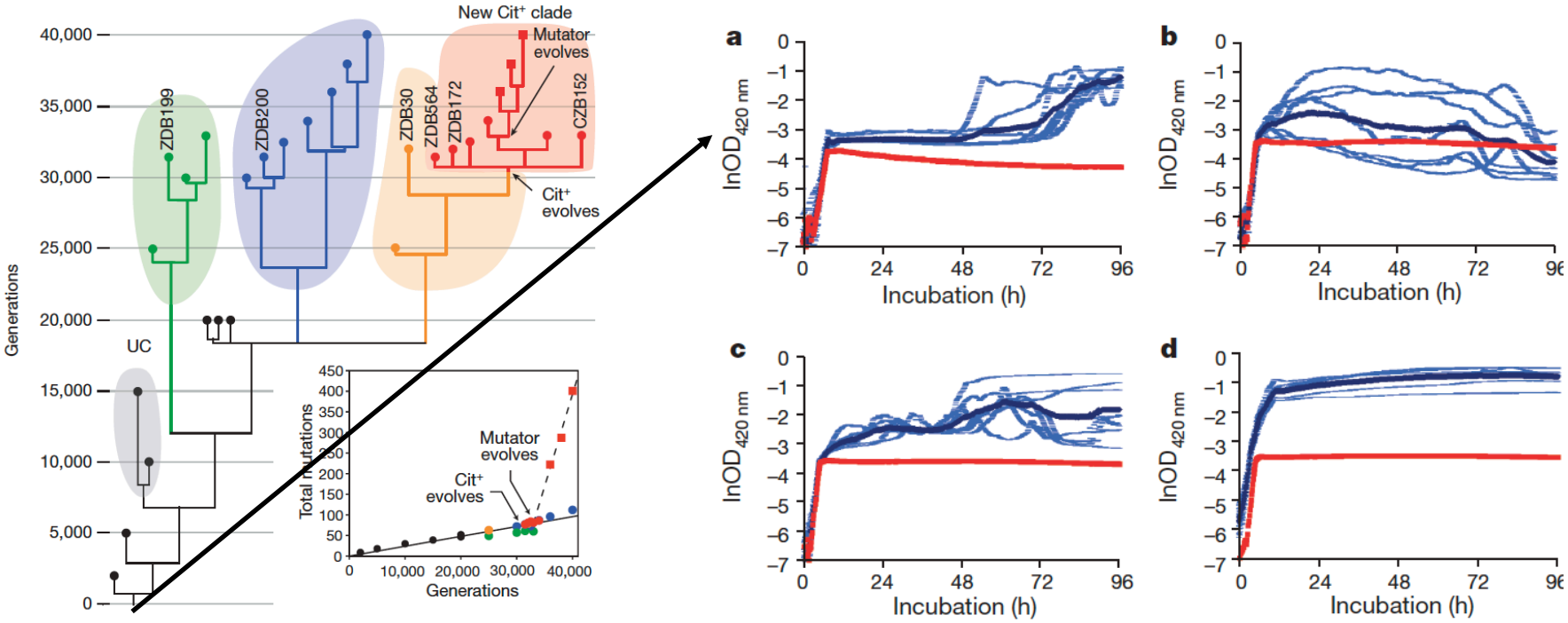
More copies : Greater metabolic ability (+ greater instability?)
Reaches equilibrium at three copies of the promoter/gene unit

Blount et al, Nature, 2012

Evolutionary experiments

Evolution of citrate metabolism

Benefit of module is dependent on strain background : Epistasis



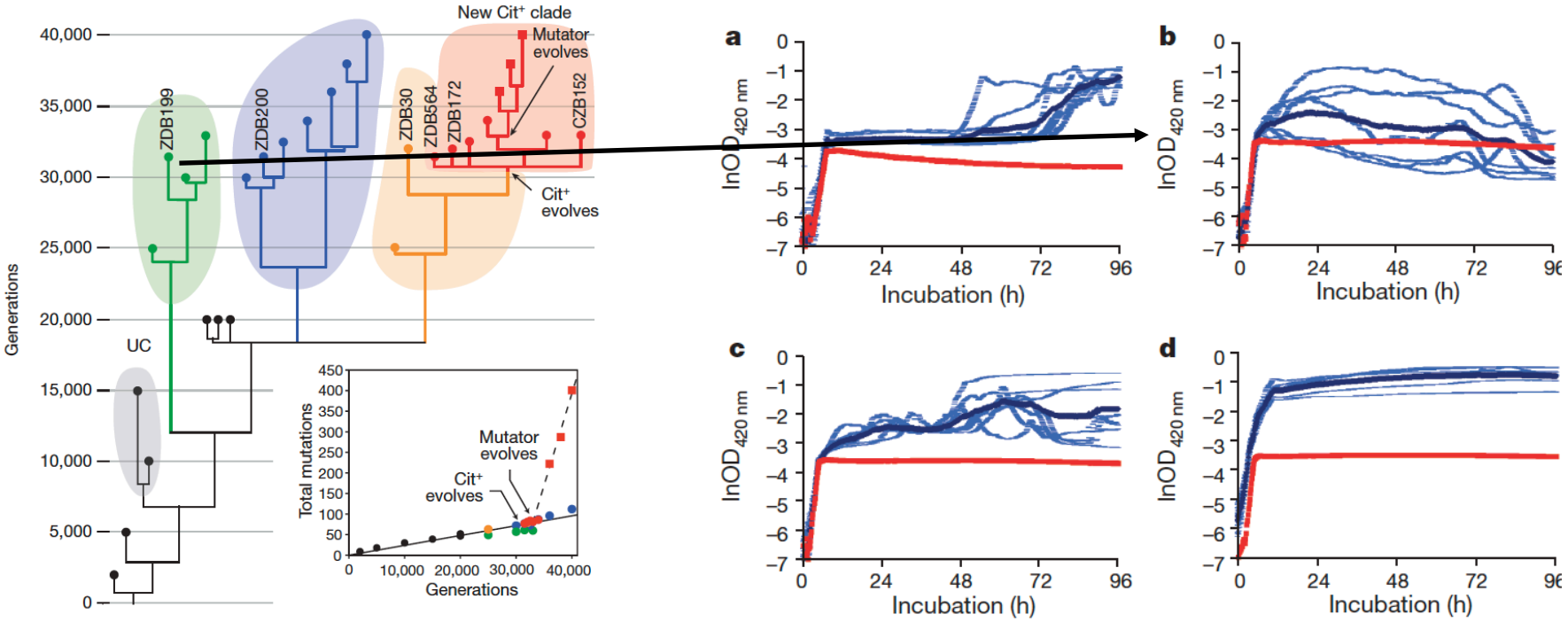
Previous mutations in the bacteria led to the module being able to convey a beneficial effect

Blount et al, Nature, 2012

Evolutionary experiments

Evolution of citrate metabolism

Benefit of module is dependent on strain background : Epistasis



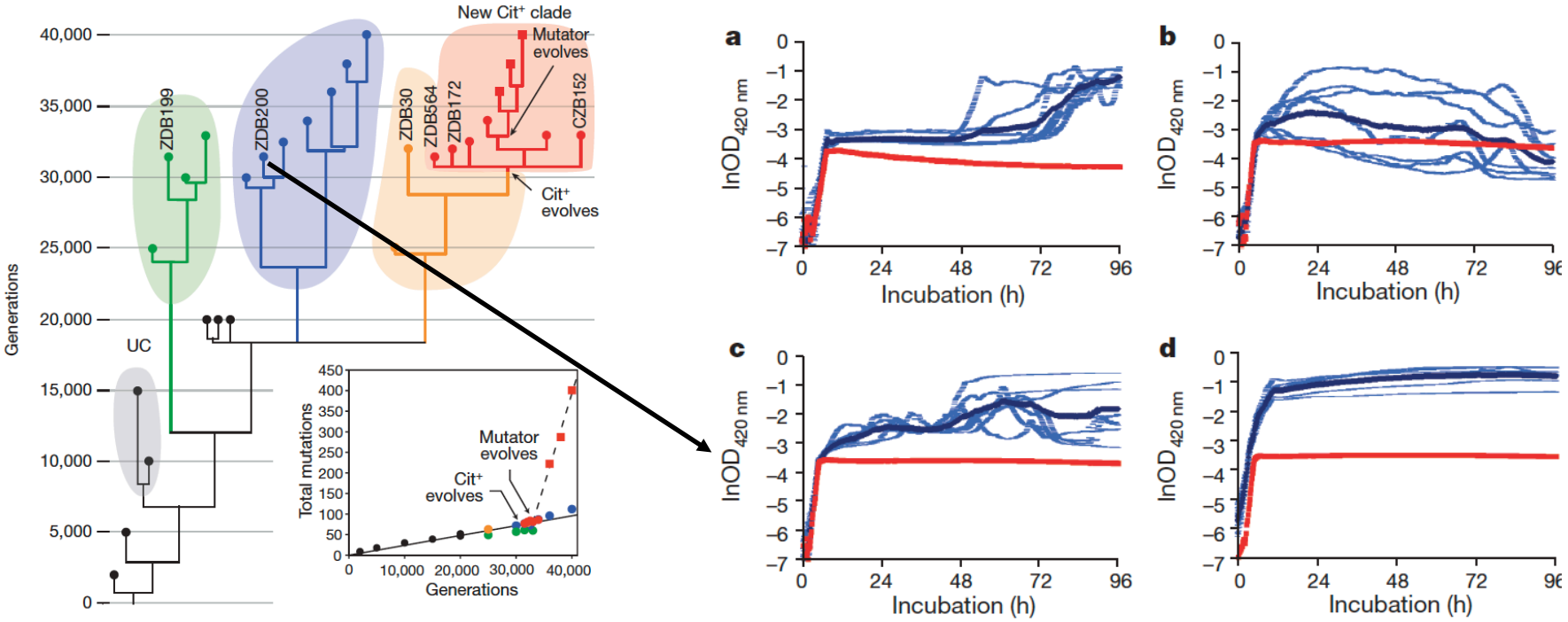
Previous mutations in the bacteria led to the module being able to convey a beneficial effect

Blount et al, Nature, 2012

Evolutionary experiments

Evolution of citrate metabolism

Benefit of module is dependent on strain background : Epistasis



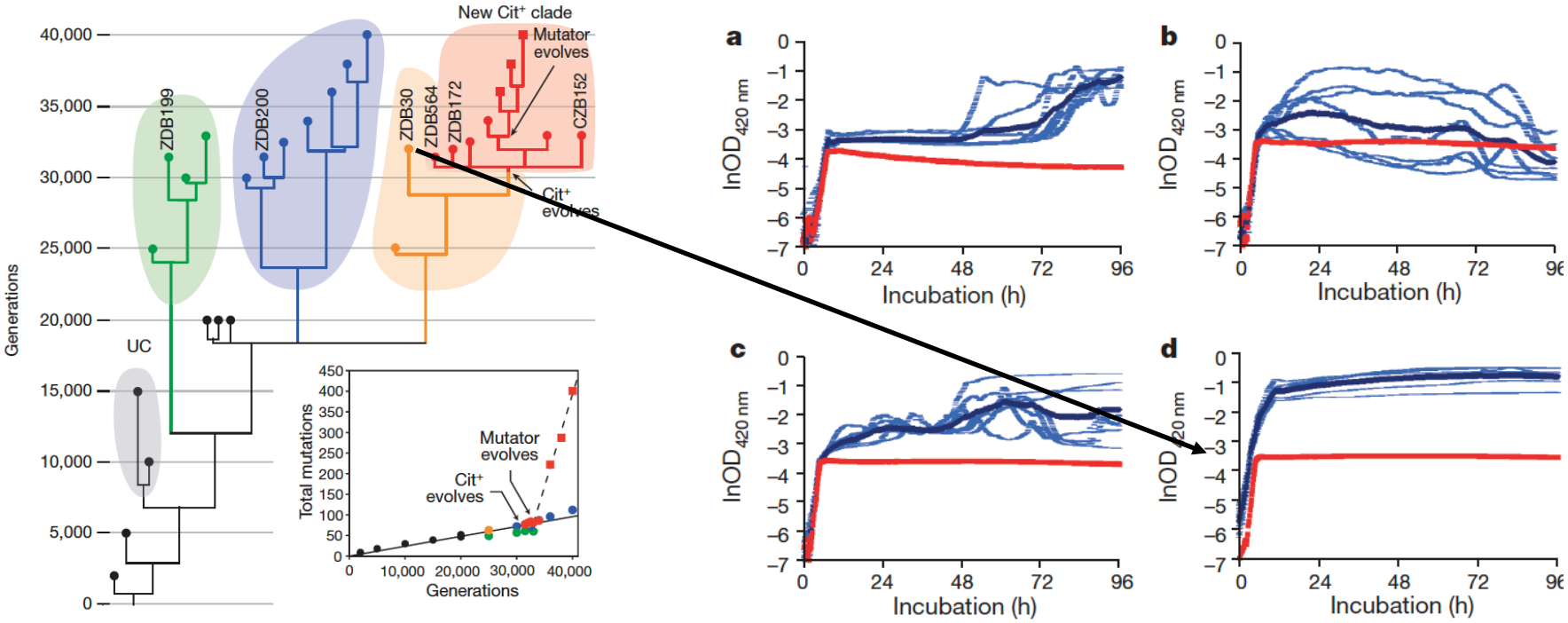
Previous mutations in the bacteria led to the module being able to convey a beneficial effect

Blount et al, Nature, 2012

Evolutionary experiments

Evolution of citrate metabolism

Benefit of module is dependent on strain background : Epistasis



Previous mutations in the bacteria led to the module being able to convey a beneficial effect

Blount et al, Nature, 2012

Evolutionary experiments

Lab evolution of Escherichia coli : One Disadvantage

Evolution happens relatively slowly (>25 years). Can it be made to happen faster?

Factors in slowing adaptation:

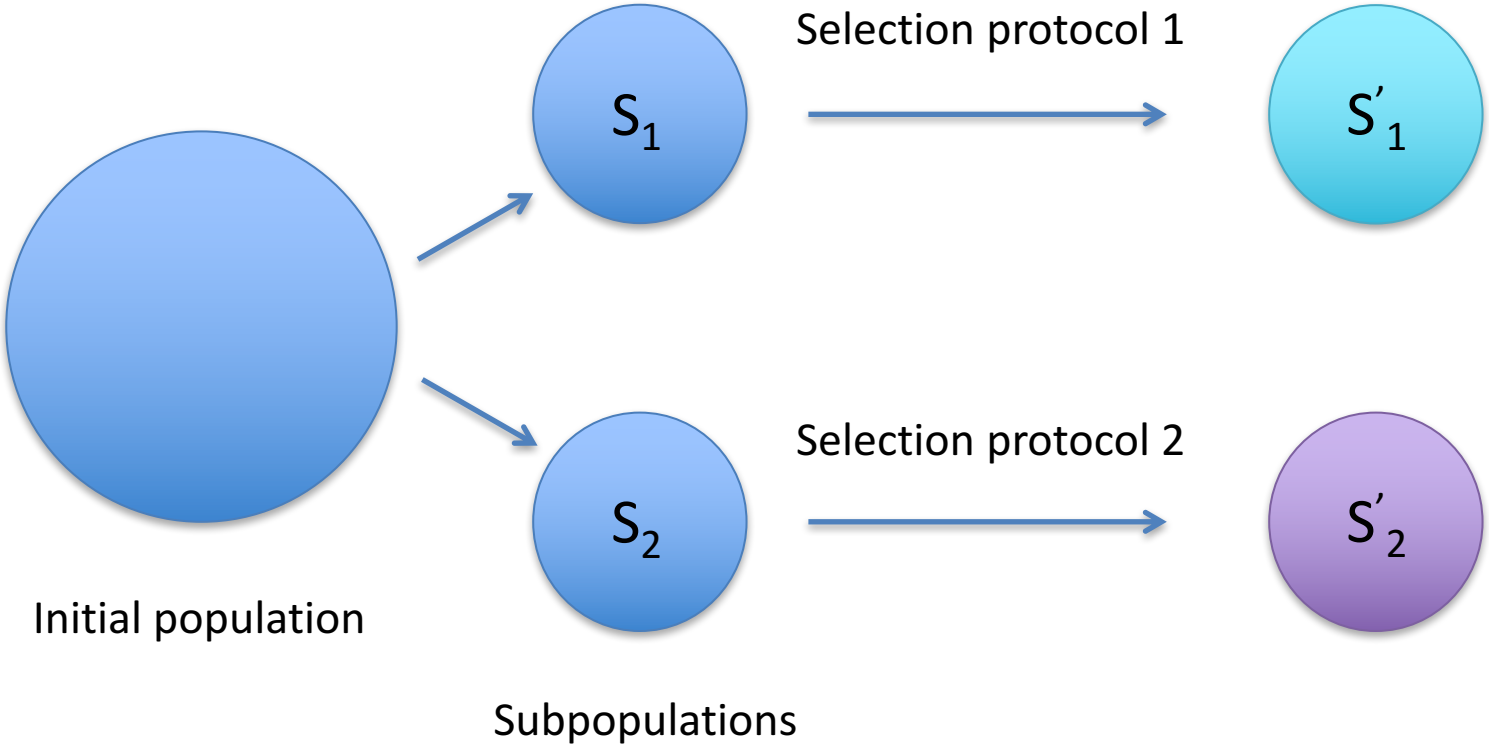
Diluting the population into new bottles: population bottleneck reduces the potential for adaptation

Constant environment: Relative lack of selection pressure

Evolutionary experiments

Speeding up evolution: Apply selection pressure

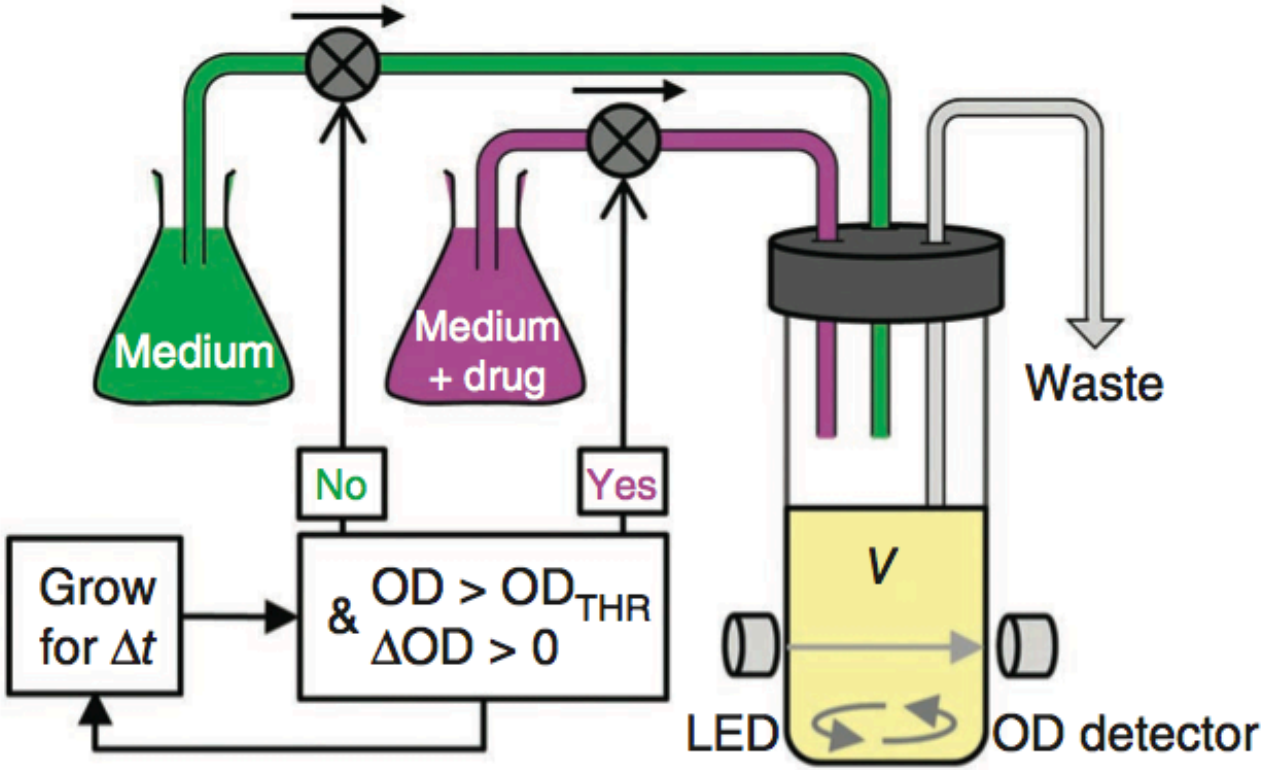
Experimental design:



Evolutionary experiments

Example: Morbidostat

Alter drug concentration with time to maintain population size



Toprak et al, Nature Genetics, 2011

Evolutionary experiments

Observed mutations

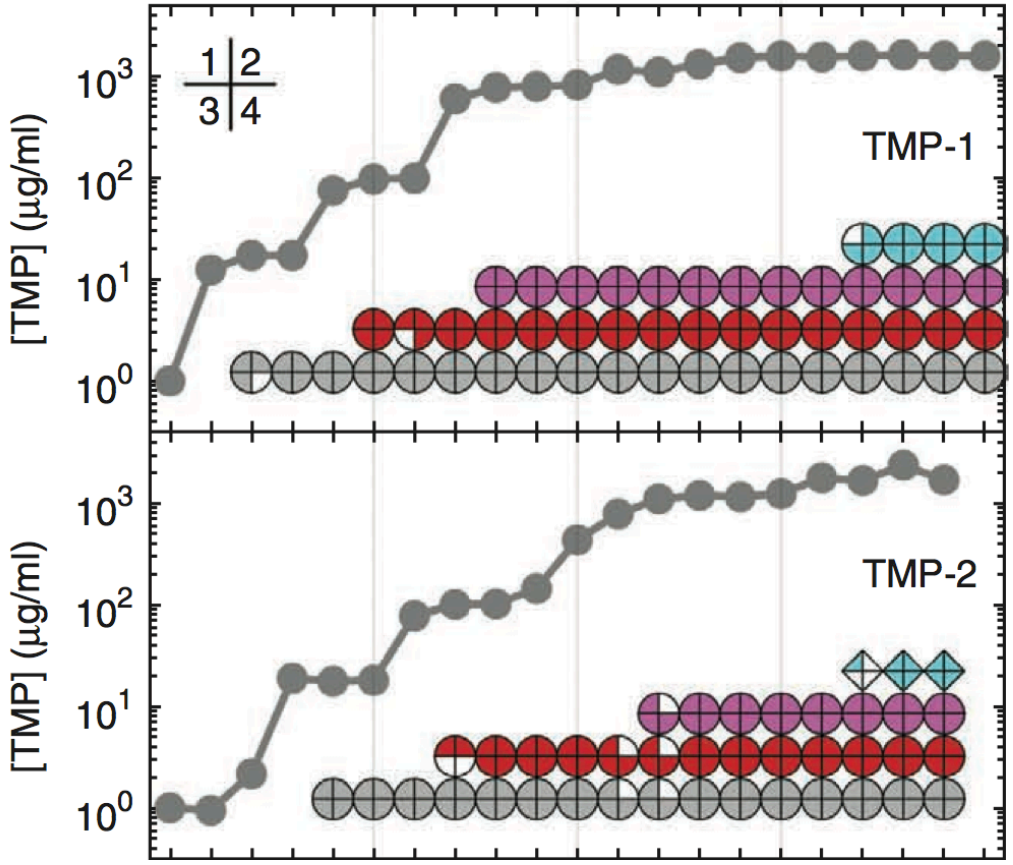
Drug binds to DHFR: observe mutations in this protein

Toprak et al, Nature Genetics, 2011

Evolutionary experiments

Observed mutations

Order of mutations preserved between experiments

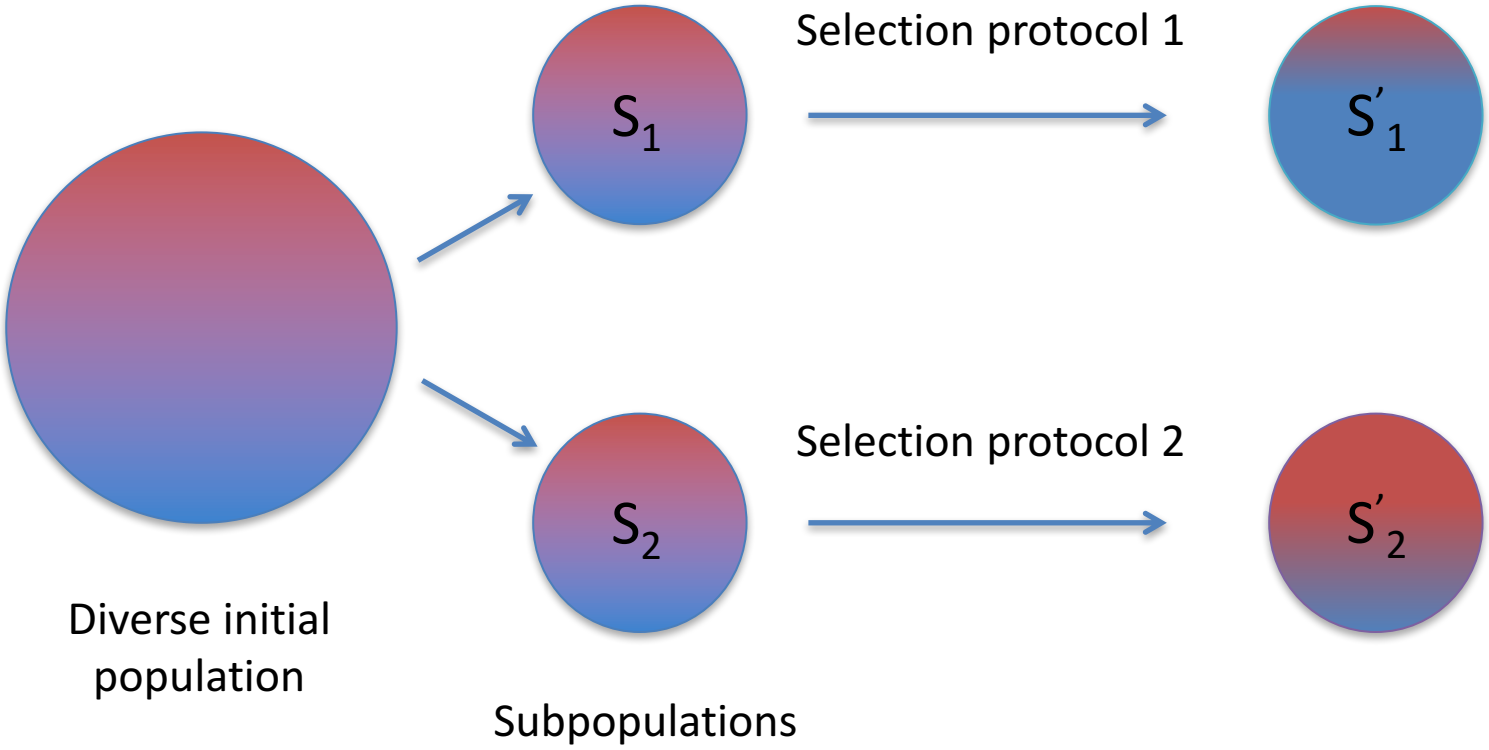


Toprak et al, Nature Genetics, 2011

Evolutionary experiments

Speeding up evolution: Variation in the initial population

Beneficial mutations exist in the initial population



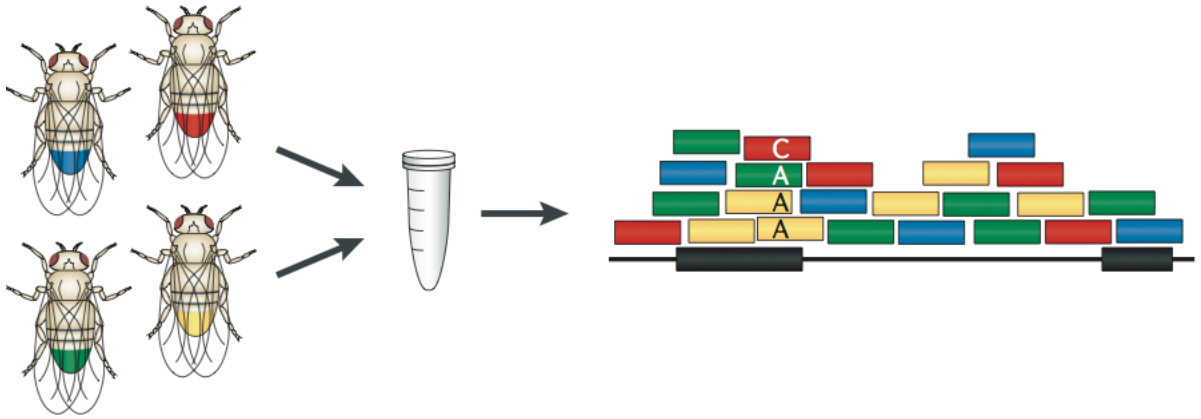
Evolutionary experiments

Example: “Evolve-and-resequence” experiment

Collect flies from a wild population

Grow under selection for 56 days

Sequence pooled population

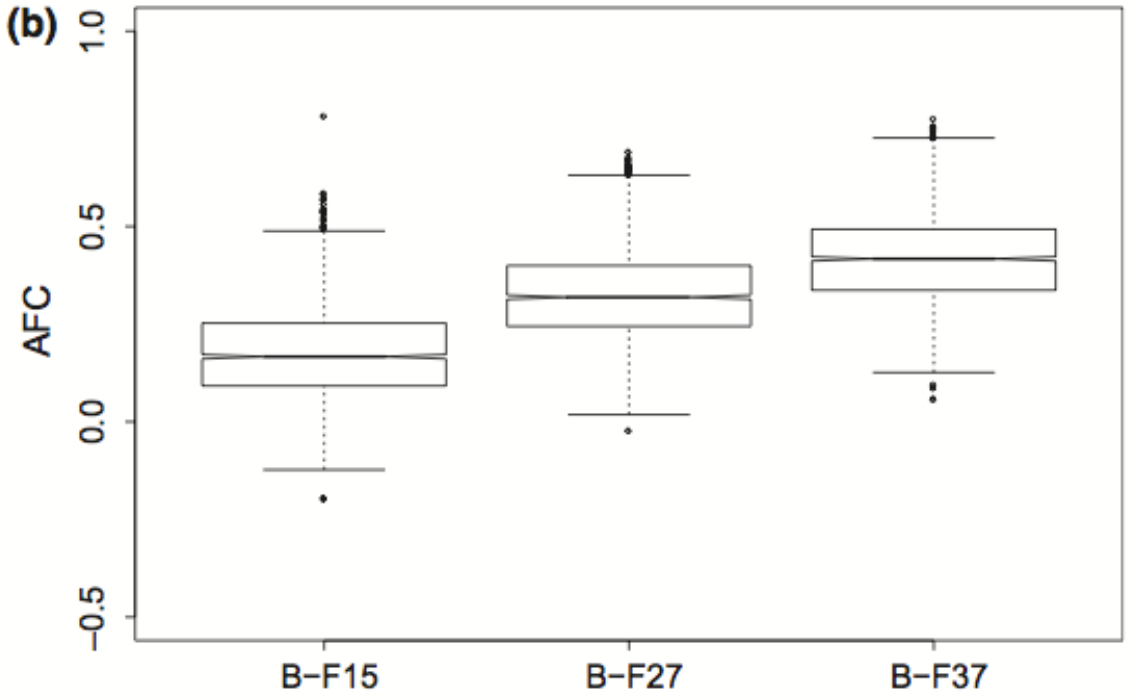


Schlotterer et al, Nature Reviews Genetics, 2014

Evolutionary experiments

Example: "Evolve-and-resequence" experiment

Changes seen in allele frequencies (top 2000 loci):

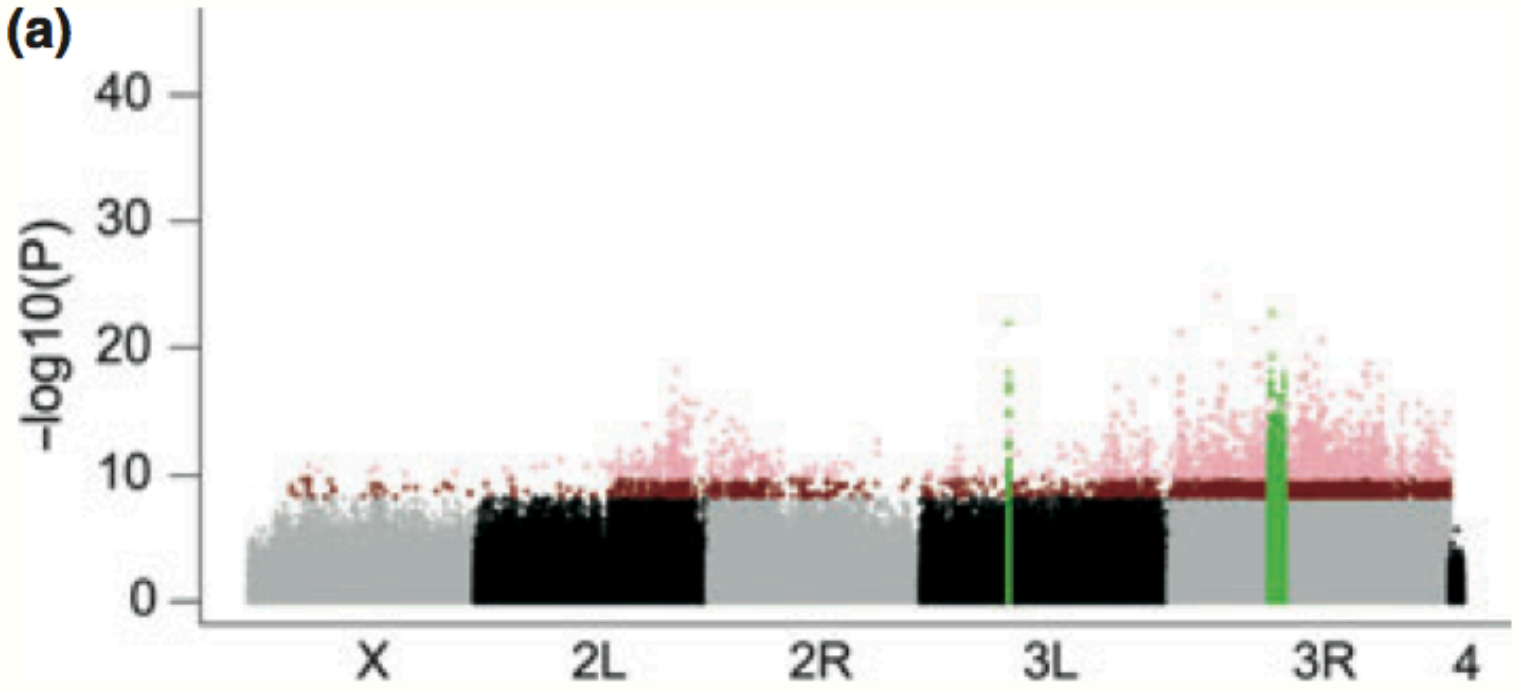


Orozco-terWengel et al, Mol. Ecol., 2012

Evolutionary experiments

Example: "Evolve-and-resequence" experiment

Test for selection: Alleles that move more than expected given the effects of genetic drift

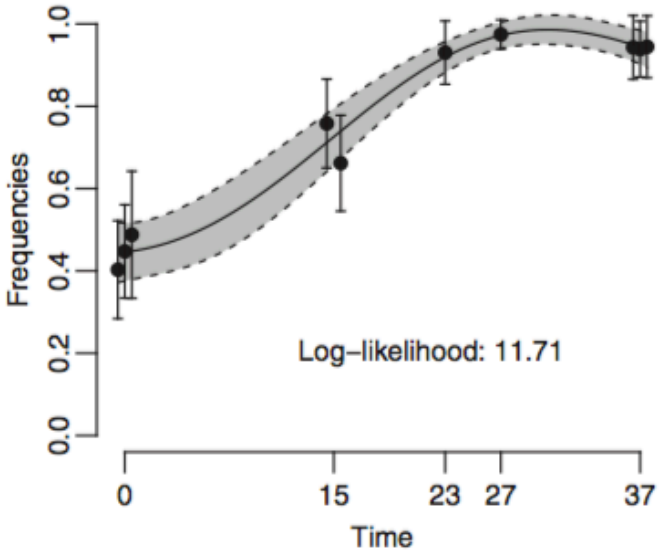


Orozco-terWengel et al, Mol. Ecol., 2012

Improved analysis

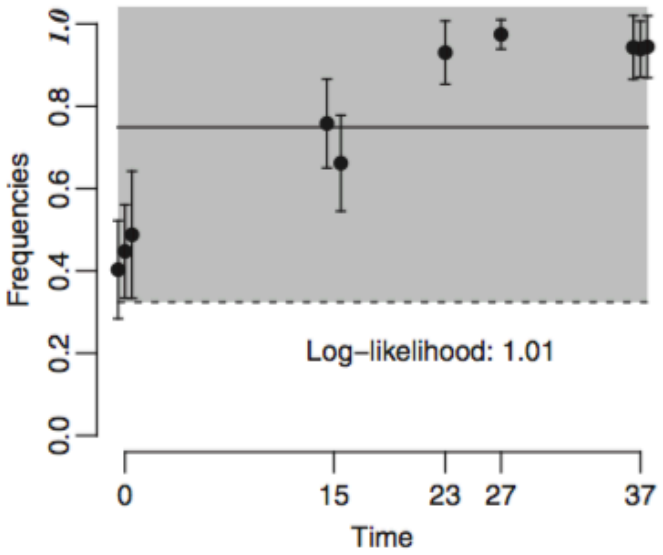
Use of Gaussian processes

Test for selection: Alleles that move more than expected given the effects of genetic drift



Time-dependent model:

$$m_{ij} = f_i\,(t_j) + \mu_{m_i} + \epsilon$$



Time-independent model:

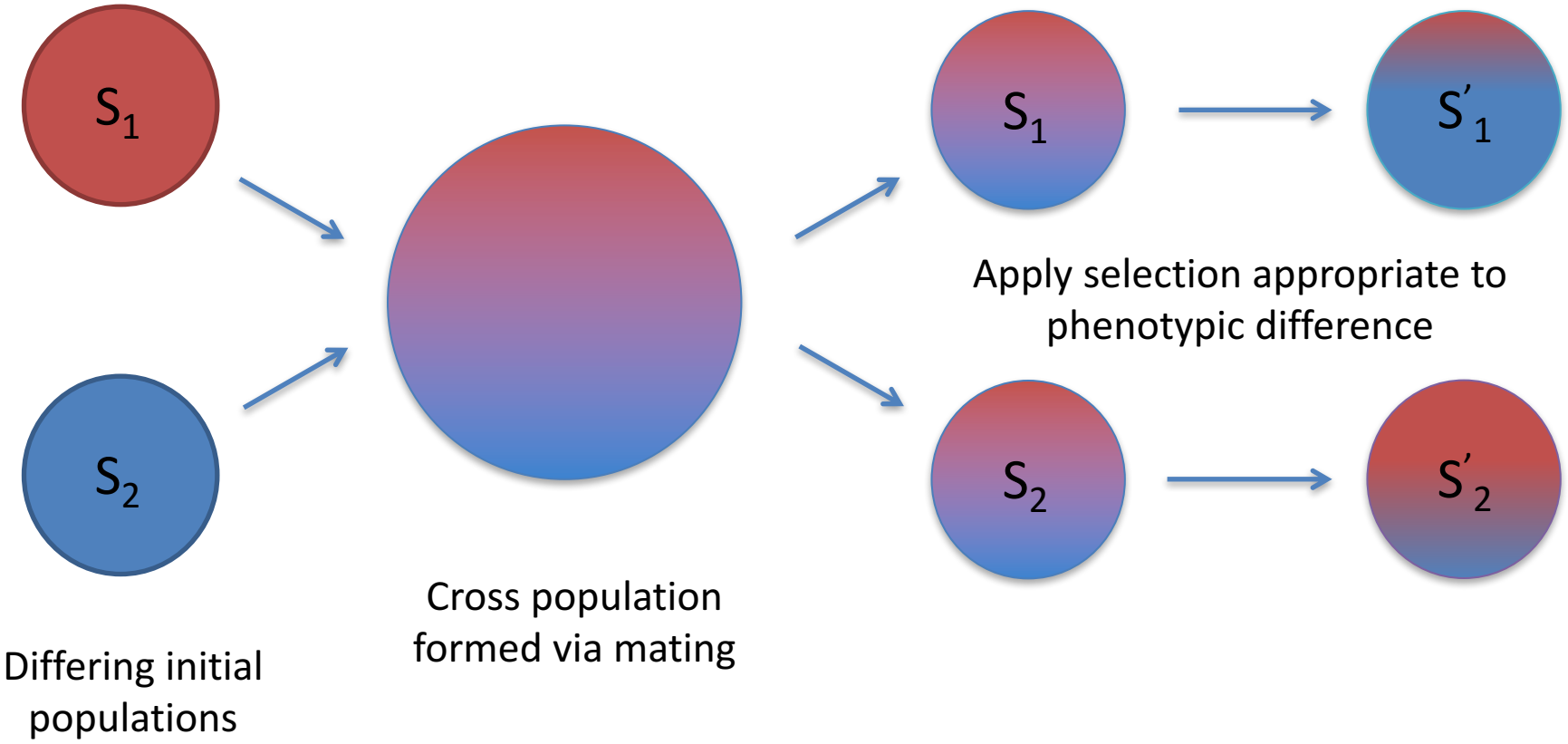
$$m_{ij} = \mu_{m_i} + \epsilon$$

Topa et al, Bioinformatics, 2015

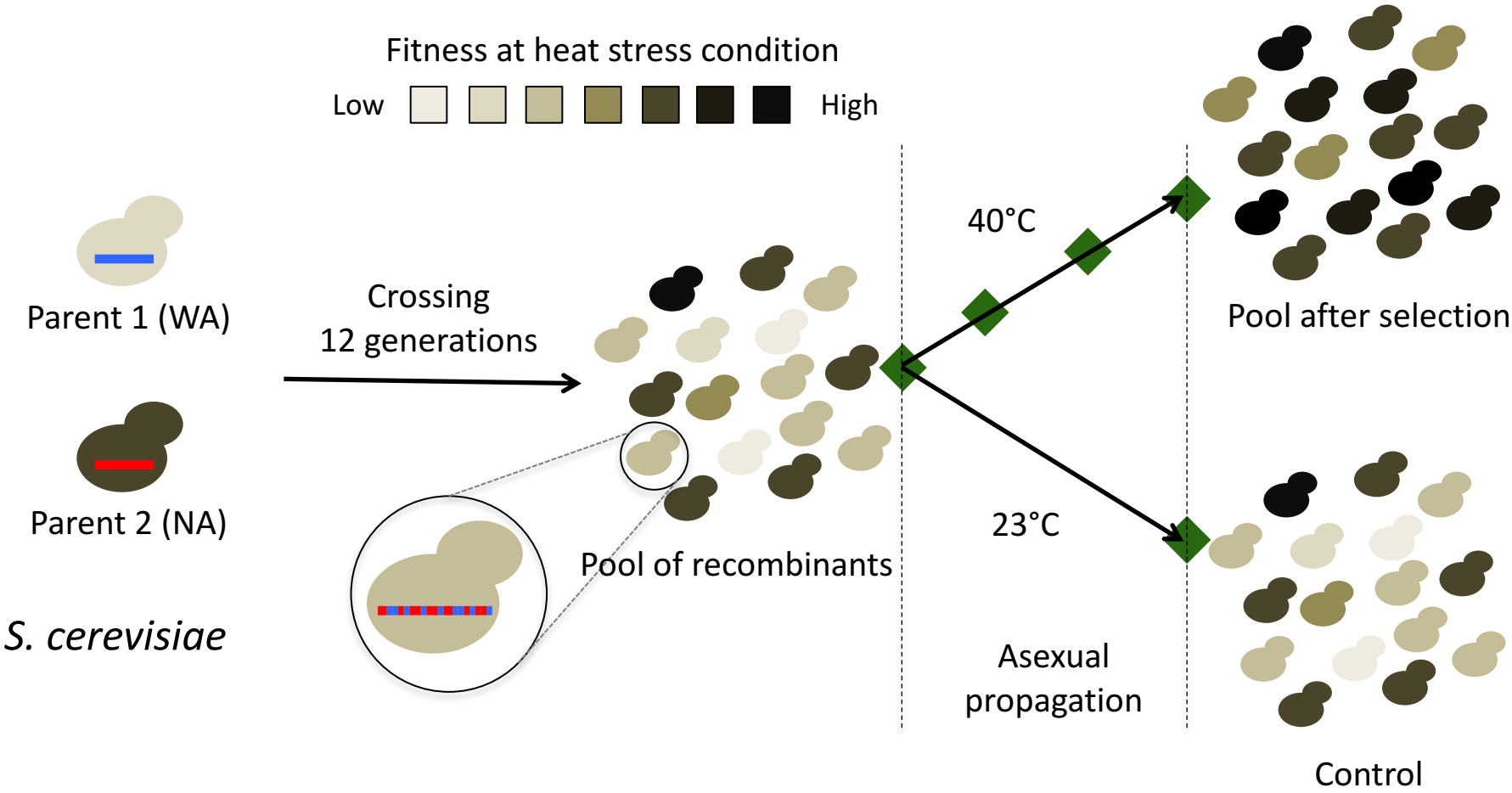
Evolutionary experiments

Easier case: Experimental cross

Can cross individuals with known phenotypes



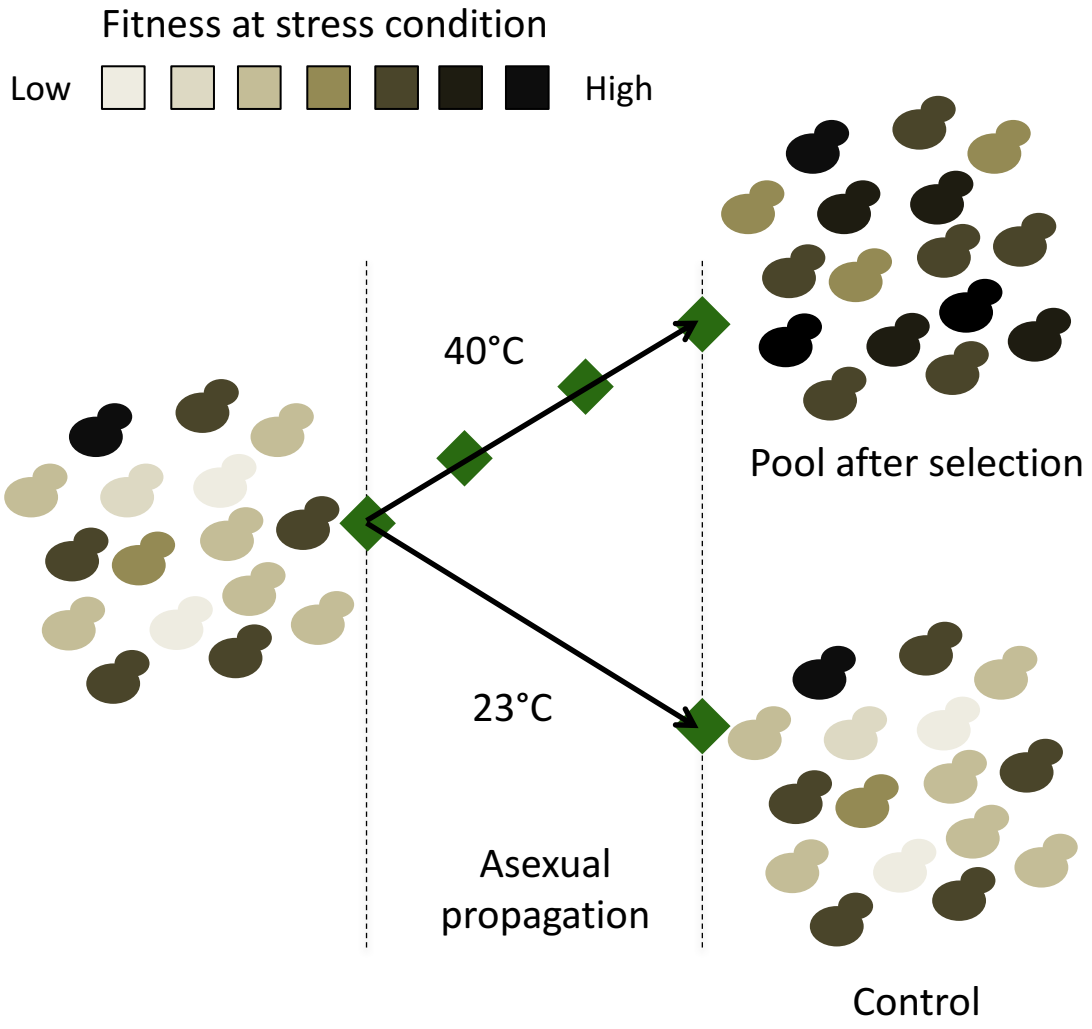
Yeast Experiment



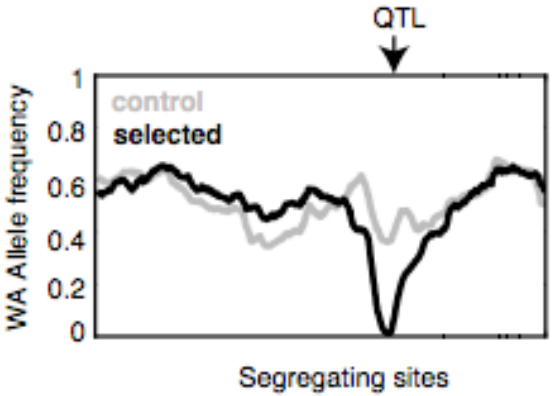
◆ Whole-genome sequencing of the pool

Figure adapted from Parts et al, Genome Research 2011

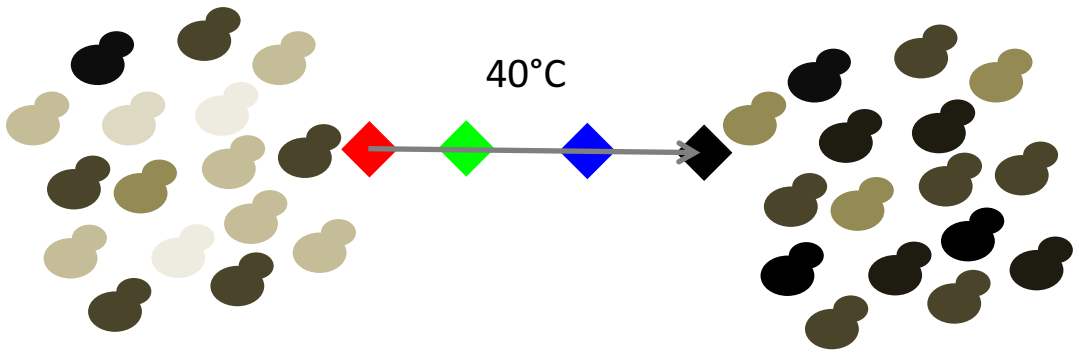
Yeast Experiment



Obtain allele frequency data at around 30 000 segregating sites



Time-resolved data

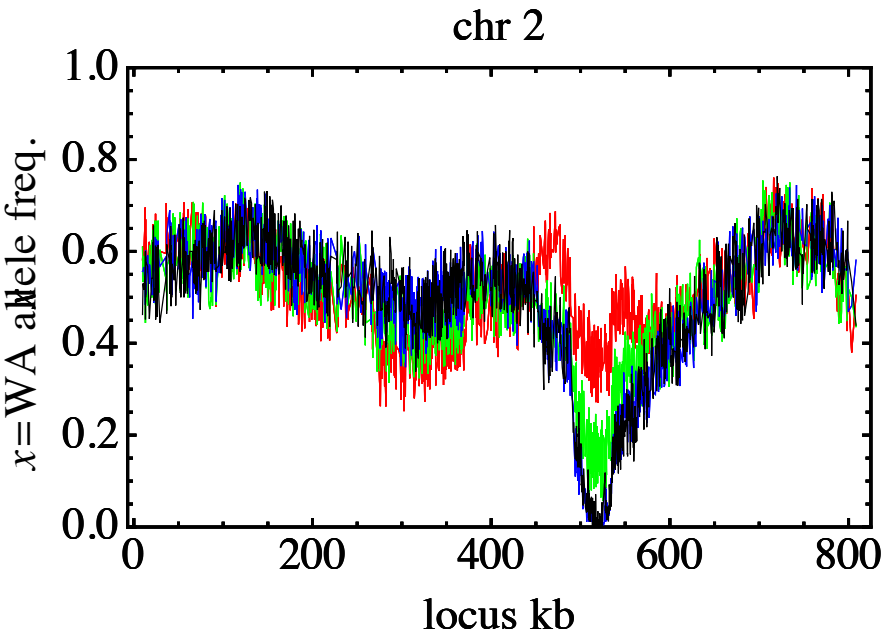


Pool of recombinants

Pool after selection

96h intervals

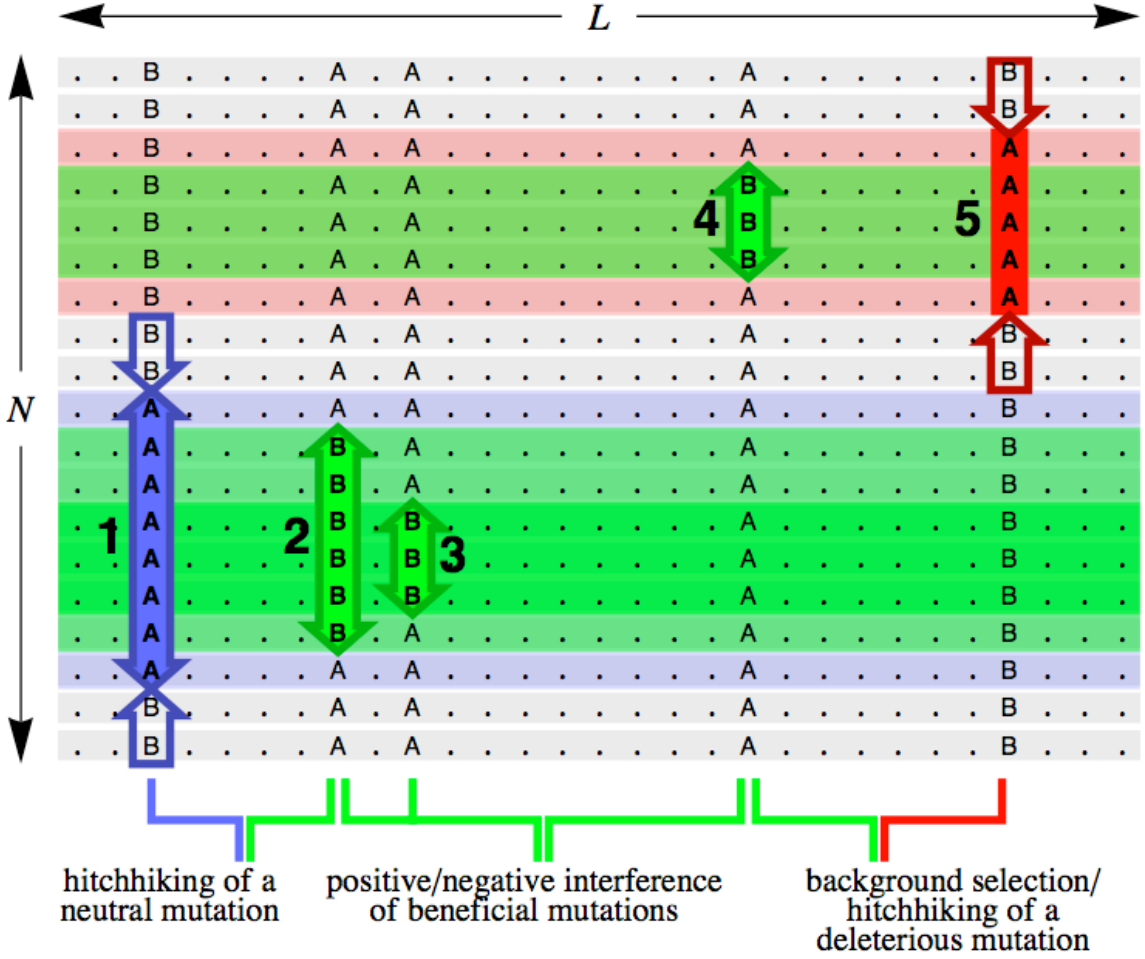
Data from chromosome 2



Inferring selection across multiple loci

Inferring selection

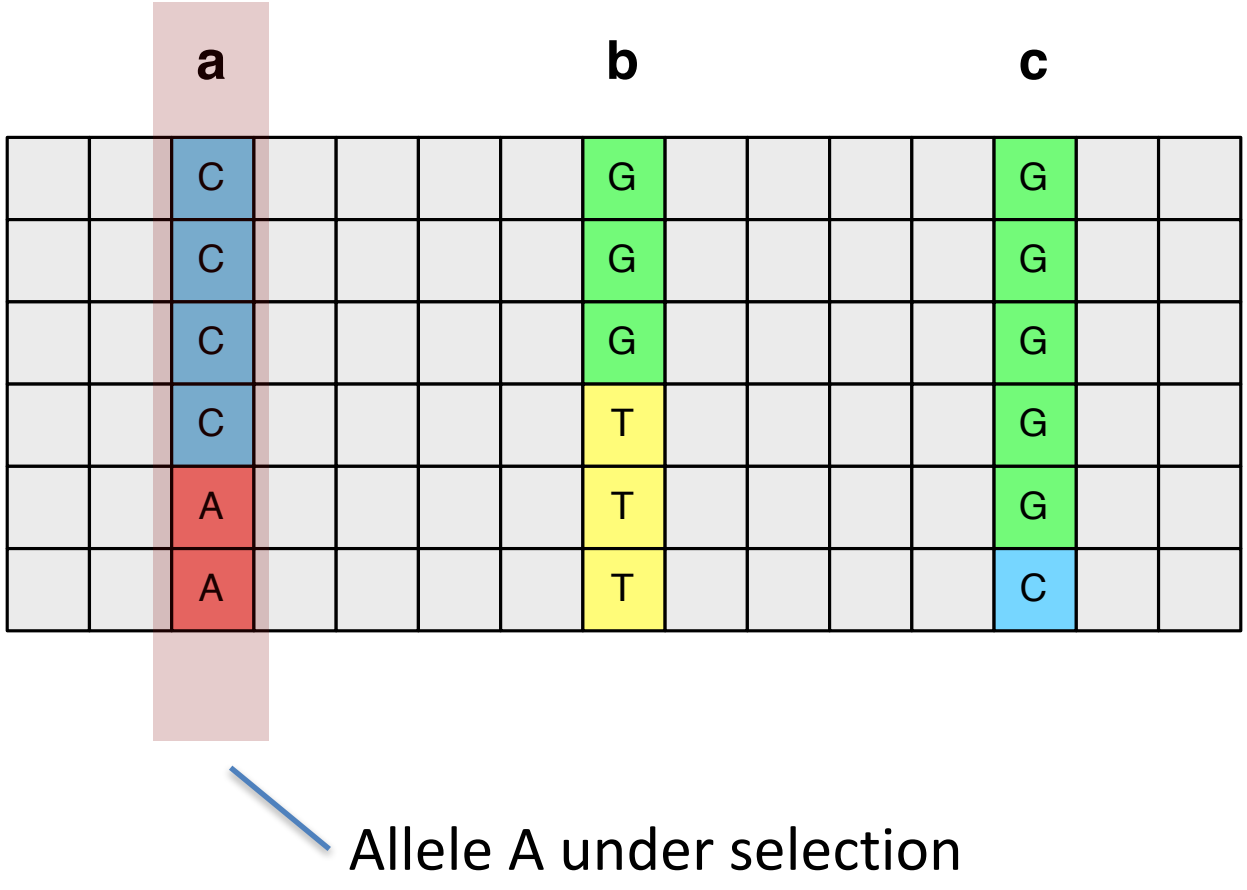
Problem of inferring selection at multiple loci



Schiffels et al., Genetics 2011

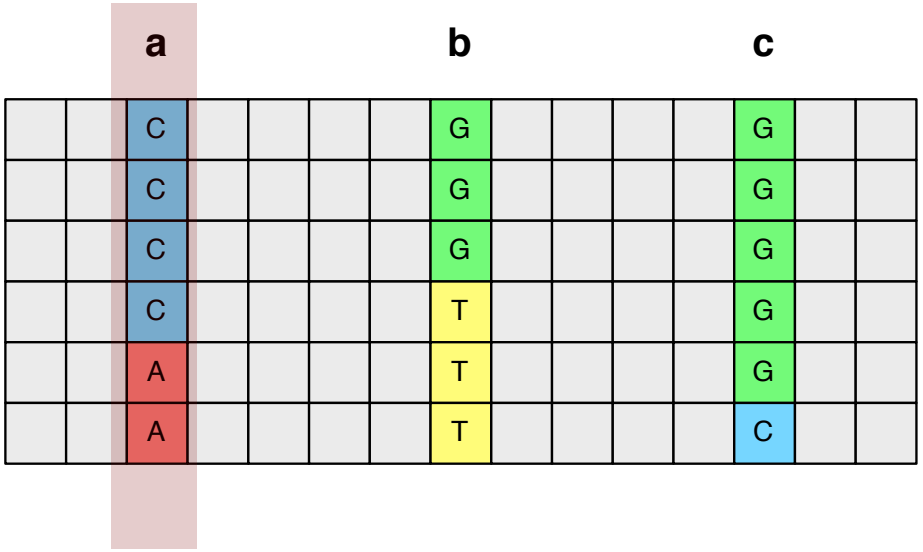
Inferring selection

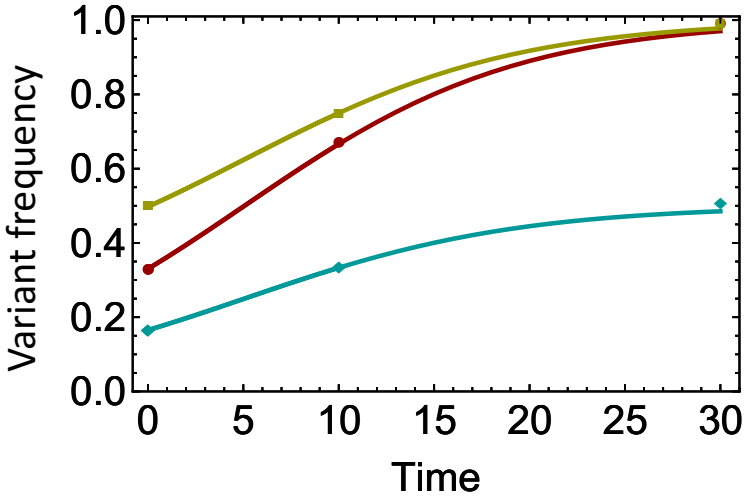
Challenge: Linked, simultaneous polymorphisms



Inferring selection

Challenge: Linked, simultaneous polymorphisms

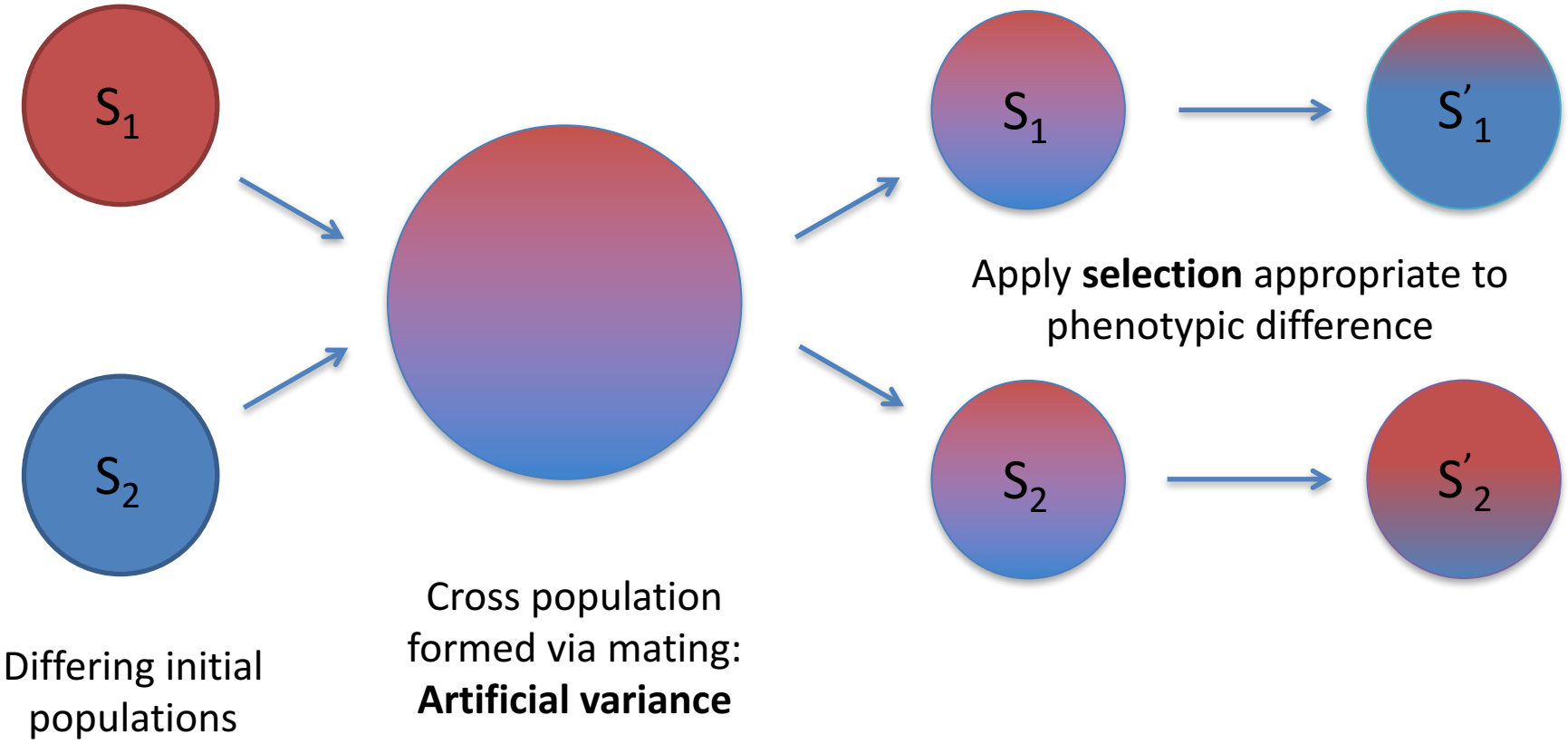




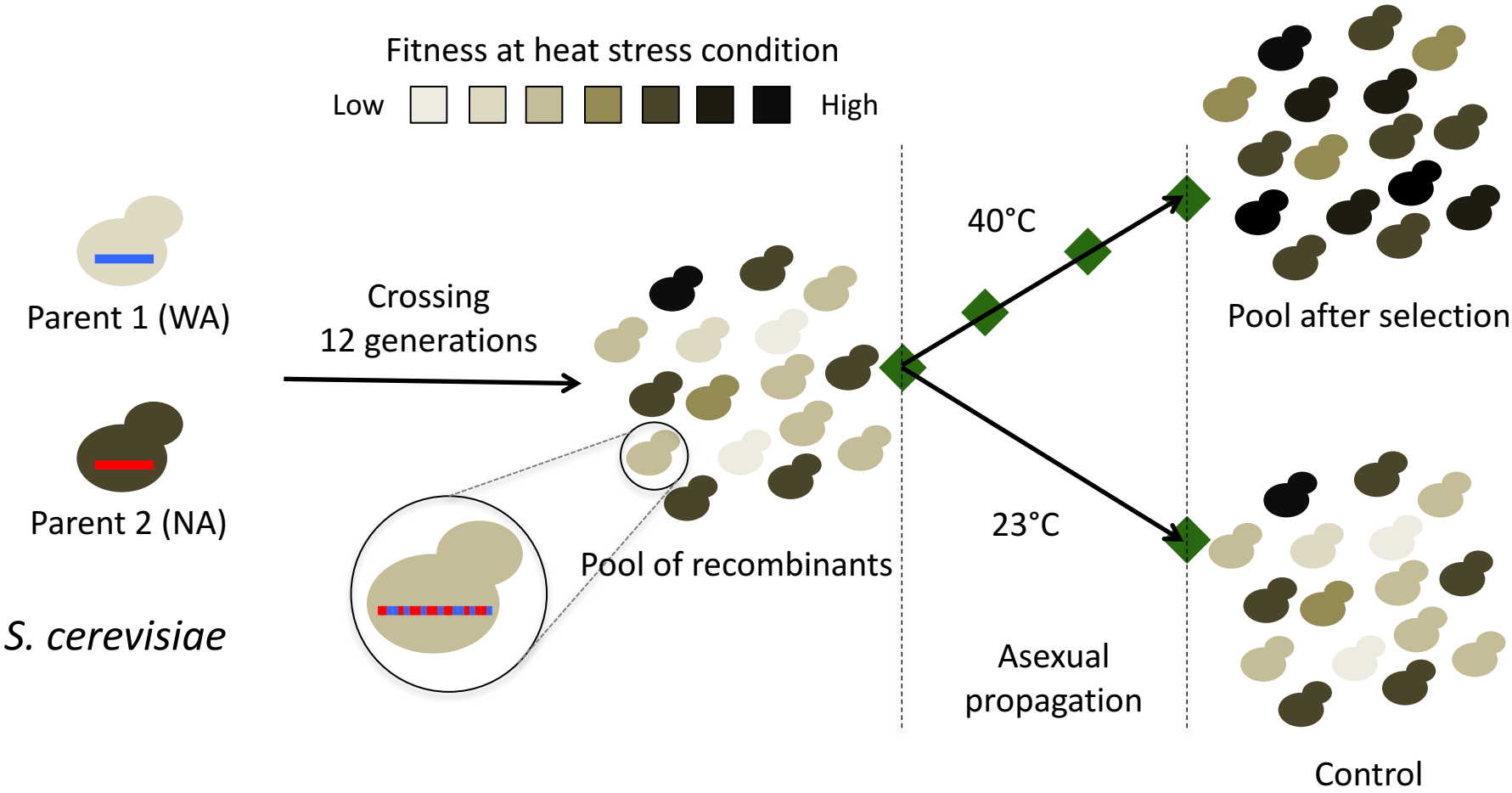
Evolutionary experiments

Speeding up evolution: Experimental cross

Can cross individuals with known phenotypes



Example III: Selection across multiple loci



◆ Whole-genome sequencing of the pool

Figure adapted from Parts et al, Genome Research 2011

Example III: Selection across multiple loci

Liquid medium: well-mixed population

Population size $N \approx 10^7$-10^8

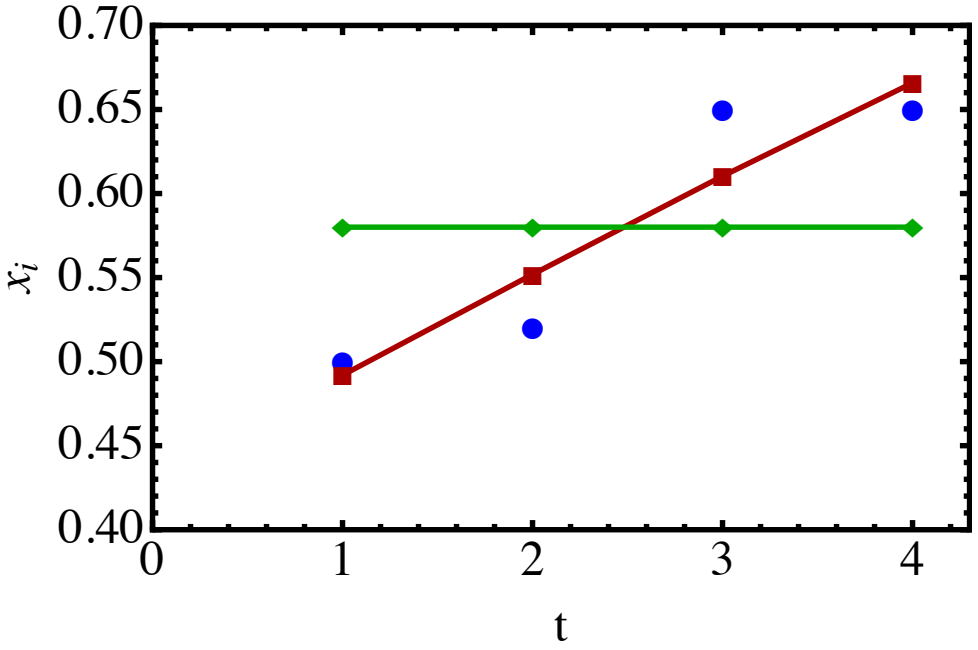
Time under selection = 288 hours (c. 25-100 generations)

Pool sequencing depth c. 100x

Example III: Selection across multiple loci

Single-locus analysis

Look for deviation from neutrality



Observed frequencies

$$x_i(t_k) = n_i(t_k)/N_i(t_k)$$

Neutral model

$$dx_i/dt = 0$$

Constant selection model

$$dx_i/dt = \sigma x_i(1 - x_i)$$

Obtain log likelihood under each model

Likelihood difference between models ΔL

Example III: Selection across multiple loci

More complicated models fit data more closely

$$\mathcal{M}_1 : x = a_0$$

$$\mathcal{M}_2 : x = a_0 + a_1 x$$

$$\mathcal{M}_3 : x = a_0 + a_1 x + a_2 x^2$$

$$\mathcal{M}_\infty : x = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots$$

Penalise for model complexity: New parameters must improve the model by more than the penalty to be accepted

Example III: Selection across multiple loci

AIC: Akaike Information Criterion

$$AIC = 2k - 2\log(L)$$

L: Likelihood
k: # parameters

Minimum AIC gives best model

BIC: Bayesian Information Criterion

$$BIC = k\log(n) - 2\log(L)$$

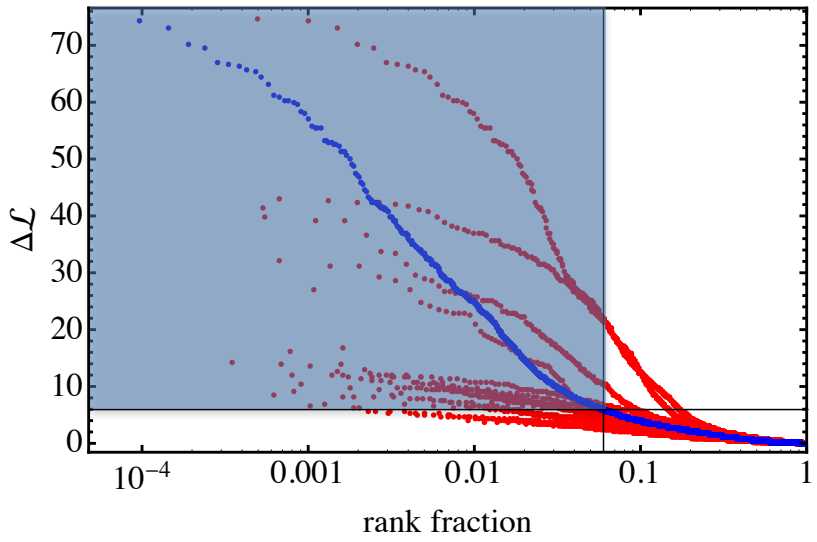
L: Likelihood
k: # parameters
n: # data points

Minimum BIC gives best model

Usually more conservative…

Example III: Selection across multiple loci

How many allele frequencies move in a significantly non-neutral manner as a result of heat selection?



Individual chromosomes

Whole genome

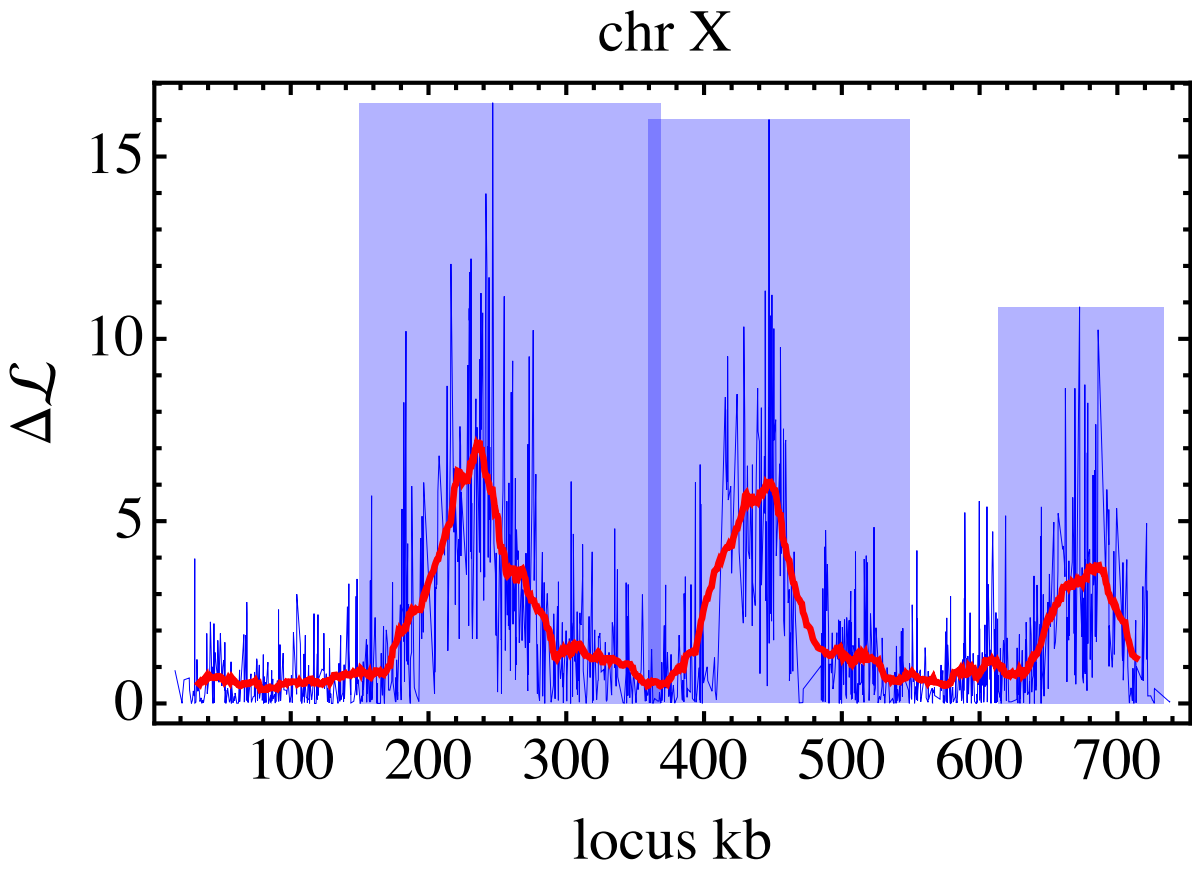
Cutoff ΔL>6 (AIC score difference >10)

Around 6% of alleles move non-neutrally

Compares to 0.02% for control experiment

Example III: Selection across multiple loci

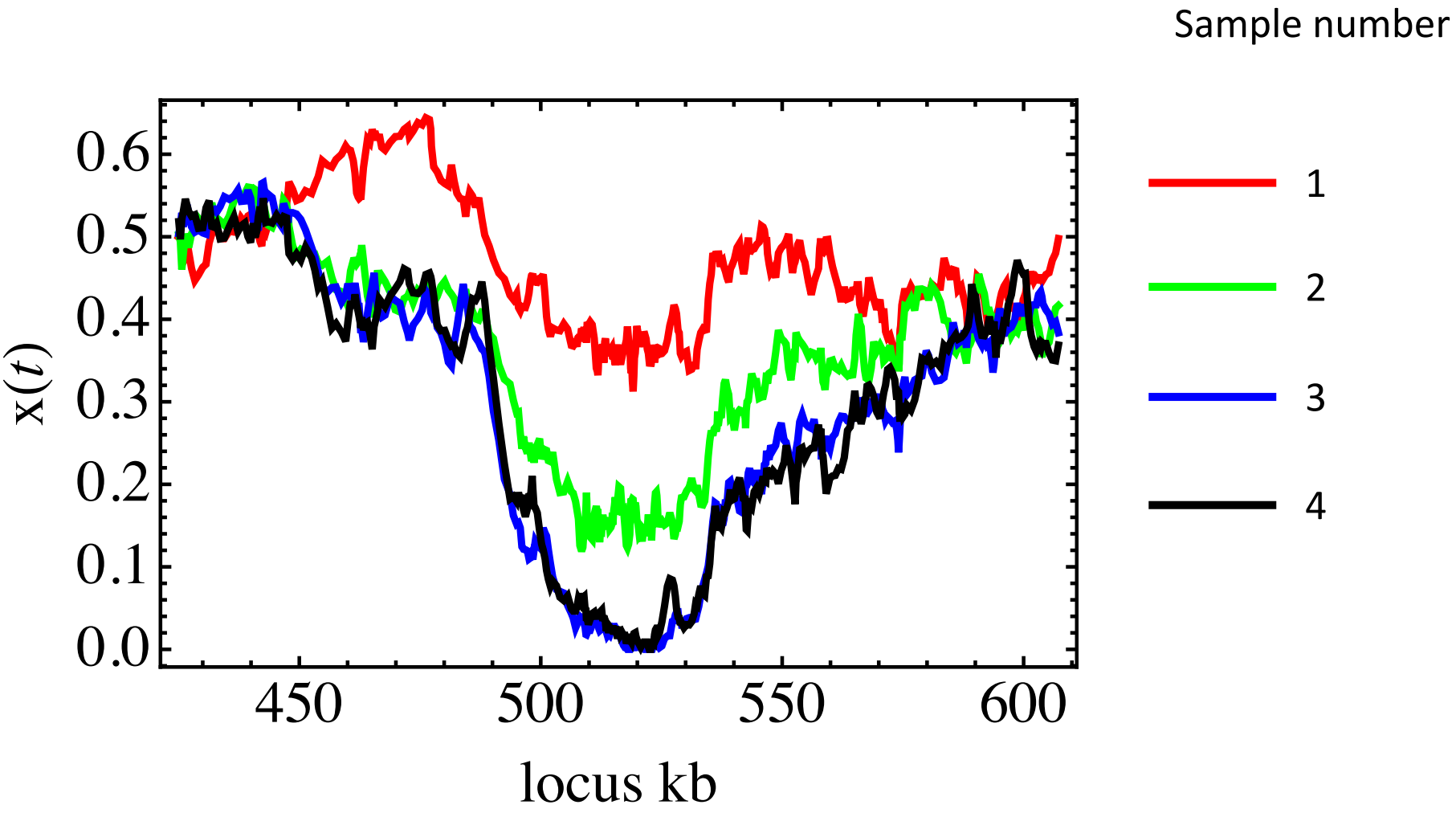
Regions of “non-neutral” behaviour



Consider each region in turn

Model comparison

Region in chromosome II

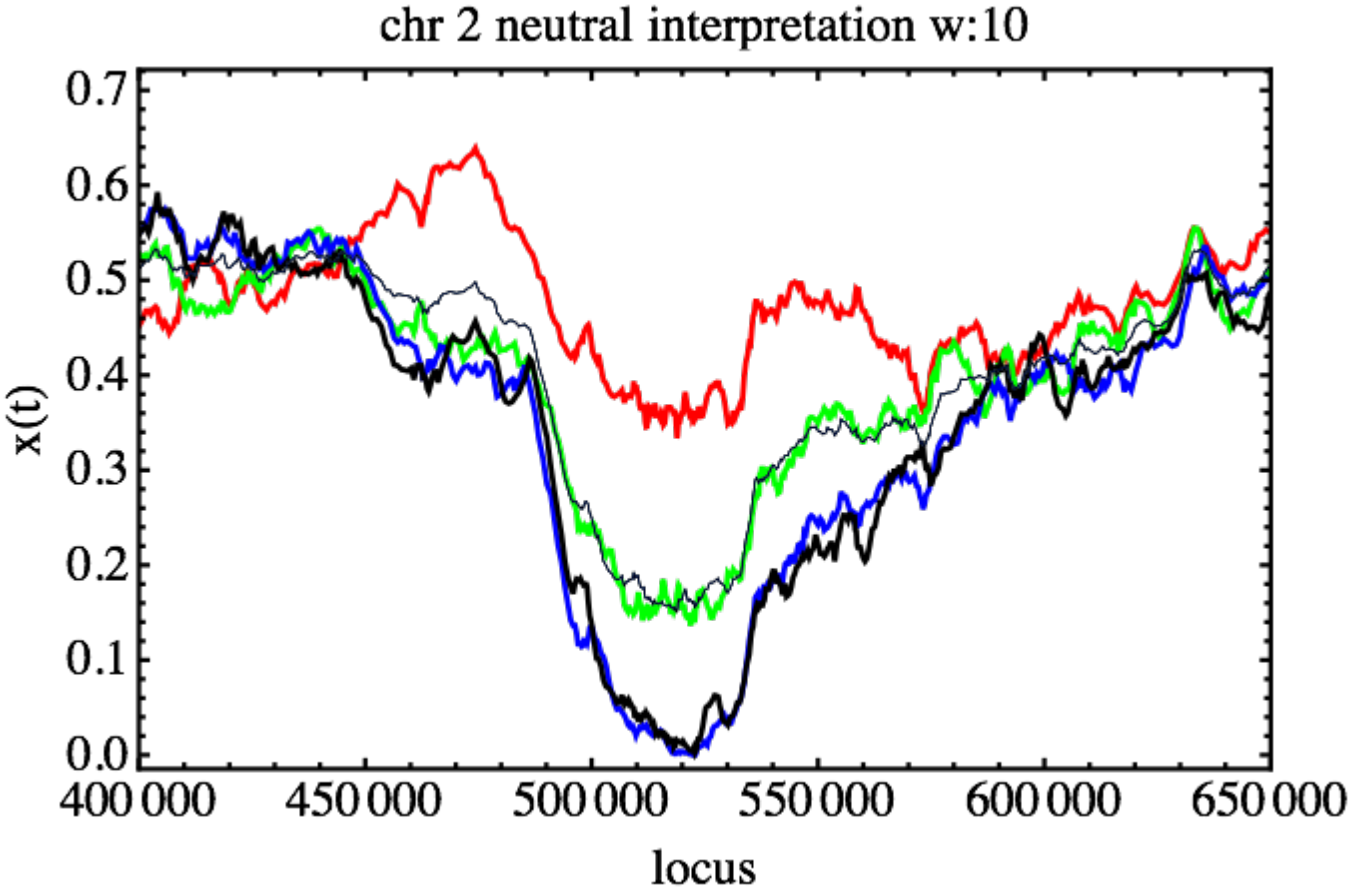


Model comparison

Region in chromosome II

Neutral model: No change in allele frequency

Solid black line: model fit

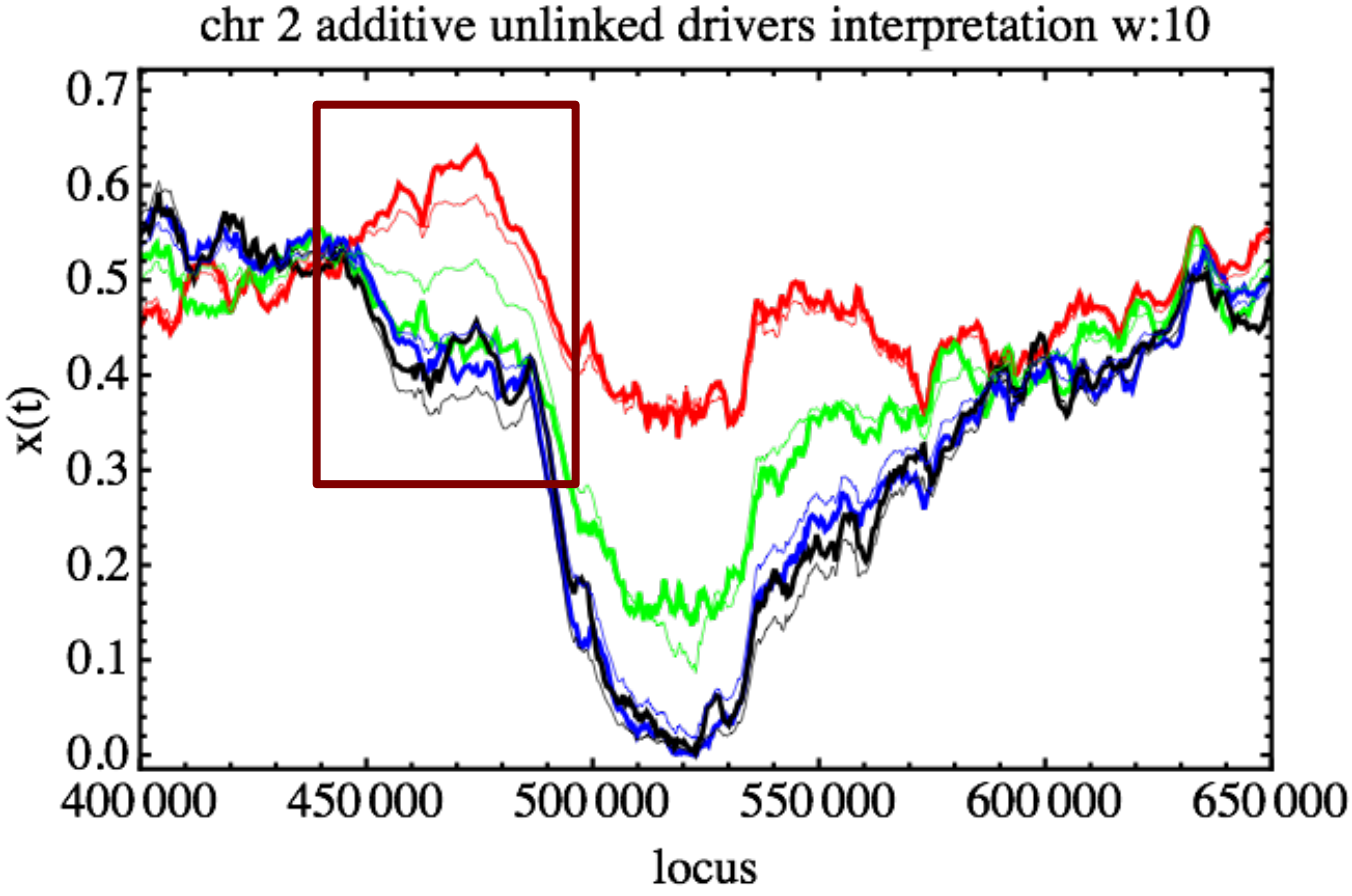


Model comparison

Region in chromosome II

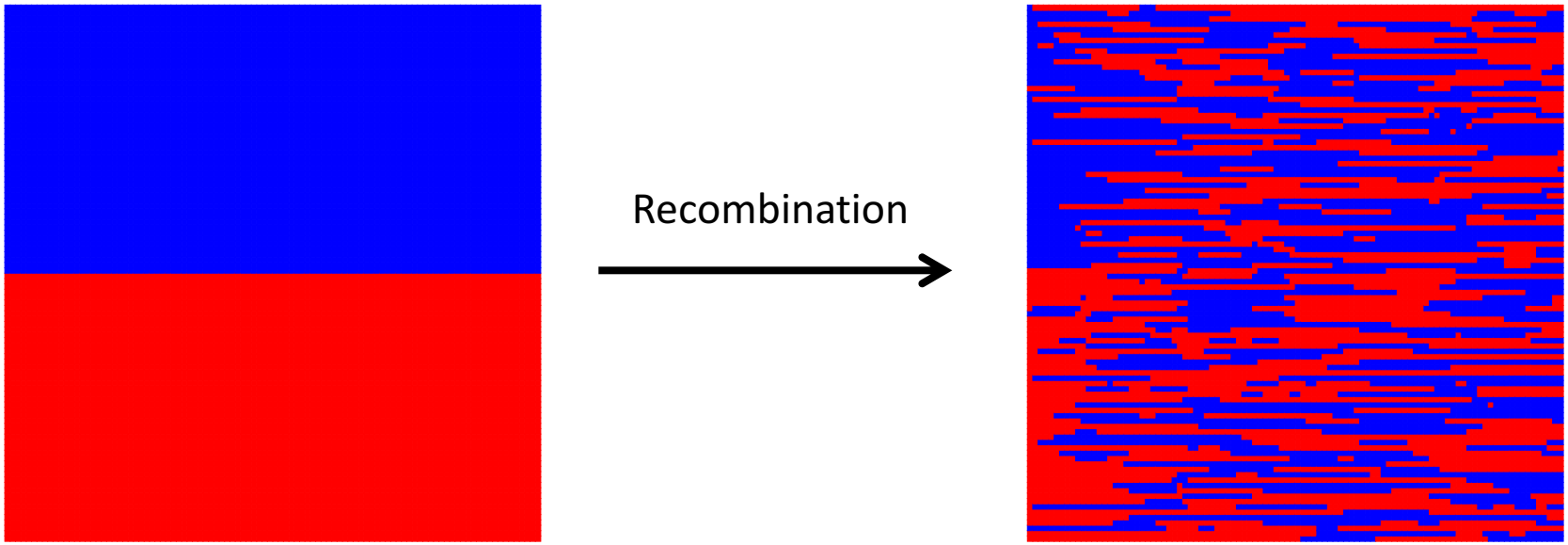
Thin coloured lines: model fit

Single-locus model: Each allele moves under its own selection



Genetic structure following recombination

Links between alleles are statistically comprehensible



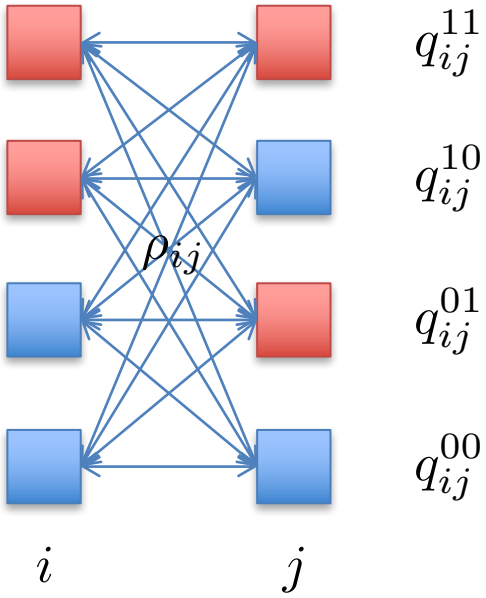
Pure genotypes

Crossed population

Linkage disequilibrium

LD and Recombination

Recombination breaks linkage disequilibrium



$$q_{ij}^{11}(t+1) = (1-\rho_{ij})q_{ij}^{11}(t) + \rho_{ij}q_i^1(t)q_j^1(t)$$

$$D_{ij}(t+1) = (1-\rho_{ij})D_{ij}(t)$$

Linkage disequilibrium decays to zero exponentially over time **+ distance**

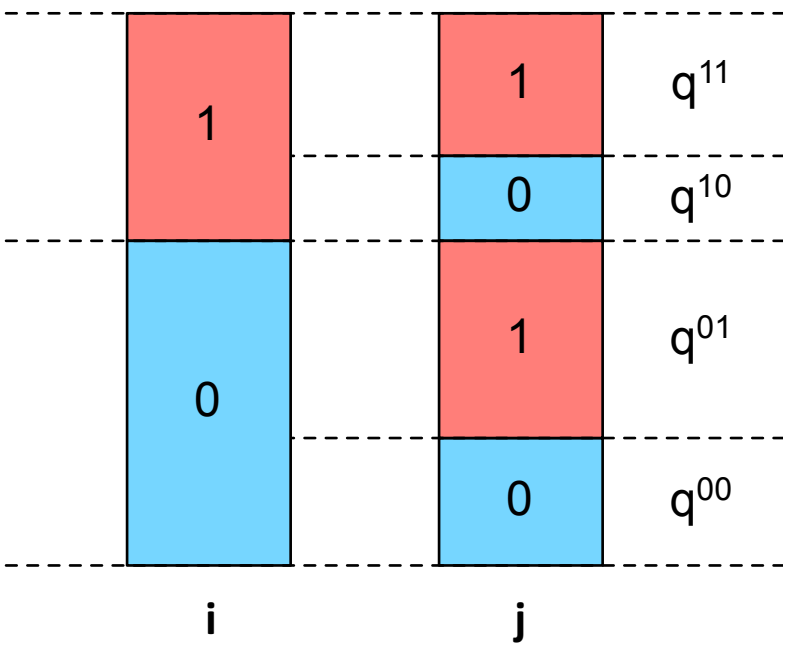
$$D_{ij} = q_{ij}^{11} - q_i^1 q_j^1$$

Example III: Selection across multiple loci

Driver and passengers model

Driver: $$\frac{dq_i^1}{dt} = \sigma q_i^1 (1 - q_i^1)$$

Passengers:



$$q_j^1(t) = q_i^1(t)\frac{q_{ij}^{11}(t_0)}{q_i^1(t_0)} + q_i^0(t)\frac{q_{ij}^{01}(t_0)}{q_i^0(t_0)}$$

$$q_{ij}^{11}(t_0) = q_i^1(t_0)q_j^1(t_0) + D_{ij}$$

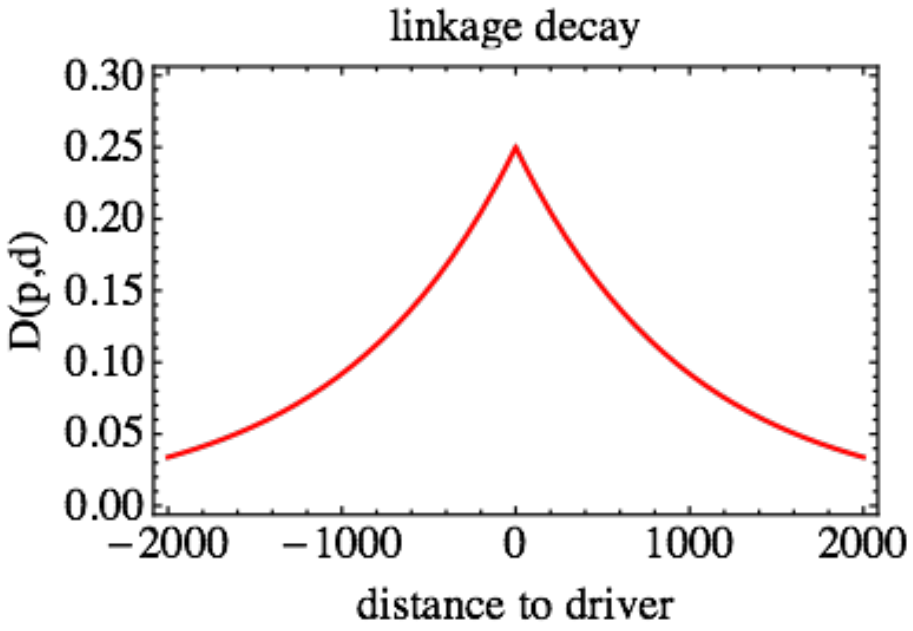
$$D_{ij}(\rho, \Delta_{ij}) = D'_{ij}(1 - \rho\Delta_{ij})^{N_c}$$

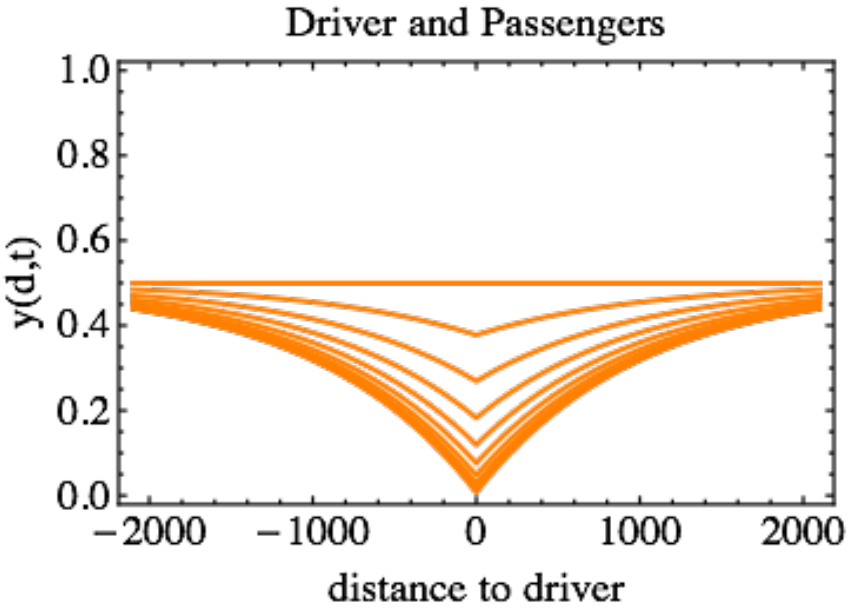
N_c: Generations of crossing
D_{ij}': Initial LD between i and j
ρ: Recombination rate
Δ_{ij}: Distance in genome between i and j

Model comparison

Decay in D_{ij} with distance and time

$$D_{ij}(\rho, \Delta_{ij}) = D'_{ij}(1 - \rho\Delta_{ij})^{N_c}$$



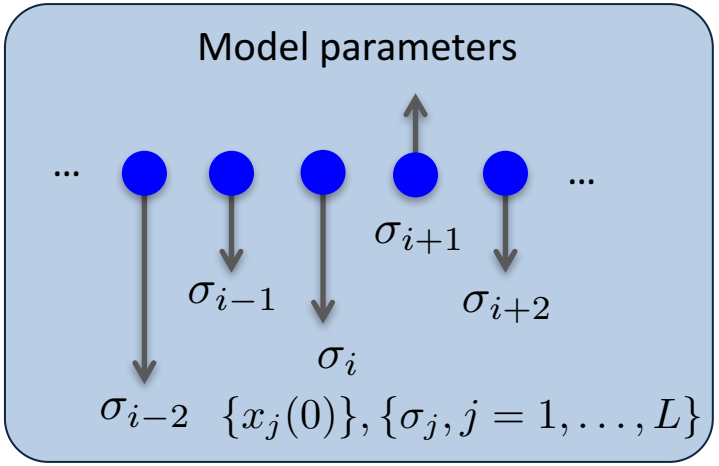


Learn position of driver, magnitude of selection, rate of recombination

Model comparison

Region in chromosome II

Single-locus model: Each allele moves under its own selection

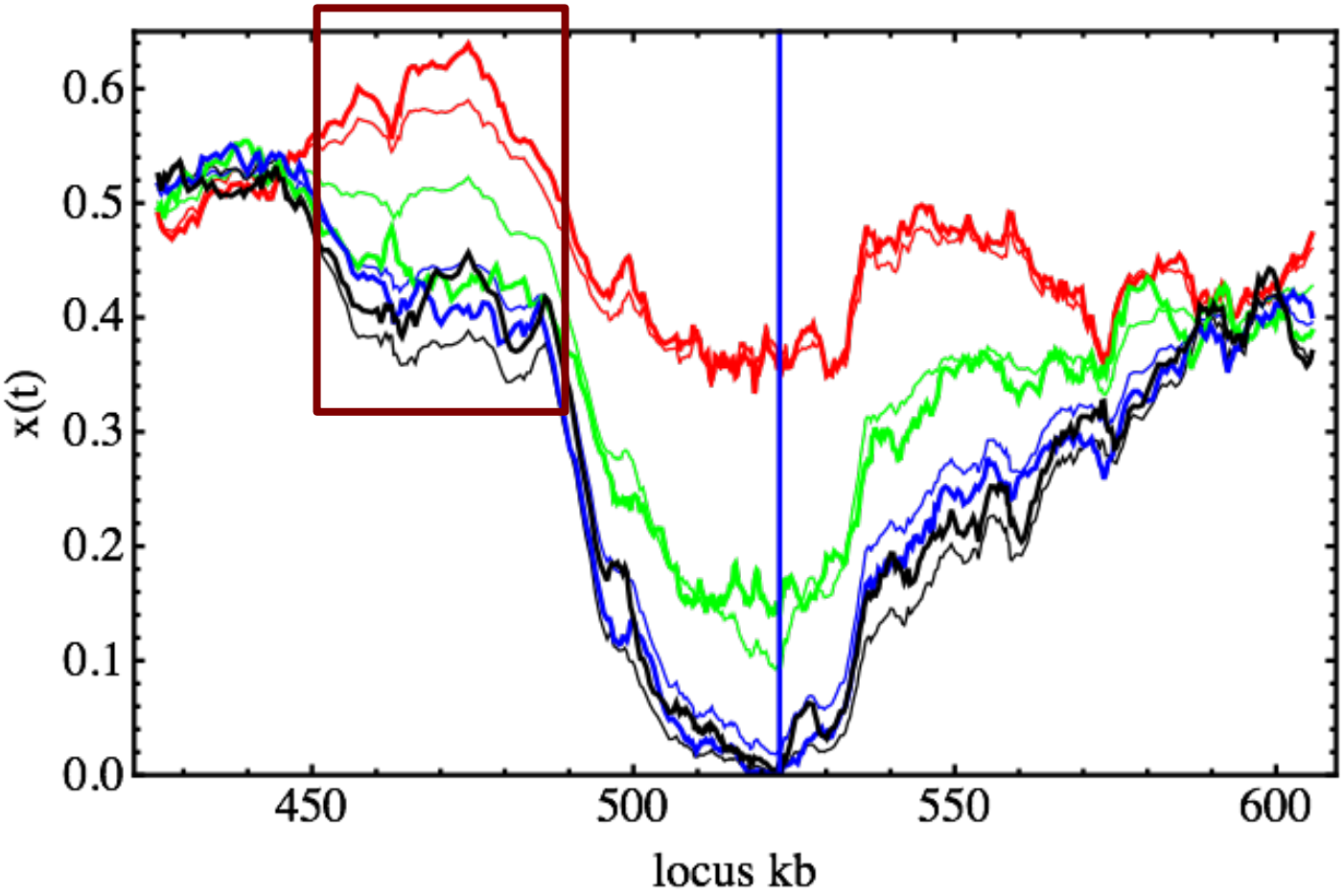


$$\Delta x_j = \sigma_j x_j (1 - x_j)$$

Model comparison

Region in chromosome II

Single-locus model: Each allele moves under its own selection

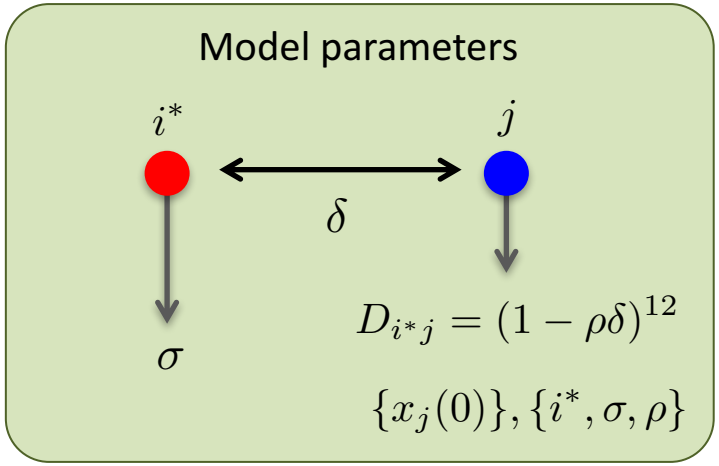


Least “neutral” allele

Model comparison

Region in chromosome II

Driver-passenger model: All linked to driver



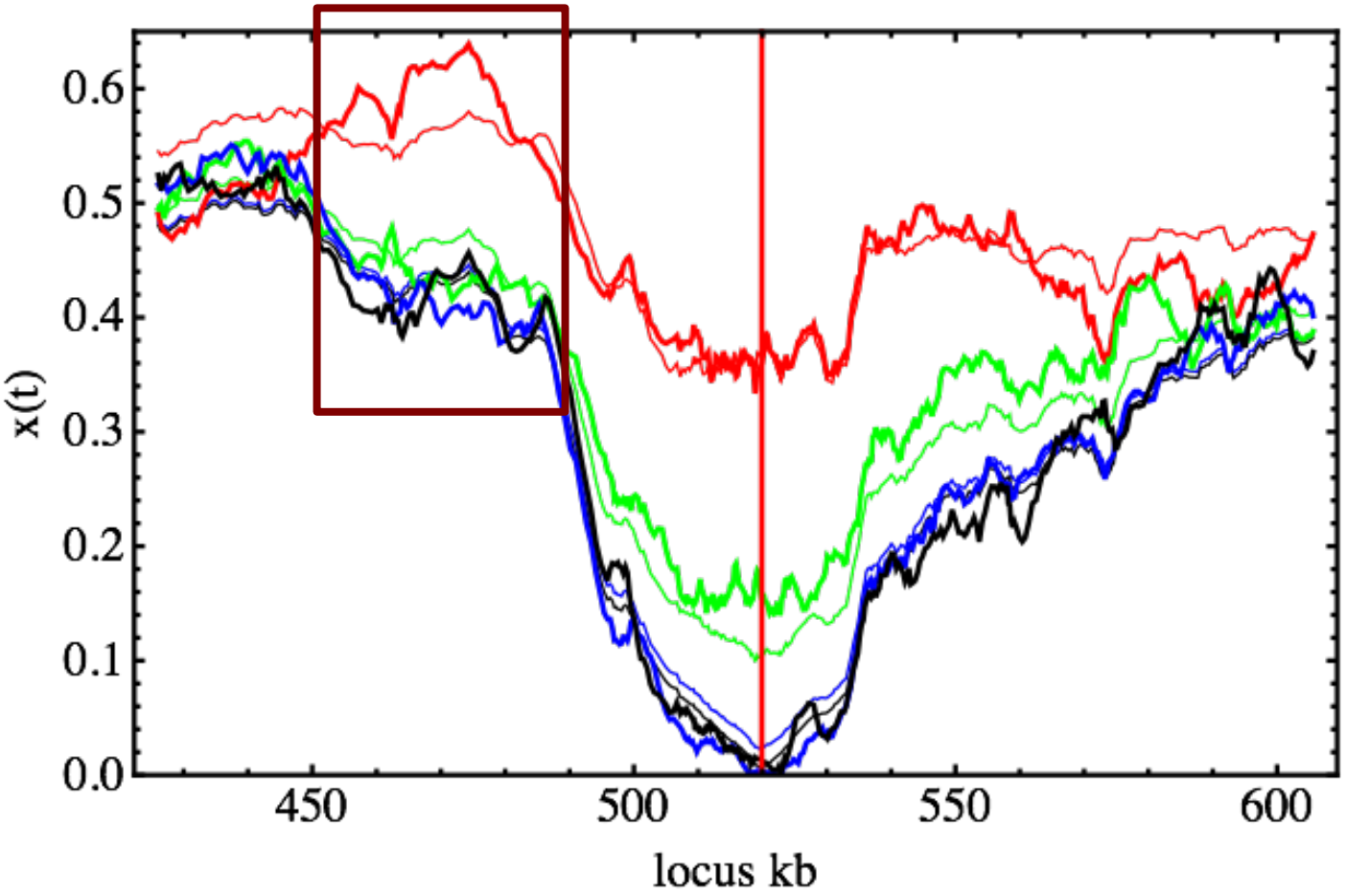
$$\Delta x_i^* = \sigma x_i^*(1 - x_i^*)$$

$$\Delta x_j = \sigma D_{i*j}$$

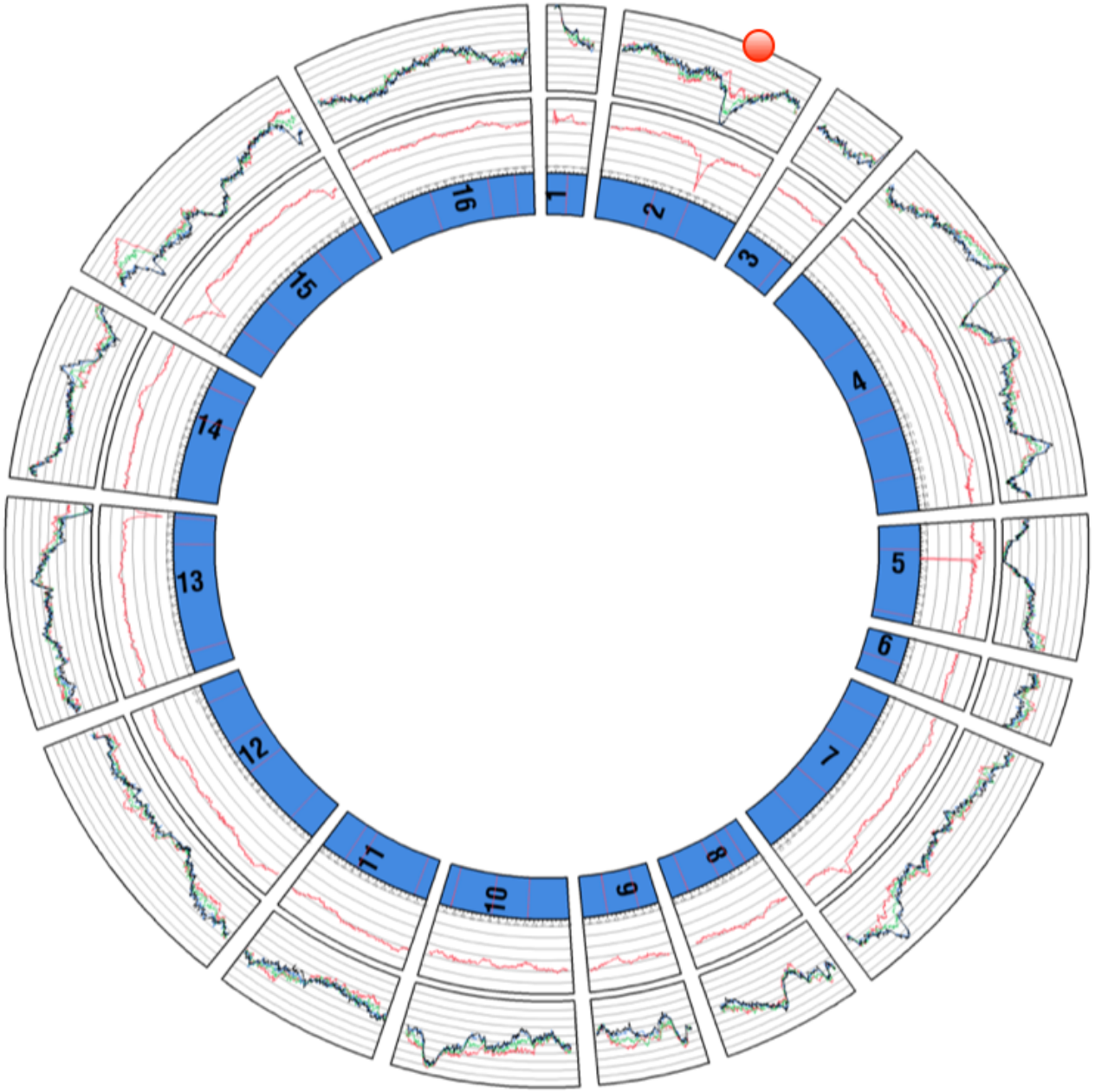
Model comparison

Region in chromosome II

Driver-passenger model: All linked to driver

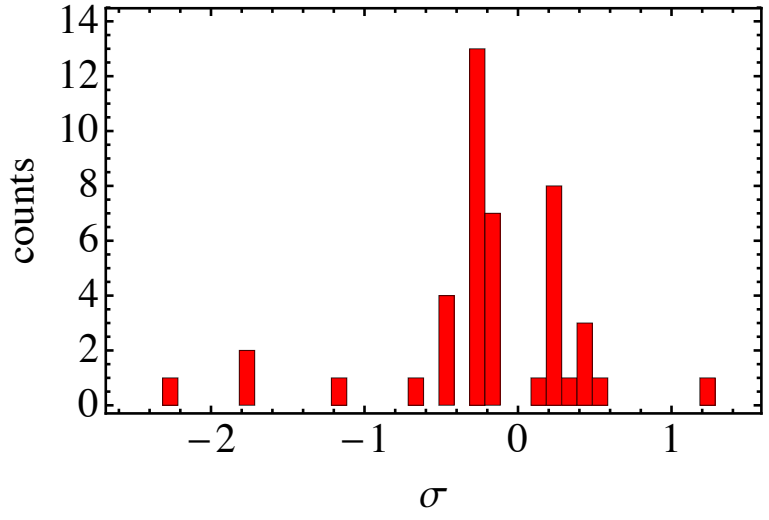


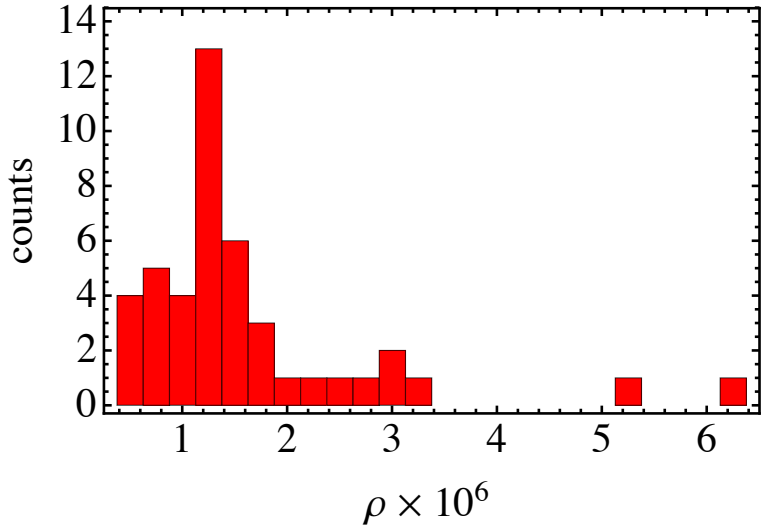
Position of inferred driver


16
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

Inference of selection genome-wide

44 alleles under selection

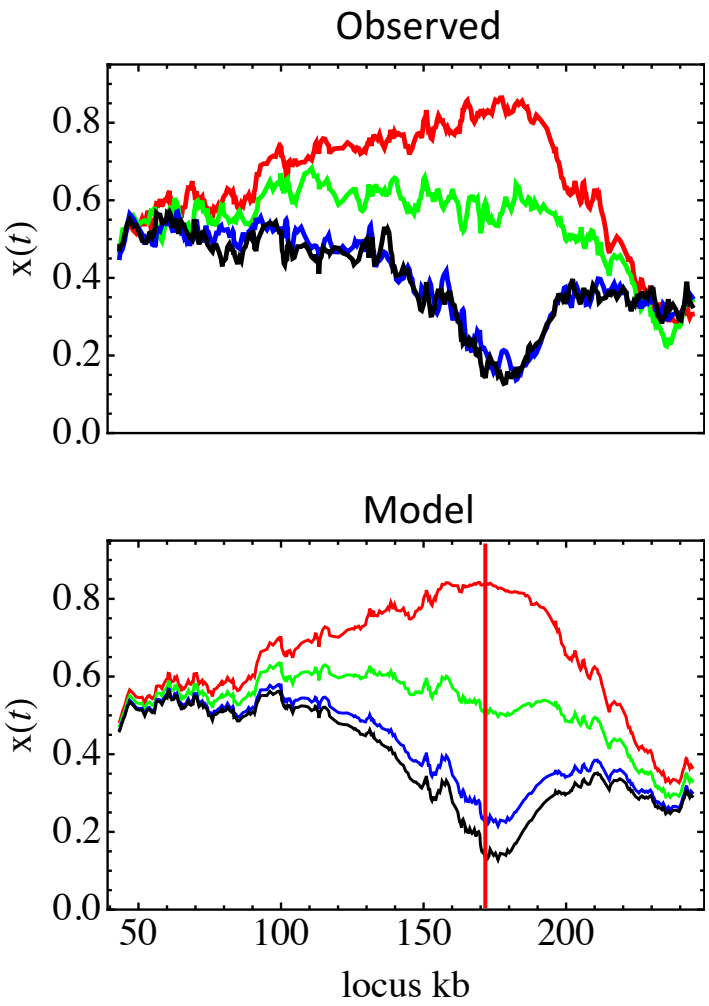




Selection for both strains: Low heat-tolerance strain has some beneficial genes for high temperatures

Limitations to the model

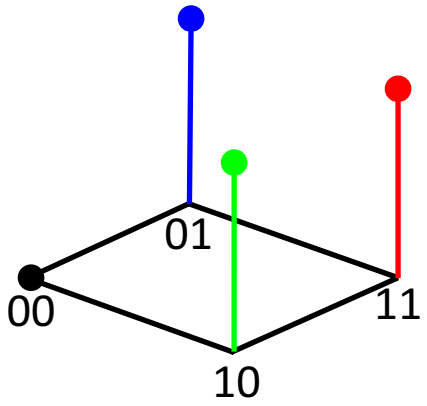
Some data visibly do not fit the model



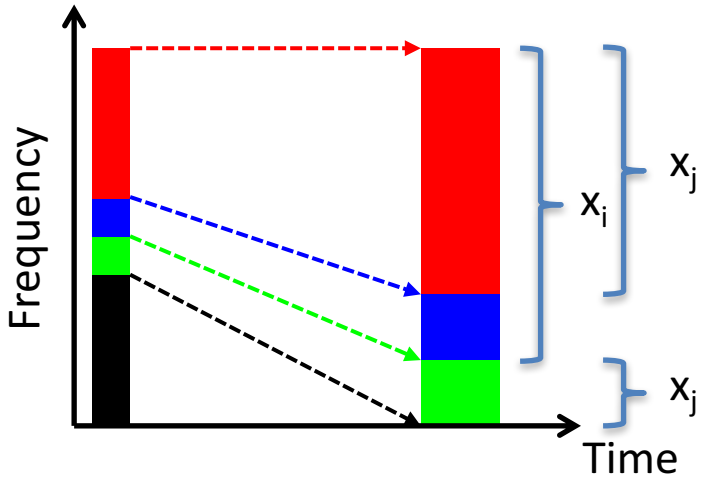
Chromosome XV: Apparent driver stops at intermediate frequency

Epistasis

Epistasis can lead to changes in allele frequency which stop at intermediate values:



Example haplotype fitnesses

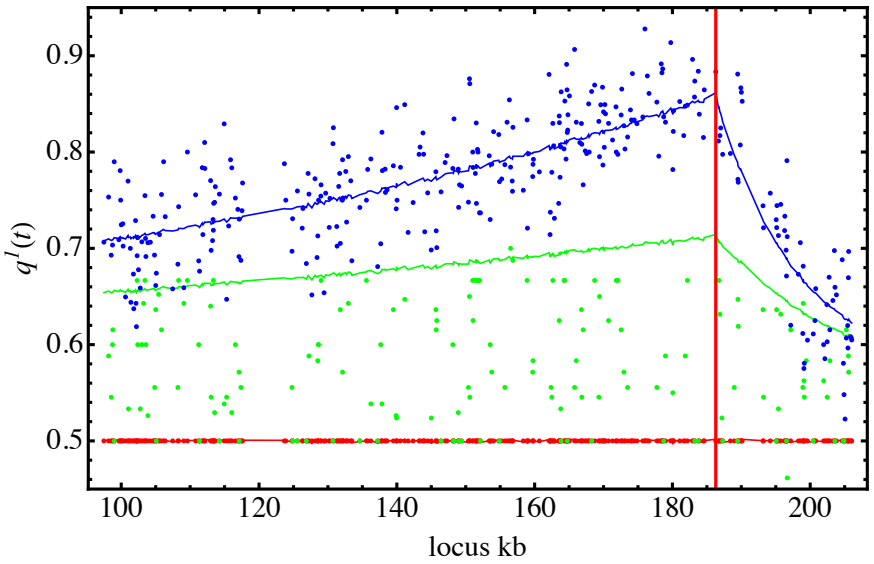


Evolution of haplotype frequencies

Variation in recombination rate

Extension of the model to incorporate multiple recombination rates

Data from chromosome XV before, during, and after the cross

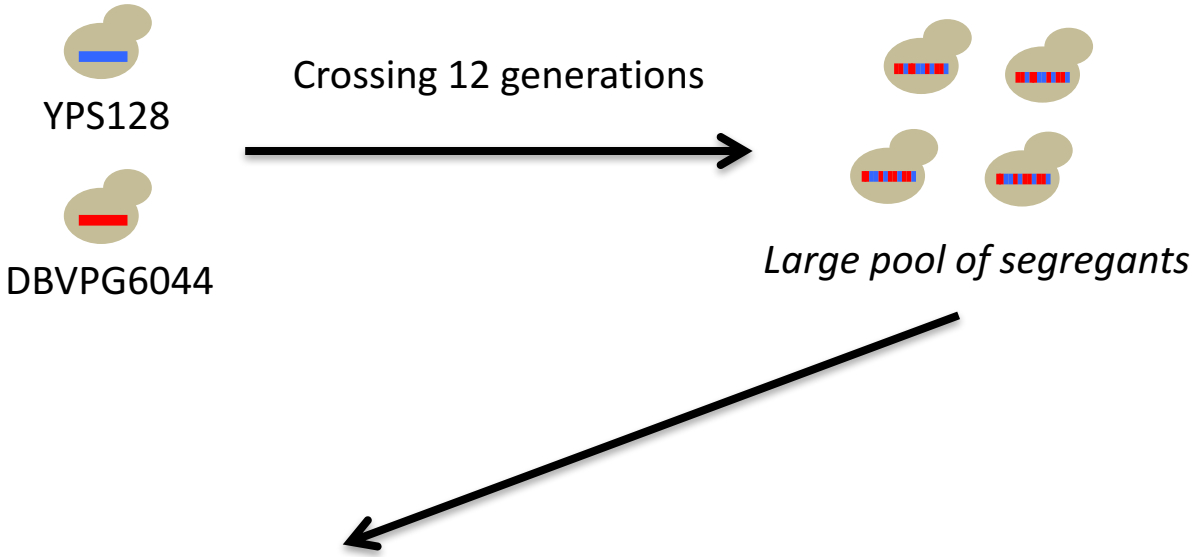


Model optimised for two recombination rates, either side of the driver

Substantial variation in recombination rate across the region of Chr XV

Variation in recombination rate

Use the cross itself to infer the recombination rate



Sequence individual genomes. Count values $\{n_{ij}^{00}, n_{ij}^{01}, n_{ij}^{10}, n_{ij}^{11}\}$ for all positions i,j.

Variation in recombination rate

Use the cross itself to infer the recombination rate

Haplotype frequencies depend upon the linkage disequilibrium between loci

$$q_{ij}^{ab}(\rho_{ij}) = q_i^a(t_0)q_j^b(t_0) + (-1)^{a+b}D_{ij}(\rho_{ij})$$

Known parameters

LD after the cross is a function of LD before the cross, and recombination

$$D_{ij}(\rho_{ij}) = D_{ij}^{\text{init}}(1-\rho_{ij}^{\text{tot}})^{N_c}$$

Known parameters

Given observations of haplotype frequencies, can infer the recombination rate

$$L(q_{ij}^{ab}(\rho_{ij}^{\text{tot}})|\mathbf{n}_{ij}) = \frac{N_{\text{s}}!}{\prod_{a,b\in\{0,1\}} n_{ij}^{ab}!} \prod_{a,b\in\{0,1\}} (q_{ij}^{ab}(\rho_{ij}^{\text{tot}}))^{n_{ij}^{ab}}.$$

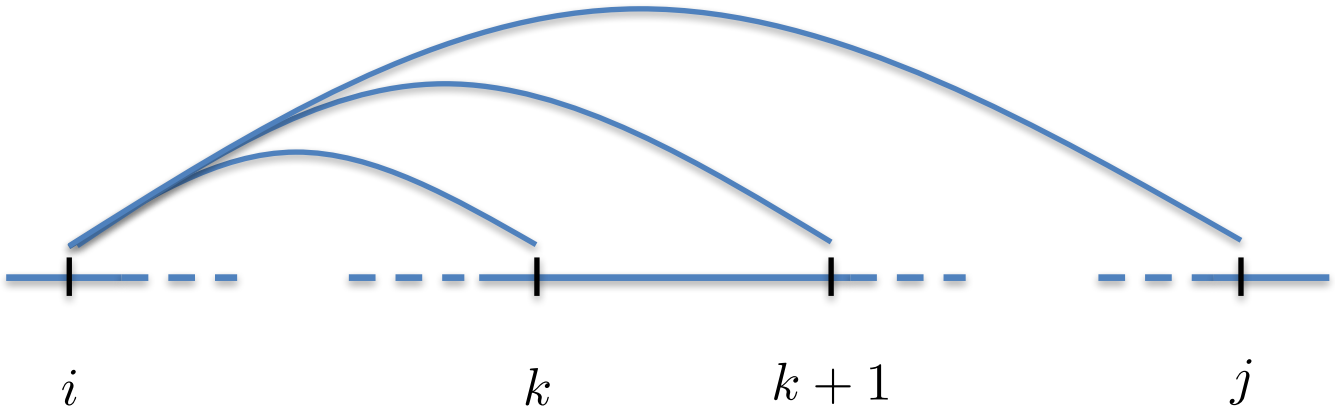
Known parameters

Learn the rate of recombination between alleles i and j

Variation in recombination rate

Use the cross itself to infer the recombination rate

Repeat pairwise inference across all pairs of loci

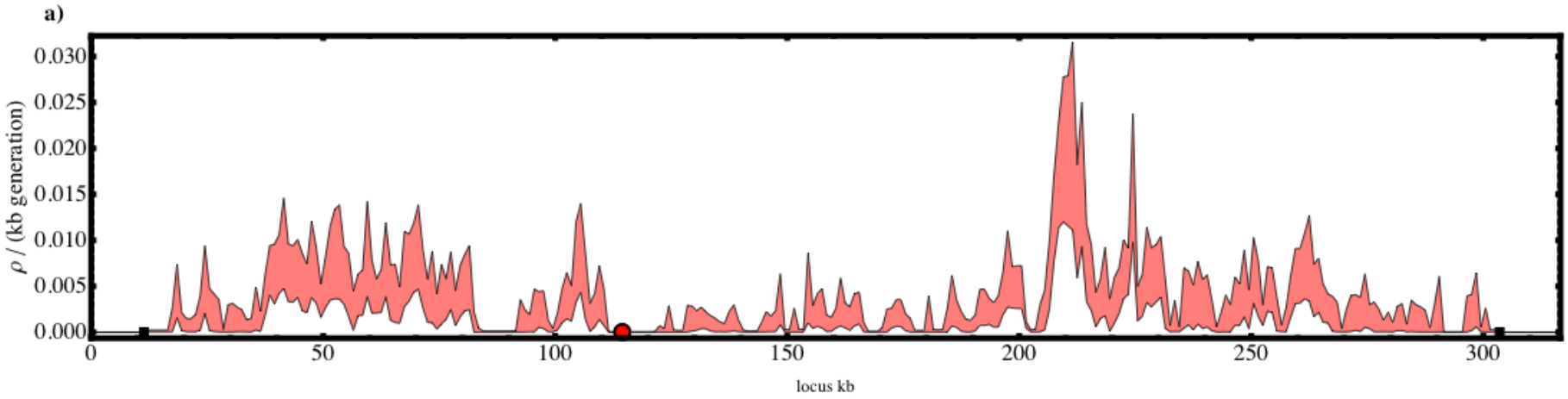


Calculate composite likelihood (product of likelihoods for all pairs)

Variation in recombination rate

Obtain map of recombination rate genome-wide

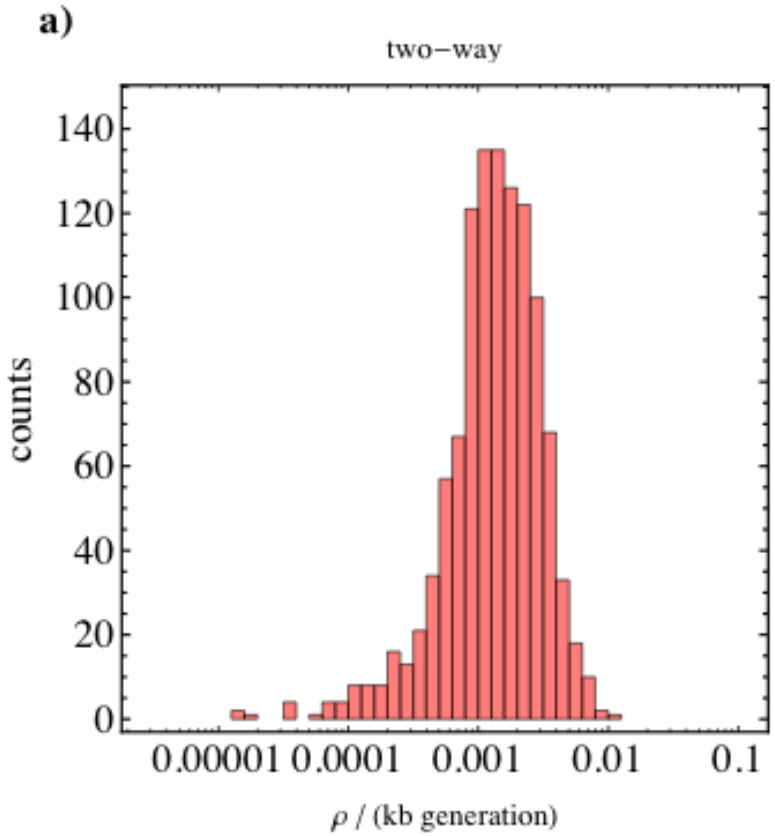
Data for chromosome III:



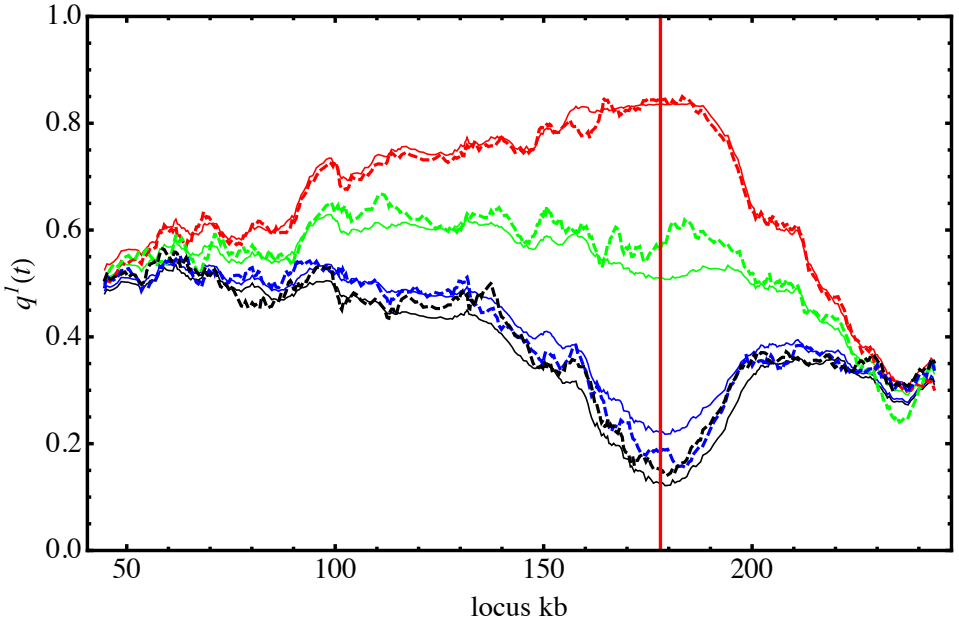
Variation in recombination rate

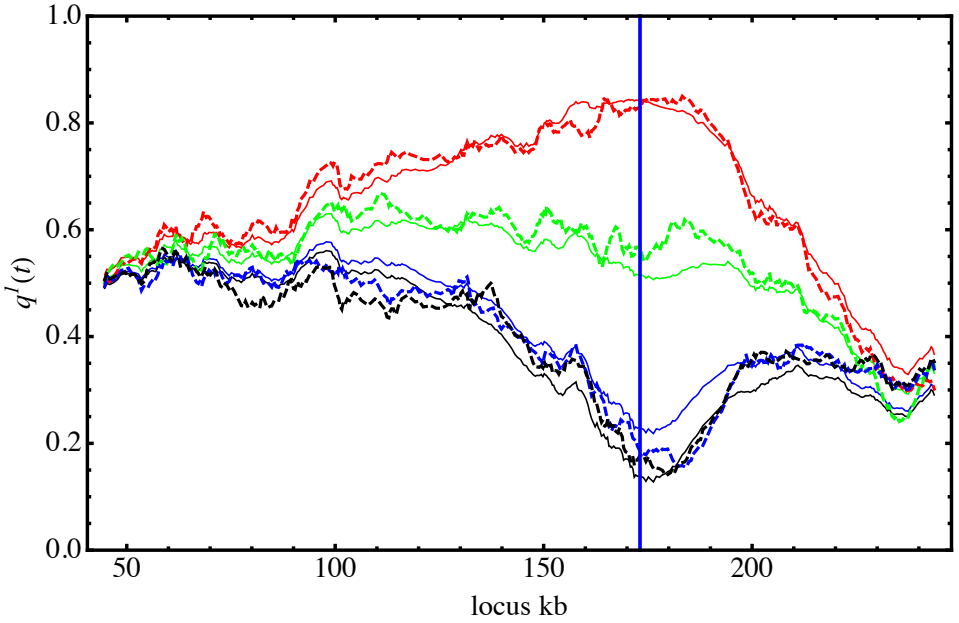
Obtain map of recombination rate genome-wide

Distribution of recombination rates genome-wide



Use recombination map to find selection





Chromsome XV data

With recombination map

Driver location shifted to the right

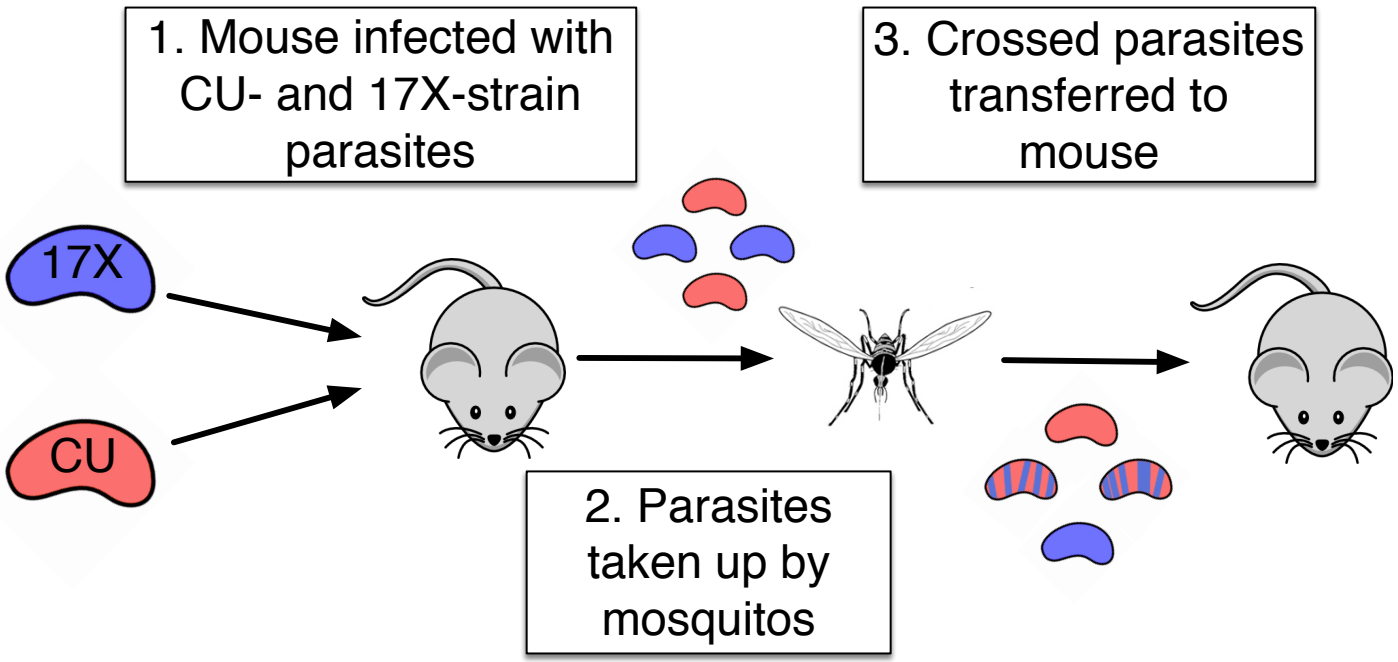
Improvement in overall model fit

Without recombination map

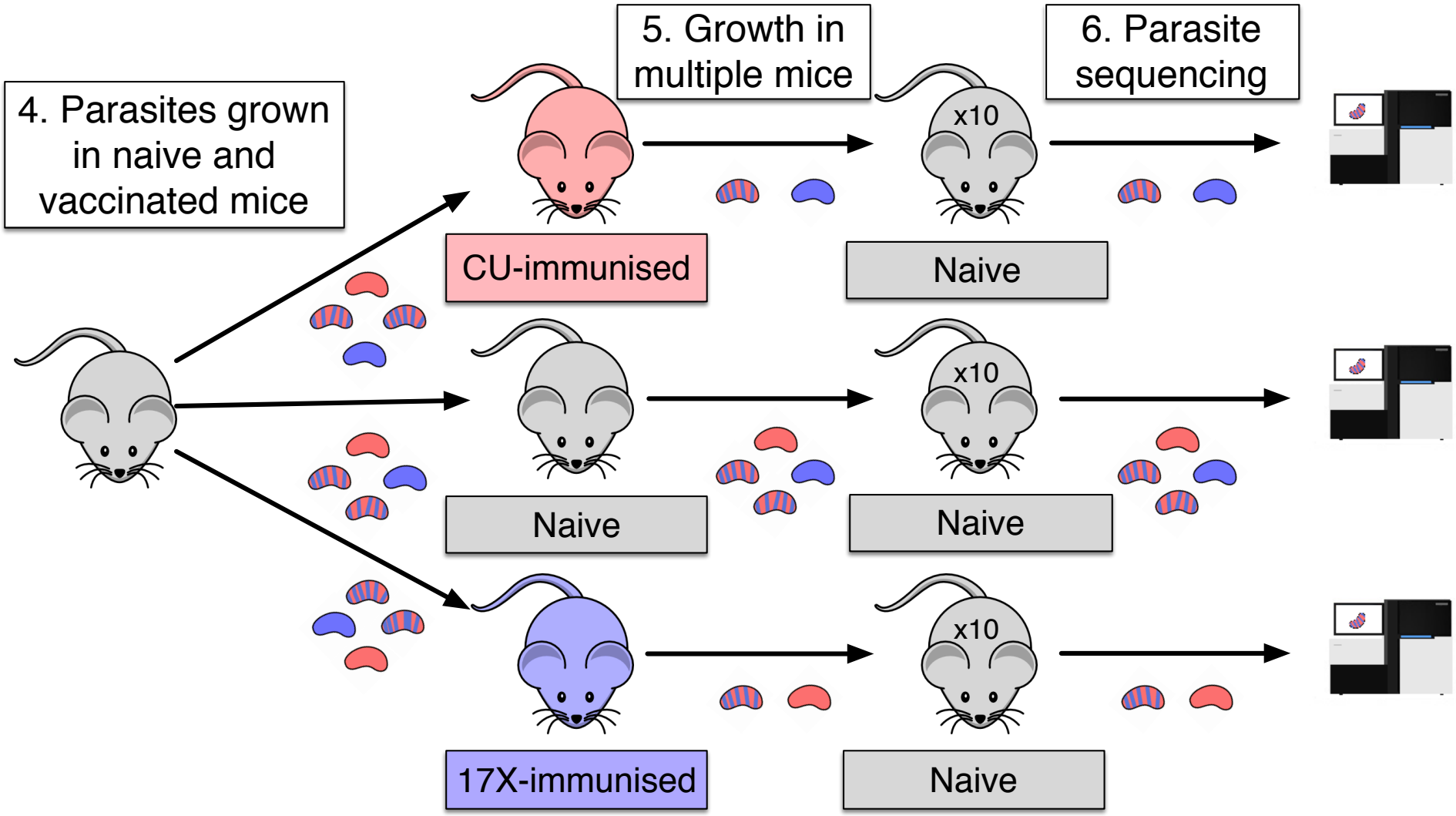
More complex multi-locus scenarios

Example IV: Crossed malaria parasites

Protocol:

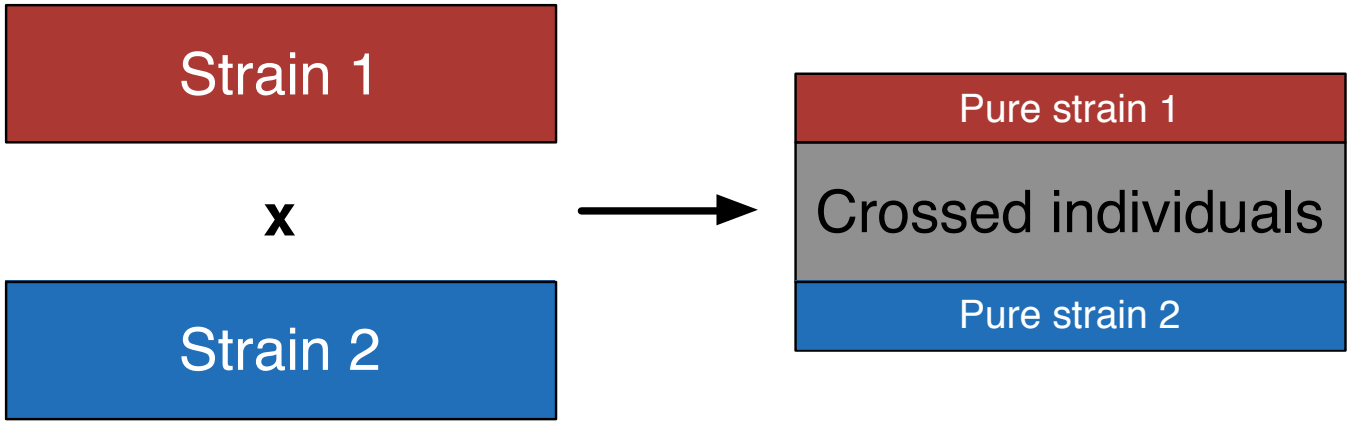


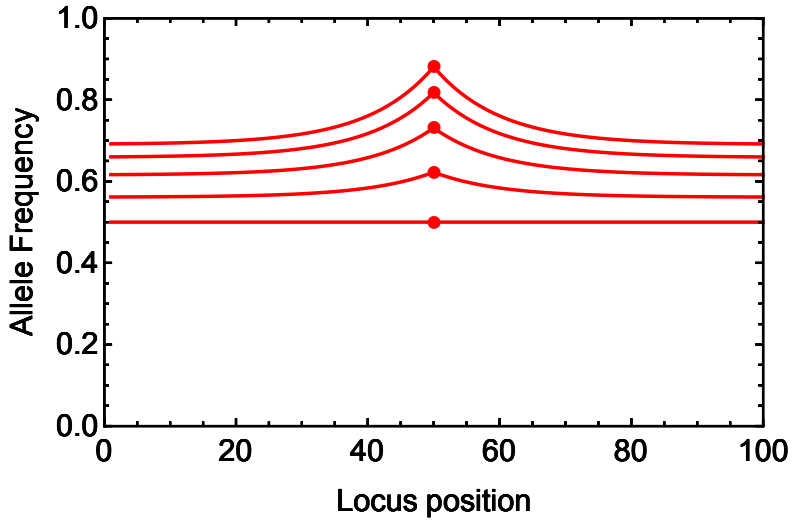
Example IV: Crossed malaria parasites



Example IV: Crossed malaria parasites

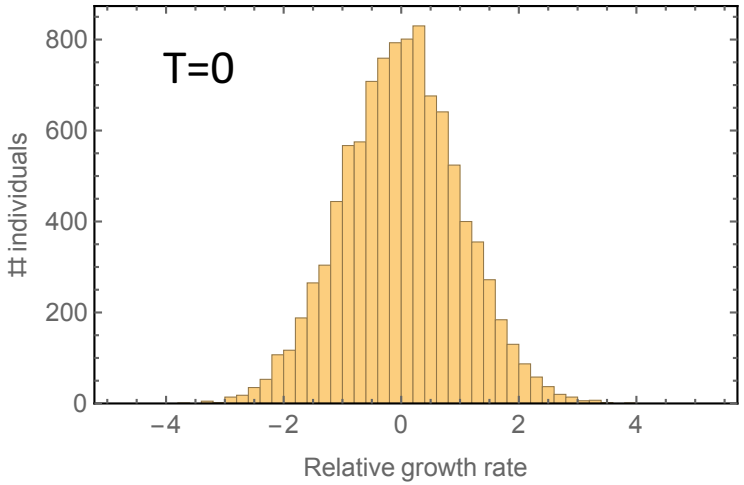
Single generation of crossing:

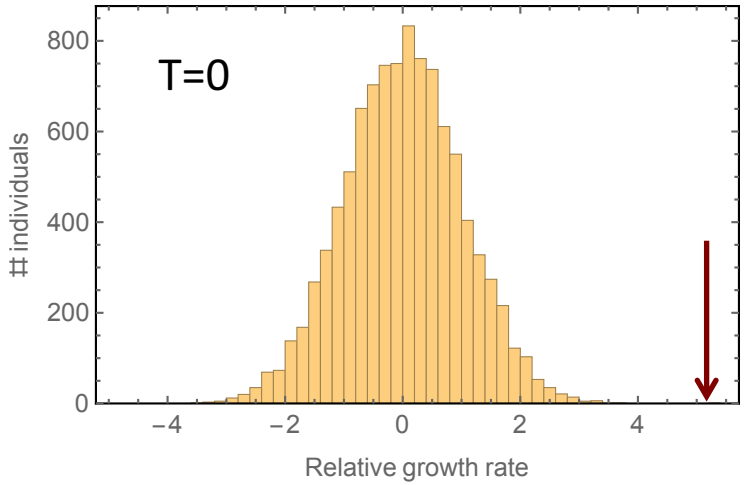




Example IV: Crossed malaria parasites

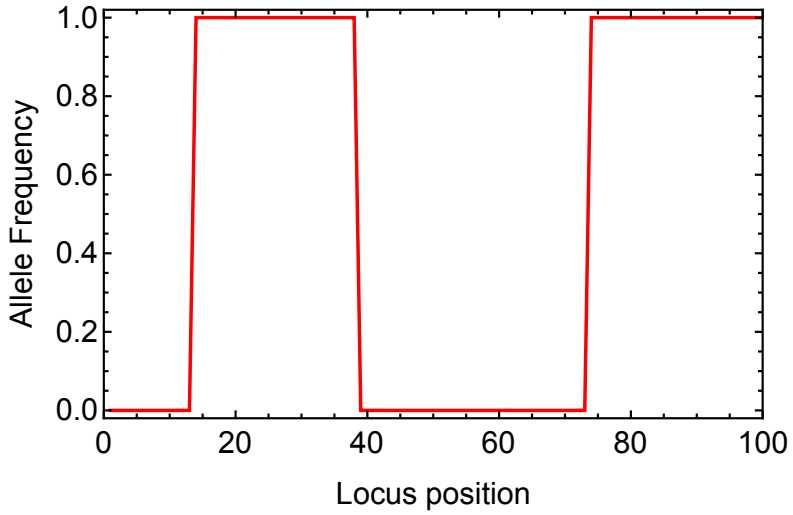
Clonal growth: high fitness individuals

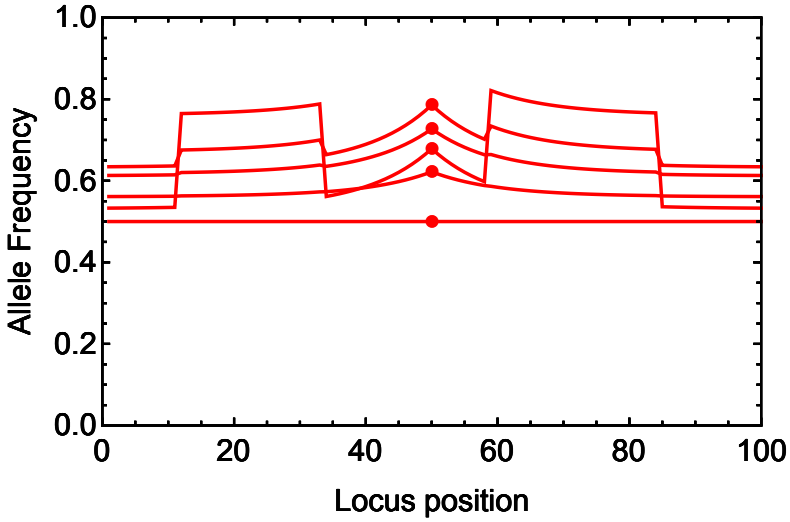




Example IV: Crossed malaria parasites

Clonal genotype switches between parental alleles

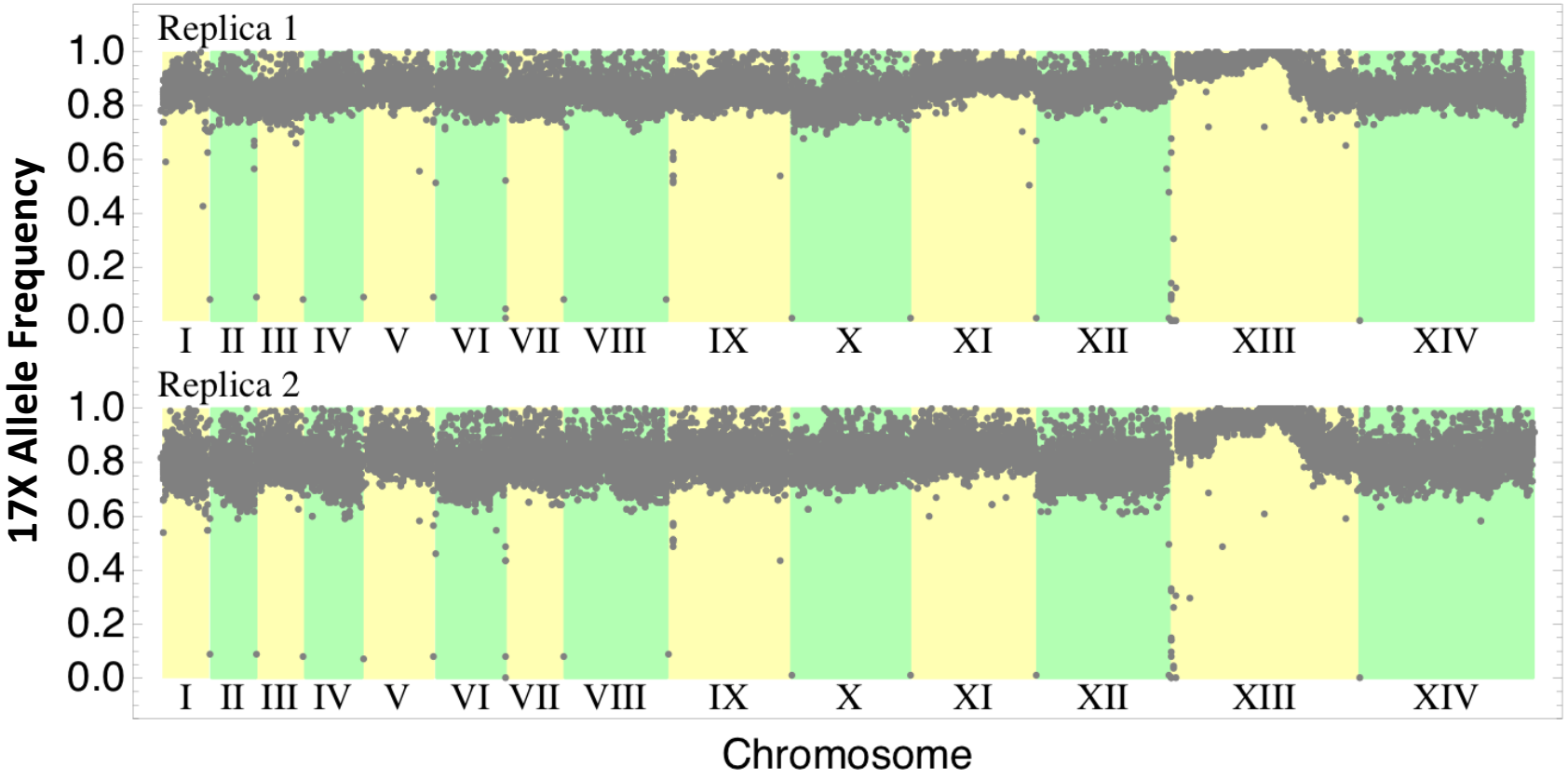




Comprises an increasing fraction of the population

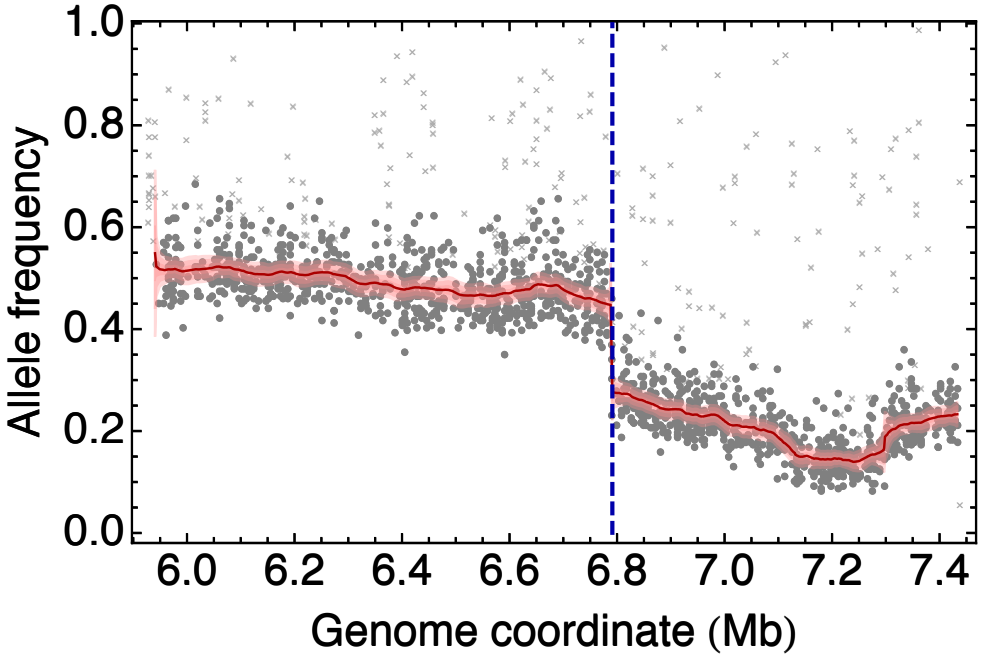
Generation of allele frequencies

Around 24,000 SNPs genome-wide



Detect clonal growth

Replica 1: Apparent sudden changes in frequency



Jump-diffusion model

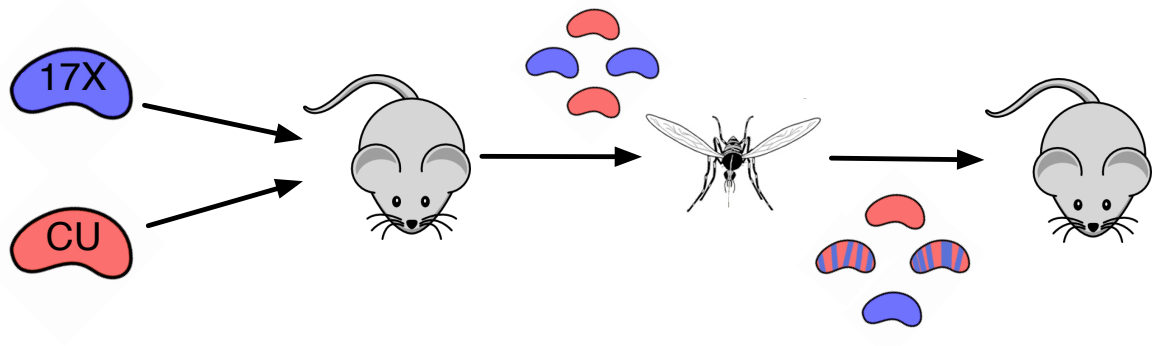
$$x_{i+1} = x_i + \mathcal{N}(0, s\sqrt{\Delta_{ij}})$$

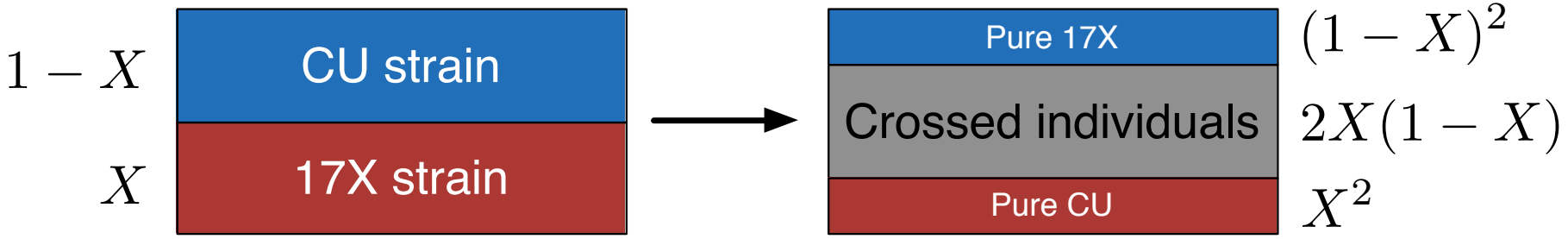
$$x_{i+1} \sim \mathcal{U}(0,1)$$

Consider regions separately

Location of selection

Model the process leading to the observed frequencies





Location of selection

Change in selected allele (t_c is time of cross):

$$x_i^1(t) = \frac{Xe^{\sigma(t-t_c)}}{1 - X + Xe^{\sigma(t-t_c)}}$$

Change in nearby allele:

$$x_j^1(t) = x_i^1(t)\frac{x_{ij}^{11}(t_c)}{x_i^1(t_c)} + x_i^0(t)\frac{x_{ij}^{01}(t_c)}{x_i^0(t_c)}$$

Pure genotypes: contribution of X^2 to x^{11}

contribution of 0 to x^{01}

Location of selection

Crossed genotypes:

$$\tilde{x}_{ij}^{11}(t_c) = \tilde{x}_i^1(t_c)\tilde{x}_j^1(t_c) + D'_{ij}e^{-\rho\Delta_{ij}}$$

$$\tilde{x}_{ij}^{11}(t_c) = \frac{1}{4}(1 + e^{-\rho\Delta_{ij}})$$

Frequency of crossed genotypes is 2X(1-X):

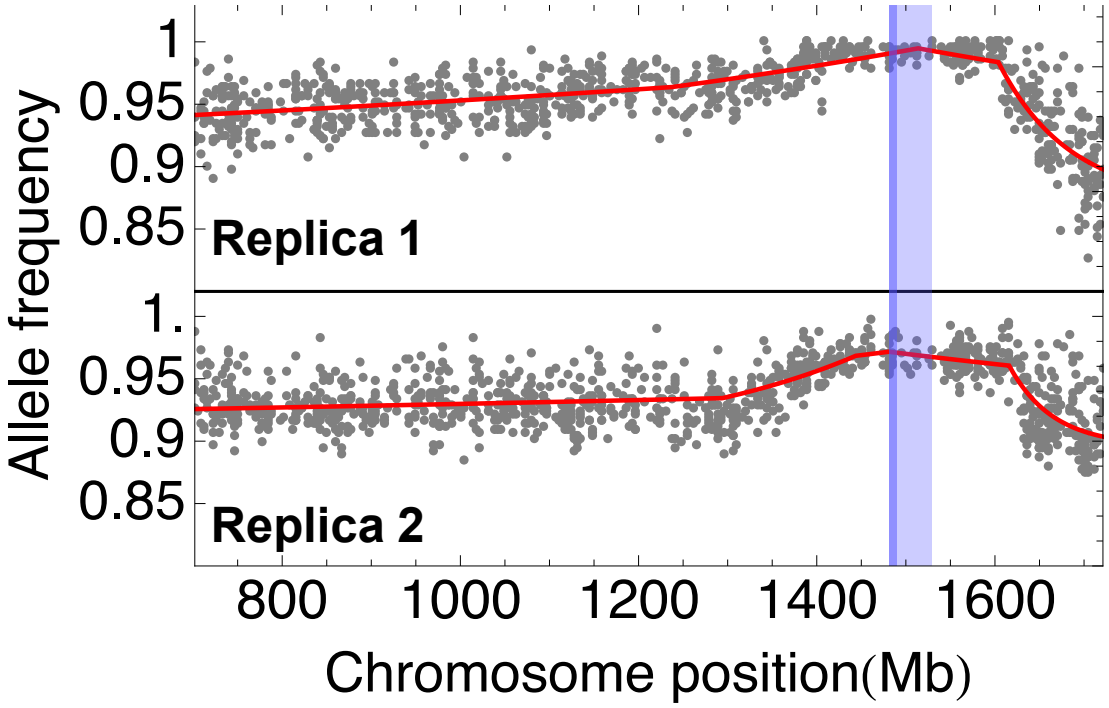
$$x_{ij}^{11}(t_c) = X^2 + \frac{1}{2}X(1-X)(1 + e^{-\rho\Delta_{ij}}).$$

$$x_{ij}^{01}(t_c) = \frac{1}{2}X(1-X)(1 - e^{-\rho\Delta_{ij}}).$$

$$x_j^1(t_o) = \left[X + \frac{1}{2}(1-X)(1 + e^{-\rho\Delta_{ij}})\right]x + \left[\frac{1}{2}X(1 - e^{-\rho\Delta_{ij}})\right](1-x) + e$$

Evolutionary model

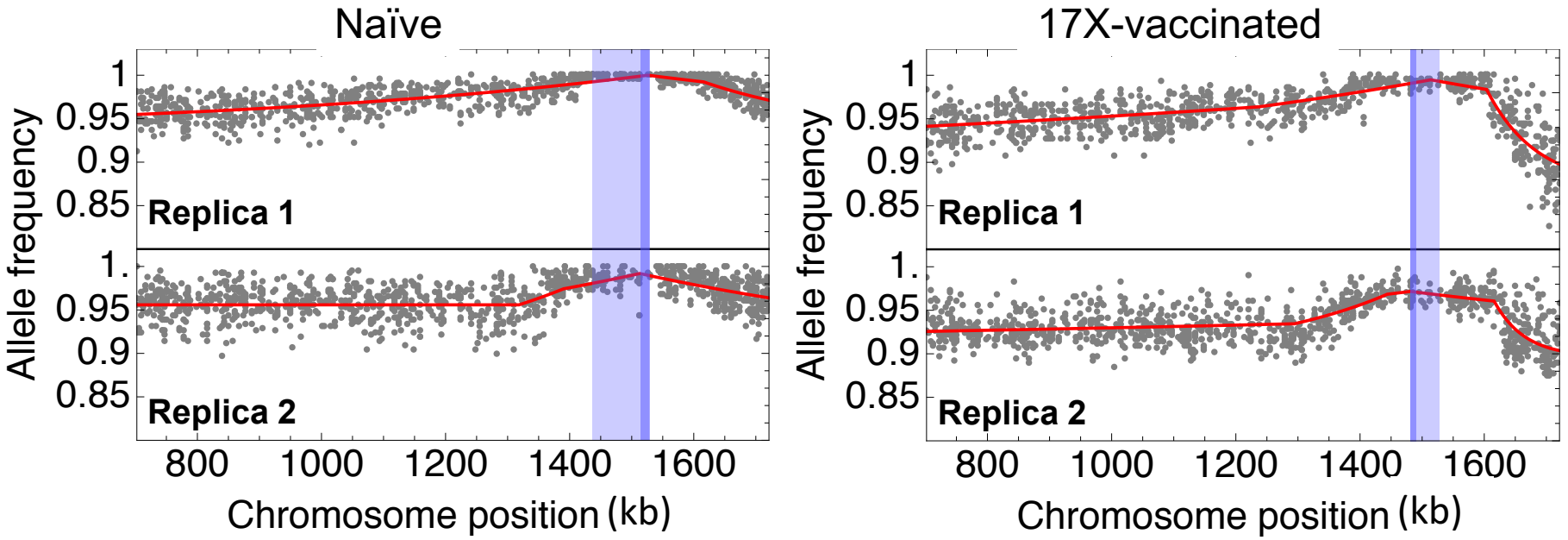
Identify confidence intervals by likelihood



More and less conservative intervals

Model outcome

Chromosome XIII



Growth allele: Position close to gene PyEBL

Erythrocyte Binding Ligand

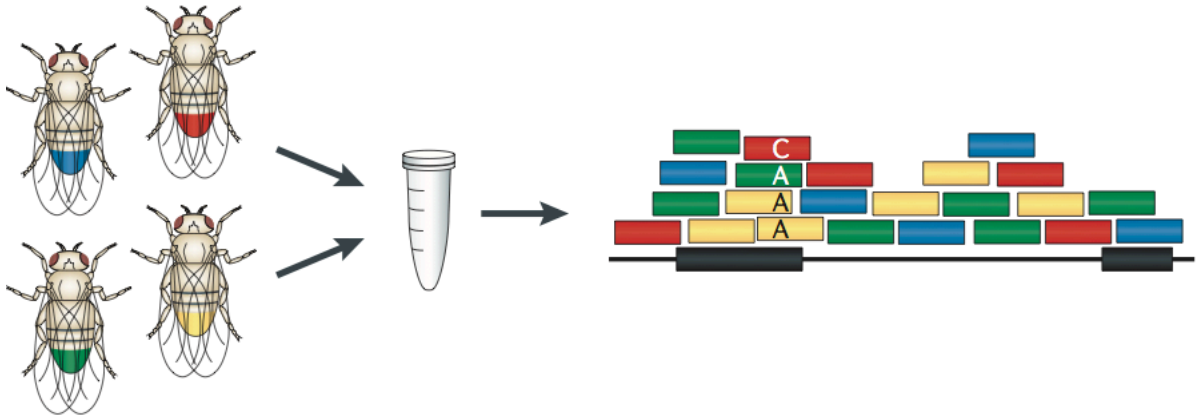
Evolutionary experiments

Example: “Evolve-and-resequence” experiment

Collect flies from a wild population

Grow under selection for 56 days

Sequence pooled population

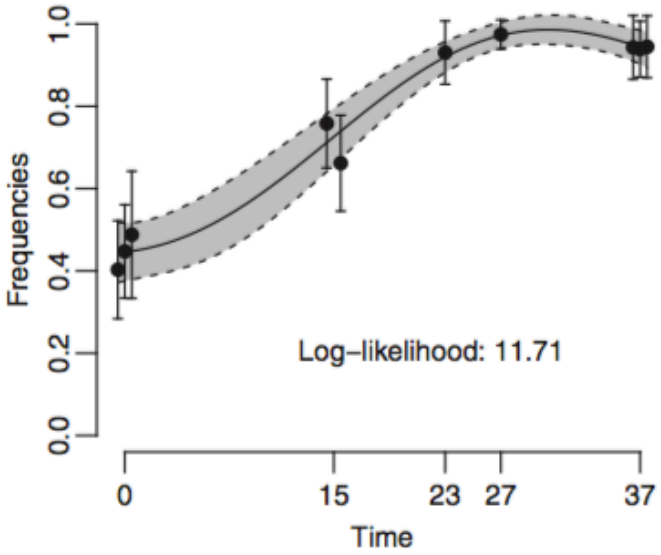


Schlotterer et al, Nature Reviews Genetics, 2014

Improved analysis

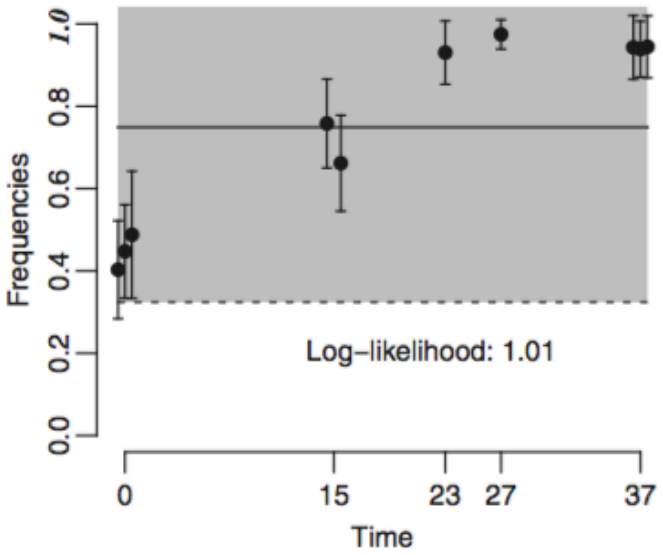
Use of Gaussian processes

Test for selection: Alleles that move more than expected given the effects of genetic drift



Time-dependent model:

$$m_{ij} = f_i\,(t_j) + \mu_{m_i} + \epsilon$$



Time-independent model:

$$m_{ij} = \mu_{m_i} + \epsilon$$

Topa et al, Bioinformatics, 2015

Improved analysis

Multi-locus inference

Use multiple (5-7) points in genome to conduct analysis

RESEARCH ARTICLE

Multi-locus Analysis of Genomic Time Series Data from Experimental Evolution

Jonathan Terhorst [1], Christian Schlötterer [2], Yun S. Song [1,3,4*]

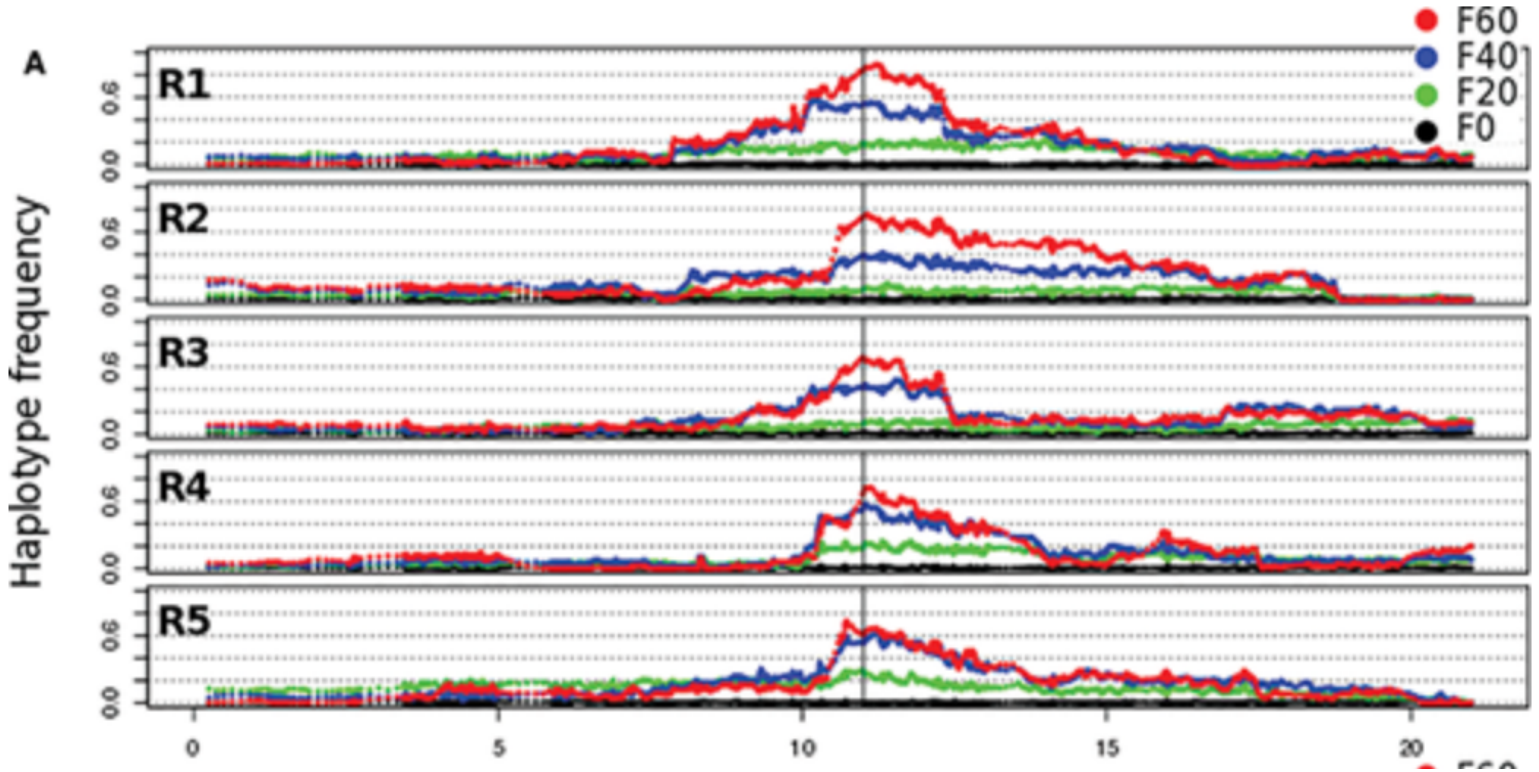
1 Department of Statistics, University of California, Berkeley, Berkeley, California, United States of America, **2** Institut für Populationsgenetik, Vetmeduni Vienna, Vienna, Austria, **3** Computer Science Division, University of California, Berkeley, Berkeley, California, United States of America, **4** Department of Integrative Biology, University of California, Berkeley, Berkeley, California, United States of America

Terhorst et al, PLoS Genetics, 2015

Improved analysis

Identification of haplotypes

Test for selection: Correlated allele frequency changes



Franssen et al, MBE, 2017

Phylogenetics + Coalescent theory

Recombination in humans

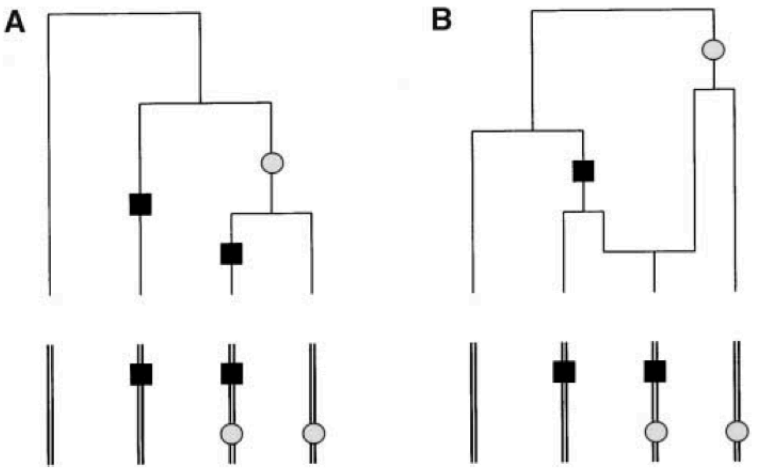
Inference from genealogy

Differences can arise through mutation or recombination

1. Estimate scaled mutation rate* θ

$$4N_e\mu \approx \hat{\theta}^*_W = \left(\sum_{k=1}^{n-1} \frac{1}{k}\right)^{-1} \ln\left(\frac{L}{L-S}\right)$$

2. Identify allele pairs for each site in the genome



Paired data Pair 1 : {AA, AT, TA, TA, AA}
Pair 2 : {GG, CG, CG, GG, GC}

L = length of genome
S = number of segregating sites

Allele pairs in each case {00, 00, 10, 10, 01}

McVean et al, Genetics, 2002

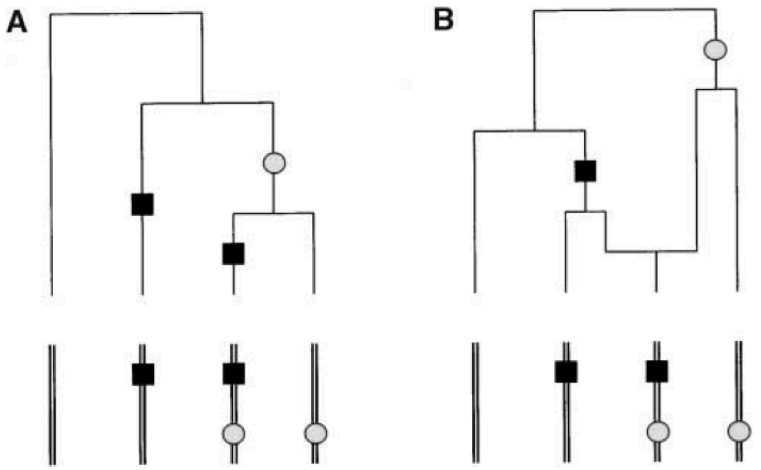
Recombination in humans

Inference from genealogy

Differences can arise through mutation or recombination

3. Calculate likelihood of observing each set of allele pairs given θ and a range of scaled recombination rates $4N_e r$

4. Calculate composite likelihood: sum of log likelihoods across all pairs of sites



Recombination rate between sites i and j $r_{ij} = \frac{r d_{ij}}{L-1}$ L = sequence length
i.e. constant recombination rate

Composite likelihood $\mathcal{L}(4N_e r) = \sum_{i,j} \mathrm{L}(X_{ij}|4N_e r_{ij})$

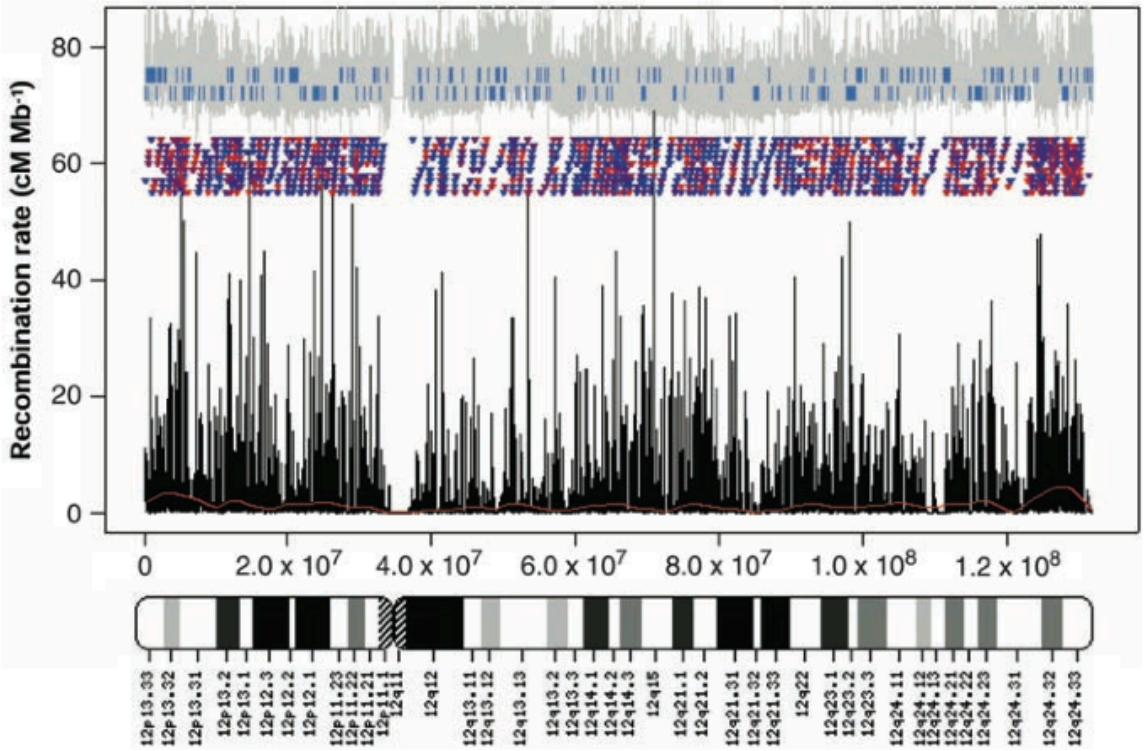
McVean et al, Genetics, 2002

Recombination in humans

Inference in human genome

Similar approach to that described above

Identified large variation in recombination rate across the genome



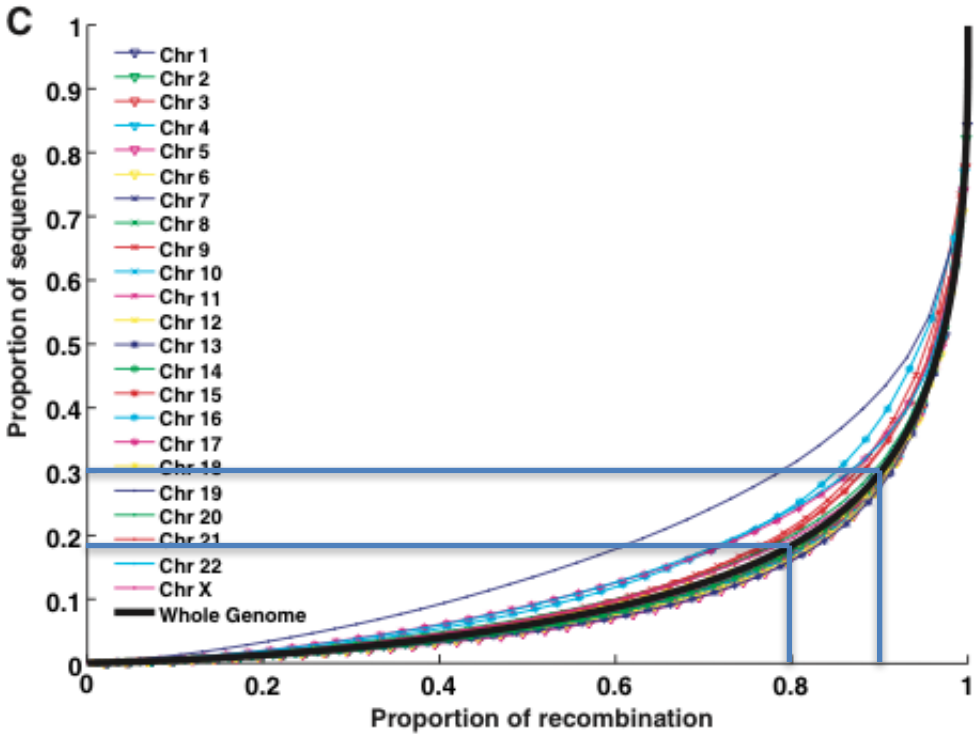
Myers et al, Science, 2005

Recombination in humans

Inference in human genome

Similar approach to that described above

Identified large variation in recombination rate across the genome



Recombination ‘hotspots’:

Regions of very frequent recombination

Myers et al, Science, 2005

Recombination in humans

Inference in human genome

Evaluated hotspots: search for common sequence motifs:

Short stretches of sequence that are over-represented

Repeat element	Hot spots containing element	Cold spots containing element	Corrected *P* value	Hot spot–motif region
THE1B	1,196	606	$<2 \times 10^{-16}$	TGTGAGG**CCTCCCTAGCCAC**GTGGAACT
THE1A	234	89	1.8×10^{-13}	GAGG**CCTCCCTAGCCAC**GT
GA-rich/ CT-rich	976	662	4.5×10^{-12}	**CCTCCCTT**
	1,005	737	7.5×10^{-8}	**CCTCCTT**
L2	10,113	9,271	7.8×10^{-7}	GAGGC**CCTCCCTGACCAC**C
AluY	3,634	3,284	1.4×10^{-2}	GATCCGC**CCNCCTTGGCCTC**CCA

Repeated motif CCTCCCTAGCCAC

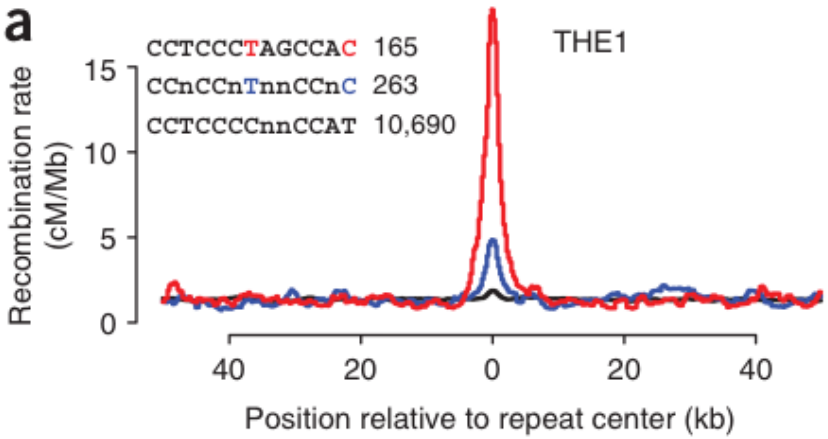
Myers et al, Nat Genet, 2008

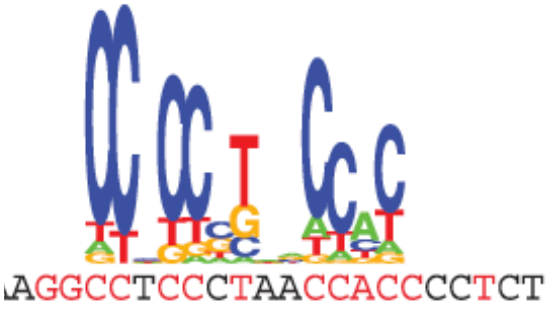
Recombination in humans

Inference in human genome

Recombination rate near identified motifs

Degeneracy in motif: Extent of enrichment in hotspots





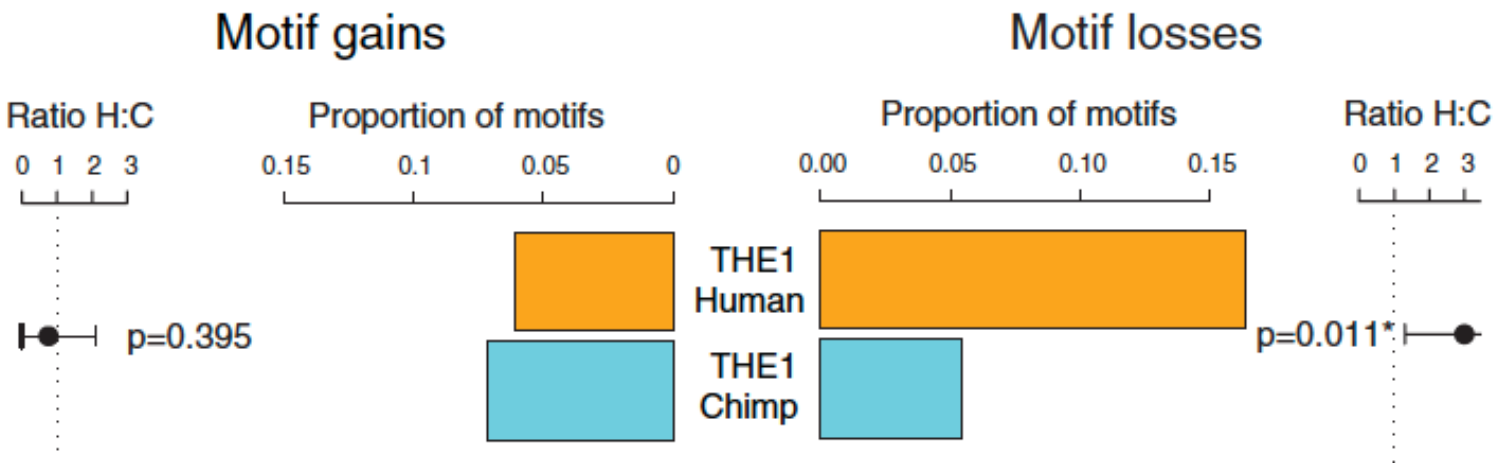
Repeated motif CCTCCCTAGCCAC

Myers et al, Nat Genet, 2008

Recombination in humans

Inference in human genome

Motif is less present in humans relative to chimpanzee genome

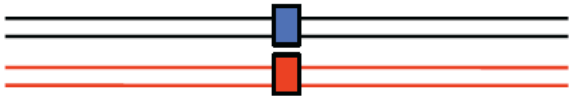


Currently being lost within the human genome

Myers et al, Science, 2010

Recombination in humans

Hotspot drive



Recombination hotspots disappear from the genome

Hotspot initiates double strand break and is eliminated from the genome

Repair does not restore the hotspot

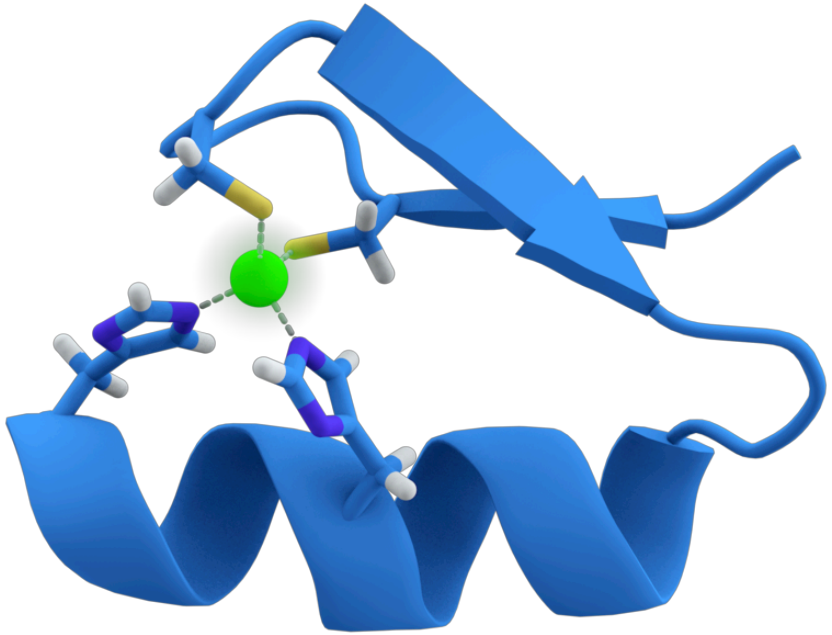
Resulting genome has lost the recombination hotspot

Webster and Hurst, Trends in Genetics, 2012

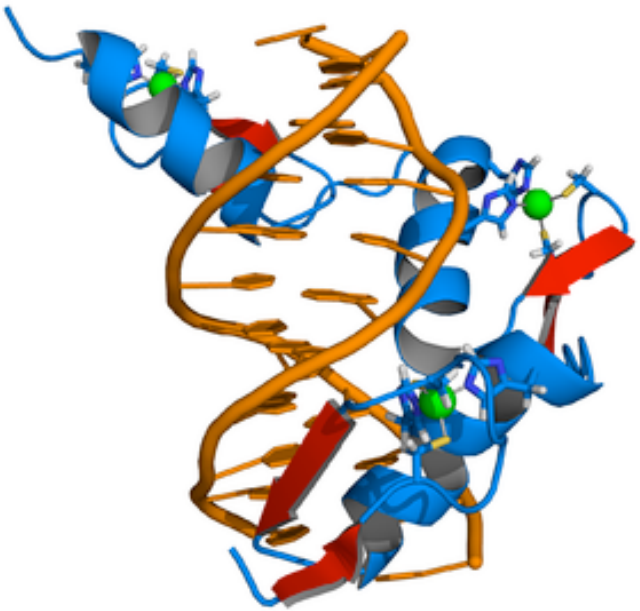
Recombination in humans

Inference in human genome

Motif is bound by a specific protein, PRDM9



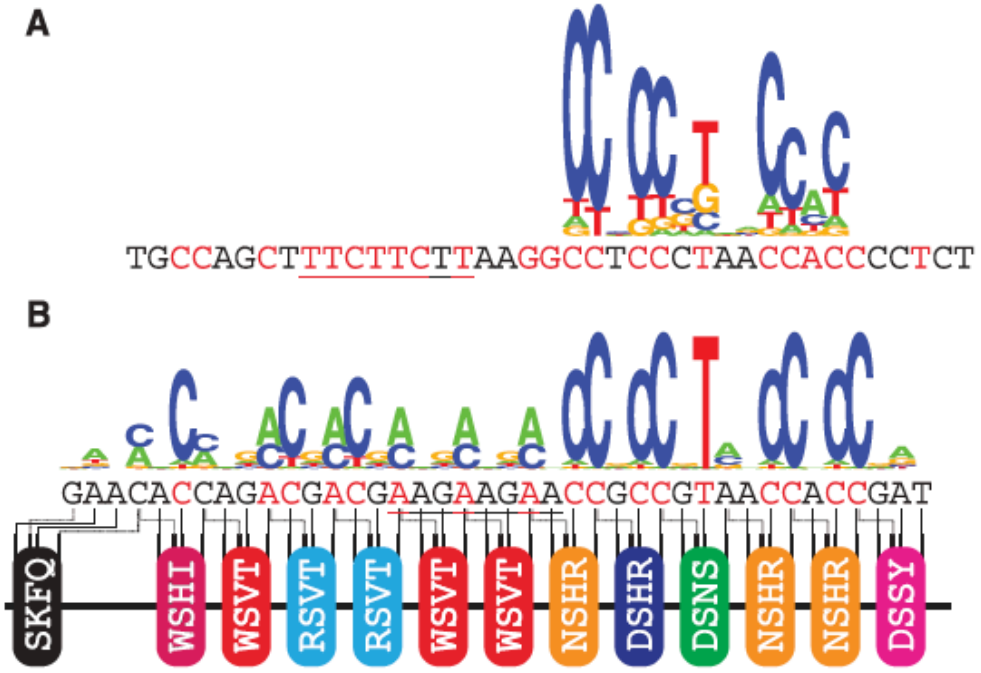
Zinc finger domain : DNA binding



Recombination in humans

Inference in human genome

Motif is bound by a specific protein, PRDM9



Equivalent gene in chimp does not bind the same motif

Gene shows signs of selection within humans

Continual selection for new recombination motifs?

Myers et al, Science, 2010