



UNIVERSITY OF  
CAMBRIDGE

# Localising Selection from Resequencing Data: Linking Genes to Phenotypes in Malaria Parasites

Chris Illingworth

University of Cambridge

# Acknowledgements

## **Nagasaki University**

Megumi Inoue  
Hussein Abkallo  
Richard Culleton

## **KAUST, Saudi Arabia**

Ho. Y. Shwen  
Arnab Pain

## **Sanger Institute**

Andrej Fischer  
Ville Mustonen  
  
Axel Martinelli

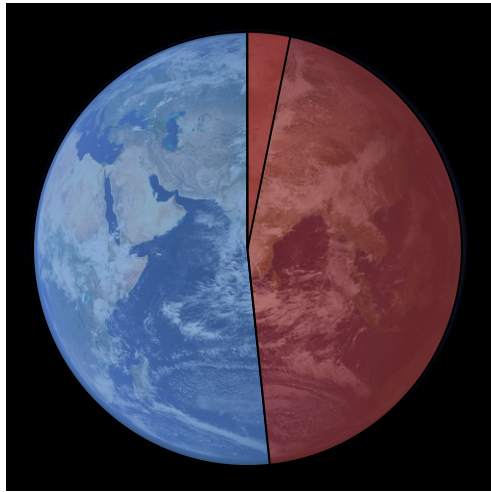
## **Edinburgh University**

Paul Hunt

**welcome**trust

THE  
ROYAL  
SOCIETY

# Malaria

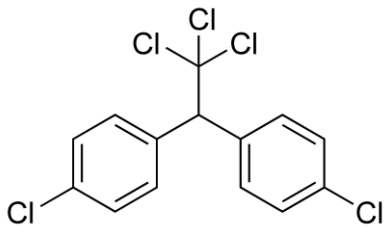


3.4 billion people at risk

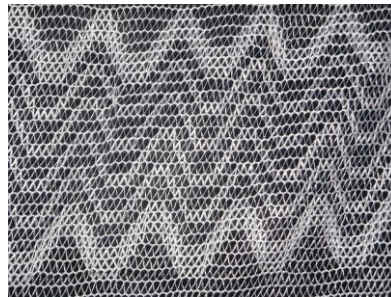
c. 200 million infections  
per year

600,000 deaths

## Control strategies



DDT



Mosquito netting



Drug therapy

# Malaria

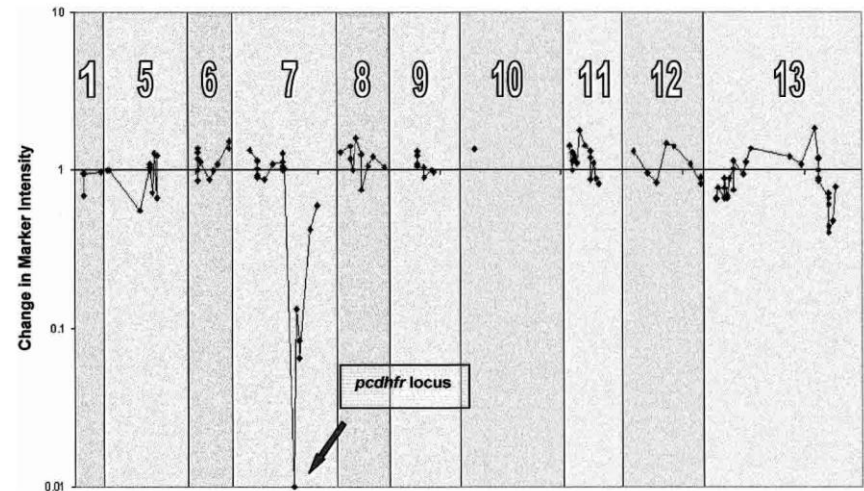
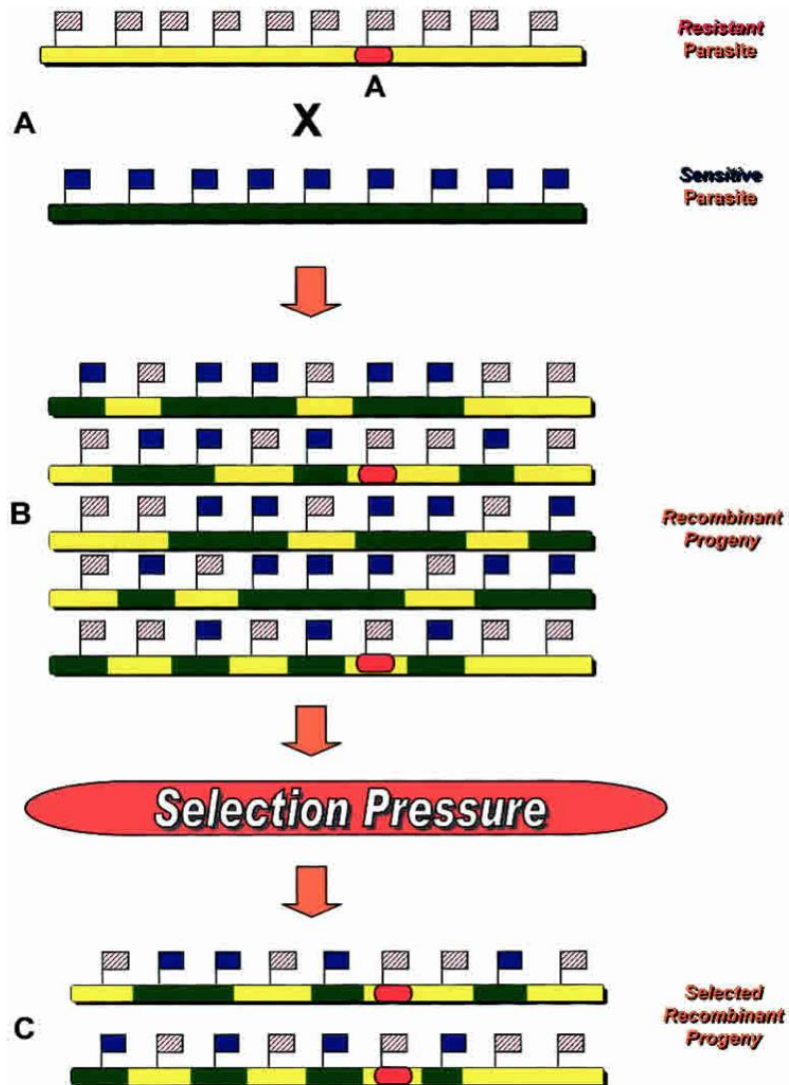
## **Genetic diversity and control**

Identify genetic factors underlying medically important phenotypes

Growth rate

Strain-specific immunity

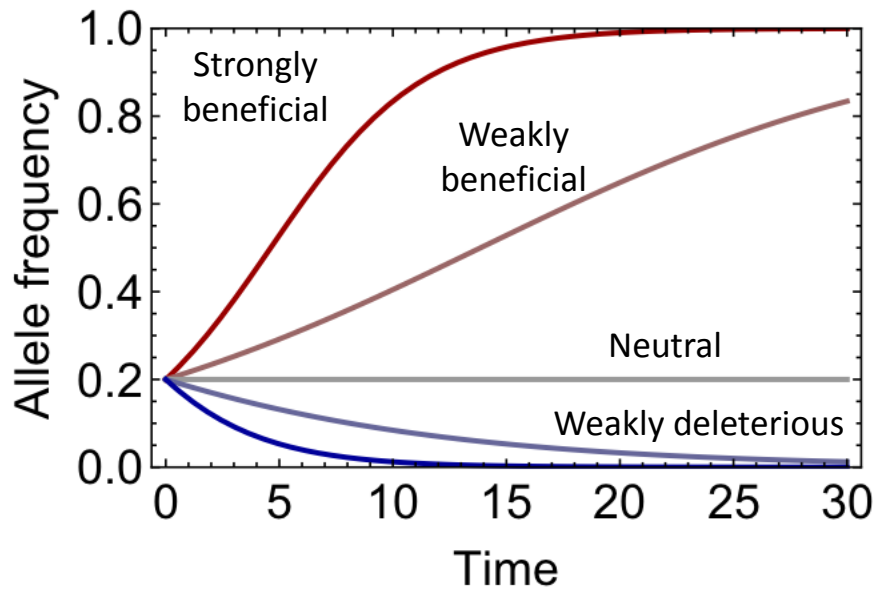
# Linkage Group Selection



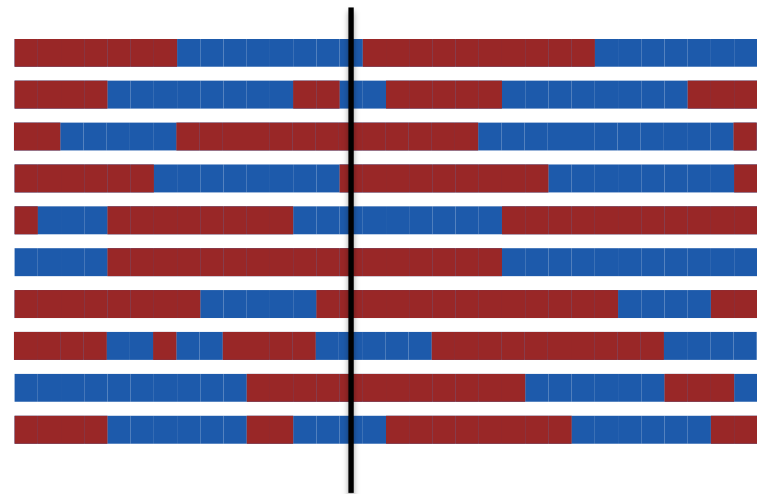
Identify selected alleles

# What do you get if you cross?

Effect of selection upon an allele frequency



Single locus

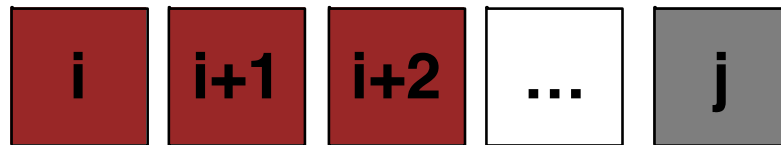


Multiple correlated loci

$$\frac{dx_i}{dt} = \sigma x_i (1 - x_i)$$

# Evolutionary dynamics

Correlation between allele frequencies



Even number of recombination events in  $[i,j]$



Odd number of recombination events in  $[i,j]$

$$P(j \text{ is red}) = 1 - \exp(-\rho\Delta_{ij})$$

$\rho$ : Recombination rate

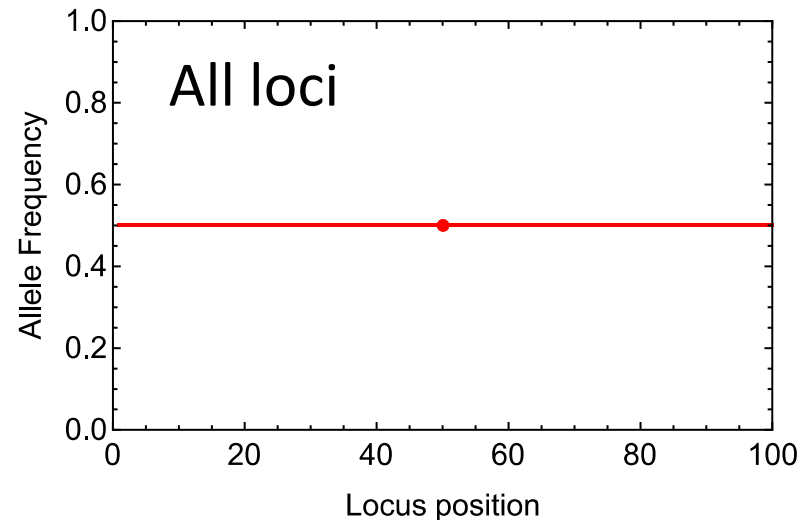
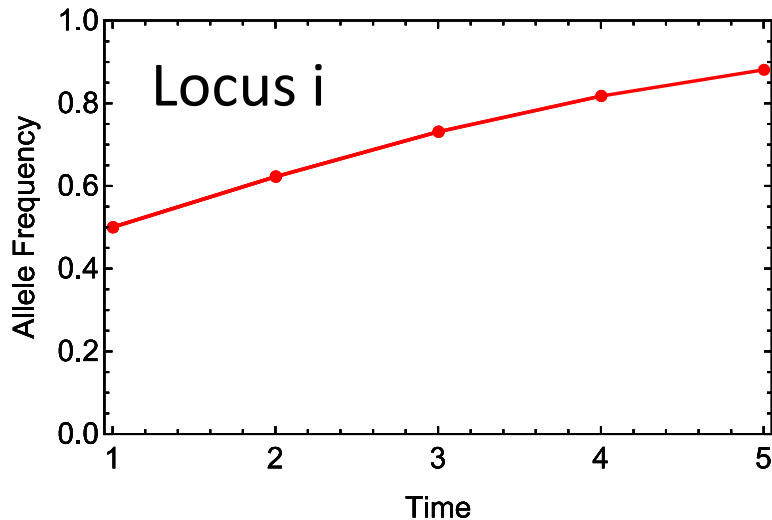
$\Delta_{ij}$ : Distance in sequence between  $i$  and  $j$

# Evolutionary dynamics

Selection for allele at locus  $i$

Change in frequency at locus  $i = \Delta x_i$

Change in frequency at locus  $j = \Delta x_i \exp(-\rho \Delta_{ij}) D'_{ij}$

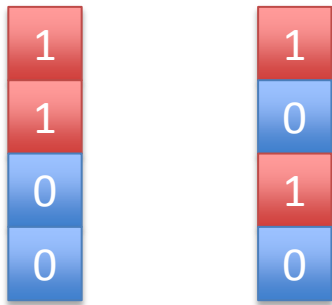




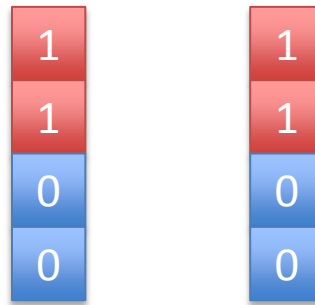
# Linkage disequilibrium $D_{ij}$

Extent of association of alleles at two loci

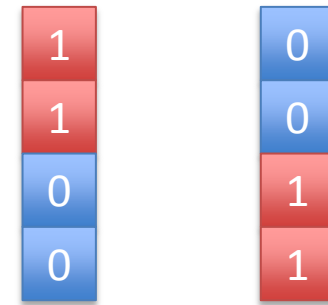
$$D_{ij} = q_{ij}^{11} - q_i^1 q_j^1$$



$$D_{ij} = 0$$



$$D_{ij} = 0.25$$



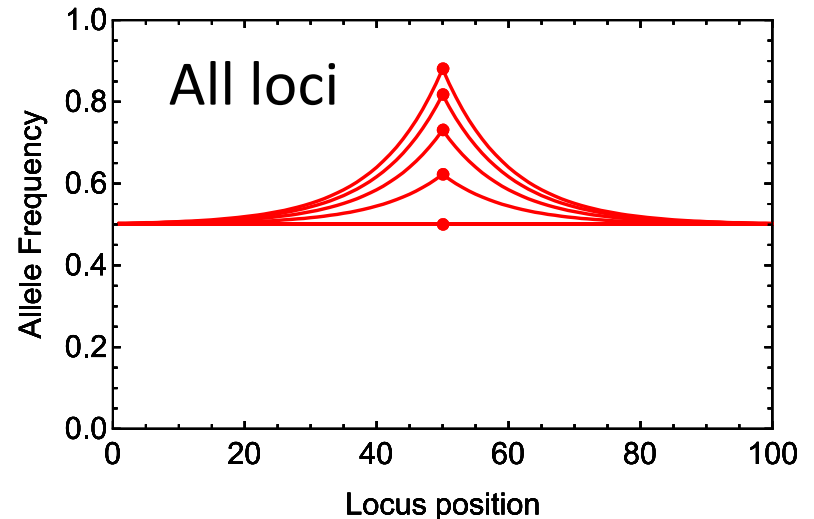
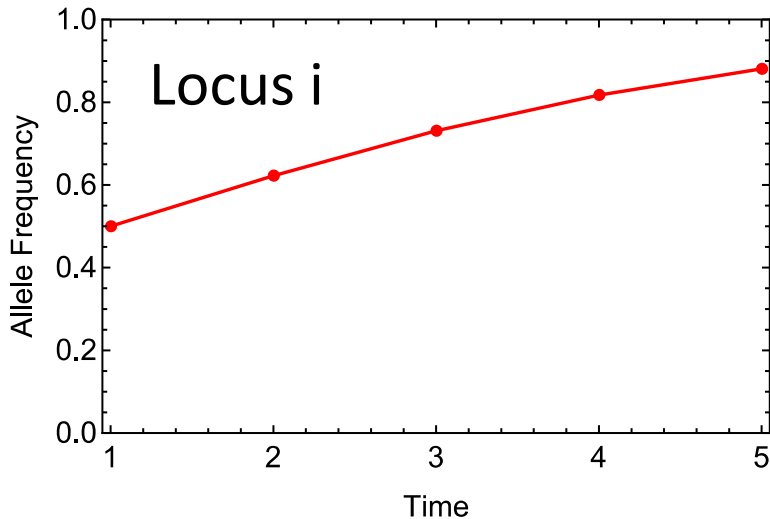
$$D_{ij} = -0.25$$

# Evolutionary dynamics

Selection for allele at locus  $i$

Change in frequency at locus  $i = \Delta x_i$

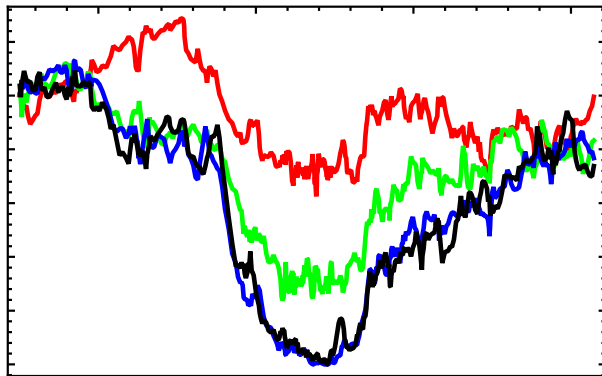
Change in frequency at locus  $j = \Delta x_i \exp(-\rho \Delta_{ij}) D'_{ij}$



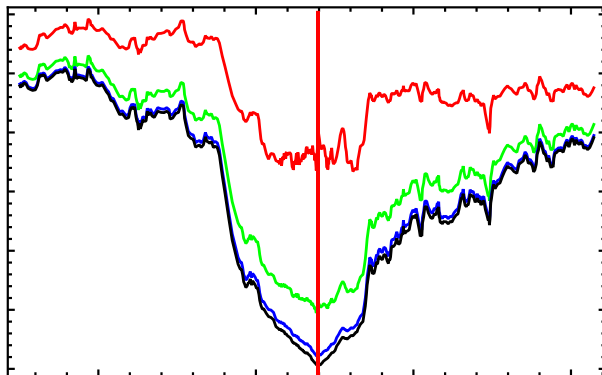
# Evolutionary dynamics

Example: fitting time-resolved data

Observed



Model



Time

1

2

3

4

Learn:

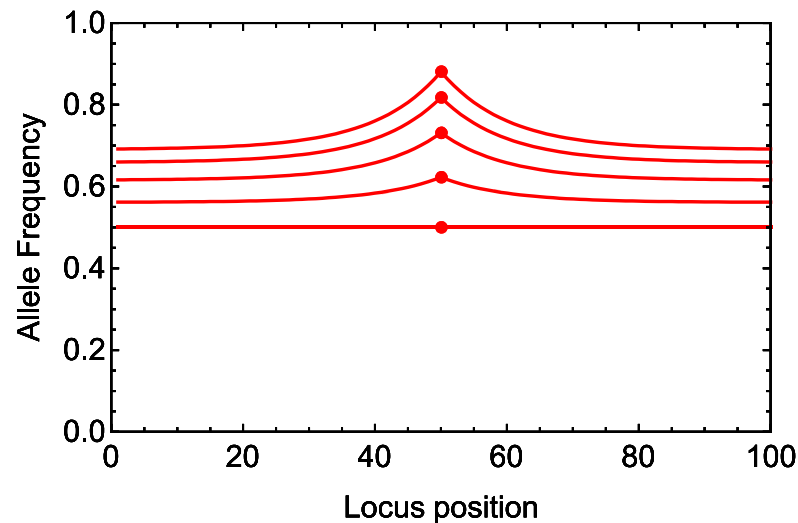
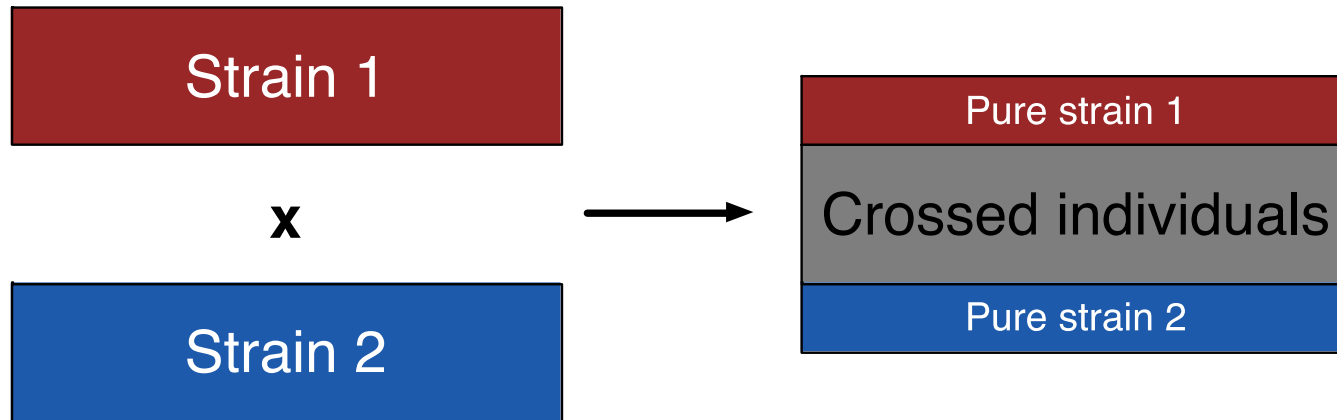
Initial allele frequencies

Locus and strength of selection

Rate of recombination

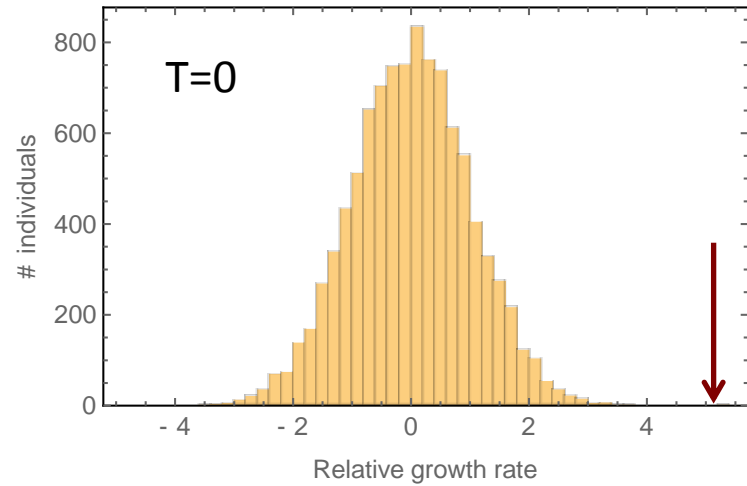
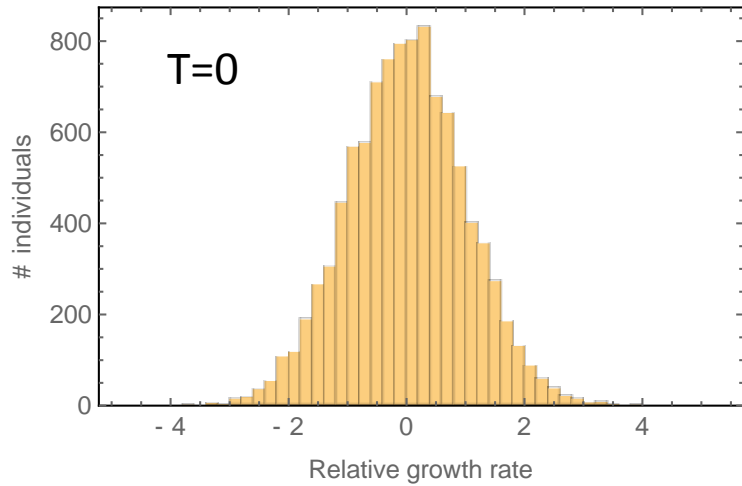
# Further considerations

Pure genotypes:



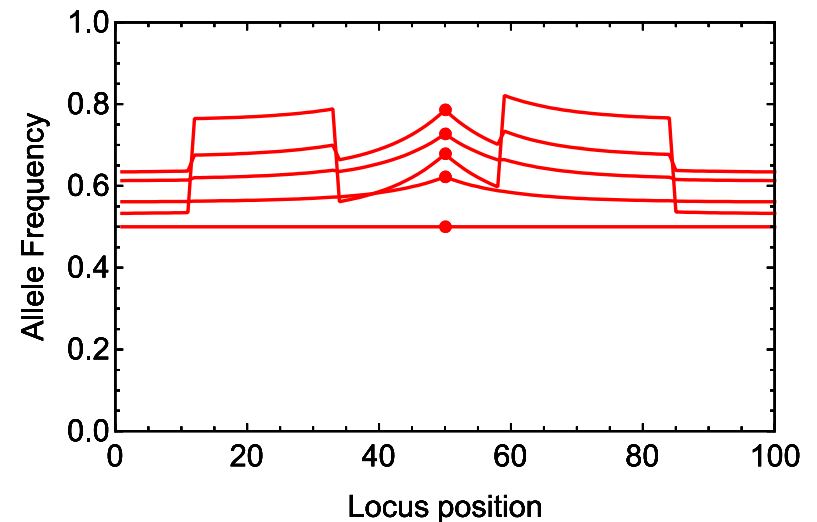
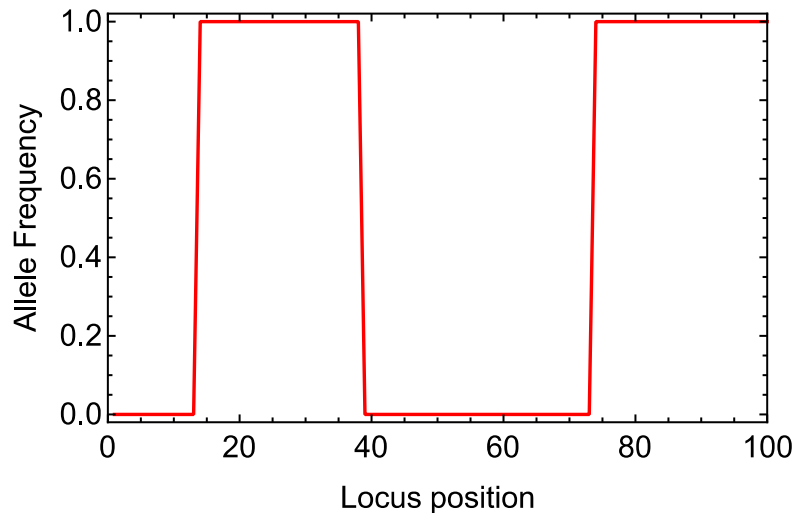
# Further considerations

## Clonal growth: high fitness individuals



# Further considerations

Clonal genotype switches between parental alleles

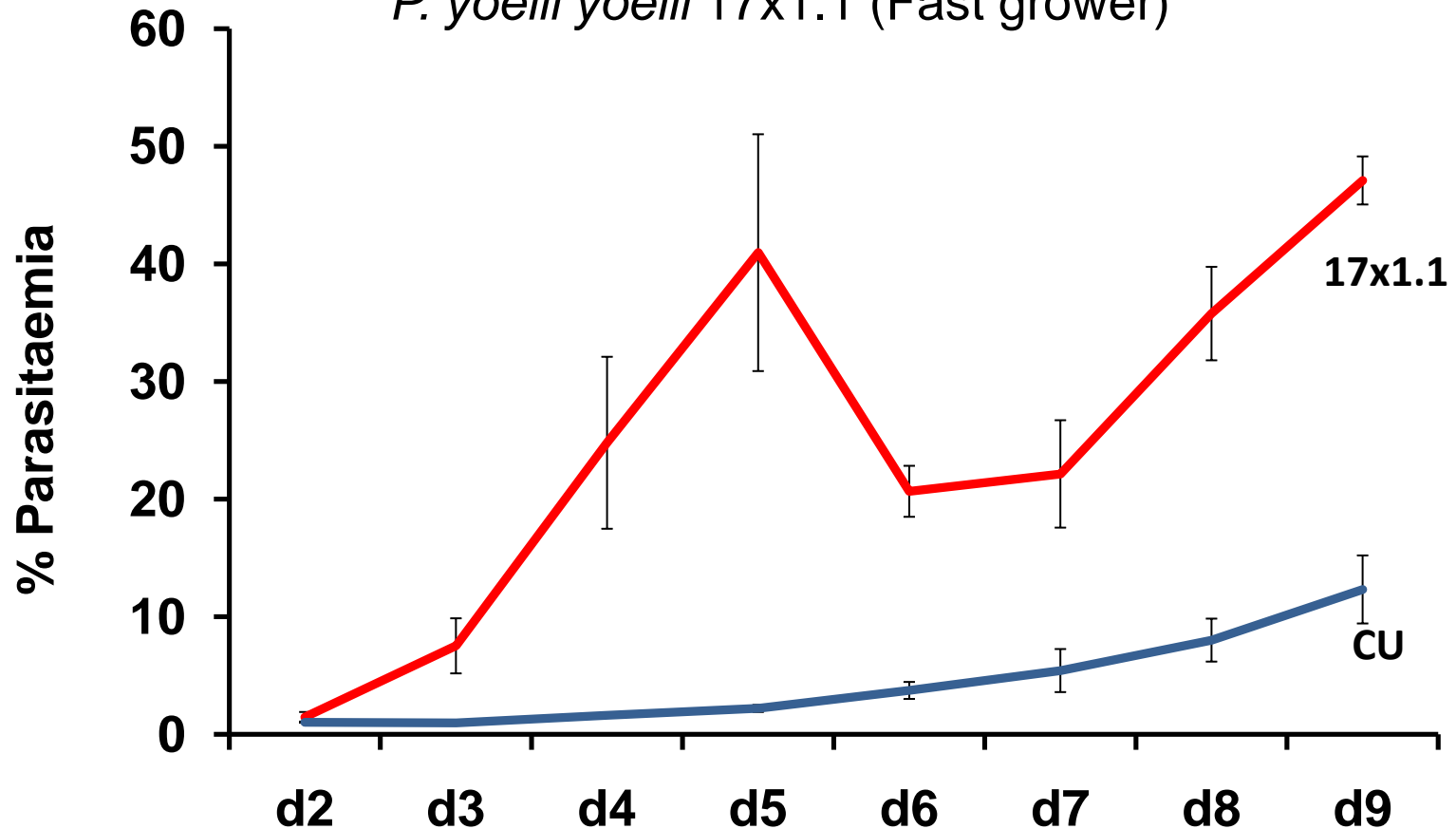


Comprises an increasing fraction of the population

# Malaria crossing experiment

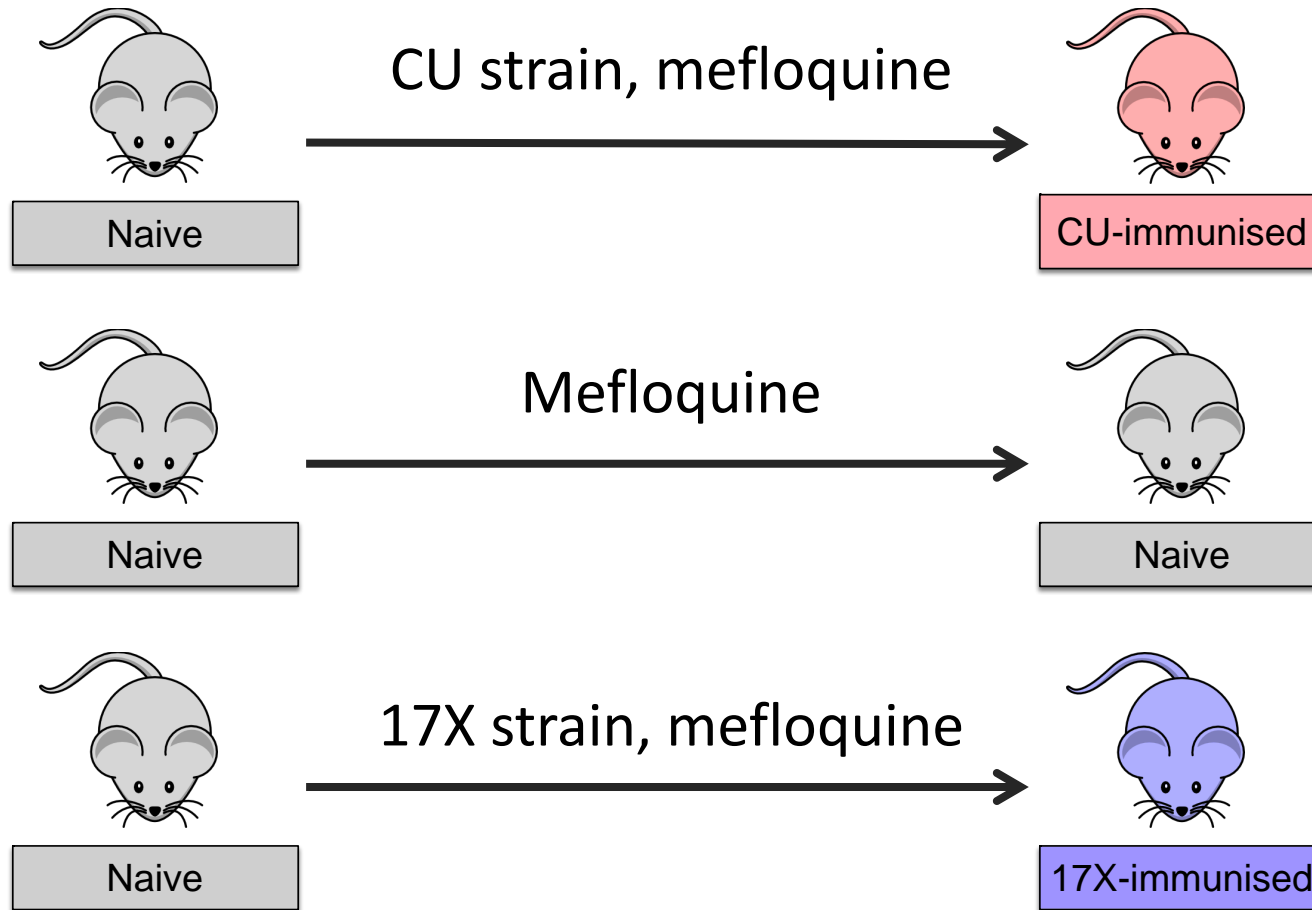
*P. yoelii yoelii* CU (Slow grower)

*P. yoelii yoelii* 17x1.1 (Fast grower)



# Malaria Crossing Experiment

## Creation of immunised strains:

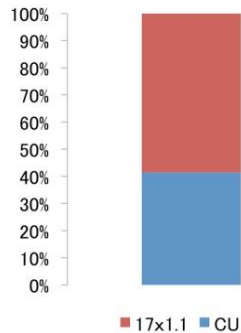




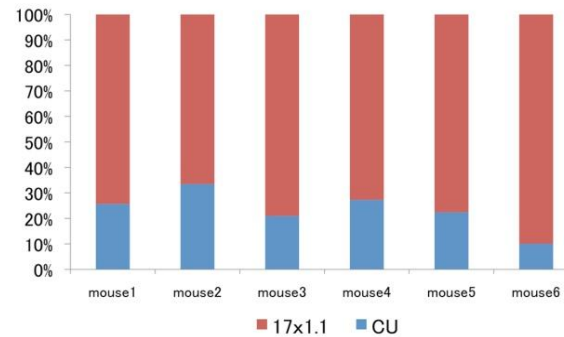
# Malaria Crossing Experiment

## Testing of immunised strains:

A) Initial Inoculum



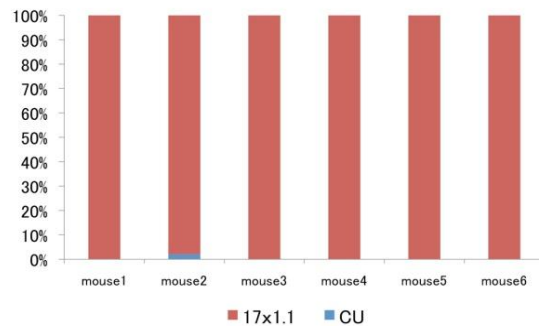
B) Mock-immunised



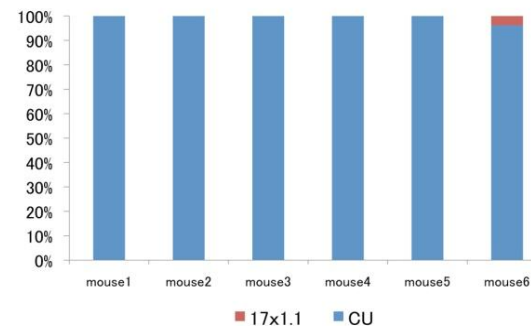
17X strain

CU strain

C) CU immunised

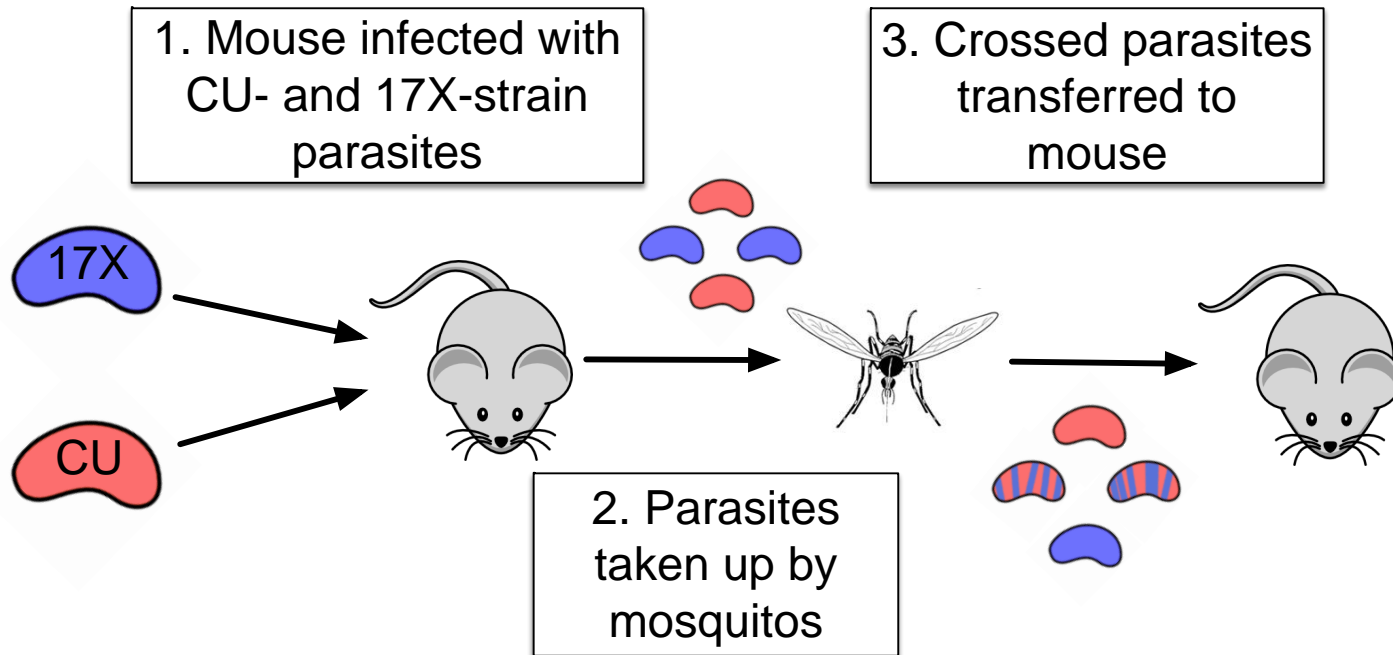


D) 17x1.1 immunised

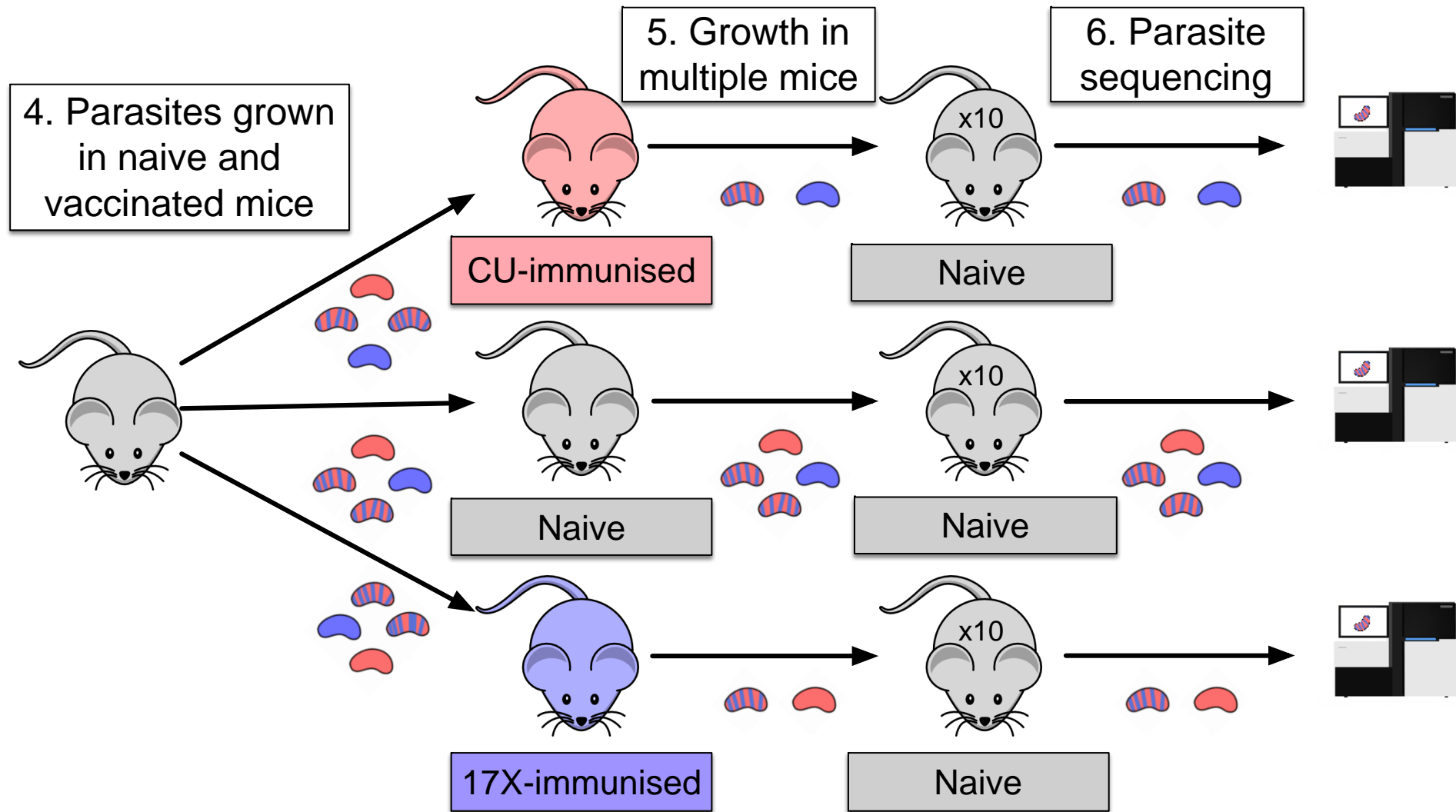


# Malaria Crossing Experiment

## Protocol:



# Malaria Crossing Experiment

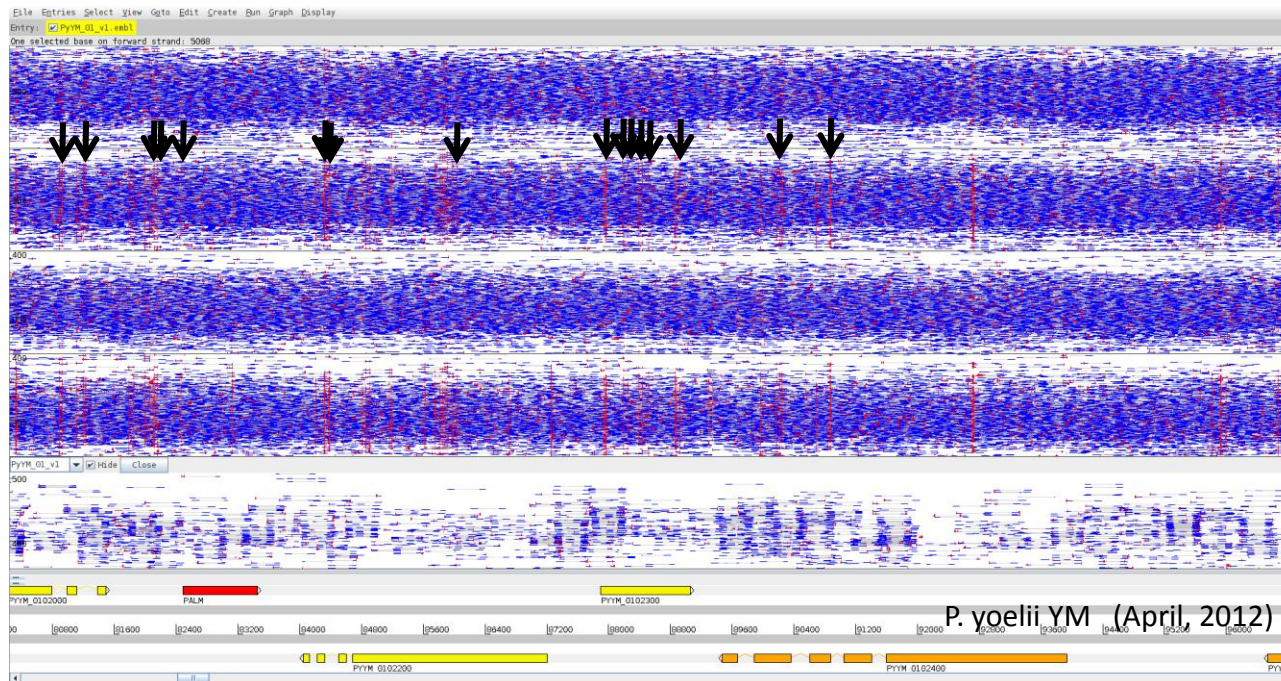


# Generation of allele frequencies

Identify SNPs from sequencing

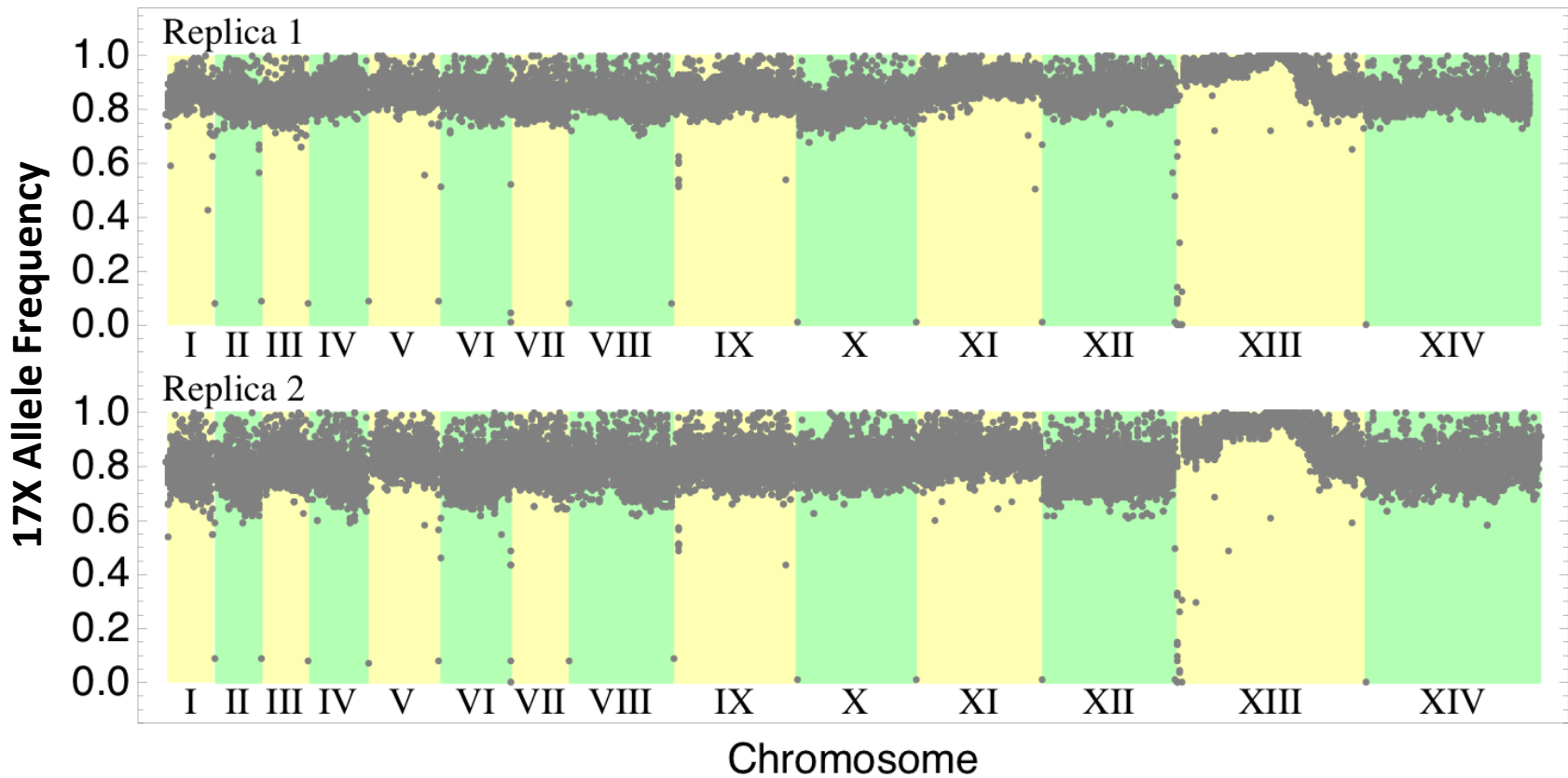
Map SNPs to newly-assembled genome

Measure proportion of parental strain at each SNP



# Generation of allele frequencies

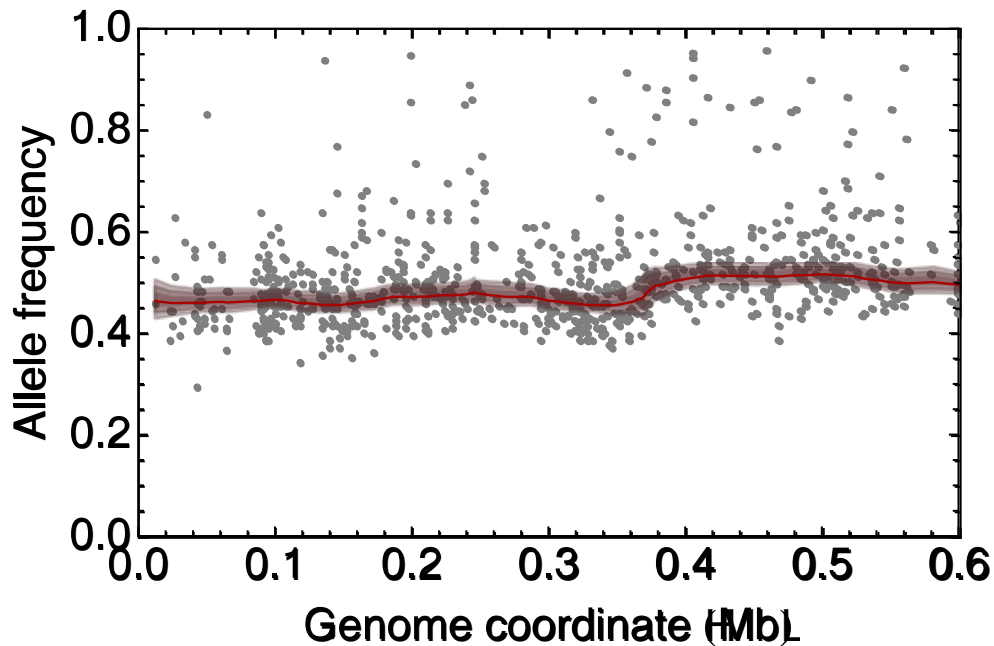
Around 24,000 SNPs genome-wide



# Further data processing

Sequencing error: Incorrect frequency

Mapping error: Incorrect location



Assume underlying frequency changes slowly

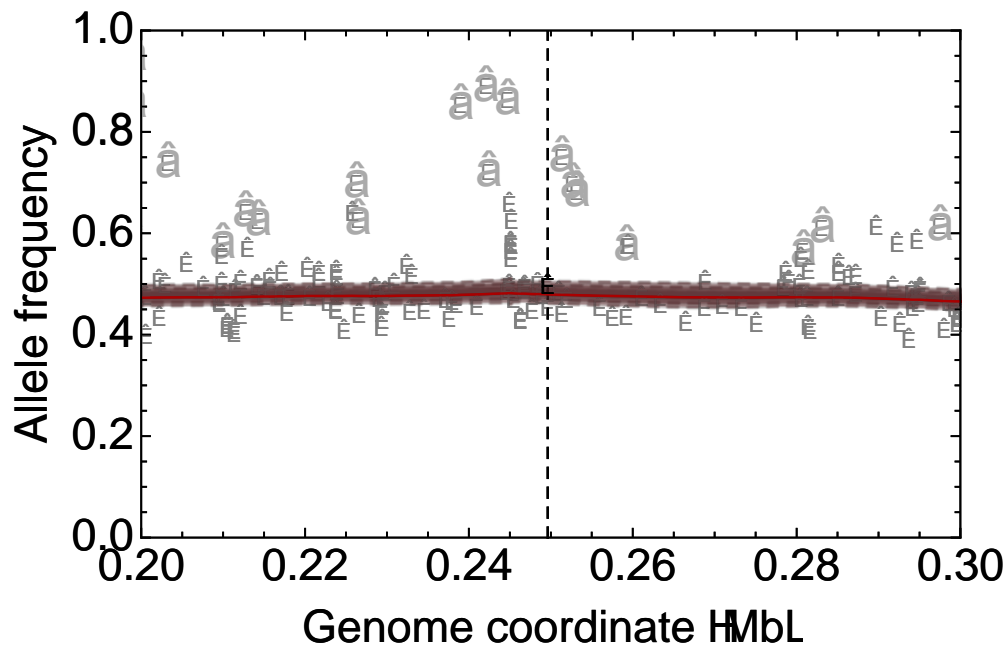
$$x_{i+1} = x_i + \mathcal{N}(0, s\sqrt{\Delta_{ij}})$$

Probability of error  $r$

Beta-binomial sampling model

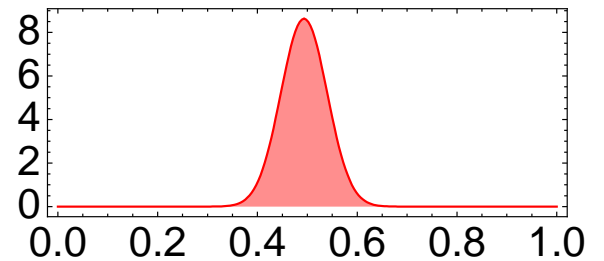
# Further data processing

Which points result from mapping error?



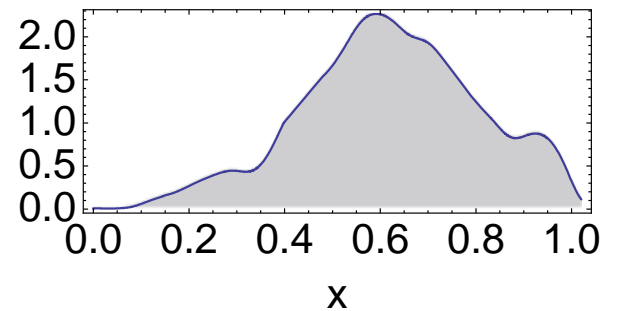
?

Local frequency distribution  $f_L$



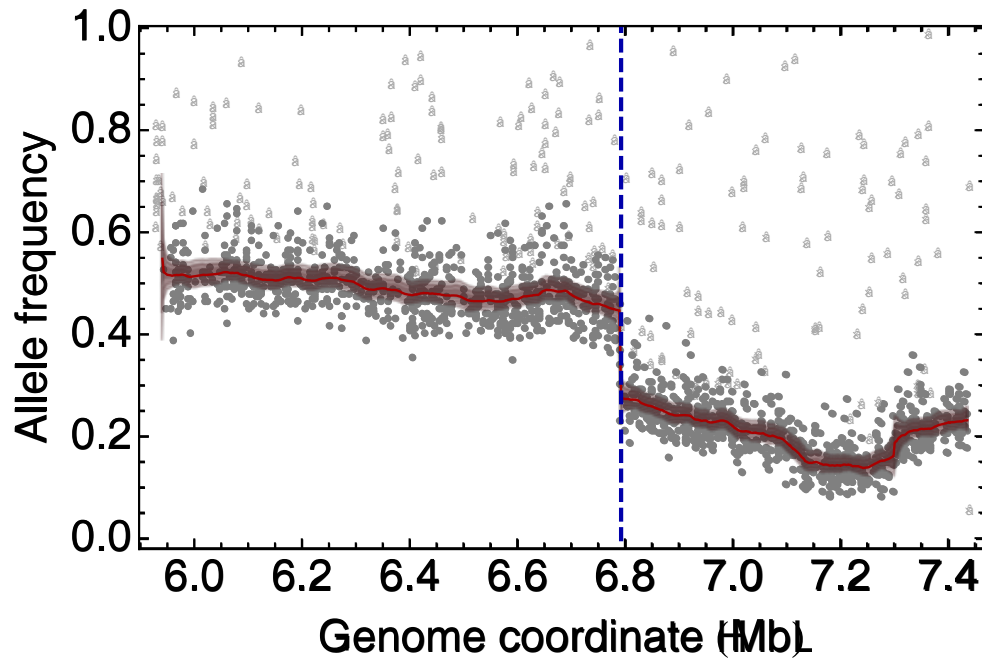
?

Global frequency distribution  $f_G$



# 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

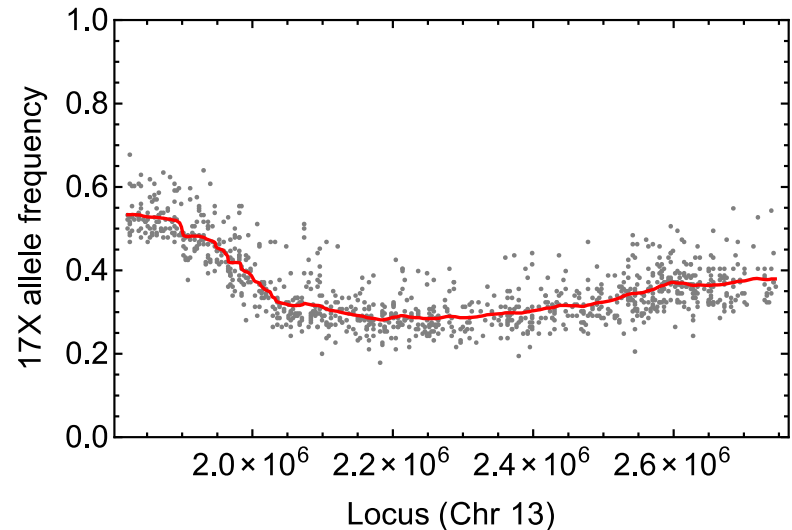
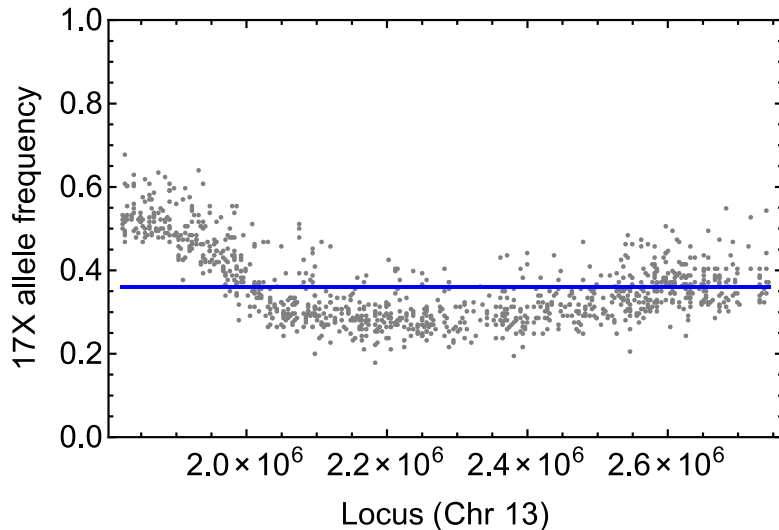


# Evolutionary model

Rule 1: Selection causes local deviations in frequencies

Measure likelihood difference per SNP between flat and diffusion models

At least 0.1 in both replicates



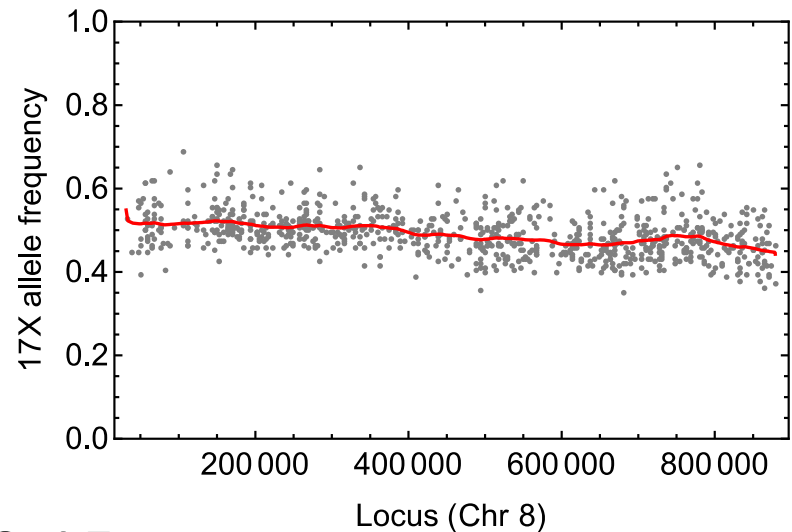
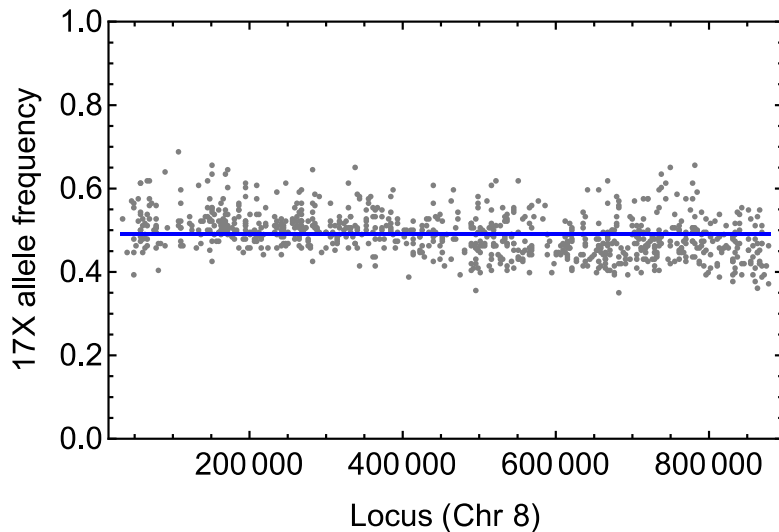
$\Delta L=8.0$

# Evolutionary model

Rule 1: Selection causes local deviations in frequencies

Measure likelihood difference  $D$  per SNP between flat and diffusion models

At least 0.1 in both replicates



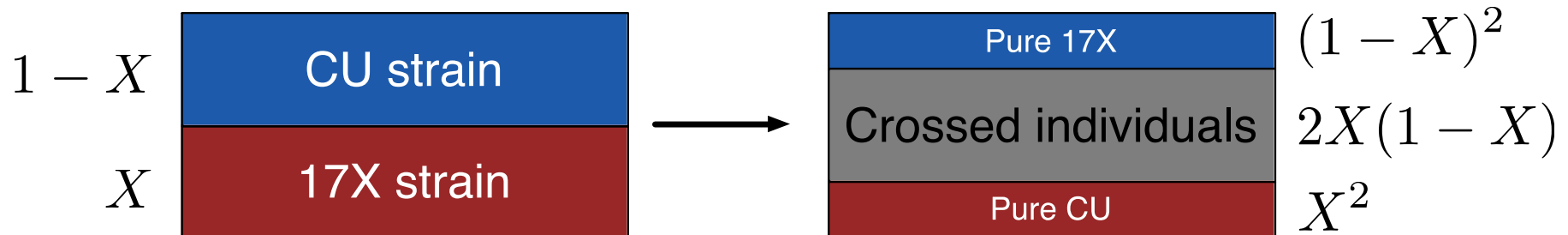
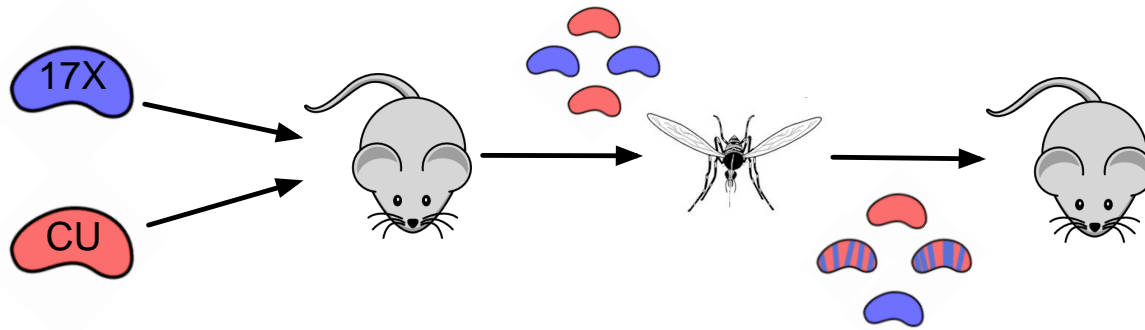
$\Delta L=0.15$

# Evolutionary model

Rule 2: Infer selection to occur in the same place in both replicates

# Location of selection

Model the process leading to the observed frequencies



# Location of selection

Model the process leading to the observed frequencies

$$x_i^1(t_o) = x + e$$

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

$X$  Frequency of 17X strain before cross

$i$  Locus of allele under selection

$x$  Frequency of allele under selection

$\rho$  Local rate of recombination

$e$  Effect of selection at distant loci

# Location of selection

Model the process leading to the observed frequencies

$$x_i^1(t_o) = x + e$$

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

$X$  Frequency of 17X strain before cross

$i$  **Locus of allele under selection**

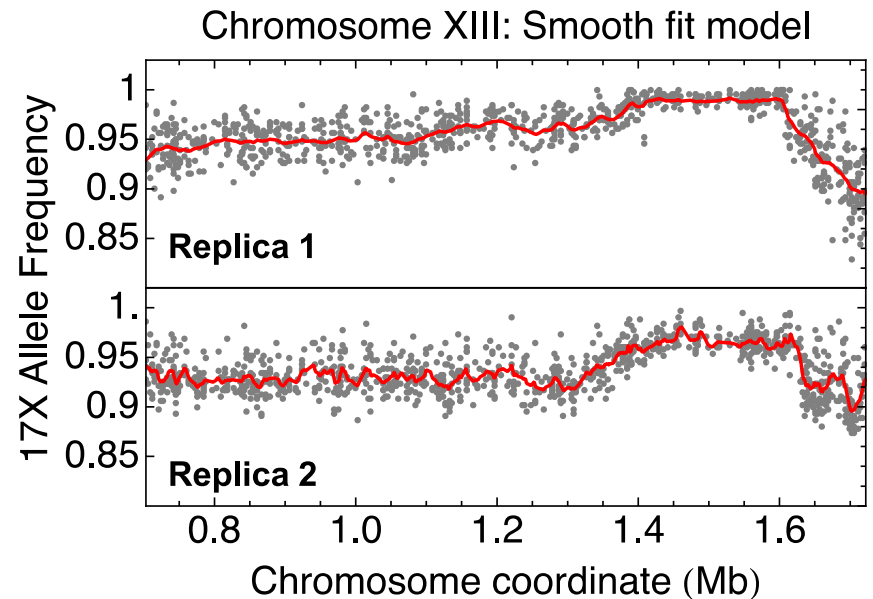
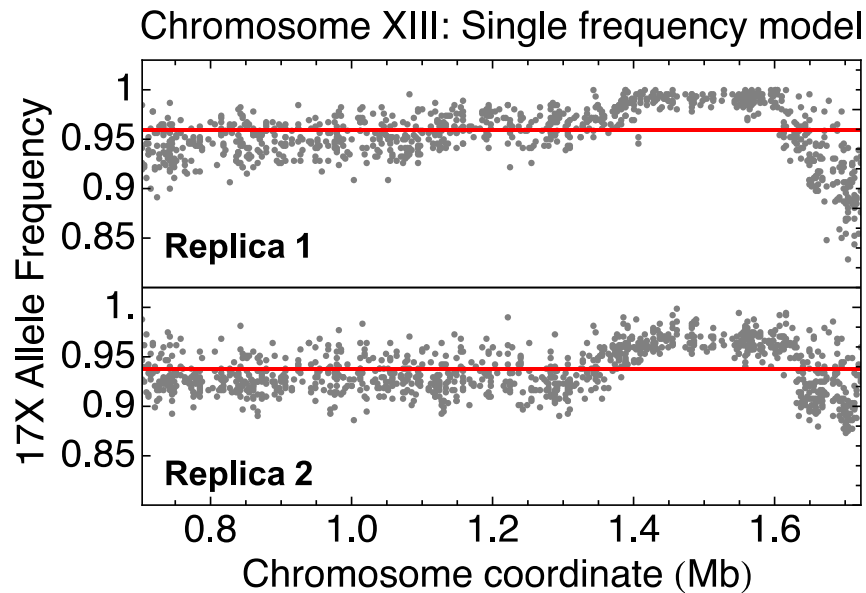
$x$  Frequency of allele under selection

$\rho$  Local rate of recombination

$e$  Effect of selection at distant loci

# Evolutionary model

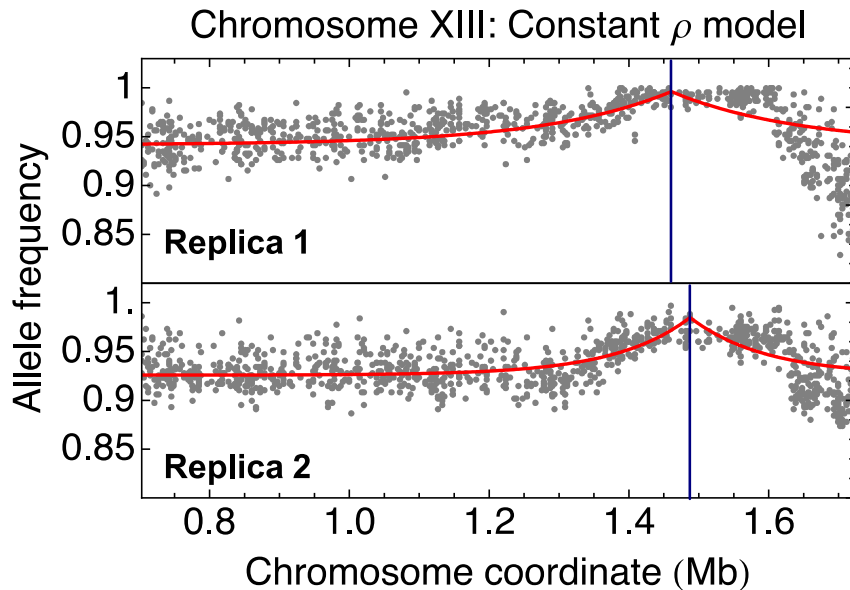
Example:



$\Delta L > 0.1$  in each case

# Evolutionary model

Example:



Location of driver:

R1: 1513 kb

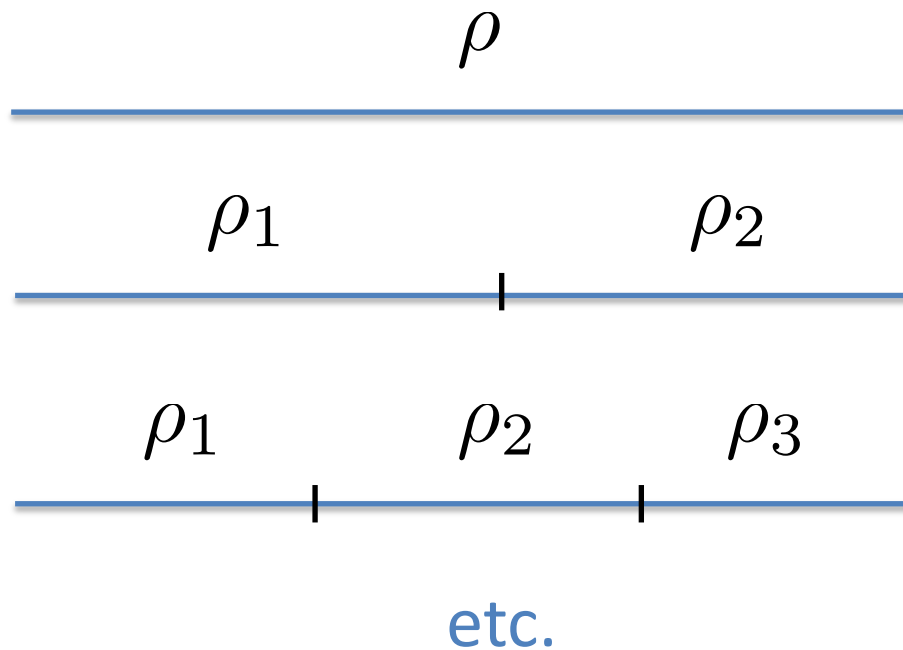
R2: 1529 kb

Identify candidate region for selection



# Overview of analysis

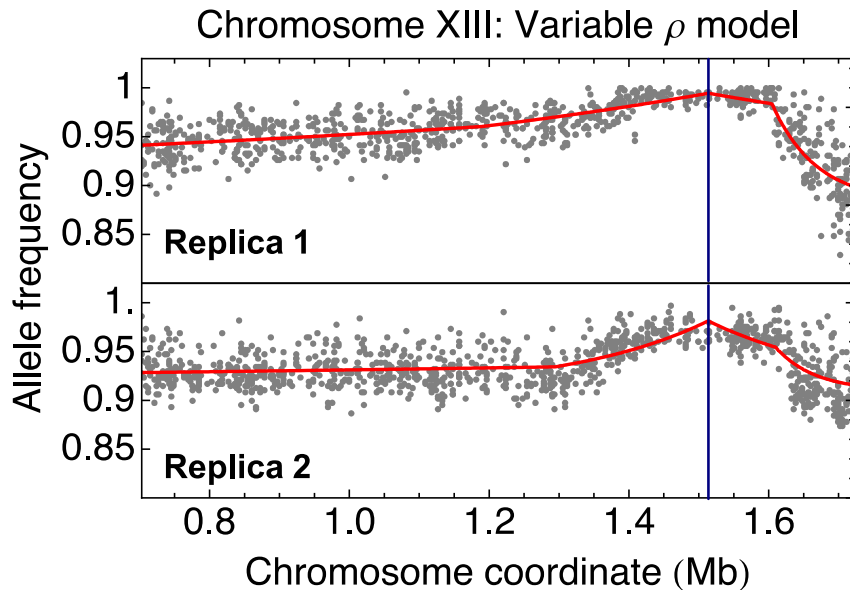
Refine model: Variable recombination



Model selection: Bayesian Information Criterion

# Evolutionary model

Refine model: Variable recombination



Location of driver:

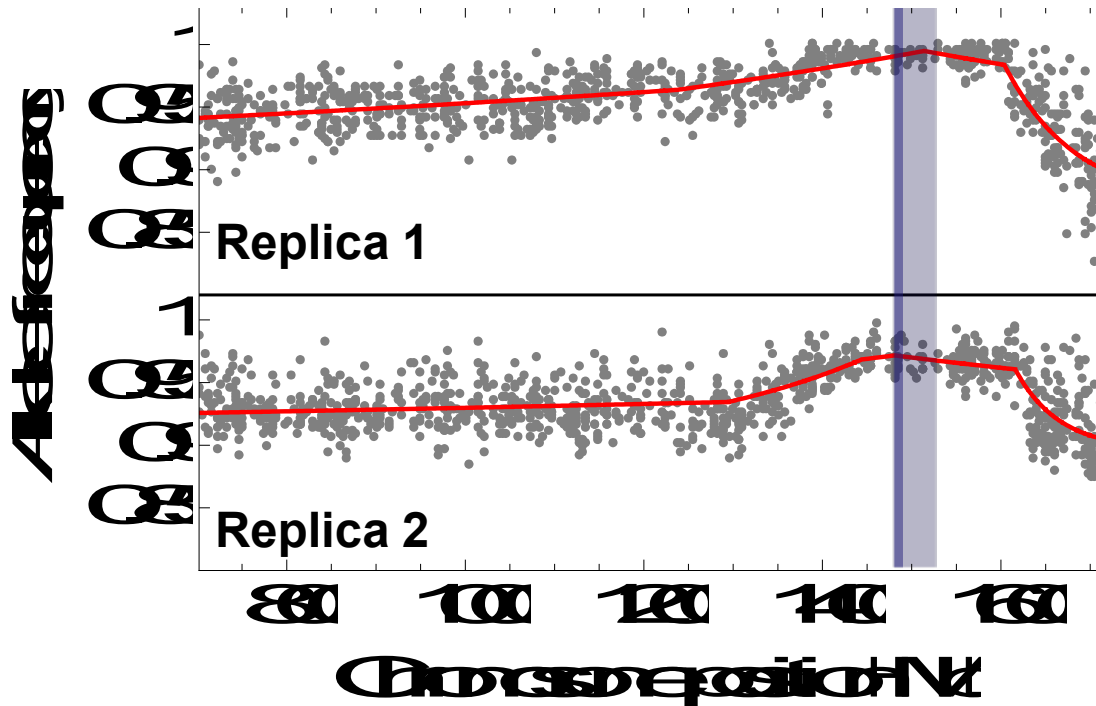
R1: 1514 kb

R2: 1514 kb

Two- and three-recombination rate models

# Evolutionary model

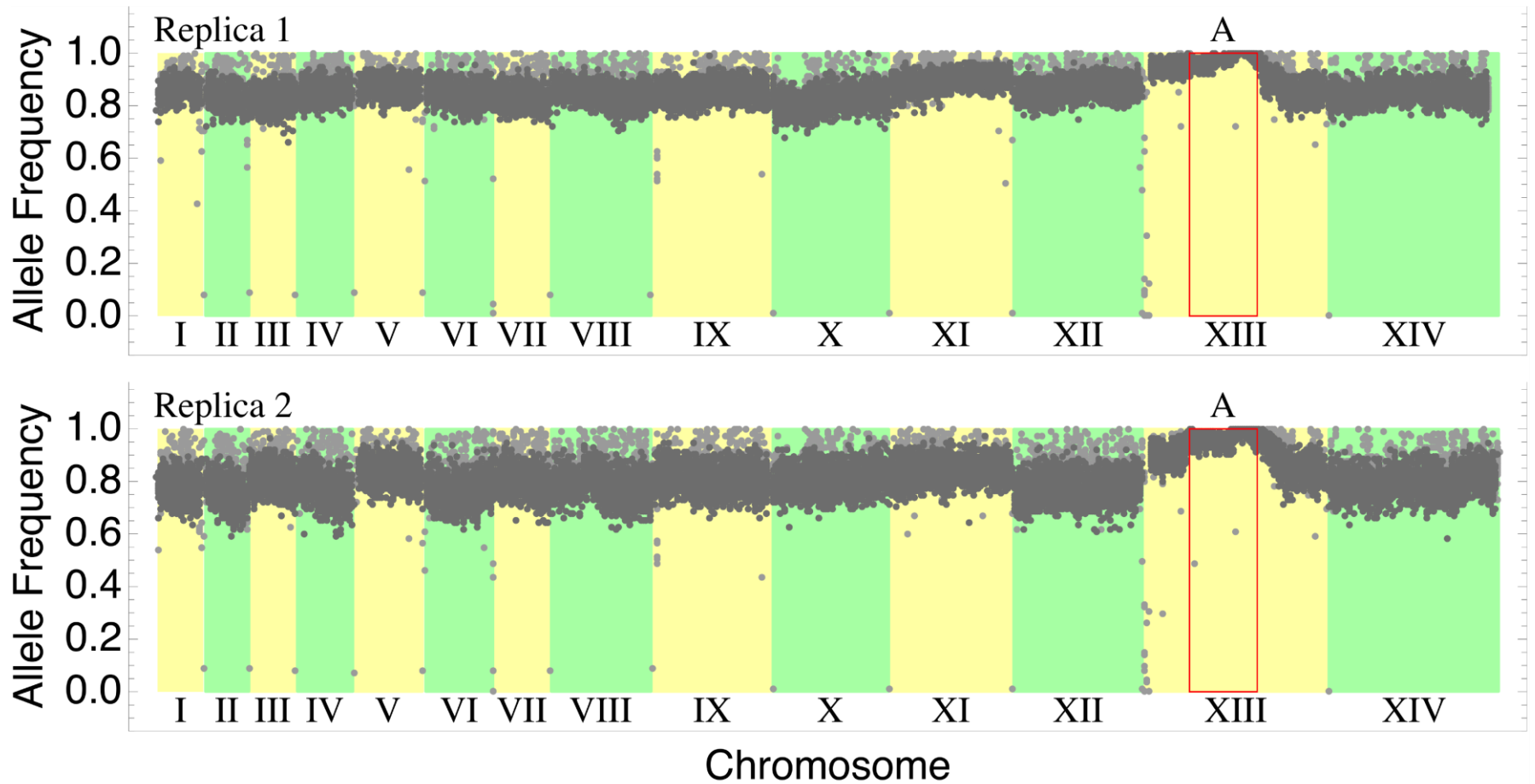
Identify confidence intervals by likelihood



More and less conservative intervals

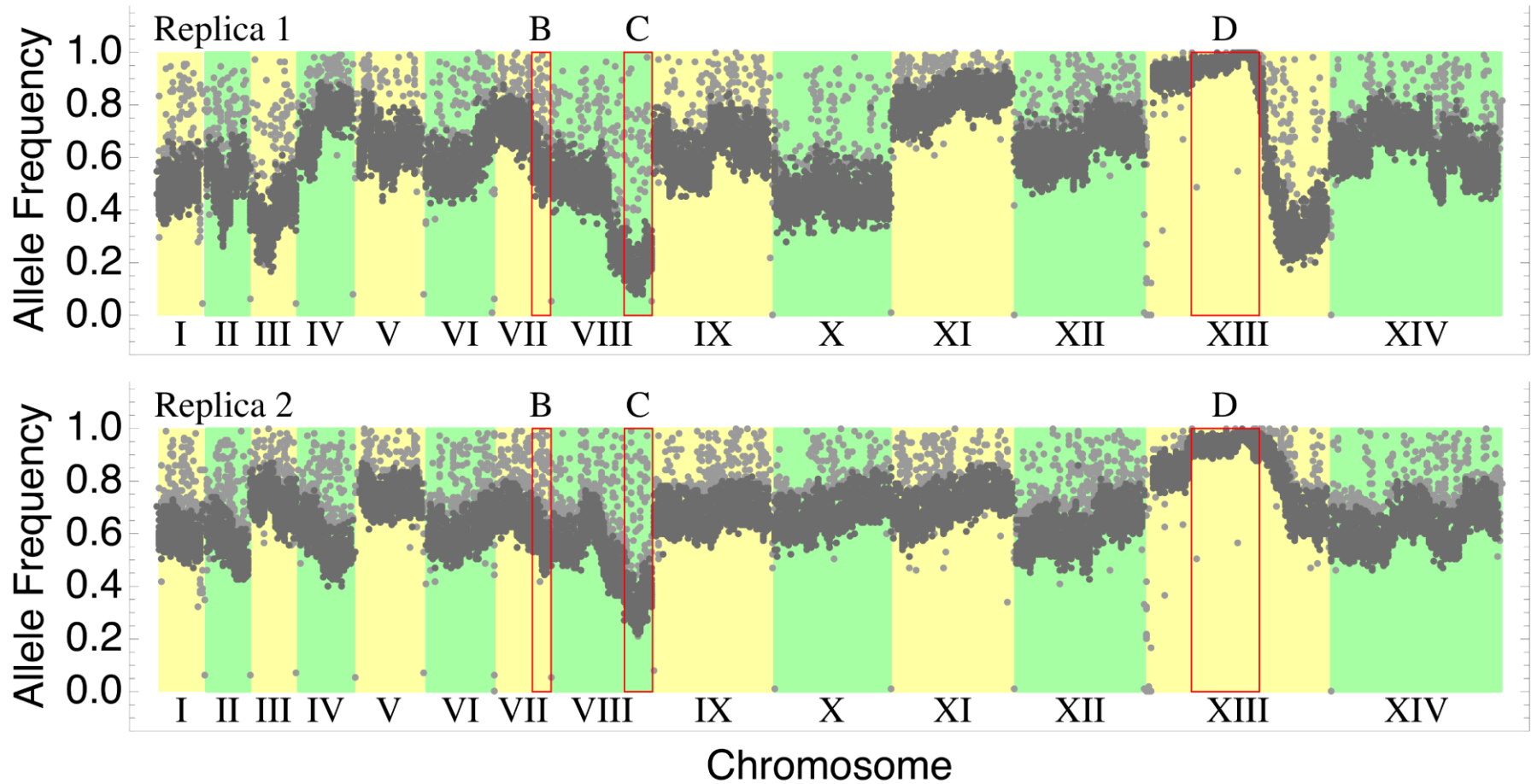
# Model outcome

## Naïve mice



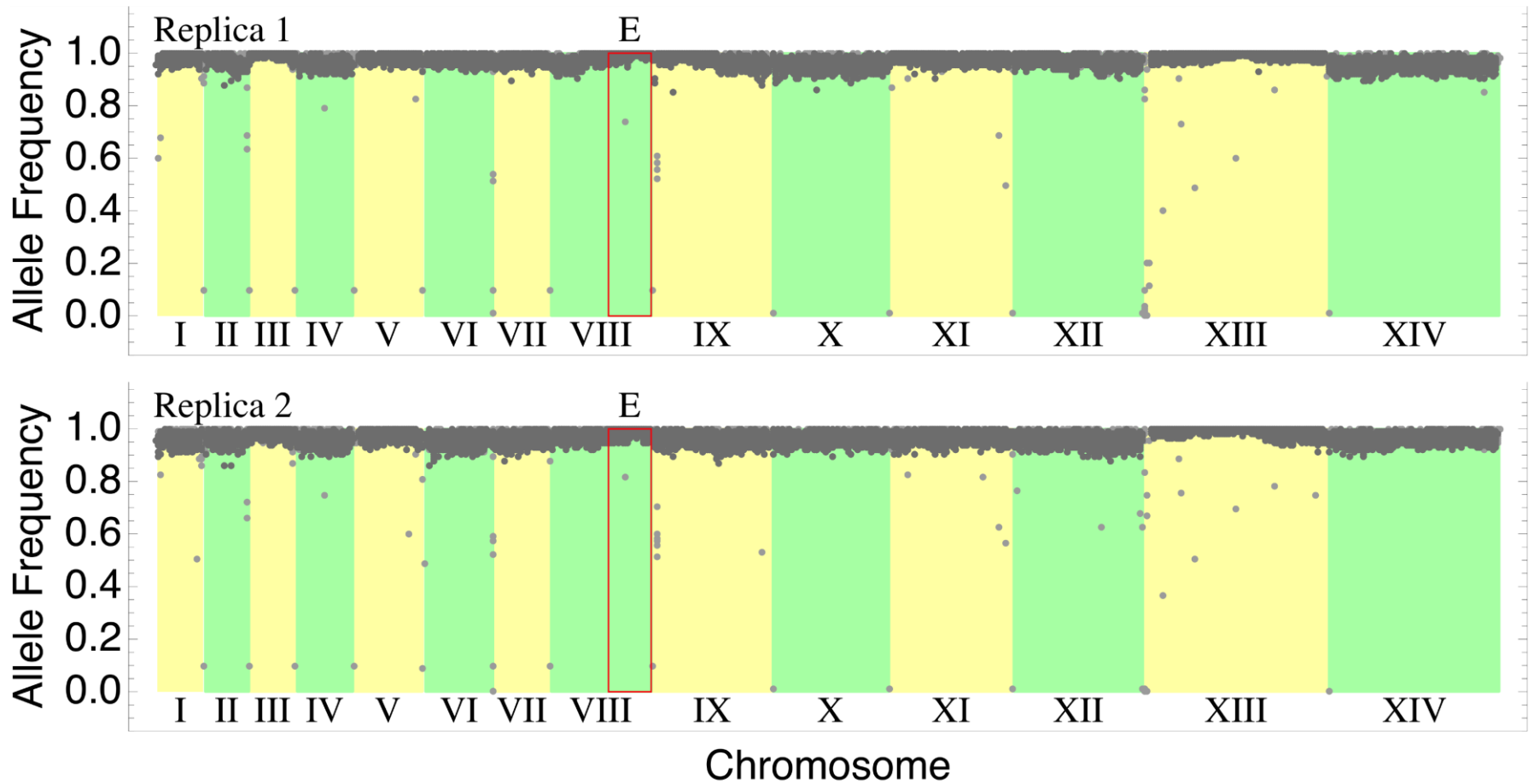
# Model outcome

17X-immunised mice



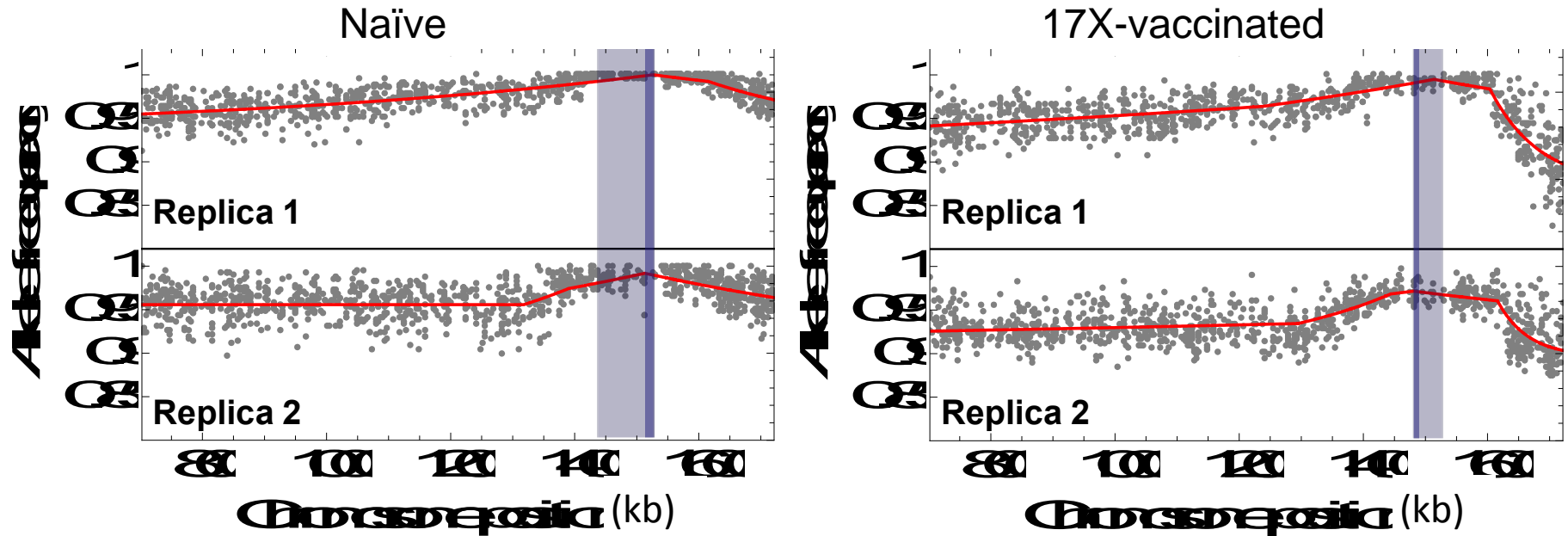
# Model outcome

CU-immunised mice



# Model outcome

## Chromosome XIII



Growth allele: Position close to gene PyEBL

Erythrocyte Binding Ligand

# Experimental validation

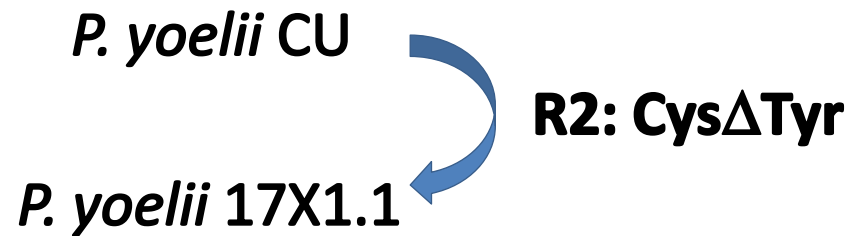
## Single SNP in region 2 of PyEBL gene

*P. yoelii* CU:

```
... aaa aat gcg tgt aat aca tat ...  
... K N A C N T Y ...
```

*P. yoelii* 17X1.1:

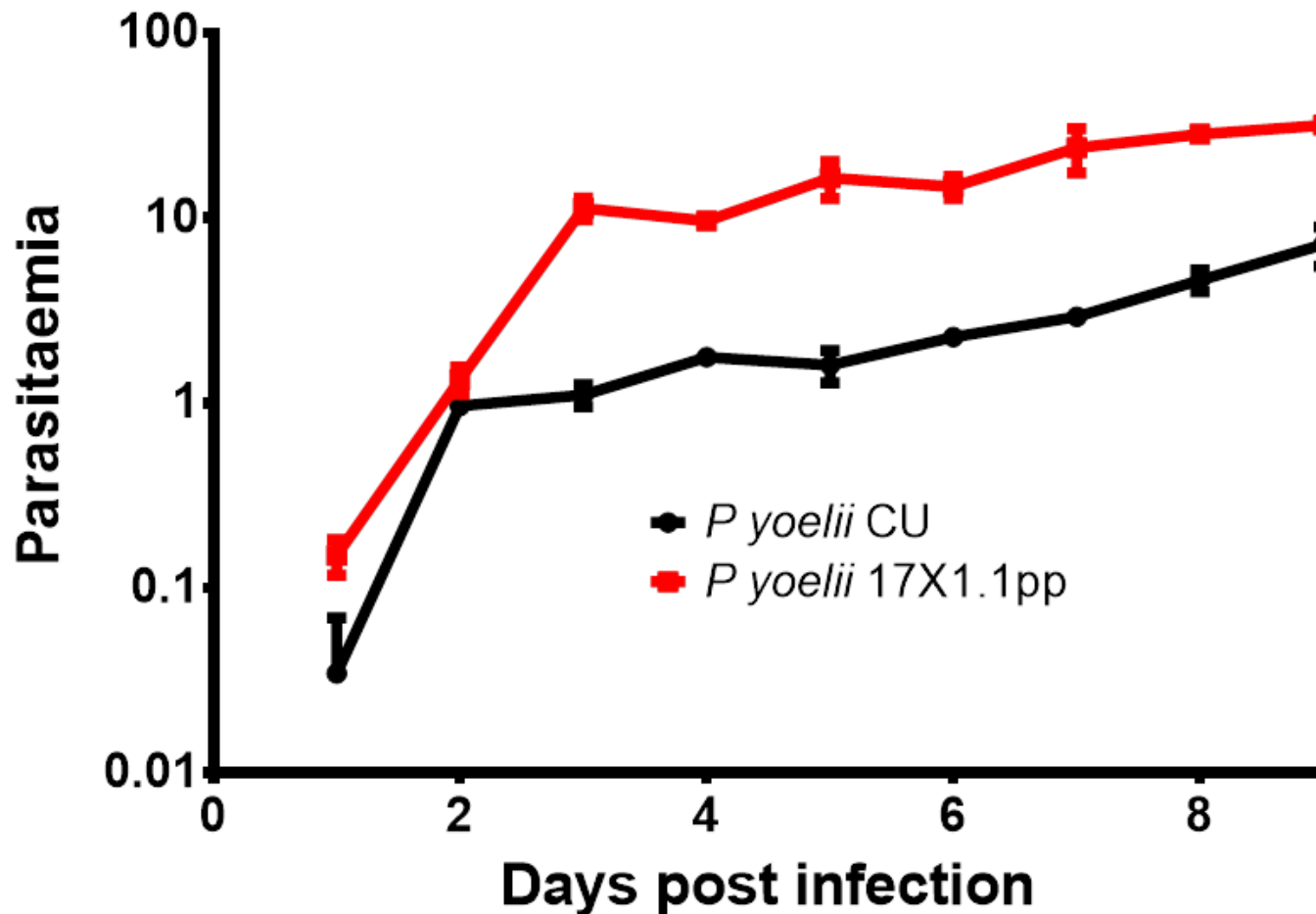
```
... aaa aat gcg tat aat aca tat ...  
... K N A Y N T Y ...
```





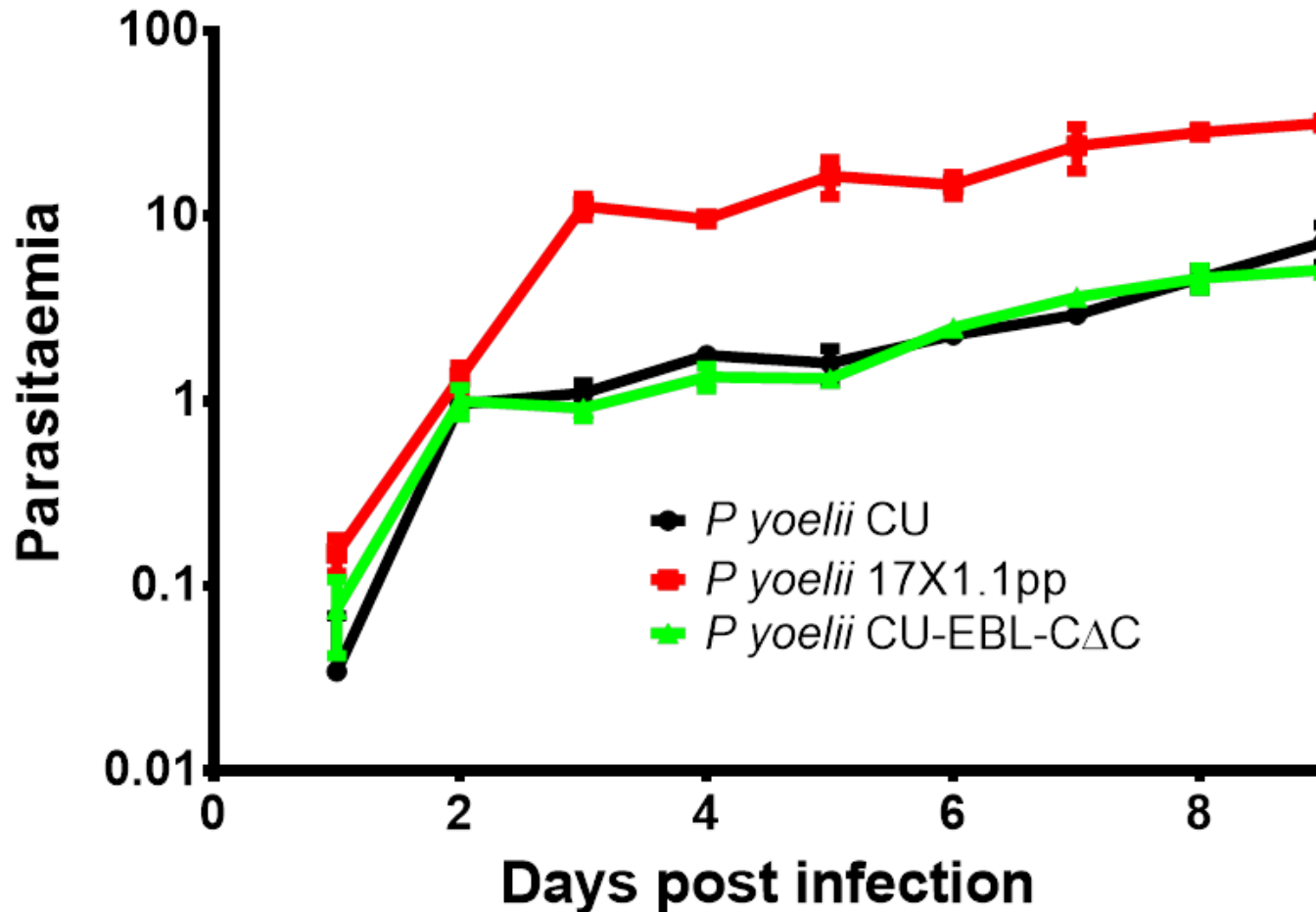
# Experimental validation

Reverse transfection: Fast gene into slow grower



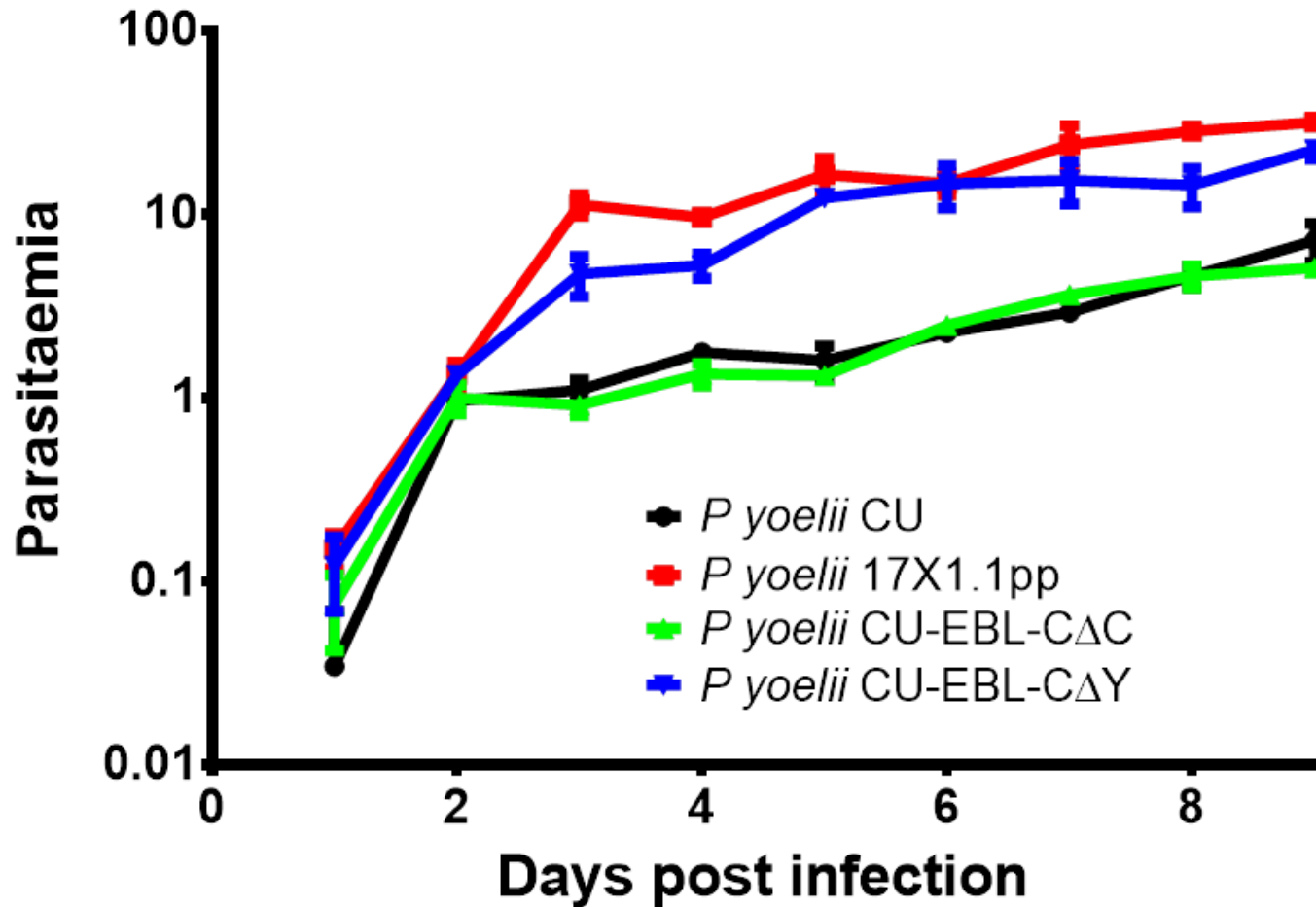
# Experimental validation

Reverse transfection: Fast gene into slow grower



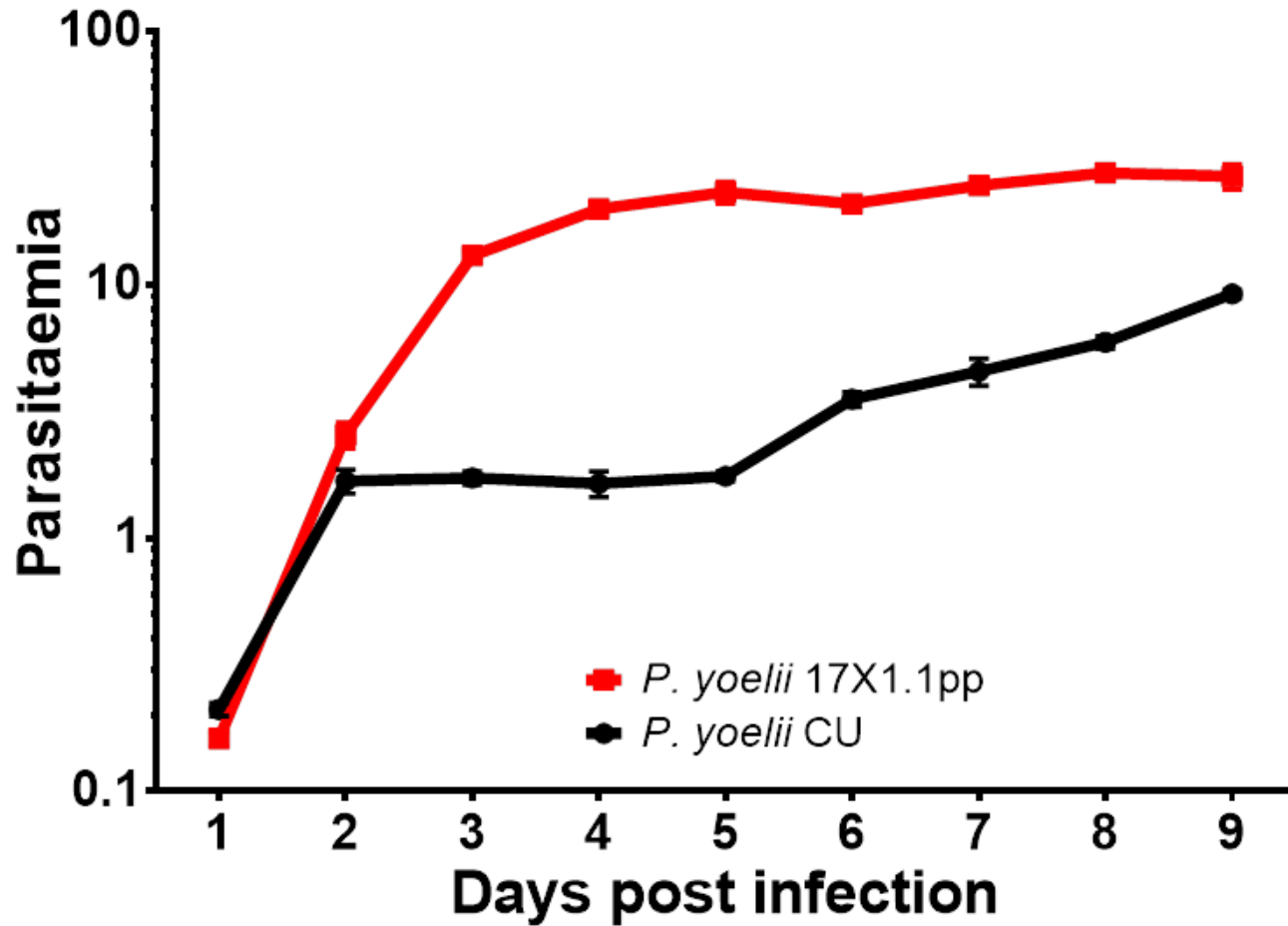
# Experimental validation

Reverse transfection: Fast gene into slow grower



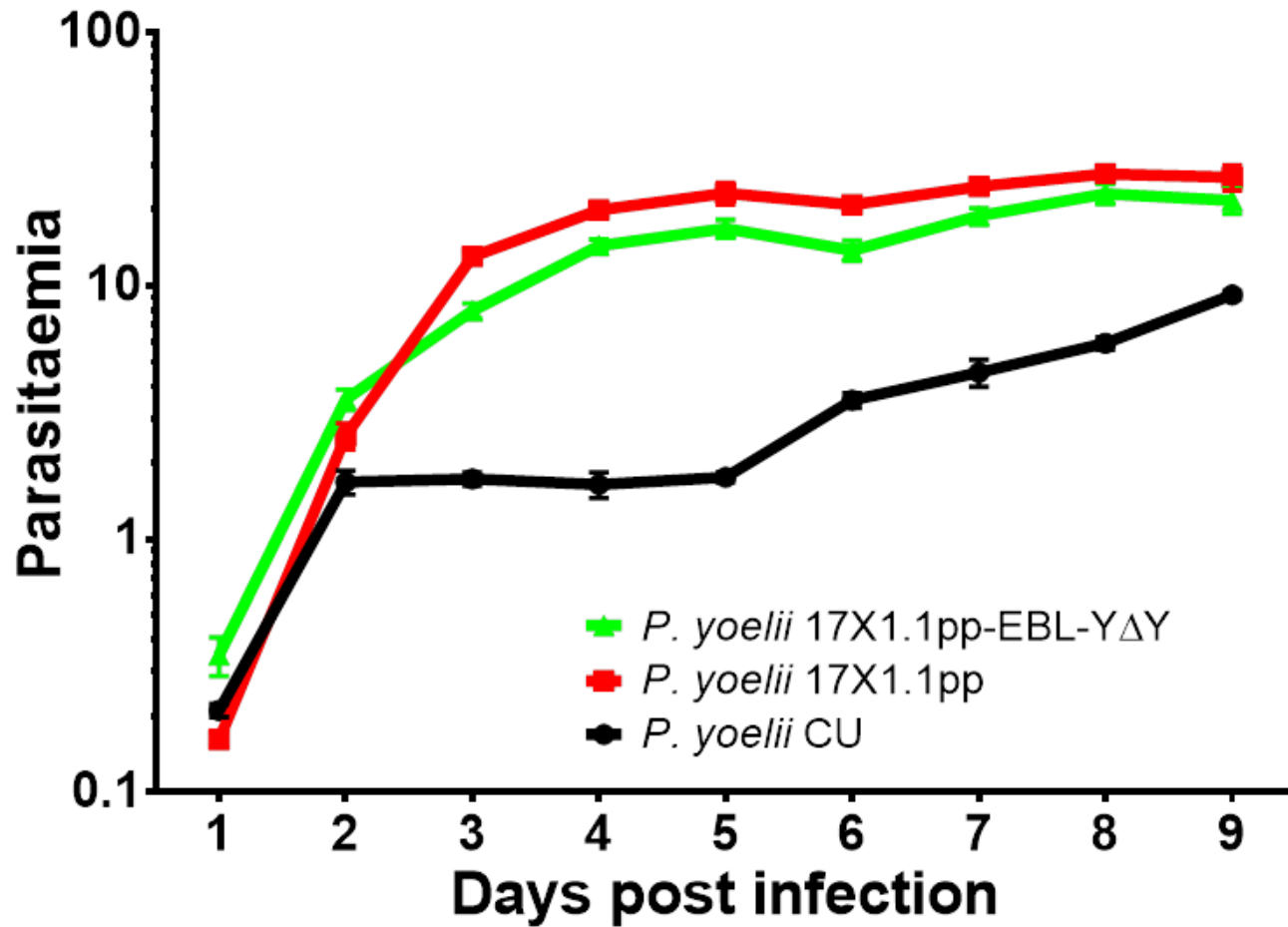
# Experimental validation

Reverse transfection: Slow gene into fast grower



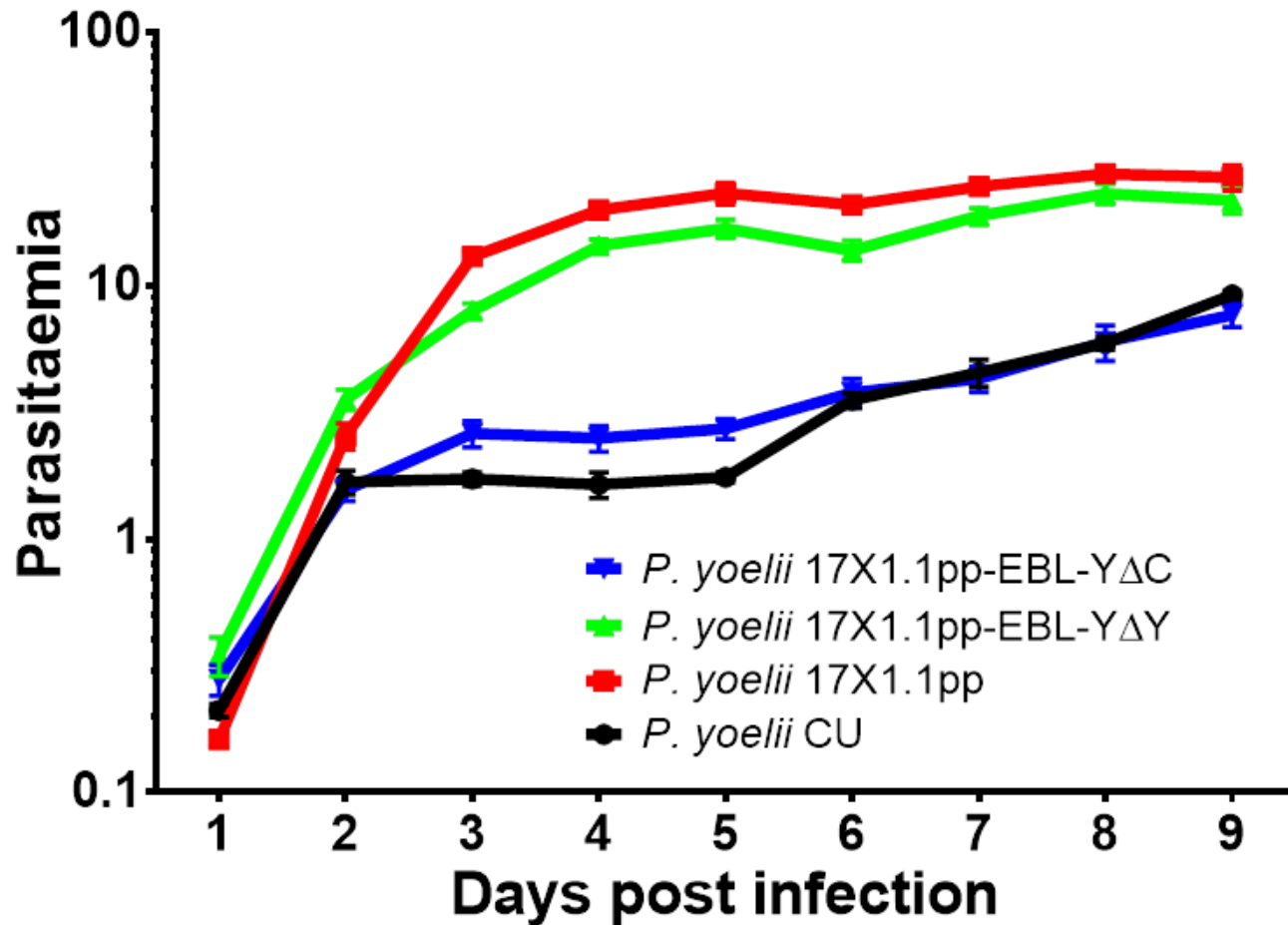
# Experimental validation

Reverse transfection: Slow gene into fast grower



# Experimental validation

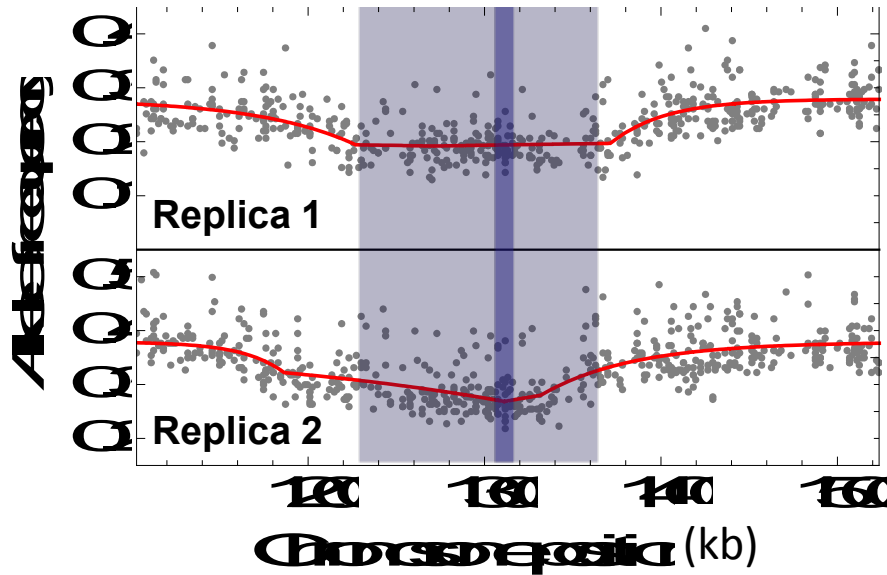
Reverse transfection: Slow gene into fast grower



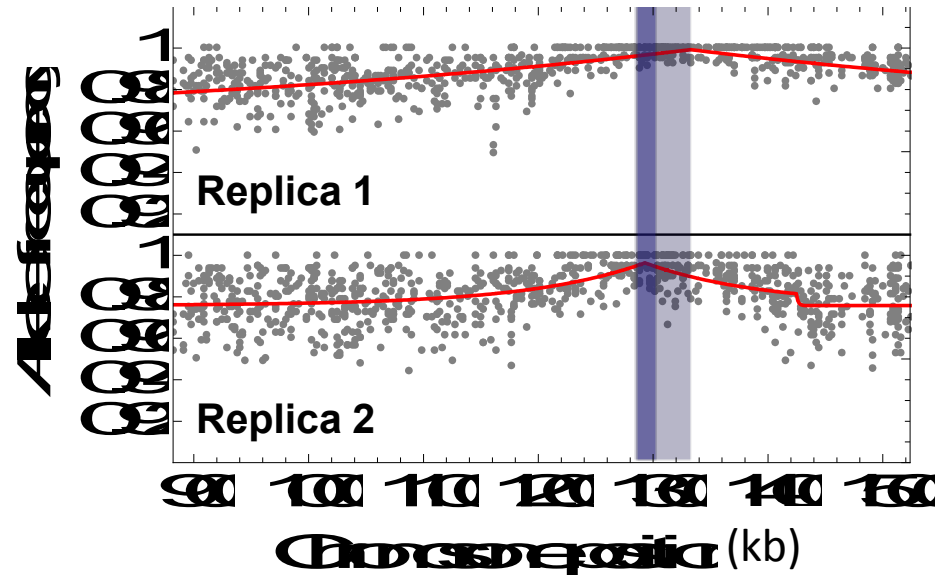
# Model outcome

## Chromosome VIII

17X-immunised



CU-immunised

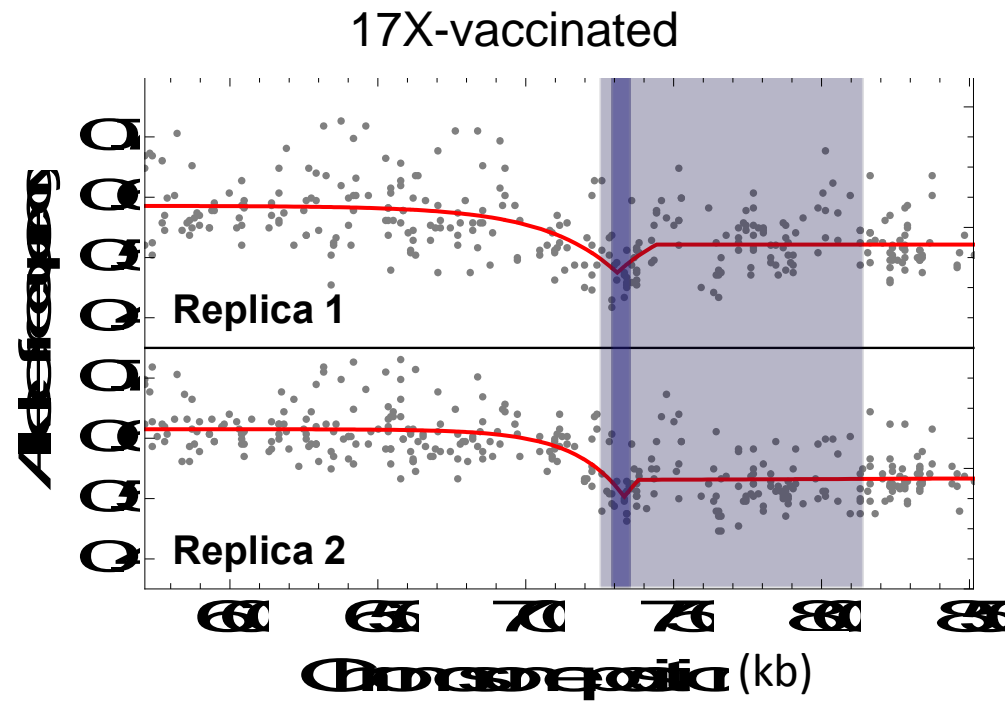


Immune-selection allele: Position close to gene MSP1

Work in progress...

# Model outcome

## Chromosome VII



What is this?



# Conclusions

Analysed data from a malaria crossing experiment

Identified SNP responsible for increased virulence

Identified SNP underlying strain-specific immunity

Ongoing aim: Identify genes responsible for clinically relevant phenotypic differences, in order to combat malarial disease