

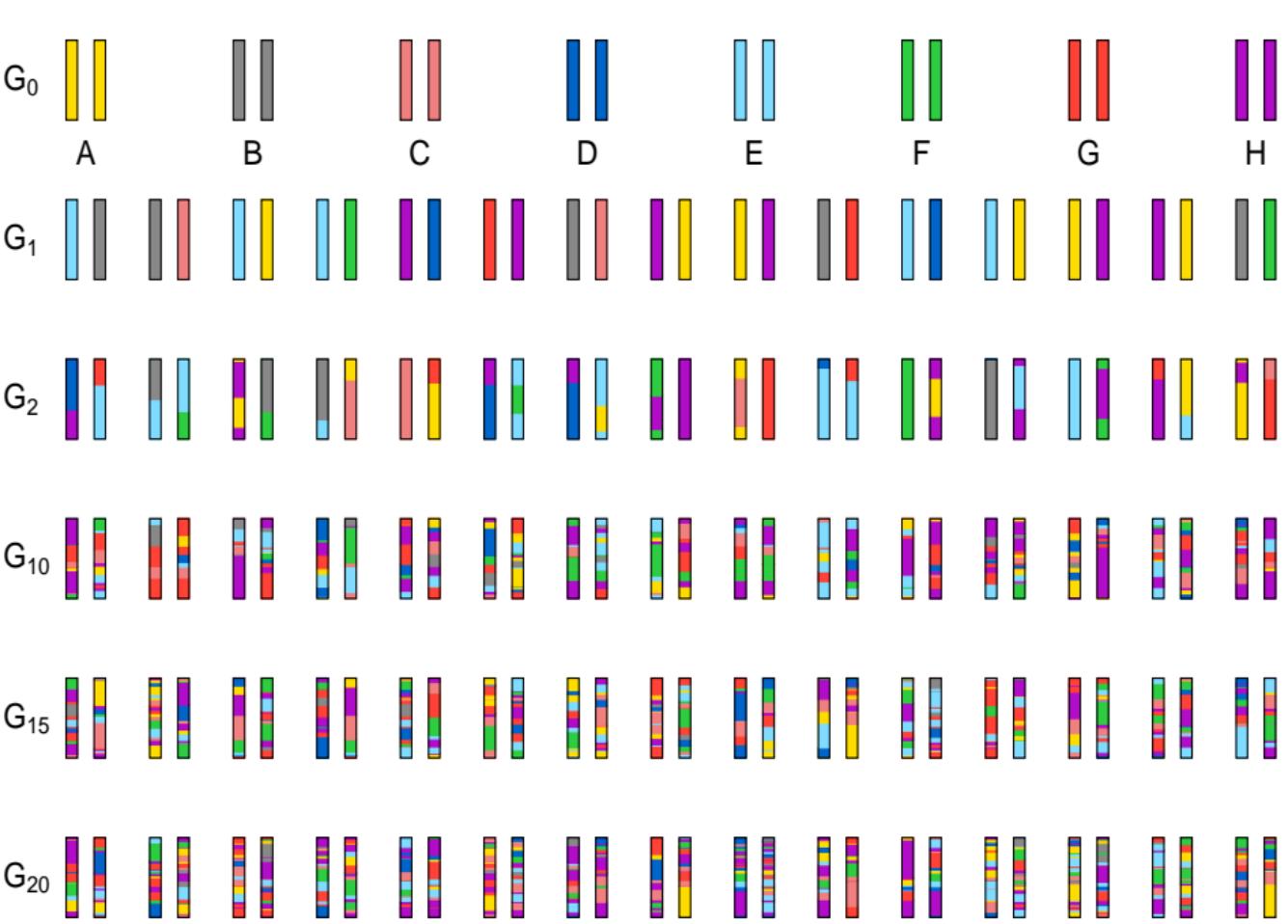
Genotype Reconstruction for Diversity Outbred Mice

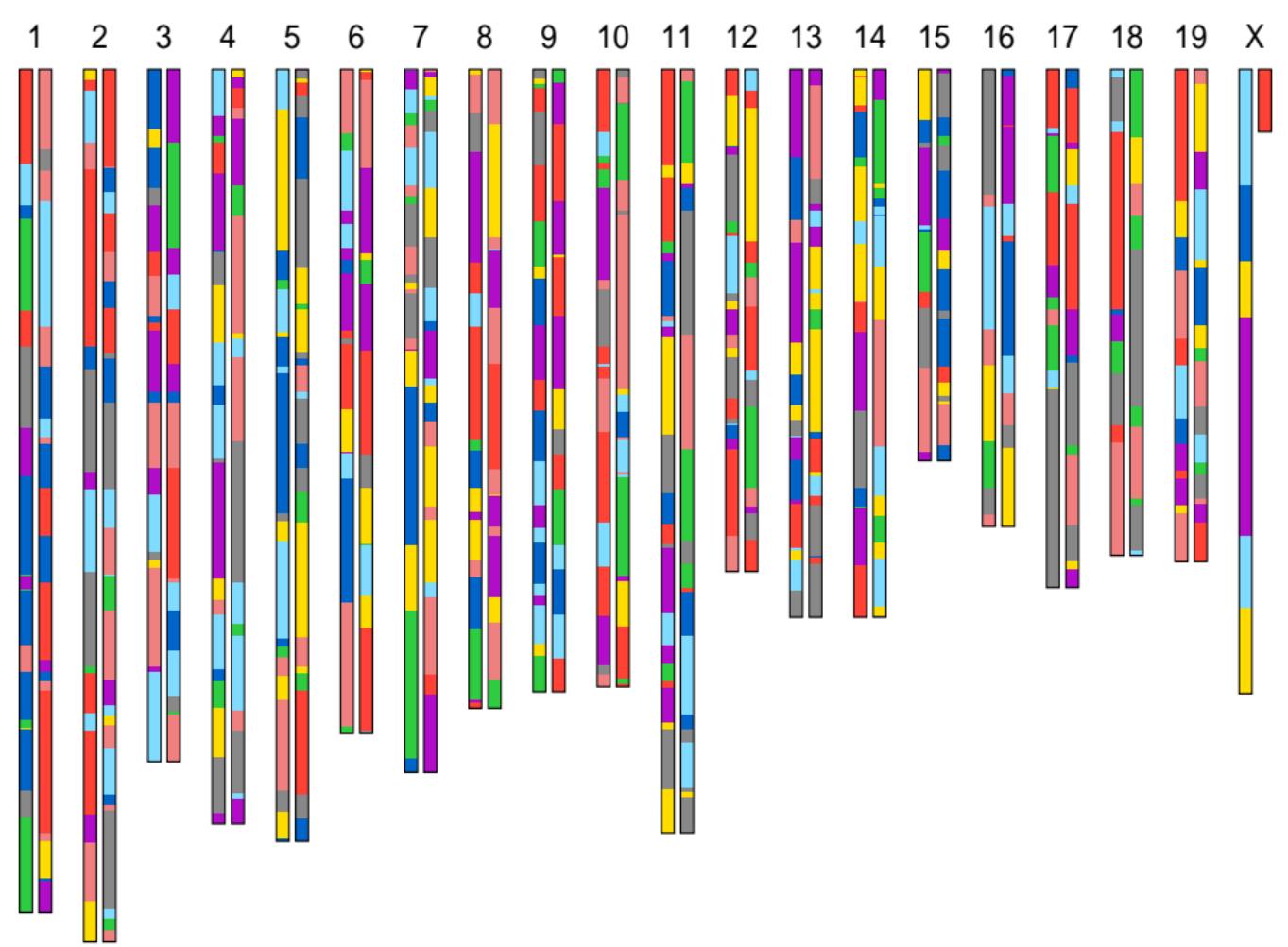
A comparison of r/qt12 and DOQTL

John Spaw¹
Karl Broman²

[github.com/JohnPSpaw¹](https://github.com/JohnPSpaw)
[github.com/kbroman²](https://github.com/kbroman)







Hidden Markov Model

Hidden Markov Model

Probabilities are generated from:

Hidden Markov Model

Probabilities are generated from:

1. Transition Model

Hidden Markov Model

Probabilities are generated from:

1. Transition Model
2. **Emission Model**

Hidden Markov Model

Probabilities are generated from:

1. Transition Model
2. **Emission Model**

Conditional probability of observed data given underlying diplotype state

Founders

A ●
B ○
C ●●
D ●●●
E ●●
F ○
G ○
H ○

○ ● ○
● ○ ○
○ ○ ○
○ ○ ○
○ ○ ○
○ ○ ○
○ ○ ○
○ ○ ○

○ ○ ○ ○
○ ○ ○ ○
○ ○ ○ ○
○ ○ ○ ○
○ ○ ○ ○
○ ○ ○ ○
○ ○ ○ ○
○ ○ ○ ○

○ ○ ○
○ ○ ○
○ ○ ○
○ ○ ○
○ ○ ○
● ● ○
○ ○ ○
○ ○ ○

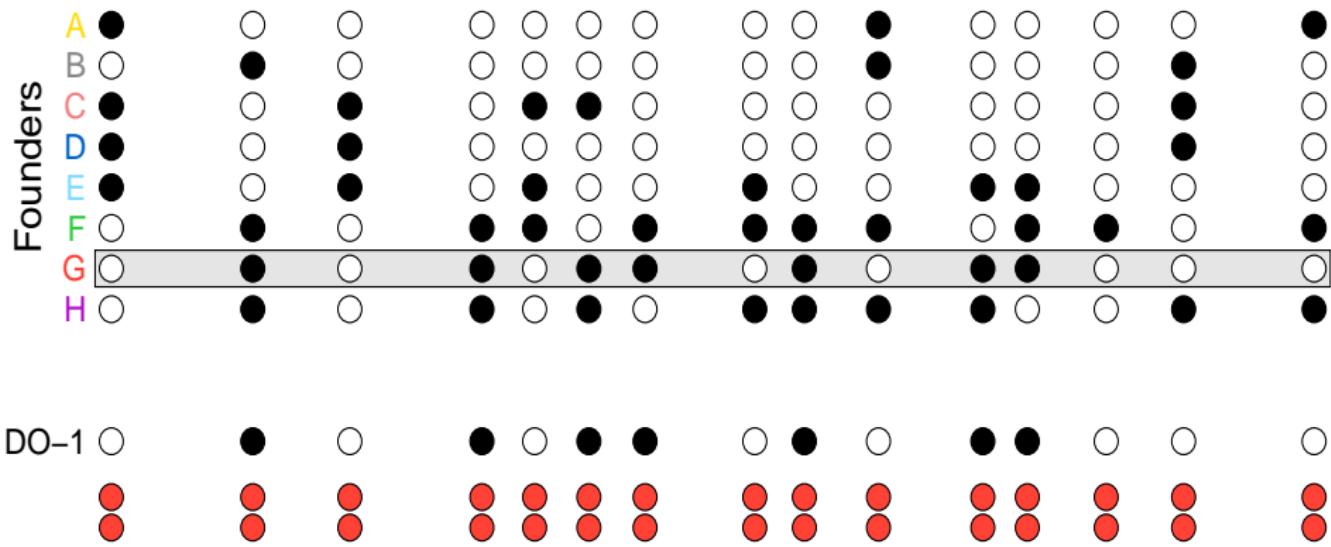
○ ○ ○
○ ○ ○
○ ○ ○
○ ○ ○
○ ○ ○
● ● ○
○ ○ ○
○ ○ ○

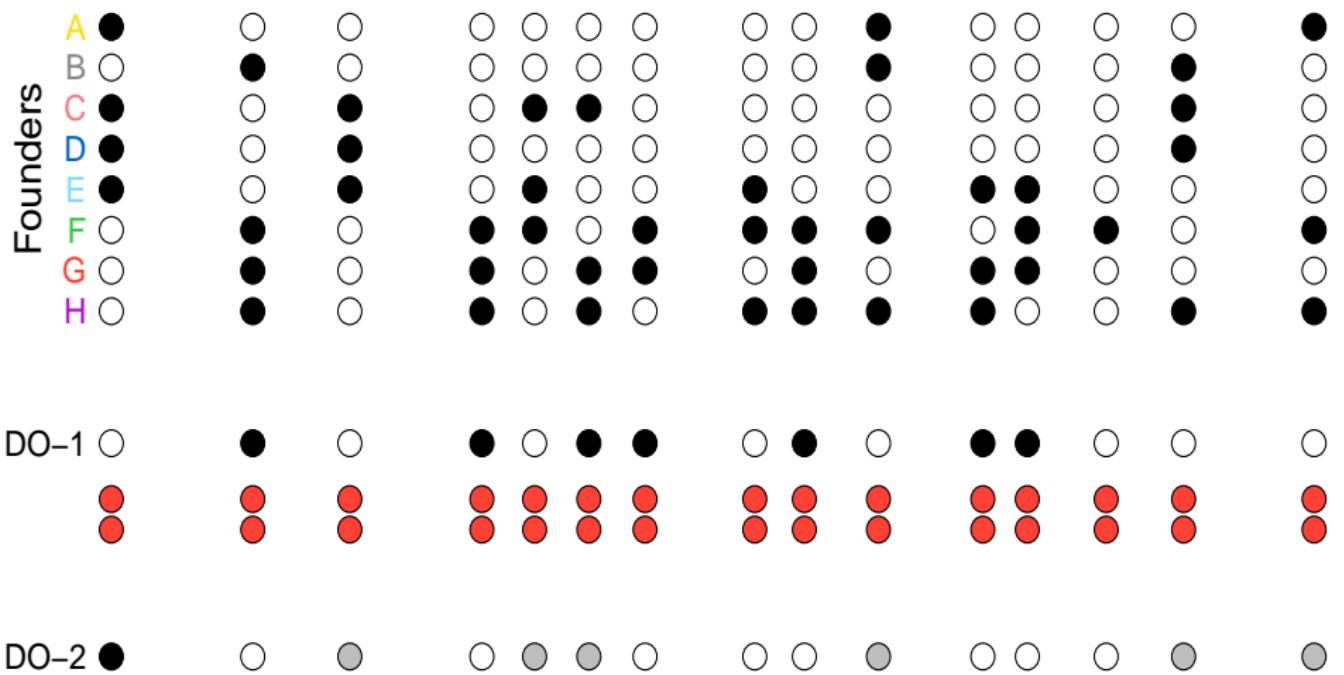
○ ○ ○
○ ○ ○
○ ○ ○
○ ○ ○
○ ○ ○
● ● ○
○ ○ ○
● ○ ○

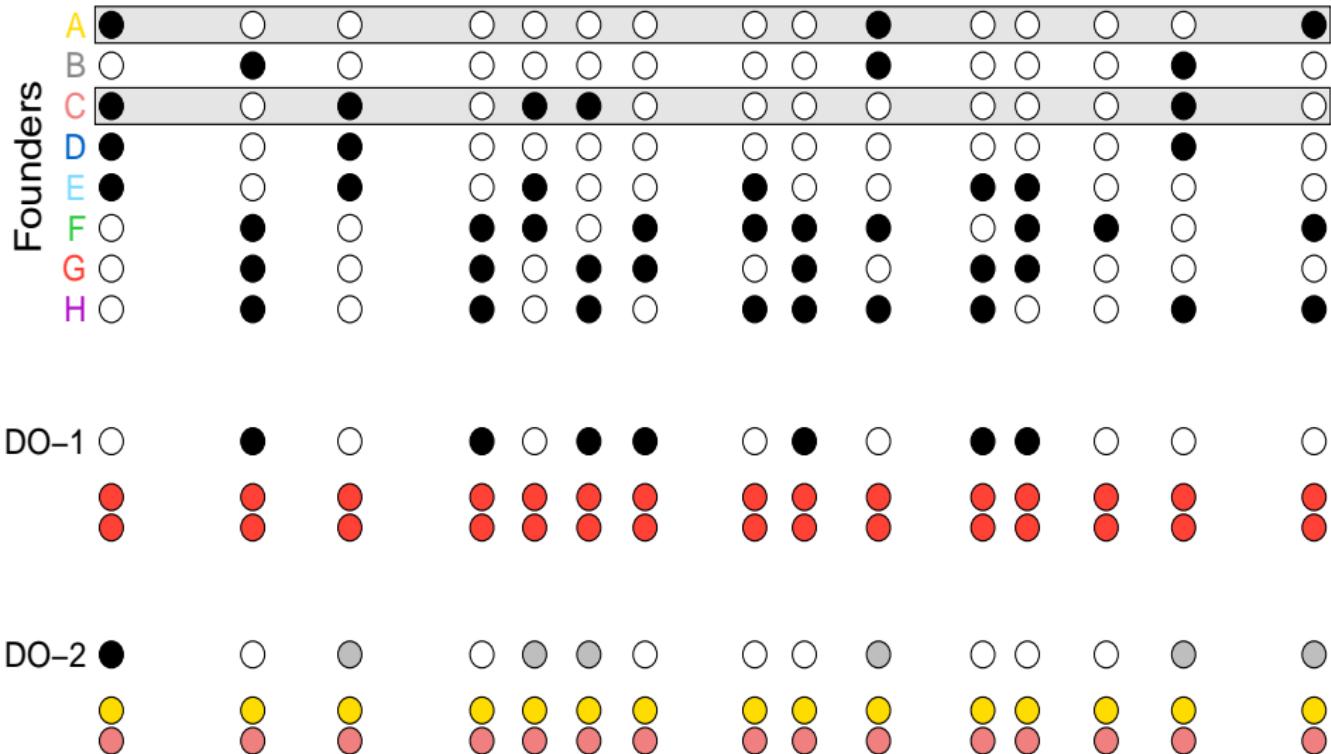
○ ○ ○
○ ○ ○
○ ○ ○
○ ○ ○
○ ○ ○
○ ○ ○
○ ○ ○
● ○ ○

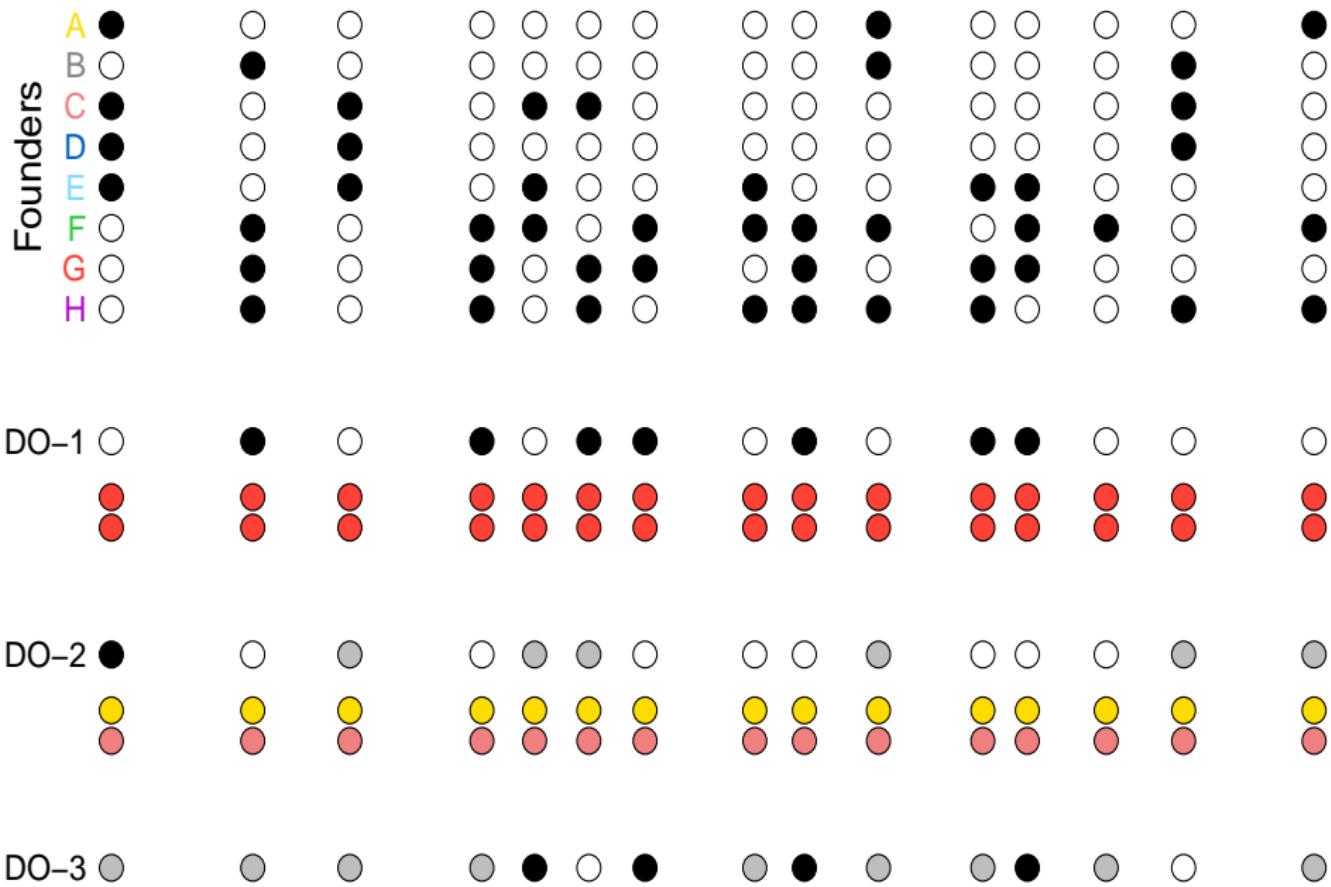
DO-1 ○

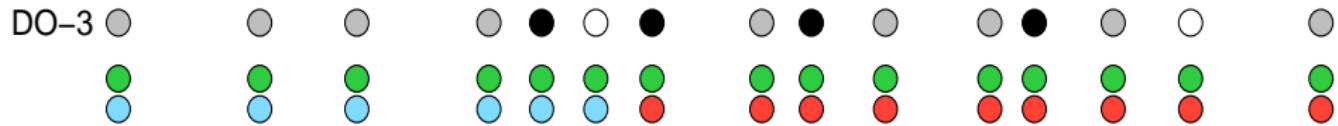
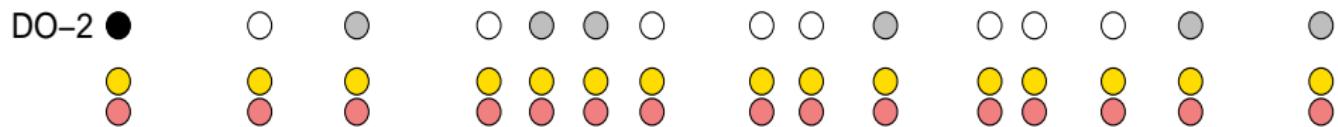
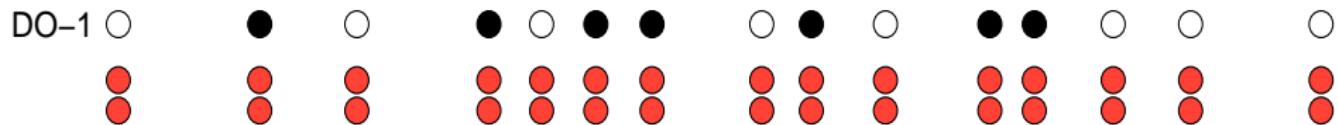
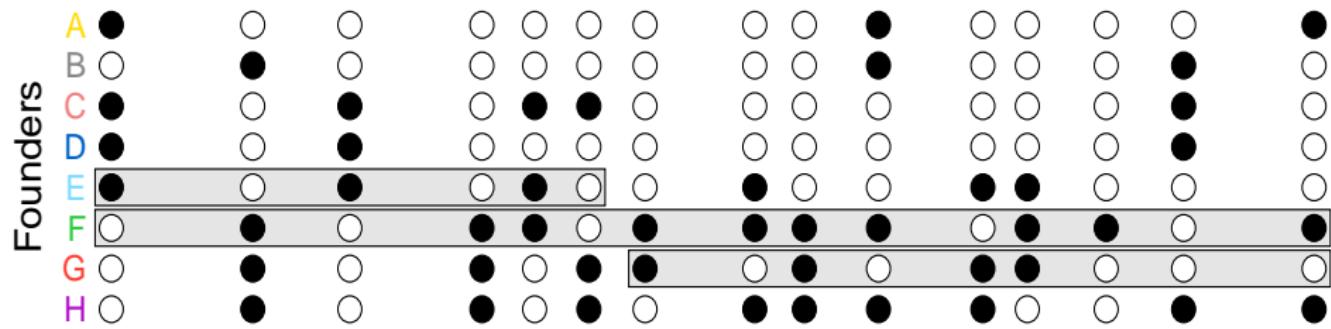
● ○ ● ○ ● ● ○ ● ● ○ ○ ○



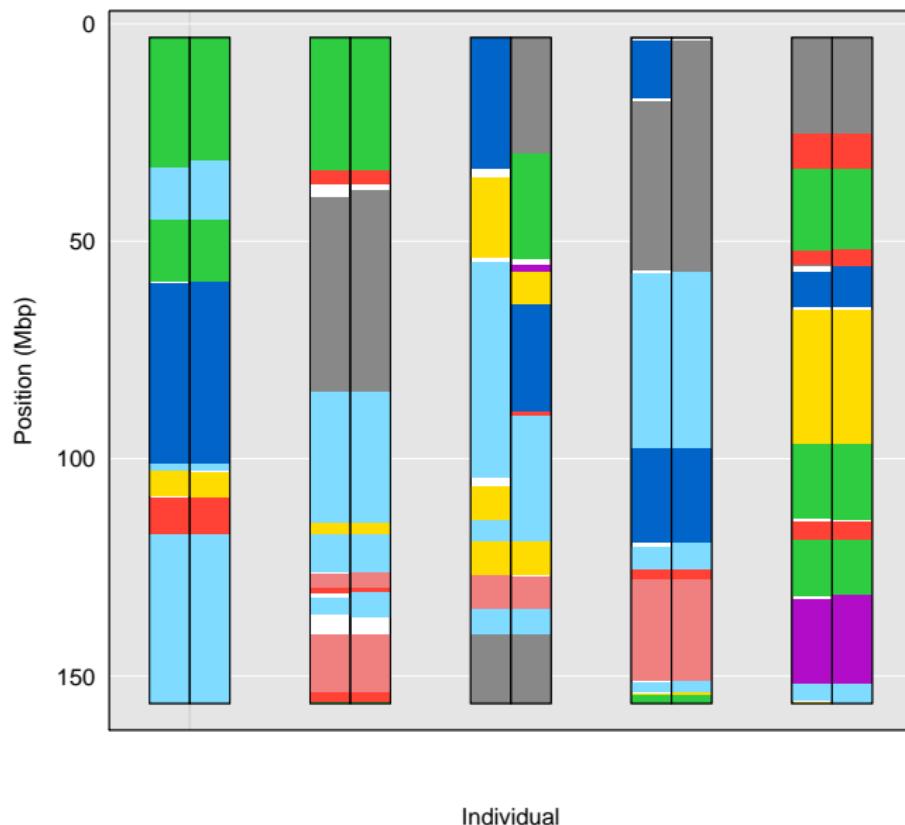








Inferred Haplotypes



Data

Two *large* 3D arrays of emission probabilities

- ▶ `r\qt12` (Broman)
- ▶ `DOQTL` (Gatti)

Data

Two *large* 3D arrays of emission probabilities

- ▶ r\qt12 (Broman)
- ▶ DOQTL (Gatti)

$500 \times 120,000 \times 36$ (Individual \times Markers \times Diplotype)

Data

Two *large* 3D arrays of emission probabilities

- ▶ r\qt12 (Broman)
- ▶ DOQTL (Gatti)

$500 \times 120,000 \times 36$ (Individual \times Markers \times Diplotype)

How are they different?

Data

Two *large* 3D arrays of emission probabilities

- ▶ r\qt12 (Broman)
- ▶ DOQTL (Gatti)

$500 \times 120,000 \times 36$ (Individual \times Markers \times Diplotype)

How are they different?

How can we visualize this?

Measure of distance

Measure of distance

For each individual at each marker:

Compute *sum of absolute differences*

$$\sum_{i=1}^{36} |p_{1,i} - p_{2,i}|$$

Measure of distance

For each individual at each marker:

Compute *sum of absolute differences*

$$\sum_{i=1}^{36} |p_{1,i} - p_{2,i}|$$

Each entry represents distance between methods

Measure of distance

For each individual at each marker:

Compute *sum of absolute differences*

$$\sum_{i=1}^{36} |p_{1,i} - p_{2,i}|$$

Each entry represents distance between methods

Reduces problems to two-dimensions:

Measure of distance

For each individual at each marker:

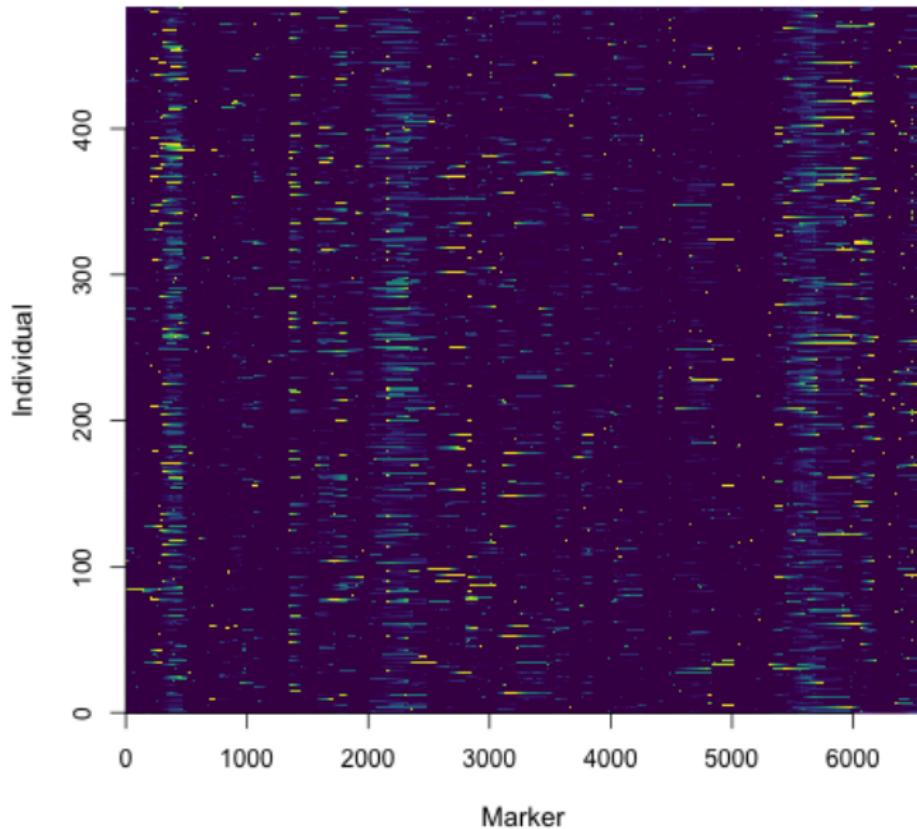
Compute *sum of absolute differences*

$$\sum_{i=1}^{36} |p_{1,i} - p_{2,i}|$$

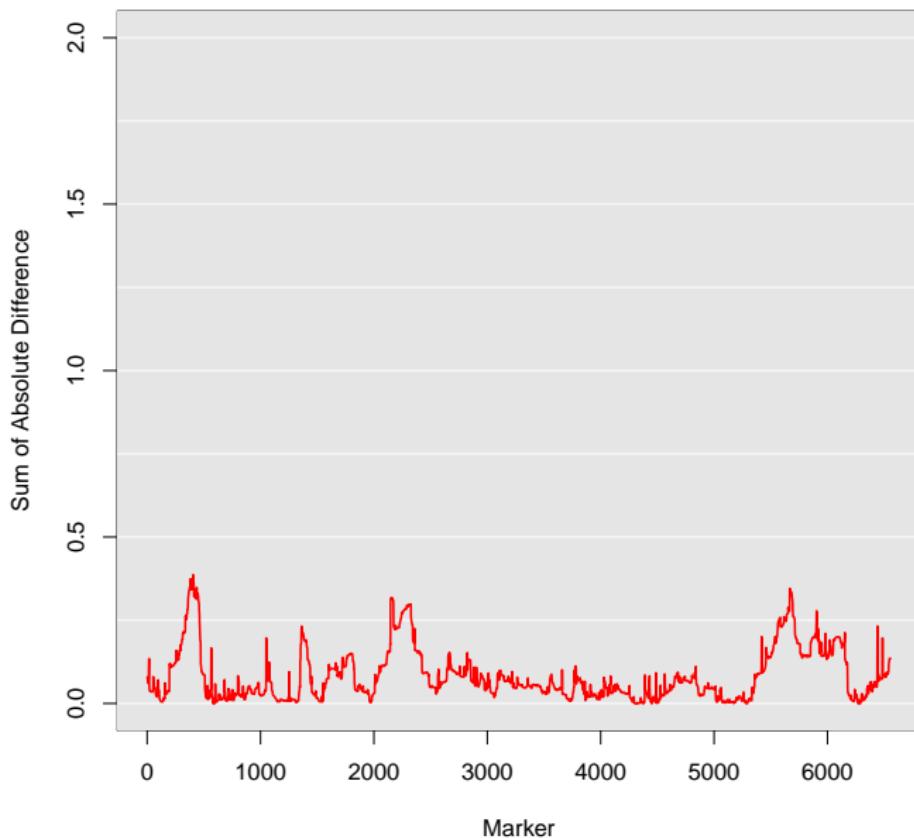
Each entry represents distance between methods

Reduces problems to two-dimensions: $500 \times 120,000$

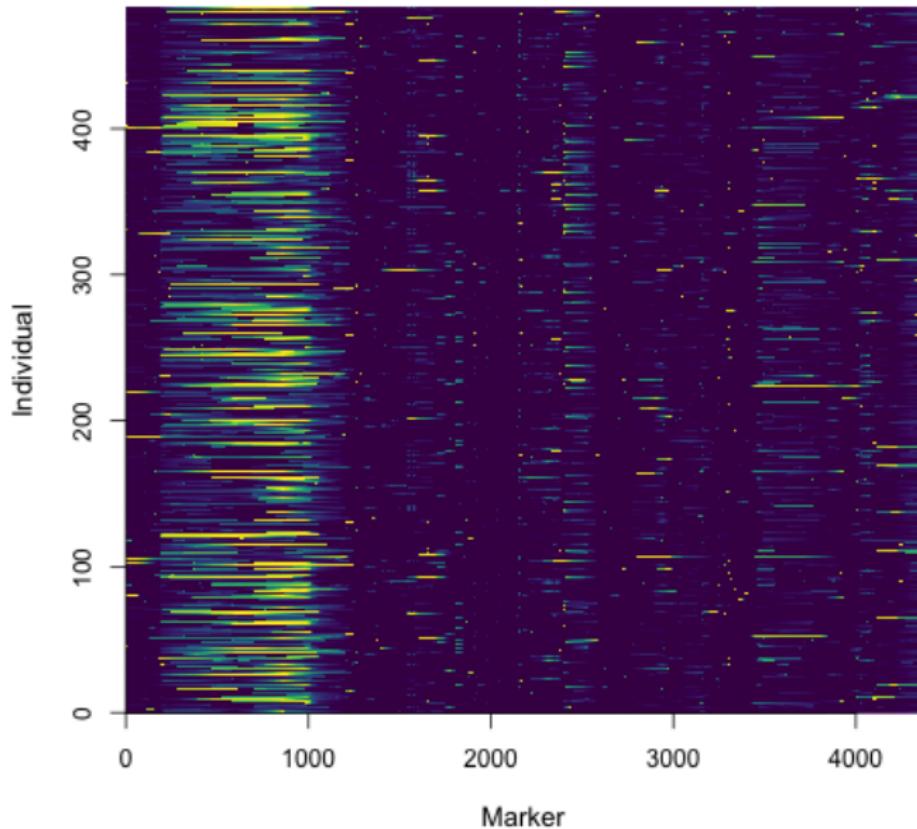
Chromosome 5



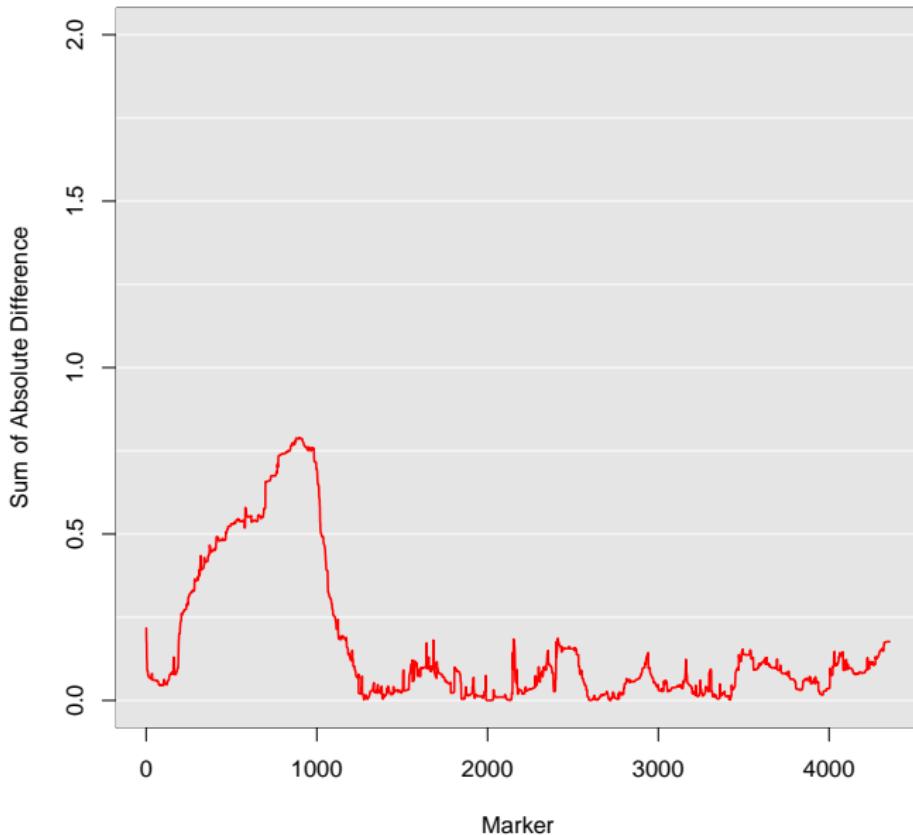
Chromosome 5

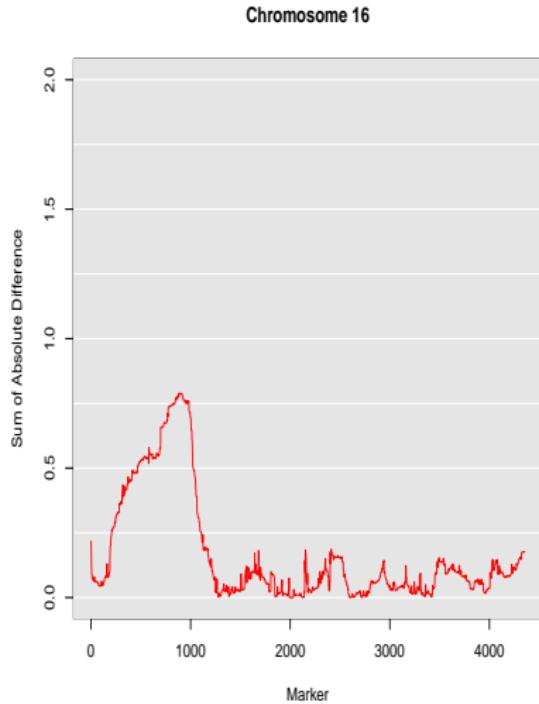
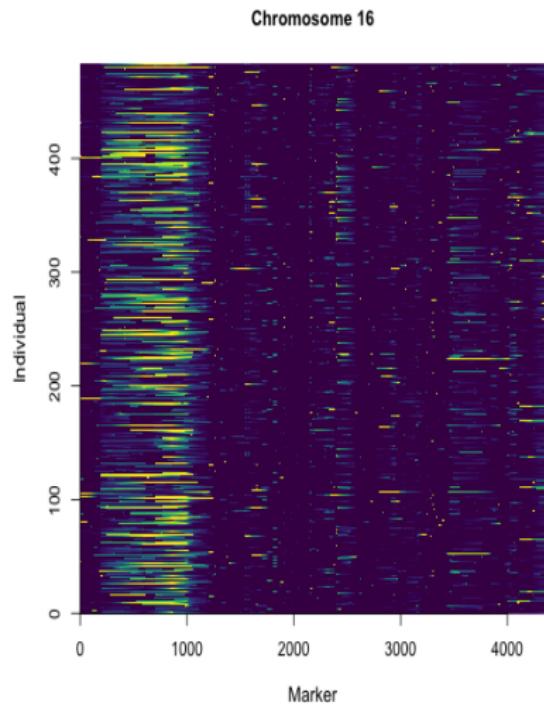


Chromosome 16

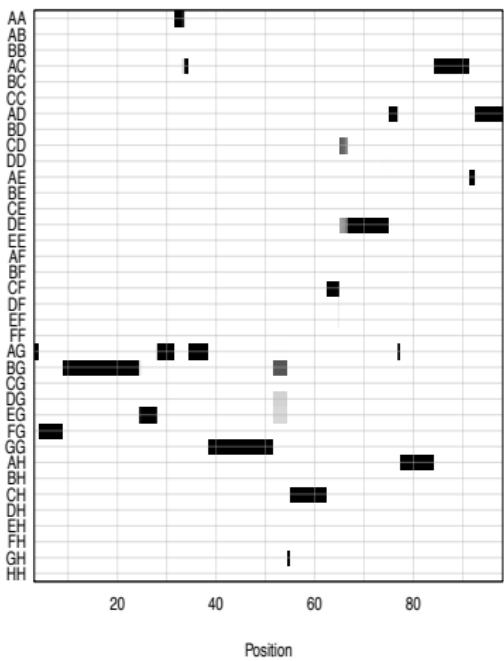


Chromosome 16

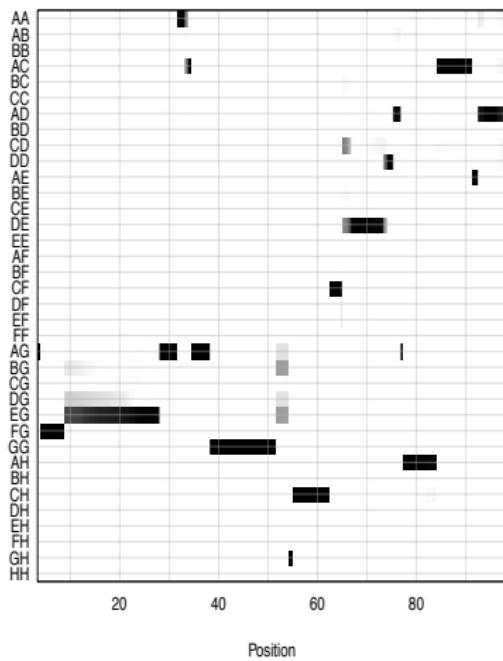




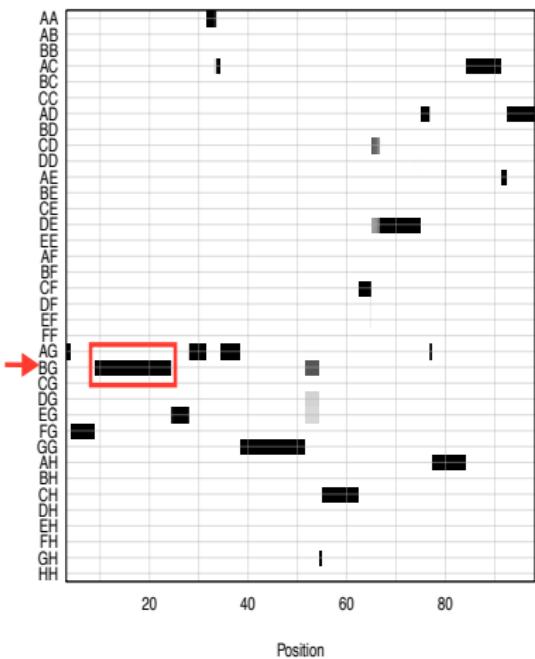
DOQTL DO-171



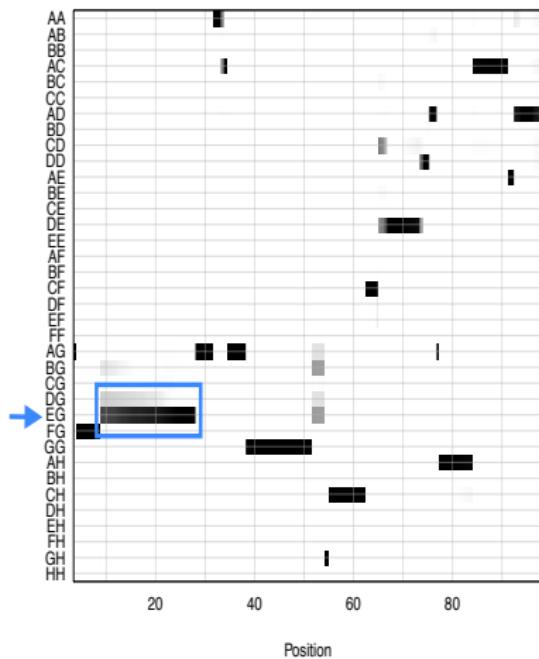
r/qtl2 DO-171



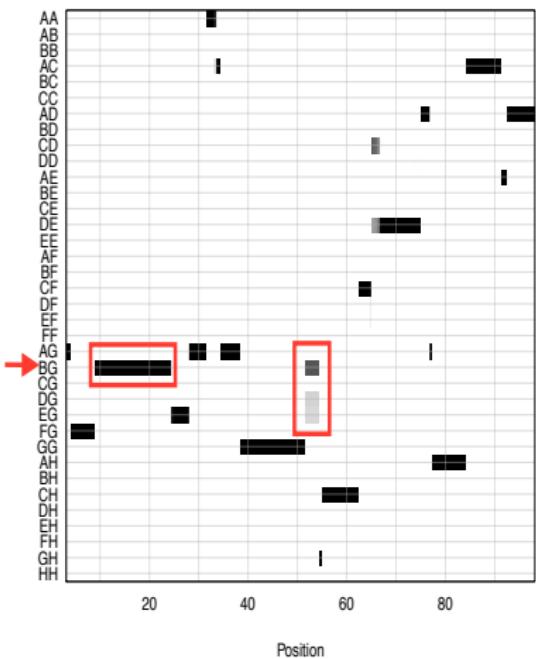
DOQTL DO-171



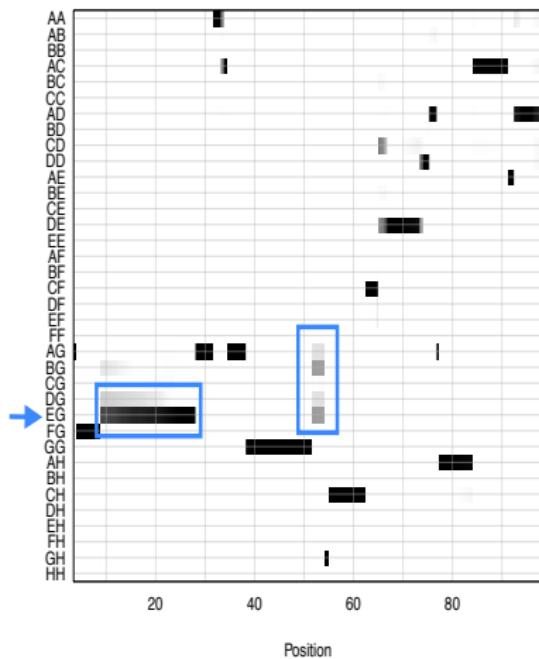
r/qtl2 DO-171

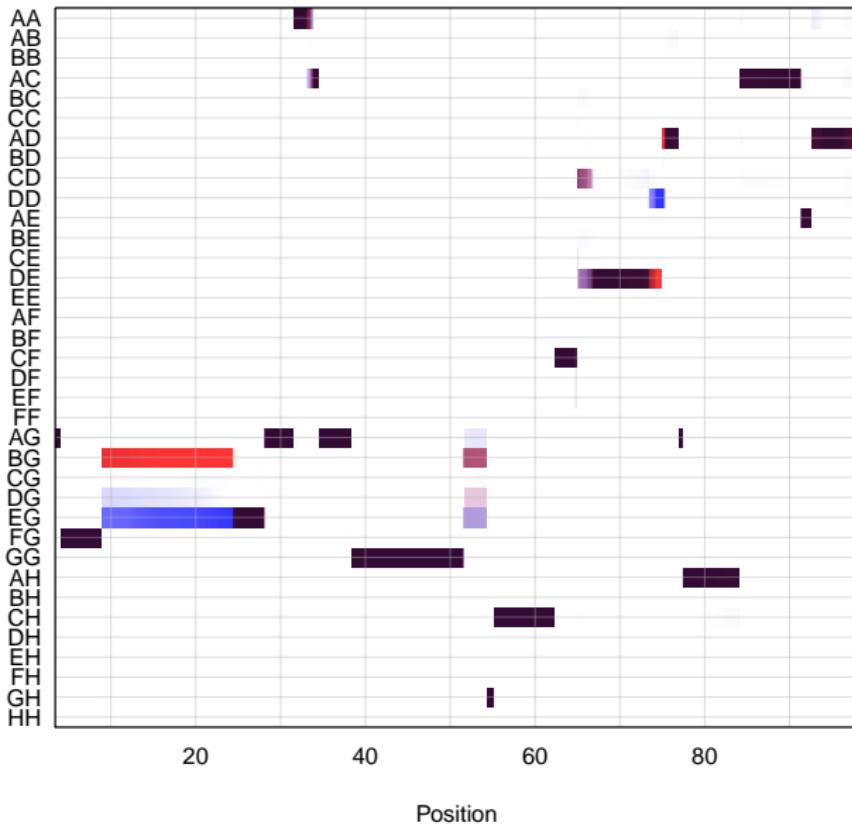


DOQTL DO-171



r/qtl2 DO-171





Which answer is 'correct'?

- ▶ Pick an interval with differing answer
- ▶ Compare SNP genotypes to expected genotypes