Multiparent populations & R/qtl2

Karl Broman

Biostatistics and Medical Informatics
University of Wisconsin – Madison

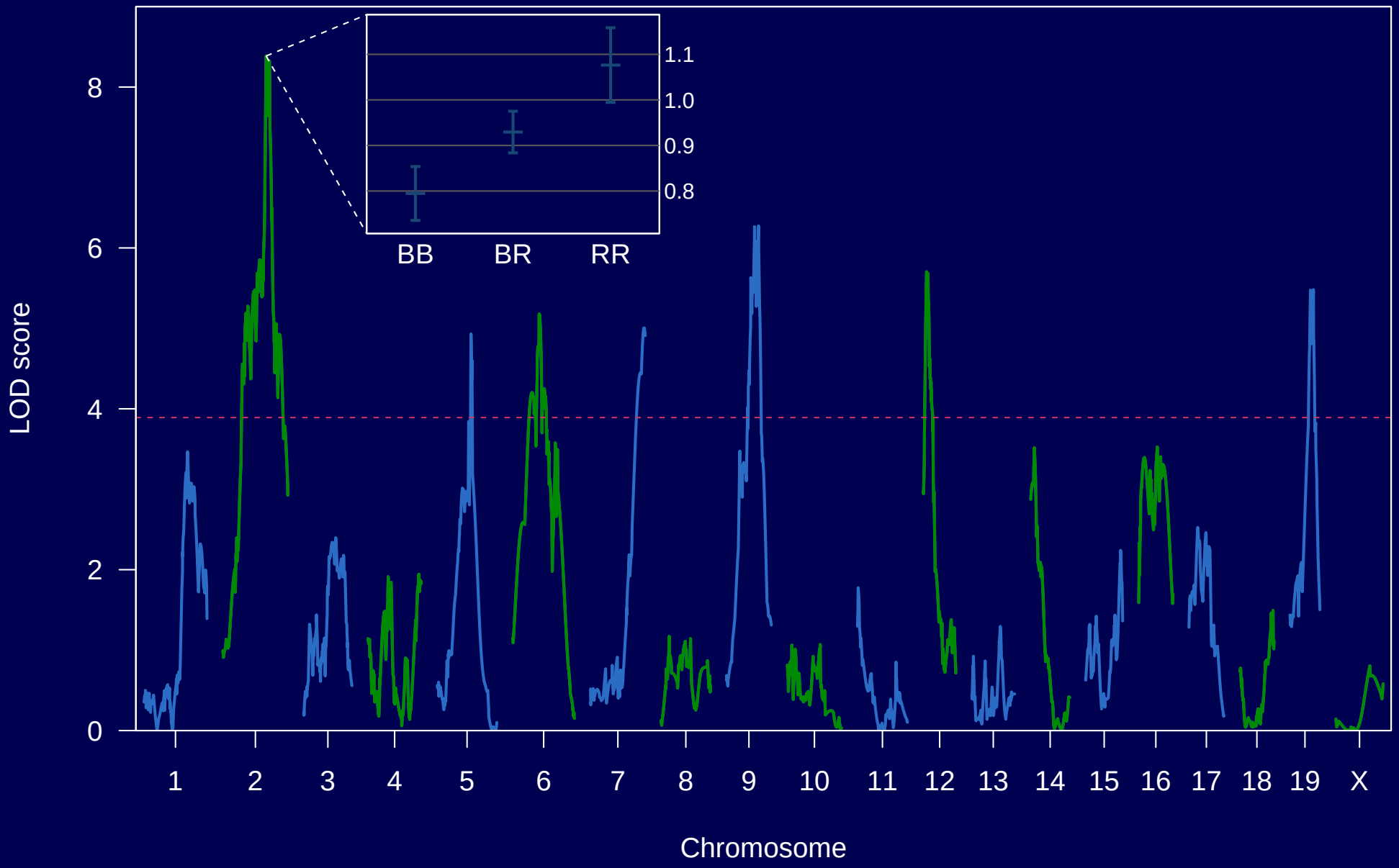
kbroman.org/qtl2

kbroman.org

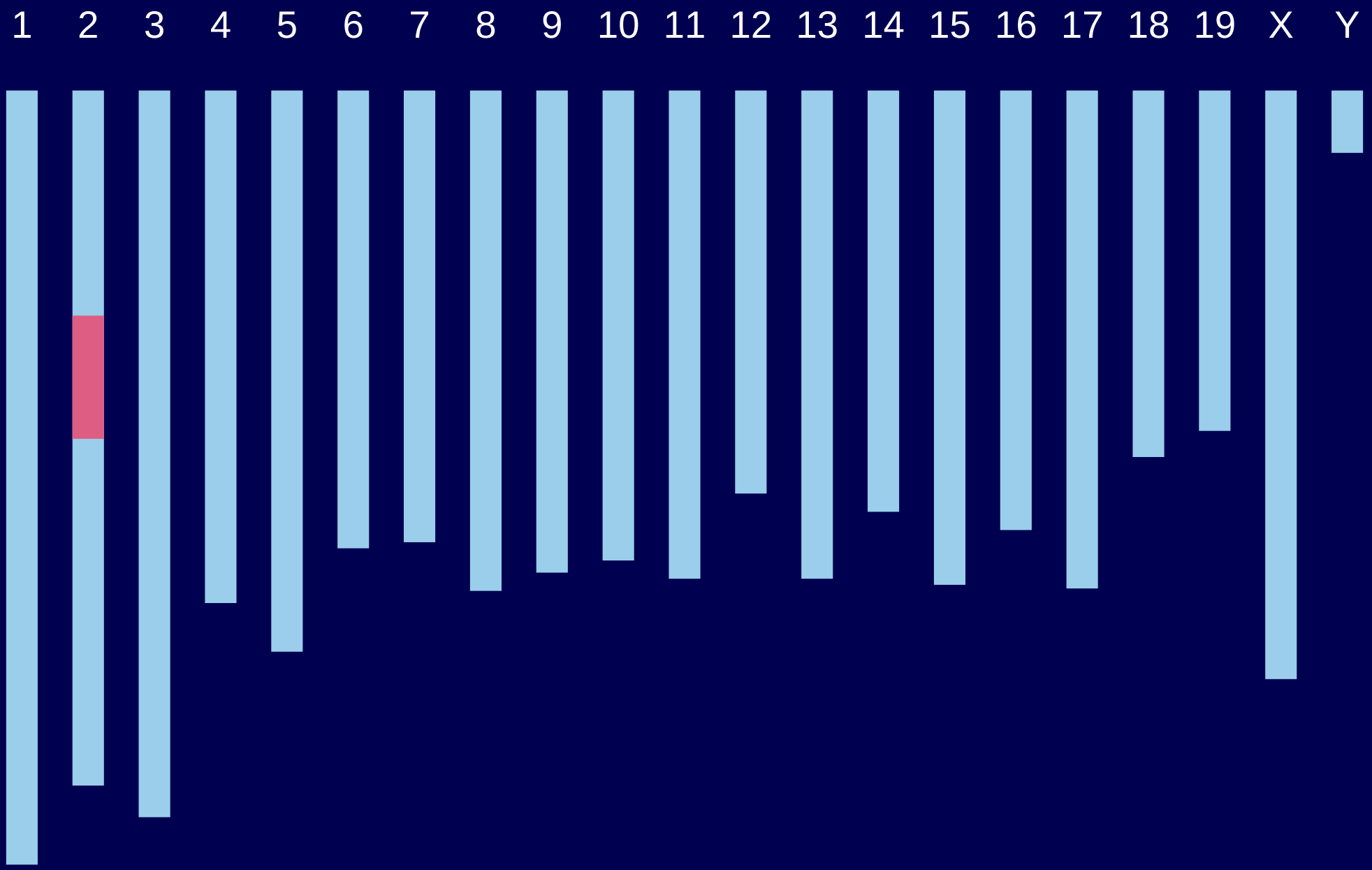
github.com/kbroman

@kwbroman

QTL mapping


LOD score
0
2
4
6
8
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
X
Chromosome
1.1
1.0
0.9
0.8
BB
BR
RR

Congenic line / NIL

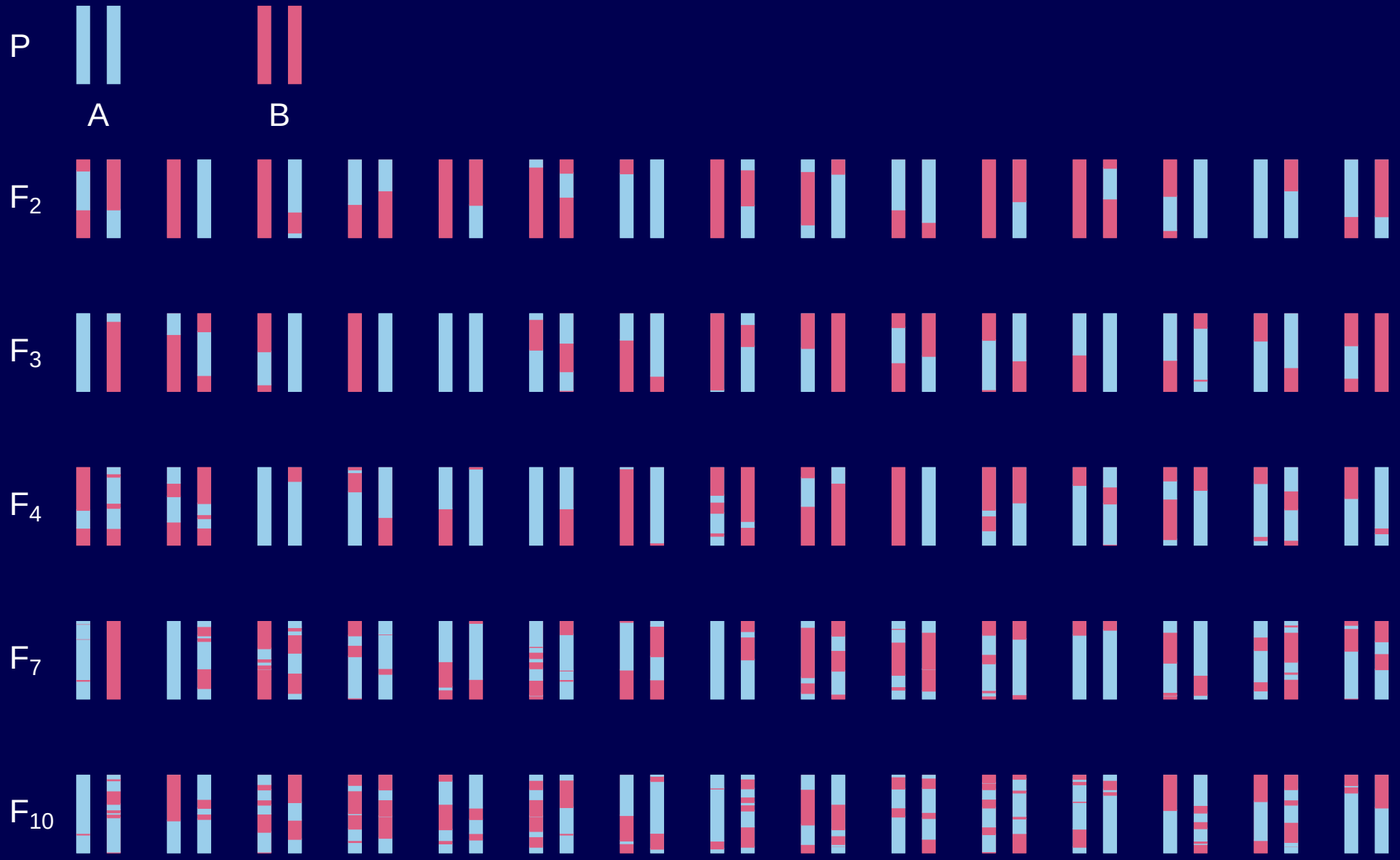


Improving precision

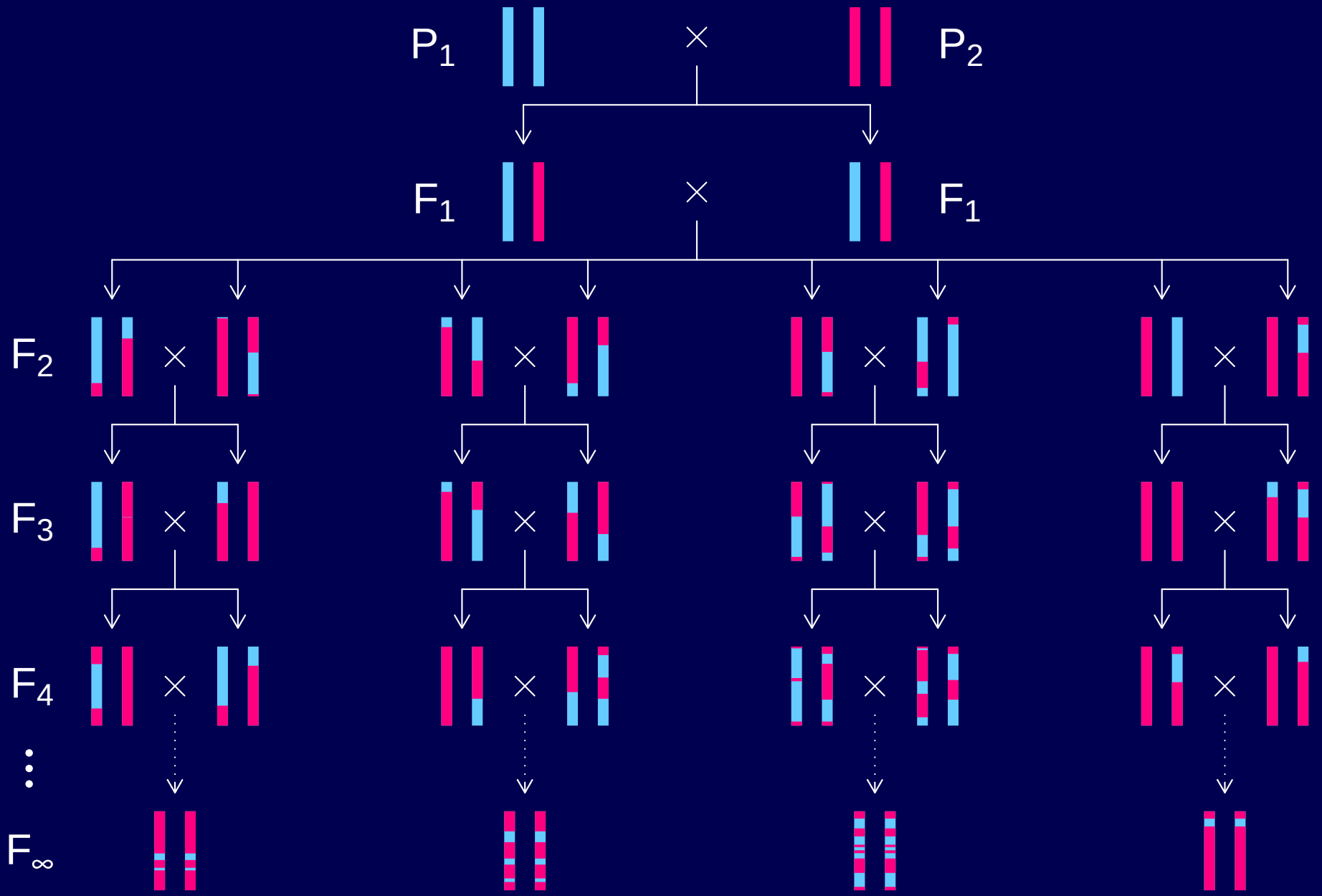
- more recombinations
- more individuals
- more precise phenotypes
- lower-level phenotypes

  transcripts, proteins, metabolites

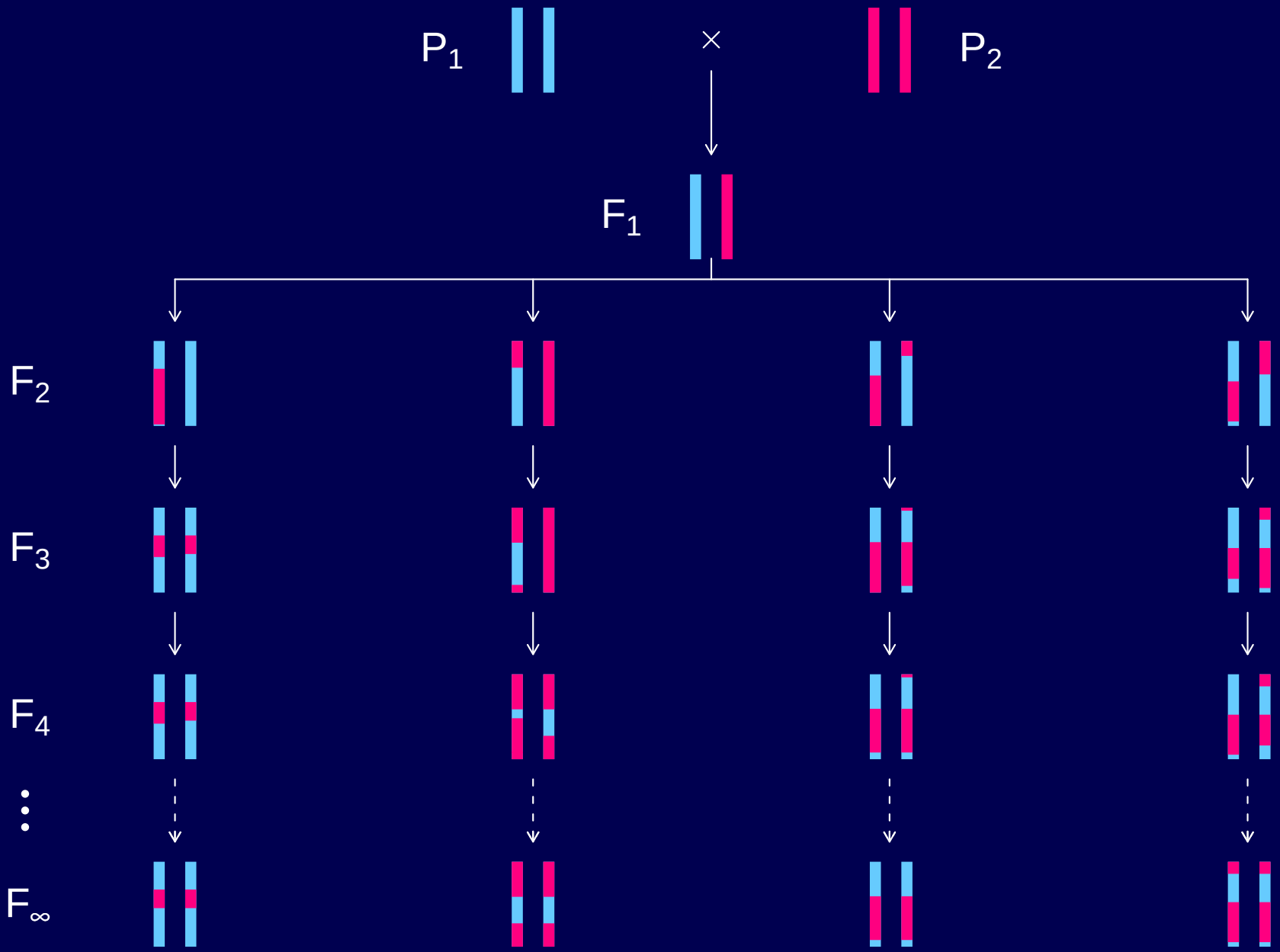
Advanced intercross lines



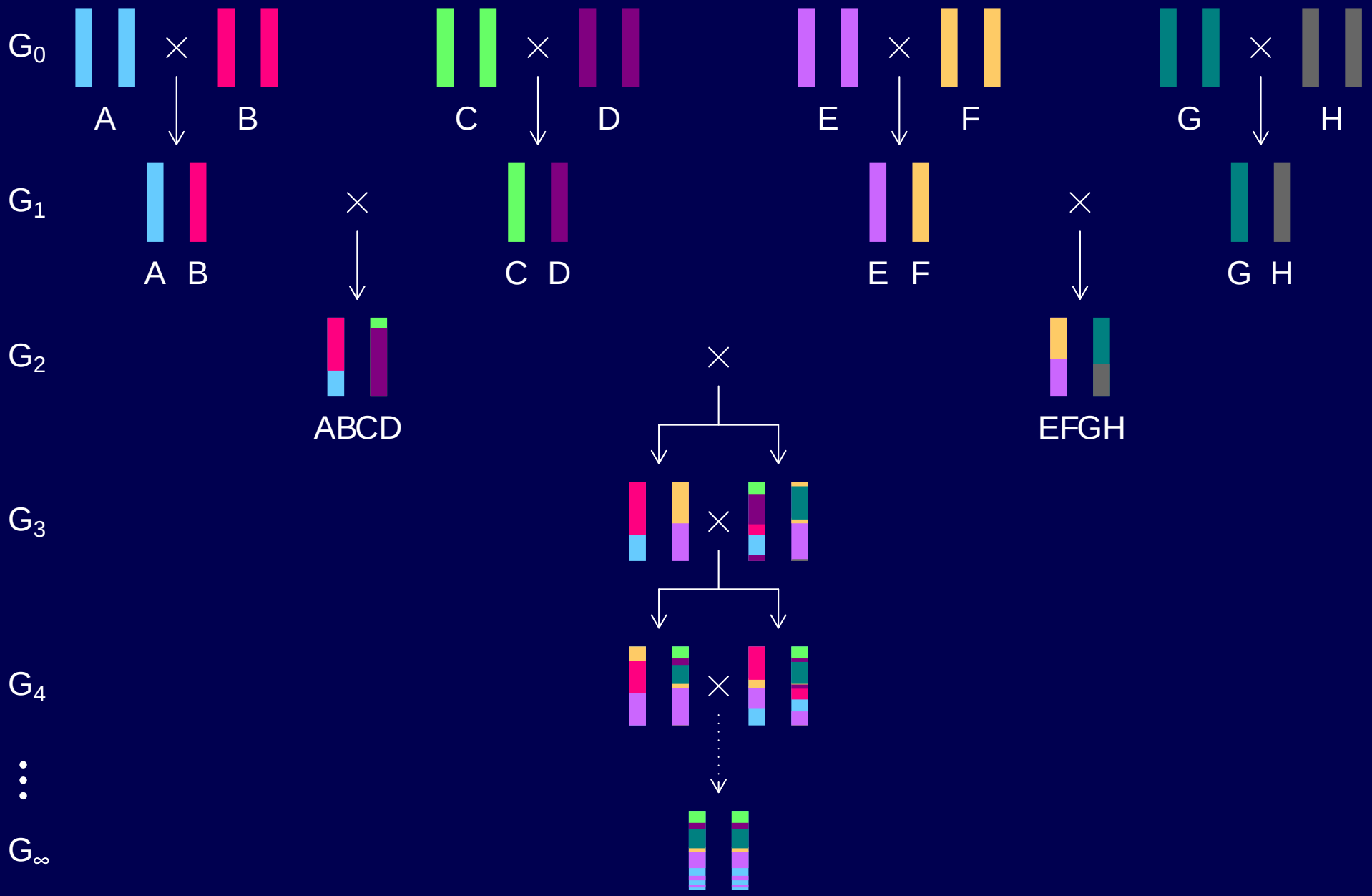
Recombinant inbred lines



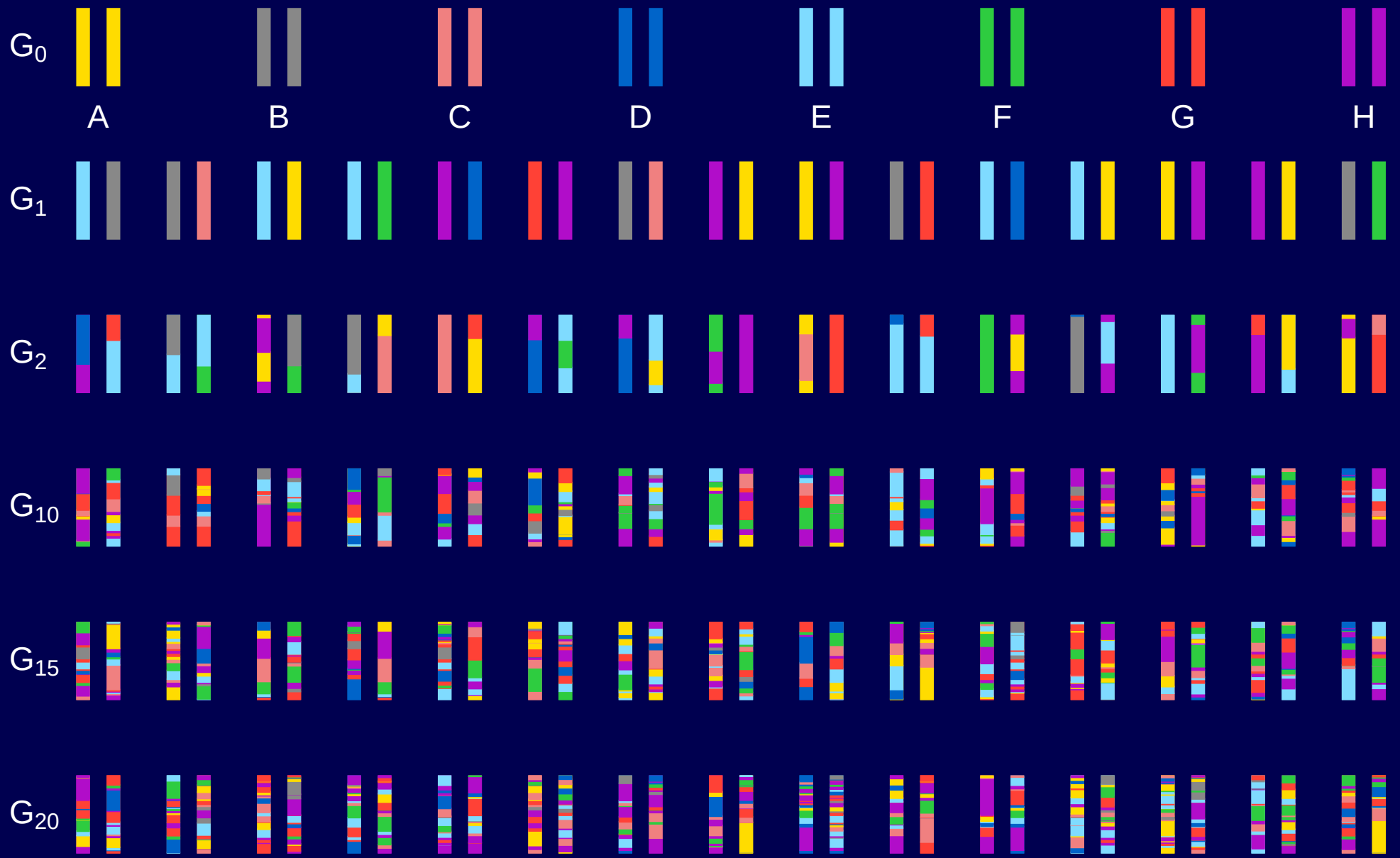
Recombinant inbred lines



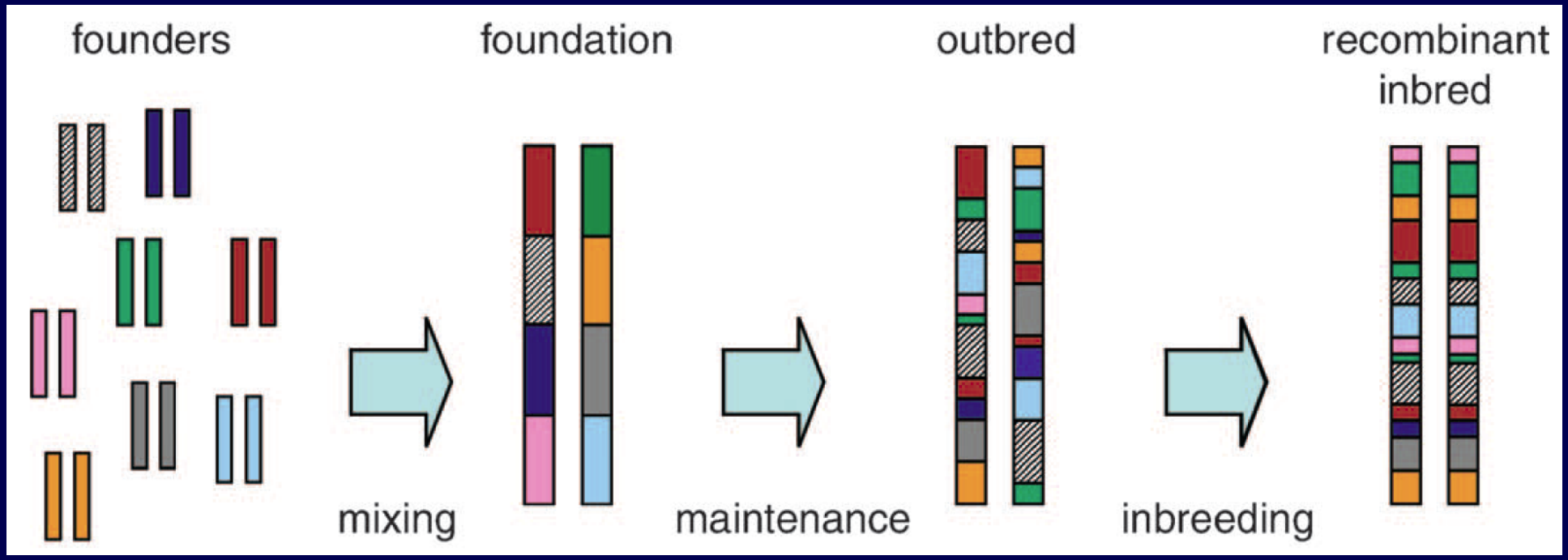
Collaborative Cross/MAGIC



Heterogeneous stock



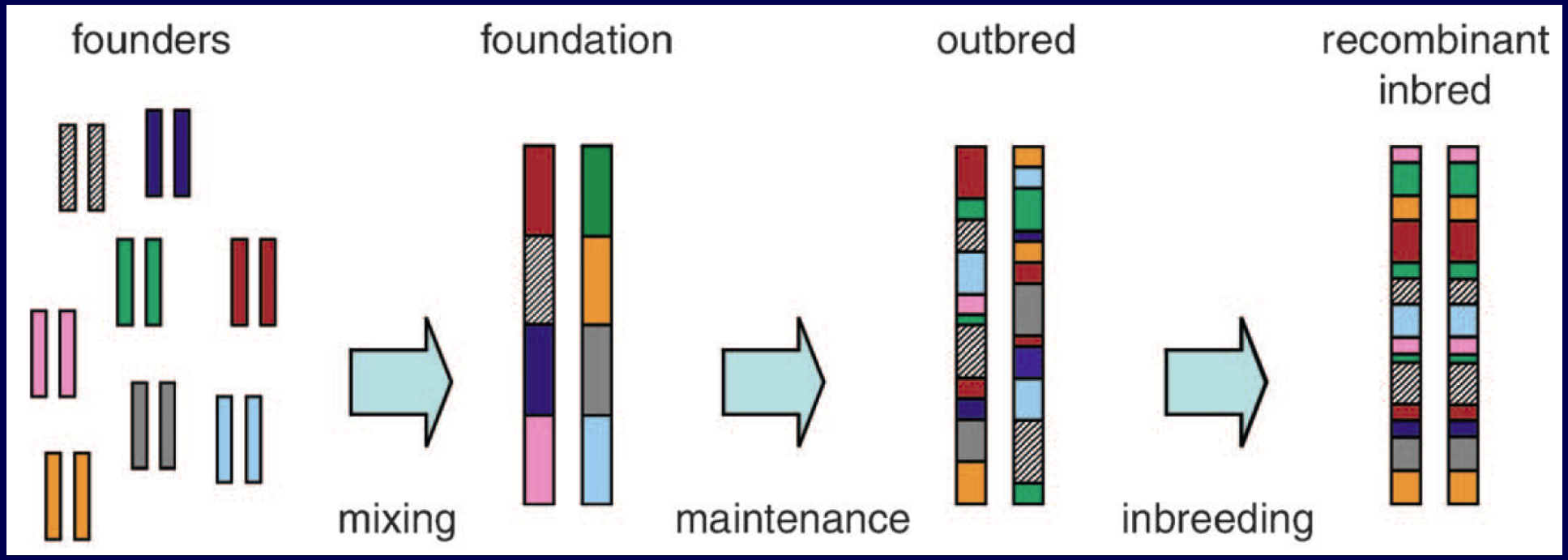
MAGIC lines



combine mix fix

Valdar et al., Genetics 172:1783, 2006

MAGIC lines

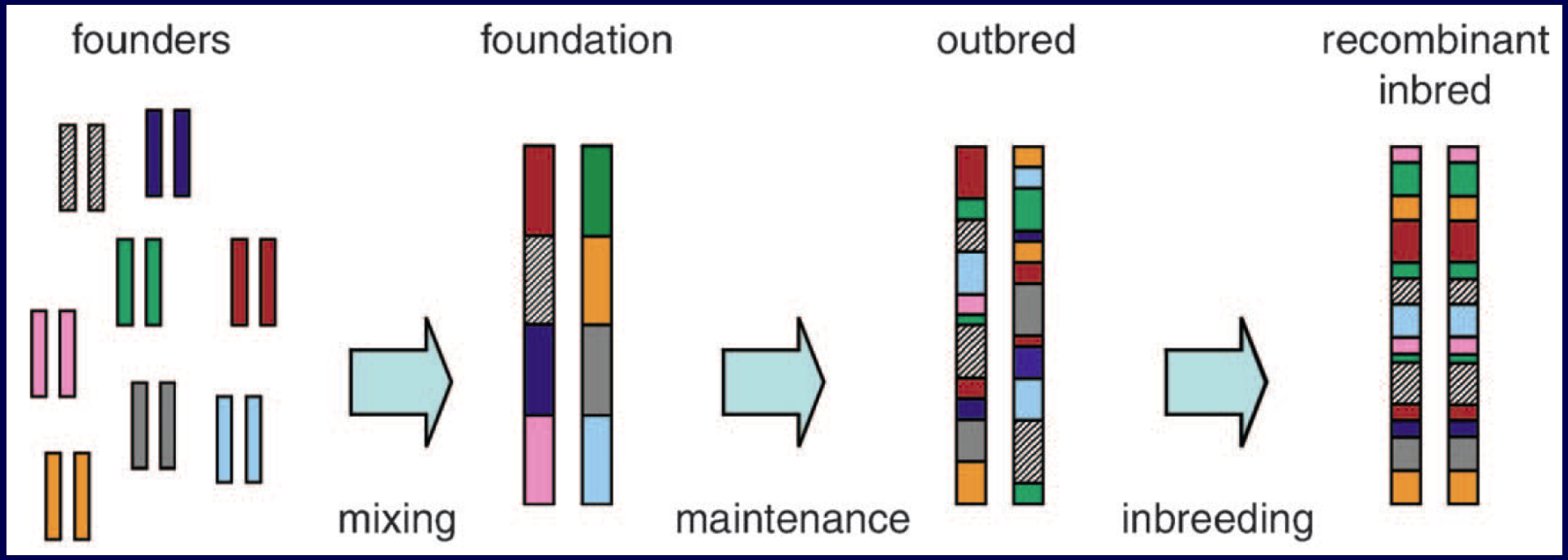


combine mix fix

How many?

Valdar et al., Genetics 172:1783, 2006

MAGIC lines



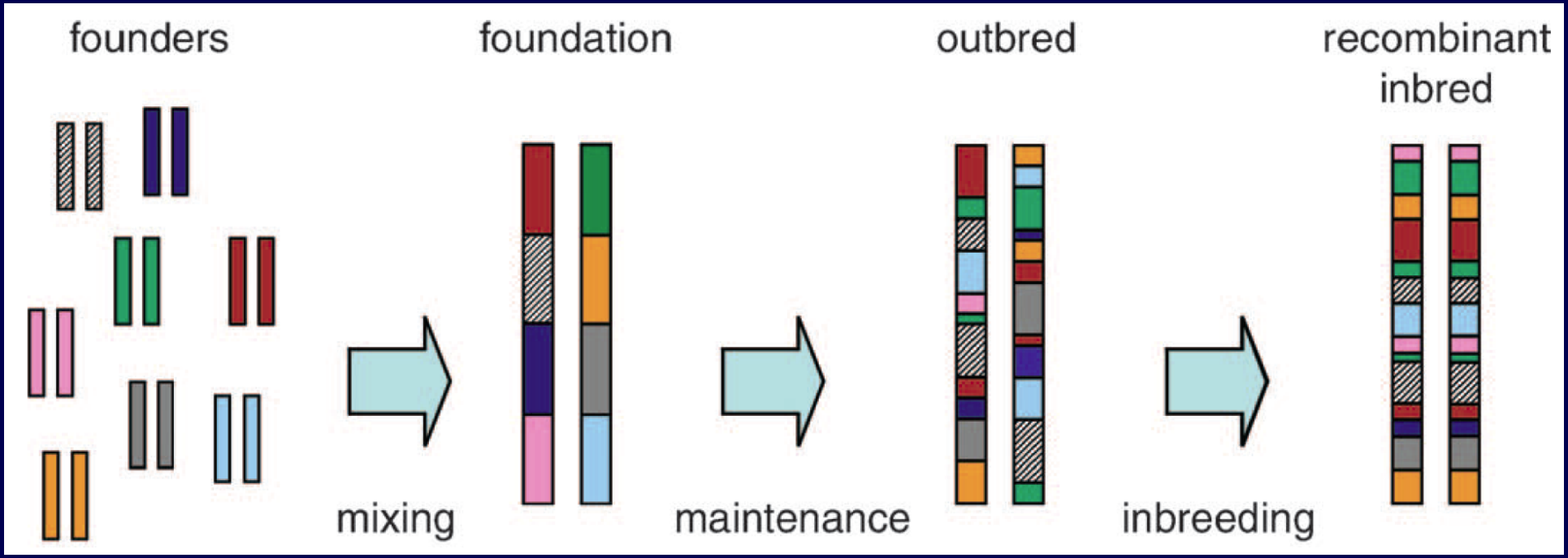
combine mix fix

How many?

Which?

Valdar et al., Genetics 172:1783, 2006

MAGIC lines



combine mix fix

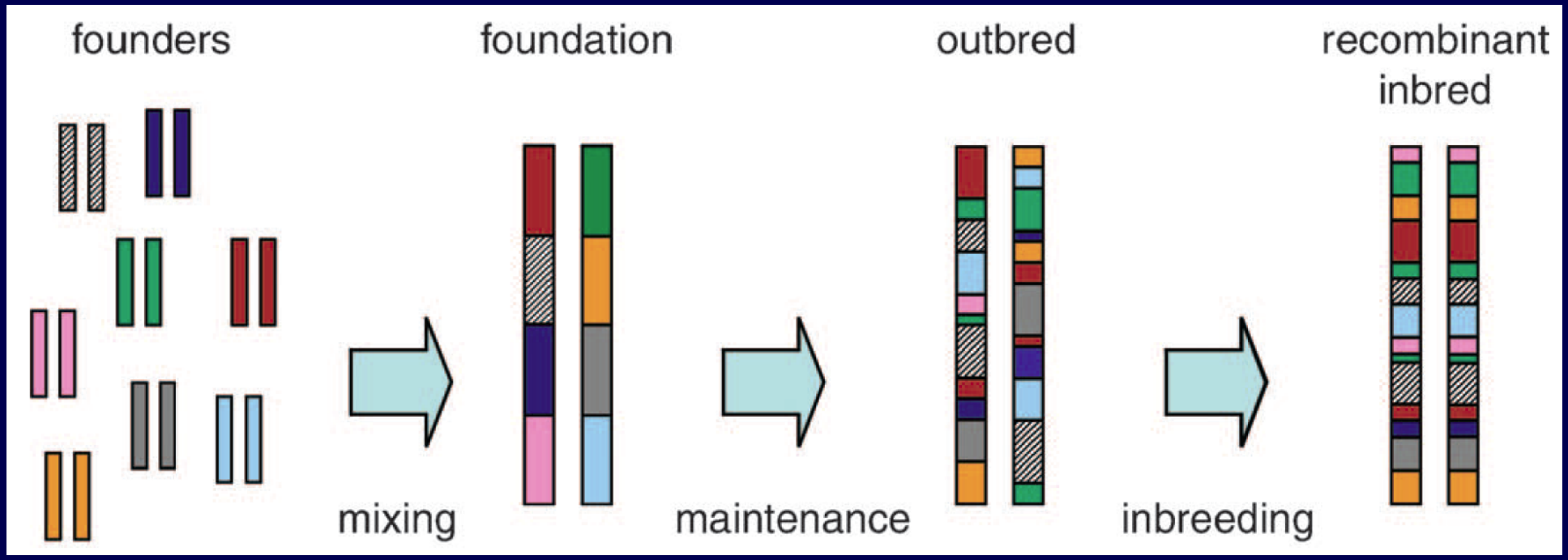
How many?

Which?

How long?

Valdar et al., Genetics 172:1783, 2006

MAGIC lines



combine mix fix

How many?

Which?

How long?

How?

Valdar et al., Genetics 172:1783, 2006

MAGIC is magic

- Genetic diversity
- High-precision mapping
- Predictable linkage disequilibrium
- Phenotype replicates to reduce individual variation
- Pool phenotypes from multiple labs, environments, treatments
- Genotype once

The goal

Identify QTG

- Power
- Mapping precision
- Estimate QTL allele frequencies

Principles

- Avoid population structure
- Tradeoff between power for *de novo* discovery and mapping precision
- More QTL to find $\Rightarrow$ more QTL getting in the way?
- More QTL alleles $\Rightarrow$ less information about each
- Are QTL alleles common or rare?

How many founders?

More

- More general use
- More QTL
- Greater precision
- Estimate allele frequencies
- Haplotype analysis in founders

Fewer

- Lower residual variance
- Greater power for a particular QTL?
- Better power for epistasis
- Rare alleles are less rare

Which founders?

- Diverse
- Interesting
- No breeding problems
- Balanced: star phylogeny

How much mixing?

- More mixing $\Rightarrow$ Greater mapping precision
- ...but lower power for *de novo* mapping
- Potential for population structure, missing alleles
- Random mating or curated mating?
- Start with many random cross directions?

Selfing or DH?

- Inbreeding gives added recombination
- But not so much as at the mixing stage
- If doubled haploids are feasible, use them

Key analysis issues

How to deal with the multiple alleles?

- Full model (an effect for each allele)
- Diallelic QTL model
- Random effects model (like BLUP)

How to account for multiple QTL?

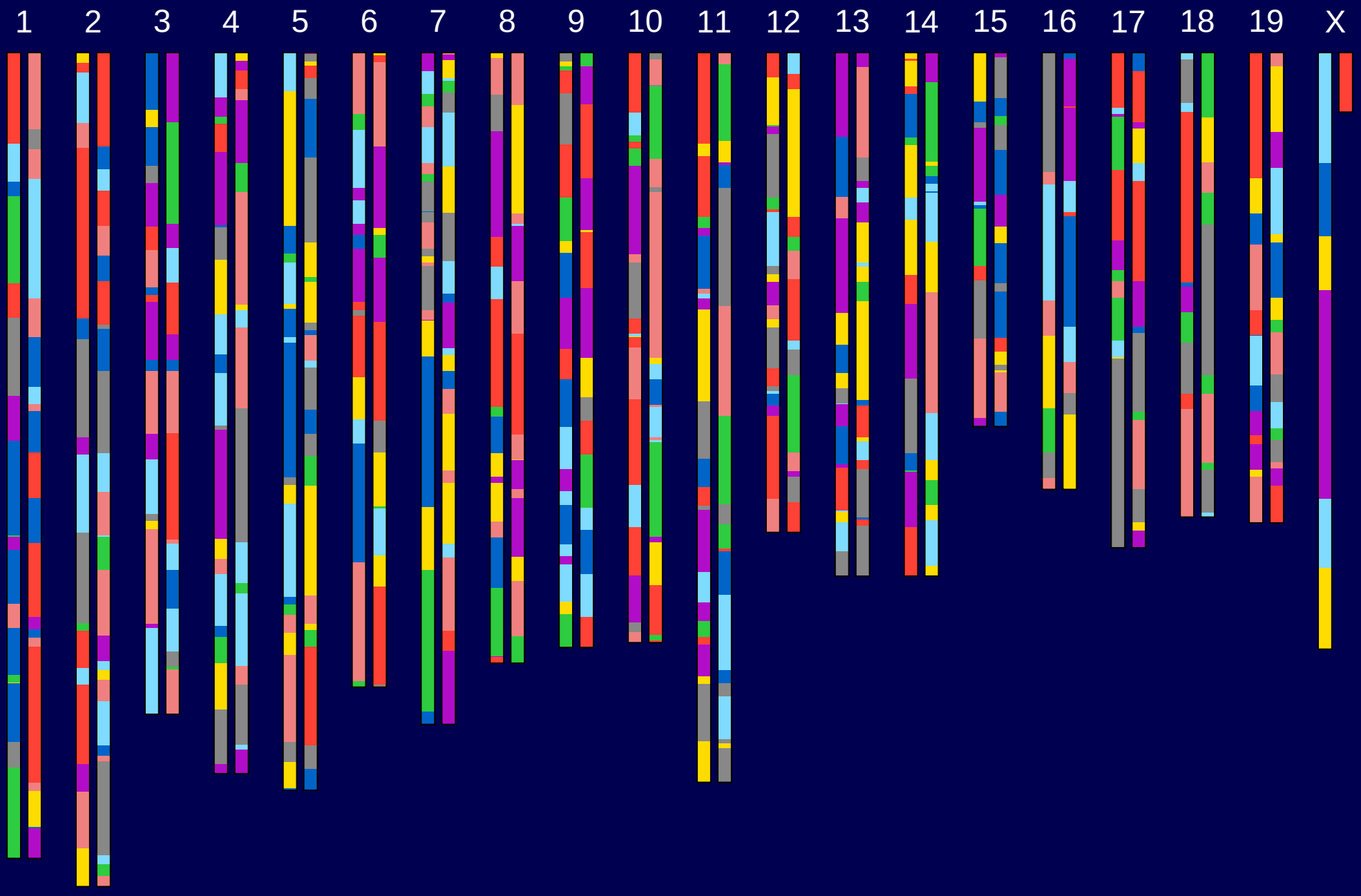
- Stepwise selection
- Bayesian model averaging
- Random effect for polygenes

Linear mixed models

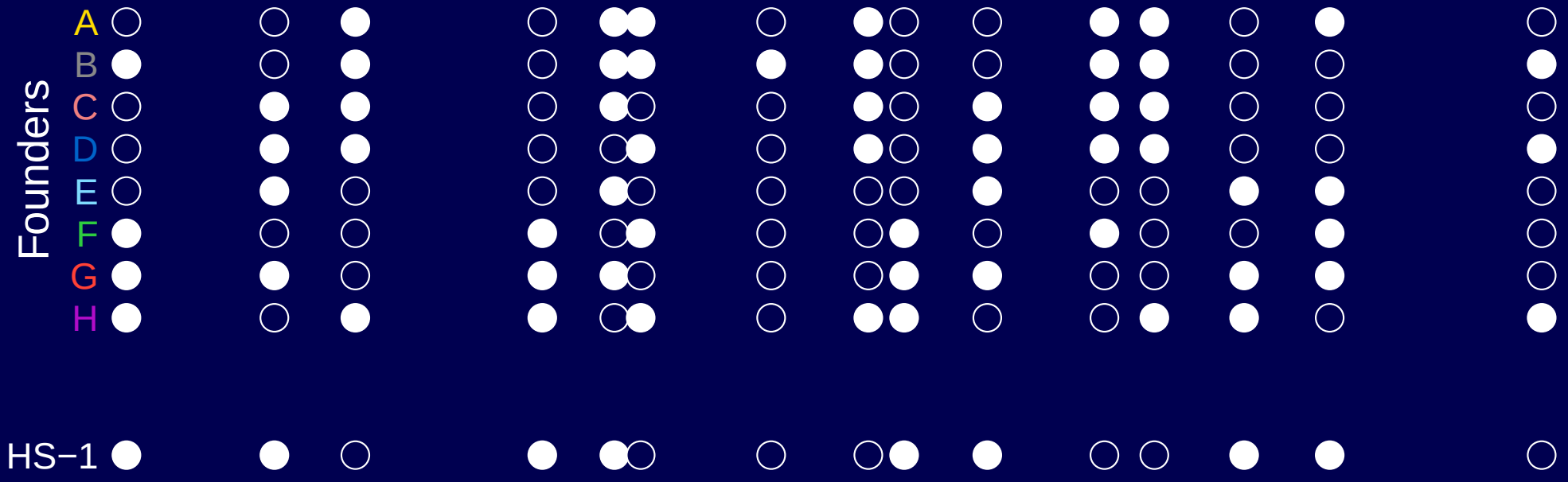
$$
\begin{aligned}
y_i &= \mu + \sum\nolimits_k \beta_k q_{ik} + \epsilon_i & \epsilon_i &\sim \mathsf{N}(0, \sigma_e^2) \\
&= \mu + \eta_i + \epsilon_i & \eta_i &\sim \mathsf{N}(0, \sigma_p^2)
\end{aligned}
$$

$$
\mathsf{COV}(\eta_i, \eta_j) = \sigma_p^2 \, (2k_{ij})
$$

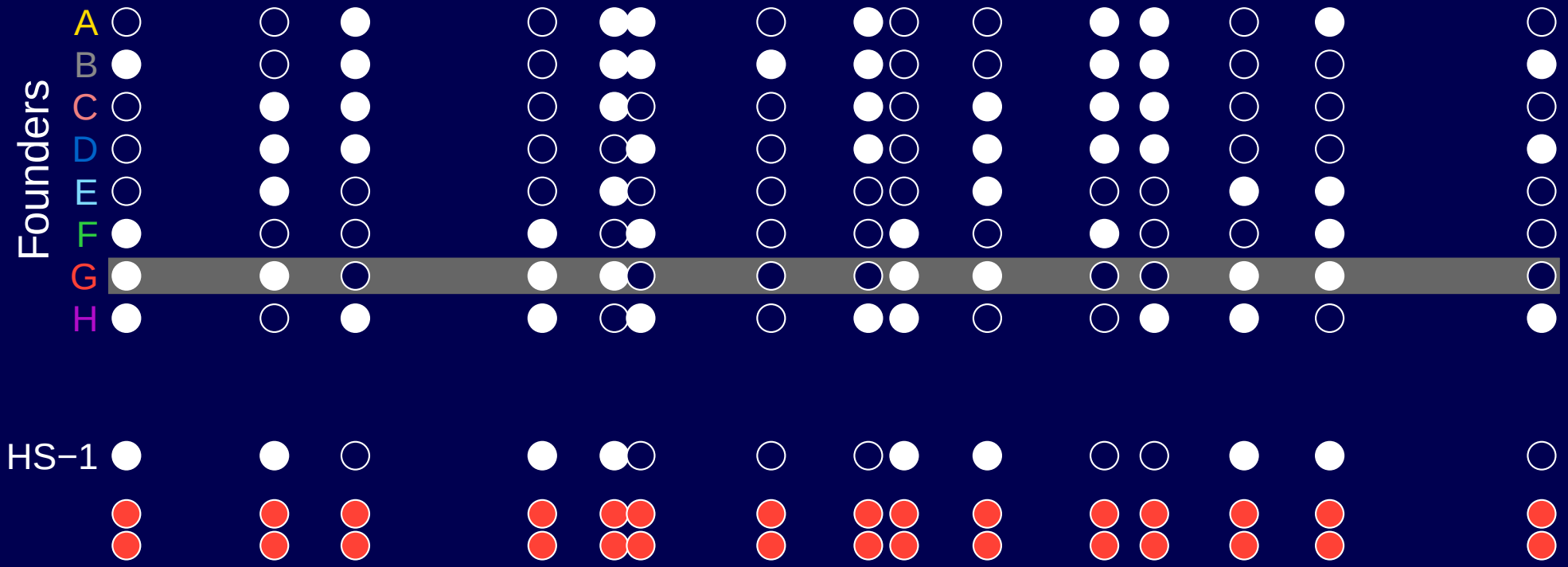
HS genome



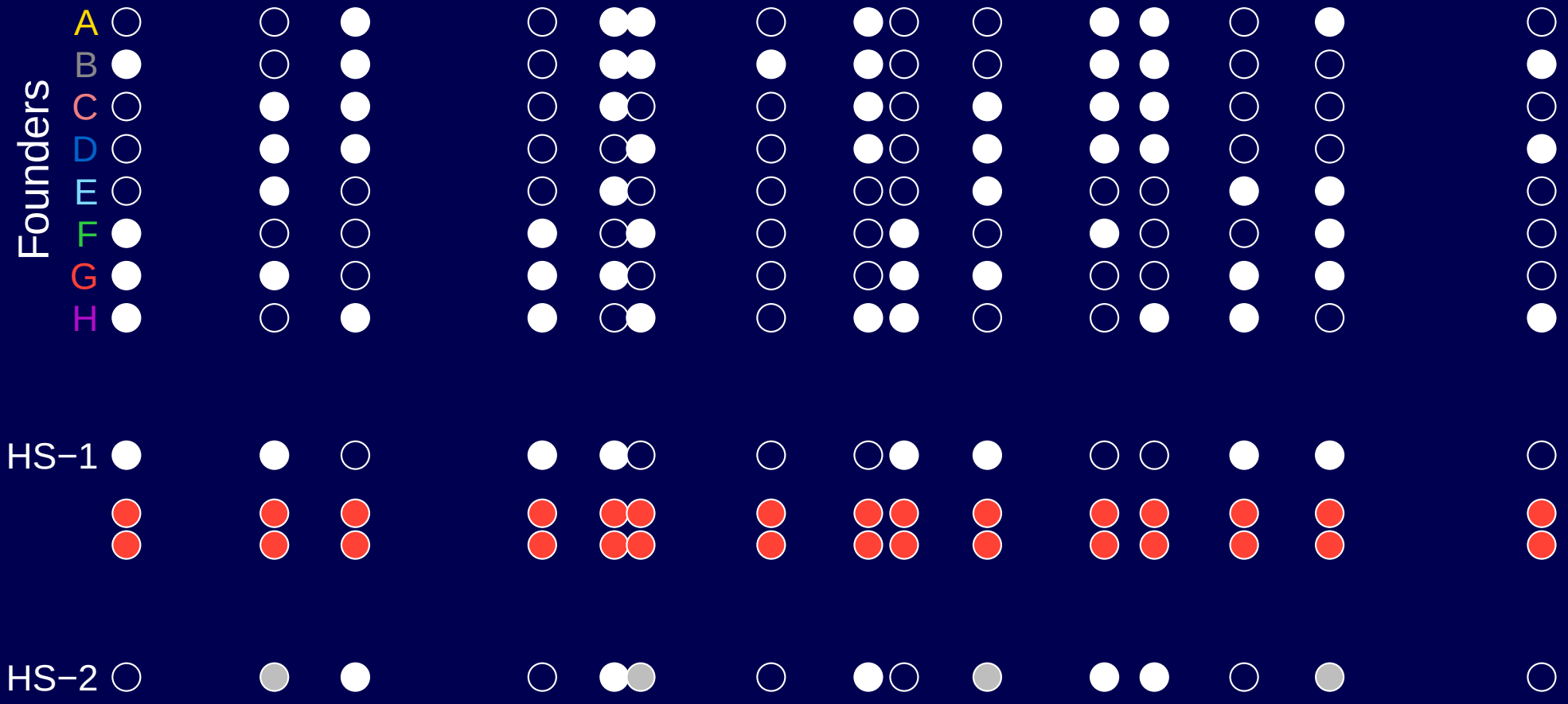
HS genotype reconstruction



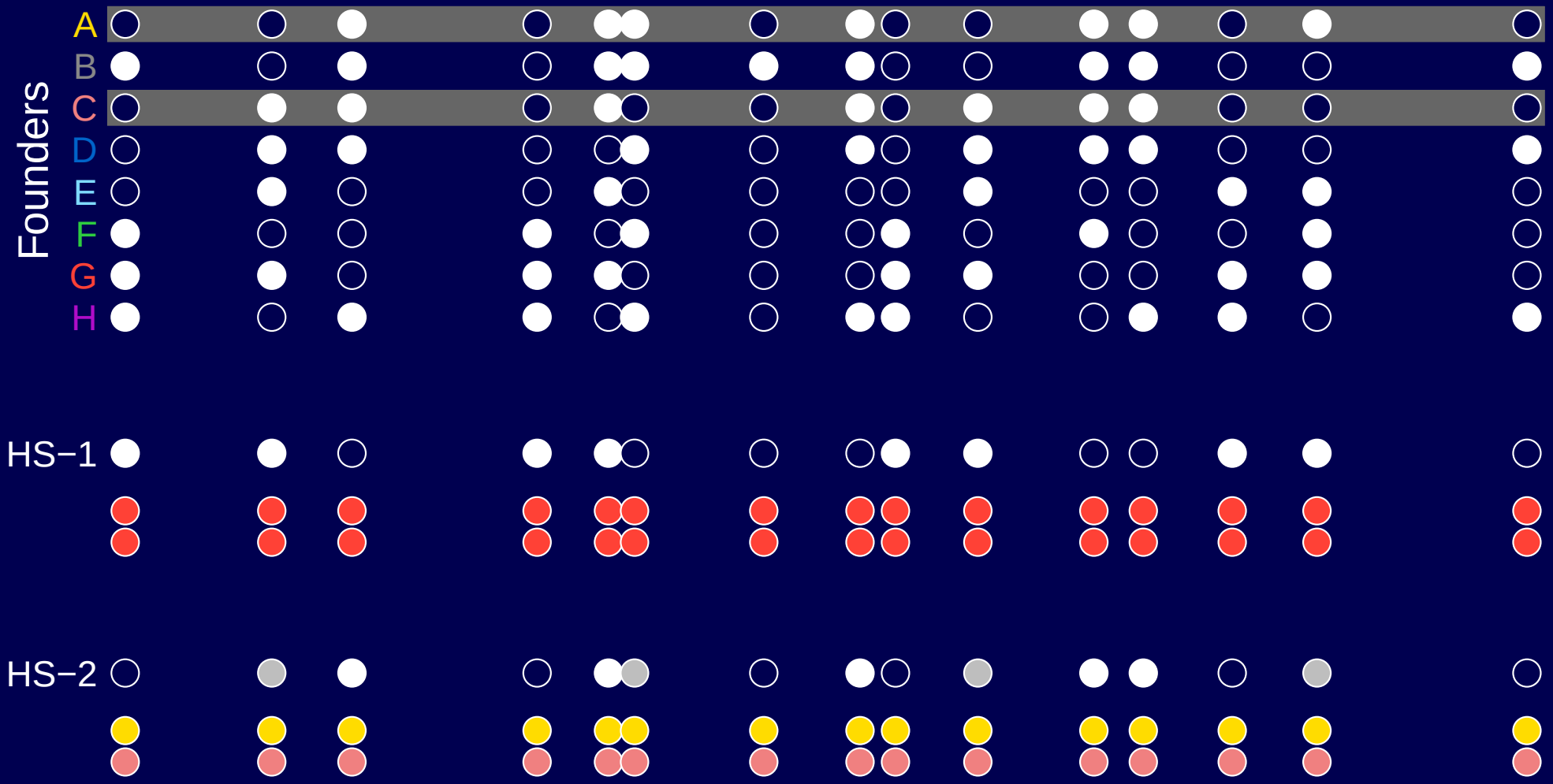
HS genotype reconstruction



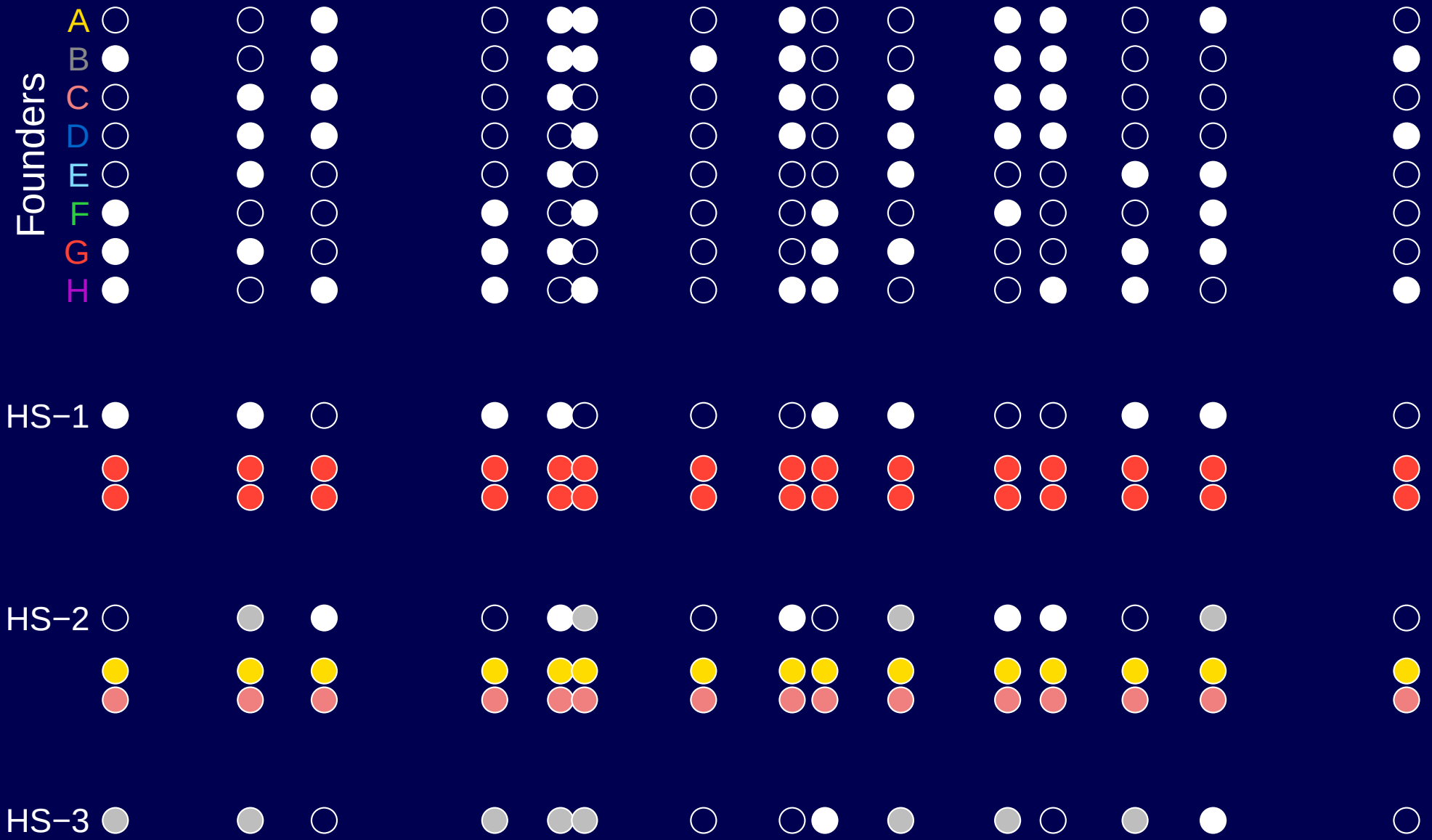
HS genotype reconstruction



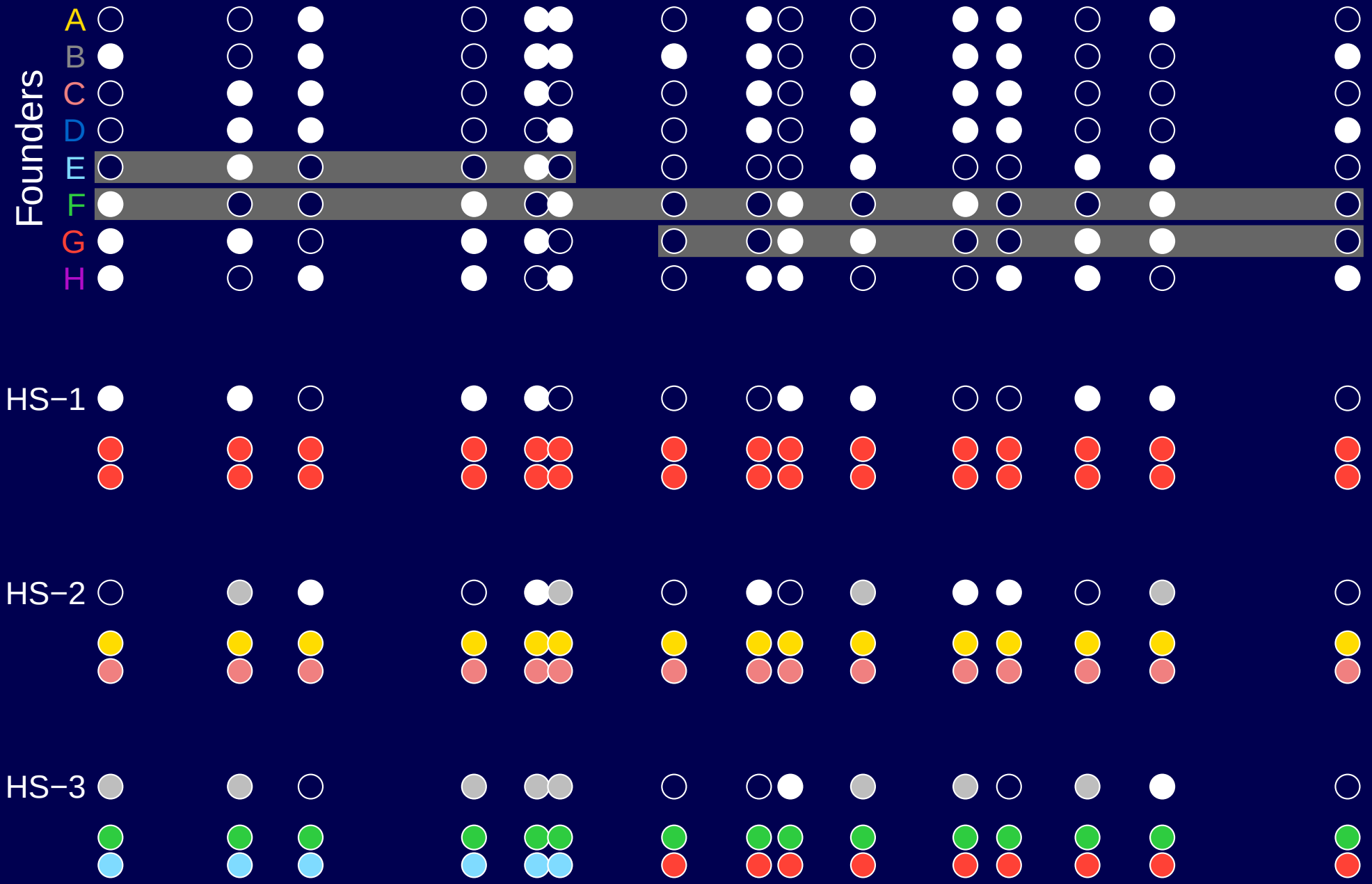
HS genotype reconstruction



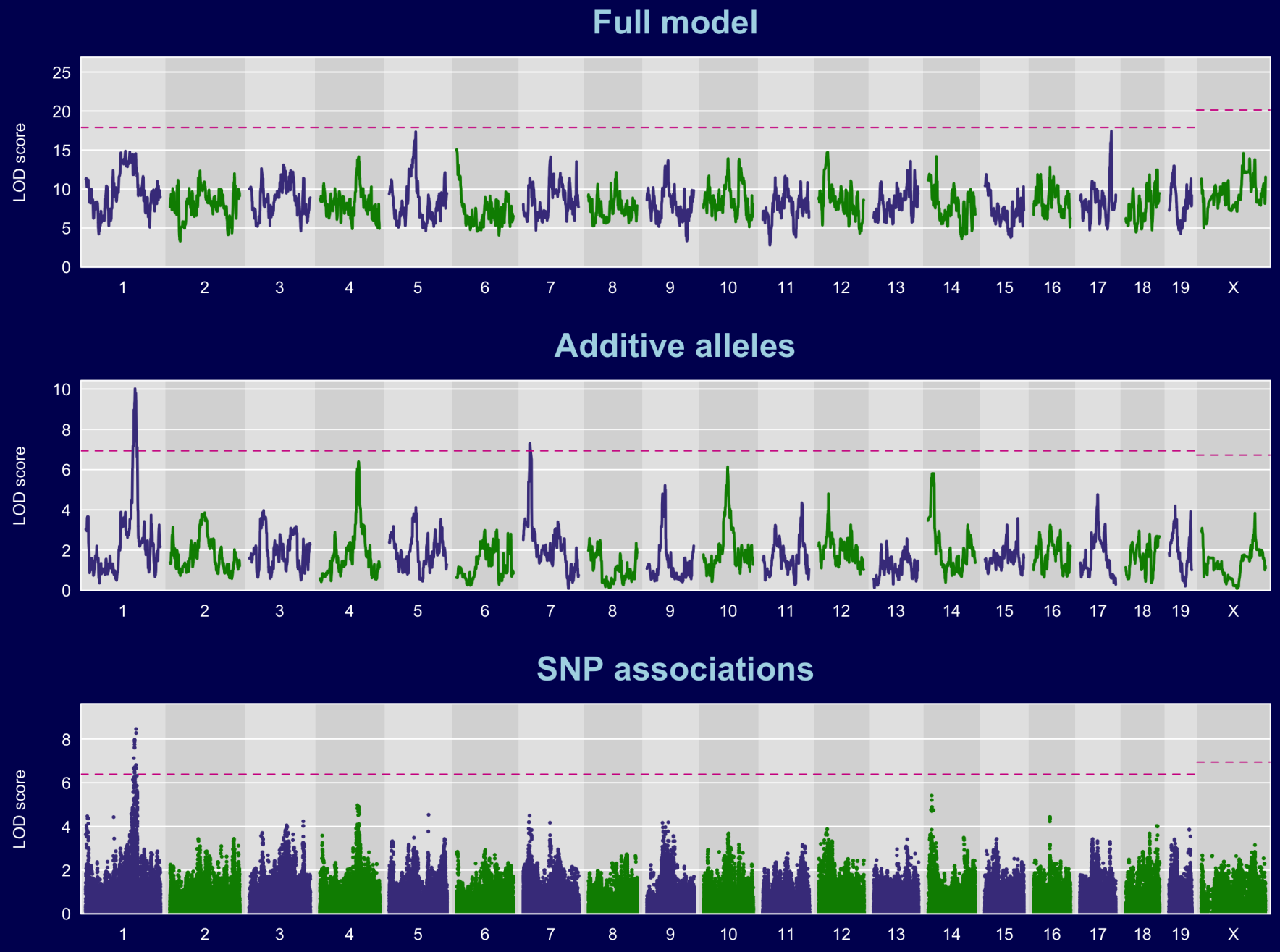
HS genotype reconstruction



HS genotype reconstruction



HS genome scans



Why R/qtl2?

- High-dimensional data
  - genotypes and phenotypes
- More diverse crosses
  - especially multi-parent populations
- Linear mixed models
  - especially in HS/AIL

R/qtl → R/qtl2

- See `kbroman.org/qtl2/assets/vignettes/rqtl_diff.html`
- New data file formats
- New data structures
- New function names

  `read.cross()` → `read_cross2()`

  `calc.genoprob()` → `calc_genoprob()`

  `scanone()` → `scan1()`
- Different treatment of intermediate calculations
- Use of individual IDs for aligning data
- Order of args when subsetting cross objects

  `cross[chr,ind]` → `cross2[ind,chr]`