QTL mapping
in MAGIC populations
with R/qtl2

Karl Broman

Biostatistics & Medical Informatics, UW–Madison

kbroman.org
github.com/kbroman
@kwbroman
Slides: bit.ly/msu2019-12
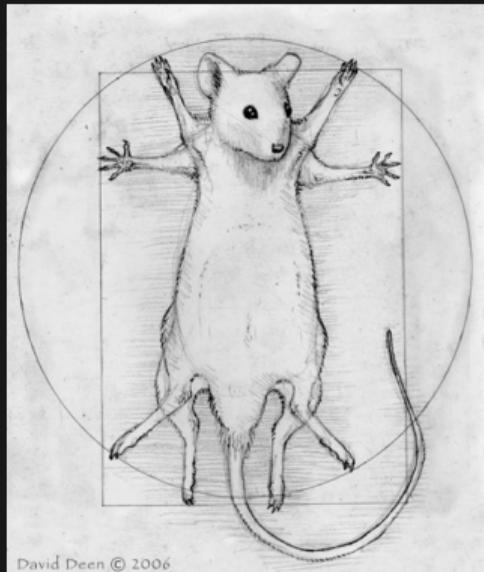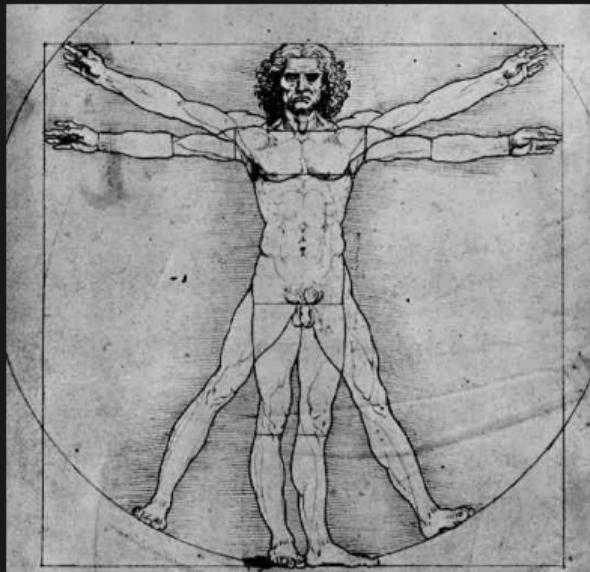
daviddeen.com

Intercross

QTL mapping

5
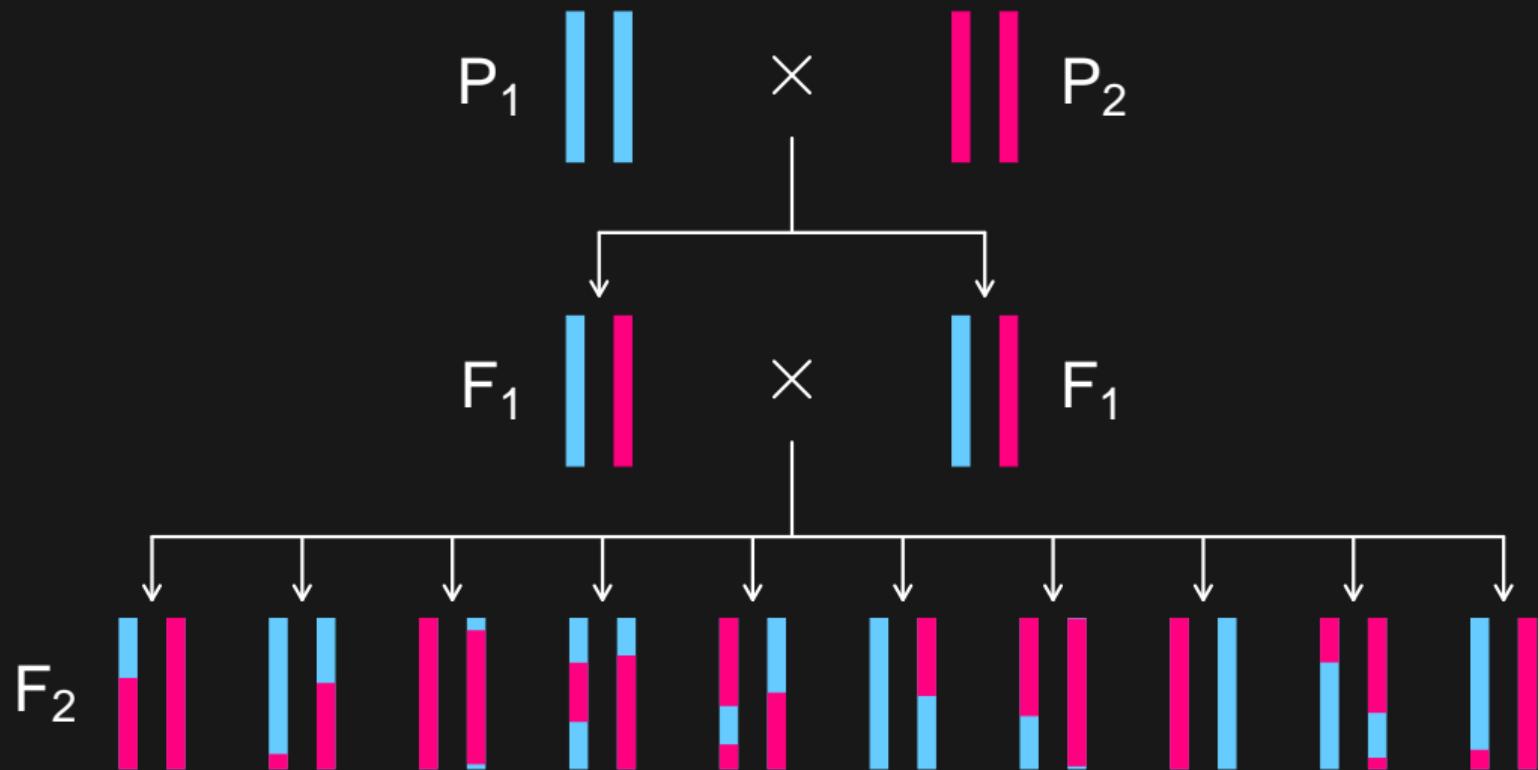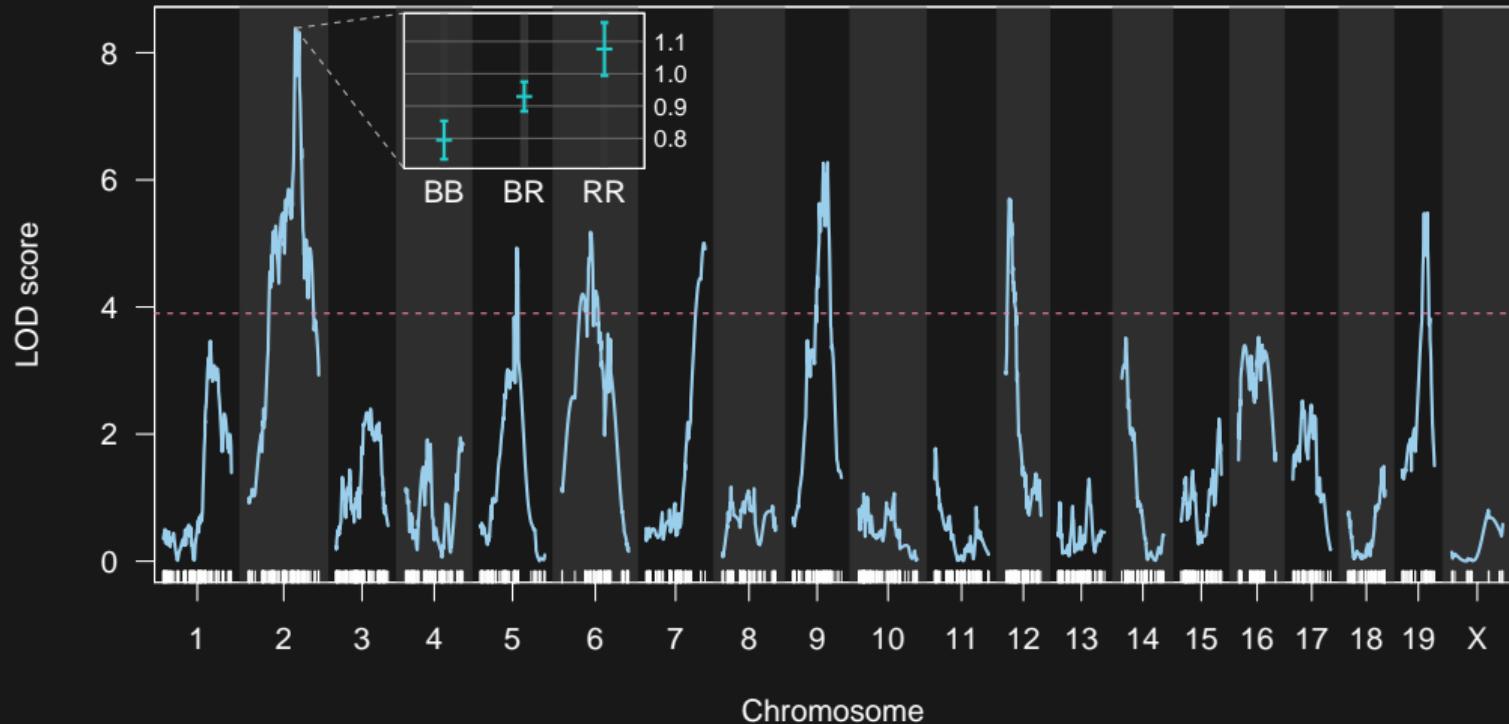
Congenic line/NIL
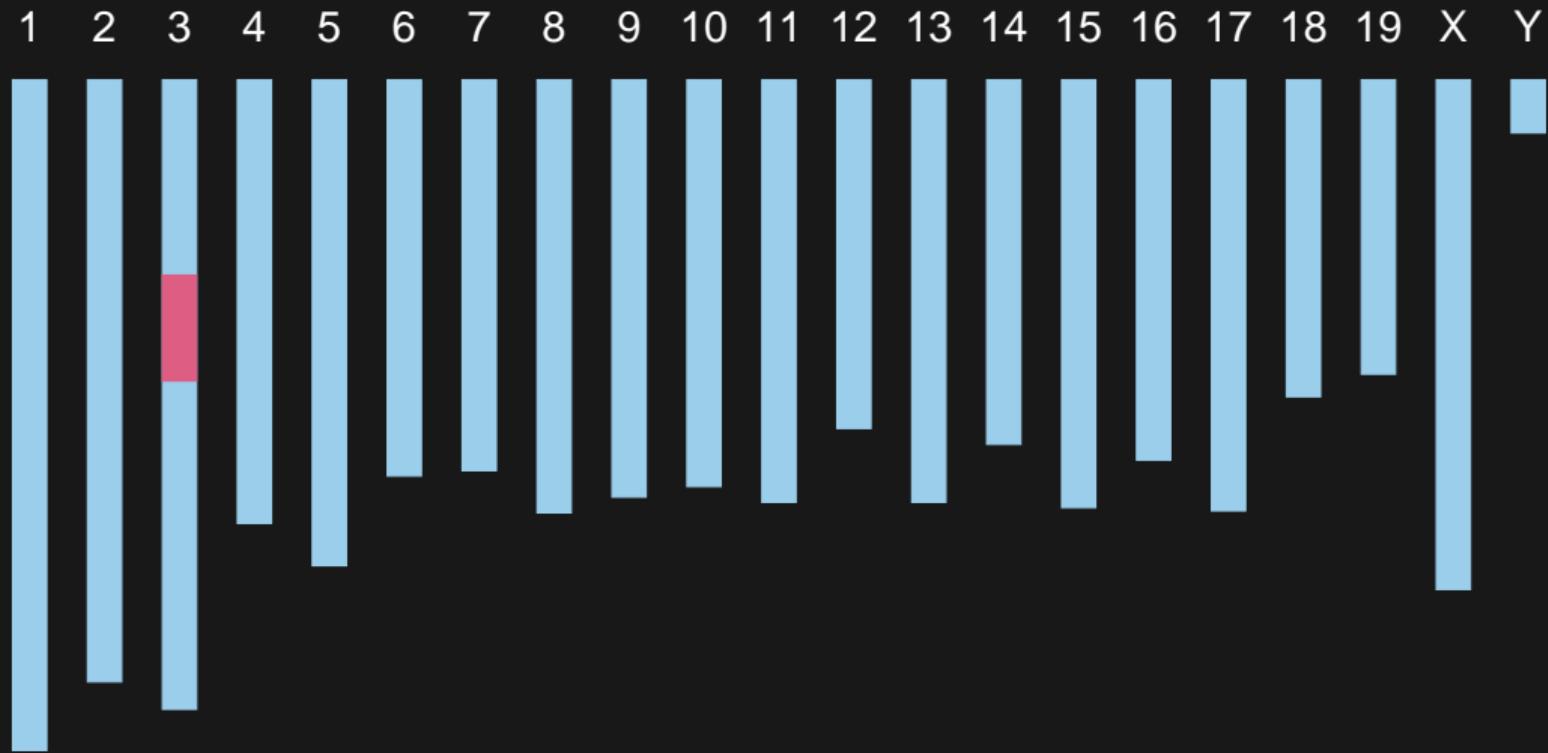
Improving precision

- ▶ more recombinations
- ▶ more individuals
- ▶ more precise phenotype
- ▶ lower-level phenotypes
  – transcripts, proteins, metabolites

Genome-scale phenotypes

Alan Attie

Advanced intercross lines
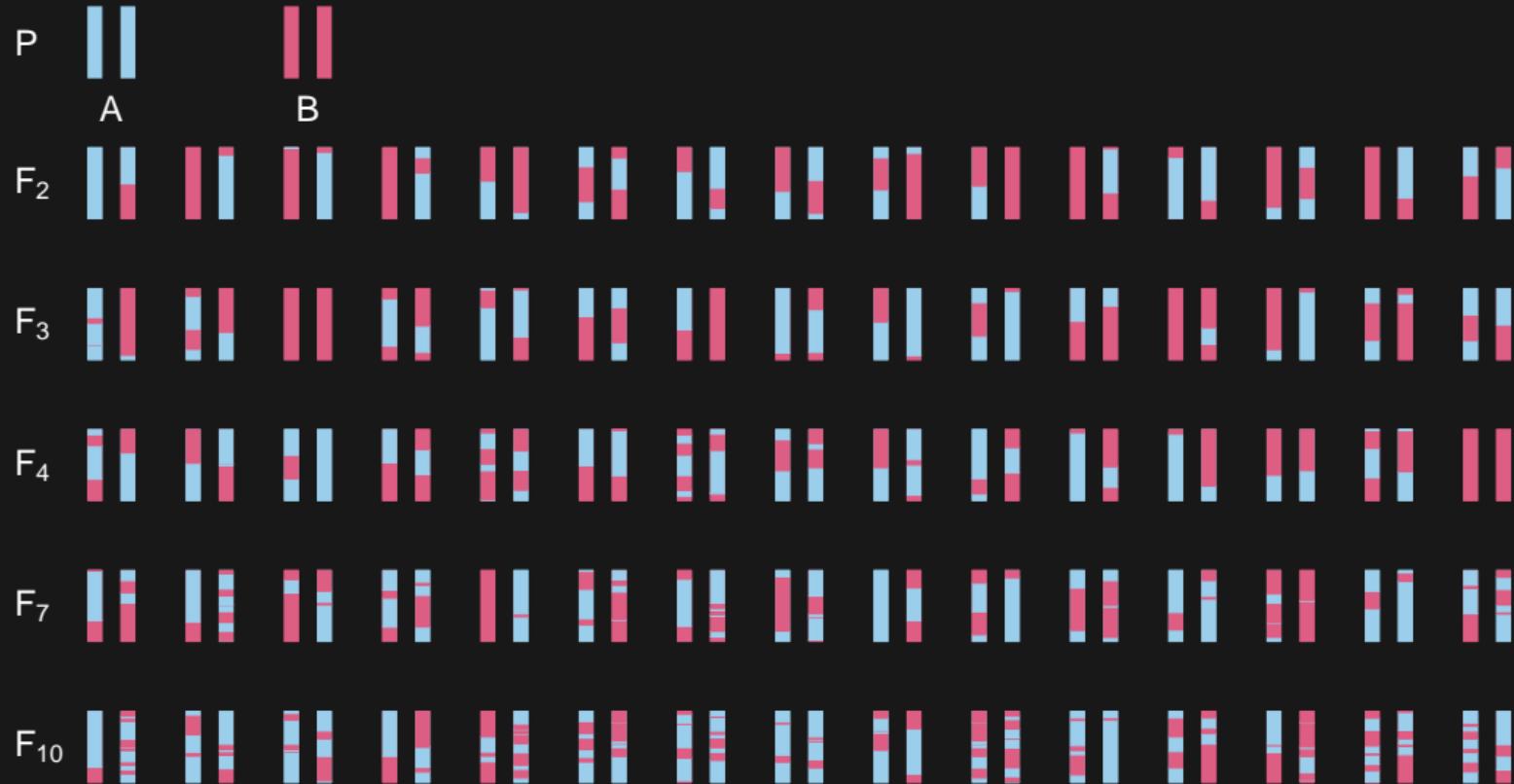
Recombinant inbred lines
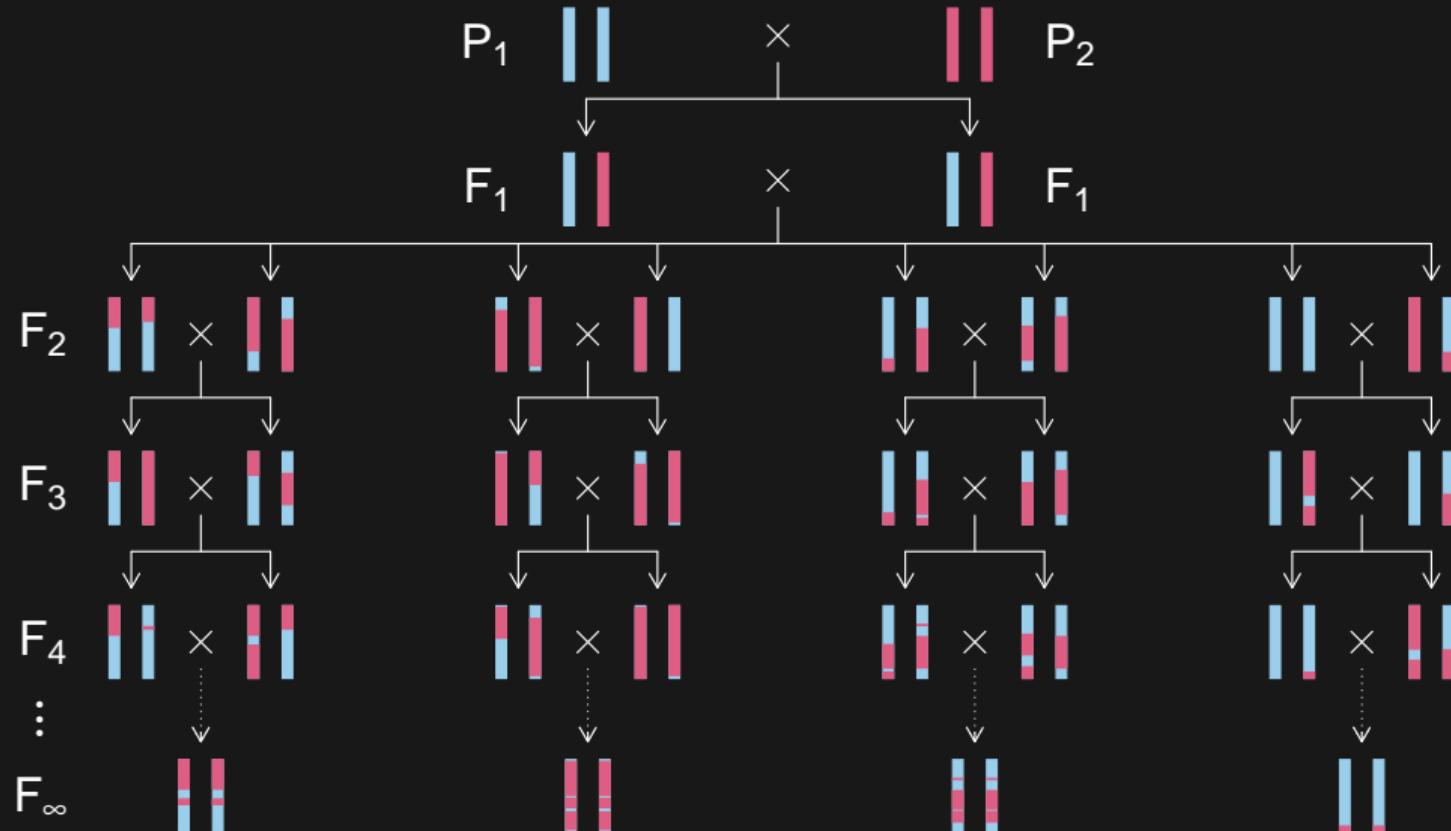
Recombinant inbred lines
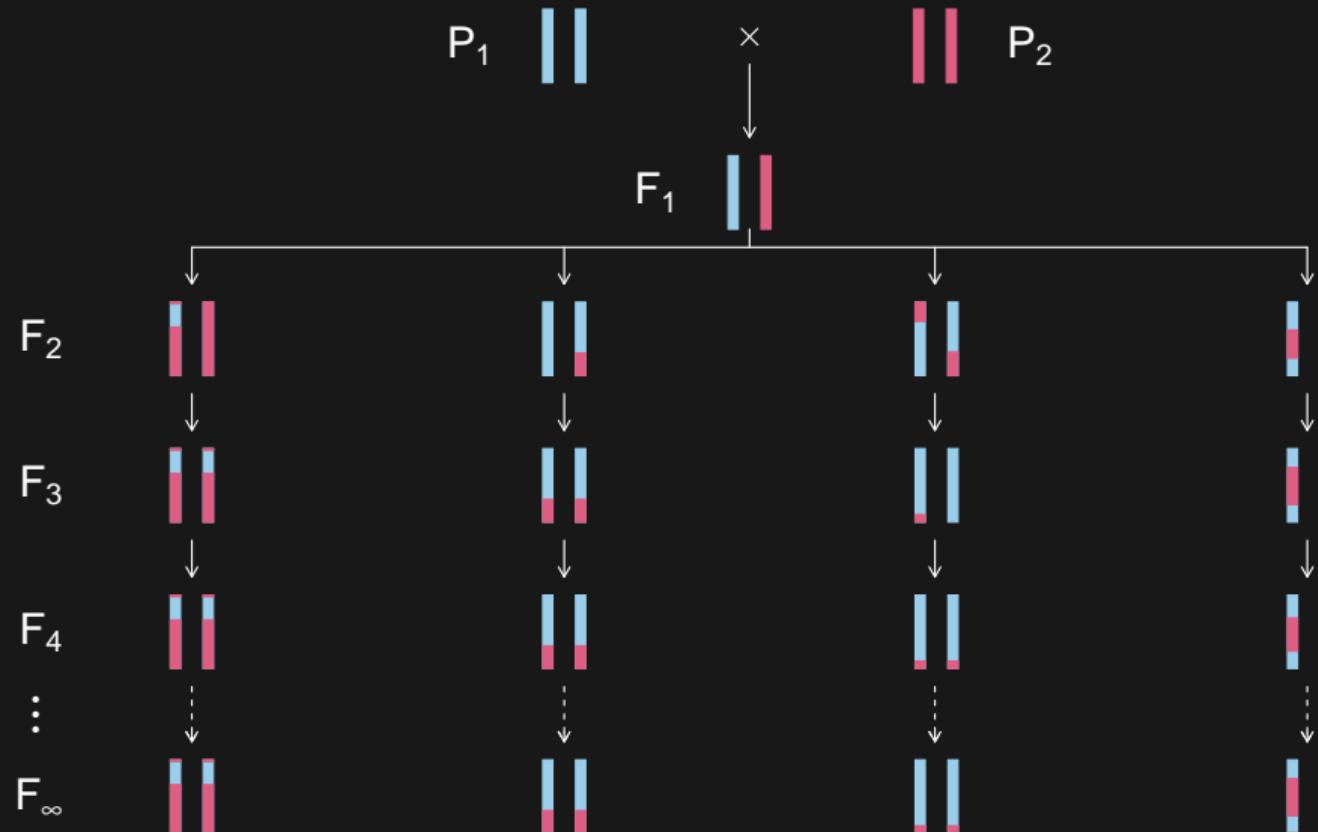
Collaborative Cross
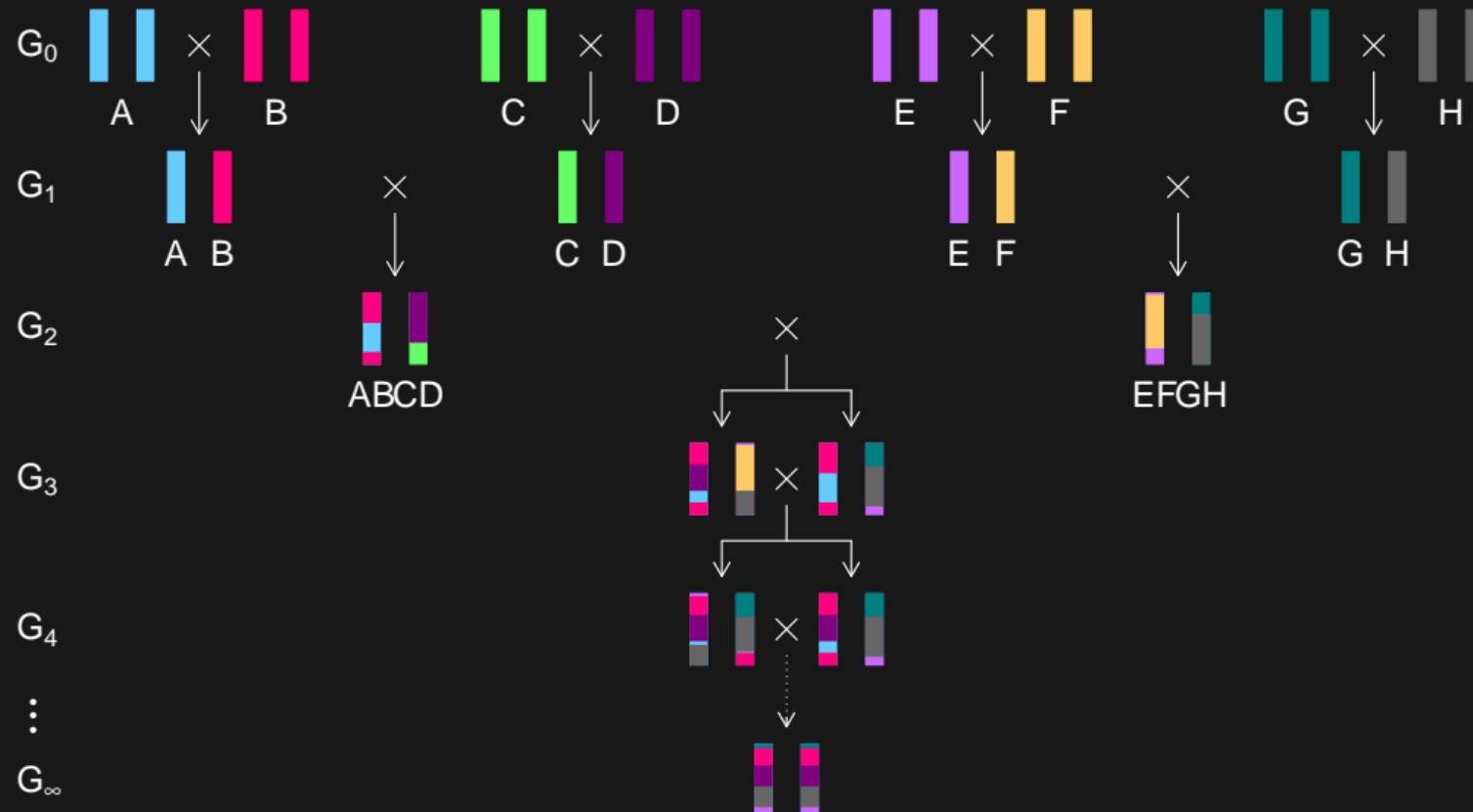
MAGIC
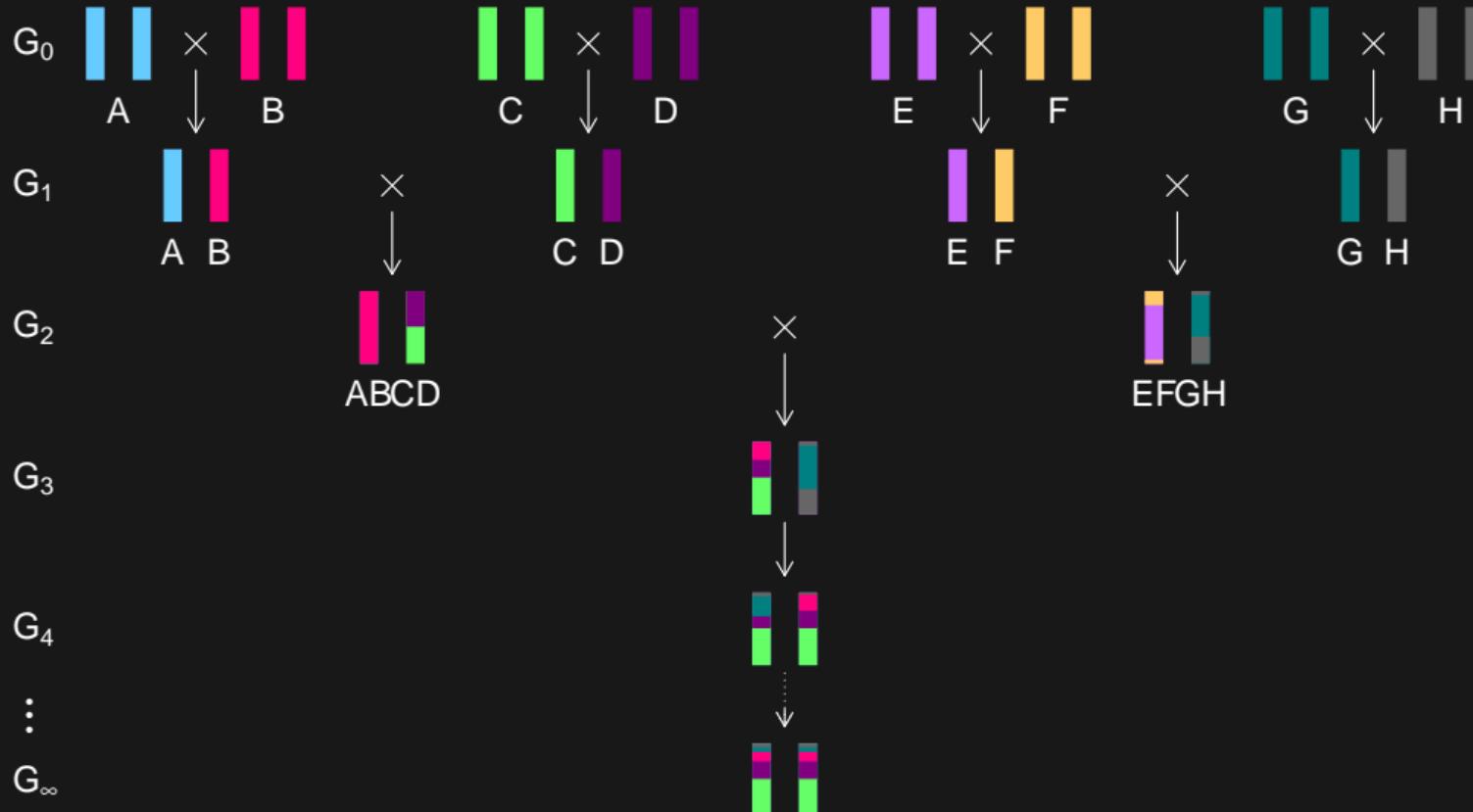
Heterogeneous stock
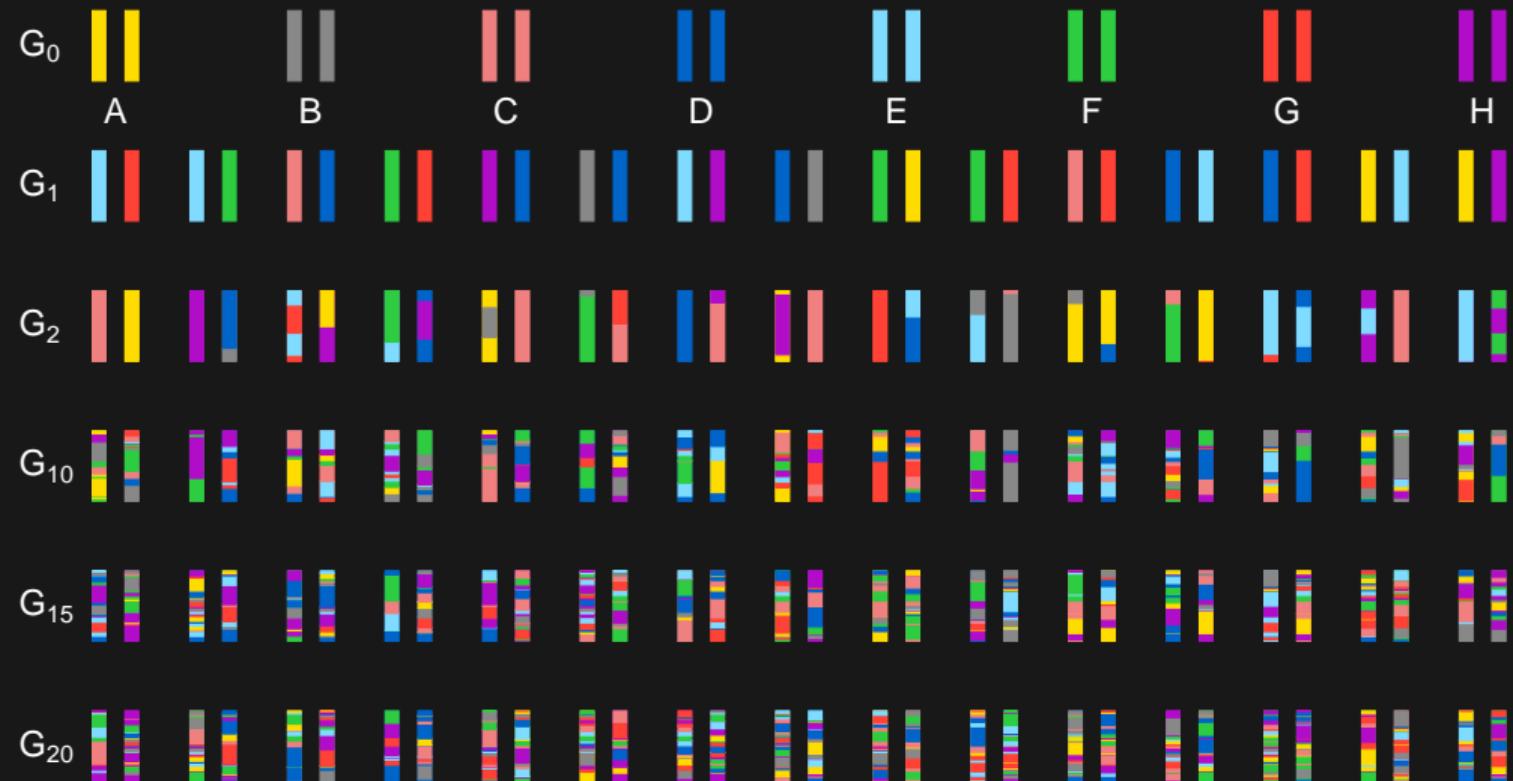
MAGIC is magic

- ► Genetic diversity

- ► High-precision mapping

- ► Predictable linkage disequilibrium

- ► Phenotype replicates to reduce individual variation

- ► Pool phenotypes from multiple labs, environments, treatments

- ► Genotype once

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
- ▶ Cool name

MAGIC lines

founders   foundation   outbred   recombinant inbred

mixing   maintenance   inbreeding

Valdar et al., Genetics 172:1783, 2006

MAGIC lines

Valdar et al., Genetics 172:1783, 2006

MAGIC lines

founders   foundation   outbred   recombinant inbred

mixing   maintenance   inbreeding

combine   mix   fix

How many?

MAGIC lines

founders   foundation   outbred   recombinant inbred

mixing   maintenance   inbreeding
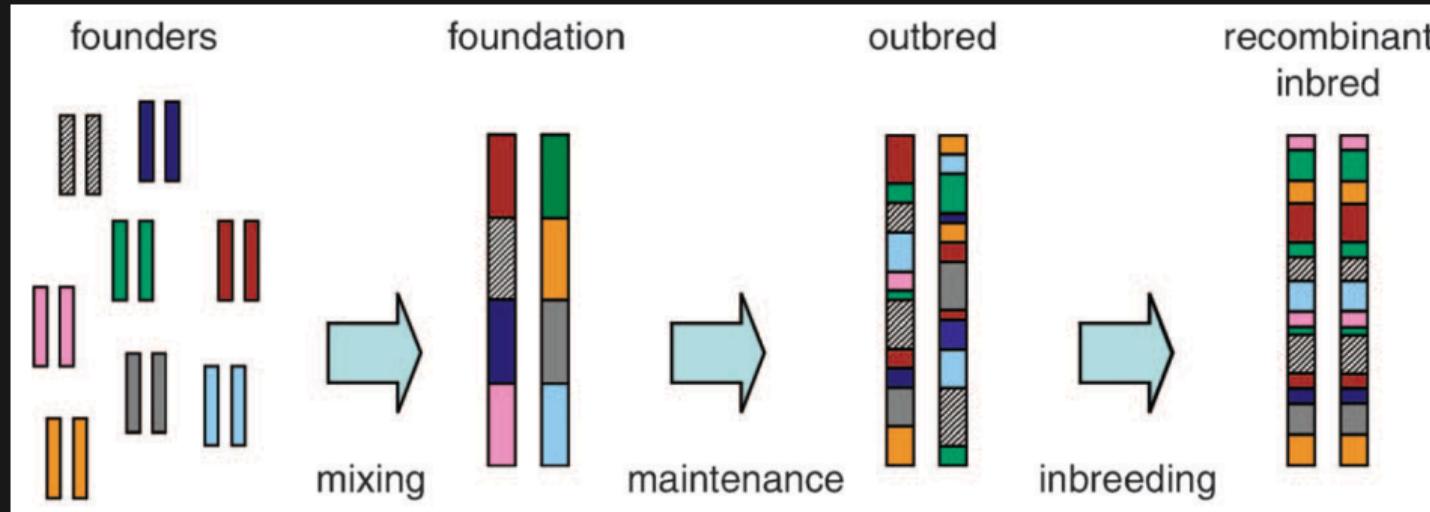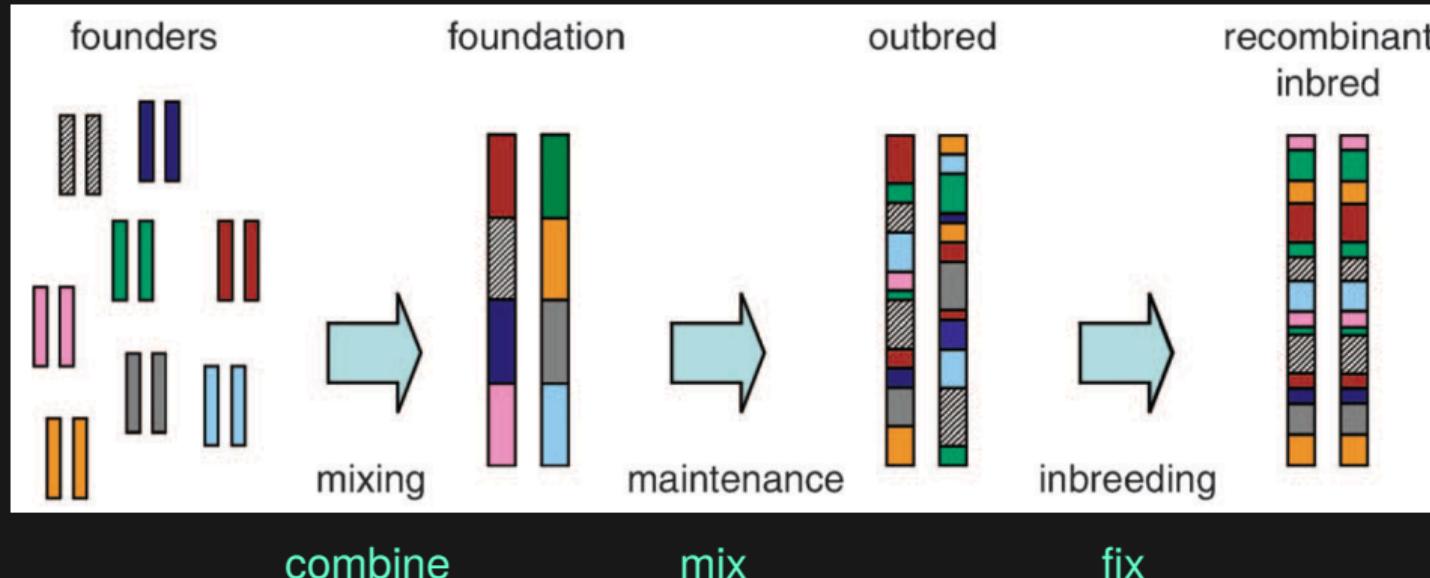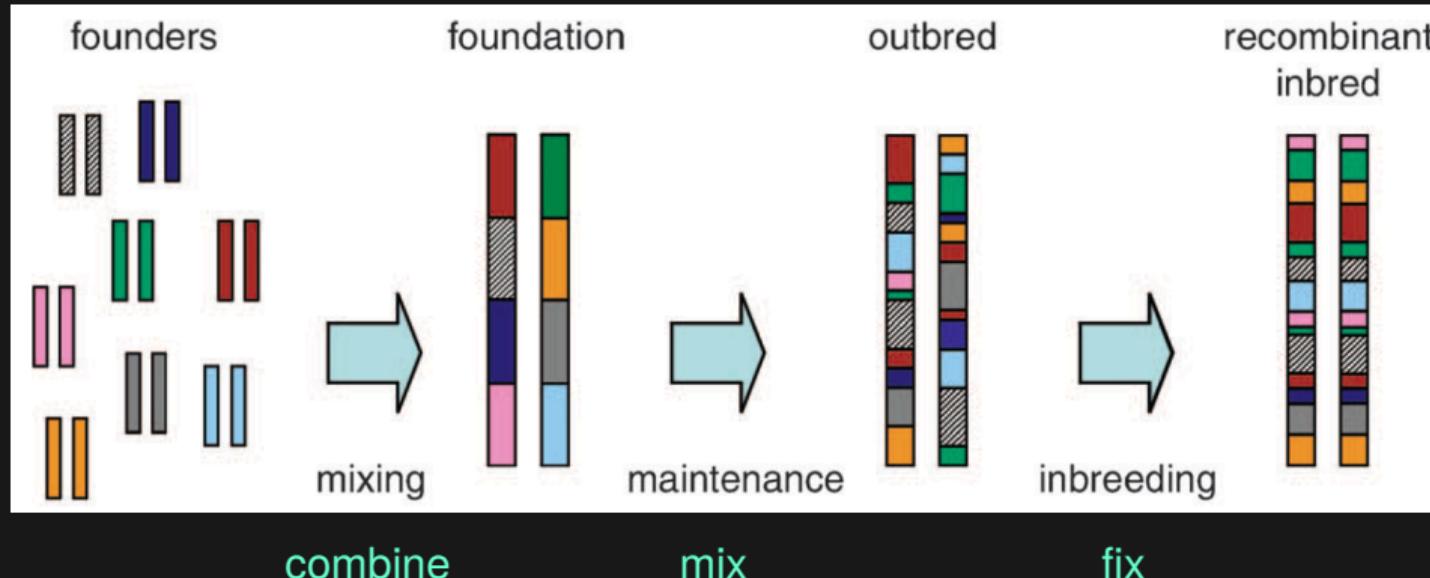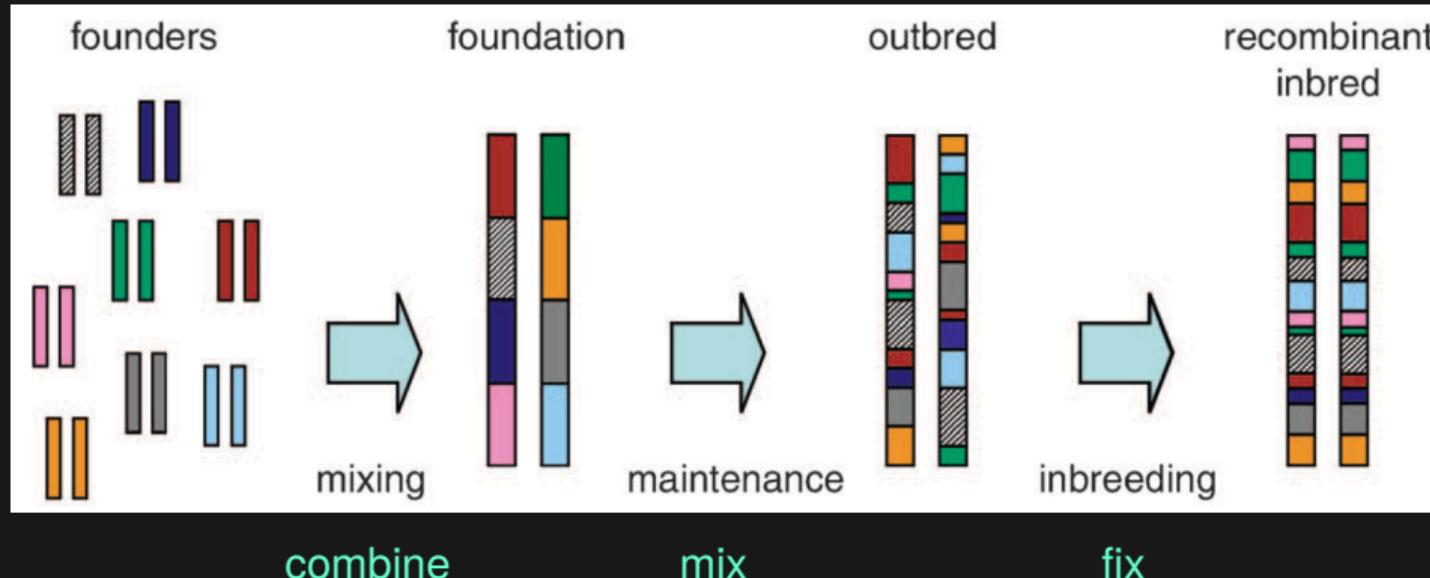
combine   mix   fix

How many?

Which?

Valdar et al., Genetics 172:1783, 2006
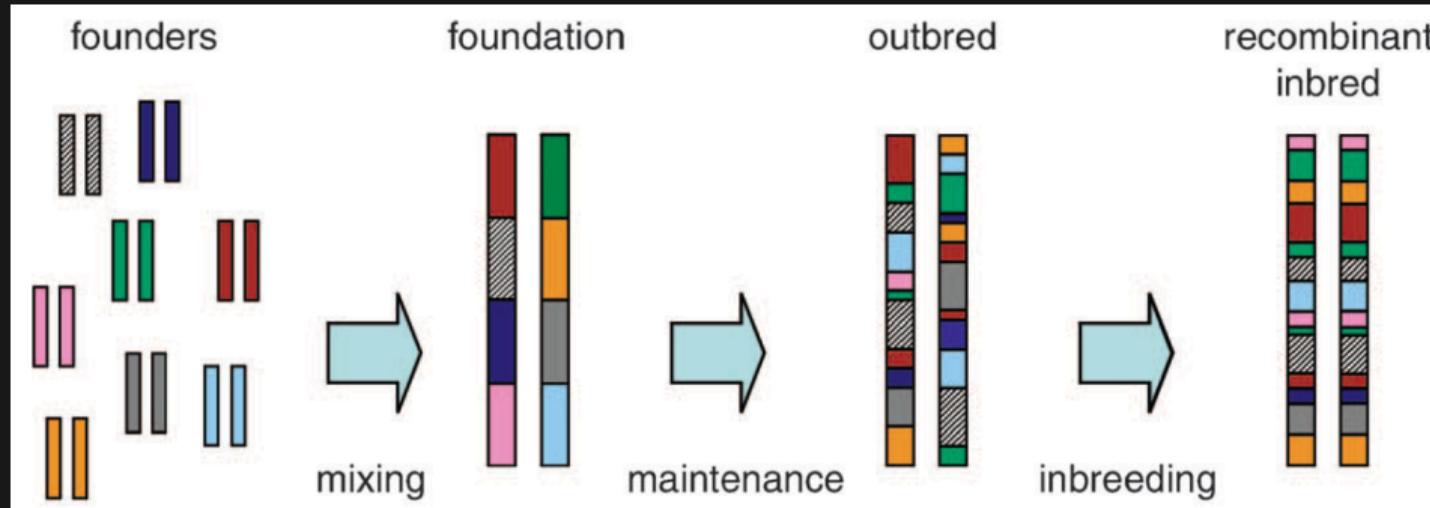
MAGIC lines

founders   foundation   outbred   recombinant inbred

mixing   maintenance   inbreeding

combine   mix   fix

How many?   How long?

Which?

Valdar et al., Genetics 172:1783, 2006

MAGIC lines

founders　　foundation　　outbred　　recombinant inbred

mixing　　maintenance　　inbreeding
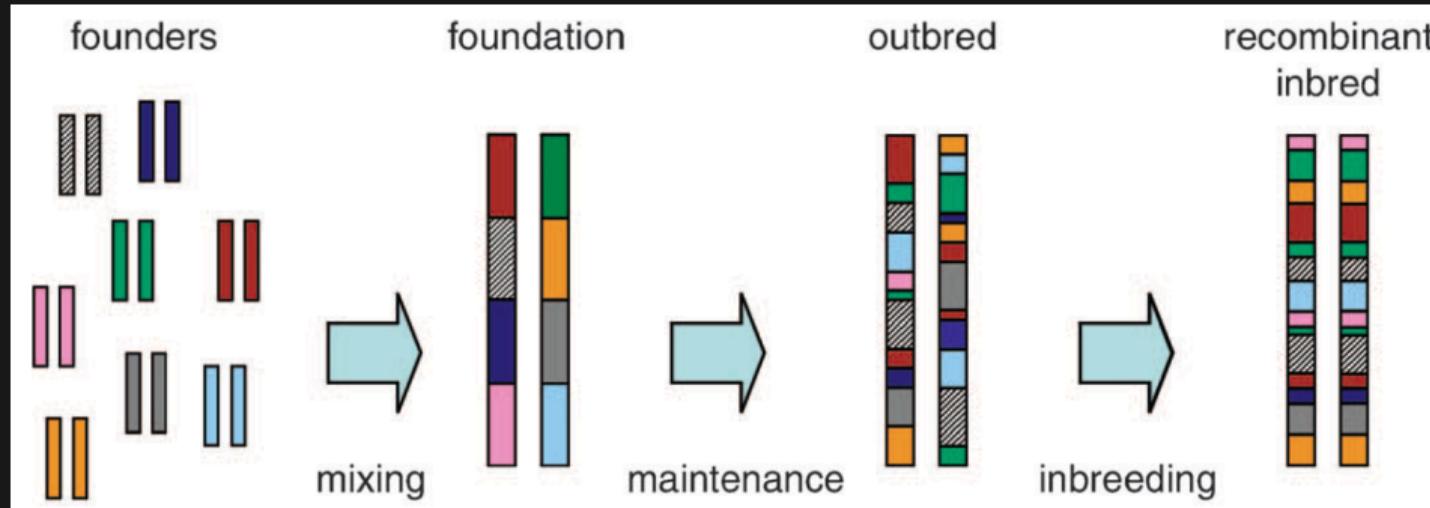
combine　　mix　　fix

How many?　　How long?　　How?

Which?

The goal

Identify QTL

- ▶ Power
- ▶ Mapping precision

The goal

Identify QTG

- ▶ Power
- ▶ Mapping precision

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

Sharing is also key

- ▶ The greatest power of MAGIC comes from sharing
  Pooling data, exploring multiple environments/treatments

- ▶ Common software needs
  Analysis software, database infrastructure

- ▶ Many students need to learn the same stuff
  Joint training opportunities

19 years of R/qtl

23
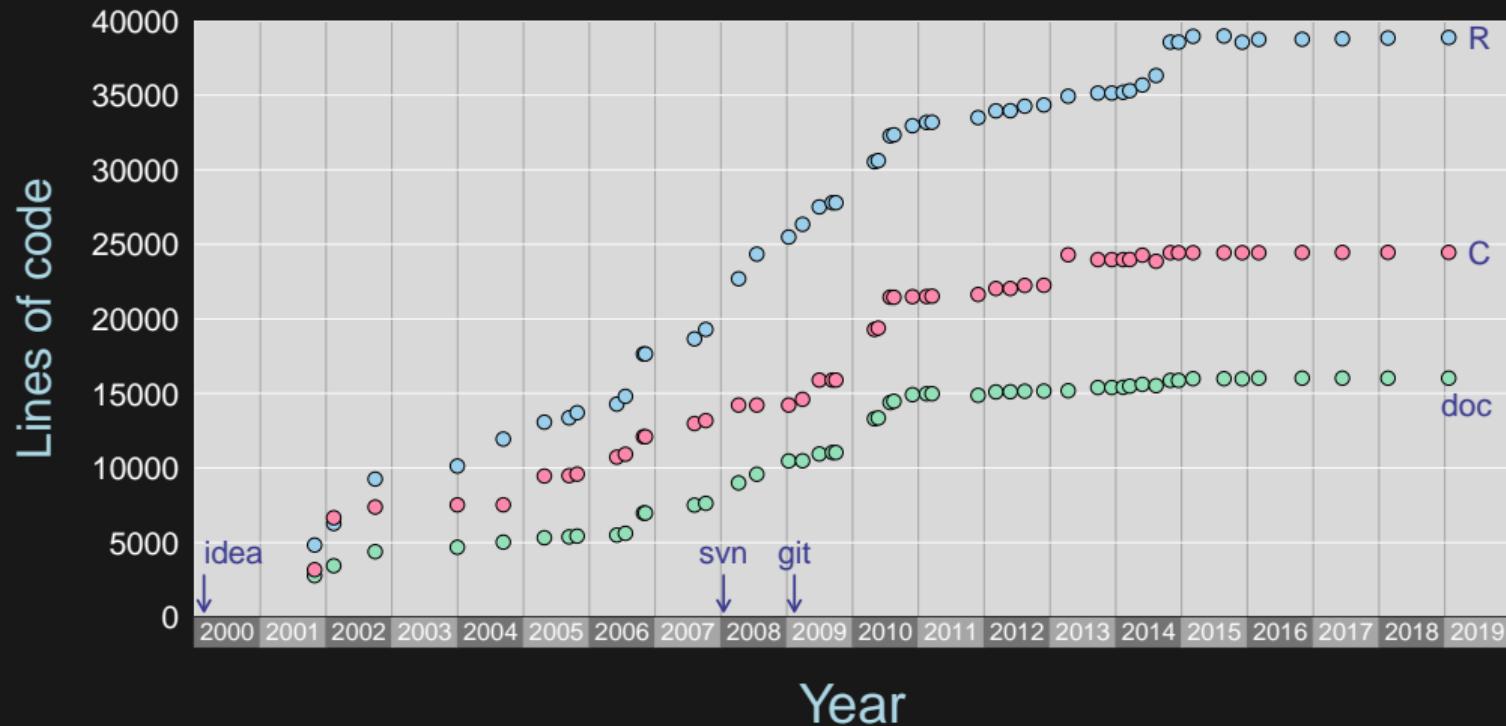
R/qtl cross types

- ▶ **backcross**, doubled haploids, haploid
- ▶ intercross
- ▶ **2-way RIL** by selfing or sibling mating
- ▶ phase-known 4-way cross

R/qtl2 cross types

- backcross, doubled haploids, haploid
- intercross
- 2-, 4-, 8-, 16-way RIL by selfing
- 2-, 4-, 8-way RIL by sibling mating
- 2-, 3-, 8-way advanced intercross
- 6- and 19-way MAGIC
- Diversity Outbred (DO) mice
- F_1 of DO $\times$ inbred
- general RIL or AIL

Data files

	A	B	C	D	E	F	G
1	id	bolting_days	seed_weight	seed_area	ttl_seedspfruit	branches	height
2	MAGIC.1	15.33	17.15	0.64	45.11	10.5	NA
3	MAGIC.2	22	22.71	0.75	49.11	4.33	42.33
4	MAGIC.3	23	21.03	0.68	57	4.67	50
5	MAGIC.4	18.67	22.45	0.74	54.33	6.33	NA
6	MAGIC.5	18.67	25.36	0.82	38.33	5.67	42.25
7	MAGIC.6	25	21.53	0.71	52	4.33	NA
8	MAGIC.7	15.33	20.92	0.71	39	4	37.35
9	MAGIC.8	14.33	24.2	0.79	50.56	7.33	43.23
10	MAGIC.9	16.33	18.86	0.63	75.78	5.33	42.4
11	MAGIC.10	30	30.46	0.91	56.33	3	48.5
12	MAGIC.11	14	20.58	0.66	41.56	5.67	36.25
13	MAGIC.12	21.33	19.05	0.67	52.33	4.67	49.35
14	MAGIC.14	18.67	21.14	0.72	49.78	6	48.5
15	MAGIC.15	15	21.71	0.73	49	6.33	41.9
16	MAGIC.16	18.33	21.37	0.71	66.67	3.33	18.17

Data files

	A	B	C	D	E	F	G
1	id	bolting_days	seed_weight	seed_area	ttl_seedspfruit	branches	height
2							

	A	B	C	D	E	F	G
1	marker	MAGIC.1	MAGIC.10	MAGIC.100	MAGIC.101	MAGIC.102	MAGIC.103
2	MN1_29291	A	A	A	B	B	A
3	MN1_29716	A	A	A	B	B	A
4	MN1_112907	A	A	A	B	B	A
5	MASC03771	A	B	A	A	B	A
6	MN1_197787	A	B	A	B	B	A
7	MN1_340810	A	B	A	A	A	A
8	MN1_395107	A	A	A	A	B	A
9	MN1_444820	A	A	–	A	A	A
10	MN1_494205	A	A	A	A	A	A
11	MN1_592863	A	B	A	A	A	A
12	MN1_592760	A	A	A	A	A	A
13	BKN118	A	A	A	A	A	A
14	MN1_1042427	A	B	B	A	B	B
15	CRY2_1021	A	A	A	A	A	A
16	CRY2_429	A	A	A	A	B	A

27

Data files

	A	B	C	D	E	F	G
1	id	bolting_days	seed_weight	seed_area	ttl_seedspfruit	branches	height
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							

	A	B	C	D	E	F	G
1	marker	MAGIC.1	MAGIC.10	MAGIC.100	MAGIC.101	MAGIC.102	MAGIC.103
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							

	A marker	B Bur	C Can	D Col	E Ct	F Edi	G Hi	H Kn	I Ler	J Mt	K No	L Oy	M Po	N Rsch	O Sf	P Tsu	Q Wil	R Ws	S Wu	T Zu
2	MN1_29291	B	A	A	B	A	A	B	A	A	A	B	B	A	B	A	A	A	A	A
3	MN1_29716	B	A	A	A	A	A	B	A	A	A	A	B	B	A	B	A	A	A	A
4	MN1_112907	B	A	A	A	B	B	B	B	A	A	B	B	A	B	B	A	A	A	A
5	MASC03771	A	B	A	A	B	A	B	A	A	A	A	A	A	A	A	A	A	A	A
6	MN1_197787	B	B	A	A	A	B	B	A	A	A	A	B	A	A	A	A	A	B	B
7	MN1_340810	B	B	A	B	A	B	A	A	A	A	A	A	A	A	A	B	B	A	A
8	MN1_395107	A	A	A	A	A	A	A	B	A	A	A	A	A	A	A	A	A	A	A
9	MN1_444820	B	A	A	B	B	A	A	A	A	A	A	A	B	A	B	B	B	A	A
10	MN1_494205	B	A	A	A	B	A	A	A	A	A	A	A	A	A	A	A	B	A	A
11	MN1_592863	A	A	A	B	A	A	A	B	A	A	A	A	B	A	B	A	B	A	A
12	MN1_592760	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
13	BKN118	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	A	A	A
14	MN1_1042427	A	B	A	B	A	B	B	A	A	B	A	A	A	B	A	A	B	A	A
15	CRY2_1021	A	A	A	A	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A
16	CRY2_429	A	A	A	A	A	B	B	B	A	B	A	A	A	A	A	A	B	A	A
17	MASC07014	A	A	B	A	A	A	A	A	A	A	A	A	B	A	A	B	A	A	A
18	MASC03609	A	A	A	A	A	A	B	B	B	B	A	A	A	A	A	A	A	A	A
19	MN1_1296068	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	A	B	B	A
20	MN1_1399466	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	–	B	A	A
21	AXR1_381	A	B	A	A	A	B	A	A	A	B	A	A	A	A	A	A	B	A	B
22	MASC07424	A	B	B	A	A	A	A	A	A	A	A	A	B	–	A	A	A	A	A

Spreadsheet 1:

	A	B	C	D	E	F	G
1	id	bolting_days	seed_weight	seed_area	ttl_seedspfruit	branches	height

Spreadsheet 2:

	A	B	C	D	E	F	G
3 (1)	marker	MAGIC.1	MAGIC.10	MAGIC.100	MAGIC.101	MAGIC.102	MAGIC.103

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

Spreadsheet 3:

	A	B	C
1	marker	chr	pos
2	MN1_29291	1	0.029291
3	MN1_29716	1	0.029757
4	MN1_112907	1	0.112907
5	MASC03771	1	0.174605
6	MN1_197787	1	0.197787
7	MN1_340810	1	0.34081
8	MN1_395107	1	0.395107
9	MN1_444820	1	0.444764
10	MN1_494205	1	0.494205
11	MN1_592863	1	0.592867
12	MN1_592760	1	0.592984
13	BKN118	1	0.761584
14	MN1_1042427	1	1.042428
15	CRY2_1021	1	1.187841
16	CRY2_429	1	1.188433
17	MASC07014	1	1.189374
18	MASC03609	1	1.22655

Spreadsheet 4 (genotype matrix):

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	marker	Bur	Can	Col	Ct	Edi	Hi	Kn	Ler	Mt	No	Ov	Po	Rsch	Sf	Tsu	Wil	Ws	Wu	Zu
2	MN1_292...												B	A	B	A	A	A	A	A
3	MN1_297...												B	A	B	A	A	A	A	A
4	MN1_112...												B	A	B	B	A	A	A	A
5	MASC03...												A	A	A	A	A	A	A	A
6	MN1_197...												A	A	B	A	A	A	B	B
7	MN1_340...												A	A	A	A	A	B	B	A
8	MN1_395...												A	A	A	B	A	B	B	A
9	MN1_444...												A	A	A	A	A	A	B	A
10	MN1_494...												A	A	A	B	A	B	B	A
11	MN1_592...												A	B	A	A	B	A	B	A
12	MN1_592...												A	A	A	A	A	A	B	A
13	BKN11...												A	B	A	A	B	A	A	A
14	MN1_1042...												A	A	A	A	A	B	A	A
15	CRY2_10...												A	A	A	–	A	B	A	B
16	CRY2_4...												A	A	B	–	A	A	A	A
17	MASC070...																			
18	MASC036...																			
19	MN1_1296...																			
20	MN1_1399...																			
21	AXR1_3...																			
22	MASC074...																			

27

Control file (json or yaml)

```json
{
  "description": "Arabidopsis MAGIC data, Gnan et al (2014)",
  "crosstype": "magic19",
  "sep": ",",
  "na.strings": ["-", "NA"],
  "comment.char": "#",
  "geno": "arabmagic_geno.csv",
  "founder_geno": "arabmagic_foundergeno.csv",
  "gmap": "arabmagic_pmap_tair9.csv",
  "pmap": "arabmagic_pmap_tair9.csv",
  "pheno": "arabmagic_pheno.csv",
  "genotypes":
    "A": 1
    "H": 2
    "B": 3
  },
  "geno_transposed": true,
  "founder_geno_transposed": true
}
```

Control file (json or yaml)

```
{
  "description": "Arabidopsis MAGIC data, Gnan et al (2014)",
  "crosstype": "magic19",
  "sep": ",",
  "na.strings": ["-", "NA"],
  "comment.char": "#",
  "geno": "arabmagic_geno.csv",
  "founder_geno": "arabmagic_foundergeno.csv",
  "gmap": "arabmagic_pmap_tair9.csv",
  "pmap": "arabmagic_pmap_tair9.csv",
  "pheno": "arabmagic_pheno.csv",
  "genotypes":
    "A": 1
    "H": 2
    "B": 3
  },
  "geno_transposed": true,
  "founder_geno_transposed": true
}
```

Control file (json or yaml)

```json
{
  "description": "Arabidopsis MAGIC data, Gnan et al (2014)",
  "crosstype": "magic19",
  "sep": ",",
  "na.strings": ["-", "NA"],
  "comment.char": "#",
  "geno": "arabmagic_geno.csv",
  "founder_geno": "arabmagic_foundergeno.csv",
  "gmap": "arabmagic_pmap_tair9.csv",
  "pmap": "arabmagic_pmap_tair9.csv",
  "pheno": "arabmagic_pheno.csv",
  "genotypes":
    "A": 1
    "H": 2
    "B": 3
  },
  "geno_transposed": true,
  "founder_geno_transposed": true
}
```

Control file (json or yaml)

```json
{
  "description": "Arabidopsis MAGIC data, Gnan et al (2014)",
  "crosstype": "magic19",
  "sep": ",",
  "na.strings": ["-", "NA"],
  "comment.char": "#",
  "geno": "arabmagic_geno.csv",
  "founder_geno": "arabmagic_foundergeno.csv",
  "gmap": "arabmagic_pmap_tair9.csv",
  "pmap": "arabmagic_pmap_tair9.csv",
  "pheno": "arabmagic_pheno.csv",
  "genotypes":
    "A": 1
    "H": 2
    "B": 3
  },
  "geno_transposed": true,
  "founder_geno_transposed": true
}
```

Control file (json or yaml)

```json
{
  "description": "Arabidopsis MAGIC data, Gnan et al (2014)",
  "crosstype": "magic19",
  "sep": ",",
  "na.strings": ["-", "NA"],
  "comment.char": "#",
  "geno": "arabmagic_geno.csv",
  "founder_geno": "arabmagic_foundergeno.csv",
  "gmap": "arabmagic_pmap_tair9.csv",
  "pmap": "arabmagic_pmap_tair9.csv",
  "pheno": "arabmagic_pheno.csv",
  "genotypes":
    "A": 1
    "H": 2
    "B": 3
  },
  "geno_transposed": true,
  "founder_geno_transposed": true
}
```

Reading data into R

```
library(qtl2)
arab <- read_cross2("arab_magic.json")
```

Reading data into R

```
library(qtl2)
arab <- read_cross2("arab_magic.json")
```

19-way Arabidopsis MAGIC
Kover et al. (2009) PLoS Genet
Gnan et al. (2014) Genetics
github.com/rqtl/qtl2data

Data diagnostics

See Broman et al. (2019) Cleaning genotype data from
Diversity Outbred mice. G3 9:1571–1579

doi: 10.1534/g3.119.400165

Genotype reconstruction

Chr 3 position (Mbp)
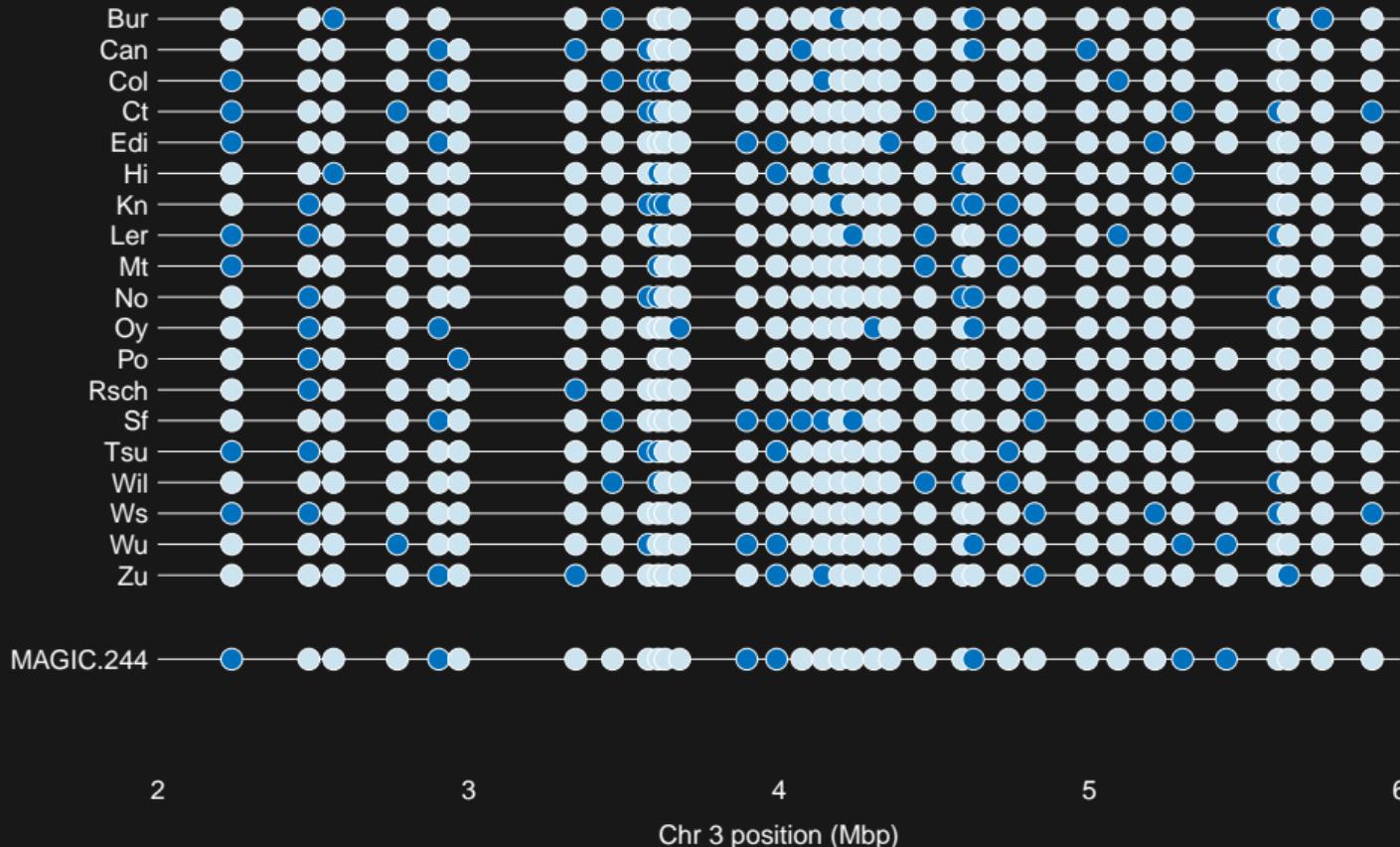
Genotype reconstruction
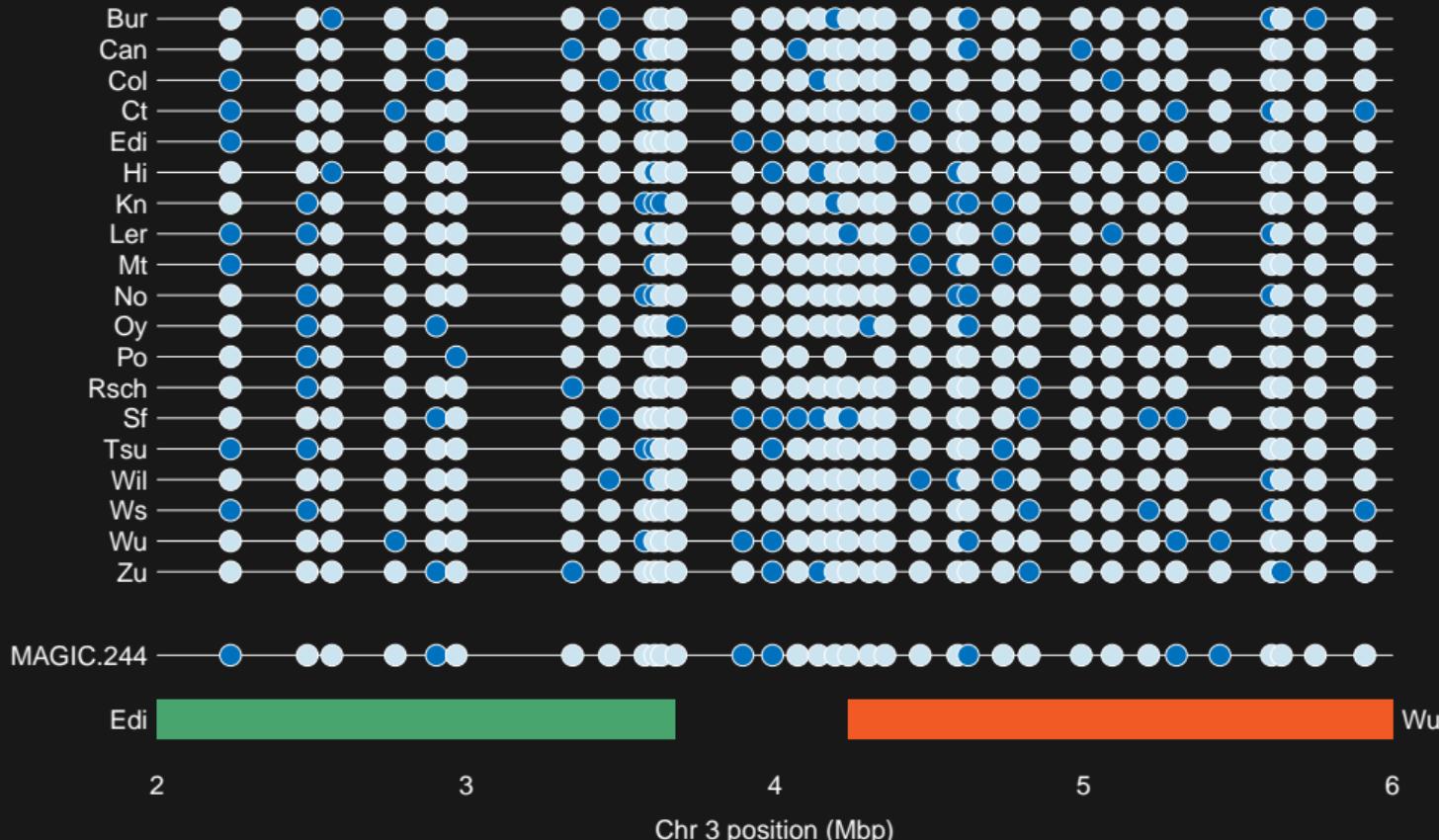
Genotype reconstruction

```
gmap <- insert_pseudomarkers(arab$gmap, step=0.2, stepwidth="max")
pmap <- interp_map(gmap, arab$gmap, arab$pmap)

pr <- calc_genoprob(arab, gmap, error_prob=0.002, cores=24)
```

Genome scan
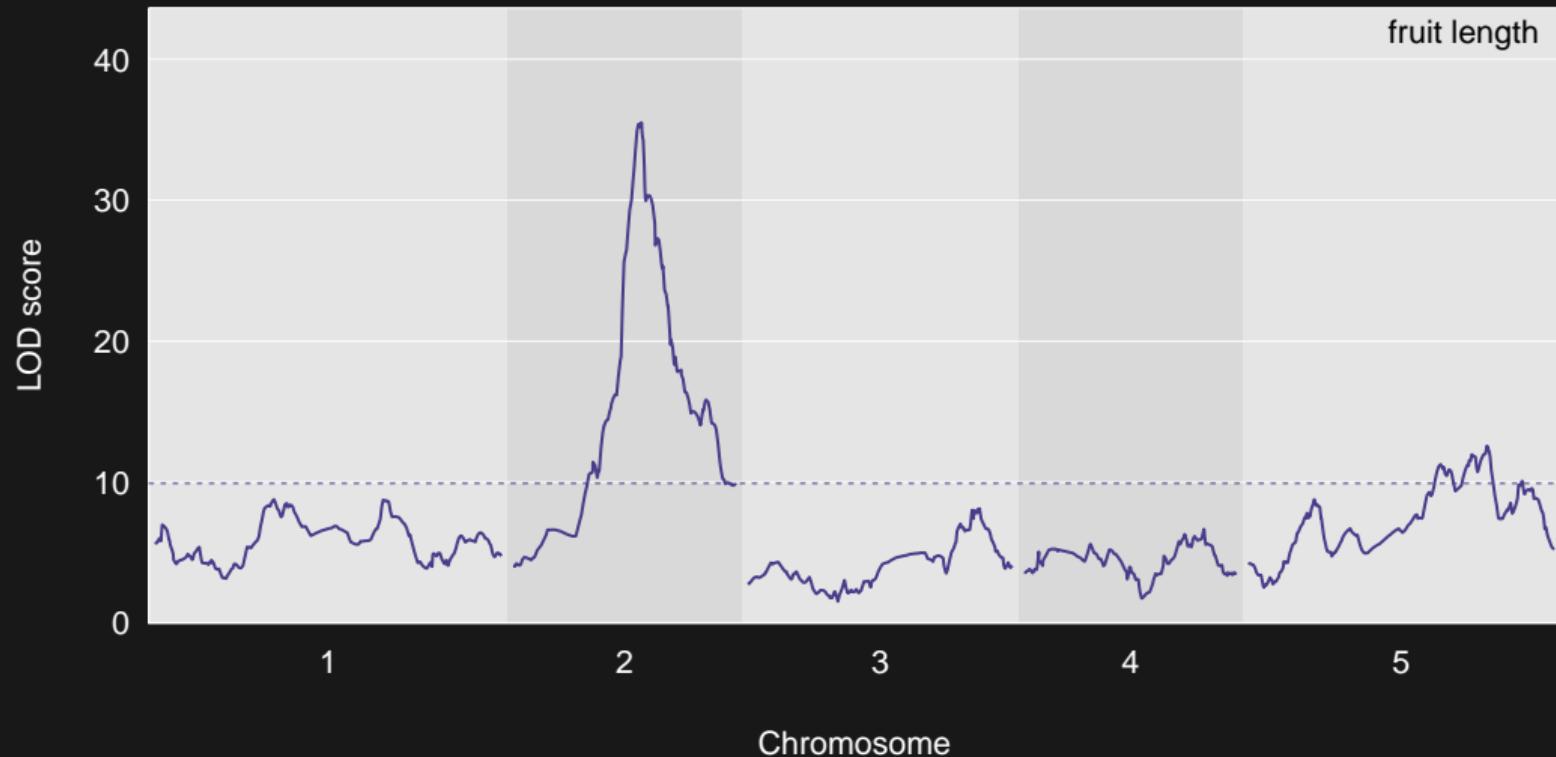
Genome scan
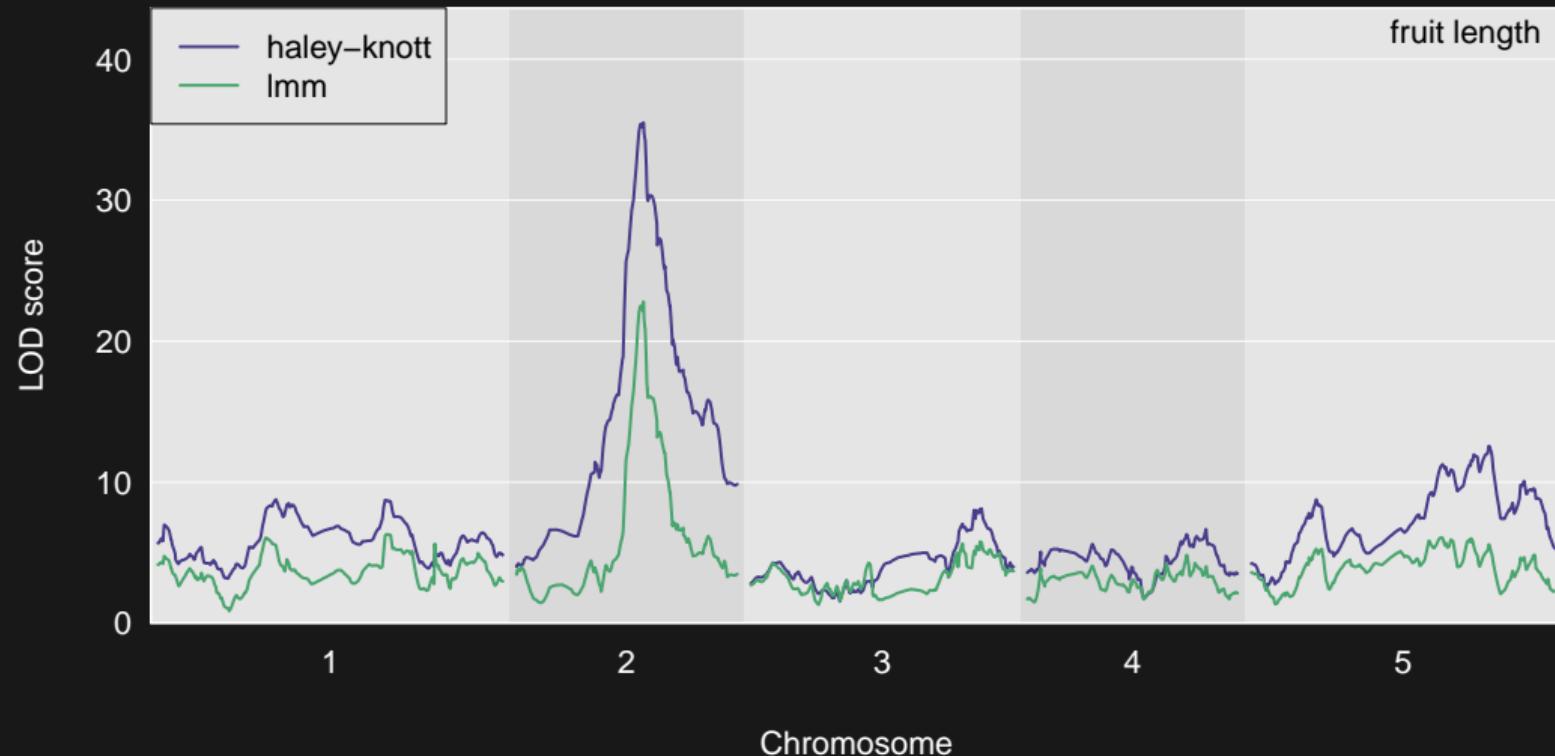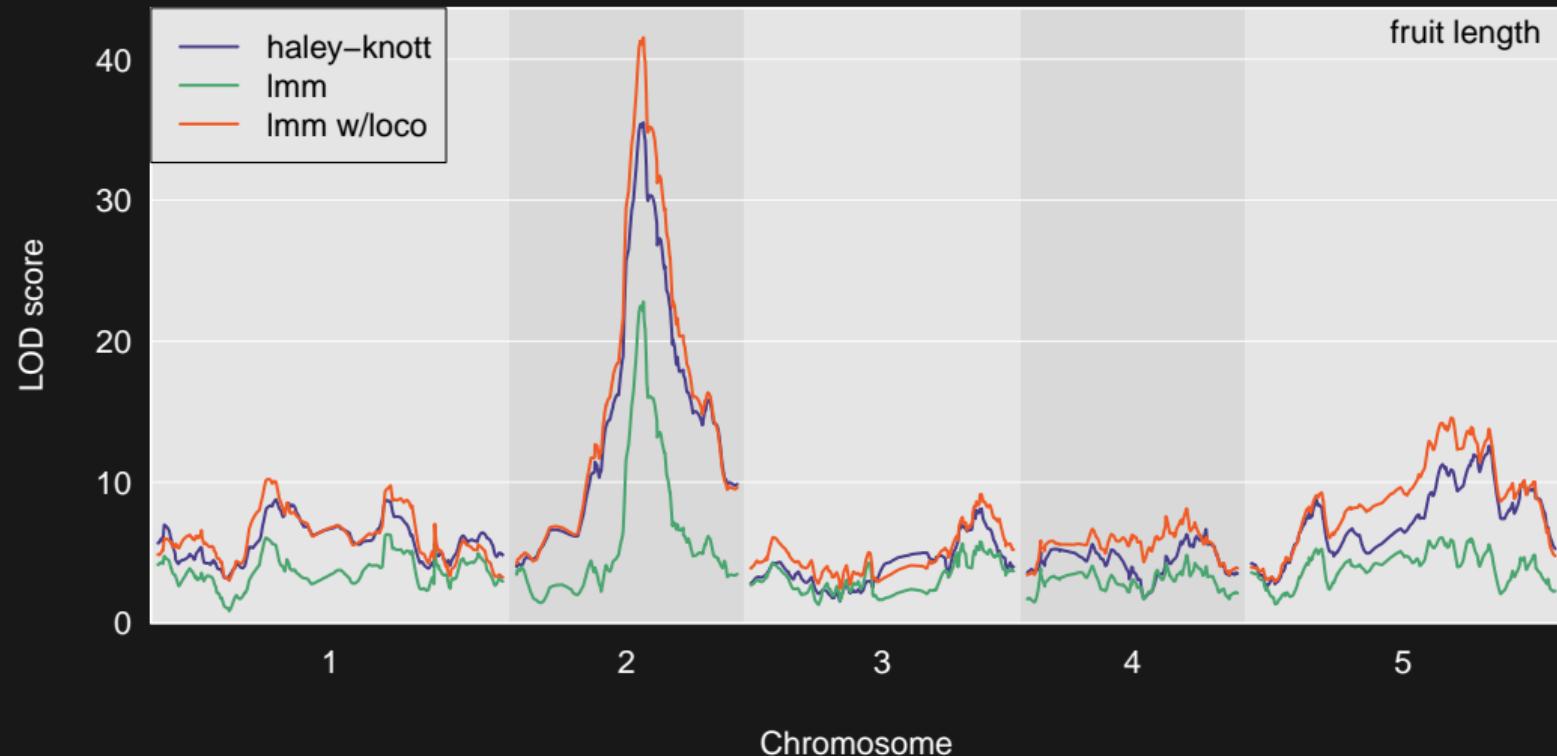
Genome scan
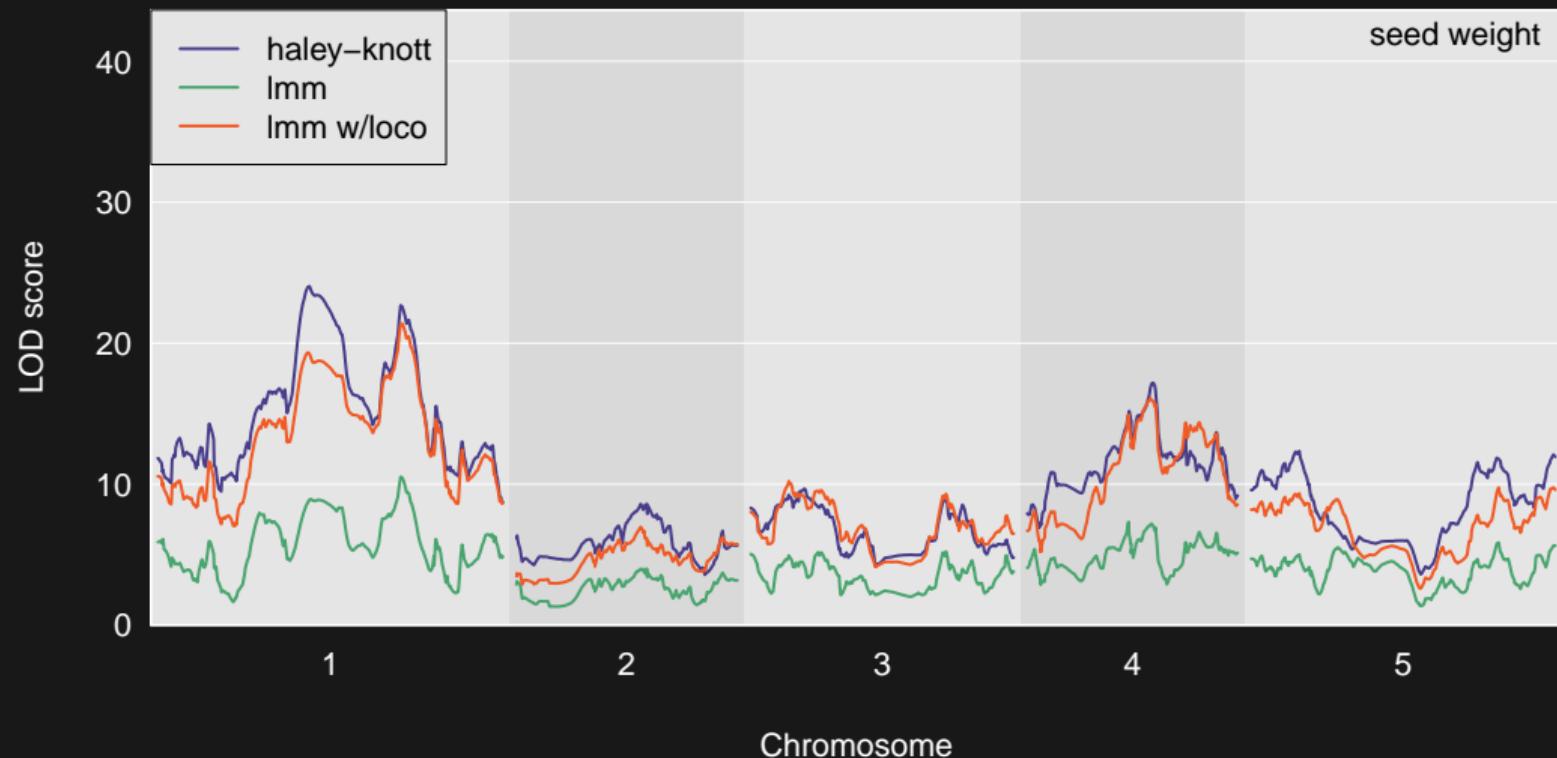
Genome scan

Genome scan

```
out_hk <- scan1(pr, arab$pheno, cores=24)

operm_hk <- scan1perm(pr, arab$pheno, n_perm=1000, cores=24)

k <- calc_kinship(pr, cores=24)
out_lmm <- scan1(pr, arab$pheno, k, cores=24)

k_loco <- calc_kinship(pr, "loco", cores=24)
out_loco <- scan1(pr, arab$pheno, k_loco, cores=24)
```

SNP association scan
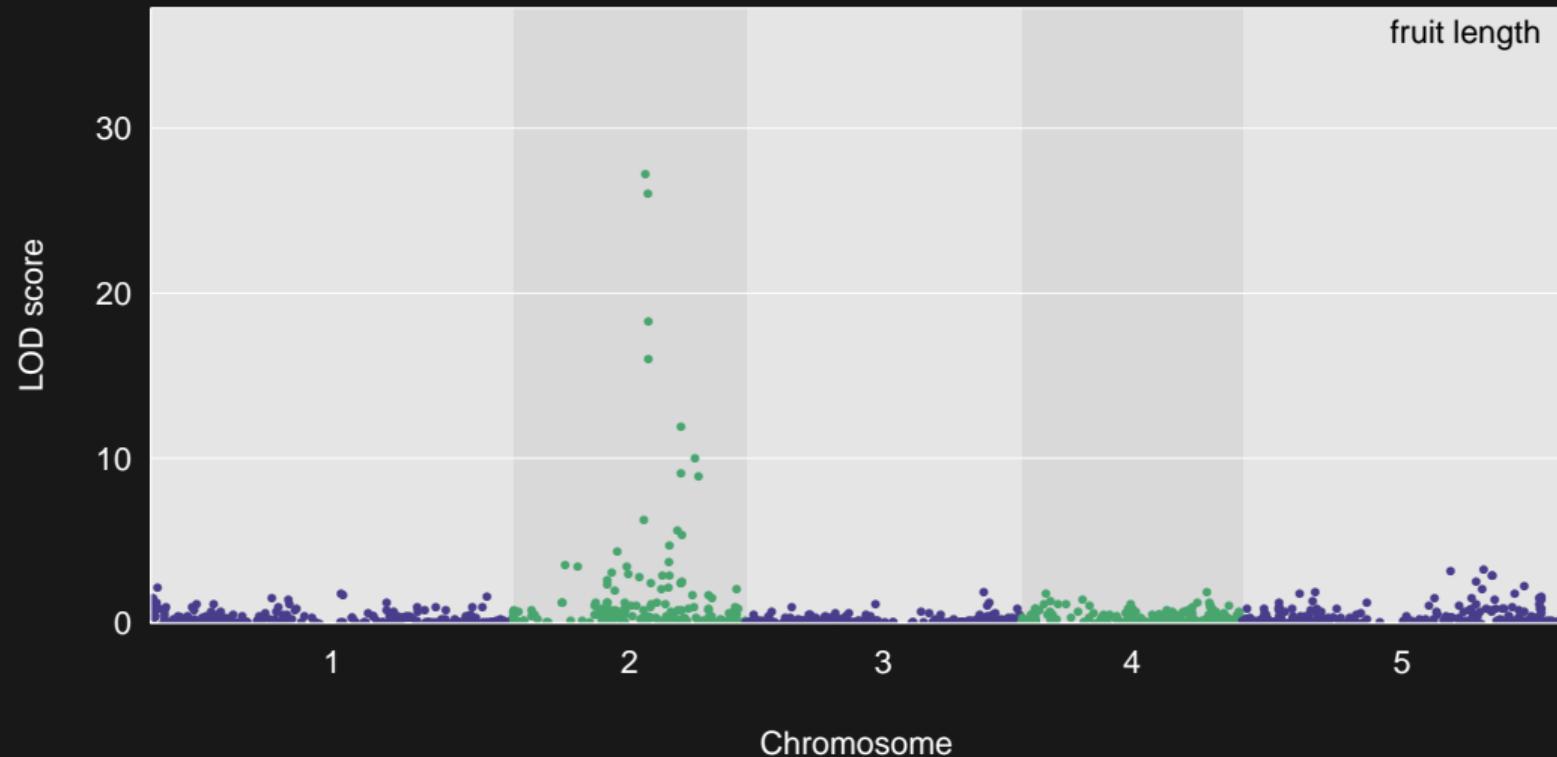
fruit length

SNP association scan

fruit length

LOD score vs Chromosome
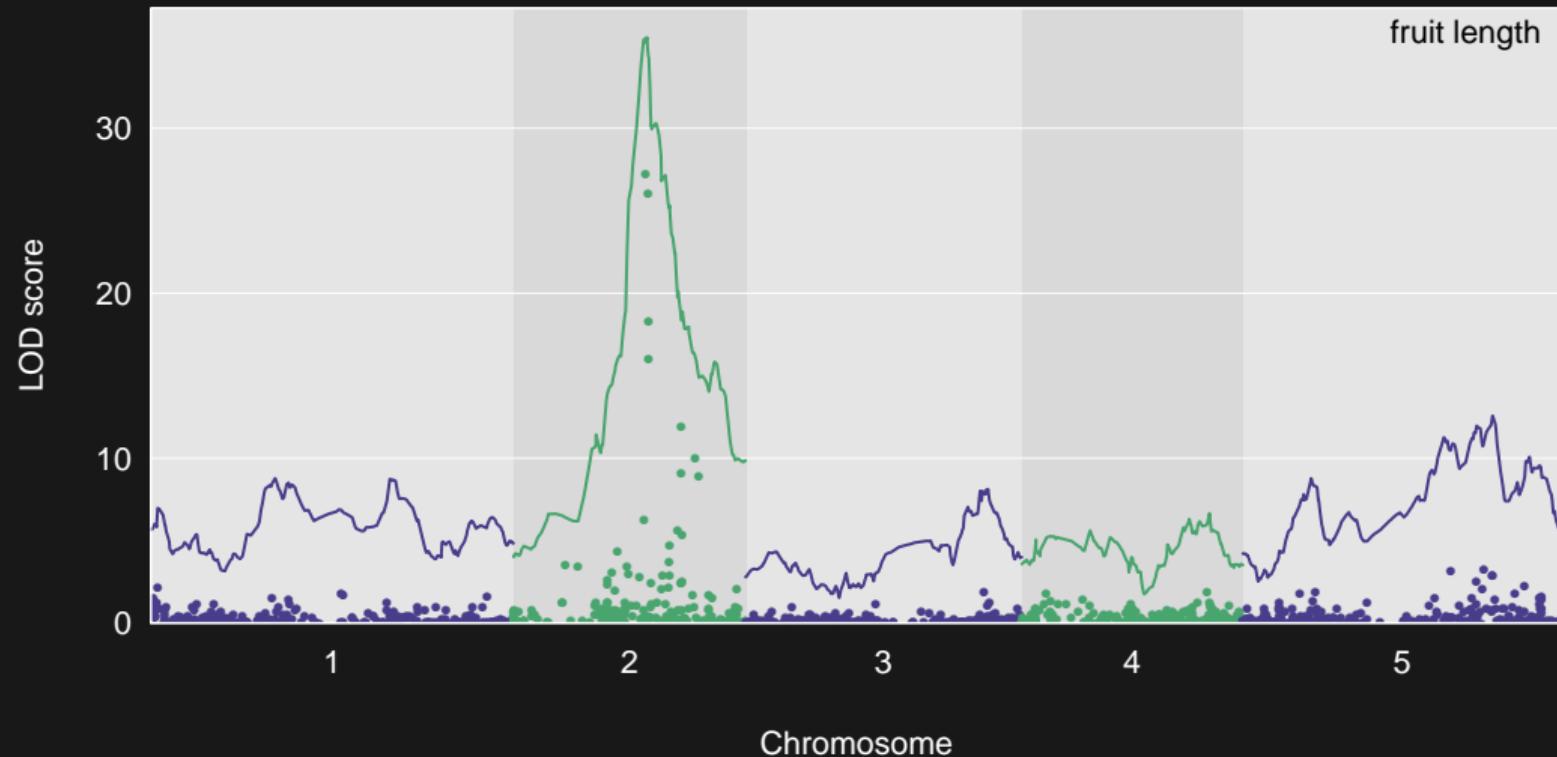
SNP association scan
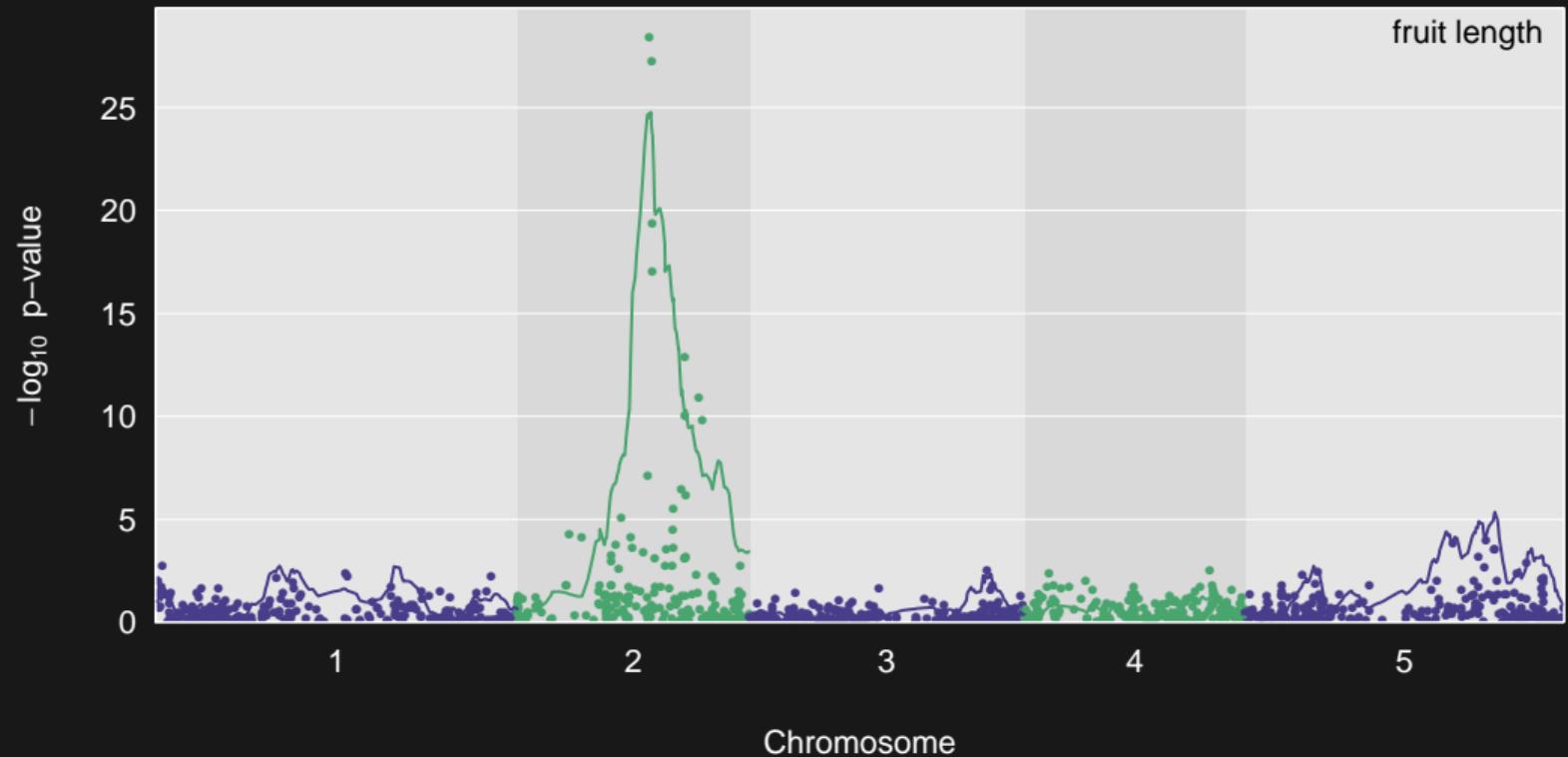
SNP association scan
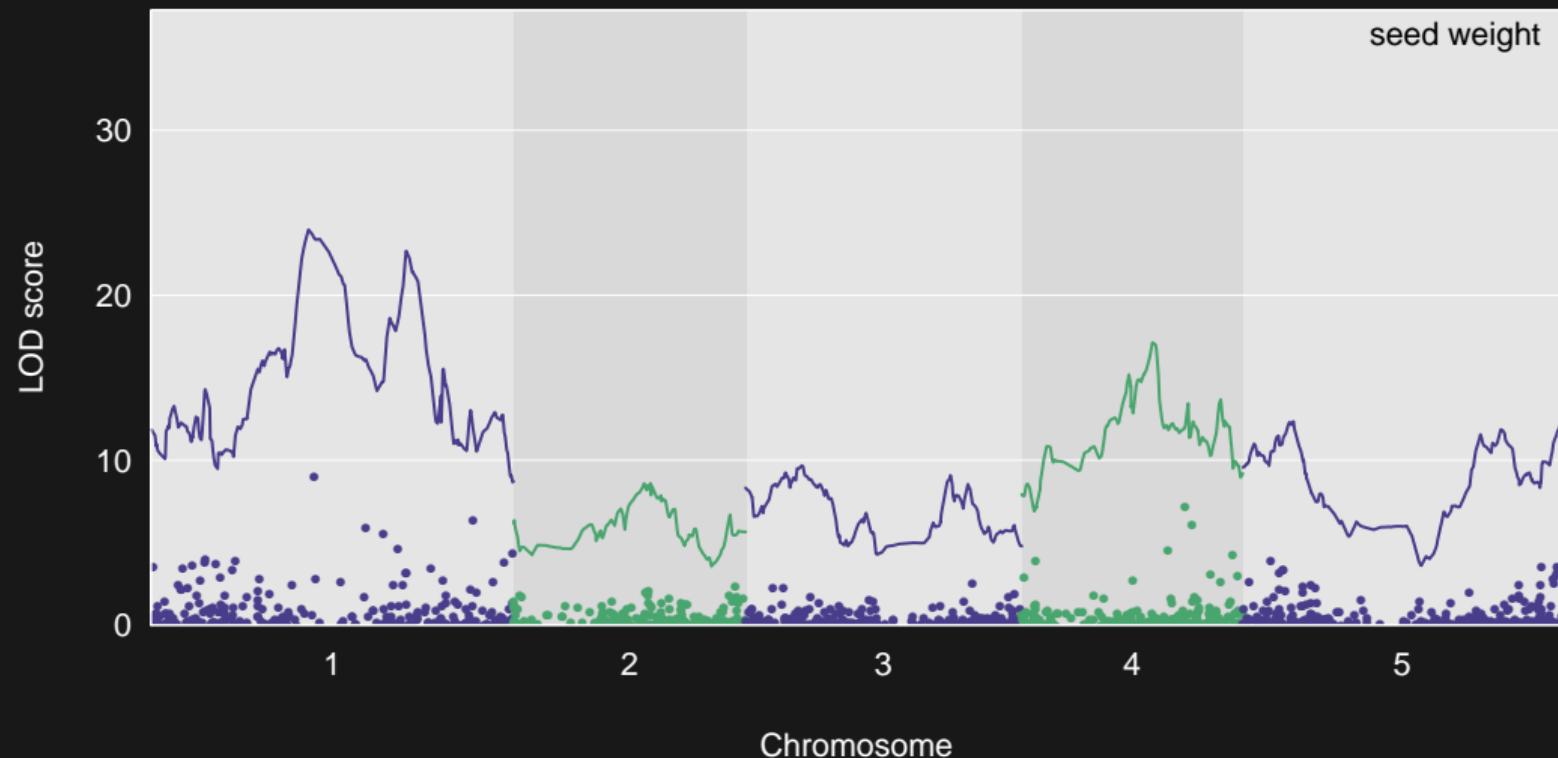
# SNP association scan
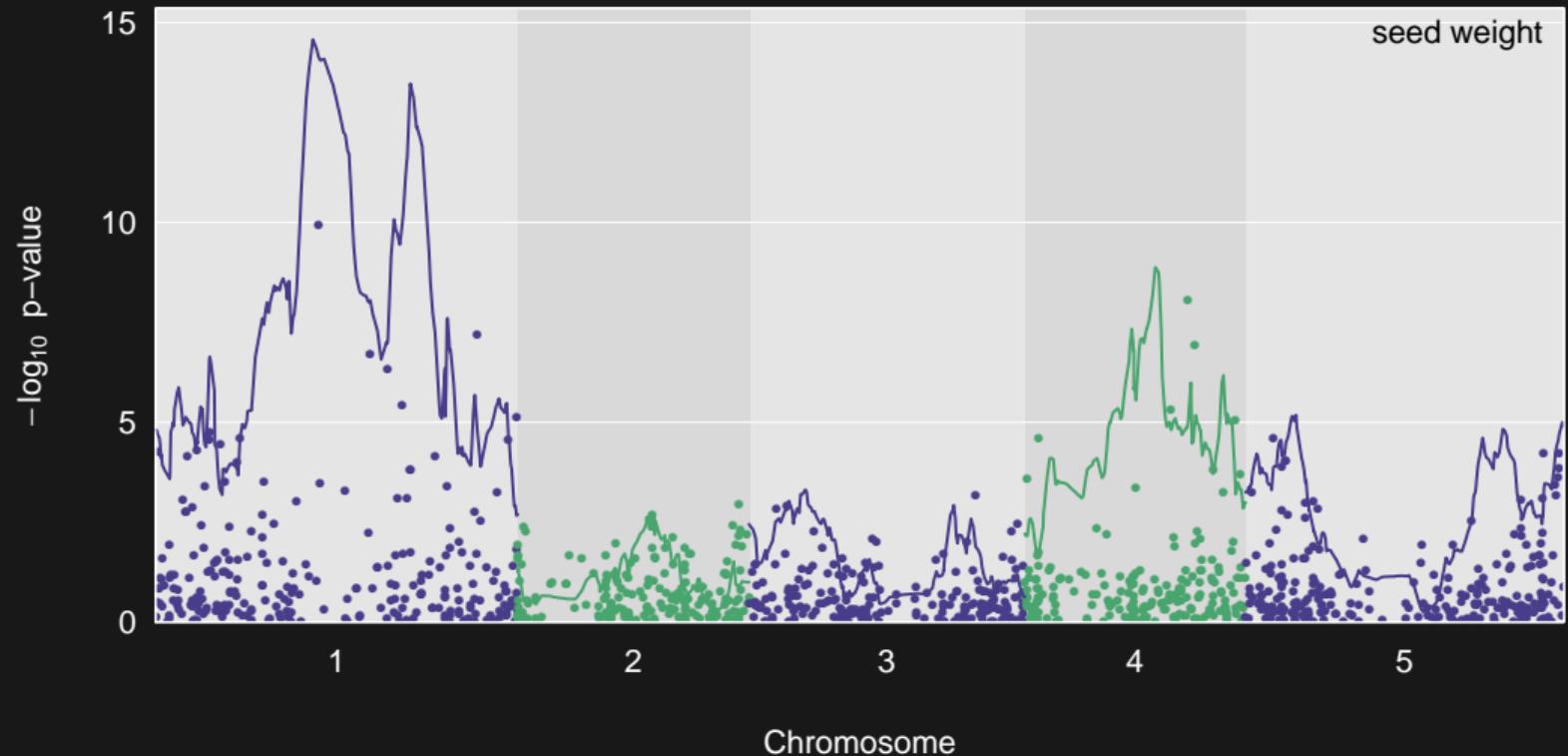
SNP association scan

```
snp_pr <- genoprob_to_snpprob(pr, arab)

out_snps <- scan1(snp_pr, arab$pheno, cores=24)
```

QTL effects

Fruit length (chr 2 @ 11.4 Mbp)

QTL effects

Fruit length (chr 2 @ 11.4 Mbp)

QTL effects

Seed weight (chr 1 @ 21.4 Mbp)
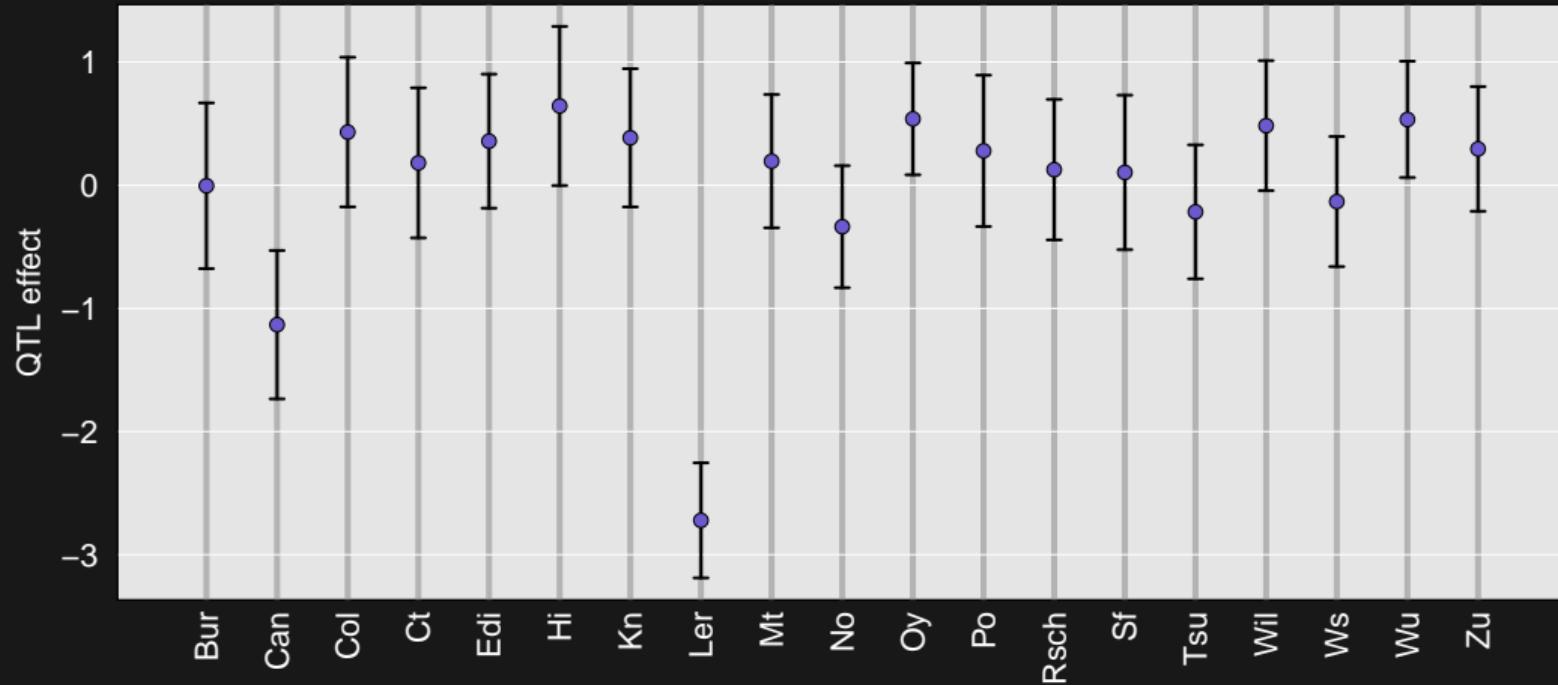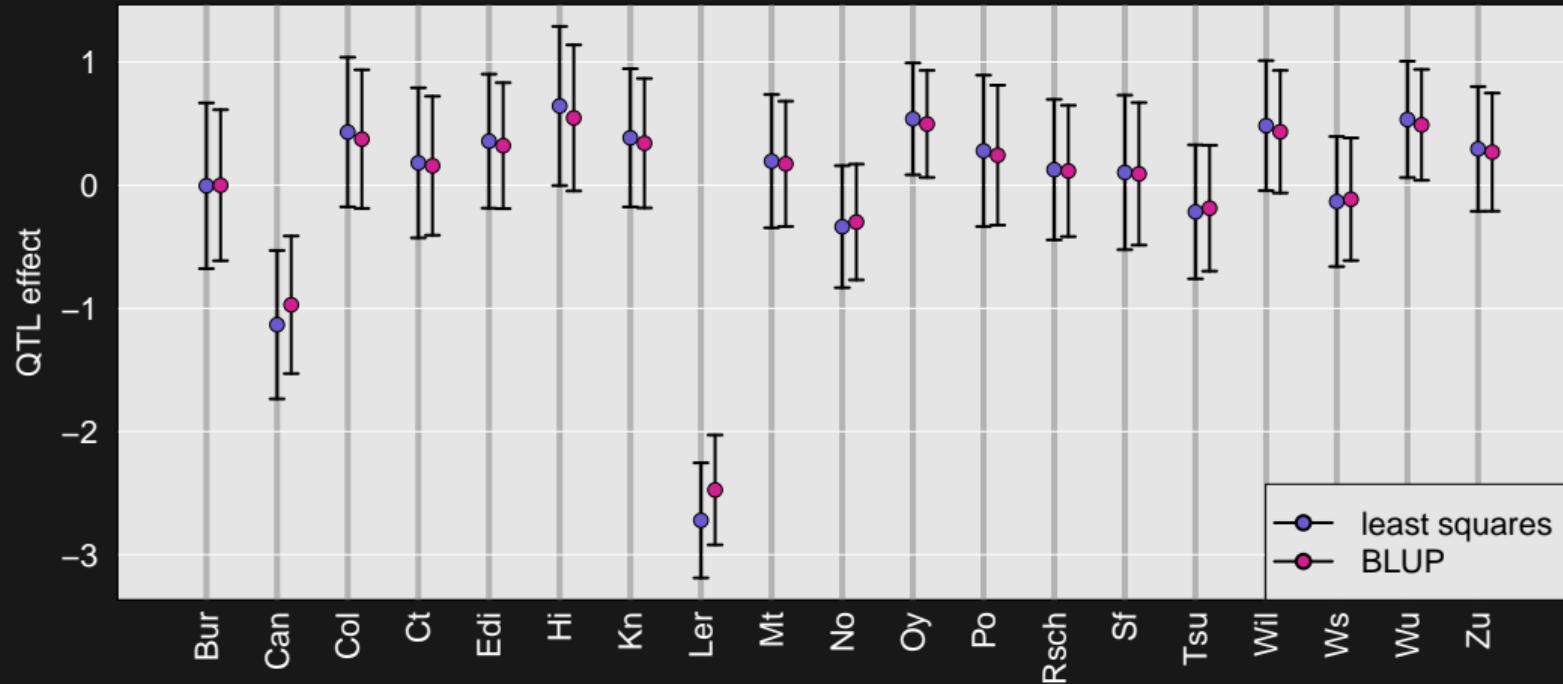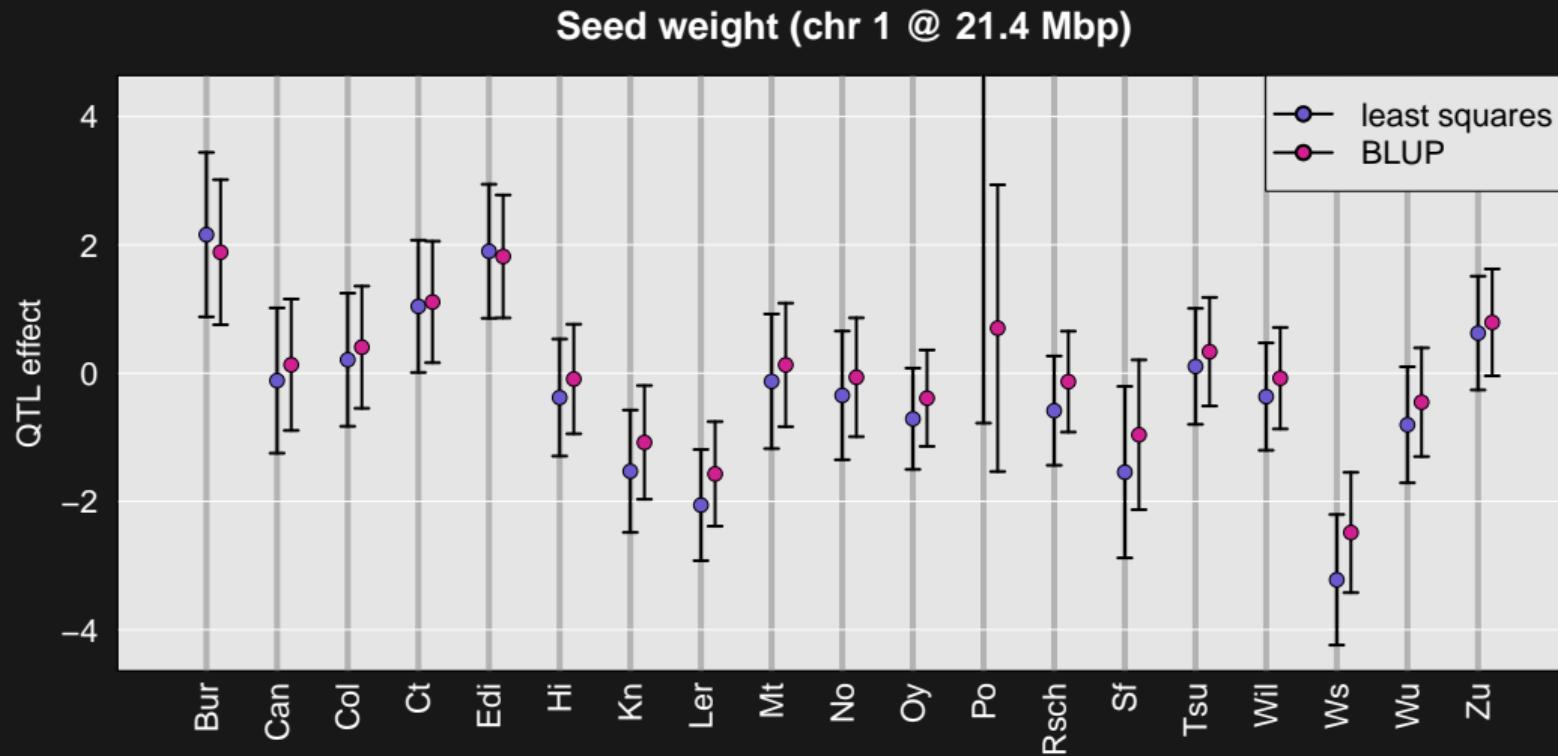
QTL effects

```
fl_peak <- max(out_hk, pmap, lodcolumn="fruit_length")
fl_pr <- pull_genoprobpos(pr, pmap, fl_peak$chr, fl_peak$pos)

fl_fit1 <- fit1(fl_pr, arab$pheno[,"fruit_length"])
fl_blup <- fit1(fl_pr, arab$pheno[,"fruit_length"], blup=TRUE)
```

Goals

- ▶ Genotype reconstructions from external software

- ▶ General models for RIL and AIL

- ▶ Sequencing-based genotype data

- ▶ Multiple-QTL models

- ▶ QTL $\times$ environment interactions

- ▶ Interactive data visualization

Slides: `bit.ly/msu2019-12` PUBLIC DOMAIN

`kbroman.org`

`kbroman.org/qtl2`

`github.com/kbroman`

`@kwbroman`