R/qtl2

high-dimensional data & multi-parent populations

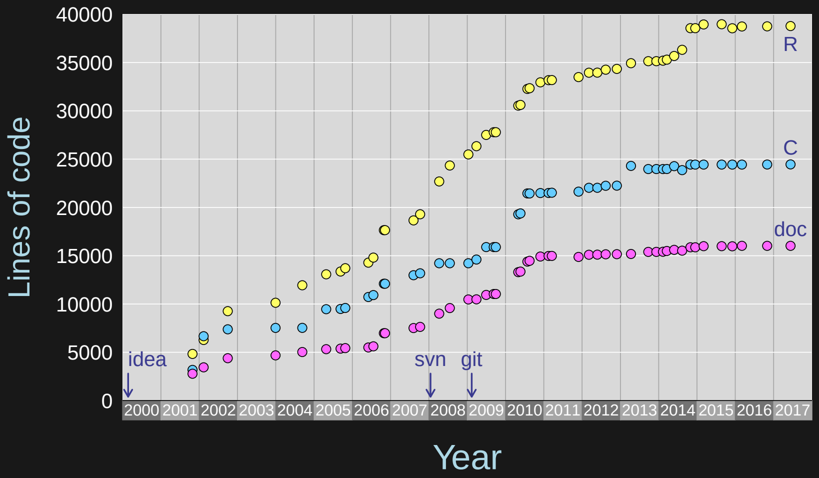
Karl Broman

Biostatistics & Medical Informatics, UW–Madison

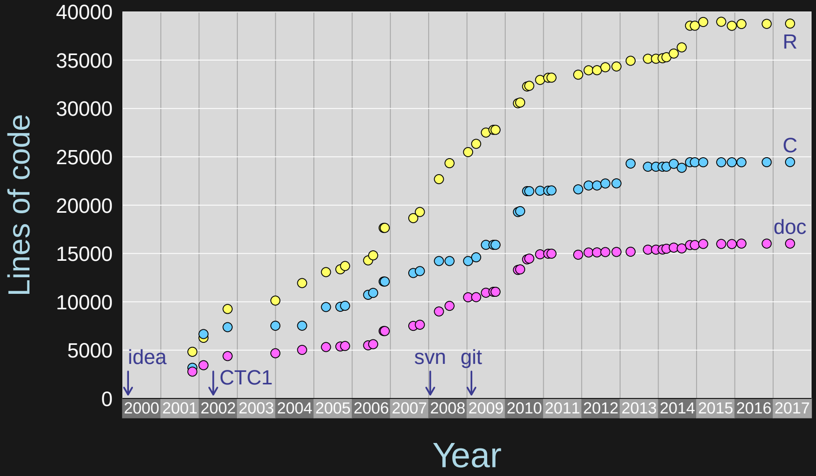
`kbroman.org`
`github.com/kbroman`
`@kwbroman`

Slides: `bit.ly/ctc_2017`

17 years of R/qtl



17 years of R/qtl



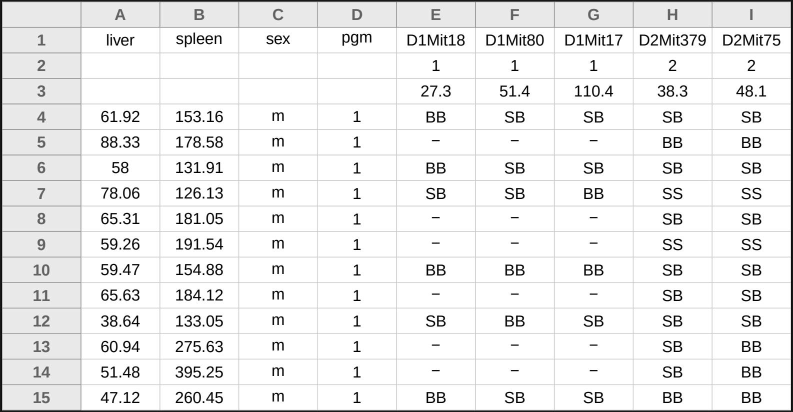
Good things

Good things

- HMM code
- basics of the user interface
- diagnostics and data visualization
- quite comprehensive
- quite flexible

Bad things

Input file



	A	B	C	D	E	F	G	H	I
1	liver	spleen	sex	pgm	D1Mit18	D1Mit80	D1Mit17	D2Mit379	D2Mit75
2					1	1	1	2	2
3					27.3	51.4	110.4	38.3	48.1
4	61.92	153.16	m	1	BB	SB	SB	SB	SB
5	88.33	178.58	m	1	–	–	–	BB	BB
6	58	131.91	m	1	BB	SB	SB	SB	SB
7	78.06	126.13	m	1	SB	SB	BB	SS	SS
8	65.31	181.05	m	1	–	–	–	SB	SB
9	59.26	191.54	m	1	–	–	–	SS	SS
10	59.47	154.88	m	1	BB	BB	BB	SB	SB
11	65.63	184.12	m	1	–	–	–	SB	SB
12	38.64	133.05	m	1	SB	BB	SB	SB	SB
13	60.94	275.63	m	1	–	–	–	SB	BB
14	51.48	395.25	m	1	–	–	–	SB	BB
15	47.12	260.45	m	1	BB	SB	SB	BB	BB

Stupidest code ever

```
n <- ncol(data)
temp <- rep(FALSE,n)
for(i in 1:n) {
  temp[i] <- all(data[2,1:i]=="")
  if(!temp[i]) break
}
if(!any(temp)) stop("...")
n.phe <- max((1:n)[temp])
```

kbroman.wordpress.com/2011/08/17/the-stupidest-r-code-ever

Open source means
everyone can see my stupid mistakes

Open source means
everyone can see my stupid mistakes

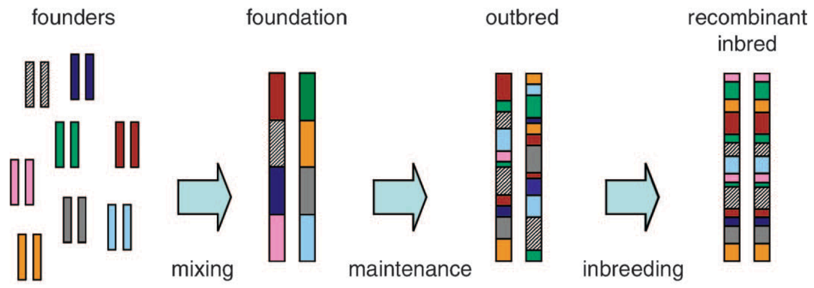
Version control means
everyone can see every stupid mistake I’ve ever made

Support

The “future”

Alan Attie

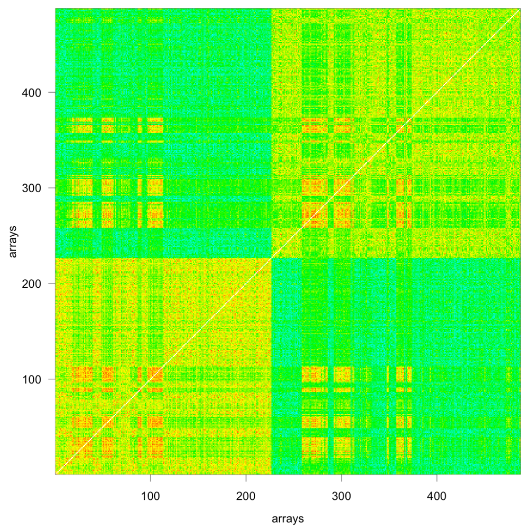
Multi-parent populations



combine mix fix

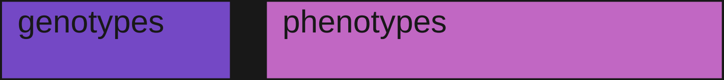
Valdar et al., Genetics 172:1783, 2006

Challenges: diagnostics



kbroman.wordpress.com/2012/04/25/microarrays-suck

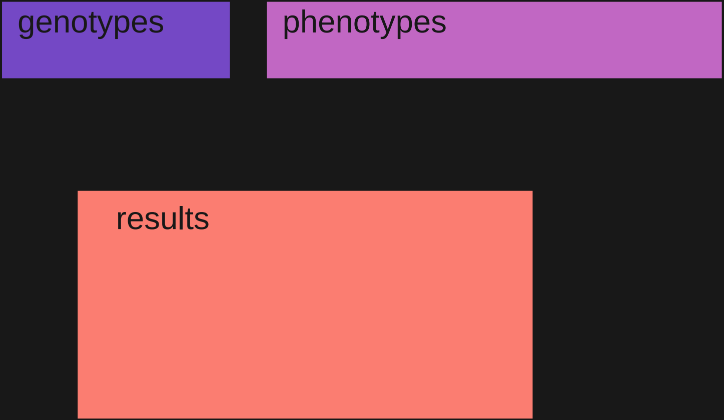
Challenges: scale of results



genotypes

phenotypes

Challenges: scale of results



genotypes

phenotypes

results

Challenges: organizing, automating

genotypes

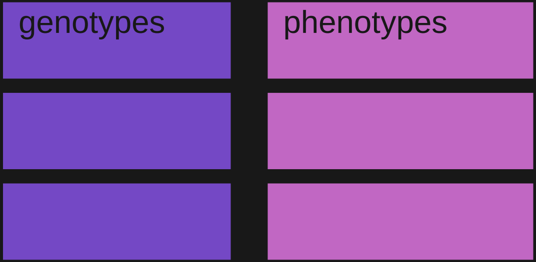
phenotypes

Challenges: organizing, automating

genotypes

phenotypes

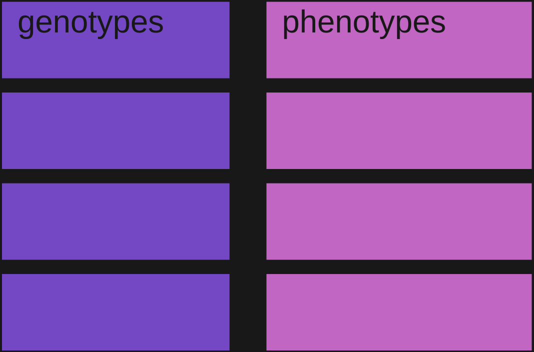
Challenges: organizing, automating



genotypes

phenotypes

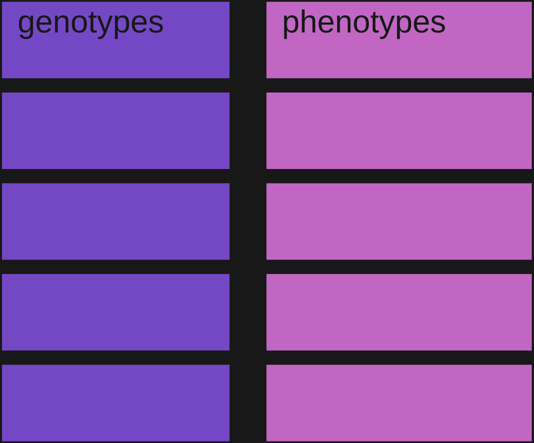
Challenges: organizing, automating



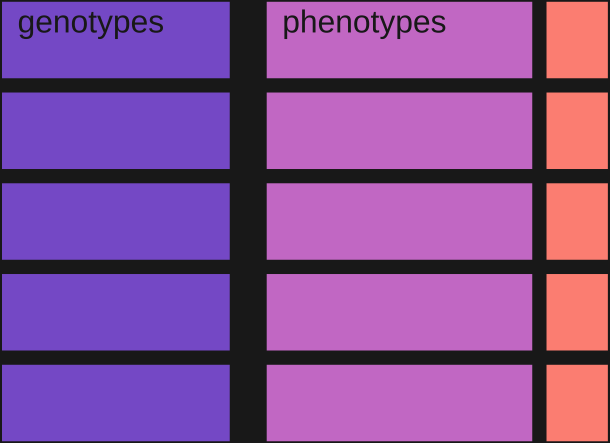
genotypes

phenotypes

Challenges: organizing, automating



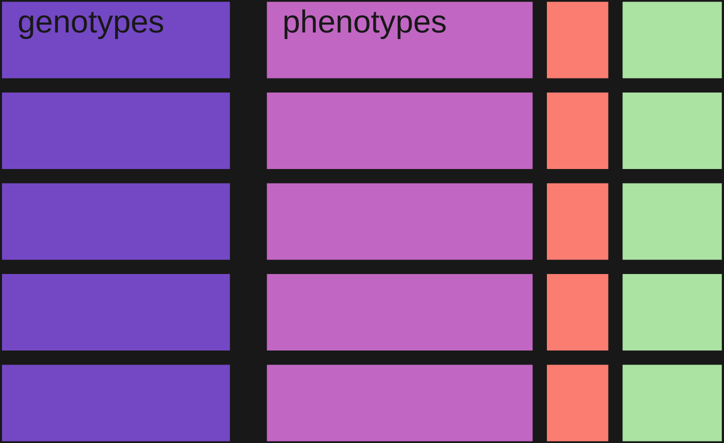
Challenges: organizing, automating



genotypes

phenotypes

Challenges: organizing, automating



Challenges: metadata

What the heck is `"FAD_NAD SI 8.3_3.3G"`?

What was the question again?


R/qt12
Now in 3D

R/qtl2

- High-density genotypes
- High-dimensional phenotypes
- Multi-parent populations
- Linear mixed models

`kbroman.org/qtl2`

R/qtl2: Let's not make the same mistakes

- C++ and Rcpp
- Roxygen2 for documentation
- Unit tests
- A single "switch" for cross type

R/qtl2: Let’s not make the same mistakes

- C++ and Rcpp
- Roxygen2 for documentation
- Unit tests
- A single “switch” for cross type
- Split into multiple packages
- Yet another data input format
- Flatter data structures, but still complex

R/qtl2 cross types

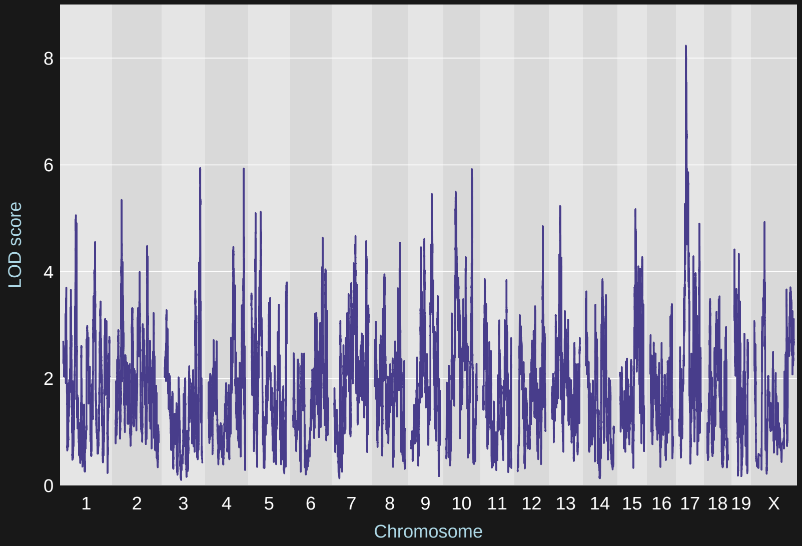
- backcross
- intercross
- 2-, 4-, 8-, and 16-way RIL by selfing
- 2-, 4-, and 8-way RIL by sib-mating
- 2-way AIL
- Heterogeneous Stock (HS)
- Diversity Outbreds (DO)
- DO-F1
- 19-way MAGIC

Initial goal

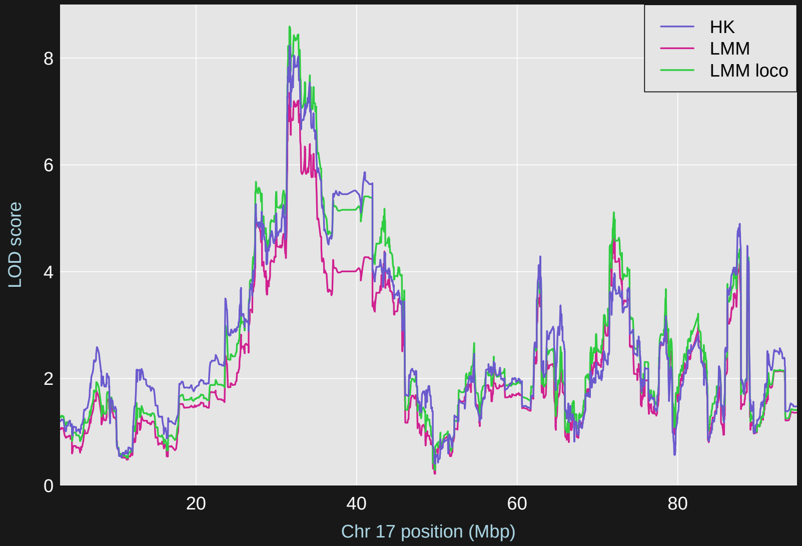
Everything in Dan Gatti's DOQTL,

but for more general crosses

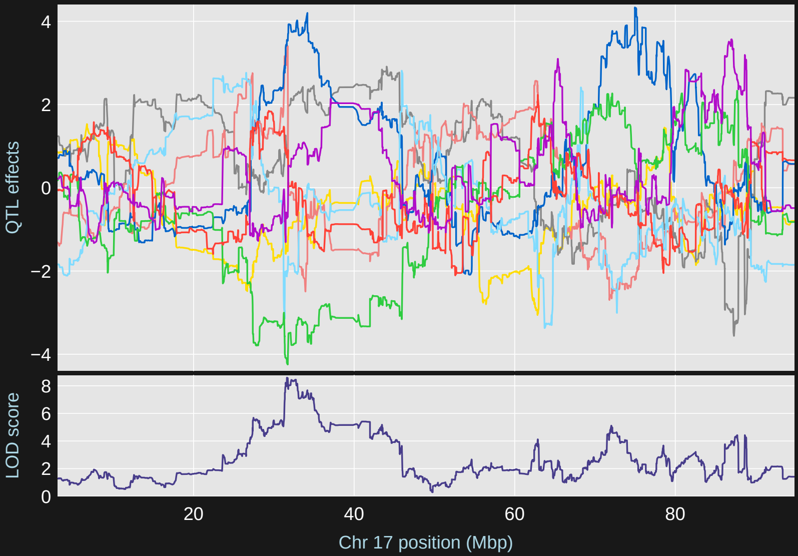
Genome scan


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

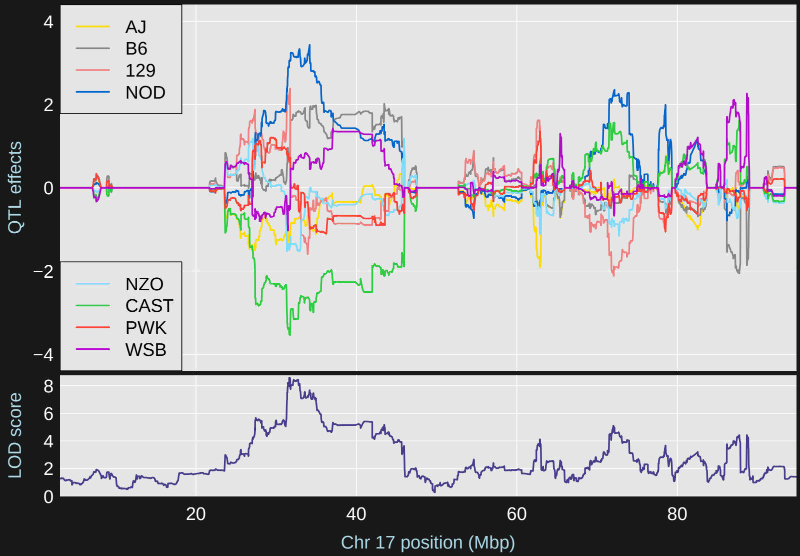
LMMs



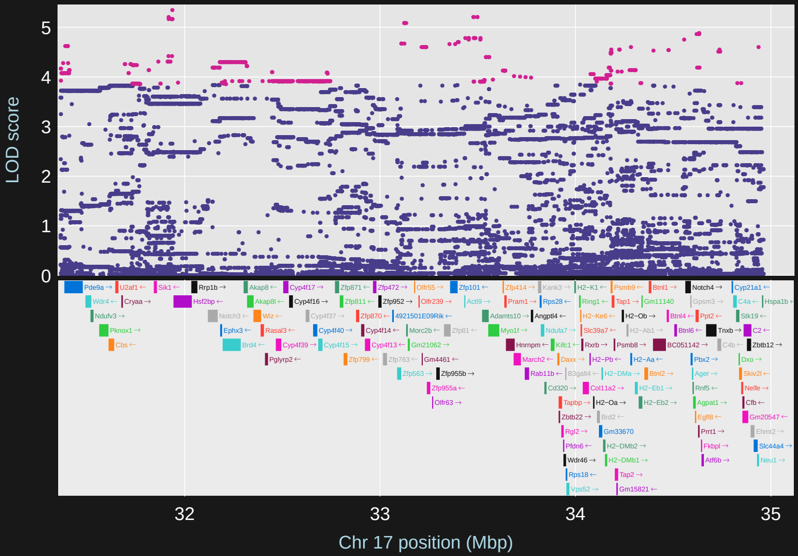
QTL effects


4
2
0
−2
−4
QTL effects
8
6
4
2
0
LOD score
20
40
60
80
Chr 17 position (Mbp)

BLUPs


AJ
B6
129
NOD
NZO
CAST
PWK
WSB
QTL effects
4
2
0
−2
−4
LOD score
8
6
4
2
0
20
40
60
80
Chr 17 position (Mbp)

SNP associations



Acknowledgments

Danny Arends
Gary Churchill
Nick Furlotte
Dan Gatti
Ritsert Jansen
Pjotr Prins
Śaunak Sen
Petr Simecek
Artem Tarasov
Hao Wu
Brian Yandell

Robert Corty
Timothée Flutre
Lars Ronnegard
Rohan Shah
Laura Shannon
Quoc Tran
Aaron Wolen

NIH/NIGMS

Slides: `bit.ly/ctc_2017`

`kbroman.org`

`kbroman.org/qtl2`

`github.com/kbroman`

`@kwbroman`