R/qtl2: progress & plans

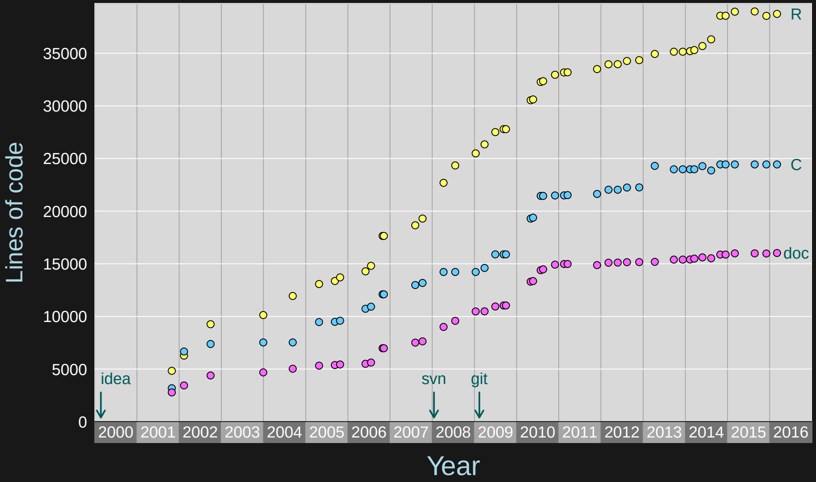
Karl Broman

Biostatistics & Medical Informatics, UW–Madison

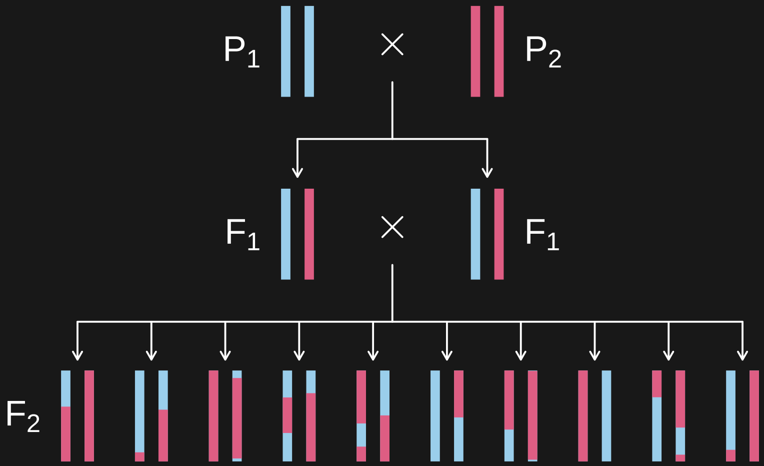
kbroman.org
github.com/kbroman
@kwbroman

Slides: bit.ly/Memphis2016

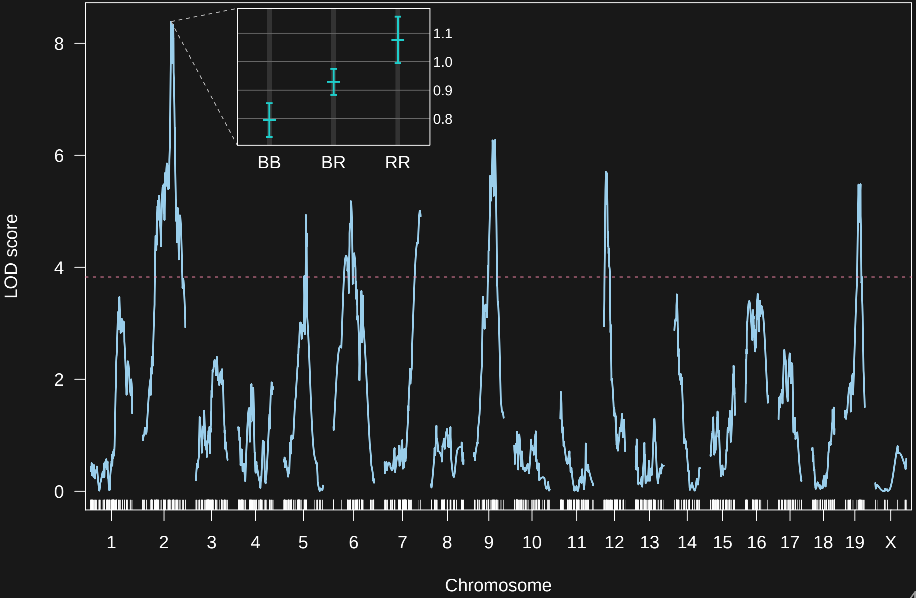
R/qtl


Lines of code
35000
30000
25000
20000
15000
10000
5000
0
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
Year
R
C
doc
idea
svn
git

Intercross



QTL mapping

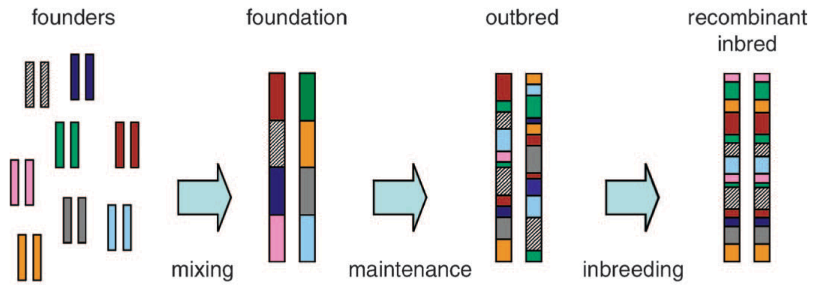


8
6
4
2
0
LOD score
1.1
1.0
0.9
0.8
BB
BR
RR
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
X
Chromosome

The “future”

Alan Attie

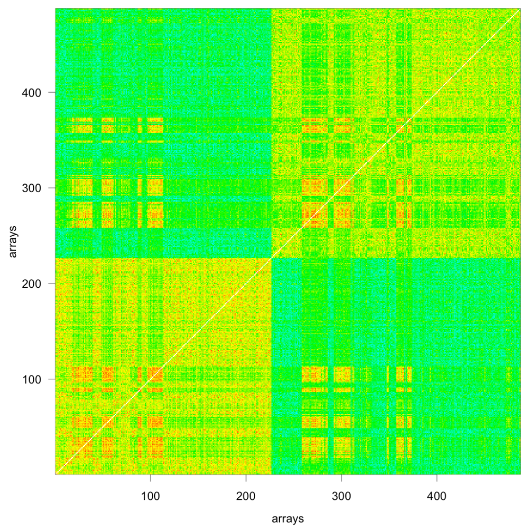
Multi-parent populations



combine mix fix

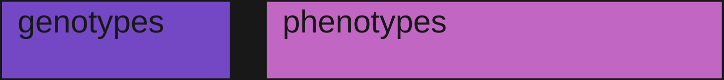
Valdar et al., Genetics 172:1783, 2006

Challenges: diagnostics



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

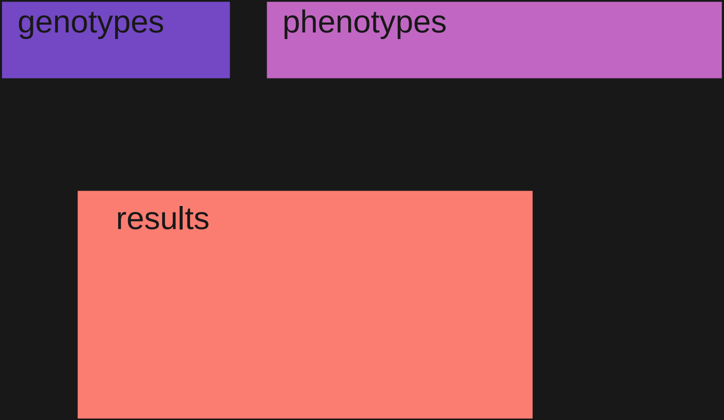
Challenges: scale of results



genotypes

phenotypes

Challenges: scale of results



Challenges: organizing, automating

genotypes

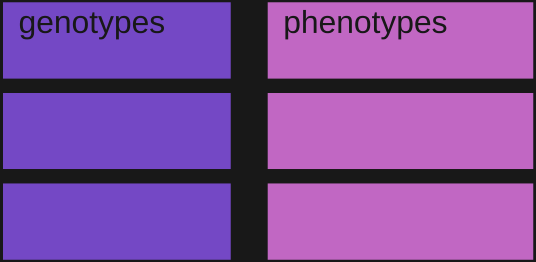
phenotypes

Challenges: organizing, automating

genotypes

phenotypes

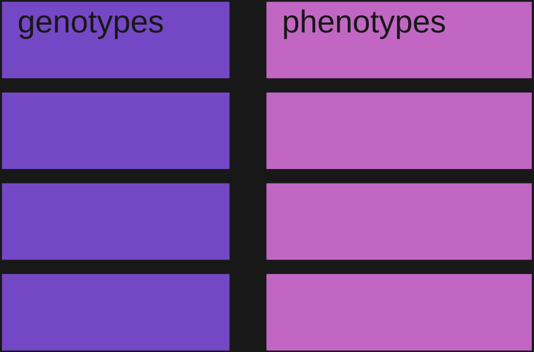
Challenges: organizing, automating



genotypes

phenotypes

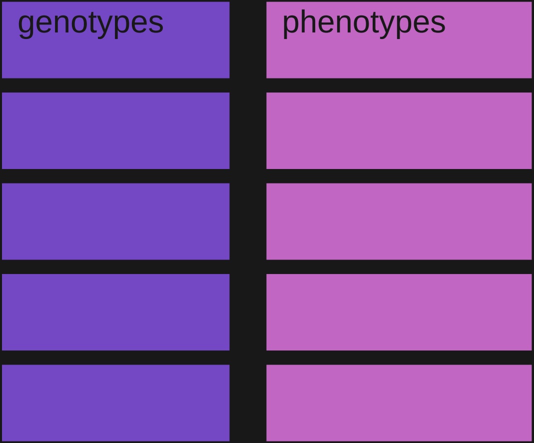
Challenges: organizing, automating



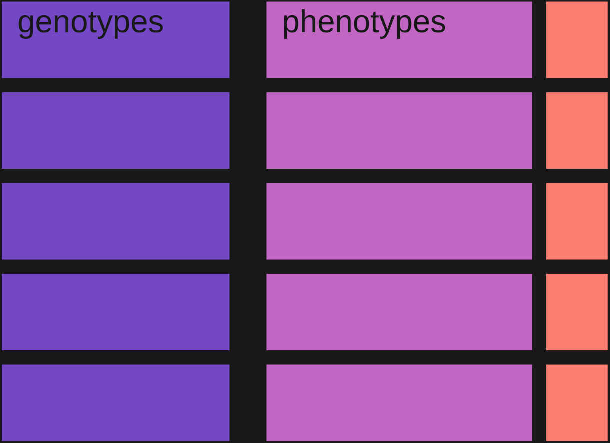
genotypes

phenotypes

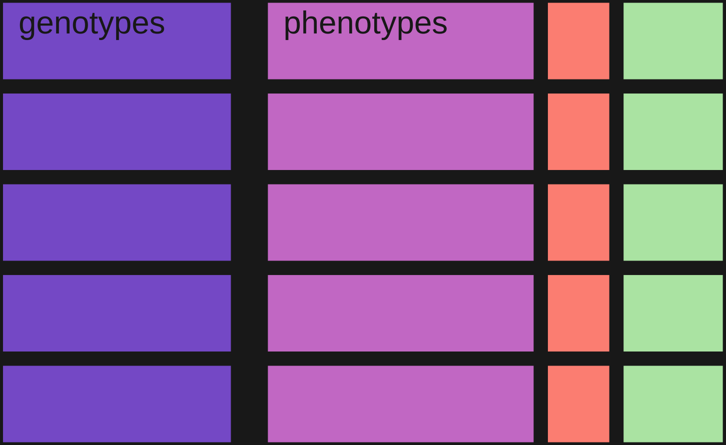
Challenges: organizing, automating



Challenges: organizing, automating



Challenges: organizing, automating



genotypes

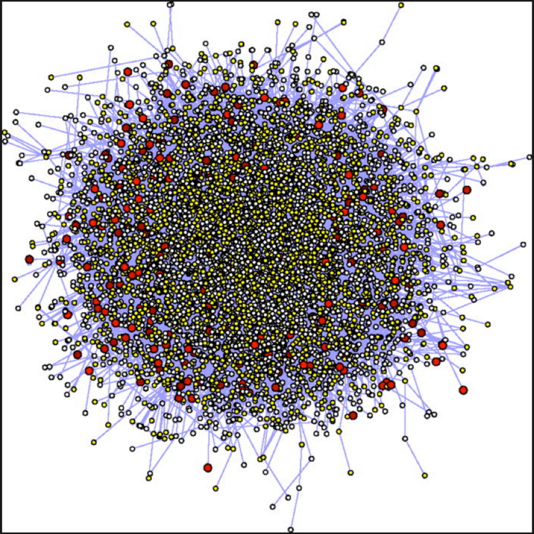
phenotypes

Challenges: metadata

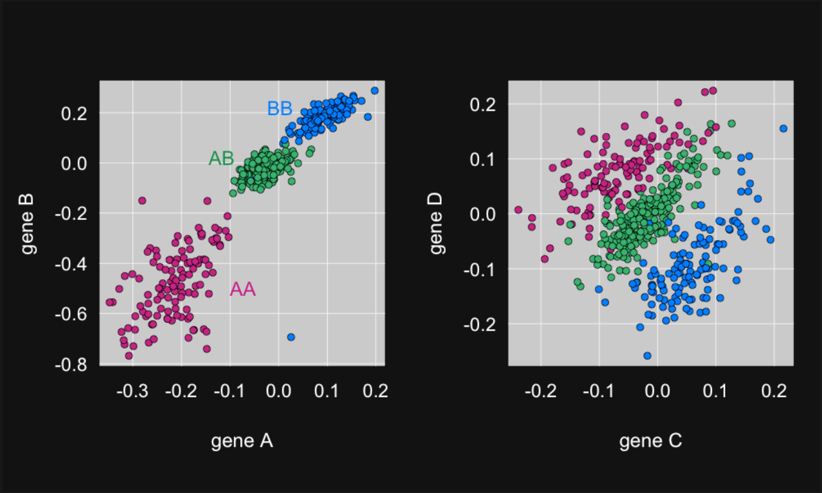
What the heck is `FAD_NAD SI 8.3_3.3G`?

What was the question again?

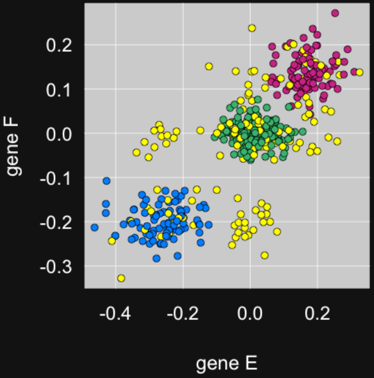
The ridiculome



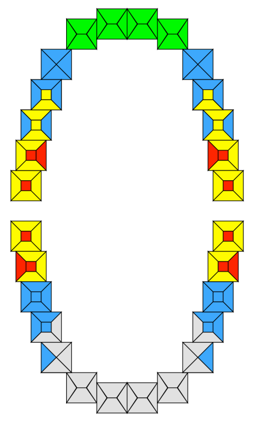
Multivariate phenotypes



Multivariate phenotypes

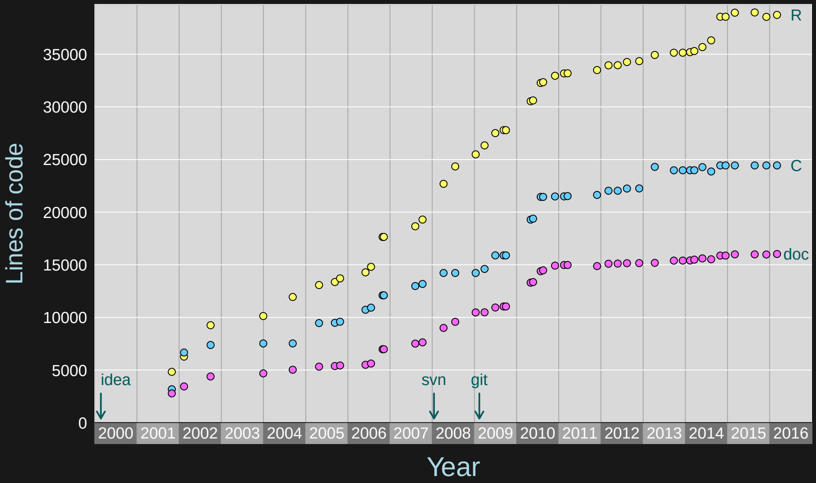


Composite phenotypes



Shaffer et al. (2013) J Dent Res 92:32-37

R/qtl


R
C
doc
idea
svn
git
Lines of code
0
5000
10000
15000
20000
25000
30000
35000
Year
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016

Good things

Good things

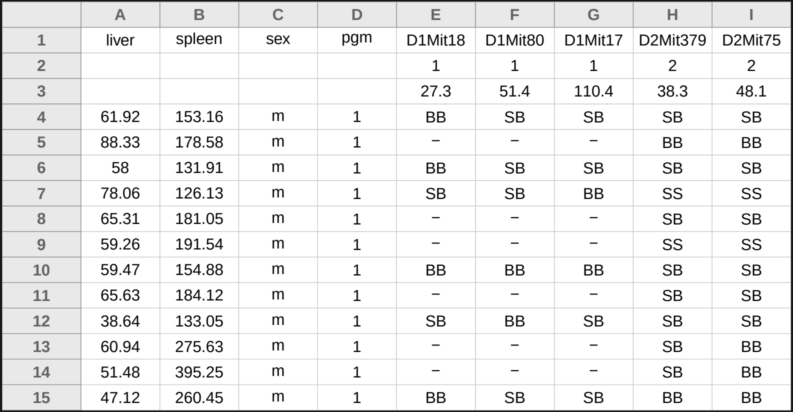
- HMM code
- basic user interface
- comprehensive
- diagnostics and data visualization
- quite flexible

Bad things

Stupidest code ever

```r
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])
```

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

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

Baroque data structures

```
attr(mycross$geno[["X"]]$probs, "map")
```

R/qtl2

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

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

- C++ and Rcpp
- Simpler documentation
- Unit tests
- A single “switch” for cross type

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

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

Slides: `bit.ly/Memphis2016`

`kbroman.org`

`github.com/kbroman`

`@kwbroman`