18 years of R/qtl
maintaining, supporting, and sustaining scientific software

Karl Broman
Biostatistics & Medical Informatics, UW–Madison

kbroman.org
github.com/kbroman
@kwbroman
Slides: bit.ly/qBio2018
18 years of R/qtl
Intercross
Data

![Heatmap of marker expression across different mice and markers. The phenotype is represented on the right side of the heatmap.](image)
QTL mapping

Chromosome
LOD score

BB  BR  RR
0.8  0.9  1.0  1.1
18 years of R/qtl
Why?
Good things
Good things

- some of the code
- basics of the user interface
- diagnostics and data visualization
- quite comprehensive
- quite flexible
Bad things
## Input file

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

kbroman.org/blog/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.
Documentation
Support
QTL mapping
Congenic line
Improving precision

- more recombinations
- more individuals
- more precise phenotype
- lower-level phenotypes
  - transcripts, proteins, metabolites
Advanced intercross lines

P
A
B
F2
F3
F4
F7
F10
Recombinant inbred lines

\[
P_1 \quad \times \quad P_2
\]

\[
P_1 \quad \times \quad P_2
\]

\[
F_1 \quad \times \quad F_1
\]

\[
F_2 \quad \times \quad F_2
\]

\[
F_3 \quad \times \quad F_3
\]

\[
F_4 \quad \times \quad F_4
\]

\[
F_\infty \quad \times \quad F_\infty
\]

...
Collaborative Cross

G₀
A ↓ B
C ↓ D

G₁
A B
C D

G₂
ABCD

G₃

G₄

G∞
Heterogeneous stock

G₀
A  B  C  D  E  F  G  H
G₁
G₂
G₁₀
G₁₅
G₂₀
Genome-scale phenotypes
Challenges: diagnostics
Challenges: diagnostics

[Link to many_boxplots]

Array
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

- genotypes
- phenotypes
Challenges: organizing, automating

genotypes

phenotypes
Challenges: organizing, automating

- genotypes
- phenotypes
Challenges: metadata

What the heck is "FAD_NAD SI 8.3_3.3G"?
What was the question again?
R/9t12
Now in 3D

ropenscilabs.github.io/miner_book
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
- 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
- Yet another data input format
- Flatter data structures, but still complex
Sustainable academic software
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/qBio2018

kbroman.org

kbroman.org/qtl2

github.com/kbroman

@kwbroman