21 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/pitt2021
21 years of R/qtl

Lines of code

Year

idea  svn  git  doc

R
C
doc
Intercross

P₁ \times P₂

F₁ \times F₁

F₂
Data
QTL mapping
21 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>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>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>1</td>
<td>260.45</td>
<td>47.12</td>
<td>15</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>395.25</td>
<td>51.48</td>
<td>14</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>275.63</td>
<td>60.94</td>
<td>13</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>133.05</td>
<td>38.64</td>
<td>12</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>184.12</td>
<td>65.63</td>
<td>11</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>154.88</td>
<td>59.26</td>
<td>10</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>131.91</td>
<td>78.06</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>126.13</td>
<td>59.47</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>181.05</td>
<td>65.31</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>191.54</td>
<td>59.26</td>
<td>6</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>11</td>
<td>154.88</td>
<td>59.47</td>
<td>5</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>184.12</td>
<td>65.63</td>
<td>4</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>13</td>
<td>275.63</td>
<td>60.94</td>
<td>3</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>14</td>
<td>395.25</td>
<td>51.48</td>
<td>2</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>260.45</td>
<td>47.12</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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.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

![QTL mapping graph](image)
Congenic line
Improving precision

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

P
A
B

F_2

F_3

F_4

F_7

F_{10}
Recombinant inbred lines

P1  ||  ||  P2
   ↓  ×  ↓
  F1  ||  ||  F1
   ↓  ×  ↓
 F2  ×  ×  F2
   ↓  ↓  ↓
 F3  ×  ×  F3
   ↓  ↓  ↓
 F4  ×  ×  F4
   ↓  ↓  ↓
 ...  ↓  ↓  ...
 F∞  ||  ||  F∞
# Heterogeneous stock

<table>
<thead>
<tr>
<th>$G_0$</th>
<th>$G_1$</th>
<th>$G_2$</th>
<th>$G_{10}$</th>
<th>$G_{15}$</th>
<th>$G_{20}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_{10}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_{15}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_{20}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Genome-scale phenotypes

Alan Attie
Challenges: diagnostics

kbroman.org/blog/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

genotypes
phenotypes
### Challenges: organizing, automating

<table>
<thead>
<tr>
<th>genotypes</th>
<th>phenotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Challenges: organizing, automating

<table>
<thead>
<tr>
<th>genotypes</th>
<th>phenotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Challenges: metadata

What the heck is "FAD_NAD SI 8.3_3.3G"?
What was the question again?
r/9t12
Now in 3D
- 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
Šaunak 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/pitt2021

kbroman.org

kbroman.org/qtl2

github.com/kbroman

@kwbroman