R/qtl: Just barely sustainable

Karl Broman

Biostatistics & Medical Informatics, UW–Madison

kbroman.org
github.com/kbroman
@kwbroman

Intercross
QTL mapping

![QTL mapping graph](image)
QTL mapping

![QTL mapping graph](image-url)
Good things
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**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<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></td>
<td></td>
<td></td>
<td></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></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>
<td>48.1</td>
<td>51.4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></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></td>
<td></td>
<td></td>
<td></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></td>
<td></td>
<td></td>
<td></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></td>
<td></td>
<td></td>
<td></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></td>
<td></td>
<td></td>
<td></td>
<td>65.31</td>
<td>181.05</td>
<td>m</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>SB</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></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></td>
<td></td>
<td></td>
<td></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></td>
<td></td>
<td></td>
<td></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></td>
<td></td>
<td></td>
<td></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></td>
<td></td>
<td></td>
<td></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></td>
<td></td>
<td></td>
<td></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></td>
<td></td>
<td></td>
<td></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>
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
More typically bad code

The `scantwo()` function is **1446 lines** long.

The related C code is 20% of the C code in R/qtl.
Baroque data structures

attr(mycross$geno[["X"]]+$probs, "map")
User support
“I tried X and it didn’t work.”
“Could you look at the attached 25-page Word document containing code and output and tell me if I’m doing something wrong?”
Incorporating others’ code
Version control
Tests
R/qtl2: Let’s not make the same mistakes

- C++ and Rcpp
- Roxygen2
- testthat
- A single “switch” for cross type
R/qtl2: Let’s not make the same mistakes

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

- Sustainable software
  - Easier if you are your primary user

- Take care of your users
  - Answer each question as if it’s the first
  - Primary goal is to be helpful
  - Write many focused vignettes solving real problems

- Version control

- Tests
Acknowledgments

▶ R/qtl users

▶ Hao Wu, Gary Churchill, Śaunak Sen, Pjotr Prins, Danny Arends, Brian Yandell, Robert Corty, Timothée Flutre, Ritsert Jansen, Lars Ronnegard, Rohan Shah, Laura Shannon, Quoc Tran, Aaron Wolen

▶ NIH/NIGMS