Seminar III: R/Bioconductor

"HardyWeinberg, an R Package that provides a graphical approach to
Hardy-Weinberg equilibrium"

Mariana Ruiz Velasco Leyva
Student of the Undergraduate Program on Genomic Sciences (LCG),
UNAM, Cuernavaca, Mexico
mruizvel@lcg.unam.mx
September 23, 2009

Abstract
Introducing into the usage of the HardyWeinberg package to improve the
visualization of a Hardy-Weinberg equilibrium test.

1 General Characteristics
As described in ’help(package = HardyWeinberg)’:

This package explores bi-allelic marker data focusing on the graphical representa-
tion of the results of tests for Hardy-Weinberg equilibrium.
Routines for several tests are included in this package.
The author is Jan Graffelman (who is also the maintainer).
Requires an R version >= to 1.8.0.
CRAN repository.
Some related packages are graphics and stats.
2 Why is it worthy to consider this package?

One of the most appealing features of HardyWeinberg package is that it contains both, functions that realize classical statistical tests and some other that help in the graphical representation of a 'de Fineti' or ternary plot.

In the first group we can find functions to calculate chi square value, conditional probability, and obviously the Hardy Weinberg equilibrium.

3 An easy example

We have a study where the following data saved in hardyw indicates the number of people containing the 3 possible alleles. We apply a Hardy -Weinberg equilibrium test.

```r
> library(HardyWeinberg)
> hardyw <- c(15, 24, 10)
> hwtest <- HWExact(hardyw, verbose = TRUE,
+                   singleterms = TRUE)
```

```
Exact test for Hardy-Weinberg equilibrium
sample counts: nAA = 15 nAB = 24 nBB = 10
H0: HWE (D=0), H1: D <> 0
D = -0.1224490 p = 1
```

Probabilities and statistics for all possible samples:

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AB</th>
<th>BB</th>
<th>Single term</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27</td>
<td>0</td>
<td>22</td>
<td>0.000000000</td>
<td>0.000000000</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>2</td>
<td>21</td>
<td>0.000000000</td>
<td>0.000000000</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>4</td>
<td>20</td>
<td>0.000000000</td>
<td>0.000000000</td>
</tr>
<tr>
<td>23</td>
<td>5</td>
<td>44</td>
<td>0</td>
<td>0.000000000</td>
<td>0.000000000</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>6</td>
<td>19</td>
<td>0.000000005</td>
<td>0.000000005</td>
</tr>
<tr>
<td>22</td>
<td>6</td>
<td>42</td>
<td>1</td>
<td>0.00000017</td>
<td>0.00000022</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>8</td>
<td>18</td>
<td>0.00000152</td>
<td>0.00000174</td>
</tr>
<tr>
<td>21</td>
<td>7</td>
<td>40</td>
<td>2</td>
<td>0.00000529</td>
<td>0.00000703</td>
</tr>
<tr>
<td>6</td>
<td>22</td>
<td>10</td>
<td>17</td>
<td>0.00002795</td>
<td>0.00003498</td>
</tr>
<tr>
<td>20</td>
<td>8</td>
<td>38</td>
<td>3</td>
<td>0.00008602</td>
<td>0.00012100</td>
</tr>
<tr>
<td>7</td>
<td>21</td>
<td>12</td>
<td>16</td>
<td>0.00031675</td>
<td>0.00043775</td>
</tr>
<tr>
<td>19</td>
<td>9</td>
<td>36</td>
<td>4</td>
<td>0.00083989</td>
<td>0.00127764</td>
</tr>
<tr>
<td>8</td>
<td>20</td>
<td>14</td>
<td>15</td>
<td>0.00233906</td>
<td>0.00361670</td>
</tr>
<tr>
<td>18</td>
<td>10</td>
<td>34</td>
<td>5</td>
<td>0.00529133</td>
<td>0.00890803</td>
</tr>
</tbody>
</table>

2
For the plot, we use the function HWData which generates data sets with both, the matrix of genotypic counts 'Xt' and the matrix with relative frequencies 'Xc', all keeping a Hardy-Weinberg equilibrium.

Two plots are created to see how the parameters can be changed while the plot is the same. The function to create a triangle plot is HwTernaryPlot.

```r
> p <- 100
> q <- 100
```
\[ \text{hweq} \leftarrow \text{HWData}(p, q) \]
\[ \text{xc} \leftarrow \text{hweq}$Xc \]
\[ \text{plotting} \leftarrow \text{HWternaryPlot}(Xc, 100, \]
\[ \quad \text{region} = 2, \text{hwcurve} = \text{FALSE}, \]
\[ \quad \text{vbounds} = \text{FALSE}, \text{signifcolour} = \text{TRUE}, \]
\[ \quad \text{curtyp} = \text{"dashed"}, \text{ssf} = \text{"min"}, \]
\[ \quad \text{main} = \text{"Hardy Weinberg Equilibrium Test"} ) \]

**Hardy Weinberg Equilibrium Test**

Note that green points represent the non-significant region of a Chi-square test and the red markers represent the significant.

\[ \text{plots} \leftarrow \text{HWternaryPlot}(Xc, 100, \]
\[ \quad \text{region} = 2, \text{hwcurve} = \text{TRUE}, \]
\[ \quad \text{vbounds} = \text{TRUE}, \text{signifcolour} = \text{TRUE}, \]
\[ \quad \text{ssf} = \text{"min"}, \text{curtyp} = \text{"solid"}, \]
\[ \quad \text{curvecols} = \text{"blue"}, \text{markerlab} = \text{"*"}) \]
If you wish to learn more about this package you can read the following paper: Graffelman, J. and Morales-Camarena, J. (2008) Graphical tests for Hardy-Weinberg equilibrium based on the ternary plot. Human Heredity 65(2): 77-84.