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 representation 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 mantainer).

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.

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

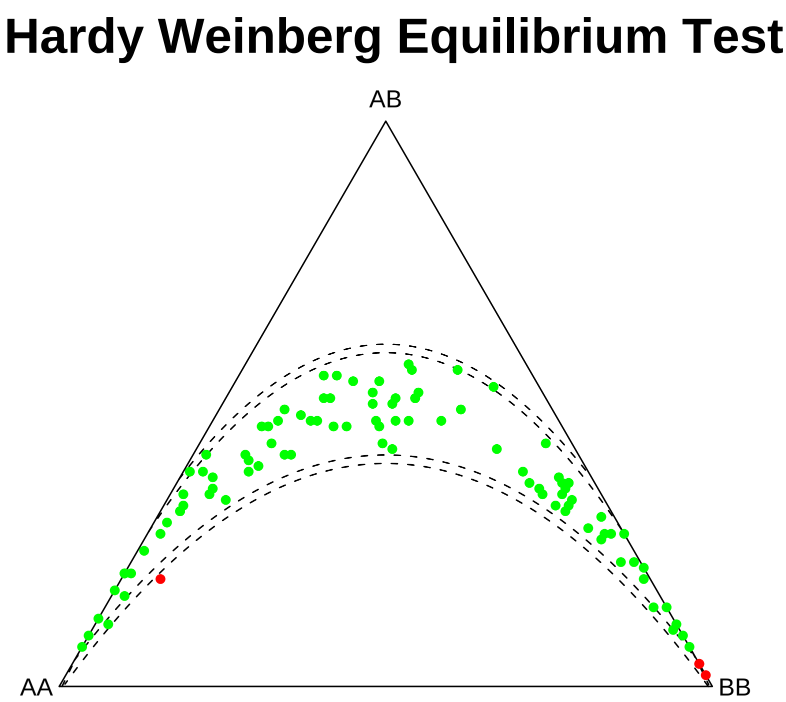
```
Exact test for Hardy-Weinberg equilibrium
sample counts: nAA =  15 nAB =  24 nBB =  10
H0: HWE (D=0), H1: D <> 0
D =  -0.1224490 p =  1

Probabilities and statistics for all possible samples:
   AA AB BB Single term       Prob
1  27  0 22  0.00000000 0.00000000
2  26  2 21  0.00000000 0.00000000
3  25  4 20  0.00000000 0.00000000
23  5 44  0  0.00000000 0.00000000
4  24  6 19  0.00000005 0.00000005
22  6 42  1  0.00000017 0.00000022
5  23  8 18  0.00000152 0.00000174
21  7 40  2  0.00000529 0.00000703
6  22 10 17  0.00002795 0.00003498
20  8 38  3  0.00008602 0.00012100
7  21 12 16  0.00031675 0.00043775
19  9 36  4  0.00083989 0.00127764
8  20 14 15  0.00233906 0.00361670
18 10 34  5  0.00529133 0.00890803
```

```
9  19 16 14  0.01169529 0.02060332
17 11 32  6  0.02248815 0.04309147
10 18 18 13  0.04066599 0.08375746
16 12 30  7  0.06639358 0.15015104
11 17 20 12  0.10016675 0.25031779
15 13 28  8  0.13885195 0.38916974
12 16 22 11  0.17691790 0.56608764
14 14 26  9  0.20827793 0.77436557
13 15 24 10  0.22563442 0.99999999
            X2       pval           D
1  49.00000000 0.00000000 -12.1224490
2  41.24926311 0.00000000 -11.1224490
3  34.16540262 0.00000001 -10.1224490
23 32.53223594 0.00000001   9.8775510
4  27.74841853 0.00000014  -9.1224490
22 26.27856851 0.00000030   8.8775510
5  21.99831083 0.00000273  -8.1224490
21 20.69177748 0.00000539   7.8775510
6  16.91507953 0.00003909  -7.1224490
20 15.77186285 0.00007146   6.8775510
7  12.49872462 0.00040723  -6.1224490
19 11.51882461 0.00068895   5.8775510
8   8.74924611 0.00309730  -5.1224490
18  7.93266277 0.00485503   4.8775510
9   5.66664399 0.01729050  -4.1224490
17  5.01337732 0.02515219   3.8775510
10  3.25091827 0.07138346  -3.1224490
16  2.76096827 0.09658976   2.8775510
11  1.50206895 0.22035329  -2.1224490
15  1.17543561 0.27828718   1.8775510
12  0.42009602 0.51688914  -1.1224490
14  0.25677935 0.61234147   0.8775510
13  0.00499949 0.94363089  -0.1224490
```

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.

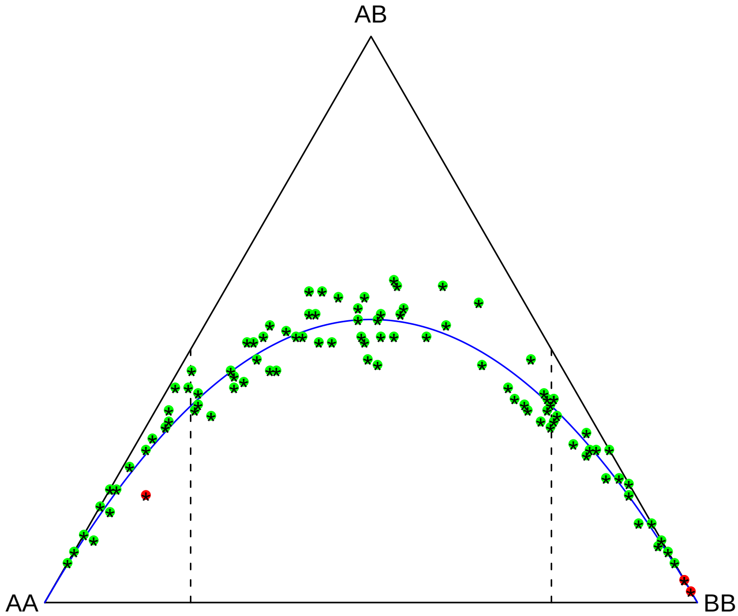
```
> p <- 100
> q <- 100
```

```
> hweq <- HWData(p, q)
> xc <- hweq$Xc
> plotting <- HWTernaryPlot(Xc, 100,
+     region = 2, hwcurve = FALSE,
+     vbounds = FALSE, signifcolour = TRUE,
+     curtyp = "dashed", ssf = "min",
+     main = "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.

```
> plots <- HWTernaryPlot(Xc, 100,
+     region = 2, hwcurve = TRUE,
+     vbounds = TRUE, signifcolour = TRUE,
+     ssf = "min", curtyp = "solid",
+     curvecols = "blue", markerlab = "*")
```



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.