Generating Marker Summary Reports Using the *GeneticsBase*

Gregory Warnes
`gregory_warnesurmc.rochester.edu`,
Nitin Jain
`nitin.jainpfizer.com`

April 25, 2007

1 Introduction

This document demonstrates how to use the *GeneticsBase* (version 0.6.0) package to generate marker summary tables. It is written as a step-by-step tutorial. For additional details on each of the R functions utilized, please see the individual help pages

2 Example

2.1 Prepare phenotype data

The first step is to prepare the phenotype data. It may be in the form of a SAS dataset, SAS export file, comma-delimited text file (CSV), tab-delimited text file (TSV), or Microsoft Excel spreadsheet file (XLS). It should have one row per observation and one column per variable, and must contain a subject identifier variable that can be used to match observations with the corresponding genotype data.

2.2 Prepare genotype data

You also need to store the genetic call data in a file that can be read into R. GeneticsBase packge accepts genotype data in a variety of formats:

- standard pedigree (ped) format.

```
a2m apoe
    50103    5010004    5090005    5090004          2          2          1
      2          3          4
    50103    5010005    5090005    5090004          2          2          1
      1          3          4
    50105    5010049    5090021    5090022          2          2          1
      1          4          4
    50105    5010070    5090020    5090019          1          2          1
      1          3          4
```

- hapmap format : The hapmap .ped format is a variant of the standard pedigree format. A portion of the first two lines of the hapmap file for chrmosome 1 are shown below:

```
rs2298011 rs1320571 rs11721 rs4018608 rs6685064 rs604618 ....
1347 14 0 0 1 1 1 1 3 3 3 3 1 1 2 2 3 3 3 3 2 2 3 3 2  .....
```

- Pfizer format: First few lines of an example file in Pfizer's data format are shown below:

```
Locus,Gene Marker,Locus Start,Project,Protocol,Sample ID,Donor ID,Genotype
A1A,C1556G,-1243,P234,103,1028022.1,1028022,G/G
A1B,T127A,20141,P234,103,1028022.1,1028022,A/T
A1B,T5094A,102358,P234,103,1028022.1,1028022,A/T
A1A,C1556G,-1243,P234,103,1035130.1,1035130,G/G
```

- Perlegen format: A portion of first two lines of data in the Perlegen format are shown below:

```
snp_id  genotype
753527  rrrrrrrrrrrrrrrhrrrrrrrrrrrrrrrrrrrrrrrrr......
752848  hhahaararhhhahrhhhahaharhaahhahrhhahhahnhha......
```

2.3 Load the genotype and phenotype data

In GeneticsBase, both genotype and phentype data is loaded by a single function *readgenes*. This function *readGenes* has four primary arguments: *gfile*, *gfromat*, *file*, and *pformat*. Arguments *gfile* and *pfile* are the names of files containing the genotype and phenotype data (respectively), and arguments *gformat* and *pformat* are the corresponding file formats for the genotype and phenotype data.

Various types of data can be loaded by readGenes function. Example files are provided in `data` subdirectory of the `GeneticsBase` package. To access these execute

```
> library(GeneticsBase)
> setwd(file.path(.path.package("GeneticsBase"), "data"))
```

Supported file formats include:

- fbat .ped format

  The Alzheimer's example dataset is stored in the Fbat variant of the .ped Pedigree Format. As it does not include phenotype data, we only use the genotyep filename and file type arguments:

  ```
  > ALZH <- readGenes(gfile = "ALZH.ped", gformat = "fbat.ped")
  ```

  The CAMP example dataset is from from the 'Childhood Asthma Management Program (CAMP)' and incudes both genotype and phenotype information. It can be loaded by:

  ```
  > CAMP <- readGenes(gfile = "CAMP.ped", gformat = "fbat.ped", pfile = "CAMPZ.phe",
  +     pformat = "fbat.phe")
  ```

- HAPMAP .ped format

  A subset of the data for the International HapMap project is available in the hapmap example data set. This file can be loaded via:

  ```
  > hapmapchr1 <- readGenes(gfile = "hapmapchr1.ped", gformat = "hapmap")
  ```

- Pfizer genetics data format

  ```
  > PfizerExample <- readGenes.pfizer("PfizerExample.txt", format = "Listing")
  ```

- Perlegen data format

  ```
  > PerlegenExample <- readGenes("PerlegenExample.txt", gformat = "perlegen")
  ```

2.4 Reviewing the data

For the purpose of this example, we are going to use CAMP data set, which can be loaded manually as shown in the previous sub-section, or via

```
> library(GeneticsBase)
> data(CAMP)

Reading 8 markers and 2011 subjects from ` CAMP.ped ' ...
generating 'geneSet' object...

Successfully read the pedigree file ` CAMP.ped '.

Number of Markers:  8
Number of Subjects: 2011
Number of Families: 651


Reading 12 vars from `CAMPZ.phe' ... Done.

Number of Phenotype Variables: 12
Number of Observations       : 2011
```

Now you can see a brief summary of the data that was loaded by simply entering the name of the object on a line by itself:

```
> CAMP

geneSet object
------------------

Number of Markers:      8
Number of Observations: 2011

Sample variables: family, pid, father, mother, sex, affected, zposfevp,                  zposfvcp, zlog

Genetic data:

     1.1900 1.1667 1.978 2.1391 2.1988 2.109     649.1837 650.1736 650.1908
m709 1/1    1/1    1/1   1/1    1/1    1/1   ... 1/1      1/1      1/1
m654 1/1    1/1    1/1   1/1    2/1    1/1   ... 2/1      <NA>     1/1
m47  1/1    1/2    1/2   1/2    2/2    1/2   ... 2/2      1/1      1/2
p46  2/2    2/2    2/2   2/2    1/2    2/2   ... 1/2      2/2      2/2
p79  2/2    2/1    2/1   2/1    1/1    2/1   ... 1/1      <NA>     2/1
p252 2/2    1/2    1/2   1/2    1/2    <NA>  ... 1/2      2/2      1/2
p491 1/1    1/1    1/1   1/1    1/1    1/1   ... 1/1      <NA>     1/1
p523 1/1    1/2    1/2   1/2    1/2    1/2   ... 1/2      1/1      1/2
     650.1675 651.568 651.1725
m709 1/1      1/1     1/1
m654 1/1      2/2     2/1
m47  1/2      2/2     2/2
p46  1/2      1/1     1/2
p79  2/1      1/1     1/1
```

```
p252 1/2      2/2     1/2
p491 1/1      1/1     1/1
p523 1/2      1/1     1/2

Warning messages:
1: geneSet Object has 121 observations.  Only first and last 6 displayed
in: .local(object, ...)
```

The phenotype data can be extracted from the CAMP data object using the `sampleInfo` command:

```
> pdata <- sampleInfo(CAMP)
> summary(pdata)

     family           pid             father           mother
 Min.   :  1.0   Min.   :   1.0   Min.   :   0.0   Min.   :   0.0
 1st Qu.:165.0   1st Qu.: 503.5   1st Qu.:   0.0   1st Qu.:   0.0
 Median :327.0   Median :1006.0   Median :   0.0   Median :   0.0
 Mean   :326.9   Mean   :1006.0   Mean   : 340.9   Mean   : 367.7
 3rd Qu.:489.0   3rd Qu.:1508.5   3rd Qu.: 521.0   3rd Qu.: 637.5
 Max.   :651.0   Max.   :2011.0   Max.   :2009.0   Max.   :2010.0

      sex           affected         zposfevp             zposfvcp
 Min.   :1.000   Min.   :0.0000   Min.   :-3.234e+00   Min.   :-2.880e+00
 1st Qu.:1.000   1st Qu.:0.0000   1st Qu.:-6.790e-01   1st Qu.:-6.250e-01
 Median :1.000   Median :0.0000   Median : 1.000e-03   Median :-1.600e-02
 Mean   :1.453   Mean   :0.7041   Mean   : 1.431e-05   Mean   :-3.433e-05
 3rd Qu.:2.000   3rd Qu.:2.0000   3rd Qu.: 6.275e-01   3rd Qu.: 6.005e-01
 Max.   :2.000   Max.   :2.0000   Max.   : 4.021e+00   Max.   : 4.041e+00

...
```

2.5 Generate the tables

We can generate a variety of summary tables on our genetics data.

- Allele information

  ```
  > alleleSummary(CAMP)

   Gene Marker Position Group Allele Count  Freq  CI-Lower CI-Upper
   ALL  m709   ?        ALL   1       2536 0.998 0.996    0.999
                        ALL   2          6 0.002 0.001    0.004
                        ALL   NA         0

        m654   ?        ALL   1       1618 0.640 0.621    0.658
                        ALL   2        912 0.360 0.342    0.379
                        ALL   NA         0

        m47    ?        ALL   1        935 0.374 0.355    0.393
                        ALL   2       1565 0.626 0.607    0.645
                        ALL   NA         0
  ```

```
      p46    ?        ALL   1        1000 0.400 0.380    0.419

...
```

- Genotype information

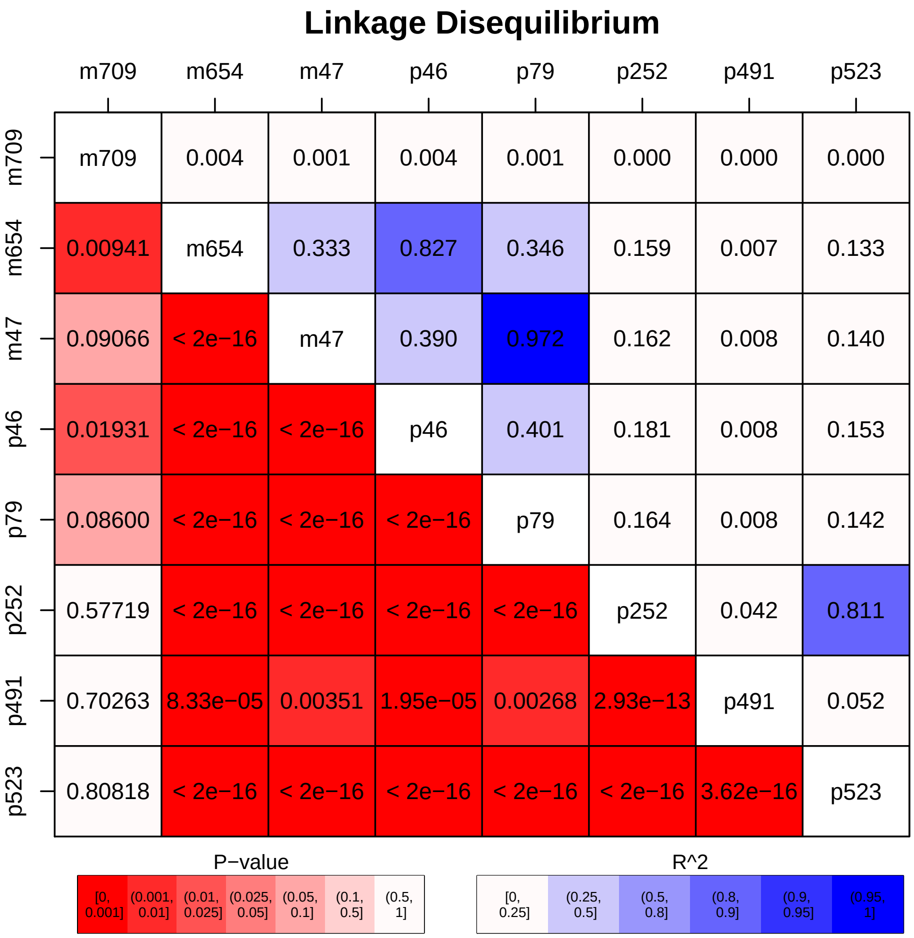
```
> genotypeSummary(CAMP)

 Gene Marker Position Group Genotype Count  Freq  CI-Lower CI-Upper Expected
 ?    m709   ?        ALL   1/1        1265 0.995 0.991    0.998    1265.007
                            1/2           6 0.005 0.002    0.009    5.986
                            2/2           0 0.000                   0.007
                            NA            0

 ?    m654   ?        ALL   1/1         537 0.425 0.398    0.452    517.376
                            1/2         544 0.430 0.403    0.458    583.247
                            2/2         184 0.145 0.126    0.165    164.376
                            NA            0

 ?    m47    ?        ALL   1/1         175 0.140 0.121    0.159    174.845
                            1/2         585 0.468 0.440    0.496    585.310
                            2/2         490 0.392 0.365    0.419    489.845

...
```

- Marker information
- Linkage disequilibrium, text view

```
> ld <- LD(CAMP)
> ld

-----------
Pairwise LD
-----------
            m709   m654    m47     p46     p79    p252    p491    p523
m709 D            0.002 -0.001   0.001  -0.001   0.000   0.000   0.000
m709 D'           1.000  0.999   1.000   0.999   0.641   1.000   0.288
m709 Corr.        0.066 -0.038   0.061  -0.039  -0.018  -0.005  -0.007
m709 R^2          0.004  0.001   0.004   0.001   0.000   0.000   0.000
     LD X^2       6.000  2.000   5.000   2.000   0.000   0.000   0.000
     P-value    0.00941 0.0907  0.0193   0.086   0.577   0.703   0.808
m709 LOD          1.464  0.622   1.188   0.640   0.067   0.032   0.013
m709 n             1303   1303    1303    1303    1303    1303    1303

m654 D                  -0.134   0.214  -0.137  -0.080  -0.004  -0.069

...
```

- Linkage disequilibrium, matrix plot

```
> plot(ld)
```

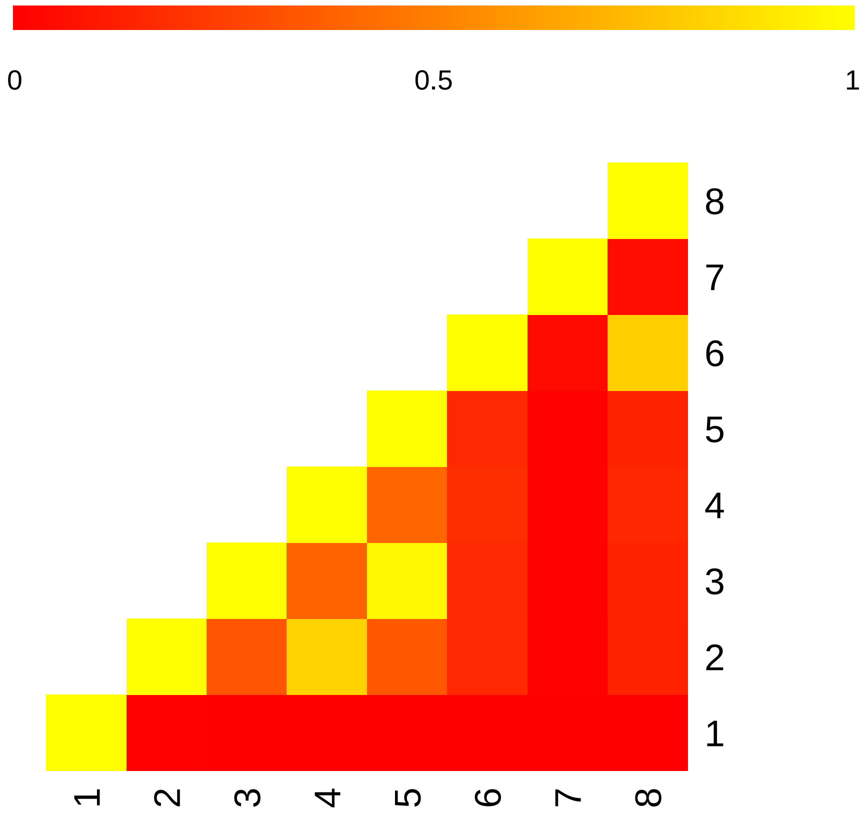


Linkage Disequilibrium

	m709	m654	m47	p46	p79	p252	p491	p523
m709	m709	0.004	0.001	0.004	0.001	0.000	0.000	0.000
m654	0.00941	m654	0.333	0.827	0.346	0.159	0.007	0.133
m47	0.09066	< 2e−16	m47	0.390	0.972	0.162	0.008	0.140
p46	0.01931	< 2e−16	< 2e−16	p46	0.401	0.181	0.008	0.153
p79	0.08600	< 2e−16	< 2e−16	< 2e−16	p79	0.164	0.008	0.142
p252	0.57719	< 2e−16	< 2e−16	< 2e−16	< 2e−16	p252	0.042	0.811
p491	0.70263	8.33e−05	0.00351	1.95e−05	0.00268	2.93e−13	p491	0.052
p523	0.80818	< 2e−16	< 2e−16	< 2e−16	< 2e−16	< 2e−16	3.62e−16	p523

- Linkage disequilibrium, graphical view using LDView

```
> r2 <- t(ld@"R^2")
> diag(r2) <- 1
> LDView(r2, labRow = markerNames(CAMP))
```



3 Generating tables for inclusion in reports

To make it simple to include the summary tables in written reports, they can be written to files in a variety of formats, including plain text, html, and LaTeX.

3.1 Plain text files

```
> aS <- alleleSummary(CAMP)
> txt(aS, file = "CAMP_alleleSummary.txt")
```

3.2 LaTex files

```
> aS <- alleleSummary(CAMP)
> latex(aS)
```

	Gene	Marker	Position	Group	Allele	Count	Freq	CI-Lower	CI-Upper
1	ALL	m709	?	ALL	1	2536	0.998	0.996	0.999
2				ALL	2	6	0.002	0.001	0.004
3				ALL	NA	0			
4									
5		m654	?	ALL	1	1618	0.640	0.621	0.658
6				ALL	2	912	0.360	0.342	0.379
7				ALL	NA	0			
8									
9		m47	?	ALL	1	935	0.374	0.355	0.393
10				ALL	2	1565	0.626	0.607	0.645
11				ALL	NA	0			
12									
13		p46	?	ALL	1	1000	0.400	0.380	0.419
14				ALL	2	1502	0.600	0.581	0.620
15				ALL	NA	0			
16									
17		p79	?	ALL	1	1544	0.621	0.601	0.639
18				ALL	2	944	0.379	0.361	0.399
19				ALL	NA	0			
20									
21		p252	?	ALL	1	517	0.221	0.205	0.238
22				ALL	2	1819	0.779	0.762	0.795
23				ALL	NA	0			
24									
25		p491	?	ALL	1	2505	0.988	0.983	0.992
26				ALL	2	31	0.012	0.008	0.017
27				ALL	NA	0			
28									
29		p523	?	ALL	1	2057	0.810	0.795	0.825
30				ALL	2	481	0.190	0.175	0.205
31				ALL	NA	0			
32									

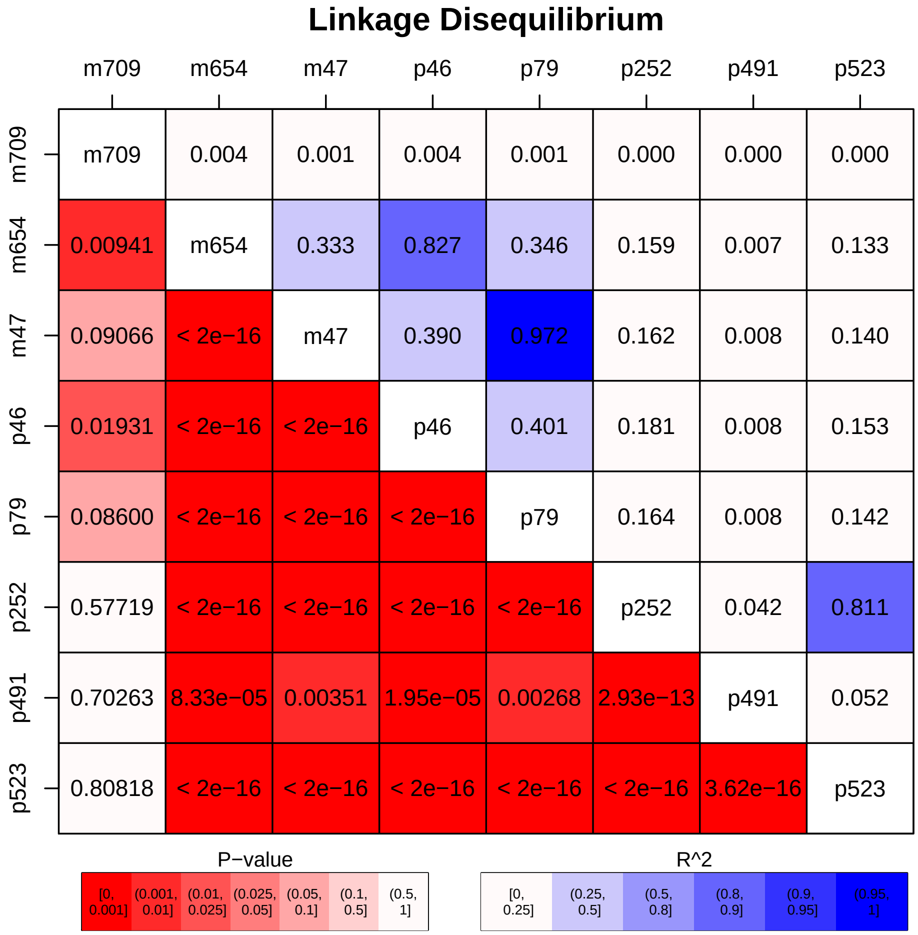
```
> gs <- genotypeSummary(CAMP[-2, ])
> latex(gS)
```

	Gene	Marker	Position	Group	Genotype	Count	Freq	CI-Lower	CI-Upper	Expected	Obs-Exp	HWE X\^2	P-value
1	?	m709	?	ALL	1/1	1265	0.995	0.991	0.998	1265.007	-0.007	0.007	1
2					1/2	6	0.005	0.002	0.009	5.986	0.014		
3					2/2	0	0.000			0.007	-0.007		
4					NA	0							
5													
6	?	m47	?	ALL	1/1	175	0.140	0.121	0.159	174.845	0.155	0.000	1
7					1/2	585	0.468	0.440	0.496	585.310	-0.310		
8					2/2	490	0.392	0.365	0.419	489.845	0.155		
9					NA	0							
10													
11	?	p46	?	ALL	1/1	209	0.167	0.146	0.188	199.840	9.160	1.165	0.288
12					1/2	582	0.465	0.437	0.493	600.320	-18.320		
13					2/2	460	0.368	0.341	0.395	450.840	9.160		
14					NA	0							
15													
16	?	p79	?	ALL	1/1	486	0.391	0.363	0.418	479.087	6.913	0.693	0.437
17					1/2	572	0.460	0.432	0.487	585.826	-13.826		
18					2/2	186	0.150	0.130	0.170	179.087	6.913		
19					NA	0							
20													
21	?	p252	?	ALL	1/1	70	0.060	0.047	0.074	57.211	12.789	4.715	0.0331
22					1/2	377	0.323	0.296	0.349	402.578	-25.578		
23					2/2	721	0.617	0.589	0.645	708.211	12.789		
24					NA	0							
25													
26	?	p491	?	ALL	1/1	1237	0.976	0.967	0.983	1237.189	-0.189	0.194	1
27					1/2	31	0.024	0.017	0.033	30.621	0.379		
28					2/2	0	0.000			0.189	-0.189		
29					NA	0							
30													
31	?	p523	?	ALL	1/1	840	0.662	0.636	0.688	833.579	6.421	1.377	0.241
32					1/2	377	0.297	0.272	0.322	389.841	-12.841		
33					2/2	52	0.041	0.031	0.052	45.579	6.421		
34					NA	0							
35													

```
> ld <- LD(CAMP)
> latex(ld)
```

	m709	m654	m47	p46	p79	p252	p491	p523
m709 D		0.002	-0.001	0.001	-0.001	0.000	0.000	0.000
m709 D'		1.000	0.999	1.000	0.999	0.641	1.000	0.288
m709 Corr.		0.066	-0.038	0.061	-0.039	-0.018	-0.005	-0.007
m709 R^2		0.004	0.001	0.004	0.001	0.000	0.000	0.000
LD χ^2		6	2	5	2	0	0	0
P-value		0.00941	0.09066	0.01931	0.08600	0.57719	0.70263	0.80818
m709 LOD		1.464	0.622	1.188	0.640	0.067	0.032	0.013
m709 n		1303	1303	1303	1303	1303	1303	1303
m654 D			-0.134	0.214	-0.137	-0.080	-0.004	-0.069
m654 D'			0.992	0.979	0.996	0.991	1.000	1.000
m654 Corr.			-0.577	0.909	-0.588	-0.399	-0.084	-0.365
m654 R^2			0.333	0.827	0.346	0.159	0.007	0.133
LD χ^2			620	1758	676	303	15	277
P-value			$< 2e-16$	$< 2e-16$	$< 2e-16$	$< 2e-16$	8.33e-05	$< 2e-16$
m654 LOD			134.831	381.769	146.932	65.982	3.362	60.366
m654 n			1303	1303	1303	1303	1303	1303
m47 D				-0.148	0.231	-0.081	-0.005	-0.071
m47 D'				0.996	0.991	0.973	0.998	0.990
m47 Corr.				-0.625	0.986	-0.403	-0.087	-0.374
m47 R^2				0.390	0.972	0.162	0.008	0.140
LD χ^2				750	2240	267	8	245
P-value				$< 2e-16$	$< 2e-16$	$< 2e-16$	0.00351	$< 2e-16$
m47 LOD				162.928	486.549	58.183	1.851	53.335
m47 n				1303	1303	1303	1303	1303
p46 D					-0.150	-0.087	-0.005	-0.075
p46 D'					0.996	0.977	1.000	0.991
p46 Corr.					-0.633	-0.426	-0.090	-0.391
p46 R^2					0.401	0.181	0.008	0.153
LD χ^2					783	319	18	283
P-value					$< 2e-16$	$< 2e-16$	1.95e-05	$< 2e-16$
p46 LOD					170.105	69.424	3.960	61.586
p46 n					1303	1303	1303	1303
p79 D						-0.081	-0.005	-0.072
p79 D'						0.974	0.999	1.000
p79 Corr.						-0.404	-0.088	-0.377
p79 R^2						0.164	0.008	0.142
LD χ^2						275	9	266
P-value						$< 2e-16$	0.00268	$< 2e-16$
p79 LOD						59.797	1.958	57.918
p79 n						1303	1303	1303
p252 D							0.009	0.146
p252 D'							1.000	0.994
p252 Corr.							0.204	0.901
p252 R^2							0.042	0.811
LD χ^2							53	1321
P-value							2.93e-13	$< 2e-16$
p252 LOD							11.564	286.890
p252 n							1303	1303
p491 D								0.010
p491 D'								1.000
p491 Corr.								0.228
p491 R^2								0.052
LD χ^2								66
P-value								3.62e-16
p491 LOD								14.426
p491 n								1303
p523 D								
p523 D'								
p523 Corr.								
p523 R^2								
LD χ^2								
P-value								
p523 LOD								
p523 n								

```
> plot(ld)
```



	m709	m654	m47	p46	p79	p252	p491	p523
m709	m709	0.004	0.001	0.004	0.001	0.000	0.000	0.000
m654	0.00941	m654	0.333	0.827	0.346	0.159	0.007	0.133
m47	0.09066	< 2e−16	m47	0.390	0.972	0.162	0.008	0.140
p46	0.01931	< 2e−16	< 2e−16	p46	0.401	0.181	0.008	0.153
p79	0.08600	< 2e−16	< 2e−16	< 2e−16	p79	0.164	0.008	0.142
p252	0.57719	< 2e−16	< 2e−16	< 2e−16	< 2e−16	p252	0.042	0.811
p491	0.70263	8.33e−05	0.00351	1.95e−05	0.00268	2.93e−13	p491	0.052
p523	0.80818	< 2e−16	< 2e−16	< 2e−16	< 2e−16	< 2e−16	3.62e−16	p523

3.3 HTML files

3.4 Graphics files

As usual, plots can be generated in any format R supports.

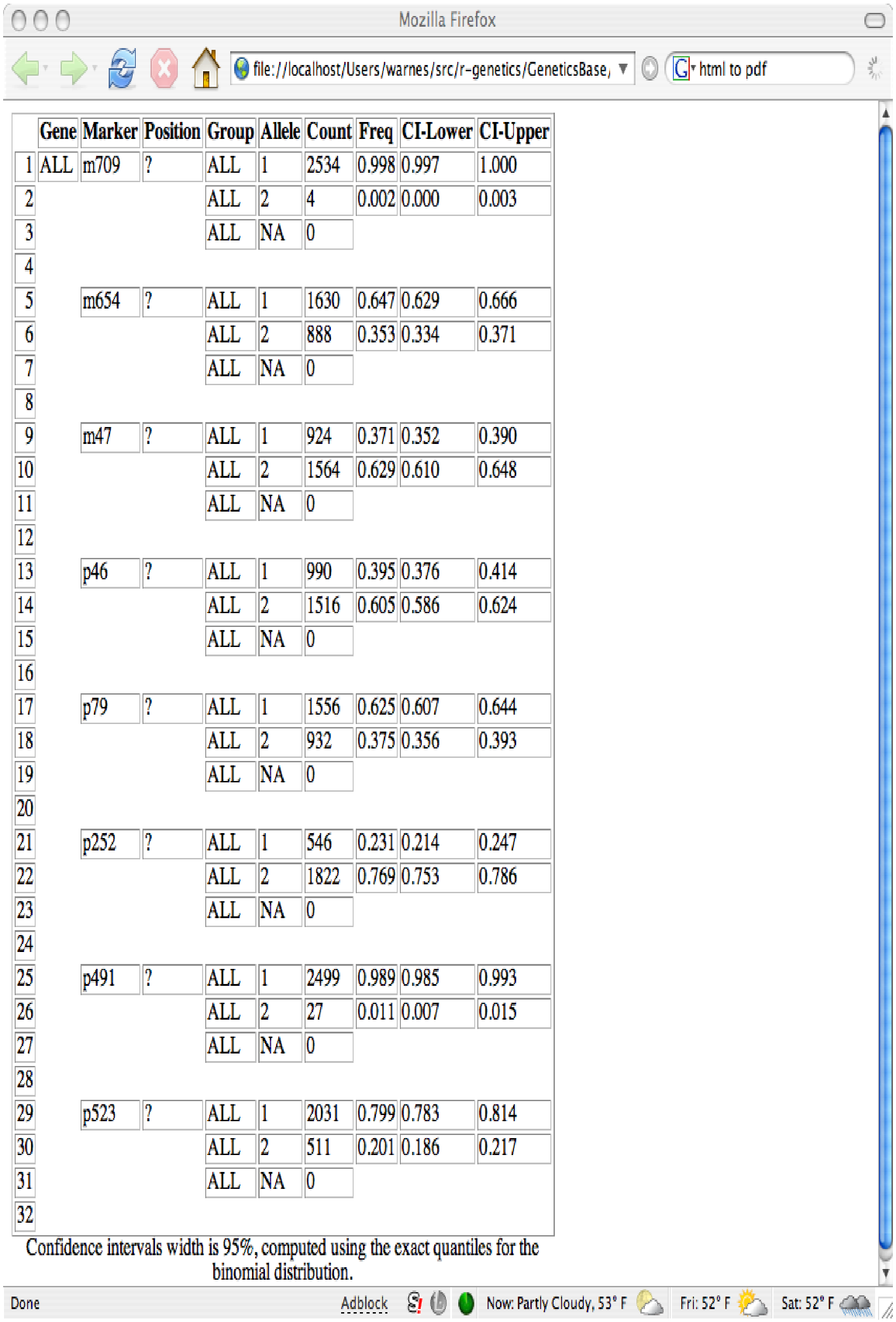
We can also output everything all at once to a set of files, encoded as plain text (format="print"), html (format="html"), or LaTeX (format="latex"):

```
> PGtables(CAMP, filename = "CAMP", sep = "_", format = "html")
```

```
Creating  CAMP_alleleSummary.html ...
Creating  CAMP_genotypeSummary.html ...
Creating  CAMP_LD.html ...
Creating  CAMP_LD.pdf ...
Done.
```

which creates a set of html and a PDF files in the current directory.

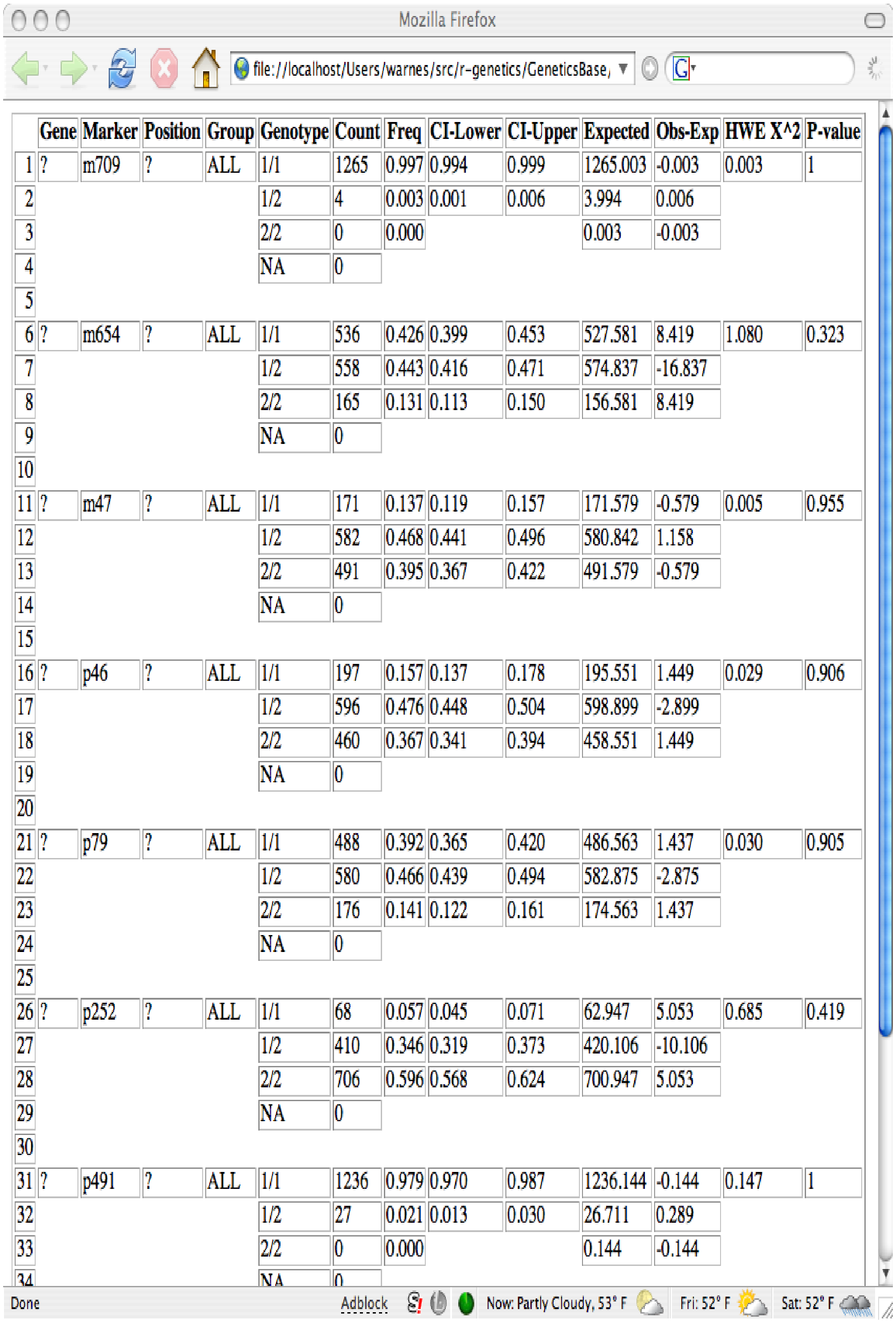
Figure 1: HTML allele summary table



	Gene	Marker	Position	Group	Allele	Count	Freq	CI-Lower	CI-Upper
1	ALL	m709	?	ALL	1	2534	0.998	0.997	1.000
2				ALL	2	4	0.002	0.000	0.003
3				ALL	NA	0			
4									
5		m654	?	ALL	1	1630	0.647	0.629	0.666
6				ALL	2	888	0.353	0.334	0.371
7				ALL	NA	0			
8									
9		m47	?	ALL	1	924	0.371	0.352	0.390
10				ALL	2	1564	0.629	0.610	0.648
11				ALL	NA	0			
12									
13		p46	?	ALL	1	990	0.395	0.376	0.414
14				ALL	2	1516	0.605	0.586	0.624
15				ALL	NA	0			
16									
17		p79	?	ALL	1	1556	0.625	0.607	0.644
18				ALL	2	932	0.375	0.356	0.393
19				ALL	NA	0			
20									
21		p252	?	ALL	1	546	0.231	0.214	0.247
22				ALL	2	1822	0.769	0.753	0.786
23				ALL	NA	0			
24									
25		p491	?	ALL	1	2499	0.989	0.985	0.993
26				ALL	2	27	0.011	0.007	0.015
27				ALL	NA	0			
28									
29		p523	?	ALL	1	2031	0.799	0.783	0.814
30				ALL	2	511	0.201	0.186	0.217
31				ALL	NA	0			
32									

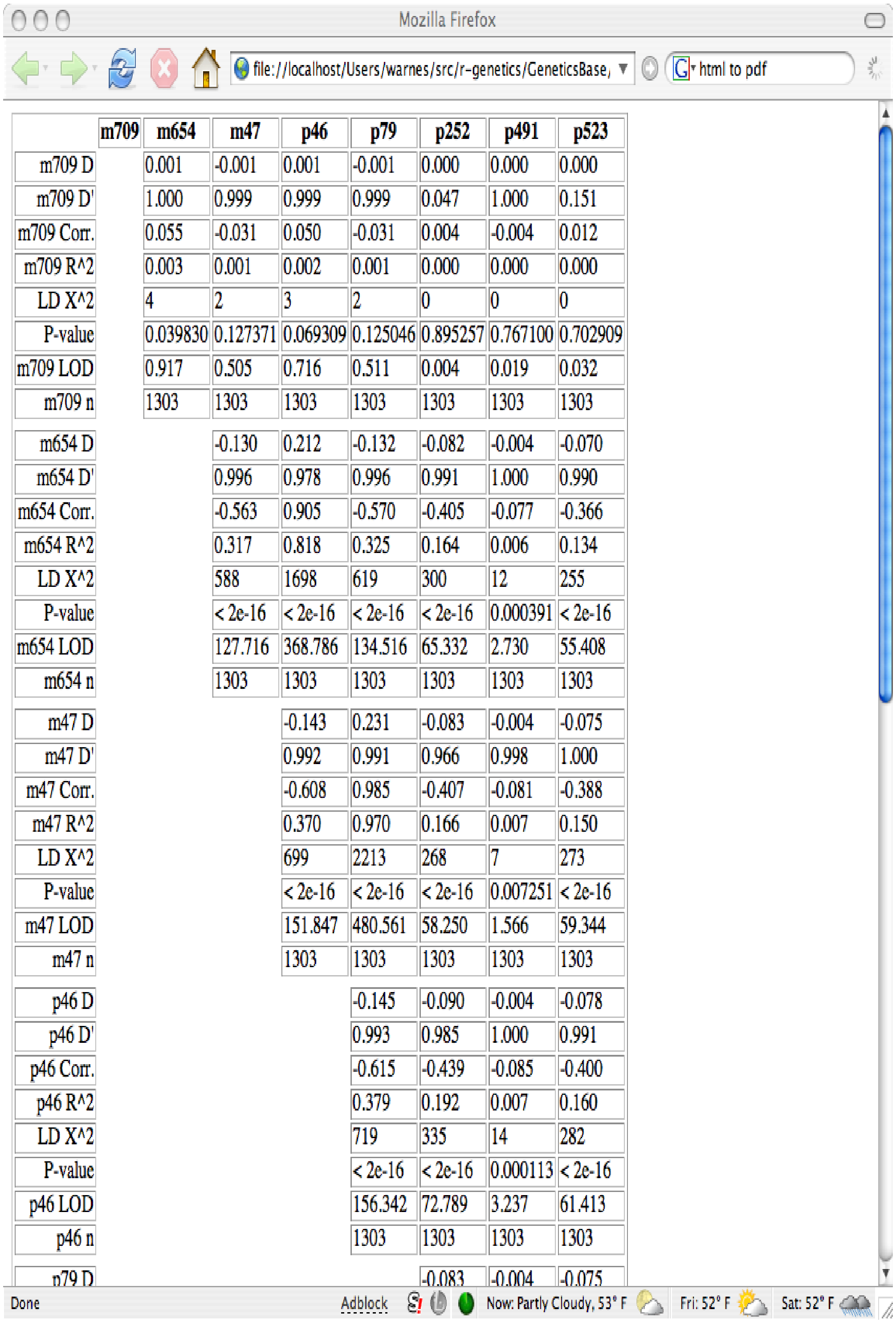
Confidence intervals width is 95%, computed using the exact quantiles for the binomial distribution.

Figure 2: HTML genotype summary table



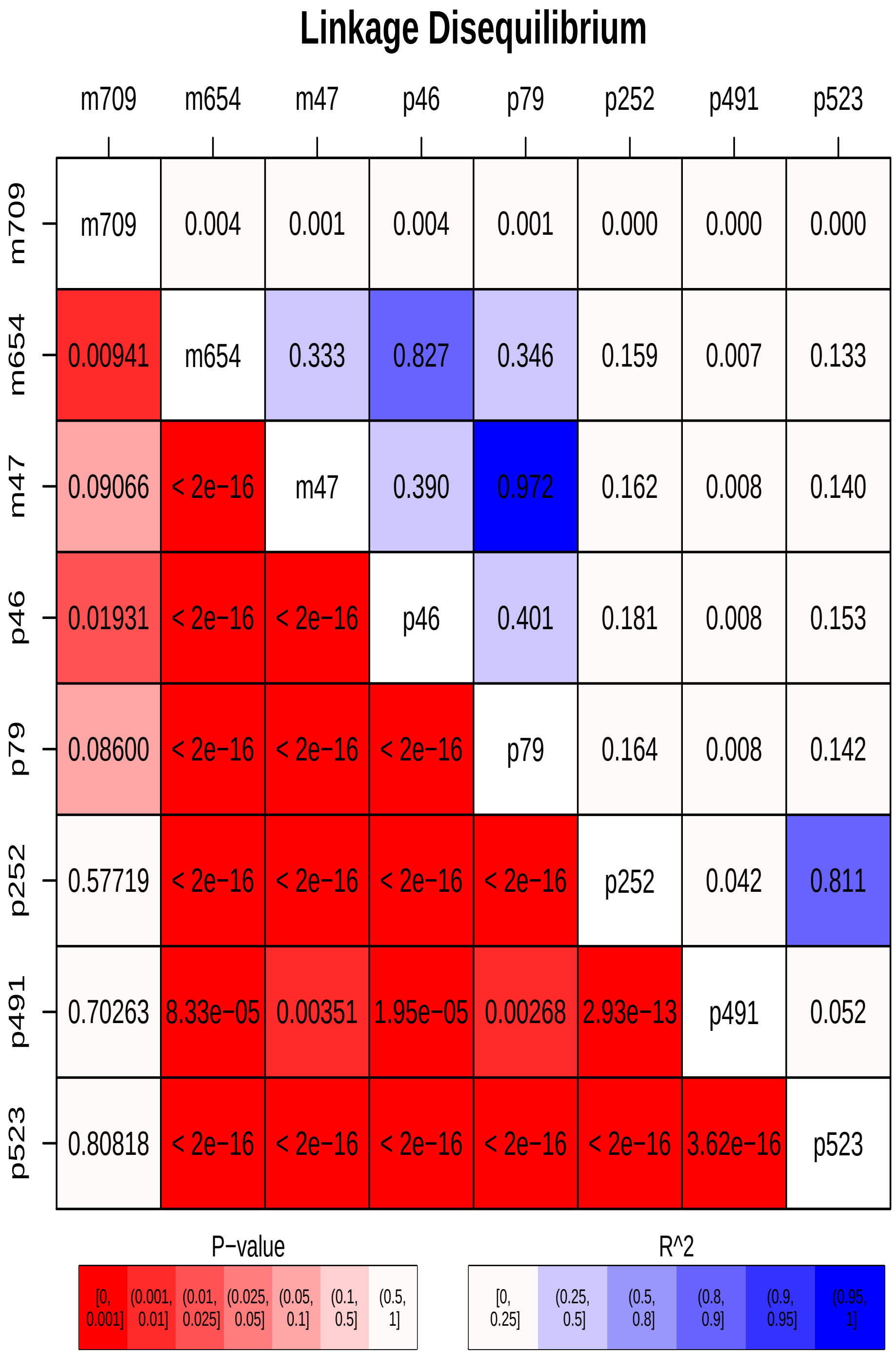
	Gene	Marker	Position	Group	Genotype	Count	Freq	CI-Lower	CI-Upper	Expected	Obs-Exp	HWE X^2	P-value
1	?	m709	?	ALL	1/1	1265	0.997	0.994	0.999	1265.003	-0.003	0.003	1
2					1/2	4	0.003	0.001	0.006	3.994	0.006		
3					2/2	0	0.000			0.003	-0.003		
4					NA	0							
5													
6	?	m654	?	ALL	1/1	536	0.426	0.399	0.453	527.581	8.419	1.080	0.323
7					1/2	558	0.443	0.416	0.471	574.837	-16.837		
8					2/2	165	0.131	0.113	0.150	156.581	8.419		
9					NA	0							
10													
11	?	m47	?	ALL	1/1	171	0.137	0.119	0.157	171.579	-0.579	0.005	0.955
12					1/2	582	0.468	0.441	0.496	580.842	1.158		
13					2/2	491	0.395	0.367	0.422	491.579	-0.579		
14					NA	0							
15													
16	?	p46	?	ALL	1/1	197	0.157	0.137	0.178	195.551	1.449	0.029	0.906
17					1/2	596	0.476	0.448	0.504	598.899	-2.899		
18					2/2	460	0.367	0.341	0.394	458.551	1.449		
19					NA	0							
20													
21	?	p79	?	ALL	1/1	488	0.392	0.365	0.420	486.563	1.437	0.030	0.905
22					1/2	580	0.466	0.439	0.494	582.875	-2.875		
23					2/2	176	0.141	0.122	0.161	174.563	1.437		
24					NA	0							
25													
26	?	p252	?	ALL	1/1	68	0.057	0.045	0.071	62.947	5.053	0.685	0.419
27					1/2	410	0.346	0.319	0.373	420.106	-10.106		
28					2/2	706	0.596	0.568	0.624	700.947	5.053		
29					NA	0							
30													
31	?	p491	?	ALL	1/1	1236	0.979	0.970	0.987	1236.144	-0.144	0.147	1
32					1/2	27	0.021	0.013	0.030	26.711	0.289		
33					2/2	0	0.000			0.144	-0.144		
34					NA	0							

Figure 3: HTML linkage disequilibrium table



	m709	m654	m47	p46	p79	p252	p491	p523
m709 D		0.001	-0.001	0.001	-0.001	0.000	0.000	0.000
m709 D'		1.000	0.999	0.999	0.999	0.047	1.000	0.151
m709 Corr.		0.055	-0.031	0.050	-0.031	0.004	-0.004	0.012
m709 R^2		0.003	0.001	0.002	0.001	0.000	0.000	0.000
LD X^2		4	2	3	2	0	0	0
P-value		0.039830	0.127371	0.069309	0.125046	0.895257	0.767100	0.702909
m709 LOD		0.917	0.505	0.716	0.511	0.004	0.019	0.032
m709 n		1303	1303	1303	1303	1303	1303	1303
m654 D			-0.130	0.212	-0.132	-0.082	-0.004	-0.070
m654 D'			0.996	0.978	0.996	0.991	1.000	0.990
m654 Corr.			-0.563	0.905	-0.570	-0.405	-0.077	-0.366
m654 R^2			0.317	0.818	0.325	0.164	0.006	0.134
LD X^2			588	1698	619	300	12	255
P-value			< 2e-16	< 2e-16	< 2e-16	< 2e-16	0.000391	< 2e-16
m654 LOD			127.716	368.786	134.516	65.332	2.730	55.408
m654 n			1303	1303	1303	1303	1303	1303
m47 D				-0.143	0.231	-0.083	-0.004	-0.075
m47 D'				0.992	0.991	0.966	0.998	1.000
m47 Corr.				-0.608	0.985	-0.407	-0.081	-0.388
m47 R^2				0.370	0.970	0.166	0.007	0.150
LD X^2				699	2213	268	7	273
P-value				< 2e-16	< 2e-16	< 2e-16	0.007251	< 2e-16
m47 LOD				151.847	480.561	58.250	1.566	59.344
m47 n				1303	1303	1303	1303	1303
p46 D					-0.145	-0.090	-0.004	-0.078
p46 D'					0.993	0.985	1.000	0.991
p46 Corr.					-0.615	-0.439	-0.085	-0.400
p46 R^2					0.379	0.192	0.007	0.160
LD X^2					719	335	14	282
P-value					< 2e-16	< 2e-16	0.000113	< 2e-16
p46 LOD					156.342	72.789	3.237	61.413
p46 n					1303	1303	1303	1303
p79 D						-0.083	-0.004	-0.075

Figure 4: Linkage disequilibrium plot



	m709	m654	m47	p46	p79	p252	p491	p523
m709	m709	0.004	0.001	0.004	0.001	0.000	0.000	0.000
m654	0.00941	m654	0.333	0.827	0.346	0.159	0.007	0.133
m47	0.09066	< 2e−16	m47	0.390	0.972	0.162	0.008	0.140
p46	0.01931	< 2e−16	< 2e−16	p46	0.401	0.181	0.008	0.153
p79	0.08600	< 2e−16	< 2e−16	< 2e−16	p79	0.164	0.008	0.142
p252	0.57719	< 2e−16	< 2e−16	< 2e−16	< 2e−16	p252	0.042	0.811
p491	0.70263	8.33e−05	0.00351	1.95e−05	0.00268	2.93e−13	p491	0.052
p523	0.80818	< 2e−16	< 2e−16	< 2e−16	< 2e−16	< 2e−16	3.62e−16	p523

4 Subsetting by Group

The `alleleSummary` and `genotypeSummary` functions also allow you to create tables which show the summary information separated out by a grouping variable, which must be discrete “factor” variables (such as Sex).

To accomplish this, add the argument `by=Sex` to the function call. For example:

```
> alleleSummary(CAMP, by = "sex")

 Gene Marker Position Group Allele Count  Freq  CI-Lower CI-Upper
 ALL  m709   ?        1     1       2536 0.998 0.996    0.999
                      1     2          6 0.002 0.001    0.004
                      1     NA         0

                      2     1       2536 0.998 0.996    0.999
                      2     2          6 0.002 0.001    0.004
                      2     NA         0

      m654   ?        1     1       1618 0.640 0.621    0.658
                      1     2        912 0.360 0.342    0.379
                      1     NA         0

                      2     1       1618 0.640 0.621    0.658
                      2     2        912 0.360 0.342    0.379
                      2     NA         0

      m47    ?        1     1        935 0.374 0.355    0.393
                      1     2       1565 0.626 0.607    0.645
                      1     NA         0

                      2     1        935 0.374 0.355    0.393
                      2     2       1565 0.626 0.607    0.645
                      2     NA         0

      p46    ?        1     1       1000 0.400 0.380    0.419
                      1     2       1502 0.600 0.581    0.620
                      1     NA         0

                      2     1       1000 0.400 0.380    0.419
                      2     2       1502 0.600 0.581    0.620
                      2     NA         0

      p79    ?        1     1       1544 0.621 0.601    0.639
                      1     2        944 0.379 0.361    0.399
                      1     NA         0

                      2     1       1544 0.621 0.601    0.639
                      2     2        944 0.379 0.361    0.399
                      2     NA         0

      p252   ?        1     1        517 0.221 0.205    0.238
                      1     2       1819 0.779 0.762    0.795
                      1     NA         0
```

```
                        2     1          517 0.221 0.205    0.238
                        2     2         1819 0.779 0.762    0.795
                        2     NA           0

        p491   ?        1     1         2505 0.988 0.983    0.992
                        1     2           31 0.012 0.008    0.017
                        1     NA           0

                        2     1         2505 0.988 0.983    0.992
                        2     2           31 0.012 0.008    0.017
                        2     NA           0

        p523   ?        1     1         2057 0.810 0.795    0.825
                        1     2          481 0.190 0.175    0.205
                        1     NA           0

                        2     1         2057 0.810 0.795    0.825
                        2     2          481 0.190 0.175    0.205
                        2     NA           0

Footer:
          Confidence intervals width is 95%, computed using
          the exact quantiles for the binomial
          distribution.
```

This will display a table within a separate block within each marker for each level of the variable `Sex`.

To control whether the summary table for entire data in addition to individual factor levels, add `includeOverall=TRUE` or `includeOverall=FALSE` (the default) as appropriate.

A Example R script

```
> library(GeneticsBase)
> data(CAMP)
> PGtables(CAMP, filename = "test", format = "html")
> PGtables(CAMP, filename = "test", format = "latex")
```

References

Warnes GR. “The Genetics Package: Utilities for handling genetic data” `R News`, Volume 3, Issue 1, June 2003.

Warnes GR. “genetics”, a package for handling marker-based genetic data within the open-source statistical package “R”. The package includes function to compute allele frequencies, use genetic markers in statistical models, estimate disequilibrium, and test for departure from Hardy-Weinberg equilibrium. `http://cran.us.r-project.org/src/contrib/PACKAGES.html#genetics`, 2002-