polySegratio: An R library for autopolyploid segregation analysis

Peter Baker

January 9, 2008

It is well known that the dosage level of markers in autopolyploids and allopolyploids can be characterised by their observed segregation ratios. On the other hand, contrary to methods employed in several studies, segregation ratios are not a good indicator of polyploid type (Qu and Hancock, 2002).

The `polySegratio` package provides standard approaches to assess marker dosage in autopolyploids although the functions could equally well be applied to allopolyploids with specified expected segregation ratios. In addition, simulated sets of markers may be generated with specified dosages, ploidy and levels of oversidpersion.

To use the library, you need to attach it with

```
> library(polySegratio)
```

1 Expected segregation ratios

Haldane (1930) outlined the derivation of the expected numbers and ratios of offspring for various parental configurations of autopolyploids. Expected gametic series for polyploids of various sizes were produced, along with expected ratios of gametic series for crosses and selfing and the equilibrium distribution under random mating. Haldane provides expected gametic series when one parent is nulliplex for polyploids up to order 16 (heccaidecaploid).

For an autooctaploid with bivalent pairing and in the absence of double reduction [1] with A being the dominant allele and a the recessive, then the expected gametic series formed are outlined in Table 1. Employing the notation that A^s represents s copies of allele A, then if a heterozygous parent $A^r a^{8-r}$ is crossed with a recessive nulliplex (a^8) octaploid then the results of crossing can be calculated by symbolic manipulation. For instance, if a parent with a

[1] Double reduction: if separation for any locus is equational the two chromatids from one chromosome may be present together in one interphase nucleus but joined to separate centromeres allowing them to enter the same gamete. Sister chromatids in the same gamete, reducing the genetic content of a gamete twice, instead of once. Normally, two of the four chromosomes end up together in a gamete, reducing the genetic content in half. With double reduction gametes, the two chromosomes in the gamete are the same, at least at some loci; i.e., they are sister chromatids, and genetic content is reduced to 1/4 when compared to the parental plant. See Mather (1936)

Table 1: The gametic segregation in an autooctaploid of a heterozygous cross ($A^s a^{8-s}, s = 1 \ldots 7$) with a nulliplex ($a^8$) assuming bivalent pairing and no double reduction. The ratio is of dominants to recessives and ω_k is the proportion of dominants.

Heterozygous	Gametes					Segregation Ratio	
Parent	A^4	A^3a	A^2a^2	Aa^3	a^4	$A^s a^{8-s}$:a^8	ω_k
Aa^7				1	1	1:1	0.500
A^2a^6			3	8	3	11:3	0.786
A^3a^5		1	6	6	1	13:1	0.929
A^4a^4	1	16	36	16	1	69:1	0.986
A^5a^3	1	6	6	1			
A^6a^2	3	8	3				
A^7a	1	1					

single dose marker Aa^7 is crossed with a nulliplex parent a^8 then $Aa^7 \times a^8$ yields $(1.Aa^3 + 1.a^4) \times (a^4)$ or zygotes $(1.Aa^7 + 1.a^8)$ with ratios $1.Aa^7 : 1.a^8$.

Although published previously in slightly different forms, the general formula of Ripol et al. (1999) is employed for $p(k)$ or the expected segregation proportion given dosage k which is

$$p(k|m,x) = 1 - \frac{\binom{m-k}{mx}}{\binom{m}{mx}}, k = 0 \ldots m/2 \quad (1)$$

where m is the ploidy level or number of homologous chromosomes and the monoploid number x is the number of chromosomes in a basic set. Note that for diploids $m = 2$, tetraploids $m = 4$, octaploids then $m = 8$ and so on.

To obtain such theoretical segregation proportions or probabilities using `expected.segRatio` is straightforward by specifying the ploidy level either numerically or by name. The function `expected.segRatio` employs Equations 1 and 2 to compute expected segregation proportions. For instance

```
> print(unlist(expected.segRatio(2)))

      ratio.SD    ploidy.level     ploidy.name    type.parents
         "0.5"             "2"       "Diploid" "heterogeneous"

> print(unlist(expected.segRatio("Tetraploid")))

          ratio.SD            ratio.DD        ploidy.level
             "0.5" "0.833333333333333"                 "4"
       ploidy.name        type.parents
      "Tetraploid"     "heterogeneous"

> print(expected.segRatio("Octa")$ratio)

    SD     DD     TD     QD
0.5000 0.7857 0.9286 0.9857
```

In the case where, an AFLP band is present in both parents but not in all offspring, there must be less than four copies of the dominant allele in both parents. For instance, crossing the two genetically similar autooctoploid lines Aa^7 results in 1 nulliplex in 4 since $(1.Aa^3 + 1.a^4)^2$ is simply $(1.A^2a^6 + 2.Aa^7 + 1.a^8)$. For alternate autooctoploid parental configurations result in segregation proportions of around 0.9 or above and would apparently therefore be indistinguishable via segregation ratios alone. Similarly to Equation 1 we deduce that if both parents contain at least one copy of the dominant marker than a general equation for then for the dosage j in the first parent and dosage k in the second parent then the expected segregation proportion $p(j,k)$ is

$$p(j,k|m,x) = 1 - \frac{\binom{m-k}{mx}\binom{m-j}{mx}}{\binom{m}{mx}^2}, j,k = 0 \ldots m/2 \quad (2)$$

where m and x are defined in Equation 1, noting that neither parent is nulliplex. Such segregation ratios may be computed using `expected.segRatio` as follows:

```
> print(unlist(expected.segRatio("tetra", type = "homoz")))

        ratio.SDxSD          ratio.SDxDD          ratio.DDxDD
              "0.75" "0.916666666666667" "0.972222222222222"
      ploidy.level          ploidy.name         type.parents
               "4"        "Tetraploid"         "homozygous"

> print(expected.segRatio("Octa", type = "homoz")$ratio)

 SDxSD  SDxDD  DDxDD  DDxTD  TDxTD  TDxQD  QDxQD
0.7500 0.8929 0.9643 0.9929 0.9969 0.9990 0.9998
```

Note that Equations 1 and 2 are defined for m even but that a warning is issued and results still calculated if m is odd. As an example

```
> a <- expected.segRatio(9)

Warning: ploidy level not even - results may be unexpected

> print(a$ratio)

    SD     DD     TD     QD
0.5556 0.8333 0.9524 0.9921
```

2 Simulating a set of markers

Functions `sim.autoMarkers` and `sim.autoCross` may be used to simulate marker data for a collection of markers where either one of the parents is nulliplex or where both parents contain at least one dose of a marker. The data are only simulated to produce appropriate segregation ratios but other genetic parameters such as recombination, degree of preferential pairing or a genetic map are not considered. The proportions in each marker dosage need to be specified.

`sim.autoMarkers` may be used to simulate dominant markers from an autopolyploid cross given the ploidy level, specified parental marker alleles, the expected segregation ratios and the proportions in each dosage marker class. The ploidy level may be chosen from tetraploid to heccaidecaploid and the segregation ratios may be specified explicitly or generated automatically.

`sim.autoCross` is a wrapper to `sim.autoMarkers` which is used to generate markers for parents with markers that are 10, 01 or 11. The proportions of markers for each of these three parental types must be specified.

Both functions return S3 class objects (class `simAutoCross` and class `simAutoMarkers`) which have associated print and plot methods.

For instance, to generate and plot the segregation proportions for 200 markers for 100 progeny from a tetraploid cross where one of the parents is nulliplex and there are 70% single dose markers and 30% dose markers then use

```
> mark.sim4 <- sim.autoMarkers(4, dose.proportion = c(0.7,
+     0.3), n.markers = 200, n.individuals = 200)
> print(mark.sim4)

Autopolyploid dominant markers generated at Thu Jan 21 21:05:06 2010
with call:
sim.autoMarkers(ploidy.level = 4, dose.proportion = c(0.7, 0.3),
    n.markers = 200, n.individuals = 200)

Ploidy level is: 4 ( Tetraploid )
Parents were set as heterogeneous for the markers
Theoretical segregation proportions:
           ratio.SD            ratio.DD        ploidy.level
              "0.5" "0.833333333333333"                 "4"
        ploidy.name        type.parents
       "Tetraploid"     "heterogeneous"

Proportions in each dosage class:
 SD  DD
0.7 0.3
No. of markers generated from multinomial distribution:
   No.markers
SD        133
DD         67

Data were generated for 200 individuals with 200 markers
A subset is:
     X.1 X.2 X.3 X.4 X.5 X.6 X.7 X.8 X.9 X.10 r   n   ratio dose
M.1  0   1   0   1   1   1   1   1   1   0    101 200 0.505 SD
M.2  1   0   0   1   0   1   1   0   0   1    96  200 0.48  SD
M.3  1   1   0   1   0   1   1   0   1   0    91  200 0.455 SD
M.4  1   0   1   1   1   0   0   1   1   1    106 200 0.53  SD
M.5  0   1   0   1   0   1   0   0   0   0    97  200 0.485 SD
M.6  0   1   0   0   0   1   1   0   1   0    100 200 0.5   SD
```

```
M.7  1   0   0   1   1   0   1   1   0   0    102 200 0.51  SD
M.8  1   1   0   0   1   0   1   0   0   0    108 200 0.54  SD
M.9  0   0   0   0   1   1   1   1   0   0    98  200 0.49  SD
M.10 1   0   1   1   0   0   1   1   0   0    100 200 0.5   SD
```

```
> plot(mark.sim4)
```

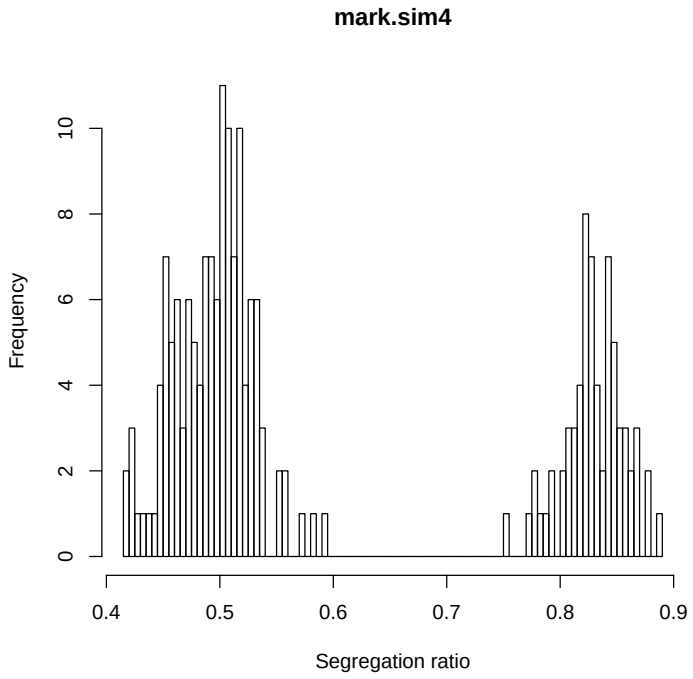

Figure 1: Segregation ratios from simulated marker data for 200 markers for a autotetraploid cross with 100 offspring

Figure 1 shows a histogram of segregation proportions for a tetraploid cross. Other plots, may be produced. For instance, the number of missing values is useful when looking at real data to determine if some markers are not well measured (See Figure 2).

Often in molecular marker studies, a small percentage of markers may be missing or misclassified. The functions `addMissing` and `addMisclass` allow marker data to be modified accordingly. The rate may be specified either as a proportion of missing at random or a proportion of columns and rows with specified proportions of missings or misclassified. Not that if markers are randomly misclassified then the expected segregations ratios are still the same and so we may not expect to see much difference to perfectly classified markers.

`addMissing` adds missing data at random to objects of class `autoMarker` or `autoCross`. `addMisclass` misclassifies marker data in objects of class `autoMarker` or `autoCross` at a specified rate. Parental marker data may also be

misclassified. An example might be

```
> miss.sim4 <- addMisclass(mark.sim4, misclass = 0.1)
> miss.sim4 <- addMissing(miss.sim4, na.proportion = 0.2)
> print(miss.sim4, col = c(1:6))

Autopolyploid dominant markers generated at Thu Jan 21 21:05:06 2010
with call:
sim.autoMarkers(ploidy.level = 4, dose.proportion = c(0.7, 0.3),
    n.markers = 200, n.individuals = 200)

Ploidy level is: 4 ( Tetraploid )
Parents were set as heterogeneous for the markers
Theoretical segregation proportions:
           ratio.SD            ratio.DD        ploidy.level
              "0.5" "0.833333333333333"                 "4"
        ploidy.name        type.parents
       "Tetraploid"     "heterogeneous"

Proportions in each dosage class:
 SD  DD
0.7 0.3
No. of markers generated from multinomial distribution:
   No.markers
SD        133
DD         67

Data were generated for 200 individuals with 200 markers
A subset is:
     X.1  X.2  X.3  X.4  X.5  X.6  X.7  X.8 X.9  X.10 r  n
M.1  0    0    <NA> <NA> <NA> 1    1    1   1    0    66 160
M.2  0    <NA> 1    1    0    1    0    0   0    1    79 156
M.3  1    1    0    1    1    1    1    0   1    0    87 165
M.4  <NA> 0    1    1    1    0    1    1   1    1    88 163
M.5  <NA> <NA> 0    <NA> 0    1    0    0   0    <NA> 77 162
M.6  <NA> 1    0    0    0    1    1    1   1    0    74 155
M.7  <NA> 0    0    1    1    <NA> 1    0   <NA> 0    78 158
M.8  <NA> 1    <NA> <NA> <NA> <NA> 1    0   0    1    91 160
M.9  <NA> <NA> 0    <NA> <NA> 1    <NA> 1   0    0    75 152
M.10 <NA> <NA> 1    1    0    0    0    1   0    0    80 164
     ratio             dose
M.1  0.4125            SD
M.2  0.506410256410256 SD
M.3  0.527272727272727 SD
M.4  0.539877300613497 SD
M.5  0.475308641975309 SD
M.6  0.47741935483871  SD
M.7  0.493670886075949 SD
```

```
M.8  0.56875           SD
M.9  0.493421052631579 SD
M.10 0.48780487804878  SD
Missing data generated for 20 % markers at random
```

```
> plot(miss.sim4, type = "all")
```

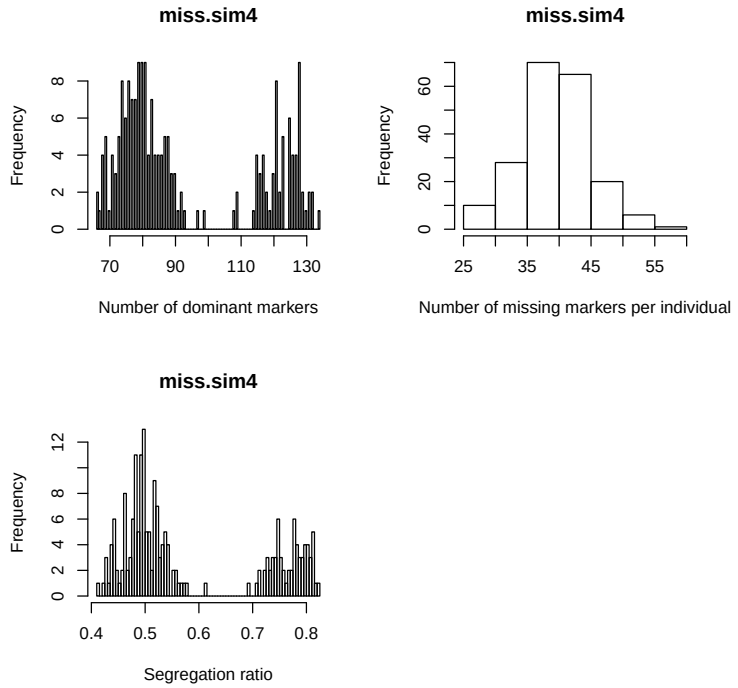

Figure 2: Histograms of the number of markers labelled 1, numbers of missing values per marker and segregation ratios

2.1 Overdispersion

Since markers are correlated and may be subject to different types of measurement errors, then the segregation ratios may follow an overdispersed Binomial distribution. Such markers may be simulated with `sim.autoMarkers` by setting the parameter `overdispersion` to `TRUE`. The amount of overdispersion or extra–binomial variation may be specified by setting the `shape1` parameter. Larger values imply less overdispersion. Typically, the `R` command would be like `sim.autoMarkers(4,c(0.8,0.2), overdisp=TRUE, shape1=20)`

Overdispersed marker data are simulated from the Beta–Binomial distribution where the Binomial proportion p is generated from a Beta distribution. Note that if p is generated from a $\beta(a,b)$ distribution, then $E(p) = a/(a+b)$

and $\mathrm{Var}(p) = ab/((a+b)^2(a+b+1))$. Thus constraining $E(p)$ to be the appropriate segregation proportion and setting the first shape parameter a implies that $b = a(1-p)/p$. Tetraploid marker data generated for a range of `shape1` or a values is shown in Figure 3.

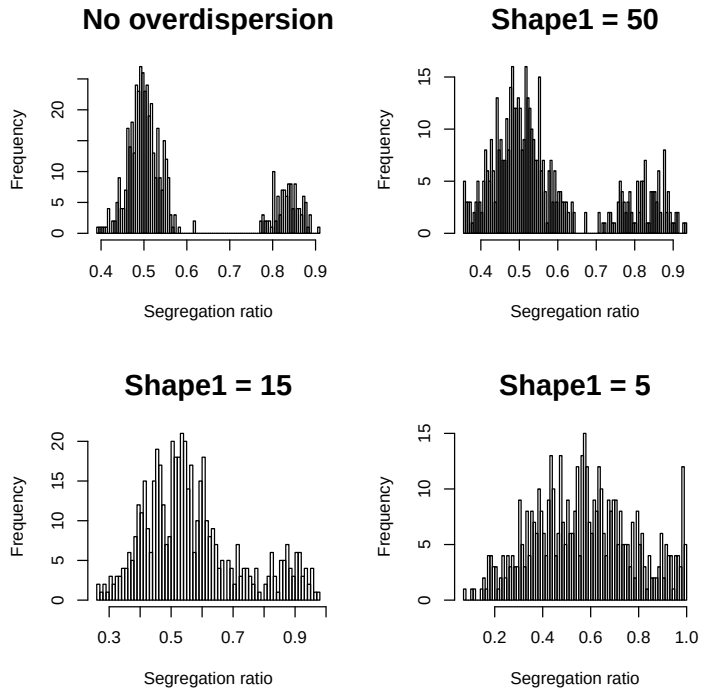

Figure 3: Histograms of the number of dominant markers simulated for 500 overdispersed markers from 200 autotetraploids. Data were generated from the Beta–Binomial distribution with a range of shape parameters. Overdispersion increases as `shape1` decreases.

3 Standard approaches for assessing marker dosage

The most widely used test for assessing marker dosage is the standard χ^2 test. Following Mather (1951), this test is often employed to compare the observed segregation ratio against its expected value. More recently, Ripol et al. (1999) proposed that the observed segregation proportion be compared to the appropriate Binomial confidence interval given the sample size and the expected segregation proportion.

Both tests may be carried out by means of the function `test.segRatio`. Note that if the tests reveal that a marker may be more than one dosage then it is not allocated a marker dosage.

3.1 χ^2 tests

The default method of assessing marker dosage in `test.segRatio` is the χ^2 test. The function requires that the segregation proportions are given in the form of object of S3 class `segRatio`. These are automatically produced for simulated data created with functions `sim.autoMarkers` and `sim.autoCross` and may be calculated from observed marker data either manually or by applying `segregationRatios` to a matrix of observed marker data.

For instance, to calculate χ^2 tests and allocate dosage for an autooctoploid then

```
> a <- sim.autoMarkers(ploidy = 8, c(0.7, 0.2, 0.09, 0.01),
+     n.markers = 200, n.individuals = 100)
> print(a)

Autopolyploid dominant markers generated at Thu Jan 21 21:05:06 2010
with call:
sim.autoMarkers(ploidy.level = 8, dose.proportion = c(0.7, 0.2,
    0.09, 0.01), n.markers = 200, n.individuals = 100)

Ploidy level is: 8 ( Octaploid )
Parents were set as heterogeneous for the markers
Theoretical segregation proportions:
           ratio.SD            ratio.DD            ratio.TD
              "0.5" "0.785714285714286" "0.928571428571429"
           ratio.QD        ploidy.level         ploidy.name
"0.985714285714286"                 "8"         "Octaploid"
       type.parents
    "heterogeneous"

Proportions in each dosage class:
  SD   DD   TD   QD
0.70 0.20 0.09 0.01
No. of markers generated from multinomial distribution:
   No.markers
SD        138
DD         47
TD         14
QD          1

Data were generated for 100 individuals with 200 markers
A subset is:
     X.1 X.2 X.3 X.4 X.5 X.6 X.7 X.8 X.9 X.10 r  n   ratio dose
M.1  0   1   1   1   0   1   0   1   1   1    42 100 0.42  SD
M.2  0   0   1   1   1   0   1   0   1   1    50 100 0.5   SD
M.3  0   1   0   0   0   0   0   0   1   1    44 100 0.44  SD
M.4  1   0   0   0   0   1   1   1   1   1    56 100 0.56  SD
M.5  1   1   0   1   0   0   1   0   1   0    51 100 0.51  SD
M.6  1   0   0   1   1   0   0   0   1   1    48 100 0.48  SD
```

```
M.7  1   1   0   1   0   1   1   0   1   1    51 100 0.51  SD
M.8  0   0   1   1   1   0   0   1   1   1    44 100 0.44  SD
M.9  1   0   1   1   1   1   1   1   1   1    55 100 0.55  SD
M.10 1   0   1   0   0   0   1   1   1   1    45 100 0.45  SD
```

Note that `a` is an object of S3 class `simAutoMarkers` and that the segregation ratios may be obtained as the list component `seg.ratios`. Since `a` is simulated we can also extract the true dosage obtain the number of correctly classified markers.

```
> ac <- test.segRatio(a$seg.ratios, ploidy = 8, method = "chi.squared")
> print(ac)

Segregation ratio test:
Method: chi.squared at the 0.05 level
Expected segregation ratios:
           ratio.SD            ratio.DD            ratio.TD
              "0.5" "0.785714285714286" "0.928571428571429"
           ratio.QD        ploidy.level         ploidy.name
"0.985714285714286"                 "8"         "Octaploid"
       type.parents
    "heterogeneous"
Proportion of markers classified at 0.05 level: 0.94
Classified: 188 , not classified: 12
Markers doubly classified: 2
      SD DD TD QD
M.189  0  0  1  1
M.198  0  0  1  1
Number in each marker dosage class (classified once):
 SD  DD  TD  QD
130  44  13   1
Dosage of first 10 markers (where dosage unique):
 M.1  M.2  M.3  M.4  M.5  M.6  M.7  M.8  M.9 M.10
   1    1    1    1    1    1    1    1    1    1
Call: test.segRatio(seg.ratio = a$seg.ratios, ploidy.level = 8, method = "chi.squared")

> print(addmargins(table(a$true.doses$dosage, ac$dosage,
+     exclude = NULL)))

         1   2   3   4 <NA> Sum
  1    130   0   0   0    8 138
  2      0  44   1   0    2  47
  3      0   0  12   0    2  14
  4      0   0   0   1    0   1
  <NA>   0   0   0   0    0   0
  Sum  130  44  13   1   12 200
```

Note that for segregation ratios near to one the χ^2 approximation may not hold and so `R` will produce a warning.

3.2 Binomial confidence intervals

The Binomial confidence interval approach of Ripol et al. (1999) is obtained by setting the `method` parameter to "`binomial`". The α level may be set in either method by setting the parameter `alpha`. For instance,

```
> ab <- test.segRatio(a$seg.ratios, ploidy = 8, method = "bin",
+     alpha = 0.01)
> print(ab)

Segregation ratio test:
Method: binomial at the 0.01 level
Expected segregation ratios:
           ratio.SD            ratio.DD            ratio.TD
              "0.5" "0.785714285714286" "0.928571428571429"
           ratio.QD        ploidy.level         ploidy.name
"0.985714285714286"                 "8"         "Octaploid"
       type.parents
    "heterogeneous"
Proportion of markers classified at 0.01 level: 0.97
Classified: 194 , not classified: 6
Markers doubly classified: 5
      SD DD TD QD
M.159  0  1  1  0
M.177  0  1  1  0
M.189  0  0  1  1
M.193  0  0  1  1
M.198  0  0  1  1
Number in each marker dosage class (classified once):
 SD  DD  TD  QD
137  45  11   1
Dosage of first 10 markers (where dosage unique):
 M.1  M.2  M.3  M.4  M.5  M.6  M.7  M.8  M.9 M.10
   1    1    1    1    1    1    1    1    1    1
Call: test.segRatio(seg.ratio = a$seg.ratios, ploidy.level = 8, method = "bin",
    alpha = 0.01)

> print(addmargins(table(a$true.doses$dosage, ab$dosage,
+     exclude = NULL)))

        1   2   3   4 <NA> Sum
  1   137   0   0   0    1 138
  2     0  45   0   0    2  47
  3     0   0  11   0    3  14
  4     0   0   0   1    0   1
  <NA>  0   0   0   0    0   0
  Sum 137  45  11   1    6 200
```

4 Utility functions

Several utility functions are included for use with real or simulated data.

When marker data are stored in spreadsheets repetitive parts of marker names may be left blank or columns containing parts of names may need to be combined. To aid the process of constructing unique marker labels, `autoFill` automatically fills out blanks of a vector with the preceding label and `makeLabel` generates labels from two columns where blanks in first column are replaced by preceding non-blank label.

```
> x <- data.frame(col1 = c("agc", "", "", "", "gct5", "",
+     "ccc", "", ""), col2 = c(1, 3, 4, 5, 1, 2, 2, 4,
+     6))
> print(x)

  col1 col2
1  agc    1
2         3
3         4
4         5
5 gct5    1
6         2
7  ccc    2
8         4
9         6

> print(makeLabel(x))

[1] "agc1"  "agc3"  "agc4"  "agc5"  "gct51" "gct52" "ccc2"  "ccc4"
[9] "ccc6"

> print(cbind(x, lab = makeLabel(x, sep = ".")))

  col1 col2    lab
1  agc    1  agc.1
2         3  agc.3
3         4  agc.4
4         5  agc.5
5 gct5    1 gct5.1
6         2 gct5.2
7  ccc    2  ccc.2
8         4  ccc.4
9         6  ccc.6
```

`divide.autoMarkers` will split up a set of markers depending on the parental alleles. This is useful when extracting markers to be used in constructing a marker map for one parent say or in obtaining those markers present in both parents but segregating in the offspring.

```
> p2 <- sim.autoCross(4, dose.proportion = list(p01 = c(0.7,
+     0.3), p10 = c(0.7, 0.3), p11 = c(0.6, 0.2, 0.2)))
> print(p2, row = c(1:5))

Autopolyploid dominant markers for crosses generated at Thu Jan 21 21:05:07 2010
with call:
sim.autoCross(ploidy.level = 4, dose.proportion = list(p01 = c(0.7,
    0.3), p10 = c(0.7, 0.3), p11 = c(0.6, 0.2, 0.2)))

Ploidy level is: 4 ( Tetraploid )
The proportion of markers of each parental type were
p10 p01 p11
0.4 0.4 0.2
Theoretical segregation proportions:
p10:
          ratio.SD            ratio.DD        ploidy.level
             "0.5" "0.833333333333333"                 "4"
       ploidy.name        type.parents
      "Tetraploid"     "heterogeneous"
p01:
          ratio.SD            ratio.DD        ploidy.level
             "0.5" "0.833333333333333"                 "4"
       ploidy.name        type.parents
      "Tetraploid"     "heterogeneous"
p11:
       ratio.SDxSD         ratio.SDxDD         ratio.DDxDD
            "0.75" "0.916666666666667" "0.972222222222222"
      ploidy.level         ploidy.name        type.parents
               "4"        "Tetraploid"        "homozygous"

Proportions in each dosage class:
p10:
 SD  DD
0.7 0.3
p01:
 SD  DD
0.7 0.3
p11:
SDxSD SDxDD DDxDD
  0.6   0.2   0.2
No. of markers generated from multinomial distribution:
p10:
   No.markers
SD        139
DD         61
p01:
   No.markers
SD        127
```

```
DD         66
p11:
      No.markers
SDxSD         66
SDxDD         15
DDxDD         26

Overall: data were generated for 200 individuals with 500 markers
A subset is:
     P.1 P.2 X.1 X.2 X.3 X.4 X.5 X.6 X.7 X.8 r   n   ratio dose
M.1  1   0   0   0   1   1   0   0   0   0   94  200 0.47  SD
M.2  1   0   0   0   0   0   0   0   1   0   92  200 0.46  SD
M.3  1   0   1   1   1   0   0   0   0   0   94  200 0.47  SD
M.4  1   0   1   0   1   0   0   0   0   1   94  200 0.47  SD
M.5  1   0   1   0   0   1   1   0   0   1   93  200 0.465 SD
M.6  1   0   1   1   1   0   0   1   1   0   94  200 0.47  SD
M.7  1   0   1   1   0   1   1   0   1   1   90  200 0.45  SD
M.8  1   0   1   0   0   1   0   1   0   1   101 200 0.505 SD
M.9  1   0   1   0   1   1   0   0   1   0   105 200 0.525 SD
M.10 1   0   1   0   0   1   0   1   1   1   105 200 0.525 SD

> ss <- divide.autoMarkers(p2$markers)
> print(ss, row = c(1:5))

Markers split for p2$markers

**** data set: Parent with 1 is P.1 and 0 is P.2
Dimension of marker data: 200 200
Data:
     X.1 X.2 X.3 X.4 X.5 X.6 X.7 X.8 X.9 X.10   r   n ratio
M.1    0   0   1   1   0   0   0   0   0    1  94 200 0.470
M.2    0   0   0   0   0   0   1   0   0    1  92 200 0.460
M.3    1   1   1   0   0   0   0   0   0    0  94 200 0.470
M.4    1   0   1   0   0   0   0   1   1    0  94 200 0.470
M.5    1   0   0   1   1   0   0   1   0    1  93 200 0.465
M.6    1   1   1   0   0   1   1   0   0    1  94 200 0.470
M.7    1   1   0   1   1   0   1   1   1    0  90 200 0.450
M.8    1   0   0   1   0   1   0   1   0    1 101 200 0.505
M.9    1   0   1   1   0   0   1   0   0    1 105 200 0.525
M.10   1   0   0   1   0   1   1   1   1    1 105 200 0.525
No. markers
    0     1
15638 24362
**** data set: Parent with 0 is P.1 and 1 is P.2
Dimension of marker data: 193 200
Data:
      X.1 X.2 X.3 X.4 X.5 X.6 X.7 X.8 X.9 X.10   r   n ratio
M.201   0   1   0   0   1   0   1   0   0    1 116 200 0.580
```

```
M.202   0   0   0   0   0   0   0   1   0    1  91 200 0.455
M.203   1   0   1   0   0   1   1   0   1    0  91 200 0.455
M.204   1   0   1   0   1   1   1   1   1    0 102 200 0.510
M.205   0   0   1   0   0   0   0   0   0    1 101 200 0.505
M.206   0   1   0   0   0   1   0   1   0    1  92 200 0.460
M.207   0   1   0   1   0   0   1   0   0    1  94 200 0.470
M.208   1   0   0   1   1   1   1   1   0    0  95 200 0.475
M.209   1   0   0   0   0   0   0   1   1    1 110 200 0.550
M.210   1   0   0   0   0   0   0   1   0    1  94 200 0.470
No. markers
    0     1
15183 23417
**** data set: Parents both  with 1 - P.1 & P.2
Dimension of marker data: 107 200
Data:
      X.1 X.2 X.3 X.4 X.5 X.6 X.7 X.8 X.9 X.10   r   n ratio
M.394   1   0   1   1   1   1   1   1   1    1 154 200 0.770
M.395   1   1   1   1   1   1   1   0   1    1 158 200 0.790
M.396   1   1   0   1   1   0   0   0   1    0 142 200 0.710
M.397   0   1   1   0   1   0   1   1   1    1 147 200 0.735
M.398   1   1   0   1   1   1   1   1   0    1 162 200 0.810
M.399   1   1   1   1   1   1   0   1   0    0 147 200 0.735
M.400   0   1   1   1   1   1   1   1   1    0 146 200 0.730
M.401   1   1   1   1   1   1   1   1   0    0 148 200 0.740
M.402   1   0   1   1   1   1   1   1   1    1 144 200 0.720
M.403   1   1   1   1   1   1   0   1   1    0 145 200 0.725
No. markers
    0     1
 3685 17715
Call:
divide.autoMarkers(markers = p2$markers)
```

References

Haldane, J. B. S. (1930). Theoretical genetics of autopolyploids. *Jornal of genetics*, 22:359–372.

Mather, K. (1936). Segregation and linkage in autotetraploids. *Journal of Genetics*, 32:287–314.

Mather, K. (1951). *The measurement of linkage in heredity.* Methuen, London.

Qu, L. and Hancock, J. (2002). Pitfalls of genetic analysis using a doubled–haploid backcrossed to its parent. *Theoretical and Applied Genetics*, 105(2-3):392–396.

Ripol, M. I., Churchill, G. A., da Silva, J. A., and Sorrells, M. (1999). Statistical aspects of genetic mapping in autopolyploids. *Gene*, 235(1-2):31–41.

4.1 Acknowledgments

Karen Aitken, given her experience in tetraploids and sugarcane marker maps, has provided many valuable insights into marker dosage in autopolyploids. David Lovell, Andrew George and Phil Jackson provided useful comments and discussions.