GENETICS

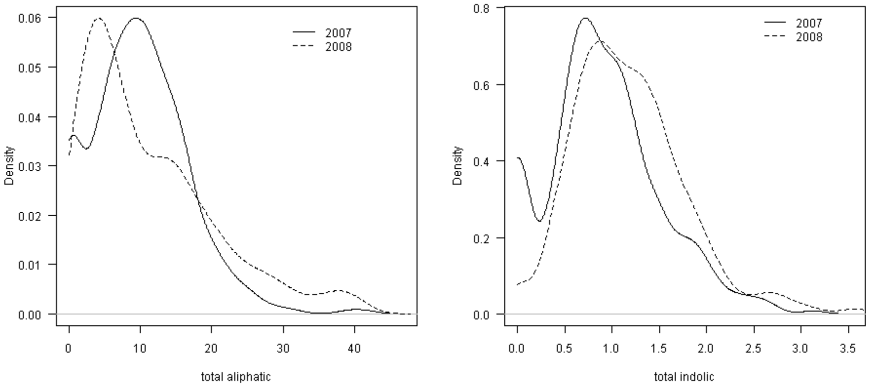
Supporting Information
http://www.genetics.org/cgi/content/full/genetics.109.108522/DC1

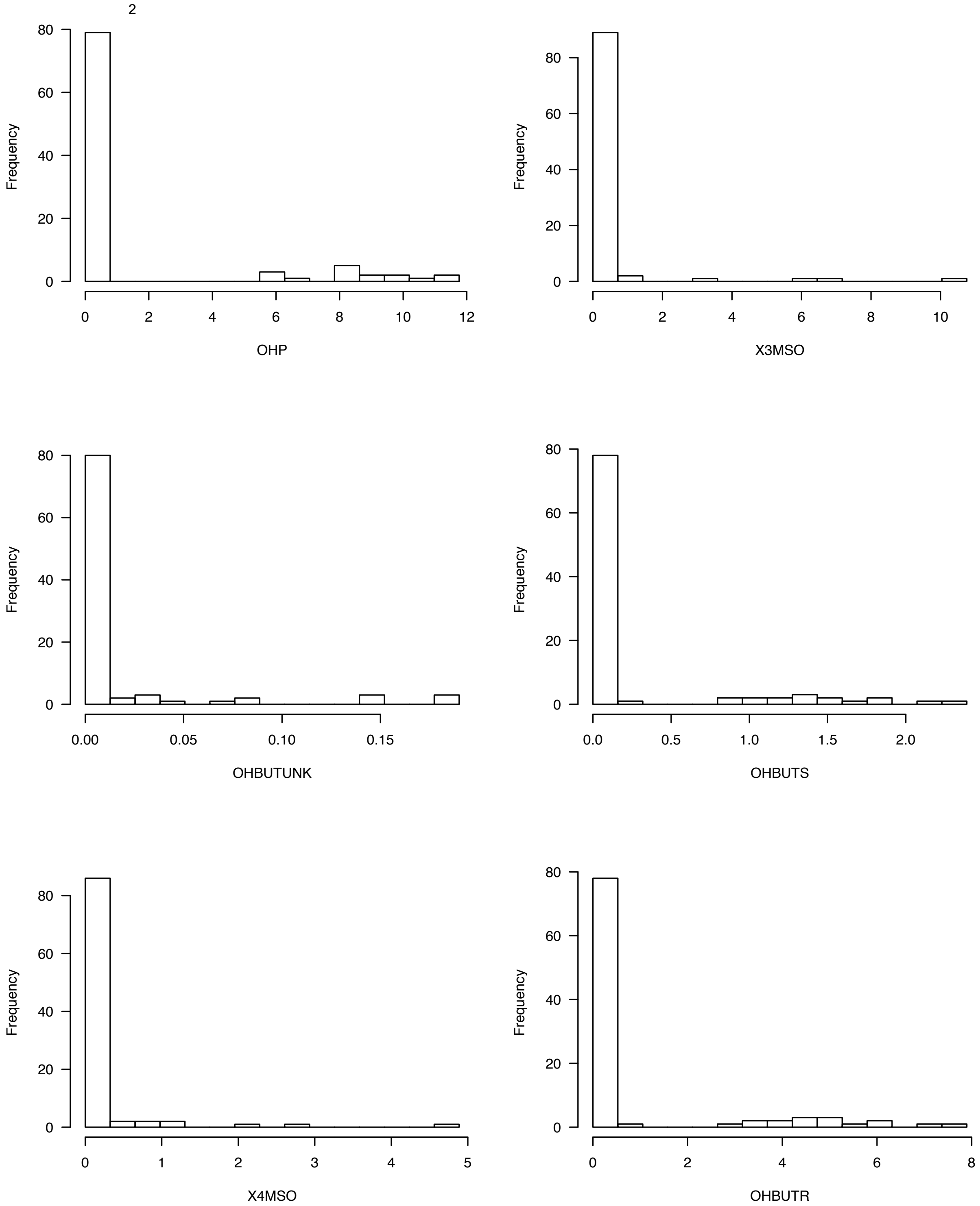
Understanding the Evolution of Defense Metabolites in *Arabidopsis thaliana* Using Genomewide Association Mapping

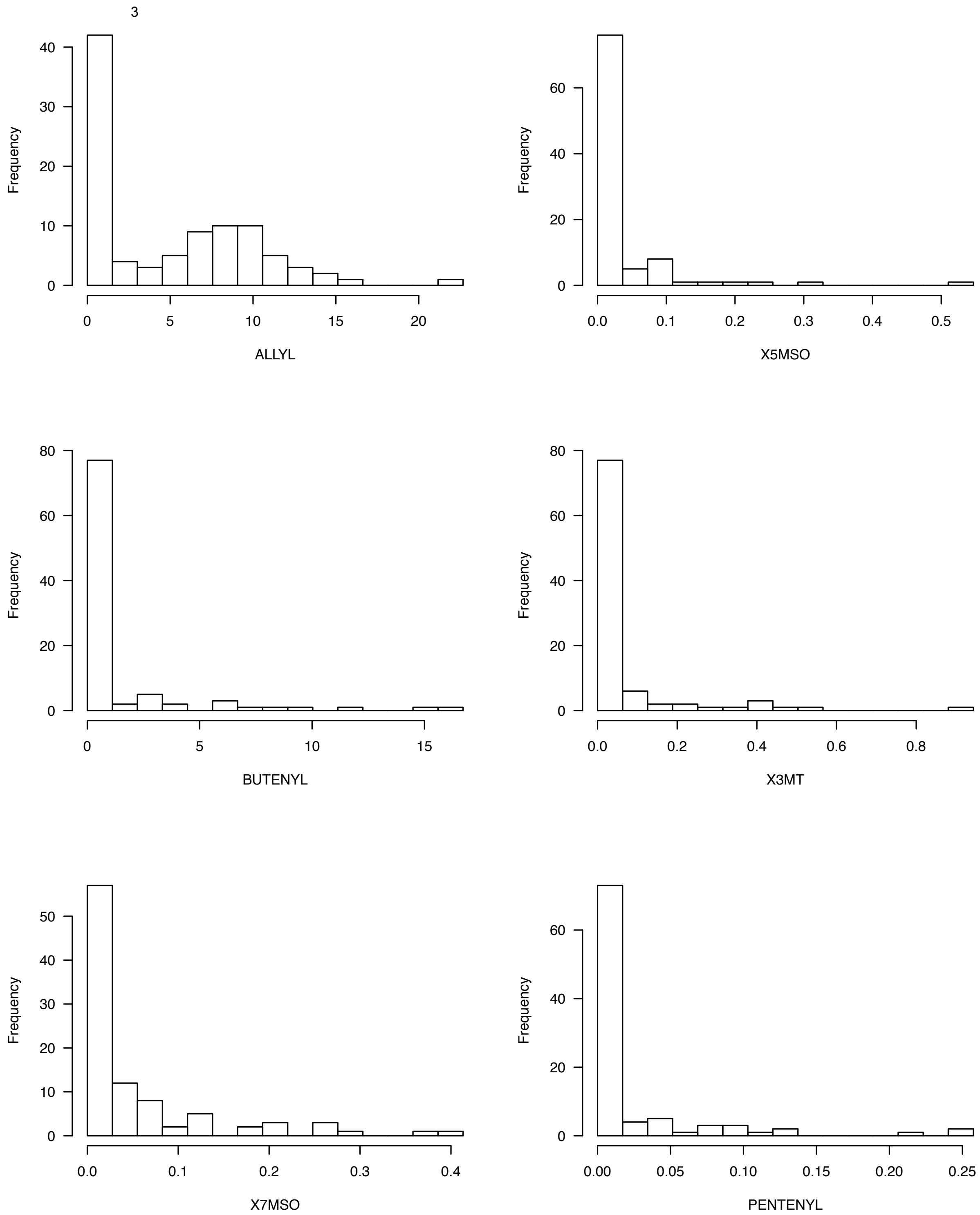
Eva K. F. Chan, Heather C. Rowe and Daniel J. Kliebenstein

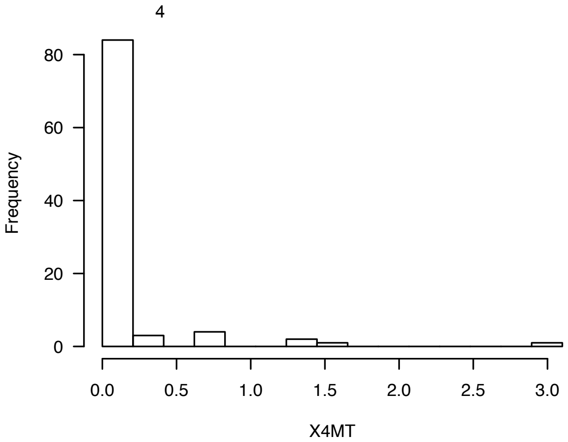
DOI: 10.1534/genetics.109.108522

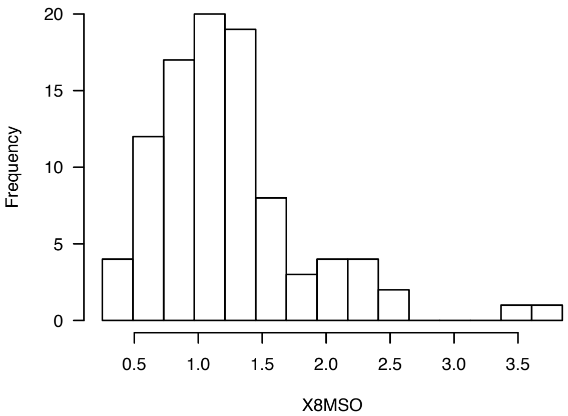
1

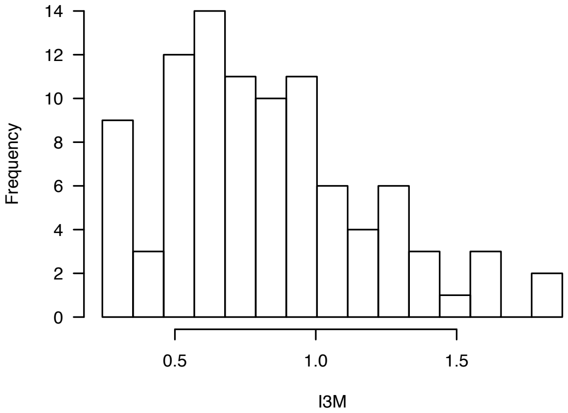

0.06
0.05
0.04
0.03
0.02
0.01
0.00
Density
2007
2008
0
10
20
30
40
total aliphatic
0.8
0.6
0.4
0.2
0.0
Density
2007
2008
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
total indolic

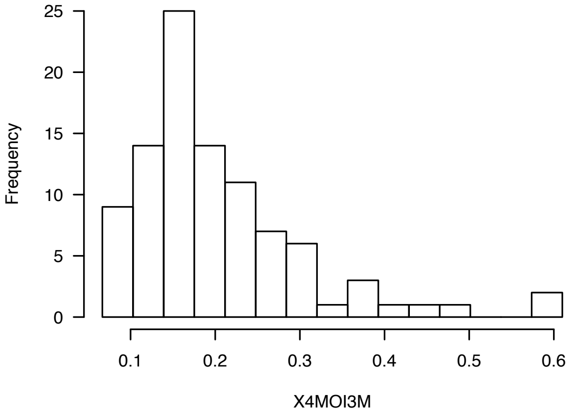

2
Frequency
80
60
40
20
0
0
2
4
6
8
10
12
OHP
Frequency
80
60
40
20
0
0
2
4
6
8
10
X3MSO
Frequency
80
60
40
20
0
0.00
0.05
0.10
0.15
OHBUTUNK
Frequency
80
60
40
20
0
0.0
0.5
1.0
1.5
2.0
OHBUTS
Frequency
80
60
40
20
0
0
1
2
3
4
5
X4MSO
Frequency
80
60
40
20
0
0
2
4
6
8
OHBUTR

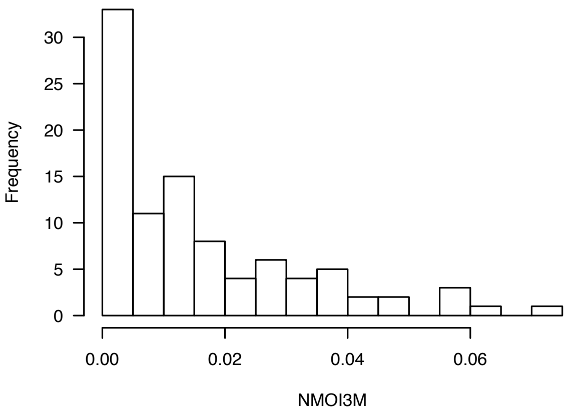

3
Frequency
0
10
20
30
40
0
5
10
15
20
ALLYL
Frequency
0
20
40
60
0.0
0.1
0.2
0.3
0.4
0.5
X5MSO
Frequency
0
20
40
60
80
0
5
10
15
BUTENYL
Frequency
0
20
40
60
80
0.0
0.2
0.4
0.6
0.8
X3MT
Frequency
0
10
20
30
40
50
0.0
0.1
0.2
0.3
0.4
X7MSO
Frequency
0
20
40
60
0.00
0.05
0.10
0.15
0.20
0.25
PENTENYL

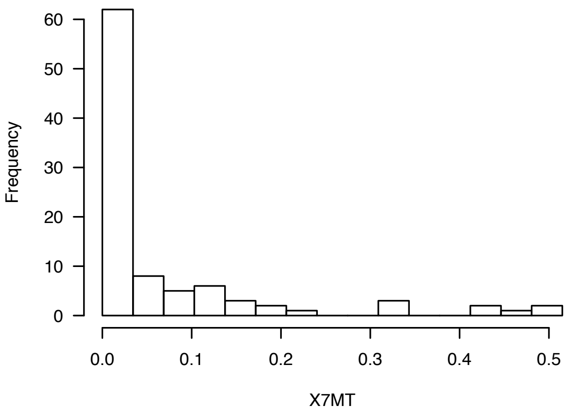

4
Frequency
0
20
40
60
80
0.0
0.5
1.0
1.5
2.0
2.5
3.0
X4MT

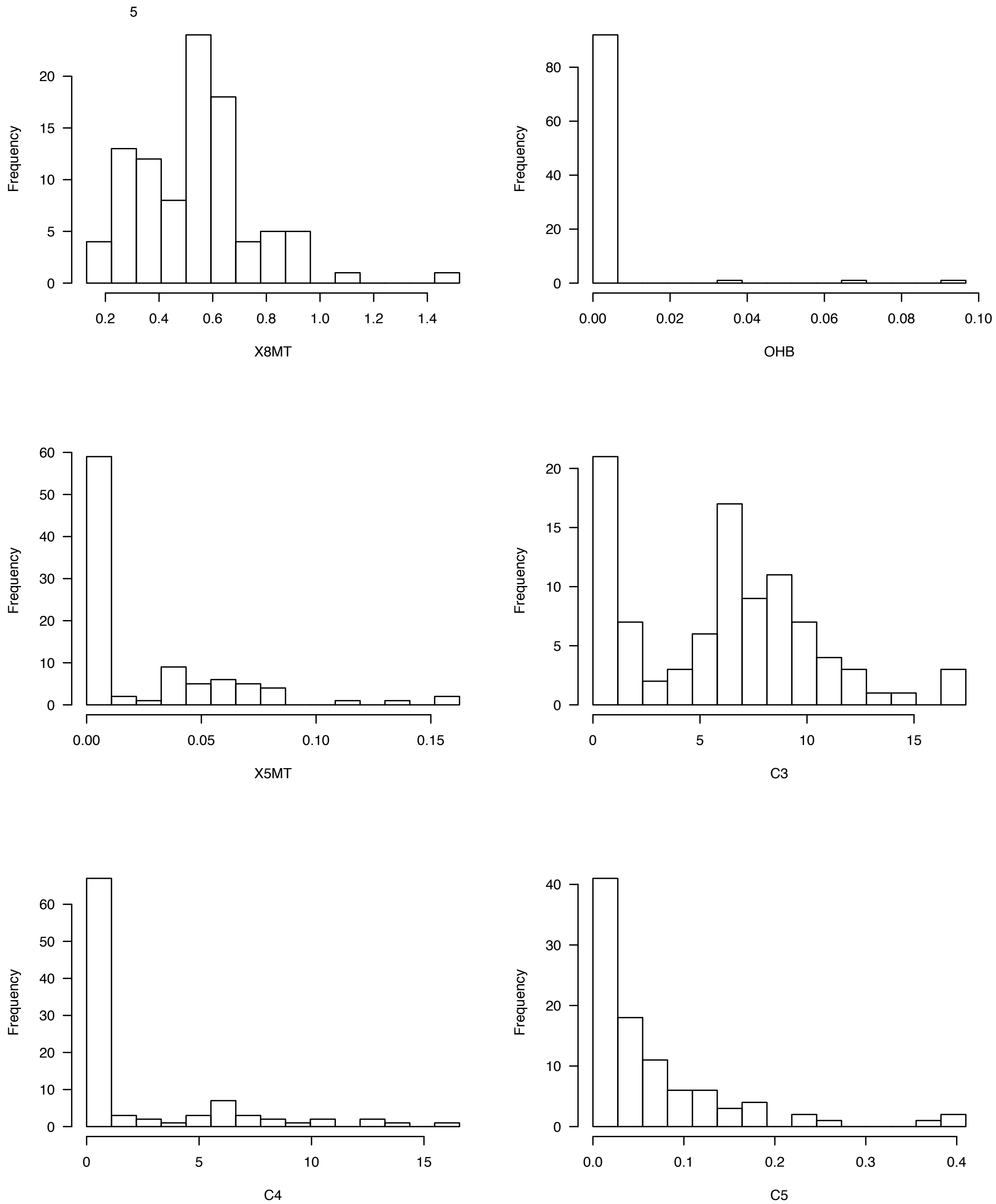

Frequency
0
5
10
15
20
0.5
1.0
1.5
2.0
2.5
3.0
3.5
X8MSO

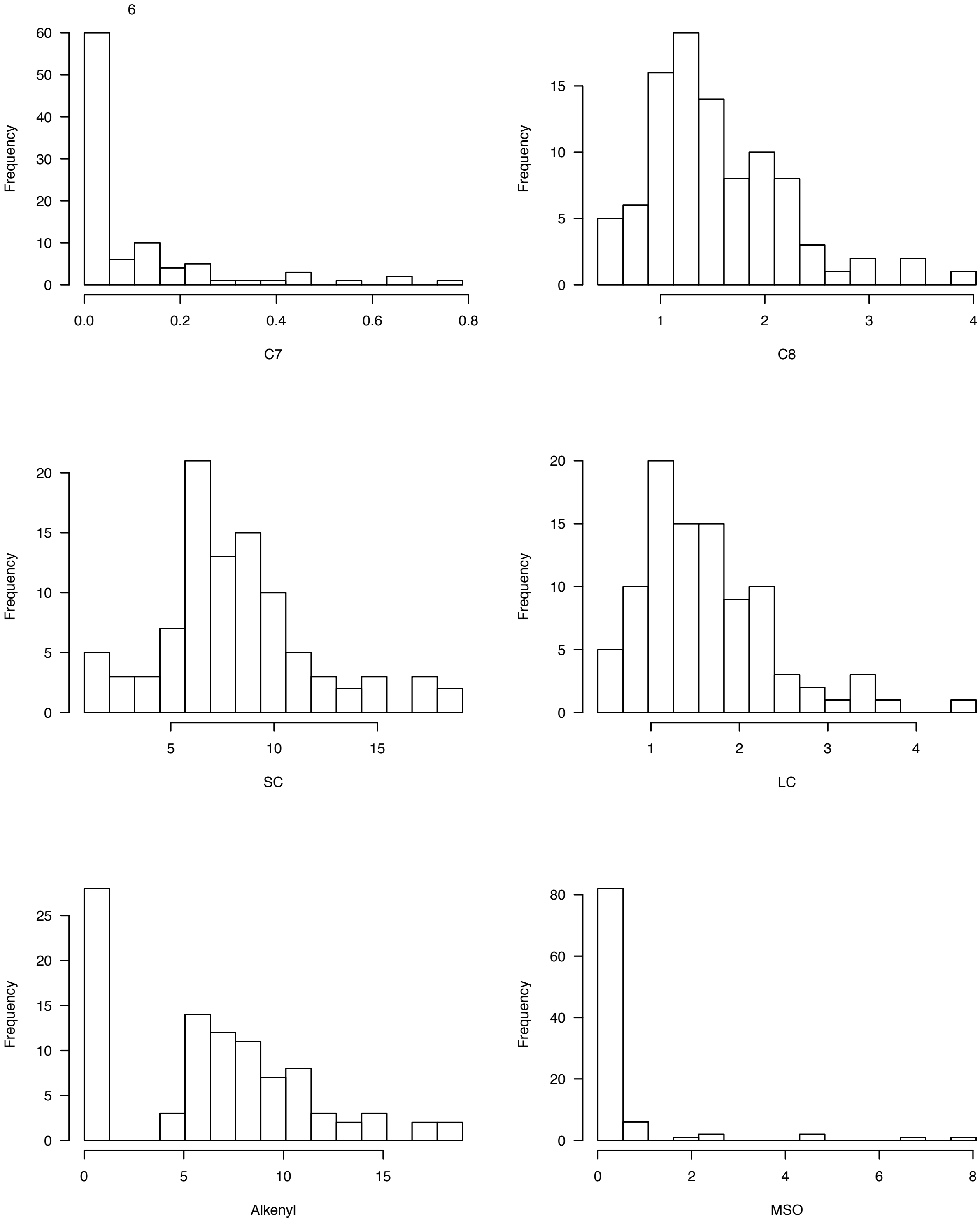

Frequency
0
2
4
6
8
10
12
14
0.5
1.0
1.5
I3M

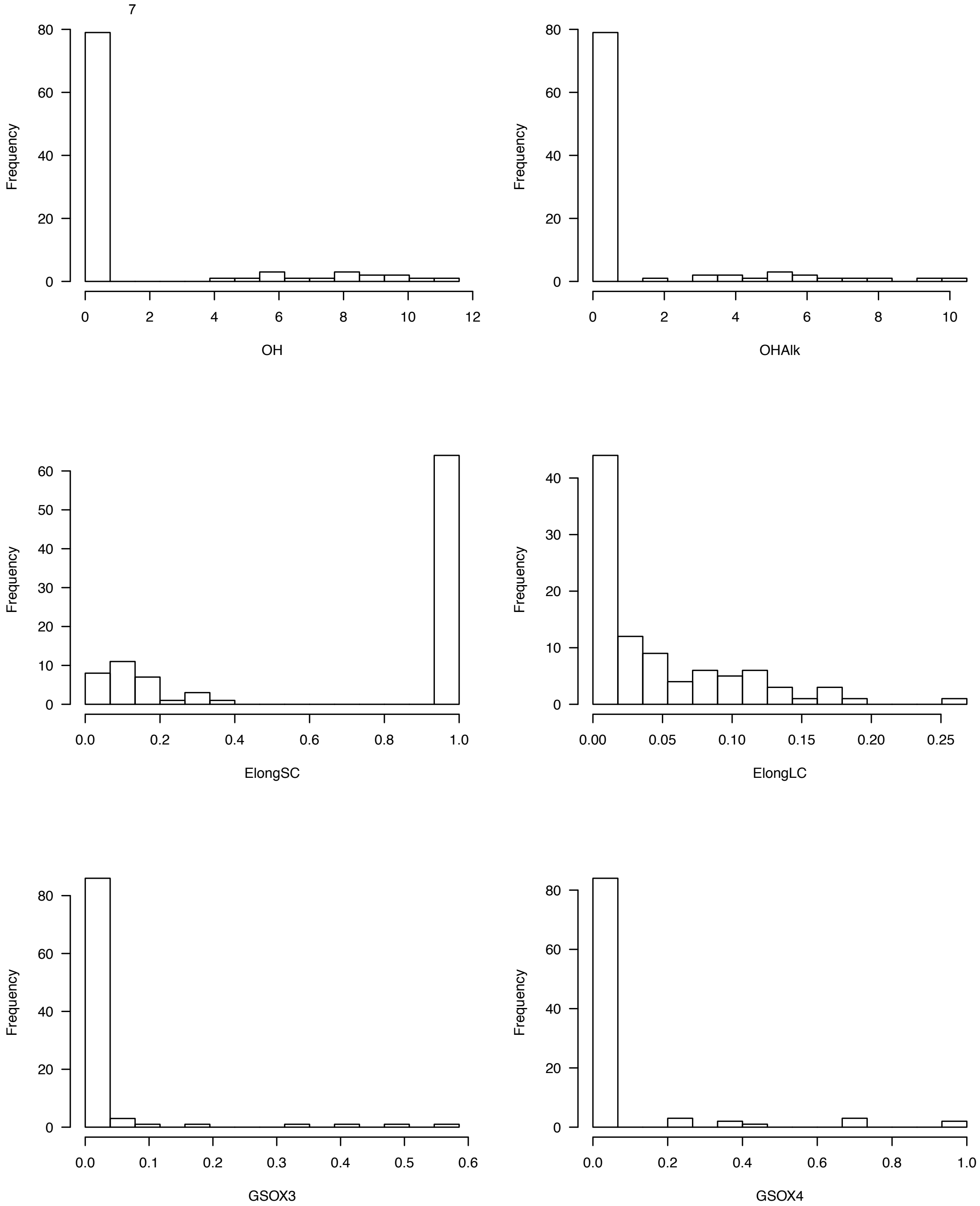

Frequency
0
5
10
15
20
25
0.1
0.2
0.3
0.4
0.5
0.6
X4MOI3M

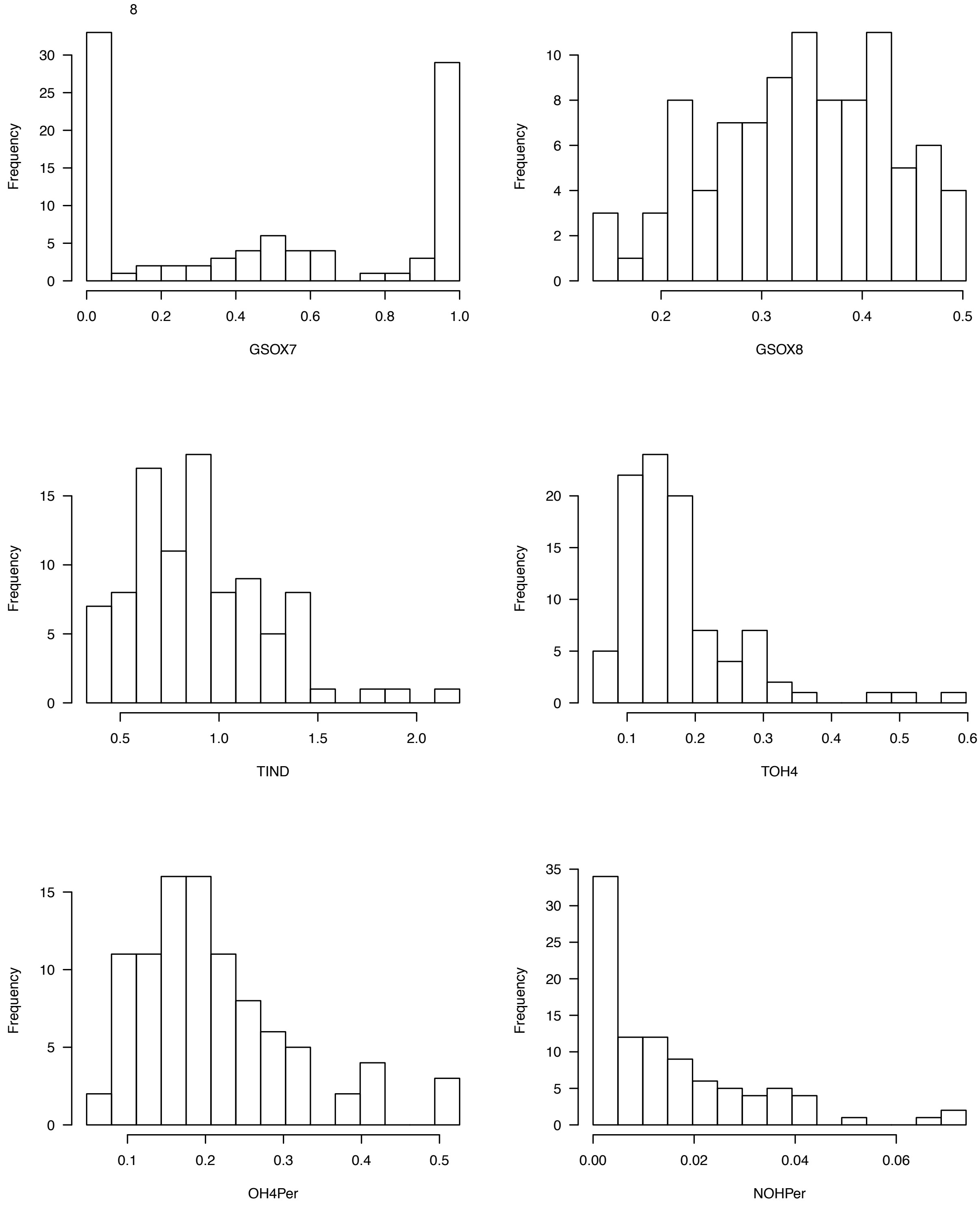

Frequency
0
5
10
15
20
25
30
0.00
0.02
0.04
0.06
NMOI3M

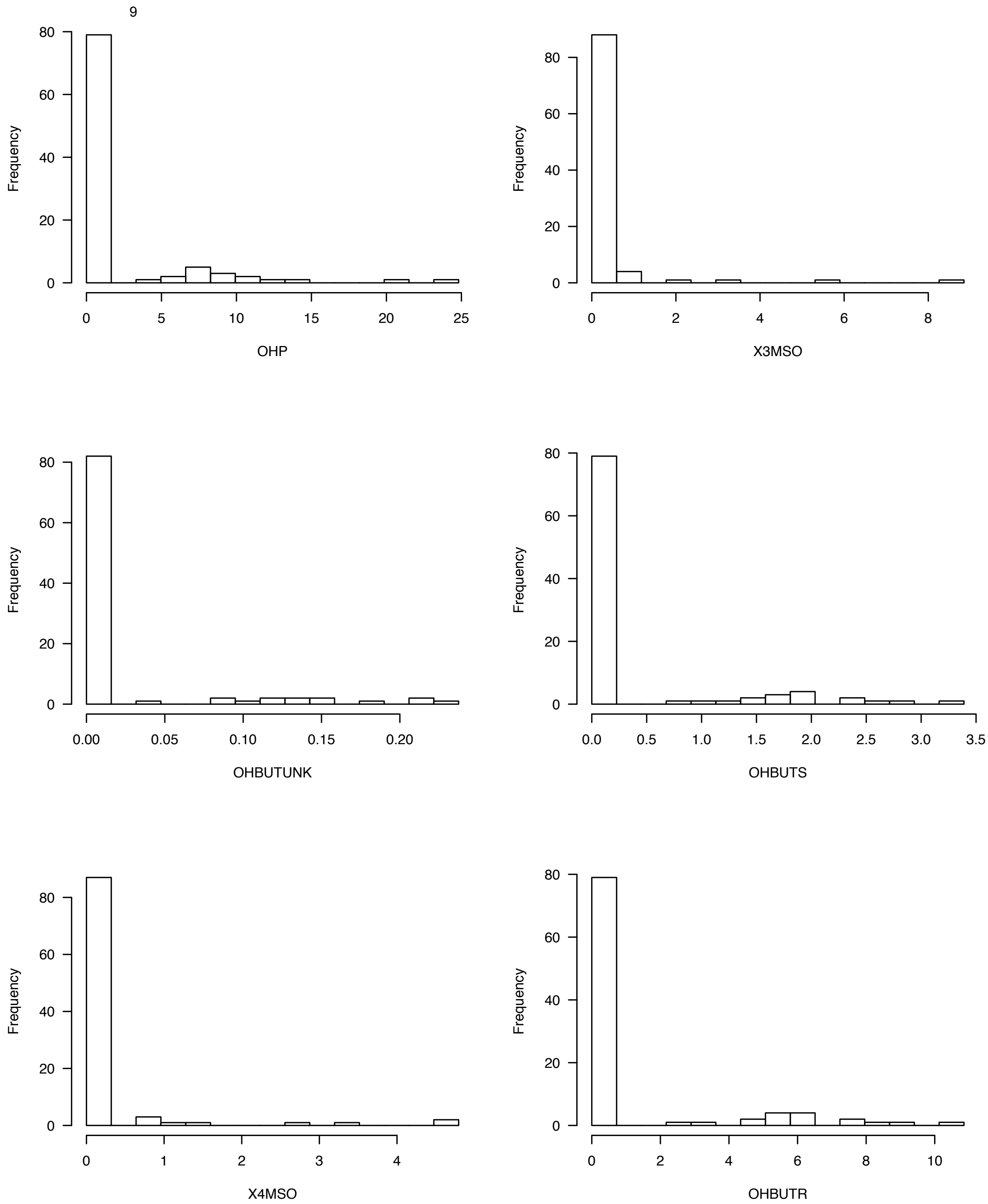

Frequency
0
10
20
30
40
50
60
0.0
0.1
0.2
0.3
0.4
0.5
X7MT


5
Frequency
0
5
10
15
20
0.2
0.4
0.6
0.8
1.0
1.2
1.4
X8MT
Frequency
0
20
40
60
80
0.00
0.02
0.04
0.06
0.08
0.10
OHB
Frequency
0
10
20
30
40
50
60
0.00
0.05
0.10
0.15
X5MT
Frequency
0
5
10
15
20
0
5
10
15
C3
Frequency
0
10
20
30
40
50
60
0
5
10
15
C4
Frequency
0
10
20
30
40
0.0
0.1
0.2
0.3
0.4
C5

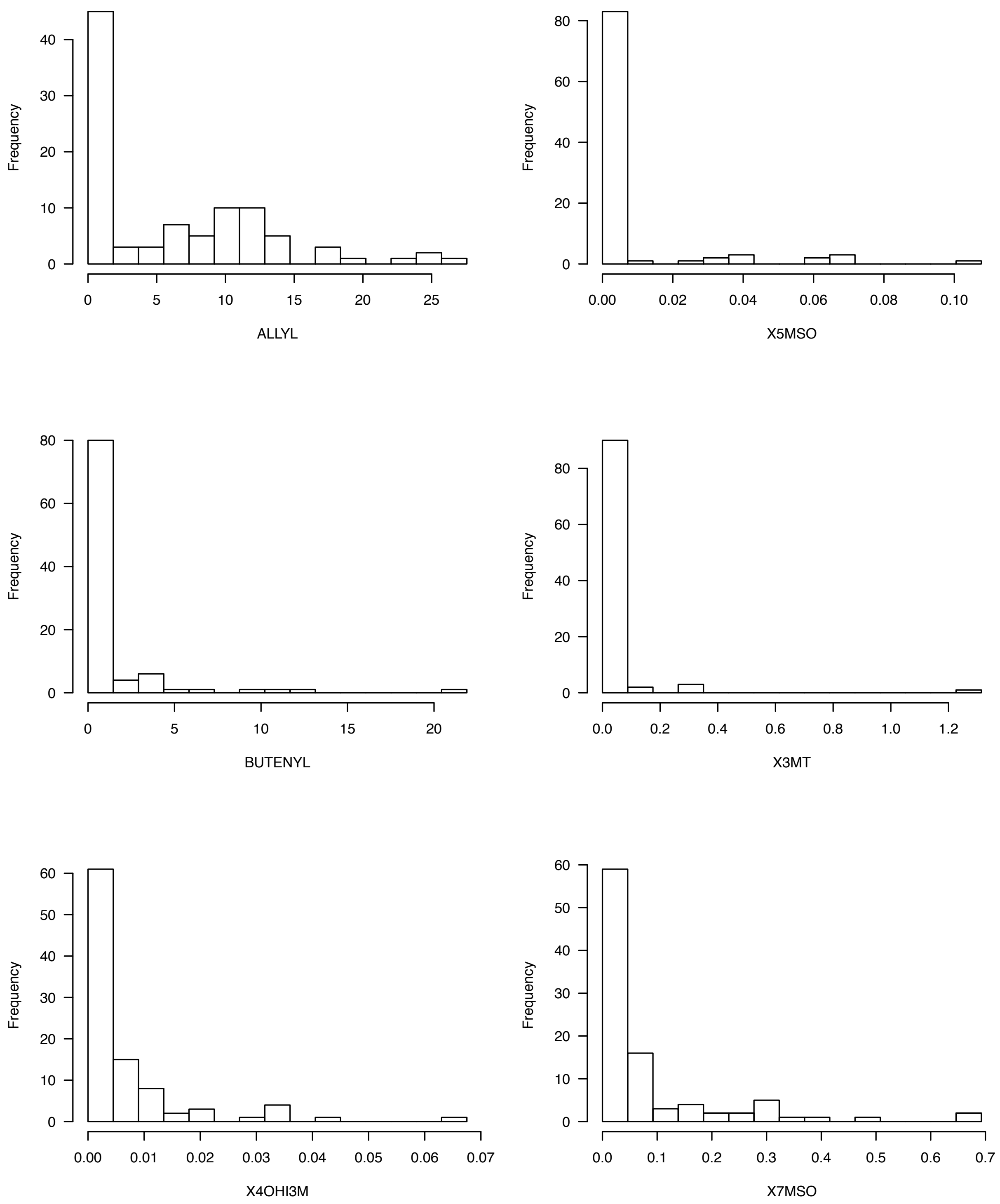

6
Frequency
0
10
20
30
40
50
60
0.0
0.2
0.4
0.6
0.8
C7
Frequency
0
5
10
15
1
2
3
4
C8
Frequency
0
5
10
15
20
5
10
15
SC
Frequency
0
5
10
15
20
1
2
3
4
LC
Frequency
0
5
10
15
20
25
0
5
10
15
Alkenyl
Frequency
0
20
40
60
80
0
2
4
6
8
MSO

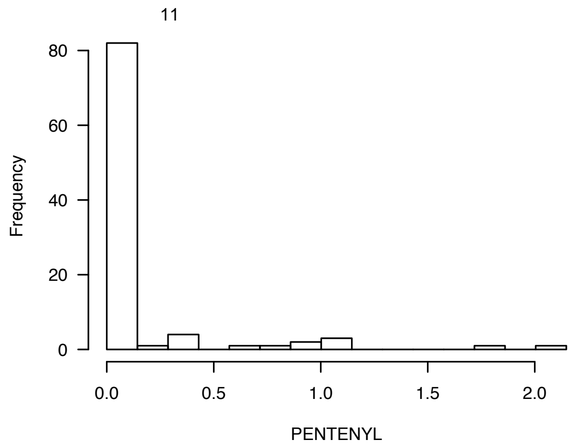

7
Frequency
80
60
40
20
0
0
2
4
6
8
10
12
OH
Frequency
80
60
40
20
0
0
2
4
6
8
10
OHAlk
Frequency
60
50
40
30
20
10
0
0.0
0.2
0.4
0.6
0.8
1.0
ElongSC
Frequency
40
30
20
10
0
0.00
0.05
0.10
0.15
0.20
0.25
ElongLC
Frequency
80
60
40
20
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
GSOX3
Frequency
80
60
40
20
0
0.0
0.2
0.4
0.6
0.8
1.0
GSOX4

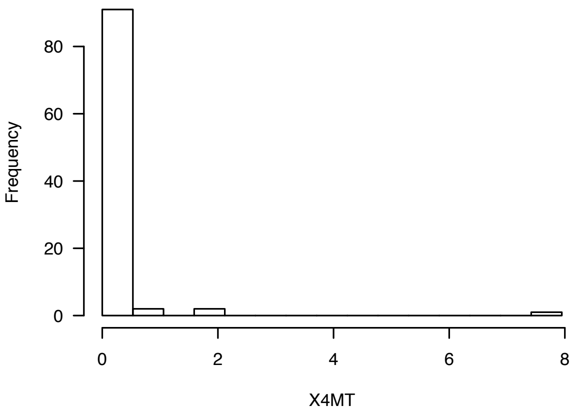

8
Frequency
0
5
10
15
20
25
30
0.0
0.2
0.4
0.6
0.8
1.0
GSOX7
Frequency
0
2
4
6
8
10
0.2
0.3
0.4
0.5
GSOX8
Frequency
0
5
10
15
0.5
1.0
1.5
2.0
TIND
Frequency
0
5
10
15
20
0.1
0.2
0.3
0.4
0.5
0.6
TOH4
Frequency
0
5
10
15
0.1
0.2
0.3
0.4
0.5
OH4Per
Frequency
0
5
10
15
20
25
30
35
0.00
0.02
0.04
0.06
NOHPer

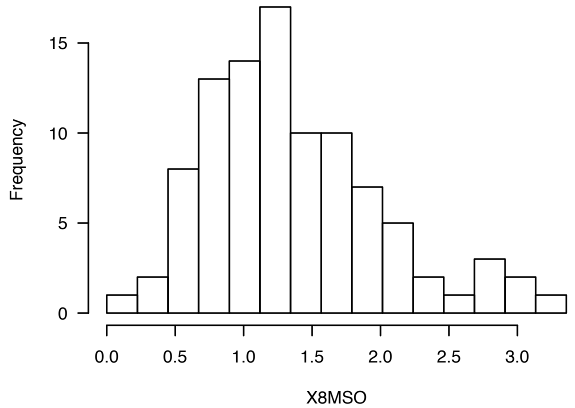

9
Frequency
0
20
40
60
80
0
5
10
15
20
25
OHP
Frequency
0
20
40
60
80
0
2
4
6
8
X3MSO
Frequency
0
20
40
60
80
0.00
0.05
0.10
0.15
0.20
OHBUTUNK
Frequency
0
20
40
60
80
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
OHBUTS
Frequency
0
20
40
60
80
0
1
2
3
4
X4MSO
Frequency
0
20
40
60
80
0
2
4
6
8
10
OHBUTR

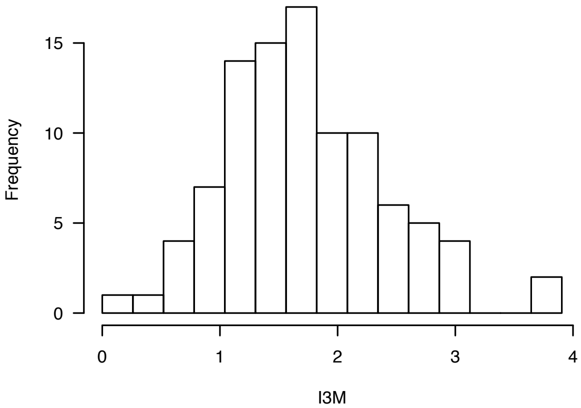
10

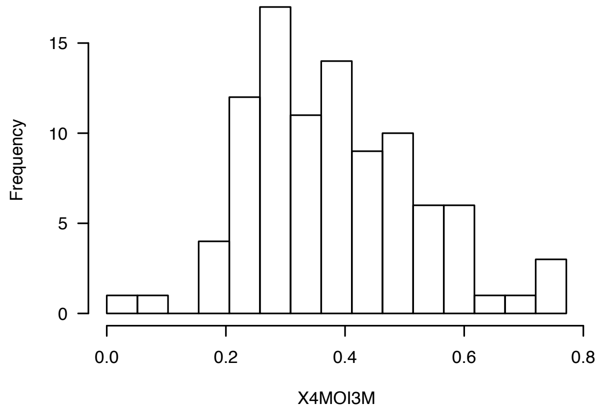

Frequency
0
10
20
30
40
0
5
10
15
20
25
ALLYL
Frequency
0
20
40
60
80
0.00
0.02
0.04
0.06
0.08
0.10
X5MSO
Frequency
0
20
40
60
80
0
5
10
15
20
BUTENYL
Frequency
0
20
40
60
80
0.0
0.2
0.4
0.6
0.8
1.0
1.2
X3MT
Frequency
0
10
20
30
40
50
60
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
X4OHI3M
Frequency
0
10
20
30
40
50
60
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
X7MSO

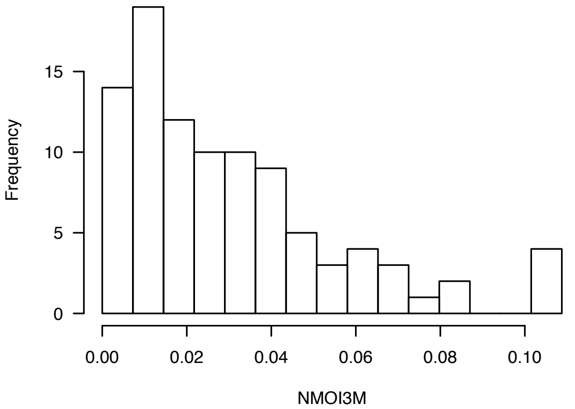

11
Frequency
0
20
40
60
80
0.0
0.5
1.0
1.5
2.0
PENTENYL

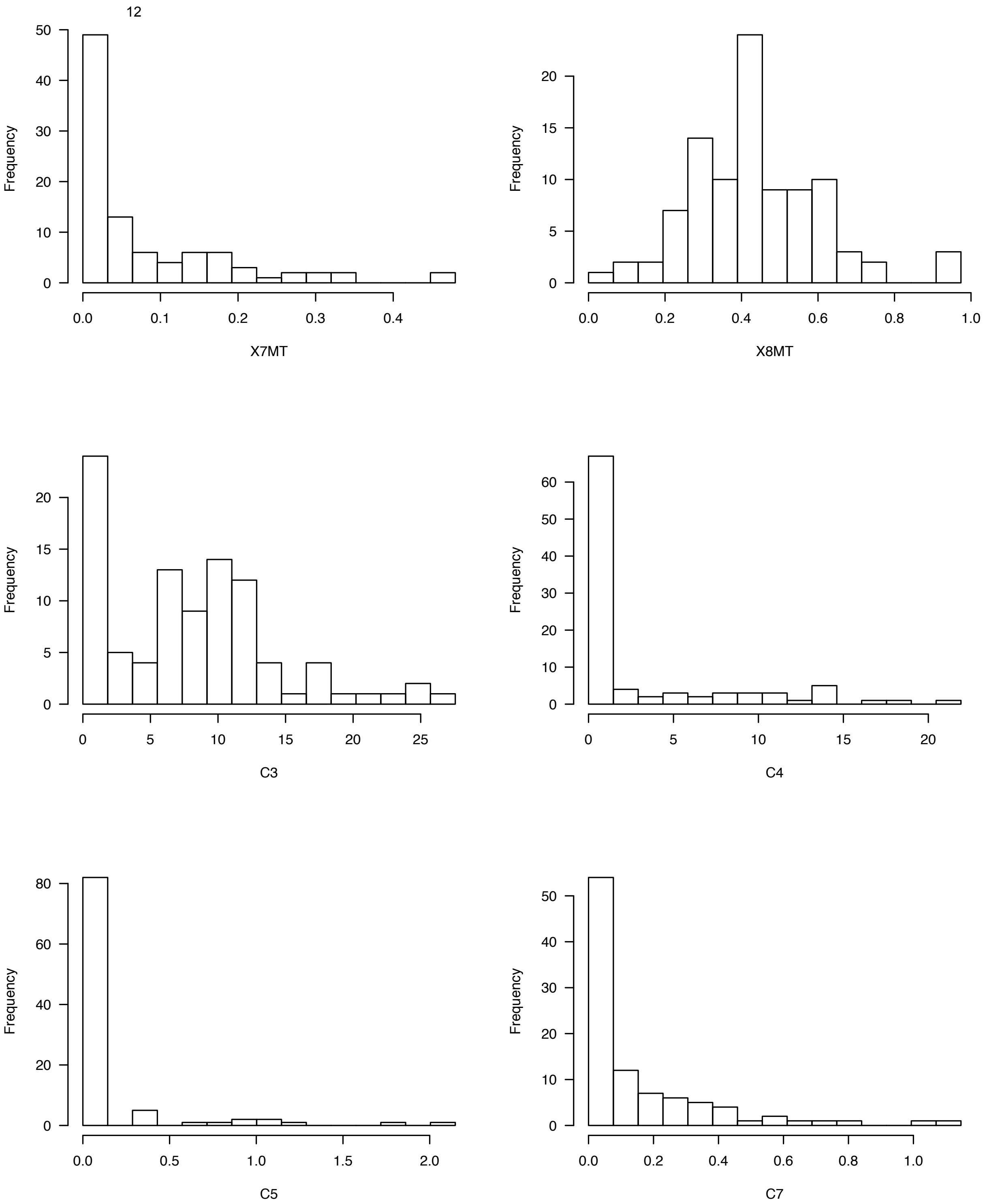

Frequency
0
20
40
60
80
0
2
4
6
8
X4MT

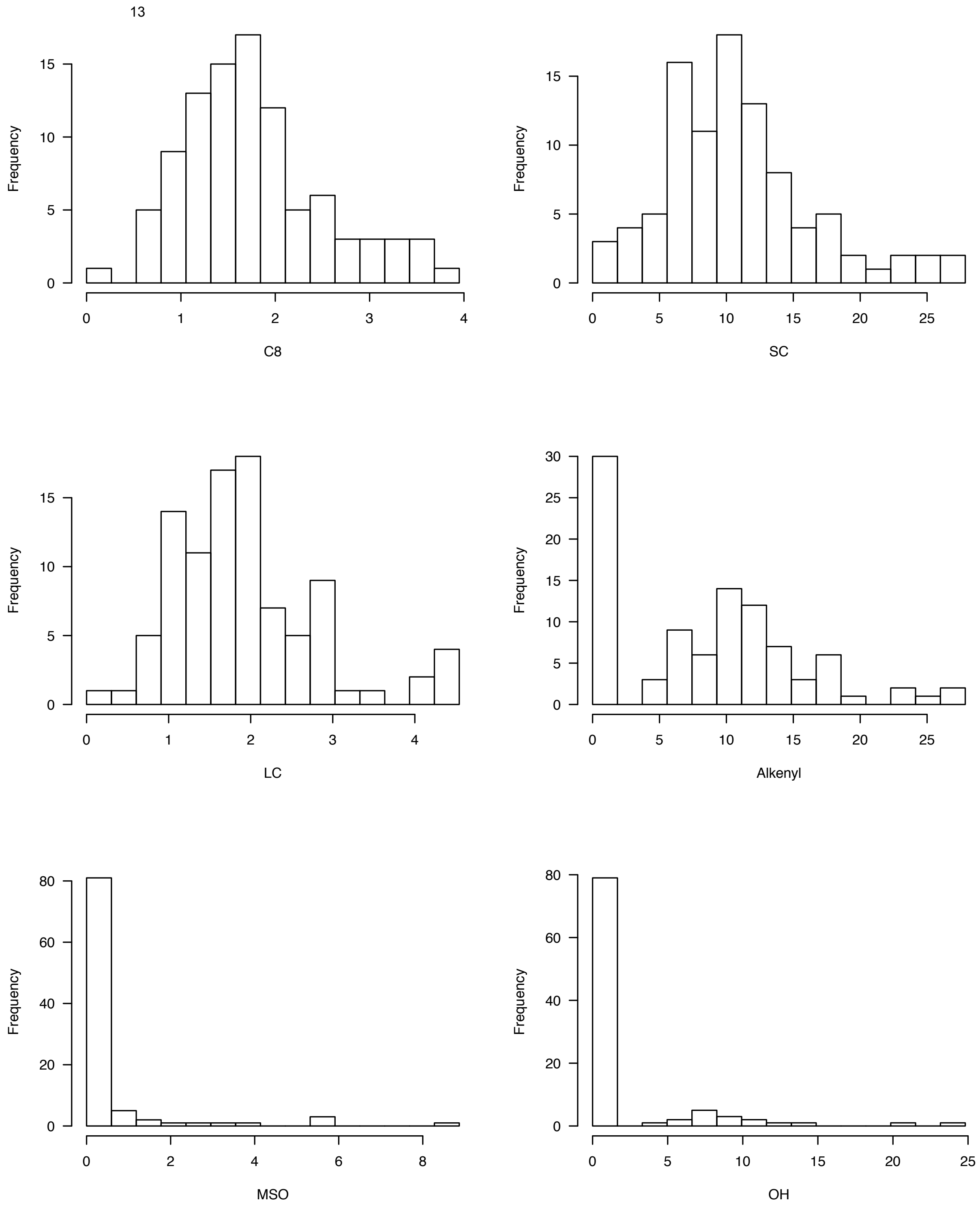

Frequency
0
5
10
15
0.0
0.5
1.0
1.5
2.0
2.5
3.0
X8MSO

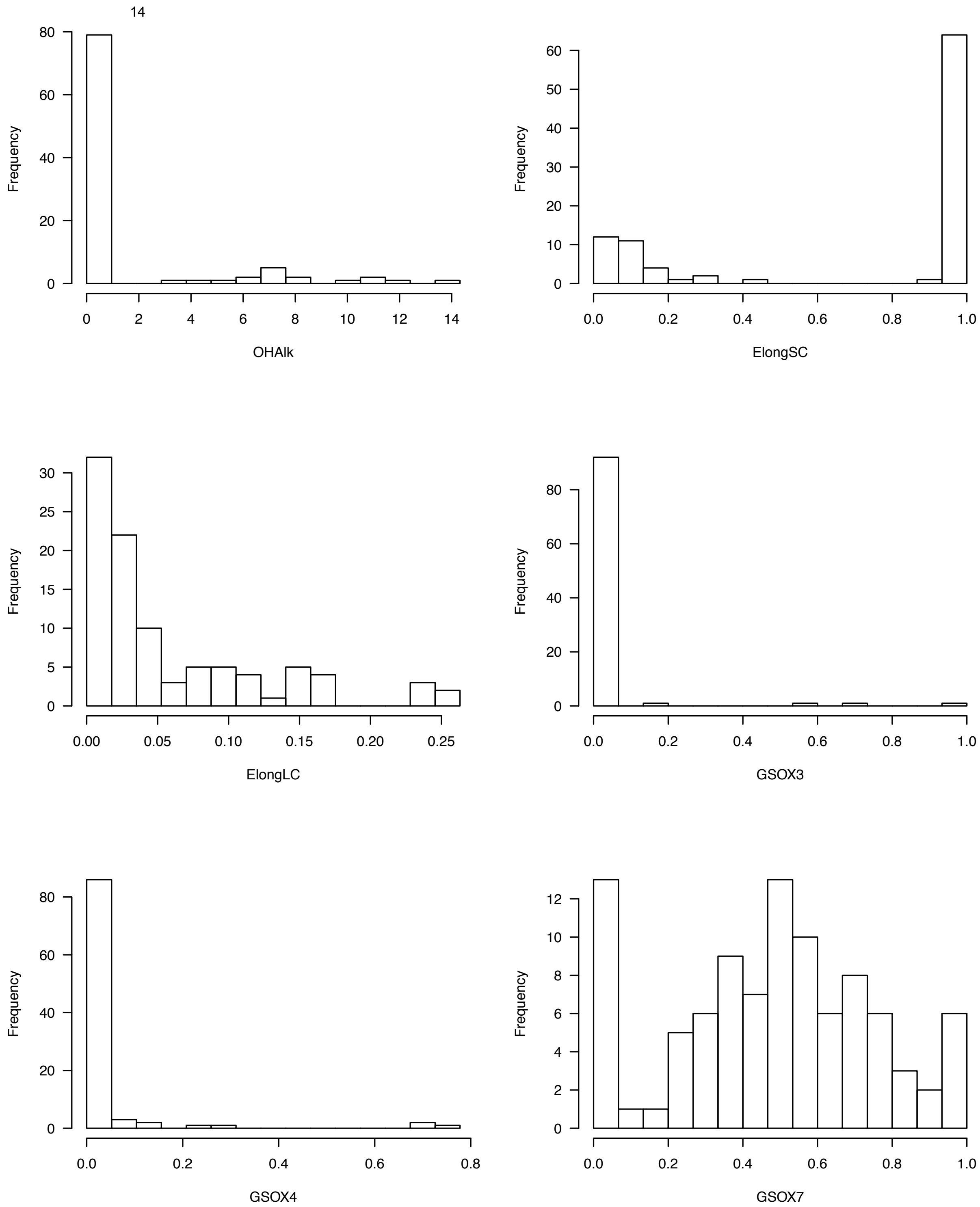

Frequency
0
5
10
15
0
1
2
3
4
I3M


Frequency
0
5
10
15
0.0
0.2
0.4
0.6
0.8
X4MOI3M


Frequency
0
5
10
15
0.00
0.02
0.04
0.06
0.08
0.10
NMOI3M


12
Frequency
0
10
20
30
40
50
0.0
0.1
0.2
0.3
0.4
X7MT
Frequency
0
5
10
15
20
0.0
0.2
0.4
0.6
0.8
1.0
X8MT
Frequency
0
5
10
15
20
0
5
10
15
20
25
C3
Frequency
0
10
20
30
40
50
60
0
5
10
15
20
C4
Frequency
0
20
40
60
80
0.0
0.5
1.0
1.5
2.0
C5
Frequency
0
10
20
30
40
50
0.0
0.2
0.4
0.6
0.8
1.0
C7


13
Frequency
0
5
10
15
0
1
2
3
4
C8
Frequency
0
5
10
15
0
5
10
15
20
25
SC
Frequency
0
5
10
15
0
1
2
3
4
LC
Frequency
0
5
10
15
20
25
30
0
5
10
15
20
25
Alkenyl
Frequency
0
20
40
60
80
0
2
4
6
8
MSO
Frequency
0
20
40
60
80
0
5
10
15
20
25
OH


14
Frequency
0
20
40
60
80
0 2 4 6 8 10 12 14
OHAlk
Frequency
0
10
20
30
40
50
60
0.0 0.2 0.4 0.6 0.8 1.0
ElongSC
Frequency
0
5
10
15
20
25
30
0.00 0.05 0.10 0.15 0.20 0.25
ElongLC
Frequency
0
20
40
60
80
0.0 0.2 0.4 0.6 0.8 1.0
GSOX3
Frequency
0
20
40
60
80
0.0 0.2 0.4 0.6 0.8
GSOX4
Frequency
0
2
4
6
8
10
12
0.0 0.2 0.4 0.6 0.8 1.0
GSOX7

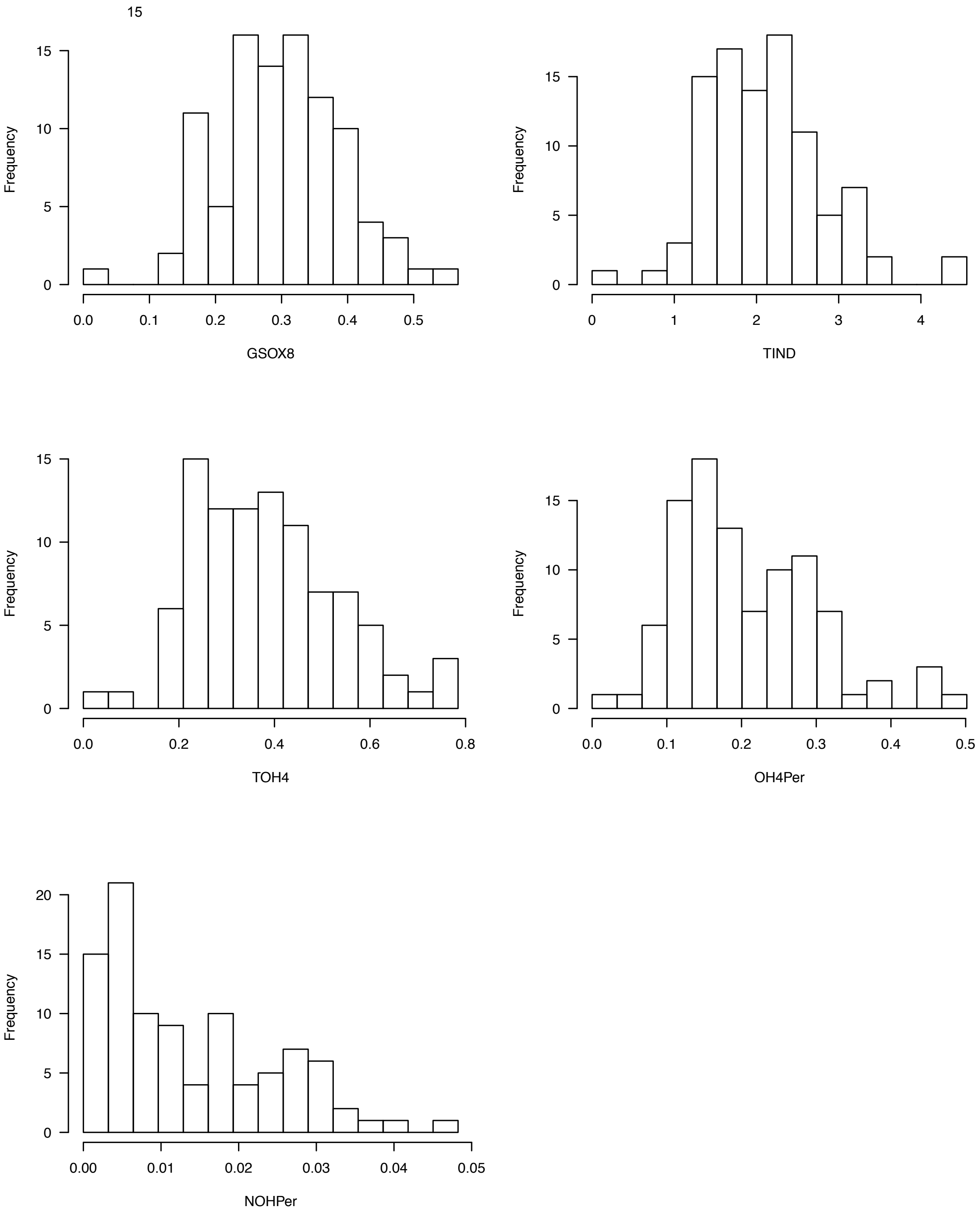
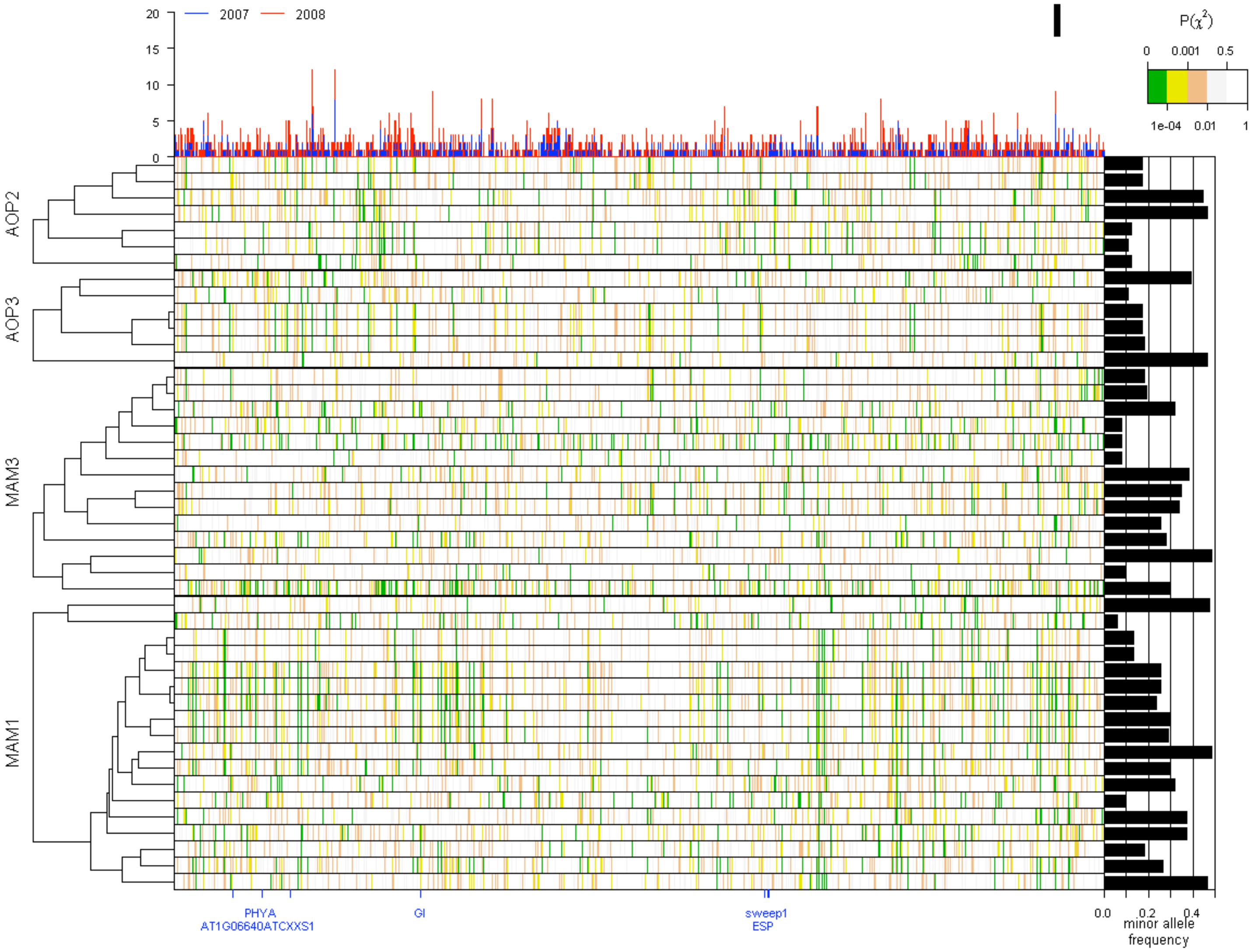
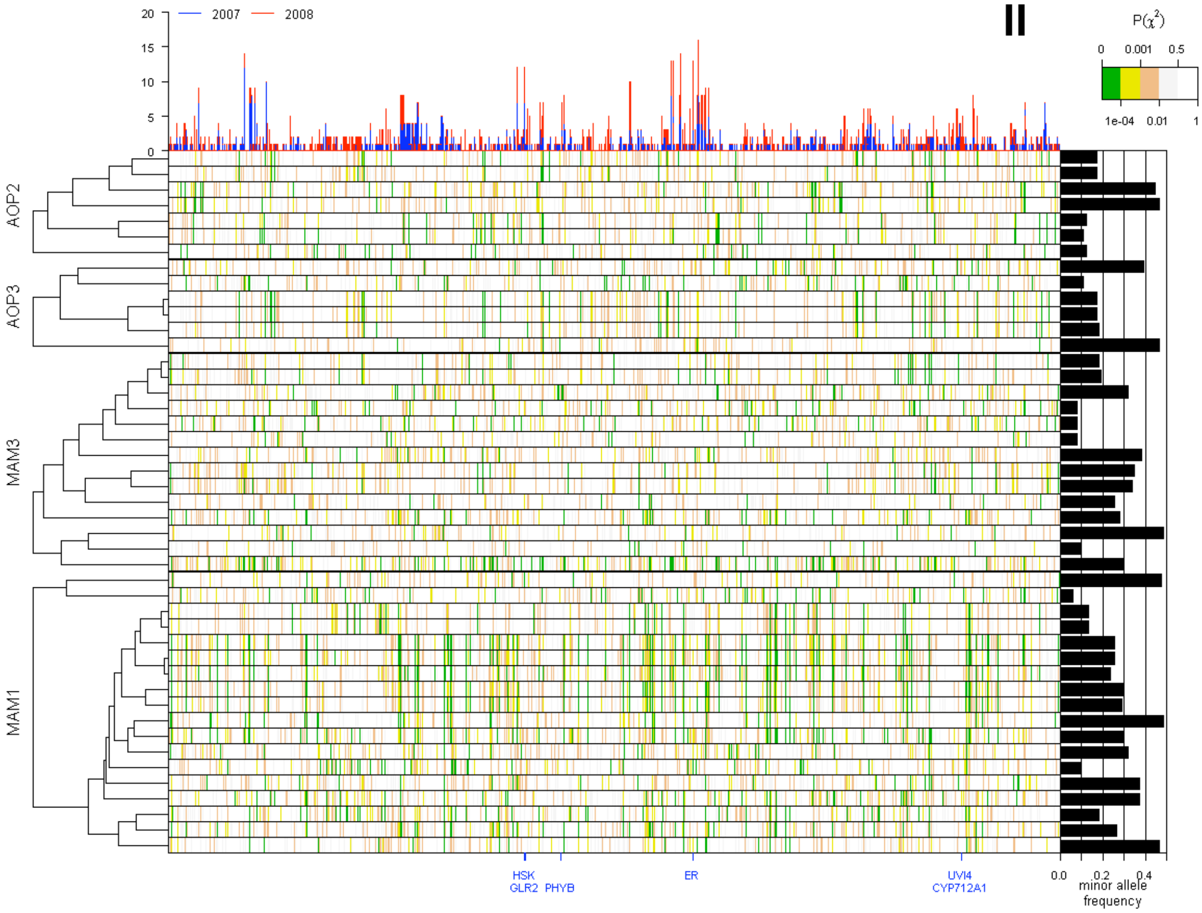
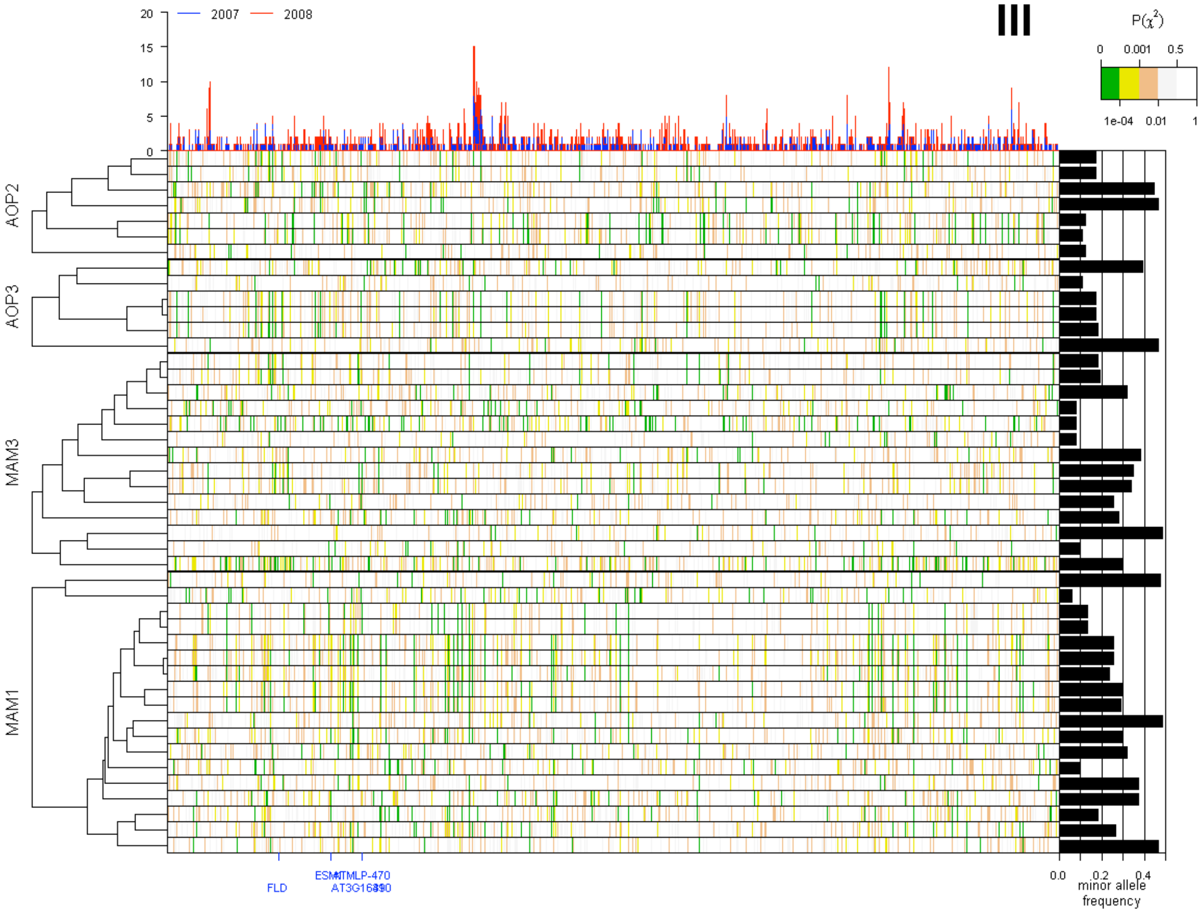
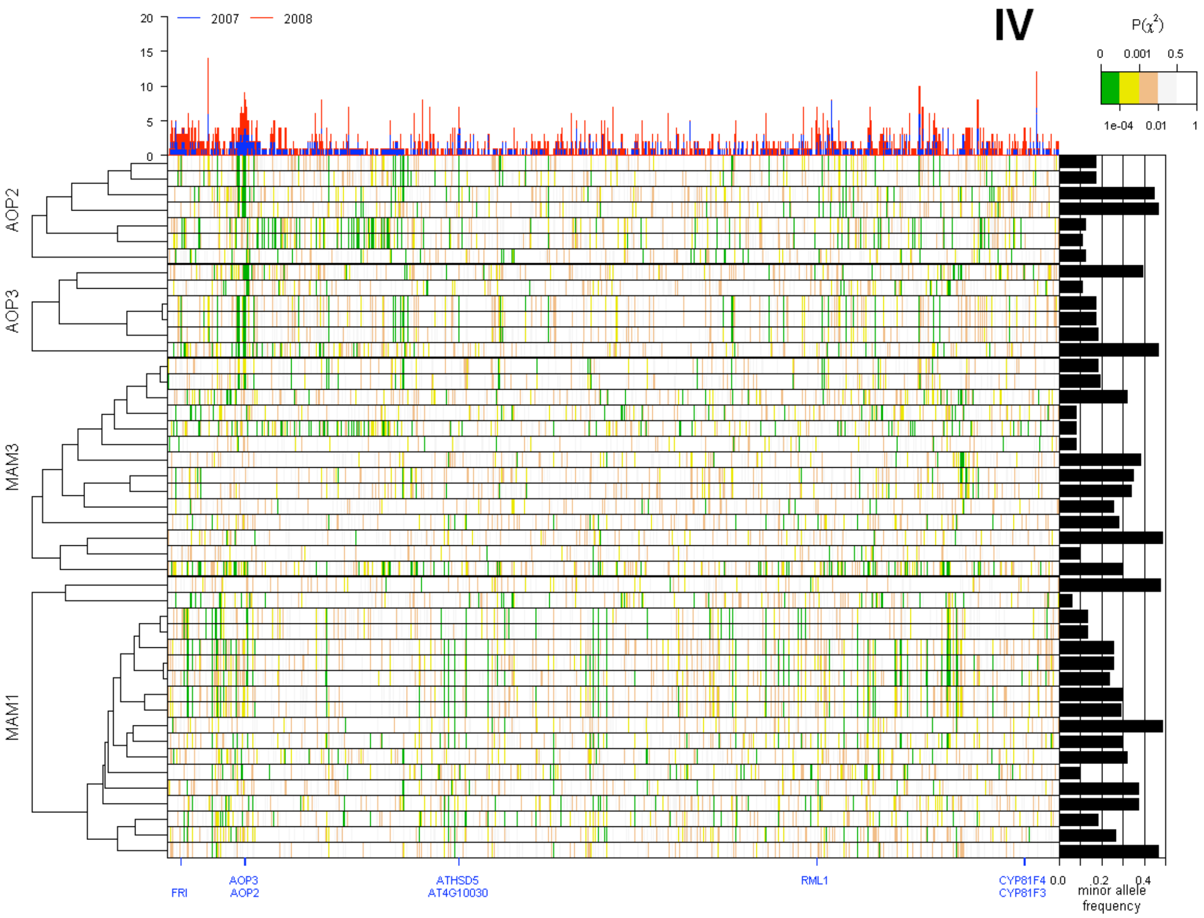


FIGURE S1.—GSL distributions in 2007 and 2008. Shown on page 1 is a direct comparison for the distribution of the total aliphatic and indolic glucosinolates for the two datasets. The graphs on the remaining pages are individual histograms for all utilized glucosinolate phenotypes within each dataset: pages 2-8 are for the 2007 dataset and 9-15 are for the 2008 dataset.


20
15
10
5
0
2007
2008
P(χ^2)
0
0.001
0.5
1e-04
0.01
1
AOP2
AOP3
MAM3
MAM1
PHYA
AT1G06640ATCXXS1
GI
sweep1
ESP
0.0
0.2
0.4
minor allele
frequency


20
15
10
5
0
2007
2008
II
P(χ²)
0
0.001
0.5
1e-04
0.01
1
AOP2
AOP3
MAM3
MAM1
HSK
GLR2
PHYB
ER
UVI4
CYP712A1
0.0
0.2
0.4
minor allele frequency


III
20
15
10
5
0
2007
2008
P(χ^2)
0
0.001
0.5
1e-04
0.01
1
AOP2
AOP3
MAM3
MAM1
FLD
ESM1
ATMLP-470
AT3G16390
0.0
0.2
0.4
minor allele frequency


IV
2007
2008
20
15
10
5
0
P(χ²)
0
0.001
0.5
1e-04
0.01
1
AOP2
AOP3
MAM3
MAM1
FRI
AOP3
AOP2
ATHSD5
AT4G10030
RML1
CYP81F4
CYP81F3
0.0
0.2
0.4
minor allele frequency

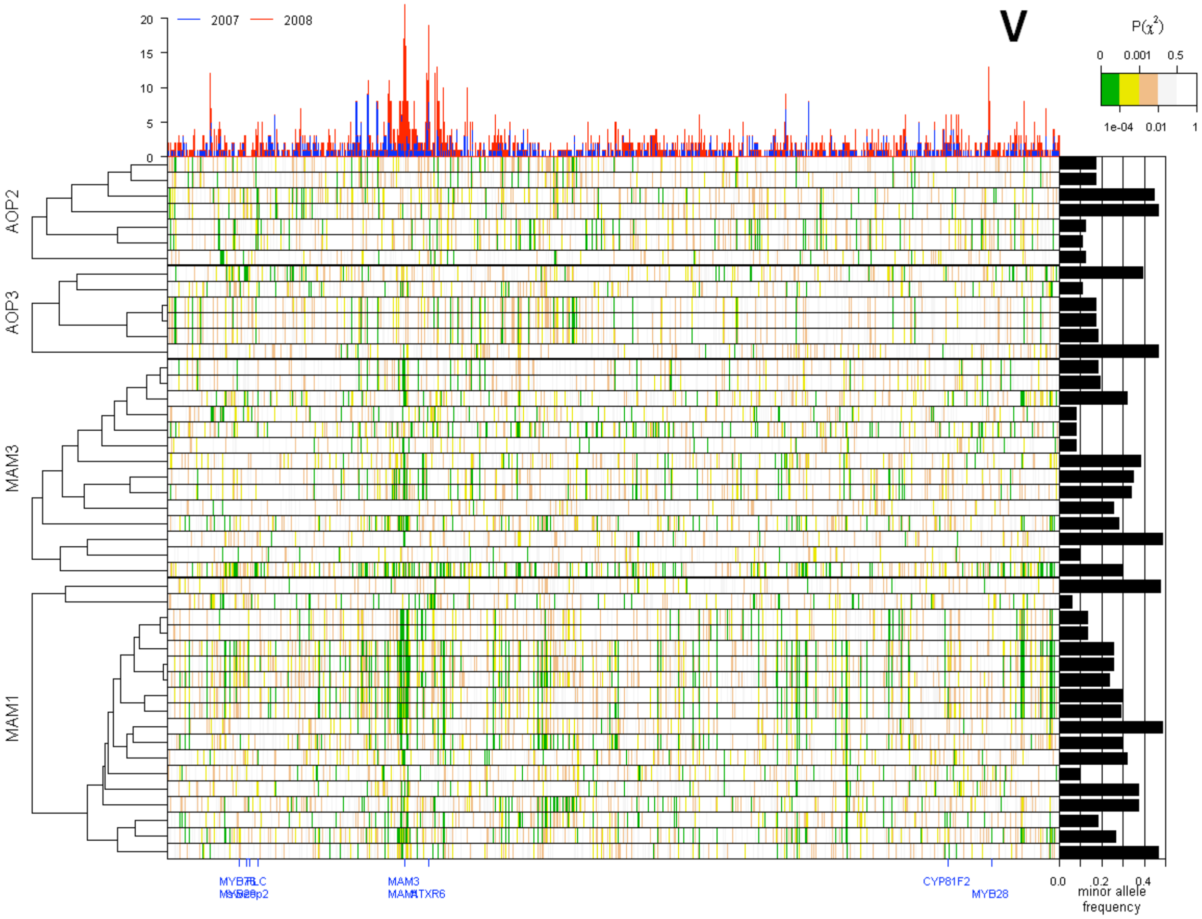
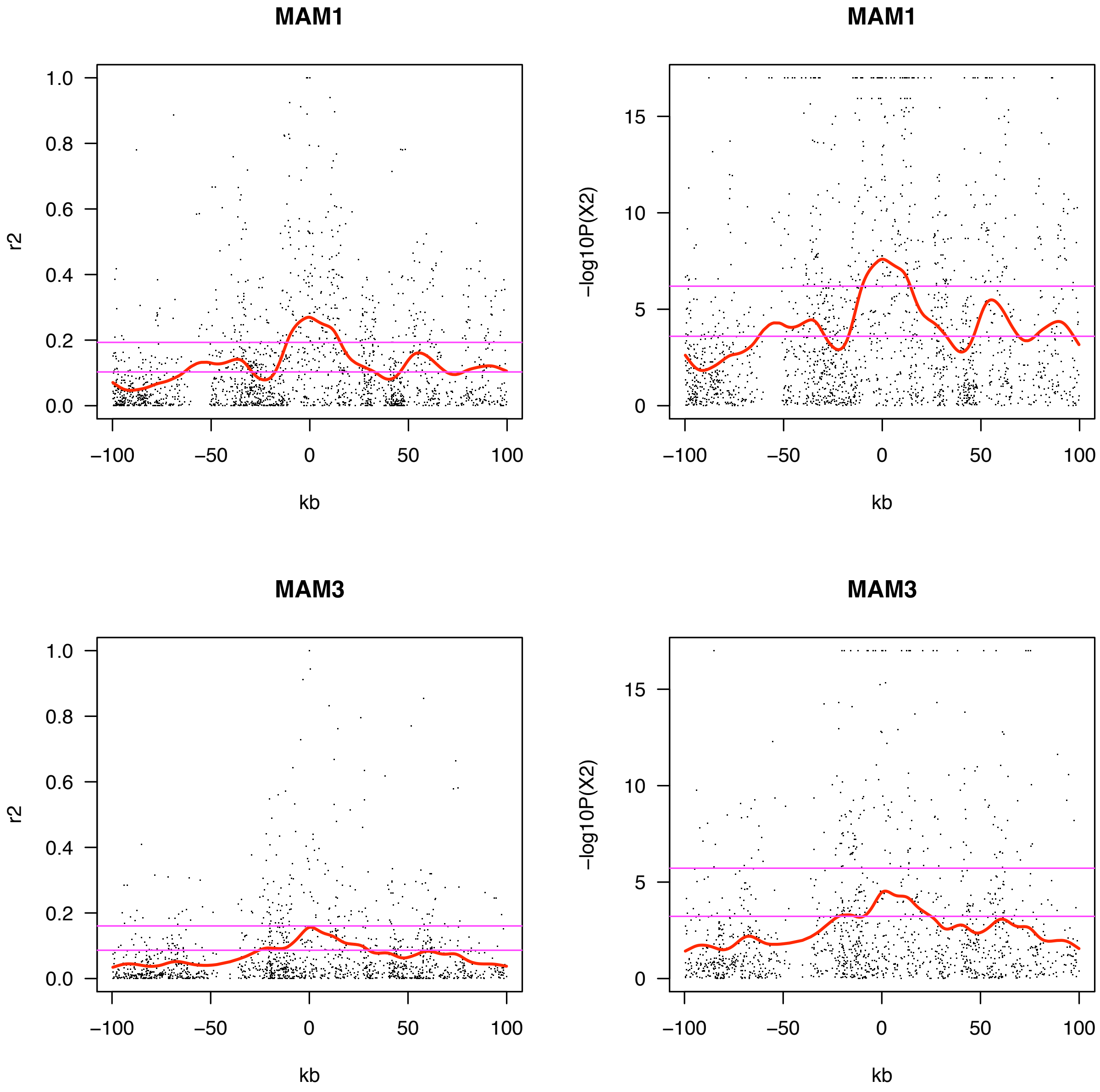
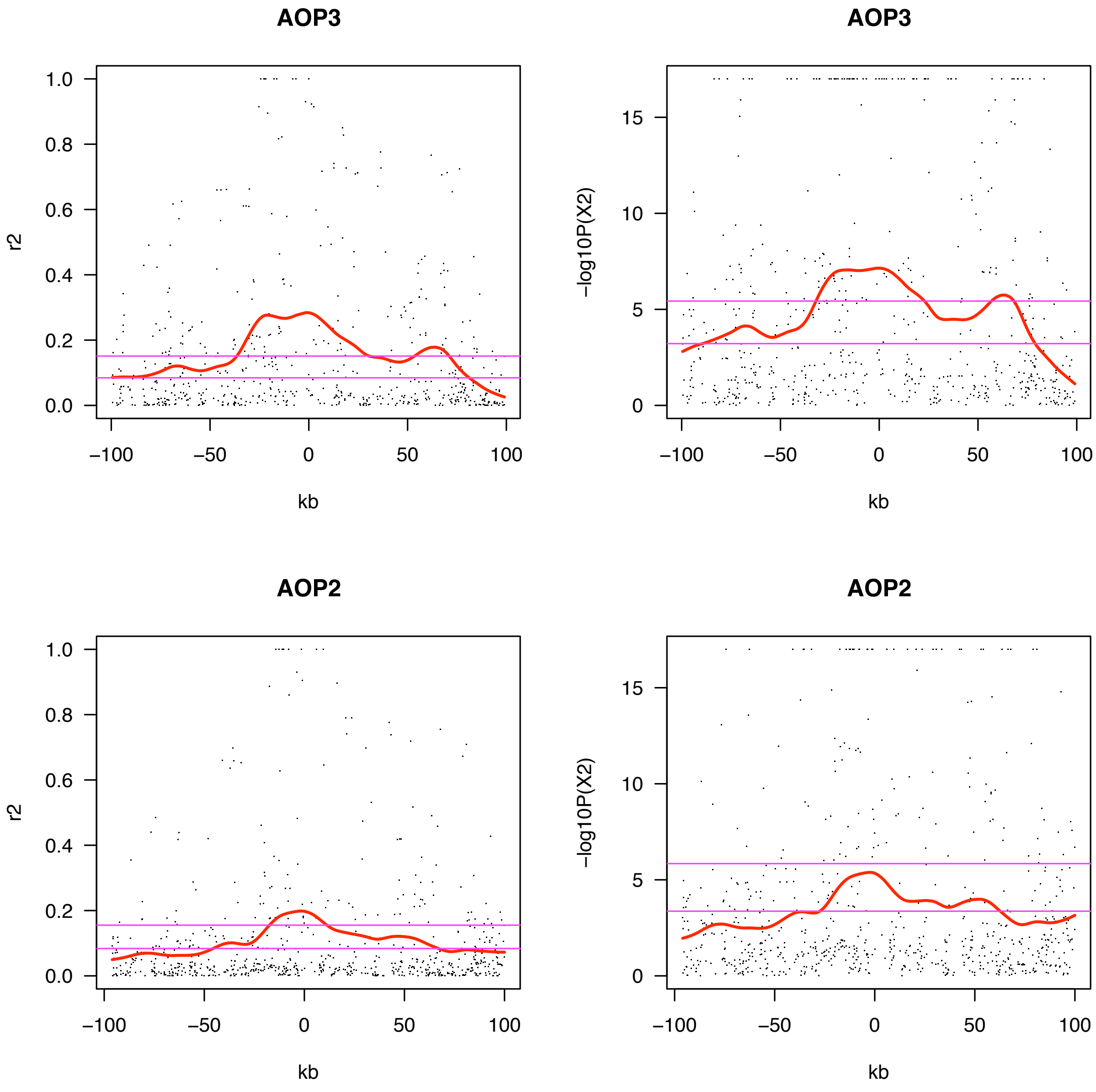


FIGURE S2.—Genome-wide LD with SNPs in the *AOP* and *MAM* loci. Pairwise linkage disequilibrium between SNPs within coding regions of query genes (*AOP2, AOP3, MAM1, and MAM3*) and all ~240,000 SNPs is shown as a heatmap with significantly strong to weak LD indicated by green to white respectively. The top histogram represents the number of GSL phenotypes showing a significant association with the individual SNP in this region for both the 2007 and 2008 datasets. The right bar-graph shows the minor allele frequency of each SNP within the query genes. The cladogram on the left indicates the hierarchical clustering of SNPs within the query genes.


MAM1
r2
1.0
0.8
0.6
0.4
0.2
0.0
−100
−50
0
50
100
kb
MAM1
−log10P(X2)
15
10
5
0
−100
−50
0
50
100
kb
MAM3
r2
1.0
0.8
0.6
0.4
0.2
0.0
−100
−50
0
50
100
kb
MAM3
−log10P(X2)
15
10
5
0
−100
−50
0
50
100
kb


AOP3
r2
1.0
0.8
0.6
0.4
0.2
0.0
−100
−50
0
50
100
kb
AOP3
−log10P(X2)
15
10
5
0
−100
−50
0
50
100
kb
AOP2
r2
1.0
0.8
0.6
0.4
0.2
0.0
−100
−50
0
50
100
kb
AOP2
−log10P(X2)
15
10
5
0
−100
−50
0
50
100
kb

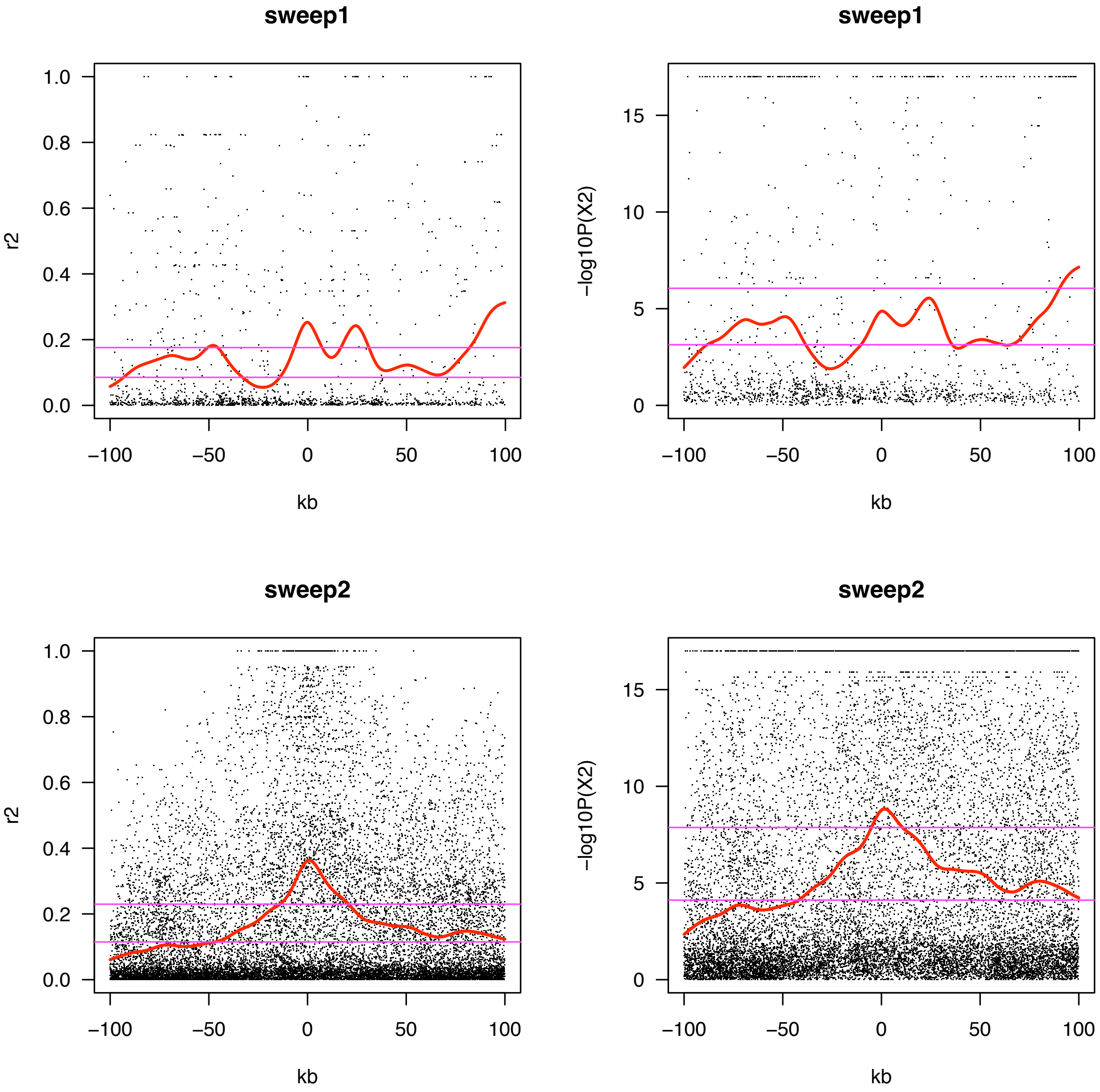
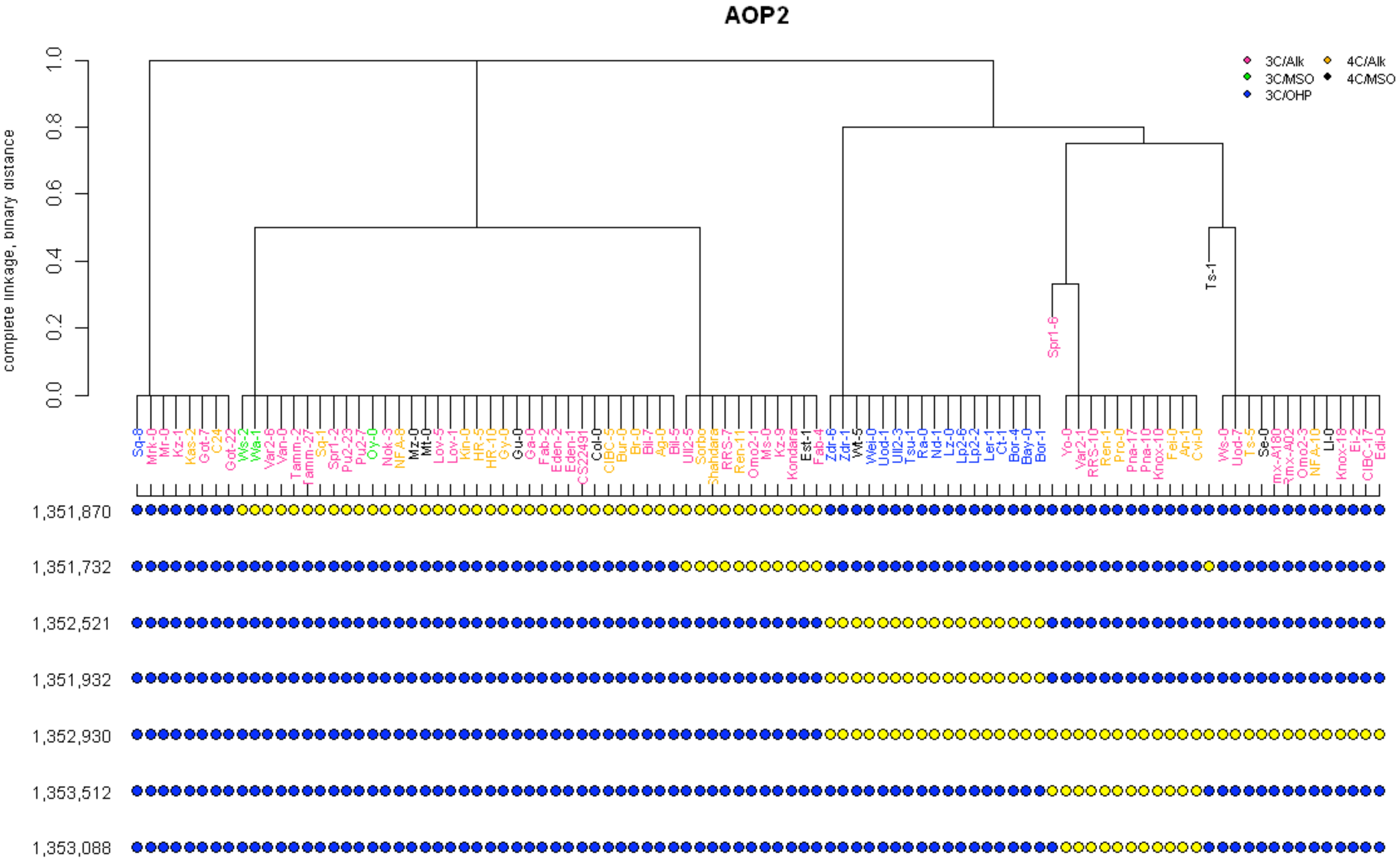
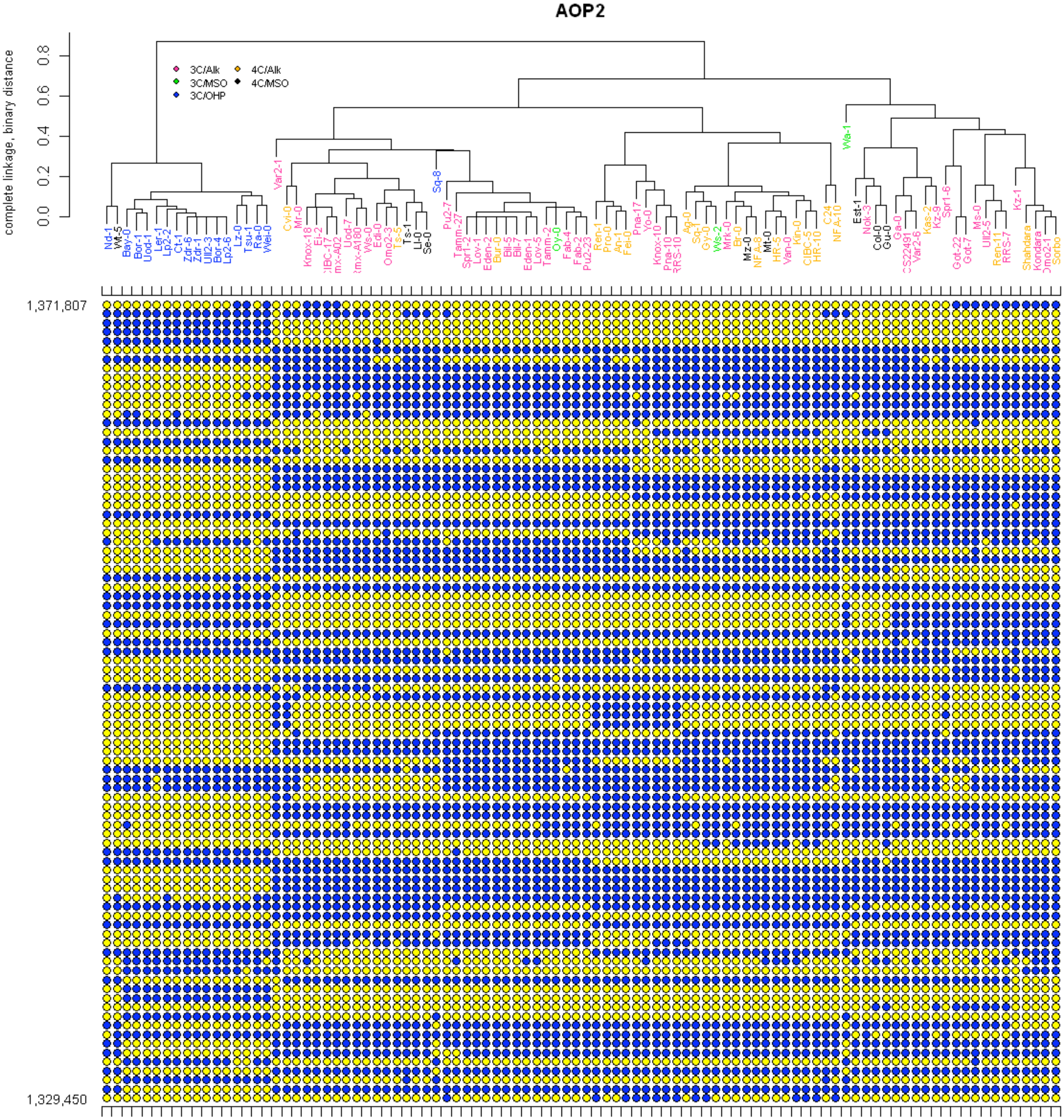
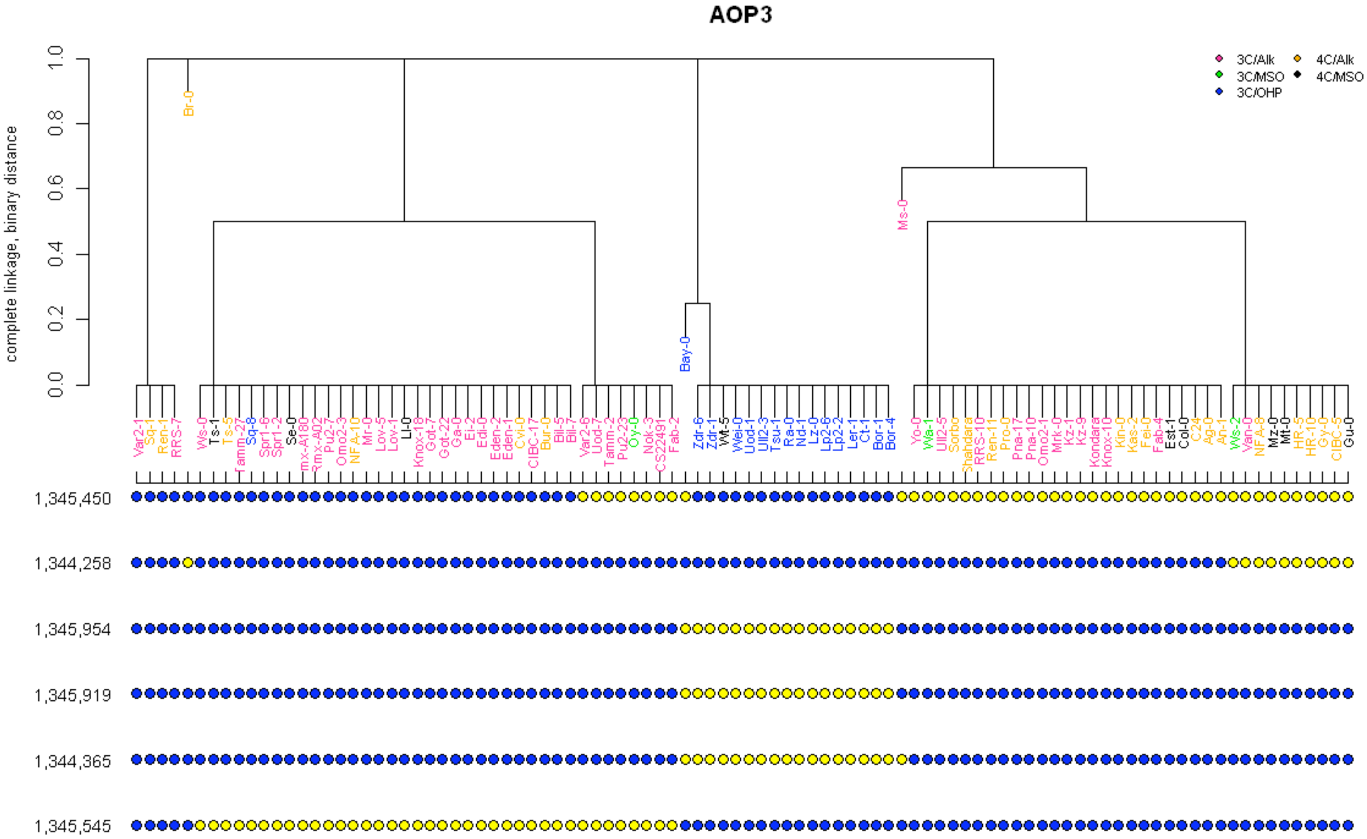
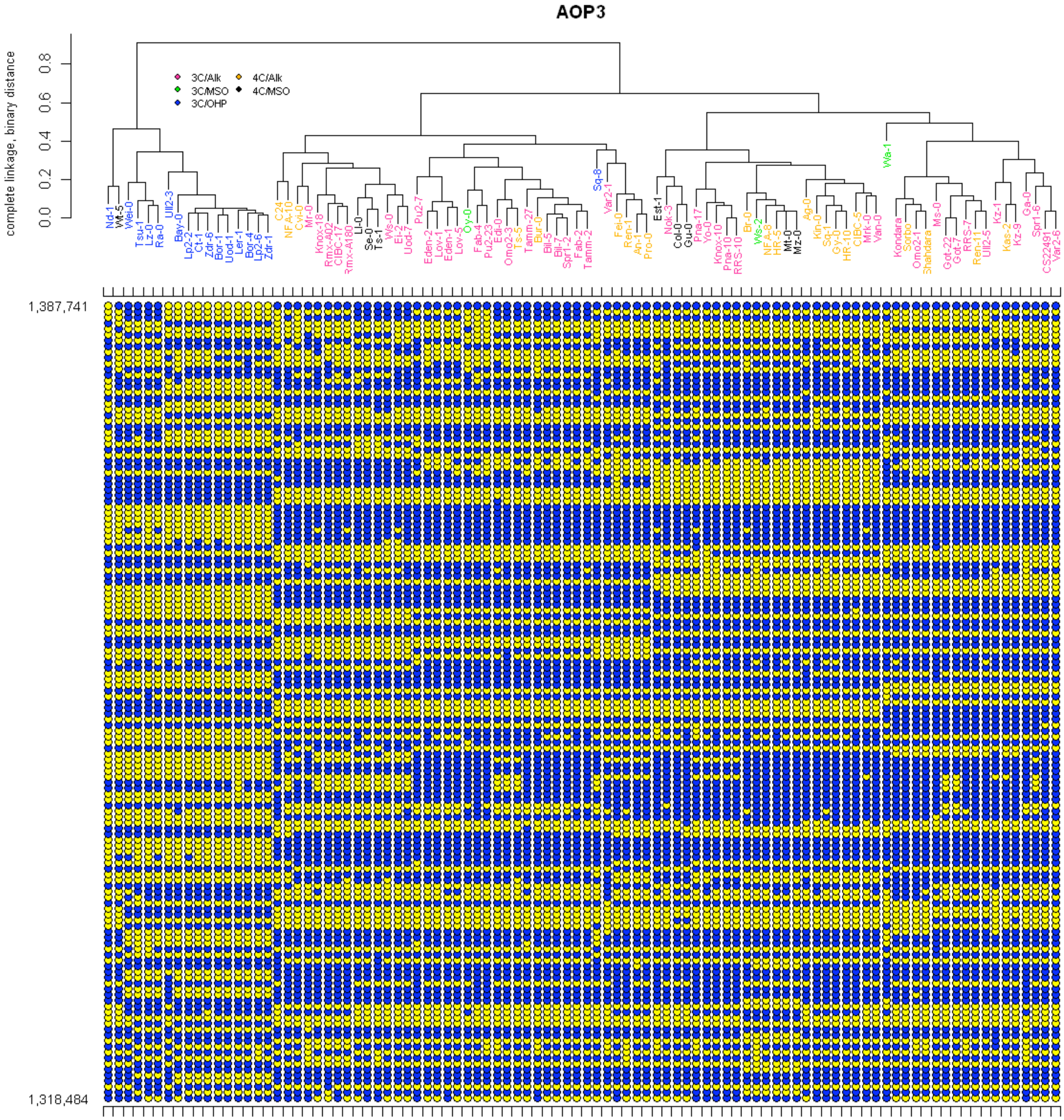


FIGURE S3.—LD decay around *MAM*, *AOP*, and recent selective sweeps. LD, as measured by r^2 and $-\log_{10}P(\chi^2)$, between SNPs within a query gene/region and SNPs within ±100kb around of the query gene is plotted against the physical distance (kilobases) between each SNP-pairs. The four query genes are *MAM1, MAM3, AOP2,* and *AOP3*, and the query regions are two previously proposed recent selective sweep (CLARK *et al.* 2007). The red line shows the fitted cubic smoothing spline (*spar*=1) between LD and distance. The top and bottom pink lines show the 99th and 95th percentiles of non-syntenic LD; i.e. LD between SNPs within the query gene and SNPs on chromosomes outside of the gene. The center of the graph is at the center of the query genes.


AOP2
complete linkage, binary distance
0.0
0.2
0.4
0.6
0.8
1.0
3C/Alk
3C/MSO
3C/OHP
4C/Alk
4C/MSO
Spr1-6
Ts-1
Sq-8
Mrk-0
Mr-0
Kz-1
Kas-2
Got-7
C24
Got-22
Ws-2
Wa-1
Var2-6
Van-0
Tamm-2
Tamm-27
Sq-1
Spr1-2
Pu2-23
Pu2-7
Oy-0
Nok-3
NFA-8
Mz-0
Mt-0
Lov-5
Lov-1
Kin-0
HR-5
HR-10
Gy-0
Gu-0
Ga-0
Fab-2
Eden-2
Eden-1
CS22491
Col-0
CIBC-5
Bur-0
Br-0
Bil-7
Ag-0
Bil-5
Ull2-5
Sorbo
Shahdara
RRS-7
Ren-11
Omo2-1
Ms-0
Kz-9
Kondara
Est-1
Fab-4
Zdr-6
Zdr-1
Wt-5
Wei-0
Uod-1
Ull2-3
Tsu-1
Ra-0
Nd-1
Lz-0
Lp2-6
Lp2-2
Ler-1
Ct-1
Bor-4
Bay-0
Bor-1
Yo-0
Var2-1
RRS-10
Ren-1
Pro-0
Pna-17
Pna-10
Knox-10
Fei-0
An-1
Cvi-0
Ws-0
Uod-7
Ts-5
Se-0
mx-A180
Rmx-A02
Omo2-3
NFA-10
Ll-0
Knox-18
Ei-2
CIBC-17
Edi-0
1,351,870
1,351,732
1,352,521
1,351,932
1,352,930
1,353,512
1,353,088


AOP2
complete linkage, binary distance
0.0
0.2
0.4
0.6
0.8
3C/Alk
3C/MSO
3C/OHP
4C/Alk
4C/MSO
Nd-1
Wt-5
Bay-0
Bor-1
Uod-1
Ler-1
Lp2-2
Ct-1
Zdr-6
Zdr-1
Ull2-3
Bor-4
Lp2-6
Lz-0
Tsu-1
Ra-0
Wei-0
Var2-1
Cvi-0
Mr-0
Knox-18
Ei-2
:IBC-17
mx-A02
Uod-7
:mx-A180
Ws-0
Edi-0
Omo2-3
Ts-5
Ts-1
Ll-0
Se-0
Sq-8
Pu2-7
Tamm-27
Spr1-2
Lov-1
Eden-2
Bur-0
Bil-5
Bil-7
Eden-1
Lov-5
Tamm-2
Oy-0
Fab-4
Fab-2
Pu2-23
Ren-1
Pro-0
An-1
Fei-0
Pna-17
Yo-0
Knox-10
Pna-10
RRS-10
Ag-0
Sq-1
Gy-0
Ws-2
Mrk-0
Br-0
Mz-0
NFA-8
Mt-0
HR-5
Van-0
Kin-0
CIBC-5
HR-10
C24
NFA-10
Wa-1
Est-1
Nok-3
Col-0
Gu-0
Ga-0
CS22491
Var2-6
Kas-2
Kz-9
Spr1-6
Got-22
Got-7
Ms-0
Ull2-5
Ren-11
RRS-7
Kz-1
Shahdara
Kondara
Omo2-1
Sorbo
1,371,807
1,329,450


AOP3
complete linkage, binary distance
0.0
0.2
0.4
0.6
0.8
1.0
3C/Alk
3C/MSO
3C/OHP
4C/Alk
4C/MSO
Br-0
Bay-0
Ms-0
Var2-1
Sq-1
Ren-1
RRS-7
Ws-0
Ts-1
Ts-5
Tamm-27
Sq-8
Spr1-6
Spr1-2
Se-0
mx-A180
Rmx-A02
Pu2-7
Omo2-3
NFA-10
Mr-0
Lov-5
Lov-1
Ll-0
Knox-18
Got-7
Got-22
Ga-0
Ei-2
Edi-0
Eden-2
Eden-1
Cvi-0
CIBC-17
Bur-0
Bil-5
Bil-7
Var2-6
Uod-7
Tamm-2
Pu2-23
Oy-0
Nok-3
CS22491
Fab-2
Zdr-6
Zdr-1
Wt-5
Wei-0
Uod-1
Ull2-3
Tsu-1
Ra-0
Nd-1
Lz-0
Lp2-6
Lp2-2
Ler-1
Ct-1
Bor-1
Bor-4
Yo-0
Wa-1
Ull2-5
Sorbo
Shahdara
RRS-10
Ren-11
Pro-0
Pna-17
Pna-10
Omo2-1
Mrk-0
Kz-1
Kz-9
Kondara
Knox-10
Kin-0
Kas-2
Fei-0
Fab-4
Est-1
Col-0
C24
Ag-0
An-1
Ws-2
Van-0
NFA-8
Mz-0
Mt-0
HR-5
HR-10
Gy-0
CIBC-5
Gu-0
1,345,450
1,344,258
1,345,954
1,345,919
1,344,365
1,345,545


AOP3
complete linkage, binary distance
0.0
0.2
0.4
0.6
0.8
◆ 3C/Alk
◆ 3C/MSO
◆ 3C/OHP
◆ 4C/Alk
◆ 4C/MSO
Nd-1
Wt-5
Wei-0
Tsu-1
Lz-0
Ra-0
Ull2-3
Bay-0
Lp2-2
Ct-1
Zdr-6
Bor-1
Uod-1
Ler-1
Bor-4
Lp2-6
Zdr-1
C24
NFA-10
Cvi-0
Mr-0
Knox-18
Rmx-A02
CIBC-17
Rmx-A180
Ll-0
Se-0
Ts-1
Ws-0
Ei-2
Uod-7
Pu2-7
Eden-2
Lov-1
Eden-1
Lov-5
Oy-0
Fab-4
Pu2-23
Edi-0
Omo2-3
Ts-5
Tamm-27
Bur-0
Bil-5
Bil-7
Spr1-2
Fab-2
Tamm-2
Sq-8
Var2-1
Fei-0
Ren-1
An-1
Pro-0
Est-1
Nok-3
Col-0
Gu-0
Pna-17
Yo-0
Knox-10
Pna-10
RRS-10
Br-0
Ws-2
NFA-8
HR-5
Mt-0
Mz-0
Ag-0
Kin-0
Sq-1
Gy-0
HR-10
CIBC-5
Mrk-0
Van-0
Wa-1
Kondara
Sorbo
Omo2-1
Shahdara
Ms-0
Got-22
Got-7
RRS-7
Ren-11
Ull2-5
Kz-1
Kas-2
Kz-9
Ga-0
Spr1-6
CS22491
Var2-6
1,387,741
1,318,484

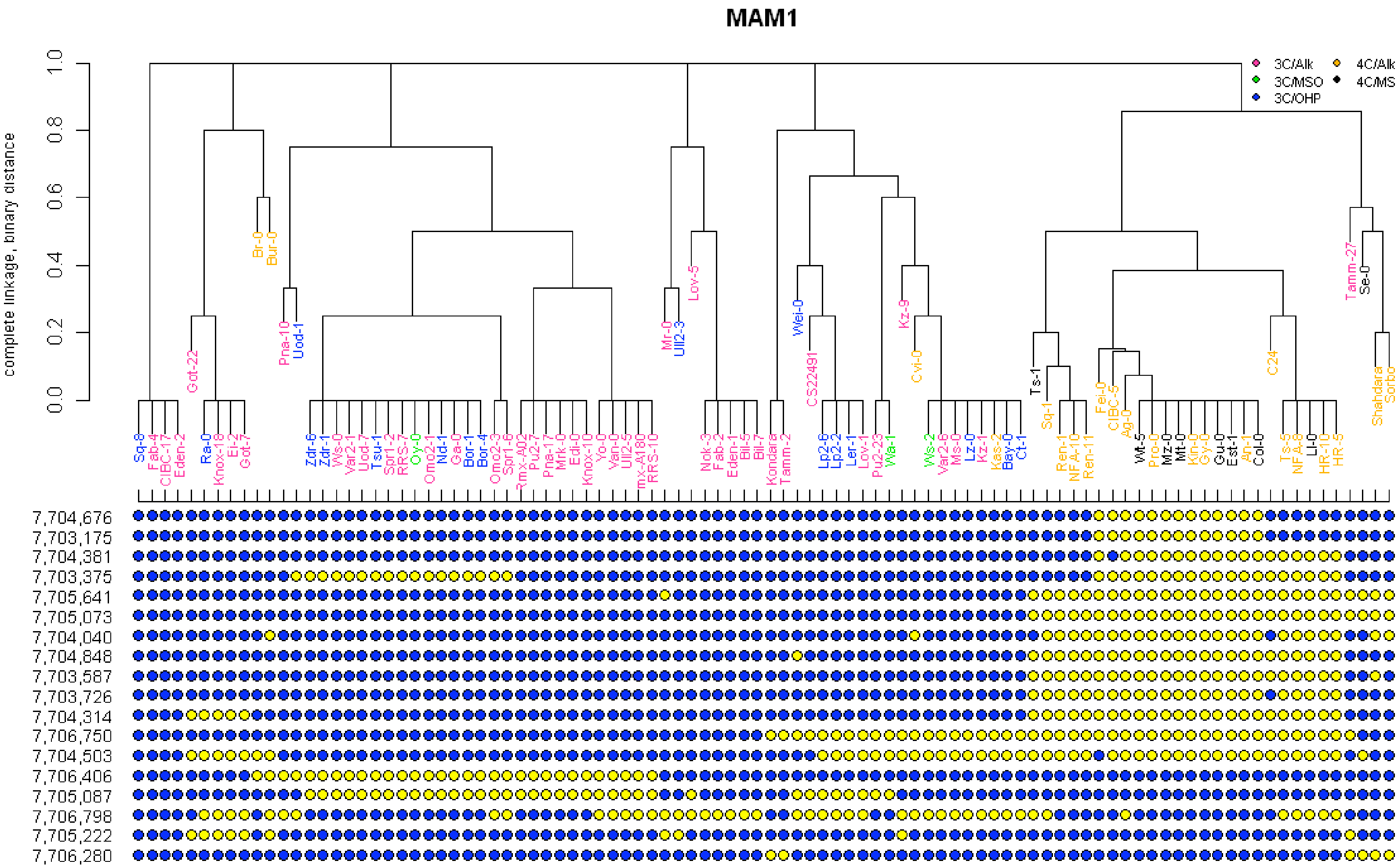


FIGURE S4.—LD Structure at and around the *AOP*,*MAM* and sweep loci. Shown are cladograms of 96 accessions using SNPs within the *AOP2, AOP3, MAM1,* and *MAM* genes, as well as the two sweep regions. The main GSL structure within each accession is indicated by color of the accession name. Each gene is shown twice, the first figure displays the allelic values of the SNPs within the specific gene itself; blue represents the majority allele and yellow the minority allele. The second figure displays the allelic values of the SNPs that are encompassed by the region of extended LD surrounding the corresponding gene as defined using the 99th (or 98th or *MAM3*) percentile boundaries from Supplemental Figure 3.

TABLE S1

ANOVA of accession and population structure for GSL 2007

Table S1 is available for download as an Excel file at http://www.genetics.org/cgi/content/full/genetics.109.108522/DC1.

For each GSL trait the following model was examined: $y_{sar} = \beta_0 + S_s + A(S)_{sa} + R_r + \varepsilon_{sar}$ where s = 1,…, 8 (NORDBORG *et al.* 2002; NORDBORG *et al.* 2005); r = 1,…4; and a = 1,…,95. The main effects are denoted as S, A and R to represent structure, accession, and replicate block respectively, and error, $\varepsilon_{sar} \sim N(0, \sigma_\varepsilon^2)$. F: F-statistic of the model, df_{num}: numerator degrees of freedom, df_{denom}: denominator d.f., P(F): nominal P-value of the F-statistic, R^2: fraction of total variance explained by the model; $\bar{x}$: observed mean, β_0: model mean, η^2: partial R^2 of the corresponding predictor variable, and P: P-value of the corresponding predictor variable. Description shows the chemical structure of the glucosinolate or the algebraic approximation of the trait.

TABLE S2

ANOVA of accession and population structure for GSL 2008

Table S2 is available for download as an Excel file at http://www.genetics.org/cgi/content/full/genetics.109.108522/DC1.

For each GSL trait the following model was examined: $y_{sar} = \beta_0 + S_s + A(S)_{sa} + R_r + \varepsilon_{sar}$ where s = 1,…, 8 (NORDBORG *et al.* 2002; NORDBORG *et al.* 2005); r = 1,…4; and a = 1,…,95. The main effects are denoted as S, A and R to represent structure, accession, and replicate block respectively, and error, $\varepsilon_{sar} \sim N(0, \sigma_\varepsilon^2)$. F: F-statistic of the model, df_{num}: numerator degrees of freedom, df_{denom}: denominator d.f., P(F): nominal P-value of the F-statistic, R^2: fraction of total variance explained by the model; $\bar{x}$: observed mean, β_0: model mean, η^2: partial R^2 of the corresponding predictor variable, and P: P-value of the corresponding predictor variable. Description shows the chemical structure of the glucosinolate or the algebraic approximation of the trait.

TABLE S3

Average glucosinolate values per accession for both experiments

Table S3 is available for download as an Excel file at http://www.genetics.org/cgi/content/full/genetics.109.108522/DC1.

All metabolite values are in nmol per mg fresh weight tissue. This is the average of four independent plants per accession.

TABLE S4

Correlation of glucosinolate phenotypes

Table S4 is available for download as an Excel file at http://www.genetics.org/cgi/content/full/genetics.109.108522/DC1.

Shown are the significant spearman rank correlations for all glucosinolate phenotypes used in this study.

TABLE S5

Gene to Trait Associations as identified from GWA

Table S5 is available for download as an Excel file at http://www.genetics.org/cgi/content/full/genetics.109.108522/DC1.

Logical table indicating whether each of 31,505 genes is significantly associated to each of the 43 traits. For each trait, a gene is significantly associated if at least two SNPs within ±1kb flanking the coding region has a P-value in the bottom 0.1 percentile of the P-value distribution. T: is significant in either 2007 or 2008 set; F: not significant in both sets; -: not significant in either 2007 or 2008 set and untested in the reverse set.

TABLE S6

Known and putative genes involved in the GSL biosynthetic pathway

Table S6 is available for download as an Excel file at http://www.genetics.org/cgi/content/full/genetics.109.108522/DC1.

List of genes either known or predicted to play a role in GSL metabolism and regulation. AGI: shows the AGI (Arabidopsis Genome Initiative) code for each gene; Pathway: specific part of the GSL metabolic system the gene is thought to function, this includes the core biosynthetic pathways as well as pathways required for the production of immediate precursors such as methionine, tryptophan, etc.; Pseudogene: whether or not the gene is predicted to be a pseudogene; Evidence: experimental evidence (Genetic or Biochemical) or sequence evidence base on homology to validated GSL gene (Homology).

TABLE S7

GSL genes as a group not preferentially identified in 2007 or 2008

Table S7 is available for download as an Excel file at http://www.genetics.org/cgi/content/full/genetics.109.108522/DC1.

The Kolmogorov-Smirnov D-statistics and corresponding P-values are shown for each trait and for each of 2007 & 2008 datasets from the GWA results. Each KS-test assessed the null hypothesis that: the GWA P-value distribution corresponding to SNPs within ±1kb flanking region of known GSL genes is identical to that of non-GSL SNPs.

TABLE S8

Using known GSL genes to estimate call thresholds

	Dataset	≥1 SNP per gene	≥2 SNP per gene	≥20% SNP per gene
# of Sig Genes	2007	5333	1056	1135
	2008	4703	893	960
GSL Genes in Sig Genes	2007	0.7%	1.2%	0.8%
	2008	0.5%	1.5%	0.5%
Sig GSL Genes	2007	25.3%	8.7%	6.0%
	2008	16.7%	8.7%	3.3%

The number of genes significantly associated with at least one trait in the two datasets (2007 and 2008) using three different call thresholds are presented. The percentage of significant genes overlapping with the list of known and putative GSL genes is provided as well as the percentage of all known GSL genes exceeding each threshold.

FILE S1

GWA Results

File S1 is available for download as a compressed (.zip) file at http://www.genetics.org/cgi/content/full/genetics.109.108522/DC1.

This dataset contains three files. One is the SNP information (snp.info.txt) which includes the chromosome and genomic position of each SNPs, as well as the minor allele frequencies in this dataset. The two other are tables of P-values of association tests between each GSL trait and SNPs; one file for each experiment (2007 & 2008).

FILE S2

Supporting Method: Modifications to R/EMMA

This file details the changes made to three functions of the R/EMMA package, version 1.0.7 (KANG *et al.* 2008).

emma.eigen.L.w.Z()

changed:

```
      eig <- eigen(K%*%crossprod(Z,Z),symmetric=FALSE,EISPACK=TRUE)
```

to:

```
      eig <- eigen(K%*%crossprod(Z,Z),symmetric=F,EISPACK=F)
```

emma.eigen.R.w.Z()

changed:

```
      eig <- eigen(K%*%crossprod(Z,SZ),symmetric=FALSE,EISPACK=TRUE)
```

to:

```
      eig <- eigen(K%*%crossprod(Z,SZ),symmetric=F,EISPACK=F)
```

emma.ML.LRT()

changed:

```
    for(i in 1:g) {
      vrows <- !is.na(ys[i,])
      if ( is.null(Z) ) {
        ML0[i] <-
emma.MLE(ys[i,vrows],X0[vrows,,drop=FALSE],K[vrows,vrows],NULL,ngrids,l
lim,ulim,esp)$ML
      }
      else {
        vids <- colSums(Z[vrows,]>0)
        ML0[i] <-
emma.MLE(ys[i,vrows],X0[vrows,,drop=FALSE],K[vids,vids],Z[vrows,vids],n
grids,llim,ulim,esp)$ML
      }
    }
```

to:

```
    for(i in 1:g) {     # for each trait
       vrows <- !is.na(ys[i,])
       if( var(ys[i,vrows],na.rm=T)>0 ) {
          if ( is.null(Z) ) {
          ML0[i] <-
emma.MLE(ys[i,vrows],X0[vrows,,drop=FALSE],K[vrows,vrows],NULL,ngrids,l
lim,ulim,esp)$ML
          }
          else {
            #vids <- colSums(Z[vrows,]>0)
            vids <- as.logical(colSums(Z[vrows,]>0))
            ML0[i] <-
emma.MLE(ys[i,vrows],X0[vrows,,drop=FALSE],K[vids,vids],Z[vrows,vids],n
grids,llim,ulim,esp)$ML
          }
       }
    }
```

File S3

Genotypes of 248,584 SNP for the 96 *A. thaliana* accessions

File S3 is available for download as a compressed (.zip) file at http://www.genetics.org/cgi/content/full/genetics.109.108522/DC1.

Genotypes of 248,584 SNP for the 96 A. thaliana accessions were obtained using the SNP chip described by Kim et al. (2007) and were generated by the groups of J. Bergelson, J. O. Borevitz, and M. Nordborg.