TRIMM (TRIad Multi-Marker test)

Min Shi, David M. Umbach, Clarice R. Weinberg

DESCRIPTION

This program performs max_Z^2, and sum_log(P) tests, association tests for child's or mother's genetic effects using multiple markers from triad families. Statistical significance is evaluated via permutation. It also outputs the nominated risk-haplotype-tagging alleles. For details, see *Shi M, Umbach DM, Weinberg CR "Identification of Risk-related Haplotypes Using Multiple SNPs from Nuclear Families*

USAGE

```
trimm(m.file="geno.m",f.file="geno.f",c.file="geno.c", sep =
"", na.strings = 0,with.header=F, use.missing = F, effect =
'C', min.info=10, permu.num=1000, cutoff.snp=0.1,
cutoff.global=0.1, cutoff.test='maxZ')
```

Note: This program has been developed and tested under R version 2.2.1 and may or may not be compatible with other versions of R.

ARGUMENTS

m.`file`	A character string giving the name and path of the ASCII data file containing mothers' genotype. The default file name is "geno.m". Currently, this program only handles triad data (mother, father and a child). For details about input data format see "**Genotype Input File Format**" section
f.`file`	A character string giving the name and path of the ASCII data file containing fathers' genotype. The default file name is "geno.f". Currently, this program only handles triad data (mother, father and a child). For details about input data format see "**Genotype Input File Format**" section
c.`file`	A character string giving the name and path of the ASCII data file containing children's genotype. The default file name is "geno.c". Currently, this program only handles triad data (mother, father and a child). For details about input data format see "**Genotype Input File Format**" section
`sep`	The character separator used in the genotype input file to separate "columns", where each column contains one allele of a single individual at a single marker. The default separator is "white space", *i.e.*, one or more spaces, tabs, newlines or carriage returns.
`na.strings`	The character string indicating missing data in the genotype input file. In agreement with linkage file format the default is 0

`with.header` A logical value used to determine whether there is header row in the input file. The default is FALSE (no header)

`use.missing` A logical value used to determine whether triads with missing parent should be included in the analysis. This option is only valid in testing child's genetic effect. When set to TRUE, the genotype difference vector is calculated based on genotype difference between the child and the available parent. **WARNING: When set to TRUE, the test is only valid when there is no maternal effect! We therefore recommend performing a test on maternal effect first and set the option to TRUE only when there is no maternal effect**. Currently we substitute "0" in the difference vector for random missing genotypes

`effect` The character string indicating which effects are being tested. Only two options are allowed "C" (for child effect) and "M" (for maternal effect). The default is "C"

`min.info` Sets the lower limit for the number of informative families that is required before a Z-score will be calculated for the corresponding SNP. The default is 10

`permu.num` The number of permutations that will be carried out in evaluating the p-value of the test. The default is 1000

`cutoff.snp` The cutoff p-value for nominating risk-haplotype-tagging SNPs, i.e. individual SNPs with a p-value smaller than the cutoff are nominated as the risk-haplotype-tagging SNPs. The default value is 0.1

`cutoff.global` The program will print out nominated risk-haplotype tagging SNPs only when the p-value of the global test of association (based on either max_Z^2 or sum_log(P) test, see `cutoff.test` for detail) is smaller than `cutoff.global`. The default value is 0.1

`cutoff.test` This parameter defines the test on which the global p-value cutoff (`cutoff.global`) is based. Only two input values are allowed: '**maxZ'** (for max_Z^2 test) and '**sumlogP'** (for sum_log(p) test). **Note: the input is case sensitive.** The default test is 'maxZ'

REFERENCES

Shi M, Umbach DM, Weinberg CR 2007 "Identification of Risk-related Haplotypes Using Multiple SNPs from Nuclear Families"

Genotype Input File Format

NOTE: Currently this program can only handle triad data.

Three input files are required: genotypes for mothers, fathers, and children. The file format is the same for these three files. The columns are in the order of: Pedigree_ID, allele 1 of SNP 1, allele 2 of SNP 1, allele 1 of SNP 2…

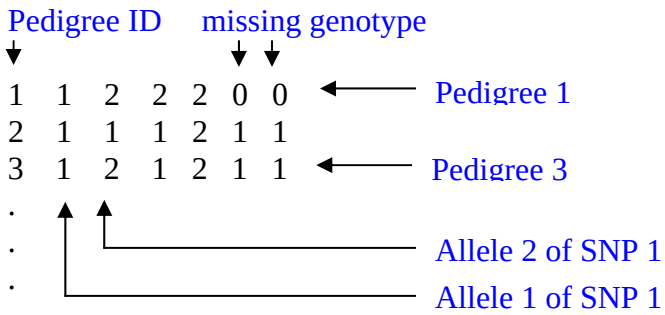
The genotype of the alleles should be coded as "1" and "2". It is very important to make sure the input file is in the correct format. Currently, the program does not check for format/coding errors in the input file.

When there is a header row in the input file, set the option **header** to TRUE.

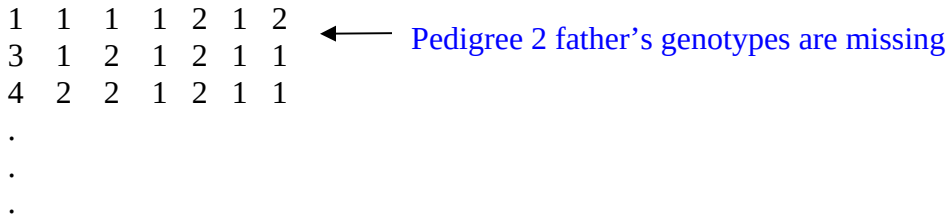
The input file allows for missing individuals in a pedigree. For example, in the example given below the father of pedigree 2 is missing.

Example:

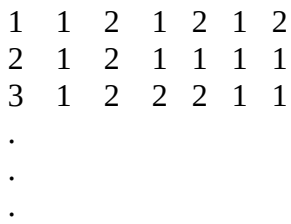
Mothers' genotype file:



Fathers' genotype file:



Children's genotype file:



Test files of input genotypes are given ('geno.c', 'geno.f', 'geno.c')

EXAMPLES

Calling the function:

All using default input

```
> trimm()
```

Specify path, estimate maternal effects:

```
>trimm("C:/genetics/geno.m","C:/genetics/geno.f","C:/genetics
/geno.c",effect='M')
```

Specify path, test for child effects, use families with missing parent, set minimum number of informative families to 10, number of permutations to 10000, the cutoff p-value for nominating risk-haplotype-tagging SNPs is 0.1:

```
>trimm(f.file='geno.f',m.file='geno.m',c.file='geno.c',effect
='C', use.missing=T, min.info=10, permu.num=10000,
cutoff.snp=0.1, cutoff.global=0.1, cutoff.test='sumlogP')
```

Example of output

Warning message

```
##############################################################################################
# WARNING: In testing child effect,the tests are valid only when there is no maternal effect.  #
#         Set use.missing to 'F' if the assumption does not hold.                              #
##############################################################################################
**********************************************************************************************

                              TESTING FOR:  Fetal Genetic Effect
                            ##############################################

 Testing conditions:

                    Using triad with missing genotype.
                    Number of permutation =  10000
                    Minimum number of informative families required =  10
                    If p-value of the global test based on  maxZ  test is less than  0.1 ,
                    print tagging alleles with individual SNP-wise p-value less than  0.1


 Global       test:

                      Max_Z^2              value:  9.657786  (max_Z= 3.107698 ) at SNP  1
                      Hotelling's T^2      value:  22.27385
                      p-value for max Z2    test:  0.0255
                      p-value for Sum_logP  test:  0.029
```

Tested effects

Testing conditions

Allele nomination conditions

Global test result

Individual SNP scores:

SNP	:	SNP_1	SNP_2	SNP_3	SNP_4	SNP_5	SNP_6	SNP_7	SNP_8	SNP_9	SNP_10
Z score	:	3.108	1.913	-1.434	0.476	0.505	2.359	1.461	-0.829	-1.992	-1.507
p-value	:	0.0034	0.0714	0.1805	0.7040	0.6757	0.0235	0.1678	0.4868	0.0604	0.1583
Tagging allele	:	2	2				2			1	

Alleles at the risk-haplotype-tagging SNPs

SNP	:	SNP_11	SNP_12	SNP_13	← SNP
Z score	:	-0.905	0.363	-1.400	← Z score at individual SNPs
p-value	:	0.43	0.79	0.19	← P values for individual SNPs
Tagging allele	:				

****

Explanation: The above output tells us: The p-value for the global test that there is no linkage or no association between the disease and the markers under study is 0.0255 based on max_Z^2 test and 0.029 based on sum_log(P) test. Using a SNP-wise cutoff of 0.1 we can nominate a haplotype composed of SNP 1, SNP 2, SNP 6, and SNP 9 with alleles (2 2 2 1)