

# Document for QSAT software

Shuanglin Zhang & Hongyu Zhao

Department of Epidemiology and public Health  
Yale University School of Medicine  
60 College street  
New Haven, CT 06520, USA

## INTRODUCTION

The program QSAT implements an association test using population based sampling and controlling the population stratification by genomic marker genotypic data (a series of unlinked marker genotypes). Detail see Zhang & Zhao (2001). This version of QSAT is only for biallelic unlinked markers and candidate loci.

We strongly encourage users to send us a brief email at either [shuanglin.zhang@yale.edu](mailto:shuanglin.zhang@yale.edu) or [hongyu.zhao@yale.edu](mailto:hongyu.zhao@yale.edu) , telling us that you have downloaded the program and what sort of the questions the program is applied to. This will help us to assess the amount of interest in QSAT, and how much effort we should put into maintaining and improving it.

## GETTING STARTED

You will need an executable file: QSAT.exe, one file of parameters : *parameter.txt* and two input data files (one file for unlinked marker genotypic data and one for phenotypes and genotypes of candidate markers). The name of the input data files is given in the file *parameter.txt*. We strongly suggest that you put the executable file, parameters file and data files into one folder. To analyze your own data, you need to prepare the data file in the appropriate format.

## FORMAT FOR INPUT DATA FILE

In the parameter file *parameter.txt* (see example file), there are description lines and number lines. For each term, only has one description line, the number following the description line can take one line or more than one line. For details, see the sample file *parameter.txt*.

The format of the data files is as follows. The input data file for unlinked marker genotypes (the file name is given in the *parameter.txt*), *unlink.dat* as an example, has a title line (mark1

mark2,..) and each subsequent line is for an individual. For each individual, each marker take two lines for the two alleles. Following is an example for 3 individuals with 5 markers.

```
mark1 mark2 mark3 mark4 mark5
0 0    1 1    0 1    0 0    1 1
0 0    1 1    0 0    0 0    1 1
0 0    1 0    1 0    0 0    1 1
```

The format of input data set for phenotypes and candidate marker data is almost the same as that of the file for unlinked markers except that there is an additional column (first column) of phenotype. Following is an example for 3 individuals with five markers.

```
mark1 mark2 mark3 mark4 mark5
0.889 0 0    1 1    0 1    0 0    1 1
1.335 0 0    1 1    0 0    0 0    1 1
0.736 0 0    1 0    1 0    0 0    1 1
```

## **FORMAT FOR OUT DATA FILE**

There is one out put file (the name was given in the file of *parameter.txt*). The results in output file include (1) exists population structure or not and the BIC values for one group of similarities (no population structure or a homologous population) and two group of similarities (exists population structure); (2) the permutation based p-value for every markers. Example of the output is given bellow.

=====  
There exists population structure.

The BIC corresponding to one group and two groups of the similarities are  
BIC(1)=-246.374823 BIC(2)=-436.241605.

marker#	P-value
14	0.08600
5	0.12200
2	0.18400
16	0.19800
9	0.24800
7	0.29400
19	0.32300
17	0.46700
18	0.50900
12	0.61000
20	0.67100
11	0.71600
6	0.72800
13	0.76800
8	0.87300
10	0.87700
1	0.88000
3	0.93100
15	0.94900
4	0.97100