

Getting started with haplotype FBAT

Steve Horvath, Xin Xu, Stephen Lake, Nan Laird

The haplotype FBAT software extends the single marker FBAT software. FBAT version 1.4 and higher contains the command `hbat` that carries out the haplotype analysis. The input files and most commands are identical to those of (single marker) FBAT. Please consult the online FBAT manual for details. The FBAT software program was implemented by Xin Xu. Please contact fbat@hsph.harvard.edu for suggestions.

Background

The haplotype software package provides a general purpose family-based testing strategy for allelic association between disease phenotypes and haplotypes when phase may be ambiguous and parental genotype data may be missing. These tests can be used in candidate gene studies with tightly linked markers. The method can be used to test for linkage between the haplotype locus and any trait influencing gene. We also describe how to test for association between haplotypes and any trait influencing gene in the presence of linkage. Our proposed weighted conditional approach extends the method described in Rabinowitz and Laird (2000) and underlying the FBAT software (Laird et al 2001; Horvath et al 2002) to multiple markers. It is attractive because it provides haplotype tests for family-based studies which are efficient and robust to population admixture, phenotype distribution specification and ascertainment based on phenotypes. It can handle missing parental genotypes and/or missing phase in both offspring and parents and multiallelic or biallelic tests.

Loading the Pedigree File and Running Haplotype FBAT

To get started, the user must call the program and load a pedigree data file. This file follows the standard linkage file format as used by genehunter or mapmaker, i.e. the columns encode

```
pedigreeID, individualID, fatherID, motherID, gender, affectionation status, first  
allele of marker 1, second allele of marker 1, first allele of marker 2, etc.
```

```
Missing values are 0  
Gender: 2 indicates female, 1 male,  
Affectionation status: 2 indicates affecteds while 1 indicates unaffecteds
```

Different from the linkage format one needs to specify the marker names in the first line! See the description of the pedigree file format for details. For example, the first few lines of a data file named `Pedigree1.txt` are:

```
Marker1 Marker2 Marker3  
1 1 0 0 1 0 0 0 3 3 3 4  
1 2 0 0 2 2 1 2 2 5 8 6  
1 3 1 2 2 2 1 1 3 5 8 4  
1 4 1 2 1 2 1 2 2 3 4 6  
3 1 0 0 1 0 1 2 1 3 3 6  
3 2 0 0 2 0 1 2 1 3 6 6  
3 3 1 2 1 2 1 2 1 3 6 6  
3 4 1 2 2 2 1 1 3 3 3 6  
many more lines...
```

Note that the method assumes that the markers are tightly linked, i.e., no recombination occurs between Marker1, Marker2, and Marker3.

Start the FBAT software.

To see a list of commands type
>> ?

To learn about the min command type
>> ? min

To start a log file type
>> log Log1.txt

To load pedigree data type

>> load ped Pedigree1.txt
read in: 3 markers from 60 pedigrees (60 nuclear families,249 persons)

To limit the output to statistics that have a p-value smaller than .1 type
>> displayp .1
current p_value is 0.100000

To test for linkage between the haplotype locus and a disease pre-disposing gene, please type

>> hbat
trait affection; model additive; test bi-allelic; minsize 10; p 0.100000

haplotype analysis for the following markers:
Marker1 Marker2 Marker3

haplotypes and EM estimates of frequency:

```
h1 : 1 3 3 0.166079
h2 : 2 3 3 0.125305
h3 : 1 3 6 0.068240
h4 : 2 5 3 0.065217
h5 : 2 1 6 0.058775 Message: estimates of haplotype frequencies
h6 : 2 3 6 0.046898
h7 : 1 5 6 0.043478
h8 : 1 3 4 0.038043
h9 : 2 3 4 0.027174
h10 : 1 5 3 0.027174
h11 : 1 1 4 0.021739
etc etc
```

Allele	afreq	fam#	S	E(S)	Var(S)	Z	P
h2	0.125	17	22.790	17.790	8.812	1.684	0.092121
h3	0.068	11	6.000	12.000	6.000	-2.449	0.014306
h4	0.065	12	17.937	12.437	6.189	2.211	0.027047

Message: the univariate test statistic that counts how often haplotype h3 is present in affecteds leads to a p.value of .014. The negative sign in the Z statistic indicates that haplotype h3 is less frequent in affecteds than expected under the null hypothesis of no linkage. These asymptotic p-values are *not* corrected for multiple comparisons, i.e. no Bonferroni corrections.

To focus on haplotypes formed by markers Marker1 and Marker3, please type

```
>> hbat Marker1 Marker3
trait affection; model additive; test bi-allelic; minsize 10; p 0.100000
```

```
haplotype analysis for the following markers:
Marker1 Marker3
```

```
haplotypes and EM estimates of frequency:
```

```
h1 : 1 3 0.220858
h2 : 2 3 0.203670
h3 : 1 6 0.148899
h4 : 2 6 0.143553
h5 : 1 4 0.072511
h6 : 2 4 0.054847
etc
```

Allele	afreq	fam#	S	E(S)	Var(S)	Z	P
h2	0.204	26	46.500	37.750	17.021	2.121	0.033931
h3	0.149	24	20.500	31.750	15.271	-2.879	0.003991
h4	0.144	21	17.500	24.917	11.726	-2.166	0.030323
h6	0.055	10	15.000	10.100	4.990	2.194	0.028268

Assume that it is already known that the locus is linked. Then it may be of interest to test for association in the presence of linkage. In this case, we recommend to use an empirical variance estimate in the test statistic, which can be invoked by typing

```
>> hbat -e Marker1 Marker3
trait affection; model additive; test bi-allelic; minsize 10; p 0.100000
```

```
haplotype analysis for the following markers:
Marker1 Marker3
```

```
haplotypes and EM estimates of frequency:
```

```
h1 : 1 3 0.220858
h2 : 2 3 0.203670
h3 : 1 6 0.148899
h4 : 2 6 0.143553
h5 : 1 4 0.072511
h6 : 2 4 0.054847
etc
```

Allele	afreq	fam#	S	E(S)	Var(S)	Z	P
h2	0.204	18	35.500	26.750	24.563	1.766	0.077477
h3	0.149	15	11.500	22.750	19.812	-2.527	0.011489
h4	0.144	16	12.500	19.917	11.924	-2.148	0.031725

Message: note that the p-values have become less significant.

New Task: Analyzing Quantitative Traits

Assume that you are dealing with a quantitative phenotype (trait), e.g. blood pressure. In this case you need to read in a phenotype file,

which has the same format as described in the FBAT manual.

Briefly the columns are pedigree ID, individual ID, trait1, trait2, etc.

Importantly, you need to specify the trait names in the first row.

Example file Pheno1.txt

```
TraitA TraitB
1 1 0.52 1
1 2 0.16 2
1 3 0.45 2
1 4 0.72 1
3 1 0.85 -      Comment: missing traits are coded by "-"
3 2 0.76 1
3 3 0.22 2
3 4 0.2 2
etc
```

To load the phenotype file, please type

```
>> load phe Pheno1.txt
2 quantitative traits have been successfully read
249 persons have been phenotyped
```

To see all the available traits type

```
>> trait
affection** TraitA TraitB
```

Message: the first (default) trait is the affection status trait specified in the pedigree file. When using a phenotype file this trait can be ignored. But for syntax reasons one still needs to provide an affection status column in the pedigree file, e.g., just fill it up with ones.

Let us now specify that TraitA is to be used in subsequent analyses.

```
>> trait TraitA
affection TraitA** TraitB
```

To test for linkage between the haplotype locus and a trait influencing gene, please type

```
>> hbat
trait TraitA; model additive; test bi-allelic; minsize 10; p 0.100000
etc
```

haplotypes and EM estimates of frequency:

```
h1  :    1    3    3    0.166079
h2  :    2    3    3    0.125305
etc
```

Allele	afreq	fam#	S	E(S)	Var(S)	Z	P
h3	0.068	11	1.890	5.790	1.737	-2.959	0.003082
h4	0.065	12	10.035	6.349	2.139	2.521	0.011713

For quantitative traits it may be useful to type

```
>> hbat -o
```

Roughly speaking the option `-o` centers each trait, i.e., it subtracts a trait “medoid” from each trait. More specifically it replaces each trait by $(\text{current trait} - \mu)$ in the test construction. The value of μ is chosen to minimize the variance of the test statistic. The `-o` option can only be used with the bi-allelic mode.

To use a multi-allelic marker coding, please type

```
>> mode m
```

To use a multi-allelic marker coding, please type

```
>> mode m
>> hbat
....
```

To use the single marker FBAT, please type

```
>> fbat
```

To turn off the log file

```
>> log off
```

For the experts:

To view the haplotype configuration of selected markers in each family type

```
>> viewhap Marker1 Marker3
```

```
>> quit
```

References

Main reference for haplotype FBAT:

Horvath S, Xu X, Lake SL, Silverman EK, Weiss ST, Laird NM (2004) Family based tests for associating haplotypes with general phenotype data: application to asthma genetics. *Genet Epidemiol*, Vol 26, No 1, 61-69

Single locus FBAT:

Horvath S, Xu X, Laird NM (2001) The family based association test method: strategies for studying general genotype-phenotype associations. *Eur J Hum Genet* 9:301-306

Laird NM, Horvath S, Xu X (2000) Implementing a unified approach to family based tests of association. *Genet Epidemiol Suppl* 19:S36-S42

Rabinowitz D, Laird NM (2000) A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. *Hum Hered* 50:227-233