

A SAS Macro "HAPEM" - a Quick Reference Sheet

- by JINGXIA LIU -

I. Macro Interpretation

This is a SAS Macro for computing haplotype frequencies at multiple linked biallelic marker (e.g., SNP) loci from their observed unphased genotype data. The EM algorithm proposed by Excoffier and Slatkin (1995) is implemented based on the assumption of Hardy-Weinberg Equilibria at these marker loci.

be less than 20.

- (3) nallele: The number of alleles at each locus. It may be a vector or a number if all loci have the same number of alleles. This macro currently works only for nallele=2 (i.e., biallelic markers)

These three arguments must be given.

- (4) The symbol of angle brackets identifies optional arguments. There are at most 7 options available and they can be given in any order after the first three. Each option is specified as a keyword followed by an equal sign and a value (not case-

1) True value setting:

```
Hap=[0 0 0 0 0 0;  
      0 0 1 0 1 0;  
      0 0 0 1 0 0;  
      0 1 1 1 0 1;  
      1 0 0 0 1 0;  
      1 1 1 0 0 1;  
      1 0 0 1 0 0;  
      1 1 1 1 1 0];
```

```
hapfreq=[0.25 0.20 0.20 0.10 0.10 0.05 0.05 0.05];
```

2) Macro Result:

```
data test2;  
input GEN01-GEN06;  
cards;
```

1	1	2	1	2	0
1	1	2	1	2	0
1	0	1	0	2	0
1	0	1	0	2	0
0	0	1	1	1	0
0	0	1	1	1	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	1	1	1	0
1	1	1	1	0	1
1	0	0	1	0	0
0	0	2	0	2	0
1	1	1	0	0	1
2	2	2	1	1	1
0	0	1	0	1	0
0	0	0	2	0	0
1	2	2	2	1	1
0	1	1	2	0	1
2	1	1	0	1	1
0	0	0	1	0	0
0	0	1	0	1	0
2	1	1	1	2	0
0	0	1	0	1	0
1	1	1	0	0	1
1	1	1	1	1	1
0	1	1	1	0	1
0	1	2	1	1	1
0	0	0	1	0	0
0	1	1	1	0	1
0	0	0	0	0	0
0	0	2	0	2	0
1	1	1	0	0	1
0	0	1	1	1	0
1	0	1	0	2	0
0	0	0	1	0	0
0	0	1	1	1	0
0	0	0	1	0	0
0	0	1	0	1	0
2	1	1	1	0	1

0	0	0	1	0	0
0	0	1	0	1	0
0	1	1	1	0	1
0	1	1	2	0	1
0	1	1	1	0	1
1	1	1	1	1	0
0	1	1	2	0	1
1	0	1	0	2	0
0	1	1	1	0	1
0	0	0	2	0	0
0	1	2	1	1	1
0	0	2	0	2	0
0	0	0	1	0	0
1	0	1	0	2	0
1	0	0	0	1	0
2	0	0	1	1	0
1	0	0	1	0	0
1	1	1	1	1	1
0	0	1	0	1	0
0	0	0	2	0	0
1	1	1	1	1	1
1	1	1	1	1	0
1	0	0	2	0	0
0	0	1	1	1	0
0	0	0	2	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	1	1	2	0	1
2	1	1	1	0	1
0	0	2	0	2	0
1	2	2	2	1	1
0	0	1	0	1	0
0	1	1	2	0	1
0	2	2	2	0	2
0	0	1	0	1	0
0	0	1	0	1	0
0	1	2	1	1	1
1	0	0	1	1	0
0	0	2	0	2	0
0	0	2	0	2	0
1	0	0	1	1	0
1	0	0	1	1	0
0	1	1	1	0	1
0	0	0	2	0	0
1	1	1	1	1	1
0	0	0	2	0	0
0	0	0	0	0	0
0	1	1	2	0	1
1	2	2	2	1	1
1	0	1	0	2	0
1	2	2	1	0	2
0	1	1	2	0	1
0	0	0	0	0	0
0	1	1	1	0	1
0	0	1	0	1	0
1	0	0	1	0	0
0	1	1	2	0	1
2	1	1	0	1	1

0	1	1	2	0	1
1	1	2	1	2	0
1	1	1	2	1	0
1	0	0	1	0	0
0	0	2	0	2	0
0	0	0	1	0	0
0	0	1	0	1	0
1	1	1	2	1	0
0	1	1	1	0	1
0	0	1	1	1	0
0	1	1	2	0	1
1	1	2	1	2	0
1	0	1	0	2	0
1	1	1	1	1	1
1	0	1	0	2	0
1	0	0	1	1	0
2	0	0	0	2	0
0	0	0	0	0	0
1	0	1	1	1	0
1	0	0	2	0	0
1	0	0	1	1	0
0	0	1	1	1	0
1	1	1	1	0	1
0	0	0	2	0	0
0	1	2	1	1	1
0	0	1	1	1	0
1	1	1	2	1	0
0	0	0	1	0	0
0	0	0	1	0	0
1	0	1	0	2	0
2	1	1	0	1	1
1	1	2	0	1	1
0	1	1	1	0	1
0	0	1	0	1	0
0	0	1	1	1	0
0	0	1	1	1	0
0	0	2	0	2	0
0	0	0	1	0	0
0	0	0	1	0	0
1	1	1	0	0	1
2	2	2	0	0	2
1	0	1	0	2	0
1	0	0	0	1	0
1	1	1	1	1	1
1	1	1	1	1	1
0	0	0	0	0	0
0	0	1	0	1	0
0	0	0	1	0	0
0	0	0	2	0	0
0	0	0	1	0	0
0	0	1	0	1	0
0	0	1	1	1	0
0	1	2	1	1	1
1	0	0	2	0	0
1	0	0	1	0	0
0	0	1	1	1	0
0	0	1	1	1	0

```

0      0      0      1      0      0
1      0      0      1      1      0
0      0      0      0      0      0
2      1      1      0      1      1
1      1      2      1      2      0
0      0      0      1      0      0
0      0      0      1      0      0
1      0      0      0      1      0
0      0      0      2      0      0
0      2      2      2      0      2
0      0      0      1      0      0
0      0      1      0      1      0
0      0      1      1      1      0
0      1      1      1      0      1
1      0      0      1      1      0
1      0      1      0      2      0
0      0      1      1      1      0
1      1      2      1      2      0
0      0      0      0      0      0
2      1      1      0      1      1
1      0      0      0      1      0
1      1      2      0      1      1
0      1      1      2      0      1
1      1      2      1      2      0
1      0      0      2      0      0
0      0      0      0      0      0
0      0      1      1      1      0
0      0      0      0      0      0
0      0      0      2      0      0
0      1      2      1      1      1
0      0      1      1      1      0
0      0      1      1      1      0
0      0      0      0      0      0
1      1      2      0      1      1
0      0      0      2      0      0
1      1      1      1      1      0
0      0      0      2      0      0
0      0      1      0      1      0
0      1      1      2      0      1
0      0      1      1      1      0
1      0      0      2      0      0
0      1      1      2      0      1
1      0      1      1      1      0
0      0      2      0      2      0
0      0      1      1      1      0
0      1      2      1      1      1
0      0      1      0      1      0
;
run;

%HAPEM(test2, nloci=6, nallele=2, haptop=8, outdata=out, itsumm=Y,
convrr=10##(-3))

proc print data=out;
title "Print Outdata Set";

```


The SAS System
2

The Convergent Criterion Used Here is 0.0010000

The Maximum Step Used Here is 500

True Iteration Step is 7

The length of the top haplotype frequency used here is 8

100 means the haplotypes not shown above have been collapsed

Marker Haplotype & Haplotype Frequency

0	0	0	0	0	0	0.2149424
0	0	0	1	0	0	0.2300870
0	0	1	0	1	0	0.2149993
0	1	1	1	0	1	0.1124639
1	0	0	0	1	0	0.0949594
1	0	0	1	0	0	0.0374369
1	1	1	0	0	1	0.0499910
1	1	1	1	1	0	0.0449995
100	100	100	100	100	100	0.0001206

Print Outdata Set

3

Obs	decimal	probmarkers	hap1	hap2	hap3	hap4	hap5	hap6
1	0	0.21494	0	0	0	0	0	0
2	4	0.23009	0	0	0	1	0	0
3	10	0.21500	0	0	1	0	1	0
4	29	0.11246	0	1	1	1	0	1
5	34	0.09496	1	0	0	0	1	0
6	36	0.03744	1	0	0	1	0	0
7	57	0.04999	1	1	1	0	0	1
8	62	0.04500	1	1	1	1	1	0
9	100	0.00012	100	100	100	100	100	100