

```

/*Supplemental material : SAS codes for building haplotype data using FASTCLUS
procedure, Sample files
are provided with the macro. The sample genotypic data are for spring barley
MAGIC population */

```

```

/*****
      /*Import the map and genotypic data */
*****/

```

```

/*the file directory should be given by the user*/

```

```

PROC IMPORT OUT= WORK.Map_Dataset
      DATAFILE= "D:\MAGIC_Haplotyping\Map_MAGIC_J_Exp_Bot.txt"
      DBMS=TAB REPLACE;
      GETNAMES=YES;
      DATAROW=2;

```

```

RUN;

```

```

PROC IMPORT OUT= WORK.Genotyping_dataset
      DATAFILE= "D:\MAGIC_Haplotyping\GENO_J_Exp_Bot.txt"
      DBMS=TAB REPLACE;
      GETNAMES=YES;
      DATAROW=2;

```

```

RUN;

```

```

Proc sort data=map_Dataset; by Marker; run;
data map_Dataset; set map_Dataset; by Marker;
      N=_N_; call symput ('no_mark',trim(left(N)));
      if first.Marker then do; i+1;
      ii=left(put(i,7.));
      call symput ('marker' || ii, trim(Marker)); end;
      drop n i ii;

```

```

run;

```

```

      /** the parental marker information are needed (the given codes for
      parents should be according the data.
      The number of parents corresponds to the population. Sample genotypic data is
      for spring barley MAGIC population
      which is constructed of eight parents ****/

```

```

%let parent_1=Z577; %let parent_2=Z579; %let parent_3=Z581; %let
parent_4=Z583;
%let parent_5=Z587; %let parent_6=Z589; %let parent_7=Z593; %let
parent_8=Z577;

```

```

data MAGIC_parents; set WORK.Genotyping_dataset;
      if genotype eq "&parent_1" or genotype eq "&parent_2" or genotype eq
"&parent_3" or genotype eq "&parent_4"
      or genotype eq "&parent_5" or genotype eq "&parent_6" or genotype eq
"&parent_7" or genotype eq "&parent_8";

```

```

run;

```

```

/*****

```

```

      /** Note: Regarding the following example of macro settings for the
      number of parents: two parents were very similar,
      therefore instead of eight, you will find seven parents in this example.
      ****/

```

```

/*****/

/***** Macro *****/

%Macro Haplotyping_with_Fastclus(Window_Size,no_parents);
/** a file containing information of the map is necessary (sample file:
map_dataset) with variables: marker chr pos ****/

%do j=1 %to    &no_mark    ;

    /***** Hide log information *****/
    %if &j gt 3 %then %do;
        proc printto log=outlog;                run;
        proc options;                            run;
        proc printto print = outdoc; run;
        ODS Results OFF;
    %end;
    /*****/

    data map_select_marker; set Map_dataset;  if Marker eq "&marker&j";
run;
    proc sql noprint;
        select chr, pos into :chr, :Chr_pos from map_select_marker;
    quit; run;

    %let posl=%sysevalf(&Chr_pos - &Window_Size); %let
posh=%sysevalf(&Chr_pos + &Window_Size);

    data map_low;      set Map_dataset; if chr eq "&chr"; if pos gt &posl;
run;
    data map;      set map_low;  if pos lt &posh; run;

    proc sql noprint;
        select Marker into :markerlist separated by ' ' from map;
    quit; run;

    data seeds; set MAGIC_parents;
        Keep genotype &markerlist;
    run;

    /*****/

    data cluster; set Genotyping_dataset; keep genotype &markerlist; run;
    proc fastclus data=cluster seed=seeds maxclusters = 8 least=1 out=out
maxc=30 maxiter=20
        outstat=outstat outs=outmean NOPRINT; var &markerlist;
    Run;
    data out; set out; keep genotype cluster; run;
    Proc sort Data=cluster; by genotype; run;
    Proc sort Data=out; by genotype; run;
    data cluster; merge cluster out; by genotype; run;
    Proc sort Data=cluster; by cluster; run;
    Data Haplo; set out; &marker&j=cluster; keep genotype &marker&j; run;
    %do p=1 %to    &no_parents    ;

        Data Haplo_&p; set out; if genotype eq "&parent_&p"; run;

```

```

proc sql noprint;
    select Cluster into :Haplo_&p from Haplo_&p;
quit; run;

%end;

Data Haplo; set out; &&marker&j=1;
    %do p=1 %to &no_parents ;
        if cluster eq "&Haplo_&p" then &&marker&j=&p;
    %end;
    keep genotype &&marker&j;
run;
%if &j eq 1 %then %do;
    Data Haplotype; set Haplo; run;
%end;
%else %do;
    Data Haplotype; merge Haplotype Haplo; by genotype; run;
%end;
%end;

proc delete data = work.cluster; run;
proc delete data = work.haplo; run;
proc delete data = work.map_low; run;
proc delete data = work.map_select_marker; run;
proc delete data = work.out; run;
proc delete data = work.outmean; run;
proc delete data = work.outstat; run;
proc delete data = work.seeds; run;
proc delete data = work.map; run;
%do p=1 %to &no_parents ;
    proc delete data = work.Haplo_&p; run;
%end;

PROC TRANSPOSE DATA=Haplotype OUT=Haplotype_Final name= Marker Label=Marker;
ID genotype; VAR _numeric_ ; RUN;

proc sort data=Haplotype_Final; by Marker; run;
proc sort data=Map_dataset; by Marker; run;
Data Haplotype_Final; merge Map_dataset Haplotype_Final; by Marker; run;
proc sort data=Haplotype_Final; by Chr pos; run;

%Mend Haplotyping_with_Fastclus;

%Haplotyping_with_Fastclus(Window_Size=5,no_parents=7);

```