

BASEPOP- a tool to determine individuals or strains to be included in a future breeding program

Financed by: EU-FP7, grant (KBBE.2013.1.2-10) under grant agreement n° 613611(FISHBOOST)

Version: 27.10.2014

Authors: Anna K. Sonesson, Nofima, and Theo H.E. Meuwissen, Norwegian University of Life Sciences

INTRODUCTION

The overall idea with the optimisation method of BASEPOP is to estimate genomic relationships (using genetic markers) within and between strains using the method 2 of VanRaden, 2008, (P.M. VanRaden, Efficient methods to compute genomic predictions. J. Dairy Sci, 2008) and use this information to optimise individual or strain contributions to a basepopulation for a new breeding program.

There is an option either to minimise the rate of inbreeding (*minimise*) or perform a selection where the phenotypic values for traits of interest are maximised with a restriction on the rates of inbreeding (*selection*). For both *minimise* and *selection* there is an option either to perform the optimisation at the strain level (*strain*) or at the individual level (*individual*). In both cases, it is the average coancestry of the candidates that is used to control the rate of inbreeding in the offspring population. For the option *selection*, the software will (automatically) first perform *minimise*, and use the minimum average coancestry value obtained as a base for calculating the restricted level.

USAGE

BASEPOP can be either run from the DOS-prompt or by double clicking on the *basepop.exe* file using Windows-explorer. *basepop.exe* will search for a steer file called *basepop.inp*. The steer file is made by the user and has a SAS-like format. It should be a txt-file. The ordering of the lines in *basepop.inp* is arbitrary. *basepop.inp* may contain blank lines, and lines starting with the % sign are ignored.

The format of the lines is:

<keyword> <parameter1 parameter2 ... >

The keywords are (compulsory keywords are underlined):

constraint deltaf/relationship/minimise value

For *selection* optimisation, **constraint deltaf** indicates that the constraint should be on ΔF and that the desired $\Delta F = value$. The resulting ΔF in the offspring population is the given value. **constraint relationship** indicates that the constraint should be on the average relationship and that the desired $average_relationship = value$. The resulting average coancestry in the offspring population is the given value. For *minimisation* optimisation, **constraint minimise** indicates that the average relationship should be minimised, i.e. ΔF should be minimised (*value* is not needed here).

strains value

indicates that optimisation is at the *strain* or *individual* level, where 1 indicates optimisation at the *strain* level and 2 indicates optimisation at the *individual* level.

input datfilename param1name param2name paramNname

id (=id number of candidate, which agrees with id numbers in the *genotypefile*) and is either individual ID (for *individual* optimisation) or strain ID (for *strain* optimisation)

pheno (=phenotype of candidate, may be set to 0 for non-available candidates)

sex (=sex of animal, compulsory for *individual* optimisation)

cmax (same as the **cmax**-keyword, except that it applies to individual animals here; these individual **cmax** values override any global values defined by the **cmax**-keyword)

cmin (same as the **cmin**-keyword, except that it applies to individual animals here; these individual **cmin** values override any global values defined by the **cmin**-keyword)

ncand value

value indicates the total number of candidates in *datfile*. *value* is only used to allocate memory for the arrays and may thus be an overestimate of the actual number of candidates.

Example line in input file

ID datfilename param1name param2name paramNname. N can be any number.

e.g.

ID data.inp body-weight

Example corresponding datfile

ID param1data param2data paramNdata .

In file data.inp for fish with ID number 1 and a body-weight of 2500 g

1 2500

genotypefile genotypefilename number of fish or number of strains in the genotypefile number of markers

ID marker1 marker2markerX N describes the N columns of markers in the genotypefile.

For SNPs, the genotypes are set as 0, 0.5 and 1 for homozygous aa, heterozygous Aa and homozygous AA. For microsatellites, the genotypes are set as frequency of each allele. For *strain* optimisation, the population frequencies should be given for each marker.

Example genotype file

For *individual* optimisation:

individual_ID genotypes_SNP1, genotypes_SNP2, genotypes_x, genotypes_X allele2, etc

e.g. with three SNPs for individual_ID= 1, genotype_SNP1=0, genotype_SNP2=1, genotype_SNP3=0.5; individual_ID= 2, genotype_SNP1=0, genotype_SNP2=1, genotype_SNP3=1;

1 0 1 0.5

2 0 1 1

For *strain* optimisation:

strain_ID genotypes_SNP1, genotypes_SNP2, genotypes_x, genotypes_X allele2, etc

e.g. with three SNPs for strain_ID= 1, genotype_SNP1=0, genotype_SNP2=1, genotype_SNP3=0.5; strain_ID= 1, genotype_SNP1=0, genotype_SNP2=1, genotype_SNP3=1;

1 0 1 0.5

1 0 1 1

In addition to the above input, constraints on min/max number of selected sires/dams or min/max possible contributions, can be given. These four constraints can be added to the basepop.inp-file.

nfemales *value*

fixes the number of selected females to *value*, and each female contributes a fraction of offspring of $1/\textit{value}$. This is a useful option for mass-spawning species.

nmales *value*

fixes the number of selected males to *value*, and each male contributes a fraction of $1/\textit{value}$. This is a useful option for mass-spawning species.

cmax *value1 value2*

indicates the maximum fraction of the offspring that could be made by any male (*value1*) and by any female (*value2*). *value1* = 0.0 or *value1* = 1.0 indicates that any fraction between 0 and 1 could be achieved. This option can only be used for the *individual* option.

cmin *value1 value2*

indicates the minimum fraction of the offspring that could be made by any selected male (*value1*) and by any selected female (*value2*). Thus, if a (fe)male is selected, (s)he gets at least a fraction of *value1*(*value2*) of the offspring, otherwise (s)he gets zero offspring. *value1* = 0.0 indicates that any fraction between 0 and 1 could be achieved. This option can only be used for the *individual* option.

Output

The output file is called *basepop.out*. This file *basepop.out* will contain the (readable) lines of *basepop.inp*.

For *individual* optimisation - the numbers of available males and available female candidates. The phenotypic means of the selected parents, and the average relationship constraint, together with the realised average relationship of the population. The latter average relationship should equal the constraint, which is also given. The numbers of selected males and females are given together with the optimum fraction of progeny for each animal. Next to this, the phenotype and the average relationship of this animal with the rest of the population are provided.

For *strain* optimisation- the number of available strains, the phenotypic means of the strains together with the realised average relationship of the population of strains. The latter average relationship should equal the constraint, which is also given. The number of strains are given together with the optimum fraction of matings of each strain. Next to this, the average phenotype and the average relationship of each strain with the remaining selected strains.

SOME MORE EXAMPLES

Example input file

```
constraint deltax 0.01
strains 1
ncand 328
input input.dat id pheno
genotfile gen.dat 3 99
```

Example genotfile (note that all individuals here come from strain 1)

```
1 0 0.5 0 0.5 0 0.5 0 0 0 0 0.5 0 0.5 0 0 0 0 0 0 0 0 0 0 0.5 0 0 0 0 0 0 1 1 0 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 1 0 0 1 0 1 0 0 0 1 0 0 0 0 1 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0.5 0.5 0.5 0.5 0 0.5 0 0 0
0 0 0 0 0 0.5 0 1 0 0 0
1 0 0.5 0 0.5 0 0.5 0 0 0 0 0.5 0 0.5 0 0 0 0 0 0 0 0 0 0 0.5 0 0 0 0 0 0 1 1 0 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 1 0 0 1 0 1 0 0 0 1 0 0 0 0.5 1 0 0 0.5 1 0 0.5 0 0 0 0 0 0 0 0.5 0.5 0.5 0 0 0.5 0
0 0 0 0 0 0 0 0 0 0.5 0.5 0 0
1 0 0.5 0 0.5 0 0.5 0 0 0 0 0.5 0 0.5 0 0 0.5 0.5 0 0 0 0.5 0 0 0 0 0 0 0 0 0.5 0.5 0 0.5 0.5 0
0 0.5 0.5 0 0 0 0 0 0 0 0 1 0 0 1 0 1 0 0 0 1 0 0 0 0 0.5
```