**OHVsim - OHV and GS Simulations**

Documentation, June 2015

**Purpose of algorithm:**

Test the difference in genetic gain between optimal haploid value (OHV) and genomic selection in a simulated multi-generation wheat breeding program that relies heavily on crossing.

**Reference:**

Daetwyler, H.D., Hayden, M.J, Spangenberg, G.C., and B.J. Hayes. 2015. Selection on Optimal Haploid Value Increases Genetic Gain and Preserves More Genetic Diversity Relative to Genomic Selection. Genetics in press.

**Availability:**

This Fortran 90 executable is available free of charge for research purposes from GIT LINK. The Fortran program was written by Hans Daetwyler, Department of Environment and Primary Industries Victoria, Bundoora, Australia, hans.daetwyler(at)ecodev.vic.gov.au to simulate the specific breeding programs in the above paper reference. Please cite the above paper when using OHVsim or the ideas therein for publications. Example input files are also available on GIT.

**Current Version:** OHVsim-1.0.f90 (June, 2015)

Version History:

1.0 Original release

**Disclaimer:**

The authors make no guarantees on the performance of the program. Use at your own risk.

**Description of algorithm:** Described in reference above.

**Limitations of OHVsim:**

The program was written specifically to run the simulations described in the above paper. It is therefore specific to the number of chromosomes in wheat (i.e. 21, across all subgenomes). It will be possible to add in different genotypes at the same number of chromosomes but it is not very flexible for a large variety of datasets.

**Compiling:**

The OHVsim1.0.out executable has been compiled with gfortran.

**Using OHVsim:**

Run executable as ./OHVsim1.0.out

**Several input files are needed (e.g. provided on GIT).**

Parameter File (“param.txt”)

This file provides key information to OHVsim which is needed for reading in data, allocating arrays etc.

1110 !number of base generation lines

4788 !number of snp

3 !number of segments per chrom for haplotype value

8 !number of generations after F1 (so base,F1, and N gens of Fplus)

0.3 !proportion selected in generation 1

0.02 !proportion selected in generation 2+ (e.g. F1's)

100 !number of offspring per outbred cross

10 !number of elite plants to create doubled haploids from

100 !number of DH per elite plant

50 !number of top DH selected

Important

The SNP per chromosome must be dividable by the number of segments per chromosome (e.g. 228/3 = 76 SNP per segment). As a default, the program will run 2 generations (base and F1), the user can specify how many additional generations it runs in addition.

The program is designed and tested to keep population size constant per generation. It is not tested for scenarios where this is not the case. Constant population size is accomplished by keeping the proportion selected in generation 2+ and the number of offspring per cross in specific ratios. For example, in the param.txt file example above, the proportion of lines selected in 2+ generations is 0.02 and the number of offspring per (outbred) cross is 100. If you would like to change the number of offspring to 20 (i.e. 100/5), then the proportion selected in 2+ generation needs to be increased to 0.1 (i.e. 0.02\*5).

The example files provided on GIT and using the param.txt settings above will produce a constant 55200 lines/individuals per generation.

genotypes.txt

* Contains genotypes coded 0, 1, or 2 for 00, 02, 22.
* Missing genotypes set to 5.
* First column is line ID, following columns are genotypes ordered from first to last on the genome

chrom\_start\_stop.txt

* Gives starting and stopping SNP per chromosome
* Needed to allocate SNP to chromosomes

Random Number Seed

* When running the first time on a new system, it is likely best to run OHVsim without giving it a random number seed. This will automatically get it from the computer clock and create a file fort.12 that will record the seed used. You can also provide a seed in a file called seed.txt and OHVsim will use that to initiate the random number generator. fort.12 will always record the seed used.

**The program makes the following output files**:

marker-effects.txt

* Contains the sampled marker effects per marker
* Col1 = Chromosome
* Col2 = SNP number on chromosome
* Col3 = Sampled marker effect

GeneticLevelPerGeneration-GS.txt and 'GeneticLevelPerGeneration-OHV.txt

* Contain summary statistics on simulated populations for GS and OHV program respectively
* Col1 = Generation
* Col2 = Number of lines in generation
* Col3 = meanGEBV
* Col4 = GEBV variance
* Col5 = GEBV standard deviation
* Col6 = mean GEBV top 10 lines in generation
* Col7 = (Col6-Col3)/Col5

DoubledHaploidStats-GS.txt and DoubledHaploidStats-OHV.txt

* Contain summary statistics on all simulated doubled haploids GS and OHV program respectively
* Col1 = Generation
* Col2 = Consecutive number of doubled haploid restarted in each generation
* Col3 = doubled haploid GEBV
* Col4 = (Col3-meanGEBVgeneration)/(GEBV standard deviation)
* Col5 = doubled haploid OHV
* Col6 = (Col5-meanGEBVgeneration)/(GEBV standard deviation)

DH-parent-stats-GS.txt and DH-parent-stats-OHV.txt

* Contain summary statistics on all elite plant chosen for doubled haploid production for GS and OHV program respectively
* Col1 = Generation
* Col2 = Consecutive number of elite plant chosen for doubled haploid production
* Col3 = GEBV
* Col4 = (Col3-meanGEBVgeneration)/(GEBV standard deviation)
* Col5 = OHV
* Col6 = (Col5-meanGEBVgeneration)/(GEBV standard deviation)

DH-stats-best-GS.txt and DH-stats-best-OHV.txt

* Contain summary statistics only on best doubled haploids in GS and OHV program respectively
* Col1 = Generation
* Col2 = Consecutive number of best doubled haploid
* Col3 = consecutive number of doubled haploid in a generation (corresponds to DoubledHaploidStats-GS.txt Col2)
* Col4 = GEBV
* Col5= (Col3-meanGEBVgeneration)/(GEBV standard deviation)
* Col6 = OHV
* Col7 = (Col5-meanGEBVgeneration)/(GEBV standard deviation)