Friday, May 26, 2017

9:14 AM

QU-GENE variables for engine file (Z indicates integer)

**Z** number of models

* Should be 1 if using all gene effects in GE system. But random effects will likely mimic real genetic effects of a large number of genes. (Page 4 of QuCimUserManual)

**Z** random seed of random gene effects

* Ensures same gene effects are used so they are repeatable.

**Z** number of genes (includes markers and gtls)

**Z** number of environment types in TPE

**Z** Number of traits (not including marker scores)

* All traits used in selection should be counted

**Z Z Z Z Z Z Z** specify names (Ets, Trts, Genes, Alls, EPN, GPM, pop)

* 7 indicators, one dimension array with size 7, each element either 1 or 0, reporting whther the names will be specified in the file (probably not necessary for IM)

Environment Info

* "TPE (target population of environments) consists of a set of distinct, relatively homogeneous environment types, each with a frequency of occurrence. Each environment type has its own gene action and interaction providing the framework for defining GE interactions."

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Row 1: Number |  |  |  | 1 |

|  |  |
| --- | --- |
| Row 2: Name (if indicator is 1, otherwise blank) | ET1 |

* Left blank if indicator is 0. Must be filled if indicator is 1.

|  |  |
| --- | --- |
| Row 3: Frequency of occurrence in TPE | 0.5000 |

* Sum of all frequencies should be 1.

"(Environment data are stored in the Env(:) array data type in the GE system definition (Fig. 2). The number of environment types can be calculated from the array size.)"

Trait INFO

* "Genotypic value of an individual can be calculated from the definition of gene actions in the GE system and from its genotypic combination. However, breeders select based on the phenotypic value in the field. So, the phenotypic value of a genotype in a specific environment needs to be defined from its genotypic value and associated environmental errors."

Example:

C:\ED529265\EA00C09A-8B3E-4F11-89BB-CF5F62F53CC2_files\image001.png

C:\ED529265\EA00C09A-8B3E-4F11-89BB-CF5F62F53CC2_files\image002.png

C:\ED529265\EA00C09A-8B3E-4F11-89BB-CF5F62F53CC2_files\image003.png

C:\ED529265\EA00C09A-8B3E-4F11-89BB-CF5F62F53CC2_files\image004.png

C:\ED529265\EA00C09A-8B3E-4F11-89BB-CF5F62F53CC2_files\image005.png

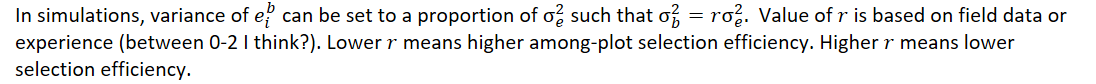
C:\ED529265\EA00C09A-8B3E-4F11-89BB-CF5F62F53CC2_files\image006.png

C:\ED529265\EA00C09A-8B3E-4F11-89BB-CF5F62F53CC2_files\image007.png

C:\ED529265\EA00C09A-8B3E-4F11-89BB-CF5F62F53CC2_files\image008.png

C:\ED529265\EA00C09A-8B3E-4F11-89BB-CF5F62F53CC2_files\image009.png

C:\ED529265\EA00C09A-8B3E-4F11-89BB-CF5F62F53CC2_files\image010.png



Define genotypic value of a genotype, then add a random affect for between-plot and within-plot error following normal distribution to the genotypic value to create the phenotypic value.

C:\ED529265\EA00C09A-8B3E-4F11-89BB-CF5F62F53CC2_files\image012.png

QuLine simulates within-family selection using phenotypic values.

Simulates among-family selection using family means.

Trait data used to define genotypic value from genotypic value.

Major trait info - environmental effects on traits (within-plot variance and among-plot variance) in each environment type.

Specify variance or individual plant level heritability in a broad sense.

For heritability, convert the specified heritability into an estimate of environmental variance based on the reference population.

Environmental variance stays constant thru simulation. Heritability varies with the changes in the genetic variation within the population from generation to generation.

Parameters for Trait Data

Row 1: Trait Number

Row 2: Trait Name (if indicator is 1, otherwise blank)

Row 3: Error Specification Type (for within, among, mixture)

1 = heritability for within-plot, and proportion of within-plot error variance for among-plot; 2 = error variance itself

* One-dimension array size 3 for 3 kinds of error (w,a,m).

Row 4+: Within/Among/Mixture errors [each ET]

Gene Info

Used to generate progeny genotypes from crossing/propagation.

Defines the genotypic value of any genotype for each trait.

Consists of:

* Location of genes on wheat chromosomes
* Number of alleles for each gene locus
* Number of traits affected by each gene
* The genotypic effects in each defined environment type
  + Linkage, multiple alleles, pleiotropy, epistasis, and GE interaction

Gene Parameters

Row 1: Gene Number

Row 2: Gene Name (if defined) - (if indicator is 1)

Row 3: Chromosome, recombination frequency, number of alleles, and number of traits the gene affects.

* All genes and markers in GE system are to be arranged in order on the chromosomes. Recombination frequency of a gene is the crossover rate between the gene and the gene located just before it (two flanking genes). If a gene is located at the beginning of a chromosome, its recombination frequency should be 0.5.

Row 4: Name of each allele (if indicator is 1, otherwise blank)

Row5+: QTL information

* Column 5: traits the gene affects
* Column 6: Environment type
* Column 7: Genotype to phenotype mapping (gene effect type)
  + 1 = additive and dominant gene without epistasis
  + 2 = epistatic gene
  + 3 = QU-GENE plug-in gene
* Column 8: How gene effects are stored
  + For additive genes (Option 1 for Column 7)
    - C:\ED529265\EA00C09A-8B3E-4F11-89BB-CF5F62F53CC2_files\image013.png
    - 0 = genotypic values in the order of AA, Aa, and aa, where A-a are two alternative alleles on the gene locus. If three alleles (A1-A2-A3), the values are arranged A1A1, A1A2, A1A3, A2A2, A2A3, A3A3. Similar for more than three alleles.
    - 1 = random gene effects with no dominance. AA and aa have random values (*AA* and *aa*) between 0.0 and 1.0. and value of genotype Aa(*Aa*) is the midpoint of *AA* and *aa*.
    - 2 = random gene effects with no over-dominance. Same as option 1 but *Aa* is just between *AA* and *aa*.
    - 3 = Random gene effects with partial/over-dominance. Same as options 1 and 2 but will result in partial or over dominance depending on chance.
  + For epistatic genes (Option 2 Column 7)
    - Which epistatic network the gene is in
  + For QU-GENE plugin genes (Option 3 for Column 7)
    - Which plugin the gene is in

Row 5+: Marker Information

* Column 5: Trait number has to be set at 0. Trait number 0 is reserved to identify which gene locus is a marker
* Column 7: Marker definition type
  + 0 = explicit values for marker types MM, Mm, and mm
  + 1 = marker is dominant
  + 2 = marker is co-dominant
* Column 8: Which gene has the marker (if required)
* Column 9: Values for each allelic combination (if required)

Population Info

Most important info: allelic combination for each individual in the population.

Two way to define population in QUG file:

1. Define the gene frequencies in the population, then qugene generates specified number of genotypes based on freqs.
2. Specify genotype of each individual in the population.

Population Parameters

Row 1: Population number

Row 2: Output file name for the population

Row 3: Population name. (indicator 1 blah blah)

Row 4: Population size

Row 5: Population Creation Type

* 1 = create population based on gene frequencies
* 2 = define stuff (not sure)
* 3 = specify alleles, all allelic combinations in population need to be known.

*For creation types 1 and 2*

Row 6+: Information for each gene in the population

* Column 1: Gene number, if the number is 0 all genes have the same specifics
* Column 2: Allele order type
  + 1 = Normal order
  + 2 = Allele of interest first
  + 3 = Allele of interest under background first
* Column 3: Gene sampling type
  + 1 = Population has the exact gene frequencies
  + 2 = Population has the sampled gene frequencies
* Column 4: Zygosity Type
  + 1 = Both hetero-zygosity and homo-zygosity are in population. Genotypes AA, Aa and aa have frequencies *p2, 2pq* and *q2*, where *p* is the allele frequency of A and *q* is the allele frequency of a.
  + 2 = Homozygous only. AA and aa have frequencies *p* and *q*.
* Column 5: Linkage type in heterozygosity. Tells how genes in double hetero-zygosity (AaBb) in the population are arranged on the chromosome.
  + 1 = Coupling (for AB/ab)
  + 2 = Repulsive (for Ab/aB)
* Column 6: Genotype combination manipulation type
  + 0 = No manipulation
  + 1+ = genotype combination is the same as gene defined
* C:\ED529265\EA00C09A-8B3E-4F11-89BB-CF5F62F53CC2_files\image014.png

Genotype parameters for type 3 creation:

* Genotype number
* Genotype name
* Allele combination on the first set of homologous chromosomes
* Allele combination on the second set of homologous chromosomes

Diagnostic information

* 0 = no information output
* 1 = summary information output
* 2 = detailed information output

General Information

* Number of strategies
* Number of runs
* Number of cycles
* Number of crosses
* Indicator for crossing block update
  + 0 = Only selected are used as parents for next cycles. Current parents will not be considered
  + 1 = Some current parents, some from selected lines
* Indicator for GE system output
  + 0 = No output of GE system information
  + 1 = Output the GE system information model by model
* Indicator for population output
  + 0 = No output
  + 1 = Output information about population before selection and after each cycle.
* Indicator for selection history output
  + 0 = No output
  + 1 = Output rogers distance for each cross, lines from each cross after each selection, lines above the mid-parent and line above the better parent at end of cycle.
* Indicator for genetic variance components output
  + 0 = No output
  + 1 = output in each environment type and among environment types for each trait in initial/selected populations.
* Parents for all crosses
  + If 'random' is specified, all parent info not required for all crosses.
  + Otherwise, input with an existing file containing the parents for each cross. Six parents per cross (2 for single, 1 for backcrossing, topcrossing, 2 for double)

NOTE: After this it is different strategies. The differences are really differences in strategies, they all start from the same input.

Parameters/Definition of a Strategy

* Strategy title
* Number of generations in the strategy
* Definition of each generation

Parameters for each selection round

* Title of the round
* Seed propagation type for the selection round
  + *clone*: asexual reproduction
  + *DH*: doubled haploid.
  + *self*: self-pollination.
  + *backcross*: backcross to one of the two parents.
  + *topcross*: cross to a third parent.
  + *doublecross*: cross to another F1.
  + *random*: random mating among the plants in a family.
  + *noself*: random mating but self-pollination is eliminated.
* Generation advance method for each selection
  + Pedigree: selected plants harvested individually, making several families
  + Bulk: harvested in bulk, making one family
* Field experiment design for each selection round
  + The number of replications for each family.
  + The number of individual plants in each replication.
  + The number of test locations.
  + The environment type for each test location.
    - Be consistent with whats in GE system. If 0, randomly assign
* Among-family selection/within-family selection
  + Num of traits used
  + Definition of each trat
    - Trait number
    - Selected proportion (proportion retained)
    - Selection mode
      * T for top
      * B for bottom
      * M for middle
      * R for random
      * TV for top cutoff
      * BV for bottom cutoff

Outputs from running QuLine

* Crosses retained after each selection (.CRO)
  + # of crosses
* Detailed GE sys info (.DEB)
* Error message (.ERR)
* Mean genotypic value in final selected pop (.FIT)
  + Mean "fitness" :
    - C:\ED529265\EA00C09A-8B3E-4F11-89BB-CF5F62F53CC2_files\image015.png
    - C:\ED529265\EA00C09A-8B3E-4F11-89BB-CF5F62F53CC2_files\image016.png
* Genes fixed in the final selected population (.FIX)
  + Percentage fixed for all traits and each trait
* Gene frequency in the initial and selected pops (.FRE)
  + Change in frequency for each gene before and after selection
* Hamming distance in the final selected pop (.HAM)
  + Measure of the difference in allelic combinations present in any genotype compared to highest target genotype in GE system.
* Selection history of all selected lines (.HIS)
* Population informaton (.POU)
  + Allele combo genotypic value, phenotypic value of each line in initial and final selected pops
* Allocated resources for all simulated strategies (.RES)
  + Number of families and indivisual plants in each generation
* Rogers distance between two parents of each cross (.ROG)
  + Rogers dist and number of lines retained from each cross
* Genetic variance in the final selected pop
  + Genetic variance and heritability in initial and selected population