

GOplan Manual

GOplan is an R package designed to manage animal breeding programs, including the breeding program of the core population and the whole breeding strategy of the crossbreeding system. Moreover, it allows optimization of the crossbreeding system.

The package has three main functions. Parameter “prm_path” is the full path of the input parameter file:

runCore(prm_path) is the function used to manage core population, you can compare the genetic progress as well as the monetary profit under different key elements in constructing populations such as the longevity of breeding stocks, mating ratio, mating methods, the number of dams, cull rate, the method of breeding value estimation, the number of individuals measured or genotyped and so on. It returns five aspects of results, the relative genetic progress, the increase of population inbreeding, the decrease of population genetic variance, the monetary profit and the phenotype.

runWhole(prm_path) can manage the crossbreeding system, it allows the evaluation of system with sire station or not, and returns the predicted value of hybrids’ phenotype and monetary profit.

runOpt(prm_path) is used to find an optimized draft of the whole crossbreeding system. By defining some uncertain parameters and giving reasonable variation ranges, it uses the Bayesian Optimization method to draw up a superior solution under fixed iteration times.

Description of parameter file:

All needed information is determined by editing the parameter files. It is worth noting that the parameters you do not need should be specified by 0, and do not use any other separators except blank to separate different numbers, or it may cause unknow errors. Do not change the sequence of each parameter.

The parameter file contains six parts, each part would be illustrated below.

Analyse Parameters

This part used to determine some basic information when run all three functions.

parameters	value	explanation
QUICK	FALSE	#FALSE, TRUE
Method	blup	#blup, ssgblup, gblup
Mate	rand	#Mate method: rand, MC or the path of extra program
Ctype	33	#type of crossbreeding: 1, 21, 22, 31, 32, 33, 41, 42
Nrep	50	#number of repetitions
Time	20	#time budget of breeding cycle
Ncores	10	#number of threads
nChr	18	#number of simulated chromosomes

nSnpPerChr	3000	#number of SNP per chr
nQtlPerChr	100	#number of QTL per chr
out_path		#path of output file

QUICK: when set to TRUE, the program would simplify the simulation of core population by replacing the process of breeding value estimation with simulation, which means just calculating the first generation's estimated breeding value (EBV) (estimate the breeding value using DMU for one time, then use its accuracy to simulate the EBV in other repetitions) and then use Cholesky method to simulate the other generations' breeding value according to the accuracy got before. We assume TBV and EBV $\sim N(u, s^2)$, so the equation of simulating EBV is:

$$EBV_{sim} = u + r * (TBV - u) + \sqrt{1 - r^2} * e * s$$

Where EBV_{sim} is the vector of simulated EBV of all individuals, u is the mean value of all individuals' TBV, r is the accuracy (Pearson's correlation coefficient between TBV and EBV), TBV is the vector of true breeding value, e is a randomly generated vector that follows a normal distribution $\sim N(0,1)$, s is the standard deviation of TBV. Noticed that

It can save time and quickly get a result, if you just want to know which factors matter more and do not care about the exact profit, it would be time-saving to use it. We highly recommend setting **QUICK** to TRUE when the **Method** is set to 'gblup' or 'ssgblup', or it would be time-consuming.

Method: now the program supports three methods to evaluate the breeding value, and call DMU to run it. blup: best liner unbiased prediction, use only pedigree information; **ssgblup**: single-step blup, use both genomic and pedigree information; gblup: genomic blup, use only genomic information.

Mate: rand: random mating; MC: minimal-coancestry mating, using annealing algorithm to find the solution; or you can give the full path of your own mating program, while our package would write out a file called "plans.txt" containing four columns, id of dam, id of sire, relationship coefficient and the expected breeding value of offspring. See implementation information of MC in the additional file.

Ctype: crossbreeding type, see detail in Figure 1, 1 means single population, 2* means tow-way crossbreeding, 3* means three-way crossbreeding and 4* means four-way crossbreeding.

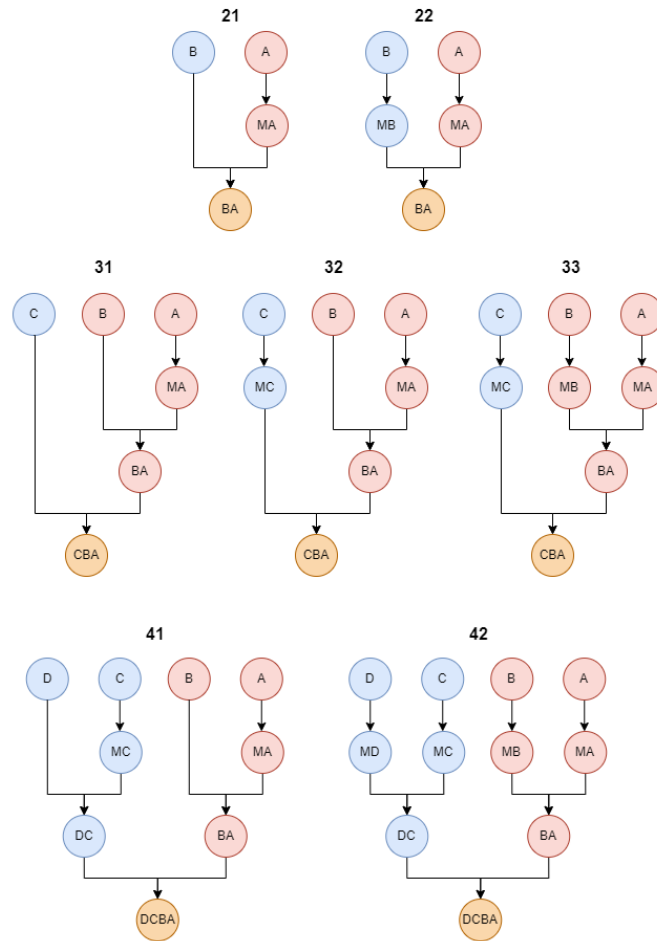


Figure1: the sketch map of different crossbreeding types.

Nrep: the number of repetitions. Final results would be the average of all repetitions.

Time: time budget of breeding cycle.

Ncores: number of threads. Our package supports multithreaded running. But you should make sure that do not set a number exceed the maximum available thread of your system.

nChr: number of simulated chromosomes. Change this parameter to adapt to different breeds.

nSnperChr: number of SNP per chr.

nQtlPerChr: number of QTL per chr.

out_path: path of output file, do not forget add "/" in the end.

VARIABLES

This part can either be specified or not, if you do not need to set variables, just skip this part. The number of variables does not have limitation, but more variables means more combinations, which may cause long time running. We recommend less than three variables each time, and one variable one row, with all the levels listed separated by blank.

Variables for **runCore(prm_path)**: Ys, Yd, Sor, nfam, nfam_F, nfam_M, ngeno, nGeno_F, nGeno_M.

Using nfam or ngeno means adjust nfam_F and nfam_M or nGeno_F and nGeno_M simultaneously.

Variables for *runWhole(prm_path)*: Asec, Bsec, Csec, Dsec, YABd, YCDs, SorAB, SorCD, SorP

Optimize Parameters

This part used to manage the range of each optimize parameter. If you don't need some parameters, just delete the row of it. When use *runCore(prm_path)*, *runWhole(prm_path)* you can let this whole part blank.

parameters	value	value	explanation
Pars	lower	upper	
Asec	2	6	#longevity of multiplier A
Bsec	1	4	#longevity of multiplier B
Csec	2	6	#longevity of multiplier C
Dsec	1	4	#longevity of multiplier D
YABd	2	6	#longevity of hybrid dam AB
YCDs	1	4	#longevity of hybrid dam CD
NAd	600	3000	#female size of nucleus A
NBd	300	1000	#female size of nucleus B
NCd	300	1000	#female size of nucleus C
NDd	300	1000	#female size of nucleus D

Population Structure

This part used to specify the structure of population in the breeding program.

parameters	value	value	value	value	explanation
dam0	2				#age of dams when first born
sire0	2				#age of sires when first born
YABd	6				#longevity of hybrid dam AB
cull_AB	0	0	0		#cull rates of each age in AB
YCDs	4				#longevity of hybrid sire CD
cull_CD	0	0	0		#cull rates of each age in CD
SorAB	100				#Mating ratio when AB cross
SorCD	100				#Mating ratio when CD cross
SorP	100				#Mating ratio of terminal cross
ABSR	0.85				#survival rate of AB
CDSR	0.85				#survival rate of CD
PSR	0.85				#survival rate of products
N_P	0				#expected number of final products

nPoints	0				#number of selected points each iteration
nIter	0				#number of iterations
Breed	A	B	C	D	
n_female	500	500	500	0	#female size
Sor	50	50	50	0	#mate ratio
Ys	2	2	2	0	#longevity of sires
Yd	4	4	4	0	#longevity of dams
n_progeny	12	12	12	0	#litter size
SR	0.85	0.85	0.85	0	#survival rate of offspring
nSec	3000	0	0	0	#size of multiplier
Ysec	6	2	2	0	#longevity of multiplier
nfam_F	2	2	2	0	#number of phenotyping females per litter
nfam_M	2	2	2	0	#number of phenotyping males per litter
nGeno_F	0	0	0	0	#number of genotyping females per litter
nGeno_M	0	0	0	0	#number of genotyping males per litter

Breeds Details

This part used to specify the information of each breed. Just list the information of breed that needed.

parameters	value	value	value	value	value	explanation
Breedname	A					
cull_s	0					#cull rates of each age in sires
cull_d	0	0	0			#cull rates of each age in dams
cull_sec	0	0	0	0	0	#cull rates of each age in multipliers
trait	JZRL					#names of considering traits
mean	180					#means of traits
var	100					#phenotype variances of traits
heri	0.3					#heritabilities of traits
weigh	-1					#weight coefficients of traits
analyse	1					#analyse method
## phenotype covariance ##						
JZRL	1					
Breedname	B					
cull_s	0					

cull_d	0	0	0			
cull_sec	0	0	0	0	0	
trait	JZRL					
mean	170					
var	100					
heri	0.3					
weigh	-1					
analyse	1					
## phenotype covariance ##						
JZRL	1					
Breedname	C					
cull_s	0					
cull_d	0	0	0			
cull_sec	0	0	0	0	0	
trait	JZRL					
mean	160					
var	100					
heri	0.3					
weigh	-1					
analyse	1					
## phenotype covariance ##						
JZRL	1					

analyse: "1" means single trait, other same numbers mean multiple traits. For example, if there are three traits, T1, T2 and T3, T1 need to be estimated by single trait, while T2 and T3 need to be estimated using multi-trait model. Then we set analyse: 1, 2, 2.

phenotype covariance ##: this part determines the covariance of traits. For example, when there are three traits, T1, T2 and T3. Then it needs to be adjusted as below:

T1	1	0	0
T2	0	1	0
T3	0	0	1

Noticed that we the column names are constant to the row names, though we omit them.

The parameter number of cull rate should always be one less than its corresponding longevity. For example, if the Yd of breed C is 4, then the cull_d of C must have three values (0, 0, 0). Because when parent exceed their longevity, they would be all culled.

Basic Economic Parameters

This part defines the economic information of the breeding program.

parameters	value	value	explanation
trait	0	JZRL	
jb_cost	906	1.5	#basic cost per individual
dam_cost	1500	0	#dam cost per individual
sire_cost	1000	0	#sire cost per individual
other_cost	10	0	#other cost each season
ind_sale	2040	0	#income of selling a product
cull_sale	1500	0	#income of culling
sire_sale	3000	0	# income of selling a sire
dam_sale	2000	0	# income of selling a dam
meas_cost	20		#phenotyping cost per individual
geno_cost	200		#genotyping cost per individual

trait: first column fixed with 0, means the initial value of all economic parameters. Then list the traits concerned, each trait corresponds to a column, means the monetary change when each trait changes one unit. The value can be negative.

jb_cost: the basic cost for raising an individual.

dam_cost: cost for keep a dam in the population each season.

sire_cost: cost for keep a sire in the population each season.

other_cost: total other cost each season.

ind_sale: income of selling a product.

cull_sale: income of culling a breeding stock.

sire_sale: income of selling a sire, a male who had been measured but not be selected can be sold as a sire.

dam_sale: income of selling a dam, a female who had been measured but not be selected can be sold as a dam.

meas_cost: phenotyping cost per individual.

geno_cost: genotyping cost per individual

Output File

runCore() would print out three files, coreOut.xlsx, coreOut_detail.xlsx and detailInfo.txt.

coreOut.xlsx has six types of information listed below, and it contains the changes of different aspects after the whole breeding process.

G	Relative genetic progress
Inb	Inbreeding increase
Profit	The economic profit
Phenotype	Predicted phenotype
Vg_Phenotype	Population genetic variance decrease of phenotype
(Last column)	The name of variation combination

coreOut_detail.xlsx contains the information of all seasons, and shows different types in separate sheets.

varg	Population genetic variance of phenotype
acc	Estimation accuracy of breeding value of phenotype
acc_index	Accuracy of select index
index	Average select index of population
dfInb	Inbreeding of population
pop0_G	Predicted phenotype

detailInfo.txt shows information about the population structure and breeding details under different variation combinations.

runWhole() would print out four types' file.

WholeOut.xlsx contains predicted phenotype of final hybrids and economic profit under different breeding programs.

*out.xlsx contains information about different subpopulations' phenotype.

detailInfo.txt is the same as described before.

p.csv is the P matrix used in gene flow method.

Additional file

1 Implementation of minimal-coancestry mating (MC) using annealing algorithm

- a) Generate an initial mate plan stochastically, calculate the overall coefficient of relationship: E_0 , set $T=1.0$, $N_{re} = 0$, $N_{eva} = 0$, where T is the annealing temperature, N_{re} is the number of plan's replace times, N_{eva} is the number of evaluation times;
- b) Change the plan by randomly select two mate pairs, and exchange the dams, the new plan's E_i is equal with $E_0 + \delta$, where $\delta = a_{s1d2} + a_{s2d1} - a_{s1d1} - a_{s2d2}$, a_{s1d2} is the coefficient of relationship between sire 1 and dam 2 (Fig.2);
- c) If $\delta < 0$, accept the new plan, or accept it with the probability of $e^{(-\delta/T)}$;
- d) Update the T value with 10% reduction each time when either the plan has been replaced $10 \cdot n_{max}$ times or has already changed plan $100 \cdot n_{max}$ times, where n_{max} is the maximum between the number of dams and the number of sires;
- e) Stop if there is no improvement between the last and the second last change of T , or repeat the above steps.

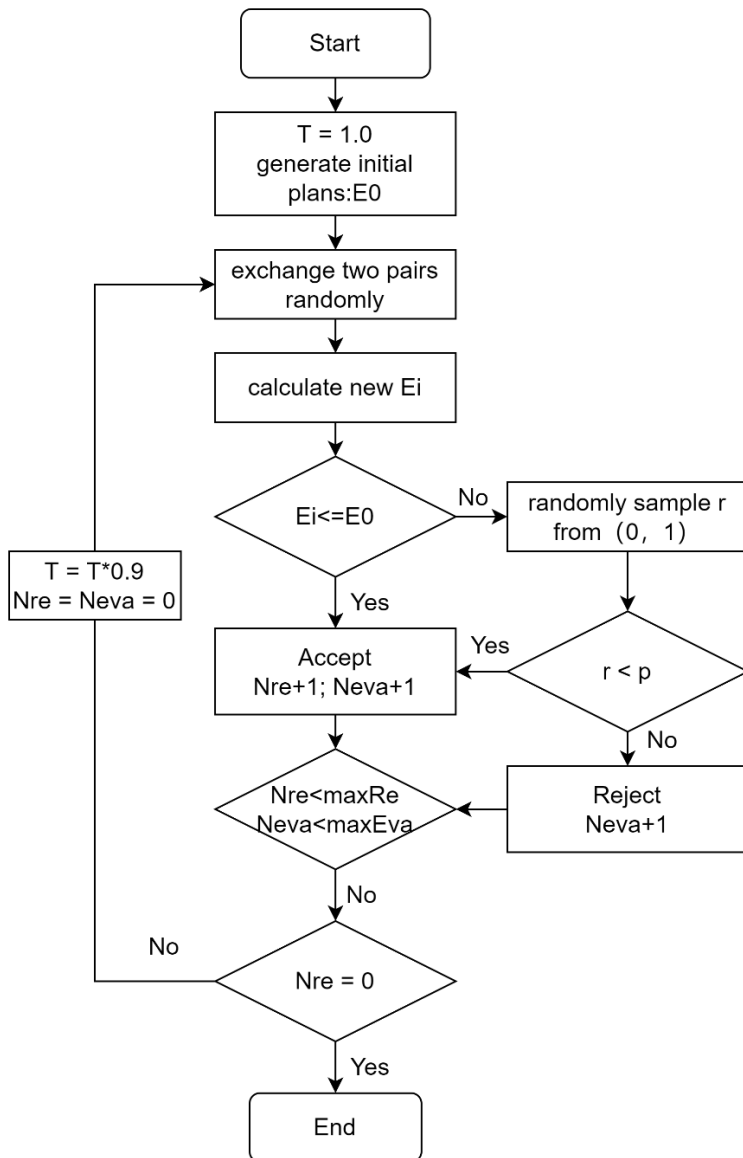


Fig. 1 A diagram illustrate the whole process of MC mate.

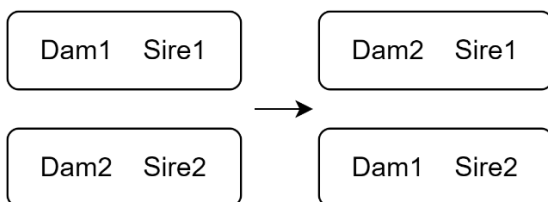


Fig. 2 Shows exchange of two mate pairs.

2 Calculation of monetary profit

In this package, we calculate the profit of each season separately and do not consider about the discount.

Main steps:

- a) Calculate the phenotype progress in time i , and figure out each economic part's value;
- b) Calculate the number of individuals of different types:

TYPE	CALCULATION METHOD
Weaning Number	The number of all dams * litter size
Culling Number	Total number of culled animals
Measuring Number	Total number of measured animals
Genotyping Number	Total number of genotyped animals
Sold Dams	Measuring Number of female - number of culled dams in core population or multiplier
Sold Sires	Measuring Number of male - number of culled sires in core population or multiplier
Product	The number of products' dams * litter size * survival rate
All Product	Weaning number * survival rate – max(Culling Number, (Measuring Number + number of culled animals of products' parents))

- c) Calculate the economic parts:

ECONOMIC PART	CALCULATION METHOD
All Basic Cost	Weaning Number * jb_cost'
All Sire Cost	Total number of sires * $sire_cost'$
All Dam Cost	Total number of dams * dam_cost'
All Measure Cost	Measuring Number * $meas_cost'$
All Genotyping Cost	Genotyping Number * $geno_cost'$
Income of selling products	All Product * ind_sale'
Income of Selling Dams	Sold Dams * dam_sale'
Income of Selling Sires	Sold Sires * $sire_sale'$
Income of Culling Animals	Culling Number * $cull_sale'$

- d) Profit = (Income of selling products + Income of Selling Sires + Income of Selling Dams + Income of Culling Animals) - (All Basic Cost + All Sire Cost + All Dam Cost + All Measure Cost + All Genotyping Cost + $other_cost$)