DESIGNING ALGORITHMS FOR MATE SELECTION WHEN MAJOR GENES OR QTL ARE IMPORTANT

R.K. Shepherd[1] and B.P. Kinghorn[2]

[1]Central Queensland University, Rockhampton, QLD 4702
[2]University of New England, Armidale, NSW 2351

SUMMARY
Look Ahead Mate Selection (LAMS) is required when future mate interactions are expected to be important. Two approaches to designing algorithms for LAMS are discussed. Deterministic group LAMS involves following the progress of selected cohorts across generations. Stochastic individual LAMS considers mating individual animals using stochastic sampling from realised QTL groups. Stochastic individual LAMS can allow variation in future response to be taken into account.
Keywords: breeding program, mate selection, QTL, evolutionary algorithms.

INTRODUCTION
Mate selection is a tactical approach for designing a breeding program in which the key issues facing animal breeders can be handled in an integrated manner (Kinghorn and Shepherd 1999). The breeding program is implemented tactically in an opportunistic manner that takes account of prevailing resources, available information and logistical constraints. Mate selection simultaneously considers the decisions of selection and mate allocation by proceeding to :

- develop a mate selection index (MSI) which describes net economic merit in terms of selection and mating decisions, and then to
- develop and implement a mate selection algorithm in order to maximise the MSI.

Total Genetic Resource Management (TGRM) is the banner under which mate selection technology has been implemented in the Australian livestock industries (Kinghorn and Shepherd 1999).

Economically important genes have been, and will continue to be, detected. These genes will vary in terms of frequency in the population, size of effect, accuracy of evaluation and pattern of expression (eg. additive/dominant/recessive/maternally imprinted/etc). Look Ahead Mate Selection (LAMS) is required to properly account for individual genes, especially non-additive genes, in mate selection (Shepherd and Kinghorn 1998). Although TGRM takes into account some breeding issues in the future (eg. long term inbreeding is handled by including coancestory in the MSI), it only considers mate selection among current candidates. Interactions between mates in later generations are not considered. This paper will discuss issues involved in designing algorithms for LAMS.

MATE SELECTION INDEX (MSI)
As LAMS involves mate selection among predicted progeny, grandprogeny, etc, the MSI will usually involve a measure of predicted merit of these descendants. Shepherd and Kinghorn (1998) consider this issue in a LAMS algorithm for crossbreeding which Hayes (2000) evaluated deterministically using a two generation MSI in which

$$MSI_{LAMS} = \sum_{i=1}^{n} PPM_i + \sum_{i=1}^{n} PGM_i$$

where n is the number of offspring per generation, *PPM* and *PGM* are predicted progeny and grandprogeny merit, respectively. This formulation can be extended for any number of generations and would usually involve consideration of expressions of production (e.g. where the main commercial tier follows 2 generations behind a nucleus), and discounting of future returns. Efficient search techniques are needed as a combinatorial number of future possibilities exist.

DETERMINISTIC GROUP LAMS

A deterministic approach to predicting future merit is to follow the progress of cohorts across generations. For example, for one known QTL with two alleles there would be 3 parental cohorts (QQ, Qq, qq), while for m QTL there would be 3^m parental cohorts. However the number of possible cohorts grows exponentially across generations when additive polygenic variation (ie. a normal curve) is present in each cohort and account is taken of the accumulation over generations of additive polygenic variation due to selection. For example, Figure 1 shows that selecting parents from two parental QTL cohorts (QQ, Qq) will result in the possibility of six progeny cohorts (3 QQ, 2 Qq and 1 qq) which differ in cohort mean due to the accumulated selection. In Figure 1, p_i (i = 1,..,4) is the proportion selected from each QTL group and there are 4 male and 4 female selected groups. If the predicted progeny are selected for breeding grandprogeny, there would be the possibility of 35 grandprogeny cohorts with different means. Hence to simultaneously consider many generations into the future, it will be necessary to restrict the number of possible cohorts available for mating each generation. This would probably be done on the basis that the cohort (or its descendants) will not be competitive due to merit (i.e. the average merit of the cohort may be judged too small to be competitive in the future).

Having decided on the number of possible cohorts available each generation then an evolutionary algorithm (EA) could be used to evolve the best mating set among current candidates based on the chosen MSI. An important component of an EA is the genetic representation of a feasible solution. A genetic representation of a feasible solution (mating set) for figure 1 could be a single array (or chromosome) consisting of 12 numbers:

Male cohort				Female cohort				Mating order			
QQ		**Qq**		**QQ**		**Qq**					
p_1^M	p_2^M	p_3^M	p_4^M	p_1^F	p_2^F	p_3^F	p_4^F	3	1	4	2

where p is the proportion of the parental male (p^M) or female (p^F) cohort selected for breeding. The mating order of the selected groups is given in the EA chromosome by the number of the selected male group as the female groups are in standard order (ie. 3, 1, 4, 2 means the male x female mating group combinations are 3x1, 1x2, 4x3 and 2x4). Starting with an initial population of EA chromosomes, new generations of the population evolve using the concepts of survival (or selection) of the fittest and breeding new chromosomes using genetic operators, all in a probabilistic framework. Differential evolution (Kinghorn and Shepherd 1999) has been found to be a suitable EA for mate selection. In general the genetic representation of a solution for LAMS would involve all the possible cohorts across the generations included in the MSI. As this number can grow

exponentially if many generations are involved, an alternative is to consider mating individual animals in a stochastic approach as there is a fixed number of matings allowed each generation.
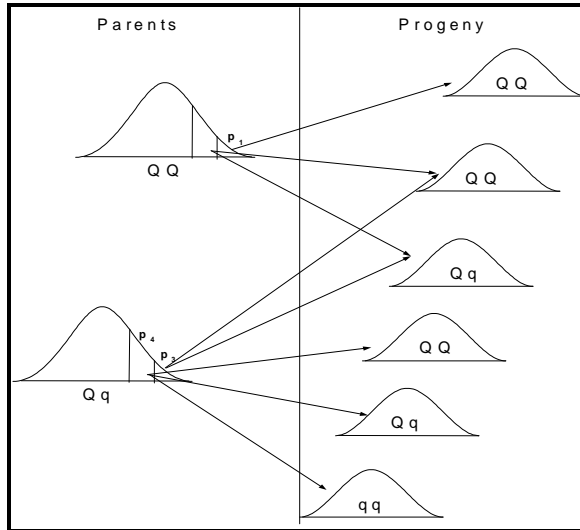

Figure 1. Six possible progeny cohorts from two parental QTL cohorts.

STOCHASTIC INDIVIDUAL LAMS

In the approach proposed, a genetic representation of a feasible solution (mating set) is represented by t chromosomes, one for each generation, with each consisting of two major components; one component gives the number of times each female candidate (given in standard order) is mated while the other gives the number of the male candidate used in each mating.

Chromosome	No. matings of female candidate j					Male candidate used in mating i				
Current candidates, gen 0	1	2	0	1	………………	4	5	4	2	………………
Predicted progeny, *gen 1*	3	0	0	1	………………	2	5	1	5	………………
. .	.	.	.	.	.	.	.	.	.	.
. .	.	.	.	.	.	.	.	.	.	.
Generation *t - 1*	2	3	1	1	………………	8	3	8	3	………………

Figure 2. Genetic representation (*t* chromosomes) of a feasible LAMS mating set.

For example if progeny were being predicted t generations ahead then a possible LAMS solution is displayed in Figure 2. This means that male x female mate selections among the current candidates are 4x1, 5x2, 4x2, 2x4, etc, while mate selections among the predicted progeny are 2x1, 5x1, 1x1, 5x4, etc. To determine the value of the MSI we calculate the merit of the proposed matings starting with the current candidates.

Predicted progeny merit (*PPM*) from a mating between current candidates is calculated as follows:

$$PPM_i = QTLeffect_i + \hat{u}_i$$

$$\hat{u}_i = \tfrac{1}{2}(\hat{u}_s + \hat{u}_d) + order_i \times s^*_{Aw}$$

where $\hat{u}_i$, $\hat{u}_s$ and $\hat{u}_d$ are the polygenic EBVs of predicted progeny i and its sire and dam, $order_i$ is the normal order statistic for individual i which characterises the deterministic Mendelian sampling for each sex, and where s^{*2}_{Aw} is the within family additive genetic variation [$= \tfrac{1}{4}(2 - F_s - F_d)s^2_A$]. The term $QTLeffect_i$ is the merit of the progeny QTL genotype which the offspring inherits stochastically from the Mendelian sampling of known QTL genotypes of its sire and dam. This formulation of merit in which the progeny polygenic component is predicted deterministically using order statistics, while the progeny QTL group can vary in a probabilistic manner, allows future mate selection opportunities to be properly accounted for. Once the n progeny of the current generation are simulated then a similar process is used to predict grandprogeny, etc. according to the proposed mating plan specified by the t chromosomes.

An appropriate EA can be designed to find an optimal LAMS solution for this genetic representation (Shepherd and Kinghorn 1999). However the difficulty in this case is that different values of the MSI will be realised for the same LAMS solution due to the stochastic nature of predicting progeny. Hence a number of replicates should be performed for each LAMS solution and then some measure of the distribution of MSI values (e.g. mean, median, mean − 0.1 x standard deviation) can be used as the fitness criterion of the EA.

CONCLUSIONS
 The deterministic group LAMS approach will be useful for evaluating breeding programs for very large populations when future mate interactions are expected to be important (eg. non-additive QTL). Techniques for handling the exponential explosion in the number of groups will be required if many generations are to be handled simultaneously. The stochastic LAMS approach will be preferred for smaller populations when running repetitions of each LAMS solution is possible in real time. The stochastic nature is advantageous as it will provide a measure of possible outcome variation from that particular LAMS mating plan. An appropriate fitness function could incorporate some measure of this risk. In both approaches the amount of computation will be considerable and so more research is required into developing efficient LAMS algorithms for real applications.

REFERENCES
Hayes, B.J. (2000) PhD thesis, University of Central Queensland.
Kinghorn, B.P. and Shepherd, R.K. (1999) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **13**: 130.
Shepherd, R.K. and Kinghorn, B.P. (1998) *Proc. 6th World Cong. Genet. App. Livest. Prod.* **25**: 431.
Shepherd R.K. and Kinghorn B.P. (1999) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **13**:126.