Efficient Breeding by Genomic Mating

Deniz Akdemir *
Department of Plant Breeding & Genetics
Cornell University
Ithaca, NY USA
Julio Isidro Sánchez
Agriculture & Food Science
University College Dublin
Dublin, Ireland

February 8, 2016

Abstract

Many breeders use selection as a first step for improvement of complex traits. The theory of selection is well studied in the genetics literature. In this article, we propose a new approach to breeding which focuses on mating instead of selection, we refer it to as genomic mating. We believe breeding by genomic mating uses the genomic information better than a recently proposed method of breeding methodologies like genomic selection. We used concepts of risk and diversification from the economics for formulating the optimum mating theory. The results from our simulations point to the efficiency of our method to selection based breeding.

Keywords & Phrases: Breeding, phenotypic selection, genomic selection, genomic mating, complex traits, genome-wide markers, inbreeding, genomic diversity, portfolio optimization

1 Introduction

The role of mating as an evolutionary force has been described by many thinkers. However, this evolutionary information have been in part ignored by breeding methods that focus only on improvement by selection.

In this article we discuss that both PS and GS are in a sense inefficient for improving complex traits in the long run, mainly because they are methods of directional selection and both of these methods, lose genetic variation (i.e., increase the level of inbreeding). The importance of inbreeding for long term success of selection has been expressed by few authors ([11, 13, 27]). Some

*Corresponding author: deniz.akdemir.work@gmail.com

approaches to control inbreeding in the selection paradigm have also been proposed ([13]). For example, in [11], it is shown that the gain in selection response from using individual genes was proportional to the variance they explained. Goddard ([11]) argues that, GS is likely to lead to a more rapid decline in the selection response unless new alleles are continuously added to the calculation of GEBVs. Goddard also recommended a weighted selection scheme, where favorable alleles with low allele frequency receive more weight. In the same article, the relationship between the effective population size and response to selection is studied. A small but informative simulation study in [13] also stresses the importance of balancing short and long term gains by controlling inbreeding in selection.

We argue in this paper that, the solution lies in replacing the selection paradigm (GS or PS) with optimal mating designs. This implies that instead of only inferring which lines in the breeding population will become the parents of the next generation, our solution should directly give the list of mates from current breeding population for each of the progeny in the next generation. When put in this way, the problem of breeder of finding the crosses that needs to be made to improve the genetic level at each cycle while controlling inbreeding, can be expressed as an optimization problem similar to the ones used in the selection paradigm [24, 12, 1]. This is a different approach than selection because the solution set includes the possibility of all lines in the breeding population contributing to the next generation. This approach brings a breakthrough method for plant breeding programs.

Except for a few approaches like assignment of mates using sequential selection of least-related mates ([22]) or linear programming ([14]) and more recent work of [27], which deal with selection of parents followed by mating, there is a scarcity of compelling approaches to combine both of these steps. The proposed genomic mating approach combines these two steps. The aim of this paper is to develop a mathematical theory for GM and compare its performance to that of the traditional breeding approaches, PS and GS.

2 Methods

Selection is an evolutionary phenomenon that affects the phenotypic distribution of a population. From a breeding point of view, selection means breeding from the "best" individuals ([10]). The simplest form of selection is to those individuals on the basis of their own phenotypic values, which is called phenotypic selection (PS). Phenotypic selection has been in use since domestication of crops thousand of years ago. In PS, phenotypes of individuals are used for selection of individuals for advancement in the breeding cycle.

The development of molecular marker techniques has offered new selection possibilities and new breeding schemes approaches ([9]). Marker assisted selection ([16]), marker assisted introgression ([8]), and marker assissted recurrent selection ([5]) have been used as methods to make selection.

With the advent of high throughput genotyping, a new tool for animal and

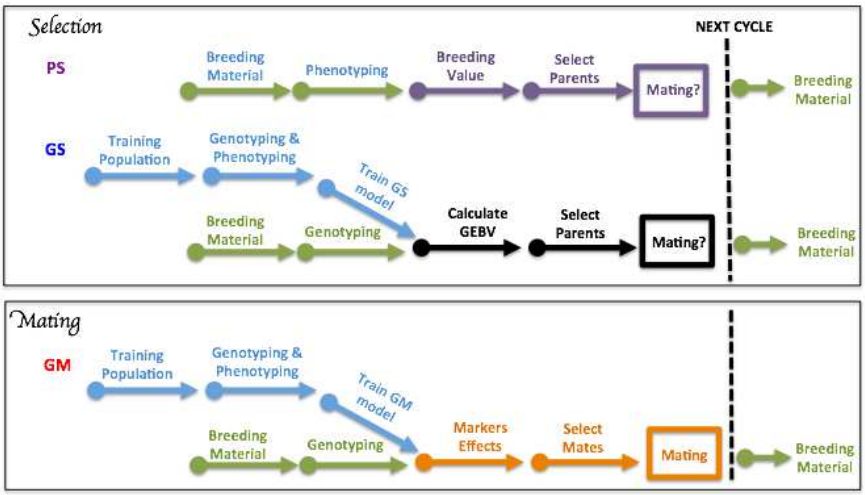

Figure 1: Diagram for the different generation interval approaches in plant breeding. Phenotypic selection (PS) and genomic selection (GS) are selection methods, and genomic mating (GM) is the mating approach. Arrows indicate the different stages during the generation interval. In PS, starting with a set of parents as breeding material, selection is performed based on phenotypes. In GS, the breeding value is predicted using a statistical model based on phenotypes and the using of whole-genome markers, and selection is based on GEBVs. GM is similar to GS in terms of training a model, but with the difference is that this information is used to decide on mates not just for selecting the top parents.

plant breeding have been developed, called GS. Genomic selection use genome-wide markers to estimate the effects of all genes or chromosome positions simultaneously ([21]) by calculating genomic estimated breeding values (GEBVs). This GEBVs are used for selection of individuals for advancement in the breeding cycle. This process involves use of genotypic and phenotypic data to build a prediction model, that will estimate GEBV's for selected individuals with only genotypic data. The adaptation of GS in many breeding programs, is due not just because allows increasing the genetic gain by reducing the generation interval in the breeding program, but also because GS is more suitable for complex quantitative traits with small effects. Many factors are involved in the relative per unit of time efficiency of these methods and their short and long time performance.

2.1 Genetic Gain and Inbreeding

It is widely accepted that short term gains from selection increases with increased selection intensity. However, increasing selection reduces the genetic variability, which increases the rates of inbreeding and may reduce gains in the long term run. Most of the selection in plant breeding are designed to maximize genetic gain. A few optimized selection schemes have been proposed to balance the gain from selection and variability ([29, 6, 20, 21]). We will give a brief review of these approaches since they relate to the mating theory.

Let A be a matrix of additive genetic relationships (by definition this is equal to twice the matrix of coefficients of co-ancestry) between the individuals in the genetic pool (this matrix can be obtained from a pedigree of genome-wide markers for the individuals) and let $\boldsymbol{c}$ be the vector of proportional contributions of individuals to the next generation under a random mating scheme. The average relatedness for a given choice of $\boldsymbol{c}$ can be defined as $r = \frac{1}{2}\boldsymbol{c}'A\boldsymbol{c}$. If $\boldsymbol{b}$ is the vector of GEBV's, i.e., the vector of BLUP estimated breeding values of the candidates for selection. The expected gain is defined as $g = \boldsymbol{c}'\boldsymbol{b}$. Without loss of generality, we will assume that the breeders long term goal is to increase the value of g.

In [29, 6, 20] an approach that seeks maximizing the genetic gain while restricting the average relationship is proposed. The optimization problem can be stated as

$$\begin{aligned} \text{minimize} \quad & r = \boldsymbol{c}'\tfrac{A}{2}\boldsymbol{c} \\ \text{subject to} \quad & \boldsymbol{c}'\boldsymbol{b} = \rho \\ & \boldsymbol{c}'\mathbf{1} = 1 \\ & \boldsymbol{c} \geq 0 \end{aligned} \tag{1}$$

This problem is easily recognized as Quadratic Optimization problem (QP). There are many efficient algorithms that solves QP's so there is in practice little difficulty in calculating the optimal solution for any particular data set. Recently, several allocation strategies were tested using QP's in [11, 22, 25]. It is easy to extend these formulations to introduce additional constraints as positiveness,minimum-maximum for proportions, minimum-maximum for number of lines (cardinality constraints).

2.1.1 Balancing Genetic Level and Inbreeding: Efficiency Frontier

By solving the QP in (1) given for varying values of ρ we can trace out an efficient frontier, a smooth non-decreasing curve that gives the best possible trade off of genetic variance against gain, i.e., the curve represents the set of Pareto-optimal allocations. This curve is called the efficiency frontier (EF) in economics literature. One such EF curve is shown in Figure 2 for a set of 50 simulated lines.

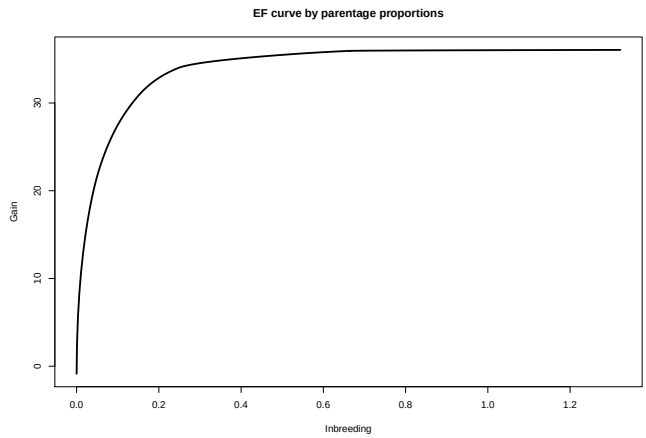

Figure 2: The EF curve for 50 simulated genotypes and a trait. The points along the curve are the "Pareto" optimal points balancing gain and inbreeding. Any point below this line is sub-optimal and any point above the line is unattainable.

We would like to bring to the attention of the readers, that there is a big similarity between the efficient portfolio selection problem and the solution offered by Matkowitz in the 1950's ([18, 19]), with the GS theory we have reviewed in this section. Portfolio theory attempts to maximize portfolio expected return for a given amount of portfolio risk, or equivalently minimize risk for a given level of expected return, by finding optimal portfolio proportions. It is in fact beneficial to consider investments in the parents as a capital investment problem, since the economical ideas such as risk-return relationships, risk diversification, etc,.. are applicable also to the investment in the populations by the breeding programs. However, we believe that this theory needs to be appropriately modified because of the specific considerations related to poliploidy, meiosis, and more generally, to reproduction in living organisms which use DNA as a mode of information transfer in reproduction.

2.2 Optimal Genomic Mating

The role of mating as an evolutionary force has been described by many thinkers since Darwin. Mating choice is a major force in the evolution of many characters seen in sexually reproducing organism ([4, 26, 7]0. What governs the operation of sexual selection is the relative parental investments in their offspring 9[15, 2, 3, 28]0. However this evolutionary information have been in part ignored by breeding methods that focus only on improvement by selection.

In this section, we will describe our proposed methodology. We introduce measures of expected inbreeding and gain for a breeding population that are the descendants of genotyped / ungenotyped individuals.

Let $\boldsymbol{b} = (\boldsymbol{b}_1', \boldsymbol{b}_2', \boldsymbol{b}_3')'$ denote the vector of genetic effects corresponding to the

parents and progeny, where $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are the genetic effects of the N parents and $\boldsymbol{b}_3$ are the genetic effects of the N_c progeny . Let the pedigree based numerator relationship matrix for the individuals in $\boldsymbol{b}$ is A and A is partitioned as

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}$$

corresponding to the partitions of $\boldsymbol{b}$. Suppose, we also have the markers for the parents in the second partition, and $\boldsymbol{u}_2 = M\boldsymbol{a}$ where M is the matrix of minor allele frequencies. Let M_c be the marker allele frequency centered incidence matrix and $\boldsymbol{a}$ is the vector of marker effects. Covariance of $\boldsymbol{b}_2$ can be written as

$$cov(\boldsymbol{b}_2) = \frac{M_c M_c}{k}\sigma_b^2 = G\sigma_b^2$$

where k is twice the sum of heterozygosities of the markers (VanRaden, 2008).

Following [23] and [17], let P be a matrix containing the transitions from ancestors to offspring. We will refer P as the mating or parentage matrix. Then, we can write $\boldsymbol{b} = P\boldsymbol{b} + \boldsymbol{\psi}$ where $\boldsymbol{\psi}$ is the vector of Mendellian samplings and founder effects with a diagonal variance D. In particular, using only the rows of P corresponding to the $\boldsymbol{b}_3$ the relationship is written as

$$\boldsymbol{b}_3 = \begin{bmatrix} P_{31} & P_{32} & P_{33} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \boldsymbol{b}_3 \end{bmatrix} + \boldsymbol{\psi}_3$$

which can also be written as a regression equation of the form $\boldsymbol{b}_3 = (I - P_{33})(P_{31}\boldsymbol{b}_1 + P_{32}\boldsymbol{b}_2 + \boldsymbol{\psi}_3)$ ([23]). The covariance matrix of $\boldsymbol{b}_3$ is given by

$$cov(\boldsymbol{b}_3) = (I - P_{33})(P_{31}A_{11}P_{31}' + P_{32}GP_{32}' + D_3)(I - P_{33})'. \tag{2}$$

The variances caused by Mendelian sampling in D_3 are related to inbreeding in the parents via

$$var(\psi) \propto (1/2 - (F_1 + F_2)/4)$$

where F_1 and F_2 are the inbreeding coefficients of the two parents which can be extracted from the diagonals of G. The covariance formula reduces to

$$cov(\boldsymbol{b}_3) = P_{32}GP_{32}' + D_3$$

if all the founders are genotyped, and a relatively simple mating strategy is assumed where founders are the only parents and no back-crossing is allowed. This is the assumption made for the remaining of this paper and in this case P_{32} is a N_c children from N parents matrix with each row having two 1/2 values at positions corresponding to two distinct parents or only a value of 1 at the position corresponding to the selfed parent. All the other elements of this matrix are zero. Nevertheless, one can easily imagine situations where some of the founders are not genotyped or when some of the progeny also have progeny,

then the formula in (2) will be relevant. Construction of the matrix P for more complex mating plans is described in [23].

The $cov(\boldsymbol{b}_3)$ gives us the expected covariance of the progeny given the transition matrix P_{32} and the realized relationship matrix G of the parents. This can be used as to measure the expected genetic diversity of a mating plan: we can use a measure in the spirit of $c'Ac$ in (1) by

$$Inbreeding(P_{32}) = \mathbf{1}'_{N_c} cov(\boldsymbol{b}_3)\mathbf{1}_{N_c} = \mathbf{1}'_{N_c}(P_{32}GP'_{32} + D_3)\mathbf{1}_{N_c}.$$

We also need a measure for genetic gain. A simple measure of gain for a given mating plan expressed in P_{32} can be constructed from the expected value of $\boldsymbol{b}_3$:

$$E(\boldsymbol{b}_3) = P_{32}M\boldsymbol{a}$$

and an overall measure can be written as

$$Gain(P_{32}) = \mathbf{1}'_{N_c} E(\boldsymbol{b}_3).$$

We want to complement the measure "gain" with a measure of within cross variance for the genetic levels of children of the parent pairs. Under the infinitesimal model, within cross variance arises from poliploidy and heterozygosity.

Suppose the organism under study is diploid. We can decode the markers matrix M coded as -1, 0, and 1 into a matrix M^* using the information in the marker effects vector $\boldsymbol{a}$ such that markers are coded as the number of beneficial alleles. We can also obtain a related marker effects vector $\boldsymbol{a}^*$ by replacing the original marker effects by the effects of the beneficial alleles so that we have $M\boldsymbol{a} = (M^* - 1)\boldsymbol{a}^*$. For a given parent pair of parents, we can calculate the vector expected number of beneficial alleles of the children of these parents using a transition vector $\boldsymbol{p}$ as $\boldsymbol{\mu} = E(\boldsymbol{m}) = \boldsymbol{p}'M^*$. In addition, for each locus we can calculate the variance for the number of beneficial alleles from the number of alleles the parents have and put them in a vector which we will denote by $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_m)$. Calculation of elements of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ from the coding in M^* can be as in Table 1. We define risk measure for this parent pair as

$$risk(\lambda_1) = (\boldsymbol{\mu} + \lambda_1 * \begin{pmatrix} \sqrt{\sigma_1/m} \\ \sqrt{\sigma_2/m} \\ \ldots \\ \sqrt{\sigma_m/m} \end{pmatrix} - \mathbf{1}_m)'\boldsymbol{a}^*$$

where $\lambda_1 \geq 0$ is the risk parameter and m is the number of markers. The risk of a mating plan (which is expressed in P_{32}) is the sum of all the risk scores for all mate pairs in that plan which we will denote by $Risk(P_{32}, \lambda_1)$.

If the risk parameter is set to zero then we have $Risk(P_{32}, \lambda_1 = 0) = \mathbf{1}'_{N_c} E(\boldsymbol{b}_3) = \mathbf{1}'_{N_c} P_{32} M\boldsymbol{a}$. The magnitude of this parameter is related to the desire of breeder to take advantage of within cross variances and increasing it gives more weight to high variance mate pairs and heterozygosity.

Table 1: Calculation of mean number and variance of the beneficial alleles of progeny at each locus from the beneficial allele code (-1, 0, 1) of the parents at the same locus.

Parent 1	**Parent 2**	**# Beneficial Allele**	**Variance**
1	1	2	0
1	0	1.5	0.5
0	1	1.5	0.5
1	-1	1	0
-1	1	1	0
0	0	1	2/3
0	-1	0.5	0.5
-1	0	0.5	0.5
-1	-1	0	0

In this sense, the efficient mating problem can be stated as an optimization problem as follows:

$$\begin{aligned} &\text{minimize} \quad Inbreeding(P_{32}) = \mathbf{1}'_{N_c}(P_{32}GP'_{32} + D_3)\mathbf{1}_{N_c} \\ &\text{subject to} \quad Risk(P_{32}, \lambda_1) = \rho \end{aligned} \tag{3}$$

where the minimization is over the matrices P_{32}. In the above optimization problem, we are trying to minimize the inbreeding in the progeny while the risk is set at the level $\rho \geq 0$. In the remaining of this paper we will prefer to use the the following equivalent formulation of the mating problem:

$$\text{minimize} \quad r = -Risk(P_{32}, \lambda_1) + \lambda_2 * Inbreeding(P_{32}) \tag{4}$$

where $\lambda_2 \geq 0$ is the parameter whose magnitude controls the amount of inbreeding in the progeny, and the minimization is again over the space of the mating matrices P_{32}. λ_1 controls allele heterozygosity weighted by the marker effects and λ_2 controls allele diversity.

The problem stated in Equation (1) and the mating problem are similar. However, the differences are noteworthy: the solution to (1) only will give the breeder the proportional contributions of parents, and leaves the mating problem unanswered. A solution of the mating problem will give us the matrix P_{32} from which an enumeration of all mates leading the progeny in the next breeding cycle can be obtained.

The main advantage of switching from an inbreeding measure based on parental contributions to a formulation based on mating plans is that, the later allows a better control of inbreeding by completely controlling who mates with who. In this way, there is an intrinsic limit to the amount of selfing or crosses of closely related lines instead of leaving the decision to a roulette wheel assignment of parents as mates in the formal method. A probabilistic assignment of mates might lead to too much inbreeding. For example, if the parental contribu-

tion proportion of a parent is 0.50, then we expect to have 25% to be obtained by selfing this parent.

The optimization problem in two is a combinatorial problem whose order increases with the number of individuals in the breeding population. We have used an evolutionary algorithm to tackle this optimization problem and found that the algorithm we have devised is very efficient for finding good solutions in reasonable computing time. We did not explore any alternatives to our algorithm but similar evolutionary algorithms like particle swarm, genetic algorithms, tabu search, and simulated annealing are some other popular algorithms which can solve this problem.

2.2.1 Efficiency frontier curve, risk and return

The frontier curve drawn using the optimal mating algorithm in Figure 3displays the same kind of information as in the EF curve in Figure 2 . The coordinates of the points on the curve are the values of estimated genetic gain (risk) and inbreeding (diversity) for optimal set of mates. By changing the λ_2 we move along this curve. Since these are the optimal points, the breeder will choose one of these points.

We want to stress the differences: First, as opposed to the continuous parentage contribution proportions solutions in the GS method , the mating method gives discrete solutions. That is to say, the solutions of the mating algorithm are the list of parent mates of the progeny. Second, there is no real guideline for choosing where to operate while using GS method. On the other side, since the mating algorithm is discrete and the number of genotypes contributing to the next generation increase starting from one as we increase the λ_2, we can identify a point to operate on this curve by slowly increasing the λ_2 until a desired minimum number of genotypes are included in the solution. This is the method we have used in our simulations where we have run simulations of several cycles of mating. we included the minimum number of parents as a paremeter: "minparents" in simulations. This allowed us to run the simulations many times without interference. However, a better approach in practical situations would be to plot the whole frontier curve and select a solution that has a good risk to diversity ratio (somewhere close to the kink of the frontier curve). The decision for the λ_1 depends on how much more risk a breeder wants to take.

To explore the effect of changing parameters for the mating algorithm, we have devised the following simulations: A marker data was created for N genotypes by randomly generating 1000 markers for each genotype. By introducing independent and identical normally distributed marker effects at 500 of randomly selected the loci we have defined a trait. Using the generated marker effects, the genetic marker data we have identified optimal the mates for the progeny at changing values of λ_1 and λ_2. A selected subset of solutions for N=50 are displayed in Figure 3. In addition, for changing values of these parameters we have plotted the gain and the inbreeding values for the optimal mates. Each curve in this figure display the efficient solutions at a fixed level of λ_1 for changing values of the λ_2 parameter.

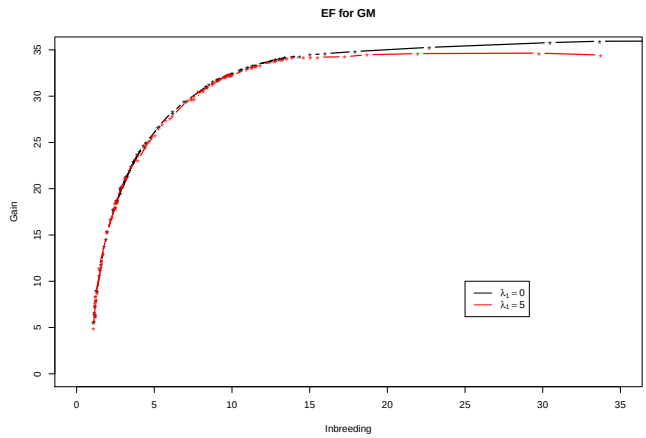

Figure 3: The EF curve for same 50 simulated genotypes and a trait in Figure 2. The points along the curve are the "Pareto" optimal points balancing gain (or risk) and inbreeding at two levels of λ_1. The effect of increasing λ_1 is forcing more heterozygosity and reducing inbreeding and reduced gains at λ_2 close to zero.

Finally, in Figure 5 the results from simulations for the study of the long term behavior of PS, GS, Efficient GS and GM. Starting from 2 founders we have formed a population of 150 genotypes with 1000 snps at 3 chromosomes each and carried this population through 200 generations of random mating and 100 generations of phenotypic selection based on a complex trait (300 QTL at random locations on each chromosome) with 0.5 heritability generated based on the infinitesimal model. Starting from this initial population, we have simulated 10 rounds of PS, and 20 rounds of GS and GM (assuming one cycle of PS and two cycles of GS and GM per year). The results of 5 replication of this simulation are presented in Figure 5. Each thin line represents the genetic gains over cycles by different methods over a replication of the experiment. The thick lines show the mean improvement for each of the methods over 5 replications. In this simulation study there is a clear advantage of using GM as a breeding method.

3 Conclusions and Discussions

In this article we have proposed a new methodology for breeding living organisms based on optimal genomic determination of mating plans, our approach can be contrasted with the selection paradigm where only proportional contributions of parents to the progeny are the main focus. To this end we have adopted economical concepts as risk and return to the specific case of breeding by mating. Although similar to GS in its information requirements, our approach offers a better use of the available genotypic and phenotypic information.

We have provided several examples and compared our method by simulations to the selection methodologies. We have found that the optimal genetic mating approach very promising for improving short and long term gains. We believe that successful application of GM will increase the rates of gains per cycle.

Under the optimal mating breeding scheme some concepts in statistical genetics like selection intensity will have to be adopted so that the choice between gain and genetic variability of the next generation become the main focus, not the cut off point approach in selection.

Although we have only demonstrated the use of GM using marker estimates, it is possible to adjust the methodology to work with phenotypic records or the BV's obtained from these. it can be argued that where PS is relatively more efficient than GS, mating using BV's and the marker data of the parents will be beneficial for balancing risks and returns.

Now, we list a short list of factors that will be important to the success of GM. As its true for almost all tools, blind use of GM can be problematic. The success of GM lean on the improvement of accuracy of marker effect estimates, which in turn depends on the following factors:

- Heritability of the trait,
- Linkage disequilibrium,
- Complexity of the trait,
- Statistical model,
- Size of breeding and training populations,
- Inbreeding in the breeding population,
- Relevance of training population to the breeding population.

The expected genetic gain and effectiveness of GM is directly proportional to the prediction accuracy of GM models. Prediction accuracy is defined as the correlation between the GEBVs and the true breeding value divided by the square root of heritability (h2) [**?**, **?**] Trait heritability, genetic architecture and LD are the only factors that cannot be controlled by the researcher. Among the factors that are under control of the researcher, the size of the training and breeding population, and the strength of genetic relationships between them are the most important factors affecting prediction accuracy. Under most circumstances, GM accuracies increase with increasing on the values of the different vectors. For example, increasing marker density will help to increase the probability of finding markers that are in consistent LD with the trait of interest.

References

[1] D. Akdemir, J. I. Sanchez, and J.-L. Jannink. Optimization of genomic selection training populations with a genetic algorithm. *Genet Sel Evol*, 47:38, 2015.

[2] S. J. Arnold. Is there a unifying concept of sexual selection that applies to both plants and animals? *American Naturalist*, pages S1–S12, 1994.

[3] S. J. Arnold and D. Duvall. Animal mating systems: a synthesis based on selection theory. *American Naturalist*, pages 317–348, 1994.

[4] S. C. Barrett. The evolution of plant sexual diversity. *Nature Reviews Genetics*, 3(4):274–284, 2002.

[5] R. Bernardo and A. Charcosset. Usefulness of gene information in marker-assisted recurrent selection: a simulation appraisal. *Crop Science*, 46(2):614–621, 2006.

[6] J. Brisbane and J. Gibson. Balancing selection response and rate of inbreeding by including genetic relationships in selection decisions. *Theoretical and Applied Genetics*, 91(3):421–431, 1995.

[7] H. L. Carson. Mate choice theory and the mode of selection in sexual populations. *Proceedings of the National Academy of Sciences*, 100(11):6584–6587, 2003.

[8] A. Charcosset and F. Hospital. Marker-assisted introgression of quantitative trait loci. *Genetics*, 147(3):1469–1485, 1997.

[9] A. Charcosset and L. Moreau. Use of molecular markers for the development of new cultivars and the evaluation of genetic diversity. *Euphytica*, 137(1):81–94, 2004.

[10] D. S. Falconer, T. F. Mackay, and R. Frankham. Introduction to quantitative genetics (4th edn). *Trends in Genetics*, 12(7):280, 1996.

[11] M. Goddard. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*, 136(2):245–257, 2009.

[12] J. Isidro, J.-L. Jannink, D. Akdemir, J. Poland, N. Heslot, and M. E. Sorrells. Training set optimization under population structure in genomic selection. *Theoretical and Applied Genetics*, 128(1):145–158, 2015.

[13] J.-L. Jannink. Dynamics of long-term genomic selection. *Genetics Selection Evolution*, 42(1):35, 2010.

[14] G. Jansen and J. Wilton. Selecting mating pairs with linear programming techniques. *Journal of dairy science*, 68(5):1302–1305, 1985.

[15] R. Lande and S. J. Arnold. The measurement of selection on correlated characters. *Evolution*, pages 1210–1226, 1983.

[16] R. Lande and R. Thompson. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, 124(3):743–756, 1990.

[17] A. Legarra, I. Aguilar, and I. Misztal. A relationship matrix including full pedigree and genomic information. *Journal of dairy science*, 92(9):4656–4663, 2009.

[18] H. Markowitz. Portfolio selection. *The journal of finance*, 7(1):77–91, 1952.

[19] H. M. Markowitz. *Portfolio selection: efficient diversification of investments*, volume 16. Yale university press, 1968.

[20] T. Meuwissen. Maximizing the response of selection with a predefined rate of inbreeding. *Journal of animal science*, 75(4):934–940, 1997.

[21] T. H. E. Meuwissen, B. J. Hayes, and M. E. Goddard. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4):1819–1829, 2001.

[22] J. Pryce, B. Hayes, and M. Goddard. Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information. *Journal of dairy science*, 95(1):377–388, 2012.

[23] R. Quaas. Additive genetic model with groups and relationships. *Journal of Dairy Science*, 71(5):1338–1345, 1988.

[24] R. Rincent, D. Laloë, S. Nicolas, T. Altmann, D. Brunel, P. Revilla, V. M. Rodriguez, J. Moreno-Gonzalez, A. Melchinger, E. Bauer, et al. Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (zea mays l.). *Genetics*, 192(2):715–728, 2012.

[25] S. Schierenbeck, E. Pimentel, M. Tietze, J. Körte, R. Reents, F. Reinhardt, H. Simianer, and S. König. Controlling inbreeding and maximizing genetic gain using semi-definite programming with pedigree-based and genomic relationships. *Journal of dairy science*, 94(12):6143–6152, 2011.

[26] I. Skogsmyr and Å. Lankinen. Sexual selection: an evolutionary force in plants? *Biological Reviews of the Cambridge Philosophical Society*, 77(04):537–562, 2002.

[27] C. Sun, P. VanRaden, J. O'Connell, K. Weigel, and D. Gianola. Mating programs including genomic relationships and dominance effects. *Journal of dairy science*, 96(12):8014–8023, 2013.

[28] M. J. Wade. The ecology of sexual selection: Mean crowding of females and resource-defence polygyny. *Evolutionary Ecology*, 9(1):118–124, 1995.

[29] N. Wray and M. Goddard. Moet breeding schemes for wool sheep 1. design alternatives. *Animal Production*, 59(01):71–86, 1994.

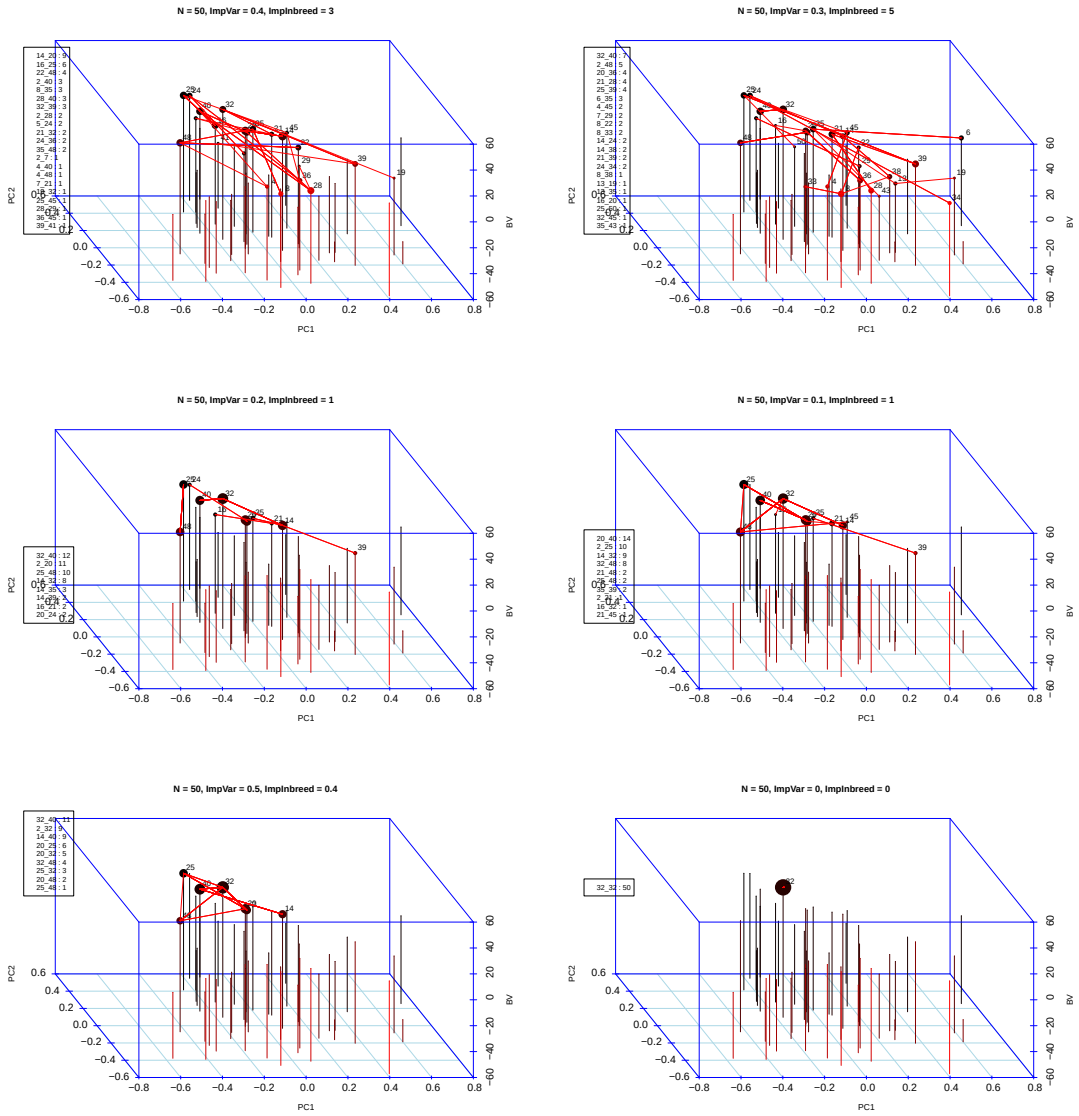

Figure 4: For a the set of 50 simulated lines, the optimal solutions to the mating problem at a few selected values of λ_1 and λ_2.

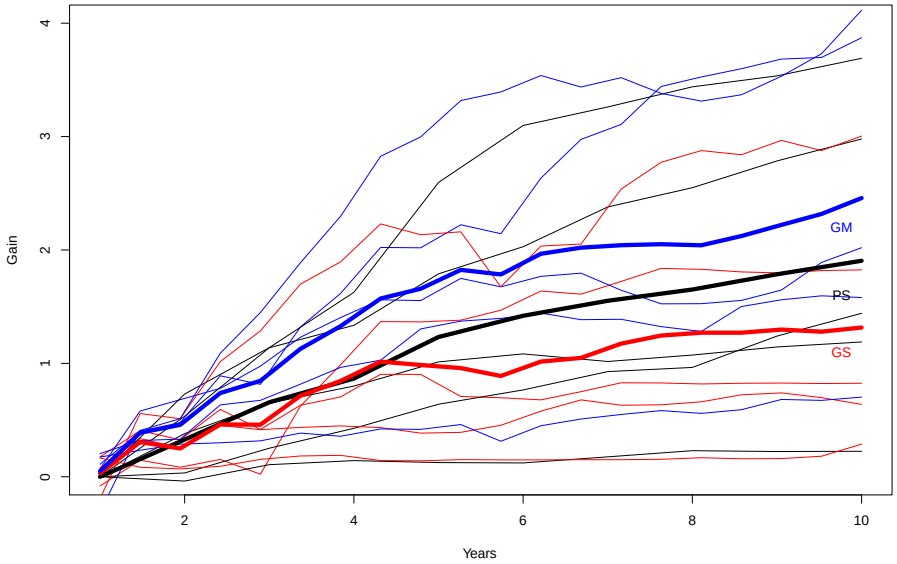

Figure 5: The long term behavior of PS, GS, Efficient GS and GM. Starting from this initial population, we have simulated 10 rounds of PS, and 20 rounds of GS and GM (assuming one cycle of PS and two cycles of GS and GM per year). Each thin line represents the genetic gains over cycles by different methods over a replication of the experiment. The thick lines show the mean improvement for each of the methods over 5 replications. In this simulation study there is a clear advantage of using GM as a breeding method.