Efficient Breeding by Genomic Mating

Deniz Akdemir[*,1] **and Julio Isidro Sánchez**[†]

[*]Cornell University, Ithaca, NY USA, [†]University College Dublin, Dublin, Ireland

ABSTRACT In this article, we propose an approach to breeding which focuses on mating instead of truncation selection, our method uses genome-wide marker information in a similar fashion to genomic selection so we refer it to as genomic mating. Using concepts of estimated breeding values, risk (usefulness) and inbreeding, an efficient mating approach is formulated for improvement of breeding values in the long run. We have used a genetic algorithm to find solutions to this optimization problem. Results from our simulations point to the efficiency of genomic mating for breeding complex traits compared to truncation selection.

KEYWORDS Breeding; Complex traits; Genomic selection; Phenotypic selection; Genome-wide markers

Selection is an evolutionary phenomenon that affects the phenotypic distribution of a population. From a breeding point of view, truncation selection means breeding from the "best" individuals (Falconer *et al.* 1996). Breeders have been selecting on the basis of phenotypic values since domestication of plants and animals or, more recently, breeders have substantially used the pedigree-based prediction of genetic values for the genetic improvement of complex trait (**????**); this is called phenotypic selection (PS).

Since the invention of the polymerase chain reaction by Mullis in 1983, the enhancements in high throughput genotyping (Lander *et al.* 2001; Margulies *et al.* 2005; Metzker 2010) have transformed breeding pipelines through marker-assisted selection (MAS) (Lande and Thompson 1990), marker assisted introgression (Charcosset and Hospital 1997), marker assisted recurrent selection (Bernardo and Charcosset 2006), and genomic selection (GS) (Meuwissen *et al.* 2001). The latter use genome-wide markers to estimate the effects of all genes or chromosome positions simultaneously (Meuwissen *et al.* 2001) to calculate genomic estimated breeding values (GEBVs), which are used for selection of individuals. This process involves the use of phenotypic and genotypic data to build prediction models that would be used to estimate GEBV's from genome wide marker data. It has been proposed that GS increases the genetic gains by reducing the generation intervals and also by increasing the accuracy of estimated breeding values. However, many factors are involved in the relative per unit of time efficiency of GS and its short and long time performance (Jannink *et al.* 2010; Daetwyler *et al.* 2007).

Some optimized parental contribution calculation schemes have been proposed to balance the gain from selection and variability (Wray and Goddard 1994; Brisbane and Gibson 1995; Meuwissen 1997; Meuwissen *et al.* 2001; Sonesson *et al.* 2012; Clark *et al.* 2013). Approaches that seek for an optimal subset of mates among potential male and female candidates have been formulated from an animal breeding perspective in Allaire (1980); Jansen and Wilton (1985); Kinghorn (1998) and in subsequent articles (Kinghorn and Shepherd 1999; Fernández *et al.* 2001; Berg *et al.* 2006; Kinghorn 2011; Pryce *et al.* 2012; Sun *et al.* 2013). These approaches also seek solutions that attain a balance between genetic gains and inbreeding and most developments in this area have been focusing on animals.

Marker assisted breeding to stack genes using complementary crosses has been useful for breeders when the trait of interest is regulated by only a few loci. For complex traits, on the other hand, there is a scarcity of methods available to breeders. Both of PS and GS focus on improvement by truncation selection, mainly ignoring the role of mating and complementation as an evolutionary force (Figure 1). For this reason both PS and GS are, in a sense, inefficient for improving complex traits in the long term. Methods that seek only a balance between genetic gains and inbreeding are incomplete because they ignore the variances in the genetic values; measures of gain do not completely capture the full potential of a mate pair.

In this article, we propose an optimal genomic mating (GM) approach for breeding (Figure 1). Our approach is focused on mainly on plant breeding scenarios. We believe it uses genomic information more completely than the recently proposed genomic selection and reinforces mating complementary individuals. Given a set of individuals in the current breeding population, their corresponding markers and related marker effects, our solution is a list of mates
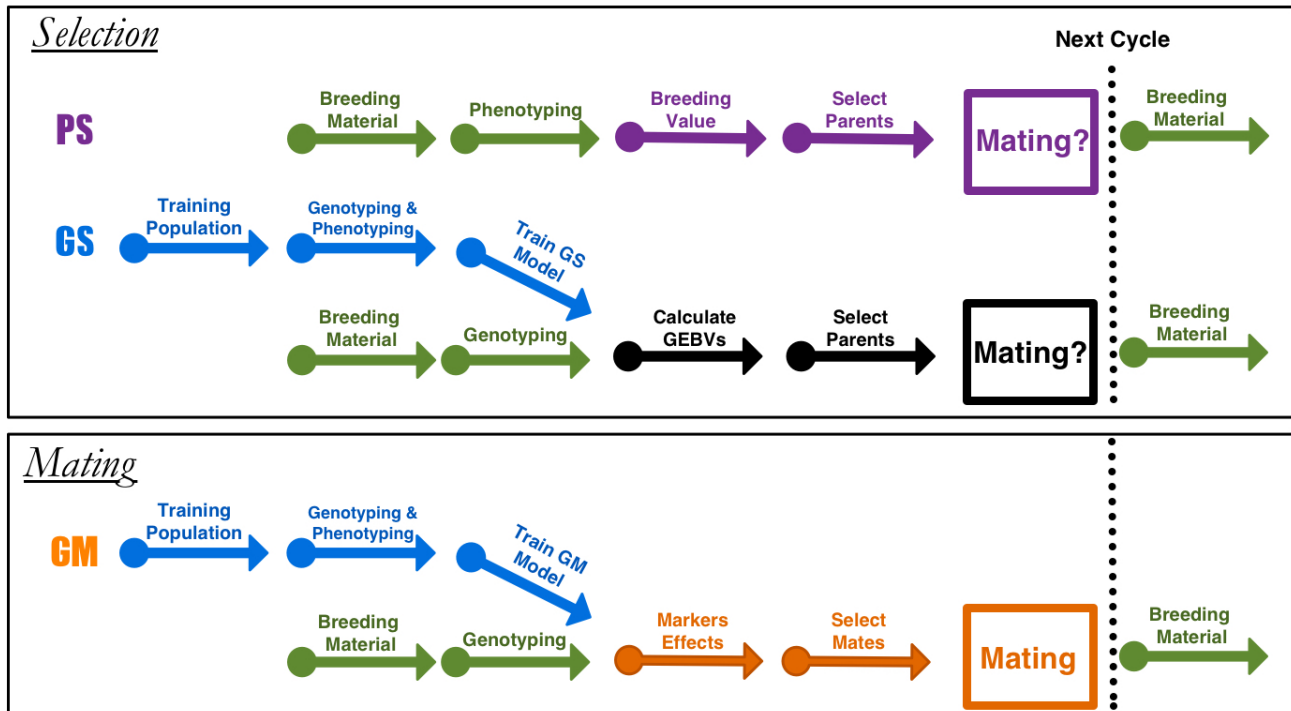
Figure 1 Diagram for the different breeding approaches. Phenotypic selection (PS) and genomic selection (GS) are truncation selection methods, and genomic mating (GM) is the mating approach. Arrows indicate the different stages in a breeding cycle. In PS, starting with a set of parents as breeding material, selection is performed based on phenotypes. In GS, the breeding value is predicted using a statistical model based on phenotypes and whole genome marker data (obtained within an experiment that is repeated in every few cycles, blue arrows) selection is based on GEBVs. Genomic mating is similar to GS in terms of estimating marker effects, but in GM the genetic information and the estimated marker effects are used to decide for the list of mates that should be crossed to obtain the next breeding population. Genomic mating is the only approach that gives an answer to the mating question: "Who is mating whom?".

that should be crossed for obtaining the next breeding population instead of a list of individuals in the current breeding population which will become the parents of the next generation. Unlike selection methods, GM approach does not exclude the possibility of contribution of all individuals to the next generation. A cross-variance term is included in the objective function along with genetic gains and inbreeding to account for potential benefits from including mates with higher estimated genetic variance. To this end, we provide a method that uses marker effect estimates to estimate within cross-variances assuming independence among loci and additive effects. The difficult computational problem of finding the optimal set of mates have been handled by an efficient genetic algorithm. Using simulations, we have compared the long range performance of GM to PS, GS and an optimal parentage contribution approach. Results from these simulations point to the viability and efficiency of genomic mating for breeding complex traits.

Methods

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 but some approaches try to balance the gain from selection and variability. We will give a brief review of these approaches since they relate to the mating theory.

Current methodology.

Many authors (Goddard 2009; Jannink 2010; Sonesson *et al.* 2012; Sun *et al.* 2013; Clark *et al.* 2013) have expressed the importance of reducing inbreeding in PS and GS for long-term success. They argued 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, stressing the importance of balancing short and long term gains by controlling inbreeding in selection.

Let A be a matrix of pedigree based numerator relationships or the additive genetic relationships between the individuals in the genetic pool (this matrix can be obtained from a pedigree of genome-wide markers for the individuals) and let 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 c can be defined as $r = \frac{1}{2}c'Ac$. If 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 $g = c'b$. Without loss of generality, we will assume that the breeders long term goal is to increase the value of g.

In (Wray and Goddard 1994; Brisbane and Gibson 1995; Meuwissen 1997) an approach that seeks maximizing the genetic gain while restricting the average relationship is proposed. The optimization problem can be stated as

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

This problem is easily recognized as a 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 (Goddard 2009; Pryce *et al.* 2012; Schierenbeck *et al.* 2011). 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).

Some authors recommended mate selection approaches that also seek a balance between gain and inbreeding from an animal breeding perspective (Allaire 1980; Jansen and Wilton 1985; Kinghorn 1998). Kinghorn in a series of articles (Kinghorn 1998; Kinghorn and Shepherd 1999; Kinghorn 2011) describes an algorithmic approach that separates the optimization and the objective function for the mate selection approach and therefore can be used for a wide array of optimization criteria (mate selection index) with hard and soft constraints. Similar algorithmic approaches were recommended in Fernández *et al.* (2001); Pryce *et al.* (2012); Sun *et al.* (2013). However, none of these methods include a term for the genotypic variance of the crosses, such as described in this paper.

By solving the QP in (1) for varying values of ρ, or using the similar criteria in the mate selection approaches, we can trace out an efficient frontier curve, a smooth non-decreasing curve that gives the best possible trade-off of genetic variance against gain. This curve represents the set of optimal allocations and it is called the efficiency frontier (EF) curve in finance (Markowitz 1952) and breeding literature.

Optimal genomic mating.

There are several alternative measures of inbreeding based on mating plans (Leutenegger *et al.* 2003; Wang 2011). In this article, we have used a measure derived under the infinitesimal genetic effects assumption proposed by (Quaas 1988) and (Legarra *et al.* 2009). A measure of gain, i.e., the total expected breeding value of the progeny, can also be calculated from the results of the same authors. However, in our belief, the expected value by itself is not a good measure of possible gains since it carries no information about the variability of breeding values (BV's) among full-sibs. Therefore, we have derived a measure called the risk of a mating plan (this is related to the concept of "usefulness") by increasing the expected BV's of the progenies by a small amount (the intensity is controlled by the parameter λ_1) proportional to their expected variance (standard deviation) calculated under the infinitesimal effects assumption. Other measures of expected variance could also been used. For example, it is possible to calculate this variance by simulating progenies for parent pairs, and one can easily include information about the LD in these simulations. Another measure of risk was proposed in (Zhong and Jannink 2007). The measures of inbreeding and risk we chose are computationally efficient and this makes the optimization over the mates feasible.

Combining the measures of inbreeding and risk into one leads to the formulation of the mating problem:

$$\underset{P_{32}}{\text{minimize}} \quad r(\lambda_1, \lambda_2, P_{32}) = -Risk(\lambda_1, P_{32}) + \lambda_2 * Inbreeding(P_{32}) \tag{2}$$

where $\lambda_2 \geq 0$ is the parameter whose magnitude controls the amount of inbreeding in the progeny, and the minimization is 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. When $\lambda_1 = 0$ the risk measure is the same as total expected gain.

Now, we give the details of how the measures $Risk(\lambda_1)$ and $Inbreeding$ are defined in this paper. 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 $\boldsymbol{b}$ be 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, coded as $0, 1$, and 2. Let M_c be the $N \times m$ marker allele frequency centered incidence matrix ($M_c = M - 2\mathbf{1}_N(p_1, p_2, \ldots, p_m)$) and $\boldsymbol{a}$ be the vector of marker effects. Variance-covariance of $\boldsymbol{b}_2$ can be written as

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

where $k = \sum_{j=1}^{m} 2p_j(1 - p_j)$ is twice the sum of heterozygosities of the markers (VanRaden 2008).

Following Quaas (1988) and Legarra *et al.* (2009), 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 stated as a regression equation of the form $\boldsymbol{b}_3 = (I - P_{33})^{-1}(P_{31}\boldsymbol{b}_1 + P_{32}\boldsymbol{b}_2 + \boldsymbol{\psi}_3)$ (Quaas 1988). The variance-covariance matrix of $\boldsymbol{b}_3$ is given by

$$Var(\boldsymbol{b}_3) = (I - P_{33})^{-1}(P_{31}A_{11}P_{31}' + P_{32}GP_{32}' + P_{32}A_{21}P_{31}' + P_{31}A_{12}P_{32}' + D_3)(I - P_{33})'^{-1}. \tag{3}$$

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 variance-covariance formula reduces to

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

4   Akdemir *et al.*

if all the founders are genotyped (no P_{31}), and a relatively simple mating strategy is assumed where founders are the only parents and no back-crossing is allowed ($P_{33} = \mathbf{0}$). This is the assumption made for the remainder of this paper and in this case P_{32} is a $N_c \times N$ matrix (N_c children from N parents) 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 (3) will be relevant. If some founders are not genotyped but a pedigree is available relating them to the rest of the founders then the variance-covariance for the founders, $Var(\boldsymbol{b}_1, \boldsymbol{b}_2)$, can be calculated using the relationship matrix in Legarra *et al.* (2009). Furthermore, construction of the mating matrices for more complex mating plans is described in (Quaas 1988).

$Var(\boldsymbol{b}_3)$ gives us the expected variance-covariance of the progeny given the mating 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 line with the inbreeding term $\boldsymbol{c}'A\boldsymbol{c}$ in (1) by

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

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

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

Finally, we want to complement the measure "gain" with a measure of within cross-variance for the genetic levels of children of the parent pairs. Suppose the organism under study is diploid. We can recode the markers matrix M coded as -1, 0, and 1 into a $N \times m$ matrix M^* using the information in the marker effects vector $\boldsymbol{a}$ such that markers are coded as the number of beneficial alleles as 0,1, or 2. This is achieved by first obtaining marker effects estimates and then using the sign of the estimates to determine what is a beneficial allele. We can also obtain a related marker effects vector $\boldsymbol{a}^*$ by replacing the original marker effects by the effects of the beneficial alleles ($\boldsymbol{a}^* = |\boldsymbol{a}|$) so that we have $M\boldsymbol{a} = (M^* - \mathbf{1}_{N \times m})\boldsymbol{a}^*$. For a given parent pair, 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 $\boldsymbol{\sigma_p} = (\sigma_{p1}, \sigma_{p2}, \ldots, \sigma_{pm})$. Calculation of elements of $\boldsymbol{\sigma_p}$ from the coding in M^* can be as in Table 1. We define risk measure for this parent pair as

$$Risk(\lambda_1) = \left(\boldsymbol{p}'M^* + \lambda_1 * \begin{pmatrix} \sqrt{\sigma_{p1}} \\ \sqrt{\sigma_{p2}} \\ \ldots \\ \sqrt{\sigma_{pm}} \end{pmatrix} - \mathbf{1}_m \right)' \boldsymbol{a}^*$$

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 λ_1 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 λ_1 is related to the desire of the breeder to take advantage of within cross variances and encourages mating parents that are heterozygotes at QTL.

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

$$\underset{P_{32}}{\text{minimize}} \quad Inbreeding(P_{32}) = \mathbf{1}'_{N_c}(P_{32}GP'_{32} + D_3)\mathbf{1}_{N_c}$$

$$\text{subject to} \qquad Risk(P_{32}, \lambda_1) = \rho. \tag{4}$$

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 remainder of this paper, we will use the the following equivalent formulation of the mating problem in Equation (2).

The optimization problem in (2) is a combinatorial problem whose order increases with the number of individuals in the breeding population and the number of progeny. We have devised a genetic algorithm to tackle this optimization

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	Expected Number of Beneficial Alleles	Variance
1	1	2	0
1	0	1.5	0.25
0	1	1.5	0.25
1	-1	1	0
-1	1	1	0
0	0	1	0.5
0	-1	0.5	0.25
-1	0	0.5	0.25
-1	-1	0	0

problem and found that the algorithm is very efficient for finding good solutions in reasonable computing time. Genetic algorithms (Holland (1973); Davis *et al.* (1991); Goldberg (2006)) are particularly suitable for optimization of combinatorial problems. The idea is to use a population of candidate solutions that is evolved toward better solutions. At each iteration of the algorithm, a fitness function is used to evaluate and select the elite individuals and subsequently the next population is formed from the elites by genetically motivated operations like crossover and mutation. It should be noted that the solutions obtained by a genetic algorithm will usually be sub-optimal and different solutions can be obtained given a different starting population of candidate solutions. We did not explore any alternatives to our mating optimization algorithm, but similar evolutionary algorithms like differential evolution, particle swarm, tabu search, and simulated annealing or hill climbing methods like the exchange method can be useful to solve this problem. As stated by other authors Kinghorn (2011) and Pryce *et al.* (2012), the mate selection problem has two independent components: A mate selection index (MSI), i.e., the optimization function and a mate selection algorithm that can be used to optimize the MSI. In our article, we have provided new approaches to both of these components: First, the MSI we have used differed from previous authors and included terms for gain, variance and inbreeding, and secondly, we have adopted a genetic algorithm that can efficiently look for good solutions.

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. Additionally, there is no real guideline for choosing where to operate while using GS method. Conversely, 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 surface 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 several cycles of mating. We included the minimum number of parents as a parameter: "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 surface and select a solution that has a good risk to diversity ratio.

There is an intrinsic limit to the amount of selfing or crosses of closely related lines in GM. Although it is hard to imagine that this is what is done in practice, theoretically, leaving the decision to a "roulette wheel" assignment of parents as mates as in the selection approach might lead to too much inbreeding. For example, if the parental contribution proportion of a parent is 0.50, then we expect to have 25% obtained by selfing this parent. GM allows a better control of inbreeding by completely controlling who mates with whom.

Results

For a set of 50 simulated lines, we have identified optimal mates for the progeny at changing values of λ_1 and λ_2. The frontier surface is drawn using the optimal mating algorithm (Figure 2). The coordinates of the points on the curve are the values of estimated risk, inbreeding and the difference between risk and gain. for the optimal sets of mates. The blue surface represents the optimal values of the objective function in Equation (2) Points below this surface correspond to sub-optimal regions and points above this surface are unattainable. The points along the surfaces are the optimal
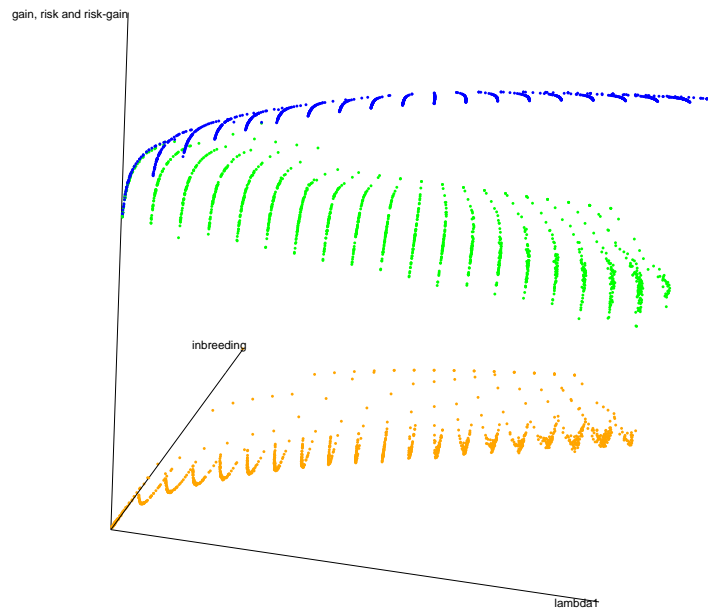
Figure 2 A marker data was created for 50 genotypes by randomly generating 1000 markers for each genotype. By introducing independent and identically normally distributed marker effects at 500 of randomly selected the loci we have defined a trait. Three surfaces are given in the figure. The blue surface represents the optimal values of the objective function in Equation (2) Points below this surface correspond to sub-optimal regions and points above this surface are unattainable. The points along the surfaces are the optimal points balancing gain, risk and inbreeding. The green surface is the expected average genetic value of the progeny and the orange surface is the value of the cross-variance term, these two surfaces add up to the blue surface.

points balancing gain, risk and inbreeding. The green surface is the expected average genetic value of the progeny and the orange surface is the value of the cross-variance term, these two surfaces add up to the blue surface. By changing λ_1 and λ_2 we move on this surface. Since the points on this surface correspond the optimal solutions, the breeder should operate on the surface. The optimal solutions to the mating problem at a few selected values of λ_1 and λ_2 are in Figures 3a-3d.

Efficient frontier surface is the basis for GM. A feasible mating plan is one that meets specified constraints. The EF surface allows breeders to understand how a mating plan's expected risk vary with the amount of inbreeding. However, the decision depends on how much more or less risk a breeder wants to take. Most breeders will be willing to assume a greater inbreeding for a greater risk. Breeders differ in the amount of inbreeding they are willing to take for a given risk. Breeders who are inbreeding averse require lower inbreeding for a given amount of risk than breeders who are risk seekers.

Figure 4a and 4b show the results from simulations for the study of the long term behavior of PS, GS, and GM. In this simulation study, there is a clear advantage of using GM as a breeding method.

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 approach where only proportional contributions of parents to the progeny are the main focus. A major novelty in GM approach as compared to the other methods is the utilization of within cross-variances (usefulness) in the objective function along with genetic gains and inbreeding.

Although similar to GS in its information requirements, our approach offers a better utilization of the genotypic and phenotypic information. Under the optimal mating breeding scheme some concepts in statistical genetics like selection
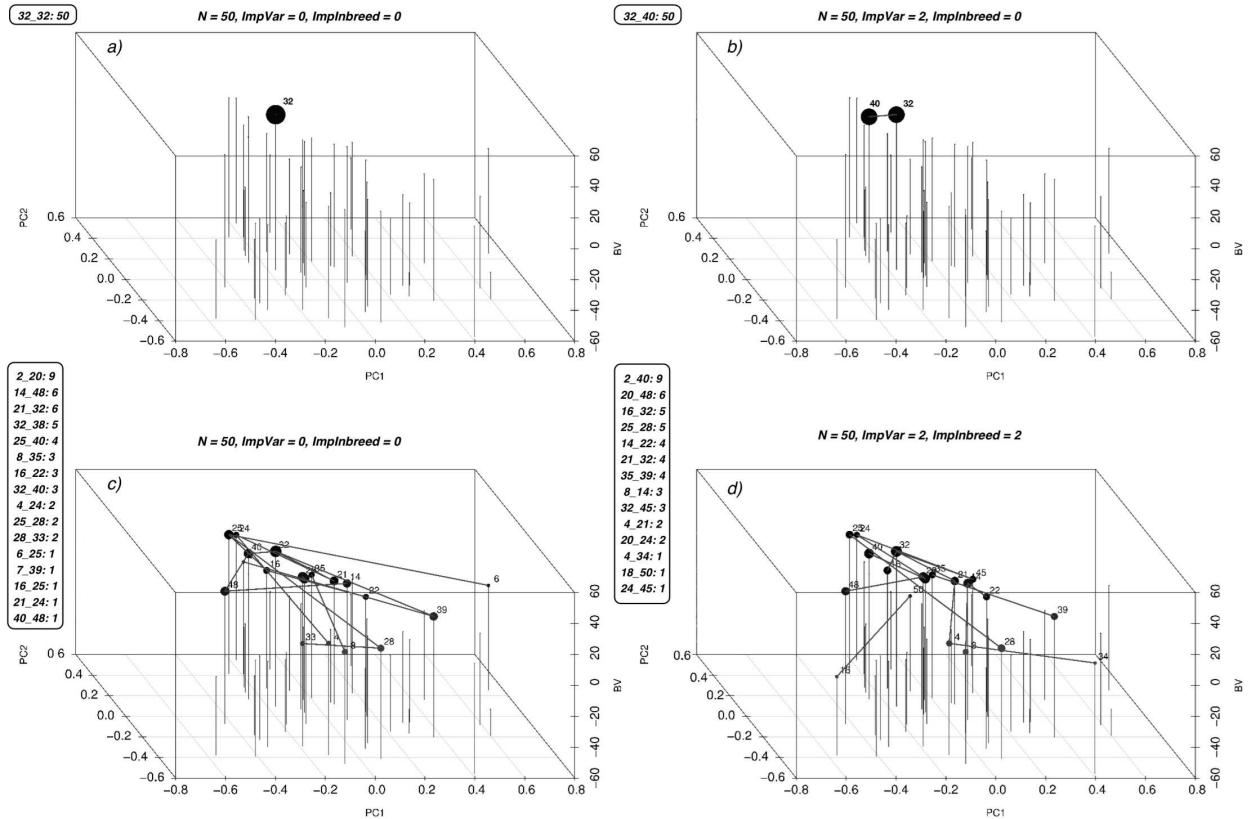
Figure 3 Optimal solutions to the mating problem at a few selected values of λ_1 and λ_2 are in (a), (b), (c), and (d). The list of mates and the number of crosses for each mate is given along the figures. The first two coordinates are used to display the genetic relationships of the lines using the first two principal components, the third coordinate displays the breeding values of the parents. Each parent is represented by a vertical bar. The lines between the vertical bars represent the matings and the size of the points on the bars are proportional to the number of crosses between that parent and any other.
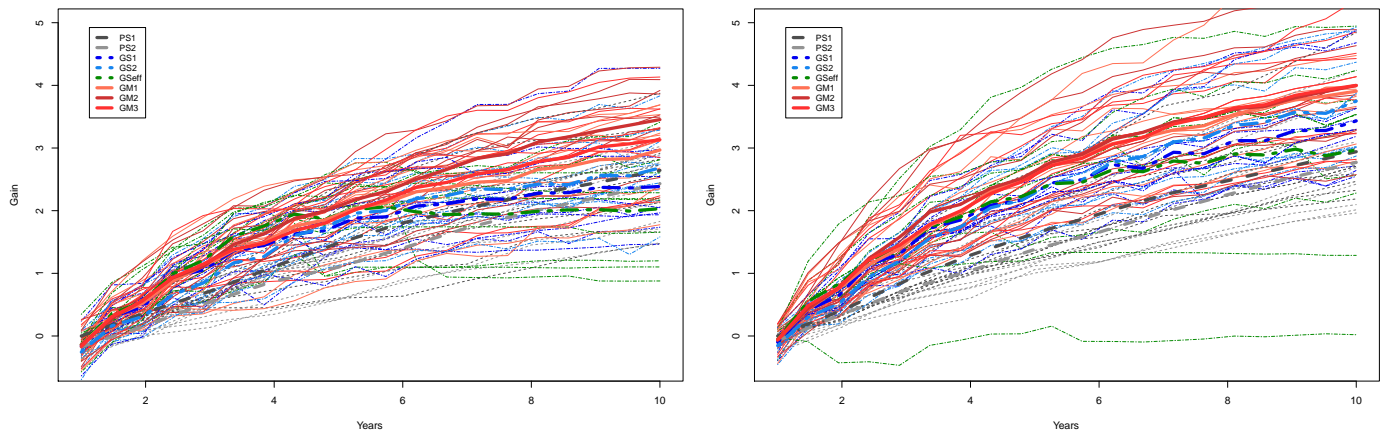
(a) 150 lines

(b) 300 lines

Figure 4 The long term behavior of PS, GS, Efficient GS and GM. Starting from 2 founders we have formed a population of 150 (4a) and 300 (4b) 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). For GS and GM, the marker effects were estimated from data once per year. The results of 10 replication of this simulation with selection intensity 10% (PS1, GS1) and 20% (PS2, GS2) for PS and GS; Efficient GS (GSeff); and GM with $\lambda_2 = 0, 5, 10$ (GM1, GM2, GM3). 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 10 replications. In these simulation studies there is a clear advantage of using GM as a breeding method.

intensity needs to be changed 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.

We have provided several examples and compared our method by simulations to the selection methodologies. We have found 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.

It is possible to adjust the GM methodology to work with either phenotypic records or the BV's when there are no marker effect estimates. Where PS is relatively more efficient than GS, mating using BV's and the marker data of the parents might be beneficial. In this manuscript, we have only considered additive effects. It would be desirable to extend the objective function to include effects and variances related to dominance, heterosis and epistasis. The optimization procedures described in this paper can be used to optimize over a variety of objective functions with hard and soft constraints.

Supplementary File S1 includes the code used for the simulating the data and applying the breeding schemes. Genetic data is simulated using the CRAN package "hypred" (Technow 2014). Mixed model software is publicly available via CRAN (Package EMMREML) (Akdemir and Godfrey 2015). The rest of the software were written using C++ and R.

Literature Cited

Akdemir, D. and O. U. Godfrey, 2015 *EMMREML: Fitting Mixed Models with Known Covariance Structures*. R package version 3.1.

Allaire, F., 1980 Mate selection by selection index theory. Theoretical and Applied Genetics **57**: 267–272.

Berg, P., J. Nielsen, M. K. Sørensen, *et al.*, 2006 Eva: Realized and predicted optimal genetic contributions. In *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13-18 August, 2006.*, pp. 27–09, Instituto Prociência.

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

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

Charcosset, A. and F. Hospital, 1997 Marker-assisted introgression of quantitative trait loci. Genetics **147**: 1469–1485.

Clark, S. A., B. P. Kinghorn, J. M. Hickey, and J. H. van der Werf, 2013 The effect of genomic information on optimal contribution selection in livestock breeding programs. Genetics Selection Evolution **45**: 1.

Daetwyler, H. D., B. Villanueva, P. Bijma, and J. A. Woolliams, 2007 Inbreeding in genome-wide selection. Journal of Animal Breeding and Genetics **124**: 369–376.

Davis, L. *et al.*, 1991 *Handbook of genetic algorithms*, volume 115. Van Nostrand Reinhold New York.

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

Fernández, J., M. Toro, and A. Caballero, 2001 Practical implementation of optimal management strategies in conservation programmes: a mate selection method. Animal biodiversity and conservation **24**: 17–24.

Goddard, M., 2009 Genomic selection: prediction of accuracy and maximisation of long term response. Genetics **136**: 245–257.

Goldberg, D. E., 2006 *Genetic algorithms*. Pearson Education India.

Holland, J. H., 1973 Genetic algorithms and the optimal allocation of trials. SIAM Journal on Computing **2**: 88–105.

Jannink, J.-L., 2010 Dynamics of long-term genomic selection. Genetics Selection Evolution **42**: 35.

Jannink, J.-L., A. J. Lorenz, and H. Iwata, 2010 Genomic selection in plant breeding: from theory to practice. Briefings in functional genomics p. elq001.

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

Kinghorn, B. and R. Shepherd, 1999 Mate selection for the tactical implementation of breeding programs. Assoc Advmt AnimBreed Genet **13**: 130–133.

Kinghorn, B. P., 1998 Mate selection by groups. Journal of dairy science **81**: 55–63.

Kinghorn, B. P., 2011 An algorithm for efficient constrained mate selection. Genetics Selection Evolution **43**: 1.

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

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, *et al.*, 2001 Initial sequencing and analysis of the human genome. Nature **409**: 860–921.

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

Leutenegger, A.-L., B. Prum, E. Génin, C. Verny, A. Lemainque, F. Clerget-Darpoux, and E. A. Thompson, 2003 Estimation of the inbreeding coefficient through use of genomic data. The American Journal of Human Genetics **73**: 516–523.

Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y.-J. Chen, Z. Chen, *et al.*, 2005 Genome sequencing in microfabricated high-density picolitre reactors. Nature **437**: 376–380.

Markowitz, H., 1952 Portfolio selection. The journal of finance **7**: 77–91.

Metzker, M. L., 2010 Sequencing technologies-the next generation. Nature reviews genetics **11**: 31–46.

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

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

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

Quaas, R., 1988 Additive genetic model with groups and relationships. Journal of Dairy Science **71**: 1338–1345.

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

Sonesson, A. K., J. A. Woolliams, and T. H. Meuwissen, 2012 Genomic selection requires genomic control of inbreeding. Genetics Selection Evolution **44**: 1.

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

Technow, F., 2014 *hypred: Simulation of Genomic Data in Applied Genetics*. R package version 0.5.

VanRaden, P., 2008 Efficient methods to compute genomic predictions. Journal of dairy science **91**: 4414–4423.

Wang, J., 2011 Coancestry: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. Molecular ecology resources **11**: 141–145.

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

Zhong, S. and J.-L. Jannink, 2007 Using quantitative trait loci results to discriminate among crosses on the basis of their progeny mean and variance. Genetics **177**: 567–576.