

Chapter 2

Marker Assisted Selection Made Cheap and Easy

H. Riday

Abstract Molecular markers in perennial forages are becoming ubiquitous for marker assisted selection (MAS). Nevertheless, widespread implementation of MAS in perennial forage breeding programs for highly quantitative traits has not yet occurred. A primary reason is likely the cost associated with genotyping plants using molecular markers. Presented are cost-effective MAS strategies that are immediately implementable in most diploid or polyploidy perennial forage breeding programs. Breeding methods developed during the pre-MAS era maximize the ratio between additive variance and the square root of the phenotypic variance (i.e. $h\sigma_A$). With the advent of molecular markers, many new breeding methods based primarily on correlated selection responses between molecular markers and quantitative traits have been proposed (i.e. $r\sigma_A$). MAS strategies should be considered using molecular markers to improve $h\sigma_A$ by enhancing traditional breeding methods. A selection strategy pyramid is envisioned with traditional maternal halfsib selection serving as the base followed by marker assisted paternity selection. With no additional cost within maternal and paternal maximum linkage disequilibrium based MAS strategies can be added. Finally, as the total cost to genotype one individual decreases, residual linkage disequilibrium based MAS strategies can be added to the above mentioned strategies or at some price point supplant these strategies.

2.1 Introduction

In most cases forage varieties available today perform better than varieties developed in the past; much of this increased performance can be attributed to plant breeding. Plant breeding has been part of the human experience since the dawn of agriculture. Starting in the twentieth century, modern plant breeding began exploiting the theory and knowledge of the mode of inheritance to accelerate selection gains. Breeding methods developed during this era typically seek to maximize the ratio between additive variance and the square root of the phenotypic variance (i.e. $h\sigma_A$). With the advent of molecular markers, many new breeding methods based primarily on correlated selection responses between molecular markers and quantitative traits have been

H. Riday (✉)

USDA-ARS, US Dairy Forage Research Center, Madison, WI 53706, USA

e-mail: Heathcliffe.Riday@ars.usda.gov

proposed (i.e. $r\sigma_A$). The advantage of breeding systems built upon correlated selection responses based on molecular markers are that theoretically selection can occur in a non-target environment due to elimination of the phenotypic variation (i.e. σ_P) from the denominator of the selection gain equation and correlation values potentially being up to one (if enough molecular markers are used).

So why have marker assisted selection (MAS) schemes as of yet not been widely incorporated into forage breeding programs, resulting in new varieties? A primary reason is likely the cost associated with genotyping plants using molecular markers. Compared to animal breeding, as well as plant breeding systems involving inbred line development, phenotyping costs in outbred forage species are comparatively much lower. Therefore, MAS schemes are usually not cost competitive, especially in cases where many molecular markers are required. Furthermore, due to the slow pace of genetic gain in highly quantitative traits, such as biomass yield and persistence, it is often difficult to distinguish between elite varieties. This may make it difficult for seed companies to justify increasing costs to accelerate improvement of these traits, as it may be more strategic to invest only enough resources in these traits to remain competitive in the marketplace rather than to excel. Additionally, competitors will almost always be only one selection cycle behind the breeder, making extra cost recuperation for such core quantitative traits difficult, further depressing the incentive to excel. For value-added traits in forages, which are often more expensive to phenotype, MAS schemes become more feasible, especially if premiums for value-added traits can be obtained during seed sales. However, in markets allowing transgenic traits, it is often more beneficial to pursue transgenic approaches, which may have more dramatic phenotypic changes and which usually afford better intellectual property protection and associated ability to recapture costs through premiums.

End users of forages still have an immense interest in core quantitative trait improvement, such as for biomass yield and persistence. In the interest of end users of forage varieties, as well as for the greater good of agriculture, efforts should be made to improve the rate of plant breeding progress. The challenge is to make these improvements in a cost neutral way (i.e., increased efficiency at the same price) so that market-driven plant breeders have an incentive to adopt new methods in order to retain competitive plant breeding programs.

A further challenge and opportunity is the reality of the diseconomy of scale in plant breeding that is a result of the nature of selection intensity (k) (Fig. 2.1).

There is a disincentive to increase selection fractions beyond 1 in 1000 because around this selection fraction, exponential increases in selection fractions increase k -values at sub-linear rates. Therefore, increasing program size is not a cost effective option; rather, selection gains at lower selection intensities need to be increased. Traditionally, molecular markers have been incorporated in schemes to increase the r -values in correlated selection responses in order to eliminate the need for extensive phenotyping. Less attention has been given to the use of molecular markers in increasing selection efficiency, especially in schemes that would necessitate phenotyping. Since forage breeding phenotyping is relatively inexpensive, it makes sense to explore such schemes.

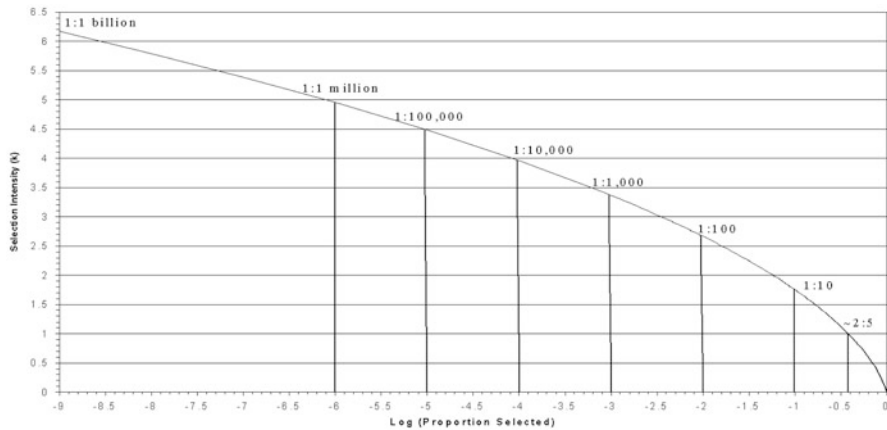


Fig. 2.1 Selection intensity (k) value at various selection proportions (log scale)

2.2 Family Based Selection

In outbred forage species, maternal identity of plants is often known. Paternity, however, is usually unknown. Riday (2011) describes the use of paternity testing to increase selection gains in red clover, a diploid forage legume species. Since the completion of this work, a pilot study was implemented to conduct paternity testing in an autotetraploid alfalfa breeding program. Software was developed and paternal identity of progeny of a 16-parent polycross were ascertained using an exclusion analysis and 16 SSR markers (Riday 2012). These two studies establish that paternity testing is now feasible in any forage species, no matter the ploidy configuration. In most cases, paternity testing should be accomplishable using less than 30 polymorphic molecular markers. Increased selection gain is made by increasing parental control and it contributes to increased efficiency in selection schemes based on $kh\sigma_A$. Riday (2011) describes various ways such information could be incorporated into a breeding program. He suggests that the risk-reward consideration of using this scheme is low while confidence in determining paternity with a set of SSR is very high and genetic gains based on paternal halfsib selection are not affected by any type of marker-quantitative trait linkage considerations.

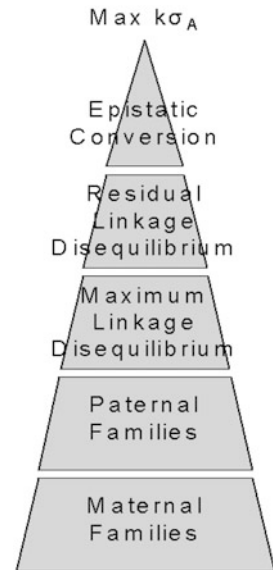
An extension of this strategy could be applied to the less informative situation where potential number and genotypes of parents of a polycross are unknown. Progeny genotypes from a polycross would be clustered into groups based on molecular markers; these groups would then be defined as pseudo-families. Breeding values would then be assigned to pseudo-families and selection would be conducted among the pseudo-families. Computer programs, such as 'Structure' (Pritchard et al. 2000), for example, could be used to cluster genotypes into groups, with genotypes being assigned to each potential cluster at a given probability and multiple runs of 'Structure' used to determine the most probable number of clusters. (It should be

noted that the instructions for using ‘Structure’ say the program should not be used on genotypes with known family structure; however since in this case *some* form of clustering is desired and “true” population structure is not relevant, this violation may be acceptable). This procedure was run on 550 progeny of a 19-mother by 96-father polycross. ‘Structure’ determined that 19 clusters were the most probable number of clusters. There was some correlation between genotype assignment to the 19 clusters and the actual 19 maternal halfsib families of the genotypes. Between pseudo-halfsib family variance in this case was approximately 70 % of the actual known maternal halfsib variance. Due to violations of halfsib family inheritance of additive genetic variance, it is unclear how much, if any, of the 70 % of the variance determined among the pseudo-halfsib families would be predictive of progeny performance in the next generation. If further study shows that some of the between pseudo-halfsib variance is heritable, this would be another potentially quick and easy use of molecular markers to increase selection response.

2.3 Correlated Selection Response Based Selection

With known maternal and paternal identity of plants and with molecular marker information collected to accomplish the paternity test, it becomes a risk free proposition to pursue molecular marker-based correlated selection responses (i.e. even if no linkage is evident between molecular markers and the quantitative traits the molecular markers used for paternity testing). It is currently popular to investigate and describe whole genome selection strategies (Jannink et al. 2010). This is indeed a good strategy, but becomes riskier at lower molecular-marker-genome coverage depending on residual linkage disequilibrium between markers and traits. Such linkage disequilibrium has been reported to be very low in outbred forage species (Isobe et al. 2009). Presented is a potentially less informative correlated selection response strategy that can be used at lower molecular marker density. Rather than selecting on residual linkage disequilibrium (i.e. whole genome selection), maximum linkage disequilibrium is targeted instead, theoretically allowing molecular markers to be used at much lower molecular marker density. The traditional problem with maximum linkage disequilibrium approaches is that trait marker linkage rapidly breaks down at greater recombination distances, making breeding values assigned to specific markers unreliable. However, if the correlated selection response is restricted to distinguishing between the two possible homologues inherited from the same halfsib parent, targeted selection can be conducted at lower risk. Essentially this is a molecular marker-correlated selection response on within halfsib family additive variance. The limitation of this method is that some accompanying phenotypic data is necessary to assign breeding values to heterozygous markers associated with common halfsib parent-derived homologues. Furthermore, the maximum amount of additive genetic variance that can be captured per halfsib parent is one-quarter (i.e. $r1/4\sigma_A$ for the maternally derived marker and $r1/4\sigma_A$ for the paternally derived marker).

Fig. 2.2 Selection strategy pyramid with *maternal halfsib family* selection at the base followed by various molecular marker assisted selection strategies including: *paternal halfsib selection*, *maximum linkage disequilibrium* based within halfsib family selection, *residual linkage disequilibrium* based selection, and finally *epistatic conversion*



2.4 Combined Methodologies

Halfsib seed-derived maternal family information, molecular marker-derived paternal halfsib information, and maximum linkage disequilibrium based within halfsib family correlated selection response information can all be combined into a common strategy using the four orthogonal or independent partitions of the additive genetic variance. Strategically, the four terms can be viewed as building blocks with maternal identity and accompanying breeding values forming the foundation. Additional resources can then be used to determine paternal identity and within halfsib family maximum linkage disequilibrium correlated selection responses (Fig. 2.2).

The advantage of this strategy is that, depending on the availability of breeding resources and the perceived value of a particular polycross, the resources commitment can be increased or decreased with some additional selection gains realized at each increased level of resource commitment.

If a few molecular markers are known *a priori* to be very close to quantitative trait loci or specific genes such that residual linkage disequilibrium-correlated selection response strategies can be pursued (i.e. $r\sigma_A$), then these could be added to a combined selection strategy in combination with maternal halfsib selection, paternal halfsib selection, and within maternal and paternal molecular marker-based halfsib selection. Based on the cost of each strategy and amount of additive variance described for each selection strategy, optimization equations could be written to pursue the most efficient selection strategy on a cost and selection gain basis. Obviously as genotyping costs decrease and molecular marker densities increase the breeding strategy would move more and more towards whole genome selection. A critical point to consider is that

before committing from the start to a potentially high cost whole genome selection strategy one should consider that there are many molecular marker assisted selection strategies in conjunction with traditional family based selection strategies that can be used in stead of or in conjunction with whole genome selection strategies that may be more cost effective or strategic to achieve improved selection gains (Fig. 2.2).

2.5 Conclusions

A final caution needs to be discussed in relation to the use of molecular marker assisted breeding strategies. Molecular markers used to describe family structure or used to identify correlations with traits under selection are *always* descriptive of genotypic potential; molecular markers *never* determine a plant's genotypic potential. A plant's genotypic potential is always determined during the random genome reshuffling of parental genomes during reproduction (unless transgenic approaches are taken). This means that the perfect genotype will only be observed rarely, irrespective of how many molecular markers are used to identify it (i.e., even if the perfect theoretical genotype is known *a priori*, it still has to be created via random mating and physically identified from among all products of random mating). The core process of extracting DNA from a genotype is not drastically decreasing in cost. Therefore, from a practical perspective, forage breeders will still be evaluating a limited number of selection units in their programs (i.e. 100–100,000). It is therefore unnecessary to have an unlimited number of molecular markers for selection among a smaller number of genotypes since only a certain amount of molecular marker information will be necessary to rank selection units, with less information needed for accurate ranking among fewer genotypes.

It should be apparent that in the combined molecular marker strategy described above, the greatest amount of information per molecular marker will be obtained from an initial set of markers. However, as the number of markers utilized is increased, the amount of information per marker at some point will decrease. The number of markers utilized in a selection strategy should be optimized so the maximum amount of resources can be applied to increase the number of selection units under evaluation. Even under the perfect correlated selection response scheme (i.e. $k\sigma_A$) selection intensity and the amount of additive variance present are still paramount drivers of selection gain.

References

- Isobe S, Köllicker R, Hisano H, Sasamoto S, Wada T, Klimenko I, Okumura K, Tabata S (2009) Construction of a consensus linkage map for red clover (*Trifolium pratense* L.). BMC Plant Bio 9:57
- Jannink J, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Briefings Funct Genomics Proteomics 9:166–177

- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Riday H (2011) Paternity testing a non-linkage based marker assisted selection scheme for outbred forage species. *Crop Sci* 51:631–641

Breeding strategies for sustainable forage and turf
grass improvement

Barth, S.; Milbourne, D. (Eds.)

2013, XII, 392 p., Hardcover

ISBN: 978-94-007-4554-4