

# Molecular Markers in the Seed Industry

Dr. Clay Sneller<sup>1</sup>

## Introduction

The concept of using genetic markers in genetic studies is not new, having been first used 80 years ago (Mendelian genetics was rediscovered only 102 years ago). But the use of genetic markers was very limited due to poor technology until 20 years ago. That has changed dramatically in the past 10 years. For example, 10 years ago there were perhaps 200 markers on the soybean map, and these markers were difficult for a breeder to use. Today there are thousands of markers for the soybean genome and thousands more will be added in the next few years. Similar examples could be used for all major crops. Use of genetic markers now dominates much of the basic plant breeding research and is finding increasing application in commercial variety development. The driving force behind this is the advancement of high-throughput molecular technology and its application to markers. My objective is to present how these changes might impact the seed industry.

## Concept of using Markers in Genetics

It is beyond the scope of this talk to cover the concepts of marker analyses in any detail. Those familiar with breeding theory likely already know, and those who do not know the theory already, do not need to know details to understand what markers can do. For the non-geneticist, markers are tags (signs, identifiers) that identify pieces of chromosomes. The presence or absence of a marker (and thus the chromosome piece associated with it) can be determined by performing laboratory assays.

Markers can be aligned by their order along a chromosome to form a map of the chromosome (See Figure 1). For example, a marker may reside on chromosome 1 of maize and can be used to follow chromosome 1 in genetic studies if the parents of the study have different forms of the marker. By itself, this is of no practical interest, but can be very powerful when we also measure valuable traits. For example, we cross a resistant parent and a susceptible parent and develop progeny. We would collect disease resistant data and marker data on all progeny. If the progeny with marker genotype “MM” from the resistant parent are on average more resistant than the progeny with marker genotype “mm” from the susceptible parent, then we can conclude that there is a gene on chromosome 1 that controls disease resistance (Figure 2). Elaborate statistics are used in such analyses, but the basic idea is as simple as described.

---

<sup>1</sup> Department of Horticulture and Crop Science, The Ohio State University/OARDC, 1680 Madison Av. Wooster, OH 44691, 330-263-3944, sneller.5@osu.edu

Marker studies can reveal many important aspects of crop genetics:

- Location of genes controlling a trait
- Number of genes controlling a trait
- Identify genes that affect multiple traits
- Determine interactions among genes
- Estimate the effect of the genes
- Determine the relationship among beneficial genes in different cultivars
- Estimate the genetic relationship among cultivars and species
- Structure of chromosomes

This knowledge has been relatively easy to determine without markers in the past for traits controlled by one (or few) gene, such as resistance to some diseases that are easy to evaluate. Marker techniques are making analysis of simple traits more practical. But many traits that are essential for developing varieties are not simply inherited. Several (or many) genes control traits such as yield, adaptation, stress tolerance, quality, and resistance to many diseases, and the expression of these traits can be greatly influenced by the environment. The genetics of these complex traits have been essentially unknown until marker analyses were readily available in the early 1980s.

The genetic knowledge listed above could greatly improve the efficiency of breeding for any trait. Of course efficiency can only be improved if the information is accurate. Accuracy of marker studies depend on the following:

- Number of markers used in the study as more markers increase the accuracy of locating genes and estimates of their effects
- Number of progeny in the study (more is better)
- Accuracy and relevancy of trait data (number of replications, environments, representative value of environments, genetic backgrounds)

Until very recently though, the number of markers available has been very limited and genotyping was expensive, so few progeny were included in studies. Despite these shortcomings that limited accuracy, great effort was expended on marker studies of many traits in many species in the past 20 years. Tremendous amounts of data and hypotheses were presented. Many of the findings to date are probably not very accurate, but do provide at least a fuzzy estimate of the genetics of many traits, and a clear estimate of the genetics of some simply inherited traits. The results to date have forever changed quantitative genetics and breeding. In addition, the results from the past 20 years have generated tremendous talk about using marker information to improve selection.

### **Concept of Marker Assisted Selection (MAS)**

Markers can improve selection of superior cultivars if they are reliably associated with beneficial genes. If markers can be strongly associated with genes controlling a trait, then progeny from crosses can be screened for the presence of the desired markers instead of screening the progeny for the trait itself. An example can be drawn from Figure 2, any progeny derived from crosses with the

resistant parent that have the marker genotype “MM” would be assumed to be resistant and would be selected; progeny with the “Mm” and “mm” genotypes would be discarded.

In theory, selection based on markers can be more efficient than traditional selection based on trait data for traits with complex inheritance such as yield (controlled by multiple genes, and/or greatly affected by environment). MAS involves simple laboratory assays that can be more accurate and less expensive than traditional trait assays. MAS can also be cost effective for simply inherited traits when trait assays are expensive. For example, in the USA many companies use MAS to select for soybean cyst nematode resistance (SCN). Greenhouse assays for SCN resistance are very reliable but cost perhaps \$5.00 per progeny while MAS for SCN resistance may cost as little as \$0.20 per progeny. A system for using MAS to improve SCN resistance has been patented in the USA.

### **Markers in Current Variety Development**

With so much promise, MAS has seemingly become the holy grail of plant breeding. Most marker studies to date report on the location and effect of genes and conclude that the markers associated with these genes would be suitable for MAS. How much of this promise has been realized to date? Unfortunately, very little, especially when we consider complex traits, the very traits where MAS should have the greatest impact in theory. Few (none?) breeders, commercial or public, use MAS for improving key traits such as yield, stress tolerance, or crop quality. There are two main reasons for this and both relate to the ability of past basic experiments to predict future breeding performance. First many markers associated with these complex traits in one environment are not associated with the trait in another environment. Thus selection for these markers does not guarantee improved performance in future environments. This is directly related to the well-known phenomenon of cultivars doing well in one environment but not another, called genotype x environment interaction.

The second reason why MAS has not been extensively applied to breeding for complex traits is that the trait value associated with many markers can be very population specific. For example marker “M” from parent 1 may be associated with increased yield of progeny when parent 1 is crossed to parent 2, but when parent 1 is crossed to parents 3, 4, or 5. Thus there are a limited number of populations where MAS for marker “M” would work and it is difficult (impossible?) to predict which populations these are. The problems associated with using MAS for complex traits may be partly overcome as high throughput genotyping for markers becomes more widely adapted. This will allow for more predictive basic studies to be conducted.

While MAS for complex traits has not really taken hold, MAS for traits controlled by few genes is common and having a large impact on breeding. The genes controlling these traits are easy to tag, even with relatively few markers, and the markers tend to be associated with such traits in most environments and predictable genetic backgrounds. For an example drawn from Figure 2, progeny from most crosses of the resistant parent with any susceptible parent, selected for having marker genotype “MM” are likely to be resistant in most environments. There are many important traits that are controlled by single (or very few) genes, particularly disease resistance, and large seed companies are using markers extensively in selecting for these traits. Several seed companies in the US routinely use MAS to select soybean lines with major genes controlling resistance to SCN,

phytophthora root rot, and brown stem rot. These companies will generate millions of marker data points per year on soybean alone.

### **Cultivar Identification**

Issues of ownership have become very important in the modern seed industry. Industry is investing millions of dollars in developing new cultivars, hybrids and technologies and wants a return on this investment. This is achieved in part by releasing more products and protecting them with greater stringency. It is common now for seed companies to seek utility patents on inbred cultivars. Markers can be used to prove the uniqueness of a cultivar, a characteristic that is essential in protecting it with PVP or patents. Until recently, considerable work was required to show that one cultivar was unique from another using morphological traits or old marker systems. For example in wheat, the old marker technology of RFLPs could barely discern one cultivar from another, a task that is now quite easy with SSR and AFLP markers. The abundance of new easy to use, highly polymorphic (e.g. readily differentiates lines) markers are making markers a common component of cultivar identification. In fact, collecting marker data can now be easier than assaying many of the morphological traits that have been used in the past, traits that did not always clearly separate two cultivars. Marker data is presented in Figure 3 that shows that maize line B73 is different from 10 other lines. Lines A632, H99, GT119, and F1 appear nearly identical and additional markers would be needed to distinguish among them.

It is common to use multiple markers to distinguish among lines. The number of markers needed to reliably differentiate two cultivars varies by species and marker type. Currently, most breeders are using simple sequence repeat (SSR) markers because they are easy to use, plentiful for most species, and are quite polymorphic and informative. In wheat and soybean, SSR markers are nearly twice as informative as the older RFLP markers, and perhaps 10 times less expensive. Estimates of the utility of markers in distinguishing between lines vary by species as some species inherently have more marker differences. Any comparison of two species is imprecise, though the data in Table 1 suggests that US soybean and maize have more differences (thus easier to tell one cultivar from another) than rice and European wheat cultivars.

### **Future of Molecular Markers in the Seed Industry**

We have seen that markers are routinely used by the seed industry for simple traits and identifying cultivars, hybrids, and seed lots. While markers and MAS have not been very useful for complex traits, better days may lie ahead. The shortcomings of past marker technology has limited application of MAS to breeding for complex traits. We simply have not had the genotyping capabilities to use enough markers, on enough large populations that are relevant to breeding to provide precise estimates of marker effects for breeding. This limitation is quickly disappearing with the advent of genomics and high-throughput genotyping.

Marker discovery is increasing at a tremendous rate due to the ability to clone and sequence (even partially) tens of thousands genes. The same tools used for mass-cloning and sequencing of genes can be applied to discover a new and abundant class of markers called single nucleotide polymorphisms (SNPs) as well as numerous SSR markers. These markers, and others being

developed, will be particularly useful as they are very amenable to automated, non-gel based genotyping. Automated genotyping involves the use of large-scale sequencers, robotics, and computers to greatly mechanize data generate, collection, and analysis, thereby increasing throughput and lowering costs.

Industry is heavily investing in genomics and its application to using markers in breeding. The technology requires very expensive equipment and highly skilled people, facts that may make this technology beyond the reach of small companies and some university programs. But genomics technology will allow for the execution of the types of studies that are needed to successfully apply information from basic marker studies to broad improvement of crops. This may allow for realization of the full potential of molecular markers to improve crops.

### References

- Bohn, M., H.F. Utz, and E. Melchinger. 1999. Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs, and SSRs and their use for predicting progeny variance. *Crop Sci.* 39:228-237.
- Diwen, N., and P.B. Cregan. 1997. Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theor. Appl. Genet.* 95: 723-733.
- Lu, H., and R. Bernardo. 2001. Molecular marker diversity among current and historical maize inbreds. *Theor. Appl. Genet.* 103:613-617.
- Manifesto, M.M., A.R. Schlatter, H.E. Hopp, E.Y. Suárez, and J. Dubcovsky. 2001. Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. *Crop Sci.* 41:682-690.
- Donini, P., J.R.Law, R.M.D.Koebner, J.C. Reeves, and R.J. Cooke. 2000. Temporal trends in the diversity of UK wheat. *Theor. Appl. Genet.* 100:912-917.
- Plaschke, J., M.W. Ganal, and M.S. Röder. 1995. Detection of genetic diversity in closely related bread wheat using microsatellite markers. *Theor. Appl. Genet.* 91:1001-1007.
- Mackill, D.J., Z.Zhang, E.D. Redoña, and P.M. Colowit. 1996. Level of polymorphism and genetic mapping of AFLP markers in rice. *Genome* 39:969-977.
- Roder, M., V. Korzun, K. Wendehake, J. Plascjke, M-H. Tixier, P. Leroy, and M.W.Ganal. 1998. A microsatellite map of wheat. *Genetics* 149:2007-2023.

Table 1. Genetic similarity among elite cultivars and inbred lines based on SSR markers. Genetic similarity is the probability that a marker is identical between two lines. Genetic similarities values range from 0 (no markers in common between two lines) to 1 (all markers are common between two lines: lines are identical).

Crop	Genetic Similarity	Reference
US Soybeans	0.36-0.42	Diwan and Cregan, 1997
Maize, Modern USA inbreds	0.35-0.38	Lu and Bernardo, 2001
Rice, japonica	0.64	Mackill et al., 1996
Wheat, Argentina	0.29	Manifesto et al., 2001
Wheat, Europe	0.44	Plaschke et al., 1995
Wheat, Germany and Austria	0.57	Bohn et al., 1999
Wheat, UK, 1990s	0.47	Donini et al., 2000

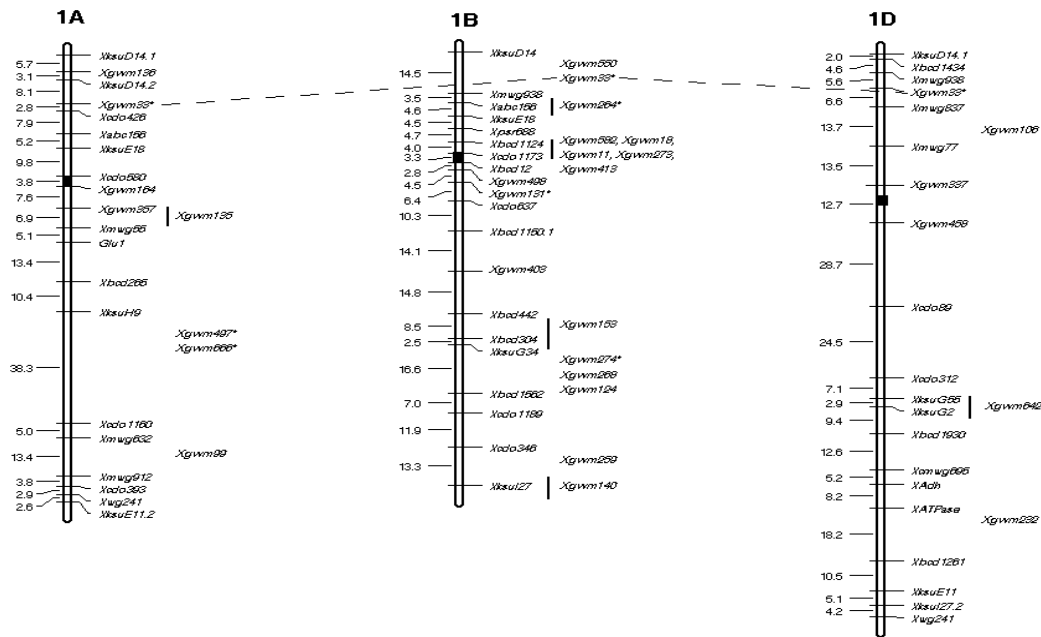


Figure 1. Marker map of three wheat chromosomes (1A, 1B, and 1D). Lines cutting across the chromosomes show the position of the markers (identified as Xgwm). The numbers to the left of each chromosome indicate the relative distance between markers. From Roder et al., Genetics149:2007-2023.

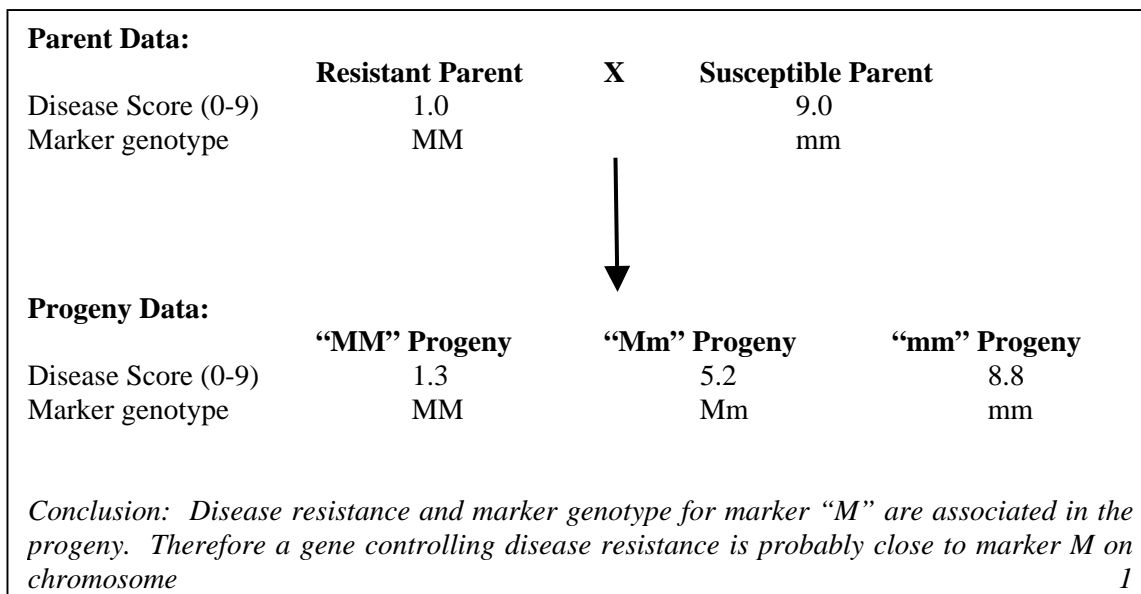


Figure 2. Diagram of a genetic study to test the association of disease resistance with a marker on chromosome 1. The resistant and susceptible parents were crossed. The progeny from the crosses were tested for disease resistance and genotyped to see if they had markers only from the resistant parent (MM), only the susceptible parent (mm), or both. The progeny with only markers from the resistant parent (MM) had an average disease score of 1.3 while those with markers from the susceptible parent (mm) had an average disease score of 8.8.

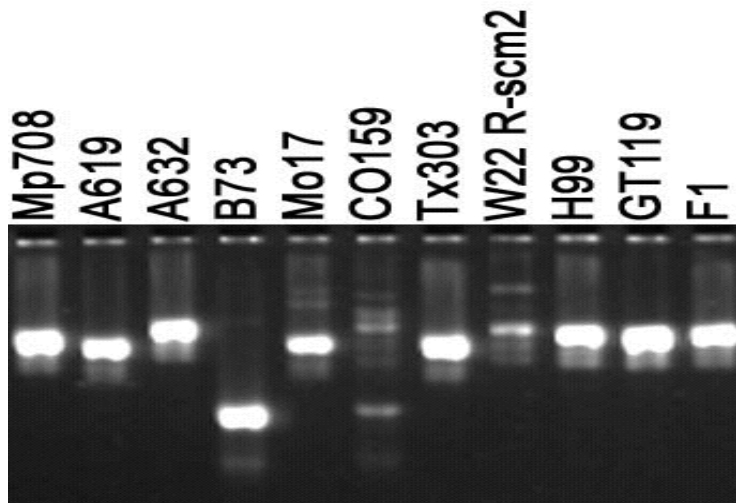


Figure 3. Simple sequence repeat marker genotype of 11 maize lines. At least four different marker genotypes are apparent among the 11 lines.