RosBREED's Community Breeders' Page

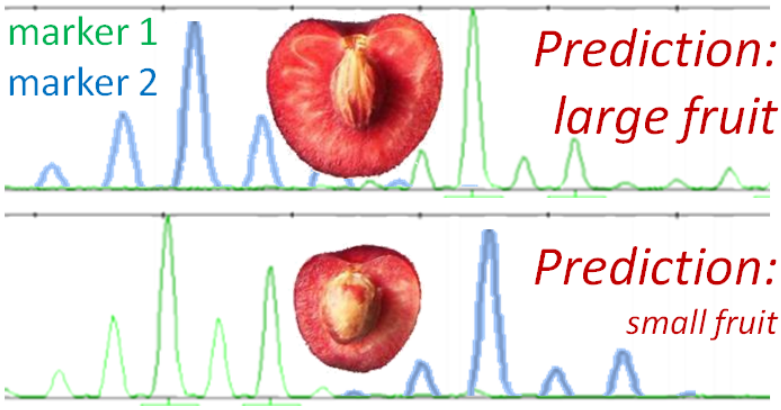
Routine DNA Testing

By Cameron Peace and Nahla Bassil, MAB Pipeline and Genotyping Team Leaders

Routine DNA testing. It's done once you've MAB Pipelined promising QTLs within your own breeding program and thereby established the performance-predictive power of each DNA test for your germplasm under your conditions. By then you are ready to screen your parent pool and/or various seedling families to gain insights into genetic potential that help with your crossing and selection decisions.

What is a DNA test?

A DNA test is one or more predictive DNA markers* *that must be run together* to provide a sensible prediction about the genetic potential of any individual screened. For example, predicting apple fruit storability by targeting the *Md-ACS1* gene is currently a DNA test that uses just a single marker. The DNA test for sweet cherry fruit size prediction relies on simultaneous genotyping with two markers flanking a valuable QTL.



* where such markers are usually simple PCR-based simple sequence repeats (SSRs), sequence-characterized amplified regions (SCARs) that especially target insertions-deletions (indels), or single nucleotide polymorphisms

But how are DNA tests run routinely? Mostly when we refer to DNA tests we're talking about locus-specific markers predictive for particular traits - that is, targeting one trait locus at a time. For example, in the Washington State University apple and sweet cherry breeding programs, thousands of seedlings each year (and dozens of parents) are routinely screened with locus-specific SSRs and SCARs for superior texture, storability, fruit size, and other traits. For seedling selection, the most economical way to conduct these tests is to run the most stringent DNA test on a family, cull the seedlings predicted to be inferior, and then run the next test on the remaining seedlings. But genotyping costs are dropping to the point where the multi-locus approach of SNP mini-array genotyping is becoming attractive.

What are SNPs?

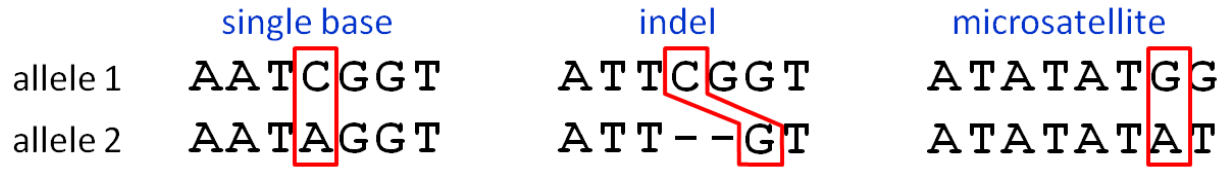


Figure 1. All DNA sequence polymorphisms can be considered as SNPs

SNP (pronounced "snip" or less commonly "S.N.P.") stands for single nucleotide polymorphism, an allelic difference of a DNA base at a locus. Any DNA sequence polymorphism can be considered to be (and detected) as a SNP (Figure 1), so there is nothing inherently unusual about SNPs at the DNA sequence level other than that they are very generic. The practical difference between SNPs and other marker types is simply how their polymorphism is detected. Actually, SNPs can be detected by the same methods as SSRs (capillary or gel electrophoresis or high-resolution melting). But the buzz about SNPs is because their inherent simplicity allows efficient, high throughput streamlined detection. While one SNP usually carries less genetic information than an SSR (a general rule of thumb in our crops is that 5 SNPs = 1 SSR in information content), running 9000 peach SNPs with the International Peach SNP Consortium peach SNP array v1 is much, much easier and cheaper than running even 1000 SSRs.

Why would more than one SNP target a given trait locus? If there are only two functional alleles (contrasting for their performance prediction), then a single SNP can be enough. A second SNP at the locus might be included for safe redundancy. But many trait loci are revealing themselves to have more than two functional alleles - for example, one conditioning a high level of the trait, one medium, and another low. Multiple SNPs are then used to distinguish among these functional alleles by revealing the distinct multi-SNP haplotypes associated with each SNP (Figure 2).

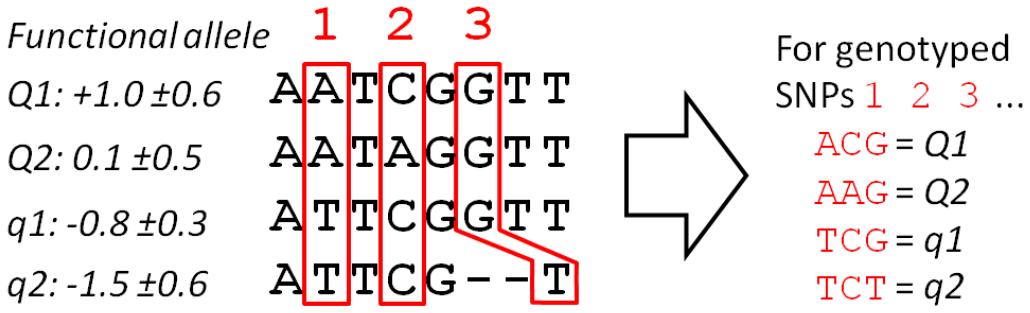


Figure 2. An example of a set of three SNPs targeting one trait locus to distinguish among four functional alleles

A SNP *mini-array* is a locked-in set of SNPs numbering in the dozens to hundreds (but not in the thousands – that would be an *array* that loses the "mini"). Effectively, the whole mini-array is a single DNA test. A SNP mini-array will target multiple trait loci usually with multiple SNPs targeting each trait locus. For example, a 24-SNP mini-array might be established for six trait loci where each locus is marked with an average of four SNPs. A major *disadvantage* of SNP mini-arrays over individuals SSRs (or SCARs or even individual SNPs) is lack of versatility. Because all the SNPs included are locked in, they all must be run on each breeding individual tested and therefore all of the SNPs must be paid for whether they are expected to be informative or not. For example, while one family might be segregating for all six trait loci targeted by the mini-array, the next family might only segregate for three of them. And the seedlings can't be whittled down by the first few SNPs

so that only the remaining seedlings are run with the next SNPs. It's all or nothing. As a minor consolation, even the unnecessary information paid for may be useful in confirming homozygosity or parentage. Another concern is that DNA extract quality might need to be substantially better than that acceptable by PCR-based genotyping approaches. We are currently investigating whether SNP mini-array and array genotyping can accept the quick and dirty (and convenient) extraction methods that currently allow DNA extraction for thousands of samples to be a relatively painless procedure.

The major *advantage* of mini-arrays is their simplicity in some situations. Those situations are when informative SNPs are known for numerous valuable trait loci (a situation that coordinated RosBREED efforts in peach, apple, and cherry have propelled us into) and when one or more breeders have many individuals they want screened. Because of the way commercial services are set up, the higher the number of individuals to screen the cheaper the cost per individual. So it pays for multiple breeders to coordinate their needs - as long as they agree on a generally informative set of SNPs comprising a mini-array. For example, currently the Demonstration Peach Breeders of RosBREED are developing two 24-SNP mini-arrays - for proof of concept of this approach to routine DNA testing as well as to gain some valuable DNA information for the participating breeders. These two mini-arrays will together target QTLs known, recently discovered, or anticipated for bacterial spot, fruit size, sweetness, maturity date, acidity, softening characteristics including slow melting flesh, blush, and flesh browning. This practical plan was hashed out during last month's Peach Breeder Workshop.

Why don't we use 23 SNPs for each mini-array? Or 20? Or 10? Currently, 24 is the minimum multiple offered by U.S. commercial service providers. Current methods for assaying a smaller number of SNPs (<24) are not cost-efficient. We expect that technological advances over the next few years will increase the cost efficiency for maximum versatility. However, there is a strong case to be made for Rosaceae breeders to develop mini-arrays with *more* SNPs …

As long as genotyping costs for dozens to hundreds of SNPs are at affordable levels for seedling screening (i.e., a total of several dollars per seedling), which will be more readily achievable if commercial service providers are approached with orders for single mini-arrays to be screened with many thousands of samples, larger mini-arrays make more sense logistically. The mini-array design (choosing the specific SNPs) and downstream data processing for intelligible breeding- relevant performance predictions can be efficiently conducted at a central coordination point - much like how RosBREED is coordinating various analytical steps for multiple Demonstration Breeders. Imagine a SNP mini-array targeting all known QTLs for high-value traits for a crop, costing just a few dollars per individual tested, that all breeders of that crop could take advantage of by sending their tissue samples to a central receiver and then soon after obtaining translated results of functional genotypes. The SNP composition of such a mini-array could be updated regularly, perhaps once a year, to reflect new QTL discoveries and agreed changes in trait locus priorities.

Which brings us to routine DNA testing of parents and advanced selections. We already have SNP arrays for peach, apple, and cherry each with thousands of SNPs, developed and evaluated recently in international partnerships coordinated by RosBREED and now being widely used in many research projects including RosBREED's QTL discovery and validation efforts involving Demonstration Breeding Programs. While these arrays are much too expensive (at ~$70-$110 per individual) for routine DNA testing of seedlings, they are quite affordable for those few individuals that breeders are most interested in: parents and advanced selections. Thanks to coordinated QTL characterization using reference sets of U.S. breeding germplasm, the genotypic data coming out of the use of these arrays is providing a wealth of performance-predictive information. We expect to be mining more and more practical breeding knowledge from these datasets for years to come. Several U.S. breeders have so far seized the opportunity to gain this relatively cheap information on their critical germplasm and connected with RosBREED's central analytical expertise. It's not too late as these arrays are still commercially available until the end of the year. In the future, we expect that refined composition of these thousands-of-SNP arrays and the practical DNA information they provide, with streamlined centralized sample and data processing, will make SNP array use on parents and advanced selections an extremely compelling routine operation for all Rosaceae breeders.