# Marker Assisted Selection: What MAS May Do for the Future of Cannabis Breeding

By John Brunstein

## A Quick History Lesson

Humans have been engaged in selective breeding of plant species – that is, the fostering of human desired traits and the suppression of undesired ones – for millennia. For corn, one of the best studied examples, published evidence suggests Mesoamerican farmers began selectively replanting the seeds of the best examples of a grass with 5-10 hard, not very tasty grain kernels per ¾" cob around 7,000 years ago. After a mere 3,000 years, average cob sizes were a whopping 1", meaning something a bit over 25-percent more yield per cob. Things picked up speed a bit after that and modern corn cobs yield around 1,000x the food per cob compared to the original wild plants and have gained a host of other desirable attributes as well (softer, faster cooking ears, higher sugar content). Corn is one of the modern world's most important starch crops, a role only made possible by all of the diligent work of breeders across that time span. Lest we think this is a one-off story, it could be repeated on similar lines for apples, rice, and many other common agricultural crops.

Cannabis is estimated to have diverged from its closest cousin, hops (of beer fame), at around 28 million years ago, somewhere roughly in the region of modern day Mongolia, and we have archaeological evidence of its use by humans – probably initially only for fiber and oil – more than 4,000 years ago in China. There are some claims in the literature of evidence of ritual burning of cannabis going back to earliest usage, but analysis of the associated residue in these sites suggests these plants contained too little psychoactive compounds (THC, THCV) to have had any noticeable effect on those inhaling the smoke. By 2,500 years ago, however, we have good evidence that at least in parts of western China, cannabis plants with significant levels of THC were being selected and burnt in enclosed spaces in funerary rituals. Whether this was primarily for psychoactive results or for the pungent aroma as a means to mask the smell of decaying corpses is currently debated, but it seems quite plausible selective breeding for aroma might have happened first with discovery of psychoactive properties coming later. In any case, it was around this time that there appears to have been spread of the use of cannabis as a drug from the Far East through the Middle East and into eastern Europe.

## The Problem with Selective Breeding

The first moral of this history lesson is that

Photo: iStockphoto.com/jxfzsy

cannabis, just like any other plant crop, can be subjected to selective breeding to alter its chemotypic profile to something more desirable to humans. The second lesson is that if you're just doing this by propagating seeds of the "best" plants from a crop cycle after cycle, it's a long process. Today's cannabis breeder isn't interested in diligently propagating plant lineages a few thousand years to get a desirable novel cultivar! Of course, they also have the benefit of a better grasp of biology, meaning selective manually-directed crossing between varieties each having some desirable traits is now the norm, followed by growth and assessment of progeny for best combinations of parental traits. These small number of progeny are then normally backcrossed and/or self crossed in an effort to "fix" the phenotype (make it reliably appear in all, or at least most, progeny). From a genetics perspective, what's actually being done in fixing is to remove heterozygosity and in effect make relevant parts of the resulting genome all either maternally or paternally derived (depending on which parent of the cross provided the desirable gene form). In other circles we call this "inbreeding," and it's generally frowned upon, because it often also leads to emergence of diseases and lack of health. Having diversity in one's genes leads to something biologists call 'hybrid vigor', which is a good thing; generally, there is a balance between amount of backcrossing/fixing, and plant viability.

Compared to a multi-millennia timescale, this more modern directed crossing, selection, and fixing is orders of magnitude faster – but it can still take years of diligent effort to grow hundreds or thousands of cross seedlings to maturity, assess each for properties, select the best, and then repeat as needed to optimize genetic stability versus health. In addition to the time investment, there's actual resource costs as well (grow space, lighting, fertigation, and the like) for all of the plants being screened. As the majority of these will end up being discarded in favor of the few selected progeny, that's effectively all wasted resources. In other words, on a modern business timescale, traditional selective breeding programs remain both costly and slow.

### Enter "Marker Assisted Selection"

To address this, we can look to apply molecular biology techniques. Well established in other aspects of agricultural breeding programs, Marker Assisted Selection (MAS) works on a relatively simple principle. Many physically expressed traits (phenotypes) are influenced by variant sequence forms (alleles) of single genes (we call these monogenic traits, as opposed to polygenic traits where multiple genes interact to create the phenotype). Cannabis is normally a diploid organism, meaning it has two copies of each chromosome (one from the father and one from the mother), and thus two copies of each gene.

For sake of argument, let's imagine there's a cannabinoid called CBX, and it's produced by an enzyme called CBX synthase from CBGA. There are two alleles of the CBX synthase gene: CBX-H (it's a fast, efficient enzyme, which produces a lot of CBX); and CBX-null (this is an inactive form of the enzyme, which toothlessly gums on CBGA but doesn't catalyze any CBX formation). These two alleles vary from each other in only one amino acid, meaning their respective DNA gene versions each has a single distinct nucleotide difference from the other. Now, we don't start knowing any of this – what we do know is among all of our cannabis cultivars on hand, we find some have a lot of CBX, some have about half that amount, and some have none. What we can do through DNA sequencing is uncover that the varieties with two copies of CBX-H (annotated as CBX-H/CBX-H) are the high CBX ones; the varieties with one copy of each allele (CBX-H/CBX-null) are the "half CBX yield" varieties; and the CBX-null/CBX-null varieties, as expected, produce no CBX.

Photo: iStockphoto.com/zokru

Now that we know these particular gene forms are responsible (or "markers") for certain phenotypes, what can we do with this information? Imagine you have a cannabis variety "Purple Space Monkeys" which has great characteristics, but produces no CBX, and you'd like to breed a new variety CBX Space Monkeys which is pretty much like Purple Space Monkeys except it has high CBX content. With traditional breeding, you'd cross a Purple Space Monkey with some CBX-containing variety, get thousands of seeds, then spend time growing them up and looking at phenotype in hopes of finding one plant with the right combination. Of course, to do that you'd have to grow them all the way through flowering and pay for chemotypic analysis on each; slow and costly. Where MAS can assist would firstly be in selecting the CBX-containing parent – you'd sequence the two alleles to confirm it's a CBX-H/CBX-H plant as opposed to a CBX-H/CBX-null. This will, as a first step, ensure that all of your progeny – the F1 generation – will be CBX-H/CBX-null, since we know Purple Space Monkeys is CBX-null/CBX-null and they had to get one CBX-H copy from the other parent. That's already better odds than if you'd blindly used what was a CBX-H/CBX-null parent, where only 50 percent of the F1 progeny would express any CBX.

("Aha, but I would obviously have used a high CBX parent," you say. Yes, in this perfect imaginary scenario that would have told you it's CBX-H/CBX-H but reality is never so clear cut. Variable penetrance and variable gene expression and things called epistatic effects affect the real world, so this hard knowledge that the parent is homozygous for the allele wanted and all progeny will carry one copy of this gene, is very useful).

Now, if you're happy with a mid-level of CBX expression, you can proceed to pick any one of these clones to propagate onward as your new variety; but what if you want high CBX levels? Now you'll want to start crossing F1 x F1 progeny (both CBX-H/CBX-null), and this is where MAS really begins to get helpful. Only one quarter – 25-percent – of these F2 progeny will be CBX-H/CBX-H. Without MAS, you're stuck doing the cross and growing up, let's say, 1,000 progeny for flowering and chemotyping. With MAS, once you have little sprouts of plants, you can sample a tiny piece of each and immediately detect which 250 are the progeny of interest worth keeping; you just saved all of the space, trouble, and expense of growing the 750 plants you know won't be high CBX. If you want to also test for a gender – another simple example of MAS – you can cull a further 125 male plants leaving only the females to follow. You've cut your work (and propagation and phenotype testing expenses) down by a factor of 7/8.

## MAS in Practice

So, is that how MAS works in real life? Well, sort of. It's presented above as an overly simplified example with perfect numbers, just like Gregor Mendel's original work*. In cannabis, the so-called THCA synthases and CBDA synthases both, really, seem to produce a mixture of products with minor allelic variations (changes in single amino acids in the protein sequence) changing the yield ratios. MAS will have the ability to distinguish all of these and their combinations in potential parents where simply looking at the THCA/CBDA levels can't, giving better insight into what crosses to set up and what allele forms to track in progeny to get desired results with regard to these major cannabinoids. Similar data is being uncovered for the complex web of terpene synthases, many of which as well are capable of producing multiple products. Many other traits such as resistance to particular pests will likely be amenable to MAS. By combining examination of multiple trait markers in a single cross, MAS becomes increasingly useful. Want a particular combination of multiple specific traits split between two parental varieties? The bigger the list, the more selection can be done on sprouts, narrowing the field smaller and smaller to just the handful of likely candidate offspring. Sequential crosses between multiple cultivars aimed at bringing in alleles unique to each source are similarly possible in an informed fashion.

To add a bit more complexity, while we've considered an imaginary marker here which is directly, mechanistically responsible for the phenotype, it's possible to have 'linked markers' – things like single nucleotide polymorphisms or SNPs – which may in and of themselves have no direct impact on a gene, but are physically closely associated on a chromosome with the gene; different forms of the marker will then statistically associate with certain allelic forms of the gene, allowing them to act as surrogate markers. Polygenic traits – those influenced by many genes – are another layer of complexity, where often several markers will have to be tracked together to get a desired result.

By combining biochemistry, MAS, and aspects of traditional breeding, the development of new cannabis varieties with directed traits can be done in orders of a few crossing and growth cycles (months), as opposed to thousands of years. For the breeder trying to make the next "big thing," it's a quantum shift in feasibility. ❖

**Mendel's "selective" observation practices, while frowned upon now, were essential at the time in allowing him to determine major statistical trends in crosses leading to the formulation of a viable (and correct) theory describing genetic trait transmission. Since then we've uncovered many of the nuances behind the samples he "didn't record observing."*