Modern Biotechnology in the Improvement of Malaysian Commodity Crop Plants

H.Y. Yeang

Formerly Head of the Biotechnology and Strategic Research Unit, Rubber Research Institute of Malaysia.
Email: yeanghy@gmail.com

This article presents a non-technical picture of how modern biotechnology can be employed to advance Malaysia's main commodity crops. The roles that DNA, genes and the genome play in crop productivity are appraised in relation to their involvement in generating molecular markers that are important new tools for the plant breeder. The potential of genetic transformation in boosting crop productivity is discussed in the light of the regulation that products derived from genetically modified crops are likely to face in future.

Malaysia has embarked on an ambitious thrust into the field of biotechnology, declaring it one of the economic engines for the country. A major objective of the national biotechnology policy is to transform and enhance the value creation of the agricultural sector through biotechnology. How might recent progress in genetics and molecular biology research be harnessed towards the achievement of this ambitious goal? This paper looks at the roles modern biotechnology can play in advancing the country's main export revenue earners in the agricultural sector.

DNA, Genes and Markers of Agronomic Traits

In classical genetics, a gene is a hereditary unit that determines a specific characteristic of an organism. At the cellular level, genes occupy specific locations on chromosomes that are visible under the microscope. At the molecular level, genes take on a different perspective. They are molecules of deoxyribonucleic acid (DNA) that play a role in the synthesis of components essential for cellular viability such as proteins and various types of ribonucleic acids (RNA). DNA molecules are made up of nucleotides linked together linearly in chains thus forming a specific DNA sequence. With protein-coding genes, the arrangement of the nucleotides along the DNA codes for corresponding amino acids. These amino acids, when strung together in the order determined by the DNA sequence, constitute the protein. Just as a biscuit mould determines the form and shape of the biscuit cast from it, the DNA determines the structure and amino acid composition of the protein. How do we reconcile the traditional perception of the gene at the cellular level with that at the molecular level as understood in modern biotechnology? Essentially, how do proteins perform the functions of genes as hereditary units as depicted in classical genetics?

Proteins exist in many forms and they are required for the structure, function, and regulation of an organism's tissues and organs. Among the most recognisable are the structural proteins that include our dietary proteins (the meat that we eat). A different and very important class of proteins is the enzymes that catalyse the various biochemical steps in the complex network of an organism's biochemical pathways. The organism produces different enzymes – numbering in the thousands in plants and animals – to cater to different metabolic needs and functions. In plants for example, genes control the synthesis of the enzymes that convert sugar derived from photosynthesis into starch. Another series of enzymes divert the metabolic products of sugars into the lipid pathway to produce the oil in oil palm fruits. Yet other groups of enzymes convert the sugars into isoprene that end up as natural rubber.

"One of the most useful tools recently made available to plant breeders is the genetic marker."

Ultimately, the gene and the protein it synthesises are responsible for all the characteristics of the plant, whether it is the texture of the bark, the shape of the leaves or the developmental behaviour of its flowers and fruits. Directly or indirectly, genes determine every trait the plant exhibits, other than those that are influenced by the environment. The DNA code laid down in the nucleotide sequences is heritable and in this way, genes impart the individual hereditary characteristics of an individual.

One of the most useful tools recently made available to plant breeders is the genetic marker. A genetic marker is a DNA sequence that can itself be a gene of interest which confers to the plant some advantageous trait. Often, however, a marker is a stretch of DNA of unknown function that happens to be located close to a desirable gene on the same chromosome. Consider this analogy. Names are arranged alphabetically in the phone directory. If the pages are ripped off the directory and shuffled, the original order of names would be jumbled, but the order of names on each page would still be retained. In other words, names that are close together alphabetically tend to remain together even if the pages are scrambled. During sexual crossing, segments of the chromosomes from the male and female parents recombine (i.e. they are re-shuffled) so that the progeny receives genes from both parents. A marker that is known to lie close to a gene of interest indicates the probable presence of that gene since both are likely to be inherited together on the same re-shuffled segment of chromosome. It is therefore possible to perform selection based on a molecular marker, without the actual gene (the 'linked gene') being even precisely located or identified.

If there is already a known plant gene for a desired characteristic, it can be directly screened for in young plants at an early age in the nursery. For this approach to work, not only must there be a pre-identified DNA sequence to search for, it is also essential that the trait is controlled by a single gene. In reality, the challenge to the plant breeder is likely to be far more formidable. The genes responsible for most of the agronomic characteristics of the oil palm, rubber tree or cocoa tree are, in the main, unknown. Moreover, many agronomic traits do not arise from the action of a gene, but are controlled by multiple genes, with each contributing a small interactive influence. Each individual gene acting in isolation may not even display the desired characteristic to a significant extent. Until recently, the only way to select for such multi-genic traits is through careful field observation in the course of conventional breeding and selection. However, emerging molecular techniques now accommodate multiple gene selection.

A DNA marker that is currently increasing in popularity is the Single Nucleotide Polymorphism (SNP), which refers to the variation occurring in a single nucleotide within a stretch of DNA. Commonly, a nucleic acid extract of the test plant is hybridized with an SNP microarray (DNA chip) that features thousands of microscopic spots of SNP markers. By comparing the microarray hybridization patterns between plants carrying a particular trait and those that do not, researchers can infer which combinations of SNPs might be associated with the agronomic characteristic being selected for. SNP microarrays therefore enable high-throughput assessment of genetic variation with the microarray profile displaying the associations between multiple genes (rather than a single gene) that contribute to the agronomic trait. In addition, such interrogation of SNPs is also useful in developing unique fingerprints for identity verification of plant genotypes.

The Genome and Genetic Variation in Commodity Crops

All plant genes are made up of DNA, but not all DNA are components of genes (that encode functional proteins or other cellular products). In man, functional genes make up some 3 to 7% of the total DNA, according to different estimates. If we expect the same of commodity crop plants in general (the rubber tree, cocoa tree, oil palm), then up to 97% of the DNA of these plants consists of DNA with unknown function. Many sections of DNA that do not code for proteins can be cut out or inserted into the plant's genome with no apparent effect on the plant. In the past, such DNA has been commonly called 'junk DNA', but these days many molecular biologists are hesitant to use that appellation because they feel that at least some proportion of this DNA may play regulatory or other roles that have yet to be properly elucidated.

The 95%, or thereabouts, of the plant's DNA that are not functional genes are the 'unexpressed' or 'non-coding' DNA sequences. For that reason, researchers in R&D often concentrate on the 5% of the plant's DNA (the 'expressed sequences') that make up genes which may be assigned putative function based on their sequence

similarities with those in other plant species. This choice is often pressed upon research groups that are limited in their research budget and manpower resources. The rationale here is that DNAs that make up genes are deemed, rightly or wrongly, to be more relevant and important than DNAs of unknown function. When genes are expressed, they first form an intermediate nucleic acid called messenger-RNA (m-RNA) that conserves the genetic code. Researchers make use of this m-RNA to discriminate the expressed DNA from the non-coding DNA. For example, the Rubber Research Institute in collaboration with the Malaysia Genome Institute and Universiti Kebangsaan Malaysia has used this approach to compile a database of 35,000 'expressed sequence tags' (ESTs) for rubber tree latex that represent partial gene sequences.

The recent big news in commodity crop research is the near whole genome sequencing of the oil palm independently by two research groups from Asiatic Development and Sime Darby. Among crop plants, only the rice plant (in 2002), sugar cane (in 2003) and maize plant (in 2008) have had almost their whole genomes sequenced prior to this. There are other crops as well that are in varying stages of completeness in their genome sequencing. (Since there is continual refinement of the initial draft sequence of a crop plant, when the genome sequence is deemed 'complete' may vary between researchers.) The oil palm genome is about four times the size of the rice genome and 90% that of the maize genome. This achievement is a very significant high point in research for the country. The Malaysian success in deciphering the oil palm whole genome sequence took the scientific world by surprise. Right until August 2008, a commentator did not hide his scepticism in writing that "Cheah (Suan Choo of Asiatic) also claimed that her company has completed the sequencing of oil-palm". Needless to say, Asiatic is having the last laugh. Of the other major Malaysian commodity crops, sequencing of the rubber tree genome is being given serious thought. The rubber genome is comparable in size to that of the oil palm and is four times the size of the cocoa genomes.

In sequencing the whole genome, researchers do not restrict themselves to the 5% of the plant's expressed DNA that make up the genes. They go the whole hog and sequence the entire complement of the plant's genes. In the case of the oil palm, this amounts to some 1.7 or 1.8 billion pairs of nucleotides that include all of the approximately 31,000 genes plus the remaining 95% unexpressed DNA of uncharacterised function that are not involved in encoding proteins. In whole genome sequencing, it is usual to begin by using only a small number of plants, or even one plant, to serve as a working model. Since roughly 98% of the DNA sequences between individual plants of the same species are identical, the whole genome of one individual tenera palm, for example, pretty much represents that of all tenera palms. For many areas of genetic research, such as in determining which genes are present or absent, on which chromosomes a particular gene is located, the genetic distance between two selected genes, etc., this level of similarity is more than adequate. The remaining approximately 2% of the DNA difference has very little bearing on such studies.

From where does this 2% variation between one palm tree and the next arise? This genetic variation is made up of SNPs, already mentioned earlier, and another attribute of the genome known as Copy Number Variation (CNV). Unlike SNPs which affect single nucleotides, CNV occurs when segments of DNA, generally longer than 1000 nucleotides, are either repeated or are deleted. CNVs extend over more than 12% of the human genome sequences in diverse populations. In an individual human genome, however, only some 100 CNVs (comprising 25 million nucleotides) are normally encountered. This makes up about 0.8% of an individual genome. Add to this the fact that (in humans at least) one in every 100 to 300 nucleotides contains a variation (i.e. an SNP) leading to single nucleotide variations making up somewhere between 0.3 to 1% of the genome. From these figures, one individual oil palm might differ from the next in less than 2% of its DNA, whether as SNPs or as CNVs..

If the DNA variations in an individual palm are a mere 2% of the total DNA, should we even be bothered with them? Can we not simply ignore them? Well, we really shouldn't. All genetic variation – whether it is between a thick or thin palm fruit mesocarp, between resistance or susceptibility to cocoa vascular streak dieback disease, between short latex flow or prolonged latex flow in the rubber tree, etc. – lies in this 2% variation in the DNA. Essentially, what separates an elite specimen from the run of the mill lies in this 2% DNA variation. It is also within this 2% of the DNA that molecular markers are to be found that will enable plant the breeder to carry out selection. Of course, new insights into the nature of the genome emerge all the time and it remains to be seen if the figures cited above will hold up in the light of future discoveries.

The whole genome sequence in itself is not the final goal of the researchers who have toiled hard and long to reach that milestone. It is what is to be done with the information that will really make a difference to the industry. Both Asiatic and Sime Darby declare that information obtained from the whole genome sequence will enable oil palm yields to double, with the former promising commercialisation of the improved varieties in six to seven years. However, neither company has disclosed in any detail how this might come about; they are keeping their research strategies very close to their chest. It is safe to say that the availability of the oil palm

whole genome sequence would contribute immensely to the discovery of selection markers.

The gene variations are certainly out there in the genome sequence, but pinning down the SNPs and the CNVs from among the 1.8 billion nucleotides is something else again. To identify SNPs or CNVs, it is necessary to compare at the very least the DNAs of two different individuals. Even with that done, the differences elucidated are confined only to those between these two individuals. The extent of success in identifying them would therefore depend in part on how many individual palms Asiatics and Sime Darby have sequenced. If the whole genome sequencing has been performed on a small number of palms, the repertoire of DNA differences would not be exhaustively represented in the genome sequences already at hand. Moreover, it might be difficult to attribute specific agronomic traits to whatever markers that are identified. Therefore, the availability of genome sequences from multiple individuals exhibiting widely varying traits, both good and bad, would be useful in identifying a large number of useful genotype markers to compile a comprehensive marker library. In this connection, the genome sequences obtained separately for the dura and tenera varieties of the oil palm by Asiatic should provide useful contrasts in agronomic characteristics that allow gene markers to be identified and assigned to them. In searching for human CNVs, researchers initially compared the two human genome maps: one assembled by Celera Genomics, Inc. and one from the public Human Genome Project. Similarly for Asiatic and Sime Darby, it might be worthwhile considering the pooling of their data to increase the count of individual palm sequences (and consequently DNA variation) that are available to identify more DNA differences that are potentially valuable selection markers.

As impressive as the complete genome sequence is, one should not go away with the idea that it is indispensable for identifying molecular markers. Markers such as SNPs can be picked out from segments of DNA that are representative of a broad cross-section of the entire genome. Therefore, what is often referred to as 'whole genome selection' need not necessarily require the entire genome of the crop plant to be sequenced. SNPs are found both in the expressed portions of the genome, i.e. the genes, as well as the non-expressed portions of the genome. In fact, many researchers who employ microarray technology choose to limit themselves to the genes, while leaving out the non-expressing portion of the genome in their analysis.

It bears repetition – whether or not the whole genome sequence is available – that success of the gene marker technology will depend to a large extent on the number of markers, such as SNPs, that are available for screening and the number of individual plants exhibiting a wide range of traits that have been screened using these markers. Ideally, all single nucleotide variations that occur at a frequency greater than 1% in the crop plant population should be represented in the SNP library. (In the US Human Genome Project, the target is an SNP map of at least 100,000 markers.) When the markers are available, large numbers of plants, comprehensively indexed for their agronomic characteristics, need to be screened to establish which microarray profile is associated with which agronomic trait. In cattle breeding, for example, the US Meat Animal Research Center plans to screen more than 7,000 head of cattle against an SNP microarray in its quest to breed cattle that produce more milk, more meat and less fart (a worrisome greenhouse gas).

Tissue Culture and Genetic Transformation

Other than when it is used as a parent plant for breeding purposes, a single superior specimen on its own is not going to make much impact to crop productivity. To establish a cultivar for commercial planting, its principal agronomic traits must be fixed and reproducible in the field. In tree crops such as oil palm, rubber and cocoa where the breeding cycle is lengthy, repeated back-crossing to fix a selected trait is inevitably a long, drawn out affair. Clonal multiplication allows the selected genotype to be adopted in commercial planting with minimal time lag. In this regard, tissue culture has, in many cases, supplanted the traditional techniques of plant clonal propagation such as bud-grafting, cuttings, marcotting, etc. However, tissue culture of tree crops, especially *via* embryogenesis (i.e. not by microcuttings), is generally slow, difficult and costly. Rubber, oil palm and cocoa are not exceptions to this rule. With bud-grafting available to rubber and cocoa, it makes little economic sense to venture into tissue culture for the purpose of mass clonal propagation. The oil palm is another story. With a solitary vegetative bud at its growing apex, there are simply no other buds to spare for bud-grafting. Tissue culture is therefore the method of necessity, rather than of choice, for clonal multiplication of the oil palm.

Besides vegetative propagation, tissue culture has another useful application in all three commodity crop plants. One of the best-known applications of modern plant biotechnology is in genetic transformation whereby selected genes are inserted into the plant's genetic constitution. In conventional plant breeding, very large numbers of genes are unavoidably reshuffled. Hence, while a progeny from sexual crossing could gain a desired trait, it might, at the same time, also inherit unwanted characteristics or lose some of the good traits displayed by its parents. In genetic transformation, on the other hand, the gene controlling a specific trait

is inserted into the plant, leaving the plant's other characteristics generally unaltered. With genetic transformation, the species barrier can be – and often is – traversed. Genes from bacteria, plants and animals alike can be inserted into plants through such genetic manipulation. Hence, a genetically transformed plant is also known as a transgenic plant, and come under the classification of what are popularly termed genetically modified organisms (GMOs).

Several methods have been developed for the genetic transformation of crop plants. A commonly adopted approach to transfer a desired gene into the plant is through mediation of the soil bacterium *Agrobacterium.* In this technique, the DNA that encodes the desired genetic elements is first inserted into *Agrobacterium* which is then allowed to infect the host plant tissue. In the process, the desired DNA is transferred from the bacterium into the genetic makeup of the plant. With crop plants, the common objective of genetic transformation is to improve crop productivity and quality. For the rubber tree, high latex output for increased rubber production is desirable, as is high girthing rate to reduce the period of immaturity and increase its value in timber. For the oil palm, modified fatty acid composition of the oil, disease resistance and amenability to mechanised harvesting are among the desirable traits. Flavour, high levels in health-related compounds and resistance to pests are among the characteristics that are sought for the cocoa plant.

There are practical limits to the extent to which genetic transformation can be applied in crop improvement. Since genetic transformation involves a plant regeneration step *via* tissue culture, the considerable complications associated with tree tissue culture are also experienced in genetic transformation. That is therefore true for the rubber tree, the oil palm and the cocoa tree. The immediate product of the gene is the protein that it encodes. Hence, genetic transformation is simplest where the goal is to have the transformed plant express a single target protein that is controlled by the inserted gene. With the well-known Bt gene derived from the insecticidal bacterium *Bacillus thuringiensis*, for example, the gene produces the protein toxin that is the target agronomic characteristic itself. Where an agronomic trait is controlled by multiple genes (a common occurrence), genetic transformation to confer such a trait to the transgenic plant becomes far more problematic. Even if such genes are identified and available, their insertion in a single act of genetic transformation, while not impossible, makes an already difficult task even more daunting. In fast growing herbaceous plants, researchers often transform genes separately into different plants, and then obtain plants with a combination of inserted foreign genes by conventional sexual crossing. With the commodity tree crops, however, the duration of the sexual cycle makes this an unattractive proposition.

How much increase in crop productivity can reasonably be expected from genetic transformation? The degree of success from genetic transformation would depend on the crop, the inserted gene and the environment where the transgenics are cultivated. Although transgenic crops have been cultivated for more than 15 years now, many early transgenics were inserted with a gene for glyphosate resistance that caters for weed control rather than improvement in crop production. These are therefore not good examples by which to gauge improvements in crop productivity. In India, cotton bioengineered with the Bt gene has been experimented on since 1977. Yield from cotton plants transformed with the Bt gene rendering the plant resistant to three species of bollworm can increase by as much as 80%. While such results may seem spectacular, it should be noted that the yield gains of similar transgenic cotton crops in the United States and China average less than 10%. So why the vast difference? Essentially, the increment in crop production depends very much on the prevalence of insect pests and the extent to which such insect pests are being controlled by existing methods. If good pest management is already practised, involvement of the Bt gene can add only marginally to the crop. In this case, the savings in crop management are as important as any actual increase in crop productivity. The first Malaysian commodity crop to be genetically transformed is the rubber tree. However, data on productivity improvement is lacking because the early research in this area concentrated on the production of pharmaceutical proteins in the latex rather than in enhancing productivity of rubber or rubberwood.

Regulating Transgenic Crops and their Products

Field planting of transgenic plants in Malaysia are now regulated through provisions laid down in the Biosafety Act. This is being done to protect the environment so that transgenic plants do not spread indiscriminately and adversely affect the ecological well-being. The regulations are in place also to protect farmers who might not wish to have non-transgenic planting materials contaminated with their genetically modified counterparts. The Biosafety Act regulates not only the planting of transgenic plants, but also the products derived from such plants. This is a critical issue when it comes to the trade of these commodities on the market.

Whereas oil palm, rubber, cocoa form the backbone of Malaysian commodity agriculture, the country does not export oil palms, rubber trees or cocoa trees. The country's exports are their produce, *viz.* palm oil, rubber,

cocoa and timber. Hence, if genetically modified (GM) crops are planted, it is not so much the transgenic plants themselves that are of concern for the export market, but the commodities from these genetically modified trees that would be subject to regulation such as additional testing for product safety, mandatory labelling and other restrictions. According to the Biosafety Act, products of genetically modified organisms (GMOs, that would include all transgenic plants) are defined as 'any product derived from a living modified organism or part of a living modified organism—
(a) if the product contains detectable recombinant deoxyribonucleic acid (DNA); or
(b) where the profile, characteristic or properties of the product is or are no longer equivalent to its conventional counterpart irrespective of the presence of the recombinant deoxyribonucleic acid (DNA).

Let's take a closer look at this particular ruling and its implications. 'Recombinant DNA' refers to the foreign DNA that is inserted into the plant by genetic transformation. All DNA in higher organisms – whether naturally occurring or recombinant – is found in the nuclei of its cells. Therefore, any plant or animal product that contains cellular material would contain DNA and would be subject to regulation as a GM product under Malaysian law. For example, cocoa beans are cellular and those harvested from transgenic plants contain recombinant DNA. On the other hand, cocoa butter is the fat extracted from cocoa beans. If sufficiently purified, it should not contain significant cellular material and hence need not come under regulation as a product of a GMO. However, much depends on how stringently the ruling is applied since trace contaminants of residual cellular material in cocoa butter may be difficult to avoid. Similarly the highly refined oil from the transgenic oil palm should escape regulation. Nevertheless, if the oil concerned is modified in its composition (for example, if the fatty acid composition is altered to enhance health benefits) through transgenesis, it would be subject to regulation under the part of the ruling covering 'characteristic or properties of the product is or are no longer equivalent to its conventional counterpart'.

The genetically modified rubber tree is treated differently again because of the peculiarity of natural rubber latex. Natural rubber latex is exuded from latex vessels (laticifers), making latex a cellular product. However, the nuclei of latex vessels are not exuded together with the latex when the tree is tapped. As mentioned above, DNA is found in nuclei of cells and tapped natural rubber latex that contains no nuclei therefore contains no DNA (other than in trace quantities). Certainly, the latex would contain recombinant ribonucleic acid (RNA), which is another nucleic acid closely related to DNA, and not confined to the nucleus. But the definition in the Malaysian Biosafety Act for the products of GMOs does not mention recombinant RNA or recombinant nucleic acids. The definition specifies recombinant DNA, thus allowing natural rubber to escape regulation for the products of GMOs under Malaysian law. This does not mean that rubber from transgenic trees can slip innocuously into export markets, however. Importing countries have their own definitions for the products of GMOs that can be different from Malaysia's. In the United Nations Environment Programme (UNEP) Biosafety Protocol, for example, the definition of a GMO makes a reference to 'replicable genetic material obtained through the use of modern biotechnology'. In simple terms, that means recombinant nucleic acids, whether DNA or RNA. Here, there's no escape for natural rubber sourced from transgenic trees. The other major product of the rubber tree is rubber wood. As is true of any timber, rubberwood is cellular in nature and contains DNA. Timber from transgenic rubber tress will therefore be regulated in Malaysia.

Transgenic research in Malaysia has not reached the stage where the country is ready to export commodities derived from transgenic crops. When the time comes, careful thought must be given before embarking on this path because of the negative perception of transgenics in many countries, especially in Europe, that import our products. Many consumers distrust GM foods even though no harmful effects from their consumption have been proven. Not so much in the case of rubber which is not a food, but cocoa and palm oil can be targeted for denigration should these products from transgenic cultivars be offered for sale. Once in the market, buyers may not discriminate between products from transgenic or non-transgenic sources, leaving the entire industry under a cloud of suspicion. Increasing familiarity with products from transgenic crops in recent years could perhaps persuade consumers to review their position on transgenic crops in the future. We would be in a better position to judge when the time comes.

Acknowledgement

I thank Dr Chow Keng See, Rubber Research Institute of Malaysia, for her helpful comments and criticisms on this article.