Development and application of molecular DNA markers in Africa: a South African view

A-M Botha1*, E Venter1, C van der Vyver2 and KJ Kunert2

1 Department of Genetics, Forestry and Agriculture Biotechnology Institute, University of Pretoria, Hillcrest, Pretoria 0002, South Africa
2 Department of Botany, Forestry and Agriculture Biotechnology Institute, University of Pretoria, Hillcrest, Pretoria 0002, South Africa
* Corresponding author, e-mail: ambothao@postino.up.ac.za

Received 25 September 2003, accepted 15 October 2003

The development of molecular DNA markers for genetic analysis has greatly increased our understanding of the structure and behaviour of plant genomes. Different DNA marker technologies exist, however, information suggests that restriction fragment polymorphisms, amplified fragment length polymorphisms and also microsatellites are currently the most effective techniques for the detection of polymorphisms in plant genomes. Several other technologies, such as cDNA-amplified fragment length polymorphism, representational difference analysis or suppression subtractive hybridisation analysis, and cDNA microarrays, belong to a range of novel techniques very useful for the analysis of transcriptome variation, under specific stressful environmental conditions. However, the application of these technologies especially in Africa is not trivial, as execution of these techniques is largely dependent upon the availability of a technological advanced infrastructure, technical skills in plant molecular biology and substantial funding of the research groups involved in this type of research. Different molecular DNA-based marker techniques and their application to genome analysis and molecular breeding, and suitability for use in Africa, are discussed in this review.

Introduction

Use of DNA-based genetic markers has forever changed the practice of genetics. In the 20 years since its discovery, many different types of DNA-based genetic markers have been used predominantly in the developed world for the construction of genetic maps, for the analysis of genetic diversity, trait mapping, as well as for applied diagnostic purposes. Molecular DNA markers are in general useful for a variety of purposes relevant to crop improvement. The most important of these uses is indirect marker-assisted selection (MAS) exercised during plant breeding. Once a target trait has been genetically unravelled, breeders use molecular DNA markers to accelerate breeding efforts by tagging favourable alleles in the selected genetic background to be improved (Dreher et al. 2003). The markers are then tested in small subsets of material to ensure that the desired loci are present through repeated selection cycles. Sampling for the desired trait is done in a large segregating population through identification of the individuals that carry the favourable trait (Ribaut et al. 2002).

The transition of molecular markers from the genetic analysis phase to validation and implementation can be very problematic, especially on the less-developed African continent, and validation of a molecular DNA marker for trait selection involves several factors. The first step would be to confirm that genes at the marked locus in the breeding lines of interest confer the trait and not genes at another locus. Secondly, the polymorphism for the markers between parent breeding lines of interest needs to be established. Thirdly, the linkage needs to be proved in breeding populations other than the one in which the trait was developed. This linkage needs to be sufficient in these populations for breeding purposes. Finally, the methodology needs to be adapted to facilitate large sample collections or population sizes (Christopher et al. 2003). In addition, it is also helpful if the developed molecular DNA marker is simplistic, and thus amenable to automation and high throughput approaches.

DNA-based marker systems have the general advantage that the DNA content of a cell is independent of environmental conditions, organ specificity or growth stage. Each cell of a living individual contains DNA as genetic material. This DNA determines the individual characteristics via the control of protein synthesis in the cell. DNA marker development has also recently taken advantage of the ease of the polymerase chain reaction (PCR) method. This entails that specific DNA fragments can be amplified million-fold in a
short time period in a thermocycler.

However, a frequently asked question is: What marker system is most useful? The answer to this question lies with the required technology, available infrastructure and technical expertise of the researcher and is therefore a critical question for African scientists. In general, two broadly defined systems exist, namely those that derive from genomic DNA, which often requires less sophisticated technologies and skills, and those that function on gene coding (transcriptome/proteome) level, and thus relating to gene products. The latter are normally based on rather expensive consumables and highly sophisticated equipment unaffordable to research laboratories in less-developed countries. In general, isolation of DNA from plants is seldom a major technically demanding process and can be carried out in any laboratory with a minimum of chemical reagents and basic DNA isolation and characterisation equipment. This renders systems based on characterisation of genomic DNA much easier to use. Such systems usually target variable regions of the genome. These variable genome sequences include single copy and repetitive genomic regions. These genomic regions may be functional or 'silent' without any obvious function. The second type of marker targets DNA coding regions in the plant genome is making use of variation in gene expression in response to internal or external signals. This, however, requires the technically much more demanding process of cDNA synthesis from isolated purified total RNA. Often the need exists to purify mRNA using commercial isolation kits. A further issue to consider is the marker type. Presently, a number of PCR-based systems for marker development exist and have been reported in the literature. Using the potential of PCR is certainly an advantage in Africa due to its technical ease in final application and demand for rather standard chemicals and enzymes available in any basic molecular biology laboratory.

The aim of this paper is to explain briefly the different molecular marker technologies, as well as discuss the application potential of specific technologies. Further, the current activities in Africa using the different technologies are highlighted. Ultimately, we also express our views concerning current limitations and future prospects for marker development in Africa.

### Marker Systems Using Genomic DNA (gDNA)

#### Random amplified polymorphic DNA

In 1990, Williams and coworkers published a paper on the arbitrary amplification of DNA without any prior knowledge of the amplified sequence (Williams et al. 1990, Welsh et al. 1992). This discovery revolutionised DNA marker development, as it opens the way for low technology and high throughput marker development. The characterisation of a DNA sample by Random Amplified Polymorphic DNA (RAPD) analysis using PCR has attracted considerable attention in the last ten years. This technique is often referred to as DNA 'fingerprinting', and is easy to apply to different sample sets. It has also been used in South African laboratories for that purpose, and has been successfully applied to cultivar identification and determination of genetic diversity for a variety of crops including cassava (Zacharias unpublished results), wheat (Myburg et al. 1997, Labuschagne et al. 1998), pumpkin (Gwanama et al. 2000), and sugarcane (Huckett and Botha 1995a, 1995b, Harvey and Botha 1995, 1996, Harvey et al. 1996). This technique, possibly the simplest of all recently applied DNA-based technologies, still plays and will continue to play a major role in the future as an affordable, technically less demanding marker system for plant cultivar identification in Africa with limited resources and an under-developed skills base.

In general, the RAPD technique consists of the production of duplicate segments of target plant DNA (Table 1). These specific DNA segments are amplified by PCR until several million copies are produced. For plant DNA duplication, pairs of very short DNA nucleotides (10-mers), called primers, are required. These primers bind randomly to the master-copy of plant DNA. Several biotechnology companies, including local South African companies, manufacture these primers. A thermo-stable enzyme, called Taq DNA polymerase, and a thermal cycler, which carries out rapid temperature fluctuation cycles, is necessary for the DNA duplication process. Finally, amplified DNA segments are separated on either agarose or polyacrylamide gels and visualised by staining. The detected polymorphism between individual plants may result from either a DNA sequence difference in the binding site for the primer on the plant DNA, or a deletion of parts of the plant DNA. These markers are

<table>
<thead>
<tr>
<th>Marker system</th>
<th>PCR based</th>
<th>Loci/ assay</th>
<th>DNA amount (µg)</th>
<th>Time/ assay</th>
<th>Technical skills required</th>
<th>Amenable to automation</th>
<th>Marker type</th>
<th>Development costs</th>
<th>Cost per analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>RFLP</td>
<td>No</td>
<td>1–3</td>
<td>5.0–10.0</td>
<td>5 days</td>
<td>High</td>
<td>Low</td>
<td>Co-dominant, reliable</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>PCR-RFLP</td>
<td>Yes</td>
<td>1</td>
<td>0.05</td>
<td>2 days</td>
<td>Low</td>
<td>High</td>
<td>Co-dominant, reliable</td>
<td>Moderate</td>
<td>Low–Moderate</td>
</tr>
<tr>
<td>RAPD</td>
<td>Yes</td>
<td>1.5–50</td>
<td>0.02</td>
<td>5h</td>
<td>Low</td>
<td>Moderate</td>
<td>Dominant, unreliable</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>AFLP</td>
<td>Yes</td>
<td>20–100</td>
<td>0.5–1.0</td>
<td>2 days</td>
<td>Medium</td>
<td>Moderate–high</td>
<td>Mostly dominant, reliable</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>SCARS/STS</td>
<td>Yes</td>
<td>1</td>
<td>0.05</td>
<td>5h</td>
<td>Low</td>
<td>High</td>
<td>Co-dominant, reliable</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>SSR</td>
<td>Yes</td>
<td>1–3</td>
<td>0.05</td>
<td>5h</td>
<td>Low</td>
<td>High</td>
<td>Co-dominant, reliable</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>SNPs</td>
<td>Yes</td>
<td>1</td>
<td>0.05</td>
<td>5h</td>
<td>Low</td>
<td>High</td>
<td>Co-dominant, reliable</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>RDA</td>
<td>Yes</td>
<td>Many</td>
<td>1.0</td>
<td>2 days</td>
<td>High</td>
<td>High</td>
<td>Co-dominant, reliable</td>
<td>Moderate</td>
<td>Low</td>
</tr>
</tbody>
</table>

* During development phase
inherited in a Mendelian fashion. Since RAPD analysis requires only a high quality thermocycler and a gel assembly, this technique has the advantage of being simple, rapid and requiring only a small amount of isolated genomic DNA from any suitable part of the plant. Virtually unlimited numbers of different RAPDs can be obtained by simply changing a primer pair without changing any of the other experimental conditions. After optimisation of the technique, a semi-qualified technical person trained in basic molecular biology techniques can theoretically carry out RAPD analysis on a routine basis.

Since 1995, RAPDs have been utilised for the development of several markers for application in MAS for a variety of agronomic important traits. These include markers for Russian wheat aphid resistance genes Dn1, Dn2 and Dn5 (Botha et al. 1995, Myburg et al. 1998, Venter and Botha 2000, Venter et al. 1998), and leaf rust resistance Lr37 (Troskie et al. 1999) and Lr42 (Lottering et al. 1999, 2002). However, it was found that transfer of the technology, for example to commercial enterprises, is extremely difficult. Thus, the major drawback of application of the basic RAPD technology in the search for molecular markers, is that reproducibility between laboratories is relatively low due to the sensitivity of the amplification step determined often by the brand of the polymerase and thermocycler used. This can severely limit RAPD for routine application in MAS unless the fragment is converted into a more stable marker type, e.g. sequence characterised amplified regions (SCARs) or sequence tag sites (STS) (Venter et al. 1998, Botha and Venter 2000, Venter and Botha 2000) (Figure 1).

**Amplified fragment length polymorphisms**

Amplified fragment length polymorphism (AFLP), developed by Zabeau and Vos (1993), is a reproducible, multiplex assay (Table 1). In comparison to other DNA marker technologies, AFLPs have the ability to generate a much larger number of polymorphic genomic fragments in a relatively short time. The technology involves the restriction digestion of genomic DNA, ligation of DNA adaptors to the digested DNA, followed by PCR rounds of pre-selective and selective amplification of the restricted fragments (Vos et al. 1995). The resulting fragment sizes range from 60–1 500bp, and amplified fragments can be fluorescently or radioactively labelled and separated on sequencing gels. AFLPs seem to overcome the major pitfalls present in RAPDs and other arbitrary methods, and appear to be as reproducible, heritable and intraspecific as RFLPs (Botha and Venter 2000 and references therein, Powell et al. 1996). Because of their large genome coverage (AFLPs on average gives 50–100 bands compared to 20 for RAPDs, Table 1), AFLPs appear to be particularly useful for ‘DNA fingerprinting’ (identification of genotypes), genetic diversity studies, local marker saturation, construction of genetic maps and quantitative trait loci (QTL) mapping in plants (Brugmans et al. 2003 and references therein). AFLPs are frequently applied for DNA fingerprinting, genetic mapping, marker development, determination of genetic diversity and heterosis in South African laboratories for a variety of crops, e.g. potato (Jacoby et al. 2003), wheat (Lottering et al. 1999, Troskie et al. 1999), and maize (LW Beyene, personal communication). The technology has also been applied for the identification of markers

![Profiles obtained after RAPD (A) or AFLP (B) amplification](image)

![Profile obtained after SCAR/STS analysis](image)

**Figure 1:** Development of sequence characterised amplified region (SCAR) or sequence tag site (STS) markers. The cloning and sequencing of a RAPD or AFLP amplicon allows for the design of specific primers for the amplification of a target-specific locus
and QTLs against root-knot nematode (Minnie et al. 2002), gray leaf spot resistance in maize (Lehmensiek et al. 2001), and wood properties in Eucalyptus and Arabidopsis (AA Myburg pers. comm.). AFLP markers have further been developed for leaf rust resistance genes Lr42 (Lottering et al. 1999, 2002), Lr37 (Troskie et al. 1999) and Lr19 (Prins et al. 2001) in wheat. Markers linked to the Aegilops ventricosa-derived chromosome segment in the wheat cultivar ‘VPM1’ on which the eyespot resistance gene PcH1, and the endopeptidase gene Ep-D1b occur, were also identified (Groenewald et al. 2003).

However, data indicate that AFLP bands in a genome may cluster around the centromeres. This is clearly related to the selected primers being used. Additionally, AFLPs offer the opportunity to compare diversity of hypo- versus hyper-methylated portions of the genome. This is done by comparing data from restriction enzyme combinations that are methylation-sensitive with methylation-insensitive combinations (Botha and Venter 2000). Evidence indicates that DNA sequences are transcribed more readily when hypo-methylated (Cedar 1988).

One of the major drawbacks of AFLPs is that it is mostly a dominant marker type and thus cannot always determine the heterozygous individuals. Also, in the absence of pedigree analysis, the identity of individual bands in a multi-band profile is not known, and there is no certainty in assigning markers to a specific locus. The presence of a fragment of identical molecular weight in different individuals cannot be taken as evidence that the two individuals share the same homologous fragment, although this assumption is generally true. A further difficulty is that single bands on a gel sometimes comprise several co-migrating amplification products. This is a major obstacle in AFLP fragment characterisation, when cloning of specific fragments becomes necessary after separation of the fragments (Botha and Venter 2000). The complexity of banding patterns, as well as the uncertainty relating to fragment identity, makes AFLP less suitable for high throughput screening in MAS. It is thus highly recommended that fragments of interest be converted into SCARs or other suitable systems if MAS is the objective in the study (Figure 1).

**Sequence characterised amplified regions**

With the problems associated with the transfer of technology and complexity of RAPD and AFLP fragments, there is a need, which might also be relevant for Africa, to develop an alternative simplified and more robust system for use in routine screening (Xu and Bakalinsky 1996). Sequence characterised amplified regions (SCARs), also known as sequence tagged sites (STS) or allele-specific associated primers (ASAP), are genomic DNA fragments at single genetically defined loci that are identifiable by PCR amplification using oligonucleotide specific primer pairs (Paran and Michelmore 1993) (Table 1, Figure 1). SCARs differ from RAPDs and AFLPs in that it amplifies a specific region in the genome, usually from low copy regions. SCARs are developed through the cloning and sequencing of a specific fragment derived from RAPD or AFLP, that is linked to a trait of interest. When designing a SCAR primer, the RAPD or AFLP primer, together with the next ten to fifteen bases, are used to design a more robust and specific primer set. This primer set results in a less complex banding pattern than the original profile obtained after RAPD or AFLP analysis (Figure 1). It is also possible to design primers from internal sequences where there are point mutations, deletions or insertions (Feuillet et al. 1995).

In South African laboratories, SCAR markers have been so far developed for root-knot nematode resistance in soybean (Minnie et al. 2002) and leaf rust resistance gene Lr19 (Prins et al. 2001). It also has been proved useful in the development of markers linked to Russian wheat aphid resistance locus Dn2 (Myburg et al. 1998), as well as for the leaf rust resistance gene Lr42 (Lottering et al. 2002). In another study, conversion of RAPD into SCARs however failed and an alternative strategy had to be deployed (Venter and Botha 2000). Such failure has also been reported by Paran and Michelmore (1993) during the development of markers linked to downy mildew resistance in lettuce. The failure to convert the primers to SCAR primers is seemingly attributed to the loss of a point mutation between the different individuals when using the longer primer and that the longer primer very likely tolerates the internal mismatches (Paran and Michelmore 1993, Venter and Botha 2000). The solution to this problem is by changing the length of the primer to obtain allele specificity (Gu et al. 1995). If internal deletions, insertions or point mutations exist in the amplified fragment, it may be appropriate to conduct restriction digests of the amplified product to differentiate between individuals (Venter and Botha 2000). In this approach, the best results are usually obtained by using four base pair cutters, as they have a better chance to restrict the fragment of interest.

SCARs have many advantages over other marker systems for high throughput screening. This includes the reduction of complexity and a high level of robustness. Thus, it is less variable among different PCR cyclers and different DNA polymerases. SCAR data can also be submitted to databases as sequence information, and isolated fragments can be used as probes to screen genomic libraries. DNA fragments can further be mapped on a genetic map as ‘physical landmarks’ (Paran and Michelmore 1993). The main drawback of SCARs is the need for sequence information to design the primers, also implying cloning of the fragment where appropriate.

**Polymerase chain reaction-restriction fragment length polymorphisms**

One of the earlier techniques for DNA marker development is the creation of restriction fragment length polymorphisms (RFLPs) where differences in the length of restriction fragments are detected. Complete digestion of genomic DNA with DNA restriction endo-nucleases generates the restriction fragments (Sambrook et al. 1989). However, RFLPs have the disadvantage of being laborious (Table 1). The technique further requires a DNA-DNA hybridisation process using an isolated and labelled DNA probe for detection of any DNA polymorphism (Southern 1975). Also, more importantly perhaps for Africa, the RFLP technique is technically demanding when applied in any routine breeding applica-
Microsatellites are generated by highly specific PCR amplification and, therefore, should not suffer from the reproducibility problems experienced with RAPD analysis (Table 1). They can be easily visualised after standard DNA agarose or polyacrylamide gel electrophoresis. Microsatellites have recently emerged as an important source of ubiquitous genetic markers for many eukaryotic genomes (Wang et al. 1994).

For plants, it has been demonstrated that microsatellites are highly informative, locus-specific markers in many species, including *Olivia* (Swanevelder unpublished results), wheat (Swanepoel et al. 2003), and maize (LW Beyene unpublished results). Since they are multi-allelic, microsatellites have a potential use in any evolutionary studies and investigation of genetic relationships in plants (Botha and Venter 2000 and references therein, Ritland 2000 and references therein). In a recent study on a pine hybrid species, Doyle et al. (2002) further demonstrated that microsatellites could be applied as a useful marker tool to determine gene flow and parentage. Overall, the isolation of microsatellites is becoming increasingly routine, especially in well-equipped molecular biology laboratories, due to the availability of automated DNA sequencing facilities, advanced techniques for the construction of genomic libraries enriched for microsatellites, and better techniques for screening of the appropriate clones.

**Single nucleotide polymorphisms**

A new generation of DNA markers, called single nucleotide polymorphisms (SNPs), are the most abundant of all markers in both animal and plant genomes. With the availability of genomic sequence, SNPs can be detected on a large scale. In humans, single nucleotide differences occur on average every 1–2kb, and a map of 1.42 million SNPs has already been produced (Wang et al. 1998, The International SNP Mapping Working Group 2001).

When considering SNPs at closely linked sites, it has been observed that the number of different combinations, or haplotypes, is limited. This makes it feasible to scan populations for haplotype variants and to test for association between a haplotype and a disease phenotype (Table 1). The probability of detecting such an association will depend on the extent of linkage disequilibrium (LD). Recombination, gene conversion, random genetic drift and selection all influence LD in a specific population (Brookes 1999). It is also not surprising that LD may vary indifferently between parts of the genome. The extent of LD will determine the number of loci that are needed to conduct a successful genome scan. In plants, where dense SNP maps are not available, whole genome-association studies will yield results only if LD extends over regions of several hundreds of kb (Devos and Beales 2003). At present, data on the extent of LD in plants are further highly dependent on the species, genome region, and population structure (Nordborg et al. 2002, Rafalski 2002, Zhu et al. 2003). This limits any short-term implementation of the technology. In a study on the SNPs associated with the vernalisation response in wheat, it was shown that association studies alone would not be sufficient to confirm the identity of candidate genes (Devos and Beales 2003).

---

**Microsatellites/simple sequence repeats**

Another widely used technique for marker development is the detection of microsatellites or simple sequence repeats (SSRs) (Table 1). Microsatellites are genetic markers derived from tandemly repeated basic DNA motifs of <6bp such as (GA)$_n$, (GT)$_n$, (TG)$_n$ or (AAT)$_n$. Such repeated sequences are widely dispersed throughout the eukaryotic genome. They are often highly polymorphic due to variation in the number of such repeats, which seemingly arise through slippage errors during DNA replication. This may occur even between closely related individuals.

**Figure 2:** Polymerase chain reaction-restriction fragment length polymorphism (PCR–RFLP) derives from specific PCR amplicons after cleavage with restriction endonucleases

![DNA template](image)

- PCR cycle 1
- PCR cycle 2
- PCR cycle 3

25-35 PCR cycles produce sufficient template for PCR-RFLP analysis

- Restriction of PCR product with restriction enzymes

- Analysis of restricted product using gel electrophoresis

**Figure 2:** Polymerase chain reaction-restriction fragment length polymorphism (PCR–RFLP) derives from specific PCR amplicons after cleavage with restriction endonucleases
Representational difference analysis

This approach follows a more unique technological path in DNA marker development than other systems (Table 1, Figure 3). Much emphasis has been focused on the evaluation of representational difference analysis (RDA) to identify and characterise more complete variable regions in the plant genome and not just single mutations. These regions might easily change in their sequence composition under stress conditions and, therefore, have potential as a DNA marker.

The RDA technique consists of subtraction of all sequences that are held in common between two individuals that differ for a single trait. RDA comprises three steps that combine into a powerful marker system. First, a representation is created through PCR amplification of a restriction-enzyme digested plant genome. Secondly, a subtractive hybridisation eliminates all DNA sequences that are identical between the two plants that are studied. Thirdly, a kinetic enrichment of all sequences unique to the plant of interest is achieved through another set of PCR cycles. A unique sequence might arise from genomic losses, rearrangements and amplifications (Lisitsyn et al. 1993, 1994). RDA has proved very useful in isolating families of repetitive sequences from diverse sets of samples (Cullis and Kunert 2000, Vorster et al. 2002). The RDA technique is especially suitable for smaller research groups lacking sophisticated equipment and the technique can be executed with basic molecular biology equipment, such as theromocycler and gel electrophoresis equipment.

The advantage of using RDA over RFLPs, RAPDs and AFLPs are in the proportion of the genome that can be scanned with each subtraction. The use of RDA can scan up to 15% of the genome of most plants in each subtraction, while the use of 300 random primers in a RAPD analysis for

**Figure 3**: Schematic representation illustrating the sequence of events during representational difference analysis (RDA), as well as an example of subtraction products obtained after separation on an agarose gel after RDA analysis.
example would scan less than 1% of the same genome. However, a subtractive technology, such as RDA, is inherently subject to several sources of bias. The representation of the genome is based on digestion of genomic DNA with single restriction enzymes. Therefore, the genomic subset obtained, depends on the sequence of the restriction site. Further, tester/driver ratios used for subtractive hybridisation are critical for the elimination of common regions and enrichment of specific sequences. Also, the initial representation is influenced by the size of the restriction enzyme-digested fragments from total genomic DNA, where PCR kinetics favours the amplification of smaller fragments over larger fragments. Since RDA has not been widely applied on genomic plant DNA, major efforts in basic research are still required to develop the technique as a powerful tool for the creation of easily applicable DNA markers.

RDA analysis is currently being performed on plant species that are of importance in Africa. Possible variable DNA regions in the banana genome are investigated with the aim to develop a DNA marker to determine true-to-typeness of micropropagated banana plants (Cullis and Kunert 2000). The use of RDA has also been extended to the detection of sequence differences between date palm varieties and the effect of environmental stress on an inland grass species (wild oat) (Vorster et al. 2002, Kunert unpublished results). This research identified a date palm-specific class of a highly variable repetitive DNA sequences. Another aspect is the detection of genome variation due to plant transformation stress. The aim is to develop a marker for plant transformation and to determine the genetic integrity of genetically modified plants (Van der Vyver unpublished results). A collaborative effort in South Africa also aims to study with RDA possible genome variations in cowpea plants that have been exposed to radiation as a breeding tool for obtaining drought-resistant plants. RDA-based research on genome variation might ultimately allow determination of the biosafety of plants that has been derived from a radiation process applied as a tool in plant breeding in Africa.

**Marker Systems Using Complementary DNA (cDNA)**

These marker systems function on the transcriptome level of the genome and they can be divided into two types. The first marker type makes use of differences between transcriptomes. These systems highlight differences in transcriptomes between different organisms or individuals on the basis of expression. Whereas the second marker type facilitates subtraction of the existing similarities in the targeted transcriptomes, and subsequent isolation of the differences between them. Non-subtractive-based systems most likely target the differences between individuals in abundantly expressed transcriptomes (e.g. housekeeping genes, stress responses), while subtractive-based systems are prone to target low or single copy differences.

**Non-subtractive-based systems**

cDNA-amplified fragment length polymorphisms

The high throughput of cDNA–AFLP (Bachem et al. 1996) makes it an excellent technique to study several samples in conjunction (Figure 4). These samples can be different individuals or a few individuals with several different treatments. The cDNA–AFLP procedure makes use of a large amount of mRNA. This mRNA is first converted into double stranded cDNA and then it is treated exactly as in genomic AFLPs. The choice of enzymes for restriction digest is of importance, as the PCR reaction will display a bias to amplifying smaller fragments. When choosing the enzymes to use, one can use the original set of Msel and EcoRI combined. If some expressed sequence tags (ESTs) knowledge is available for the organism that will be studied, the enzyme selection can be based on this knowledge. The choice of primers will depend on the complexity of the transcriptome studied. cDNA–AFLP is more labour intensive than differential display (Liang and Pardee 1992) and requires more technical experience.

Several cDNA–AFLP profiling studies have been reported from South Africa. In a recent study on cDNA–AFLP profiling of wheat upon Russian wheat aphid (RWA) infestation, using near isogenic wheat lines susceptible ('Tugela') and resistant to Russian wheat aphid ('Tugela DN'), 12 000 transcript derived fragments (TDFs) were generated using ten primer combinations. Four percent of the obtained TDFs were differentially expressed, and are in the process of further analysis (Matsiolo and Botha 2003). These TDFs potentially contain sequences linked to the RWA resistance gene Dn1, which can be used as markers in MAS. However, it may also be useful for the saturation of the genomic region on chromosome 7D S during a map-based cloning approach to obtain the RWA resistance gene Dn1. Other efforts include the analysis of Arabidopsis mutant lines for the identification of TDFs important in wood properties (LW Beyene pers. comm.). It was also used to study differentially expressed TDFs from Vitis vinifera cv. Chardonnay to illustrate the developmentally regulated sequences from ripening grape berry tissue (Venter et al. 2001).

**Subtractive-based systems**

Suppression subtractive hybridisation and representational difference analysis

The two techniques (SSH and cDNA–RDA) using cDNA as starting material follow similar protocols of restriction enzyme digestion of cDNA, DNA subtraction and amplification of subtraction products with PCR. In 1994, Hubank and Schatz adapted RDA for use with cDNA. This technique might be slightly more challenging than SSH because additional adaptor ligation and subtraction steps might be required for RDA. So far, cDNA–RDA has been widely applied in animal but rarely in plant research. When performing cDNA–RDA, the cDNA is usually digested before PCR amplification by a four base recognition sequence restriction enzyme, where DpnI is most commonly used, to produce the representations of two types of cDNAs to be subtracted. cDNA–RDA is currently being used to detect possible uniquely expressed gene(s) in already available weevil-resistant Highland banana plants (Kunert unpublished results).

In 1996, Diatchenko and co-workers proposed the SSH method that combines the power of suppression PCR with
Figure 4: cDNA–AFLP analysis is a technically demanding process, but the number of transcript derived fragments (TDFs) is highly informative, as illustrated after cDNA-analysis of Russian wheat aphid infested wheat (Matsioloko and Botha 2003)
Botha, Venter, Van der Vyver and Kunert

160

the normalisation and enrichment of subtraction hybridisation to enrich for low abundance mRNA transcripts (Diatchenko et al. 1996). In brief the method entails that cDNA from two populations are restricted with a four base cutting enzyme, usually Rsal. The tester population is split into two subsets and two different adaptors are selectively ligated onto these cDNAs (Figure 5). These two subsets are then combined with an excess of the driver cDNA in two separate hybridisation reactions. During these hybridisations the concentration of abundant and less abundant molecules are equalised. After that, the two hybridisation reactions are combined in the presence of an excess of driver and allowed to hybridise further, during which an enrichment of the low abundance molecules occurs. Finally, the difference prod-

Figure 5: Schematic representation illustrating the sequence of events during suppression subtractive hybridisation (SSH), as well as results obtained after SSH analysis in wheat (Botha unpublished results)
ucts are isolated through PCR that enrich only for molecules that contain different adaptors on either end. These products are usually cloned and a quick screen of the ensuing library easily identifies these from the rest of the clones.

Several studies indicated that SSH is useful even when working with complex genomic systems, e.g. hexaploid wheat (Lacock and Botha 2003, Van Niekerk and Botha 2003). Using nuclear isogenic wheat lines, as well as the progenitors of wheat resistant and susceptible to RWA, several expressed sequence tags (ESTs) were isolated with no significant homology (E-value ≤ 10^-3) to any previously known sequence from plants in genomic sequence databases (Van Niekerk and Botha 2003). Moreover, ESTs related to several resistance gene families (RGAs) were also isolated, including a leucine-rich-like fragment and a leucine-zipper-like domain of resistance gene (RGAs) (Lacock and Botha 2003). These ESTs are usually cloned and a quick screen of the ensuing library easily identifies these from the rest of the clones.

Role of ESTs and cDNA Microarrays in DNA Marker Development

Expressed sequence tags (ESTs) and cDNA clones represent the coding regions (transcriptome) within the DNA. These ESTs can be expressed in response to a developmental and/or environmental change. They can be utilised as markers during selection or for mapping purposes (Figure 6). The origin of ESTs is cDNA that is synthesised from mRNA. Several comparative analyses, using ESTs from a variety of material, have been done to enable a better understanding of gene regulation in response to signals. They have also been used successfully for the identification of genes conferring resistance to pests or pathogens (Lacock and Botha 2003), or genes that are differentially regulated during development (Carson and Botha 2000, Carson et al. 2000, 2002). These ESTs can then be studied further using cDNA macroarrays (Carson et al. 2000) or cDNA microarrays (Botha et al. 2003).

Microarray analysis was developed to improve the scientist’s understanding of gene regulation and expression using large sample sizes (Baldwin et al. 1999, Granjeaud et al. 1999). Microarray technology is based on well-known molecular techniques. These techniques underwent a series of modifications to adapt for combination of large datasets from different experiments (Breyne and Zabeau 2001). Microarrays have been described as second-generation dot-blot (Rickett and Dix 1999), since both techniques involve the immobilisation of single-stranded DNA on a solid support (Southern 1975) that is hybridised with single-stranded DNA or RNA populations (Hoheisel 1997, Gerhold et al. 1999, Rickett and Dix 1999). Hybridisation is possible between two such populations since single-stranded DNA or RNA can bind to its single-stranded complement that is immobilised on the solid support (Chee et al. 1996, Brazma et al. 2000).

Today, two methodically distinct methods are employed to assemble microarrays. The first technique, namely oligonucleotide microarrays (Schena et al. 1998), involves the synthesis of oligonucleotides that vary from 10 to 25 base pairs (Thieffry 1999). The synthesis of these oligonucleotides takes place either in situ (Pease et al. 1994, Schena et al. 1998) or the amino acid chains are synthesised separately and then spotted onto a glass support (Marshall and Hodgson 1998). Oligonucleotide microarrays can be used to detect polymorphisms since the sequence of the target fragment can be determined (Hoheisel 1997).

One advantage of microarray analysis is that gene expression and regulation patterns can be monitored on a large scale over several biological experiments and using samples obtained from different species (Quackenbush 2001, Schenk et al. 2000). A further advantage is that the putative function of an unknown gene can be determined when such a gene clusters together with a group of genes of known function (Kerr and Churchill 2001). The technology also enables the screening of large sample sizes for functionality, and if followed by genetic linkage analysis in e.g. F2 populations resulting from narrow or wide crosses, single seeds derived from recombinant inbred lines (RILs) or double haploids (DHs), back cross populations or near isogenic lines (NILs), it opens up opportunities for application in MAS (Figure 6). Results obtained after microarray analysis may also be useful for the identification of quantitative trait loci (QTLs, Jansen and Nap 2001). However, the genes conferring the traits of interest must be identified and cloned, and have to be available for spotting onto the array for analysis.

Recently, during analysis of a custom cDNA microarray that contained ESTs and cDNA clones of wheat (Lacock et al. 2003), flax and banana isolated previously (Cullis unpublished results), several cDNA clones that are either up- or down-regulated in response to RWA feeding were identified (Botha et al. 2003). The expression profiles of selected clones were confirmed using Northern blot analysis and quantitative PCR. The origin of the selected wheat clones and ESTs were confirmed with Southern blot analysis. Moreover, analysis of an NBS–RGA2 clone obtained from this study showed co-segregation at a linkage distance of 3.15cM with the RWA resistance gene Dn1 in a F2 segregating population (‘Tugela’ x ‘Tugela DN’) (Swanepoel et al. 2003), and thus has potential application in MAS.

Molecular Genetic Maps

The new generation of DNA markers focuses on whole genome studies rather than single traits. These markers include the genes conferring the trait in the case of monogenic traits or well-defined QTLs, where additive and complex polygenic traits are studied. Following such approaches further lays the platform for cloning the genes of interest using map-based cloning approaches. A major use for molecular markers is the construction of consensus genetic maps by analysing the co-segregation of markers and phenotypes or traits in defined populations. These populations come from a diversity of sources, e.g. F2 populations resulting from narrow or wide crosses, single seeds derived from
Figure 6: Expression profiling in combination with genetic mapping facilitates the development of markers from genes and gene products on a large-scale.
RILs or DHs, back cross populations or NILs. Markers suitable for such a map include all sequence-derived characters, e.g. SSRs, SCARs, SNPs, ESTs, genes etc. The purpose of a consensus map is to provide a basis for selection of alternative markers in a chromosome region of interest. This is needed in order to identify DNA polymorphisms that are able to track this region when a cross is made during breeding. It also enables the establishment of physical distances for map-based cloning purposes when gene isolation is the objective (Appels 2003).

In our search for 'perfect' markers — markers that are tolerant and informative in diverse genetic backgrounds — the objective will always be the isolation of the genes responsible for specific agronomic traits. One approach to identify and isolate genes conferring genetic traits, is map-based cloning. Several factors will determine the successes in a map-based cloning approach. These include the nature of the genome of interest, whether it is a diploid or polyploid species, the amount of information available regarding the genome, e.g. existing genetic and physical maps. In the past, map-based cloning was limited to diploid species, such as tomato and *Arabidopsis thaliana*, because of the lack of information in plants with more complex genomic structure, e.g. cereal species. However, with recent advances through the development of high-density maps and the possibility to perform chromosome walking on large genomic fragments, the way was opened for the application of this technology to cereal and other large genomes. In recent studies, the genes *Lr21* (Gill et al. 2003) and *Lr10* (Feuillet et al. 2003) conferring resistance to *Puccinia triticana*, as well as Q, the gene largely responsible for the domestication of the wheat species (conferring the free-threshing character and the square headed phenotype, Faris and Gill 2002), were cloned using this strategy.

**Economics and Feasibility of Application**

The advances made through biotechnological tools in genetics and crop improvement worldwide, is undeniable. However, for application of MAS in developing countries, other considerations have to be taken into account. Breeding efforts in developing countries are public domain, and thus function within a limited budget and resource framework. This is in sharp contrast to developed countries where private firms can maximise the net benefits generated through their breeding programmes. Therefore, they can opt for technologies that allow them to bring improved products to the markets faster, even if the process is more costly (Morris et al. 2003). Several other contributing factors can further hamper the application of these technologies in developing countries, such as infrastructure and technical skills. For developing countries to benefit from MAS and other biotechnological tools, the technology needs to be simplified (e.g. presence or absence, and not multiple loci) and accessible in terms of required chemicals and reagents (e.g. inexpensive and available).

**What has been achieved so far in the rest of Africa in DNA marker development?**

Attempts to develop molecular markers in Africa are still limited to a few laboratories. These laboratories have the necessary infrastructure and easy access to chemicals and enzymes required for a development programme for DNA markers. Although a number of studies have been conducted for the development of molecular markers in crops with importance to Africa, most of this research has so far been conducted outside Africa. Clearly, transfer of technical knowledge to Africa is urgently required in order that Africa becomes self-sustainable and starts to benefit fully from recent advances made in plant biotechnology/molecular biology. What has been achieved so far for crops of relevance to Africa? The RAPD and microsatellite techniques have been successfully applied to study the genetic behaviour and structure in banana regarding the fingerprinting of hybrids and the development of marker-assisted breeding systems (Vuylstke et al. 1998). For cassava, Fregene et al. (1996) already developed 132 RFLP, 30 RAPD and 3 microsatellite markers. In addition, Chavariagga-Aguirre and colleagues (1999) could identify 32 microsatellite markers from which 22 DNA primers have been developed for *Musa*. For maize, a staple crop in southern Africa, molecular maps have been generated using RFLPs, RAPDs, AFLPs and microsatellites. Molecular mapping techniques have been applied to a variety of traits of importance in maize including turgicium blight resistance (Schechert et al. 1999), anthesis-silking interval (Ribaut et al. 1996), and grey leaf spot disease resistance (Bubeck et al. 1993). For sorghum and pearl millet, several studies focused on the development of molecular markers for identification of various disease resistance traits, such as downy mildew, anthracnose, leaf blight, head smut and grain moulds (Magill et al. 1997, Hash et al. 1997).

**Conclusion**

Linked markers, if validated properly, can add great value to breeding programmes by saving on time and easing selection. However, the ‘perfect’ markers will still be the genes conferring the trait of interest. And thus, worldwide efforts have been increased to achieve this goal. To date, genes conferring resistance to leaf rust (*Puccinia triticana*) *Lr10, Lr21* and the Q gene responsible for free-threshing have been cloned from economically important cereal species, and there are expectations that many more will follow. These have mostly been done via map-based cloning approaches, and using these sequences as markers proved extremely useful in the ongoing breeding efforts. However in many of the African laboratories several of the abovementioned technologies may not be feasible (e.g. most cDNA-based technologies), mainly due to costs involved, and lack of available technical skills and infrastructure.
References


Faris JD, Gill BS (2002) Genome targeting and high-resolution mapping of the domestication gene Q in wheat. Genome 45: 706–718


Gerhold D, Rushmore T, Caskey CT (1999) DNA chips: promising tools have become powerful tools. Techniques in Biological Science 24: 168–173


Harvey M, Botha FC (1996) Use of PCR-based methodologies for the determination of DNA diversity between Saccharum varieties. Euphytica 89: 257–265

Station. Research towards efficient and sustainable production. South African Sugar Association Experiment Station, Mount Edgecombe, KwaZulu-Natal, pp 67–69


Edited by CH Bornman