THE APPLICATION OF MICROSATELLITES TO SUGARCANE PARENTAGE DETERMINATION AND VARIETAL IDENTIFICATION

by

SIMON MATTHEW HACK

Submitted in fulfilment of the academic requirements for the degree of Master of Science in the School of Life & Environmental Sciences, University of Natal, Durban, South Africa.

December 2002

ABSTRACT

The use of microsatellite markers has matured and become commonplace for plant genome analyses and is now poised for widespread practical application in sugarcane. Sequence Tagged Microsatellite Site (STMS) amplification is the most prevalent microsatellite-based approach and involves the amplification of a microsatellite by designing primers that flank and hence define the microsatellite site, revealing variation in the length of repeat motifs between individuals. Twenty-six microsatellite primer pairs received from the International Sugarcane Microsatellite Consortium (ISMC) were evaluated and the STMS protocol was optimised to ensure robust and reproducible results. The objectives of this study were to use STMS for sugarcane parentage analysis and fingerprinting. Previously, Restriction Fragment Length Polymorphism (RFLP) marker data had suggested that the parentage of a genetic mapping population, sugarcane cross AA40 (N18 x CP57/614), was incorrect. Based on the assertion that the incorrect parentage was as a result of either mislabelling at planting or at seed collection, microsatellite parentage analysis was carried out on eight potential parent pairs (13 cultivars). A total of 75 markers were scored with non-parental bands (12 on average) being observed for all of the potential parent pairs and none could be identified as the true AA40 parents. It has been suggested in other plant species that PCR artefacts could give rise to non-parental bands and to investigate this the marker data of single parent DNA reactions and pooled parent pair DNA reactions or 'synthetic offspring' were compared. The results suggested that either a certain percentage of non-parental bands, perhaps 10% (maximum value observed), should be tolerated in microsatellite parentage analysis or a marker should only be considered to be discriminating for parentage if it is absent in both the parents and the pooled parent pair amplifications. Fingerprinting of 20 cultivars using 14 microsatellite primer pairs was conducted to evaluate the potential of the STMS approach for sugarcane varietal identification. It was found that only two microsatellite primer pairs were

required to discriminate between all 20 cultivars with a theoretical number of non-differentiated pairs of cultivars (X_K) of only 0.03. This estimator was used to determine the approximate number of microsatellites necessary for large-scale sugarcane fingerprinting.

PREFACE

The experimental work described in this dissertation was conducted in the Biotechnology Department of the South African Sugar Association Experiment Station (SASEX), from March 2001 to December 2002, under the supervision of Dr BI Huckett (SASEX) and co-supervision of MK Butterfield (SASEX).

These studies represent original work by the author and have not otherwise been submitted in any form for any degree or diploma to any tertiary institution. Where use has been made of the work of others it is duly acknowledged in the text.

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENT

My sincere thanks to my supervisors Dr Barbara Huckett and Mike Butterfield for their guidance, encouragement and understanding throughout the duration of this study.

The South African Sugar Association Experiment Station and the Sugar Industry Trust Fund for Education for financial support.

Sandy Snyman, Gwethlyn Meyer, Derek Watt and Deborah Carson for their friendship and advice.

Finally and most importantly, thanks to my parents and Vanessa for their love and support.

CHAPTER 1

GENERAL INTRODUCTION

Sugarcane (*Saccharum* spp. hybrids) is grown in over 127 countries and is one of the staple sources of sweetening agents in the world; as such it is a very important crop from an agri-economic perspective (Rash, 1995; Naik, 2001). The economic importance of the sugarcane crop is not signified by its share in the world gross cropped area, which is a little over 1%, but rather as a valuable money crop being used to produce raw and refined sugar, syrups, specialized sugars and a range of by-products, notwithstanding its value as an important source of foreign exchange (Naik, 2001).

In South Africa the sugar industry produces an estimated average of 2.5 million tons of sugar per season; with about 50% of this sugar being exported to markets in Africa, the Middle East, North America and Asia. Sucrose is South Africa's third most important agricultural export, and contributed R1.9 billion to the country's foreign exchange from the 2000/2001 season (Anon., 2001; Naik, 2001).

Sugarcane cultivation and breeding for the development of superior breeding lines or cultivars has a long international history (Huckett & Botha, 1995; Butterfield et al., 2001), and in South Africa the South African Sugar Association Experiment Station (SASEX) is responsible for producing new commercial sugarcane cultivars for the local industry. The cultivars released are selected primarily for their sucrose yield but phenotypic characters such as disease and pest resistance and tolerance to adverse environmental conditions are also considered (Butterfield & Thomas, 1996).

Detailed records of most of the crosses contributing to present day varieties do exist. However, knowledge of the parents is subject to uncertainty due to the

method of crossing, which has the capacity to result in illegitimate pollination (Huckett & Botha, 1995). In addition, management of sugarcane collections is labour intensive and susceptible to error because of the regular replanting required. As a result errors have been detected in sugarcane breeding programmes (Jannoo et al., 2001) and even in international sugarcane germplasm collections (Cordeiro, 2001).

In the SASEX breeding programme, sugarcane clonal identification is based on morphological characters such as the auricle, bud wing and stalk colour (Mike Butterfield, 2002, pers comm.). However, the assessment of morphological characters is often difficult and unreliable, particularly for closely related cultivars, because of (1) the subjectivity in the analyses of characters, (2) the influence of environmental or management practices on the character and (3) the limited diversity among cultivars with a highly similar pedigree (Morell et al., 1995; Piperidis et al., 2001).

The problems associated with phenotypic identification have fostered the need for an accurate, fast, reliable and cost effective genetic method to distinguish between sugarcane varieties, verify pedigrees and estimate genetic diversity. Moreover such a technique would provide impetus for the development of new and superior cultivars. The majority of molecular markers do not suffer from the same limitations as phenotypic characters and as such could complement and facilitate the breeding programme (Kolliker et al., 2001), assign unique genetic fingerprints to varieties (Rongwen et al., 1995; Hokanson et al., 1998) and provide information on molecular-based genetic relationships (Struss & Plieske, 1998).

Microsatellites, or simple sequence repeats (SSRs), are molecular markers that have become increasingly popular for the molecular characterisation of different plant species (McCouch et al., 1997; Wunsch & Hormanza, 2002). Sequence Tagged Microsatellite Site (STMS) amplification being the most common microsatellite-based approach for genome analysis. Microsatellites are DNA

sequences composed of tandemly repeated nucleotide units of 1 to 6 base pairs (Garland et al., 1999), flanked by specific sequences that in the STMS approach are used to design PCR primers (Becker & Heun, 1995). In STMS, amplified allelic differences between genotypes are referred to as simple sequence length polymorphisms (SSLPs) and are usually the result of variable numbers of repeat units within the microsatellite sequence (McCouch et al., 1997). Microsatellites are valuable genetic markers because they are co-dominant, reproducible, able to detect high levels of allelic diversity and are easily and economically assayed by the polymerase chain reaction (PCR) (Liu et al., 1996; McCouch et al., 1997; McGregor et al., 2000).

In soybean (*Glycine max)*, SSRs have been shown to provide a higher incidence of detectable polymorphism than other techniques such as Restriction Fragment Length Polymorphism (RFLP) and Random Amplified Polymorphic DNA (RAPD) (Powell et al., 1996b). Jones et al., (1997) showed that both Amplified Fragment Length Polymorphism (AFLP) and STMS amplification in tomato (*Lycopersicon esculentum*) cultivars were more reproducible between and within laboratories than RAPDs or RFLPs. The success of using these markers in other crops species like barley, *Hordeum vulgare* (Saghai Maroof et al., 1994; Russell et al., 1997); rice, *Oryza sativa* (Yang et al., 1994; Garland et al., 1999); sorghum, *Sorghum vulgare* (Brown et al., 1996); wheat, *Triticum aestivum* (Roder et al., 1995; Fahima et al., 2002) and grape, *Vitis vinifera* (Sanchez-Escribano et al., 1999) have encouraged the development of microsatellites for sugarcane.

The high cost of characterising microsatellites is prohibitive to their use, so an International Sugarcane Microsatellite Consortium (ISMC) was instigated in 1998 to isolate SSRs and develop microsatellite primer pairs. Over a two-year period more than 200 sugarcane microsatellite primer pairs were developed and distributed to all the members, SASEX included (Cordeiro et al., 2000). However, the usefulness of these microsatellite primer pairs for genome

analysis still needed to be assessed in the context of individual breeding programmes, including the one at SASEX.

There were three major objectives in this study:

1. To evaluate a subset of the microsatellite primer pairs received from the ISMC and optimise the STMS protocol to allow for robust and reproducible results. The optimisation involved the testing of various PCR parameters and gel electrophoresis systems (Chapter 3).

2. To utilise this optimised STMS approach for investigating the parentage of the AA40 population and determining the innate incidence of non-parental bands and their impact on parentage analysis (Chapter 4).

3. To investigate the usefulness of STMS for sugarcane varietal identification (fingerprinting), as well as test the reproducibility of the results and sensitivity of this approach to DNA quality and origin (Chapter 5).

CHAPTER 2

LITERATURE REVIEW

2.1 SUGARCANE

Knowledge of the origin, genetics and breeding of modern sugarcane is important in understanding how these affect and challenge the use of molecular markers, such as Sequence Tagged Microsatellite Site (STMS) amplification, for sugarcane genome analysis.

2.1.1 Taxonomy and Origin of Sugarcane

Sugarcane is a large perennial grass (Lu et al., 1994b) cultivated in tropical and intertropical regions. It belongs to the genus *Saccharum* L. of the family *Poaceae*, which is placed in the tribe *Andropogoneae* (Jannoo et al., 1999) together with the genera *Zea* and *Sorghum* (Lu et al., 1994b). The genus *Saccharum* is characterised by both a high ploidy level and aneuploidy, and formally comprises six species: *S. spontaneum*, *S. robustum*, *S. officinarum*, *S. barberi*, *S. sinense* and *S. edule*. Despite their complex genomic structure, the *Saccharum* species (except *S. edule*) are female fertile and often male fertile (D'Hont et al., 1994).

S. officinarum, a cultivated species with 2n=80 euploids (Bremer, 1961), is the primary source of genes for sucrose accumulation and as such is referred to as the 'noble' cane (Bremer, 1930; Li & Price, 1967; Lu et al., 1994b). It is thought to be derived from 2n=60-200 forms of the wild species *S. robustum* in New Guinea, the centre of its origin (Brandes, 1958). The involvement of several sources of *S. robustum* could explain the large nuclear genetic diversity observed among *S. officinarum* clones, for the diversity within *S. robustum* is larger than that within *S. officinarum* (Daniels & Roach, 1987; Lu et al., 1994a).

S. spontaneum is a wild species that shows great variability and a wide range of chromosome numbers from 2n=40-128 (Sreenivasan et al., 1987). It produces little sugar but is adaptable to different environments, resistant to several diseases and shows high ratooning ability (Jackson, 1994). It has a large distribution from Japan and New Guinea to the Mediterranean and Africa, with India as its centre of origin (Daniels & Daniels, 1975; Lu et al., 1994b).

The other three species are probably of a secondary origin. *S. barberi* is a cultivated species indigenous to India, with 2n=81-124 aneuploids. *S. sinense* is a cultivated species indigenous to China, with 2n=111-120 aneuploids (Price, 1968; Lu et al., 1994b). These two species are derived from natural hybridisations between *S. officinarum* and *S. spontaneum* in India and China respectively (Price, 1968; D'Hont et al., 1994). This interspecific hybrid origin has been confirmed by RFLP using single copy maize nuclear probes (Lu et al., 1994b) and more recently by genomic in situ hybridization (GISH) (D'Hont & Paulet, 2000). *S. edule* (2n=60, 70, 80, and aneuploid clones) is a minor group of sterile canes, which are cultivated as a traditional vegetable in Melanesia (Lu et al., 1994b), and are supposedly derived from intergeneric hybridisation (Daniels & Roach, 1987; Burner, 1991).

Other allied genera, such as *Erianthus* (section *Ripidium* 2n= 20, 30, 40 and 60 euploid clones), *Miscanthus* (section *Diandra* 2n= 38, 40 and 76 euploids), *Narenga* (typically 2n= 30) and *Sclerostachya* (Hack.), are closely related to the *Saccharum* genus and constitute with it an interbreeding group that is termed the '*Saccharum* complex' (Daniels & Roach, 1987; Lu et al., 1994b). Members of this '*Saccharum* complex' are thought to have arisen through polyploidisation and hybridisation events (Jannoo et al., 1999).

2.1.2 Genome and Genetics of Sugarcane

The genus *Saccharum* is characterised by both a high ploidy level and aneuploidy, with relatively large genomes (Da Silva et al., 1993; Ha et al., 1999; Hoarau et al., 2001). The elevated ploidy levels and cytogenetic complexity of

interspecific hybrids, combined with the difficulty of controlled hybridisation, have limited classical genetic studies in sugarcane (Da Silva et al., 1993; Hogarth et al., 1987; Butterfield et al., 2001). Furthermore, it is often not possible to identity as many distinct alleles as the ploidy level could allow for a given locus: some alleles are simplex (only one copy is present at the locus), others duplex (two copies) or multiplex (multiple copies) (Grivet et al., 1996).

The basic chromosome number of *Saccharum* has been variously defined (Sreenivasan et al., 1987) as x = 5, 6, 8, 10 or 12 and may vary with species. It is likely that x = 10 is the basic number in *Saccharum* because this number predominates in Andropogoneae (Burner, 1991; Whalen, 1991). Recent cytogenetic studies have clarified our knowledge of the basic chromosome numbers in *Saccharum* genus, giving x=10 for *S. officinarum* and x=8 for *S. spontaneum* (D'Hont et al., 1998). For these two species an autopolyploid origin is suspected, although mainly bivalents are observed at meiosis and no close diploid relatives are known (Celarier, 1956; D'Hont & Grivet, 1996).

Nair (1975) observed that persistence of high chromosome number in present-day sugarcane cultivars indicates selective advantage of *S. officinarum* and *S. spontaneum* genetic contributions, despite opportunity for chromosome loss through meiotic irregularity. Bremer (1961) theorised that chromosome number may affect the size of sugarcane plants.

While chromosome numbers have been determined for many *Saccharum* clones, there is comparatively less data on meiotic chromosome pairing. Most euploid clones of *Saccharum* species tend to be meiotically regular with predominantly bivalent chromosome pairing (Sreenivasan et al., 1987). However, Sreenivasan & Jagathesan (1975) found heteromorphic bivalents, univalents and numeric aberrations in most of the 28 clones of *S. spontaneum* studied (2n=40-126), although multivalents were rare. The presence of heteromorphic bivalents and multivalents may indicate homoeologous chromosome pairing (Burner, 1991). It is suspected that meiotic irregularities of

Saccharum clones might be due to ploidy level, and Barnett & Carver (1967) showed that the frequencies of univalents increased with ploidy level in naturally occurring tetraploid, hexaploid and octoploid plants of *Panicum virgatum* L.
The haploid genome size of a range of anonymous *Saccharum* species has been reported as varying between 2547 and 4183 Mbp (Arumuganathan & Earle, 1991). In polyploids, the haploid chromosome number is not the same as the monoploid. The monoploid genome size for *S. officinarum* (x=10) would thus be approximately 926 Mbp and for *S. spontaneum* (x=8) 760 Mbp (Butterfield et al., 2001).

2.1.3 History of Sugarcane Cultivation

Until the end of the 19th century, *S. officinarum,* together with *S. barberi* and *S. sinense*, provided most of the commercially grown cultivars. Sugarcane breeding was limited to their collection and evaluation from native gardens, while their multiplication and propagation were performed exclusively by means of stem pieces (Lu et al., 1994a).

The discovery of sexual fertility in 1888 in Java, and the ravages caused by the 'sereh' disease at the same period, stimulated the first man-made interspecific hybrids to be produced in Java and India (Lu et al., 1994a), involving essentially *S. officinarum*, *S. spontaneum* and *S. barberi* (Lu et al., 1994b; Grivet et al., 1996). Restoration of the high sugar producing type was obtained through repeated backcrossing of the hybrids to *S. officinarum,* and this also resulted in minimising the negative effect of the wild parent. This breeding procedure is referred to as 'nobilisation' in sugarcane (Lu et al., 1994a).

Nobilisation is characterised by asymmetric chromosome transmission (Bremer, 1961), and during this procedure an *S. spontaneum* clone pollinates a *S. officinarum* clone. The female parent (*S. officinarum*) transmits 2n=80 chromosomes to the F_1 (2n=100-144), whereas the wild male parent (*S. spontaneum*) transmits n=20-64 (Bremer, 1961). When this F_1 hybrid is backcrossed to *S. officinarum*, the same phenomenon of 2n+n transmission

occurs in the BC_1 (2n=130-152) hybrids. Endoduplication or fusion of two nuclei after the second meiosis division have been suggested by Bhat & Gill (1985) to explain this peculiar chromosome transmission. It is from the BC_2 onward that the chromosome transmission becomes normal (Lu et al., 1994a). Therefore, nobilised clones are characterised by a high chromosome number (2n=100-130), with roughly 80% of the genome derived from *S. officinarum* and the remainder from *S. spontaneum*, whether directly or via *S. barberi* (Price, 1957; Roach, 1969). The most important part of the diversity among the varieties is due to the alleles brought by *S. spontaneum* (Lu et al., 1994a).

These interspecific hybrids were proven to be major breakthroughs in sugarcane improvement; solving some of the disease problems and presenting an unexpectedly high yield, high ratooning and adaptability (Roach, 1972). Modern sugarcane cultivars are derived largely from intercrossing of these first nobilised hybrids. However, as only a few clones of these species were used, the narrow genetic base or limited degree of diversity of modern hybrid varieties is one of the principal causes of the present slow rate of sugarcane breeding progress (Berding & Roach, 1987; Roach, 1989; Deren, 1995).

2.1.4 Modern Sugarcane Breeding

The Plant Breeding Department of the South African Sugar Association Experiment Station (SASEX), at Mount Edgecombe, is responsible for producing the commercial sugarcane varieties grown in South Africa. The principle aim is to produce the most profitable sugarcane varieties for the South African industry's climatic and environmental conditions, in terms of recoverable sugar per hectare, and for these varieties to have adequate resistance to important diseases and pests (Natalie Coetzee, 2002, pers. comm.).

In the breeding programme, parent varieties are chosen based on a number of different criteria. The traits of interest include high sucrose, good agronomic characteristics and freedom from diseases and pests. All five of the agro-climatic regions of the South African sugar industry have a separate breeding

scheme with parents adapted to that particular local climate and disease spectrum. Parent varieties are chosen from local varieties, imported varieties from other countries and from wild germplasm. These wild species include vigorous, low sucrose *S. spontaneum* clones and poor-growing, high sucrose *S. officinarum* clones. These species widen the genetic base of the local sugarcane germplasm, providing novel sources of disease resistance and important agronomic traits (Natalie Coetzee, 2002, pers. comm.). The parent varieties are placed in glasshouses where the optimal photoperiod and temperature for flowering is maintained, and once flowering has occurred selected parents are paired off to allow for cross-pollination.

Table 2.1 Simplified layout of the SASEX selection programme. The varieties are divided between six selection farms, where evaluation is carried out independently in stage 1 to 4. At stage 5 the best varieties from each site are combined and planted across different sites (After Butterfield & Thomas, 1996).

Selection Stage	Year	Trial Type	Number of varieties
	1	TERRACE Potted seedlings Plant crop	250000
Stage 1	2	SINGLE STOOLS Replication of families Plant crop	175000
Stage 2	3-4	SINGLE LINES One row, no replication Plant crop + 1st ratoon	20000
Stage 3	5-6	OBSERVATION TRIAL Two rows, two replications Plant crop	2000
Stage 4	7-9	PRIMARY VARIETY TRIAL 5 rows, three replications Plant crop + two ratoons	300
Stage 5	10-13	SECONDARY VARIETY TRIAL 5rows, three replications Plant + two ratoons Trials at three regional testing sites	100
	14-15	BULKING UP	1-2

Every year 250,000 candidate varieties, obtained from seed produced after cross-pollination between selected parents, enter into the selection programme. These potential new commercial varieties then undergo a five stage selection programme, which takes between 12 and 15 years to complete and involves several trials over either different years and/or locations (Mike Butterfield, 2002, pers. comm.). During the selection process (Table 2.1), clones are chosen for their sucrose content, sucrose yield per hectare and resistance to the many different diseases and pests that occur in the industry (Butterfield & Thomas, 1996).

2.2 MOLECULAR MARKERS FOR PLANT GENOME ANALYSIS

2.2.1 Background

The rapid development of molecular techniques, over the last few decades, now offers a palette of technical approaches for plant genotyping or genome analysis. Which technique is most appropriate depends upon (1) the extent of genetic polymorphism required, (2) the analytical or statistical approaches available for the technique's application and (3) the pragmatics of time and costs of materials (Parker et al., 1998). The discovery of the polymerase chain reaction (PCR) was a landmark in molecular marker evolution and has proved to be a unique process for the development and utilisation of a battery of new very sensitive and quick approaches, such as AFLP or microsatellites (Paglia & Morgante, 1998; Koreth et al., 1996).

A reproducible and informative molecular marker system has application in the following areas (Cordeiro, 2001; Lee & Henry, 2001):

- Ensuring field grown cane is true to type;
- Determination of genetic diversity in commercial sugarcane cultivars;
- Management of breeding programs through marker assisted selection;
- Determination of genetic diversity of parents in breeding programs (heterosis);
- Confirmation that parents selected in breeding programs are true to type;
- Protection of plant breeders' rights.

In this chapter the principle, the methodology, the level of polymorphism and the limitations of various molecular markers will be discussed. These approaches include isozyme analysis, Restriction Fragment Length Polymorphism (RFLP), Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP) and microsatellite analysis.

2.2.2 Isozyme / Allozyme Analysis

The term allozyme refers to different allelic forms of nuclear-encoded enzymes, whereas isozyme is a more robust term referring to different biochemical forms of an enzyme identified by electrophoresis (Parker et al., 1998).

The premise of isozyme or allozyme analysis is that a tissue extract is electrophoresed and in so doing the enzymes are separated by size, shape and/or charge along an electrical gradient (Parker et al., 1998). Enzyme-specific stains are then used to visualise the resulting electromorph bands, which can be from one to several bands depending on the number of loci, their state of homo- or heterozygosity, and the enzyme molecule configuration (Weising et al., 1995).

Obtaining allozyme data is relatively inexpensive and straightforward once the basic procedure has been perfected for a given species (Lebot & Aradhya, 1992). To provide suitable statistical confidence for many applications of allozymes, however, requires at least 10-20 independently segregating polymorphic loci (more than 2 alleles) (Parker et al., 1998). Unfortunately some species are monomorphic for most allozymes that can be screened by standard procedures, due to the redundancy of the genetic code and the similarity of certain amino acids (Weising et al., 1995).

In sugarcane, isozymes have been used to differentiate between wild and noble canes and to show progeny-parent relationships (Glaszmann et al., 1989). Isozyme research with sugarcane has also revealed that most of the diversity within sugarcane varieties is related to the presence or absence of *S. spontaneum* genes (Eksomtramage et al., 1992). However, Glaszmann et al. (1989) found that the use of isozymes in sugarcane is often beset with practical difficulties due to the high number of bands that may migrate at similar distances and the occurrence of multiple bands of unequal intensities, both of which arising due to the high ploidy of sugarcane. This is further compounded by the fact that isozyme detection is often weak and unreliable, and consequently may produce different results in different laboratories (Gallacher et al., 1995). In addition, their wider application is limited by the requirement of local cultivation (Eksomtramage et al., 1992), their sensitivity to environmental conditions and management practices (Dawson et al., 1993) and their tissue specificity (Godwin et al., 2001).

2.2.3 Restriction Fragment Length Polymorphism (RFLP)

Restriction fragment length polymorphisms (RFLPs) were originally used in man (Botstein et al., 1980), but were rapidly adopted as a plant DNA marker (Beckmann & Soller, 1986). They are codominant, simply inherited and naturally occurring Mendelian characters; which exhibit environmental stability and nearly unlimited availability, making them a useful tool for genome analysis (Graner et al., 1990,1991). In RFLP the genetic material itself is screened and the same

RFLPs will be detected in DNA isolated from organs and tissues in the plant, irrespective of the age of the particular tissue. This is a significant advantage over biochemical tests, such as isozymes, that assay gene products (Beckmann & Soller, 1986; Ainsworth & Sharp, 1989).

RFLP involves firstly the extraction of genomic DNA, followed by its digestion with specific restriction endonucleases, which cut the DNA into fragments (Morell et al., 1995). An RFLP results when variation in restriction enzyme cleavage sites, arising due to base substitutions, insertions, deletions or translocations in the genomic DNA (Gupta et al., 2002), is detected by Southern hybridisation using either a pre-exiting probe for a specific gene from a closely related species or a probe generated for a specific sequence that occurs in the region of interest (Ainsworth & Sharp, 1989; Parker et al., 1998). It is a common practise to screen a number of probes for RFLP analysis, as well as to utilise an array of restriction enzymes to determine the most suitable combination. If two individuals differ at a restriction site this will affect the length of a particular DNA fragment, homologous to the probe, bringing about a screenable polymorphism. In this way, a restriction site polymorphism at the DNA level is detected as a restriction fragment length polymorphism (Beckmann & Soller, 1986; Weising et al., 1995).

In sugarcane, RFLPs have been used to show a strong molecular differentiation between *S. officinarum* and *S. spontaneum* (Lu et al., 1994b, Jannoo et al., 1999), and that the major part of the diversity among sugarcane cultivars arises solely from the *S. spontaneum* chromosomes (Lu et al., 1994a). Moreover, RFLP maps have been or are being constructed for numerous crop plants (Graner et al., 1991), including sugarcane (D'Hont et al., 1994; Grivet et al., 1996), to assess genetic variability (characterisation of germplasm stocks), determine correlations between RFLP markers and qualitative or quantitative traits (Tang et al., 2000; Xu et al., 2002) and in some cases to maximise the benefits of marker-assisted selection or elucidate phylogenetic relationships (Debener et al., 1990; Da Silva et al., 1993; Besse et al., 1997).

Nevertheless, RFLP research is most often hampered by the requirement of large quantities of DNA, by the fact that Southern hybridisation is time-consuming and expensive and lastly by the lack of suitable probes (Ainsworth & Sharp, 1989). Consequently, it has not been widely adopted for fingerprinting purposes (D'Hont et al., 1994; Parker et al., 1998).

2.2.4 PCR-Based Techniques

PCR is an *in vitro* technique that allows the amplification of a specific DNA region that lies between two regions of known DNA sequence (Saiki et al., 1985). This technique allows for the amplification of any DNA sequence of interest to high copy numbers, thereby circumventing the need for molecular cloning (Erlich, 1989; Weising et al., 1995). Some of the advantages of PCR-based marker systems are that: (1) PCR requires only small amounts of DNA, and often crude miniprep procedures yield DNA of sufficient quantity and quality; (2) PCR is relatively quick to perform and technically straightforward, once PCR conditions have been established and (3) the range of primer sequences possible gives PCR-based techniques great diagnostic power (Morell et al., 1995).

Various molecular techniques have developed utilising PCR, namely Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP) and microsatellites or Simple Sequence Repeats (SSR).

2.2.4.1 Random Amplified Polymorphic DNA (RAPD)

The PCR-based technique, Random Amplified Polymorphic DNA (RAPD) is conceptually simple and RAPD markers are produced by PCR using random oligonucleotide primers, usually decamers, of known sequence (Williams et al., 1990). RAPD primers are used singly, not in combination with a second primer as in standard PCR (Erlich, 1989), and amplification occurs when the same sequence, complementary to the primer, is present in inverse orientation within an amplifiable distance (Gupta et al., 2002). The amplification products are

generally resolved on agarose gels, and different RAPD patterns (polymorphisms) arise when genomic regions vary for the presence or absence of complementary primer annealing sites (Tinker et al., 1993). These markers serve as dominant genetic markers, which are inherited in a Mendelian manner (Welsh et al., 1991; Dawson et al., 1993).

RAPD analysis has many potential applications, and may be used to assess kinship relationships (Tinker et al., 1993) or genetic diversity (Dawson et al., 1993), construct genetic maps (Saliba-Colombani, 2000) or create specific probes (Xu et al., 1995). The main advantages of the RAPD technology include: (1) suitability for work on anonymous genomes, (2) applicability to problems where only limited quantities of DNA are available and (3) efficiency and low expense (Hadrys et al., 1992).

In sugarcane, RAPDs have been used to evaluate genetic diversity in 20 commercial sugarcane hybrids as well as between members of the *Saccharum* complex (Harvey & Botha, 1996); to resolve taxonomical groups in sugarcane cluster analyses (Nair et al., 1999) and to determine correlations between certain RAPD markers and resistance to diseases, such as sugarcane mosaic virus (Barnes et al., 1997).

There are various limitations and considerations in RAPD analysis. Firstly, the primer size will determine the degree of specificity in genome scanning i.e. short length will result in an unreasonably large number of sequences, while larger primers will amplify too few sequences to be informative. Increasing primer length may also increase non-specific primer annealing and hence reduce reproducibility (Hadrys et al., 1992; Siles et al., 2000). Secondly, RAPD analysis can be sensitive to PCR reaction conditions because of the low-stringency annealing temperatures and short length of the primers used (Carlson et al., 1991; Klein-Lankhorst et al., 1991). However Williams et al., (1990) found that if the RAPD amplification was repeated two or more times, the majority of markers were reproducible and scorable. Thirdly, RAPD markers are generally short and may lead to artifactual amplification products, arising from non-

specific priming or from heteroduplex formation between related amplification products, that results in unclear and non-reproducible fragments (Klein-Lankhorst et al., 1991; Pan et al., 1997).

Consequently, the analytical power of RAPD markers is not competitive with analyses using sequence information or single locus probe fingerprinting technologies; and as such is not suitable for applications such as extensive fingerprinting projects (Siles et al., 2000) or definitive parentage determination.

2.2.4.2 Amplified Fragment Length Polymorphism (AFLP)

The AFLP method was originally developed as a universal DNA fingerprinting method (Zabeau & Vos, 1993; Vos et al., 1995) and is robust and relatively insensitive to reaction conditions. Consequently, reproducibility is high (Jones et al., 1997; McGregor et al., 2000) and genetic background is less likely to result in artifactual polymorphisms (Williams et al., 1990; Maughan et al., 1996; Mueller & Wolfenbarger, 1999). AFLP analysis is able to assay a larger number of DNA loci, to reveal more polymorphic bands in one gel lane, than RAPDs, RFLPs or microsatellites (Cho et al., 1998; Saliba-Colombani et al., 2000); yet of these techniques AFLP does not offer the highest level of polymorphism (Pejic et al., 1998; Lima et al., 2002).

The AFLP technique basically consists of three steps: (1) digestion of total cellular DNA with two restriction enzymes and ligation of restriction halfsite-specific adaptors to all restriction fragments; (2) selective amplification of only a subset of the restriction fragments with two PCR primers that have corresponding adaptor- and restriction-site-sequences as their target sites and (3) electrophoretic separation of the PCR products on a denaturing polyacrylamide gel (Janssen et al., 1996). When the AFLP method is applied to complex genomes, like plants, two cycles of selective amplification are performed. In the first step, named preamplification, the genomic DNA is amplified with AFLP primers both having a single selective nucleotide. This pool of PCR products is then amplified with primers both having three selective

nucleotides. This two-step amplification process reduces the amplicons to a manageable number, with only 1 out of every 4096 possible amplicons being amplified (Vos et al., 1995). Segregation analysis and linkage studies indicate that AFLP markers are inherited in a Mendelian manner (Maughan et al., 1996), but AFLP markers cannot distinguish heterozygotes and homozygotes, and as such are classed as dominant markers (Bradshaw et al., 1998).

The basic difference between RFLP and AFLP polymorphisms is that for RFLPs an area is scanned that is defined by the number of nucleotides in the restriction sites; whereas for the AFLP technique an additional number of nucleotides defined by the selective nucleotides is scanned. Therefore, it is expected that AFLP makers will detect more point mutations per 100 nucleotides than RFLPs, but should detect more or less the same frequency of insertions or deletions (Becker et al., 1995).

AFLPs have been used for the analysis of genetic diversity (Russell et al, 1997), DNA fingerprinting (Powell et al., 1996b), the construction of linkage-maps (Becker et al., 1995; Cho et al., 1998; Hoarau et al., 2001) and to locate traits of interest or track their transmission (Gupta et al., 1999; Hartl et al., 1999). For genome analysis in sugarcane, AFLPs have proved highly informative as a means of elucidating genetic diversity (Lu et al., 1994a; Lima et al., 2002) and Besse et al., (1998) and Hoarau et al., (2001) have used AFLP markers for mapping in sugarcane and were able to reveal the major *Saccharum* complex groups. Their results were in accordance with those obtained by RFLP markers (Lu et al., 1994b). This indicates that, with respect to sugarcane, AFLPs have the resolving power to analyse relationships within and between *Saccharum* species, with the additional benefit of being a multi-locus marker system (Cordeiro, 2001).

However, although AFLP is a powerful molecular marker, some reproducibility issues have been raised in sugarcane, and it is believed that these stem from: (1) partial digestion of the template DNA as a result of insufficient enzymatic

conditions or due to unexpected or inconsistent methylation of template DNA; (2) poor amplification of PCR fragments or (3) DNA contamination (Cordeiro, 2001).

2.3 MICROSATELLITE MARKERS

"History sometimes takes ironical twists and the history of science is no exception. Microsatellites have been detected in eukaryote genomes for over 15 years, although they were regarded as sequences of no particular interest. It was realised in the late 1980s with the rise of PCR that microsatellites may be the most powerful Mendelian markers ever found" (Jarne & Lagoda, 1996).

The various aspects of microsatellites, from their origin to their development and utilisation in various plant species, will be discussed in the following sections. This should provide background and a solid understanding of the complexities of microsatellites; while also giving an indication of the status of microsatellite development and utilisation.

2.3.1 Structure and Organisation

Microsatellites, or simple sequence repeats (SSRs), are hypervariable DNA sequences (Wilder & Hollocher, 2001) composed of stretches of monotonously repeated short nucleotide motifs that are arranged head-to-tail (Hancock, 1999) within euchromatin (Koreth et al., 1996). Almost all permutations of di- and tri-nucleotide motifs, but also several longer ones, can be found as building blocks of simple sequence repeats (Koreth et al., 1996).

The most polymorphic and therefore the most useful for many applications are uninterrupted microsatellites (Weber, 1990), but many microsatellite loci amplified by PCR contain interruptions. Interruptions within the core sequence seem to stabilise arrays of repeats, making interrupted microsatellites less variable than pure ones (Goldstein et al., 1995). Microsatellites may also be

compound or made up of contiguous or adjacent tandem arrays of different motifs (Hancock, 1999) (Table 2.2).

Table 2.2 Ten-repeat dinucleotide arrays showing the three families of microsatellites, namely pure, compound and interrupted repeats.

Type	Ten-repeat dinucleotides
Pure	CACACACACACACACACACA
Compound	CACACACACA**GAGAGAGAA**
Interrupted	CACA**TT**CACACA**TT**CA**TT**CA

The informativeness (number of alleles) of microsatellite loci might be expected to increase with array length (Weber, 1990), but analyses testing this prediction have not provided clear results (Valdes et al., 1993; Rongwen et al., 1995; Guilford et al., 1997), even though microsatellite mutation rates seem to show the expected dependence on repeat unit size (Weber & Wong, 1993).

Pedigree analyses and population studies have indicated that microsatellites are codominant and inherited in a Mendelian fashion (Edwards et al., 1992). They are considered as neutral and are highly polymorphic in natural populations with average expected heterozygosity (H) or gene diversity (Nei, 1973) well above 50% in general, peaking virtually at 100%; though compound and interrupted loci tend to be less polymorphic (Tautz, 1989; Jarne and Lagoda, 1996).

It was initially thought that microsatellite sequences possessed a functional role in the genome, either directly via a role in gene regulation (Hamada et al., 1984) or indirectly as hot spots for recombination (Slighton et al., 1980). While in general no definitive function can be ascribed to microsatellite sequences, in specific instances CAG trinucleotide repeats have been shown to be transcribed; DNA binding proteins specific to di- and trinucleotide repeats have been identified (Richards et al., 1993) and some repeats have been identified as sites of nucleosome assembly *in vitro* (Wang et al., 1994a). However, for the

most part, the ubiquitous occurrence of simple sequences in organismal genomes has so far not been satisfactorily resolved, and the majority may be maintained solely by their ability to replicate and expand in the genome within the limits established by negative selection pressure (Koreth et al., 1996; Hancock, 1999).

2.3.2 Frequency and Distribution

Microsatellites occur as interspersed repetitive elements in all eukaryotic, and to a lesser extent in prokaryotic and eubacterial genomes (Tautz, 1989), at higher frequencies than would be expected purely on the basis of base composition (Hancock, 1999). The human genome is estimated to contain on average 10 fold more microsatellites than plant genomes (Powell et al., 1996a) or at least one simple sequence stretch every 10kb of DNA sequence (Tautz, 1989).
In the last few years, surveys of DNA sequence databases have revealed an abundance of SSR loci in plants, and subsequent studies have demonstrated the informativeness of these markers in several genera (Liu et al., 1995). Database searches indicate that (AT)n, (A)n, (GA)n, (TAT)n and (CA)n repeats are the most frequently occurring SSRs among the plant species examined (Lagercrantz et al., 1993; Morgante and Olivieri, 1993; Wang et al., 1994b), and that tetranucleotide repeats are rarer than trinucleotide, which are in turn rarer than dinucleotide repeats (Hokanson et al.,1998).

The frequencies of the (GA)n and (CA)n repeats based on DNA library screening have been reported for several plant genomes, and there is one (GA)n repeat every 125-250kb and one (CA)n repeat every 250-480kb in *Arabidopsis thaliana* (Bell & Ecker, 1994); *Brassica napus* (Lagercrantz et al., 1993); rice, *Oryza sativa* (Wu & Tanksley, 1993) and seashore paspalum, *Paspalum vaginatum* Swartz (Liu et al., 1995). In the barley (*Hordeum vulgare*) genome it is estimated that a (GA)n will be present every 330kb and one (CA)n repeat every 620kb, which means that there are a total of 1.5×10^4 (GA)n and 7.9×10^3 (CA)n repeats in the genome (Liu et al., 1996). While in wheat (*Triticum aestivum*) these repeats are observed every 440 kb and 704 kb,

respectively (Roder et al., 1995). The most frequent trinucleotide and tetranucleotide repeats found in plant genomes are (AAT)n, (AAC)n, (AGC)n, (AAG)n, (AATT)n and (AAAT)n (Wang et al., 1994b; Gupta et al., 1996). Initial studies utilising fluorescent *in situ* hybridisation (FISH) (Schmidt & Heslop-Harrison, 1996) and Southern hybridisation (Broun & Tanksley, 1996) showed a clustering of microsatellites around the centromere of chromosomes. More recently, in contrast to these earlier reports, genetic and physical mapping have shown that microsatellites are not clustered in specific regions but rather are uniformly distributed in different regions (Panaud et al., 1996; Senior et al., 1996; McCouch et al., 1997; Gianfranceschi et al., 1998; Roder et al., 1998a, 1998b, Cregan et al., 1999). However, although mapping suggests a more or less even (i.e. random) distribution of microsatellites at the gross level, even the highest resolution maps contain some long gaps and low-density regions, many near telomeres (Wu & Tanksley, 1993; Dib et al., 1996; Dietrich et al., 1996).

Little is known in any detail about the factors that might restrict the distribution of microsatellites within non-coding regions. However, considering that sequence affects the flexibility of DNA (El Hassan & Calladine, 1996a, 1996b) and its folding into chromatin (Sivolob & Khrapunov, 1995) it may constrain microsatellites somewhat. Furthermore the locations of replication origins, although ill-defined in eukaryotes, may affect the organisation of microsatellites in the genome, as direction of replication appears to influence the stability of microsatellites (Maurer et al., 1996; Freudenreich et al., 1997).

Microsatellites may be found within expressed regions of the genome, although this is a rare event particularly for microsatellites not based on repeat units of three or more nucleotides, such as $(CA)_n$, as these can give rise to frameshifts if they mutate, a situation seen in some genetic diseases (Weber, 1990; Bruland et al., 1999). Furthermore the detection of a size ceiling on allele size among microsatellites in exons, suggests that these loci are under selective pressure. As a result microsatellites might be excluded from the immediate vicinity of coding regions, as well as from the coding regions themselves (Broun &

Tanksley, 1996). Whether size limitation also applies to loci within introns is unclear; however, they may be more prone than loci located outside genes to selective influence acting on nearby exons through background selection (Charlesworth et al., 1993).

2.3.3 Mutation Mechanism

Microsatellite mutation rates in *in vitro* systems are estimated around 10^{-2} events per locus per replication in *E.coli* (Levinson & Gutman, 1987a) and 10^{-4} to 10^{-5} in yeast (Henderdon & Petes, 1992; Strand et al., 1993), which is high compared to rates of point mutation that are of the order of 10^{-9} - 10^{-10} (Hancock, 1999). Estimates from pedigree analysis in humans suggest a microsatellite mutation rate of around 10^{-3} events per locus per generation (Weber & Wong, 1993). There are two proposed schemes to explain these high rates of mutation. The first involves only a single DNA double helix and slipped strand mispairing (slippage) during DNA replication (Levinson & Gutman, 1987b), the second involves recombination between DNA molecules (Jefferys et al., 1994).

Slippage during replication can take place when the nascent DNA strand dissociates from the template strand. When non-repetitive sequences are being replicated this does not pose a problem because there is only one way in which the nascent strand can reanneal precisely to the template strand before replication is recommenced. If the replicated sequence, however, is repetitive in nature the nascent strand may reanneal out-of-phase with the template strand. When replication is continued after such a misannealing, the eventual nascent strand will be longer or shorter than the template, depending on whether the misannealing gives rise to looped-out bases in the template strand, in which case the product will be shorter; or the nascent strand, in which case it will be longer (Levinson & Gutman, 1987a; Hancock, 1999).

Recombination could potentially alter the lengths of microsatellites in two ways, by unequal crossing-over or by gene conversion. Unequal crossing-over

involves crossing over between chromosome strands (DNA molecules) that are misaligned, giving rise to a deletion in one DNA molecule and insertion in another and can occur both between chromatids in the same chromosome or between chromosomes (Smith, 1976). This occurs most easily for long, tandemly repeated sequences where the recombination machinery cannot easily determine the correct register between the two strands. Gene conversion involves unidirectional transfer of information by recombination, probably as a response to DNA damage, and can transfer sequences in an out-of-phase manner from one allele to another. This has been suggested to generate diversity at minisatellite loci (Jefferys et al., 1994), which are tandemly repeated arrays of basic motifs longer than those found in microsatellites (Hancock, 1999).

Nonetheless, slippage is the commonly accepted mutation model for microsatellites and evidence for the primary role of replication slippage in the generation of length mutation in microsatellites comes from genetic analyses of the process in yeast and *E.coli*. In both systems, length instability of tandem repeats is unaffected by mutants with greatly decreased recombination frequencies (Henderson & Petes, 1992). Furthermore length mutations in microsatellites represent gains or losses of single repeat units, while recombination based mutation would be expected to give rise to a wider range of novel mutants (Hancock, 1999).

Microsatellites exhibit high mutation rates even in species otherwise characterised by low levels of genetic diversity, and consequently they are useful for many applications from plant varietal identification to population studies because a single locus with numerous alleles can be examined (Saghai Maroof et al., 1994). Informative microsatellite variability has been found in insect species with little or no allozyme variability (Gupta et al., 1994), and the utility of SSR loci is also apparent in self-breeding plants. In highly inbred soybean cultivars (*Glycine max.*), Rongwen et al., (1995) reported 11 to 26

alleles per locus and an average heterozygosity of 0.87 at seven SSR loci, substantially exceeding that obtained with allozyme and RFLP markers.

2.3.4 Signature Tagged Microsatellite Site (STMS) Amplification

A number of strategies (both hybridisation based and PCR based) have been designed to exploit microsatellite sequences for the study of plant genomes (Joshi, 1999; Gupta & Varshney, 2000). Signature Tagged Microsatellite Site (STMS) amplification is the most widely used method utilising microsatellite DNA or simple sequence repeats. It involves the amplification of a SSR by designing primers that flank and hence define the microsatellite site, revealing variation in the length of the repeat motifs between individuals, following electrophoresis through an acrylamide or agarose gel (Parker et al., 1998). This method is referred to commonly, though incorrectly, as microsatellites or simple sequence repeats.

Due to their ubiquity, PCR typability, Mendelian co-dominant inheritance, and extreme polymorphism, microsatellites or STMS markers have assumed an increasingly important role as markers in genome analysis (Koreth et al., 1996).

2.3.4.1 STMS Marker Development

The sequences flanking microsatellite loci in a genome are believed to be conserved within a particular species, across species within a genus and rarely even across related genera (Gupta et al., 2002). These flanking sequences can therefore be used for designing primers, and the resultant STMS usually identifies a single locus that, because of the high mutation rate of SSRs, is often multi-allelic (Tautz, 1989; Jones et al., 1997; McCouch et al., 1997). However, the major drawback of microsatellites is that they need to be isolated *de novo* from species being examined for the first time (Zane et al., 2002). The simplest way of finding SSRs is to conduct a search in public sequence databases such as Genbank or the European Molecular Biology Laboratory (EMBL), and this is the least costly in terms of time and resources (Brown et al., 1996). In species

where databases of expressed sequence tag (EST) data have been compiled, identification of microsatellites is also possible. The advantage of using SSRs present in EST sequences is that genes of known function can be mapped (Holton, 2001). However SSRs derived from ESTs are generally less polymorphic (Da Silva, 2001) than those from other approaches and of the 8678 sugarcane sequences scanned by Cordeiro et al., (2001) only approximately 250 (2.9%) revealed microsatellites.

Alternatively, SSR primers designed for closely related species to the particular species in question can be used to obtain polymorphic bands. This is called cross-genetic amplification and the taxonomic distance of the species of interest and the conservation of the flanking sequences, determines whether the microsatellite sequence is amplified and the level of variation observed. Often the reactions need to be optimised and the products sequenced to verify the presence of microsatellite regions (Maguire, 2001).

The percentage of cross genetic amplification is zero for *Paspalum* SSR primers used on *Sorghum*, 18% for *Zea* SSR primers used on *Sorghum* and 22% for *Picea* SSR primers used to amplify regions on the *Pinus* genome (Brown et al., 1996; Peakall, 1997). The screening of public libraries and the use of SSR primers of related species are the least costly methods in terms of time and resources and are, therefore, useful starting points in the search for SSR primers, considering that the SSRs of many species have already been characterized, including maize (*Zea mays)* (Senior & Heun, 1993) and soybean (*Glycine max*) (Akkaya et al., 1992).

A third option involves searching the genome of interest for SSRs. Standard methods for the isolation of plant microsatellite loci (Figure 2.1) involve digestion of genomic DNA, ligation into a plasmid vector, which is then transformed into E.coli (constructing genomic libraries), followed by hybridisation with a labelled microsatellite oligonucleotide probe, DNA

sequencing of positive clones, primer design, locus-specific PCR amplification and identification of polymorphisms (Liu et al., 1996; Holton, 2001).

The relatively low frequency of microsatellites in plant genomes presents some technical problems for the large-scale isolation of microsatellites (Panaud et al., 1995; Maguire, 2001).The development of microsatellites in sorghum (*Sorghum vulgare*) from standard (un-enriched) libraries was shown by Brown et al., (1996) to be inherently inefficient as only 0,2% of the clones hybridised to the SSR probes. Of these clones, 70% of their sequences were unsuitable for primer design, and futhermore following primer synthesis up to 65% of the primer pairs failed to show polymorphisms.

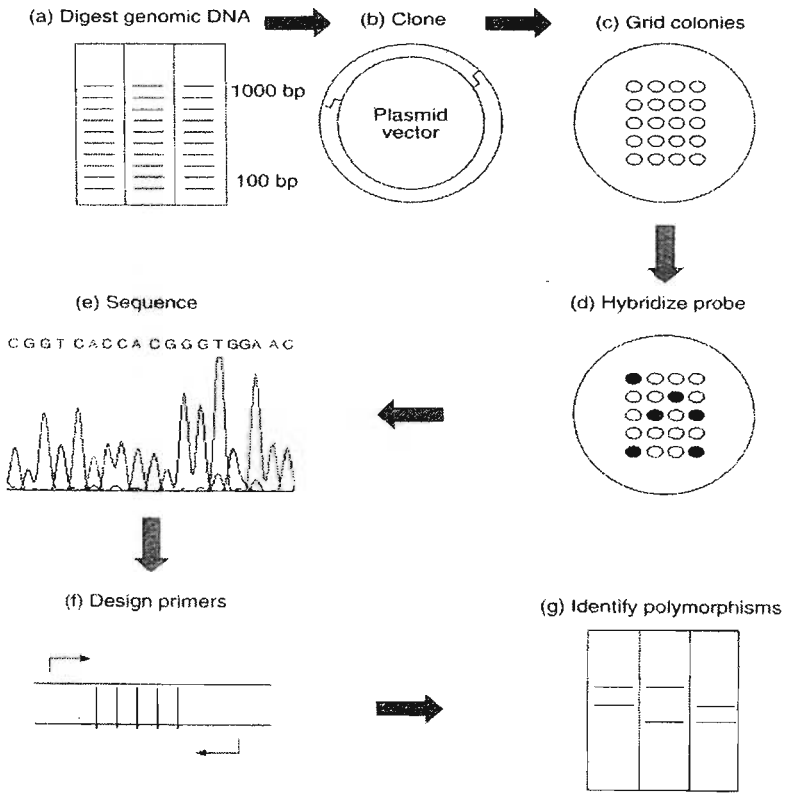

Figure 2.1 Standard method for the isolation of plant microsatellite loci (Maguire, 2001). This involves the (a) digestion of genomic DNA, (b) ligation into a plasmid vector, (c) transformation into *E.coli* and grid colonies, (d) hybridisation with a labelled microsatellite oligonucleotide probe, (e) DNA sequencing of positive clones, (f) primer design and locus-specific PCR amplification, and (g) identification of polymorphisms.

Although much of the early characterization of SSRs has relied on database searches of published sequences or on the construction of genomic libraries, the recent development of new microsatellite enrichment techniques (Edwards *et* al., 1996; Cordeiro *et* al., 1999) has, however, increased the efficiency of microsatellite characterization in species in which little or no previous sequence knowledge is available. The enrichment techniques can be categorised by the mode of enrichment such as: (1) enrichment by colony/plaque hybridisation; (2) enrichment by primer extension; (3) enrichment by hybridisation; and (4) enrichment by screening random amplified polymorphic DNA (RAPD) or Amplified Fragment Length Polymorphisms (AFLP) profiles (Hakki & Akkaya, 2000; Maguire, 2001; Zane et al., 2002).

Microsatellite isolation by enrichment hybridisation

Enrichment by hybridisation (Figure 2.2) is the most popular approach for the isolation of microsatellites. There are several advantages to this method: (1) it is applicable to many plant species, (2) it is quick and relatively inexpensive, and (3) it results in the production of a large number of clones containing many different microsatellite repeats, thus eliminating the need for further library construction with different microsatellite oligonucleotides (Maguire, 2001).

As in standard or traditional methods, the first step in microsatellite isolation by enrichment hybridisation is DNA fragmentation either by sonication or by digestion with restriction enzymes. These fragments are then ligated to adapters or into a vector. Selective hybridisation is performed using multiple probes, containing different tandem repeats, which are bound to a nylon membrane and this results in the simultaneous enrichment of pools of microsatellite repeats. The bound fragments are eluted and amplified by PCR, using primers designed for the adapter sequence. The resulting amplicons are digested, size-selected and ligated into a plasmid vector and transformed into *E.coli.* Finally recombinant colonies are selected and directly sequenced (Maguire, 2001; Zane et al., 2002). Various modifications of this protocol exist

for a large variety of taxa from plants to vertebrates, with an enrichment efficiency generally ranging from 20% to 90% (Zane et al., 2002).

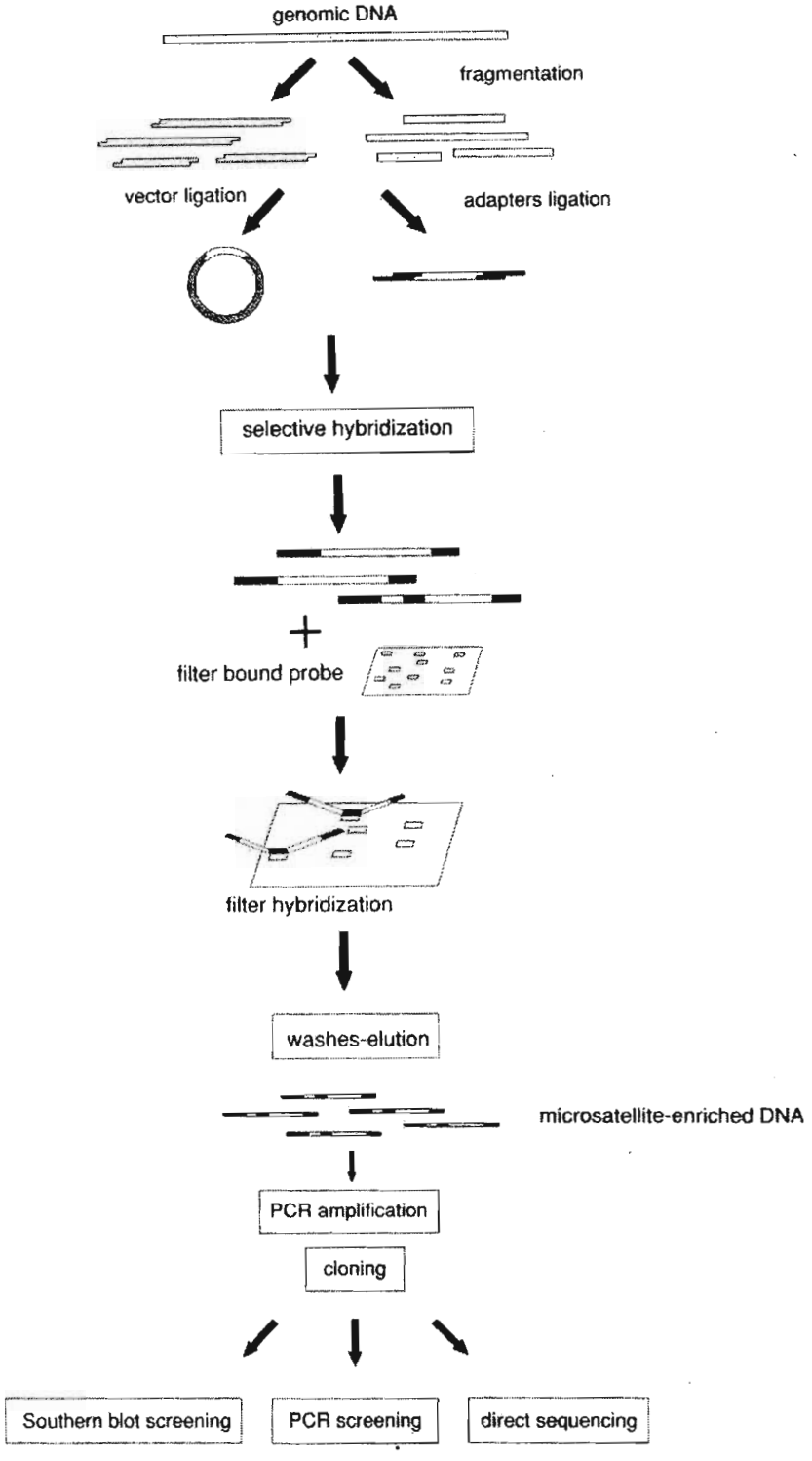

Figure 2.2 Schematic representation of microsatellite isolation by enrichment hybridisation (after Zane et al., 2002)

2.3.4.2 STMS Methodology

In STMS amplification variation in the number of tandem repeats at a microsatellite site, which are primarily due to slippage-based mutations, gives rise to simple sequence length polymorphisms (SSLPs) (Brown et al., 1996). Such variations in tandem repeat number accumulate in populations more rapidly than point mutations, insertions or deletion events, which are events responsible for RFLPs (McCouch et al., 1997).

These differences in length of PCR products or SSLPs are resolved using either agarose, polyacrylamide gel electrophoresis (PAGE), denaturing PAGE or capillary electrophoresis (Jones et al., 1997). The allele size differences are difficult to resolve on agarose gels with ethidium bromide staining (Becker & Heun, 1995; Holton, 2001), but high resolutions can be achieved through the use of polyacrylamide gels in combination with either ethidium bromide staining, silver staining (Scrimshaw, 1992), radiolabelling or fluorescence labelling (Holton, 2001). Although both denaturing and non-denaturing PAGE have been used to resolve small size differences between alleles (Lagoda et al., 1998), single nucleotide resolution of DNA fragments requires the use of denaturing PAGE or capillary electrophoresis (Holton, 2001).

The use of fluorescent primers in combination with a semi-automated DNA sequencer has been shown to be a very promising alternative (Ziegle et al., 1992; Schwengel et al., 1994), and has greatly increased the throughput of microsatellite based systems used to assay variation in humans (Levitt et al., 1994); soybean, *Glycine max* (Diwan & Cregan, 1997); *Brassica* (Mitchell et al., 1997) and tomato, *Lycopersicon esculentum* (Bredemeijer et al., 1998). However, certain groups have raised doubts as to the accuracy of the internal size standard based sizing in automated electrophoresis systems (Schwengel et al., 1994; Delmotte et al., 2001).

Yang et al., (1994) referred to the sequence amplified by each microsatellite primer pair as a particular locus, and any variants thereof (these will be

detected by a difference in length) as an allele of the particular locus under consideration. In polyploids, such as sugarcane, a specific banding pattern is generated per microsatellite primer pair/template combination with the origin of each band being unknown. The bands visualised may be PCR amplicons from the same microsatellite locus, multiple microsatellite loci or perhaps even non-specific PCR products, and as a result these bands are not referred to as alleles (Kaye et al., 1999).

2.3.4.3 Applications of Microsatellite Marker Systems

In plants STMS amplification is particularly attractive as a molecular marker system and its development is accelerating (Donini et al., 1998). In fact, it provides a higher incidence of detectable polymorphisms in relatively unpolymorphic species, such as wheat (*Triticum aestivum*) (Roder et al., 1995) and barley (*Hordeum vulgare*) (Liu et al., 1996), than RFLPs and is more reproducible than RAPDs (Powell et al., 1996b).

There are numerous applications available to utilise the polymorphism detected by microsatellites, but only the most significant applications are explored below.

Fingerprinting and genotyping

In plant species morphological or phenotypic characteristics have long been used to classify or distinguish plant genotypes, however their screening is subjective and often influenced by the environment (McGregor et al., 2000; Russell et al., 1997). Furthermore, examination of morphological characters is labour intensive; for example, over 80 separate morphological markers are examined for a barley (*Hordeum vulgare*) genotype (Cooke, 1984), while at present 52 phenotypic characters have been suggested by the Union for the Protection of New Varieties (UPOV) (http://www.upov.int) for the establishment of Plant Breeders' Rights in sugarcane. Moreover, with an ever-increasing number of cultivars and the finite number of morphological characters, it has

become apparent that such traits will not suffice to establish uniqueness in the future (Rongwen et al., 1995).

DNA markers offer a superior approach for varietal identification revealing genotypic rather then phenotypic polymorphisms, with STMS amplification detecting a large number of alleles accurately and repeatedly. This means that microsatellite data from a number of loci has the potential to provide unique allelic profiles that can be used in fingerprinting and varietal identification (Cordeiro et al., 2000).

Microsatellites, in the STMS approach, have been used in numerous different plant species for varietal identification, such as barley, *Hordeum vulgare*; (Russell et al., 1997); wheat, *Triticum aestivum* (Donini et al., 1998; Gupta et al., 1999); potato, *Solanum tuberosum* (Ashkenazi et al., 2001); rice, *Oryza sativa* (Garland et al., 1999); grapevine, *Vitis vinifera* (Thomas & Scott, 1993); and conifers (Paglia & Morgante, 1998). Bredemeijer et al., (1998) found that four microsatellites were sufficient to differentiate between all 16 cultivars of tomatoes (*Lycopersicon esculentum*) investigated; while Rongwen et al., (1995) used seven microsatellites to discriminate between 94 diverse soybean (*Glycine max)* genotypes and McGregor et al., (2000) was able to use two microsatellites to yield unique profiles for 20 potato (*Solanum tuberosum*) cultivars. In sugarcane Piperidis et al., (2001), have demonstrated that by using only five microsatellite primer pairs in the STMS approach, 40 Australian varieties could be resolved.

It is significant to note, however, that at present UPOV does not accept the use of DNA fingerprints as a standard technique to show a variety's distinctiveness (D), uniformity (U) and stability (S). This DUS-testing forms the basis on which Plant Breeders' Rights (PBR) are granted (Morell et al., 1995). However the fact that DNA fingerprinting is accepted as supplementary character information, with protein fractionation by electrophoresis having been incorporated into tests for barley, *Hordeum vulgare*; wheat, *Triticum aestivum*; and maize, *Zea mays*;

varieties (Law et al., 1998), does allude to the possible inclusion of these fingerprints for such tests in the future. In fact, plant breeders are often keen to strengthen their applications for protection of plant varieties by presenting molecular data, which corroborate their claims of the distinctness of their variety (Lee & Henry, 2001).

Verification of pedigrees

In crop plants, the utilisation of germplasm in the process of developing new breeding lines or cultivars is complicated, requiring generally a number of cycles (one cycle in sugarcane) of crossing and selection. This provides the opportunity for human error and incorrect record keeping, which could potentially result in a recorded pedigree being incorrect (Warburton & Hoisington, 2001). Molecular markers, such as microsatellites, provide a means of verifying pedigrees of valuable germplasm. The term 'parentage analysis' refers to the process whereby the identity of both parents or the seed parent is revealed using the genotype of the progeny, the genotype of the seed parent (if known), and the genotypes of all potential parents at a defined set of gene loci (Gillet, 1999). This approach has been used in humans (Jeffreys & Pena, 1993), chimpanzees (*Pan troglodytes*) (Morin et al., 1994) and even in plant species such as sweetpotato (*Ipomoea batatas)*, which is a polyploid species (Buteler et al., 2002).

Microsatellites can also be used to screen the potential progeny of a cross to ensure that all are legitimate. Jannoo et al., (2001) used one microsatellite to screen 186 sugarcane progeny and successfully detected the presence of 16 illegitimate clones.

Gene tagging and marker-assisted selection

Plant improvement, either by natural selection or through the efforts of breeders, has always relied upon creating, evaluating and selecting the right combination of alleles. However, various obstacles hinder conventional plant

breeding during selection of desirable plants from a segregating population, such as having to screen a large segregating population for a desirable trait e.g. disease resistance and the associated difficulty in screening the population for a desired trait, when the environment influences the trait. In view of these difficulties the concept of indirect marker-aided selection at the seedling stage in early generations is very appealing. The availability of a tightly linked molecular marker for a trait will facilitate plant breeding by saving time and expense, although, in many cases the occurrence of linkage disequilibrium will make gene tagging difficult (Gupta & Varshney, 2000).

A large number of monogenic and polygenic loci for various traits have been identified in a number of plants, which are currently being exploited in marker-assisted selection (McCouch et al., 1997). A number of genes for disease resistance have already been tagged in wheat (*Triticum aestivum*) (Fahima et al., 1998; Korzun et al., 1998; Prasad et al., 1999), and rice (*Oryza sativa*) using microsatellite markers. In soybean (*Glycine max)*, an $(AT)_{15}$ repeat was located within a soybean (*Glycine max)* heat shock protein gene, which is about 0.5cM from (Rsv), a gene conferring resistance to soybean mosaic virus. Furthermore, several other resistance genes including peanut mottle virus (Rpv), phytopthera (Rps3) and Javanese rootknot nematode are clustered in this region of the soybean (*Glycine max)* genome (Joshi et al., 1999).

2.3.4.4 STMS Amplification Limitations and Considerations

In contrast to such methods as RFLPs that do not require previous sequence knowledge, the development of SSRs requires an initial high cost and labour intensive development, which is the most significant drawback of microsatellites (Cordiero, 2001). Although cross-species amplification has been experimented with, its success has been very limited and as such SSR development will remain costly and time consuming (Brown et al., 1996).

A further technical complication for SSRs is that high-resolution electrophoresis is required, particularly for dinucleotide repeat SSRs, where amplification often

leads to the production of numerous stutter bands beside the microsatellite band, differing by 1 or 2 bp (Koreth et al., 1996). The major mechanism postulated for this is 'slipped strand mispairing' (akin to the slippage mutation mechanism for microsatellite size expansion and sequence evolution), and involves the SSR sequences permitting slippage of the copied strand on the template producing fragments with two-nucleotide spacing. Other postulated mechanisms are: failure of the polymerase to read through the repeats and the 3' terminal addition of nucleotides by the polymerase (Hauge, 1993; Shibata et al., 1994).

The presence of these extraneous bands can lead to difficulties in scoring gels or to confusion when determining the true allele size from either gel electrophoresis, or an electropherogram derived from an automated DNA sequencer (Figure 2.3). The use of SSRs containing trinucleotide or higher order repeats eliminates this problem, however these markers generally show less polymorphism (Holton, 2001).

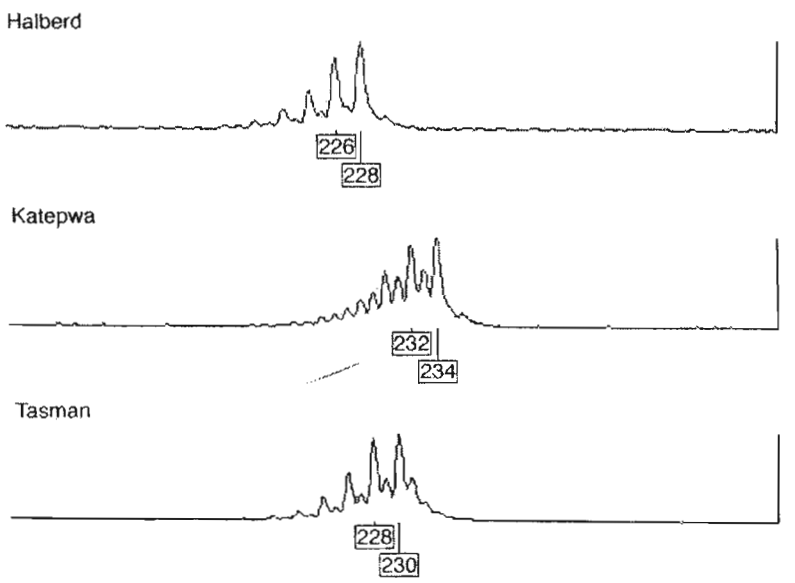

Figure 2.3 Artefacts due to 'stutter' and non-template A-addition to PCR amplification products (Holton, 2001). Wheat varieties 'Halberd', 'Katepwa' and 'Tasman' show stutter peaks, which differ by 2 bp increments (eg. 226 and 228). Additional peaks between the stutter peaks are due to non-template addition of a single A nucleotide to the 3' end of the PCR products by *Taq* polymerase.

Variation in flanking regions is also an important consideration in STMS as point mutation in primer annealing sites may cause microsatellite loci to not amplify, giving rise to a null allele (Gupta & Varshney, 2000). Null alleles have been reported in many plant species, with frequencies of null alleles of up to 15% having been reported, and they are found in up to 25% of loci (Callen et al., 1993; Paetkau & Strobeck, 1995; Prasad et al., 2000); though data sets are still too restricted to assess the generality of these trends. Null alleles may be detected in population studies by careful testing against frequencies that are expected under the Hardy-Weinberg equilibrium, provided that heterozygote deficiencies have no other origin, such as the mating system (Powell et al., 1996b).

The presence of null alleles will lead to an underestimation of heterozygosity within a population because in the heterozygote, the presence of a null allele on one homologue cannot be recorded. In some cases, new primers can be designed or the PCR modified to allow for amplification of the null allele (Gupta & Varshney, 2000). This problem seems to be more prevalent in highly outbred heterozygous species (Powell et al., 1996b; Gianfranceschi et al., 1998).

2.3.4.5 Development and Use of STMS in Sugarcane

Microsatellite development is costly and time consuming and as a result an International Sugarcane Microsatellite Consortium (ISMC) was established by Southern Cross University (SCU), New South Wales, Australia in 1998 and was conducted under the auspices of the International Consortium for Sugarcane Biotechnology (ICSB). The International Sugarcane Microsatellite Consortium was made up of 13 members from 8 different countries and its objective was to develop and characterise sugarcane microsatellites. In this regard, two members, the Centre for Plant Conservation Genetics (CPCG) in Australia and the Centre de Coopération Internationale en Recherche Agronomique pour le Developpement (CIRAD) in France developed two different microsatellite-enriched libraries; and various members, including the South African Sugar Association Experiment Station (SASEX), were involved in sequencing the

resulting clones. From the CPCG library, it was determined that the most common (76.8%) repeat motifs were either dinucleotide or trinucleotide repeats, while the remaining 15.8% of the repeat motifs comprised eight tetranucleotides, two pentanucelotides and 57 mononucleotides. The most common microsatellite motif was the (TG)n/(CA)n group, which represents over 29.5% of all microsatellites, and these ranged between 6 and 57 repeats in length with an average of 20. The (TAC)n/(GTA)n group represented the most common trinucleotide at 5.5% of the detected microsatellites and the number of repeats ranged between 8 and 92 with an average of 39 repeats. This work was finished in 2000 and only the dinucleotides and trinucleotides were used to yield more than 200 primer pairs suitable for sequence tagged microsatellite site (STMS) amplification (Cordeiro, 2000; Jannoo et al., 2001; Piperidis et al., 2001).

Testing of these primer pairs on a small sample population of five sugarcane genotypes revealed each microsatellite had between three and 12 alleles with an average of eight. Markers showing polymorphisms had a polymorphism information content (PIC) (Weir, 1996) value of between 0.48 to 0.8 with a mean value of 0.72. (Cordeiro et al., 2000). These PIC values, a determination of the value of a marker in detecting polymorphism, indicate that sugarcane microsatellite markers are, for the most part, highly informative.

Members of the International Sugarcane Microsatellite Consortium (ISMC) have already begun to utilise the SSR primer pairs. For instance, Jannoo et al., (2001) used a single consortium derived microsatellite primer to perform a legitimacy test on 186 progeny derived from a cross, and determined that 16 were illegitimate clones. In the same study, Jannoo et al., (2001) utilised five microsatellites to ensure that a set of cultivars were correctly identified. The high polyploidy of sugarcane might be an advantage in the use of STMS for varietal identification, allowing unique fingerprints to be generated for a sample population amplified using only a single microsatellite primer pair (Cordeiro, 2001). However, the amplification of a large number of over-lapping

indistinguishable bands can also occur due to the presence of multiple priming sites, as a consequence of high ploidy.

CHAPTER 3

MICROSATELLITE EVALUATION AND OPTIMISATION OF PCR AMPLIFICATION AND GEL ELECTROPHORESIS

3.1 INTRODUCTION

The sequence tagged microsatellite site (STMS) approach utilises the presence of simple sequence repeats/microsatellites in the genome, and involves amplifying genomic DNA with two unique oligonucleotide primers that flank, and hence define, the microsatellite site revealing variation in the length of the repeat motifs between individuals, following electrophoresis through an acrylamide or agarose gel to resolve the differences in allele size (Parker *et al.*, 1998). However, optimisation of the PCR and gel electrophoresis conditions of the protocol is crucial in ensuring adequate sensitivity and overall reproducibility while at the same time preventing the occurrence of non-specific or faint amplicons.

Previous work by Cordeiro et al., (2000) using STMS, evaluated the usefulness of 124 sugarcane microsatellite primer pairs, derived from the International Sugarcane Microsatellite Consortium (ISMC), and demonstrated both their ability to provide unique allelic profiles and their reproducibility. Consequently, the initial PCR protocol used in this study was taken from Cordeiro et al., 2000 and primer pairs were chosen from those provided by the ISMC.

The objectives of this study were two-fold, first of which was the evaluation of the application of a set of sugarcane microsatellite primer pairs to South African sugarcane germplasm. Second, was the optimisation of the various parameters of the PCR reaction, for example $MgCl_2$ concentration, and the assessment of the suitability of various gel electrophoresis systems.

3.2 MATERIALS AND METHODS

3.2.1 Sugarcane Microsatellite Primers

Twenty-six primer pairs (Table 3.1) were chosen from the 260 supplied by the International Sugarcane Microsatellite Consortium (ISMC). The ISMC supplied heterozygosity data for all the primers; and from this, twenty-one primer pairs were chosen with 100% heterozygosity values, while the remainder were suggested (George Piperidis, 2001, pers. comm.). The primers were resuspended in water at 50µM and then pooled into 20 µl sub-aliquots. All primers were stored at –20°C.

3.2.2 Sugarcane Cultivars

Five cultivars (NCo376, N24, N21, N17, and N14) were used for the microsatellite evaluation and protocol optimisation. DNA was isolated following a modification of the procedure of Dellaporta et al. (1983). The DNA stocks were quantified using a fluorometer (Hoefer DyNA Quant 200) and 50 ng/µl dilutions were made and stored at –20°C.

3.2.3 Isolation of Sugarcane DNA

Modified Dellaporta method

DNA was routinely extracted from sugarcane following a modification of the Dellaporta method (Dellaporta et al., 1983). Five stalks of sugarcane were collected from different individuals of a single variety. The soft inner leafroll tissue was sliced into discs approximately 2mm thick. Six grams of pooled tissue were weighed out, frozen with liquid nitrogen and ground to a fine powder in a mortar and pestle. The tissue was transferred to 35ml extraction buffer (100mM Tris-Cl, 500mM NaCl, 50mM EDTA and 1% (v/v) β-mercaptoethanol, pH 8.0) and 3.5ml of 20% (w/v) sodium dodecyl sulphate (SDS) added. Samples were mixed, incubated at 70°C for 1 hour, 7ml of 5M potassium acetate added and tubes placed on ice for 30 min. The samples were

centrifuged for 15min at 10300Xg in a Beckman Avanti™ J-251 centrifuge and the supernatant filtered through four layers of gauze cloth into a new tube. Isopropanol (¾ volume) was added to each tube, mixed gently, and incubated at 4°C until the DNA formed a precipitate. DNA was spooled out using sterile glass hooks and dissolved in 3ml TE buffer (10mM Tris-Cl, 1mM EDTA, pH 8.0) at 37°C overnight. An equal volume of chloroform:isoamyl alcohol (24:1) was added, the tubes mixed, and centrifuged at 3840Xg for 10min. The DNA-containing aqueous phase was drawn off, aliquoted into 1.5 ml tubes, and stored at –20°C.

3.2.4 DNA Quality and Quantity Assessment

The extracted DNA was quantified by fluorometry. A HOEFER® DyNA Quant 200 fluorometer was used to determine the concentration of 2µl of DNA in 2ml of assay solution. A low range assay solution, 0.1µg/ml Hoeschst 33258 dye in 1X TNE (0.2M NaCl, 10mM Tris-Cl, 1mM EDTA, pH 7.4), was prepared fresh from concentrated stocks for DNA estimated to be 10-500ng/ml. The high range assay solution, 1µg/ml Hoeschst 33258 dye in 1X TNE, was required for higher concentrations of DNA (100-5000ng/ml). The fluorometer was calibrated using an undiluted 1mg/ml DNA standard (Calf thymus DNA, ultrapure, Sigma D4764) for the high range assay or a 1:10 dilution of the DNA standard for the low range assay. The instrument was zeroed with 2ml of assay solution before and after calibration, and between DNA samples. An average of three fluorometer readings per sample was recorded as the DNA concentration.

The quality of the extracted DNA was assessed by measuring the absorbance of a 1:10 dilution of the sample at 260nm (A_{260}) and 280nm (A_{280}) on a Beckman DU 7500 spectrophotometer. The purity is indicated by $(A_{260})/(A_{280})$, and this ratio is 1.8 for pure DNA free from protein.

Table 3.1 The twenty-six sugarcane microsatellite primers evaluated and optimised for use on South African sugarcane varieties.

Name	Repeat	Forward Primer Sequence	Reverse Primer Sequence	Used T_A	Size Expected
SMC334BS*	$(TG)_{36}$	CAA TTC TGA CCG TGC AAA GAT	CGA TGA GCT TGA TTG CGA ATG	50	213
SMC336BS*	$(TG)_{23}(AG)_{19}$	ATT CTA GTG CCA ATC CAT CTC A	CAT GCC AAC TTC CAA ACA GAC	50	233
SMC355BS	$(CCG)_7$	ATC CGA TGA CGT TCA CCC AC	CGT TGT GCT CTA CCC GAT GAA	55	188
SMC378BS	$(TG)_{57}$	TGT GCC AAA TTA TCT GTG GAC	GCC ATT GCT ATT TTT CCT TCA	50	210
SMC703BS	$(CA)_{12}$	GCC TTT CTC CAA ACC AAT TAG T	GTT GTT TAT GGA ATG GTG AGG A	50	215
SMC805BS	$(TGC)_7$	TGA GAA TGC TGT CAT AGG GCT TG	TCG AGG TGA ATC ATT GCC TTC A	55	264
SMC843BS	$(TG)_{16}CAA(GT)_9$	GGT CCC ATT GAT GTG GCA	CTA GGA CCT TGT GGT TAC CGT	50	201
SMC17CG	$(GA)_{22}$	AAG GTA GCA CGA AAC ACG TCG AT	AAC CCC AGC GGC GAT CTC T	55	194
SMC226CG	$(CA)_{10}$	GAG GCT CAG AAG CTG GCA T	ACC CTC TAT TTC CGA GTT GGT	50	136
SMC371CG	$(CA)_{13}$	GGA TAT GGT TTT CAT TGC CAC TTG	CAT TTT AAG CGT ATG GGG TAA CAA	55	205
SMC1039CG	$(TG)_{17}$	AGG TGA GAG TTC CTG GCT TTC CA	TGT GCT GGC AAG ATA CCC CTA CTT	55	189
SMC278CS	$(TG)_{19}(AG)_{25}$	TTC TAG TGC CAA TCC ATC TCA GA	CAT GCC AAC TTC CAA ACA GAC T	50	236
SMC280CS	$(CA)_{21}$	TGA TCG CAC GTT GTA TCC AAC A	TTT GAC CAC GCC ACG GTA GAT	55	238
SMC286CS*	$(TG)_{43}$	TCA AAT GGG ACC TTA TTG GAG	TCC CTC GAT CTC CGT TGT T	50	202
SMC569CS*	$(TG)_{37}$	GCG ATG GTT CCT ATG CAA CTT	TTC GTG GCT GAG ATT CAC ACT A	50	273
SMC640CS	$(CT)_5/(CA)_8$	TTA AGA GAC CCG CCT TTG GAA	TGC CAG AAG TGG TTG TGC TCA	55	230
SMC662CS	$(AC)_{31}$	GAC TGC ATG GCT TGC TGA TCG	GGA CCT TGG CGG TGA TGG G	55	241
SMC687CS	$(CAG)_8$	AGC CAT GCA GGC AGG CAT	CGC ACA ATC TGC AAG TGC ATC A	55	203
SMC213MS	$(CA)_{12}$	GCA GGG AGA CGA ACA CGA GT	TGC AAC CTT CTT CAG GCT TGA	55	148
SMC219MS	$(CA)_{40}$	TCT CCC TCG ATC TCC GTT GT	GGA GTG TCT TCA GCT ATC GGA	50	185

Name	Repeat	Forward Primer Sequence	Reverse Primer Sequence	Used T_A	Size Expected
SMC221MS	$(TG)_{25}$	CAT GCC AAC TTC CAA ACA GAC T	GGT GAT GCG AAG AGA TTG GA	50	204
SMC238MS	$(TG)_{19}C(GA)_{25}$	TTG GAT TGG ATT CTA GTG CCA A	AGG AAA TGG ATT GCT CAG GTG T	55	206
SMC519MS	$(TGC)_{18}$	CGA TGG ACG CCA ATG CAA	GTG CCG CCG CAC CTC ATA	55	208
SMC17AUQ	$(TG)_{16}$	CGT AGG CGA GAG GCT TAT CAA A	TGT CGG TCA CCC TCC AAG GA	55	224
SMC36BUQ*	$(TTG)_7$	GGG TTT CAT CTC TAG CCT ACC	TCA GTA GCA GAG TCA GAC GCT T	50	118
SMC31CUQ	$(TC)_{18}(AC)_{22}$	CAT GCC AAC TTC CAA TAC AGA CT	AGT GCC AAT CCA TCT CAG AGAN	50	225

* Primers suggested by Dr. Piperidis, BSES, Australia.

3.2.5 Initial PCR Protocol

PCR reactions were performed in 25 µl volume made in sterile water containing 10mM Tris-HCl pH 9.0 and 50mM KCl, 2.5mM MgCl, 1.0U Taq polymerase (Promega), 0.2 mM of each dNTP, 0.2 µM of each forward and reverse primer, 25 ng of template DNA. Reactions were run on a Perkin Elmer 9700 thermocycler. Cycling conditions were: 94°C for 3 min, followed by 30 cycles of 94°C for 30 sec, appropriate annealing temperature (either 50°C or 55°C) for 30 sec, 73°C for 30 sec and a final extension step at 73°C for 3 min (after Cordeiro et al., 2000).

3.2.6 PCR Optimisation

Initially the PCR reactions were carried as described in Section 3.2.5, but the protocol was amended after each subsequent reaction parameter optimisation. For all the PCR optimisation experiments the PCR products were resolved on standard agarose gels (Section 3.2.7.1), except for the PCR products of the buffer concentration experiment that were resolved on a denaturing polyacrylamide gel (Section 3.2.7.3).

3.2.7 Gel Electrophoresis

3.2.7.1 Standard Agarose Gels

Amplification products were resolved using 3% agarose gels (Seakem LE) run at 4.5V/cm in 0.5X TBE buffer. Staining was as described in Section 3.2.8.

3.2.7.2 High Quality Agarose Gels

High quality MS-8 agarose (Whitehead Scientific) gels were pre-chilled to 4°C and were run on ice at 4.5V/cm in 1XTAE buffer. Staining was as described in Section 3.2.8.

3.2.7.3 Denaturing Polyacrylamide Gels

Eight percent denaturing polyacrylamide (27:1) gels were prepared following the protocol of Sambrook & Russell (2001), except that gels were run in a Hoefer SE-600 dual cooled gel electrophoresis unit. The gels were run at 200V and the running time was varied between 2hrs and 2hr30min, depending on the size of the amplification products, and the temperature was maintained at 20°C.

Equal volumes of formamide loading buffer (80% formamide, 10mM EDTA and 0.1% bromophenol blue and xylene cyanol) was added to each PCR sample. Prior to electrophoresis, the samples were heated to 95°C for 5 minutes and then snap-cooled on ice (Sambrook & Russell, 2001)

3.2.8 Staining and Visualisation

3.2.8.1 Agarose Gel Staining

After electrophoresis, gels were stained in an ethidium bromide solution (0.5µg/ml) for 15 minutes, followed by destaining for 30 minutes in 0.5X TBE or 1X TAE (appropriate to gel). DNA fragments were visualised on a Hoefer Mighty Bright UV transillumionator at 300 nm and photographed using a Vilber Lourmat photographic system. Pictures were printed on Sony high density, light sensitive paper using a Sony video graphic printer.

3.2.8.2 Denaturing Polyacrylamide Gel Staining

SYBR Gold stain (Molecular Probes) was prepared as per manufacturers instructions and stored at 4°C. Following electrophoresis the polyacrylamide gels were stained for 30 minutes and a digital image was captured using a Vilber Lourmat photographic system.

3.3 RESULTS AND DISCUSSION

3.3.1 Genomic DNA Yield and Purity

Sugarcane genomic DNA was extracted from selected sugarcane varieties using the modified Dellaporta method (Section 3.2.3), and the average yield was 285.0±44.7ng/ul. The purity values were determined by calculating the A_{260}/A_{280} ratio. The average purity of the five cultivars was 1.67±0.03, which is relatively pure considering that pure DNA has an A_{260}/A_{280} ratio of 1.8. The purity value being below 1.8 indicates the presence of residual protein contamination in the samples (Sambrook & Russell, 2001).

3.3.2 Microsatellite Evaluation

The twenty-six microsatellite primer pairs (Table 3.1) were tested on the five cultivars (NCo376, N24, N21, N17, and N14). PCR was carried out using a modified method of Cordeiro et al., (2000), as described in Section 3.25, and the amplicons were resolved on Seakem LE agarose gels (Section 3.2.7.1). Five of the microsatellites (SMC355BS, SMC378BS, SMC371CG, SMC238MS and SMC519MS) failed to amplify any bands and were excluded from further investigation (Results not included).

3.3.3 PCR Optimisation

The initial PCR protocol described in Section 3.2.5, a modified PCR protocol of Cordeiro et al., (2000), revealed amplification patterns that showed the presence of both faint PCR products and smearing (Figure 3.1). The robustness of microsatellites is dependent on PCR, and as such it is imperative for microsatellite-based fingerprinting or genotyping that the PCR conditions and program utilised amplify specific products reproducibly. In an attempt to improve upon the stringency of the PCR amplification, various reaction conditions were tested and varied.

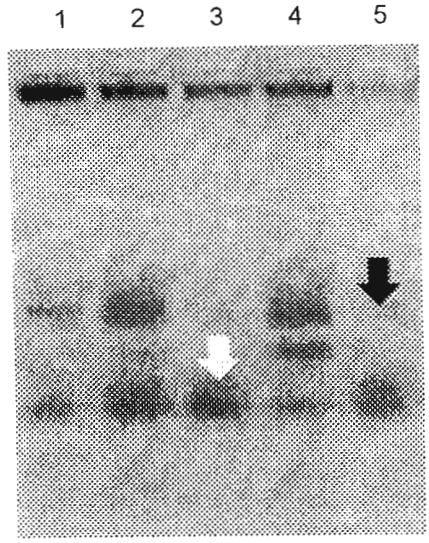

Figure 3.1 Agarose gel separation of the amplification products of microsatellite primer pair, SMC17CG. The cultivars used were as follows: lane 1, Nco376; lane 2, N24; lane 3, N21; lane 4, N17; lane 5, N14. The white arrow indicates the presence of smearing, while the black arrow indicates a faint PCR product. PCR was carried out as described in Cordeiro et al., (2000), except that 30 cycles and 1.0U Taq polymerase were used.

3.3.3.1 Magnesium Concentration

The effect of MgCl concentration on PCR is two-fold, firstly it affects the overall stringency of the reaction and secondly it affects the polymerase itself (activity, processivity). Generally, excess Mg^{2+} will result in accumulation of non-specific amplification products and insufficient Mg^{2+} will reduce yield (Saiki, 1989). An Mg^{2+} titration experiment, from 1mM to 3mM with 0.5mM increments, was carried out using six microsatellites (SMC703BS, SMC805BS, SMC17CG, SMC226CG, SMC687CS and SMC219MS) and three cultivars (Nco376, N14 and N21), in order to determine the optimal concentration to be used for all the microsatellites. The PCR reactions were carried out as described in Section 3.2.5. Figure 3.2 shows the amplification results for microsatellite primer pair SMC805BS.

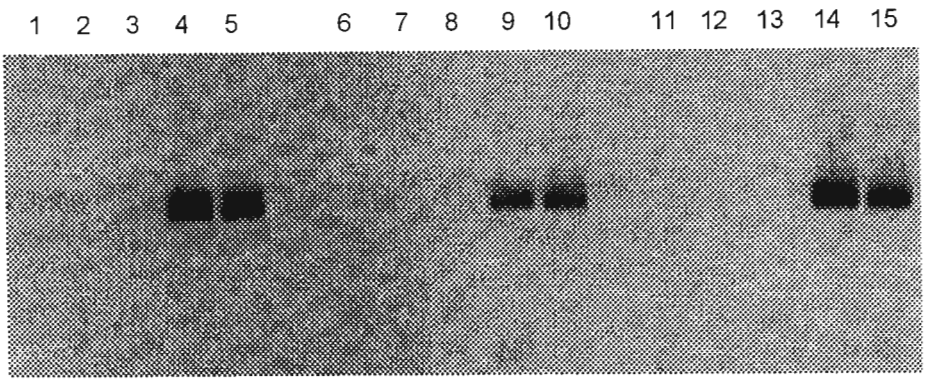

Figure 3.2 Effect of increasing magnesium concentration on the amplification of microsatellite, SMC805BS. Microsatellite primer pair SMC805BS, was used to amplify genomic DNA from N14 (lane 1-5), N21 (Lane 6-10) and NCo376 (lane 11-15), and reactions were titrated from 1mM Mg^{2+} to 3mM Mg^{2+} in increments of 0.5 mM. The amplification products were separated on an agarose gel.

For the majority of the microsatellites tested no PCR products were observed at $MgCl_2$ concentrations below 2.5mM, with smearing being visible in certain microsatellites at 3mM $MgCl_2$. It was concluded that a 2.5mM final concentration of $MgCl_2$ was the most suitable for sugarcane microsatellite amplification, and its use was made standard in subsequent optimisation experiments.

3.3.3.2 Primer Concentration and Template Quantity

Within limits, increasing primer concentration may improve the outcome of the PCR reaction; however if the primer or template concentration is too high, the reaction may be inhibited due to a reduction in free Mg^{2+}. PCR reactions were carried out using varying amounts of both template and primer as a means to optimise both parameters in tandem according to Table 3.2.

Table 3.2 Array used to optimise primer concentration (μM) and template quantity (ng).

Primer Concentration (μM)	Template (ng)		
0.1	25	50	75
0.2	25	50	75
0.3	25	50	75
0.4	25	50	75

The amount of primer and template were titrated in tandem using cultivar N14 and microsatellite primer pair, SMC226CG (Figure 3.3).

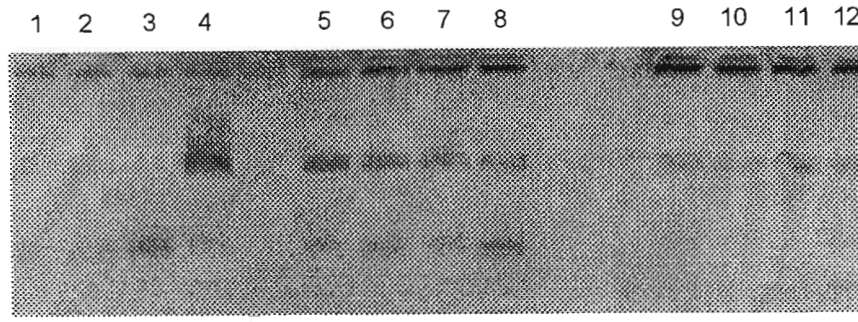

Figure 3.3 Effect of differing amounts of template and primers on the amplification of microsatellite primer pair, SMC226CG. Microsatellite primer pair, SMC226CG was used to amplify genomic DNA from N14 with differing amounts of template and primers: 25 ng (lane 1-4), 50 ng (lane 5-8) and 75 ng (lane 9-12) template with primer concentrations from 0.1μM – 0.4μM with increments of 0.1μM. The amplification products were separated on an agarose gel.

On the agarose gel (Figure 3.3) two PCR products were resolved for SMC226CG. At 50 ng of template, two bands were amplified for each primer concentration. However 25ng of template was insufficient to amplify the larger band, except when using 0.4μM primers; while at 75ng of template, the larger band was preferentially amplified. Furthermore at 50 ng of template, there seems to be a change in amplification product intensity as the primer concentration is increased from 0.1μM (favours larger band) to 0.4μM (favours smaller band) with balanced amplification occurring at 0.2 μM. It was found that

the most ideal primer concentration and template quantity was 0.2 µM and 50 ng respectively and these values were used for further PCR optimisation experiments.

3.3.3.3 PCR Cycles and *Taq* Polymerase

Increasing the *Taq* polymerase concentration can improve the yield of the desired products; however, if increased beyond its optimum concentration this can result in greater production of non-specific PCR products and reduced yield of the desired target fragments (Saiki, 1989). Twenty-five to thirty-five cycles usually amplifies sufficient PCR products for visualisation, with little being gained by increasing the cycle number up to 60 (Henegariu et al., 1997). It is significant to note that if non-specific PCR products are amplified in the initial cycles of the PCR, and a high number of PCR cycles are used, it could result in the spurious products being of the same intensity as the specific products - thereby severely affecting microsatellite scoring.

PCR reactions were replicated using different numbers of PCR cycles (30, 35 and 40) and two different (1U and 1.5U) *Taq* polymerase quantities, according to Table 3.3

Table 3.3 Array used to optimise PCR cycle number and Taq polymerase concentrations.

Taq polymerase	PCR Cycle Number		
1U	30	35	40
1.5U	30	35	40

Figure 3.4 shows the results of cultivar N14 being amplified by microsatellite primer pair SMC17CG using the above conditions.

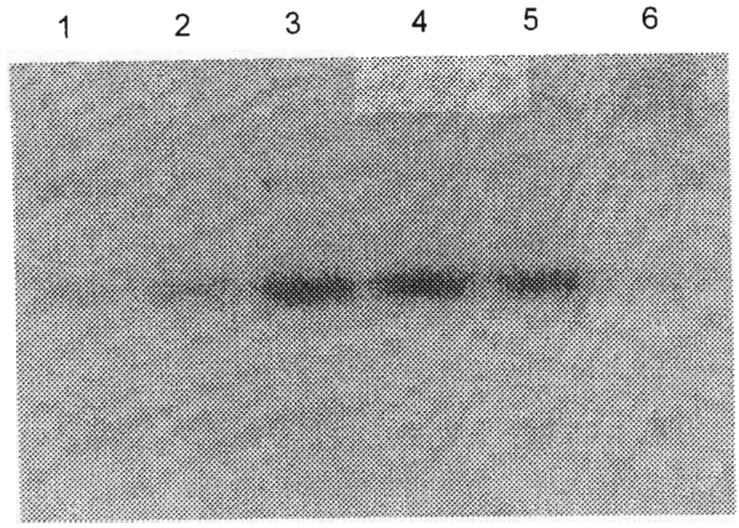

Figure 3.4 Effect of differing amounts of Taq polymerase and number of PCR cycles on the amplification of microsatellite primer pair, SMC17CG. Microsatellite primer pair SMC17CG was used to amplify genomic DNA from N14: lane 1-2 (30 cycles), lane 3-4 (35 cycles) and lane 5-6 (40 cycles) with 1U Taq polymerase in lane 1, 3 and 5 and 1.5U Taq polymerase in lane 2,4, and 6.

It was determined that 1.5U of *Taq* polymerase and 35 cycles yielded the best results, and the PCR protocol was amended to use these new values.

3.3.3.4 Hot Start PCR Using AmpliTaq Gold

Hot start PCR entails the withholding of at least one reagent (typically Taq polymerase) from the reaction mixture until the reaction tube has reached a suitably high temperature. The benefit of this is that primer annealing to non-target sequences, at low temperatures, is prevented and this reduces the occurrence of PCR artefacts and improves the yield of low copy number amplifications (Chou et al., 1992).

Amplification reactions of microsatellite primer pair SMC280CS and cultivars N14, N17 and N24, respectively, were replicated using two different polymerases, namely *Taq* polymerase (Promega) and AmpliTaq Gold (Applied Biosystems). The optimised PCR protocol, determined thus far, was used and the amplicons were separated on an agarose gel (Figure 3.5).

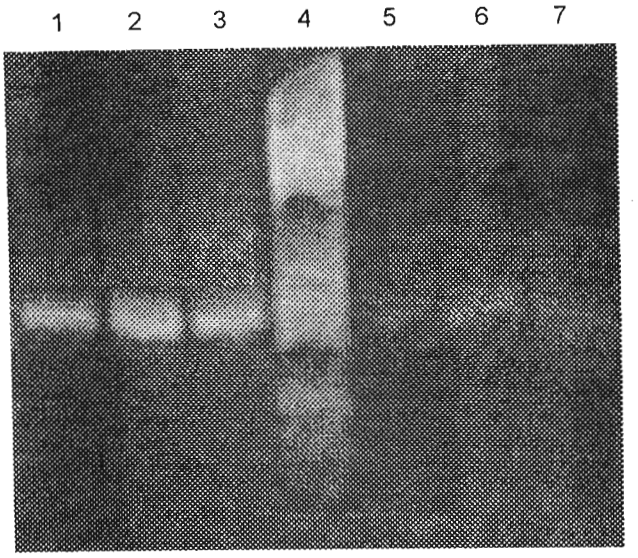

Figure 3.5 Effect of different types of DNA polymerase on the amplification of microsatellite primer pair, SMC280CS. AmpliTaq Gold (lane 1-3) and Taq polymerase (lane 5-7) amplifying N14 (lane 1 and 5), N17 (lane 2 and 6) and N24 (lane 3 and 7). Lane 4 is molecular weight marker 5. Amplification products were separated on an agarose gel.

Although AmpliTaq Gold did result in greater yields, its price was prohibitive and it was decided that it would not be used in subsequent PCR reactions. However the use of cheaper PCR additives and adjuncts was explored in Section 3.3.3.5 as an alternative.

3.3.3.5 The Use of PCR Additives / Adjuncts

The use of PCR adjuncts/additives has been suggested to improve amplification efficiency (improved yield) and specificity (no unspecific products). However it has been found that their use can improve, reduce or have no effect on the efficiency or specificity of a PCR reaction, and as a result their usefulness needs to be tested in each case. Microsatellite primer pair SMC805BS and cultivar N14 were used to assess the usefulness of various adjuncts according to Table 3.4, and Figure 3.6 shows the results.

Table 3.4 PCR additives/adjuncts tested for their effect on microsatellite amplification and their final concentrations.

Name of Adjunct	Quantity or final concentration
Formamide	5% (v/v)
Bovine Serum albumin (BSA)	100 µg/ml
Dimethylsulfoxide (DMSO)	5% (v/v)
Glycerol	15% (v/v)
PCR supermix (Gibco)	1X
Q-solution (Qiagen)	1X
Betaine (Sigma)	1X

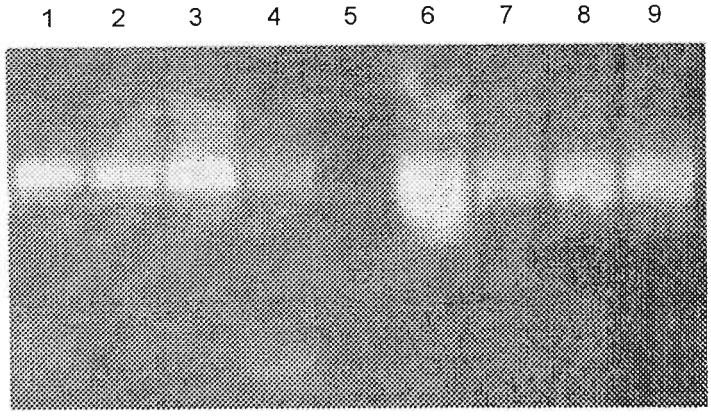

Figure 3.6 Effect of various PCR additives on the amplification of N14 by microsatellite primer pair, SMC805BS. The reactions were run on an agarose gel and were as follows: 5% (v/v) formamide (Lane 1); 0.1mg/ml BSA (Lane 2); 5% (v/v) DMSO (Lane 3); 15% (v/v) glycerol (Lane 4); 1X PCR Supermix (Lane 5); 1X Q-solution (Lane 6); no additives (lane 7); AmpliTaq Gold (Lane 8) and 1X Betaine (Lane 9).

From Figure 3.6, it was decided that 5% (v/v) formamide was the most ideal additive. However, in subsequent reactions the amplification failed. It was suspected that the presence of formamide was denaturing the Taq polymerase.

Microsatellite primer pair SMC17AUQ and cultivars N14 and N17, respectively, were used to ascertain firstly, whether the order in which the PCR cocktail was prepared - with respect to the addition of the formamide - had an effect on

amplification (Figure 3.7) and secondly, whether different formamide percentages were more suitable for PCR (Figure 3.8).

PCR reactions of microsatellite primer pair, SMC17AUQ and cultivar N14 were replicated, except that the addition of the Taq polymerase and the formamide was varied such that: (1) Taq polymerase was added after the 5% (v/v) formamide was added to the PCR reaction mix in one reaction; (2) 5% (v/v) formamide was added to the Taq polymerase prior to incorporation into the reaction mix in another reaction, and (3) in the final reaction no 5% (v/v) formamide was added (Figure 3.7).

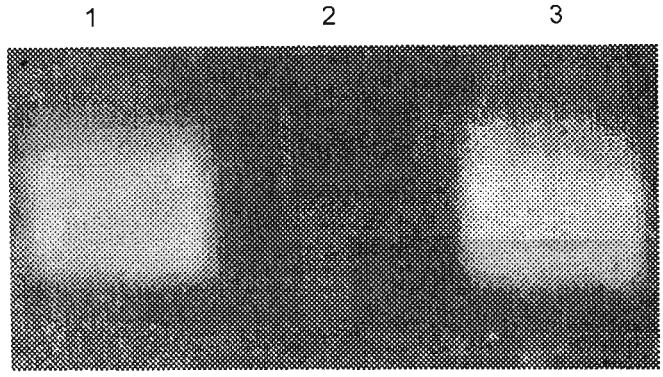

Figure 3.7 Effect of formamide addition on the amplification of N14 by microsatellite primer pair, SMC17AUQ. The PCR reactions were run on an agarose gel as follows: Taq polymerase added last to PCR cocktail containing 5% (v/v) Formamide (Lane 1); 5% (v/v) Formamide added to Taq polymerase prior to incorporation into PCR cocktail (Lane 2) and formamide not added (Lane 3).

It was found that if the formamide was added to the PCR cocktail prior to the Taq polymerase, the PCR was successful. However, if the Taq polymerase and the formamide were mixed together prior to their addition to the PCR cocktail, the amplification failed (Figure 3.7).

PCR reactions with microsatellite primer pair, SMC17AUQ and cultivar N17 were replicated using differing final concentrations of formamide: 2% (v/v), 4% (v/v), 5% (v/v), 6% (v/v), 8% (v/v) and 10% (v/v) (Figure 3.8). It was found that

lower levels of formamide were more suitable, while 5% (v/v) formamide incorporation was shown to be inconsistent.

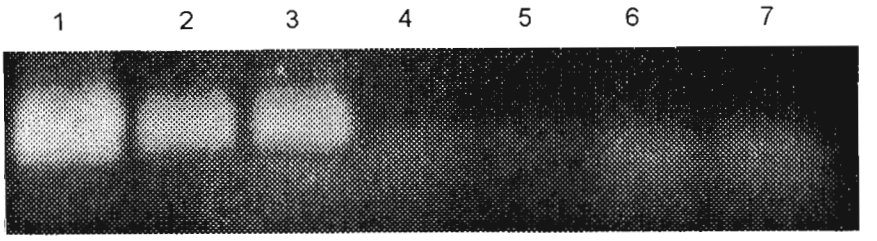

Figure 3.8 Effect of differing concentrations of formamide on the amplification of cultivar N17 by microsatellite primer pair, SMC17AUQ. Reactions were run on a agarose gel as follows: No formamide added (Lane 1); 2% (v/v) formamide present (Lane 2); 4% (v/v) formamide present (Lane 3); 5% (v/v) formamide present (Lane 4); 6% (v/v) formamide present (Lane 5); 8% (v/v) formamide present (Lane 6); 10% (v/v) formamide present (Lane 7).

It was therefore concluded that the addition of formamide was not reliable in improving amplification. Consequently, the use of additives was not included into the optimised PCR protocol.

3.3.3.6 Extension Temperature

Henegariu et al. (1997) found that reduced extension temperature resulted in visually higher yields of PCR products, while higher temperatures resulted in the absence of certain PCR products and the presence of some unspecific products. Two different extension temperatures (65°C and 73°C) were investigated in an attempt to improve the yield of the PCR amplification. PCRs were performed using cultivar N14 and microsatellite primer pairs SMC280CS, SMC662CS and SMC17AUQ and were replicated for the two different extension temperatures (Figure 3.9).

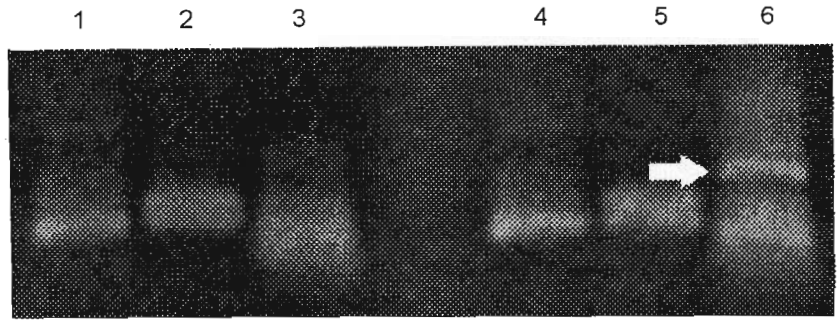

Figure 3.9 Effect of extension temperatures, 65°C and 73°C, on the amplification of cultivar N14 by microsatellite primer pairs SMC280CS, SMC17AUQ and SMC662CS. The reactions were run on an agarose gel and were as follows: lane 1-3 (65°C extension temperature) and lane 4-6 (73°C extension temperature); SMC280CS (lane 1 and 4); SMC17AUQ (lane 2 and 5) and SMC662CS (lane 3 and 6). The white arrow indicates an additional amplification product.

Unexpectedly, it was found that the higher extension temperature of 73°C resulted in improved yields and the amplification of a band that was absent at 65°C. The extension temperature originally used was 73°C and as such this PCR parameter remained unchanged.

3.3.3.7 Buffer Concentration

Generally the amplification of longer products is favoured at lower salt concentrations (lower buffer concentration), while smaller amplicons are favoured at higher buffer concentrations. This is due to longer DNA templates not denaturing at high salt concentrations (Henegariu et al., 1997). PCR reactions of cultivar N14 and microsatellite primer pairs SMC280CS and SMC17AUQ, respectively, were replicated and a 10X PCR buffer (100mM Tris-HCl, 500mM KCl, pH 8.3) was aliquoted to give four different PCR buffer concentrations: 0.5X, 1.0X, 1.5X and 2.0X (Figure 3.10).

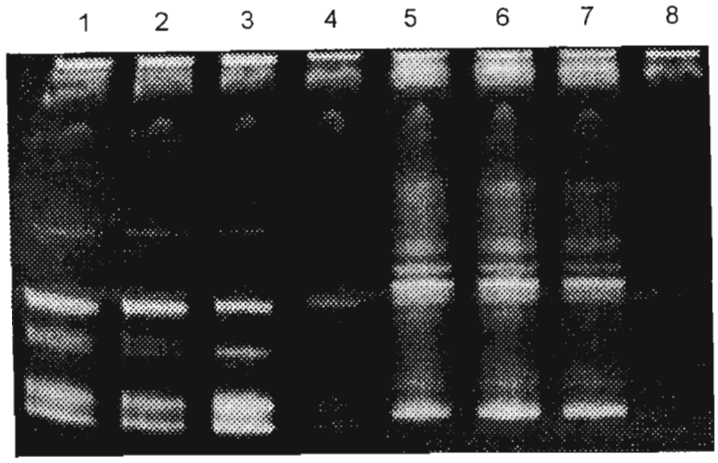

Figure 3.10 Effect of PCR buffer concentration on the amplification of cultivar N14 by microsatellite primer pairs SMC280CS and SMC17AUQ. The reactions were separated on a 8% denaturing polyacrylamide gel as follows: SMC280CS with 0.5X buffer (lane 1); SMC280CS with 1.0X buffer (lane 2); SMC280CS with 1.5X buffer (lane 3); SMC280CS with 2.0X buffer (lane 4); SMC17AUQ with 0.5X buffer (lane 5); SMC17AUQ with 1.0X buffer (lane 6); SMC17AUQ with 1.5X buffer (lane 7) and SMC17AUQ with 2.0X buffer (lane 8).

The results in Figure 3.10 showed that 1X PCR buffer (10mM Tris-HCl pH 9.0 and 50mM KCl) was the most suitable for microsatellite amplification, and its use was made standard for subsequent PCR.

3.3.4 Evaluation of Different Gel Electrophoresis Systems

Initially, Seakem LE agarose gels (Section 3.2.7.1) were used to evaluate the 26 SSR primers. Although these gels were able to eliminate those microsatellite primer pairs that produced no amplification products, they proved incapable of providing sufficient resolution to allow for accurate microsatellite scoring (Results not shown). As a result, high quality agarose gels (Section 3.2.7.2) were used in their stead. However it was found that these gels were not able to resolve the amplicons as discrete bands, or resolve small base pair differences between amplicons of different cultivars.

Consequently, a denaturing polyacrylamide gel system using SYBR Gold Stain (Molecular Probes) was developed and was found to resolve bands more discretely and clearly than high quality agarose gel electrophoresis with ethidium bromide staining. Figure 3.11 shows the comparative resolution of the two gel systems.

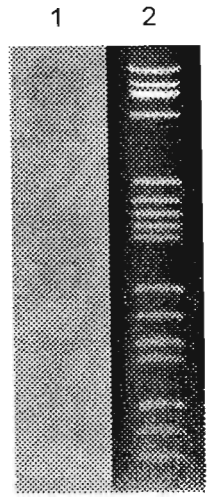

Figure 3.11 Molecular weight marker 5 run on: (1) 3% MS-8 high quality agarose gel stained with ethidium bromide and (2) 8% denaturing polyacrylamide gel visualised with SYBR Gold stain.

This denaturing polyacrylamide gel system (8% denaturing polyacrylamide gel visualised with SYBR Gold stain) is better suited to the separation of microsatellite amplification products because of the presence of a denaturing agent, urea, which suppresses base pairing and the formation of secondary structure in the DNA. As a result denatured DNA migration through the gel is almost completely independent of base composition and sequence (Sambrook & Russell, 2001). Furthermore, SYBR Gold staining was used in place of ethidium bromide staining for the visualising of the separated amplification products because of its superior sensitivity, ease of use and its comparatively low level of background.

3.4 CONCLUSION

In this chapter, various aspects of the microsatellite protocol from PCR reaction components and cycling conditions to gel electrophoresis and visualisation parameters were explored in an attempt to increase PCR yield and stringency, and improve gel separation and visualisation respectively. It was found for the microsatellite primers tested (presumed to be representative of all primers) that the majority of experimentally derived optimised PCR parameters were accordant with those of the initial modified PCR protocol derived from Cordeiro et al., (2000) except that:

1. 50ng of template amplified more reproducibly than the original value of 25ng.
2. 1.5U of Taq polymerase (Promega) amplified better than the initial quantity of 1U.
3. PCR yield from 35 cycles was considerably superior to that achieved with the initial value of 30 cycles.

Optimised Sugarcane Microsatellite Protocol

PCR reaction: 25µl volume containing 50ng of DNA, 200 µM dNTPs, 0.2 µM of each primer, 2.5mM $MgCl_2$, 1.5 U Taq polymerase (Promega), 1X PCR Buffer (10mM Tris-HCl pH 9, 50mM KCl and 0.1% Triton® X-100) and made up with sterile water. Cycling conditions (run on a GeneAmp 9700 thermocycler): 94°C for 5 min; followed by 35 cycles of 94°C for 30 sec, appropriate annealing temperature (either 50°C or 55°C) for 30 sec, 73°C for 30 sec and a final extension step of 73°C for 3 min. Electrophoretic separation: 8% denaturing polyacrylamide gel (27:1) according to Section 3.2.7.3, visualised using SYBR GOLD Stain as described in Section 3.2.8.2.

In this chapter it was shown that optimisation of the PCR parameters and gel system yielded a protocol that is robust and informative. This protocol was used for parentage determination as reported in Chapter 4.

CHAPTER 4

APPLICATION OF MICROSATELLITES TO PARENTAGE ANALYSIS: ELUCIDATING THE AA40 PROGENITORS

4.1 INTRODUCTION

A sugarcane genetic mapping project was initiated at SASEX in 1993, in order to identify quantitative trait loci (QTLs) related to important phenotypic characters. Six biparental crosses were planted in the field and the progeny of one cross, AA40 (N18 x CP57/614), that showed good segregation for a number of important traits, was selected for the mapping project (Barnes & Bester, 2000). Although, at the outset of the mapping project, RAPD analysis of the AA40 cross provided no evidence of marker incompatibility between parents and progeny, subsequent RFLP data showed not only that the AA40 progeny arose from a single crossing event, but that there was a relatively high proportion of bands that were present in the progeny but not in either of the two parents (non-parental bands) (Harvey & Huckett, 1998). Furthermore, the relatively high percentage of band sharing observed in the RFLP data between the AA40 progeny and cultivar N18 (female parent) and the improbability that the female parent would be misidentified, implied that the male parent of AA40, thought to be CP57/614, had been misidentified at some stage prior to crossing, or during the crossing procedure (Harvey & Huckett, 1998; Barnes & Bester, 2000). An attempt was made to identify the male parent of the AA40 population by carrying out RFLP analysis on 10 AA40 progeny, the female N18 and 14 putative males, present in the glasshouse at the time of the AA40 cross. However this was unsuccessful and no definitive identification of a single variety as the male parent could be made (Barnes & Bester, 2000).

It is recognised that incorrect pedigree or parentage is not unexpected or uncommon in sugarcane breeding programmes, either at SASEX or

internationally, as there is ample opportunity for both the misidentification of parental material and illegitimate pollination.

Microsatellites provide a means of verifying pedigrees of valuable germplasm by applying the "principle of exclusion" whereby nearly conclusive proof of non-paternity is derived from parent-offspring marker genotype data incompatibility (Chakraborty et al., 1988; Buteler et al., 1997). To expand on this, markers possessed by the progeny at the involved loci permit all those potential parent pairs to be ruled out that could not have contributed to these markers (presence of non-parental bands). If all but one pair of potential parents can be eliminated, then the remaining parent pair is unambiguously assigned. When more than one remains, which is very often the case, then more loci need to be investigated until an unambiguous assignment is achieved (Gillet, 1999). This type of elimination or exclusion strategy is the only one possible in the case of a complex polyploid, such as sugarcane.

In this investigation microsatellite analysis was carried out, to resolve the AA40 parentage, on 13 potential parent cultivars and 10 AA40 progeny using the optimised STMS protocol and evaluated primer pairs from Chapter 3. The 13 cultivars were composed of six potential parent pairs (eleven cultivars) that were present in the breeding glasshouse at the time of the AA40 cross and were chosen to investigate the possibility of mislabelling at planting, and two crosses or potential parent pairs (four cultivars) that were planted next to the AA40 progeny in the field and were selected to investigate the assertion of mislabelling at seed collection.

Recently, the validity of parentage analysis based on parent-progeny genotype incompatibility (presence of non-parental bands) has been questioned (Crouch et al., 1999a, 1999b) and in polyploids it has been suggested that PCR recombination could occur, due to genic redundancy, leading to non-specific chimeric amplicons (Cronn et al., 2002). In this study PCR reactions were carried out on both single parent DNA and pooled parent pair DNA samples, as

a means to determine the incidence of non-parental bands in sugarcane varieties and their impact on parentage analysis. Pooled parent pair DNA reactions are referred to as 'synthetic offspring' and their amplification products, it is believed, contain the full complement of bands that could be in any single progeny of the parent pair (Carlson, 1991; Hadrys et al., 1992).

4.2 MATERIALS AND METHODS

4.2.1 Sugarcane Cultivars

Twenty-four cultivars (Table 4.1) were used for the parentage investigation, and were composed of 13 potential parent cultivars representing 8 crosses, cultivar 84F2753 (progeny of another cross between N18 and CP57/614) and 10 AA40 progeny.

Table 4.1 Sugarcane cultivars used for parentage analysis, both potential parents and progeny.

No.	Name	No.	Name
1	N14	13	84F2753*
2	N17	14	76F0879
3	N18	15	AA40 02
4	NCo376	16	AA40 09
5	CP57/614	17	AA40 41
6	81W50	18	AA40 43
7	CP56/59	19	AA40 61
8	Co1001	20	AA40 82
9	75F2297	21	AA40 90
10	MZC74/275	22	AA40 111
11	77F790	23	AA40 126
12	79F2011	24	AA40 143

*84F2753 is the progeny of a previous cross made between N18 and CP57/614

4.2.2 Sugarcane Microsatellite Primers

Twenty-one SSR primers that were previously shown to amplify PCR products (Section 3.3.2) were selected (Table 4.2). The primers were diluted to 50µM with sterile distilled water and then pooled into 20 µl sub-aliquots. All primers were stored at –20°C.

4.2.3 Isolation of Sugarcane DNA

Fresh sugarcane leafroll was collected and ground in liquid nitrogen and total DNA extracted following a modification of the procedure of Dellaporta et al., (1983). The protocol used is described in Section 3.2.3, except that Whatmann 91 filter paper was used, instead of four layers of gauze cloth, to filter the supernatant.

4.2.4 DNA Quality and Quantity Assessment

The DNA stocks (Table 4.1) were quantified using a fluorometer (Hoefer DyNA Quant 200) as described in Section 3.2.4 and 50 ng/µl dilutions were made and stored at –20°C.

4.2.5 Polymerase Chain Reaction

The PCR amplifications were carried out in 25 µl reaction volumes and run on a GeneAmp 9700 thermocycler, as described in Section 3.4.

4.2.6 Denaturing Polyacrylamide Gels

Eight percent denaturing polyacrylamide (27:1) gels were prepared as described in Section 3.2.7.3 and SYBR Gold (Molecular Probes) staining was carried out according to Section 3.2.8.2.

Table 4.2 The twenty-one sugarcane microsatellite primer pairs assessed for use in the parentage analysis of the AA40 population.

Name	Repeat	Forward Primer Sequence	Reverse Primer Sequence	Used T_A	Size Expected
SMC334BS*	$(TG)_{36}$	CAA TTC TGA CCG TGC AAA GAT	CGA TGA GCT TGA TTG CGA ATG	50	213
SMC336BS*	$(TG)_{23}(AG)_{19}$	ATT CTA GTG CCA ATC CAT CTC A	CAT GCC AAC TTC CAA ACA GAC	50	233
SMC703BS	$(CA)_{12}$	GCC TTT CTC CAA ACC AAT TAG T	GTT GTT TAT GGA ATG GTG AGG A	50	215
SMC805BS	$(TGC)_7$	TGA GAA TGC TGT CAT AGG GCT TG	TCG AGG TGA ATC ATT GCC TTC A	55	264
SMC843BS	$(TG)_{16}CAA(GT)_9$	GGT CCC ATT GAT GTG GCA	CTA GGA CCT TGT GGT TAC CGT	50	201
SMC17CG	$(GA)_{22}$	AAG GTA GCA CGA AAC ACG TCG AT	AAC CCC AGC GGC GAT CTC T	55	194
SMC226CG	$(CA)_{10}$	GAG GCT CAG AAG CTG GCA T	ACC CTC TAT TTC CGA GTT GGT	50	136
SMC1039CG	$(TG)_{17}$	AGG TGA GAG TTC CTG GCT TTC CA	TGT GCT GGC AAG ATA CCC CTA CTT	55	189
SMC278CS*	$(TG)_{19}(AG)_{25}$	TTC TAG TGC CAA TCC ATC TCA GA	CAT GCC AAC TTC CAA ACA GAC T	50	236
SMC280CS*	$(CA)_{21}$	TGA TCG CAC GTT GTA TCC AAC A	TTT GAC CAC GCC ACG GTA GAT	55	238
SMC286CS*	$(TG)_{43}$	TCA AAT GGG ACC TTA TTG GAG	TCC CTC GAT CTC CGT TGT T	50	202
SMC569CS*	$(TG)_{37}$	GCG ATG GTT CCT ATG CAA CTT	TTC GTG GCT GAG ATT CAC ACT A	50	273
SMC640CS	$(CT)_5/(CA)_8$	TTA AGA GAC CCG CCT TTG GAA	TGC CAG AAG TGG TTG TGC TCA	55	230
SMC662CS	$(AC)_{31}$	GAC TGC ATG GCT TGC TGA TCG	GGA CCT TGG CGG TGA TGG G	55	241
SMC687CS	$(CAG)_8$	AGC CAT GCA GGC AGG CAT	CGC ACA ATC TGC AAG TGC ATC A	55	203
SMC213MS*	$(CA)_{12}$	GCA GGG AGA CGA ACA CGA GT	TGC AAC CTT CTT CAG GCT TGA	55	148
SMC219MS*	$(CA)_{40}$	TCT CCC TCG ATC TCC GTT GT	GGA GTG TCT TCA GCT ATC GGA	50	185
SMC221MS	$(TG)_{25}$	CAT GCC AAC TTC CAA ACA GAC T	GGT GAT GCG AAG AGA TTG GA	50	204
SMC17AUQ*	$(TG)_{16}$	CGT AGG CGA GAG GCT TAT CAA A	TGT CGG TCA CCC TCC AAG GA	55	224
SMC36BUQ*	$(TTG)_7$	GGG TTT CAT CTC TAG CCT ACC	TCA GTA GCA GAG TCA GAC GCT T	50	118
SMC31CUQ	$(TC)_{18}(AC)_{22}$	CAT GCC AAC TTC CAA TAC AGA CT	AGT GCC AAT CCA TCT CAG AGAN	50	225

Microsatellites primers used for the parentage analysis of the AA40 population.

4.2.7 Gel Analysis and Scoring

For each genotype, the bands on the gels were scored as either present (1) or absent (0), and the marker data of potential parental pairs were compared to that of ten randomly chosen AA40 progeny.

4.3. RESULTS

4.3.1 Genomic DNA Yield and Purity

Sugarcane genomic DNA was extracted from selected sugarcane varieties using the modified Dellaporta method (Section 3.2.3), and the average yield was 338.5±123.4ng/µl. The purity values were determined by calculating the A_{260}/A_{280} ratio. The average purity of the 24 cultivars was 1.69±0.11, which is relatively pure considering that pure DNA has an A_{260}/A_{280} ratio of 1.8. The purity value being below 1.8 indicates the presence of residual protein contamination in the samples (Sambrook & Russell, 2001).

4.3.2 Evaluation of Microsatellite Primer Pairs for Parentage Analysis

The 24 cultivars were amplified using the set of 21 microsatellite primer pairs (Table 4.2) and their amplification products were resolved on 8% denaturing polyacrylamide gels. Two gels were used for each microsatellite evaluation, one progeny gel and one parental gel, with cultivars N14, N18 and CP57/614 being duplicated to standardise the gels and ensure comparability. The gels for each microsatellite primer pair were evaluated and of the 21 primer pairs, 19 (90%) generated amplification products and of these; nine (47%) produced monomorphic (Figure 4.1) or un-scorable banding patterns.

Ten SSR primers (SMC36BUQ, SMC280CS, SMC17AUQ, SMC278CS, SMC336BS, SMC213MS, SMC334BS, SMC286CS, SMC219MS and

SMC569CS) of the 21 tested were found to generate polymorphic scorable banding patterns, and these primers were used for the parentage analysis.

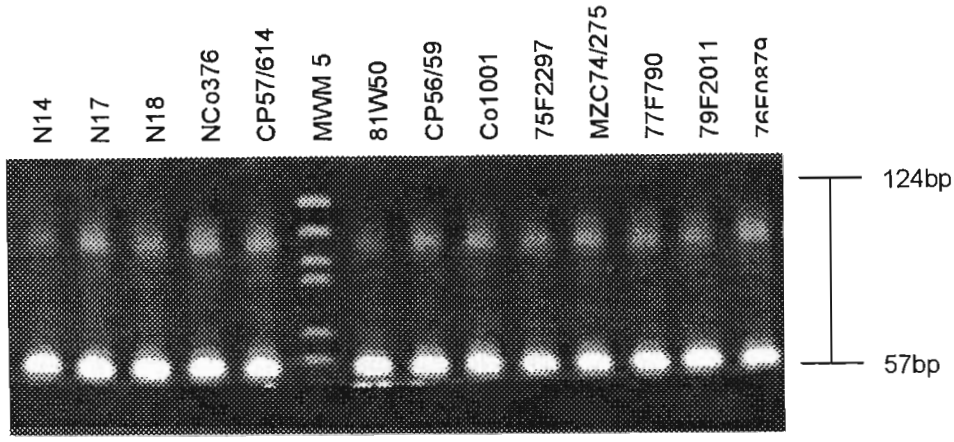

Figure 4.1 Polyacrylamide gel separation of the monomorphic amplification products of microsatellite primer pair, SMC1039CG.

Figure 4.2 shows the separated PCR amplification products, from one of the ten polymorphic microsatellite primer pairs (SMC336BS), on a denaturing polyacrylamide gel.

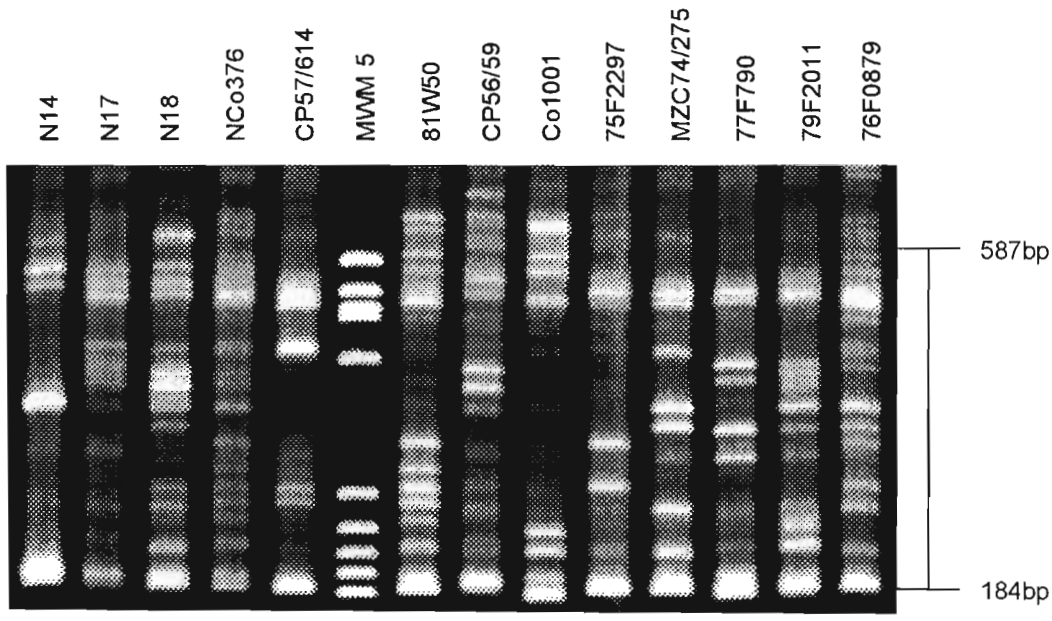

Figure 4.2 Polyacrylamide gel separation of the amplification products of microsatellite primer pair, SMC336BS.

4.3.3 Parentage Analysis

It was suspected that mislabelling at planting or at seed collection could have led to the incorrect pedigree of the AA40 population. To investigate these assertions of both mislabelling at planting (six potential parent pairs) and in a restricted manner that of mislabelling at seed collection (two potential parent pairs), the 24 cultivars were fingerprinted with 10 SSR primer pairs (Appendix 1). For each parent pair, if a particular marker was absent (0 x 0) then it would be considered to be a discriminating marker. If present (1) in any of the progeny, these discriminating markers would constitute a mismatch between the parental pair and the progeny. In parentage analysis the occurrence of these mismatches, referred to as non-parental markers, could be used to eliminate that parent pair from further consideration. The number of discriminating and non-parental markers found across all microsatellites for each potential parent combination is presented in Table 4.3.

Table 4.3 The number of discriminating and non-parental markers for the eight potential parent pairs.

Parents	Cross Name	No. of discriminating markers	No. of non-parental markers
N18 x CP57/614[1]	AA40	16	16
81W50 x N14[1]	Z586	8	8
CP56/59 x Co1001[1]	AA1312	14	13
Nco376 x 75F2297[1]	AA1292	7	6
MZC74/275 x 77F790[1]	Z927	15	14
79F2011 x N14[1]	AA157	18	18
N17 x CP57/614[2]	N/A	14	13
MZC74/275 x 76F0879[2]	N/A	10	9
Average	---	13	12

[1]Parental pairs included to investigate mislabelling at planting. [2]Parental pairs included to investigate mislabelling at seed collection.

The non-parental marker data were used to calculate what percentage of the potential parent pairs was excluded for each microsatellite (Table 4.4).

Table 4.4 The ten microsatellites used for the parentage, including the repeat sequence, number of markers scored and percentage of parent pairs excluded.

Name	Repeat	No. of markers	% of parent pairs excluded
SMC336BS	$(TG)_{23}(AG)_{19}$	7	100%
SMC569CS	$(TG)_{37}$	9	75%
SMC278CS	$(TG)_{19}(AG)_{25}$	6	75%
SMC213MS	$(CA)_{12}$	6	63%
SMC17AUQ	$(TG)_{16}$	11	50%
SMC219MS	$(CA)_{40}$	8	50%
SMC36BUQ	$(TTG)_{7}$	11	25%
SMC280CS	$(CA)_{21}$	8	25%
SMC286CS	$(TG)_{43}$	7	25%
SMC334BS	$(TG)_{36}$	2	13%
Total		75	---

4.3.4 Cultivar 84F2753 as a Parentage Analysis Control

Cultivar 84F2753 is one of the progeny from a cross made previously between N18 and CP57/614. This cultivar functioned as a 'positive control' for the parentage analysis, in that if the microsatellite data showed that N18 and CP57/614 were the progenitors and that the pedigree was correct (no non-parental bands) then this approach would be suitable for determining the true parents of the AA40 population. The banding patterns of cultivars N18 and CP57/614 were compared to those of cultivar 84F2753 and each individual AA40 progeny, of the set of ten, and the results are presented in Table 4.5.

4.3.5 Origin of the Non-Parental Bands

It was postulated that some of the non-parental markers observed in the parentage analysis could be found to be of a PCR artifactual nature (Crouch et al., 1999a, 1999b; Cronn et al., 2002) and could be scored or accounted for in 'synthetic offspring'. A synthetic offspring is produced by mixing equal amounts of DNA from both the female and male parent and is a complete representation of both parental genomes, matching the profile of a large sample of offspring,

since in both the 'synthetic' and real offspring allele combinations that may cause amplification artefacts are represented (Hadrys et al., 1992).

Table 4.5 The number of discriminating and non-parental markers for each AA40 progeny and cultivar 84F2753, determined from the banding patterns of parent pair N18 and CP57/614.

Name of variety	Number of Discriminating markers	Number of non-parental markers
AA40 02	16	11
AA40 09	16	9
AA40 41	16	4
AA40 43	16	9
AA40 61	16	5
AA40 82	16	12
AA40 90	16	10
AA40 111	16	5
AA40 126	16	12
AA40 143	16	10
84F2753	16	0

To test this, six microsatellite primer pairs (SMC36BUQ, SMC278CS, SMC336BS, SMC569CS, SMC17AUQ and SMC280CS) were used to amplify reactions containing 50 ng of each potential parent pairs DNA (25 ng DNA of each potential parent cultivar pooled according to the crosses in Table 4.3.) as well as 6 AA40 progeny (AA40 09, AA40 43, AA40 90, AA40 111, AA40 126 and AA40 143). Figure 4.3 shows an example of the differences between the single template PCR and the pooled template PCR.

In most of the pooled template reactions, as in Figure 4.3, the presence of novel markers (not present in single PCR), in conjunction to the absence of certain single template PCR bands (absent markers) were observed. The results are presented into Table 4.6.

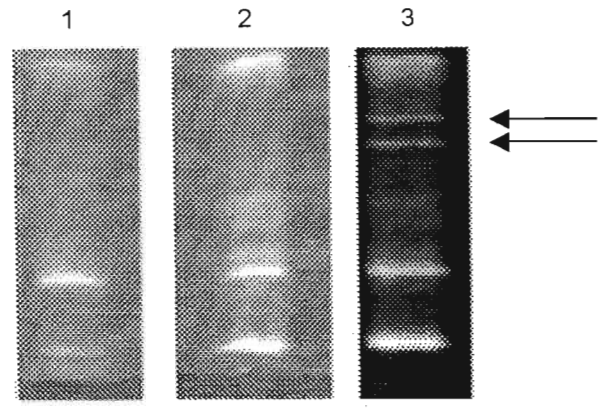

Figure 4.3 Differing banding patterns of microsatellite SMC569CS from both single template PCR and pooled template PCR. The reactions were as follows: 81W50 (Lane 1), N14 (Lane 2) and a pooled template reaction of N14 and 81W50 (Lane 3). Two bands that are not observed in both 81W50 (Lane 1) or N14 (Lane 2) are present in the pooled template reaction (Lane 3).

Table 4.6 The pooled template PCR results for the eight putative parent pairs.

Name	No. of markers scored	% of novel markers*	% of absent markers**
N18 + CP57/614	41	2%	0%
81W50 + N14	48	4%	2%
CP56/59 + Co1001	41	2%	0%
Nco376 + 75F2297	45	0%	0%
MZC74/275 + 77F790	44	5%	0%
79F2011 + N14	40	10%	0%
N17 + CP57/614	39	3%	3%
MZC74/275 + 76F0879	45	4%	0%

* Percentage of markers found in the pooled template reaction but absent in the single temple PCR reactions. ** Percentage of markers present in single template PCR but absent in the pooled template PCR.

4.4 DISCUSSION

4.4.1 Identity of AA40 Parents

Ten microsatellites were used for the parentage determination of the AA40 population and generated a total of 75 markers, of which 19 (25%) were monomorphic. The number of markers recorded per microsatellite primer pair

across the 13 potential parent cultivars tested, ranged between two and eleven with an average of eight. Microsatellite primer pair SMC336BS was the most informative for the parentage determination, as its seven scored markers found non-parental bands for all eight of the potential parent pairs. For each parent combination a significant number of non-parental markers, 12 on average, were observed in the progeny and this suggests that none of these parent pairs are likely to be the true parents of AA40. Furthermore, considering the high number of non-parental markers observed, it is unlikely that microsatellite mutations could have led to misinterpreting the true parents as being incorrect. Consequently, it appears that the mislabelling of the cross did not occur at the time of planting; suggesting that perhaps, mislabelling occurred when the cross was made.

4.4.2 Non-Mendelian Mechanisms in Polyploids

Relatively little is known about the nature and scope of the genetic interactions which occur during and immediately after allopolyploid formation. However recent studies in *Brassica* (Song et al., 1995) and *Triticum* (Liu et al., 1998a, 1998b), for example, have shown that allopolyploid formation has been associated with an unexplained appearance of novel genomic fragments in Southern blots, less than full additivity of parental genomes (disappearance of parental fragments) and methylation changes. Furthermore, Hanson et al., 1998, 1999 has shown that some repetitive DNAs, such as rDNA and retrotransposons, in allopolyploid cotton (Gossypium) are subject to non-Mendelian molecular evolutionary phenomena or sequence loss (Pestsova et al., 1998). These results disagree with those of Liu et al., 2001 who found genetic additivity and epigenetic stasis in *Gossypium*. These findings show that in some allopolyploids there are rapid genetic and genomic interactions during the initial stages of formation, which include non-Mendelian genetic mechanisms, for which satisfactory mechanistic explanations are lacking (Lui et al., 2001). Thus with this in mind, it could be postulated that the occurrence of non-parental markers and for that matter the disappearance of other markers in the AA40 population could be as a result of novel intergenomic interactions or

several epigenetic processes, such as nucleolar dominance (Comai, 2000), gene silencing (Pikaard, 2000), aneuploid reduction or mobile element activation (Lui & Wendel, 2000), the latter also resulting in genetic change (Liu & Wendel, 2002).

Cultivar 84F2753 was included in the set of 24 cultivars used in the parentage investigation as a means to verify the usefulness and authenticity of this microsatellite approach as well as to ascertain whether non-Mendelian genetic mechanisms occurred during sugarcane hybridisation. Cultivar 84F2753 and the 10 AA40 progeny were analysed with 10 microsatellite primer pairs and the banding patterns compared to that of N18 and CP57/614. The data suggested that N18 and CP57/614 were the progenitors of 84F2753, with no observed non-parental bands; while for the AA40 progeny a minimum number of four non-parental bands were observed. The significance of this is that the STMS approach was able to confirm the pedigree of 84F2753 and verify that N18 and CP57/614 were not the parents of the AA40 population. This is in accordance with the results of Buteler et al., (2002), who observed no allelic incompatibility between progeny and their true parents in polyploid sweetpotato (*Ipomoea batatas*).

Although this limited experiment suggests that non-Mendelian phenomena are not a common aspect of sugarcane hybridisation, more investigation is required to confirm this.

4.4.3 Legitimacy of Parentage Analysis

In microsatellite-based parentage analysis of complex genomes, such as sugarcane, the presence of non-parental bands is grounds to exclude a parental pair as being the true parents. However in certain species the guidelines for exclusion are being relaxed as more and more evidence mounts that even in situations when the true parents are compared to their progeny, a percentage of non-parental markers persist, although there origin is unclear. Crouch et al., (1999b) found that in Plantain (*Musa* spp., AAB group) 20% of the

markers scored were not present in either parent of a cross, yet were found to segregate in the progeny.

Based on the observations of Crouch et al., (1999a, 1999b), it was postulated that some of the non-parental bands observed in the parentage analysis could be found to be of a PCR artifactual nature. The comparison between the banding patterns of the single template reactions and those of the pooled template samples or 'synthetic offspring' (Hadrys et al., 1992) revealed that the mean of novel markers was 3.8%±2.9% with a maximum of 10%, and that the mean for absent markers was 0.6%±1.1%. Although the process responsible for the different amplification results between the single template and pooled template reaction products is not completely understood, various causal factors could be responsible such as: miss-priming, heteroduplex formation, incomplete denaturing of PCR amplicons prior to gel electrophoresis or PCR recombination. The latter occurs in polyploids due to the presence of numerous highly similar, non-identical target sequences and results in the formation of intergenic chimeras of specific PCR products (Cronn et al., 2002). The characteristics of the template such as purity, degree of degradation, conformation and complexity can also play a significant part in PCR artefact formation.

Considering the propensity of polyploids, such as sugarcane, to amplify non-specific products it is advisable that a certain percentage of non-parental bands, perhaps 10% (maximum value), be tolerated in microsatellite-based parentage determination. Alternatively, the banding pattern of a 'synthetic offspring' for a potential parent pair should be able to account for any PCR artifactual non-parental markers that are observed in the progeny and if not then, perhaps, these markers represent genuine parent-progeny marker incompatibilities and the parentage is incorrect.

CHAPTER 5

MICROSATELLITE FINGERPRINTING OF SUGARCANE

5.1 INTRODUCTION

The identification of sugarcane varieties is of importance to the South African Sugar Industry for efficient maintenance of their large germplasm collections, as well as ensuring that the plants selected as parents in the breeding programme, and the resulting progeny, are correct. Moreover, the desire for the protection of intellectual property, specifically Plant Breeders' Rights (PBR), has fostered the need for irrefutable identification of cultivars.

At SASEX, sugarcane morphological or phenotypic markers are used for varietal identification. However the assessment of these phenotypic markers is often time-consuming, labour intensive and requires mature sugarcane for assessment. In recent years, microsatellite markers have become increasingly popular as the DNA marker system used in many crops for varietal identification (Cordeiro, 2001), with the most commonly used approach being Sequence Tagged Microsatellite Site amplification (STMS). The STMS approach has been used successfully for fingerprinting in numerous plant species, such as barley, *Hordeum vulgare* (Russell et al., 1997); wheat, *Triticum aestivum* (Gupta et al., 1999); and sugarcane (Piperidis et al., 2001).

STMS-based fingerprinting or genotyping requires that microsatellite primer pairs generate unique banding patterns or fingerprints for each individual or line assessed (Warburton & Hoisington, 2001) and in this regard the high ploidy number in sugarcane can be an advantage, potentially allowing unique fingerprints to be created using only a single microsatellite. Determining the discriminating power (D) of a microsatellite provides a means of evaluating its efficiency for varietal identification and is defined as the probability that two randomly chosen individuals have different patterns or are distinguishable from

one another (Tessier et al., 1999). Prudence however, is always necessary when using a small number of primers to identify a set of varieties, as the same set may not be sufficient to distinguish individuals from a larger sample group. The determination of the theoretical number of non-differentiated pairs of cultivars (X_K) provides an estimator for the number of microsatellites required for large scale fingerprinting (Tessier et al., 1999).

The aim of this study was to investigate the potential of STMS for sugarcane varietal identification, using a set of cultivars and a subset of the microsatellite primer pairs received from the ISMC. However, this study coincided with a five member international initiative, consisting of BSES (Australia), CIRAD (France), Copersucar (Brazil), MSIRI (Mauritius) and SASEX (South Africa), to verify the reproducibility of the STMS approach across different laboratories. As a result it was decided, for expedience, that the 20 cultivars and 14 microsatellite primer pairs that were compiled for the reproducibility test be adopted for the SASEX fingerprinting investigation. Consequently, the protocols used in this study, namely P^{33} radiolabelling and PAGE, were prescribed by the international initiative and differed from those used in Chapters 3 and 4.

Before an approach such as STMS can be endorsed or widely accepted for fingerprinting, it must be shown to be insensitive to slight variations in reaction conditions and be able to generate reproducible fingerprints. To investigate the impact environmental variation (different origins), type and condition of plant material and DNA extraction protocol have on STMS banding pattern reproducibility, a total of 16 different DNA samples of cultivar NCo376 were amplified by six microsatellite primer pairs.

In this study: (1) twenty cultivars were fingerprinted by 14 microsatellite primer pairs, (2) the theoretical number of microsatellite markers required for large scale fingerprinting of South African sugarcane germplasm was determined and (3) the sensitivity of the STMS approach to DNA quality, origin and DNA extraction protocol, was investigated by comparing the banding patterns of

various DNA samples of cultivar NCo376. As part of the aforementioned international initiative, the fingerprinting data in this chapter were passed onto BSES for assessment of the reproducibility across laboratories, however BSES has not completed the analysis and the results are still outstanding.

5.2 Materials and Methods

5.2.1 Sugarcane Microsatellite Primers

Fourteen sugarcane microsatellite primer pairs (Table 5.1) were used for the fingerprinting investigation. The primers were diluted to 50µM with sterile distilled water and then pooled into 20 µl sub-aliquots. All primers were stored at –20°C.

5.2.2 Cultivars for Fingerprinting

Twenty cultivars (Table 5.2) were selected for the fingerprinting investigation. Total DNA was extracted from the 16 cultivars available at SASEX, following a modified procedure of Dellaporta et al., (1983). Pelleted DNA samples of the four cultivars that were unavailable at SASEX were received from BSES (LCP85-384) and CIRAD (Mandalay, B46364 and RB72-454). The pelleted DNA samples were re-suspended in sterile distilled water and left overnight to dissolve. The DNA stocks were quantified and 20 ng/µl dilutions were made and stored at –20°C.

Table 5.2 The 20 sugarcane cultivars used for the fingerprinting investigation.

No.	Name	No.	Name	No.	Name	No.	Name
1	Coimbatore	6	Co270	11	NCo376	16	Q96
2	Mandalay	7	Co290	12	POJ2878	17	Q117
3	Badila	8	Co475	13	Q115	18	R570
4	Black Cheribon	9	LCP85-384	14	Q124	19	RB72-454
5	B46364	10	NCo310	15	Q136	20	SP70-1143

Table 5.1 The 14 sugarcane microsatellite primer pairs used for the fingerprinting.

Name	Repeat	Forward Primer Sequence	Reverse Primer Sequence	Used T_A	Size Expected
SMC36BUQ	$(TTG)_7$	GGG TTT CAT CTC TAG CCT ACC	TCA GTA GCA GAG TCA GAC GCT T	50	118
SMC119CG	$(TTG)_{12}$	TTTC ATC TCT AGC CTA CCC CAA	AGC AGC CAT TTA CCC AGG A	50	119
SMC286CS	$(TG)_{43}$	TCA AAT GGG ACC TTA TTG GAG	TCC CTC GAT CTC CGT TGT T	50	202
mSSCIR19	$(GA)_{23}$	GGT TCC AAA ATA CAC AAA	CAA TCT TAT CTA CGC ACT T	50	178
SMC334BS	$(TG)_{36}$	CAA TTC TGA CCG TGC AAA GAT	CGA TGA GCT TGA TTG CGA ATG	50	213
SMC336BS	$(TG)_{23}(AG)_{19}$	ATT CTA GTG CCA ATC CAT CTC A	CAT GCC AAC TTC CAA ACA GAC	50	233
mSSCIR10	$(CT)_{17}$	ACA CCA CTC ACA TCC ACT TG	TGA TAC ACC ATT GTT GAT GC	50	186
SMC278CS	$(TG)_{19}(AG)_{25}$	TTC TAG TGC CAA TCC ATC TCA GA	CAT GCC AAC TTC CAA ACA GAC T	50	236
mSSCIR56	$(GT)_{24}$	ATT TGA CGC TAC GAT GGT G	ATC CGT TTT TCA GCA GAG C	55	184
SMC749BS	$(TC)_{24}(CA)_{23}$	CAT GCC AAC TTC CAA ACA GAC T	CGT CTT GTG GAT TGG ATT GGA	55	263
mSSCIR60	$(GT)_{18}(GA)_9$	GGC TGC TGG CTG GGT TG	CAT CAT TCC GCC TGT CAT TG	55	229
mSSCIR12	$(GA)_{13}$	AAG AAG CGG AGG AGG ACA GAA T	CCC GTT CCA AGT TAC AGA CCA G	55	279
mSSCIR25	$(GA)_{24}$	TTG CCG TTG CCT GCT CT	CAC GCA CTC CAC TCA CAC C	55	300
mSSCIR55	$(GT)_{30}$	ARA TGT AGC AGT AGG ACC AA	CAA CAG GTT TCA GTA TAT TT	55	333

5.2.3 Differing Samples of Cultivar NCo376 for fingerprinting

A total of 16 DNA samples of cultivar NCo376 were extracted (Table 5.3) either from leaf material or from leafroll. A modified procedure of Dellaporta et al., (1983) was used to extract DNA from the fresh leafroll of sugarcane stalks, while DNA was extracted from leaf material according to Section 5.2.4.2. The DNA stocks were quantified and ~10 ng/µl dilutions were made and stored at –20°C.

Table 5.3 Sixteen DNA samples of cultivar NCo376, including their origin and extraction material.

No.	Origin	Extraction Material
1	Field 43, SASEX	1 stalk
2	Field 43, SASEX	1 stalk
3	Field 43, SASEX	1 stalk
4	Field 43, SASEX	1 stalk
5	Field 43, SASEX	1 stalk
6	Field 43, SASEX	5 stalks
7	Kearsney Expt. Station	5 stalks
8	Crookes Bros. Ltd Farm	5 stalks
9	Gingindlouw Farm	5 stalks
10	Brynus Hill Farm	5 stalks
11	Field 43, SASEX	5 stalks cut and left at room temperature for 3 days.
12	Field 16, SASEX	Young leaf
13	Field 16, SASEX	Old leaf
14	Field 16, SASEX	Leaf cut and left at room temperature for 4 days.
15	Field 16, SASEX	Leaf cut and left at –80°C for 4 days.
16	Field 16, SASEX	Leaf infected with sugarcane mosaic virus.

5.2.4 Isolation of Sugarcane DNA

5.2.4.1 Modified Dellaporta DNA Extraction Method

Fresh sugarcane leafroll was collected and ground in liquid nitrogen and total DNA extracted following a modification of the procedure of Dellaporta et al., (1983). The protocol used is described in Section 3.2.3, except that Whatman 91 filter paper was used, instead of four layers of gauze cloth, to filter the supernatant.

5.2.4.2 Genomic DNA Extraction from Leaf Material

Fresh leaf tissue (4-5cm long, minus midrib) was frozen in liquid nitrogen and then ground to a fine powder in a mortar and pestle. A spatula full of ground material was then transferred to a 1.5ml tube containing 500µl of extraction buffer (200mM Tris-HCl pH 7.5, 250mM NaCl, 25mM EDTA, 0.5%SDS), mixed and placed at 65°C for 30 minutes. The tube was then centrifuged for 20min at 2105Xg in a Sorvall® MC12V bench top centrifuge, and 300µl of the supernatant was transferred to a new 1.5ml tube containing 300µl isopropanol. The tube was mixed gently and then left for 15 minutes at room temperature, to allow DNA precipitation. The tube was then centrifuged for 30min at 2105Xg in a Sorvall® MC12V bench top centrifuge, and the supernatant removed. The pellet was then washed with 70% ethanol, air dried and then re-suspended in 100µl TE buffer (10mM Tris-Cl, 1mM EDTA, pH 8.0) at 37°C overnight (http://plantdev.bio.wzw.tum.de/methods/dnaisolation/genomicquickprepPCR2.html).

5.2.5 DNA Quality and Quantity Assessment

The DNA stocks (Table 5.2 and 5.3) were quantified using a fluorometer (Hoefer DyNA Quant 200) as described in Section 3.2.4 and dilutions were made and stored at –20°C.

5.2.6 Polymerase Chain Reaction

PCR amplification was carried out in a 20µl reaction volume, containing 30ng of DNA, 50 µM dNTPs, 0.25 µM of each primer, 2.5mM $MgCl_2$, 0.9U AmpliTaq Gold (Applied Biosystems), 1µCi [α-^{33}P] dCTP and 1X GeneAmp PCR Buffer (10mM Tris-HCl pH 8.3, 50mM KCl). The reactions were run on a GeneAmp 9700 thermocycler. Cycling conditions were: 94°C for 10 min; followed by 33 cycles of 94°C for 30 sec, appropriate annealing temperature (either 50°C or 55°C) for 30 sec, 73°C for 30 sec and a final extension step of 73°C for 3 min.

An equal volume of formamide loading buffer (98% formamide, 10mM EDTA pH 8.0, 0.05% bromophenol blue and xylene cyanol) was added to each PCR sample. Prior to electrophoresis, the samples were heated to 95°C for 3 minutes and then snap-cooled on ice.

5.2.7 Preparation of radiolabelled DNA ladder

A 30-330 bp AFLP DNA ladder (GibcoBRL) was radiolabelled as per manufacturers instructions, except that T4 polynucleotide kinase (Roche) and 10X Phosphorylation buffer (Roche) was used. Prior to electrophoresis the ladder was heated at 70°C for 5 minutes and then loaded onto the polyacrylamide gels as a molecular weight marker.

5.2.8 Denaturing Polyacrylamide Gels

Five percent denaturing (7.5M urea/ 1xTBE) polyacrylamide (19:1) gels were prepared following Sambrook & Russell (2001). The samples were run on either a BaseAce Sequencer (Stratagene) or a Dual Dedicated Height Nucleic Acid Sequencer (C.B.S. Scientific). The polyacrylamide gels were pre-electrophoresed at constant wattage (90 – 100W) for ~ 1 hour or until the gels reached a temperature of 55°C. The samples (4ul) were then loaded, and the gels were run at sufficient constant power to maintain a temperature of 50°C ~ 80W. Following electrophoresis the gels were dried at 80°C for 1h30min in a Model 583 Gel Dryer (BioRad) using a HydroTech Vacuum Pump (BioRad).

5.2.9 Autoradiography

The dried gel was exposed to Biomax MR X-ray film (Kodak) in a cassette for 3 days at room temperature. Following exposure, the X-ray was developed for 5 min in developer (Ilford Phenisol), briefly rinsed in stop solution (5% acetic acid) and fixed for 5min in fixative solution (Ilford Hypam). The X-ray was washed under tap water and allowed to dry before being viewed on a visible light box (Hoefer).

5.2.10 Gel Scoring and Analysis

The banding patterns of the 20 cultivars used in the fingerprinting study were scored by two independent scorers as either present (1) or absent (0), and the scoring compiled together. Markers were omitted if differences in scoring were observed.

To compare the efficiency of the microsatellite primer pairs for varietal identification, the discriminating power (D) of each primer pair was estimated. If *C* is the confusion probability, i.e. the probability that two randomly chosen individuals have the same banding patterns, then $D = 1 - C$ represents the probability that two randomly chosen individuals have different patterns, and thus are distinguishable from one another.

In a set of *N* individuals, it is possible to draw $N(N-1)/2$ different pairs. For the *i*th pattern of a given *j*th primer, present at frequency p_i in this set of varieties, the confusion probability c_i is:

$$c_i = p_i \frac{(Np_i - 1)}{N - 1} \qquad (1)$$

(Tessier et al., 1999)

For the jth primer, the confusion probability C_j is equal to the sum of the different c_i for all I patterns generated by the primer:

$$C_j = \sum_{i=1}^{I} c_i = \sum_{i=}^{I} p_i \frac{(Np_i - 1)}{N - 1} \quad (2)$$

(Tessier et al., 1999)

The discriminating power, D, of the jth microsatellite primer pair is determined as follows:

$$D_j = 1 - C_j = 1 - \sum_{i=1}^{I} p_i \frac{(Np_i - 1)}{N - 1} \quad (3)$$

(Tessier et al., 1999)

where N is the number of varieties being compared and p_i is the frequency of the ith microsatellite pattern. Theoretically, the total number of non-differentiated pairs of varieties for the jth primer is given by $X_j = (N(N-1)/2)C_j$. Therefore for a given combination of k primers, under the hypothesis of independence of the considered primers' patterns, X_k is calculated as:

$$X_k = \frac{N(N-1)}{2} \prod_{j=1}^{k} C_j \quad (4)$$

(Tessier et al., 1999)

5.3 RESULTS

5.3.1 Genomic DNA Extractions

The average DNA yields from leafroll extractions was 221.0±97.0ng/µl and 13.2±2.5ng/ul for the leaf extractions. The purity values of the DNA samples were determined by calculating the A_{260}/A_{280} ratio (Section 5.2.5). The average purity of the DNA extractions was 1.75 and 1.71 for the leafroll material (stalk) and the leaf extracts, respectively.

5.3.2 Polymorphisms Observed Between the 20 cultivars

For varietal identification each individual or line assessed must generate it's own unique banding pattern or fingerprint. In the STMS approach these banding patterns are composed of microsatellite-based PCR amplicons or markers. In general, the more unique banding patterns a microsatellite primer pair generates for a set of cultivars, the greater it's efficacy for fingerprinting. To examine the suitability of the STMS approach for fingerprinting or varietal identification, a set of 20 cultivars (Table 5.2) were fingerprinted using 14 microsatellite primer pairs (Table 5.1).

Using these microsatellite primer pairs a total of 86 markers (Appendix 2) were amplified and scored across the 20 cultivars. The number of polymorphic markers per microsatellite primer pair varied from 3 to 10 and generated 3 – 18 different banding patterns per primer pair. The discriminating power (D) (see equation 5.2.10.3) of each microsatellite primer pair was estimated to compare their informativeness and to evaluate their suitability for varietal identification (Table 5.4).

Discriminating power (D) is defined as the probability that two randomly chosen individuals have different banding patterns (Tessier et al., 1999). The relationship between discriminating power (D) and the number of unique banding patterns revealed for each microsatellite primer pair is shown in Figure 5.1

Table 5.4 Primer discriminating power (D) calculated for the 14 microsatellite primer pairs used to fingerprint a set of 20 sugarcane cultivars.

	Number of markers	Number of unique banding patterns	Discriminating power (D)	Non-differentiated pairs of varieties (X)
MSSCIR12	9	18	0.99	2
MSSCIR56	10	18	0.98	3
MSSCIR19	9	14	0.96	7
SMC336BS	8	15	0.95	9
SMC119CG	10	15	0.94	11
SMC278CS	7	11	0.93	13
SMC334BS	7	11	0.93	14
SMC286CS	6	7	0.84	30
SMC36BUQ	3	5	0.80	38
SMC749BS	3	6	0.76	46
MSSCIR25	4	6	0.68	61
MSSCIR10	3	3	0.62	73
MSSCIR55	4	5	0.56	84
MSSCIR60	3	4	0.28	136

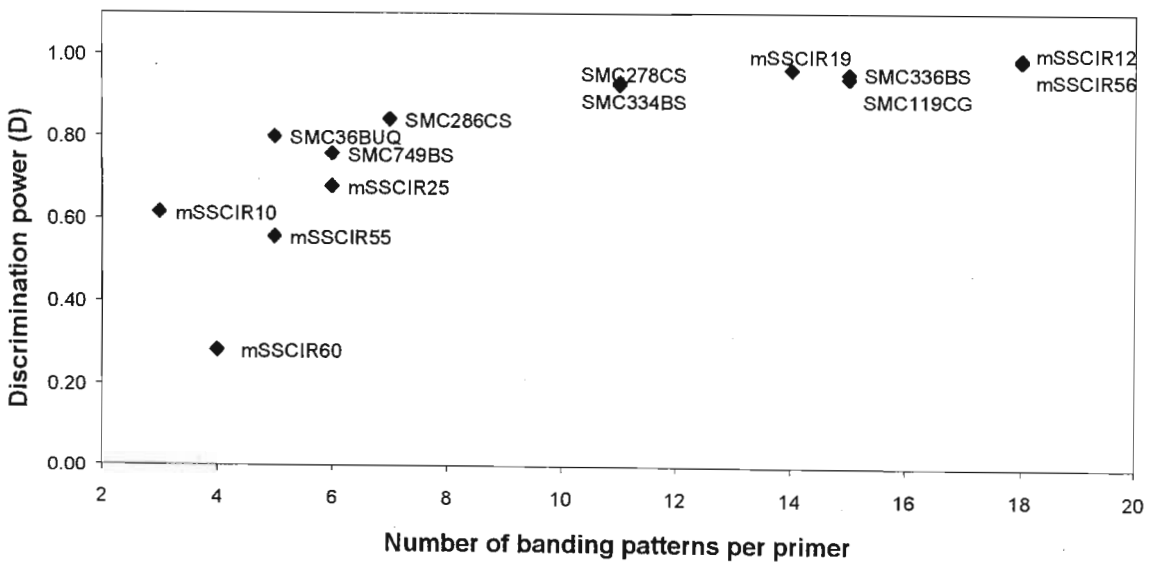

Figure 5.1 Value of the discriminating power (D) of the primers as a function of their number of banding patterns.

5.3.3 Evaluation of Combinations of Microsatellites

The discriminating power was determined for each microsatellite (Table 5.4), as reported previously, and the five most discriminating microsatellites were chosen to estimate the number of microsatellites necessary for large-scale sugarcane varietal identification. The theoretical number of non-differentiated pairs of cultivars (X_K) (see equation 5.2.10.4) was calculated for various combinations of these microsatellites (Table 5.5).

Table 5.5 Theoretical numbers of non-differentiated pairs of cultivars (X_K) for various microsatellite combinations, determined for 20, 100 and 1000 cultivars.

Primer combinations	No of markers	X_K Number of cultivars 20	100	1000
mSSCIR12	9	2.000	52.105	5257.895
mSSCIR12 + mSSCIR56	18	0.032	0.823	83.019
mSSCIR12 + mSSCIR56 + mSSCIR19	28	0.001	0.030	3.059
mSSCIR12 + mSSCIR56 + mSSCIR19 + SMC336BS	36	0.000	0.001	0.145
mSSCIR12 + mSSCIR56 + mSSCIR19 + SMC336BS + SMC119CG	46	0.000	0.000	0.008

5.3.4 Reproducibility of Microsatellite Markers

The application of a microsatellite approach, such as STMS, to sugarcane varietal identification relies on the fact that the method must be reproducible and robust. To investigate the potential impact environmental variation (different origins), type and condition of plant material used, and DNA extraction protocol

have on banding pattern reproducibility, a total of 16 DNA samples of cultivar NCo376 were prepared (Table 5.3). The NCo376 DNA samples were sourced from different locations and/or extracted from either leaf or leafroll of fresh or aged plant material. The various samples were amplified with six microsatellite primer pairs (SMC336BS, mSSCIR56, mSSCIR12, SMC119CG, SMC749BS and SMC278BS). The markers scored for NCo376 in the fingerprinting study were scored in this reproducibility study, and for each microsatellite primer pair no variation in the 47 scored markers was observed. However, variation in the extent of stutter and intensity of stutter bands was evident (Figure 5.2).

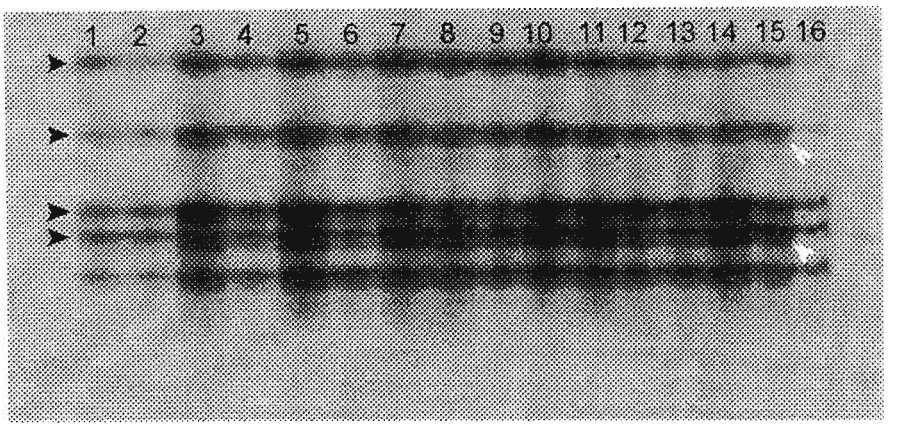

Figure 5.2 Autoradiograph showing the amplification of the various samples of cultivar NCo376 with microsatellite primer pair SMC336BS. Cultivar NCo376 amplification reactions were as follows: DNA of individual stalks from Field 43, SASEX (lane 1-5); pooled DNA from five stalks from Field 43, SASEX (lane 6), Kearsney Expt. Station (lane 7), Crookes Bros. Farm (lane 8), Gingindlouw Farm (lane 9) and Brynus Hill Farm (lane 10); pooled DNA of five stalks from Field 43, SASEX that had been cut and left at room temperature for 3 days (lane 11); DNA of young leaf (lane 12); old leaf (lane 13); leaf cut and left at room temperature for 4 days (lane 14); leaf cut and left at –80°C for 4 days (lane 15) and leaf infected with sugarcane mosaic virus (lane 16). Black arrows indicate scored markers, while white arrows indicate stutter bands.

5.4 DISCUSSION

The objectives of this study were to demonstrate the potential of STMS for sugarcane varietal identification, determine the discriminating power and

reproducibility of the microsatellite primer pairs used and finally suggest the number of microsatellites required for large scale fingerprinting.

Fourteen microsatellite primer pairs (Table 5.1) were used to fingerprint 20 cultivars (Table 5.2). A total of 86 polymorphic markers were scored and the number of polymorphic markers per primer pair varied from 3 to 10 (average of 6) and generated 3 – 18 different banding patterns per primer pair (Table 5.4). The number of markers detected per microsatellite primer pair across the 20 cultivars was approximately the same as found in a previous study of sugarcane (Cordeio et al., 2000), but lower then that found in other crops, such as rice, *Oryza sativa* (Garland et al., 1999); apple, *Malus pumila* (Hokanson et al., 1998); and emmer wheat, *Triticium dicoccoides* (Fahima et al., 2002).

It is significant to note that the use of P^{33} radiolabelling improved the marker visualisation, in some cases more markers were revealed per microsatellite, and reduced smearing in the gel lanes when compared to that of ethidium bromide and SYBR Gold staining.

As a measure of polymorphic information content, values for discriminating power (D) were calculated for each microsatellite primer pair. The average discriminating power for the 14 microsatellite primer pairs tested was 0.80, which was similar to the value in grape varieties of 0.79 (Tessier et al., 1999). The analysis of discrimination power revealed that microsatellite primer pair mSSCIR12 was the most efficient microsatellite (D = 0.99) for discriminating the 20 varieties. Interestingly, it was found that the efficiency of a given microsatellite primer pair does not depend only on the number of patterns it generates (Figure 5.1 and Table 5.4). For example, even if 2 microsatellite primer pairs produce the same numbers of unique patterns, such as SMC36BUQ and mSSCIR55, they can have very different discriminating powers, 0.80 and 0.56, respectively. Moreover, 2 primers with quite different numbers of patterns can have similar discrimination powers such as primer pairs SMC119CG and SMC278CS, with 15 and 11 patterns, respectively.

Microsatellite markers, in the STMS approach, have been used in various plant species, including sugarcane, to estimate genetic diversity (Jannoo et al., 2001), determine genetic relationships (Zhang et al., 2001) and to generate fingerprints for varietal identification (Piperidis et al., 2001). However no estimates have been made on the number of markers that are required for large scale sugarcane fingerprinting or, on the probability that two randomly chosen individuals will have different banding patterns. The level of desired statistical confidence is project specific and should take into account: (1) the severity of the consequences of incorrectly assigning a match to different individuals and (2) the proportion of closely related individuals that will probably be sampled.

The theoretical number of non-differentiated pairs of varieties (X_k) was determined as an estimator for the efficiency of various combinations of microsatellite primer pairs for varietal identification. These values are derived, in part, from the product of the individual microsatellite primer pairs D values, under the hypothesis of independence of the considered primers banding patterns, for a combination of primer pairs (Tessier et al., 1999).

It was determined for the 20 cultivars assessed, that if the marker data of microsatellites mSSCIR12 and mSSCIR56 were used in combination, that the risk of confusion or theoretical number of non-differentiated pairs of varieties (X_K) was only 0.03 (Table 5.5). Although, two microsatellite primer pairs were found to be sufficient to discriminate or uniquely identify all 20 cultivars, it was found that the total number of confusions or theoretical number of non-differentiated pairs of varieties (X_K) grows exponentially with the size of the sample or number of varieties (Tessier et al., 1999), such that 0.83 confusions are expected in a sample of 100 cultivars, and 83.02 confusions in a sample of 1000 cultivars (Table 5.5). Nevertheless, it was estimated that the use of the four most informative microsatellite primer pairs (mSSCIR12, mSSCIR56, mSSCIR19 and SMC336BS) in combination would successfully discriminate 1000 sugarcane varieties with only 0.14 non-differentiated pairs of varieties (Table 5.5).

However in circumstances where the microsatellite primer pairs, in combination, generate banding patterns which are not independent or which exhibit linkage disequilibrium, the product rule is violated and the estimation of the theoretical number of non-differentiated pairs of cultivars from the X_K values will be lower then the 'true' value. In the study by Tessier et al., (1999) 31 confusions were expected on the basis of the D values, yet 47 were observed. This underestimation is significant for sugarcane varietal identification, because of sugarcane's narrow genetic base and high instance of related individuals. Consequently, for large scale fingerprinting of sugarcane germplasm it is advisable to determine the risk of confusion or number of non-differentiated pairs of varieties (X_K), as a guideline to the number of markers that are required for statistical confidence.

However, before STMS can be accepted as a fingerprinting technique for large-scale sugarcane varietal identification, it is necessary that the reproducibility and robustness of the approach and scored markers be verified. To assess the reproducibility of the STMS approach, six microsatellite primer pairs were used to amplify 16 different samples of cultivar NCo376 (Table 5.3). These DNA samples were sourced from different locations and/or extracted from either the leaf or leafroll (two different extraction protocols) of fresh or aged plant material. It was found that the banding patterns for the various samples showed variation in extent and intensity of stutter bands, but these bands generally fell outside of the size range of the scored markers or were not scored as markers in the fingerprinting study.

The formation of 'stutter bands' is an innate quality of microsatellite marker systems, but the high ploidy level in sugarcane can result in an impressive number of 'stutter bands', which can overlap causing scoring to be difficult (Kaye et al., 1999). The origin of stutter bands is not completely understood although they are believed to be the result of polymerase slippage during amplification, leading to the formation of a series of non-specific PCR products,

which vary from the 'true' amplicons by a set number of nucleotides- two for dinucleotide repeats. The variation observed for the stutter bands might be due to slight differences in DNA quality or purity (A_{260} / A_{280} ratio) or slight differences in DNA quantity between the DNA samples. However no relationship for the extent of stuttering could be observed between the DNA extraction protocol, source or condition of plant material used for extraction or origin of the sugarcane. Therefore, considering the suitability and ease of extracting DNA from leaf material, as opposed to that of leafroll, it is advisable to use this approach for large-scale fingerprinting of sugarcane germplasm in the future.

In this study, sequence tagged microsatellite site (STMS) amplification was successfully applied to the identification of 20 sugarcane cultivars. Moreover these results confirmed the usefulness and suitability of the STMS approach for large-scale varietal identification and provided a guideline for the number of markers required for a particular statistical confidence.

CHAPTER 6

GENERAL DISCUSSION AND CONCLUSION

Microsatellites or simple sequence repeats have been shown to be powerful molecular markers, which are now poised to make a great contribution to sugarcane breeding and the sugar industry as a whole. The sequence tagged microsatellite site (STMS) approach is recognised as a robust and reproducible means of revealing the polymorphism or informativeness of microsatellites for various applications, such as parentage analysis or fingerprinting.

The suitability of the STMS approach and PCR typability of a set of 26 ISMC microsatellite primer pairs were investigated using a STMS protocol of Cordeiro et al., (2000). Of the 26 primer pairs evaluated, five (19%) failed to amplify and it was postulated that this was as a result of loss of, or mutation in, regions homologous to the primer sequence (i.e. null alleles). The STMS approach was refined for use on South African sugarcane germplasm by optimising the PCR parameters and testing various gel electrophoresis systems.

The parentage of the AA40 population was examined using this optimised STMS approach and a subset of the evaluated microsatellite primer pairs. The results revealed that none of the eight putative parent pairs, chosen to investigate possible mislabelling at seed collection or planting, were the true parents of the AA40 population. Furthermore, due to the significant number of non-parental markers scored across all the putative parent pairs, it was doubtful that microsatellite mutation could have led to misinterpreting the data and incorrectly discounting the true parents. Consequently, it was hypothesized that the incorrect parentage of the AA40 population did not occur due to mislabelling at planting or seed collection but rather as a result of mislabelling at the time of crossing. Therefore in the future the scope of the AA40 parentage analysis must

be widened and the number of cultivars assessed increased in an effort to identity the true parents.

Recently, Crouch et al., (1999a, 1999b) has suggested that the relatively high proportion of non-parental alleles segregating in both diploid and tetraploid *Musa* progeny could be as a consequence of heteroduplex formation (Crouch et al., 1998). Furthermore, it has been suggested that PCR recombination could be a frequent outcome of amplification from complex eukaryotic genomes, given their typically high levels of genic redundancy, and that this might be especially acute in polyploids (Cronn et al., 2002). As a result, it was hypothesised that a portion of the observed non-parental markers in the sugarcane varieties could be of a PCR artifactual nature and were not true parent-progeny marker data incompatibilities.

This hypothesis was tested by investigating marker variation between single parent and pooled parent pair (synthetic offspring) PCR reactions. STMS analyses of all eight putative parent pairs using six microsatellite primer pairs revealed that the mean of novel markers, not observed in the single template reactions, was 3.8%±2.9% with a maximum of 10%, although their origin was unknown. In future, based on these data, it is advisable that either putative parent pairs only be excluded from parentage analyses if the number of non-parental markers exceeds 10%, or that a marker is only considered to be discriminating for parentage if it is absent from both parents and the pooled template PCR reaction or 'synthetic offspring'. Although parentage analysis in sugarcane needs to be refined, it could make a great contribution to sugarcane breeding programmes, where it is recognised that illegitimate pollination and incorrect parentage are not uncommon.

The application of the STMS approach to fingerprinting is of considerable value to the sugar industry because it offers a rapid and reliable means to uniquely identify sugarcane varieties, develop a fingerprinting database and manage germplasm collections. The potential of the STMS approach for sugarcane

varietal identification was investigated by fingerprinting a set of 20 cultivars with 14 microsatellite primer pairs. The efficiency of each microsatellite primer pair for fingerprinting was evaluated by determining their discriminating power (D) and the average was found to be 0.80.

Although there have been numerous reports of the use of STMS in sugarcane, none have focused on the way to optimally apply this approach to large-scale sugarcane varietal identification. In particular, the greatest challenges are to reduce the cost of analysis (i.e. the number of amplifications and thus the number of primers) as well as the likelihood of two randomly chosen varieties being indistinguishable. With this in mind, the theoretical number of non-differentiated pairs of cultivars (X_K) was determined for various combinations of the five most discriminating microsatellites.

From the X_K values, it was found that a combination of the two most discriminating microsatellite primer pairs (mSSCIR12, mSSCIR56) would be sufficient to uniquely identify up to 100 cultivars, provided that the 20 fingerprinted varieties were representative of the larger sample of 100 and that the microsatellites were independent. It is significant to note that the X_K values have been found to underestimate the risk of confusion and as such they may be regarded as a guideline for the minimum number of microsatellites required for fingerprinting. The sensitivity of this STMS approach to DNA of different origin, quality, source material and extraction protocol was assessed by conducting a reproducibility investigation on various DNA samples of cultivar NCo376. Variation in the extent and intensity of stutter bands was observed, although the amplification of scored markers was robust and reproducible with no observable relationship between the extent of stutter and some DNA characteristic, such as extraction protocol.

In summary, it was concluded that although the STMS approach was unsuccessful at resolving the AA40 parents, it shows great promise for future parentage investigations. Furthermore, evidence in this study also suggests that

the STMS approach is a viable technology for large-scale sugarcane varietal identification, being both robust and reproducible, although determining the number of microsatellites required is still an obstacle.

REFERENCES

Ainsworth, C.C. and Sharp, P.J. (1989). The potential role of DNA in plant variety identification. *Plant Varieties and Seeds* **2**: 27-34.

Akkaya, M.S., Bhagwat, A.A. and Cregan, P.B. (1992). Length polymorphisms of simple sequence repeat DNA in soybean. *Genetics* **132**: 1131-1139.

Anonymous (2001). *2001/2002 South African Sugar Industry Directory*. SASA External Affairs Division, South Africa.

Arumuganathan, K. and Earle, E.D. (1991). Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter* **9**: 208-219.

Ashkenazi, V., Chani, E., Lavi, U., Levy, D., Hillel, J. and Veilleux, R.E. (2001). Development of microsatellite markers in potato and their use in phylogenetic and fingerprinting analyses. *Genome* **44**: 50-62.

Barnes, J.M. and Bester, A.E. (2000). Genetic mapping in sugarcane: prospects and progress in the South African sugar industry. *Proceedings of the South African Sugar Technologists Association* **74**: 177-179

Barnes, J.M., Rutherford, R.S. and Botha, F.C. (1997). The identification of potential genetic markers in sugarcane varieties for the prediction of disease and pest resistance ratings. *Proceedings of the South African Sugarcane Technologists Association* **71**: 57-61.

Barnett, F.L. and Carver, R.F. (1967). Meiosis and pollen stainability in switchgrass, *Panicum virgatum* L. *Crop Science* **7**: 301-304.

Becker, J. and Heun, M. (1995). Barley microsatellites: allele variation and mapping. *Plant Molecular Biology* **27**: 835-845.

Becker, J., Vos, P., Kuiper, M., Salamini, F. and Heun, M. (1995). Combined mapping of AFLP and RFLP markers in barley. *Molecular and General Genetics* **249**: 65-73.

Beckmann, J.S. and Soller, M. (1986). Restriction fragment length polymorphisms and genetic improvement of agricultural species. *Euphytica* **35**: 111-124.

Bell, C.J. and Ecker, J.R. (1994). Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* **19**: 137-144.

Berding, N. and Roach, B.T. (1987). Germplasm collection, maintenance, and use. In: Heinz, D.J. (ed.), *Sugarcane improvement through Breeding*. Elsevier, Amsterdam, pp. 143-210.

Besse, P., McIntyre, C.L. and Berding, N. (1997). Characterisation of *Erianthus* sect. Ripidium and *Saccharum* germplasm (Andropogpnease - Saccharinae) using RFLP markers. *Euphytica* **93**: 283-292.

Besse, P., Taylor, G., Carroll, B., Berding N., Burner, D. and McIntyre, C.L. (1998). Assessing genetic diversity in a sugarcane germplasm collection using automated AFLP analysis. *Genetica* **104**: 143-153.

Bhat, S.R. and Gill, S.S. (1985). The implication of 2n egg gametes in nobilisation and breeding of sugarcane. *Euphytica* **34**: 377-384.

Botstein, D., White, R.L., Skolnick, W.M. and Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**: 314-331.

Bradshaw, J.E., Hackett, C.A., Meyer, R.C., Milbourne, D., McNicol, J.W., Philips, M.S. and Waugh, R. (1998). Identification of AFLP and SSR markers associated with quantitative resistance to *Globodera pallida* (Stone) in tetraploid potato (*Solanum tuberosum* subsp. *tuberosum*) with a view to marker-assisted selection. *Theoretical and Applied Genetics* **97**: 202-210.

Brandes, E.W. (1958). Origin, classification and characteristics. In: Artschwager, E. and Brandes, E.W. (eds.), *Sugarcane (Saccharum officinarum L.)*. USDA Agriculture Handbook no. 122, pp. 1-35.

Bredemeijer, G.M.M., Arens, P., Wouters, D., Visser, D. and Vosman, B. (1998). The use of semi-automated fluorescent microsatellite analysis for tomato cultivar identification. *Theoretical and Applied Genetics* **97**: 584-590.

Bremer, G. (1930). The cytology of *Saccharum*. *Proceedings of the International Society of Sugarcane Technologists* **3**: 408-415.

Bremer, G. (1961). Problems in breeding and cytology of sugarcane. *Euphytica* **10**: 59-78.

Broun, P. and Tanksley, S.D. (1996). Characterization and genetic mapping of simple repeat sequences in the tomato genome. *Molecular and General Genetics* **250**: 39-49.

Brown, S.M., Hopkins, M.S., Mitchell, S.E., Senior, M.L., Wang, T.Y., Duncan, R.R., Gonzalez-Candelas, F. and Kresovich, S. (1996). Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. *Theoretical and Applied Genetics* **93**: 190-198.

Bruland, O., Almqvist, E.W., Goldberg, Y.P., Boman, H., Hayden, M.R. and Knappskog, P.M. (1999). Accurate determination of the number of CAG repeats in the Huntington disease gene using a sequence-specific internal DNA standard. *Clinical Genetics* **55**: 198-202.

Burner, D.M. (1991). Cytogenetic analyses of sugarcane relatives (Andropogoneae: Saccharinae). *Euphytica* **54**: 125-133.

Buteler, M.I., LaBonte, D.R., Jarret, R.L. and Macchiavelli, R.E. (2002). Microsatellite-based paternity analysis in polyploid sweetpotato. *Journal of the American Society of Horticultural Science* **127**: 392-396.

Buteler, M.I., Labonte, D.R. and Macchiavelli, R.E. (1997). Determining paternity in polyploids: hexaploid simulation studies. *Euphytica* 96: 353-361.

Butterfield, M.K. (2002). Personnel communication. Department of Plant Breeding, SASEX, South Africa.

Butterfield, M.K., D'Hont, A. and Berding, N. (2001). The sugarcane genome: a synthesis of current understanding, and lessons for breeding and biotechnology. *Proceedings of the South African Sugarcane Technologists Association* **75**: 1-5.

Butterfield, M.K. and Thomas, D.W. (1996). Sucrose, yield and disease resistance characteristics of sugarcane varieties under test in the SASEX selection programme. *Proceedings of the South African Sugarcane Technologists Association* **70**: 103-105.

Callen, D.F., Thompson, A.D., Shen, Y., Philips, H.A., Richards, R.I., Mulley, J.C. and Sutherland, G.R. (1993). Incidence and origin of 'null' alleles in the $(AC)_n$ microsatellite markers. *American Journal of Human Genetics* **52**: 922-927.

Carlson, J.E., Tulsieram, L.K., Glaubitz, J.C., Luk, V.W.K., Kauffeldt, C. and Rutledge, R. (1991). Segregation of random amplified DNA markers in F1 progeny of conifers. *Theoretical and Applied Genetics* **83**: 194-200.

Celarier, R.P. (1956). Cytotaxonomy of the Andropogoneae. I. Subtribes Dimarinae and Saccharinae. *Cytologia* **21**: 272-291.

Chakraborty, R., Meagher, T.R. and Smouse, P.E. (1988). Parentage analysis with genetic markers in natural populations. I. The expected proportion of offspring with unambiguous paternity. *Genetics* **118**: 527-536.

Charlesworth, B., Morgan, M.T. and Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289-1303.

Cho, Y.G., McCouch, S.R., Kuiper, M., Kang, M.R., Pot, J., Groenen, J.T.M. and Eun, M.Y. (1998). Integrated map of AFLP, SSLP and RFLP markers using a recombinant inbred population of rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* **97**: 370-380.

Chou, Q., Russell, M., Birch, D.E., Raymond, J. and Bloch, W. (1992). Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Nucleic Acids Research* **20**: 1717-1723.

Coetzee, N. (2002). Personnel communication. Department of Plant Breeding, SASEX, South Africa.

Comai, L. 2000. Genetic and epigenetic interactions in allopolyploid plants. *Plant Molecular Biology* **43**: 387-399.

Cooke, R.J. (1984). The characterisation and identification of crop cultivars by electrophoresis. *Electrophoresis* **5**: 59-72.

Cordeiro, G.M. (2001). Molecular marker systems for sugarcane germplasm analysis. In: Henry, R.J. (ed.), *Plant genotyping: the DNA fingerprinting of plants*. CABI International, Wallingford, UK, pp. 129-146.

Cordeiro, G.M., Casu, R., McIntyre, C.L., Manners, J.M. and Henry, R.J. (2001). Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross-transferable to *Erianthus* and *Sorghum*. *Plant Science* **160**: 1115-1123.

Cordeiro, G.M., Maguire, T.L., Edwards, K.J. and Henry, R.J. (1999). Optimisation of a microsatellite enrichment technique in *Saccharum* spp. *Plant Molecular Biology Reporter* **17**: 225-229.

Cordeiro, G.M., Taylor, G.O. and Henry, R.J. (2000). Characterisation of microsatellite markers from sugarcane (*Saccharum* sp.), a highly polyploid species. *Plant Science* **155**: 161-168.

Cregan, P.B., Jarvik, T., Bush, A.L., Shoemaker, R.C., Lark, K.G., Kabler, A.I., Kaya, N., VanToai, T.T., Lohnes, D.G., Chung, J. and Specht, J. (1999). An integrated genetic linkage map of the soybean genome. *Crop Science* **39**: 1464-1490.

Cronn, R., Cedroni, M., Haselkorn, T., Grover, C. and Wendel, J.F. (2002). PCR-mediated recombination in amplification products derived from polyploid cotton. *Theoretical and Applied Genetics* **104**: 482-489.

Crouch, J.H., Crouch, H.K., Constandt, H., Van Gysel, A., Breyne, P., Van Montagu, M., Jarret, R.L. and Ortiz, R. (1999a). Comparison of PCR-based molecular marker analyses of *Musa* breeding populations. *Molecular Breeding* **5**: 233-244.

Crouch, H.K., Crouch, J.H., Jarret, R.L., Cregan, P.B. and Ortiz, R. (1998). Segregation of microsatellite loci from haploid and diploid gametes in *Musa*. *Crop Science* **38**: 211-217.

Crouch, J.H., Crouch, H.K., Tenkouano, A. and Ortiz, R. (1999b). VNTR-based diversity analysis of 2x and 4x full-sib *Musa* hybrids. *EJB Electronic Journal of Biotechnology* [online], **2**, 3.
Available from: www.ejb.org/content/vol2/issue3/full/1/index.htm

D'Hont, A. and Paulet, F. (2000). The origin of "North Indian and Chinese sugarcanes" investigated by genomic in situ hybridization and RFLP markers. 6th International Society of Sugarcane Technologists (ISSCT) Breeding Workshop.
Available from http://issct.intnet.mu/bridabs.htm#2

D'Hont, A. and Grivet, L. (1996). Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Molecular and General Genetics* **250**: 405-413.

D'Hont, A., Ison, D., Alix, K., Roux, C. and Glaszmann, J.C. (1998). Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. *Genome* **41**: 221-225.

D'Hont, A., Lu, Y.H., De leon, D.G., Grivet, L., Feldmann, P., Lanuad, C. and Glaszmann, J.C. (1994). A molecular approach to unravelling the genetics of sugarcane, a complex polyploid of the Andropogoneae tribe. *Genome* **37**: 222-230.

Da Silva, J. (2001). Preliminary analysis of microsatellite markers derived from sugarcane expressed sequence tags (ESTs). *Genetics and Molecular Biology* **24**: 155-159.

Da Silva, J., Burnquist, W.L. and Tanksley, S.D. (1993). RFLP linkage map and genome analysis of *Saccharum spontaneum*. *Genome* **36**: 782-791.

Daniels, J. and Daniels, C.A. (1975). Geographical, historical and cultural aspects of the origin of the Indian and Chinese sugarcanes *S. barberi* and *S. sinense*. *Sugarcane Breeding Newsletter* **36**: 4-23.

Daniels, J. and Roach, B.T. (1987). Taxonomy and evolution. In: Heinz, D.J. (ed.), *Sugarcane improvement through breeding.* Elsevier Press, Amsterdam, pp. 7-84.

Dawson, K., Chalmers, K.J., Waugh, R. and Powell, W. (1993). Detection and analysis of genetic variation in *Hordeum spontaneum* populations from Israel using RAPD markers. *Molecular Ecology* **2**: 151-159.

Deberer, T. (1990). Phylogeny of wild and cultivated *Solanum* species based on nuclear restriction fragment length polymorphisms (RFLPs). *Theoretical and Applied Genetics* **79**: 360-368.

Dellaporta, S.L., Wood, J. and Hicks, J.B. (1983). A plant DNA mini-preparation. Version II. *Plant Molecular Biology Reporter* **1**: 19-21.

Delmotte, F., Leterme, N. and Simon, J.C. (2001). Microsatellite allele sizing: difference between automated capillary electrophoresis and manual technique. *Biotechniques* **31**: 810-818.

Deren, C.W. (1995). Genetic base of U.S. mainland sugar cane. *Crop Science* **35**: 1195-1199.

Dib, C., Faure S., Fizames C., Samson D., Drouot N., Vignal A., Millasseau P., Marc S., Hazan J., Seboun E., Lathrop M., Gyapay G., Morissette J., and Weissenbach J. (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**: 152-154.

Dietrich, W.F., Miller, J., Steen, R., Mechant, M.A., Damron-Boles, D. and Husain, Z. (1996). A comprehensive genetic map of the mouse genome. *Nature* **380**: 152-154.

Diwan, N. and Cregan, P.B. (1997). Automated sizing of fluorescent-labelled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theoretical and Applied Genetics* **95**: 723-733.

Donini, P., Stephenson, P., Bryan, G.J. and Koebner, R.M.D. (1998). The potential of microsatellites for high throughput genetic diversity assessment in wheat and barley. *Genetic Resources and Crop Evolution* **45**: 415-421.

Edwards, A., Hammond, H.A., Jin, L., Caskey, C.T. and Chakraborty, R. (1992). Genetic variation in five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* **12**: 241-253.

Edwards, K.J., Barker, J.H.A., Daly, A., Jones, C. and Karp, A. (1996). Microsatellite libraries enriched for several microsatellite sequences in plants. *Biotechniques* **20**: 759-760.

Eksomtramage, T., Paulet, F., Noyer, J.L., Feldmann, P. and Glaszmann, J.C. (1992). Utility of isozymes in sugar cane breeding. *Sugar Cane* **3**: 14-21.

El Hassan, M.A. and Calladine, C.R. (1996a). Propeller-twisting of base-pairs and the flexibility of dinucleotide steps. *Journal of Molecular Biology* **259**: 95-103.

El Hassan, M.A. and Calladine, C.R. (1996b). Structural mechanics of bent DNA. *Endeavour* **20**: 61-68.

Erlich, H.A. (1989). Basic methodology. In: Erlich, H.A. (ed.), *PCR technology: principles and applications for DNA amplification.* Stockton Press, New York, pp. 1-6.

Fahima, T., Roder, M.S., Grama, A. and Nevo, E. (1998). Microsatellite DNA polymorphism divergence in *Triticum dicoccoides* accessions highly resistant to yellow rust. *Theoretical and Applied Genetics* **96**: 187-195.

Fahima, T., Roder, M.S., Wendehake, K., Kirzhner, V.M. and Nevo, E. (2002). Microsatellite polymorphism in natural populations of wild emmer wheat, *Triticium dicoccoides*, in Israel. *Theoretical and Applied Genetics* **104**: 17-29.

Freudenreich, C.H., Stavenhagen, J.B. and Zakian, V.A. (1997). Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Molecular and Cellular Biology* **17**: 2090-2098.

Gallacher, D.J. and Berding, N. (1995). Use of isozyme phenotypes for rapid discrimination among sugarcane clones. *Australian Journal of Agricultural Research* **46**: 601-609.

Garland, S.H., Lewin, L., Abedinia, M., Henry, R. and Blakeney, A. (1999). The use of microsatellite polymorphisms for the identification of Australian breeding lines of rice (*Oryza sativa* L.). *Euphytica* **108**: 53-63.

Gianfranceschi, L., Seglias, N., Tarchini, R., Komjanc, M. and Gessler, C. (1998). Simple sequence repeats for the genetic analysis of apple. *Theoretical and Applied Genetics* **96**: 1069-1076.

Gillet, E.M. (1999). Purposes: classification and desired marker characteristics. In: Gillet, E.M. (ed.), *Which DNA Marker for which purpose?* Institut für Forstgenetik und Forstpflanzenzüchtung, Universität Göttingen, pp. 1-5.

Glaszmann, J.C., Fautret, A., Noyer, J.L., Feldmann, P. and Lanaud, C. (1989). Biochemical genetic markers in sugarcane. *Theoretical and Applied Genetics* **78**: 537-543.

Goldstein, D.B. and Clark, A.E. (1995). Microsatellite variation in North American populations of *Drosophila melanogaster*. *Nucleic Acids Research.* **23**: 3882-3886.

Godwin, I.D., Mace, E.S. and Nurzuhairawaty (2001). Genotyping Pacific island Taro (*Colocasia Esculenta* (L.) Schott) germplasm. In: Henry, R.J. (ed.), *Plant genotyping: the DNA fingerprinting of plants*. CABI International , UK, pp. 109-128.

Graner, A., Jahoor, A., Schondelmaier, J., Siedler, H., Pillen, K., Fischbeck, G., Wenzel, G. and Herrmann, R.G. (1991). Construction of an RFLP map of barley. *Theoretical and Applied Genetics* **83**: 250-256.

Graner, A., Siedler, H., Jahoor, A., Herrmann, R.G. and Wenzel, G. (1990). Assessment of the degree and type of restriction fragment length polymorphism in barley (*Hordeum vulgare*). *Theoretical and Applied Genetics* **80**: 826-832.

Grivet, L., D'Hont, A., Roques, D., Feldmann, P., Lanaud, C. and Glaszmann, J.C. (1996). RFLP mapping in cultivated sugarcane (*Saccharum* spp.): genome organisation in a highly polyploid and aneuploid interspecific hybrid. *Genetics* **142**: 987-1000.

Guilford, P., Prakesh, S., Zhu, J.M., Rikkerink, E., Gardiner, S., Bassett, H. and Forster, R. (1997). Microsatellites in *Malus* x *Domestica* (Apple): abundance, polymorphism and cultivar identification. *Theoretical and Applied Genetics* **94**: 249-254.

Gupta, M., Chyi, Y.S., Romero-Severson, J. and Owen, L. (1994). Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats. *Theoretical and Applied Genetics* **89**: 998-1006.

Gupta, P.K., Balyan, H.S., Sharma, P.C. and Ramesh, B. (1996). Microsatellites in plants: a new class of molecular markers. *Current Science* **70**: 45-54.

Gupta, P.K. and Varshney, R.K. (2000). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* **113**: 163-185.

Gupta, P.K., Varshney, R.K. and Prasad, M. (2002). Molecular Markers: principles and methodology. In: Jain, S.M., Brar, D.S. and Ahloowalia, B.S. (eds.), *Molecular techniques for crop improvement.* Kluwer Academic Publishers, Boston, pp. 9-54.

Gupta, P.K., Varshney, R.K., Sharma, P.C. and Ramash, B. (1999). Molecular markers and their application in wheat breeding. *Plant Breeding* **118**: 369-390.

Ha, S., Moore, P.H., Heinz, D., Kato, S., Ohmido, N. and Fukui, K. (1999). Quantitative chromosome map of the polyploid *Saccharum spontaneum* by multicolour fluorescence in situ hybridisation and imaging methods. *Plant Molecular Biology* **39**: 1165-1173.

Hadrys, H., Balick, M. and Schierwater, B. (1992). Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Molecular Ecology* **1**: 55-63.

Hakki, E.E. and Akkaya, M.S. (2000). Microsatellite isolation using amplified fragment length polymorphism markers: no cloning, no screening. *Molecular Ecology* **9**: 2152-2154.

Hamada, H., Seidman, M., Howard, B.H. and Gorman, C.M. (1984). Enhanced gene expression by the poly(dT-dG).poly(dC-dA) sequence. *Molecular and Cellular Biology* **4**: 2622-2630.

Hancock, J.M. (1999). Microsatellite and simple sequences: genomic context and mutational mechanisms. In: Goldstein, D.B., Schlotterer, C. (eds.), *Microsatellites: evolution and applications*. Oxford University Press, Oxford, pp. 1-10.

Hanson, R.E., Zhao, X.-P., Islam-Faridi, M.N., Paterson, A.H., Zwick, M.S., Crane, C.F., McKnight, T.D., Stelly, D.M., and Price, H.J. (1998). Evolution of interspersed repetitive elements in *Gossypium* (Malvaceae). *American Journal of Botany* **85**: 1364-1369.

Hanson, R.E., Islam-Faridi, M.N., Crane, C.F., Zwick, M.S., Czeschin, D.G., Wendel, J.F., McKnight, T.D., Price, H.J., and Stelly, D.M.(1999). Ty1-*copia*-retrotransposon behaviour in a polyploid cotton. *Chromosome Research* **8**: 73-76.

Hartl, L., Mohler, V., Zeller, F.J., Hsam, S.L.K. and Schweizer, G. (1999). Identification of AFLP markers closely linked to the powdery mildew resistance genes *Pm1c* and *Pm4a* in common wheat (*Triticum aestivum* L.). *Genome* **42**: 322-329.

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

Harvey, M. and Huckett, B.I. (1998). Preliminary genetic analysis of the sugarcane population AA40. *Proceedings of the South African Sugarcane Technologists Association* **72**: 316-317.

Hauge, X.Y. and Litt, M. (1993). A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by PCR. *Human Molecular Genetics* **2**: 411-415.

Henderson, S.T. and Petes, T.D. (1992). Instability of simple sequence DNA in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **12**: 2749-2757.

Henegariu, O., Heerema, N.A., Dlouhy, S.R., Vance, G.H. and Vogt, P.H. (1997). Multiplex PCR: Critical parameters and step-by-step protocol. *Biotechniques* **23**: 504-511.

Hoarau, J.Y., Offmann, B., D'Hont, A., Risterucci, A.M., Roques, D., Glaszmann, J.C. and Grivet, I. (2001). Genetic dissection of modern sugarcane cultivar (*Saccharum* spp.). Part 1. Genome mapping with AFLP markers. *Theoretical and Applied Genetics* **103**: 84-97.

Hogarth, D.M. (1987). Genetics of sugarcane. In: Heinz, D.J. (ed.), *Sugarcane improvement through breeding*. Elsevier Press, Amsterdam, pp. 255-271.

Hokanson, S.C., Szewc-McFadden, A.K., Lamboy, W.F. and McFerson, J.R. (1998). Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus* X *domestica* borkh. core subset collection. *Theoretical and Applied Genetics* **97**: 671-683.

Holton, T.A. (2001). Plant genotyping by analysis of microsatellites. In: Henry, R.J. (ed.), *Plant genotyping: the DNA fingerprinting of plants*. CABI Publishing, Wallingford, UK. pp. 15-28.

Huckett, B.I. and Botha, F.C. (1995). Stability and potential use of RAPD markers in a sugarcane genealogy. *Euphytica* **86**: 117-125.

Jackson, P. (1994). Genetic relationships between attributes in sugarcane clones closely related to *Saccharum spontaneum*. *Euphytica* **79**: 101-108.

Jannoo, N., Forget, L. and Dookun, A. (2001). Contribution of microsatellites to the sugar cane breeding program in Mauritius. *Proceedings of the International Society of Sugarcane Technologists* **24**, 637-639.

Jannoo, N., Grivet, L., Seguin, M., Paulet, F., Domaingue, P.S.R., Rao, A., Dookun, A., D'Hont, A. and Glaszmann, J.C. (1999). Molecular investigation of the genetic base of sugarcane cultivars. *Theoretical and Applied Genetics* **99**: 171-184.

Janssen, P., Coopman, R., Huys, G., Swings, J., Bleeker, M., Vos, P., Zabeau, M. and Kersters, K. (1996). Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. *Microbiology* **142**: 1881-1893.

Jarne, P. and Lagoda, P.J.L. (1996). Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* **11**: 424-429.

Jeffreys, A.J. and Pena, S.D.J. (1993). Brief Introduction to human DNA fingerprinting. In: Pena, S.D.J., Chakraborty, R., Epplen, J.T. and Jeffreys, A.J. (eds.), *DNA fingerprinting: state of the science.* Birkhauser Verlag, Switzerland, pp. 1-20.

Jeffreys, A.J., Tamaki, K., MacLeod, A., Monckton, D.G., Neil, D.L. and Armour, J.A.L. (1994). Complex gene conversion events in germline mutation at human minisatellites. *Nature Genetics.* **6**: 136-145.

Jones, C.J., Edwards, K.J., Castaglione, S., Winfield, M.O., Sala, F., van de Wiel, C., Bredemeijer, G.M.M., Vosman, B., Matthes, M., Daly, A., Brettschneider, R., Bettini, P., Buiatti, M., Maestri, E., Malcevschi, A., Marmiroli, N., Aert, R., Volckaert, G., Rueda, J., Linacero, R., Vazquez, A. and Karp, A. (1997). Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding* **3**: 381-390.

Joshi, S.P., Ranjekar, P.K. and Gupta, V.S. (1999). Molecular markers in plant genome analysis. *Current Science* **77**: 230-240.

Kaye, C., Gay, C., Rodies-Goud, M. and D'Hont, A. (1999). The development of SSRs in sugarcane at CIRAD, France. *The Satellite* **2**: 2-3.

Klein-Lankhorst, R.M., Vermunt, A., Weide, R., Liharska, T. and Zabel, P. (1991). Isolation of molecular markers for tomato (*L. esculentum*) using random amplified polymorphic DNA (RAPD). *Theoretical and Applied Genetics* **83**: 108-114.

Kolliker, R., Jones, E.S., Drayton, M.C., Dupal, M.P. and Forster, J.W. (2001). Development and characterisation of simple sequence repeat (SSR) markers for white clover (*Trifolium repens* L.). *Theoretical and Applied Genetics* **102**: 416-424.

Koreth, J., O'Leary, J.J. and McGee, J. (1996). Microsatellites and PCR genomic analysis. *Journal of Pathology* **178**: 239-248.

Korzan, V., Roder, M.S., Ganal, M.W., Worland, A.J. and Law, C.N. (1998). Genetic analysis of the dwarfing gene *Rht8* in wheat. Part 1. Molecular mapping of *Rht8* on the short arm of chromosome 2D of bread wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics* **96**: 1104-1109.

Lagercrantz, U., Ellegren, H. and Andersson, L. (1993). The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Res.* **21**: 1111-1115.

Lagoda, P.J.L., Dambier, D., Graphin, A., Bauens, F.C., Lanaud, C. and Noyer, J.L. (1998). Non-radioactive sequence-tagged microsatellite site analysis: a method transferable to the tropics. *Electrophoresis* **19**: 152-157.

Law, C.N. (1995). Genetic manipulation in plant breeding- prospects and limitations. *Euphytica* **85**: 1-12.

Lebot, V. and Aradhya, K.M. (1992). Isozyme variation in Taro (*Colocasia esculenta* L.) in Asia and Oceania. *Euphytica* **56**: 55-66.

Lee, L.S. and Henry, R.J. (2001). Commercial applications of plant genotyping. In: Henry, R.J. (ed.), *Plant genotyping: the DNA fingerprinting of plants*. CABI Publishers, Wallington, UK, pp. 265-273.

Levinson, G. and Gutman, G.A. (1987a). High frequency of short frameshifts in poly-CA/GT tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Research* **15**: 5323-5338.

Levinson, G. and Gutman, G.A. (1987b). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology Evolution*. **4**: 203-221.

Levitt, R.C., Kiser, M.B., Dragwa, C., Jedlicka, A.E., Xu, J., Meyers, D.A. and Hudson, J.R. (1994). Fluorescence-based resource for semi-automated genomic analyses using microsatellite markers. *Genomics* **24**: 361-365.

Li, H.W. and Price, S. (1967). Chromosome numbers of noble sugarcane clones. *Proceedings of the International Society of Sugar Cane Technologists*. **12**: 884-886.

Lima, M.L.A., Garcia, A.A.F., Oliveira, K.M., Matsuoka, S., Arizono, H., de Souza C.L. and de Souza, A.P. (2002). Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.). *Theoretical and Applied Genetics* **104**: 30-38.

Liu, B., Brubaker, C.L., Mergeai, G., Cronn, R.C., and Wendel, J.F. 2001. Polyploid formation in cotton is not accompanied by rapid genomic changes. *Genome* **44**: 321-330.

Liu, B., Vega, J,M., and Feldman, M. (1998a). Rapid genomic changes in newly synthesised ampliploids of *Triticum* and *Aegilops*. II. Changes in low-copy coding DNA sequences. *Genome* **41**: 535-542.

Liu, B., Vega, J,M., Segal, G., Abbo, S., Rodova, M., and Feldman, M. (1998b). Rapid genomic changes in newly synthesised ampliploids of *Triticum* and *Aegilops*. I. Changes in low-copy coding DNA sequences. *Genome* **41**: 272-277.

Liu, B., and Wendel, J.F. 2000. Retroelement activation followed by rapid repression in interspecific hybrid plants. *Genome* **43**: 874-880.

Liu, B., and Wendel, J.F. 2002. Non-mendelian phenomena in Allopolyploid genome evolution. *Current Genomics* **3**: 1-17.

Liu, Z.W., Biyashev, R.M. and Saghai Maroof, M.A. (1996). Development of simple sequence repeat DNA markers and their integration into a barley linkage map. *Theoretical and Applied Genetics* **93**: 869-876.

Liu, Z.W., Jarret, R.L., Kresovich, S. and Duncan, R.R. (1995). Characterization and analysis of simple sequence repeat (SSR) loci in seashore paspalum (*Paspalum vaginatum* Swartz.). *Theoretical and Applied Genetics* **91**: 47-52.

Lu, Y.H., D'Hont, A., Paulet, F., Grivet, L., Arnaud, M. and Glaszmann, J.C. (1994a). Molecular diversity and genome structure in modern sugarcane varieties. *Euphytica* **78**: 217-226.

Lu, Y.H., D'Hont, A., Walker, D.I.T., Rao, P.S., Feldmann, P. and Glaszmann, J.C. (1994b). Relationship among ancestral species of sugarcane revealed with RFLP using single copy maize nuclear probes. *Euphytica* **78**: 7-18.

Maguire, T.L. (2001). Producing and exploiting enriched microsatellite libraries. In: Henry, R.J. (ed.), *Plant genotyping: the DNA fingerprinting of plants*. CABI Publishing, Wallington, UK. pp. 193-210.

Maughan, P.J., Saghai Maroof, M.A. and Buss, G.R. (1995). Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. *Genome* **38**: 715-723.

Maurer, D.J., O'Callaghan, B.L. and Livingston, D.M.1. (1996). Mapping the polarity of changes that occur in interrupted CAG repeat tracts in yeast. *Molecular and Cellular Biology* **18**: 4597-4604.

McCouch, S.R., Chen, X., Panaud, O., Temnykh, S., Xu, Y., Gu Cho, Y., Huang, N., Ishii, T. and Blair, M. (1997). Microsatellite marker development, mapping and applications in rice genetics and breeding. *Plant Molecular Biology* **35** : 89-99.

McGregor, C.E., Greyling, M.M. and Warnich, L. (2000). The use of simple sequence repeats (SSRs) to identify commercially important potato (*Solanum tuberosum* L.) cultivars in South Africa. *Suid Afrikaanse Tydskrif van Plant Grond* **17**: 177-179.

Mitchell, S.E., Kresovich, S., Jester, C.A., Hernandez, C.J. and Szewc-McFadden, A.K. (1997). Application of multiplex PCR and fluorescence-based, semi-automated allele sizing technology for genotyping plant genetic resources. *Crop Science* **37**: 617-624.

Morell, M.K., Peakall, R., Appels, R., Preston, L.R. and Lloyd, H.L. (1995). DNA profiling techniques for plant variety identification. *Australian Journal of Experimental Agriculture* **35**: 807-819.

Morgante, M. and Olivieri, A.M. (1993). PCR-amplified microsatellites as markers in plant genetics. *Plant Journal* **3**: 175-182.

Morin, P.A., Wallis, J., Moore, J.J. and Woodruff, D.S. (1994). Paternity exclusion in a community of wild chimpanzees using hypervariable simple sequence repeats. *Molecular Ecology* **5**: 469-478.

Mueller, U.G. and Wolfenberger, L.L. (1999). AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution* **14**: 389-394.

Naik, G.R. (2001). *Sugarcane Biotechnology*. Science Publishers, Enfield (NH), USA.

Nair, M.K. (1975). Cytogenetics of *Saccharum officinarum* L. and *S. spontaneum* L. IV. Chromosome number and meiosis in *Saccharum officinarum* X *S. spontaneum* hybrids. *Caryologia* **28**: 1-14.

Nair, N.V., Nair, S., Sreenivasaan, T.V. and Mohan, M. (1999). Analysis of genetic diversity and phylogeny in *Saccharum* and related genera using RAPD markers. *Genetic Resources and Crop Evolution* **46**: 73-79.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Science USA* **70**: 3321-3323.

Paetku, D. and Strobeck, C. (1995). The molecular basis and evolutionary history of microsatellite null alleles in bears. *Molecular Ecology* **4**: 519-520.

Paglia, G. and Morgante, M. (1998). PCR-based multiplex DNA fingerprinting techniques for the analysis of conifer genomes. *Molecular Breeding* **4**: 173-177.

Pan, Y.B., Burner, D.M., Ehrlich, K.C., Grisham, M.P. and Wei, Q. (1997). Analysis of primer-derived, non-specific amplification products in RAPD-PCR. *Biotechniques* **22**: 1071-1077.

Panaud, O., Chen, X. and McCouch, S.R. (1995). Frequency of microsatellite sequences in rice (*Oryza sativa* L.). *Genome* **38**: 1170-1176.

Panaud, O., Chen, X. and McCouch, S.R. (1996). Development of microsatellites markers and characterisation of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L). *Molecular and General Genetics* **252**: 597-607.

Parker, P.G., Snow, A.A., Schug, M.D., Booton, G.C. and Fuerst, P.A. (1998). What molecules can tell us about populations: choosing and using a molecular marker. *Ecology* **79**: 361-382.

Peakall, R. (1997). PCR based genetic markers and their application in turfgrass breeding. *International Turfgrass Society Research Journal* **8**: 243-257.

Pejic, L., Ajmone-Marsan, P., Morgante, M., Kozumplick, V., Castiglioni, P., Taramino, G. and Motto, M. (1998). Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs and AFLPs. *Theoretical and Applied Genetics* **97**: 1248-1255.

Pestsova, E.G., Goncharov, N.P. and Salina, E.A. 1998. Elimination of a tandem repeat of telomeric heterochromatin during the evolution of wheat. *Theoretical and Applied Genetics* **97**: 1380-1386.

Pikaard, C.S. 2002. Genomic change and gene silencing in polyploids. *Trends in Genetics* **17**: 675-677.

Piperidis, G. (2001). Personnel communication. David North Plant Research Centre, Bureau of Sugar Experimental Stations, Brisbane, Australia.

Piperidis, G., Taylor, G. O. and Smith, G. R. (2001). A microsatellite marker database for fingerprinting sugarcane clones. *Proceedings of the International Society of Sugar Cane Technologists* **24**: 637-639.

Powell, W., Mahray, G.C. and Provan, J. (1996a). Polymorphism revealed by simple sequence repeats. *Trends in Plant Science* **1**: 215-222.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S. and Rafalski, A. (1996b). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* **2**: 225-238.

Prasad, M., Varshney, R.K., Kumar, A., Balyan, H.S., Sharma, P.S., Edwards, K.J., Singh, H., Dhaliwal, H.S., Roy, J.K. and Gupta, P.K. (1999). A microsatellite marker associated with a QTL for grain protein content on chromosome arm 2DL of bread wheat. *Theoretical and Applied Genetics* **99**: 341-345.

Prasad, M., Varshney, R.K., Roy, J.K., Balyan, H.S. and Gupta, P.K. (2000). The use of microsatellites for detecting DNA polymorphisms, genotype identification and genetic diversity in wheat. *Theoretical and Applied Genetics* **100:** 584-592.

Price, S. (1957). Cytological studies in *Saccharum* and allied genera. Part III. Chromosome numbers in interspecific hybrids. *Botanical Gazette* **118**: 146-159.

Price, S. (1968). Cytology of Chinese and North Indian sugarcanes. *Economic Botany* **22**: 155-164.

Rash, K. (1995). The structure of technology in Brazilian sugarcane production, 1975-1987 - an application of modified symmetric generalised McFadden cost function. *Journal of Applied Econometrics* **10**: 221-232.

Richards, R.I., Holman, K., Yu, S. and Sutherland, G.R. (1993). Fragile X syndrome unstable element, poly(CCG)$_n$, and other simple tandem repeat sequences are binding sites for specific nuclear proteins. *Human Molecular Genetics* **2**: 1429-1435.

Roach, B.T. (1969). Cytological studies in *Saccharum*. Chromosome transmission in interspecific and intergeneric crosses. *Proceedings of the International Society of Sugar Cane Technologists* **13**: 901-920.

Roach, B.T. (1972). Nobilisation of sugarcane. *Proceedings of the International Society of Sugar Cane Technologists* **14**: 206-216.

Roach, B.T. (1989). Origin and improvement of the genetic base of sugarcane. *Proceedings of the Australian Society of Sugar Cane Technologists* **11**: 34-47.

Roder, M.S., Korzan, B.S., Gill, B.S. and Ganal, M.W. (1998a). The physical mapping of microsatellite markers in wheat. *Genome* **41**: 278-283.

Roder, M.S., Korzan, K., Wendehake, J. and Plaschke, J. (1998b). A microsatellite map of wheat. *Genetics* **149**: 2007-2023.

Roder, M.S., Plaschke, J., Konig, S.U., Borner, A., Sorrells, M.E., Tankley, S.D. and Ganal, M.W. (1995). Abundance, variability and chromosomal location of microsatellites in wheat. *Molecular and General Genetics* **246**: 327-333.

Rongwen, J., Akkaya, M.S., Bhagwat, A.A., Lavi, U. and Cregan, P.B. (1995). The use of microsatellite DNA markers for soybean genotype identification. *Theoretical and Applied Genetics* **90**: 43-48.

Russell, J., Fuller, J.D., Young, G., Thomas, B., Taramino, G., Macaulay, M., Waugh, R. and Powell, W. (1997). Discriminating between barley genotypes using microsatellite markers. *Genome* **40**: 442-450.

Saghai Maroof, M.A., Biyashev, R.M., Yang, G.P., Zhang, Q. and Allard, R.W. (1994). Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations and population dynamics. *Proceedings of the National Academy of Science USA* **91**: 5466-5470.

Saiki, R.K. (1989). The design and optimisation of the PCR. In: Erlich, H.A. (ed.), *PCR technology: principles and applications for DNA amplification*. Stockton Press, New York, pp. 7-16.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. (1985). Enzymatic amplification of ß-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia. *Science* **230**: 1350-1354.

Saliba-Colombani, V., Causse, M., Gervais, L. and Philouza, J. (2000). Efficiency of RFLP, RAPD and AFLP markers for the construction of an intraspecific map of the tomato genome. *Genome* **43**: 29-40.

Sambrook, J. and Russell, D.W. (2001). *Molecular cloning: a laboratory manual.* Cold Spring Harbour Laboratory Press, New York.

Sanchez-Escribano, E.M., Martin, J.R., Carreno, J. and Cenis, J.L. (1999). Use of sequence tagged microsatellite site markers for characterising table grape cultivars. *Genome* **42**: 87-93.

Schmidt, T. and Heslop-Harrison, J.S. (1996). High resolution mapping of repetitive DNA by *in situ* hybridisation: molecular and chromosomal features of prominent dispersed and discretely localised DNA families from the wild beet species *Beta procumbens. Plant Molecular Biology* **30**: 1099-1114.

Schwengel, D.A., Jedlicka, A.E., Nanthakumar, E.J., Weber, J.L. and Levitt, R.C. (1994). Comparison of fluorescent based semi-automated genotyping of multiple microsatellite loci with autoradiographic techniques. *Genomics* **22**: 46-54.

Scrimshaw, B.J. (1992). A simple non-radioactive procedure for visualisation of $(dC\text{-}dA)_n$ dinucleotide repeat length polymorphisms. *Biotechniques* **13**: 189.

Senior, M.L., Chin, E.C.L., Lee, M. and Smith, J.S.C. (1996). Simple sequence repeat markers developed from maize sequences found in the Genebank database: map construction. *Crop Science* **36**: 1676-1683.

Senior, M.L. and Heun, M. (1993). Mapping maize microsatellites and Cr conformation of target repeats using a CT primer. *Genome* **36**: 584-589.

Shibata, D. (1994). Extraction of DNA from paraffin embedded tissues for analysis by polymerase chain reaction: new tricks from an old friend. *Human Pathology* **25**: 561-563.

Siles, B.A., O'Neil, K.A., Fox, M.A., Anderson, D.E., Kuntz, A.F., Ranganath, S.C. and Morris, A.C. (2000). Genetic fingerprinting of grape plant (*Vitis vinifera*) using random amplified polymorphic DNA (RAPD) analysis and dynamic size-sieving capillary electrophoresis. *Journal of Agricultural and Food Chemistry* **48**: 5903-5912.

Sivolob, AN. and Khrapunov, S.N. (1995). Translational positioning of nucleosomes on DNA: the role of sequence-dependent isotropic DNA bending stiffness. *Journal of Molecular Biology* **247**: 918-931.

Slighton, J.L., Blechl, A.E. and Smithies, O. (1980). Human foetal Gγ- and Aγ-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* **21**: 627-638.

Smith, G.P. (1976). Evolution of repeated DNA sequences by unequal crossover. *Science* **191**: 528-535.

Song, K., Lu, P., Tang, K. and Osborn, T.C. (1995). Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proceedings of the National Academy of Science USA* **92**: 7719-7723.

Sreenivasan, T.V., Ahloowalia, B.S. and Heinz, D.J. (1987). Cytogenetics. In: Heinz, D.J. (ed.), *Sugarcane improvement through breeding.* Elsevier Press, Amsterdam, pp. 211-253.

Sreenivasan, T.V. and Jagathesan, D. (1975). Meiotic abnormalities in *Saccharum spontaneum. Euphytica* **24**: 543-549.

Strand, M., Prolla, T.A., Liskay, R.M. and Petes, T.D. (1993). Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**: 274-276.

Struss, D. and Plieske, J. (1998). The use of microsatellite markers for detection of genetic diversity in barley populations. *Theoretical and Applied Genetics* **97**: 308-315.

Tang, Y., Sorrells, M.E., Kochian, L.V. and Garvin, D.F. (2000). Identification of RFLP markers linked to the barley aluminium tolerance gene *Alp*. *Crop Science* **40**: 778-782.

Tautz, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* **17**: 6463-6471.

Tessier, C., David, J., This, P., Boursiquot, J.M. and Charrier, A. (1999). Optimisation of the choice of molecular markers for varietal identification in *Vitis vinifera* L. *Theoretical and Applied Genetics* **98**: 171-177.

Thomas, M.R. and Scott, N.S. (1993). Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence-tagged sites (STSs). *Theoretical and Applied genetics* **86**: 985-990.

Tinker, N.A., Fortin, M.G. and Mather, D.E. (1993). Random amplified polymorphic DNA and pedigree relationships in spring barley. *Theoretical and Applied Genetics* **85**: 976-984.

Valdes, A.M., Slatkin, M. and Freimer, N.B. (1993). Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics* **133**: 737-749.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**: 4407-4414.

Wang, Y.H., Amirhaeri, S., Kang, S., Wells, R.D. and Griffith, J.D. (1994a). Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene. *Science* **265**: 669-671.

Wang, Z., Weber, J.L., Zhong, G. and Tanksley, S.D. (1994b). Survey of plant short tandem DNA repeats. *Theoretical and Applied Genetics* **88**: 1-6.

Warburton, M. and Hoisington, D. (2001). Applications of molecular marker techniques to the use of international germplasm collections. In: Henry, R.J. (ed.), *Plant genotyping: the DNA fingerprinting of plants*. CABI Publishing, Wallington, UK, pp. 83-93.

Weber, J.L. (1990). Informativeness of Human $(dC\text{-}dA)_n$ - $(dG\text{-}dT)_n$ polymorphisms. *Genomics* **7**: 524-530.

Weber, J.L. and Wong, C. (1993). Mutation of human short tandem repeats. *Human Molecular Genetics* **2**: 1123-1128.

Weir, B.S. (1996). *Genetic-Data Analysis II: methods of discrete population genetic data*. Sinauer Assoc, Inc., Sunderland, Massachusetts.

Weising, K., Nybom, H., Wolff, K. and Meyer, W. (1995). *DNA fingerprinting in plants and fungi*. CRC Press, USA.

Welsh, J., Honeycutt, R.J., McClelland, M. and Sobral, B.W.S. (1991). Parentage determination in maize hybrids using the arbitrarily primed polymerase chain reaction (AP-PCR). *Theoretical and Applied Genetics* **82**: 473-476.

Whalen, M.D. (1991). Taxonomy of *Saccharum* (Poaceae). *Baileya* **23**: 109-125.

Wilder, J. and Hollocher, H. (2001). Mobile Elements and the genesis of microsatellites in Dipterans. *Molecular Biology and Evolution* **18**: 384-392.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**: 6531-6535.

Wu, K.S. and Tanksley, S. (1993). Abundance, polymorphism and genetic mapping of microsatellites in rice. *Molecular and General Genetics* **241**: 225-235.

Wunsch, A. and Hormaza, J.I. (2002). Cultivar identification and genetic fingerprinting of temperate fruit tree species using DNA markers. *Euphytica* **125**: 59-67.

Xu, H., Wilson, D.J., Arulsekar, S. and Balkalinsky, A.T. (1995). Sequence-specific polymerase chain-reaction markers derived from randomly amplified polymorphic DNA markers for fingerprinting grape (*Vitis*) rootstocks. *Journal of the American Society for Horticultural Science* **120**: 714-720.

Xu, X.F., Mei, H.W., Luo, L.J., Cheng, X.N. and Li, Z.K. (2002). RFLP-facilitated investigation of the quantitative resistance of rice to brown plant hopper (*Nilaparvata lugens*). *Theoretical and Applied Genetics* **104**: 248-253.

Yang, G.P., Saghai Maroof, M.A., Xu, C.G., Zhang, Q. and Biyashev, R.M. (1994). Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. *Molecular and General Genetics* **245**: 187-194.

Zabeau, M. and Vos, P. (1993). Selective restriction fragment amplification: a general method for DNA fingerprinting. European patent application number 92402629.7 (Publication number 0 534 858 A1).

Zane, L., Bargelloni, L. and Patarnello, T. (2002). Strategies for microsatellites isolation: a review. *Molecular Ecology* **11**: 1-16.

Zhang, M., Yu, A.and Chen, R. (2001). Utility of SSRs for determining genetic similarity and relationships in *Saccharum* and its related genera. *Proceedings of the International Society of Sugar Cane Technologists* **24**: 630-361.

Ziegle, J.S., Su, Y., Corcoran, K.P., Nie, L., Mayrand, E., Hoff, L.B., Mcbride, L.J., Kronick, M.N. and Diehl, S.R. (1992). Application of automated DNA sizing technology for genotyping microsatellite loci. *Genomics* **14**: 1026-1031.

APPENDICES

Appendix 1. The scoring data for the 10 microsatellite primer pairs used in the parentage analysis

The microsatellite markers were scored from the bottom of the gel upwards (smallest to largest amplicons) with marker 1 being at the bottom.

Microsatellite primer pair: SMC336BS																								
Marker No.	N14	N17	N18	NCo376	CP57/614	81W50	CP56/59	Co1001	75F2297	MZC74/275	77F790	79F2011	84F2753	76F0879	AA40 02	AA40 09	AA40 41	AA40 43	AA40 61	AA40 82	AA40 90	AA40 111	AA40 126	AA40 143
1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0	1	1	0	0
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	0	0	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1
4	0	0	0	0	0	1	0	1	0	0	0	1	0	0	1	1	0	0	1	0	1	0	1	1
5	0	0	1	1	1	1	1	0	0	1	0	0	0	1	0	1	1	1	1	0	1	1	1	0
6	0	0	1	1	1	1	1	0	1	0	0	0	0	1	0	1	0	1	1	0	1	1	1	0
7	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	0

Microsatellite primer pair: SMC278CS

Marker No.	1	2	3	4	5	6
N14	1	0	0	0	0	0
N17	1	1	0	0	0	0
N18	1	1	0	1	1	0
NCo376	1	1	1	1	1	1
CP57/614	1	0	0	1	1	1
81W50	1	1	1	1	1	1
CP56/59	1	1	0	1	1	0
Co1001	0	0	0	0	0	0
75F2297	1	1	0	0	1	0
MZC74/275	1	1	0	1	0	0
77F790	1	1	0	0	0	0
79F2011	1	1	1	0	0	0
84F2753	1	0	0	0	0	0
76F0879	1	1	0	1	1	0
AA40 02	1	1	1	0	0	0
AA40 09	1	1	1	1	1	1
AA40 41	1	1	0	1	0	0
AA40 43	1	1	0	1	1	0
AA40 61	1	1	1	1	1	0
AA40 82	1	1	1	1	1	0
AA40 90	1	1	1	1	1	1
AA40 111	1	1	0	1	1	0
AA40 126	1	1	1	1	1	1
AA40 143	1	1	1	0	0	0

Microsatellite primer pair: SMC334BS

Marker No.	1	2
N14	1	1
N17	1	1
N18	1	1
NCo376	1	1
CP57/614	1	1
81W50	1	1
CP56/59	1	0
Co1001	1	0
75F2297	1	1
MZC74/275	1	1
77F790	1	1
79F2011	1	1
84F2753	1	1
76F0879	1	1
AA40 02	1	1
AA40 09	0	1
AA40 41	0	1
AA40 43	1	1
AA40 61	1	1
AA40 82	0	1
AA40 90	1	1
AA40 111	0	1
AA40 126	1	1
AA40 143	0	1

Microsatellite primer pair: SMC213MS																								
Marker No.	N14	N17	N18	NCo376	CP57/614	81W50	CP56/59	Co1001	75F2297	MZC74/275	77F790	79F2011	84F2753	76F0879	AA40 02	AA40 09	AA40 41	AA40 43	AA40 61	AA40 82	AA40 90	AA40 111	AA40 126	AA40 143
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	0	1	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	1
5	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	1	1	1	1	1	0
6	1	0	0	1	1	0	1	1	1	1	0	1	1	0	0	1	1	1	0	1	1	1	1	1

Microsatellite primer pair: SMC286CS

Marker No.	N14	N17	N18	NCo376	CP57/614	81W50	CP56/59	Co1001	75F2297	MZC74/275	77F790	79F2011	84F2753	76F0879	AA40 02	AA40 09	AA40 41	AA40 43	AA40 61	AA40 82	AA40 90	AA40 111	AA40 126	AA40 143
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	0	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	0	1	1	0	1	1
4	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	0	0	1	1	1	1	1	1	1
5	1	1	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	1	1	0
6	1	1	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	0
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1

Microsatellite primer pair: SMC280CS																							
Marker No.	N14	N17	N18	NCo376	CP57/614	81W50	CP56/59	Co1001	75F2297	MZC74/275	77F790	79F2011	84F2753	76F0879	AA40 02	AA40 09	AA40 41	AA40 43	AA40 61	AA40 82	AA40 90	AA40 111	AA40 126	AA40 143
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0	1
4	1	0	1	1	1	0	1	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1
5	0	1	1	1	0	1	0	1	0	0	0	1	1	0	1	1	1	1	1	1	1	1	0	1
6	0	1	1	1	1	1	0	1	0	0	0	1	1	0	1	1	1	1	1	1	1	1	0	1
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Microsatellite primer pair: SMC36BUQ																							
Marker No.	N14	N17	N18	NCo376	CP57/614	81W50	CP56/59	Co1001	75F2297	MZC74/275	77F790	79F2011	84F2753	76F0879	AA40 02	AA40 09	AA40 41	AA40 43	AA40 61	AA40 82	AA40 90	AA40 111	AA40 126	AA40 143
1	1	1	1	1	1	1	1	1	0	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	0	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1
4	1	0	1	1	0	0	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	0
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	1	0	0	0	0	0	0	0	1	1	0	0	0	1	1	0	0	0	0	1	1	0	1	1
7	1	0	0	0	0	0	0	0	1	1	0	0	0	1	1	0	0	0	0	1	1	0	1	1
8	0	0	1	0	0	1	0	1	1	1	0	0	1	0	0	0	0	1	0	0	1	1	1	1
9	0	0	1	0	0	1	0	1	1	1	0	0	1	0	0	0	0	1	0	0	1	1	1	1
10	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Microsatellite primer pair: SMC219MS

Marker No.	N14	N17	N18	NCo376	CP57/614	81W50	CP56/59	Co1001	75F2297	MZC74/275	77F790	79F2011	84F2753	76F0879	AA40 02	AA40 09	AA40 41	AA40 43	AA40 61	AA40 82	AA40 90	AA40 111	AA40 126	AA40 143
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	0	1	1	1	1	0	1	1	0	0	1	0	0	1	1	1	1	1	0	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	1	0	1	0	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
6	1	1	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0
7	1	1	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0
8	1	1	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0

Microsatellite primer pair: SMC17AUQ

Marker No.	N14	N17	N18	NCo376	CP57/614	81W50	CP56/59	Co1001	75F2297	MZC74/275	77F790	79F2011	84F2753	76F0879	AA40 02	AA40 09	AA40 41	AA40 43	AA40 61	AA40 82	AA40 90	AA40 111	AA40 126	AA40 143
1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	0	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	1	1	0	1	0	0	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1
5	0	1	1	1	0	1	0	1	1	1	0	1	1	1	0	1	0	0	1	1	1	1	1	0
6	1	1	1	1	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
7	0	0	0	1	1	1	0	0	1	0	1	0	1	0	0	1	1	1	1	1	1	1	1	1
8	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	1	0	1	1	1	0	0	0	0	1	0	1	0	0	1	1	1	1	1	1	1	1	0
10	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Microsatellite primer pair: SMC569CS																								
Marker No.	N14	N17	N18	NCo376	CP57/614	81W50	CP56/59	Co1001	75F2297	MZC74/275	77F790	79F2011	84F2753	76F0879	AA40 02	AA40 09	AA40 41	AA40 43	AA40 61	AA40 82	AA40 90	AA40 111	AA40 126	AA40 143
1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
2	1	0	1	0	0	0	1	1	0	1	0	1	0	1	1	1	0	1	0	1	0	0	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	0	1	0	0	0	0	1	0	1	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1
5	0	1	0	0	0	0	1	0	1	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1
6	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	1	0	1	0	0	1	1
7	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	1	0	1	0	0	1	1
8	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	1	0	1	0	0	1	1
9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Appendix 2. The scoring data for the 14 microsatellite primer pairs used in the fingerprinting study of 20 sugarcane cultivars

The microsatellite markers were scored from the bottom of the gel upwards (smallest to largest amplicons) with marker 1 being at the bottom.

Microsatellite primer pair: mSSCIR12

Marker No.	Coimbatore	Mandalay	Badila	B.Cheribon	B46364	Co270	Co290	Co475	LCP85-384	NCo310	NCo376	POJ2878	Q115	Q124	Q136	Q96	Q117	R570	RB72-454	SP70-1143
1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1
2	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1
3	0	0	1	1	0	1	0	0	0	1	1	1	1	1	1	0	1	1	1	0
4	1	1	0	1	1	1	1	0	1	1	1	1	0	1	0	1	0	0	0	0
5	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	1
6	1	0	0	0	0	1	0	1	1	1	1	1	0	0	0	1	1	1	1	1
7	1	0	0	0	0	1	1	1	0	0	0	0	1	1	0	0	1	1	1	1
8	0	1	0	1	0	0	1	1	1	1	1	0	0	0	1	1	1	0	0	0
9	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0

Microsatellite primer pair: mSSCIR25

Marker No.	Coimbatore	Mandalay	Badila	B.Cheribon	B46364	Co270	Co290	Co475	LCP85-384	NCo310	NCo376	POJ2878	Q115	Q124	Q136	Q96	Q117	R570	RB72-454	SP70-1143
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	1	1	1	0	1	1	1	1	1	0	0	1	1	1	1	1	1
4	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	0	0	0

Microsatellite primer pair: SMC336BS

Marker No.	Coimbatore	Mandalay	Badila	B.Cheribon	B46364	Co270	Co290	Co475	LCP85-384	NCo310	NCo376	POJ2878	Q115	Q124	Q136	Q96	Q117	R570	RB72-454	SP70-1143
1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	1	0	0	0	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1
4	0	0	0	1	0	1	1	1	1	1	1	0	0	1	1	1	1	0	1	1
5	1	1	0	0	1	0	1	0	0	0	0	0	1	1	0	0	1	0	1	1
6	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	1	0	0	1	1	0	1	1	1	0	0	0	1	1
8	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0

Microsatellite primer pair: mSSCIR56

Marker No.	Coimbatore	Mandalay	Badila	B.Cheribon	B46364	Co270	Co290	Co475	LCP85-384	NCo310	NCo376	POJ2878	Q115	Q124	Q136	Q96	Q117	R570	RB72-454	SP70-1143
1	0	0	0	0	1	1	1	1	0	1	1	0	1	1	1	0	0	1	0	1
2	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	1	1	1	0
4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
6	1	0	0	1	1	1	0	0	0	0	0	1	1	1	1	0	1	1	1	1
7	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	1
8	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0
9	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	0
10	0	0	0	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0

Microsatellite primer pair: mSSCIR19

Marker No.	Coimbatore	Mandalay	Badila	B.Cheribon	B46364	Co270	Co290	Co475	LCP85-384	NCo310	NCo376	POJ2878	Q115	Q124	Q136	Q96	Q117	R570	RB72-454	SP70-1143
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	1	0
3	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
5	0	1	0	0	1	0	1	0	1	1	1	0	1	0	1	1	1	0	1	0
6	1	0	0	1	0	0	1	0	1	1	0	0	1	1	1	1	1	1	0	0
7	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	1	0	1	0	1	0	0	0	0	1	0	1	1	1	0	0	0	1	0
9	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0

Microsatellite primer pair: SMC119CG

Marker No.	Coimbatore	Mandalay	Badila	B.Cheribon	B46364	Co270	Co290	Co475	LCP85-384	NCo310	NCo376	POJ2878	Q115	Q124	Q136	Q96	Q117	R570	RB72-454	SP70-1143
1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1
2	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
3	0	0	1	1	1	0	1	0	0	1	0	1	1	1	1	1	1	0	0	0
4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
6	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0	1	0
7	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
8	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1
9	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	1	1	0	0	0	0	0	1	1	0	1	1	1	1	1	0	0	0

Microsatellite primer pair: SMC278CS

Marker No.	Coimbatore	Mandalay	Badila	B.Cheribon	B46364	Co270	Co290	Co475	LCP85-384	NCo310	NCo376	POJ2878	Q115	Q124	Q136	Q96	Q117	R570	RB72-454	SP70-1143
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	1	0	1	1	1	1	1	1	0	0	1	1	1	1	0	1	1
3	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	1	0	0	1	1	0	1	1	1	0	0	0	1	1
5	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
6	0	0	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0	0	1	0
7	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Microsatellite primer pair: SMC334BS

Marker No.	Coimbatore	Mandalay	Badila	B.Cheribon	B46364	Co270	Co290	Co475	LCP85-384	NCo310	NCo376	POJ2878	Q115	Q124	Q136	Q96	Q117	R570	RB72-454	SP70-1143
1	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	0	0	0	1	0
2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	1	0	0	0	0	0	1	1	1	1	1	0	0	0
4	0	0	0	0	0	0	1	1	0	1	1	0	1	1	1	1	1	0	0	0
5	0	0	0	0	0	0	1	1	0	1	1	0	1	1	1	0	1	0	0	0
6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	1	0	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1

Microsatellite primer pair: SMC36BUQ

Marker No.	Coimbatore	Mandalay	Badila	B.Cheribon	B46364	Co270	Co290	Co475	LCP85-384	NCo310	NCo376	POJ2878	Q115	Q124	Q136	Q96	Q117	R570	RB72-454	SP70-1143
1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	1
2	0	0	0	0	1	1	0	0	1	1	0	0	1	1	1	0	0	0	0	0
3	0	1	0	0	1	1	0	1	1	1	0	0	0	1	1	0	0	1	0	0

Microsatellite primer pair: SMC749BS

Marker No.	Coimbatore	Mandalay	Badila	B.Cheribon	B46364	Co270	Co290	Co475	LCP85-384	NCo310	NCo376	POJ2878	Q115	Q124	Q136	Q96	Q117	R570	RB72-454	SP70-1143
1	0	0	0	0	1	0	0	1	1	0	1	1	0	0	0	1	0	1	0	0
2	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	0

Microsatellite primer pair: SMC286CS

Marker No.	Coimbatore	Mandalay	Badila	B.Cheribon	B46364	Co270	Co290	Co475	LCP85-384	NCo310	NCo376	POJ2878	Q115	Q124	Q136	Q96	Q117	R570	RB72-454	SP70-1143
1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
3	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
4	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0
5	0	0	0	1	1	1	1	0	1	0	1	0	0	0	1	0	1	1	1	1
6	0	0	1	0	1	1	1	0	0	0	1	1	1	1	1	1	0	0	0	1

Microsatellite primer pair: mSSCIR10

Marker No.	Coimbatore	Mandalay	Badila	B.Cheribon	B46364	Co270	Co290	Co475	LCP85-384	NCo310	NCo376	POJ2878	Q115	Q124	Q136	Q96	Q117	R570	RB72-454	SP70-1143
1	1	0	1	0	0	1	1	1	0	0	0	0	0	1	1	1	1	1	1	0
2	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Microsatellite primer pair: mSSCIR55

Marker No.	Coimbatore	Mandalay	Badila	B.Cheribon	B46364	Co270	Co290	Co475	LCP85-384	NCo310	NCo376	POJ2878	Q115	Q124	Q136	Q96	Q117	R570	RB72-454	SP70-1143
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	1	1	0	1	1	0	1	1	1	1	0	0	1	1	1	1	1
4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Microsatellite primer pair: mSSCIR60

Marker No.	Coimbatore	Mandalay	Badila	B.Cheribon	B46364	Co270	Co290	Co475	LCP85-384	NCo310	NCo376	POJ2878	Q115	Q124	Q136	Q96	Q117	R570	RB72-454	SP70-1143
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0