Implementation of molecular markers for triticale cultivar identification and marker-assisted selection

by

Daphne Nyachaki Bitalo

Presented in partial fulfillment of the requirements for the degree of

Master of Science at the Department of Genetics, University of Stellenbosch.

Study Leader: Willem C. Botes

Department of Genetics

Faculty of AgriScience

December 2012

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Name: Daphne Nyachaki Bitalo

Date: 16/01/2012

Abstract

Triticale is an amphidiploid that consists of wheat (A and B) and rye (R) genomes. This cereal is fast becoming important on a commercial basis and warrants further assessment for the better management and breeding of the hybrid. The assessment of the genetic diversity among the wheat and rye genomes within triticale can be obtained by using molecular markers developed in both donor genomes. Simple sequence repeats markers (SSRs) and amplified fragment length markers (AFLPs) have been previously used to assess the genetic diversity among triticale lines.

SSRs are highly polymorphic markers that are abundant and which have been shown to be highly transferable between species in previous studies while AFLP markers are known to generate plenty of data as they cover so many loci.

Thus, the aim of this study was to develop a marker system suitable to assess the genetic diversity and relationships of advanced breeding material (and cultivars) of the Stellenbosch University's Plant Breeding Laboratory (SU-PBL).

Therefore, both AFLP and SSR markers were initially analysed using eight triticale cultivars (with known pedigrees) to facilitate cultivar identification. Fourty-two AFLP primer combinations and 86 SSR markers were used to assess the genetic diversity among the Elite triticale cultivars.

The AFLP primer combinations generated under average polymorphism information content (PIC) values. Furthermore, these markers generated neighbour-joining (NJ) and unweighted pair group method with arithmetic average (UPGMA) dendograms that displayed relationships that did not correspond with the available pedigree information. Therefore, this marker system was found not to be suitable.

A set of 86 SSRs previously identified in both wheat and rye, was used to test the genetic diversity among the eight cultivars. The markers developed in wheat achieved 84% transferability while those developed in rye achieved 79.3% transferability. A subset of SSR markers was able to distinguish the cultivars, and correctly identify them by generating NJ and UPGMA dendograms that exhibited relationships that corroborated the available pedigree data. This panel of markers was therefore chosen as the most suitable for the assessment of the advanced breeding material.

The panel of seven SSR markers was optimised for semi-automated analysis and was used to screen and detect the genetic diversity among 306 triticale entries in the F6, Senior and Elite phases of the SU-PBL triticale breeding programme. An average PIC value of 0.65 was detected and moderate genetic variation was observed. NJ and UPGMA dendograms generated showed no clear groupings. However, the panel of markers managed to accurately identify all cultivars within the breeding program.

The marker panel developed in this study is being used to routinely distinguish among the advanced breeding material within the SU-PBL triticale breeding programme and as a tool in molecular-assisted backcross.

Abbreviations

A	Adenine
AFLPs	Amplified fragment length polymorphisms
ATP	Adenosine triphosphate
ß-ME	ß-Mercaptoethanol
bp	Base pairs
C	Cytosine
CAPS	Cleaved amplified polymorphic sequences
cDNA	Complimentary DNA
cm	centimetres
CTAB	Cetyl trimethylalmmonium bromide
^{o}C	Degrees Celsius
DAFF	Department of Agriculture, Forestry and Fisheries
ddH$_2$O	Double distilled water
DNA	Deoxyribonucleic acid
gDNA	Genomic deoxyribonucleic acid
dNTPs	Deoxyribonucleotidetriphosphate
ds	Double stranded
DUS	Distinctness, uniformity and stability
*EcoR*I	Restriction enzyme from *Escherichia coli* strain R
EST	Expressed sequence tag
EST-SSRs	Expressed sequence tag derived simple sequence repeats
EtBr	EthidiumBromide
IP	Intellectual Property
IRAP	Inter-retrotransposon amplified polymorphism
m	metres

MAS	Marker-assisted selection
MI	Marker index
mins	Minutes
NJ	Neigbour-joining
ng	Nano grams
PAGE	Polyacrylamide gel electrophoresis
RAMPO	Retrotransposon-microsatellite amplified polymorphism
RAPD	Random amplification of polymorphic DNA
RFLP	Random amplified Length polymorphisms
PBR	Plant Breeders' Rights
PIC	Polymorphism information content
SCARS	Sequence characterized amplified region
sec	Seconds
SNPs	Simple sequence repeats
SSCP	Single strand conformational polymorphism
SSRs	Simple sequence repeats
STS	Sequence tagged site
SU-PBL	Stellenbosch University's plant breeding laboratory
UPGMA	Unweighted pair group method with arithmetic average
UPOV	International Convention for the Protection of New Varieties of Plants
USA	United States of America
VDAs	Variant detector arrays

List of Tables

List of Figures

Contents

Language and style used in this thesis are in accordance with the requirements of the South African Journal of Plant and Soil. This thesis represents a compilation of manuscripts where each chapter is an individual entity and some repetition between chapters has, therefore, been unavoidable.

Chapter 1:   A review of marker assisted selection strategies, progress and current applications in triticale and rye

Introduction

The aim of this study is to implement a minimum marker set for the introduction of marker-assisted selection (MAS) in the triticale breeding programme at Stellenbosch University's Plant Breeding Laboratory (SU-PBL).

Since triticale bares three genomes; two wheat (AABB) and rye (RR), the first two portions of the study objectives are based on the assessment of the three genomes within triticale. The objectives are;

- AABB genomes;
  - Identify AFLP markers and assess their suitability in the screening of triticale; and
  - Optimise microsatellite markers previously identified in the wheat genome and in triticale advanced breeding lines and cultivars.
- RR genome;
  - Identify AFLP markers and their suitability in the screening of triticale; and
  - Optimise microsatellite markers previously identified in the rye genome and in triticale advanced breeding lines and cultivars.

The final objective of the study is to assess the implementation of the most suitable marker system for high throughput analysis that can be routinely performed by the breeding programme on a seasonal basis. This will be achieved by;

- Optimizing the most polymorphic marker system for semi-automated analysis; and
- Using this marker system to test current advanced breeding material within the SU-PBL

The resulting evidence from these evaluations will be used to propose the most robust, commercially convenient and reliable methodology for the testing of genetic diversity, to identify material for backcrosses and fingerprint material for intellectual property (IP) purposes

A review of marker-assisted selection strategies, progress and current applications in triticale and rye

D. Bitalo

Stellenbosch University Plant Breeding Laboratory, Department of Genetics, Private Bag X1, Matieland 7602, South Africa

Abstract

Triticale is an intergeneric hybrid derived from a wheat (Triticum *turgidum* L.) rye (Secale *cereale* L.) cross. Early work with triticale focused on primary triticales, which were characterised by poor agronomical traits and cytological instability. However, advancements in biotechnology aided in the development of improved triticale genotypes that gave rise to commercial cultivars.

The progression of marker-assisted selection (MAS) strategies in the cereal breeding industry is becoming vital for the potential gain and effective preservation of the genetic resources at hand. Using molecular marker data plant varieties can reliably be identified when running distinctness, uniformity and stability (DUS) tests for the awarding and protection of plant breeders' rights (PBR). Molecular marker data is also accurately used to assess the genetic diversity among cultivars without biasing results especially for species that lack reliable pedigree information.

Keywords: Triticale, rye, molecular markers, Restriction fragment length polymorphisms (RFLPs), Random amplified polymorphic DNA (RAPDs), Single nucleotide polymorphism (SNPs), Simple sequence repeats (SSRs) genetic diversity, marker-assisted selection (MAS), plant breeders' rights (PBR), genetic diversity.

Triticale

Origin

The origin of triticale dates back to 1873 when Scottish scientist, A. Stephen Wilson, made the first cross between wheat and rye (Oettler, 2005). Using diploid rye (*Secale cereale* L.) as the male parent, he applied its pollen on the stigma of emasculated tetraploid durum wheat (*Triticum turgidum* L.) and hexaploid wheat (*Triticum aestivum* L.) to produce triticale hybrids (Hulse, 1976). With this cross, Wilson succeeded in obtaining plants that exhibited characteristics intermediate to the two parental species and presented these results at a meeting of the Royal Botanical Society. However, both plants were sterile and produced completely dysfunctional pollen grains (Ammar *et al.,* 2004).

During the years subsequent to the report by Wilson, many publications in the 1800's on wheat- rye hybrids were recorded; most notably the spontaneous doubling of chromosomes in the partially fertile hybrids grown by Rimpau in 1888 (Oettler, 2005). In 1918, the agricultural experimental station at Saratov in Russia, reported several thousand naturally occurring wheat-rye hybrids which were male sterile, and had an inability to reproduce themselves (Hulse, 1976). This became a persistent trend for most of the wheat-rye hybrids in the early years of triticale's development, pushing researchers and plant breeders to rely on spontaneous chromosome doubling, and the natural viability of the hybrid embryo (Oettler, 2005). The development of improved techniques of embryo culture, to rescue the aborting embryo (Laibach, 1925) and colchicine, to double the chromosome number (Pierre Givaudon, 1937 cited by Hulse, 1976), brought the dawn of commercial scale triticale breeding.

The nomenclature and naming of triticale created so much confusion and several names were proposed (Oettler, 2005). In 1899, the researcher Wittmack suggested that the hybrid be named after a merge between the names of its donor parents. Finally, the name *Triticosecale* or short-form, triticale, was agreed upon. Almost 70 years later, the scientist Bernard R. Baum suggested that the full name should be x *Triticosecale* Wittmack in honour of Wittmack (Oettler, 2005). Finally, a consensus was reached. Following the international code of botanical nomenclature, the name x *Triticosecale* Wittmack ex A. Camus or common name 'triticale' was agreed upon (Oettler, 2005). The name was first quoted in literature published in Germany in 1935, and had been coined by Lindschau and Oehler (1935) as a fusion between the Latin words for wheat (*Triticum*) and rye (*Secale*) (Dogan *et al.,* 2009).

Triticale development and domestication

Triticale is an allopolyploid (amphiploid), implying that it stably bares the genomes of wheat (*Triticum* sp.) and rye (*Secale* sp.). Therefore, the initial triticales are fertile, true-breeding progenies that result from an intergeneric hybridization which is followed by chromosome doubling between a seed parent from the genus *Triticum* and a pollen parent from the genus *Secale* (Ammar *et al.,* 2004). The development of triticale was motivated by the concept of unifying the positive characteristics exhibited by its donor parents. For instance, wheat is mainly used in food products while rye thrives in non-optimal environments (McGoverin *et al.,* 2011).

Advancements in the development of the wheat-rye hybridization technology resulted in triticale variants that exhibit different genome structures and ploidy levels, with respect to the wheat (AABBDD) and rye (RR) genomes (McGoverin *et al.,* 2011). Therefore, Kiss (1966) suggested that there be a clear distinction in the naming of these triticale variants. He then introduced the terms primary and secondary triticale, which are now generally accepted (Oettler, 2005).

Primary triticales are newly synthesised allopolyploids generated from wheat-rye crosses. Earlier work on triticale focused on octoploid primary triticale ($2n = 8x = 56$, AABBDDRR), which result from a cross between hexaploid *Triticum* spp. and rye (Oettler, 2005; Thiemt and Oettler, 2008). These lines however, didn't meet the expected agronomical performance and so more work was done with hexaploid primary triticale (Ammar *et al.,* 2004; Oettler, 2005).

Hexaploid primary triticale are synthesised from crossing tetraploid *Triticum* spp. with rye (Thiemt and Oettler, 2008). These progenies play a vital part in the provision of starting material for breeding programs in North America and Europe. The production of hexaploid primaries was due to the improvement in embryo rescue culture techniques, and the discovery of colchicine (Ammar *et al.,* 2004). Furthermore, researchers produced hexaploid primaries, because it performed better agronomically and commercially than the octoploid primaries (Ammar *et al.,* 2004; Oettler, 2005; Mergoum *et al.,* 2009).

Much interest was taken in the morphology, cytology and agronomical performance of the primary triticale progenies (Oettler, 2005). Early on, researchers noticed that the practical superiority expected from the primary triticale was not being realised and so studies were undertaken to improve it (Ammar *et al.,* 2004; Oettler, 2005). The focus was to develop

13

germplasm that did not present with the cytological instability exhibited by the octoploids in particular and progeny that could perform agronomically as a commercial crop (Oettler, 2005; Mergoum *et al.,* 2009).

In 1954, Kiss made his first octoploid x hexaploid crosses and by 1960, he obtained secondary triticale that were superior to the donor parents but lacked straw strength (Ammar *et al.,* 2004). However, such shortfalls did not deter Kiss, who continued improving his material using octoploid x hexaploid crosses. Eventually, two secondary triticale selections, 'Triticale No. 57' and 'Triticale No.64' were the first to be released for commercial production and were grown at a scale of 40,000 ha in Hungary (Zillinsky, 1974; Ammar *et al.,* 2004).

Today, secondary hexaploid triticale are the most commercially grown triticale worldwide and these can be generated by crossing primary triticale cultivars of similar or different ploidy levels, or by crossing triticale cultivars with wheat or rye cultivars (McGoverin *et al.,* 2011; Oettler, 2005).

Commercially, two types of hexaploid triticale are grown; complete triticale, which carry all seven pairs of unchanged chromosomes from rye, and substituted triticale, which have one or more of the rye chromosomes replaced with D-genome chromosomes from hexaploid wheat (Fox *et al.,* 1990). One such substitution involves the replacement of chromosome 2R of rye by 2D from wheat, which is thought to have arisen from a natural recombination event (Fox *et al.,* 1990).

Triticale cultivars can be further classified into three basic types; spring, winter, and intermediate (facultative). Spring triticale types are day length insensitive, exhibit upright growth and produce abundant amounts of forage early in their growth due to their short growth period. These types are mostly bred in warmer areas like South Africa and Australia (Santiveri *et al.,* 2002; Mergoum *et al.,* 2004; Salmon *et al.,* 2004).

Winter triticale types require cold conditions to induce floral differentiation and are generally planted in the fall. These types have a slow and long growth cycle and yield more forage than the spring types. Winter types have achieved a bigger cultivation level than spring types because of the extensive progression of these types in Poland, Northern Europe and North America (Mergoum *et al.,* 2009).

The facultative triticale cultivars have low vernaliszation requirements and can be grown in both spring and winter (Mergoum *et al.,* 2004; Salmon *et al.,* 2004).

Production and utilisation of triticale

Triticale commercial breeding programmes were initially started in the 1950s by Sánchez-Monge (1958) in Spain. However, the most extensive and successful breeding program to date is the collaboration between Centro Internacional de Mejoramiento de Maiz y Trigo (CIMMYT) in Mexico and the universities of Manitoba and Guelph in Canada (Hulse, 1976; Oettler, 2005). South Africa established a triticale breeding program in 1974 at the Department of Genetics, Stellenbosch University, with the aims to improve the disease resistance ability of triticale to increase yield and nutrient content (Pienaar *et al.,* 1991).

As of 2005, 199 cultivars of triticale were reported by the Official Journal of the European Union (Tams, 2006). And in 2009, triticale breeding was recorded in 24 European countries and grossed 4.3 million hectares of harvest worldwide, according to the Food and Agriculture Organisation (FAOSTAT) report. In many triticale growing European countries such as Hungary, the harvest areas of rye and triticale are equal (Tams, 2006).

Triticale has become a commercially established crop internationally and is mainly used in the farming industry as feed for animals, forage, silage, grain-feed and hay (Mergoum *et al.,* 2009). However, the crop is also being exploited as a potential bio-energy and bio-ethanol source (Eudes, 2006; McGoverin *et al.,* 2011).

Triticale is used as a feed- grain by poultry and pig farmers. Triticale mainly replaced rye in the feeding of poultry. Rye contains high levels arabinoxylans that induce the excessive consumption of water by poultry (Boros 2002; Mergoum *et al.,* 2009). Therefore, studies were done by Boros (1998), and these showed that no noticeable negative effects occurred within a cohort of broiler chickens that were fed with hexaploid or octoploid types of winter triticale in Poland. Triticale has also performed well in the feeding of hogs even those at a tender age and a newly developed digestive system. This is due to the higher level of lysine found within the cereal that makes it a good substitution for a maize-based diet for pigs. Triticale is further being used as a source of feed for other livestock such as cattle, sheep and goats due to this very reason (Salmon *et al.,* 2004; McGoverin *et al.,* 2011).

The use of triticale as a source of forage compares favourably with other small-grain cereals (Varughese *et al.,* 1996). The crop can be grown for green forage and silage either as a mono-crop or the winter and spring types can be blended (Baron *et al.,* 1992). It can also be mixed with other legumes cereals or annual rye grass (Carnide *et al.,* 1998). Overall, triticale can increase the rotation of crops in a maize-based forage industry (Oettler, 2005; Mergoum *et al.,* 2009).

Furthermore, triticale is used as a rotation crop to prevent the infestation of pests building up within other crops (Mergoum *et al.,* 2009).

Although the major purpose for developing triticale was to increase the source of cereal foods consumed by humans, triticale has failed to reach that goal. The cereal lacks the baking quality of wheat due to low gluten contents but has a higher content of proteins. Therefore, the baking quality of this cereal needs further improvement (Salmon *et al.,* 2004)

Triticale presents genetic uses as well. Triticale incorporates the superior nutritional properties of wheat with the tolerance of rye to survive in adverse surroundings (Dogan *et al.,* 2009), and has the potential to be used as an intermediate to transfer genes of interest from rye into wheat, particularly those related to biotic and abiotic stresses (da Costa *et al.,* 2007). For instance, a cross between triticale and wheat was employed to transfer desirable Hessian fly resistance genes, found on chromosome 2RL of rye, to wheat (Vaillancourt *et al.,* 2007).

Rye

Rye (*Secale cereale* L.) is a cross pollinated, diploid (2n=14) species belonging to the grass tribe of *Triticeae,* which is shared by wheat, and is grown in temperate regions (Persson and Bothmer, 2002). The genus *Secale* is divided in to three broad groups; wild, weedy and cultivated and only S. *cereale* is cultivated (Khush, 1963). The scientist Vavilov (1917) was the first to postulate that cultivated rye originated from weedy rye, and in the succeeding years, he and his associates undertook a systematic collection and classification of the various types of weedy rye. Parts of Afghanistan, central Asia, northern Iran and Turkey were explored to analyse the morphological appearance of spikes in weedy ryes and the distribution of different grain.

Although these studies provided useful information in establishing the centre of variability of weedy ryes in the countries surrounding the Caspian Sea, they also resulted in taxonomic

confusion. Some were regarded as varieties, subspecies or independent species by various researchers (Khush, 1963).

Past studies on rye had not paid much attention to the cytogenetic relationship between weedy and cultivated rye species and it wasn't until the discovery by Schiemann and Nürnberg-Krüger (1952) that cultivated rye, *S. cereale*, differs from wild perennial rye, *S. montanum*, by two reciprocal translocations, that much fascination was aroused in studying the chromosomal arrangements of the other *Secale* species. Studies by Khush and Stebbins (1961) and Khush (1962) showed that five species of *Secale* exhibit differences in the end arrangements of their chromosomes.

Despite the number of studies run on the genus *Secale,* plenty of disparity was noted over its classification. Finally, a consensus was reached and the genus was classified in the modern taxonomic systems adopted by the American Germplasm Resources Information Network, (GRIN) (http://www.ars-grin.gov; sourced October, 2011). The genus *Secale* is now known to comprise of four species; *S. cereale* L., *S. sylvestre* Host, *S. vavilovii* Grossh, and *S. strictum* (Tang *et al.,* 2011).

Cultivated rye is thought to have originated in the Anatolean Plateau of Turkey (Feuillet *et al.,* 2007), and although its primary centre of origin is not known, the centre of diversity is commonly accepted as Southwest Asia (which includes: Turkey, Armenia, and Iran). This is essentially the same area of origin as common wheat, barley, and oats (Bushuk, 2001; Shang *et al.,* 2006). Several authors have speculated on how cultivated rye penetrated the borders of Europe; Khush (1962) concluded that it probably entered through the northern Caucuses or through central Asia, Bushuk (1976) proposed that it was probably distributed from south-western Asia to Russia, and subsequently into Poland and Germany from where it gradually spread throughout most of Europe (Ma *et al.,* 2004) and throughout the globe in all the major cereal producing countries.

Production and utilisation of rye

Rye covers an extensive range of purposes making it a versatile crop. It is utilised as both animal feed and fertilizer in crop rotations as a green plant. As grain, it is used as animal feed and feedstock in the production of alcohol beverages. Rye rates second to wheat as the most commonly used grain crop in the production of leavened bread and is favoured over wheat in

17

the production of "black" bread (enjoyed by most people because of the characteristic rye flavour), which is a widespread staple in Eastern Europe and parts of Asia. Rye exhibits winter hardiness, can be grown in cool temperate and semi-arid regions and can easily flourish in areas with temperatures too severe for wheat or barley. Most cultivated rye is sown during fall and generally called "winter rye" but however, some spring rye cultivars are grown in warmer areas (e.g. South Africa), but spring cultivars are generally known for their inferior agronomic characteristics (Bushuk, 2001).

Archaeobotanical evidence places the existence of cultivated rye in Europe around the Bronze Age (1800–1500 BC), in the old Czechoslovakia, Moldavia and Ukraine; when rye only occurred as a weed and was later used as a cultivated crop in the Nordic region around 500 AD (Persson and Von Bothmer, 2002). This then led to the development of landraces. The evidence further illustrated that the spread of rye cultivation in Europe was mostly in areas where farming was difficult, and that it was cultivated due to its superior yield in ecologically subsidiary areas, and developments in farming techniques and the growing demand for cereals (Behre, 1992; Persson and von Bothmer, 2002).

Modern cultivated rye is distributed in parts of central Europe, the western Mediterranean region, the Caucasus to Central Asia, with isolated populations in South Africa (Chikmawati et al., 2005). The amount of rye cultivated worldwide has shown a decline; from approximately 10.8 million hectares in 1995 to 6.6 million hectares as of 2009 (FAOSTAT). Rye is an important source of useful genes for wheat and triticale improvement and is a constituent of triticale (Varshney et al., 2007). This makes rye vital to the cereal industry (Khlestkina et al., 2009).

Molecular markers

The use of genetic markers is not novel; Gregor Mendel used phenotype-based genetic markers in an experiment in the nineteenth century (Agarwal *et al.,* 2008). These phenotype-based markers are referred to as morphological markers and are one of the three main types of markers used in plant breeding. Morphological markers show mutations of morphological characteristics like plant height, which is noted in the plant's phenotype. However, previous studies of genetic diversity in triticale done by Royo *et al.,* (1995) and Furman *et al.,* (1997) using morphological markers, showed that morphological characteristics are prone to environmental conditions and are limited in number. Thus, the minimal coverage of the genome by phenotypic markers and their resulting estimated diversity values are not an actual reflection of the genetic difference of the populations under study (Kuleung *et al.,* 2006).

The second type of marker system is the protein marker system. These markers detect the presence, absence or abundance of a specific protein. In other words, these markers are associated with gene products. They are classified as either isozymes or allozymes. Isozymes are the most frequently used of the two, and show a varying effect in an electric field (in a resolving medium) on enzymes that catalyse the same chemical reactions but have different amino acid sequences and charge (Weising *et al.,* 2005). Allozymes (alloenzymes) are variants of an enzyme that are coded by different alleles at the same locus. These markers exhibit high levels of functional conservation and low levels of polymorphism. Both isozymes and allozymes have a neutral effect on the phenotype of a plant and have the ability to discriminate between heterozygotes and homozygotes, are inexpensive in comparison to molecular markers and are involved in well-known roles in the metabolic system. However, protein markers cover a limited number of loci, exhibit low levels of polymorphism and require the use of a different protocol for each isozyme system (Farooq and Azam, 2002).

The application of molecular markers in some manner has become routine in most crops nowadays. However, it is vital to identify the most efficient and cost efficient markers to apply in breeding programmes (Gupta and Varshney, 2000).

 DNA markers are fragments of DNA sequences that exhibit differences, or then polymorphism, among different individuals or different fragments of DNA. It is the most versatile of all markers as it is not affected by the environment, more stable than morphological markers, and present in all tissues regardless of growth and differentiation

19

phases (Agarwal *et al.,* 2008). These markers have been put to use in genome mapping (Hackauf and Wehling, 2002), DNA fingerprinting (Gupta and Varshney, 2000), and the study of genetic diversity (Kuleung *et al.,* 2006).

According to Weising *et al.,* 2005 and Agarwal *et al.,* 2008, a molecular marker is considered ideal based on the following criteria;

- Even and frequent occurrence in the genome;
- Provision of adequate resolution of genetic differences;
- Generation of multiple and reliable, independent markers;
- Quick, easy and inexpensive to assay techniques;
- Exhibition of moderate to high levels of polymorphism;
- Exhibition of linkage to distinct phenotypes;
- Exhibition of codominant inheritance;
- Unambiguous assignment of alleles; and
- No need for prior information on the genome under study.

Developing and deciding upon the most ideal molecular markers is not without challenges. However, it would be best to take into consideration factors such as reproducibility, technical requirement and many others before settling on the most ideal molecular marker for any given application. Table 1.1 summarises some of the characteristics of the most frequently used molecular markers in the applied plant sciences, and more specifically plant breeding.

DNA markers can be analysed by hybridization-based or polymerase chain reaction (PCR) based techniques. However, PCR-based markers are favoured over hybridization-based markers, because of their simplicity, sensitivity, the low amounts of DNA used and their high amenability to automation hence producing more accurate, reliable, cost effective and high throughput genetic information (Röder *et al.,* 1998; Tams *et al.,* 2004; 2005).

PCR-based molecular markers

Restriction fragment length polymorphism

Restriction fragment length polymorphism (RFLP) is a technique based on the PCR amplification of specific fragments, which are subsequently subjected to endonuclease digestions using DNA restriction enzymes that recognise specific sequences in the PCR amplicons. These restriction enzymes catalyse endonucleolytic cleavages to yield fragments of defined lengths. The resulting restriction fragments may be visualised on agarose gels after being separated according to their molecular sizes. The resulting differences in molecular sizes create a differential profile which could be as a result of the absence/ presence of a cleavage site, or the insertion or deletion of blocks of DNA. The variations can also alter the length of the DNA fragments and are detected as a discrete marker directly linked to the genotype of an individual organism (Botstein *et al.,* 1980).

Due to the frequent occurrence of RFLPs in the genome, the RFLP technique is considered a relatively superior marker which is robust, relatively polymorphic, reproducible and exhibits codominant inheritance (Agarwal *et al.,* 2008). Nonetheless the technique isn't short of limitations; it is quite laborious and time consuming, involves expensive assay techniques, requires large amounts of high quality DNA (about 50-200μg, microgram), involves use of radioactive reagents (Farooq and Azam, 2002) and requires prior sequence information for the generation of probes (Agarwal, *et al.,* 2008).

Random amplified polymorphic DNA

The Random amplified polymorphic DNA (RAPD) technique is based on the random PCR amplification of genomic DNA using single primers of arbitrary nucleotide sequence. RAPDs infer DNA polymorphisms produced by rearrangements or deletions at or between oligonucleotide primer binding sites in the genome using short random oligonucleotide primers (about ten bases long) (Agarwal *et al.,* 2008). This technique has the ability to detect polymorphisms without prior sequence knowledge making it easily applicable across species using universal primers (Williams *et al.,* 1990).

The RAPD technique is comprised of two variant methodologies. The first one is the arbitrarily primed polymerase chain reaction (AP-PCR) (Welsh and McClelland, 1990),

21

which requires a single primer of 10-15 nucleotides long. This technique requires the amplification of DNA at a low stringency for the first two PCR cycles and amplification at high stringency annealing temperatures for the rest of the cycles (Agarwal *et al.,* 2008).

Table 1.1: Characteristics of frequently used molecular markers (adapted from Agarwal *et al.,* 2008)

Marker	Abun-dance	Repro-ducibility	Polymor-phism	Locus spe-cific	Technical require-ment	DNA re-quired	Major appli-cation
RAPD	High	High	Medium	Yes	High	High	Physical mapping
RFLP	High	Low	Medium	No	Low	Low	Gene tagging
SSR	Medium	Medium	Medium	Yes	Medium	Low	Genetic diversity
SSCP	Low	Medium	Low	Yes	Medium	Low	SNP mapping
CAPS	Low	High	Low	Yes	High	Low	Allelic diversity
SCAR	Low	High	Medium	Yes	Medium	Low	Gene tagging, Physical mapping
AFLP	High	High	Medium	No	Medium	Low	Gene tagging, Genetic diversity
IRAP	High	High	Medium	Yes	High	Low	Genetic diversity
RAMPO	Medium	Medium	Medium	Yes	High	Low	Genetic diversity

RFLP restriction fragment length polymorphism, RAPD random amplified polymorphic DNA, SSR simple sequence repeats, SSCP single strand conformational polymorphism, CAPS cleaved amplified polymorphic sequence, SCAR sequence characterized amplified region, AFLP amplified fragment length polymorphism, IRAP inter-retrotransposon amplified polymorphism, RAMPO retrotransposon-microsatellite amplified polymorphism.
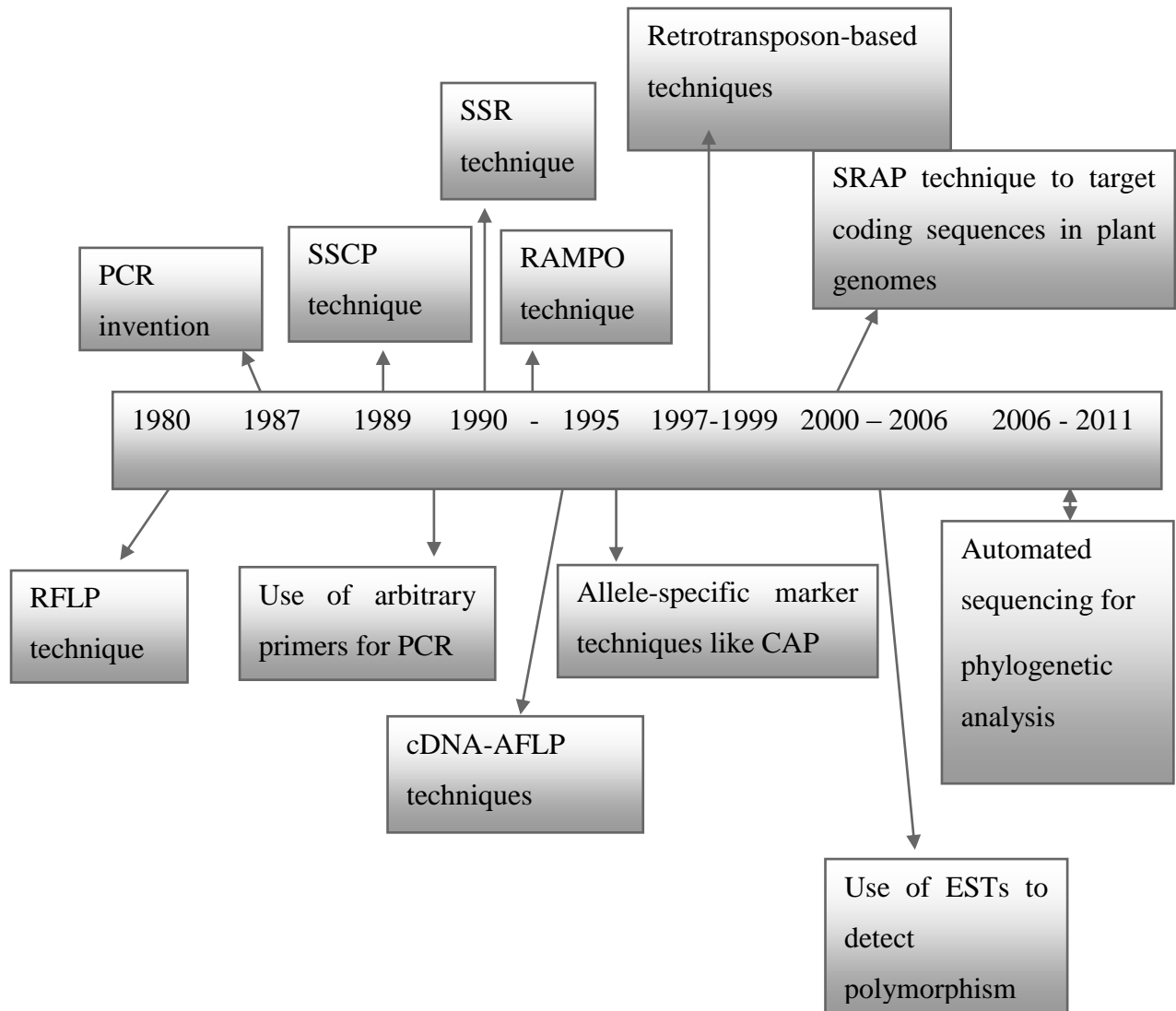
Figure 1.1: A schematic illustration of the advancements in molecular techniques over a period of two decades (adapted from Agarwal *et al.,* 2008).

The second RAPD marker variant is the DNA amplification fingerprinting (DAF) (Caetano-Anolles and Bassam, 1993) technique, which entails the application of a single arbitrary primer shorter than ten nucleotides for amplification (Agarwal *et al.,* 2008). The resulting amplicons are visualised on silver stained polyacrylamide gels.

RAPD markers are well suited for DNA fingerprinting and have also been used in studying the genetic diversity among populations like the *Jatropha* species (Ram *et al.,* 2008). RAPD markers exhibit an efficient assessment of polymorphisms for rapid identification, are locus specific and relatively cheaper than other PCR-based DNA markers. Other benefits this technique boasts over some of the PCR-based DNA markers are;

- The use of a universal set of primers for the genomic analysis across a wide variety of species;

- No need for nucleotide sequencing or isolation of cloned DNA; and

- Each RAPD marker is the equivalent of a sequence tagged site (STS), making it useful for the development of physical maps (Williams *et al.,* 1990).

However, RAPDs show low levels of polymorphism, and fail to distinguish between heterozygotes and homozygotes. RAPD markers can however exhibit codominance when each RAPD fragment is amplified by PCR using specific primers in a technique called sequence characterised amplified regions (SCARs) but this technique requires prior sequencing data of the RAPD band which can be a limiting factor (Farooq and Azam, 2002).

Single nucleotide polymorphisms

Single nucleotide polymorphisms (SNPs) are units of genetic variation and represent the DNA sequence differences between alleles. These polymorphisms are easily used as molecular markers as they are widely distributed throughout the genome with variations in occurrence and distribution across species (Rafalski, 2002). For instance, maize is estimated to have 1 SNP per 60-120 bp (Ching *et al.,* 2002) and the human genome is estimated to have a SNP every 1000 bp (Sachidanandam *et al.,* 2001). SNPs are currently the most popular marker platform in plant breeding because of their genetic stability over SSRs and their ability to construct genetic maps with 100 times more marker density than is possible with SSRs (Varshney *et al.,* 2007).

In most of the genomes studied thus far, SNPs were found to occur mostly in the coding regions of the genome (Edwards *et al.,* 2007). These SNPs are either synonymous or non-synonymous. Synonymous SNPs do not alter the amino acid sequence, but can modify mRNA splicing, hence causing differences in phenotype (Agarwal *et al.,* 2008). Non-synonymous SNPs result in an alteration in the amino acid sequence and these mutations have a neutral effect on the phenotype (Soleiman *et al.,* 2003).

Prior to their use, SNPs must undergo detection and validation. Detection can be done using any of the following four techniques (some of which are facilitated by the increase in the number of EST sequences deposited in public domains);

- Identification of single strand conformation polymorphisms (SSCP);

- Heteroduplex analysis;

- Direct DNA sequencing; or

- Variant detector arrays (VDAs) (Shah and Kusiak, 2004).

Detection of SNPs by SSCP involves the amplification of the DNA fragment spanning the putative SNP. This fragment is then denatured and analysed on a non-denaturing polyacrylamide gel where the single stranded fragment take on a secondary structure (Gray *et al.,* 2000). The migration rates of various fragments are used to identify SNPs. This is the most commonly used detection technique because of its simplicity, high sensitivity and low cost (Liu *et al.,* 2007). However, this technique is sensitive to the size of fragments and exhibits limited efficiency for amplicons above 200 bp and lower than 500 bp (Shahinnia and Sayed-Tabatabaei, 2009).

Single nucleotide polymorphisms can also be detected by re-sequencing of amplicons with or without pre-screening, searching genomic libraries and searching expressed sequence tag (EST) deposits available from various public domains such as the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov; cited October, 2011). The direct sequencing of DNA amplicons from different individuals is one of the more popular ways of detecting SNPs. Primers are designed to amplify about 400-700 bp of DNA segments derived from EST deposits or genes of interest. The PCR products are then sequenced directly in both directions and the resulting sequences aligned either manually or using various sequence alignment software such as CAP3 (Huang and Madan, 1999) to discern true polymorphisms from sequencing errors (Duran *et al.,* 2009).

The validation of SNPs is executed using various techniques, for instance primer extension, allele-specific hybridization, oligonucleotide ligation and restriction enzyme digestion (Soleiman *et al.,* 2003; Agarwal *et al.,* 2008). There have been many advancements in the techniques used to genotype SNPs (e.g. DNA chips) in order to make the technique more attractive for high throughput use in (Rafalski, 2002) the construction of high-density genetic maps (Liu *et al.,* 2007), identification of crop cultivars, assessment of genetic diversity (Varshney *et al.,* 2007), detection of linkage disequilibrium across genomes (Rafalski, 2002) and marker-assisted breeding. The SNP technique exhibits other advantages, which include;

- The ability to genotype small PCR amplicons (less than 100 bp) with accuracy irrespective of DNA degradation or presence of PCR inhibitors;
- Sample processing is mostly automated.

This technique is ailed with one major disadvantage, SNP markers produce less alleles than SSR markers hence more SNPs need to be genotyped to obtain a distinctive DNA profile. This can be overcome by the amount of SNPs available and the low costs involved in automated high throughput genotyping of SNPs (Yan *et al.,* 2009).

Simple sequence repeat markers

Simple sequence repeats (SSRs) or microsatellites are variable tandem repeat sequences of DNA spanning one to six base pairs and they serve as highly informative genetic markers. It was hence recommended that a merger be brought between the informative nature of these tandem repeats, and the rapid and sensitive nature of PCR (Akkaya *et al.,* 1992). SSRs have the ability to exhibit high levels of polymorphisms due to the rapid evolution of the repeat units, they occur frequently in the genome of most organisms making them highly polymorphic and multiallelic, are codominant and mostly chromosome specific (Röder *et al.,* 1998). Though they are time consuming and expensive to develop and optimise (Weising *et al.,* 2005), they also allow for automated analysis and high reproducibility (Chen *et al.,* 2007).

There are two different types of SSR markers. Expressed sequence tags (ESTs) are sequences found in the transcribed regions of the genome and these are used to develop EST-derived microsatellites or EST-SSRs. This marker system detects the variation in the expressed portion of the genome. The markers found in the non-transcribed regions of the genome are called genomic SSRs (Weising *et al.,* 2005). Although EST-SSRs have been found to be less polymorphic than genomic SSRs (Leigh *et al.,* 2003), their increase in public databases such as GenBank (http://www.ncbi.nlm.nih.gov; cited October, 2011) and their ability to analyse transcribed regions of even the most redundant genomes like wheat (Leigh *et al.,* 2003) and rye (Hackauf and Wehling, 2002), makes them a valuable tool for genetic diversity studies.

Simple sequence repeat markers exhibit transferability between populations and have made studies on the evaluation of the genetic diversity among triticale more feasible (Tams *et al.,* 2004; Kuleung *et al.,* 2006). The usefulness of SSRs in analysing the structure of triticale

27

germplasm is reliant on the SSR markers developed in wheat and rye. Several groups have embarked on this task with varying success (Röder *et al.,* 1995; Saal and Wricke, 1999).

Design and analysis of SSR markers

Microsatellites are amplified by PCR using primers (18-25 bp long) specific for sequences flanking varying regions of tandem repeats of about two to six base pairs. The varying number of tandem repeats causes fragments of different sizes to be amplified and this exhibits the polymorphic nature of SSR markers (Manifesto *et al.,* 2001). There are a number of user-friendly software that can be used to design primers, for instance, Primer3 (Rözen and Skaletsky, 2000) and Oligo Analyzer version 1.0.2 (Teemu Kuulasmaa, 2000). The resulting PCR amplicons can be visualised on silver stained denaturing polyacrylamide gels (PAGE) as banding patterns or haplotypes.

Simple sequence repeat marker analysis can alternatively be automated. The PCR protocol in this analysis is similar to the one with conventional primers; however, automated analysis requires the use of fluorescent labelled primers, which are genotyped on an automated DNA sequencer. The PCR reaction contains only one fluorescently labelled primer and this can allow for multiplexing of primers with different colour labels and size ranges. The possibility to multiplex primers in a PCR reaction or in a run increases which increases the number of samples genotyped which subsequently cuts costs. Therefore, more data is generated from one lane on a DNA sequencer and alleles are scored more accurately since the automated sequencer has an internal size standard (Jewell *et al.,* 2010)

However, SSR marker analysis presents with some genotyping errors. For instance, in cases where the DNA is of low quality/ quantity, the number of target DNA molecules in the extract is low leading to only a few intact molecules. This situation will most likely lead to allele dropouts which results from the preferential amplification of certain DNA fragments. Stutter bands can be formed and these appear as faint bands below or before the actual band. Stutter bands are as a result of the repeat structure of SSRs, a phenomenon known as PCR slippage. The appearance of stutter bands is worsened with the use of di-nucleotide repeats and their intensity can vary across different germplasm. This problem can be remedied by optimising PCR conditions like altering annealing temperatures, altering the number of

amplification cycles, using higher quality *Taq* polymerase or by selecting SSRs with higher repeat units (Pompanon *et al.,* 2005).

Data analysis

Amplified DNA band fragments are scored as alleles and their sizes determined using a 50 bp or 100 bp ladder. Repeats of SSRs are scored as haplotypes with each locus representing the genetic information of each SSR marker while the bands produced by each AFLP primer combination are scored as present (1) or absent (0) (Vyhnánek *et al.,* 2009). This data can then be recorded in excel and imported in to PowerMarker version 3.25 (Liu and Muse, 2005) to be converted into frequency data for example. The frequency data is used to calculate frequency-based distances, which give an indication of the genetic distance between two populations or individuals. Calculating the genetic distance between two populations, gives a relative estimate of the time that has elapsed since the populations were established. Therefore, as the amount of time which separates two populations increases, the difference in the allele frequencies is expected to increase. This is probably due to the differentiation in the allele frequencies at selectively neutral loci as a result of mutation and genetic drift. A commonly used distance measure in this programme is CSChord (Cavalli-Sforza and Edwards, 1967). CSChord computes the chord distance between two populations without involving any evolutionary models (Cavalli-Sforza and Edwards, 1967).

The resulting frequency-based distances are usually used to construct phylogenetic trees using several methods. The unweighted pair group method with arithmetic average (UPGMA) (Sneath and Sokal, 1973) trees progressively cluster the most closely related taxa until all the taxa form a rooted tree. These trees are called ultrametric trees since the UPGMA method assumes a constant rate of molecular clock so that all genetic distances fit on a clock-like tree (Nei, 1987). On the other hand, the neighbour-joining (NJ) (Saitou and Nei, 1987) tree clusters taxa to form an unrooted tree and requires that the genetic distances only be additive. Therefore, a NJ tree doesn't assume a constant rate of molecular clock as it makes more precise estimates of branch lengths (Nei, 1991).

Both NJ and UPGMA are distance-matrix methods that assemble the observed taxonomic units (OTUs) in to a phylogenetic tree based on the distances calculated between the OTUs. These generated phylogenetic trees exhibit just the minimal information about the OTUs being tested and have been used in combination in previous studies as this generates a more accurate grouping of the OTUs (Tamura *et al.,* 2004)

To calculate the discriminatory power of each marker, Weir's (1996) gene diversity and its alternative polymorphism information content (PIC) can be computed using the software

30

PowerMarker v3.25. This depicts the aptitude of a marker to detect polymorphisms in the population under study.

The gene diversity, which is referred to as the expected heterozygosity, defines the probability that any two alleles; chosen randomly from a population, are different (Moose and Mumm, 2008). The unbiased estimator of gene diversity at the *l*th locus is;

$D = 1 - \frac{1}{L} \sum_l \sum_i P_i^2$ (Liu and Muse, 2005). The value P is the frequency of the *i*th allele at the *l*th locus and L is the number of loci.

The increase in PIC can be a substantial indicator to the discriminatory power of a marker (Botstein et al., 1980) and is estimated as;

$PIC = 1 - \sum_i^n P_i^2$ (Liu and Muse, 2005). The value P_i is the frequency of the *i*th allele and n is the number of loci examined.

Note that the PIC obtained when using AFLP markers cannot exceed 0.5 and so when calculating this value, an arithmetic mean is computed;

$PIC = \sum_{j=1}^N PIC_j / N$ (Vuylsteke *et al.*, 2000).

PIC_j is the PIC value calculated at locus j of the AFLP marker and N is the total number of AFLP markers generated by a primer combination.

Degree of variability

The concept of plant breeding is simple; cross the best parents and recover the progeny that outperforms the parents. This principle however, is more complex in practice as it entails three main steps.

First, germplasm with useful genetic variation must be selected or created. Secondly, the germplasm with superior phenotypes are identified. Finally, improved cultivars are developed from the selected germplasm (Moose and Mumm, 2008). Plant breeders are therefore tasked with so many steps that have to fit the objectives of the breeding scheme. Plant breeders unfortunately don't have as much exposure to survey methods needed and hence may work in relative isolation from the multitude of farmers' preferences (Morris and Bellon, 2004). Most plant breeders focus on mere crop yield and resistance to diseases and pests, and don't investigate deeper into other factors such as ease of harvest and storage, taste and cooking qualities, crop maturity and suitability of crop residues as livestock feed, as these factors have proved difficult to improve in conventional plant breeding programs. Hence, the need for diversity both phenotypically and genetically must be balanced by Elite performance to maximise the probability for successful improvement in breeding (Moose and Mumm, 2008).

Over 60% of the global population consumes cereal crops as staple food (FAO, 2007). The end consumption of these cereals varies depending on their carbohydrate, lipid, protein and vitamin percentage composition. The levels at which these components vary in the cereals has been shown to be in correlation to the species, genotypes and to a lesser degree, the production environment. Therefore, it is essential to reliably identify specific cereal species and cultivars to control the handling, marketing and processing of these items for consumption (Ko et *al.,* 1994; Terzi *et al.,* 2005).

The variation in DNA of species is dependent on recombination, mutation within a genome, migration rates of a population, selection within a population, the overall population size and the subdivisions within, and random genetic drift (Talbert *et al.,* 1998). Ultimately, high levels of variability equate to an increased ability to respond to threats such as disease, parasites and environmental change, and low levels of variability tend to limit a species' ability to respond to such threats over time (Amos and Harwood, 1998).

Much concern is raised over the effects of habitat fragmentation, which almost always leads to population size reduction resulting in increased loss of genetic variability through genetic drift. In extreme cases of fragmentation, genetic bottlenecks are created due to the

32

exaggerated loss in population number. At levels near or at mutation drift equilibrium, high levels of variability imply either low rates of gain or low rates of loss, just as low levels of diversity imply either low rates of gain or rapid loss (Amos and Harwood, 1998). Diversity can also be gained through mutation or gene flow between neighbouring populations and can be lost actively through natural selection, which is manifested as inbreeding depression (Amos and Harwood, 1998).

To broaden the available genetic variation in future cereal breeding, it's imperative that the genetic diversity among cultivars and species be investigated. Genetic diversity can be assessed by various means ranging from morphological traits to molecular marker data. Furthermore, investigating the genetic diversity of germplasm could result in efficient management of breeding material and improvement of crop productivity (Huang and Rozelle, 2002).

To ensure that consumers get the varieties that the manufacturers claim to be selling, the quality of food must be maintained to strengthen consumer trust. Analytical tools such as DNA-based techniques have been developed to scrutinise the composition of raw materials, to identify cereal species, and to fingerprint genotypes and varieties in an effort to determine authenticity (Popping, 2002). For instance, the European Union allows for the composition of 3% non-durum wheat (e.g. T. *aestivum*) in pasta. Bryan *et al.,* 1998 used real-time PCR to quantify the amount of non-durum wheat by targeting the D-genome sequences present only in T. *aestivum*.

The identification of plant varieties also profits plant breeders and commercial companies as well as protecting their interests. Therefore, when measuring for the amount of variability in a population, it's crucial to use a technique that is accurate, allows for high throughput analysis, and is cost effective and automated. Measuring the variability in a population is incentivised by the need to examine the conservation dynamics of the population, necessity for assessing the amount of variability available for breeding purposes such as back crossing, and a call to protect the rights of released varieties for intellectual purposes (Terzi *et al.,* 2005).

33

Plant Breeders' Rights

Plant Breeders' Rights (PBR) is considered as a straight forward method of plant variety protection. In other words, it is the intellectual property right granted to a breeder of a new plant variety (Weising *et al.,* 2005). Though the protection of intellectual property by way of copyright, trademarks and patents have been well established, the intellectual protection by way of PBR is relatively recent (Evans and Haines, 2007). This right was first written into law in South Africa in 1976. Plant Breeders' Rights is governed by the department of Agriculture, Forestry and Fisheries (DAFF) under Act No 15 of 1976. South Africa's PBR Act was harmonised with the international world when it was included in the International Convention for the Protection of New Varieties of Plants (UPOV) in 1977.

The UPOV convention is an international agreement that was established in 1961 to standardise PBR laws, determine standardised procedures to test new varieties and to establish stronger ties between its member countries. In October 2010, the forty-fourth ordinary UPOV council session was held in Geneva and there were about 68 member countries represented in UPOV. UPOV has gone so far as to implement ease of access to PBR application (by any person) within any of the UPOV member countries. UPOV together with all its member countries has listed over 250 kinds of plants whose varieties can be granted PBR. These varieties are required to have certain aspects in order to be considered as 'new' and to be granted a PBR;

- Propagating material of the variety must not have been sold in the country for more than a year;
- Propagating material of a variety of a tree must not have been available in another country (in trade or to the public) for more than 6 years, or more than 4 years in the case of a different plant;
- Propagating material must comply with the distinctness, uniformity and stability (DUS) requirements; it must clearly be discernible from any other variety of the same species, all the plants in a planting must look similar and exhibit the same characteristics, repeated cultivation of the said variety must produce plants similar to the original; and
- Plant varieties must have an acceptable value (the cost must match the variety's value) (URL:http://www.upov.int/; cited October, 2011);

(http://www.nda.agric.za/docs/GeneticResources/variety_control.htm; cited October, 2011).

The PBR Act makes provision for the owner of a new variety to financially gain from efforts in the development of the given new variety. The validity of a PBR license stretches from a period of 20 to 25 years in South Africa, depending on the type of plant. According to DAFF, the first 5 to 8 years of the PBR give the owner the sole right to produce and market propagated material of the variety and only during the next 15 to 17 years of the PBR, is the owner compelled to issue licenses to other persons (who wish to use and market said material) and has the added advantage of claiming royalties from all propagated and marketed produce (URL:http://www.daff.gov.za; cited October, 2011); (http://www.nda.agric.za/docs/GeneticResources/variety_control.htm; cited October, 2011).

Thus with much to gain or lose during the production of a new variety, DUS tests are certainly the pivotal point of a breeder's hard work. DUS tests could be run over a year or a couple of years (depending on the plant species) and are based on morphological characteristic comparisons between old and new varieties. By implementing molecular marker techniques, DUS tests could become more accurate at determining distinction between varieties and be hastened along (Weising *et al.,* 2005).

Marker-assisted assessment of breeding material

Despite the progress advanced plant breeding techniques have made to increasing crop yield, plant breeders are still concerned with better ways to cope with the constant changing environments plants are exposed to. Consequently, changes in agricultural practices create an opportunity to develop genotypes with certain agronomic traits, which can cope in the target environments, keeping in mind that the organisms in each environment are constantly changing. For instance, fungal and pest communities continually evolve to eventually overcome host plant resistance. Another factor plant breeders have to cope with is the change in consumer preferences and requirements (Collard and Mackill, 2008).

However, it would be important to focus more on improving traits related to crop yield, stability and sustainability as this would address issues such as; disease resistance, abiotic stress tolerance and nutrient and water-use efficiency (Collard and Mackill, 2008). Using just conventional breeding techniques isn't going to improve any of these traits and hence plant breeders need to incorporate biotechnological techniques to achieve the stated goals (Huang and Rozelle, 2002).

Currently, DNA markers are being incorporated in plant breeding in a procedure commonly referred to as marker-assisted selection (MAS). Bearing in mind that the principle behind plant breeding is the selection of particular plants with desirable traits, and the assembly of a desirable combination of the genes coding for these traits in to new varieties, selection for these traits can be long, laborious and expensive if pedigree breeding techniques are used.

For instance, selection of plants with superior traits involves optical assessments for agronomic traits or resistance to stresses. Those traits of higher heritability would be selected for in plants in the early generations but those of low heritability would be selected in the later generations (F5 or F6) that have become more homozygous. This could take up to 5 to 10 years of constant harvesting and evaluating replicated field trials in order to identify Elite lines. Perhaps, the scope and difficulty of selection can be scaled down by the incorporation of MAS (Collard and Mackill, 2008).

The previously discussed advantages of molecular markers play a big role in the way in which plant breeding is practiced today especially when focusing on the implementation of

36

MAS. The implementation of MAS in the evaluation of backcrossing material and breeding material is henceforth explored.

Backcrossing aims to integrate one (or more) gene(s) into the adapted or Elite varieties. This usually involves a donor parent with a large number of desirable traits, albeit a deficiency in only a few characteristics (Allard, 1999). One example of marker-assisted backcrossing in plants is the incorporation of the *Pm22* genes encoding for powdery mildew in to wheat using AFLPs (Zhou et *al.,* 2005). Using three Chinese wheat cultivars as intermittent parents, Zhou et *al.,* 2005, developed 33 near-isogenic lines (NILs) carrying 22 powdery mildew resistance genes. In the agronomic trait findings, all the NILs developed showed no significant difference to their intermittent parents and the results generated by AFLP analysis indicated that the NILs had high genetic similarity to their intermittent parents. Also, the resistance to powdery mildew was stably expressed in the relevant NILs and a further screening of eleven of the NILs using molecular markers linked to the resistance genes *Pm1c*, *Pm4b*, *Pm13*, *Pm21*, *PmP*, *PmE*, *PmPS5A*, *PmPS5B*, *PmY39*, *PmY150*, and *PmH*, showed that the screened NILs all carried the targeted genes. Thus, using DNA markers in backcrossing increases the efficiency of selection as it is also used to trace the introgressed transgenes in the Elite cultivars (Collard and Mackill, 2008).

MAS can also be used to quantify breeding material for the identification or confirmation of cultivars. One of the predicaments encountered in the routines of seed handling, is the unintentional mixing of different seed strains within or between crop breeding programs, especially when handling cereal crops from the same family such as the *Triticae* family. Traditionally, such predicaments have been rectified using visual selection and data based on morphological characteristics. This however, would prove futile as the morphological tests would not be accurate (Collard and Mackill, 2008). For instance, Yashitola et *al.,* 2002 used simple sequence repeat (SSR) and sequence tagged site (STS) markers to confirm the purity of hybrid rice. This assessment was used to replace the morphological testing of the material, thereby providing accurate data and efficiency. Most importantly, the plants didn't have to be grown to maturity as MAS techniques test for genetic differences and not morphological differences. Therefore, more accurate data was generated in a short time span.

Genetic diversity evaluation

An increment in the genetic diversity in the traditional varieties, modern cultivars and landraces, could boost the levels of food production by elevating plants' resilience to disease, pest attack and environmental changes. However, the available genetic diversity in most plants is decreasing and this is believed to be as a result of the rigorous human selection for agronomic traits and the use of small population sizes (Tanksley and McCouch, 1997). Although cultivation, domestication and modern breeding methods have increased the presence of Elite cultivars, these have been noted to hold only a fraction of the available diversity in species. One such extreme example was the mass growth of the cultivar 'Turkey' (a hard red winter wheat variety) in the United States of America (USA) in 1919, which covered almost the entire acreage. This cultivar and a second cultivar called 'Tenmarq', were in 1949 replaced by four other cultivars which were less closely interrelated even though they were derivatives of Turkey. Over the years, further addition of less related germplasm was made, but this maintained the gene pool at a relatively restricted level (Feuillet *et al.,* 2008).

Plant breeders have noticed that the exploitation of regional landraces, local cultivars and related species could be the most viable source for the improvement and sustenance of the domesticated crop species. These wild ancestors of the currently domesticated and economically viable plants offer a reservoir of genetic variation (Feuillet *et al.,* 2008). The exploitation of these germplasm for genetic gain has been in the works for over a century as is noted in the creation of the triticale genome (Feuillet *et al.,* 2008) and has become a vital prerequisite for the existence of most breeding programs aimed at crop productivity (Chen and Li, 2007).

Genetic variability can be attained by introgressing major genes from wild species and this can be achieved by following two strategic steps; by using sexual hybridisation in order to convey the wild genome in to a cultivated background or by homologous and/ or homeologous recombination to dispense with or replace the deleterious alleles and/ or genes that 'come along for the ride' with the selected locus (linkage drag). There have been some major successes with the introgression of genes, like the incorporation of dwarfing genes (Reduced Height loci *Rht-B1* and *Rht-D1*) and genes conferring durable resistance against a wide spectrum of insects and diseases into wheat by Nobel laureate Dr. Norman Borlaug that fuelled the Green Revolution (Feuillet *et al.,* 2008). Introgression is however faced by numerous barriers; such as being laborious and, for complex characters, largely a failure. The

use of wild genetic resources in breeding tends to lead to linkage drag which limits efficiency as time is consumed trying to eliminate the negative alleles (Feuillet *et al.,* 2008).

The genetic variation of triticale must be sought by using its parent species due to its lack of natural evolution processes (Haesaert and De Baets, 1994). According to studies done by Góral *et al.* 1999, Góral 2002 and Oettler *et al.* 2001; 2003, the breeding of hybrid triticale cultivars gives a possible 20% yield increment in the crop's production. This is further fostered by the existence of the rye genome in triticale, which nurtures the expectation that triticale has the potential for commercial use of heterosis in hybrids (Tams *et al.,* 2004). Thus, it goes without saying that the basis for fully exploiting heterosis in triticale is the extensive depiction of its parents with regard to the development of heterotic groups. These heterotic groups can be established based on geographical origin, agronomic traits, and recorded pedigree data or on biochemical and/ or molecular marker data (Tams *et al.,* 2004).

Studies run by Royo *et al.,* 1995 based on the morphological traits exhibited by American triticale cultivars, divulged only two groups; winter and spring triticale, but failed to group the cultivars according to geographical origin. A later study done by Furman *et al.,* 1997 investigating the genetic diversity among 3,000 genotypes of triticale from USA, Canada and Mexico based on agronomical traits. These genotypes were grouped as either 'complete' or 'incomplete'. The evaluation of genetic diversity based on agronomical traits is therefore hindered by the limited number of traits and environmental changes.

When measuring for genetic diversity based on pedigree information, the co-ancestry coefficient (f); which is an indirect measure of the relative genetic similarity of related individuals, is calculated (Lübberstedt *et al.,* 2000). The estimates of genetic diversity calculated whilst using f are often unrealistic and tend to bias the estimations of the genetic diversity made. For instance, assumptions relating to the inheritance of each parental genome by descendants, genetic drift, selection pressure and relatedness of ancestors with known pedigree, are violated (Bohn *et al.,* 1999; Lübberstedt *et al.,* 2000). This shouldn't cause problems for species such as maize, whose pedigrees are reliable and suitably documented (Smith *et al.,* 1985). This adversely, is not the case for triticale whose pedigrees are mostly complicated or unreported, making it nearly impossible to acquire estimates of genetic diversity from them (Kuleung *et al.,* 2006). This lapse in information stems from the synthesis of the different types of triticale; primary types were synthesized using tetraploid or hexaploid wheat and rye populations, and secondary types were often backcrossed to wheat

and rye. However, these manipulations/ studies don't have well documented pedigree data (Tams *et al.,* 2004).

The use of direct genetic diversity estimates based on molecular markers, nevertheless managed to curb the inherit problems faced by both morphological markers and genetic diversity estimates based on pedigrees in triticale (Kuleung *et al.,* 2006). Molecular markers are not influenced by weather and exhibit accurate genetic diversity based on polymorphic DNA markers such as single nucleotide polymorphisms (SNPs), amplified length polymorphisms (AFLPs), simple sequence repeats (SSRs), among others (Chen and Li, 2007) as is illustrated in figure 1.1, which is a schematic representation of the progression of molecular marker systems over a period of two decades. Prior studies have shown that hybridisation-based markers such as RFLPs exhibit limited intraspecific polymorphisms and less success in genetic diversity studies than PCR-based markers (AFLPs, SSRs and SNPs) (Röder *et al.,* 1998).

Consequently, most plant diversity studies have incorporated the use of AFLPs, SSRs and SNPs. Examples include;

- Use of SSRs (developed in rye and wheat) and AFLPs to study the genetic diversity of European winter triticale by Tams *et al.* (2004; 2005), which detected moderate to low diversity levels among the triticale cultivars using data generated by SSRs and AFLPs respectively;
- Detection of genetic diversity among 80 hexaploid triticale accessions of European origin using SSRs (developed in wheat and rye) by Kuleung *et al.,* (2006), results showed that these markers managed to group the accessions in to 5 clusters that corresponded with the available pedigree information;
- da Costa *et al.,* 2007 determined the genetic variability of 54 triticale genotypes sourced from two breeding programs in Brazil, using 42 wheat genomic microsatellites. The set of microsatellites used grouped the genotypes in to seven clusters. The study also found a 71.42% transferability of these markers indicating a high possibility of exploiting these transferable markers in further triticale genetic studies;
- Varshney *et al.,* used 48 barley EST primer pairs to elucidate the potential of SNP markers in determining the genetic variability among five rye inbred lines. A sum of

96 SNPs and 26 indels were defined from the sequences of 14 resulting amplicons and a mean expected heterozygosity of 0.66 was observed; and many more.

References

- Agarwal, M, Shrivastava, N and Padh, H (2008) 'Advances in molecular marker techniques and their applications in plant sciences', *Plant Cell Reports*, 27: 617-631

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

- Allard, R, W, (1999) '*Principles of plant breeding*', 2nd edition. New York, NY: Wiley

- Ammar, K, Mergoum, M and Rajaram, S (2004) "The history of triticale. In: *Triticale improvement and production* (FAO plant production and protection paper 179) (edited by Mergoum, M and Gómez-Macpherson, H), 1-9. Rome: Food and Agriculture Organisation of the United Nations

- Amos, W and Harwood, J (1998) "Factors affecting levels of genetic diversity in natural populations", *Philosophical Transactions of the Royal Society of London, Biological Sciences*, 353: 177-186

- Baron, V, S, Najda, H, G, Salmon, D, F and Dick, A, C (1992) "Post-flowering forage potential of spring and winter cereal mixtures", Canadian Journal of Plant Science, 72:137-145

- Behre, K-E (1992) 'The history of rye cultivation in Europe', *Vegetation History and Archaeobotany*, 1: 141-156

- Bohn, M, Utz, H, F and Melchinger, A, E (1999) 'Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs and SSRs and their use for predicting progeny variance', *Crop Science*, 39: 228–237

- Boros, D (2002) "Physico-chemical quality indicators suitable in selection of triticale for high nutritive value. In: *Proceedings of the 5th International Triticale Symposium*, 1: 239-224

- Boros, D (1998) "Nutritive value of different forms of triticale for monogastric animals. *In:* P. Juskiw, ed. *Proceedings of the 4th International. Triticale Symposium*, Red Deer, Alberta, Canada, I: 177-184. International Triticale Association

- Botstein D, White R, L, Skolnick 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

- Bryan, G, J, Dixon, A, Gale, M, D and Wiseman, G, (1998) "A PCR-based method for the detection of hexaploid bread wheat adulteration of durum wheat and pasta", *Journal of Cereal Science,* 28: 135-146

- Bushuk, W (Editor) (1976) 'Rye: Production, chemistry and technology', *American Association of Cereal Chemists, Incorporated.* St. Paul, Minnesota. Pp 1-11

- Bushuk, W (2001) 'Rye production and uses worldwide', *American Association of Cereal Chemists, Incorporated*, 46 (2): 70-73

- Caetano-Anollés, G and Bassam, B, J, D, N, A (1993) "Amplification fingerprinting using arbitrary oligonucleotide primers", *Applied Biochemistry and Biotechnology*, 42:189-200

- Carnide, V, Guedes-Pinto, H, Miguel-Rodrigues, Sequeira, C and Mascarenhas-Ferreira, A (1998) 'Forage yield and quality of triticale-vetch mixtures', *In: Proceedings of the 4th International Triticale Symposium* (Ed. P. Juskiw), Vol. II, pp. 252-255. Red Deer, Alberta, Canada

- Cavalli-Sforza, L, L and Edwards, A, W, F (1967) 'Phylogenetic analysis: models and estimation procedures', *American Journal of Human Genetics*, 19: 233-257

- Chen, G and Li, L (2007) 'Detection of genetic diversity in synthetic hexaploid Wheat using microsatellite markers', *Agricultural Sciences in China*, 6(12): 1403-1410

- Chikmawati, T, Skovmand, B and Gustafson, J, P (2005) 'Phylogenetic relationships among *Secale* species revealed by amplified fragment length polymorphisms', *Genome*, 48: 792-801

- Ching, A, D, A, Caldwell K, S, Jung, M, Dolan, M, Smith, O, S, Tingey, S, Morgante, M, and Rafalski, A (2002) "SNP frequency, haplotype structure and linkage disequilibrium in Elite maize inbred lines", *BioMed Central Genetics*, 3: 19

- Collard, B, C, Y and Mackill, D, J (2008) 'Marker assisted selection: an approach for precision plant breeding in the twenty-first century', *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363: 557-572

- da Costa, C, T, Albuquerque, A, C, S, Junior, A, N, Marcelino, F, C and Pereira, J, F (2007) 'Genetic diversity of Brazilian triticale evaluated with genomic wheat microsatellites', *Pesquisa Agropecuária Brasília*, 42(11): 1577-1586

- Department of Agriculture, Forestry and Fisheries (DAFF) (2011), Homepage [online] (cited October 2011) Available from <URL:http://www.daff.gov.za>

- DAFF<http://www.nda.agric.za/docs/GeneticResources/variety_control.htm> sourced 18 July, 2012

- Dogan, R, Kacar, O, Coplu, N and Azkan, N (2009) 'Characteristics of new breeding lines of triticale', *African journal of Agricultural Research*, 4(2): 133-138

- Duran, C, Appleby, N, Vardy, M, Imelfort, M, Edwards, D and Batley, J (2009) 'Single nucleotide polymorphism discovery in barley using autoSNPdb', *Plant Biotechnology Journal*, 7: 326-333

- Edwards, D, Forster, J, W, Chagné, D and Batley, J (2007) 'What are SNPs?' *Association Mapping in Plants,* pg 41-52

- Eudes, F (2006) "Canadian triticale biorefinery initiative. In: *Proceedings of the 6th International Triticale Symposium*. Pp. 85-88. Stellenbosch, South Africa

- Evans, P and Haines, T (2007) 'Plant breeder's rights', *Legal Issues in Business,* 9: 3-12.

- Farooq, S and Azam, F (2002) 'Molecular markers in Plant Breeding-I: Concepts and Characterisation', *Pakistan Journal of Biological Sciences*, 5(10): 1135-1140

- Feulliet, C, Langridge, P and Waughn, R (2007) 'Cereal breeding takes a walk on the wild side', *Trends in Genetics*, 24(1): 24-32

- Food and Agriculture Organisation of the United Nations (FAO) (2011) Homepage [online] (cited November, 2011) Available from <http://www.fao.org>

- Food and Agriculture Organisation of the United Nations Statistics (FAOSTAT) (2011) [online] (cited November, 2011) Available from <http://www.faostat.fao.org/site/399/default.aspx>

- Fox, P, N, Skovmand, B, Thompson B, K, H, Braun, H-J and Cormierl, R (1990) 'Yield and adaptation of hexaploid spring triticale', *Euphytica* 47: 57-64

- Furman B, J, Qualset C, O, Skovmand B, Heaton J, H, Corke H and Wesenberg D, M (1997) 'Characterisation and analysis of North American triticale genetic resources', *Crop Science*, 37: 1951–1959

- Góral H, Wêgrzyn S, Spiss L (1999) 'Heterosis and combining ability in spring triticale (× *Triticosecale* Wittm.)', *Plant Breeding and Seed Science*, 43: 25-34.

- Góral H (2002) 'Biological-breeding aspects of utilisation of heterosis in triticale (× *Triticosecale* Wittmack)', *Zesz Nauk Akademii Rolniczej w Krakowie,* 283: 1-11

- Gray, I, C, Campbell, D, A and Spurr, N, K (2000) 'Single nucleotide polymorphisms as tools in human genetics', *Human Molecular Genetics*, 9(16): 2403-2408

- Germplasm Resources Information Network (GRIN) (http://www.ars-grin.gov; cited October, 2011).

- Gupta, P, K and Varshney, R, K (1999) 'Molecular markers for genetic fidelity during micropropagation and germplasm conservation?' *Current Science*, 76: 1308-1310.

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

- 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 *et al.,* (2002) 'Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat', *Theoretical and Applied Genetics*, 105: 413-422

- Hackauf, B and Wehling, P (2002) 'Identification of microsatellite polymorphisms in an expressed portion of the rye genome', *Plant Breeding*, 121: 17-25

- Haesaert, G and De Baets, A (1994) 'Triticale breeding: Problems and Possibilities', *Plant Breeding for Mankind*, 355: 121-134

- Huang, X and Madan, A (1999) 'CAP3: a DNA assembly program', *Genome Research*, 9: 868-877

- Huang, J, Pray, C and Rozelle, S (2002) 'Enhancing the crops to feed the poor', *Nature*, 418: 678-684

- Hulse, J, H (1976) 'Triticale', *Personal communication*

- Jewell, M, C, Frere, C, H, Prentis, P, J, Lambrides, C, J and Godwin, I, D (2010) 'Characterization and multiplexing of EST-SSR primers in *Cynodon* (Poaceae) species', *American Journal of Botany*, 97: 99-101

- Khlestkina, E, K, Than, M, H, M, Pestsova, E, G, Röder, M, S, Malyshev, S, V, Korzun, V and Börner, A (2004) 'Mapping of 99 new microsatellite-derived loci in rye (Secale cereale L.) including 39 expressed sequence tags', *Theoretical and Applied Genetics*, 109: 725-732

- Khlestkina, E, K, Tereshchenko, O, Y and Salina, E, A (2009) 'Anthocyanin biosynthesis genes location and expression in wheat-rye hybrids', *Molecular Genetics and Genomics*, 282: 475-485

- Khush, G, S and Stebbins, G, L (1961) 'Cytogenetic and evolutionary studies in *secale*. I. Some new data on the ancestry of S. *cereale*', *American Journal of Botany*, 48: 723-730

- Khush, G, S (1962) 'Cytogenetic and evolutionary studies in Secale. II. Interrelationships of wild species', *Evolution* 16 (In press).

- Khush, G, S (1963) 'Cytogenetic and Evolutionary Studies in Secale III. Cytogenetics of Weedy Ryes and Origin of Cultivated Rye', *Economic Botany*, 17(1): 60-7

- Kiss, Á, (1966) 'Neue Richtung in der Triticale-Züchtung', *Zeitschrift für Pflanzenzüchtung*, 55: 309-329

- Kiss, G, B, Csanádi, G, Kálmán, K, Kaló, P and Ökresz, L (1993) 'Construction of a basic genetic map for alfalfa using RFLP, RAPD, isozyme and morphological markers', *Molecular and General Genetics*, 238: 129-137

- Ko, H, L, Henry, R, J, Graham, G, C., Fox, G, P, Chadbone, D, A, Haak, I, C, (1994) 'Identification of cereals using the polymerase chain reaction', *Journal of Cereal Science,* 19: 101-106

- Kuleung, C, Baezinger, P, S, Kachman, SD and Dweikat, I (2006) 'Evaluating the genetic diversity of triticale with wheat and rye SSR markers', *Crop Science*, 46: 1692-1700

- Laibach, F (1925) 'Das Taubwerden von Bastardsamen und die künstliche Aufzucht früh absterbender Bastardembryonen', *Zeitschrift für Botanik*, 17: 417-459

- Leigh, F, Lea, V, Law, J, Wolters, P, Powell, W and Donini, P (2003) 'Assessment of EST- and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat', *Euphytica*, 133: 359-366

- Lindschau, M and Oehler, E (1935) 'Untersuchungen am konstant intermediären additiven Rimpau'schen Weizen-Roggenbastard', *Der Züchter*, 7: 228-233

- Liu, K, Goodman, M, Muse, S, Smith, S, Buckler, E and Doebley, J (2003) 'Genetic Structure and Diversity Among Maize Inbred Lines as Inferred From DNA Microsatellites', *Genetics*, 165: 2117-2128

- Liu, K and Muse, S, V (2005) 'PowerMarker: Intergrated environment for genetic marker data', *Bioinformatics*, 21(9): 2128-2129

- Liu, Q, Ni, Z, Peng, H, Song, W, Liu, Z and Sun, Q (2007) 'Molecular mapping of a dominant non-glaucousness gene from synthetic hexaploid wheat (*Triticum aestivum* L.)', *Euphytica*, 155: 71-78

- Lübberstedt, Melchinger, T, Dussle, A, E, Vuylsteke, C, Kuiper, M and Martin (2000) 'Relationships among Early European Maize Inbreds: IV. Genetic Diversity Revealed with AFLP Markers and Comparison with RFLP, RAPD, and Pedigree Data', *Crop Science*, Vol 40

- Ma, R, Yli-mattila, T and Pulli, S (2004) 'Phylogenetic relationships among genotypes of worldwide collection of spring and winter ryes (*Secale cereale* L.) determined by RAPD-PCR Markers', *Hereditas,* 140: 210-221

- Mackill, D, J, Nguyen, H, T and Zhang, J (1999) 'Use of molecular markers in plant improvement programs for rainfed lowland rice', *Field Crops Research*, 64: 177-185.

- Manifesto, M, M, Schlatter, A, R, Hopp, H, E, Suárez, E, Y and Dubcovsky, J (2001) 'Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers', *Crop Science*, 41: 682-690

- McDonough, C & Cooper, K, V (1998) 'New marketing strategies for triticale in southern Australia', *In Proceedings of the 4th International Triticale Symposium (Ed. P. Juskiw)*, 1: 284–289. Red Deer, Alberta, Canada.

- McGoverin, M, C, Snyders, F, Muller, N, Botes, W, Fox, G and Manley, M (2011) 'A review of triticale uses and the effect of growth environment on grain quality', *Journal of the Science of Food and Agriculture*, 91: 115-1165

- Mergoum, M, Pfeiffer, W, H, Peña, R, J, Ammar, K and Rajaram, S (2004) 'Triticale improvement: the CIMMYT programme. In: *Triticale improvement and production* (FAO plant production and protection paper 179) (edited by Mergoum, M and Gómez-Macpherson, H)', Pp. 11-26. Rome: Food and Agriculture Organisation of the United Nations

- Mergoum, M, Singh, P, K, Peña, R, J, Lazano-del Rí0, A, J, Cooper, K, V, Salmon, D, F and Gómez-Macpherson, H (2009) 'Triticale: A "new" crop with old challenges', Handbook of Plant Breeding: Cereals, Vol 3, ed. By Carena, M, J, Springer, NewYork, pp 267-290.

- Moose, S and Mumm, R (2008) 'Molecular Plant Breeding as the Foundation for 21st Century Crop Improvement', *Plant Physiology*, 147: 969-977

- Morris, M, L and Bellon, M, R (2004) 'Participatory plant breeding research: opportunities and challenges for the international crop improvement system', *Euphytica*, 136: 21-35.

- Nei, M, (1987) 'Molecular Evolutionary Genetics*, Columbia University Press*, New York

- Nei, M (1991) "Relative efficiencies of different tree-making methods for molecular data", *Phylogenetic Analysis of DNA Sequence, pp.* 90-128 edited by Miyamoto, M, M and Cracraft, J. Oxford University Press, New York.

- Oettler, G, Becker, H, C and Hoppe, G (2001) 'Heterosis for yield and other agronomic traits of winter triticale F1 and F2 hybrids', *Plant Breeding*, 120: 351-353

- Oettler, G, Burger, H, Melchinger, A, E (2003) 'Heterosis and combining ability for grain yield and other agronomic traits in winter triticale', *Plant Breeding*, 122: 318-321

- Oettler, G (2005) 'The future of a botanical curiosity - Triticale; past, present and future', *Journal of Agricultural Science*, 143: 329-346

- Persson, K and von Bothmer, R (2002) 'Genetic diversity amongst landraces of rye (*Secale cereale* L.) from Northern Europe', *Hereditas*, 136: 29-38

- Pienaar, R de.v, Littlejohn, GM., and Roux, HS (1991) '*Breeding spring triticale for South Western Cape Province*', Department of Genetics, Stellenbosch University

- Pompanon, F, Bonin, A, Bellemain, E and Taberlet, P (2005) 'Genotyping errors: causes, consequences and solutions', *Nature Reviews Genetics*, 6: 847-859

- Popping, B (2002) 'The application of biotechnological methods in authenticity testing', *Journal of Biotechnology*, 98: 107-112

- Rafalski, A (2002) 'Applications of single nucleotide polymorphisms in crop genetics', *Plant Biology*, 5: 94-100

- Ram, S, G, Parthiban, K, T, Kumar, R, S, Thiruvengadam, V and Paramathma, M (2008) 'Genetic diversity among Jatropha species as revealed by RAPD markers', *Genetic Resources of Crop Evolution*, 55: 803-809

- Röder, M, S, Plaschke, J, Konig, S, U and Sorrells, M, S (1995) 'Abundance, variability and chromosomal location of microsatellite in wheat,' *Molecular and General Genetics*, 246: 327-333

- Röder, MS, Korzun, V, Wendehake, K, Plaschke, J, Tixier, MH, Leroy, P and Ganal, MW (1998) 'A microsatellite map of wheat', *Genetics,* 149: 2007-2023.

- Royo, C, Soler, C and Ramogosa, I (1995) 'Agronomical and morphological differentiation among winter and spring triticale,' *Plant Breeding*, 114: 413-416

- Rözen and Skaletsky, (2000) 'Primer3 on the www for general users and for biologist programmers', *Methods in Molecular Biology*, 132: 365-86

- Saal, B and Wricke, G (1999) 'Development of simple sequence repeat markers in rye (*Secale cereale* L.)', *Genome*, 42: 964-972

- Sachidanandam, R, Weissman, D, Schmidt, S, C, Kakol, J, M, Stein, L, D, Marth, G, Sherry, S, Mullikin, J, C, Mortimore, B, J and Willey, D, L (2001) 'A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms,' *Nature*, 409: 928-933

- Saitou, N and Nei, M (1987) 'The neighbour-joining method; a new method for reconstructing phylogenetic trees', *Molecular Biology and Evolution*, 4: 406-42

- Salmon, D, F, Mergoum, M and Gómez-Macpherson (2004) "Triticale production and management. In: *Triticale improvement and production* (FAO plant production and protection paper 179) (edited by Mergoum, M and Gómez-Macpherson, H), 27-36. Rome: Food and Agriculture Organisation of the United Nations

- Sánchez-Monge, E (1958) 'Hexaploid triticale. In: *Proceedings of the First International Wheat Genetics Symposium*', Winnipeg, Manitoba, Canada (Edited by B, C Jenkins), Pp. 181-194. University of Manitoba, Winnipeg, Canada.

- Santiveri, F, Royo, C and Romagosa (2002) 'Patterns of grain filling of spring and winter hexaploid triticales', *European Journal of Agronomy*, 16: 219-230

- Schiemann, E and Nürnberg-Krüger, U (1952) 'Neue Untersuchungen an *Secale* africanum Stapf. II. Secale africanum und seine Bastarde mit Secale montanum und *Secale cereale*', *Naturwissenshaften*, 39: 136-137

- Schwarz, G, Michalek, W, Mohler, V, Wenzel, G and Jahoor, A (1999) 'Chromosome landing at the *Mla* locus in barley (*Hordeum vulgare* L.) by means of high-resolution mapping with AFLP markers', *Theoretical and Applied Genetics*, 98: 521-530

- Shah, S, C and Kusiak, A (2004) 'Data mining and genetic algorithm based gene/SNP selection', *Artificial Intelligence in Medicine*, 31: 183-196

- Shahinnia, T and Sayed-Tabatabaei, E, B (2009) 'Conversion of barley SNPs into PCR-based markers using dCAPS method', *Genetics and Molecular Biology*, 32(3): 564-567

- Shang, H-Y, Wei1,Y-M, Wang, X-R and Zheng, Y-L (2006) 'Genetic diversity and phylogenetic relationships in the rye genus *Secale* L. (rye) based on Secale cereale microsatellite markers', *Genetics and Molecular Biology*, 29(4): 685-691

- Schubert, A, C (2009) '*Quantitative trait loci analysis in spring wheat comparing two advanced backcross populations derived from an exotic wheat accession*', University of Bonn. (PhD dissertation)

- Singh, N, K, Shepherd, K, W and McIntosh, R, A (1990) 'Linkage mapping of genes for resistance to leaf, stem and stripe rusts and ω-secalins on the short arm of rye chromosome 1R,' *Theoretical and Applied Genetics*, 80: 609-616

- Smith J, S, C, Goodman M, M and Stuber C, W (1985) 'Genetic variability within U.S. maize germplasm: I. Historically important lines,' *Crop Science*, 25: 551-555

- Sneath, P, H, A and Sokal, R, R (1973) "Numerical Taxonomy", W, H, Freeman, San Franciso.

- Soleimani, V, D, Baum, B, R and Johnson, D, A (2003) 'Efficient Validation of Single Nucleotide Polymorphisms in Plants by Allele-Specific PCR, With an Example From Barley', *Plant Molecular Biology Reporter,* 21: 281-288

- Somers D, J, Kirkpatrick, R, Moniwa, M and Walsh, A (2003) 'Mining single nucleotide polymorphisms from hexaploid wheat ESTs', *Genome*, 49: 431-437

- Talbert, L, E, Smith, L, Y and Blake, N, K (1998) 'More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA', *Genome*, 41: 402-407

- Tams, SH, Bauer, E, Oettler, G and Melchinger, AE (2004) 'Genetic diversity in European winter triticale determined with SSR markers and coancestry coefficient', *Theoretical and Applied Genetics*, 108:1385-1391

- Tams, S, H, Melchinger, A, E and Bauer, E (2005) 'Genetic similarity among European winter triticale Elite germplasms assessed with AFLP and comparisons with SSR and pedigree data', *Plant Breeding*, 124: 154-160

- Tams, S, H, Bauer, Oettler, G, Melchinger, A, E and Schön, C-C (2006) 'Prospects for hybrid breeding in winter triticale: II. Relationship between parental genetic distance and specific combining ability', *Plant Breeding,* 125: 331-336

- Tams, S, H (2006) 'Molecular and agronomic assessment of genetic diversity and hybrid breeding in triticale', Stuttgart: University of Hohenheim. (PhD dissertation)

- Tamura, K, Nei, M and Kumar, S (2004) "Prospects for inferring very large phylogenies by using the neighbour-joining method", *Proceedings of the National Academy of Sciences*, 101: 11030-11035

- Tamura, K, Peterson, D, Peterson, N, Stecher, G, Nei, M, and Kumar, S (2011) 'MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods', *Molecular Biology and Evolution* (submitted)

- Tang, Z, X, Ross, K, Ren, Z, L, Yang, Z, J, Zhang, H, Y, Chikmawati, T, Miftahudin and Gustafson, J, P (2011) '*Secale*', *Wild Crop Relatives Genomic and Breeding Resources*, 367-396

- Tanksley, S (1993) 'Mapping polygenes', *Annual Review of Genetics*, 27: 205-233

- Tanksley, S and McCouch, S (1997) 'Seed banks and molecular maps: Unlocking genetic potential from the wild', *Science,* 277: 1063-1066

- Teemu Kuulasmaa, OligoAnalyzer (version 1.0.2), Homepage [online]. Available from <URL:http://www.uku.fi/~kuulasma/OligoSoftware>

- Terzi, V, Morcia, C, Gorrini, A, Stanca, A, M, Shewry, P, R and Faccioli, P (2005) 'DNA-based methods for identification and quantification of small grain cereal mixtures and fingerprinting of varieties', *Journal of Cereal Science*, 41: 213-220

- Thiemt, E, M and Oettler, G (2008) "Agronomic performance of anther-derived doubled haploid and single seed descent lines in crosses between primary and secondary winter triticale", *Plant Breeding*, 127: 476-479

- Trethowan, R, M, Reynolds, M, Sayre, K and Ortiz-Monasterio, I (2005) 'Adapting wheat cultivars to resource conserving farming practices and human nutritional needs', *Annual Review of Applied Biology*, 146: 405-413

- UPOV (International Convention for the Protection of New Varieties of Plants), Homepage [online] Available from <http://www.upov.int/>

- Vaillancourt, A, Nkongolo, K, K, Michael, P and Mehes, M (2007) 'Identification, characterization, and chromosome locations of rye and wheat specific ISSR and SCAR markers useful for breeding', *Euphytica,* 159: 297-306

- Varshney, R, K, Thiel, T, Stein, N, Langridge, P and Graner, A (2002) 'In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species', *Cellular and Molecular Biology Letters*, 7: 537-546

- Varshney, R, K, Beier, U, Khlestkina, E, K, Kota, R, Korzun, V, Graner, A and Börner, A (2007) 'Single nucleotide polymorphisms in rye (*Secale cereale* L.): discovery, frequency, and applications for genome mapping and diversity studies', *Theoretical and Applied Genetics*, 114: 1105-1116

- Varughese, G, Pfeiffer, W, H and Peña, R, J (1996) "Triticale: a successful alternative crop (Part 2)", *Cereal Foods World*, 41: 635-645

- Vyhnánek, T, Nevrtalova, E, and Slezakova, K (2009) "Detection of the genetic variability of triticale using wheat and rye SSR markers", *Cereal Research Communication*, *37*: 23-29

- Weir, B, S, (1996) '*Genetic data analysis II*', MA: Sinauer Associates, Incorporated, Sunderland

- Weising, K, Nybom, H, Wolff, K and Kahl G (Ed) (2005) *DNA fingerprinting in plants: principles, methods and applications* (2nd edition), CRC Press, Taylor & Francis Group, Boca Raton, USA

- Welsh, J and McClelland, M (1990) 'Fingerprinting genomes using PCR with arbitrary primers', *Nucleic Acids Research*, 18:7213-7218

- Williams, J, G, 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

- Yan, J, Yang, X, Shah, T, Sánchez-Villeda, H, Li, J, Warburton, M, Zhou, Y, Crouch, J, H and Xu, Y (2009) 'High-throughput SNP genotyping with the GoldenGate assay of maize', *Molecular Breeding*, 25:441-451

- Yashitola, J, Thirumurugan, T, Sundaram, R, M, Naseerullah, M, K, Ramesha, M, S, Sarma, N, P and Sonti, R, V (2002) 'Assessment of purity of rice hybrids using microsatellite and STS markers,' *Crop Science*, 42: 1369-1373

- Zhou, R, Zhu, Z, Kong, X, Huo, N, Tian, Q, Li, P, Jin, C, Dong, Y and Jia, J (2005) 'Development of wheat near-isogenic lines for powdery mildew resistance', *Theoretical and Applied Genetics*, 110: 640-648

- Zillinsky, F, J (1974) "The development of triticale", *Advances in Agronomy*, 26: 315-348

Chapter 2:   **Evaluation of the suitability of AFLP markers for genetic diversity assessment in triticale breeding**

Evaluation of the suitability of AFLP markers for genetic diversity assessment in triticale breeding

D. Bitalo

Stellenbosch University Plant Breeding Laboratory, Department of Genetics, Private Bag X1, Matieland 7602, South Africa

Abstract

The amplified fragment length polymorphism (AFLP) marker system is a PCR-based technique that is based on the selective amplification of restriction fragments. AFLPs are recommended as a suitably efficient marker system for the analysis of genetic studies in plant breeding because of their ability to score numerous loci without any prior knowledge of the genome under investigation.

The genetic diversity of commercially viable and Elite triticale cultivars was assessed using molecular data generated by 42 AFLP primer combinations (18 *Eco*RI/*Mse*I, 24 *Pst*I/*Mse*I).

Up to 61% of the *Eco*RI/*Mse*I primer combinations amplified across the cultivars in comparison to 67% of the *Pst*I/*Mse*I primer combinations that generated amplicons. The *Pst*I/*Mse*I primer combinations produced more bands (534) than the *Eco*RI/*Mse*I primer combinations (324).

Polymorphism information content (PIC) values generated by both AFLP primer combinations ranged from 0.11 to 0.39 for *Eco*RI/*Mse*I and from 0.20 to 0.27 for *Pst*I/*Mse*I primers respectively, indicative of a low to moderate genetic variation detected among the cultivars. The *Pst*I/*Mse*I primers had a slightly higher average Marker Index (MI) value than the *Eco*RI/*Mse*I primers but both primer combinations detected similar levels of average polymorphisms.

Subsets of seven AFLP primer combinations (3 EcoRI/MseI and 4 PstI/MseI primers) were selected and were used to create neighbour-joining (NJ) and unweighted pair group method with arithmetic average (UPGMA) dendograms for the cultivars. The molecular data generated by these primers showed differences between the rye and triticale cultivars but the relationships displayed among the triticale cultivars didn't corroborate available pedigree data.

Keywords: Triticale, rye, Amplified fragment length polymorphism (AFLP), *Mse*I, *Eco*RI, *Pst*I, polyacrylamide (PAGE) gels, Polymorphism information content (PIC), genetic diversity.

Introduction

The amplified fragment length polymorphism (AFLP) technique was first implemented by Vos *et al*. (1995), and was based on the selective amplification of restriction fragments of a total digest of genomic DNA. The technique follows three steps; template DNA restriction by two different restriction enzymes (rare cutters e.g. *Mse*I and frequent cutters e.g. *Eco*RI) and ligation of oligonucleotide adapters to the ends of the digestion fragments, selective amplification of the restriction fragments using preselective primers (primers that recognise the adapter sequences) (Vos *et al.,* 1995) and finally gel electrophoresis-based analysis of the amplified fragments. The AFLP technique produces fingerprints of any organism in spite of its source and without any previous knowledge of the genome (Agarwal *et al.,* 2008). A limited set of generic primers is used to optimise the number of fragments detected in one reaction (Vos *et al*., 1995; Schwarz *et al*., 2000).

The AFLP technique displays several advantages for the analysis of complex genomes (Liu *et al.,* 2009) and these include among others (Vos *et al.,* 1995, Schwarz *et al.,* 2000, Tams *et al.,* 2005);

- The generation of large numbers of anonymous and randomly distributed molecular markers in a single reaction;
- Generally a high level of polymorphism; and
- Robustness and reliability due to the rigorous reaction conditions used for primer annealing as a result of combining the RFLP technique with the power of PCR.

AFLP analysis employs the use of oligonucleotides as adapters and primers, as described by Vos *et al.,* (1995). The adapters used consist of core and enzyme-specific sequences; examples of which are *Eco*RI adapters, *Mse*I-adapters and *Pst*I-adapters. On the other hand, AFLP primers consist of core (CORE), enzyme-specific (ENZ) and selective extension (EXT) sequences. The principle of the technique is based on the amplification of subsets of genomic restriction fragments using PCR (Vos *et al.,* 1995).

Firstly enzymes restrict DNA and double-stranded (ds) adapters are ligated to the restricted DNA fragments to generate template DNA for amplification. The sequences of the adapters and the adjacent restriction sites serve as primer binding sites for the later amplification of the restriction fragments. The selective nucleotides at the 3'-ends of the primers can therefore only prime DNA synthesis from the restriction sites, and only those restriction fragments that have nucleotides flanking the restriction site, while matching the selective nucleotides, will be amplified (Vos *et al.,* 1995).

According to Vos *et al,* 1995, the objectives of using two restriction enzymes during the restriction/ ligation phase of the AFLP technique are:

- The number of fragments amplified are reduced by using the rare cutter as the number of selective nucleotides needed for selective amplification are limited in the rare/frequent cutters;
- Frequent cutter generates small DNA fragments which amplify satisfactorily and are in the most favourable size range for separation on polyacrylamide gels;
- 'Doublets' on gels due to uneven mobility of the two strands of amplified products are reduced with the use of two restriction enzymes which make it possible to label one strand of the ds PCR products;
- The number of fragments to be amplified are regulated by the use of two restriction enzymes;
- Moderation of the number of fragments amplified is made flexible by use of two restriction enzymes; and
- Generation of large numbers of different fingerprints is made possible with the use of a low number of primers which generate various primer combinations.

AFLP bands can be detected by different techniques including silver-stained polyacrylamide (PAGE) gels, autoradiography and semi-automated means. Conventional detection by autoradiography incorporates the labelling of one of the primers with a radiolabel such as [γ-^{33}P] ATP. However, this technique has a low throughput as only one sample per lane is analysed, allele-typing is not accurate, can be time-consuming and prone to errors.

When using gels, both selective primers (frequent and rare) are kept unlabelled, and bands are detected on a silver-stained denaturing PAGE gel. This technique, though the cheapest of the three, shares limitations similar to detection by autoradiography. The use of semi-automated

methods offers a much better means to analyse AFLP fragments as it incorporates the use of fluorescence dyes that are much safer and easier to distinguish, exhibits a higher efficiency, has an internal size standard with accurate allele size calling and allows for the analysis of more than one sample in a single lane (Schwarz *et al.,* 2000).

AFLP bands are either scored as present or absent as each polymorphic fragment is scored as a locus with two allelic classes (Manifesto *et al.,* 2001) but this can sometimes be problematic especially when using conventional means of detection as the AFLP technique can generate between 50 and 100 DNA fragments. This is generally seen in the linear correlation between numbers of amplified fragments and genome size. Fingerprints of complex genomes consist predominantly of unique AFLP fragments, but are characterised by the presence of small numbers of more intense repeated fragments (Vos *et al.,* 1995)

Previous studies on the assessment of the genomes of the parental species of triticale were predominantly done using random amplified length polymorphism (RFLP) markers (Bohn *et al.,* 1999; Korzun *et al.,* 2001) and random amplification of polymorphic DNA (RAPDs) (Gupta *et al.,* 1999; Masojć *et al.,* 2001) however, these didn't produce the high levels of band discrimination on PAGE gels, or dense coverage of the genome as AFLP markers do.

Despite the fact that AFLP markers display dominance, they come highly recommended as a suitably efficient marker system for the analysis of genetic studies in plant breeding programmes due to the numerous loci scored by each primer combination (Powell *et al.,* 1996). Therefore, AFLP markers were chosen for this study especially considering that the genotypes under study were expected to be closely related when considering the available pedigree information.

Plant breeding programmes rely on maintaining a high level of genetic diversity (Collard and Mackill, 2008) in order to attain successful progress from selection. To achieve this, the genetic base of the breeding material requires an efficient technique to identify the diverse genotypes for hybridisation with the Elite cultivars and/or each other (Collard and Mackill, 2008). A considerable number of studies have managed to saturate both the parental genomes of triticale with AFLP markers (Korzun *et al.* 2001; Ma *et al.* 2001; Hackauf and Wehling, 2002; Saal and Wricke, 2002; Vieira *et al.*, 2007) in a less laborious manner than if the conventionally favoured RFLPs had been used. For instance, the average coverage of the rye

map density increased from 4.0 to 2.9 cM and the number of marker loci from 282 to 480 (Bednarek, *et al.,* 2003).

Although the use of AFLPs and other DNA-based marker systems in plant breeding offer such advantages, there seems to be tentativeness in their application. According to Collard and Mackill, 2008, this could be instigated by the following reasons;

- Limited number of markers and few detections of polymorphisms within breeding material;
- High costs incurred when using the advanced marker techniques;
- Insufficient association between data generated by markers and the QTLs;
- The "application gap" between plant breeders and science institutes; and
- The "knowledge gap" between molecular biologists, plant breeders and other disciplines.

Therefore, this study aimed to screen AFLP primer combinations to assess the genetic diversity amongst Elite and commercially available cultivars sourced from the SU-PBL triticale breeding programme taking in to consideration cost and other feasibility factors. In order to reach this aim the following objectives have been identified:

- Screen AFLP primer combinations; and
- Compare the data obtained with that obtained from assessments done using SSR markers.

Materials and Methods

Plant material

A set of eight Elite and commercially released cultivars of spring triticale ('US2007', 'US2008' (marketed as 'AgBeacon'), 'US2009', 'US2010', 'Bacchus', 'Tobie', 'Ibis' and 'Rex'), two rye entries ('Duiker' and 'Henoch', a landrace) and a wheat cultivar, 'SST88' (Sensako, South Africa), were sourced from the SU-PBL at Welgevallen, and grown in 3 litre pots in the greenhouse for two weeks. The leaves from each plant were used to extract genomic DNA (gDNA).

Genomic DNA extraction

DNA was extracted using a modified protocol described by Doyle and Doyle, 1990. About 0.1g of leaf tissue was cut into separate 2.2 ml microcentrifuge tubes together with three stainless steel bearings. Approximately 800μl of 2% (w/v) CTAB [5M NaCl, 0.5M EDTA (pH 8), 1M Tris-Cl (pH 8)] and 1.6 μl 0.2% (v/v) of β-Mercaptoethanol (β-ME) were added. The leaf tissue was ground using a Qiagen Tissuelyser (company), (three sessions of 90 sec at 30Hz). The tubes were placed in a 60°C water bath for one hour.

A mixture of 800μl Chloroform: Isoamyl alcohol (C: I) made to a 24:1 ratio, was added to the incubated gDNA. Tubes were then centrifuged at 12000rpm for 10 min. Clear supernatants were transferred to clean tubes and 0.5 volume of phenol and 0.5 volume of C:I were added. Centrifugation proceeded at 12000rpm for 5 min. An equal volume of C:I was added to the extracted aqueous phase and the mixture centrifuged at 12000rpm for 5 min. The final supernatant was incubated overnight in an equal volume of isopropanol at -20°C.

Subsequent to incubation, samples were centrifuged at 12000rpm, 4°C, for 10 min and the supernatant was discarded cautiously. A millilitre of ice cold 70% (v/v) ethanol was added to the pellets in the tube and the tubes were centrifuged at 12000rpm, 4 ^{o}C, for 10 min. Again the supernatant was discarded with care, without discarding the pellets. The air-dried pellets were then dissolved in 50 μl of TE [1M Tris-Cl, 0.5M EDTA (pH 8)] containing 40 μg/ml RNaseA and placed in a 37°C water bath for 30 minutes.

Approximately one tenth volume of sodium acetate [3M (pH 5)] and 2.5 volume 100% (v/v) ice cold ethanol were added and the tubes were centrifuged at 12000rpm, 4 ^{o}C, for 10 min. Further discarding of the resulting supernatants was made and the pellets were washed twice with 1 ml of ice cold 70% (v/v) ethanol. A final 5 min centrifugation phase at 12000rpm, 4 ^{o}C, was made. The supernatant was again discarded and the pellet was left to air dry, after which it was dissolved in 40 μl distilled (SABAX) water.

The extracted genomic DNA was quantified using the NanoDrop® ND-1000 Spectrophotometer following the Thermo SCIENTIFIC user's manual guide. Dilutions to 100ng/ μl were made and samples were stored at 4 ^{o}C. All stock samples were stored at -20 ^{o}C till further use.

58

AFLP marker evaluation

The AFLP assays followed a four-step protocol as previously described in chapter 2. A revised version of the protocol described by Kahn *et al.,* (2000) was used and the amplification of restriction fragments with selective primers was done using cycling conditions described by Vos *et al.,* (1995).

The restriction-ligation reaction was performed using approximately 300 ng of gDNA that were subjected to restriction-ligation in a single 20µl step containing; 5 units of *Eco*RI (Roche) and 5 units of *Mse*I (New England Biolabs) restriction enzymes or 5 units of *Pst*I (Fermentas) and 5 units of *Mse*I (New England Biolabs), equal volumes of *Eco*RI (5'-CTC GTA GAC TGC GTA CC-3') (1µl) and *Mse*I (5'-GAC GAT GAG TCC TGA G-3'/ 3'-TA CTC AGG ACT CAT-5') (1µl) adaptors, or *Mse*I (1µl) and *Pst*I (5'- (1µl) adaptors, 1X One-Phor-All buffer (USB), 1mM ATP (Pharmacia Biotech), 0.1µg/µl of Bovine Serum Albumin (BSA) (New England Biolabs), 1 unit of T4 DNA ligase (USB) and ddH$_2$O. The contents in the digestion-ligation reaction tube were then mixed by gentle vortexing, centrifuged briefly and incubated overnight at room temperature. Restriction and ligation were confirmed by electrophoresing 5µl of restriction-ligation DNA on a 1.5% agarose gel at 100V for 45 minutes. The remaining 15µl of the reaction was diluted with 1X TE$_{0.1}$ buffer (10mM Tris-HCl, 0.1mM NNA$_2$EDTA, pH 8) in a 1:9 ratio.

The amplification of preselective primers was done in a thin-walled PCR tube to a total volume of 20 µl per sample containing 5.2µl of the diluted restriction-ligation DNA, 1X PCR NH$_4$ buffer (Bioline [16mM (NH4)2SO4, 67mM Tris-HCl (pH 8.8) at 25°C, 0.01% (v/v) Tween-20]), 200µM of each dNTP (Bioline), 1.5mM of MgCl$_2$ (Bioline), 1 unit of BIOTAQ™ DNA polymerase (Bioline), 1.5µl of preselective primer mix [E00 primer (5'-GAC TGC GTA CCA ATT C-3')/M00 primer (5'-GAT GAG TCC TGA GTA A-3') or P01 primer (5'-GAC TGC GTA CAT GCA GA-3')/M01 (5'-GAT GAG TCC TGA GTA AA-3')] and AFLP-grade ddH$_2$O. The PCR cycling profile was as follows: 5 min elongation at 72 ℃, 30 cycles of 30 seconds at 94 ℃, 56 ^{o}C for 1 min, 1 min at 72 ℃ and a final elongation of 5 min at 72 ℃. 5µl of the preamplified PCR products were separated on a 1.5% agarose gel at 100V for 45 mins to confirm amplification. A 1:9 dilution was then made using the remaining 15µl of the amplification products using 1X TE$_{0.1}$ buffer. These diluted samples were then stored at 4°C until needed.

The selective amplification of preamplified products was done to a total volume of 10µl and contained; 2.5µl of the diluted preselective DNA, 1X PCR NH_4 buffer (Bioline [16mM (NH4)2SO4, 67mM Tris-HCl (pH 8.8) at 25°C, 0.01% (v/v) Tween-20]), 0.2mM of each dNTP (Bioline), 0.5µM of the *Mse*I selective primer (5'-GAT GAG TCC TGA GTA ANN N-3', where N conveys the specific selective nucleotides), 0.05µM of *Eco*RI selective primer (5'-GAC TGC GTA CCA ATT CNN N-3', Table 2.1) or *Pst*I selective primer (5'-GAC TGC GTA CAT GCA GAN NN-3', Table 2.2), 1.5mM $MgCl_2$ (Bioline), 0.25 units of BIOTAQ™ DNA polymerase (Bioline) and ddH_2O. The PCR profile was partitioned into two sections; a 13 cycle reaction entailing 30 s of 94 $^{\circ}$C, 30 s at an annealing temperature of 65 $^{\circ}$C, 1 min at 72 $^{\circ}$C with a 0.7 $^{\circ}$C drop in annealing temperature with each cycle for 12 cycles, and a 23 cycle reaction entailing 30 s at 94°C, 30 s at 56°C, 1 min at 72°C and, a final soak temperature at 4°C. The amplified samples were stored at 4°C until needed for electrophoresis.

Agarose gels were prepared and run following the Sambrook *et al.,* (1991) protocol. DNA fragments from restriction digest reactions in AFLP assays were separated on 1.5% (w/v) agarose gels in 1X TBE buffer (Tris-HCl, EDTA, Boric acid) and run at 100V for 45 minutes. A 6X loading buffer (Ficoll Orange G) was used and the gels were stained with ethidium bromide (0.5g/ml) at the end of the run.

Table 2.1: *Eco*RI/*Mse*I primer combinations used in the study (adopted from Groenewald *et al.,* 2005).

*Mse*I primers	*Eco*RI primers		
	*Eco*RI-ACA	*Eco*RI-AAC	*Eco*RI-AGG
*Mse*I-CAG	1	2	3
*Mse*I-CAT	4	5	6
*Mse*I-CTC	7	8	9
MseI-CAC	10	11	12
*Mse*I-CAA	13	14	15
*Mse*I-CTA	16	17	18

Table 2.2: *Pst*I/*Mse*I primer combinations used in the study (adopted from Manifesto *et al.,* 2001).

*Mse*I primers	*Pst*I primers			
	*Pst*I-ACC	*Pst*I-AGC	*Pst*I-AAA	*Pst*I-AGG
*Mse*I-CAC	19	20	21	22
*Mse*I-CAA	23	24	25	26
*Mse*I-CTA	27	28	29	30
*Mse*I-CAG	31	32	33	34
*Mse*I-CAT	35	36	37	38
*Mse*I-CTC	39	40	41	42

Amplification products were separated on a 6% (w/v) denaturing polyacryalmide gel (acrylamide: bisacrylamide, 19:1) containing 6M urea and 1X TBE buffer (Tris-HCl, EDTA, Boric acid). 15µl of each sample (10µl PCR product: 10µl loading buffer [98% formamide, 10mM Na_2EDTA (pH 8), 0.05% (w/v) bromophenolblue, 0.05% (w/v) xylene cyanol FF] were loaded following 5 min of denaturing at 95°C and immediate quenching on ice. Samples along with a 100bp ladder (Promega, Madison, Wisconsin, USA) were separated by electrophoresis in a 1X TBE buffer at 60W for 4 hours. Band fragments were revealed by following a modified silver staining procedure described by Tixier *et al.,* 1997; (1) 20 min in fixing solution (0.5% (v/v) acetic acid, 10% (v/v) ethanol) (2) two 5 min rinses in distilled water (3) 20 min in 0.1% (w/v) silver nitrate (4) 10 s rinse in distilled water (5) 10-15 min in developing solution (1.5% (w/v) sodium hydroxide, 0.16% (v/v) formaldehyde).

Gels were then rinsed in distilled water and images recorded using a Nikon digital camera on a light-table.

Data analysis

The resulting DNA fragments were scored as either present (1) or absent (0) at each locus representing the genetic information of each AFLP primer combination.

The created haplotypes were used to compute the frequency-based distances between cultivars using the CSChord (Cavalli-Sforza and Edwards, 1967) method; Weir's (1996) gene diversity and polymorphism information content (PIC) values were computed using PowerMarker v3.25 (Liu and Muse, 2005). Due to the scoring methodology of the band fragments generated at each locus, PIC values expected at each locus cannot exceed 0.5 (Manifesto *et al.,* 2001). Therefore average PIC values for each primer combination were computed and were labelled PIC (Tams *et al.,* 2005)

The marker index (MI) of each primer combination was also calculated as;

$$MI = PIC \; X \; n\text{ß}$$

Where n is the total number of loci detected and ß is the proportion of polymorphic bands (Tams *et al.,* 2005).

The frequency-based distances were used to generate unweighted pair group method with arithmetic average (UPGMA) (Sneath and Sokal, 1973) and neighbour-joining (NJ) (Saitou and Nei, 1987) trees to illustrate the relationships among the cultivars. All trees were imported into MEGA v5.05 (Tamura *et al.,* 2011) from PowerMarker v3.25 (Liu and Muse, 2005).

Results and discussion

Levels of polymorphisms revealed by EcoRI/MseI primer combinations

A total of 18 *Eco*RI/*Mse*I primer combinations were screened for polymorphisms in the controls. Primer combinations were run in duplicate on PAGE gels and out of 18 combinations, only 11 combinations (61%) showed repeatability. Band fragments were scored in the regions ranging from 100bp to 500 bp; where markers could be explicitly scored. A statistical summary of the detected loci was computed and the average PIC values across each loci of each primer combination were computed (Table 2.2). The lowest average PIC value noted was 0.11 using the primer combination *Mse*I-CAG/*Eco*RI-ACA (rxn1) and the highest was 0.39 using the primer combinations *Mse*I-CAG/*Eco*RI-AAC (rxn2), *Mse*I-

CAC/*Eco*RI-ACA (rxn10) and *Mse*I-CAC/*Eco*RI-AGG (rxn12) respectively. Based on the PIC values generated by the band fragments at each locus of each primer combination, the percentage level of polymorphism was calculated. This was done by computing the ratio between the number of loci exhibiting a PIC value ≥0.25 across the cultivars and the total number of loci generated by each primer combination. Primer combinations rxn10, rxn11 and rxn12 were selected as having the most polymorphic loci (≥50% polymorphism) and the ability to exhibit genetic diversity among the control lines.

An average MI value of 3.09 was obtained. Primer combination rxn1 recorded the lowest MI at 0.34 and rxn12 yielded at the highest at 8.58.

Only the frequency-based distances generated by the three primer combinations noted as being most polymorphic were used in the construction of both the NJ and UPGMA dendograms to show the relationships among the lines.

Levels of polymorphisms revealed by PstI/MseI primer combinations

A total of 24 *Pst*I/*Mse*I primer combinations were screened for polymorphisms in the Elite cultivars. Primer combinations were run in duplicate on PAGE gels and out of 24 combinations, only 18 were repeatable in the 100 to 500 bp region. The polymorphism levels of each primer combination were noted and are listed in Table 2.2.

Up to 534 bands were detected in 16 of the 24 primer combinations, with numbers ranging from 6 to 52 bands per marker clearly scored.

The lowest recorded average PIC value was 0.20 generated by primer combination *Mse*I-CTA/*Pst*I-AGG (rxn30), which also had the lowest observed polymorphism 12%. The highest PIC was noted at 0.27 and was generated by primer combinations *Mse*I-CAA/*Pst*I-AGC (rxn24), *Mse*I-CAG/*Pst*I-AGC (rxn32) and *Mse*I-CTC/*Pst*I-AGC (rxn40). However, rxn24 generated only 45% polymorphic band fragments which didn't qualify it for the cluster analysis. MI values averaged at 3.65 and ranged from 0.52 to 7.33.

The frequency-based distances of the primer combinations that generated 50% or more polymorphic bands and had a PIC value ≥0.25 were used in the generation of NJ and UPGMA dendograms and cluster analysis. These primer combinations comprised of rxn24, rxn32, rxn34 and rxn40.

Table 2.3: Levels of polymorphisms detected by 11 EcoRI/MseI and 16 *Pst*I/*Mse*I primer combinations.

PC	Av. Genetic Diversity	PIC	Total No. Loci	Polymorphic loci	% Polymorphism	MI
rxn1	0.13	0.11	19	3	16	0.34
rxn2	0.39	0.31	12	8	67	2.49
rxn3	0.25	0.21	29	7	24	1.47
rxn5	0.20	0.17	27	4	15	0.68
rxn6	0.25	0.21	32	12	38	2.48
rxn7	0.28	0.23	49	19	39	4.32
rxn8	0.26	0.21	22	10	45	2.14
rxn9	0.23	0.19	29	4	14	0.76
rxn10	0.39	0.31	34	23	68	7.02
rxn11	0.33	0.27	30	14	47	3.77
rxn12	0.39	0.31	41	28	68	8.58
rxn22	0.32	0.26	29	12	41	3.15
rxn23	0.32	0.26	30	14	47	3.64
rxn24	0.33	0.27	31	14	45	3.74
rxn25	0.32	0.26	25	13	52	3.41
rxn26	0.31	0.26	32	14	44	3.58
rxn 28	0.33	0.26	29	14	48	3.70
rxn 29	0.31	0.26	6	2	33	0.52
rxn 30	0.23	0.20	25	3	12	0.61
rxn 31	0.29	0.24	52	22	42	5.32
rxn32	0.34	0.27	52	27	52	7.33
rxn34	0.33	0.26	47	24	51	6.29
rxn35	0.29	0.24	37	12	32	2.90
rxn36	0.29	0.24	38	13	34	3.10
rxn40	0.34	0.27	34	19	56	5.15
rxn41	0.29	0.24	31	10	32	2.41
rxn42	0.31	0.25	31	14	45	3.54

PC primer combination, **rxn** Reaction, **Av**. Average, *PIC* Average polymorphism information content, **MI** marker index

Primer combinations highlighted in grey were chosen for the construction of the UPGMA dendogram.

Cluster analysis using frequency-based distances

Using the CSChord (1967) distance method, frequency-based distances were generated for the chosen 3 *Eco*RI/MseI and 4 *Pst*I/MseI primer combinations. These distances are listed in Table 2.4 and show the genetic variability between the cultivars. The highest distance of 0.90 was observed between the triticale lines and the wheat and rye lines expressing the expected genetic differences among these lines. However, the two rye cultivars 'Duiker' and 'Henoch' demonstrated an extensive variation in genetic distance (0.83). This was not expected as 'Duiker' is a cultivar that was selected from the landrace 'Henoch'.

The CSChord distances between the triticale cultivars 'Tobie', 'Bacchus', 'Ibis' and 'Rex' range from 0.67 to 0.73. This is indicative of a relatively high genetic variation among these cultivars implying that the selected AFLP primer combinations are able to make distinction among the triticale cultivars.

The frequency-based distances in Table 2.4 were then used to generate UPGMA and NJ dendograms (figures 3.1 and 3.2) bootstrapped to 1000 permutations. These trees were imported into MEGA v5.05 (Tamura *et al.,* 2011) from PowerMarker v3.25 (Liu and Muse, 2005) and were used to infer the relationships amongst the cultivars.

Table 2.4: Frequency-based distances computed using the CSChord, 1967 distance method

OTU	Bacchus	Duiker	Henoch	Ibis	Rex	SST	Tobie	US2007	US2008	US2009	US2010
Bacchus	0.00	0.90	0.90	0.73	0.73	0.90	0.67	0.87	0.90	0.87	0.80
Duiker	0.90	0.00	0.83	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90
Henoch	0.90	0.83	0.00	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90
Ibis	0.73	0.90	0.90	0.00	0.67	0.90	0.70	0.90	0.90	0.90	0.83
Rex	0.73	0.90	0.90	0.67	0.00	0.90	0.70	0.90	0.90	0.90	0.83
SST	0.90	0.90	0.90	0.90	0.90	0.00	0.90	0.90	0.90	0.90	0.90
Tobie	0.67	0.90	0.90	0.70	0.70	0.90	0.00	0.87	0.87	0.87	0.80
US2007	0.87	0.90	0.90	0.90	0.90	0.90	0.87	0.00	0.87	0.83	0.90
US2008	0.90	0.90	0.90	0.90	0.90	0.90	0.87	0.87	0.00	0.77	0.83
US2009	0.87	0.90	0.90	0.90	0.90	0.90	0.87	0.83	0.77	0.00	0.80
US2010	0.80	0.90	0.90	0.83	0.83	0.90	0.80	0.90	0.83	0.80	0.00

OTU Observed taxonomic unit

The UPGMA dendogram generated was bootstrapped to 1000 permutations and a consensus tree with an SBL (smallest length value) of 3.31 was recorded and the consensus value was 70%. Most of the bootstrap values generated were below 70% indicating that the generated phylogenies were not well supported.

The NJ dendogram generated exhibited an SBL value of 3.34 with a consensus value of 70%. However, as is noted in the UPGMA dendogram generated, most of the bootstrap values are below 70% and therefore the dendogram is not well supported.

Both UPGMA and NJ dendograms generated using data generated by the seven AFLP primer combinations, showed clear differences among the Elite and commercially viable spring triticale cultivars but the relationship fingerprints generated failed to corroborate available pedigree data.

Despite the fact that the AFLP markers used in this study were sensitive to minor genetic differences, generated adequate polymorphic markers from a just a few primer combinations and were cost efficient, the phylogenetic trees generated using data from these markers failed to infer the expected relationships. There a couple of reasons that could explain this;

When inferring phylogenetic relationships in species of higher taxonomic levels, the high variability of AFLP markers can reduce the similarities between distant taxa to the level of chance (Mueller and Wolfenbarger, 1999). Therefore, this limits the utility of this marker system in a genetic diversity study;

The largely dominant nature of these markers provided less powerful genetic data;

The polymorphic AFLP markers were generated at random and their distribution across the genome was not investigated. Hence, very many markers were used to generate data and this lowered the precision of the estimation of genetic distances among the taxa leading to an inaccurate inference of phylogenetic relationships.

Therefore, it would be best to use map-based AFLP markers in order to minimise the number of markers genotyped and to infer more precise phylogenetic relationships (Vuylsteke *et al.,* 2000).

Figure 2.1: A UPGMA dendogram showing the grouping of eight spring triticale cultivars based on CSChord (1967) frequency-based distances generated by seven AFLP primer combinations

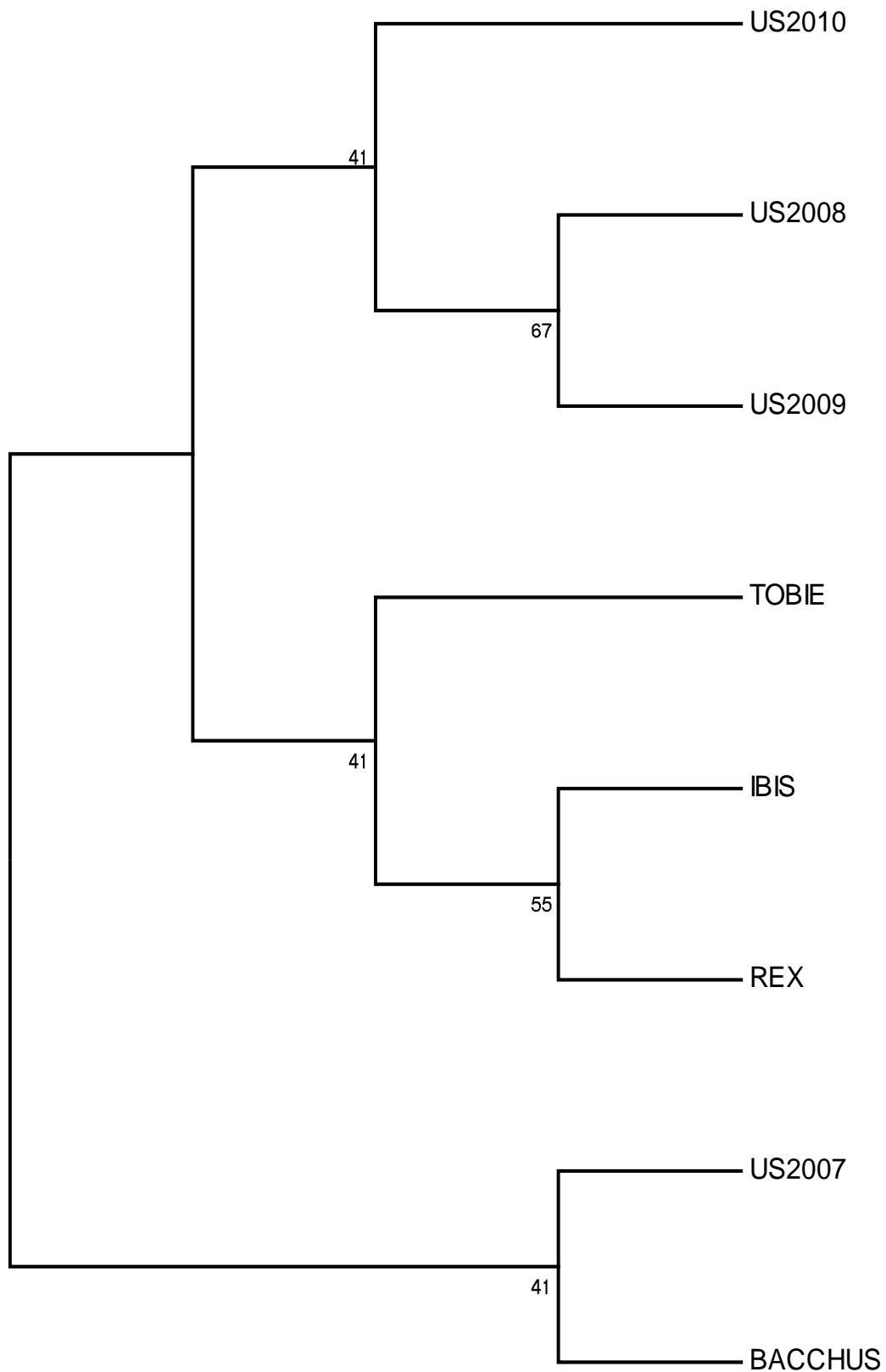**Figure 2.2**: A NJ dendogram showing the grouping of eight spring triticale cultivars based on CSChord (1967) frequency-based distances generated by seven AFLP primer combinations
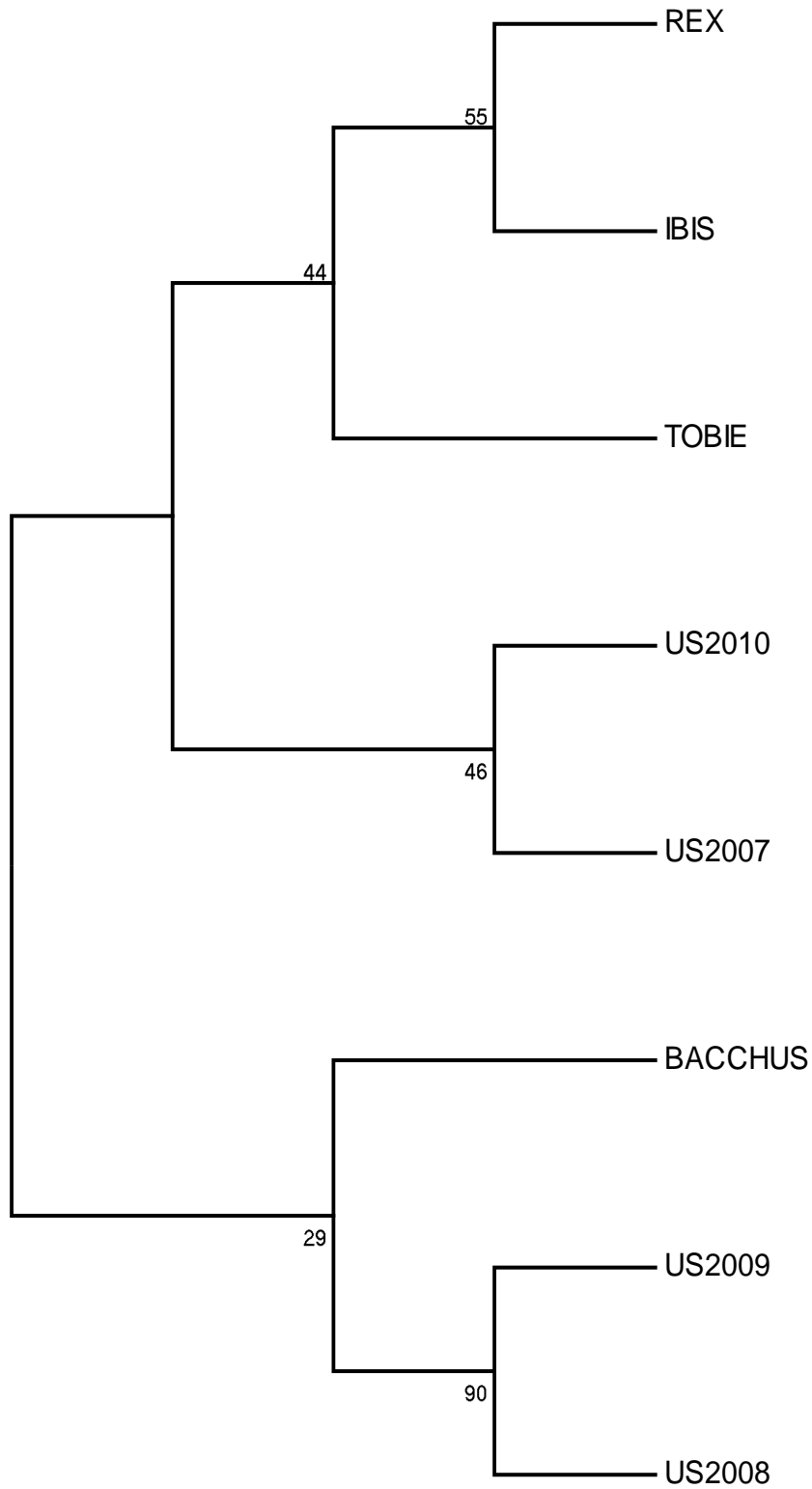
Contrast between EcoRI/MseI and PstI/MseI primer combinations

Based on the average MI values recorded, the *Pst*I/*Mse*I primer combinations show a slightly higher level of MI than the *Eco*RI/*Mse*I primer combinations (3.65 versus 3.09). This slight variation in MI values could be biased due to the fact that all 16 *Pst*I/*Mse*I primer combinations generated band fragments that were easy to score unambiguously as opposed to the band fragments generated by the 11 *Eco*RI/*Mse*I primer combinations scored.

Each *Eco*RI/*Mse*I primer combination exhibited an average of 40.09% level of polymorphism and each *Pst*I/*Mse*I primer combination averaged a 41.63% level of polymorphism. Based on the average PIC values, *Eco*RI/*Mse*I primer combinations generated an average value of 0.23 that is close to the 0.25 averaged by *Pst*I/*Mse*I primer combinations. Therefore, both primer combinations detect a moderate genetic variability among the control cultivars.

Additionally, the PIC values detected among the control cultivars vary slightly from PIC values obtained by Manifesto *et al.,* 2001 when using the same *Pst*I/*Mse*I primer combinations among Argentine bread wheat cultivars. In their study, Manifesto *et al.,* 2001 obtained PIC values that ranged from 0.26 to 0.38 with an average of 0.30 as opposed to the 0.24 to 0.27 PIC range and 0.25 average values, obtained in this study.

In a study by Chikmawati *et al.* (2005) when analysing the phylogenetic relationships among 29 accessions of rye using 18 *Eco*RI/*Mse*I and *Pst*I/*Mse*I primer combinations, the average percentage polymorphism detected among the rye accessions by *Pst*I/*Mse*I primer combinations was 16% higher than that detected by the *Eco*RI/*Mse*I primer combinations. This difference in polymorphism percentages is much higher than the 1.53% difference recorded between the *Pst*I/*Mse*I markers and the *Eco*RI/*Mse*I markers in this study.

These findings could suggest that the AFLP primer combinations used in this study may show varying results perharps due to the clustering of some AFLP primer combinations in specific regions of the genome and also due to the species under study. Many other studies in plant species have shown that *Eco*RI/*Mse*I primer combinations mainly detect tightly linked markers. Up to 80% of AFLPs detected with *Eco*RI/*Mse*I primer combinations in the mapping of the tomato genome, tended to cluster around the centromere of the chromosome. Only about 16% of the AFLPs were detected with *Pst*I/*Mse*I primer combinations in the same region (Haanstra *et al.,* 1999). This similar type of clustering was observed in barley (Qi *et al.,* 1998) and maize (Vuylsteke *et al.,* 1999).

Thus, it is very likely that the *Eco*RI and *Mse*I selective primers tend to cluster in regions that are AT-rich and these AT-rich sequence blocks are recognised by the *Eco*RI/*Mse*I primer combinations in the pericentric regions of the chromosomes in connection with suppressed recombination. This begs the question as to whether to opt for the *Pst*I/*Mse*I primer combinations over *Eco*RI/*Mse*I primer combinations. Although it may seem tempting, the substantial levels of methylation at symmetric CpXpG sites in higher plants have to be kept in consideration as these result in fewer bands being produced (Saal and Wricke, 2002).

Overall, the *Pst*I/*Mse*I markers have the ability to provide additional genetic information in the less covered distal regions of the rye genome (Saal and Wricke, 2002) in genetic studies of both rye and triticale. However, the *Eco*RI/*Mse*I primer combinations should not be ruled out altogether even if clustering of AFLPs isn't beneficial for many genetic studies among rye and triticale, nonetheless, this clustering does present a valuable tool for rapid saturation in regions of the genome that lack RFLP, RAPD or SSR markers (Bednarek *et al.,* 2003).

Since there was not much difference in the PIC values detected by both primer combinations among the material studied, it is preferable to continue implementing both in the assessment of the genetic diversity within the SU-PBL in order to ensure a relatively extensive coverage of the genome.

The selected polymorphic AFLP primer combinations (3*Eco*RI/*Mse*I, 4 *Pst*I/*Mse*I) have managed to detect moderate differences among the closely related triticale cultivars and the rye cultivars. The markers also managed to show some differences between certain cultivars, but unfortunately this information does not corroborate available pedigree data.

Conclusion

The AFLP primer combinations used during the study managed to amplify so many loci in triticale, which were difficult to score manually and hence the generated data failed to substantiate the available pedigree information. Therefore, this marker system could not be utilised in a high throughput analysis.

However, running a pilot study such as this that involved the assessment of Elite cultivars could open avenues to overcome some of the impediments faced by both the rye and triticale breeding programs in terms of molecular marker implementation.

Marker techniques need to be modified to increase the sensitivity and resolution to detect genetic distinctiveness and diversity. So it is highly recommended that the AFLP marker systems used in this study should implement different forms of detection e.g. the use of fluorescent primers for automated fragment analysis as opposed to PAGE gels, or the detection of amplified AFLP fragments using an automated ALF DNA sequencer with the fragment option (Huang and Sun, 1999) in order to reduce the problems encountered during practical handling, data generation and analysis.

References:

- Agarwal, M, Shrivastava, N and Padh, H (2008) 'Advances in molecular marker techniques and their applications in plant sciences', *Plant Cell Reports*, 27: 617-631

- Bednarek, P, T, Masojc, P, Lewandowska, R and Mysków, B (2003) 'Saturating rye genetic map with amplified fragment length polymorphism (AFLP) and random amplified polymorphic DNA (RAPD) markers', *Journal of applied genetics*, 44(1): 21-33

- Bohn, M, Utz, H, F and Melchinger, A, E (1999) 'Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs and SSRs and their use for predicting progeny variance', *Crop Science*, 39: 228–237

- Cavalli-Sforza, L, L and Edwards, A, W, F (1967) 'Phylogenetic analysis: models and estimation procedures', *American Journal of Human Genetics*, 19: 233-257

- Chikmawati, T, Skovmand, B and Gustafson, J, P (2005) 'Phylogenetic relationships among Secale species revealed by amplified fragment length polymorphisms', *Genome*, 48: 792-801

- Collard, B, C, Y and Mackill, D, J (2008) 'Marker assisted selection: an approach for precision plant breeding in the twenty-first century', *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363: 557-572

- Doyle, J, J and Doyle, J, L (1990) 'Isolation of plant DNA from fresh tissue', *Focus* 12: 13-15

- Groenewald, J, Z, Fourie, M, Marais, A, S, Marais, G, F (2005) 'Extension and use of a physical map of the Thinopyrum-derived Lr19 Translocation', *Theoretical and Applied Genetics*, 112: 131-138

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

- Haanstra, J, P, Wye, W, C, Verbakel, H, Meijer-Dekens, F, van denBerg, P, Odinot, P, van Heusden, A, W, Tanksley, S, Lindhout, P and Peleman, J (1999) 'An integrated high-density RFLP–AFLP map of tomato based on two *Lycopersicon esculentum* X L. *pennellii* F2 populations', *Theoretical and Applied Genetics*, 99: 254-271

- Hackauf, B and Wehling, P (2002) 'Identification of microsatellite polymorphisms in an expressed portion of the rye genome', *Plant Breeding*, 121: 17-25

- Huang, J and Sun, M (1999) 'A modified fluorescence-labelled primers and automated DNA sequencer detection for efficient fingerprinting analysis in plants', *Biotechnology techniques*, 13: 277-278

- Khan, J, Procunier, D, Humphreys, G, Tranquilli, G, Schlatter, A, R, Marcucci Poltri, S, Joppa, L, R, Frohberg, R and Dubcovsky, J (2000) 'Development of PCR based markers for a high grain protein content gene from *Triticum turgidum* ssp. *dicoccoides* transferred to bread wheat', Crop Science, 40: 518-524

- Korzun, V, Malyshev, S, Voylokov, A, V and Börner, A (2001) 'A genetic map of rye (*Secale cereale* L.) combining RFLP, isozyme, protein, microsatellite and gene loci', *Theoretical and Applied Genetics*, 102: 709-717

- Liu, K and Muse, S, V (2005) 'PowerMarker: Intergrated environment for genetic marker data', *Bioinformatics*, 21(9): 2128-2129

- Ma, X-F, Wanous, M, K, Houchins, K, Rodriguezmilla, M, A, Goicoechea P, G, Wang, Z, Xie M and Gustafson, J, P (2001) 'Molecular linkage mapping in rye (*Secale cereale* L.)', *Theoretical and Applied Genetics*, 102: 517-523

- Manifesto, M, M, Schlatter, A, R, Hopp, H, E, Suárez, E, Y and Dubcovsky, J (2001) 'Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers', *Crop Science*, 41: 682-690

- Masojć, P, Myoeków, B and Milczarski, P, (2001) 'Extending a RFLP-based genetic map of rye using random amplified polymorphic DNA (RAPD) and isozyme markers', *Theoretical and Applied Genetics*, 102: 1273-1279

- Mueller, U, G and Wolfenbarger, L, L (1999) "AFLP genotyping and fingerprinting" *Tree,* 14: 389-393

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

- Qi, X, Stam, P and Lindhout, P (1998) 'Use of locus-specific AFLP markers to construct a high-density molecular map in barley', *Theoretical and Applied Genetics*, 96: 376-384

- Saal, B and Wricke, G (1999) 'Development of simple sequence repeat markers in rye (*Secale cereale* L.)', *Genome*, 42: 964-972

- Saal, B and Wricke, G (2002) 'Clustering of amplified fragment length polymorphism markers in a linkage map of rye', *Plant Breeding*, 121: 117-123

- Saitou, N and Nei, M (1987) "The neighbour-joining method; a new method for reconstructing phylogenetic trees", *Molecular Biology and Evolution*, 4: 406-425

- Sambrook, J, Fritsch, EF and Maniatis, T (1991) '*Molecular cloning: A laboratory manual*. Trends in Biotechnology', 9: 213-214

- Schwarz, G, Michalek, W, Mohler, V, Wenzel, G and Jahoor, A (1999) 'Chromosome landing at the *Mla* locus in barley (*Hordeum vulgare* L.) by means of high-resolution mapping with AFLP markers', *Theoretical and Applied Genetics*, 98: 521-530

- Schwarz, G, Herz, M, Huang, X, Q, Michalek, W, Jahoor, A, Wenzel, G and Mohler, V (2000) 'Application of fluorescence-based semi-automated AFLP analysis in barley and wheat', *Theoretical and Applied Genetics*, 100: 545-551

- Sneath, P, H, A and Sokal, R, R (1973) "Numerical Taxonomy", W, H, Freeman, San Franciso

- Tams, S, H, Melchinger, A, E and Bauer, E (2005) 'Genetic similarity among European winter triticale Elite germplasms assessed with AFLP and comparisons with SSR and pedigree data', *Plant Breeding*, 124: 154-160

- Tamura, K, Peterson, D, Peterson, N, Stecher, G, Nei, M, and Kumar, S (2011) 'MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods', *Molecular Biology and Evolution* (submitted)

- Teemu Kuulasmaa, OligoAnalyzer (version 1.0.2), Homepage [online]. Available from <URL:http://www.uku.fi/~kuulasma/OligoSoftware>

- Tixier, M, H, Sourdille, P, Röder, M, Leroy, P and Bernard M (1997) "Detection of wheat microsatellites using a non-radioactive silver nitrate staining method", *Journal of Genetics and Breeding*, 51: 175-177

- Vieira, E, A et al., (2007) 'Association between genetic distances in wheat (*Triticum aestivum* L.) as estimated by AFLP and morphological markers, *Genetics and Molecular Biology*, 30 (2): 392-399

- Vos, P et al., (1995) 'AFLP: a new technique for DNA fingerprinting', *Nucleic Acids Research*, 23(21): 4407-4414

- Vuylsteke, M *et al.,* (1999) 'Two high-density AFLP linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers', *Theoretical and Applied Genetics*, 99: 921-935

- Vuylsteke, M, Mank, R, Brugmans, B, Stam, P and Kuiper, M (2000) "Further characterization of AFLP data as a tool in genetic diversity assessments among maize (*Zea mays* L.) inbred lines", *Molecular Breeding,* 6: 265-276

- Weir, B, S, (1996) '*Genetic data analysis II*', MA: Sinauer Associates, Incorporated, Sunderland

Chapter 3:   Development and evaluation of the suitability of SSR markers for genetic diversity assessment in triticale breeding

Development and evaluation of the suitability of SSR markers for genetic diversity assessment in triticale breeding

D. Bitalo

Stellenbosch University Plant Breeding Laboratory, Department of Genetics, Private Bag X1, Matieland 7602, South Africa

Abstract

Simple sequence repeat (SSR) markers show a relatively good transferability between closely related species. Therefore, this study focused on using SSR markers developed in wheat and rye to measure the genetic diversity among Elite triticale entries.

A set of 29 EST and genomic SSR primer pairs developed in rye were used to test the genetic variability among eight triticale cultivars and two rye cultivars. Only 23 of the primer sets produced amplicons among the Elite entries expressing a 79.3% transferability of these SSRs to triticale.

Fifty-seven EST and genomic SSR markers were used to analyse the genetic variation expressed by the wheat (A and B) genomes in the triticale entries. The SSR markers tested achieved transferability levels of 84% across the wheat genomes in the Elite triticale entries, which imply that these markers could serve as a diagnostic tool for the genetic assessment of triticale.

Therefore, two EST R-genome SSRs were selected to create relationship fingerprints for the Elite entries, which corresponded with available pedigree information. Furthermore, these primer sets were selected to test the semi-automated DNA sequence detection technique.

A subset of five wheat SSRs (two EST and three genomic) were also chosen for their high discriminatory ability and were allocated to test the semi-automated DNA sequencer detection technique.

Keywords: Triticale, wheat, rye, simple sequence repeat (SSR) markers, Polymorphism information content (PIC), genetic diversity.

Introduction

The distribution and organisation of microsatellites in a number of plant genomes have been studied using fluorescence *in situ* hybridization (FISH) and in-gel hybridization techniques (Gupta *et al.,* 1999). Röder *et al.* (1993) for instance, using FISH analysis showed that there was a regular occurrence of $(GAA)_n$ repeats in barley and other *Triticeae* species and Gupta *et al.* (1999) using in-gel hybridization showed that there were numerous microsatellites in long stretches of repeated DNA of bread wheat.

However, later studies showed that FISH and in-gel hybridization techniques only detected the extended regions of the repetitive DNA and did not focus on the regions of the DNA uniquely prone to containing microsatellite motifs (Gupta *et al.,* 1999). However, on the basis of genetic and physical mapping techniques, short regions of repetitive DNA were amplified using PCR amplification.

Genetic maps of wheat constructed using SSRs showed that the microsatellites were not clustered in particular regions but were evenly distributed in the different regions of the wheat chromosomes (Röder *et al.,* 1998). The average distances between two SSRs for wheat were recorded as 704 kb between two $(AC)_n$ repeat motifs and 440 kb between two $(GA)_n$ repeat motifs (Röder *et al.,* 1995).

Rye has been noted to be a vital genetic source for wheat improvement programs due to its agronomical traits; the resistance genes *Pm8* to powdery mildew (*Erysiphe graminis* f. sp. *Tritici*), *Lr26* to leaf rust (*Puccinia recondite* f. sp. *Tritici*), *Sr31* to stem rust (*Puccinia graminis* f. sp. *Tritici*) and *Yr9* to yellow rust (*Puccinia stiiformis* f. sp. *Tritici*) which map on chromosome 1RS (Singh *et al.,* 1990), contributed to the success of wheat cultivars bearing the 1BL.1RS translocation.

These and other traits of rye that cannot be analysed in other related crop species, necessitate a more detailed analysis of the rye genome using microsatellite markers in order to recognise economically important regions with molecular markers (Hackauf and Wehling, 2002). These maps are also necessary in triticale breeding programmes as rye is a constituent of the synthesised hybrid (Khlestkina *et al.,* 2004). The pioneer linkage maps of rye were based on RFLP markers (Devos *et al.,* 1993; Senft and Wricke, 1996), and later improved by combining morphological genes and selected anchor probes of rye and other cereals like wheat and barley (Korzun *et al.,* 2001) to saturate the rye linkage maps.

78

Due to the high level of reproducibility, robustness and ease of high throughput analysis of SSRs (Hackauf and Wehling, 2002), they have become the favoured marker choice for mapping genes in rye. However rye SSR databases are not as well covered as those of other cereal genomes like wheat (Röder *et al.,* 1998) or barley (Ramsey *et al.,* 2000).

Efforts to saturate rye linkage maps with SSR markers includes studies done by Hackauf and Wehling (2002) that used public sequencing resources to access rye cDNA sequences from anthers, cold-stressed leaves and aluminium-stressed roots and used these as a resource for SSR marker development. One other effort was the utilisation of mapped wheat and barley microsatellites which may cross-amplify in rye. Khlestkina *et al.,* (2004) searched 8,930 EST sequences deposited in Genbank <http://www.ncbi.nlm.nih.gov> for rye microsatellite sequences with di-, tri- and tetra-nucleotide repeats and also utilised wheat microsatellite markers (WMS) of the Getersleben collection (Röder et al., 1998) to saturate the rye genomic map. Such efforts opened the gateway to screening the R genome in triticale using SSR markers.

The wheat genomic and EST databases have accumulated plenty sequences of SSR markers that were previously shown to be successful in assessing the genetic diversity in wheat and were also reported to exhibit transferability to triticale. Triticale shares the A and B genomes with wheat and the wheat genomic and EST-SSRs usually detect a single locus in each of those genomes (Schubert, 2009). These databases are publicly available via domains like the Wheat Microsatellite Consortium (WMC) primer sequences published online at <http://wheat.pw.usda.gov/ggpages/SSR/WMC/>, and the BARC markers that were developed for the United States Wheat and Barley Scab Initiative to map and characterise genes for fusarium resistance (Song *et al.,* 2002; 2005). Other database sources have a large deposit of ESTs developed for wheat (Qi *et al.,* 2004) and the EST-SSRs can be developed from the *in silico* analysis of these databases and can improve gene-based maps and marker assisted selection (MAS) studies (Zhang *et al.,* 2005). In wheat, *in silico* analysis of the EST-SSR bread wheat database <http://www.ncbi.nlm.nih.gov/dbEST> showed a presence of one EST-SSR every 6.2 kb (Varshney *et al.,* 2002).

Previous studies have successfully implemented SSRs developed in both wheat and rye to assess the genetic diversity among triticale.

Tams *et al.,* (2004; 2005) assessed the genetic diversity of winter triticale using SSRs from wheat and rye and concluded that these SSRs provide a powerful tool in the study of genetic diversity among triticale lines.

Kuleung *et al.,* (2006) used 43 gSSRs developed in wheat, and 14 gSSRs and EST-SSRs developed in rye, to assess the genetic diversity and relationship of 80 hexaploid triticale accessions pooled from global breeding programs. Cluster analysis grouped the accessions in to five groups that corresponded with the pedigree information, country of origin, growth habits and release year, and this was attributed to the exchange of material between breeding programs.

At the SU-PBL, results of work done using wheat SSRs was used to analyse 40 wheat genotypes sourced from five breeding programs, including SU-PBL. Data generated by these markers managed to group the genotypes into seven clusters and managed to accurately identify different cultivars (Honing, 2007).

This study aims to extend the use of these SSR markers and others developed, in the wheat and rye genomes, to evaluate the genetic diversity among triticale cultivars of an Elite repute, to identify polymorphic SSRs, and to optimise a cost effective, high throughput technique that implements these polymorphic markers in the day-to-day typing of the breeding material.

Materials and methods

Plant material

Spring triticale cultivars ('US2007', 'US2008' (AgBeacon), 'US2009', 'US2010', 'Bacchus', 'Tobie', 'Ibis' and 'Rex'), two rye cultivars ('Duiker' and 'Henoch') sourced from the SU-PBL, were evaluated in this study. A commercial wheat cultivar 'SST88' (Sensako, South Africa), was also evaluated.

The breeding codes for these cultivars are not availed in this paper due to the protection of intellectual property (IP).

Genomic DNA extraction

DNA was extracted following a modified CTAB protocol by Doyle and Doyle, (1990) as described in chapter 3 of this document.

All extracted DNA was then quantified and its quality checked using the NanoDrop® ND-1000 Spectrophotometer following the Thermo SCIENTIFIC user's manual guide. Each DNA sample was then diluted to 100ng/ μl and stored at 4 ^{o}C while all stock DNA was stored at -20 ^{o}C until needed.

SSR marker evaluation

Fifty-seven SSR markers developed in wheat (Table 6.1 shown in the Appendix) were screened for polymorphisms in this study. Röder *et al.* (1995; 1998) developed SSR markers that were genomic specific and mapped to the A, B and D genomes of hexaploid wheat. Clones containing microsatellite repeats GA and GT were purified from genomic phage library fragments and primers were designed for these sequenced clones and designated Xgwm for "Getersleben Wheat Microsatellite." Ma *et al.,* (1996) screened two wheat (*Triticum aestivum* L.) genomic libraries for di-, tri-, and tetranucleotide tandem repeats. Clones containing $(AC)_n$, $(AG)_n$, $(TCT)_n$ and $(TTG)_n$ repeats were isolated and sequenced. Primers flanking these repeat sequences were designed using the software MACVECTOR (Kodak, New Haven, Conn.) and were designated Xcnl. Other genomic specific SSRs developed for the wheat genome include:

- Xwmc primers which were supplied to all members of the WMC (Wheat Microsatellite Consortium) for genotyping and characterisation (Rampling *et al.,* 2001; Gupta *et al.,* 2002; 2005);
- Xcfa (Pierre Sourdille microsatellite) primers (Sourdille *et al.,* 2001); and
- Xbarc (Beltsville Agriculture Research Center) and Xcfd (Pierre Sourdille microsatellite) (Somers *et al.,* 2004; Kong *et al.,* 2005).

Zhang *et al.,* (2005; 2007) screened over 46,000 EST contigs to identify microsatellites. The microsatellites were defined as sequences having a minimum of three repeats (with each repeat motif containing 1 to 6 nucleotides) and a total length of no less than 12 nucleotides. Primers spanning these microsatellites were then designed using PRIMER software (version 0.5, Whitehead Institute for Biomedical Research, Cambridge, Mass). Primers with an optimal length (18 to 22 bp), an optimal product size (100 to 400 bp), melting temperatures

81

ranging from 57°C and 63°C and other factors were selected and these primer sequences subjected to a BLAST analysis to evade redundancy. These primers were designated cfe and each forward primer was M13-tailed with the sequence 5'-CACGACGTTGTAAAACGAC-3'.

Rye EST sequences deposited in GenBank (http://www.ncbi.nlm.nih.gov) were screened for microsatellite sequences with di, tri and tetra-nucleotide repeats. Primers flanking these repeats were designed using Primer3.0 Input (version 0.4.0) software (Rözen and Skaletsky, 2000) and designated rye expressed microsatellite sites (REMS) (Khlestkina *et al.,* 2004) and *Secale cereale* microsatellite (SCM) (Hackauf and Wehling, 2002).

An 18bp sequence (5'-TGT AAA ACG ACG GCC AGT-3') identical to the sequence of an M13 primer, was added to the 5'-end of each forward primer as described by Hackauf and Wehling (2002) and Khlestkina *et al.* (2004). Oligo Analyzer version 1.0.2 (Teemu Kuulasmaa, 2000) was employed to analyse the designed primers for self-annealing. An additional group of SSRs developed by Saal and Wricke, (1999) and labelled "SCM" were also screened.

These primers were designed from the flanking sequences of (GT/CA)$_n$ and (CT/GA)$_n$ repeats in the genomic library using the software OLIGO (Rychlik and Rhoads, 1989), choosing primers between 18 and 26 bp in length and annealing temperatures between 50°C and 65°C. All the 29 EST and genomic SSRs developed in rye are listed in Table 6.2 (Appendix).

PCR protocol

PCR reactions were run with a Thermal Cycler 2720 (Applied Biosystems) with a total reaction volume of 20µl following conditions as described in Röder *et al.,* (1998); 1 X Bioline NH$_4$ buffer [16mM (NH4)2SO4, 67mM Tris-HCl (pH 8.8) at 25°C, 0.01% (v/v) Tween-20], 0.2mM of each dNTP, 0.5µM of each primer, 1.5mM MgCl$_2$, 1U *Taq* polymerase (Bioline BIOTAC^{TM}) and 300ng template DNA. Thermal cycling conditions were (1) 94°C for 3 min (2) 94°C for 1 min (45 cycles) (3) annealing temperature for 1 min (4) 72°C for 2 min (4) 72°C for 10 min.

PCR primers designed for the wheat genome were used in the PCR amplification of wheat and triticale cultivars while those designed for the rye genome were used in the PCR amplification of rye and triticale cultivars.

DNA separation and analysis

Amplification products were separated on a 6% (w/v) denaturing polyacrylamide gel (acrylamide: bisacrylamide, 19:1) containing 6M urea and 1X TBE buffer (Tris-HCl, EDTA, Boric acid). 15µl of each sample (10µl PCR product: 10µl loading buffer [98% formamide, 10mM Na_2EDTA (pH 8), 0.05% (w/v) bromophenolblue, 0.05% (w/v) xylene cyanol FF] were loaded after 5 min of denaturing at 95°C and immediate quenching on ice. Samples along with a 50bp or 100bp ladder were separated by electrophoresis in a 1X TBE buffer at 60W for 4 hours. Band fragments were revealed by following silver staining procedures described by Tixier *et al.,* 1997; (1) 20 min in fixing solution (0.5% (v/v) acetic acid, 10% (v/v) ethanol) (2) two 5 min rinses in distilled water (3) 20 min in 0.1% (w/v) silver nitrate (4) 10 s rinse in distilled water (5) 10-15 min in developing solution (1.5% (w/v) sodium hydroxide, 0.16% (v/v) formaldehyde).

Gels were then rinsed in distilled water and digitally recorded using a Nikon digital camera and a light-table.

Data analysis

The resulting band fragments were scored as alleles and their sizes determined using the 50bp or 100bp ladder (Promega, Madison, Wisconsin, USA). Repeats of SSRs were scored as haplotypes with each locus representing the genetic information of each SSR marker. The scored band sizes were used to compute the frequency-based distances between cultivars using the CSChord (Cavalli-Sforza and Edwards, 1967) method, Weir's (1996) gene diversity and its alternative polymorphism information content (PIC) were calculated using PowerMarker v3.25 (Liu and Muse, 2005) to indicate the ability of each marker to detect polymorphisms in the study population.

The PIC (Botstein *et al.,* 1980) values were calculated in PowerMarker v 3.25 (Liu and Muse, 2005).

The scored fragments were also used in a cluster analysis using both the unweighted pair group method with arithmetic average (UPGMA) (Sneath and Sokal, 1973) and Neighbour-joining (NJ) (Saitou and Nei, 1987) trees to show the relationships among the cultivars.

Results and discussion

Of the 57 wheat microsatellite primer pairs tested, 48 markers (35 located on the A genome chromosomes and 13 located on the B genome chromosomes) produced amplicons in the nine accessions (eight spring triticale and one wheat). These results were summarised in Table 3.1. The 48 markers were chosen after excluding those markers that showed ambiguity and no cross-amplification. Overall 84% of the wheat markers amplified microsatellite loci in triticale, suggesting a good transferability of these wheat markers to triticale.

A total of 197 alleles were detected with these primer sets, located on 12 chromosomes with six chromosomes from each of the genomes as is shown in Table 3.1. The average number of alleles per locus was four with the minimum number detected at microsatellite loci Xbarc1021, Xgwm285 and Xwmc25. In the A genome, 136 alleles were detected that varied from the 61 alleles detected in the B genome. This shows a difference in the amount of allelic variation contributed by each genome in triticale. However, the average alleles detected per locus between the two genomes didn't vary significantly; five in the B genome and four in the A genome. This expresses a slight difference in the allelic variation detected in these two genomes in triticale.

PIC and gene diversity values were calculated to better comprehend the genetic differentiation among the screened lines. These values are listed in Table 3.1. Not much difference was noted in the PIC values between the A genome (0.5) and B genome (0.59) implying that both genomes express relatively the same genetic variation in triticale.

Table 3.1: A statistical summary generated in PowerMarker v.3.25

Genome	Marker	M.A.F	Allele No	Gene diversity	PIC
A-genome	cfe2-6A,6B,6D	0.11	9	0.89	0.88
	cfe22-4A	0.67	3	0.49	0.44
	cfe34-4BL,5A,5B	0.78	3	0.37	0.34
	cfe37-5A,5B	0.78	3	0.37	0.34
	cfe53-2AL	0.22	8	0.86	0.85
	cfe110-4A,4BL	0.44	4	0.67	0.61
	cfe125-6A,6B	0.44	6	0.74	0.71
	cfe132-6A	0.56	5	0.64	0.61
	cfe166-7A,7B,7D	0.56	3	0.57	0.49
	cfe260-7A,7B,7D	0.78	3	0.37	0.34
	cfe270-4D,5A,5B	0.56	4	0.62	0.57
	cfe284-7A	0.78	3	0.37	0.34
	Xbarc12-3A	0.44	3	0.64	0.57
	Xbarc19-3A	0.89	2	0.20	0.18
	Xbarc37-6A	0.89	2	0.20	0.18
	Xbarc206-6A	0.78	3	0.37	0.34
	Xbarc263-1AS	0.56	4	0.62	0.57
	Xbarc1021-3A	1.00	1	0.00	0.00
	Xcfa2019-7AL	0.56	4	0.62	0.57
	Xcfa2153-1A	0.78	3	0.37	0.34
	Xcfd58-1A,1D	0.22	5	0.79	0.76
	Xcfd79-3A	0.22	6	0.81	0.79
	Xgwm2-3A	0.44	5	0.72	0.68
	Xgwm5-3A	0.89	2	0.20	0.18
	Xgwm135-1A	0.33	5	0.77	0.73
	Xgwm136-1A	0.67	3	0.49	0.44
	Xgwm160-4A	0.44	6	0.74	0.71
	Xgwm357-1A	0.67	2	0.44	0.35
	Xgwm369-3A	0.22	7	0.84	0.82
	Xgwm666-1A	0.67	3	0.49	0.44
	Xwmc59-1A	0.56	3	0.57	0.49
	Xwmc167-2A	0.56	4	0.62	0.57
	Xwmc169-3A	0.89	2	0.20	0.18
	Xwmc254-1A	0.67	3	0.49	0.44
	Xwmc256-6A	0.56	4	0.62	0.57

Table 3.1 continued...

Genome	Marker	M.A.F	Allele No	Gene Diversity	PIC
B-genome	cfe274-4BL	0.56	5	0.64	0.61
	Xcn13-6BS	0.33	6	0.79	0.76
	Xgwm18-1B,4B	0.44	5	0.72	0.68
	Xgwm46-7BS	0.22	7	0.84	0.82
	Xgwm108-3B	0.22	6	0.81	0.79
	Xgwm165-4BL/S	0.78	3	0.37	0.34
	Xgwm285-3B	0.89	2	0.20	0.18
	Xgwm340-3B	0.44	5	0.72	0.68
	Xgwm369-3A	0.22	7	0.84	0.82
	Xgwm389-3B	0.56	4	0.62	0.57
	Xgwm429-2B	0.33	6	0.79	0.76
	Xgwm493-3B	0.44	4	0.67	0.61
	Xgwm550-1B	0.33	6	0.79	0.76
	Xwmc25-2B	0.89	2	0.20	0.18

M.A.F Major allele frequency, No**.** Number, **PIC** Polymorphism information content

Primers highlighted in grey were chosen for the construction of the UPGMA dendogram.

A subset of five markers (cfe2, cfe53, Xcfd79, Xgwm108, Xgwm369) covering chromosomes 2AL, 3A, 6A, 3B and 6B were selected to calculate the frequency-based distances (Table 3.2) amongst the cultivars and consequently generate UPGMA and NJ dendograms. These markers were chosen based on the following criteria;

- High PIC value (≥0.8);
- Repeatable and distinct banding pattern; and
- Ability to detect differences among cultivars compared to other SSRs.

Out of the 29 rye microsatellite markers tested, 23 (9 ESTs and 14 genomic) cross-amplified showing a 79.3% transferability of these markers to triticale. This percentage far surpasses the 39% transferability of 28 rye microsatellites among 80 global accession of triticale assessed by Kuleung *et al.* (2006).

Approximately 20.7% of the microsatellites were excluded due to their inability to amplify within the rye genome or due to the ambiguity of the band fragments produced on the PAGE

gels. According to Leonova *et al.* (2005), such band fragments may not even contain any SSR sequences.

A sum total of 102 alleles were detected spanning over chromosomes 1R, 2R, 3R, 5R, 6R and 7R. Primer set SCM80 detected only one allele while SCM120 detected eight alleles and an average of four alleles was noted. These results are summarised in Table 3.3. Furthermore, the computed major allele frequencies (M.A.F) ranged from 0.2 to 1 and exhibited an indirectly proportional relationship to the PIC values expressed by each primer set. For instance, SCM80 had the highest M.A.F (1) but the lowest PIC (0) while Xrems1266 which had a low M.A.F (0.3), had a higher PIC value (0.8).

An average PIC value of 0.55 was detected for all loci conveying moderate variability among the rye and triticale accessions. This is similar to the 0.54 average variability detected among 80 hexaploid triticale accessions sourced from 17 countries (South Africa included) by Kuleung *et al.,* 2006.

When comparing the genetic diversity detected by the EST primers and by the genomic primers, it was found that the nine ESTs exhibited an average PIC of 0.58 while the 14 genomic microsatellites exhibited an average of 0.55. This hints at the detection of moderate polymorphisms in the rye genome with fewer EST SSRs than genomic SSRs.

Table 3.2: Frequency-based distances computed for wheat genome SSRs using the CSChord, 1967 distance method

OTU	Bacchus	Ibis	Rex	SST	Tobie	US2007	US2008	US2009	US2010
Bacchus	0.00	0.72	0.90	0.90	0.72	0.72	0.90	0.72	0.90
Ibis	0.72	0.00	0.72	0.90	0.90	0.90	0.90	0.90	0.90
Rex	0.90	0.72	0.00	0.90	0.90	0.90	0.90	0.90	0.90
SST	0.90	0.90	0.90	0.00	0.72	0.90	0.90	0.90	0.72
Tobie	0.72	0.90	0.90	0.72	0.00	0.72	0.90	0.72	0.90
US2007	0.72	0.90	0.90	0.90	0.72	0.00	0.90	0.90	0.90
US2008	0.90	0.90	0.90	0.90	0.90	0.90	0.00	0.90	0.90
US2009	0.72	0.90	0.90	0.90	0.72	0.90	0.90	0.00	0.90
US2010	0.90	0.90	0.90	0.72	0.90	0.90	0.90	0.90	0.00

O.T.U Observed taxonomic unit

Table 3.3: A statistical summary generated in PowerMarker v.3.25

Marker	M.A.F	Allele No.	Gene Diversity	PIC
SCM2-6RL	0.70	2	0.42	0.33
SCM4-1R	0.50	4	0.66	0.61
SCM5-3RL	0.60	4	0.58	0.54
SCM9-1RS	0.80	3	0.34	0.31
SCM36-1R	0.40	6	0.76	0.73
SCM38-2R	0.30	6	0.80	0.77
SCM39-1R	0.80	3	0.34	0.31
SCM40-7R	0.40	7	0.78	0.76
SCM43-2R	0.60	4	0.58	0.54
SCM66-3R	0.80	2	0.32	0.27
SCM69-2R	0.30	7	0.82	0.80
SCM75-2RL	0.80	3	0.34	0.31
SCM80-1R	1.00	1	0.00	0.00
SCM83-2R	0.60	2	0.48	0.36
SCM86-7R	0.40	3	0.64	0.56
SCM109-5RL	0.40	5	0.72	0.68
SCM112-3R	0.40	6	0.76	0.73
SCM120-5RL	0.20	8	0.86	0.84
SCM206-3RS	0.70	4	0.48	0.45
Xrems1135-7R	0.30	5	0.76	0.72
Xrems1162-7R	0.40	5	0.72	0.68
Xrems1197-7R	0.50	5	0.68	0.64
Xrems1266-5R	0.30	7	0.82	0.80

M.A.F Major allele frequency, **No.** Number, **PIC** Polymorphism information content

Primers highlighted in grey were chosen for the construction of the UPGMA dendogram.

Table 3.4: Frequency-based distances computed for rye genome SSRs using the CSChord, 1967 distance method

OTU	Bacchus	Duiker	Henoch	Ibis	Rex	Tobie	US2007	US2008	US2009	US2010
Bacchus	0.00	0.90	0.90	0.90	0.90	0.00	0.45	0.90	0.90	0.45
Duiker	0.90	0.00	0.90	0.90	0.45	0.90	0.90	0.90	0.90	0.90
Henoch	0.90	0.90	0.00	0.90	0.90	0.90	0.90	0.90	0.90	0.90
Ibis	0.90	0.90	0.90	0.00	0.90	0.90	0.90	0.90	0.45	0.90
Rex	0.90	0.45	0.90	0.90	0.00	0.90	0.90	0.90	0.90	0.90
Tobie	0.00	0.90	0.90	0.90	0.90	0.00	0.45	0.90	0.90	0.45
US2007	0.45	0.90	0.90	0.90	0.90	0.45	0.00	0.90	0.90	0.90
US2008	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.00	0.45	0.90
US2009	0.90	0.90	0.90	0.45	0.90	0.90	0.90	0.45	0.00	0.90
US2010	0.45	0.90	0.90	0.90	0.90	0.45	0.90	0.90	0.90	0.00

O.T.U Observed taxonomic unit

Dendograms

UPGMA and NJ dendograms (figures 4.1 and 4.2) were generated for the wheat genome SSRs. Both trees were bootstrapped to 1000 permutations and a consensus value of 70% was computed. Both trees generated bootstrap values below 70% indicative of unsupported clustering. The SBL (shortest branch length) values were 3.84 and 3.71 for the UPGMA and NJ dendograms respectively.

The wheat cultivar "SST88" is an out-group in this study cohort and was used as such when generating both trees. This cultivar was also used to root both trees.

 Five SSRs were chosen based on their distinct and repeatable banding patterns, PIC values (≥0.8) and the ability to distinguish among the closely related rye and triticale cultivars. Frequency-based distances (Table 3.4) computed for primer sets SCM38 and Xrems1266 spanning chromosomes 3R and 5R respectively were used to generate UPGMA and NJ dendograms (figures 4.3 and 4.4).

Both trees were bootstrapped to 1000 permutations. An SBL value of 2.70 was generated for both trees. Rye cultivars, "Henoch" and "Duiker" were used as out-groups in the generated trees.

Again, both dendograms exhibited the same relationships among the Elite cultivars.

Therefore, a consensus tree was generated to show the overall relationships among the Elite spring triticale while using data generated by all seven markers developed in wheat and rye. This NJ tree is shown in figure 3.5. This tree manages to cluster cultivars "US2008" and "US2010" which according to the available pedigree, are sister lines. Also shown, are the clusterings of "Bacchus" and "US2007" and "Tobie" and "US2009" which share parents with similar pedigrees. Cultivar "Rex" is also more closely related to the other triticale cultivars unlike cultivar "Ibis" which is closer to durum wheat than triticale and this is shown in the NJ dendogram generated.

Therefore, the panel of seven SSR markers developed in wheat and rye, managed to show differences among the commercial triticale cultivars. These markers also generated phylogenetic fingerprints that exhibited relationships which are consistent with the available pedigree data and ultimately managed to accurately identify the commercial cultivars.

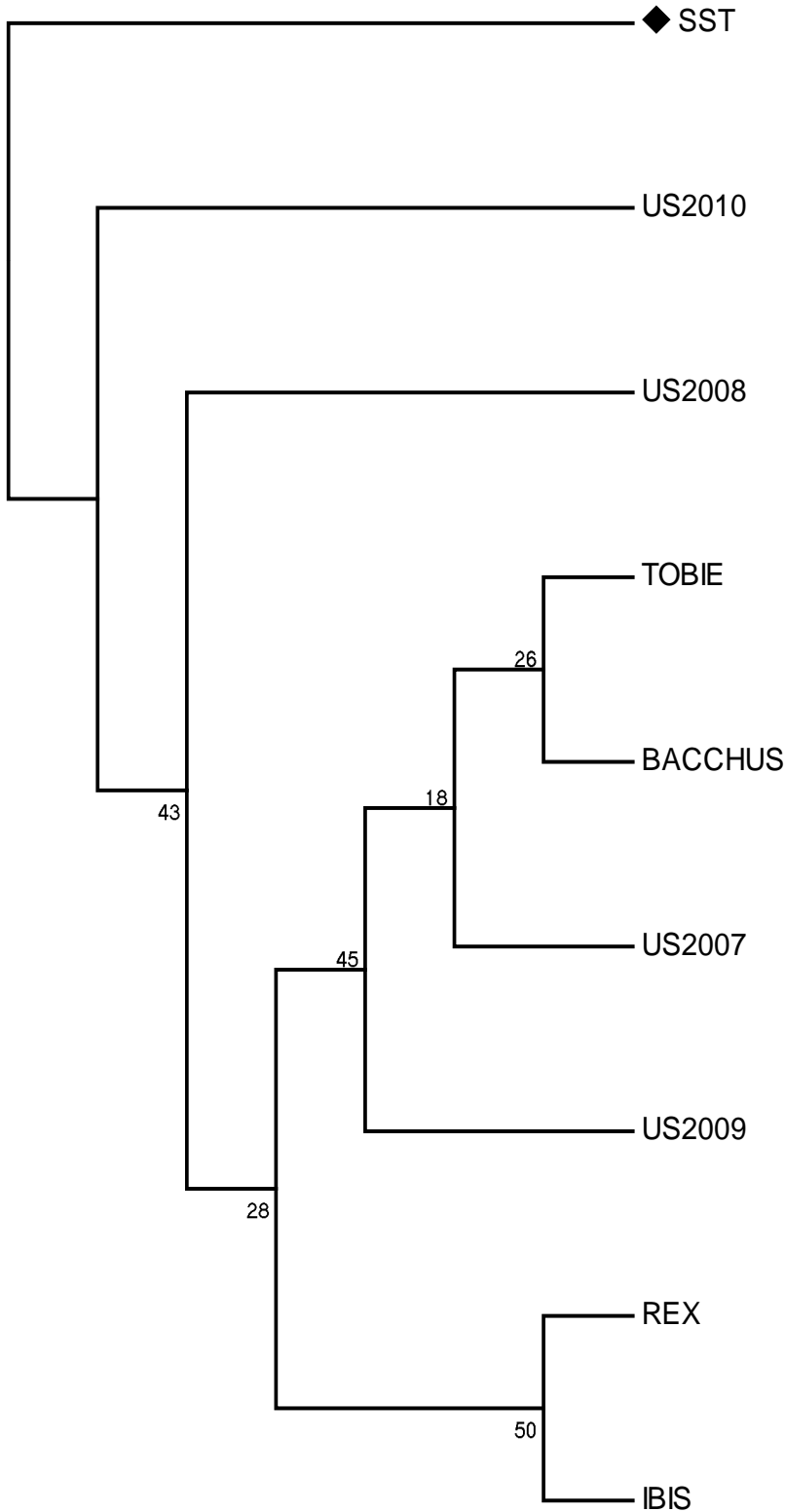**Figure 3.1**: A UPGMA dendogram showing the relationship of eight spring triticale cultivars based on CS Chord (1967) frequency-based distances generated by five SSRs developed in wheat. "SST" is used as the out-group and is used to root the tree.

Figure 3.2: A NJ dendogram showing the relationship of eight spring triticale cultivars based on CSChord (1967) frequency-based distances generated by five SSRs developed in wheat. "SST" is the out-group and is used to root the tree.

Figure 3.3: A UPGMA dendogram showing the relationship of eight spring triticale cultivars based on CSChord (1967) frequency-based distances generated by two SSRs developed in rye. Rye cultivars "Henoch" and "Duiker" are out-groups and were used to root the tree.
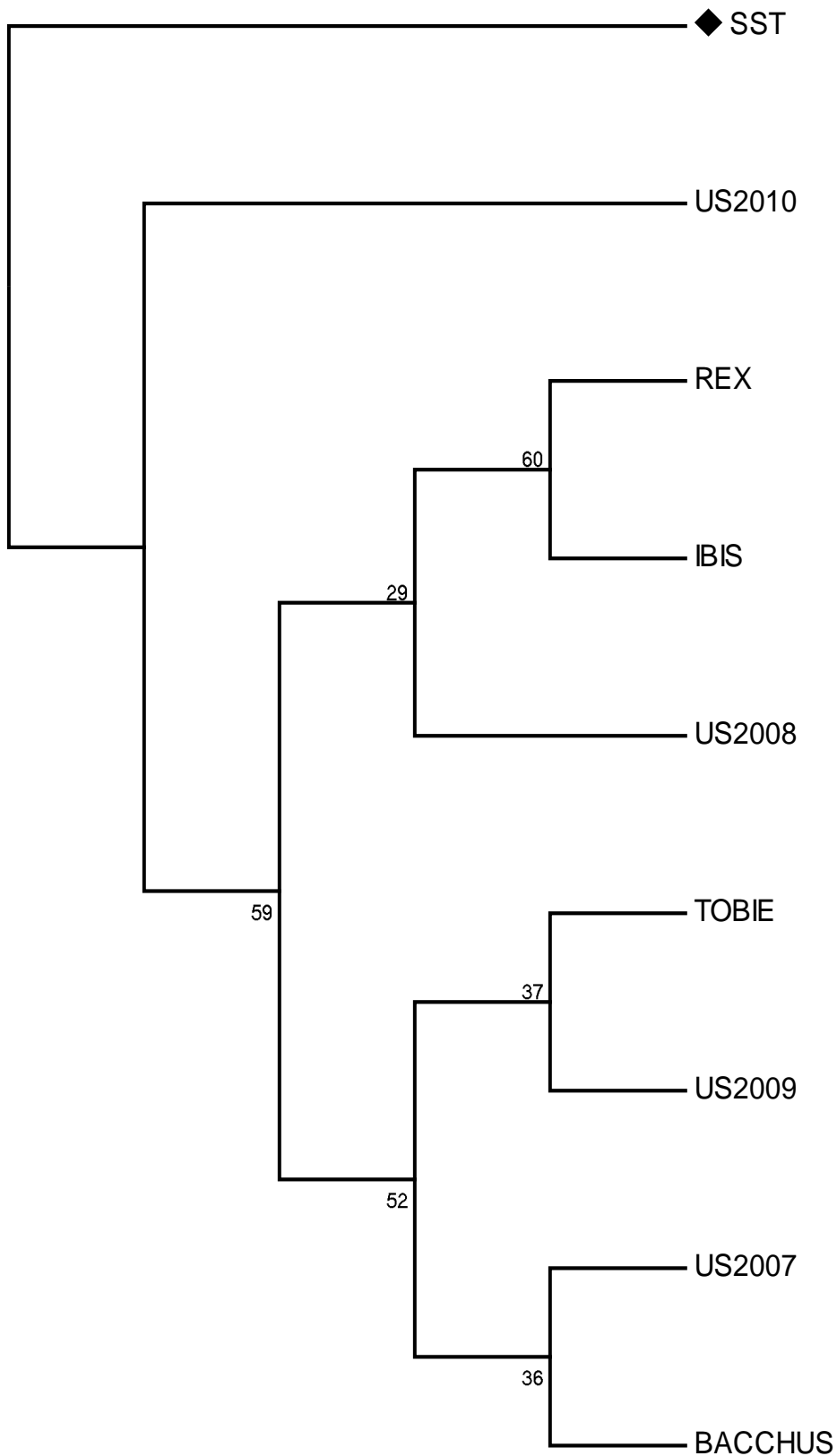
94

Figure 3.4: A NJ dendogram showing the relationship of eight spring triticale cultivars based on CSChord (1967) frequency-based distances generated by two SSRs developed in rye. Rye cultivars "Henoch" and "Duiker" are out-groups and were used to root the tree.
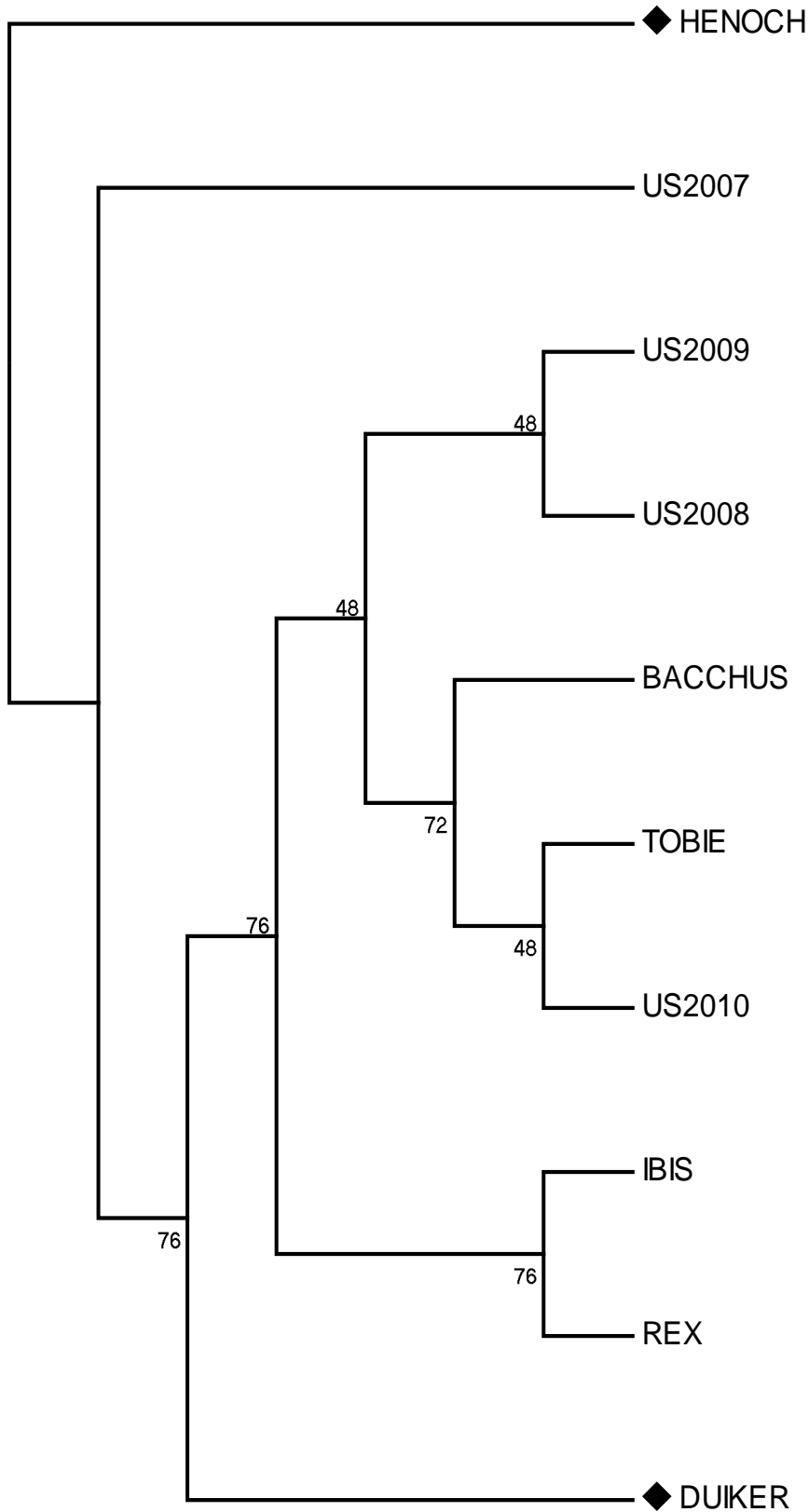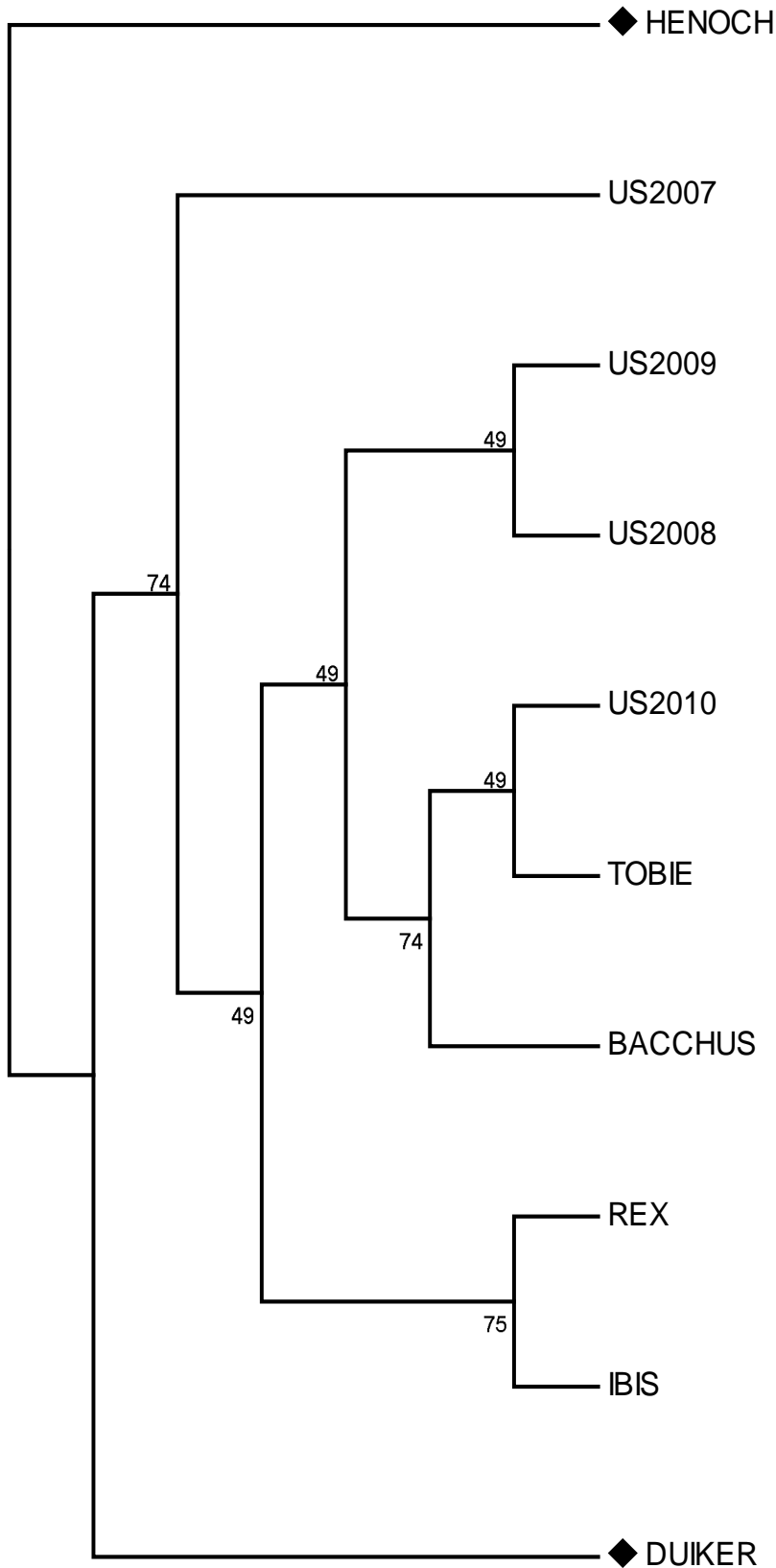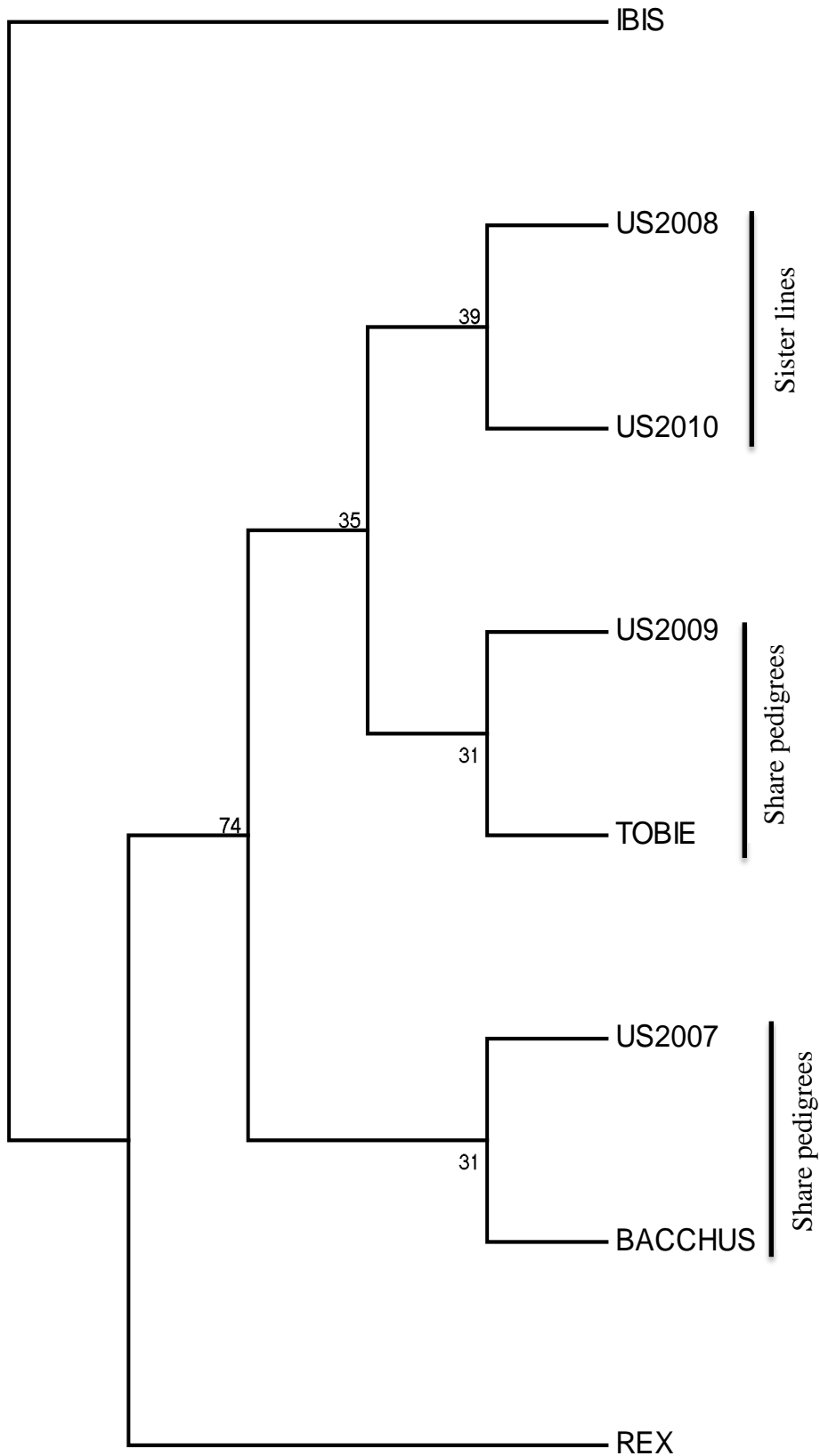
95

Figure 3.5: A NJ dendogram showing the relationships among eight spring triticale cultivars using frequency-based distances generated by seven SSRs developed in wheat and rye.

Polymorphisms detected in the A, B and R genomes

Interestingly, only a few B genome SSRs are needed to detect the moderate genetic variation among the triticale which isn't the case for the A genome SSRs. This could be due to the fact that eight of the 35 A genome markers that produced amplicons did not show exclusive clustering to the A genome. These eight SSRs also show clustering in the D and B genomes. This could explain why the scale gets a little swayed to the B genome side in terms of genetic variability detected.

Of note, is the actual amplification of some of the D genome SSR fragments in hexaploid triticale which essentially lacks this genome. Tams *et al.,* (2004) attained amplification of hexaploid triticale DNA fragments while using D genome SSRs mapping to all but chromosomes 1D and 2D. Leonova *et al.,* (2005) suggests that this may be associated to wheat-rye or wheat-wheat translocations within triticale. Similarly, it has been shown that the change in triticale genomic composition and gene expression results from the effect of rye chromosomes and not the wheat-rye translocation composition. Both ways, confirmatory studies need to be done using *in situ* hybridization and C-banding (Zhang *et al.,* 2007)

The detected overall average PIC value (0.52) in this study resembles the 0.54 value obtained by Kuleung *et al.* (2006) when screening 80 hexaploid triticale accessions of global origin, using 43 wheat developed SSRs. This implies that the Elite and commercially available triticale cultivars bred in the SU-PBL exhibit a genetic diversity equivalent to that of the global triticale germplasm screened by Kuleung *et al.* (2006). Comparatively in a study assessing the genetic diversity among 54 Brazilian triticale accessions using 42 wheat SSRs, Da Costa *et al.* (2007) recorded an average PIC value of 0.36 which is much lower than that obtained in this study. This was not surprising as most of the accessions were of Mexican origin. This implies that the South African spring triticale screened have a higher level of genetic variability than the triticale accessions that were of Mexican origin.

An average PIC value of 0.55 was detected in the R-genome conveying moderate variability among the rye and triticale accessions. This is similar to the 0.54 average variability detected among 80 hexaploid triticale accessions sourced from 17 countries (South Africa included) by Kuleung *et al.* (2006).

In comparing the genetic diversity detected by the EST primers as opposed to the genomic primers, it was found that the nine ESTs exhibited an average PIC of 0.58 while the 14

genomic microsatellites exhibited an average of 0.55. This hints at the detection of moderate polymorphisms in the rye genome with fewer EST SSRs than genomic SSRs.

Most EST SSRs either have a gene function or a close linkage to coding regions, which is not established in most genomic SSRs. For instance, Hackauf and Wehling (2002) did a BLASTX search of the 70 rye ESTs and found 44.5% of the SCM loci could be associated with proteins of known or unknown function, and primer set SCM80 was found to be associated with a putative RNA-binding protein the *Oryza sativa* species.

This implies that EST SSRs have the potential to increase the efficiency of SSR marker implementation in plant breeding as they express a higher transferability among related species, and have an ability to detect variation in the transcribed regions genomes such as rye (Hackauf and Wehling, 2002; Zhang *et al.,* 2005).

Conclusions

SSR markers that are specific to the wheat genome were successfully optimised and transferred to the Elite and commercially thriving triticale cultivars in the SU-PBL. These markers showed an overall transferability of 80% and managed to detect a moderate genetic variability (PIC 0.52) among the genotypes. SSR markers specific to the rye genome were also optimised and showed 79.3% transferability across the triticale cultivars. These markers also detected moderate levels of genetic variability among the genotypes (PIC 0.55).

Data generated by a panel of seven SSR markers specific to both the wheat and rye genomes, was used to generate phylogenetic fingerprints for the cultivars. These markers were chosen because of their ability to discern the differences among the triticale cultivars and their ability to exhibit high levels of polymorphisms. The generated fingerprints managed to group those cultivars that shared parents with the same pedigrees and effectively managed to accurately identify the cultivars. Therefore, this panel of seven SSR markers was chosen for the implementation of a high throughput analysis.

References

- Cavalli-Sforza, L, L and Edwards, A, W, F (1967) 'Phylogenetic analysis: models and estimation procedures', *American Journal of Human Genetics*, 19: 233-257
- da Costa, C, T, Albuquerque, A, C, S, Junior, A, N, Marcelino, F, C and Pereira, J, F (2007) 'Genetic diversity of Brazilian triticale evaluated with genomic wheat microsatellites', *Pesquisa Agropecuária Brasília*, 42(11): 1577-1586

- Devos, K, M, Atkinson, M, D, Chinoy, C, N, Francis, H, A, Hartcourt, R.L., Koebner, R, M, D, Liu, C, J, Masojc, P, Xie, D, X, and Gale, M, D (1993) 'Chromosomal rearrangements in the rye genome relative to that of wheat', *Theoretical and Applied Genetics*, 85: 673-680

- Doyle, J, J and Doyle, J, L (1990) 'Isolation of plant DNA from fresh tissue', *Focus* 12: 13-15

- GenBank (2009), Homepage [online] (cited October 2011) Available from <URL:http://www.ncbi.nlm.nih.gov/>

- Goyal, A, Bandopadhyay, Sourdille, P, Endo, T, R, Baylan, H, S and Gupta, K (2005) 'Physical maps of wheat chromosomes', Functional and Integrative Genomics, 5 (4): 260-263

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

- Gupta, P, K *et al.,* (2002) 'Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat', *Theoretical and Applied Genetics*, 105: 413-422

- Gupta, P, K, Rustgi, S, Sharma, S, Singh, R, Kumar, N and Balyan, H, S (2003) 'Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat', *Molecular Genetics and Genomics*, 270: 315-323

- Hackauf, B and Wehling, P (2002) 'Identification of microsatellite polymorphisms in an expressed portion of the rye genome', *Plant Breeding*, 121: 17-25

- Honing, J (2007) '*Evaluation and implementation of DNA-based diagnostic methodology to distinguish wheat genotypes*', Stellenbosch: University of Stellenbosch. (Master's thesis).

- Khlestkina, E, K, Than, M, H, M, Pestsova, E, G, Röder, M, S, Malyshev, S, V, Korzun, V and Börner, A (2004) 'Mapping of 99 new microsatellite-derived loci in rye (Secale cereale L.) including 39 expressed sequence tags', *Theoretical and Applied Genetics*, 109: 725-732

- Korzun, V, Malyshev, S, Voylokov, A, V and Börner, A (2001) 'A genetic map of rye (*Secale cereale* L.) combining RFLP, isozyme, protein, microsatellite and gene loci', *Theoretical and Applied Genetics*, 102: 709-717

- Kong, L, Ohm, H, W, Cambron, S, E and Williams, C, E (2005) 'Molecular mapping determines that Hessian fly resistance gene *H9* is located on chromosome 1A of wheat', *Plant Breeding*, 124: 525-531

- Kuleung, C, Baezinger, P, S and Dweikat, I (2004) 'Transferability of SSR markers among wheat, rye and triticale', *Theoretical and Applied Genetics*, 108: 1147-1150

- Kuleung, C, Baezinger, P, S, Kachman, SD and Dweikat, I (2006) 'Evaluating the genetic diversity of triticale with wheat and rye SSR markers', *Crop Science*, 46: 1692-1700

- Leonova, I, N, Dobrovolskaya, O, B, Kaminskaya, L, N, Adonina, I, G, Koren, L, V, Khotyljova, L, V and Salina, E, A (2005) 'Molecular analysis of the triticale lines with different *vrn* gene systems using microsatellite markers and hybridisation in situ', *Russian Journal of Genetics,* 41: 1014-1020

- Liu, K and Muse, S, V (2005) 'PowerMarker: Intergrated environment for genetic marker data', *Bioinformatics*, 21(9): 2128-2129

- Ma, Z, Q, Röder, M and Sorrells, M, E (1996) 'Frequencies and sequence characteristics of di-, tri-, and tetra-nucleotide microsatellites in wheat', *Genome*, 39: 123-130

- Qi, L *et al.,* (2004) 'A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat', *Genetics*, 168: 701-712

- Rampling, L, R, Harker, N, Shariflou, M, R and Morrel, M, K (2001) 'Detection and analysis systems for microsatellite markers in wheat', *Australian Journal of Agricultural Research*, 52: 1131-1141

- Ramsay, L *et al.,* (2000) 'A simple sequence repeat-based linkage map of barley', *Genetics*, 156: 1997-2005

- Röder, M, S, Lapitan, N, L, V, Sorrells, M, E and Tanksley, S, D (1993) 'Genetic and physical mapping of barley telomeres', *Molecular and General Genetics,* 238: 294-303

- Röder, M, S, Plaschke, J, Konig, S, U and Sorrells, M, S (1995) 'Abundance, variability and chromosomal location of microsatellite in wheat,' *Molecular and General Genetics*, 246: 327-333

- Röder, MS, Korzun, V, Wendehake, K, Plaschke, J, Tixier, MH, Leroy, P and Ganal, MW (1998) 'A microsatellite map of wheat', *Genetics,* 149: 2007-2023

- Rözen and Skaletsky, (2000) 'Primer3 on the www for general users and for biologist programmers', *Methods in Molecular Biology*, 132: 365-86

- Rychlik, W and Rhoads, R, E (1989) "A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA", *Nucleic Acids Research*, 17: 8543-8551

- Saitou, N and Nei, M (1987) "The neighbour-joining method; a new method for reconstructing phylogenetic trees", *Molecular Biology and Evolution*, 4: 406-425

- Saal, B and Wricke, G (1999) 'Development of simple sequence repeat markers in rye (*Secale cereale* L.)', *Genome*, 42: 964-972

- Senft, P, and Wricke, G (1996) 'An extended genetic map of rye (*Secale cereale* L.),' *Plant Breeding*, 115: 508-510

- Schubert, A, C (2009) '*Quantitative trait loci analysis in spring wheat comparing two advanced backcross populations derived from an exotic wheat accession*', University of Bonn. (PhD dissertation)

- Singh, N, K, Shepherd, K, W and McIntosh, R, A (1990) 'Linkage mapping of genes for resistance to leaf, stem and stripe rusts and ω-secalins on the short arm of rye chromosome 1R,' *Theoretical and Applied Genetics*, 80: 609-616

- Sneath, P, H, A and Sokal, R, R (1973) "Numerical Taxonomy", W, H, Freeman, San Franciso

- Somers, D, Isaac, P and Edwards, K (2004) 'A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.)', *Theoretical and Applied Genetics*, 109: 1105-1114

- Song, Q, J, Fickus, E, W and Cregan, P, B (2002) 'Characteristics of trinucleotide markers in wheat', *Theoretical and Applied Genetics*, 104: 286-293

- Song, Q, Shi, J, Singh, S, Fickus, E, Costa, J, Lewis, J, Gill, B, Ward, R and Cregan, P (2005) 'Development and mapping of microsatellite (SSR) markers in wheat', *Theoretical and Applied Genetics*, 110: 550-560

- Sourdille, P, Tavaud, M, Charmet, G and Bernard, M (2001) 'Transferability of wheat microsatellites to diploid Triticeae species carrying the A, B and D genomes', *Theoretical and Applied Genetetics*, 103: 346-352

- Tixier, M, H, Sourdille, P, Röder, M, Leroy, P and Bernard, M (1997) 'Detection of wheat microsatellites using a non radioactive silver-nitrate staining method', *Journal of Genetics Breeding*, 51: 175-177

- Tams, SH, Bauer, E, Oettler, G and Melchinger, AE (2004) 'Genetic diversity in European winter triticale determined with SSR markers and coancestry coefficient', *Theoretical and Applied Genetics*, 108:1385-1391

- Tams, S, H, Melchinger, A, E and Bauer, E (2005) 'Genetic similarity among European winter triticale Elite germplasms assessed with AFLP and comparisons with SSR and pedigree data', *Plant Breeding*, 124: 154-160

- Tamura, K, Peterson, D, Peterson, N, Stecher, G, Nei, M, and Kumar, S (2011) 'MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods', *Molecular Biology and Evolution* (submitted)

- Teemu Kuulasmaa, OligoAnalyzer (version 1.0.2), Homepage [online]. Available from <URL:http://www.uku.fi/~kuulasma/OligoSoftware>

- Varshney, R, K, Thiel, T, Stein, N, Langridge, P and Graner, A (2002) 'In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species', *Cellular and Molecular Biology Letters*, 7: 537-546

- Vyhnánek, T, Nevrtalova, E, and Slezakova, K (2009) 'Detection of the genetic variability of triticale using wheat and rye SSR markers' *Cereal Research Communication*, *37*: 23-29

- Weir, B, S, (1996) '*Genetic data analysis II*', MA: Sinauer Associates, Incorporated, Sunderland

- Zhang, L, Y, Bernard, M, Leroy, P, Feuillet, C and Sourdille, P (2005) 'High transferability of bread wheat EST-derived SSRs to other cereals', *Theoretical and Applied Genetics*, 111: 677-687

- Zhang, L, Y, Bernard, M, Raveli, C, Balfourier, F, Leroy, P, Feuillet, C and Sourdille, P (2007) 'Wheat EST-SSRs for tracing chromosome segments from a wide range of grass species', *Plant Breeding*, 126: 251-258

Chapter 4:   Implementation and assessment of SSR marker based high-throughput approach in triticale breeding

Implementation and assessment of SSR marker based high-throughput approach in triticale breeding

D. Bitalo

Stellenbosch University Plant Breeding Laboratory, Department of Genetics,
Private Bag X1, Matieland 7602, South Africa

Abstract

The objective of this study was to use the previously identified polymorphic simple sequence repeat (SSR) markers developed in wheat and rye, to screen a selection of the pure breeding material of the Stellenbosch University Plant Breeding Laboratory's (SU-PBL) triticale breeding programme using a semi-automated DNA technique.

The forward primers of the seven SSR primer sets were fluorescently labelled and their PCR protocols were optimised to allow for the multiplexing of primers sets. A total of 306 triticale entries (30 Elite, 20 Seniors and 256 F6 entries) were screened using all seven SSR primers sets previously developed in the wheat and rye genomes.

An average PIC value of 0.65 was observed among the triticale entries implying a moderate to high level of diversity among the triticale breeding material. Neighbour-joining (NJ) dendograms and unweighted pair group method with arithmetic average (UPGMA) were then generated for the 30 Elite, 20 Seniors and 256 F6 entries. No clear groupings were observed among the advanced breeding material but rather sub-clusters were observed.

Keywords: Triticale, rye, simple sequence repeat (SSR) markers, fluorescence-based semi-automated analysis, genetic diversity, fingerprint.

Introduction

To implement the routine use of SSR markers a genotyping technique that is both efficient and cost effective is required (Pesoa-Filhoa *et al.,* 2007). Due to the possibility of automating the analysis of microsatellites, fluorescently labelled primer sets and a DNA sequencer can be used in a procedure called fluorescent-based semi-automated analysis. This technique was first used in the genotyping of restriction fragments (Carrano *et al.,* 1989) and incorporates the use of one fluorescently labelled primer and one unlabelled primer whilst maintaining the same PCR conditions implemented when running conventional PCR reactions that require only unlabelled primers (Coburn *et al*., 2002).

The labelling of primers permits multiplexing which increases the production of information from one lane or sample on a DNA sequencer. Consequently, primers that are labelled with different fluorescent dyes can be analysed in one lane without hindrance, reducing the overall costs incurred. Furthermore, using a semi-automated technique to genotype SSRs improves the accuracy in the size calling of alleles due to the addition of an internal size standard to each lane and the use of automated allele-calling algorithms (Coburn *et al*., 2002).

The SU-PBL programme provided triticale breeding material from three different phases; the Elite, Senior and F6 phases of the breeding programme. This section of the study was aimed at implementing the findings from the previous chapters (3 and 4). This was done by identifying those markers that showed genetic diversity amongst the control cohort, optimising these markers for semi-automated analysis and using the optimised protocol to test the breeding material within the breeding programme.

Breeding Material

Fresh leaf tissue was excised from 306 advanced triticale breeding lines and cultivars sourced from the breeding program at Welgevallen.

Twenty entries were randomly chosen from the Elite triticale breeding program (labelled E1-E20), 30 entries were chosen from the Senior triticale breeding program (labelled S1-S30) and 256 entries were chosen from the triticale F6 generation (labelled F1-F256).

Genomic DNA extraction

In a 2.2 ml microcentrifugetube, about 0.1g of leaf tissue and three steel ball bearings were added. Following a modified CTAB protocol described by Doyle and Doyle (1990), the extraction procedure is described in chapter 3.

All extracted DNA samples were quantified using the NanoDrop® ND-1000 Spectrophotometer following the Thermo SCIENTIFIC user's manual guide. Each DNA sample was then diluted to 100ng/ µl and stored at 4 ^{o}C while all stock DNA was stored at -20 ^{o}C until needed.

Fluorescence-based semi-automated analysis

Only one of the microsatellite primers (i.e. forward primer) was labelled with a fluorescent dye as is shown in Table 4.1. The primers with differing fragment lengths were labelled with the same dyes and those that have similar fragment lengths were labelled with different dyes to avoid overlapping, and enable efficient multiplexing of the primers in one PCR amplification reaction.

PCR protocol

Approximately 100ng of template genomic DNA, 0.5µM of each fluorescently labelled forward primer, 0.5µM of each unlabelled reverse primer, 1X KAPA2G™ multiplex kit (kapabiosystems) containing 0.2mM of each dNTP, 3mM of $MgCl_2$, 1.5X KAPA2G buffer A, 1U of KAPA2G Fast HotStart DNA Polymerase, and ddH_2O were added per 20µl reaction.

PCR cycling conditions performed in a Thermal Cycler 2720 (Applied Biosystems) PCR machine and adopted from the KAPA2G™ multiplex kit manual, were modified; 3 mins at 95°C, 30 cycles of 15 sec at 95°C, 30sec at 60°C, 30 sec at 72°C, 30 mins at 72°C and a final holding temperature of 4°C. All PCR products were stored at 4°C until needed. The products were then analysed on an automated DNA sequencer (Applied Biosystems 3130xl Genetic Analyzer).

Table 4.1: Fluorescently labelled primers used for the semi-automated analysis

Marker	Forward/reverse	Expected length	Dye
cfe2-6A,6B,6D F	CTTCGCCGACAAGAAGAAGT	243-298	NED Yellow
cfe2-6A,6B,6D R	CGGCACGTACTCCACCTC		
cfe53-2AL F	TGGACCGCAGAGACTTCG	100-124	PET Red
cfe53-2AL R	GTCCGCCCAAACCCTACC		
Xcfd79-3A F	TCTGGTTCTTTGGAGGAAGA	190-250	FAM Blue
Xcfd79-3A R	CATCCAACAATTTGCCCAT		
Xgwm108-3B F	CGACAATGGGGTCTTAGCAT	125-178	FAM Blue
Xgwm108-3B R	TGCACACTTAAATTACATCCGC		
Xgwm369-3A F	CTGCAGGCCATGATGATG	184-295	PET Red
Xgwm369-3A R	ACCGTGGGTGTTGTGAGC		
Xrems1266-5R F	ACGACGGCAGTGAGAGAGAG	168-271	VIC Green
Xrems1266-5R R	TCGGCTTCATCGTCTACTCC		
SCM38-2R F	TGACCTGCACACCTCATCTCA	106-217	NED Yellow
SCM38-2R R	GCTTTGCTCCTGATGTCGAT		

Automated sequence electrophoresis

Automated sequencing was done on an Applied Biosystems 3130xl Genetic Analyzer (California, USA) that is suited for fluorescence-based capillary electrophoresis. The capillary electrophoresis system allows for the use of an electrokinetic injection to load the samples on to the capillaries in less than 30 seconds. A flowable polymer is loaded into the capillaries prior to each run. Samples are injected and run according to the user's specified conditions to detect polymorphisms among samples. A GeneScan™ size standard- 500 LIZ® was added to each of the PCR amplified samples. This size standard is designed for sizing DNA fragments ranging from 35 to 500 bp. This size standard emits 16 single-stranded labelled fragments of 35, 50, 75, 100, 139, 150, 160, 200, 250, 300, 340, 350, 400, 450, 490, 500 bases. The ensuing results can be integrated with several software packages for the analysis of data and in this case, the software program GeneMapper® version 4.0 was used.

Data analysis

The ensuing sequence data was analysed using GeneMapper® version 4.0, which allows for allele calling and annotation. The annotated allele data was exported to PwerMarker v 3.25 25 (Liu and Muse, 2005) for a computation of statistical summaries.

Results and Discussion

Triticale breeding material

The triticale breeding material was tested using seven SSR primers specific to the wheat and rye genomes. The scored alleles were then imported into PowerMarker v 3.25 (Liu and Muse, 2005) and a statistical summary was computed showing the M.A.F, genetic diversity and PIC (Table 4.2).

The average M.A.F was 0.45 ranging from 0.17 (marker cfe2) to 0.84 (marker Xrems1266). The average PIC value was 0.65 ranging from 0.27 to 0.90. This illustrates a moderate to high level of genetic diversity among the triticale breeding material.

The SSR primers specific to the rye genome displayed PIC values much lower than those displayed by the SSRs specific to the wheat genome, as is shown in Table 4.2. This could be interpreted as a demonstration of lower genetic variability in the rye genome within the triticale breeding material but because only two chromosomes (2R and 5R) are covered by the SSRs used in screening the material, this interpretation would be biased.

Table 4.2: A summary of the statistical analysis performed in PowerMarker v 3.25

Marker	M.A.F	Sample Size	Gene Diversity	PIC
cfe2-6A,6B,6D	0.17	314	0.90	0.90
cfe53-2AL	0.48	314	0.62	0.55
SCM38-2R	0.63	314	0.56	0.53
Xcfd79-3A	0.47	314	0.69	0.65
Xgwm369-3A	0.48	314	0.72	0.70
Xgwm108-3B	0.09	314	0.95	0.95
Xrems1266-5R	0.84	314	0.28	0.27
Mean	0.45	314	0.68	0.65

M.A.F Major allele frequency, **PIC** Polymorphism information content

Based on the computed frequency-based distances, Unweighted pair group method with arithmetic average (UPGMA) and Neighbour-joining (NJ) dendograms were generated (figures 4.1-4.5) to show whether the genetic relationship among triticale entries from the Elite, Senior and F6 phases relate to the commercially viable cultivars that were used as controls in this study.

In figure 4.1, the triticale advanced breeding material from the Senior breeding phase was grouped into four main clusters (Cluster I-IV). Within these clusters, it was possible to postulate which of the Senior triticale entries share a pedigree with any of the Elite and commercially viable cultivars that were used as controls within the study. Cluster II shows that triticale entries labelled S1, S16, S25 and S9 share a pedigree with the Elite cultivar "Ibis" (tagged with the black diamond).

Figure 4.2 shows the UPGMA dendogram created for the Senior triticale entries. This dendogram shows a varying genetic fingerprint from the one exhibited by the NJ dendogram as it generates five and not four clusters. However, both dendograms exhibit the clustering of the Elite and commercially thriving cultivars. These cultivars cluster with each other and away from the advanced breeding material in the Senior phase of breeding. This pattern shows that the panel of seven SSR markers managed to discern between the breeding material and the Elite cultivars but generated trees that were not statistically supported by the bootstrap values.

In figure 4.3, the advanced breeding material from the Elite breeding phase was grouped in to five main clusters (Cluster I-V). This is almost similar to the number of clusters generated by the UPGMA dendogram among the Elite triticale entries (figure 4.5).

When analysing the genetic diversity among the F6 triticale entries, the panel of seven SSR markers failed to group the material into any major clusters. Plenty of sub-clusters were generated (figure 4.5). This illustrates that there's more genetic diversity among the F6 triticale entries than any of the other advanced breeding material within the Senior and Elite phases of the breeding program.

Figure 4.1: A NJ dendogram showing the relationship of the Senior triticale entries based on data generated by seven SSRs developed in wheat and rye.

Figure 4.2: A UPGMA dendogram showing the relationship of the Senior triticale entries based on data generated by seven SSRs developed in wheat and rye

Figure 4.3: A NJ dendogram showing the relationships of the Elite triticale entries based on data generated by seven SSRs developed in wheat and rye

Figure 4.4: A UPGMA dendogram showing the relationships of the Elite triticale entries based on data generated by seven SSRs developed in wheat and rye
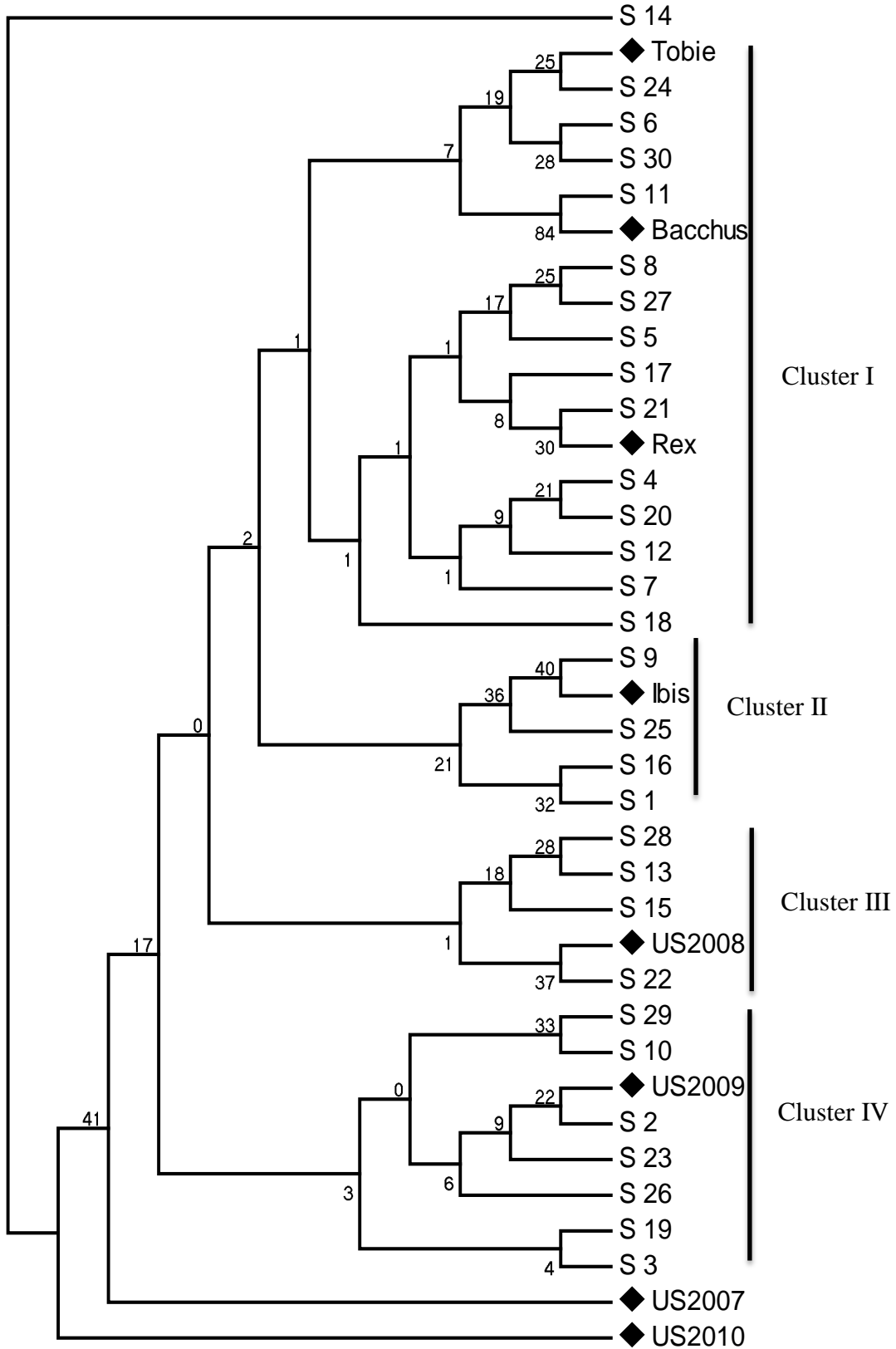
Figure 4.5: A NJ dendogram showing the relationships of the F6 triticale entries based on data generated by seven SSRs developed in wheat and rye

Figure 4.5 cont.

116

Figure 4.5 cont.

117

Figure 4.5 cont.

118
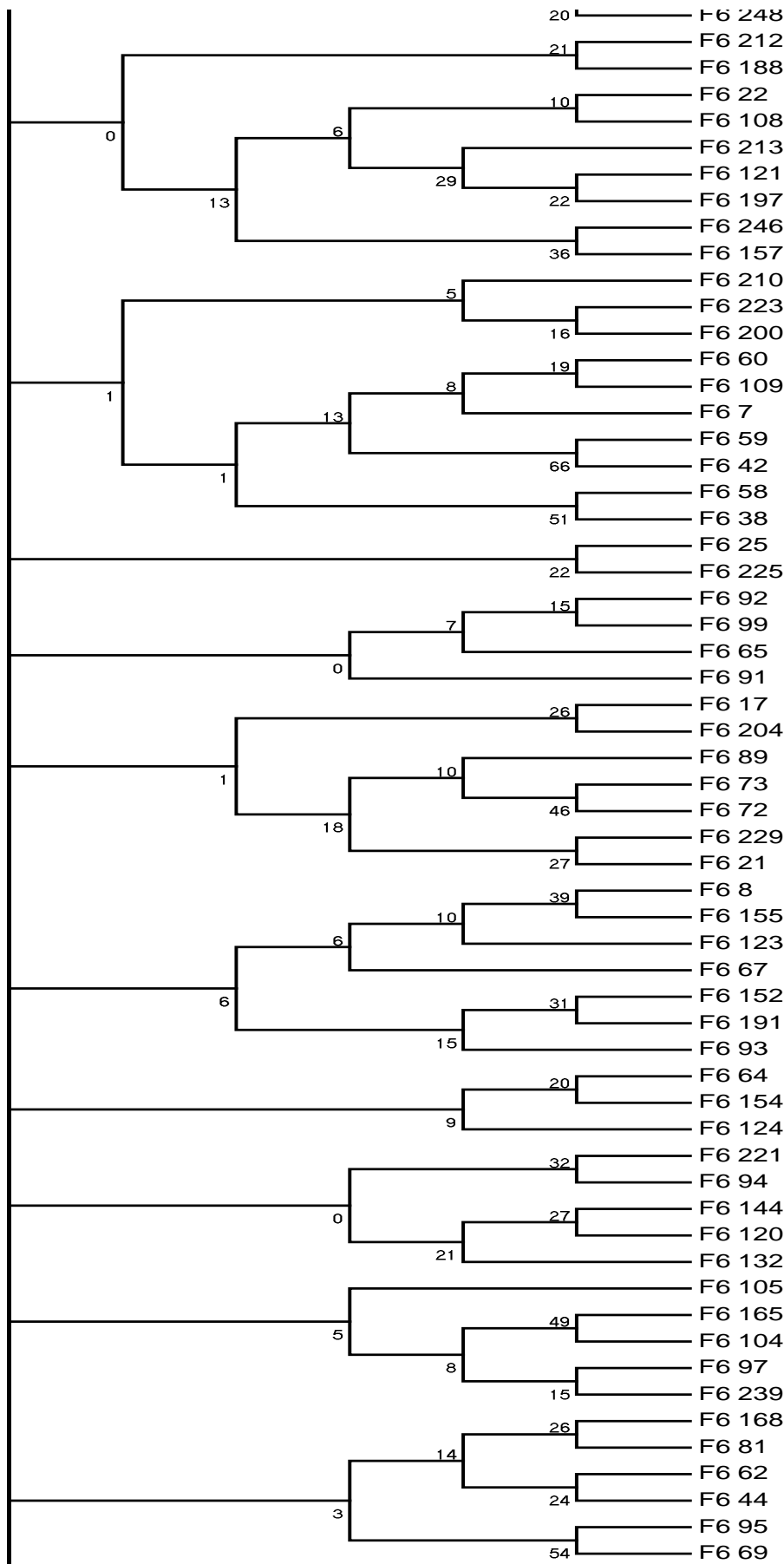
Figure 4.5 cont.

Figure 4.5 cont.

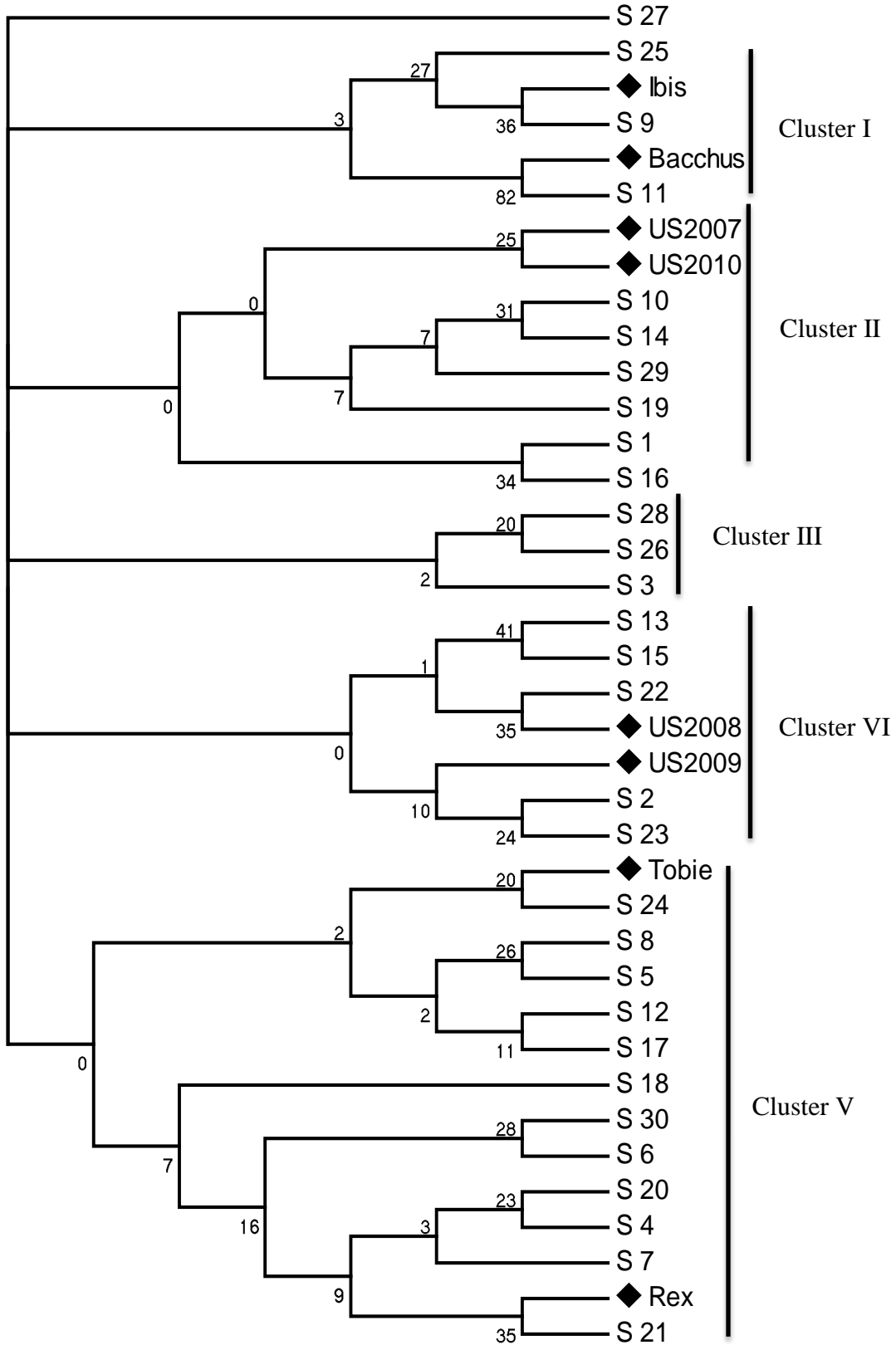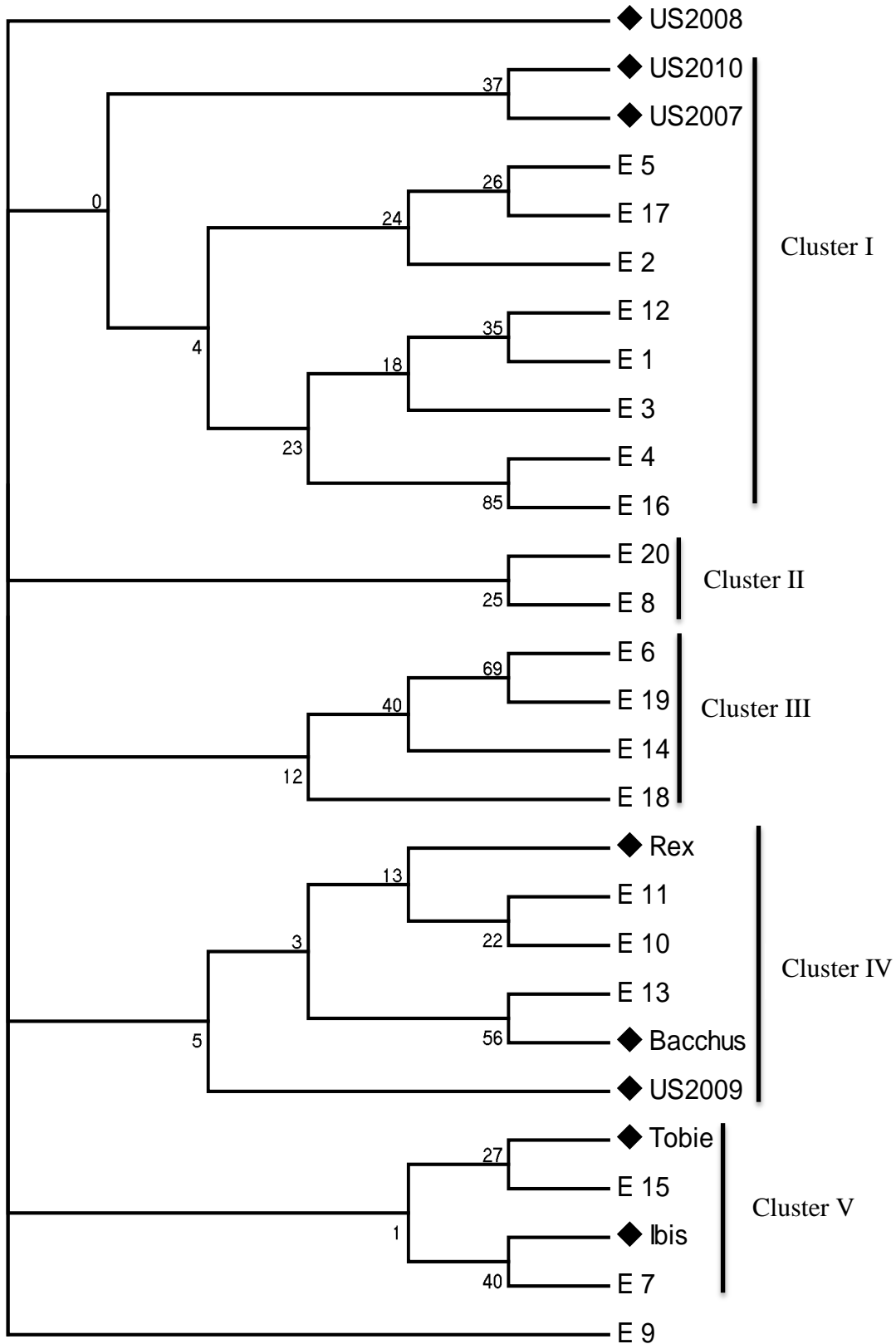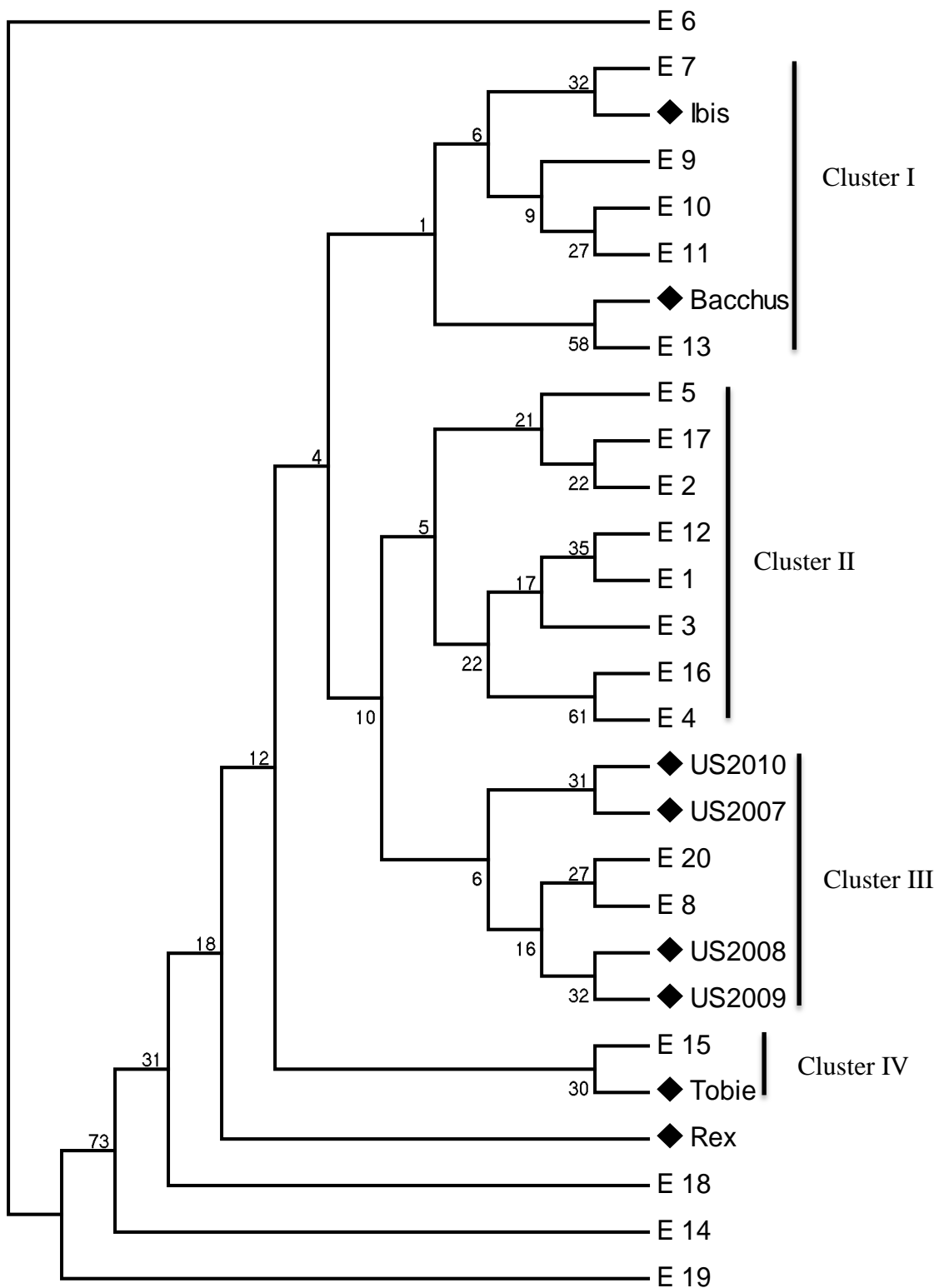120

Figure 4.5 cont.

Figure 4.6: A UPGMA dendogram showing the relationships of all the advanced breeding triticale entries based on data generated by seven SSRs developed in wheat and rye
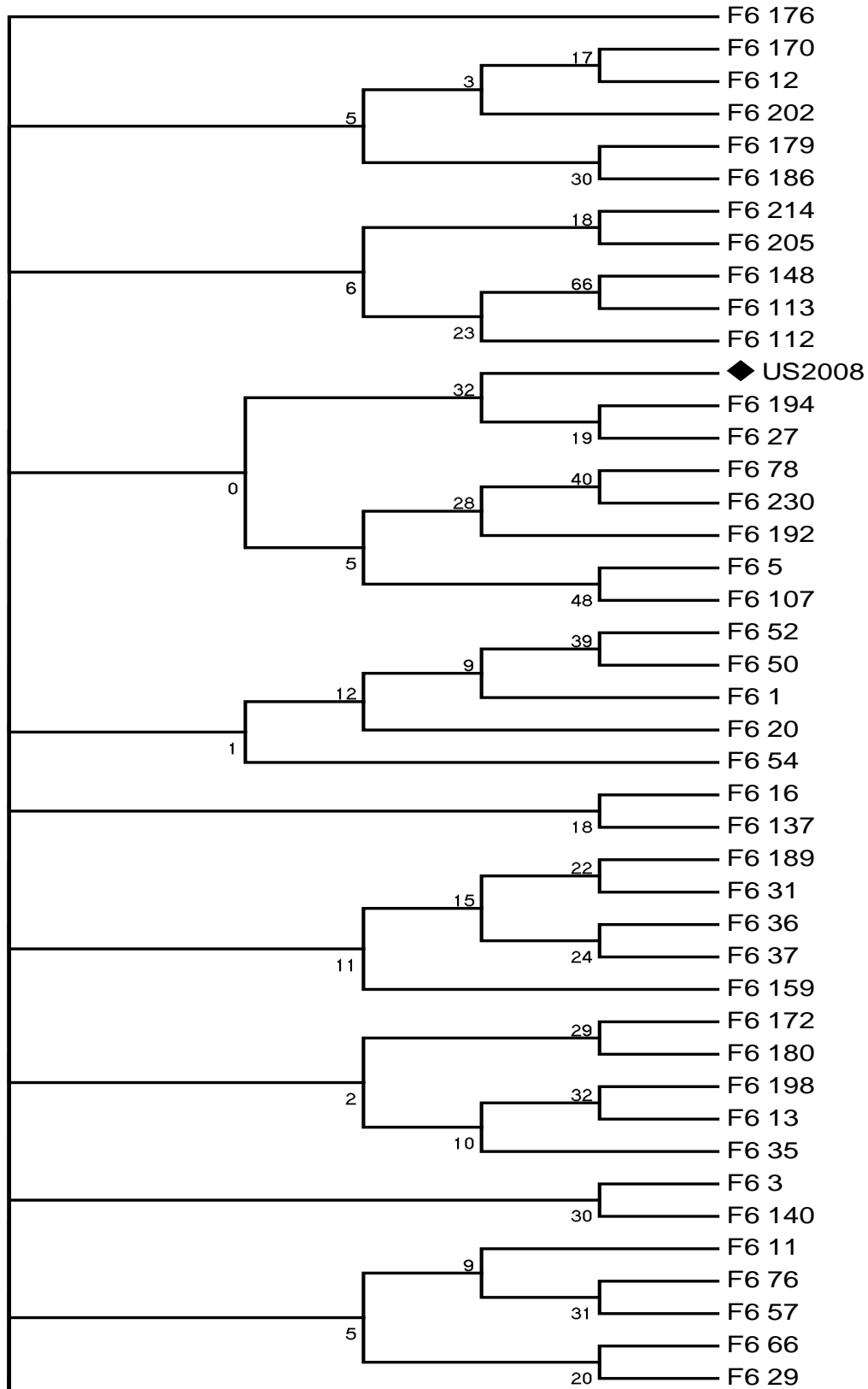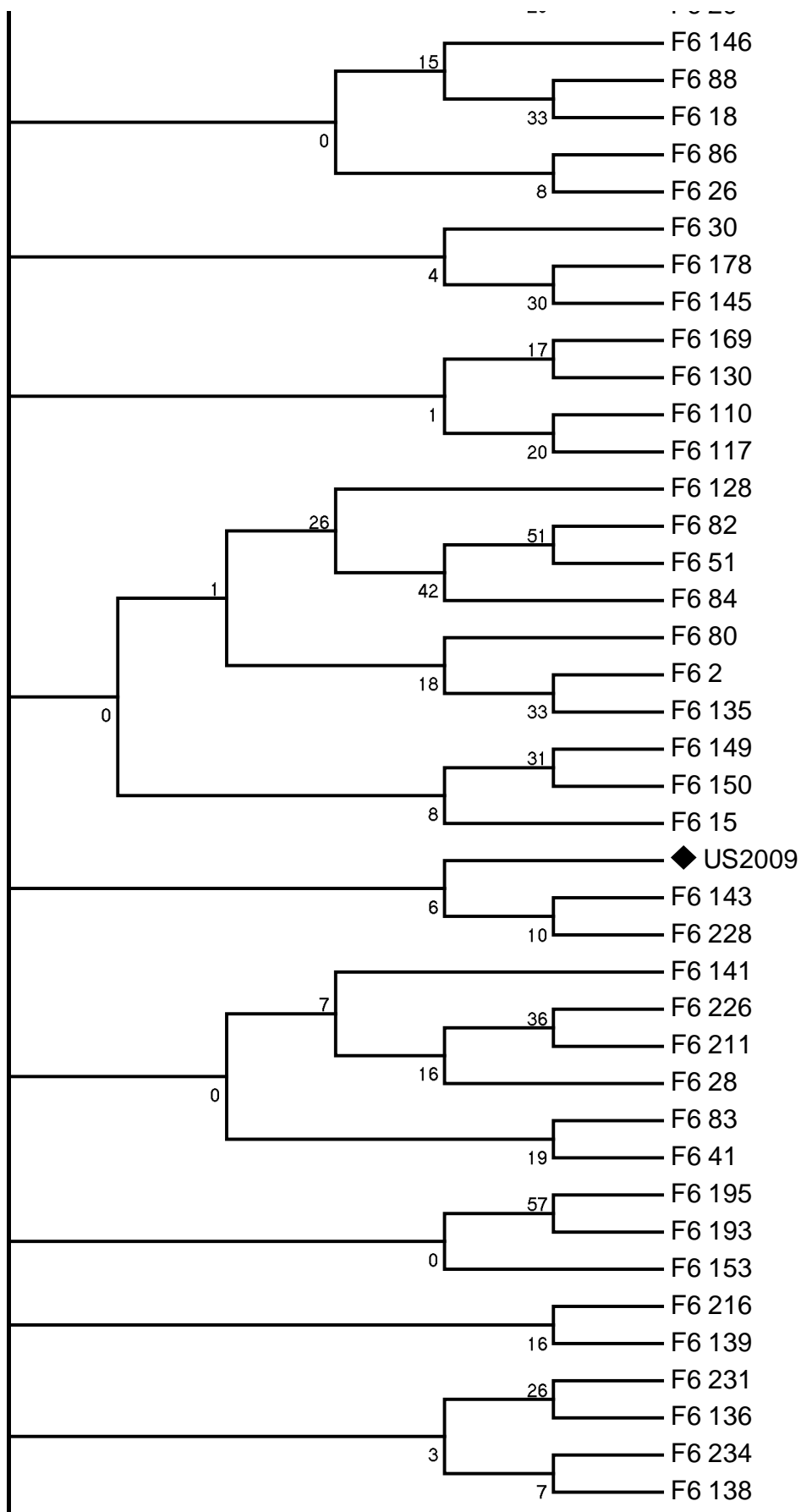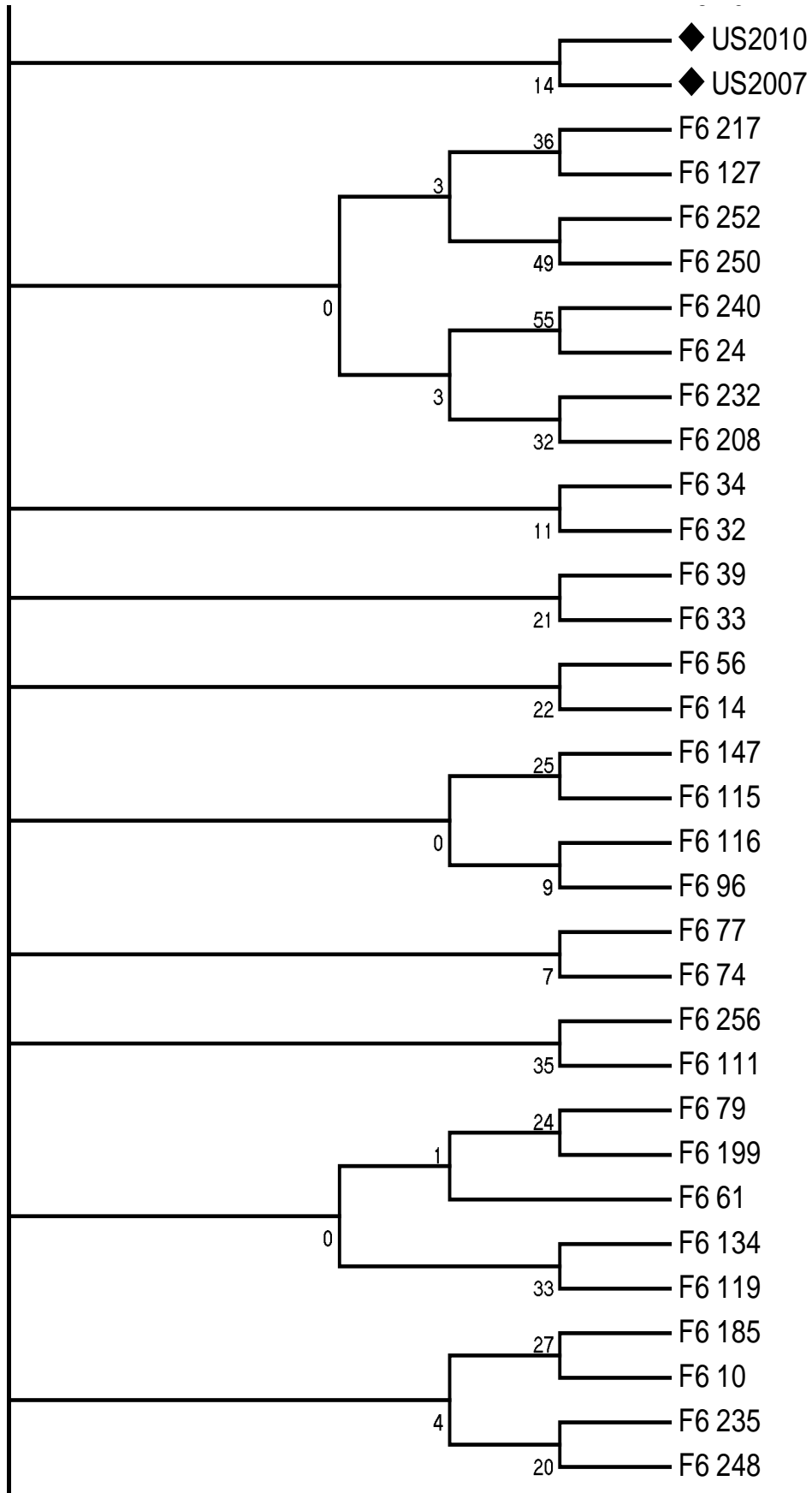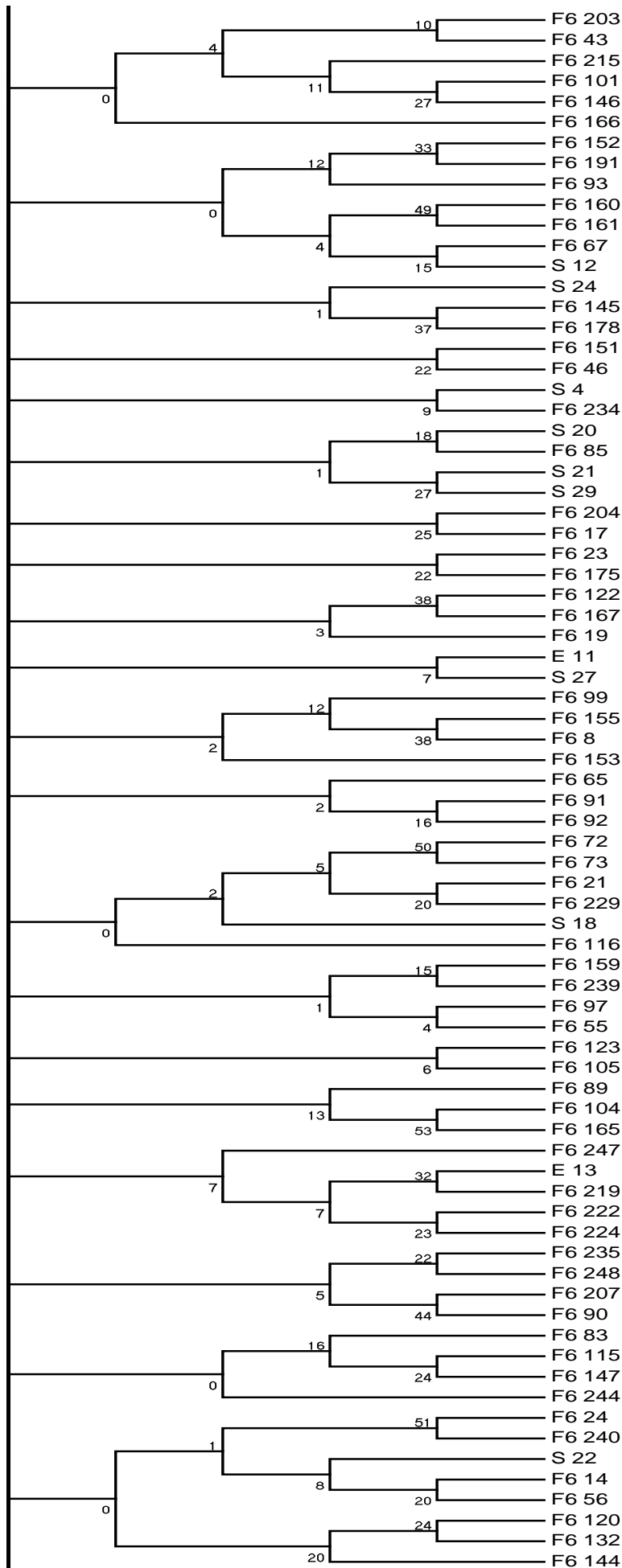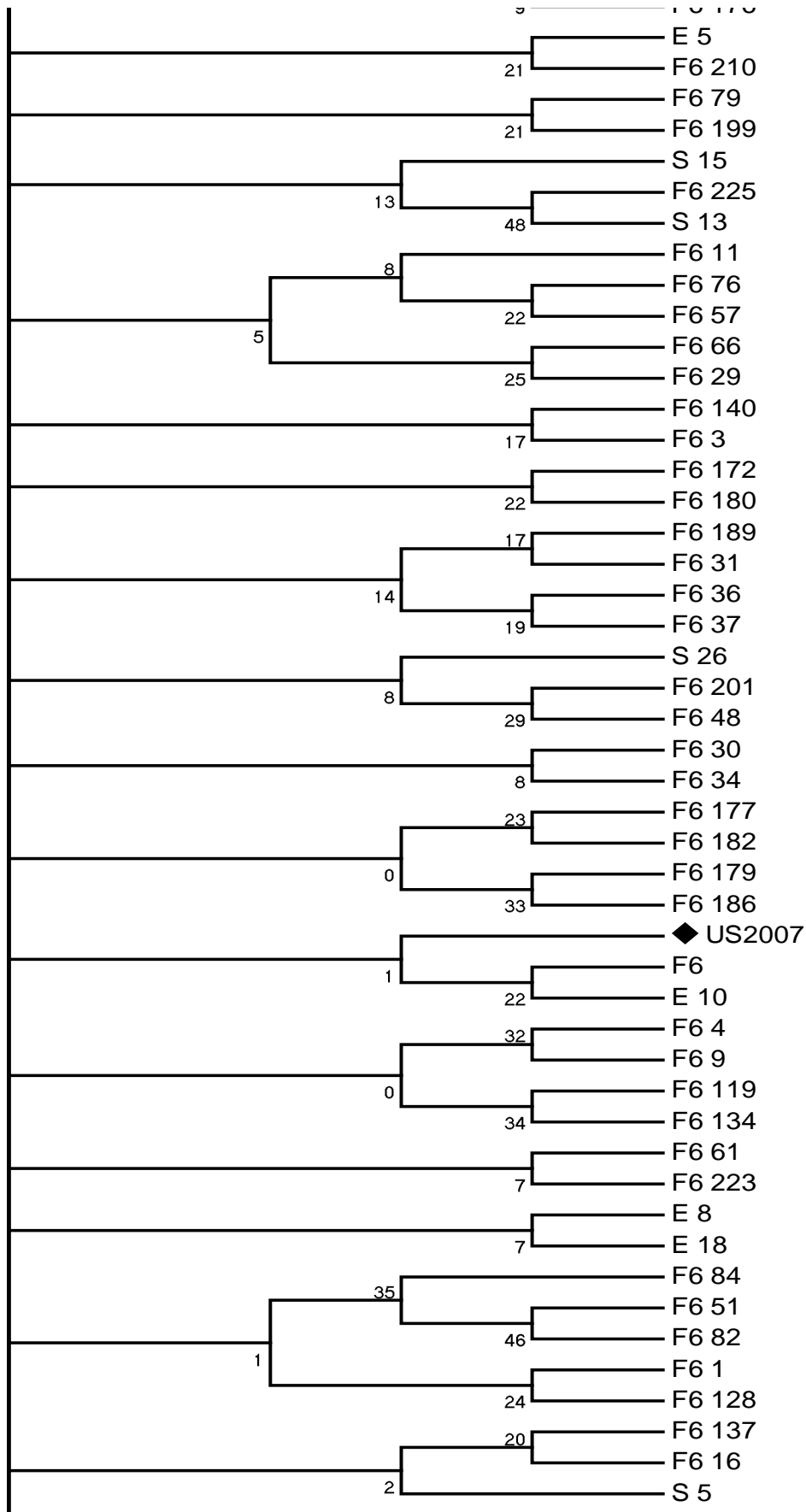
Figure 4.6 cont.

Figure 4.6 cont.

124
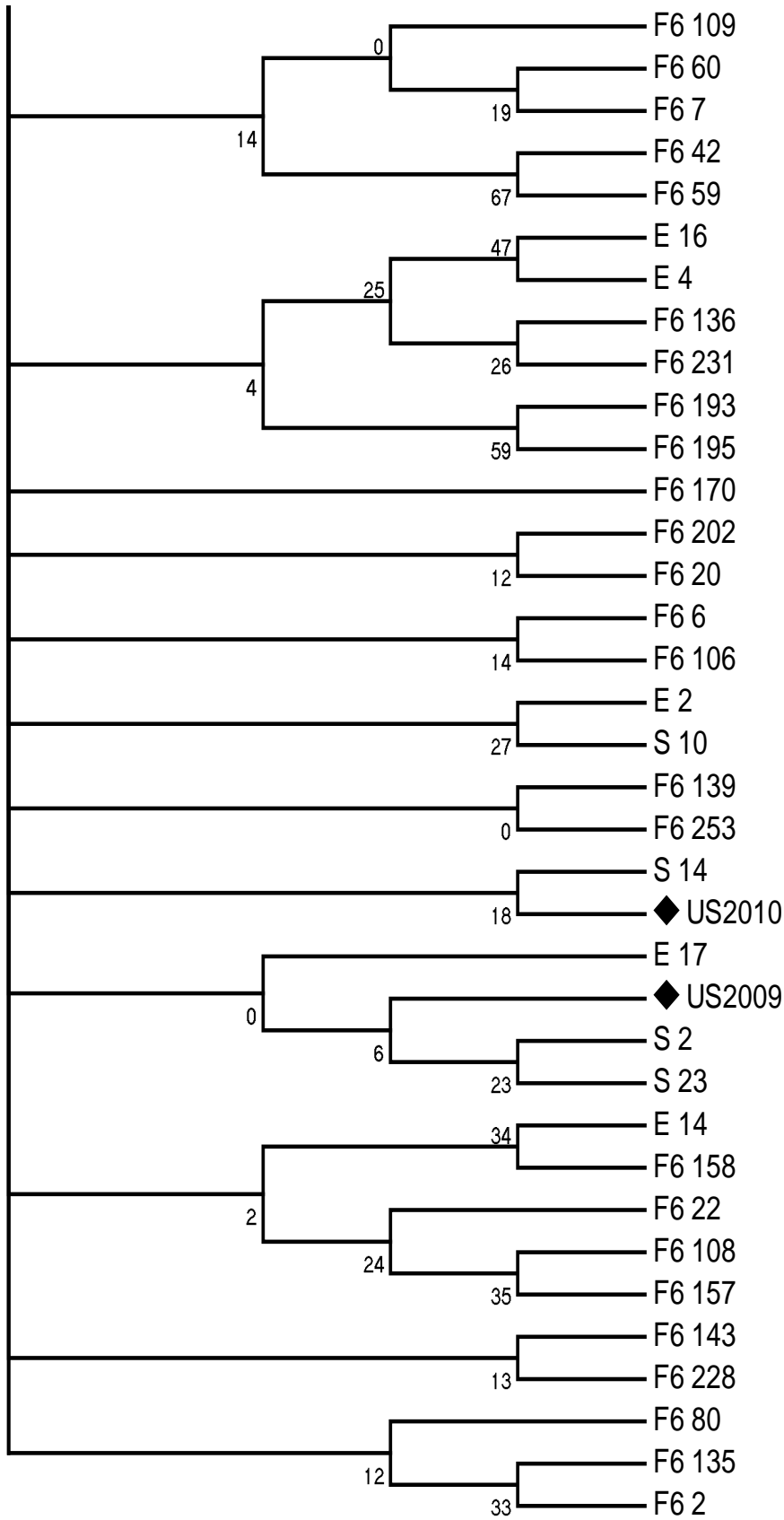
Figure 4.6 cont.
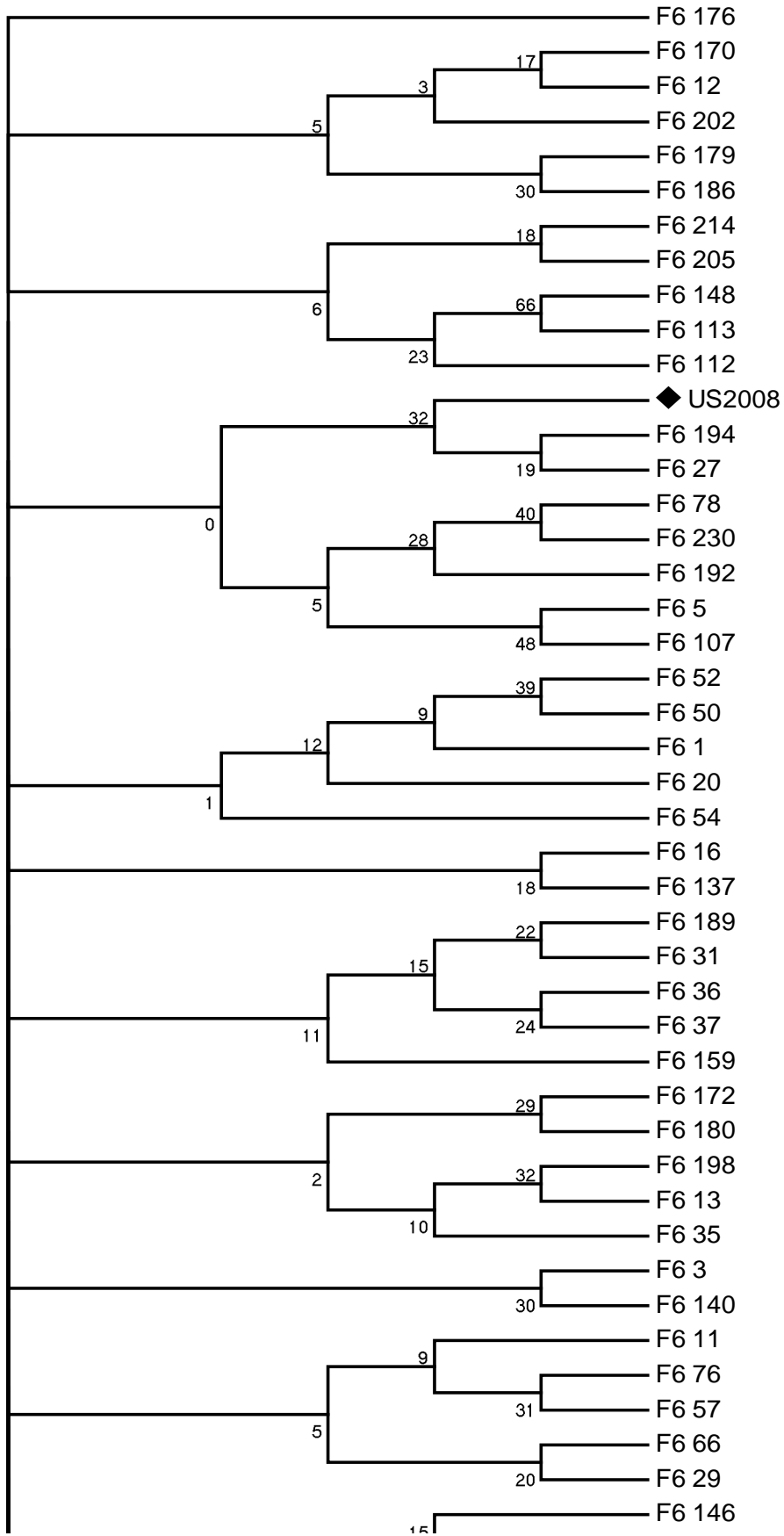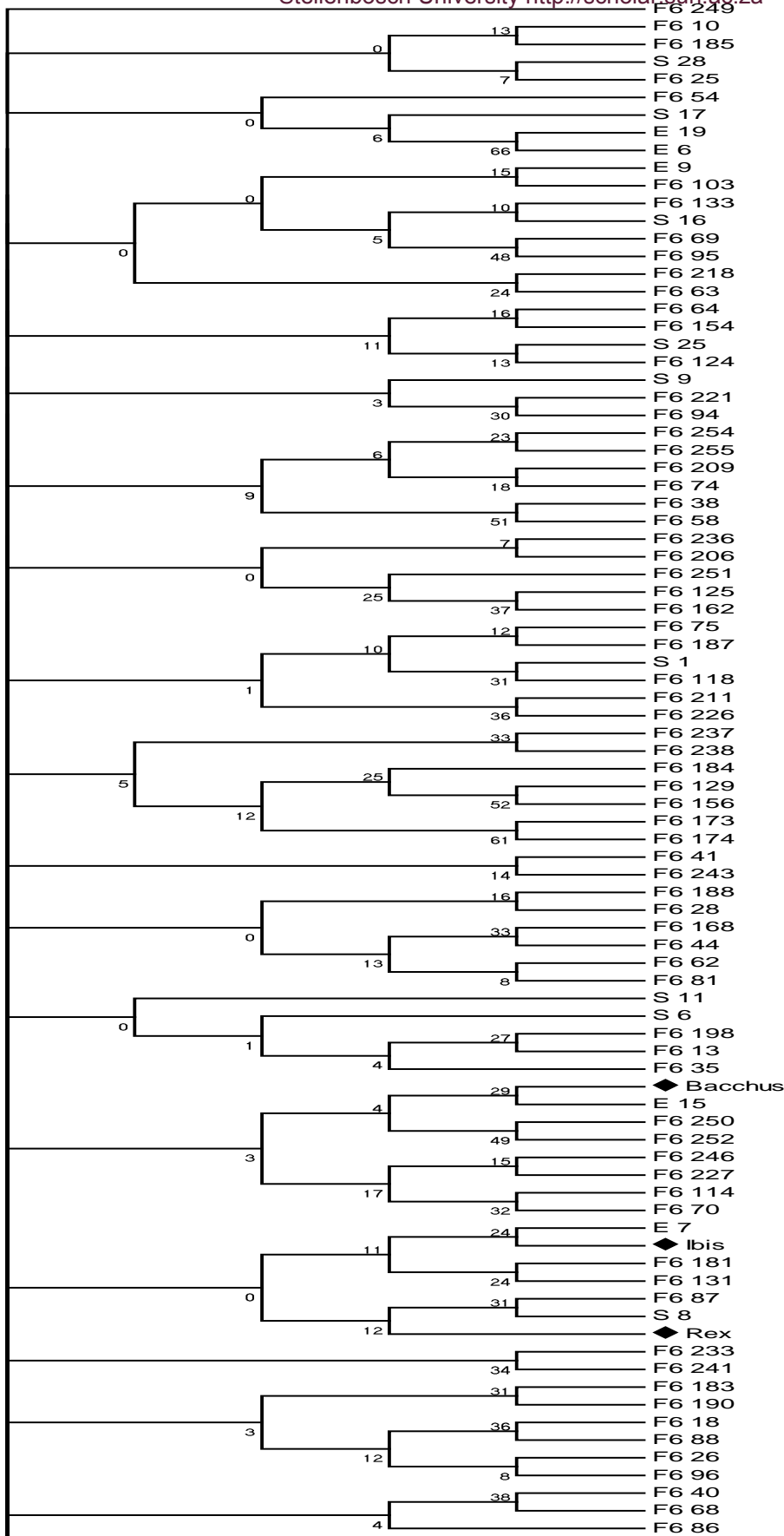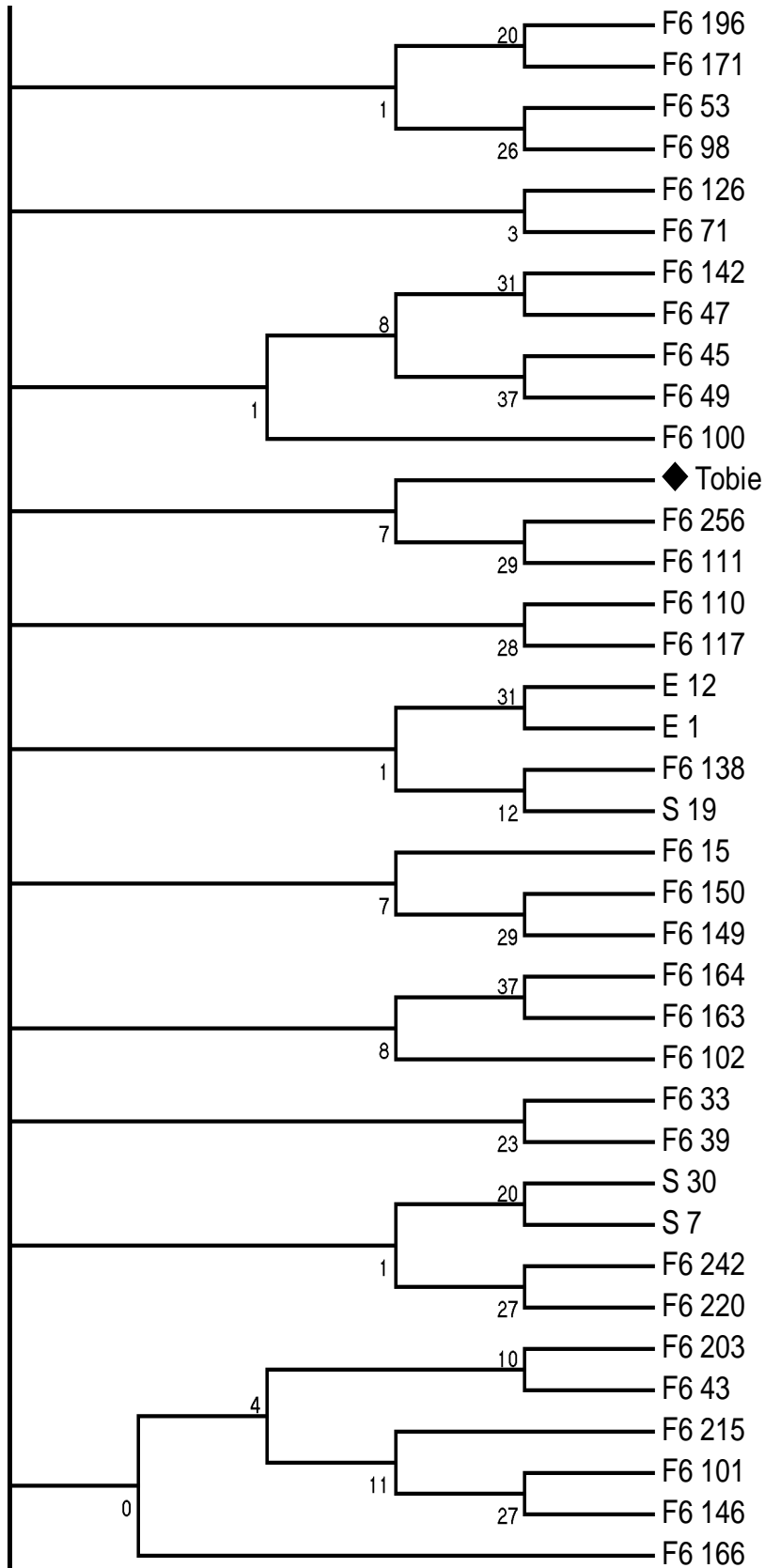
Finally a consensus UPGMA dendogram was generated for all the advanced breeding material from all three breeding phases (figure 4.6). This dendogram generated several small sub-clusters and there was no major groups observed among the advanced breeding material.

The small sub-clusters grouped the advanced breeding material (from all three breeding phases) and the Elite triticale cultivars together. This was done with the exception of the observed clustering of cultivars 'US2009' and 'US2010'. Therefore, the data generated by the seven SSR markers managed to show the genetic relationships among the advanced breeding material and the Elite cultivars.

Conclusion

The panel of SSR markers chosen in this study was found to be informative in the assessment of the genetic diversity of the advanced breeding material and in the subsequent identification of the Elite cultivars and advanced breeding material. Though this panel represents a very small coverage of the available SSR markers developed for both the wheat and rye genomes, it can be utilised effectively in conjuction with the available deposited microsatellite information in both wheat and rye microsatellite databases, to fingerprint the material at the SU-PBL season after season.

Effectively, the optimisation of the semi-automated technique allows for the easy assembly of multiplex combinations which improved the efficiency of the genetic analysis and cut the costs involved. For instance as is shown in Table 4.3, charges run up to 25 rand for a sample to be electrophoresed in one lane (capillary) on the ABI DNA sequencer. This charge worked out effectively because for each DNA sample screened, all seven SSR markers could be analysed due to the optimised multiplex PCR reactions developed in the study.

Using the ABI DNA sequencer was more cost effective when screening the many DNA samples taken from the breeding material as it only took a couple of hours from when the DNA was amplified to the generation of data in comparison to if the conventional PCR and gel electrophoresis technique had been used. Table 4.3 shows a comparative summary of the laboratory and economic aspects involved in both techniques.

The optimisation of the multiplex PCR ensured a consistency in PCR primer concentration and ensured a lowering in errors resulting from allele calling when analysing data using GeneMapper® version 4.0. Furthermore, all PCR amplicons underwent a post-PCR clean-up. This was done to eliminate the chances of contamination without having to change the

126

established PCR protocol. In conclusion, once properly optimised, the multiplex-PCR technique is a robust technique for high throughput analysis.

Table 4.3: A comparison of the laboratory and economic aspects involved in conventional and multiplex PCR techniques

	Conventional PCR	**Multiplex PCR**
Laboratory aspects		
Sample size	11/396	396
DNA quantity per reaction	3µl	1µl
PCR primers	Unlabelled	Labelled
Volume primers per reaction	0.5 µl of each primer	0.4µl of each primer
Technical steps	Many	Few
Project time	Days	Hours
Economic aspects*		
Cost per primer	R120 @ R0.24 per sample	R1,500 @ R6 per sample
KAPA2G™ Fast Multiplex Mix per sample		**R5.39**
Sequencing per lane		R25
Optimisation	yes	yes

*All economic aspects exclude the plastic materials and DNA extraction chemicals.

**All economic aspects were calculated during November 2011.

References:

- Carrano, A, V *et al.,* (1989) 'A high-resolution, fluorescence-based, semi-automated method for DNA fingerprinting', *Genomics*, 4: 129-136

- Coburn, J, R, Temnykh, S, V, Paul, E, M and McCouch, S, K (2002) 'Design and application of microsatellite marker panels for semiautomated genotyping of rice (*Oryza sativa* L.)', *Crop Science*, 42: 2092-2099

- Doyle, J, J and Doyle, J, L (1990) 'Isolation of plant DNA from fresh tissue', *Focus* 12: 13-15

- Liu, K and Muse, S, V (2005) 'PowerMarker: Integrated environment for genetic marker data', *Bioinformatics*, 21(9): 2128-2129

- Pessoa-Filho, M, Beló, A, Alcochete, A, A, N, Rangel, P, H, N and Ferreira, M, E (2007) 'A set of multiplex panels of microsatellite markers for rapid molecular characterization of rice accessions', *BMC Plant Biology*, 21: 7-23

Chapter 5: Conclusion

Molecular marker assessment

The initial assessment of molecular markers was done using triticale, wheat and rye cultivars sourced from the SU-PBL and Sensako (wheat). The development of the SSR markers though usually time consuming and laborious, was fastened in this study by using SSR markers developed in the parental genomes of triticale. The utilisation of these markers in triticale was cost effective, showed specific and repeatable banding patterns which were informative and generated reliable data. Overall, 84% of the 57 SSRs developed in wheat and 79.3% of the 29 SSRs developed in rye, produced amplicons in the triticale cultivars. Therefore, the assessment of the diversity of the triticale genome was assessed in totality as well as separately for the wheat and rye genomes, and this was done after confirming that SSR markers specific for the rye genome rarely amplify fragments in the wheat genome and *vice versa* (Röder *et al.,* 1995).

Microsatellite typing of the Elite cultivars showed a moderate genetic variability among the cultivars and this is thought to be due to the cultivars being developed from parents sharing the same pedigrees. Of note is the detection of higher polymorphisms by the SSR markers specific to the rye genome. This suggests that the SSRs developed in rye have a superior ability to identify cultivars in both the rye and triticale breeding programmes, and should be used as a routine diagnostic tool.

In addition, this study also managed to successfully implement EST-derived rye SSR markers. These markers also detected higher levels of polymorphism than the genomic markers from both the wheat and rye genomes. The EST markers used in this study were either linked with a known gene function or were closely linked to coding regions. This increased their efficiency in being transferred to the triticale species and also detection of genetic variation. Hence, combining the use of genomic and EST SSRs in the triticale breeding programme will be a valuable tool for assessing genetic diversity and identification of cultivars.

The utilisation of AFLP markers was time consuming and laborious. This is due to the generation of many bands by AFLP primer combinations, which when scored, on the PAGE gels leaves room for error. However, these markers were much cheaper to develop than SSR markers and also managed to amplify more loci per assay. *Eco*RI/*Mse*I primer combinations showed 61% transferability to rye and triticale in comparison to 75% showed by *Pst*I/*Mse*I

129

primer combinations. Both were however lower than the transferability of SSR markers to triticale. This difference in transferability is suggested to be as a result of the uniform distribution of the *Pst*I/*Mse*I primer combinations in the genome as opposed to the clustering of *Eco*RI/*Mse*I primer combinations around the centromere and the detection of variations in different parts of the genome by SSR and AFLP markers (Qi *et al.,* 1998; Haanstra *et al.,* 1999; Vuylsteke *et al.,* 1999).

Overall, the *Pst*I/*Mse*I markers have the ability to provide additional genetic information in the less covered distal regions of the rye genome (Saal and Wricke, 2002) while the *Eco*RI/*Mse*I primer combinations have the ability to provide genetic information in tightly linked regions that don't have SSRs (Bednarek *et al.,* 2003). The AFLP markers detected low to moderate levels of genetic variation among the cultivars.

Even though the AFLP markers showed an ability to discriminate than that of the SSR markers, they generated dendograms that did not correspond with data generated by both the rye and wheat genome specific SSR markers. Hence, the AFLP markers were not utilised in the further screening of the breeding material.

In conclusion, the first aim of the study was achieved and the SSR marker system was chosen for the second part of the study due to its reliability, ease of utilisation and cost.

Implementation of the SSR marker system for high throughput analysis

The second aim of the study was to optimise the most polymorphic marker system for semi-automated analysis in order to screen the rye and triticale breeding material. Hence, the SSR marker system was chosen, optimised and implemented in a high throughput environment.

The utilisation of the semi-automated analysis was more efficient and easier than using the conventional PAGE gel analysis, allowed for the multiple PCR amplification of primer sets, hence curbing costs, and was more accurate in terms of data generation and analysis. Overall, it was a successfully optimised for the high throughput genetic fingerprinting of the triticale material.

The seven SSRs developed in the rye and wheat genomes showed a moderate genetic variability among the F6, Seniors and Elite accessions of triticale. Again, it is highly likely that the breeding material screened share parents with similar pedigrees. The wheat SSRs detected slightly higher polymorphisms than the rye SSRs but this is slightly biased as there was an unbalanced coverage of all three genomes (only two rye genome SSRs were used).

130

Nevertheless, all seven SSRs had the ability to detect polymorphisms among the triticale breeding material and present a precise platform for the continuous fingerprinting of the triticale breeding material.

The AFLP marker technique used in this study did not successfully distinguish among the closely related commercial triticale and rye cultivars. However, the detection techniques and data analysis were conventional and should be further developed. For accurate and repeatable data analysis and generation, better allele scoring techniques should be applied. The *Pst*I/*Mse*I primer combinations used in this study have been noted to be more uniformly distributed across the genome while the *Eco*RI/*Mse*I primer combinations show clustering in specific regions not covered by SSR markers. Hence when assessing the genetic diversity among material in the SU-PBL, combining *Pst*I/*Mse*I and *Eco*RI/*Mse*I primers would be more ideal than using one over the other.

The utilisation of SSR markers developed in wheat and rye in the assessment of genetic diversity among triticale cultivars is more prudent as it allows for the distribution of at least one pair of primers per chromosome arm. Hence, more SSRs developed in rye should be added to any further genetic studies.

To gain complete insight into the diversity patterns of the lines, the utilisation of the pooled data from both wheat and rye SSRs, including those specific to the expressed regions (ESTs) of both genomes should be standard in order to provide a broader coverage of the triticale genome for further manipulations.

Overall, a minimal marker set (seven SSRs) that could accurately identify and distinguish material between the cultivars and advanced breeding material, was developed. This marker set also generated data that corroborated available pedigree data.

In future, this marker set can be used to:

- Distinguish advanced breeding material;
- Identify cultivars for the identification of parental heritage; and
- A tool in molecular-assisted backcross.

Future studies should exploit the ESTs developed in this study and other ESTs deposited in public domains to develop SNP markers for a wider diagnostic coverage of rye and triticale genomes.

References

- Bednarek, P, T, Masojc, P, Lewandowska, R and Mysków, B (2003) 'Saturating rye genetic map with amplified fragment length polymorphism (AFLP) and random amplified polymorphic DNA (RAPD) markers', *Journal of applied genetics*, 44(1): 21-33

- Feulliet, C, Langridge, P and Waughn, R (2007) 'Cereal breeding takes a walk on the wild side', *Trends in Genetics*, 24(1): 24-32

- Haanstra, J, P, Wye, W, C, Verbakel, H, Meijer-Dekens, F, van denBerg, P, Odinot, P, van Heusden, A, W, Tanksley, S, Lindhout, P and Peleman, J (1999) 'An integrated high-density RFLP–AFLP map of tomato based on two *Lycopersicon esculentum* X L. *pennellii* F2 populations', *Theoretical and Applied Genetics*, 99: 254-271

- Röder, M, S, Plaschke, J, Konig, S, U and Sorrells, M, S (1995) 'Abundance, variability and chromosomal location of microsatellite in wheat,' *Molecular and General Genetics*, 246: 327-333

- Saal, B and Wricke, G (2002) 'Clustering of amplified fragment length polymorphism markers in a linkage map of rye', *Plant Breeding,* 121: 117-123

- Qi, X, Stam, P and Lindhout, P (1998) 'Use of locus-specific AFLP markers to construct a high-density molecular map in barley', *Theoretical and Applied Genetics*, 96: 376-384

- Vuylsteke, M *et al.,* (1999) 'Two high-density AFLP linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers', *Theoretical and Applied Genetics*, 99: 921-935

Chapter 6: Appendix

This section is a compilation of the reagents and PCR components used during the study.

DNA extraction reagents:

TE buffer

10mM Tris-Cl (pH8)
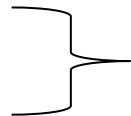
1mM EDTA (pH8)

To 100ml

Add 1ml of 1M Tris-Cl

Add 200μl of 0.5M EDTA

Make to volume (100ml) and autoclave

0.5M EDTA (100ml)

0.5M EDTA (100ml)

18.612g of EDTA

Add 80ml dH$_2$O

Set to pH8 using NaOH

Make to volume

Autoclave

1M Tris-Cl (250ml)

30.275g of Tris

Set to pH8 using HCl

Make to volume, autoclave

2% CTAB buffer (200ml)

2.4226g Tris-Cl

16.3632g NaCl

1.4890g EDTA

4g CTAB

Add 150ml dH$_2$O, set to pH8

Make to volume, autoclave

C:I solution

24 parts chloroform

1 part isoamylalcohol

5X TBE buffer (2 litres)

121.1g Tris (pH8)

7.44g EDTA

61.83g Boric acid

Table 6.1: SSR marker sequences (A & B genomes), annealing temperatures and repeat lengths

Marker name	Forward/ Reverse sequence (5'-3')	T_A (^{o}C)	Repeat
cfe2-6A,6B,6D F	CTTCGCCGACAAGAAGAAGT	60	$(CGC)_4$
cfe2-6A,6B,6D R	CGGCACGTACTCCACCTC		
cfe13-2AL,2D F	AAATCCAAGATGTGCCAAGG	60	$(CGG)_5$
cfe13-2AL,2D R	TCGCCGCCAACTACTACC		
cfe22-4A F	AGGACGTGAAGATCCATTGC	60	$(TTC)_4$
cfe22-4A R	GGTGGCTGGGAGACTATTGA		
cfe34-4BL,5A,5B F	ACAGCAGGCATCCACTATACG	60	$(TG)_8$
cfe34-4BL,5A,5B R	GCCTAGTTCGATGACAAGCA		
cfe37-5A,5B F	TCCTCGTCAACTACTGCGG	60	$(TACG)_3$
cfe37-5A,5B R	GGATGGATTGTAGATCATGCG		
cfe53-2AL F	TGGACCGCAGAGACTTCG	60	$(TC)_{10}$
cfe53-2AL R	GTCCGCCCAAACCCTACC		

Marker name	Forward/ Reverse sequence (5'-3')	T_A (^{o}C)	Repeat
cfe110-4A,4BL F	AGAGCCGAAATAGTCTCGCA	60	$(AC)_7$
cfe110-4A,4BL R	TGCCACCATCCTGAGCTAC		
cfe125-6A,6B F	AGGACTTGGCCCTAACGC	60	$(CAG)_6$
cfe125-6A,6B R	AACGATGGAACAAGGAAACG		
cfe132-6A F	ATATGTCGAGCTTCGGCG	60	$(GGC+GGC)_4$
cfe132-6A R	GGATTGAAACTGGCAAAGGA		
cfe 166-7A,7B F	ACCAGCTCAACAAAATGCG	59.8	$(CGT)_7$
cfe 166-7A,7B R	CACGTTCCACGCTACTACCA		
cfe260-7A,7B F	GAAGCCTCCGCTGCTAAAC	60.1	$(CCT)_7$
cfe260-7A,7B R	TGGAAACTGATACAAGGCAGC		
cfe270- 5A,5B F	CCCGTAGAAATGGTACTGGTG	59.4	$(GTG)_7$
cfe270- 5A,5B R	GGCATGGGCTCGCACTAC		
cfe274-4BL F	GACACACTTGCCGTGGAAC	60	$(CCA)_5$
cfe274-4BL R	GAAGATCACGAGGACGAAGC		
cfe284-7A F	CCCACTAGCCAGCCACTC	59.4	$(AATC)_4$

Marker name	Forward/ Reverse sequence (5'-3')	T_A (^oC)	Repeat
cfe284-7A R	TGCGGTACTATTCCACGACA		
cfe297-6A,6B F	TCCCTCGCTCGCTATGATT	60	$(CGG)_4$
cfe297-6A,6B R	GAAGTAGGAGCACACCTCGC		
Xbarc12-3A F	CGACAGAGTGATCACCCAAATATAA	53	$(TAA)_{28}$
Xbarc12-3A R	CATCGGTCTAATTGTCAATGTA		
Xbarc19-3A F	GCGACCCGAGTAGCCTGAA	58	$(TAA)_{18}$
Xbarc19-3A R	GGTGGACCATTAGACGCTTACTTG		
Xbarc37-6A F	CAGCGCTCCCCGACTCAGATCCTT	55	(TTA)
Xbarc37-6A R	GCGCCATGTTTCTTTTATTACTCACTTT		
Xbarc206-6A F	GCTTTGCCAGGTGAGCACTCT	61	$(CT)_{10}$
Xbarc206-6A R	TGGCCGGGTATTTGAGTTGGAGTTT		
Xbarc263-1AS F	GGAAGCGCGTCAGCACTAGGCAAC	55	
Xbarc263-1AS R	GGCTTCTAGGTGCTGCGGCT		
Xbarc1021-3A F	GGAAGGACCTGACTGACTGCATCTG	58	$(TAGA)_8$
Xbarc1021-3A R	GCGATCACAACCATTCTTTTTAACTA		

136

Marker name	Forward/ Reverse sequence (5'-3')	T_A (^{o}C)	Repeat
Xcfa2019-7AL F	GACGAGCTAACTGCAGACCC	60	
Xcfa2019-7AL R	CTCAATCCTGATGCGGAGAT		
Xcfa2153-1A F	TTGTGCATGATGGCTTCAAT	60	
Xcfa2153-1A R	CCAATCCTAATGATCCGCTG		
Xcfd58-1A,1D F	AATGGGCCTTTAAGAGCAAAA	55	$(CT)_5(CA)_{10}$
Xcfd58-1A,1D R	AGGGGTGAAAGGTTGGAGAC		
Xcfd79-3A F	TCTGGTTCTTTGGAGGAAGA	60	$(GA)_{26}$
Xcfd79-3A R	CATCCAACAATTTGCCCAT		
Xcn13-6BS F	AGAACAGTCTTCTAGGTTAG	50	
Xcn13-6BS R	CGAGGGACAGACGAATC		
Xgwm2-3A F	CTGCAAGCCTGTGATCAACT	52	$(CA)_{18}$
Xgwm2-3A R	CATTCTCAAATGATCGAACA		
Xgwm5-3A F	GCCAGCTACCTCGATACAACTC	50	$(TC)_{23}(T)_4(GT)_{12}(GA)_{10}$
Xgwm5-3A R	AGAAAGGGCCAGGCTAGTAGT		
Xgwm18-1B,4B F	TGGCGCCATGATTGCATTATCTTC	57	$(CA)_{17}GA (TA)_4$

137

Marker name	Forward/ Reverse sequence (5'-3')	T_A (^oC)	Repeat
Xgwm18-1B,4B R	GGTTGCTGAAGAACCTTATTTAGG		
Xgwm32-3A F	TATGCCGAATTTGTGGACAA	55	$(GA)_{19}$
Xgwm32-3A R	TGCTTGGTCTTGAGCATCAC		
Xgwm46-7BS F	GCACGTGAATGGATTGGAC	60	$(GA)_2GC(GA)_{33}$
Xgwm46-7BS R	TGACCCAATAGTGGTCA		
Xgwm99-1A F	AGGATGGACGTATGCATCACA	53	$(CA)_{21}$
Xgwm99-1A R	GCCATATTTGATGACGCATA		
Xgwm108-3B F	CGACAATGGGGTCTTAGCAT	55	$(GT)_{35}imp$
Xgwm108-3B R	TGCACACTTAAATTACATCCGC		
Xgwm135-1A F	TGTCAACATCGTTTTGAAAAGG	56	$(GA)_{20}$
Xgwm135-1A R	ACACTGTCAACCTGGCAATG		
Xgwm136-1A F	GACAGCACCTTGCCCTTTG	56	$(CT)_{58}$
Xgwm136-1A R	CATCGGCAACATGCTCATC		
Xgwm155-3A F	CAATCATTTCCCCCTCCC	53	$(CT)_{19}$
Xgwm155-3A R	AATCATTGGAAATCCATATGCC		

Marker name	Forward/ Reverse sequence (5'-3')	T_A (^{o}C)	Repeat
Xgwm160-4A F	TTCAATTCAGTCTTGGCTTGG	54	$(GA)_{21}$
Xgwm160-4A R	CTGCAGGAAAAAAAGTACACCC		
Xgwm165-4BL/S F	TGCAGTGGTCAGATGTTTCC	60	$(GA)_{20}$
Xgwm165-4BL/S R	CTTTTCTTTCAGATTGCGCC		
Xgwm285-3B F	ATGACCCTTCTGCCAAACAC	55	$(GA)_{27}$
Xgwm285-3B R	ATCGACCGGGATCTAGCC		
Xgwm340-3B F	GCAATCTTTTTTCTGACCACG	55	$(GA)_{26}$
Xgwm340-3B R	ACGAGGCAAGAACACACATG		
Xgwm357-1A F	TATGGTCAAAGTTGGACCTCG	55	$(GA)_{18}$
Xgwm357-1A R	AGGCTGCAGCTCTTCTTCAG		
Xgwm369-3A F	CTGCAGGCCATGATGATG	55	$(CT)_{11}(T)_2(CT)_{21}$
Xgwm369-3A R	ACCGTGGGTGTTGTGAGC		
Xgwm389-3B F	ATCATGTCGATCTCCTTGACG	55	$(CT)_{14}(GT)_{16}$
Xgwm389-3B R	TGCCATGCACATTAGCAGAT		
Xgwm429-2B F	TTGTACATTAAGTTCCCATTA	50	$(CT)_{25}$

Marker name	Forward/ Reverse sequence (5'-3')	T_A (^oC)	Repeat
Xgwm429-2B R	TTTAAGGACCTACATGACAC		
Xgwm493-3B F	TTCCCATAACTAAAACCGCG	60	$(CA)_{43\ imp}$
Xgwm493-3B R	GGAACATCATTTCTGGACTTTG		
Xgwm513-4B F	ATCCGTAGCACCTACTGGTCA	55	$(CA)_{12}$
Xgwm513-4B R	GGTCTGTTCATGCCACATTG		
Xgwm550-1B F	CCCACAAGAACCTTTGAAGA	54	$(CT)_8 (GT)_{18}$
Xgwm550-1B R	CATTGTGTGTGCAAGGCAC		
Xgwm666-1A F	GCACCCACATCTTCGACC	60	$(CA)_{13}$
Xgwm666-1A R	TGCTGCTGGTCTCTGTGC		
Xrems1135-1R,3R,7R F	AGTCTTTTGGAGGTGCATCAG	54	$(GA)_6$
Xrems1135-1R,3R,7R R	CACATCAATCCTCGCTGCTA		
Xrems1162-7R F	TATGACCTGGTGAGGTTCG	54	$(GCC)_5$
Xrems1162-7R R	CGTCTTGAAAGGCTCGTTGT		
Xrems1167-5R F	TGTAAACGACCAGTCCTGAT	54	$(CGG)_5$
Xrems1167-5R R	GAAGGAATTTGGCAGCAGAG		

Marker name	Forward/ Reverse sequence (5'-3')	T_A (°C)	Repeat
Xrems1197-7R F	TGTTCGTCTACGGCTCCATC	54	$(CGC)_5$
Xrems1197-7R R	CAACTCCCACATGGGTCTG		
Xrems1264-5R F	TGTAACGAGGATGAAGAGCG	54	$(CGTC)_5$
Xrems1264-5R R	TCACCTTCTTCTTCGCCAAC		
Xrems1266-5R F	ACGACGGCAGTGAGAGAGAG	54	$(GA)_8$
Xrems1266-5R R	TCGGCTTCATCGTCTACTCC		
Xwmc11-1A F	TTGTGATCCTGGTTGTGTTGTGA	56	(CT)
Xwmc11-1A R	CACCCAGCCGTTATATATGTTGA		
Xwmc25-2B,2D F	TCTGGCCAGGATCAATATTACT	50	$(GT)_{26}$
Xwmc25-2B,2D R	TAAGATACATAGATCCAACACC		
Xwmc59-1A,6A F	TCATTCGTTGCAGATACACCAC	58	$(CA)_{19}$
Xwmc59-1A,6A R	TCAATGCCCTTGTTTCTGACCT		
Xwmc167-2A F	AGTGGTAATGAGGTGAAAGAAG	52	N/A
Xwmc167-2A R	TCGGTCGTATATGCATGTAAAG		
Xwmc169-3A F	TACCCGAATCTGGAAAATCAAT	54	$(CA)_{25}$

Marker name	Forward/ Reverse sequence (5'-3')	T_A (^{o}C)	Repeat
Xwmc169-3A R	TGGAAGCTTGCTAACTTTGGAG		
Xwmc177-2A F	AGGGCTCTCTTTAATTCTTGCT	52	$(CA)_{21}$
Xwmc177-2A R	GGTCTATCGTAATCCACCTGTA		
Xwmc254-1A F	AGTAATCTGGTCCTCTCTTCTTCT	51	$(AC)_{28}...(AC)_{27}$
Xwmc254-1A R	AGGTAATCTCCGAGTGCACTTCAT		
Xwmc256-6A F	CCAAATCTTCGAACAAGAACCC	56	$(CA)_{12}$
Xwmc256-6A R	ACCGATCGATGGTGTATACTGA		
Xwmc532-3A F	GATACATCAAGATCGTGCCAAA	56	$(GA)_{11}$
Xwmc532-3A R	GGGAGAAATCATTAACGAAGGG		

Table 6.2: SSR marker sequences (R genome), annealing temperatures and repeat length

Marker name	Forward/ Reverse sequence (5'-3')	T_A (^{o}C)	Repeat
SCM2-6RL F	GATGACTATGACTACCAGGATGAA	55	$(GT)_{10}$
SCM2-6RL R	GGAGTGAGAAGGCCGAGAAG		
SCM4-1R F	AAGGTGTTGCCTCCATGTTC	57	$(AC)_8$
SCM4-1R R	TTGATTCCGTTGGACTTGAA		
SCM5-3RL F	TCGCGATACATCAAGATCGTG	50	$(GA)_{16}$
SCM5-3RL R	CTAGCATCGACGTAACCCTTT		
SCM9-1RS F	TGACAACCCCCTTTCCCTCGT	60	$(GT)_8$
SCM9-1RS R	TCATCGACGCTAAGGAGGACCC		
SCM36-1R F	TCCTCGCTCTCTCCTCCTTC	59	$(AGC)_6$
SCM36-1R R	CATGAAGATGTCGCTGTTGC		
SCM38-2R F	TGACCTGCACACCTCATCTCA	59	$(AGC)_6$
SCM38-2R R	GCTTTGCTCCTGATGTCGAT		
SCM39-1R F	GACCTCAGTGGAGCCTCTAGGT	60	$(GT)_8(GC)_6\ldots(GT)_{53}$
SCM39-1R R	GGACATCTGCCGTGACAATACC		

143

Marker name	Forward/ Reverse sequence (5'-3')	T_A (^{o}C)	Repeat
SCM40-7R F	CCCTTCAGCGGTCATTGTTG	60	$(GT)_{18}$
SCM40-7R R	CACATCTTGGGCCTGACACC		
SCM43-2R F	CTAGGGGATTACAGGGAGGGCA	60	$(GT)_{11}$
SCM43-2R R	GTTCCCTTGTCCTACTCGTTACCG		
SCM66-3R F	TGTAAGCTTCTCCATGTTCCCC	59	$(CCG)_7$
SCM66-3R R	GAGCACGTGGGAGATGAAGT		
SCM69-2R F	CTACCTGCTGTTCCCATTTGG	60	$(CA)_{10}\ldots(CA)_7$
SCM69-2R R	GTGTGTAGAAGATGTTGTCCTGG		
SCM75-2RL F	TTTTCTATCTCAGCGATTCATGC	60	$(CA)_7(CT)_{15}\ldots(CA)_{10}$
SCM75-2RL R	TCCTGAGATCAAGTGCGTGTG		
SCM80-1R F	TATCTGTGCGCGTCCTTTTAC	59	$(CGG)_5$
SCM80-1R R	AGTATCGCGGACCTGAACTG		
SCM83-2R F	TGTTCCCCTTTGCTCCTTTTC	59	$(CGG)_5$
SCM83-2R R	CAACACGCCACCAACAATAC		
SCM86-7R F	CAGATAGATGGGTGTTGTGCG	60	$(GT)_{20}$

Marker name	Forward/ Reverse sequence (5'-3')	T_A (^oC)	Repeat
SCM86-7R R	CTCTTCTCGACATCCACACTCC		
SCM109-5RL F	AACCCCCTTTCGTACCTTGT	60	$(GT)_9$
SCM109-5RL F	TAAAGCAAACCACCAGAGCC		
SCM112-3R F	TACGTACAATGGGTGCAAATG	56	$(GGC)_5$
SCM112-3R R	TTTGAACTTTCCAGGGCATC		
SCM120-5RL F	CATTGTTGCGAGTGTTGAAGC	60	$(AC)_{10}$
SCM120-5RL R	TGTGCTGTCGTCGATGTTGTC		
SCM152(F1)-4R	TAAAACGACGGCCAGTGACGA	68	$(AG)_7$
SCM152(F2)-4R	ACGGCCAGTGGAGCAGCAGCAG		
SCM152-4R R	ATGTAGCCGAGGATGGTGAG		
SCM159(F1)-4R	AAGAGCCAGTTTGGACTTGGAG	68	$(GAAA)_5$
SCM159(F2)- 4R	CGGCCAGTGGTTCCTTGGAT		
SCM159(F2)- 4R R	CGGGAAGGAAAAACAGAAAACT		
SCM206-3RS F	TCATCGAAAGACGGGACACCT	60	$(TA)_{28}GA(CA)_9$
SCM206-3RS R	ACGATTGCTGCTAATAGCCATGTG		

Marker name	Forward/ Reverse sequence (5'-3')	T_A (^oC)	Repeat
Xrems1135-1R,3R,7R F	AGTCTTTTGGAGGTGCATCAG	54	$(GA)_6$
Xrems1135-1R,3R,7R R	CACATCAATCCTCGCTGCTA		
Xrems1162-7R F	TATGACCTGGTGAGGTTCG	54	$(GCC)_5$
Xrems1162-7R R	CGTCTTGAAAGGCTCGTTGT		
Xrems1167-5R F	TGTAAACGACCAGTCCTGAT	54	$(CGG)_5$
Xrems1167-5R R	GAAGGAATTTGGCAGCAGAG		
Xrems1197-7R F	TGTTCGTCTACGGCTCCATC	54	$(CGC)_5$
Xrems1197-7R R	CAACTCCCACATGGGTCTG		
Xrems1264-5R F	TGTAACGAGGATGAAGAGCG	54	$(CGTC)_5$
Xrems1264-5R R	TCACCTTCTTCTTCGCCAAC		
Xrems1266-5R F	ACGACGGCAGTGAGAGAGAG	54	$(GA)_8$
Xrems1266-5R R	TCGGCTTCATCGTCTACTCC		

PCR reagents:

Table 6.3: AFLP restriction-ligation reactions using EcoRI/MseI and PstI/MseI enzymes and adaptors

Reagent	Vol/rxn	Final conc	Reagent	Vol/rxn	Final conc
gDNA (100ng/µl)	3	300ng	gDNA (100ng/µl)	3	300ng
ddH$_2$O	9.6		ddH$_2$O	9.6	
One-Phor-All buffer (10X)	2	1X	One-Phor-All buffer (10X)	2	1X
ATP (10mM)	2	1mM	ATP (10mM)	2	1mM
*EcoR*I adoptor (10nM)	1	10nM	*Pst*I adoptor (10nM)	1	10nM
*Mse*I adaptor (10nM)	1	10nM	*Mse*I adaptor (10nM)	1	10nM
*EcoR*I restriction enzyme (10U/µl)	0.5	5U	*Pst*I restriction enzyme (10U/µl)	0.5	5U
*Mse*I restriction enzyme (10U/µl)	0.5	5U	*Mse*I restriction enzyme (10U/µl)	0.5	5U
Bovime Serum Albumin (10mg/ml)	0.2	0.1µg/µl	Bovime Serum Albumin (10mg/ml)	0.2	0.1µg/µl
T$_4$ DNA ligase (5U/µl)	0.2	1U	T$_4$ DNA ligase (5U/µl)	0.2	1U
Total volume	20		Total volume	20	

Table 6.4: AFLP restriction-ligation reaction using PstI and MseI enzymes and adaptors

Reagent	Vol/rxn	Final con
gDNA (100ng/µl)	3	300ng
ddH$_2$O	9.6	
One-Phor-All buffer (10X)	2	1X
ATP (10mM)	2	1mM
*Pst*I adoptor (10nM)	1	10nM
*Mse*I adaptor (10nM)	1	10nM
*Pst*I restriction enzyme (10U/µl)	0.5	5U
*Mse*I restriction enzyme (10U/µl)	0.5	5U
Bovime Serum Albumin (10mg/ml)	0.2	0.1µg/µl
T$_4$ DNA ligase (5U/µl)	0.2	1U
Total volume	20	

Table 6.5: Pre-amplification using pre-selective primers E00/M00 and M01/P01

Reagent	Vol/ rxn	Final conc	Reagent	Vol/ rxn	Final conc
Diluted restriction-ligation DNA	5.2		Diluted restriction-ligation DNA	5.2	
ddH$_2$O	8.9		ddH$_2$O	8.9	
PCR buffer (10X)	2	1X	PCR buffer (10X)	2	1X
dNTPs(2.5mM)	1.6	0.2mM	dNTPs(2.5mM)	1.6	0.2mM
E00 primer (10nM)	1.5		P01 primer (10nM)	1.5	
M00 primer (10nM)			M01 primer (10nM)		
MgCl$_2$ (50mM)	0.6	1.5mM	MgCl$_2$ (50mM)	0.6	1.5mM
Taq(5U/µl)	0.2	1U	Taq(5U/µl)	0.2	1U
Total volume	20		Total volume	20	

Table 6.6: Amplification using selective EcoRI/MseI primers and PstI/MseI primers

Reagent	Vol/rnx	Final conc	Reagent	Vol/rnx	Final conc
Diluted pre-selective DNA	2.5		Diluted pre-selective DNA	2.5	
ddH$_2$O	4.8		ddH$_2$O	4.8	
Buffer (10X)	1	1X	Buffer (10X)	1	1X
dNTPs (2.5nM)	0.8	0.2mM	dNTPs (2.5nM)	0.8	0.2mM
*Mse*I primer (10µM)	0.5	0.5µM	*Mse*I primer (10µM)	0.5	0.5µM
*EcoR*I primer (10µM)	0.05	0.05µM	*Pst*I primer (10µM)	0.05	0.05µM
MgCl$_2$ (50mM)	0.3	1.5mM	MgCl$_2$ (50mM)	0.3	1.5mM
Taq (5U/µl)	0.05	0.25U	Taq (5U/µl)	0.05	0.25U
Total volume	10µl		Total volume	10µl	

Table 6.7: PCR reaction for conventional SSR primer sets

Reagent	Vol/rnx	Final conc
DNA(100ng/µl)	3	300ng
ddH$_2$O	10.6	
Buffer (10X)	2	1X
dNTPs (2.5mM)	1.6	0.2mM
Forward primer (10µM)	1	0.5µM
Reverse primer (10µM)	1	0.5µM
MgCl$_2$ (50mM)	0.6	1.5mM
Taq (5U/µl)	0.2	1U
Total volume	20µl	

Table 6.8: PCR reaction for labelled SSR primer sets

Reagent	Vol/rxn	Final conc
DNA(100ng/µl)	1	100ng
ddH$_2$O	3.4	
KAPA2G multiplex kit (2X)	10	1X
cfe2 F (10µM)	0.4	0.2µM
cfe2 R (10µM)	0.4	0.2µM
cfe53 F (10µM)	0.4	0.2µM
cfe53 R (10µM)	0.4	0.2µM
Xcfd79 F (10µM)	0.4	0.2µM
Xcfd79 R (10µM)	0.4	0.2µM
SCM38 F (10µM)	0.4	0.2µM
SCM38 R (10µM)	0.4	0.2µM
Xgwm108 F (10µM)	0.4	0.2µM
Xgwm108 R (10µM)	0.4	0.2µM
Xgwm369 F (10µM)	0.4	0.2µM
Xgwm369 R (10µM)	0.4	0.2µM
Xrems1266 F (10µM)	0.4	0.2µM
Xrems1266 R (10µM)	0.4	0.2µM
Total volume	20	

Gel eletropherosis reagents:

1.5% agarose gel

1.275g agarose

85ml 1X TBE

Microwave till dissolved

Cool before casting on gel tray

40% Acrylamide (200ml)

76g acrylamide

4g bisacrylamide

Make up to 150ml and dissolve

Make up to volume

Cover flask with foil, store at 4°C

6% Gel mix (500ml)

75ml of 40% acrylamide

181.8g urea

100ml TBE (5X)

Dissolve and make up to volume

Fixing Solution (2.1 litres)

210ml ethanol

10.5ml acetic acid

Make to volume with dH_2O

Staining Solution (2.1 litres)

2.1g $AgNO_3$

Developing solution (2.1 litres)

31.5g NaOH