EVALUATION AND IMPLEMENTATION OF DNA-BASED DIAGNOSTIC METHODOLOGY TO DISTINGUISH WHEAT GENOTYPES

by

Jennifer Honing

Presented in partial fulfilment of the requirements for the degree of Master of Science at the Department of Genetics, University of Stellenbosch.

December 2007

Supervisor: Prof GF Marais

Co-Supervisor: Mr WC Botes

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

_____ Date: _____

Jennifer Honing

ABSTRACT

The aim of this study was to develop a DNA-based diagnostic system that can be used to distinguish between genotypes in the wheat breeding program at the University of Stellenbosch. Known marker systems were investigated and the chosen marker system would then be implemented to determine its utility in the breeding program.

Three marker systems were considered, i.e. microsatellites, Amplified Fragment Length Polymorphisms (AFLPs) and various retrotransposon-based markers. Each system is based on polymerase chain reaction (PCR) amplification from specific primer pairs. The multitude of primer options was narrowed down during a review of published literature regarding wheat molecular markers. Thirty nine microsatellite primer pairs and nine AFLP primer combinations were chosen for the initial genotype evaluation. Four different retrotransposon-based techniques were investigated; namely Inter-Retrotransposon Amplified Polymorphism (IRAP), REtrotransposon-Microsatellite Amplified Polymorphism (REMAP), Sequence-Specific Amplified Polymorphism (SSAP) and, a derivative of these developed in this study, *Wis-2* Retrotransposon Amplification.

The study started with twenty genotypes which included varieties/breeding lines from five breeding programmes. The genotypes were chosen as representative of the respective breeding populations and were used in the initial testing of the marker systems. Eighteen microsatellites were evaluated using the panel of twenty genotypes. From this, six primer pairs (*Xgwm190*, *Xgwm437*, *Xgwm539*, *Xwmc11*, *Xwmc59* and *Xwmc177*) were chosen to test the semi-automated DNA sequencer detection system. A single band/peak in each microsatellite profile was used for genotyping. Four of the primer pairs were labelled with different fluorochromes which enabled them to be multiplexed. The differences in amplification products of the six microsatellites meant that all six could be detected in one electrophoresis run.

The banding pattern produced by microsatellite *Xwmc177* was complex and highly polymorphic and was therefore also analysed in the same way as the AFLP patterns. When analyzed in this manner it proved to be more informative than the combination of six microsatellites (with a single prominent band scored in each).

Three AFLP primer combinations could also be multiplexed and visualised together. The three *Eco*RI selective primers were labelled with different dyes and used with one *Mse*I selective primer. The SSAP system also used fluorescently labelled primers and proved to be the most useful of the retrotransposon-based methods. However, this system produced such a large amount of data that it made analysis too time consuming. Therefore the six microsatellites and three AFLP primer combinations (*Mse*I-CTC and *Eco*RI-ACA, -AAC, -AGG) were selected for routine genotyping. Due to the numerous highly polymorphic bands produced by the SSAP system it could be very useful to differentiate very closely related genotypes that cannot be distinguished with the markers proposed for routine use.

A panel of 119 breeding lines were then used to implement the two chosen marker systems. The results obtained for these markers were used to produce a dendrogram of the lines using the SAS cluster analysis function. The clusters showed that most of the lines could be distinguished from each other. The *Mse*I-CTC and *Eco*RI-AGG primer combination was the most informative. It produced the largest number of clusters (53) and could therefore discriminate between more of the lines than any other method.

The dendrograms and clusters allowed sixteen of the breeding lines to be selected to test the optimal number of seeds to represent an entire population (variety/breeding line) as one seed was not sufficient. It was decided that eight seeds could provide a good representation of the intra-line variability.

OPSOMMING

Die doel van hierdie studie was om ‘n DNS-gebaseerde diagnostiese sisteem te ontwikkel wat gebruik kan word om koringgenotipes uit die Universiteit van Stellenbosch se teelprogram te onderskei. Nadat bekende merkersisteme ondersoek is, is daar op 'n gepaste sisteem besluit wat daarna geimplementeer en ge-evalueer is.

Drie merkersisteme is oorweeg, d.i. mikrosatelliete, ge-Amplifiseerde-Fragment-Lengte-Polimorfismes (AFLPs) en verskeie retrotransposon-gebaseerde merkers. Elke sisteem is gebaseer op polimerase-ketting-reaksie- (PKR-) amplifikasie van spesifike inleier-pare. Die menige moontlike inleierstelle is eers tot ’n werkbare aantal verminder deur gepubliseerde literatuur rakende koring molekulêre merkers te raadpleeg. Nege-en-dertig mikrosatelliet inleier-pare en nege AFLP inleier-kombinasies is vir die aanvanklike genotipering gekies. Vier verskillende retrotransposon-gebaseerde tegnieke is ondersoek; naamlik Inter-Retrotranposon-ge-Amplifiseerde-Polimorfisme (IRAP), REtrotransposon-Mikrosatelliet-ge-Amplifiseerde-Polimorfisme (REMAP), Volgorde-Spesifike-ge-Amplifiseerde-Polimorfisme (VSAP) en, ’n alternatiewe protokol wat in hierdie studie ontwikkel is, nl. *Wis-2* Retrotransposon Amplifikasie.

Die studie het aanvanklik op twintig genotipes gefokus wat variëteite/teellyne van vyf teelprogramme ingesluit het. Die genotipes is gekies as verteenwoordigend van die teelpopulasies en is gebruik in die aanvanklike toetsing van die merkersisteme. Agtien mikrosatelliete is met behulp van die paneel van twintig lyne ge-evalueer. Hieruit is ses pare inleiers (*Xgwm190*, *Xgwm437*, *Xgwm539*, *Xwmc11*, *Xwmc59* en *Xwmc177*) gekies om die semi-ge-outomatiseerde DNS-volgorde-bepaler waarneming-sisteem te toets. ’n Enkel band/piek in elke mikrosatelliet-profiel is vir genotipering gebruik. Vier van die inleierpare is met verskillende fluorochromes gemerk vir gelyktydige waarneming. Verskille in amplifikasie-produkte van die ses mikrosatelliete het beteken dat almal in een elektroforese-lopie gedoen kon word.

Die bandprofiel van mikrosatelliet *Xwmc177* was kompleks en hoogs polimorfies en is daarom ook op dieselfde wyse as die AFLP-profiele ontleed. Die volle *Xwmc177* profiel was meer informatief as die kombinasie van ses mikrosatelliete (met gebruik van ’n enkel prominente band in elk).

Drie AFLP inleierkombinasies kon in een reaksie geamplifiseer en gevisualiseer word. Die drie *Eco*RI selektiewe inleiers is met verskillende kleurstowwe gemerk en saam met een *Mse*I selektiewe inleier gebruik. Die VSAP-sisteem het ook fluoresserende inleiers gebruik en was die nuttigste retrotransposon-gebaseerde metode. Maar, hierdie sisteem het soveel data geproduseer dat dit omslagtig was om te ontleed. Die ses mikrosatelliete en drie AFLP inleierkombinasies (*Mse*I-CTC en *Eco*RI-ACA, -AAC, -AGG) blyk daarom die nuttigste te wees vir roetine genotipering. Omdat die VSAP sisteem soveel hoogs polimorfiese bande produseer kan dit egter nuttig wees om baie nabyverwante genotipes, wat nie onderskeibaar is met die roetine-merkers nie, te onderskei.

'n Paneel van 119 teellyne is gebruik om die twee gekose roetine merkersisteme te evalueer. Die verkreë resultate is gebruik om 'n dendrogram van die lyne saam te stel met gebruik van die SAS tros-analise funksie. Die groeperings het gewys dat meeste van die lyne van mekaar onderskei kon word. Die *Mse*I-CTC en *Eco*RI-AGG inleierkombinasie was die mees informatiewe. Dit het die grootste aantal groeperings (53) geproduseer en kon dus meer van die lyne onderskei as enige van die ander metodes.

Die dendrogramme en groeperings is gebruik om sestien teellyne te kies waarmee die optimale aantal indiwidue vir betroubare genotipering bepaal kon word aangesien een indiwidu nie genoeg is nie. Aanduidings is dat agt sade per genotipe 'n goeie monstering van die intra-lyn variasie verseker.

ABBREVIATIONS

A	Adenine
AFLP	Amplified Fragment Length Polymorphism
ALP	Amplicon Length Polymorphism
AP-PCR	Arbitrarily-Primed PCR
AS-PCR	Allele-Specific PCR
ATP	Adenosine triphosphate
β-ME	β-Mercaptoethanol
bp	Base pairs
BSA	Bovine serum albumin
C	Cytosine
CAPS	Cleaved Amplified Polymorphic Sequences
cDNA	Complementary DNA
cm	Centimetre
CPVO	Community Plant Variety Office
CTAB	Cetyl trimethylammonium bromide
°C	Degrees Celsius
DAF	DNA Amplification Fingerprinting
DArT	Diversity Array Technology
DGGE	Denaturing Gradient Gel Electrophoresis
ddH_2O	Double distilled water
dH_2O	Distilled water
DIG	Digoxigenin
DNA	Deoxyribonucleic Acid
dNTPs	Deoxyribonucleotidetriphosphate
DUS	Distinctness-uniformity-stability
*Eco*RI	Restriction enzyme isolated from *Escherichia coli* strain R
EST-SSR	Expressed Sequence Tag derived Simple Sequence Repeat
EtBr	Ethidium bromide
F	Forward
FISSR-PCR	Fluorescent Inter-Simple Sequence Repeat PCR
G	Guanine
g	Grams
gDNA	Genomic Deoxyribonucleic Acid

ha	Hectare
*Hae*III	Restriction enzyme isolated from *Haemophilus aegyptius*
Hz	Hertz
IRAP	Inter-Retrotransposon Amplified Polymorphism
ISA	Inter-simple Sequence repeat Amplification
ISF	International Seed Federation
ISSR-PCR	Inter-Simple Sequence Repeat PCR
λ	Lambda
LINE	Long Interspersed Element
LTR	Long Terminal Repeat
M	Molarity
Mb	Mega bases
$MgCl_2$	Magnesium Chloride
µg/µl	Microgram per microlitre
µg/ml	Microgram per millilitre
µl	Microlitre
µM	Micromolar
mins	Minutes
ml	Millilitre
mM	Millimolar
mm	Millimetre
*Mse*I	Restriction enzyme isolated from *Micrococcus* sp.
n	Haploid
2n	Diploid
Na_2EDTA	Disodium ethylenediaminetetraacetic acid
NaCl	Sodium Chloride
ng	Nanograms
ng/µl	Nanogram per microlitre
$(NH_4)_2SO_4$	Ammonium sulphate
PAGE	Polyacrylamide gel electrophoresis
PBR	Plant breeders' rights
PCR	Polymerase Chain Reaction
pH	Percentage hydrogen
*Pst*I	Restriction enzyme isolated from *Providencia stuartii 164*
QTL	Quantitative Trait Loci
R	Reverse

RAPD	Random Amplified Polymorphic DNA
REMAP	Retrotransposon-Microsatellite Amplified Polymorphism
RFLP	Restriction Fragment Length Polymorphism
RGAP	Resistance Gene Analogue Polymorphism
RNA	RiboNucleic Acid
rpm	Revolutions Per Minute
SAS	Statistical Analysis System
SCAR	Sequence Characterized Amplified Region
SGI	Small Grain Institute
SINE	Short Interspersed Element
SNP	Single Nucleotide Polymorphism
SPLAT	Single Polymorphic Length Amplification Test
SSAP	Sequence-Specific Amplified Polymorphism
SSR	Simple Sequence Repeat
STS	Sequence-Tagged Sites
T	Thymine
TBE	Tris-Borate-EDTA
TE	Tris-EDTA
TEMED	N,N,N,N- Tetramethylenediamine
Tris	2-amino-2-hydroxymethylpropane-1,3-diol
U	Units
UPOV	Union Internationale pour la Protection des Ortention Végétales
US	University of Stellenbosch
USA	United States of America
UV	Ultraviolet
V	Volts
v.v	Volume for volume
W	Watts
w/v	Weight per volume
Xbarc	Beltsville Agriculture Research Center microsatellite
Xcfd	Pierre Sourdille microsatellite
Xgwm	Gatersleben wheat microsatellite
Xwmc	Wheat microsatellite consortium

ACKNOWLEDGEMENTS

I would like to thank the following people:

- Prof GF Marais for his guidance, insight and constant patience,
- Mrs AS Marais for her technical assistance and support,
- Mrs Sadie for her help with the analysis of the data,
- Harry Crossley for funds,
- Everybody in Lab 239 and the Genetics Department (Stellenbosch University) for all the laughs and shared knowledge,
- Aletta Eksteen for her help, friendship and shoulders to lean on,
- My family, especially my dad and mom for their never ending love and funds and for always supporting me,
- Geoffrey Thompson for his love and allowing me to be his scientist,
- God, who makes everything possible.

CONTENTS

1. INTRODUCTION

The purpose of the study was to develop and evaluate a DNA-based technique that can be used to distinguish between different common or bread wheat (*Triticum aestivum*) genotypes. Bread wheat is a hexaploid plant that has a comparatively limited gene pool, thus, identifying different genotypes based on heterogeneity in its DNA is restricted. There are numerous examples in literature where organisms of the same species were identified based on their DNA. This was the starting point of this study. Examples of variety identification or discrimination were researched for ideas of what to use to distinguish bread wheat genotypes.

Three different types of molecular markers were investigated: Microsatellites, Amplified Fragment Length Polymorphisms and Retrotransposons. Each of these was tested on a population of twenty plants from five different breeding programmes. The markers were evaluated on the following factors: level of polymorphism, cost of the method, time of the method and how the results could be analysed. The information was then used to propose a methodology that can be used routinely for genotyping of new elite material developed in the Stellenbosch University wheat breeding program. Finally, the procedure was evaluated by applying it to a larger population of breeding lines.

2. LITERATURE REVIEW

2.1. COMMON WHEAT

Wheat is the most extensively cultivated crop in the world. It occupies about 225 million ha, which constitutes about 17% of the cultivated land in the world. About 1.2 million ha is allocated to wheat cultivation in sub-Saharan Africa. Bread wheat is the staple food for 35% of the world's population. It provides 55% of the carbohydrates and 20% of the calories consumed around the world. This is more than any other crop plant, which includes rice, maize and potatoes (Evenson & Gollin 2003; Gupta *et al*. 1999; Hart 2001; Huang *et al*. 2003).

It is a very important source of food for so many that it needs to be protected. Any of its many pathogens can destroy a vast portion of a crop very quickly after infection. As is the case with many crop species, wheat is also susceptible to drought and is affected by many other environmental factors. Therefore potential wheat cultivars have to be tested extensively to find the best type for each geographical area. This means that once a farmer or breeder has found a cultivar that can perform really well in a particular environment it is wise to maintain and protect the cultivar. Identification is the first and most important step in this regard.

2.1.1. Genetics and origin

Common wheat, *Triticum aestivum*, is a self-fertilizing, allohexaploid (2n = 6x = 42) with the A, B and D genomes. Each genome consists of seven chromosomes, thus, there are seven homoeologous groups; each group has three chromosomes, one from each genome. This adds up to approximately 16 billion base pairs of nuclear genome, which is about 810 Mb per chromosome. However, almost 80% of this genome consists of repetitive DNA which makes genome-wide studies very difficult (Gao *et al*. 2004; Gupta *et al*. 1999; Gupta *et al*. 2002; Huang *et al*. 2003; Knott 1989; Leigh *et al*. 2003b; Röder *et al*. 1998).

The wheat genome is characterized by a low level of genetic polymorphism which can be explained by the way in which hexaploid wheat originated. About 5000 to 6000 years ago, allohexaploid bread wheat (*T. aestivum;* 2n = 42, AABBDD) came into existence. It developed from the natural hybridisation of a tetraploid wheat, *T. turgidum* (2n = 28, AABB),

and a diploid wheat, *T. tauschii* (2n = 14, DD). This hybridisation event probably involved a limited number of genotypes from each species. This caused an evolutionary bottleneck which is responsible for hexaploid wheat's limited genetic base relative to its wild ancestors. It possibly caused some important genes to be absent from the hexaploid wheat gene pool (Fritz *et al*. 1995; Pestsova *et al*. 2000). However, many research groups have been able to overcome this limitation in polymorphism and found differences (particularly at the DNA level using molecular markers) between wheat genotypes. Identifying differences in the DNA sequences of plants makes it possible to analyze them at any stage of their development and very quickly (Talbert *et al*. 1998).

2.1.2. Degree and measurement of variability

The level of DNA variation in a population is determined by the following factors: recombination and mutation in the genome, migration and selection in the population, a change in population size and population subdivision, random genetic drift, etc. Change in population size is generally given as the reason for the limited amount of variation in wheat. This is based on the assumption that a single, bottleneck event occurred during the origin of hexaploid wheat. Thus, it is the limited amount of variation taken from its ancestors and the fact that it was formed about 10 000 years ago that predict a very small amount of variation. Wheat does, however, have abundant genetic variation for certain traits, such as: winter versus spring growth pattern, response to day length, cold hardiness, insect and disease resistance, etc. The wide adaptability and variation of hexaploid wheat has caused many authors to suggest that wheat might actually have originated from many hybridization events involving different genotypes of its ancestors (Talbert *et al*. 1998).

In addition to varietal identification, genotyping also finds application in wheat breeding programmes. It might be easy to measure genetic differences between plants, but the totality of differences is very limited in wheat and other cultivated plants. Conventional breeding programmes together with current agricultural and social practices have caused a remarkable loss in genetic diversity in many cultivated species. The narrowing genetic base of elite germplasm has increased their potential vulnerability to pests and diseases. Better knowledge of the genetic similarity of the breeding material could help to preserve genetic diversity and extend long-term selection gain. Assessing the genetic diversity in the gene pool of elite

breeding material could improve germplasm management, speed up crop improvement and make it more efficient (Huang *et al*. 2002; Li *et al*. 2001).

The commercial and manufacturing fields have also found a need for identifying plants. The emphasis is on how fast, reliable and cost effective the high throughput technique for plant identification could be. McIntosh *et al*. (2005) found a conserved single copy nuclear gene in the grass family that could be used for grass fingerprinting. PCR primers were designed for this region and they found that each grass species had a different polymorphism. Therefore these primers could be used to identify the presence of different grass species, such as rice, barley, maize and wheat, in a mixed genomic DNA sample. The main function of this protocol would be to identify any contaminants or multiple sources in processed products. This will help to ensure the purity and quality of all grain products.

2.1.3. Types and uses of common wheat

Common wheat can be classified in many ways. One such distinction is between spring and winter types. Spring wheat does not have a cold requirement. It is planted in spring and allowed to grow into summer. This type of crop produces a rather low grain yield and therefore, is grown to a lesser extent than winter wheat throughout the world. Winter wheat is planted late in the summer or early in autumn. Germination takes place and growth continues until there is a small wheat plant. During winter the plant is dormant and growth only continues in spring. Better soil moisture and an extended growing period result in higher yields than is the case with spring wheat, but only where the soil does not freeze (Shellenberger 1971).

It is also customary to differentiate between hard and soft wheat. This refers to grain texture. Wheat types with hard kernels are used to make flour used in the production of bread. Softer kernel types are used to make flour for cakes, cookies and pastries (Giroux *et al*. 1998). Other foods that are made from common wheat include: macaroni, breakfast cereals, gluten (used for the production of flavour extenders), beer, etc. Wheat is also used as animal and poultry feed and is currently being investigated for use in the production of bio-ethanol (Kim & Dale 2004).

2.1.4. Plant breeders' rights (PBR)

Plant breeders' rights (PBR) are the intellectual property rights granted to the breeder of a new plant variety. These rights allow a breeder to charge royalties on the sale of that plant material especially if it was bought to be multiplied. The Community Plant Variety Office (CPVO) allocates these rights in Europe. Two other organisations that issue these rights are the International Union for the Protection of New Varieties of Plants (Union Internationale pour la Protection des Ortention Végétales [UPOV]) and the International Seed Federation (ISF) (Weising *et al*. 2005).

In order to obtain PBR the new variety has to pass the distinctness-uniformity-stability (DUS) requirements. It has to be unique from all other varieties that have already been described, it has to be uniform so that all individuals are as identical as possible, and the variety has to be stable so that it stays true to its particular traits over successive generations. The DUS characteristics are assessed in trials conducted over one to numerous years. During this time morphological traits are investigated and compared between new and old varieties. Molecular markers could aid the DUS test by determining distinctness and deciding which of the existing varieties should be used as references in future tests (Weising *et al*. 2005).

2.2. MARKERS

There are three main types of marker that can be used in plant breeding. These are: morphological markers, protein markers and DNA markers (Chahal & Gosal 2002). Each has specific characteristics, applications, advantages and disadvantages. DNA markers are more versatile than the other marker types and test plants based on their DNA, which can be done at any stage of the plant's development. This allows researchers to analyse the plants quicker and easier (Talbert *et al*. 1998). Non-DNA markers include isozymes, storage proteins and phenotypic markers. Generally, the latter markers are not very useful because of their limited number and that they are influenced by the environment (Sud *et al*. 2005).

2.2.1. Morphological markers

Mutations in morphological marker loci (plant height, presence of awns, etc) can clearly be seen in the phenotype of the plant and the loci involved are indicated on many linkage maps.

The only problem with this marker type is that there are a limited number of them which means that only a small portion of the genome can be checked for its contribution to the complex characters that are visible (Chahal & Gosal 2002).

2.2.2. Protein markers

Protein markers are associated with the gene product. Different alleles produce different proteins therefore this marker system is more informative than the morphological marker. The most common example of protein markers is isozymes. This term refers to variations of an enzyme that have the same catalytic activity but different molecular weights/conformations/charges and are affected differently by an electric field in a resolving medium. Differences observed in the mobility of isozymes are mostly caused by point mutations in the genes that produced them. The mutations cause amino acids to be substituted and result in differences in the electric charge of the enzymes. Isozymes have a codominant nature allowing for discrimination between heterozygotes and homozygotes. The main advantages of this marker type are that they are codominant, easy to use and cheap to test. Major disadvantages are their limited number of loci, need for fresh tissue and often limited variation (Chahal & Gosal 2002; Weising *et al*. 2005).

Allozymes are alternative types of an enzyme that are produced by alleles of the same genes (Stenesh 1989; Walker 1989). These variants are also known as allelozymes and have different amino acid sequences (Kahl 2004). Thus, an allozyme is a specific category of isozyme and its variation therefore has the same basis, nature and application. Similar to isozymes, allozymes are not very polymorphic but they make up for this by being affordable and quick to use (Mueller & Wolfenbarger 1999).

2.2.3. DNA markers

The term DNA marker refers to a small region of DNA that shows sequence polymorphism in different individuals within a species or group of individuals. Base pair changes in DNA are much more frequent than rearrangements, and polymorphisms in the base sequence at DNA level are a lot more frequent than changes in the charge of proteins. About 30 to 90 percent of DNA in organisms consists of non-coding areas. This means that many mutations can occur in these areas causing no phenotypic effect to the organism. The ultimate progression to

studying DNA polymorphisms would be to sequence the whole genome of wild type individuals and then identify any changes. This unfortunately would be very time and money consuming (Chahal & Gosal 2002).

There are three groups of DNA markers. The first group is the hybridization-based DNA markers which include: Restriction Fragment Length Polymorphisms (RFLPs) and oligonucleotide fingerprinting. The second is PCR-based DNA markers, including: Random Amplified Polymorphic DNAs (RAPDs), Sequence Characterized Amplified Regions (SCARs), Simple Sequence Repeats (SSRs) or microsatellites, Sequence-Tagged Sites (STS), Amplified Fragment Length Polymorphisms (AFLPs), Inter-simple Sequence repeat Amplification (ISA), Cleaved Amplified Polymorphic Sequences (CAPS) and Amplicon Length Polymorphisms (ALPs). The last group includes DNA chips and sequencing-based DNA markers such as Single Nucleotide Polymorphisms (SNPs). These are not the only DNA marker techniques, but this covers the three main groups. There are many subdivisions of these types (Gupta *et al*. 1999).

Weising *et al*. (2005) made a list of eleven traits that would be desired in a molecular marker. This list contains many things that will be considered and focused on in this study. The list includes:

1. moderate to high levels of polymorphism;
2. codominant inheritance (allows homo- and heterozygotes to be distinguished in diploid organisms);
3. allele assignment should be unambiguous;
4. frequent occurrence in the genome;
5. evenly distributed throughout the genome;
6. selectively neutral behaviour;
7. easy access for purchasing and fast procedures;
8. easy and fast assay (e.g. automated detection);
9. highly reproducible;
10. easy data sharing between laboratories;
11. low cost for both marker development and assay.

Certain characteristics are common to all DNA markers. These include: their abundance in the genome, their independence of tissue and the fact that they are not influenced by the environment. Consequently, DNA markers are useful for the physiological and morphological characterisation of genotypes, variety identification, tracking traits in breeding and generating both physical and genetic chromosome maps (Manifesto *et al*. 2001; Perry 2004; Rampling *et al*. 2001).

Several DNA markers were compared in Avise (2004). This comparison is given in Table 2-1 and considers some of the traits listed above. It looks at each marker's ability to generate qualitative markers at a particular locus in a population. The table illustrates the differences and characteristics of DNA markers. It is clear from this comparison that each marker system has its own qualities and functions in a population study.

Table 2-1 Comparison of specific molecular marker techniques based on the given criteria (adapted from Avise 2004; Mueller & Wolfenbarger 1999).

CRITERIA	Allozymes	RAPD	RFLP	SSR (Microsatellites)	AFLP
Number of loci assayed	Many	Many	Few	Several	Many
Many alleles identifiable per locus?	Yes	No	Yes	Yes	No
Repeatability of assays	High	Variable	High	High	High
Resolution of genetic differences	Moderate	Moderate	High	High	High
Nature of markers	Codominant	Dominant	Codominant	Codominant	Dominant
Ease of use and development	Easy	Easy	Difficult	Difficult	Moderate
Laboratory development time	Short	Short	Long	Long	Short
Quantity of information	Low	High	Low	High	High

2.2.4. PCR-based DNA markers

The techniques used in this study are all PCR- (polymerase chain reaction) based. Target DNA is denatured, primers are annealed and then the fragment is extended. This cycle is repeated many times causing mass amplification of the fragment. This fragment can be visualised by using ethidium bromide (EtBr) and ultraviolet (UV) light, fluorescently labelled primers with detection on an automated sequencer or polyacrylamide gel electrophoresis and silver staining (Brown 2002; Sambrook *et al*. 1991; Weising *et al*. 2005).

Any technique that will be used to identify (fingerprint) or distinguish wheat cultivars/lines has to be rapid, consistent and reliable. PCR methods should be applicable and due to its sensitivity, very small amounts of DNA samples can be employed. PCR methodology is simple and can be automated, which means that results can be obtained quicker and at a low cost (Ko *et al*. 1994). The cost of these reactions can be further reduced when using automated product detection. The PCR primers can be labelled with different colours and the reactions can then be multiplexed to reveal more polymorphisms from one sample than previously possible. This also allows for a higher throughput of samples and more reliable results (Blair *et al*. 2002).

Random Amplification of Polymorphic DNA (RAPD) is a PCR reaction that amplifies a DNA sequence between random primers. The Sequence Characterized Amplified Region (SCAR) marker is a RAPD-derived molecular marker that was converted to overcome some of the problems experienced with RAPD markers. SCAR markers are more reliable than RAPDs and can also be visualised on an agarose gel (Chahal & Gosal 2002). SCAR development is not limited to RAPD fragments. Cloning and sequencing has also been applied to AFLP fragments resulting in their conversion to SCAR markers (Shan *et al*. 1999; Weising *et al*. 2005).

Amplified Fragment Length Polymorphism (AFLP) is a method that combines elements from the other techniques to detect polymorphisms between DNA sequences. The DNA sample is cut with restriction enzymes and adaptors are ligated to the fragments. Preselective primers that bind to the adaptors are used to amplify these fragments and then specific primers with one, two or three selective bases on the 3' end are used to selectively amplify fragments, resulting in a simpler banding pattern (Kalendar *et al*. 1999).

The Resistance Gene Analogue Polymorphism (RGAP) technique targets the coding region of the genome. Primers are designed that target and amplify differences in the conserved domains that are found in resistance genes. Therefore these primers can be used between species and in various combinations to produce informative results. The banding pattern obtained using these primers are very similar to that seen in AFLPs (Chen *et al*. 1998; Diaz & Ferrer 2003; Weising *et al*. 2005).

Another PCR technique amplifies SSRs (short sequence repeats) or microsatellites using primers developed around the repetitive sequence. The most powerful PCR-based marker system is AFLP, which has a high multiplexing ratio (Waugh *et al*. 1997). Another marker type is one that uses PCR primers to target and amplify retrotransposons. These primers are used in a variety of techniques in combination with other primers. For example, retrotransposon primers are used with microsatellite primers in the REMAP (Retrotransposon-Microsatellite Amplified Polymorphism) technique. Another technique, SSAP (Sequence-Specific Amplified Polymorphism), combines AFLP selective primers with primers that target the long terminal repeat (LTR) of a retrotransposon (Weising *et al*. 2005).

Other DNA marker techniques that are used to identify polymorphic genetic loci and are mostly PCR-based include: Arbitrarily-Primed PCR (AP-PCR), Allele-Specific PCR (AS-PCR), DNA Amplification Fingerprinting (DAF), Denaturing Gradient Gel Electrophoresis (DGGE), Single Strand Conformational Polymorphism (SSAP) and Single Polymorphic Length Amplification Test (SPLAT) (Chahal & Gosal 2002).

2.2.5. DNA markers in plant breeding

According to Chahal & Gosal (2002) there are eight possible uses for DNA markers in plant breeding. The first of these is in the construction of high density genetic maps. This will allow genes to be detected based on their position in the map using DNA markers rather than looking at their phenotypic effect. The phenotype of a crop does not illustrate the genetic structure of the plant completely. The genetic map also helps to narrow down the area of interest when looking for a specific gene.

The second application is in comparative gene mapping. This involves using different plant species to determine gene order and evolutionary relationships between the species. Many similarities have been observed between the locus order in wheat, rice, barley and sorghum. Thus DNA markers can be used to construct a map of the common ancestor of the cereal genomes.

The next application is in marker-assisted selection. This allows new varieties to be developed and released quicker and more efficiently. Time is not wasted waiting for the crop to grow and mature so that its phenotypic traits can be observed. Very small plants can be

tested and the desired plants can be selected and used in further studies. The plants with unwanted traits, such as susceptibility to diseases, can be removed and/or replaced early-on.

DNA markers also aid in the transfer of genes from wild germplasm to cultivated varieties. Marking the desired genes will help to ensure that unwanted DNA can be eliminated early and only the desired genes can be retained. Often very large sections of DNA are co-transferred from a wild species when only a specific gene was desired. Thus, DNA markers can help to identify the desired gene carried over fairly quickly and accurately.

Quantitative Trait Loci (QTL) are very prominent in plant genomes but they cannot be selected and analysed individually due to their small effect and modification by the environment. DNA markers can be used in their genetic analysis, help to identify their chromosomal location and subsequently aid their effective manipulation in breeding programmes.

Gene pyramiding is another breeding objective that is difficult to achieve relying only on the plant's phenotype. However, if DNA markers can be coupled to the desired genes and/or characteristics in the genome it becomes easier to manipulate and combine the different desired traits in one genotype.

DNA markers can also be used very effectively in DNA fingerprinting. DNA fingerprints may serve to enforce plant breeder's rights. Also, fingerprinting of varieties/germplasm can help to characterise genetic diversity, so that germplasm resources can be managed and utilised more effectively. Using DNA-based markers to identify genetic relationships between genotypes will help to select parents for crosses, help to select a base population and will help to maximise heterosis.

The last application of DNA markers in plant breeding is for map-based cloning of genes. DNA markers will help to target a section of the genome containing the desired gene. The section of DNA can then be used for positional cloning.

2.2.6. Measuring genetic variability/diversity

Mueller & Wolfenbarger (1999) recommended that the best system to investigate genetic diversity should satisfy the following criteria:

- it should be cost effective and not take too long to test and analyse;
- multiple, independent markers should be produced;
- the genetic differences should be clear;
- it should be reliable and repeatable;
- it should use small amounts of DNA, even slightly degraded samples;
- it should need relatively little expertise in the molecular field; and
- no prior knowledge of the organism's genome should be needed.

Terzi *et al*. (2005) pointed out that variety identification benefits and/or protects farmers, importers, exporters and consumers. This methodology can also be used to protect the rights of the plant breeder. Therefore a high throughput technique that requires very little cost and can be automated would be essential. AFLPs and microsatellites were recommended as the systems that meet all of these requirements. Microsatellites have a more specific type of polymorphism whereas AFLP is genome wide and less specific. Both of these techniques were investigated in their report.

There are four possible reasons for measuring the amount of variability or diversity in a population. These include:

1. Studying the dynamics of a population, for example in conservation or a germplasm collection, etc.
2. Assessing the amount of variability or diversity available for breeding, for example concerning yield or disease resistance, etc.
3. Being able to measure the variability could also help to protect the rights of released varieties and the breeders that developed them, and
4. Being able to monitor the quality of commercially traded products is also important.

Terzi *et al*. (2005) developed a technique that could distinguish between different cereal species in a mixture of flour. This was done to limit the amount of bread wheat flour

contamination found in pasta flour, which should only be durum wheat flour. The big difference between durum and bread wheat is the D-genome. This was exploited in their study by designing PCR primers that target the D genome. The primers were very effective and could identify *T. aestivum* DNA in a predominantly durum wheat sample.

2.2.7. Reliability and reproducibility of DNA markers

The results obtained from DNA markers have to be reliable and repeatable. Therefore it is important that the system developed to measure variability/diversity has these very important traits. The reliability of band scoring (results obtained from a DNA marker) can be increased by using only the comparatively strong bands. The weaker bands will not ensure repeatability. The largest and smallest bands should also be excluded because these are also not very reliable to be repeated (Weising *et al*. 2005).

Microsatellites are considered one of the most reproducible markers. This is due to their specificity and the small number of bands produced in a single PCR reaction. Microsatellite results have two potential problems. The results are influenced by stutter bands or a non-amplifying allele occurs which is called a null allele. This can affect the accuracy of the results, but the specificity and other advantages of this technique compensate for these problems (Weising *et al*. 2005).

2.3. DNA FINGERPRINTING IN PLANTS

A commonly used tool for plant variety differentiation is DNA fingerprinting. This is a reliable tool that uses highly polymorphic and abundant DNA markers. It removes the restrictions that are associated with biochemical and morphological characteristics of, especially, very closely related plant varieties (Dograr *et al*. 2000).

2.3.1. Early development of this technique

The first DNA fingerprinting methods used in plants were based on southern blotting and RFLP analysis. Ryskov *et al*. (1988) first reported DNA fingerprinting in plants by demonstrating differences in the DNA fragment pattern of two barley varieties. They digested the DNA with *Hae*III and then used the M13 probe in Southern blot hybridisation.

Dallas (1988) used the human 33.6 minisatellite probe to distinguish between different rice cultivars.

PCR amplification of microsatellites was first demonstrated in plants (soybean) by Akkaya *et al*. (1992). It was later discovered that microsatellites could be used for discriminating plant genotypes in population studies, gene tagging and linkage mapping (Weising *et al*. 2005).

2.3.2. Use of DNA fingerprinting to identify genotypes

Genotype identification using DNA fingerprinting is important in humans as well. It is especially used in forensics to help identify individuals at a crime scene. This is also the case in forensic botany. Plant material can be used to solve a criminal case. Plant DNA fingerprints have been used as evidence to connect a person with the crime scene using the plant sample that was found on that person. In these cases, where two samples are found to be identical, it is necessary to investigate the overall variation present in the group where the sample originated. RAPDs and AFLPs are the most suited techniques for forensic botany. They are quite easy to perform and require no prior knowledge of the DNA sequence of the organism (Korpelainen & Virtanen 2003).

DNA fingerprinting was also used in a court case to discriminate between strawberry varieties. A patented variety was being produced illegally and this could be identified using DNA fingerprinting (Congiu *et al*. 2000)

Unfortunately these techniques (RAPD and AFLP) cannot be used if the template DNA is degraded. In that case microsatellites can be used in forensic botany due to the small amount of DNA it is capable of recognising and amplifying. They also have the advantage of being species specific (not influenced by contamination) and are very reproducible. Microsatellite results are also easy to manage and compare in databases (Weising *et al*. 2005).

The two most informative DNA markers to use for DNA fingerprinting are therefore AFLP and microsatellites. These are very often used in studies of genetic diversity within and among populations. The two marker systems differ considerably. For example, AFLP is based on the selective amplification of fragments produced by genomic DNA digestion with restriction enzymes. It is the technique to use if no prior knowledge of the genome is

available. There are a large number of AFLP markers available and they can be used fairly quickly. AFLP markers are codominant or biallelic (either present or absent). On the contrary microsatellites are short tandem repeats about 1-6 bp long and can show large differences in length between individuals. They are mostly codominant and multiallelic. Thus, there are two possible routes for a genetic diversity study: one path has a large number of less informative markers (AFLP) and the other has a small number of more informative markers (microsatellites) (Gaudeul *et al*. 2004).

2.3.3. Examples of DNA fingerprinting in plants

The most abundant and polymorphic marker in plant species was shown to be microsatellites. They have been useful for genotype identification in rice and soybean (Plaschke *et al*. 1995).

Microsatellites and AFLPs have contrasting characteristics. The AFLP technique produces numerous markers spread randomly throughout the genome whereas microsatellites are very specific. In a diversity study the level of polymorphism revealed by microsatellites would be very low. Many studies use microsatellites together with AFLPs to measure the level of diversity in a population. This is done to reveal the balance between the evolutionary forces behind each of the markers. Thus, a comparative analysis study is performed (Mariette *et al*. 2001).

Mariette *et al*. (2001) tested twenty three populations of *Pinus pinaster* with three microsatellite loci and 122 AFLP loci. This study was aimed at assessing the genetic diversity both within the population and between populations using both techniques and then comparing the results. They also wanted to know if the markers gave the same information. The microsatellites showed a higher level of diversity within the population than the AFLPs.

Bohn *et al*. (1999) assessed the genetic diversity of both German and Austrian winter wheat (*T. aestivum*) varieties. They used three different techniques; namely 117 RFLP probes, 16 AFLP primer combinations and 21 SSR primer pairs. This team recommended that based on the results of their study, the technique to use for DNA fingerprinting in wheat is AFLP.

Jakše *et al*. (2001) described the characterisation and use of microsatellites for determining genetic identities and assessing the genetic variability of hop varieties. They found 32

polymorphic alleles in 55 diploid hop genotypes. Only 81% of the genotypes could be distinguished from each other using microsatellites. Forty-one of these genotypes were assessed using AFLPs as well. The results from the two techniques were compared and it was concluded that microsatellites and AFLPs should be used for different aspects of cultivar evaluation.

2.4. RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLP)

The RFLP technique involves the generation of fragments produced by restriction enzymes. The length differences observed between these fragments are caused by one of two things. First, the presence and/or absence of the restriction sites of the enzymes used; and secondly one or many insertions or deletions between the restriction sites (Mueller & Wolfenbarger 1999).

DNA variations can be evaluated by cutting the entire genome with restriction enzymes. This produces a large number of DNA fragments depending on the frequency of the recognition sites of the enzymes used. The length of each fragment depends on the distance between the restriction sites. To reduce the number of fragments that must be analysed, probes are used that hybridise to specific fragments. The fragments recognised by the probe are visualised using autoradiography or DIG detection. Any differences in the restriction sites between individuals will produce a different banding pattern using the same restriction enzyme and hybridising probe. The most effective probes to use are unique sequences or low copy-number sequences. These will hybridise to the least number of fragments and thus make analysis much easier than a high copy-number probe (Chahal & Gosal 2002).

The use of RFLP markers in wheat varietal identification is restricted because of low intraspecific polymorphism. This occurs because most hexaploid wheat subspecies are very closely related to *T. aestivum* (Röder *et al*. 1995).

2.5. RANDOM AMPLIFIED POLYMORPHIC DNA (RAPD)

RAPD markers are produced by amplifying random DNA fragments using short primers that consist of a random sequence of nucleotides. This is a dominant marker type but homologous

alleles can sometimes be observed with the help of pedigrees (Mueller & Wolfenbarger 1999).

In plants, the primer is usually 10 bp long and can amplify between two and ten fragments. The fragments can easily be separated on an agarose gel and visualised with ethidium bromide. Polymorphisms are caused either by point mutations in the primer's binding site or by insertions or deletions that alter the length of the fragment or prevent it from amplifying at all (Chahal & Gosal 2002).

RAPDs can be used in DNA fingerprinting. It is a relatively simple technique that does not require any sequence information or a probe library and does not use radioactivity. It can be performed at very low costs and has a high marker output. RAPDs have already been used to determine the phylogenetic relationship between species, subspecies and cultivars. It can also be used to measure the genetic variation in populations and species and can identify cultivars, varieties, breeding lines and clones. Identifying plant breeding material accurately is important for protecting breeder's rights and can speed up plant breeding programmes. Unfortunately many problems (in particular pertaining to repeatability) have arisen in the use of RAPD markers due to the complexity of the hexaploid wheat genome. This has limited the use of this marker system for cultivar identification. However, RAPD markers can be converted to more specific and reliable SCAR markers (Myburg *et al*. 1997).

SCAR markers are designed by cloning and sequencing the original, desired RAPD fragments. Primers of about 24 bp are then designed that target the ends of the RAPD fragment. Therefore these primers will amplify a single locus or a sequence-characterised amplified region when used in a PCR with the original template DNA (Weising *et al*. 2005). The disadvantage of this conversion is the time and effort needed. On the other hand, the main advantage of a SCAR is that it uses very stringent PCR conditions. This means that amplification is very specific causing results to be reliable and reproducible. SCAR markers are locus specific and codominance can be identified. Thus, differentiating between heterozygotes and homozygotes may be possible. Codominant SCAR markers are much more informative than dominant RAPD markers especially in genetic mapping (Chahal & Gosal 2002; Weising *et al*. 2005).

2.6. DIVERSITY ARRAY TECHNOLOGY (DArT)

Diversity array technology is a basic genotyping technique, which is both cost effective and reproducible. This technique was invented by Dr Andrzej Kilian as a means to surmount some of the restrictions of the other molecular markers including RFLP, AFLP and SSR (www.diversityarrays.com/index.html).

The DArT technology has six steps, which include: reducing the complexity of the DNA of interest using restriction enzymes and ligating adaptors; creating a library of fragments cloned into *Escherichia coli*; microarraying libraries by amplifying clones and spotting them onto glass slides; hybridisation of fluorescently labelled DNA onto slides (using different colours helps to produce different patterns); scanning the slides for any hybridisation signal, which comprises removing excess dye and capturing the final pattern; and finally extracting and analysing the data using specific DArT software (www.diversityarrays.com/index.html).

This technique has various advantages and applications. It is reliable, requires no prior knowledge of the sequence investigated, it has a high throughput due to its ability for multiplexing, and it can be performed and reproduced at a relatively low cost. These traits aid in its use for genome profiling and in background screening. It can also be used to construct genetic linkage maps rather quickly, to identify QTLs, to accelerate the introduction of genomic regions during backcrossing programmes, to facilitate marker-assisted selection by targeting numerous traits at once, to assess the genetic diversity of a population, and it can examine the make-up of a complex DNA sample (www.diversityarrays.com/index.html).

The species that can currently be genotyped by the DArT technology include: wheat, barley, rice, apple, cassava, tomato, sorghum, pigeon pea, ryegrass and Arabidopsis. These species have known and efficient DNA complexity-reduction methods. Microarrays are currently being developed for the following species: chickpea, sugarcane, lupins, quinoa, banana and coconut (www.diversityarrays.com/index.html).

Akbari *et al*. (2006) showed that diversity array technology could successfully be applied to the 16 billion bp genome of hexaploid bread wheat. They found that the data obtained from the DArT markers was comparable to that obtained from RFLP, AFLP and SSR markers.

2.7. MICROSATELLITES

Microsatellites are tandem repeats of very short sequences. The first record of a microsatellite was early in the 1970s in a hermit crab. It was a $(TAGG)_n$ repeat found in the crab's satellite DNA (Skinner *et al*. 1974). Many microsatellites have since been reported in bacteria, fungi, plants, animals and humans (Weising *et al*. 2005). Microsatellites were first discovered in plants using RFLP fingerprinting and oligonucleotide probes (Beyermann *et al*. 1992). They were first cloned in plants in 1991 and the first PCR-generated, locus-specific microsatellite marker in plants was reported in 1992 (Akkaya *et al*. 1992; Condit & Hubbell 1991).

Their high level of polymorphism is the best characteristic of microsatellites. Each marker has the potential to have a large number of alleles because one less/extra repeat sequence is a polymorphism. It is this trait that makes them a commonly used marker to derive genetic fingerprints or to study the genetic relationship between genotypes. Microsatellites show codominant inheritance which is important to discriminate between closely related lines. At present microsatellite markers are being widely used to identify genotypes, map quantitative trait loci (QTL) and assess genetic diversity (Mahmood *et al*. 2004).

PCR primers (18 – 25 bp) specific to sequences that flank the tandem repeats of 2 – 4 bp are used to amplify the microsatellite loci. Variation in the number of repeats causes fragments of different sizes to be amplified. These size differences give microsatellites (short sequence repeats) their polymorphic nature. This method is useful for genotype identification in self-pollinating species that have a low genetic variability. The highest number of microsatellite loci in the existing microsatellite map of wheat occurs in the B genome and the lowest amount in the D genome (Manifesto *et al*. 2001; Pestsova *et al*. 2000).

Microsatellites have great potential to be used as genetic markers in hexaploid wheat. $(GA)_n/(GT)_n$ sequences are found approximately every 270 kb of DNA; $(AC)_n$ sequences are found roughly every 292 kb and the $(AG)_n$ microsatellite sequence is found almost every 212 kb. Dinucleotide repeats are the most frequent and can be up to 40 repeats long. The trinucleotide repeat sequences are about one tenth as frequent and the tetranucleotide repeats are the rarest sequence. They are all distributed quite randomly throughout the genome but most are found on the B genome. These markers can be used for testing genetic diversity

among varieties of bread wheat and should be the future marker of choice in practical wheat breeding because of its locus-specificity and high level of polymorphism (Gupta *et al*. 1999).

SSRs are rapidly replacing RFLP mapping technology in humans due to their ease of use and high information content. For the same reasons these markers are now being used increasingly to measure the genetic diversity of plant breeding material and cultivars. SSRs are also being used to expand the genetic maps of the major cereal crops. Microsatellites furthermore find application in marker-assisted screening. Therefore methods have to be developed that allow for high throughput while maintaining the high level of information provided by microsatellites. Such a system has to be reliable and should be available at a reasonable cost to the consumer/researcher (Donini *et al*. 1998).

2.7.1. Advantages and disadvantages

Microsatellites are easy to use but one disadvantage is that most are found in the non-coding regions of the genome. The process to develop a microsatellite marker is also very complex and involves isolation and sequencing of clones that contain the simple sequence repeat motif followed by the design and testing of the flanking primers. Thus, it is time consuming and very expensive and only about 30% of the primer pairs developed are functional and useful for genetic analysis (Gao *et al*. 2004).

As DNA markers, microsatellites have numerous desirable features such as high levels of polymorphism and information content, unambiguous designation of alleles, even dispersal, selective neutrality, high reproducibility, codominance and rapid and simple genotyping assays. They have a wide range of applications; such as genetic mapping and genome analysis, genotype identification and variety protection, seed purity evaluation and germplasm conservation, paternity determination and pedigree analysis, gene and QTL analysis, marker-assisted breeding, measurement of genetic diversity, assignment of lines to heterotic groups and genetic fingerprinting. Microsatellites provide a power of discrimination equal to or better than that of RFLP in a more cost effective manner (Li *et al*. 2001).

Microsatellites are up to ten times more variable than other marker systems like RFLP. This is important for the genetic mapping of species that have little intraspecific polymorphism; like wheat (Röder *et al*. 1995). In wheat, microsatellites are genome-specific and seem to be

evenly distributed throughout the genome. They are regularly used for mapping agronomically important genes, studying the genetic diversity of bread wheat and other related species; verify the identity of cytogenetic stocks and to investigate the genetic diversity and evolutionary history of specific genes in bread wheat (Korzun *et al*. 1997, 1999). Other uses of wheat microsatellites include: tagging resistance genes, identifying QTLs, marker-assisted selection, verifying integrity and genetic stability of gene bank accessions and to outline the origin and relationship of varieties (Huang *et al*. 2002; Schmidt *et al*. 2004).

The robustness, high level of information and the fact that microsatellites require very small amounts of DNA template ensure that these markers can be used for genetic mapping. Using them in marker assisted selection and diversity studies depends greatly on their ability to allow large sample throughput while maintaining the advantages of being robust and highly informative. Microsatellites are useful as an added system to AFLP analysis. It has greater genome coverage and is more sensitive to sample contamination than AFLP analysis. Microsatellites also have the highest level of polymorphism and are the most informative marker system in hexaploid wheat (Donini *et al*. 1998; Kobiljski *et al*. 2002; Perry 2004).

Microsatellites are multiallelic which makes them great for evolutionary studies and genetic diversity and relationship studies (Ahmad 2002). Simple sequence repeats (SSR) or simple tandem repeats (STR) are short repeats of 2 – 6 nucleotides. More than 1000 wheat genomic SSR markers have been mapped and are available. They are used in wheat for genome and physical mapping, gene tagging and genetic diversity estimates (Bandopadhyay *et al*. 2004; Kuleung *et al*. 2004).

2.7.1.1. Fluorescent-based semi-automated analysis

Microsatellite analysis can be automated. This is achieved using fluorescently labelled primers and a DNA sequencer. The PCR reaction includes one fluorescently labelled primer and one unlabelled primer. The reaction conditions are exactly the same as for unlabelled primers. Fluorescently labelled microsatellite markers coupled with genotyping on an automated DNA sequencer have certain advantages over the silver staining method. The first advantage is the increase in throughput because the amount of information received from one lane/sample is increased through multiplexing. In other words, markers that are labelled with different colours can be analysed together because the colours can be separated after

electrophoresis. Another advantage is the improved accuracy in sizing of alleles. This is achieved with an internal size standard added to every lane analysed as well as the automated allele-calling algorithm. Thus automation increases the speed and accuracy of data collection and processing and reduces cost. This is done by decreasing the volume of the PCR reaction while maintaining the ability to detect loci that are often difficult to amplify (Coburn *et al*. 2002).

The semi-automated microsatellite genotyping method is replacing the manual techniques in plant breeding and genetics research. Microsatellite markers can now be used in high throughput mapping, pedigree analysis, fingerprinting of accessions and assaying genetic diversity and it can improve the efficiency of managing a germplasm collection, deliver purity-proven seed stocks to growers and provide the basis of intellectual property protection (Coburn *et al*. 2002).

Blair *et al*. (2002) used four multiplexed reactions to test 72 rice cultivars with twenty seven simple sequence repeats. The microsatellite markers were fluorescently labelled which made semi-automated detection of the fragments and sizing of the alleles possible. This system was useful for fingerprinting varieties and could be used to facilitate mapping of a segregating population, and to identify genes and the QTLs underlying traits of interest. This technique was found to be easy to perform, optimise and analyse because rice has a diploid genome.

Coburn *et al*. (2002) also used high throughput microsatellites to analyse rice. They were able to combine one hundred and fifty nine microsatellite primer pairs into twenty-one multiplexed semi-automated genotyping reactions.

There are certain limitations to detecting allelic diversity in rice using microsatellites in association with conventional visualisation techniques. Some of these limitations pertain to: the specific loci being assayed, the number and diversity of genotypes that are sampled and the sensitivity of the technique that is used to detect the small molecular weight differences. Automation of microsatellite detection on a DNA sequencer using fluorescently labelled primers is more cost effective and an efficient genotyping method. Multiplexing increases throughput. It allows a lot of DNA fragments and multiple loci to be tested in a single lane. This is achieved by labelling markers of the same size with different coloured dyes, but those with different sizes can be labelled with the same coloured dye. Microsatellites can either be

pooled together after amplification or amplified in the same PCR reaction. Fluorescently labelled markers, with dyes of different colours, and automating the detection on a DNA sequencer has been shown to be an inexpensive and effective method for genotyping (Blair *et al*. 2002).

There are limitations to the use of multiplex PCR reactions for microsatellite markers. The first of these is that either no multiplex reactions have been developed or only a small number of markers have been developed for crop plants. Secondly, the number of microsatellite loci that are required for genetic mapping or other plant breeding applications is too many to form a single multiplex PCR reaction. The third limitation is that some microsatellite primer pairs do not function correctly in multiplex reactions. Lastly, the microsatellites that are used in a particular mapping problem may have to change quite often because of practicality or necessity (Tang *et al*. 2003).

2.7.2. Different types of microsatellites

Two recently developed variations in microsatellite amplification methodology exist. The first is Expressed Sequence Tag derived Simple Sequence Repeats (EST-SSR). These microsatellites are found in the transcribed region of the wheat genome (Eujayl *et al*. 2002). They are less polymorphic than genomic simple sequence repeats, but can still be used as informative markers in the assessment of genetic relationships (Leigh *et al*. 2003b)

The other method is called Inter-Simple Sequence Repeat PCR (ISSR-PCR) or microsatellite primed PCR or Simple Sequence Repeat (SSR-) Anchored PCR. This system uses anchored simple sequence repeat primers and has the potential to characterise complex genomes. It has a simple concept, which is that the 5' or 3' end of the primer is anchored with two or four purine or pyrimidine residues. This causes PCR amplification of genomic fragments that are surrounded by closely spaced inverse microsatellites. The reaction either contains an ATP labelled ISSR primer or one of the dNTPs is radioactively labelled. The system can be adapted to use fluorescently labelled primers and is then called Fluorescent Inter-Simple Sequence Repeat PCR (FISSR-PCR) (Nagaraju *et al*. 2002).

2.7.3. Different uses

There are many examples of microsatellites being used to test genetic diversity. This has been done in humans, animals and plants.

Gyapay *et al*. (1996) described two genotyping techniques in humans using microsatellites. The first technique was aimed at researchers who do not have access to an automated DNA sequencer. A sequencing gel was used to separate the PCR products which were then transferred to a nylon membrane. The products were then detected by hybridization with nonradioactive probes. The other technique used fluorescently labelled primers. The PCR products were detected on an automated DNA sequencer. These techniques are very similar to those described in this study except that the denaturing polyacrylamide gels in this study were silver stained.

Li *et al*. (2001) evaluated the genetic diversity of 90 cowpea breeding lines using twenty-seven microsatellite primer pairs. Five polymorphic microsatellites could distinguish 88 of the 90 cowpea breeding lines.

Ahmad (2002) tested the genetic diversity of thirteen wheat genotypes grown in New Zealand using 43 simple sequence repeats (SSRs). They were able to analyse various wheat cultivars and showed that SSRs could be used on its own as a mechanism to investigate genetic diversity between cultivars. Ahmad analysed a smaller number of elite wheat genotypes with a higher number of SSRs. This meant that the genome coverage was much better. A collection of highly polymorphic SSRs could be chosen to test genetic diversity, to identify cultivars, and plant variety protection could be established for wheat.

Dograr *et al*. (2000) was also able to distinguish between durum wheat lines. They used seven microsatellites to discriminate between four well-adapted landrace selections, five cultivars and seven recently developed advanced lines.

Prasad *et al*. (2000) used twenty wheat microsatellites on 55 elite wheat genotypes. They wanted to examine the effectiveness of microsatellites to detect DNA polymorphisms, identify genotypes and to estimate the genetic diversity among the wheat genotypes. The 55 wheat genotypes originated in twenty nine countries from six continents. A subset of four of

the twenty microsatellites was able to distinguish between 41 genotypes. The ability of such a small number of microsatellites to distinguish between so many genotypes illustrates the great potential of these markers to be used in genotype discrimination. These four microsatellites were therefore also used in the present study.

Huang *et al*. (2002) used a set of twenty four microsatellites to assess the genetic diversity of 998 bread wheat (*T. aestivum* L.) varieties. The 998 varieties were sampled from 68 countries on five continents. The genetic diversity was shown to be different between the geographic regions. The most diverse varieties were those taken from the East and Middle East. It was also seen that the varieties from Southeast Europe were more diverse than those from North and Southwest Europe.

Another study that focused on bread wheat from Europe was reported by Plaschke *et al*. (1995). They used twenty three microsatellites to estimate the genetic diversity of 40 wheat cultivars and lines. These lines were mostly elite material from Europe. These microsatellites were enough to distinguish all of the cultivars, except two. Therefore they showed that relatively few microsatellites are needed to estimate the genetic diversity of wheat and to identify elite hexaploid bread wheat cultivars.

Roussel *et al*. (2005) also investigated the variation in bread wheat in Europe. They took 480 bread wheat varieties from fifteen European geographical areas that were released between 1840 and 2000. These were tested with 39 microsatellite markers. The analysis of qualitative variation showed that the more recent the release of the European variety the more similar they are to each other. The western European countries including France, The Netherlands, Great Britain and Belgium showed a smaller number of alleles than the south-eastern European countries of former Yugoslavia, Greece, Bulgaria, Romania and Hungary. The Mediterranean area which includes Italy, Spain and Portugal had a higher number of alleles than the other two areas. This suggests that the diversity in European wheat types is not randomly distributed. This team suggested that this could be explained by chronological and geographical variation which can be linked to differences in breeding practices and agricultural policies of the countries involved in the study.

2.8. AMPLIFIED FRAGMENT LENGTH POLYMORPHISM (AFLP)

Amplified fragment length polymorphism is also a highly polymorphic marker system. It can be used in all organisms without prior knowledge of its genome and produces very informative DNA fingerprints (Weising *et al*. 2005).

It was first described by Vos *et al*. (1995). The technique is based on selectively amplifying fragments formed by specific restriction enzyme digestion. The main steps in the production of these fragments are the following. First the whole genome is cut with two restriction enzymes. One is a rare cutter (usually *Eco*RI) and the other is a frequent cutter (usually *Mse*I). This refers to the relative number of recognition sites in an organism's genome. Two different adaptors bind to the ends of the digestion fragments. Each restriction enzyme has a different adaptor sequence. These sequences are recognised by the preselective primers. The final amplification step involves primers that recognise the adaptor sequence and have different extensions on the 3' end. These primers can be used to amplify various fragments by changing the one, two or three nucleotides on the 3' end of the primer. For each selective nucleotide added the number of fragments amplified is reduced to 1/16 of the restriction fragments produced in the first step of this technique. This number is then further reduced to only 1/4096 restriction fragments that are amplified with primers that have three selective nucleotides. This very small number of fragments makes visualisation and analysis much simpler. Also, only the final amplification fragments that have different enzyme recognition sites on each end are visualised in this technique (Vos *et al*. 1995; Weising *et al*. 2005).

Fragment visualisation was previously achieved using radioactively labelling of one of the selective primers, but new technology has made this technique safer by eliminating the radioactive labels. These have been replaced by fluorescently labelled primers that are visualised on a DNA sequencer. The appeal of this technique is that it has a high multiplexing ratio when using fluorescently labelled primers. This allows a lot of information to be generated from a rather small number of reactions (Schwarz *et al*. 2000; Waugh *et al*. 1997).

AFLP is a very effective technique in identifying polymorphisms. The large number of fragments generated and its high reproducibility makes this a good technique for detecting polymorphisms and for determining linkages by analysing the individuals of a population.

Compared to RFLP and RAPD, AFLP is the most efficient marker for polymorphism detection. AFLP can therefore be used to evaluate genetic diversity and genotype relationships in wheat (Gupta *et al*. 1999; Manifesto *et al*. 2001)

AFLP can detect numerous polymorphisms between samples. The reason for the polymorphism cannot be detected, but it can be any of the following: the creation or loss of a restriction enzyme recognition site, mutation within the priming site on the 3' end of the AFLP primers or insertion/deletion within a locus in the genome. The mutations can be divided into two broad categories: (a) mutations that are inherited by all the individuals in a cultivar, but are different between the cultivars and can thus be used to identify the cultivar; and (b) mutations that occur in different plants of the same cultivar (Soleimani *et al*. 2002; Weising *et al*. 2005).

2.8.1. Advantages and disadvantages

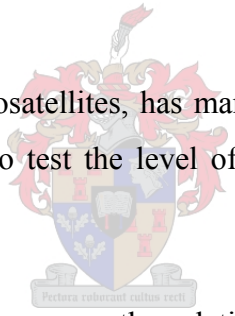
Only a limited amount of the amplified fragments are visualised in this technique. This is controlled by the one, two or three base pair extensions of the selective primers. The main advantage of AFLP is that multiple bands are produced from all over the genome. The main disadvantage with AFLP is that alleles can not be recognised very easily (Abdel-Satar *et al*. 2003).

The new fluorescently labelled system provides some added advantages to the AFLP system. The first is that the final fragments generated in this system are sized automatically. This enables the fragments of different reactions to be compared very accurately. The number of samples that can be compared together is also relatively unlimited compared to the older radioactive system. Another limitation of isotope visualisation is the very low throughput achieved by analysing only one sample at a time. This is overcome by the multiplexing ability of the new system. Other advantages to the fluorescent technique include: a large number of anonymous and randomly scattered markers are generated in a single lane; the stringent PCR conditions produce a high level of polymorphism that is reproducible and reliable. An internal size standard with a different colour is added to every lane/sample, which provides correct size-calling within and between gel runs (Schwarz *et al*. 2000).

There are numerous advantages to the AFLP technique in general. It is as versatile as any other PCR reaction. No prior knowledge of the genome is needed for this technique. The conditions of the required PCR reaction can be set in such a way that the results are robust and reproducible. Any bands of interest that are produced in this method can be removed from the gel, cloned, sequenced and converted to a SCAR marker, which is more specific. Different AFLP primer combinations can be put together to provide a large amount of information for the samples tested. Multiplexing with these primers is also possible as well as using selective primers of different lengths i.e. the number of selective nucleotides added to the 3' end differ. The enzymes that are used can also be changed (Weising *et al*. 2005).

There are also, however, disadvantages to this technique. These include: dominance of the markers, clustering of markers, limited polymorphisms in certain cultivated species, the need for both good quality and medium quantities of template DNA (Weising *et al*. 2005).

2.8.2. Different uses



The AFLP technique, like with microsatellites, has many applications. It is often used as a DNA fingerprinting method and/or to test the level of variation or genotype diversity in a population.

Incirli & Akkaya (2001) were able to assess the relationship of nine winter and six spring durum wheat cultivars grown in Turkey. Eighteen AFLP primer combinations produced 189 polymorphic loci. Only *Eco*RI and *Mse*I selective primers were used with three nucleotide extensions on the 3' end. The relationship between these cultivars was known beforehand and the expected results were obtained using the AFLP technique.

Powell *et al*. (1997) studied quantitative traits in barley using AFLPs. Both *Eco*RI and *Pst*I selective primers were used in combination with *Mse*I selective primers. They reported that the AFLP marker had wider genome coverage than the RAPDs and RFLPs that were also used. Their study produced evidence that the *Eco*RI and *Pst*I enzymes cover different parts of the genome.

2.9. RETROTRANSPOSONS

Retrotransposons are mobile genetic elements that are found everywhere in the plant kingdom. They are found on all chromosomes, have variable copy numbers and are randomly spread throughout the genome. They move via RNA intermediates and then insert cDNA copies anywhere in the genome. This increases the size of the plant's genome and its total amount of DNA. Large differences in retrotransposon copy numbers can be seen at very short evolutionary time scales (Leigh *et al*. 2003a).

Retrotransposons form part of the Class I transposons that spread via a RNA intermediate and is then reverse transcribed into cDNA. The retrotransposons can be divided into four types: retroviruses, Long Terminal Repeat (LTR) retrotransposons, Long Interspersed Elements (LINEs) and Short Interspersed Elements (SINEs). LINEs and SINEs are also referred to as non-LTR retrotransposons (Weising *et al*. 2005).

LTR retrotransposons are identified by the 300 to 500 bp-long direct repeat that is found at both ends of this element. These sequences control the initiation and termination of transcription and polyadenylation. The DNA sequence found between the LTRs is usually about 3 to 5 kb in length but it can be more than 10 kb. This sequence encodes for all the enzymes and proteins that the retrotransposon will need to replicate and spread to another site in its host's genome (Leigh *et al*. 2003a; Weising *et al*. 2005).

Retrotransposons have a high copy number in plant genomes and show a certain amount of sequence heterogeneity and insertional polymorphism not only between but also within species. Retrotransposons are the most common form of eukaryotic transposable elements. They replicate via a RNA intermediate and then convert it to DNA via reverse transcription before reinsertion. Retrotransposons are generally found in the regions flanking known plant genes (Waugh *et al*. 1997).

The LTRs of each retrotransposon family is unique and this has been used in genetic analysis by designing primers in the LTRs that amplify their flanking regions. Each insertion of a new daughter copy acts like an evolutionary stop watch and generates a marker of that exact time and at that exact place. LTRs can also rearrange themselves. Rearrangement between different LTRs causes the element to be removed, leaving just the LTRs. Or a rearrangement

can occur to disrupt the internal sequence. This maintains the relationship between the LTR and the integration point and the flanking DNA. A nested integration can provide information on the historical activity of the different families (Leigh *et al*. 2003a).

It is the error-prone nature of their replication by reverse transcriptase, the mutagenic potential of transpositional integration and the effects of accumulation and recombination that make active retrotransposons big contributors of genetic diversity in plants. Genomic changes caused by retrotransposons can be monitored by the joints that form between the flanking DNA and the LTR that was created during integration. PCR-based marker systems have been developed that amplify sequences between retrotransposons and flanking DNA (Kalendar *et al*. 2000; Manninen *et al*. 2000).

2.9.1. Different types of retrotransposon-based DNA marker techniques

Inter-Retrotransposon Amplified Polymorphisms (IRAPs), Retrotransposon-Microsatellite Amplified Polymorphisms (REMAPs) and Sequence-Specific Amplified Polymorphisms (SSAPs) are all based on retrotransposon activity and are being used extensively as markers. The fact that each transposition event is different and each transposon family has a different history suggests that multiple retrotransposon families can be used for genetic analysis in applications such as mapping, fingerprinting, marker-assisted selection and evolutionary studies (Leigh *et al*. 2003a). Each of the techniques is represented in Figure 2-1 at the end of this section.

2.9.1.1. Inter-Retrotransposon Amplified Polymorphism (IRAP)

IRAP exploits the detection of a chance association between the LTR and other retrotransposons (Leigh *et al*. 2003a). IRAP markers are generated between two LTRs using outward facing primers that anneal to the LTR target sequence. These markers have been used to distinguish barley varieties and create fingerprints of the species. The dispersion, ubiquity and prevalence of retrotransposons in plant genomes give them an excellent base for a marker system. Retrotransposons replicate via successive transcription, reverse transcription and insertion of new cDNA copies into the genome just like retroviruses. The new insertions into the genome will create polymorphisms (Kalendar *et al*. 1999).

2.9.1.2. Retrotransposon-Microsatellite Amplified Polymorphism (REMAP)

REMAP markers are generated between a LTR and a simple sequence repeat (Kalendar *et al*. 1999). Similar to IRAP, REMAP exploits a chance association (Leigh *et al*. 2003a). One of the primers bind to the LTR of the retrotransposon and the other binds to a repetitive section found inside a microsatellite. Therefore the polymorphisms found in this technique could be caused by the retrotransposon or by the number of repeats in the microsatellite (Kalendar *et al*. 1999).

2.9.1.3. Sequence-Specific Amplified Polymorphism (SSAP)

The two LTR-containing groups (Ty1-*copia* and *gypsy*) and the non-LTR retrotransposons (LINE elements) are all found in plant genomes. The most popular transposon-based marker method is SSAP and is also known as transposon display (Queen *et al*. 2004).

The Ty1-*copia* group of retrotransposons are found randomly spread throughout the barley genome. They have a very high copy number and are widely dispersed over the chromosomes. These characteristics endow the retrotransposons with significant potential to be developed into a multiplexed DNA-based marker system. This is what was described by Waugh *et al*. in 1997. They used the LTR sequence of the *BARE*-1-like retrotransposons in barley to detect DNA polymorphisms. This was based on the relative distances between retrotransposons and restriction enzyme cutting sites. They demonstrated that this retrotransposon-based technique was more polymorphic than AFLP. The main difference between the two techniques is the selective primers that are used in the final amplification reaction. In AFLP the selective primers are based on both of the adaptor sequences (e.g. *Eco*RI and *Mse*I). The SSAP technique uses one selective (*Eco*RI or *Mse*I) primer and one primer that recognises the retrotransposon sequence (Leigh *et al*. 2003a).

Waugh *et al*. (1997) used one primer that recognises the *BARE*-1 retrotransposon and the other primer was either a selective *Mse*I primer or a selective *Pst*I primer. These were used to amplify fragments in the barley genome. The selective primers used had one, two or three additional selective nucleotides. The *BARE*-1 primer was radioactively labelled to allow the fragments to be visualised.

A primer facing outward from the *BARE*-l LTR was used together with an AFLP adaptor primer in the SSAP technique in barley (Kalendar *et al*. 1999). A variation to this SSAP technique used in barley was developed by Queen *et al*. in 2004. They found that the nineteen nucleotides used as the target site in the *BARE*-1 element found in barley correspond exactly to the target site selected in the *Wis-2* element found in wheat. Therefore retrotransposon markers were developed that are based on the *BARE*-1 element but that recognises and binds to the *Wis-2* retrotransposon found in wheat. They showed that four different LTR retrotransposons could be used in SSAP as molecular markers for mapping and diversity studies in a broad variety of wheat species.

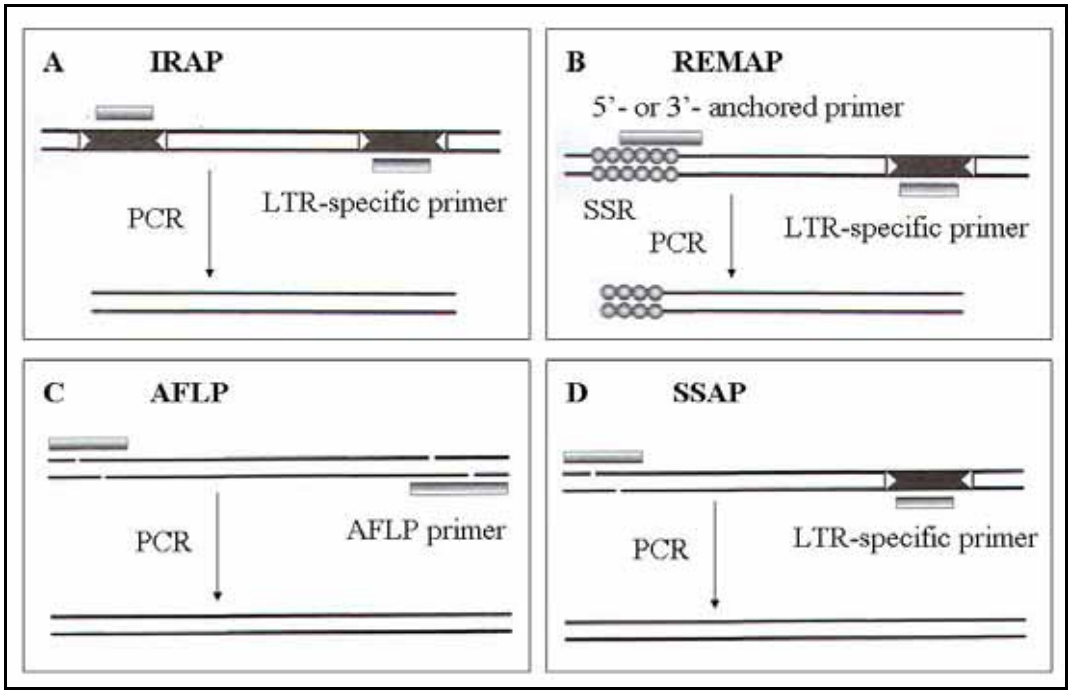


Figure 2-1 Schematic representation of the various retrotransposon-based techniques. (**A**) Inter-Retrotransposon Amplified Polymorphism: two LTR-specific primers amplify fragments between the retrotransposons. (**B**) Retrotransposon-Microsatellite Amplified Polymorphism: amplifies fragments between a microsatellite and a retrotransposon. (**C**) Amplified Fragment Length Polymorphism. This was inserted to compare it to the SSAP technique. (**D**) Sequence-Specific Amplified Polymorphism: one AFLP selective primer and one LTR-specific primer.

2.9.2. Different uses

Gribbon *et al*. (1999) were the first to describe an SSAP method used on wheat. This technique focused on the Ty1-*copia* group of retrotransposons. This group is found in most crops including barley. They used *Pst*I selective primers and a *BARE*-1 retrotransposon primer to show that the level of polymorphism of the *BARE*-1 retrotransposon in *T. aestivum* was not very high. The conclusion or reason given for this was that it was due to the very low level of overall polymorphism seen in *T. aestivum*.

Porceddu *et al*. (2002) used SSAP in *Medicago sativa* L to amplify the LTR of the *Tms1* retrotransposon. They used the *Eco*RI and *Mse*I restriction enzymes in the initial step of this technique. Forty nine polymorphic markers were produced that could be reliably scored and used for mapping.

2.10. STATISTICAL PROCEDURES FOR THE ANALYSIS OF MULTIPLE MARKER INFORMATION

Discriminant and cluster analysis are valuable tools in the interpretation of marker data. A brief description of each is given below.

2.10.1. Discriminant Analysis

Discriminant analysis focuses on the development of a set of rules that can be used to classify an observation into one or other class/group that was previously decided upon. The classification of samples is done using a set of data that was provided for each group in the initial development of the discriminant rules. This analysis test is used when a researcher has two sets of samples. The first set of samples has certain observations and the group classification of each sample is known. The other set of samples also has certain observations but the grouping of these samples is unknown. Therefore the first sample set is used to develop limitations/rules to each group to discriminate it from the other groups. These limitations/rules are then applied to the second sample set so as to classify each sample as a member of one of the known groups. Each limitation/rule is derived so that as few as possible misclassifications are achieved when these rules are reapplied to the original set of samples (Der & Everitt 2002).

Beharav & Nevo (2003) describes the discriminant function analysis as a test to distinguish between two or more groups. These groups are dependent on the specific research condition. A researcher accumulates various observations or *discriminating variables* that evaluate certain features that are expected to differ between the groups. The mathematics behind discriminant analysis weighs each discriminating variable and combines them in a linear manner so that the groups are forced to be as statistically dissimilar as possible.

2.10.2. Cluster Analysis

According to Der & Everitt (2002) cluster analysis refers to a method of grouping of observations into clusters. This starts with single observations that can not be grouped and ending with a single group that contains all of the observations. The groups are determined using the data and the probability percentage that the observations can be grouped. A visual illustration of the clusters or groupings is given in a dendrogram which looks very much like a tree. Each branch of the tree represents the relationship between the observations. In other words how closely related the clusters are to each other. The ideal situation for a genotyping study is that all the samples tested appear in separate clusters to illustrate that each genotype is different to all the other genotypes.

2.11. STUDY OBJECTIVES

This study aims to develop methodology to readily and routinely distinguish between commercial wheat cultivars and elite germplasm that were developed by the University of Stellenbosch. Such methodology should also be powerful enough to distinguish among the majority of advanced breeding lines developed annually in the University of Stellenbosch recurrent mass selection program. Marker technologies shown by the literature to be suited best to wheat genotyping (AFLPs, SSRs and retransposon-based) will be evaluated in order to propose and implement a routine genotyping procedure that is economical, reliable and robust. The system developed will be evaluated making use of advanced breeding lines developed in the Stellenbosch University breeding program.

3. MATERIALS AND METHODS

This study was conducted in three stages:

Firstly, an initial screen was done to identify a subset of primer pairs that has the potential to be the most powerful within each of the three marker systems (SSRs, AFLPs and retrotransposon-based). Based on previously reported results, collections of primer sets were obtained that were shown to be the most likely to produce useful polymorphisms. These primers were used to amplify marker fragments in five genotypes that represent five different breeding germplasm pools. In order to minimise costs and to obtain detailed results, the microsatellite amplification products were first visualised on denaturing polyacrylamide gels. The AFLP amplification products on the other hand were visualised directly on the automated sequencer to avoid the use of radioactive isotopes. The amplification products obtained in the retrotransposon-based systems were visualised either on agarose gels (as described in the literature) or on the automated sequencer. These results were used to identify primer pairs in each marker system that gave the best discrimination between the genotypes.

Secondly, the selected primer pairs in each marker system were then tested on a panel of twenty genotypes of released varieties/advanced lines. This was done to compare the power of each system to discriminate between genotypes, to evaluate the usefulness of semi-automated visualisation of polymorphisms and to suggest a routine fingerprinting method.

Finally, the proposed fingerprinting method was tested on 119 advanced breeding lines from the breeding program at the University of Stellenbosch. The primary objectives were:

- To determine to what extent it would be possible to distinguish between lines.
- Propose a simple system to compare and record individual profiles.
- Suggest an appropriate sample size that would result in the consistent assessment of identity.

3.1. PLANT MATERIAL

Seeds of released varieties and experimental breeding lines (Table 3-1) of five different breeding programmes were used. Initially, two seeds from each variety/line were planted in a greenhouse and the leaves were used separately for genomic DNA (gDNA) extraction.

Table 3-1 The wheat genotypes employed in the first part of this study.

PLANT NUMBER		BREEDING LINE/VARIETY	PEDIGREE[1]	BREEDER
1	**21**	00K180	*Lr19*-149/7*W84-17//Capp Desprez/2* W84-17/3/SST66/5*W84-17/4/SST57	**US[2]**
2	**22**	00K268	92M30/3/SST66/2*W84-17//Alpha/4/ SST66/2*W84-17//Alpha/SST57	
3	**23**	97K1	SST66/2*W84-17//Alpha/SST57	
4	**24**	98K120	Kariega/91K3//SST66/2*W84-17	
5	**25**	00K60	W84-10/Palmiet//93 Sel1/Kariega/4/ SST66/2*W84-17//Alpha/3/SST57	
6	**26**	PAN 3404	Commercial variety	**PANNAR**
7	**27**	PAN 3490	Commercial variety	
8	**28**	PAN 3492	Commercial variety	
9	**29**	Kariega	Commercial variety	**SGI[3]**
10	**30**	Biedou	Commercial variety	
11	**31**	SST 015	Commercial variety	**Monsanto**
12	**32**	SST 027	Commercial variety	
13	**33**	SST 57	Commercial variety	
14	**34**	SST 65	Commercial variety	
15	**35**	SST 88	Commercial variety	
16	**36**	SST 94	Commercial variety	
17	**37**	12th HRWYT 46	Croc1/*Ae squarrosa*-213//PGO/3/ Sodat/Swm3//NG8201/Ning8647	**CYMMIT**
18	**38**	16 HRWSN-136	Dagua/Tota//BB/WT/3/Caqueta/ Arauca/4/Yuriya 79/5/Irena	
19	**39**	16 HRWSN-89	Chibia/5/CNDO/R143//Ente/Mexi_2/3/ *Aegilops squarrosa* (*Tauschii*)/4/Weaver	
20	**40**	16 HRWSN-28	PANDION	

[1] The pedigree information of the PANNAR, SGI and Monsanto cultivars is confidential due to company policy.

[2] University of Stellenbosch

[3] ARC Small Grain Institute

In the final stages of this study, DNA was extracted again. Only this time four seeds of each breeding line in Table 3-9 were grown in a five squares by five squares (25 compartments) plastic seed holder. Each square was layered with two layers of filter paper and wet with distilled water. The seeds were allowed to germinate for four to five days. Germination

could be slowed by placing the seeds at 4°C. All samples germinated successfully and the first leaves from the four seeds were collectively used for gDNA extraction.

Similarly, the use of multiple seeds was tested to determine the optimal amount to represent the population. This time 22 seeds of each highlighted breeding line in Table 3-9 were grown in separate Petri dishes. The seeds were left to germinate as before on two layers of filter paper and wet with distilled water. All seeds germinated successfully and it was possible to use eight of the first leaves collectively for gDNA extraction as well as another twelve of the first leaves. The other two seeds were added to ensure that twenty first leaves would be available for gDNA extraction.

3.2. GENOMIC DNA EXTRACTION AND QUANTIFICATION

An adaptation of the Doyle and Doyle (1990) protocol was used. Twenty millilitres CTAB [1.4 M NaCl, 20 mM Na_2EDTA (pH 8), 100 mM Tris-Cl (pH 8)] and 40 µl β-Mercaptoethanol (β-ME) were preheated in a waterbath at 60°C. Approximately 0.1 g of new wheat leaves were cut into pieces and placed in 2.2 ml centrifuge tubes containing two steel balls. Eight hundred microlitres 2% (w/v) CTAB and 1.6 µl 0.2% (v.v) β-ME were pipetted into each 2.2 ml centrifuge tube. The tissue was ground in a Tissuelyser (Qiagen) for three times 90 seconds at 30 Hz. The tubes were rotated after every 90 second step. The mixture was incubated in a waterbath at 60°C for 60 minutes. Eight hundred microlitres chloroform: isoamyl alcohol (24:1) were added to the mixture and centrifuged (12000 rpm, 8 mins). This was followed by one phenol: chloroform: isoamyl alcohol (25:24:1) extraction of equal volume and centrifugation (12000 rpm, 3 min) and then a chloroform: isoamyl alcohol (24:1) extraction of equal volume and centrifugation (12000 rpm, 5 min). One volume room temperature isopropanol was added and incubated (-20°C, overnight). This was followed by centrifugation (12000 rpm, 10 min, 4°C). The pellet was washed with 70% ethanol, centrifuged (12000 rpm, 5 min, 4°C) and then allowed to air dry. The dry pellet was dissolved in 45 µl TE and 40 µg/ml RNase A and incubated (37°C, 30 min). Sodium acetate (3 M, pH 5, 0.1 volume) and 100% ethanol (2.5 volume) were added and centrifuged (12000 rpm, 10 min, 4°C). The pellet was washed twice with 70% ethanol and centrifuged (12000 rpm, 5 min, 4°C). The pellet was air dried and resuspended in 50 µl dH_2O (SABAX) and stored at -20°C until required.

The quality and quantity of all extracted DNA was established with a NanoDrop® ND-1000 Spectrophotometer according to the manufacturer's instructions. The concentration of the 39 samples in Table 3-1 (number 40 did not germinate) was confirmed on a 1% agarose gel (specifications given in section 3.7.1) using 0.5 µg/µl Lambda (λ) DNA standards (Promega). A concentration series of 0.1 µg/µl and 0.3 µg/µl was used for extracted gDNA quantification.

Each sample was diluted to 100 ng/µl for microsatellite analysis and to 50 ng/µl for AFLP analysis. Each sample used in the final phase of this study was diluted to 50 ng/µl which is sufficient for both microsatellite and AFLP methods.

3.3. MICROSATELLITE ANALYSIS

Five plant samples, each from a different breeding program (Table 3-1), were tested using thirty nine microsatellite primer pairs (Table 3-2). This was to test for polymorphisms between the breeding programmes. The reaction conditions were the same for each microsatellite. Only the annealing temperature of each primer pair differs and these are given in Table 3-2.

3.3.1. Primer annealing temperature

The annealing temperature of the microsatellite primers used is given in Table 3-2. The temperatures were calculated on the Integrated DNA Technologies website (http://www.idtdna.com/Home/Home.aspx), as well as on the PCR machines in the laboratory. Primer and repeat sequences as well as product sizes are from Graingenes 2.0 (http://wheat.pw.usda.gov/GG2/index.shtml) or from the literature referenced after Table 3-2.

Most of the microsatellites used in this study were dinucleotide repeats such as $(CT)_n$ or $(GA)_n$ or $(CA)_n$ or $(GT)_n$. This can be seen in Table 3-2. The $(AT)_n$ sequence is the most abundant microsatellite found in plants, but was not used in this study.

Table 3-2 Microsatellite primers used in this study.

Primer Name	Primer Sequence	Annealing Temp. (°C)	Repeat Sequence/SSR Size
***Xbarc12-3A* F**	CGA CAG AGT GAT CAC CCA AAT ATA A	53	$(TAA)_{28}$
***Xbarc12-3A* R**	CAT CGG TCT AAT TGT CAA TGT A		
***Xbarc19-3A* F**	GCG ACC CGA GTA GCC TGA A	58	$(TAA)_{18}$
***Xbarc19-3A* R**	GGT GGA CCA TTA GAC GCT TAC TTG		
***Xbarc206-6A* F**	GCT TTG CCA GGT GAG CAC TCT	61	$(CT)_{10}$
***Xbarc206-6A* R**	TGG CCG GGT ATT TGA GTT GGA GTT T		
***Xbarc1021-3A* F**	GGA AGG ACC TGA CTG ACT GCA TCT G	58	$(TAGA)_8$
***Xbarc1021-3A* R**	GCG ATC ACA ACC ATT CTT TTT AAC TA		279
***Xcfd58-1A, 1D* F**	AAT GGG CCT TTA AGA GCA AAA	55	$(CT)_5(CA)_{10}$
***Xcfd58-1A, 1D* R**	AGG GGT GAA AGG TTG GAG AC		179
***Xcfd65-1D* F**	AGA CGA TGA GAA GGA AGC CA	55	$(CT)_{32}$
***Xcfd65-1D* R**	CCT CCC TTG TTT TTG GGA TT		199
***Xcfd233-2D* F**	GAA TTT TTG GTG GCC TGT GT	60	$(GA)_{38}$
***Xcfd233-2D* R**	ATC ACT GCA CCG ACT TTT GG		289
***Xgwm2-3A* F**	CTG CAA GCC TGT GAT CAA CT	52	$(CA)_{18}$
***Xgwm2-3A* R**	CAT TCT CAA ATG ATC GAA CA		132
***Xgwm3-3D* F**	GCA GCG GCA CTG GTA CAT TT	55	$(CA)_{18}$
***Xgwm3-3D* R**	AAT ATC GCA TCA CTA TCC CA		79
***Xgwm18-1B, 4B* F**	TGG CGC CAT GAT TGC ATT ATC TTC	57	$(CA)_{17}GA(TA)_4$
***Xgwm18-1B, 4B* R**	GGT TGC TGA AGA ACC TTA TTT AGG		183
***Xgwm99-1A* F**	AAG ATG GAC GTA TGC ATC ACA	53	$(CA)_{21}$
***Xgwm99-1A* R**	GCC ATA TTT GAT GAC GCA TA		117
***Xgwm108-3B* F**	CGA CAA TGG GGT CTT AGC AT	55	$(GT)_{35}$imp
***Xgwm108-3B* R**	TGC ACA CTT AAA TTA CAT CCG C		135
***Xgwm135-1A* F**	TGT CAA CAT CGT TTT GAA AAG G	56	$(GA)_{20}$
***Xgwm135-1A* R**	ACA CTG TCA ACC TGG CAA TG		153
***Xgwm136-1A* F**	GAC AGC ACC TTG CCC TTT G	56	$(CT)_{58}$
***Xgwm136-1A* R**	CAT CGG CAA CAT GCT CAT C		278
***Xgwm155-3A* F**	CAA TCA TTT CCC CCT CCC	53	$(CT)_{19}$
***Xgwm155-3A* R**	AAT CAT TGG AAA TCC ATA TGC C		144
***Xgwm160-4A* F**	TTC AAT TCA GTC TTG GCT TGG	54	$(GA)_{21}$
***Xgwm160-4A* R**	CTG CAG GAA AAA AAG TAC ACC C		182
***Xgwm190-5D* F**	GTG CTT GCT GAG CTA TGA GTC	55	$(CT)_{22}$
***Xgwm190-5D* R**	GTG CCA CGT GGT ACC TTT G		209
***Xgwm285-3B* F**	ATG ACC CTT CTG CCA AAC AC	55	$(GA)_{27}$
***Xgwm285-3B* R**	ATC GAC CGG GAT CTA GCC		222

Primer Name	Primer Sequence	Annealing Temp. (°C)	Repeat Sequence/SSR Size
***Xgwm314-3D* F**	AGG AGC TCC TCT GTG CCA C	55	$(CT)_{25}$imp
***Xgwm314-3D* R**	TTC GGG ACT CTC TTC CCT G		182
***Xgwm325-6D* F**	TTT CTT CTG TCG TTC TCT TCC C	55	$(CT)_{16}$
***Xgwm325-6D* R**	TTT TTA CGC GTC AAC GAC G		137
***Xgwm337-1D* F**	CCT CTT CCT CCC TCA CTT AGC	55	$(CT)_5(CACT)_6(CA)_{43}$
***Xgwm337-1D* R**	TGC TAA CTG GCC TTT GCC		177
***Xgwm340-3B* F**	GCA ATC TTT TTT CTG ACC ACG	55	$(GA)_{26}$
***Xgwm340-3B* R**	ACG AGG CAA GAA CAC ACA TG		159
***Xgwm341-3D* F**	TTC AGT GGT AGC GGT CGA G	55	$(CT)_{26}$
***Xgwm341-3D* R**	CCG ACA TCT CAT GGA TCC AC		166
***Xgwm357-1A* F**	TAT GGT CAA AGT TGG ACC TCG	55	$(GA)_{18}$
***Xgwm357-1A* R**	AGG CTG CAG CTC TTC TTC AG		122
***Xgwm369-3A* F**	CTG CAG GCC ATG ATG ATG	55	$(CT)_{11}(T)_2(CT)_{21}$
***Xgwm369-3A* R**	ACC GTG GGT GTT GTG AGC		184
***Xgwm389-3B* F**	ATC ATG TCG ATC TCC TTG ACG	55	$(CT)_{14}(GT)_{16}$
***Xgwm389-3B* R**	TGC CAT GCA CAT TAG CAG AT		129
***Xgwm437-7D* F**	GAT CAA GAC TTT TGT ATC TCT C	47	$(CT)_{24}$
***Xgwm437-7D* R**	GAT GTC CAA CAG TTA GCT TA		107
***Xgwm458-1D* F**	AAT GGC AAT TGG AAG ACA TAG C	55	$(CA)_{13}$
***Xgwm458-1D* R**	TTC GCA ATG TTG ATT TGG C		112
***Xgwm513-4B* F**	ATC CGT AGC ACC TAC TGG TCA	55	$(CA)_{12}$
***Xgwm513-4B* R**	GGT CTG TTC ATG CCA CAT TG		140
***Xgwm539-2D* F**	CTG CTC TAA GAT TCA TGC AAC C	60	$(GA)_{27}$
***Xgwm539-2D* R**	GAG GCT TGT GCC CTC TGT AG		143
***Xgwm550-1B* F**	CCC ACA AGA ACC TTT GAA GA	54	$(CT)_8(GT)_{18}$
***Xgwm550-1B* R**	CAT TGT GTG TGC AAG GCA C		156
***Xwmc11-1A, 3A* F**	TTG TGA TCC TGG TTG TGT TGT GA	56	(CT)
***Xwmc11-1A, 3A* R**	CAC CCA GCC GTT ATA TAT GTT GA		177
***Xwmc25-2B, 2D* F**	TCT GGC CAG GAT CAA TAT TAC T	50	$(GT)_{26}$
***Xwmc25-2B, 2D* R**	TAA GAT ACA TAG ATC CAA CAC C		166
***Xwmc59-1A, 6A* F**	TCA TTC GTT GCA GAT ACA CCA C	58	$(CA)_{19}$
***Xwmc59-1A, 6A* R**	TCA ATG CCC TTG TTT CTG ACC T		197
***Xwmc167-2A* F**	AGT GGT AAT GAG GTG AAA GAA G	52	NA
***Xwmc167-2A* R**	TCG GTC GTA TAT GCA TGT AAA G		185
***Xwmc169-3A* F**	TAC CCG AAT CTG GAA AAT CAA T	54	$(CA)_{25}$
***Xwmc169-3A* R**	TGG AAG CTT GCT AAC TTT GGA G		167
***Xwmc177-2A* F**	AGG GCT CTC TTT AAT TCT TGC T	52	$(CA)_{21}$
***Xwmc177-2A* R**	GGT CTA TCG TAA TCC ACC TGT A		184

Primer Name	Primer Sequence	Annealing Temp. (°C)	Repeat Sequence/SSR Size
***Xwmc256-6A* F**	CCA AAT CTT CGA ACA AGA ACC C	56	$(CA)_{12}$
***Xwmc256-6A* R**	ACC GAT CGA TGG TGT ATA CTG A		117
***Xwmc532-3A* F**	GAT ACA TCA AGA TCG TGC CAA A	56	$(GA)_{11}$
***Xwmc532-3A* R**	GGG AGA AAT CAT TAA CGA AGG G		176

BARC – Beltsville Agriculture Research Center (Kong *et al*. 2005)

CFD – Pierre Sourdille microsatellite (Kong *et al*. 2005)

GWM – Gatersleben Wheat Microsatellite (Ahmad 2002)

WMC – Wheat Microsatellite Consortium (Gupta *et al*. 1999; Rampling *et al*. 2001)

imp = imperfect repeat

The following nine microsatellites from Table 3-2 were recommended by Huang *et al*. (2002) and Röder *et al*. (2002) as being very polymorphic: *Xgwm3-3D*, *Xgwm18-1B*, *Xgwm155-3A*, *Xgwm160-4A*, *Xgwm190-5D*, *Xgwm325-6D*, *Xgwm357-1A*, *Xgwm437-7D* and *Xgwm513-4B*. The following four microsatellites were recommended by Prasad *et al*. (2000): *Xwmc25-2D*, *Xwmc167-2D*, *Xwmc169-3A* and *Xwmc177-2A*.

3.3.2. Polymerase chain reaction (PCR)

All PCR reactions were performed in either a GeneAmp® PCR System 2700, 9700 or a Thermal Cycler 2720 (Applied Biosystems). PCR conditions were based on Röder *et al*. (1998). PCR conditions used for each microsatellite reaction were as follows: approximately 300 ng of template DNA, 2x Bioline PCR NH_4 reaction buffer [16 mM $(NH_4)_2SO_4$, 67 mM Tris-HCl pH 8.8 at 25°C, 0.01% Tween-20], 0.2 mM dNTPs, 0.5 µM forward primer, 0.5 µM reverse primer, 1.5 mM $MgCl_2$ and 1 U Bioline BIOTAQ™ DNA polymerase per 20 µl reaction.

PCR cycling conditions were as follows: 3 min denaturation at 94°C followed by 45 cycles of 1 min at 94°C, 1 min at temperature given in Table 3-2, 2 min at 72°C and a final step of 10 min at 72°C. The samples were kept at 4°C until needed.

An agarose gel was used to confirm PCR amplification and a denaturing polyacrylamide gel was used to detect polymorphisms in the microsatellites.

3.3.3. Marker evaluation on panel of released varieties/advanced lines

One plant from each of the 20 genotypes (samples 1-20, Table 3-1) was screened with the eighteen microsatellites that showed polymorphism between the breeding programmes. The following microsatellites were used: *Xbarc12-3A, Xbarc19-3A, Xcfd58-1A & 1D, Xgwm18-1B & 4B, Xgwm155-3A, Xgwm160-4A, Xgwm190-5D, Xgwm285-3B, Xgwm325-6D, Xgwm340-3B, Xgwm437-7D, Xgwm539-2D, Xwmc11-1A & 3A, Xwmc25-2B & 2D, Xwmc59-1A & 6A, Xwmc177-2A, Xwmc256-6A* and *Xwmc532-3A*.

3.3.4. Fluorescence-based semi-automated analysis

The reaction conditions for semi-automated analysis are the same as for microsatellites analysed on a denaturing polyacrylamide gel. The only difference is that one of the primers in the reaction is labelled with a fluorescent dye. These microsatellite primers could also be multiplexed because of the different colours of the labels.

Table 3-3 Labelled microsatellite primers used in this study.

Primer Name	Primer Sequence	Dye
***Xgwm190-5D* F**	GTG CTT GCT GAG CTA TGA GTC	NED (yellow)
***Xgwm190-5D* R**	GTG CCA CGT GGT ACC TTT G	-
***Xgwm437-7D* F**	GAT CAA GAC TTT TGT ATC TCT C	NED (yellow)
***Xgwm437-7D* R**	GAT GTC CAA CAG TTA GCT TA	-
***Xgwm539-2D* F**	CTG CTC TAA GAT TCA TGC AAC C	PET (red)
***Xgwm539-2D* R**	GAG GCT TGT GCC CTC TGT AG	-
***Xwmc11-1A, 3A* F**	CAC CCA GCC GTT ATA TAT GTT GA	FAM (blue)
***Xwmc11-1A, 3A* R**	GTT GTG ATC CTG GTT GTG TTG TGA	-
***Xwmc59-1A, 6A* F**	TCA TTC GTT GCA GAT ACA CCA C	PET (red)
***Xwmc59-1A, 6A* R**	GTC AAT GCC CTT GTT TCT GAC CT	-
***Xwmc177-2A* F**	AGG GCT CTC TTT AAT TCT TGC T	VIC (green)
***Xwmc177-2A* R**	GGT CTA TCG TAA TCC ACC TGT A	-

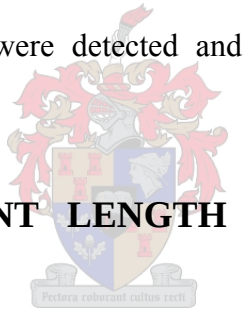
All PCR reactions were performed in either a GeneAmp® PCR System 2700, 9700 or a Thermal Cycler 2720 (Applied Biosystems). PCR conditions were based on Röder *et al*. (1998). PCR conditions used for the multiplex microsatellite reaction were as follows: approximately 100 ng of template DNA, 2x Bioline PCR NH_4 reaction buffer, 0.2 mM

dNTPs, 1.5 mM $MgCl_2$ and 1 U Bioline BIOTAQ™ DNA polymerase per 20 µl reaction. The primer concentrations were 0.5 µM of each labelled forward primer and 0.5 µM of each unlabelled reverse primer except for the *Xgwm190* primer pair. The concentration used for that primer was 0.6 µM for both the labelled forward primer and unlabelled reverse primer.

Microsatellites *Xgwm437* and *Xgwm539* were not included in the multiplex reaction. These loci were amplified in separate PCR reactions with the same reaction and cycling conditions as described in section 3.3.2. The only differences were the following: 100 ng of template DNA, one labelled primer (Table 3-3) and 0.5 U Bioline BIOTAQ™ DNA polymerase per 10 µl reaction.

PCR cycling conditions for the multiplex reaction were as follows: 3 min denaturation at 94°C followed by 45 cycles of 1 min at 94°C, 1 min at 55°C, 3 min at 72°C and a final step of 10 min at 72°C. The samples were kept at 4°C until needed.

The product from these reactions were detected and visualised on an automated DNA sequencer.



3.4. AMPLIFIED FRAGMENT LENGTH POLYMORPHISM (AFLP) ANALYSIS

A modified version of the Donini *et al*. (1997) technique was employed using reagents and concentrations given in the AFLP™ Plant Mapping Kit protocol (Applied Biosystem 1997). Schwarz *et al*. (2000) was consulted to automate the technique using a multiplex selective PCR reaction and fluorescently labelled primers.

This technique has four steps: 1) the restriction digestion of genomic DNA and the ligation of adaptors; 2) the amplification with preselective primers; 3) the amplification of the restriction fragments with selective primers and 4) the detection of the amplified products.

3.4.1. Restriction digestion of genomic DNA and ligation of adaptors

About 300 ng of total genomic DNA was digested with 5 units of *Eco*RI (Roche) and 5 units *Mse*I (New England Biolabs) restriction enzymes in a total reaction volume of 20 µl. One

microlitre *Eco*RI adaptor (5' – CTC GTA GAC TGC GTA CC – 3'/3' – CAT CTG ACG CAT GGT TAA – 5') and 1 µl *Mse*I adaptor (5' – GAC GAT GAG TCC TGA G – 3'/3' – TA CTC AGG ACT CAT – 5') was ligated to the digested DNA in the same reaction. The reaction also included 0.1 µg/µl BSA (New England Biolabs), 1x One-Phor-All buffer (USB), 1 mM ATP (Pharmacia Biotech), 1 unit T4 DNA ligase (USB) and ddH_2O water. The reaction was mixed, centrifuged briefly and incubated at room temperature overnight.

A 1:9 dilution was made using 15 µl of each of the restriction digestion and ligation reactions using 1x $TE_{0.1}$ buffer (10 mM Tris-HCl, 0.1 mM Na_2EDTA, pH 8). The remaining 5 µl was tested on a 1.5% agarose gel to confirm digestion and ligation.

3.4.2. Amplification with preselective primers

The amplification with preselective primers was performed in a thin-walled 0.2 ml PCR tube. This reaction has a total volume of 20 µl per sample, which consists of 5.2 µl of the diluted restriction-ligation template, 1.5 µl preselective primer mix [E00 primer (5' – GAC TGC GTA CCA ATT C – 3'), M00 primer (5' – GAT GAG TCC TGA GTA A – 3')], 1x Bioline PCR NH_4 reaction buffer, 0.2 mM dNTPs, 1.5 mM $MgCl_2$ and 1 U Bioline BIOTAQ™ DNA polymerase.

PCR cycling conditions were as follows: 5 min elongation at 72°C followed by 30 cycles of 30 seconds at 94°C, 1 min at 56°C, 1 min at 72°C and a final elongation of 5 min at 72°C. The samples were kept at 4°C until needed.

A 1:9 dilution was made using 15 µl of each of the amplification reactions using 1x $TE_{0.1}$ buffer. The remaining 5 µl was tested on a 1.5% agarose gel to confirm amplification.

3.4.3. Amplification of restriction fragments with selective primers

The selective amplification reaction was performed in a thin-walled 0.2 ml PCR tube. The total reaction volume was 10 µl which consisted of 2.5 µl template DNA (diluted preselective amplification product), 0.5 µM unlabeled *Mse*I primer (5' – GAT GAG TCC TGA GTA ANN N – 3', N denotes the specific selective nucleotides given in Table 3-4), 0.05 µM of each labelled *Eco*RI primer {5' – GAC TGC GTA CCA ATT CNN N – '3, N denotes the

specific selective nucleotides given in Table 3-4 (three different colours used in one reaction)}, 1x Bioline PCR NH_4 reaction buffer, 0.2 mM dNTPs, 1.5 mM $MgCl_2$ and 0.25 U Bioline BIOTAQ™ DNA polymerase.

Table 3-4 Selective primer combinations used for AFLP analysis. The numbers represent the primer combinations.

***Mse*I Primer**	***Eco*RI Primers**		
	*Eco*RI – ACA FAM	*Eco*RI – AAC NED	*Eco*RI – AGG JOE
*Mse*I – CAG	1	2	3
*Mse*I – CAT	4	5	6
*Mse*I – CTC	7	8	9
Label Colour	blue	yellow	green

Groenewald *et al*. (2005) showed that the selective primers in Table 3-4 were polymorphic in wheat.

PCR cycling conditions were as follows: 1 cycle of 30 seconds at 94°C, 30 seconds at 65°C, 1 min at 72°C, followed by 12 cycles in which the annealing temperature decreases by 0.7°C each cycle. This was followed by 23 cycles of 30 seconds at 94°C, 30 seconds at 56°C, 1 min at 72°C and a final soak temperature of 4°C. The samples were kept at 4°C until electrophoresis.

3.4.4. Detection of amplified products and data analysis

The selective amplification products can be detected either by separation on a denaturing polyacrylamide gel using radioactively {γ-^{33}P} labelled primers or on an automated DNA sequencer using fluorescently labelled primers. Fluorescently labelled *Eco*RI primers were used in this thesis to selectively amplify target sequences. The final amplification products were detected on an automated DNA sequencer and analysed using GeneMapper. The results from GeneMapper were further analysed using Microsoft Excel and SAS.

3.5. RETROTRANSPOSONS

There are various techniques that use transposons as either the target element or part of the target DNA. Primers were chosen from those given in literature that were designed for the *BARE*-1 retrotransposon found in barley (Kalendar *et al*. 1999) or designed based on the *Wis-2* retrotransposon found in the wheat genome. The three techniques are described in the following subsections.

3.5.1. Inter-Retrotransposon Amplified Polymorphism (IRAP) analysis

This technique amplifies a region of DNA between two adjacent retrotransposons. Part of the primers given in Table 3-5 bind to a sequence in the *Wis-2* retrotransposon and the rest of the primer binds to a sequence that was added to the fragment to assist in the making and cloning of the marker. The primers were tested in various combinations and used individually in a single primer PCR.

Table 3-5 IRAP primers recommended by Kalendar *et al*. (1999) and tested in this study.

Primer Name	Primer Sequence	Annealing Temp (°C)
LTR-R 6150	*CTGGTTCGGCCC*ATGTCTATGTATCCACACATGGTA	60
LTR-F 6149	*CTCGCTCGCCCA*CTACATCAACCGCGTTTATT	

All PCR reactions were performed in either a GeneAmp® PCR System 2700, 9700 or a Thermal Cycler 2720 (Applied Biosystems). PCR conditions were adapted from Kalendar *et al*. (1999). PCR conditions used were as follows: approximately 50 ng of template DNA, 1x Bioline PCR NH_4 reaction buffer, 0.2 mM dNTPs, 0.5 µM LTR-F 6149 primer, 0.5 µM LTR-R 6150 primer, 2 mM $MgCl_2$ and 1 U Bioline BIOTAQ™ DNA polymerase per 20 µl reaction.

PCR cycling conditions were as follows: 2 min denaturation at 94°C followed by 30 cycles of 30 seconds at 94°C, 30 seconds at 60°C, 2 min at 72°C and a final step of 5 min at 72°C. The PCR products were kept at 4°C until needed.

A 1.5% agarose gel was used to visualise the PCR products.

3.5.2. Retrotransposon-Microsatellite Amplified Polymorphism (REMAP) analysis

This technique amplifies a region of DNA between a microsatellite and a retrotransposon, which makes primer selection very specific. The one primer is a short sequence complementary to the microsatellite that will be amplified plus one anchoring base pair. The other primer will target the retrotransposon and bind to it. Thus a piece of DNA between the retrotransposon and microsatellite will be amplified. Table 3-6 indicates the primers that were recommended and used in this study. Each microsatellite primer was used together with the LTR-R 7286 primer at the annealing temperature given in the table.

Each primer, number 2 to 8, was used in combination with the LTR primer in a PCR reaction numbered the same as the primer in Table 3-6.

Table 3-6 REMAP primers recommended by Kalendar *et al*. (1999) and tested in this study.

Primer / Reaction Number	Primer Name	Primer Sequence	Annealing Temp (°C)
	LTR-R 7286	*GGAATTC*ATAGCATGGATAATAAACGATTATC	
2	**8081**	$(GA)_9C$	55
3	**8082**	$(CT)_9G$	55
4	**8385**	$(CAC)_7G$	58
5	**8386**	$(GTG)_7C$	58
6	**8387**	$(CA)_{10}G$	55
7	**8564**	$(CAC)_7T$	58
8	**8565**	$GT(CAC)_7$	58

All PCR reactions were performed in either a GeneAmp® PCR System 2700, 9700 or a Thermal Cycler 2720 (Applied Biosystems). PCR conditions were adapted from Kalendar *et al*. (1999). PCR conditions used for primer combinations 2, 3, 6, 7 and 8 (Table 3-6) were as follows: approximately 25 ng of template DNA, 1x Bioline PCR NH_4 reaction buffer, 0.2 mM dNTPs, 0.4 µM microsatellite primer, 0.4 µM LTR-R 7286 primer, 2 mM $MgCl_2$ and 1 U Bioline BIOTAQ™ DNA polymerase per 20 µl reaction. PCR conditions used for primer combinations 4 and 5 (Table 3-6) were the same as those given above except that approximately 50 ng of template DNA was used per 20 µl reaction.

PCR cycling conditions were as follows: 2 min denaturation at 94°C followed by 30 cycles of 30 seconds at 94°C, 30 seconds at temperature given in Table 3-6, 2 min at 72°C and a final step of 5 min at 72°C. The samples were kept at 4°C until needed.

A 1.5% agarose gel was used to visualise the PCR products.

3.5.3. Sequence-Specific Amplified Polymorphism (SSAP) analysis

This reaction is based on similar principles to the AFLP technique. It amplifies a fragment between the *Wis-2* retrotransposon and a restriction enzyme recognition site. The restriction-ligation reaction and preselective primer amplification reaction for this technique is identical to those in the AFLP technique. Thus the preselective PCR product used in AFLP analysis was amplified with the primers given in Table 3-7.

Table 3-7 SSAP primers recommended by Leigh *et al*. (2003a), Queen *et al*. (2004) and Waugh *et al*. (1997) and used in this study.

Primer Name	Primer Sequence	Annealing Temperature
***Wis-2* L** (blue)	FAM-CTAGGGCATAATTCCAACA	65°C to 56°C
***Mse*I-CAG**	GATGAGTCCTGAGTAACAG	
***Mse*I-CAT**	GATGAGTCCTGAGTAACAT	
***Mse*I-CTA**	GATGAGTCCTGAGTAACTA	
***Mse*I-CTC**	GATGAGTCCTGAGTAACTC	
***Mse*I-CTG**	GATGAGTCCTGAGTAACTG	
***Mse*I-CTT**	GATGAGTCCTGAGTAACTT	
***Wis-2* U**	CTAGGGCATAATTCCAACA	65°C to 56°C
***Eco*RI-AAC** (yellow)	NED-GACTGCGTACCAATTCAAC	
***Eco*RI-ACA** (blue)	FAM-GACTGCGTACCAATTCACA	
***Eco*RI-AGG** (green)	JOE-GACTGCGTACCAATTCAGG	

The *Mse*I and *Wis-2* L selective amplification reaction was performed in a thin-walled 0.2 ml PCR tube. The total reaction volume was 20 µl which consisted of 1 µl template DNA (diluted preselective amplification product), 0.15 µM unlabeled *Mse*I primer (5' – GAT GAG TCC TGA GTA ANN N – 3', N denotes the specific selective nucleotides given in Table 3-

7), 0.25 µM labelled *Wis-2* L primer (given in Table 3-7), 1x Bioline PCR NH_4 reaction buffer, 0.25 mM dNTPs, 2 mM $MgCl_2$ and 0.5 U Bioline BIOTAQ™ DNA polymerase.

The *Eco*RI and *Wis-2* U selective amplification reaction was performed in a thin-walled 0.2 ml PCR tube. The total reaction volume was 20 µl which consisted of 2 µl template DNA (diluted preselective amplification product), 0.5 µM unlabeled *Wis-2* U primer (given in Table 3-7), 0.03 µM labelled *Eco*RI primer (5' – GAC TGC GTA CCA ATT CNN N – '3, N denotes the specific selective nucleotides given in Table 3-7), 1x Bioline PCR NH_4 reaction buffer, 0.25 mM dNTPs, 1.5 mM $MgCl_2$ and 0.5 U Bioline BIOTAQ™ DNA polymerase.

PCR cycling conditions for both reactions were as follows: 1 cycle of 30 seconds at 94°C, 30 seconds at 65°C, 1 min at 72°C, followed by 12 cycles in which the annealing temperature decreased by 0.7°C each cycle. This was followed by 23 cycles of 30 seconds at 94°C, 30 seconds at 56°C, 1 min at 72°C and a final soak temperature of 4°C. The samples were kept at 4°C until electrophoresis.

The final SSAP products were electrophoresed and visualised using the automated DNA sequencer and GeneMapper. The results were treated in the same way as the AFLP results. Microsoft Excel was used to analyse the results from GeneMapper. Final analysis was done using SAS.

3.5.4. *Wis-2* Retrotransposon Amplification analysis

The *Wis-2* retrotransposon, which is found in the wheat genome, has a similar sequence to the *BARE*-1 retrotransposon in barley. This technique uses the *Wis-2* retrotransposon sequence in a modified version of the IRAP technique (section 3.5.1.). These modifications were not taken from previously published literature, but were investigated for the first time in this study.

A forward and reverse primer was designed based on the *Wis-2* retrotransposon present in wheat. The retrotransposon sequence was found on the GrainGenes website (http://wheat.pw.usda.gov/GG2/index.shtml) and BioEdit was used to design the primers given in Table 3-8.

Table 3-8 Characteristics of primers developed from the *Wis-2* retrotransposon.

Primer Name	Primer Sequence	Annealing Temperature
***Wis-2* F**	GAG TAG AAC ACA AAG TA	55°C
***Wis-2* R**	TCA AGG TCA CAT CCC CA	

All PCR reactions were performed in either a GeneAmp® PCR System 2700, 9700 or a Thermal Cycler 2720 (Applied Biosystems). PCR conditions used were as follows: approximately 25 ng of template DNA, 1x Bioline PCR NH_4 reaction buffer, 0.2 mM dNTPs, 0.5 µM *Wis-2* F primer, 0.5 µM *Wis-2* R primer, 2 mM $MgCl_2$ and 1 U Bioline BIOTAQ™ DNA polymerase per 20 µl reaction.

PCR cycling conditions were as follows: 1 min denaturation at 94°C followed by 40 cycles of 30 seconds at 94°C, 30 seconds at 55°C, 1 min at 72°C and a final step of 5 min at 72°C. The samples were kept at 4°C until needed.

A 1% agarose gel was used to visualise the PCR products.

3.6. IMPLEMENTATION OF SELECTED MARKERS

A population of 119 breeding lines (Table 3-9) was used to confirm the diagnostic methodology obtained in this study. Selected microsatellite and AFLP primer sets were used.

Table 3-9 List of 119 breeding lines that were screened.

Sample Number	Breeding Line	Sample Number	Breeding Line	Sample Number	Breeding Line	Sample Number	Breeding Line
1	03H1	**31**	03H93-3	**61**	03H211-2	**91**	03H313-1
2	03H8	**32**	03H95-1	**62**	03H212-5	**92**	03H315-1
3	03H12	**33**	03H100-1	**63**	03H213-7	**93**	03H318-2
4	03H26	**34**	03H102-3	**64**	03H216-7	**94**	03H322-3
5	03H29	**35**	03H103-2	**65**	03H217-2	**95**	03H327-3
6	03H30	**36**	03H104-3	**66**	03H222	**96**	03H328-2
7	03H31	**37**	03H105-3	**67**	03H228	**97**	03H333-2
8	03H34-2	**38**	03H106-2	**68**	03H230-1	**98**	03H339-2

Sample Number	Breeding Line	Sample Number	Breeding Line	Sample Number	Breeding Line	Sample Number	Breeding Line
9	03H45-1	**39**	03H108-7	**69**	03H235-1	**99**	03H342-2
10	03H49-2	**40**	03H109-1	**70**	03H237-2	**100**	03H344-5
11	03H56-1	**41**	03H110-9	**71**	03H242-2	**101**	03H345-4
12	03H57-1	**42**	03H112-4	**72**	03H244-2	**102**	03H349-2
13	03H59-1	**43**	03H118-3	**73**	03H257-1	**103**	03H350-8
14	03H61-1	**44**	03H128	**74**	03H260-1	**104**	03H352
15	03H65-1	**45**	03H132	**75**	03H265-1	**105**	03H364-1
16	03H66-2	**46**	03H148	**76**	03H266-2	**106**	03H365-1
17	03H68-1	**47**	03H149	**77**	03H272-1	**107**	03H369-1
18	03H69-4	**48**	03H178-3	**78**	03H277-2	**108**	03H370-2
19	03H70-1	**49**	03H183-1	**79**	03H278-1	**109**	03H375-2
20	03H72-1	**50**	03H185-1	**80**	03H279-2	**110**	03H377-4
21	03H74-4	**51**	03H189-2	**81**	03H280-4	**111**	03H379-3
22	03H75-1	**52**	03H191-2	**82**	03H281-3	**112**	03H380-3
23	03H76-1	**53**	03H192-2	**83**	03H282-4	**113**	03H381-4
24	03H77-2	**54**	03H195-3	**84**	03H283-2	**114**	03H402-2
25	03H80-2	**55**	03H196-1	**85**	03H285-1	**115**	03H409-1
26	03H82-2	**56**	03H200-1	**86**	03H293	**116**	03H412-3
27	03H85-4	**57**	03H203-2	**87**	03H297	**117**	03H414-2
28	03H86-5	**58**	03H204-3	**88**	03H301	**118**	03H415-3
29	03H88-1	**59**	03H206-4	**89**	03H310-1	**119**	03H418-1
30	03H89-2	**60**	03H207-3	**90**	03H312-2		

The samples highlighted in Table 3-9 were used to determine the minimum sample size that can accurately represent the line.

3.6.1. Microsatellite analysis

The 119 plants given in Table 3-9 were analysed with the six microsatellite primer pairs listed in Table 3-3. These are: *Xgwm190-5D, Xgwm437-7D, Xgwm539-2D, Xwmc11-1A & 3A, Xwmc59-1A & 6A* and *Xwmc177-2A*. The reaction conditions were the same as those described in section 3.3.4.

3.6.2. Amplified Fragment Length Polymorphism (AFLP) analysis

The 119 samples were tested with only three of the previously described AFLP primer combinations (combinations 7, 8 and 9 in Table 3-4). Only the *Mse*I-CTC primer was used in combination with the three *Eco*RI primers. The technique was performed using the same protocol as described in section 3.4.

The sixteen samples highlighted in Table 3-9 were used to test the sample size that is adequate to assess the identity of a population. Eight and twelve seeds per breeding line were again tested with the three AFLP primer combinations.

3.7. ELECTROPHORESIS AND VISUALISATION

3.7.1. Agarose gel electrophoresis

Agarose gel electrophoresis was performed according to Sambrook *et al*. (1991).

DNA fragments were separated on 1 or 1.5% (w/v) agarose (Bioline) gels in 1x TBE [5x TBE stock: 0.5 M Tris-Cl, 0.5 M boric acid, 0.5 M Na_2EDTA (pH 8)] at 90 to 120 V for 30 to 45 minutes. A 5x TBE stock solution was prepared and diluted to 1x for the working stock. Ethidium bromide (0.5 µg/ml) was added to each agarose gel to visualise the DNA using a ultra-violet (UV) transilluminator. Six times (6x) loading buffer (Ficoll Orange G) was used to facilitate DNA loading into the wells.

3.7.2. Polyacrylamide gel electrophoresis

Five microlitres of the PCR product were electrophoresed on a 1.5% (w/v) agarose gel in 1x TBE buffer (100 V, 45 min). The remaining 15 µl of the reaction were prepared for polyacrylamide gel electrophoresis which was used to separate the DNA fragments. Fifteen microlitres formamide loading buffer [98% formamide, 10 mM Na_2EDTA pH 8, 0.05% (w/v) bromophenolblue, 0.05% (w/v) xylene cyanol FF] were added to each sample. These were denatured at 95°C for five minutes and immediately quenched on ice before loading into the polyacrylamide gel. The 6% w/v acrylamide: bis-acrylamide denaturing gel containing 6 M urea and 1x TBE was prepared as follows: 800 µl ammonium persulphate (10% w/v) and 160

µl N,N,N,N- Tetramethylenediamine (TEMED) to 160 ml 6% stock solution. The gel was cast in 1 mm spacers and comb and allowed to set for one hour. The polymerised gel was placed on a Model S 2001 sequencing electrophoresis apparatus (Life-Technologies™) and pre-run in 1x TBE buffer for 30 minutes at 70 W. The wells were flushed with 1x TBE buffer before loading the samples. Ten to twelve microlitres of each denatured sample were loaded into the polyacrylamide gel as well as 2 µl of a 100 bp ladder. The gel was electrophoresed at a constant 70 W for 5 – 6 hours. After electrophoresis, the gel was silver stained, covered in plastic and photographed using a digital camera.

3.7.3. Silver staining

Following electrophoresis, the glass plates were separated and the gel containing plate was placed in fixative (0.5% v/v glacial acetic acid and 10% v/v ethanol) for 20 minutes. This was followed by two five minute washing steps with dH_2O at room temperature. Next the gel was stained in 0.1% w/v $AgNO_3$ for 20 minutes at room temperature followed by a ten second rinse in dH_2O to remove the excess staining solution. The gel was placed in developing solution (1.5% w/v NaOH and 0.16% v/v formaldehyde added just before use) at 8°C until the bands appeared. This was followed by a final rinse in dH_2O before the gel was covered in plastic to prevent it from drying out.

3.7.4. Automated sequencer electrophoresis

An Applied Biosystems 3130XL Genetic Analyzer (California, USA) with a 50 cm capillary and POP-7 Polymer (California, USA) was used to detect polymorphisms between samples. The size standard GeneScan™ - 500 LIZ was added to each microsatellite sample. GeneScan™ - 500 ROX was added to every AFLP sample. Both size standards have a range of 75 to 500 bp. They do, however, emit at different wavelengths and are visualised as orange for LIZ and red for ROX. Due to manufacture's restrictions, microsatellites and AFLPs need to be analysed using different size standards. GeneMapper software was used to visualise and analyse the results obtained from the Genetic Analyzer. The genetic analyzer was used according to manufacturer's instructions.

3.8. DATA ANALYSIS

The SAS software program (Version 9.1) was used to perform a discriminant analysis and a cluster analysis. The microsatellite data obtained from the sequencer were analysed separately from the AFLP data taken from the sequencer. Discriminant analysis is a classification method that requires previous knowledge of the group membership of each sample. A cluster analysis is a classification method used when group membership is unknown.

The microsatellite data of the initial population of twenty released varieties/advanced lines were used to perform a cluster analysis. A discriminant analysis was performed on the AFLP data.

Microsatellite *Xwmc177* was analysed separately as an AFLP banding pattern. A discriminant analysis was performed on the initial population of twenty genotypes. A cluster analysis was performed on the 119 breeding lines.

The microsatellite and AFLP data obtained from the 119 breeding lines were subjected to cluster analysis using the SAS software.

The sixteen samples that were used to test the optimal sample size was also analysed with the SAS cluster method.

4. RESULTS AND DISCUSSION

The aim of this study was to develop a method to distinguish between common wheat genotypes. DNA marker techniques were evaluated based on polymorphism, cost, ease of use and ease of analysis.

Following a literature study it was decided to focus on and compare three techniques: microsatellites, AFLPs and retrotransposon-based markers. All of these are PCR based. Microsatellites were chosen due to their high level of specific polymorphism in wheat. It is also a rather easy technique to perform and analyse. AFLPs are very polymorphic and fairly easy to analyse. The technique itself is more time consuming than the microsatellite technique. Retrotransposon-based markers were investigated as an alternative to the two well known techniques and to test their level of polymorphism in wheat.

This study starts by evaluating a large group of marker loci and narrows them down to a subset that can be used routinely to investigate and identify bread wheat genotypes at the University of Stellenbosch.

The results of this study are given and discussed in the following sections:

1. Initial evaluation of each marker system
    i. Microsatellites;
    ii. Amplified Fragment Length Polymorphisms (AFLPs);
    iii. Retrotransposons.
2. Implementation of the selected markers on a population of breeding lines
    i. Microsatellites;
    ii. Amplified Fragment Length Polymorphisms (AFLPs);
3. System for genotyping wheat at Stellenbosch University.

4.1. INITIAL EVALUATION OF EACH MARKER SYSTEM

4.1.1. Microsatellites

The first marker system investigated was microsatellites. Two different detection methods were investigated. The first method was using polyacrylamide gel electrophoresis. This was

used to detect the banding pattern of each microsatellite; to see if there was any structure in the pattern and to identify representative bands that could be scored consistently. Denaturing polyacrylamide gels were used for the initial marker evaluation as well as the marker evaluation on the panel of released varieties/advanced lines.

The second detection method uses fluorescently labelled primers and the DNA sequencer to detect the amplified fragments. This system is easier and more precise, and allows for multiplexing, but the banding pattern of each microsatellite is not as clear as on the denaturing polyacrylamide gel. However, the bands targeted for scoring should be carefully chosen so that they do not overlap with the products of other loci. Only six microsatellites were chosen for this system and they were evaluated on all 39 samples of released varieties/advanced lines.

The initial marker evaluation included only five samples, one from each of the five breeding programmes (Table 3-1). The level of polymorphism of the marker could be evaluated and if there was any observed between the five samples then it was applied to the panel of twenty released varieties/advanced lines.

The five plants were tested with the 39 microsatellite primer pairs given in Table 3-2. Most of these primers were previously used in the laboratory and showed an easy to score banding pattern and a visible degree of polymorphism. Thirteen of the 39 microsatellite primer pairs were chosen from literature. Nine of these were recommended by Huang *et al*. (2002) and Röder *et al*. (2002). These microsatellites were chosen because they were used by both studies, showed a high level of polymorphism and covered most of the wheat chromosomes. The other four microsatellites were recommended by Prasad *et al*. (2000). According to the latter authors these four microsatellites showed a high level of polymorphism and together they could discriminate between 41 of the wheat genotypes that were tested in their study.

The genomic DNA extracted from the set of twenty genotypes (Table 3-1) was of good quality as can be seen in Figures 7-1 to 7-2 in the Addendum A. A subset of five plants (Table 3-1) was used in PCR amplifications using 39 microsatellite primer sets and the amplification products confirmed on agarose gels (Addendum A Figures 7-3 to 7-5, 7-8, 7-9 and 7-13). The amplified microsatellite products were visualised on denaturing polyacrylamide gels. Photos of these gels are given in Figures 7-6 to 7-7 and 7-10 to 7-12 and 7-14 in Addendum A. Based on the polymorphism detected, the polymorphic markers

were selected and applied to the panel of twenty released varieties/advanced lines. Eighteen of the 39 microsatellites were chosen to test this panel and are indicated (boxes) in the respective photographs.

4.1.1.1. Evaluation of the full set of 20 genotypes with eighteen selected microsatellite primer sets (visualized on denaturing polyacrylamide gels)

Polyacrylamide detection is not a very quick and easy system. It requires some experience and approximately 10 hrs to complete. Also, the denaturing polyacrylamide gel system can only accommodate 43 samples. This number gets less when controls and molecular sizing ladders need to be added. Determining allele size is also difficult with this detection system. Therefore allele scoring and reliability and reproducibility of the microsatellite data are not as exact using polyacrylamide gel electrophoresis. Here polyacrylamide gel electrophoresis was performed first in order to visualize all amplified products produced by a specific primer combination. The banding pattern thus obtained is easier to interpret on these gels compared to what is seen on the Genetic Analyzer making it easier to detect patterned (structured) profiles.

Photos of the denaturing polyacrylamide gels that were obtained are given in Figures 7-15 to 7-34 in Addendum A. With respect to each gel an area (pattern) or single band was selected for scoring and assignment of an allele symbol. This was done as the bands that occur in a pattern convey identical information and should therefore only be scored once. Also, areas where the band patterns of different genomic regions may overlap need to be avoided as it will complicate scoring and compromise accuracy. The results from these gels were summarised in Table 4-1. Only the data pertaining to a set of polymorphic bands that could be consistently scored are represented in Table 4-1. The bands scored are indicated with boxes on the appropriate gel photos and the name of each allele is given at the top of the gel photo. Letters indicate the different alleles observed for those microsatellites. The alleles were named from smallest to largest; where the smallest fragment is an A and the other letters indicate larger amplification products/alleles. The different alleles recognized are indicated on the gel photos. The positive/negative symbols indicate the presence or absence of an amplified product for the relevant microsatellites.

Table 4-1 Summary of polymorphisms detected among a panel of 20 released varieties and breeding lines. Highlighted microsatellites were chosen to test automated visualisation.

Markers	BREEDER AND VARIETY/LINE NUMBER																			
	US					PANNAR			SGI		Monsanto						CYMMIT			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Xbarc12	B	A	B	A	B	E	E	E	E	E	E	C	-	C	C	C	C	C	D	F
Xbarc19	B	B	A	A	A	A	A	A	A	A	A	A	B	A	A	-	A	A	A	A
Xcfd58	+	-	+	+	+	+	-	+	+	+	-	-	+	-	-	+	-	-	-	-
Xgwm18	-	+	-	-	-	+	-	+	+	-	+	+	+	+	+	+	-	+	-	+
Xgwm155	C	A	A	A	A	A	A	A	A	A	C	B	A	A	C	A	C	B	-	-
Xgwm160 (A)	+	-	-	-	-	+	-	-	+	+	+	-	+	+	-	+	-	-	+	-
Xgwm160 (B)	B	A	A	B	B	B	B	B	B	B	B	C	B	B	C	B	A	A	B	A
Xgwm190	B	B	B	A	B	A	A	A	A	A	A	A	A	A	A	A	B	A	A	-
Xgwm285	A	A	A	B	A	B	B	B	B	B	B	B	B	AB	A	B	A	A	A	A
Xgwm325	B	A	-	C	B	B	A	A	B	B	B	B	B	B	B	B	B	B	B	B
Xgwm340	C	A	A	A	A	C	A	A	C	C	C	C	B	C	C	B	C	B	C	D
Xgwm437	E	E	E	B	E	E	D	D	E	E	A	D	E	A	C	C	B	B	A	B
Xgwm539	D	C	C	B	F	B	B	B	B	B	B	B	C	B	D	C	D	A	E	G
Xwmc11	-	A	D	B	A	B	B	B	B	B	B	A	A	B	A	A	C	A	B	-
Xwmc25	-	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	-	-	-	-
Xwmc59	+	+	-	+	-	+	+	+	+	+	+	+	+	+	+	+	-	+	+	-
Xwmc177	C	B	C	A	D	A	D	D	A	A	A	B	C	A	C	C	A	C	A	B
Xwmc256	B	B	A	B	A	B	B	B	B	B	C	B	A	C	A	A	B	B	A	A
Xwmc532	B	C	C	A	A	B	A	A	B	B	A	B	B	A	B	B	B	A	A	B

The set of eighteen microsatellites clearly distinguished among the twenty varieties/ breeding lines. The six microsatellites that showed the most polymorphism and an amplification product larger than 75 bp were singled out for automated visualization and are given in Figures 4-1 to 4-6. One primer of these microsatellites was fluorescently labelled. Thus the results of the denaturing polyacrylamide gels could be compared to those received from the Applied Biosystems 3130XL Genetic Analyzer (automated DNA sequencer). The blocks shown on these figures and those in Addendum A (Fig. 7-15 to 7-34) highlight/indicate the alleles that were used to score each microsatellite. Also, shown (arrow) is the specific band in each pattern/ region that was targeted for scoring with the automated detection system; this was always the smallest molecular size band in a patterned polymorphism. The alleles allocated to each sample (Table 4-1) as well as the sample number (Table 3-1) are given at the top of each gel photo.

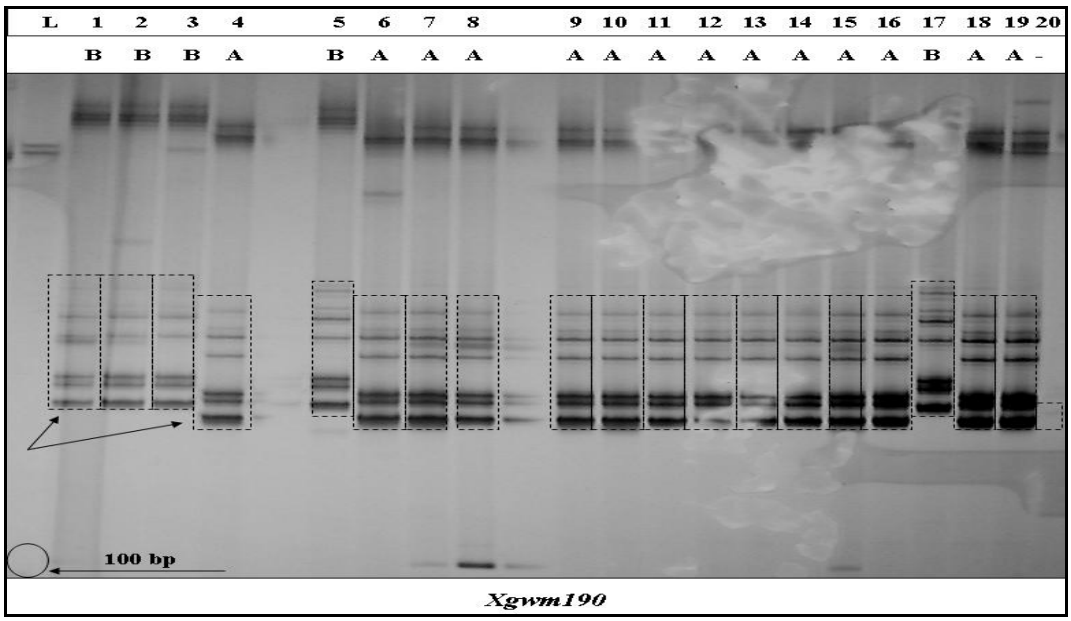


Figure 4-1 Gel photo illustrating the amplification of the *Xgwm190* microsatellite (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The sample numbers and alleles are given above each lane. The specific band targeted for automated detection is arrowed.

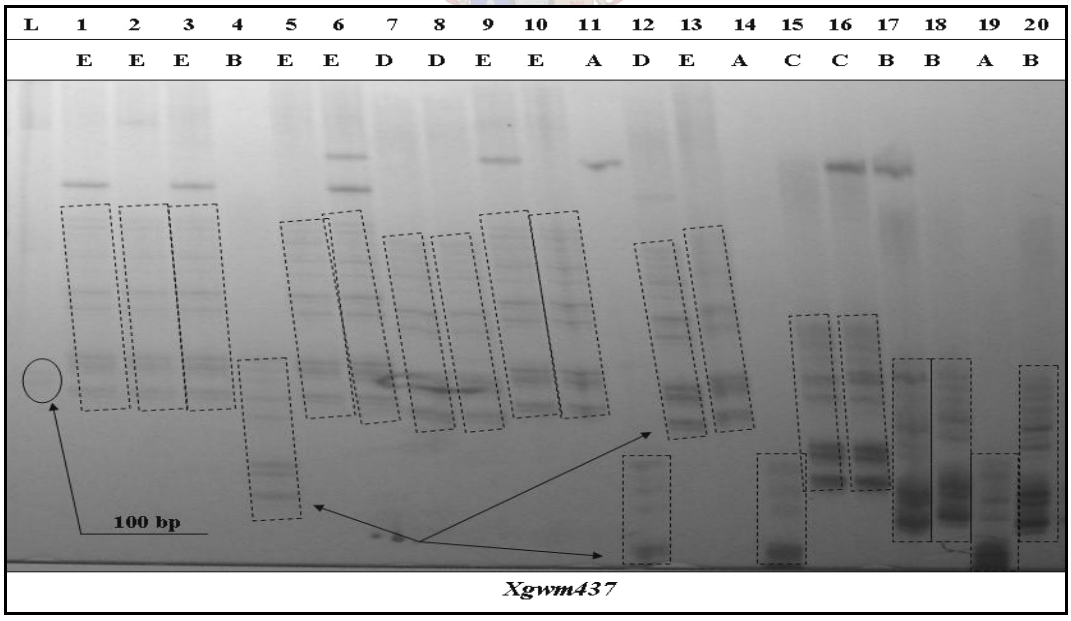


Figure 4-2 Gel photo illustrating the amplification of the *Xgwm437* microsatellite (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The sample numbers and alleles are given above each lane. The specific band targeted for automated detection is arrowed.

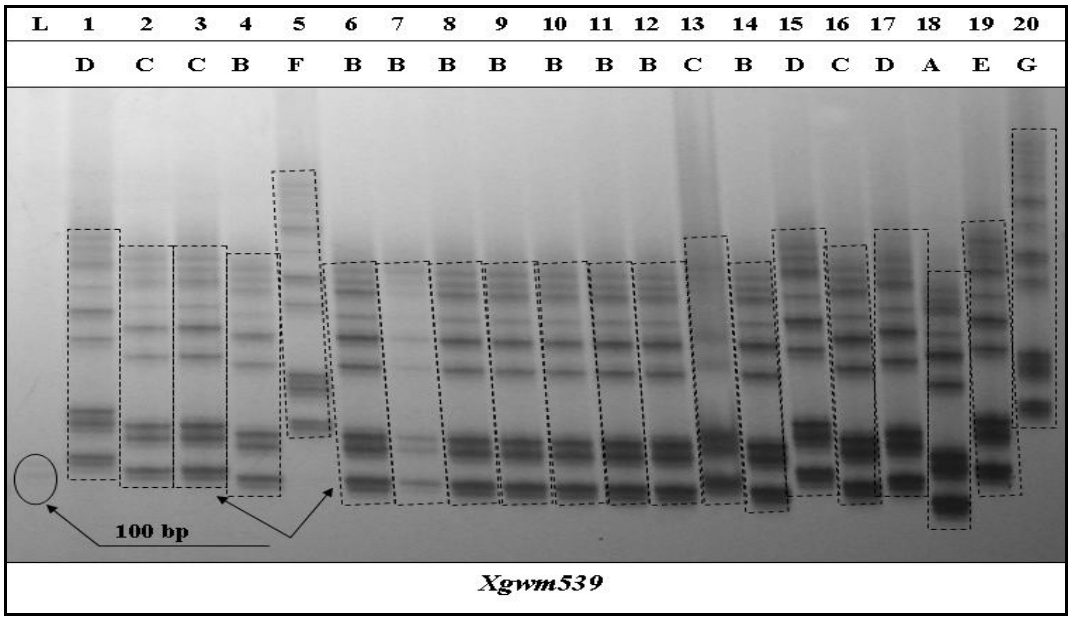


Figure 4-3 Gel photo illustrating the amplification of the *Xgwm539* microsatellite (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The sample numbers and alleles are given above each lane. The specific band targeted for automated detection is arrowed.

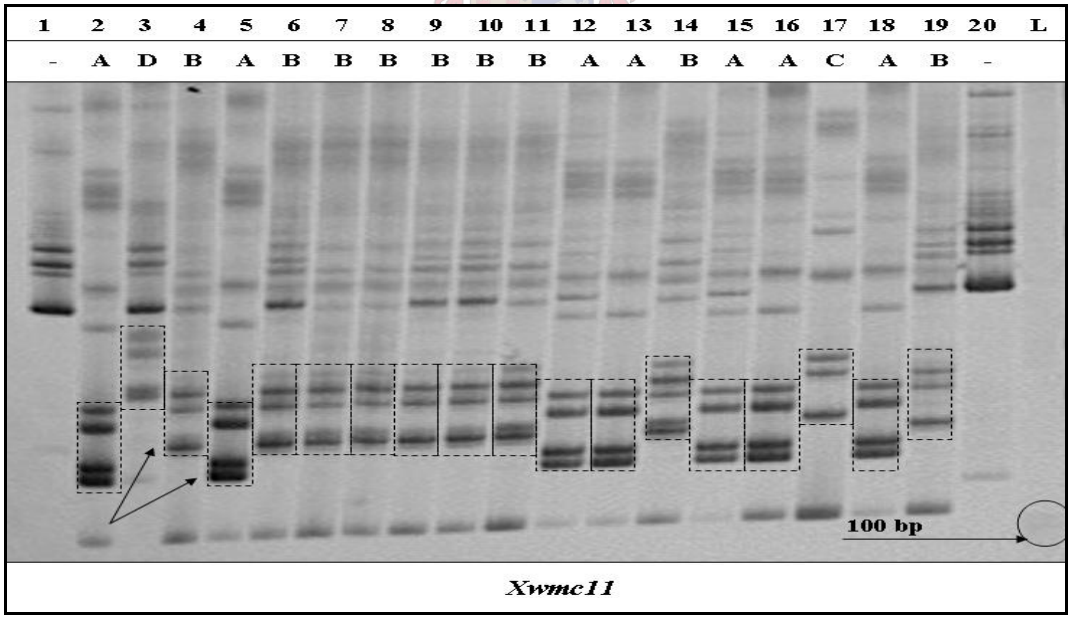


Figure 4-4 Gel photo illustrating the amplification of the *Xwmc11* microsatellite (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The sample numbers and alleles are given above each lane. The specific band targeted for automated detection is arrowed.

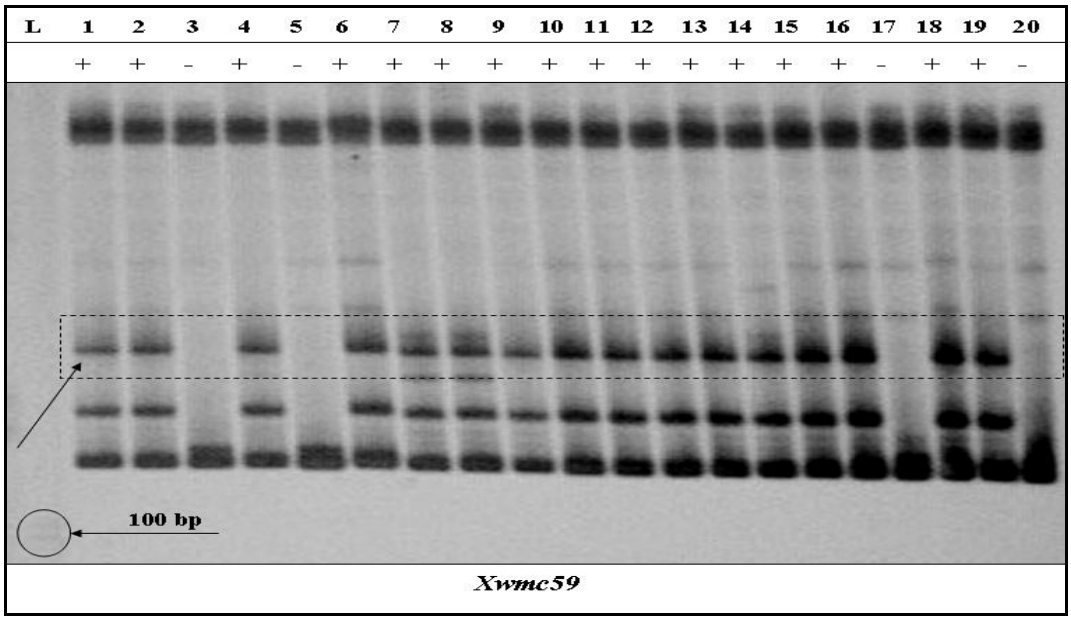


Figure 4-5 Gel photo illustrating the amplification of the *Xwmc59* microsatellite (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The sample numbers and alleles are given above each lane. The specific band targeted for automated detection is arrowed.

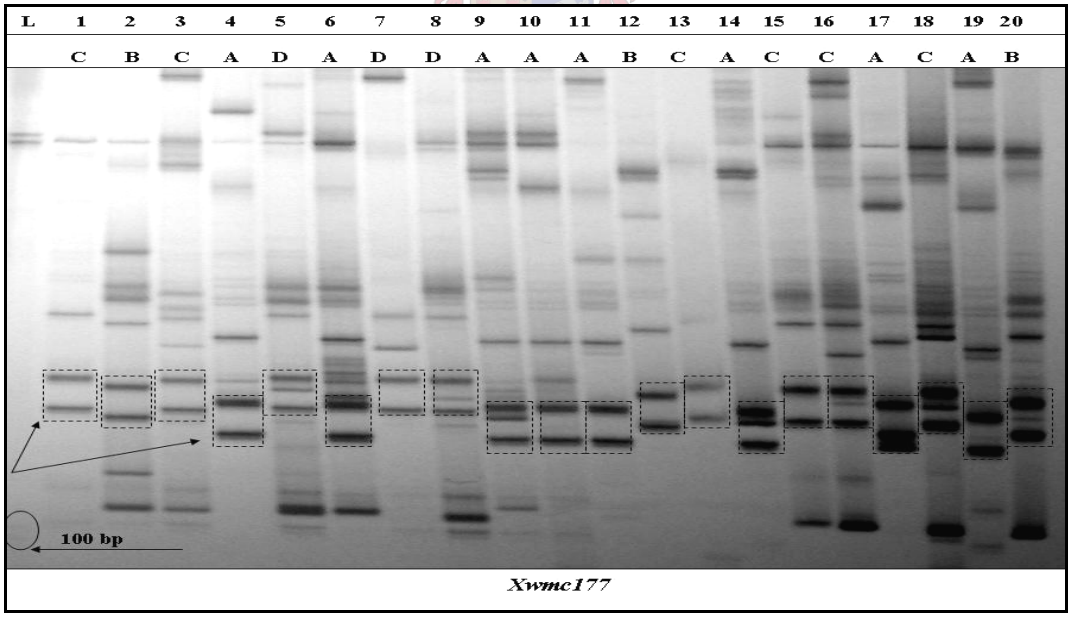


Figure 4-6 Gel photo illustrating the amplification of the *Xwmc177* microsatellite (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The sample numbers and alleles are given above each lane. The specific band targeted for automated detection is arrowed.

4.1.1.2. Fluorescence-based semi-automated analysis

The six microsatellite primer pairs chosen to be fluorescently labelled (Table 3-3) were electrophoresed on an Applied Biosystems 3130XL Genetic Analyzer and the results were analysed using GeneMapper. The chromatograms are given in Figures 7-35 to 7-97 in Addendum A and the selected peaks are identified by the name blocks under each one. These peaks correspond to the targeted bands identified with an arrow on the acrylamide gel photos.

The respective graphs were analyzed and the alleles of each sample and marker were recorded. The size range of the alleles per marker is given in Table 4-2. The alleles present in each of the twenty samples were recorded and named according to their size in base pairs and these sizes are given in Table 4-3. These results are more precise than those obtained from a denaturing polyacrylamide gel.

Using fluorescently labelled primers to amplify microsatellites is easier and quicker than denaturing polyacrylamide gels. The quality of the results obtained from the automated DNA sequencer is much better. Analysing the results can also be automated to ensure that numerous samples can be tested very quickly. The cost of primers and analysis is worth the amount of time, effort and reagents saved to obtain highly accurate and reproducible results. A more complete cost comparison is given in Table 4-6.

Table 4-2 Alleles for each fluorescently labelled microsatellite marker.

Marker	Allele Size Range (bp)
Xgwm190 (yellow)	205 – 213
Xgwm437 (yellow)	89 – 139
Xgwm539 (red)	132 – 142
Xwmc11 (blue)	161 – 181
Xwmc59 (red)	159
Xwmc177 (green)	180 – 194

As shown in Table 4-2, the *Xgwm190* and the *Xgwm437* primer pairs are both labelled with NED, the yellow dye and *Xwmc59* and *Xgwm539* are both labelled with PET, the red dye. These microsatellites can, however, be electrophoresed together because of the differences in the sizes of their amplified fragments. This means that all six microsatellites in Table 4-2

could be electrophoresed together. They have different colour dyes and/or differences in the size of their amplified products. This also helped to reduce the cost of this technique and increase the amount of information obtained from each sample.

The alleles for each microsatellite are given in Table 4-3. These values correspond to the size of the amplified fragment. To simplify interpretation, it was assumed that sequencer readings were sufficiently accurate to distinguish alleles with an accuracy of one base pair. The two plants for each released variety/advanced line are given together to illustrate the possible differences between them. Most of the genotypes of these two plants are the same, which illustrates the uniformity of these lines. The differences that are observed between the two plants illustrate the power of these microsatellites to not only distinguish between different lines, but also between plants of the same line.

Two different types of microsatellites were observed in this study. The first type works on an amplification versus no amplification (null allele) principle and therefore only has two alleles. This was observed with *Xwmc59* and the second *Xwmc11* amplification product (243 bp). The other type of microsatellite that was observed has more than two alleles and these are observed as a difference in the size of the amplified fragment. This can be seen in *Xgwm190*, *Xgwm437*, *Xgwm539*, *Xwmc11* (the first product) and *Xwmc177* (both products). These microsatellites can have any number of alleles. Differences in the size of each allele can be as little as two base pairs for a dinucleotide repeat microsatellite, which is the case for the six used here.

Table 4-3 Microsatellite results from the Genetic Analyzer for all 39 samples (Table 3-1). Numbers indicate the size of alleles in base pairs.

Samples (Table 3-1)	**Markers**						**Breeding Lines**
	Xgwm 190	***Xgwm 437***	***Xgwm 539***	***Xwmc 11***	***Xwmc 59***	***Xwmc 177***	
1	213	134	140	175	159	192	**US**
21	213	-	-	173	159	192	
2	213	134	136	163	159	188	
22	213	132	136	163	159	192	
3	213	139	138	161	-	192	
23	205	134	136	163	159	180	

Samples (Table 3-1)	Markers						Breeding Lines
	Xgwm 190	*Xgwm 437*	*Xgwm 539*	*Xwmc 11*	*Xwmc 59*	*Xwmc 177*	
4	205	104	134	173	159	180	**US**
24	205	104	-	173	159	180	
5	213	134	134	163	-	194	
25	213	134	132	163	159	194	
6	205	134	134	173	159	180	**PANNAR**
26	205	132	132	173	159	180	
7	205	130	134	173	159	194	
27	205	130	134	173	159	194	
8	205	130	134	173	159	194	
28	205	130	134	173	159	194	
9	205	134	134	173	159	180	**SGI**
29	205	134	134	173	159	180	
10	205	134	134	173	159	180	
30	205	134	134	163	159	180	
11	205	89	134	173	159	180	**MONSANTO**
31	205	89	134	173	159	180	
12	205	130	134	163	159	188	
32	205	130	134	163	159	188	
13	205	134	136	163	159	192	
33	205	134	136	163	159	192	
14	205	89	134	173	159	180	
34	-	89	134	173	159	180	
15	205	114	140	163	159	192	
35	205	114	140	163	159	192	
16	205	114	136	163	159	192	
36	205	130	134	163	159	188	
17	213	-	-	181	-	186	**CYMMIT**
37	213	104	138	181	-	186	
18	205	-	132	163	159	192	
38	-	106	132	-	159	180	
19	205	-	142	173	159	180	
39	-	89	142	173	159	180	
20	209	104	-	175	-	188	

It is clear that these six microsatellites are very effective for genotype discrimination. These six microsatellites are capable of distinguishing between the twenty different genotypes as

well as identifying some differences between the two plants of the genotypes (Table 4-3). The alleles of each microsatellite are thus not consistent between some of the plants of the same breeding line. The observed variation among plants within breeding lines means that it will be necessary to involve several plants during sampling so as to obtain a more representative profile of a particular line. The presence of intra-line variability is reflective of the fact that wheat breeding lines may contain considerable genetic variation while being relatively uniform for obvious phenotypic characteristics such as days to maturity, plant height, ear type, etc. During inbreeding and selection a limited number of single plant selection steps (often 3 – 4) are normally introduced thus preserving considerable genetic variation between derived lines.

4.1.1.3 Microsatellite Xwmc177

This microsatellite was one of the four recommended by Prasad *et al*. (2000). These authors found that this one microsatellite marker could discriminate between sixteen of the genotypes that were tested. The discriminatory power of this microsatellite could be increased by using it in combination with one, two or three other microsatellite markers. This made the microsatellite an understandable choice for this study.

From Figure 4-6 it is clear that this microsatellite on its own can distinguish among all of the twenty cultivars/ breeding lines. Apart from the indicated area there is little evidence of the predominantly structured patterns seen with the other microsatellites. *Xwmc177* could therefore be extremely useful when used singly, either with polyacrylamide or automated detection systems. It can also be used in a two-phase typing system where material is first typed with this microsatellite alone, analyzed with GeneMapper and only those cases that remain unresolved can then be typed with a further set of microsatellites or AFLP markers.

4.1.1.3. Marker detection method comparison

Six microsatellites were visualised using denaturing polyacrylamide gels and the Genetic Analyzer. Figure 4-7 shows the amplification product of five samples for microsatellite *Xwmc11*. The denaturing polyacrylamide gel was turned on its side with the bottom of the gel now on the left. This means that the smallest PCR product is on the far left. The arrows help

to indicate this band. This gel is now in the same orientation as the chromatogram of this fluorescently labelled microsatellite primer. The chromatogram is given in Figure 4-8.

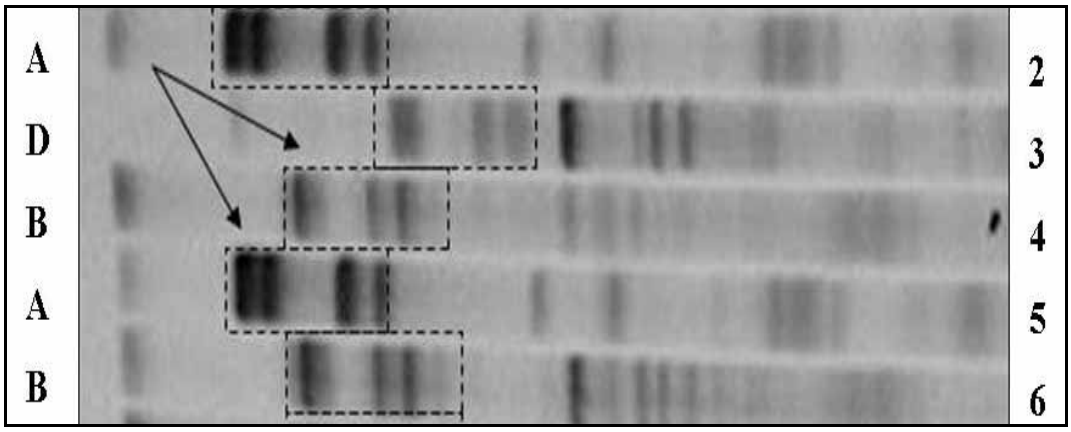


Figure 4-7 Microsatellite *Xwmc11* visualised on a denaturing polyacrylamide gel. The alleles are indicated by the blocks and arrows. Sample numbers (Table 3-1) are indicated on the right and allele names on the left of each sample.

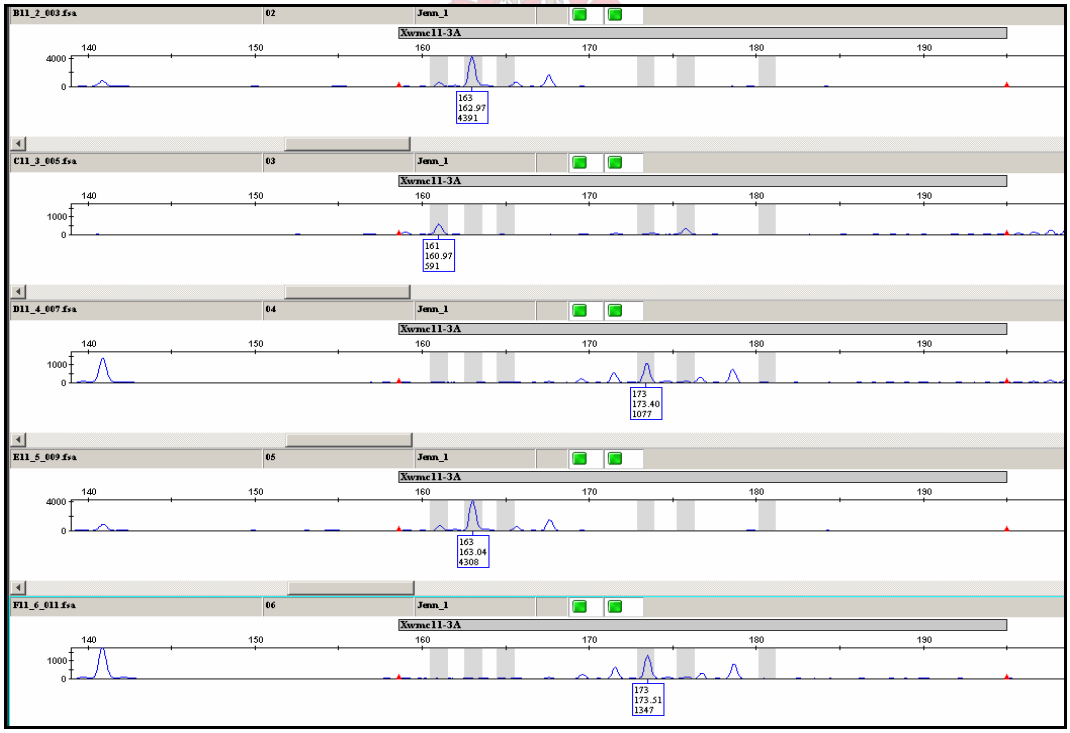


Figure 4-8 Microsatellite *Xwmc11* visualised on the Genetic Analyzer. The alleles are indicated by the blocks under the peaks. These are samples 2, 3, 4, 5, and 6 (Table 3-1).

Comparing these two photos (Figure 4-7 and 4-8) strengthens the fact that these hexaploid wheat samples are all homozygous. The additional peaks seen in Figure 4-8 are also seen in Figure 4-7 and should therefore not be included in the scoring of this microsatellite. The five samples in Figure 4-7 and 4-8 are examples and representative of the banding pattern that is seen in the other samples. The rest of the samples can be seen in Figure 4-4 for the denaturing polyacrylamide gel and in Figures 7-39 to 7-77 for the chromatograms (Addendum A). This banding pattern led the study to choose and score only the smallest amplification product in the pattern and ignore the rest. The samples used are commercially utilized varieties of a self-fertilizing, hexaploid crop, so heterozygotes are not expected and not included in the final results.

4.1.2. Amplified Fragment Length Polymorphisms (AFLPs)

Ten samples were used in the initial optimisation of this technique. This included two from each of the five breeding programmes (Table 3-1). These were used for each step of the AFLP technique. Individual selective PCR reactions were performed to ensure that the previous steps in the technique were functioning correctly. Thereafter, the *Eco*RI selective primers (with different colour labels) were used in one reaction with each *Mse*I selective primer.

The twenty different released varieties/advanced lines were tested following the steps described in section 3.4. All nine selective *Eco*RI and *Mse*I primer combinations were evaluated on the panel of released varieties/advanced lines.

4.1.2.1. Restriction digestion of genomic DNA and ligation of adaptors

Each DNA sample was digested using *Eco*RI and *Mse*I restriction enzymes. The enzyme digestion and ligation of the adaptors was performed in the same reaction. This was checked on a 1.5% agarose gel (Figure 4-9).

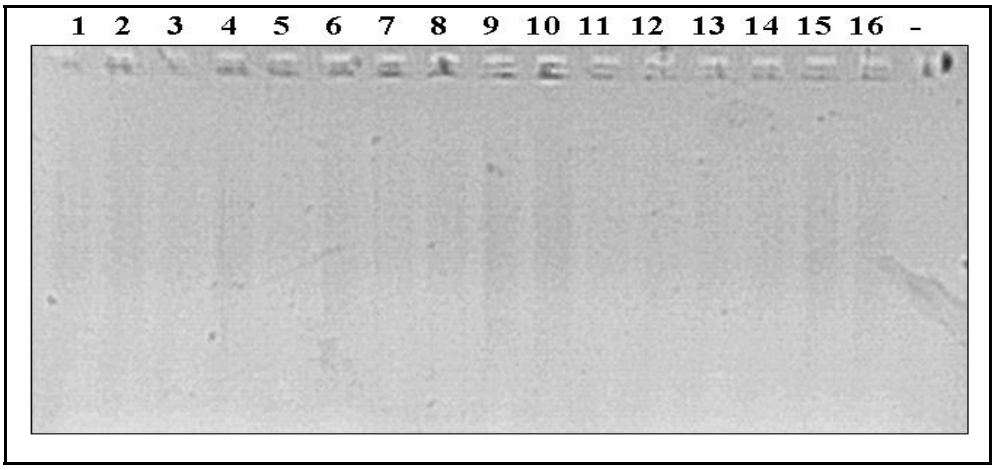


Figure 4-9 Gel photo showing the restriction digest and adaptor ligation of sixteen samples (1.5% agarose gel electrophoresed at 100 V for 30 minutes). Lane 17: negative water control.

All that can be seen in the gel photo in Figure 4-9 are faint smears. This is consistent with what is expected from this type of reaction. The entire genome is digested with a rare and a frequent cutting restriction enzyme. This produces a large amount of fragments. Only one quarter of the reaction volume was electrophoresed in the gel photo in Figure 4-9. The rest of the reaction was diluted ten times and then preselectively amplified using primers that will recognise and anneal to the adaptors.

4.1.2.2. Amplification with preselective primers

Each restriction-ligation product was amplified using preselective primers. These primers recognise the adaptor sequences and anneal to them. Each primer has one additional nucleotide that binds to one nucleotide of the original DNA sequence that is next to the adaptor. This reduces the amount of amplified fragments, so they can start to separate on an agarose gel. This can be seen in Figure 4-10.

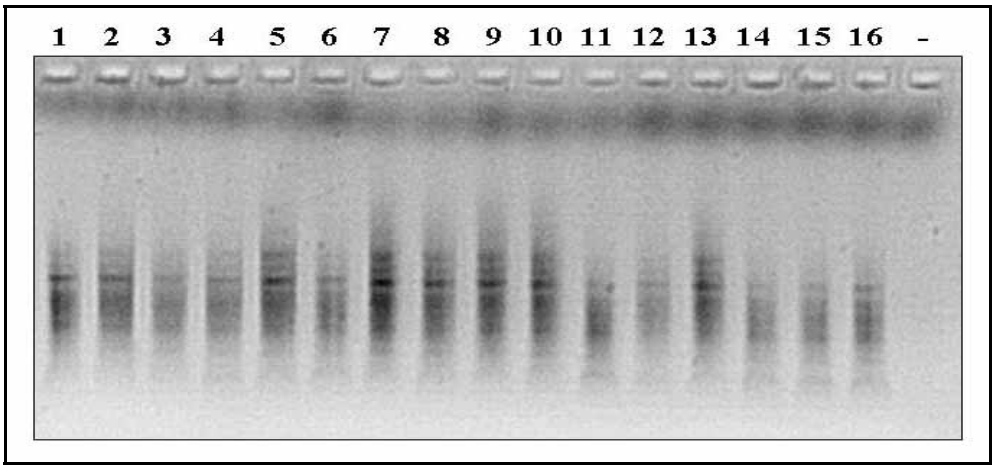


Figure 4-10 Gel photo showing the preselective amplification product of sixteen samples (1.5% agarose gel electrophoresed at 100 V for 30 minutes). Lane 17: negative water control.

4.1.2.3. Amplification of restriction fragments with selective primers

The amplification product from the preselective PCR was diluted ten times. This was then used as the template DNA to amplify fragments using the selective primers. These primers included the three labelled selective *Eco*RI primers and one unlabelled selective *Mse*I primer. The three *Eco*RI primers could be used in the same reaction because they are labelled with different dyes. This maximises the amount of information obtained from each sample and reduces the cost of electrophoresis.

4.1.2.4. Detection of amplified products and data analysis

The picture given in Figure 4-11 illustrates the results obtained from the automated DNA sequencer and analysed using GeneMapper software. The picture shows the entire range of amplified fragments. Each peak indicates a fragment with a different number of nucleotides. This study focused on fragments between 100 and 500 bp in length. Many limitations can be defined in the GeneMapper software to simplify the results. These include specifying the focus area of the alleles and selecting fragments according to their peak intensity.

The grey lines in Figure 4-11 represent the allele bins that are generated by the computer according to specifications/limitations selected using the software. Some of these include eliminating alleles that occur in every sample, peaks with an intensity less than 50 units, only alleles between 100 and 500 bp in size, etc.

Each of the three colours is a different primer pair combination (see Table 3-4), which means that the different colours cannot be analysed together. This explains the differences observed in Figure 4-11. The primer combination labelled with FAM (blue) has a lot more alleles than the NED labelled primer combination. The yellow fluorescent label (NED) appears black in GeneMapper. This is because black is easier to see on a computer screen than yellow.

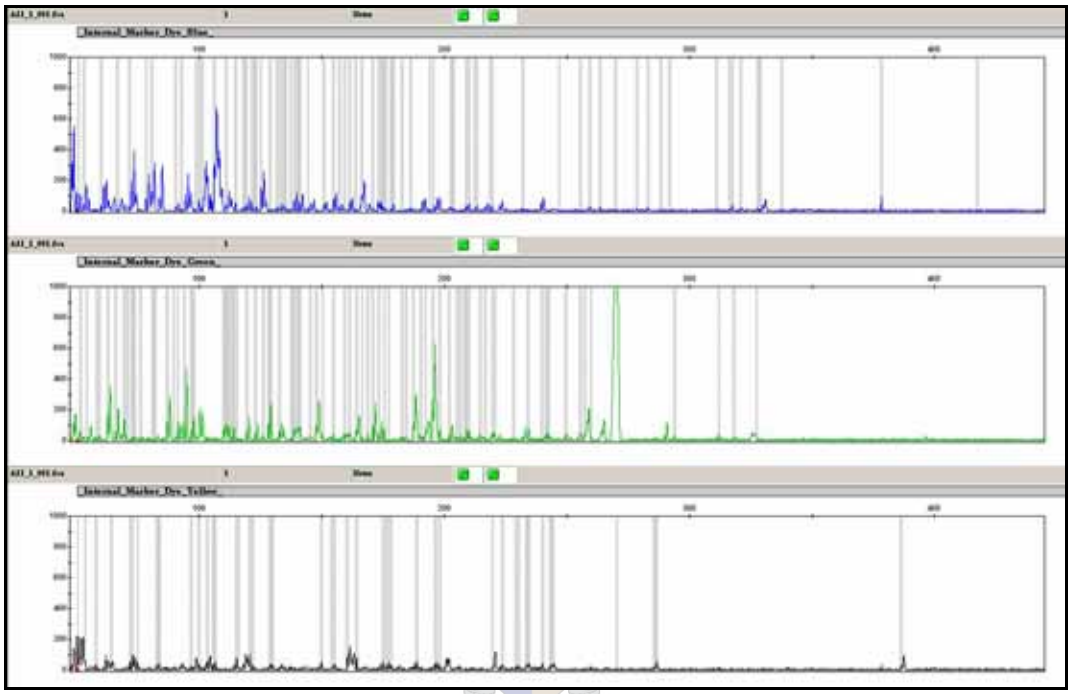


Figure 4-11 Photo illustrating the peaks generated by GeneMapper using the data obtained from the genetic analyzer. This is the same sample with each of the three *Eco*RI primers combined with the *Mse*I-CTC primer. These peaks are set at an intensity of 1000. The primer combination labelled with FAM (blue) is at the top. The JOE (green) labelled primer is in the middle and the NED (yellow, appears black on screen) labelled primer is at the bottom.

This technique detects amplified fragments that have one labelled and one unlabelled end (*Eco*RI/*Mse*I). A completely unlabelled fragment (*Mse*I/*Mse*I) can not be detected. Thus the number of fragments visualised is less than those amplified, which makes analysis much easier. A fragment that is labelled on both sides (*Eco*RI/*Eco*RI) is observed during this technique, but there are very few of these fragments. This is because the *Eco*RI restriction enzyme is a rare cutting enzyme. *Mse*I is the frequent cutter which produces a lot of *Mse*I/*Mse*I fragments. This explains why the *Mse*I selective primer has to be unlabelled.

Each primer combination produced a different number of polymorphic alleles. These are based on a present (1) and absent (0) principle. This means that every allele indicates that an amplified product was observed in at least one of the samples. All alleles that are common to every sample were eliminated from the analysis. This reduces and simplifies the amount of information obtained for the specific primer combination.

The analysis of AFLPs is not as easy as microsatellites because there are more bands to look at and compare. The large number of bands does, however, decrease the chance that two samples will have the same banding pattern. The level of polymorphism is higher than with microsatellites but it is less specific. The large number of possible primer combinations also helps to make this technique very informative.

Nine different primer combinations were used in this technique. The three *Eco*RI selective primers were used in combination with each of the three *Mse*I selective primers. The number of polymorphic alleles observed for each combination is given in Table 4-4. This table also provides information of the most polymorphic individual primer which is *Eco*RI-AGG. The most polymorphic primer combination is *Eco*RI-AGG and *Mse*I-CTC.

Table 4-4 Total number of polymorphisms/alleles for each AFLP selective primer pair combination.

Primer Combination	***Eco*RI-AAC**	***Eco*RI-ACA**	***Eco*RI-AGG**	**Total**
***Mse*I-CAG**	16	16	41	73
***Mse*I-CAT**	32	26	38	96
***Mse*I-CTC**	22	32	62	116
Total	70	74	141	285

The *Mse*I-CTC primer was used in combination with all three *Eco*RI selective primers for the implementation section (4.2) of this study. These are the most polymorphic primer combinations and thus have a greater chance of showing differences between closely related breeding lines such as those used in section 4.2. This is the only reason for their continued use.

The next best primer combinations to use would be the *Mse*I-CAT and *Eco*RI selective primers. The *Mse*I-CAG primer did not provide sufficient numbers of polymorphic bands to be used further in this study. Each primer combination did however seem to have a sufficient level of polymorphism to discriminate between the initial twenty cultivars/ breeding lines.

The exact level of discrimination will be discussed further in section 4.1.5 with the rest of the data analysis information.

4.1.3. Retrotransposons

Kalendar *et al*. (1999) developed various retrotransposon-based techniques for barley using the *BARE*-1 retrotransposon as a target sequence. This retrotransposon has a similar sequence to the *Wis-2* retrotransposon that is found in the wheat genome (Queen *et al*. 2004). Therefore the techniques that were able to distinguish between a number of barley varieties were investigated for their ability to also discriminate between wheat varieties. The same primer sequences were used in the initial testing of these techniques.

The four different retrotransposon-based techniques were first tested using five genotypes, one from each breeding program (Table 3-1). This was done to optimise the reaction conditions before testing the panel of twenty released varieties/advanced lines. Only the Sequence-Specific Amplified Polymorphism (SSAP) technique and the reaction based on the *Wis-2* retrotransposon in wheat were eventually used to test the entire panel.

4.1.3.1. Inter-Retrotransposon Amplified Polymorphism (IRAP) analysis

This technique involves the amplification of a fragment between two LTRs (long terminal repeats), which are found on either side of retrotransposons. The primer sequences were based on the *BARE*-1 (barley retrotransposon) sequence. The reaction conditions given by Queen *et al*. (2004) were attempted but no amplification was observed. The conditions were altered and optimisation was attempted but as can be seen in Figure 4-12, the reaction was not successful.

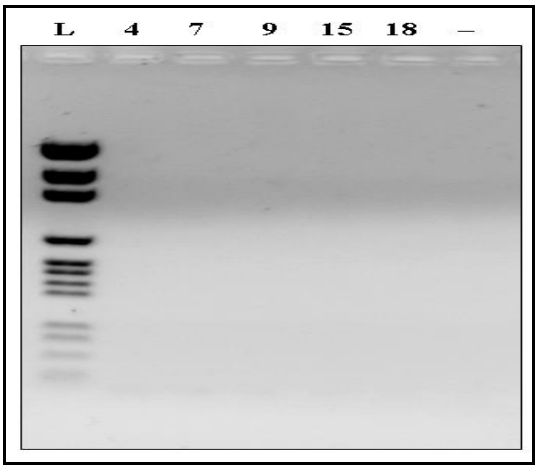


Figure 4-12 Gel photo illustrating the lack of amplification using the IRAP primers (1.5% agarose gel electrophoresed at 100 V for 30 minutes). The L is pGEM ladder. Numbers above each lane correspond to the samples in Table 3-1.

A possible reason for the lack of fragment amplification is that the primer sequences include a cloning sequence (given in italics in Table 3-5). This was done in barley to facilitate the production of the primers. An attempt was therefore made to locate the primer sequences (Table 3-5) in the wheat retrotransposon (*Wis-2*). The entire primer sequence in Table 3-5 was used to find a complimentary fragment in the wheat retrotransposon as well as the sequence excluding the italics section. Neither sequence showed a positive alignment which would imply that the primers developed for barley are unlikely to generate IRAP fragments in wheat. The IRAP technique was therefore adapted in this study to produce the *Wis-2* Retrotransposon Amplification technique (described in section 3.5.4, also see 4.1.3.4 below) which is based on similar principles to the IRAP technique. The primers bind to the ends of the *Wis-2* retrotransposon and amplify it to produce various fragments.

4.1.3.2. Retrotransposon-Microsatellite Amplified Polymorphism (REMAP) analysis

The fragment produced in this technique is found between an LTR and a simple repeat sequence. Thus one primer targets a part of the LTR and the other primer targets and consists of a simple repeat sequence. Seven different repeat sequence primers were used in combination with one LTR primer. The LTR primer sequence was also based on the *BARE*-1 retrotransposon found in barley. This technique did however produce an amplification product in wheat which can be seen in Figure 4-13. The seven primer combinations, numbered 2 to 8, are illustrated in Figure 4-13. This photo shows amplification in just two

samples per primer combination. Each combination was also tested on the panel of twenty released varieties/advanced lines (Table 3-1). These photos are given in Figures 7-98 to 7-104 in Addendum A.

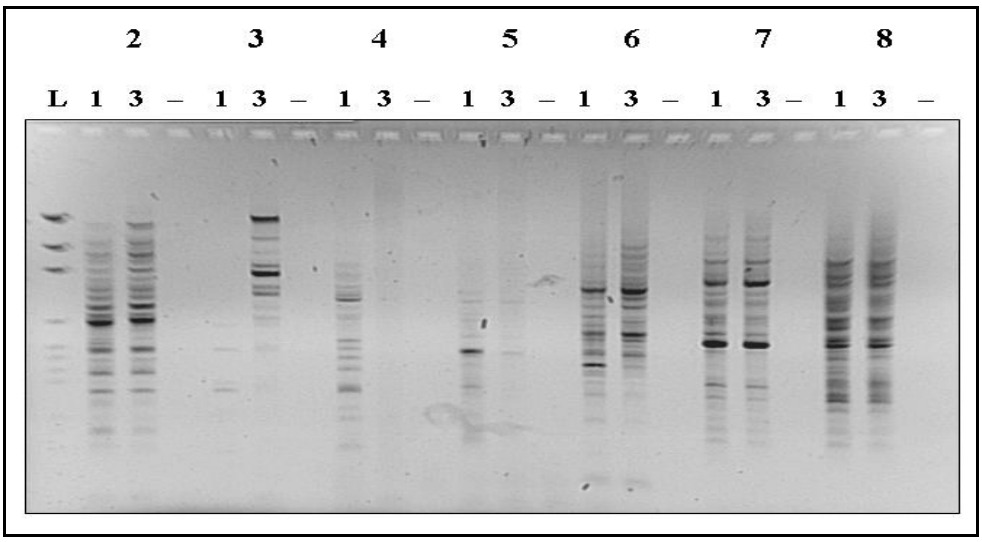


Figure 4-13 Gel photo illustrating the seven different REMAP primer combinations corresponding to Table 3-6 (1.5% agarose gel electrophoresed at 100 V for 45 minutes). Lane L: pGEM ladder. Samples 1 and 3 and a negative control (water) were used for each combination. The primer combination is given above each group of three samples.

The problem with this technique, however, is that the results were not reproducible. This can be seen when the photos in Figures 7-98 to 7-104 in Addendum A are compared with the photo in Figure 4-13. The results of the same samples using the same primer combinations do not correspond. The PCR reactions were optimised to produce the banding patterns seen above. These conditions were then used to test the panel of released varieties/advanced lines and different banding patterns were observed for the same samples. Therefore this technique was not used to test the 119 breeding lines in section 4.2.

4.1.3.3. Sequence-Specific Amplified Polymorphism (SSAP) analysis

The SSAP technique was developed by Waugh *et al*. (1997). The protocol for the technique was identical to that of the AFLP technique developed by Vos *et al*. (1995) except for one selective primer. The traditional *Mse*I or *Pst*I or *Eco*RI selective primers were used in combination with a *Wis-2* retrotransposon-based primer.

Two types of the *Wis-2* primer were used. These have identical sequences but the one was labelled and the other was not, as shown in Table 3-7. This was done so that the labelled primer can be used with the unlabelled *Mse*I selective primers and the unlabelled primer could be used with the labelled *Eco*RI selective primers. Thus, the *Eco*RI primers could again be multiplexed because each has a different coloured label.

Therefore two different reactions had to be optimised for this technique. First the individual selective PCR had to be optimised, which included one *Mse*I selective primer and the labelled *Wis-2* primer. Two samples were initially tested with all six *Mse*I and *Wis-2* primer combinations. The reaction conditions were based on those given by Waugh *et al*. (1997). The other reaction was the multiplex reaction including the three different *Eco*RI selective primers and the unlabelled *Wis-2* primer. The conditions of the multiplex reaction were similar to those for the selective amplification reaction in the AFLP technique. Four samples were tested in the initial reaction of the multiplexed selective amplification reaction. The final optimised reaction conditions for both PCR reactions are given in section 3.5.3.

The photo in Figure 4-14 illustrates the three *Eco*RI primers amplified in the same reaction with the *Wis-2* primer. Each colour represents a different *Eco*RI primer. The colours are given in Table 3-7. Figure 4-14 shows fragments of between 100 and 500 bp in size. It is clear from the figure that the *Eco*RI primers do not produce a large number of fragments when in combination with the *Wis-2* primer. This could be explained by the fact that *Eco*RI is a restriction enzyme that recognises and cuts the genome in relatively few places.

The *Mse*I and *Wis-2* combinations, on the other hand, each produced a rather large number of alleles at very high intensities. This can be seen in the photo in Figure 4-15. This illustrates the polymorphism of these combinations. *Mse*I is a restriction enzyme that cuts the genome more frequently than the *Eco*RI restriction enzyme. This explains the higher number of alleles produced from the *Mse*I combinations compared to the *Eco*RI primers. The size of the fragments in Figure 4-15 are also between 100 and 500 bp in length.

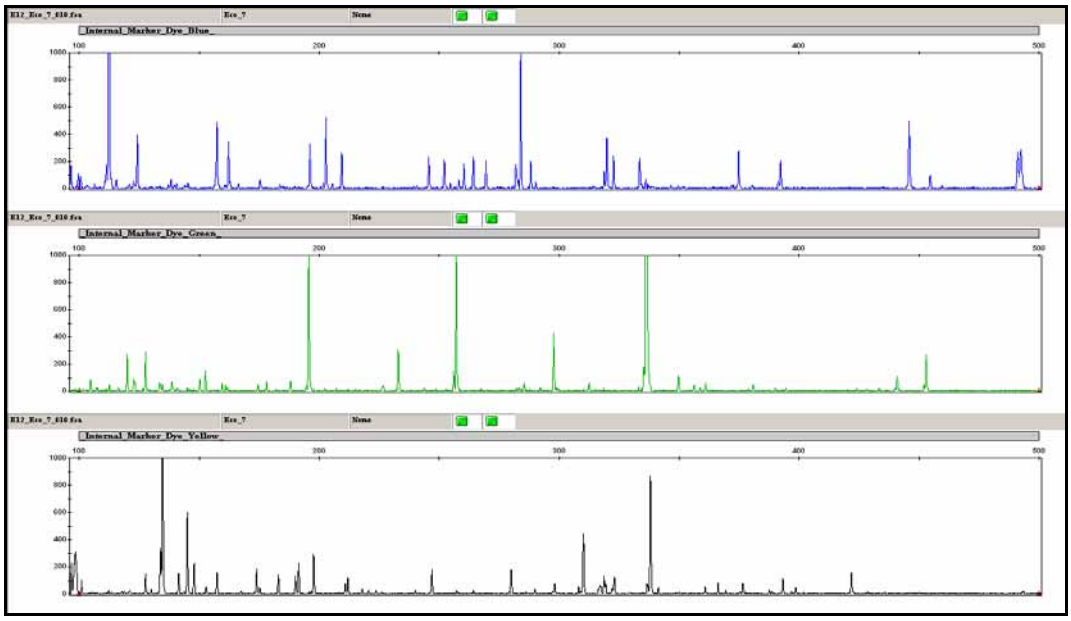


Figure 4-14 Photo illustrating the peaks generated by GeneMapper using the data obtained from the genetic analyzer. This is the same sample with each of the three *Eco*RI primers combined with *Wis-2* U primer. These peaks are set at an intensity of 1000.

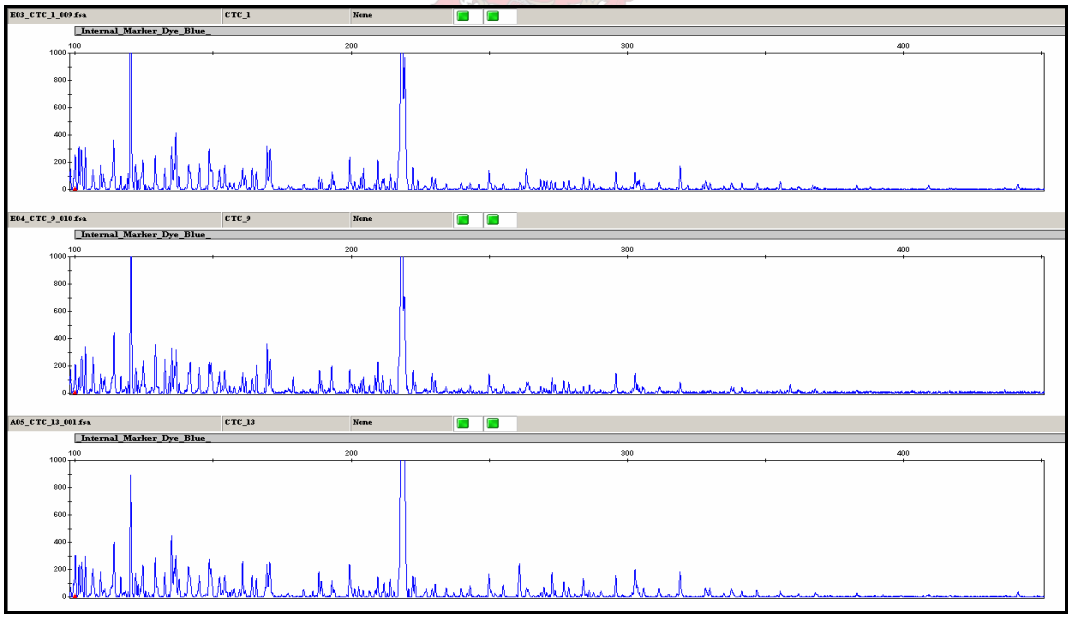


Figure 4-15 Photo illustrating the peaks generated by GeneMapper using the data obtained from the genetic analyzer. This picture shows three samples amplified with the same *Mse*I-CTC primer combined with the *Wis-2* U primer. These peaks are set at an intensity of 1000.

The SSAP data were analysed in exactly the same way as the AFLP data. A red fluorescently labelled size standard is loaded and electrophoresed with every sample. The same size

standard is used in the AFLP technique and has a size range of 75 to 500 bp. Fragments between 100 and 500 bp were focused on in this study. These are more reliable and reproducible than the smaller and larger fragments and neither technique produced many fragments between 400 and 500 bp.

As with AFLP, each primer combination produced a different number of polymorphic alleles. These are based on a present (1) and absent (0) principle. This means that every allele indicates that an amplified product was observed in one of the samples. All alleles that are common to every sample were eliminated from the analysis. The large number of bands produced in this technique decreases the chance that two samples will have the same banding pattern.

Table 4-5 shows the number of alleles observed for each of the primer combinations. The large number of alleles emphasises how frequent the *Mse*I enzyme cuts the wheat genome. Each *Mse*I primer has to be used in a separate PCR reaction with the *Wis-2* primer because the *Mse*I primers are not labelled. This makes these reactions more time and reagent consuming than the *Eco*RI reactions and the AFLP combinations. These can both be multiplexed which helps to reduce the time it takes to prepare each PCR sample and the time that is needed to complete each PCR reaction.

Table 4-5 Total number of polymorphisms/alleles for each combination of SSAP selective primer pairs.

Primer Combination			
	***Wis*-2 L**		***Wis*-2 U**
***Mse*I-CAG**	106	***Eco*RI-AAC**	26
***Mse*I-CAT**	143	***Eco*RI-ACA**	32
***Mse*I-CTA**	109	***Eco*RI-AGG**	24
***Mse*I-CTC**	75		
***Mse*I-CTG**	130		
***Mse*I-CTT**	137		
Total	700	**Total**	82

The level of polymorphism observed in the *Eco*RI reaction is sufficient to discriminate between the twenty samples in Table 3-1. The SSAP reactions produced a lot more alleles than the AFLP reactions. This can only be seen in the *Mse*I reactions because it cuts the

genome in more places than the *Eco*RI restriction enzyme. Therefore more fragments have a *Mse*I recognition site on one side and the targeted LTR sequence on the other side. This technique also detects and visualises less fragments than were amplified fragments.

4.1.3.4. Wis-2 Retrotransposon Amplification analysis

This technique was inspired by the other retrotransposon-based techniques (IRAP and REMAP techniques described in sections 3.5.1 and 3.5.2 and above). These were first described and tested in barley. Therefore the primer sequences were designed according to sequences found in barley, specifically the *BARE*-1 retrotransposon. This new technique uses the *Wis-2* retrotransposon that is found in the wheat genome. The primers recognise and bind to the ends of the retrotransposon sequence. Thus the complete retrotransposon is amplified using these primers.

The twenty released varieties/advanced lines given in Table 3-1 were tested with the *Wis-2* primers. The result of this PCR can be seen in the photo in Figure 4-16. The primers produced a large number of bands. The more prominent bands are seen in every sample whereas the lesser bands were difficult to score and reproduce. In view of the low level of polymorphism among genotypes this technique was not evaluated further.

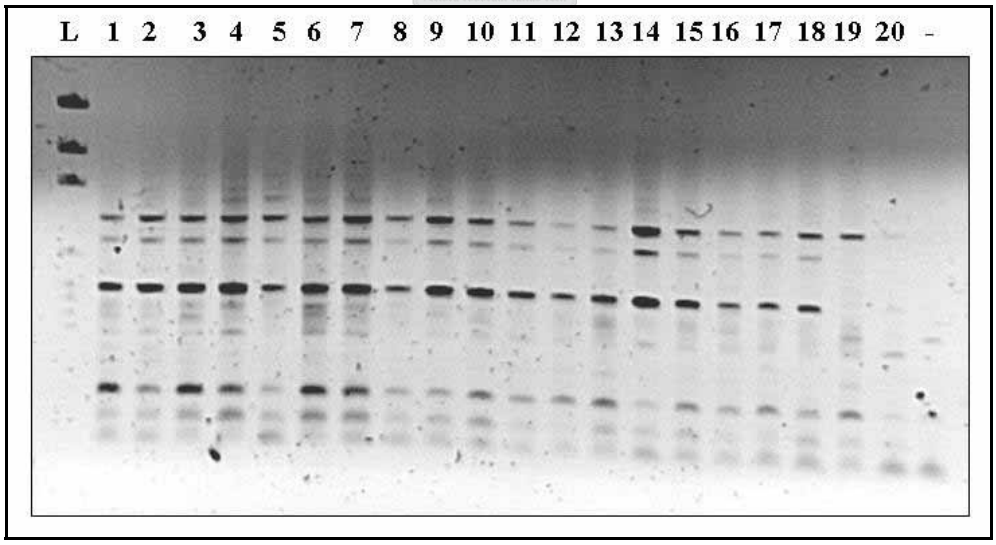


Figure 4-16 Gel photo illustrating the amplification produced by the *Wis-2* primers (1% agarose gel electrophoresed at 100 V for 45 minutes). The L is pGEM ladder. Lane 1 – 20: numbers correspond to the samples in Table 3-1. Lane 21: negative water control.

4.1.4. Marker comparison

The markers that were chosen to screen the panel of released varieties/advanced lines were microsatellites, Amplified Fragment Length Polymorphisms (AFLPs) and Sequence-Specific Amplified Polymorphisms (SSAPs). The three sets of markers have been compared in Table 4-6. It is clear that the easiest system to use is microsatellites, but it does have the smallest number of polymorphic loci. SSAP has the highest number of polymorphic loci but it can be rather cumbersome for analysis. The number of loci produced by SSAP reactions were, for some primer pairs, more than double the number produced by AFLP primers. Thus AFLP could actually be the best choice for this study. The level of polymorphism observed was more manageable for analysis and yet still sufficient to discriminate between the genotypes.

Table 4-6 Comparison of the three marker systems used in this study.

CRITERIA	MICROSATELLITES	AFLP	SSAP
DNA required (ng)	100	300	
DNA quality	Good		
Number polymorphic loci analyzed	8	10-70	20-150
Ease of use	Easy	Moderate	
Manual/Automated detection	Both systems	Automated	
Reproducibility	High		
Time required	1 day	3 days	
Cost for analysis:	Moderate	Moderate to high	
(a) Extractions	Moderate		
(b) Preselective reactions	None	3 (gDNA to results)	
(c) Primers & PCR components	Moderate	High	
(d) PCR reactions	1 for results	2 for results	
(e) Automated visualisation	6 data sets per sample	3 data sets per sample	
Ease of analysis	Easy	Moderate	High

One aspect that is very different for the marker systems is the cost involved. The cost of reagents per reaction, the cost of visualisation and the cost of analysis for each marker was compared and given relative to one another in Table 4-6. Microsatellites are definitely the cheapest markers to use of the three investigated. All that is needed is basic PCR reagents including either fluorescently labelled primers for automated detection or unlabelled primers for PAGE and silver staining detection.

AFLP and SSAP require many reagents, such as very expensive restriction enzymes and fluorescently labelled primers. These techniques involve three steps which each require different reagents. Therefore the technique to use depends on the samples that have to be analysed. For a smaller group of unrelated genotypes microsatellites are the easiest and cheapest option, but for a large group of very closely related genotypes AFLP would be the marker of choice.

4.1.5. Data analysis

The results of the initial marker evaluation using the 39 samples in Table 3-1 were compared statistically. The data from the ABI 3130XL were used for each marker type. A discriminant analysis was performed using the SAS software. This analysis type allows for two types of tests. Firstly, data is tested from a sample set where the samples were already classified into known and definite groups. This provides each classification with a linear function. This function can then be used in the second type of test to reclassify the previous data and check the original classification of the samples, or data taken from a new collection of samples can be used to classify the new samples into the known groups using these functions.

A cluster analysis was also performed using the SAS software. Cluster analysis divides samples into groups that are the most similar according to the data. These groups are referred to as clusters. Cluster analysis produces dendrograms that represent the relationship of each sample to every other sample. Unfortunately these could not be included in this written piece as the size of the dendrogram and the amount of detail that appears on it made it illegible. Therefore no dendrograms for any of the cluster analyses performed were included in this report.

4.1.5.1. Microsatellites

Six microsatellites were used to evaluate the fluorescently labelled primers and automated DNA sequencer technique. The initial population of 39 plants was tested with these six microsatellites. The data obtained from the sequencer for these samples were used to perform a cluster analysis using SAS. The data used are given in Table 4-3.

In doing the cluster analysis, the 39 genotypes were split into various groups or clusters. These clusters are given in Table 4-7. The very small number of clusters is due to the limited amount of variables obtained with the six microsatellites. The ampersand combines samples with the same genotype. It is clear that these six microsatellites are consistent between the two samples with the same genotype. There is very little intra-line variability characterised with these microsatellites. The clusters in Table 4-7 also show that there is some similarity between the Pannar, Small Grain Institute and Monsanto genotypes. A complete list of the clusters is given in Table 4-8. The Tie refers to samples that are identical according to this system. The highlighted section indicates the 95% accuracy cut-off used to compile the information given in Table 4-7.

Table 4-7 Thirty nine samples in clusters at 95% accuracy for the microsatellite data.

Marker Type	**Microsatellites**	
Total Clusters	7	
Cluster Number	**Samples per cluster**	
1	2 & 22	11 & 31
	6 & 26	12 & 32
	7 & 27	13 & 33
	8 & 28	15 & 35
	9 & 29	16 & 36
	10 & 30	1, 4, 14, 23
2	5 & 25, 3, 37	
3	34, 39	
4	17, 21, 24	
5	18, 19	
6	20	
7	38	

Table 4-8 Cluster procedure results for the 39 initial genotypes.

MICROSATELLITES						
NCL	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**
38	7	27	2	0.0000	1.000	T
37	CL38	8	3	0.0000	1.000	T
36	CL37	28	4	0.0000	1.000	T
35	6	9	2	0.0000	1.000	T
34	CL35	29	3	0.0000	1.000	T
33	CL34	10	4	0.0000	1.000	T
32	11	31	2	0.0000	1.000	T
31	13	33	2	0.0000	1.000	T
30	CL32	14	3	0.0000	1.000	T
29	15	35	2	0.0000	1.000	T
28	32	36	2	0.0000	1.000	
27	5	25	2	0.0000	1.000	T
26	23	30	2	0.0000	1.000	
25	12	CL28	3	0.0000	1.000	
24	CL33	26	5	0.0000	1.000	
23	2	22	2	0.0000	1.000	
22	CL29	16	3	0.0000	1.000	
21	34	39	2	0.0001	1.000	
20	3	CL27	3	0.0001	1.000	
19	21	17	2	0.0001	1.000	
18	CL25	CL31	5	0.0001	0.999	
17	CL23	CL18	7	0.0003	0.999	
16	1	CL36	5	0.0003	0.999	
15	CL26	CL24	7	0.0004	0.999	
14	4	CL30	4	0.0004	0.998	
13	18	19	2	0.0004	0.998	
12	CL16	CL17	12	0.0010	0.997	
11	CL12	CL15	19	0.0016	0.995	
10	CL11	CL22	22	0.0027	0.992	
9	CL20	37	4	0.0027	0.990	
8	CL10	CL14	26	0.0128	0.977	
7	CL19	24	3	0.0189	0.958	
6	CL9	20	5	0.0392	0.919	
5	CL21	38	3	0.0520	0.867	
4	CL7	CL13	5	0.0620	0.805	
3	CL8	CL4	31	0.1852	0.620	
2	CL3	CL6	36	0.2858	0.334	
1	CL2	CL5	39	0.3338	0.000	

Seven clusters are not sufficient to discriminate between all twenty genotypes. Too few variables were used to be able to distinguish between 39 samples. The *Xwmc177* microsatellite, however, produced a multitude of bands that was similar to an AFLP banding pattern. Therefore this microsatellite was analysed as an AFLP.

4.1.5.2. Microsatellite Xwmc177

The multiple bands amplified by this microsatellite were analysed as many variables of the same marker, similar to the way in which the AFLP data were analysed. A discriminant analysis using the SAS software was performed for this data. The variables used for the analysis are given in Table 8-1 in Addendum B. All the other information obtained during the analysis is given in Addendum B. These include the reclassification of all 39 samples in Table 8-2. Figure 8-1 illustrates the distribution of the samples. The labels on the graph correspond to the cultivar names given in Table 3-1. The summary of the reclassifications is given below in Table 4-11. Number 1 to 20 corresponds to the name of the line given on the left of Table 4-9 and the number given in Table 3-1.

Table 4-9 Number of observations per line (Table 3-1) and percent reclassified into each line for microsatellite *Xwmc177-2A*.

From Line	Into Line																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	19	17	18	19	20	Total
00K180	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
00K268	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
97K1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2
	0	0	50	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	100
98K120	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2
	0	0	0	50	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	100
00K60	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
PAN3404	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
PAN3490	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	100
PAN3492	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	2
	0	0	0	0	0	0	50	50	0	0	0	0	0	0	0	0	0	0	0	0	100
Kariega	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	2
	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	100
Biedou	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	2
	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	50	0	0	0	0	100
SST015	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	2
	0	0	0	0	0	0	0	0	50	0	50	0	0	0	0	0	0	0	0	0	100
SST027	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	2
	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	100
SST57	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2
	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	100
SST65	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	2
	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	100
SST88	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	100

From Line	Into Line																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	19	17	18	19	20	Total
SST94	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	2
	0	0	0	0	0	0	0	0	0	0		0	0	0	0	100	0	0	0	0	100
12HRWYT4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	2
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	100
16HRWSN1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	100
16HRWSN8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	2
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	50	0	100
16HRWSN2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	100
Total	2	2	1	1	2	2	3	1	5	1	1	2	2	2	2	4	2	2	1	1	39
	5.13	5.13	2.56	2.56	5.13	5.13	7.69	2.56	12.82	2.56	2.56	5.13	5.13	5.13	5.13	10.26	5.13	5.13	2.56	2.56	100

The discriminant analysis using this marker produced a reclassification of only six genotypes. These are the additional highlighted cells in Table 8-2 in Addendum B. This shows a much better discrimination between the genotypes. The different samples were accurately identified between each other.

Therefore this microsatellite could be used on its own to distinguish between different genotypes. If it is a small number of genotypes then this microsatellite should be able to correctly distinguish between all of them. The number of samples to discriminate between has an influence on which marker to use.

4.1.5.3. Amplified Fragment Length Polymorphism (AFLP)

Nine different primer combinations (Table 3-4) were tested on the original population of twenty genotypes (Table 3-1). The final amplification product was visualised on the ABI 3130XL and results were analysed and recorded using GeneMapper. Each set of results is given in Addendum B in Tables 8-3, 8-6, 8-9, 8-12, 8-15, 8-17, 8-21, 8-24 and 8-27 according to primer combination. These tables contain the size of each band that was amplified in each sample. Only those bands that were chosen for final analysis are given in Addendum B.

The tables listed above were used in a discriminant analysis using SAS. This test was performed per primer combination to determine which primer pairs were the most suited for this study. The twenty samples were used to formulate an equation that can then be used to classify the genotypes into a breeding program. This is done to check the accuracy of the variables obtained from each sample. Table 4-12 illustrates the number of samples that were reclassified in each primer combination. The percentage of reclassifications per breeding

program is given for each primer combination in the Addendum together with the linear functions of the variables of each primer combination. The results in Table 4-12 can be used together with the results given in Table 4-4, regarding the number of polymorphisms/alleles observed per primer combination. The number of polymorphic bands and the effectiveness of the bands demonstrate the value of each primer combination in this study. The number of polymorphic bands observed is also a good indication of the primers that have good discriminatory power among genotypes.

Table 4-10 The number of correct classifications obtained from each primer combination using discriminant analysis and SAS.

Primer Combination	***Eco*RI-ACA**	***Eco*RI-AAC**	***Eco*RI-AGG**	**Total**
***Mse*I-CAG**	14	17	18	49
***Mse*I-CAT**	19	20	20	59
***Mse*I-CTC**	20	19	20	59
Total	53	56	58	167

The results given in Table 4-12 shows that the *Mse*I-CTC and *Mse*I-CAT selective primers are the best for discriminating between genotypes of the breeding programmes. These primers in combination with the three *Eco*RI selective primers managed to correctly classify 59 of the 60 genotypes. This translates to 98% accuracy for each *Mse*I selective primer. Table 4-12 also shows that the most accurate *Eco*RI selective primer is *Eco*RI-AGG. This primer showed 97% accuracy.

4.1.5.4. Sequence-Specific Amplified Polymorphism (SSAP)

This technique produced an exceedingly large amount of polymorphic bands that were problematic to analyse and it was therefore not done. It can be expected that these data would have been even more powerful in discriminating among the breeding programmes.

4.2. IMPLEMENTATION OF THE SELECTED MARKERS ON A POPULATION OF BREEDING LINES

Intra-line variability was detected in the first part of this study, so for this section DNA from four plants per genotype was pooled. This will provide a better representation of the alleles in each of the breeding lines, particularly as the breeding lines are F_5-derived and may contain considerable heterogeneity.

The better marker systems were chosen to test 119 breeding lines. A list of these lines is given in Table 3-9. Microsatellites and AFLPs were chosen from the techniques tested. These showed the highest level of polymorphism between the panel of released varieties/advanced lines and are the easiest to analyse. All six of the fluorescently labelled microsatellites were used to screen the 119 breeding lines. However, the complete profiles generated with the microsatellite *Xwmc177* primers could have been used instead. Three AFLP primer combinations were used. The three labelled *Eco*RI primers were used in combination with the *Mse*I-CTC primer. The highest number of polymorphic alleles was observed with these three primer combinations. Similarly, the SSAP primers could have been used instead of the selected group of AFLP primers.

4.2.1. Microsatellites

It was possible to multiplex four of the labelled primers: *Xgwm190*, *Xwmx11*, *Xwmc59* and *Xwmc177*. These four primers were each labelled with a different colour. This made amplifying them together that much easier. The other two microsatellites (*Xgwm437* and *Xgwm539*) were amplified separately. These two were labelled with different colours. It was possible to electrophorese all six microsatellites together. The fragments of the same colour had different size ranges so they did not influence each other's results.

4.2.2. Amplified Fragment Length Polymorphisms (AFLPs)

Three primer combinations (Table 4-13) were used to screen the 119 breeding lines. These showed the highest number of polymorphic loci in the panel of released varieties/advanced lines.

Table 4-11 Total number of polymorphisms/alleles for the AFLP selective primer pair combinations used to screen the 119 breeding lines.

Primer Combination	*Eco*RI-AAC	*Eco*RI-ACA	*Eco*RI-AGG	Total
***Mse*I-CTC**	35	36	47	118

The number of polymorphic bands obtained from the breeding population was considerably less than was obtained in the evaluation of this marker type. This is exactly why the primer combinations with the highest number of polymorphic bands were chosen to screen the breeding lines. It ensured that there would be a number of alleles detected between the lines.

4.2.3. Data analysis

The SAS program was used to perform a cluster analysis on the results for the microsatellites, all three AFLP primer combinations separately and microsatellite *Xwmc177* on its own. These clusters were developed in five different tests; each time using the same 119 samples but the different results for each marker. The samples per cluster are provided in Table 4-12. The number of clusters for each marker is at an accuracy of 95%. The cluster history for the microsatellites and *Xwmc177* are given in Table 8-30 and the three AFLP primer combinations are given in Table 8-31 in Addendum B.

Table 4-12 Samples in the clusters achieved at 95% accuracy for each of the five markers used to test the 119 breeding lines.

Marker Type	Microsatellites	*Xwmc177* only	*Mse*I-CTC & *Eco*RI-ACA	*Mse*I-CTC & *Eco*RI-AAC	*Mse*I-CTC & *Eco*RI-AGG
Total Clusters	12	20	36	33	53
Cluster Number	**Samples per cluster**				
1	21, 25, 65, 3, 30, 80, 70, 76	1-8, 11-24, 26-36, 39, 41-45, 47, 48, 49, 50, 51, 52, 55-60, 62-64, 67, 70, 71, 72, 73, 74, 75, 76, 78, 79, 80, 81, 82, 83, 86, 87, 88, 89, 92, 94, 96, 97, 100, 103, 105-107, 111, 112	4, 12, 7, 45	2, 4, 5, 6, 7, 12, 13, 14, 20, 23, 26, 28, 29, 30, 36, 38, 41, 42, 44, 46, 50, 54, 55, 58, 60, 61, 110, 97, 45	10, 16, 23, 27, 43, 109, 7

Marker Type	Microsatellites	*Xwmc177* only	*Mse*I-CTC & *Eco*RI-ACA	*Mse*I-CTC & *Eco*RI-AAC	*Mse*I-CTC & *Eco*RI-AGG
Total Clusters	12	20	36	33	53
Cluster Number	**Samples per cluster**				
2	49, 57, 45, 79, 44	9, 25, 38, 65, 66, 99	3, 13, 15, 19, 29, 34, 35, 57, 58, 60, 44	3, 8, 15, 19, 22, 31, 33, 47, 48, 105, 106, 107	13, 19, 12
3	55, 62, 52, 78, 32, 67, 63, 81, 31, 59, 92, 64, 71	10, 98	14, 22, 38, 62	10, 16, 21, 40, 102, 109, 119, 118	17, 21, 118, 119
4	27, 89, 51, 93, 22, 17, 47	37, 102, 117	18, 26, 50, 51, 55, 108, 39	9, 18, 24, 25, 34, 35, 39	11, 25, 28, 34, 35, 40, 44, 54, 61, 51
5	83, 103, 96, 91, 2, 5, 85, 105, 98, 4, 7, 84, 66, 107, 112, 33, 86, 13, 82, 99, 106	101, 108	5, 30, 37, 105, 11	11, 27, 43, 59	14, 26, 29, 39, 41, 42, 46, 50, 55, 58, 59, 62, 107
6	8, 11, 37, 15, 100, 9, 97, 14, 95	68, 84, 54	1, 31, 32, 33, 56, 49, 9	51, 53, 57, 62, 68, 72, 83, 92, 101	2, 38, 53, 60, 110, 18, 22
7	18, 20, 19, 24, 87, 56, 68, 28, 1, 48, 58	61, 69	28, 36, 61, 52	1, 49, 56, 108	6, 47, 30, 102
8	26, 94, 36, 69, 118, 104, 77, 101, 115, 102, 109, 53, 16, 117, 116, 46, 119, 113, 12, 110, 38, 42, 108, 114	104, 119	23, 41, 2, 107, 102, 110, 16	32, 69, 98, 104, 37, 52	95, 106
9	72, 90, 111, 23, 29, 41, 50, 60	116, 118	25, 42, 10, 40	64, 70, 90, 85	75, 96, 98, 68, 103
10	35, 43, 74, 88, 40, 54	114, 115	46, 53, 54	74, 75, 91, 73	67, 82, 93, 97
11	6, 10, 61, 73	93	68, 69, 20, 97	66, 67	4, 99, 33, 15
12	39, 75, 34	95	73, 74, 89	86, 112	52, 105
13		77	63, 81, 85, 84, 96, 92, 79, 82	71, 116	81, 85, 70
14		110	70, 83, 64	76, 103	66, 101
15		46	99, 100, 59, 106	81, 93, 88	3, 31
16		40	21, 109, 118, 17	111, 117	74, 88, 83
17		113	111, 112, 117, 86	65, 77	64, 65
18		85	66, 67	78, 80	37, 108
19		53	77, 80	98	1, 49
20		109	8, 47, 24	99	63, 84
21			76, 78, 93, 101	100	76, 77
22			88, 90, 27	84	94, 115
23			65, 91	94	5, 24
24			6, 43	113	32
25			98, 104	63	71

Marker Type	Microsatellites	*Xwmc177* only	*Mse*I-CTC & *Eco*RI-ACA	*Mse*I-CTC & *Eco*RI-AAC	*Mse*I-CTC & *Eco*RI-AGG
Total Clusters	12	20	36	33	53
Cluster Number	**Samples per cluster**				
26			95, 103	95	48
27			94, 115	89	79
28			114	79	80
29			113	17	69
30			116	82	91
31			119	87	113
32			75	114	104
33			48	115	56
34			72		73
35			87		90
36			71		36
37					100
38					111
39					112
40					45
41					20
42					89
43					78
44					86
45					92
46					116
47					9
48					72
49					8
50					57
51					87
52					114
53					117

The numbers highlighted in Table 4-12 correspond to the numbers of the breeding lines in Table 3-9. These samples were used to test the number of seeds that can accurately represent the genotype of the population.

Table 4-12 displays the differences in the marker systems. AFLP is clearly the marker with the greatest discriminatory power between these breeding lines. The most powerful primer combination is *Mse*I-CTC and *Eco*RI-AGG. The other two AFLP primer combinations are very similar. This table also illustrates that using *Xwmc177* as an AFLP is more informative than the six microsatellites. This adds to the idea that AFLP is a better marker system to distinguish between a large number of genotypes.

4.2.4. Number of individuals to use per sample

AFLP was chosen to test the thirty two samples; sixteen samples consisting of eight seeds and the same sixteen samples but consisting of twelve seeds. The results (banding pattern) for the three AFLP primer combinations per sixteen samples and number of seeds is given in Table 8-32 to 8-40 in Addendum B. Included in these tables is the cluster history for each primer combination for both eight and twelve seeds. The data were analysed using the SAS program and a cluster analysis function. The clusters and the samples per cluster for the three AFLP primer combinations are given in Table 4-13. The clusters are also given according to the number of seeds pooled in the extraction.

Table 4-13 Samples in the clusters achieved at a 95% accuracy for each of the three primer combinations used to test the sixteen breeding lines with eight or twelve seeds.

Primer Combination	***Mse*I-CTC & *Eco*RI-ACA**		***Mse*I-CTC & *Eco*RI-AAC**		***Mse*I-CTC & *Eco*RI-AGG**	
Number of seeds	**8**	**12**	**8**	**12**	**8**	**12**
Total Clusters	**13**	**12**	**10**	**11**	**13**	**11**
Cluster Number	**Samples per cluster**					
1	95, 113, 86	86, 95, 104	72, 104, 73	72, 78, 95	86, 114, 95	86, 95, 113, 114
2	80, 114	78, 100	89, 114, 115	80, 114	89, 113	78, 80
3	89	72, 80	78, 116	86, 104	71	72, 73
4	104	73	48, 87	89, 116	80	100
5	48	89	71	71	72	104
6	71	114	95	100	115	71
7	87	113	113	113	48	116
8	116	87	80	48	104	89
9	115	116	100	73	73	87
10	72	48	86	87	78	48
11	73	115		115	100	115
12	78	71			116	
13	100				87	

Table 4-13 shows that a sample of eight seeds could be adequate. For two of the primer pairs, increasing the number of seeds per sample results in slightly fewer clusters. This would suggest that a sample of twelve seeds is more representative of all the bands that occur in the variety/line. Thus fewer distinct classes are picked up. However, the difference is small and would suggest that while it may be better to involve more seeds in a sample, a sample of eight seeds would still give a good account of the intra-line variability. Therefore this system

proves to be the best for distinguishing between genotypes. It also demonstrated a good overall representation of the entire line. This system is not just based on an individual plant from a line but it shows discrimination for a large portion of the line. Table 4-13 also illustrates that if one primer combination is not able to separate specific lines then one of the others can be used.

4.3. SYSTEM FOR GENOTYPING WHEAT AT STELLENBOSCH UNIVERSITY.

The final information obtained during this study illustrates that the marker system to use is determined by two very clear issues. The first is the number of different genotypes to be tested and the other is the funds available for the test. These two factors will determine whether microsatellites and/or AFLP/SSAP should be used to fingerprint a wheat collection.

In the case of the wheat breeding program at the University of Stellenbosch, both systems are recommended. The microsatellites would be used when one or two lines are to be compared with lines/released varieties from other breeding programmes. This system is cheaper and provides clear and specific differences between genotypes. With few genotypes being compared it is generally easy to score the microsatellite markers from sequencing gels. Wheat microsatellite markers vary in their utility for genotyping and may broadly be grouped into two types. Firstly, some microsatellite primers produce complex profiles that include structured patterns. All the bands that are part of the pattern derive from the same locus and convey the same information. The patterns are normally easy to score on a sequencing gel, however, for automated scoring a specific band in the pattern needs to be targeted. To ensure that alleles are identified correctly, the pattern (band) that is scored should not overlap with other patterns/bands in the profile. The type of microsatellite primers such as *Xwmc177* generate a large number of bands from seemingly different loci that are better suited to automated detection and analysis. These microsatellites are generally more versatile. Microsatellites are also cheaper to perform and can be tested and analysed quicker than AFLPs. If the number of samples tested is not too great then data analysis is not required. Microsatellites generate less information so comparisons can be made just by looking at the banding patterns.

However, if a larger number of samples are to be tested or if these lines are known to be closely related then AFLPs should be used. The three primer combinations used in section 4.2 should be sufficient when used either singly or in combination. SSAP analysis compares cost wise with AFLPs yet is more powerful and can similarly be used on its own or supplemented with AFLP. This is an expensive technique but the high level of polymorphism makes it appropriate for genotype discrimination.

The optimal number of plants that should be pooled during gDNA extraction is clearly a function of the level of heterogeneity within lines/varieties. In the present study the F_5-derived breeding lines can be expected to have a relatively high level of intra-line variability. It appears that increasing the number of individual samples per line from eight to twelve results only in small changes in the ability of the technique to detect inter-line differences. While a bigger sample will be more representative, a bulk of eight plants will still provide a relatively good indication of real differences among lines.

5. CONCLUSION

The main objective of this study was to develop a DNA-based diagnostic technique that could distinguish between various *T. aestivum* genotypes. Other objectives were considered during the course of the study. These included, the effective implementation of the technique; the number of seeds that could represent an entire population; and the time and cost needed for the technique.

Each of these objectives was investigated and it was clear that the diagnostic technique would depend on a few things. Firstly, the technique would be either microsatellites or AFLPs. The retrotransposons proved to be less helpful than the other well-known techniques, in particular when the primer sequence was derived from transposons identified in barley. However, complementing a wheat transposon based primer with an AFLP primer in a SSAP application proved to be very successful, yet the vast number of polymorphic loci detected made it difficult to analyze. Secondly, the number of genotypes that will be compared is important. Microsatellites would be more effective when comparing a smaller group of unrelated genotypes. This is also a much cheaper and easier technique to implement. There are many, many microsatellites in the wheat genome and the six used in this study are only the start. More polymorphic and applicable microsatellites can be added at any time to increase the discriminatory power of the marker system.

Amplified fragment length polymorphism is a more expensive technique to use than microsatellites. The level of polymorphism observed for this marker system makes it the better choice for more closely related genotypes. It will also be more effective when comparing a large number of samples as it is well-suited to automated detection.

Therefore, the final thing to consider when using this diagnostic technique is this: do you have a few genotypes that can be compared for relatively low cost and produce easy to analyse data. Or do you have a large number of closely related genotypes that must be compared so cost is not a factor.

6. REFERENCES

- Abdel-Satar MA, Khalil MS, Mohmed IN, Abd-Elsalam KA, Verreet JA (2003) Molecular phylogeny of *Fusarium* species by AFLP fingerprint. African Journal of Biotechnology 2: 51-55

- Ahmad M (2002) Assessment of genomic diversity among wheat genotypes as determined by simple sequence repeats. Genome 45: 646-651

- Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes N, Sharp P, Vaughan P, Rathmell B, Huttner E, Kilian A (2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor Appl Genet 113: 1409-1420

- Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132: 1131-1139

- Avise JC (ed) (2004) Molecular Techniques. In: Molecular Markers, Natural History, and Evolution (2nd Edition). Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts, USA. pp 105

- Bandopadhyay R, Sharma S, Rustgi S, Singh R, Kumar A, Balyan HS, Gupta PK (2004) DNA polymorphism among 18 species of *Triticum-Aegilops* complex using wheat EST-SSRs. Plant Science 166: 349-356

- Beharav A, Nevo E (2003) Predictive validity of discriminant analysis for genetic data. Genetica 119: 259-267

- Beyermann B, Nürnberg P, Weihe A, Meixner M, Epplen JT, Börner T (1992) Fingerprinting plant genomes with oligonucleotide probes specific for simple repetitive DNA sequences. Theor Appl Genet 83: 691-694

- Blair MW, Hedetale V, McCouch SR (2002) Fluorescent-labeled microsatellite panels useful for detecting allelic diversity in cultivated rice (*Oryza sativa* L.). Theor Appl Genet 105: 449-457

- Bohn M, Utz HF, Melchinger AE (1999) Genetic similarities among Winter wheat cultivars determined on the basis of RFLPs, AFLPs, and SSRs and their use for predicting progeny variance. Crop Sci 39: 228-237

- Brown TA (ed) (2002) Studying DNA. In: Genomes (2nd Edition). BIOS Scientific Publishers Ltd, Oxford, UK. pp 121-122

- Chahal GS, Gosal SS (2002) Molecular approaches to crop improvement. In: Principles and procedures of plant breeding: Biotechnological and conventional approaches. Alpha Science International Ltd, Pangbourne, England. Edited by Chahal GS, Gosal SS. pp 473-485

- Chen XM, Line RF, Leung H (1998) Genome scanning for resistance-gene analogs in rice, barley, and wheat by high-resolution electrophoresis. Theor Appl Genet 97: 345-355

- Coburn JR, Temnykh SV, Paul EM, McCouch SR (2002) Design and application of microsatellite marker panels for semiautomated genotyping of rice (*Oryza sativa* L.). Crop Sci 42: 2092-2099

- Condit R, Hubbell SP (1991) Abundance and DNA sequence of two-base repeat regions in tropical tree genomes. Genome 34: 66-71

- Congiu L, Chicca M, Cella R, Rossi R, Bernacchia G (2000) The use of random amplified polymorphic DNA (RAPD) markers to identify strawberry varieties: a forensic application. Molecular Ecology 9: 229-232

- Dallas JF (1988) Detection of DNA "fingerprints" of cultivated rice by hybridization with a human minisatellite DNA probe. Proc Natl Acad Sci USA 85: 6831-6835

- Der G, Everitt BS (2002) A handbook of statistical analyses using SAS (2nd Edition), Chapman & Hall/CRC, New York, USA.

- Diaz V, Ferrer E (2003) Genetic variation of populations of *Pinus oocarpa* revealed by resistance gene analog polymorphism (RGAP). Genome 46: 404-410

- Diversity Arrays Technology (www.diversityarrays.com/index.html) 5 September 2007

- Dograr N, Akin-Yalin S, Akkaya MS (2000) Discriminating durum wheat cultivars using highly polymorphic simple sequence repeat DNA markers. Plant Breeding 119: 360-362

- Donini P, Elias ML, Bougourd SM, Koebner RMD (1997) AFLP fingerprinting reveals pattern differences between template DNA extracted from different plant organs. Genome 40: 521-526

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

- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12: 13-15

- Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104: 399-407

- Evenson RE, Gollin D (2003) Crop genetic improvement in developing countries: Overview and summary. In: Crop variety improvement and its effect on productivity: The impact of international agricultural research. CABI Publishing, Cambridge, USA. Edited by Evenson RE, Gollin D. pp 16-19

- Fritz AK, Cox TS, Gill BS, Sears RG (1995) Molecular marker-facilitated analysis of introgression in Winter wheat x *Triticum tauschii* populations. Crop Sci 35: 1691-1695

- Gao LF, Jing RL, Huo NX, Li Y, Li XP, Zhou RH, Chang XP, Tang JF, Ma ZY, Jia JZ (2004) One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. Theor Appl Genet 108: 1392-1400

- Gaudeul M, Till-Bottraud I, Barjon F, Manel S (2004) Genetic diversity and differentiation in *Eryngium alpinum* L. (Apiaceae): Comparison of AFLP and microsatellite markers. Heredity 92: 508-518

- Giroux MJ, Morris CF (1998) Wheat grain hardness results from highly conserved mutations in the friabilin components puroindoline a and b. Prac Natl Acad Sci 95: 6262-6266

- Grain Genes 2.0 (http://wheat.pw.usda.gov/GG2/index.shtml) 2 May 2007

- Gribbon BM, Pearce SR, Kalendar R, Schulman AH, Paulin L, Jack P, Kumar A, Flavell AJ (1999) Phylogeny and transpositional activity of Ty1-*copia* group retrotransposons in cereal genomes. Mol Gen Genet 261: 883-891

- Groenewald JZ, Fourie M, Marais AS, Marais GF (2005) Extension and use of a physical map of the *Thinopyrum*-derived *Lr19* translocation. Theor Appl Genet 112: 131-138

- Gupta PK, Varshney RK, Sharma PC, Ramesh B (1999) Molecular markers and their applications in wheat breeding. Plant Breeding 118: 369-390

- Gupta PK, Balyan HS, Edwards KJ, Isaac P, Korzun V, Röder M, Gautier M-F, Joudrier P, Schlatter AR, Dubcovsky J, De la Pena RC, Khairallah M, Penner G, Hayden MJ, Sharp P, Keller B, Wang RCC, Hardouin JP, Jack P, Leroy P (2002) Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat. Theor Appl Genet 105: 413-422

- Gyapay G, Ginot F, Nguyen S, Vignal A, Weissenbach J (1996) Genotyping procedures in linkage mapping. Methods: A Companion to Methods in Enzymology 9: 91-97

- Hart GE (2001) Molecular-marker maps of the cultivated wheat and other *Triticum* species. In: DNA-based markers in plants (2nd Edition). Kluwer Academic Publishers, London, UK. Edited by Phillips RL, Vasil IK. pp 421-441

- Huang L, Brooks SA, Li W, Fellers JP, Trick HN, Gill BS (2003) Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. Genetics 164: 655-664

- Huang XQ, Börner A, Röder MS, Ganal MW (2002) Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor Appl Genet 105: 699-707

- Incirli A, Akkaya MS (2001) Assessment of genetic relationships in durum wheat cultivars using AFLP markers. Genetic Resources and Crop Evaluation 48: 233-238

- Integrated DNA Technologies (http://www.idtdna.com/Home/Home.aspx) 2 May 2007

- Jakše J, Kindlhofer K, Javornik B (2001) Assessment of genetic variation and differentiation of hop genotypes by microsatellite and AFLP markers. Genome 44: 773-782

- Kahl G (ed) (2004) The Dictionary of Gene Technology: Genomics, Transcriptomics, Proteomics (3rd Edition). Wiley-VCH, Germany.

- Kalendar R, Grob T, Regina M, Suoniemi A, Schulman A (1999) IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theor Appl Genet 98: 704-711

- Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE*-1 retrotransposon dynamics in response to sharp microclimatic divergence. PNAS 97: 6603-6607

- Kim S, Dale BE (2004) Global potential bioethanol production from wasted crops and crop residues. Biomass & Bioenergy 26: 361-375

- Knott DR (ed) (1989) Monographs on theoretical and applied genetics. Vol 12. Springer-Verlag, Berlin, Germany

- Ko HL, Henry RJ, Graham GC, Fox GP, Chadbone DA, Haak IC (1994) Identification of cereals using the polymerase chain reaction. Journal of Cereal Science 19: 101-106

- Kobiljski B, Quarrie S, Denčić S, Kirby J, Ivegeš M (2002) Genetic diversity of the novi sad wheat core collection revealed by microsatellites. Cellular & Molecular Biology Letters 7: 685-694

- Kong L, Ohm HW, Cambron SE, Williams CE (2005) Molecular mapping determines that Hessian fly resistance gene *H9* is located on chromosome 1A of wheat. Plant Breeding 124: 525-531

- Korpelainen H, Virtanen V (2003) DNA fingerprinting of mosses. J Forensic Sci 48: 804-807

- Korzun V, Börner A, Worland AJ, Law CN, Röder MS (1997) Application of microsatellite markers to distinguish inter-varietal chromosome substitution lines of wheat (*Triticum aestivum* L.). Euphytica 95: 149-155

- Korzun V, Röder MS, Wendehake K, Pasqualone A, Lotti C, Ganal MW, Blanco A (1999) Integration of dinucleotide microsatellites from hexaploid bread wheat into a genetic linkage map of durum wheat. Theor Appl Genet 98: 1202-1207

- Kuleung C, Baenziger PS, Dweikat I (2004) Transferability of SSR markers among wheat, rye, and triticale. Theor Appl Genet 108: 1147-1150

- Leigh F, Kalendar R, Lea V, Lee D, Donini P, Schulman AH (2003a) Comparison of the utility of barley retrotransposon families for genetic analysis by molecular marker techniques. Mol Gen Genomics 269: 464-474

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

- Li C-D, Fatokun CA, Ubi B, Singh BB, Scoles GJ (2001) Determining genetic similarities and relationships among cowpea breeding lines and cultivars by microsatellite markers. Crop Sci 41: 189-197

- Mahmood A, Baenziger PS, Budak H, Gill KS, Dweikat I (2004) The use of microsatellite markers for the detection of genetic similarity among winter bread wheat lines for chromosome 3A. Theor Appl Genet 109: 1494-1503

- Manifesto MM, Schlatter AR, Hopp HE, Suárez EY, Dubcovsky J (2001) Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. Crop Sci 41: 682-690

- Manninen O, Kalendar R, Robinson J, Schulman AH (2000) Application of *BARE*-1 retrotransposon markers to the mapping of a major resistance gene for net blotch in barley. Mol Gen Genet 264: 325-334

- Mariette S, Chagné D, Lézier C, Pastuszka P, Raffin A, Plomion C, Kremer A (2001) Genetic diversity within and among *Pinus pinaster* populations: comparison between AFLP and microsatellite markers. Heredity 86: 469-479

- McIntosh SR, Pacey-Miller T, Henry RJ (2005) A universal protocol for identification of cereals. Journal of Cereal Science 41: 37-46

- Mueller UG, Wolfenbarger LL (1999) AFLP genotyping and fingerprinting. Tree 14: 389-394

- Myburg AA, Botha A-M, Wingfield BD, Wilding WJM (1997) Identification and genetic distance analysis of wheat cultivars using RAPD fingerprinting. Cereal Research Communications 25: 875-882

- Nagaraju J, Kathirvel M, Subbaiah EV, Muthulakshmi M, Kumar LD (2002) FISSR-PCR: a simple and sensitive assay for highthroughput genotyping and genetic mapping. Molecular and Cellular Probes 16: 67-72

- Perry DJ (2004) Identification of Canadian durum wheat varieties using a single PCR. Theor Appl Genet 109: 55-61

- Pestsova E, Ganal MW, Röder MS (2000) Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. Genome 43: 689-697

- Plaschke J, Ganal MW, Röder MS (1995) Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor Appl Genet 91: 1001-1007

- Porceddu A, Albertini E, Barcaccia G, Marconi G, Bertoli FB, Veronesi F (2002) Development of S-SAP markers based on an LTR-like sequence from *Medicago sativa* L. Mol Genet Genomics 267: 107-114

- Powell W, Thomas WTB, Baird E, Lawrence P, Booth A, Harrower B, McNicol JW, Waugh R (1997) Analysis of quantitative traits in barley by the use of Amplified Fragment Length Polymorphisms. Heredity 79: 48-59

- Prasad M, Varshney RK, Roy JK, Balyan HS, Gupta PK (2000) The use of microsatellites for detecting DNA polymorphism, genotype identification and genetic diversity in wheat. Theor Appl Genet 100: 584-592

- Queen RA, Gribbon BM, James C, Jack P, Flavell AJ (2004) Retrotransposon-based molecular markers for linkage and genetic diversity analysis in wheat. Mol Gen Genomics 271: 91-97

- Rampling LR, Harker N, Shariflou MR, Morell MK (2001) Detection and analysis systems for microsatellite markers in wheat. Aust J Agric Res 52: 1131-1141

- Röder MS, Plaschke J, König SU, Börner A, Sorrells ME, Tanksley SD, Ganal MW (1995) Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet 246: 327-333

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

- Röder MS, Wendehake K, Korzun V, Bredemeijer G, Laborie D, Bertrand L, Isaac P, Rendell S, Jackson J, Cooke RJ, Vosman B, Ganal MW (2002) Construction and analysis of a microsatellite-based database of European wheat varieties. Theor Appl Genet 106: 67-73

- Roussel V, Leisova L, Exbrayat F, Stehno Z, Balfourier F (2005) SSR allelic diversity changes in 480 European bread wheat varieties released from 1840 to 2000. Theor Appl Genet 111: 162-170

- Ryskov AP, Jincharadze AG, Prosnyak MI, Ivanov PL, Limborska SA (1988) M13 phage DNA as a universal marker for DNA fingerprinting of animals, plants and microorganisms. FEBS Letters 233: 388-392

- Sambrook J, Fritsch EF, Maniatis T (1991) Molecular cloning: A laboratory manual. Trends in Biotechnology 9: 213-214

- Schmidt AL, Gale KR, Ellis MH, Giffard PM (2004) Sequence variation at a microsatellite locus (XGWM261) in hexaploid wheat (*Triticum aestivum*) varieties. Euphytica 135: 239-246

- Schwarz G, Herz M, Huang XQ, Michalek W, Jahoor A, Wenzel G, Mohler V (2000) Application of fluorescence-based semi-automated AFLP analysis in barley and wheat. Theor Appl Genet 100: 545-551

- Shan X, Blake TK, Talbert LE (1999) Conversion of AFLP marker to sequence-specific PCR markers in barley and wheat. Theor Appl Genet 98: 1072-1078

- Shellenberger JA (1971) Production and utilization of wheat. In: Wheat chemistry and technology (2nd Edition). American Association of Cereal Chemists Inc, St. Paul, Minnesota, USA. Edited by Pomeranz Y. pp 11-17

- Skinner DM, Beattie WG, Blattner FR, Stark BP, Dahlberg JE (1974) The repeat sequence of a hermit crab satellite deoxyribonucleic acid is $(\text{-T-A-G-G-})_n \cdot (\text{-A-T-C-C-})_n$. Biochemistry 13: 3930-3937

- Soleimani VD, Baum BR, Johnson DA (2002) Identification of Canadian durum wheat [*Triticum turgidum* L. subsp. *Durum* (Desf.) Husn.] cultivars using AFLP and their STS markers. Can J Plant Sci 82: 35-41

- Stenesh J (ed) (1989) Dictionary of biochemistry and molecular biology (2nd edition). John Wiley & Sons, New York, USA.

- Sud S, Bains NS, Nanda GS (2005) Genetic relationships among wheat genotypes, as revealed by microsatellite markers and pedigree analysis. J Appl Genet 46: 375-379

- Talbert LE, Smith LY, Blake NK (1998) More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA. Genome 41: 402-407

- Tang S, Kishore VK, Knapp SJ (2003) PCR-multiplexes for a genome-wide framework of simple sequence repeat marker loci in cultivated sunflower. Theor Appl Genet 107: 6-19

- Terzi V, Morcia C, Gorrini A, Stanca AM, Shewry PR, 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

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

- Walker PMB (ed) (1989) Chambers Biology Dictionary. Chamber Cambridge, UK.

- Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BBT, Powell W (1997) Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol Gen Genet 253: 687-694

- Weising K, Nybom H, Wolff K, Kahl G (ed) (2005) DNA Fingerprinting in Plants: Principles, Methods, and Applications (2nd Edition). CRC Press, Taylor & Francis Group, Boca Raton, USA.

7. ADDENDUM A

This addendum contains all the gel photos that were obtained and then used in this study.

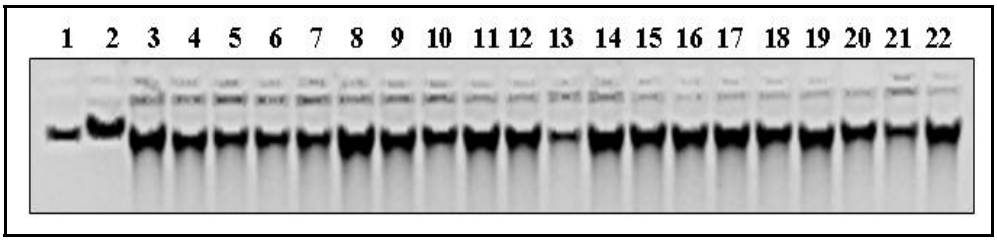


Figure 7-1 Gel photo showing the concentration of extracted gDNA. Lane 1 and 2 contains two different concentrations of lambda DNA. Lane 1: 0.1 µg/µl (1 µl λ and 4 µl dH_2O). Lane 2: 0.3 µg/µl (3 µl λ and 2 µl dH_2O). Lane 3 – 22: samples 1 to 20 as given in Table 3-1 (1 µl DNA stock and 4 µl dH_2O).

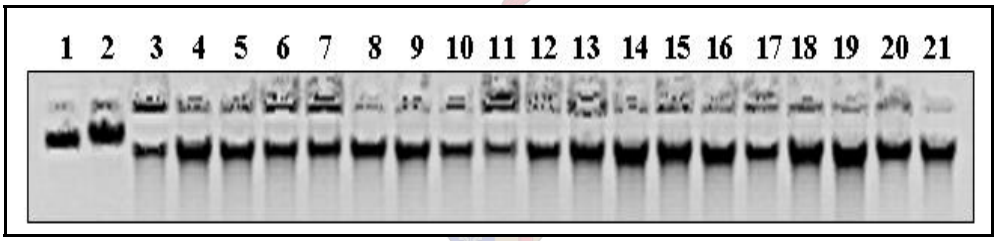


Figure 7-2 Gel photo showing the concentration of extracted gDNA. Lane 1 and 2 contains two different concentrations of lambda DNA. Lane 1: 0.1 µg/µl. Lane 2: 0.3 µg/µl. Lane 3 – 21: samples 21 to 39 as given in Table 3-1.

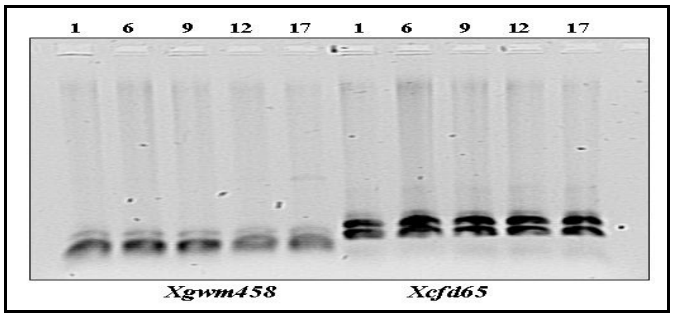


Figure 7-3 Gel photo confirming the amplification of the *Xgwm458* and *Xcfd65* microsatellites (1.5% agarose gel electrophoresed for 45 min at 100V). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

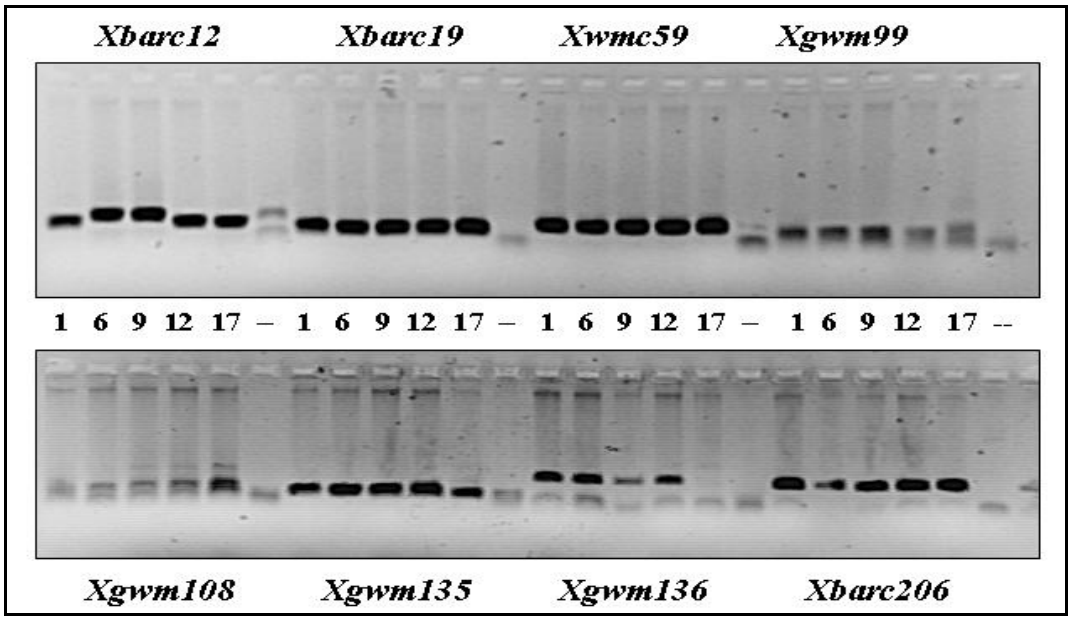


Figure 7-4 Gel photo confirming the amplification of the *Xbarc12, Xbarc19, Xwmc59, Xgwm99, Xgwm108, Xgwm135, Xgwm136* and *Xbarc206* microsatellites (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

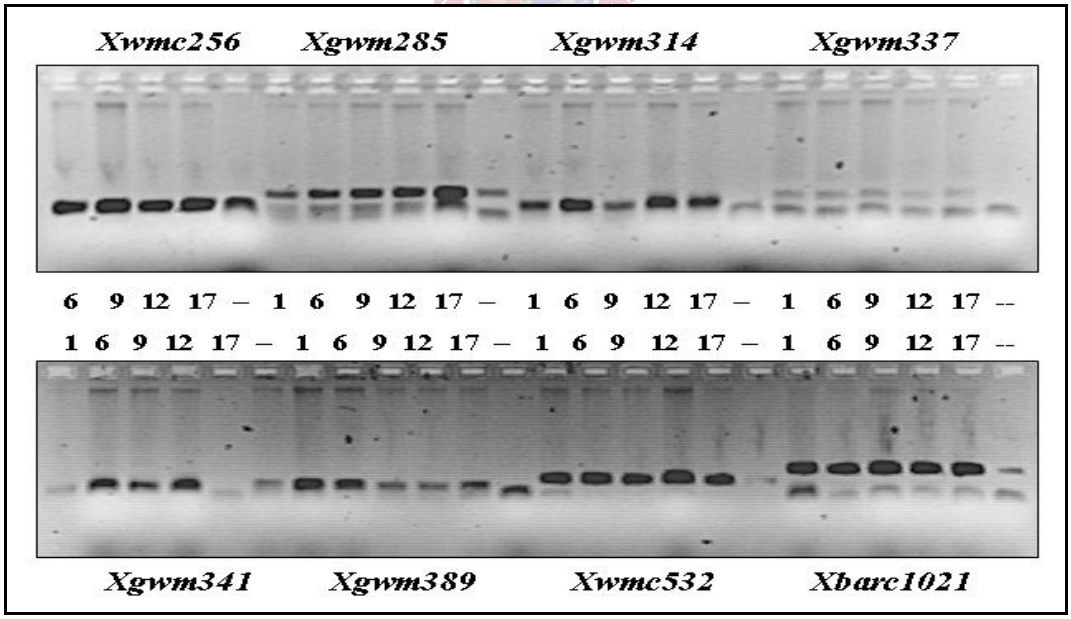


Figure 7-5 Gel photo confirming the amplification of the *Xwmc256, Xgwm285, Xgwm314, Xgwm337, Xgwm341, Xgwm389, Xwmc532* and *Xbarc1021* microsatellites (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

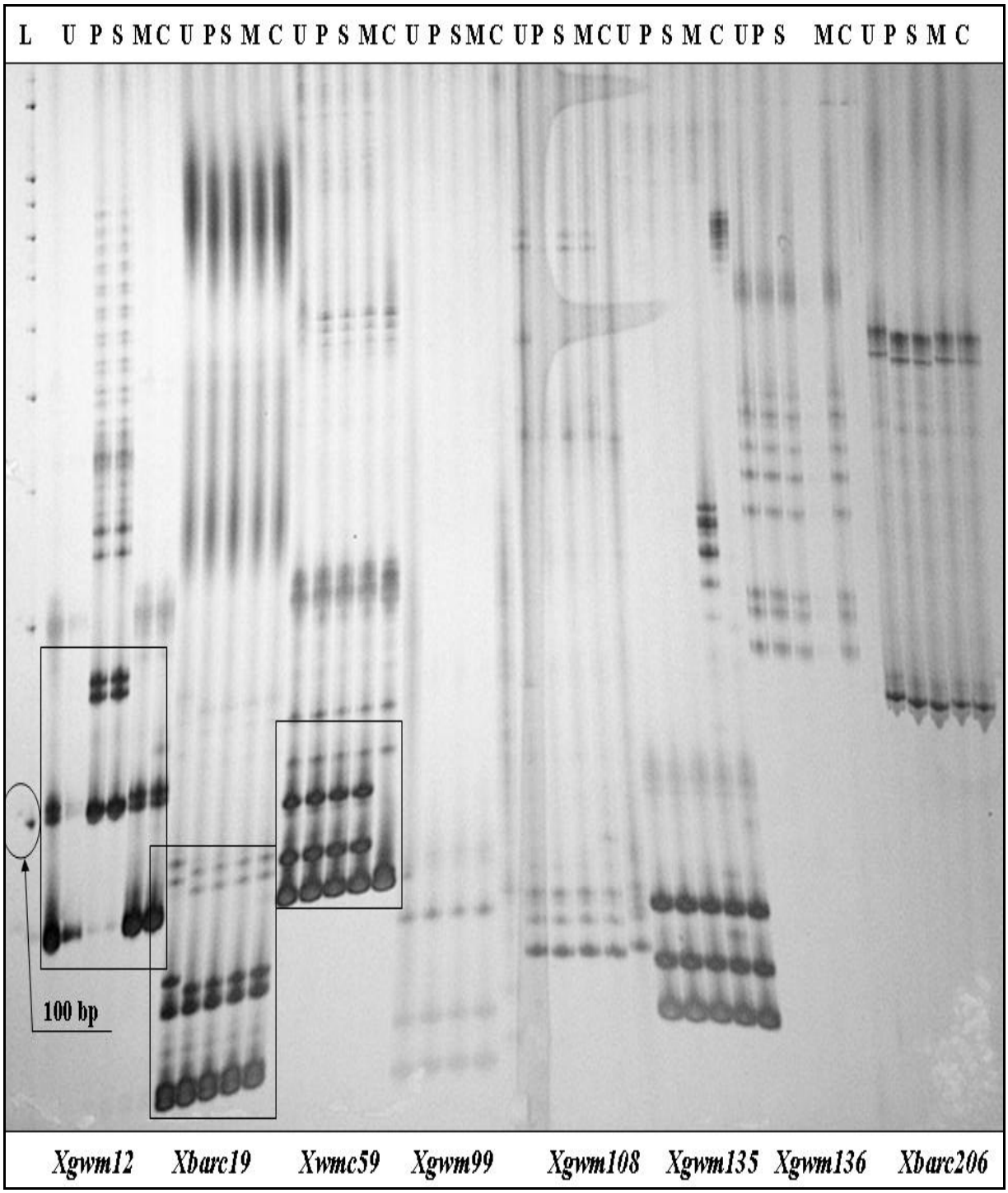


Figure 7-6 Gel photo illustrating the amplification of the *Xgwm12, Xbarc19, Xwmc59, Xgwm99, Xgwm108, Xgwm135, Xgwm136* and *Xbarc206* microsatellites (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The letters given above each lane correspond to the five breeding programmes (Table 3-1). L: 100 bp ladder.

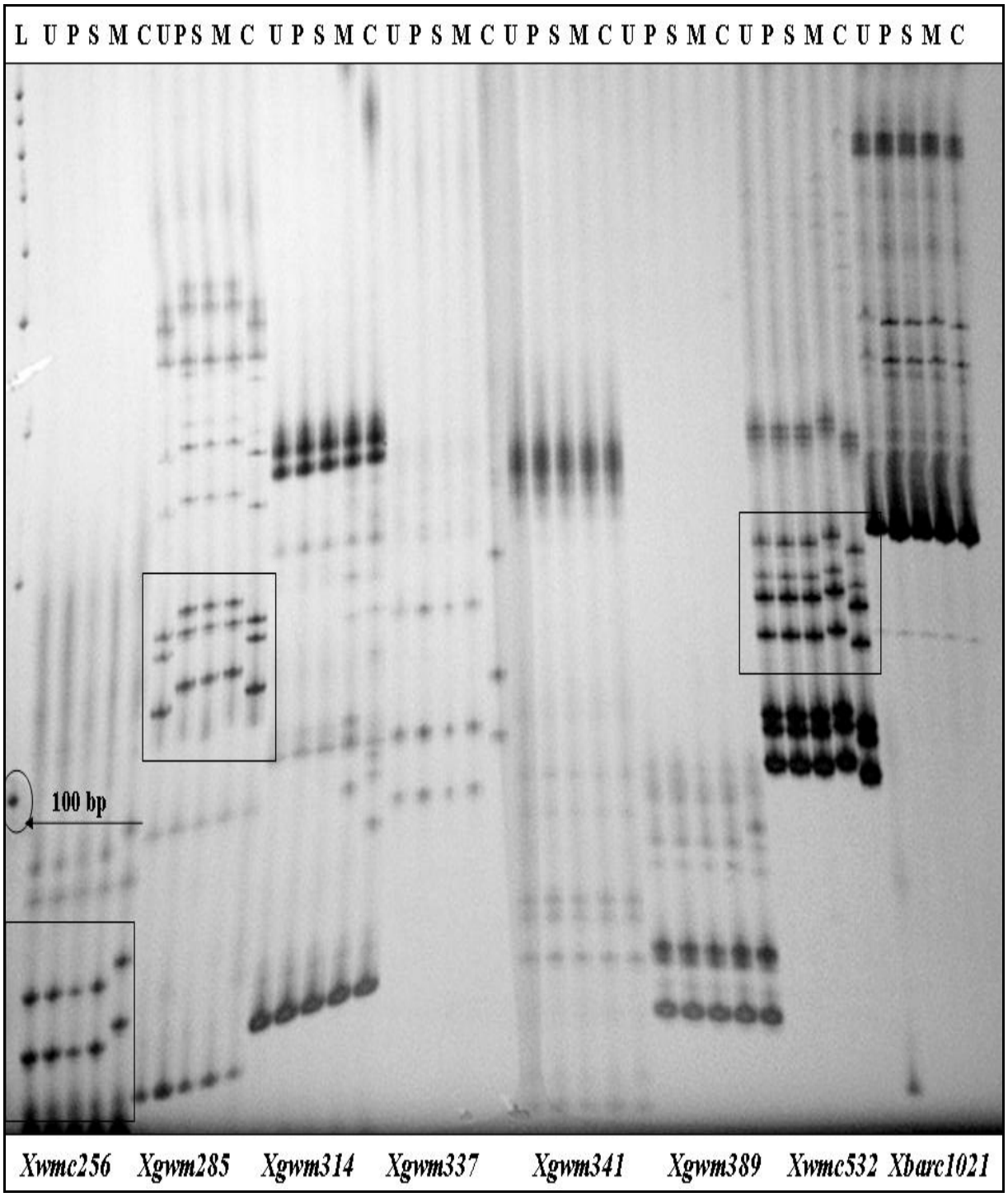


Figure 7-7 Gel photo illustrating the amplification of the *Xwmc256, Xgwm285, Xgwm314, Xgwm337, Xgwm341, Xgwm389, Xwmc532* and *Xbarc1021* microsatellites (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The letters given above each lane correspond to the five breeding programmes (Table 3-1). L: 100 bp ladder.

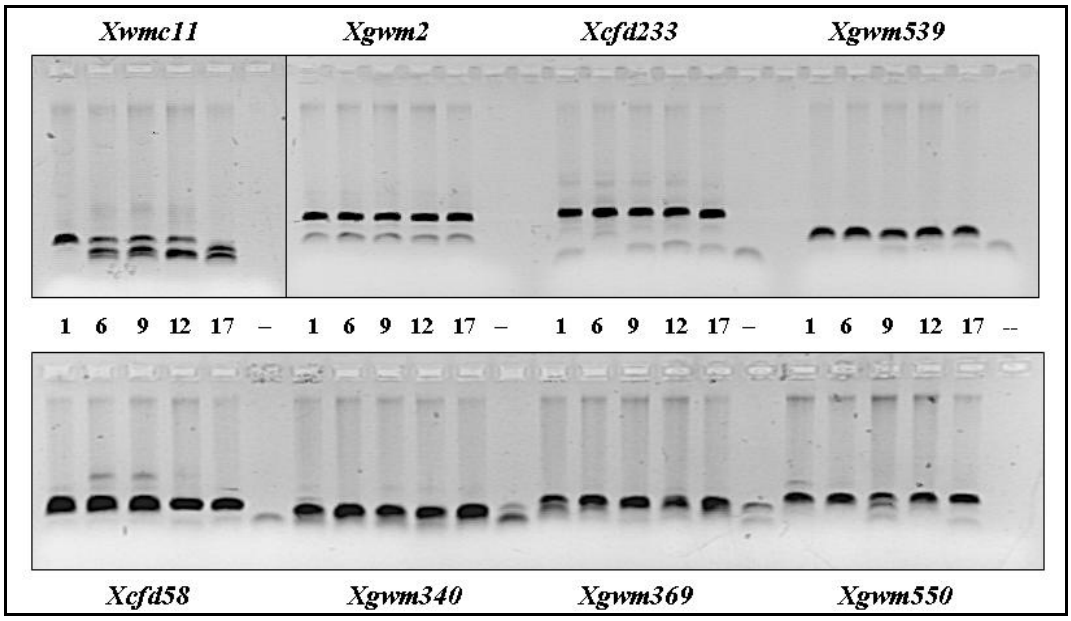


Figure 7-8 Gel photo confirming the amplification of the *Xwmc11, Xgwm2, Xcfd233, Xgwm539, Xcfd58, Xgwm340, Xgwm369* and *Xgwm550* microsatellites (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

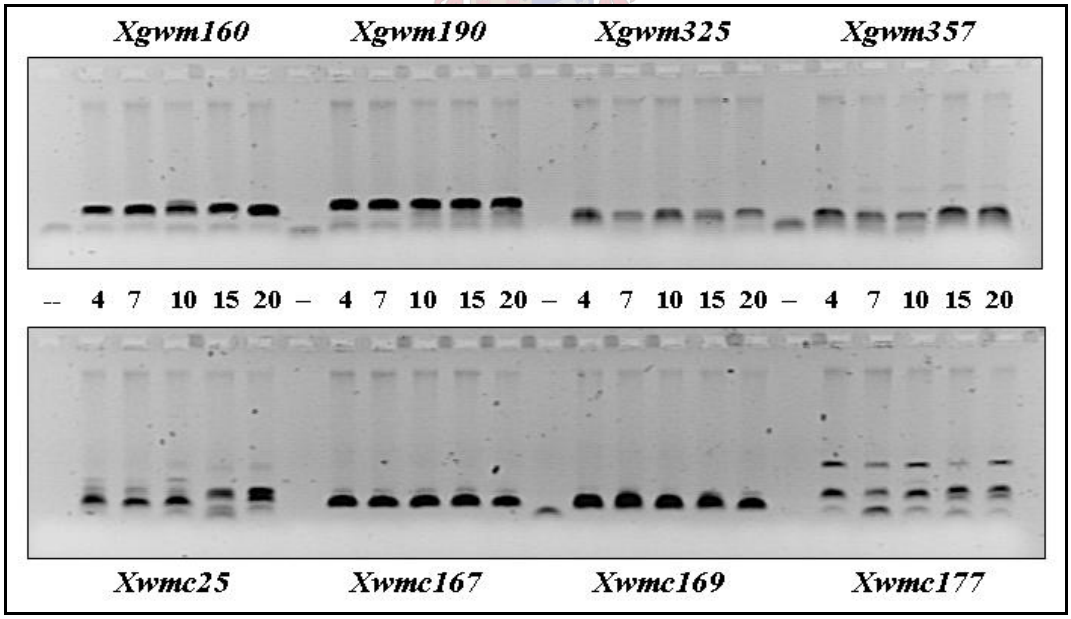


Figure 7-9 Gel photo confirming the amplification of the *Xgwm160, Xgwm190, Xgwm325, Xgwm357, Xwmc25, Xwmc167, Xwmc169* and *Xwmc177* microsatellites (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

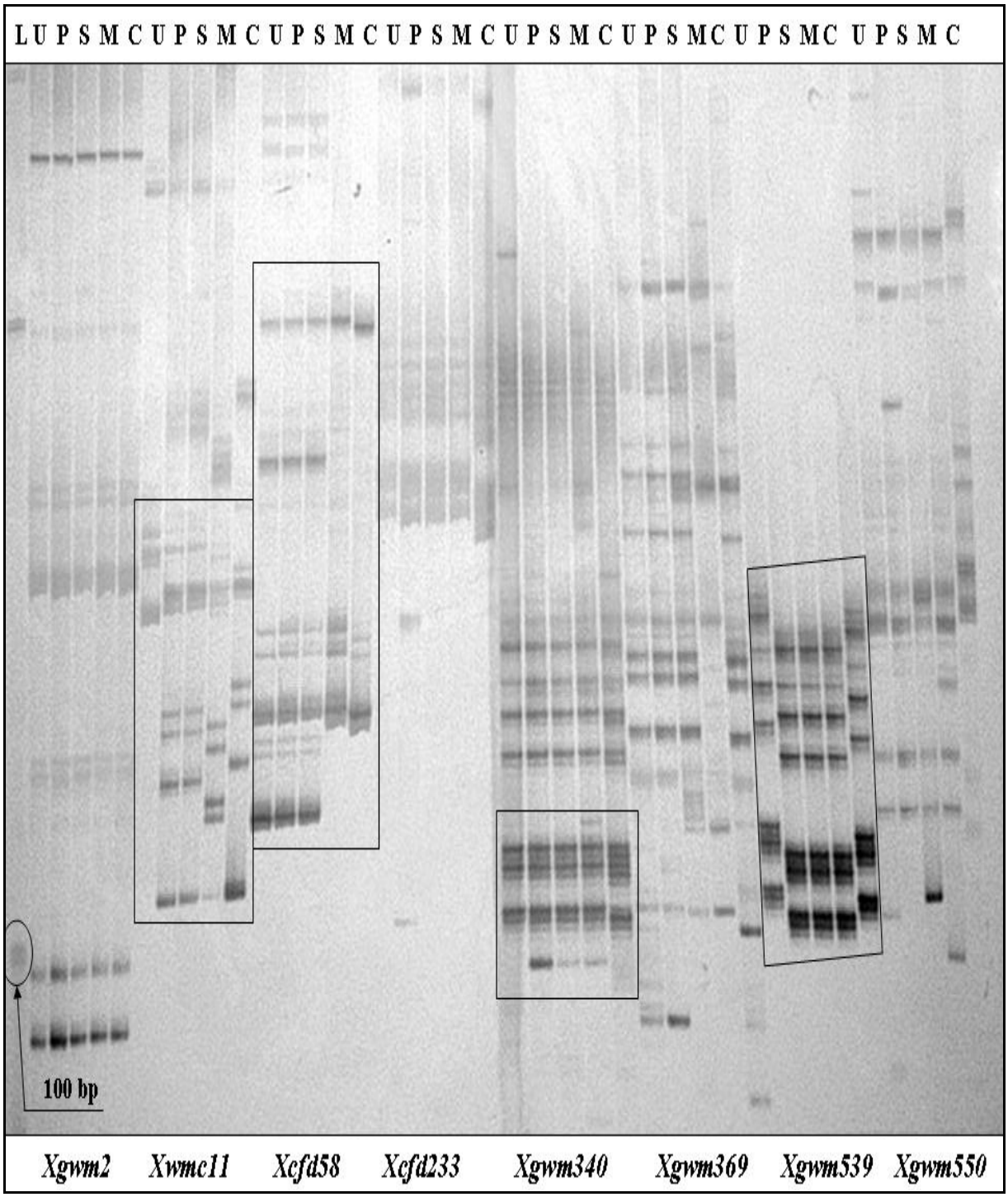


Figure 7-10 Gel photo illustrating the amplification of the *Xgwm2, Xwmc11, Xcfd58, Xcfd233, Xgwm340, Xgwm369, Xgwm539* and *Xgwm550* microsatellites (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The letters given above each lane correspond to the five breeding programmes (Table 3-1). L: 100 bp ladder.

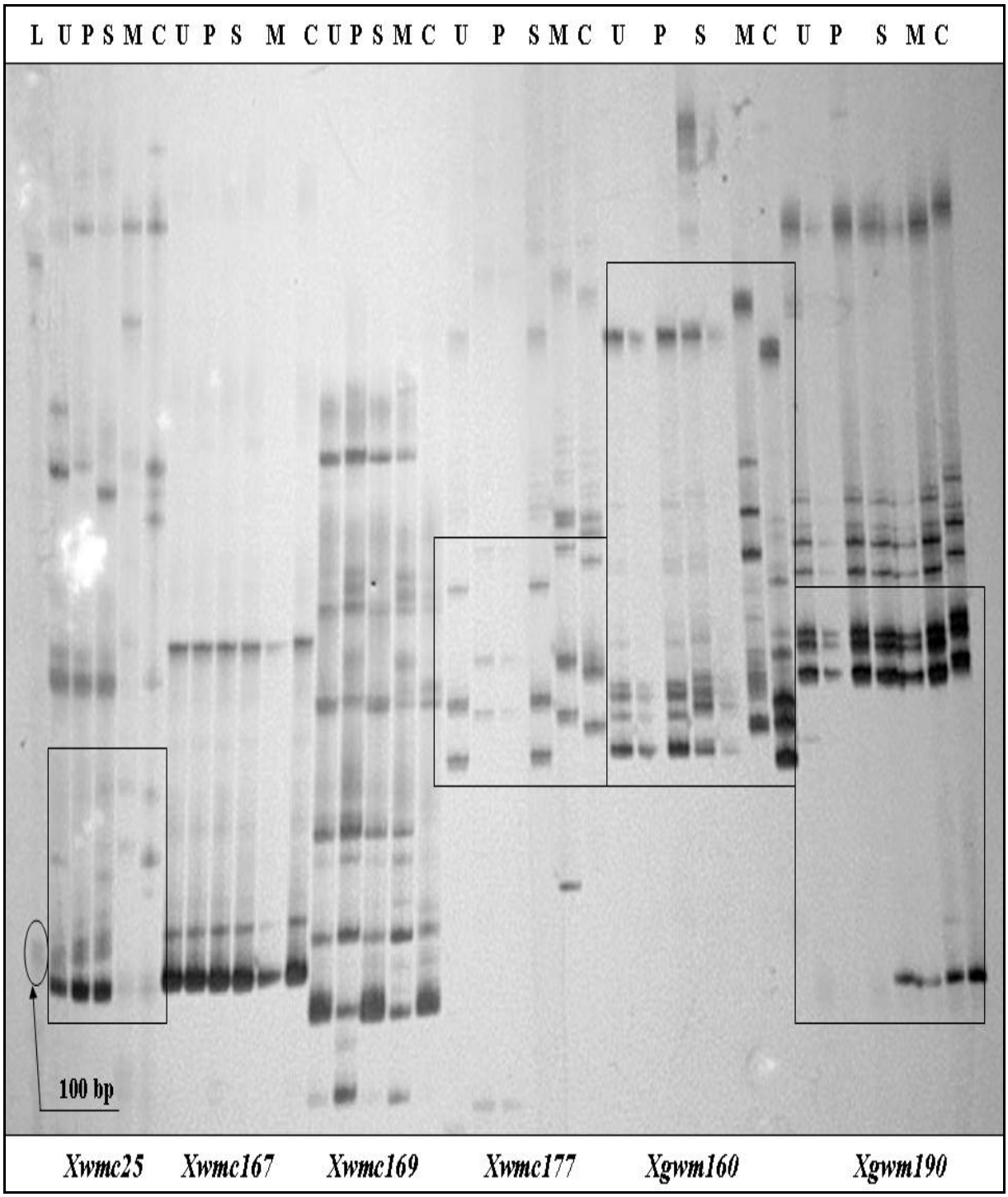


Figure 7-11 Gel photo illustrating the amplification of the *Xwmc25, Xwmc167, Xwmc169, Xwmc177, Xgwm160* and *Xgwm190* microsatellites (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The letters given above each lane correspond to the five breeding programmes (Table 3-1). L: 100 bp ladder.

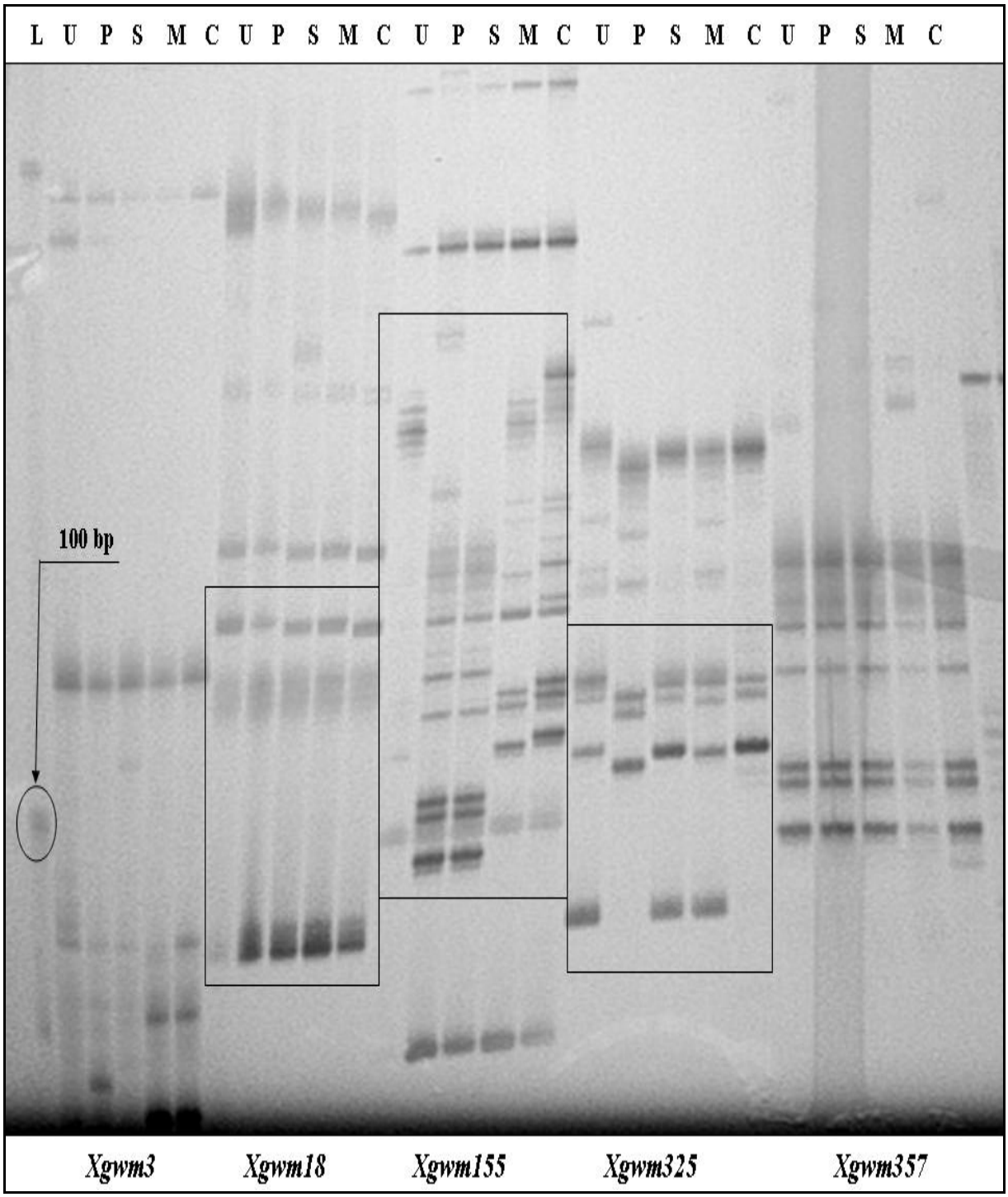


Figure 7-12 Gel photo illustrating the amplification of the *Xgwm3, Xgwm18, Xgwm155, Xgwm325* and *Xgwm357* microsatellites (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The letters given above each lane correspond to the five breeding programmes (Table 3-1). L: 100 bp ladder.

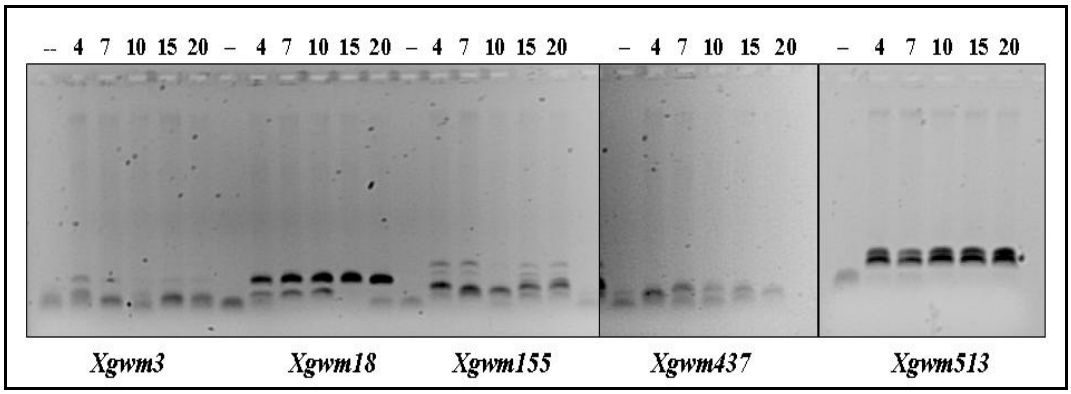


Figure 7-13 Gel photo confirming the amplification of the *Xgwm3, Xgwm18, Xgwm155, Xgwm437,* and *Xgwm513* microsatellites (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

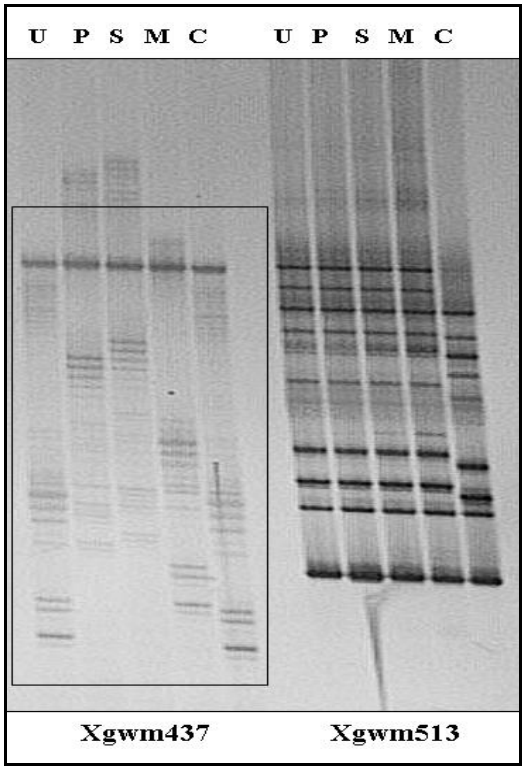


Figure 7-14 Gel photo illustrating the amplification of the *Xgwm4372* and *Xgwm513* microsatellites (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The letters given above each lane correspond to the five breeding programmes (Table 3-1).

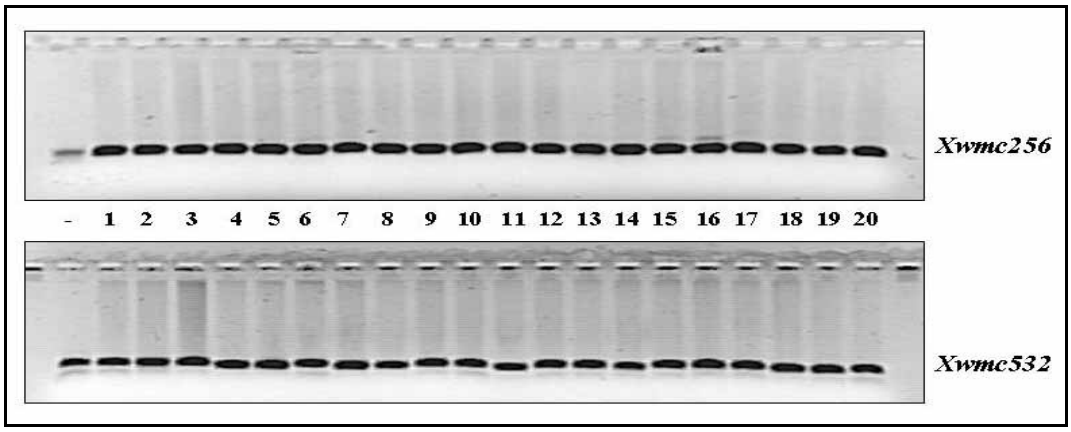


Figure 7-15 Gel photo confirming the amplification of the *Xwmc256* and *Xwmc532* microsatellites (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

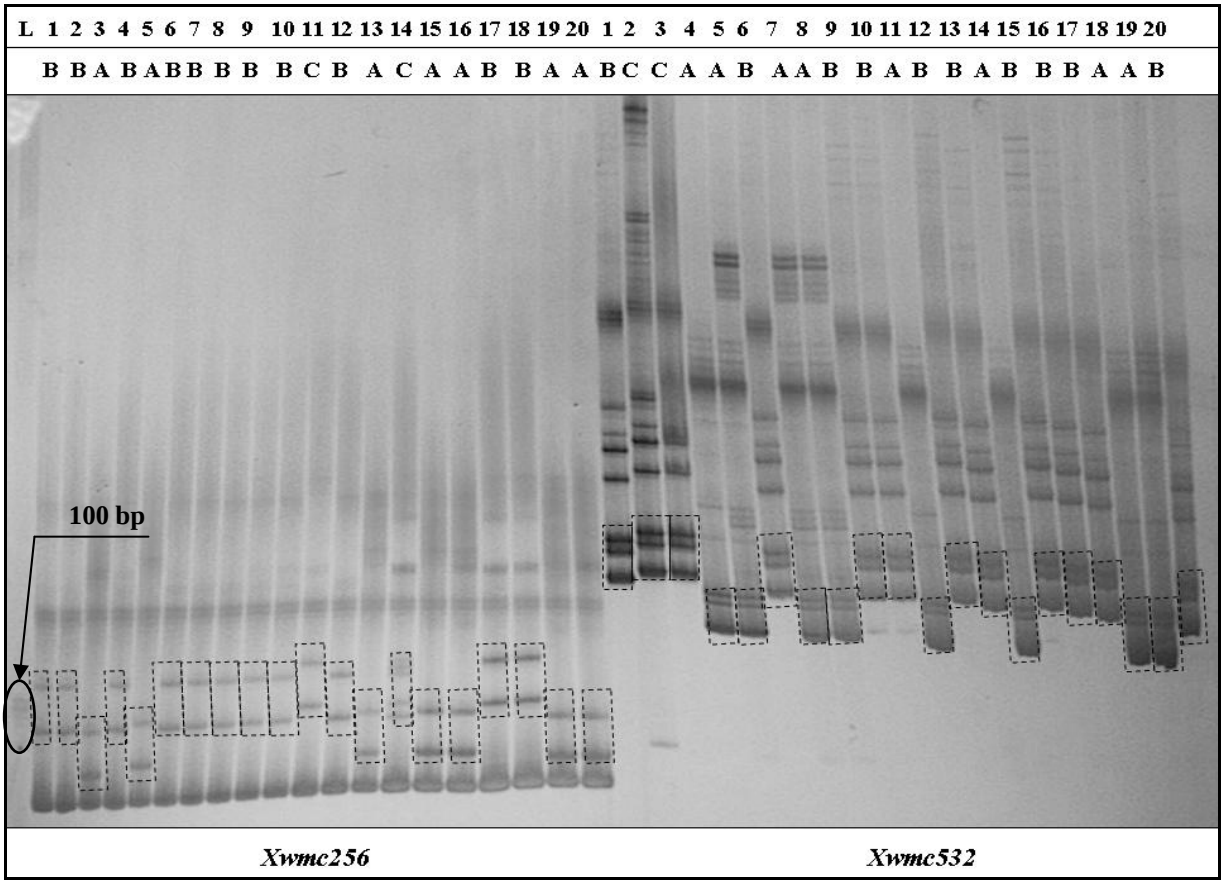


Figure 7-16 Gel photo illustrating the amplification of the *Xwmc256* and *Xwmc532* microsatellites (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1**.** L: 100 bp ladder.

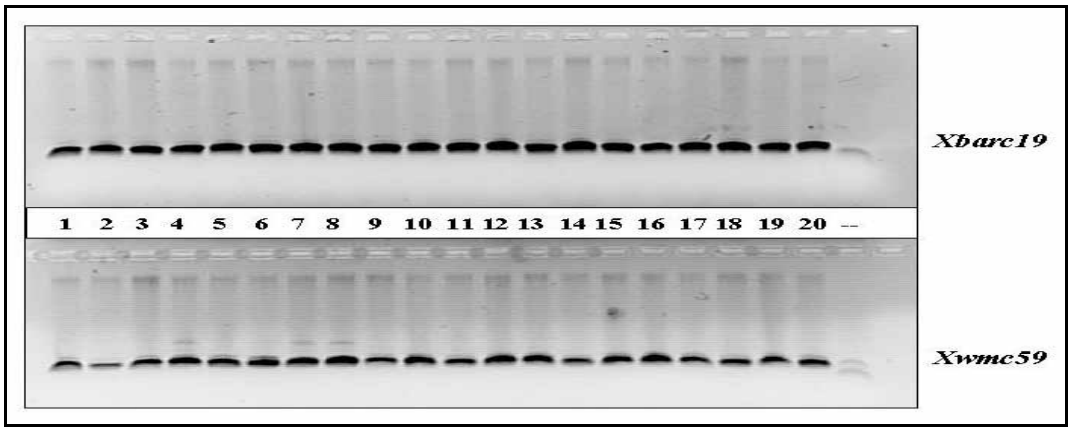


Figure 7-17 Gel photo confirming the amplification of the *Xbarc19* and *Xwmc59* microsatellites (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

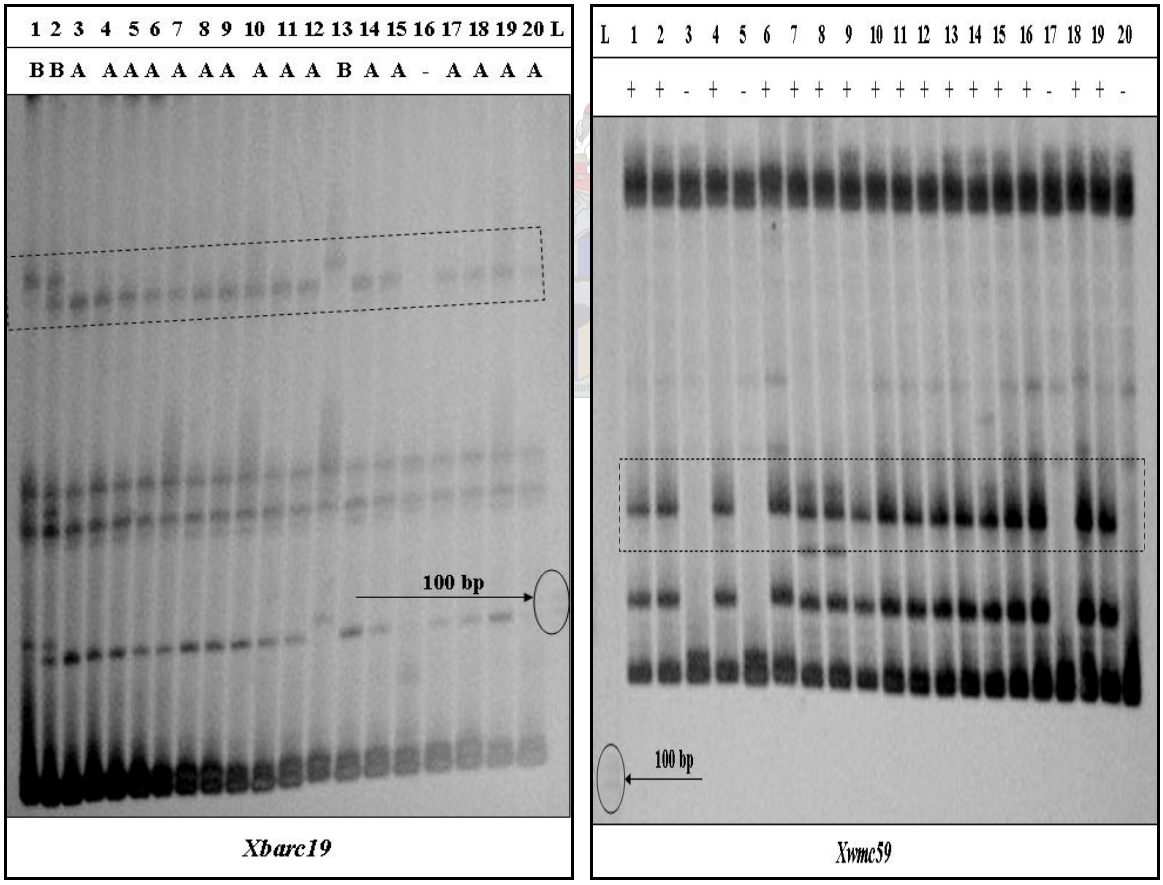


Figure 7-18 Gel photo illustrating the amplification of the *Xbarc19* and *Xwmc59* microsatellites (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1. L: 100 bp ladder.

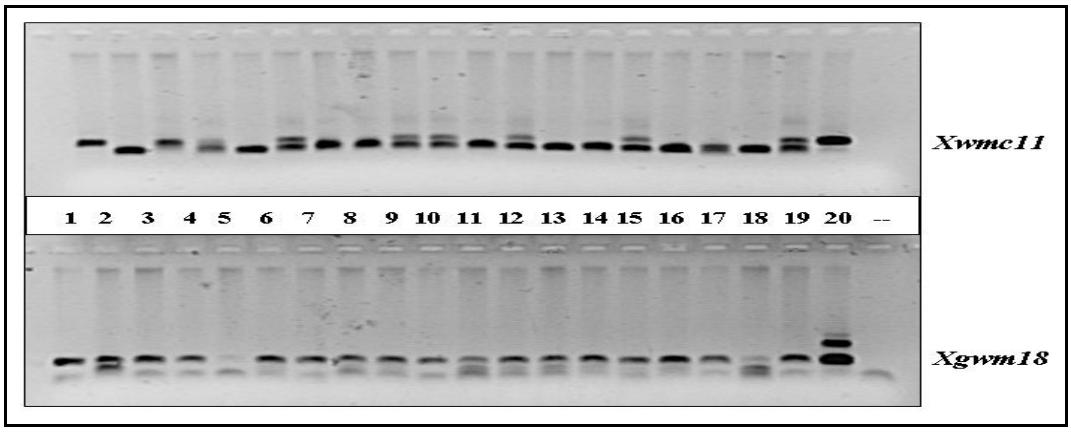


Figure 7-19 Gel photo confirming the amplification of the *Xwmc11* and *Xgwm18* microsatellites (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

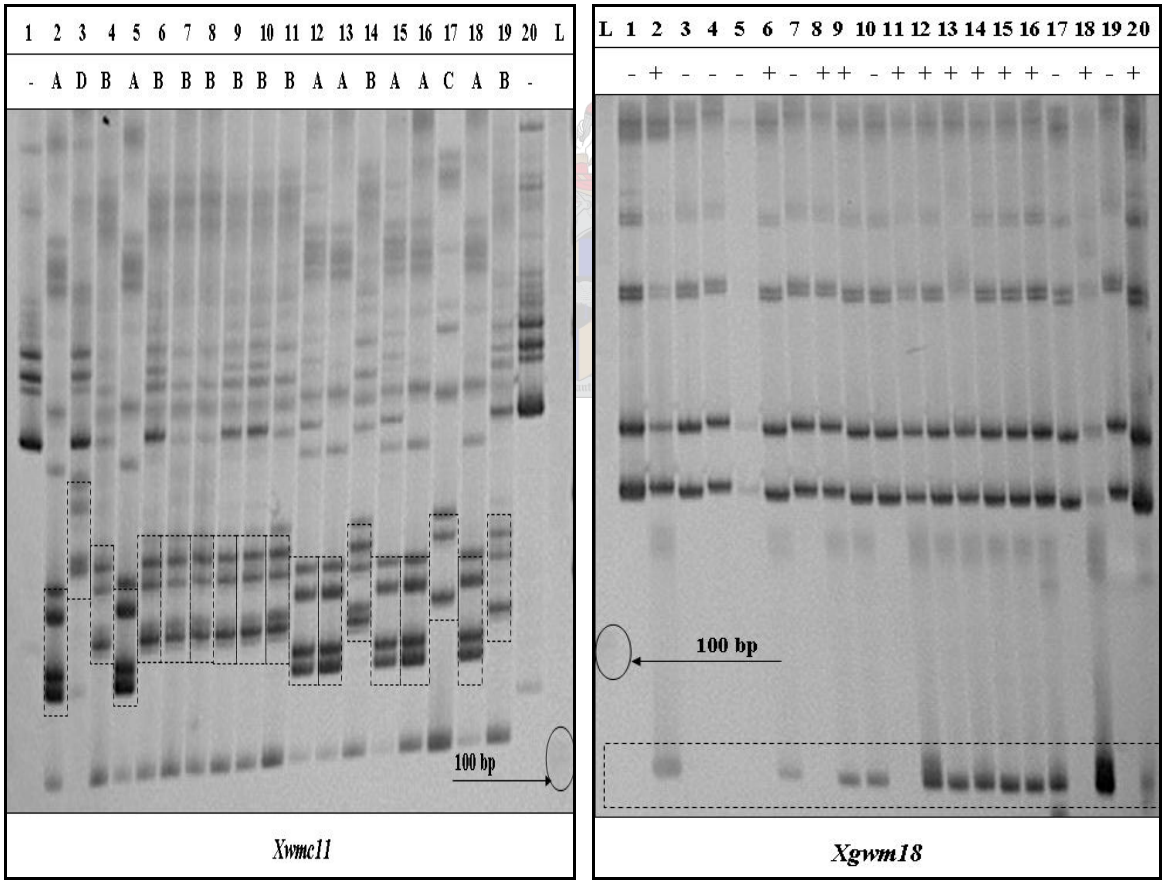


Figure 7-20 Gel photo illustrating the amplification of the *Xwmc11* and *Xgwm18* microsatellites (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1. L: 100 bp ladder.

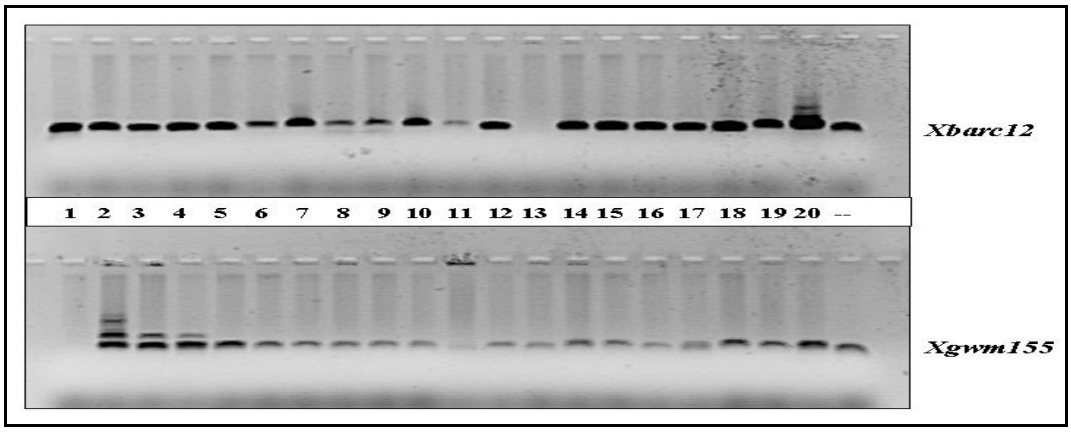


Figure 7-21 Gel photo confirming the amplification of the *Xbarc12* and *Xgwm155* microsatellites (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

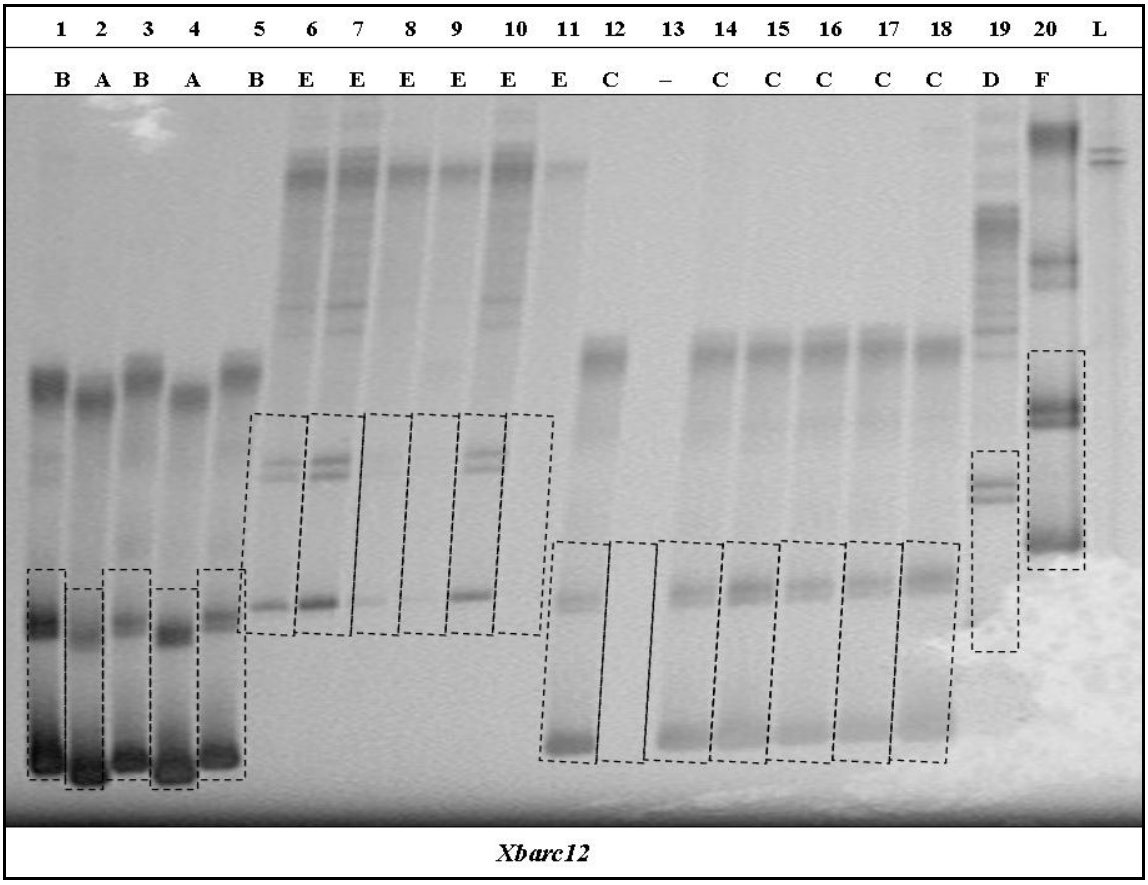


Figure 7-22 Gel photo illustrating the amplification of the *Xbarc12* microsatellite (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1. L: 100 bp ladder.

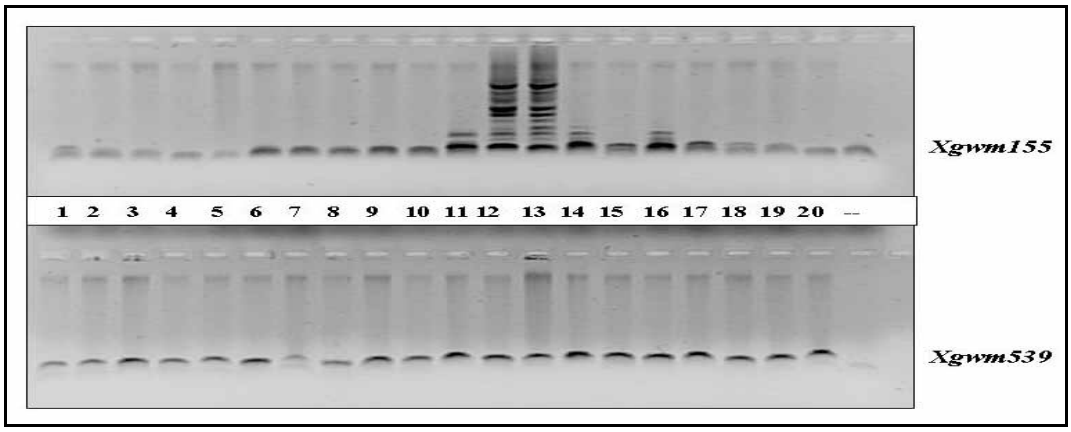


Figure 7-23 Gel photo confirming the amplification of the *Xgwm155* and *Xgwm539* microsatellites (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

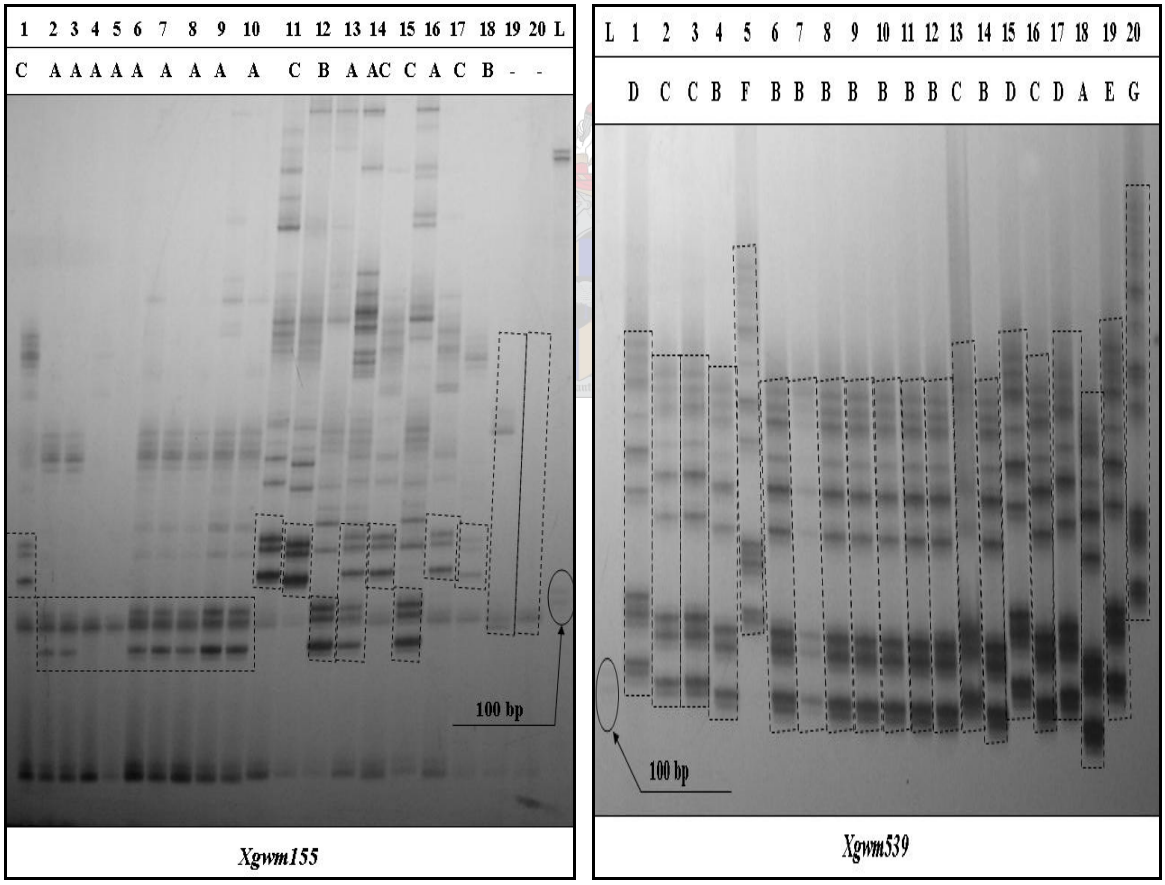


Figure 7-24 Gel photo illustrating the amplification of the *Xgwm155* and *Xgwm539* microsatellites (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1. L: 100 bp ladder.

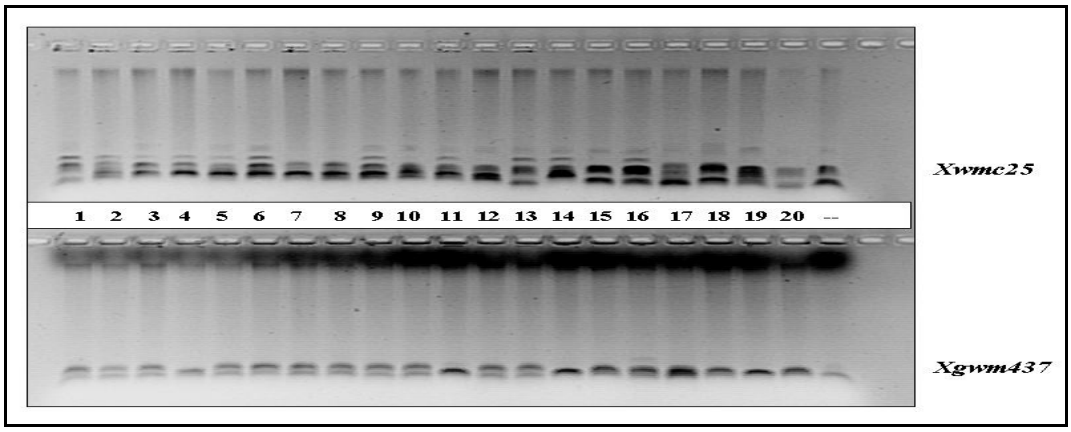


Figure 7-25 Gel photo confirming the amplification of the *Xwmc25* and *Xgwm437* microsatellites (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

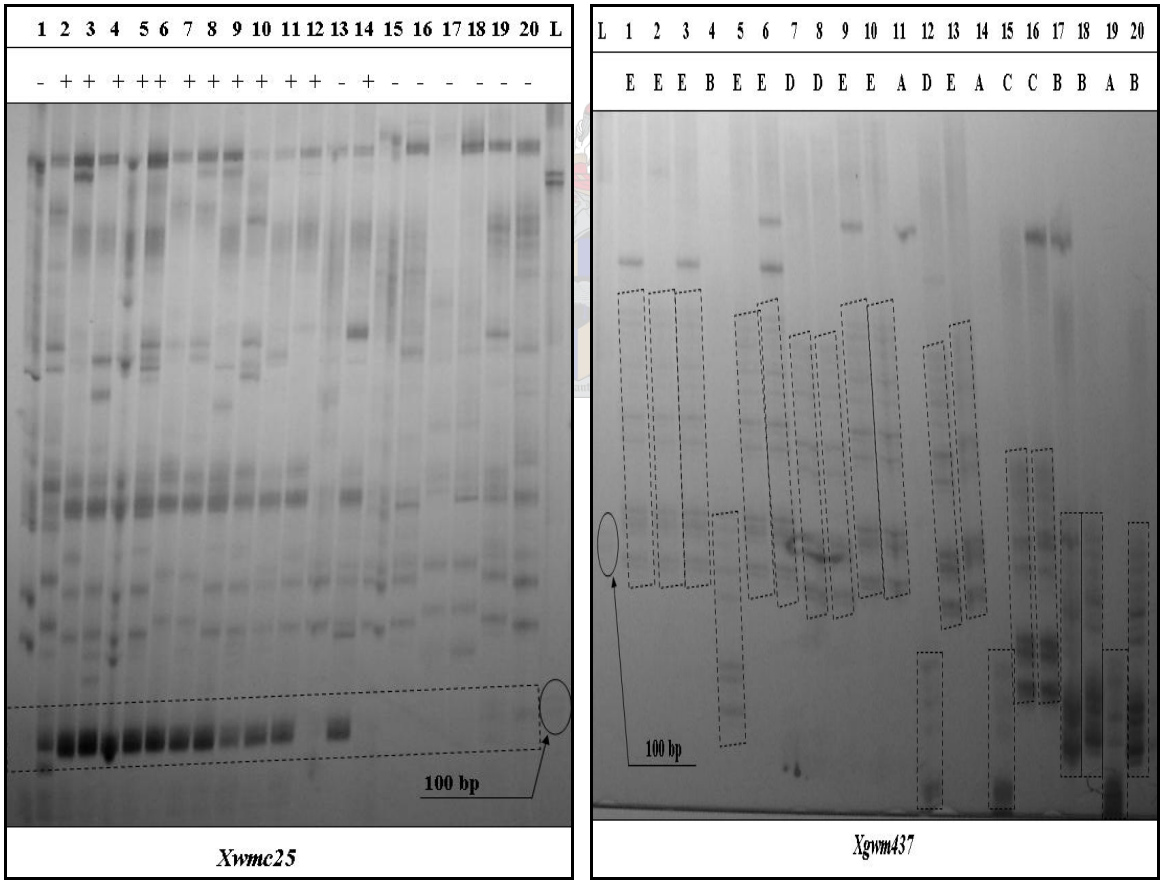


Figure 7-26 Gel photo illustrating the amplification of the *Xwmc25* and *Xgwm437* microsatellites (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1. L: 100 bp ladder.

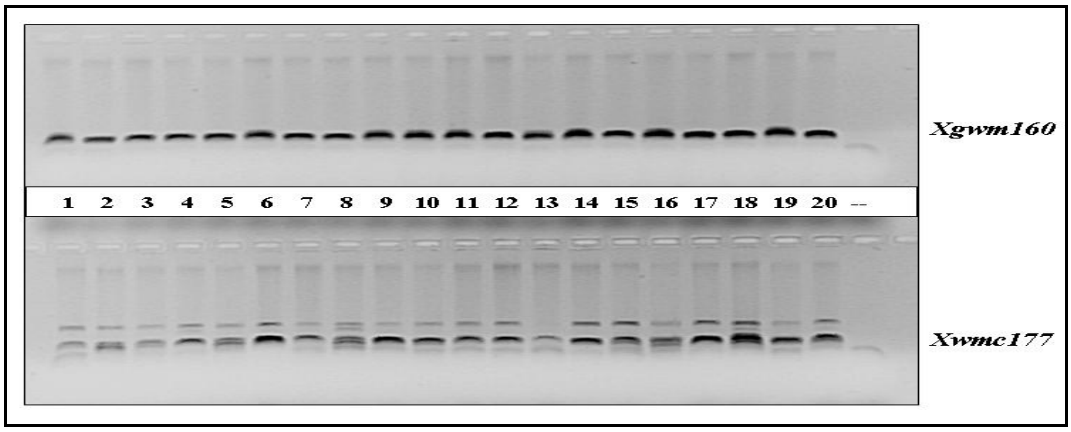


Figure 7-27 Gel photo confirming the amplification of the *Xgwm160* and *Xwmc177* microsatellites (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

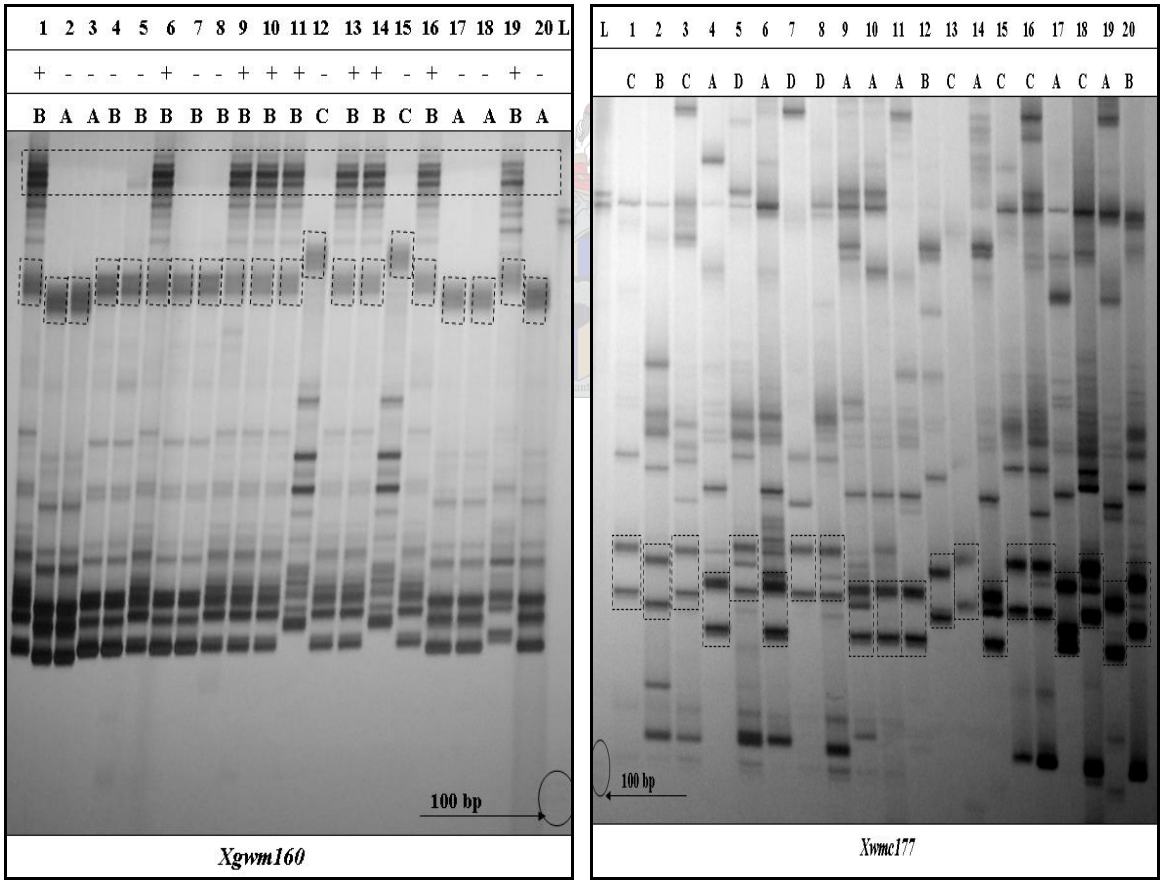


Figure 7-28 Gel photo illustrating the amplification of the *Xgwm160* and *Xwmc177* microsatellites (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1. L: 100 bp ladder.

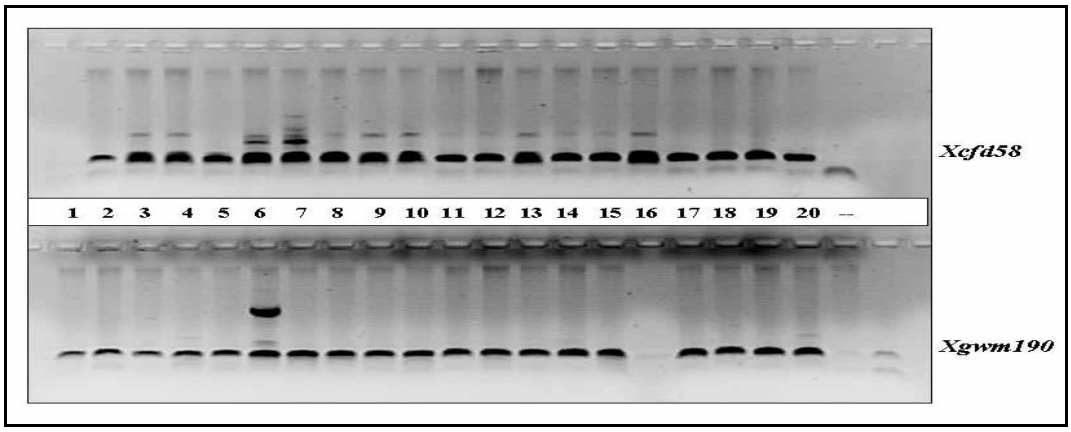


Figure 7-29 Gel photo confirming the amplification of the *Xcfd58* and *Xgwm190* microsatellites (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

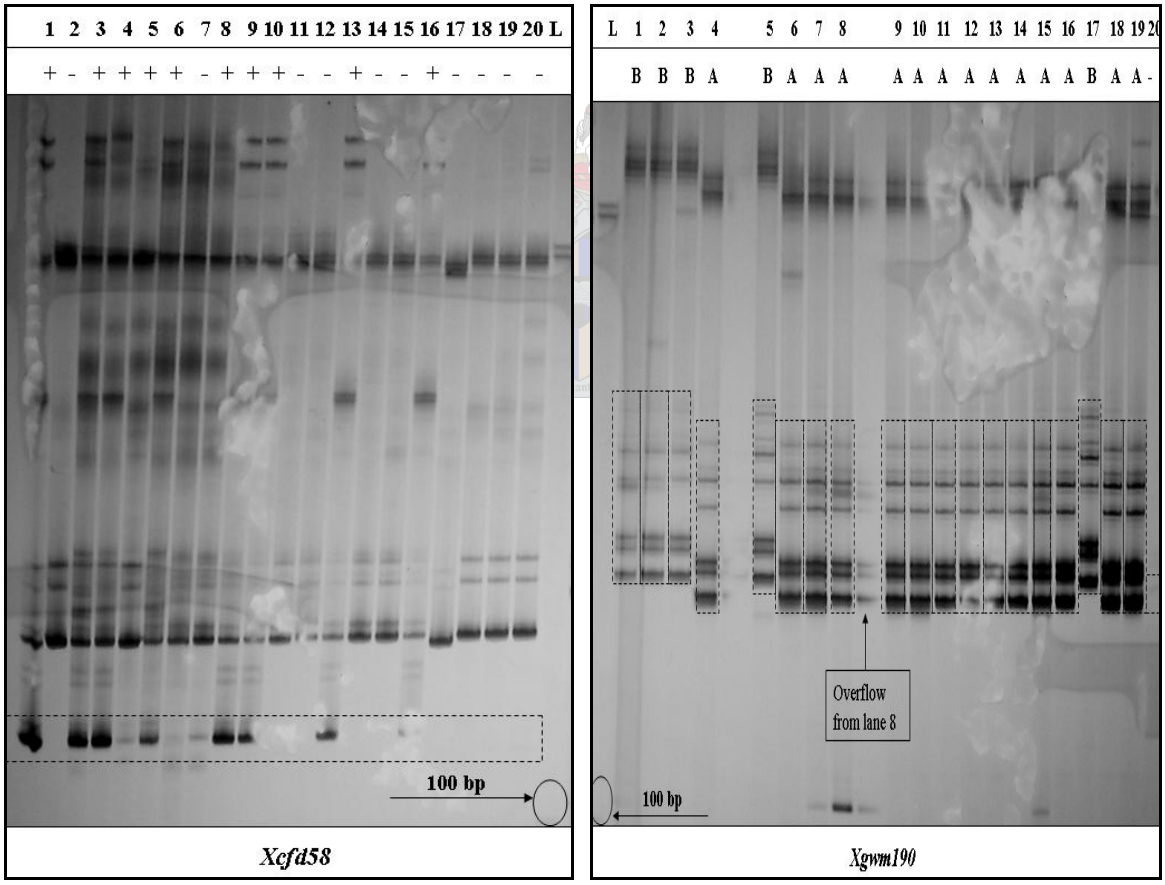


Figure 7-30 Gel photo illustrating the amplification of the *Xcfd58* and *Xgwm190* microsatellites (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1. L: 100 bp ladder.

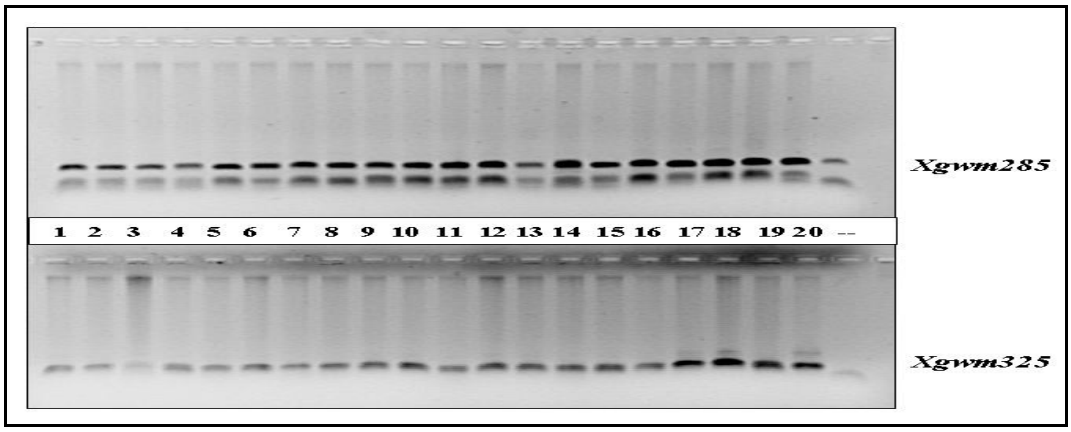


Figure 7-31 Gel photo confirming the amplification of the *Xgwm285* and *Xgwm325* microsatellites (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

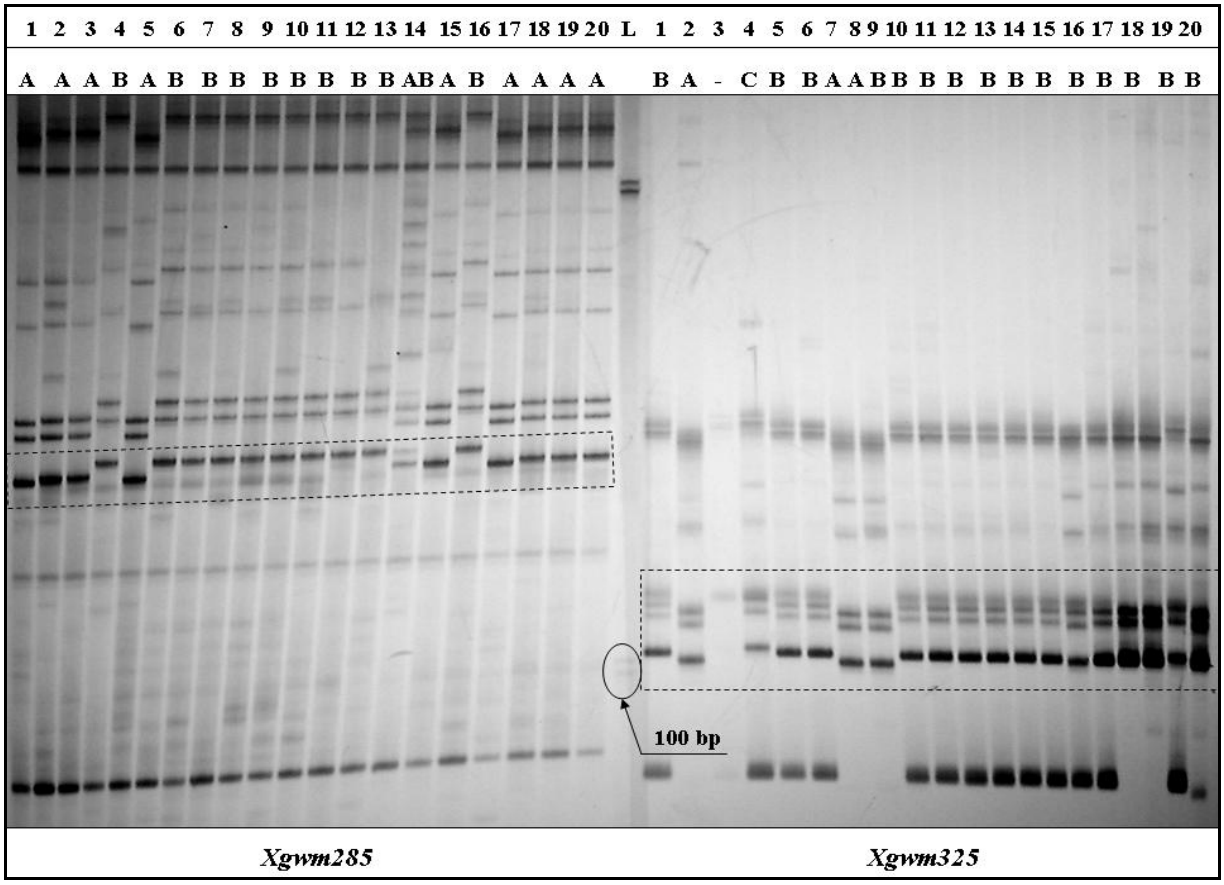


Figure 7-32 Gel photo illustrating the amplification of the *Xgwm285* and *Xgwm325* microsatellites (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1. L: 100 bp ladder.

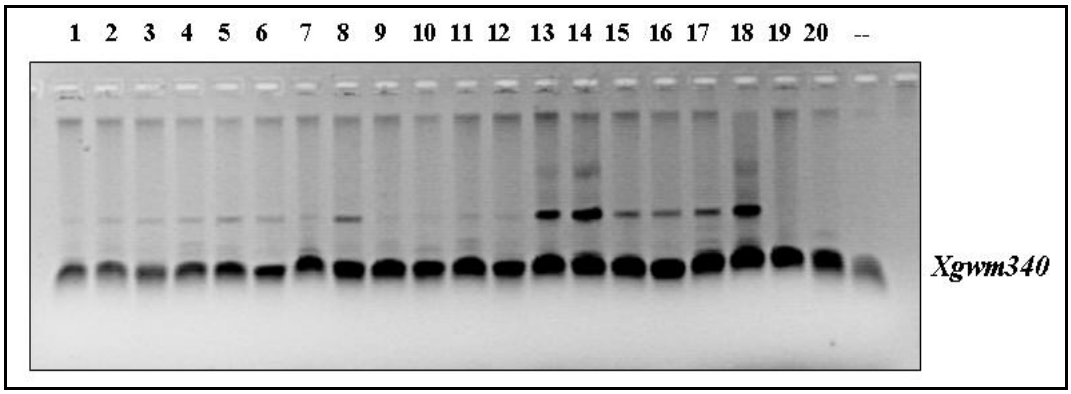


Figure 7-33 Gel photo confirming the amplification of the *Xgwm340* microsatellites (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

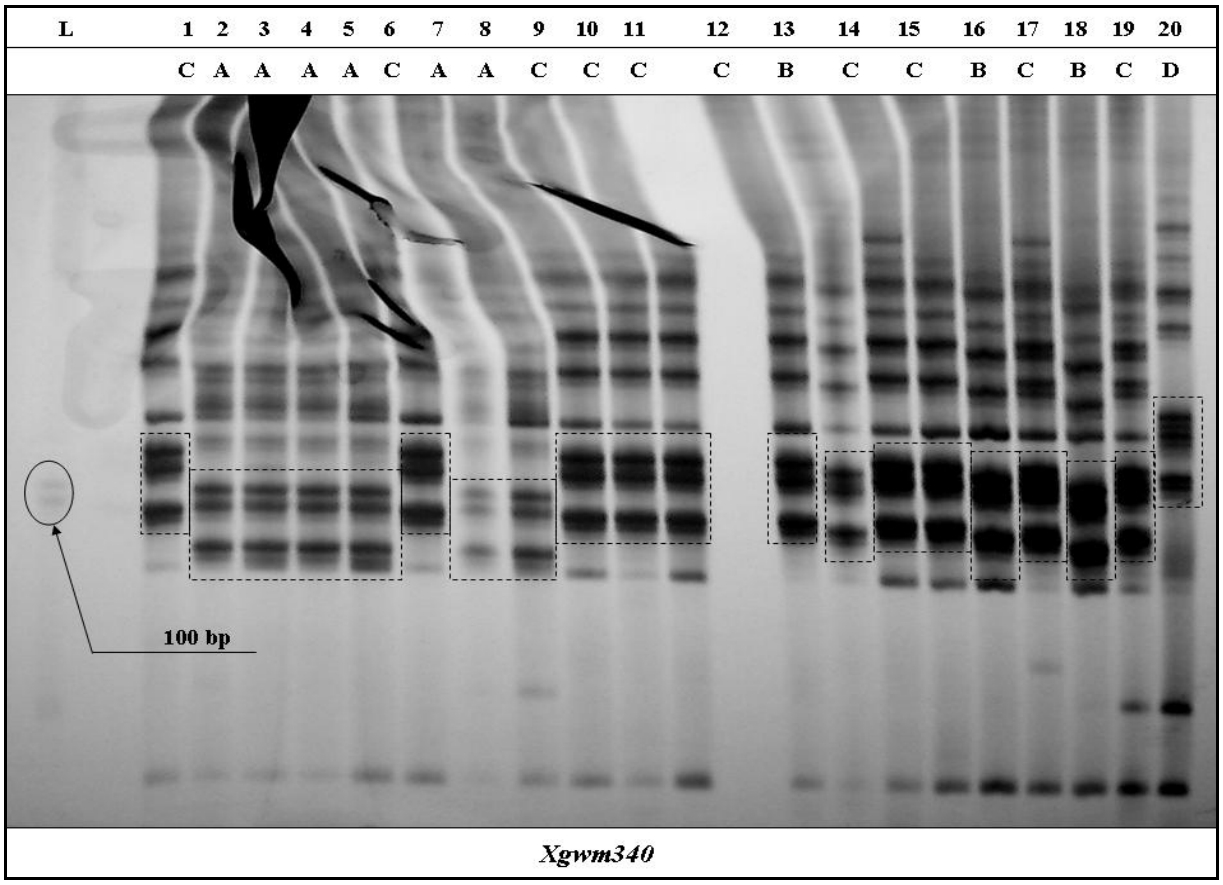


Figure 7-34 Gel photo illustrating the amplification of the *Xgwm340* microsatellite (6% denaturing polyacrylamide gel electrophoresed at 70W for 5-6 hours). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1. L: 100 bp ladder.

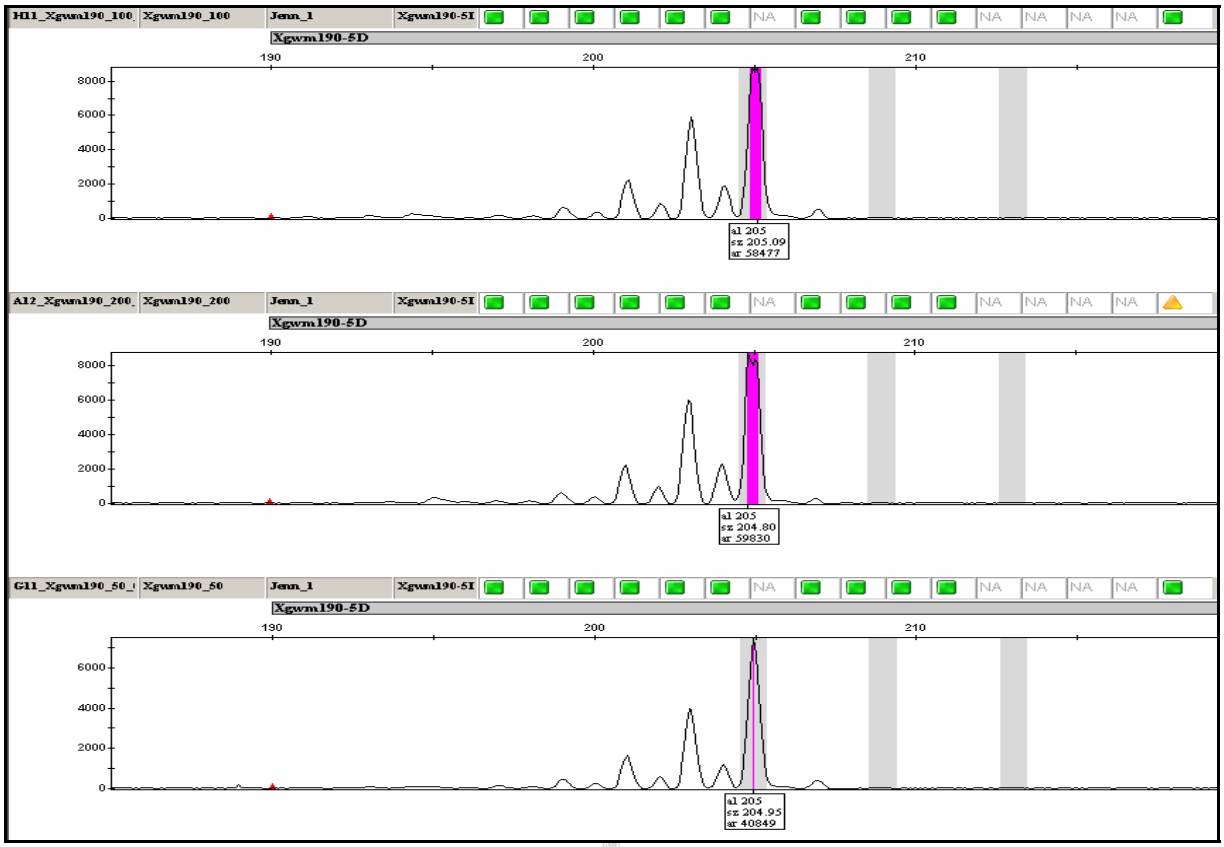


Figure 7-35 Template DNA concentration range (100 ng, 200 ng, 50 ng) using primer *Xgwm190*.

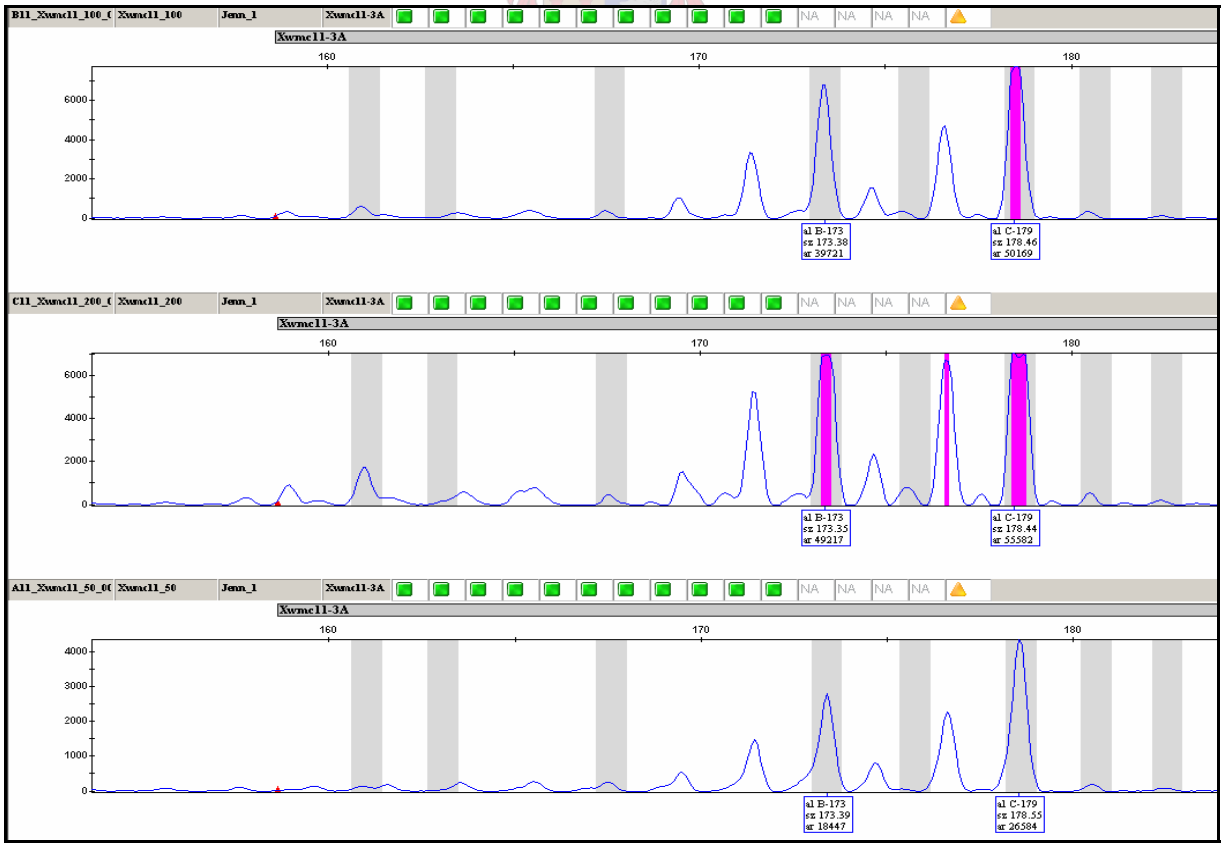


Figure 7-36 Template DNA concentration range (100 ng, 200 ng, 50 ng) using primer *Xwmc11*.

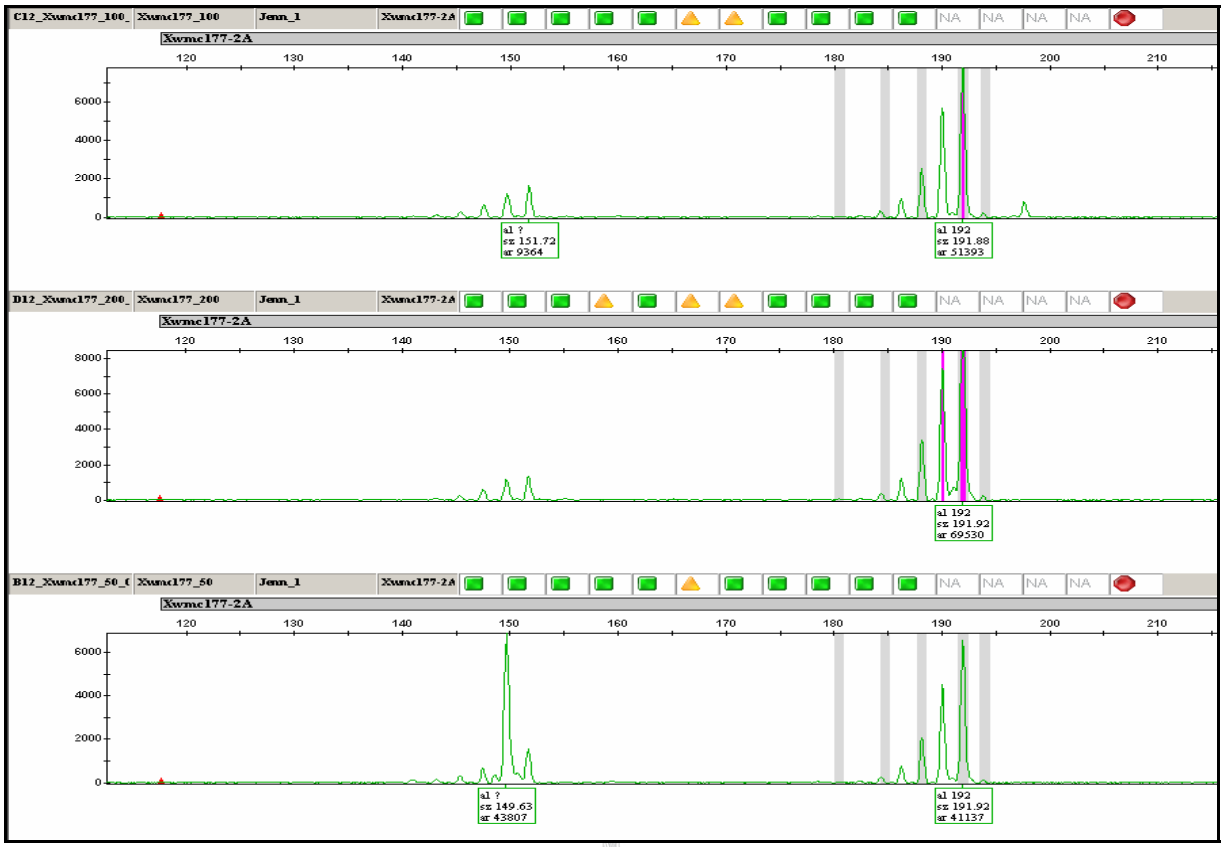


Figure 7-37 Template DNA concentration range (100 ng, 200 ng, 50 ng) using primer *Xwmc177*.

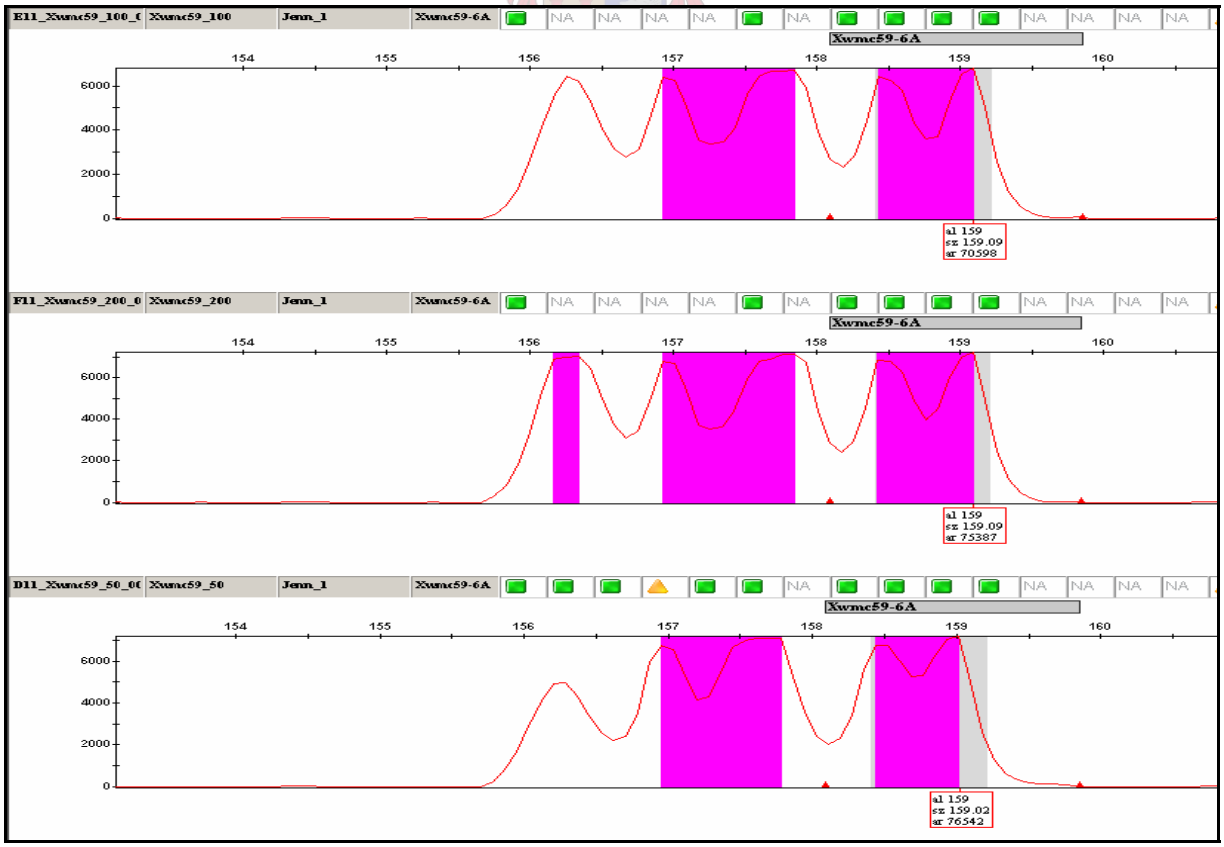


Figure 7-38 Template DNA concentration range (100 ng, 200 ng, 50 ng) using primer *Xwmc59*.

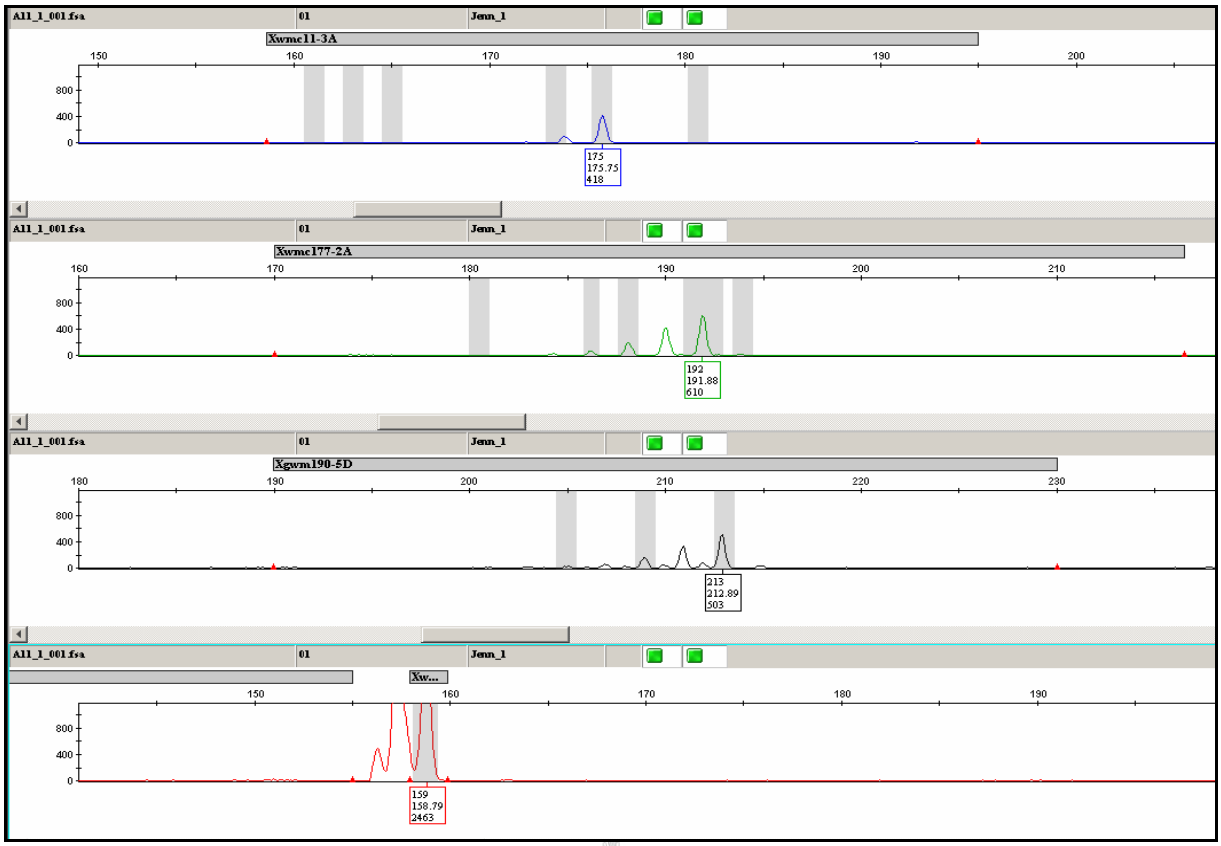


Figure 7-39 Sample number 1. (*Xwmc11*, *Xwmc177*, *Xgwm190*, *Xwmc59*)

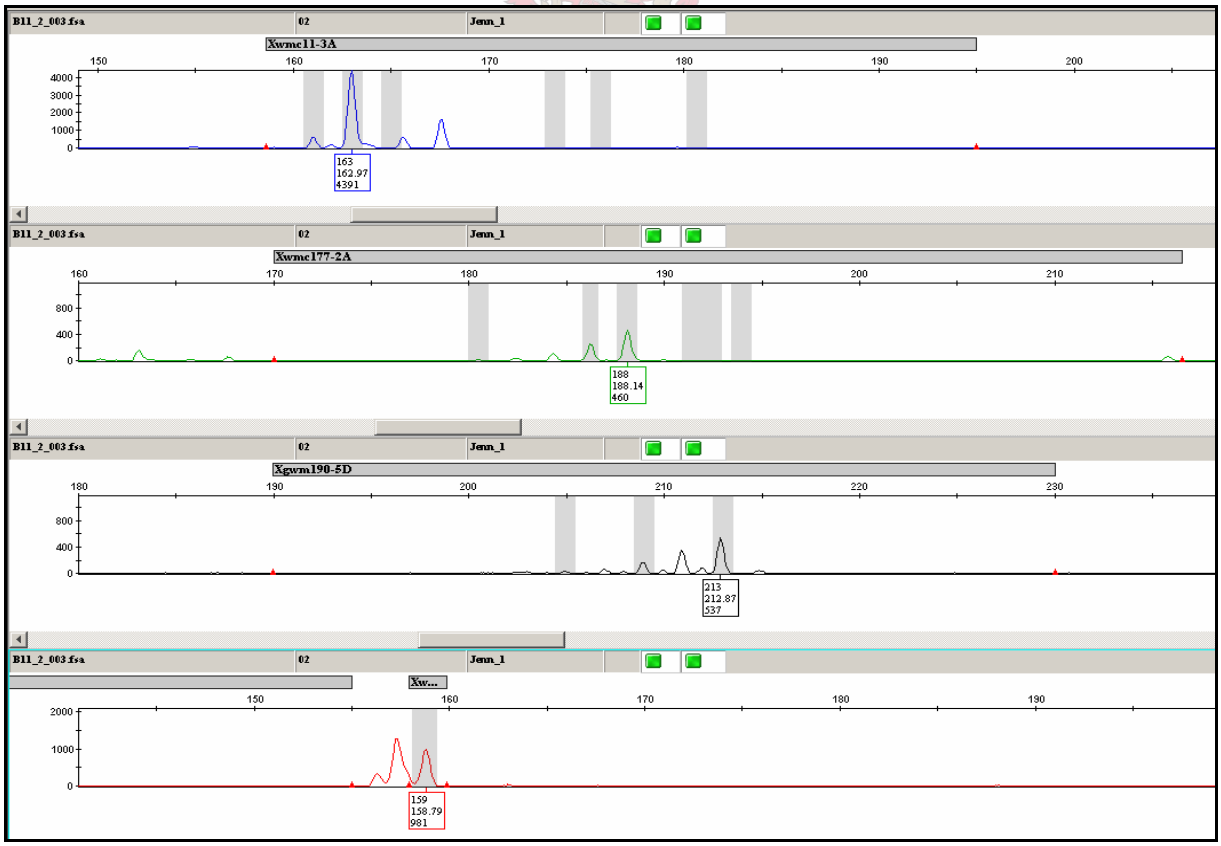


Figure 7-40 Sample number 2. (*Xwmc11*, *Xwmc177*, *Xgwm190*, *Xwmc59*)

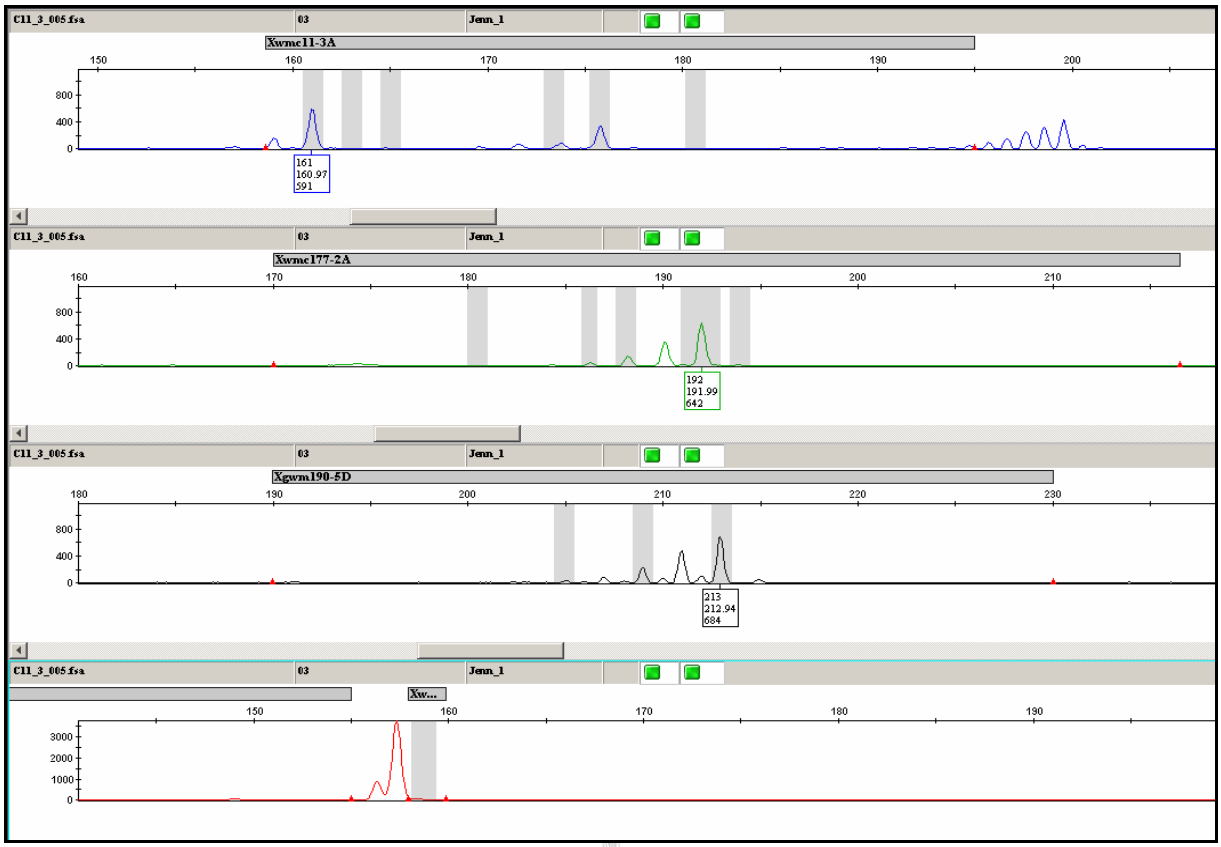


Figure 7-41 Sample number 3. (*Xwmc11*, *Xwmc177*, *Xgwm190*, *Xwmc59*)

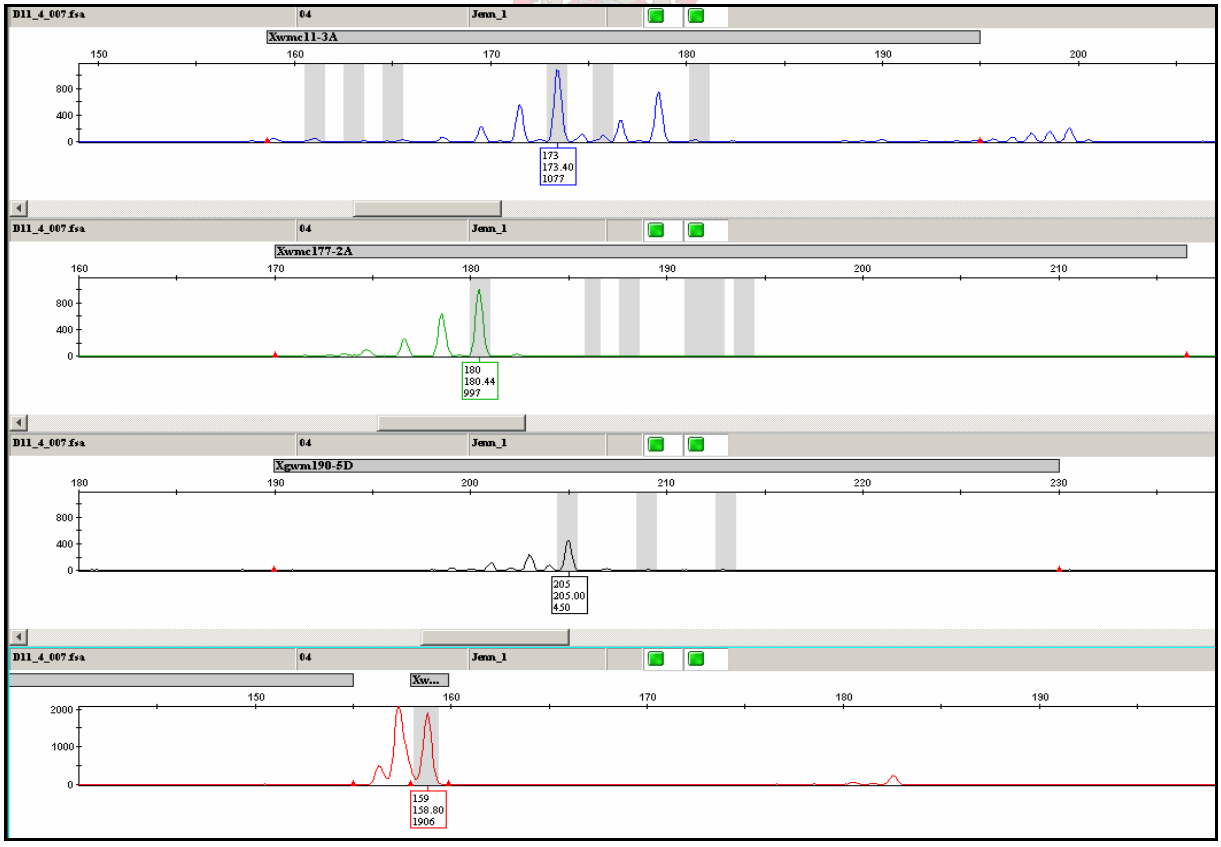


Figure 7-42 Sample number 4. (*Xwmc11*, *Xwmc177*, *Xgwm190*, *Xwmc59*)

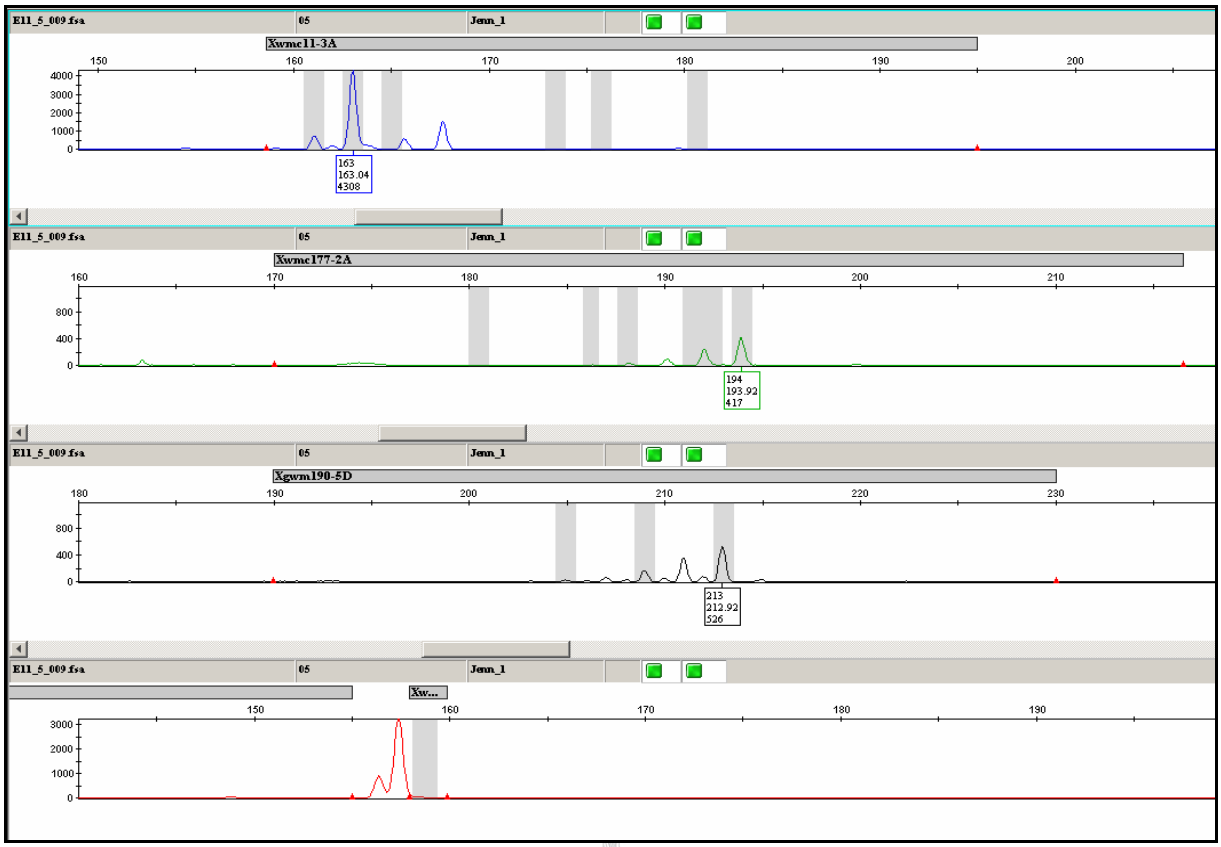


Figure 7-43 Sample number 5. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

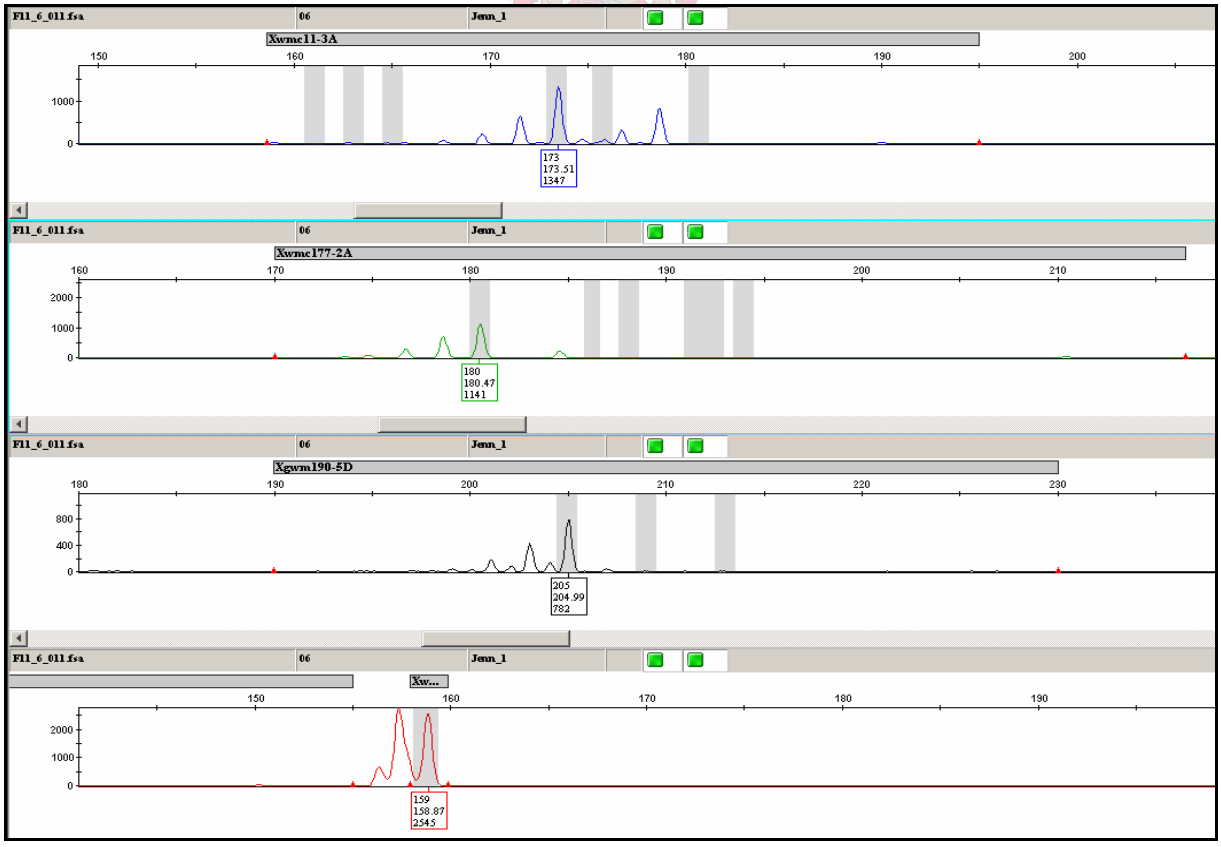


Figure 7-44 Sample number 6. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

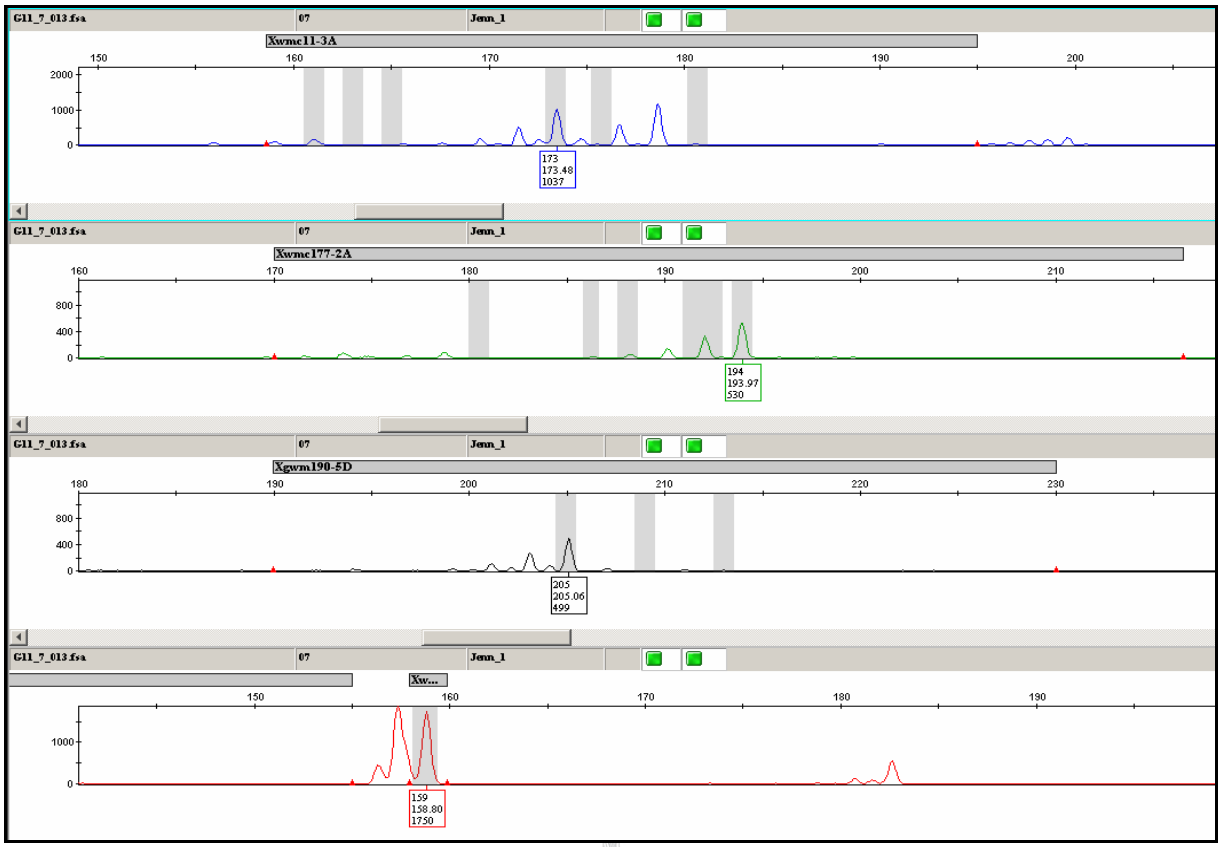


Figure 7-45 Sample number 7. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

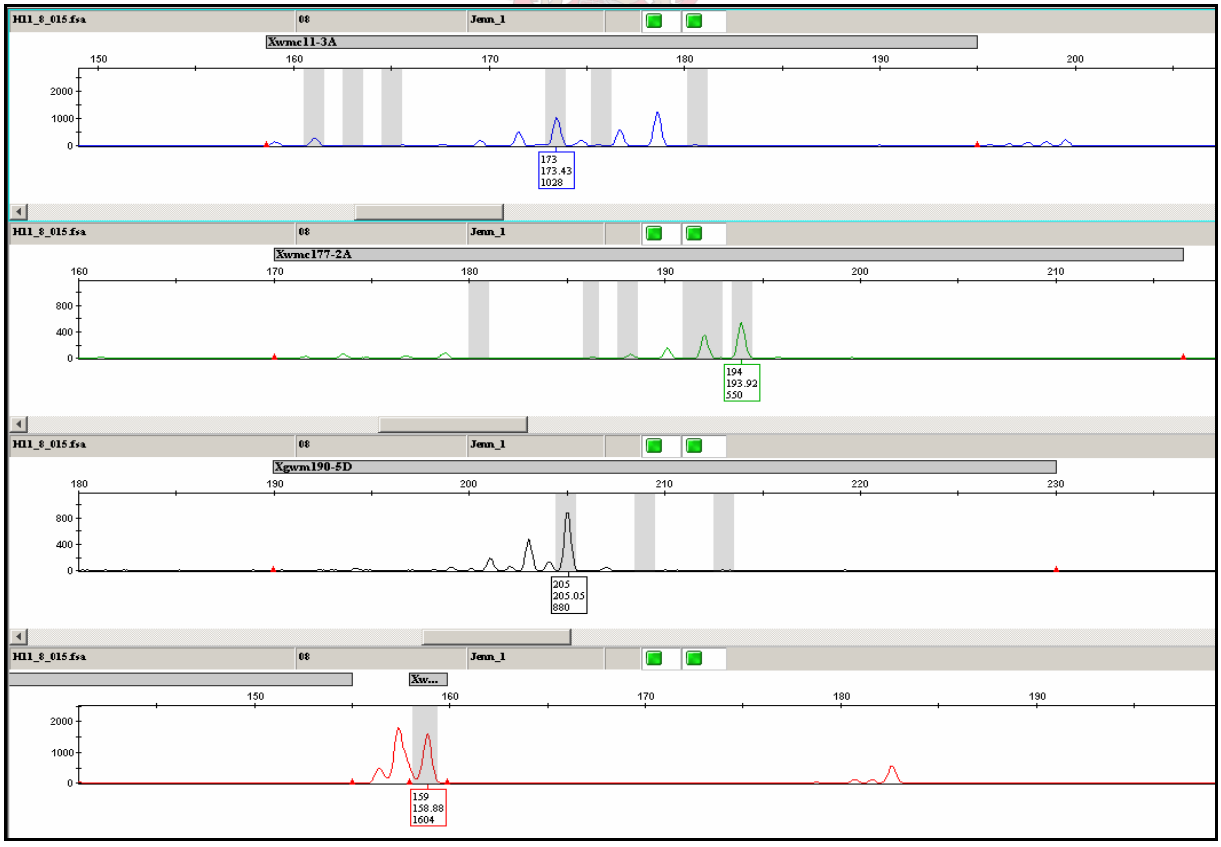


Figure 7-46 Sample number 8. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

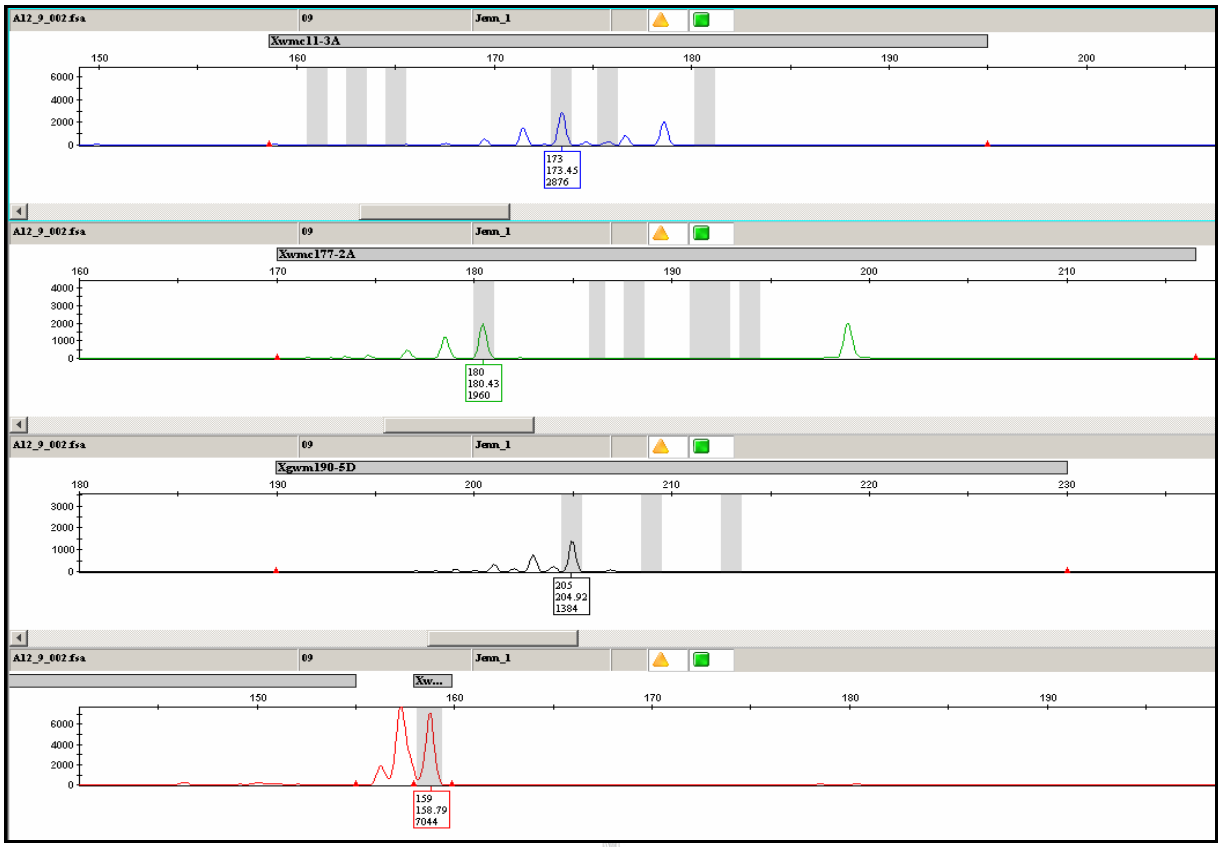


Figure 7-47 Sample number 9. (*Xwmc11*, *Xwmc177*, *Xgwm190*, *Xwmc59*)

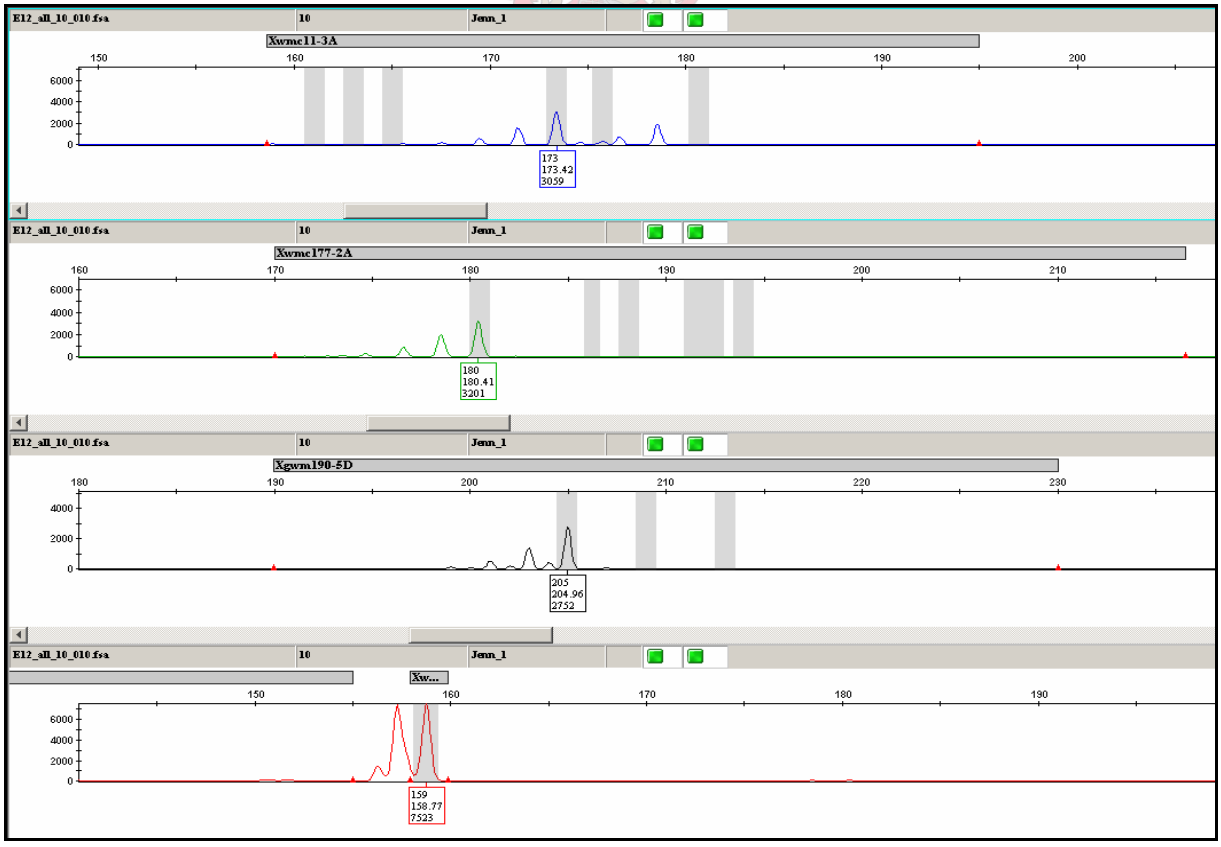


Figure 7-48 Sample number 10. (*Xwmc11*, *Xwmc177*, *Xgwm190*, *Xwmc59*)

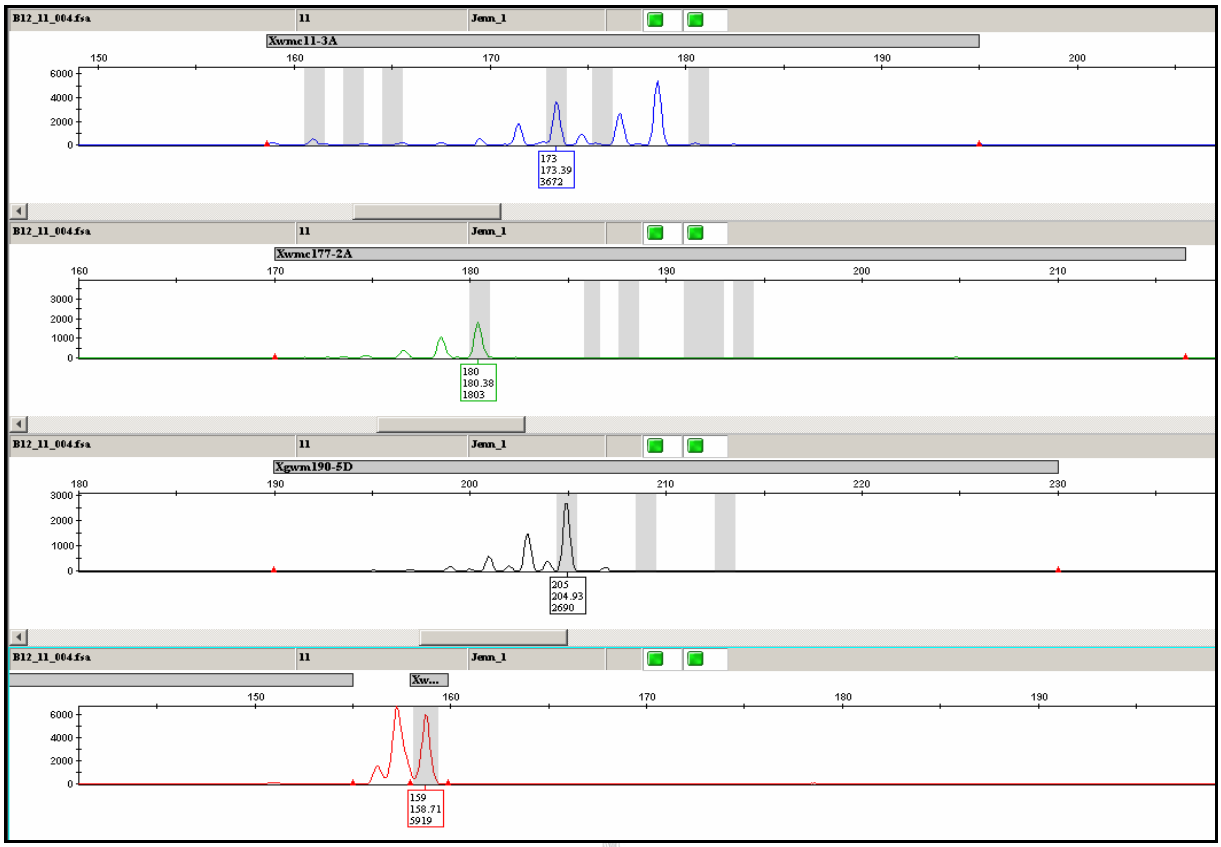


Figure 7-49 Sample number 11. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

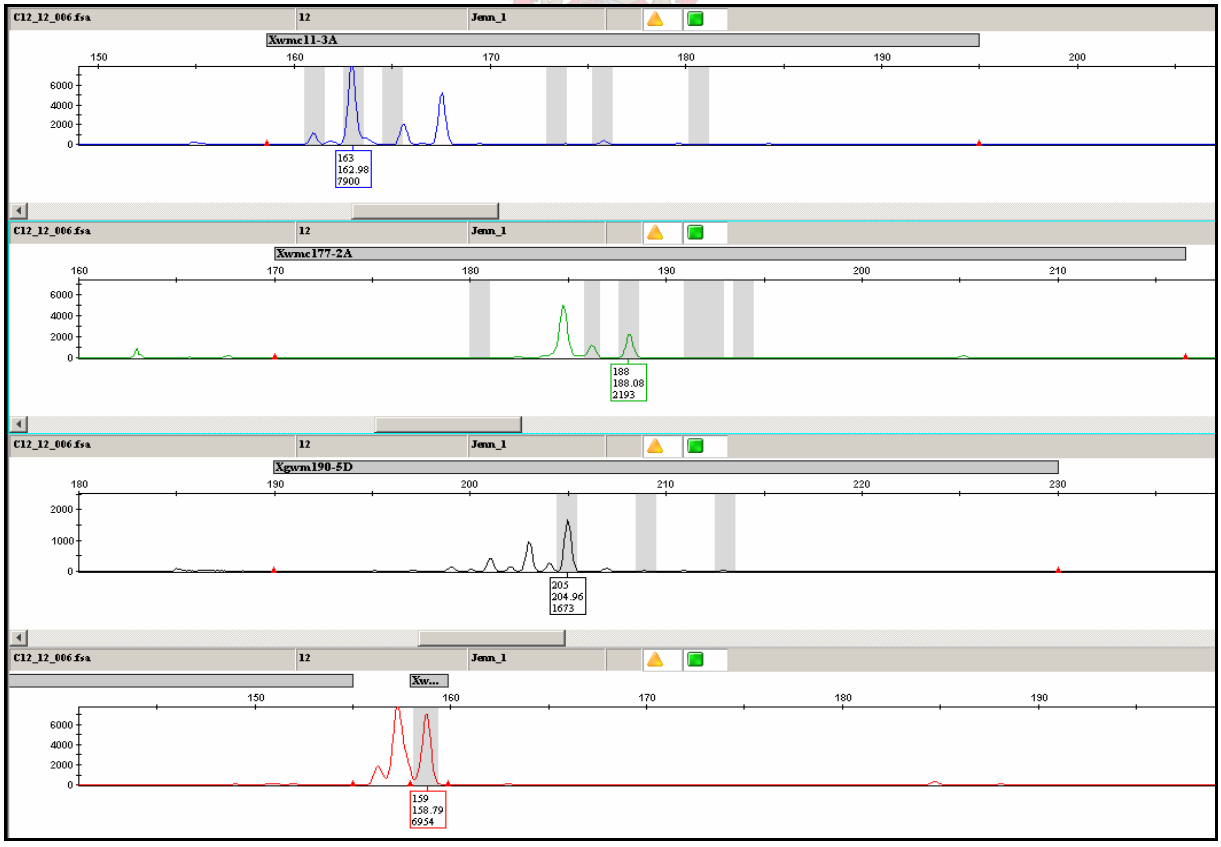


Figure 7-50 Sample number 12. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

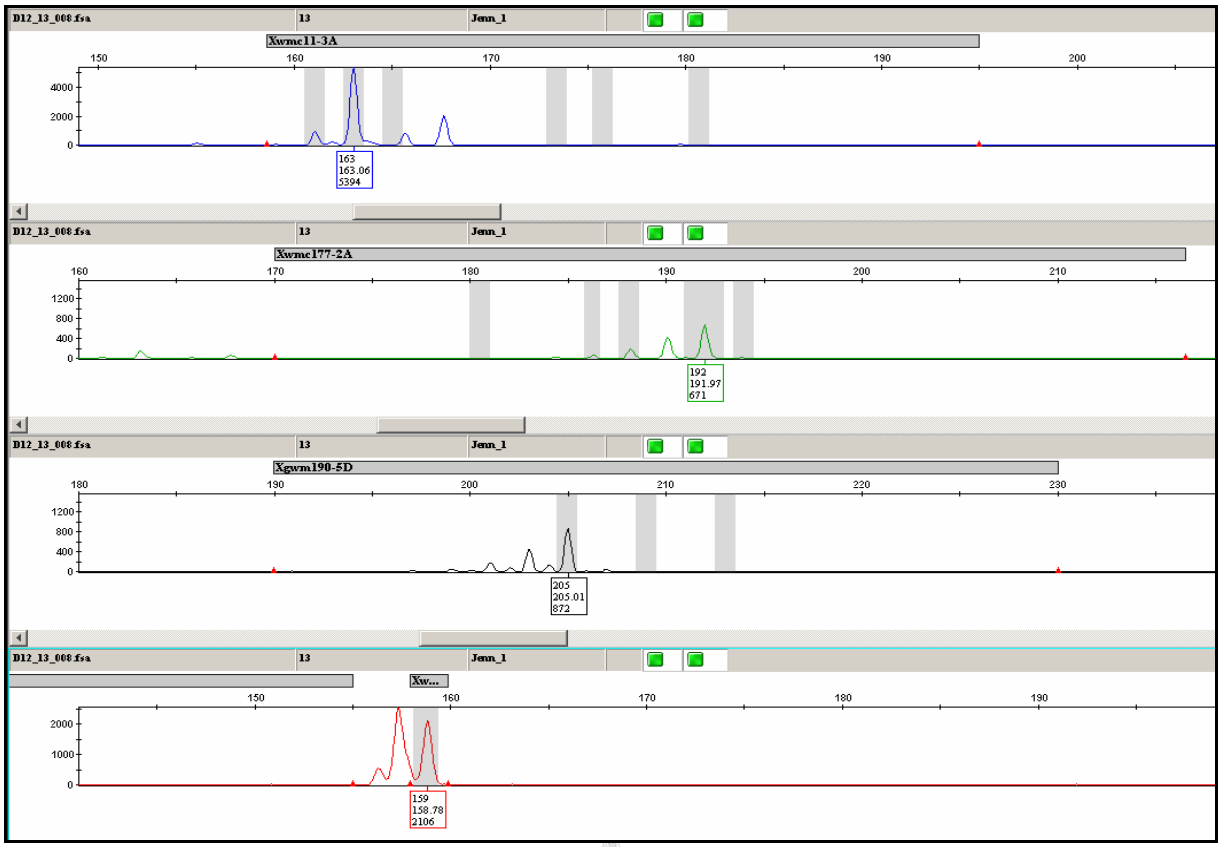


Figure 7-51 Sample number 13. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

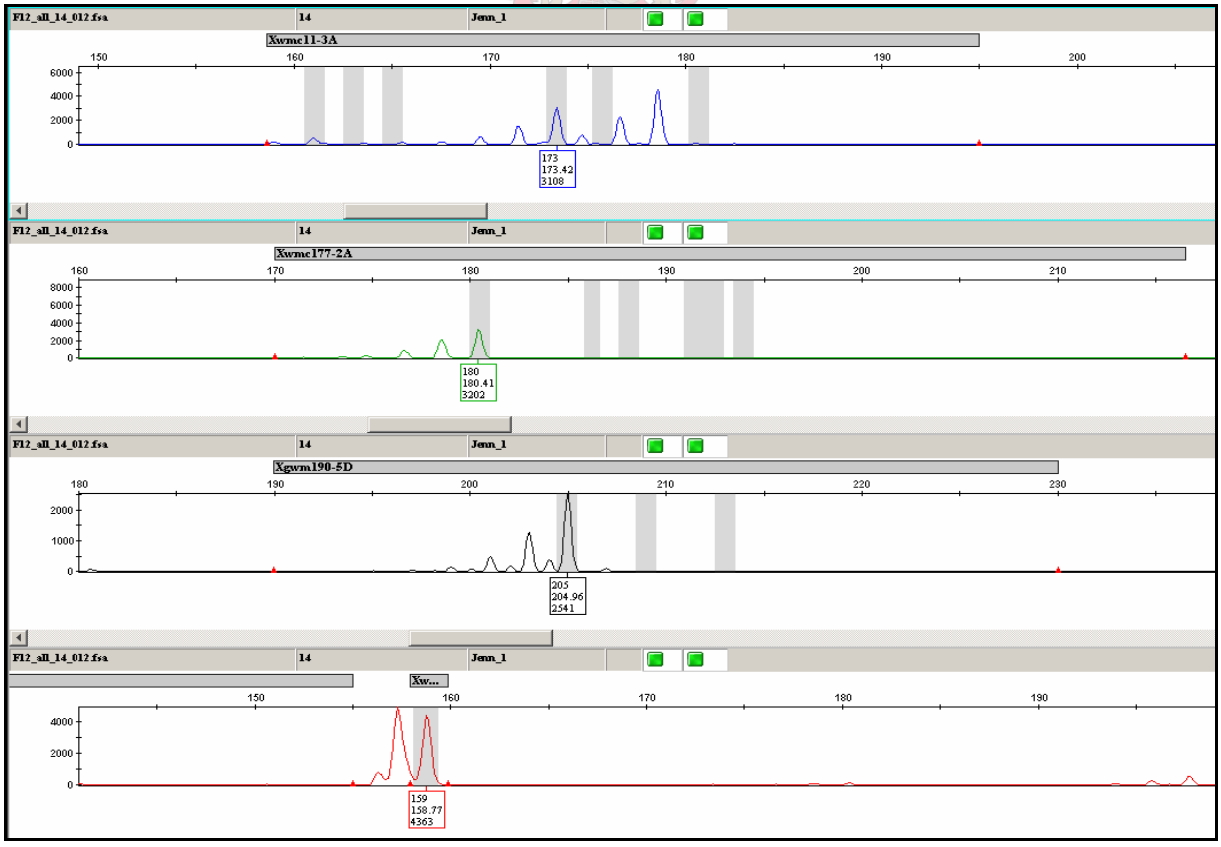


Figure 7-52 Sample number 14. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

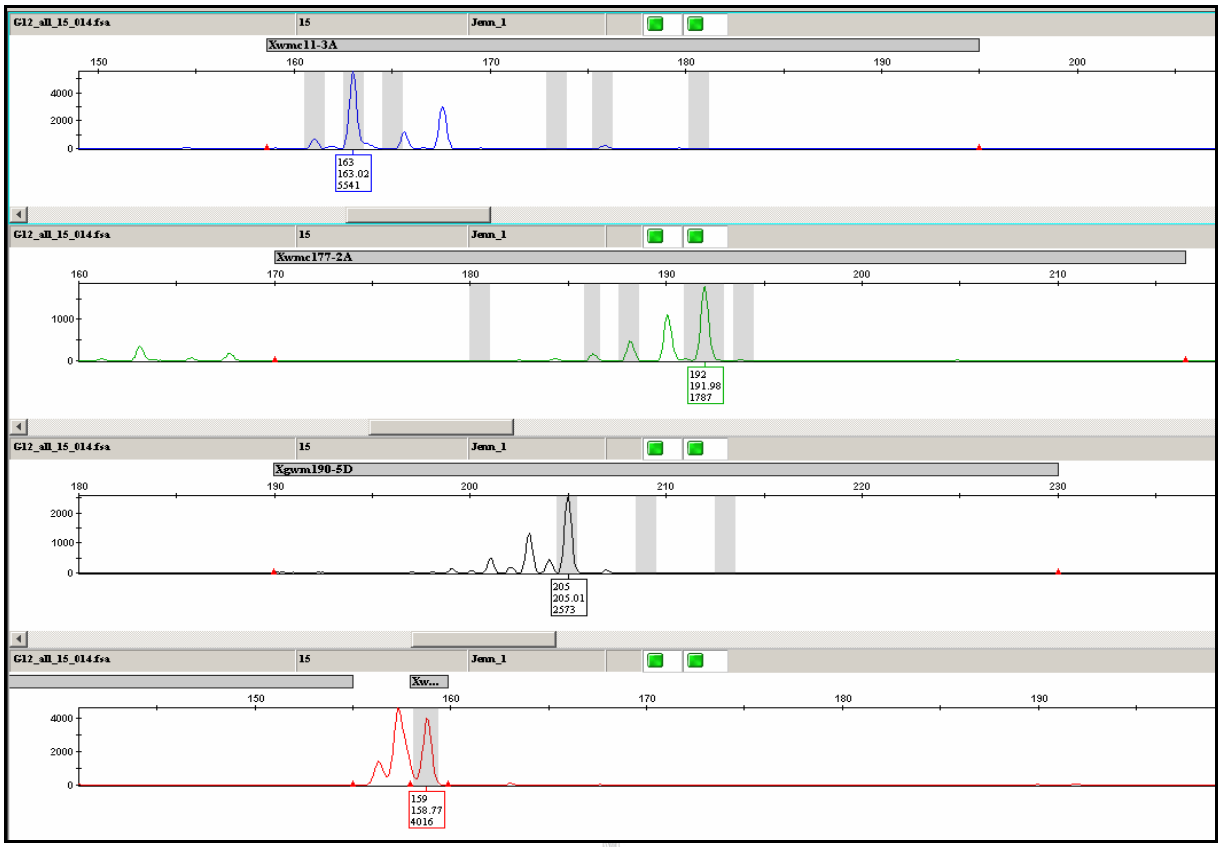


Figure 7-53 Sample number 15. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

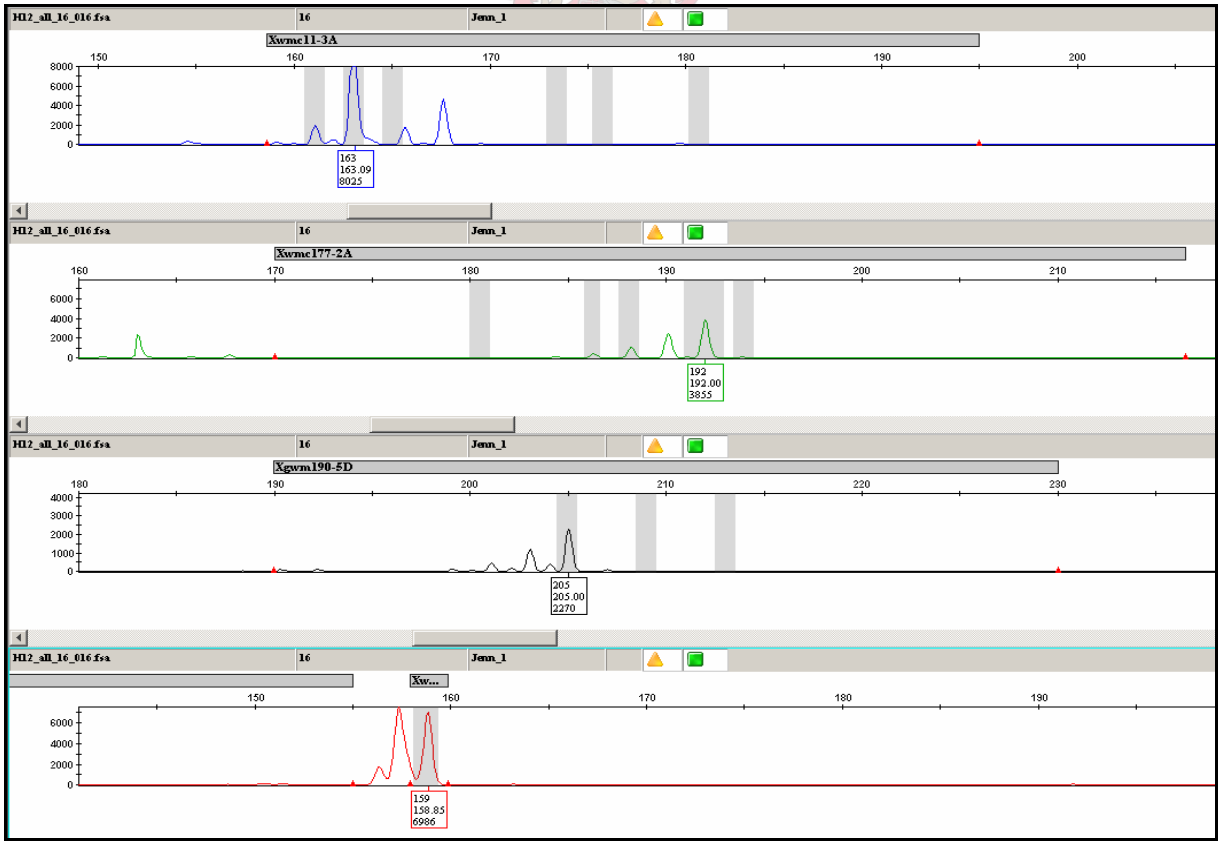


Figure 7-54 Sample number 16. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

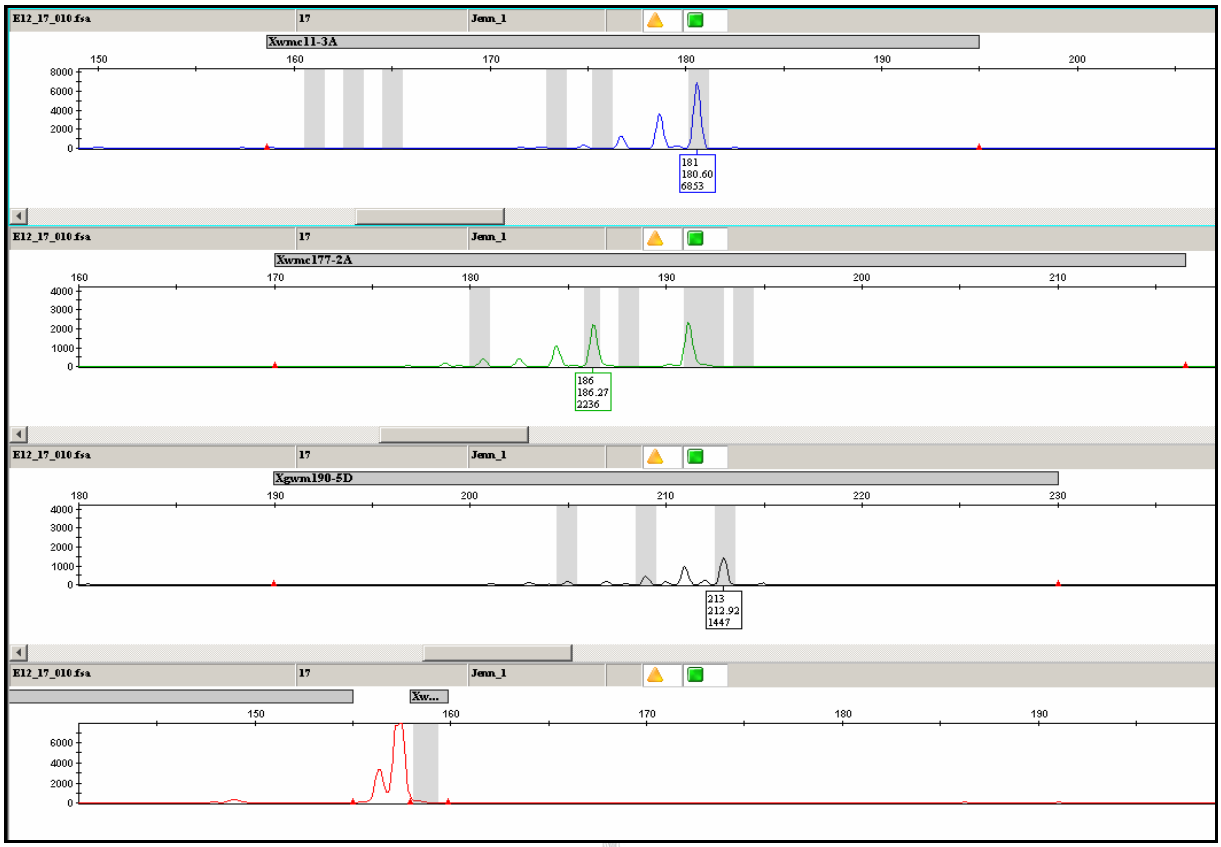


Figure 7-55 Sample number 17. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

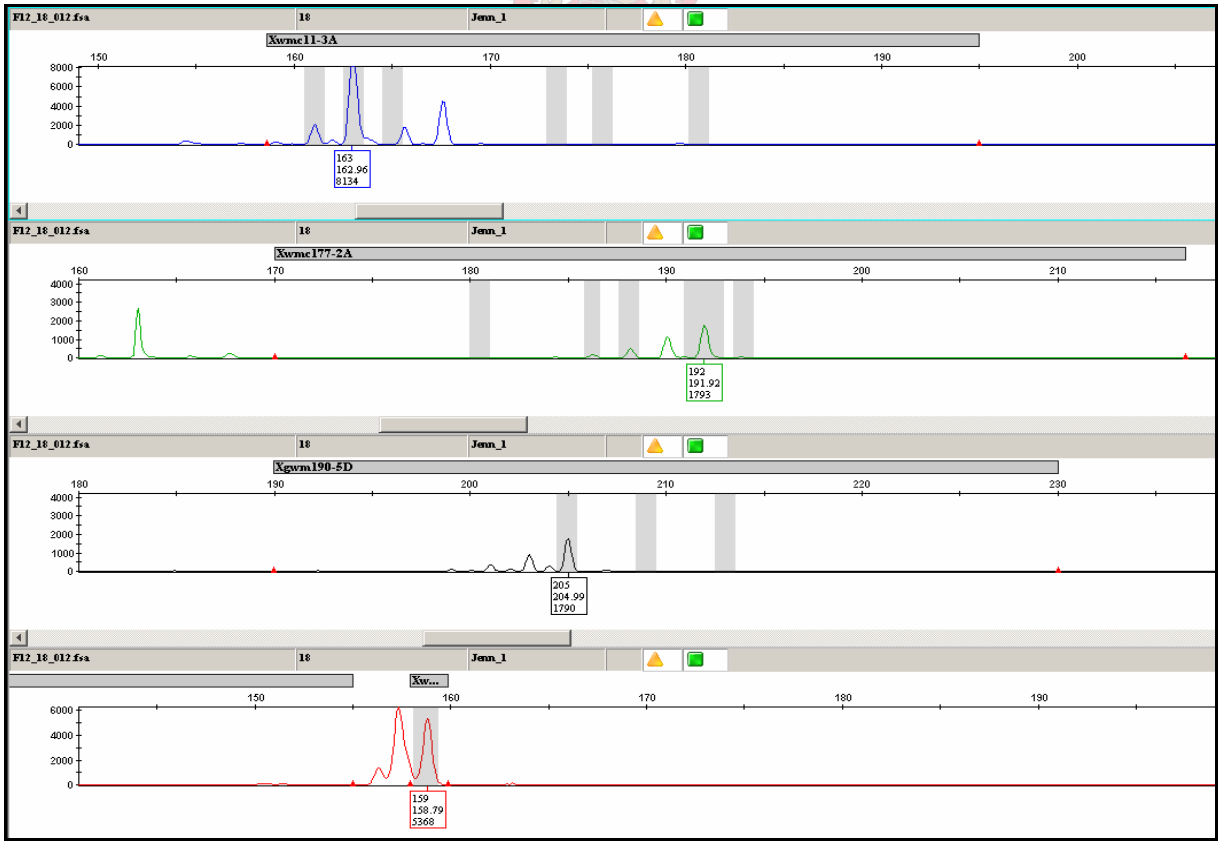


Figure 7-56 Sample number 18. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

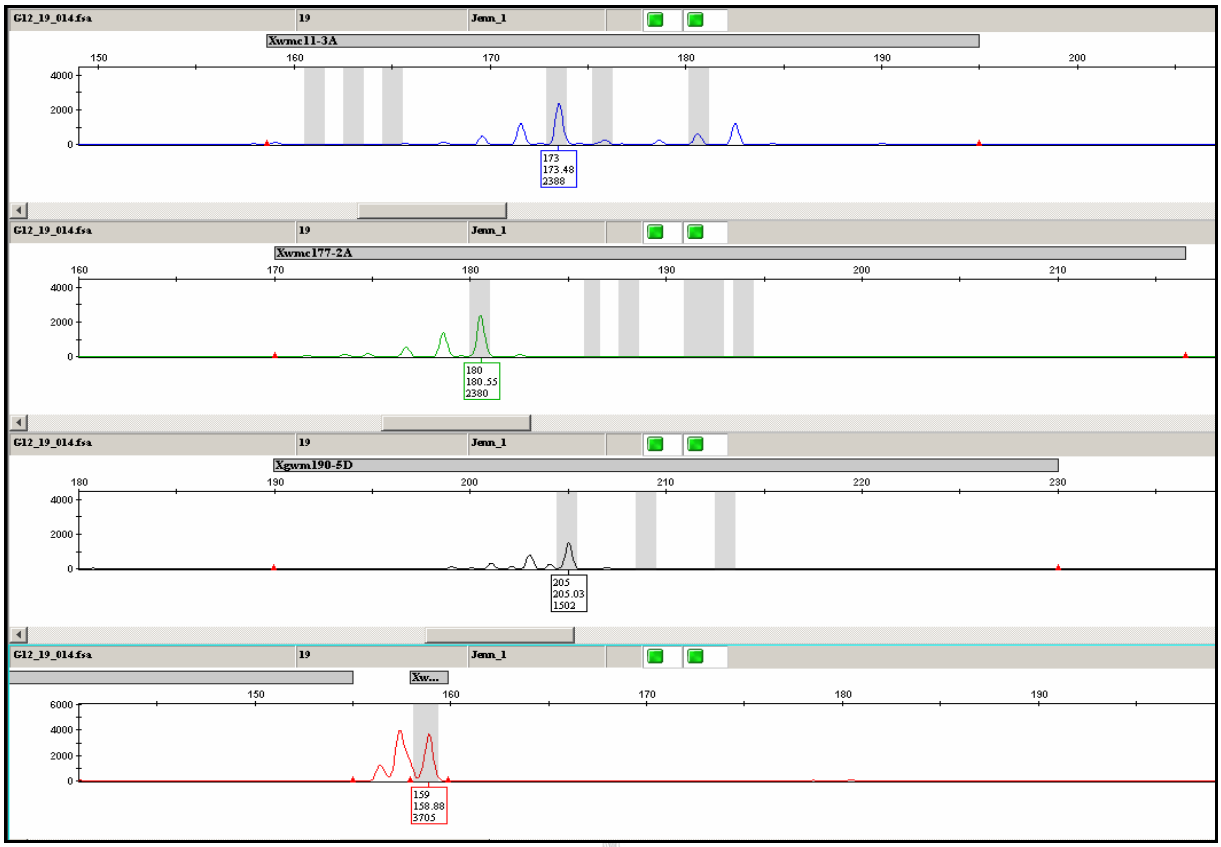


Figure 7-57 Sample number 19. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

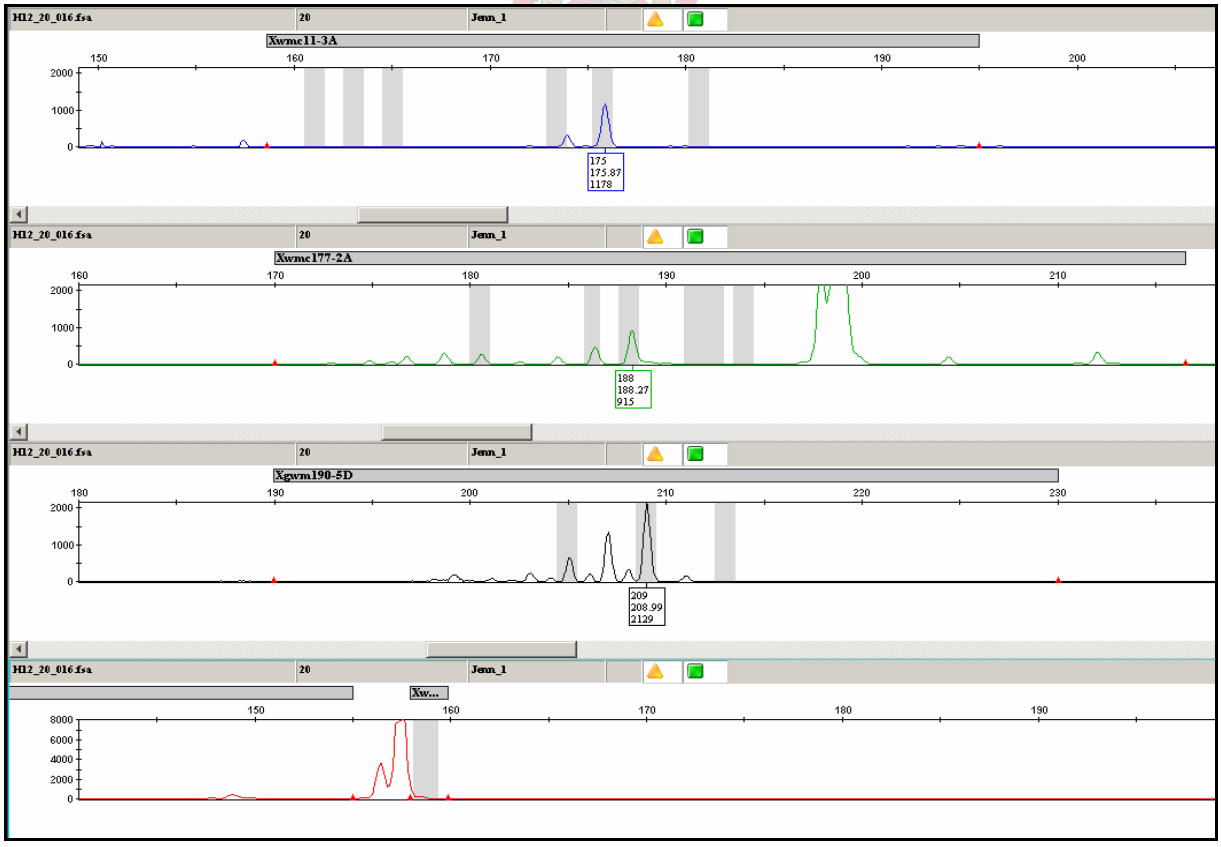


Figure 7-58 Sample number 20. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

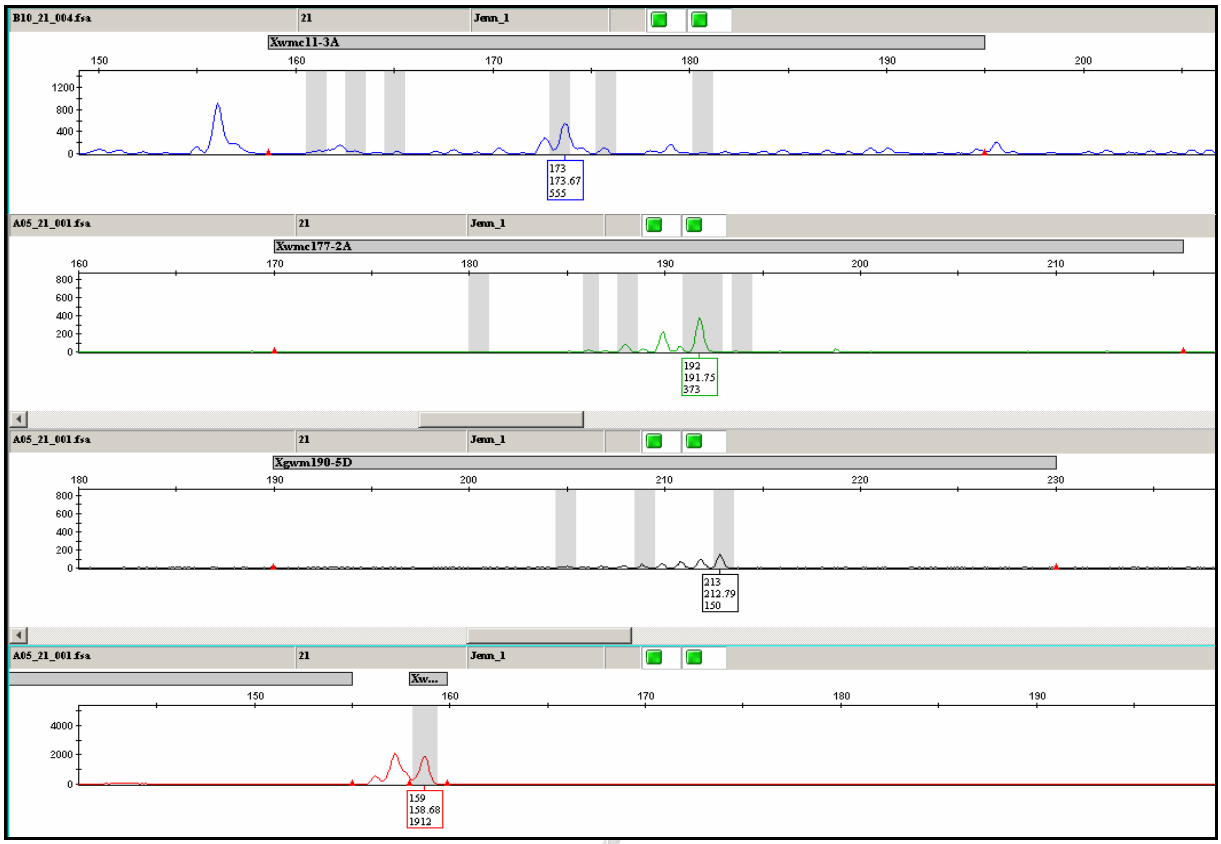


Figure 7-59 Sample number 21. (*Xwmc11*, *Xwmc177*, *Xgwm190*, *Xwmc59*)

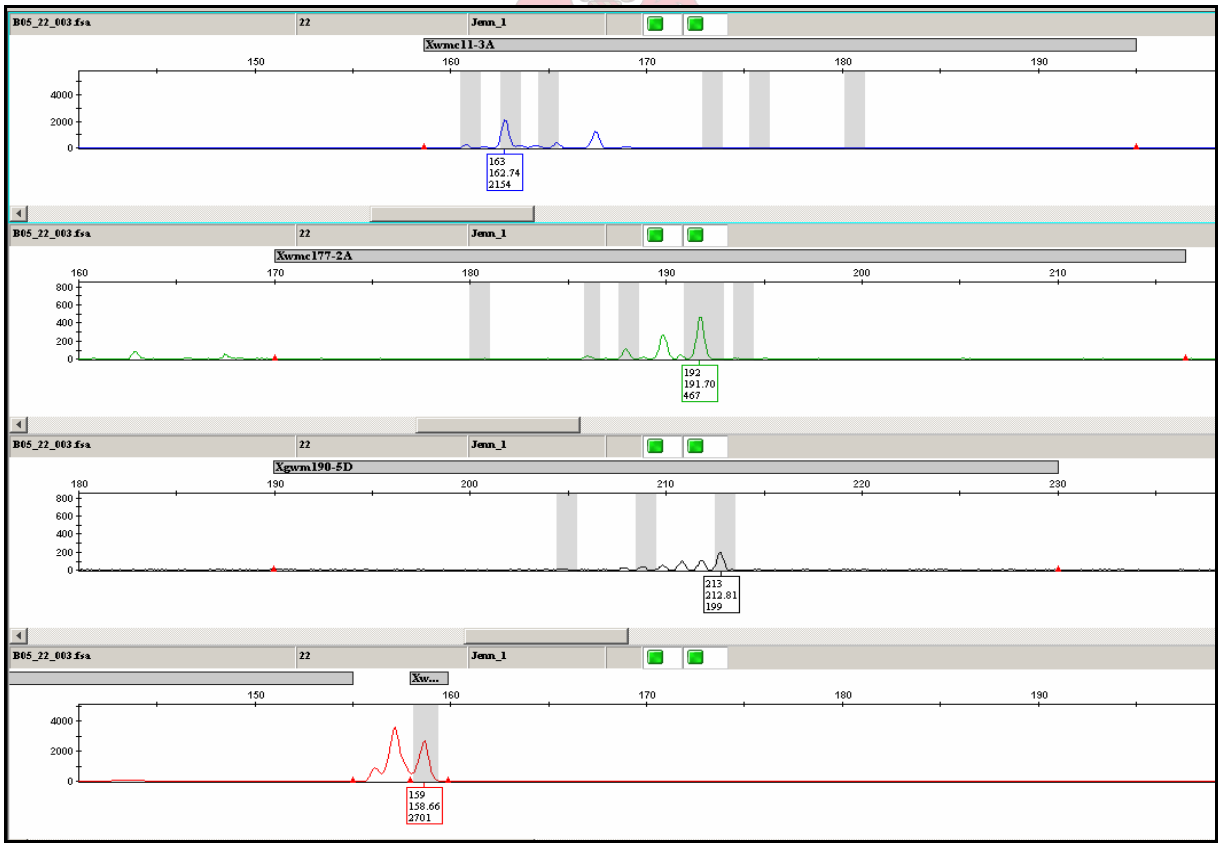


Figure 7-60 Sample number 22. (*Xwmc11*, *Xwmc177*, *Xgwm190*, *Xwmc59*)

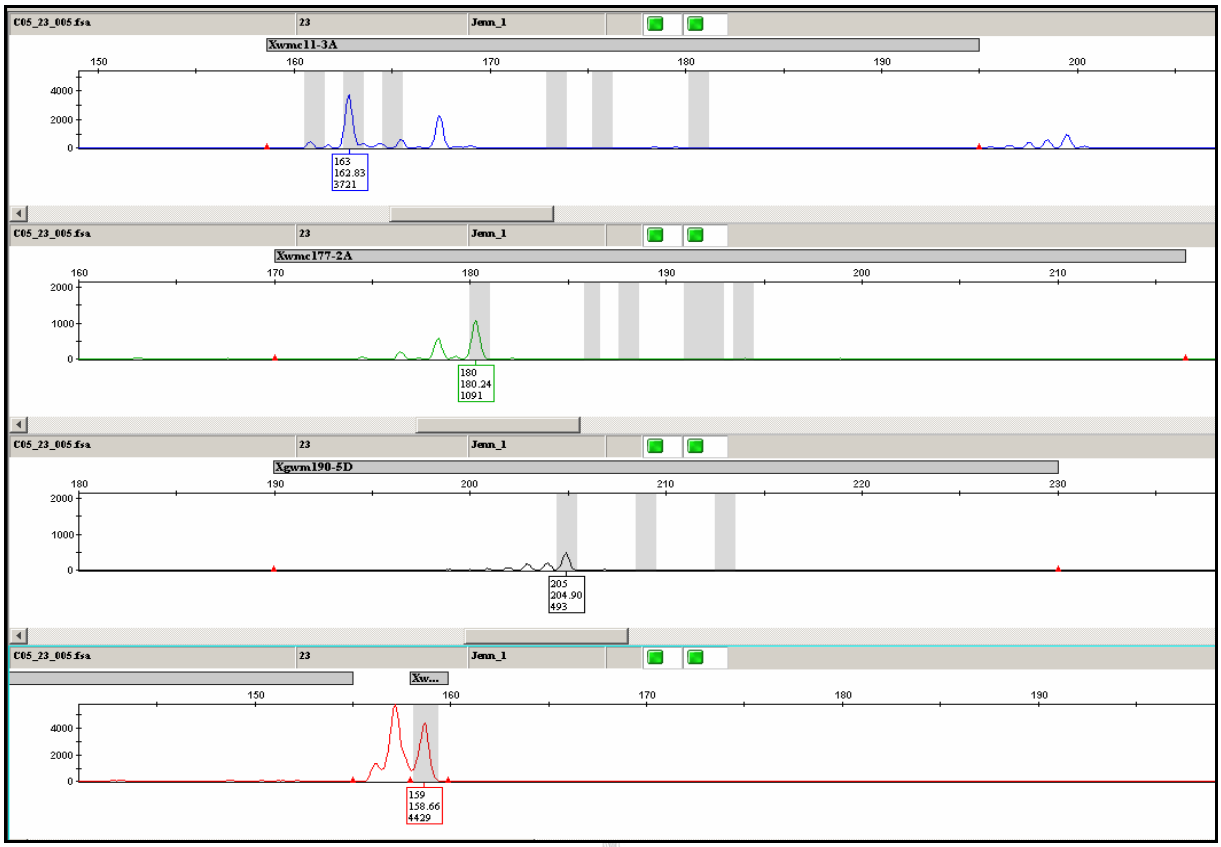


Figure 7-61 Sample number 23. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

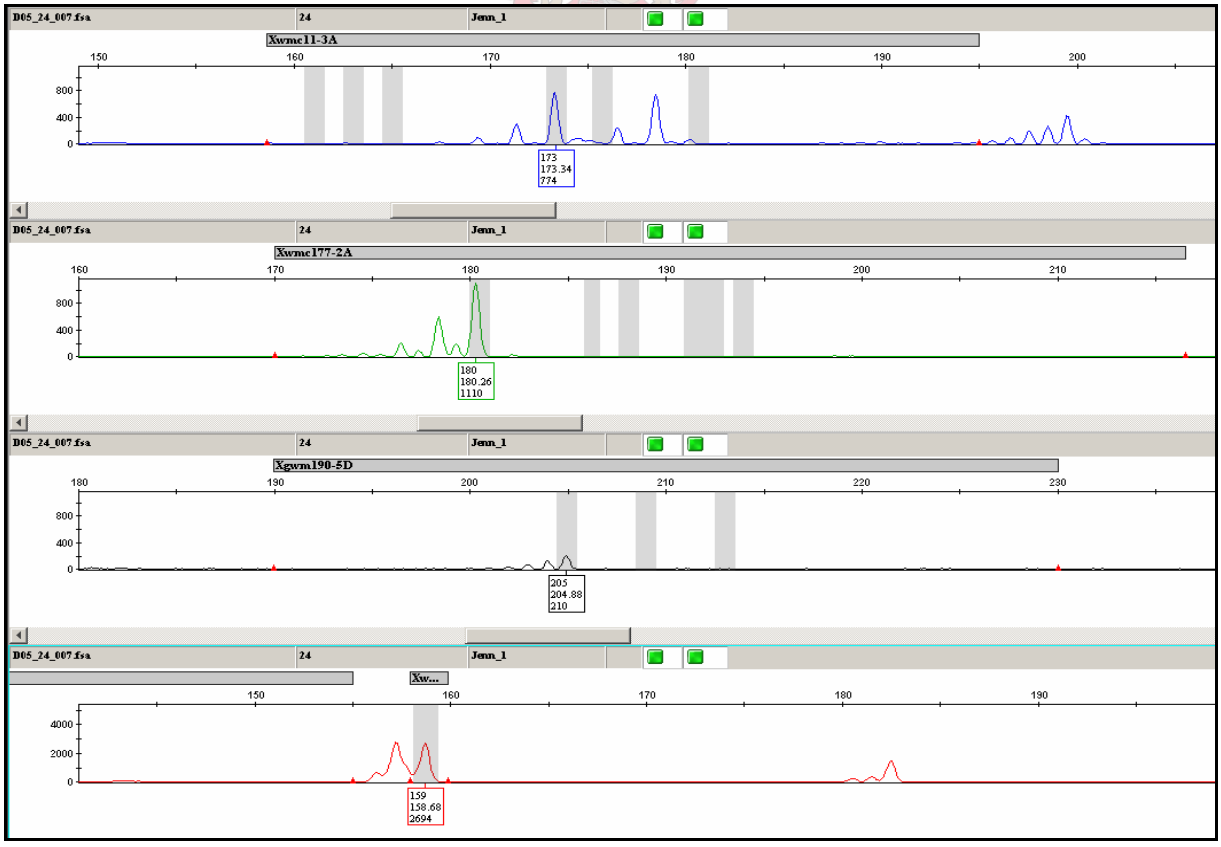


Figure 7-62 Sample number 24. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

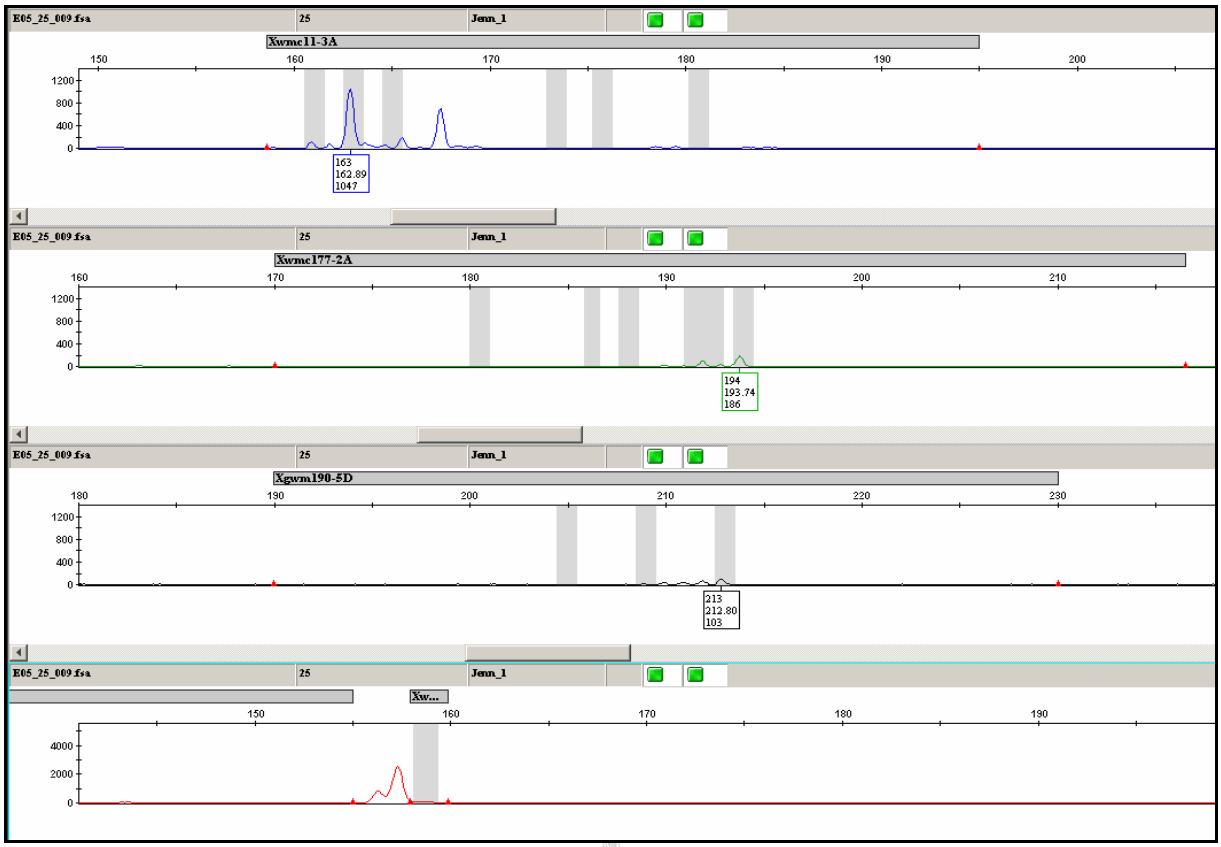


Figure 7-63 Sample number 25. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

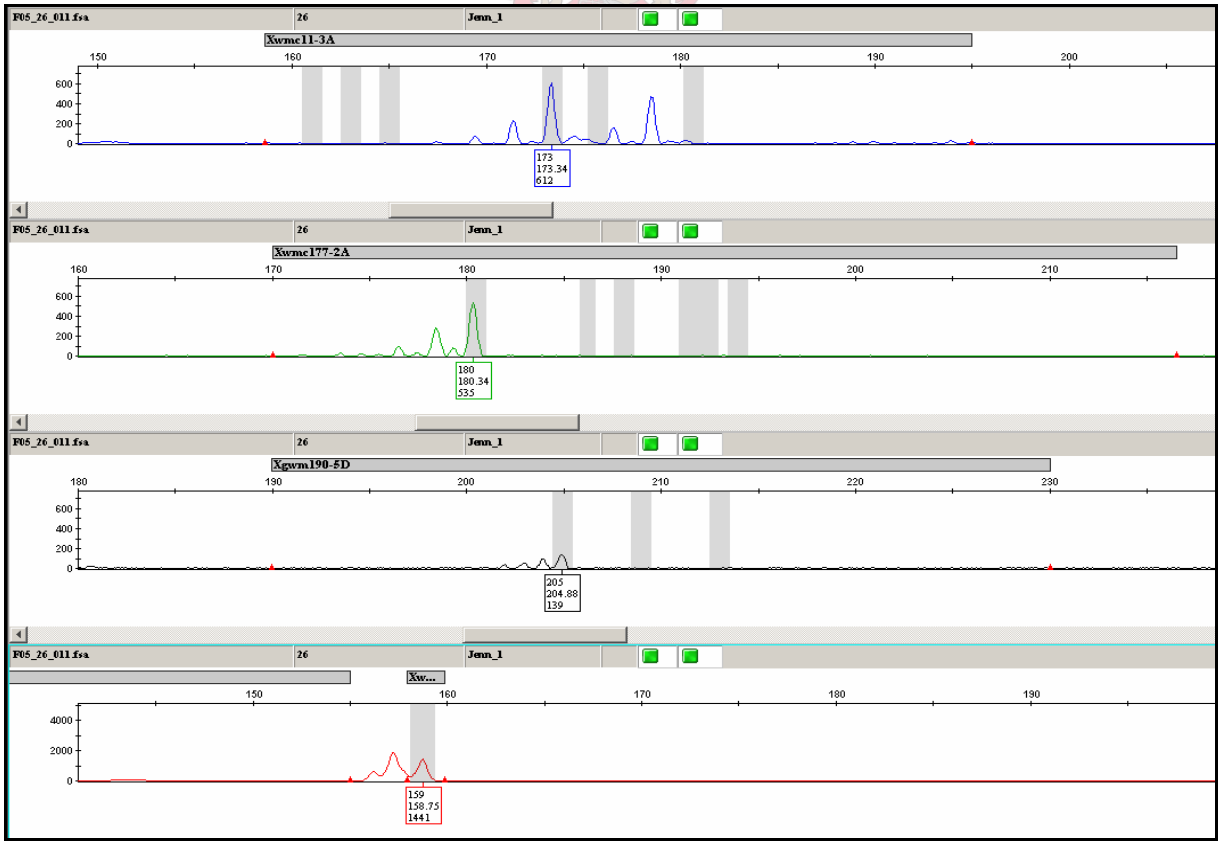


Figure 7-64 Sample number 26. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

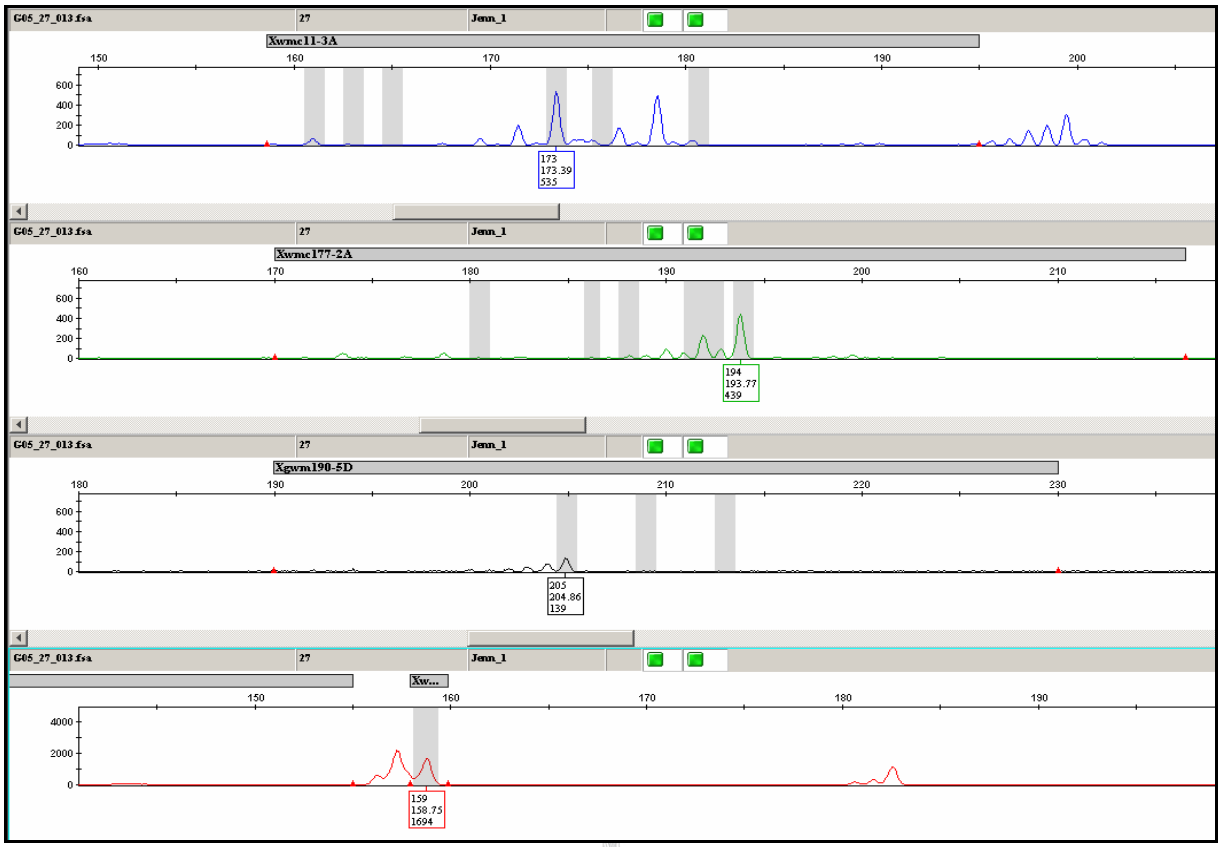


Figure 7-65 Sample number 27. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

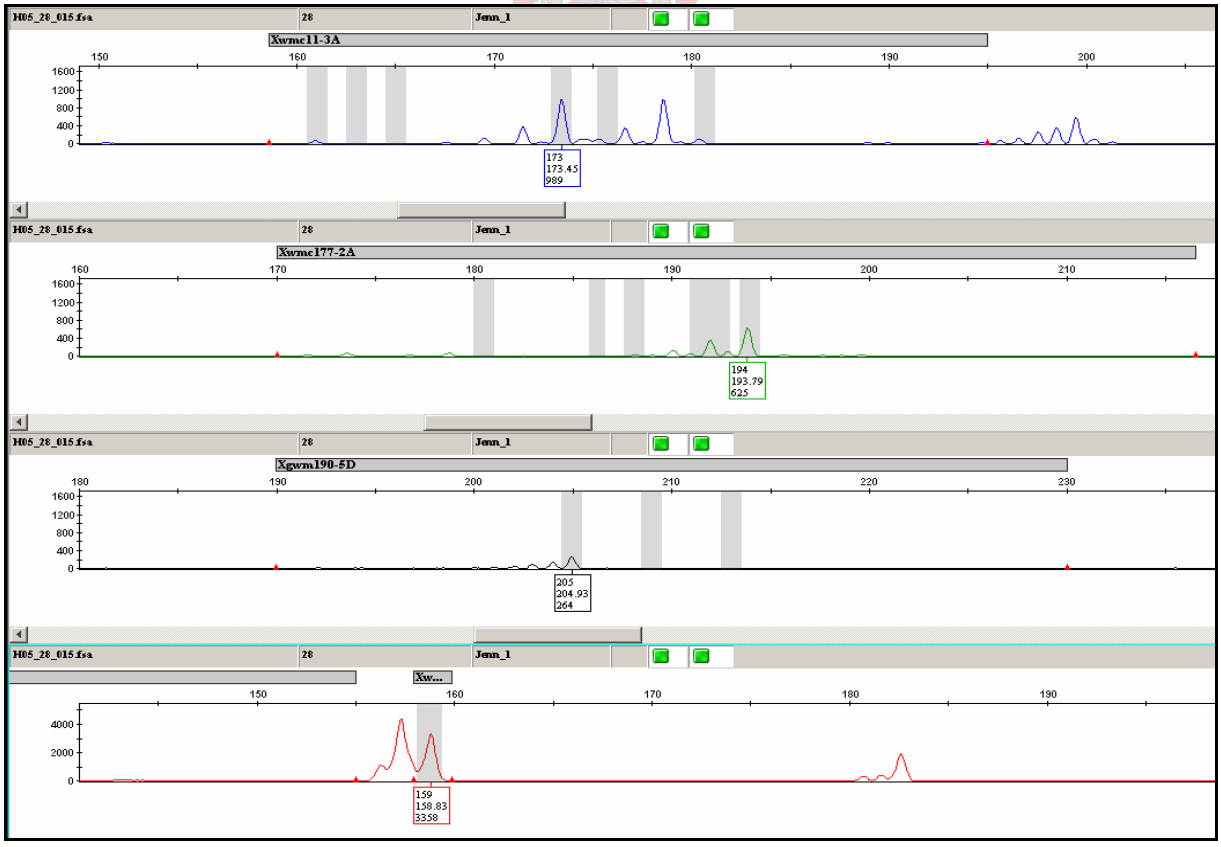


Figure 7-66 Sample number 28. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

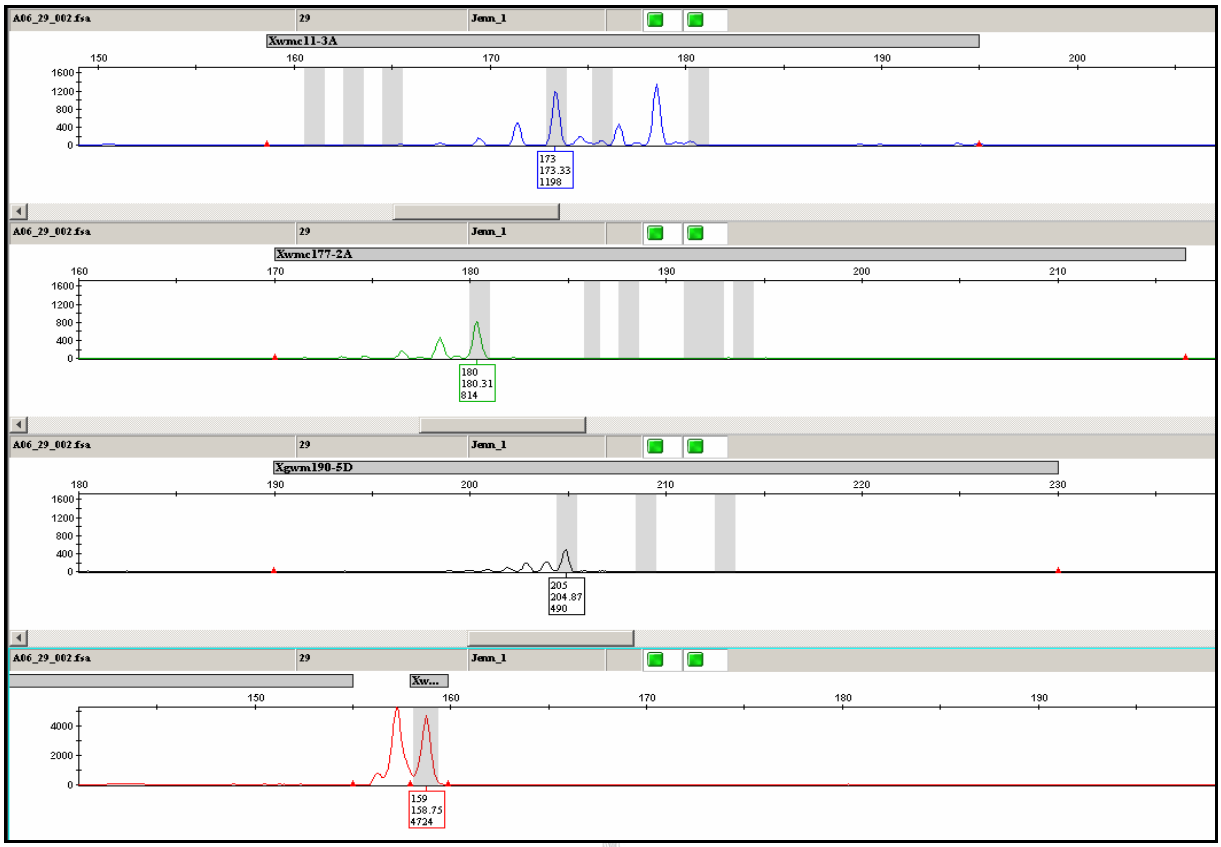


Figure 7-67 Sample number 29. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

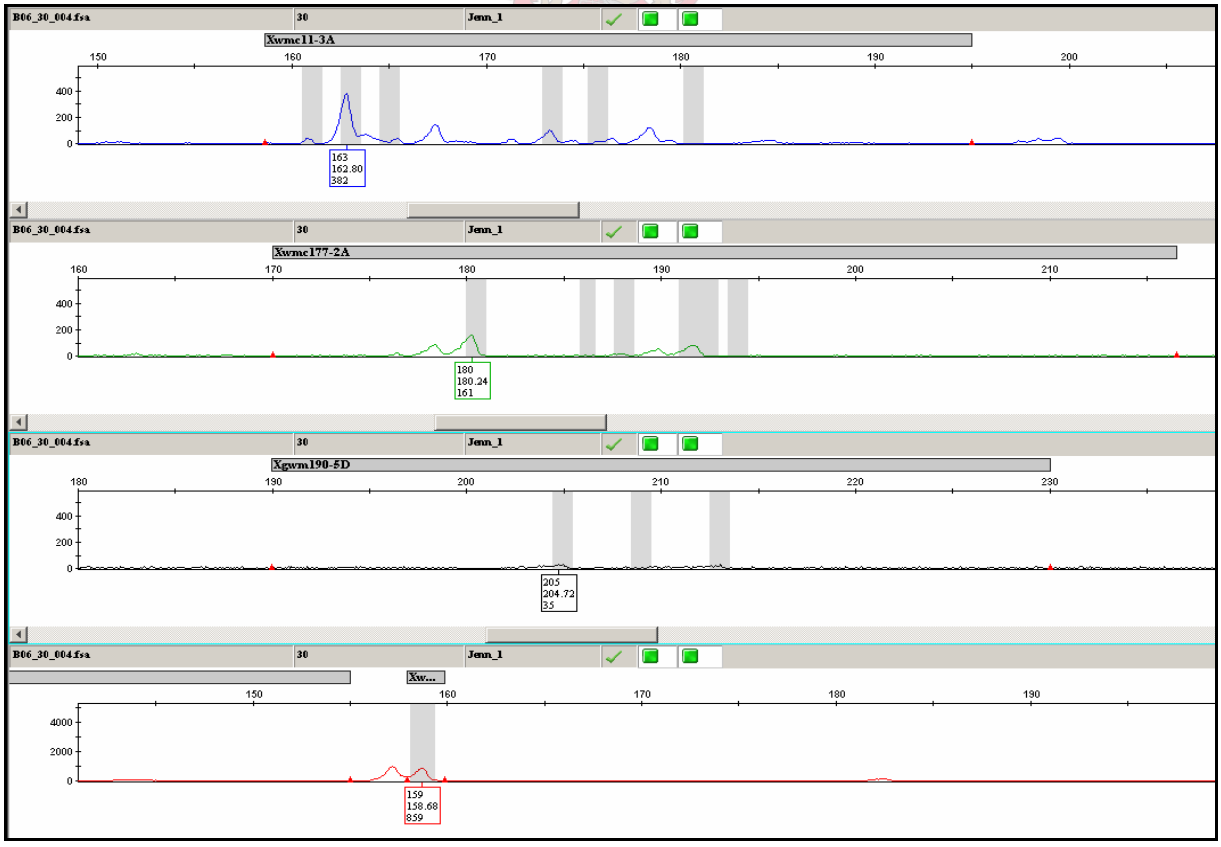


Figure 7-68 Sample number 30. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

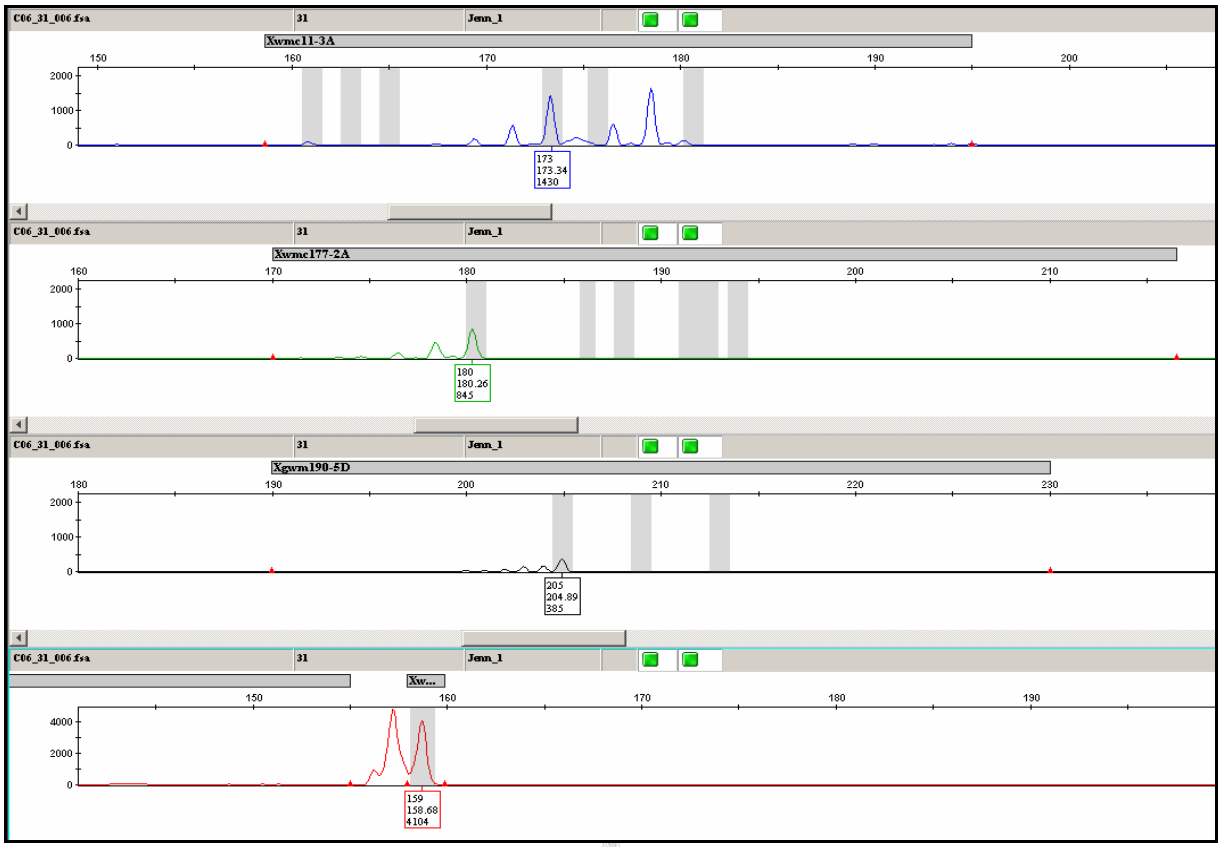


Figure 7-69 Sample number 31. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

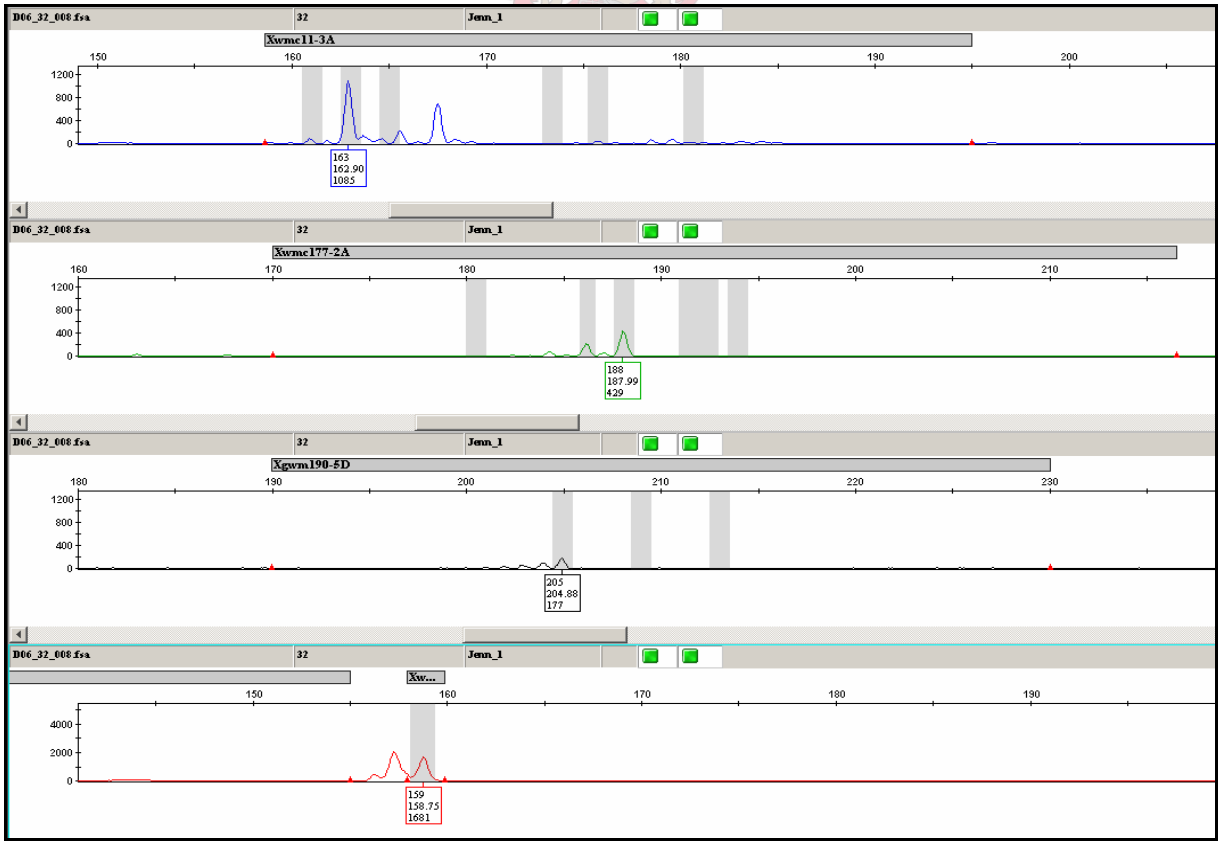


Figure 7-70 Sample number 32. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

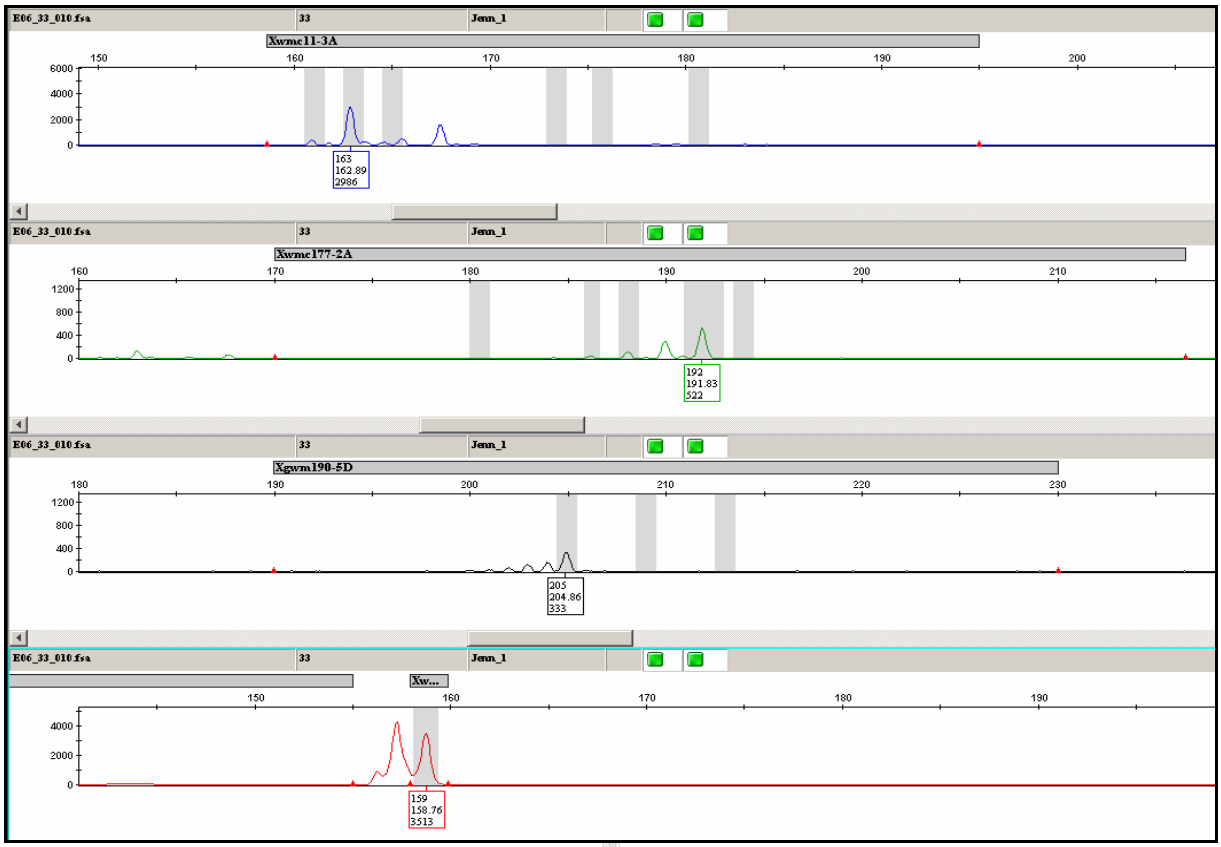


Figure 7-71 Sample number 33. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

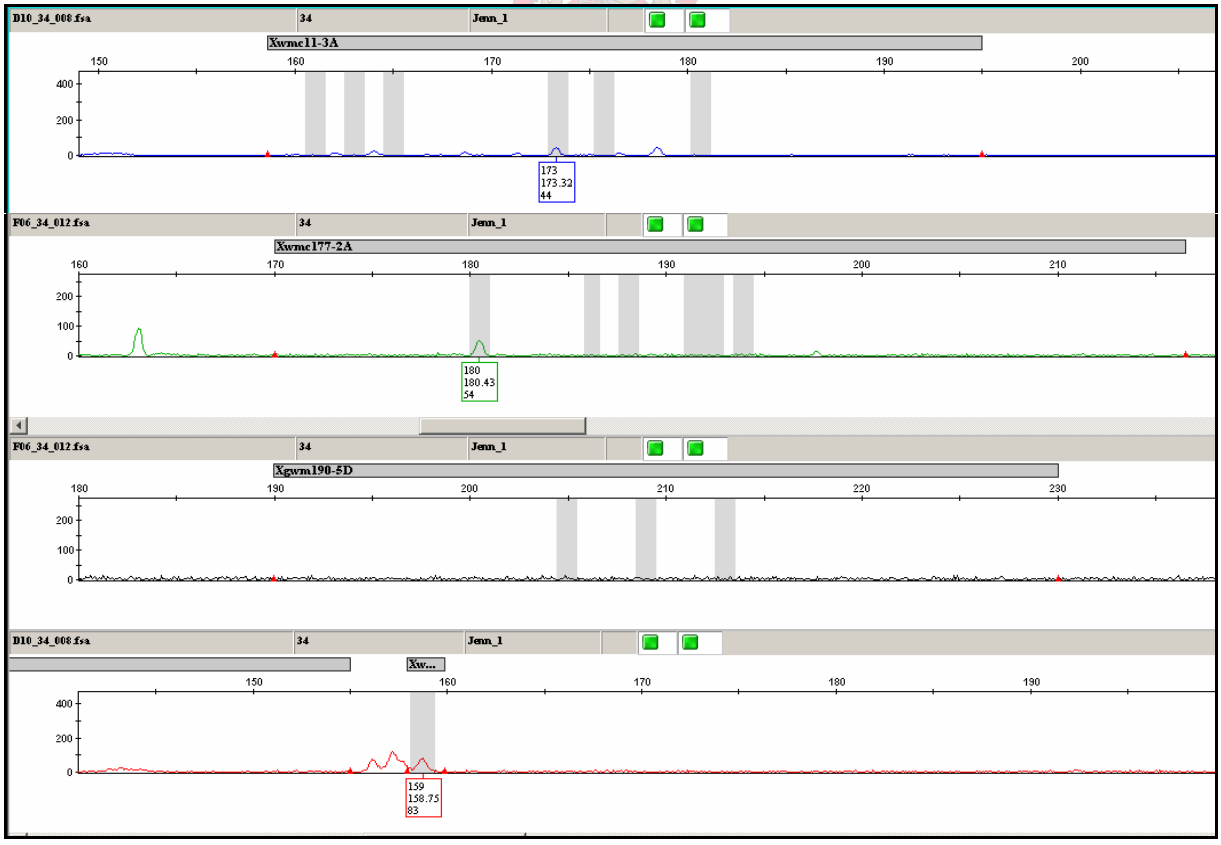


Figure 7-72 Sample number 34. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

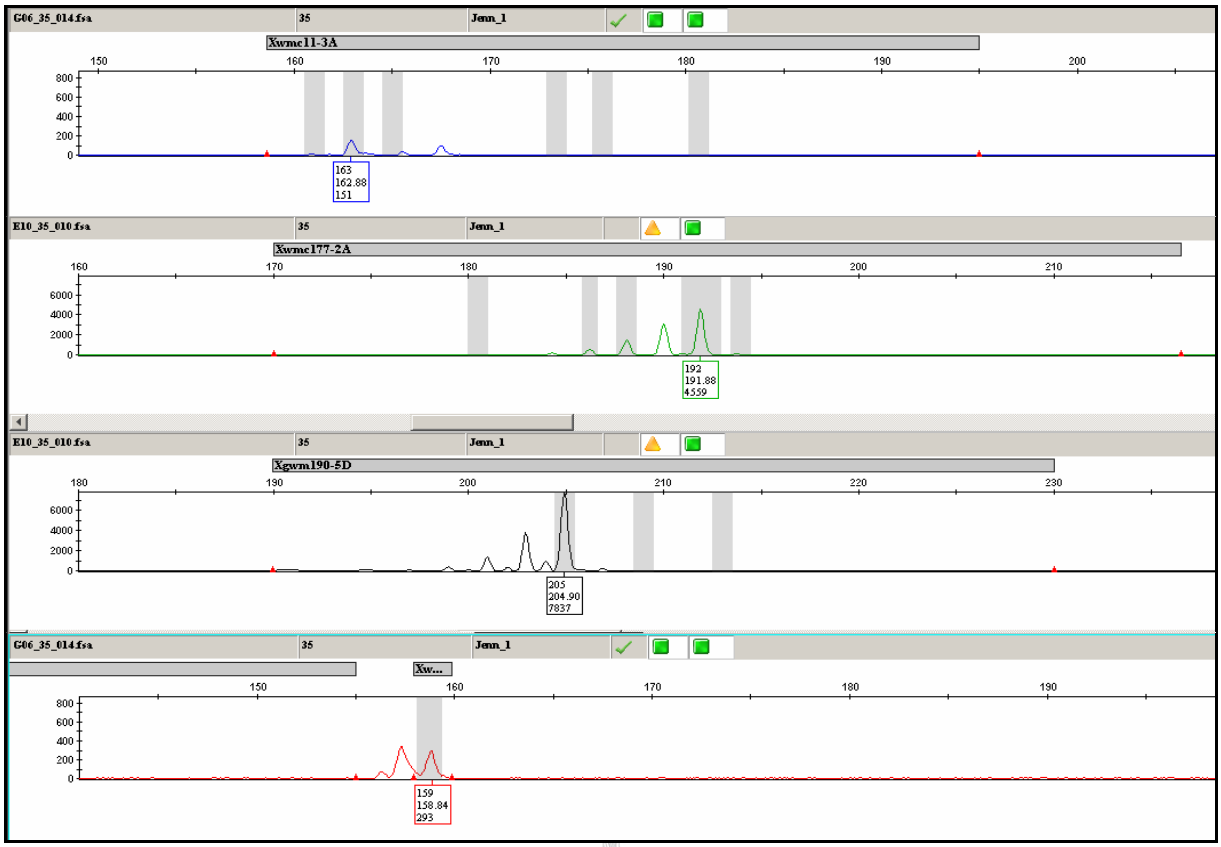


Figure 7-73 Sample number 35. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

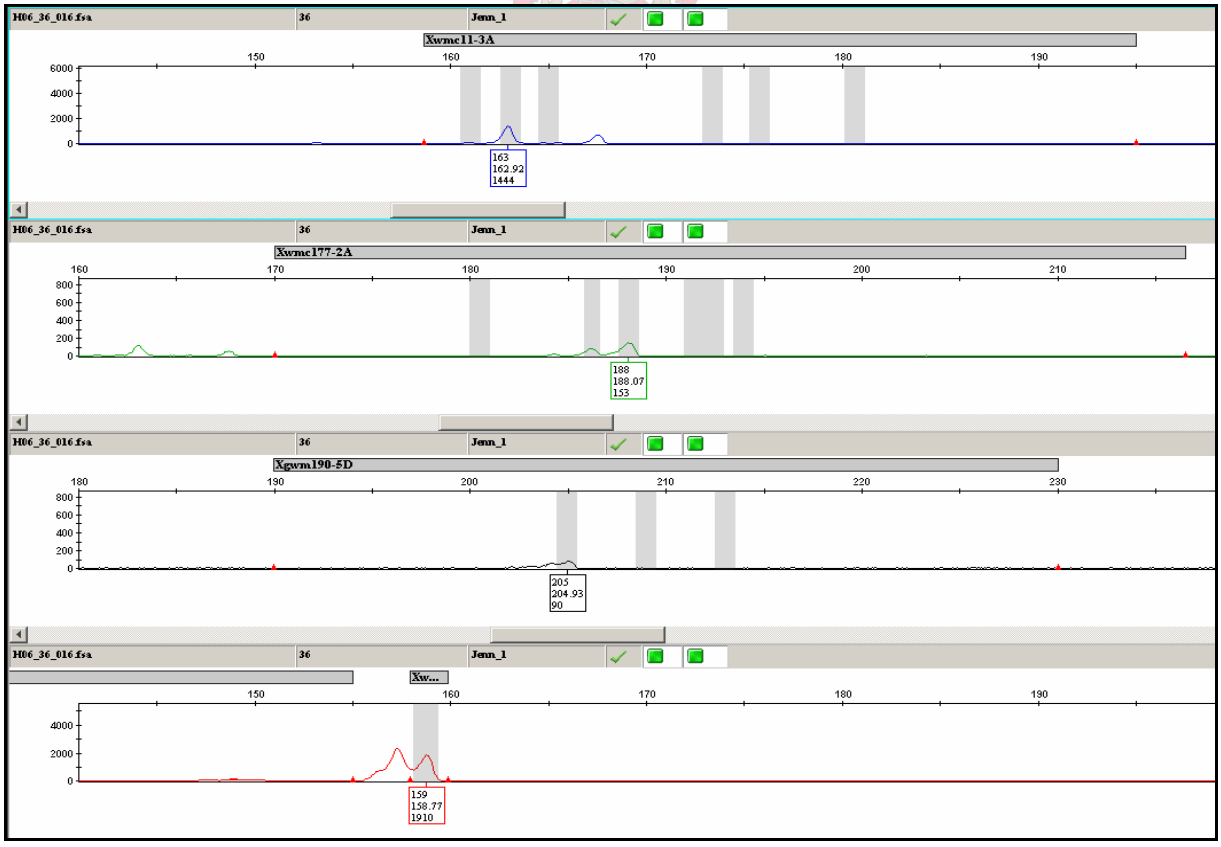


Figure 7-74 Sample number 36. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

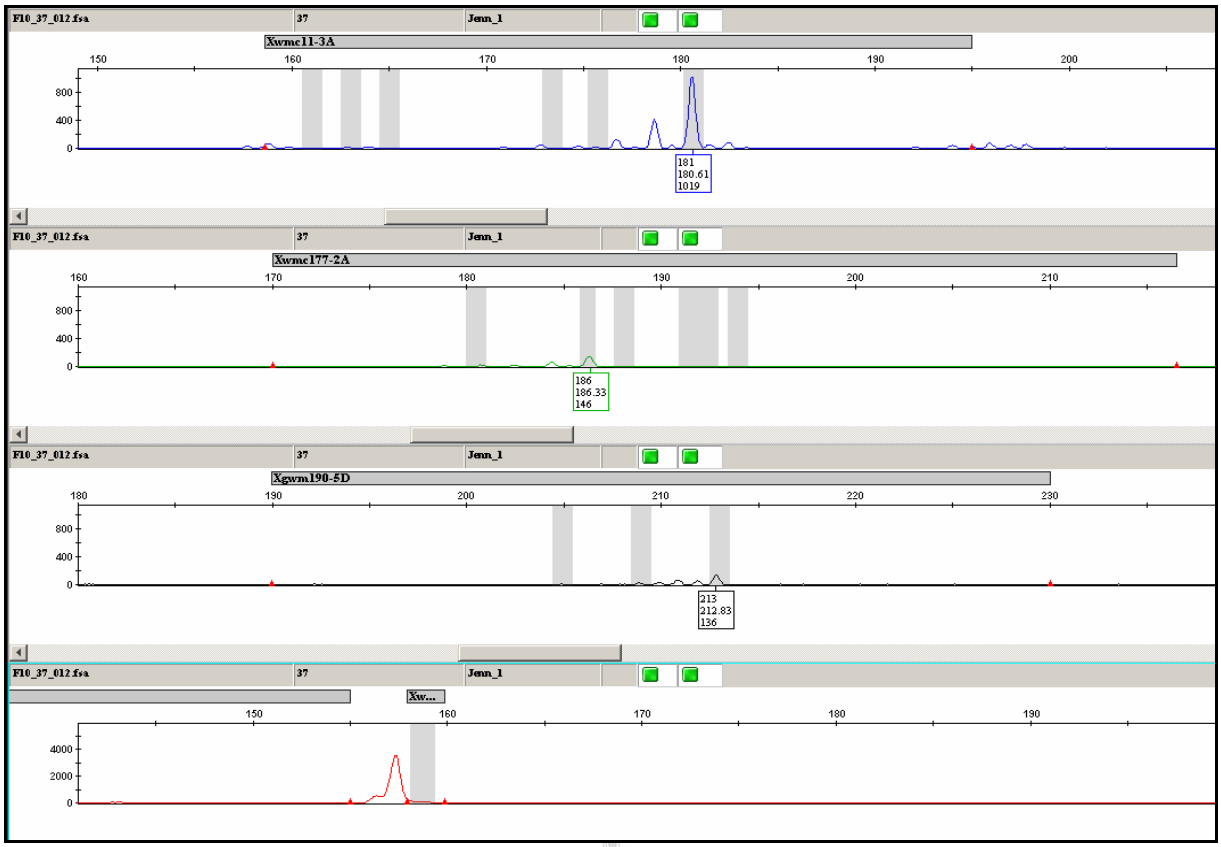


Figure 7-75 Sample number 37. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

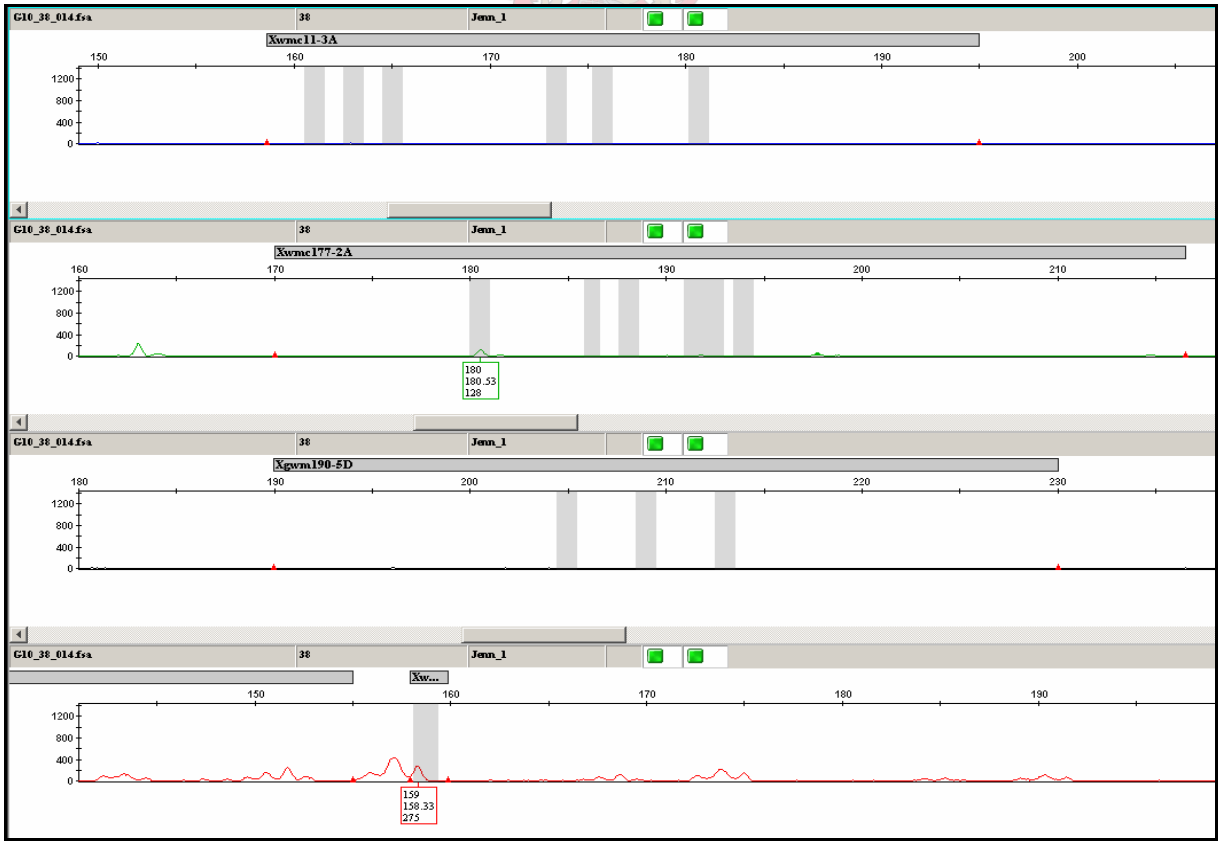


Figure 7-76 Sample number 38. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

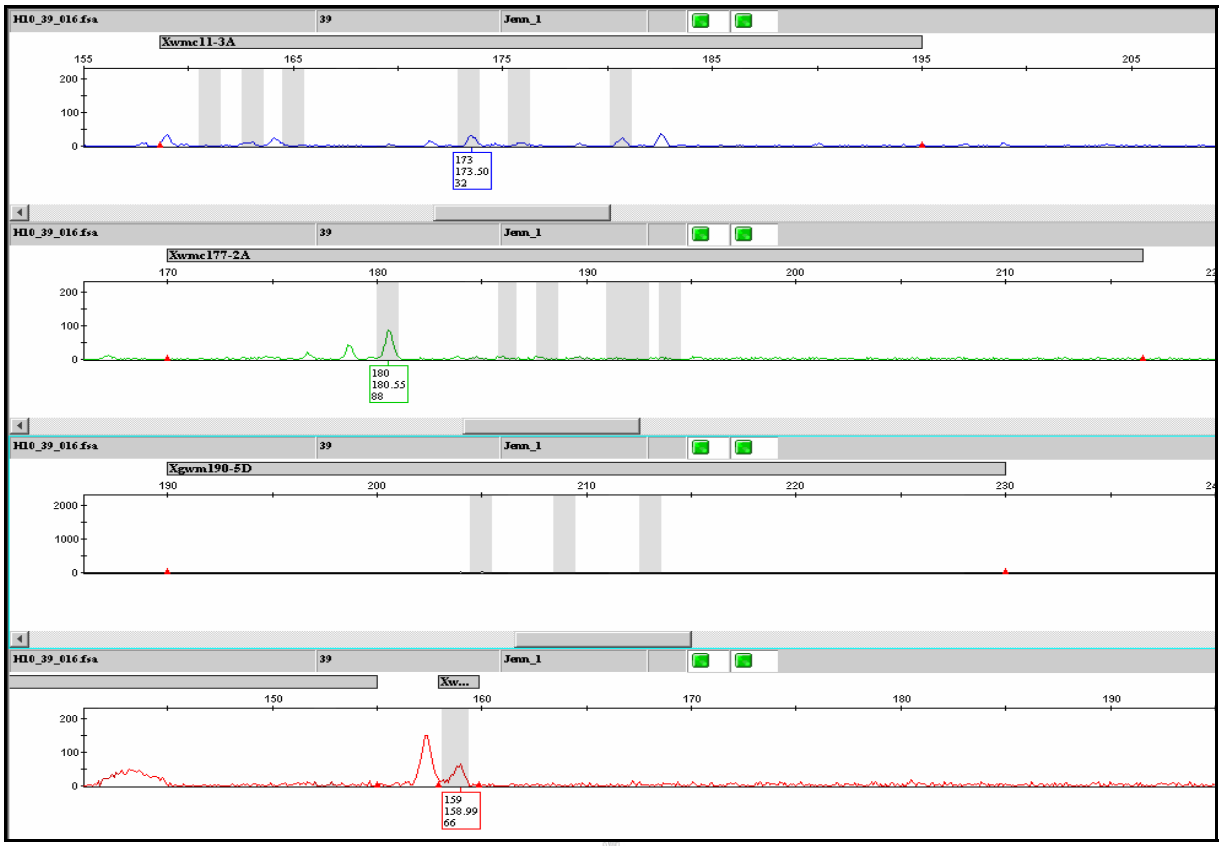


Figure 7-77 Sample number 39. (*Xwmc11, Xwmc177, Xgwm190, Xwmc59*)

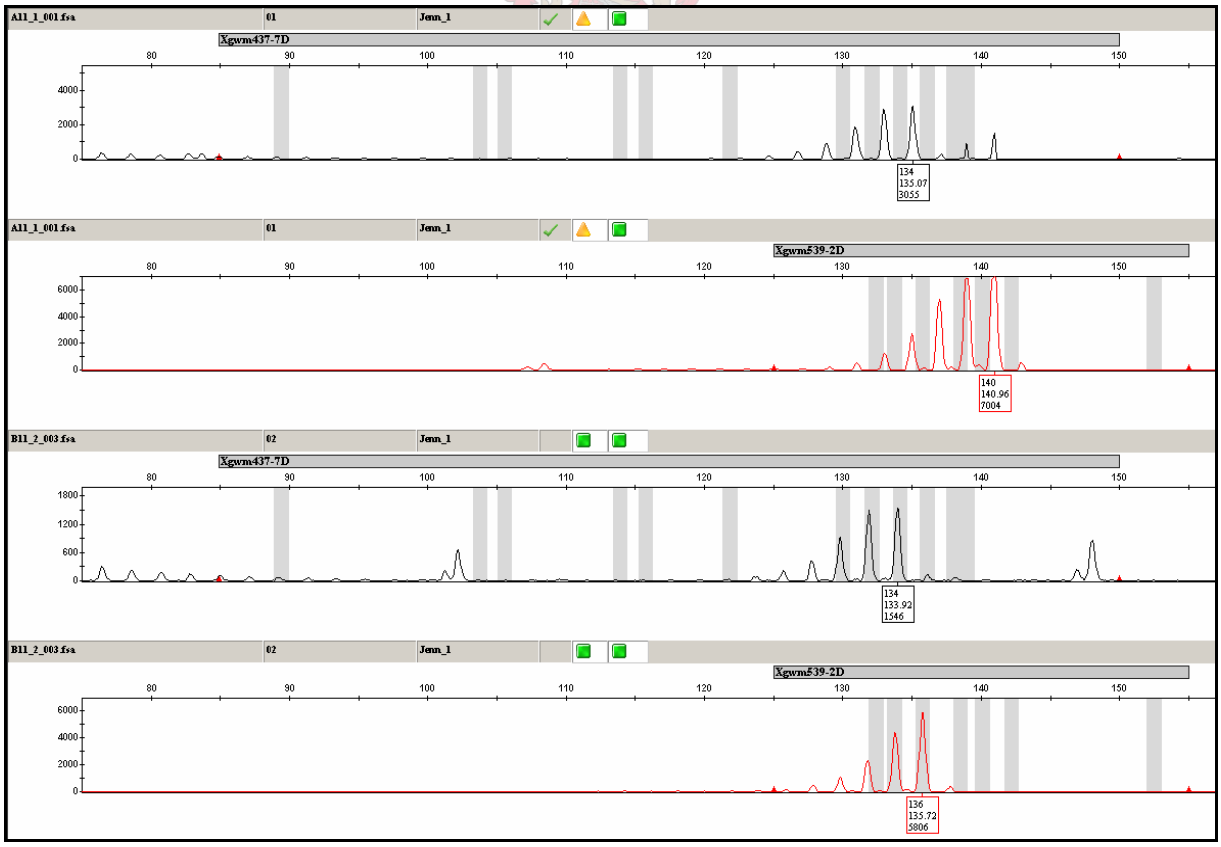


Figure 7-78 Sample number 1 and 2. (*Xgwm437, Xgwm539*)

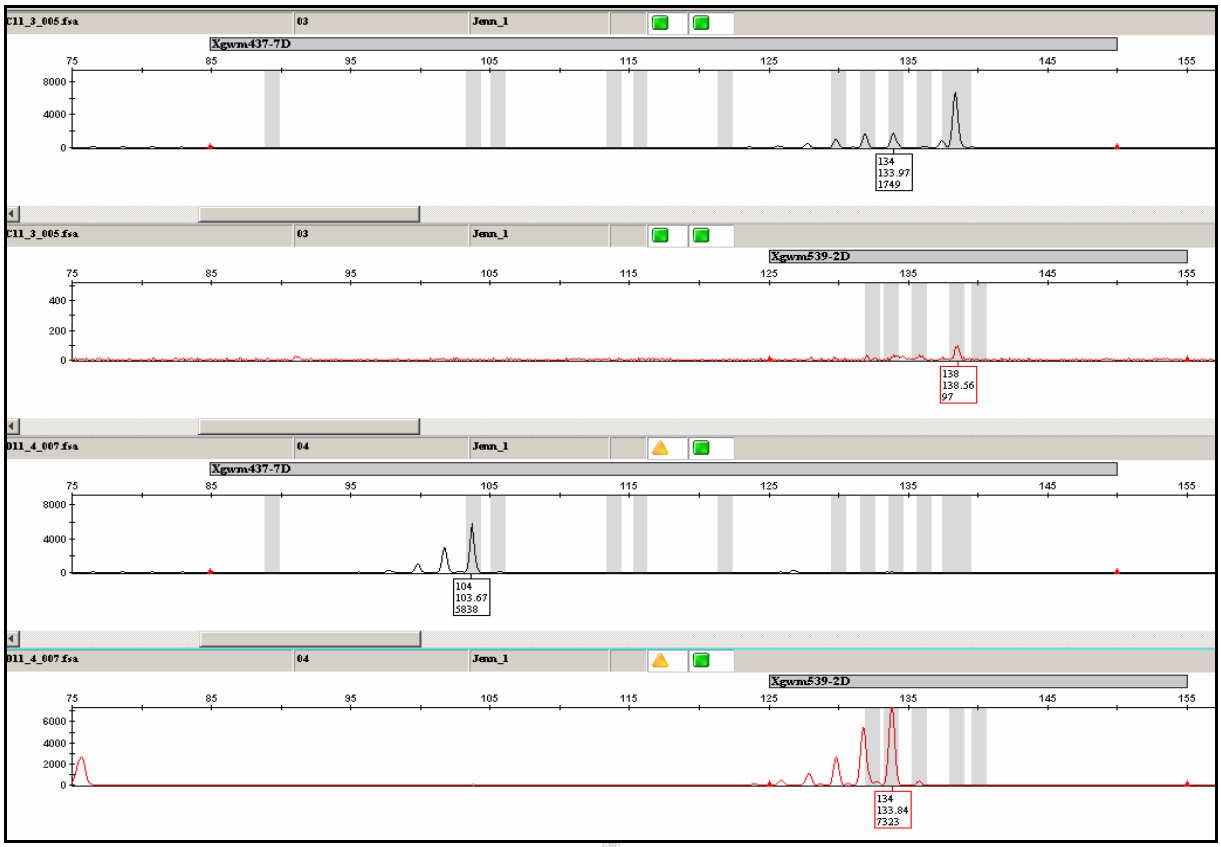


Figure 7-79 Sample number 3 and 4. (*Xgwm437*, *Xgwm539*)

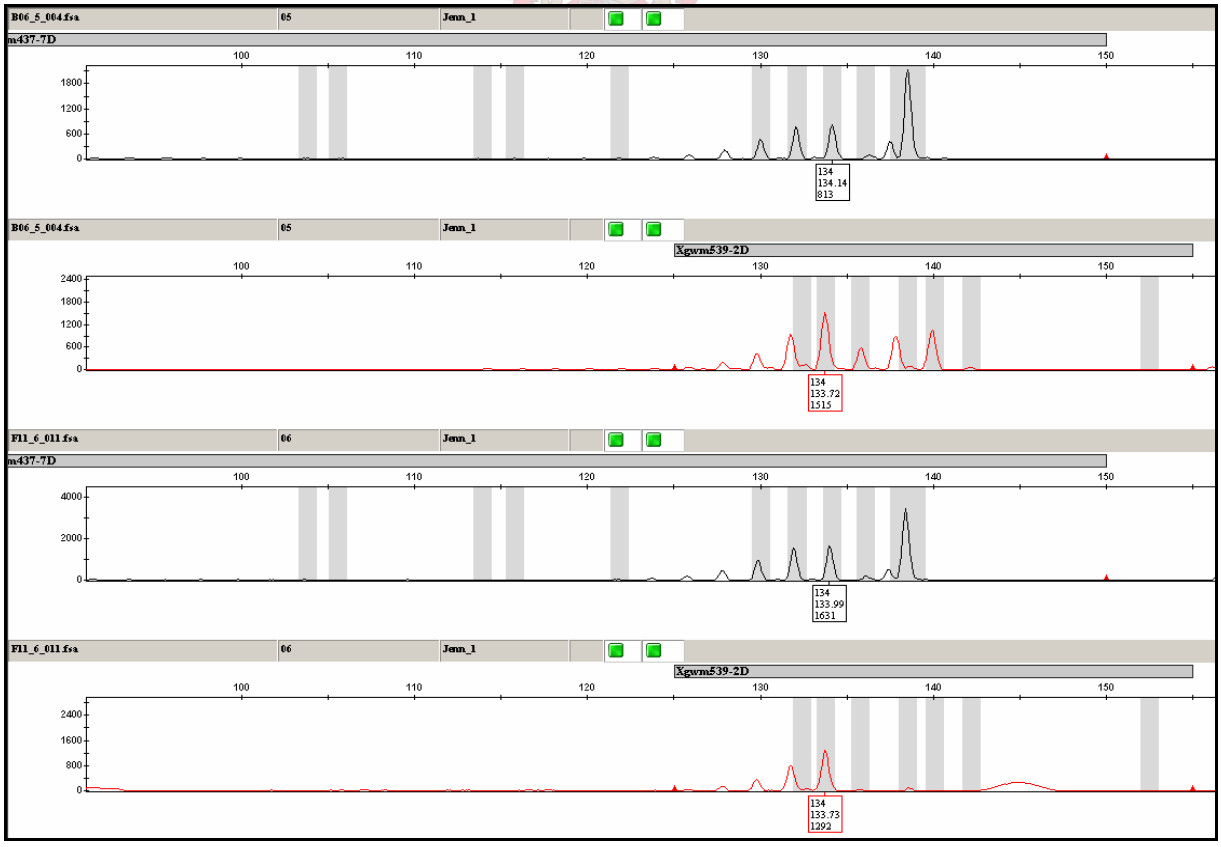


Figure 7-80 Sample number 5 and 6. (*Xgwm437*, *Xgwm539*)

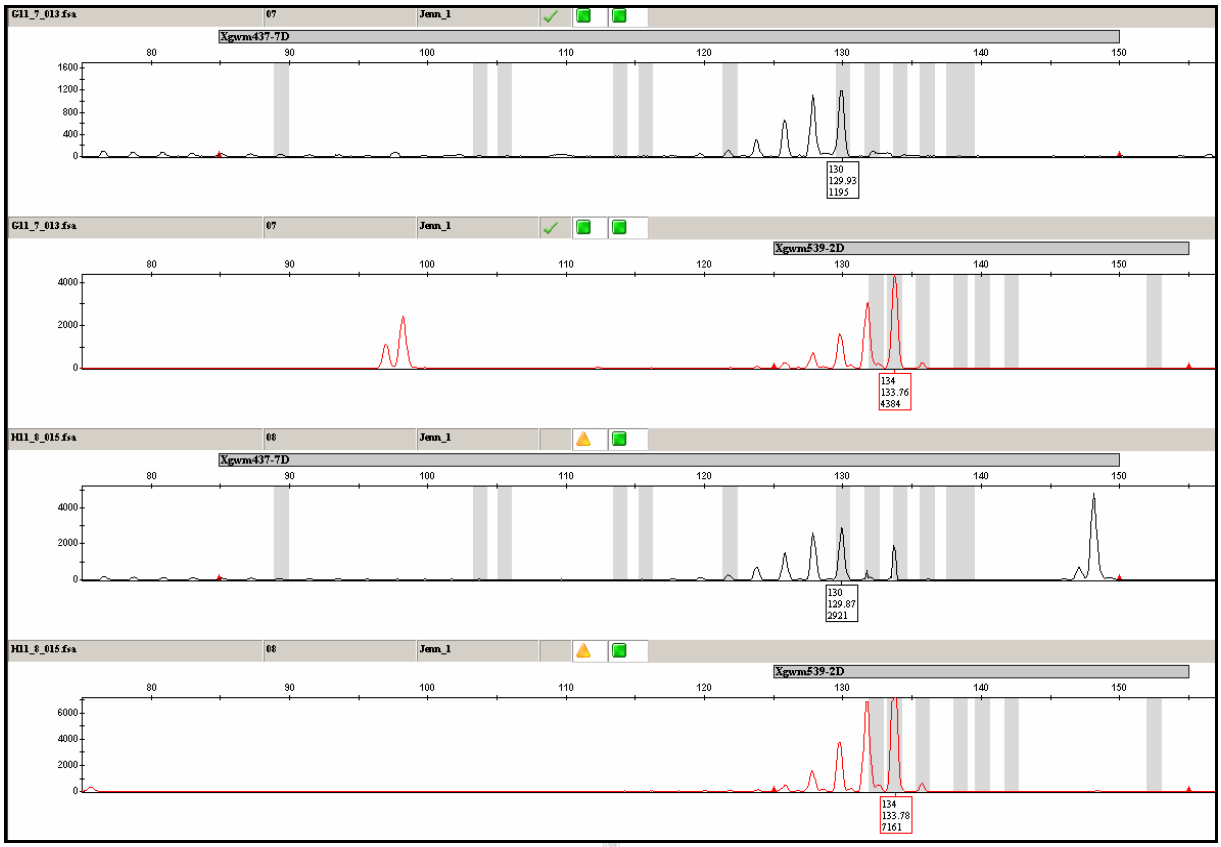


Figure 7-81 Sample number 7 and 8. (*Xgwm437, Xgwm539*)

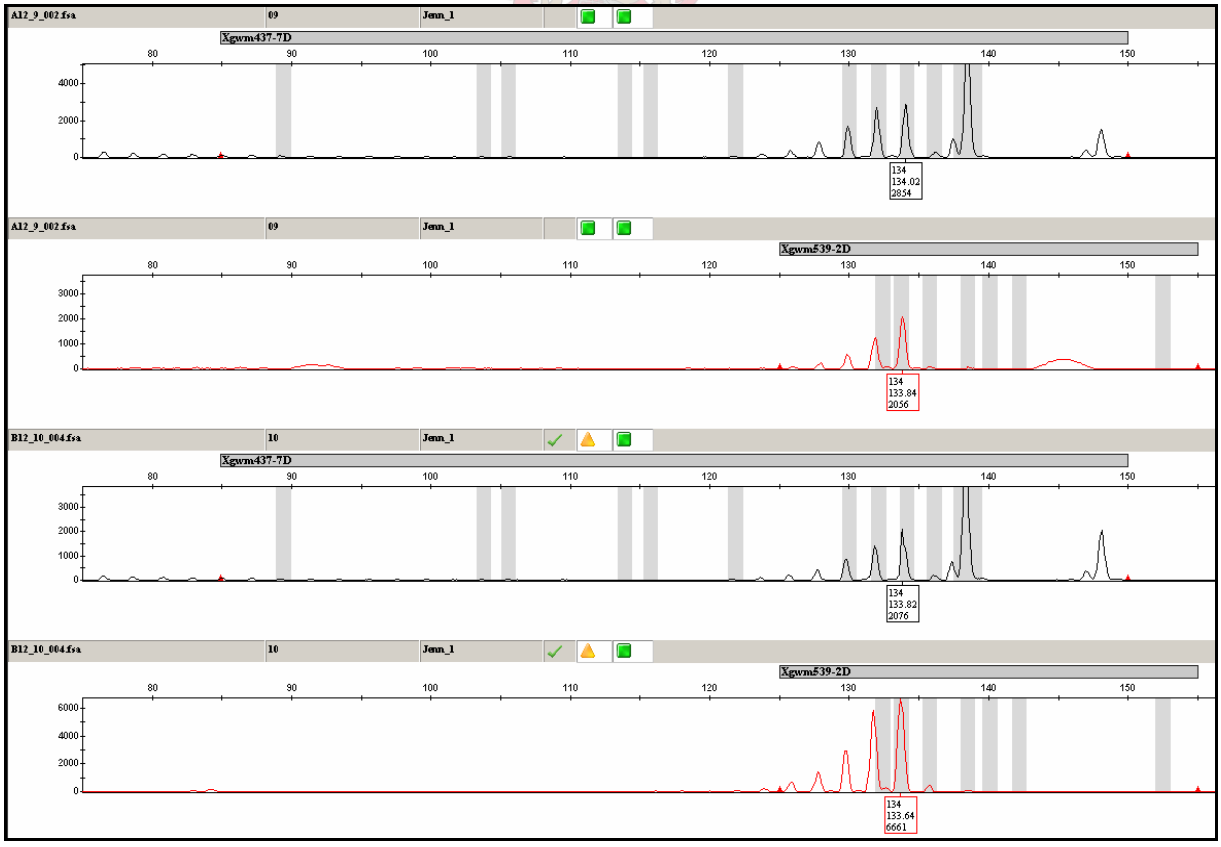


Figure 7-82 Sample number 9 and 10. (*Xgwm437, Xgwm539*)

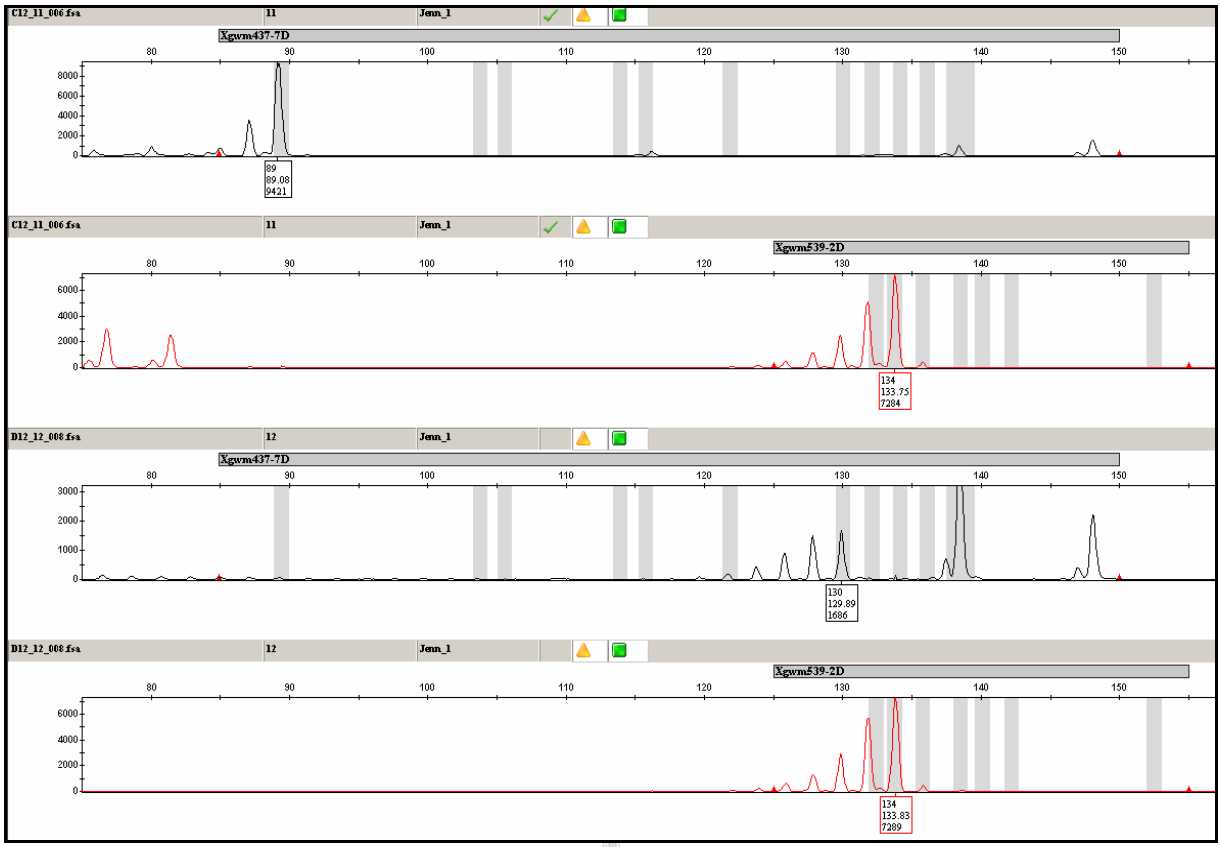


Figure 7-83 Sample number 11 and 12. (*Xgwm437*, *Xgwm539*)

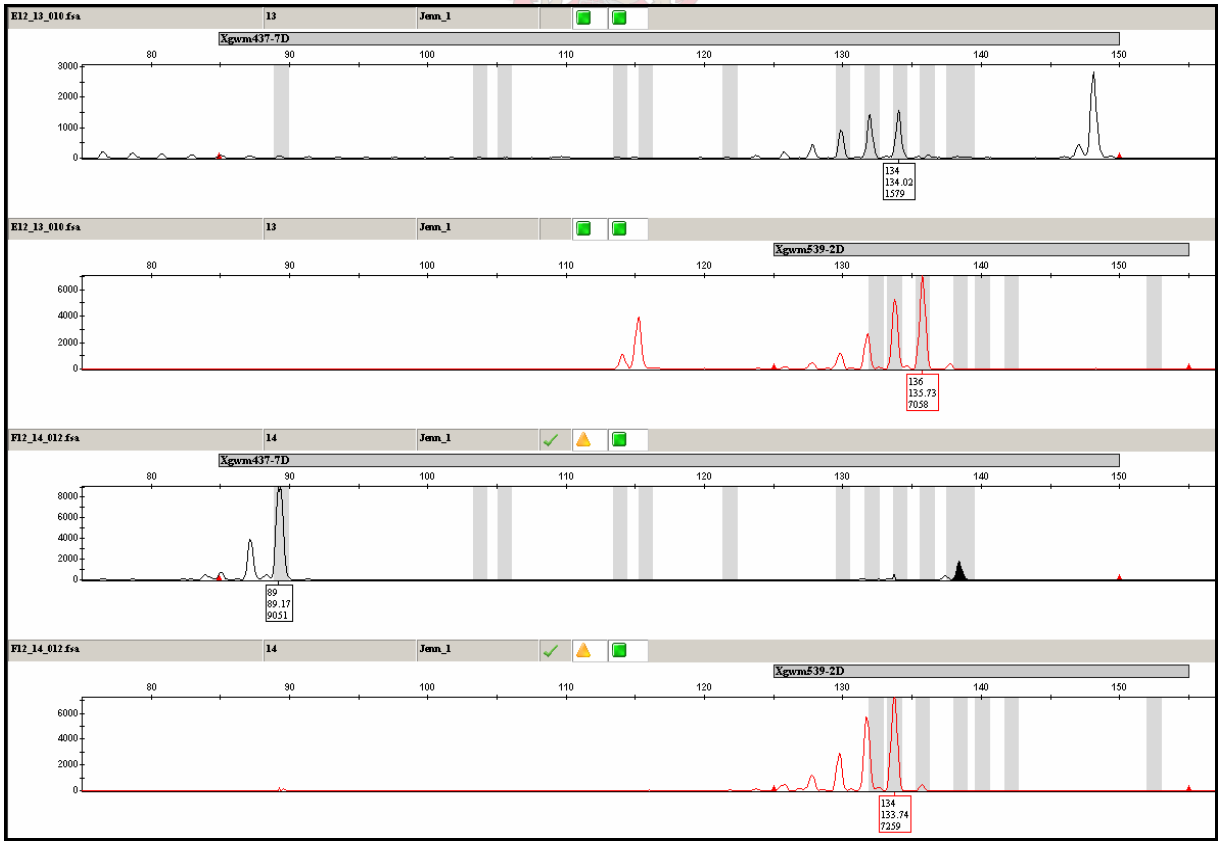


Figure 7-84 Sample number 13 and 14. (*Xgwm437*, *Xgwm539*)

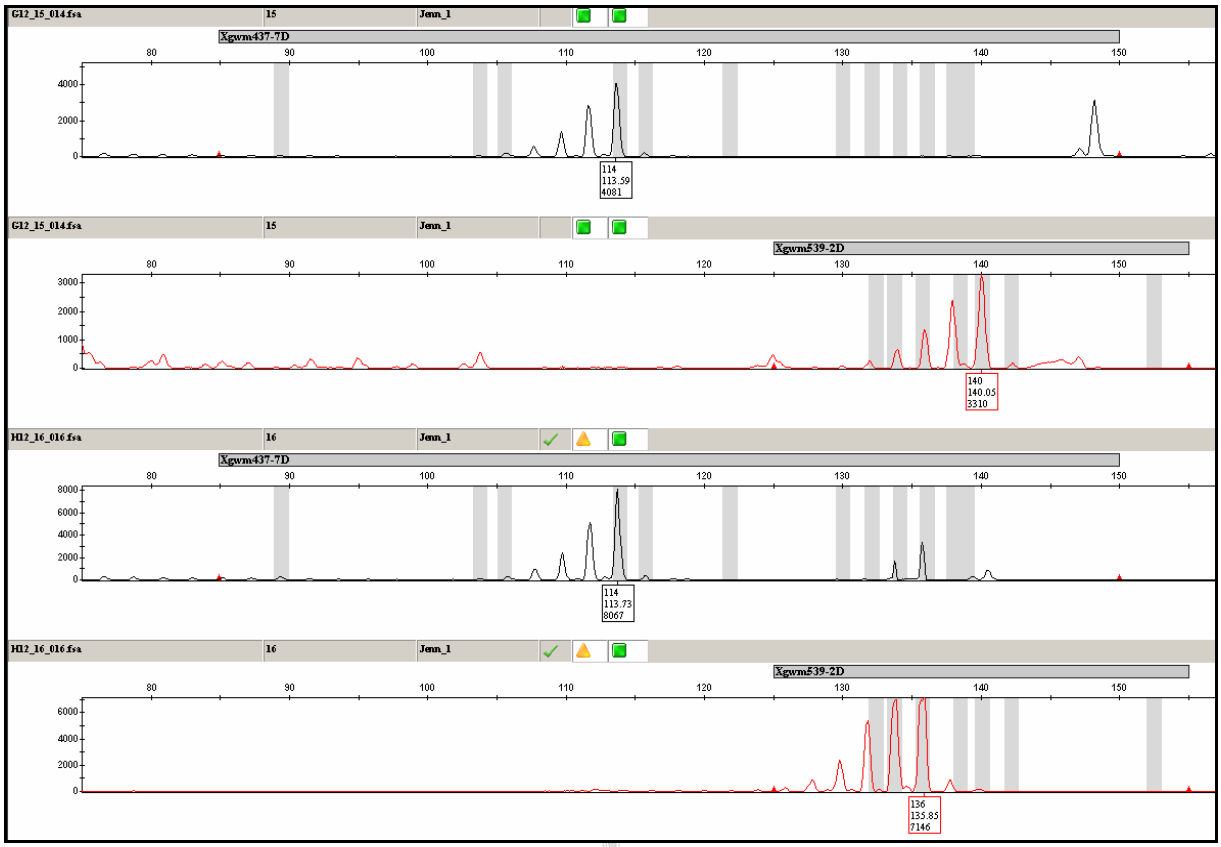


Figure 7-85 Sample number 15 and 16. (*Xgwm437*, *Xgwm539*)

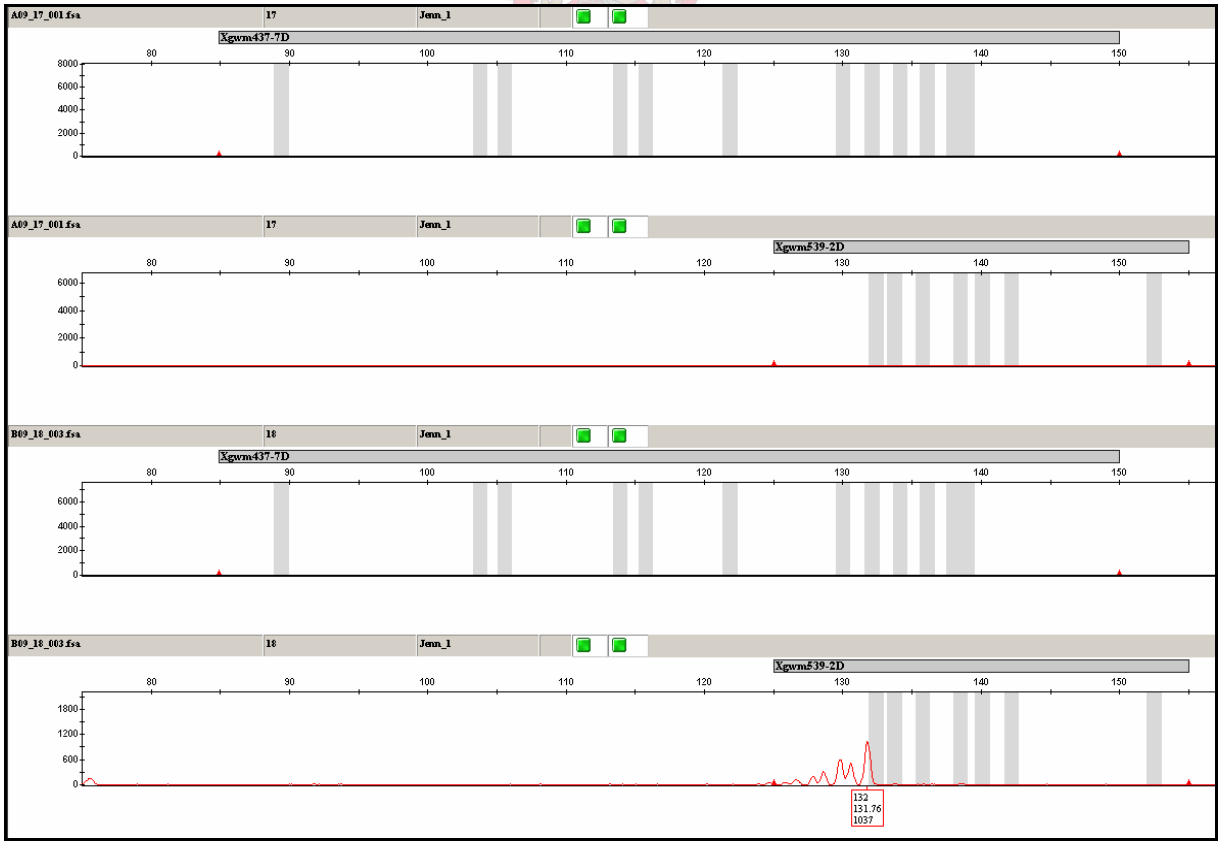


Figure 7-86 Sample number 17 and 18. (*Xgwm437*, *Xgwm539*)

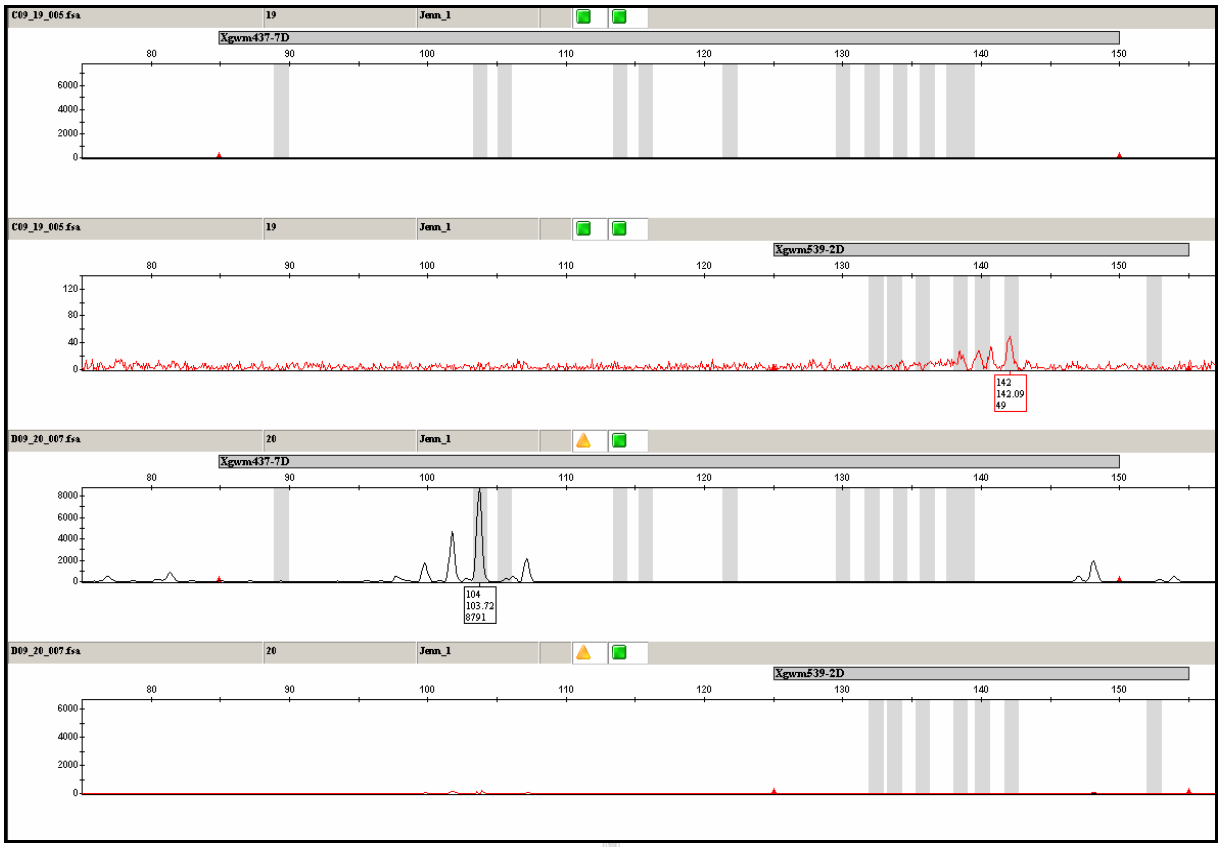


Figure 7-87 Sample number 19 and 20. (*Xgwm437*, *Xgwm539*)

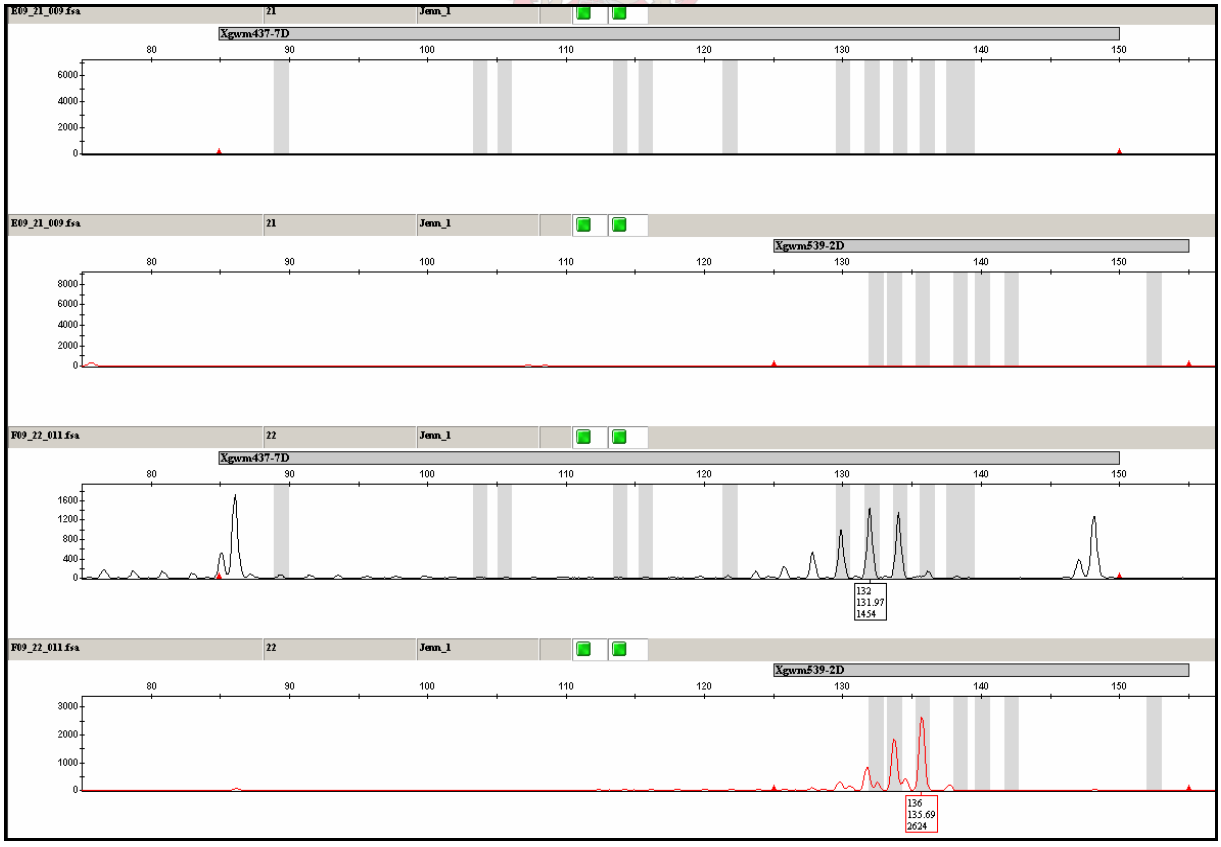


Figure 7-88 Sample number 21 and 22. (*Xgwm437*, *Xgwm539*)

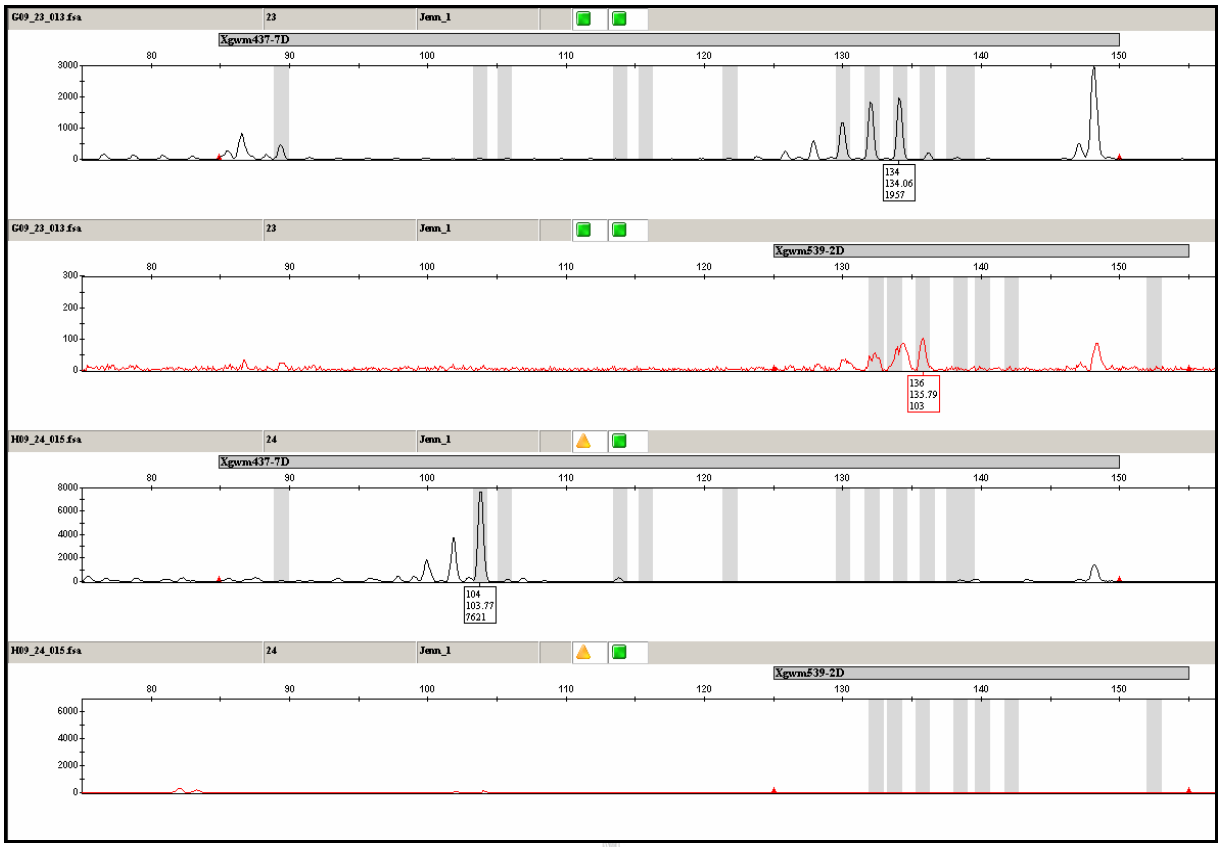


Figure 7-89 Sample number 23 and 24. (*Xgwm437*, *Xgwm539*)

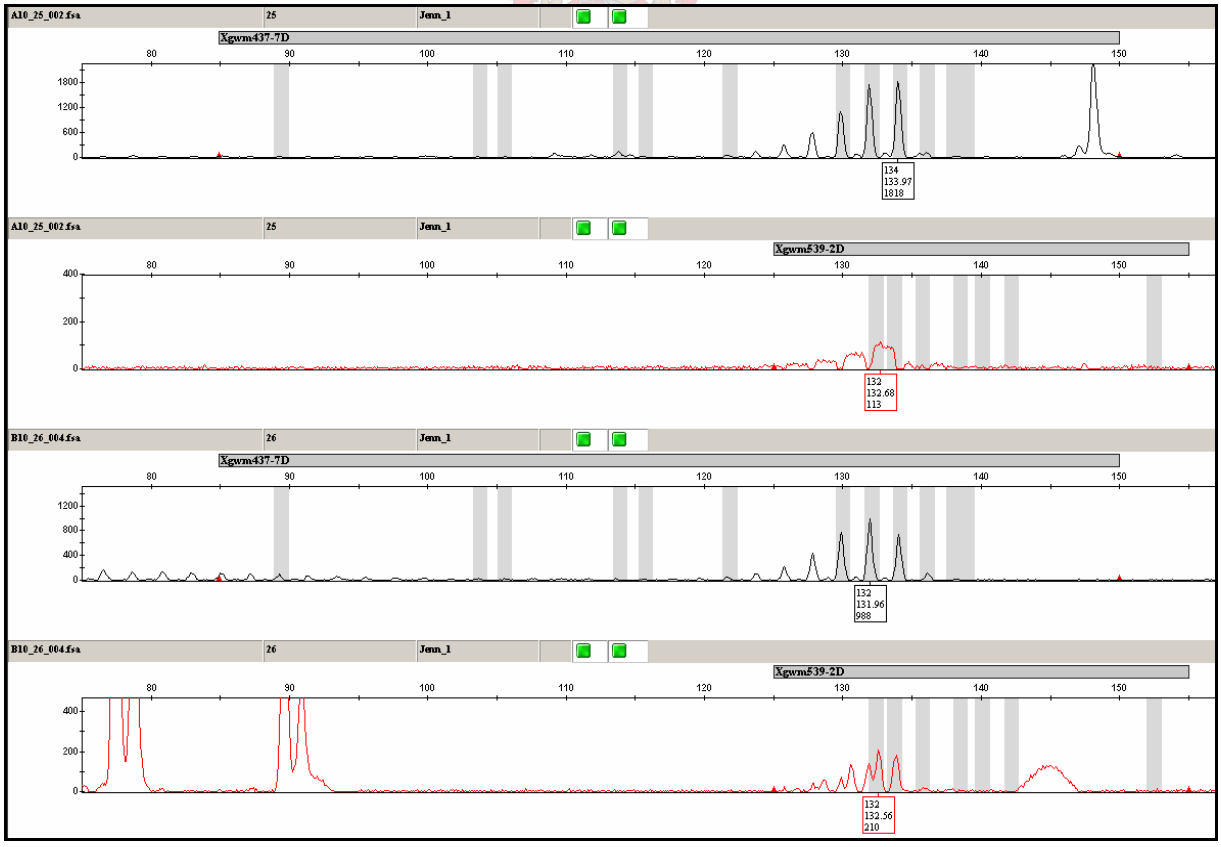


Figure 7-90 Sample number 25 and 26. (*Xgwm437*, *Xgwm539*)

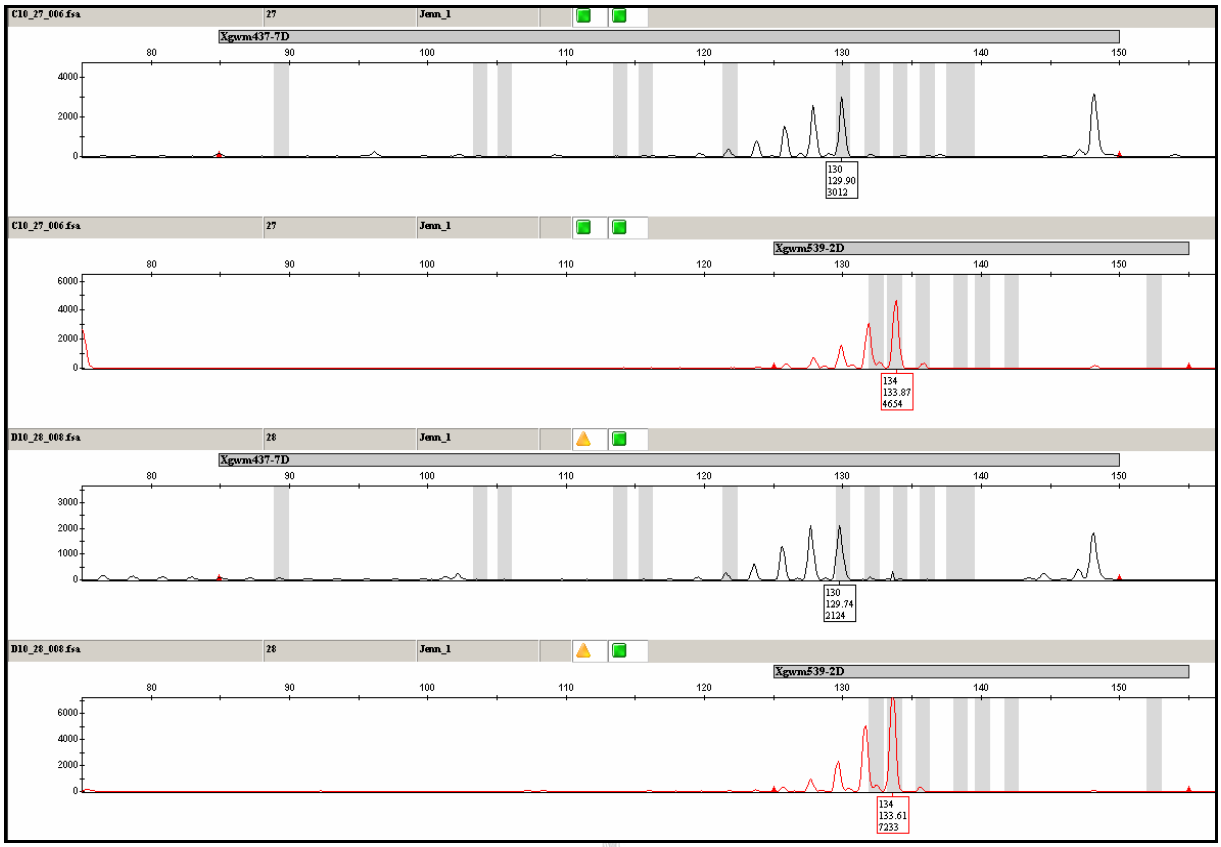


Figure 7-91 Sample number 27 and 28. (*Xgwm437*, *Xgwm539*)

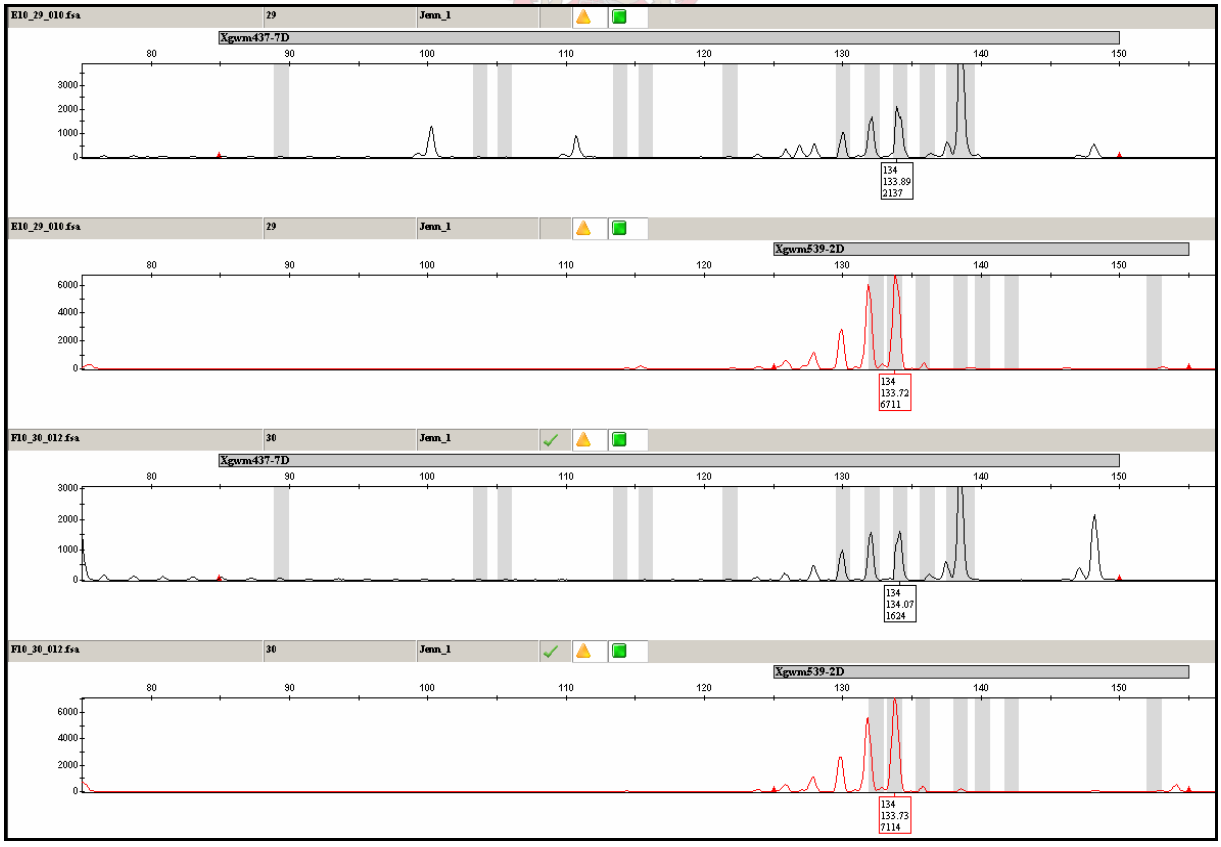


Figure 7-92 Sample number 29 and 30. (*Xgwm437*, *Xgwm539*)

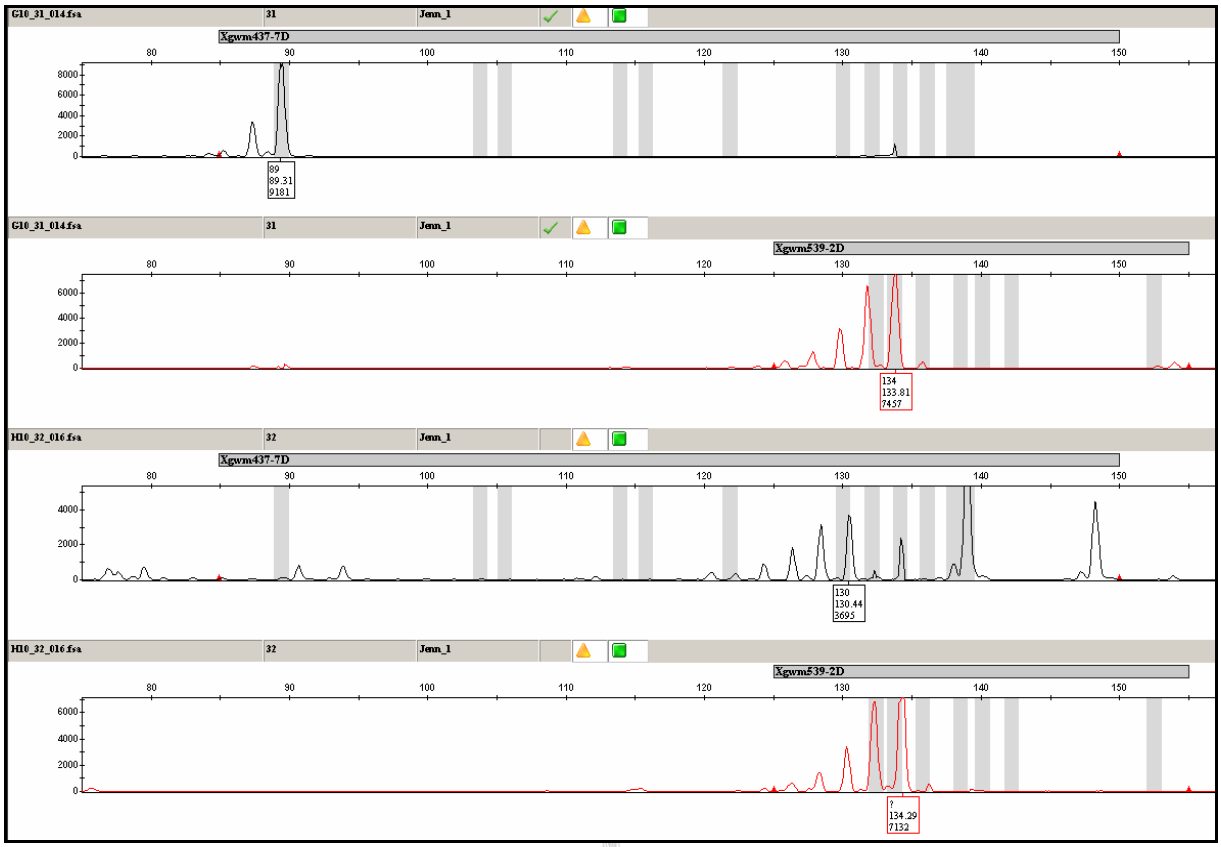


Figure 7-93 Sample number 31 and 32. (*Xgwm437*, *Xgwm539*)

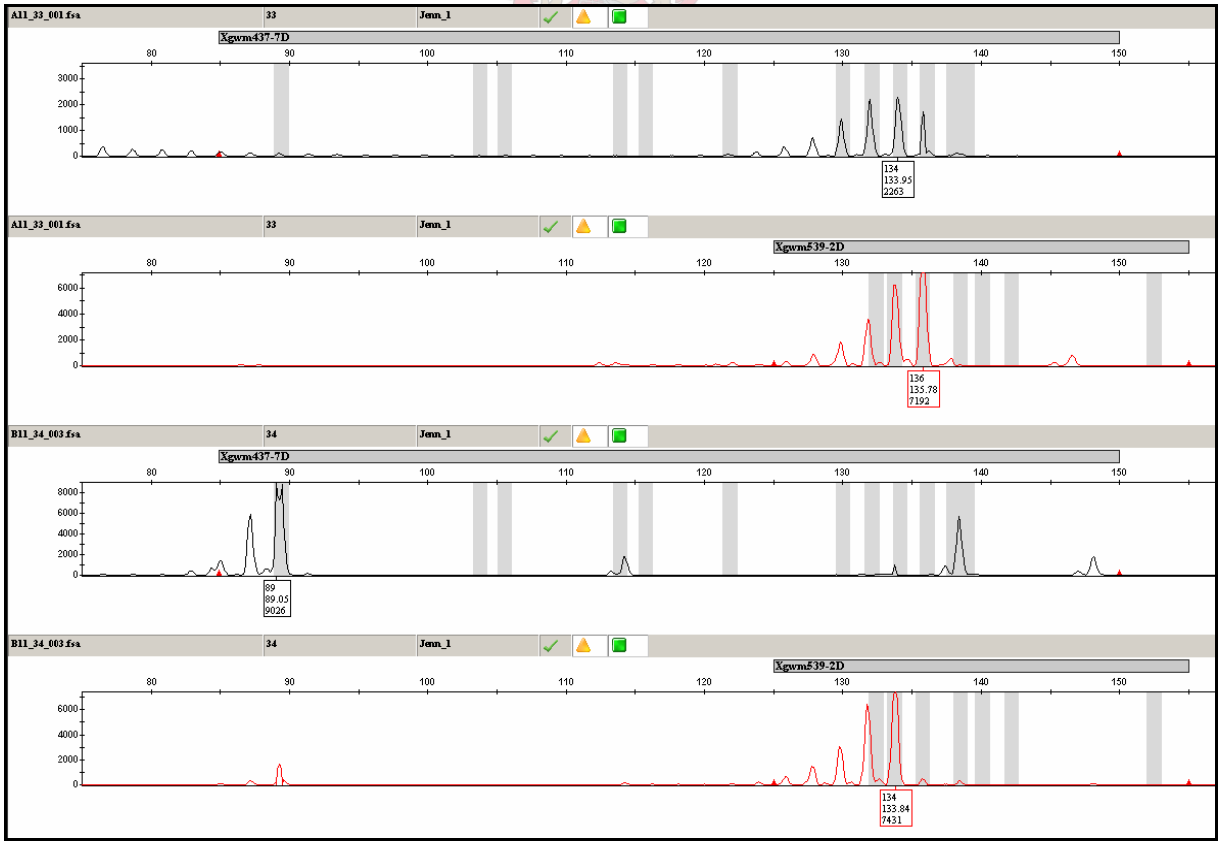


Figure 7-94 Sample number 33 and 34. (*Xgwm437*, *Xgwm539*)

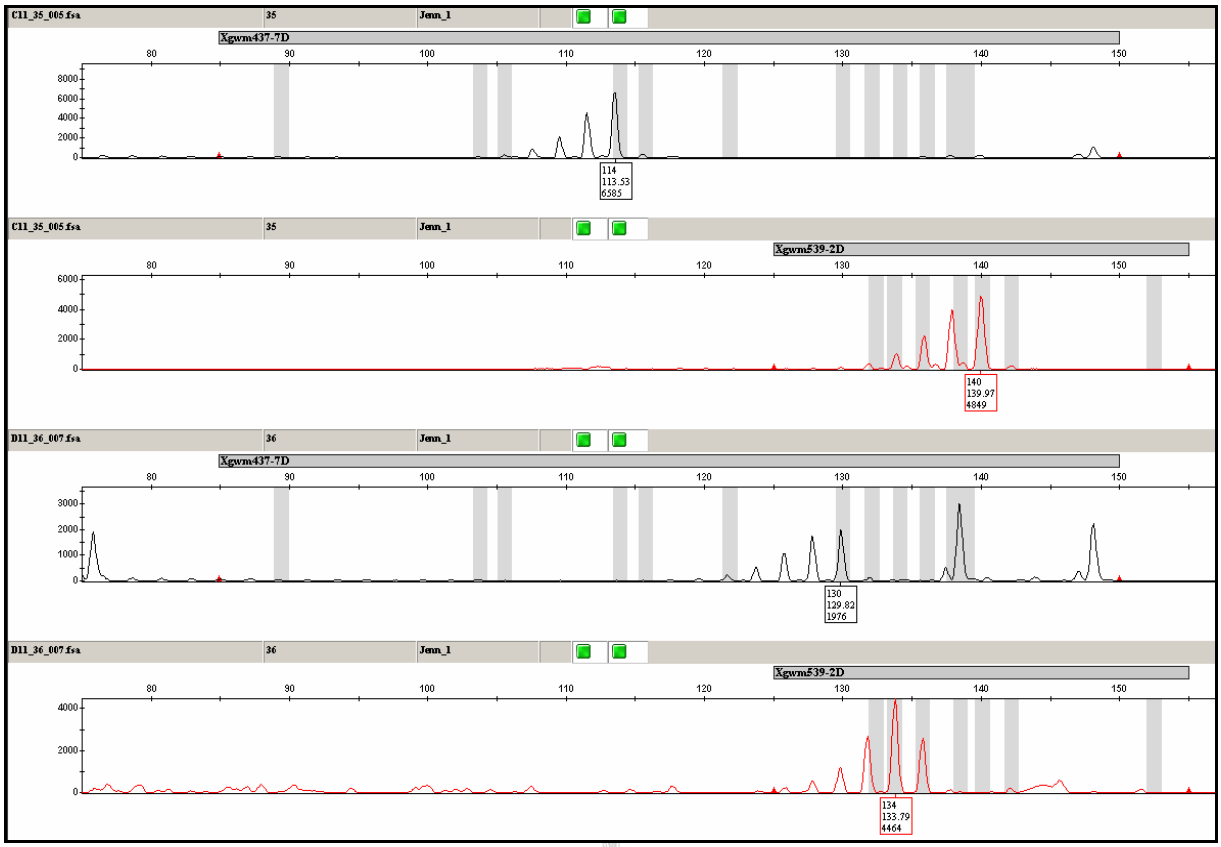


Figure 7-95 Sample number 35 and 36. (*Xgwm437*, *Xgwm539*)

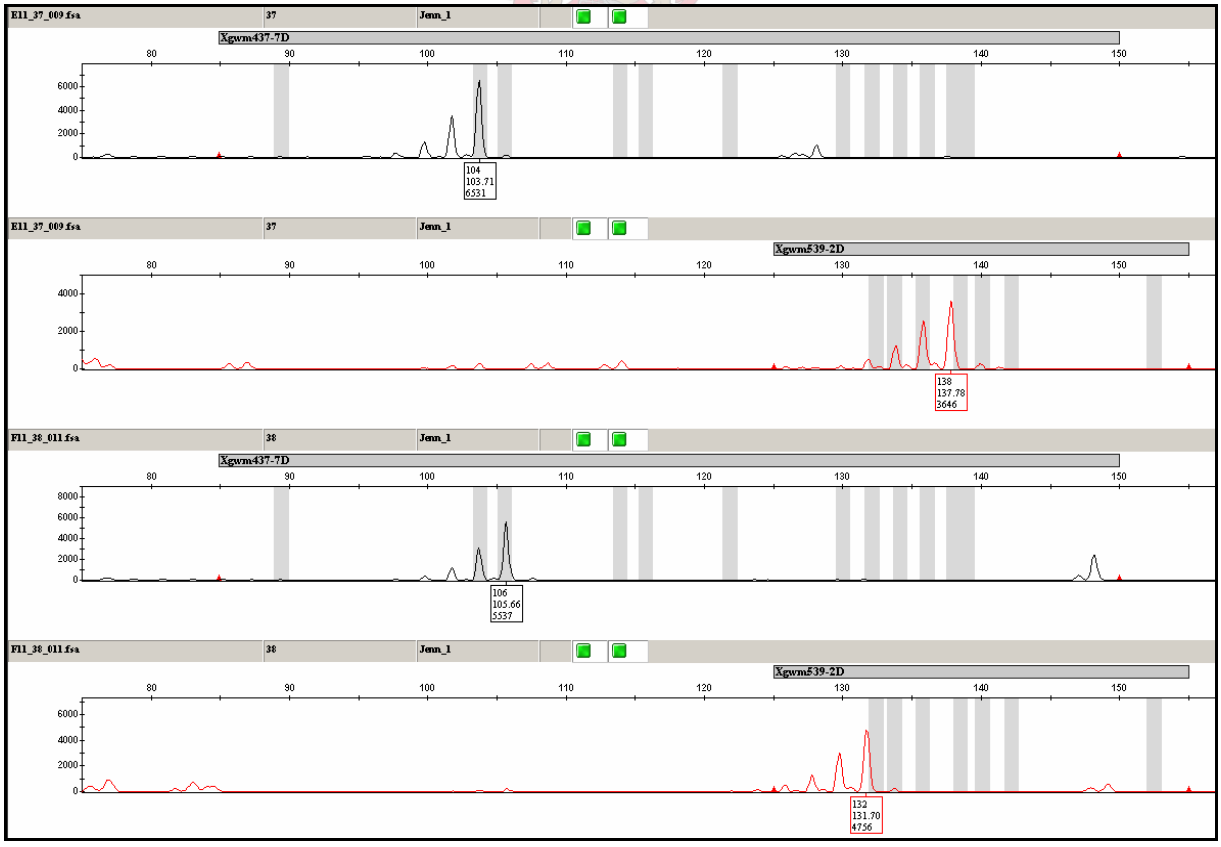


Figure 7-96 Sample number 37 and 38. (*Xgwm437*, *Xgwm539*)

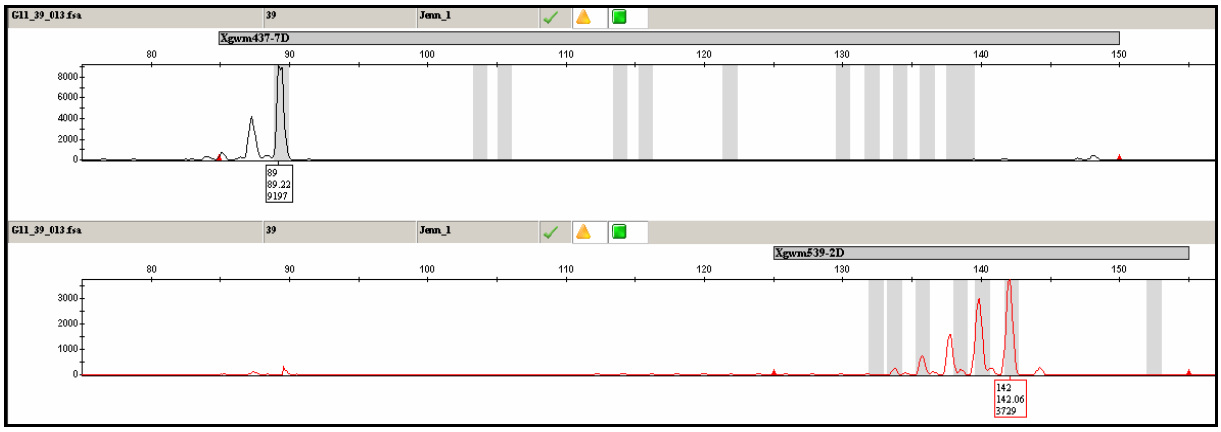


Figure 7-97 Sample number 39. (*Xgwm437*, *Xgwm539*)

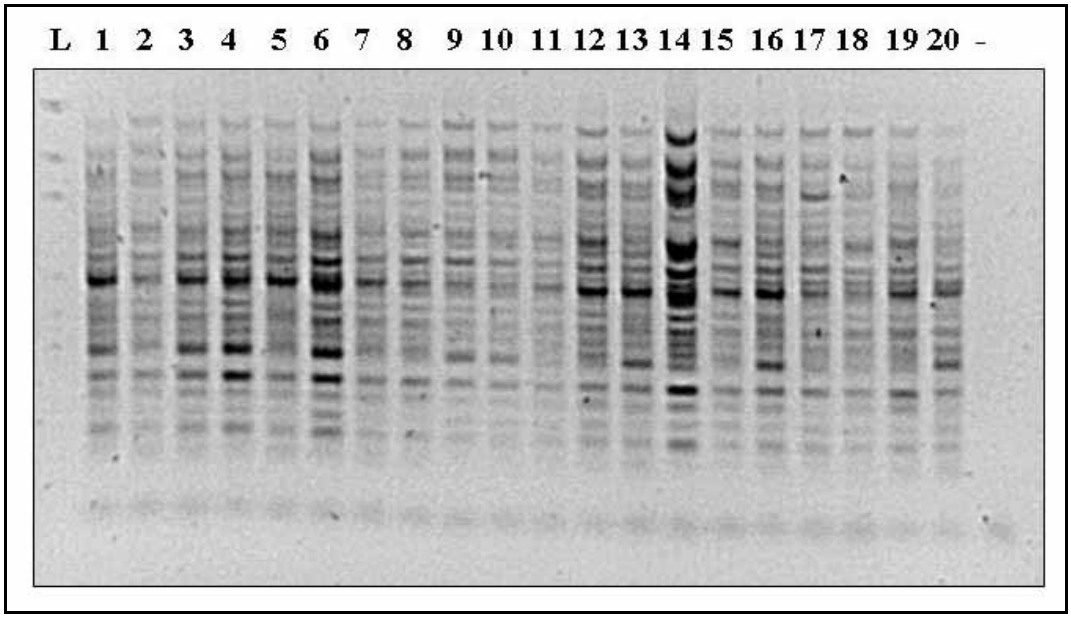


Figure 7-98 Gel photo illustrating the amplification of REMAP combination 2 (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

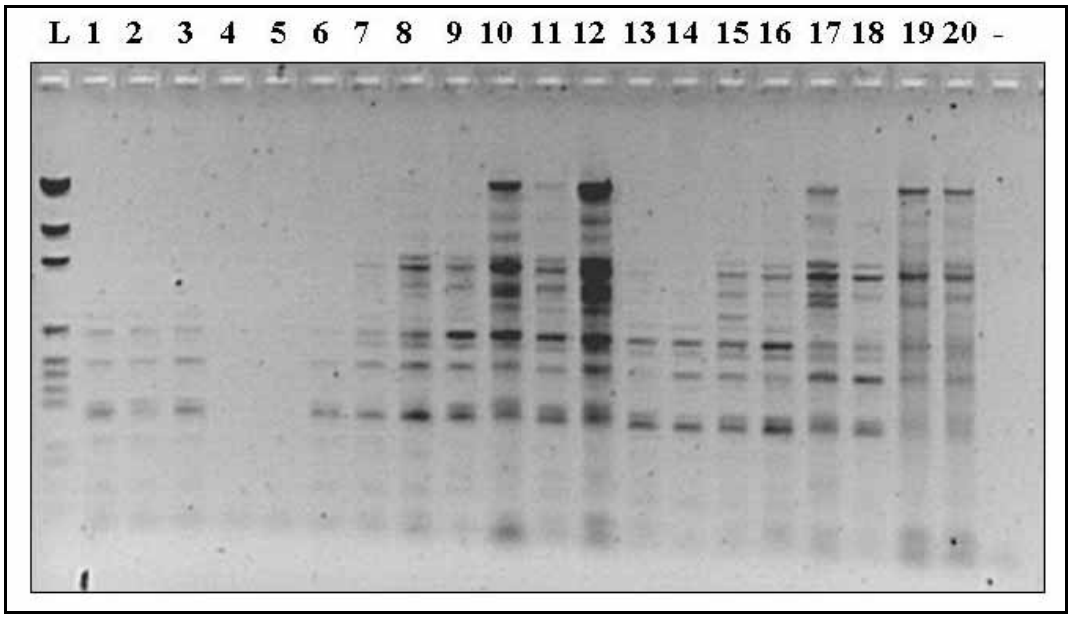


Figure 7-99 Gel photo illustrating the amplification of REMAP combination 3 (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

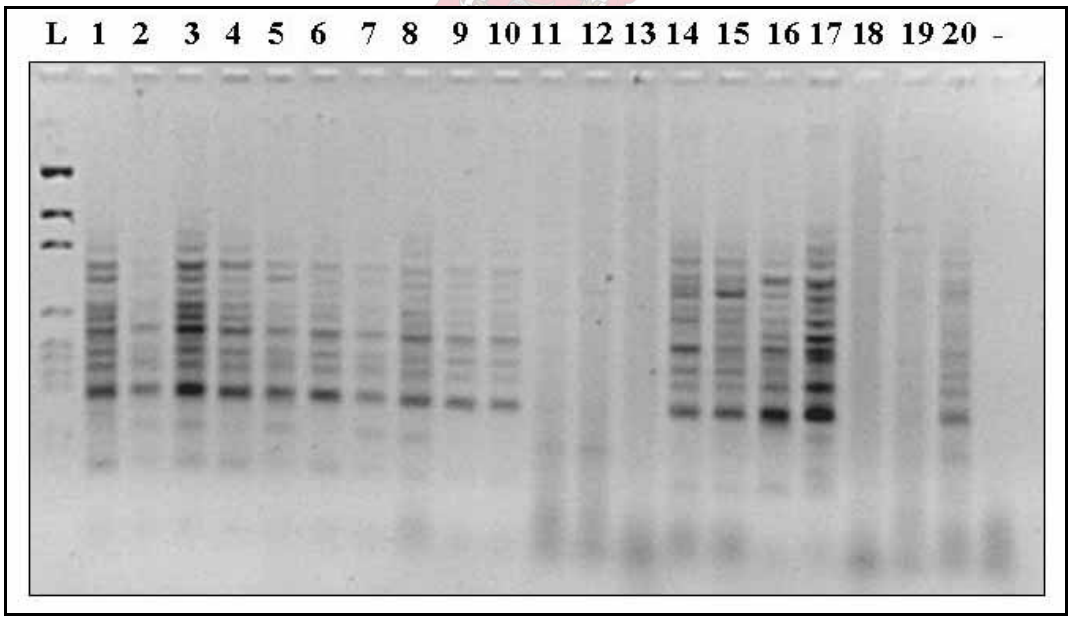


Figure 7-100 Gel photo illustrating the amplification of REMAP combination 4 (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

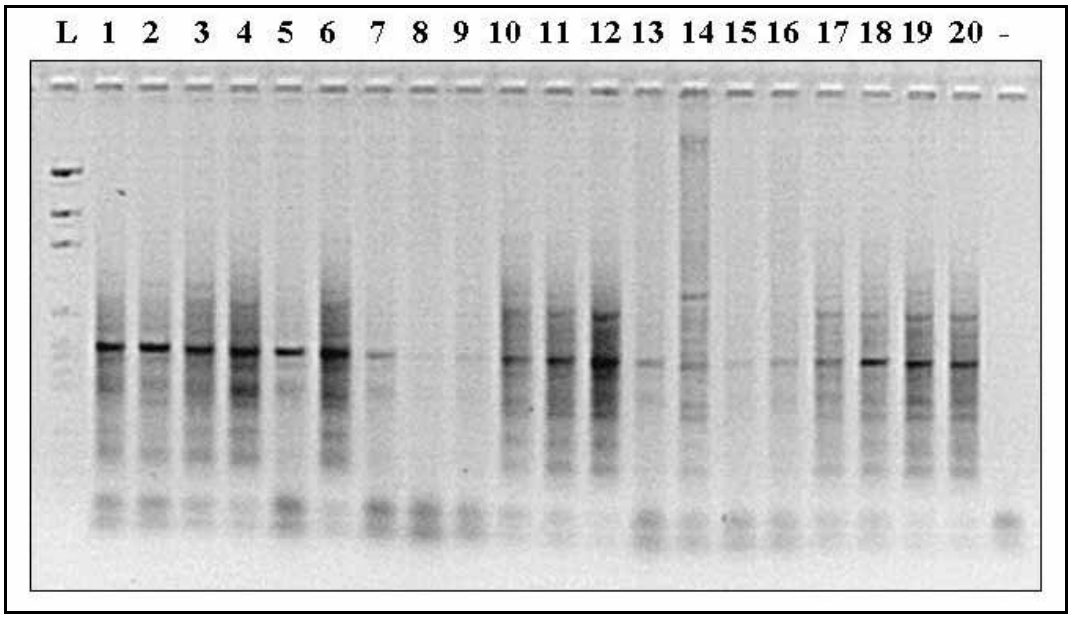


Figure 7-101 Gel photo illustrating the amplification of REMAP combination 5 (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

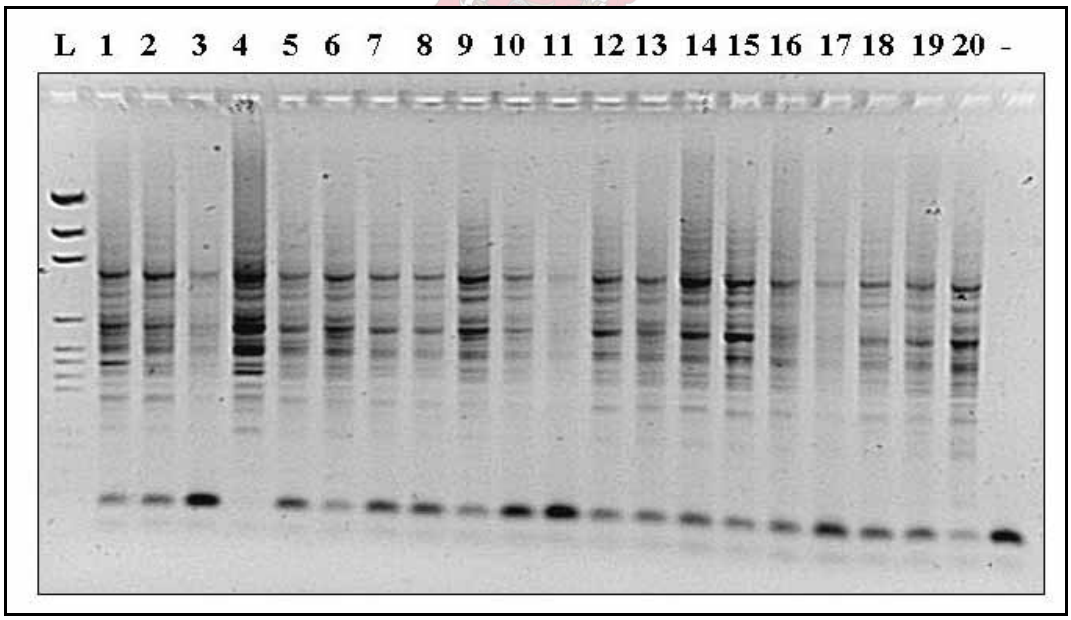


Figure 7-102 Gel photo illustrating the amplification of REMAP combination 6 (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

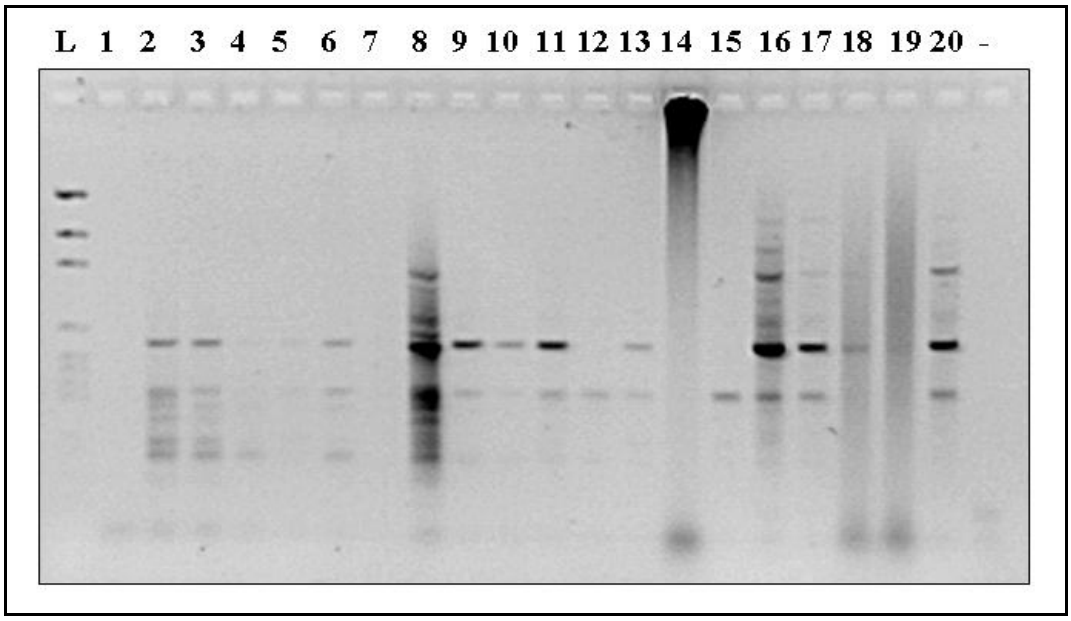


Figure 7-103 Gel photo illustrating the amplification of REMAP combination 7 (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

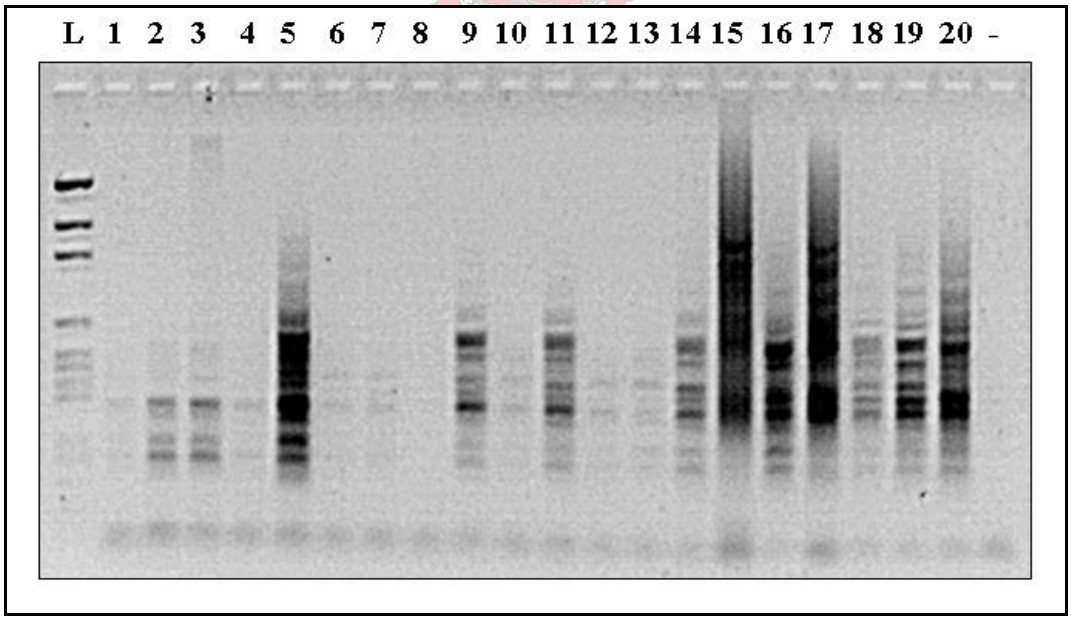


Figure 7-104 Gel photo illustrating the amplification of REMAP combination 8 (1.5% agarose gel electrophoresed at 100V for 45 min). The sample numbers are given above each lane. These numbers correspond to those in Table 3-1.

8. ADDENDUM B

This addendum contains tables of data obtained from the SAS program after a discriminant analysis and a cluster analysis was used for the two plant populations. Discriminant analysis was used for the first population on the *Xwmc177* data and all nine sets of AFLP data. The population of 119 breeding lines and the microsatellite data for the first population were compared with cluster analysis.

Table 8-1 Variables for microsatellite *Xwmc177-2A* when this marker was considered to be more like an AFLP.

Line	US										PANNAR						SGI			
	00K180		00K268		97K1		98K120		00K60		PAN 3404		PAN 3490		PAN 3492		Kariega		Biedou	
	1	21	2	22	3	23	4	24	5	25	6	26	7	27	8	28	9	29	10	30
V1	102	0	102	0	102	0	102	0	102	0	102	0	102	0	102	0	0	0	0	0
V2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V3	0	0	0	0	0	0	0	0	106	0	0	0	0	0	0	0	0	0	0	0
V4	0	0	0	0	0	0	0	0	109	0	0	0	0	0	0	0	0	0	0	0
V5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V7	0	0	0	0	0	0	0	0	0	0	119	0	0	0	0	0	0	0	0	0
V8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V11	0	0	0	0	0	0	0	0	132	0	0	0	0	0	0	0	0	0	0	0
V13	139	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	139	0
V14	141	0	0	0	0	0	0	0	0	0	141	0	141	0	141	0	141	0	141	0
V16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	146	0	0	0
V17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	149	0	0	0
V20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0
V21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V23	0	0	163	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V26	0	0	0	0	0	0	175	0	0	0	0	0	0	0	0	0	0	0	175	0
V27	0	0	0	0	0	177	177	177	0	0	177	0	0	0	0	0	177	177	177	0
V28	0	0	0	0	0	179	179	179	0	0	179	179	0	0	0	0	179	179	179	0
V29	0	0	0	0	0	180	180	180	0	0	180	180	0	0	0	0	180	180	180	0
V30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V31	0	0	184	0	0	0	0	0	0	0	184	0	0	0	0	0	0	0	0	0
V32	0	0	186	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V33	188	0	188	188	188	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V34	190	190	0	190	190	0	0	0	190	0	0	0	190	0	190	190	0	0	0	0
V35	192	192	0	192	192	0	0	0	192	0	0	0	192	192	192	192	0	0	0	0
V36	0	0	0	0	0	0	0	0	0	0	0	0	0	193	0	193	0	0	0	0
V37	0	0	0	0	0	0	0	0	194	194	0	0	194	194	194	194	0	0	0	0
V38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	199	0	0	0
V40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V41	237	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	386	0
V45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	424	0

Line	MONSANTO												CYMMIT						
	SST 015		SST 027		SST 57		SST 65		SST 88		SST 94		12th HRWYT 46		16 HRWSN-136		16 HRWSN-89		16 HRWSN-28
	11	31	12	32	13	33	14	34	15	35	16	36	17	37	18	38	19	39	20
V1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V2	0	0	0	0	0	0	0	0	105	0	0	0	0	0	0	0	0	0	0
V3	0	0	0	0	0	0	0	0	106	0	0	0	0	0	0	0	0	0	0
V4	0	0	0	0	0	0	0	109	109	0	0	0	0	0	0	109	0	0	109
V5	110	0	0	0	0	0	0	0	0	0	0	0	0	0	0	110	0	0	0
V7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V8	0	0	0	0	0	0	0	126	0	0	0	0	0	0	0	126	0	0	0
V9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	127	0	0	0
V10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	129	0	0	0	0
V11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V14	141	0	0	0	0	0	141	0	141	0	0	0	141	0	141	0	141	0	0
V16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	146	0	0	0
V17	0	0	0	0	0	0	0	0	0	147	0	0	0	0	0	0	0	0	0
V18	0	0	0	0	0	0	0	0	0	0	0	0	148	0	0	0	0	0	0
V19	0	0	0	0	0	0	0	0	0	149	0	0	0	0	0	0	0	0	0
V20	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0	150
V21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	151	0	0	0	151
V22	0	0	0	0	0	0	0	0	0	152	152	0	0	0	0	0	0	0	0
V23	0	0	163	0	163	163	0	0	163	0	163	0	0	0	163	163	0	0	0
V24	0	0	168	0	0	0	0	0	168	0	0	0	0	0	168	0	0	0	0
V25	0	0	0	0	0	0	174	0	0	0	0	0	0	0	0	0	174	0	0
V26	175	0	0	0	0	0	175	0	0	0	0	0	0	0	0	0	175	0	0
V27	177	177	0	0	0	0	177	0	0	0	0	0	0	0	0	0	177	0	0
V28	179	179	0	0	0	0	179	0	0	0	0	0	179	0	0	0	179	0	0
V29	180	180	0	0	0	0	180	0	0	0	0	0	180	0	0	180	180	0	0
V30	0	0	0	0	0	0	0	0	0	0	0	0	182	0	0	0	182	0	0
V31	0	0	184	0	0	0	0	0	0	184	0	0	184	0	0	0	0	0	0
V32	0	0	186	186	0	0	0	0	186	186	186	0	186	186	186	0	0	0	186
V33	0	0	188	188	188	188	0	0	188	188	188	0	0	0	188	0	0	0	188
V34	0	0	0	0	190	190	0	0	190	190	190	0	0	0	190	0	0	0	0
V35	0	0	0	0	192	192	0	0	192	192	192	0	0	0	192	0	0	0	0
V36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	198
V39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	199
V40	0	0	205	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V42	0	0	243	0	0	0	0	0	243	0	0	0	0	0	0	0	243	0	243
V43	0	0	0	0	0	0	0	0	0	0	320	0	0	0	0	0	0	0	0
V44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V45	0	0	0	0	0	0	424	0	424	0	424	0	0	0	0	0	0	0	0

Table 8-2 Probabilities and reclassifications for each sample using only microsatellite *Xwmc177-2A* as determined by discriminant analysis and SAS.

Sample Name	Name of Line	Classified	00K180	00K268	97K1	98K120	00K60	PAN3404	PAN3490	PAN3492
			Kariega	Biedou	SST015	SST027	SST57	SST65	SST88	SST94
			12HRWYT4	16HRWSN1	16HRWSN8	16HRWSN2				
1	00K180	00K180	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
21	00K180	00K180	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
2	00K268	00K268	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
22	00K268	00K268	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				

Sample Name	Name of Line	Classified	00K180	00K268	97K1	98K120	00K60	PAN3404	PAN3490	PAN3492
			Kariega	Biedou	SST015	SST027	SST57	SST65	SST88	SST94
			12HRWYT4	16HRWSN1	16HRWSN8	16HRWSN2				
3	97K1	97K1	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
23	97K1	Kariega	0.000	0.000	0.250	0.250	0.000	0.000	0.000	0.000
			0.250	0.000	0.250	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
4	98K120	98K120	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
24	98K120	Kariega	0.000	0.000	0.250	0.250	0.000	0.000	0.000	0.000
			0.250	0.000	0.250	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
5	00K60	00K60	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
25	00K60	00K60	0.000	0.000	1.000	0.000	1.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
6	PAN3404	PAN3404	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
26	PAN3404	PAN3404	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
7	PAN3490	PAN3490	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.500
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
27	PAN3490	PAN3490	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
8	PAN3492	PAN3490	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.500
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
28	PAN3492	PAN3492	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
9	Kariega	Kariega	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
29	Kariega	Kariega	0.000	0.000	0.250	0.250	0.000	0.000	0.000	0.000
			0.250	0.000	0.250	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
10	Biedou	Biedou	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
30	Biedou	SST94	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.333	0.000	0.000	0.000	0.000	0.000	0.333
			0.000	0.000	0.333	0.000				
11	SST015	SST015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
31	SST015	Kariega	0.000	0.000	0.250	0.250	0.000	0.000	0.000	0.000
			0.250	0.000	0.250	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
12	SST027	SST027	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
32	SST027	SST027	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
13	SST57	SST57	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
33	SST57	SST57	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000				
14	SST65	SST65	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
			0.000	0.000	0.000	0.000				
34	SST65	SST65	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000

Sample Name	Name of Line	Classified	00K180	00K268	97K1	98K120	00K60	PAN3404	PAN3490	PAN3492
			Kariega	Biedou	SST015	SST027	SST57	SST65	SST88	SST94
			12HRWYT4	16HRWSN1	16HRWSN8	16HRWSN2				
			0.000	0.000	0.000	0.000				
15	SST88	SST88	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
			0.000	0.000	0.000	0.000				
35	SST88	SST88	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
			0.000	0.000	0.000	0.000				
16	SST94	SST94	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
			0.000	0.000	0.000	0.000				
36	SST94	SST94	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.333	0.000	0.000	0.000	0.000	0.000	0.333
			0.000	0.000	0.333	0.000				
17	12HRWYT4	12HRWYT4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			1.000	0.000	0.000	0.000				
37	12HRWYT4	12HRWYT4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			1.000	0.000	0.000	0.000				
18	16HRWSN1	16HRWSN1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	1.000	0.000	0.000				
38	16HRWSN1	16HRWSN1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	1.000	0.000	0.000	0.000	0.000		
19	16HRWSN8	16HRWSN8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	1.000	0.000				
39	16HRWSN8	SST94	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.333	0.000	0.000	0.000	0.000	0.000	0.333
			0.000	0.000	0.333	0.000				
20	16HRWSN2	16HRWSN2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			0.000	0.000	0.000	1.000				

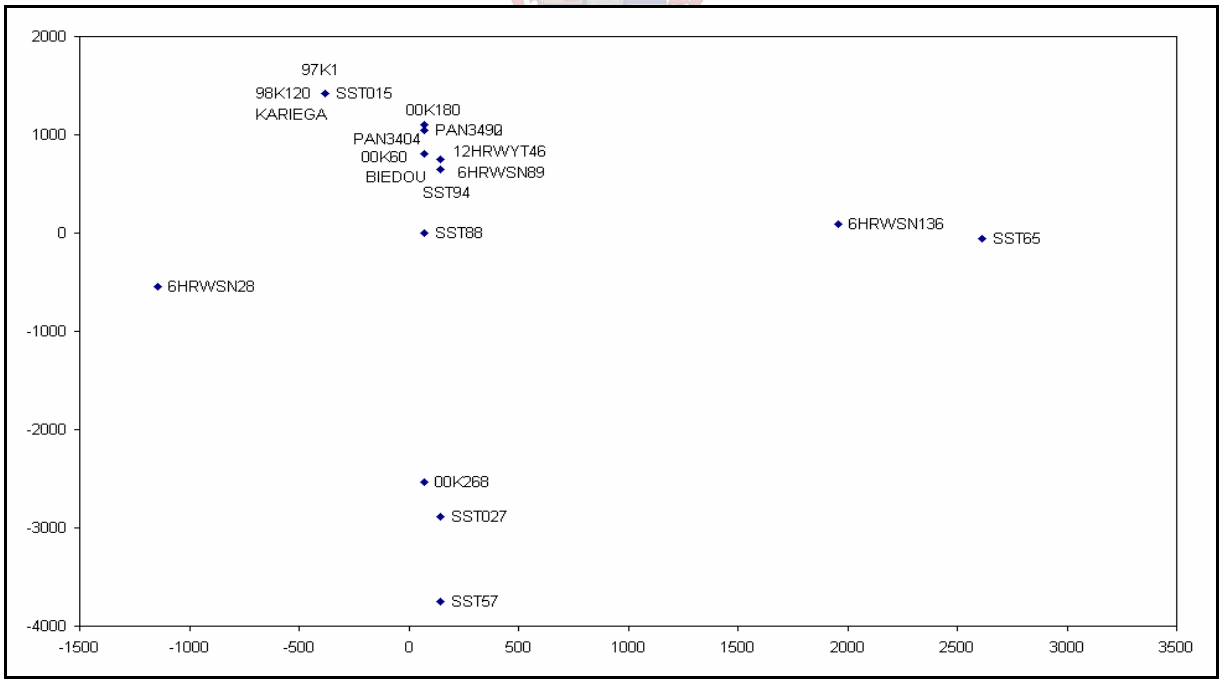


Figure 8-1 Initial sample distribution using only microsatellite *Xwmc177-2A* and SAS software.

Table 8-3 Banding pattern results for primer combination *Mse*1-CAG and *Eco*R1-ACA.

CAG-ACA	US					PANNAR			SGI		MONSANTO						CYMMIT			
LINE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
V1	0	0	0	0	59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V2	60	0	0	0	60	0	0	0	0	60	0	60	0	0	0	60	0	0	60	0
V3	0	64	0	0	0	0	0	0	0	0	0	0	0	64	0	0	0	64	0	0
V4	65	0	65	65	65	65	65	65	65	0	65	65	65	0	65	65	0	0	65	65
V5	0	68	0	0	0	0	0	0	0	0	0	0	0	68	0	0	0	68	0	0
V6	0	84	0	0	84	0	0	0	0	0	0	0	0	84	0	0	84	84	0	0
V7	89	0	89	89	0	89	0	89	89	0	0	89	0	0	0	89	0	0	89	89
V8	0	0	0	0	96	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V9	0	0	0	0	0	0	102	0	0	0	0	0	0	0	0	0	0	0	0	0
V10	107	0	107	107	0	107	0	107	107	0	107	107	107	0	107	107	0	0	107	107
V11	0	110	0	0	110	0	0	0	0	110	0	0	0	110	0	0	110	110	0	0
V12	0	0	0	0	129	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	134	0
V14	181	0	181	181	0	181	0	181	181	0	181	181	181	0	181	181	181	0	181	181
V15	0	203	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V16	233	0	233	233	0	233	233	233	233	0	0	233	0	0	0	233	0	0	0	233

Table 8-4 Linear discriminant function for each breeding programme obtained using SAS software for primer combination *Mse*1-CAG and *Eco*R1-ACA.

VARIABLE	US	PANNAR	SGI	MONSANTO	CYMMIT
Constant	-78346332	-78346331	-78346322	-78346371	-78346343
V1	-281988	-281988	-281988	-281989	-281988
V2	0.45177	0.45177	0.45177	1.15177	0.68510
V3	212984	212984	212984	212984	212984
V4	118729	118729	118729	1187290	1187290
V5	200456	200456	200455	200456	200456
V6	-324546	-324546	-324546	-324546	-324546
V7	-74456	-74456	-74456	-74456	-74456
V8	-173305	-173305	-173305	-173305	-173305
V9	714630	714630	714630	714630	714630
V10	488382	488382	488382	488382	488382
V11	1424478	1424478	1424478	1424478	1424478
V12	-128971	-128971	-128971	-128971	-128971
V13	49452	49452	49452	49451	49452
V14	150618	150618	150618	150619	150619
V15	-0.03714	-0.07162	-0.07162	-0.20955	-0.14059
V16	28440	28440	28440	28440	28440

Table 8-5 Number of observations and percent classified into each line for primer combination *Mse*1-CAG and *Eco*R1-ACA.

	Into Line					
From Line	**US**	**PANNAR**	**SGI**	**MONSANTO**	**CYMMIT**	**Total**
US	2 50.00	2 50.00	0 0.00	0 0.00	0 0.00	4 100.00
PANNAR	0 0.00	3 100.00	0 0.00	0 0.00	0 0.00	3 100.00
SGI	0 0.00	1 50.00	1 50.00	0 0.00	0 0.00	2 100.00
MONSANTO	0 0.00	0 0.00	0 0.00	5 83.33	1 16.67	6 100.00
CYMMIT	0 0.00	1 25.00	0 0.00	0 0.00	3 75.00	4 100.00
Total	4 10.53	7 36.84	1 5.26	5 26.32	4 21.05	19 100.00

Table 8-6 Banding pattern results for primer combination *Mse*1-CAG and *Eco*R1-AAC.

CAG-AAC	**US**					**PANNAR**			**SGI**		**MONSANTO**						**CYMMIT**			
LINE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
V1	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V2	57	0	0	0	57	0	0	0	0	57	57	57	57	0	57	57	0	0	57	57
V3	58	0	58	58	58	58	58	58	58	58	58	58	58	0	58	58	0	0	58	58
V4	65	0	65	65	0	65	65	65	65	0	0	0	0	0	0	0	0	0	0	65
V5	67	0	67	67	67	67	67	67	67	67	67	67	67	0	67	67	67	0	67	67
V6	69	0	69	69	69	69	69	69	69	69	69	69	69	0	69	69	69	0	69	69
V7	83	0	83	83	0	83	83	83	83	0	0	83	83	0	83	83	0	0	83	83
V8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	84	0
V9	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V10	106	0	106	106	106	0	0	106	106	0	0	106	106	0	0	106	0	0	0	0
V11	0	0	0	0	107	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V12	112	0	0	112	0	112	0	0	112	0	112	112	0	0	112	112	0	0	112	112
V13	0	0	0	0	221	0	221	0	0	0	0	0	221	0	0	0	0	0	0	0
V14	0	0	0	0	0	0	0	310	0	0	0	0	0	0	0	0	0	0	0	0
V15	0	0	0	0	386	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V16	0	0	0	0	0	0	0	388	0	0	0	0	388	0	0	0	0	0	0	0

Table 8-7 Linear discriminant function for each breeding programme obtained using SAS software for primer combination *Mse*1-CAG and *Eco*R1-AAC.

VARIABLE	**US**	**PANNAR**	**SGI**	**MONSANTO**	**CYMMIT**
Constant	-4.12500	-17.77778	-6.25000	-5.48611	-4.23611
V1	0.02500	-0.02222	-0.03333	-0.05556	0.03056
V2	-2.324E-16	-0.23392	-0.08772	-0.05848	0.05848
V3	0.06466	0.20115	0.30172	0.17960	-0.24425

VARIABLE	US	PANNAR	SGI	MONSANTO	CYMMIT
Constant	-4.12500	-17.77778	-6.25000	-5.48611	-4.23611
V4	0.09615	-0.00855	-0.01282	-0.19444	0.09829
V5	0.01866	0.04561	0.01244	0.00207	0.09121
V6	0.01812	0.04428	0.01208	0.00201	0.08857
V7	-0.04518	0.06024	-0.09036	0.01506	0.03012
V8	0.04464	-0.07937	-0.02976	-0.15377	0.10913
V9	0.01250	-0.01111	-0.01667	-0.02778	0.01528
V10	0.03538	-0.12579	0.02358	0.00393	-0.03931
V11	0.01168	-0.01038	-0.01558	-0.02596	0.01428
V12	-0.03348	0.02976	-0.02232	0.04092	-0.00744
V13	-0.01697	0.08296	-0.01131	0.02074	-0.00377
V14	-0.02016	0.13620	-0.01344	0.01389	0.00627
V15	0.00324	-0.00288	-0.00432	-0.00720	0.00396
V16	-0.00322	-0.02721	-0.00215	-0.0003580	0.00358

Table 8-8 Number of observations and percent classified into each line for primer combination *Mse*1-CAG and *Eco*R1-AAC.

	Into Line					
From Line	**US**	**PANNAR**	**SGI**	**MONSANTO**	**CYMMIT**	**Total**
US	5 100.00	0 0.00	0 0.00	0 0.00	0 0.00	5 100.00
PANNAR	0 0.00	3 100.00	0 0.00	0 0.00	0 0.00	3 100.00
SGI	1 50.00	0 0.00	1 50.00	0 0.00	0 0.00	2 100.00
MONSANTO	1 16.67	0 0.00	0 0.00	5 83.33	0 0.00	6 100.00
CYMMIT	1 25.00	0 0.00	0 0.00	0 0.00	3 75.00	4 100.00
Total	8 40.00	3 15.00	1 5.00	5 25.00	3 15.00	20 100.00

Table 8-9 Banding pattern results for primer combination *Mse*1-CAG and *Eco*R1-AGG.

CAG-AGG	US					PANNAR			SGI		MONSANTO						CYMMIT			
LINE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
V1	51	0	0	0	51	0	0	0	0	51	51	51	51	0	51	51	51	0	51	0
V2	52	0	52	52	52	52	52	52	52	52	52	52	52	0	52	52	52	0	52	52
V3	58	0	0	0	0	58	0	0	0	0	58	0	58	0	58	58	58	0	58	58
V5	62	0	0	0	0	0	0	0	0	0	62	62	0	0	62	62	0	0	62	0
V6	0	0	0	0	0	0	0	0	0	0	67	67	67	0	67	67	67	0	67	67
V7	68	0	68	68	0	68	0	68	68	0	0	68	68	0	68	68	0	0	68	68
V8	0	0	0	0	70	0	0	0	0	70	0	0	0	0	0	0	70	0	0	0
V9	0	0	0	0	71	0	0	0	0	71	0	0	0	0	0	71	71	0	0	0
V11	80	0	80	80	0	80	80	80	80	0	80	80	80	0	80	80	0	0	80	80

CAG-AGG	US					PANNAR			SGI		MONSANTO						CYMMIT			
LINE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
V12	83	0	83	83	0	83	0	83	0	0	0	0	0	0	0	0	0	0	83	83
V14	85	0	85	85	0	85	85	85	85	85	85	85	85	0	85	85	0	0	85	85
V16	91	0	0	0	91	0	0	0	0	91	91	91	91	0	0	91	91	0	91	0
V17	95	0	95	95	95	95	0	95	0	95	95	95	95	0	95	95	0	0	95	95
V18	96	0	96	96	96	96	0	96	96	96	96	96	96	0	96	96	96	0	96	96
V19	100	0	100	100	100	100	100	100	100	0	100	0	100	0	100	100	100	0	100	100
V20	101	0	0	101	0	0	101	101	0	0	101	0	0	0	101	0	0	0	101	101
V21	0	0	0	0	110	0	0	0	0	110	0	0	0	0	110	0	110	0	0	0
V22	111	0	0	111	111	111	0	0	0	111	111	111	111	0	111	111	111	0	111	111
V23	112	0	112	112	0	112	0	112	112	0	112	112	112	0	112	112	0	0	112	112
V24	0	0	0	0	0	0	113	0	0	0	113	113	113	0	0	0	113	0	0	0
V25	0	0	0	0	0	0	0	0	0	117	0	117	0	0	0	0	117	0	0	0
V26	118	0	118	118	118	118	0	0	0	118	118	118	118	0	118	118	0	0	118	118
V28	0	0	0	0	128	0	0	0	0	128	0	128	0	0	0	128	128	0	0	0
V34	0	0	0	0	175	0	0	0	0	175	0	175	175	0	0	0	175	0	0	0
V36	183	0	0	0	0	0	0	0	0	0	0	183	0	0	0	183	0	0	183	0
V37	0	0	0	207	0	207	0	207	207	0	0	0	0	0	0	0	0	0	0	0
V40	0	0	380	0	0	0	380	380	380	0	0	0	0	0	0	0	0	0	0	0

Table 8-10 Linear discriminant function for each breeding programme obtained using SAS software for primer combination *Mse*1-CAG and *Eco*R1-AGG.

VARIABLE	US	PANNAR	SGI	MONSANTO	CYMMIT
Constant	-6.00000	-27852507	-69891608	-6.25000	-5.62500
V1	-0.01012	-234093	-314639	0.06146	-0.06146
V2	0.08591	674237	1152659	-0.05414	0.05414
V3	-0.00869	761691	385788	0.01948	-0.01948
V5	0.00225	82574	-290304	0.03656	-0.03656
V6	-0.22388	-0.31970	-0.35145	-2.907E-16	0.22388
V7	0.00393	-380781	-497239	0.02307	-0.02307
V8	0.02158	210102	880394	-0.08209	0.08209
V9	-0.00161	-387544	-552049	0.05317	-0.05317
V11	0.02244	154466	-12821	0.02225	-0.02225
V12	0.0068	20112	-669786	-0.10777	0.10777
V14	0.00616	261831	977629	0.05509	-0.05509
V16	-0.00344	-109565	323826	-0.00239	0.00239
V17	-0.00399	407328	-556144	0.04500	-0.04500
V18	0.01479	-353739	425485	0.00728	-0.00728
V19	0.03010	-257124	-560485	-0.03012	0.03012
V20	0.00790	-114128	-44463	-0.00644	0.00644
V21	0.00801	77168	-195880	0.00595	-0.00595
V22	0.01215	-17530	111143	-0.01013	0.01013
V23	0.00136	-286836	-51886	0.03315	-0.03315
V24	0.01293	-65789	-750683	0.02159	-0.02159
V25	0.00223	-180355	5438	-0.00508	0.00508
V26	0.01353	-195372	-76114	-0.01102	0.01102
V28	0.01107	137953	-163184	0.01273	-0.01273

VARIABLE	US	PANNAR	SGI	MONSANTO	CYMMIT
Constant	-6.00000	-27852507	-69891608	-6.25000	-5.62500
V34	0.00383	111347	64701	0.0002751	-0.0002751
V36	0.00377	91747	25013	-0.02833	0.02833
V37	-0.00244	213421	108095	0.00546	-0.00546
V40	0.00432	80804	79531	-0.00170	0.00170

Table 8-11 Number of observations and percent classified into each line for primer combination *Mse*1-CAG and *Eco*R1-AGG.

	Into Line					
From Line	**US**	**PANNAR**	**SGI**	**MONSANTO**	**CYMMIT**	**Total**
US	4 80.00	0 0.00	0 0.00	0 0.00	1 20.00	5 100.00
PANNAR	0 0.00	3 100.00	0 0.00	0 0.00	0 0.00	3 100.00
SGI	0 0.00	0 0.00	2 100.00	0 0.00	0 0.00	2 100.00
MONSANTO	0 0.00	0 0.00	0 0.00	5 83.33	1 16.67	6 100.00
CYMMIT	0 0.00	0 0.00	0 0.00	0 0.00	4 100.00	4 100.00
Total	4 20.00	3 15.00	2 10.00	5 25.00	6 30.00	20 100.00

Table 8-12 Banding pattern results for primer combination *Mse*1-CAT and *Eco*R1-ACA.

CAT-ACA	US					PANNAR			SGI		MONSANTO						CYMMIT			
LINE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
V1	0	54	0	0	0	54	0	0	0	0	0	0	0	0	0	54	0	0	54	0
V3	0	65	0	0	65	0	0	0	0	0	0	0	0	0	0	0	0	65	0	0
V4	78	0	78	78	0	78	78	78	78	78	78	78	78	78	78	78	78	78	78	78
V6	80	0	80	80	0	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
V7	0	0	0	0	0	82	82	0	82	0	82	82	0	82	82	82	82	82	82	82
V8	0	84	0	0	84	0	0	0	84	0	0	0	0	0	0	0	0	0	0	0
V9	0	0	0	0	0	0	0	0	0	0	0	98	0	0	0	98	0	0	0	0
V10	0	99	99	99	99	99	99	0	0	99	99	99	0	99	0	99	0	99	0	99
V12	102	0	102	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V16	0	112	0	0	112	0	0	0	0	0	0	0	0	0	0	0	0	112	0	0
V17	0	0	0	0	0	0	0	0	0	125	0	0	0	0	0	125	0	0	0	0
V18	0	0	0	0	0	0	128	128	128	0	0	0	128	0	0	0	0	0	0	0
V19	139	0	139	139	0	139	139	139	139	139	0	0	139	139	0	0	139	0	139	0
V20	148	0	148	148	0	148	148	148	148	148	148	0	148	148	0	0	148	0	0	0
V21	0	158	0	0	158	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V23	0	0	0	0	0	0	0	0	0	0	165	165	0	165	0	0	165	0	0	165
V24	0	174	0	0	0	0	0	0	0	0	0	174	0	0	174	174	0	0	174	0
V25	182	0	182	182	0	182	182	182	182	182	182	182	182	182	182	182	0	0	0	182
V26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	219	219	0	0	0	219

Table 8-13 Linear discriminant function for each breeding programme obtained using SAS software for primer combination *Mse*1-CAT and *Eco*R1-ACA.

VARIABLE	US	PANNAR	SGI	MONSANTO	CYMMIT
Constant	-399862412	-337570709	-496131506	-337570727	-314919256
V1	-1468154	-838945	-2097363	-838945	419473
V3	0.35497	0.69735	0.35875	0.69735	0.57440
V4	3601819	3747021	3456618	3747021	4037426
V6	3511774	3653345	3370203	3653346	3936491
V7	-0.70755	0.11833	-0.61881	-0.06460	-0.42502
V8	1887629	1078644	2696613	1078645	-539320
V9	-0.36571	-0.34017	-1.61573	-0.64629	-1.20747
V10	0.32464	-0.07423	0.11745	-0.07423	0.03712
V12	0.20235	-0.13647	-0.04706	-0.13647	-0.07882
V16	0.20601	0.40471	0.20820	0.40471	0.33336
V17	634243	362425	906062	362425	-181211
V18	-619378	-353930	-884825	-353930	176965
V19	0.24016	0.21432	0.26601	0.21432	0.27054
V20	535678	306102	765253	306101	-153051
V21	4058000	3986318	4129681	3986318	3842953
V23	-480487	-274564	-686409	-274563	137283
V24	455634	260362	650906	260363	-130181
V25	871212	497836	1244589	497836	-248918
V26	-362010	-206863	-517158	-206863	103432

Table 8-14 Number of observations and percent classified into each line for primer combination *Mse*1-CAT and *Eco*R1-ACA.

	Into Line					
From Line	**US**	**PANNAR**	**SGI**	**MONSANTO**	**CYMMIT**	**Total**
US	5 100.00	0 0.00	0 0.00	0 0.00	0 0.00	5 100.00
PANNAR	0 0.00	3 100.00	0 0.00	0 0.00	0 0.00	3 100.00
SGI	0 0.00	0 0.00	2 100.00	0 0.00	0 0.00	2 100.00
MONSANTO	0 0.00	1 16.67	0 0.00	5 83.33	0 0.00	6 100.00
CYMMIT	0 0.00	0 0.00	0 0.00	0 0.00	4 100.00	4 100.00
Total	5 25.00	4 20.00	2 10.00	5 25.00	4 20.00	20 100.00

Table 8-15 Banding pattern results for primer combination *Mse*1-CAT and *Eco*R1-AAC.

CAT-AAC	US					PANNAR			SGI		MONSANTO						CYMMIT			
LINE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
V2	0	51	0	0	0	0	0	0	0	0	51	51	0	0	51	51	0	0	51	51
V3	52	0	52	52	0	52	52	52	52	52	52	52	52	52	52	52	52	52	52	52
V4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	57	0	0	0	0	57
V5	58	0	58	58	0	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58
V7	65	65	0	65	0	65	65	65	65	65	65	65	65	65	65	65	65	65	65	65
V10	0	96	0	0	0	0	96	0	0	0	0	0	0	0	0	0	0	0	0	0
V11	98	0	98	98	0	0	98	98	0	0	0	0	98	0	0	0	0	0	0	0
V12	99	0	99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V13	0	0	0	0	0	107	107	0	0	0	0	0	0	0	0	0	0	0	0	0
V14	0	111	0	111	0	111	111	0	0	0	111	111	0	111	111	111	0	0	0	0
V15	0	0	0	0	0	0	112	0	0	0	0	0	0	0	0	112	0	0	0	0
V19	0	0	0	0	0	0	0	0	0	0	0	173	0	0	173	173	173	0	0	173
V20	174	0	174	174	0	174	174	174	174	174	174	174	174	174	174	174	174	174	174	174
V22	0	221	0	0	221	0	0	0	0	221	0	0	0	0	0	0	0	221	0	0
V23	298	0	298	298	0	298	298	298	298	298	298	0	298	298	0	0	298	0	0	0
V24	299	0	299	299	0	299	299	299	299	299	299	0	299	299	0	0	299	0	0	0
V25	300	0	300	300	0	0	0	300	300	300	0	0	300	0	0	0	0	0	0	0
V26	0	0	302	302	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V27	0	0	0	0	0	0	0	0	0	384	0	0	384	0	0	0	0	0	0	0
V28	0	387	0	0	387	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V29	0	388	388	0	388	0	0	0	0	0	0	0	0	0	388	0	0	388	0	0
V31	493	0	493	0	0	0	0	493	493	493	0	0	493	493	0	0	493	493	0	0
V32	0	0	0	0	0	0	0	0	0	499	0	0	499	0	0	0	499	0	0	0

Table 8-16 Linear discriminant function for each breeding programme obtained using SAS software for primer combination *Mse*1-CAT and *Eco*R1-AAC.

VARIABLE	US	PANNAR	SGI	MONSANTO	CYMMIT
Constant	-308032343	-363338781	-436338507	-388758577	-465553272
V2	-315851	1705209	947554	1580583	1658938
V3	3126656	4705696	5594083	4776794	5328599
V4	64861	-199658	-194584	-692266	-62448
V5	2803209	4218900	5015384	4282643	4777365
V7	460209	-1050093	-1380626	-1068183	234450
V10	-481210	77199	1443629	74186	817327
V11	471389	-75623	-1414167	-72672	-800647
V12	829394	-1946696	-2488183	-2098995	-2537722
V13	-115993	706404	347982	384585	757442
V14	291810	-235321	-875432	-164862	-1606382
V15	0.04020	0.03767	0.14726	0.09883	0.03240
V19	-21370	65783	64112	228088	20575
V20	934403	1406300	1671795	1427548	1592455
V22	16729	-51495	-50187	-178548	-16106
V23	-6203	19095	18610	66207	5972
V24	-6182	19031	18547	65985	5952

VARIABLE	US	PANNAR	SGI	MONSANTO	CYMMIT
Constant	-308032343	-363338781	-436338507	-388758577	-465553272
V25	-33693	-100300	101080	-168790	-344682
V26	111294	-263697	-333881	-360567	38674
V27	-28884	88910	86651	308275	27809
V28	1591898	1379728	1260358	1370174	1296030
V29	-9529	29331	28586	101699	9174
V31	-32674	176401	98023	163509	171614
V32	14818	-45613	-44454	-158153	-14267

Table 8-17 Number of observations and percent classified into each line for primer combination *Mse*1-CAT and *Eco*R1-AAC.

	Into Line					
From Line	**US**	**PANNAR**	**SGI**	**MONSANTO**	**CYMMIT**	**Total**
US	5 100.00	0 0.00	0 0.00	0 0.00	0 0.00	5 100.00
PANNAR	0 0.00	3 100.00	0 0.00	0 0.00	0 0.00	3 100.00
SGI	0 0.00	0 0.00	2 100.00	0 0.00	0 0.00	2 100.00
MONSANTO	0 0.00	0 0.00	0 0.00	6 100.00	0 0.00	6 100.00
CYMMIT	0 0.00	0 0.00	0 0.00	0 0.00	4 100.00	4 100.00
Total	5 25.00	3 15.00	2 10.00	6 30.00	4 20.00	20 100.00

Table 8-18 Banding pattern results for primer combination *Mse*1-CAT and *Eco*R1-AGG.

CAT-AGG	US					PANNAR			SGI		MONSANTO						CYMMIT			
LINE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
V1	50	50	0	0	50	0	50	0	50	0	0	0	0	0	0	0	0	0	0	0
V2	51	51	0	51	51	51	0	51	0	51	51	51	51	51	51	51	51	51	51	51
V4	0	0	0	0	65	0	0	0	0	0	0	0	0	0	65	65	0	65	0	0
V5	69	0	69	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V6	0	70	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	70	70
V7	0	73	0	0	73	0	0	0	0	0	73	73	0	0	73	73	73	73	73	73
V8	0	0	0	0	0	0	76	0	0	0	0	0	0	0	0	0	76	0	0	0
V10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	87	87	0	87	0	0
V11	0	89	0	0	89	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V13	103	0	103	103	0	103	103	103	103	103	103	0	103	103	0	0	103	0	0	0
V15	0	109	109	109	0	109	109	109	109	109	109	109	0	109	0	0	0	109	0	0
V16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	111	0	111	111
V17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	116	0	116	0	116
V18	119	0	119	119	0	119	119	119	119	119	119	119	119	119	119	119	119	119	119	119
V20	0	0	0	0	0	0	0	0	0	0	0	0	125	0	0	0	125	0	0	0
V23	0	0	0	145	0	0	145	0	0	0	0	0	0	0	0	0	0	0	0	0

CAT-AGG	US					PANNAR			SGI		MONSANTO						CYMMIT			
LINE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
V25	196	0	196	196	0	196	196	196	196	196	196	196	196	196	196	196	196	0	0	196
V27	0	0	0	0	0	0	0	0	0	0	0	0	0	201	201	201	0	201	0	0
V28	0	0	0	0	0	204	204	204	0	204	204	204	0	204	204	204	0	204	204	204
V29	0	0	0	0	0	0	0	0	222	222	0	0	0	0	0	0	0	0	0	0
V31	0	0	0	229	0	0	229	0	229	229	0	0	229	229	0	0	229	0	0	0
V32	0	0	0	0	0	0	0	0	233	233	0	0	233	233	0	0	0	0	0	0
V33	0	0	0	0	0	0	0	0	234	234	0	0	234	234	0	0	234	0	0	0
V34	236	0	236	236	0	236	236	236	236	236	236	0	236	236	0	0	236	236	0	0
V35	0	0	0	0	0	0	0	0	0	237	0	0	0	237	0	237	237	0	0	0
V36	0	0	0	0	0	0	262	262	0	0	0	262	262	0	262	0	262	262	0	262
V37	0	0	0	0	0	0	0	0	0	0	279	279	0	0	279	0	279	0	0	279
V38	280	0	280	280	0	280	280	280	280	280	280	280	280	280	280	0	280	0	0	280

Table 8-19 Linear discriminant function for each breeding programme obtained using SAS software for primer combination *Mse*1-CAT and *Eco*R1-AGG.

VARIABLE	US	PANNAR	SGI	MONSANTO	CYMMIT
Constant	-401631583	-448561046	-927342730	-483025649	-657736902
V1	743671	1725127	2211636	1583783	1164661
V2	1665199	701698	174272	338778	1299392
V4	3095630	2800502	3192189	2423321	2857584
V5	2879958	-1296578	-646887	-308963	142898
V6	1130104	856996	1257466	872129	874871
V7	1605217	1082135	2692317	2965669	2929215
V8	-869084	983956	-633827	-306781	847960
V10	-406804	-116066	1160117	-406624	-816417
V11	4075929	3032446	1952672	2057270	1333571
V13	229868	281838	191142	214907	-429034
V15	1120263	1119660	1096052	885017	1142221
V16	782104	958924	650341	731200	2826476
V17	-470863	-419193	-508977	-246430	705045
V18	2615554	3395974	42203536	4125307	4666561
V20	245214	980966	257686	1061644	1010028
V23	556307	-713225	-708883	-458907	-756296
V25	614642	845385	632418	569265	-257464
V27	-217489	-248219	-940064	-127725	48693
V28	-178981	582905	527281	570174	515056
V29	5.6398E-10	5.0864E-10	3036075	7.4073E-12	3.7221E-10
V31	144661	-95340	215131	122033	232549
V32	89941	121806	728726	617383	479734
V33	6306	98931	466295	281398	501080
V34	-100324	-123005	-83422	-93794	187248
V35	-47382	-288788	-24774	-229553	-477247
V36	-0.01117	0.06375	-0.01919	-0.00744	0.08332
V37	-417800	-600834	-226496	-211211	-124975
V38	181131	180584	-6144	-85936	12663

Table 8-20 Number of observations and percent classified into each line for primer combination *Mse*1-CAT and *Eco*R1-AGG.

	Into Line					
From Line	**US**	**PANNAR**	**SGI**	**MONSANTO**	**CYMMIT**	**Total**
US	5 100.00	0 0.00	0 0.00	0 0.00	0 0.00	5 100.00
PANNAR	0 0.00	3 100.00	0 0.00	0 0.00	0 0.00	3 100.00
SGI	0 0.00	0 0.00	2 100.00	0 0.00	0 0.00	2 100.00
MONSANTO	0 0.00	0 0.00	0 0.00	6 100.00	0 0.00	6 100.00
CYMMIT	0 0.00	0 0.00	0 0.00	0 0.00	4 100.00	4 100.00
Total	5 25.00	3 15.00	2 10.00	6 30.00	4 20.00	20 100.00

Table 8-21 Banding pattern results for primer combination *Mse*1-CTC and *Eco*R1-ACA.

CTC-ACA	US					PANNAR			SGI		MONSANTO						CYMMIT			
LINE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
V2	0	0	0	51	51	51	0	0	51	51	51	51	51	51	51	51	51	51	51	51
V3	0	54	0	0	54	54	54	54	54	54	54	54	54	54	54	54	54	54	54	54
V4	0	61	0	0	0	0	61	0	0	0	0	0	61	0	0	61	0	0	0	0
V6	0	0	0	69	0	0	0	0	0	0	69	0	0	0	69	0	0	0	69	0
V7	0	0	0	0	0	0	74	74	74	74	0	0	0	74	0	0	74	74	0	0
V8	0	78	0	0	0	78	0	0	0	0	0	78	0	0	0	78	0	78	0	78
V9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	94	0	0	0	0
V10	0	95	0	0	95	95	95	95	95	95	95	95	95	95	95	95	95	95	95	95
V11	0	96	0	0	0	0	0	0	0	0	0	0	0	0	96	0	0	0	96	0
V12	0	101	0	0	0	101	0	0	0	0	0	0	0	0	0	101	0	0	0	0
V13	104	0	104	104	104	0	104	104	104	0	104	0	104	0	104	0	0	0	104	0
V14	0	106	0	0	106	106	0	0	0	0	106	106	106	106	106	106	106	106	0	106
V19	0	0	0	140	0	140	0	0	140	140	140	0	140	140	0	140	0	0	0	0
V20	0	0	0	0	0	142	142	142	142	142	142	0	142	142	0	0	142	142	0	0
V22	0	0	0	0	0	0	0	0	0	0	0	155	0	0	0	155	0	0	0	0
V24	0	166	0	0	0	0	166	166	0	0	0	166	0	0	0	0	166	166	0	166
V26	0	192	0	0	192	192	192	192	192	192	192	192	192	192	192	0	192	192	192	0
V27	0	0	0	0	0	0	0	0	198	198	0	0	0	0	0	0	0	0	0	0
V28	203	0	203	203	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V29	0	0	0	223	223	0	0	0	0	0	0	0	223	0	0	0	223	0	223	0
V30	0	0	0	240	240	0	240	240	240	240	0	0	0	0	0	0	0	240	0	0
V32	331	0	331	331	0	0	331	331	331	331	331	0	331	331	331	0	331	331	331	0

Table 8-22 Linear discriminant function for each breeding programme obtained using SAS software for primer combination *Mse*1-CTC and *Eco*R1-ACA.

VARIABLE	US	PANNAR	SGI	MONSANTO	CYMMIT
Constant	-222482219	-270614599	-786632039	-503118717	-458483849
V2	553974	2402929	3764060	2742905	3544751
V3	4030141	4682155	5152642	5507576	5609201
V4	-0.04982	-0.06694	0.07515	0.30677	-0.05870
V6	-300339	-474417	-1098529	-840579	-250546
V7	-280045	-442361	-1024303	-783782	-233617
V8	45147	266956	21996	810545	1662250
V9	-36429	-96829	315038	908943	1121564
V10	2290817	2661436	2928870	3130622	3188388
V11	-334906	-205077	316569	586531	-78260
V12	-69970	-4299	-610557	-2012115	-1438806
V13	0.22618	0.28904	0.29873	0.28507	0.19526
V14	174132	-368669	-293115	499758	-367851
V19	267421	-25161	265315	695244	5258
V20	-490070	-113837	-47560	-288924	-58092
V22	201385	272715	695814	1134084	368904
V24	114343	214195	410929	267228	239328
V26	41493	-253203	-328764	-337545	-410858
V27	5.9973E-10	4.9718E-11	3978119	9.2787E-10	1.18622E-9
V28	1666862	1493420	1368266	1273850	1246816
V29	-92930	-146793	-339904	-260089	-77523
V30	-101020	-223156	-322976	-505094	-612263
V32	322028	456411	674287	861203	759410

Table 8-23 Number of observations and percent classified into each line for primer combination *Mse*1-CTC and *Eco*R1-ACA.

	Into Line					
From Line	**US**	**PANNAR**	**SGI**	**MONSANTO**	**CYMMIT**	**Total**
US	5 100.00	0 0.00	0 0.00	0 0.00	0 0.00	5 100.00
PANNAR	0 0.00	3 100.00	0 0.00	0 0.00	0 0.00	3 100.00
SGI	0 0.00	0 0.00	2 100.00	0 0.00	0 0.00	2 100.00
MONSANTO	0 0.00	0 0.00	0 0.00	6 100.00	0 0.00	6 100.00
CYMMIT	0 0.00	0 0.00	0 0.00	0 0.00	4 100.00	4 100.00
Total	5 25.00	3 15.00	2 10.00	6 30.00	4 20.00	20 100.00

Table 8-24 Banding pattern results for primer combination *Mse*1-CTC and *Eco*R1-AAC.

CTC-AAC	US					PANNAR			SGI		MONSANTO						CYMMIT			
LINE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
V1	0	50	0	50	0	50	50	50	50	50	50	50	50	50	50	50	50	0	50	50
V2	0	51	0	0	51	51	51	51	51	51	0	51	51	51	51	51	51	51	51	51
V3	0	52	0	52	52	52	52	52	52	52	52	52	52	52	52	52	52	52	52	52
V4	0	53	0	0	0	53	0	53	53	53	53	53	53	0	0	53	0	0	0	0
V5	0	62	0	0	0	62	0	0	0	0	62	62	0	0	62	62	0	0	0	62
V6	0	0	0	0	0	63	0	0	0	0	0	63	0	0	63	63	0	0	0	0
V7	0	0	0	0	0	0	74	74	74	0	0	0	74	74	0	0	74	74	0	0
V8	99	0	99	99	0	0	0	0	99	0	0	0	0	0	0	0	0	0	0	0
V10	104	104	0	104	104	104	104	104	104	104	104	104	104	104	104	104	104	104	104	0
V12	0	0	0	0	119	0	0	0	0	0	0	0	0	0	0	0	119	119	0	0
V13	0	0	0	0	0	0	0	0	120	120	0	0	0	0	0	0	0	0	0	0
V14	161	161	161	161	161	0	161	161	161	161	161	0	161	161	161	0	161	161	0	0
V15	163	0	163	163	0	0	163	163	163	0	163	0	163	163	163	0	163	0	163	0
V16	164	0	164	164	0	0	164	0	0	0	164	0	164	0	164	0	0	0	164	0
V17	0	0	0	0	0	0	0	0	201	0	201	0	0	201	0	0	0	0	0	0
V18	202	0	202	202	202	0	0	0	202	202	202	0	0	202	202	0	0	0	202	0
V19	0	0	0	0	0	0	221	0	0	221	0	0	0	0	0	221	221	221	0	0
V20	0	0	0	0	0	0	270	270	0	0	0	0	0	0	0	0	0	0	0	0
V22	0	388	0	0	0	388	0	0	0	388	0	388	0	0	388	388	0	0	0	0

Table 8-25 Linear discriminant function for each breeding programme obtained using SAS software for primer combination *Mse*1-CTC and *Eco*R1-AAC.

VARIABLE	US	PANNAR	SGI	MONSANTO	CYMMIT
Constant	-117083410	-24762919	-685154743	-24762936	-33772871
V1	1608830	1485771	3094601	1485771	716019
V2	2394988	583538	2978526	583538	872094
V3	-1546951	-1428626	-2975577	-1428626	-688479
V4	505922	467224	973147	467225	225164
V5	-163258	-95049	-258307	-95049	372080
V6	-2279415	421492	-1857923	421492	-1576036
V7	-1062685	187463	-875222	187463	33237
V8	2094474	291564	2386038	291564	1010997
V10	0.07823	-0.04265	-0.10865	-0.04265	0.04453
V12	941364	200565	1141929	200565	567610
V13	1.26523E-8	8.64616E-9	8161345	1.28724E-8	1.29826E-8
V14	229415	190410	419825	190410	-69163
V15	-493506	-455758	-949264	-455758	-219638
V16	163499	150993	314492	150993	72766
V17	50358	29319	79677	29319	-114771
V18	215375	216003	431378	216003	232358
V19	0.19805	0.14378	0.34183	0.21165	0.20951
V20	99310	91714	191025	91714	44199
V22	150502	-66130	84372	-66130	112573

Table 8-26 Number of observations and percent classified into each line for primer combination *Mse*1-CTC and *Eco*R1-AAC.

	Into Line					
From Line	**US**	**PANNAR**	**SGI**	**MONSANTO**	**CYMMIT**	**Total**
US	5 100.00	0 0.00	0 0.00	0 0.00	0 0.00	5 100.00
PANNAR	0 0.00	3 100.00	0 0.00	0 0.00	0 0.00	3 100.00
SGI	0 0.00	0 0.00	2 100.00	0 0.00	0 0.00	2 100.00
MONSANTO	0 0.00	1 16.67	0 0.00	5 83.33	0 0.00	6 100.00
CYMMIT	0 0.00	0 0.00	0 0.00	0 0.00	4 100.00	4 100.00
Total	5 25.00	4 20.00	2 10.00	5 25.00	4 20.00	20 100.00

Table 8-27 Banding pattern results for primer combination *Mse*1-CTC and *Eco*R1-AGG.

CTC-AGG	**US**					**PANNAR**			**SGI**			**MONSANTO**					**CYMMIT**			
LINE	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
V1	0	51	0	0	0	51	51	0	0	0	0	0	0	0	0	0	0	0	0	0
V2	0	54	0	0	0	0	54	0	0	0	0	54	0	0	0	54	0	0	0	54
V3	0	55	0	0	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55
V4	0	59	0	0	59	0	0	0	0	0	0	0	0	0	0	0	0	59	0	0
V5	0	62	0	0	62	62	62	0	0	0	0	62	0	0	0	62	0	0	0	62
V6	0	0	0	0	0	0	66	0	0	0	0	66	0	0	0	66	0	0	0	66
V9	0	69	0	0	69	69	69	0	0	0	0	69	0	0	0	0	69	0	69	69
V11	0	72	0	0	0	0	72	0	0	0	0	0	0	0	0	0	0	0	0	72
V12	0	73	0	0	73	0	73	73	0	0	0	0	0	0	0	0	0	73	0	73
V15	88	88	0	0	88	0	88	0	0	0	0	0	0	88	0	88	88	88	0	0
V17	0	0	0	0	0	91	0	0	0	91	0	91	0	91	0	0	0	0	0	91
V18	0	92	0	0	92	0	92	0	0	0	0	0	0	0	0	92	92	92	0	92
V19	0	94	0	0	0	94	94	0	0	0	0	94	0	0	0	94	0	0	0	94
V21	0	0	0	0	0	0	98	98	0	0	0	0	0	0	0	0	0	0	98	98
V22	0	100	0	0	100	100	100	100	100	100	100	100	100	100	0	100	100	100	100	100
V25	0	112	0	0	112	0	112	0	0	0	0	0	0	0	0	0	0	0	0	0
V26	113	113	113	0	113	0	113	0	0	113	0	0	0	0	0	0	0	0	113	0
V29	0	124	0	0	124	0	124	0	0	0	0	0	0	124	0	0	0	0	0	0
V33	0	134	0	0	134	134	134	0	0	134	0	134	0	0	0	134	0	0	0	0
V34	0	0	0	0	0	136	0	0	0	0	0	0	0	0	0	136	0	0	0	136
V35	0	0	0	137	0	137	0	0	137	137	0	0	137	0	0	137	0	0	0	0
V36	139	139	139	0	0	0	139	0	0	0	0	0	0	0	0	0	0	0	0	0
V39	0	0	0	0	0	148	0	0	0	0	0	148	0	0	0	148	0	0	0	148
V42	0	165	0	165	165	0	165	165	165	165	0	0	0	0	0	165	165	165	165	165
V45	0	0	0	174	0	174	174	174	0	0	174	0	174	174	174	174	174	174	174	0
V47	0	0	0	0	0	187	0	0	0	0	0	187	0	0	0	0	0	0	0	187

CTC-AGG	US					PANNAR			SGI		MONSANTO						CYMMIT			
LINE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
V48	0	193	193	0	193	193	193	0	193	193	0	193	193	193	0	193	193	193	193	193
V49	0	0	0	0	0	195	195	195	0	0	0	195	195	0	0	195	195	195	0	195
V50	0	203	0	0	203	0	203	203	203	203	0	0	203	203	0	0	203	203	0	0
V51	0	209	0	0	209	209	0	0	209	209	0	0	209	0	0	209	209	209	0	0
V52	0	210	0	0	210	0	210	0	0	0	210	0	210	0	210	210	210	210	0	0
V53	220	0	220	220	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V55	233	233	233	233	233	0	233	233	233	233	0	0	0	0	0	0	233	233	233	0
V56	0	0	0	0	234	0	0	0	234	234	0	0	0	0	0	0	0	234	0	0
V58	0	0	265	265	0	265	265	265	265	265	265	0	265	265	265	0	0	265	265	0
V59	0	0	291	291	0	0	291	0	291	291	0	0	0	291	0	0	0	291	0	0
V60	326	0	326	326	0	0	0	0	326	0	326	0	0	0	326	0	0	0	0	0

Table 8-28 Linear discriminant function for each breeding programme obtained using SAS software for primer combination *Mse*1-CTC and *Eco*R1-AGG.

VARIABLE	US	PANNAR	SGI	MONSANTO	CYMMIT
Constant	-299445351	-516824153	-594358361	-320748530	-451716719
V1	-1018697	946885	-393481	-1394232	-169912
V2	-453296	-1280099	227138	706286	-624588
V3	2122027	4772319	5379519	5361491	4379010
V4	436248	-1637856	-2319704	-570784	-15013
V5	304953	247455	-543754	-67596	-622927
V6	-1139051	-222285	-130008	754565	-241379
V9	684128	658889	55390	-190812	1364639
V11	179572	-1041361	953070	-523549	403468
V12	753700	937436	-400664	-2212	306311
V15	89631	779419	325682	740469	1030278
V17	912279	833261	1491399	852790	1156431
V18	73494	-99868	-305322	-495130	283973
V19	-159098	92162	34834	154028	-43602
V21	60769	1194114	183500	460041	1016827
V22	0.08628	0.18306	0.19799	0.20676	0.17419
V25	-44197	-486840	-651521	-626704	-1014970
V26	503589	212282	588014	203265	518572
V29	21381	-292364	-598851	-81032	-754762
V33	-32560	129384	115071	-11466	-504819
V34	347190	619239	147999	-139650	282536
V35	249308	-457773	609693	-64578	-290310
V36	377524	550213	495995	247872	290145
V39	208298	440981	69007	337664	241021
V42	488381	92480	669850	-121654	330513
V45	58063	759648	-384487	627032	718585
V47	422893	568252	299938	270777	693862
V48	80102	-81577	-42676	117930	211747
V49	18585	705479	-175383	318540	405829
V50	234493	311620	493555	195236	33449
V51	166119	-24956	365065	-79547	43859

VARIABLE	US	PANNAR	SGI	MONSANTO	CYMMIT
Constant	-299445351	-516824153	-594358361	-320748530	-451716719
V52	-57731	21088	-135818	242371	118905
V53	981382	318809	167009	171516	417137
V55	558617	510815	613823	70841	476990
V56	-111033	-45596	568532	-92859	97811
V58	96639	446905	112871	330480	296626
V59	-114026	-117170	24005	-155245	-98998
V60	415828	359362	667430	303792	295055

Table 8-29 Number of observations and percent classified into each line for primer combination *Mse*1-CTC and *Eco*R1-AGG.

	Into Line					
From Line	**US**	**PANNAR**	**SGI**	**MONSANTO**	**CYMMIT**	**Total**
US	5 100.00	0 0.00	0 0.00	0 0.00	0 0.00	5 100.00
PANNAR	0 0.00	3 100.00	0 0.00	0 0.00	0 0.00	3 100.00
SGI	0 0.00	0 0.00	2 100.00	0 0.00	0 0.00	2 100.00
MONSANTO	0 0.00	0 0.00	0 0.00	6 100.00	0 0.00	6 100.00
CYMMIT	0 0.00	0 0.00	0 0.00	0 0.00	4 100.00	4 100.00
Total	5 25.00	3 15.00	2 10.00	6 30.00	4 20.00	20 100.00

Table 8-30 Cluster procedure analysis results for the 119 breeding lines. On the left are the microsatellite results and on the right, the *Xwmc177* microsatellite tested as an AFLP.

MICROSATELLITES							*Xwmc177* as AFLP						
NCL	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**	**NCL**	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**
118	19	24	2	0.0000	1.000	T	**118**	1	2	2	0.0000	1.000	T
117	21	25	2	0.0000	1.000	T	**117**	CL118	3	3	0.0000	1.000	T
116	29	41	2	0.0000	1.000	T	**116**	CL117	4	4	0.0000	1.000	T
115	49	57	2	0.0000	1.000	T	**115**	CL116	5	5	0.0000	1.000	T
114	55	62	2	0.0000	1.000	T	**114**	CL115	6	6	0.0000	1.000	T
113	CL117	65	3	0.0000	1.000	T	**113**	CL114	7	7	0.0000	1.000	T
112	32	67	2	0.0000	1.000	T	**112**	CL113	8	8	0.0000	1.000	T
111	64	71	2	0.0000	1.000	T	**111**	CL112	11	9	0.0000	1.000	T
110	70	76	2	0.0000	1.000	T	**110**	CL111	12	10	0.0000	1.000	T
109	63	81	2	0.0000	1.000	T	**109**	CL110	13	11	0.0000	1.000	T
108	27	89	2	0.0000	1.000	T	**108**	CL109	14	12	0.0000	1.000	T
107	83	103	2	0.0000	1.000	T	**107**	CL108	15	13	0.0000	1.000	T
106	99	106	2	0.0000	1.000		**106**	CL107	16	14	0.0000	1.000	T
105	2	5	2	0.0000	1.000	T	**105**	CL106	17	15	0.0000	1.000	T

MICROSATELLITES							*Xwmc177* as AFLP						
NCL	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**	**NCL**	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**
104	8	11	2	0.0000	1.000	T	**104**	CL105	18	16	0.0000	1.000	T
103	18	20	2	0.0000	1.000	T	**103**	CL104	19	17	0.0000	1.000	T
102	13	82	2	0.0000	1.000	T	**102**	CL103	20	18	0.0000	1.000	T
101	59	92	2	0.0000	1.000	T	**101**	CL102	21	19	0.0000	1.000	T
100	26	94	2	0.0000	1.000		**100**	CL101	22	20	0.0000	1.000	T
99	CL110	80	3	0.0000	1.000		**99**	CL100	23	21	0.0000	1.000	T
98	CL105	91	3	0.0000	1.000		**98**	CL99	24	22	0.0000	1.000	T
97	108	114	2	0.0000	1.000		**97**	9	25	2	0.0000	1.000	T
96	38	42	2	0.0000	1.000		**96**	CL98	26	23	0.0000	1.000	T
95	50	60	2	0.0000	1.000	T	**95**	CL96	27	24	0.0000	1.000	T
94	33	86	2	0.0000	1.000		**94**	CL95	28	25	0.0000	1.000	T
93	45	79	2	0.0000	1.000	T	**93**	CL94	29	26	0.0000	1.000	T
92	12	110	2	0.0000	1.000		**92**	CL93	30	27	0.0000	1.000	T
91	52	CL114	3	0.0000	1.000	T	**91**	CL92	31	28	0.0000	1.000	T
90	101	115	2	0.0000	1.000		**90**	CL91	32	29	0.0000	1.000	T
89	CL91	78	4	0.0000	1.000	T	**89**	CL90	33	30	0.0000	1.000	T
88	CL107	85	3	0.0000	1.000		**88**	CL89	34	31	0.0000	1.000	T
87	48	58	2	0.0000	1.000		**87**	CL88	35	32	0.0000	1.000	T
86	CL100	36	3	0.0000	1.000		**86**	CL87	36	33	0.0000	1.000	T
85	CL98	96	4	0.0000	1.000		**85**	CL97	38	3	0.0000	1.000	T
84	4	7	2	0.0000	1.000	T	**84**	CL86	39	34	0.0000	1.000	T
83	56	68	2	0.0000	1.000		**83**	CL84	41	35	0.0000	1.000	T
82	72	90	2	0.0000	1.000	T	**82**	CL83	42	36	0.0000	1.000	T
81	15	100	2	0.0000	1.000		**81**	CL82	43	37	0.0000	1.000	T
80	CL81	37	3	0.0000	1.000		**80**	CL81	44	38	0.0000	1.000	T
79	77	CL90	3	0.0000	1.000		**79**	CL80	45	39	0.0000	1.000	T
78	CL101	CL111	4	0.0000	1.000		**78**	CL79	47	40	0.0000	1.000	T
77	35	43	2	0.0000	1.000		**77**	CL78	48	41	0.0000	1.000	T
76	3	CL113	4	0.0000	1.000		**76**	CL77	49	42	0.0000	1.000	T
75	CL93	CL115	4	0.0000	1.000		**75**	CL76	50	43	0.0000	1.000	T
74	17	47	2	0.0000	1.000		**74**	CL75	51	44	0.0000	1.000	T
73	46	119	2	0.0000	1.000		**73**	CL74	52	45	0.0000	1.000	T
72	CL116	CL95	4	0.0000	1.000		**72**	CL73	55	46	0.0000	1.000	T
71	CL112	CL89	6	0.0000	1.000		**71**	CL72	56	47	0.0000	1.000	T
70	CL103	CL118	4	0.0000	1.000		**70**	CL71	57	48	0.0000	1.000	T
69	69	118	2	0.0000	1.000		**69**	CL70	58	49	0.0000	1.000	T
68	CL104	CL80	5	0.0000	1.000		**68**	CL69	59	50	0.0000	1.000	T
67	CL92	116	3	0.0000	1.000		**67**	CL68	60	51	0.0000	1.000	T
66	CL97	113	3	0.0000	1.000		**66**	CL67	62	52	0.0000	1.000	T
65	107	112	2	0.0000	1.000		**65**	CL66	63	53	0.0000	1.000	T
64	CL85	CL88	7	0.0000	1.000		**64**	CL65	64	54	0.0000	1.000	T
63	CL96	117	3	0.0000	1.000		**63**	CL85	65	4	0.0000	1.000	T
62	9	97	2	0.0000	1.000		**62**	CL63	66	5	0.0000	1.000	T
61	CL77	74	3	0.0000	1.000		**61**	CL64	67	55	0.0000	1.000	T
60	CL82	111	3	0.0000	1.000		**60**	CL61	70	56	0.0000	1.000	T
59	CL79	104	4	0.0000	1.000		**59**	CL60	71	57	0.0000	1.000	T
58	CL70	87	5	0.0000	1.000		**58**	CL59	72	58	0.0000	1.000	T
57	16	CL66	4	0.0000	1.000		**57**	CL58	73	59	0.0000	1.000	T
56	6	10	2	0.0000	1.000		**56**	CL57	74	60	0.0000	1.000	T
55	CL102	66	3	0.0000	1.000		**55**	CL56	75	61	0.0000	1.000	T
54	CL64	105	8	0.0000	1.000		**54**	CL55	76	62	0.0000	1.000	T
53	CL76	30	5	0.0000	1.000		**53**	CL54	78	63	0.0000	1.000	T

MICROSATELLITES							*Xwmc177* as AFLP						
NCL	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**	**NCL**	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**
52	CL69	CL59	6	0.0000	0.999		**52**	CL53	79	64	0.0000	1.000	T
51	31	CL78	5	0.0000	0.999		**51**	CL52	80	65	0.0000	1.000	T
50	CL61	88	4	0.0000	0.999		**50**	CL51	81	66	0.0000	1.000	T
49	CL54	98	9	0.0000	0.999		**49**	CL50	82	67	0.0000	1.000	T
48	CL73	109	3	0.0000	0.999		**48**	CL49	83	68	0.0000	1.000	T
47	CL58	CL83	7	0.0000	0.999		**47**	CL48	86	69	0.0000	1.000	T
46	CL74	22	3	0.0000	0.999		**46**	CL47	87	70	0.0000	1.000	T
45	CL63	102	4	0.0000	0.999		**45**	CL46	88	71	0.0000	1.000	T
44	CL84	CL65	4	0.0000	0.999		**44**	CL45	89	72	0.0000	1.000	T
43	39	75	2	0.0000	0.999		**43**	CL44	90	73	0.0000	1.000	T
42	CL49	CL94	11	0.0000	0.999		**42**	CL43	91	74	0.0000	1.000	T
41	1	CL87	3	0.0000	0.999		**41**	CL42	92	75	0.0000	1.000	T
40	CL67	53	4	0.0000	0.999		**40**	CL41	94	76	0.0000	1.000	T
39	23	CL60	4	0.0001	0.999		**39**	CL40	96	77	0.0000	1.000	T
38	CL57	CL48	7	0.0001	0.999		**38**	10	98	2	0.0000	1.000	T
37	CL55	CL106	5	0.0001	0.999		**37**	CL62	99	6	0.0000	1.000	T
36	CL50	40	5	0.0001	0.999		**36**	CL39	100	78	0.0000	1.000	T
35	CL47	28	8	0.0001	0.999		**35**	37	102	2	0.0000	1.000	T
34	CL44	84	5	0.0001	0.999		**34**	CL36	103	79	0.0000	1.000	T
33	CL68	CL62	7	0.0001	0.999		**33**	CL34	105	80	0.0000	1.000	T
32	CL51	CL109	7	0.0001	0.998		**32**	CL33	106	81	0.0000	1.000	T
31	CL108	51	3	0.0001	0.998		**31**	CL32	107	82	0.0000	1.000	T
30	44	CL75	5	0.0002	0.998		**30**	101	108	2	0.0000	1.000	T
29	34	CL43	3	0.0002	0.998		**29**	CL31	111	83	0.0000	1.000	T
28	CL86	CL52	9	0.0002	0.998		**28**	CL29	112	84	0.0000	1.000	T
27	CL36	54	6	0.0002	0.998		**27**	CL35	117	3	0.0000	1.000	
26	CL33	14	8	0.0002	0.997		**26**	68	84	2	0.0038	0.996	
25	CL34	CL37	10	0.0002	0.997		**25**	CL28	97	85	0.0048	0.991	
24	CL46	93	4	0.0002	0.997		**24**	61	69	2	0.0063	0.985	
23	CL40	CL38	11	0.0003	0.997		**23**	104	119	2	0.0066	0.979	
22	CL23	CL45	15	0.0003	0.996		**22**	116	118	2	0.0069	0.972	
21	CL53	CL99	8	0.0003	0.996		**21**	54	CL26	3	0.0077	0.964	
20	CL41	CL35	11	0.0005	0.996		**20**	114	115	2	0.0086	0.955	
19	CL39	CL72	8	0.0005	0.995		**19**	CL21	93	4	0.0110	0.944	
18	CL32	CL71	13	0.0008	0.994		**18**	95	CL23	3	0.0110	0.933	
17	CL22	CL28	24	0.0011	0.993		**17**	77	110	2	0.0112	0.922	
16	CL42	CL25	21	0.0017	0.992		**16**	46	CL30	3	0.0126	0.909	
15	61	73	2	0.0046	0.987		**15**	CL25	40	86	0.0127	0.897	
14	CL26	95	9	0.0070	0.980		**14**	113	CL20	3	0.0164	0.880	
13	CL56	CL15	4	0.0093	0.971		**13**	CL24	85	3	0.0179	0.863	
12	CL24	CL31	7	0.0119	0.959		**12**	CL27	CL22	5	0.0199	0.843	
11	CL29	CL27	9	0.0132	0.946		**11**	CL15	CL37	92	0.0221	0.821	
10	CL13	CL30	9	0.0152	0.930		**10**	CL38	53	3	0.0226	0.798	
9	CL20	CL19	19	0.0302	0.900		**9**	CL10	CL17	5	0.0329	0.765	
8	CL10	CL12	16	0.0308	0.869		**8**	CL9	CL19	9	0.0404	0.725	
7	CL21	CL18	21	0.0323	0.837		**7**	CL12	CL14	8	0.0418	0.683	
6	CL14	CL17	33	0.0477	0.789		**6**	CL7	CL16	11	0.0557	0.627	
5	CL16	CL11	30	0.0500	0.739		**5**	CL11	CL8	101	0.0593	0.568	
4	CL9	CL7	40	0.0730	0.666		**4**	CL6	CL18	14	0.0686	0.499	
3	CL8	CL6	49	0.1041	0.562		**3**	CL4	109	15	0.0916	0.408	
2	CL4	CL5	70	0.1852	0.377		**2**	CL5	CL13	104	0.1067	0.301	
1	CL2	CL3	119	0.3771	0.000		**1**	CL2	CL3	119	0.3009	0.000	

Table 8-31 Cluster procedure analysis results for the 119 breeding lines for the three AFLP combinations as specified.

***Mse1*-CTC & *Eco*R1-ACA**							***Mse1*-CTC & *Eco*R1-AAC**							***Mse1*-CTC & *Eco*R1-AGG**						
NCL	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**	**NCL**	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**	**NCL**	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**
118	4	12	2	0.0000	1.00	T	**118**	2	4	2	0.0000	1.00	T	**118**	10	16	2	0.0000	1.00	T
117	3	13	2	0.0000	1.00	T	**117**	5	6	2	0.0000	1.00	T	**117**	13	19	2	0.0000	1.00	T
116	CL117	15	3	0.0000	1.00	T	**116**	CL118	7	3	0.0000	1.00	T	**116**	17	21	2	0.0000	1.00	T
115	CL116	19	4	0.0000	1.00	T	**115**	3	8	2	0.0000	1.00	T	**115**	CL118	23	3	0.0000	1.00	T
114	14	22	2	0.0000	1.00	T	**114**	CL116	12	4	0.0000	1.00	T	**114**	11	25	2	0.0000	1.00	T
113	18	26	2	0.0000	1.00	T	**113**	CL114	13	5	0.0000	1.00	T	**113**	14	26	2	0.0000	1.00	T
112	CL115	29	5	0.0000	1.00	T	**112**	CL113	14	6	0.0000	1.00	T	**112**	CL115	27	4	0.0000	1.00	T
111	5	30	2	0.0000	1.00	T	**111**	CL115	15	3	0.0000	1.00	T	**111**	CL114	28	3	0.0000	1.00	T
110	1	31	2	0.0000	1.00	T	**110**	10	16	2	0.0000	1.00	T	**110**	CL113	29	3	0.0000	1.00	T
109	CL110	32	3	0.0000	1.00	T	**109**	9	18	2	0.0000	1.00	T	**109**	CL111	34	4	0.0000	1.00	T
108	CL109	33	4	0.0000	1.00	T	**108**	CL111	19	4	0.0000	1.00	T	**108**	CL109	35	5	0.0000	1.00	T
107	CL112	34	6	0.0000	1.00	T	**107**	CL112	20	7	0.0000	1.00	T	**107**	2	38	2	0.0000	1.00	T
106	CL107	35	7	0.0000	1.00	T	**106**	CL110	21	3	0.0000	1.00	T	**106**	CL110	39	4	0.0000	1.00	T
105	28	36	2	0.0000	1.00	T	**105**	CL108	22	5	0.0000	1.00	T	**105**	CL108	40	6	0.0000	1.00	T
104	CL114	38	3	0.0000	1.00	T	**104**	CL107	23	8	0.0000	1.00	T	**104**	CL106	41	5	0.0000	1.00	T
103	23	41	2	0.0000	1.00	T	**103**	CL109	24	3	0.0000	1.00	T	**103**	CL104	42	6	0.0000	1.00	T
102	25	42	2	0.0000	1.00	T	**102**	CL103	25	4	0.0000	1.00	T	**102**	CL112	43	5	0.0000	1.00	T
101	50	51	2	0.0000	1.00	T	**101**	CL104	26	9	0.0000	1.00	T	**101**	CL105	44	7	0.0000	1.00	T
100	46	53	2	0.0000	1.00	T	**100**	11	27	2	0.0000	1.00	T	**100**	CL103	46	7	0.0000	1.00	T
99	CL100	54	3	0.0000	1.00	T	**99**	CL101	28	10	0.0000	1.00	T	**99**	6	47	2	0.0000	1.00	T
98	CL108	56	5	0.0000	1.00	T	**98**	CL99	29	11	0.0000	1.00	T	**98**	CL100	50	8	0.0000	1.00	T
97	CL106	57	8	0.0000	1.00	T	**97**	CL98	30	12	0.0000	1.00	T	**97**	CL107	53	3	0.0000	1.00	T
96	CL97	58	9	0.0000	1.00	T	**96**	CL105	31	6	0.0000	1.00	T	**96**	CL101	54	8	0.0000	1.00	T
95	CL96	60	10	0.0000	1.00	T	**95**	CL96	33	7	0.0000	1.00	T	**95**	CL98	55	9	0.0000	1.00	T
94	CL105	61	3	0.0000	1.00	T	**94**	CL102	34	5	0.0000	1.00	T	**94**	CL95	58	10	0.0000	1.00	T
93	68	69	2	0.0000	1.00	T	**93**	CL94	35	6	0.0000	1.00	T	**93**	CL94	59	11	0.0000	1.00	T
92	73	74	2	0.0000	1.00	T	**92**	CL97	36	13	0.0000	1.00	T	**92**	CL97	60	4	0.0000	1.00	T
91	63	81	2	0.0000	1.00	T	**91**	CL92	38	14	0.0000	1.00	T	**91**	CL96	61	9	0.0000	1.00	T
90	70	83	2	0.0000	1.00	T	**90**	CL93	39	7	0.0000	1.00	T	**90**	CL93	62	12	0.0000	1.00	T
89	CL91	85	3	0.0000	1.00	T	**89**	CL106	40	4	0.0000	1.00	T	**89**	CL90	107	13	0.0000	1.00	T
88	84	96	2	0.0000	1.00	T	**88**	CL91	41	15	0.0000	1.00	T	**88**	CL92	110	5	0.0000	1.00	
87	99	100	2	0.0000	1.00	T	**87**	CL88	42	16	0.0000	1.00	T	**87**	95	106	2	0.0002	1.00	
86	59	106	2	0.0000	1.00	T	**86**	CL100	43	3	0.0000	1.00	T	**86**	CL116	118	3	0.0003	0.999	
85	2	107	2	0.0000	1.00	T	**85**	CL87	44	17	0.0000	1.00	T	**85**	CL88	18	6	0.0004	0.999	
84	55	108	2	0.0000	1.00	T	**84**	CL85	46	18	0.0000	1.00	T	**84**	CL91	51	10	0.0004	0.999	
83	21	109	2	0.0000	1.00	T	**83**	CL95	47	8	0.0000	1.00	T	**83**	CL86	119	4	0.0004	0.998	
82	111	112	2	0.0000	1.00	T	**82**	CL83	48	9	0.0000	1.00	T	**82**	75	96	2	0.0005	0.998	
81	CL83	118	3	0.0000	1.00		**81**	CL84	50	19	0.0000	1.00	T	**81**	67	82	2	0.0007	0.997	T
80	66	67	2	0.0004	1.00	T	**80**	51	53	2	0.0000	1.00	T	**80**	93	97	2	0.0007	0.996	
79	77	80	2	0.0004	0.999		**79**	CL81	54	20	0.0000	1.00	T	**79**	4	99	2	0.0009	0.995	
78	10	40	2	0.0004	0.999	T	**78**	CL79	55	21	0.0000	1.00	T	**78**	52	105	2	0.0010	0.994	
77	8	47	2	0.0004	0.998	T	**77**	49	56	2	0.0000	1.00	T	**77**	81	85	2	0.0010	0.993	
76	76	78	2	0.0004	0.998	T	**76**	CL78	58	22	0.0000	1.00	T	**76**	CL85	22	7	0.0010	0.992	
75	88	90	2	0.0004	0.998	T	**75**	CL76	60	23	0.0000	1.00	T	**75**	C82	98	3	0.0011	0.991	
74	20	97	2	0.0004	0.997		**74**	CL75	61	24	0.0000	1.00	T	**74**	66	101	2	0.0012	0.990	
73	64	CL90	3	0.0005	0.997	T	**73**	CL80	62	3	0.0000	1.00	T	**73**	CL81	CL80	4	0.0012	0.989	
72	CL88	92	3	0.0005	0.996		**72**	66	67	2	0.0000	1.00	T	**72**	3	31	2	0.0012	0.988	
71	CL113	39	3	0.0005	0.996		**71**	CL72	68	3	0.0000	1.00	T	**71**	74	88	2	0.0013	0.986	

Mse1-CTC & *Eco*R1-ACA							*Mse1*-CTC & *Eco*R1-AAC							*Mse1*-CTC & *Eco*R1-AGG						
NCL	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**	**NCL**	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**	**NCL**	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**
70	CL103	102	3	0.0005	0.995		**70**	32	69	2	0.0000	1.00	T	**70**	12	CL117	3	0.0013	0.985	
69	CL111	37	3	0.0005	0.995	T	**69**	64	70	2	0.0000	1.00	T	**69**	64	65	2	0.0013	0.984	
68	CL85	110	3	0.0005	0.994		**68**	74	75	2	0.0000	1.00	T	**68**	37	108	2	0.0013	0.983	
67	CL89	79	4	0.0005	0.994		**67**	CL71	83	4	0.0000	1.00	T	**67**	CL102	109	6	0.0014	0.981	
66	CL118	7	3	0.0006	0.993		**66**	CL69	90	3	0.0000	1.00	T	**66**	68	103	2	0.0015	0.980	
65	CL98	49	6	0.0006	0.992		**65**	CL68	91	3	0.0000	1.00	T	**65**	CL79	33	3	0.0016	0.978	
64	CL94	52	4	0.0006	0.992	T	**64**	CL67	92	5	0.0000	1.00	T	**64**	1	49	2	0.0016	0.976	
63	CL104	62	4	0.0006	0.991		**63**	CL70	98	3	0.0000	1.00	T	**63**	63	84	2	0.0017	0.975	
62	65	91	2	0.0007	0.991		**62**	CL89	102	5	0.0000	1.00	T	**62**	CL71	83	3	0.0018	0.973	
61	CL69	105	4	0.0007	0.990		**61**	CL63	104	4	0.0000	1.00	T	**61**	76	77	2	0.0018	0.971	
60	CL65	9	7	0.0007	0.989		**60**	CL82	105	10	0.0000	1.00	T	**60**	70	CL77	3	0.0020	0.969	
59	CL92	89	3	0.0007	0.988		**59**	CL60	106	11	0.0000	1.00	T	**59**	CL99	30	3	0.0021	0.967	
58	CL101	CL84	4	0.0008	0.988		**58**	CL59	107	12	0.0000	1.00	T	**58**	7	CL67	7	0.0021	0.965	
57	6	43	2	0.0008	0.987	T	**57**	CL77	108	3	0.0000	1.00	T	**57**	94	115	2	0.0022	0.963	
56	98	104	2	0.0008	0.986		**56**	CL62	109	6	0.0000	1.00	T	**56**	5	24	2	0.0025	0.960	
55	CL78	CL102	4	0.0009	0.985		**55**	CL74	110	25	0.0000	1.00		**55**	CL65	15	4	0.0026	0.958	
54	CL82	117	3	0.0010	0.984		**54**	CL73	57	4	0.0007	0.999		**54**	CL59	102	4	0.0029	0.955	
53	CL95	44	11	0.0010	0.983		**53**	CL117	97	3	0.0008	0.998		**53**	CL66	CL75	5	0.0029	0.952	
52	93	101	2	0.0011	0.982	T	**52**	1	CL57	4	0.0009	0.998		**52**	32	71	2	0.0029	0.949	
51	95	103	2	0.0011	0.981		**51**	CL66	85	4	0.0009	0.997		**51**	CL72	CL70	5	0.0029	0.946	
50	82	CL72	4	0.0013	0.980		**50**	CL61	37	5	0.0010	0.996		**50**	CL74	CL61	4	0.0030	0.943	
49	17	CL81	4	0.0013	0.978		**49**	86	112	2	0.0011	0.994		**49**	CL64	48	3	0.0032	0.940	
48	CL68	CL70	6	0.0014	0.977		**48**	71	116	2	0.0014	0.993		**48**	79	80	2	0.0032	0.936	
47	CL66	45	4	0.0019	0.975		**47**	72	101	2	0.0016	0.991		**47**	CL73	69	5	0.0032	0.933	
46	CL67	CL50	8	0.0019	0.973		**46**	CL86	59	4	0.0017	0.990		**46**	91	113	2	0.0033	0.930	
45	CL71	CL58	7	0.0019	0.971		**45**	76	103	2	0.0018	0.988		**45**	CL53	104	6	0.0035	0.926	
44	CL74	CL93	4	0.0019	0.969		**44**	81	93	2	0.0020	0.986		**44**	CL56	56	3	0.0035	0.923	
43	27	CL75	3	0.0019	0.967		**43**	73	CL65	4	0.0021	0.984		**43**	73	CL62	4	0.0035	0.919	
42	CL77	24	3	0.0020	0.965		**42**	111	117	2	0.0026	0.981		**42**	CL78	90	3	0.0037	0.916	
41	86	CL54	4	0.0020	0.963		**41**	CL55	45	26	0.0026	0.979		**41**	CL55	CL68	6	0.0038	0.912	
40	CL48	16	7	0.0024	0.961		**40**	65	77	2	0.0028	0.976		**40**	36	100	2	0.0039	0.908	
39	CL76	CL52	4	0.0024	0.959		**39**	CL56	119	7	0.0029	0.973		**39**	CL52	CL87	4	0.0039	0.904	
38	CL86	CL87	4	0.0027	0.956		**38**	CL64	CL47	7	0.0030	0.970		**38**	CL63	CL69	4	0.0039	0.900	
37	CL61	11	5	0.0028	0.953		**37**	CL41	CL53	29	0.0030	0.967		**37**	CL76	CL84	17	0.0041	0.896	
36	94	115	2	0.0028	0.950		**36**	CL50	52	6	0.0030	0.964		**36**	111	112	2	0.0043	0.892	
35	CL53	CL45	18	0.0030	0.947		**35**	CL39	118	8	0.0031	0.961		**35**	CL89	45	14	0.0044	0.887	
34	CL51	114	3	0.0033	0.944		**34**	78	80	2	0.0035	0.957		**34**	20	CL42	4	0.0044	0.883	
33	CL57	CL59	5	0.0037	0.940		**33**	CL44	88	3	0.0038	0.954		**33**	CL50	89	5	0.0044	0.879	
32	113	116	2	0.0038	0.936		**32**	CL40	96	3	0.0042	0.949		**32**	78	86	2	0.0046	0.874	
31	CL46	CL73	11	0.0042	0.932		**31**	CL58	CL90	19	0.0042	0.945		**31**	CL46	92	3	0.0048	0.869	
30	CL49	119	5	0.0042	0.928		**30**	99	100	2	0.0043	0.941		**30**	CL38	116	5	0.0049	0.864	
29	CL60	CL42	10	0.0043	0.924		**29**	84	94	2	0.0046	0.936		**29**	CL49	9	4	0.0049	0.859	
28	CL62	CL79	4	0.0045	0.919		**28**	CL48	113	3	0.0048	0.931		**28**	72	CL57	3	0.0049	0.855	
27	75	CL39	5	0.0048	0.914		**27**	63	95	2	0.0052	0.926		**27**	CL58	CL35	21	0.0050	0.850	
26	CL35	48	19	0.0052	0.909		**26**	CL38	89	8	0.0054	0.921		**26**	CL54	8	5	0.0051	0.844	
25	CL44	CL56	6	0.0054	0.904		**25**	CL32	79	4	0.0054	0.915		**25**	CL39	57	5	0.0057	0.839	
24	CL40	CL30	12	0.0055	0.898		**24**	CL35	17	9	0.0066	0.909		**24**	CL30	CL48	7	0.0057	0.833	
23	CL99	72	4	0.0057	0.892		**23**	82	CL29	3	0.0067	0.902		**23**	CL43	CL32	6	0.0058	0.827	
22	CL43	87	4	0.0063	0.886		**22**	CL28	CL42	5	0.0073	0.895		**22**	CL51	CL27	26	0.0060	0.821	
21	CL37	CL63	9	0.0063	0.880		**21**	CL51	CL33	7	0.0078	0.887		**21**	CL28	CL31	6	0.0065	0.815	
20	CL23	71	5	0.0068	0.873		**20**	CL27	CL49	4	0.0080	0.879		**20**	CL33	CL47	10	0.0067	0.808	
19	CL21	CL38	13	0.0074	0.866		**19**	CL37	CL54	33	0.0082	0.871		**19**	CL34	CL40	6	0.0070	0.801	
18	CL27	CL41	9	0.0076	0.858		**18**	CL23	87	4	0.0082	0.863		**18**	CL44	CL26	8	0.0075	0.794	
17	CL25	CL80	8	0.0076	0.850		**17**	CL52	CL31	23	0.0083	0.854		**17**	87	114	2	0.0078	0.786	

Mse1-CTC & *Eco*R1-ACA							*Mse1*-CTC & *Eco*R1-AAC							*Mse1*-CTC & *Eco*R1-AGG						
NCL	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**	**NCL**	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**	**NCL**	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**
16	CL26	CL55	23	0.0083	0.842		**16**	CL36	CL45	8	0.0086	0.846		**16**	CL21	117	7	0.0079	0.778	
15	CL17	CL64	12	0.0091	0.833		**15**	CL34	CL18	6	0.0103	0.836		**15**	CL16	CL36	9	0.0085	0.769	
14	CL24	CL47	16	0.0102	0.823		**14**	CL46	CL30	6	0.0106	0.825		**14**	CL29	CL41	10	0.0088	0.761	
13	CL33	CL20	10	0.0102	0.813		**13**	CL26	CL43	12	0.0138	0.811		**13**	CL45	CL60	9	0.0091	0.752	
12	CL31	CL28	15	0.0113	0.801		**12**	CL21	CL22	12	0.0143	0.797		**12**	CL25	CL17	7	0.0105	0.741	
11	CL13	CL22	14	0.0142	0.787		**11**	CL25	114	5	0.0150	0.782		**11**	CL20	CL13	19	0.0107	0.730	
10	CL12	CL36	17	0.0175	0.770		**10**	CL16	CL20	12	0.0190	0.763		**10**	CL14	CL18	18	0.0126	0.718	
9	CL15	CL34	15	0.0176	0.752		**9**	CL12	115	13	0.0242	0.739		**9**	CL19	CL12	13	0.0158	0.702	
8	CL10	CL18	26	0.0223	0.730		**8**	CL11	CL13	17	0.0243	0.714		**8**	CL22	CL83	30	0.0187	0.683	
7	CL11	CL32	16	0.0232	0.707		**7**	CL9	CL15	19	0.0263	0.688		**7**	CL24	CL15	16	0.0203	0.663	
6	CL29	CL16	33	0.0237	0.683		**6**	CL19	CL24	42	0.0270	0.661		**6**	CL11	CL23	25	0.0243	0.639	
5	CL6	CL14	49	0.0512	0.632		**5**	CL6	CL14	48	0.0400	0.621		**5**	CL37	CL8	47	0.0335	0.605	
4	CL7	CL9	31	0.0583	0.573		**4**	CL10	CL8	29	0.0514	0.570		**4**	CL10	CL9	31	0.0396	0.566	
3	CL5	CL19	62	0.0678	0.506		**3**	CL17	CL5	71	0.0596	0.510		**3**	CL7	CL6	41	0.0760	0.490	
2	CL4	CL8	57	0.1496	0.356		**2**	CL4	CL7	48	0.0606	0.449		**2**	CL4	CL5	78	0.1109	0.379	
1	CL3	CL2	119	0.3559	0.000		**1**	CL3	CL2	119	0.4494	0.000		**1**	CL2	CL3	119	0.3787	0.000	

Table 8-32 Variables of primer combination *Mse*1-CTC and *Eco*R1-ACA used for cluster analysis. Eight seeds were pooled.

Line	03H 178-3	03H 242-2	03H 244-2	03H 257-1	03H 277-2	03H 279-2	03H 293	03H 297	03H 310-1	03H 327-3	03H 344-5	03H 352	03H 381-4	03H 402-2	03H 409-1	03H 412-3
V1	102	0	102	0	102	102	102	102	0	102	0	102	102	102	102	0
V2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	106	0
V3	108	108	0	108	0	108	108	108	108	108	0	108	0	108	108	108
V4	0	0	0	127	0	0	0	127	0	0	0	0	0	0	0	0
V5	0	0	0	140	0	0	0	0	0	0	0	0	0	0	0	0
V6	142	142	0	142	142	0	0	142	142	0	0	0	0	0	142	142
V7	0	0	0	147	0	0	0	0	0	0	0	0	0	0	0	147
V8	156	0	0	0	0	0	0	156	0	0	0	0	0	0	0	0
V9	0	169	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V10	0	0	0	0	0	0	0	183	0	0	0	0	0	0	0	0
V11	192	192	192	192	192	0	192	192	0	0	0	192	0	0	0	0
V12	0	0	0	0	196	0	0	0	0	0	0	0	0	0	0	0
V13	0	0	0	0	0	219	0	0	0	0	0	0	0	0	0	0
V14	0	0	224	224	0	0	0	224	0	0	224	224	0	224	224	224
V15	240	240	0	240	240	240	0	240	240	0	240	240	0	240	240	240
V16	270	0	270	270	270	270	0	0	270	0	270	270	0	270	0	0
V17	0	0	0	0	0	0	0	312	0	0	0	0	0	0	0	0
V18	318	318	318	318	318	318	0	318	0	0	318	0	0	318	0	318
V19	0	0	0	0	321	0	0	321	0	0	321	0	0	0	0	321
V20	0	0	0	0	0	0	0	0	0	0	326	0	0	0	0	0
V21	0	328	0	0	0	0	0	0	0	0	0	0	0	0	0	328
V22	0	0	0	329	0	0	0	329	0	0	0	0	0	0	0	0

Table 8-33 Variables of primer combination *Mse*1-CTC and *Eco*R1-ACA used for cluster analysis. Twelve seeds were pooled.

Line	03H 178-3	03H 242-2	03H 244-2	03H 257-1	03H 277-2	03H 279-2	03H 293	03H 297	03H 310-1	03H 327-3	03H 344-5	03H 352	03H 381-4	03H 402-2	03H 409-1	03H 412-3
V1	0	0	102	0	0	0	0	0	102	0	0	0	102	102	0	0
V2	103	103	0	103	103	103	103	103	0	103	103	103	103	103	103	103
V3	0	0	0	108	0	0	0	0	108	0	0	0	0	108	0	0
V4	0	0	0	0	0	0	0	113	0	0	113	0	0	0	0	0
V5	0	0	0	0	0	0	0	121	0	0	0	0	0	0	121	0
V6	127	0	0	0	0	0	0	127	0	0	0	0	0	0	127	127
V7	0	0	0	140	0	0	0	0	0	0	0	0	0	0	0	0
V8	142	142	142	142	142	142	0	142	142	0	142	0	0	142	142	142
V9	145	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V10	147	0	0	0	0	0	0	147	0	0	0	0	0	0	0	147
V11	152	152	0	0	0	0	0	152	0	0	0	0	0	0	0	0
V12	0	156	0	0	0	0	0	156	0	0	0	0	0	0	156	156
V13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	163
V14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	166	0
V15	0	169	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V16	182	182	0	0	0	0	0	182	0	0	0	0	0	0	182	0
V17	192	192	192	192	192	192	192	192	192	0	192	0	0	192	192	192
V18	0	0	0	0	196	0	0	0	0	0	196	0	0	196	0	0
V19	0	0	0	0	0	0	0	219	0	0	0	0	0	0	0	0
V20	220	220	220	220	220	220	220	220	220	220	220	220	0	220	220	220
V21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	222	0
V22	0	223	223	223	0	223	0	223	0	0	223	0	0	223	223	223
V23	240	240	0	240	240	0	0	240	240	240	240	0	0	240	240	240
V24	0	0	0	0	259	259	0	0	0	0	259	0	0	259	0	0
V25	270	270	270	270	270	270	0	0	270	0	270	0	270	270	270	0
V26	0	0	0	0	0	0	0	311	0	0	0	0	0	0	0	0
V27	317	317	0	317	317	0	0	317	317	0	317	0	0	317	317	317
V28	321	0	0	0	321	0	0	321	0	0	321	0	0	321	321	321
V29	0	0	0	0	326	326	0	0	0	0	326	0	0	326	0	0
V30	0	328	0	0	328	0	0	328	0	0	328	0	0	0	0	328
V31	0	329	0	0	0	0	0	329	0	0	0	0	0	0	329	0

Table 8-34 Cluster history for the *Mse1*-CTC & *Eco*R1-ACA primer combination.

***Mse1*-CTC & *Eco*R1-ACA**													
8 seeds							**12 seeds**						
NCL	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**	**NCL**	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**
15	10	13	2	0.0029	0.997		**15**	7	12	2	0.0058	0.994	
14	7	CL15	3	0.0131	0.984		**14**	5	11	2	0.0099	0.984	
13	6	14	2	0.0242	0.960		**13**	3	6	2	0.0139	0.970	
12	9	12	2	0.0291	0.931		**12**	CL15	10	3	0.0141	0.956	
11	1	2	2	0.0336	0.897		**11**	4	9	2	0.0143	0.942	
10	8	16	2	0.0353	0.862		**10**	CL14	14	3	0.0306	0.912	
9	CL12	15	3	0.0374	0.824		**9**	CL12	13	4	0.0337	0.878	
8	3	4	2	0.0388	0.786		**8**	8	16	2	0.0421	0.836	
7	CL8	CL13	4	0.0497	0.736		**7**	1	15	2	0.0458	0.790	
6	CL11	CL7	6	0.0605	0.675		**6**	CL13	CL11	4	0.0644	0.726	
5	5	11	2	0.0648	0.611		**5**	2	CL8	3	0.0650	0.661	

Mse1-CTC & *Eco*R1-ACA													
8 seeds							12 seeds						
NCL	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**	**NCL**	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**
4	CL5	CL10	4	0.0801	0.530		**4**	CL7	CL5	5	0.0777	0.583	
3	CL14	CL9	6	0.0926	0.438		**3**	CL6	CL9	8	0.0968	0.486	
2	CL6	CL4	10	0.1589	0.279		**2**	CL4	CL10	8	0.1396	0.347	
1	CL2	CL3	16	0.2789	0.000		**1**	CL2	CL3	16	0.3465	0.000	

Table 8-35 Variables of primer combination *Mse*1-CTC and *Eco*R1-AAC used for cluster analysis. Eight seeds were pooled.

Line	03H 178-3	03H 242-2	03H 244-2	03H 257-1	03H 277-2	03H 279-2	03H 293	03H 297	03H 310-1	03H 327-3	03H 344-5	03H 352	03H 381-4	03H 402-2	03H 409-1	03H 412-3
V1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	103	0
V2	0	0	0	0	0	0	0	0	0	0	107	0	0	0	0	0
V3	0	0	0	0	0	119	0	0	0	0	119	0	0	0	0	119
V4	0	154	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V5	162	162	0	162	162	0	0	162	0	0	162	0	0	0	0	162
V6	0	0	0	0	0	0	0	163	0	0	163	0	0	0	0	0
V7	201	201	201	201	201	201	201	201	201	201	201	201	0	201	201	201
V8	230	230	0	0	230	230	0	230	230	0	230	0	0	230	230	230
V9	245	245	245	245	245	245	245	245	245	0	245	245	0	245	245	245
V10	0	0	0	0	0	0	0	0	0	0	270	0	0	0	0	0
V11	286	0	0	0	0	286	0	286	0	0	0	0	0	0	0	0
V12	287	287	287	287	287	287	0	287	287	0	287	287	0	287	287	287

Table 8-36 Variables of primer combination *Mse*1-CTC and *Eco*R1-AAC used for cluster analysis. Twelve seeds were pooled.

Line	03H 178-3	03H 242-2	03H 244-2	03H 257-1	03H 277-2	03H 279-2	03H 293	03H 297	03H 310-1	03H 327-3	03H 344-5	03H 352	03H 381-4	03H 402-2	03H 409-1	03H 412-3
V1	0	0	0	103	0	0	0	0	0	0	0	0	0	0	103	0
V2	0	0	0	0	0	0	0	0	0	0	107	0	0	0	0	0
V3	119	119	0	0	0	0	0	119	0	0	0	0	0	0	119	0
V4	121	121	0	0	0	0	0	0	0	0	0	0	0	0	121	0
V5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	141
V6	154	154	0	154	0	0	0	0	154	0	0	0	0	0	0	0
V7	161	161	161	161	161	161	0	161	161	161	161	0	0	161	161	161
V8	0	0	0	0	0	0	0	163	0	0	0	0	0	0	163	0
V9	0	0	0	0	0	0	0	177	0	0	0	0	0	0	0	0
V10	197	0	0	0	0	0	0	0	0	0	0	0	0	0	197	0
V11	201	201	201	201	201	201	0	201	201	0	201	201	201	201	201	201
V12	0	0	0	0	0	0	0	224	224	0	0	0	0	0	224	224
V13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	229	0
V14	230	230	0	0	0	230	0	230	230	0	230	0	0	230	230	230
V15	0	234	0	0	0	0	0	234	0	0	0	0	0	0	0	0
V16	0	244	244	0	244	244	0	244	244	244	244	0	0	244	244	244
V17	286	286	0	286	0	0	0	286	0	0	286	0	0	0	286	0
V18	287	287	287	287	287	287	287	287	287	287	287	287	0	287	287	287

Table 8-37 Cluster history for the *Mse1*-CTC & *Eco*R1-AAC primer combination.

Mse1-CTC & *Eco*R1-AAC													
8 seeds							12 seeds						
NCL	Clusters Joined		FREQ	SPRSQ	RSQ	Tie	NCL	Clusters Joined		FREQ	SPRSQ	RSQ	Tie
15	3	12	2	0.0000	1.000	T	**15**	3	5	2	0.0000	1.000	T
14	9	14	2	0.0000	1.000		**14**	6	14	2	0.0000	1.000	
13	CL14	15	3	0.0073	0.993		**13**	7	12	2	0.0129	0.987	
12	5	16	2	0.0073	0.985		**12**	9	16	2	0.0139	0.973	
11	1	8	2	0.0137	0.972		**11**	CL15	10	3	0.0172	0.956	
10	CL15	4	3	0.0181	0.954		**10**	2	11	2	0.0379	0.918	
9	2	CL12	3	0.0188	0.935		**9**	CL14	CL12	4	0.0390	0.879	
8	10	13	2	0.0209	0.914		**8**	CL13	13	3	0.0393	0.840	
7	CL11	6	3	0.0324	0.882		**7**	1	4	2	0.0415	0.798	
6	CL9	11	4	0.0363	0.845		**6**	CL10	8	3	0.0586	0.739	
5	7	CL8	3	0.0483	0.797		**5**	CL11	CL9	7	0.0795	0.660	
4	CL6	CL13	7	0.0616	0.735		**4**	CL6	15	4	0.0841	0.576	
3	CL4	CL10	10	0.1236	0.612		**3**	CL7	CL4	6	0.1097	0.466	
2	CL7	CL3	13	0.2131	0.399		**2**	CL5	CL8	1	0.1585	0.308	
1	CL2	CL5	16	0.3987	0.000		**1**	CL3	CL2	16	0.3076	0.000	

Table 8-38 Variables of primer combination *Mse*1-CTC and *Eco*R1-AGG used for cluster analysis. Eight seeds were pooled.

Line	03H 178-3	03H 242-2	03H 244-2	03H 257-1	03H 277-2	03H 279-2	03H 293	03H 297	03H 310-1	03H 327-3	03H 344-5	03H 352	03H 381-4	03H 402-2	03H 409-1	03H 412-3
V1	0	101	0	0	0	0	0	0	101	0	0	0	0	0	101	0
V2	0	0	0	0	0	0	0	0	0	0	105	0	0	0	0	0
V3	0	129	0	0	0	129	0	129	0	0	0	0	0	0	129	129
V4	0	138	0	0	0	0	0	138	0	0	0	0	0	0	0	0
V5	149	149	0	149	149	149	0	149	0	0	149	149	0	0	149	0
V6	0	0	0	0	0	0	0	0	159	0	0	159	0	0	0	0
V7	165	165	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V8	0	0	0	0	0	0	0	172	0	0	0	0	0	0	172	172
V9	189	189	0	189	0	0	0	189	0	0	0	189	0	0	0	189
V10	209	209	0	209	0	209	0	0	0	0	0	209	0	0	0	0
V11	210	0	0	0	0	0	0	210	0	0	0	210	0	0	0	0
V12	0	0	0	221	0	0	0	221	0	0	221	0	0	0	0	0
V13	233	233	0	233	0	233	0	233	233	0	0	233	233	0	0	0
V14	0	0	0	0	0	0	0	256	0	0	0	0	0	0	0	0
V15	0	0	258	258	258	258	0	258	0	0	258	0	258	0	258	258
V16	265	265	265	265	265	265	265	265	265	0	265	265	265	265	265	265
V17	291	291	0	0	0	291	0	291	0	0	0	0	0	0	0	291
V18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	314	314
V19	318	0	318	318	0	0	0	0	0	0	0	0	0	0	318	0
V20	0	0	0	0	334	0	0	0	0	0	0	0	0	0	0	0
V21	0	0	0	0	0	0	0	0	0	0	396	0	0	0	0	396

Table 8-39 Variables of primer combination *Mse*1-CTC and *Eco*R1-AGG used for cluster analysis. Twelve seeds were pooled.

Line	03H 178-3	03H 242-2	03H 244-2	03H 257-1	03H 277-2	03H 279-2	03H 293	03H 297	03H 310-1	03H 327-3	03H 344-5	03H 352	03H 381-4	03H 402-2	03H 409-1	03H 412-3
V1	0	101	0	0	0	0	0	0	101	0	0	0	0	0	101	0
V2	104	0	0	0	0	0	0	0	0	0	104	0	0	0	0	0
V3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	113	0
V4	0	120	0	0	0	0	0	0	0	0	0	0	0	0	0	120
V5	0	124	0	124	0	0	0	0	0	0	0	0	0	0	0	0
V6	129	129	0	0	0	0	0	0	0	0	0	0	0	0	0	129
V7	0	138	0	0	0	0	0	138	0	0	0	0	0	0	0	138
V8	149	149	149	149	149	149	149	149	149	149	149	0	0	0	149	149
V9	159	0	0	0	0	0	0	159	159	0	0	0	0	0	159	0
V10	165	165	165	165	0	0	0	165	0	0	0	165	0	0	0	165
V11	189	189	0	189	0	0	0	189	189	0	0	189	0	189	189	189
V12	0	194	0	0	0	0	0	194	0	0	0	0	0	0	194	0
V13	0	0	0	0	0	0	0	0	202	0	0	0	0	0	0	0
V14	209	0	0	209	0	209	0	0	0	0	0	0	0	0	0	0
V15	210	0	0	0	0	0	0	210	0	0	0	0	0	0	0	0
V16	0	0	220	220	0	0	0	220	0	0	220	0	0	0	0	0
V17	233	233	233	233	0	0	233	233	233	233	233	0	233	233	233	233
V18	0	0	0	0	0	0	0	234	0	0	0	0	0	0	234	0
V19	256	256	0	0	0	0	0	0	0	0	0	0	0	0	256	256
V20	291	291	0	0	0	0	0	291	291	0	0	0	0	0	291	291
V21	313	0	0	0	0	0	0	0	0	0	0	0	0	0	313	313
V22	318	0	318	318	0	0	0	0	0	0	0	0	0	0	318	0
V23	0	396	0	0	0	0	0	396	0	0	0	0	0	0	396	396

Table 8-40 Cluster history for the *Mse1*-CTC & *Eco*R1-AGG primer combination.

***Mse1*-CTC & *Eco*R1-AGG**													
8 seeds							**12 seeds**						
NCL	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**	**NCL**	**Clusters Joined**		**FREQ**	**SPRSQ**	**RSQ**	**Tie**
15	7	14	2	0.0000	1.000		**15**	7	10	2	0.0000	1.000	
14	CL15	10	3	0.0177	0.982		**14**	CL15	13	3	0.0053	0.995	
13	9	13	2	0.0193	0.963		**13**	5	6	2	0.0078	0.987	
12	2	6	2	0.0301	0.933		**12**	CL14	14	4	0.0123	0.975	
11	3	15	2	0.0336	0.899		**11**	3	4	2	0.0170	0.958	
10	CL14	CL13	5	0.0398	0.860		**10**	CL12	11	5	0.0192	0.938	
9	1	12	2	0.0451	0.814		**9**	CL13	12	3	0.0230	0.915	
8	CL9	4	3	0.0553	0.759		**8**	2	16	2	0.0289	0.886	
7	5	CL10	6	0.0597	0.699		**7**	CL9	CL10	8	0.0439	0.843	
6	11	16	2	0.0658	0.634		**6**	CL7	9	9	0.0618	0.781	
5	CL8	CL12	5	0.0663	0.567		**5**	CL8	8	3	0.0740	0.707	
4	CL5	8	6	0.0836	0.484		**4**	1	15	2	0.0744	0.632	
3	CL11	CL7	8	0.1036	0.380		**3**	CL11	CL6	11	0.1074	0.525	
2	CL3	CL6	10	0.1538	0.226		**2**	CL4	CL5	5	0.1133	0.412	
1	CL4	CL2	16	0.2262	0.000		**1**	CL2	CL3	16	0.4115	0.000	