(2)

FINGERPRINTING OF FULL AND HALF-SIB BLACK WATTLE (*ACACIA MEARNSII*) PROGENIES USING RANDOM AMPLIFIED POLYMORPHIC DNA (RAPD)

RIANN NAGURAN

Submitted in fulfillment of the academic requirements of the degree of

Master of Science

in the School of Molecular and Cellular Biosciences,

University of KwaZulu-Natal

Pietermaritzburg 2005

PREFACE

The experimental work described in this dissertation was conducted at the University of KwaZulu-Natal, Pietermaritzburg, under the supervision of Professor Annabel Fossey.

The results have not been submitted in any other form to another university and except where the work of others is acknowledged in the text, the results are of my own investigation.

.....

Riann Naguran
November 2005

I certify that the above statement is correct.

.....

Professor Annabel Fossey
Supervisor

ACKNOWLEDGEMENTS

I wish to express my sincere thanks and appreciation to the following people, who have assisted me in compiling this thesis:

- Professor Annabel Fossey, my supervisor for her expert advice, support and encouragement.
- Mr Rob Dunlop, for supplying me with plant material, his co-supervision and enthusiastic involvement in the research.
- ICFR, for making this project possible.
- My fiancé Rolan, for your understanding, enthusiasm and patience in all things, friendship and growing love.
- My friends Kesh and Esh for their support.
- My parents and brother for their unwavering support, understanding and financial assistance during my university career.

TABLE OF CONTENTS

Acknowledgments		iii
Table of contents		iv
List of figures		viii
List of tables		x
List of abbreviations		xiii
Abstract		xiv
CHAPTER ONE:	**LITERATURE REVIEW**	1
1.1	INTRODUCTION	1
1.2	BLACK WATTLE INDUSTRY IN SOUTH AFRICA	2
1.2.1	History of black wattle	2
1.2.2	Characteristics and growth habits of black wattle	3
1.2.3	Distribution of black wattle plantations in South Africa	3
1.2.4	Uses of black wattle	4
A.	Black wattle bark	4
B.	Black wattle timber	5
1.2.5	Cultivation of black wattle	6
1.2.6	Breeding techniques	7
1.3	USE OF MOLECULAR FINGERPRINTING IN FORESTRY	8
1.3.1	Introduction	8
1.3.2	Isozyme or allozyme markers	10
1.3.3	DNA markers	11
A.	Single locus marker - Restriction Fragment Length Polymorphisms (RFLPs)	12
B.	Single locus marker - Simple Sequence Repeats (SSRs)	14
C.	Multi-locus marker - Amplified Fragment Length Polymorphism (AFLP)	16

D. Multi-locus marker - Random Amplified Polymorphic DNA (RAPD) 16

1.4 DETERMINATION OF GENETIC VARIATION 18

1.4.1 Determination of population genetic variation 19

1.4.2 Determination of inter-population genetic relationships 21

1.4.3 Estimation of population differentiation 23

A. Wright's *F*-statistic 23

B. Analysis of Molecular Variance (AMOVA) 25

1.5 AIMS 26

CHAPTER TWO: MATERIALS AND METHODS 28

2.1 INTRODUCTION 28

2.2 MATERIALS 28

2.3 METHODS 31

2.3.1 Leaf material 31

2.3.2 DNA extraction 31

2.3.3 Assessment of DNA purity 33

A. Agarose gel electrophoresis 33

B. Spectrophotometric analysis 34

2.3.4 RAPD fingerprinting 34

A. Preparation of template DNA stock 34

B. RAPD primers 35

C. Generation of RAPD fingerprints 36

2.3.5 Resolution of the amplified DNA 37

2.3.6 Repeatability and reproducibility 37

2.4 ANALYSIS OF RAPD FINGERPRINTS 38

2.4.1 Construction of digital fingerprints 39

2.4.2 Determination of primer performance 41

2.4.3 Estimation of genetic variation within groups 42

A. Estimated heterozygosity 42

B. Nei's gene diversity 43

C. Shannon's Information Index 43
2.4.4 Estimation of the genetic relatedness between and within groups 44
2.4.5 Estimation of group differentiation 44
2.5 UTILISATION OF COMPUTER SOFTWARE 45
2.5.1 POPGENE 45
2.5.2 ARLEQUIN 47
A. Arlequin input file 48
B. Execution of analysis 49
2.5.3 Treeview 52

CHAPTER THREE: RESULTS 53

3.1 INTRODUCTION 53
3.2 DNA ISOLATION 54
3.2.1 Confirmation of DNA isolation 54
3.2.2 Determination of DNA concentrations 55
3.3 GENERATION OF RAPD FINGERPRINTS 57
3.3.1 Optimization of RAPD protocol 57
3.3.2 Evaluation of RAPD fingerprints 58
3.3.3 Conversion of molecular RAPD fingerprints into digital fingerprints 60
3.4 ANALYSIS OF PRIMER PERFORMANCE 63
3.5 ANALYSIS OF GENETIC VARIATION 67
3.5.1 Estimation of genetic variation within groups 68
3.5.2 Estimation of genetic relatedness and differences between and within groups 72
3.5.3 Estimation of group differentiation 78

CHAPTER FOUR: DISCUSSION AND CONCLUSION **80**

4.1 INTRODUCTION 80
4.2 RAPD ANALYSES AND GENETIC VARIATION 80
4.3 POPULATION AND GROUP ANALYSIS 82
4.4 CONCLUSION 86
4.4.1 Recommendations for future work 86

REFERENCES **88**

APPENDIX A **a**

APPENDIX B **d**

APPENDIX C **w**

LIST OF FIGURES

Figure 1.1	Distribution of black wattle in KwaZulu-Natal.	4
Figure 2.1	Method of scoring a RAPD fingerprint.	40
Figure 2.2	*Notepad* display of partial digital genotypes of eleven individuals.	41
Figure 2.3	Example of a POPGENE input file containing digitized RAPD fingerprints.	46
Figure 2.4 (a)	Example of an ARLEQUIN input file.	48
Figure 2.4 (b)	(Cont.) Example of an ARLEQUIN input file.	49
Figure 2.5	ARLEQUIN programme showing the project wizard window.	50
Figure 2.6	ARLEQUIN programme showing the calculation settings window.	51
Figure 3.1	Detection of high molecular weight DNA on agarose.	55
Figure 3.2	Gels showing suitable and unsuitable primers tested.	59
Figure 3.3	Examples of RAPD fingerprint profiles displaying monomorphic and polymorphic bands.	64

Figure 3.4	Scatter plot depicting the relationship between the percentage GC content of the primers and the respective number of bands produced.	67
Figure 3.5	A graphical representation of the mean percentage of individuals displaying the dominant phenotype.	69
Figure 3.6	Range of pairwise identities (similarities) for each black wattle group.	73
Figure 3.7	Collection of pairwise analyses of Nei's identities and distances. Identity values above diagonal and distance values below diagonal.	74
Figure 3.8	Genetic trees showing genetic relationships between individuals of groups based on Nei's genetic distance.	76
Figure 3.9	Genetic tree showing the relationship between the seven groups based on Nei's genetic distance.	77
Figure 3.10	Genetic tree for individuals in the population based on Nei's gene diversity	78

LIST OF TABLES

Table 1.1	Formulae calculating mean heterozygosity (Hartl and Clark, 1997).	24
Table 1.2	Formulae for the calculation of *F*-statistics (Hartl and Clark, 1997).	25
Table 2.1	Black wattle group information.	30
Table 2.2	Primer information.	35
Table 2.3	Optimal reaction composition used to generate RAPD fingerprints.	36
Table 3.1	DNA purity ratios and concentrations.	56
Table 3.2	RAPD reaction mixture composition.	57
Table 3.3	Digital fingerprints of all individuals.	61
Table 3.4	Number of polymorphic and monomorphic loci generated by the selected primers.	65
Table 3.5	Percentage of GC content and the number of loci amplified by the selected primers.	66
Table 3.6	Frequencies of dominant phenotypes for each locus of the selected primers.	68

Table 3.7 Quantification measures of genetic variation. 70

Table 3.8 Estimated frequencies of the *null-alleles* (q) for each locus. 71

Table 3.9 Estimated heterozygosity (2pq) values for each locus. 72

Table 3.10 Nei's genetic identity (above diagonal) and genetic distance (below diagonal) obtained for the seven groups. 75

Table 3.11 Partitioning of total variation using AMOVA. 79

LIST OF ABBREVIATIONS

AFLP	Amplified Fragment Length Polymorphism
AMOVA	Analysis Of Molecular Variance
AP-PCR	Arbitrary-Primed Polymerase Chain Reaction
bp	Base Pairs
C-TAB	N – Cetyl –N, N, N –Trimethyl Ammonium Bromide
DNA	Deoxyribose Nucleic Acid
dNTP	Deoxynucleoside Triphosphate
EDTA	Ethylene Diamine Tetra –Acetic Acid di –Sodium Salt
ICFR	Institute for Commercial Forestry Research
NaCl	Sodium Chloride
OD	Optical Density
PCR	Polymerase Chain Reaction
PVP	Polyvinylpyrrolidone
RAPD	Randomly Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
rpm	Revolutions Per Minute
SSRs	Simple Sequence Repeats
TAE buffer	Tris–Sodium Acetate–Ethylene Diamene Tetra–Acetic Acid di–Sodium Salt
TBE buffer	Tris–Boric Acid–Ethylene Diamene Tetra–Acetic Acid di–Sodium Salt
TE buffer	Tris–Ethylene Diamene Tetra–Acetic Acid di–Sodium Salt
Tris-HCl	Tris Hydrochloride
UPGMA	Unweighted Pair Group Method With Arithmetic Mean
VNTR	Variable Number Tandem Repeats
WRI	Wattle Research Institute

ABSTRACT

Black wattle (*Acacia mearnsii)*, which belongs to the genus *Acacia,* is one of the many species of trees or hardwoods grown commercially in South Africa. Black wattle is a species indigenous to Australia and was introduced into South Africa by the van der Plank brothers in 1864. These trees are grown in South Africa because of its tannin-rich bark, the extract of which is used by the leather tanning industry. Black wattle is also grown for its timber, timber products and pulp. The introduction and cultivation history of black wattle suggests that the South African plantations contain limited genetic variation with relatedness amongst groups estimated to be high, thus implying a narrow genetic base in the South African black wattle population. In this investigation, Random Amplified Polymorphic DNA (RAPD) was used to estimate the genetic variation between seven different black wattle groups. A total number of 34 individuals obtained from different areas in South Africa were examined; Piet Retief (group 47 and 50: half-sibs), Kumbula (group 85: unrelated individuals), Howick (group 400: unrelated individuals) and an unknown area (groups 88, 89, 91: full-sibs). As this investigation was the first of its kind, a DNA isolation method as well as a PCR-RAPD protocol had to be modified. Total genomic DNA was successfully extracted using the CTAB DNA extraction method. This method removed large amounts of tannin present in the cells of the black wattle leaves and extracted high quality DNA to conduct between 50-100 RAPD reactions. The DNA purities ranged from 0.1 to 1.8, with an average of 1.46. A total of fourteen 10-mer RAPD primer sequences were randomly selected from the Operon Technologies primer list A, and tested in this investigation. Of the 14 primers used, only nine primers produced clear, single and repeatable bands. Therefore nine primers were selected for subsequent analyses. Ninety one loci that generated bands ranging from 300-3050 base pairs were produced. Seven to 13 loci per primer were generated. A total of 95.6 % of the loci were polymorphic. The overall expected mean heterozygosity (H = 0.3) obtained in this study was high in comparison to other studies conducted on acacias. The high levels of genetic variation were attributed to mating systems, dissortative mating and geographic distribution. The statistical packages POPGENE and ARLEQUIN were used to analyse the RAPD fingerprints. The genetic measures, Nei's diversity and Shannon's Information Index, showed that there was greater diversity exhibited (Nei's gene

diversity = 32.09 % and Shannon's = 48.31 %), in the whole population than in each of the groups (with average of Nei's gene diversity = 20.33 % and Shannon's = 34.64 %). With regards to individual group analyses, low levels of genetic variation was obtained in group 400 (unrelated), from the Howick region, and group 85 (unrelated), from the Kumbula region, (mean 0.14 and 0.17 respectively). The low genetic values were attributed to limited gene exchange occurring in these two areas, bottlenecks and selection pressures. Groups 88, 89 and 91, from the unknown region (full-sib groups), were the most variable in comparison to the other groups, with means of (0.27, 0.24 and 0.18 respectively). These high genetic variation values could be due to the fact that gene migration could have occurred between these groups and others in the area. It is thought that most acacias are insect-pollinated and this could have lead to gene migration between groups or populations, thereby explaining the high mean values. The gene flow obtained for the seven groups (F_{ST} = 0.174) indicated that great genetic differentiation existed in this population of black wattle studied. This value is higher in comparison to other woody species; however it is similar to other acacia species. UPGMA cluster analysis using Nei's unbiased genetic distance, revealed four distinct clusters of groups corresponding to the distribution areas represented in this study. The Howick (group 400: unrelated) and Kumbula (group 85: unrelated) were more closely related to each other than to the other groups, since both these groups are from Natal. The Piet Retief groups (groups 47 and 50: half-sibs), branched-off together, indicating that they are distinct from the other groups. The pairwise analysis of identity showed that the relationship between the group from Howick (group 400: unrelated) and all the other groups from the other regions was the lowest, ranging from 64 % to 79 %. The relationship between all the groups beside the group from Howick (group 400: unrelated) was reasonably high, ranging from 78 % to 90 %. This distance displayed by group 400 (unrelated) from Howick in relation to the groups, is attributed to the fact that it is frost resistant and the other groups not. Genetic variation was also detected and partitioned, between and within groups, by Analysis of Molecular Variance (AMOVA). Majority of the variation existed within groups (82.65 %) but significant differentiation was recorded between groups (17.44 %). This high level of within group differentiation may be explained by many aspects, such as the species breeding system, genetic drift or genetic isolation of groups or populations. The application of RAPD fingerprinting in black wattle has provided a more in depth understanding of the genetic variation

residing in the South African population. The results achieved implementing this technique has shown that significant genetic variation exists within the black wattle population in South Africa. The results obtained in this study are also important since it is contrary to the expectation that the black wattle population in South Africa has low genetic variation. This knowledge is of great value to genetically discriminate between individuals or groups, to improve the selection of superior genotypes and allowing improved quality control in breeding programmes and seed orchard management.

CHAPTER ONE

INTRODUCTION

1.1 INTRODUCTION

South Africa has the third largest and one of the oldest plantation resource areas in the southern hemisphere (Owen and van der Zel, 2000). South Africa is also a world leader in management of, and research on cultivated timber plantations (Owen and van der Zel, 2000). Though South Africa only has 0.07 % of the world's productive forest area, it has been estimated that it produces 1.2 % of the global industrial output. The Department of Forestry and Water Affairs (2003) estimated that a total area of about 1.5 million hectares of South Africa (1.4 %) is used for commercial forestry.

Forestry's importance to South Africa is largely due to the income gained from forestry products, for example paper, wood products and pulp. During 1999 and 2000 it was estimated that South African forests consisted of 38.9 % eucalypts, 52.2 % pine, 8.3 % wattle and 0.6 % other species (Department of Forestry and Water Affairs, 2003).

In South Africa, the wattle industry is based on the cultivation of black wattle, an acacia native to Australia. The total area of commercially grown wattle amounts to approximately 130 000 hectares, situated in a belt from south-eastern Mpumulanga through to KwaZulu-Natal. Black wattle is grown for its timber and bark products, namely tannin extracts, which is in great demand locally and internationally. Due to this demand, black wattle has increased in value to the forestry industry (Kevin, 2000).

1.2 BLACK WATTLE INDUSTRY IN SOUTH AFRICA

1.2.1 History of black wattle

Acacia mearnsii or black wattle belongs to the largest genus in the subfamily Mimosoideae and the second largest genus in the family Legumionosae (Sherry, 1971; Ross, 1972; Carr, 1976). This genus consists of 1100–1200 species of trees, shrubs and climbers (Harsh, 2000).

Black wattle was first introduced to South Africa, from Australia, in 1864 by Charles and John van der Plank (Sherry, 1971) and grown in KwaZulu-Natal for shade trees, firewood and shelterbelts. Around the 1880s, it was discovered that between 36–44 %, of the bark consisted of tannin, which was used to tan hides for leather. This stimulated a great demand for wattle bark extract internationally and initiated the widespread establishment of plantations of black wattle throughout South Africa (Sherry, 1971).

By the 1900s the growth and export of wattle bark extract had increased greatly. In 1907, the first wattle organization, the Natal Wattle Growers' Union was formed. This organization aimed at furthering the interests of wattle growers in Natal. In 1937, the South African Wattle Growers' Union was founded, to represent growers' interests in the industry (Sherry, 1971).

In 1947, the Wattle Research Institute (WRI) was founded which aimed at conducting research into wattle growing, silviculture, genetics, chemistry and entomology. WRI focused mainly on the reduction of growing costs, the improvement of yield, the selection of disease resistant genotypes and researching the major pests and diseases (Dunlop, 2002). In 1984, the WRI became known as the Institute for Commercial Forestry Research (ICFR), which conducted research on black wattle as well as eucalypts and pines. This institute, together with other organizations has, over the years, contributed greatly to the wattle industry (Dunlop, 2002).

1.2.2 Characteristics and growth habits of black wattle

Black wattle is usually between 10-15 metres tall, sometimes reaching 25 metres with a stem diameter of up to 60 centimetres. The bark is usually brownish-black, hard and fissured. Unlike other acacia, *Acacia mearnsii* has no thorns. The leaves are usually fern-like with an inflorescence globular head, consisting of 20-30 pale yellow white flowers (Joker, 2000).

Black wattle grows well from seed and is often one of the first species to re-colonize disturbed areas. Therefore it is regarded as a pioneer species. This ability to spread fast has lead to it being regarded as a weed in many countries (Kevin, 2000). It is also a fast growing species, increasing two-three metres per year, and is often used as a nurse species protecting and assisting the silviculture and management of slower growing species such as black wood (*Acacia melanoxylon*) (Kevin, 2000).

1.2.3 Distribution of black wattle plantations in South Africa

Black wattle plantations are found scattered throughout KwaZulu-Natal on a range of sites (Figure 1.1). Black wattle are found in areas with deep, moist alluvial soils to areas of low fertility skeletal soils. Generally, the life span of black wattle is between 15-20 years. However in regions where rainfall is more than 700 millimetres per year, wattle may live up to several decades, thus creating a better opportunity for commercial production (Kevin, 2000).

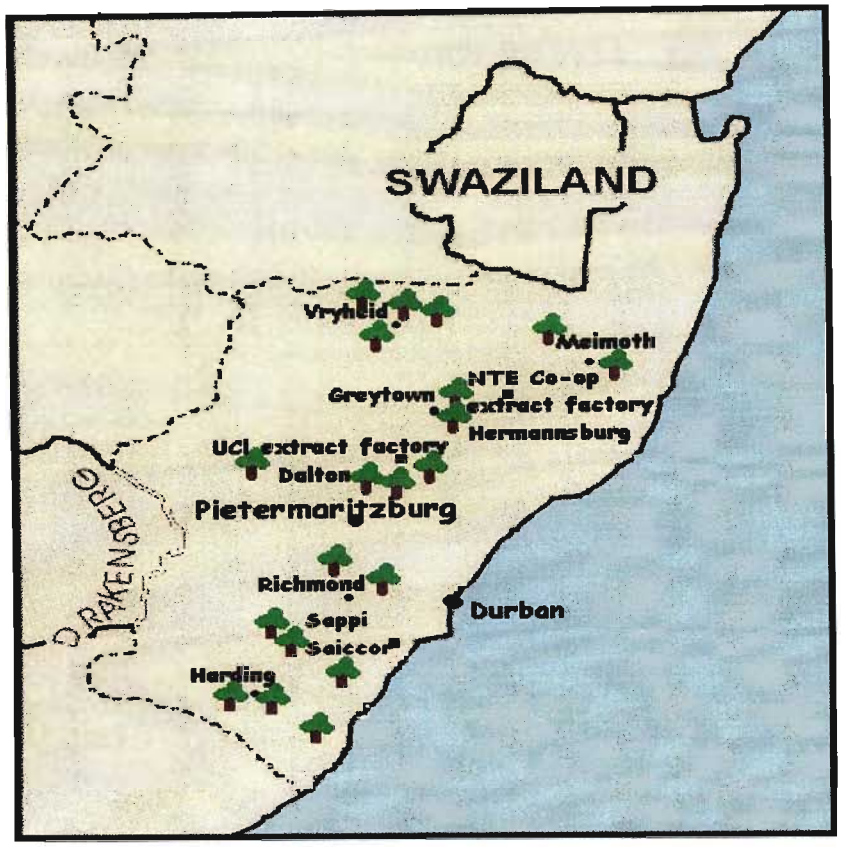

Figure 1.1 **Distribution of black wattle in KwaZulu-Natal.**

1.2.4 Uses of black wattle

After introduction to South Africa, black wattle was used for shelter, firewood, building and fencing poles. However, as the black wattle industry progressed over the years, so too its uses. Black wattle is today used mainly for its bark and timber products.

A. Black wattle bark

The main use of black wattle bark lies in its tannin content, which provides the greatest proportion of the financial return from the wattle growing industry in South Africa. The barks of some acacia species are among the richest sources of vegetable tannins known. More than 30 % of their dry weight consists of tannins (Sherry, 1971). Tannin

content varies with bark thickness and the age of the tree; the older the tree the more tannin it contains (Sherry, 1971). Tannins are used for tanning material, mainly leather. The tannin quickly penetrates the hide giving it firm, durable and light coloured leather (Dunlop, 2002). Tannins are also used for plywood and particle-board adhesives.

Apart from tanning, black wattle is also used for a number of other purposes (Sherry, 1971):

- A compound from wattle extract can be used for the conditioning of clay in the ceramic industry,
- Plastic mouldings,
- Manufacture of adhesives and bonding agents as a basis for surface coatings of metal,
- The production of ion-exchange resins and chelating,
- The flotation treatment of ores,
- A source of phenolic chemicals,
- Oil-well drilling, and
- The control of virus diseases in plants.

B. Black wattle timber

Many products are made from black wattle timber. These include parquet flooring blocks, furniture, hardwood, rayon, charcoal, mining timber, structural timber and fencing poles. Traditionally, black wattle were used for building purposes and firewood, which was usually produced from thinning plantations destined for timber production (Dunlop, 2002). More recently, black wattle has been grown for pulp (Dunlop, 2002). It is a high yielding pulp species that requires low chemical input in the pulping process, and provides favourable paper strength properties. The current demand of black wattle pulpwood is 60 000-100 000 green tonnes per year (Dunlop, 2002).

1.2.5 Cultivation of black wattle

Black wattle is usually established by implementing three different methods. This is done firstly, by natural regeneration of existing seed; secondly, by line-sowing and thirdly, by planting nursery raised seedlings. The seedling method was only adopted in South Africa after 1970, when genetically improved seed became available from seed orchards. Black wattle seed can remain dormant in the soil for over 50 years. When the conditions become favourable black wattle will germinate. In the industry, germination is induced by heating or scarifying the seed. In the field this is achieved by either burning the plantation residue after clear-felling, or by ripping lines across the plantation where the trees are intended to grow. This will result in a vast number of seedlings germinating, which require a series of spacing and thinning operations to achieve an acceptably well stocked plantation (Smith and Dunlop, 2002).

The line-sowing method entails making shallow drill-lines in the soil; (five to 10 centimetres deep) three metres apart where the trees are intended to grow. Seed is then poured into these lines at a rate of approximately three and a half kilograms per hectare. The removed soil is then used to cover the seeds. When these seeds germinate, a 'spacing and thinning' regime is required to produce a plantation (Smith and Dunlop, 2002).

The seedling method involves planting, nursery raised seedlings in lines, three metres apart with one and a half metres between each seedling. This gives a stocking of 2 222 stems per hectare, which is followed up with two thinning operations during a 10 year rotation. This method allows for the use of genetically improved seed and the resulting plantations, if managed properly; should yield more bark and timber than plantations established using natural regeneration or line-sowing (Smith and Dunlop, 2002). The ideal final stocking, when using any one of the three methods of establishment, is 1 500 stems per hectare (Smith and Dunlop, 2002).

1.2.6 Breeding techniques

Breeding techniques in black wattle date back to long before the WRI was established. As early as 1927, Osborn, an early breeder of black wattle, had planted 300 open-pollinated trees and noted that black wattle differed in a range of characteristics (Dunlop, 2002). Since then, a number of conventional methods of genetic improvement through selection, testing and breeding have been practised with dramatic results.

Genetic techniques have, in the past, been used to produce superior trees for commercial forestry operations. These programmes rely on the broad natural variability of genotypes within populations. Traditionally selections were made for 'investment' traits such as straightness, form and superior growth yield, although 'insurance' traits such as disease resistance are also included. However recently, tree breeders have begun to include wood quality characteristics in their selection as well (Trotter, 1990).

Today, the main aims of the black wattle breeding programme are (Dunlop, 2002):

- To produce disease and frost resistant trees,
- To increase the fibre yield and quality of black wattle wood, while maintaining an acceptable bark quality, and
- To supply the industry with improved germplasm.

In order for these characteristics to be improved, the genetic make-up of the plant need to be known. It is here that molecular techniques play an important role. Molecular techniques can be applied in the breeding programme of black wattle to further increase genetic gains by way of (Dunlop, 2002):

- Construction of genetic linkage maps that lead to marker-assisted selection,
- Germplasm identification,
- Genotyping and accurate identification of genotypes,
- Detection of genetic variation, and
- Early selection using molecular markers.

Genotyping has the potential to identify clones in seed orchards where mislabelling errors of 10-15 % are common (Wheeler and Jech, 1992). The characterisation of markers linked to traits of commercial interest such as wood quality and disease resistance will allow selection of progeny at the nursery stage (Butcher *et al.*, 1998).

DNA markers may be valuable as it provides an understanding of the genetic variation and diversity present between different black wattle populations and among individuals in black wattle populations, in South Africa. This information is important in the breeding of black wattle, so that appropriate sampling strategies can be implemented to create a broad genetic base. This is important since black wattle in South Africa is presumed to have a narrow genetic base. DNA markers are powerful tools for addressing genetic questions which relate to breeding programmes.

1.3 USE OF MOLECULAR FINGERPRINTING IN FORESTRY

1.3.1 Introduction

Evolution is a necessary process for the survival of a species. All organisms face stresses from natural sources such as competition, predation, human pollution, habitat destruction, and exploitation of commercially valuable species (Wright, 1978; Magurran and May, 1999). As a result of these stresses, the environment that most organisms live in is highly variable. If all individuals within a population were genetically similar, it is likely that they would all be equally susceptible to various biological forces. These forces include human intervention, natural phenomena such as drought, floods, drastic fluctuations in temperature, disease and insect attack (Wright, 1978; Magurran and May, 1999). It is therefore important that individuals within populations have the ability to adapt to these changing environments. This adaptive potential is the basis for natural selection and is largely genetically determined (Avise, 1994). In order for a species to survive, some populations of a species must exhibit sufficient genetic variability so that individuals within that population are able to adapt to the changing environment. The measurement of the inherent variation within populations and between populations is

therefore the business of geneticists, breeders and evolutionists (Wright, 1978; Avise, 1994).

The measurement of genetic variation can be undertaken by using traditional Mendelian methods of making crosses and scoring the phenotypes of the offspring in one or more generations (Avise, 1994). This is, however, too inefficient for a detailed estimate of genetic variation. The process is primarily restricted to phenotypic characters, which are limited in number (Avise, 1994). The process is time consuming and does not always yield precise information on genotypes (homozygous dominant *versus* heterozygote). These limitations are overcome by employing molecular techniques that are able to identify genetic differences directly, by analysing DNA or gene products through allozyme marker analysis (Parker *et al.*, 1998).

DNA fingerprinting provides a permanent and reliable method of individual identification. Another useful application of DNA fingerprinting is that it can provide an understanding of the genetic variation present in the population and among individuals. DNA fingerprinting, identification analysis, profiling or typing all refers to the characterization of one or more relatively rare features of an individual's genome.

There are a vast number of molecular tools available for analysing genetic variation in populations. The following factors need to be taken into account before selecting the appropriate technique (Parker *et al.*, 1998):

- The extent of genetic polymorphism required to best answer the question posed,
- The analytical or statistical approach available for the techniques applied, and
- The pragmatics of time and cost of material.

1.3.2 Isozyme or allozyme markers

Allozymes are defined as variant forms of an enzyme that are coded for by different alleles at the same locus. Isozymes, on the other hand, are defined as variant forms of an enzyme, that are coded for by genes located at different loci. Electrophoretic separation of allelic variants of enzymes has been used to study genetic variation in populations of a wide variety of species for many years (Avise, 1994; Magurran and May, 1999).

Isozyme analysis is relatively inexpensive and straightforward and allows for data to be collected relatively quickly from large samples (Chamberlain and Hubert, 1998). However, despite the advantages, isozymes are not without their shortcomings. Firstly, the enzymes used in the isozyme analysis are a limited and non-representative sample in comparison to the total array of proteins present in an organism. Secondly, electrophoretic differences are only one kind of difference that exists between genetically related proteins (Soltis and Soltis, 1990). Thirdly, the number of isozyme markers available is limited, and their expression is often restricted to specific tissue or developmental states (Tanksley and Orton, 1983; Winter and Kahl, 1995).

Allozyme variation has been studied in numerous plants. Harsh (2000) used allozyme analysis to determine the population genetic structure of two *Acacia brevispica* populations in Kenya. This study was motivated by the fact that several acacia species flower simultaneously and share pollinators; therefore allowing interspecific pollen transfer. There was suspicion of a gene transfer between the two populations that are ten kilometres apart. The results from the allozyme analysis showed that individuals exhibited a high heterozygosity frequency (27 %), while differences between populations were low (1.3 %). This proved that high levels of gene movement had occurred, between populations located within 10 kilometres of one another.

Broadhurst and Coates (2002) conducted an investigation into the genetic variation within and between rare and geographically widespread taxa of the *Acacia acuminate* Benth (Mimosaceae), a West Australian complex, by using allozymes. The genetic structure of 25 populations was investigated using 16 polymorphic allozyme loci. All taxa

and variants exhibited relatively high levels of genetic variation compared with other woody angiosperms with similar geographic distribution, which may represent a series of hybridization events between the various taxa.

1.3.3 DNA markers

There are different molecular markers that are used to fingerprint individuals. Some of these markers are coding genes that are sequenced and analysed for nucleotide differences (Avise, 1994). This is an expensive and time consuming method to measure genetic variation. Other markers are not genes in the classical sense, that is, these markers do not encode for a particular protein product. However, these markers represent useful and constant 'landmarks' in the genome that displays a large amount of polymorphism (Butcher *et al.,* 1998). The different polymorphic DNA types are termed 'alleles' and these are the essential ingredients required for the fingerprinting process (Butcher *et al.,* 1998). The type of genetic markers available for fingerprinting individuals can be sub-divided into two groups according to the number of loci involved. The first group or class of genetic markers, known as the multi-locus markers, analyze several loci simultaneously, yielding a DNA fingerprint in one step. The single-locus markers, on the other hand, aim at one locus and therefore require the combination of several locus-specific assays to achieve a similar multi-locus type DNA profile (Krawczak and Schmidtke, 1998). At present a number of techniques are available that fall under either multi-locus methods of detection or single-locus methods.

The most widely used DNA markers consist of the specific markers, which require prior knowledge of the genome, while the random markers do not. Examples of specific DNA markers include Restriction Fragment Length Polymorphisms (RFLPs) (Tanksley *et al.,* 1989) and Simple Sequence Repeats (SSRs), while the random DNA markers include Random Amplified Polymorphic DNA (RAPD) (Welsh and McClelland, 1990) and Amplified Fragment Length Polymorphisms (AFLPs) (Vos *et al.,* 1995).

A. Single locus marker - Restriction Fragment Length Polymorphisms (RFLPs)

RFLP, a single-locus marker, was the first DNA marker used in DNA fingerprinting (Parker *et al.*, 1998; Matyas, 1999). RFLPs are based on heritable polymorphisms in the length of fragments produced by digestion of DNA with restriction enzymes. If a restriction site is present on a strand of DNA, the DNA strand will be cleaved in the presence of the corresponding restriction enzyme that would recognize the site. This would result in the strand splitting into two and thus showing up as two different bands on the gel. Differences in the lengths of fragments generated occur as a result of mutations in the base sequence of the DNA through insertions, deletions and base substitutions, which is carried through to the offspring. The number of fragments obtained and their sizes would depend upon the number of restriction sites and their positions in the DNA, which in turn, would be characteristic of the genome and the genomic sequence being analysed (Cheliak and Rogers, 1990; Krawczak and Schmidtke, 1998; Parker *et al.*, 1998; Matyas, 1999).

RFLPs have a number of properties, which make them amenable for use as genetic markers (Beckmann and Soller, 1986; Tanksley *et al.*, 1989). RFLPs are suitable for parentage analysis and mapping (Cheliak and Rogers, 1990). It has also been found that RFLP analysis is useful in estimating relatedness among groups and within groups (Parker *et al.*, 1998). Another advantage of RFLPs is the ability to generate fingerprints with a number of restriction enzymes employed simultaneously. The result of such simultaneous digestions is typical DNA fingerprints that may be highly polymorphic due to the large number of different sized fragments (Parker *et al.*, 1998).

RFLP markers have many uses and have been applied in a variety of applications in forestry (Neale and Williams, 1991):

- Studies of genome organization,
- Population and evolutionary genetic studies,
- Genetic identification in tree improvement programmes, and
- Markers for gene isolation and cloning.

RFLPs are co-dominant markers, and are useful to distinguish between homozygous dominant individuals and heterozygotes. However, in order to amplify a specific region, sequence information is required to develop the RFLPs, which are species specific, time consuming and very expensive. (Parker *et al.*, 1998; Matyas, 1999).

RFLP markers have been used in the domestication and breeding programmes of acacia, to assess genetic variation present in natural seed produced populations (Butcher *et al.*, 1998). Only 56 % of genetic variation detected in the natural populations was represented in the seed production area in Indonesia. The majority of trees originated from a region characterized by low levels of diversity, indicating that the breeding programme could be improved by expanding the genetic base (Butcher *et al.*, 1998).

Another use of RFLP analyses is to discriminate between individuals, allowing improved quality control in breeding programmes and seed orchard management. In *Acacia mangium,* full-sib pedigrees have been screened using RFLP markers to remove progeny derived from foreign pollen or selfing (Butcher *et al.*, 1998). The effect of different pollination techniques has also been investigated with these markers (Butcher *et al.*, 1998). A linkage map is currently being developed for *Acacia mangium* using RFLPs and microsatellites. The potential of these markers lies in marker assisted selection, where markers linked to quantitative traits such as disease resistance and pulp yield may be utilised in the selection of offspring at the nursery stage, thereby improving the efficiency of breeding programmes. RFLPs have also been used to detect genetic diversity in coconut (*Cocos nucifera*) (Lebrun *et al.*, 1998).

B. Single locus marker - Simple Sequence Repeats (SSRs)

Simple Sequence Repeats (SSRs) are loci containing few to many repeat sequences, and are found in numerous areas in the genome of an organism (Avise, 1994). Due to the large number of alleles found for many SSRs, a large amount of variation can be detected, therefore making these markers popular fingerprinting tools (Avise, 1994). They are usually sub-divided into two classes according to the number of nucleotides making up a single repeat unit, namely, the length of the actual repeat (Krawczak and Schmidtke, 1998). Minisatellites, otherwise known as variable number tandem repeats (VNTRs), comprise short tandem repeats of about 40 base pairs in length, whereas microsatellites are simpler in that it constitutes between one and four nucleotides per repeat unit (Avise, 1994).

Minisatellites, first discovered in 1985 by Alec Jeffreys and his research team, were one of the early DNA markers employed to generate what is now known as a DNA fingerprint. The minisatellites 33.6 and 33.15 are composed of several single-locus probes that work simultaneously to produce a barcode-like multi-locus DNA fingerprint (Jeffreys *et al.,* 1985). Minisatellites each detect approximately 17 highly polymorphic bands, making them still a popular method to utilise today (Jeffreys *et al.,* 1985).

Microsatellites are a type of DNA marker, which are important for individual genotyping, and studies of gene flow in forest trees. Microsatellites consist of segments of DNA containing numerous tandem repeats of a short 'motif' sequence, usually one to six bases (for example: CACACA... or TGTGTG...) (Mullis and Faloona, 1987; Goldstein and Schlotterer, 1999). The repeating T-G or C-A units on one DNA strand will be complementary to repeating A-C or G-T units on the other strand, respectively. Microsatellite loci are analysed by amplifying the target region using Polymerase Chain Reaction (PCR) (Williams *et al.,* 1990), followed by electrophoresis through a polyacrylamide gel to allow resolution of alleles that may differ in size by as few as two base pairs (Parker *et al.*, 1998).

Microsatellites have many uses and have several advantages over other techniques (Parker *et al.*, 1998):

- Microsatellites are the preferred markers to assign parentage,
- Microsatellites are also useful to determine relatedness between individuals and the amount of diversity between populations,
- Microsatellites provide information from individual loci and therefore are amenable to calculations of allele frequencies necessary for population studies,
- Microsatellites typically exhibit high levels of gene diversity and therefore are expected to routinely assay high levels of polymorphism, and
- Once primers have been developed, the protocol consists of PCR and electrophoresis, eliminating the time consuming gel blot hybridizations.

A disadvantage of microsatellites, lies in the time consuming process of identifying these regions from genomic libraries, which are newly constructed for each species. Known primers are not thought to be likely to amplify the same locus across related taxa, unless the microsatellite region is flanked by highly conserved sequences where priming sites are located (Matyas, 1999). Another disadvantage is the effort and expense required for the development of microsatellites.

Microsatellites are also extremely sensitive; under optimal conditions these markers can detect loci at digest concentrations of less than 100 ng (Krawczak and Schmidtke, 1998). However, one major drawback of these markers is that it requires prior sequence information, which is more often than not available. This means that detailed research about the organism's sequence composition is necessary before the fingerprinting analysis can be conducted (Bruford and Wayne, 1993).

In a study undertaken by Lefort *et al.* (1998), microsatellites were used to determine the genetic relationship among selected oaks (*Quercus robur*). Using nine loci it was established that five selected trees were not closely related and would therefore form a suitable seed source for an advanced breeding programme. In another study, nine

microsatellite loci were used to detect seed contaminants and confirm half-sib relationships in seed lots from single *Quercus robur* trees (Lexer *et al.,* 1999).

C. Multi-locus marker - Amplified Fragment Length Polymorphism (AFLP)

Amplified Fragment Length Polymorphism (AFLP) is based upon the selective amplification of a number of genomic fragments (Vos *et al.,* 1995). This is achieved by digesting DNA with two restriction enzymes, followed by ligation of complementary adapter sequences, extended with a variable number of 3' 'selective' nucleotides as primers (Matyas, 1999). The subset of amplified fragments is then analysed by denaturing polyacrylamide gel electrophoresis to generate the fingerprint. The DNA polymorphisms identified using AFLP are typically inherited in Mendelian fashion and may therefore be used for typing, identification of molecular markers and mapping of genetic loci (Matyas, 1999)

D. Multi-locus marker - Random Amplified Polymorphic DNA (RAPD)

Random Amplified Polymorphic DNA (RAPD) is a method that utilises a short primer of arbitrary nucleotide sequence of approximately ten base pairs long, which no specific knowledge of a particular DNA sequence is required (Williams *et al.,* 1990; Parker *et al.,* 1998). The primer is annealed in the first few cycles of PCR at low stringency. The low stringency ensures the generation of products by allowing priming with mismatches between primer and template. The subsequent PCR cycles are then performed at a higher stringency after the generation of some initial products that now have ends complementary to the primer. Allelic variation consists of the presence or absence of particular amplification products, which can be separated on agarose gels stained with ethidium bromide. The RAPD process typically reveals several polymorphic fragments per primer within populations; other fragments may appear as monomorphic bands within or across populations (Parker *et al.,* 1998).

The use of RAPD markers has a number of advantages over methods involving RFLPs or isozymes. Prior sequence information of the DNA is not required; the protocol is quick and simple to perform and is amenable to automation; the use of radioactivity is not necessary and the RAPD technique requires only a nanogram quantity of DNA as opposed to microgram quantities for RFLPs (Powell *et al.*, 1996; Parker *et al.*, 1998). The RAPD procedure has been proved to be very valuable and is extensively employed especially in forestry, since it is rapid, inexpensive and answers particular questions of interest. Repeatability can be assured by strictly maintaining laboratory conditions (Williams *et al.*, 1990).

The main limitation of RAPD markers is their dominant nature, resulting in the inability to determine whether an individual locus is homozygous or heterozygous (Smith and Wayne, 1996). This is particularly important when dealing with recessive traits. This can be overcome by obtaining two closely linked RAPD markers; each specific for one of the alleles and using them as a pair (Williams *et al.*, 1990). Amplification of both markers of the pair would be diagnostic for a heterozygous genomic region. Alternatively, a RAPD band can be excised and used as a probe to establish sequence similarity of two segregating RAPD bands of different molecular weights, thereby allowing pairs of bands to be scored as a single co-dominant marker (Williams *et al.*, 1990).

There are a number of applications of RAPDs. RAPDs have been found to be useful in the assignment of parentage (Hadrys *et al.*, 1993). RAPDs have also been useful in estimating relatedness among individuals or differentiation among groups (Ashburner *et al.*, 1997; Parker *et al.*, 1998). In a study conducted by Shrestha *et al.* (2002), RAPDs were used to assess patterns of genetic variation within and between 12 populations of *Acacia radiana* from the Arava valley and western Negev. A high level of genetic polymorphism was recorded within populations. These populations were also highly differentiated from one another. Therefore from a conservation point of view, it is important to conserve each population separately as loss of one population could lead to loss of gene variation. RAPDs have also proven effective in the detection of hybridization and introgression in several groups of plants. Hybridization and introgression between rare and common species has significant consequences for the preservation of genetic diversity. Rare species may be endangered either by

outbreeding depression or by genetic assimilation. The detection of genetic mixing therefore plays an important role in conservation biology (Smith and Wayne, 1996). RAPDs are also used for detecting variation in non-coding regions of nuclear or cytoplasmic DNA and are used in the construction of genetic maps (Matyas, 1999). RAPDs were also used in studies to determine genetic variation conducted on *Chamaecyparis formosensis* and *Chamaecyparis taiwanensis* by Hwang *et al.* (2001); *Picea abies* population (Troggio *et al.*, 2001); *Pinus resinosa* (Mosseler *et al.*, 1992); *Prunus mahaleb* (Rosaceae) (Jordano and Godoy, 2000); *Eucalyptus urophyllaa* and *Eucalyptus grandis* clones (Muro-Abad *et al.*, 2001); *Araucaria araucana* trees (Bekessey *et al.*, 2002).

1.4 DETERMINATION OF GENETIC VARIATION

Genetic diversity refers to the variation at the level of individual genes and provides the basis of the ability of organisms to adapt to changes in their environment through natural selection (Chamberlain and Hubert, 1998). Genetic variation within a population occurs when there is more than one allele present at a given locus. Some loci may have only one allele in an entire population. When the alleles are identical, the individual is homozygous for that trait. This locus is said to be fixed and the population to be monomorphic (Avise, 1994; Chamberlain and Hubert, 1998). When the pair of alleles consists of two different alleles, the individual within the population is said to be heterozygous. The population is therefore said to be both segregating and polymorphic for the locus (Avise, 1994; Chamberlain and Hubert, 1998). When the number of alleles at the locus increases, the proportion of heterozygous individuals in the population increases as well. The proportion of heterozygotes therefore indicates the proportion of genetic variation. The more variation, the better the chance that at least some of the individuals will have an allelic variant that is suited for the new environment, and will produce offspring with the variant that will in turn reproduce and continue the population into subsequent generations (Avise, 1994; Chamberlain and Hubert, 1998). The inability to adapt to changing conditions greatly increases the risk of extinction. Populations with little genetic variation are more vulnerable to the arrival of new pests or diseases, pollution, changes in climate and habitat destruction due to human activities or other

catastrophic events (Chamberlain and Hubert, 1998). Gene conservation management aimed to save adaptive genetic diversity should be based on the knowledge of the genetic basis of adaptation (Chamberlain and Hubert, 1998).

There are many ways to determine genetic variation. The following are a few of the available formulae.

1.4.1 Determination of population genetic variation

Many methods exist to measure genetic variation, example proportion of polymorphic loci and determining the genetic heterozygosity.

The proportion of polymorphic loci or polymorphism (P) indicates the proportion of loci that show evidence of more than one allele (Hartl and Clark, 1997). This is calculated by dividing the total number of polymorphic loci present by the total number of loci:

$$P = \frac{\text{Number of polymorphic loci}}{\text{Total number of loci scored}}$$

Another measure of genetic variation in a population is the average frequency of heterozygous individuals per locus (H) of the population (Hartl and Clark, 1997). This measure is known as genetic heterozygosity. It is calculated by first determining the frequency of heterozygotes at each locus and then averaging these frequencies over all loci (Bader, 1998). This value may be determined from the raw data, or it may be estimated from the frequency data where:

$$H = 1 - \sum_{i=1}^{m} p_i^2$$

H is the heterozygosity estimated at locus i and p^2 is the frequency of the i^{th} allele at that locus.

Another method of calculating the genetic heterozygosity was proposed by Lynch and Milligan (1994), which is specifically applicable to dominant markers. According to their method, the frequency of the number of bands absent at a particular locus, the null-phenotype, represents the frequency of the recessive genotype (q^2). The frequency of the *null-allele* (q) may be estimated from this value, by taking the square root of the observed proportion of individuals exhibiting this recessive genotype. The frequency of p is then calculated, by applying the formula p+q=1. From the estimated allele frequencies it is then possible to estimate heterozygosity by applying the formula 2pq. It is however important to mention that Hardy-Weinberg equilibrium can not be empirically obtained using dominant markers, therefore this calculated heterozygosity represents the expected heterozygosity given Hardy-Weinberg equilibrium (Lynch and Milligan, 1994). The mean heterozygosity is obtained by summing all the heterozygosity values of all the individuals and dividing by the total number of individuals.

$$H = \frac{H1+H2+H3...Hn}{\text{Total number of individuals}}$$

Lynch and Milligan (1994) also suggested that only those loci whose frequency of the null-phenotype exceeds a value of 3/n (where n is the number of individuals in the population sampled), should be included in the analysis, and therefore taken as being polymorphic. If the frequency of the null-phenotype is less than 3/n, then the particular locus is regarded as being monomorphic.

1.4.2 Determination of inter-population genetic relationships

Genetic distance measures, summarize the genetic differences between any two individuals, populations or species. These differences are quantified by applying a variety of different formulae to provide an estimate of genetic relatedness. The methods measuring Nei's (Nei, 1972), Roger's (Wright, 1978) and Shannon's (Magurran, 1988) diversity index will be examined.

The statistical calculations formulated by Nei (1972), include genetic identity (I), which estimates the alleles that are identical between two populations, thus similarity; while genetic distance (D), estimates the number of allele differences between two populations. Genetic identity is calculated as follows:

$$I = \sum xiyi \,/\, (\sum xi^2 \sum yi^2)0.5$$

Where x and y are two populations, and *xi* and *yi* are the frequencies of the *i*th alleles in the populations x and y respectively.

When dealing with multiple loci, the identity value is calculated as follows:

$$I = Jxy/\ (Jx\ Jy)0.5$$

Where Jx, Jy and Jxy are the arithmetic means across all loci of $\sum xi^2$. $\sum yi^2$ and $\sum xiyi$ respectively. Genetic distance is given by:

$$D = -1n$$

The values for I may range from zero to one and the values for genetic distance from zero to infinity. Closely related populations usually have I values > 0.9 and D values < 0.1. Distant related populations have I values < 0.8 and D values > 0.2.

Another method of genetic distance measure was proposed by Roger (Wright, 1978). For this calculation, certain assumptions are made. These include that for a given locus with *m* alleles, let *xi* and *yi* be the frequencies of the *i*th allele in populations x and y. Roger's D (distance) and S (similarity) is calculated as follows:

$$D = [0.5 \sum(xi - yi)^2]0.5$$

and

$$S = 1\text{-}D$$

Roger's D value ranges between zero and one.

Shannon's diversity index (Magurran, 1988) is another measure of genetic diversity between species. The measure of Shannon's diversity index (Magurran, 1988) is similar to that of Nei's measure (Nei, 1972) in that the higher the value, the greater the genetic diversity within the population. Shannon's (Magurran, 1988) index is scaled on a range of zero to one, with one representing the maximum possible diversity for the species. It is represented by the following equation:

$$H = -\sum_{i=1}^{s} p_i \ln p_i$$

where: H' = Shannon's index, p_i = proportional contribution by species, and s = number of species (Magurran, 1988).

1.4.3 Estimation of population differentiation

A. Wright's *F*-statistic

When a population is divided into isolated subpopulations, there is less heterozygosity in each subpopulation than there would be if the population was undivided. Founder effects that act upon the different subpopulations cause the allele frequencies to be different from the larger population (Hartl and Clark, 1997). The subpopulations are also smaller in size than the larger population; therefore, since the allele frequency in each generation represents a sample of the previous generation's allele frequency, there will be greater sampling error in these small groups than there would be in a larger undifferentiated population. Therefore, genetic drift will cause these smaller groups to have different allele frequencies and allele fixation at a higher rate than a larger undifferentiated population (Hartl and Clark, 1997).

The decrease in heterozygosity due to subdivision within a population is known as Wright's *F*-statistic, or the fixation index. This index, therefore, measures the degree of differentiation, between groups or subpopulations, by measuring the loss of heterozygosity. This is done by looking at the genotypic proportion (Hartl and Clark, 1997). The fixation index ranges from zero (indicating no differentiation between the overall population and its subpopulations) to a maximum of value one.

The fixation index can be obtained at three differentiated hierarchical levels of a population structure, which consist of demes. A deme is a group of locally inter-breeding population. These include (Hartl and Clark, 1997):

- The degree of differentiation within a population among groups of demes (F_{SG}),
- Within groups among demes (F_{GT}), and
- Within a population among demes (F_{ST}).

To determine the fixation index, the mean heterozygosity at each of these three levels should be calculated. For a locus with two alternate alleles, allele frequency is symbolized as *p* and the alternative form of the allele, *q*, is equal to 1–*p*. The mean heterozygosity for each level, for a population subdivided into three hierarchical levels, is then determined (Table 1.1). At each of the three levels of hierarchy, the mean allele frequency *p* is determined. This *p* value is multiplied by 2(1–*p*); this product is the frequency of heterozygotes for that allele (Hartl and Clark, 1997).

Table 1.1 Formulae calculating mean heterozygosity (Hartl and Clark, 1997).

Level of population hierarchy	Heterozygosity
Demes	$H_S = \sum_i \sum_j 2p_{ij}(1-p_{ij})$
Groups of demes	$H_G = \sum_i 2\frac{\sum_j p_{ij}}{n_i}\left(1-\frac{\sum_j p_{ij}}{n_i}\right)$
Total population	$H_T = \sum 2\frac{\sum_i \sum_j p_{ij}}{n_i}\left(1-\frac{\sum_i \sum_j p_{ij}}{n_i}\right)$

F-statistics (Table 1.2) can be calculated once the heterozygosity at each hierarchical level is obtained (Hartl and Clark, 1997):

Table 1.2 Formulae for the calculation of *F*-statistics (Hartl and Clark, 1997).

Level of population hierarchy	*F*-statistic
Among demes within group	$F_{SG} = \frac{H_G - H_S}{H_G}$
Among groups within population	$F_{GT} = \frac{H_T - H_G}{H_T}$
Among demes within population	$F_{ST} = \frac{H_T - H_S}{H_T}$

B. Analysis of Molecular Variance (AMOVA)

Wright's *F*-statistic is based upon comparison of gene frequencies among demes; however, molecular data reveals not only the frequency of molecular markers, but also the number of mutational differences between different genes. A technique that could be used to estimate population differentiation by analyzing differences between molecular sequences, rather than assumed Mendelian gene frequencies, would therefore be very useful (Excoffier *et al.,* 1992).

Analysis of Molecular Variance (AMOVA) is a method of estimating and testing differences between many subpopulations from molecular data. A variety of molecular marker data (for example, RFLP or AFLP), direct sequence data, or phylogenetic trees may be analysed using this method (Excoffier *et al.,* 1992). AMOVA works by using raw data, obtained from gels; example the value of one denotes the presence of a band and zero, its absence. AMOVA uses this data to create a Euclidean distance matrix between samples to measure the genetic structure of the population (Excoffier *et al.,* 1992). This is done by subtracting the Boolean vector of one haplotype from another according to the formula (p_j–p_k). The squared Euclidean distances are calculated from the equation:

$$\theta^2_{jk} = (p_j - p_k)'W\,(p_j - p_k).$$

where W is a weighting matrix.

These values obtained are then arranged into a matrix and divided into submatrices equivalent to subdivisions within the population (Excoffier *et al.*, 1992).

AMOVA is a hierarchical analysis of variance that partitions the total variance into (Excoffier *et al.*, 1992):

- Diversity among groups of populations,
- Diversity among the populations within groups, and
- Diversity among the individuals within a population.

1.5 AIMS

When black wattle was first introduced into South Africa, it was brought into the country 'in a bag of seeds'. No information regarding its genetic variability was known. As the years progressed, breeding of black wattle on a large scale had occurred and breeding programmes involved crossing trees with each other. Therefore, many of the groups in the breeding population were related. This implies that the genetic variability between and within groups should be low; and therefore the South African black wattle population is presumed to have a narrow genetic base. This is problematic in the wattle breeding industry as improved productivity relies upon the exploitation of genetic variation.

The aim of this investigation was to estimate the genetic variation between different black wattle groups, using the RAPD fingerprinting. The RAPD technique was chosen as it has proven to be successful in determining genetic differences between individuals of many species. It also provides a powerful tool for addressing genetic questions related to breeding and domestication programmes. RAPDs are valuable in determining the

level and structuring of diversity and the identification of provenance effects, which may influence sampling strategies to encompass a broad genetic base. Thus an understanding of the extent of the genetic variation would greatly contribute to the breeding and cultivation programmes in the black wattle industry.

This research involved the following components:

- Extraction of DNA suitable for PCR from a number of black wattle groups,
- Identification of polymorphic RAPD primers, and
- Determination of the genetic variation within and amongst the different groups using suitable statistical procedures.

CHAPTER TWO

MATERIALS AND METHODS

2.1 INTRODUCTION

An investigation was undertaken to assess the genetic variation in the South African black wattle population, as it is expected that genetic variability is low due to the limited number of seed that the South African plantations developed from. In this investigation, an analysis of genetic variation was conducted on seven different black wattle groups containing full-sib, half-sib and unrelated progenies, using Randomly Amplified Polymorphic DNA (RAPD). A total of 34 individuals were studied in this investigation. The genetic variation was determined within and amongst the different groups using suitable statistical procedures.

The recipes of all solutions have been taken up in Appendix A.

2.2 MATERIALS

Knowledge of the introduction history of black wattle into South Africa has indicated that the trees participating in the black wattle breeding programme should show a high degree of relatedness. The major objective of this investigation was to distinguish between closely related South African black wattle individuals. Thus, a range of relatives at different levels of relatedness were included in the investigation.

The following characteristics and requirements of the current black wattle breeding programme, therefore, motivated the inclusion of full-sib, half-sib, distantly related and unrelated individuals:

- The base breeding population of the black wattle improvement programme of the Institute of Commercial Forestry Research (ICFR) is relatively small when compared to other forestry breeding programmes in South Africa,
- Many of the groups in the breeding population are expected to be related,
- A knowledge of the degree of relatedness, while seedlings are still in the nursery, would be of benefit to the breeding programme by allowing early selection of genotypes, and
- The need to be able to detect and measure any relatedness in new selections.

Seven groups of individuals were selected from the black wattle breeding programme of the ICFR for this investigation. Three groups consisted of full-sibs, two of half-sibs, one group of unrelated individuals and one group of unrelated individuals that originated from the same provenance in Australia (Table 2.1). Five individuals within each group were studied, except for group 400, the Australian provenance, where only four individuals were available.

Table 2.1 Black wattle group information.

Group identity	Group composition
47	The individuals in this group were grown from seeds collected from a tree (number 47) in the Piet Retief area, in southeastern Mpumalanga. Group 47 occurred in block A23 of the Derby plantation. The individuals in this group were all half-sibs as the paternal parent was unknown.
50	The individuals in this group were grown from seeds collected from a tree (number 50) in the Piet Retief area, in southeastern Mpumalanga. Group 50 occurred in block A23 of the Derby plantation. The individuals in this group were all half-sibs as the paternal parent was unknown.
85	The individuals of this group were expected to be unrelated, as seed were collected off the ground of a plantation in the Kumbula area, near Dundee. These individuals were regarded to have neither a maternal nor a paternal parent in common.
88	The individuals in this group were full-sibs that arose from a controlled cross between tree 117 (mother / seed parent) and tree 272 (father / pollen parent).
89	The individuals of this group were full-sibs that arose from a controlled cross between tree 75 (mother / seed parent) and tree 136 (father / pollen parent).
91	The individuals of this group were full-sibs that arose from a controlled cross between tree 18 (mother / seed parent) and tree 84 (father / pollen parent).
400	This group consisted of four trees that were obtained from a wattle plantation outside of Howick. These trees originated from the Lake George provenance in Australia. This provenance is known for its frost tolerance. The trees sampled are currently part of the ICFR breeding group, designed to add a degree of frost tolerance to the commercial seed distributed to growers. These plants are presumed to be unrelated and common characteristics should be regarded as being provenance effects rather than due to genetic relatedness.

2.3 METHODS

2.3.1 Leaf material

Fresh leaf material was used as the source of DNA and obtained from the ICFR nursery, where the seeds of all the different groups were grown. Approximately six-eight of the youngest, greenest leaves, at the tip of a branch were picked at more or less 8:00 in the morning, an hour before DNA extraction commenced. The leaves were stored in clear plastic bags, labeled and kept on ice to prevent them from withering, until they were required for extraction.

2.3.2 DNA extraction

DNA was extracted from freshly collected leaf material using the DNA isolation-CTAB (N-cetyl-N, N, N-Trimethyl ammonium bromide) method of Doyle and Doyle (1990), with minor modifications.

The following protocol was standardized for 0.6–1 g leaf material for a 30 ml chloroform resistant plastic centrifuge tube.

Standard CTAB DNA extraction procedure:

- Preheat CTAB isolation buffer [2 % CTAB, 1.4 M NaCl, 20 mM EDTA, 100 mM Tris-HCl (pH 8.0), 1 % PVP, and 0.2 % 2-mercaptoethanol (added just before use)] at 65 °C for approximately 15 minutes.
- Wipe the leaf material with 70 % ethanol and grind fresh leaf tissue to a powder with liquid nitrogen.
- Add 3 ml 2 % CTAB isolation buffer to the grindate and pour into 30 ml chloroform resistant plastic centrifuge tubes. Rinse the mortar with 2 ml additional 2 % CTAB buffer and add to tube. Place tubes on ice.
- Incubate samples at 65 °C for one hour with occasional gentle swirling, every 15 minutes.

- Cool samples for 15 minutes at room temperature, and then add an equal volume (5 ml) of chloroform: isoamyl alcohol (24:1), mix well and incubate for 15 minutes at room temperature.
- Centrifuge for 10 minutes at 8 000 rpm in a refrigerated super speed centrifuge (Eppendorf model 5415C centrifuge, 13 000 rpm).
- Add top layer to a chloroform resistant centrifuge tube and add 500 µl of CTAB wash buffer (10 % CTAB and 0.7 M NaCl).
- Incubate samples at 65 °C for 30 minutes with occasional gentle swirling, every 15 minutes.
- Cool samples for 15 minutes at room temperature, add equal volume (5 ml) of chloroform: isoamyl alcohol (24:1), mix well and incubate for five minutes at room temperature.
- Centrifuge for 10 minutes at 8 000 rpm in a refrigerated super speed centrifuge (Eppendorf model 5415C centrifuge, 13 000 rpm).
- Add top layer to a chloroform resistant centrifuge tube and add an equal volume (5 ml) of ice-cold isopropanol. Mix gently and precipitate the nuclear acids for 60 minutes at a temperature of –20 °C. At this point the precipitation process may be left to continue overnight at –20 °C.
- Spool DNA with glass hooks; use flamed glass hooks for sterility. Samples can also be centrifuged at 10 000 rpm for 10 minutes as in step 6 to precipitate the DNA.
- Pour off supernatant and remove isopropanol by inverting tube when centrifuged. Remove isopropanol with pipette when spooled.
- Add 400 µl of 1 x TE buffer (10 mM Tris (pH 7.5) and 1 mM EDTA) and incubate in a water bath at 50 °C for 15 minutes to dissolve the DNA pellet.
- Transfer the solution to an eppendorf tube, add 400 µl 1 M NaCl and incubate at room temperature for 30 minutes with occasional inversion of tube.
- Add 400 µl isopropanol, precipitate the DNA for one hour at –20 °C and collect DNA by centrifugation for 15 minutes in a bench centrifuge at 4 °C (Eppendorf model 5415C centrifuge, 13 000 rpm).
- Pour off supernatant, centrifuge for one minute and remove excess ethanol with a pipette.
- Add 500 µl of 1 x TE buffer and incubate in a water-bath at 50 °C for 15 minutes to dissolve the DNA pellet.
- Add 2.5 µl 10 mg/ml RNAse to tube and incubate for two hours to overnight at 37 °C.

- Add 500 µl isopropanol, precipitate the DNA for one hour at –20 °C and collect DNA by centrifugation for 15 minutes in a bench centrifugation at 4 °C.
- Decant all liquid and add 500 µl of 70 % ice cold ethanol.
- Centrifuge for 10 minutes in a bench centrifugation at 4 °C (Eppendorf model 5415C centrifuge, 13 000 rpm).
- Remove liquid from pellet with a pipette and air-dry for 15 minutes.
- Dissolve pellet in 300 µl 1 x TE buffer at 50 °C for 15 minutes.
- Solution can be stored at 4 °C for short-term use or at –20 °C for extended periods of time.

2.3.3 Assessment of DNA purity

To confirm the presence of DNA, two methods were utilised. Firstly, an agarose gel was run to verify that the DNA was of high molecular weight, and secondly a spectrophotometric analysis was carried out to determine the concentration and purity of the extracted DNA samples.

A. Agarose gel electrophoresis

The presence of DNA was confirmed using gel electrophoresis. It should be mentioned that both TBE and TAE were used to test the buffering capacity of each. An 0.8 % agarose gel containing 1 × TBE or TAE buffer, and ethidium bromide was cast in a 6.5 cm × 10 cm gel tray. DNA samples of a volume of 10 µl were mixed with 2 µl loading dye, and loaded into the wells. Molecular weight marker III (Roche Applied Science) was loaded together with the samples, to determine the size of the DNA sample.

B. Spectrophotometric analysis

The concentrations of the DNA extracts were determined with a Beckman DU® 640 Spectrophotometer. The spectrophotometer was first blanked using 1 ml of TE buffer. Thereafter, absorbance values (A) consisting of a dilution of 5 µl DNA sample in 495 µl TE buffer were recorded at 260 nm and 280 nm.

The purity of the DNA extracts were calculated by applying the following equation and compared with pure DNA with a value of 1.8 (Sambrook *et al.*, 1989):

$$\text{Purity} = A_{260\ nm} / A_{280\ nm}$$

The concentration of DNA extracts was determined using the following equation, using the constant $1A_{260\ nm}$ unit of double stranded DNA = 50 $ng.\mu l^{-1}$ (Sambrook *et al.*, 1989):

$$\text{DNA concentration} = A_{260\ nm} \times \text{dilution factor} \times 50\ ng.\mu l^{-1}$$

2.3.4 RAPD fingerprinting

RAPD fingerprints were generated for all 34 samples in the following steps. Firstly, template DNA stock solutions were prepared, and secondly suitable primers identified. Finally, the RAPD-PCR was optimized.

A. Preparation of template DNA stock

The original extracted DNA solutions were stored at –20 °C. Working stock solutions of template DNA were made up to 100 ng/µl for all the samples. These solutions were stored at 4 °C for a maximum of two months before being discarded, due to the reduced stability of DNA at low concentrations.

B. RAPD Primers

Fourteen 10-mer RAPD primer sequences were randomly selected from the Operon Technologies primer list A, obtained from the Operon Technologies website (www.bioresearchonline.com/storefronts/operon.html). The selected primers all had a GC content that ranged between 60 and 70 %, because primers with a high GC content may lead to secondary priming artifacts and 'noisy' sequences. The selected primers (Table 2.2) were synthesized by the Molecular and Cell Biology Synthetic DNA Laboratory in Cape Town. Working stock solutions of 100 ng/µl template DNA were prepared for each of the primers according to the formula $C_1V_1 = C_2V_2$.

Table 2.2 Primer information.

Primer number	Sequence	OD
A 01	5'- CAG GCC CTT C-3'	459.5
A 02	5'- TGC CGA GCT G-3'	507.6
A 04	5'- AAT CGG GCT G-3'	443.7
A 06	5' -GGT CCC TGA C-3'	473.4
A 10	5'- GTG ATC GCA G-3'	439.1
A 12	5'- TCG GCG ATA G-3'	479.1
A 13	5'- CAG CAC CCA C-3'	465.5
A 17	5'- GAC CGC TTG T-3'	501.6
A 20	5'- GTT GCC ATC C-3'	470.0
C 06	5'- GAA CGG ACT C-3'	401.1
C 05	5'- GAT GAC CGC C-3'	195.8
A 19	5'- CAA ACG TCG G-3'	144.6
A 09	5'- GGG TAA CGC C-3'	143.9
17	5'- AAA CGG GCG G-3'	171.5

OD=optical density of primer

C. Generation of RAPD fingerprints

RAPD fragments were generated by a Gene Amp® PCR System 2700 thermocycler. The PCR reaction mixture of Shrestha *et al.* (2002) was utilised and resulted in satisfactory fingerprints, however minor adjustments were made to the protocol in order to obtain more defined fingerprints. Firstly, different concentrations of primer, 2 μM and 4 μM, and different amounts of template DNA, 1-4 μl, were tested. The optimal combination of reagents that produced fingerprints that were clear with singular bands, were utilised for all subsequent amplification reactions (Table 2.3).

Table 2.3 Optimal reaction composition used to generate RAPD fingerprints.

Reagents	Concentration	Required concentration	Volume used (μl)
PCR reaction buffer	10 ×	1 ×	2.5
$MgCl_2$	25 mM	2.5 mM	2.5
dNTPs	10 mM	200 μM	0.5
Primer	100 μM	2 μM	1.0
Taq polymerase	5 U	1 U	0.2
Sterilized distilled water	–	–	16.3
DNA template	100 ng/μl	100 ng	2.0

A reaction volume of 25 μl was used which contained 2.5 μl of 10 × PCR reaction buffer without magnesium chloride, 0.5 μl of 200 μM of each deoxynucleotide triphosphate (dNTP), 2.5 μl of 2.5 mM magnesium chloride, 1 μl of 2 μM primer, 0.2 μl of 5 units *Taq*, and 2 μl of 10 ng DNA. A volume of 16.3 μl sterilized distilled water was finally added to bring the volume up to 25 μl. The RAPD fingerprints were amplified by applying the

following PCR cycling conditions; a cycle that consisted of an initial denaturating step at 94 °C for three minutes, followed by 41 cycles at 94 °C for one minute, 37 °C for two minutes and finally four minutes at 72 °C completed the programme.

2.3.5 Resolution of the amplified DNA

The amplified DNA fragments were electrophoretically separated on 2 % agarose gels. The gels were prepared using 0.5 × TBE or TAE buffer and 2.5 g of agarose, in a 15 cm × 20 cm gel tray. Gel loading buffer (4 μl) was added to each 25 μl product, of which the entire sample was loaded into the wells of the agarose gels. Molecular weight marker X (Roche Applied Science) was loaded in order to determine the sizes of the amplified fragments. Electrophoresis of the amplified fragments was carried out by applying 100 volts at 400 amperes in 0.5 × TBE or TAE buffer for approximately six hours, or until the loading dye reached about a centimetre from the end of the gel. The gels were then viewed under an ultra violet light.

2.3.6 Repeatability and reproducibility

The RAPD fingerprinting technique is widely employed to estimate genetic variability. For this reason the results need to be repeatable and reproducible. Each of the polymorphic primers was tested at least twice in order to verify the reproducibility of the results obtained. Only those fingerprints that produced reproducible, clear bands were used in the investigation and scored.

2.4 ANALYSIS OF RAPD FINGERPRINTS

The RAPD fingerprints generated in this investigation were analysed in terms of the following:

- Description of each individual's phenotype,
- Determination and composition of the extent of genetic variation within groups,
- Determination of the group differentiation, and
- Comparison of the genetic variation between groups and within groups.

Two software packages developed to analyse molecular genetic data and one to construct dendrograms were employed:

- POPGENE (Version 1.30) by Yeh *et al.*, (1999), a software package which is freely available from the internet (www.ualberta.ca/~fyeh/) was used to calculate genetic variation between and within groups using dominant markers.
- ARLEQUIN (Version 2.000) by Schneider *et al.*, (2000) is a software package which is widely used and freely available from the internet (http://lgb.unige.ch/arlequin/). The ARLEQUIN software package version 2.000 (Schneider *et al.*, 2000) was employed to generate Wright's *F*-statistic value and an AMOVA table, which indicates the genetic variation between and within group variation to the total variation.
- The Treeview programme was used to view the dendrograms generated from POPGENE.

It should be mentioned that ARLEQUIN and POPGENE are able to do most molecular genetic analyses. However, POPGENE was used for most of the analyses because of its user-friendly interface. ARLEQUIN was employed only for the assessment of group differentiation, as this routine of ARLEQUIN provided more information than that of POPGENE.

The following steps were undertaken to analyse the RAPD fingerprints:

STEP 1: Digital fingerprints (phenotypes) of each individual were constructed by converting RAPD gel fingerprints into digital fingerprints.

STEP 2: POPGENE was employed to calculate different measures of genetic variation within groups and the whole population; the polymorphic content, Shannon's Information Index and estimate the mean heterozygosity.

STEP 3: The polymorphic loci, estimated with POPGENE were employed to assess primer performance.

STEP 4: POPGENE was utilised to estimate the genetic differences between the groups by calculating Nei's Unbiased Measures of Genetic Identity (I) and Genetic distance (D) (Nei, 1972).

STEP 5: The genetic relationships were represented by constructing dendrograms based upon Nei's Genetic distance using UPGMA (Sneath and Sokal, 1973).

STEP 6: ARLEQUIN was used to determine differentiation between the groups by selecting AMOVA (Analysis of Molecular Variance).

STEP 7: The data were interpreted.

2.4.1 Construction of digital fingerprints

A digital fingerprint was constructed for each individual by listing the primer results of each fingerprint in a particular order. Each locus of each primer was assessed for a band being either present or absent. A value of '1' was awarded to a band that was present (presence-phenotype) and the value of '0' to one that was absent (null-phenotype) from the fingerprint (Figure 2.1).

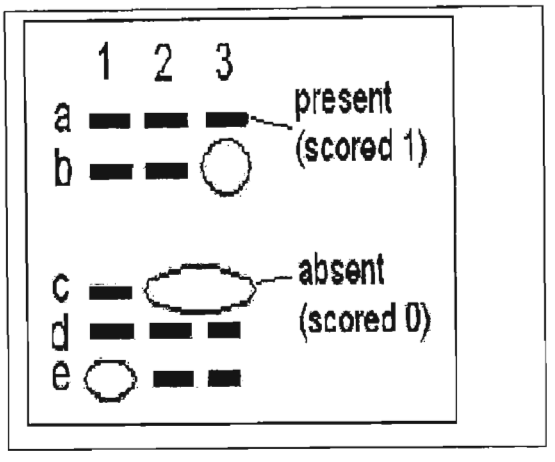

Figure 2.1 Method of scoring a RAPD fingerprint.

In order to score fingerprints accurately, a standard had to be set as to which bands should or should not be included in the digital fingerprint. RAPD fingerprints were evaluated in the following way:

- Primers that amplified too few loci, thus producing too few bands (less than six bands) were excluded,
- Primers that produced faint bands were excluded,
- Only fingerprints of primers that produced clear, discernable and reproducible bands were scored, and
- Loci were identified as being polymorphic when bands were present in some individuals, but absent in others, while monomorphic loci were those that produced bands in all the individuals.

The phenotypes of the different loci of the fingerprints were recorded in *Notepad*, the internal word processor of *Windows*, because both POPGENE and ARLEQUIN software packages were compatible with its format. The digital fingerprints were constructed by listing the fingerprint information generated by each primer sequentially in digital format for each individual (Figure 2.2).

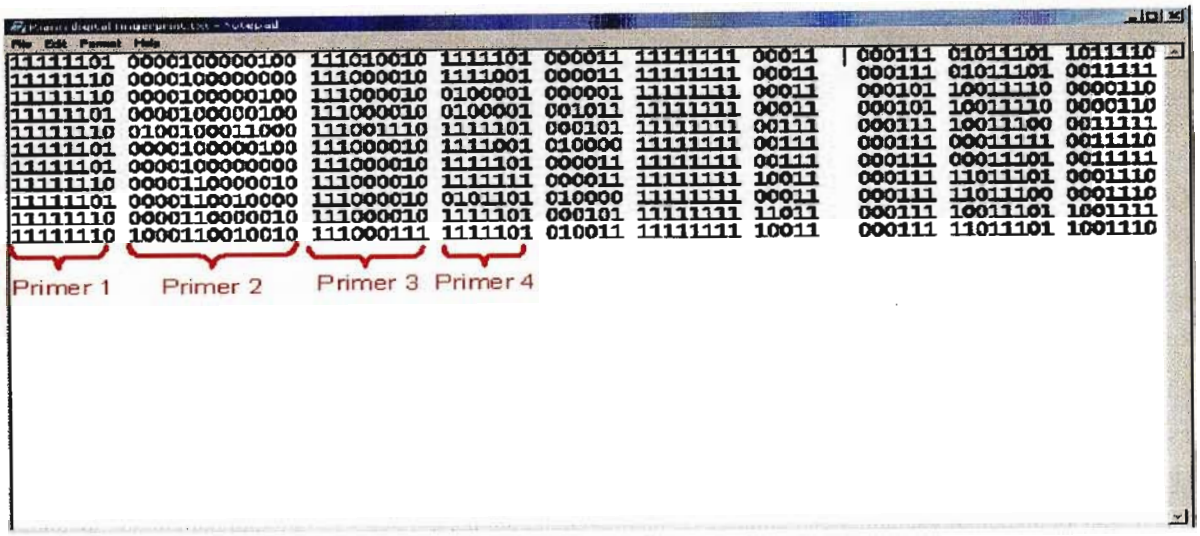

Figure 2.2 ***Notepad*** **display of partial digital fingerprints of eleven individuals.**

2.4.2 Determination of primer performance

Arbitrary primers do not all produce suitable results for a particular investigation. Some primers do not anneal sufficiently and produce faint and/or too few bands. Others may produce only monomorphic bands. Those primers that produce scorable fingerprints provide valuable comparative information about the primers. It is therefore imperative to determine the performance of each primer. The performances of the primers were determined by assessing their ability to generate clear and readable fingerprints and by the number of polymorphic and monomorphic loci that they amplified. The primer performance was also assessed in terms of the GC content, as the GC content often influences the number of loci that a primer amplifies.

The primers were evaluated according to the number of monomorphic and polymorphic loci that they amplified. Primers that produced a combination of polymorphic and monomorphic bands were all included in this investigation. The frequency of polymorphic or monomorphic loci for each primer was computed by dividing the total number of polymorphic or monomorphic loci by the total number of loci:

$$\text{Frequency of polymorphic loci per primer} = \frac{\text{Number of polymorphic loci}}{\text{Total number of loci}}$$

The GC content of a primer is thought to have some effect on the performance of a primer. This implies that the number of bands a primer produces could be affected by the number of 'G's and 'C's present in its sequence. This relationship was assessed by computing the number of bands amplified by each primer, as well as the percentage of GC of each primer.

2.4.3 Estimation of genetic variation within groups

The genetic variation was estimated for the different groups and the population as a whole using a number of measures. These measures included the estimated amount of heterozygosity, the gene diversity of Nei (1972) and Shannon's (Magurran, 1988) Information Index.

A. Estimated heterozygosity

The heterozygosity (2pq) of the group was estimated by determining the frequency p and q from the frequency of the null-phenotype. The *null-allelic* frequencies (q) for each locus were estimated from the frequency of the recessive phenotypes (null-phenotypes), q^2. Thereafter, the frequency of the *presence-allele* was estimated by applying the formula p=1−q. The estimated heterozygosity could then be calculated with the formula 2pq.

B. Nei's gene diversity

The quantification of genetic variation was estimated within each group. This was undertaken by calculating Nei's gene diversity (1972). This measure examines the mean allelic frequency within each of the different groups. Nei's gene diversity was computed by applying the following equation:

$$D = -\ln (I)$$

Where: $I = J_{XY} / \sqrt{J_X J_Y}$

$$J_{XY} = \sum x_i y_i$$

$$J_X = \sum x^2_i$$

$$J_y = \sum y^2_i$$

where x_i y_i are the frequencies of the i^{th} allele in groups X and Y.

The values obtained here range between 0 and 1, where a value close to 1 indicates great genetic variation.

C. Shannon's Information Index

Shannon's Information Index (Magurran, 1988), another measure of genetic variation, is similar to that of Nei's measure in that the higher the value, the greater the genetic variation within the group. Shannon's Information Index (Magurran, 1988) is scaled on a range of 0 to 1, with 1 representing the maximum possible diversity for the group. It is computed by applying the following equation:

$$H_0 = -\sum Pi \log_2 Pi$$

where: H = Shannon's Index,

Pi = proportional contribution by individuals.

2.4.4 Estimation of the genetic relatedness between and within groups

An estimation of the genetic relatedness within and between groups was obtained by calculating Nei's genetic identity and diversity values (Nei, 1972). Nei's genetic diversity (D) that estimates differences and Nei's genetic identity (I) that estimates similarity between groups were calculated. The relationship between identity and diversity is simply I+D=1.

Nei's (1972) genetic identity and diversity values were calculated using UPGMA (Unweighted pair group method with arithmetic mean) (Sneath and Sokal, 1973). These values were used to construct dendrograms for each group, between groups, and between all 34 individuals in the whole population.

2.4.5 Estimation of group differentiation

The population of black wattle used in this investigation consisted of a number of groups. In order to get a better understanding of the genetic differentiation between and within these groups, an Analysis of Molecular Variance (AMOVA) and Wright's *F*-statistics was determined by employing the software package ARLEQUIN version 2.000 (Schneider *et al.*, 2000).

2.5 UTILISATION OF COMPUTER SOFTWARE

There are a number of computer packages that exist that calculate genetic variation in groups. The two statistical packages used in this investigation were POPGENE version 1.31 (Yeh *et al.*, 1999) and ARLEQUIN version 2.000 (Schneider *et al.*, 2000)

2.5.1 POPGENE

POPGENE, a MICROSOFT® Window-based computer package, analyses genetic variation within groups using co-dominant and dominant markers, as well as quantitative data (Yeh *et al.*, 1999). POPGENE was selected because of its ability to perform certain required calculations, and its user-friendly interface. It performs a number of calculations, such as, allele frequency, gene diversity, genetic distance, *F*–statistics, Shannon's Information Index (Magurran, 1988), neutrality tests and it also provides genetic trees or dendrograms.

The POPGENE programme was employed to estimate genetic distances in the following way:

- An input file containing the digitized data was created in *Notepad* (Appendix B).
- The data of the input file was then imported by POPGENE to carry out the various calculations, by selecting the dominant marker option.

The input file generated in *Notepad* consisted of a header section, which contained specific information about the dataset that the programme required to run the analyses, namely, the type of data, the number of groups and the names of the loci. The header section is followed by the data. The data of the seven groups were given a letter of the alphabet (A to F for 47, 50, 85, 88, 89, 91 and 400 respectively). Figure 2.3 provides a representative input file of data used in this investigation.

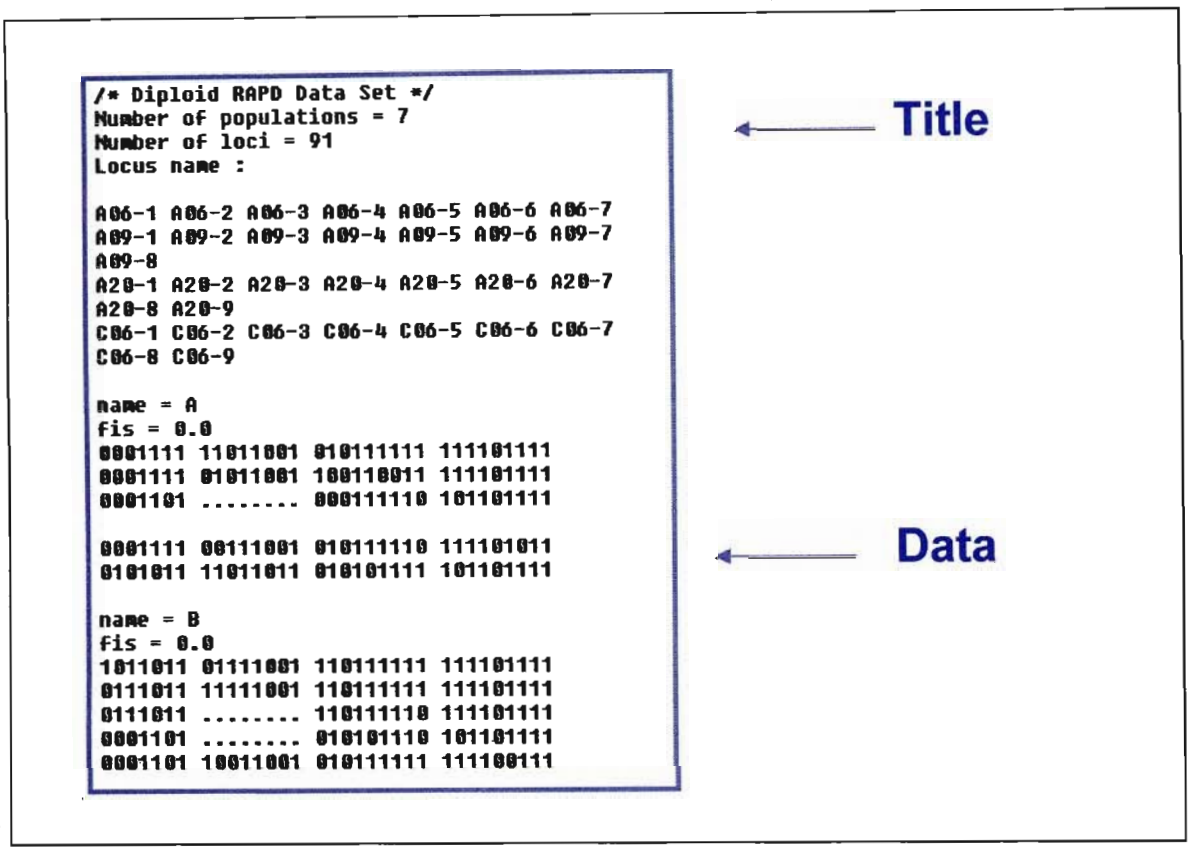

Figure 2.3 Example of a POPGENE input file containing digitized RAPD fingerprints.

Once the input file was created and saved, the analyses were conducted. POPGENE was implemented in the following manner:

- F_{IS} value of zero was assigned to each group to denote Hardy-Weinberg equilibrium,
- The files were loaded into POPGENE as 'dominant marker data',
- Diploid data analyses were generated for Shannon's Information Index (Magurran, 1988), polymorphic loci, Nei's genetic distance (1972) and identity, and
- Two output files were generated, one containing the data of the dendrograms and another containing all other genetic analyses.

2.5.2 ARLEQUIN

ARLEQUIN is a group genetics software package, which is able to handle large samples of molecular data (RFLPs, DNA sequences, microsatellites), while retaining the capacity of analyzing conventional genetic data (standard multi-locus data or allele frequency data). ARLEQUIN deals with a variety of population genetics methods, which can be implemented, either at the intra-population or at the inter-population level, and they can be conveniently selected and parameterized through a graphical interface (Schneider *et al.*, 2000).

The ARLEQUIN programme does not provide RAPD specific-functions to analyse RAPD data. However, ARLEQUIN's RFLP functions may be used to analyse RAPD data. The reason being, the programme's RFLP functions require a specific data format file containing the RFLP data. This format uses a '1' to denote a presence of a restriction site and a '0' to denote a lack thereof. Similarly, general RAPD data denotes the presence of a band using a '1' and '0' to denote an absence of a band. Hence, the RAPD data was manually hard coded into a RFLP data file using the '1' and '0' format. Therefore, the programme's RFLP functions easily processed the RAPD data file.

The ARLEQUIN software package version 2.000 (Schneider *et al.*, 2000) was employed to generate Wright's *F*-statistic value and an AMOVA table, which indicated the genetic variation between and within group variation to the total variation. AMOVA works with digital data by creating a distance matrix between samples in order to measure the genetic structure of the group from which the samples are drawn. In statistical terms, AMOVA is a testing procedure based on permutational analysis and involves few assumptions about the statistical properties of the data. AMOVA is like a hierarchical analysis of variance in that it separates and tests tiers of genetic diversity:

- Diversity among different populations,
- Diversity among the groups within a population, and
- Diversity among the individuals within a group.

ARLEQUIN is quite a demanding software package and is not very user-friendly, therefore most of the calculations were conducted with POPGENE except for the AMOVA, which was undertaken with ARLEQUIN. The ARLEQUIN analysis involved the following sequential steps:

- Create an input file according to the RFLP data type,
- Open the project file and save the input file as specified,
- Execute the analysis, choosing the preferred statistical methods, and
- Interpret the results.

A. Arlequin input file

The input file generated in *Notepad* consisted of a profile, the data, details of the samples and the structure of the groups as shown in Figure 2.4. The profile section of the input file was made up of the title, number of samples or groups, and data type. ARLEQUIN also makes specifications for locus separators and missing data. The data section consisted of the digital data of all 34 samples. The samples section was made up of the different number of groups containing information about the sample name and sample size. The structure section consisted of the number of groups involved in the study. The input file was then saved.

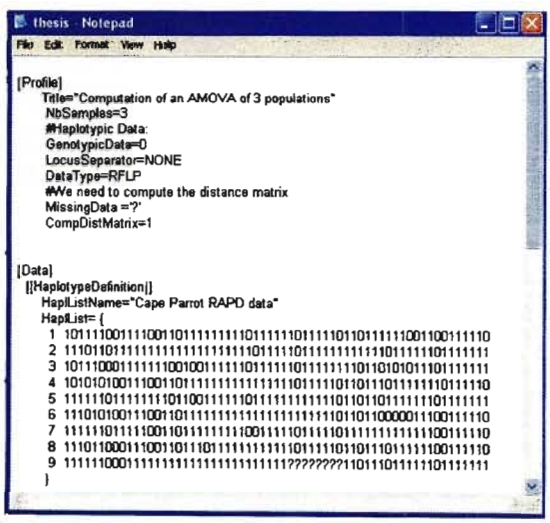

Figure 2.4 (a) **Example of an ARLEQUIN input file.**

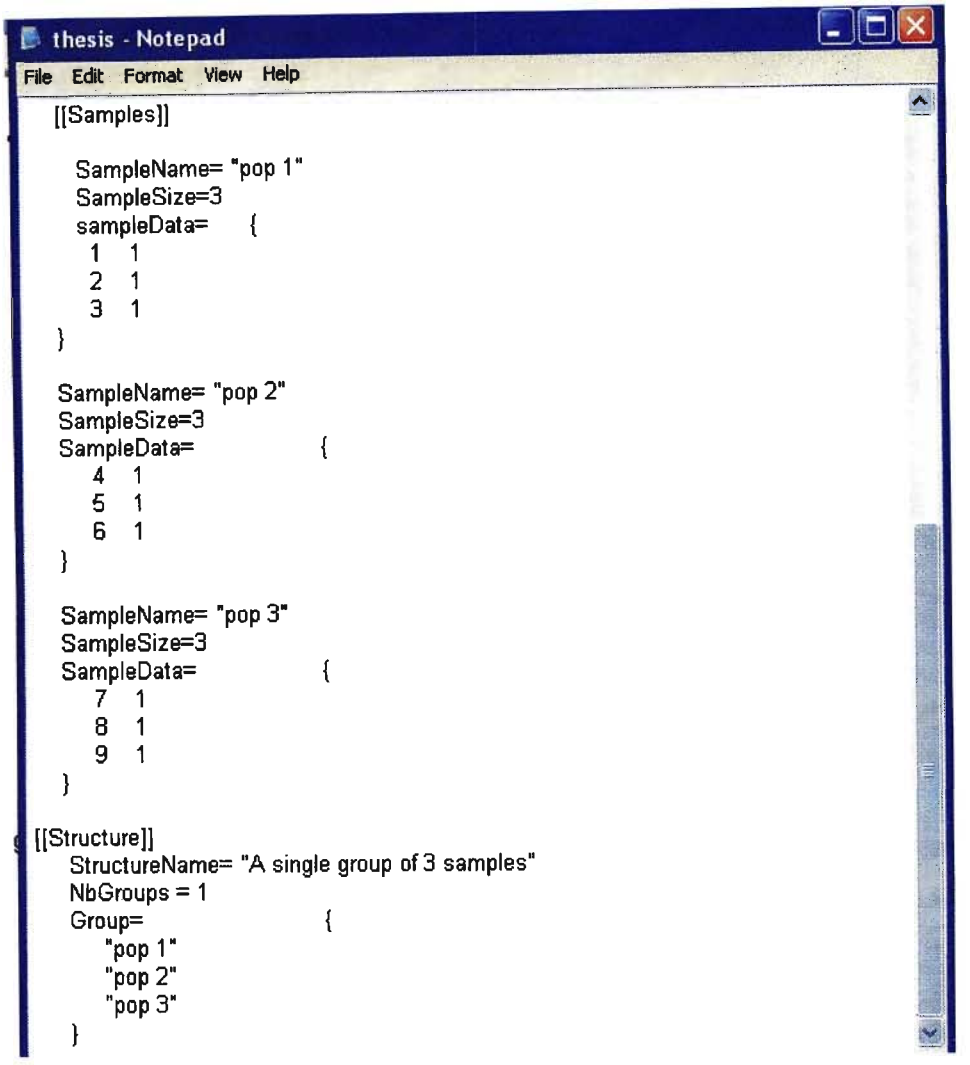

Figure 2.4 (b) **(Cont.) Example of an ARLEQUIN input file.**

B. Execution of analysis

Once the input file was created, the following steps were followed to conduct the analysis:

- The ARLEQUIN programme was opened and the project wizard option was selected (Figure 2.5), to create an ARLEQUIN project file,
- The browse button was clicked. A 'save as' window popped up which gave the option to save the file,

- The file was saved as a 'txt.arp' file, in the folder where the original input data file was contained. It is important to create an 'arp' file as ARLEQUIN will only function if the input data is in an 'arp' format,
- The 'open outline as project' option was chosen,

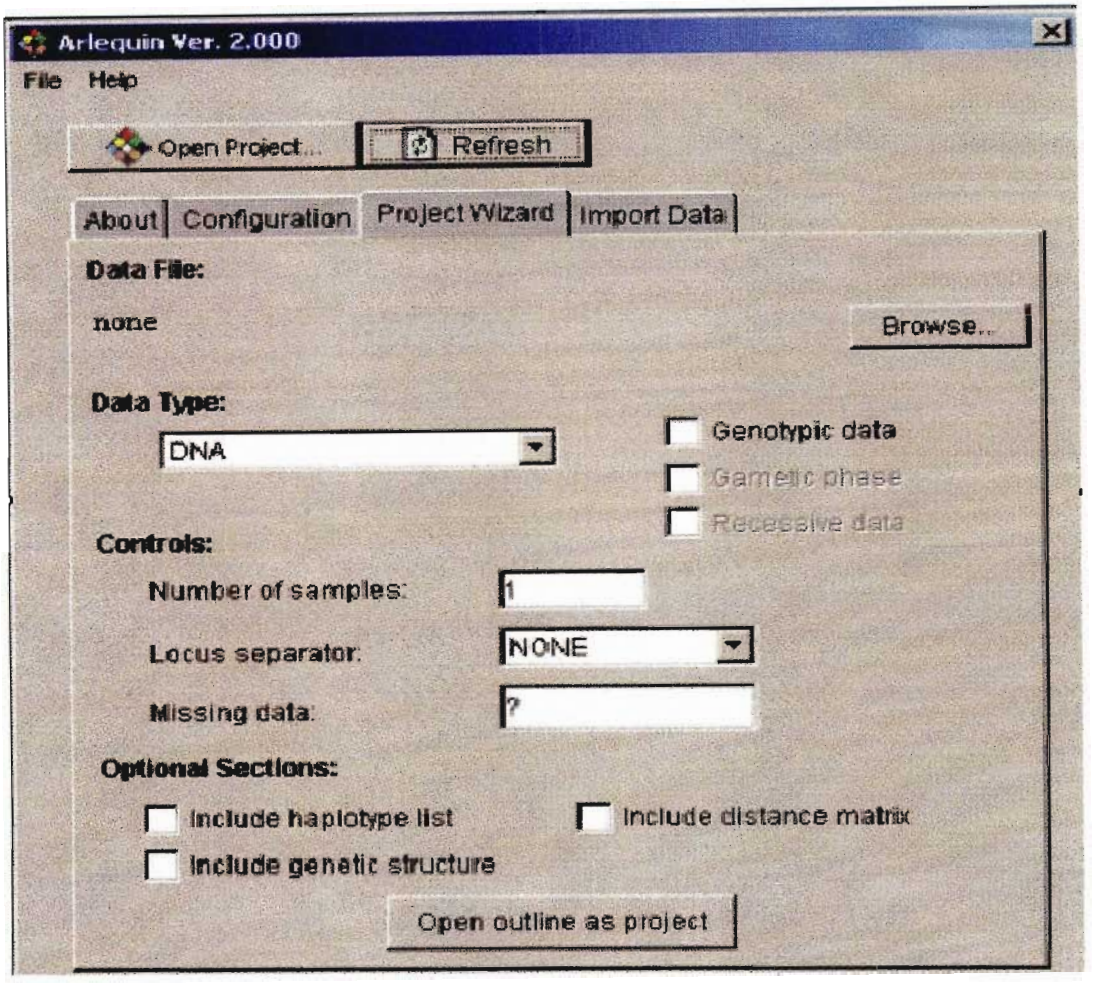

Figure 2.5 ARLEQUIN programme showing the project wizard window.

- A second window, consisting of the calculation settings appeared (Figure 2.6),
- This window was minimized, and the folder containing both the original input data and the 'txt.arp' file was opened,
- The contents of the 'txt.arp' file were deleted (the file contains examples data generated by the programme) and the original input data was copied and pasted into the 'txt.arp' file,

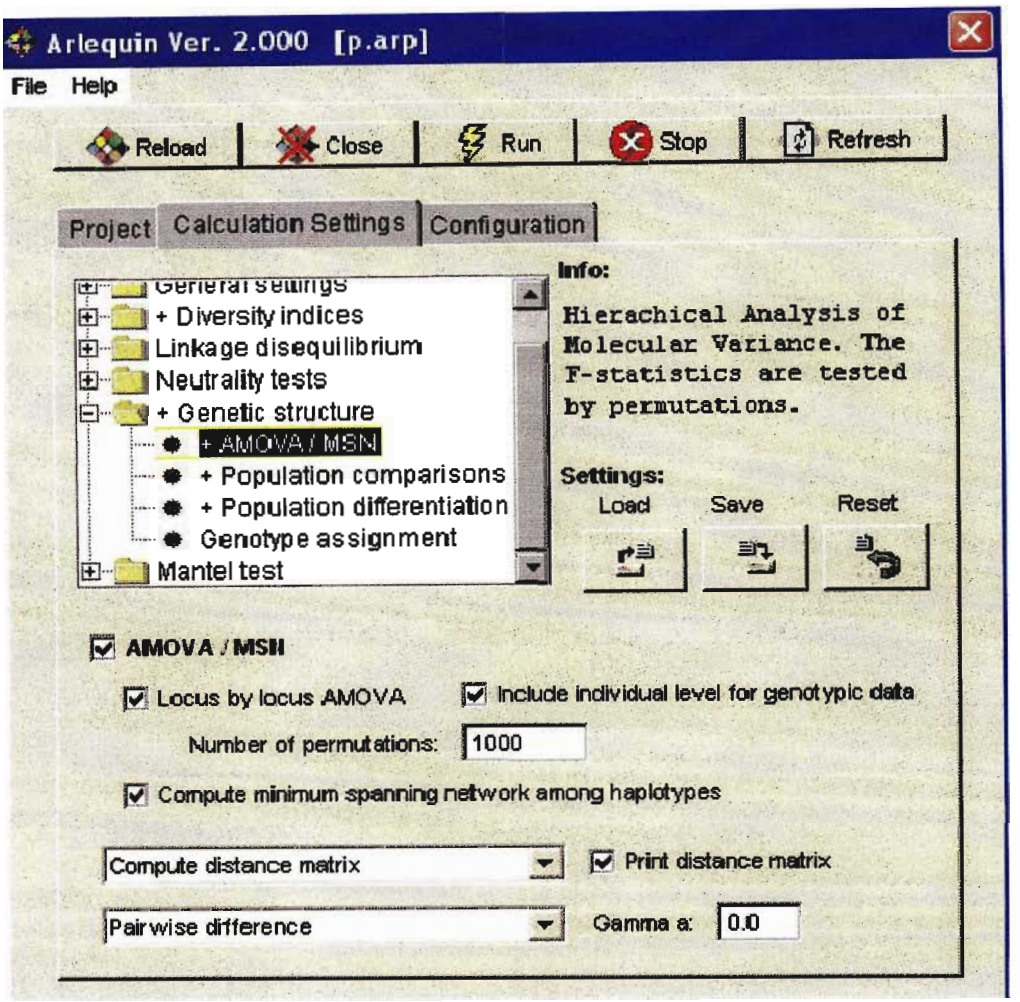

Figure 2.6 ARLEQUIN programme showing the calculation settings window.

- This was saved as a new 'txt.arp' file
- The calculation settings window was restored,
- The genetic structure option was chosen which allowed the AMOVA settings to appear,
- The desired settings were clicked and the analysis was executed,
- The results were automatically saved in the same folder as the input file, and
- The results were interpreted.

2.5.3 Treeview

The TREEVIEW programme (Page, 1996) was employed to view the dendrograms generated from POPGENE. POPGENE generates a file named *.dgm that contains the dendrogram information. When POGENE is opened and this file loaded a graphical display of the dendrogram is obtained.

CHAPTER THREE

RESULTS

3.1 INTRODUCTION

In this investigation RAPD analyses were undertaken to determine the genetic variation within the black wattle population of South Africa. Black wattle trees were collected from different regions covering the black wattle plantation area. The trees belonged to seven groups of which two were half-sib groups, three full-sib groups, and two groups, which consisted of, unrelated individuals from different areas.

The CTAB isolation technique of Doyle and Doyle (1990) was employed to extract DNA from these samples, but considering that this investigation was undertaken for the first time, the PCR procedure had to be optimized, and various statistical packages explored. A total of 14 primers were initially selected to test their polymorphic content in a representative sample containing individuals of varying levels of genetic relatedness.

The results of this investigation are presented in the following order:

- Results of DNA isolation,
- Results of RAPD fingerprinting, and
- Results of statistical analysis of genetic variation.

Raw data and computer printouts of statistical analyses have been taken up in Appendix B and C respectively.

3.2 DNA ISOLATION

The CTAB method (Doyle and Doyle, 1990), a DNA isolation technique commonly used in plants, was employed to extract DNA from the samples. This protocol separated secondary metabolites, such as polysaccharides, polyphenolics and tannins, released during cell lysis, from the DNA. These brown coloured and sticky substances, which abound in many woody plants, irreversibly adhere to DNA, often inhibiting digestion and PCR amplification (Plessis *et al.,* 1999). In some instances the leaves were older and contained tough cell walls and large amounts of tannin. Therefore it was difficult to extract DNA and the amount of DNA extracted was insufficient. Hence additional extractions were required to increase the amount of DNA until sufficient amounts were obtained. The CTAB DNA isolation procedure (Doyle and Doyle, 1990) successfully yielded large amounts of good quality DNA to conduct between 50-100 RAPD reactions.

3.2.1 Confirmation of DNA isolation

The presence of DNA was confirmed by gel electrophoresis. A gel of 0.8 % agarose was left to run for 30 minutes at 80 volts and thereafter viewed under ultraviolet light. An example of an agarose gel used to detect the presence of DNA is presented in Figure 3.1. The clear, bright DNA bands are evidence of high molecular weight DNA with little or no shearing of the DNA.

It should also be mentioned that it was found that in the trial runs, TBE performed better than TAE. TBE had a better buffering capacity of smaller DNA fragments such as those produced in RAPD reactions, thus the gel could be run for a longer period of time, in this case six hours. It was also found that a 1 % agarose gel proved to create a clearer, higher resolution of the bands when viewed under the ultra violet light than that of a 2 % gel.

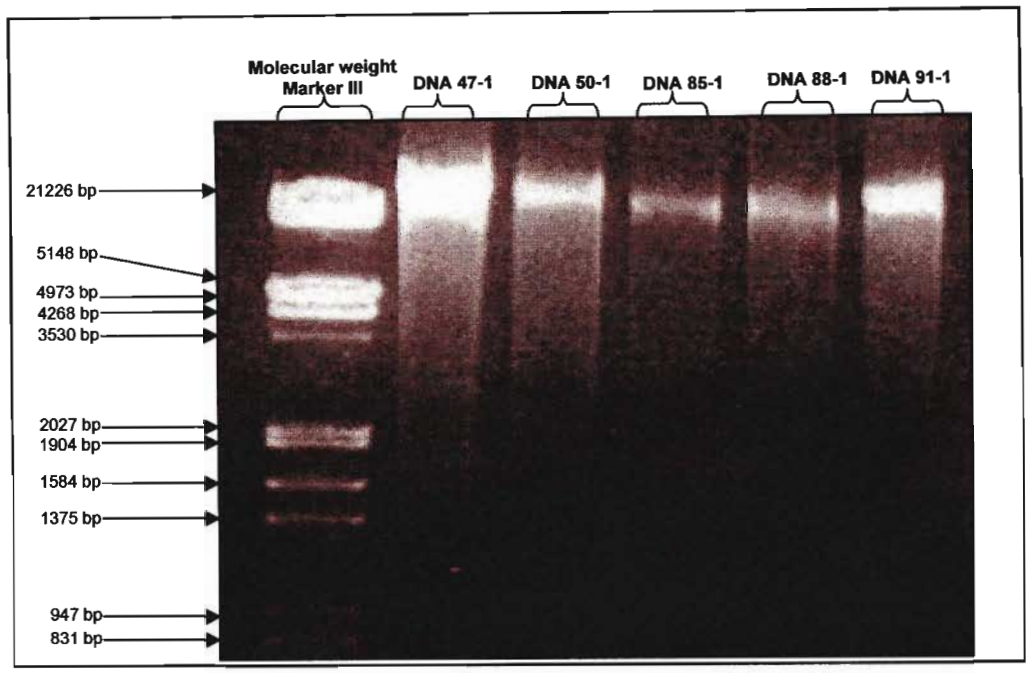

Figure 3.1 **Detection of high molecular weight DNA on agarose.**

Lane 1: Molecular weight marker III.

Lane 2- 6: Lanes 2-6 contains good quality DNA extracted from the respective individuals.

3.2.2 Determination of DNA concentrations

Once the presence of the DNA was confirmed by the running of an agarose gel, a spectrophotometric analysis was undertaken to determine the DNA concentration of all the DNA samples. The DNA isolated from all samples was of an acceptable purity to conduct RAPD analyses. The DNA purities ranged from 0.1 to 1.8, with a mean of 1.46. The DNA concentrations, on the other hand, ranged from 6.5 $\mu g.\mu l^{-1}$ to 687 $\mu g.\mu l^{-1}$, with a mean concentration of 359.33 $\mu g.\mu l^{-1}$. Table 3.1 provides a list of DNA purities and concentrations.

Table 3.1 DNA purity ratios and concentrations.

Plant number	Absorption 260 nm	Absorption 280 nm	Purity	Concentration µg/µl
47-1	0.123	0.081	1.500	0.618
47-2	0.001	0.011	0.100	0.006
47-3	0.115	0.071	1.600	0.578
47-4	0.120	0.073	1.600	0.601
47-5	0.108	0.066	1.600	0.540
50-1	0.137	0.078	1.800	0.687
50-2	0.106	0.062	1.700	0.530
50-3	0.115	0.067	1.700	0.579
50-4	0.103	0.064	1.600	0.517
50-5	0.100	0.055	1.800	0.500
85-1	0.096	0.055	1.700	0.484
85-2	0.094	0.054	1.700	0.471
85-3	0.101	0.059	1.700	0.508
85-4	0.098	0.056	1.700	0.490
85-5	0.095	0.057	1.700	0.479
88-1	0.109	0.063	1.700	0.547
88-2	0.088	0.051	1.700	0.441
88-3	0.103	0.060	1.700	0.517
88-4	0.120	0.068	1.800	0.604
88-5	0.013	0.012	1.000	0.065
89-1	0.012	0.009	1.300	0.062
89-2	0.010	0.007	1.400	0.054
89-3	0.001	0.003	0.400	0.065
89-4	0.021	0.013	1.600	0.106
89-5	0.005	0.005	1.100	0.027
91-4	0.014	0.016	0.800	0.072
91-5	0.039	0.043	1.000	0.196
91-6	0.014	0.016	1.000	0.071
91-7	0.002	0.001	1.400	0.013
91-8	0.033	0.023	1.400	0.167
401	0.088	0.051	1.700	0.441
402	0.094	0.054	1.700	0.471
403	0.120	0.068	1.800	0.604
404	0.021	0.013	1.600	0.106

3.3 GENERATION OF RAPD FINGERPRINTS

RAPD fingerprints were generated for the 34 selected black wattle plants using 14 different primers, to determine genetic variation. Ten of the primers were randomly selected from the Operon Technologies primer list A, obtained from the Operon Technologies website, while the other four were self-constructed.

3.3.1 Optimization of RAPD protocol

A RAPD-PCR protocol was implemented to test its suitability to generate RAPD fingerprints with black wattle DNA. After varying the quantities of the components of the RAPD-PCR reaction mixture, an optimized composition was obtained (Table 3.2). The components that required adjustment included the amount of template DNA and primer concentration. The amount of template DNA tested ranged from 1 µl to 4 µl, with 2 µl producing the best results. For the optimization of the concentration of the primer, volumes were varied from 1µl to 4 µl, where a concentration of 2 µl primer produced RAPD fingerprints with bright and distinguishable bands.

Table 3.2 RAPD reaction mixture composition.

Reagents	Concentration	Required concentration	Volume used (µl)
PCR reaction buffer	10 x	1 x	2.5
$MgCl_2$	25 mM	25 mM	2.5
dNTPs	10 mM	200 µM	0.5
Primer	100 µM	2 µM	1.0
Taq polymerase	5 U	1 U	0.2
Sterilized distilled water	-	-	16.3
DNA template	100 ng/µl	100 ng	2.00

3.3.2 Evaluation of RAPD fingerprints

Scoring of the bands of the RAPD fingerprints is a time consuming process and could introduce inaccuracies if the scoring process is not well described and standardized. Therefore, to ensure that scoring was consistent throughout the process, a standard was designed that included specifications on how clear or bright a band should to be, in order for it to be included in the analysis. Once this was decided upon, this standard was applied throughout the scoring process to limit the amount of errors. In this investigation only bands that were clear, singular and that were repeatable, were included in the analyses. Figure 3.2 shows RAPD fingerprints with clear and singular bands, as well as fingerprints of which the bands were too faint to be included in the subsequent analyses.

It was found that five of the primers (primers A01, A04, A12, A17 and A19) produced faint and unreliable fingerprints with some of the samples, and were therefore excluded from subsequent analyses. The sequences of these primers were probably too different from those of black wattle DNA, thus annealing insufficiently to produce sufficient amplification product to produce scorable bands. The total number of scorable loci utilised in all subsequent analyses produced by the remainder of the nine primers was 91, and ranged in size from 3 00 bp to 3 050 bp.

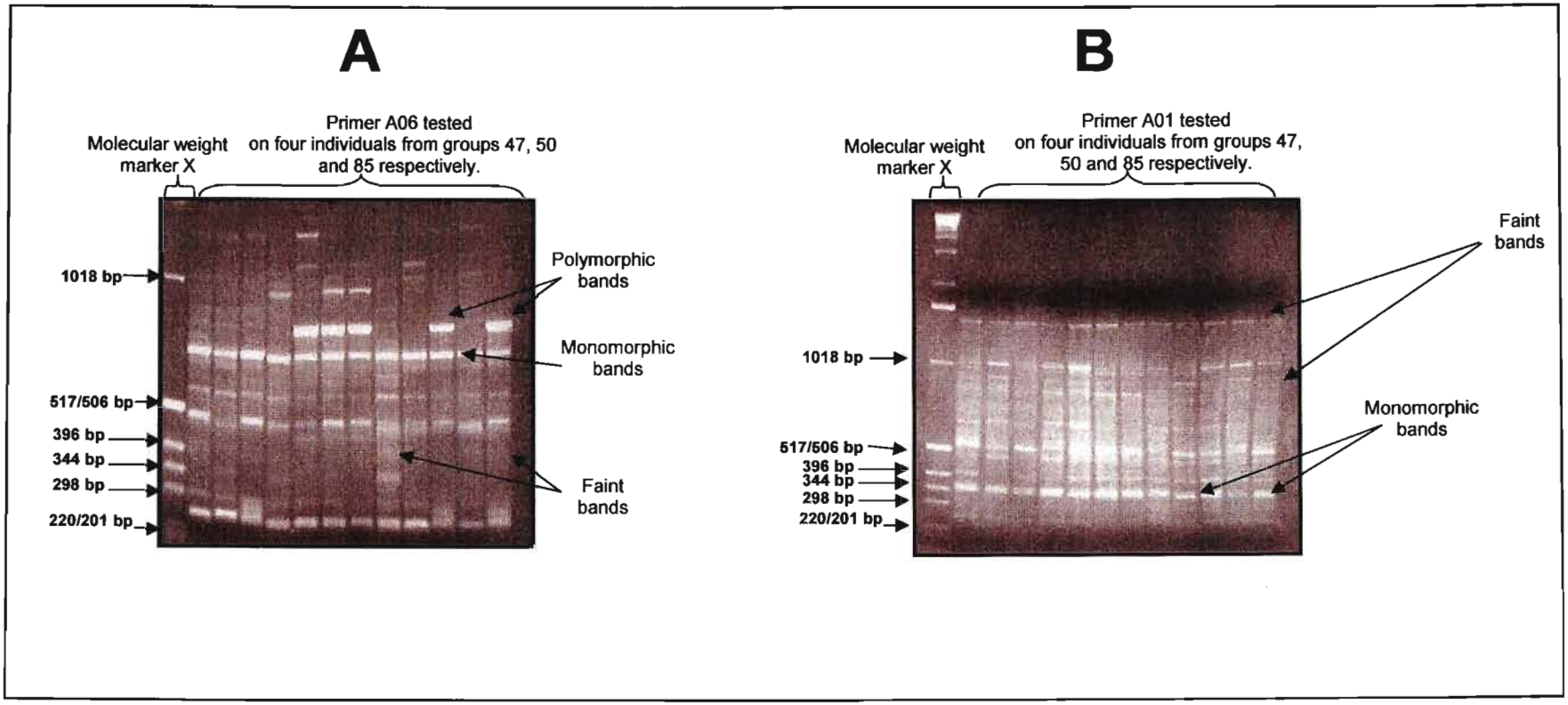

Figure 3.2 Gels showing suitable and unsuitable primers tested.

A- Gel showing monomorphic and polymorphic bands.

B- Gel showing bands of individuals too faint to score.

3.3.3 Conversion of molecular RAPD fingerprints into digital fingerprints

The RAPD molecular fingerprints that were generated for all individuals by the different primers were used to compile a digital fingerprint that could be used in all subsequent analyses. The RAPD fingerprints of the nine primers were scored by employing a binary scoring procedure. The presence of a band of a particular locus was scored as '1' and the absence of a band as '0'. The binary values of each locus of each primer were entered into *Notepad*, the Windows internal word processor, in a sequential manner (Table 3.3). In a number of individuals, a few primers were not able to generate amplification products (marked by asterisks) after at least three attempts. It was interesting to note that primers A02, A09, C05 and 17 did not amplify any loci of the members of group 400 after numerous attempts. Furthermore, primer C06 produced a band of 272 bp in size that occurred in all individuals (highlighted in pink). Three other loci produced bands that occurred in all individuals accept the members of group 400; locus four of primer A02 (highlighted in brown) produced a band of size 893 bp; locus eight of primer 17 (highlighted in blue) produced a band of size 317 bp and locus 10 of primer 17 (highlighted in red) produced a band of size 436 bp. All the other loci produced varied results for which no specific pattern could be identified.

Table 3.3 Digital fingerprints of all individuals.

INDIVIDUAL*	PRIMER A02	PRIMER A06	PRIMER A09	PRIMER A10	PRIMER A13	PRIMER A20	PRIMER C06	PRIMER C05	PRIMER 17
47-1	0101101111100	0001111	11011001	000100010010	01000100110	010111111	111101111	111010101000	000000010 1
47-2	1101101110110	0001111	01011001	000000000010	00000101110	100110011	111101111	011111111101	1111101111
47-3	0101001111011	0001101	.....**	101011010101	11000100110	000111110	101101111	..... **	..... **
47-4	0001111111011	0001111	00111001	000100110010	01010101110	010111110	111101011	111100011000	000001010 1
47-5	0001011111011	0101011	11011011	100011110011	01000101110	010101111	101101111	011100101101	1111101101
50-1	1111111111011	1011011	01111001	100011010111	11001101110	110111111	111101111	111101000001	1111010101
50-2	0111011111011	0111011	11111001	100001011011	11000110110	110111111	111101111	111101101100	1111101101
50-3	0111011101011	0111011	.....**	100001011010	11000110110	110111110	111101111	.....**	..... **
50-4	0011011111111	0001101	..... **	001001100001	01000101110	010101110	101101111	..... **	..... **
50-5	0011011111111	0001101	10011001	001001010111	11100110111	010111111	111100111	001010000000	000001010 1
85-1	0011011111100	0011010	.....**	111101001111	11000001011	010111110	111111111	.....**	..... **
85-2	0011011111100	0001111	01011000	001001001111	11000110111	010101110	101101011	..... **	1110100101
85-3	0011011111100	0011111	.....**	001000000010	11000101111	010111111	001101111	.....**	..... **
85-4	0001001111100	0001111	11011100	001000100010	11000000111	001110110	001110111	010110101100	1111101101
85-5	0001001111100	0001111	10011100	011101100011	11000110110	010101100	101111111	010110101101	0100100101
88-1	0011110100011	1001000	01011001	101010000100	11100010110	001010000	001100111	011111110001	0001011101
88-2	0011010011111	0001001	01001101	101011100000	11100110110	010111111	101100011	000011001100	0110100101
88-3	0011011100000	0001000	01011000	000000000010	00000000110	000110110	011100111	000000110000	..... **
88-4	0111001011111	0001111	00011000	100011110011	01001101111	110111111	111101111	111101000010	0010011111
88-5	0011001111111	0001001	.....**	100001100110	01001110110	010110110	011100011	.....**	..... **

..... continued

89-1	0111001011111	0001111	00011101	000010011011	01000100111	010111111	111100111	001011001100	111010111
89-2	0011101111111	0001111	01011001	001100000010	01000000110	001110000	011101101	101001000000	000000011
89-3	1111001110011	0001011	00011001	100001010011	01001111111	110111111	111101111	001001010110	111010111
89-4	0101111100000	0001010	00011000	000000000010	00010000110	100110001	011110111	111101000010	001000111
89-5	0101001111100	0001011	00011100	100011011101	01000100110	000111111	101101111	001011001110	000010111
91-4	0011001100000	0001010	01010011	000100100110	00000000110	010111100	101101011	..... **	..... **
91-5	0101001111111	0101011	01011101	101011000111	11000100110	000101100	101111111	001011100111	010010101
91-6	0011001100000	0001110	01010110	001100100110	00000000110	010111100	101111111	001111100110	110010111
91-7	0001001111011	0101001	01011101	100011110010	01000101110	000111100	111101111	001001000101	010010101
91-8	1011001111101	0101111	01010100	001100110010	01000000110	010111100	101111111	001111000110	010010111
401	.....**	0011010	.....**	000000111010	11000001110	000110110	000010011	.....**	.....**
402	..... **	0010010	.....**	000000000010	00000001111	000110100	101101011	.....**	.....**
403	..... **	0011011	.....**	000000000010	00000001110	000110000	101101011	.....**	.....**
404	..... **	0001011	.....**	000000101010	00000001101	000110100	101101011	.....**	.....**

* Numbers 47-, 50-, 85-, 88-, 89-, 91- and 400s refer to the different groups.

** No results.

3.4 ANALYSIS OF PRIMER PERFORMANCE

The performance of any primer is determined not only by means of its ability to generate clear and readable fingerprints, but also by the number of loci amplified and the number of polymorphic and monomorphic bands that it produces. Another factor that may also influence the number of loci that a primer amplifies is the GC content of a primer.

The monomorphic bands represent species specific alleles of a particular locus, where the primer annealing sequence is the same for all the individuals assessed; every individual therefore produces a band for the particular locus. On the other hand, polymorphic loci produce bands in some individuals, while others do not amplify the locus, indicating mutations in either the primer annealing sequences or intervening sequences, resulting in an amplification product that has either an increased or a decreased size and thus not scored for the particular locus. Figure 3.3 provides a representation of RAPD fingerprints that contain examples of loci that were monomorphic and loci that were polymorphic.

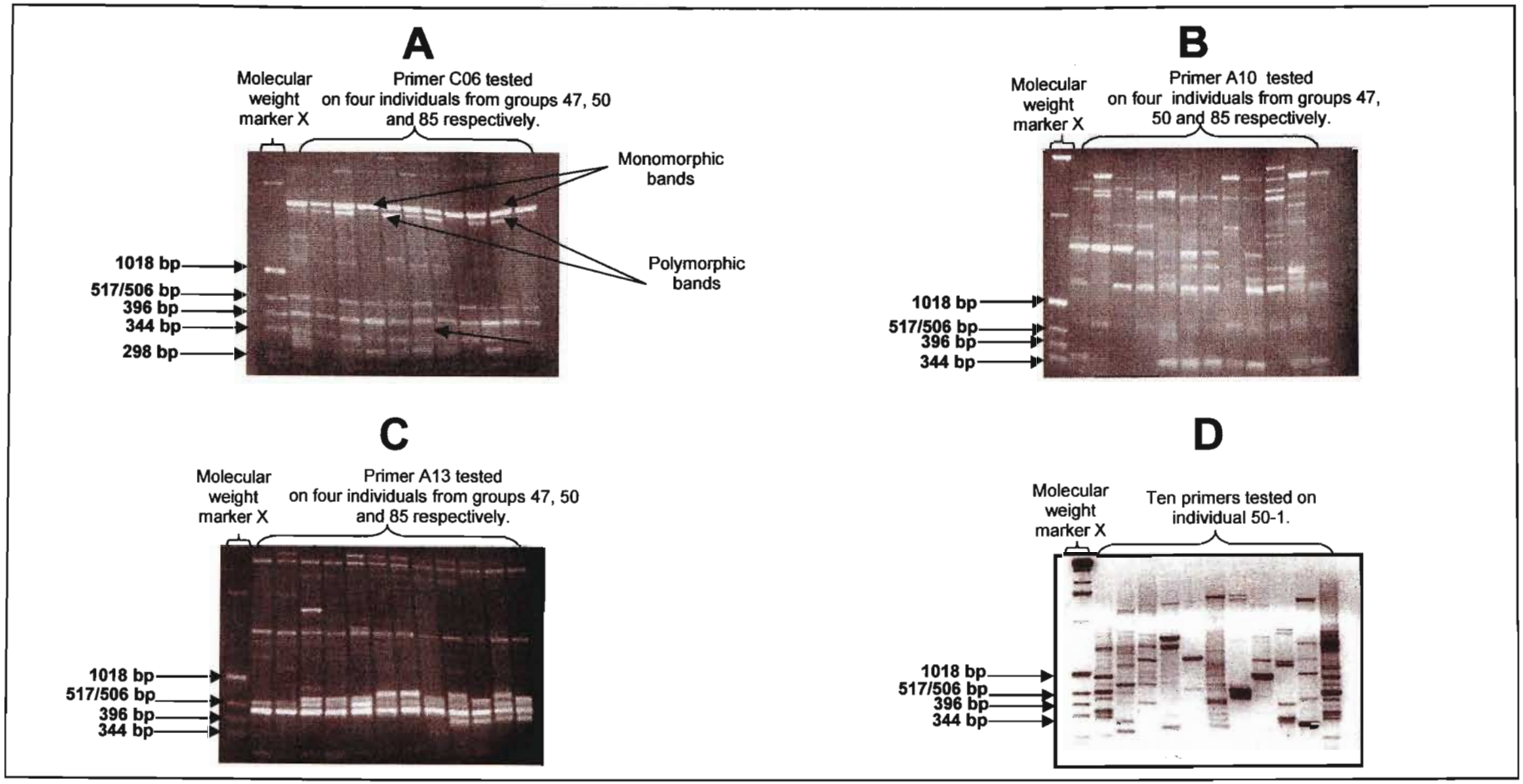

Figure 3.3 Examples of RAPD fingerprint profiles displaying monomorphic and polymorphic bands.

A- Gel showing DNA samples that were tested on primer C06.

B- Gel showing DNA samples that were tested on primer A10.

C- Gel showing DNA samples that were tested on primer A13.

D- Gel showing fingerprints of individual 50-1 using ten primers. This profile indicates how variable the different primers are on a single DNA sample.

The selected primers were evaluated in terms of the number of loci, as well as the number of monomorphic and polymorphic loci they amplified. A summary of the individual primer performance is presented in table 3.4. The number of loci that the primers amplified ranged, from seven (primer A06) to 13 (primer A02). When considering all the loci together, a total of 95.6 % of the loci were polymorphic. However, the loci of six of the primers were all polymorphic (primers A06, A09, A10, A20, A13 and C05) while the remaining three primers amplified both monomorphic as well polymorphic loci.

Table 3.4 Number of polymorphic and monomorphic loci generated by the selected primers.

Selected primer	Primer sequence	Number of scorable loci	Number of polymorphic loci	Percentage polymorphic loci (%)	Number of monomorphic loci	Percentage monomorphic loci (%)
A01	5'- CAG GCC CTT C-3'	*	-	-	-	-
A02	5'- TGC CGA GCT G-3'	13	12	92	1	8
A04	5'- AAT CGG GCT G-3'	*	-	-	-	-
A06	5' -GGT CCC TGA C-3'	7	7	100	0	0
A10	5'- GTG ATC GCA G-3'	12	12	100	0	0
A12	5'- TCG GCG ATA G-3'	*	-	-	-	-
A13	5'- CAG CAC CCA C-3'	11	11	100	0	0
A17	5'- GAC CGC TTG T-3'	*	-	-	-	-
A20	5'- GTT GCC ATC C-3'	9	9	100	0	0
A19	5'- CAA ACG TCG G-3'	*	-	-	-	-
A09	5'- GGG TAA CGC C-3'	8	8	100	0	0
C06	5'- GAA CGG ACT C-3'	9	8	89	1	11
C05	5'- GAT GAC CGC C-3'	12	12	100	0	0
17	5'- AAA CGG GCG G-3'	10	8	80	2	20
Total		91	87	95.6	4	4.3

* Primers excluded from analyses due to low quality fingerprints.

The GC content of the sequence of a primer may influence its ability to anneal and thus its ability to amplify loci and to produce bands. Therefore, the performances of the primers were also assessed with regards to their GC content. Table 3.5 lists a summary of the relationship between the GC content and the number of bands produced by each of the selected primers. No particular pattern could be identified with this limited number of primers. The ranges of the number of loci amplified by primers with a GC content of 60 % and those with 70 % were very similar; nine to 12 loci for the primers with 60 % GC content and 10 to 13 loci for the primers with 70 % GC content.

Table 3.5 Percentage of GC content and the number of loci amplified by the selected primers.

Primer name	Percentage of GC content	Number of bands produced
A02	70	13
A06	70	7
A10	60	12
A13	70	11
A20	60	9
A09	70	8
C06	60	9
C05	70	12
17	70	10

In order to obtain a better understanding of the effect of the GC content of a primer on the number of loci it amplified, a graphical representation was constructed (Figure 3.4). As mentioned, no clear relationship could be identified between the GC content and the number of bands that a primer produced.

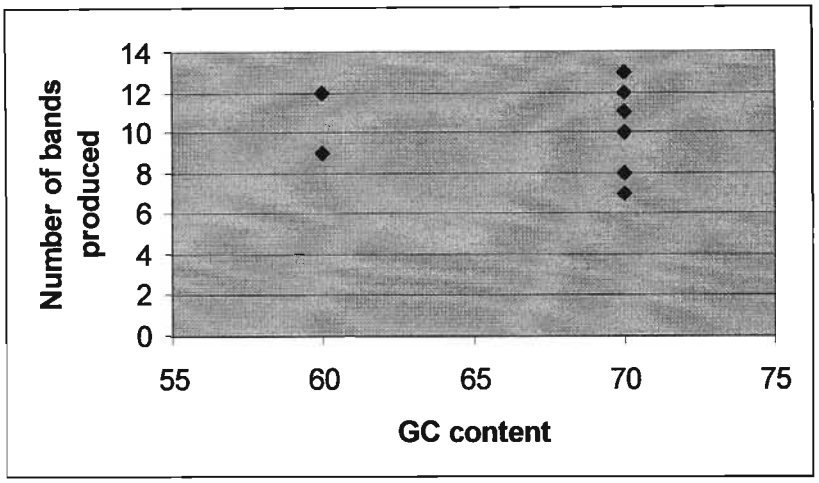

Figure 3.4 **Scatter plot depicting the relationship between the percentage GC content of the primers and the respective number of bands produced.**

3.5 ANALYSIS OF GENETIC VARIATION

Genetic variation of the 34 black wattle plants of the seven groups were estimated by the assessment of RAPD fingerprints generated by the selected primers in the following way:

- The genetic variation was estimated for each of the groups as well as for the population as a whole,
- The genetic relatedness (similarity) between the different groups was estimated, and
- The genetic differentiation (distance) between the different groups was estimated.

3.5.1 Estimation of genetic variation within groups

The genetic variation for the different groups and the whole population was estimated by calculating the mean number of dominant phenotypes, polymorphic content, genetic diversity of Nei (1972), Shannon's Information Index (Magurran, 1988) and the amount of heterozygosity. The frequency of the dominant phenotypes for each locus of the different primers was determined for the population (Table 3.6). Primer C06 displayed the highest percentage of dominant phenotypes in the population (76 %), while primer A10 the lowest (38 %), which indicated that this primer also amplified a large number of monomorphic loci. Thus, primer A10 proved to be the most variable primer with only a few individuals exhibiting bands. In the population as a whole, the mean number of dominant phenotypes over all the loci was 54 %, indicating a degree of genetic variability.

Table 3.6 Frequencies of dominant phenotypes for each locus of the selected primers.

Primer	Frequency of dominant phenotypes: Locus number													Total Ind*	Mean**
	1	2	3	4	5	6	7	8	9	10	11	12	13		
A02	0.13	0.40	0.63	1.00	0.23	0.47	0.93	0.90	0.80	0.77	0.57	0.60	0.60	30	0.63
A06	0.06	0.17	0.24	0.97	0.44	0.76	0.76							34	0.50
A10	0.41	0.06	0.41	0.24	0.29	0.50	0.41	0.44	0.24	0.32	0.85	0.41		34	0.38
A13	0.41	0.76	0.09	0.06	0.12	0.62	0.26	0.41	0.97	0.97	0.29			34	0.44
A20	0.21	0.59	0.09	0.97	0.85	0.68	0.85	0.65	0.38					34	0.59
A09	0.25	0.67	0.13	0.96	0.88	0.38	0.13	0.63						24	0.50
C06	0.76	0.47	0.97	0.97	0.24	0.74	0.74	0.97	1.00					34	0.76
C05	0.32	0.50	0.82	0.55	0.55	0.68	0.45	0.23	0.45	0.59	0.32	0.32		22	0.50
17	0.39	0.61	0.48	0.26	0.61	0.26	0.61	1.00	0.39	1.00				23	0.57

Mean frequency of the dominant phenotypes: 0.54

* Total number of individuals for which a particular primer produced usable results.

** Mean frequency of individuals with the dominant phenotype for a particular primer.

A graphical representation of the mean frequency of dominant phenotypes identified by each primer is presented in figure 3.5. Here too, it is clear that primer C06 resulted in the greatest frequency of individuals displaying the dominant phenotypes. In contrast, primer A10 displayed the lowest frequency.

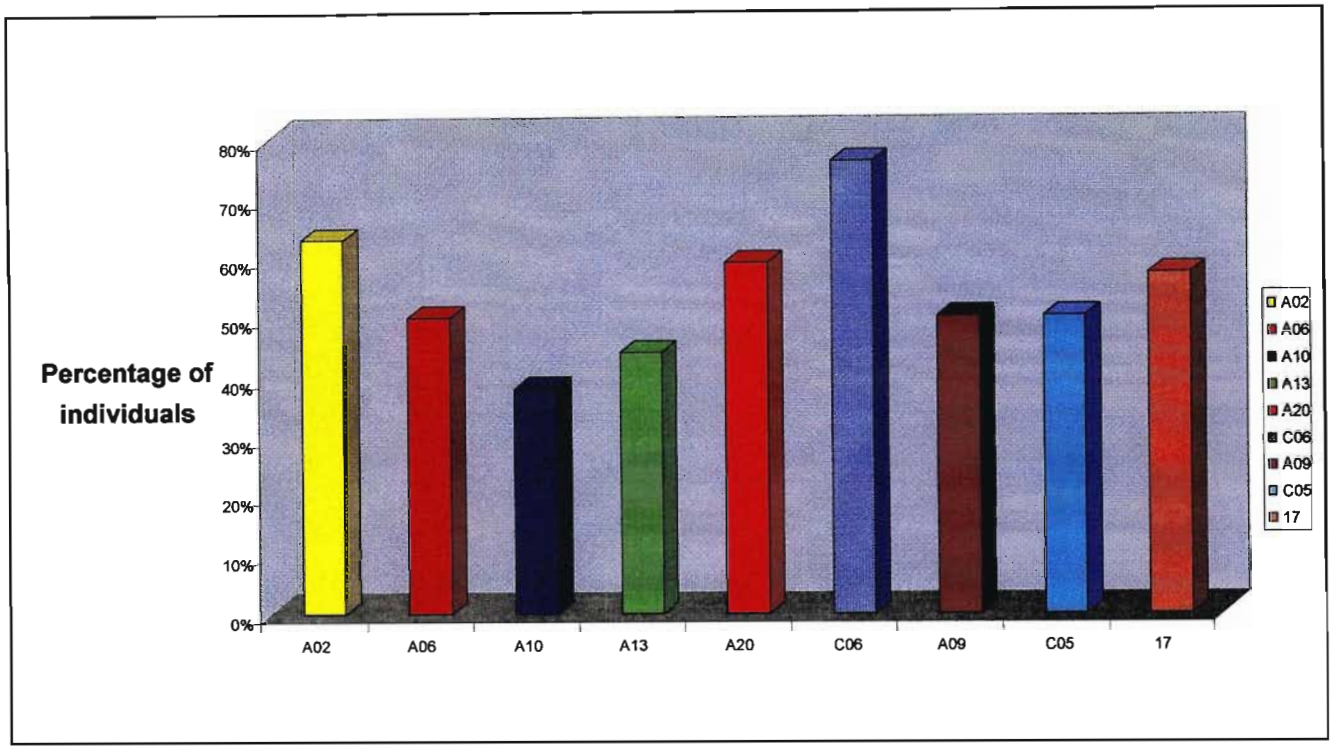

Figure 3.5 A graphical representation of the mean percentage of individuals displaying the dominant phenotype.

The quantification of genetic variation was also estimated within a group. This was undertaken by calculating Nei's gene diversity (1972), Shannon's Information Index (Magurran, 1988), and percentage of polymorphic loci. The genetic measures, Nei's diversity and Shannon's Information Index, showed that the whole population, contained the greatest amount of variation (Nei's gene diversity = 32.09 % and Shannon's Information Index = 48.31 %), when compared to each of the groups (Table 3.7). This was reflected by the percentage of polymorphic loci (95.60 %). The mean Nei's genetic diversity (1972) for individual groups ranged from 14.20 % to 26.83 %, with a mean value across groups of 0.20 ± 0.21 (mean ± 1SE).

These data revealed a number of surprising outcomes. Firstly the two groups, containing 'unrelated' members (groups 400 and 85) displayed the least amount of variation. Group 400 (unrelated) from Howick was the most uniform of all groups investigated. This was also true for group 85 (unrelated) from Kumbula, which displayed the second least variation. These data probably indicate that the area, from which the two groups originated from, developed from a limited number of seeds. The two half-sib groups (47 and 50), from Piet Retief, and three full-sib groups (88, 89 and 91), from an unknown area, unexpectedly displayed the greatest amount of variability, clearly indicated by the number of polymorphic loci, Nei's gene diversity (1972) and Shannon's Information Index (Magurran, 1988). These data emphasise the need for an understanding of the developmental history of plantations. The high variability displayed by the half-sibs and full-sib groups probably reflects the variation within the parents that formed these plantations, thus making the members of these progenies so variable.

Table 3.7 Quantification measures of genetic variation.

Group	Polymorphic loci (%)	Nei's gene diversity (%)	Shannon's Information Index (%)
Group			
47	58.24	22.46 (0.22±0.21)*	33.08 (0.33±0.30)*
50	53.85	21.01 (0.21±0.21)*	30.82 (0.31±0.30)*
85	40.66	16.67 (0.17±0.21)*	24.21 (0.24±0.30)*
88	72.53	26.83 (0.27±0.19)*	39.86 (0.40±0.27)*
89	63.74	23.50 (0.24±0.20)*	34.92 (0.35±0.29)*
91	45.05	17.62 (0.18±0.21)*	25.87 (0.26±0.30)*
400	18.68	14.20 (0.14±0.21)*	20.69 (0.21±0.30)*
Whole population	95.60	32.09 (0.32±0.16)*	48.31 (0.48±0.20)*

* Denotes mean and ± 1 standard error values.

The heterozygosity (2pq) of the population was estimated by determining the frequency of the null-phenotype at each locus across all primers and estimating the frequency of p and q. The number of individuals in this investigation was 34; therefore loci with a null-phenotypic frequency less than 0.1 were regarded as monomorphic and excluded from the analyses (Table 3.8). In this population, besides the four true monomorphic loci, A02.4, C06.9, 17.8 and 17.10, an additional nine loci were reclassified as monomorphic according to the recommendations of Lynch and Milligan (1994); loci A02.7, A06.4, A13.9, A13.10, A20.4, A09.4, C06.3, C06.4 and C06.8. The calculated mean frequency of *null-allele* (q) was estimated to be 0.61.

Table 3.8 Estimated frequencies of the *null-alleles* (q) for each locus.

	Primer								
	A02	**A06**	**A10**	**A13**	**A20**	**A09**	**C06**	**C05**	**17**
1	0.93	0.97	0.77	0.77	0.89	0.87	0.49	0.83	0.78
2	0.77	0.91	0.97	0.49	0.64	0.58	0.73	0.71	0.63
3	0.61	0.87	0.77	0.95	0.95	0.94	0.00	0.43	0.72
4	0.00	0.00	0.87	0.97	0.00	0.00	0.00	0.67	0.86
5	0.88	0.75	0.84	0.94	0.38	0.35	0.87	0.67	0.63
6	0.73	0.49	0.71	0.62	0.57	0.79	0.51	0.56	0.86
7	0.00	0.49	0.77	0.86	0.38	0.94	0.51	0.74	0.63
8	0.32		0.75	0.77	0.59	0.61	0.00	0.88	0.00
9	0.45		0.87	0.00	0.79		0.00	0.74	0.78
10	0.48		0.82	0.00				0.64	0.00
11	0.66		0.38	0.84				0.83	
12	0.63		0.77					0.83	
13	0.63								
Mean frequency of *null-allele* (q):	**0.61**								

After the frequencies of the *null-alleles* (q) were determined, the frequencies of the *presence-alleles* (p) were calculated by applying the formula p=1-q. Thereafter the frequencies of heterozygosities could be estimated by applying the formula 2pq (Table 3.9). The average heterozygosity values for the different

primers ranged from 0.2 to 0.4. The primer exhibiting the largest amount of heterozygosity was primer C05, with a mean value of 0.4. Primers C06 and A13 exhibited the least amount of heterozygosity, 0.2. The total heterozygosity frequency across all loci for the total population as a whole was estimated as 0.3.

Table 3.9 Estimated heterozygosity (2pq) values for each locus.

Locus	Primer								
	A02	**A06**	**A10**	**A13**	**A20**	**C06**	**A09**	**C05**	**17**
1	0.13	0.06	0.35	0.35	0.20	0.50	0.27	0.28	0.34
2	0.35	0.16	0.06	0.50	0.46	0.39	0.49	0.46	0.47
3	0.48	0.23	0.35	0.10	0.10	0.00	0.11	0.49	0.40
4	0.00	0.00	0.23	0.06	0.00	0.00	0.00	0.44	0.24
5	0.21	0.38	0.27	0.11	0.47	0.23	0.46	0.44	0.47
6	0.39	0.50	0.41	0.47	0.49	0.50	0.33	0.49	0.24
7	0.00	0.50	0.35	0.24	0.47	0.50	0.11	0.38	0.47
8	0.44		0.38	0.35	0.48	0.00	0.48	0.21	0.00
9	0.49		0.23	0.00	0.33	0.00		0.38	0.34
10	0.50		0.30	0.00				0.46	0.00
11	0.45		0.47	0.27				0.28	
12	0.47		0.35					0.28	
13	0.47								
Mean heterozygosity: 0.3									

Average = average heterozygosity value for each primer.

3.5.2 Estimation of genetic relatedness and differences between and within groups

Genetic relatedness was estimated between members of groups, between groups and between members of the population. This was done by calculating Nei's gene diversity (1972) and Nei's gene identity (1972). Figure 3.6 depicts the ranges of identities for each of the different groups collected from the pairwise analyses of the identities and distances displayed in figure 3.7. These pairwise comparisons support the variation

measures obtained for each of the groups using percentage of polymorphism, Nei's gene diversity (1972) and Shannon's Information Index (Magurran, 1988). The unrelated groups (groups 400 and 85) both displayed the greatest level of similarity with the difference between the lowest and highest pairwise estimate being only 10 % for group 400, but 19 % for group 85. The full-sib groups (88, 89 and 91), as found with the quantification of genetic variation within these groups, displayed a wide range of identities from as low as 47 % in group 88, to as high as 86 % in group 91. As expected, the half-sib groups' range of identities (47 and 50), were somewhat in-between that of the two unrelated groups (groups 85 and 400) and the full-sib groups (88, 89 and 91).

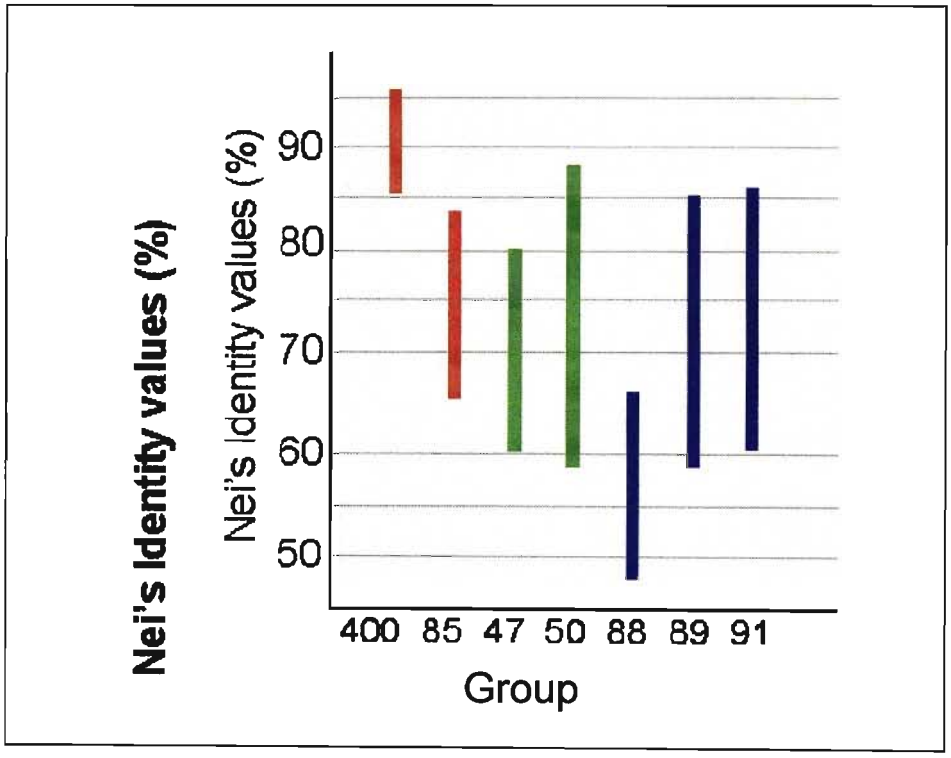

Figure 3.6 **Range of pairwise identities (similarities) for each black wattle group.**

a) Group 47

Family Identification	47-1	47-2	47-3	47-4	47-5
47-1	****	0.7253	0.6264	0.8022	0.6813
47-2	0.3212	****	0.6813	0.6374	0.6923
47-3	0.4678	0.3837	****	0.6044	0.7473
47-4	0.2204	0.4504	0.4504	****	0.6813
47-5	0.3837	0.3677	0.2914	0.3837	****

b) Group 50

Family Identification	50-1	50-2	50-3	50-4	50-5
50-1	****	0.8022	0.7253	0.6374	0.6593
50-2	0.2204	****	0.8791	0.7033	0.6593
50-3	0.3212	0.1288	****	0.7802	0.5824
50-4	0.4504	0.3520	0.2482	****	0.6264
50-5	0.4165	0.4165	0.5406	0.4678	****

c) Group 85

Family Identification	85-1	85-2	85-3	85-4	85-5
85-1	****	0.7473	0.8352	0.6484	0.6484
85-2	0.2914	****	0.7802	0.7033	0.7692
85-3	0.1801	0.2482	****	0.7473	0.6593
85-4	0.4333	0.3520	0.2914	****	0.7802
85-5	0.4333	0.2624	0.4165	0.2482	****

d) Group 88

Family Identification	88-1	88-2	88-3	88-4	88-5
88-1	****	0.6044	0.6593	0.4725	0.5934
88-2	0.5035	****	0.5714	0.5495	0.6593
88-3	0.4165	0.5596	****	0.5495	0.6484
88-4	0.7497	0.5035	0.5988	****	0.6374
88-5	0.5219	0.4165	0.4333	0.4504	****

e) Group 89

Family Identification	89-1	89-2	89-3	89-4	89-5
89-1	****	0.6593	0.7912	0.5824	0.8022
89-2	0.4165	****	0.6044	0.7033	0.6154
89-3	0.2342	0.5035	****	0.6154	0.7253
89-4	0.5406	0.3520	0.4855	****	0.6264
89-5	0.2204	0.4855	0.3212	0.4678	****

f) Group 91

Family Identification	91-1	91-2	91-3	91-4	91-5
91-1	****	0.6044	0.8462	0.6044	0.7033
91-2	0.5035	****	0.6923	0.8242	0.7473
91-3	0.1671	0.3677	****	0.6484	0.8571
91-4	0.5035	0.1934	0.4333	****	0.7473
91-5	0.3520	0.2914	0.1542	0.2914	****

g) Group 400

Family Identification	401	402	403	404
401	*****	08571	0.8571	0.8571
402	0.1542	****	0.9560	0.9341
403	0.1542	0.0450	*****	0.9341
404	0.1542	0.0682	0.0682	*****

Figure 3.7 **Collection of pairwise analyses of Nei's identities and distances. Identity values above diagonal and distance values below diagonal.**

The relatedness between the seven different groups was also established with Nei's identity and distance measures (1972). The pairwise analysis of identity showed that the relationship between group 400 and all the other groups was the lowest in the population, ranging from 64 % to 79 % (Table 3.10). The relationship between all the groups besides group 400 was reasonably high, ranging from 78 % to 90 %, which supported all previously presented data. This suggests that the groups are closely related to each other.

Table 3.10 Nei's genetic identity (above diagonal) and genetic distance (below diagonal) obtained for the seven groups.

Group	47	50	85	88	89	91	400
47	****	0.8961	0.8361	0.8562	0.8798	0.8387	0.6889
50	0.1097	****	0.7782	0.8559	0.8404	0.7949	0.7949
85	0.1790	0.2508	****	0.7932	0.7800	0.8147	0.6683
88	0.1552	0.1556	0.2317	****	0.9047	0.8388	0.6379
89	0.1280	0.1739	0.2485	0.1002	****	0.8744	0.6627
91	0.1759	0.2295	0.2050	0.1758	0.1342	****	0.6708
400	0.3727	0.4459	0.4031	0.4495	0.4114	0.3993	****

Nei's pairwise analysis of identity and similarity were determined and dendrograms constructed using UPGMA. The dendrograms were constructed between group members within a group to provide a visual account of the genetic relationships (Figure 3.8).

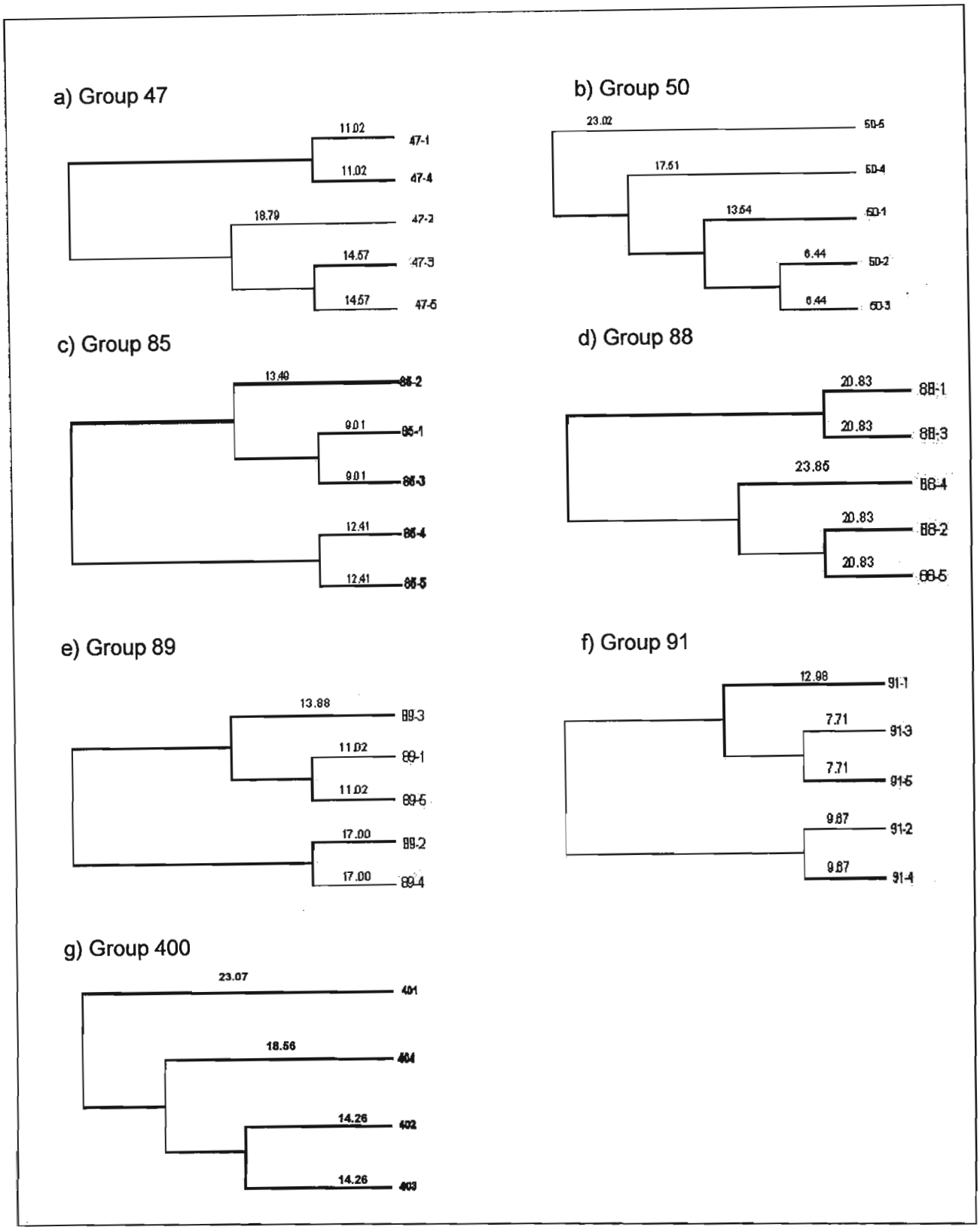

Figure 3.8 **Genetic trees showing genetic relationships between individuals of groups based on Nei's genetic distance.**

A dendrogram was prepared to display the genetic relationships between the different groups. In support of all data previously presented, group 400 was the most distantly placed in the tree. Interestingly the two half-sib groups (groups 47 and 50) were grouped together as well as two of the full-sib groups (groups 88 and 89), showing the close genetic relationship between these groups.

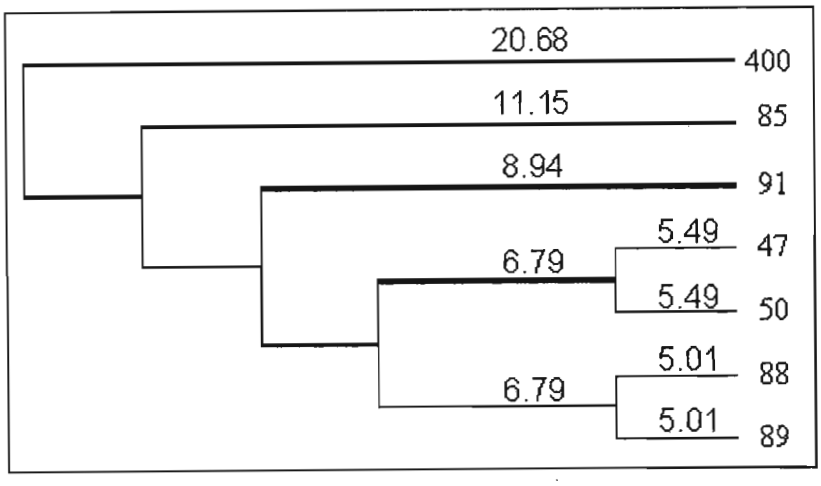

Figure 3.9 Genetic tree showing the relationship between the seven groups based on Nei's genetic distance.

Genetic relatedness of all 34 individuals were compared by applying Nei's identity and distance pairwise analysis (1972) and constructing a dendrogram using UPGMA (Sneath and Sokal, 1973). Contrary to expectation, only group 400 (highlighted in pink) had all its individuals clustered together. Members of the half-sib and full-sib groups appeared to be more randomly distributed in the tree with members of the half-sib groups, as well as full-sib groups, tending to cluster into smaller groups. By these groups' members branching off together, suggests that the members of the groups are genetically very similar to the other groups and genetically different from the members within it's group.

The group information revealed that group 400 was frost tolerant, whilst the other groups were not. This could explain the divergent results that this group displayed throughout this investigation, as its genetic composition was rather different to the frost intolerant

individuals, which probably explains the difficulty experienced in amplifying the loci of some primers. The two half-sib groups (47 and 50) did not display perfect clustering although many of their members tended to cluster together, which can be explained by the fact that they both originated from the Piet Retief area in Mpumalanga and probably share similar genetic backgrounds.

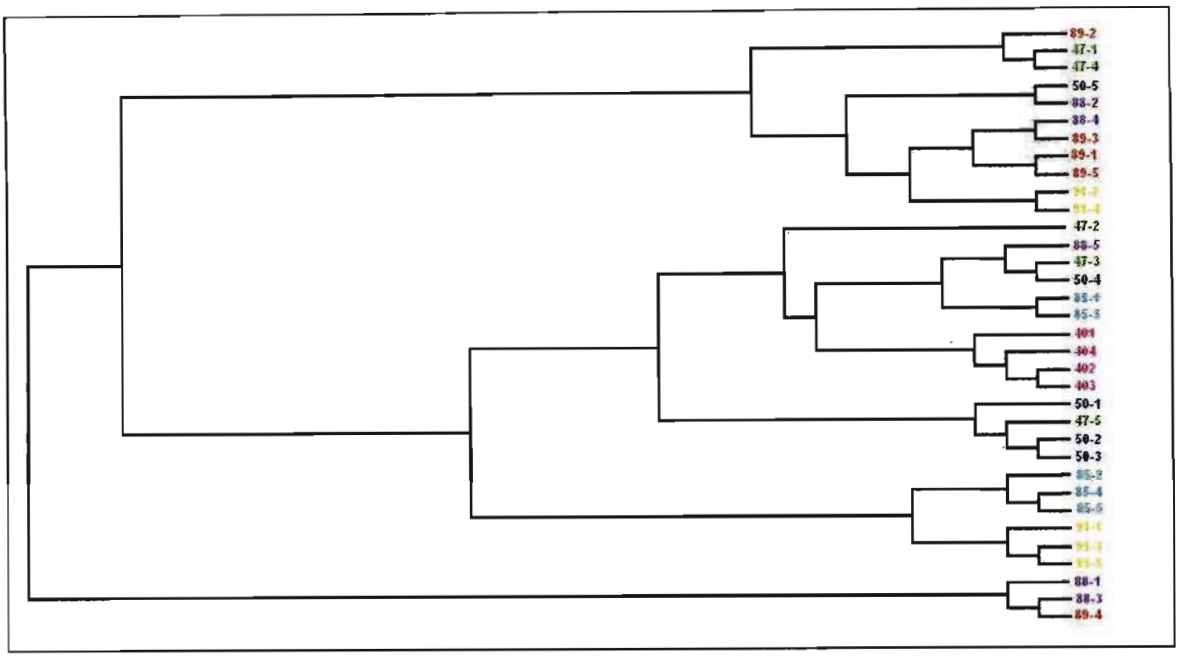

Figure 3.10 Genetic tree for individuals in the population based on Nei's gene diversity.

3.5.3 Estimation of group differentiation

The contribution of the within group variation and between group variation to the total variation was estimated to obtain an idea of the group differentiation. This analysis was undertaken by conducting an Analysis of Molecular Variance (AMOVA), by utilising the software programme ARLEQUIN version 2.000 (Schneider *et al.*, 2000). AMOVA estimated the relative group differentiation directly from molecular data. Random permutations (1023) were used to test the significance of the covariance components. The results of the AMOVA partitioning of RAPD variance among and within groups, revealed a total genetic variation of 82.65 % was attributable to within group variation, and only 17.44 % to between groups variation in black wattle (Table 3.11). However,

the proportion of variation attributed to between groups and within groups variation were both found to be highly significant (P<0.001) thus indicating that there were significant genetic differentiation between groups.

Table 3.11 Partitioning of total variation using AMOVA.

Source of variation	df*	Sum of squares	Variance components	Percentage of variation
Among groups	6	79.985	1.39040	17.44
Within groups	27	177.750	6.58333	82.56
Total	33	257.735	7.97374	

Fixation Index F_{ST}: 0.17437

* df = degrees of freedom

Wright's fixation index (F_{ST}) was also computed to estimate group differentiation, but unlike AMOVA that estimates group differentiation directly from molecular data, this index estimates the differentiation by assuming Mendelian gene frequencies. The calculation of Wright's fixation index (F_{ST}), confirmed the results displayed in the AMOVA analysis, revealing a significant F_{ST} value of 0.17437, which is the same value as that revealed by AMOVA. According to Wright's guidelines, a value between 0.05 and 0.15 indicates great genetic differentiation; therefore these results indicate that there has been some extensive differentiation between the different groups.

CHAPTER FOUR

DISCUSSION AND CONCLUSION

4.1 INTRODUCTION

Black wattle, indigenous to Australia, is grown for its timber and mainly its bark products. The black wattle industry in South Africa has in recent times become a popular and competitive industry, thus recognising the need for research into various industry-related aspects such as cultivation and product quality. Research is mostly aimed at improving the quality of the product and introducing disease resistance by devising breeding programmes, as well as investigating new uses of the species. Due to the introduction history of the species, it is expected that the genetic base of the South African population is narrow. Therefore, this investigation was undertaken to obtain an idea of the breadth of the genetic variability of the Southern African black wattle population. This knowledge will greatly facilitate the breeders when searching and selecting superior genotypes for breeding purposes.

4.2 RAPD ANALYSES AND GENETIC VARIATION

In recent times, newly developed molecular markers such as Restriction Fragment Length Polymorphisms (RFLP) (Tanksley *et al.,* 1989), Randomly Amplified Polymorphic DNA (RAPD) (Williams *et al.*, 1990), Amplified Fragment Length Polymorphism (AFLP) (Vos *et al.,* 1995) and microsatellites and minisatellites (Tingey and Tufo, 1993), have contributed extensively to the measurement and understanding of genetic variation in populations and species. In this investigation RADP fingerprints were generated from the seven different groups of black wattle, each consisting of four or five individuals, from different regions in South Africa.

The introduction history of the black wattle into South Africa suggests that it was brought into the country 'in a bag of seeds'. As a result very little is known of the origins and the genetic variability of these plants. However what is known is that the genetic variability between and within groups should be low due to inbreeding between these plants. However from the results in this investigation it was revealed that a considerable amount of genetic variation exists, with 95.6 % of the loci being polymorphic across the nine primers. These results are similar to that found in the study by Shrestha *et al.* (2002), who used RAPD fingerprinting to investigate genetic variability in 12 populations of *Acacia raddiana* from the Arava valley and western Negev. The *Acacia raddiania* populations also showed a high degree of polymorphism (90.69 %). These results are contrary to conventional expectation of small, isolated populations. Playford *et al.* (1993) had also discovered that a significant amount of polymorphism (69 %) existed in the *Acacia melaoxylon* population. A similar result was also seen in study on the *Acacia anamala* (Coates, 1988), where the level of polymorphism was 61.1 %. These studies, both conducted on Australian acacia, suggested that the high increase in polymorphism values was due to the increase in latitude, with respect to geographical distribution.

An additional measure used to assess genetic variation in a population is the calculation of the mean heterozygosity. The overall mean expected heterozygosity estimated for *Acacia mearnsii* in this study was found to be 0.3. According to Moran (1992), who conducted a collective study on a range of acacia species and other hardwoods; the average heterozygosity value for acacia species was found to be 0.132. Therefore the heterozygosity value (H=0.3), in this study, is higher than other acacia species. However, there are many studies that support this high heterozygosity value. Moran (1992) found high levels of heterozygosity (0.208) in *Acacia melanoxylon* populations in association with a great genetic differentiation among geographic areas. High values of heterozygosity were also found in African populations of *Acacia albida* (Joly *et al.*, 1992), with values ranging from 0.260 to 0.442. Harsh (2000) used allozyme analysis to determine the population genetic structure of two *Acacia brevispica* populations. The results from the allozyme analysis showed that individuals exhibited a high heterozygosity value (0.27).

Although high values of heterozygosity were found in this study, there are many other studies on hardwoods and acacia that exhibited low levels of heterozygosity. Tropical angiosperms, such as *Eucalyptus* (0.174) and *Casuarina* (0.211) displayed low heterozygosity values (Moran, 1992). Estimated heterozygosity values were also lower in Australian species of acacia, for example, *Acacia auriculiformis* and *Acacia crassicarpa* (0.146 and 0.141 respectively) (Moran *et al.,* 1989), *Acacia aulacocarpa* (0.112) (McGranahan *et al.,* 1997) and *Acacia mangium,* which had the lowest value (0.017) (Moran *et al.,* 1989). The low levels of genetic variation found in these studies occurred due to these populations been founded on a narrow genetic base or these populations may have experienced a decline in diversity because of isolation and small size.

The high levels of genetic variability present in this study, as compared to the other studies, may be related to mating systems and geographic distribution. Species with cross fecundation and a wide geographic range, as presented in this study, have higher levels of genetic diversity than do selfer or endemic species. Similarly, species whose seeds are dispersed by animals or wind, maintain high levels of within-population genetic variability (Hamrick and Godt, 1989). Disassortative mating also could have led to the increase in heterozygosity. It is also thought that rare allele advantage could be another contributing factor to an increased heterozygosity. In some mating systems a male bearing a rare allele will have a mating advantage. Rare allele advantage will tend to increase the frequency of the rare allele and hence increase heterozygosity (Chamberlain and Hubert, 1998).

4.3 POPULATION AND GROUP ANALYSIS

The data generated in this investigation revealed a number of surprising outcomes. This study demonstrated low levels of genetic diversity in group 400 (unrelated), from the Howick region, and group 85 (unrelated), from the Kumbula region, (mean 0.14 and 0.17 respectively). These values are low and indicate that little genetic variation exists in

these two groups. These values probably indicate the limited gene exchange between groups in the Kumbula regions and the Howick regions. These groups being genetically different from the other groups could mean that they have been reared from plants that came from the same area. The low levels of genetic diversity in these areas could also suggest that these groups have been founded on a narrow genetic base or have experienced a decline in diversity because of isolation and small size (Butcher *et al.*, 1998). Bottlenecks associated with long-distance founding events, the absence of migration as a source of genetic enrichment and the possibility of selection pressures in new environments are all also likely to lead to a loss of genetic variation during colonizing events (Barret and Husband, 1989). It should be mentioned, that group 400 is frost tolerant while the other groups are not. As a result, it was expected that this group branched off separately from the other groups, as presented in the dendrogram.

The two half-sib groups (groups: 47 and 50) from Piet Retief, and three full-sib groups (88, 89 and 91), from the unknown area, unexpectedly displayed the greatest amount of variability. The cluster analysis also revealed that groups 47 and 50 (half-sibs) from Piet Retief branched off together. This revealed that groups 47 and 50 (half-sibs) from Piet Retief are genetically different from group 400 (unrelated) from Howick, and group 85 (unrelated) from Kumbula, and very little or no gene exchange could have occurred between these groups. The three full-sib groups, 88, 89 and 91, were the most variable in comparison to other groups, revealing means of (0.27, 0.24 and 0.18 respectively). According to Moran (1992), most acacias are insect-pollinated and this could lead to gene migration between groups or populations, thus explaining the high genetic variations present in these groups. The high genetic variation could also be due to possible inclusion of introduced populations, in this case the groups' parents, which were of unknown and possibly mixed origins.

From a conservation point of view, and based on the cluster analysis, to increase or maintain the high genetic variation of some of the groups, it would be best to conserve each group separately. On the other hand, genetic variation can be increased or introduced into some of the groups, if the groups that are distantly placed from each

other on the tree, are crossed with each other. For example, group 400 has a particular trait that is not present in the other groups; group 400 is frost tolerant and the other groups are not. This would benefit the plant breeder, in that the frost tolerant attribute would protect black wattle trees against the harsh winter climate in South Africa.

From the above results it can be noted that gene flow could have occurred between groups 47 and 50 (half-sibs) from Piet Retief and groups 88 and 89 (full-sibs) from the unknown region, since they are genetically similar to each other. One way in which gene flow can be measured is to calculate the fixation index, F_{ST}. This value varies from zero to one. If a species is divided into many strongly differentiated subpopulations, as a result of gene flow, the F_{ST} value will be large. On the other hand, if the subpopulations within a species are similar and undifferentiated due to extensive gene flow between them, this value will be small (Hartl and Clark, 1997).

Moran (1992) reviewed the gene flow of three acacia species, *Acacia melaoxylon* (F_{ST}= 0.377), *Acacia anamala* (F_{ST}= 0.056), and *Acacia mangium.* (F_{ST}= 0.311). The two species of acacia, *Acacia melaoxylon* and *Acacia mangium,* both displayed high F_{ST} values. These values are quite high in comparison to other woody species which rarely exceed 10-15 % (Hamrick and Godt, 1989). Therefore from the studies conducted by Moran (1992) it can be concluded that substantial gene flow is present in acacias with the subpopulations in this species being highly differentiated. In this study the estimated F_{ST} value obtained was 0.174 between the seven groups investigated. According to Hartl and Clark (1997) an F_{ST} value between 0.15 and 0.25 indicates great genetic differentiation. Therefore the value (F_{ST} = 0.174) obtained in this study indicates that quite a significant amount of genetic variation exists among these seven groups. As mentioned previously, it was expected that the black wattle population in South Africa has a narrow genetic base due to the introductory history of this species. However, these results have contradicted this expectation since the South African black wattle population does have a considerable amount of genetic variation. This is imperative as breeders can now use this information for sampling strategies and for future breeding and domestication programmes.

The results of the AMOVA partitioning of RAPD variance between and within groups revealed a total genetic variation of 82.65 % ascribed to within group variation, and 17.44 % to between groups variation in *Acacia mearnsii*. The high genetic variation found within groups explains why the full-sib and half-sib groups displayed the greatest amount of genetic variation as compared to the other groups. It also clarifies why these groups were distributed randomly in the population dendrogram as opposed to being grouped in their individual groups as was initially expected. This level of group differentiation is far higher than is usually encountered in outcrossing species (Shrestha *et al.,* 2002). This high level of group differentiation may be explained by several factors, including the species breeding system, genetic drift or genetic isolation of populations or groups. Hamrick and Godt (1989), whose study support the findings in this study, have also revealed that acacia species generally possess traits that are correlated with high variability within groups/populations and small genetic differences between groups/populations. There are few studies that exist that test this hypothesis. Moreover, most studies have been conducted on Australian species with limited research on the African species. There is a wide range of levels of genetic diversity and unexpectedly high levels of diversity between populations, in the Australian acacia species studied to date (Moran *et al.*, 1990; Playford *et al.,* 1993; Wickneswari and Norwati, 1993). The one African species studied, *Acacia albida*, which support the findings in this study, also had a high genetic diversity within groups and moderate differences between populations (Joly *et al.,* 1992). Another study with similar findings conducted by Shrestha *et al.* (2002), on *Acacia raddiana,* also revealed that an extensive population differentiation of 59.4 % attributed between populations and a lower value of 40.8 % occurred within populations.

4.4 CONCLUSION

The focus of this research was to assess genetic variation between different black wattle groups, using RAPD fingerprinting. RAPD fingerprinting to assess genetic variability in black wattle is the first of its kind undertaken in the South African population. This research used RAPD fingerprinting with a fixed sample size (maximum of five individuals), limited group representation (seven groups), and minimal primers. Although these were limitations, the research outcome has produced valuable results. These results have shown that some degree of genetic variation exists, between different groups of black wattle, from the different areas of South Africa. For example, some groups have shown that they are distinct from others. Also some groups have shown that they are related to each other to some degree. A benefit of determining this distinctness and relatedness is to improve quality control in breeding programmes and seed orchard management. Hence, this investigation has proven that RAPD fingerprinting is a powerful tool that aids in understanding the inherent variability of black wattle. These results have also correlated with other similar research and extends in the research of black wattle in South Africa.

4.4.1 Recommendations for future work

The knowledge obtained in this research will be able to make a significant contribution to selection of genotypes for breeding activities. It should however be stressed that in order to obtain a more accurate assessment of genetic variability and relatedness of South African black wattle, a number of factors should be taken into consideration for future studies.

- More individuals may be included in the study, for example 20-30 individuals per group,
- The number of groups may be increased,

- The groups may be included from all provinces in South Africa, in an attempt to include a wider, better representation of black wattle in South Africa, and
- Different primers may be tested.

This study approach focused on using the RAPD technique that is widely employed. Because the use of RAPDs for the estimation of nucleotide variability requires that results be repeatable, a number of technical aspects should be considered. Band consistency of 90-100 % is attainable if several precautions are taken. Following Welsh and McClelland (1990) and from experience gained from this study it is recommended that:

- All template concentrations must be uniform across all samples. Consistent results are obtained if all template samples are titrated to concentration and measured against known quantitative DNA standards on agarose gels (Borowsky *et al.*, 1995).
- The type of buffers used makes a marked difference with regards to band distinction. TAE buffer has shown to have a better buffering capacity of smaller DNA fragments than TBE buffer.
- Running the products on a less concentrated gel. In this case a 1% gel as opposed to a 2 % gel, provides a sensitive and quantitative measure of band presence.

Thus using RAPDs, it is possible to reliably assess gene diversity in populations using DNA samples from as few as five individuals. This approach should prove useful in large scale comparisons of species and in the study of populations that are endangered or relatively inaccessible.

REFERENCES

Ashburner G R, Thompson W K and Halloran G M. (1997). RAPD Analysis of South Pacific Coconut Palm Populations. *Crop Science.* 37: 992-997.

Avise J C. (1994). Molecular Markers, Natural History and Evolution. Chapman and Hall, New York.

Bader J M. (1998). Measuring Genetic Variability in Natural Populations by Allozyme Electrophoresis. *In*: Tested Studies for Laboratory Teaching. Karcher S J (Editor). Proceedings of the 19th Workshop/Conference of the Association for Biology Laboratory Education. 19: 25-42.

Barret S and Husband B C. (1989). The Genetics of Plant Migration and Colonisation. *In:* Plant Population Genetics, Breeding and Genetic Resources. Brown A H D, Clegg M T, Kahler A L and Weir B S (Editors). Sinauer Associates, Sunderland, MA. 254-277.

Beckmann J S and Soller M. (1986). Restriction Fragment Length Polymorphism in Plant Genetic Improvement. *Oxford Surveys of Plant Molecular and Cell Biology.* 3: 196-225.

Bekessey S A, Allnutt T R, Premoli A C, Lara A, Ennos R A, Burgman M A, Cortes M and Newton A C. (2002). Genetic Variation in the Vulnerable and Endemic Monkey Puzzle Tree, Detected Using RAPDs. *Heredity.* 88: 243-249.

Borowsky R, McClelland M, Cheng R and Welsh J. (1995). Arbitrarily Primed DNA Fingerprinting for Phylogenetic Reconstruction in Vertebrates. *Molecular Biology and Evolution.* 12: 1022-1032.

Broadhurst L and Coates D. (2002). Genetic Diversity Within and Divergence Between Rare and Geographically Widespread Taxa of the *Acacia acuminata* Benth. (Mimosaceae) Complex. *Heredity.* 88: 250-257.

Bruford M W and Wayne R K. (1993). Microsatellites and their Application to Population Genetic Studies. *Current Opinion in Genetics and Development.* 3: 939-943.

Butcher P A, Moran G F and DeCroocq S. (1998). Use of Molecular Markers in Domestication and Breeding Programs for Acacias. *In*: Recent Developments in Acacia Planting. Turnbull J W, Crompton H R and Pinyopusarerk K (Editors). Proceedings of an International Workshop, Hanoi, Vietnam. 82: 211-215.

Carr J D. (1976). The South African Acacias. First Edition. Conservation Press (PTY) LTD.

Chamberlain J R and Hubert J D. (1998). Molecular Analysis of Genetic Variation. *Tropical Forestry Paper.* 4: 67-73.

Cheliak W M and Rogers D L. (1990). Integrating Biotechnology into Tree Improvement Programs. *Canadian Journal of Forest Research.* 20: 452-463.

Coates D J. (1988). Genetic Diversity and Population Genetic Structure in the Rare Chittering Grass Wattle *Acacia anomala. Australian Journal of Botany.* 36: 273-286.

Department of Forestry and Water Affairs. (2003). South African Forestry Facts for the Year 2001/2002. Newsletter. Unpublished.

Doyle J J and Doyle J L. (1990). Isolation of Plant DNA from Fresh Tissue. *Focus.* 12: 13-15.

Dunlop R W. (2002). Tree Improvement. *In:* Black Wattle: The South African Research Experience. Dunlop R W and MacLennan L A (Editors). Insititute for Commercial Forestry Research. University of Natal Press, Pietermaritzburg.

Excoffier L, Smouse P E and Quattro J M. (1992). Analysis of Molecular Variance Inferred from Metric Distances Among DNA Haplogroups: Applications to Human Mitochondrial DNA Restriction Data. *Genetics.* 131: 479-491.

Goldstein D and Schlotterer C. (1999). Microsatellites: Evolution and Applications. Oxford University Press.

Hadrys H, Schierwater B, Dellaporta S L and Buss L W. (1993). Determination of Paternity in Dragonflies by Random Amplified Polymorphic DNA Fingerprinting. *Molecular Ecology*. 2: 79-87.

Hamrick J I and Godt M J W. (1989). Allozyme Diversity in Plant Species. *In*: Plant Population Genetics, Breeding and Genetic Resources. Brown A H D, Clegg M T, Kahler A L and Weir B S (Editors). Sinauer Associates, Sunderland, MA. 43-44.

Harsh J. (2000). Population Genetic Structure of *Acacia brevispica* from East Africa. *Australian Journal of Botany*. 40: 59-73.

Hartl D L and Clark A G. (1997). Principles of Population Genetics. Third Edition. Sinauer Associates, Sunderland MA.

Hwang S Y, Lin H W, Kuo Y S and Lin T P. (2001). RAPD Variation in Relation to Population of *Chamaecyparis formosensis* and *Chamaecyparis taiwanensis. Botanical Bulletin of Academia Sinica*. 42: 173-179.

Jeffreys A J, Wilson V and Thein S L. (1985). Hypervariable Minisatellite Regions in Human DNA. *Nature*. 314: 67-73.

Joker D. (2000). *Acacia mearnsii* De Wild. Danida Forest Seed Centre. Seed Leaflet: No. 4. Unpublished.

Joly H I, Zeh-Nlo M, Danthu P and Aygalent C. (1992). Population Genetics of an African Acacia, *Acacia albida:* Genetic Diversity of Populations from West Africa. *Australian Journal of Botany*. 40: 59-73.

Jordano P and Godoy J A. (2000). RAPD Variation and Population Genetic Structure in *Prunus mahaleb* (Rosaceae), an Animal-Dispersed Tree. *Molecular Ecology*. 9: 1293-1305.

Kevin T. (2000). Black Wattle for Farm Forestry. State of Victoria, Department of Natural Resources and Environment. Unpublished.

Krawczak M and Schmidtke J. (1998). DNA Fingerprinting. Second Edition. BIOS Scientific Publishers, Oxford.

Lebrun P, N'Cho N P, Seguin M, Grivet L and Baudouin L. (1998). Genetic Diversity in Coconut (*Cocos nucifera*) Revealed by Restriction Fragment Length Polymorphism (RFLP) Markers. *Euphytica.* 101: 103–108.

Lefort F, Lally M, Thompson D and Douglas G C. (1998). Morphological Traits, Microsatellite Fingerprinting and Genetic Relatedness of a Stand of Elite Oaks (*Quercus Robur*) at Tullynally, Ireland. *Silvae Genetica.* 47: 257-262.

Lexer C, Heinze B, Steinkellner H, Kampfer S, Ziegenhagen B and Glosssl J. (1999). Microsatellite Analysis of Maternal Half-Sib Families of *Quercus Robur*, Pedunculate Oak: Detection of Seed Contaminations and Inference of the Seed Parents from the Offspring. *Theoretical and Applied Genetics.* 99: 185-191.

Lynch M and B G Milligan. (1994). Analysis of Population Genetic Structure with RAPD Markers. *Molecular Ecology*. 3: 91-99.

Magurran A E. (1988). Ecological Diversity and its Measurement. Princeton University Press, Princeton, NJ.

Magurran A E and May R M. (1999). Evolution of Biological Diversity. Oxford University Press.

Matyas C. (1999). Forest Genetic Sustainability. Kluwer Academic Publishers.

McGranahan M, Bell J C, Moran G F and Slee M. (1997). High Genetic Divergence Between Geographic Regions in the Highly Outcrossing Species *Acacia aulocarpa* (Cunn. ex Benth). *Forest Genetics.* 4: 1-13.

Moran G F, Muona O and Bell J C. (1989). Breeding Systems and Genetic Diversity in *Acacia auriculiformis* and *Acacia crassicarpa. Biotropica.* 21: 250-256.

Moran G F, Muona O and Bell J C. (1990). *Acacia mangium,* A Tropical Forest Tree of the Coastal Lowlands with Low Genetic Diversity. *Evolution.* 43: 231-235.

Moran G F. (1992). Patterns of Genetic Diversity in Australian Tree Species. *New Forests.* 6: 49-66.

Mosseler A, Egger K N and Hughs G A. (1992). Low Levels of Genetic Diversity in Red Pine Confirmed by Random Amplified Polymorphic DNA Markers. *Canadian Journal of Forest Research.* 22: 1332-1337.

Mullis K B and Faloona F A. (1987). Specific synthesis of DNA *in vitro* Through a Polymerase-Catalysed Chain Reaction. *Methods in Enzymology.* 155: 335-350.

Muro-Abad J I, Gomes E A, Cancio O N and DeAraujo E F. (2001). Genetic Analysis of *Eucalyptus urophylla* and *Eucalyptus grandis* Clones Selected in Commercial Crops from the Brazalian Amazon by RAPD Markers. *Silvae Genetica.* 50: 177-187.

Neale D B and Williams C G. (1991). Restriction Fragment Length Polymorphism Mapping in Conifers and Applications to Forest Genetics and Tree Improvement. *Canadian Journal of Forest Research.* 21: 545-554.

Nei M. (1972). Genetic Distance Between Populations. *American Naturalist.* 106: 283-292.

Owen D L and van der Zel D W. (2000). Trees, Forests and Plantations in Southern Africa. *In:* South African Forestry Handbook. Owen D L (Editor). South African Institute of Forestry, Pretoria. 1: 3-8.

Page R D M. (1996). TREEVIEW: An Application to Display Phylogenetic Trees on Personal Computers. *Computer Applications in the Biosciences.* 12: 357-358.

Parker P G, Snow A A, Schug M D, Booton G C and Fuerst P A. (1998). What Molecules Can Tell Us About Populations: Choosing and Using a Molecular Marker. *Ecology.* 79: 361–382.

Playford J, Bell J C and Moran G F. (1993). A Major Disjunction in Genetic Diversity over the Geographic Range of *Acacia melanoxylon. Australian Journal of Botany.* 41: 355-368.

Plessis S, Buys M H and Nel M. (1999). Optimised DNA Isolation from *Acacia karroo* (Fabaceae). *South African Journal of Botany.* 65: 437.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S and Rafalski A. (1996). The Comparison of RFLP, RAPD, AFLP and SSR (microsatellite) Markers for Germplasm Analysis. *Molecular Breeding.* 2: 225-238.

Ross J H. (1972). The Acacia Species of Natal. Second Edition. The Natal Branch of the Wildlife Protection and Conservation Society of South Africa.

Sambrook J, Fritsch E F and Maniatis T. (1989). Molecular Cloning - A Laboratory Manual. Second Edition. Cold Spring Harbour Laboratory Press, New York.

Schneider S, Roessli D and Excoffier L. (2000). Arlequin Version. 2000: A Software For Population Genetic Analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Sherry S P. (1971). The Black Wattle. University of Natal Press, Pietermaritzburg.

Shrestha M K, Goldhirsh A G and Ward D. (2002). Population Genetic Structure and the Conservation of Isolated Populations of *Acacia raddiana* in Negev Desert. *Biological Conservation.* 108: 119-127.

Smith C W and Dunlop R W. (2002). Plantation Establishment, Re-establishment and Regenertaion. *In:* Black Wattle: The South African Research Experience. Dunlop R W and MacLennan L A (Editors). Insititute for Commercial Forestry Research. University of Natal Press, Pietermaritzburg.

Smith T B and Wayne R K. (1996). Molecular Genetic Approaches in Conservation. Oxford University Press.

Sneath P H A and Sokal R R. (1973). Numerical Taxonomy. W H Freeman and Company, San Francisco.

Soltis D and Soltis P. (1990). Isozymes and Analyses of Genetic Structure of Plant Populations: *In:* Isozymes in Plant Biology. Soltis D (Editor). Discorides Press, Portland Oregon. 87-105.

Tanksley S D and Orton T J. (1983). Isozymes in Plant Genetics and Breeding. Amsterdam, Elsevier.

Tanksley S D, Young N D, Patterson A H and Bonierbale M W. (1989). RFLP Mapping in Plant Breeding: New Tools for an Old Science. *Biotechnology.* 7: 257-263.

Tingey S V and del Tufo J P. (1993). Genetic Analysis with Random Amplified Polymorphic DNA Markers. *Plant Physiology.* 101: 349-352.

Troggio M, Kubisiak T L and Bucci G. (2001). Randomly Amplified Polymorphic DNA Linkage Relationships in Different Norway Spruce Populations. *Canadian Journal of Forest Research.* 31(8): 1456-1461.

Trotter P C. (1990). Biotechnology in the Pulp and Paper Industry: A Review - Tree Improvement, Pulping and Bleaching, and Dissolving Pulp Applications. *Tappi Journal.* 73(4): 198-204.

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

Welsh J and McClelland M. (1990). Fingerprinting Genomes Using PCR with Arbitrary Primers. *Nucleic Acids Research.* 18: 7213-7218.

Wheeler N and Jech K. (1992). The Use of Electrophoretic Markers in Seed Orchard Research. *Canadian Journal of Forest Research.* 27: 311–328.

Wickneswari R and Norwati M. (1993). Genetic Diversity of Natural Populations of *Acacia auriculiformis. Australian Journal of Botany.* 41: 65-77.

Williams J G K, Kubelik A R, Livak K J, Rafalski J A and Tingey S V. (1990). DNA Polymorphisms Amplified by Arbitrary Primers are Useful as Genetics Markers. *Nucleic Acids Research.* 18(2): 6531-6535.

Winter P and Kahl G. (1995). Molecular Marker Technologies for Plant Improvement. *World Journal of Microbiology and Biotechnology.* 11(4): 438-448.

Wright S. (1978). Evolution and the Genetics of Populations: Variability Within and Among Natural Populations. University of Chicago Press, Chicago.

Yeh F C, Yang R C and Boyle T (1999). POPGENE, Version 1.31. Microsoft Windows-Based Software for Population Genetics Analysis. University of Alberta and Centre for International Forestry Research, Alberta,Canada.

APPENDIX A

Laboratory Solution Recipes

2% CTAB isolation buffer (500 ml)

10 g CTAB (2%)

40.9 g NaCl (1.4 M)

3072 g EDTA (20 mM)

50 ml Tris-HCl (100 mM)

5 g PVP (1%)

Add 1ml of 2-mercapto ethanol (0.2%) after autoclaving

70% Ethanol (500 ml)

150 ml distilled water

350 ml absolute ethanol (99.5%)

Chloroform: isoamyl alcohol (24:1) (500 ml)

480ml chloroform

20ml isoamyl alcohol

Store in light sensitive bottle

10% CTAB (100 ml)

4.1 g NaCl (0.7 M)

10 g CTAB

100 ml distilled water

Autoclave

1 × TE (100 ml)

10ml 10 × TE

90 ml distilled water

Autoclave

5 × TBE buffer (1 litre)

54 g Tris base

27.5 g boric acid

20 ml EDTA (0.5 M-pH 8.0)

Autoclave

1 × TBE (1 litre)

100 ml TBE (5 ×)

900 ml distilled water

Autoclave

0.5 M EDTA (pH 8.0)

73.0625 g EDTA

300 ml distilled water

Adjust with NaOH pellets to pH 8.0

EDTA should dissolve once pH is adjusted.

Autoclave

50 × TAE (1 litre)

242 g Tris base

57.1 ml glacial acetic acid

100 ml EDTA (0.5 M-pH 8.0)

Autoclave

10 × TE (pH 7.5) (500 ml)

6.055 g Tris (100 mM)

1.86 g EDTA (100 mM)

Autoclave

Tris-HCl (pH 8.0) (500 ml)

60.55 g Tris

500 ml distilled water

Autoclave

1 M NaCl (500 ml)

29.22 g NaCl

500 ml distilled water

Autoclave

EtBr (10 mg / ml)

1 g ethidium bromide

100 ml distilled water

Loading buffer

0.0125 g bromophenol blue

0.0125 g xylene cyanol

1.5 ml glycerol

3.5 ml distilled water

Agarose gel (0.8 %)

0.4 g agarose

50 ml 1 × TAE

2.5 µl EtBr (10 mg / ml)

Agarose gel (1.0 %)

2.75 g agarose

250 ml 1 × TB`E

12 µl EtBr (10 mg / ml)

APPENDIX B

INPUT DATA FILES

POPGENE

Group 47

/* Diploid RAPD Data Set */

Number of populations = 5

Number of loci = 91

Locus name :

A02-1 A02-2 A02-3 A02-4 A02-5 A02-6 A02-7 A02-8 A02-9 A02-10 A02-11 A02-12 A02-13

A06-1 A06-2 A06-3 A06-4 A06-5 A06-6 A06-7

A09-1 A09-2 A09-3 A09-4 A09-5 A09-6 A09-7 A09-8

A10-1 A10-2 A10-3 A10-4 A10-5 A10-6 A10-7 A10-8 A10-9 A10-10 A10-11 A10-12

A13-1 A13-2 A13-3 A13-4 A13-5 A13-6 A13-7 A13-8 A13-9 A13-10 A13-11

A20-1 A20-2 A20-3 A20-4 A20-5 A20-6 A20-7 A20-8 A20-9

C06-1 C06-2 C06-3 C06-4 C06-5 C06-6 C06-7 C06-8 C06-9

C05-1 C05-2 C05-3 C05-4 C05-5 C05-6 C05-7 C05-8 C05-9 C05-10 C05-11 C05-12

P17-1 P17-2 P17-3 P17-4 P17-5 P17-6 P17-7 P17-8 P17-9 P17-10

name = Aone

fis = 0.0

0101101111100 0001111 11011001 000100010010 01000100110 010111111 111101111 111010101000 0000000101

name = Atwo

fis = 0.0

1101101110110 0001111 01011001 000000000010 00000101110 100110011 111101111 011111111101 1111101111

name = Athree

fis = 0.0

0101001111011 0001101 ..... 101011010101 11000100110 000111110 101101111

....

name = Afour

fis = 0.0

0001111111011 0001111 00111001 000100110010 01010101110 010111110

111101011 111100011000 0000010101

name = Afive

fis = 0.0

0001011111011 0101011 11011011 100011110011 01000101110 010101111

101101111 011100101101 1111101101

Group 50

/* Diploid RAPD Data Set */

Number of populations = 5

Number of loci = 91

Locus name :

A02-1 A02-2 A02-3 A02-4 A02-5 A02-6 A02-7 A02-8 A02-9 A02-10 A02-11 A02-12 A02-13

A06-1 A06-2 A06-3 A06-4 A06-5 A06-6 A06-7

A09-1 A09-2 A09-3 A09-4 A09-5 A09-6 A09-7 A09-8

A10-1 A10-2 A10-3 A10-4 A10-5 A10-6 A10-7 A10-8 A10-9 A10-10 A10-11 A10-12

A13-1 A13-2 A13-3 A13-4 A13-5 A13-6 A13-7 A13-8 A13-9 A13-10 A13-11

A20-1 A20-2 A20-3 A20-4 A20-5 A20-6 A20-7 A20-8 A20-9

C06-1 C06-2 C06-3 C06-4 C06-5 C06-6 C06-7 C06-8 C06-9

C05-1 C05-2 C05-3 C05-4 C05-5 C05-6 C05-7 C05-8 C05-9 C05-10 C05-11 C05-12

P17-1 P17-2 P17-3 P17-4 P17-5 P17-6 P17-7 P17-8 P17-9 P17-10

name = Bone

fis = 0.0

1111111111011 1011011 01111001 100011010111 11001101110 110111111 111101111 111101000001 1111010101

name = Btwo

fis = 0.0

0111011111011 0111011 11111001 100001011011 11000110110 110111111 111101111 111101101100 1111101101

name = Bthree

fis = 0.0

0111011101011 0111011 ..... 100001011010 11000110110 110111110 111101111 ....

name = Bfour

fis = 0.0

0011011111111 0001101 ..... 001001100001 01000101110 010101110 101101111 ....

name = Bfive

fis = 0.0

0011011111111 0001101 10011001 001001010111 11100110111 010111111 111100111 001010000000 0000010101

Group 85

/* Diploid RAPD Data Set */

Number of populations = 5

Number of loci = 91

Locus name :

A02-1 A02-2 A02-3 A02-4 A02-5 A02-6 A02-7 A02-8 A02-9 A02-10 A02-11 A02-12 A02-13

A06-1 A06-2 A06-3 A06-4 A06-5 A06-6 A06-7

A09-1 A09-2 A09-3 A09-4 A09-5 A09-6 A09-7 A09-8

A10-1 A10-2 A10-3 A10-4 A10-5 A10-6 A10-7 A10-8 A10-9 A10-10 A10-11 A10-12

A13-1 A13-2 A13-3 A13-4 A13-5 A13-6 A13-7 A13-8 A13-9 A13-10 A13-11

A20-1 A20-2 A20-3 A20-4 A20-5 A20-6 A20-7 A20-8 A20-9

C06-1 C06-2 C06-3 C06-4 C06-5 C06-6 C06-7 C06-8 C06-9

C05-1 C05-2 C05-3 C05-4 C05-5 C05-6 C05-7 C05-8 C05-9 C05-10 C05-11 C05-12

P17-1 P17-2 P17-3 P17-4 P17-5 P17-6 P17-7 P17-8 P17-9 P17-10

name = Cone

fis = 0.0

0011011111100 0011010 ..... 111101001111 11000001011 010111110 111111111 ....

name = Ctwo

fis = 0.0

0011011111100 0001111 01011000 001001001111 11000110111 010101110 101101011 1110100101

name = Cthree

fis = 0.0

0011011111100 0011111 ..... 001000000010 11000101111 010111111 001101111 ....

name = Cfour
fis = 0.0
0001001111100 0001111 11011100 001000100010 11000000111 001110110
001110111 010110101100 1111101101

name = Cfive
fis = 0.0
0001001111100 0001111 10011100 011101100011 11000110110 010101100
101111111 010110101101 0100100101

Group 88

/* Diploid RAPD Data Set */
Number of populations = 5
Number of loci = 91
Locus name :
A02-1 A02-2 A02-3 A02-4 A02-5 A02-6 A02-7 A02-8 A02-9 A02-10 A02-11 A02-12 A02-13
A06-1 A06-2 A06-3 A06-4 A06-5 A06-6 A06-7
A09-1 A09-2 A09-3 A09-4 A09-5 A09-6 A09-7 A09-8
A10-1 A10-2 A10-3 A10-4 A10-5 A10-6 A10-7 A10-8 A10-9 A10-10 A10-11 A10-12
A13-1 A13-2 A13-3 A13-4 A13-5 A13-6 A13-7 A13-8 A13-9 A13-10 A13-11
A20-1 A20-2 A20-3 A20-4 A20-5 A20-6 A20-7 A20-8 A20-9
C06-1 C06-2 C06-3 C06-4 C06-5 C06-6 C06-7 C06-8 C06-9
C05-1 C05-2 C05-3 C05-4 C05-5 C05-6 C05-7 C05-8 C05-9 C05-10 C05-11 C05-12
P17-1 P17-2 P17-3 P17-4 P17-5 P17-6 P17-7 P17-8 P17-9 P17-10

name = Done
fis = 0.0
0011110100011 1001000 01011001 101010000100 11100010110 001010000
001100111 011111110001 0001011101

name = Dtwo

fis = 0.0

0011010011111 0001001 01001101 101011100000 11100110110 010111111 101100011 000011001100 0110100101

name = Dthree

fis = 0.0

0011011100000 0001000 01011000 000000000010 00000000110 000110110 011100111 000000110000 .....

name = Dfour

fis = 0.0

0111001011111 0001111 00011000 100011110011 01001101111 110111111 111101111 111101000010 0010011111

name = Dfive

fis = 0.0

0011001111111 0001001 ..... 100001100110 01001110110 010110110 011100011 ....

Group 89

/* Diploid RAPD Data Set */

Number of populations = 5

Number of loci = 91

Locus name :

A02-1 A02-2 A02-3 A02-4 A02-5 A02-6 A02-7 A02-8 A02-9 A02-10 A02-11 A02-12 A02-13

A06-1 A06-2 A06-3 A06-4 A06-5 A06-6 A06-7

A09-1 A09-2 A09-3 A09-4 A09-5 A09-6 A09-7 A09-8

A10-1 A10-2 A10-3 A10-4 A10-5 A10-6 A10-7 A10-8 A10-9 A10-10 A10-11 A10-12

A13-1 A13-2 A13-3 A13-4 A13-5 A13-6 A13-7 A13-8 A13-9 A13-10 A13-11

A20-1 A20-2 A20-3 A20-4 A20-5 A20-6 A20-7 A20-8 A20-9

C06-1 C06-2 C06-3 C06-4 C06-5 C06-6 C06-7 C06-8 C06-9

C05-1 C05-2 C05-3 C05-4 C05-5 C05-6 C05-7 C05-8 C05-9 C05-10 C05-11 C05-12

P17-1 P17-2 P17-3 P17-4 P17-5 P17-6 P17-7 P17-8 P17-9 P17-10

name = Eone

fis = 0.0

0111001011111 0001111 00011101 000010011011 01000100111 010111111 111100111 001011001100 1110101111

name = Etwo

fis = 0.0

0011101111111 0001111 01011001 001100000010 01000000110 001110000 011101101 101001000000 0000000111

name = Ethree

fis = 0.0

1111001110011 0001011 00011001 100001010011 01001111111 110111111 111101111 001001010110 1110101111

name = Efour

fis = 0.0

0101111100000 0001010 00011000 000000000010 00010000110 100110001 011110111 111101000010 0010001111

name = Efive

fis = 0.0

0101001111100 0001011 00011100 100011011101 01000100110 000111111 101101111 001011001110 0000101111

Group 91

/* Diploid RAPD Data Set */
Number of populations = 5
Number of loci = 91
Locus name :
A02-1 A02-2 A02-3 A02-4 A02-5 A02-6 A02-7 A02-8 A02-9 A02-10 A02-11 A02-12 A02-13
A06-1 A06-2 A06-3 A06-4 A06-5 A06-6 A06-7
A09-1 A09-2 A09-3 A09-4 A09-5 A09-6 A09-7 A09-8
A10-1 A10-2 A10-3 A10-4 A10-5 A10-6 A10-7 A10-8 A10-9 A10-10 A10-11 A10-12
A13-1 A13-2 A13-3 A13-4 A13-5 A13-6 A13-7 A13-8 A13-9 A13-10 A13-11
A20-1 A20-2 A20-3 A20-4 A20-5 A20-6 A20-7 A20-8 A20-9
C06-1 C06-2 C06-3 C06-4 C06-5 C06-6 C06-7 C06-8 C06-9
C05-1 C05-2 C05-3 C05-4 C05-5 C05-6 C05-7 C05-8 C05-9 C05-10 C05-11 C05-12
P17-1 P17-2 P17-3 P17-4 P17-5 P17-6 P17-7 P17-8 P17-9 P17-10

name = Fone
fis = 0.0
0011001100000 0001010 01010011 000100100110 00000000110 010111100
101101011 ....

name = Ftwo
fis = 0.0
0101001111111 0101011 01011101 101011000111 11000100110 000101100
101111111 001011100111 0100101101

name = Fthree
fis = 0.00
0011001100000 0001110 01010110 001100100110 00000000110 010111100
101111111 001111100110 1100101111

name = Ffour
fis = 0.0
0001001111011 0101001 01011101 100011110010 01000101110 000111100
111101111 001001000101 0100101101

name = Ffive
fis = 0.0
1011001111101 0101111 01010100 001100110010 01000000110 010111100
101111111 001111000110 0100101111

Group 400

/* Diploid RAPD Data Set */
Number of populations = 4
Number of loci = 91
Locus name :
A02-1 A02-2 A02-3 A02-4 A02-5 A02-6 A02-7 A02-8 A02-9 A02-10 A02-11 A02-12 A02-13
A06-1 A06-2 A06-3 A06-4 A06-5 A06-6 A06-7
A09-1 A09-2 A09-3 A09-4 A09-5 A09-6 A09-7 A09-8
A10-1 A10-2 A10-3 A10-4 A10-5 A10-6 A10-7 A10-8 A10-9 A10-10 A10-11 A10-12
A13-1 A13-2 A13-3 A13-4 A13-5 A13-6 A13-7 A13-8 A13-9 A13-10 A13-11
A20-1 A20-2 A20-3 A20-4 A20-5 A20-6 A20-7 A20-8 A20-9
C06-1 C06-2 C06-3 C06-4 C06-5 C06-6 C06-7 C06-8 C06-9
C05-1 C05-2 C05-3 C05-4 C05-5 C05-6 C05-7 C05-8 C05-9 C05-10 C05-11 C05-12
P17-1 P17-2 P17-3 P17-4 P17-5 P17-6 P17-7 P17-8 P17-9 P17-10

name = Gone
fis = 0.0
..... 0011010 ..... 000000111010 11000001110 000110110 000010011 .....
.....

name = Gtwo

fis = 0.0

..... 0010010 ..... 000000000010 00000001111 000110100 101101011 .....

.....

name = Gthree

fis = 0.0

..... 0011011 ..... 000000000010 00000001110 000110000 101101011 .....

.....

name = Gfour

fis = 0.0

..... 0001011 ..... 000000101010 00000001101 000110100 101101011 .....

.....

Between Group Analysis

/* Diploid RAPD Data Set */

Number of populations = 7

Number of loci = 91

Locus name :

A02-1 A02-2 A02-3 A02-4 A02-5 A02-6 A02-7 A02-8 A02-9 A02-10 A02-11 A02-12 A02-13

A06-1 A06-2 A06-3 A06-4 A06-5 A06-6 A06-7

A09-1 A09-2 A09-3 A09-4 A09-5 A09-6 A09-7 A09-8

A10-1 A10-2 A10-3 A10-4 A10-5 A10-6 A10-7 A10-8 A10-9 A10-10 A10-11 A10-12

A13-1 A13-2 A13-3 A13-4 A13-5 A13-6 A13-7 A13-8 A13-9 A13-10 A13-11

A20-1 A20-2 A20-3 A20-4 A20-5 A20-6 A20-7 A20-8 A20-9

C06-1 C06-2 C06-3 C06-4 C06-5 C06-6 C06-7 C06-8 C06-9

C05-1 C05-2 C05-3 C05-4 C05-5 C05-6 C05-7 C05-8 C05-9 C05-10 C05-11 C05-12

P17-1 P17-2 P17-3 P17-4 P17-5 P17-6 P17-7 P17-8 P17-9 P17-10

name = A

fis = 0.0

0101101111100 0001111 11011001 000100010010 01000100110 010111111
111101111 111010101000 0000000101
1101101110110 0001111 01011001 000000000010 00000101110 100110011
111101111 011111111101 1111101111
0101001111011 0001101 ..... 101011010101 11000100110 000111110 101101111
....
0001111111011 0001111 00111001 000100110010 01010101110 010111110
111101011 111100011000 0000010101
0001011111011 0101011 11011011 100011110011 01000101110 010101111
101101111 011100101101 1111101101

name = B

fis = 0.0

1111111111011 1011011 01111001 100011010111 11001101110 110111111
111101111 111101000001 1111010101
0111011111011 0111011 11111001 100001011011 11000110110 110111111
111101111 111101101100 1111101101
0111011101011 0111011 ..... 100001011010 11000110110 110111110 111101111
....
0011011111111 0001101 ..... 001001100001 01000101110 010101110 101101111
....
0011011111111 0001101 10011001 001001010111 11100110111 010111111
111100111 001010000000 0000010101

name = C

fis = 0.0

0011011111100 0011010 ..... 111101001111 11000001011 010111110 111111111
....
0011011111100 0001111 01011000 001001001111 11000110111 010101110
101101011 1110100101

0011011111100 0011111 ..... 001000000010 11000101111 010111111 001101111
....
0001001111100 0001111 11011100 001000100010 11000000111 001110110
001110111 010110101100 1111101101
0001001111100 0001111 10011100 011101100011 11000110110 010101100
101111111 010110101101 0100100101

name = D
fis = 0.0
0011110100011 1001000 01011001 101010000100 11100010110 001010000
001100111 011111110001 0001011101
0011010011111 0001001 01001101 101011100000 11100110110 010111111
101100011 000011001100 0110100101
0011011100000 0001000 01011000 000000000010 00000000110 000110110
011100111 000000110000 .....
0111001011111 0001111 00011000 100011110011 01001101111 110111111
111101111 111101000010 0010011111
0011001111111 0001001 ..... 100001100110 01001110110 010110110 011100011
....

name = E
fis = 0.0
0111001011111 0001111 00011101 000010011011 01000100111 010111111
111100111 001011001100 1110101111
0011101111111 0001111 01011001 001100000010 01000000110 001110000
011101101 101001000000 0000000111
1111001110011 0001011 00011001 100001010011 01001111111 110111111
111101111 001001010110 1110101111
0101111100000 0001010 00011000 000000000010 00010000110 100110001
011110111 111101000010 0010001111
0101001111100 0001011 00011100 100011011101 01000100110 000111111
101101111 001011001110 0000101111

name = F

fis = 0.0

0011001100000 0001010 01010011 000100100110 00000000110 010111100 101101011 ....

0101001111111 0101011 01011101 101011000111 11000100110 000101100 101111111 001011100111 0100101101

0011001100000 0001110 01010110 001100100110 00000000110 010111100 101111111 001111100110 1100101111

0001001111011 0101001 01011101 100011110010 01000101110 000111100 111101111 001001000101 0100101101

1011001111101 0101111 01010100 001100110010 01000000110 010111100 101111111 001111000110 0100101111

name = G

fis = 0.0

..... 0011010 ..... 000000111010 11000001110 000110110 000010011 ....

..... 0010010 ..... 000000000010 00000001111 000110100 101101011 ....

..... 0011011 ..... 000000000010 00000001110 000110000 101101011 ....

..... 0001011 ..... 000000101010 00000001101 000110100 101101011 ....

Between Individual Analysis

/* Diploid RAPD Data Set */

Number of populations = 34

Number of loci = 91

Locus name :

A02-1 A02-2 A02-3 A02-4 A02-5 A02-6 A02-7 A02-8 A02-9 A02-10 A02-11 A02-12 A02-13

q

A06-1 A06-2 A06-3 A06-4 A06-5 A06-6 A06-7

A09-1 A09-2 A09-3 A09-4 A09-5 A09-6 A09-7 A09-8

A10-1 A10-2 A10-3 A10-4 A10-5 A10-6 A10-7 A10-8 A10-9 A10-10 A10-11 A10-12

A13-1 A13-2 A13-3 A13-4 A13-5 A13-6 A13-7 A13-8 A13-9 A13-10 A13-11

A20-1 A20-2 A20-3 A20-4 A20-5 A20-6 A20-7 A20-8 A20-9

C06-1 C06-2 C06-3 C06-4 C06-5 C06-6 C06-7 C06-8 C06-9

C05-1 C05-2 C05-3 C05-4 C05-5 C05-6 C05-7 C05-8 C05-9 C05-10 C05-11 C05-12

P17-1 P17-2 P17-3 P17-4 P17-5 P17-6 P17-7 P17-8 P17-9 P17-10

name = Aone

fis = 0.0

01011011111100 0001111 11011001 000100010010 01000100110 010111111
111101111 111010101000 0000000101

name = Atwo

fis = 0.0

1101101110110 0001111 01011001 000000000010 00000101110 100110011
111101111 011111111101 1111101111

name = Athree

fis = 0.0

0101001111011 0001101 ..... 101011010101 11000100110 000111110 101101111
....

name = Afour

fis = 0.0

0001111111011 0001111 00111001 000100110010 01010101110 010111110
111101011 111100011000 0000010101

name = Afive

fis = 0.0

0001011111011 0101011 11011011 100011110011 01000101110 010101111
101101111 011100101101 1111101101

r

name = Bone

fis = 0.0

1111111111011 1011011 01111001 100011010111 11001101110 110111111 111101111 111101000001 1111010101

name = Btwo

fis = 0.0

0111011111011 0111011 11111001 100001011011 11000110110 110111111 111101111 111101101100 1111101101

name = Bthree

fis = 0.0

0111011101011 0111011 ..... 100001011010 11000110110 110111110 111101111 ....

name = Bfour

fis = 0.0

0011011111111 0001101 ..... 001001100001 01000101110 010101110 101101111 ....

name = Bfive

fis = 0.0

0011011111111 0001101 10011001 001001010111 11100110111 010111111 111100111 001010000000 0000010101

name = Cone

fis = 0.0

0011011111100 0011010 ..... 111101001111 11000001011 010111110 111111111 ....

name = Ctwo
fis = 0.0
0011011111100 0001111 01011000 001001001111 11000110111 010101110
101101011 1110100101

name = Cthree
fis = 0.0
0011011111100 0011111 ..... 001000000010 11000101111 010111111 001101111
....

name = Cfour
fis = 0.0
0001001111100 0001111 11011100 001000100010 11000000111 001110110
001110111 010110101100 1111101101

name = Cfive
fis = 0.0
0001001111100 0001111 10011100 011101100011 11000110110 010101100
101111111 010110101101 0100100101

name = Done
fis = 0.0
0011110100011 1001000 01011001 101010000100 11100010110 001010000
001100111 011111110001 0001011101

name = Dtwo
fis = 0.0
0011010011111 0001001 01001101 101011100000 11100110110 010111111
101100011 000011001100 0110100101

t

name = Dthree

fis = 0.0

0011011100000 0001000 01011000 000000000010 00000000110 000110110

011100111 000000110000 .....

name = Dfour

fis = 0.0

0111001011111 0001111 00011000 100011110011 01001101111 110111111

111101111 111101000010 0010011111

name = Dfive

fis = 0.0

0011001111111 0001001 ..... 100001100110 01001110110 010110110 011100011

....

name = Eone

fis = 0.0

0111001011111 0001111 00011101 000010011011 01000100111 010111111

111100111 001011001100 1110101111

name = Etwo

fis = 0.0

0011101111111 0001111 01011001 001100000010 01000000110 001110000

011101101 101001000000 0000000111

name = Ethree

fis = 0.0

1111001110011 0001011 00011001 100001010011 01001111111 110111111

111101111 001001010110 1110101111

u

name = Efour
fis = 0.0
0101111100000 0001010 00011000 000000000010 00010000110 100110001
011110111 111101000010 0010001111

name = Efive
fis = 0.0
0101001111100 0001011 00011100 100011011101 01000100110 000111111
101101111 001011001110 0000101111

name = Fone
fis = 0.0
0011001100000 0001010 01010011 000100100110 00000000110 010111100
101101011 ....

name = Ftwo
fis = 0.0
0101001111111 0101011 01011101 101011000111 11000100110 000101100
101111111 001011100111 0100101101

name = Fthree
fis = 0.00
0011001100000 0001110 01010110 001100100110 00000000110 010111100
101111111 001111100110 1100101111

name = Ffour
fis = 0.0
0001001111011 0101001 01011101 100011110010 01000101110 000111100
111101111 001001000101 0100101101

name = Ffive

fis = 0.0

1011001111101 0101111 01010100 001100110010 01000000110 010111100 101111111 001111000110 0100101111

name = Gone

fis = 0.0

..... 0011010 ..... 000000111010 11000001110 000110110 000010011 ....

name = Gtwo

fis = 0.0

..... 0010010 ..... 000000000010 00000001111 000110100 101101011 ....

name = Gthree

fis = 0.0

..... 0011011 ..... 000000000010 00000001110 000110000 101101011 ....

name = Gfour

fis = 0.0

..... 0001011 ..... 000000101010 00000001101 000110100 101101011 ....

APPENDIX C

Popgene Analysis

Within group analysis:

Group 47

Summary Statistics:

Locus	Sample Size	h*	I*
Mean	5	0.2413	0.3508
St. Dev		0.2141	0.3055

* h = Nei's (1973) gene diversity

* I = Shannon's Information index [Lewontin (1972)]

Locus	Sample Size	Ht	Hs	Gst
Mean	5	0.1246	0.0000	1.0000
St. Dev		0.0782	0.0000	

The number of polymorphic loci is : 53

The percentage of polymorphic loci is : 58.24

pop ID	1	2	3	4	5
1	****	0.7253	0.6264	0.8022	0.6813
2	0.3212	****	0.6813	0.6374	0.6923
3	0.4678	0.3837	****	0.6044	0.7473
4	0.2204	0.4504	0.5035	****	0.6813
5	0.3837	0.3677	0.2914	0.3837	****

Group 50

Summary Statistics:

Locus	Sample Size	h*	I*
Mean	4	0.2229	0.3244
St. Dev		0.2133	0.3067

* h = Nei's (1973) gene diversity

* I = Shannon's Information index [Lewontin (1972)]

Locus	Sample Size	Ht	Hs	Gst
Mean	4	-0.1629	0.0000	1.0000
St. Dev		0.3696	0.0000	

The number of polymorphic loci is : 49

The percentage of polymorphic loci is : 53.85

pop ID	1	2	3	4	5
1	****	0.8022	0.7253	0.6374	0.6593
2	0.2204	****	0.8791	0.7033	0.6593
3	0.3212	0.1288	****	0.7802	0.5824
4	0.4504	0.3520	0.2482	****	0.6264
5	0.4165	0.4165	0.5406	0.4678	****

y

Group 85

Summary Statistics :

Locus	Sample Size	h*	I*
Mean	4	0.1698	0.2464
St. Dev		0.2115	0.3036

* h = Nei's (1973) gene diversity

* I = Shannon's Information index [Lewontin (1972)]

Locus	Sample Size	Ht	Hs	Gst
Mean	4	-0.5756	0.0000	1.0000
St. Dev		2.2268	0.0000	

The number of polymorphic loci is : 37

The percentage of polymorphic loci is : 40.66

pop ID	1	2	3	4	5
1	****	0.7473	0.8352	0.6484	0.6484
2	0.2914	****	0.7802	0.7033	0.7692
3	0.1801	0.2482	****	0.7473	0.6593
4	0.4333	0.3520	0.2914	****	0.7802
5	0.4333	0.2624	0.4165	0.2482	****

z

Group 88

Summary Statistics :

Locus	Sample Size	h*	I*
Mean	5	0.3031	0.4399
St. Dev		0.1975	0.2802

* h = Nei's (1973) gene diversity

* I = Shannon's Information index [Lewontin (1972)]

Locus	Sample Size	Ht	Hs	Gst
Mean	5	0.1171	0.0000	1.0000
St. Dev		0.1856	0.0000	

The number of polymorphic loci is : 66

The percentage of polymorphic loci is : 72.53

pop ID	1	2	3	4	5
1	****	0.6044	0.6593	0.4725	0.5934
2	0.5035	****	0.5714	0.6044	0.6593
3	0.4165	0.5596	****	0.5495	0.6484
4	0.7497	0.5035	0.5988	****	0.6374
5	0.5219	0.4165	0.4333	0.4504	****

Group 89

Summary Statistics :

Locus	Sample Size	h*	I*
Mean	5	0.2620	0.3815
St. Dev		0.2086	0.2974

* h = Nei's (1973) gene diversity

* I = Shannon's Information index [Lewontin (1972)]

Locus	Sample Size	Ht	Hs	Gst
Mean	5	0.2620	0.0000	1.0000
St. Dev		0.0435	0.0000	

The number of polymorphic loci is : 58

The percentage of polymorphic loci is : 63.74

pop ID	1	2	3	4	5
1	****	0.6593	0.7912	0.5824	0.8022
2	0.4165	****	0.6044	0.7033	0.6154
3	0.2342	0.5035	****	0.6154	0.7253
4	0.5406	0.3520	0.4855	****	0.6264
5	0.2204	0.4855	0.3212	0.4678	****

Group 91

Summary Statistics :

Locus	Sample Size	h*	I*
Mean	5	0.1943	0.2796
St. Dev		0.2214	0.3150

* h = Nei's (1973) gene diversity

* I = Shannon's Information index [Lewontin (1972)]

Locus	Sample Size	Ht	Hs	Gst
Mean	5	0.1147	0.0000	1.0000
St. Dev		0.0886	0.0000	

The number of polymorphic loci is : 41

The percentage of polymorphic loci is : 45.05

pop ID	1	2	3	4	5
1	****	0.6044	0.8462	0.6044	0.7033
2	0.5035	****	0.6923	0.8242	0.7473
3	0.1671	0.3677	****	0.6484	0.8571
4	0.5035	0.1934	0.4333	****	0.7473
5	0.3520	0.2914	0.1542	0.2914	****

Group 400

Summary Statistics :

Locus	Sample Size	h*	I*
Mean	4	-	0.2101
St. Dev		-	0.2886

* h = Nei's (1973) gene diversity

* I = Shannon's Information index [Lewontin (1972)]

pop ID	1	2	3	4
1	*****	0.8571	0.8571	0.8571
2	0.1542	*****	0.9560	0.9341
3	0.1542	0.0450	*****	0.9341
4	0.1542	0.0682	0.0682	*****

Between group analysis:

Population ID : 1

Population name : A

Summary Statistics :

Locus	Sample Size	h*	I*
Mean	5	0.2246	0.3308
St. Dev		0.2086	0.2972

* h = Nei's (1973) gene diversity

* I = Shannon's Information index [Lewontin (1972)]

The number of polymorphic loci is : 53

The percentage of polymorphic loci is : 58.24 %

Population ID : 2

Population name : B

Summary Statistics :

Locus	Sample Size	h*	I*
Mean	4	0.2101	0.3082
St. Dev		0.2128	0.3023

* h = Nei's (1973) gene diversity

* I = Shannon's Information index [Lewontin (1972)]

The number of polymorphic loci is : 49

The percentage of polymorphic loci is : 53.85 %

Population ID : 3

Population name : C

Summary Statistics :

Locus	Sample Size	h*	I*
Mean	4	0.1667	0.2421
St. Dev		0.2124	0.3028

* h = Nei's (1973) gene diversity

* I = Shannon's Information index [Lewontin (1972)]

The number of polymorphic loci is : 37

The percentage of polymorphic loci is : 40.66 %

Population ID : 4

Population name : D

Summary Statistics :

Locus	Sample Size	h*	I*
Mean	5	0.2683	0.3986
St. Dev		0.1935	0.2717

* h = Nei's (1973) gene diversity
* I = Shannon's Information index [Lewontin (1972)]

The number of polymorphic loci is : 66
The percentage of polymorphic loci is : 72.53 %

Population ID : 5
Population name : E

Summary Statistics :

Locus	Sample Size	h*	I*
Mean	5	0.2350	0.3492
St. Dev		0.2021	0.2866

* h = Nei's (1973) gene diversity
* I = Shannon's Information index [Lewontin (1972)]

The number of polymorphic loci is : 58
The percentage of polymorphic loci is : 63.74 %

Population ID : 6

Population name : F

Summary Statistics :

Locus	Sample Size	h*	I*
Mean	5	0.1762	0.2587
St. Dev		0.2077	0.2982

* h = Nei's (1973) gene diversity

* I = Shannon's Information index [Lewontin (1972)]

The number of polymorphic loci is : 41

The percentage of polymorphic loci is : 45.05 %

Population ID : 7

Population name : G

Summary Statistics :

Locus	Sample Size	h*	I*
Mean	4	0.1420	0.2069
St. Dev		0.2056	0.2927

* h = Nei's (1973) gene diversity

* I = Shannon's Information index [Lewontin (1972)]

The number of polymorphic loci is : 17

The percentage of polymorphic loci is : 18.68 %

Multiple Summary Statistics :

Locus	Sample Size	h*	I*
Mean	30	0.3209	0.4831
St. Dev		0.1565	0.1995

* h = Nei's (1973) gene diversity

* I = Shannon's Information index [Lewontin (1972)]

Locus	Sample Size	Ht	Hs	Gst
Mean	30	0.2399	0.1937	0.1926
St. Dev		0.0340	0.0089	

The number of polymorphic loci is : 87

The percentage of polymorphic loci is : 95.60

pop ID	1	2	3	4	5	6	7
1	****	0.8961	0.8361	0.8562	0.8798	0.8387	0.6889
2	0.1097	****	0.7782	0.8559	0.8404	0.7949	0.6402
3	0.1790	0.2508	****	0.7932	0.7800	0.8147	0.6683
4	0.1552	0.1556	0.2317	****	0.9047	0.8388	0.6379
5	0.1280	0.1739	0.2485	0.1002	****	0.8744	0.6627
6	0.1759	0.2295	0.2050	0.1758	0.1342	****	0.6708
7	0.3727	0.4459	0.4031	0.4495	0.4114	0.3993	****

AMOVA ANALYSIS:

AMOVA design and results :

Reference: Weir, B.S. and Cockerham, C.C. 1984.
Excoffier, L., Smouse, P., and Quattro, J. 1992.
Weir, B. S., 1996.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among populations	6	79.985	1.39040 Va	17.44
Within populations	27	177.750	6.58333 Vb	82.56
Total	33	257.735	7.97374	

Fixation Index FST : 0.17437

Significance tests (1023 permutations)

Va and FST : P(rand. value > obs. value) = 0.00000
P(rand. value = obs. value) = 0.00000
P(rand. value >= obs. value) = 0.00000+-0.00000

AMOVA design and results (average over 47 loci):

Source of variation	Sum of squares	Variance components
Among populations	79.985	1.39040
Within populations	177.750	6.58333
Total	257.735	7.97374

Average F-Statistics over all loci

Fixation Indices

FST : 0.17437

Significance tests (1023 permutations)

Va and FST : P(rand. value > obs. value) = 0.00000

P(rand. value = obs. value) = 0.00000

P(rand. value >= obs. value) = 0.00000

==========

==========

== ANALYSES AT THE INTRA-POPULATION LEVEL

==========

==========

==========

==========

== Sample : fortyseven pop 1

==========

==========

==========

== Molecular diversity indices : (fortyseven pop 1)

==========

Reference: Tajima, F., 1983.

Tajima, F. 1993.

Nei, M., 1987.

Zouros, E., 1979.

Ewens, W.J. 1972.

Sample size	: 5.00000
No. of haplotypes	: 5
Allowed level of missing data	: 5.00000 %
Number of observed indels	: 0
Number of polymorphic sites	: 33
Number of usable nucleotide sites	: 61

Distance method : Pairwise difference

Mean number of pairwise differences : 16.200001 +/- 8.737138

Average gene diversity over loci : 0.265574 +/- 0.167444

(Standard deviations are for both sampling and stochastic processes)

Unable to compute theta(Hom) and theta(k) when all gene copies are different

==========
== Sample : fifty pop 2
==========

==========
== Molecular diversity indices : (fifty pop 2)
==========

Reference: Tajima, F., 1983.
Tajima, F. 1993.
Nei, M., 1987.
Zouros, E., 1979.
Ewens, W.J. 1972.

Sample size : 5.000000
No. of haplotypes : 5

Allowed level of missing data : 5.000000 %
Number of observed indels : 0
Number of polymorphic sites : 30
Number of usable nucleotide sites : 61

Distance method : Pairwise difference

Mean number of pairwise differences : 14.800000 +/- 8.011333

Average gene diversity over loci : 0.242623 +/- 0.153535

(Standard deviations are for both sampling and stochastic processes)

Unable to compute theta(Hom) and theta(k) when all gene copies are different

==========

== Sample : eightyfive pop 3

==========

==========

==========

== Molecular diversity indices : (eightyfive pop 3)

==========

Reference: Tajima, F., 1983.

Tajima, F. 1993.

Nei, M., 1987.

Zouros, E., 1979.

Ewens, W.J. 1972.

Sample size : 5.000000

No. of haplotypes : 5

Allowed level of missing data : 5.000000 %

Number of observed indels : 0

Number of polymorphic sites : 29

nn

Number of usable nucleotide sites : 61

Distance method : Pairwise difference

Mean number of pairwise differences : 14.800000 +/- 8.011333

Average gene diversity over loci : 0.242623 +/- 0.153535

(Standard deviations are for both sampling and stochastic processes)

Unable to compute theta(Hom) and theta(k) when all gene copies are different

==========

== Sample : eightyeight pop 4

==========

==========

== Molecular diversity indices : (eightyeight pop 4)

==========

Reference: Tajima, F., 1983.

Tajima, F. 1993.

Nei, M., 1987.

Zouros, E., 1979.

Ewens, W.J. 1972.

Sample size : 5.000000

No. of haplotypes : 5

Allowed level of missing data : 5.000000 %

Number of observed indels : 0
Number of polymorphic sites : 43
Number of usable nucleotide sites : 61

Distance method : Pairwise difference
Mean number of pairwise differences : 22.000000 +/- 11.743187

Average gene diversity over loci : 0.360656 +/- 0.225054

(Standard deviations are for both sampling and stochastic processes)

Unable to compute theta(Hom) and theta(k) when all gene copies are different

==========
== Sample : eightynine pop 5
==========

==========
== Molecular diversity indices : (eightynine pop 5)
==========

Reference: Tajima, F., 1983.
Tajima, F. 1993.
Nei, M., 1987.
Zouros, E., 1979.
Ewens, W.J. 1972.
Sample size : 5.000000
No. of haplotypes : 5

Allowed level of missing data : 5.000000 %
Number of observed indels : 0
Number of polymorphic sites : 58
Number of usable nucleotide sites : 91

Distance method : Pairwise difference
Mean number of pairwise differences : 29.799999 +/- 15.784710

Average gene diversity over loci : 0.327473 +/- 0.202781

(Standard deviations are for both sampling and stochastic processes)

Unable to compute theta(Hom) and theta(k) when all gene copies are different

==========
== Sample : ninetyone pop 6
==========

==========
== Molecular diversity indices : (ninetyone pop 6)
==========

Reference: Tajima, F., 1983.
Tajima, F. 1993.
Nei, M., 1987.
Zouros, E., 1979.
Ewens, W.J. 1972.

Sample size : 5.000000
No. of haplotypes : 5

Allowed level of missing data : 5.000000 %
Number of observed indels : 0
Number of polymorphic sites : 34
Number of usable nucleotide sites : 69

Distance method : Pairwise difference
Mean number of pairwise differences : 18.200001 +/- 9.773833

Average gene diversity over loci : 0.263768 +/- 0.165595

(Standard deviations are for both sampling and stochastic processes)

Unable to compute theta(Hom) and theta(k) when all gene copies are different

==========
== Sample : fourhundred pop 7
==========

==========
== Molecular diversity indices : (fourhundred pop 7)
==========

Reference: Tajima, F., 1983.
Tajima, F. 1993.
Nei, M., 1987.

Zouros, E., 1979.

Ewens, W.J. 1972.

Sample size : 4.000000

No. of haplotypes : 4

Allowed level of missing data : 5.000000 %

Number of observed indels : 0

Number of polymorphic sites : 17

Number of usable nucleotide sites : 48

Distance method : Pairwise difference

Mean number of pairwise differences : 9.166667 +/- 5.354497

Average gene diversity over loci : 0.190972 +/- 0.133207