ESTIMATION OF GENETIC VARIATION AND MARKER IDENTIFICATION IN BLACK WATTLE (*Acacia mearnsii* De Wild.) WITH RAPD FINGERPRINTING

by

YAKSHA SEWPERSAD

Submitted in fulfilment of the academic

requirements of the degree of

Master of Science

in the School of Molecular and Cellular Biosciences,

University of KwaZulu-Natal

Pietermaritzburg

2004

PREFACE

The experiment work described in this dissertation was conducted at the University of 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 other is acknowledged in the text, are the results of my own investigation.

December 2004

I certify the above statement is correct.

Professor Annabel Fossey

Supervisor

ACKNOWLEDGEMENTS

I wish to sincerely thank the following people for contributing to the completion of this thesis:

- To Professor Annabel Fossey, my supervisor and friend, thank you for your expertise, understanding, and encouragement. It was an absolute honour to be working with you.

- To Rob Dunlop, my co-supervisor, thank you for your support, patience, and guidance in helping me to complete this project. It was a pleasure to be working with you.

- To Mrs Pat Wallis, thank you for your patience, advise, help, and chemicals.

- To Colleen LeRoux, thank you for your encouragement and support.

- To Tammy Swain from the Institute of Commercial Forestry Research in Sabie. Thank you for your advice and guidance

- To my friends, Kumi and Seshni, your support and encouragement throughout this degree really means a lot to me. A large part of my completion of this degree is due to you!

- To my sisters, thank you for always supporting me in whatever way you could and for always being there.

- To my parents, somehow thank you does not seem enough to express the sincere gratitude I feel towards you. You have stood by and supported me through some very difficult times and I definitely would not be where I am today without you.

- To my husband, Owen, thank you for always believing in and supporting me through some trying times. Your understanding and patience throughout this study has been remarkable and I could not have done it without you.

- Last but not least, my beautiful baby girl, Reeya. You where my motivation throughout this degree, so this one is for you!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AFLP	-	Amplified fragment length polymorphism
AMOVA	-	Analysis of molecular variance
bp	-	base pairs
CTAB	-	Hexadecyltimethylammonium bromide
DNA	-	Deoxyribonucleic Acid
DNTPs	-	Deoxyribonucleoside triphosphates
EDTA	-	Ethylenediaminetetraacetic acid
GC	-	Guanine Cytosine
ICFR	-	Institute for Commercial Forestry Research
$MgCl_2$	-	Magnesium chloride
NaCl	-	Sodium chloride
OD	-	Optical density
PCR	-	Polymerase chain reaction
RAPD	-	Random amplified polymorphic DNA
RFLP	-	Restriction fragment length polymorphism
SAWGU	-	South African Wattle Growers' Union
SDS	-	Sodium dodecyl sulphate
Taq	-	*Thermus aquaticus*
UPGMA	-	Unweighted pair-group method using arithmetic mean
UV	-	Ultra violet

ABSTRACT

Black wattle was introduced into South Africa from Australia in 1864 as a tree to be used primarily for shade for livestock, shelterbelts, windbreaks, and for fuel wood. Today black wattle is a major plantation species that contributes substantially to the South African forestry and forest products sector making up approximately 7 % of South African plantation forestry. The demand for black wattle bark products and timber has cemented both locally and internationally, its place in South African forestry. Due to the limited number of seeds that formed the genetic base of South African black wattle, it is expected that the South African population is genetically rather uniform.

The aims of this investigation were to generate RAPD fingerprints of black wattle trees in order to identify putative markers linked to the disease complex gummosis. Together with this research, an assessment of the genetic variation and genetic relatedness of the trees included in this investigation was also undertaken. As this was the first investigation of its kind, the DNA isolation method, as well as the PCR conditions needed optimization before the final fingerprints could be generated and analysed.

The plant material used in this investigation was obtained from the Institute for Commercial Forestry Research (ICFR) nursery in Pietermaritzburg, KwaZulu-Natal. Leaf samples where taken from four families, of which three where susceptible to the disease complex gummosis and one was resistant to the disease. These four families included three half-sib families and one full-sib family.

In all molecular techniques, the efficient isolation of DNA is an important step in the process and it is therefore important that suitable DNA isolation procedures are employed that yield pure and high molecular weight DNA, which is used in subsequent analyses. Furthermore, it would be advantageous if the selected procedure were quick, easy to apply and cost effective and, if possible, devoid of dangerous chemicals. The isolation of DNA from black wattle has proven to be problematic because of the presence of high levels of polysaccharides, polyphenols, and tannins in the cells, making it difficult to separate the DNA out. Thus, the first

part of this study involved the optimization of a suitable DNA isolation method that was assessed both quantitatively and qualitatively.

Three DNA isolation methods were compared; CTAB method, salting out method, and Chelex©-100 method which was a modification of the CTAB method. The CTAB method was selected as being the most effective isolation method because it produced the best quantitative and qualitative results as compared to the other two methods, which produced degraded, low purity and low concentration DNA. The CTAB method was optimized by increasing the concentration of polyvinylpolypyrrolidone (PVP) and 2-mercaptoethanol, which are reagents responsible for the binding of contaminants. The DNA concentration isolated by the CTAB isolation method ranged from 0.001 µg/ml to 0.0015 µg/ml with an average concentration of 0.001364 µg/ml and the purity ranged from 1.7 to 2.1 with an average purity of 1.9, which was within the standard range of pure DNA.

Successful amplification of DNA through PCR is species specific and dependent upon obtaining the most optimal combination of reagents and conditions. In this investigation the DNA template, primer and magnesium chloride concentrations were optimized. Several DNA concentrations, ranging from 50 ng/µl to 100 ng/µl were assessed and found that 100 ng/µl of genomic DNA gave the most consistent and brightest DNA fingerprint bands for black wattle. Likewise, the optimal primer concentration was found to be 2 µM and for $MgCl_2$ 2.5 mM. The optimization of the PCR settings to amplify DNA was undertaken by varying the annealing temperature from 29°C to 41°C. 37°C produced the most reproducible and clear fingerprint bands and was selected as the optimal annealing temperature.

RAPD fingerprints were evaluated to identify loci that were suitable to score for all subsequent analyses. Loci were selected if the bands that they produced were clear, repeatable and consistent. A total of 94 loci produced consistently clear and repeatable bands and were scored according to a binary format, where the presence of a band (presence-phenotype) was given a value of "1" and the absence of a band (null-phenotype) was awarded a value of "0", thereby producing a digital fingerprint.

The RAPD fingerprints were assessed for the presence of a band linked to the disease complex gummosis, by comparing the fingerprints of the disease resistant individuals to those of the fingerprints of the disease susceptible individuals. One such band was identified.

The performance of the primers was assessed for the number of loci that they amplified and the number of polymorphic and monomorphic loci. The number of loci amplified by the different primers ranged from five to 13. Five of the primers amplified only monomorphic loci, while the remaining nine amplified monomorphic as well as polymorphic loci. Of the 94 loci that were scored 27.7 % (26 loci) were monomorphic, while 72.3 % (68 loci) were polymorphic. It was found that the GC content of the primers did not influence the primers used in this investigation in any particular way.

Genetic variation within this group of black wattle individuals was quantified for each family individually as well as for the entire population. The quantification of genetic variation was undertaken by determining the percentage polymorphic loci, Nei's gene diversity, Shannon's Information Index and estimated heterozygosity. All these measures of variation clearly demonstrated a rather limited amount of genetic variation in this population, which was also true for each of the families. The polymorphic content of the entire population was estimated as 58.51 %. However, family 101 displayed the highest values for Nei's gene diversity (15.56 %) and Shannon's Information Index (23.64 %). As expected, the only family containing full-sibs (101x77) contained consistently the least amount of genetic variation (Nei's gene diversity = 10.64 %; Shannon's information index = 15.89 %). Second to family 101x77 was family 18, expected for a half-sib family with a small number of individuals. The two larger families, 99 and 101, both displayed nearly double the number of polymorphic loci (43.62 %; 46.81 %) than the two smaller families (28.72 % for both), probably indicating the need to evaluate larger families or more loci.

The estimated heterozygosity for the population was determined by firstly estimating q, the frequency of the recessive allele, from the frequency of the null-phenotype at each locus by computing the square root, and thereafter calculating 2pq, the estimated heterozygosity. For the entire population the estimated heterozygosity is averaged over all loci. The estimated heterozygosity of this population was low, an

average of 24 %, with the lowest percentage estimated at a particular locus of 4 % (2 loci) and the highest of 50 % (4 loci). The mean heterozygosity of the loci of the different primers individually ranged from 11 % to 39 %.

Genetic relationships between families and among individuals within families were quantified according to Nei's unbiased estimate of genetic identity (I) and distance (D). A pairwise analysis between the individuals within the different families revealed in general that all identities were high, between 71% and 99 %. The similarity ranges were narrow for the two small families; 80 % to 93 % for family 18 and 80 % to 96 % for family 101x77. The two larger families, on the other hand, displayed a larger range in similarity comparisons; 72 % to 98 % for family 99 and 71 % to 99 % for family 101.

A comparison of genetic identity between the four families also displayed extensive genetic similarity ranging from 92 % to 99 %, which was expected for each of the families; however, it was an unexpected result that the families were so closely related. These data therefore clearly support the notion that the genetic base of South African black wattle population is narrow.

An analysis of family differentiation was undertaken by conducting an analysis of molecular variance (AMOVA) by employing the software programme ARLEQUIN. The results of the AMOVA partitioning of RAPD variance among and within families revealed of the total genetic variation, 87.86 % was attributable to within family variation and only 12.14 % to between families variation in black wattle, both highly significant ($P<0.001$) thus indicating that there is a significant genetic differentiation between and within families.

Wright's fixation index (F_{ST}) is also computed to estimate population differentiation, but unlike AMOVA that estimates population differentiation directly from molecular data, this index estimates the differentiation by assuming Mendelian segregation of alleles. The calculation of Wright's fixation index (F_{ST}), confirmed the results displayed in the AMOVA analysis, revealing a significant F_{ST} value of 0.12140, which is the same as what the AMOVA estimated. According to Wright's guidelines, a value between 0.05 and 0.15 indicates great genetic differentiation between subfamilies.

The data generated in this investigation supports the notion that South African black wattle is genetically rather uniform, due to its introduction history from Australia were a limited number of seeds were brought into the country. Furthermore, these data clearly emphasize the importance of such analyses as marker-assisted selection is a major activity in general breeding programmes and the knowledge of the extent of available variation is required in order to identify and select suitable superior genotypes to use in breeding programmes.

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Plant breeding - the art and science of improving the heredity of plants and trees for the benefit of humankind has been practised for many centuries. Plant breeding today is a complex and sophisticated industry, generating vast amounts of revenue. The recent introduction of modern technologies such as, mutation breeding, tissue culture, marker-assisted selection, and various other biotechnologies has greatly facilitated the rapid growth of the industry.

Traditional plant breeding involves the selection of parental plants with desired traits, crossing, the collection, and planting of seed. The results of such crosses are then evaluated after the cultivation of the progeny (Sedjo, 1999). This is a relatively slow and labour-intensive process. It is important to note that plant breeders work with entire genomes, and when a cross is made to introduce desirable traits, one or more undesirable traits are also introduced; and these must then be painstakingly "bred out" (Helentjaris et al., 1985). The techniques of biotechnology can be used to speed up the process and improve the precision of plant breeding.

Biotechnology success in agriculture has been tremendous over the past few years (Hayami and Rultan, 1985). Techniques have been developed that allow specialised genes to be implanted in plants. These can be naturally occurring or modified genes. Recent applications of genetic engineering in agriculture have been successful in transferring to agricultural crops a host of desired traits including resistance to insects, herbicides, frost, disease, and so forth. The success of genetic approaches for increasing yields and generating desired modifications in agricultural plants is unquestionable.

Biotechnological issues in commercial forestry revolve around the scientific and technical ability to develop superior trees, using traditional breeding techniques to take advantage of the variability provided in the species. One important area is the potential to use DNA to determine desirable wood qualities. Disease tolerance is also one of the exciting prospects for the future (Mohan et al., 1997). Tools of this form could drastically improve the process of selecting disease tolerant, fast growing trees with excellent wood qualities in various tree species.

Currently, most of the world's industrial wood is drawn from commercially planted forests in what is essentially a foraging operation. In the past, harvests were taken from indigenous forests as humans simply collected the bounty of nature (Sedjo, 1999). Table 1.1 indicates progress made over time in the plantation forestry industry.

Table 1.1 Progress in tree breeding strategies over time (adapted from Sedjo, 1999).

Type	Period		
Wild forests	10,000 B.C.	-	current
Managed forests	100 B.C.	-	current
Planted forests	1800	-	current
Planted, Intensively managed	1960	-	current
Planted, Superior trees, Traditional Breeding techniques	1970	-	current
Planted, Superior trees, Genetic Modification	1999?	-	future

Even though forest management began in China as early as 100 BC (Menzies, 1985), significant areas of managed forest were not common until the Middle Ages. Planted forests began in some earnest in the 19th century in Europe, and the middle of the 20th century in North America (Sedjo, 1999).

In South Africa, by utilising the genetic variation and different provenances of a commercially available forestry species, both government and private forestry

companies have been able to convert 1 518 138 hectares (1997) of previously unforested land throughout South Africa. Today the industry is worth R12 billion and it directly provides about 68 000 jobs with a further 40 000 through contractorization. Total export earnings of the associated forestry processors represent R5.5 billion per annum. Nett export earnings from this industry are R2 billion per annum (Harvett, 2001).

Indigenous or plantation forestry in South Africa stretches from Mpumalanga in the north through to the Western Cape in the south and includes numerous environmentally diverse sites, each with its own unique influence on tree growth (Harvett, 2001). The major species that were initially introduced over the past 30 years by Mondi Tree Improvement Research (TIR) was pine from Mexico, a number of different pure eucalyptus species, and black wattle from Australia (Dvorak et al., 2000). The introduction and evaluation of new eucalyptus species is currently being co-ordinated through the Institute of Commercial Forestry Research (ICFR). New species of pine and provenances of eucalypts have been introduced through CAMCORE and professional contacts with individuals within the forestry industry (Dvorak and Donahue, 1992).

The great variation in the South African environment has had a substantial influence on the development of forestry in the country. Selection of species suited to areas with distinctly different climates and soil types has involved great effort (Hagedorn, 1993). *Acacia mearnsii* (black wattle), native in Australia, is an important plantation tree in South Africa. The tannins from the bark is used in leather tanning which was previously primarily why black wattle was grown, however this use is now considerably less due to the utilisation of vegetable tannins in the leather tanning industry. The timber is used for mine props, poles and more recently and to a greater extent in the pulp and rayon industry. South Africa produces about one million tons of black wattle timber per annum (Dunlop et al., 2000).

1.2 BLACK WATTLE

1.2.1 Introduction

Research conducted over the past 50 years has played an important role in transforming black wattle (*Acacia mearnsii*) introduced into South Africa from Australia, as a tree to be used for shade for livestock, shelterbelts, windbreaks and for fuelwood, into a major plantation species that contributes substantially to the South African forestry and forest products sector. Today, commercial black wattle (*Acacia mearnsii*) plantations make up approximately 7 % of the South African plantation forestry estate and provide employment, directly and indirectly, to over 36 000 people (Dunlop, 2002).

In Australia black wattle occurs naturally in the south eastern region along the coastal lowlands, and at moderate altitudes of the adjacent interior of southern New South Wales and Victoria, as well as at lower altitudes in eastern Tasmania. Over the relatively wide natural range of the species, the annual rainfall varies from as low as 450 mm per annum to as high as 1 500 mm. In Australia, black wattle appears to grow best where the rainfall ranges from 625 to 875 mm, the mean maximum and minimum temperatures for the warmest and coolest months are 28°C and 0°C respectively, and frosts are limited (Sherry, 1971).

It is reported that Charles and John van der Plank introduced the first black wattle seeds into South Africa from Australia in 1864 (Sherry, 1971), and the first black wattle plantings were carried out about ten years later. By 1884 there was enough black wattle bark in South Africa, and therefore tannin, to conduct the first tanning tests and in 1888 the results indicated that the bark was rich in vegetable tannins, which could be used to tan hides for producing leather of excellent quality (Dunlop, 2002). This led to the establishment of plantations for the production of bark for export and widespread planting of black wattle throughout South Africa, with the greatest success in KwaZulu-Natal (Jarmaine and Lloyd-Jones, 1982).

In the mid 1900's the total area under black wattle plantation had risen to approximately 300 000 hectares when the demand for black wattle bark extract rose dramatically due to the need for saddles, boots and other leather requirements for the soldiers of World War II. The subsequent decline in demand for vegetable tanning agents both locally and internationally led to the area under black wattle shrinking to about 130 000 hectares by the turn of the 19th century, making up about 7 % of the South African commercial forestry areas (Dunlop, 2002).

In 1937 the South African Wattle Growers' Union (SAWGU) was founded to represent the interests of black wattle growers in South Africa and in 1947, the Wattle Research Institute, now called the Institute for Commercial Forestry Research (ICFR), was inaugurated to conduct research into particular aspects of black wattle growing, such as silviculture, genetics, chemistry and entomology (Haigh, 1993). At that stage the black wattle industry was driven mainly by the demand for vegetable tannin, and most of the wood was used for firewood or discarded. Research activities aimed to develop silviculture techniques that reduced the cost of growing the trees and improved the yields. Tree breeding research selected strains with increased disease resistance, and good vigour while entomology and plant pathology researchers studied the major pests and diseases affecting the black wattle (Gibson, 1964).

The demand for black wattle bark products and timber has cemented both locally and internationally, its place in South African forestry (Sherry, 1971). The emergence of competitive markets for black wattle timber and new bark products has brought the need for new research initiatives to provide innovative, multi-disciplinary solutions.

1.2.2 Black wattle products

Black wattle is a very versatile and useful forestry tree. Apart from the commercial value of black wattle extract obtained from the bark, the timber has in the past few years also been identified as a source of raw material for high quality pulp (Barnes, 2001).

Black wattle extract

Black wattle extract is used mainly as a vegetable tanning material because of its relatively high tannin content. The tannin combines with skin protein to form leather, which under normal conditions is not putrescible, even in the wet state, whereas raw skin is readily putrescible in the wet state (Stubbings, 1977). Vegetable tanned leather for shoes, belts, bags, saddles and other products is the main application for black wattle extract using well-known brands, Mimosa ME, Powder and Elephant, Unicorn and Ostrich Solid Mimosa. An important derivative of black wattle extract is the thermo-setting industrial adhesive sold under the Bondtite brand name. It is extremely popular because vegetable-based extract products are considered to be environmentally friendly (Dobson and Feely, 2002).

Black wattle timber

For a long time, black wattle timber was regarded as little more than a by-product, but over the past few years its value has dramatically increased. Black wattle timber was traditionally used for firewood and building purposes and products such as parquet flooring blocks, furniture, and hardboard, as well as mining timber, structural timber and fencing poles. Black wattle timber is relatively hard and knotty, and combined with the small log sizes, is generally not suited for saw timber. However, these properties, combined with the dense timber properties, favourable prices for black wattle and exports, has seen the demand for black wattle timber improve in various markets and other uses (Sherry, 1971). In recent years, it has become a popular source of high quality fibre for pulp production (Gourlay and Barnes, 1994). Black wattle timber is also well suited to the manufacture of charcoal. As long as the 'braaivleis' remains a South African way of life there will always be a demand for firewood and in this field the wattle is eminently suitable (Jarmain and Lloyd-Jones, 1982).

In addition to the uses already mentioned, a further use of black wattle timber is being investigated, and that is its possible conversion to liquid fuel for internal combustion engines (Dobson and Feely, 2002).

1.2.3 Threats to black wattle plantations

Black wattle as a crop is subjected to injury by numerous agencies both organic and inorganic. The main threats to black wattle plantations are abiotic injury, mammalian injury, insect pests, and diseases.

Abiotic injury can be caused by fire, wind, hail and frost. Black wattle is most vulnerable to fire during the winter months when strong, dry winds may prevail. Fires from burning slash and natural grasslands near the plantations are also responsible for majority of plantation fires. Wind damage is generally limited to the breakage of one or both stems of forked trees or to uproot trees with weak root systems. In the summer rainfall region of South Africa there are clearly defined zones known as 'hail belts' where the probability of hail is above average (Sherry, 1971). Where possible, black wattle should not be planted within these zones as the species is susceptible to hail damage. In South Africa, temperatures below freezing injure the stem cambium and may cause swelling near the base of the stem. If the temperatures fall below -4°C the trees usually die, or are killed down to ground level (Beck and Goodricke, 2002).

Mammalian injury is another threat to black wattle plantations. Monkeys have been known to cause stem damage by searching for gum, which exudes from black wattle trees (Sherry, 1971). Wild animals, cattle and goats cause browsing damage to young seedlings in the first months after planting. Current trials have indicated that a simple wire cage with netting, protecting each seedling, will prevent it from being damaged (Hagedorn, 2001).

All the insect pests that attack and affect the silviculture of black wattle seedlings and established black wattle trees in South Africa are indigenous. These insects can be grouped into those pests that affect the establishment of black wattle seedlings, pests of establishment, and those pests that damage trees after they become established, post-establishment pests (Govender, 2002). The whitegrubs and cutworm are pests that affect the establishment of black wattle seedlings in the re-establishment of previous black wattle sites (Govender, 1995). More often than not these pests cause death of seedlings if effective insecticides are not administered in time. Wattle bagworm and

wattle mirid are known as post-establishment pests. Both of these pests cause defoliation of the black wattle trees but can be controlled using insecticides (Govender, 2002).

A number of diseases caused by biotic agents have been identified for black wattle in the past 100 years. Some of these diseases, caused by pathogenic fungi, have resulted in considerable losses to the black wattle industry (Roux, 2002). The Albert Falls disease is characterised by the rapid wilt and death of affected trees as well as branch death and yellowing of the foliage. Although many different fungi have been isolated from infected material, the causal agent has not been identified (Wingfield and Roux, 2000). Another disease, Ceratocystis wilt, is characterised by the rapid wilting and eventual death of trees (Wingfield et al., 1996). The only control measure available for this disease is the selection of disease tolerant trees. With the use of molecular tools, the selection of disease tolerant trees has become possible (Roux, 2002).

The term gummosis refers to a disease condition where gum is exuded from the tree (Roux et al., 1995). The exudation of gum is an inherent reaction of the tree to stressful conditions, either from external injury or infection by a pathogen. In South Africa, the term gummosis has been applied to a complex of diseases associated with black wattle, for example the black butt disease (Zeijlemaker and Margot, 1971). There are two means by which its incidence can be eliminated. The first of these is by trying to prevent rapid growth in plantations, and to avoid mechanical injury to the trees when fungal attacks are at an optimum. The second method for the control of gummosis has been achieved by selecting and breeding trees resistant to the disease complex, including black butt disease (Zeijlemaker and Margot, 1970).

Problems associated with pests and diseases must be actively addressed in order to control the survival and success of the black wattle industry in South Africa. It can be envisaged that diseases of black wattle and other forestry species will increase in the future, due to the movement of people around the world and the increased trade between countries. It is therefore imperative that effective tree breeding strategies be implemented in order to protect and conserve the black wattle species.

1.2.4 Improvement of black wattle

Unlike most forestry species, black wattle was generally cultivated for its bark with timber production being a secondary consideration (Sherry, 1971). Initially the aims of black wattle improvement programmes were to improve the quality and quantity of the bark yields of black wattle and to supply the industry with improved seed (Dunlop, 2002). However, with the recent increase in the use of black wattle timber in local and international pulp mills, the emphasis in the breeding programme has changed from improving bark yield and quality, to improving timber yield and quality while maintaining an acceptable bark quality, with one of the main breeding criteria being the introduction of disease resistance (Dunlop et al., 2003).

Any successful long-term forest tree improvement programme has a well planned strategy. This strategy usually encompasses two aspects (Dunlop, 2002):

- the *breeding* activities employed to improve the species, and
- *production* plans to establish production seed orchards (seedlings or clonal) or to mass produce clones.

Black wattle breeding

Provenance testing, vegetative propagation, pollination mechanisms and progeny testing are the different breeding activities used to improve the black wattle species.

Provenance testing

Provenance testing is important in tree breeding as it evaluates the variation that exists within a species when the parent trees are selected from different geographical regions. Such species have strains within them, which have evolved and adapted themselves to various conditions. Therefore trees of a particular species growing in one area may not be suitable to grow in another area. These different 'populations' of trees within the same species are referred to as provenances of the same species. Therefore

provenance selection forms one of the basic activities in tree improvement (Dunlop, 2002).

Various trials have been performed in order to determine which Australian provenances are the best for obtaining seed material for South African conditions (Dunlop and Hagedorn, 1998). These trials have succeeded in identifying certain provenances in Australia, such as Stratford, Bateman's Bay and Cann River, as being more suitable than others for growth under South African conditions. In the colder parts of the black wattle growing areas of South Africa, there is also a need for frost tolerant trees. Hagedorn (1993) suggested the use of the Australian Lake George material to create a cold tolerant population. Australian material from Lake George and Mittagong and various local seed sources have been used to establish a frost tolerance trial near Sheepmoor in the south eastern Mpumalanga area (Dunlop, 2002). These results will act as indicators of where to obtain seed in Australia for inclusion in the South African breeding programmes in the future.

Vegetative propagation

Vegetative propagation is an option employed to speed up genetic improvement programmes of forestry trees and is fast becoming a very important nursery management tool (Garbutt, 1971). The main uses of the vegetative propagation in improvement programmes are to multiply selected parent trees with superior genotypes so as to avoid the risk of loss, to multiply material for experimental purposes, to produce clonal material of superior genotypes, and to provide stocks on a commercial scale (Dunlop, 2002).

The first rooted cuttings experiments with black wattle started in 1944 through the employment of air-layers, however by 1964 the limited value of the air-layers became apparent, as they were labour-intensive and short lived. Numerous unsuccessful attempts had been made since then to vegetatively propagate black wattle. It was only in 1997 that Ken Liesegang, a private nurseryman, successfully cultivated rooted cuttings from black wattle seedlings (Dunlop, 2002). Subsequently, intensive research

has continued to investigate vegetative propagation through rooted cuttings (Beck, 1999).

Pollination mechanism

There are two possible pollination strategies employed in black wattle breeding and improvement programmes; namely, open-pollination and controlled-pollination.

Open pollination relies mainly on the pollination brought about by insect activity. Moncur et al. (1991), found a wide range of insect species on black wattle flowers with honey bees having the most polyads (carries the pollen) on their bodies. This system is currently used in the black wattle production seed orchards. If trees with superior genotypes are planted in close proximity to one another, an acceptable level of cross pollination is achieved without any human intervention, resulting in sufficient improved seed to satisfy the requirements of the industry (Dunlop et al., 2003). Controlled pollination on the other hand, is a technique to produce offspring, which receive genes from each of two known parent trees. It consists of transferring pollen manually from one tree to receptive female strobili (cones) of another tree, while excluding all other pollen (Dunlop, 2002). This method is not being used and has not been used in the breeding programme to date.

Progeny testing

Progeny testing was introduced into the black wattle industry in the 1940's to test genotypes in order to determine the breeding values for individual trees. Superior trees were selected using morphological and chemical characteristics and their progeny evaluated to estimate the value of the parental genotype (WRI, 1948). The intention was to select trees whose progeny showed outstanding characteristics, propagate them vegetatively, and then to intermix the clones in isolated plantations. The cross pollination between these selected clones should in theory produce superior strains. This method was a sound basis for a breeding strategy; however, with the limited success achieved with vegetative propagation, this breeding strategy still remains theoretical (Dunlop, 2002).

In 1950 the selection of superior phenotypes began and seed collected from these trees were planted in progeny trials (WRI, 1950). This was thus the beginning of many selections and subsequent progeny tests. The half- and full-sibling progeny trials that were established were used to estimate genetic parameters and to provide superior seed for the plantations. Results from these trials indicated that mass selection was suitable. These results also indicated that individual tree heritability rates were estimated for tannin content as 0.55, for gummosis as 0.29, for bark thickness as 0.22 and for diameter as 0.14 (WRI, 1963). These trials showed the potential for selection and their results showed that early selection could be employed with confidence. The many progeny tests planted over the years have identified, and continue to identify, material for inclusion in seed orchards that have since the late 1960's supplied the industry with improved black wattle seed (Dunlop, 2002).

Black wattle production

The production of improved seed is central to the black wattle industry. Improved seed is produced by the establishment of seed orchards for black wattle. The first seed orchard was established in 1965 with two objectives in mind: firstly, to bring together the best parents of families which were outstanding in one-parent progeny tests, and secondly, to form the basis of a plantation for seed production (WRI, 1965). Since 1965, numerous seed orchards of black wattle have been established.

The performance of seed orchard material is tested against unimproved commercial material available to ensure that improvement is in fact occurring. In the black wattle improvement programmes, this involves the inclusion of improved seed in progeny trials. In 1970, the Wattle Research Institute tested its seventeenth progeny trial and found that there was no difference between the commercial and seed orchard material for height, but the seed orchard material was 6.7 % more disease free than the commercial family (WRI, 1970). From these trials and results it became clear that the improvement programmes were in fact producing improved material (Dunlop, 2002).

In the black wattle industry it has been recognized that early identification of superior genotypes, in juveniles, would greatly facilitate and speed up breeding programmes.

With the recent development of techniques such as the polymerase chain reaction (PCR) (Kary and Mullis, 1985) and DNA fingerprinting techniques, it has become possible to assess the genetic make-up of individuals directly. Such molecular technologies have been introduced into the forestry industry and are currently being used on a growing scale. In the black wattle industry, molecular technology is still in its infancy and requires extensive basic research.

1.3 MOLECULAR FINGERPRINTING IN FOREST TREE BREEDING

DNA fingerprinting, identification analysis, profiling or typing all refers to the characterisation of an individual's genome. Every organism has a characteristic phenotype or physical appearance because it possesses a unique, hereditary genotype or make-up (Kirby, 1990). The exception to this rule is identical twins, or clones, as they have the same unique genotype with only subtle differences in phenotype, brought about through complex developmental events and environmental conditions. The DNA of any individual is identical whether it is extracted from roots, leaves, or stems specimens (Bruford et al., 1992). These principles of individual uniqueness and identical DNA structure within all tissues of an organism provide the basis of DNA fingerprinting.

The genetic material present in every cell of an organism, through the employment of molecular technologies, produces a DNA fingerprint unique to the individual. Most of an individual's genetic material is found in the nucleus; however, a small portion occurs in cellular organelles, which are the mitochondria and chloroplasts (Snustad and Simmons, 2000). The small size, relatively rapid rate of evolutionary change and maternal inheritance of mitochondrial DNA, has made it useful for examining population structure and evolutionary relationships (Bruford et al., 1992).

Both sources of DNA, nuclear and cytoplasmic, are broadly made up of two types of sequences: (i) sequences that encode for proteins that determine phenotypic traits such as plumage, height and weight, and (ii) sequences that do not appear to encode a particular product (Snustad and Simmons, 2000). DNA fingerprinting is based upon identifying polymorphic DNA that varies from individual to individual. Genetic markers,

or molecular markers as they are otherwise known, are identified through amplification of their specific sequences by short DNA probes called primers. These primers produce different sized amplification products, which may be accompanied by cleavage. The resulting products are separated in a suitable matrix and visualized as a fragment profile, called a fingerprint (Avise, 1994). The genome of most eukaryotes is so vast, that it contains many mutations in sequence composition, brought about by substitutions, insertions and deletions. The probability therefore, that two individuals contain an identical set of markers is low, providing a means to produce unique molecular fingerprints for individuals (Krawczak and Schmidtke, 1994). Thus, the probability of producing unique fingerprints depends on the type and number of molecular markers used to produce the fingerprints.

Many DNA fingerprinting activities have been undertaken on organisms including micro-organisms, plants and animals. These studies have focused on intra-population studies, where there is a need to identify individuals or determine paternity. It has however, also been applied to other population-based studies, such as looking at population structure and evolutionary relationships between populations and species. Interesting investigations such as the assessment of genetic stability in somatic embryogenesis-derived populations of black spruce has also been undertaken and have revealed that no variation was detected within clones (Isabel et al., 1993).

In recent times DNA fingerprinting has become a standard technology in many commercially utilised species, as well as a range of other species. However, the number of investigations that relate to black wattle is rare. In other forestry species a number of molecular investigations have been published. For instance, in pine, species-specific markers were developed and used to characterise eight pine species (Nkongolo et al., 2002). Similarly, markers were used in eucalyptus to obtain fingerprints for the 15 genotypes (Rodrigo et al., 2002). Diversity studies is another very popular area where fingerprinting is used (Luciane and Cavalli-Molina, 2000). Hwang et al. (2001) conducted a study on two species of *Chamaecyparis,* a valuable timber producing tree in Taiwan, to determine the genetic variation. High levels of genetic variation were estimated, thus indicating extensive genetic differentiation through evolution in the two species.

1.3.1 DNA molecular markers

Molecular markers are not genes in the classical sense in that molecular markers usually do not encode for a particular protein product. Molecular markers are however DNA sequences that show different levels of variation and are extremely useful in a number of fields of genetics such as mapping and fingerprinting, as molecular markers are constant 'landmarks' in the genome, which display a certain amount of polymorphism. These are differences in sequences between individuals that reside at a particular site and it is these different polymorphic types that are termed 'alleles', which are the essential ingredient required for the fingerprinting process (Parker et al., 1998).

Polymorphism detection or DNA variation is used to compare individuals, populations and species. This has developed into a number of different technologies all targeting different parts of the genome in different ways. Much of this genetic variation may be associated with non-coding regions of the DNA and therefore has no impact on phenotypic expression of variability, nor is it subject to natural selection, which acts at the level of the gene product (Cooke and Buckley, 1987).

The physical aspects of DNA make it attractive to be used for marker analysis. DNA is found in nearly all cells of all organisms and can be recovered from both living and dead tissue. Tissues can also be stored more easily under field conditions and in many cases only nanograms of DNA are needed for analysis, when using the polymerase chain reaction (PCR) for amplification (Williams et al., 1990). Therefore, for inter- and intra-population studies', studying the DNA itself is far more useful and accurate in studies on individual identification, paternity testing, inbreeding assessment, genetic diversity and population structure analysis or phenotypic traits.

In the investigation of genetic differences between individuals, the challenge is however, to find an appropriate method that will reliably reveal sufficient genetic variation to answer a particular question, with a minimum amount of effort and expense (Parker et al., 1998). The type of genetic markers available for these investigations 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, analyses 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 1994). At present, a number of methods are available that fall under either multi-locus method of detection or single-locus methods. The more popular methods make use of PCR, which allows for the production of large amounts of a specific DNA fragment with a particular length and sequence, from small amounts of template DNA (Innis et al., 1990).

PCR, which was developed in 1985 by Kary and Mullis, has revolutionised molecular biology and genetics by permitting the isolation and characterisation of specific DNA fragments and has resulted in the development of diverse applications in many fields (White et al., 1989). In order for PCR to function effectively, however, primers that flank the target sequence and initiate the chain reaction are required. Some of the single-locus molecular markers such as restriction fragment length polymorphisms (RFLPs), micro- and minisatellites which are generally referred to as simple sequence repeats (SSRs), require specific primers to function, whereas multi-locus markers, such as random amplified polymorphic DNAs (RAPDs) and amplified fragment length polymorphisms (AFLPs), use arbitrary primers. Due to the specific nature of the single-locus markers, only the particular locus in question is amplified, whereas in the case of multi-locus markers a number of arbitrary loci are amplified.

Single-locus markers

There is a wide range of different single-locus molecular markers that have been utilised in the development of fingerprinting technologies. Single-locus markers are sequences of DNA that show variation at a particular locus in the genome. Some of these markers are genes in the traditional sense, while others are repeat sequences. A range of these markers relates to differences in the annealing capabilities of a DNA probe or primer, where the ability to anneal requires a complementary sequence. In the case where there is a mutation in the annealing sequence, the probe or primer does not anneal and in that way indicates a difference. Due to their polymorphic content, single-locus markers have gained popularity in a range of different investigations, such as studies of

complex mating systems and comparisons of genetic variation between populations (Bruford et al., 1992).

There are mainly two types of single-locus markers. They are restriction fragment length polymorphisms (RFLPs) and simple sequence repeats (SSRs).

Restriction fragment length polymorphisms (RFLPs) were one of the first DNA fingerprinting procedures developed and are today widely used as a means for determining paternity in various species (Kirby, 1990). RFLPs are markers detected by treating DNA with restriction enzymes (Burke et al., 1991). If a restriction site is present on a strand of DNA, the DNA strand will be cleaved by the presence of the corresponding restriction enzyme that would recognise the site. This would result in the strand decreasing in size and thus showing up as two different bands on a gel. Variations in the characteristic pattern of a RFLP digest can be caused by base pair deletions, substitutions, inversions and translocations, which result in the loss or gain of a recognition site resulting in cleavage of a fragment.

The advantage of using RFLPs lies in the simultaneous use of a few restriction enzymes in an analysis resulting in DNA fingerprints that vary from one individual to another due to the number of different sized fragments produced by the digestion (Parker et al., 1998). RFLPs can thus generate sufficient variation to investigate genetic questions about within and between population genetic variations. Furthermore, they are co-dominant markers, which make them very useful to distinguish between homozygous dominant individuals and heterozygous individuals (Parker et al., 1998). Unfortunately, this method is very labour-intensive and expensive to develop.

Another type of single-locus marker, known as simple sequence repeats (SSRs), exhibits variation in tandem repeat numbers and are found in numerous areas in the genome of an organism (Avise, 1994). Genetic variation or polymorphisms are evident as the number of core sequences present in an individual. An allele is represented by a certain number of repeats of the core sequence (Parker et al., 1998). In a number of instances however, these loci occur at more than one site in the genome, so that with a single primer set a multi-locus type fingerprint results.

These SSRs can be sub-divided into two classes according to the number of nucleotides making up the repeat unit (Krawczak and Schmidtke, 1994). Minisatellites, otherwise known as variable number tandem repeats (VNTRs), comprise short tandem repeats of about 40 base pairs in length (Avise, 1994). Discovered in 1985 by Alec Jeffreys and his research team (Jeffreys et al., 1985), they were first used to establish what is known today as a DNA fingerprint (Davies, 1988). Microsatellites, on the other hand, are composed of varying numbers of tandemly repeated one to five base pair motifs, for example, mononucleotide repeats (A, T) or dinucleotide repeats (AT, GA). The discovery that these sequences are ubiquitous in various animals and plants has contributed to their usefulness and beneficial use in population biology, parentage determination, and individual assessment.

Single-locus markers are extremely powerful tools in DNA fingerprinting (Bruford et al., 1992). Comparison between individuals on the same gel is simple as only a few, usually clear bands are present in an analysis gel. Furthermore, under optimal conditions, very little DNA template is required; in most cases less than 100 ng (Krawczak and Schmidtke, 1994). However, one major drawback of these markers is that they require prior sequence information, which is more than often not available. This means that extensive and expensive research on the organism's sequence structure is necessary before the fingerprinting analysis can take place. Furthermore, for multi-locus profiling, time-consuming amplification and screening of each locus needs to be undertaken (Bruford et al., 1992).

Multi-locus markers

Multi-locus fingerprints are primarily viewed as fingerprints that are generated from a number of loci and viewed in a single lane on a gel. They can either be generated by the amplification of several single-locus markers, using the primers specific for each of the loci, or by amplifying many loci using one or a few arbitrary primers (Krawczak and Schmidtke, 1994). The use of arbitrary primers does not require any prior knowledge of the DNA composition and is much easier and less time-consuming than having to amplify several markers independently. Multi-locus fingerprints have high information content and can be applied for individual identification and parentage analyses.

There are mainly two types of multi-locus markers. They are randomly amplified polymorphic DNA (RAPD) (Williams et al., 1990) and amplified fragment length polymorphism (AFLP) (Vos et al., 1995).

Williams et al. (1990) developed a PCR based genetic marker, named randomly amplified polymorphic DNA (RAPD). This technology does not require any prior knowledge of the genome of the organism and is based on the application of random primers. The primers are arbitrary, approximately 10 to 11 nucleotide bases in length and anneal to multiple sites on the template DNA due to their short nature (Williams et al., 1990). If these primers anneal in the correct orientation and are a suitable distance apart, the unknown sequence between them is amplified, resulting in visualised bands on an agarose gel (Welsh and McClelland, 1990). Polymorphisms between genotypes are due to the occurrence of single nucleotide polymorphisms (SNPs) that are present in the annealing region of the primer or due to insertions or deletions of segments in the sequence between the annealing primers. By changing the annealing sequence through inserting, deleting or substituting single bases, these SNPs alter the ability of the primer to anneal at a specific location. Without primer binding, sequence amplification cannot occur, resulting in the absence of a band in the analysis. RAPD alleles are therefore seen as *'presence'* or *'absence'* alleles *(null-alleles)*, where a fragment will be amplified if the primer anneals; or will not, if the primer is not able to anneal. Polymorphisms are seen as presence or absence of specific bands on a gel (Bowditch et al., 1991). Most loci generated by RAPD amplification are assumed to be dominant for band presence in the absence of segregation analysis.

RAPDs are often used to identify individuals and to analyse populations. They thus require high reproducibility. Once the ideal or optimised conditions regarding primer constraints, stringency of the reaction, DNA quantity and concentration, *Taq* polymerase concentration and PCR cycle, have been determined for a particular species, it is imperative to keep them constant in order to be able to compare results from different reactions (Bowditch et al., 1991). Careful attention to laboratory techniques will result in reproducible reactions from run to run, provided the same thermocycler is used.

Most successful results are obtained when many primers are screened so that multiple polymorphisms are detected. Not all primers will yield reliable and reproducible results, therefore only those that do should be included in the analysis and used for scoring (Bowditch et al., 1991; Fritsch and Rieseberg, 1995).

RAPDs have become very important genetic markers because the assay is simple, fast and relatively inexpensive, allowing many loci to be identified in a single reaction. Furthermore, only small amounts of DNA are required and no prior knowledge of the genome in question is necessary (Fritsch and Rieseberg, 1995). They do however, have a few disadvantages. RAPDs are dominant markers; therefore heterozygotes cannot be distinguished from homozygous dominant genotypes (Williams et al., 1990). Furthermore, the results obtained may be inconsistent because PCR is performed using very short, arbitrary primers that may lead to different patterns depending on the PCR conditions used. Therefore, in order to get consistent and reliable results, RAPD conditions need to be carefully standardised and maintained from run to run (Bowditch et al., 1993).

The advantages RAPDs offer over the other molecular techniques, has led to a broad range of applications in fields of biology such as ecology, conservation and population biology (Haig, 1998). Although RAPD analysis is relatively new it has already proven useful for genetic mapping, species and individual identification, pedigree analysis and parentage determination (Fritsch and Rieseberg, 1995). These markers have been used with success in numerous studies and are becoming increasingly more popular.

Another molecular marker, that also uses the multi-locus system, is amplified fragment length polymorphism (AFLP). The AFLP technique developed by Vos et al. (1995) is an efficient molecular technique for generating a large number of DNA-based genetic markers. It is very similar to RAPDs in that it requires no prior sequence knowledge and is based on random amplification and identification of polymorphisms at many DNA loci (Pejic et al., 1999).

The AFLP technique is based upon repeated amplification using PCR on a subset of restriction fragments from a total digest of genomic DNA (Desmarais et al., 1998). This

is achieved by digesting DNA with two restriction enzymes, a rare and a frequent cutter, according to the length of their restriction site. This is followed by each fragment being ligated to adaptors that serve as a binding site for primers. Only fragments that have both restriction sites (one at either end) will be amplified by PCR. This initial amplification reduces the total number of restricted fragments present in the reaction and is therefore pre-selective (Questiau et al., 1999). The PCR product is then used as a template for a second round of amplification (selective) using primers with three additional selective nucleotides included at the 3' end. The amplification is selective because one selective nucleotide on each primer results in the amplification of only one out of 16 fragments, while two selective primers will reduce this value to 1/256. Therefore, the number of resulting amplified fragments is determined by the number and composition of the selective nucleotides used. The final product of the whole process is a multi-locus fingerprint-like pattern on a gel that can be scored with an automated sequencer (Questiau et al., 1999).

Since Vos et al. (1995) first published this method; it has become increasingly more popular in studies on genetics. At first it was mainly used in plant mapping, as seen in a study by Schondelmaier et al. (1996), and in studies on crop and wild plant diversity (Travis et al., 1996). A study by Ziegenhagen et al. (1999), showed the usefulness of AFLPs in kinship determination and paternity analysis in oak trees and a study by Pejic et al. (1999) analysed the genetic similarity among maize inbred lines, using AFLPs.

The advantages of using AFLPs are that it is PCR based and thus does not require probe hybridisation, many bands (approximately 50 - 100) are displayed and only a small amount of DNA is needed to get an accurate result (Desmarais et al., 1998). The banding patterns also seem to be more consistent than those found in RAPDs. The major disadvantage of this method however, is that it is significantly more expensive than RAPD analysis and is technically more challenging. Furthermore, like RAPDs, AFLPs cannot distinguish between heterozygotes and homozygous dominant individuals as the process is based on the *'presence'* and *'absence'* of bands. They are thus dominant markers and not co-dominant markers (Pejic et al., 1999).

Table 1.2 shows a detailed summary of the comparison of the single- and multi-locus markers.

Table 1.2 Comparative descriptions of common DNA molecular genetic markers.

Characteristic	RFLP	SSR	RAPD	AFLP
Origin	Anonymous/ Genic	Anonymous	Anonymous	Anonymous
Maximum theoretical number of possible loci in analysis	Limited by the restriction site (nucleotide) polymorphism (tens of thousands)	Limited by the size of genome and number of simple repeats in a genome (tens of thousands)	Limited by the size of genome, and by nucleotide polymorphism (tens of thousands)	Limited by the restriction site (nucleotide) polymorphism (tens of thousands)
Dominance	Co-dominant	Co-dominant	Dominant	Dominant
Null-alleles	Rare to extremely rare	Occasional to common	Not applicable (presence/ absence type of detection)	Not applicable (presence/ absence type of detection)
Transferability	Across genera	Within genus or species	Within species	Within species
Reproducibility	High to very high	Medium to high	Low to medium	Medium to high
Amount of sample required per sample	2-10 mg DNA	10-20 ng DNA	2-10 ng DNA	0.2-1 µg DNA

Ease of development	Difficult	Difficult	Easy	Moderate
Ease of assay	Difficult	Easy to moderate	Easy to moderate	Moderate to difficult
Automation/ Multiplexing	Difficult	Possible	Possible	Possible
Genome and QTL mapping potential	Good	Good	Very good	Very good
Comparative mapping potential	Good	Limited	Very limited	Very limited
Candidate gene mapping potential	Limited	Not good	Not good	Not good
Potential for studying adaptive genetic variation	Limited	Limited	Limited	Limited
Development	Moderate	Expensive	Inexpensive	Moderate
Assay	Moderate	Moderate	Inexpensive	Moderate to expensive
Equipment	Moderate	Moderate to expensive	Moderate	Moderate to expensive

1.3.2 Genetic analysis of fingerprints

Molecular markers provide powerful means for analysis of variation between species, populations, groups of individuals within populations, and between individuals both related and unrelated. The variation produced by the various markers is visible in DNA fingerprints; however these results need to be transformed into numerical format so that various measures of variation can be determined.

Through the years, various tests, comparative indices and formulae have been developed to assess molecular variation (Avise, 1994). These mathematical and statistical formulae include the assessment of genetic variation and genetic relatedness between two entities, be they individuals or populations. All of these methods, although different from one another, have their own particular advantages and disadvantages. They all calculate or estimate numerical values that can be used to determine genetic variation or genetic relatedness.

Assessment of genetic variation

Genetic variation refers to the variation at the level of individual genes (polymorphism), and provides a mechanism for populations to adapt to their ever-changing environments. The greater the genetic variation, the greater the probability that at least some individuals in a population will have an allelic variant that is suited to the new or changing environment. These individuals will produce offspring with variant alleles, thereby assuring that such beneficial alleles are inherited from generation to generation (Sun et al., 1998).

Numerous measures have been developed to quantify genetic variation. These indices include genetic polymorphism, heterozygosity and Shannon's Information Index.

A genetic polymorphism is defined as the occurrence of two or more alleles at one locus in the same population (Cavalli-Sforza and Bodmer, 1971). A locus is considered polymorphic if the frequency of the most common allele is less than 0.99 or in some cases less than 0.95. The proportion of polymorphic loci is calculated by counting the

number of polymorphic loci and then dividing this number by the total number of loci examined.

$$\text{Frequency of polymorphism} = \frac{\text{Total number of polymorphic loci}}{\text{Total number of loci studied}}$$

The number of loci examined and the number of individuals examined in a population directly affect the accuracy of this measure.

The determination of polymorphic content as a measure of genetic variation is not satisfactory on its own, being both arbitrary and imprecise in that slightly polymorphic loci are treated in much the same way as very polymorphic loci (Ayala, 1982). A more suitable measure of genetic variation is that of the average frequency of the heterozygous individuals per locus present in the population (Cooke and Buckley, 1987). This measure is called genetic heterozygosity and can be calculated in a number of different ways, depending on whether variation between individuals, populations or loci is being examined.

The heterozygosity values are determined by using either observed frequencies, such as in co-dominant markers, or through estimation with dominant markers, where Hardy-Weinberg equilibrium is assumed and the frequency of the recessive allele (q) calculated (Ayala, 1982). Although the Hardy-Weinberg equilibrium rarely exists in nature, it is useful in indicating the presence of selective processes in a population. Once the frequency of the *null-allele* (q) has been estimated from the frequency of the recessive phenotype, the dominant allele frequency (p) can be estimated and then the frequency of the heterozygotes (2pq) (Ayala, 1982).

Shannon's Information Index is another measure of genetic variation. The higher the Shannon's Information Index, the greater the genetic variation within the population and is scaled from a value of zero to one, where a value of one equals the maximum diversity attainable. Shannon's Information Index can be calculated using the following formula (Hwang, 2001):

- $H_0 = -\sum P_i \log_2 P_i^2$, (Hwang, 2001)
  Where P_i is the frequency of the presence or absence amplified fragment.

As Shannon's Information Index is based on information theory it is sensitive to the presence of rare types or individuals.

Another diversity index known as Simpson's Index is less sensitive to rare types and the interpretation thereof is more intuitive than Shannon's Information Index (Szczepaniak et al., 2002).

Simpson's diversity Index can be calculated using the formula:

- $D = 1 - \sum n_i (n_i - 1)/ N (N - 1)$, (Szczepaniak et al., 2002)
  Where n_i is the number of individuals with the phenotype i and N the sample size.

Similar to Shannon's Information Index, a value of zero indicates no diversity and a value of one the maximum diversity.

Originally, Simpson's Index represented the probability that any two individuals, selected at random would be different. Pielou's (1975) modification of Simpson's Index however, transformed this probability-based index into the class of indices to which Shannon's Information Index belongs. For this reason these two indices are similar in many aspects and have the same applicability.

Assessment of genetic relatedness

A number of statistical methods have been formulated in an attempt to quantify the degree of genetic relatedness between individuals and populations. Two such statistical calculations have been formulated by Nei (1972). They are (1) genetic identity (I), which estimates the proportion of genes that are identical in the two entities being compared, and (2) genetic distance (D), which estimates the accumulated number of gene

differences per locus that have occurred over evolutionary time. If the rate of gene substitutions per year is constant, the genetic distance is linearly related to evolutionary time.

One advantage to these measures of genetic differentiation is that they apply to any population, whether they be haploid, diploid, tetraploid or selfing. This is because the definitions of I and D depend solely on gene frequencies rather than on genotype frequencies as in the Hardy-Weinberg equation.

The formulation for genetic identity is derived as follows:

- $I = \Sigma x_i y_i / (\Sigma x_i^2 \Sigma y_i^2)0.5$, (Nei, 1972)
  Where the frequencies of the *i*th alleles in populations X and Y respectively, are x_i and y_i.

For multiple loci, the overall similarity or identity is :

- $I = J_{xy} / (J_x J_y)0.5$, (Nei, 1972)
  Where J_x, J_y and J_{xy} are the arithmetic means across all loci of Σx_i^2, Σy_i^2, and $\Sigma x_i y_i$ respectively. The formulation for genetic distance (D), is given by:

- $D = -\ln I$, (Nei, 1972)

The values for I may range from 0 (no alleles in common) to 1 (the same alleles, and in the same frequencies are found in both populations). Genetic distance, D, may range from a value of 0 (no allelic substitutions) to infinity. Generally speaking, closely related populations tend to show I values > 0.9 and D values < 0.1. Divergent populations or separate species tend to have genetic identities I < 0.8 and genetic distances D > 0.2.

Proper interpretation of these values may even be used to determine whether populations are undergoing genetic differentiation during speciation. Local populations,

which are genetically similar, have I values in excess of 0.97. As speciation takes place, the I value decreases and the D value increases.

Another measure of genetic distance has been proposed by Roger (1972). Rogers' D is defined as:

- $D = [0.5\Sigma(x_i - y_i)^2]0.5$, (Roger, 1972)
  Where for a given locus with m alleles, x_i and y_i are the frequencies of the ith allele in populations X and Y, where the summation is over all alleles.

When data from more than one gene is considered, the arithmetic mean of such values across loci provides the overall genetic distance estimate. Rogers' D values are between zero and one and similarity is given by:

- $S = 1 - D$, (Roger, 1972)

Assessment of population differentiation

There are two methods to assess within and between population differentiation, Wright's fixation index (F_{ST}) and the analysis of molecular variance (AMOVA).

Wright's fixation index (F_{ST})

When a population is divided into isolated subpopulations, there is less heterozygosity than there would be if the population was undivided. The decline in heterozygosity due to subdivision within a population is due to the phenomenon of genetic drift and has usually been quantified using an index known as Wright's *F* statistic, also known as the fixation index. The fixation index is a measure of the difference between the mean heterozygosity among the subdivisions in a population and the potential frequency of heterozygotes if all members of the population mixed freely and non-assortatively (Hartl and Clark, 1997). The fixation index ranges from 0, thus indicating no differentiation between the overall population and its subpopulations, to a theoretical maximum of one,

though in practice the observed fixation index is much less than one even in highly differentiated populations.

Fixation indices can be determined for three differentiated hierarchical levels of a population structure, with the most commonly used of the three being F_{ST}. F_{ST} is concerned with subpopulations relative to the population as a whole and therefore gives a measure of the extent to which a species is organised into subpopulations with restricted gene flow (Balding et al., 2001).

Wright proposed the following guidelines for the interpretation of F_{ST}:

- A value between 0 and 0.05 indicates little genetic differentiation.
- A value between 0.05 and 0.15 indicates moderate differentiation.
- A value between 0.15 and 0.25 indicates great genetic differentiation.
- A value above 0.25 indicates very great genetic differentiation.

Wright also mentioned that a value less than 0.05, although indicating little differentiation, were by no means negligible.

Analysis of molecular variance (AMOVA)

Analysis of molecular variance (AMOVA), is another measure of population differentiation. Unlike Wright's fixation index that estimates population differentiation by assuming Mendelian gene frequencies, this index estimates the population differentiation directly from molecular data. A variety of molecular data such as molecular marker data, direct sequence data or phylogenetic trees based on such molecular data may be analysed using this method (Excoffier et al., 1992).

AMOVA treats any kind of raw molecular data as a Boolean vector p_i, that is, a 1 x n matrix of ones and zeros, one indicating the presence of a phenotype and zero its absence. A marker could be a nucleotide base, a base sequence, a restriction fragment or a mutational event (Excoffier et al., 1992). From this information Euclidean distances, defined as the shortest distance between two points, are calculated by subtracting the

Boolean vectors of one entity from another. A matrix consisting of all pair-wise squared Euclidean distances between all entities is constructed and used in the analysis (Excoffier et al., 1992).

Squared Euclidean distances are calculated according to the following formula:

- $\theta^2_{jk} = (p_j - p_k)' W (p_j - p_k)$, (Excoffier et al., 1992)
  Where W is a matrix of differential weights for the various sites.

Using the constructed matrix, a hierarchical analysis of variance is performed to determine the subdivision. Usually, in the simple cases, the total variance is partitioned into between populations and within populations to give an idea as to how much of the genetic diversity can be attributed to each of these components. Variance can however be further subdivided into within individual differences, between individuals within populations, between populations within groups and between groups.

1.4 APPLICATIONS OF FINGERPRINTING IN FOREST TREE BREEDING

In recent years great strides have been made in the field of molecular analyses in forestry species. Fingerprinting has been employed to determine genetic variation within populations as well as between populations of forestry species (Ahuja, 1995). It has been reported that the global forestry plantation area has increased by an estimated 32 million hectares while the area of natural forests has declined by 126 million hectares. Molecular markers have became invaluable in the quantification of the genetic variation to aid in sampling strategies and especially to identify specific genotypes that contain particular genes, such as genes for resistance to diseases and environmental conditions such as drought and frost (McAree, 2002). Markers are especially advantageous for agronomic traits that are otherwise difficult to tag such as resistance to pathogens, insects and nematodes, tolerance to abiotic stresses, quality parameters and quantitative traits. Molecular marker studies have accelerated the mapping of many genes in different plant species. Examples of successes with molecular markers are provided in Table 1.3.

Table 1.3 Selected examples of association of molecular markers with the desired traits in different crops.

Characteristics	Examples	References
Wheat		
Powdery mildew resistance genes *Pm-1* and *Pm-2*	RFLP markers 3cM from *Pm-1* and 3cM from *Pm-2*	Hartl et al., 1995
Maize		
Leaf blight resistance *rhm* gene	Tight linkage of *rhm* with RFLP loci UMC85 and p144	Zaitlin et al., 1993
Northern corn blight resistance gene *Htn-1*	*Htn-1* gene 0.8cM distal to RFLP UMC117	Simcox and Bennetzen, 1993
Sorghum		
Head smut resistance *Shs*	Linkage with RFLP loci detected by probes pFBT, xS560 and xS1294. One RAPD locus from primer OPG5	Oh et al., 1994
Barley		
Stem rust resistance gene *Rpg 1*	RFLP marker ABG704 on chromosome 1	Killian et al., 1994 Penner et al., 1995
Stem rust resistance gene *rpg4*	3 RAPD markers on chromosome 7M	Borovkova et al., 1995
Resistance to *Rhynchosporium secalis*	Cosegregation with RFLP markers on chromosome 3	Graner and Tekauz, 1996
Soybean		
Cyst nematode resistance	Two RFLP markers associated resistance	Skorupska et al., 1994

Resistance to soybean mosaic virus *Rsv* gene	Marker at 0.5, 1.5 and 2.1 cM	Yu et al., 1994
Pea		
Powdery mildew resistance	RFLP markers at 11 cM	Dirlewanger et al., 1994
Alfalfa		
Somatic embryogenesis	RAPD markers	Yu and Pauls, 1993
Tomato		
Insect resistance mediated by 2-tridecanone (2-TD)	Direct selection for RFLP loci increased the frequency of 2-TD-mediated resistance	Nienhuis et al., 1987
Nematode resistance *Mi* gene	RFLP markers tightly linked	Klein-Lankhorst et al., 1991
Resistance to powdery mildew caused by *Oidium lycopersicum O1-1* gene	Near *Aps-1* region on chromosome 6 close to *Mi* and *Cf-2/Cf-5* genes	Van de Beek, 1994
Soluble solid content (SS)	RFLPs linked to SS	Osborn et al., 1987
Potato		
Cyst nematode resistance *H1* gene	RFLP marker at 2.7 cM from H1	Pineda et al., 1993
Cyst nematode *GroV1* locus in *solanum vernei*	RFLP markers on chromosome 5	Jacobs et al., 1996

The main area where microsatellite markers are being applied in forestry trees include studies of genetic variation in natural and breeding populations, particularly in species with low levels of gene flow, pollen and/or seed dispersal and mating systems (Kostia et al., 1995). As these parameters are relevant to the conservation of forest genetic resources, microsatellites are being used to monitor genetic impacts of forest management practices and of fragmentation. The first microsatellites developed in forestry trees were for *Pinus radiata* (Smith and Devey, 1994), in which 24 loci were

characterised. They have since been developed from a range of temperate and tropical forestry trees, including *Eucalyptus sieberi* (Glaubitz et al., 1999), *Eucalyptus grandis* (Brondani et al., 1998), *Eucalyptus nitens* (Byrne et al., 1997), and in several pine species (Kostia et al., 1995; Echt et al., 1996; Hicks et al., 1998) to name a few. Bundock et al. (2000), constructed linkage maps using microsatellite markers in *Eucalyptus globulus*. They found that the male parent had 13 linkage groups covering 1013 cM, while the female parent had 11 linkage groups covering 701 cM.

Restriction fragment length polymorphism (RFLP) mapping in tree improvement has many useful applications. It adds to the number of genetic markers known in trees, facilitates the assessment of genomic organisation, is used to determine population genetic variation, and is employed to evaluate evolutionary relationships. One of the greatest limitations of this technology is the amount of prior research required before the technology can be employed. Specific probes or primers have to be developed which is a lengthy and expensive process (Neale and Williams, 1990).

The applications of RAPD markers in forestry have been very successful. The application of RAPDs to forest trees includes identification of genotypes and genetic relatedness within and between populations. RAPD markers have been used for identity studies in *Populus* (Sanchez et al., 1998) where 25 poplar clones were screened for RAPD markers in order to evaluate the use of RAPD analysis to distinguish between species. The marker produced characteristic bands for every species. De Laia et al. (2000) also used RAPD markers to analyse the genotypes of *Eucalyptus* clones hybrids obtained by vegetative micropropagation. Various diversity studies have also employed RAPD technology. Mosseler et al. (1992) used RAPD markers to confirm the low levels of genetic diversity in red pine which demonstrated the long time periods required for recovery following a loss of genetic diversity in long-lived, long-generation organisms like trees. In 1995, Bucci and Menozzi investigated the genetic variation of RAPD markers in an Italian population of Norway spruce. Here they found that the expected genetic diversity ranged from 0.119 to 0.508 and the single-locus inbreeding estimates were calculated at –0.136 indicating an excess of heterozygotes.

Research achievements that relate to the employment of molecular technologies in black wattle are limited. Studies include the examination of genetic variation in natural populations of *Acacia mearnsii* (Searle and Bell, 1998). Twenty-three isozyme loci were examined within and between 19 natural populations of *Acacia mearnsii*. *Acacia mearnsii* was found to have moderate genetic variation (0.201) with the majority (89.2 %) of variation occurring within populations. Another study conducted by Butcher et al. (1998) estimated the genetic variation in *Acacia mangium* using 57 anonymous RFLP loci for ten individuals from each of ten natural populations. The level of genetic variation varied significantly among the populations, ranging from 0.01 on the island of Ceram in Indonesia to 0.21 in Muting, New Guinea. The small, geographically isolated populations of Daintree, Townsville, Ceram and Sidei had low levels of diversity (0.01 to 0.09) whereas the large New Guinea and the Cape York Peninsula populations had higher levels of variation (0.16 to 0.21).

1.5 AIMS

In the forestry industry, one of the most important complicating factors is that of a long generation interval. Many of the economically important characteristics are identifiable once the trees are adult, thus causing a delay in detection (Zobel and Talbert, 1991). This delay impacts on the effective execution of many production and breeding activities and is accompanied by large financial contributions that could be avoided in the case of early identification. In black wattle, it would be beneficial to the industry if early detection tools could be developed to identify important genetic traits in the seedling. Gummosis is an important disease, causing extensive losses. Identification of resistance genotypes in juvenile plants would assist early selection of superior genotypes, thereby facilitating production and breeding of the tree (Bousquet and Lalonde, 1990).

Another very important factor in black wattle breeding programmes is that of the knowledge of the extent of available variation in a population. Black wattle was introduced into South Africa from Australia and therefore, due to the limited number of seeds that formed the genetic base of South African black wattle, it is expected that the South African population is genetically rather uniform. Thus an assessment of the

genetic variation and genetic relatedness of the trees included in this investigation was undertaken.

As this was the first comprehensive molecular population investigation of its kind, in black wattle, optimization of the technologies required to generate RAPD fingerprints, such as optimization of a DNA isolation method and RAPD amplification conditions were firstly undertaken.

RAPD fingerprints were generated for individuals of a number of black wattle families and assessed in terms of the following:

- Identification of RAPD fragments linked to gummosis.
- Assessment of genetic variation within families and the population.
- Assessment of family relatedness.
- Assessment of family differentiation.

CHAPTER 2

MATERIALS AND METHODS

2.1 INTRODUCTION

In black wattle, one of the most important forestry species in South Africa, the disease gummosis is of great importance as it causes extensive losses each year. This investigation was undertaken to identify RAPD markers, which could be developed into identification procedures to identify resistant genotypes in juvenile plants. This would greatly facilitate selection procedures in the production and breeding programmes of black wattle. Furthermore, the data generated from this investigation would also be used to estimate the genetic variation in black wattle as the extent of its genetic base is unknown.

This investigation therefore comprised of three main parts, namely:

- To find a suitable protocol for extracting high molecular weight DNA from various ages of black wattle trees,
- To test various primers in order to find markers linked to disease resistance, and
- To estimate the genetic variation using molecular data.

All recipes of solutions are taken up in the Appendix.

2.2 MATERIAL

The screening for markers linked to disease resistance of the disease, gummosis, in black wattle was carried out by first isolating the DNA and then running a series of RAPD fingerprints using different primers in an attempt to identify a band that would be present in the disease resistant samples, but absent in the susceptible samples.

2.2.1 Plant material

Plant material was obtained from the Institute for Commercial Forestry Research (ICFR) nursery in Pietermaritzburg, KwaZulu-Natal. Leaf samples were taken from four families, of which three where susceptible to the disease complex gummosis and one was resistant to the disease, and placed in a plastic bag and then on ice. Details of the families are displayed in Table 2.1.

Table 2.1 Detailed histories of the four black wattle families used in this investigation.

Family	Number of seedlings per family	History
99	**11**	These seedlings were half-sibs grown from open pollinated seed collected from tree 99. This tree had proven to be *susceptible* to the disease complex known as gummosis.
101	**13**	These seedlings were half-sibs grown from open pollinated seed collected from tree 101. This tree had proven to be *susceptible* to the disease complex known as gummosis.
101x77	**6**	These seedlings were full sibs grown from seed produced by crossing tree 101 (seed parent - mother) with tree 77 (pollen parent - father). In the ICFR progeny tests, they had proven to be *susceptible* to disease.
18	**5**	These seedlings were half-sibs grown from open pollinated seed collected from tree 18. This tree had proven to be superior for the traits under selection in the black wattle breeding programme at the ICFR, in particular for the *disease resistance* to gummosis.
Total	**35**	

2.2.2 DNA source

DNA isolation is a very important part of plant molecular genetic studies. The cells of many plant species contain phenolics and polysaccharides that interfere with the DNA isolation process. Therefore, in order to isolate high integrity DNA, the isolation protocols used in plant molecular investigations are far more complex than that of animals and often require extensive optimisation. In black wattle, a woody species, large amounts of polysaccharides and tannins are present in cells making it difficult to separate the DNA out. As only a limited number of molecular investigations have been undertaken with black wattle, no suitable protocol has been developed. Thus, the first part of this study involved the optimisation of a DNA isolation protocol.

High molecular weight DNA is required for DNA amplification. Therefore, for the optimisation of the DNA isolation protocol, plant material was collected from trees of a range of ages (Table 2.2). Leaves were collected at Bloemedal Research Station in KwaZulu-Natal from one-, five- and eleven-year old black wattle trees. Approximately ten young and fleshy leaves, depending on the number available, were picked from tips of branches. The leaves were immediately placed in plastic bags, sealed, labelled and placed on ice.

Table 2.2 **Collection information of the material used in the DNA isolation optimisation process.**

ID	Label	Age (years)	Collection date
B19.09	G1	1	31/01/2003
B19.09	G2	1	31/01/2003
B19.09	G3	5	31/01/2003
B19.09	G4	5	31/01/2003
B19.09	G5	11	31/01/2003
B19.09	G6	11	31/01/2003

2.3 METHODS

2.3.1 DNA isolation

Various DNA isolation procedures were tested in order to find a suitable method for the isolation of DNA from the black wattle leaves of different ages. These included a standard hexadecyltimethylammonium bromide (CTAB) procedure of Doyle and Doyle (1990), a standard salting out method from Aljanabi and Martinez (1997) and a Chelex©-100 procedure modified from Doyle and Doyle (1990). The protocols were standardised for 0.5-1 g leaf material that could be handled in a 30 ml chloroform resistant plastic centrifuge tube.

DNA isolation protocol 1 - Standard CTAB procedure (Doyle and Doyle, 1990)

This protocol takes three days to complete and is as follows:

1. Preheat 2 % CTAB isolation buffer to 65°C (± 15 minutes).
2. Wipe the leaf material with ethanol (EtOH) and grind fresh leaf tissue to a powder with liquid nitrogen.
3. Add 3 ml CTAB isolation buffer to grindate and pour into chloroform resistant plastic centrifuge tubes. Rinse the mortar with 1 ml additional CTAB buffer into the tubes; place tubes on ice.
4. Incubate samples at 65°C for one hour with occasional gentle swirling (± every 15 minutes).
5. Cool samples for 15 minutes at room temperature, add an equal volume (5 ml) of chloroform:isoamyl alcohol (24:1), mix well and incubate for five minutes at room temperature.
6. Centrifuge for 10 minutes at 8 000 rpm in a refrigerated superspeed centrifuge.
7. Remove the top layer containing the DNA and place it in a chloroform resistant centrifuge tube, thereafter 500 µl of CTAB wash buffer was added.
8. Incubate the tubes at 65°C for 30 minutes with occasional gentle swirling (± every 15 minutes).
9. Cool the tubes for 15 minutes at room temperature; add equal volume (5 ml) of chloroform:isoamyl alcohol (24:1); mix well and incubate for a further five minutes at room temperature.

10. Centrifuge the tubes for ten minutes at 8 000 rpm in a refrigerated superspeed centrifuge.
11. Add top layer containing the DNA to a chloroform resistant centrifuge tube and an equal volume (5 ml) ice cold isopropanol and mix gently. Precipitate the DNA for 60 minutes at -20°C (can leave overnight at -20°C).
12. The tubes may then be centrifuged for 10 minutes at 10 000 rpm as in step 6.
13. Pour off the supernatant and remove isopropanol by inverting tube.
14. Add 400 µl of 1 x TE to the DNA pellet in the tube and incubate in a waterbath at 50°C for 15 minutes to dissolve the pellet.
15. 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.
16. 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.
17. Pour off the supernatant; centrifuge for one minute and remove excess EtOH with a pipette.
18. Add 500 µl of 1 x TE to the DNA pellet in the tube and incubate in a waterbath at 50°C for 15 minutes to dissolve the DNA pellet.
19. Add 2.5 µl 10 mg/ml RNAse to the tube and incubate for two hours or overnight at 37°C.
20. Add 500 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.
21. Decant all liquid and add 500 µl 70 % ice cold EtOH. Centrifuge for 10 minutes in a bench centrifuge at 4°C.
22. Remove all liquid from pellet with a pipette and dry for 15 minutes in the speedyvac.
23. Dissolve pellet in 300 µl 1 x TE at 50°C for 15 minutes.
24. Store solution at 4°C for short-term use or, at -20°C for extended periods of time.

DNA isolation protocol 2 - Standard salting out procedure (Aljanabi and Martinez, 1997)

This protocol takes one and a half days to complete and is follows:

1. Wipe the leaf material with ethanol (EtOH) and grind fresh leaf tissue to a powder with liquid nitrogen.
2. Add 3 ml lysis buffer to grindate and pour into chloroform resistant plastic centrifuge tubes. Rinse the mortar with 1 ml additional lysis buffer and add into the tubes; place tubes on ice.

3. The plant material grindate in the lysis buffer was incubated overnight at 37°C.
4. 300 µl 5 M NaCl was then added to the tubes and then shaken rigorously.
5. The tubes were then centrifuged for 15 minutes at 5 000 rpm.
6. The DNA containing supernatant was removed and transferred to another tube.
7. The last two steps (numbers 5 and 6) were repeated until the supernatant was clear.
8. The DNA was EtOH-precipitated by adding two volumes of 100 % ice-cold ethanol.
9. Tubes were mixed by inversion and centrifuged at 13 000 rpm for 15 minutes.
10. The supernatant was discarded and the pellet washed in 70 % EtOH.
11. After a final centrifugation at 13 000 rpm for 10 minutes, the final washing step (as in number 9) was repeated, the supernatant discarded, and the pellet allowed to dry.
12. The pellet was then resuspended overnight at 37°C in 20-100 µl 10 mM Tris (pH8), depending on the size of the pellet.
11. The DNA containing solutions was then stored at 4°C.

DNA isolation protocol 3 - Chelex©-100 procedure (Modified from Doyle and Doyle, 1990)

1. This procedure is a slight modification of the standard CTAB protocol (Doyle and Doyle, 1990), where step three was modified.
2. The 2 % CTAB was replaced with 20 % Chelex©-100. The protocol was then followed as previously described.

Selection and optimisation of DNA isolation protocol

The CTAB DNA isolation protocol of Doyle and Doyle (1990) proved to be the most suitable protocol as it constantly produced high quantities of high integrity DNA as revealed by 2 % agarose gel electrophoresis. Various modifications of the CTAB DNA isolation protocol were tested on five-year old leaf material in order to improve yield and DNA quality. The results of these modifications are presented in Table 2.3. The concentration of polyvinylpolypyrrolidone (PVP), a polysaccharide binder, and 2-mercapto-ethanol, which prevent sample oxidation and inhibit the release of RNAses from the tissues prior to the phenol extraction, were adjusted.

Table 2.3 **Comparison of the CTAB protocol according to Doyle and Doyle (1990) and optimised CTAB protocol.**

Reagent	Original conc. (%)	Optimal conc. (%)	Quantity
PVP	1	2	10 g
2-mercapto-ethanol	0.2	0.4	2 ml

Estimation of purity and concentration of DNA

Prior to the amplification of the DNA to produce RAPD fingerprints the DNA was assessed for purity and concentration. A Beckman DU 640 Spectrophotometer was used to determine the purity and concentration of the DNA in each sample. Samples were blanked and diluted in 1 x TE, with a dilution factor of 10 times, consisting of 5 µl sample in 45 µl 1 x TE. The purity was calculated using the formula A_{260} / A_{280} and the concentration calculated with the formula: A_{260} × dilution factor × 50 µg/ ml.

2.3.2 Generation of RAPD fingerprints

Prior to the generation of RAPD fingerprints of all the samples, a number of tests were conducted on a limited number of samples using only a few primers to determine whether the isolated DNA was of sufficiently high quality to be used in RAPD-PCR. This also involved the optimisation of the PCR conditions.

Determination of PCR quality of isolated DNA

The amplification of DNA through PCR requires good quality DNA. In order to determine the PCR ability of the extracted DNA, a number of test RAPD analyses were conducted on a few of the black wattle samples. Three Unit Standard Primers from the University of British Columbia were employed; namely Primer 182 (5' -GTT CTC GTG T-3'), Primer 332 (5' -AAC GCG TAG A-3') and Primer 409 (5' -TAG GCG GCG G-3').

Once the PCR conditions were standardised, RAPD fingerprints were generated for all the samples using 14 different primers and assessed for their ability to generate polymorphic fingerprints. One arbitrary primer was designed and 13 were selected from the Operon Technologies RAPD 10-mer primers Set A and Set C (Table 2.4). The Molecular and Cell Biology Synthetic DNA Laboratory at the University of Cape Town synthesized the primers. Working stock solutions of 100 μM were prepared with 1 x TE for all the primers.

Table 2.4 Primers used to generate RAPD fingerprints of black wattle individuals.

REFERENCE NUMBER	SEQUENCE	OD*	C_1* (μM)	V_1* (μl)	C_2* (μM)	V_2* (μl)	Vol 1 xTE*
Arbitrary primer							
P17: 03 1251	AAA CGG GCG G	171.5	1923	10	100	192.3	182.3
Operon Set A primers							
A-01: 03 0096	CAG GCC CTT C	459.5	5152	10	100	515.2	505.2
A-02: 03 0997	TGC CGA GCT G	507.6	5692	10	100	569.2	559.2
A-04: 03 0998	AAT CGG GCT G	443.7	4975	10	100	497.5	487.5
A-06: 03 0999	GGT CCC TGA C	473.4	5308	10	100	530.8	520.8
A-09: 03 1253	GGG TAA CGC C	143.9	2735	10	100	273.5	263.5
A-10: 03 1000	GTG ATC GCA G	439.1	4923	10	100	492.3	482.3
A-12: 03 1001	TCG GCG ATA G	479.1	5372	10	100	537.2	527.2
A-13: 03 1002	CAG CAC CCA C	465.5	5219	10	100	521.9	511.9
A-17: 03 1003	GAC CGC TTG T	501.6	5624	10	100	562.4	552.4
A-19: 03 1254	CAA ACG TCG G	144.6	1621	10	100	162.1	152.1
A-20: 03 1004	GTT GCC ATC C	470.0	5269	10	100	526.9	516.9
Operon Set C primers							
C-05: 03 1252	GAT GAC CGC C	195.8	2195	10	100	219.5	209.5
C-06: 03 1005	GAA CGG ACT C	401.1	4496	10	100	449.6	439.6

* OD = optical density given by the laboratory involved in the synthesis of the primers.

* C_1 = initial concentration

* C_2 = final concentration

* V_1 = initial volume

* V_2 = final volume

* VOL 1 x TE = amount of TE added to 10 μl of primer to obtain 100 μM concentration.

PCR conditions

RAPD-PCR conditions were set up according to the PCR Core Kit of Roche Diagnostics. Reagents were added together to make up a 25 µl reaction solution and consisted of: 2.5 µl 10 × Buffer (with $MgCl_2$), 0.5 µl of 200 µM dNTPs, 2.5 µl of 2.5 mM $MgCl_2$, 3.3 µl of 2 µM primer, 5 µl distilled H_2O, 10 µl DNA (100 ng) and 0.2 µl (1 unit) *Taq* polymerase. Distilled water was added to obtain a final volume of 25 µl in each PCR tube. Master solutions, consisting of all reagents excluding the template DNA, were used to speed up the process and make it more accurate. The PCR reagents, excluding the DNA template, were pipetted into an eppendorf tube and gently shaken in order to equally distribute the reagents within the tube. *Taq* polymerase was added last so that minimal hybridisation would take place between the primer sequences. Thereafter, 24 µl of the master solution was added to each PCR tube on ice, containing 10 µl of template DNA. The PCR tubes were then placed in the GENE AMP PCR 9700 machine to complete the PCR.

The PCR conditions were optimised for black wattle by testing a number of different combinations of the PCR reagents using different concentrations and volumes, as well as conditions such as temperatures. A range of concentrations, between 1 mM and 5 mM were tested in order to determine the optimum concentration of magnesium chloride of 2.5 mM. Similarly, five annealing temperatures, ranging from 34°C to 42°C, were tested to obtain the optimum annealing temperature of 37°C. The optimum DNA concentration was found to be 100 ng using 41 cycles in the PCR reaction. This combination of reagents and settings were found to be the optimum PCR conditions for black wattle and produced the brightest and most reliable bands on a 2 % agarose gel. The concentrations and combinations that generated the best result were then selected as the optimum composition for the PCR as listed in Table 2.5.

Table 2.5 Optimised PCR reaction for black wattle.

Reagents	Supplier's concentration	Optimal concentration	Volume (µl)
10 x PCR buffer	**10 x**	**1 x**	**2.5**
$MgCl_2$	**25 mM**	**2.5 mM**	**2.5**
***Taq* polymerase**	**5 U/µl**	**1 U/µl**	**0.2**
Primer	**15 µM**	**2 µM**	**3.3**
dH_2O	**-**	**-**	**6.5**
DNA template	**10 ng/µl**	**100 ng**	**10.0**

The optimised PCR cycle conditions started with an initial denaturation at 94°C, 41 cycles consisting of one minute denaturation at 94°C, two minutes annealing at 37°C, and two minutes extension at 72°C were performed prior to a final extension of four minutes at 72°C and subsequent cooling to 4°C. This PCR took just over four and a half hours.

Generation and visualisation of RAPD fingerprints

RAPD-PCR amplification products were visualised through gel electrophoresis on 2 % agarose slab gels (250 ml 0.5 x TBE) to which 6.25 µl ethidium bromide (20 mg/ml) was added. The DNA amplification products were allowed to run at 100 volts (V) for approximately five and a half hours in one litre 0.5 x TBE and 25 µl ethidium bromide (20 mg/ ml). Thereafter, the gel was placed on an ultraviolet (UV) transilluminator to visualise the bands.

During the investigation it was found that TBE, as compared to other buffers such as TAE, had a better buffering capacity for smaller fragments, such as those produced in RAPD reactions, particularly in gels running over longer periods of time. Furthermore, a 2 % agarose gel separated the RAPD fragments more effectively than gels that had a higher concentration of agarose.

Assessment of reproducibility of RAPD fingerprints

Reproducibility is one of the main concerns when employing the RAPD assay. With population investigations such as this, it is very important that the reproducibility is high so that comparisons can be conducted accurately (Bowditch et al., 1993).

In this investigation each RAPD reaction was repeated twice for each of the identified polymorphic primers and assessed for reproducibility. Gels that displayed identical fingerprints with each repeat were included in all subsequent analyses.

2.4 SCORING OF RAPD FINGERPRINTS

RAPD fingerprints are considered to be the phenotypes of a set of loci generated with a particular primer. The phenotypes of a particular locus are the fragments of a particular size produced by a particular primer and can have only two different alleles, namely, a *presence-allele* and an *absence-allele* (*null-allele*), where the *presence-allele* produces a fragment of the particular size, while the *null-allele* does not, due to mutation(s) in the annealing site(s) of the primers or intervening sequences between the primers. As RAPD markers are dominant in nature, a genotype homozygous (AA) for the *presence-allele* cannot be distinguished from a genotype that is heterozygous (Aa). A fragment of a particular size, therefore, indicates a dominant homozygote (AA) or heterozygote (Aa), while its absence, indicates the recessive genotype (aa). Figure 2.1 illustrates the presence or absence of bands in RAPD fingerprints of homozygous and heterozygous loci. Polymorphic primers therefore, produce both types of phenotypes in different individuals.

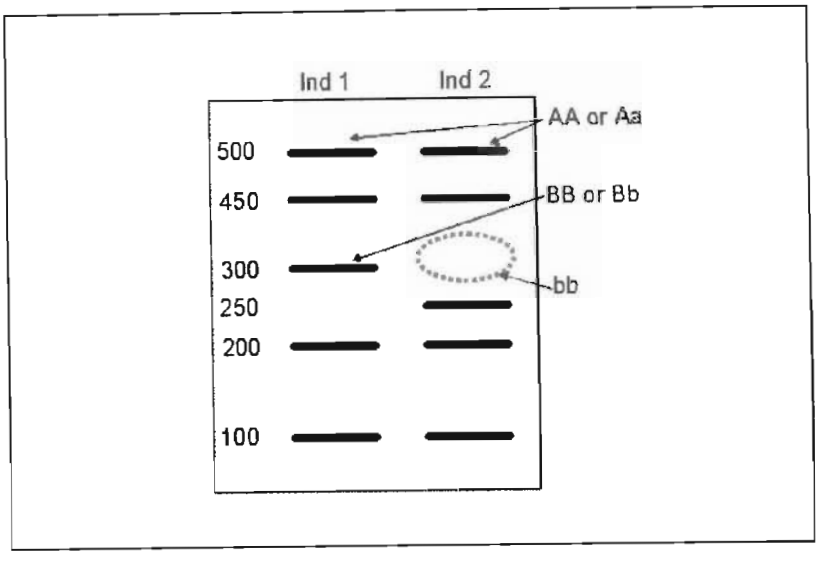


Figure 2.1 **Different phenotypes generated by different genotypes of polymorphic primers.**

2.4.1 Conversion of molecular fingerprints into digital profiles

In order to analyse the molecular fingerprints, it was necessary to convert the gel fingerprints into a digital format. Thus, before scoring the agarose gel fingerprints, suitable loci for scoring, were identified in the following manner:

- Bands had to be clear and single in nature, and
- Bands had to be reproducible.

It was found that most individuals produced reliable and reproducible bands within the size range of 3 054 bp and 298 bp. Bands larger than 3 054 bp and smaller than 298 bp were generally faint and their presence inconsistent when comparing fingerprints between repeated gels. Therefore, only loci that produced bands within this range were included in this investigation.

An individual's fingerprint was scored according to the binary number system. For each locus a value of "1" was given to a band that was present in a fingerprint, while a value of "0" was given when the particular band was absent. The resulting digital profiles for an individual of a particular primer therefore consisted of a series of ones and zeros ranging from as many as ten digits to as few as five as illustrated in Figure 2.2.

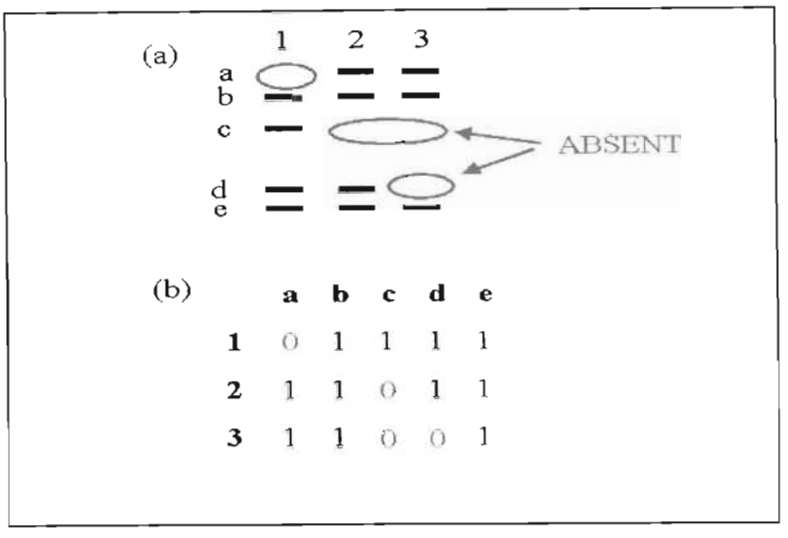


Figure 2.2 Binary scoring of RAPD fingerprints. (a) Molecular fingerprint, and (b) Corresponding binary profile.

The numerical series of zeros and ones, termed a binary profile (fingerprint), was recorded in *Notepad*, Windows' internal word processor. These digital profiles were used in all subsequent statistical analyses. Figure 2.3 displays a portion of a *Notepad* file containing digital profiles of a number of individuals.

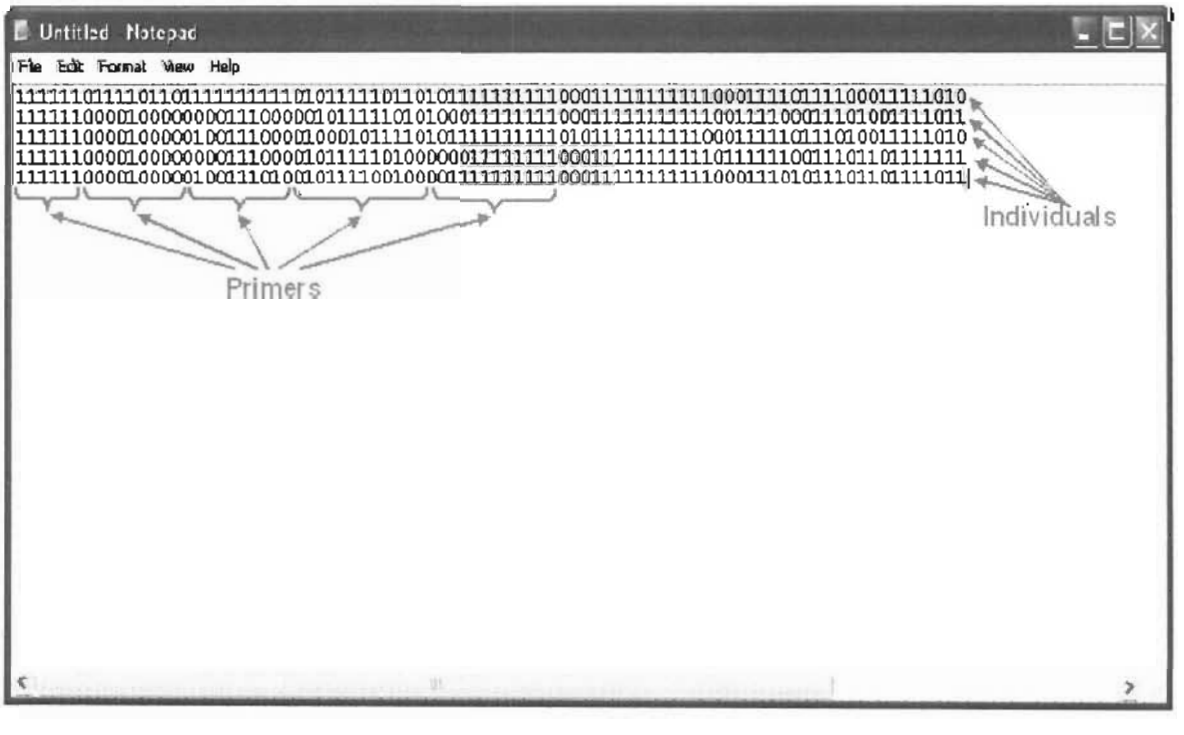


Figure 2.3 Digital fingerprint of arbitrary individuals prepared with *Notepad*.

2.5 ANALYSIS OF PRIMER PERFORMANCE

Primers that produce mostly monomorphic bands are not suitable for most molecular analyses as it indicates genetic homogeneity. It was therefore important to identify the primers that produced polymorphic bands. Therefore, an investigation was initially conducted on a limited number of individuals in order to determine which of the 14 primers were most suited to continue with in this investigation. The 14 primers were tested on a limited number of individuals that were representative of each of the families of black wattle trees.

The fourteen primers were firstly assessed for their ability to produce reproducible bands. Primers that produced clear and reproducible polymorphic bands were identified. If a band was present in all thirty five individuals, the locus was termed monomorphic, but if it were absent in some individuals, it was termed polymorphic. The number of bands and proportion of polymorphic loci of the different primers was then calculated and compared between the different primers. Primers that did not display any polymorphism were excluded from any subsequent analyses.

Secondly, the primers that were used in this investigation had different percentages of GC content, so, an analysis was performed where the effect of GC content on the number of bands and polymorphic bands were assessed and compared.

2.6 ANALYSIS OF RAPD FINGERPRINTS

Knowledge of statistical analysis of molecular data is fast becoming a vital tool to tree breeders. There are numerous statistical methods available to analyse molecular data. It is therefore important that before software is selected to analyse the data, knowledge of the various types of analyses is imperative so that suitable software is selected.

The steps to follow when analyzing molecular data include:

- Decide what type of analyses will be undertaken at the individual level and at the intra- and inter-population levels.
- Select appropriate software that would be employed in the various analyses and calculations.
- Import molecular data into selected software, and
- Analyse and interpret results.

In this investigation a number of analyses were conducted with the data. These included firstly, an assessment of possible associations between RAPD fragments and the disease gummosis, and secondly, the determination of genetic variation and relationships. Genetic variation was assessed by determining the proportion polymorphism, heterozygosity, as well as the calculation of genetic relationship statistics such as genetic distance and identity measures within and between families. An analysis of molecular variance (AMOVA) was also conducted in order to determine the contribution of the between and within population or group variation to the total variation.

2.6.1 Identification of markers linked to the disease gummosis

Markers linked to the disease complex gummosis were identified by comparing the RAPD fingerprints on the gels. The disease resistant samples were compared to the disease susceptible samples for each primer and a marker was identified as being present in the disease resistant samples and absent in the disease susceptible samples.

2.6.2 Quantification of genetic variation

There are a number of ways that are used to estimate the genetic variation between individuals of a population and between families. One of these is to calculate the amount of polymorphic loci present in the population. In this investigation the proportion of polymorphism was determined for the population as a whole as well as for each of the different families.

Another way to estimate the genetic variation between individuals of a population and between families is by calculating the amount of heterozygous loci present in the population, a preferred estimate over proportion polymorphism. Heterozygosity can be estimated in a number of different ways, each method giving a value indicative of the genetic variation in the population. As RAPD loci are dominant in nature, it was necessary to calculate the frequency of the alleles at each locus and then estimate the heterozygosity. This was done by estimating the frequency of the null-phenotype at each locus across all primers and then determining the frequencies of the alleles (p) and (q). Thereafter, the heterozygosity was estimated by applying the formula (2pq). The estimated heterozygosities were then averaged for a particular group, namely, the population and families.

Another method of quantifying genetic variation (Nei, 1978) is by determining the gene diversity and genic variation in a population as well as in families.

2.6.3 Quantification of genetic relationships

Nei's (1972) genetic identity and genetic distance, as well as Nei's (1978) unbiased genetic identity and genetic distance were calculated for the four families as well as for the population as a whole. However, as Nei's (1978) unbiased estimate of genetic distance and identity is used for sample sizes less than 50 (Orlay et al., 1997), it was decided to use the unbiased estimate in this investigation. Nei's (1978) unbiased estimate served to determine the genetic distance as well as the genetic identity between the families and individuals. From the results of the genetic identity and genetic distance measures, a dendogram depicting the relationship between the four families as well as for the population was constructed.

2.6.4 Analysis of molecular variance (AMOVA)

Analysis of molecular variance (AMOVA) is a method of estimating population differentiation directly from molecular data and testing various hypotheses about the differentiation (Excoffier et al., 1992). It is a widely used method of analyses that is included in almost every investigation in order to determine the contribution of the between and within population or group variation to the total variation.

AMOVA works on binary matrix data to create a distance matrix between samples in order to measure the genetic structure of the population from which the samples are drawn. It treats any raw molecular data, consisting of ones and zeros, where one indicates the presence of a marker and a zero its absence, as a Boolean vector. From this information Euclidean distances are calculated by subtracting the Boolean vectors of one entity from another. A matrix consisting of all pairwise squared Euclidian distances between all entities is then constructed and used in the analyses (Excoffier et al., 1992).

The constructed matrix is then employed in a hierarchical analysis of variance to determine the subdivision, where the total variance is partitioned into between populations and within populations to give an idea as to how much of the genetic variation can be attributed to each of these components.

2.6.5 Usage of software

There are various software packages available, which can be employed to analyse molecular data. In this investigation two software packages, POPGENE version 1.31 (Yeh et al., 1997) and ARLEQUIN version 2.000 (Schneider et al., 2000) were utilized.

POPGENE

POPGENE is a free user-friendly, Microsoft Window-based computer package which is specifically designed for the population genetic analysis of co-dominant as well as dominant markers, using either haploid or diploid data. Most types of population genetic

measures, such as allele and genotypic frequencies, diversity indices, neutrality tests and genetic distances could be computed using POPGENE. POPGENE was obtained from the website www.ualberta.ca/~fyeh/. Before POPGENE could be employed the digital fingerprints recorded in *Notepad* were reconfigured into an input file to contain programme specific instructions according to the programme's specifications.

The RAPD digital profiles were entered into the programme as two separate input files, one that would be used to evaluate between family population diversity and the other to compare the diversity between all individuals as well as between members of a single family as illustrated in Figure 2.4. In both cases the file consisted of a header section (specifying the title, number of populations, number of loci and the locus names) as well as a data section. The data were entered in columns representing the digital profiles obtained for each individual.

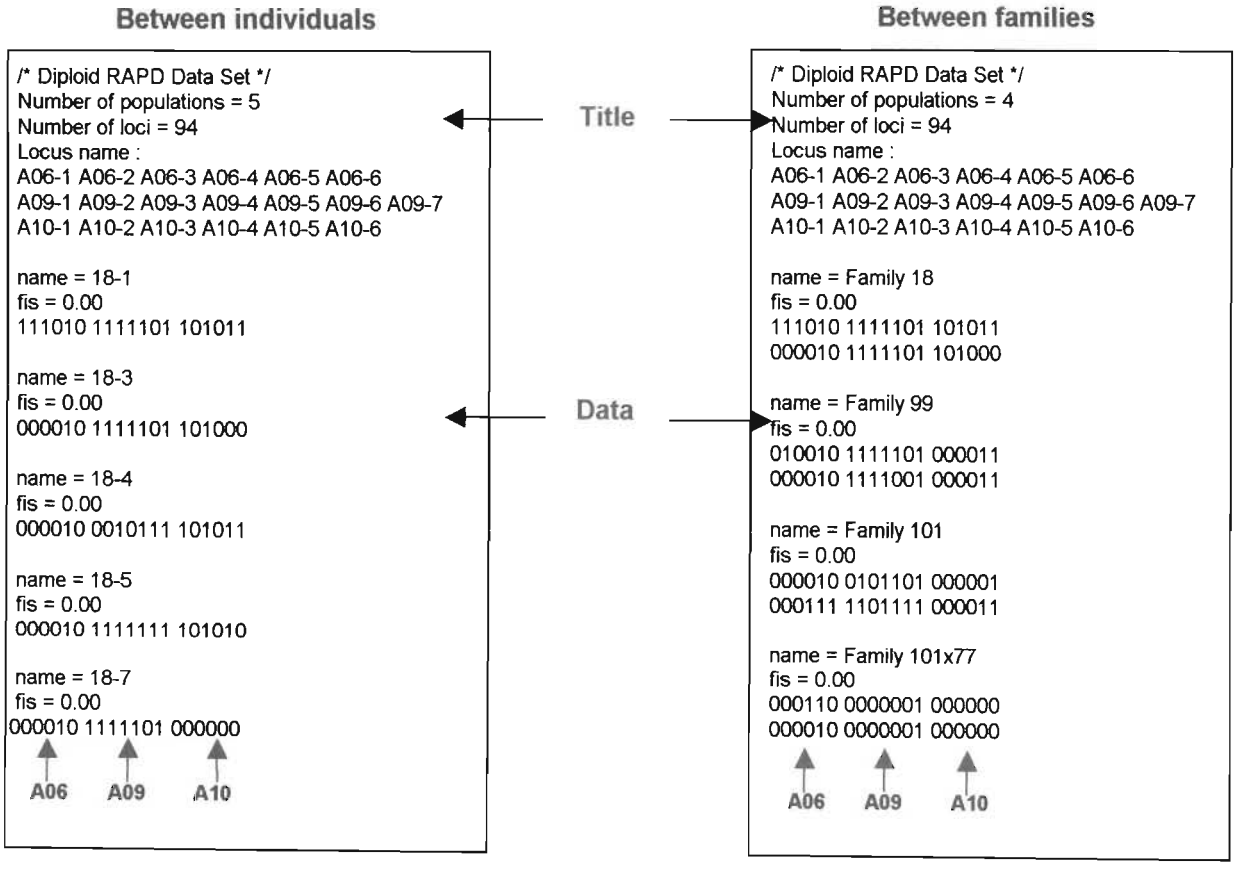


Figure 2.4 **Example of the two different types of POPGENE input files created for analyses between families and between individuals within a family.**

In the case of the between family input file, the data were grouped into four sections representing the different families and labelled accordingly. The four families were labelled: 'Family 18' for those 5 individuals originating from this family, 'Family 101' for the thirteen individuals obtained from this family, 'Family 99' for the 11 individuals from this family, and 'Family 101 x 77' for the remaining six samples where as in the case of the population as a whole, all 35 individuals were grouped together. In both of the input files, each family was given a F_{IS} value of zero to denote Hardy-Weinberg equilibrium. Once the data was loaded into the programme an analysis of the genetic variation and genetic relationship between and within families was determined using dominant marker analysis on diploid data. Examples of the input files for POPGENE are taken up in the Appendix B.

ARLEQUIN

The ARLEQUIN software package version 2.000 (Schneider et al., 2000) is the preferred software for an AMOVA analyses. It consists of a Java graphical interface that allows the user to rapidly select the analyses required to perform with the data. Many different types of raw data can be handled by ARLEQUIN, ranging from DNA sequences to microsatellite data and RFLP to allele frequency data. This package does, however, not deal with dominant marker analysis directly but can still be used because the functions originally written for RFLPs are directly applicable for dominant markers such as RAPDs and AFLPs, which also require binary data (values of "0" and "1"). The ARLEQUIN software can be obtained from the website www.unige.ch/arlequin/.

ARLEQUIN molecular data analyses are very similar, and in some cases identical, to that of POPGENE. ARLEQUIN can perform a wide range of statistical tests and measures that can be applied to data however, in this investigation, ARLEQUIN was only used to conduct an AMOVA and calculate a Eucilidian distance matrix. POPGENE was used for the rest of the statistical analyses, as this software package is more user-friendly and less challenging to the casual computer user as compared to the ARLEQUIN software package.

An input file, meeting the specific requirements for the ARLEQUIN package, was constructed using the raw score data saved in the text file created earlier. Similarly to POPGENE, the input file consisted of a profile section specifying the properties of the data such as the title of the project, type of data, number of samples and other specifications, as well as a data section containing the raw data from the text file. Following the raw data was a section detailing the structure of the data. Here the frequency of each specified profile in each family was given as well as an idea as to how the families were compiled and grouped. An example of an ARLEQUIN input file is shown in Figure 2.5. The original input file constructed for ARLEQUIN is taken up in the Appendix C.

```
[Profile]

     Title="AMOVA analysis of 4 black wattle families"
     #This project contains a sample of 4 famlies:
NbSamples=4
     #Haplotypic Data:
     GenotypicData=0
     #Each character is an Allele:
     LocusSeparator=NONE
     #No explicit haplotype constitution for the individuals
     DataType=RFLP
     #We need to compute the distance matrix
     MissingData ='?'
     CompDistMatrix=1

[Data]
[[HaplotypeDefinition]]
      HaplListName="RAPD data of four families"
      HaplList= {
 1  11111101101001000001111110 1
 2  11111100001000000011100001
 3  11111100001000001011100001
 4  11111100001000001011100001
 5  11111100001000000011100001
 6  11111100001000001011101001
 7  11111100001000000011100001
 8  11111100001000001011100001
```

```
      [[Samples]]

SampleName= "Family 18"
SampleSize=2
 SampleData=                    {
1   1
2   1
    }
 SampleName= "Family 99"
SampleSize=2
 SampleData=                    {
 3   1
  4   1
    }
 SampleName= "Family 101"
  SampleSize=2
SampleData=         {
 5    1
 6    1
        }
   SampleName= "Family 101 x 77"
   SampleSize=2
   SampleData=        {
    7   1
    8   1
              }
  [[Structure]]

 StructureName= "A single group of 4 samples"
 NbGroups = 1
Group=                {
 "Family 18"
 "Family 99"
 "Family 101"
 "Family 101 x 77"
```

Figure 2.5 An example of an ARLEQUIN input file displaying the three sections.

ARLEQUIN analyses are undertaken by creating a Project file with the extension 'arp.'. This file contains the input file with all the instructions for the analyses and may contain the data section (Procedure 1) or may have the data section in an external file, in which case it is simply referred to as (Procedure 2). Figure 2.6 illustrates the possible routes that can be followed using the procedures mentioned to create a project file.

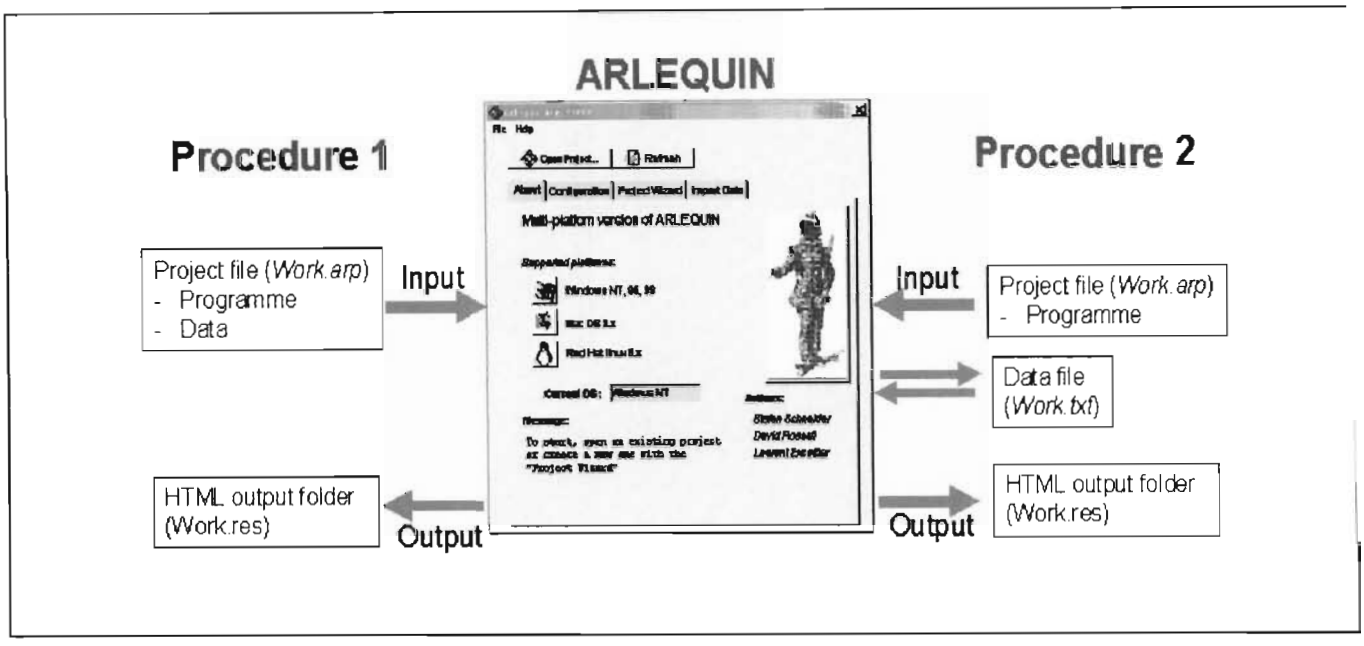


Figure 2.6 Two possible routes that can be followed to create a project file.

In this investigation, procedure 1 was followed and a project file, containing the data section, was created using the 'project wizard'. The project wizard allows the user to create an 'arp' folder in which the completed input file can be pasted. A 'mock' folder, whose contents and specifications needed not be accurate due to editing later, was thus created using project wizard. The input file was copied into this folder and the original contents, deleted. This resulted in a project file (with an 'arp' extension) containing the input file for the analysis that would be recognised by the ARLEQUIN programme.

Once the project file was saved, it was opened by ARLEQUIN using the 'Open Project' option and loaded into the programme. If it was successfully loaded, an outline of the project file and its contents were displayed in the programme window. By selecting the 'Calculation Settings' option the various statistical tests offered by ARLEQUIN were presented in the interface. In this investigation, the 'Genetic structure' calculation

containing the AMOVA analysis was selected and the appropriate settings, including whether or not a distance matrix should be computed and displayed, specified. An AMOVA analysis was conducted on the data from the specified project file, by constructing a Euclidean distance matrix, using 1 023 permutations. Wright's fixation index (F_{ST}) was also calculated by the programme. The 'Run' function was then selected, after which the specified computations were undertaken by the programme. The results were viewed in the AMOVA folder as a separate folder, with the extension '.res', and interpreted.

CHAPTER 3

RESULTS

3.1 INTRODUCTION

In this investigation the optimisation of a DNA extraction method for use in RAPD fingerprinting was undertaken in black wattle trees, to identify putative markers linked to the disease complex gummosis. Together with this research an assessment of the genetic variation and genetic relatedness of the trees used in this investigation was also undertaken.

As this was the first investigation of its kind, the PCR protocol and conditions had to be optimised for this particular species before the final fingerprints could be generated and analysed.

RAPD fingerprints were generated for all the individuals of the different families. The fingerprints were assessed for putative markers that could be linked to disease resistance against the disease gummosis. Also, the fingerprints were employed in the determination of genetic variation, genetic relatedness and family differentiation.

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

- Optimisation of a suitable DNA extraction method for black wattle.
- Optimisation of a RAPD-PCR protocol.
- Screening of 14 primers for polymorphic content.
- Analysis of primer performance.
- Identification of markers linked to the disease gummosis.
- Quantification of genetic variation within families.
- Quantification of genetic relationships between families, and
- Quantification of family differentiation.

The various statistical print outs have been taken up in the Appendix.

3.2 ASSESSMENT OF DNA ISOLATION METHODOLOGY

DNA isolation is the first, and important, step in molecular genetic studies. In black wattle, a woody species, large amounts of polysaccharides and tannins are present in cells making it difficult to separate the DNA out. Therefore, a DNA isolation protocol needs to be designed in such a way so that all contaminants are separated from the DNA. As only a limited number of molecular investigations have been undertaken with black wattle, an optimal DNA isolation protocol has not been described. Thus, the first part of this study involved the optimisation of a suitable DNA isolation method that was then assessed both quantitatively and qualitatively.

Three DNA isolation methods were assessed in this investigation. The CTAB method (Doyle and Doyle, 1990) was tested because it is a very comprehensive protocol that includes a number of reagents known to remove polysaccharides and other contaminants from plant tissue. The second method assessed was the salting out method (Aljanabi and Martinex, 1997), which is known to precipitate proteins and carbohydrates. Furthermore, salt also provides a favourable environment for the DNA as it contributes positively charged atoms that neutralise the normal negative charge of the DNA. The third method assessed was the Chelex©-100 method (modified from Doyle and Doyle, 1990) which is a DNA release method. Chelex©-100 is a binding resin that binds to cellular components that interfere with subsequent analysis by releasing the DNA.

The three DNA isolation methods were assessed for their ability to isolate both high quality and quantity of DNA from three black wattle leaf samples. The quality of the isolated DNA was assessed by running the DNA on an agarose gel and the quantity was assessed by spectrophotometry. The average purity and concentration of DNA isolated from the three samples by the three methods are shown in Table 3.1. The average purity between the three DNA isolation methods ranged from 0.64 by the Chelex©-100 method to 1.78 by the CTAB method. Likewise, the average concentration ranged from 0.014 to 0.350. The CTAB method constantly showed a much higher average of purity and concentration than the Chelex©-100 and the salting out method.

Table 3.1 Comparison of purity and concentration of DNA extraction by CTAB method, Chelex©-100 method, and salting out method.

DNA sample	DNA isolation method					
	CTAB		Chelex©-100		Salting out	
	Purity	Concentration (µg/µl)	Purity	Concentration (µg/µl)	Purity	Concentration (µg/µl)
1	1.80	0.115	0.83	0.010	1.30	0.025
2	1.85	0.125	1.00	0.013	1.10	0.020
3	1.70	0.110	0.10	0.019	1.15	0.020
Average	**1.78**	**0.350**	**0.64**	**0.014**	**1.18**	**0.022**

The quality of the isolated DNA was also assessed by gel electrophoresis to verify whether the DNA was of high molecular weight or whether it was degraded. The CTAB method produced consistently high molecular weight DNA without degradation, while the DNA isolated by the Chelex©-100 and salting out methods produced largely degraded DNA of low molecular weight.

Although three different ages of leaf samples were employed in the investigation of a suitable DNA isolation protocol, it was found that the concentration of DNA was more or less similar in all age groups investigated.

After the preliminary evaluation of the three DNA isolation methods, the CTAB method proposed by Doyle and Doyle (1990) was selected as being the most effective method in isolating DNA from black wattle leaves. However, slight optimisations of the original method were undertaken by varying the concentrations of polyvinylpolypyrrolidone (PVP) and 2-mercapto-ethanol. These reagents are vital in the binding of polysaccharides and tannins of black wattle, thereby releasing the DNA. The optimal concentrations of the CTAB protocol are presented in Table 3.2.

Table 3.2 Optimised CTAB DNA isolation method.

Reagent	Initial concentration (%)	Optimal concentration (%)	Quantity
PVP	1	2	10 g
2-mercapto-ethanol	0.2	0.4	2 ml

DNA was isolated from the different individuals by applying the CTAB method. The DNA concentrations ranged from 0.001 µg/ml to 0.00150 µg/ml with an average concentration of 0.001364 µg/ml. The purity of the isolated DNA ranged from 1.7 to a high of 2.1 with an average purity of 1.9 (Table 3.3). This purity indicated that the isolated DNA was very pure as the purity values ranged within the standard range of pure DNA; 1.8 and 2.0.

Table 3.3 Spectrophotometric readings, DNA purity, and concentrations obtained from the optimised CTAB DNA isolation method.

Individual	A260	A280	Purity	Conc. (µg/ml)
18-1	0.002858	0.001588	1.8	0.001429
18-3	0.002850	0.001425	2.0	0.001425
18-4	0.002799	0.001646	1.7	0.001400
18-5	0.003000	0.001579	1.9	0.001500
18-7	0.003010	0.001505	2.0	0.001505
99-1	0.002750	0.001310	2.1	0.001375
99-2	0.002745	0.001525	1.8	0.001373
99-3	0.002000	0.001111	1.8	0.001000
99-4	0.002200	0.001200	1.8	0.001100
99-6	0.002740	0.001612	1.7	0.001370
99-8	0.002699	0.001499	1.8	0.001349
99-9	0.002857	0.001587	1.8	0.001429
99-10	0.002853	0.001585	1.8	0.001427
99-11	0.002845	0.001422	2.0	0.001422
99-12	0.002849	0.001425	2.0	0.001425
99-13	0.002840	0.001495	1.9	0.001420
101-1	0.002839	0.001494	1.9	0.001419
101-2	0.002830	0.001348	2.1	0.001415
101-3	0.002765	0.001455	1.9	0.001382
101-4	0.002350	0.001300	1.8	0.001175
101-5	0.002833	0.001349	2.1	0.001417
101-7	0.002835	0.001668	1.7	0.001418
101-8	0.002837	0.001669	1.7	0.001419
101-9	0.002836	0.001576	1.8	0.001418
101-10	0.002749	0.001527	1.8	0.001374
101-11	0.002763	0.001382	2.0	0.001382
101-12	0.002760	0.001453	1.9	0.001380
101-13	0.002863	0.001507	1.9	0.001432
101-14	0.002731	0.001517	1.8	0.001366
101x77-1	0.002352	0.001307	1.8	0.001176
101x77-2	0.002743	0.001372	2.0	0.001372
101x77-3	0.002769	0.001538	1.8	0.001384
101x77-4	0.002746	0.001373	2.0	0.001373
101x77-5	0.002100	0.001150	1.8	0.001050
101x77-8	0.002860	0.001505	1.9	0.001430
Mean			**1.9**	**0.001364**

3.3 GENERATION OF RAPD FINGERPRINTS

RAPD-PCR is species specific and is dependent on obtaining the optimal combination of reagents and conditions. It is therefore important to optimise these reagent concentrations and PCR settings when analysing a species for the first time. In this investigation the DNA, primer and magnesium chloride ($MgCl_2$) concentrations of the PCR reaction mixture was optimised by varying the concentrations of one particular reagent, whilst keeping the other reagents in the reaction mixture constant. The resulting RAPD amplification products were then separated on a 2 % agarose gel and the fingerprints evaluated for clear and singular bands.

Several DNA concentrations, ranging from 50 ng/µl to 100 ng/µl were assessed. It is very important to determine the optimal amount of template, as too little DNA may result in faint or no bands at all, whereas too much DNA may produce smeared bands. It was found that 100 ng/µl of genomic DNA produced the most consistent and brightest DNA fingerprints for this particular species. Likewise, the optimal primer concentration was found to be 2 µM and for $MgCl_2$ 2.5 mM. Table 3.4 presents the different options tested to generate RAPD fingerprints; the DNA, primer, and $MgCl_2$, and the optimised concentrations are indicated in italics.

Table 3.4 Options tested to determine optimised concentrations for PCR.

Content	Options		
	1	2	3
DNA (ng/µl)	50	75	*100*
Primer (µM)	1	*2*	4
$MgCl_2$ (mM)	1	*2.5*	5

The optimisation of PCR settings to amplify DNA to produce RAPD fingerprints was undertaken by varying the annealing temperature. Temperatures ranging from 29°C to 41°C were assessed and the optimum annealing temperature was found to be 37°C. The optimum temperature was selected according to the quality of the

fingerprint generated by comparing which temperature produced the most reproducible and clear bands. Figure 3.1 presents a representation of gels displaying poor amplification and high quality amplification.

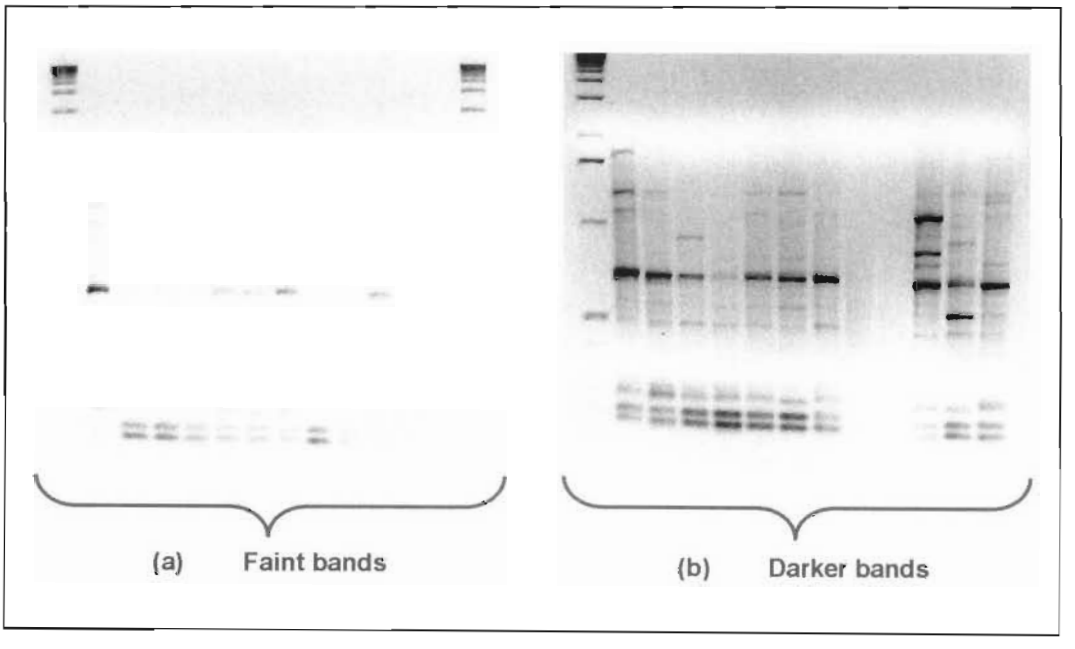


Figure 3.1 Representation of DNA fingerprints showing (a) poor quality amplification, and (b) high quality amplification.

The optimised PCR settings used to amplify DNA to produce RAPD fingerprints in this investigation are presented in Figure 3.2.

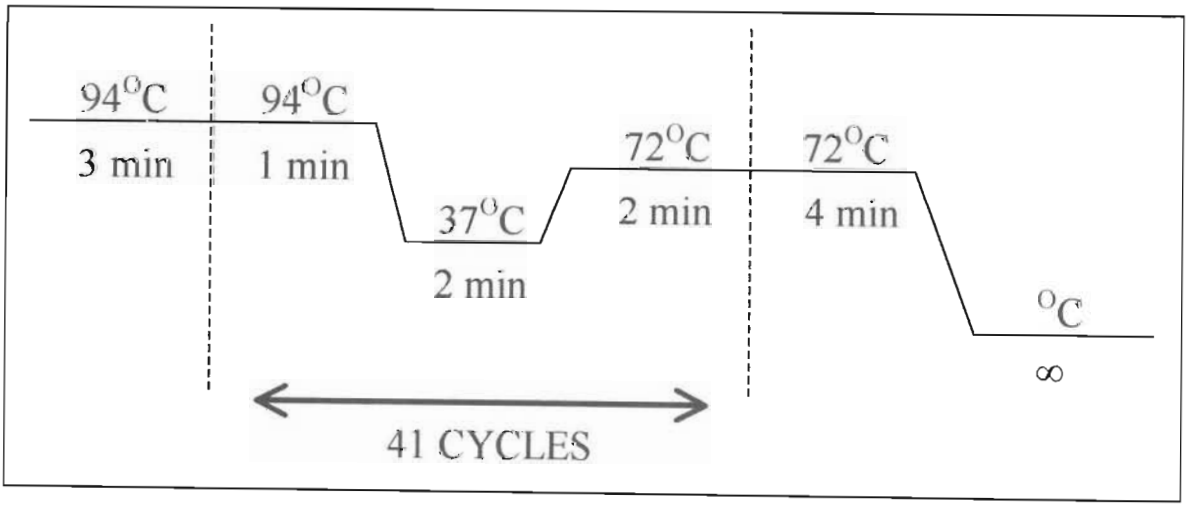


Figure 3.2 Representation of optimised RAPD-PCR settings and cycle numbers.

3.4 ANALYSIS OF RAPD FINGERPRINTS

3.4.1 Identification of suitable loci

The first step undertaken in the analysis of RAPD fingerprints was to identify those loci that were suitable to use in all subsequent analyses. Suitable loci were selected according to how clear, repeatable and consistent the bands they produced were. A total of 94 loci produced by the 14 primers were consistently clear and repeatable and thus suitable for all subsequent analyses.

3.4.2 Scoring of RAPD fingerprints

The molecular RAPD fingerprints generated by agarose gel electrophoresis where converted into digital profiles. A digital profile was representative of the phenotype. The presence of a band indicated the dominant phenotype of the genotypes AA or Aa, and was scored as a "1". The absence of a band indicated the recessive phenotype of the genotype aa, scored as a "0". Fingerprint scores were recorded in *Notepad*, Windows' internal word processor, by listing the loci in a specific order for each primer which was also sequentially arranged (Figure 3.3).

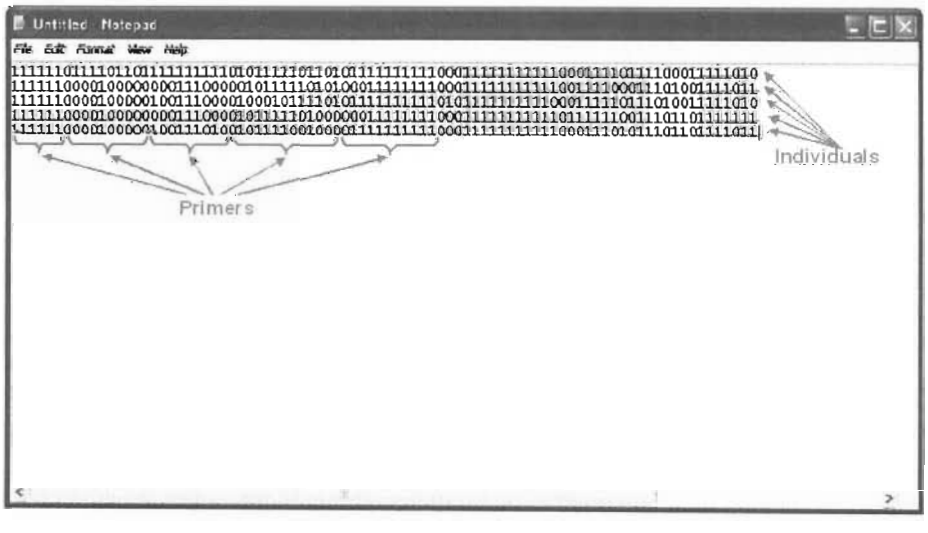


Figure 3.3 Example of digital profiles scoring of five black wattle individuals.

The phenotypic digital profiles of all individuals assessed in this investigation are represented in Table 3.5.

Table 3.5 Digital profiles of the phenotypes of each individual showing the presence-phenotype and the null-phenotype of each of the 94 loci.

Individuals	Primers													
	A-01	A-02	A-04	A-06	A-09	A-10	A-12	A-13	A-17	A-19	A-20	C-05	C-06	P17
18-1	111111	0111101101111	111	111010	1111101	101011	11111111	00011	1111	11111	000111	10111100	0111110	1000101111
18-3	111111	0000100000000	111	000010	1111101	101000	11111111	00011	1111	11111	001111	00011101	0011110	1100101111
18-4	111111	0000100000100	111	000010	0010111	101011	11111111	00011	1111	11111	000111	11011101	0011111	0100101111
18-5	111111	0000100000100	111	000010	1111111	101010	11111111	00011	1111	11111	000111	00011100	0011111	0100101101
18-7	111111	0000100000000	111	000010	1111101	000000	11111111	00011	1111	11111	011111	10011101	1011111	1100101101
99-1	111111	0000100000100	111	010010	1111001	000011	11111111	00011	1111	11111	000111	01011101	1011110	1100100001
99-2	111111	0000100000000	111	000010	1111001	000011	11111111	00011	1111	11111	000111	01011101	0011111	0001101101
99-3	111111	0000100000100	111	000010	0100001	000001	11111111	00011	1111	11111	000101	10011110	0000110	0010100101
99-4	111111	0000100000100	111	000010	0100001	001011	11111111	00011	1111	11111	000101	10011110	0000110	0010100101
99-6	111111	0100100011000	111	001110	1111101	000101	11111111	00111	1111	11111	000111	10011100	0011111	1100100111
99-8	111111	0000100000100	111	000010	1111001	010000	11111111	00111	1111	11111	000111	00011111	0011110	1100101111
99-9	111111	0000100000000	111	000010	1111101	000011	11111111	00111	1111	11111	000111	00011101	0011111	1100101011
99-10	111111	0000110000010	111	000010	1111111	000011	11111111	10011	1111	11111	000111	11011101	0001110	1100101001
99-11	111111	0000110010000	111	000010	0101101	010000	11111111	00011	1111	11111	000111	11011100	0001110	0100111011
99-12	111111	0000110000010	111	000010	1111101	000101	11111111	11011	1111	11111	000111	10011101	1001111	1100100001
99-13	111111	1000110010010	111	000111	1111101	010011	11111111	10011	1111	11111	000111	11011101	1001110	1100101011
101-1	111111	0100100001000	111	000010	0101101	000001	11111111	00011	1111	11111	000111	11011100	0010110	0000101001
101-2	111111	0001110000010	111	000111	1101111	000011	11111111	10011	1111	11111	100111	11011101	1001111	1100101011
101-3	111111	0000100000010	111	000010	0101101	000011	11111111	00011	1111	11111	000101	00011101	0001111	0001100001
101-4	111111	0000100000001	111	000010	0100011	000000	11111111	00011	1111	11111	000101	10011111	0000110	0010101001
101-5	111111	0000110010001	111	000011	1111101	010010	11111111	10011	1111	11111	000111	11011101	1001111	1100111001
101-7	111111	0000110000001	111	000011	1111101	010011	11111111	10011	1111	11111	000111	11011101	1001111	0100101001
101-8	111111	0000110000001	111	000011	1111101	010000	11111111	10011	1111	11111	000111	11011101	1001111	1100101011
101-9	111111	0000110010000	111	001111	1111101	010011	11111111	00011	1111	11111	010111	01011101	1011110	1100111001
101-10	111111	1000101000000	111	000110	1101101	001001	11111111	00011	1111	11111	000111	11011101	0101111	1110110111
101-11	111111	0000100000000	111	000010	0000001	000000	11111111	00011	1111	11111	000101	10011101	0000110	0001100111
101-12	111111	0100100101000	111	000111	1101101	001001	11111111	00011	1111	11111	000111	11011100	0101110	1100110111
101-13	111111	0000100000000	111	000011	1101101	000001	11111111	00011	1111	11111	000111	01011101	0001111	0001110111
101-14	111111	0000100000000	111	000011	1101101	000001	11111111	00011	1111	11111	000111	01011101	0001111	1001110101
101x77-1	111111	0000100000100	111	000011	0011101	000010	11111111	00011	1111	11111	000101	01011101	1001110	0000100001
101X77-2	111111	0000100000000	111	000110	0000001	000000	11111111	00011	1111	11111	000101	00011101	0000110	0001100111
101X77-3	111111	0000100000000	111	000010	0000001	000000	11111111	00011	1111	11111	000111	00011101	0000110	0000100111
101X77-4	111111	0000100000100	111	000110	1000101	000010	11111111	00011	1111	11111	000111	10011100	0000111	0000100111
101X77-5	111111	0000100000100	111	000010	0000001	000000	11111111	00011	1111	11111	000101	10011101	0000110	0000110111
101X77-8	111111	0001100000100	111	000110	1111101	001001	11111111	10011	1111	11111	000111	01011101	1001111	1100110111

*Numbers 18, 99, 101 and 101x77 refer to the different families where families 18 (resistant family), 99 and 101 are half-sibs and family 101x77 are full-sibs.

3.4.3 Primer performance

The performance of the 14 primers employed in this investigation was assessed for their ability to amplify polymorphic loci by evaluating all the fingerprints. The monomorphic as well as polymorphic loci were identified for each of the primers. Primers were named polymorphic if some of the loci produced bands that were present in some individuals and absent in others, while primers were named monomorphic if all the loci produced bands in all the individuals. Figure 3.4 displays a typical black wattle RAPD fingerprint, with monomorphic and polymorphic loci.

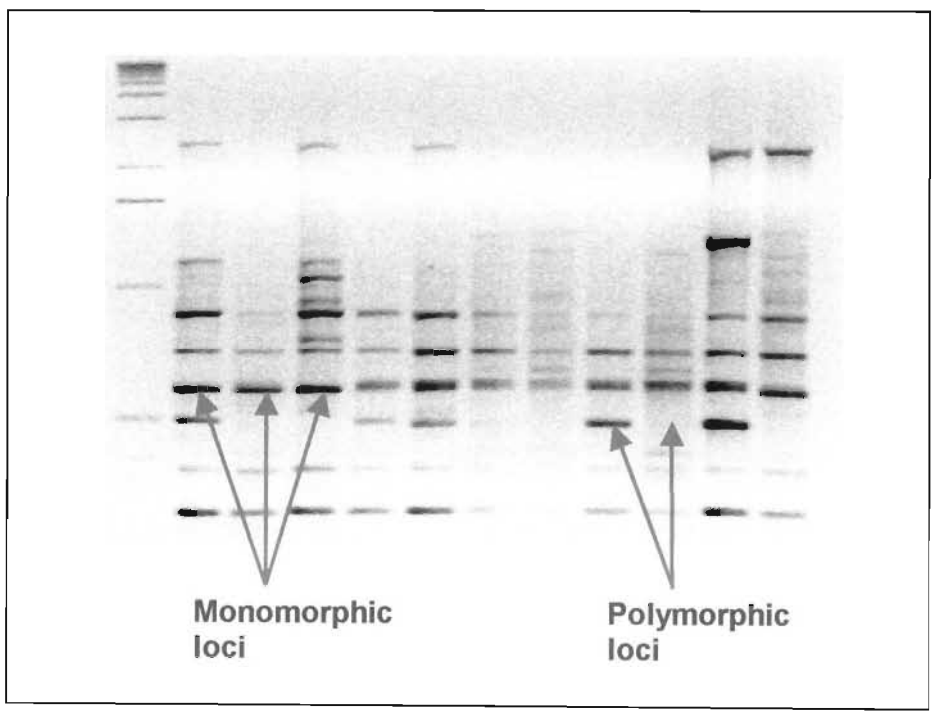


Figure 3.4 Representation of typical fingerprints showing monomorphic and polymorphic loci.

In this investigation it was found that five of the 14 primers were monomorphic and thus excluded from subsequent analyses, leaving 68 analysable loci. The remaining nine polymorphic primers produced between three and 12 polymorphic loci (Table 3.6).

Table 3.6 **Primer nucleotide sequence, GC content, number of loci produced, and number and percentage of monomorphic and polymorphic loci.**

Primer	Nucleotide sequence	GC content (%)	Number of loci observed	Number of monomorphic loci	Number of polymorphic loci	Percentage of monomorphic loci	Percentage of polymorphic loci
*A-01	CAG GCC CTT C	70	-	-	-	-	-
A-02	TGC CGA GCT G	70	13	1	12	8	92
*A-04	AAT CGG GCT G	60	-	-	-	-	-
A-06	GGT CCC TGA C	70	6	1	5	17	83
A-09	GGG TAA CGC C	70	7	1	6	14	86
A-10	GTG ATC GCA G	60	6	0	6	0	100
*A-12	TCG GCG ATA G	60	-	-	-	-	-
A-13	CAG CAC CCA C	70	5	2	3	40	60
*A-17	GAC CGC TTG T	60	-	-	-	-	-
*A-19	CAA ACG TCG G	60	-	-	-	-	-
A-20	GTT GCC ATC C	60	6	3	3	50	50
C-05	GAT GAC CGC C	70	8	3	5	37	63
C-06	GAA CGC ACT C	60	7	2	5	29	71
P17	AAA CGG GCG G	70	10	2	8	20	80
Total			**68**	**15**	**53**	**22**	**78**

*Monomorphic primer.

In order to establish whether or not the GC content of the primers affected the number of loci amplified in a fingerprint, the number of amplified loci of the primers with different GC contents was compared. From Table 3.6 it became clear that no particular relationship could be established between the percentage GC content of the primers and the number of bands they produced. It should however be mentioned that the primers were originally selected on the basis of having a relatively high GC content (≥ 60 %) because it has been genetically found that primers with an increased GC content has a greater probability to generate bands than primers with a lower GC content (Benita et al., 2003).

3.5 ASSESSMENT OF GENETIC VARIATION

3.5.1 RAPD markers associated with gummosis

RAPD markers linked to the disease complex gummosis in black wattle were investigated by comparing the RAPD fingerprints of the disease resistant individuals (Family 18) to the fingerprints of the disease susceptible individuals. If a band is found to be present in the fingerprints of the disease resistant individuals, but absent from the fingerprints of susceptible individuals, this could be identified as a putative disease resistant marker.

In this investigation, no band was present only in the resistant individuals. In the same way, no specific band was present only in the susceptible individuals. However, one band specific to one resistant individual was identified (Figure 3.5).

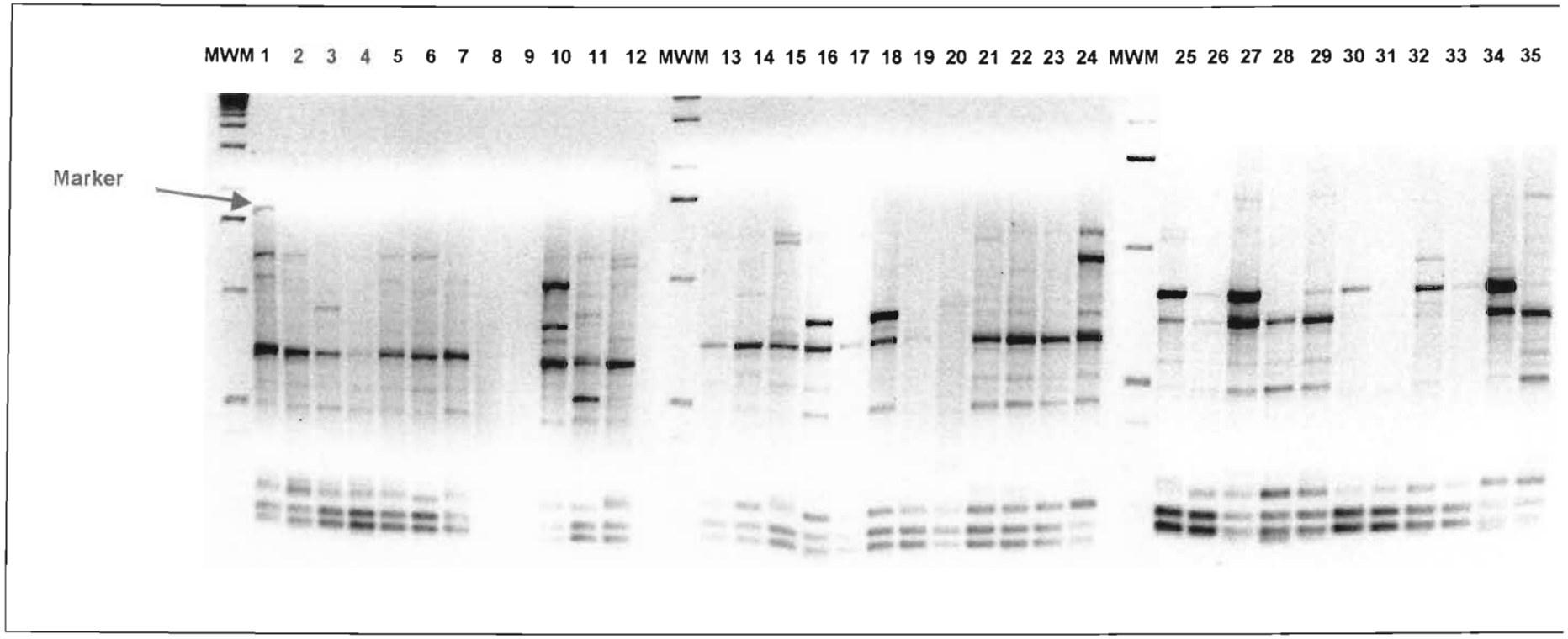


Figure 3.5 **DNA fingerprint showing marker linked to the disease gummosis, which is present in the disease resistant individual (in blue) but absent in the disease susceptible individuals (black).**

3.5.2 Quantification of genetic variation

There are a number of different methods of quantifying genetic variation within a population or family group. In this investigation the quantification of genetic variation was undertaken by calculating the percentage polymorphic loci, Nei's gene diversity (1973), Shannon's Information Index (Lewontin, 1972), and heterozygosity, which were determined for the population, as well as for each family individually.

The three measures of genetic variation, polymorphic percentage, Nei's diversity and Shannon's information index, clearly demonstrated a rather limited amount of genetic variation in this population, which was also reflected by each of the families (Table 3.7). It is interesting to note that when the polymorphic content was calculated, the population returned the highest percentage (58.51%) polymorphic loci. Interestingly, family 101 displayed the highest values for Nei's gene diversity and Shannon's Information Index. The mean for Nei's gene diversity for all families ranged from 0.1064±0.1754 to 0.1556±0.1887 and for the population was 0.1405±0.1789. Likewise, the mean for Shannon's Information Index ranged from 0.1589±0.2574 to 0.2364±0.2738 and for the population was 0.2213±0.2540.

Table 3.7 Various measures for the quantification of genetic variation.

Group	Polymorphic loci (%)	Nei's gene diversity		Shannon's Information Index	
		(%)	mean±*SD	(%)	mean±(1*SD)
Family:					
18	28.72	10.89	0.1089 ± 0.1775	16.21	0.1621 ±0.2606
99	43.62	14.98	0.1498 ± 0.1900	22.61	0.2261 ± 0.2759
101	46.81	15.56	0.1556 ± 0.1887	23.64	0.2364 ± 0.2738
101x77	28.72	10.64	0.1064 ± 0.1754	15.89	0.1589 ± 0.2574
Whole population	58.51	14.05	0.1405 ± 0.1789	22.13	0.2213 ± 0.2540

*Standard Deviation

As expected, the only family containing full-sibs (101x77) contained consistently the least amount of genetic variation (Nei's gene diversity = 10.64 % and Shannon's Information Index = 15.89 %). Second to family 101x77 was family 18. The limited amount of genetic variability for family 18 was expected for a half-sib family; however,

the small number of individuals in this family probably contributed to this low value as well.

It was interesting to find that the two larger families, 99 and 101, both displayed nearly double the number of polymorphic loci than the two smaller families, probably indicating the need to evaluate larger families, even with the large number of loci investigated (68). This result was also supported by Nei's gene diversity and Shannon's Information Index.

The heterozygosity (2pq) of the population was estimated by firstly determining the frequency of the null-phenotype at each locus across all primers. Thereafter, the frequency of the *null-allele* (q) was calculated by computing the square root of the frequency of the null-phenotype. The frequency of the *presence-allele* (p) was then estimated by applying the formula p = 1-q. Thereafter, the frequency of the heterozygotes (2pq) was estimated for each locus, after which the mean was calculated for the population (Table 3.8). The mean estimated heterozygosity of this population was only 24 %, supporting the notion that genetic variation is rather low. Although this is expected for such closely related individuals, it was not expected to be so across the population. The lowest percentage estimated at a particular locus was 4 % (2 loci) and the highest 50 % (4 loci) (Table 3.8).

When the mean heterozygosity of the loci of the different primers was quantified separately, it was found that the average heterozygosity ranged from 11 % to 39 %, affirming the importance of the pre-evaluation of primers.

Table 3.8 Estimated heterozygosity for the population.

Locus	Null-phenotype frequency (q^2)	Estimated *null-allele* frequency (q)	Estimated *presence-allele* (p)	Estimated heterozygosity (2pq)
A02-1	0.9409	0.97	0.03	**0.06**
A02-2	0.8836	0.94	0.06	**0.11**
A02-3	0.9409	0.97	0.03	**0.06**
A02-4	0.9409	0.97	0.03	**0.06**
A02-5	0.7396	0.86	0.14	**0.24**
A02-6	0.8836	0.94	0.06	**0.11**
A02-7	0.9409	0.97	0.03	**0.06**
A02-8	0.8649	0.93	0.07	**0.13**
A02-9	0.9095	0.95	0.05	**0.10**
A02-10	0.7056	0.84	0.16	**0.27**
A02-11	0.8649	0.93	0.07	**0.13**
A02-12	0.8836	0.94	0.06	**0.11**
Mean heterozygosity - Primer A02				**0.12**
A06-1	0.9604	0.98	0.02	**0.04**
A06-2	0.9409	0.97	0.03	**0.06**
A06-3	0.9095	0.95	0.05	**0.10**
A06-4	0.7396	0.86	0.14	**0.24**
A06-5	0.7056	0.84	0.16	**0.27**
Mean heterozygosity - Primer A06				**0.14**
A09-1	0.3364	0.58	0.42	**0.49**
A09-2	0.2025	0.45	0.55	**0.50**
A09-3	0.4624	0.68	0.32	**0.44**
A09-4	0.2601	0.51	0.49	**0.25**
A09-5	0.2601	0.51	0.49	**0.50**
A09-6	0.8649	0.93	0.07	**0.13**
Mean heterozygosity - Primer A09				**0.39**
A10-1	0.8836	0.94	0.06	**0.11**
A10-2	0.7921	0.89	0.11	**0.20**
A10-3	0.7744	0.88	0.12	**0.11**
A10-4	0.9409	0.97	0.03	**0.06**
A10-5	0.5329	0.73	0.27	**0.39**
A10-6	0.3969	0.63	0.37	**0.47**

Locus	Null-phenotype frequency (q^2)	Estimated *null-allele* frequency (q)	Estimated *presence-allele* (p)	Estimated heterozygosity (2pq)
Mean heterozygosity - Primer A10				**0.22**
A13-1	0.7744	0.88	0.12	**0.21**
A13-2	0.9409	0.97	0.03	**0.06**
A13-3	0.9095	0.95	0.05	**0.10**
Mean heterozygosity - Primer A13				**0.12**
A20-1	0.9095	0.95	0.05	**0.10**
A20-2	0.9409	0.97	0.03	**0.06**
A20-3	0.9409	0.97	0.03	**0.06**
A20-4	0.2601	0.51	0.49	**0.50**
Mean heterozygosity - Primer A20				**0.18**
C05-1	0.3969	0.63	0.37	**0.47**
C05-2	0.4900	0.70	0.30	**0.42**
C05-3	0.9604	0.98	0.02	**0.04**
C05-4	0.8836	0.94	0.06	**0.11**
C05-5	0.2601	0.51	0.49	**0.50**
Mean heterozygosity - Primer C05				**0.31**
C06-1	0.6889	0.83	0.17	**0.28**
C06-2	0.9095	0.95	0.05	**0.10**
C06-3	0.6561	0.81	0.19	**0.31**
C06-4	0.2601	0.51	0.49	**0.50**
C06-5	0.5041	0.71	0.29	**0.41**
Mean heterozygosity - Primer C06				**0.32**
P17-1	0.4900	0.70	0.30	**0.30**
P17-2	0.4356	0.66	0.34	**0.45**
P17-3	0.8836	0.94	0.06	**0.11**
P17-4	0.8281	0.91	0.09	**0.16**
P17-5	0.7396	0.86	0.14	**0.24**
P17-6	0.4900	0.70	0.30	**0.42**
P17-7	0.4356	0.66	0.34	**0.45**
P17-8	0.4624	0.68	0.32	**0.44**
Mean heterozygosity - Primer P17				**0.32**
Average heterozygosity for the population				***0.24***

3.5.3 Quantification of genetic relationships

Genetic relationships between families and among individuals within families were quantified according to Nei's (1978) unbiased estimate of genetic identity and distance by means of a pairwise analysis. Figure 3.6 displays matrices of identity and distance measures that were estimated with genetic identity (I) above the diagonal and distance (D) below the diagonal.

In a pairwise analysis of genetic differences (diversity) and similarities (identity) between individuals within the different families revealed that in general all identities were high, between 71 % and 98 %. It was interesting to note that the identity ranges were narrow for the two small families; 80 % to 92 % for family 18 and 80 % to 97 % for family 101x77. The two larger families, on the other hand, displayed a larger range in identity comparisons; 72 % to 98 % for family 99 and 71 % to 98 % for family 101.

(a) Family 18

Individual identification	18-1	18-3	18-4	18-5	18-7
18-1	****	0.8511	0.8085	0.8298	0.7979
18-3	0.1613	****	0.8723	0.9149	0.9255
18-4	0.2126	0.1366	****	0.9149	0.8404
18-5	0.1866	0.0889	0.0889	****	0.8830
18-7	0.2258	0.0774	0.1738	0.1245	****

(b) Family 101x77

Individual identification	101x77-2	101x77-3	101x77-4	101x77-5	101x77-8	101x77-1
101x77-2	****	0.9681	0.9043	0.9468	0.7979	0.8511
101x77-3	0.0324	****	0.9149	0.9574	0.8085	0.8617
101x77-4	0.1006	0.0889	****	0.9149	0.9298	0.8404
101x77-5	0.0547	0.0435	0.0889	****	0.8085	0.8617
101x77-8	0.2258	0.2126	0.1866	0.2126	****	0.8191
101x77-1	0.1613	0.1488	0.1738	0.1488	0.1995	****

(c) Family 99

Individual identification	99-1	99-2	99-3	99-4	99-6	99-8	99-9	99-10	99-11	99-12	99-13
99-1	****	0.8936	0.8085	0.8085	0.8191	0.8723	0.9149	0.8936	0.8191	0.8723	0.8511
99-2	0.1125	****	0.8298	0.8298	0.8191	0.8723	0.9149	0.8723	0.8191	0.8298	0.8085
99-3	0.2126	0.1866	****	0.9787	0.7766	0.8298	0.7872	0.7872	0.8191	0.7872	0.7234
99-4	0.2126	0.1866	0.0215	****	0.7553	0.8085	0.7872	0.7872	0.7979	0.7660	0.7234
99-6	0.1995	0.1995	0.2528	0.2806	****	0.8404	0.8830	0.7979	0.8085	0.8404	0.7979
99-8	0.1366	0.1366	0.1866	0.2126	0.1738	****	0.9149	0.8298	0.8404	0.8085	0.8085
99-9	0.0889	0.0889	0.2392	0.2392	0.1245	0.0889	****	0.8936	0.8404	0.8723	0.8511
99-10	0.1125	0.1366	0.2392	0.2392	0.2258	0.1866	0.1125	****	0.8617	0.9149	0.9149
99-11	0.1995	0.1995	0.1995	0.2258	0.2126	0.1738	0.1738	0.1488	****	0.8191	0.8617
99-12	0.1366	0.1866	0.2392	0.2666	0.1738	0.2126	0.1366	0.0889	0.1995	****	0.8723
99-13	0.1613	0.2126	0.3238	0.3238	0.2258	0.2126	0.1613	0.0889	0.1488	0.1366	****

(d) Family 101

Individual identification	101-1	101-2	101-3	101-4	101-5	101-7	101-8	101-9	101-10	101-11	101-12	101-13	101-14
101-1	****	0.7872	0.8617	0.8617	0.7872	0.8298	0.8085	0.7979	0.7979	0.8511	0.8511	0.8511	0.8511
101-2	0.2392	****	0.8191	0.7553	0.8723	0.8936	0.8936	0.8404	0.8191	0.7447	0.8085	0.8298	0.8298
101-3	0.1488	0.1995	****	0.8511	0.7979	0.8404	0.7979	0.7872	0.8085	0.8830	0.7766	0.9043	0.9043
101-4	0.1488	0.2806	0.1613	****	0.7766	0.7979	0.7979	0.7234	0.7660	0.9043	0.7340	0.7979	0.7979
101-5	0.2392	0.1366	0.2258	0.2528	****	0.9574	0.9574	0.9043	0.7979	0.7447	0.7872	0.8298	0.8511
101-7	0.1866	0.1125	0.1738	0.2258	0.0435	****	0.9574	0.8830	0.7979	0.7660	0.7872	0.8511	0.8511
101-8	0.2126	0.1125	0.2258	0.2258	0.0435	0.0435	****	0.8617	0.8191	0.7872	0.8085	0.8511	0.8511
101-9	0.2258	0.1738	0.2392	0.3238	0.1006	0.1245	0.1488	****	0.7872	0.7128	0.7979	0.8191	0.8404
101-10	0.2258	0.1995	0.2126	0.2666	0.2258	0.2258	0.1995	0.2392	****	0.7979	0.9043	0.8830	0.8830
101-11	0.1613	0.2948	0.1245	0.1006	0.2948	0.2666	0.2392	0.3386	0.2258	****	0.7872	0.8723	0.8511
101-12	0.1613	0.2126	0.2528	0.3092	0.2392	0.2392	0.2126	0.2258	0.1006	0.2392	****	0.8723	0.8723
101-13	0.1613	0.1866	0.1006	0.2258	0.1866	0.1613	0.1613	0.1995	0.1245	0.1366	0.1366	****	0.9787
101-14	0.1613	0.1866	0.1006	0.2258	0.1613	0.1613	0.1613	0.1738	0.1245	0.1613	0.1366	0.0215	****

Figure 3.6 **Genetic identity (above diagonal) and distance (below diagonal) measures within families ((a) – (d)).**

A comparison of genetic identity between the four families also displayed extensive genetic similarity ranging from 92 % to 99 % (Figure 3.7). It was expected that within each of the families the genetic variation would be low, however, it was an unexpected result that the families were so closely related.

Table 3.9 Genetic identity (above diagonal) and distance (below diagonal) measures between families.

Family Identity	18	99	101	101x77
18	****	0.9536	0.9423	0.9227
99	0.0475	****	0.9903	0.9662
101	0.0594	0.0098	****	0.9721
101x77	0.0804	0.0344	0.0283	****

These data clearly support the notion that the genetic base in South African black wattle is low as was estimated by the polymorphic percentage, Nei's genetic diversity, Shannon's Information Index, and the estimated heterozygosity.

A popular way to display genetic relationships estimated by Nei's genetic identity and genetic distance is through a dendogram. The dendograms in this investigation were based upon Nei's (1978) unbiased genetic distance using the unweighted pair group method with arithmetic mean (UPGMA) method and were modified from NEIGHBOR procedure of PHYLIP version 3.5 (Sneath and Snokal, 1973). The trees provided a representation of the overall genetic relationships between individuals within a family and between families, where the branch lengths indicates an estimate of frequency differences (Figure 3.7). The branch lengths of the different trees are all supportive of the closeness of the genetic relationship amongst this collection of black wattle individuals.

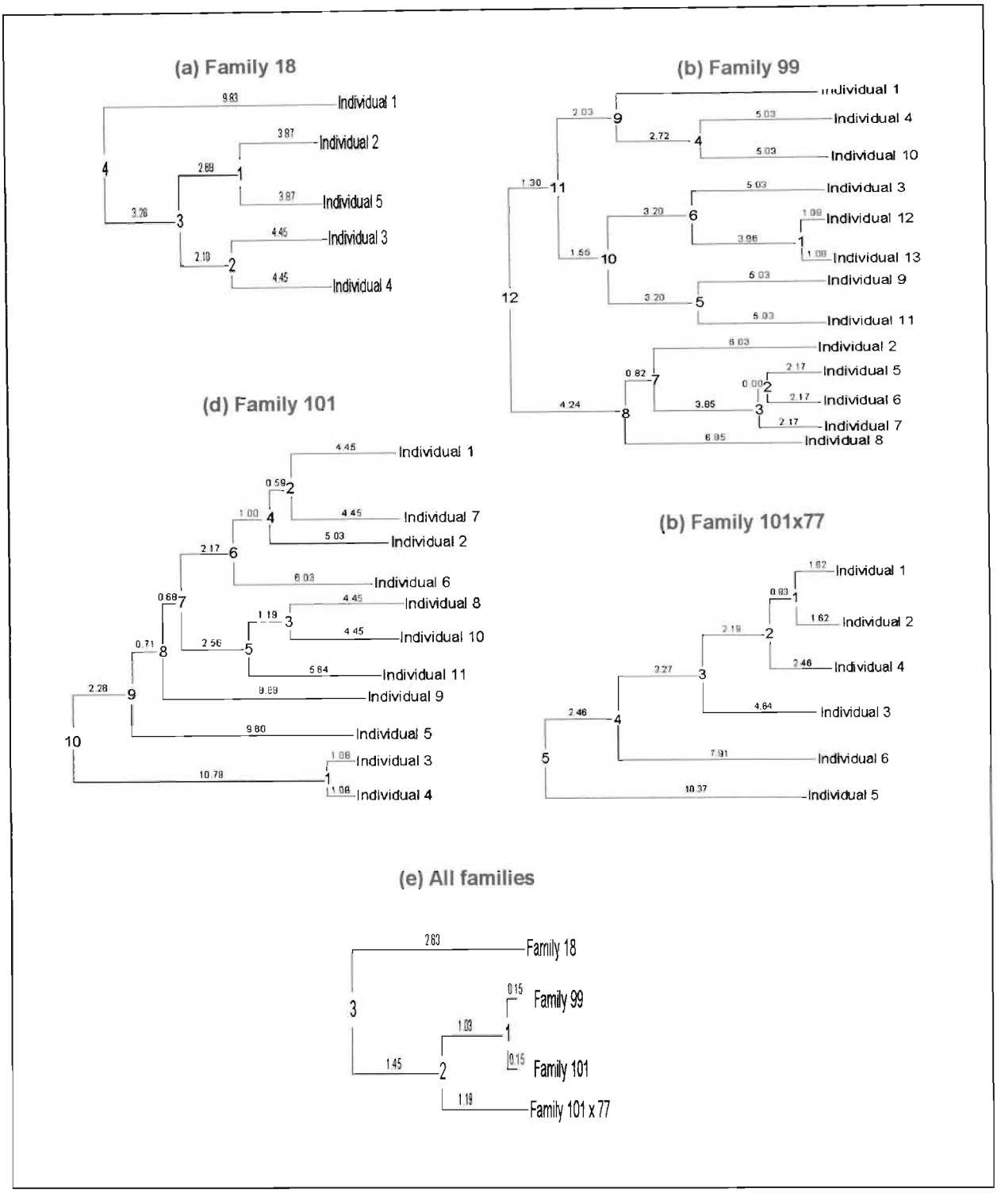


Figure 3.7 **Dendograms representing the overall genetic relationships between individuals within a family and between families, (a) Half-sib family 18, (b) Half-sib family 99, (c) Half-sib family 101, (d) Full-sib family 101x77 and (e) All families.**

A comparison of the genetic relatedness between the different families revealed that the three gummosis susceptible families clustered together in the tree with short arm lengths, while family 18, gummosis resistant, was separated from these families on a separate branch.

A tree was constructed using TREEVIEW (Page, 1996) to relate the relationships between all the individuals of the population (Figure 3.8). No consistent pattern of relationship could be identified. The individuals of the different families appeared to be scattered amongst one another. This random-type of distribution is probably explained by the extremely close genetic relationship that was identified amongst all individuals assessed.

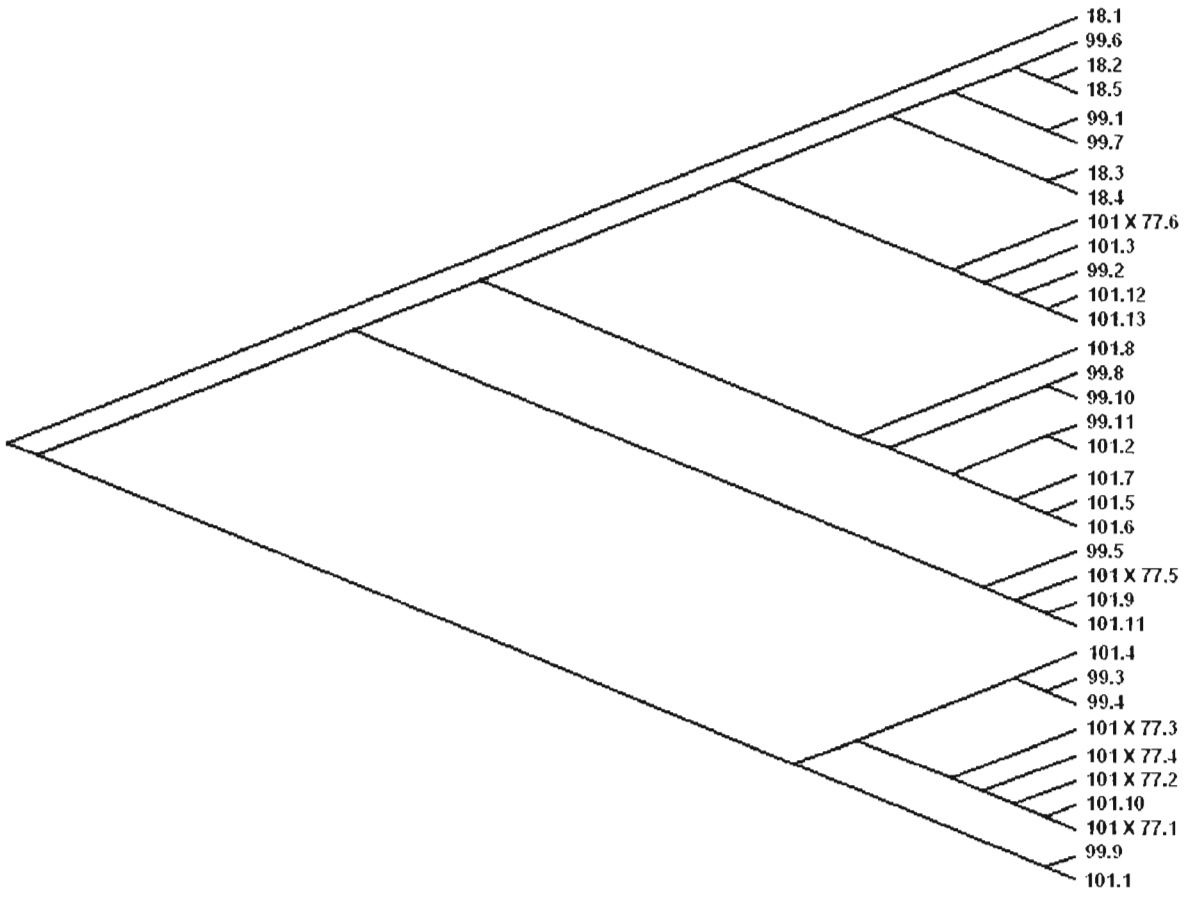


Figure 3.8 Dendogram representing overall genetic relationships between individual of the black wattle population examined.

3.5.4 Quantification of population differentiation

A hierarchical analysis of molecular variance (AMOVA) was conducted to quantify and partition the levels of variability into among- and within-form components. AMOVA is a method of estimating population differentiation directly from molecular data and testing hypotheses about such differentiation. To calculate and assess the molecular variance among and within the families, the software programme ARLEQUIN version 2.000 (Schneider et al., 2000) was employed. The results of the AMOVA partitioning of RAPD variance among and within families are shown in Table 3.9. Of the total genetic variation, 87.86 % was attributable to within family's variation and only 12.14 % to among families variation in black wattle. However, the proportion of variation attributed to among families and within families variation was both found to be highly significant (P<0.001) thus indicating that there was a significant genetic differentiation between families and within families.

Table 3.10 Analysis of molecular variance for one population of black wattle individuals based on RAPD variations.

Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage of variation	Probability* value (P)
Among families	3	47.469	1.01681	12.14	<0.001
Within families	31	228.131	7.35908	87.86	<0.001
Total	34	275.600	8.37589		
Fixation index (F_{ST})			**0.12140**		

*Nonparametric randomisation test (1023 permutations).

Wright's fixation index (F_{ST}) another measure of population differentiation, was also calculated. But, unlike AMOVA that estimates population 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 by the AMOVA analysis. In this investigation a significant F_{ST} value of 0.12140 was obtained. According to Wright's guidelines (Hartl and Clarke, 1997), a value between 0.05 and 0.15 indicates extensive genetic differentiation between families. The F_{ST} calculated for these families was approximately 12 % thus indicating

that genetic differentiation was extensive between different black wattle families analysed in this investigation.

CHAPTER 4

DISCUSSION AND CONCLUDING REMARKS

Black wattle was introduced into South Africa from Australia in 1864 as a tree to be used primarily for shade for livestock, shelterbelts, windbreaks, and for fuel wood (Sherry. 1971). Today black wattle is a major plantation species that contributes substantially to the South African forestry and forest products sector making up approximately 7 % of the South African plantation forestry estate. The demand for black wattle bark products and timber has cemented both locally and internationally, its place in South African forestry (Dunlop, 2002).

Recently, it has been recognised that breeding programmes will benefit greatly with the inclusion of molecular technologies. Such techniques could reduce the rather long generation interval by the identification of superior genotypes in juveniles. The possibility to identify superior genotypes, in juveniles, would greatly facilitate and speed up all breeding programmes by including molecular technologies.

The introduction of molecular technologies into tree breeding programmes facilitates these programmes in various ways. DNA fingerprinting is utilised to identify genotypes in terms of specific advantageous traits. In many instances it is not possible to identify the presence of a particular allele and thus the identification of linked DNA markers is utilised. Such markers play a significant role in marker assisted selection (MAS), where sort after genotypes are identified at any age. In tree breeding the identification of superior genotypes at the seedling stage is of great value.

DNA fingerprinting also provides valuable information about the breadth of the genetic base of a population. Such investigations are of importance in especially introduced species displaying dwindling numbers.

In all molecular techniques, the efficient isolation of DNA is an important step in the process (Csaikl et al., 1998). It is therefore important that suitable DNA isolation procedures are employed that yield pure and high molecular weight DNA. Furthermore,

it would be advantageous if the selected procedure is quick, easy to apply and cost effective and, if possible, devoid of dangerous chemicals (Puchooa, 2004).

The isolation of DNA from black wattle has proven to be problematic because of the presences of high levels of polysaccharides, polyphenols, and tannins present in the cells. These contaminants are very difficult to remove, and tend to inhibit the activity of certain DNA-modifying enzymes, such as *Taq* polymerase. They may also interfere in the quantification of the purity of nucleic acids by spectrophotometric methods (Wilkie et al., 1993). Therefore, a suitable isolation method needs to be optimised to yield pure DNA for subsequent DNA amplifications by PCR to generate RAPDs, AFLPs, RFLPs, and microsatellites fingerprints (Paterson et al., 1993).

The CTAB DNA isolation method of Doyle and Doyle (1990) has proven to be a successful isolator of DNA in many tree species. The cells for many tree species contain a range of different contaminants that are separated from DNA by the CTAB method. In lychee (Puchooa, 2004) and pine (Strange et al., 1998) that contain high quantities of polysaccharides and phenolics, high molecular weight DNA of high quality was isolated with the CTAB method. In black wattle, the CTAB method successfully isolated DNA that was pure and could be utilised in PCR. The success of this method is attributed to its comprehensive contaminant binding ability through the use of antioxidants 2-mercapto-ethanol and polyvinylpolypyrrolidone (PVP), which were able to bind polyphenolics residing in the cells of black wattle. Furthermore, the high salt concentration employed by the CTAB method, assisted with the removal of the polysaccharides and tannins.

A few important points regarding the CTAB isolation method need to be emphasised. Firstly, it is advisable to use freshly cut leaves, however if this is not possible, leaf material should be refrigerated. Secondly, it is important that liquid nitrogen be used to grind the plant tissue as it bursts the plant cell walls and allows the DNA to mix with the CTAB isolation buffer. It is also vital that the plant grindate be placed on ice immediately so as to prevent the DNA from degrading.

In forestry, where some form of genotyping is often required, the next step after DNA isolation entails a PCR. PCR is utilised to amplify a range of DNA markers; random markers such as RAPD and AFLP as well as specific markers such as RFLP and SSR. Recent developments in DNA marker technology together with traditional breeding practices provide new solutions for selecting and maintaining desirable genotypes. Markers closely linked to desirable traits facilitate the identification and selection of superior genotypes, marker assisted selection, and is regularly performed on early segregating populations during early stages of plant development (Mohan et al., 1997). Through the application of marker-assisted selection, it is now possible for a breeder to conduct many rounds of selection in a year (Adam-Blondon et al., 1994), thereby reducing the long generation interval in trees (Butcher, 2002). Molecular technologies has in recent times become fully integrated into existing tree breeding programmes all over the world allowing breeders to access, transfer and combine genes with precision at a rate not previously possible.

Extensive efforts have been devoted to the construction of molecular marker maps for the major commercial genera, such as for eucalypts, pines, and acacias by employing RFLPs, SSRs, RAPDs and AFLPs. These molecular maps have been used to locate markers associated with a variety of forestry traits of commercial interest, such as growth, frost tolerance, wood properties, vegetative propagation, leaf oil composition and disease resistance (Dekkers and Hospital, 2002). Bulked segregant analysis (Michelmore et al., 1991) was employed by Devey et al. (1995) to isolate ten RAPD markers flanking the resistance gene to white pine blister rust in Sugar pine, six of them within 5 cM of the gene. In Loblolly pine, a dominant gene that confers resistance to fusiform rust disease was identified by genomic mapping (Wilcox et al., 1995). Lehner et al. (1995) identified in Norway spruce a RAPD marker closely linked to the pendula gene. These findings thus indicate how molecular marker technology has contributed to general breeding programmes in a variety of forestry tree species accelerating breeding success (Rafalski et al., 1991; Rafalski and Tingey, 1993). In black wattle, in this investigation for the first time, one putative marker was identified linked to disease resistance of the disease gummosis.

Genetic variation has become one of the keywords of the scientists and silviculturists who are concerned about the sustainable management of forests. Genetic variation refers to the variation at the level of individual genes (polymorphism), and provides a mechanism for populations to adapt to their ever-changing environment. The greater the genetic variation, the greater the probability that at least some individuals in a population will have an allelic variant that is suited to the new and changing environment. These individuals will produce offspring with variant alleles, thereby assuring that such beneficial alleles are inherited from generation to generation (Sun et al., 1998).

An understanding of the genetic structure of forest tree species is fundamental for an appropriate utilisation of forest genetic resources, either for genetic improvement of plantations or for management or conservation of natural communities (Sun et al., 1998). In a tree improvement programme, knowledge of the apportionment of genetic variation within and among natural tree populations is important for initial selection. Genetic variation in forest trees has been investigated using morphological, physiological, biochemical and molecular marker traits (Wickneswari and Norwati, 1993). Knowledge of genetic variation will aid in making reasoned decisions on conservation of various tree species (Runo et al., 2004).

Forest geneticists have long used field experiments and, to a lesser extent, molecular markers to study patterns of adaptation in forest trees. Field experiments are employed to estimate genetic parameters of measurable traits, but they are unable to provide information about particular genes and how many of them are involved in adaptation, nor how much of phenotypic variation can be explained by genetic variation amongst these genes. Another and generally complementary approach for estimating adaptive genetic variation is to measure the genetic variation using molecular genetic markers.

Molecular markers, which detect variation at the DNA level, provide a way to characterise genomes accurately at a faster rate. Recently, an array of molecular marker techniques has been developed. Molecular markers like restriction fragment length polymorphism (RFLP) (Lebrun et al., 1998), randomly amplified polymorphic DNA (RAPD) (Ashburner et al., 1997; Rodriguez et al., 1997), amplified fragment length polymorphism (AFLP) (Perera et al., 1998) and microsatellites or simple sequence

repeats (SSRs) (Perera et al., 2000) have been successfully employed for assessing genetic variation in a variety of species. Although newer techniques like AFLP and SSRs are gaining importance due to their ability to detect more polymorphism, RAPD markers remain popular because of their simplicity and low development cost. The RAPD technique is an efficient tool for identifying variation and estimating diversity in different biological systems (Tingey and Tufo, 1993). RAPD markers are generated by PCR amplification of random genomic segments with a single short primer (10 -11 mer) of arbitrary sequence (Williams et al., 1990). Since no a priori knowledge of genome structure is needed, they are especially useful for analysis of less studied genomes like black wattle.

Many studies have been conducted on the determination of the genetic variation within and among populations in a variety of species. In 2001, Hwang et al. investigated the population differentiation between two populations of each of the two species of *Chamaecyparis* trees and found that there was a greater population differentiation (15.13 %) between the populations of the one species as compared to the other (14.73 %). The genetic diversity was also calculated using Shannon's Information Index and it was found that one population had a greater genetic diversity (species one, 0.390 and species two, 0.448) than the other (species one, 0.267 and species two, 0.409) in both the species of *Chamaecyparis.* High levels of population differentiation and genetic variation indicated dynamic evolution in these two *Chamaecyparis* species. Pedro and Jose (2000) investigated the genetic variation among seven Rosaceae trees, an animal dispersed tree, and found that the genetic diversity for the individual populations ranged from 8.9 % to 14.9 %. According to Hwang et al. (2001), these high levels of genetic diversity are most probably associated with long-distance seed dispersal by efficient medium-sized frugivorous birds and mammals.

A similar study showing high levels of genetic diversity was done by Moran (1992) who studied patterns of genetic diversity in three major Australian tree species; casuarina, eucalypts and acacia, by estimating the mean percentage of polymorphic loci, expected heterozygosity, and Nei's gene diversity for several populations of each of the species. The *eucalypts* gave a mean percentage of polymorphic loci of 56.8 %, 47.76 % for *acacia*, and 74.5 % for *casuarina*. The mean expected heterozygosity was estimated at

17.4 % for *eucalypts,* 14.5 % for *acacia*, and 21.1 % for *casuarina*. Nei's mean gene diversity of each species was calculated at 21.3 % for *eucalypts*, 19.9 % for *acacia*, and 28.7 % for *casuarina*. From these quantifications of genetic variation, it was deduced that *casuarina* has a higher level of genetic variation compared to *eucalypts* and *acacia*.

In most studies into genetic variation in trees, the levels of genetic variation were found to be very high; however this was not the case in black wattle. In this investigation, the data clearly support the notion that the genetic base in South African black wattle is rather narrow as estimated by Nei's unbiased estimate of genetic identity (I) and distance (D), quantification of differentiation (AMOVA and Wright's fixation index), percentage polymorphic loci, Nei's genetic diversity, Shannon's information index, and the estimated heterozygosity.

Genetic relationships between families and among individuals within families were quantified according to Nei's unbiased estimate of genetic identity (I) and distance (D). A pairwise analysis between the individuals within the different families revealed in general that all similarities were high, between 71% and 99 %. The similarity ranges were narrow for the two small families; 80 % to 93 % for family 18 and 80 % to 96 % for family 101x77. The two larger families, on the other hand, displayed a larger range in identity comparisons; 72 % to 98 % for family 99 and 71 % to 99 % for family 101.

A comparison of the genetic identity between the four families displayed extensive genetic similarity ranging from 92 % to 99 %, which was expected for each of the families; however, it was an unexpected result that the families were so closely related. These data clearly support the notion that the genetic base of South African black wattle population is narrow.

An analysis of family differentiation was undertaken by conducting an analysis of molecular variance (AMOVA). The results of the AMOVA partitioning of RAPD variance among and within families revealed that of the total genetic variation, 87.86 % was attributable to within family variation and only 12.14 % to between families variation in black wattle, both highly significant ($P<0.001$) thus indicating that there is a significant genetic differentiation between the families.

Wright's fixation index (F_{ST}) was also computed to estimate population differentiation, but unlike AMOVA that estimates population 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 0.12140, which is the same as what the AMOVA estimated.

Genetic variation within this group of black wattle individuals was quantified for each family individually as well as for the entire population. The quantification of genetic variation was undertaken by determining the percentage polymorphic loci, Nei's gene diversity, Shannon's Information Index and estimated heterozygosity. All these measures of variation clearly demonstrated a rather limited amount of genetic variation in this population, which was also true for each of the families. The polymorphic content of the entire population was estimated as 58.51 %. However, family 101 displayed the highest values for Nei's gene diversity (15.56 %) and Shannon's Information Index (23.64 %). As expected, the only family containing full-sibs (101x77) contained consistently the least amount of genetic variation (Nei's gene diversity = 10.64 %; Shannon's information index = 15.89 %). Second to family 101x77 was family 18, expected for a half-sib family with a small number of individuals. The two larger families, 99 and 101, both displayed nearly double the number of polymorphic loci (43.62 %; 46.81 %) than the two smaller families (28.72 % for both), probably indicating the need to evaluate larger families or more loci.

The estimated heterozygosity for the population was determined by firstly estimating q from the frequency of the null-phenotype at each locus by computing the square root, and thereafter calculating 2pq, the estimated heterozygosity. For the entire population the estimated heterozygosity was averaged over all loci. The estimated heterozygosity of this population was low, an average of 24 %, with the lowest percentage estimated at a particular locus of 4 % (2 loci) and the highest of 50 % (4 loci). The mean heterozygosity of the loci of the different primers individually ranged from 11 % to 39 %.

The various estimations of genetic variation within black wattle generally indicated limited variability. These findings support the introduction history of black wattle, where

a limited number of seeds, thus a limited amount of variability, were introduced from Australia. Together with the limited original genetic base, extensive selection has been undertaken in recent times therefore reducing the limited variability available even further.

Although only a few families were investigated, the results clearly demonstrate the importance of obtaining extensive knowledge of the introduction history and genetic variation of a species, as this knowledge impacts greatly on breeding activities such as selection of superior genotypes and marker-assisted selection. It is therefore important that an extensive investigation into genetic variation is undertaken across the plantation distribution area of black wattle in order to obtain a more accurate evaluation of the extent and complexities of genetic variability in black wattle.

REFERENCES

Adam-Blondon A. F., Sevignac M., Bannerot H. and Dron M., 1994. SCAR, RAPD AND RFLP markers linked to a dominant gene (*Are*) conferring resistance to anthracnose in common bean. *Theoretical and Applied Genetics*, 88: 865-870.

Ahuja M. R., Devey M. E., Groover A. T., Jermstad K. D. and Neale D. B., 1994. Mapped DNA probes from loblolly pine can be used for restriction fragment length polymorphism mapping in other conifers. *Theoretical and Applied Genetics,* 88: 279-282.

Aljanabi S. M. and Martinez I., 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research,* 25: 4692-4693.

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*. New York: Chapman & Hall, Inc.

Ayala F. J., 1982. *Population and evolutionary genetics: A Primer*. Philippines: The Benjamin/Cummings Publishing Company, Inc.

Barnes R. D., 2001. The African acacias - A thorny subject. *South African Forestry Journal*, 190: 9-17.

Beck S. L., 1999. Micropropagation of *Acacia mearnsii* (De Wild). PhD Thesis. University of Natal, Pietermaritzburg.

Beck S. L. and Goodricke T. G., 2002. Threats to wattle plantations I: Abiotic and mammalian. In: Dunlop R. W. and Maclennan L. A. (eds) *Black Wattle: The South African research experience*, Pg. 101-104.

Borovkova I. G., Steffenson B. J., Jin Y., Rasmussen J. B., Kilian A., Kleinhofs A., Rossnagel B. G. and Kao K. N., 1995. Identification of molecular markers linked to the stem rust resistance gene rpg4 in barley. *Phytopathology*, 85: 181-185.

Bowditch B. M., Albright D. G., Williams J. G. K. and Braun M. J., 1991. Use of Randomly Amplified Polymorphic DNA Markers in Comparative Genome Studies. *Methods in Ezymology*, 224: 294-309.

Brondani R. P. V., Brondani C., Tarchini R. and Grattapaglia D., 1998. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *Theoretical and Applied Genetics*, 97: 816-827.

Bruford M. W., Hanotte O., Brookfield J. F. Y. and Burke T., 1992. Single-locus and multilocus DNA fingerprinting. In: *Molecular genetic analysis of populations*, (Hoelzel, A. R., ed). Oxford: IRL Press; 225-269.

Bucci G. and Menozzi P., 1995. Genetic variation of RAPD markers in a *Picea abies* Karst. Population. *Heredity,* 75: 188-197.

Bundock P. C., Hayden M. and Vaillancourt R. E., 2000. Linkage maps of *Eucalyptus globulus* using RAPD and Microsattelite markers. *Silvae Genetica,* 49: 223-232.

Burke T., Dolf G., Jeffreys A. J. and Wolff R., 1991. *DNA Fingerprinting: Approaches and Applications*. Switzerland: Birkhauser Verlag.

Burdon R. D., 1992. Tree breeding and the new biotechnology - in damaging conflict or constructive synergism? In: Paper to IUFRO S2.02-08 Conference, *Breeding Tropical Trees*, Cali, Columbia.

Butcher P. A., Moran G. F. and Perkins H. D., 1998. RFLP diversity in the nuclear genome of *Acacia mangium. Heredity,* 81: 205-213.

Butcher P. A., 2002. Molecular breeding of tropical trees. Paper presented at the conference on "Advances in Genetic Improvement of Tropical Tree Species", Yogyakarta, Indonesia, 1-3 October 2002.

Byrne M., Murrel J. C., Owen J. V., Williams E. R. and Moran G. F., 1997. Mapping of quantitative trait loci influencing frost tolerance in *Eucalyptus nitens. Theoretical and Applied Genetics,* 95: 975-979.

Cavalli-Sforza L. L. and Bodmer W. F., 1971. *The genetics of human populations*. W. H. Freeman and Company, San Fransisco, California.

Cavalli-Sforza L. L. and Edwards A. W. F., 1967. Phylogenetic Analysis: Models and estimation procedures. *Evolution,* 32: 550-570.

Cheliak W. M. and Rogers D. L., 1990. Integrating biotechnology into tree improvement programs. *Canadian Journal of Forestry Research,* 20: 452-463.

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

Csailk U. M., Bastian H., Brettschneider R., Gauch S., Meir A., Schauerte M., Scholtz F., Sperisen C., Vornam B. and Ziegenhagen B., 1998. Comparative analysis of different DNA extraction protocols: A fast, universal maxi-preparation of high quality plant DNA for genetic evaluation and phylogenetic studies. *Plant Molecular Biology Reporter,* 16: 69-86.

Davies K. E., 1988. Genome Analysis: A practical approach. UK: IRL Press Ltd.

Dekkers J. C. M. and Hospital F., 2002. The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews: Genetics* 3: 22-32.

Desmarais E., Lanneluc I. and Lagnel J., 1998. Direct Amplification of Length Polymorphisms (DALP), or how to get and characterize new genetic markers in many species. *Nucleic Acid Research,* 26(6): 1458-1465.

Devey M. E., Delfino-mix A., Kinloch B. B. and Neale D. B., 1995. Random amplified polymorphic DNA markers tightly linked to a gene for resistance to white pine blister rust in sugar pine. *Proceedings of the National Academy of Science* (USA), 92: 2066-2070.

De Laia M. L., Gomes E. A., Esbrisse E. J. and De Araujo E. F., 2000. Random Amplified Polymorphic DNA (RAPD) analysis of genotypic identities in *Eucalyptus* clones. *Silvae Genetica,* 49: 239-243.

Dirlewanger E. P., Isaac G., Ranade S., Belajouza M., Cousin R. and de Vienne D., 1994. Restriction fragment length polymorphism analysis of loci associated with disease resistance genes and developmental traits in *Pisum sativum* L. *Theoretical and Applied Genetics,* 88: 17-27.

Dobson D. and Feely J., 2002. The wattle forest products industry. In: Dunlop R. W. and Maclennan L. A. (eds) Black Wattle: The South African research experience, Pg. 3-9.

Doyle J. J. and Doyle J. L., 1990. Isolation of plant DNA from fresh tissue. *Focus*, 12: 13-15.Dunlop R. W., 2002. Introduction. In: Dunlop R. W. and Maclennan L. A. (eds) Black Wattle: The South African research experience, Pg. 1-2.

Dunlop R. W., 2002. Tree Improvement. In: Dunlop R. W. and Maclennan L. A. (eds) Black Wattle: The South African research experience, Pg. 11-32.

Dunlop R. W., Barnes R. D., Nixon K. M. and Hagedorn S. F., 2003. A breeding strategy for *Acacia mearnsii* (black wattle) in South Africa. *South African Forestry Journal,* 199: 55-63.

Dunlop R. W., Goodricke T. G. and Clarke C. R. E., 2000. Open-pollinated family variation in growth, wood and dissolving pulp properties of *Acacia mearnsii.* ICFR Bulletin series 16/2000.

Dunlop R. W. and Hagedorn S., 1998. Final report on two Australian *Acacia mearnsii* (black wattle) provenance trials established in KwaZulu-Natal and south eastern Mpumalanga. ICFR Bulletin 7/98.

Dvorak W. S. and Donahue J. K., 1992. CAMCORE review 1980-1992. College of Forest Resources, NC State University. Raleigh, Pg. 97.

Dvorak W. S., Hodge G. R., Bester C., Kietzka E., Osorio L. F. and Stanger T. K., 2000b. *Pinus Patula.* In: the conservation and testing of tropical and subtropical forest species by the CAMCORE Cooperative (in press).

Echt C. S., May-Marquardt P., Hseih M. and Zahorchak R., 1996. Characterization of microsatellite markers in eastern white pine. *Genome*, 39: 1002–1008.

Excoffier L., Smouse P. E. and Quattro J. M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics,* 131: 479-491.

Fritsch P. and Rieseberg L. H., 1995. The Use of Randomly Amplified Polymorphic DNA (RAPD) in Conservation Genetics. In: Molecular Genetics Approaches in Conservation. 1996. Smith, T. B. and R. K. Wayne. New York: Oxford University Press.

Garbutt D. C. F., 1971. Budding as a method for the clonal propagation of black wattle. Wattle Research Report 1971, Pg. 54-57.

Gibson I. A. S., 1964. The impact of disease on forest production in Africa. In: FAO/IUFRO Symposium on internationally dangerous forest diseases and insects, Oxford.

Glaubitz J. C., Garnier-Gere P. and Moran G. F., 1999. Assessment of options and research priorities for the practical, sensitive and cost effective monitoring of Montreal Indicator 1.3.a – 'the amount of genetic diversity within and between populations of representative forest dwelling species'. Scoping study prepared for the Forest and Wood Products Research and Development Corporation.

Gourlay I. D. and Barnes R. D., 1994. Seasonal growth zones in the wood of *Acacia karoo* Hayne: their definition and implications. *Commonwealth Forestry Review,* 73: 121-127.

Govender P., 2002. Threats to wattle plantations II: Insects. In: Dunlop R. W. and Maclennan L. A. (eds) Black Wattle: The South African research experience, Pg. 105-120.

Govender P., 1995. The pest status and chemical control of whitegrubs and cutworms in forestry in the Natal Midlands. M.Sc thesis, University of Natal, Pietermaritzburg.

Graner A. and Tekauz A., 1996. RFLP mapping in barley of a dominant gene conferring resistance to scald (*Rhynchosporium secalis*). *Theoretical and Applied Genetics,* 93: 421-425.

Hagedorn S. F., 1993. Half rotation results of Australian black wattle provenances grown in Natal and the south eastern Transvaal. ICFR Bulletin Series 11/93.

Haig S. M., 1998. Molecular Contributions to Conservation. *Ecology,* 79(2): 413-425.

Haigh H., 1993. Growing black wattle. Forestry Development, Department of Water Affairs and Forestry, Pretoria. Extension Leaflet 1/93 (15).

Hagedorn S. F., 2001. Results from a black wattle browsing control trial. ICFR Bulletin 11/93.

Haines R. J., 1992. Mass propagation by cuttings, biotechnologies, and the capture of genetic gain. In: Paper presented to IUFRO Symposium "Mass production technology for genetically improved fast growing forest tree species", Bordeaux, France.

Hartl D. L., 2000. A primer of population genetics (3rd edition). USA: Sinauer Associates.

Hartl D. L. and Clark A. G., 1997. Principles of population genetics. 3rd edition. Sunderland (MA): Sinauer Associates.

Hartl L., Weiss H., Stephan U., Zeller F. J. and Jahoor A., 1995. Molecular identification of powdery mildew resistance genes in common wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics,* 90: 601-606.

Harvett C., 2001. A manager's view of tree improvement - towards designer fibres: maximising value from applied tree improvement. *South African Forestry Journal,* 190: 25-29.

Hayami Y. and Ruttan V. W., 1985. Agricultural development: An international perspective. Johns Hopkins University Press, Baltimore, Pg. 115.

Heinkel R., Hartmann W. and Stoesser R., 2000. On the Origin of the plum cultivars 'Cacaks Beauty', 'Cacaks Early' and 'Cacaks Fruitful' as investigated by the inheritance of random amplified polymorphic DNA (RAPD) fragments. Scientia Horticulturae 83: 149-155.

Helentjaris T., King G., Slocum M. K., Siedenstrang C. and Wegman S., 1985. Restriction fragment polymorphisms as probes for plant diversity, and their development as tools for applied plant breeding. *Plant Molecular Biology,* 5: 109-118.

Hicks M., Adams D., O'Keefe S., Macdonald E. and Hodgetts R., 1998. The development of RAPD and microsatellite markers in lodgepole pine. *Genome*, 41: 797-805.

Hillis D. M., 1984. Misuse and modification of Nei's genetic distance. Systematic Zoology, 33: 238-240.

Hwang S., Lin H., Kuo Y. and Lin T., 2001. RAPD variation in relation to population differentiation of *Chamaecyparis formosensis* and *Chamaecyparis taiwanensis. Botanical Bulletin of Academia Sinica,* 42: 173-178.

Innis M. A., Gelfand D. H., Sninsky J. J. and White T. J., 1990. PCR Protocols: A Guide to Methods and Applications. San Diego: Academic Press, Inc.

Isabel N., Tremblay L., Michaud M., Tremblay F. M. and Bousquet J., 1993. RAPDs as an aid to evaluate the genetic integrity of somatic embryogenesis-derived populations of *Picea mariana* (Mill.) B.S.P. *Theoretical and Applied Genetics,* 86: 81-87.

Jacobs J. M. E., Herman J. V. C., Horsman K., Arens P. F. P., Verkerk-Bakker B., Jacobsen E., Pereira A. and Stiekema W. J., 1996. Mapping

of resistance to the potato cyst nematode *Globodera rostochiensis* from the wild potato species *Solanum vernei*. *Molecular Breeding,* 2: 51-56.

Jarmain R. M. and Lloyd-Jones G. A., 1982. The wattle industry in South Africa - A survey for students. Wattle Research Institute Document 12/82. Typescript.

Kang H-W., Park D-S., Go S-J. and Eun M-Y., 2002. Fingerprinting of Diverse Genomes using PCR with Universal Rice Primers Generated from Repetetive Sequence of Korean Weedy Rice. *Molecules and Cells* 13(2): 281-287.

Killian A., Steffenson B. J., Saghai Maroof M. A. and Kleinhofs A., 1994. RFLP markers linked to the durable stem rust resistance gene *Rpg 1* in barley. *Molecular Plant-Microbe Interact,* 7: 298-301.

Kirby L. T., 1990. DNA Fingerprinting: An Introduction. New York: Stockton Press.

Klein-Lankhorst R., Rietveld P., Machiels B., Verkerk R., Weide R., Gebhardt C., Kornneef M. and Zabel P., 1991. RFLP markers linked to the root knot nematode resistance gene *Mi* in tomato. *Theoretical and Applied Genetics,* 81: 661-667.

Kostia S., Varvio S. L., Vakkari P. and Pulkkinen P., 1995. Microsatellite sequences in a conifer, *Pinus sylvestris*. *Genome*, 38: 1244–1248.

Krawczak M. and Schmidtke J., 1994. DNA Fingerprinting. Oxford: Bios Scientific Publishers Ltd.

Lebrun P., N'cho Y. P., Seguin M., Grivet L. and Baudouin L., 1998. Genetic diversity in coconut (*Cocos nucifera* L.) revealed by restriction

fragment length polymorphism (RFLP) markers. *Euphytica*, 101: 103-108.

Lehner A., Campbell M. A., Wheeler N. C., Poykko T., Glossl J., Kreike J. and Neale D. B., 1995. identification of a RAPD marker linked to the pendula gene in Norway spruce (*Picea abies* (L) Karst. f. *pendula*). *Theoretical and applied Genetics*, 91: 1092-12094.

Lewontin R. C., 1972. The apportionment of human diversity. *Evolutionary Biology*, 6: 381-398.

Lougheed S. C., Gibbs H. L, Prior K. A. and Weatherhead P. J., 2000. A comparison of RAPD versus Microsatellite DNA Markers in Population Studies of the Massasauga Rattlesnake. *The American Genetic Association* 91: 458-463.

Luciane G. and Cavalli-Molina S., 2000. Genetic variation in natural populations of maté (*Ilex paraguariensis* A. St.-Hil., Aquifoliaceae) using RAPD markers. *Heredity,* 84 (6): 647-656.

Lynch M. and Milligan G., 1994. Analysis of population genetic structure with RAPD markers. *Molecular Ecology,* 3: 91-99.

McAree D., 2002. The forest service diversity plan. Biology and Environment: Proceedings of the Royal Irish Academy, 102B(3): 183-184.

Menzies N., 1985. Land tenure and resource utilization in China: A historical perspective. In: Paper presented at a conference on managing renewable resources in Asia, Sapporo, Japan.

Michelmore R. W., Paran I. and Kesseli R. V., 1991. Identification of markers linked to disease-resistance genes by bulked segregant

analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Science*, 88: 9828-9832.

Mohan M., Nair S., Bhagwat A., Krishna T. G., Yano M., Bhatia C. R. and Sasaki T., 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular Breeding,* 3: 87-103.

Mosseler A., Egger K. N. and Hughes G. A., 1992. Low levels of genetic diversity in red pine confirmed by random amplified polymorphic DNA markers. *Canadian Journal of Forestry Research,* 22: 1332-1337.

Moncur M. W., Moran G. F. and Grant J. E., 1991. Factors limiting seed production in *Acacia mearnsii*. In: Turnbull J (ed.) Advances in Tropical Acacia Research. ACIAR Proceedings Series, No. 35, Pg. 20-25.

Moran G. F., 1992. Patterns of genetic diversity in Australian tree species. *New Forests*, 6: 49-66.

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 Forestry Research,* 21: 545-554.

Nei M., 1972. Genetic distance between populations. *American Naturalist,* 106: 283-292.

Nei M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics,* 89: 583-590.

Nienhuis J., Helentjaris T., Slocum M., Ruggero B. and Schaefer A., 1987. Restriction fragment length polymorphism analysis of loci associated with insect resistance in tomato. *Crop Science,* 27: 797-803.

Nkongolo K. K., Michael P. and Gratton W. S., 2002. Identification and characterization of RAPD markers inferring genetic relationships among Pine species. *Genome,* 45: 51-58.

Oh B. J., Frederiksen R. A. and Magill C. W., 1994. Identification of molecular markers linked to head smut resistance gene (*Shs*) in sorghum by RFLP and RAPD analyses. *Phytopathology,* 84: 830-833.

Osborn T. C., Alexander D. C. and Fobes J. F., 1987. Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruit. *Theoretical and Applied Genetics,* 73: 350-356.

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(2): 361-382.

Paterson A. H., Brubaker C. L. and Wendel J. F., 1993. A rapid method for extraction of cotton (*Gossypium spp*) genomic DNA suitable for RFLP or PCR analysis. *Plant Molecular Biology Reporter,* 11: 122-127.

Pejic I., Ajamone-Marsan P., Morgante M., Kozumplick V., Castiglioni P., Taramino G. and Motto M., 1999. Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs and AFLPs. http://www.agron.missouri.edu/mnl/72/65pejic.html.

Penner G. A., Stebbing J. A. and Legge B., 1995. Conversion of an RFLP marker for the barley stem rust resistance gene *Rpg 1* to a specific PCR-amplifiable polymorphism. *Molecular Breeding,* 1: 349-354.

Perera L., Russell J. R., Proven J., NcNicol J. W. and Powell W., 1998. Evaluating genetic relationship between indigenous coconut (*Cocos nucifera* L.) accessions from Sri Lanka by means of AFLP profiling. *Theoretical and Applied Genetics*, 96: 545-550.

Perera L., Russell J. R., Proven J. and Powell W., 2000. Use of microsatellite DNA markers to investigate the level of genetic diversity and population genetic structure of coconut (*Cocos nucifera* L.). *Genome*, 43: 15-21.

Pielou E. C., 1975. Ecological diversity. John Wiley, New York.

Pineda O., Bonierbale M. W., Plaisted R. L., Brodie B. B. and Tanksley S. D., 1993. Identification of RFLP markers linked to the *H1* gene conferring resistance to the potato cyst nematode *Globodera rostochiensis*. *Genome,* 36: 152-156.

Puchooa D., 2004. A simple, rapid and efficient method for the extraction of genomic DNA from lychee (*Litchi chinensis* Sonn.). *African Journal of Biotechnology,* 3(4): 253-255.

Rafalski J. A. and Tingey S. V., 1993. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. *Trends in Genetics*, 9: 275-280.

Rafalski J. A., Tingey S. V. and Williams J. G. K., 1991. RAPD markers – a new technology for genetic mapping and plant breeding. *AgBiotech News and Information*, 3: 645-648.

Rodriguez M. J. B., Estioko L. P., Namia T. I. and Soniega J. A., 1997. Analysis of genetic diversity in coconut by RAPD. *The Philippine Journal of Coconut Studies XXII*, 1-7.

Rodrigo B. R., Jupiter I. M. A., Ismael E. P., Elza F. A., 2002. Fingerprint and genetic diversity analysis of *Eucalyptus* spp. Genotypes using RAPD and SSR markers. *Scientia Forestalis,* 62: 24-31.

Roger J. S., 1972. Measures of genetic similarity and genetic distance. University of Texas Publication, 7213: 145-153.

Roux J., 2002. Threats to wattle plantations III: Diseases. In: Dunlop R. W. and Maclennan L. A. (eds) Black Wattle: The South African research experience, Pg. 121-126.

Roux J., Kemp G. H. J. and Wingfield M. J., 1995. Diseases of black wattle in South Africa - A review. *South African Journal of Forestry,* 174: 35-40.

Runo M. S., Muluvi G. M. and Odee D. W., 2004. Analysis of genetic structure in *Melia volkensii* (Gurke.) populations using random amplified polymorphic DNA. *African Journal of Biotechnology*, 3(8): 421-425.

Sanchez N., Grau J. M., Manzanera J. A. and Bueno M. A., 1998. RAPD markers for the identification of *Populus* species. *Silvae Genetica,* 47: 67-71.

Schneider S., Roessli D. and Excoffier L., 2000. Arlequin: A software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Schondelmaier J., Steinruecken G. and Jung C., 1996. Integration of AFLP markers into the linkage map of sugar beet (*Beta vulgaris* L.). *Plant Breeding,* 115: 231-237.

Searle S. D., Bell J. C. and Moran G. F., 2000. Genetic diversity in natural populations of *Acacia mearnsii*. *Australian Journal of Botany,* 48: 279-286.

Sedjo R. A., 1999. Biotechnology and planted forests: Assessment of potential and possibilities. Resources for the future, Washington.

Sherry S. P., 1971. The black wattle (*Acacia mearnsii de Wild.*). University of Natal Press, Pietermaritzburg, South Africa.

Simcox K. D. and Bennetzen J. L., 1993. The use of molecular markers to study *Setosphaeria turcica* resistance in maize. *Phytopathology,* 83: 1326-1330.

Skorupska H. T., Choi I. S., Rao-Arelli A. P. and Bridges W. C., 1994. Resistance to soybean cyst nematode and molecular polymorphism in various sources of Peking soybean. *Euphytica,* 75: 63-70.

Smith D. N. and Devey M. E., 1994. Occurrence and inheritance of microsatellites in *Pinus radiata*. *Genome,* 37: 977–983.

Sneath P. H. A. and Snokal R. R., 1973. Numerical Taxonomy. W.H. Freeman and Company, San Francisco, pp 230-234.

Snustad D. P. and Simmons M. J., 2000. Principles of Genetics, 2nd Edition. John Wiley & Sons Inc., USA.

Strange C., Prehn D. and Arce-Johnson P., 1998. Isolation of *Pinus radiate* genomic DNA suitable for RAPD analysis. *Plant Molecular Biology Reporter*, 16: 1-8.

Stubbings J. A., 1977. A.C.A.C.I.A. *South African Forestry Journal,* 102: 15-16.

Sun M., Wong K. C. and Lee J. S. Y., 1998. Reproduction biology and population genetic structure of *Kandelia* candela (Rhizophoraceae) a viviparous mangrove species. *American Journal of Botany*, 85: 1631-1637.

Tingey S. V. and del Tufo J. P., 1993. Genetic analysis with random amplified polymorphic DNA markers. *Plant Physiology*, 101: 349-352.

Travis S. E., Maschinski J. and Keim P., 1996. An analysis of genetic variation in *Astragalus cremnophylax* var. *cremnophylax*, a critically endangered plant, using AFLP markers. *Molecular Ecology,* 5: 735-745.

Van der Beek J. G., Pet G. and Lindhout P., 1994. Resistance to powdery mildew (*Oidium lycopersicum*) in *Lycopersicon hirsutum* is controlled by an incompletely-dominant gene *O1-1* on chromosome 6. *Theoretical and Applied Genetics,* 89: 467-473.

Vos P., Hodgers R. and Bleeker M., 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acid Research,* 23: 4407-4414.

Welsh J. and McClellend M., 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acid Research,* 18(24): 7213-7218.

White T. J., Arnheim N. and Erlich H. E., 1989. The polymerase chain reaction. *Annual Review of Biochemistry,* 5(6): 185-189.

Wickneswari R. and Norwati M., 1993. Genetic diversity of natural populations of *Acacia auriculiformis*. *American Journal of Botany*, 41: 65-77.

Wilcox P. L., 1995. Genetic dissection of fusiform rust resistance in loblolly pine. Ph.D thesis, North Carolina State University, 125p.

Wilkie S. E., Issac P. G. and Slater R. J., 1993. Random amplification of polymorphic DNA (RAPD) markers for genetic analysis in *Allium*. *Theoretical and Applied Genetics*. 86: 497-504.

Williams J. G. K., Kubelik A. R., Livak K. J., Rafalski J. A. and Tingey S. V., 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acid Research,* 18(22): 6531-6535.

WRI, 1948. Report for 1948. First year.

WRI, 1950. Wattle Research Institute Report. Second year.

WRI, 1963. Wattle Research Institute Report for 1962-1963.

WRI, 1965. Wattle Research Institute Report for 1964-1965.

WRI, 1970. Wattle Research Institute Report for 1969-1970.

Wingfield M. J. and Roux J., 2000. Diseases of pines, eucalypts and wattles. In: Van der Sijde (ed.) Forestry Handbook, The South African Association of Forestry, Pretoria.

Wingfield M. J., De Beer C., Visser C. D. and Wingfield B. D., 1996. A new Ceratocystis species defined using morphological and ribosomal DNA comparisons. *Systematic and Applied Microbiology,* 19: 191-202.

Yu K. and Pauls K. P., 1993. Identification of a RAPD marker associated with somatic embryogenesis in alfalfa. *Plant Molecular Biology,* 22: 269-277.

Yu Y. G., Saghai Maroof M. A., Buss G. R., Maughan P. J. and Tolin S. A., 1994. RFLP and microsatellite mapping of a gene for soybean mosaic virus resistance. *Phytopathology,* 84: 60-64.

Zhivotovsky L. A., 1999. Estimating population structure in diploids with multi-locus dominant DNA markers. *Molecular Ecology,* 8: 907-913.

Ziegenhagen B., Kuhlenkamp V., Brettschneider R., Scholz F., Stephan B. R. and Degen B., 1999. DIG-labelled AFLPs in oaks- A DNA marker for reconstruction of Full- or half-sib family relationships. http://webdoc.gwdg.de/ebook/y/1999/whichmarker/m05/Chap5.htm.

Zeijlemaker F. C. J. and Margot P., 1971. Black-Butt disease of black wattle. Wattle Research Institute Report for 1970-1971: 49-50.

Zeijlemaker F. C. J. and Margot P., 1970. Physical nature and origin of gummosis. Wattle Research Institute Report for 1969-1970: 28-29.

Zaitlin D., DeMars S. and Ma Y., 1993. Linkage of *rhm*, a recessive gene for resistance to southern corn leaf blight to RFLP marker loci in maize (*Zea mays*) seedlings. *Genome,* 36: 555-564.

YEH FRANCIS C., YANG R-C., TIMOTHY B. J., YE Z-H. and MAO, JUDY X, 1997. POPGENE, the user-friendly shareware for population genetic analysis. Molecular Biology and Biotechnology Centre, University of Alberta, Canada.

APPENDIX

A. LABORATORY SOLUTION RECIPES

B. POPGENE ANALYSIS

C. ARLEQUIN ANALYSIS

APPENDIX A

1. DNA isolation protocol 1

CTAB isolation buffer

- 10 g 2 % CTAB
- 40.9 g 1.4 M NaCl
- 2 ml 0.4 % 2-mercaptoethanol
- 3.72 g 20 mM EDTA
- 50 ml 100 mM Tris-HCl (pH 8.0)
- 10 g 2 % PVP

Make up to 500 ml with distilled water.

CTAB wash buffer

- 4.1 g NaCl
- 10 g CTAB

Make up to 100 ml with distilled water.

Chloroform:Isoamyl alcohol (24:1)

- 480 ml chloroform
- 20 ml isoamyl alcohol

Makes 500 ml. Store in a light sensitive bottle.

1 x TE

- 0.65 g 10 mM Tris
- 0.186 g 1 mM EDTA

Add water to one litre and adjust pH to 7.5.

2. DNA isolation protocol 2

lysis buffer

- 500 µl 1xTNE
- 50 µl 1M Tris-HCL pH 8
- 1 unit Proteinase K (1 µl)
- 7.5 µl 25 % SDS
- 7.5 µl Triton X-100

Make up to one litre.

5 M NaCl

- 29.22g NaCl

Add water to one litre.

10 mM Tris (pH8)

- 5 ml 0.5 M EDTA
- 10 ml 0.5 M Tris-HCl, pH 8.0
- 5ml 10% SDS solution

Make up to 100ml with water and autoclave.

3. DNA isolation protocol 3

20% Chelex©-100

- 0.12 g 10 mMTris
- 20 g Chelex

Make up to 100 ml.

APPENDIX B

1. All Families

Input file

/* Diploid RAPD Data Set */

Number of populations = 4

Number of loci = 94

Locus name:

A01-1 A01-2 A01-3 A01-4 A01-5 A01-6

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

A04-1 A04-2 A04-3

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

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

A10-1 A10-2 A10-3 A10-4 A10-5 A10-6

A12-1 A12-2 A12-3 A12-4 A12-5 A12-6 A12-7 A12-8

A13-1 A13-2 A13-3 A13-4 A13-5

A17-1 A17-2 A17-3 A17-4

A19-1 A19-2 A19-3 A19-4 A19-5

A20-1 A20-2 A20-3 A20-4 A20-5 A20-6

C05-1 C05-2 C05-3 C05-4 C05-5 C05-6 C05-7 C05-8

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

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

name = Family 18

fis = 0.00

111111 0110100100000 111 111010 1111101 101011 11111111 00011 1111 11111 000111 10111100 0111110 1000101111

111111 0000100000000 111 000010 1111101 101000 11111111 00011 1111 11111 001111 00011101 0011110 1100101111

111111 0000100000100 111 000010 0010111 101011 11111111 00011 1111 11111 000111 11011101 0011111 0100101111

111111 0000100000100 111 000010 1111111 101010 11111111 00011 1111 11111 000111 00011100 0011111 0100101101

111111 0000100000000 111 000010 1111101 000000 11111111 00011 1111 11111 011111 10011101 1011111 1100101101

name = Family 99

fis = 0.00

111111 0000100000100 111 010010 1111101 000011 11111111 00011 1111 11111 000111 01011101 1011110 1100100001

111111 0000100000000 111 000010 1111001 000011 11111111 00011 1111 11111 000111 01011101 0011111 0001101101

111111 0000100000100 111 000010 0100001 000001 11111111 00011 1111 11111 000101 10011110 0000110 0010100101

111111 0000100000100 111 000010 0100001 001011 11111111 00011 1111 11111 000101 10011110 0000110 0010100101

111111 0100100011000 111 001110 1111101 000101 11111111 00111 1111 11111 000111 10011100 0011111 1100100111

111111 0000100000100 111 000010 1111001 010000 11111111 00111 1111 11111 000111 00011111 0011110 1100101111

111111 0000100000000 111 000010 1111101 000011 11111111 00111 1111 11111 000111 00011101 0011111 1100101011

111111 0000110000010 111 000010 1111111 000011 11111111 10011 1111 11111 000111 11011101 0001110 1100101001

111111 0000110010000 111 000010 0101101 010000 11111111 00011 1111 11111 000111 11011100 0001110 0100111011

111111 0000110000010 111 000010 1111101 000101 11111111 11011 1111 11111 000111 10011101 1001111 1100100001

111111 1000110010010 111 000111 1111101 010011 11111111 10011 1111 11111 000111 11011101 1001110 1100101011

name = Family 101

fis = 0.00

111111 0100100001000 111 000010 0101101 000001 11111111 00011 1111 11111 000111 11011100 0010110 0000101001

111111 0001110000010 111 000111 1101111 000011 11111111 10011 1111 11111 100111 11011101 1001111 1100101011
111111 0000100000010 111 000010 0101101 000011 11111111 00011 1111 11111 000101 00011101 0001111 0001100001
111111 0000100000001 111 000010 0100011 000000 11111111 00011 1111 11111 000101 10011111 0000110 0010101001
111111 0000110010001 111 000011 1111101 010010 11111111 10011 1111 11111 000111 11011101 1001111 1100111001
111111 0000110000001 111 000011 1111101 010011 11111111 10011 1111 11111 000111 11011101 1001111 0100101001
111111 0000110000001 111 000011 1111101 010000 11111111 10011 1111 11111 000111 11011101 1001111 1100101011
111111 0000110010000 111 001111 1111101 010011 11111111 00011 1111 11111 010111 01011101 1011110 1100111001
111111 1000101000000 111 000110 1101101 001001 11111111 00011 1111 11111 000111 11011101 0101111 1110110111
111111 0000100000000 111 000010 0000001 000000 11111111 00011 1111 11111 000101 10011101 0000110 0001100111
111111 0100100101000 111 000111 1101101 001001 11111111 00011 1111 11111 000111 11011100 0101110 1100110111
111111 0000100000000 111 000011 1101101 000001 11111111 00011 1111 11111 000111 01011101 0001111 0001110111
111111 0000100000000 111 000011 1101101 000001 11111111 00011 1111 11111 000111 01011101 0001111 1001110101

name = Family 101x77

fis = 0.00

111111 0000100000000 111 000110 0000001 000000 11111111 00011 1111 11111 000101 00011101 0000110 0001100111
111111 0000100000000 111 000010 0000001 000000 11111111 00011 1111 11111 000111 00011101 0000110 0000100111
111111 0000100000100 111 000110 1000101 000010 11111111 00011 1111 11111 000111 10011100 0000111 0000100111
111111 0000100000100 111 000010 0000001 000000 11111111 00011 1111 11111 000101 10011101 0000110 0000110111
111111 0001100000100 111 000110 1111101 001001 11111111 10011 1111 11111 000111 01011101 1001111 1100110111
111111 0000100000100 111 000011 0011101 000010 11111111 00001 1111 11111 000101 01011101 1001110 0000100001

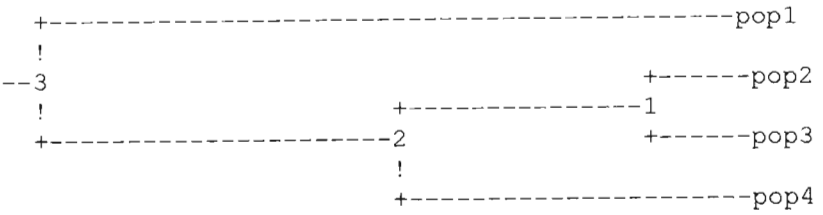
RESULTS – ALL FAMILIES

```
**************************************************************************
************
   **
**
   **          Nei's Analysis of Gene Diversity in Subdivided
Populations           **
   **          [See Nei (1987) Molecular Evolutionary Genetics (p. 187-
192)]         **
   **
**

**************************************************************************
************
```

Locus	Sample Size	Ht	Hs	Gst	Nm*
A01-1	35	0.0000	0.0000	****	****
A01-2	35	0.0000	0.0000	****	****
A01-3	35	0.0000	0.0000	****	****
A01-4	35	0.0000	0.0000	****	****
A01-5	35	0.0000	0.0000	****	****
A01-6	35	0.0000	0.0000	****	****
A02-1	35	0.0420	0.0410	0.0222	21.9921
A02-2	35	0.1094	0.1063	0.0286	17.0051
A02-3	35	0.0514	0.0472	0.0813	5.6481
A02-4	35	0.0612	0.0586	0.0420	11.4063
A02-5	35	0.0000	0.0000	****	****
A02-6	35	0.1871	0.1652	0.1169	3.7782
A02-7	35	0.0194	0.0188	0.0297	16.3267
A02-8	35	0.0698	0.0661	0.0533	8.8758
A02-9	35	0.1072	0.0996	0.0707	6.5676
A02-10	35	0.0613	0.0590	0.0373	12.9032
A02-11	35	0.3348	0.2900	0.1338	3.2368
A02-12	35	0.1072	0.0996	0.0707	6.5676
A02-13	35	0.0804	0.0699	0.1315	3.3028
A04-1	35	0.0000	0.0000	****	****
A04-2	35	0.0000	0.0000	****	****
A04-3	35	0.0000	0.0000	****	****
A06-1	35	0.0514	0.0472	0.0813	5.6481
A06-2	35	0.0732	0.0694	0.0514	9.2201
A06-3	35	0.0911	0.0882	0.0313	15.4883
A06-4	35	0.2395	0.2166	0.0955	4.7357
A06-5	35	0.0000	0.0000	****	****
A06-6	35	0.2238	0.1797	0.1969	2.0400
A09-1	35	0.4855	0.4468	0.0797	5.7698
A09-2	35	0.4836	0.2636	0.4549	0.5991
A09-3	35	0.4964	0.2695	0.4570	0.5942
A09-4	35	0.4992	0.4400	0.1185	3.7197
A09-5	35	0.4889	0.3424	0.2996	1.1688
A09-6	35	0.1605	0.1463	0.0885	5.1520
A09-7	35	0.0000	0.0000	****	****
A10-1	35	0.2382	0.1236	0.4811	0.5393

A10-2	35	0.1452	0.1326	0.0863	5.2960
A10-3	35	0.3098	0.2224	0.2821	1.2721
A10-4	35	0.0466	0.0432	0.0734	6.3166
A10-5	35	0.3970	0.3855	0.0290	16.7334
A10-6	35	0.4441	0.3729	0.1605	2.6161
A12-1	35	0.0000	0.0000	****	****
A12-2	35	0.0000	0.0000	****	****
A12-3	35	0.0000	0.0000	****	****
A12-4	35	0.0000	0.0000	****	****
A12-5	35	0.0000	0.0000	****	****
A12-6	35	0.0000	0.0000	****	****
A12-7	35	0.0000	0.0000	****	****
A12-8	35	0.0000	0.0000	****	****
A13-1	35	0.1809	0.1724	0.0470	10.1361
A13-2	35	0.0230	0.0222	0.0353	13.6587
A13-3	35	0.0709	0.0628	0.1146	3.8624
A13-4	35	0.1833	0.1208	0.3410	0.9663
A13-5	35	0.0000	0.0000	****	****
A17-1	35	0.0000	0.0000	****	****
A17-2	35	0.0000	0.0000	****	****
A17-3	35	0.0000	0.0000	****	****
A17-4	35	0.0000	0.0000	****	****
A19-1	35	0.0000	0.0000	****	****
A19-2	35	0.0000	0.0000	****	****
A19-3	35	0.0000	0.0000	****	****
A19-4	35	0.0000	0.0000	****	****
A19-5	35	0.0000	0.0000	****	****
A20-1	35	0.0194	0.0188	0.0297	16.3267
A20-2	35	0.0698	0.0661	0.0533	8.8758
A20-3	35	0.1064	0.0873	0.1791	2.2910
A20-4	35	0.0000	0.0000	****	****
A20-5	35	0.4814	0.3507	0.2715	1.3413
A20-6	35	0.0000	0.0000	****	****
C05-1	35	0.4540	0.4343	0.0433	11.0500
C05-2	35	0.3919	0.3435	0.1236	3.5456
C05-3	35	0.0514	0.0472	0.0813	5.6481
C05-4	35	0.0000	0.0000	****	****
C05-5	35	0.0000	0.0000	****	****
C05-6	35	0.0000	0.0000	****	****
C05-7	35	0.0889	0.0816	0.0817	5.6217
C05-8	35	0.4998	0.4759	0.0479	9.9422
C06-1	35	0.2728	0.2694	0.0123	40.0468
C06-2	35	0.0885	0.0841	0.0505	9.3982
C06-3	35	0.4458	0.1334	0.7008	0.2135
C06-4	35	0.4904	0.3220	0.3434	0.9560
C06-5	35	0.0000	0.0000	****	****
C06-6	35	0.0000	0.0000	****	****
C06-7	35	0.4061	0.3896	0.0406	11.8281
P17-1	35	0.4144	0.3846	0.0718	6.4623
P17-2	35	0.4606	0.3970	0.1379	3.1253
P17-3	35	0.0839	0.0800	0.0466	10.2260
P17-4	35	0.1394	0.1318	0.0546	8.6607
P17-5	35	0.0000	0.0000	****	****
P17-6	35	0.2173	0.1948	0.1038	4.3150
P17-7	35	0.4844	0.2187	0.5484	0.4117
P17-8	35	0.4994	0.3019	0.3955	0.7641
P17-9	35	0.4671	0.4312	0.0768	6.0123

```
 P17-10            35    0.0000    0.0000      ****      ****

Mean               35    0.1404    0.1121    0.2018    1.9781
St. Dev                  0.0325    0.0204
=============================================================

* Nm = estimate of gene flow from Gst or Gcs. E.g., Nm = 0.5(1 -
Gst)/Gst;
  See McDermott and McDonald, Ann. Rev. Phytopathol. 31:353-373 (1993).
The number of polymorphic loci is :     55
The percentage of polymorphic loci is : 58.51



**************************************************************************
************
   **
**
   **         Nei's Original Measures of Genetic Identity and Genetic
distance      **
   **                       [See Nei (1972) Am. Nat. 106:283-292)]
**
   **
**

**************************************************************************
************


=================================================
pop ID    1         2         3         4
=================================================
1           ****    0.9536    0.9423    0.9227
2         0.0475      ****    0.9903    0.9662
3         0.0594    0.0098      ****    0.9721
4         0.0804    0.0344    0.0283      ****
=================================================

Nei's genetic identity (above diagonal) and genetic distance (below
diagonal).



**************************************************************************
************
   **
**
   **          Dendrogram Based Nei's (1972) Genetic distance: Method =
UPGMA         **
   **              --Modified from NEIGHBOR procedure of PHYLIP Version
3.5            **
   **
**

**************************************************************************
************
```



* File Name: dgram1.plt

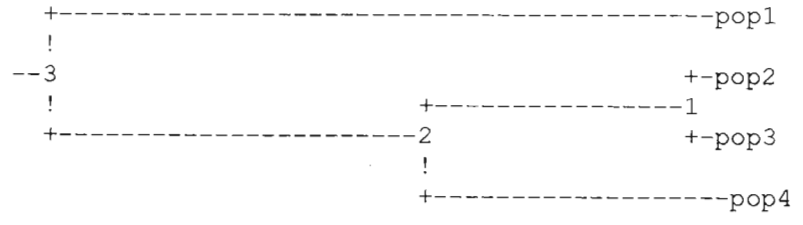
Between	And	Length
3	pop1	3.12342
3	2	1.55677
2	1	1.07867
1	pop2	0.48799
1	pop3	0.48799
2	pop4	1.56665

****
****
** **
** Nei's Unbiased Measures of Genetic Identity and Genetic distance **
** [See Nei (1978) Genetics 89:583-590] **
** **
****
****

pop ID	1	2	3	4
1	****	0.9628	0.9511	0.9323
2	0.0379	****	0.9969	0.9737
3	0.0502	0.0031	****	0.9794
4	0.0701	0.0266	0.0208	****

Nei's genetic identity (above diagonal) and genetic distance (below diagonal).

```
*************************************************************************
************
   **
**
   **          Dendrogram Based Nei's (1978) Genetic distance: Method =
UPGMA         **
   **              --Modified from NEIGHBOR procedure of PHYLIP Version
3.5            **
   **
**

*************************************************************************
************
```



* File Name: dgram2.plt

Between	And	Length
3	pop1	2.63639
3	2	1.45132
2	1	1.03027
1	pop2	0.15481
1	pop3	0.15481
2	pop4	1.18508

2. Family 18

INPUT FILE

/* Diploid RAPD Data Set */
Number of populations = 5
Number of loci = 94
Locus name :
A01-1 A01-2 A01-3 A01-4 A01-5 A01-6
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
A04-1 A04-2 A04-3
A06-1 A06-2 A06-3 A06-4 A06-5 A06-6
A09-1 A09-2 A09-3 A09-4 A09-5 A09-6 A09-7
A10-1 A10-2 A10-3 A10-4 A10-5 A10-6
A12-1 A12-2 A12-3 A12-4 A12-5 A12-6 A12-7 A12-8
A13-1 A13-2 A13-3 A13-4 A13-5
A17-1 A17-2 A17-3 A17-4
A19-1 A19-2 A19-3 A19-4 A19-5
A20-1 A20-2 A20-3 A20-4 A20-5 A20-6
C05-1 C05-2 C05-3 C05-4 C05-5 C05-6 C05-7 C05-8
C06-1 C06-2 C06-3 C06-4 C06-5 C06-6 C06-7
P17-1 P17-2 P17-3 P17-4 P17-5 P17-6 P17-7 P17-8 P17-9 P17-10

name = 18-1

fis = 0.00

111111 0110100100000 111 111010 1111101 101011 11111111 00011 1111 11111 000111 10111100 0111110 1000101111

name = 18-3

fis = 0.00

111111 0000100000000 111 000010 1111101 101000 11111111 00011 1111 11111 001111 00011101 0011110 1100101111

name = 18-4

fis = 0.00

111111 0000100000100 111 000010 0010111 101011 11111111 00011 1111 11111 000111 11011101 0011111 0100101111

name = 18-5

fis = 0.00

111111 0000100000100 111 000010 1111111 101010 11111111 00011 1111 11111 000111 00011100 0011111 0100101101

name = 18-7

fis = 0.00

111111 0000100000000 111 000010 1111101 000000 11111111 00011 1111 11111 011111 10011101 1011111 1100101101

RESULTS - FAMILY 18

```
**************************************************************************
************
   **
**
   **          Nei's Analysis of Gene Diversity in Subdivided
Populations           **
   **         [See Nei (1987) Molecular Evolutionary Genetics (p. 187-
192)]         **
   **
**

**************************************************************************
************
```

Locus	Sample Size	Ht	Hs	Gst	Nm*
A01-1	5	0.0000	0.0000	****	****
A01-2	5	0.0000	0.0000	****	****
A01-3	5	0.0000	0.0000	****	****
A01-4	5	0.0000	0.0000	****	****
A01-5	5	0.0000	0.0000	****	****
A01-6	5	0.0000	0.0000	****	****
A02-1	5	0.0000	0.0000	****	****
A02-2	5	0.3200	0.0000	1.0000	0.0000
A02-3	5	0.3200	0.0000	1.0000	0.0000
A02-4	5	0.0000	0.0000	****	****
A02-5	5	0.0000	0.0000	****	****
A02-6	5	0.0000	0.0000	****	****
A02-7	5	0.0000	0.0000	****	****
A02-8	5	0.3200	0.0000	1.0000	0.0000
A02-9	5	0.0000	0.0000	****	****
A02-10	5	0.0000	0.0000	****	****
A02-11	5	0.4800	0.0000	1.0000	0.0000
A02-12	5	0.0000	0.0000	****	****
A02-13	5	0.0000	0.0000	****	****
A04-1	5	0.0000	0.0000	****	****
A04-2	5	0.0000	0.0000	****	****
A04-3	5	0.0000	0.0000	****	****
A06-1	5	0.3200	0.0000	1.0000	0.0000
A06-2	5	0.3200	0.0000	1.0000	0.0000
A06-3	5	0.3200	0.0000	1.0000	0.0000
A06-4	5	0.0000	0.0000	****	****
A06-5	5	0.0000	0.0000	****	****
A06-6	5	0.0000	0.0000	****	****
A09-1	5	0.3200	0.0000	1.0000	0.0000
A09-2	5	0.3200	0.0000	1.0000	0.0000
A09-3	5	0.0000	0.0000	****	****
A09-4	5	0.3200	0.0000	1.0000	0.0000
A09-5	5	0.0000	0.0000	****	****
A09-6	5	0.4800	0.0000	1.0000	0.0000
A09-7	5	0.0000	0.0000	****	****
A10-1	5	0.3200	0.0000	1.0000	0.0000

A10-2	5	0.0000	0.0000	****	****
A10-3	5	0.3200	0.0000	1.0000	0.0000
A10-4	5	0.0000	0.0000	****	****
A10-5	5	0.4800	0.0000	1.0000	0.0000
A10-6	5	0.4800	0.0000	1.0000	0.0000
A12-1	5	0.0000	0.0000	****	****
A12-2	5	0.0000	0.0000	****	****
A12-3	5	0.0000	0.0000	****	****
A12-4	5	0.0000	0.0000	****	****
A12-5	5	0.0000	0.0000	****	****
A12-6	5	0.0000	0.0000	****	****
A12-7	5	0.0000	0.0000	****	****
A12-8	5	0.0000	0.0000	****	****
A13-1	5	0.0000	0.0000	****	****
A13-2	5	0.0000	0.0000	****	****
A13-3	5	0.0000	0.0000	****	****
A13-4	5	0.0000	0.0000	****	****
A13-5	5	0.0000	0.0000	****	****
A17-1	5	0.0000	0.0000	****	****
A17-2	5	0.0000	0.0000	****	****
A17-3	5	0.0000	0.0000	****	****
A17-4	5	0.0000	0.0000	****	****
A19-1	5	0.0000	0.0000	****	****
A19-2	5	0.0000	0.0000	****	****
A19-3	5	0.0000	0.0000	****	****
A19-4	5	0.0000	0.0000	****	****
A19-5	5	0.0000	0.0000	****	****
A20-1	5	0.0000	0.0000	****	****
A20-2	5	0.3200	0.0000	1.0000	0.0000
A20-3	5	0.4800	0.0000	1.0000	0.0000
A20-4	5	0.0000	0.0000	****	****
A20-5	5	0.0000	0.0000	****	****
A20-6	5	0.0000	0.0000	****	****
C05-1	5	0.4800	0.0000	1.0000	0.0000
C05-2	5	0.3200	0.0000	1.0000	0.0000
C05-3	5	0.3200	0.0000	1.0000	0.0000
C05-4	5	0.0000	0.0000	****	****
C05-5	5	0.0000	0.0000	****	****
C05-6	5	0.0000	0.0000	****	****
C05-7	5	0.0000	0.0000	****	****
C05-8	5	0.4800	0.0000	1.0000	0.0000
C06-1	5	0.3200	0.0000	1.0000	0.0000
C06-2	5	0.3200	0.0000	1.0000	0.0000
C06-3	5	0.0000	0.0000	****	****
C06-4	5	0.0000	0.0000	****	****
C06-5	5	0.0000	0.0000	****	****
C06-6	5	0.0000	0.0000	****	****
C06-7	5	0.4800	0.0000	1.0000	0.0000
P17-1	5	0.4800	0.0000	1.0000	0.0000
P17-2	5	0.3200	0.0000	1.0000	0.0000
P17-3	5	0.0000	0.0000	****	****
P17-4	5	0.0000	0.0000	****	****
P17-5	5	0.0000	0.0000	****	****
P17-6	5	0.0000	0.0000	****	****
P17-7	5	0.0000	0.0000	****	****
P17-8	5	0.0000	0.0000	****	****
P17-9	5	0.4800	0.0000	1.0000	0.0000

P17-10	5	0.0000	0.0000	****	****
Mean	5	0.1089	0.0000	1.0000	0.0000
St. Dev		0.0315	0.0000		

* Nm = estimate of gene flow from Gst or Gcs. E.g., Nm = 0.5(1 - Gst)/Gst;
See McDermott and McDonald, Ann. Rev. Phytopathol. 31:353-373 (1993).
The number of polymorphic loci is : 27
The percentage of polymorphic loci is : 28.72

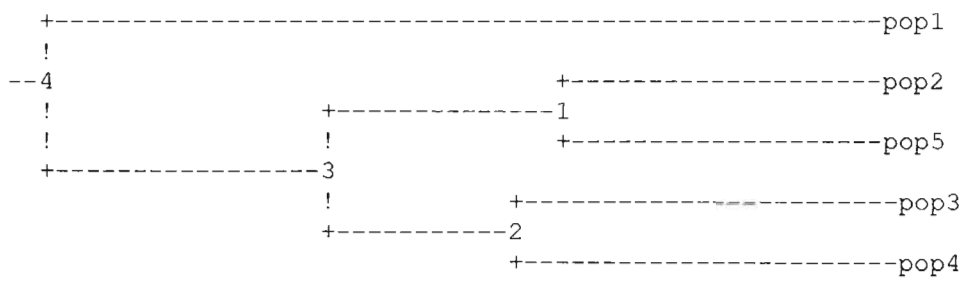
****
** **
** Nei's Original Measures of Genetic Identity and Genetic distance **
** [See Nei (1972) Am. Nat. 106:283-292)] **
** **
**

****

pop ID	1	2	3	4	5
1	****	0.8511	0.8085	0.8298	0.7979
2	0.1613	****	0.8723	0.9149	0.9255
3	0.2126	0.1366	****	0.9149	0.8404
4	0.1866	0.0889	0.0889	****	0.8830
5	0.2258	0.0774	0.1738	0.1245	****

Nei's genetic identity (above diagonal) and genetic distance (below diagonal).

```
*************************************************************************
************
   **
**
   **          Dendrogram Based Nei's (1972) Genetic distance: Method =
UPGMA         **
   **              --Modified from NEIGHBOR procedure of PHYLIP Version
3.5            **
   **
**

*************************************************************************
************
```



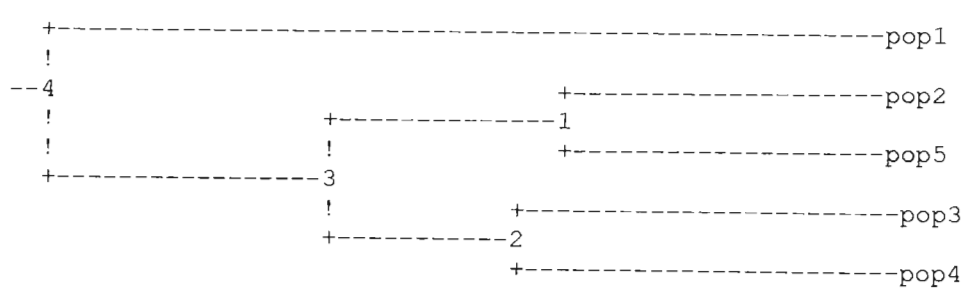

```
* File Name: dgram1.plt

Between         And              Length
-------         ---              ------
   4           pop1              9.82778
   4              3              3.27998
   3              1              2.67847
   1           pop2              3.86933
   1           pop5              3.86933
   3              2              2.10043
   2           pop3              4.44737
   2           pop4              4.44737



*************************************************************************
************
   **
**
   **          Nei's Unbiased Measures of Genetic Identity and Genetic
distance      **
   **                        [See Nei (1978)  Genetics 89:583-590]
**
   **
**
```

```
**************************************************************************
************
```

pop ID	1	2	3	4	5
1	****	0.8511	0.8085	0.8298	0.7979
2	0.1613	****	0.8723	0.9149	0.9255
3	0.2126	0.1366	****	0.9149	0.8404
4	0.1866	0.0889	0.0889	****	0.8830
5	0.2258	0.0774	0.1738	0.1245	****

Nei's genetic identity (above diagonal) and genetic distance (below diagonal).

```
**************************************************************************
************
   **
**
   **          Dendrogram Based Nei's (1978) Genetic distance: Method =
UPGMA         **
   **             --Modified from NEIGHBOR procedure of PHYLIP Version
3.5            **
   **
**

**************************************************************************
************
```




```
* File Name: dgram2.plt
```

Between	And	Length
4	pop1	9.82778
4	3	3.27998

3	1	2.67847
1	pop2	3.86933
1	pop5	3.86933
3	2	2.10043
2	pop3	4.44737
2	pop4	4.44737

3. Family 99

INPUT FILE

/* Diploid RAPD Data Set */
Number of populations = 11
Number of loci = 94
Locus name :
A01-1 A01-2 A01-3 A01-4 A01-5 A01-6
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
A04-1 A04-2 A04-3
A06-1 A06-2 A06-3 A06-4 A06-5 A06-6
A09-1 A09-2 A09-3 A09-4 A09-5 A09-6 A09-7
A10-1 A10-2 A10-3 A10-4 A10-5 A10-6
A12-1 A12-2 A12-3 A12-4 A12-5 A12-6 A12-7 A12-8
A13-1 A13-2 A13-3 A13-4 A13-5
A17-1 A17-2 A17-3 A17-4
A19-1 A19-2 A19-3 A19-4 A19-5
A20-1 A20-2 A20-3 A20-4 A20-5 A20-6
C05-1 C05-2 C05-3 C05-4 C05-5 C05-6 C05-7 C05-8
C06-1 C06-2 C06-3 C06-4 C06-5 C06-6 C06-7
P17-1 P17-2 P17-3 P17-4 P17-5 P17-6 P17-7 P17-8 P17-9 P17-10

name = 99-1

fis = 0.00

111111 0000100000100 111 010010 1111101 000011 11111111 00011 1111 11111 000111 01011101 1011110 1100100001

name = 99-2

fis = 0.00

111111 0000100000000 111 000010 1111001 000011 11111111 00011 1111 11111 000111 01011101 0011111 0001101101

name = 99-3

fis = 0.00

111111 0000100000100 111 000010 0100001 000001 11111111 00011 1111 11111 000101 10011110 0000110 0010100101

name = 99-4

fis = 0.00

111111 0000100000100 111 000010 0100001 001011 11111111 00011 1111 11111 000101 10011110 0000110 0010100101

name = 99-6

fis = 0.00

111111 0100100011000 111 001110 1111101 000101 11111111 00111 1111 11111 000111 10011100 0011111 1100100111

name = 99-8

fis = 0.00

111111 0000100000100 111 000010 1111001 010000 11111111 00111 1111 11111 000111 00011111 0011110 1100101111

name = 99-9

fis = 0.00

111111 0000100000000 111 000010 1111101 000011 11111111 00111 1111 11111 000111 00011101 0011111 1100101011

name = 99-10
fis = 0.00
111111 0000110000010 111 000010 1111111 000011 11111111 10011 1111 11111 000111 11011101 0001110 1100101001

name = 99-11
fis = 0.00
111111 0000110010000 111 000010 0101101 010000 11111111 00011 1111 11111 000111 11011100 0001110 0100111011

name = 99-12
fis = 0.00
111111 0000110000010 111 000010 1111101 000101 11111111 11011 1111 11111 000111 10011101 1001111 1100100001

name = 99-13
fis = 0.00
111111 1000110010010 111 000111 1111101 010011 11111111 10011 1111 11111 000111 11011101 1001110 1100101011

RESULTS – FAMILY 99

```
**********************************************************************
************
   **
**
   **          Nei's Analysis of Gene Diversity in Subdivided
Populations           **
   **          [See Nei (1987) Molecular Evolutionary Genetics (p. 187-
192)]         **
   **
**

**********************************************************************
************
```

Locus	Sample Size	Ht	Hs	Gst	Nm*
A01-1	11	0.0000	0.0000	****	****
A01-2	11	0.0000	0.0000	****	****
A01-3	11	0.0000	0.0000	****	****
A01-4	11	0.0000	0.0000	****	****
A01-5	11	0.0000	0.0000	****	****
A01-6	11	0.0000	0.0000	****	****
A02-1	11	0.1653	0.0000	1.0000	0.0000
A02-2	11	0.1653	0.0000	1.0000	0.0000
A02-3	11	0.0000	0.0000	****	****
A02-4	11	0.0000	0.0000	****	****
A02-5	11	0.0000	0.0000	****	****
A02-6	11	0.4628	0.0000	1.0000	0.0000
A02-7	11	0.0000	0.0000	****	****
A02-8	11	0.0000	0.0000	****	****
A02-9	11	0.3967	0.0000	1.0000	0.0000
A02-10	11	0.1653	0.0000	1.0000	0.0000
A02-11	11	0.4628	0.0000	1.0000	0.0000
A02-12	11	0.3967	0.0000	1.0000	0.0000
A02-13	11	0.0000	0.0000	****	****
A04-1	11	0.0000	0.0000	****	****
A04-2	11	0.0000	0.0000	****	****
A04-3	11	0.0000	0.0000	****	****
A06-1	11	0.0000	0.0000	****	****
A06-2	11	0.1653	0.0000	1.0000	0.0000
A06-3	11	0.1653	0.0000	1.0000	0.0000
A06-4	11	0.2975	0.0000	1.0000	0.0000
A06-5	11	0.0000	0.0000	****	****
A06-6	11	0.1653	0.0000	1.0000	0.0000
A09-1	11	0.3967	0.0000	1.0000	0.0000
A09-2	11	0.0000	0.0000	****	****
A09-3	11	0.3967	0.0000	1.0000	0.0000
A09-4	11	0.2975	0.0000	1.0000	0.0000
A09-5	11	0.4628	0.0000	1.0000	0.0000
A09-6	11	0.1653	0.0000	1.0000	0.0000
A09-7	11	0.0000	0.0000	****	****
A10-1	11	0.0000	0.0000	****	****

A10-2	11	0.3967	0.0000	1.0000	0.0000
A10-3	11	0.1653	0.0000	1.0000	0.0000
A10-4	11	0.2975	0.0000	1.0000	0.0000
A10-5	11	0.4959	0.0000	1.0000	0.0000
A10-6	11	0.2975	0.0000	1.0000	0.0000
A12-1	11	0.0000	0.0000	****	****
A12-2	11	0.0000	0.0000	****	****
A12-3	11	0.0000	0.0000	****	****
A12-4	11	0.0000	0.0000	****	****
A12-5	11	0.0000	0.0000	****	****
A12-6	11	0.0000	0.0000	****	****
A12-7	11	0.0000	0.0000	****	****
A12-8	11	0.0000	0.0000	****	****
A13-1	11	0.3967	0.0000	1.0000	0.0000
A13-2	11	0.1653	0.0000	1.0000	0.0000
A13-3	11	0.3967	0.0000	1.0000	0.0000
A13-4	11	0.0000	0.0000	****	****
A13-5	11	0.0000	0.0000	****	****
A17-1	11	0.0000	0.0000	****	****
A17-2	11	0.0000	0.0000	****	****
A17-3	11	0.0000	0.0000	****	****
A17-4	11	0.0000	0.0000	****	****
A19-1	11	0.0000	0.0000	****	****
A19-2	11	0.0000	0.0000	****	****
A19-3	11	0.0000	0.0000	****	****
A19-4	11	0.0000	0.0000	****	****
A19-5	11	0.0000	0.0000	****	****
A20-1	11	0.0000	0.0000	****	****
A20-2	11	0.0000	0.0000	****	****
A20-3	11	0.0000	0.0000	****	****
A20-4	11	0.0000	0.0000	****	****
A20-5	11	0.2975	0.0000	1.0000	0.0000
A20-6	11	0.0000	0.0000	****	****
C05-1	11	0.4628	0.0000	1.0000	0.0000
C05-2	11	0.4959	0.0000	1.0000	0.0000
C05-3	11	0.0000	0.0000	****	****
C05-4	11	0.0000	0.0000	****	****
C05-5	11	0.0000	0.0000	****	****
C05-6	11	0.0000	0.0000	****	****
C05-7	11	0.3967	0.0000	1.0000	0.0000
C05-8	11	0.4628	0.0000	1.0000	0.0000
C06-1	11	0.3967	0.0000	1.0000	0.0000
C06-2	11	0.0000	0.0000	****	****
C06-3	11	0.4959	0.0000	1.0000	0.0000
C06-4	11	0.2975	0.0000	1.0000	0.0000
C06-5	11	0.0000	0.0000	****	****
C06-6	11	0.0000	0.0000	****	****
C06-7	11	0.4628	0.0000	1.0000	0.0000
P17-1	11	0.4628	0.0000	1.0000	0.0000
P17-2	11	0.3967	0.0000	1.0000	0.0000
P17-3	11	0.2975	0.0000	1.0000	0.0000
P17-4	11	0.1653	0.0000	1.0000	0.0000
P17-5	11	0.0000	0.0000	****	****
P17-6	11	0.1653	0.0000	1.0000	0.0000
P17-7	11	0.4959	0.0000	1.0000	0.0000
P17-8	11	0.4959	0.0000	1.0000	0.0000
P17-9	11	0.4959	0.0000	1.0000	0.0000

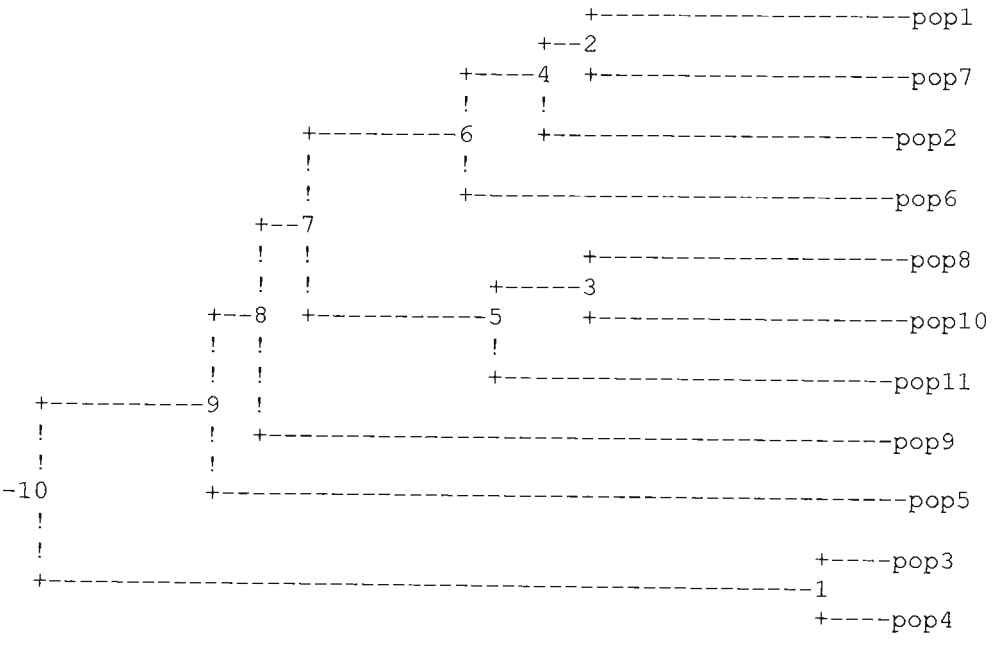
```
 P17-10            11     0.0000     0.0000       ****       ****

Mean               11     0.1498     0.0000     1.0000     0.0000
St. Dev                   0.0361     0.0000
=============================================================

* Nm = estimate of gene flow from Gst or Gcs. E.g., Nm = 0.5(1 -
Gst)/Gst;
  See McDermott and McDonald, Ann. Rev. Phytopathol. 31:353-373 (1993).
The number of polymorphic loci is :      41
The percentage of polymorphic loci is : 43.62


*************************************************************************
************
   **
**
   **         Nei's Original Measures of Genetic Identity and Genetic
distance      **
   **                      [See Nei (1972) Am. Nat. 106:283-292)]
**
   **
**

*************************************************************************
************


=========================================================================
=============================================
pop ID    1         2         3         4         5         6         7
8         9         10        11
=========================================================================
=============================================
1           ****    0.8936    0.8085    0.8085    0.8191    0.8723
0.9149    0.8936    0.8191    0.8723    0.8511
2         0.1125      ****    0.8298    0.8298    0.8191    0.8723
0.9149    0.8723    0.8191    0.8298    0.8085
3         0.2126    0.1866      ****    0.9787    0.7766    0.8298
0.7872    0.7872    0.8191    0.7872    0.7234
4         0.2126    0.1866    0.0215      ****    0.7553    0.8085
0.7872    0.7872    0.7979    0.7660    0.7234
5         0.1995    0.1995    0.2528    0.2806      ****    0.8404
0.8830    0.7979    0.8085    0.8404    0.7979
6         0.1366    0.1366    0.1866    0.2126    0.1738      ****
0.9149    0.8298    0.8404    0.8085    0.8085
7         0.0889    0.0889    0.2392    0.2392    0.1245    0.0889
****    0.8936    0.8404    0.8723    0.8511
8         0.1125    0.1366    0.2392    0.2392    0.2258    0.1866
0.1125      ****    0.8617    0.9149    0.9149
9         0.1995    0.1995    0.1995    0.2258    0.2126    0.1738
0.1738    0.1488      ****    0.8191    0.8617
10        0.1366    0.1866    0.2392    0.2666    0.1738    0.2126
0.1366    0.0889    0.1995      ****    0.8723
11        0.1613    0.2126    0.3238    0.3238    0.2258    0.2126
0.1613    0.0889    0.1488    0.1366      ****
```

==========
==========

Nei's genetic identity (above diagonal) and genetic distance (below diagonal).

****
****
** **
** Dendrogram Based Nei's (1972) Genetic distance: Method = UPGMA **
** --Modified from NEIGHBOR procedure of PHYLIP Version 3.5 **
** **
**

****
****



* File Name: dgram1.plt

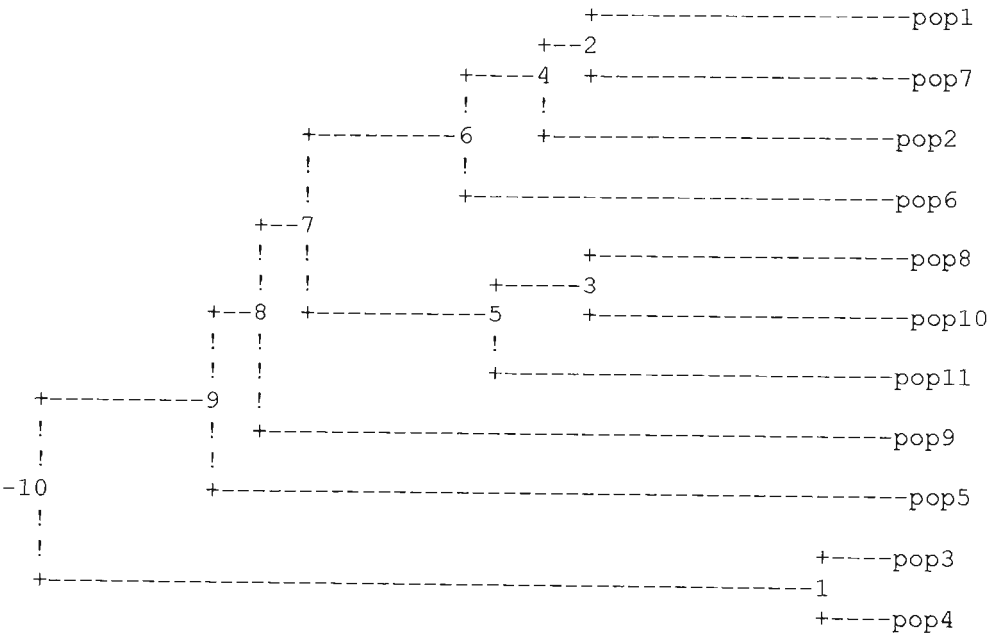
Between	And	Length
10	9	2.25586
9	8	0.71097
8	7	0.68435

7	6	2.16534
6	4	0.99934
4	2	0.58826
2	pop1	4.44737
2	pop7	4.44737
4	pop2	5.03564
6	pop6	6.03498
7	5	2.56224
5	3	1.19070
3	pop8	4.44737
3	pop10	4.44737
5	pop11	5.63808
8	pop9	8.88467
9	pop5	9.59563
10	1	10.77619
1	pop3	1.07531
1	pop4	1.07531

```
**************************************************************************
************
   **
**
   **          Nei's Unbiased Measures of Genetic Identity and Genetic
distance       **
   **                       [See Nei (1978)  Genetics 89:583-590]
**
   **
**

**************************************************************************
************
```

```
==========================================================================
==========================================
pop ID    1         2         3         4         5         6         7
8         9         10        11
==========================================================================
==========================================
1           ****    0.8936    0.8085    0.8085    0.8191    0.8723
0.9149    0.8936    0.8191    0.8723    0.8511
2         0.1125      ****    0.8298    0.8298    0.8191    0.8723
0.9149    0.8723    0.8191    0.8298    0.8085
3         0.2126    0.1866      ****    0.9787    0.7766    0.8298
0.7872    0.7872    0.8191    0.7872    0.7234
4         0.2126    0.1866    0.0215      ****    0.7553    0.8085
0.7872    0.7872    0.7979    0.7660    0.7234
5         0.1995    0.1995    0.2528    0.2806      ****    0.8404
0.8830    0.7979    0.8085    0.8404    0.7979
6         0.1366    0.1366    0.1866    0.2126    0.1738      ****
0.9149    0.8298    0.8404    0.8085    0.8085
7         0.0889    0.0889    0.2392    0.2392    0.1245    0.0889
****    0.8936    0.8404    0.8723    0.8511
```

```
8          0.1125    0.1366    0.2392    0.2392    0.2258    0.1866
0.1125       ****    0.8617    0.9149    0.9149
9          0.1995    0.1995    0.1995    0.2258    0.2126    0.1738
0.1738     0.1488      ****    0.8191    0.8617
10         0.1366    0.1866    0.2392    0.2666    0.1738    0.2126
0.1366     0.0889    0.1995      ****    0.8723
11         0.1613    0.2126    0.3238    0.3238    0.2258    0.2126
0.1613     0.0889    0.1488    0.1366      ****
================================================================================
================================================

Nei's genetic identity (above diagonal) and genetic distance (below
diagonal).



********************************************************************************
************
   **
**
   **          Dendrogram Based Nei's (1978) Genetic distance: Method =
UPGMA         **
   **              --Modified from NEIGHBOR procedure of PHYLIP Version
3.5            **
   **
**

********************************************************************************
************
```



* File Name: dgram2.plt

Between	And	Length
10	9	2.25586
9	8	0.71097
8	7	0.68435
7	6	2.16534
6	4	0.99934
4	2	0.58826
2	pop1	4.44737
2	pop7	4.44737
4	pop2	5.03564
6	pop6	6.03498
7	5	2.56224
5	3	1.19070
3	pop8	4.44737
3	pop10	4.44737
5	pop11	5.63808
8	pop9	8.88467
9	pop5	9.59563
10	1	10.77619
1	pop3	1.07531
1	pop4	1.07531

4. Family 101

INPUT FILE

/* Diploid RAPD Data Set */
Number of populations = 13
Number of loci = 94
Locus name :
A01-1 A01-2 A01-3 A01-4 A01-5 A01-6
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
A04-1 A04-2 A04-3
A06-1 A06-2 A06-3 A06-4 A06-5 A06-6
A09-1 A09-2 A09-3 A09-4 A09-5 A09-6 A09-7
A10-1 A10-2 A10-3 A10-4 A10-5 A10-6
A12-1 A12-2 A12-3 A12-4 A12-5 A12-6 A12-7 A12-8
A13-1 A13-2 A13-3 A13-4 A13-5
A17-1 A17-2 A17-3 A17-4
A19-1 A19-2 A19-3 A19-4 A19-5
A20-1 A20-2 A20-3 A20-4 A20-5 A20-6
C05-1 C05-2 C05-3 C05-4 C05-5 C05-6 C05-7 C05-8
C06-1 C06-2 C06-3 C06-4 C06-5 C06-6 C06-7
P17-1 P17-2 P17-3 P17-4 P17-5 P17-6 P17-7 P17-8 P17-9 P17-10

name = 101-1
fis = 0.00
111111 0100100001000 111 000010 0101101 000001 11111111 00011 1111 11111 000111 11011100 0010110 0000101001

name = 101-2
fis = 0.00
111111 0001110000010 111 000111 1101111 000011 11111111 10011 1111 11111 100111 11011101 1001111 1100101011

name = 101-3
fis = 0.00
111111 0000100000010 111 000010 0101101 000011 11111111 00011 1111 11111 000101 00011101 0001111 0001100001

name = 101-4
fis = 0.00
111111 0000100000001 111 000010 0100011 000000 11111111 00011 1111 11111 000101 10011111 0000110 0010101001

name = 101-5
fis = 0.00
111111 0000110010001 111 000011 1111101 010010 11111111 10011 1111 11111 000111 11011101 1001111 1100111001

name = 101-7
fis = 0.00
111111 0000110000001 111 000011 1111101 010011 11111111 10011 1111 11111 000111 11011101 1001111 0100101001

name = 101-8
fis = 0.00
111111 0000110000001 111 000011 1111101 010000 11111111 10011 1111 11111 000111 11011101 1001111 1100101011

name = 101-9
fis = 0.00
111111 0000110010000 111 001111 1111101 010011 11111111 00011 1111 11111 010111 01011101 1011110 1100111001

name = 101-10
fis = 0.00
111111 1000101000000 111 000110 1101101 001001 11111111 00011 1111 11111 000111 11011101 0101111 1110110111

name = 101-11
fis = 0.00
111111 0000100000000 111 000010 0000001 000000 11111111 00011 1111 11111 000101 10011101 0000110 0001100111

name = 101-12
fis = 0.00
111111 0100100101000 111 000111 1101101 001001 11111111 00011 1111 11111 000111 11011100 0101110 1100110111

name = 101-13
fis = 0.00
111111 0000100000000 111 000011 1101101 000001 11111111 00011 1111 11111 000111 01011101 0001111 0001110111

name = 101-14

fis = 0.00
111111 0000100000000 111 000011 1101101 000001 11111111 00011 1111 11111 000111 01011101 0001111 1001110101

RESULTS – FAMILY 101

```
**************************************************************************
***********
   **
**
   **          Nei's Analysis of Gene Diversity in Subdivided
Populations          **
   **          [See Nei (1987) Molecular Evolutionary Genetics (p. 187-
192)]         **
   **
**

**************************************************************************
***********
```

Locus	Sample Size	Ht	Hs	Gst	Nm*
A01-1	13	0.0000	0.0000	****	****
A01-2	13	0.0000	0.0000	****	****
A01-3	13	0.0000	0.0000	****	****
A01-4	13	0.0000	0.0000	****	****
A01-5	13	0.0000	0.0000	****	****
A01-6	13	0.0000	0.0000	****	****
A02-1	13	0.1420	0.0000	1.0000	0.0000
A02-2	13	0.2604	0.0000	1.0000	0.0000
A02-3	13	0.0000	0.0000	****	****
A02-4	13	0.1420	0.0000	1.0000	0.0000
A02-5	13	0.0000	0.0000	****	****
A02-6	13	0.4734	0.0000	1.0000	0.0000
A02-7	13	0.1420	0.0000	1.0000	0.0000
A02-8	13	0.1420	0.0000	1.0000	0.0000
A02-9	13	0.2604	0.0000	1.0000	0.0000
A02-10	13	0.2604	0.0000	1.0000	0.0000
A02-11	13	0.0000	0.0000	****	****
A02-12	13	0.2604	0.0000	1.0000	0.0000
A02-13	13	0.4260	0.0000	1.0000	0.0000
A04-1	13	0.0000	0.0000	****	****
A04-2	13	0.0000	0.0000	****	****
A04-3	13	0.0000	0.0000	****	****
A06-1	13	0.0000	0.0000	****	****
A06-2	13	0.0000	0.0000	****	****
A06-3	13	0.1420	0.0000	1.0000	0.0000
A06-4	13	0.4260	0.0000	1.0000	0.0000
A06-5	13	0.0000	0.0000	****	****
A06-6	13	0.4734	0.0000	1.0000	0.0000
A09-1	13	0.4260	0.0000	1.0000	0.0000
A09-2	13	0.1420	0.0000	1.0000	0.0000
A09-3	13	0.4260	0.0000	1.0000	0.0000
A09-4	13	0.2604	0.0000	1.0000	0.0000
A09-5	13	0.2604	0.0000	1.0000	0.0000
A09-6	13	0.2604	0.0000	1.0000	0.0000
A09-7	13	0.0000	0.0000	****	****
A10-1	13	0.0000	0.0000	****	****

A10-2	13	0.4260	0.0000	1.0000	0.0000
A10-3	13	0.2604	0.0000	1.0000	0.0000
A10-4	13	0.0000	0.0000	****	****
A10-5	13	0.4734	0.0000	1.0000	0.0000
A10-6	13	0.4260	0.0000	1.0000	0.0000
A12-1	13	0.0000	0.0000	****	****
A12-2	13	0.0000	0.0000	****	****
A12-3	13	0.0000	0.0000	****	****
A12-4	13	0.0000	0.0000	****	****
A12-5	13	0.0000	0.0000	****	****
A12-6	13	0.0000	0.0000	****	****
A12-7	13	0.0000	0.0000	****	****
A12-8	13	0.0000	0.0000	****	****
A13-1	13	0.4260	0.0000	1.0000	0.0000
A13-2	13	0.0000	0.0000	****	****
A13-3	13	0.0000	0.0000	****	****
A13-4	13	0.0000	0.0000	****	****
A13-5	13	0.0000	0.0000	****	****
A17-1	13	0.0000	0.0000	****	****
A17-2	13	0.0000	0.0000	****	****
A17-3	13	0.0000	0.0000	****	****
A17-4	13	0.0000	0.0000	****	****
A19-1	13	0.0000	0.0000	****	****
A19-2	13	0.0000	0.0000	****	****
A19-3	13	0.0000	0.0000	****	****
A19-4	13	0.0000	0.0000	****	****
A19-5	13	0.0000	0.0000	****	****
A20-1	13	0.1420	0.0000	1.0000	0.0000
A20-2	13	0.1420	0.0000	1.0000	0.0000
A20-3	13	0.0000	0.0000	****	****
A20-4	13	0.0000	0.0000	****	****
A20-5	13	0.3550	0.0000	1.0000	0.0000
A20-6	13	0.0000	0.0000	****	****
C05-1	13	0.4260	0.0000	1.0000	0.0000
C05-2	13	0.3550	0.0000	1.0000	0.0000
C05-3	13	0.0000	0.0000	****	****
C05-4	13	0.0000	0.0000	****	****
C05-5	13	0.0000	0.0000	****	****
C05-6	13	0.0000	0.0000	****	****
C05-7	13	0.1420	0.0000	1.0000	0.0000
C05-8	13	0.2604	0.0000	1.0000	0.0000
C06-1	13	0.4734	0.0000	1.0000	0.0000
C06-2	13	0.2604	0.0000	1.0000	0.0000
C06-3	13	0.2604	0.0000	1.0000	0.0000
C06-4	13	0.3550	0.0000	1.0000	0.0000
C06-5	13	0.0000	0.0000	****	****
C06-6	13	0.0000	0.0000	****	****
C06-7	13	0.4734	0.0000	1.0000	0.0000
P17-1	13	0.4970	0.0000	1.0000	0.0000
P17-2	13	0.4970	0.0000	1.0000	0.0000
P17-3	13	0.2604	0.0000	1.0000	0.0000
P17-4	13	0.4260	0.0000	1.0000	0.0000
P17-5	13	0.0000	0.0000	****	****
P17-6	13	0.4970	0.0000	1.0000	0.0000
P17-7	13	0.4970	0.0000	1.0000	0.0000
P17-8	13	0.4734	0.0000	1.0000	0.0000
P17-9	13	0.4970	0.0000	1.0000	0.0000

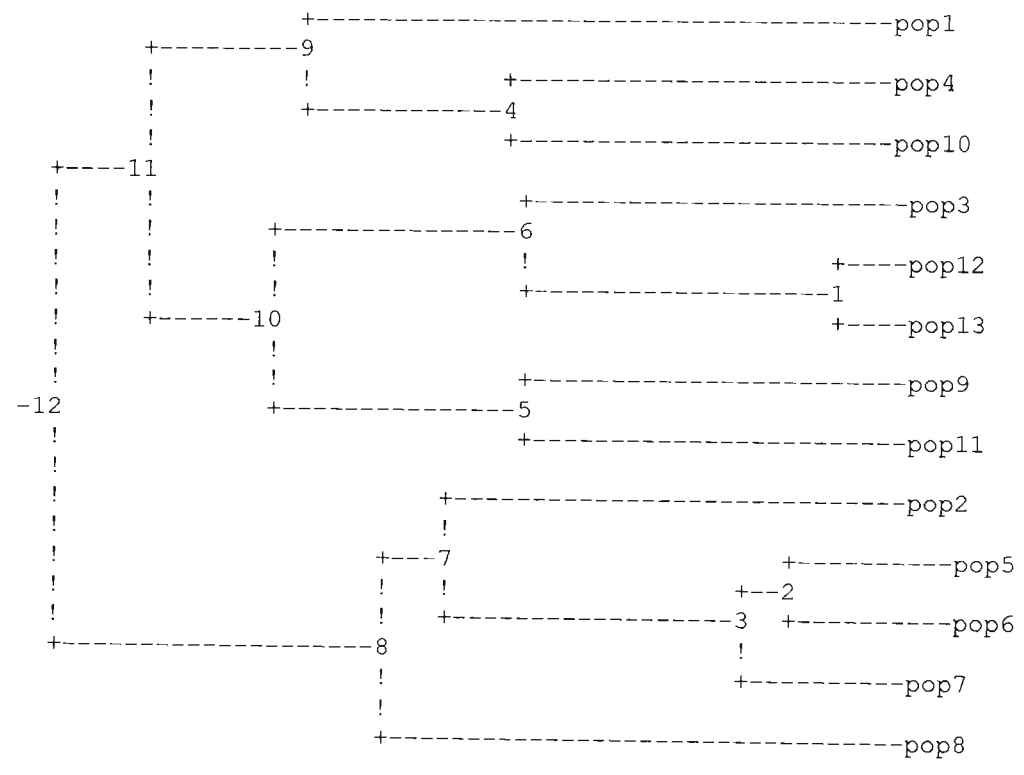
P17-10	13	0.0000	0.0000	****	****
Mean	13	0.1556	0.0000	1.0000	0.0000
St. Dev		0.0356	0.0000		

* Nm = estimate of gene flow from Gst or Gcs. E.g., Nm = 0.5(1 - Gst)/Gst;
See McDermott and McDonald, Ann. Rev. Phytopathol. 31:353-373 (1993).
The number of polymorphic loci is : 44
The percentage of polymorphic loci is : 46.81

****
**
**
** Nei's Original Measures of Genetic Identity and Genetic distance **
** [See Nei (1972) Am. Nat. 106:283-292)]
**
**
****

pop ID	1	2	3	4	5	6	7
8	9	10	11	12	13		
1	****	0.7872	0.8617	0.8617	0.7872	0.8298	
0.8085	0.7979	0.7979	0.8511	0.8511	0.8511	0.8511	
2	0.2392	****	0.8191	0.7553	0.8723	0.8936	
0.8936	0.8404	0.8191	0.7447	0.8085	0.8298	0.8298	
3	0.1488	0.1995	****	0.8511	0.7979	0.8404	
0.7979	0.7872	0.8085	0.8830	0.7766	0.9043	0.9043	
4	0.1488	0.2806	0.1613	****	0.7766	0.7979	
0.7979	0.7234	0.7660	0.9043	0.7340	0.7979	0.7979	
5	0.2392	0.1366	0.2258	0.2528	****	0.9574	
0.9574	0.9043	0.7979	0.7447	0.7872	0.8298	0.8511	
6	0.1866	0.1125	0.1738	0.2258	0.0435	****	
0.9574	0.8830	0.7979	0.7660	0.7872	0.8511	0.8511	
7	0.2126	0.1125	0.2258	0.2258	0.0435	0.0435	
****	0.8617	0.8191	0.7872	0.8085	0.8511	0.8511	
8	0.2258	0.1738	0.2392	0.3238	0.1006	0.1245	
0.1488	****	0.7872	0.7128	0.7979	0.8191	0.8404	
9	0.2258	0.1995	0.2126	0.2666	0.2258	0.2258	
0.1995	0.2392	****	0.7979	0.9043	0.8830	0.8830	
10	0.1613	0.2948	0.1245	0.1006	0.2948	0.2666	
0.2392	0.3386	0.2258	****	0.7872	0.8723	0.8511	
11	0.1613	0.2126	0.2528	0.3092	0.2392	0.2392	
0.2126	0.2258	0.1006	0.2392	****	0.8723	0.8723	

```
12          0.1613    0.1866    0.1006    0.2258    0.1866    0.1613
0.1613      0.1995    0.1245    0.1366    0.1366      ****    0.9787
13          0.1613    0.1866    0.1006    0.2258    0.1613    0.1613
0.1613      0.1738    0.1245    0.1613    0.1366    0.0215      ****
=============================================================================
=============================================================================

Nei's genetic identity (above diagonal) and genetic distance (below
diagonal).



*****************************************************************************
***********
   **
**
   **          Dendrogram Based Nei's (1972) Genetic distance: Method =
UPGMA         **
   **             --Modified from NEIGHBOR procedure of PHYLIP Version
3.5           **
   **
**

*****************************************************************************
***********
```



* File Name: dgram1.plt

Between	And	Length
12	11	1.30480
11	9	2.02879
9	pop1	7.75284
9	4	2.72067
4	pop4	5.03218
4	pop10	5.03218
11	10	1.55283
10	6	3.19663
6	pop3	5.03218
6	1	3.95687
1	pop12	1.07531
1	pop13	1.07531
10	5	3.19663
5	pop9	5.03218
5	pop11	5.03218
12	8	4.23906
8	7	0.82185
7	pop2	6.02552
7	3	3.85127
3	2	0.00000
2	pop5	2.17426
2	pop6	2.17426
3	pop7	2.17426
8	pop8	6.84738

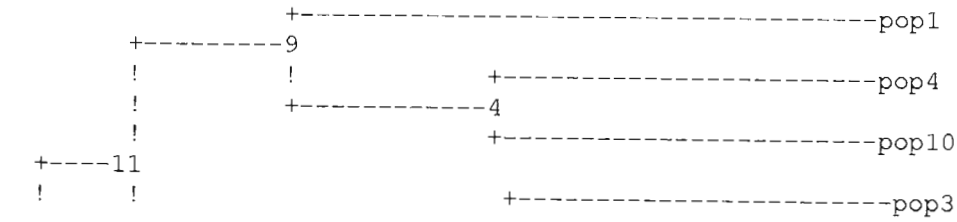
```
*************************************************************************
***********
   **
**
   **         Nei's Unbiased Measures of Genetic Identity and Genetic
distance      **
   **                      [See Nei (1978)  Genetics 89:583-590]
**
   **
**

*************************************************************************
***********
```

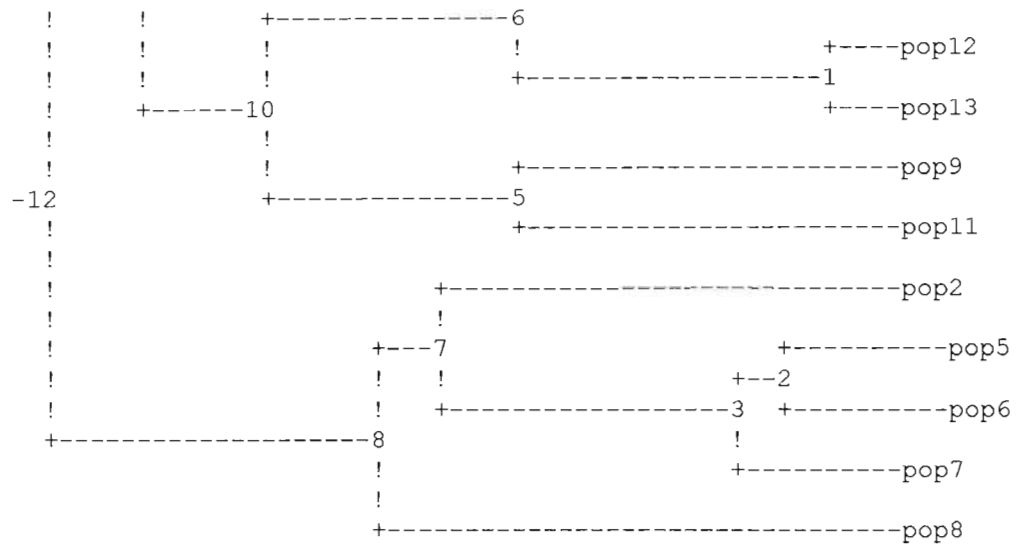
pop ID	1	2	3	4	5	6	7
8	9	10	11	12	13		
1	****	0.7872	0.8617	0.8617	0.7872	0.8298	
0.8085	0.7979	0.7979	0.8511	0.8511	0.8511	0.8511	

```
2          0.2392       ****     0.8191     0.7553     0.8723     0.8936
0.8936     0.8404     0.8191     0.7447     0.8085     0.8298     0.8298
3          0.1488     0.1995       ****     0.8511     0.7979     0.8404
0.7979     0.7872     0.8085     0.8830     0.7766     0.9043     0.9043
4          0.1488     0.2806     0.1613       ****     0.7766     0.7979
0.7979     0.7234     0.7660     0.9043     0.7340     0.7979     0.7979
5          0.2392     0.1366     0.2258     0.2528       ****     0.9574
0.9574     0.9043     0.7979     0.7447     0.7872     0.8298     0.8511
6          0.1866     0.1125     0.1738     0.2258     0.0435       ****
0.9574     0.8830     0.7979     0.7660     0.7872     0.8511     0.8511
7          0.2126     0.1125     0.2258     0.2258     0.0435     0.0435
****     0.8617     0.8191     0.7872     0.8085     0.8511     0.8511
8          0.2258     0.1738     0.2392     0.3238     0.1006     0.1245
0.1488       ****     0.7872     0.7128     0.7979     0.8191     0.8404
9          0.2258     0.1995     0.2126     0.2666     0.2258     0.2258
0.1995     0.2392       ****     0.7979     0.9043     0.8830     0.8830
10         0.1613     0.2948     0.1245     0.1006     0.2948     0.2666
0.2392     0.3386     0.2258       ****     0.7872     0.8723     0.8511
11         0.1613     0.2126     0.2528     0.3092     0.2392     0.2392
0.2126     0.2258     0.1006     0.2392       ****     0.8723     0.8723
12         0.1613     0.1866     0.1006     0.2258     0.1866     0.1613
0.1613     0.1995     0.1245     0.1366     0.1366       ****     0.9787
13         0.1613     0.1866     0.1006     0.2258     0.1613     0.1613
0.1613     0.1738     0.1245     0.1613     0.1366     0.0215       ****
================================================================================
================================================================
```

Nei's genetic identity (above diagonal) and genetic distance (below diagonal).

```
********************************************************************************
***********
   **
**
   **          Dendrogram Based Nei's (1978) Genetic distance: Method =
UPGMA         **
   **              --Modified from NEIGHBOR procedure of PHYLIP Version
3.5            **
   **
**

********************************************************************************
***********
```





* File Name: dgram2.plt

Between	And	Length
12	11	1.30480
11	9	2.02879
9	pop1	7.75284
9	4	2.72067
4	pop4	5.03218
4	pop10	5.03218
11	10	1.55283
10	6	3.19663
6	pop3	5.03218
6	1	3.95687
1	pop12	1.07531
1	pop13	1.07531
10	5	3.19663
5	pop9	5.03218
5	pop11	5.03218
12	8	4.23906
8	7	0.82185
7	pop2	6.02552
7	3	3.85127
3	2	0.00000
2	pop5	2.17426
2	pop6	2.17426
3	pop7	2.17426
8	pop8	6.84738

5. Family 101x77

INPUT FILE

/* Diploid RAPD Data Set */
Number of populations = 6
Number of loci = 94
Locus name :
A01-1 A01-2 A01-3 A01-4 A01-5 A01-6
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
A04-1 A04-2 A04-3
A06-1 A06-2 A06-3 A06-4 A06-5 A06-6
A09-1 A09-2 A09-3 A09-4 A09-5 A09-6 A09-7
A10-1 A10-2 A10-3 A10-4 A10-5 A10-6
A12-1 A12-2 A12-3 A12-4 A12-5 A12-6 A12-7 A12-8
A13-1 A13-2 A13-3 A13-4 A13-5
A17-1 A17-2 A17-3 A17-4
A19-1 A19-2 A19-3 A19-4 A19-5
A20-1 A20-2 A20-3 A20-4 A20-5 A20-6
C05-1 C05-2 C05-3 C05-4 C05-5 C05-6 C05-7 C05-8
C06-1 C06-2 C06-3 C06-4 C06-5 C06-6 C06-7
P17-1 P17-2 P17-3 P17-4 P17-5 P17-6 P17-7 P17-8 P17-9 P17-10

name = 101x77-2
fis = 0.00
111111 0000100000000 111 000110 0000001 000000 11111111 00011 1111 11111 000101 00011101 0000110 0001100111

name = 101x77-3
fis = 0.00
111111 0000100000000 111 000010 0000001 000000 11111111 00011 1111 11111 000111 00011101 0000110 0000100111

name = 101x77-4
fis = 0.00
111111 0000100000100 111 000110 1000101 000010 11111111 00011 1111 11111 000111 10011100 0000111 0000100111

name = 101x77-5
fis = 0.00
111111 0000100000100 111 000010 0000001 000000 11111111 00011 1111 11111 000101 10011101 0000110 0000110111

name = 101x77-8
fis = 0.00
111111 0001100000100 111 000110 1111101 001001 11111111 10011 1111 11111 000111 01011101 1001111 1100110111

name = 101x77-1
fis = 0.00
111111 0000100000100 111 000011 0011101 000010 11111111 00001 1111 11111 000101 01011101 1001110 0000100001

RESULTS – FAMILY 101X77

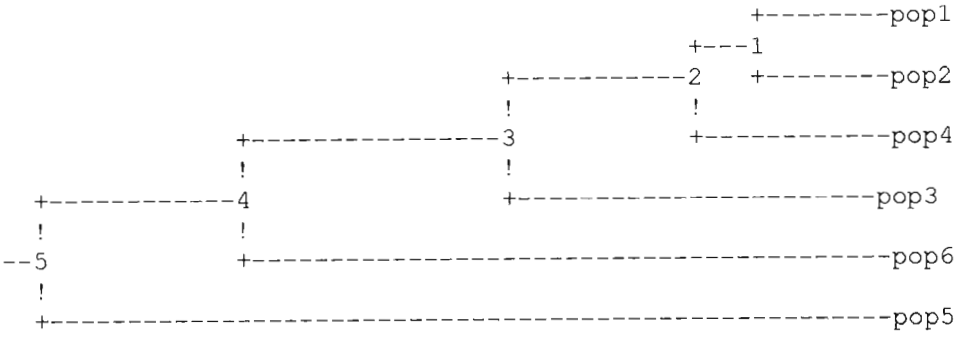
```
**************************************************************************
************
   **
**
   **          Nei's Analysis of Gene Diversity in Subdivided
Populations           **
   **         [See Nei (1987) Molecular Evolutionary Genetics (p. 187-
192)]         **
   **
**

**************************************************************************
************
```

Locus	Sample Size	Ht	Hs	Gst	Nm*
A01-1	6	0.0000	0.0000	****	****
A01-2	6	0.0000	0.0000	****	****
A01-3	6	0.0000	0.0000	****	****
A01-4	6	0.0000	0.0000	****	****
A01-5	6	0.0000	0.0000	****	****
A01-6	6	0.0000	0.0000	****	****
A02-1	6	0.0000	0.0000	****	****
A02-2	6	0.0000	0.0000	****	****
A02-3	6	0.0000	0.0000	****	****
A02-4	6	0.2778	0.0000	1.0000	0.0000
A02-5	6	0.0000	0.0000	****	****
A02-6	6	0.0000	0.0000	****	****
A02-7	6	0.0000	0.0000	****	****
A02-8	6	0.0000	0.0000	****	****
A02-9	6	0.0000	0.0000	****	****
A02-10	6	0.0000	0.0000	****	****
A02-11	6	0.4444	0.0000	1.0000	0.0000
A02-12	6	0.0000	0.0000	****	****
A02-13	6	0.0000	0.0000	****	****
A04-1	6	0.0000	0.0000	****	****
A04-2	6	0.0000	0.0000	****	****
A04-3	6	0.0000	0.0000	****	****
A06-1	6	0.0000	0.0000	****	****
A06-2	6	0.0000	0.0000	****	****
A06-3	6	0.0000	0.0000	****	****
A06-4	6	0.5000	0.0000	1.0000	0.0000
A06-5	6	0.0000	0.0000	****	****
A06-6	6	0.2778	0.0000	1.0000	0.0000
A09-1	6	0.4444	0.0000	1.0000	0.0000
A09-2	6	0.2778	0.0000	1.0000	0.0000
A09-3	6	0.4444	0.0000	1.0000	0.0000
A09-4	6	0.4444	0.0000	1.0000	0.0000
A09-5	6	0.5000	0.0000	1.0000	0.0000
A09-6	6	0.0000	0.0000	****	****
A09-7	6	0.0000	0.0000	****	****
A10-1	6	0.0000	0.0000	****	****

A10-2	6	0.0000	0.0000	****	****
A10-3	6	0.2778	0.0000	1.0000	0.0000
A10-4	6	0.0000	0.0000	****	****
A10-5	6	0.4444	0.0000	1.0000	0.0000
A10-6	6	0.2778	0.0000	1.0000	0.0000
A12-1	6	0.0000	0.0000	****	****
A12-2	6	0.0000	0.0000	****	****
A12-3	6	0.0000	0.0000	****	****
A12-4	6	0.0000	0.0000	****	****
A12-5	6	0.0000	0.0000	****	****
A12-6	6	0.0000	0.0000	****	****
A12-7	6	0.0000	0.0000	****	****
A12-8	6	0.0000	0.0000	****	****
A13-1	6	0.2778	0.0000	1.0000	0.0000
A13-2	6	0.0000	0.0000	****	****
A13-3	6	0.0000	0.0000	****	****
A13-4	6	0.2778	0.0000	1.0000	0.0000
A13-5	6	0.0000	0.0000	****	****
A17-1	6	0.0000	0.0000	****	****
A17-2	6	0.0000	0.0000	****	****
A17-3	6	0.0000	0.0000	****	****
A17-4	6	0.0000	0.0000	****	****
A19-1	6	0.0000	0.0000	****	****
A19-2	6	0.0000	0.0000	****	****
A19-3	6	0.0000	0.0000	****	****
A19-4	6	0.0000	0.0000	****	****
A19-5	6	0.0000	0.0000	****	****
A20-1	6	0.0000	0.0000	****	****
A20-2	6	0.0000	0.0000	****	****
A20-3	6	0.0000	0.0000	****	****
A20-4	6	0.0000	0.0000	****	****
A20-5	6	0.5000	0.0000	1.0000	0.0000
A20-6	6	0.0000	0.0000	****	****
C05-1	6	0.4444	0.0000	1.0000	0.0000
C05-2	6	0.4444	0.0000	1.0000	0.0000
C05-3	6	0.0000	0.0000	****	****
C05-4	6	0.0000	0.0000	****	****
C05-5	6	0.0000	0.0000	****	****
C05-6	6	0.0000	0.0000	****	****
C05-7	6	0.0000	0.0000	****	****
C05-8	6	0.2778	0.0000	1.0000	0.0000
C06-1	6	0.4444	0.0000	1.0000	0.0000
C06-2	6	0.0000	0.0000	****	****
C06-3	6	0.0000	0.0000	****	****
C06-4	6	0.4444	0.0000	1.0000	0.0000
C06-5	6	0.0000	0.0000	****	****
C06-6	6	0.0000	0.0000	****	****
C06-7	6	0.4444	0.0000	1.0000	0.0000
P17-1	6	0.2778	0.0000	1.0000	0.0000
P17-2	6	0.2778	0.0000	1.0000	0.0000
P17-3	6	0.0000	0.0000	****	****
P17-4	6	0.2778	0.0000	1.0000	0.0000
P17-5	6	0.0000	0.0000	****	****
P17-6	6	0.4444	0.0000	1.0000	0.0000
P17-7	6	0.0000	0.0000	****	****
P17-8	6	0.2778	0.0000	1.0000	0.0000
P17-9	6	0.2778	0.0000	1.0000	0.0000

```
 P17-10            6    0.0000    0.0000      ****      ****

Mean               6    0.1064    0.0000    1.0000    0.0000
St. Dev                 0.0308    0.0000
=============================================================

* Nm = estimate of gene flow from Gst or Gcs. E.g., Nm = 0.5(1 -
Gst)/Gst;
  See McDermott and McDonald, Ann. Rev. Phytopathol. 31:353-373 (1993).
The number of polymorphic loci is :     27
The percentage of polymorphic loci is : 28.72


*************************************************************************
************
   **
**
   **         Nei's Original Measures of Genetic Identity and Genetic
distance      **
   **                      [See Nei (1972) Am. Nat. 106:283-292)]
**
   **
**

*************************************************************************
************


=========================================================================
pop ID    1         2         3         4         5         6
=========================================================================
1           ****    0.9681    0.9043    0.9468    0.7979    0.8511
2         0.0324      ****    0.9149    0.9574    0.8085    0.8617
3         0.1006    0.0889      ****    0.9149    0.8298    0.8404
4         0.0547    0.0435    0.0889      ****    0.8085    0.8617
5         0.2258    0.2126    0.1866    0.2126      ****    0.8191
6         0.1613    0.1488    0.1738    0.1488    0.1995      ****
=========================================================================

Nei's genetic identity (above diagonal) and genetic distance (below
diagonal).



*************************************************************************
************
   **
**
   **         Dendrogram Based Nei's (1972) Genetic distance: Method =
UPGMA        **
   **            --Modified from NEIGHBOR procedure of PHYLIP Version
3.5          **
   **
**
```

```
**************************************************************************
************
```




```
* File Name: dgram1.plt
```

Between	And	Length
5	4	2.45997
4	3	3.26777
3	2	2.18872
2	1	0.83182
1	pop1	1.62176
1	pop2	1.62176
2	pop4	2.45359
3	pop3	4.64231
4	pop6	7.91008
5	pop5	10.37005

```
**************************************************************************
************
   **
**
   **         Nei's Unbiased Measures of Genetic Identity and Genetic
distance      **
   **                      [See Nei (1978)  Genetics 89:583-590]
**
   **
**

**************************************************************************
************

=============================================================
pop ID    1         2         3         4         5         6
```

1	****	0.9681	0.9043	0.9468	0.7979	0.8511
2	0.0324	****	0.9149	0.9574	0.8085	0.8617
3	0.1006	0.0889	****	0.9149	0.8298	0.8404
4	0.0547	0.0435	0.0889	****	0.8085	0.8617
5	0.2258	0.2126	0.1866	0.2126	****	0.8191
6	0.1613	0.1488	0.1738	0.1488	0.1995	****

Nei's genetic identity (above diagonal) and genetic distance (below diagonal).

```
*************************************************************************
************
   **
**
   **          Dendrogram Based Nei's (1978) Genetic distance: Method =
UPGMA         **
   **             --Modified from NEIGHBOR procedure of PHYLIP Version
3.5           **
   **
**

*************************************************************************
************
                                                    +--------pop1
                                                +---1
                                    +-----------2   +--------pop2
                                    !           !
                  +-----------------3           +------------pop4
                  !                 !
  +---------------4                 +----------------------pop3
  !               !
--5               +-------------------------------------------pop6
  !
  +-----------------------------------------------------------pop5
```

* File Name: dgram2.plt

Between	And	Length
5	4	2.45997
4	3	3.26777
3	2	2.18872
2	1	0.83182
1	pop1	1.62176
1	pop2	1.62176
2	pop4	2.45359
3	pop3	4.64231
4	pop6	7.91008
5	pop5	10.37005

APPENDIX C

ARLEQUIN INPUT FILE

Programme to assess population differentiation between 4 black wattle families

[Profile]

Title="Computation of an AMOVA of 4 black wattel families"
#This project contains a sample of 4 famlies:
NbSamples=4
#Haplotypic Data:
GenotypicData=0
#Each character is an Allele:
LocusSeparator=NONE
#No explicit haplotype constitution is given for the individuals
DataType=RFLP
#We need to compute the distance matrix
MissingData ='?'
CompDistMatrix=1

[Data]

[[HaplotypeDefinition]]

HaplListName="RAPD data of four families"

HaplList= {

1 1111110110100100000111111010111110110101111111111000111111111110000111101111000111111010001011 1
2 1111110000100000000111000010111110110100011111111000111111111110011110001110100111101100101111
3 1111110000100000100111000010001011110101111111111000111111111110001111101110100111110100101111

```
 4 1111110000100000100111000010111111110101011111111100011111111111000111000111000011111010010110
 5 1111110000100000000111000010111110100000011111111000111111111111011111100111011011111110010110
 6 1111110000100000100111010010111111010000011111111100011111111111000111010111011011111011001000
 7 1111110000100000000111000010111100100001111111111000111111111111000111010111010011111000110110
 8 1111110000100000100111000010010000100000111111111000111111111111000101100111110000011000101001
 9 1111110000100000100111000010010000100101111111111000111111111111000101100111110000011000101001
10 1111110100100011000111001110111110100010111111111001111111111111000111100111000011111110010011
11 1111110000100000100111000010111100101000011111111001111111111111000111000111110011111011001011
12 1111110000100000000111000010111110100001111111111001111111111111000111000111010011111110010101
13 1111110000110000010111000010111111100001111111111001111111111111000111101110100011101100101001
14 1111110000110010000111000010010110101000011111111000111111111111000111101110000011100100111011
15 1111110000110000010111000010111110100010111111111101111111111111000111001110110011111100100001
16 1111111000110010010111000111111110101001111111111100111111111111000111101110110011101100101011
17 1111110100100001000111000010010110100000111111111000111111111111000111101110000101100000101001
18 1111110001110000010111000111110111100001111111111001111111111111000111101110110011111100101011
19 1111110000100000010111000010010110100001111111111000111111111111000101000111010001111000100001
20 1111110000100000001111000010010001100000011111111000111111111111000101100111110000110010101001
21 1111110000110010001111000011111110101001011111111001111111111111000111101110110011111100111001
22 1111110000110000001111000011111110101001111111111001111111111111000111101110110011111010010100 1
23 1111110000110000001111000011111110101000011111111001111111111111000111101110110011111100101011
24 1111110000110010000111001111111110101001111111111000111111111111010111010111011011111011001110 1
25 1111111000101000000111000110110110100100111111111000111111111111000111101110101011111110110111
26 1111110000100000000111000010000000100000011111111000111111111111000101100111010000110000110011 1
27 1111110100100101000111000111110110100100111111111000111111111111000111101110001011101100110111
28 1111110000100000000111000011110110100000111111111000111111111111000111010111010001111000111011 1
29 1111110000100000000111000011110110100000111111111000111111111111000111010111010001111100111010 1
30 1111110000100000000111000110000000100000011111111000111111111111000101000111010000110000110011 1
31 1111110000100000000111000010000000100000011111111000111111111111000111000111010000110000010011 1
32 1111110000100000100111000110100010100000101111111100011111111111000111001110000001110000100111
33 1111110000100000100111000010000000100000011111111000111111111111000101100111010000110000011011 1
34 1111110001100000100111000110111110100100111111111100111111111111000111010111011001111110011011 1
35 1111110000100000100111000011001110100001011111111000011111111111000101010111011001110000010000 1

}
        [[Samples]]

   SampleName= "Family 18"
```

```
SampleSize=5
sampleData=          {
   1   1
   2   1
   3   1
   4   1
   5   1
}

SampleName= "Family 99"
SampleSize=11
SampleData=          {
   6   1
   7   1
   8   1
      9   1
      10  1
      11  1
      12  1
      13  1
      14  1
      15  1
      16  1
}

SampleName= "Family 101"
SampleSize=13
SampleData=          {
   17  1
   18  1
   19  1
      20  1
      21  1
      22  1
      23  1
      24  1
      25  1
```

```
        26  1
        27  1
        28  1
        29  1
   }

  SampleName= "Family 101 x 77"
   SampleSize=6
   SampleData=          {
      30  1
      31  1
      32  1
         33  1
         34  1
         35  1

}

[[Structure]]
    StructureName= "A single group of 4 samples"
    NbGroups = 1
    Group=               {
        "Family 18"
        "Family 99"
        "Family 101"
        "Family 101 x 77"
    }
```

ARLEQUIN RESULTS

Project information:

NbSamples = 4
DataType = RFLP
GenotypicData = 0

==========
Settings used for Calculations
==========

General settings:

Deletion Weight = 1
Transition Weight Weight = 1
Tranversion Weight Weight = 1
Epsilon Value = 1e-07
Significant digits for output = 5
Use original haplotype definition
Alllowed level of missing data = 0.05

Active Tasks:

Standard indices:

Molecular Diversity:

Molecular Distance :Pairwise difference
GammaA Value = 0
Theta estimators :

Analysis of Molecular Variance:

No. of Permutations = 1000

Population pairwise Fst values:

Compute pairwise differences
No. of permutations for significance = 100
No. of permutations for Mantel test = 1000

Distance matrix:
Compute distance matrix
Molecular distance : Pairwise difference
Gamma a value = 0

Exact test of population differentiation:

No. of steps in Markov chain = 10000
No. of Dememorisation Steps = 1000

Required precision on Probability = 0
Print Histogram and table : Yes
Significance level = 0.05

==========
== GENETIC STRUCTURE ANALYSIS
==========

Number of usable loci for distance computation : 94
Allowed level of missing data : 0.05

List of usable loci :

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
91	92	93	94											

List of loci with too much missing data :

NONE

==========
AMOVA ANALYSIS
==========

Genetic structure to test :

No. of Groups = 1

[[Structure]]

StructureName = "A single group of 4 samples"
NbGroups = 1
IndividualLevel = 0
DistMatLabel = ""
Group={

"Family 18"
"Family 99"
"Family 101"
"Family 101 x 77"
}

Distance method: Pairwise difference

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	3	47.469	1.01681 Va	12.14
Within populations	31	228.131	7.35908 Vb	87.86
Total	34	275.600	8.37589	

Fixation Index FST : 0.12140

Significance tests (1023 permutations)

Va and FST : P(rand. value > obs. value) = 0.00098
P(rand. value = obs. value) = 0.00000
P(rand. value >= obs. value) = 0.00098+-0.00098

Histogram of variance components null distributions

	Va
-0.40027	12
-0.31349	49
-0.22670	127
-0.13992	160
-0.05313	175
0.03365	140
0.12043	132

0.20722	88
0.29400	52
0.38078	41
0.46757	30
0.55435	8
0.64114	4
0.72792	1
0.81470	0
0.90149	2
0.98827	1
1.07505	0
1.16184	1
1.24862	0

Observed values:

Va : 1.01681

==========

== Comparisons of pairs of population samples

==========

List of labels for population samples used below:

Label	Population name
1:	Family 18
2:	Family 99
3:	Family 101
4:	Family 101 x 77

Population pairwise FSTs

Distance method: Pairwise difference

	1	2	3	4
1	0.00000			
2	0.10436	0.00000		
3	0.18803	0.01796	0.00000	
4	0.29945	0.15654	0.14442	0.00000

FST P values

Number of permutations : 110

	1	2	3	4
1	*			
2	0.04505+-0.0152	*		
3	0.00000+-0.0000	0.14414+-0.0364	*	
4	0.02703+-0.0139	0.00000+-0.0000	0.05405+-0.0201	*

Matrix of significant Fst P values
Significance Level=0.0500

Number of permutations : 110

	1	2	3	4
1		+	+	+
2	+		-	+
3	+	-		-
4	+	+	-	

Population average pairwise differences

Above diagonal : Average number of pairwise differences between populations (PiXY)
Diagonal elements : Average number of pairwise differences within population (PiX)
Below diagonal : Corrected average pairwise difference (PiXY-(PiX+PiY)/2)

Distance method: Pairwise difference

	1	2	3	4
1	12.80000	16.01818	18.01538	17.66667
2	1.87273	15.49091	15.95804	16.54545
3	3.69231	0.28951	15.84615	16.58974
4	5.26667	2.80000	2.66667	12.00000

PXY P value

	1	2	3
2	0.00000		
3	0.00000	0.00000	
4	0.00000	0.00000	0.00000

Corrected PXY P value

	1	2	3
2	0.00000		
3	0.00000	0.00000	
4	0.00000	0.00000	0.00000

==========
== Exact Test of Sample Differentiation Based on Haplotype Frequencies
==========
Reference: *Raymond M. and F. Rousset. 1995 .*
Goudet, J., M. Raymond, T. de Meeüs and F. Rousset, 1996.

List of labels for population samples used below:

Label	Population name
1	Family 18
2	Family 99
3	Family 101
4	Family 101 x 77

Global test of differentiation among sample:

Non-differentiation: Exact P value = 1.00000 +- 0.00000 (6000 Markov steps done)

Differentiation test between all pairs of samples:

Markov chain length : 10000 steps)

Non-differentiation exact P values :

	1	2	3
2	1.00000+-0.0000		
3	1.00000+-0.0000	1.00000+-0.0000	
4	1.00000+-0.0000	1.00000+-0.0000	1.00000+-0.0000

Table of significant differences (significance level=0.0500):

	1	2	3	4

1		-	-	-
2	-		-	-
3	-	-		-
4	-	-	-	

Histogram of the number of significant different populations (significance level=0.0500):

	1		2		3		4
	0		0		0		0

==========
== ANALYSES AT THE INTRA-POPULATION LEVEL
==========

==========
== Sample : Family 18
==========

==========
== Standard diversity indices : (Family 18)
==========

Reference: *Nei, M., 1987.*
No. of gene copies : 5
No. haplotypes : 5
No. of loci : 94
No. of usable loci : 94 loci with less than 5.00 % missing data
No. of polymorphic sites : 27
Gene diversity : 1.0000 +/- 0.1265

(Standard deviation is for the sampling process)

==========
== Molecular diversity indices : (Family 18)
==========

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

Sample size : 5.0000
No. of haplotypes : 5

Allowed level of missing data : 5.0000 %
Number of observed indels : 0
Number of polymorphic sites : 27
Number of usable nucleotide sites : 94

Distance method : Pairwise difference
Mean number of pairwise differences : 12.800000 +/- 6.974238

Average gene diversity over loci : 0.136170 +/- 0.086736

(Standard deviations are for both sampling and stochastic processes)

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

==========
== Sample : Family 99
==========

==========
== Standard diversity indices : (Family 99)
==========

Reference: *Nei, M., 1987.*
No. of gene copies : 11
No. haplotypes : 11
No. of loci : 94
No. of usable loci : 94 loci with less than 5.00 % missing data
No. of polymorphic sites : 41
Gene diversity : 1.0000 +/- 0.0388

(Standard deviation is for the sampling process)

==========
== Molecular diversity indices : (Family 99)
==========

Reference: *Tajima, F., 1983.*
Tajima, F. 1993.
Nei, M., 1987.
Zouros, E., 1979.
Ewens, W.J. 1972.
Sample size : 11.0000
No. of haplotypes : 11

Allowed level of missing data : 5.0000 %

Number of observed indels : 0
Number of polymorphic sites : 41
Number of usable nucleotide sites : 94

Distance method : Pairwise difference
Mean number of pairwise differences : 15.490909 +/- 7.502689

Average gene diversity over loci : 0.164797 +/- 0.090044

(Standard deviations are for both sampling and stochastic processes)

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

==========
== Sample : Family 101
==========

==========
== Standard diversity indices : (Family 101)
==========

Reference: *Nei, M., 1987.*
No. of gene copies : 13
No. haplotypes : 13
No. of loci : 94
No. of usable loci : 94 loci with less than 5.00 % missing data
No. of polymorphic sites : 44
Gene diversity : 1.0000 +/- 0.0302

(Standard deviation is for the sampling process)

==========
== Molecular diversity indices : (Family 101)
==========

Reference: *Tajima, F., 1983.*
Tajima, F. 1993.
Nei, M., 1987.
Zouros, E., 1979.
Ewens, W.J. 1972.
Sample size : 13.0000
No. of haplotypes : 13

Allowed level of missing data : 5.0000 %
Number of observed indels : 0
Number of polymorphic sites : 44
Number of usable nucleotide sites : 94

Distance method : Pairwise difference
Mean number of pairwise differences : 15.846154 +/- 7.566288

Average gene diversity over loci : 0.168576 +/- 0.090479

(Standard deviations are for both sampling and stochastic processes)

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

==========
== Sample : Family 101 x 77
==========

==========
== Standard diversity indices : (Family 101 x 77)
==========

Reference: *Nei, M., 1987.*
No. of gene copies : 6
No. haplotypes : 6
No. of loci : 94
No. of usable loci : 94 loci with less than 5.00 % missing data
No. of polymorphic sites : 27
Gene diversity : 1.0000 +/- 0.0962

(Standard deviation is for the sampling process)

==========
== Molecular diversity indices : (Family 101 x 77)
==========

Reference: *Tajima, F., 1983.*
Tajima, F. 1993.
Nei, M., 1987.
Zouros, E., 1979.
Ewens, W.J. 1972.
Sample size : 6.0000
No. of haplotypes : 6

Allowed level of missing data : 5.0000 %
Number of observed indels : 0
Number of polymorphic sites : 27
Number of usable nucleotide sites : 94

Distance method : Pairwise difference
Mean number of pairwise differences : 12.000000 +/- 6.340347

Average gene diversity over loci : 0.127660 +/- 0.077885

(Standard deviations are for both sampling and stochastic processes)