Characterization of genetic diversity in selected cultivars and identification of a possible molecular marker for drought tolerance in tea *Camellia sinensis* (L.) O. Kuntze

---

By

M. Wium

Submitted in partial fulfilment of the degree:

MSc. Biochemistry

In the Faculty of Natural & Agricultural Science

University of Pretoria

Pretoria

June 2009

Declaration

I declare that the dissertation, which I hereby submit for the degree M.Sc. (Biochemistry) at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signature: ..... Date: 27 June 2009

Mariet Wium

Acknowledgements

Bedankings

I acknowledge with gratitude the valuable contributions the following people and institutes have made towards this study:

The **Tea Research Foundation of Central Africa** (TRFCA) for the collection of the morphological data, the tea leaves sample and the funding of the experimental work, in particular Dr. Nyirenda (tea breeder).

Professor Liang Chen (Tea Research Institute, Chinese Academy of Agricultural Sciences) for the donated DNA samples from 22 Chinese cultivars.

The **University of Pretoria** for educational grants.

Mrs. **Elna Cowley** (technical assistant, Department of Genetics) and Mrs. **Sandra van Wyngaardt** (technical assistant, Department of Biochemistry) for their assistance during the experimental part of the study.

The **Sequencing facility** of the University of Pretoria for the assistance during the experimental part of the study.

The **Department of Biochemistry**, University of Pretoria for granting me the opportunity to undertake this study.

My **fellow students** for their support and in particular Ezette du Rand and Christian Stutzer for the proof reading of this document.

Professor Zeno Apostolides, my supervisor for the last two years, for his leadership, input and time.

"The scientist is not the person who gives the right answers but the one asking the right questions" Claude Levi-Strauss

My family and friends:

Aan Ouma Marthie Wium (21/01/1921 - 15/03/2008)

Dankie… want elke persoon het te minste een iemand nodig wat glo jy kan niks verkeerd doen nie en jy is die beste…

Aan Pa en Ma (Manie en Cornelia Wium):

Dankie vir julle ondersteuning, liefde en geduld.

"Nine tenths of education is encouragement" Anatole France

Aan Nel-mari en Eduard Wium:

Sussies en boeties is baie spesiaal- jy mag met julle lelik wees en hulle mag nie kwaad bly nie…

Aan Ouma Nellie van Rensburg:

Elke oproep (selfs die 6:30 op 'n Sondag oggend) is waardeer

"The truest greatness lies in being kind, the truest wisdom in a happy mind"

Aan Andre en al my vriende:

"Friends are family we choose for ourselves"

Julle ondersteuning het my die jaar gedra.

Index

List of Abbreviations

°C	Degrees Celsius
°N	Degrees North
°S	Degrees South
5S rDNA	Nuclear-Encoded 5S Ribosomal RNA Genes
6PGDH	6-Phosphogluconate Dehydrogenase
A	Adenine
A%	Primer Association Rate
AAT	Aspartate Amino Transferase
AFLP	Amplified Fragment Length Polymorphisms
ALD	Aldolase
APX	Ascorbate Peroxidase
BC	Before Christ
bp	Base pairs
C	Cytosine
CAPS	Cleaved Amplified Polymorphism Sequences
cDNA	Complementary DNA
CHS	Chalcone Synthase
cm	Centimetre
cm^2	Centimetre square

Cu-Zn	Copper-Zinc
D	Nei similarity coefficient
D_A	Genetic distance
DFR	Dihydroflavonol 4-Reductase
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleoside triphosphate
DMSO	Dimethyl sulfoxide
ECG	(−)-Epicatechin-3-gallate
EGCG	(−)-Epigallocatechin-3-gallate
EGC	(−)-Epigallocatechin
EST	Expressed sequence tag
EST-SSR	Expressed Sequence Tag-Simple Sequence Repeats
EtBr	Ethidium Bromide
E-nose	Electronic nose
e_l	Mean error rate per allele
e_r	Mean error rate per locus
$\tilde{e}_l$	Overall mean error rate per allele
$\tilde{e}_r$	Overall mean error rate per locus
F1	First generation
F_{IS}	Inbreeding coefficient of the individuals in a subpopulation

F_{ST}	Variation among subgroups in relation to the total population
FUM	Fumerase
G	Guanine
G3PDH	Glyceraldehyde-3-Phosphate Dehydrogenase
G6PDH	Glucose-6-Phosphate Dehydrogenase
H_e	Expected Heterozygosity
H_o	Observed Heterozygosity
H_s	Expected Heterozygosity
H_T	Total Heterozygosity
IDH	Isocitrate Dehydrogenase
IPGRI	International Plant Genetic Resources Institute
ISSR	Inter Simple Sequence Repeats
J	Jaccards coefficient
kbp	Kilo base pair
kg/ha	Kilogram per hectare
L	Litre
MAS	Marker-assisted selection
ME	Malic Enzyme

min	Minutes
mM	Mili molar
mRNA	Mature Ribonucleic Acid
m_l	Number of single locus genotypes with at least one mismatch
m/s	Meter per second
n	Number of samples
ng	Nanograms
nm	Nanometer
ng/ml	Nanogram per millilitre
nt	Number of replicated single locus genotypes
PAL	Phenylalanine Ammonia-Lyase
PCR	Polymerase Chain Reaction
PCR-RFLP	Polymerase Chain Reaction-Restriction Fragment Length Polymorphism
PIC	Polymorphism Information Content
PGM	Phosphoglucomutase
PolyA	Poly Adenine tail
PPO	Polyphenol Oxidase
PO	Peroxidase
QTL	Quantitative Trait Loci

RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphisms
RNA	Ribonucleic Acid
RNase	Ribonuclease
S	Simple match coefficient
SCAR	Sequence Characterized Amplified Region
sec	Seconds
SKD	Shikimate Dehydrogenase
SOD	Superoxide Dismutase
SSR	Simple sequence repeats
STS-RFLP	Sequence Tagged Site-Restriction Fragment Length Polymorphism
T	Thymine
t	Number of repeats
TAE	Tris-acetate EDTA buffer
t_m	Multi-locus estimation
TRFCA	Tea Research Foundation of Central Africa
UPGMA	Unweighted Pair Group Method with Arithmetic Averages
UV	Ultra Violet
V	Volt

List of Tables

Chapter 1

Chapter 2

Chapter 3

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

CHAPTER 1

GENETIC STUDIES IN TEA (Camellia sinensis) AND BREEDING STRATEGIES

"Tea is a miraculous medicine for the maintenance of health. Tea has an extraordinary power to prolong life."

Eisai ("Father of Tea" in Japan)

Maintaining Health by Drinking Tea 1211

1.1. The Importance of Tea

Tea is made from the young tender shoots of *Camellia sinensis*. Billions of people consume this non-alcoholic beverage on a daily basis. More than 3 billion cups of tea are consumed every day. This adds up to more than 45.4 L (12 gallons) per person annually (Wolf *et al.,* 2007).

The growing interest in tea is mainly due to its medicinal properties. Tea has anticarcinogenic, anti-aging, anti-inflammatory, antibacterial and anti-fungal properties (Friedman, 2007; Yang *et al.,* 2006). The antioxidants and alkaloids (e.g. caffeine) in tea are the major contributors to these properties. Tea contains between 1 to 5% caffeine. The young tea leaves contain about 30% polyphenolic antioxidants known as catechins (Perva-Uzunalić *et al.,* 2006). The major catechins in green tea are (−)-epigallocatechin-3-gallate (EGCG), (−)-epigallocatechin (EGC), (−)-epicatechin-3-gallate (ECG) and (−)-epicatechin (EC), Figure 1.1. The minor catechins namely, catechin, gallocatechin, epigallocatechin digallate, epicatechin digallate, 3-O-methyl EC and EGC, catechin gallate and gallocatechin gallate are also found in green tea, but only contribute to a small portion of the total catechin content. In black tea, most of the catechins are converted to theaflavins and thearubigins in an enzymatic oxidation step formally called the fermentation step during production, Figure 1.1 (Shaheen *et al.,* 2006; Siddiqui *et al.,* 2004; Wheeler & Wheeler, 2004).

Tea is produced in 52 countries (Mondal *et al.,* 2004). The top 10 tea producing countries in 2005 were China, India, Sri Lanka, Kenya, Turkey, Indonesia, Vietnam, Japan, Argentina and Bangladesh, Figure 1.2 (Chen *et al.,* 2007a). Tea contributes

to both the economy and cultural traditions of these and other countries. In addition to its direct use, tea is also used in the pharmaceutical and cosmetic industries (Chen & Zhou, 2005e).

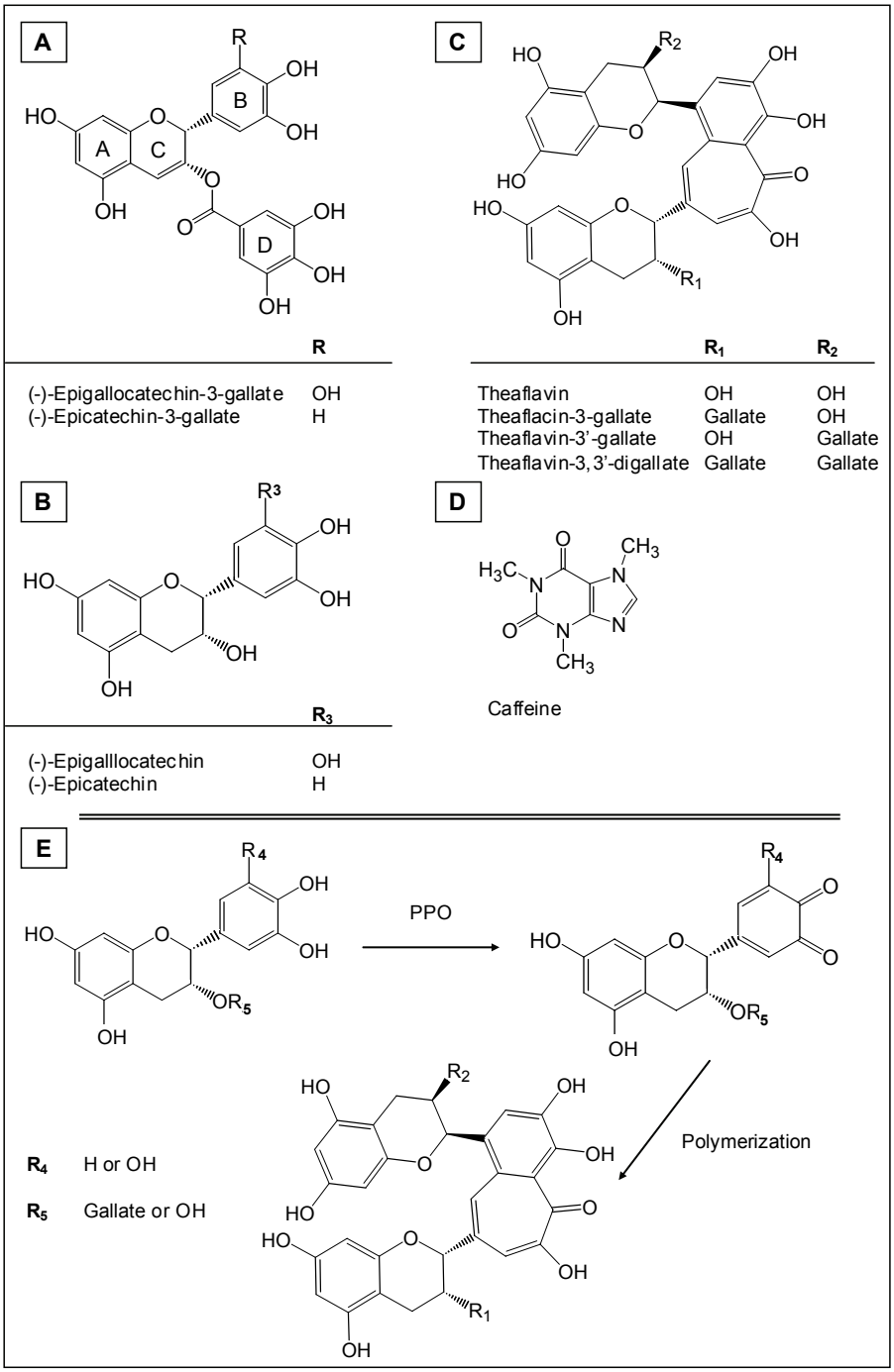


Figure 1.1: A) and B) The chemical structures of the major catechins in green, C) the theaflavins in black tea and D) caffeine. E) The enzymatic conversion of catechin to theaflavin during aerobic fermentation. Abbreviation: PPO: Polyphenol Oxidase (adapted from Chen *et al.*, 2007b and Friedman, 2007)

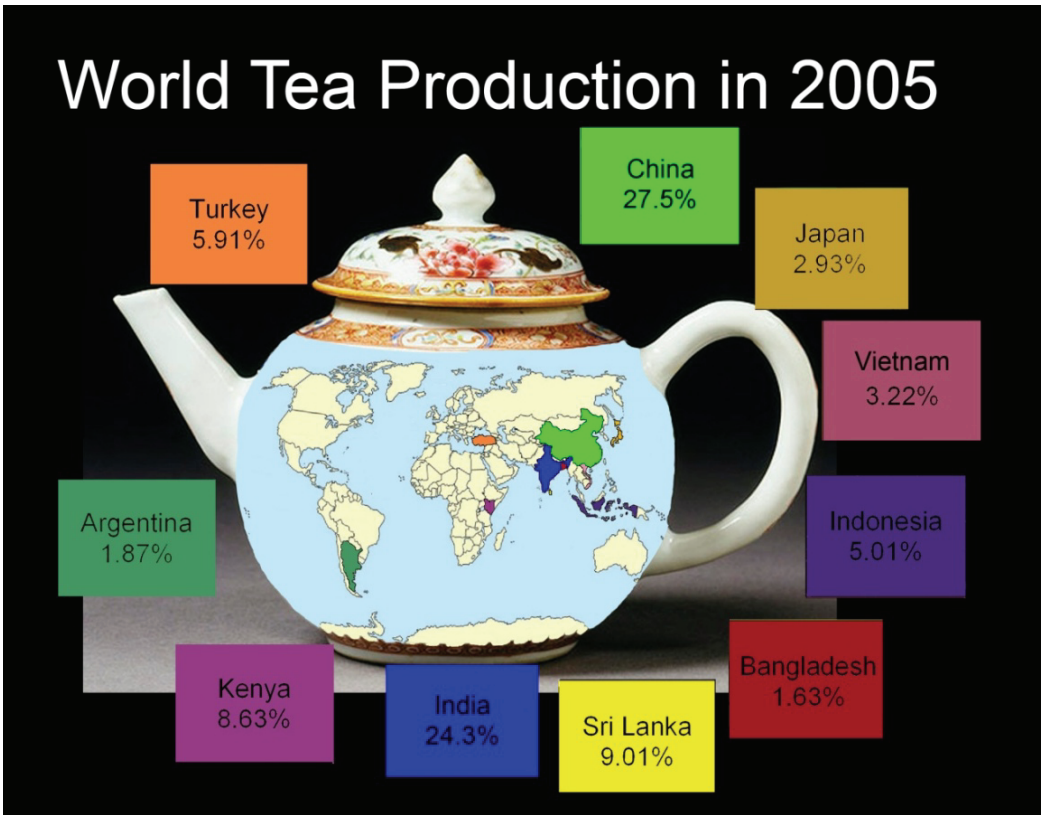


Figure 1.2: The world tea production in 2005 (Statistics from http://faostat.fao.org/site/567/default.aspx#ancor)

The origin of tea

Camellia sinensis is an evergreen tree that is indigenous to China and the Assam region in India. Tea is one of the oldest beverages in the world and was first consumed as a medicine. The legend of tea's origin dates back to 2700 B.C., when drinking boiling water was a custom in China. According to the legend, a tea leaf dropped into a pot of boiling water prepared for the Chinese emperor, Shen Nung. The emperor was enchanted by the taste, which started the custom of tea drinking (Wheeler & Wheeler, 2004; Wolf *et al.,* 2007).

This however is not the only legend about tea's origin. Bodhidharma was a Buddhist monk, who struggled to stay awake during meditation. In order to prevent him from

falling asleep he cut his eyelids off. This gained the favour of the gods and from his eyelids the first tea plant was born. By chewing only the leaves of the plant, Bodhidharma stayed awake for seven years. Today tea trees are still found around old Buddhist monasteries (Dan, 2006).

Tea classification

Camellia sinensis belong to the family *Theaceae* in the section *Thea* (Prince & Parks, 2001). *Camellia sinensis (L.) O. Kuntze* is botanically subdivided into two main subspecies. *Camellia sinensis* var. *sinensis* is a small leaved bush, indigenous to China, Figure 1.3. This subspecies is tolerant to a cold climate. The second subspecies is *assamica,* a large leaved, tall tree originally from the Assam region in India, Figure 1.3 (Kaundun & Matsumoto, 2003a). A*ssamica* prefers a semitropical climate and on average has a higher caffeine and catechin content than *sinensis.* Green tea is made mainly from *sinensis* and black tea from *assamica* (Astill *et al.,* 2001; Singh & Ahuja, 2006; Tanaka & Taniguchi., 2007). A third minor subspecies, *Camellia assamica* subspecies *lasiocalyx* (Cambod or Indochina tea) is an intermediate between *assamica* and *sinensis*. Since breeding programs have constantly crossed the different subspecies, it is nearly impossible to classify all tea cultivars into these three groups without doubt (Singh & Ahuja, 2006).

Tea cultivars each have a set of unique characteristics. The cultivar Benefuki, a hybrid tea between *sinensis* and *assamica* subspecies, is high in methylated epigallocatechin gallate (Kaundun & Matsumoto, 2003a). Luxi white tea, a Chinese cultivar, has a high polyphenol content (53.7%), while Anji white tea has a high amino acid content (6.5%). Guangdong tea is low in caffeine (0.14%) in contrast to wild tea from Yunnan that is known for its high caffeine content (6.96%) (Mondal *et al.,* 2004).

Figure 1.3: Young shoots of the two subspecies, sinensis (left) and *assamica* (right) (Dan, 2006)

The tea industry

In the industry, tea is classified into three main groups: green tea, black tea and oolong tea. The differences in these groups are in the processing methods. About 78% of the tea produced is black tea, 20% is green tea and 2% is oolong tea (Rehman *et al*., 2007).

Black tea is fermented tea. Fermentation is an oxidative process that relies on the presences of the enzymes, polyphenol oxidase (PPO; E.C. [Enzyme Commission] 1.14.18.1; monophenol mono-oxygenase (tyrosinase)) and peroxidase (PO; E.C. 1.10.3.2; o-diphenol: O_2 oxidoreductase), Figure 1.1 (Haslam, 2003; Chen *et al*., 2007b). After plucking, leaves are first rolled which breaks the cellular components and releases the enzymes. This promotes the oxidation processes (Bhattacharya *et al*., 2007). The fermentation reactions are time, temperature, pH, relative humidity and oxygen dependent (Muthumani & Kumar, 2007). In this process, catechins are oxidized to theaflavins and thearubigins. Theaflavins and thearubigins are responsible for the astringent taste of black tea and contributed to the quality of black tea.

Green tea is unfermented tea. The leaves are plucked and immediately steamed or pan-fried to inactivate PPO. Since about 90–95% of the polyphenols in black tea undergo enzymatic oxidation to theaflavins and thearubigins, green tea contains significantly more catechins than black tea (Chen *et al*., 2005a; Astill *et al*., 2001).

Oolong tea is semi fermented and is mostly produce in China and Taiwan (Kaundun & Matsumoto, 2003a). White tea is also made from *Camellia sinensis*. It is an unfermented tea, but it is only made from the first leaf and the bud giving it a lighter colour than green tea (Du *et al*., 2006). However, only a small amount of white tea is produced. Commercially teas are traded under company names, but due to the increasing demand for better quality teas some Japanese teas are sold under the cultivar's name (Kaundun & Matsumoto, 2003a). Other teas that are not made from *Camellia sinensis* include rooibos tea and honeybush tea (South African teas), as well as fruit and herbal teas that are sold under the respective plant names.

The Tea Plant

Growing wild a tea plant can reach heights of 15 metres (50 feet), however under cultivation shrubs are maintained at 60 to 100 cm for harvesting. Plantations start to flower after three to eight years and can be harvested for more than a 100 years (Fang *et al*., 2006; Kwong-Robbins, 2005, Chen *et al*., 2008). Tea is diploid with a chromosomal number of 30 (2n = 30), though some triploid cultivars have been reported (Devarumath *et al*., 2002). The genome size of tea is 4,000 million bases (Tanaka & Taniguchi, 2007) and the karyotype ranges from 1.28 μ to 3.44 μ (Mondal *et al*., 2004).

Tea Research Foundation of Central Africa (TRFCA)

The Tea Research Foundation of Central Africa (TRFCA) started a breeding program in 1956 with the aim of increasing the quality and quantity of tea produced in Africa (Ellis & Nyirenda, 1995). This breeding program succeeded to steadily increase the yield over the first 40 years. Recently however, this increase is stabilizing.

This breeding program started with a small group of selected parents and the introduction of new genetic material to the program was limited to superior offspring. It is therefore believed that expanding the gene pool by the introduction of new genetic material may lead to a continuing in the increase in yield. The genetically distant plants could be added to the gene pool which may prevent inbreeding.

Genetic diversity

Genetic relationship can be defined as the evolutionary history among the cultivars/species/taxons (Huson & Bryant, 2006). This relationship relies on the presence of a common ancestor. Numerous techniques have been applied to investigate the genetic relationships in tea. These techniques include morphological, biochemical, and molecular markers and will be discussed below.

1.2. Why Are Genetic Studies Important In Tea?

Conventional breeding relies on the natural diversity within a population for the improvement of quality and quantity. A large and diverse gene pool is important since it allows species to adapt to the changeable abiotic and biotic conditions.

Tea is expected to have a high level of diversity (Thomas *et al.,* 2006a; Chen *et al.,* 2005b) since:

1. Tea has an outbreeding nature, which favours continuous gene exchange with neighbours and in some cases with wild relatives (Katoh *et al.,* 2003).
2. Tea is cultivated in the latitudinal range from 45 °N to 34 °S, which include about 52 countries and a range of different environmental conditions. In order to survive these diverse environmental conditions a wide range of cultivars are available (Mondal *et al.,* 2004).
3. Different teas (green, black and oolong) are made from the same plant has led breeders to focus on the different characteristics, preserving diversity (Astill *et al.,* 2001).
4. To a lesser extent the existence of chromosomal duplications (Devarumath *et al.,* 2002; Wachira & Ng'etich, 1999).

The world's average yield of tea plantations increased from 783.6 kg/ha in 2000 to 987.4 kg/ha in 2006 (Yao *et al*., 2008). This is mainly due to the replacement of seedling populations with clonal populations, which can lead to a decrease in diversity. In Japan, more than 86% of the tea plantations are clonal Yabukita plantations. About 60% of the cultivars released from Kenya are derived from the cultivar '6/8'. Cultivars develop in Tocklai, India are selected mainly (75%) from Betjani, Cinnamara and Cambod stock (Yao *et al.,* 2008). The increase in clonal plantations and continuous crossbreeding with small selection of superior trees can lead to a reduction in the gene pool of tea. Furthermore, it is known that inbreeding and the loss of diversity reduces the reproducibility and the survival of outbreeding species in small laboratory populations. This leads to an increase in the risk of extinction. Although not all species show a decrease in fitness of all the

characteristics studied, reproduction is generally affected and can be seen as a warning sign (Frankham, 2005). The knowledge of the genetic diversity in the tea plant can assist tea breeders in selecting distant parental plants, preventing inbreeding. This will prevent the shrinking of the gene pool and further loss of desirable traits.

Genetic diversity in tea can be assigned using morphological markers (phenotype), biochemical markers (metabolites), allozyme markers (proteins) and/or molecular markers (genotype) (De Veinne, 2003). These markers and their application in breeding will be the focus of this section.

1.3. Morphogenesis, Biochemical and Allozyme Studies

Morphological characteristics of plants were initially used as markers to assign genetic diversity. The morphological characters of a plant (phenotype) are the product of both the environment and its genotype.

The determination of the diversity among tea by means of morphological data is done by evaluating different characteristics of the leaf, stem, flower, fruit and yields (Dan, 2006; Chen *et al.,* 2005a; Thomas *et al.,* 2006a), see Table 1.1.

The disadvantages of this method are:

1. Characteristics are influenced by environmental factors, such as, the availability of water or the soil type, implying that plants from different areas cannot be compared.
2. The development stage of the plant also influences the morphological traits.

3. Morphological characteristics have a low heritability.
4. Some characteristics are expressed only late in the plant's life, for example tea plants only flower at three to eight years of age and some cultivars are sterile (Chen *et al.*, 2008).
5. This approach is time consuming (Dan, 2006).

In the **biochemical** analysis of plants, the concentrations of different metabolites are determined and compared among the samples. In tea catechin, amino acid and chlorophyll content are compared to analyse diversity among plants (Magoma *et al.,* 2000; Magoma *et al.,* 2003; Saravanan *et al.,* 2005; Thomas *et al.,* 2006a).

In **allozyme** analysis, the total protein extract of the plants are separated by electrophoresis. The different isoforms of the enzymes are then scored to assign diversity. The following enzymes are reported to be used for allozyme analysis of tea: E.C. 1.1.1.49), 6-phosphogluconate dehydrogenase (6PGDH, E.C. 1.1.1.44), shikimate dehydrogenase (SKD, E.C. 1.1.1.25), malic enzyme (ME, E.C. aldolase (ALD, E.C. 4.1.2.13), fumerase (FUM, E.C. 4.2.1.2), isocitrate dehydrogenase (IDH, E.C. 1.1.1.42), glucose-6-phosphate dehydrogenase (G6PDH, 1.1.1.40), glyceraldehydes-3-phosphate dehydrogenase (G3PDH, E.C. 1.2.1.12), phosphoglucomutase (PGM, E.C. 5.4.2.2) and aspartate amino transferase (AAT, E.C. 2.6.1.1) (Chen *et al.,* 2005a). Allozymes reveal low levels of diversity in tea (Chen *et al.,* 2005a).

Table 1.1: An illustration of some of the characteristics used to assign diversity based on the morphology of tea plants (Dan, 2006; Chen *et al.,* 2005a)

Morphological characteristic	**Score**				
	1	**2**	**3**	**4**	**5**
Leave shape	 Ovate	 Oblong	 Elliptic	 Lancelate	Other
Leave pose	 Erect (<35°)	 Semi-erect (35° - 75°)	 Horizontal (76° -90°)	 Drooping (>90°)	
Leaf base shape	 Attenuate	 Rounded	 Blunt	Other	
Leaf colour	Light green	Green	Grayed-green	Grayed-yellow	Yellow-green
Serrula form	 Regularly acute	 Regularly blunt	 Irregularly acute	 Irregularly blunt	
Flower colour	White	Cream	White with red purple (pinkish) tinge	Purple (pink) to purple-violet	Other
The splitting of style	 Geniculation (free for greater part of their length)	 Ascending (free for about half their length)	 United for greater part of the length, the free part short, more or less horizontal		
Bud hairiness	Absent	Few	Mediate	A lot	Plenty

Morphological, biochemical and allozyme methods were compared to each other in a study by Chen *et al.,* (2005a). All the dendrograms showed similar results and it was concluded that the efficiency of these techniques to discriminate between cultivars were similar. Morphological and biochemical analysis are indirectly dependent on the level of gene expression and therefore will be directly influenced by the environment. In allozyme analysis, presence or absence of a particular isoform is detected rather than the level of a substance in the case of biochemical analysis (Chen *et al*., 2005a).

1.4. Molecular Genetic Studies

Molecular genetic markers compensate for some of the disadvantages in using morphological and biochemical markers. Molecular markers exploit variation on DNA level and are therefore unaffected by environmental changes. This provides a more direct and reliable measurement of diversity.

The ideal molecular marker should (De Vienne, 2003; Dan, 2006):

1. be highly polymorphic.
2. be co-dominant.
3. unambiguous.
4. occur frequently in the genome.
5. be evenly distributed throughout the genome.
6. be easily accessible and assays should be fast and easy.
7. be highly reproducible.
8. allow for easy exchange of data between laboratories.
9. be both low in cost and development.

With these aims in mind, the ideal marker is yet to be discovered. However, the current molecular markers combine some of these characteristics (Dan, 2006). Molecular markers can be divided into two groups, markers based on hybridization to DNA segments within the genome and markers based on the polymerase chain reaction (PCR).

1.5. Hybridization Based Markers

Hybridization based techniques rely on the ability of a probe (short labelled nucleotide sequence) to recognize a particular site on a DNA sequence. Only one marker, Restriction Fragment Length Polymorphisms (RFLP), use hybridization as sole detection technique (see Figure 1.4 and Table 1.2).

1.5.1. Restriction Fragment Length Polymorphism (RFLP)

RFLP was the first molecular technique developed for the assessment of genetic diversity. In this technique, genomic DNA is first digested with one or more restriction enzyme. Usually the recognition sites of these enzymes are four bases or more. The fragments are then separated on size followed by the annealing of a labelled probe to the digested DNA, Figure 1.5 (De Vienne, 2003). Figure 1.5 is a representation of RFLP analysis of five hypothetical cultivars, the variation of cultivar 2 and 3 from 1 indicates point mutations in the restriction enzyme recognition sites, while cultivar 4 and 5 is a deletion and an insertion, respectively. RFLPs are co-dominant and therefore it is possible to distinguish between a heterozygote and homozygote plant.

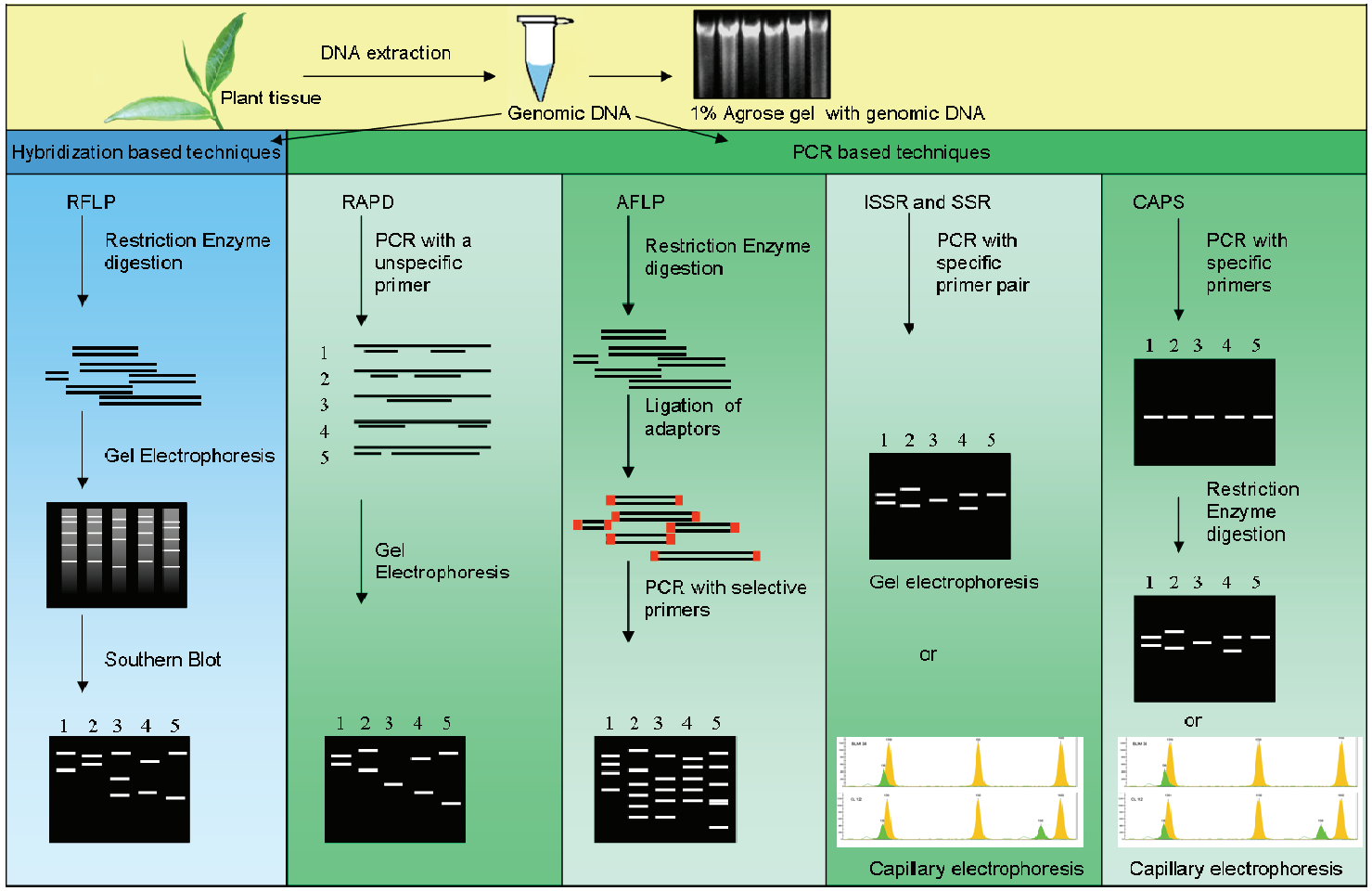


Figure 1.4: Summary of the molecular techniques discussed in the text. DNA is extracted from the leaf and run on 1% Agarose gel. The techniques are divided into two groups; hybridization based techniques (RFLP) and PCR based techniques (RAPD, AFLP, ISSR SSR and CAPS). The electrophoresis of technique shows the diff profiles of five sam RFLP: DNA is digeste restriction enzymes an southern blot is perfor RAPD: PCR is performed using unspecific primers. The fragments are analysed using gel electrophoresis. AFLP: DNA is digested with two restriction enzymes and adaptors are ligated to the fragments. PCR is performed with specific primers and analysed using electrophoresis. ISSR and SSR: PCR is performed with specific primers and analysed using electrophoresis. The schematic presentation also shows a profile of capillary electrophoresis of two cultivars. CAPS: DNA is amplified with specific primers. Fragments are digested with restriction enzymes and analysed with gel or capillary electrophoresis (Compile from Bussell *et al.*, 2005; Blears *et al.*, 1998; Dan, 2006; De Vienne, 2003; McGregor *et al.*, 2000; Singh & Ahuja, 2006)

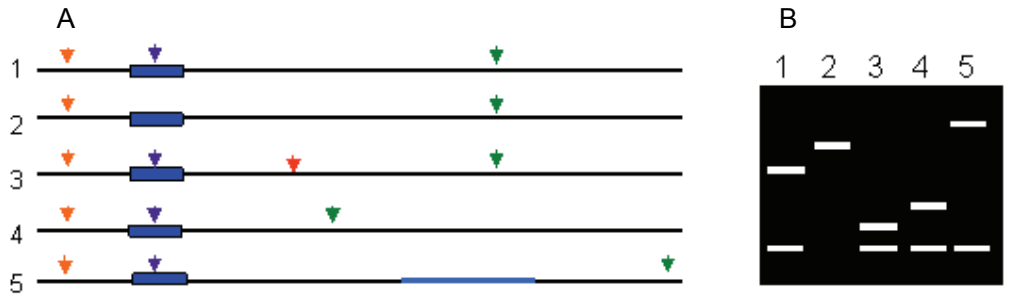


Figure 1.5: Schematic presentation of the RFLP analysis of five cultivars. A) The arrows indicate the restriction recognition sites of the enzyme, while the blue block is the portion of the sequence to which the probe will align. The variation of cultivars 2 and 3 from 1 is due to point mutations that affected the recognition sites of the restriction enzyme, cultivar 2 loses a site (purple arrow lost), while cultivar 3 gains a site (indicated with the red arrow). Variation in cultivar 4 is because of a deletion, while cultivar 5 has an insertion mutation indicated by the blue portion of the black line. B) The profile of these cultivars after electrophoresis (constructed from De Vienne, 2003; Singh & Ahuja, 2006)

The phenylalanine ammonia-lyase (PAL) gene has been used as a probe in the RFLP analysis of tea (Matsumoto *et al.,* 2002; Matsumoto *et al.,* 2004). PAL is part of the phenylpropanoid pathway (catechin producing pathway). Another probe has also been designed from the nuclear-encoded 5S ribosomal RNA genes (5S rDNA). In higher plants, this gene is usually organized into clusters of tandem repeats of which the size ranges between 200 and 900 bp (base pairs). These regions have a highly conserved coding region of 120 bp. However, a higher level of diversity is observed in the non-transcribed spacer. The spacers can differ not only in length, but also in sequence and the copy number of the repeated units, creating ample opportunity for variation. This probe has been used to discriminate the subspecies *sinensis* from *assamica* and *lasiocalyx* (Singh & Ahuja, 2006).

A disadvantage of RFLP is that it requires a large amount of relatively pure DNA limiting the amount of markers that can be analysed per plant. In addition, this technique is time consuming and difficult to automate. This resulted in a tendency

towards PCR based methods, since PCR techniques require low quantities of DNA per analysis (De Vienne, 2003).

1.6. Polymerase Chain Reaction (PCR) Base Methods

PCR is the *in vitro* amplification of DNA using an enzyme (DNA polymerase), primers (short nucleotide sequence) and free nucleotides. This reaction is divided into 3 basic temperature control steps that are repeated to achieve amplification (Figure 1.6), namely:

1. DNA template denaturation at 94 °C,
2. Primer annealing between 45 to 65 °C
3. Elongation of the new DNA strand (68 to 72 °C).

Firstly, the double stranded DNA is denatured at 94 °C to single strands. The second step involves a temperature drop to allow the primer to bind to the single stranded DNA template. The temperature of this step depends on the specific nucleotide content (specifically the GC content) and the length of the primers. The last elongation step is performed at 68 to 72 °C, depending on the optimum temperature of the specific DNA polymerase used. The reaction is usually started at 94 °C for 2 to 5 minutes, to ensure that the DNA is denatured completely. The final step of PCR is an extended elongation step (68 to 72 °C), to ensure that the amplification products are completed.

In PCR based genetic markers, fragments of the genomic or organelle DNA is amplified. The size of the fragments is then scored and analysed to assign genetic diversity. Compared to RFLPs, PCR base methods are faster and less dependent on the quality of the extracted DNA. In addition, PCR methods use less DNA making it

possible to apply many markers to a small quantity of DNA. PCR-based markers can be subdivided into arbitrary and specific markers. Arbitrary markers analyse the genome based on uncharacterized DNA fragments, while specific markers amplify known regions of the genome.

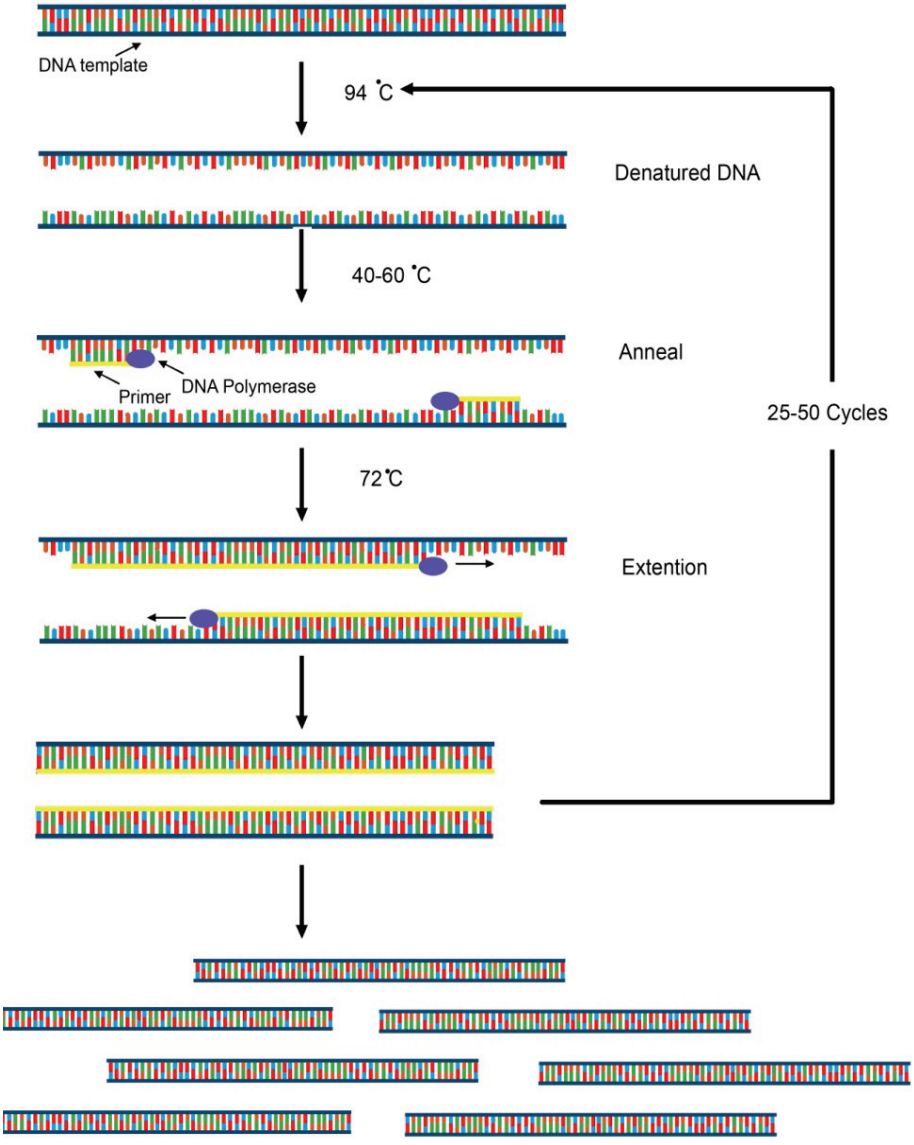


Figure 1.6: Schematic presentation of the principles of PCR. The first step is a denaturing step at 94 °C followed by 25-50 cycles of denature (94 °C), elongation at the temperature of the particular primer set and extension at 68-72 °C. See text for detail (Adapted from Semagn *et al*., 2006; Dan, 2006)

1.6.1. Arbitrary Markers

Arbitrary markers open genetic analysis to many species, since their use does not depend on the availability of sequence information (Bussell *et al.,* 2005). Primers in arbitrary markers amplify non-specific parts of the genome. Several types of markers can be classified as arbitrary and these include Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphisms (AFLP) and Inter Simple Sequence Repeats (ISSR) (see Figure 1.4 and Table 1.2).

1.6.1.1. Random Amplified Polymorphic DNA (RAPD)

Williams *et al*. developed the Random Amplified Polymorphic DNA method (RAPD) in 1990. This technique employs primers of 8 to 12 bases that anneal at multiple sites scattered randomly in the genomic DNA. The primer sequences must have a GC content higher than 50% (Bussell *et al.,* 2005). Fragments are analysed based on size and scored based on absence or presence of a particular fragment, Figure 1.4. The advantage of RAPDs is that it is fast, inexpensive, easy and does not require sequence information for development. RAPD analysis requires a small amount of DNA. However, the reliability and repeatability of this technique is problematic due to the low annealing temperature of the short primer (Blears *et al.,* 1998; Savelkoul *et al.,* 1999).

RAPDs have been successfully used to evaluate the genetic stability of micropropagated tea plants (Devarumath *et al.,* 2002), detect genetic diversity between tea (Chen & Yamaguchi, 2005c; Gul *et al.,* 2007; Wachira *et al.,* 2001), for parental analysis, to construct a genetic linkage map for tea (Hackett *et al.,* 2000), to identify possible QTL markers (Mishra & Sen-Mandi, 2004) and to investigate the relationship between *Camellia* species (Chen & Yamaguchi, 2002).

1.6.1.2. Amplified Fragment Length Polymorphism (AFLP)

Amplified Fragment Length Polymorphism (AFLP) combines the restriction site recognition elements of RFLPs with the added advantages of being PCR based markers. Genomic DNA is first digested with two restriction enzymes (recognition site 4 to 6 bp), Figure 1.7. The selection of these enzymes is crucial to the success of the analysis. The first enzyme is selected to have a high cutting frequency, while the second is a rare-cutting enzyme in the genome. Specific adaptors (one for each restriction enzyme overhang) are ligated to the fragments. Enzyme digestion of the genomic DNA and ligation of the adaptors to the fragments are carried out simultaneously. This step (step 1, Figure 1.7) is followed by PCR with a primer pair.

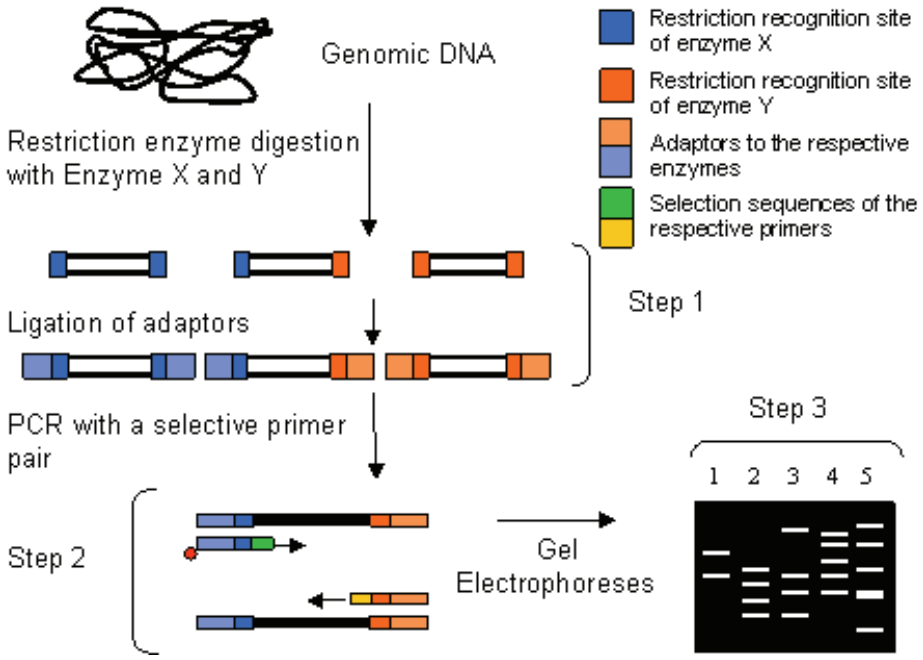


Figure 1.7: Schematic presentation of AFLP. Step 1 includes restriction enzyme digestion and ligation. Step 2 is the PCR reaction with a pair of selective primers. The red dot indicates a fluorescent label and the respective colours is explained in the key (top right). Step 3 is the gel electrophoresis of the respective fragments. 1 to 5 indicates hypothetical cultivars (Adapted from Blears *et al.,* 1998)

Each primer consists of a specific and a random region. The first is specific to the adaptor region and restriction enzyme recognition portion to which the adaptors were initially ligated to (about 15 base). The random portion is 1 to 4 bp in length (step 2, Figure 1.7). The specific portion of the primer stabilizes it, enhancing repeatability and reliability, while the random portion selects for a certain subset of loci. For example, if this random region consists of one bp, it will theoretically amplify 1 in every 4 of the fragmented DNA for each primer. Therefore 1 in every 16 (4x4) fragments will be amplified by the primer pair. If it consists of three bp, 1 in every 4096 (4^3x4^3) fragments will be amplified for the primer set. Furthermore, one of the primers in the pair is usually fluorescent labelled. Fragments are analysed with gel electrophoresis and bands are scored on the absence or presence of a particular fragment size (Blears *et al*., 1998; Bussell *et al*., 2005; Techen *et al*., 2004). AFLP is technically more demanding than RAPDs but its reproducibility and reliability is higher (Bussell *et al.,* 2005; Semagn *et al.,* 2006).

AFLP is a fast technique that requires no sequence information for development and has a high multiplex ratio where up to a 100 genetic loci may be analysed in one experiment. A drawback however, is that it cannot distinguish between a heterozygote and homozygote (dominant). Partially digested DNA can also cause problems in AFLPs (Nybom, 2004).

AFLPs have been used with success to discriminate the subspecies of tea plants (Paul *et al.,* 1997) and evaluate the genetic diversity among tea cultivars (Balasaravanan *et al.,* 2003; Wachira *et al.,* 2001).

1.6.1.3. Inter Simple Sequence Repeats (ISSR)

Inter Simple Sequence Repeats (ISSR) were develop from SSR (Simple Sequence Repeats, see Section 1.6.2.2) with the advantage that they need no sequence information (McGregor *et al.,* 2000). ISSR primers amplify parts of the genome known as microsatellites. Microsatellites are tandem repeats of 1 to 6 nucleotides. ISSR primers can be divided into two groups, unanchored and anchored primers (Figure 1.8).

Unanchored ISSR primers use the sequence of the repetitive region itself (15 to 30 bp) and successful analysis has been frequently reported. Anchored ISSR primers are usually 15 to 30 bp in length with an anchor region of 3 to 5 bases on the 3' or 5' end side. Anchoring prevents the primers of annealing throughout the simple sequence repeat, since repeatable units are used for example $(CA)_n$ (Bussell *et al.,* 2005).

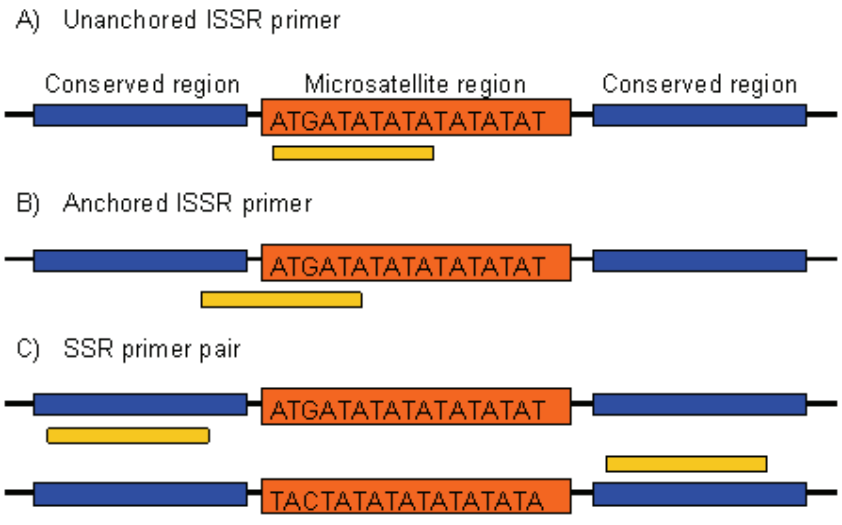


Figure 1.8: Primer design of ISSR and SSR primers. A) Unanchored ISSR primers are designed to match the repetitive region. B) In anchored ISSR primers the 5' end of the primer matches a few bases of the non-repeated sequence flanking the microsatellite region. C) SSR primers pair is designed specific to the conserved region of the adjacent microsatellite region (Adapted from Dan, 2006)

ISSR is dominant markers however, co-dominance can be achieved if the anchored region of the primer is extended. ISSR uses longer primers than RAPDs, therefore higher annealing temperatures are used, resulting in better reproducibility (Poulin *et al.,* 2005). It can be automated and is technically less demanding than AFLPs. These advantages make ISSR the superior technique compared to AFLPs and RAPDs.

In tea, ISSR has been applied to determine the mating system parameters for tea. A study confirmed that tea is an outbreeding species, based on the high multi-locus estimation (t_m) of 0.997±0.094 (Muoki *et al.,* 2007). ISSR has also been used to assess the genetic fidelity among micropropagated tea plants (Devarumath *et al.,* 2002), assigning genetic diversity (Mondal, 2002; Yao *et al.,* 2008) and to evaluate somaclonal variants (Thomas *et al.,* 2006b).

1.6.2. Sequence Specific Primers

Specific markers have an increased reproducibility and repeatability compared to arbitrary markers. Sequence information is required in the design of these markers. Specific markers included Cleaved Amplified Polymorphism Sequences (CAPS) and Simple Sequence Repeats (SSR) (see Figure 1.4 and Table 1.2),

1.6.2.1. Cleaved Amplified Polymorphism Sequence (CAPS)

Cleaved Amplified Polymorphism Sequence (CAPS) or Sequence Tagged Site-Restriction Fragment Length Polymorphism (STS-RFLP) or PCR-RFLP is a hybrid between RFLP and PCR techniques. In CAPS analysis, genomic DNA is amplified using specific primers where after the fragments are digested with restriction enzymes, Figure 1.9.

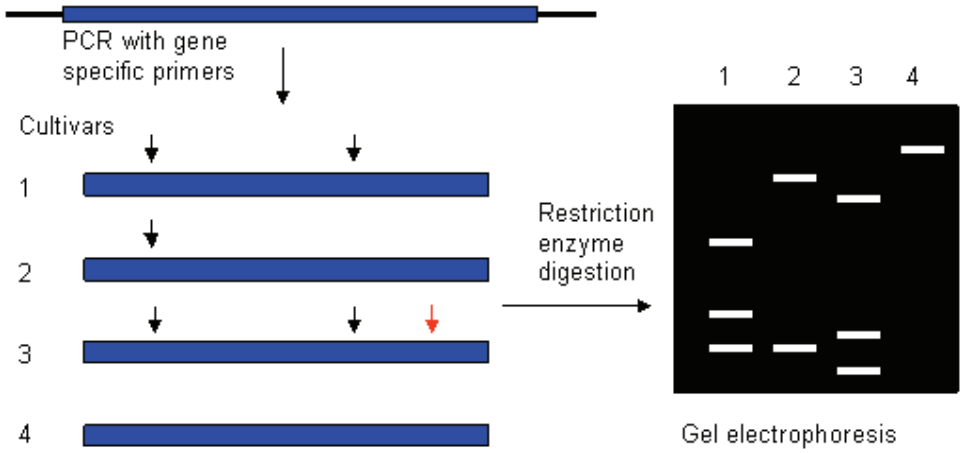


Figure 1.9: A schematic representation of CAPS analysis. First, a specific part of the DNA/ gene is amplified by PCR. The fragments are then digested by restriction enzymes. The arrows indicated the digested sites. The restriction profile is shown on the gel electrophoresis analysis, see text (constructed from Dan, 2006)

These restriction enzymes are known as secondary markers and reveal the differences between the tested individuals. The restriction enzyme usually used in this particular analysis has a 4 bp recognition site, since it is expected that there will be one recognition site every 256 (4^4) bases (Gostimsky *et al.,* 2005). In Figure 1.9, the differences between the cultivars are due to mutations that result in the loss (cultivar 2 and 4) or gain (cultivar 3, red arrow) of restriction sites. Since CAPS relies on the absence or presence of point mutations that adjust the restriction recognition sites, CAPS can therefore be used to verify the presence of a point mutation (Kaundun & Windass, 2006). CAPS is co-dominant and highly reliable and repeatable (Gostimsky *et al.,* 2005).

CAPS have been developed for tea based on the gene sequences of PAL, chalcone synthase (CHS) and dihydroflavonol 4-reductase (DFR). These three genes are involved in catechin and theaflavin synthesis and therefore, they contribute directly to

the taste and quality of a tea cultivar (Kaundun & Matsumoto, 2003a; Kaundun & Matsumoto, 2003b).

1.6.2.2. Simple Sequence Repeat (SSR)

Simple Sequence Repeats (SSRs), microsatellites or Sequence Tagged Microsatellite Sites (STMS) were first discovered by Tautz (1989) in *Drosophila*. Today SSRs are used for numerous genetic applications including genetic analysis (Xu *et al.,* 2004), gene mapping, qualitative trait analysis (Frewen *et al.,* 2000; Marques *et al.,* 2002), cultivar identification (Wünsch & Hormaza, 2002) and parental analysis (Yazdani *et al.,* 2003).

Microsatellites are regions of 1 to 6 bp tandem repeats. These repeats are spread randomly throughout the genome of most, if not all, eukaryotes and prokaryotes (Xu *et al.,* 2004). SSRs are located in both the coding (exon) and non-coding (intron) part of the genome.

Although microsatellite regions predominately occur in the non-coding regions of the genome, 7 to 10% are found within the coding regions (Chistiakov *et al.,* 2006). The non-coding part of the genome is more prone to mutations due to the absence of selection pressure. The overall content of SSRs in the genome is related to the genome size (Li & Yin, 2007). These regions have on average a higher mutation rate, ranging from 10^{-7} (1 in every 10,000,000 bp) to 10^{-3} mutations per locus per generation in eukaryotes (Buschiazzo *et al.,* 2006).

Table 1.2: A comparison of the properties of the molecular markers (Aydagn *et al.,* 2005; Dan, 2006; Bussell *et al.,* 2005; Techen *et al.,* 2004)

	Molecular marker					
	RFLP	**RAPD**	**AFLP**	**ISSR**	**SSR**	**CAPS**
Restriction enzyme used	Yes	No	Yes	No	No	Yes
PCR used	No	Yes	Yes	Yes	Yes	Yes
DNA quality	High molecular weight	Medium molecular weight	Constant quality in samples	Standard	Standard	Standard
Primer species-specific	-	No	No	No	Yes	Yes
Inheritance	Co-dominant	Dominant	Dominant	Dominant	Co-dominant	Co-dominant
Sequence information for Development	Yes	No	No	No	Yes	Yes
Developmental effort	High	Very low	Low	Low	High	High
Genotyping effort	High	Very low	Very low	Low	Low	Low
Reliability	High	Low	High	High	High	High
Accuracy	Very high	Very low	Medium	High	High	High
Possibility of automation	No	Yes	Yes	Yes	Yes	Yes
Cost after development	High	Low	Medium	Low	Low	Medium

Some SSR regions have distinct physiological functions. A telomere is a 'specialized SSR' with a 4-10 bp repetitive motif (Tóth *et al.,* 2000). This motif is repeated several hundreds or thousands of times ending in a 5' overhang at the end of the chromosome. Telomeres protect the chromosome from degradation (Chistiakov *et al.,* 2006). SSRs are known to be involved in the recombination of genes. SSRs can

also be involved in transcriptional activation when occurring in the promoter region of a gene. These regions may encode for polyglutamine (CAG_n) or polyproline (CCG_n) which may play a role in protein-protein interaction of transcription factors. It is proposed that microsatellites are generated by DNA polymerase slippage or unequal recombination (Tóth *et al.,* 2000; Bushiazzo *et al.,* 2006).

In microsatellite analysis, primer sequences are designed based on the conserved region flanking the tandem repeats, Figure 1.8. SSR are the most frequently used markers because of their high reproducibility, multiallelic nature, co-dominant inheritance, relative abundance and good genome coverage (Varshney *et al.,* 2005). This PCR technique uses low quantities of DNA, can be automated and is fast, reliable and repeatable. One of the major disadvantages of microsatellites is that they lack cross-species compatibility (Barbara *et al.,* 2007). SSRs are the most used molecular markers in current literature.

In the past development of SSR primers were expensive and time consuming. This process involved the construction of a genomic DNA library, screening for SSR clones and sequence analysis (Xu *et al.,* 2004). Figure 1.10 shows the schematic representation of the hybridization protocol. Genomic DNA is first digested into fragments and then adaptors are ligated to the fragments. Fragments with microsatellites are then selected using filter bound probes or biotinylated probes. These fragments are amplified, cloned and sequenced. The primers can then be design based on the conserved region that flanks the repeatable units (Zane *et al.,* 2002).

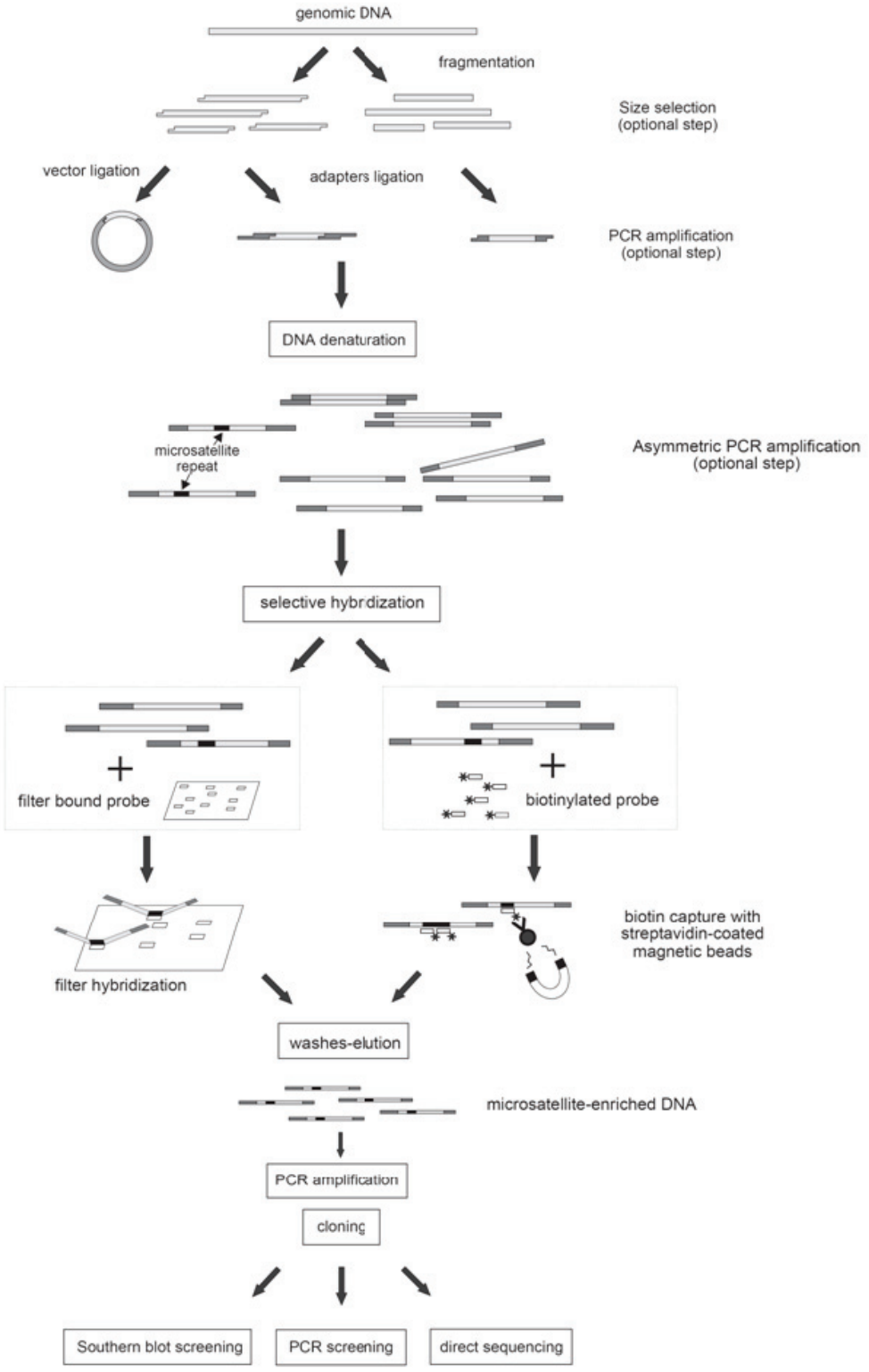


Figure 1.10: The schematic representation of selective hybridization protocols (Zane *et al.,* 2002)

However, Expressed Sequence Tag (EST) provided the necessary sequence information for the development of SSR primers as a by-product of gene discovery studies. In EST analysis, a cDNA library is first generated. The inserts is then amplified with PCR and sequenced (Figure 1.11). With the increasing number of EST projects for plants and the availability of the sequenced data online, the design of SSR primers has become less expensive (Varshney *et al.,* 2005). An additional advantage of EST-SSR generated markers is that they show a higher degree of cross-species compatibility and can be used directly in related species and even in related genera (Jiang *et al.,* 2006).

EST analysis of tea reported 1684 ESTs from a cDNA library of tender shoots (Chen *et al.,* 2005d) and 508 ESTs from a subtractive cDNA library (young shoots – mature leaves) (Park *et al.,* 2004). The genbank listed 2227 ESTs for C. *sinensis* in November 2007 this number increased to 3404 in November 2008 (http://www.ncbi.nlm.nih.gov/dbEST/). Jin *et al.* (2006) reported the data mining results from 1589 ESTs for EST-SSR (genic SSR primers) and identified 281 SSR regions in the 246 ESTs sequences. Nineteen primer pairs were designed of which three failed initial testing and only 10 showed polymorphisms (Jin *et al.,* 2006). In a more recent study, the same research group applied these 10 primers to 42 tea cultivars to prove that tea EST-SSRs are effective for the evaluation of the tea genome (Jin *et al.,* 2007). Zhao *et al.* (2007) develop 24 primers from the same set of EST sequences. These primers were tested in *Camellia sinensis* cultivars and for cross-species compatibility (Zhoa *et al.,* 2008).

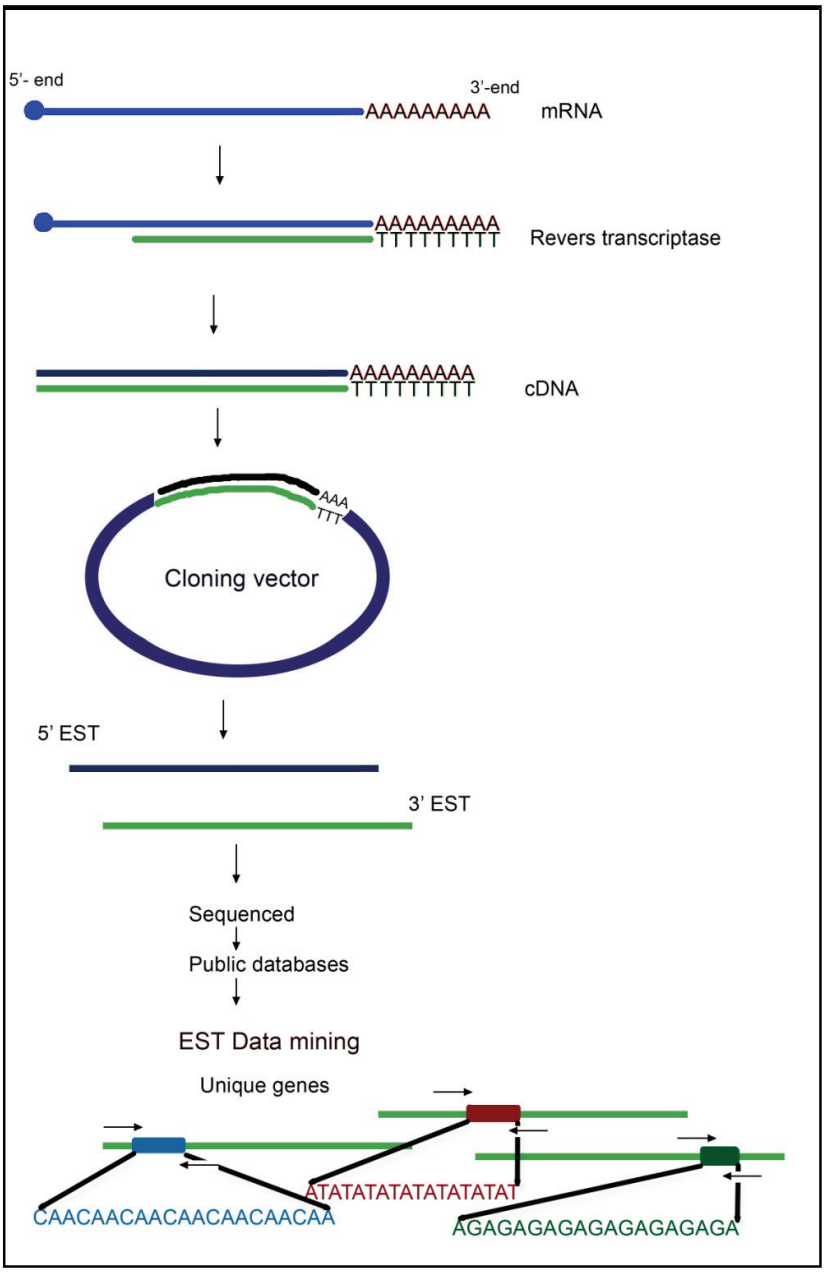


Figure 1.11: EST sequencing and data mining for SSR. mRNA (light blue line, polyA tail in red and the modified 5' guanine nucleotide cap is indicated with a circle) is isolated from tissue. A cDNA strand (green line) is synthesised using reverse transcriptase. Double stranded cDNA is produced using DNA polymerase one and RNAse. The cDNA are inserted into cloning vectors to produce a cDNA library. The inserts are sequenced from both ends with universal primers. The unique sequences are deposited into an EST-database. Unique SSR sequences are identified among the EST-sequences data and primers are designed for the flanking regions (primers indicated with horizontal arrows). Primers are then tested for polymorphism and applied in genotyping (Adapted from Bouck *et al.,* 2007)

The development of several SSR primer pairs (excluding the EST-SSRs) for tea has been reported (Kaundun and Matsumoto *et al.,* 2004; Freeman *et al.,* 2004; Hung *et al.,* 2008). In a study by Freeman *et al.* (2004), 15 SSR markers for tea were identified using a dinucleotide enriched library. This enriched library was constructed using a hybridization protocol similar to the one depicted in Figure 1.10. These primers were developed to discriminate between cultivars from different origins. SSR primers for PAL and DFR genes were designed based on the conserved flanking regions of the exons (Kaundun and Matsumoto, 2004). In the same study, the authors also tested SSR of genetically close species for cross-species compatibility. SSRs from *Camellia japonica* was tested and all of the four tested SSRs primers were compatible with *C. sinensis* (*C. japonica* is a flowering plant that belongs to the same subgenus, *Thea*, as tea). Hung *et al.* (2008) report the development of 11 SSR primer pairs for tea from Taiwan. Four of the SSR primers developed by Freeman *et a.,* (2004) and two from Kaundun and Matsumoto (2004) were applied to discriminate between 69 tea plants from Vietnam (Dan, 2006). Ohsako *et al.* (2008) used six of the primers from Freeman *et al.* (2004) to analyse the variability among Japanese tea.

1.6.3. Scoring and Evaluation of Diversity

The determination of the genetic relationship and genetic interpretation relies on characteristics that are scorable, variable, independent from each other, heritable and homologous. A homologous characteristic is one that is shared by a group within the population, this can be due to the sharing of an ancestor or/and a common trait. The methods of analysis depend not only on the type of marker, but also on the end results. Dominant and co-dominant markers are treated different, however when these markers are compared the same analysis method is used.

In order to quantify the variation between different samples, two approaches can be followed. In the conventional band pattern approach, scoring is done based on either the presence (1) or absence (0) of a particular size band. The genetic similarities can be calculated with the following three formulas (Bonin *et al.,* 2007):

- Jaccards coefficient J = a / (a+b+c)
- Dice coefficient D = 2a / (2a+b+c)
- The simple match coefficient S = (a+d) / (a+b+c+d)

Table 1.3: Scoring of bands in dominant markers

		Cultivar A	
		Band present (1)	Band absent (0)
Cultivar B	Band present (1)	a (1,1)	b (1,0)
	Band absent (0)	c (0,1)	d (0,0)

The key for the abbreviations in the formulas are indicated in Table 1.3. The Jaccards (1908) method only takes in account the bands present in one or both of the individuals. This method does not consider the absence of a band from both individuals to be relevant, thus, the absence of the same band is due to different mutations (Bonin *et al.,* 2007). The dice coefficient (Nei's similarity coefficient), also only consider bands present in at least one individual. However it gives more weight to bands present in both cultivars (Nei & Li*,* 1979). Therefore, this method highlights the similarities between the individuals rather than their differences. The simple match coefficient considers the bands present and absent to have the same value in the analysis therefore presuming that the same mutation has caused the bands to be absent. J, D and S can be any value between 0 and 1. The zero denotes no

similarities between the individuals, while 1 indicates that the individuals are identical (Bussell *et al.,* 2005; Bonin *et al.,* 2007). The output of these equations is a similarity or dissimilarity matrix that compares each sample to all the other samples.

The second approach is the allele frequency approach that can be done in a number of ways. The square-root procedure uses the inbreeding coefficient together with the frequency of null homozygotes of absent bands. However, this leads to a false prediction if the null allele frequency is low. A second procedure is the Lynch and Milligan procedure that is an improvement of the square-root procedure. This procedure estimates the null allele frequencies more accurately, although it has a restriction. The frequency of the absent band in this procedure must be higher than 3/N (N is the number of samples). The Bayesian procedure is an approach that is routinely employed (Bonin *et al.,* 2007). It uses posterior probabilities to estimate the most likely genetic relationship between samples (Huelsenbeck *et al.,* 2001). These parameters are applied to calculate the relationship between cultivars and to illustrate the genetic relationship in a tree (dendogram) form.

Genetic trees consist of leaves, branches and nodes that either can be rooted or unrooted, Figure 1.12. The leaves represent the different cultivars/species, while the length of the branches indicated the evolutionary distance between the data points (Huson & Bryant, 2006). The nodes indicate the last common ancestor.

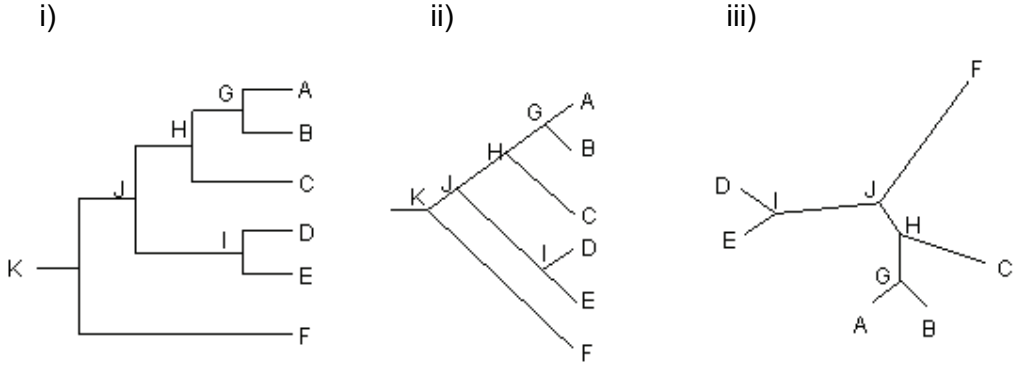


Figure 1.12: Tree types. i) and ii) are the same rooted tree drawn in different styles. iii) is an unrooted tree. A, B, C, D, E and F represent the leaves that are the samples analysed. G, H, I, J and K are internodes with K being the root node. The branches between the internodes and the leaves indicated evolutionary distance (Vandamme, 2003)

Trees are usually drawn using algorithmic methods such as Unweighted Pair-Group Method with Arithmetic mean (UPGMA) or Neighbor Joining (NJ). UPGMA is a clustering method. In this method, the two samples that have the closest distance between them are linked. The branch between these two samples is defined as half the distance between them. These two samples are now seen as a 'cluster', reducing the number of samples in the distance matrix by one. The distance matrix is recalculated and the closest related samples are again linked into a cluster. The process continues to reduce the samples one by one until only one entry remains. The tree is constructed from the root with the first nodes being the last two clusters. UPGMA assumes that all samples are equal distances from the root (ultrametric) and additive (meaning that the distance between any two nodes equals the sum of the length of all the branches between them). These assumptions are not always true (Hall, 2008).

NJ also reduces the number of entries with a stepwise approach although this method does not use clusters. The distances of the internodes in NJ are directly calculated. The original matrix is used to calculate the net diversity, which is the sum

of all the distances. These values are then applied to calculate the corrected distance matrix. The distance between the two samples with the closest corrected distances are then calculated. These distances between these two samples are not necessarily identical as with the UPGMA approach. The matrix is now recalculated and a new node is substituted from the previously linked samples. NJ does not assume that the samples are additive or that they are equal distances from the root (Hall, 2008; Vandalle *et al.,* 2003).

Diversity analysis of tea has previously be done using the UPGMA approach with the Jaccards coefficient (Mishra & Sen-Mandi, 2004; Young-Goo *et al.,* 2002; Chen *et al.,* 2005b) or with the Nei similarity coefficient (Chen & Yamaguchi, 2002; Thomas *et al.,* 2006b; Balasaravanan *et al.,* 2003; Wachira *et al.,* 2001).

1.7. Tea Breeding Strategy

Tea breeding like all tree breeding programs is a time consuming process. Breeding is complicated with self-incompatible, high inbreeding depression and long life cycle tea (Mondal, 2001). Tea plants only flower after three to eight years and are outbreeding. Tea breeding strategies include conventional, Marker-assisted selection and transgenic tea breeding.

1.7.1. Conventional Tea Breeding

In conventional breeding, it will take approximately 15 to 20 years to release new cultivars (Figure 1.13A). The conventional breeding program of the TRFCA starts with hand pollination of flowers (5,000 flowers) shown in Figure 1.13A. The seeds from these hand-pollinated flowers are harvested (3,750 seed) and planted.

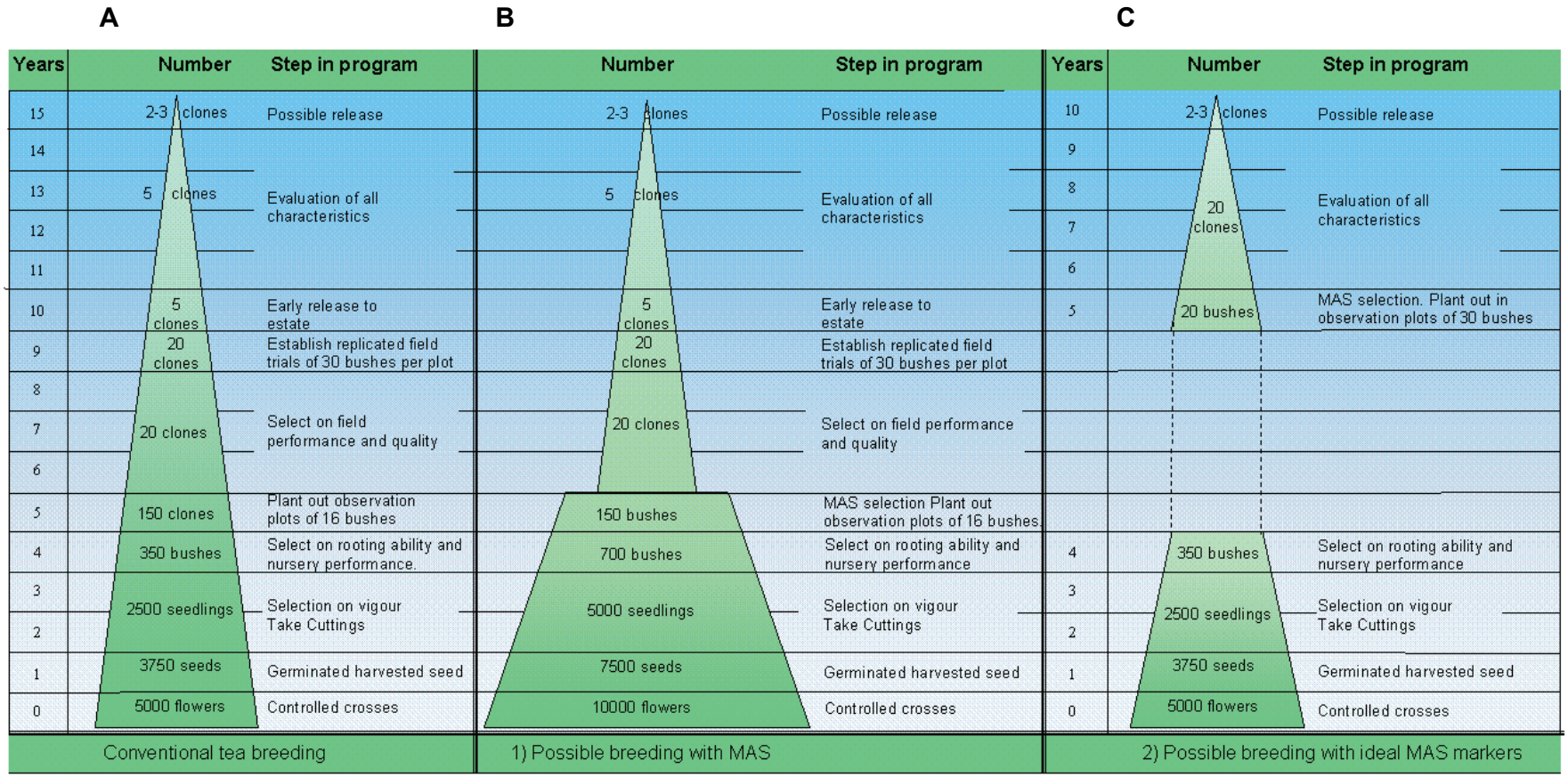


Figure 1.13: The conventional breeding and selection program of the Tea Research Foundation of Central Africa (TRFCA) and the possible breeding with MAS. A) The conventional breeding program of the TRFCA. B) Possible MAS assisted breeding program, in this program the field time of trials are not reduce. C) Possible MAS assisted breeding with ideal markers (constructed from Francia *et al.,* 2005; Pijut *et al.,* 2007; Wright, 2002)

Seedlings (2,000) are selected on vigour and cuttings are taken. These bushes are then selected on rooting ability (350). The best rooters (150) are planted into observation plots of 16 bushes each for a field trial. Plants are next selected on field performance and quality. The 20 selected clones are planted into 30 bush observation plots. The best five clones are released to estates for field evaluation under different environmental conditions. All the characteristics of these five clones are evaluated over the next four to five years and two or three of them are released as new improved cultivars (Figure 1.13A).

1.7.2. Marker-Assisted Selection (MAS) Breeding

Marker-assisted selection (MAS) is the selection of superior plants using DNA markers. This approach relies on the concept that a particular allele of a gene is linked to a certain phenotypical trait. Therefore, the phenotypic characteristic can be predicted very early in the breeding program by the presences of the allele. An additional advantage of MAS is that the evaluation of the phenotype is based on the genotype that implies that it is independent of environmental conditions. The regions on the genome where these alleles are located are referred to as quantitative trait loci (QTLs). Agronomical important traits, for example yield and quality, can usually be linked to more than one gene/ QTL.

QTL can be identified through genetic analysis. During genetic analysis, a particular size band of a marker can sometimes be associated with morphological traits by coincidence. In a RAPD study by Mishra and Sen-Mandi (2004), a band associated to drought tolerance in tea was identified (at 1400 bp with OPAH02 primer). The

particular band was only found in cultivars that were identified as drought tolerant. These cultivars were identified based on pot trials and the high activities of Cu-Zn cytosolic superoxide dismutase (Cu-Zn SOD) and ascorbate peroxidase (APX II). High activity of these enzymes is known to be associated with drought tolerance in plants. Mishra and Sen-Mandi (2004) concluded that it could be used to screen for drought tolerance in tea. RAPD markers that associate with morphological traits could be developed into sequence characterized amplified region (SCAR) markers. This involves the sequencing of the specific band followed by the design of a primer set based on the sequence. The designed primer pair is then tested. After which SCAR markers can be used to identify morphological traits. These markers have higher reproducibility and repeatability than RAPD makers due to the longer primers (Myburg *et al.,* 1998).

RAPD markers related with theanine content, date of bud sprouting, resistance to anthracnose and tolerance to cold were mapped on a linkage map (as referenced by Chen *et al.,* 2007a). QTLs have been identified for leaf angle, leaf colour and two diseases: anthracnose (caused by *Colletotrichum lindemuthianum*) and mulberry scale (caused by *Pseudaulacaspis pentagona*). However, only the QTL for mulberry scale, MSR-1 (a single dominate gene) has been developed into a MAS marker (as referenced by Tanaka & Taniguchi, 2007). Mulberry scale is a scale polyphagous insect that causes major damage to tea trees in Japan. Infected plantations should be treated with insecticide 2-5 days after the hatching peak. The larva develops a wax layer that serves as protection in the later stages of the life cycle. In addition, the insects live inside the crown of the tree giving it protection against insecticides. For these reasons, insecticides are ineffective. The development of MAS markers to

detect resistance against Mulberry scale insects can play a crucial role in tea breeding (Kaneko *et al.,* 2006; Takeda*,* 2004; Tanaka & Taniguchi*,* 2007).

If **ideal** MAS markers could be developed for tea, the development time to release new improved cultivars may be reduced as shown in Figure 1.13C. Field trials in conventional breeding take approximately 10 years in contrast to MAS breeding that will only take approximately five years. MAS could reduce the number of cultivars that should be subjected to field trials. Therefore the number of flowers pollinated and/or seeds germinated can be increased; leading to an increase in breeding efficiency. A disadvantage of MAS breeding is that marker development is time consuming and difficult, since more than one gene can be responsible for a phenotype. If only a small number of markers are available, the practical application of MAS will be expensive.

A small number of markers or markers that are not 100% effective could used in an intermediate between conventional breeding and ideal MAS breeding, Figure 1.13B. In this approach, the time of the field trials will not be reduced. However, the input will be increased and subsequently reduce with MAS markers before the field trials. This will allow for adequate field trials and the increase in input will improve the likelihood to find superior cultivars. Conventional breeding results in the release of the best two cultivars out of 5,000 flowers. This intermediate MAS approach could potentially release the best two cultivars out of 10,000 flowers, increasing the quality of released cultivars rather than the quantity (since the quantity will require an increase in field space).

1.7.3. Genetic Engineering of Transgenic Tea Breeding

Genetic engineering is the manipulation of genes and/or the introduction of foreign DNA into the genome. In transgenic plants foreign genes have been introduced into their genome of the plants. Genes from other plant species with desired traits can be introduced into the genome of the tea plant, thereby improving the performance of the plant. Both MAS and conventional breeding rely on the selection of desired traits from the existing gene pool, therefore limiting the possibilities. Transgenic breeding however, can increase the possibility of success by selecting genes from other plant species and thereby increasing the efficiency of the program.

Agrobacterium mediated gene transformation is the most widely used method for gene transfection in plants. This method is inexpensive, simple, allows little rearrangement of transgenes and is effective for transfection in plants (Sandal *et al.,* 2007). *Agrobacterium tumefaciens* belongs to the family *Rhizobiaceae* and is a soil-born bacterial pathogen that infects wounds in dicot plants. This bacterium cause gall crown disease. However, with careful engineering Agrobacterium can be used to transfect cells with genes of interest (Kumar *et al.,* 2004).

The leaves of tea plants are high in polyphenols. Polyphenols are antibacterial. This implies that Agrobacterium-mediated transfection in tea leaves are problematic. Mondal *et al.* (2001) reported the first generation of transgenic tea using Agrobacterium transformation from the somatic embryos. This cannot be applied to commercial cultivars, since somatic embryos are not available. Therefore, this cannot be applied to further improve a high quality cultivar. To solve this problem, L-glutamine was added to the leaf-culturing medium. L-glutamine allows for

Agrobacterium-mediated transfection and additionally allows the cells to retain the ability for shoot regeneration (Sandal *et al.,* 2007).

The possible improvement of tea by genetic engineering can result in the improvement of existing tea cultivars. An example of such an improvement of tea is the decaffeination of tea as reviewed by Yadav and Ahuja (2007). They proposed three possible solutions:

1. The overexpression of caffeine degrading enzymes. However the degradation pathway of caffeine in tea is not fully elucidated as yet (Ashihara *et al.,* 2008; Yadav & Ahuja, 2007).
2. The use of antisense technology for degradation/ blockage of the genes in the caffeine synthesis pathway. The antisense mRNA of the gene that needs to be silence is inserted into the genome. Antisense mRNA will be produced and will then anneal to the sense RNA. Double stranded RNA prevents transcription and is degraded. However, this approach has not been report yet in tea.
3. RNA interference for degradation of genes in the caffeine synthesis pathway. Using RNAi, Ogita *et al.* (2003) could reduce caffeine by up to 70% in coffee (*Coffea canephora*). A similar approach can be followed in tea, though it remains to be demonstrated.

Tea breeding can also benefit if resistance to biotic stress and tolerance to abiotic stress can be increased. The introduction and silencing of genes in tea may result in the development of ideal cultivars exhibiting all the desirable traits. With this approach, genes with known functions from other species can be introduced into tea, opening endless possibilities. To date, no genetically modified tea cultivar is available and it remains an avenue to be explored for future development.

1.8. Concluding Remarks

Different genetic markers may reveal different levels of genetic diversity due to the diversity in marker characteristics. Since morphological characteristics are influenced by the environment, studies show differences compared to DNA based markers. The majority of the genome is non-coding DNA (only about 2% of the genome is encoded). Therefore, the differences detected by DNA based methods will mostly be in the non-coding part. Combining different sets of data should give more precise results, since a more complete portion of the genome will be analysed.

It seems that inbreeding in tea is unlikely to happen due to the outbreeding nature of tea, however “prevention is better than cure”. Breeding programs have to be managed to prevent inbreeding, as the increase in clonal populations will have a negative effect on the gene pool.

Biotechnology provides new alternatives to conventional plant breeding that can assist in the improvement and the understanding of plant breeding. MAS and QTL can improve the efficiency of tea breeding. Genetic engineering of tea can provide the opportunity to improve tea with the gene knowledge of other plant species. This may provide resistance to pests and pathogens and increase tolerance to abiotic stress such as drought. It is important to consider all possible methods of breeding so that this field may progress as fast as possible.

1.9. Future perspectives

Tea is a major commodity that contributes to the economy of third world countries. Therefore, tea resources should be carefully managed. The breeding program of the TRFCA is reaching a yield plateau. It is believed that this can be overcome by the

introduction of new genetic material. The genetic relationship of the existing TRFCA line (in the TRFCA breeding program) should be evaluated in relation to other cultivars, in order to select genetically distant cultivars to include in the program. The management of tea breeding program could play a cardinal role in the global improvement of tea as a commercial crop plant and the competitiveness of the tea industry in Southern Africa.

The exploration and implication of new technologic can contribute to the knolage of tea. Research should focus on the development and/or the identification of QTL and MAS markers to improve the efficiency of tea breeding. Genetic engineering also has great potential for tea breeding and should be investigated for stragegical reasons with due respect to consumer resistance.

The objectives of this study therefore are:

1. To investigate the morphological diversity in the tea cultivars from the TRFCA breeding program and in seedling accessions in Malawi
2. To investigate the microsatellite diversity in the TRFCA cultivars, Malawian seedling selections and cultivars from China
3. To identify possible trait associated RAPD primers

CHAPTER 2

MORPHOLOGICAL DIVERSITY OF SELECTED CULTIVARS AND SEEDLING TEA (*Camellia sinensis*) FROM THE TEA RESEARCH FOUNDATION OF CENTRAL AFRICA

2.1. Introduction

Tea is a major commodity in developing countries where more than 95% of the world's tea is produced (http://faostat.fao.org/site/567/default.aspx#ancor). Tea contributes to economic growth, social habits and in some cases even the religion of countries (Chen *et al.,* 2005b; Devarumath *et al.,* 2002; Grigg, 2002). It is made from the young shoots of *Camellia sinensis*. *Camellia sinensis* belongs to the family Theaceace, genus *Camellia.* Tea is the most important cash crop in the genus *Camellia* and can be divided into three subspecies: *Camellia sinensis sinensis*, *Camellia sinensis assamica* and *Camellia sinensis lasiocalyx* (Singh & Ahuja, 2006), Figure 2.1. The classification of the subspecies is based on leave size, tree morphology, catechins and caffeine content, Table 2.1.

Botanical Classification of Tea

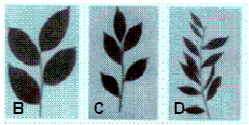
Kingdom:	Plantae
Phylum:	Anthophyta (Flowering plants/angiosperm)
Class:	Dicotyledones
Family:	Theaceace
Section:	Thea
Genus:	Camellia
Species:	Camellia sinensis
Subspecies:	Camellia sinensis sinensis
	Camellia sinensis assamica
	Camellia sinensis lasiocalyx

Figure 2.1: The botanical classification of tea (compiled from Chen *et al.,* 2005b; Singh & Ahuja, 2006; Solomon *et al.,* 2002)

Cultivated tea consists of all three subspecies that contribute to the diversity in this outbreeding species (Mishra & Sen-Mandi, 2004). Tea is cultivated in 52 countries in the world (Mondal *et al.,* 2004). This acclimatization ability is due to a high level of diversity in tea germplasm. Genetic diversity among individuals in a population is an

important factor since it is known that inbreeding and the loss of genetic diversity can lead to an increase in extinction and a decrease in fitness of small laboratory trails (Frankham, 2005).

Table 2.1: A comparison between the subspecies of tea

Subspecies	***Camellia sinensis assamica***	***Camellia sinensis lasiocalyx***	***Camellia sinensis sinensis***
Common name	Assam type	Cambod/Southern type	China type
Country of origin	Assam region, India	?	China
Leaf size	Large 	Intermediate 	Small 
Tree morphology	Tall tree	?	Bush
Caffeine	High	?	Low
Catechins	High	?	Low
Preferred climate	Semi-tropical	Intermediate	Cold
Commercial tea	Black tea	Black & Green tea	Green tea
Compiled from: Astill *et al.,* 2001; Devarumath *et al.,* 2002; Katoh *et al.,* 2003; Kaundun *et al.,* 2000; Magoma *et al.,* 2000; Singh & Ahuja, 2006.			

Genetic diversity can be assessed on three levels: morphological, biochemical and molecular levels. Morphological analysis focus on the phenotype of the plant to assess diversity, while biochemical and molecular analysis analyse the metabolites and DNA, respectively. Various authors (Chen *et al.,* 2005a; Dan, 2006; Su *et al.,* 2007; Thomas *et al*., 2006a) have described the morphological parameters of tea, Table 2.2. In tea the characteristics of the seeds, flowers, leaves and tree morphologies can be analysed to calculate diversity among tea plants. However, since seeds and flowers are expressed only after three to eight years and some tea

plants are sterile the application of seeds and flowers in morphological analysis may delay the analysis (Chen et al., 2008).

In tea, like all hardwood trees with a long reproductive cycle, breeding is a challenging and time consuming process. It takes approximately 15 years to release a tea cultivar. Tea breeding programs concentrate on the improvement of the quality, yield, abiotic and biotic tolerance. Caffeine and catechins are important quality markers in tea (Owuor *et al.,* 2006; Saravanan *et al.,* 2005), while leave parameters and shoot density play an important role in the yield. Leave size influences not only the yield but also plucking efficiency (where manual harvesting is used, the larger the leaves the less shoots/kg tea therefore faster plucking/kg tea and a higher labour efficiency).

Abiotic stress for example drought and temperature cause considerable damage to agricultural production and limit the locations that are suitable for tea production. In tea drought can reduce the yield significantly and lead to major economic losses (Sharma & Kumar, 2005). Biotic stresses in tea include Heliopeltis, *Phomopsis* and mulberry scale (Pseudaulacapis pentagona). In Malawi, the mosquito bug, *Heliopeltis schoutedeni* causes damage to the tea plantations. This insect, both in the lymph and adult stages, suck cell sap from the mature bud, young leaves and stem resulting in necrotic lesions. Heliopeltis damage also affects other crops such as cocoa and cashew nuts (Bhuyan & Bhattacharya, 2006, Ellis & Nyirenda, 1995).

Table 2.2: A summary of the morphological characteristics that could be analysed

Characteristic
Habitat
Life form
Tree shape
Stem
Single or multiple stem
Stem colour (3.5 cm above ground 3 year old plant exposed to full sunlight)
Stem pigmentation
Branch angles
Leave and shoot
Plucking point density
Date of bud first flush
Bud colour
Bud hairiness
Length of plucking shoot (cm)
Weight of plucking shoot (g/100)
Leaf shape
Leaf colour
Leaf texture
Leaf margin serration
Leaf apex shape
Acutance of leaf margin
Petiole hairiness
Main vine hairiness
Vestiture on lower surface
Leaf length (L) (cm)
Leaf breath (B) (cm)
Leaf area (LxB) (cm^2)
Leaf shape index (L/B)
Number of leaf vine
Pubescence of flush shoot
Pubescence on squama
Flower
Date of blooming
Flower bud size (diameter, mm)
Duration of blooming
Pedicel length (mm)
Pubescence on pedicel
Sepal colour

Table 2.2 continues...

Characteristic
Flower
Pubescence on sepal
Sepal diameter (mm)
Flower diameter (mm)
Petal length (mm)
Petal colour
Petal texture
Pubescence on petal
Pubescence on ovary
Length of column (mm)
Styles number
Position of styles divided on column
Height of column compare to stamen
Stamen location style
Fruit and Seed
Fruit set
Fruit size in diameter (mm)
Locule per fruit
Fruit colour
Pericarp thickness (mm)
Fruit peduncle length (mm)
Fruit peduncle diameter (mm)
Seed ripe time
Seed shape
Seed size in diameter (mm)
Seed colour
Seed weight (g/1000)
Abiotic stress susceptibility
Low temperature
High temperature
Drought
Water salinity
Reaction to soil acidity
Reaction to soil salinity
Characteristic
Biotic stress susceptibility
Viruses
Mosaic virus

Table 2.2 continues...

Diseases	
	Phomopsis theae (Collar and branch canker)
	Cephaleuros parasiticus (Red rust)
	Exobasidium vexans (Blister blight)
Pests	
	Helopeltis theivora (Tea mosquito bug)
	Pseudaulacaspis pentagona (Mulberry scale)
	Toxoptera aurantii (Thrips)

Compile from IPGRI, 1997; Chen *et al.,* 2005a

The breeding program of Tea Research Foundation of Central Africa (TRFCA) started in 1956 (Ellis & Nyirenda, 1995). In this program, the 16 parent plants were selected from field accessions and their F1 superior offspring. Since 1956 considerable progress has been made in this program with regard to yield and quality. More recently, this progress in the yield and quality of the new released cultivars seems to have reached a plateau. It appears that the maximum potential in the genetic pool of the original 16 parent plants has been reached. It is hypothesised that this is due to a lack of diversity in the parental plants and offspring. Therefore, the aim is to find new genetic material from field accessions in Malawian seedling gardens that will be genetically distant from the existing lines. This study investigate the morphological diversity of tea cultivars from the TRFCA program and field accessions with superior qualities from different Malawian tea estates.

2.2. Hypothesis

Hypothesis: The genetic distance of the seedling accessions will be wider than the genetic distance of the TRFCA cultivars.

Null hypothesis: The genetic distance of the seedling accessions will not be wider than the genetic distance of the TRFCA cultivars.

2.3. Materials and Methods

The 195 tea accessions that were examined in this study were cultivated at the TRFCA. Data were collected during March 2008 (late summer). The total rain during this month was 191.2 mm, which was received in five days. The minimum and maximum temperature during March was 17.7 °C and 27.8 °C, respectively. TRFCA is located 16° 05' south, 35° 37' east, 650 meters above sea level and has an average annual rainfall of 1500 mm. March is part of the tea harvesting season in Malawi and plants were actively growing. The accessions were from 189 field accessions in seedling gardens from Malawian tea estates and six cultivars from the TRFCA's breeding program (PC1, PC80, PC153, PC168, PC268 and SFS150).

The following six morphological traits were examined: 1) leaf colour, 2) Heliopeltis resistance, 3) shoot density, 4) leaf length, 5) leaf width and 6) internode length. The leaf colour was assigned on the overall appearance of the growing shoots into four colours groups: 1) dark green, 2) light green, 3) pale green and 4) tinged red. The Heliopeltis resistance was assigned on the severity to which the cultivar plots were affected. Healthy plots were assigned a 1, some leaf damage a 2 and stem die back a 3. The shoot densities were measured by counting the number of shoots in a square of 20X20 cm^2 in three tea bushes per accessions. The leaf length and width measurements were of the third leaf on five shoots that were randomly selected. The internode length was measured between the third and fourth leaf on five randomly selected shoots at the two and a bud stage. Cultivars were also classified by visual

inspection by Dr. H Nyirenda, based on their appearance into the three subspecies: 1) China, 2) Assam and 3) Cambod type tea.

Two methods of scoring were used to divide the traits of the accessions into groups. Each morphological trait was divided into five groups: 1) very low, 2) low, 3) medium, 4) high and 5) very high, except for colour and Heliopeltis that were divided into four and three groups, respectively. The first method is based on the scoring of the ranges of the traits. In the range scoring method data ranges of the traits were assigned into the groups using binary code, 1 for presence and 0 for absence, see Table 2.5 and Table 2.6. The ranges of specific traits were allowed to overlap into more than one group. This brings into account the large variation in morphological traits of the cultivars. The second scoring method was based on the means of the traits. The means were scored based on the presence (1) or absence (0) of the mean in the specific group, see Table 2.5 and Table 2.6. This method only allows one positive (presence, 1) in all of the groups per trait.

Data analysis was done using the software, PowerMarker 3.25 (Liu & Muse, 2005), Treeview (Page, 1996) and MEGA4 (Tamura *et al*., 2007). Genetic distances (D_A) between two cultivars were calculated by PowerMarker 3.25 using Nei dissimilarity equation of 1983:

$$D_A = 1 - \frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{a_j}\sqrt{p_{ij}q_{ij}}$$

Where p_{ij} and q_{ij} are the frequencies of i^{th} allele at the j^{th} locus in cultivar X and Y respectively, while a_j is the number of alleles at the j^{th} locus, and m is the number of loci examined (Liu & Muse, 2005). In the case of morphological data, the locus is the trait and the allele is the characteristic groups.

PowerMarker 3.25 using the following formula calculated expected heterozygosity (He):

$$\hat{D}_l = (1 - \sum_{u=1}^{k} \tilde{p}_{lu}{}^2) / (1 - \frac{1+f}{n})$$

This is the gene diversity at the l^{th} locus, k is the number of loci, P_{lu} the frequency of l at u and f is the inbreeding coefficient. The inbreeding coefficient was calculated as in Weir and Hill (2002).

The polymorphic information content (PIC) reflects the number of alleles and their distribution within a population. PIC is calculated using PowerMarker 3.25.

The genetic relationship between accessions was illustrated for each scoring method based on the Nei dissimilarity distances of 1983 and Unweighted Pair Group Method with Arithmetic Averages (UPGMA) as a clustering algorithm.

2.4. Results and Discussion

Morphological trait distribution and characteristics

Morphological traits (6) were used to assign the diversity among the accessions. Each trait, except for colour and Heliopeltis resistance, was divided into five groups, see Table 2.5, according to the two scoring methods. The distribution of the cultivars into these groups is shown in Figure 2.2 and Figure 2.3. Since the grouping of Heliopeltis resistance and colour were strict and the cultivars were only allowed to fall into one of the groups, the distribution of these two traits for both the range and the mean scoring methods are identical (a cultivar can only have one colour leaves or can only be resistant, intermediate or susceptible to Heliopeltis), see Figure 2.3 and

Table 2.6. Table 2.6 is an example of the scoring based on the range and mean scoring methods for SFS150.

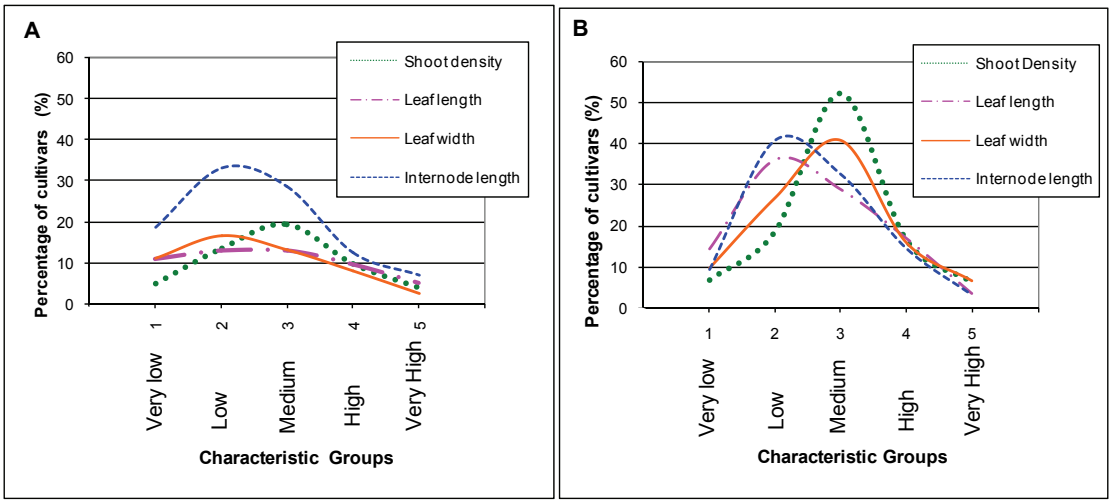


Figure 2.2: The distribution of the morphological traits into the assigned groups, based on A) the range scoring method and B) the means scoring method

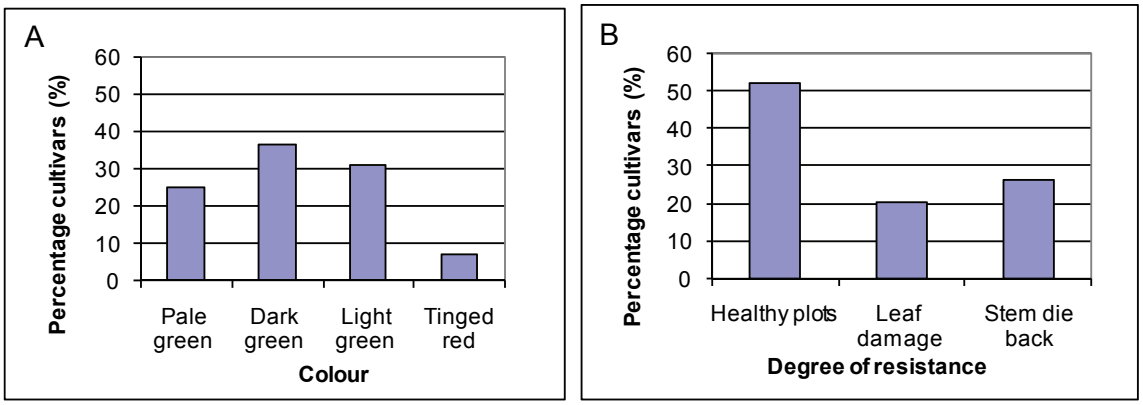


Figure 2.3: The frequency distribution of A) colour and B) Heliopeltis into the respective groups. Colour: P= Pale green, D= Dark green, L= Light green, and T= Tinged red. Heliopeltis resistance: 1 = Healthy plot, 2 = Leaf damage and 3 = Stem die back

The minimum and maximum shoots per 20X20 cm^2 within the 195 analysed cultivars are 3 and 32 shoots, respectively while the mean of the cultivars range from 4 to 27.7. The leaf length, leaf width and internode length ranges from 42 to 221 mm, 22 to 95 mm and 15 to 100 mm, respectively. The averages range from 59 to 126.8 for

leaf length, 25.6 to 65.6 for leaf width and 18 to 71.6 for internode length. The binary data of the trait analysis is listed in Electronic Appendix A for the range scoring method and Electronic Appendix B for the mean scoring method.

Table 2.3: The characteristics of the morphological traits

Trait	Range Parameters			Mean Parameters		
	Major Allele (group) Frequency	Gene (trait) Diversity	PIC	Major Allele (group) Frequency	Gene (trait) Diversity	PIC
Colour	0.3660	0.7040	0.6472	0.3660	0.7040	0.6472
Heliopeltis	0.5206	0.6192	0.5541	0.5206	0.6192	0.5541
Shoot Density	0.2526	0.8751	0.8643	0.5206	0.6590	0.6181
Leaf Length	0.2165	0.8484	0.8312	0.3608	0.7354	0.6906
Leaf Width	0.2835	0.8500	0.8351	0.4072	0.7227	0.6792
Internode Length	0.1959	0.8711	0.8584	0.4072	0.6998	0.6487
Mean	**0.3058**	**0.7946**	**0.7650**	**0.4304**	**0.6900**	**0.6396**

The characteristics of the morphological traits are listed in Table 2.3, Heliopeltis resistance shows the lowest PIC value. Since in this trait an accession can only be scored into one of the groups the value, 0.5541 is identical for both analysis methods. Most of the cultivars (52.06%) are resistant to Heliopeltis. This may be an effect of the selection by the breeding program. The highest PIC value (0.8643) was reported for shoot density with the range scoring method. The PIC values were on average higher for the range scoring method.

The Nei equation of 1983 (Nei & Chesser, 1983) was used to calculate the dissimilarities between cultivars. (Electronic Appendix C for the range scoring method

and Electronic Appendix D for the mean scoring method.) This data was used to construct the UPGMA tree in Figure 2.4 for the range scoring method and Figure 2.5 for the mean scoring method.

Range scoring method

The UPGMA tree, Figure 2.4 revealed seven clusters based on the range scoring method. The six TRFCA cultivars (Figure 2.4) seem to be scattered among the field accessions from the tea estates. Although the six morphological traits could discriminate between all the accessions, it was unable to discriminate between the subpopulations: China, Cambod and Assam type, except for the Chinese tea cultivars that cluster together in cluster E. The number of morphological traits may be too small to discriminate between subspecies. Also contributing to this may be the distribution of the cultivars in the study since 56.7% of the cultivars were classified as Cambod or Cambod hybrid types (77 of 195 cultivars (39.7%) were Cambod, 20 were Cambod/China hybrids (10.3%) and 13 Cambod/Assam hybrids (6.7%)). Cambod is generally seen as an intermediate between China and Assam type tea since the largest part of the accessions is Cambod type and overlapping morphological traits could complicate the analysis. The expected heterozygosity (H_e) for all the accessions in the analysis for this scoring method is 0.7946. The H_e for the subpopulations ranges from 0.6481 for the TRFCA cultivars to 0.7909 for all the seedling accessions, Table 2.4.

Mean scoring method

The UPGMA tree based on the mean scoring method, Figure 2.5 revealed seven clusters. Most of the Chinese type tea cultivars cluster together in cluster G, while the most Assam type cultivars are in cluster A-B. Cluster E and F contain a mix of

Cambod and Chinese type cultivars. Even though this analysis could not clearly distinguish between the different origins of cultivars it does indicate a trend. Since the cultivars were assigned into the three subpopulations based on appearance and it is known that the characteristics of the subpopulations overlap it can be expected that it would be difficult to distinguish between subpopulations. These six morphological traits could not discriminate between all the cultivars using the mean scoring method. The number of traits examined should be increased to enable discrimination among all cultivars. The H_e reported for mean scoring method in 0.6900. The H_e for the subpopulations ranges from 0.5185 for the TRFCA cultivars to 0.6839 for all the seedling accessions, Table 2.4.

Table 2.4: The morphological diversity and PIC values of the accessions grouped according to subpopulations

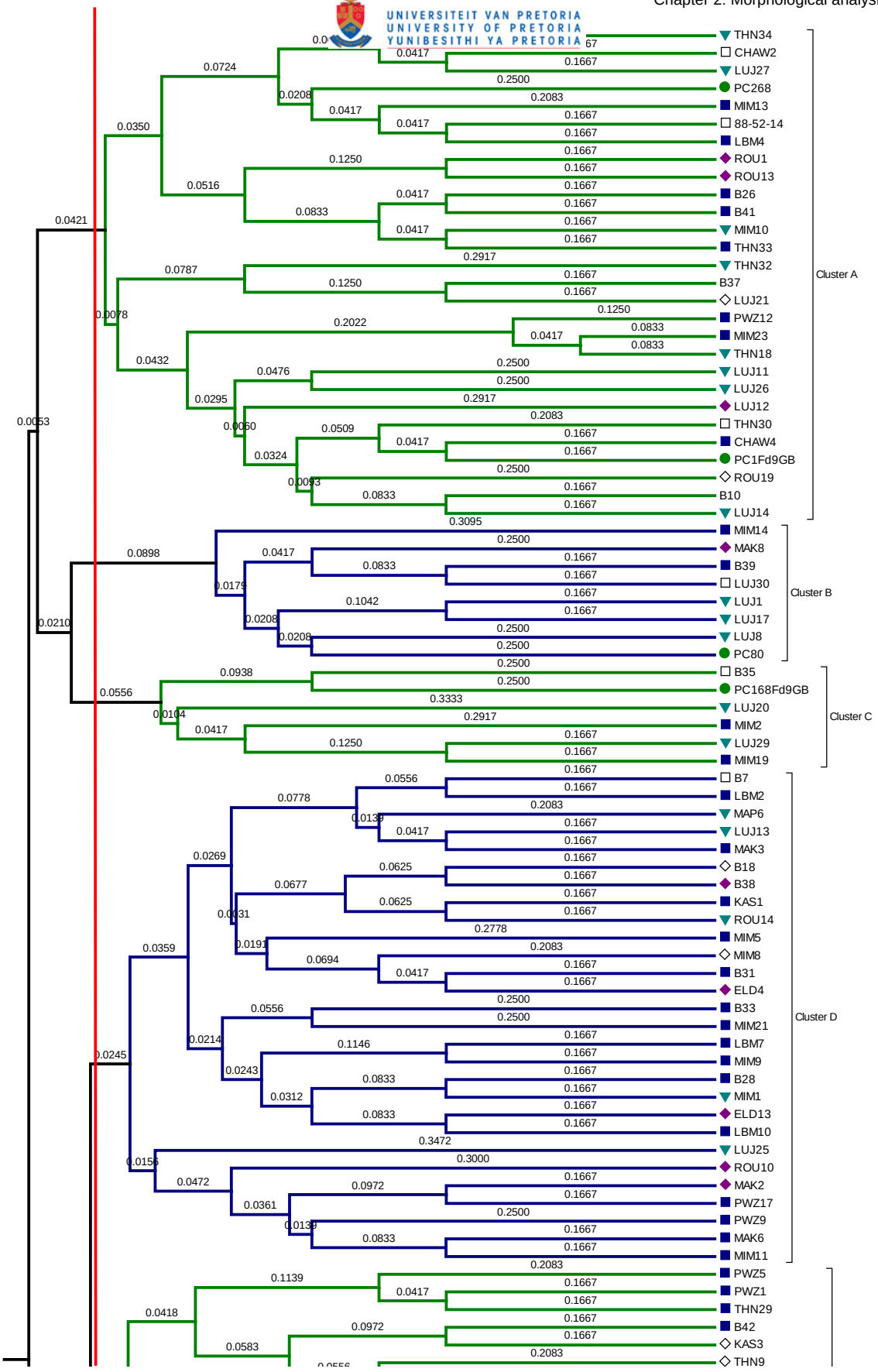
Group	Number of Accessions	Mean Scoring Method		Range Scoring Method	
		Gene (trait) Diversity	PIC	Gene (trait) Diversity	PIC
All Accessions	195	0.6900	0.6396	0.7946	0.7650
Seedling Cultivars	189	0.6839	0.63332	0.7909	0.7610
TRFCA Cultivars	6	0.5185	0.4372	0.6481	0.5756
Assam type	24	0.6389	0.5756	0.7332	0.6948
China type	55	0.6537	0.5946	0.7239	0.6905
Cambod type	77	0.6311	0.5745	0.7623	0.7274
China/Cambod type	20	0.6067	0.5513	0.7767	0.7416
Cambod/Assam type	13	0.6312	0.5654	0.7535	0.7152

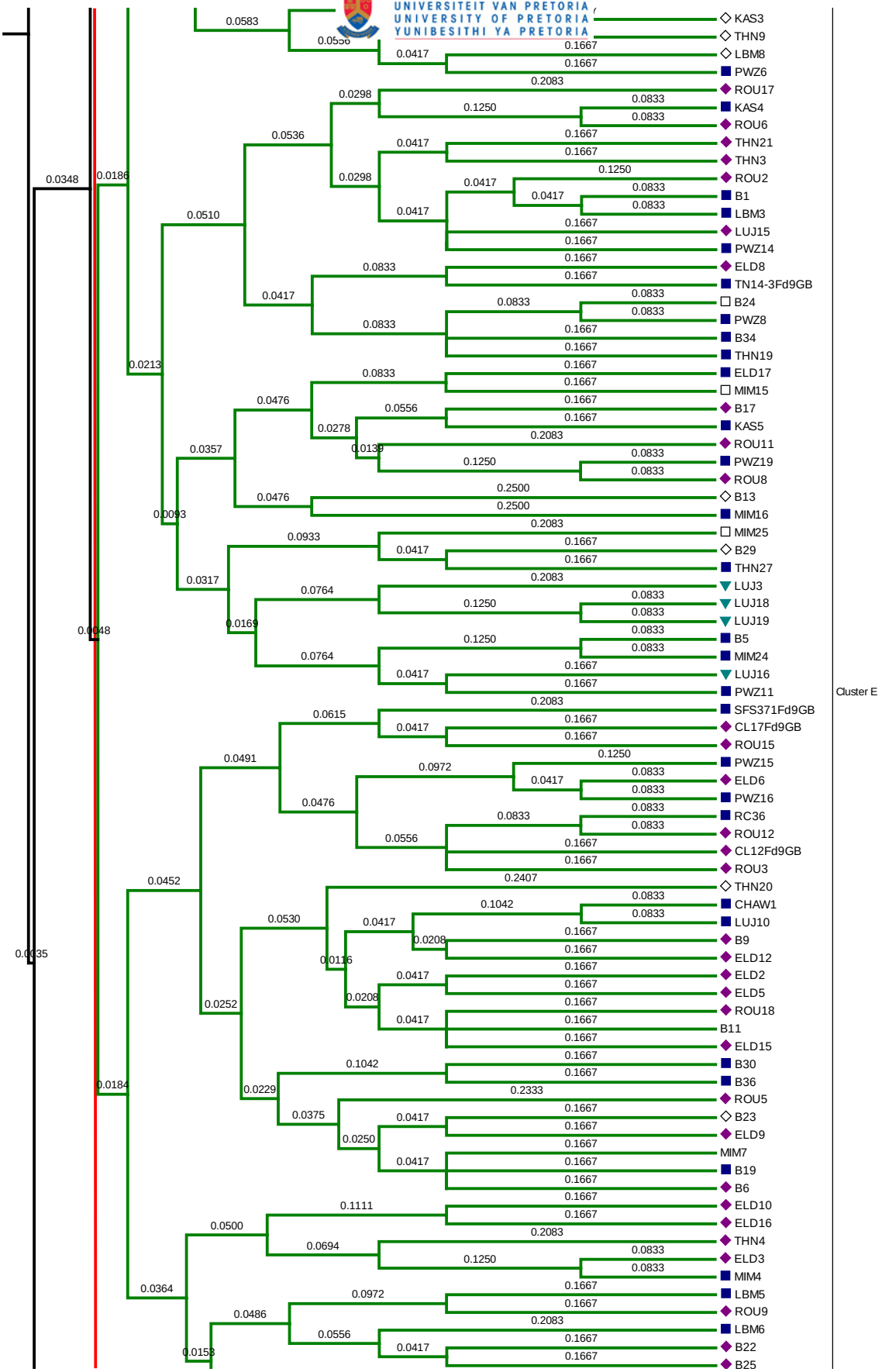
Table 2.5: The distribution of the groups within the morphological traits scored based on the ranges and mean scoring method

Morphological Trait	**Range Scoring Method**						**Mean Scoring Method**					
	Characteristic Groups						**Characteristic Groups**					
Colour	**1**	**2**	**3**	**4**			**1**	**2**	**3**	**4**		
	Pale Green	Dark Green	Light Green	Tinged Red			Pale Green	Dark Green	Light Green	Tinged Red		
Heliopeltis	**1**	**2**	**3**				**1**	**2**	**3**			
	Resistant	Intermediate	Susceptible				Resistant	Intermediate	Susceptible			
	Range	**Characteristic Groups**					**Range of Means**	**Characteristic Groups**				
		1	**2**	**3**	**4**	**5**		**1**	**2**	**3**	**4**	**5**
Shoot Density (shoots per 20x20cm^2)	3-32	0-7	8-10	11-15	16-20	20-∞	4-27.7	0-7	7.01-10	10.01-15	15.01-20	20-01-∞
Leaf Length (mm)	42-221	0-70	71-83	84-96	97-110	110-∞	59-126.8	0-70	70.01-83	83.01-96	96.01-110	110.01-∞
Leaf Width (mm)	22-95	0-30	31-37	38-44	45-50	51-∞	25.6-65.6	0-29	29.02-34	34.01-41	41.01-50	50.01-∞
Internode Length (mm)	15-100	0-30	31-40	41-50	51-60	61-∞	18-71.6	0-30	30.01-40	40.01-50	50.01-60	60.01-∞

Table 2.6: An example of the scoring of SFS150 based on the range and mean scoring methods

SFS150	**Range Scoring Method**						**Mean Scoring Method**					
Morphological Trait	**Characteristic Groups**					**Total**	**Characteristic Groups**					**Total**
Colour	**1**	**2**	**3**	**4**			**1**	**2**	**3**	**4**		
	Pale green	Dark Green	Light green	Tinged red			Pale green	Dark Green	Light green	Tinged red		
Pale green	1	0	0	0		1000	1	0	0	0		1000
Heliopeltis	**1**	**2**	**3**				**1**	**2**	**3**			
	Resistant	Intermediate	Susceptible				Resistant	Intermediate	Susceptible			
Resistant	1	0	0			100	1	0	0			100
	1	**2**	**3**	**4**	**5**		**1**	**2**	**3**	**4**	**5**	
Shoot Density (shoots per 20x20 cm^2)	0-7	8-10	11-15	16-20	20-∞		0-7	7.01-10	10.01-15	15.01-20	20-01-∞	
Range: 16-21 Mean: 18.3	0	0	0	1	1	00011	0	0	0	1	0	00010
Leaf Length (mm)	0-70	71-83	84-96	97-110	110-∞		0-70	70.01-83	83.01-96	96.01-110	110.01-∞	
Range: 42-100 Mean: 81	1	1	1	1	0	11110	0	1	0	0	0	01000
Leaf Width (mm)	0-30	31-37	38-44	45-50	51-∞		0-29	29.02-34	34.01-41	41.01-50	50.01-∞	
Range: 37-41 Mean: 38.8	0	1	1	0	0	01100	0	0	1	0	0	00100
Internode Length (mm)	0-30	31-40	41-50	51-60	61-∞		0-30	30.01-40	40.01-50	50.01-60	60.01-∞	
Range: 35-70 Mean: 58.4	0	1	1	1	1	01111	0	0	0	1	0	00010


THN34
CHAW2
LUJ27
PC268
MIM13
88-52-14
LBM4
ROU1
ROU13
B26
B41
MIM10
THN33
THN32
B37
LUJ21
PWZ12
MIM23
THN18
LUJ11
LUJ26
LUJ12
THN30
CHAW4
PC1Fd9GB
ROU19
B10
LUJ14
Cluster A
MIM14
MAK8
B39
LUJ30
LUJ1
LUJ17
LUJ8
PC80
Cluster B
B35
PC168Fd9GB
LUJ20
MIM2
LUJ29
MIM19
Cluster C
B7
LBM2
MAP6
LUJ13
MAK3
B18
B38
KAS1
ROU14
MIM5
MIM8
B31
ELD4
B33
MIM21
LBM7
MIM9
B28
MIM1
ELD13
LBM10
LUJ25
ROU10
MAK2
PWZ17
PWZ9
MAK6
MIM11
Cluster D
PWZ5
PWZ1
THN29
B42
KAS3
THN9


KAS3
THN9
LBM8
PWZ6
ROU17
KAS4
ROU6
THN21
THN3
ROU2
B1
LBM3
LUJ15
PWZ14
ELD8
TN14-3Fd9GB
B24
PWZ8
B34
THN19
ELD17
MIM15
B17
KAS5
ROU11
PWZ19
ROU8
B13
MIM16
MIM25
B29
THN27
LUJ3
LUJ18
LUJ19
B5
MIM24
LUJ16
PWZ11
SFS371Fd9GB
CL17Fd9GB
ROU15
PWZ15
ELD6
PWZ16
RC36
ROU12
CL12Fd9GB
ROU3
THN20
CHAW1
LUJ10
B9
ELD12
ELD2
ELD5
ROU18
B11
ELD15
B30
B36
ROU5
B23
ELD9
MIM7
B19
B6
ELD10
ELD16
THN4
ELD3
MIM4
LBM5
ROU9
LBM6
B22
B25
Cluster E

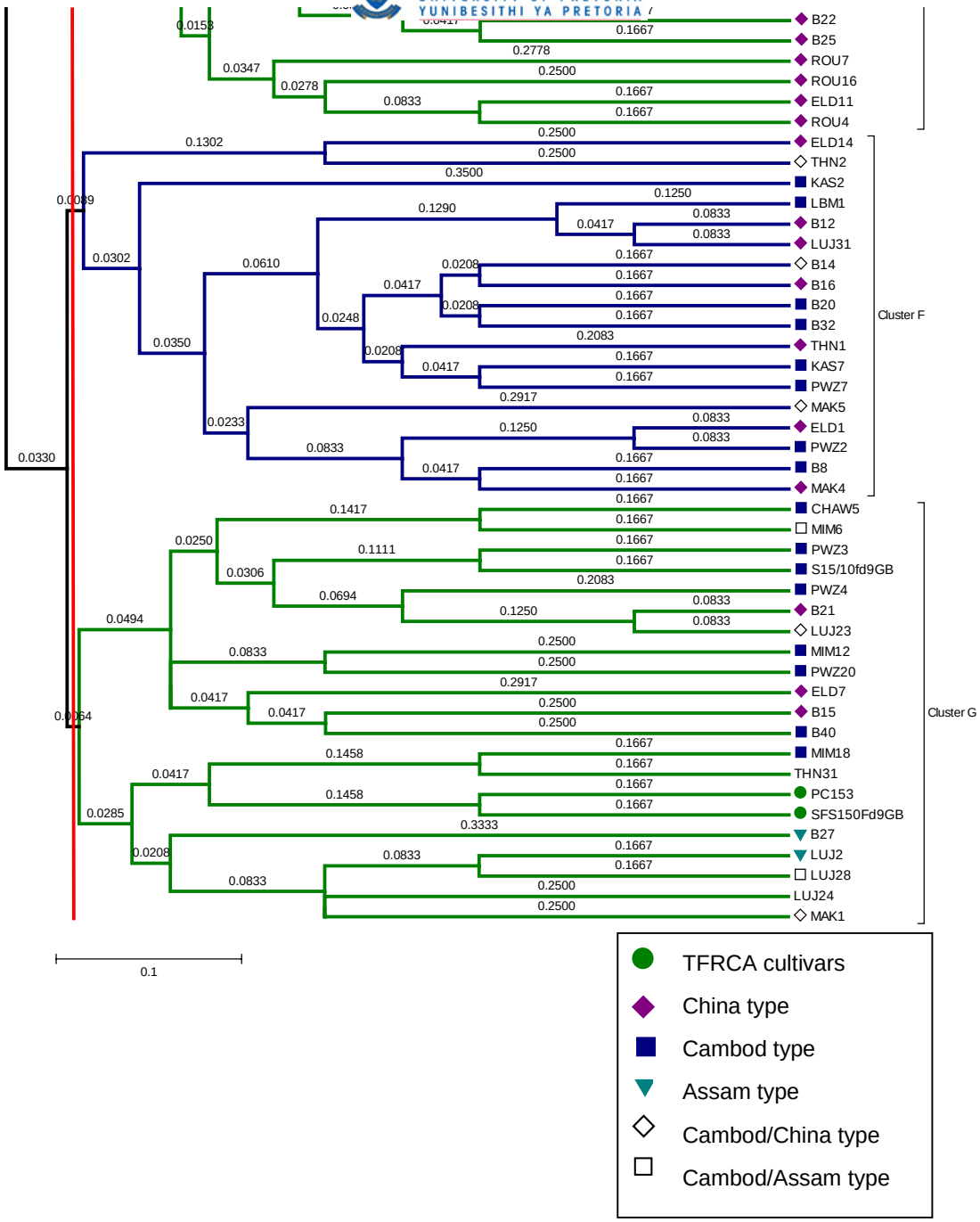
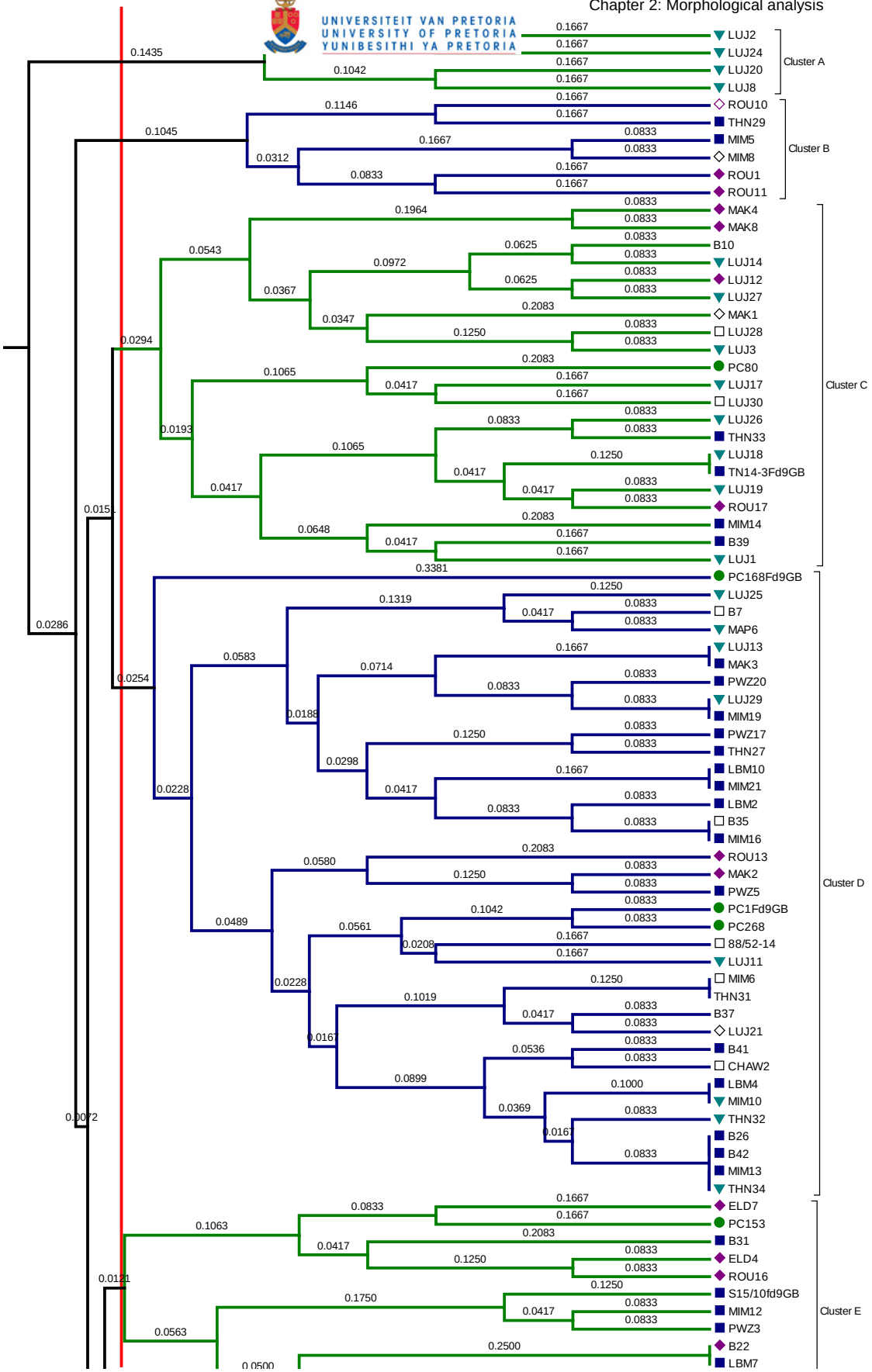
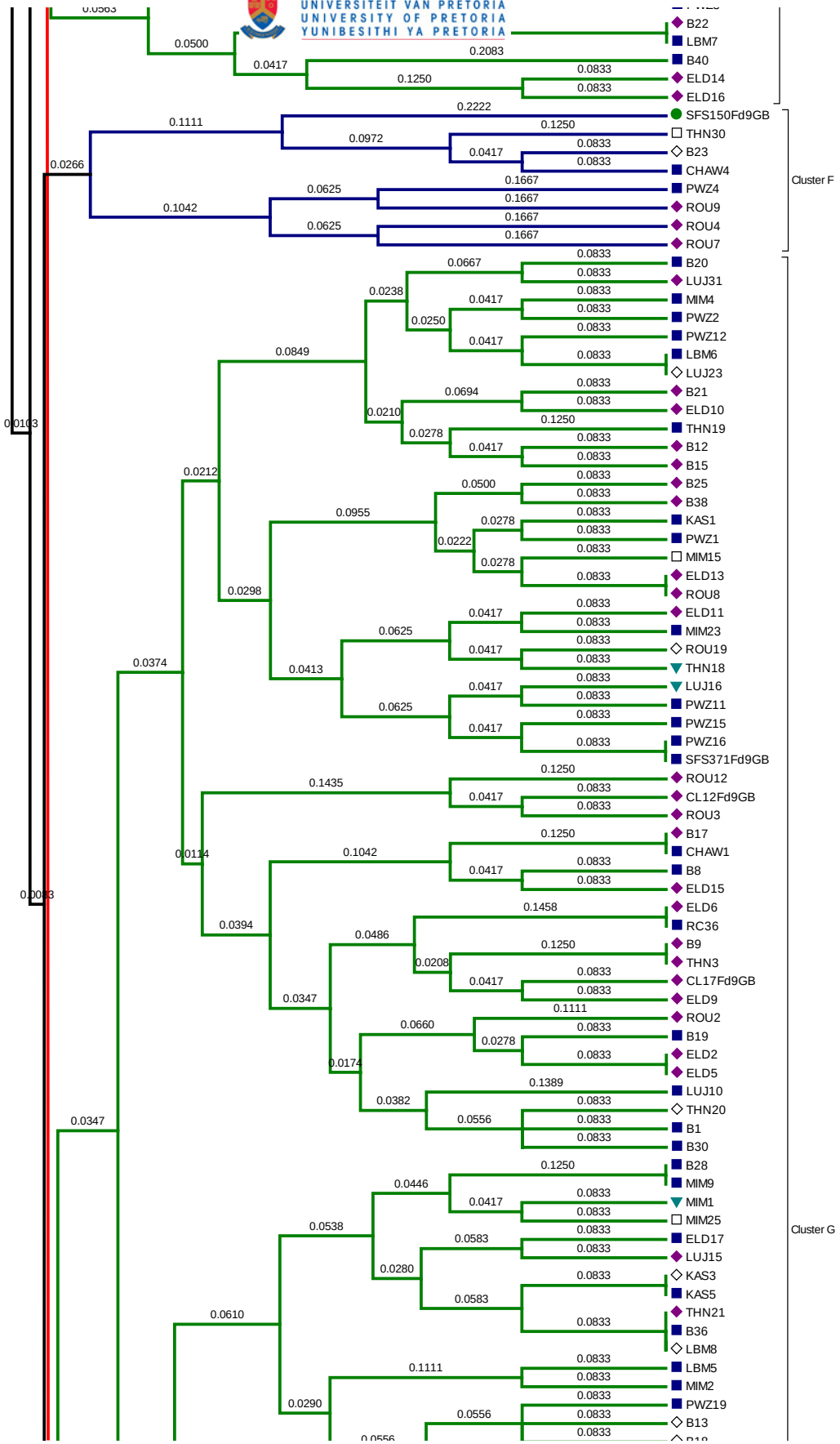


Figure 2.4: The morphological relationship between the 189 seedling accessions and six cultivars from the TRFCA, based on the range scoring method. The relationship is based on Nei distances equation of 1983 and UPGMA clustering algorithm. Seven clusters can be distinguished (in alternative colours and labelled cluster A to G)


LUJ2
LUJ24
LUJ20
LUJ8
Cluster A
ROU10
THN29
MIM5
MIM8
ROU1
ROU11
Cluster B
MAK4
MAK8
B10
LUJ14
LUJ12
LUJ27
MAK1
LUJ28
LUJ3
PC80
LUJ17
LUJ30
LUJ26
THN33
LUJ18
TN14-3Fd9GB
LUJ19
ROU17
MIM14
B39
LUJ1
Cluster C
PC168Fd9GB
LUJ25
B7
MAP6
LUJ13
MAK3
PWZ20
LUJ29
MIM19
PWZ17
THN27
LBM10
MIM21
LBM2
B35
MIM16
ROU13
MAK2
PWZ5
PC1Fd9GB
PC268
88/52-14
LUJ11
MIM6
THN31
B37
LUJ21
B41
CHAW2
LBM4
MIM10
THN32
B26
B42
MIM13
THN34
Cluster D
ELD7
PC153
B31
ELD4
ROU16
S15/10fd9GB
MIM12
PWZ3
B22
LBM7
Cluster E


B22
LBM7
B40
ELD14
ELD16
SFS150Fd9GB
THN30
B23
CHAW4
PWZ4
ROU9
ROU4
ROU7
Cluster F
B20
LUJ31
MIM4
PWZ2
PWZ12
LBM6
LUJ23
B21
ELD10
THN19
B12
B15
B25
B38
KAS1
PWZ1
MIM15
ELD13
ROU8
ELD11
MIM23
ROU19
THN18
LUJ16
PWZ11
PWZ15
PWZ16
SFS371Fd9GB
ROU12
CL12Fd9GB
ROU3
B17
CHAW1
B8
ELD15
ELD6
RC36
B9
THN3
CL17Fd9GB
ELD9
ROU2
B19
ELD2
ELD5
LUJ10
THN20
B1
B30
B28
MIM9
MIM1
MIM25
ELD17
LUJ15
KAS3
KAS5
THN21
B36
LBM8
LBM5
MIM2
PWZ19
B13
Cluster G

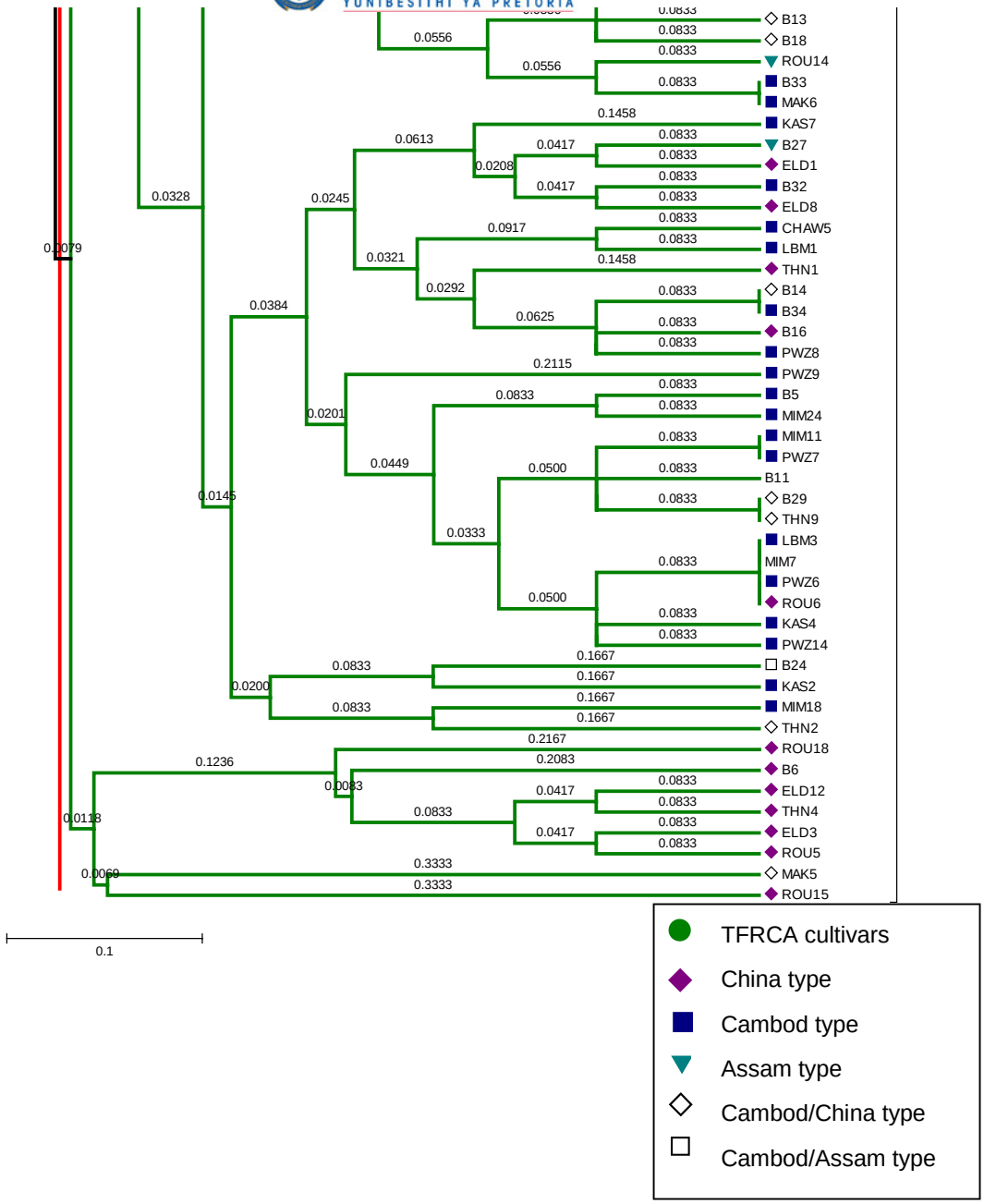


Figure 2.5: The morphological relationship between the 189 seedling accessions and six cultivars from the TRFCA, based on the mean scoring method. The relationship is based on Nei distances equation of 1983 and UPGMA clustering algorithm. Seven clusters can be distinguished (in alternative colours and labelled cluster A to G)

Comparison between the scoring methods

The range scoring method reported a higher H_e (0.7946) than the mean scoring method (0.6900). This is expected since the range scoring method reveals more diversity. Even though the mean scoring method could not discriminate between all the accessions, the distribution is in better agreement with the subpopulation classification. In addition, a low H_e is expected for morphological analysis and therefore the mean scoring method's results are more acceptable.

The genetic relationship between the populations shows simular profiles, Figure 2.6 and Figure 2.7 however, the distances differ between the mean and range scoring methods. The range scoring method reveals larger distances between the populations. The Nei dissimilarity between the populations ranges from 0.1105 to 0.4550 for the range scoring method, it only ranges between 0.0584 and 0.2419 for the mean scoring method, Table 2.7. The TRFCA cultivars do not cluster with any of the subpopulations of seedling accessions; however, the TRFCA cultivars only consist of six cultivars.

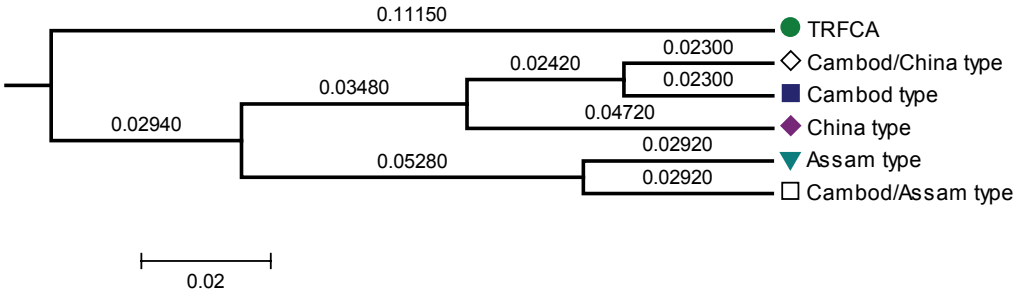


Figure 2.6: The morphological relationship of the subpopulations in the mean scoring method based on Nei dissimilarity of 1983 and UPGMA

Comparison with literature

It is difficult to compare the reported results to other studies since only a small number of morphological characteristics were scored in this study, other studies used a large number of characteristics. A study by Chen *et al.* (2005a) that used allozymes, biochemical and morphological traits as markers reported H_e between 0.114 and 0.218 among tea subpopulations (55 morphological traits, four biochemical parameters, 10 allozymes, 87 accessions). The H_e reported in the current study was higher (0.7946 for the range scoring method and 0.6900 for the mean scoring method). This may be due to the combination of the different markers in the study by Chen *et al.* (2005a), since allozymes reveal very low levels of diversity. Also, the number of traits scored will influence the H_e to a certain extent.

RAPD and AFLP analysis by Wachira *et al.* (2001) showed a H_e of 0.311 and 0.361, respectively (37 RAPD primers, 6 AFLP primers, 44 accessions). SSR studies on tea reported H_e of 0.596 (4 SSR primers, 24 accessions (Kaundun & Matsumoto, 2002)) and 0.865 (6 SSR primers, 96 accessions (Dan, 2006)). This is within the range reported here.

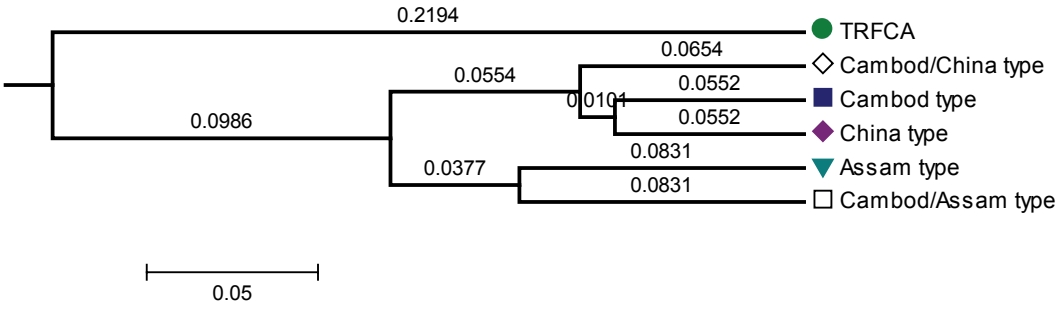


Figure 2.7: The morphological relationship of the subpopulations in the range scoring method based on Nei dissimilarity of 1983 and UPGMA

Table 2.7: The Nei dissimilarity matrix of 1983 for the range scoring method (top half) and the mean scoring method (bottom half)

	Assam type	Cambod type	Cambod/ Assam type	Cambod/ China type	China type	TRFCA
Assam type	0.0000	0.1466	0.1662	0.2372	0.2875	0.3834
Cambod type	0.1026	0.0000	0.2037	0.1166	0.1105	0.4118
Cambod/Assam type	0.0584	0.1037	0.0000	0.2413	0.3331	0.4100
Cambod/China type	0.1485	0.0461	0.1408	0.0000	0.1450	0.4550
China type	0.2397	0.0859	0.2491	0.1029	0.0000	0.5335
TRFCA	0.1963	0.1759	0.2434	0.2346	0.2646	0.0000

The TRFCA cultivars

The TRFCA cultivars and parental plants were all in clusters C to E with the mean scoring method (Figure 2.5). However, this is not the case with the range scoring method were these six cultivars are more scattered. Since the number of TRFCA cultivars in this study is small, it is difficult to draw conclusion from the available information.

In general, the TRFCA cultivars seem to be scattered among the cultivars from the tea estates. The TRFCA cultivars do not cluster together as was expected; this may be evidence for a high level of diversity and thereby supporting the null hypothesis. However, the relationship between the TRFCA cultivars and the subpopulations of the seedling accessions show that the TRFCA cultivars are distant from the seedling accessions. This contradiction may be due to the small sample size of the TRFCA and it is therefore not possible to draw conclusions from the results.

The genetic diversity for tea in Malawi may be narrow and the TRFCA cultivars well distributed within Malawian tea accessions. Therefore, cultivars from other countries should be included in future studies. Countries like China and India would be preferable since it is known that tea is indigenous to these two countries and organisms have a higher level of diversity at their centre of origin (Magoma *et al.,* 2003; Wachira & Ng'etich, 1999).

Morphological traits

Morphological trait analysis is a time consuming and labour intensive process although it requires no expensive equipment. It is known to be influenced by environmental conditions and the developmental stage of the plants (Blears *et al.,* 1998; Gunasekare, 2007; Singh & Ahuja, 2006). In this study, all cultivars were cultivated on the experimental farm of the TRFCA in Malawi therefore limiting variation due to different environmental conditions. However, some of the cultivars were still juvenile trees that could have an effect on the results. External factors can influence the outcome of morphological studies therefore morphological results are debatable. A good suggestion is to support morphological data with molecular data that are not influenced by the above mentioned factors.

2.5. Conclusion

The morphological data scored based on the ranges were able to discriminate between all the tea accessions and the H_e was 0.7946. The mean scoring method was unable to discriminate between all the accessions and the H_e was 0.6900. Of the two scoring methods, the mean scoring method is in better agreement with the plant breeder (Dr. H. Nyirenda) classification based on visual inspection because it shows a better clustering with regard to the subpopulations. The TRFCA cultivars are

scattered among the clusters. The results neither support nor reject the hypothesis since TRFCA cultivars are well distributed among the accessions from Malawi but the number of TRFCA cultivars in the study were too small. Therefore, tea plants from other countries (China and India) should be included in further studies. The small number of TRFCA cultivars is very unfortunate and decrease the credibility of the conclusions. It should be repeated with a more representative group of the 35 released TRFCA cultivars (all released TRFCA cultivars should preferably be included).

CHAPTER 3

INVESTIGATION OF THE GENETIC DIVERSITY AMONG TEA (Camellia sinensis) CULTIVARS USING MICROSATELLITE MARKERS

3.1. Introduction

Camellia sinensis is a hardwood tree indigenous to China and India (Paul *et al*., 1997). This tree's leaves are the commercial source of tea, the most widely consumed beverage in the world, second only to water.

Tea is cultivated in 52 countries (Mondal *et al.,* 2004) and plays an important role in the economies of these countries. To improve the yield and quality of tea, seedling plantations are being replaced with clonal plantations, for example about 86% of the tea in Japan is Yabukita clonal tea plantations (Kaundun *et al*., 2000). This results in a decrease of the genetic diversity among tea trees. Also contributing to the decrease in genetic diversity is tea breeding programs that usually use a small number of superior parental plants. In India, over 75% of the cultivars released by the Tocklai Experimental Station of the Tea Research Association are derived from Betja, Cinnamara or Cambod stocks, while 60% of the cultivars released by the Tea Research Foundation in Kericho, Kenya, are derived from the "6/8" cultivar (Yao *et al*., 2008).

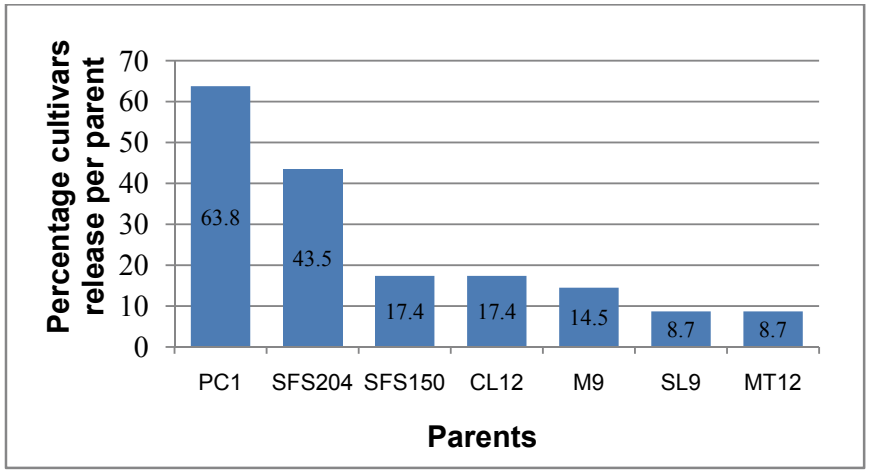


Figure 3.1: The parental statistics of the 39 released cultivars by the TRFCA from 1956 to 2007.63.8

The breeding program of the TRFCA is no exception to this, it consists of 16 parental plants and 39 released cultivars. The parental distribution is indicated in Figure 3.1 and Table 3.1. As seen in Figure 3.1, PC1 and SFS204 are the parents of 63.8% and 43.5%, respectively of the released TRFCA cultivars. It is important for breeding programs to consider the genetic diversity when selecting parental plants.

Table 3.1: Summary of the 39 TRFCA released cultivars from 1956 to 2007 and their parents

Cultivar	Parents	Cultivar	Parents
PC1	Unknown	PC168	PC1 x TRI6/8
PC79	S13 x SFS204	PC169	M9 x SL73
PC80	C5 x CL12	PC175	PC1 x Unknown
PC81	O11 x CL2	PC184	PC1 x MT12
PC105	SFS204 x PC1	PC185	PC1 x MT12
PC108	SFS204 x PC1	PC190	SFS150 x SFS204
PC110	SFS204 x PC1	PC198	SFS150 x MT12
PC113	SFS204 x PC1	PC213	SFS150 x PC1
PC114	SFS204 x PC1	PC268	SFS150 x PC1
PC115	SFS204 x PC1	RC1(MFS87)	Cambod
PC117	SFS204 x PC1	RC2(PC187)	Cambod
PC118	SFS204 x PC1	RC3(7-7)	Unknown
PC119	SFS204 x PC1	RC4(PC164)	PC1 x SL9
PC122	SFS204 x PC1	RC5(PC138)	PC1 x CL12
PC123	SFS204 x PC1	RC6(PC141)	PC1 x CL12
PC127	SFS204 x PC 1	RC7(PC153)	SL5 x SFS150
PC131	M9 x SFS204	RC13(RS233)	Cambod
PC150	M9 x SL1	RC15(PC152)	M9 x PC81
PC153	SL5 x SFS150	RC16(PC219)	I5 x CL12
PC165	PC1 x Unknown		

Genetic diversity can be analysed using molecular markers. These techniques include, RAPD, AFLP, ISSR, CAPS and microsatellites (SSR). Each of these techniques has advantages and disadvantages. The reproducibility of RAPD is problematic, while AFLP is technically demanding. An added disadvantage of RAPD, AFLP and ISSR are that these markers are dominant. CAPS and microsatellites are co-dominated with high reproducibility. Microsatellites are the method of choice (Blears *et al*., 1998; Bussell *et al*., 2005; Gostimsky *et al*., 2005; Varshney *et al*., 2007).

Microsatellites are regions of 1 to 6 bp repeats that are found in most if not all eukaryotes. These regions are spread throughout the genome of organisms in both the functional and non-functional parts of the genome (Tóth *et al*., 2000). Microsatellite primers are designed based on the flanking regions. Therefore, the design requires sequence information (Selko & Toonen, 2006). The development of microsatellites used to have the disadvantage of being costly, however with the increase in the availability of open access EST sequences, the development cost has decreased (Bouck & Vision, 2007; Jiang *et al*., 2006). Since ESTs are part of the functional DNA that is in general more conserved throughout the genus, EST microsatellite primers have the added advantage of been cross species compatible (Bouck & Vision, 2007).

Microsatellite primers were developed for Camellia sinensis (4 primers by Kaundun & Matsumoto, 2004; 15 primers by Freeman *et al*., 2004; 11 primers by Hung *et al*., 2008; 10 EST primers by Jin *et al*., 2006; 24 EST primers by Zhao *et al*., 2008). Some of these primers have been applied to assess genetic diversity of tea cultivars from Vietnam (Dan, 2006) and Japan (Ohsako *et al*., 2008).

Although no mention of microsatellite error rate has been made in previous tea studies (to the authors knowledge), it is an important factor to consider. The error rate is an indication of the reliability of the data. Errors in microsatellite analysis may be due to numerous causes, includes sampling, contamination, scoring and data analysis. Scoring and data analysis are complicated by null alleles, allelic dropout, misscoring and false alleles. Mutations in the complimentary sequence of one of the primers can lead to the occurrences of null alleles (where no alleles are amplified). Allelic dropout is the amplification of only one allele in a heterozygous locus. Stutter bands (artefacts that are a result of PCR amplification) complicate scoring of alleles, which may lead to misscoring of some alleles. False alleles occur during PCR amplification and are allele like artefacts. Allelic dropout, misscoring and false alleles especially can cause problems with a low quality and quantity DNA (Bonin *et al*., 2004; Pompanon *et al*., 2005). It is therefore important to calculate the error rate of the results.

In this study, the genetic diversity in the tea cultivars from the TRFCA (Malawi), selected seedlings from Malawian tea estates and cultivars from China will be determined.

3.2. Hypothesis

Hypothesis 1: The genetic diversity of the TRFCA cultivars will be narrow compared to the diversity in the seedling accessions.

Hypothesis 2: The China origin cultivars will cluster together.

Null hypothesis 1: The genetic diversity of the TRFCA cultivars will not be narrow compared to the seedling diversity.

Null hypothesis 2: The China origin cultivars will not cluster together.

3.3. Materials and Methods

Plant selection

TRFCA cultivars (41) were bred and selected at the TRFCA in Malawi from 1956 to 2007. The seedling accessions (233) were selected from tea estates in Malawi. These seedlings were selected from seedling fields based on special qualities such as vigour, drought tolerance and early spring budding. All TRFCA cultivars and seedling accessions were cultivated at the TRFCA (latitude 16° 05' south, longitude 35° 37' east and altitude 650 meters above sea level). Replicate genotypes of 15 samples were included in the assessments to calculate the error rate. These replicated genotypes were taken from the same clone located in two different fields. Professor Liang Chen (Tea Research Institute, Chinese Academy of Agricultural Sciences) kindly donated DNA samples from 22 Chinese cultivars.

Plant DNA extraction

DNA was extracted from fresh leaves using the Qiagen DNeasy Plant Mini Kit (QIAGEN, Valencia, CA, USA). Plant tissue samples were homogenized for 20 sec at 4.0 m/s in a FastPrep FP120 instrument (QBiogene, Carlsbad, CA, USA). All steps were performed as described in the DNeasy Plant Mini Kit manual. Agarose gel electrophoresis and the GeneQuant Pro UV/Vis Spectrophotometer (Amersham Biosciences, Uppsala, Sweden) were used to determined DNA quality and quantity.

PCR and Genescan conditions

The microsatellite primers (8) were selected from the study by Freeman *et al*., 2004, Table 3.2. Polymerase chain reactions (PCRs) were performed in a total volume of 13 µl with the following components: 15 ±5 ng of genomic DNA, 0.2 µM of each primer (Applied Biosystems, Foster City, USA), 0.2 mM of each dNTP (Fermentas,

Burlington, Canada), 1% DMSO (Sigma-Aldrich, St. Louis, USA), 0.15 U of Exsel High Fidelity DNA Polymerase (JMR Holdings, London, UK) and 1X PCR Exsel buffer. The forward primer in each of the primer set was labelled with a phosphoramidite fluorescent label (FAM, HEX, TET or VIC). DMSO was included in the reaction to improve the yield and specificity of the PCR reaction (Chakrabarti & Schutt, 2001; Kang *et al*., 2005).

PCR amplifications were performed in the Gene-AMP PCR System 9700 (Perkin Elmer, Coneticut, USA). The cycling conditions were as indicated in Table 3.3. The step-down amplification step was included to increase specificity. The additional elongation steps ensure the complete non-template addition of adenine (Magnuson *et al*., 1996).

Table 3.2: The primer information (Freeman *et al*., 2004)

Primer	Sequence	T_A	Label	Dilution[1]
CamsinM2	F: CCTCTGGTGGTCCTACACCT R: AAAGCCTTGATGCCTTTCG	60	6-FAM	1:10
CamsinM4	F: ACATTCAAGCANTCCACATATGTGAAA R: CCTGNTGCAGGACTGTCTATAGATGA	60	PET	1:10
CansinM5	F: AAACTTCAACAACCAGCTCTGGTA R: ATTATAGGATGCAAACAGGCATGA	64	NED	1:10
CamsinM6	F: TGTTTTCTTAGGGTTGGATAAAGG R: TTTTGTTGTAATGACGAAAATTC	55	VIC	1:10
CamsinM8	F: CCATCATTGGCCATTACTACAA R: CCATATGTGTGTGAATGATAAAACC	58	VIC	1:50
CamsinM10	F: TTACATCTCTTTTGCAGCTGTCGG R: CTTCGGGAACTTCTGCTTCATC	64	6-FAM	1:5
CamsinM12	F: CATTATCGTCACTTGCAAAGAGGT R: CGAGAAGAAGAGCTCTATTGGTT	62	NED	1:10
CamsinM13	F: CACATTGTGGCGTGTTATTAATTT R: ACATTGGCTATCTCTCATCATGG	60	PET	1:10

T_A = Annealing temperature
[1] Dilution of the genescan reactions

PCR products were diluted 6:1 with loading dye (0.025% (w/v) bromophenol blue, 30% (v/v) glycerol) and separated on 2% (w/v) agarose (Whitehead scientific, Cape town, RSA) gel containing 0.5 ng/ml ethidium bromide (EtBr), in 1 X TAE buffer (0.04 M Tris-acetate, 1 mM EDTA, pH 8). The Generuler^{TM} DNA ladder, Low Range (Fermentas, Burlington, Canada) was used to determine the size of fragments. Samples were electrophoresed at 100 V in a Maxicell® EC360M Electrophoretic gel System (E-C Apparatus Corporation, St. Petersburg, Florida, USA). Bands were visualised at 312 nm on a Spectroline TC-312A UV transilluminator (Spectronics Corporation, USA), as the interchelated EtBr emits fluorescence at a wavelength of 300 nm (Walker, 2000).

Table 3.3: The PCR conditions for the SSR primers analysis

Cycle conditions		Temperature	Time
Step 1 Denaturation		94 °C	5 min
Step 2 Step-down Amplification 10 cycles	Denaturation	94 °C	30 sec
	Annealing	4 °C higer than the annealing (Table 3.2) with a decrease of 0.4 °C with each cycle	30 sec
	Elongation	68 °C	20 sec
Step 3 Amplification 25 cycles	Denaturation	94 °C	30 sec
	Annealing	Annealing temperature (Table 3.2)	20 sec
	Elongation	68 °C	30 sec
Step 4 Final Elongation		68 °C	10 min
		60 °C	30 min
		20 °C	30 min
		4 °C	Infinity

PCR products were combined and diluted with dddH_2O into two sets, CamsinM2, 8, 12 and 13 together and CamsinM4, 5, 6, together. The dilution factor of each primer product is listed, in Table 3.2. The two sample mixtures were further diluted 1:11 with a formamide solution containing Genescan™ -500 LIZ™: Hi-DiTM formamide, 14:1,000 (Applied Biosystems, Foster City, USA). Samples were denatured and analyzed on an ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems, Foster City, USA) using Genescan™-500 LIZ™ as an internal standard. Electropherograms were analyzed with the ABI PRISM® Genemapper® 3.0 software (Applied Biosystems, Foster City, USA).

Data analysis

The mean error rate per allele was calculated using the following formula:

$$e_l = m_l / nt$$

Where e_l is the mean error rate per allele, m_l is the number of single locus genotypes including at least one mismatch and the number of replicated single locus genotypes is nt, where n is the number of samples and t is the number of repeats (Pompanon *et al*., 2005).

The error rate per reaction (e_r) was calculated using the following formula:

$$e_r = m_l / r$$

Where m_l is the number of single locus genotypes including at least one mismatch and r the number of reactions (Pompanon *et al*., 2005).

The polymorphic information content (PIC) reflects the number of alleles and their distribution within a population. PIC was calculated using PowerMarker 3.25 (Liu & Muse, 2005).

The Excel Microsatellite Toolkit (Park, 2001) was used to convert the microsatellite data to FSTAT format where after FSTAT version 2.9.3 (Goude, 2001) was used to calculate the Nei F-statistics for the populations. The following parameters were calculated with FSTAT version 2.9.3: the observed heterozygosity (H_O), the expected Heterozygosity (H_S), the overall gene diversity (H_T), allelic richness, Inbreeding coefficient of the individuals in a subpopulation (F_{IS}) and variation among subgroups in relation to the total population (F_{ST}).

Genetic distances (D_A) between two cultivars were calculated using PowerMarker 3.25 (Liu & Muse, 2005), using the Nei dissimilarity equation of 1983:

$$D_A = 1 - \frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{a_j}\sqrt{p_{ij}q_{ij}}$$

Where p_{ij} and q_{ij} are the frequencies of i^{th} allele at the j^{th} locus in cultivar X and Y respectively, while a_j is the number of alleles at the j^{th} locus, and m is the number of loci examined (Liu & Muse, 2005).

The genetic relationships between the accessions were illustrated based on the Nei dissimilarity distances of 1983 and Unweighted Pair Group Method with Arithmetic Averages (UPGMA) as the clustering algorithm using PowerMarker 3.25. MEGA4 (Tamura *et al*., 2007) and Treeview (Page, 1996) were used to indicate the clusters and individuals in the subpopulations.

3.4. Results and Discussion

Tea accessions

A total number of 296 accessions were included in this study. These accessions include 22 Chinese cultivars, 41 TRFCA cultivars (labelled PC, SFS or RC) and 233 seedling accessions from Malawian tea estates. The seedling accessions were classified by the tea expert, Dr. H.E. Nyirenda into groups according to their appearances by visual inspection. These groups include Cambod type (92), Assam type (29), China type (82) and the hybrids Cambod/China (16) and Cambod/Assam (14). All the accessions and the classifications are listed in Electronic Appendix E.

Primer selection

The cultivars from the TRFCA, seedling accessions as well as cultivars from China were subjected to microsatellite analysis. The primers were selected from the article by Freeman *et al*. (2004). Of the eight primers tested, listed in Table 3.2, CamsinM12 failed to produce acceptable genescan profiles due to large amounts of scatter bands.

Error rate

DNA from 15 replicated genotypes was extracted to calculate the error rate. These replicates were sampled from two different plants of the same clones growing in different fields. This is 4.7% of the accessions; it is recommended that 5 to 10% of the accessions should be repeated blindly (Pompanon *et al*., 2005). These repeats allow for the assessment of all errors in microsatellite analysis. Technical repeats were also done for four cultivars to assess the error rate of the PCR reactions, genescan and data analysis. In the technical repeats, the PCR reactions were repeated four times from the same DNA sample of each of the four cultivars, after which the data were used to calculate the error rate, Table 3.4.

Table 3.4: Error rates of the technical repeats and overall error rate

	The Error Rate of the Technical repeats					The Overall Error Rate				
Locus	**n x t**	**Error per allele**	**% e_l**	**Error per locus**	**% e_r**	**n x t**	**Error per allele**	**% e_l**	**Error per locus**	**% e_r**
CamsinM2	15	2	6.67	1	6.67	30	7	11.67	4	13.33
CamsinM4	14	13	46.43	7	50.00	30	13	21.67	8	26.67
CamsinM5	14	2	7.14	1	7.14	30	4	6.67	2	6.67
CamsinM6	14	7	25.00	7	50.00	30	24	40.00	14	46.67
CamsinM8	16	<1	<1	<1	<6.25	30	4	6.67	2	6.67
CamsinM10	14	1	3.57	1	7.14	30	4	6.67	4	13.33
CamsinM13	8	1	6.25	1	12.50	30	6	10.00	4	13.33
Mean of all loci			**13.58**		**19.07**			**14.76**		**18.10**
Mean of all loci (excluding CamsinM4 and CamsinM6)			**4.73**		**6.69**			**8.33**		**10.67**
Difference between technical and overall error rates				**e_l = 3.60**					**e_r = 3.98**	

n the number of samples
t the number of repeats
e_l the mean error rate per allele
e_r the error rate per reaction

The mean overall error rate per locus ($\bar{e}_r$) is 18.10% and error rate per allele ($\bar{e}_l$) is 14.76%, while the technical $\bar{e}_r$ is 19.07% and the technical $\bar{e}_l$ is 13.58%, Table 3.4. All these values are high. The high error rates of CamsinM4 (e_l = 21.67% and e_r = 26.67%) and CamsinM6 (e_l =40.00% and e_r = 46.67%) is leading to the high mean error rates. By excluding these two primers from the mean calculation, the overall $\bar{e}_r$ decrease to 10.67% and the overall $\bar{e}_l$ decrease to 8.33%, while the technical $\bar{e}_r$ decrease to 6.69% and the technical $\bar{e}_l$ is decreased to 4.73%. Therefore, CamsinM4 and CamsinM6 were excluded from all further calculations and analysis. The high technical error rates in these two primers were mainly due to false alleles. The exclusion of the two primers (CamsinM4 and 6) did not have a huge influence on the clustering however it did influence the relationship of the cultivars within the clusters.

Not one of the loci in the genotype replica of seedling accession of BB35 or that of TRFCA cultivar PC108 is identical. This may be due to sampling error, planting errors (leaf samples were collected from the same clone planted different fields), labelling errors and/or contamination. The total sampling, labelling and/or contamination errors are the difference between the overall and the technical error rate. Theoretically, these two values should be identical since sampling, labelling and contamination errors will affect both alleles of the locus. This difference is 3.60% for e_l and 3.98% for e_r and is therefore accpectable.

Microsatellite results

All microsatellite results are listed in Electronic Appendix E. Missing data are indicated with a “?”. Some problems were encountered with the China origin cultivars due to quantity of DNA. Of the 22 cultivars, only 11 amplified in all the loci. For the population statistics performed with FSTAT, the accessions with missing data were

excluded. Accessions with missing data were, however included in PowerMarker analysis to illustrate the genetic relationship.

The number of alleles per locus ranges from 17 (CamsinM5) to 28 (CamsinM13) in the 292 accessions, Table 3.5. The mean number of alleles is 21.8 with a total number of 109 alleles in the five loci, Table 3.5. These values are higher than those reported in literature by Ohsako *et al*. (2008) and Freeman *et al.* (2004) however, these values are similar to that reported by Dan (2006).

Table 3.5: The characteristics of the primers

Locus	Allele No[1]	Range	PIC[1]	H_0 [2]	H_S [2]	H_T [2]
CamsinM2	25	159-266	0.9072	0.617	0.893	0.920
CamsinM5	17	171-201	0.8286	0.754	0.772	0.840
CamsinM8	19	110-171	0.8556	0.629	0.802	0.867
CamsinM10	20	157-213	0.6735	0.643	0.647	0.676
CamsinM13	28	110-234	0.8707	0.518	0.861	0.890
Overall	21.8		0.8271	0.632	0.795	0.839

[1] Calculated with PowerMarker 3.25
[2] Calculated with FSTAT

PIC for the loci range from 0.9072 (CamsinM2) to 0.6735 (CamsinM10) and the mean PIC is 0.8271, Table 3.5. The PIC values correspond to that of Freeman *et al*. (2004) and indicate that the loci have a high level of polymorphism. The allelic richness indicates the level of heterozygosity within the populations. The mean allelic richness per subpopulations ranges from 6.6756 to 9.6026, Table 3.6. Overall, the allelic richness per population is low for the TRFCA cultivars.

The observed heterozygosity (H_0) within all the cultivars, ranges from 0.526 (CamsinM13) to 0.741 (CamsinM5) for the loci, Table 3.5. The level of expected

heterozygosity (H_S) within all the cultivars ranges from 0.673 (CamsinM10) to 0.904 (CamsinM2) with a mean of 0.809. In the microsatellite study by Ohsako *et al*. (2008), the reported H_S values ranging from 0.460 to 0.799, which are lower than the current study's values.

Table 3.6: Allelic Richness per population locus based on minimum sample size of 11 diploid individuals

Locus	Assam type	Cambod type	Cambod/ Assam type	Cambod/ China type	China type	China origin	Seedling Selections	TRFCA cultivars
CamsinM2	10.126	6.195	10.387	9.981	12.206	14.000	11.633	7.081
CamsinM5	5.303	5.561	6.790	8.119	8.728	10.000	8.211	6.754
CamsinM8	4.914	7.125	8.596	9.898	9.933	5.000	9.888	7.972
CamsinM10	4.381	5.127	6.257	6.638	8.387	11.000	7.317	5.109
CamsinM13	8.654	9.887	9.695	9.254	8.759	8.000	10.005	7.437
Mean	6.676	6.779	8.345	8.778	9.603	9.600	9.411	6.871

The total heterozygosity (H_T) for the loci, range from 0.701 (CamsinM10) to 0.925 (CamsinM2), is slightly higher than the H_S values. The mean H_T is 0.846, Table 3.5. Ohsako *et al*. reported H_T values ranging from 0.486 to 0.827, higher values were obtained in the current study. This indicated higher levels of heterozygosity in the current study in comparison to the study by Ohsako *et al*. (2008).

F_{IS} is the inbreeding coefficient of the individuals in the subpopulation. This value describes the variation within the individuals within the subpopulation. The F_{IS} value can range between –1 and 1, with 1 implying total inbreeding and –1 being outbreeding. The F_{IS} reported here range from 0.081 (China origin) to 0.317 (Assam type), Table 3.7. These values imply that there are moderate levels of inbreeding

within the subpopulations. It is expected that the level of inbreeding in the China origin accessions would be less than that in the seedlings and TRFCA cultivars. Since it is known that organisms are more diverse at its origin and tea is indigenous to China and India. Interestingly, the TRFCA subpopulation shows a F_{IS} value of 0.183 (Table 3.7), which is within the range and not higher than the seedling accessions (F_{IS} = 0.239).

The F_{ST} is the variation among subgroups relative to the variation within the total population, it is also referred to as the pairwise population differences. The F_{ST} values range from -0.0049 to 0.1373, (Table 3.8). The values were calculated not assuming Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium assumes that all individuals within the population mate at random with respect to genotype and that there is neither selection nor migration (Fairbanks & Andersen, 1999). Therefore, the population in the current study does not meet the requirement to assume Hardy-Weinberg equilibrium.

The significance of the F_{ST} values between the subpopulations ($P < 0.05$) is indicated in (Table 3.8). This table indicates that all the subpopulations differ significantly from both the TRFCA and China cultivars, Table 3.8. Significant differences were also found between Assam type and all other groups except for Cambod type. Although Assam type and Cambod/Assam type differ significantly, Cambod/Assam type does not differ significantly from Cambod type. Cambod/China type does not show any significant difference to that of China type or Cambod type.

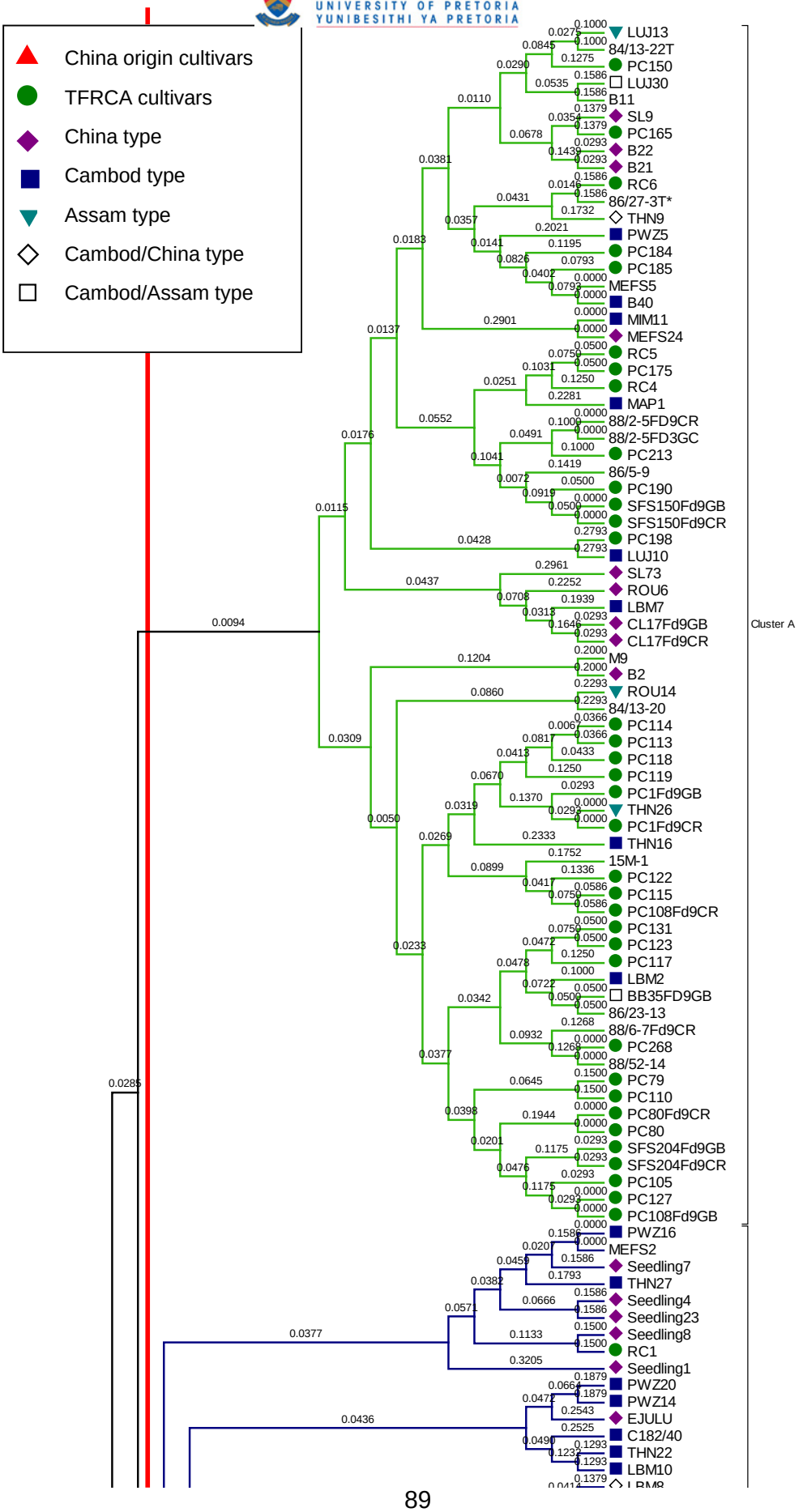
Table 3.7: Inbreeding coefficient per population (F_{IS}) with +1 being total inbreeding and -1 total outbreeding

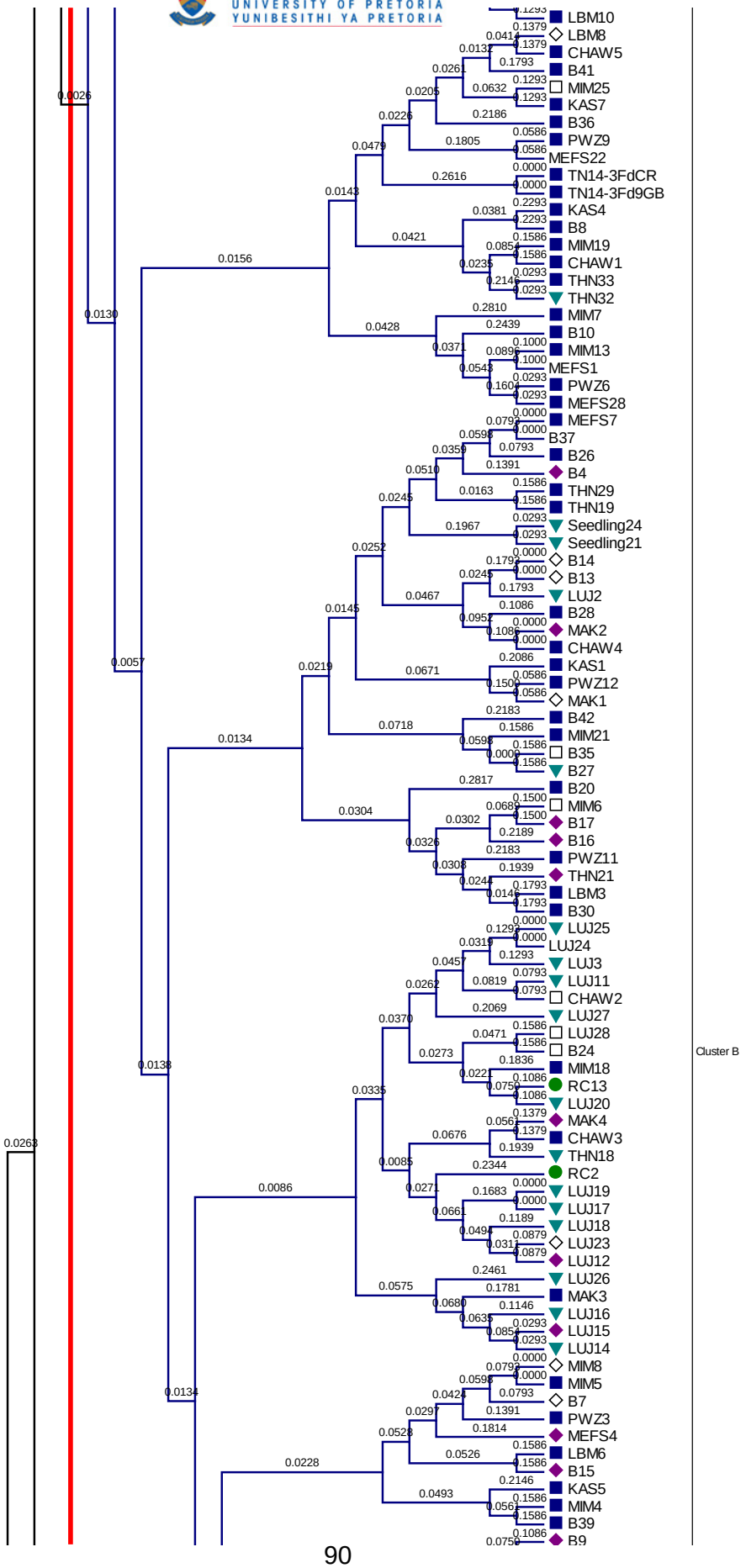
	Assam type	Cambod type	Cambod/ Assam type	Cambod/ China type	China type	China origin	Seedling Selections	TRFCA cultivars	All cultivars
CamsinM2	0.298	0.272	0.186	0.641	0.360	0.139	0.333	0.193	0.321
CamsinM5	0.353	0.050	-0.164	-0.011	0.115	0.095	0.111	-0.106	0.103
CamsinM8	0.250	0.132	0.327	0.283	0.321	-0.067	0.239	0.394	0.264
CamsinM10	0.031	-0.03	0.065	-0.036	0.126	0.053	0.071	-0.056	0.061
CamsinM13	0.529	0.369	0.498	0.517	0.332	0.149	0.387	0.408	0.397
Mean	**0.317**	**0.176**	**0.209**	**0.296**	**0.252**	**0.081**	**0.238**	**0.183**	**0.239**

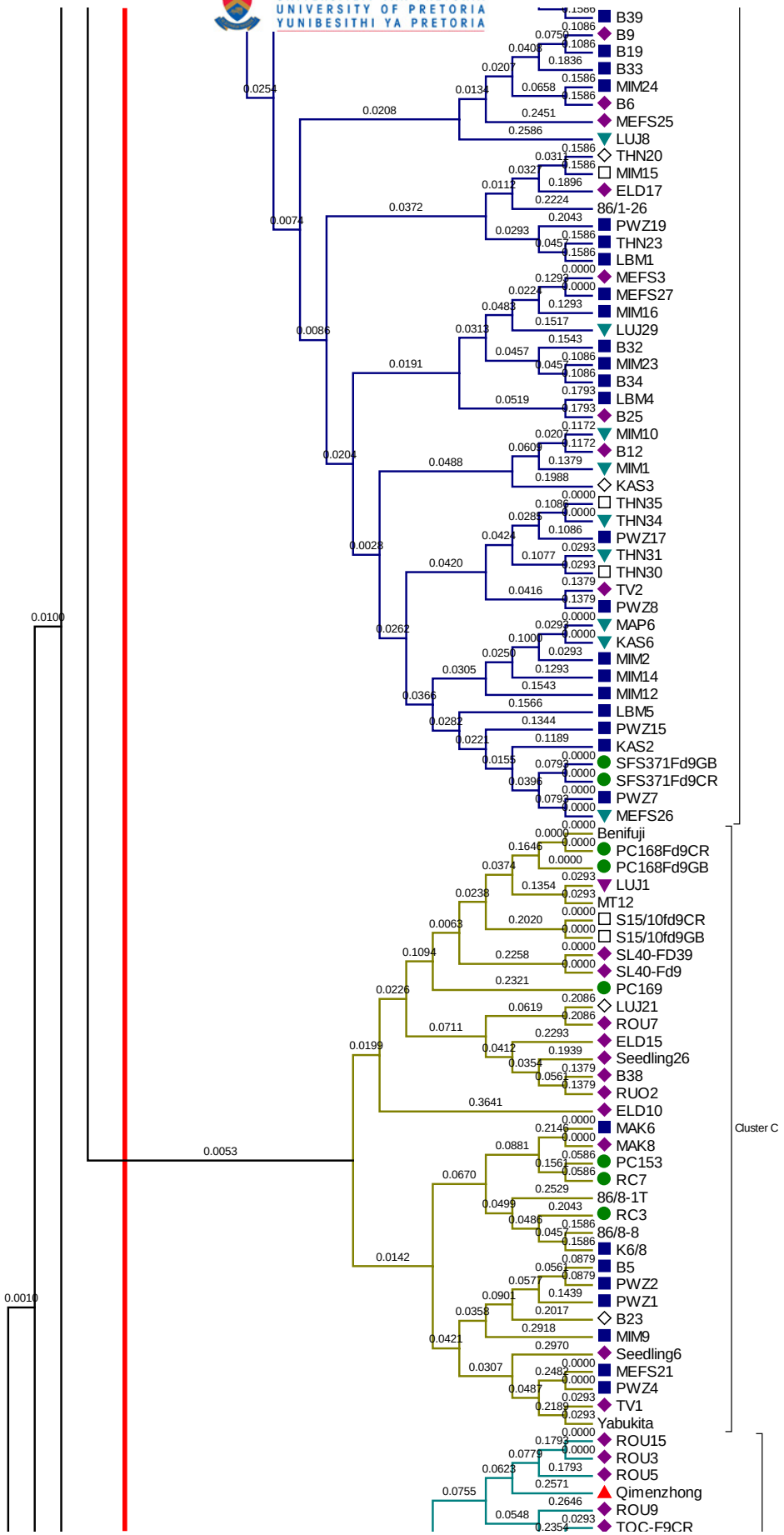
Table 3.8: The pairwise F_{ST} values (not assuming Hardy-Weinberg equilibrium within subgroups) between the subgroups (top half) with the pairwise significant (P < 0.05) after standard Bonferroni corrections

	Assam type	Cambod type	Cambod/Assam type	Cambod/China type	China type	China origin	TRFCA cultivars
Number of individuals	**29**	**92**	**14**	**16**	**82**	**11**	**41**
Assam type	0	0.0002^{NS}	0.0254*	0.0347*	0.0694*	0.1373*	0.0586*
Cambod type		0	0.0172^{NS}	0.0393^{NS}	0.0737*	0.1349*	0.0747*
Cambod/Assam type			0	0.0060^{NS}	0.0444*	0.1045*	0.0550*
Cambod/China type				0	0.0097^{NS}	0.0682*	0.0457*
China type					0	0.0596*	0.0580*
China origin						0	0.1135*
TRFCA cultivars							0

* Significance at the 5% nominal level
NS Non-significant


China origin cultivars
TFRCA cultivars
China type
Cambod type
Assam type
Cambod/China type
Cambod/Assam type
LUJ13
84/13-22T
PC150
LUJ30
B11
SL9
PC165
B22
B21
RC6
86/27-3T*
THN9
PWZ5
PC184
PC185
MEFS5
B40
MIM11
MEFS24
RC5
PC175
RC4
MAP1
88/2-5FD9CR
88/2-5FD3GC
PC213
86/5-9
PC190
SFS150Fd9GB
SFS150Fd9CR
PC198
LUJ10
SL73
ROU6
LBM7
CL17Fd9GB
CL17Fd9CR
M9
B2
ROU14
84/13-20
PC114
PC113
PC118
PC119
PC1Fd9GB
THN26
PC1Fd9CR
THN16
15M-1
PC122
PC115
PC108Fd9CR
PC131
PC123
PC117
LBM2
BB35FD9GB
86/23-13
88/6-7Fd9CR
PC268
88/52-14
PC79
PC110
PC80Fd9CR
PC80
SFS204Fd9GB
SFS204Fd9CR
PC105
PC127
PC108Fd9GB
PWZ16
MEFS2
Seedling7
THN27
Seedling4
Seedling23
Seedling8
RC1
Seedling1
PWZ20
PWZ14
EJULU
C182/40
THN22
LBM10
Cluster A


LBM10
LBM8
CHAW5
B41
MIM25
KAS7
B36
PWZ9
MEFS22
TN14-3FdCR
TN14-3Fd9GB
KAS4
B8
MIM19
CHAW1
THN33
THN32
MIM7
B10
MIM13
MEFS1
PWZ6
MEFS28
MEFS7
B37
B26
B4
THN29
THN19
Seedling24
Seedling21
B14
B13
LUJ2
B28
MAK2
CHAW4
KAS1
PWZ12
MAK1
B42
MIM21
B35
B27
B20
MIM6
B17
B16
PWZ11
THN21
LBM3
B30
LUJ25
LUJ24
LUJ3
LUJ11
CHAW2
LUJ27
LUJ28
B24
MIM18
RC13
LUJ20
MAK4
CHAW3
THN18
RC2
LUJ19
LUJ17
LUJ18
LUJ23
LUJ12
LUJ26
MAK3
LUJ16
LUJ15
LUJ14
MIM8
MIM5
B7
PWZ3
MEFS4
LBM6
B15
KAS5
MIM4
B39
B9
Cluster B


B39
B9
B19
B33
MIM24
B6
MEFS25
LUJ8
THN20
MIM15
ELD17
86/1-26
PWZ19
THN23
LBM1
MEFS3
MEFS27
MIM16
LUJ29
B32
MIM23
B34
LBM4
B25
MIM10
B12
MIM1
KAS3
THN35
THN34
PWZ17
THN31
THN30
TV2
PWZ8
MAP6
KAS6
MIM2
MIM14
MIM12
LBM5
PWZ15
KAS2
SFS371Fd9GB
SFS371Fd9CR
PWZ7
MEFS26
Benifuji
PC168Fd9CR
PC168Fd9GB
LUJ1
MT12
S15/10fd9CR
S15/10fd9GB
SL40-FD39
SL40-Fd9
PC169
LUJ21
ROU7
ELD15
Seedling26
B38
RUO2
ELD10
MAK6
MAK8
PC153
RC7
86/8-1T
RC3
86/8-8
K6/8
B5
PWZ2
PWZ1
B23
MIM9
Seedling6
MEFS21
PWZ4
TV1
Yabukita
ROU15
ROU3
ROU5
Qimenzhong
ROU9
TOC-F9CR
Cluster C

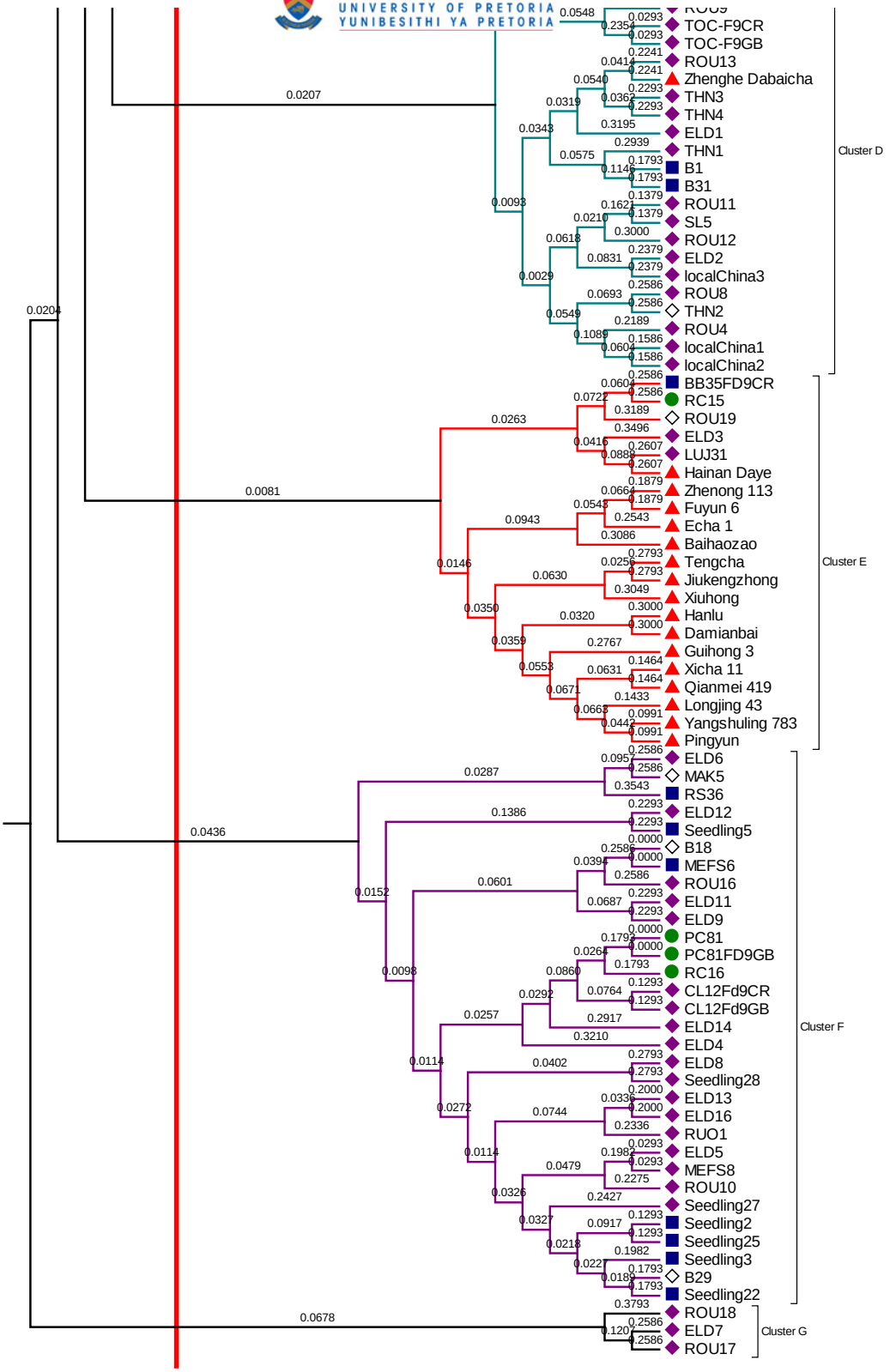


Figure 3.2: The genetic relationship among 41 TRFCA cultivars, 233 seedling accessions and 18 China origin cultivars using the UPGMA method based on the Nei dissimilarity equation of 1983

The genetic relationship among the cultivars is illustrated in Figure 3.2. Most of the TRFCA cultivars group together in cluster A. PC1 and most of its offspring cluster together within cluster A, which also includes SFS204 and its offspring. Most of the China origin cultivars can be found in cluster E. Cluster B consist mainly of Cambod type seedling accessions. The China type seedling accessions are in cluster D, F and G, however cluster G only consist of three cultivars. The Assam type seedling accessions are mostly distributed among the TRFCA cultivars. An example of genetically distant cultivars from PC1 or SFS204 (cluster A) will be the cultivars ELD14 and ROU1 (cluster F). The Nei dissimilarities matrix of 1983 for the accessions can be seen in Electronic Appendix F.

The genetic relationship among the populations is illustrated in Figure 3.3. The China origin cultivars are separated from both the seedling accessions (Assam, Cambod, China type) and the TRFCA cultivars. The TRFCA cultivars group together with the Assam and Cambod/Assam type. The Nei dissimilarity of 1983 ranges from 0.0983 for Cambod and Cambod/China type to 0.5467 for China origin and Assam type.

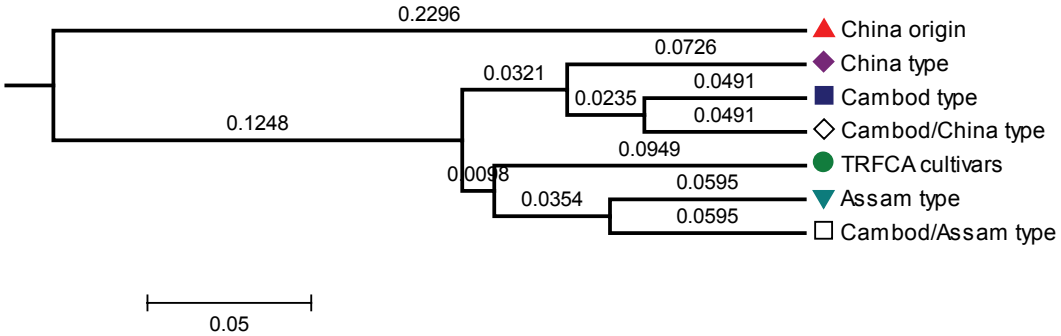


Figure 3.3: The genetic relationship among the populations based on UPGMA method with Nei dissimilarity equations of 1983.

Table 3.9: The Nei dissimilarity matrix of 1983 of the populations

Populations	Cambod type	Cambod/ Assam type	Cambod/ China type	China origin	China type	TRFCA cultivar
Assam type	0.1441	0.1190	0.1955	0.5467	0.2816	0.2113
Cambod type		0.1342	0.0983	0.4401	0.1542	0.1992
Cambod/ Assam type			0.2121	0.5374	0.3065	0.1684
Cambod/ China type				0.3920	0.1364	0.1583
China origin					0.3509	0.4880
China type						0.2540

3.5. Conclusion

Although not many studies consider the effect of error rate on the results, it is important to know the reliability of the results and conclusions. In this study, the error rate was calculated and the high error rate loci (CamsinM4 and CamsinM6) were excluded from the study to increase the reliability of the results.

The results support both hypothesis 1 and 2, since the TRFCA and China origin cultivars cluster together and the F_{ST} values indicated significant differences between the TRFCA and China origin cultivars with all other groups. The China origin cultivars have the lowest level of inbreeding, which is expected since organisms are more diverse at their origin. The level of inbreeding of the TRFCA cultivars is within the range of the seedling accessions.

The TRFCA cultivars cluster together in cluster A, Figure 3.2. The F_{IS} value for the TRFCA cultivars indicated moderate levels of inbreeding. While the F_{ST} values differ significantly for the TRFCA cultivars from all other subpopulations. The allelic

richness of the TRFCA cultivars is on the low end of the range of the allelic richness reported here.

Although the level of inbreeding in the TRFCA is moderate, the allelic richness and heterozygosity (H_0, H_S and H_T) of the TRFCA cultivars is in the low end of the ranges observed. Indicating that the TRFCA breeding program may benefit from the introduction of genetically distant tea plants, to expand the gene pool and increase the level of heterozygosity.

The tea breeding community aspire to control and improve tea as a commodity. Therefore, studies like this contribute to the knowledge of the genetic diversity among tea. Accessions from other countries could be included to assess the tea on a global scale. Preferably, the number of loci accessed should be increased for a more accurate quantification of the diversity among the accessions. Knowledge of the diversity among the accessions allows for better management of the diversity within tea as a commodity.

CHAPTER 4

THE IDENTIFICATION OF A POSSIBLE TRAIT-ASSOCIATED RANDOM AMPLIFIED POLYMORPHIC DNA MARKER FOR DROUGHT TOLERANCE IN TEA (*Camellia sinensis*)

4.1. Introduction

Camellia sinensis is an evergreen tree that is widely known because tea is made from its leaves. The impact of tea on the health of people is long known. In 1211, Eisai ("Father of Tea" in Japan) said in the book, Maintaining Health by Drinking Tea, "Tea *is a miraculous medicine for the maintenance of health. Tea has an extraordinary power to prolong life"* (Wheeler & Wheeler, 2004). Today tea is the second most widely consumed beverage in the world, second only to water (Ferruzzi & Green, 2006).

Tea breeding that focuses on the improvement of the plant's quality and yield is a worldwide initiative. In China, the world's largest tea producing country, tea trees are preserved in the China National Germplasm Tea Repositories (Chen & Yamaguchi, 2002; Chen & Yamaguchi, 2005c). This is to safeguard the genetic diversity among tea since the diversity could be decreasing because of the replacement of seedling plantations with clonal plantations. Southern Africa's own tea breeding program is situated in Malawi. The Tea Research Foundation of Central Africa is exploiting all possibilities to improve tea cultivars for the African climate and markets. Tea breeding is a long term project like all trees, mainly due to long field trial periods. The release of a new tea cultivar takes approximately 15 to 20 years.

Tea breeding and selection depends on the interplay of factors such as resistance to biotic stress, tolerance to abiotic stress, quality and yield of the processed tea leaves (Mondal, 2002). This interplay is illustrated in Figure 4.1, the improvement of a specific trait will influence the yield and/or quality of the tea produced. This will lead to an increased profit, if the other quality or quantity parameters do not decrease.

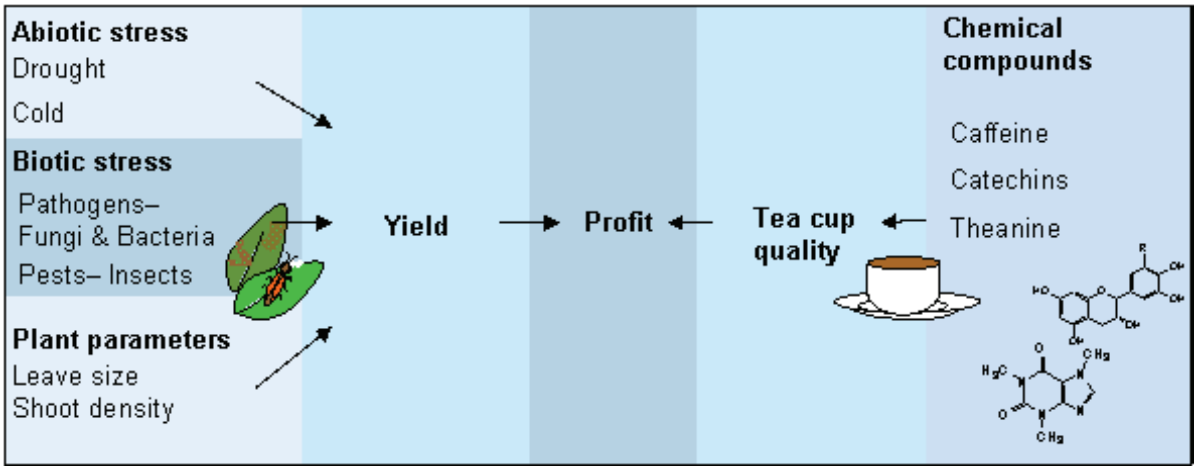


Figure 4.1: The factors influencing tea breeding

Biotic stress is pests (insects) and pathogens (bacteria and fungi). The International Plant Genetic Resources Institute (IPGRI) listed 40 insects and 28 diseases that affect *C. sinensis* (IPGRI, 1997). Among these are Phomopsis (*Phomopsis theae*), a fungal infection that cause collar and branch canker (IPGRI, 1997; Ponmurugan *et al.,* 2007) and *Helopeltis schoutedeni* or mosquito bug, an insect that cause damage to the tea in Malawi (Ellis & Nyirenda, 1995).

Abiotic stress includes temperature and drought. These parameters have a huge influence on the adaptability to climates and the survival of a tree. Interestingly, subspecies can be distinguished based on cold resistance or susceptibility (Singh & Ahuja, 2006). Tea tolerating cold will also have more off-season growth than susceptible plants. From the perspective of a tea breeder, off-season growth implies a longer growth season, which increases the total yield for the year. The selection for drought resistance is one of the most important factors in plant breeding. Drought has an extensive impact on all cultivated plants (Babu *et al.,* 2003; Bruce *et al.,* 2002; Maqbool *et al.,* 2008). Therefore, drought is considered in most, if not all, plant breeding programs around the world.

Tea quality markers include catechins, theanine and caffeine. The concentration of these compounds influences both the taste and the health properties of tea. Caffeine is responsible for the bitterness and briskness of tea (Paul, 2008), while the taste of theanine is described as umami, meaty or brothy (Deng *et al.,* 2008). Catechins are antioxidants, caffeine increase alertness and anxiety while theanine reduce blood pressure and cause relaxation (Ferruzzi & Green, 2006; Rogers *et al.,* 2008). The concentration of these metabolites influences the price of tea (Paul, 2008), Figure 4.1.

The improvement of plant selection based on DNA markers can decrease the selection time in plant species (Peleman & Van der Voort, 2003). Quality Trait Loci (QTLs) are regions on the chromosome that are linked to a specific molecular trait. These regions are either close or within the gene that governs the trait. QTLs can be exploited in plant breeding by developing them into Marker Assisted Selection (MAS) markers. MAS markers are used in breeding to select for traits since DNA markers are not influenced by the environment or developmental stage of the plant. Plants can be tested for these qualities at an early stage that can lead to an improved efficiency in the breeding program. However, in tea only one MAS marker has been developed. This marker developed by Tanaka and Taniguchi in 2003 (as reference by Tanaka & Taniguchi, 2007) for the identification of Mulberry Scale resistance in tea.

The first step in the development of MAS markers is the identification of a QTL. QTL can be identified using different approaches. One such an approach is Random Amplified Polymorphic DNA (RAPD). RAPD primers use 8 to 12 base primers to amplify non-specific random parts of the genome. RAPDs are usually applied to

assess genetic diversity within plants or animals. In tea RAPDs have been used to evaluated the genetic diversity among cultivars (Kaundun *et al.,* 2000; Wachira *et al.,* 2001), to evaluate the genetic relationship between cultivated tea and wild tea (Lai *et al.,* 2001; Singh *et al.,* 2004), to evaluate the genetic integrity of micropropagated tea plants (Devarumath *et al.,* 2002) and to identify a possible drought tolerant marker (Mishra & Sen-Mandi, 2004). However, due to the random nature of this approach the possibility exists that a fragment in the analysis may associate with specific traits. These fragments will be either close to or within the gene of interest. RAPD is an inexpensive approach for the identification of trait associated markers. However, it relies on the assumption that if enough markers are tested a trait associated fragment will be identified. In this study, RAPD markers were tested to identify association with traits.

4.2. Hypothesis

Hypothesis 1: The $RAPD7_{1400}$ will associate with all of the drought tolerant cultivars as in the study by Mishra & Sen-Mandi (2004).

Hypothesis 2: Using RAPD markers from other plants species will increase the probability of finding useful RAPD markers in tea.

Null hypothesis 1: The $RAPD7_{1400}$ will not associate with all of the drought tolerant cultivars as in the study by Mishra & Sen-Mandi (2004).

Null hypothesis 2: Using RAPD markers from other plants species will not increase the probability of finding useful RAPD markers in tea.

4.3. Materials and Methods

Plant material and DNA extraction

Tea plants were selected based on field performance data collected over many years by an expert tea breeder, Dr. H.E. Nyirenda, Table 4.1. Three traits were evaluated: Heliopeltis resistance, drought tolerance and cold tolerance. The plants were all cultivated at the TRFCA, Malawi (latitude 16° 05' south, longitude 35° 37' east and altitude 650 meters above sea level). DNA was extracted from fresh leaves using the Qiagen kit for plant DNA extraction, DNeasy mini kit (QIAGEN, Valencia, CA, USA). Plant tissue samples were homogenized for 20 sec at 4.0 m/s in a FastPrep FP120 instrument (QBiogene, Carlsbad, CA, USA). All steps were performed as described in the DNeasy Plant Mini Kit manual. DNA quality and quantity were determined by agarose gel electrophoresis and the GeneQuant Pro UV/Vis Spectrophotometer (Amersham Biosciences, Uppsala, Sweden).

RAPD primers and PCR conditions

RAPD primers were selected from previous studies, Table 4.2 and preference was given to perennial over annual plants where possible. Polymerase chain reactions (PCRs) were performed in a total volume of 13 µl with the following components: 15 ± 5 ng of genomic DNA, 0.4 µM of primer (Whitehead scientific, Cape town, RSA), 0.2 mM of each dNTP (Fermentas, Burlington, Canada), 1% DMSO (Sigma-Aldrich, St. Louis, USA), 0.15 U of Exsel High Fidelity DNA Polymerase (JMR Holdings, London, UK) and 1X PCR Exsel buffer. DMSO improves the yield and specificity of the PCR reaction (Chakrabarti & Schutt, 2001; Kang *et al.,* 2005).

PCR amplifications were performed in the Gene-AMP PCR System 9700 (Perkin Elmer, Coneticut, USA) with the following cycling conditions: an initial denaturation

step of 94 °C for 5 min, followed by 45 cycles denature (94 °C) for 1 min, annealing at 36 °C for 1 min and extension at 68 °C for 2 min, ending in a final elongation step for 10 min at 68 °C (Chen & Yamaguchi, 2002). PCR products were diluted 1:6 with loading dye (0.025% (w/v) bromophenol blue, 30% (v/v) glycerol) separated on 1.25% (w/v) agarose (Whitehead scientific, Cape town, RSA) gel containing 0.5 ng/ml ethidium bromide (EtBr), in 1X TAE buffer (0.04 M Tris-acetate, 1 mM EDTA, pH 8). A 1 kbp DNA ladder (Fermentas, Burlington, Canada) was used to determine the size of fragments. Samples were electrophoresed at 100 V in a Maxicell® EC360M Electrophoretic gel system (E-C Apparatus Corporation, St. Petersburg, Florida). Bands were visualised at 312 nm on a Spectroline TC-312A UV transilluminator (Spectronics Corporation, USA), as the interchelated EtBr emits fluorescence at a wavelength of 300 nm (Walker, 2000). Reactions were repeated 2 to 3 times to ensure that the fragments were consistent and reproducible.

Table 4.1: Cultivars selected based on field performance over several years

Traits					
Heliopeltis		**Cold**		**Drought**	
Susceptible	**Resistant**	**Susceptible**	**Tolerant**	**Susceptible**	**Tolerant**
B22	B2	PC105	LMB 7	PC105	PC165
B23	B4	PC108	PC123	PC108	PC168
CHAW4	B6	PC114	PC153	PC110	PC175
ELD3	B8	PC117	PC198	PC113	PC185
ELD11	B12	PC118	PC213	PC114	PC198
PC 1	B14	PC119	SFS 150	PC117	PC213
PC80	B15	PC131	SFS204	PC118	RC 4
PC108	PC153	86/17-2	88/6-7	PC119	RC 6

Table 4.2: A summary of the selected RAPD primers and the associated traits

RAPD	Primer	Sequence	Plant	Common name	Annual/ perennial	Linked to trait	Reference
RAPD1	UBC162	AAC TTA CCG C	*Cornus sericea*	Red osier dogwood	Perennial	Low temperature induce dormancy	Svendsen *et al.,* 2007
RAPD2	474	AGG CGG GAA C	*Brassica napus L.*	Seedrape	Annual	Winter survival	Asghari *et al.,* 2008
RAPD3	528	GGA TCT ATG C					
RAPD4	430	ATG CGG CAC C					
RAPD5	UBC218	CTC AGC CCA G	*Eucalyptus globulus*	Eucalyptus	Perennial	Freezing resistance	Fernandez *et al.,* 2006
RAPD6	UBC237	CGA CCA GAG C					
RAPD7	OPAH02	GAG ACC AGA C	*Camellia sinensis*	Tea	Perennial	Drought tolerance	Mishra & Sen-Mandi, 2004
RAPD8	Ope06	CCA CGG GAA C	*Lens culinaris Medikus*	Lentil	Annual	Anthracnose resistant	Tullu *et al.,* 2003
RAPD9	UBC704	GGA AGG AGG G					Taran *et al.,* 2003
RAPD10	P6-920	TCG GCG GTT C	*Triticum aestivum L.*	Wheat	Annual	Drought	Pakniyat & Tavakol, 2007
RAPD11	P7	CTG CAT CGT G					
RAPD12	OPAB19	ACA CCG ATG G	*Glycine max*	Soybean	Annual	*Phomopsis* resistance	Carvalho *et al.,* 2002
RAPD13	OPA10	GTG ATC CCA G	*Malus spp.*	Apple	Perennial	*Dysaphis devecta* resistance (insect)	Roche *et al.,* 1997
RAPD14	OPC08	TGG ACC GGT G					
RAPD15	OPT09	CAC CCC TGA G					
RAPD16	3-1	TTC ATA CGC G	*Triticum aestivum L.*	Wheat	Annual	*Mayetiola destructor* resistance (Hessian fly)	Dweikat *et al.,* 1997
RAPD17	5-1	CGC ATT TGC A					
RAPD18	6-1	GTT TCG CTC C					

4.4. Results and Discussion

Cultivars were selected based on field performance, listed in Table 4.1. RAPD primers were selected that were associated with specific traits in previous studies in literature, Table 4.2. It is expected that similar genes are responsible for the regulation and control of the same process in different plant species. Since QTLs are located on or near the gene that regulate the trait, markers were selected from studies that previously linked them to traits in other plant species.

Only one primer that was used in this study was previously identified to associate with a trait (drought) in *Camellia sinensis* (Mishra & Sen-Mandi, 2004). The other primers (17) were identified in other plant species, Table 4.2.

These 17 primers include six primers associated with cold tolerance of which one is linked to low temperature induced dormancy in red osier dogwood (Svendsen *et al.,* 2007), three to winter survival of seedrapes (Asghari *et al.,* 2006) and two to freezing resistance in eucalyptus (Fernandez *et al.,* 2006). Two primers were previously linked to drought tolerance in wheat (Paknivat & Tavakol, 2007). Six primers were linked to insect resistance, three in apples to the insect Dysaphis devecta (Roche *et al.,* 1997) and three in wheat to the Hessian fly (Dweikat *et al.,* 1997). The remaining three primers were linked to fungal resistance, one in soybean to Phomopsis (Carvalho *et al.,* 2002) and the other two to anthracnose in lentil (Tullu *et al.,* 2003; Taran *et al.,* 2003).

Of these 17 primers, six are from perennial plants and 11 from annual plants. The screening of the three traits with 17 primers (excluding RAPD7 because it was

previously identified in tea) resulted in the identification of one primer (RAPD8) that associate with drought resistance, present in the drought tolerant and absent in the drought susceptible cultivars, Table 4.2. Six of the 8 (75%) drought tolerant cultivars (PC175, PC185, PC198, PC213, RC4 and RC6) show a band at about 3,000 bp that was absent (0%) in the drought susceptible cultivars with RAPD8.

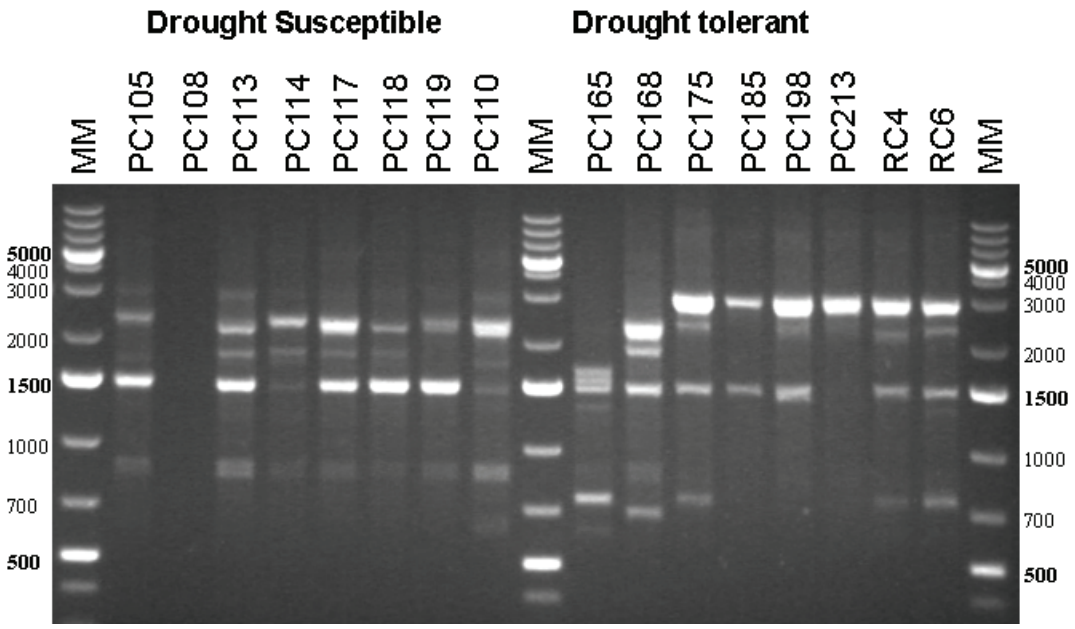


Figure 4.2: The agarose gel of RAPD8, showing the drought associated band at about 3,000 bp (MM = 1kb molecular marker, Fermentas)

The drought tolerant primer RAPD7, identified in tea by Mishra & Sen-Mandi (2004) shows a band at 1400 bp for 4 of the 8 (50%) drought tolerant cultivars (PC165, PC168, RC4 and RC6) and no band at 1400 bp was observed for the eight drought susceptible cultivars (0%), Figure 4.3.

In the study by Mishra & Sen-Mandi (2004) drought tolerant cultivars were first selected based on pot experiments and then on enzyme activity of cell protective enzymes. The enzymes, Cu-Zn cytosolic superoxide dismutase (Cu-Zn SOD) and ascorbate peroxidase (APX II) were extracted from water stressed leaves after which the activities were assayed. Cultivars that showed high levels of the enzymes and tolerance in the pot experiments were considered drought tolerant (4 cultivars out of the tested 10 cultivars). These four cultivars also showed a band at about 1400 bp with primer OPAH02, 100% (RAPD 7 in the current study). In the current study, cultivars were selected only on field performance and only 50% of the drought tolerant cultivars have a band at 1400 bp. The differences between the studies can be explained based on the selection procedure or the larger sample size of the current study (8 drought tolerant cultivars versus 4).

Using both of the fragments associated with drought ($RAPD7_{1400}$ and $RAPD8_{3000}$) it was possible to discriminate all the drought tolerant cultivars (100%) from the drought susceptible cultivars in this small sample set (16 cultivars, eight tolerant and eight susceptible to drought).

The absence of false positives (RAPD7 and RAPD8) in the drought susceptible cultivars with this small sample may be a fortunate coincidence. A larger sample size may result in a more accurate prediction of both the false positives and the false negatives. The majority of seedlings tested at the TRFCA are however classified as drought susceptible, while only a small number are drought tolerant. This implies that increasing the sample size for the drought tolerant plants may be difficult.

In this study, 17 primers (excluding primer RAPD7 because it was identified in tea previously) were tested for association with three traits. Only one association was found for drought tolerance (1/17, 5.9%), no association was found for Heliopeltis resistance and cold tolerance. Interestingly, not one of the primers associates with the same trait in tea than previously reported to associate with other plants. In literature, (Table 4.3) the primer association rate (A %) range from 0.16 to 10.0%. Therefore, the selection of primers that were previously identified for associations did not increase the A%.

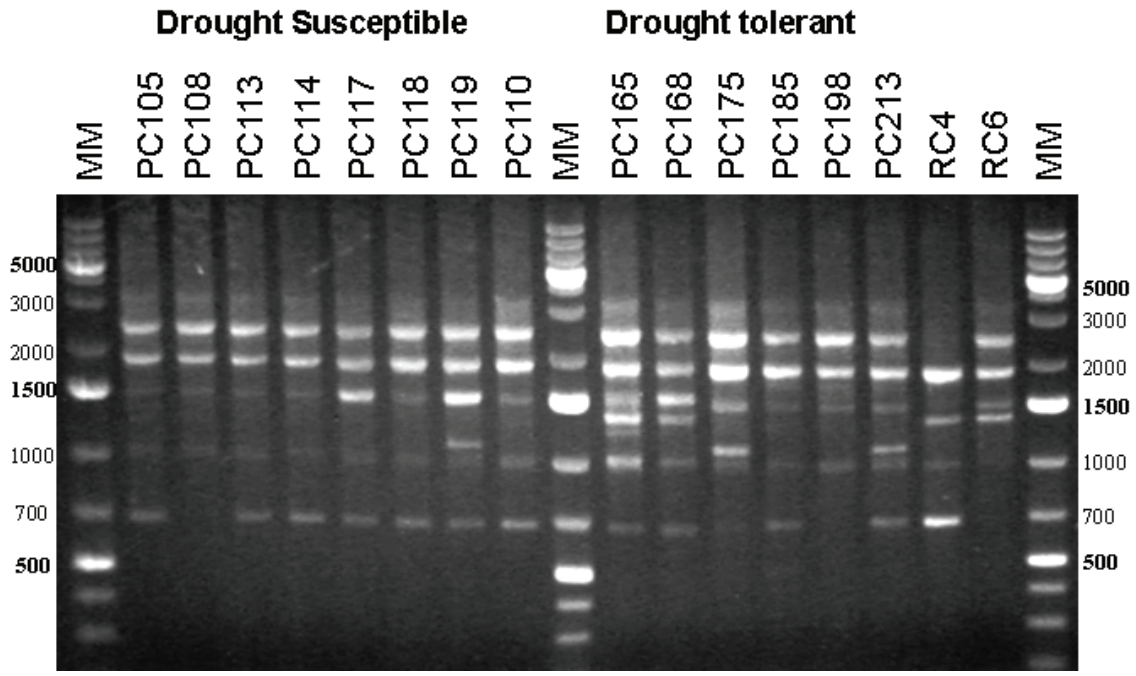


Figure 4.3: The agarose gel of RAPD7, showing the drought associated band at about 1400 bp (MM = 1kb molecular marker, Fermentas)

In addition, when comparing trait specific results with other studies, (Table 4.3) it is interesting to note that the A% of the trait drought (6.67 to 9.09% and 5.9% in the current study) is high compared to the other traits 0.16 to 1.13%, with the exception

of freeze resistance. It is expected that drought and cold tolerance are regulated by more than one gene (Parry *et al.,* 2005). Since trait association with RAPD markers are based on coincidence, the probability of finding a trait associated marker should be higher in multi gene regulated traits than in single gene regulated traits. This may explain why only one trait-associated marker was identified. In order to identify markers for all traits more primers should be tested (at least 600 primers per trait).

Table 4.3: Summary of RAPD trait associated studies

Study	Plant (Trait)	Trait	Nr[1]	NOA[2]	A%[3]
Mishra & Sen-Mandi, 2004	*Camellia sinensis*	Drought	11	1	9.09
Pakniyat & Tavakol, 2007	*Triticum aestivum* L.	Drought	30	2	6.67
Fernandez *et al.,* 2006	*Eucalyptus globules*	Freezing resistance	20	2	10.0
Svendsen *et al.,* 2007	*Cornus sericea*	Low temperature induce dormancy	515	1	0.19
Asghari *et al.,* 2007	*Brassica napus*	Winter survival	250	1	0.40
Tullu *et al.,* 2003	*Lens culinaris medikus*	Anthracnose	700	6	0.86
Carvalho *et al.,* 2000	*Glycine max*	Phomopsis	600	1	0.16
Dweikat *et al.,* 1997	*Triticum aestivum* L.	Hessian fly	1600	18	1.13

[1] Nr: Number of primers

[2] NOA: Number of primers associated with trait

[3] A% Primers associated rate

RAPDs are a simple, inexpensive approach for the identification of trait-associated markers. However, due to the short primers used in this technique the repeatability and reproducibility is low. In addition, RAPDs are a dominant technique. The major

advantage of this technique is that it requires no prior information or expensive development. RAPD have been used in several food crops to identify traits (review by Salem *et al.,* 2007) and provide a fast, cost effective way to identify molecular markers.

4.5. Conclusion and Future Perspectives

In this study, 17 primers (previously identified in other crops) were tested for association with drought tolerance, Heliopeltis resistance and cold tolerance in tea. RAPD8 was identified as a possible drought resistance marker. This primer shows a fragment at 3,000 bp in 6 of 8 drought tolerant cultivars and no false positives (no fragment at 3,000 bp in the drought susceptible cultivars). No markers were identified for Heliopeltis resistance or cold tolerance.

The possible drought marker identified by Mishra & Sen-Mandi (2004) was confirmed, the band at 1400 bp was present in 4 out of 8 (50%) drought tolerant cultivars. This rejects hypothesis 1 (supporting null hypothesis 1) that the fragment will be present in all drought tolerant cultivars as in the experiment by Mishra & Sen-Mandi (2004). Interestingly, no primer was linked to the same trait that it was associated with in other plants and the selection did not improve the efficiency. This rejects hypothesis 2 that the selection of trait-associated primers will improve the frequency of association (supporting null hypothesis 2).

For future work the trait-associated fragments should be isolated from the agarose gel and cloned into a cloning vector, from which it can be sequenced or the fragment can be sequenced directly (Hernandez *et al.,* 1999). The sequence can be subjected

to a BLASTn search against the GenBank to find the possible gene involved in drought tolerance. Drought tolerance is a complicated trait that is governed by more than one mechanism. The possible involvement of these fragments in tolerance to drought should be investigated for a better understanding of the association of the fragment to drought tolerance.

From the sequence data, SCAR primers can be designed to detect drought resistance in tea cultivars. Since numerous genes are involved in drought tolerance, the application of only one SCAR marker alone would be limited. Research should focus on the development of more markers for drought tolerance so that they can be applied to tea breeding with success.

In conclusion, droughts result in considerable losses for the agricultural community. The predicted global climate changes will lead to an annual temperature rise that may lead to unpredictable weather conditions (Svendsen *et al.,* 2007). In this light, the identification of drought tolerance in plants becomes increasingly important. This emphasizes the importance of the development of genetic markers for drought tolerance in tea to help the plant breeding program of the TRFCA.

CHAPTER 5

CONCLUDING DISCUSSION: TEA BREEDING AND GENETIC STUDIES

"It is not the strongest of the species that survive,
nor the most intelligent,
but the one most responsive to change."
- Charles Darwin

Camellia sinensis is a hardwood tree indigenous to China and India (Paul *et al.,* 1997). This tree's leaves are the commercial source of tea, the most consumed caffeine beverage in the world. More than 3 billion cups of tea are consumed every day, adding up to more than 45.4 L (12 gallons) per person annually (Wolf *et al.,* 2007).

Tea breeding programs

Tea is an outbreeding tree species, with a long reproductive cycle (three to eight years). Therefore, tea breeding is a time consuming process, it takes approximately 15 to 20 years for the release of a cultivar. Breeding programs generally focus on the improvement of quality, yield and biotic and abiotic stress tolerance. Quality parameters include theanine, catechins and caffeine that contribute to the taste of tea. Characteristics such as shoot density, leaf size and leaf angle influence the yield. Both yield and quality have a direct effect on the income of a tea estate. Abiotic stress drought and temperature has an immense effect on plant production every year. Therefore breeding programs aim at increasing drought and temperature tolerance. Resistance to biotic stress, pests and pathogens, also play a critical role in breeding.

The replacement of seedling plantations with clonal plantations has lead to an increase of the yield and quality of the tea produced. However, it may also lead to a decrease in genetic diversity. It is estimated that 85% of the 90% of the clonal tea plantations in Japan are based on a single clone, namely Yabukita (Kaundun *et al.,* 2000). The clonal plantations have increased to about 50% in India, 33% (2006) in China, about 57% in Sri Lanka and about 59% in Kenya (Ni *et al.,* 2008). The small

number of parental plants in most breeding programs may also contribute to a decrease in diversity. In India, over 75% of the cultivars released in Tocklai are derived from Betja, Cinnamara or Cambod stocks while 60% of the cultivars released in Kenya are derived from “6/8” cultivar (Yao *et al.,* 2008). The breeding program of the TRFCA is no exception to this, about 63.8% of the cultivars release from the TRFCA is derived from PC1. Therefore, breeding programs should consider the possible decrease in genetic diversity in the management of the programs.

The TRFCA started a tea breeding program in 1956, aiming at increasing the yield and quantity of tea production in Africa (Ellis & Nyirenda, 1995). The breeding program succeeded in steadily increasing the yield and quality of the released cultivars over the first 40 years. Recently however, the progress seems to plateau off. This may indicate that, the full potential of the parental plants has been reached. This program started with a small group of selected parents and the introduction of new genetic material to the program was limited to superior offspring. It is therefore, believed that expanding the gene pool by the introduction of new genetic material may solve this problem.

In the molecular age the world is in currently, it has become increasingly important for breeding programs to consider new approaches. Possibilities provided by biotechnology and breeding opportunities like MAS should therefore be exploited. The tea research in this regard is limited and the identification of QTLs and MAS markers that may contribute to the breeding of tea.

This study aimed at the evaluation of the genetic diversity among the seedling accessions and TRFCA cultivars. Cultivars with China origin were also included to analyse the diversity on a more comprehensive scale. The diversity was analysed using both morphological and molecular techniques. The association of RAPD markers with traits of interests was also investigated. Finally, it is believed that the knowledge gained by the study may assist in the improvement of the TRFCA's breeding program.

Morphological analysis

The morphological data of six traits were collected from 186 seedling accessions and six TRFCA cultivars. The data was analysed using two scoring methods, range and mean based. The range scoring method could discriminate between all the accessions in the analysis. The H_e for the range scoring method was 0.7946. The mean scoring method on the other hand was unable to discriminate between all the accessions and had a H_e value of 0.6900.

Of the two scoring methods, the mean scoring method was in better agreement to the visual classification of the expert tea breeder, Dr. H.E. Nyirenda. This suggests that the mean scoring method is superior to the range scoring method (Chapter 2, Figure 2.4 and Figure 2.5). The TRFCA cultivars seem to be scattered among the seedling accessions. Therefore, the TRFCA cultivars may be well distributed among the accessions from Malawi. It was unfortunate that only six TRFCA cultivars were included in the morphological data set. Due to the small number of TRFCA cultivars, the morphological data analysis was inconclusive.

Chen *et al.* (2005b) used allozymes, biochemical and morphological traits as markers to analysis the diversity among tea subspecies and reported H_e values that range from 0.114 to 0.218 (55 morphological traits, four biochemical parameters, 10 allozymes, 87 accessions). This is lower than the H_e values in the current study, which may be a result of the combination of the different markers used in the study by Chen *et al.* (2005b) since allozymes reveal very low levels of diversity. However, the larger effect can be contributed to the small number of traits evaluated in the current study. Also, the traits investigated all revealed high levels of diversity, which is not an accurate reflection of most morphological traits.

Future work should focus on the analysis of more morphological traits (Table 2.2, Chapter 2) to increase the effectiveness of the study. Also, a more representative sample set of the TRFCA cultivars should be included. This should preferably include all of the 41 released cultivars from the TRFCA.

Microsatellite analysis

The genetic diversity of 41 TRFCA cultivars, 233 seedling accessions (Cambod type (92), Assam type (29), China type (82) and the hybrids Cambod/China (16) and Cambod/Assam (14)) and 22 China origin cultivars were analysed using seven microsatellite primers results. Two of the seven results SSR primers were rejected due to their high error rate. The results from the remaining five SSR primers were used for population analysis. Not many studies consider the effect of error rate on the results, it is however important to know the reliability of the results and conclusions. In this study, the error rate was calculated and the high error rate loci were excluded from the study to increase the reliability of the study.

The H_o for this study ranges from 0.526 to 0.741. In the microsatellite study by Kaundun and Matsumoto (2002), the H_o ranges from 0.278 to 0.698, this is lower than the values from the current study. However, Kaundun and Matsumoto only analysed 24 cultivars and the number of alleles per locus range from 2 to 5, whereas the number of alleles per locus in this study ranged from 17 to 28 possibly explaining the differences.

The level of H_s in this study ranges from 0.673 to 0.904, this is higher than the range of 0.460 to 0.799 reported by Ohsako *et al.* (2008). Ohsako *et al.* assed the diversity in tea from Japan with microsatellites markers. Other studies also found lower levels of diversity within Japan and it is hypothesised that this may be due to long intensive selection programs in this country (Kaundun & Park, 2002; Matsumoto *et al.,* 2002; Matsumoto *et al.,* 2004; Ni *et al*., 2008).

The genetic relationship illustrated in Chapter 3, Figure 3.2 showed a close cluster of China origin cultivars. The TRFCA cultivars were also clustered together. The F_{ST} values indicate significant differences between the TRFCA cultivars and all the subpopulations. The F_{IS} value of 0.183 is evidence of a moderate inbreeding level within the TRFCA cultivars. However, the level of inbreeding within the TRFCA cultivars was not higher than that of either the seedling accessions or the subpopulations. Considering the level of inbreeding and the significant differences between the subpopulations and the TRFCA cultivars, the breeding program may benefit from the introduction of new genetic material.

The F_{ST} values of the China origin cultivars show significant differences with all other subpopulations. The inbreeding within the China origin cultivars is lower than that of the other subpopulation. This is expected since organisms are more divers at their centre of origin (Kaundun & Matsumoto, 2003a; Ni *et al.,* 2008).

The genetic relationships between the populations indicate TRFCA cultivars are closest to Cambod/Assam and Assam type and furthest apart from the China origin cultivars. Considering the lower level of inbreeding of the China origin cultivars, along with the higher level in the TRFCA cultivars and the fact that the level of inbreeding within the TRFCA is within the range of the seedling accessions, the inbreeding within tea from Malawi is higher than within tea from China cultivars.

Comparison between the morphological and microsatellite analysis

When comparing the morphological data with the microsatellite data (Electronic Appendix G shows the relationship among the same accessions as evaluated in Chapter 2 with the microsatellite data), the discrimination power of the five microsatellite loci is superior compared to the six morphological traits. The microsatellite analysis is in better agreement with the subpopulations classified by tea expert Dr. H.E. Nyirenda. However, the six TRFCA cultivars do not cluster together with either the morphological or the microsatellite analysis.

Future work should include analysis of the TRFCA cultivars on a more comprehensive scale. This would entail the inclusion of more morphological traits as well as all released TRFCA cultivars.

Possible trait associated marker

A total of 17 RAPD primers that were previously identified to associate with traits of interests in other crops were tested for association with drought tolerance, Heliopeltis resistance and cold tolerance in tea. RAPD8 was identified as a possible drought tolerance marker. This primer shows a fragment at about 3,000 bp in 6 of 8 drought tolerant cultivars (75%). Also, no false positives were observed (no fragment at 3,000 bp in the drought susceptible cultivars). No markers were identified for Heliopeltis resistance or cold tolerance.

The RAPD primer identified by Mishra and Sen-Mandi (2004) for its association with drought in tea was also tested (1400 bp band present in drought tolerant cultivars). The association was found in 4 out of 8 (50%) drought tolerant cultivars it was however found in all the drought tolerant cultivars by Mishra and Sen-Mandi (4 out of 4 drought tolerant cultivars). The difference may be due to the method of selection of drought tolerant cultivars by Mishra and Sen-Mandi (2004).

Using both the primers (RAPD8 and the primer identified by Mishra and Sen-Mandi, 2004) all the drought tolerant cultivars could be distinguished from the susceptible cultivars and no false positives were detected.

Interestingly, none of the primers were linked to the same trait that the primer was associated with in other plants. The selection of primers from previous studies did therefore not improve the frequency of association.

Markers as the one identified by this study should be further developed in order to exploit the full potential of the discovery. The gene implicated in the trait should be identified. RAPD markers have low annealing temperatures and are therefore likely to have low repeatability and reproducibility. Therefore, by sequencing the associated fragment the primer length can be increased. This marker is now called a SCAR marker and will have a higher annealing temperature and consequently a higher selectivity, reproducibility and repeatability. It can be developed into a QTL by mapping the primer's chromosomal position. Alternatively, it can be developed into a MAS marker and assist in future tea breeding.

Future work could also include the testing of more RAPD primers to identify trait-associated markers for economically important traits. Other studies suggested that up to 600 primers should to be tested (see Chapter 4, Table 4.3) to identify a marker for each trait. Also more than one marker per trait will improve the selection efficiency.

The implication on breeding

Tea is a major commodity for developing countries where 95% of the world's tea is produced (http://faostat.fao.org/site/567/default.aspx#ancor). The tea breeding community aspire to control and improve tea. Therefore, studies like this one that contribute to the knowledge of the genetic diversity can assist tea breeders. Accessions from other countries could be included to assess tea on a global scale. Preferably, the number of loci accessed should also be increased for a more accurate quantification of the diversity among the accessions. This will allow better management of the diversity within tea as a commodity.

The current study assessed the genetic diversity within the cultivars from the TRFCA in comparison to Malawian seedling accessions and China origin cultivars. The plateau that has been reached with regard to the progress in yield and quality could not completely be explained by inbreeding. Since only moderate levels of inbreeding were observed within the TRFCA cultivars. However, it remains important to preserve the genetic diversity within a population. Therefore the introduction of new parents into the program may have a positive effect. Charles Darwin said, *"It is not the strongest of the species that survive, nor the most intelligent, but the one most responsive to change."* The higher the level of diversity within a population the better the probability to survive/adapt to changes. In order to increase the genetic diversity, the introduction of new plants into the program is required.

This study may aid in the selection of new genetically distant parental plants for inclusion into the breeding program. This introduction of these plants may help with the progress of the program. However, it does not guarantee a solution to the problem with regards to an increase in yield. It could lead to increase in the heterozygosity within the TRFCA cultivars, which could lower the risk of extinction due to catastrophes such as environmental changes and diseases.

The development of markers to detect traits of interests early on in the breeding program could increase the efficiency of the breeding programs (see Chapter 1, Figure 1.13). Therefore, research focusing on this may have great benefit for the tea breeding community.

A desirable trait that should be investigated is drought tolerance since, droughts result in considerable losses for the agricultural community. The predicted global climate changes will lead to annual temperature rise that may lead to unpredictable weather conditions (Svendsen *et al.,* 2007). In this light, the identification of drought tolerance in plants becomes increasingly important. This emphasizes the importance of the development of genetic markers for drought tolerance in tea to help the plant breeding program of the TRFCA.

In conclusion, this research is the first attempt to characterize the genetic diversity found within tea from Malawi. Tea breeders in Malawi will benefit from the knowledge of the genetic relationship among the tea cultivars. This research attempts to add to the body of knowledge on the genetic diversity in tea and the identification of traits of interest.

Summary

Tea is made from the young leaves of *Camellia sinensis*, an evergreen, outbreeding tree species. The replacing of seedling with clonal plantations and the use of small numbers of parental plants in breeding programs could cause a decrease in the genetic diversity. The Tea Research Foundation of Central Africa (TRFCA), in Malawi, started tea breeding in 1956. This program has provided new cultivars with increased quality and yield over the first 40 years. However, recently the progress seems to have reached a plateau, suggesting that the maximum potential of the original genetic base has been reached.

This study aimed at the evaluation of the genetic diversity within TRFCA cultivars, Malawian seedling accessions and cultivars currently used in China using morphological and microsatellite markers.

In the morphological trait analysis, 186 seedling accessions and six TRFCA cultivars were accessed using six morphological traits. The morphological data were scored using two scoring methods, the range scoring method and the mean scoring method. Although the mean scoring method could not discriminate between all the accessions, it is in better agreement with the classification of an expert tea breeder. The six TRFCA cultivars were well distributed among the seedling accessions from Malawi. However, the results of the morphological study were not conclusive since the sample size of the TRFCA cultivars was too small.

Microsatellite analysis of 41 TRFCA cultivars, 233 seedling accessions and 22 China origin cultivars revealed only moderate levels of inbreeding. The genetic relationship

shows the TRFCA cultivars and the China origin cultivars in their own clusters. The F_{ST} values indicated significant differences between the TRFCA and China origin cultivars with all other groups. The allelic richness and the heterozygosity of the TRFCA cultivars are on the low end of the ranges of the groups studied.

The TRFCA cultivars may benefit from the introduction of genetically distant tea plants into the breeding program since the level of heterozygosity within this group is relatively low.

A second aim of this study is the identification of Random Amplified Polymorphic DNA (RAPD) markers that associate with traits of interest. Primers (17) were selected from previously trait associate plant studies. This selection did not increase the association rate of the primers. The traits that were investigated include Heliopeltis resistance, drought and cold tolerance. One of the tested primers RAPD8, associates with 6 of 8 drought tolerant cultivars. Another RAPD primer previously identified by Mishra and Sen-Mandi (2004) associated with only 4 of 8 drought tolerant cultivars in this study. Neither of the two primers associated with the 8 drought susceptible cultivars. Together these two primers could discriminate all 8 drought tolerant cultivars from the 8 susceptible cultivars and could possibly be used to identify drought tolerant plants.

This study contributes to the knowledge of the genetic diversity of tea which may contribute to the management of genetic diversity within tea. In addition a possible trait associated RAPD marker for drought was identified that could be used in the selection for drought tolerant cultivars.

Samevatting

Tee word van die jong blare van die *Camellia sinensis* plant gemaak. *Camellia sinensis* is ‘n immergroen boom spesie. Tans word die meeste van die saailing plantasies met klone plantasies vervang wat ‘n negatiewe uitwerking op die genetiese diversiteit van tee het. Tee ontwikkelingsprogramme beïnvloed ook die diversiteit negatief omdat ‘n klein hoeveelheid ouer plante gebruik word. Die Tee Navorsings van Sentrale Afrika (TRFCA) is in Malawi geleë en spesialiseer vanaf 1956 in tee kultivar ontwikkeling. Die kwaliteit en die opbrengs van die vrygestelde kultivars het oor die eerste 40 jaar van die program verbeter. Die vordering in hierdie program is besig om ‘n plato te bereik. ‘n Moontlike verduideliking hiervoor kan wees dat die maksimum kapasitieit van die bestaande genetiese basis bereik is.

Die eerste doelwit van hierdie studie is om die genetiese diversitiet van die TRFCA kultivars, geselekteerde saailinge van Malawi en kulticars wat tans in China gekweek word te evalueer deur morfologiese en mikrosateliet merkers te gebruik.

In die morfologiese evaluasie is 186 geselekteerde saailinge en ses TRFCA kultivars met behulp van ses morfologiese karaktereineskappe geëvalueer. Die data is geëvalueer deur die gebruik van twee puntestelsels, ’n reeks puntestelsels en ‘n gemiddelde puntestelsel. Alhoewel die gemiddelde puntestelsel nie tussen al die plante kon onderskei nie, het dit ‘n beter groepering ten opsigte van sub-spesies (H_e was 0.6900) gegee. Die ses TRFCA kultivars is tussen die saailinge van Malawi versprei. Weens die klein hoeveelheid TRFCA kultivars is die gevolgtrekings van die morphologiese evaluasie nie betekenisvol nie.

Mikrosateliet evaluasie is gedoen met 41 TRFCA kultivars, 233 geselekteerde saailinge en 22 Chinese kultivars. Gemiddelde vlakke van inteling is opgemerk. In die illustrasie van die genetiese verhouding tussen die plante is die TRFCA kultivars gegroepeer, so ook die Chinese kultivars. Die F_{ST} waardes toon betekenisvolle statisitiese verskille tussen TRFCA en die Chinese oorsprong kultivars en al die ander groepe. Die alleliese rykheid van die TRFCA kultivars was aan die onderpunt van die reeks.

Hierdie resultate dui daarop dat die TRFCA ontwikkellingsprogram moontlik voordeel kan trek uit die byvoeging van ouer pante wat geneties divers is in die program. Dit mag ‘n positiewe uitwerking hê as gevolg van die ondergemiddelde vlak van die “heterozygosity” in die TRFCA kultivars.

‘n Tweede doelwit van die studie was die identifisering van moontlike “Random Amplified Polymorphic DNA” (RAPD) merkers wat met belangrike kenmerke geassosieer kan word. RAPD merkers (17) wat geïdentifiseer is in vorige kenmerk-assosiasiestudies in ander gewassse is geselekteer. Die seleksie van hierdie merkers het nie die assosiasie persentasie verhoog nie. Die kenmerkke wat ondersoek is, was Heliopeltis weerstand, droogte toleransie en koue toleransie. Een van die merkers, RAPD8 assosieer met 6 van die 8 droogte toleransie kultivars. Die RAPD merker wat deur Mishra and Sen-Mandi (2004) geïdentifiseer is, assosieer met 4 van die 8 droogte toleransie kultivars. Beide die merkers assosieer met geen van die droogte sensitiewe kultivars nie. Gesamentlik kon die twee RAPD merkers al 8 van die droogte toleransie kultivars van die sensitiewe kultivars onderskei.

Hierdie studie dra by tot die kennis van die genetiese diversitiet van tee wat kan lei to die beter bestuur wan die tee ontwikkelingsprogram. Veder is ‘n moontlike merker vir droogte toleransie geïdentifiseer wat gebruik kan word in die seleksie van moontlike droogte bestande kultivars.

Reference

Astill, C., Birch, M.R., Dacombe, C., Humphrey, P.G., Martin, P.T. (2001). Factors affecting the caffeine and polyphenol contents of black and green tea infusions. Journal of Agricultural and Food Chemistry. 49: 5340-5347.

Aydagn, Y.B. (2005). Genetic analysis of traditional Ethiopian highland maize (*Zea mays L.*) using molecular markers and morphological traits: implication for breeding and conservation. PhD. Thesis. University of Pretoria. Plant breeding/Genetics. 10-50.

Ashihara, H., Sano, H., Crozier, A. (2008). Caffeine and related purine alkaloids: Biosynthesis, catabolism, function and genetic engineering. Phytochemistry. 69: 841-856.

Asghari, A., Mohammadi, S.A., Moghaddam, M., Shokuhian, A.A. (2008). Identification of SSR and RAPD markers associated with QTLs of winter survival and related traits in *Brassica napus* L. Africa Journal of Biotechnology. 7 (7): 897-903.

Babu, R.C., Nguyen, B.D., Chamarerk, V., Shanmugasundaram, P., Chezhian, P., Jeyaprakash, P., Ganesh, S.K., Palchamy, A., Sadasivam, S., Sarkarung, S., Wade, l.J., Nguyen, H.T. (2003). Genetic analysis of drought resistance in rice by molecular markers: Association between secondary traits and field performance. Crop Science. 43: 1457–1469.

Balasaravanan, T., Pius, P.K., Kumar, R.R., Muraleedharan, N., Shasany, A.K. (2003). Genetic diversity among south Indian tea germplasm (*Camellia sinensis*, *C. assamica and C. assamica spp. lasiocalyx*) using AFLP markers. Plant Science. 165: 365-372.

Barbara, T., Palma-Silva, C., Paggi, G.M., Bered, F., Fay, M.F., Lexer, C. (2007). Cross-species transfer of nuclear microsatellite markers: potential and limitations. Molecular Ecology. 16: 3759–3767.

Bhattacharya, N., Seth, S., Tudu, B., Tamuly, P., Jana, A., Ghosh, D., Bandyopadhyay, R., Bhuyan, M., Sabhapandit, S. (2007). Detection of optimum fermentation time for black tea manufacturing using electronic nose. Sensors and Actuators B. 122: 627–634.

Bhuyan, M., Bhattacharya, P.R. (2006). Feeding and oviposition preference of *Helopeltis theivora* (*Hemiptera: Miridae*) on tea in Northeast India. Insect Science. 13: 485-488.

Blears, M.J., De Grandis, S.A., Lee, H., Trevors, J.T. (1998). Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. Journal of Industrial Microbiology & Biotechnology. 21: 99–114.

Bonin, A., Bellemain, E., Eidesen, P.B., Pompanon, F., Brochmann, C., Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. Molecular Ecology. 13: 3261-3273.

Bonin, A., Ehrich, D., Manel, S. (2007). Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. Molecular Ecology. 16: 3737–3758.

Bouck, A., Vision, T. (2007). The molecular ecologist's guide to expressed sequence tags. Molecular Ecology. 16: 907–924.

Bruce, W.B., Edmeades, G.O., Barker, T.C (2002). Molecular and physiological approaches to maize improvement for drought tolerance. Journal of Experimental Botany. 53 (366): 13-25.

Buschiazzo, E., Gemmell, N.J. (2006). The rise, fall and renaissance of microsatellites in eukaryotic genomes. BioEssays. 28: 1040-1050.

Bussell, J.D., Waycott, M., Chappill, J.A. (2005). Arbitrarily amplified DNA markers as characters for phylogenetic inference. Perspectives in Plant Ecology, Evolution and Systematics. 7: 3–26.

Carvalho, G.A., Sediyama, T., Alzate-Marin, A. L., Barros, E.G., Moreira, M.A. (2002). Identification of RAPD markers linked to a soybean stem canker resistance gene. Fitopatologia Brasileira 27: 474-478.

Chakrabarti, R., Schutt, C.E. (2001). The enhancement of PCR amplification by low molecular-weight sulfones. Gene. 274: 293–298.

Chen, L., Yamaguchi, S. (2002). Genetic diversity and phylogeny of tea plant (*Camellia sinensis*) and its related species and varieties in the section Thea genus *Camellia* determined by randomly amplied polymorphic DNA analysis. Journal of Horticultural Science and Biotechnology. 77 (6): 729-73.

Chen, J., Wang, P., Xia, Y., Xu, M., Pei, S. (2005a). Genetic diversity and differentiation of *Camellia sinensis* L. (cultivated tea) and its relatives in Yunnan province of China, revealed by morphology biochemistry and allozyme studies. Genetic Resources and Crop Evolution. 5: 41-52.

Chen, L., Gao, Q.K., Chen, D.M., Xu, C.J. (2005b). The use of RAPD markers for detecting genetic diversity, relationship and molecular identification of Chinese elite tea genetic resources [*Camellia sinensis (L.) O. Kuntze*] preserved in a tea germplasm repository. Biodiversity and Conservation. 14: 1433–1444.

Chen, L., Yamaguchi, S. (2005c). RAPD markers for discriminating tea germplasms at the inter-specific level in China. Plant Breeding. 124: 404-409.

Chen, L., Zhao, L.P., Gao, Q.K. (2005d). Generation and analysis of expressed sequence tags from the tender shoots cDNA library of tea plant (*Camellia sinensis*). Plant Science 168: 359–363.

Chen, L., Zhou, Z.X. (2005e). Variations of main quality components of tea genetic resources [*Camellia sinensis* (L.) O. Kuntze] preserved in the China National Germplasm Tea Repository. Plant Foods for Human Nutrition. 60: 31–35.

Chen, L., Zhou, Z.X., Yang, Y.J. (2007a). Genetic improvement and breeding of tea plant (*Camellia sinensis*) in China: from individual selection to hybridization and molecular breeding. Euphytica. 154: 239-248.

Chen, P.C., Chang, F.S., Chen, I.Z., Lu, F.M., Cheng, T.J., Chen, R.L.C. (2007b). Redox potential of tea infusion as an index for the degree of fermentation. Analytica Chimica Acta. 594: 32–36.

Chen, L., Yu, F.L., Yao, M.Z., Lu, B., Yang, K., Du, Y.Y., (2008), Preparation of the UPOV guidelines for the conduct of tests for distinctness, uniformity and stability-Tea plant [*Camellia sinensis* (L.) O. Kuntze]. Agricultural Sciences in China. 7 (2): 224-231.

Chistiakov, D.A., Hellemans, B., Volckaert, F.A.M. (2006). Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. Aquaculture. 255: 1–29.

Dan, V.T. (2006). Assessing genetic diversity in Vietnam tea [*Camellia sinensis* (L.) O. Kuntze] using morphology, inter-simple sequence repeat (ISSR) and microsatellite (SSR) markers. Ph.D. thesis. George-August University Göttingen, Germany. 2-26, 94-119.

Deng, W.W., Ogita, S., Ashihara, H. (2008). Biosynthesis of theanine (γ-ethylamino-L-glutamic acid) in seedlings of *Camellia sinensis*. Photochemistry Letters 1: 115–119.

Devarumath, R.M., Nandy, S., Rani, V., Marimuthu, S., Muraleedharan, N., Raina, S.N. (2002). RAPD, ISSR and RFLP fingerprints as useful markers to evaluate genetic integrity of micropropagated plants of three diploid and triploid elite tea clones representing *Camellia sinensis* (China type) and C. *assamica* ssp. *assamica* (Assam-India type). Plant Cell Reports. 21: 166–173.

De Vienne, D. (2003). In Molecular markers in plant genetics and biotechnology. Science Publishers, Inc. United State of America. ISBN 1-57808-239-0. Introduction and Chapter 1: Principal sources of molecular markers. 1-40.

Du, Y.Y., Liang, Y.R., Wang, H., Wang, K.R., Lu, J.L., Zhang, G.H.W., Lin, P., Li, M., Fang, Q.Y. (2006). A study on the chemical composition of albino tea cultivars. Journal of Horticultural Science & Biotechnology. 81(5): 809–812.

Dweikat, I., Ohm, H., Patterson, F., Cambron, S. (1997). Identification of RAPD markers for 11 Hessian fly resistance genes in wheat. Theoretical applied genetics. 94: 419-423.

Ellis, R., Nyirenda, H.E. (1995). A Successful plant improvement programme on Tea (*Camellia sinensis*). Experimental Agriculture. 31: 307-323.

Fairbanks, D.J., Andersen, W.R. (1999) In: Genetics: the continuity of life. ISBN: 0-534-25272-9 Chapter 19: Population genetics. 580-590.

Fang, W.P., Jlang, C.J., Yu, M., Ye, A.H., Wang, Z.X. (2006). Differentially expression of *Tua1*, a Tubulin-encoding gene, during flowering of tea plant *Camellia sinensis* (L.) O. Kuntze using cDNA Amplified Fragment Length Polymorphism technique. Acta Biochimica et Biophysica Sinica. 38 (9): 653–662.

Fernandez, M., Valenzuela, S., Balocchi, C. (2006). RAPD and freezing resistance in *Eucalyptus globulus*. Electronic Journal of Biotechnology. 9 (3): 303-309.

Ferruzzi, M.G., Green, R.J. (2006). Analysis of catechins from milk–tea beverages by enzyme assisted extraction followed by high performance liquid chromatography. Food Chemistry. 99: 484–491.

Frankham, R. (2005). Genetics and extinction. Biological conservation. 126: 131-140.

Francia, E., Tacconi, G., Crosatti, C., Barabaschi, D., Bulagarelli, D., Dall'Aglio, E., Vale, G. (2005). Marker assisted selection in crop plants. Plant Cell, Tissue and Organ Culture. 82: 317-342.

Freeman, S., West, J., James, C., Lea, V., Mayes, S. (2004). Isolation and characterization of highly polymorphic microsatellites in tea (*Camellia sinensis).* Molecular Ecology Notes. 4: 324–326.

Frewen, B.E., Chen, T.H.H., Howe, G.T., Davis, J., Rohde, A., Boerjan, W., Bradshaw, H.D. (2000). Quantitative Trait Loci and candidate gene mapping of bud set and bud flush in *Populus*. Genetics. 154: 837–845.

Friedman, M. (2007). Overview of antibacterial, antitoxin, antiviral, and antifungal activities of tea flavonoids and teas. Molecular Nutrition & Food Research. 51: 116 – 134.

Gostimsky, S.A., Kokaeva, Z.G., Konovalov, F.A. (2005). Studying plant genome variation using molecular markers. Russian Journal of Genetics. 41(4): 378–388.

Grigg, D. (2002).The worlds of tea and coffee: Patterns of consumption. GeoJournal. 57: 283–294.

Goudet, J. (2001). FSTAT, a program to estimate and test gene diversities and fixation indices (ver. 2.9.3). Available from http://www.unil.ch/izea/software/fstat.html

Gul, S., Ahmad, H., Khan, I.A., Alam, M. (2007). Assessment of genetic diversity in tea genotypes through RAPD primers. Pakistan Journal of Biological Sciences. 10 (15): 2609-2611.

Gunasekare, M.T.K. (2007). Applications of molecular markers to the genetic improvement of Camellia sinensis L. (tea) – A review. Journal of Horticultural Science & Biotechnology 82 (2): 161–169.

Hackett, C.A., Wachira, F.N., Paul, S., Powell, W., Waugh, R. (2000). Construction of a genetic linkage map for Camellia sinensis (tea). Heredity. 85: 346-355.

Hall, G.H. (2008). In Phylogenetic trees made easy: A how to manual. 3rd Edition. ISBN 978-0-87893-310-5. Chapter 4: Neighbor-Joining Trees. 63-88.

Haslam, E. (2003). Thoughts on thearubigins. Phytochemistry. 64: 61–73.

Hernandez, P., Martin, A., Dorado, G. (1999). Development of SCARs by direct sequencing of RAPD products: a practical tool for the introgression and marker-assisted selection of wheat. Molecular breeding. 5: 245-253.

Huelsenbeck, J., Ronquist, F., Nielsen, R., Bollback, J.P. (2001). Bayseian inference of phylogeny and its impact on evolutionary biology. Science. 294: 2310-231.

Hung, C.Y., Wang, Y.H., Huang, C.C., Gong, X., Ge, X.J., Chiang, T.J. (2008). Isolation and characterization of 11 microsatellite loci from Camellia sinensis in Taiwan using PCR-based isolation of microsatellite arrays (PIMA). Conservation Genetics. 9: 779–781.

Huson, D.H., Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution. 23 (2): 254–267.

IPGRI. (1997). Descriptors for tea (*Camellia sinensis*). International Plant Genetic Resources Institute, Rome, Italy. (URL: http://www.cgiar.org/ipgri/).

Jiang, D., Zhong, G.Y., Hong, Q.B. (2006). Analysis of microsatellites in citrus. Unigenes Acta Genefica Sinica. 33 (4): 345-353.

Jin, J.Q., Cui, H.R., Chen, W.Y., Lu, M., Yao, Y., Xin, Y., Gong, X.C., (2006). Data mining for SSRs in ESTs and development of EST-SSR marker in tea plant (*Camellia sinensis*). Journal of Tea Science. 26: 17–23.

Jin, J.Q., Cui, H.R., Gong, X.C., Chen, W.Y., Xin, Y. (*2007)* Studies on tea plants (*Camellia sinensis*) germplasms using EST-SSR marker. Hereditas. 29 (1): 103-108.

Kaneko, S., Ozawa, A., Saito, T., Tatara, A., Katayama, H., Doi, M. (2006). Relationship between the seasonal prevalence of the predacious coccinellid *Pseudoscymnus hareja (Coleoptera: Coccinellidae*) and the mulberry scale *Pseudaulacaspis pentagona (Hemiptera: Diaspididae*) in tea fields: Monitoring using sticky traps. Applied entomology and zoology. 41 (4): 621–626.

Kang, J., Lee, M.S., Gorenstein, D.G. (2005). The enhancement of PCR amplification of a random sequence DNA library by DMSO and betaine: Application to in vitro combinatorial selection of aptamers. Journal of Biochemical and Biophysical Methods. 64: 147–151.

Katoh, Y., Katoh, M., Takeda, Y., Omiori, M. (2003). Genetic diversity within cultivated teas base on nucleotide sequence comparison of ribosomal RNA maturase in chloroplast DNA. Euphytica. 134: 287-295.

Kaundun, S.S., Windass, J.D. (2006). Derived cleaved amplified polymorphic sequence, a simple method to detect a key point mutation conferring acetyl CoA carboxylase inhibitor herbicide resistance in grass weeds. Weed Research. 46: 34–39.

Kaundun, S.S., Matsumoto, S. (2002). Heterologous nuclear and chloroplast microsatellite amplification and variation in tea, *Camellia sinensis*. Genome. 45: 1041–1048.

Kaundun, S.S., Matsumoto, S. (2003a). Identification of processed Japanese green tea based on polymorphisms generated by STS-RFLP Analysis. Journal of Agricultural and Food Chemistry. 5: 1765-1770.

Kaundun, S.S., Matsumoto, S. (2003b). Development of CAPS markers based on three key genes of the phenylpropanoid pathway in Tea, *Camellia sinensis* (L.) O. *Kuntze*, and differentiation between *assamica* and *sinensis* varieties. Theoretical Applied Genetics. 106: 375–383.

Kaundun, S.S., Matsumoto, S. (2004). PCR-based amplicon length polymorphisms (ALPs) at microsatellite loci and indels from non-coding DNA regions of cloned genes as a means of authenticating commercial Japanese green teas. Journal of the Science of Food and Agriculture. 84: 895–902.

Kaundun, S.S., Park, Y.G. (2002). Genetic Structure of Six Korean Tea Populations as Revealed by RAPD-PCR Markers. Crop Science. 42:594–601.

Kaundun, S.S., Zhyvoloup, A., Park, Y.G. (2000). Evaluation of the genetic diversity among elite tea (*Camellia sinensis* var. *sinensis*) accessions using RAPD markers. Euphytica. 115: 7–16.

Kumar, N., Pandey, S., Bhattacharya, A., Ahuja, P.S. (2004). Do leaf surface characteristics affect Agrobacterium infection in tea [*Camellia sinensis* (L.) O Kuntze]? Journal of Biological sciences. 29 (3): 309-317.

Kwong-Robbins, C. (2005). Tea Time: Have You Had Your Tea Yet? US Pharmacology. 10: 47-50.

Lai, J.A., Yang, W.C., Hsian, J.Y. (2001). An assessment of genetic relationship in cultivated tea clones and native wild tea in Taiwan using RAPD and ISSR markers. Botanical Bulletin of Academia Sinica. 42: 93-100.

Li, S.X., Yin, T.M. (2007). Map and analysis of microsatellites in the genome of Populus: The first sequenced perennial plant. Science in China Series C: Life Sciences. 50 (5): 690-699.

Liu, K., Muse, S.V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 21 (9): 2128-2129.

Magoma, G.N., Wachira, F.N., Obanda, M., Imbuga, M., Agong, S.G. (2000). The use of catechins as biochemical markers in diversity studies of tea (*Camellia sinensis*). Genetic Resources and Crop Evolution. 47: 107–114.

Magoma, G.N., Wachira, F.N., Imbuga, M.O., Agong, S.G. (2003). Biochemical differentiation in *Camellia sinensis* and its wild relatives as revealed by isozyme and catechin patterns. Biochemical Systematics and Ecology. 31: 995–1010.

Maqbool, A., Zahur, M., Irfan, M., Younas, M., Barozai, K., Rashid, B., Husnain, T., Riazuddin, S. (2008). Identification and expression of six drought-responsive transcripts through differentiald in Desi Cotton (*Gossypium arboreum*). Molecular Biology. 42 (4): 492–498.

Magnuson, V.L., Ally, D.S., Nylund, S.J., Karanjawala, Z.E. (1996). Substrate Nucleotide-Determined Non-Templated Addition of Adenine by *Taq*DNA Polymerase: Implications for PCR-Based Genotyping and Cloning. BioTechniques. 21: 700-709.

Marques, C.M., Brondani, R.P.V., Grattapaglia, D., Sederoff, R. (2002). Conservation and synteny of SSR loci and QTLs for vegetative propagation in four Eucalyptus species. Theoretical Applied Genetics. 105: 474–478.

Matsumoto, S., Kiriiwa, Y., Takeda, Y. (2002). Differentiation of Japanese green tea cultivars as revealed by RFLP analysis of phenylalanine ammonia-lyase DNA. Theoretical Applied Genetics. 104: 998–1002.

Matsumoto, S., Kiriiwa, Y., Yamaguchi, S. (2004). The Korean Tea Plant (*Camellia sinensis*): RFLP analysis of genetic diversity and relationship to Japanese tea. Breeding Science. 54: 231-237.

McGregor, C.E., Lambert, C.A., Greyling, M.M., Louw, J.H., Warnich, L. (2000). A comparative assessment of DNA fingerprinting techniques (RAPD, ISSR, AFLP and SSR) in tetraploid potato (*Solanum tuberosum L.*) germplasm. Euphytica. 113: 135–144.

Mishra, R.K., Sen-Mandi, S. (2004). Molecular profiling and development of DNA marker associated with drought tolerance in tea clones growing in Darjeeling. Current Science. 87 (1): 60-66.

Mondal, T.K., Bhattacharya, A., Ahuja, P.S. (2001). Transgenic tea [*Camellia sinensis* (L.) O. Kuntze cv. Kangra Jat] plants obtained by Agrobacterium-mediated transformation of somatic embryos. Genetic Transformation and Hybridization. 20: 712-720.

Mondal, T.K. (2002). Assessment of genetic diversity of tea (*Camellia sinensis* (L.) O. Kuntze) by inter-simple sequence repeat polymerase chain reaction. Euphytica. 128: 307–315.

Mondal, T.K., Bhattacharya, A., Laxmikumaran, M., Ahuja, P.S. (2004). Recent advances of tea (*Camellia sinensis*) biotechnology. Plant Cell, Tissue and Organ Culture. 76: 195–254.

Muoki, R.C., Wachira, F.N., Pathak, R.S. Kamunya, S.M. (2007). Assessment of the mating system of *Camellia sinensis* in biclonal seed orchards based on PCR markers. Journal of Horticultural Science & Biotechnology. 82 (5): 733–738.

Muthumani, T., Kumar, R.S.S. (2007). Influence of fermentation time on the development of compounds responsible for quality in black tea. Food Chemistry. 101: 98–102.

Myburg, A.A. Cawood, M., Wingfield, B.D., Botha, A.M. (1998). Development of RAPD and SCAR markers linked to the Russian wheat aphid resistance gene Dn2 in wheat. Theoretical Applied Genetics. 96: 1162-1169.

Nei, M., Chesser, R.K. (1983). Estimation of fixation indices and gene diversities. Annual Human genetics. 47: 253-259.

Nei, M., Li, W.H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences of the United States of America. 76 (10): 5269-5273.

Nybom, H. (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Molecular Ecology. 13: 1143–1155.

Ni, S., Yao, M., Chen, L., Zhao, L., Wang, X. (2008). Germplasm and breeding research of tea plant based on DNA markers approaches. Frontiers in Agricultural China. 2 (2): 200-207.

Ogita, S., Uefuji, H., Yamaguchi, Y., Koizumi, N., Sano. H. (2003). Producing decaffeinated coffee plants. Nature. 423: 823.

Ohsako, T., Ohgushi, T., Motosugi, H., Oka, K. (2008). Microsatellite variability within and among local landrace populations of tea, *Camellia sinensis* (L.) O. Kuntze, in Kyoto, Japan. Genetic Resources and Crop Evolution. 55: 1047-1053.

Owuor, P.O., Obanda, M., Nyirenda, H.E., Mphangwe, N.I.K., Wright, L.P., Apostolides, Z. (2006). The relationship between some chemical parameters and sensory evaluations for plain black tea (*Camellia sinensis*) produced in Kenya and comparison with similar teas from Malawi and South Africa. Food Chemistry. 97: 644-653.

Page, R.D.M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences. 12: 357-358.

Pakniyat, H., Tavakol, E. (2007). RAPD markers associated with drought tolerance in bread wheat (*Triticum aestivum* L.). Pakistan Journal of Biological Science. 10 (18): 3237-3239.

Park, J.S., Kim, J.B., Hahn, B.S., Kim, K.H., Ha, S.H., Kim, J.B., Kim Y.H. (2004). EST analysis of genes involved in secondary metabolism in *Camellia sinensis* (tea), using suppression subtractive hybridization. Plant Science. 166: 953–961.

Park, S.D.E. (2001). Trypano tolerance in West African Cattle and the Population Genetic Effects of Selection. Ph.D. thesis. University of Dublin.

Parry, M.A.J., Flexas, J., Medrano, H. (2005). Prospects for crop production under drought: research priorities and future directions. Annals of Applied Biology. 147: 211-226.

Paul, S., Wachira, F.N., Powell, W., Waugh, R. (1997). Diversity and genetic differentiation among populations of Indian and Kenyan tea (*Camellia sinensis* (L.) O. Kuntze) revealed by AFLP markers. Theoretical Applied Genetics. 94: 255–263.

Paul, S.K. (2008). Predictability of price of tea from sensory assessments and biochemical information using data-mining techniques. Journal of the Science of Food and Agriculture. 88: 1354–1362.

Peleman, J.D., Van der Voort, J.R. (2003). Breeding by design. Trends in Plant Science. 8 (7): 330-334.

Perva-Uzunalić, A., Skerget, M., Knez, Z., Weinreich, B., Otto, F., Gruner, S. (2006). Extraction of active ingredients of green tea (*Camellia sinensis*): Extraction efficiency of major catechins and caffeine. Food Chemistry. 96: 597-605.

Pijut, P.M., Woeste, K.E., Vengadesan, G., Michler, C.H. (2007). Technological advances in temperate hardwood tree improvement including breeding and molecular marker applications. In Vitro Cell Development Biology-Plant. 43: 283-303.

Pompanon, F., Bonin, A., Bellemain, E., Taberlet, P. (2005). Genotyping errors: causes consequences and solutions. Nature reviews. Dio: 10.1038/nrg1707.

Ponmurugan, P., Baby, U.I., Rajkumar, R. (2007). Growth, photosynthetic and biochemical responses of tea cultivars infected with various diseases. Photosynthetica. 45 (1): 143-146.

Poulin, J., Weller, S.G., Sakai, A.K. (2005). Genetic diversity does not affect the invasiveness of fountain grass (*Pennisetum setaceum*) in Arizona, California and Hawaii. Diversity and Distributions. 11: 241–247.

Prince, L.M., Parks, C.R. (2001). Phylogenetic Relationships of Theaceae Inferred from Chloroplast DNA Sequence. American Journal of Botany. 88 (12): 2309-2320.

Rehman, S., Bhatti, H.N., Iqbal, Z., Rashid, U. (2007). Essential oil composition of commercial black tea (*Camellia sinensis*). International Journal of Food Science and Technology. Doi: 10.1111/j.1365-2621.2006.01458.

Roche, P., Alston, F.H., Maliepaard, C., Evns, K.M., Vrielink, R., Dunemann, F., Markussen, T., Tartarini, S., Brown, L.M., Ryder, C., King, G.J. (1997). RFLP and RAPD markers linked to the rosy leaf curling aphid resistance gene (Sd1) in apple. Theoretical Applied Genetics. 94: 528-533.

Rogers, P.J., Smith, J.E., Heatherley, S.V., Pleydell-Pearce, C.W. (2008). Time for tea: mood blood pressure and cognitive performance effects of caffeine and theanine administered alone and together. Psychopharmacology. 195: 569-577.

Salem, H.H., Ali, B.A., Huang, T.H., Qin, D.N., Wang, X.M., Xie, Q.D. (2007). Use of random Amplified Polymorphic DNA analysis for economically important food crops. Journal of Integrative Plant Biology. 49: 1670-1680.

Sandal, I., Saini, U., Lacroix, B., Bhattacharya, A., Ahuja, P.S., Citovsky, V. (2007). Agrobacterium-mediated genetic transformation of tea leaf explants: effects of counteracting bactericidity of leaf polyphnols without loss of bacterial virulence. Genetic Transformation and Hybridization. 26: 169-176.

Saravanan, M., John, K.M.M., Kumar, R.R., Pius, P.K., Sasikumar, R. (2005). Genetic diversity of UPASI tea clones (*Camellia sinensis* (L.) O. Kuntze) on the basis of total catechins and their fractions. Phytochemistry. 66: 561–565.

Savelkoul, P.H.M., Aarts, H.J.M., De Haas, J., Dijkshoorn, L., Duim, B., Otsen, M., Rademaker, J.L.W., Schouls, L., Lenstra, J.L. (1999). Amplified-Fragment Length Polymorphism Analysis: the State of an Art. Journal of Clinical Microbiology. 37 (10): 3083–3091.

Semagn, K., Bjørnstad, A., Ndjiondjop, M.N. (2006). An overview of molecular marker methods for plants. African Journal of Biotechnology 5 (25): 2540-2568.

Selkoe, K.A., Toonen, R.J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters. 9: 615–629.

Shaheen, S.M., Hossen, N., Ahmed, M., Amran, S. (2006). Green Tea in Health Care: A natural medicine, a natural drink. Journal of Applied Science Research. 2 (6): 306-309.

Sharma, P., Kumar, S. (2005). Differential display-mediated identification of three drought-responsive expressed sequence tags in tea [*Camellia sinensis* (L.) O. Kuntze]. Journal of Biosciences. 30 (2): 231-235.

Siddiqui, I.A., Afaq, F., Adhami, V.M., Ahmad, N., Mukhtar, H. (2004). Antioxidants of the Beverage Tea in Promotion of Human Health. Antioxidants & Redox signalling. 6 (3): 571-582.

Singh, D., Ahuja, P.S. (2006). 5S rDNA gene diversity in tea (*Camellia sinensis* (L.) O. Kuntze) and its use for variety identification. Genome. 49: 91–96.

Singh, M., Saroop, J., Dhinman, B. (2004). Detection of intra-clonal genetic variability in vegetatively propagated tea using RAPD markers. Biologia Plantarum. 48 (1): 113-115.

Solomon, E.D., Berg, L.R., Martin, D.W. (2002). In Biology. 6th edition. ISBN: 0-03-033503-5. Chapter 22: Understanding diversity: Systematics. 468-472.

Su, M.H., Tshou, C.H., Hsieh, C.F. (2007). Morphological comparisons of Taiwan native wild tea plant (*Camellia sinensis* (L.) O. Kuntze forma formosensis Kitamura) and two closely related taxa using numerical methods. Taiwania. 52 (1): 70-83.

Svendsen, E., Wilen, R., Stevenson, R., Liu, R., Tanino, K.K. (2007). A molecular marker associated with low temperature induction of dormancy in red osier dogwood (*Cornus sericea*). Tree physiology. 27: 385-397.

Takeda, M. (2004). Effects of temperature on oviposition in overwintering females and hatch in first-generation larvae of *Pseudaulacaspis pentagona* (Hemiptera: *Diaspididae*) Applied entomology and zoology. 39 (1): 15–26.

Tanaka, J., Taniguchi, F. (2007). In Genome mapping and molecular breeding in plant. Volume 6. ISBN: 13 978-3-540-34537-4. Chapter 6: Tea. 119-126.

Tamura, K., Dudley, J., Nei, M., Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution. 24: 1596-1599.

Taran, B., Buchwaldt, L., Tullu, A., Banniza, S., Warkentin, T.D., Vandenberg, A. (2003). Using molecular markers to pyramid genes for resistance to ascochyta blight and anthracnose in lentil (*Lens culinaris Medik*). Euphytica. 134: 223-230.

Tautz, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research. 17 (16): 6463-6471.

Techen, N., Crockett, S.L., Khan, I.A., Scheffler, B.E. (2004). Authentication of Medicinal Plants Using Molecular Biology Techniques to Compliment Conventional Methods. Current Medicinal Chemistry. 11: 1391-1401.

Thomas, J., Kumar, R.R., Mandal, A.K.A. (2006a). Metabolite profiling and characterization of somaclonal variants in tea (*Camellia* spp.) for identifying productive and quality accession. Phytochemistry. 67: 1136-1142.

Thomas, J., Vijayan, D., Joshi, S.D., Lopez, S.J., Kumar, R.R. (2006b). Genetic integrity of somaclonal variants in tea (*Camellia sinensis* (L.) O Kuntze) as revealed by inter simple sequence repeats. Journal of Biotechnology. 123:149-154.

Tóth, G., Gáspári, Z., Jurka, J. (2000). Microsatellites in different eukaryotic genomes: Survey and analysis. Genome Research. 10: 967-981.

Tullu, A., Buchwaldt, L., Warkentin, T., Taran, B., Vandenberg, A. (2003). Genetics of resistance to anthracnose and identification of AFLP and RAPD markers linked to the resistance gene in PI 320937 germplasm of lentil (*Lens culinaris* Medikus). Theoretical applied genetics. 106: 428-434.

Vandamme, A.M. (2003). In: The Phylogenetic Handbook: A practical approach to DNA and protein phylogeny. ISBN: 0 521 80390. Chapter 1: Basic concepts of molecular evolution. 1-20.

Varshney, R.K., Graner, A., Sorrells, M.E. (2005). Genic microsatellite markers in plants: features and applications. TRENDS in Biotechnology. 23 (1): 48-55.

Varshney, R.K., Chabane, K., Hendre, P.S., Aggarwal, R.K., Graner, A. (2007). Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. Plant Science. 173: 638–649.

Wachira, F.N., Ng'etich, W.K. (1999). Dry-matter production and partition in diploid, triploid and tetraploid tea. Journal of Horticultural Science & Biotechnology. 74 (4): 507-512.

Wachira, F., Tanaka, J., Takeda, Y. (2001). Genetic variation and differentiation in tea (*Camellia sinensis*) germplasm revealed by RAPD and AFLP variation. Journal of Horticultural Science & Biotechnology. 76 (5): 557-563.

Walker, J.M. (2000). Chapter 12: Electrophoretic techniques. In Principles and Techniques of practical Biochemistry 5th edition (editor Wilson, K., Walker J.M.) Cambridge University press. ISBN 0521 65873X. 580-612.

Weir, B.S., Hill, W.G. (2002). Estimating F statistics. Annual Reviews in Genetics. 36: 721–750.

Wheeler, D.S., Wheeler, W.J. (2004). The Medicinal chemistry of tea. Drug development research. 61: 45-65.

Wolf, A., Bray, G.A., Popkin, B.M. (2007). A short history of beverages and how our body treats them. Obesity reviews. Doi: 10.1111.

Wright, L.P. (2002). Biochemical analysis for identification of quality in black tea (*Camellia sinensis*). Ph.D thesis. University of Pretoria. Biochemistry. 14-18.

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

Xu, Y., Ma, R.C., Xie, H., Liu, J.T., Cao, M.Q. (2004). Development of SSR markers for the phylogenetic analysis of almond trees from China and the Mediterranean region. Genome. 47 (6): 1091-1104.

Yadav, S.Y., Ahuja, P.A. (2007). Towards Generating Caffeine-free Tea by metabolic engineering. Plant Foods for Human Nutrition. 62: 185-191.

Yang, C.S., Lambert, J.D., Ju, J., Lu, G., Sang, S. (2006). Tea and cancer prevention: Molecular mechanisms and human relevance. Toxicology and Applied Pharmacology. Doi:10.1016/j.taap.2006.11.024.

Yao, M.Z., Chen, L., Liang, Y.R. (2008). Genetic diversity among tea cultivars from China, Japan and Kenya revealed by ISSR markers and its implication for parental selection in tea breeding programmes. Plant Breeding. 127: 166-172.

Yazdani, R., Scotti, I., Jansson, G., Plomion, C., Mathur, G. (2003). Inheritance and diversity of simple sequence repeat (SSR) microsatellite markers in various families of *Picea abies.* Hereditas. 138: 219–227.

Young-Goo, P., Kaundun, S.S., Zhyvoloup, A. (2002). Use of the bulked genomic DNA-based RAPD methodology to assess the genetic diversity among abandoned Korean tea plantations. Genetic Resources and Crop Evolution. 49 (2): 159-165.

Zane, L., Bargelloni, L., Paratnello, T. (2002). Strategies for microsatellite isolation: A review. Molecular Ecology. 11: 1-16.

Zhao, L.P., Liu, Z., Chen, L. (2007). Generation and characterization of 24 novel EST derived microsatellites from tea plants (*Camellia sinensis*) and cross species amplification in closely related species and varieties. Conserved Genetics. Dio: 10.1007/s10591-077-9476-y.

Zhao, L.P., Liu, Z., Chen, L., Yao, M.Z., Wang, X.C. (2008). Generation and characterization of 24 novel EST derived microsatellites from tea plant (Camellia sinensis) and cross-species amplification in its closely related species and varieties. Conservation Genetics. 9: 1327–1331.