THE GENETIC DIVERSITY OF *PENNISETUM GLAUCUM (L.) R. BR.* (PEARL MILLET) LANDRACES IN NAMIBIA

A THESIS SUBMITTED IN FULFILMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

OF

THE UNIVERSITY OF NAMIBIA

BY

BILLY McBENEDICT

200847074

September 2015

Main Supervisor: Prof PM. Chimwamurombe

Co-Supervisor: Dr E. G. Kwembeya

Co-Supervisor: Dr G. Maggs-Kölling (Gobabeb Research and Training Centre)

ABSTRACT

The current *Pennisetum glaucum (L.) R. Br.* cultivars in Namibia have overall poor performance and this poses a threat to the nation's food security because this crop is staple for over 70% of the Namibian population. The crop suffers from a range of undesirable production traits such as; susceptibility to diseases, low yield and prolonged reproductive cycle due to lack of genetic diversity. This study was aimed at understanding the genetic diversity of the crop in Namibia by means of Internal Transcribed Spacer (ITS) sequence data analysis, and SSR and RAPD techniques, and evaluating the most informative technique of the three. A total of 1441 genotypes were collected from the Gene Bank representing all the Namibian landraces. Samples of 96 genotypes were further analyzed using the Shannon-Wiener diversity index and revealed values of 0.45 for SSR and 0.7 for RAPD indicating low genetic diversity. Ordination results using Principal Coordinate Analysis (PCoA) for RAPD, morphology and SSR data confirmed the same clusters as generated by Un-weighted Pair Group Method with Arithmetic Mean (UPGMA) for the same data sets. UPGMA generated phenograms of 29 morphological characterized genotypes were generated for SSR, RAPD and morphology data. The ITS sequences of the same 29 accessions were used to construct a phylogenetic tree. This tree revealed clades with highest similarity with the phenogram generated from morphology (78%) data, followed by the SSR data (68%), and lastly the RAPD (50%) data. Lodging susceptibility, tillering attitude, spike density, bristle length, fodder yield potential, early vigour, number of nodal tillers and

spike shape were identified as the phenotypic characters upon which some clusters were based in all data sets used. It is recommended that efforts be made to widen the available gene pool in order to increase genetic diversity in Namibia by introducing genetically diverse accessions from other countries and maintaining the diversity through breeding programs.

Keywords: Genetic diversity, Internal Transcribed Spacers, ITS, *Pennisetum glaucum*, RAPD, SSR

TABLE OF CONTENTS

ABBREVIATIONS

µl - microlitre

µM - micromole

A – Adenine

ADH – Alcohol dehydrogenase

AFLP – Amplified Fragment Length Polymorphism

bp – Base pair

C – Cytosine

CMS – Cytoplasmic male-sterility

cpDNA – Chloroplast deoxyribonucleic acid

cpSSRs – Chloroplast Simple Sequence Repeats

CTAB - Cetyl trimethylammonium bromide

dATP - Deoxyadenosine triphosphate

dCTP - Deoxycytidine triphosphate

ddA – Dideoxyadenine

ddC – Dideoxycytosine

ddG – Dideoxyguanine

ddT - Dideoxythymine

ddTTP – Dideoxythymidine triphosphate

dGTP - Deoxyguanosine triphosphate

DNA – Deoxyribonucleic acid

DNRs - Dinucleotide repeats

dTTP - Deoxythymidine triphosphate

EDTA – Ethylenediaminetetra acetic acid

EST – Expressed sequence tags

FAO – Food and Agriculture Organisation

G – Guanine

IARI - Indian Agricultural Research Institute

ICRISAT - International Crops Research Institute for the Semi-Arid Tropics

ITS – Internal Transcribed Spacer regions

MEGA - Molecular Evolutionary Genetics Analysis

Mg^{2+} - Magnesium

NCBI - National Centre for Biotechnology Information

Ng – Nanogram

NPGRC - National Plant Genetic Resource Centre

nrDNA – Nuclear ribosomal Deoxyribonucleic acid

^{o}C – Degree Celsius

OH – Hydroxyl

PCA - Principal Component Analysis

PCoA - Principal Coordinate Analysis

PCR – Polymerase Chain reaction

PIC – Polymorphism Information Content

QTL – Quantitative trait locus

RAPD – Random Amplified Polymorphic DNA

RFLP – Restriction Fragment Length Polymorphism

RNA - Ribonucleic acid

Rpm – Revolutions per minute

RPO - Research and Publications Office

rRNA – Ribosomal ribonucleic acid

SNPs – Single Nucleotide Polymorphisms

SSLPs - Simple Sequence Length Polymorphisms

SSRs – Simple Sequence Repeats

STMS - Sequence tagged Microsatellite site

STRs - Short/simple tandem repeats

T – Thymine

TNRs – Trinucleotide repeats

TTNRs - Tetra-nucleotide repeats

UNAM – University of Namibia

UPGMA – Unweighted pair group method using arithmetic averages

VNTRs – Variable number of tandem repeats

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge the sponsors of this study: UNDP through the Benefit-Sharing Fund (BSF) for providing financial resources for this study. I am also grateful to the National Plant Genetic Resource Centre of Namibia and their staff members for providing the Pearl millet accessions that were used in this study. A special thanks to Ms. R. Moses and Ms. K. Sikute who were helpful during the collection of the *Pennisetum glaucum (L). R. Br.* genotypes.

I am deeply grateful to my main supervisor, Prof P.M. Chimwamurombe for being my mentor since the onset of my career to whom I owe my academic achievements. I also thank Prof P.M. Chimwamurombe for considering me to be his student among others. He has always extended his guidance and help beyond academics. His advice and encouragement has triggered positive energies of change in my life. Furthermore, am grateful for his availability even in times when he was busy and far, he made it possible for me to progress with this research making his absence not have a significant drawback. I must say he is one of a kind and has a heart that is willing to help, an ear that listens to predicaments and provides a solution of progress. Through him I have learnt a lot about Science, research and various issues of life that have elevated my level of maturity. My story of success can never surely be complete in the omission of his Name.

I am thankful to my Co-supervisor, Dr E. G. Kwembeya. His door was always open for consultation and advice. His input in this research was immense and

unreserved; he always advised me to work hard and remain under God's covering. I am also grateful to my Co-supervisor Dr G. Maggs-Kölling. She encouraged me and made it possible for me to engage in this research. I also want to thank her for not giving up on my academic capabilities.

I am highly grateful to whom I am indebted Dr. E. Julies who encouraged me to work hard and granted me unlimited access to the Biological Sciences Laboratory and its facilities. The success of this study is owed to her support. Besides, she ensured that I had a favorable environment to carry out my research. I would like to give my appreciation to Prof. K. Chinsembu for his unrelenting support, advice and immense knowledge in genetics that he offered me during this study and was indeed needed to accomplish this study.

I am thankful for the technical support from my colleagues and technicians of the Biological Sciences Department; Ms. M. J. Johnson, Ms. K. Kaitjizemine, Mr. A. Mbangu. They made it possible for me to have access to laboratory equipment and rendered their unlimited support. The smooth advancement of this study is owed to them. I extend my thanks to the whole Biological Sciences Department staff at large for their various support rendered during this study. I am deeply grateful to Dr. R. Bock who rendered his support and encouragement to me during my proposal writing. He was effective and informative in the progress of my proposal and the amendment's that were due to be made.

I would like to thank Ms. Celine Mukakalisa for the advice during my data analysis. Her kindness and support pushed me to work hard. I am also grateful to my

colleagues; Daniel Haiyambo, Bianca Nawases, Munyaradzi Tambo, Taonga Namate, Florence Dushimemaria, Naomi Libala, Lahja Omagano Whitney Shingenge and Fransiska Kangombe for their love, support and encouragement during this study. I am grateful to the Biochemistry and Chemistry Department of the University of Namibia for providing access to their laboratory equipment. I am also grateful to their technical staff and colleagues; Annastasia Ekandjo, Hilaria Hakwenye, Seno Namwandi and Moola Nyambe for their support. Furthermore, I am thankful to Dr. R. Steenkamp and Prof S. Singh for granting me access to use the Physics Optics lab during this study.

To my family, Beauty, Benedict, Lydia, Ben, Bertha, Britney, Constance, Tapfuma, Rumbidzai and Eveline I am very grateful for the unlimited love, support, encouragement and prayers that u continuously give to me. You have always given me the drive to pull through hardships and challenges and to you I owe the hard work implemented in this research. Furthermore, I would also love to thank my best friend Wilhelmina Hauwanga who gave me unlimited encouragement, love and support during this study.

DEDICATION

I dedicate my hard work and endurance in this study to my parents. They have always been supportive in my education and shown me love. In addition they have always encouraged me to aim to progress in all areas of my life and to put God first. I also dedicate this thesis to my entire family. I appreciate their love and prayer for me and take this opportunity to say I deeply love and appreciate God for them.

DECLARATIONS

I, Billy McBenedict, declare hereby that this study is a true reflection of my own research, and that this work, or part thereof has not been submitted for a degree in any other institution of higher education.

……. Date……..

Billy McBenedict

CONFERENCE PROCEEDINGS AND MERITS

CONFERENCE PROCEEDINGS

1. McBenedict, B., Chimwamurombe, P.M., Kwembeya, E. G. & Maggs-Kölling, G. (2014). The Genetic Diversity of *Pennisetum glaucum (L.) R. Br.* (Pearl millet) landraces in Namibia- **Power point presentation** presented at the 2nd Annual Science Research Conference at the University of Namibia Windhoek Main Campus from 30th – 31st October, 2014.
2. McBenedict, B., Chimwamurombe, P.M., Kwembeya, E. G. & Maggs-Kölling, G. (2015). The Genetic Diversity of *Pennisetum glaucum (L.) R. Br.* (Pearl millet) landraces in Namibia- The data from this research was used in a panel discussion at the Youth Environmental Summit (YES) in celebration of the International Day for Biological Diversity, Waterberg, Namibia, May, 22 - 24, 2015.

CHAPTER 1: INTRODUCTION

1.1 Orientation of proposed study

1.1.1 Genetic diversity

Genetic diversity is defined as a variety of genes in a given species which are important for adaptation, conservation of desired traits and drives survival during natural selection (Holden & Peacock, 1993). Variations can occur in many forms such as; the number and structure of chromosomes, amount of DNA found in cells, and dissimilarities in the DNA sequence. Genetic diversity occurs as an outcome of mutations, recombination, genetic drift, migration and selection. Populations evolve due to the presence of genetic diversity which is encrypted in the differences in genes and allele frequencies (De Vicente & Fulton, 2003). Genetic diversity is an important aspect in the exploration of plant genetic resources. Plant genetic resources are comprised of wild species, landraces, commercial cultivars, hybrids and breeding lines, and wild relatives of crops. These resources have proved to be beneficial in conservation genetics and have future potential applications, especially in crop improvement (De Vicente & Fulton, 2003).

Crop improvement programs are successfully implemented with the knowledge of the gene pool of the crop under investigation. This information is important because both the environment and genes play an important role on the phenotype of an organism (Sharma, 2001). The gene pool can be assessed for genetic variation by scrutinizing the

phenotype and genotype of an organism. The phenotype of an individual is a product of the collective interaction between the environment and the individual's genes while the genotype is the actual genetic make-up of an organism. Furthermore, the results of a crop genetic diversity study provide information that is necessary to its improvement, adaptation, conservation of desired traits and shows how the phenotype of the crop is related to its genetic expression. Molecular tools have demonstrated effectiveness in investigating genetic diversity within gene pools and distinguishing unique cultivars with desired traits of agro-economic importance (Barcaccia, 2009). The effectiveness of molecular tools is dependent on their ability to detect and analyze molecular markers. Molecular markers are sequences of DNA or protein that are readily detectable and whose inheritance and polymorphism can be monitored (Chakauya, 2002).

1.1.2 Pearl millet

Pennisetum glaucum (L.) R. Br. (Pearl millet) is a multi-purpose cereal widely cultivated in different parts of the world for grain, stover and green fodder on about 27 million hectares, primarily in Asia and Africa (Chakauya, 2002). It is the sixth most important cereal crop annually produced in the world following wheat, rice, maize, barley and sorghum (Sharma, 2001). Pearl millet has also been extensively described to be used for human food (Board on Science and Technology for International Development, 1996; Maundu, Ngugi & Kabuye, 1999; Vom Brocke, Weltzien, Christinck, Presterl & Geiger, 2003), animal food (Board on Science and Technology for International Development, 1996; Cunningham, Mulham, Milthorpe & Leigh, 1981;

Maheshwari, 1963; Maundu, Ngugi & Kabuye, 1999), as a fuel (Board on Science and Technology for International Development, 1996), and for medicinal purposes (Maundu, Ngugi & Kabuye, 1999).

Pearl millet is a highly nutritious staple food crop widely grown in different regions of Namibia namely; Zambezi, Kavango East, Kavango West, Otjozondjupa, Oshikoto, Kunene, Omusati, Ohangwena and Oshana. Among other uses, the crop is used to brew opaque beer and to cook traditional scones called *mun'ende* (Silozi) and *Oshikwiila* (Oshiwambo) which are also fed to chickens (Board on Science and Technology for International Development, 1996). *P. glaucum* accessions have been collected, morphologically characterized and stored in the National Plant Genetic Resource Centre of Namibia (The Gene Bank of Namibia) in the past years. However, there have been no studies on the molecular characterization of the Namibian *P. glaucum* genotypes. This study is therefore aimed at exploring the genetic diversity of the crop by means of microsatellites markers, Random Amplified Polymorphic DNA (RAPD) markers and analysis of the Internal Transcribed Spacer (ITS) sequences in order to complement the morphological data lodged in the National Gene Bank.

1.2 Statement of the problem

The current *P. glaucum* cultivars in Namibia have an overall poor performance and there is lack of knowledge regarding the genetic diversity of the crop thus limiting; the understanding and improvement of the crop, and selection of suitable techniques that

can be routinely used to effectively monitor and reveal the level of genetic diversity of *P. glaucum* in future. The poor performance poses a threat to the nation's food security since it is a staple food crop for over 70% of the population. The crop suffers from a range of undesirable production traits, such as susceptibility to diseases, low yield as well as prolonged reproductive cycle due to lack of genetic diversity. This has an adverse effect on the net yield of the crop. In addition, the Gene Bank of Namibia incurs costs and spends time on collecting and further storing germplasm without knowledge of the genetic diversity within the crop. Knowledge of genetic diversity within the crop is important in germplasm management because only genetically diverse accessions will be stored. Furthermore, it is also relevant for the Gene Bank of Namibia to have knowledge of which technique is informative in assessing the genetic diversity of the Namibian germplasm for plant improvement purposes.

1.3 Overall aim of the study

The aim of the study was to generate information useful in resource management for germplasm conservation and breeding programs by determining the genetic diversity of *P. glaucum* landraces in Namibia. In addition, this study is aimed; at revealing the source of the poor performance, developing remediation strategies and enabling scientists and managers to preserve and enhance the diversity of the crop since it is a staple food crop for over 70% of the Namibian population.

1.4 Specific objectives of the study

- To investigate the genetic diversity of *P. glaucum* using microsatellites (SSRs) markers, Random Amplified Polymorphic DNA (RAPDs) markers and analysis of the Internal Transcribed Spacer (ITS) sequences in order to identify genetically distinct accessions with desired traits to be used in breeding programs aimed at improving Namibian *P. glaucum*.
- To determine the most informative method for studying the genetic diversity of *P. glaucum* among three techniques (ITS, SSR and RAPD) to allow the selection of suitable techniques that can be routinely used to effectively monitor and reveal the level of genetic diversity of *P. glaucum* in future.

1.5 Hypotheses of the study

- *P. glaucum* landraces in Namibia lack genetic diversity. Namibian *P. glaucum* landraces display undesirable production traits, such as susceptibility to diseases, low yield as well as prolonged reproductive cycle and these traits are indicative of lack of genetic diversity (Brown, 2008). Therefore, several molecular tools: microsatellites markers, RAPD markers and nr DNA ITS sequence data were applied in this study to prove this hypothesis.

1.6 Significance of the study

The results of this study will provide information useful in resource management for germplasm conservation and breeding programs in Namibia. This will help on the choice of cultivars with desired traits that are important to improving the *P. glaucum* crop and enhancing the nation's food security by implementing crop improvement programs using genetically diverse cultivars with the desired traits. Genetically diverse accessions will be identified and used in breeding programs. The high cost of storing accessions that is currently incurred by the Gene Bank will also be reduced because duplications will be determined and only genetically diverse accessions will be stored. Furthermore, this study will provide the Gene Bank with information regarding the most informative technique of the three (SSR, RAPD and ITS sequence analysis) to be used routinely in assessing the genetic diversity of *P. glaucum*.

1.7 Limitations of study

Most of the accessions were collected between the year 1991 and 2000, and only a few were collected after the year 2000. This implies that genetic changes in the crop that occurred after the period of collection were not detected in the present study. Therefore, all results and conclusions were based on the accessions that were collected during this period.

CHAPTER 2: LITERATURE REVIEW

2.1 Classification of *P. glaucum*

The genera *Pennisetum* is currently divided into five sections based on morphological characteristics and these sections are; *Penicillaria, Brevivalvula*, *Gymnothrix*, *Heterostachya* and *Eu-Pennisetum* (Martel *et al.*, 2004). *P. glaucum*, which is the domesticated pearl millet and its wild relatives, belong to the *Penicillaria* section which contains a total of about 140 species (Brunken, 1977). The species in the *Penicillaria* section have chromosome numbers of n = 5, n = 7, n = 8 or n = 9 with varying ploidy levels which range from diploid to octoploid (Jauhar, 1981). Martel *et al.* (2004) and Jauhar (1981) argued that cytological and flow cytometry studies have revealed a relationship between the number of chromosomes and the genome size which is, the higher the number of chromosomes, the smaller the genome size and *vice versa*.

P. glaucum is a diploid, monocotyledonous plant belonging to the *Poaceae* family with a basic chromosome number of seven. Possession of a low chromosome number (2n = 14), short life cycle, high multiplication ratio (up to 1:1000), ratooning ability and the ease with which cross pollination can be done due to protogyny make it a good subject for genetic research (Sharma, 2001). Ratooning is the ability of the plant to regenerate new tillers from the main crop stubbles (Pushpavathi, 2003).

2.2 Genetic diversity studies on *P. glaucum*

Most genetic diversity studies on *P. glaucum* have been done using morphological and isozyme data analyses (Tostain, Riandley & Marchaise, 1987; Tostain & Marchais, 1989; Tostain, 1994). However, isozyme markers have a limitation in that they are influenced by environmental conditions at various developmental stages (Falkenhagen, 1985). Moreover, the isozyme loci that can be analyzed are limited and genotypic variations are not easily detected (Tobolski & Kempery, 1992). The use of microsatellites and RAPD's to successfully study genetic diversity in *P. glaucum* has been documented (Yadav, Jain, S., Jain, V., & Jain, R., 2007).

2.3 Methods used to assess genetic diversity

Inter-population and intra-population measurements of genetic diversity are usually achieved by the use of allozyme or DNA markers which measure levels of variation directly. However, Mondini, Noorani and Paguotta (2009) argued that genetic diversity can as well be evaluated using morphological, and biochemical characterization and evaluation. Genetic diversity can be measured at a molecular level through the analysis of allozymes, microsatellites, minisatellites, Random Amplified Polymorphic DNA (RAPD), Restriction Fragment Length Polymorphisms (RFLP), Amplified Fragment Length Polymorphism (AFLP), and DNA sequencing of either segments such as ITS or the whole genome (Frankham, Briscoe & Ballou, 2002).

2.3.1 Morphological characterization

Morphological characterization involves experiments that need huge pieces of land which might render them expensive. However, this is the most basic way of characterization because it involves observations and does not require expensive equipment. This technique is limited due to the fact that traits are often susceptible to phenotypic plasticity caused by environmental variations (Mondini *et al.*, 2009).

2.3.2 Biochemical markers

Biochemical analysis depends on the splitting-up of proteins into particular banding patterns. This is a fast technique that can be performed with small amounts of biological material. However, only a limited number of enzymes are available and thus, the resolution of diversity is limited (Mondini *et al.*, 2009). Protein analysis was the first molecular method used to assess genetic diversity through electrophoresis (Frankham *et al.*, 2002). The technique employs electrophoresis to separate proteins on the basis of their net charge and molecular weights. These separations distinguish diverse forms of proteins and determine the level of genetic variation for particular protein locus. The electrophoretic detection of genetic diversity is underestimated because only around 30% of the DNA base changes alter protein charges. Blood, liver or kidney are used as samples in humans to perform protein electrophoresis while leaves and root tips are

used as samples in plants because these contain sufficient amounts and a variety of soluble proteins (Frankham *et al.*, 2002).

Protein markers reveal higher genetic variation as compared to observing morphological traits. Protein markers are also referred to as biochemical markers, and isozyme or allozyme analysis. Isozymes are enzymes that have structurally different molecular forms but have the same catalytic function while allozymes are allelic variants of enzymes encoded by structural genes and are the oldest among the molecular markers (Kumar, Gupta, Misra, Modi & Pandey, 2009). Isozymes are a consequence of changes in the amino acids which results in a change in the net charge or structure of the produced protein. These alterations are detected by the differences in protein movement on the gel because proteins are made up of different charged amino acids. Gel electrophoresis and enzyme-specific stains are able to reveal allelic variants.

Advantages and disadvantages of Biochemical markers

Allozymes procedures are quick and easy, independent of DNA extraction and do not need sequence information, primers or probes. Numerous crop improvement programmes have exploited isozymes markers effectively (Baes & Custsem, 1993; Glaszmann, Fautret, Noyer, Feldman & Lanaud., 1989; Vallejos, 1983). In addition, they are co-dominant markers whose banding patterns (zymograms) can be interpreted with ease in terms of loci and alleles, or might need segregation analysis of progeny of known parental crosses for interpretation. Allozymes are highly reproducible and can be

functional at relatively low costs depending on the enzyme staining chemical consumption. Furthermore, they are a robust complement to the morphological assessment of variation and require relatively simple equipment.

Optimization of the technique is needed for some enzymes in certain species. In some instances, complex banding profiles emerge from polyploidy or genes that have been duplicated and the formation of inter-genic heterodimers which may possibly complicate interpretation (Kumar *et al.*, 2009). Allozymes display low levels of polymorphism, are limited by being influenced by the environment and changes in different developmental stages, and are comparatively limited in number. Allozymes are also limited by co-migration; proteins may possess similar electrophoretic mobility and yet come from distantly related germplasm.

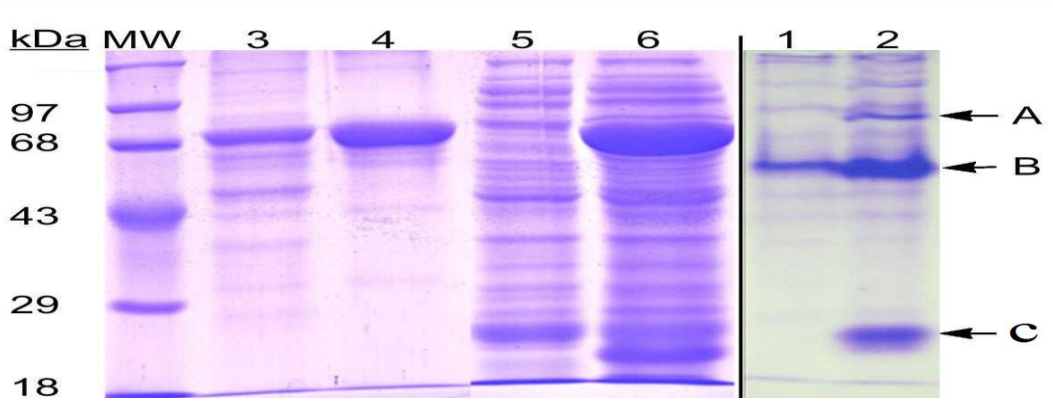


Figure 1: An illustration of the gel electrophoresis of protein markers. Adapted from http://www.bio.miami.edu/~cmallery/150/gene/Taq.htm. Retrieved on the 10th of June, 2015

2.3.3 DNA markers

Molecular analyses encompass a variety of DNA molecular markers. According to Mondini *et al.* (2009), the DNA markers can be effectively employed for analysis of variation. In addition, these markers have diverse applications and genetic qualities such as dominancy and co-dominance.

2.3.3.1 Microsatellites

Microsatellite loci are tandem repeats of short DNA segments, typically 1-5 bases in length (Frankham *et al.*, 2002). These repeats are abundant across genomes and occur in many species where they are found in different forms such as $(TG)_n$ or $(AAT)_n$ (Bruford & Wayne, 1993) and display high levels of polymorphism. Microsatellites are alternatively known as simple sequence repeats (SSRs) and short/simple tandem repeats (STRs) (Frankham *et al.*, 2002), simple sequence length polymorphisms (SSLPs) and variable number tandem repeats (VNTRs) (Bruford & Wayne, 1993), and sequence tagged microsatellite site (STMS) (Akkaya, Bhagwat & Cregan, 1992). Microsatellites exist in variable numbers due to mutation mechanisms of DNA slippage during DNA replication (slip-strand mis-pairing errors during DNA replication) and recombination during crossing over which possibly alters the SSR length or by gene conversion (Li, Korol, Fahima, Beiles & Nevo, 2002). They further argued that microsatellites make up a larger portion of the DNA noncoding regions and are rare in protein coding regions.

SSR's have been found to have functions involved in chromatin organization (chromosomal organization, DNA structure and centromere and telomere), regulation of DNA metabolic processes (DNA replication, recombination, mismatch repair system and cell cycle) and regulation of gene activity (transcription, binding protein & translation).

Microsatellite properties of hyper-variability, co-dominance and reproducibility brand them as being suitable for genome mapping and population genetics studies (Dayanandan, Rajora & Bawa, 1998). Different forms of SSRs exist, such as Inter-SSRs (a RAPD variant but more rigorous) and chloroplast microsatellites (cpSSRs), which are comparable to nuclear microsatellites. The repeat is typically only 1 base pair such as $(T)_n$ (Proven, Russell, Booth & Powell, 1999).

Microsatellites have mostly been applied in studies utilizing allozymes including investigation of gene flow and mating, systems and paternity exploration and diversity studies (Chase, Moller, Kesseli & Bawa, 1996; Rossetto, Slade, Baverstock, Henry & Lee, 1999; Streiff *et al.*, 1999). Few studies have also been conducted on phylogenetics (Robinson & Harris, 1999). In addition, SSRs have been utilized in cross species amplification in which the same primers have been used to amplify the identical regions in closely related taxa. Robinson & Harris (1999) suggest the possible existence of a connection between the degree of cross-species amplification and taxonomic distance.

A study by Mariac *et al.* (2006) investigated the existence of genetic diversity in 421 cultivated and 46 wild accessions of *P. glaucum* in Niger using microsatellites markers. Their results indicate higher gene diversity in wild accessions and lower

diversity and number of alleles in cultivated accessions. Mariac *et al.* (2006) argued that a significant variation exists between the cultivated and wild accessions in Niger. This is in agreement with Tostain (1994) who describes the morphological diversity of *P. glaucum* in Niger as the highest in West Africa with regards to spike morphology. These investigations and results complement information that both wild millet *(Pennisetum glaucum (L.) R. Br. ssp. monodii)* and cultivated pearl millet *(Pennisetum glaucum (L.) R. Br. ssp. glaucum)* are found in Niger.

Chowdari, Davierwala, Gupta, Ranjekar and Govila (1998) analyzed twenty two (22) *P. glaucum* cultivars and thirty six (36) *P. glaucum* landraces collected from the Indian Agricultural Research Institute (IARI). Their results showed the occurrence and polymorphism of di-, tri-, tetra-nucleotide and minisatellite repeats in the *P. glaucum* genome. In addition, microsatellite and minisatellite [$(GATA)_4$ and pV47] showed the most polymorphic patterns in both cultivars and landraces. However, here was no discrimination in polymorphism based on the geographic distribution of the landraces representing the eight states of India in which the samples were collected. In a genetic diversity study by Yadav *et al.* (2006) the genetic relationships/differences between the open pollinated varieties, hybrids, cytoplasmic male-sterility (CMS) and restorer lines in pearl millet, 20 pearl millet genotypes were assessed for diversity. A total of 349 clear bands were produced using 30 ISSR markers with 79.1% polymorphism across all the varieties.

Satyavathi *et al.* (2013) investigated the genetic diversity of *P. glaucum* novel set of restorer lines using SSR markers in 45. Their findings showed that the cluster

groupings were not according to pedigree but based on plant node pubescence, leaf sheath pubescence, spike exertion and spike density. The variability was noticeably based on heterozygosity displayed *via* gel bands. This highlights the importance of heterozygosity among the genotypes in breeding and the co-dominant property of microsatellites which differentiates heterozygotes from homozygotes by revealing two bands.

Advantages and disadvantages of microsatellites

Microsatellites ability to detect high levels of polymorphism has marked the technique as one of the most extensively used molecular technique in studies of genetics (Takundwa *et al.*, 2010). Microsatellites are employed in cultivar identification, paternity analysis and gene-flow research due to their high level of discrimination of individuals in an assortment of loci (Hokanson, Szewc-McFadden, Lamboy & McFerson, 1998). Furthermore, microsatellites have been found to have a high resolution in studies as compared to allozymes and RFLPs (Robinson & Harris, 1999). Microsatellites also have an advantage in that they are co-dominant markers making them able to differentiate heterozygotes from homozygotes. This intensifies their proficiency and precision in population genetic measures as compared to dominant markers such as RAPDs and AFLPs. Paternity analyses, gene-flow and hybridization exploit the identity of heterozygotes in the F_1 generation (Schlötterer & Pemberton, 1994). SSR PCR technique is relatively inexpensive in the event that primers are available. In addition, cross-species amplification of SSRs reduces the cost of

developing species specific suitable SSR primers; instead the same primers can be used in closely related taxa (Robinson & Harris, 1999). Neutral selectivity is among other properties desired in a marker for analysis and SSRs display neutral selectivity as opposed to allozymes (Robinson & Harris, 1999).

Although microsatellites are extensively utilized, they also have some disadvantages. This suggests a combination of techniques in order to achieve more informative results. The development and screening for microsatellites in organisms is complex and expensive especially in instances when the results of a screen produce less microsatellite loci. The occurrence of slippage during amplification presents a difficult in the analysis of mononucleotide and dinucleotide repeats. The manifestations of slippage leads to amplification of differently sized amplicons with a variation of about 1-5 repeat units from the expected product (Ciofi *et al.*, 1998).

The assumption that co-migrating fragments are homologous also presents a problem in the application of microsatellites in phylogenetic studies. Besides, non-homologous fragments are described as either occurring within the SSR repeat regions or the SSR flanking region. Homology is subject to mutation rates in the number of repeats, this problem is amplified with an increase in mutations and insignificant with reduced mutation rates. The homology hypothesis and microsatellites mechanisms of mutation has been studied in which amplified microsatellites were sequenced and the results displayed variations in the non-repeated flanking regions caused by point mutations and insertions or the deletion of bases (indels) (Robinson & Harris, 1999).

The effect of point mutations is unnoticed in the length of the microsatellites loci whereas a change in the length of the microsatellites amplification product is observed in the case of indels. It can be argued that these mutations could bring about the misreading of results on the basis of size separation. For instance, a 6 bp insertion in the flanking region of a $(CA)_{20}$ allele would allow co-migration with a $(CA)_{23}$ allele that did not have an insertion (Grimaldi & Crouau-Roy, 1997). Conversely some studies have shown that co-migrating bands are correctly homologous compared to the flanking sequences and microsatellites from which it was obtained (Steinkellner, Lexer, Turestschek & Glössl, 1997). Furthermore, there has been less phylogenetic analysis that have utilized the calculation of allele frequencies at each locus probably because of the problems faced in data coding and computing the distances (Buth & Murphy, 1999).

The diversity of microsatellites is detected by amplifying DNA in a PCR reaction. Primers are used to amplify the microsatellites loci. Primers are conserved sequences flanking a sequence of interest and are used for amplification in PCR. The resulting DNA amplicons are separated according to size (based on the number of repeats) on an acrylamide or agarose gel electrophoresis. Upon size separation, diversity of the amplicons is detected by any of the three methods: 1. staining the gels with ethidium bromide (used to stain DNA), 2. Using radioactively labeled primers and autoradiography of gels or 3. Using primers which are fluorescently labeled and running the PCR products on a DNA sequencing machine. Variation in the microsatellites alleles with different number of repeats are detected by the difference in the banding patterns displayed on the gel. Based on the banding patterns, genetic diversity

calculations are then performed to investigate the extent of the diversity. Phylogenetic analysis of microsatellites polymorphism is performed in either of the two methods; 1.by regarding alleles as characters when scored for presence or absence and calculating either distance or using character measures, and 2. By regarding allele frequency at loci as characters and calculating distance measures (Robinson & Harris, 1999). Allozyme phylogenetic analysis is conducted in a comparable way to that of microsatellites.

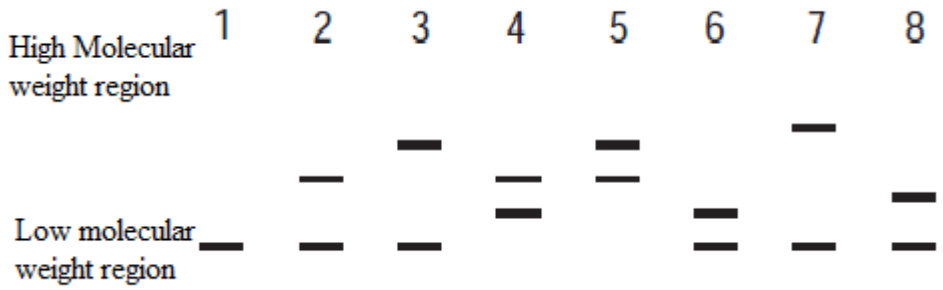


Figure 2: An illustration of the gel electrophoresis showing polymorphism in microsatellites (Modified from O'Connell & Wright, 1997)

2.3.3.2 Polymerase Chain Reaction sequencing

ITS

Internal Transcribed Spacers are sequences that are found in eukaryotic rRNA genes between the 18S and 5.8S rRNA coding regions, and between the 5.8S rRNA and 28S coding regions. ITS1 is located between the 18S and 5.8S rRNA coding regions while ITS2 is found between the 5.8S rRNA and 28S coding regions (Liu & Schardl, 1994).

The ITS1 and ITS2 regions are useful for phylogenetic analysis among related species or among populations within a species owing to their high evolution rate. The analysis of ITS1, 5.8S and ITS2 regions has led to the identification of species specific markers in non-coding regions (ITS1and ITS2) and coding regions (5.8S) that are vital in distinguishing specific species from impurities (Hershkovitz & Zimmer, 1996; Jobes & Thien, 1997; Liston, Robinson, Oliphant & Alvarez Buylla, 1996; Liu & Schardl, 1994). ITS region analysis has also been utilized in discriminating between Angiosperms, Bryophytes, algae and fungi (Jobes & Thien, 1997).

The homology in the 5.8S rRNA gene nucleotide sequence in closely related species, the absence of homology in non-related species, and the diversity in the ITS1 and ITS2 segments of related species supports the use of the ITS region in quantifying similarities between species (Chatterton, Hsiao, Asay, Wang & Jensen, 1992) .Their application is further not limited to distinguishing between fungi that is pathogenic to plants and nonpathogenic fungi (Jobes & Thien, 1997). ITS region analysis utilizes the PCR technique (Figure 3) because any genomic region can be amplified, sequenced and analyzed in several organisms eliminating the necessity to clone and isolate large amounts of ultra-pure genomic DNA (Schlötterer, 2004). Sequencing involves investigating the nucleotide sequence of the PCR- amplified DNA segment using specific primers.

Polymerase Chain Reaction sequencing (PCR-sequencing) is a technique used to investigate the order of the nucleotide bases in a DNA strand. This provides detailed information regarding the precise order of the DNA bases (T, A, G, C) that is required

for analysis. Two methods of sequencing were independently developed in 1974 by two teams (1. led by Maxam & Gilbert, and 2. led by Sanger). The Maxam and Gilbert method utilizes the chemical cleavage protocol while the Sanger method is similar to the natural process of DNA replication. Chemical degradation method and chain termination method are the respective names designated to the two methods. Both approaches were widespread and earned them the 1980 Nobel Prize which was shared by both teams. Sanger's method became the standard because of its practicality (Kumar *et al.*, 2009) and was used in this study.

Sanger's chain termination method is alternatively called the Sanger's dideoxy sequencing method. This procedure is dependent on the idea that single stranded DNA molecules that differ in length by a single nucleotide can be separated from one another by polyacrylamide gel electrophoresis. In addition, the technique utilizes modified bases that lack the -OH at the 3' carbon atom of de-oxy ribose sugar called dideoxy nucleotides. A dideoxynucleotide such as dideoxythymidine triphosphate (ddTTP) is added to the growing DNA strand until chain elongation stops due to the absence of 3' -OH for the next nucleotide to be attached. The technique is started by annealing a primer to the PCR products, followed by separating the mixture into four subsamples. DNA is then replicated in vitro by adding the four deoxynucleotides (adenine, cytocine, guanine and thymine), a single dideoxynucleotide (ddA, ddC, ddG or ddT) and the enzyme DNA polymerase to each reaction. Elongation of sequence occurs provided that deoxynucleotides are assimilated in the newly synthesized DNA strand (Kumar *et al.*, 2009). DNA replication is terminated upon the integration of a dideoxynucleotide. Each

of the four subsamples comprises sequences of varying length terminated due to the random incorporation of dideoxynucleotides to the many DNA molecules during the reaction at any incidence of the particular dideoxy base used in the subsample. The resulting fragments in each of the four subsamples are separated by gel electrophoresis.

Although the ITS region has been widely studied for diversity in various plants, animals, insects, pathogens, algae and fungi, there has been less literature on the analysis of ITS regions in *P. glaucum.* However ITS region analysis has been performed on other members of the *Poacea* family. Hence, the pieces of literature below will mainly focus on the analysis of ITS region concerning some members of the *Poacea* family, and in fungi in which it has been widely studied.

Martel *et al.* (2004) analyzed the ITS sequences of 16 *Pennisetum* species obtained from the five morphological sections (*Penicillaria, Brevivalvula, Gymnothrix, Heterostachya and Eu-Pennisetum*) of the genera *Pennisetum*. Their findings indicated that an alignment of ITS1 and ITS2 from *Pennisetum* species except for the hexaploid *Pennisetum Pedicellatum* which revealed a heterogenic region needed 46 gaps composed of 1bp to 4bp in size and had 148 variable sites that are perhaps a result of nucleotide substitution. Based on infrageneric phylogeny of *Pennisetum, Pennisetum* formed a monophyletic grouping with *Cenchrus ciliaris* and were both paraphyletic which is in agreement with literature on the analysis of chloroplast DNA sequences and inflorescence development traits (Doust & Kellogg, 2002; Giussani, Cota-Sánchez, Zuloaga & Kellogg, 2001).

Harlan and De Wet (1971) revealed that the primary, secondary and tertiary gene pools may possibly be genetically arranged on the basis of the ease with which the wild germplasm could be introgressed into the cultivated species. This observation is congruent with Martel *et al.* (2004) who established that *P. glaucum* and its wild relatives *Pennisetum glaucum* ssp. *monodii* forma *mollissimum* and forma *violaceum* clustered together by a 100% bootstrap support. These species showed high level of similarity due to their low sequence differentiation owing to them belonging to the same genetic species as described by gene flow, cytogenetic and molecular studies (Chowdhury & Smith, 1988; Clegg, Rawson & Thomas, 1984; Khalfallah, Sarr & Siljak-Yakovlev, 1993; Martel, Ricroch & Sarr, 1996; Robert, Lamy & Sarr, 1992;).

Chatterton *et al.* (1992) studied the ITS nucleotide sequence of rDNA in diploid wheat, *Triticum speltoides* L. (Tausch) Gren. ex Richter (*Gramineae*) and found that the 5.8S rRNA genes in diploid and hexaploid (*T. aestivum L.*) wheat are indistinguishable but vary from that of mung bean at 14 sites. In 1994, Sun, Skinner, Liang and Hulbert conducted a Phylogenetic analysis of 15 sorghum accessions and related taxa using ITS sequences of nuclear ribosomal DNA. The 15 accessions consisted of one from *Chaetosorghum* and one from the *Heterosorghum*, four from *Parasorghum*, two from *Stiposorghum*, and seven from the sorghum section (one accession from each of the *Sorghum propinquum* and *Sorghum halepense*, and five races of *Sorghum bicolor).* Their results showed that *Sorghum propinquum, Sorghum halepense* and *Sorghum bicolor* are highly similar to the cultivated sorghum and are perhaps its closest wild relatives. Sun *et al.* (1994) also argued that *Chaetosorghum* and *Heterosorghum* are

more comparable to sorghum than *Parasorghum, Sorgum bicolor* is most probably more related to *Sorghum nitidum* and that diversity is high in the sorghum group.

Zhang *et al.* (2002) investigated the origin and evolution of tetraploid wheat based on the ITS sequences of nuclear ribosomal DNA. The analysis of ITS1 and ITS2 regions of 24 clones from wheat species; *Triticum turgidum* (AABB) and *Triticum timopheevii* (AAGG), *Triticum monococcum* (AA), *Triticum urartu* (AA), and five species in *Aegilops* section. *Sitopsis* (SS genome) showed that *Aegilops speltoides* was distinctive compared to other species in *Aegilops* section *Sitopsis* and was probably the donor of the B and G genomes to tetraploid wheat.

Goryunova, Chikida, Gori and Kochieva (2005) carried out a research on the analysis of ITS nucleotide sequence polymorphism of ribosomal genes in diploid *Aegilops* (L.) species. Twenty two samples of five diploid *Aegilops* species were studied and it was observed that different *Aegilops* species exhibited ITS nucleotide sequence substitutions that were species specific. In addition, diploid *Aegilops* species revealed intraspecific variations in ITS structure in the same sample presumably as a result of variations in genomes of respective plants or the existence of numerous types of ribosomal genes in the genome of one plant. However, polymorphism in interspecific and intraspecific ITS nucleotide sequences of rDNA was particularly low.

Asay, Hsiao, Chatterton, Jensen and Wang (1992) studied the nucleotide sequence of the ITS region of rDNA in mountain rye, *Secale montanum* Guss.(*Gramineae*). The results indicate homology in the 5.8S rRNA gene in Mountain rye and wheat *(Triticum aestivum)* and variations in the nucleotides at position 13 in ITS1 and position 30 in

ITS2. Mountain rye was found to have 601bp in the ITS region while bread wheat *(Triticum aestivum)* had 602 bp. These results complement another study by Chatterton, Hsiao, Asay, Jensen and Wang (1992) in which they found similar results by studying the nucleotide sequence of the ITS region of rDNA in barley, *Hordeum vulgate L*. (*Gramineae*). They argued that the 5.8S rRNA gene of barley has complete homology with that of wheat (*Triticum aestivum L*.), mountain rye (*Secale montanum Guss*), and rice. Nevertheless, the ITS1 and ITS2 regions in barley have been found to reveal minor variations in comparison to wheat.

Dillon *et al.* (2004) analyzed the *Sorghum laxiflorum* and *Sorghum macrospermum*, the Australian native species most closely related to the cultivated *S. bicolor* based on ITS1 and *ndh*F sequence analysis of 25 sorghum species. The analysis revealed that *Sorghum laxiflorum* and *Sorghum macrospermum* are Australian species closely related to the cultivated sorghum. In addition, there was no noticeable variation between the sections *Chaetosorghum* and *Heterosorghum*, *Para-sorghum* and *Stiposorghum* while *Eu-sorghum* emerged as the divergent section from the five known sorghum sections. These results complement other studies (Dillon, Lawrence & Henry, 2001; Dillon, Lawrence, Henry & Price, 2007) performed on sorghum.

In a study by O'Donnell (1992) on the ribosomal DNA of Internal Transcribed Spacers (ITS) in the phytopathogenic *Ascomycete*, *Fusarium sambucinum* (*Gibberella pulicaris*) disclosed variation. The ITS regions of eighty six strains of heterothallic phytopathogenic filamentous fungus, *Fusarium sambucinum* from different geographic locations were analyzed for variations and disclosed three ITS types (A, B, C) that

contain a low variation ranging from 0-2.3 %. It was also observed that each of the 86 strains contained the three ITS types. Furthermore, it was observed that AT insertion/deletion events are prevalent in ITS1 while GC insertion/ deletions dominate in the ITS2 region. These findings are congruent with various literatures (Hirata & Takamatsu, 1996; Saunders & Druehl, 1993; Takamatsu, Hirata & Sato, 1998).

Korabečná, Liška and Fajfrlik (2003) analyzed eleven species of clinical isolates of fungi and detected within species variation in the ITS and 5.8S rRNA gene Region. Variations were identified among the 11 species of *Candida albicans*, *Candida catenulata*, *Candida colliculosa*, *Candida glabrata*, *Candida kefyr*, *Candida melinii*, *Candida parapsilosis*, *Candida guillermondii*, *Candida solanii*, *Candida tropicalis* and *Saccharomyces cerevisiae*. These findings are in agreement with Karvonen, Szmidt and Savolainen (1994), Reed and Philips (2000), and Velegraki, Logotheti, Pyrri and Kapsanaki-gotsi (2001), who argued that ITS regions evolve quickly resulting in both intra-species and interspecies variations.

Advantages and disadvantages of PCR sequencing

PCR sequencing offers the ultimate measurement of genetic variation because it reveals the nucleotide variation of amplified fragments. Additionally, PCR sequencing makes it possible to target specific sequences in a wide range of species using universal primer pairs for chloroplast, mitochondria and ribosomal genomes. PCR sequencing is highly reproducible, can detect homologous variations and requires low amounts of target

DNA, and most of the technical procedures are amenable to automation. Unlike in the past were phylogenetic analysis were based on nuclear ribosomal DNA and chloroplast DNA due to their high copy number, low copy and single nuclear DNA markers have proved to be the modern powerful tools for phylogenetic analyses (Mort & Crawford, 2004; Small, Cron & Wendel, 2004). Low-copy nuclear markers mostly overcome the difficulty of uniparental inheritance presented by plastid markers and concerted evolution found in nuclear ribosomal DNA (Arnheim, 1983; Corriveau & Coleman, 1988). This has restricted the application of nuclear ribosomal DNA and plastid markers, and their reliability in phylogenetic studies (Bailey, Carr, Harris & Hughes, 2003). In addition, Kumar *et al.* (2009), Small *et al.* (2004), and Wolfe, Li and Sharp (1987), argued that low-copy nuclear markers are preferred because; evolution is independent of paralogous sequences and tend to be stable in position and copy number, and they display higher rates of evolution compared to cpDNA and nrDNA markers which has promoted their wide spread application in closely related species.

PCR sequencing of the whole genome is expensive; hence only segments of the genomes can be analyzed. A huge cost of synthesizing primers is incurred if suitable primers for the segment of interest are not available and requires expensive equipment's. In addition, sequencing is time consuming which influences most studies to concentrate on particular segments leading to less exploration of the entire genome. Furthermore, analytical procedures involving the visualization of sequences by polyacrylamide gel electrophoresis and autoradiography are laborious and technically demanding (Kumar *et al.*, 2009).

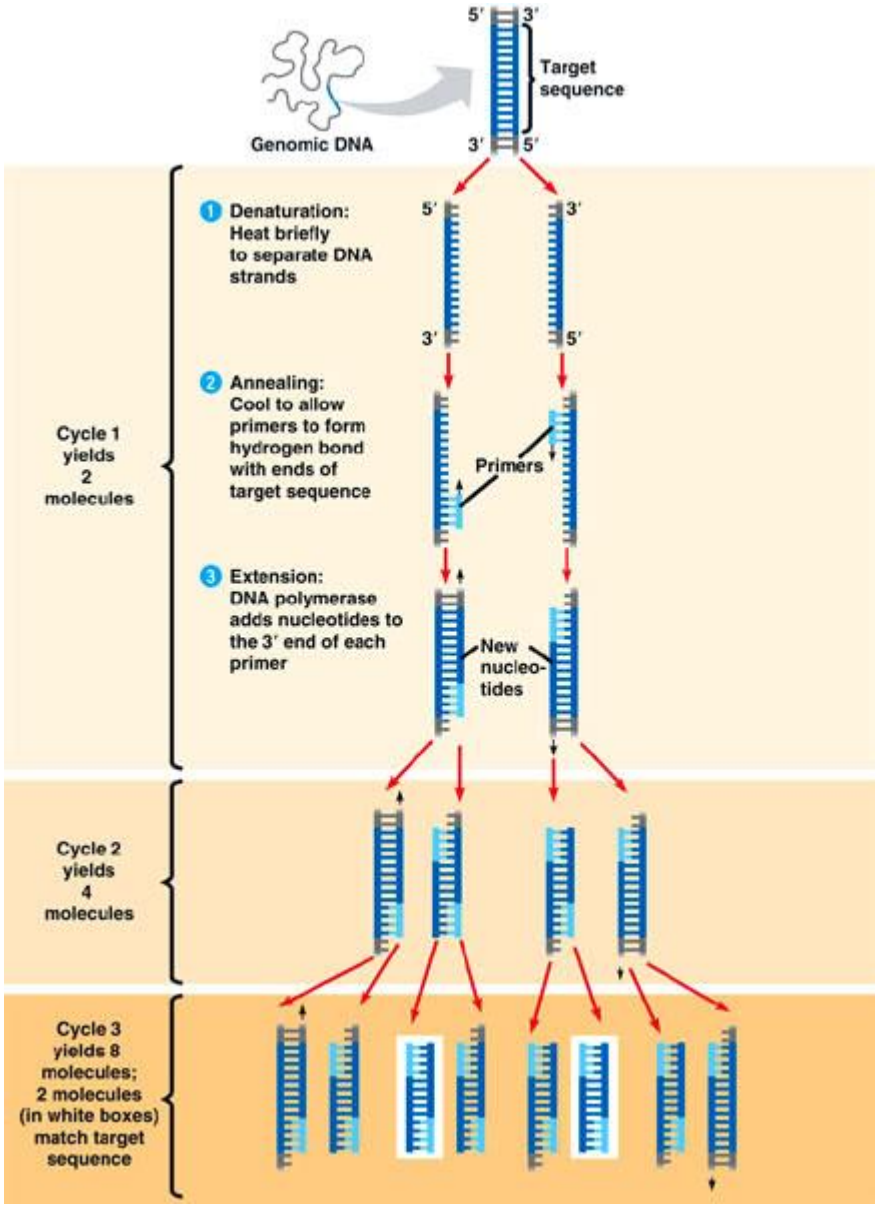


Figure 3: An illustration of the PCR technique. Adapted from http://www.bio.miami.edu/~cmallery/150/gene/Taq.htm. Retrieved on the 10th of June, 2015

2.3.3.3 Random Amplified Polymorphic DNA

Random Amplified Polymorphic DNA (RAPD) is a technique that utilizes random primer sequences to amplify DNA as opposed to microsatellites which employ specific ones. RAPD makes use of primers (mostly 10-20 base pairs in length) to amplify nuclear DNA samples generating multiple DNA fragments. The RAPD PCR products are separated on an agarose gel or by sequencing gels. Each RAPD primer amplifies multiple fragments of size range 100-200 base pairs. The detected banding patterns are recorded as presence-absence if variation exists at the priming sites in the DNA. RAPD analyze multiple loci in the genome and eliminate the need to sequence the entire genome and the establishment of specific primers. RAPDs assume a dominant inheritance for present bands and a recessive inheritance for absent bands. Polymorphism is noticed and scored as the presence and absence of a fragment. The presence-absence scores are then utilized in genetic diversity calculations.

Polymorphism is a result of deletion, nucleotide insertion or substitutions in the genomes which are detected by the high resolution displayed by the RAPD technique (Chakauya, 2002). Owing to its high resolution, RAPDS have been employed in various studies such as; finger printing in molecular ecology to determine taxonomic identity, analyzing kingship relationships, creation of probes that are specific, detection of interspecific gene flow and to analyze mixed genome sample, genetic mapping and generation of linkage maps (Hadrys, Balick & Schierwater, 1992).

Govindaraj, Selvi, Prabhu and Rajarathinam (2009) investigated the diversity of 20 *P. glaucum* genotypes and found that the genotypes clustered into 8 different groups after analysis. This study showed the ability of RAPD markers to correctly disclose genetic variability at DNA level among *P. glaucum* genotypes. RAPD is an effective tool that can be used to select the markers which are well distributed throughout the genome and to get an insight into the genetic resources that can be employed in future hybridization programs for *P. glaucum* germplasm management.

Chowdari *et al.* (1998) detected polymorphism in 36 landraces and 22 cultivars of *P. glaucum* using RAPD analysis. They argued that RAPDs can identify a greater number of genotypes by including more primers as demonstrated in their study. RAPD analysis revealed a high degree of genetic diversity among the cultivars as well as between the landraces. This is in agreement with Chandra-Shekara, Prassana, Bhat and Singh (2007) who found genetic diversity in elite pearl millet inbred lines using more RAPD primers.

Advantages and disadvantages of RAPDs

RAPD technology is inexpensive and has a wide potential to be applied in molecular biology and ecology. RAPDs are easily detected (compared to RFLPs), non-laborious, and can be effective in the analysis of highly heterozygous genomes in which they rather contend with RFLPs (Williams, Kubelik, Livak, Rafalski & Tingey, 1990). RAPDs can be employed in non-invasive sampling and are able to analyze multiple loci

in the genome without the need to sequence the entire genome. In addition, it utilizes nonspecific random primers to assay many loci. Besides, the RAPD technique is quick, easy, does not require prior sequence information and has a high power of resolution (Chakauya, 2002).

The degree of specificity is influenced by the size of the primer. Presumably, short length primers are likely to amplify an irrational large number of sequences while longer primers may amplify less informative few sequences. Longer primers which are beyond the recommended (10-20 base pairs) are possibly less specific due to increased frequencies of non-specific primer annealing. This has an implication because the occurrence of non-reproducible amplification patterns might increase. Studies encourage the use of standard RAPD conditions of 10-20 base pair primer length, a G+C content of the primers that is similar to that of the genome being analyzed in order to maximize the frequency of specific primer-template binding thereby improving the quality of the amplification results (Hadrys *et al.*, 1992).

RAPD is a PCR dependent technique and is thus sensitive to the reaction conditions. Alterations in reaction conditions affect the reproducibility of the PCR products (Williams *et al.*, 1990). PCR reaction conditions influencing the reproducibility of RAPD products are 1. The temperature profile shape which is a thermal cycler property that needs to be uniform, 2. Type of polymerase used and 3. Mg^{2+} concentrations. Reproducibility is certain under uniform reaction conditions. In addition, *Taq* or DNA concentrations also influence the amplification results (Hadrys *et al.*, 1992).

The possibility of co-migration in the RAPD technique compromises its reliability. RAPD technique assumes that amplified fragments are distinct with no possibility of amplifying distinct fragments which co-migrate on the gel due to similar size. Hadrys *et al.* (1992) argued that eluting individual PCR products from gels and reprobing the products *via* Southern analysis or instead using polyacrylamide gel electrophoresis to increase the resolution of band separation certainly detects co-migration in the RAPD technique.

Certain fragments are unclear and give problems with scoring. In some cases, non-reproducible bands result from non-specific priming or from heteroduplex formation between related amplification products (Williams *et al.*, 1990). In such instances, these fragments are not relevant for the study or use as genetic markers. Nevertheless, RAPD seems to produce reproducible bands that are mostly tested two or more times in order to confirm reproducibility upon which they are easily scored. In certain cases, PCR reactions produce amplification products in the absence of template DNA in the PCR reaction mix. These products are no longer detected once the DNA being investigated is added to the mix (Hadrys *et al.*, 1992).

RAPDs are dominant markers, do not favor the detection of heterozygosity and are the least informative of all the known DNA markers. The effect is that a fragment representing a homozygous genotype (AA) cannot be distinguished from a heterozygous genotype (Aa). This presents difficulties in exploring population genetics using RAPDs and marks RAPDs as less competitive in comparison to analytical tools that involve sequence information or single locus probe finger printing technologies.

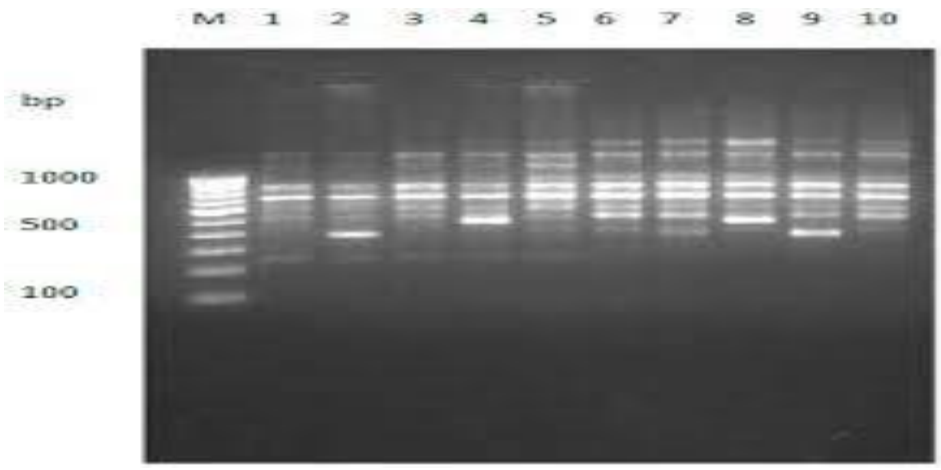


Figure 4: An illustration of an electrophoresis gel of the RAPD technique. Adapted from http://www.bio.miami.edu/~cmallery/150/gene/Taq.htm. Retrieved on the 10th June, 2015

2.3.3.4 Amplified Fragment Length Polymorphism

Amplified Fragment Length Polymorphism (AFLP) is a PCR based technique that is similar to RAPD (Vos *et al.*, 1995) because both use random sequences in their analysis. AFLPs random segments are amplified under high stringent PCR conditions which enables primers with a single base difference to produce distinct band patterns. AFLP involves the cutting of small DNA segments aided by a restriction endonuclease (enzyme) resulting in sticky and staggered ends. Short synthetic DNA fragments of known sequence which are referred to as adapters (10-30 double stranded bases) are then ligated to the cut sticky staggered ends. Amplification in a PCR reaction is then performed using primers that are complementary to those of the known adapter

sequence with added selective nucleotides (Frankham *et al.*, 2002; Mueller *et al.*, 1996). Restriction and ligation occur simultaneously and the products are scored as present (dominant) or absence (recessive) which is similar to that of RAPDs.

In population and conservation genetics, AFLP have been used to mainly investigate genetic variation below the species level. Investigating variations within populations and the population structure at species level have been widely explored (Arens, Coops, Jansen & Vosman, 1998; González *et al.*, 1998; Majer, Lewis & Mithen, 1998; Travis, Maschinski, & Keim, 1996; Winfield *et al.*, 1998) by studying parameters such as estimation of *F*ST analogs and genetic variation within populations with the utility of AFLP. In addition, AFLP markers have been useful in assessing gene flow and dispersal, outcrossing, introgression and cases of hybridization (Mueller & Wolfenbarger, 1999). AFLP have also been employed in paternity analysis because of the technique's competence in discriminating individuals in a population (Krauss, 1999).

Advantages and disadvantages of AFLP

AFLPs have a high resolution and are able to analyze the genome without prior genome sequence information. Only small amounts of DNA are required in a PCR reaction involving AFLPs and the technique is easy and employs less labor. In addition AFLPs are highly reproducible and are able to generate DNA finger prints irrespective of the DNA source. Furthermore, AFLP markers efficiently depict genetic variations in

individuals, populations and lineages that are evolving independently (Mueller & Wolfenburg, 1999). The high level of polymorphism produced by AFLPs appears to be the driving force towards their selection for analysis.

AFLPs present challenges when analyzing higher taxonomic levels due to the high variability found in its markers. This causes a reduction in the similarity between distant taxa paving way to coincidence. Comparable to RAPDs, AFLPs are also dominant markers and are therefore not able to discriminate between homozygous and heterozygous genotypes. In addition, scoring of AFLPs tends to be tedious and present problems with size homoplasy. Size homoplasy is seemingly the utmost challenge in the use of AFLPs owing to the hypothesis that co-migrating bands are homologous which may be due to the occurrence of multiple bands (increasing the possibility of a pair-wise homology of all of them) (Mace, Gebhardt & Lester, 1999).

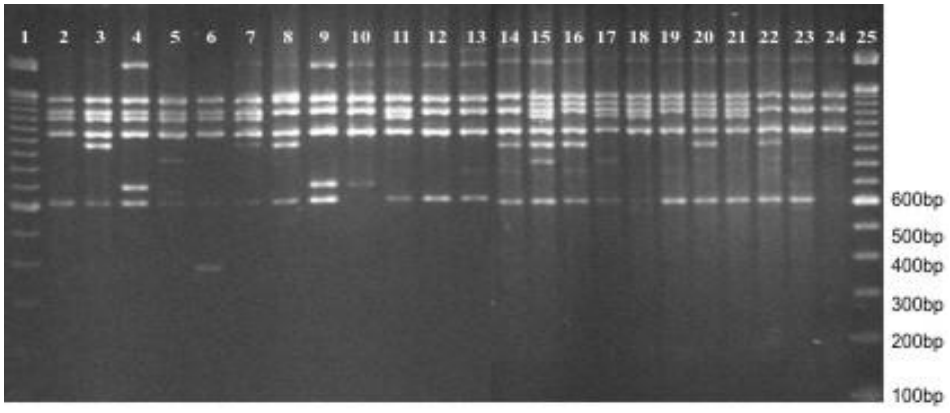


Figure 5: An illustration of an electrophoresis gel of the AFLP technique. Adapted from http://www.bio.miami.edu/~cmallery/150/gene/Taq.htm. Retrieved on 10 June, 2015

2.3.3.5 Restriction Fragment Length Polymorphism

Restriction Fragment Length Polymorphism (RFLP) is a technique used to track a specific DNA sequence as it is inherited by other cells. RFLP is able to distinguish individuals by analyzing fragments of DNA resulting from cleavage. RFLPs have various applications such as; to investigate the source of a DNA sample in paternity cases or criminal cases, determining the recombination rates and establishment of genetic maps and assessment of an individual's disease status in epidemiological studies. The method involves the purification of DNA and cleaving it with a restriction enzyme and running a gel to separate fragments of different sizes (Frankham *et al.*, 2002).

The strands of DNA are separated and transferred to a membrane which is then dried. The membrane is then transferred to a solution that comprises multiple single copies of radioactively labeled DNA (probe) segments to locate the locus under investigation. Upon complementary base pairing, the membrane is washed to remove unhybridized single stranded probe molecules. The membrane is then dried and autoradiographed. The presence of polymorphism in the DNA sequence at the restriction site is then detected by the different sizes displayed on the autoradiograph. This is due to the difference in distance between the cleaved sites of DNA.

Gepts and Clegg (1989) analyzed polymorphism in chloroplast DNA, nuclear ribosomal RNA, and alcohol dehydrogenase (ADH) sequences in 25 wild and 54 cultivated accessions of *P. glaucum* using Restriction-Fragment-Length Polymorphisms

(RFLP). Their results show no polymorphism in chloroplast DNA sequences. However, polymorphism was observed to exist in nuclear rRNA in the non-transcribed spacer region in which higher polymorphism was observed in the wild *P. glaucum* as compared to the cultivars which rather had homogenous sequences.

Advantages and disadvantages of RFLP

RFLPs are co-dominant markers which makes them more informative as compared to dominant markers such as RAPDs. The RFLPs technique is robust, enables the detection of polymorphism in large genomic regions and is relatively inexpensive. In addition, variations in known genes can be followed by the use of RFLPs. RFLPs have good transferability between laboratories and no sequence information is required. RFLP is also recommended for phylogenetic analysis of related species because it is centered on sequence homology. This technique displays a high discriminatory power which enables analysis at population and/or species level using a single locus probe, or individual level by using a multi-locus probe (De Vicente & Fulton, 2003). Moreover, RFLPs analysis is simple and can be readily applied as long as appropriate probes are available. Genetic linkage maps have been constructed using this technique.

The RFLPs technique requires large amounts of DNA and cannot therefore be used in non-invasive sampling. Moreover, the RFLPs technique is only able to detect a minor proportion of the total genetic variation in the DNA sequence as compared to sequencing the locus which reveals much information on its genetic variation (Nei,

1987). This method does not allow automation and can only detect low levels of polymorphism in some species. Besides a suitable probe library is needed and it is a long procedure especially when single copy probes are employed. In addition, the technique is costly, demanding and analysis is only done on a few loci for each assay. Probes have to be spread across all laboratories that are collaborating and various probe-enzyme combinations might be necessary (De Vicente & Fulton, 2003).

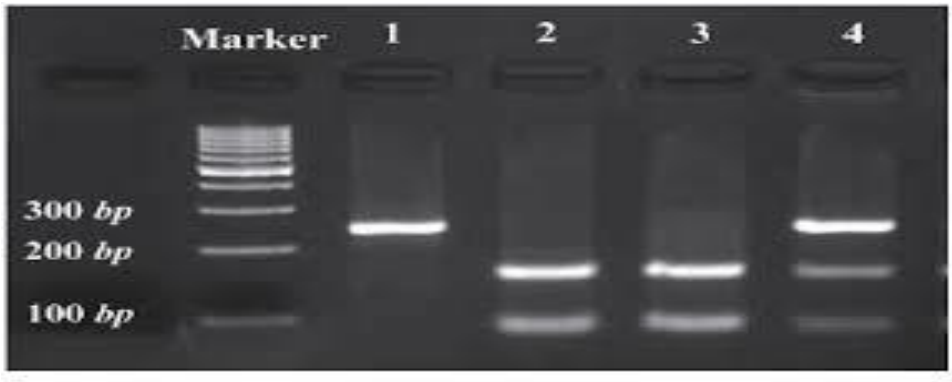


Figure 6: An illustration of an electrophoresis gel of the RFLP technique. Adapted from http://www.bio.miami.edu/~cmallery/150/gene/Taq.htm. Retrieved on the 10th of June, 2015

2.3.3.6 Single Nucleotide Polymorphisms

Single Nucleotide Polymorphisms (SNPs) are positions in the DNA of a species at which two or more alternative bases occur at a frequency more than 1% (Frankham *et al.*, 2002). SNPs are the wide spread class of DNA markers because most polymorphisms are the outcome of point mutations and SNPs are the most abundant molecular markers in the genome (Mondini *et al.*, 2009). SNPs and flanking sequences are identified by library construction and sequencing or through the screening of readily

available sequence databases and have been used to saturate linkage maps in order to trace important characters in the genome (Cho *et al.*, 2000). SNPs mostly occur in the non-coding regions of the genome while its existence in coding regions can generate either non-synonymous mutations with the consequence of an amino acid sequence change or synonymous mutations that do not amend the amino acid sequence (Sunyaev *et al.*, 1999). However, synonymous alterations can modify mRNA splicing causing phenotypic variations (Richard & Beckman, 1995). SNPs genotyping evaluation depends on allele-specific hybridization, oligonucleotide ligation, and primer extension or invasive cleavage (Sobrino, Briona & Carracedoa, 2005).

Advantages and disadvantages of SNPs

SNPs analysis is able to detect variations as low as a single base difference. This technique is possibly effective when analyzing diversity in crops where it is difficult to find polymorphisms such as in the cultivated tomato (Kumar *et al.*, 2009). SNPs provide the possibilities of automation especially when the SNPs have been located and suitable primers are available. In addition, the wide spread occurrence of SNPs across genomes and variable distribution among species make them the preferred effective mode of analyzing genetic diversity. SNPs have various genotyping methods comprising DNA chips, allele-specific PCR and primer extension approaches which are predominantly preferred due to their high data throughput. In addition, the technique

allows automation and can be used for rapid identification of crop cultivars and construction of ultra-high-density genetic maps.

SNPs technique needs sequence information in order to synthesize PCR primers that are allele specific or oligonucleotide probes. Multiplex PCR and hybridization to oligonucleotide microarrays or analysis on automated sequencers are frequently used to investigate the occurrence of SNPs in high sample throughput. This increases the cost of the technique and has discouraged Gene Banks from regularly utilizing SNPs markers.

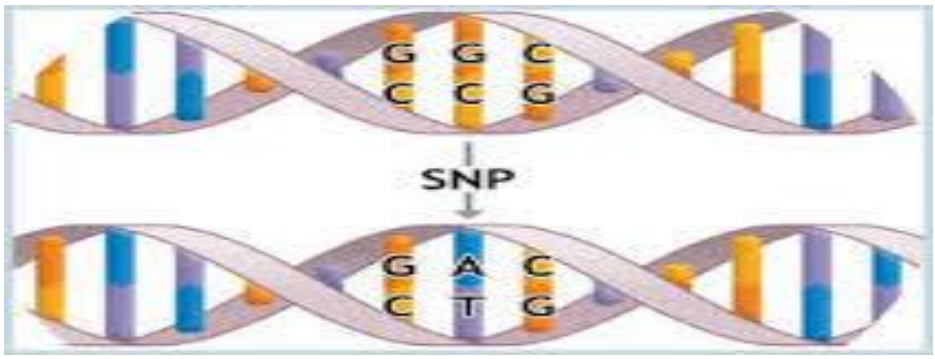


Figure 7: An illustration of the occurrence of a Single Nucleotide Polymorphism (SNP). Adapted from http://www.bio.miami.edu/~cmallery/150/gene/Taq.htm. Retrieved on the 10th June, 2015

Table 1: Comparison of different characteristics of the most frequently used molecular marker techniques (Adapted from Mondini *et al.*, 2009, pp. 24).

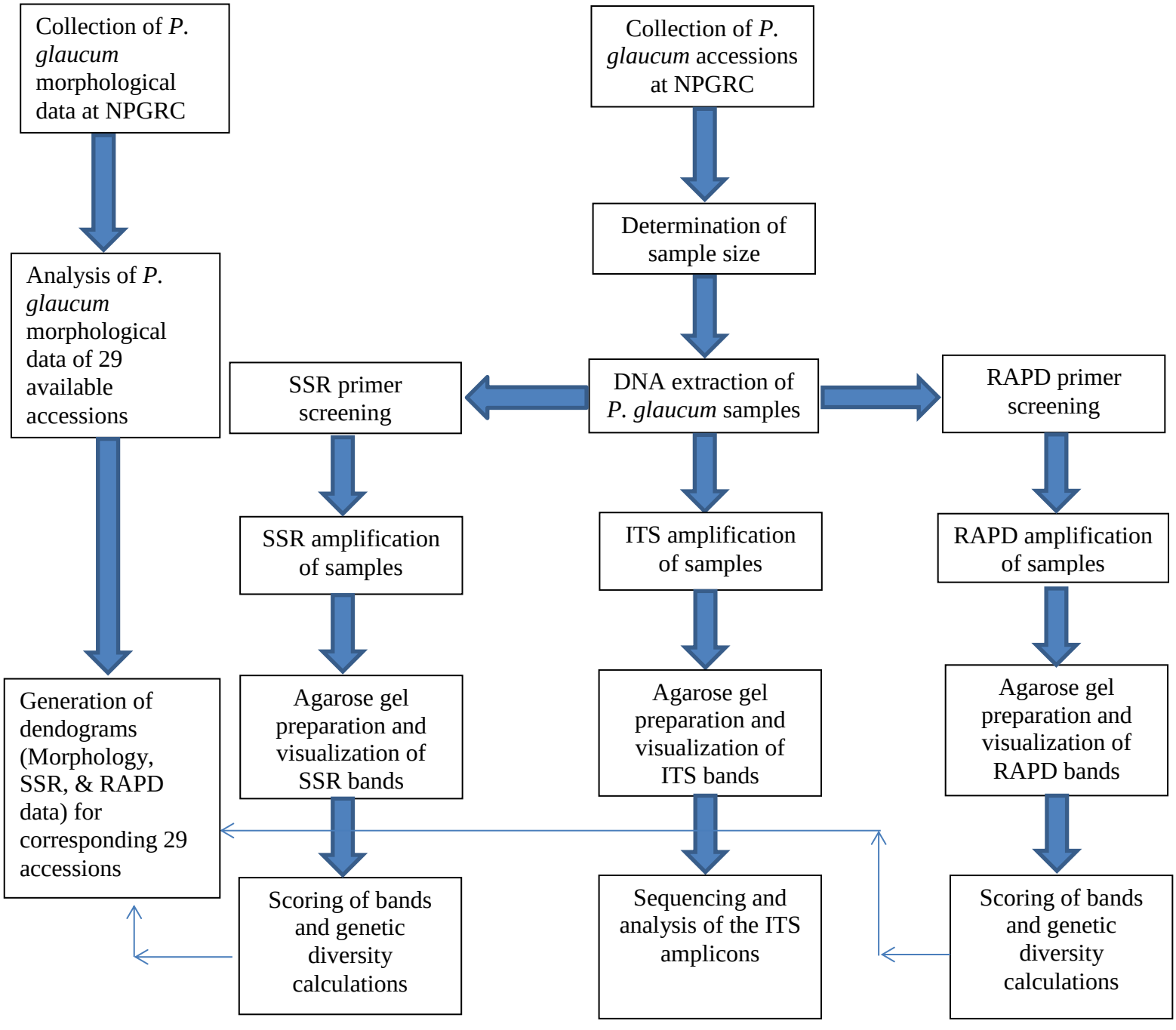
Molecular marker	RFLP	RAPD	AFLP	SSR	SNP
Degree of polymorphism	M	M	M	M	H
Locus specificity	Y	N	N	N	Y
Dominance/ Co-dominance	C	D	D	C	C
Ease of replication	H	L	H	M	H
Abundance	H	H	H	M	H
PCR-based	Y	N	N	N	Y
Quantity of DNA required	H	L	M	L	L
Automation	N	Y	Y	Y	Y
Cost per assay	H	L	M	L/M	L
Technical requirement	H	L	M	L/M	M

Key: H = High; L = Low; M = Medium; Y = Yes; N = No

CHAPTER 3: MATERIALS AND METHODS

3.1 Research design

The outline of the research design that was used to assess the genetic diversity in this study is displayed below.



3.2 Sampling strategy

All the *P. glaucum* accession, which amounted to 1441 landraces, were obtained from the Namibian National Plant Genetic Resource Centre (NPGRC) and used in the present study. However, of the 1441 landraces, only about 680 have been morphologically characterized by the Namibian NPGRC in an ongoing effort to characterize the entire landraces and a sample of 96 landraces were genetically characterized in this study. The Namibian NPGRC has a total of about 1441 accessions of *P. glaucum.* These accessions were collected in 1991 by the International Crops Research Institute of Semi-Arid Tropics (ICRISAT) and have been multiplied and characterized for gene banking purposes (R. Moses, personal communication, April 17, 2014). From a total of 1441 accessions, the sample size of 94 landraces used in the study was calculated using the following formula (Yamane, 1967):

$$\mathbf{n} = \frac{\mathbf{N}}{\mathbf{1} + \mathbf{N(e)^2}}$$

$$\mathbf{n} = \frac{1441}{1 + 1441\ (0.1)^2}$$

$$\mathbf{n} = 94$$

Where n is the sample size, N is the population size and e is the level of precision (90% confidence level and 10% level of precision). Sampling was done in such a way that all the regions from which the *P. glaucum* accessions were collected were represented in the calculated sample size. This was achieved by Random sampling from each region based on proportionate calculations. The proportions were calculated using the formula outlined below:

$$P = \frac{N_s}{N_T} \times n$$

Where P= proportion, N_s=Total number of accessions collected from a region, N_T= Total number of accessions in the entire population and n = the calculated sample size from the entire population.

Table 2: The calculated sample values of proportions from each regional.

Region	**Total Accessions per region**	**Total accessions in (Namibia)**	**Total accessions in (Namibia) calculated sample size**	**Regional sample size**
Karas	15	1441	96	1
Zambezi	292	1441	96	19
Okavango (east + west)	716	1441	96	48
Otjozondjupa	20	1441	96	1
Oshikoto	28	1441	96	2

Kunene	29	1441	96	2
Omusati	144	1441	96	10
Ohangwena	175	1441	96	12
Oshana	22	1441	96	1

A total sample size of 96 was used instead of 94 due to ease of use with the 96 well PCR-machine and to ensure that every region is represented since some regions had a low population. *P. glaucum* accessions were then planted in the green house at the University of Namibia (UNAM) in pots and left to grow for two weeks until the time of DNA extraction.

3.3 Genomic DNA extraction

P. glaucum genomic DNA was extracted at the Biological Sciences Department laboratories at the University of Namibia. DNA extraction was performed using the CTAB/chloroform-isoamyl alcohol protocol as described by Doyle (1991). For each sample, about 50 mg of leaf tissue was ground in liquid nitrogen using a motar and pestle and then transferred to respective 1.5 ml Eppendorf tubes. An amount of 500µl of CTAB buffer [1 M Tris, pH 8.0, 5 M NaCl, 0.5 M EDTA and CTAB powder (20g/l of buffer prepared)] was added to the Eppendorf tubes and incubated at 55 ºC for 1 hour. Thereafter, 500µl of 24:1 chloroform-isoamyl alcohol was added and mixed by shaking the tubes with the hands to form an emulsion. The tubes were then centrifuged for 10

minutes at 13,000 rpm resulting into three layers. The top layer which is the aqueous phase was then collected by pipetting and placed into new 1.5 ml Eppendorf tubes and 0.08 volumes of cold 7.5M ammonium acetate was added and mixed by inverting. An amount of 0.54 volumes of cold isopropanol was then added and mixed by inverting. The tubes were then kept in the freezer overnight and centrifuged the next morning at 13,000 rpm for 3 min. The supernatant was pipetted off from the tubes leaving the DNA pellet behind. A volume of 700µl of cold 70% ethanol was added and the tubes were inverted once to mix and centrifuged at 13,000 rpm for 1 minute. The liquid was then pipetted off from the tubes and 700 µl of 95% ethanol was then added to the tubes after which they were inverted once to mix and centrifuged at 13,000 rpm for 1 minute. The resulting DNA pellet was collected by pipetting the supernatant out and left to dry on a hot plate at 55 ºC. Upon drying, the DNA pellet was then suspended in 100 µl of double distilled water and stored in the freezer at -20 ºC until use. Prior to storage, the DNA concentration was determined using a Nano Drop 2000c spectrophotometer (Thermo Scientific, Massachusetts, USA).

3.4 Microsatellites data collection

3.4.1 Microsatellites primer screening

A total of 10 SSR primer (Neto, Von Pinho, Carvalho & Pereira, 2013) sets (Table 3) obtained from Inqaba Biotechnical Industries were screened for polymorphism. This was done by randomly selecting 2 DNA samples from each of the regions (Zambezi,

Karas, Kunene, Ohangwena, Okavango [East & West], Omusati, Oshana, Oshikoto and Otjozondjupa) involved in the origin of all the accessions. The primers were dissolved in nuclease-free water to make 100 µM stock solutions. The primer stock solutions were then diluted to form 20 µM aliquots of each primer pair. Extracted DNA samples with a concentration of 20 – 250 µg/µl were diluted to 10ng of DNA per µl for PCR amplification. A total of 18 DNA samples were then amplified in a 25 µl PCR reaction volume comprising 12.5 µl of 2x master mix, 1 µl of template DNA, 1 µl of a 1 µM concentration forward SSR primer, 1 µl of a 1 µM concentration reverse SSR primer and 9.5 µl of nuclease free water with thermo-cycling conditions of; initial denaturation step at 95 ºC for 4 min, followed by 35 cycles of denaturation at 94 ºC for 1 min, annealing ranged between 57 ºC and 62 ºC which was primer dependent for 1 min and an elongation at 72 ºC for 1 min. Final elongation was performed at 72 ºC for 4 min and held at 4 ºC. A 2% agarose gel electrophoresis with a 1kb generule DNA ladder was then conducted to determine the pair of primers that displayed polymorphism or monomorphism based on the amplification product band patterns of separation on the gel from the different DNA templates. The outcome of the screening was described as monomorphic, polymorphic or unable to amplify. The PCR was then conducted with the actual samples but using the primers that displayed a higher level of polymorphism.

3.4.2 Microsatellites amplification protocol

Samples (96) of DNA were amplified in a 25 µl PCR reaction volumes using primers that revealed polymorphism based on agarose gel patterns. Six (6) primer sets (3005, 3016, 3020, 3022, 3035 & 3039) were used to amplify microsatellites loci by combining 1 µl of template DNA, 1 µl of a 1 µM concentration forward SSR primer, 1 µl of a 1 µM concentration reverse SSR primer, 9.5 µl of nuclease-free water and 12.5 µl of 2x Dream *Taq* master mix which contained: Dream *Taq* DNA Polymerase, 2x Dream *Taq* buffer, dATP, dCTP, dGTP and dTTP of 0.4 mM each, and 0.4 mM $MgCl_2$. The reaction profile consisted of initial denaturation temperature of 94 ºC for 3 min, followed by 35 cycles of denaturation temperature at 94 ºC for 1 min, 1 min at respective annealing temperatures and 1 min extension temperature at 72 ºC. The final extension was then performed at 72 ºC for 4 min and lastly held at 4 ºC.

Table 3: SSR forward primer screened for polymorphism and their characteristics.

Primer	Sequence	Annealing temperature
3002F	AAAGTTACCGGGAGGGTAAAAA	58.95
3005F	CGCGGTGTTCTCACACAC	62.18
3009F	CTGTACCATGTGCGCTGATT	60.4
3013F	TGTGGGAGAGAGGAGAGTCC	64.5

3016F	TTGTGGCTGAAGAAGAGATCC	60.61
3020F	GTTCCATGGAGCTGGAAGTC	62.45
3022F	CTGGAAGTCCTTCTCGGTTG	62.45
3028F	ACGATTCTTCGTCGTTCCAG	60.4
3035F	GCCAAGGAGGTCAAGATCG	62.32
3039F	GGCACGAGGGGCTAAGTAA	62.32

3.4.3 Polymorphism Information Content

Genetic diversity calculations were performed at each locus to determine allelic Polymorphism Information Content (PIC) based on allele frequencies of the 96 accessions analyzed. The PIC values for each SSR primer were estimated by determining the frequency of alleles per locus as described by Sharma, Kantartzi and Stewart (2009) using the formula: PIC = 1 - $\sum x^2_i$, where x_i is the relative frequency of the i^{th} allele of the SSR loci. Markers were designated as being informative when PIC was ≥ 0.5. The most informative primers were then used for subsequent genetic diversity calculations.

3.5 RAPD data collection

3.5.1 RAPD primer screening

RAPD primers (12) (Govindaraj *et al*., 2009) shown below (Table 4) obtained from Inqaba Biotechnical Industries were screened for polymorphism in a similar way to that

described in section 3.4.1 under microsatellites screening. Samples (18) of DNA initially used for microsatellites primer screening were also used to screen for RAPD primers that could display polymorphism. The 18 DNA samples were amplified in a 25 µl PCR reaction volume containing 12.5 µl of 2x master mix, 2 µl of template DNA, 2 µl of a 1 µM concentration RAPD primers and 8.5 µl of nuclease-free water with PCR conditions of: initial denaturation step at 95 ºC for 3 min, the next step involved 40 cycles of denaturation at 94 ºC for 15 sec, annealing was reliant on primer specificity (Table 4) and an elongation at 72 ºC for 1 min. Final extension was carried out at 72 ºC for 7 min and held at 4 ºC. The RAPD PCR products were then visualized on a 2 % agarose gel electrophoresis and the results designated as monomorphic, polymorphic or unable to amplify based on the separation of the amplification products and banding patterns displayed on the agarose gel. The primers that could detect polymorphism were then employed to amplify the 96 samples of *P. glaucum.*

3.5.2 RAPD amplification protocol

Upon successful determination of the primers able to detect polymorphism, 96 DNA samples were then amplified using three (3) primers (OPN-16, OPE-09 & OPE-18) in a 25µl PCR reaction volumes containing 12.5 µl of 2x master mix, 2 µl of template DNA, 2 µl of a 1 µM concentration RAPD primers and 8.5 µl of nuclease-free water using the following PCR conditions: an initial denaturation step of 95 ºC for 3 min, 40 cycles of denaturation at 94 ºC for 15 sec, annealing was primer specific (Table 4) which was between 39 ºC and 43 ºC, an elongation at 72 ºC for 1 min and the final extension was

carried out at 72 ºC for 7 min and held at 4 ºC. A 2% agarose gel electrophoresis was then used to visualize the RAPD PCR products and the results were recorded for analysis. The RAPD bands were tested to confirm reproducibility by amplifying the DNA samples three times using the same reaction conditions. Only reproducible bands were considered for genetic diversity calculations.

Table.4: RAPD primers screened for polymorphism and their characteristics.

Primer	Sequence	Annealing temperature
OPL-11	ACGATGAGCC	39.5
OPAL-20	AGGAGTCGGA	39.5
OPAL-08	GTCGCCCTCA	43.6
OPAL-15	AGGGGACACC	43.6
OPM-16	GTAACCAGCC	39.5
OPN-16	AAGCGACCTG	39.5
OPF-10	GGAAGCTTGG	39.5
OPF-01	ACGGATCCTG	39.5
OPE-04	GTGACATGCC	39.5
OPE-09	CTTCACCCGA	39.5
OPE-18	GGACTGCAGA	39.5
OPAG-13	GGCTTGGCGA	43.6

3.5.3 Polymorphism Information Content

The PIC displayed by each RAPD primer were estimated by determining the frequency of alleles of the 96 DNA samples per locus as described by Sharma *et al.* (2009) using the formula: PIC = 1 - $\sum x^2_i$, where x_i is the relative frequency of the i^{th} allele of the RAPD loci. Markers were designated as being informative when the value of PIC was $\geq$ 0.5.

3.6 ITS sequence data collection

3.6.1 ITS amplification protocol

Samples (96) of DNA were amplified in 50 µl PCR reaction volumes using ITS1 and ITS4 primers. The PCR reaction mixture contained 4 µl of template DNA, 2 µl of a 0.5 µM concentration of ITS1 primer, 2 µl of a 0.5 µM concentration of ITS4 primer, 17 µl of nuclease free water and 25 µl of 2x Dream *Taq* master mix which contained: Dream *Taq* DNA polymerase, 2x Dream *Taq* buffer, dATP, dCTP, dGTP and dTTP of 0.4 mM each, and 0.4 mM $MgCl_2$. The PCR reaction profile consisted of initial denaturation temperature of 94 ºC for 4 min, followed by 40 cycles of denaturation temperature at 95 ºC for 1 min, annealing temperature of 55 ºC for 1 min, and an extension temperature at 72 ºC for 45 seconds. The final extension was then performed at 72 ºC for 10 min and lastly the PCR products were held at 4 ºC. A 2% agarose gel was then prepared in order to visualize PCR products and determine the success of the amplification.

3.7 Data collection and diversity analysis

3.7.1 SSR data collection and diversity analyses

3.7.1.1 SSR genetic diversity assessment

The overall genetic diversity within the *P. glaucum* accessions based on SSRs was measured by the Shannon-Wiener diversity index (Heip, Herman & Soetaert, 1998) which is obtained by the formula;

$$H = -\sum_{i=1}^{K} pi \log pi$$

Where P_i is the frequency of the i^{th} allele and k is the sum of alleles in the samples. The value of *H* provided the genetic diversity collectively revealed by all the SSR primer pairs initially selected (with a high PIC) to be used for further analysis.

3.7.1.2 Microsatellites data analysis

Microsatellites data (bands) for each primer pair from the gel were manually scored into a binary matrix for subsequent calculations. Each polymorphic band was regarded as a single character and was scored for the presence (1) of the character and absence (0) of the character. The degree of genetic diversity was calculated based on the loci that are polymorphic.

The binary data was then entered into Primer-E 5 for Windows (Plymouth Routines in Multivariate Ecological Research) software package (Clarke & Gorley, 2001) for similarity calculations and cluster analysis. The Bray-Curtis similarity method calculations were used to obtain similarity coefficients after which a phenogram was generated by Unweighted Pair-Group Method of arithmetic Average (UPGMA) clustering algorithm.

3.7.1.3 Ordination of SSR data

Ordination was done using Principal Coordinate Analysis (PCoA). Principal Coordinate Analysis (PCoA) was used to analyze microsatellites data using NTSYS-pc (Numerical Taxonomy and Multivariate Analysis System) software package (Rohlf, 1992). PCoA ordination plots for SSR were generated from SSRs data and used to identify trends, patterns, and asses the robustness and validity of the clusters generated from UPGMA. PCoA reduces the dimensionality of variables while maintaining their distance relationships (Gower, 1966).

3.7.2 RAPD data collection and diversity analysis

3.7.2.1 RAPD genetic diversity assessment

The overall genetic diversity detected by RAPD primers within the *P. glaucum* genotypes was also measured by the Shannon-Wiener diversity index (Heip *et al.*, 1998) using the formula below:

$$H = -\sum_{i=1}^{K} pi \log pi$$

Where P_i is the frequency of the *i*th allele and k is the sum of alleles in the samples. The value of *H* provided the diversity disclosed by all the RAPD primers employed for diversity analysis.

3.7.2.2 RAPD data analysis

The RAPD data obtained from the gel electrophoresis were scored into a binary matrix of presence (1) and absence (0). The data was then entered into Primer-E 5 for

Windows (Plymouth Routines in Multivariate Ecological Research) software package (Clarke & Gorley, 2001) for similarity calculations and cluster analysis. Similarity estimates were performed using Bray-Curtis similarity method and a phenogram was generated using Unweighted Pair-Group Method of arithmetic Average (UPGMA).

3.7.2.3 Ordination of RAPD data

Principal Coordinate Analysis (PCoA) was performed on RAPD data for Ordination purposes. NTSYS-pc (Numerical Taxonomy and Multivariate Analysis System) software package (Rohlf, 1992) was employed in the Principal Coordinate Analysis (PCoA) of the data from RAPD gels. This was done in order to confirm the clusters generated from UPGMA.

3.7.3 ITS data collection and diversity analysis

3.7.3.1 ITS data collection

The PCR products of ITS1 and ITS4 were sent for sequencing at Inqaba Biotechnical Industries (Pty) Ltd. Out of the 96 PCR samples sent for sequencing, eighty nine (89) reliable sequences were obtained from Inqaba Biotechnical Industries (Pty) Ltd. The sequences were then edited in BioEdit (Biological Sequence Alignment editor for Windows 99/98/NT/2K/XP/7) sequence alignment editor software (Hall, 1998).

3.7.3.2 ITS data analysis

A pairwise alignment in BioEdit (Hall, 1998) which allows ends to slide was performed on each ITS1 sequence from each sample with its complementary ITS4 reversed respective sequence. From each pair, a single sequence was obtained by merging the overlap and the two respective ends of the pair of sequences in order to form one edited sequence for analysis. The 89 edited sequences were then used to construct a phylogenic tree in MEGA (Molecular Evolutionary Genetics Analysis) software version 6.0 (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013).

The phylogenetic tree was resolved by using the Maximum Likelihood method based on the Tamura-Nei model (Tamura & Nei, 1993) and the bootstrap consensus tree was determined from 1000 replicates which represented the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 70% bootstrap replicates were collapsed and initial trees for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The use all sites function was used in the missing data treatment and the 1st+2nd+3rd+Non-Coding codons were included in the analysis. A *Pennisetum purpureum* strain retrieved from the National Centre for Biotechnology Information (NCBI) Website with accession number FJ626358.1 was also included in the analysis and used as the out-group to root the phylogenetic tree.

In addition, seven other *P. glaucum* ITS sequences with respective accession numbers; FJ766182.1, AY628132.1, AY628131.1, AY628130.1, AY628129.1, AY628116.1 and AY628115.1 were retrieved from the National Centre for

Biotechnology Information (NCBI) website and used in combination with the eighty nine sample sequences and a *P. pupureum* sequence. This was in order to assess the degree of relatedness based on the ITS sequences between the Namibian *P. glaucum* and other (non-Namibian) germplasm. The aforementioned *P. purpureum* strain was also included in the analysis as an out group to root the phylogenetic tree. This tree was constructed using the Tamura-Nei model (Tamura & Nei, 1993) on Maximum Likelihood method, the use all sites function was used in the missing data treatment, and the 1st+2nd+3rd+Non-Coding codons were included. The bootstrap consensus tree was molded from 1000 replicates which represented the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 70% bootstrap replicates were collapsed and initial trees for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach.

3.7.4 Morphological and molecular consistency of clusters

Unprocessed morphological data was obtained from the Gene Bank and processed in order to evaluate the relatedness of the *P. glaucum* accessions based on morphology analysis and to superimpose morphological characters onto the phylogenetic tree. Out of the ninety six (96) *P. glaucum* accessions used in this study, only twenty nine (29) *P. glaucum* accessions have been morphologically characterized (see Appendix C, for morphologically characterized data). Hence, the remaining sixty six (66) accessions of the present study are among the morphologically uncharacterized accessions. The

characters considered for analysis were early vigour, tillering altitude, number of nodal tillers, plant aspects, lodging susceptibility, fodder potential, spike shape, spikelet shattering, bristle length and spike density. Nevertheless only discrete morphological data was used in this study.

A similarity matrix was constructed using Primer-E 5 for Windows (Plymouth Routines in Multivariate Ecological Research) software package (Clarke & Gorley, 2001) for similarity calculations based on the ten (10) discrete morphological characters mentioned above. The Bray-Curtis similarity method was used and a phenogram was generated based on Unweighted Pair-Group Method of arithmetic Average (UPGMA). SSRs and RAPDs phenograms were then constructed based on the twenty nine (29) accessions that had corresponding morphological characterization data in order to investigate the relationship between the morphological characterization data derived cluster, and the SSR and RAPD data derived clusters of the twenty nine (29) *P. glaucum* accessions. Ordination was then done using Principal Coordinate Analysis (PCoA) on the discrete morphological data, SSRs data and RAPDs data in order to compare the number of respective groupings formed by different datasets and to confirm groupings with the clusters generated by UPGMA.

A phylogenetic tree was generated from the ITS sequences of the aforementioned twenty nine (29) *P. glaucum* accessions that had corresponding morphological characterization data. The *P. purpureum* strain with accession number FJ626358.1 was included in the analysis and used as the root of the phylogenetic tree. In addition, the seven *P. glaucum* ITS sequences retrieved from the National Centre for

Biotechnology Information (NCBI) website with respective accession numbers; FJ766182.1, AY628132.1, AY628131.1, AY628130.1, AY628129.1, AY628116.1 and AY628115.1 were included in the analysis. The tree was resolved by using the Maximum Likelihood method based on the Tamura-Nei model (Tamura & Nei, 1993) whereas the bootstrap consensus tree was determined from 1000 replicates which represented the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 70% bootstrap replicates were collapsed and initial trees for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The use all sites function was used in the missing data treatment and the 1st+2nd+3rd+Non-Coding codons were included.

3.8 Research ethics

The *P. glaucum* accessions obtained from the NPGRC were solely used for the purpose of this research. In addition, exclusive rights over the genetic diversity study of the Namibian *P. glaucum* germplasm was obtained from NPGRC and the results were kept confidential in line with the documentation of germplasm from the NPGRC. Throughout the research, the appropriate recognition of the intellectual, scientific support and operational contributions was applied. In addition, an ethical clearance certificate was obtained from the Research and Publications Office (RPO) in order to carry out the research at the University of Namibia Biological Sciences Department laboratories.

Chapter 4: RESULTS

4.1 DNA extraction

DNA was extracted from fresh leaves of *P. glaucum* samples and diluted to a standard concentration of 10ng/µl for each of the samples. The CTAB/chloroform-isoamyl alcohol protocol as described by Doyle (1991) was employed for extraction and produced good quality DNA. Extraction was performed on a total of 1441 accessions.

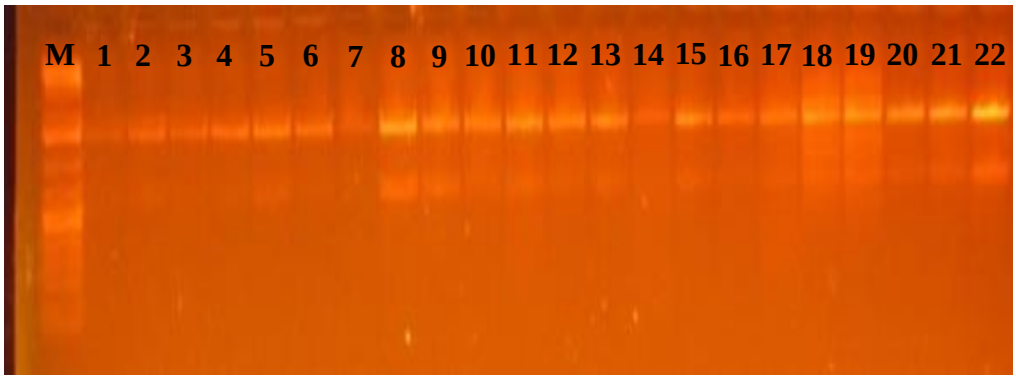


Figure 8: A 2% illustrative agarose gel electrophoresis showing 22 of the 96 *P. glaucum* DNA samples. M indicates the DNA molecular size marker (O' gene Ruler 100bp DNA ladder).

4.2 Microsatellites

4.2.1 Microsatellites amplification

The primer pairs 3005, 3016, 3020, 3022, 3035 and 3039 were used to amplify a total of ninety six (96) DNA samples employing the same reaction conditions with the exception of annealing temperatures which were primer specific. The amplicons were then run on a 2% agarose gel. Primer 3022 gave monomorphic bands and was not

considered for diversity calculations. However, the remaining five primers (3005, 3016, 3020, 3035 and 3039) produced a total of 662 reproducible bands. Reproducibility was tested by amplifying the DNA samples three times using the same reaction conditions and only reproducible bands were considered for genetic diversity calculations. Polymorphic primers based on the PIC were then used to evaluate the 96 samples. The amplicon band sizes ranged from 250 bp to 450 bp. Figures 9, 10 and 11 resulted from primer pairs; 3035, 3005 and 3039 respectively. These figures are illustrative and do not indicate the entire gel with the total number of bands produced as described in Table 5.

Table 5: The number of SSR amplicon bands produced by each SSR primer pair.

Primer	Number of bands
3005	116
3016	138
3020	208
3035	95
3039	107

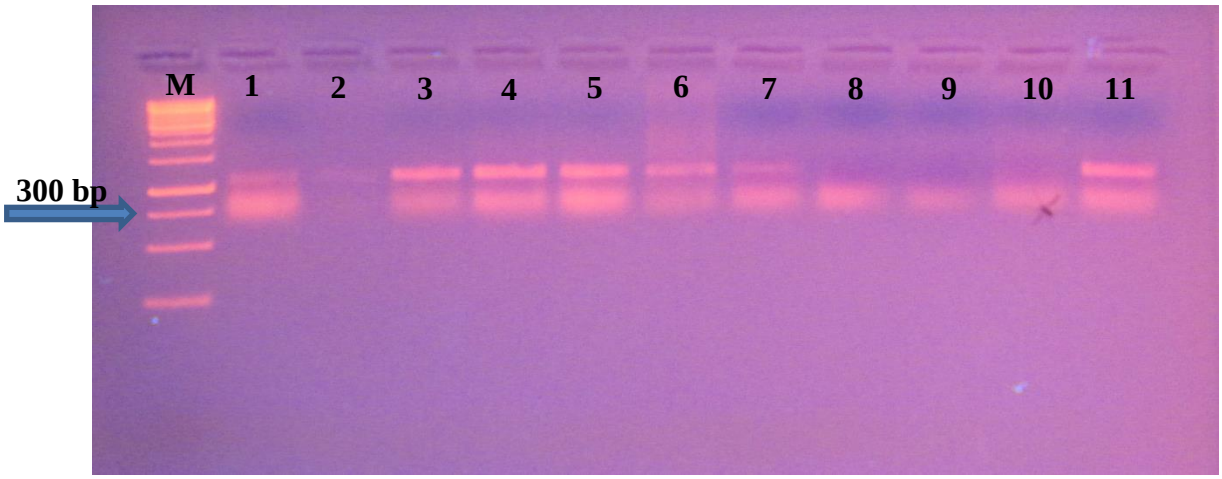


Figure 9: A 2% illustrative agarose gel electrophoresis of primer pair 3035 amplification products showing 11 from the 96 *P. glaucum* individuals. M indicates the DNA molecular size marker (O' gene Ruler 100bp DNA ladder).

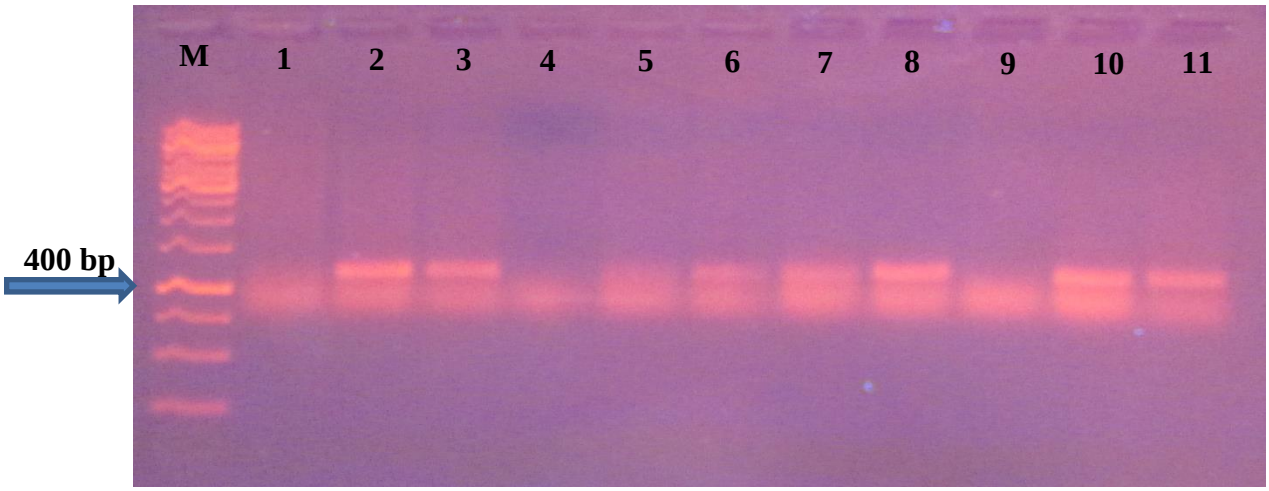


Figure 10: A 2% illustrative agarose gel electrophoresis of primer pair 3005 amplification products showing 11 from the 96 *P. glaucum* individuals. M indicates the DNA molecular size marker (O' gene Ruler 100bp DNA ladder).

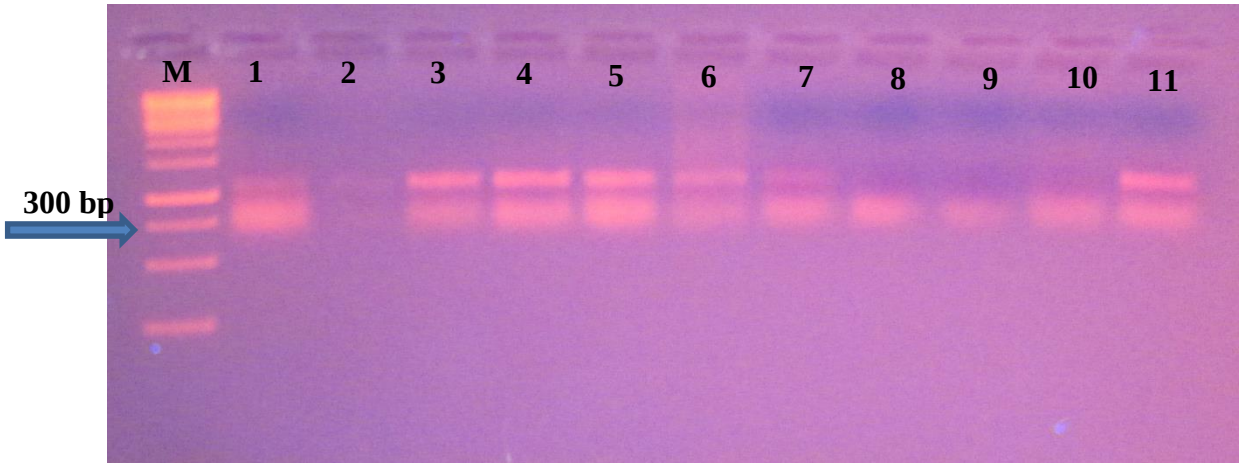


Figure 11: A 2% illustrative agarose gel electrophoresis of primer pair 3039 amplification products showing 11 from the 96 *P. glaucum* individuals. M indicates the DNA molecular size marker (O' gene Ruler 100bp DNA ladder).

4.2.2 SSR PIC assessments

Table 6: PIC values of the SSR primers: 3005, 3020, 3039, 3016 and 3035.

Primer	PIC value
3005	0.64
3016	0.48
3020	0.48
3035	0.76
3039	0.69

According to Table 6, primer 3035 was found to be the most informative of the five SSR primers (3005, 3020, 3039, 3016 and 3035). Primer 3016 and 3020 produced PIC values less than 0.5 and hereafter were not employed in successive calculations and analysis whereas primer 3005, 3035 and 3039 were used for diversity calculations because they gave a PIC value greater than 0.5.

4.2.3 SSR genetic diversity calculations

The genetic diversity of the ninety six (96) *P. glaucum* accessions was calculated using the Shannon diversity index based on SSR primers: 3005, 3035 and 3039. These primers detected a total of 6 alleles designated a1, a2, a3, a4, a5 and a6. The calculated value of the Shannon diversity index was found to be 0.45 (Table 7) indicating a low level of genetic diversity among the accessions. This value is low and agrees with the expectations of the hypothesis suggesting that: the poor performance of the Namibian *P. glaucum* landraces is due to lack of genetic diversity. . The Shannon diversity index values are usually between 1.5 and 3.5 in most ecological studies, and the index is rarely greater than 4 (Pavoine & Bonsall, 2011).

Table 7: SSR Shannon diversity index calculated from primers 3005, 3035 and 3039.

Allele	**Frequency (P_i)**	**$P_i \log P_i$**	$H = -\sum_{i=1}^{K} pi \log pi$
a1	0.29	-0.16	0.45
a2	0.92	-0.03	
a3	0.93	-0.03	
a4	0.05	-0.07	
a5	0.93	-0.03	
a6	0.18	-0.13	

4.2.4 SSR clusters

UPGMA-generated phenograms computed using the Bray-Curtis similarity revealed that at a similarity level of 68%, the 96 *P. glaucum* genotypes were divided into two main clusters labeled I and II (Figure 12). The first main cluster was composed of only one (1) cluster while the second main cluster was divided into eleven (11) sub-clusters designated A-K respectively. The first (1) main cluster was formed at a similarity level of 68% and was comprised of 7 accessions.

The second main cluster was formed by eleven (11) sub-clusters designated A-K (Figure 12). The first sub-cluster A comprised of 5 accessions (similarity level 70%), second sub-cluster B contained 3 accessions (similarity level 72 %), third sub-cluster C was the largest sub-cluster consisting of 36 accessions (Figure 12) (similarity level 84%), whereas the fourth sub-cluster D consisted of only 1 accession (Nam 2580, similarity level 85%) and the fifth sub-cluster E by 2 accessions (Nam 1163 and Nam 138; similarity level 90%). Further, the sixth sub-cluster F contained 14 accessions (similarity level 90%) and the seventh sub-cluster G by 2 accessions (Nam 2868 and Nam 129; similarity level of 90%). Finally, the eighth sub-cluster H consisted of 3 accessions with (90% similarity level). The ninth sub-cluster I contained 4 accessions (similarity level 85%). The tenth sub-cluster J contained 2 accessions (Nam 2780 and Nam 2828, similarity level 87%), and the eleventh sub-cluster K contained 17 accessions (similarity level 87%). Most of the clusters were formed at a level of similarity that is above 70%, this indicates that the level of genetic diversity is indeed low.

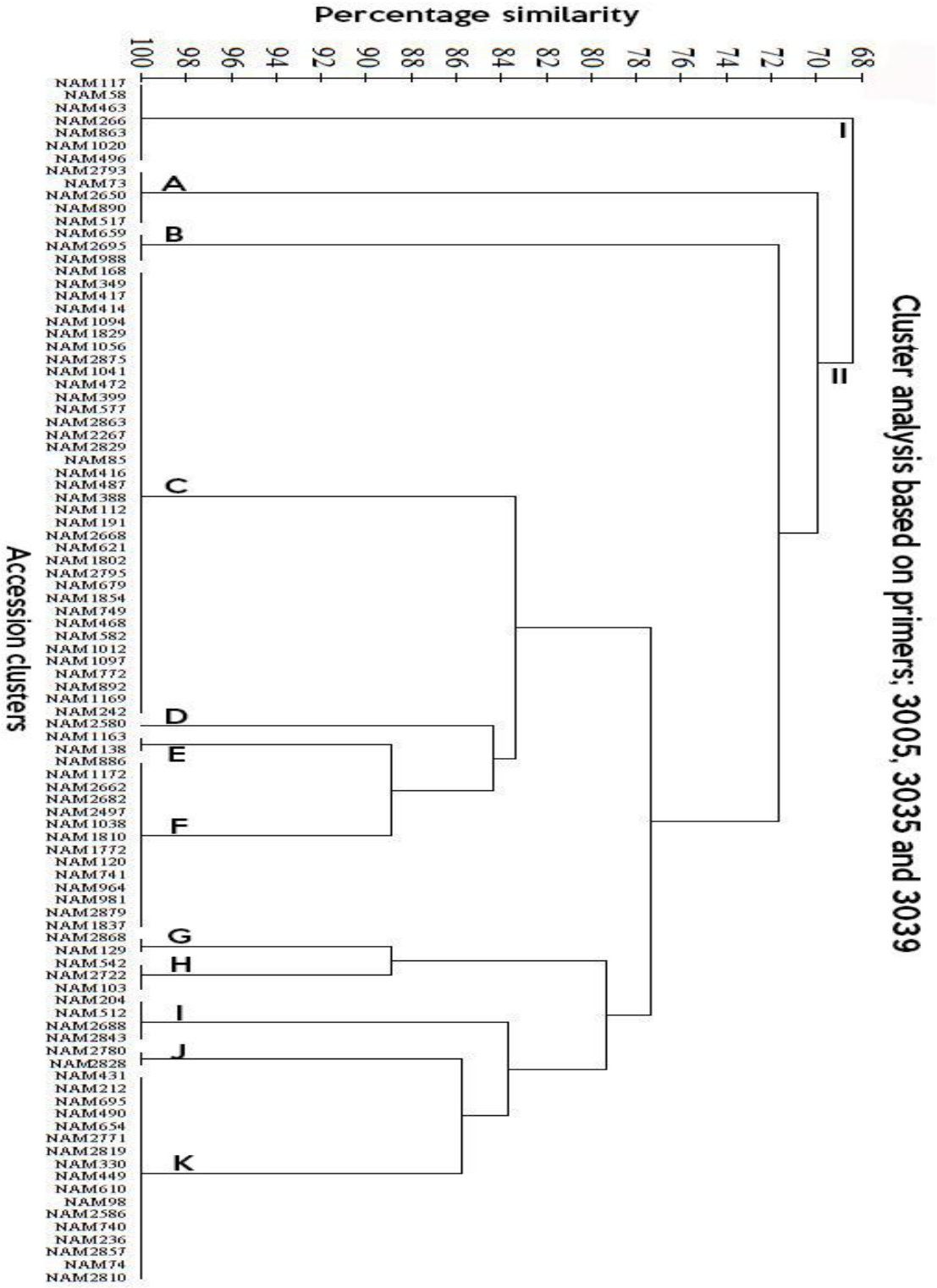


Figure 12: UPGMA phenogram generated from 3005, 3035 and 3039 SSR profiles of 96 *P. glaucum* individuals. The scale represents Bray-Curtis similarity coefficients

4.2.5 SSR data ordination

The results from PCoA analysis based on SSR markers revealed diversity of the *P. glaucum* by assessing six alleles that were detected. The ordination plot revealed twelve (12) clearly distinct groupings as also indicated in the cluster analysis.

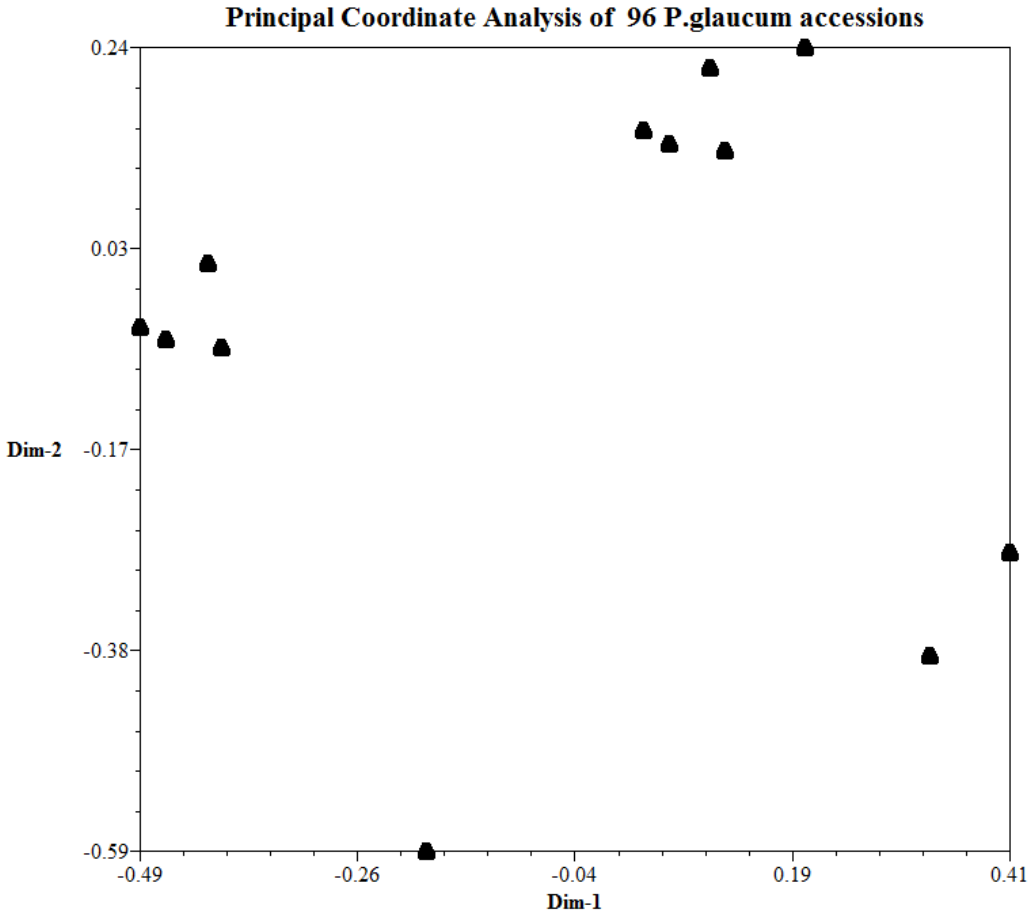


Figure 13: Principal Coordinate Analysis of the 96 *P. glaucum* genotypes based on SSRs marker data.

4.3 RAPDs

4.3.1 RAPD amplification

RAPD primers OPN-16, OPE-18 and OPE-09 revealed high levels of polymorphism based on PIC values upon which they were used to determine genetic variability in the 96 *P. glaucum* accessions using constant reaction conditions except for the annealing temperature which was changed according to individual primer specifications. The RAPD PCR products were electrophoresed on a 2% agarose gel. The three (3) RAPD primers produced a total of 275 reproducible bands with sizes ranging from 250 bp to 400 bp. Figures 14, 15 and 16 displays illustrative gels from primers; OPN-16, OPE-18 and OPE-09 respectively, and do not indicate the entire gel with the total number of bands produced as described in Table 8.

Table 8: The number of bands produced by each RAPD primer.

Primer	Number of bands
OPN-16	112
OPE-18	82
OPE-09	81

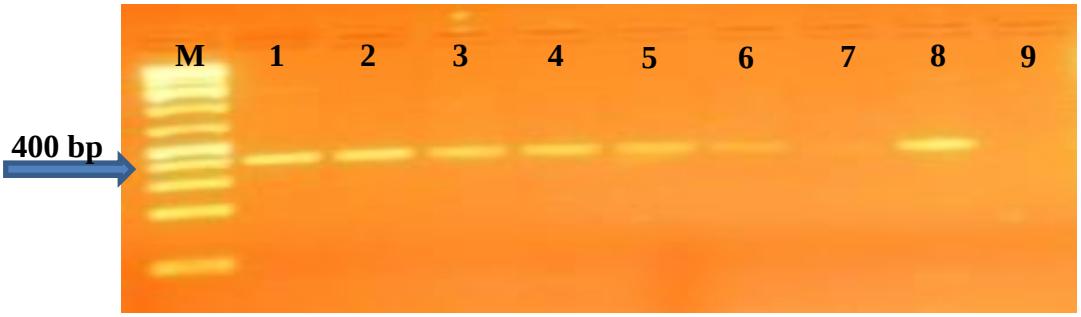


Figure 14: A 2% illustrative agarose gel electrophoresis of primer OPN-16 amplification products showing 9 from the 96 *P. glaucum* individuals. M indicates the DNA molecular size marker (O' gene Ruler 100bp DNA ladder)

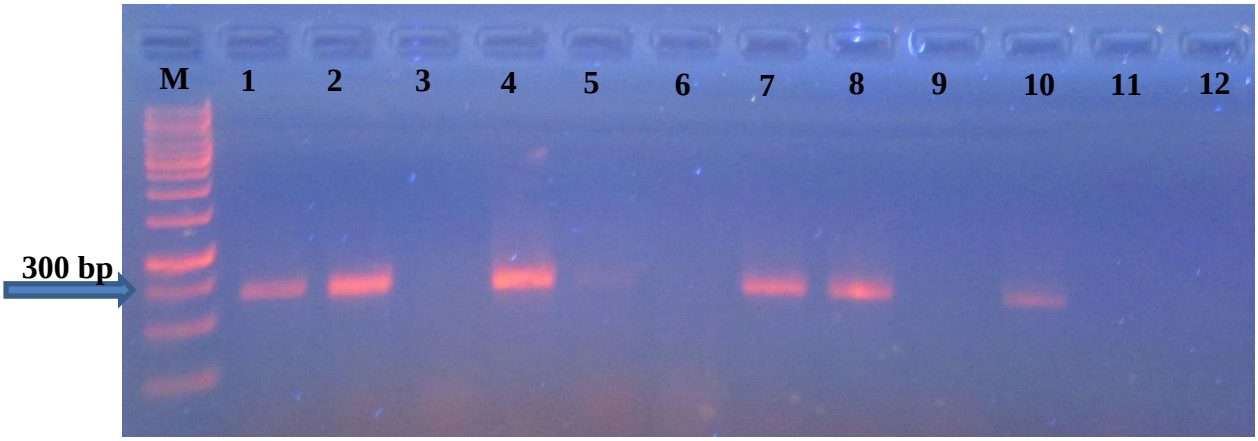


Figure 15: A 2% illustrative agarose gel electrophoresis of primer OPE-09 amplification products showing 12 from the 96 *P. glaucum* individuals. M indicates the DNA molecular size marker (O' gene Ruler 100bp DNA ladder)

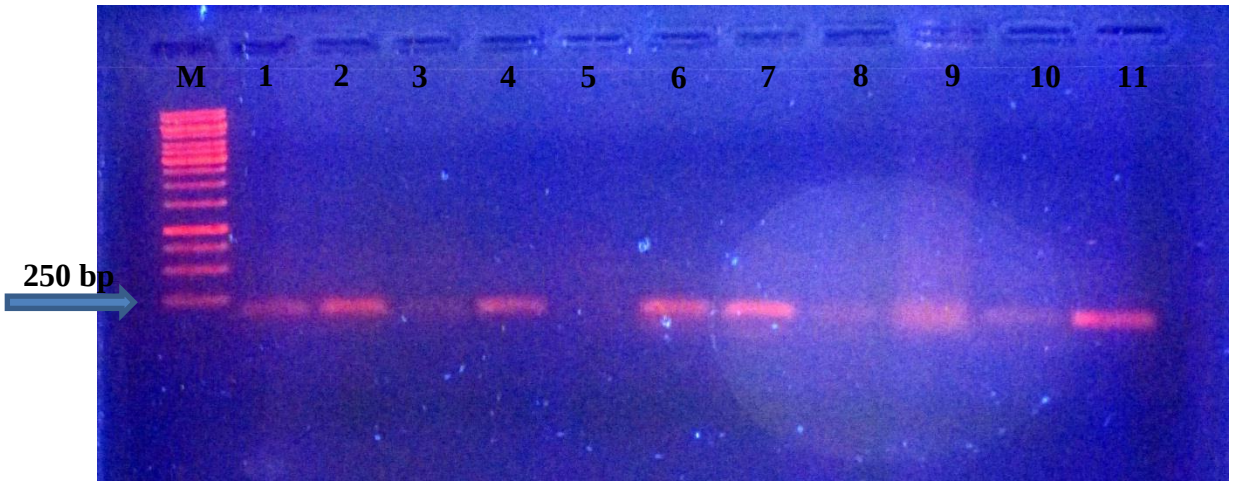


Figure 16: A 2% illustrative agarose gel electrophoresis of primer OPE-18 amplification products showing 11 from the 96 *P. glaucum* individuals. M indicates the DNA molecular size marker (O' gene Ruler 1Kb DNA ladder)

3.2 RAPD PIC assessments

The PIC values of the three RAPD primers used to amplify DNA was calculated in order to eliminate uninformative primers.

Table 9: PIC values of RAPD primers OPN-16, OPN-18 and OPE-09.

Primer	PIC value
OPN-16	0.66
OPN-18	0.82
OPE-09	0.82

According to **Table 9**, primers OPN-16, OPN-18 and OPE-09 gave a PIC value that is greater than 0.5. Hence they were all used for genetic diversity calculations.

4.3.3 Genetic diversity calculations

The Shannon diversity index of the ninety six (96) *P. glaucum* genotypes using primers OPN-16, OPN-18 and OPE-09 was calculated to measure the level of diversity. A total of six (6) alleles were revealed by the RAPD primers (OPN-16, OPN-18 and OPE-09) which were designated a7, a8, a9, a10, a11 and a12. The calculated value of the Shannon diversity index was found to be 0.7 (the value is generally between 1.5 and 3.5).

Table 10: Shannon diversity index based on RAPD primers OPN-16, OPN-18 and OPE-09.

Allele	Frequency (P_i)	$P_i \log P_i$	$H = -\sum_{i=1}^{K} pi \log pi$
a1	0.68	-0.11	0.7
a2	0.49	-0.15	
a3	0.67	-0.12	
a4	0.19	-0.14	
a5	0.76	-0.09	
a6	0.08	-0.09	

4.3.4 RAPD clusters

The ninety six (96) *P. glaucum* genotypes were grouped into two (2) main clusters designated I and II at a 50% similarity level. However, the first main cluster was divided into 15 sub-clusters designated A-O while the second main cluster was divided into four cluster designated W, X, Y and Z. Sub-cluster A , consisted of 4 accessions (similarity level 78%), sub-cluster B consisted of only 1 accession (Nam 2722, similarity level 80%). The third sub-cluster C was formed by 13 accessions (similarity level 80%; Figure 17) and the fourth sub-cluster D consisted of only 1 accession (Nam 2668; similarity level 80%).

The fifth sub-cluster E consisted of 12 accessions (similarity level 80%; Figure 17), the sixth sub-cluster F consisted of 3 accessions (similarity level 80%), the seventh sub-cluster G consisted of 18 accessions (similarity level 80%) and the eighth sub-cluster H consisted of only 1 accession (Nam 1163, similarity level 91%).

The ninth sub-cluster I was formed by 1 accession (Nam 212, similarity level 91%). One accession (Nam 2662) formed the tenth cluster (J, similarity level of 88%). The eleventh sub-cluster K was formed by 3 accessions (similarity level 88%). The twelfth sub-cluster L was formed by 2 accessions (Nam 2780 and Nam 2879: similarity level 88%). The thirteenth sub-cluster M was formed at a similarity level of 80% and consisted of 5 accessions (Figure 17). The fourteenth sub-cluster N was formed by 7 accessions (similarity level 86%), the fifteenth sub-cluster, O by 17 accessions (similarity level 86%).

The second main cluster was divided into four sub-clusters designated W, X, Y and Z. The first sub-cluster, W was formed by Nam 2497 and Nam 191 at a similarity level of 56%. The second sub-cluster, X formed at a similarity level of 62% and consisted of only Nam 2267 whereas the third sub-cluster, Y formed at a similarity level of 80% and consisted of Nam 416 and Nam 73 respectively. The fourth sub-cluster Z formed at a similarity level of 80% and consisted of accessions Nam 654 and Nam 2828.

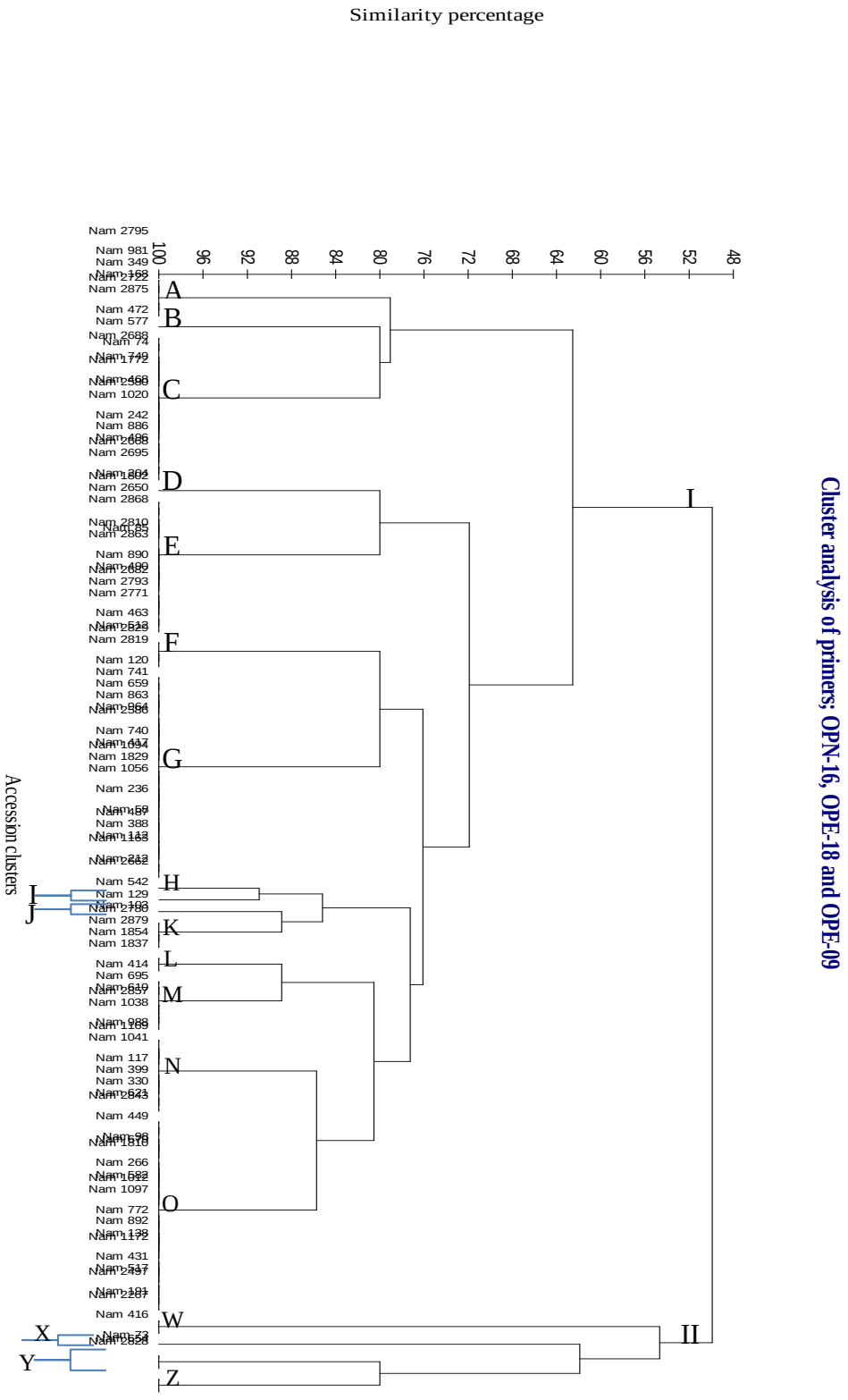


Figure 17: UPGMA phenogram generated from OPN-16, OPE-18 and OPE-09 RAPD profiles of 96 *P. glaucum* individuals. The scale represents Bray-Curtis similarity coefficients

4.3.5 RAPD Data Ordination

PCoA analysis of RAPD showed that the diversity of the *P. glaucum* could be evaluated based on the six alleles that were detected. The six alleles designated: a7, a8, a9, a10, a11 and a12 were used as the variables for the principal coordinate analysis. The six variables resulted in the construction of 19 groupings with accessions on the ordination plot consistent with those indicated in the cluster analysis above. This showed that the groupings displayed by UPGMA were reliable.

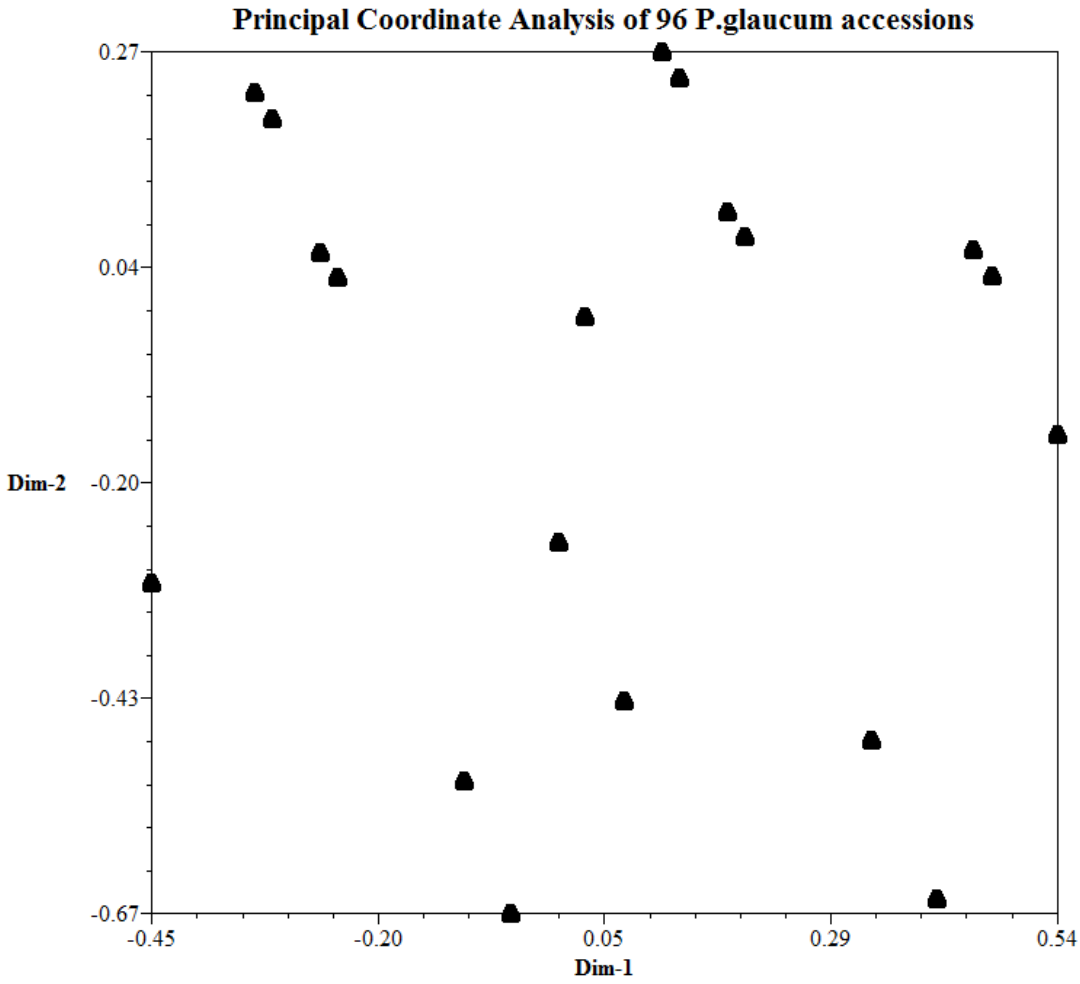


Figure 18: Principal Coordinate Analysis of the 96 *P. glaucum* genotypes based on RAPD data.

4.4 ITS

4.4.1 ITS amplification

Primers ITS1 and ITS4 were used to amplify the 96 *P. glaucum* accessions using the aforementioned reaction conditions. The PCR product sizes ranged between 700 base

pairs and 1000 base pairs. The ITS PCR amplicons were electrophoresed on a 2% agarose gel.

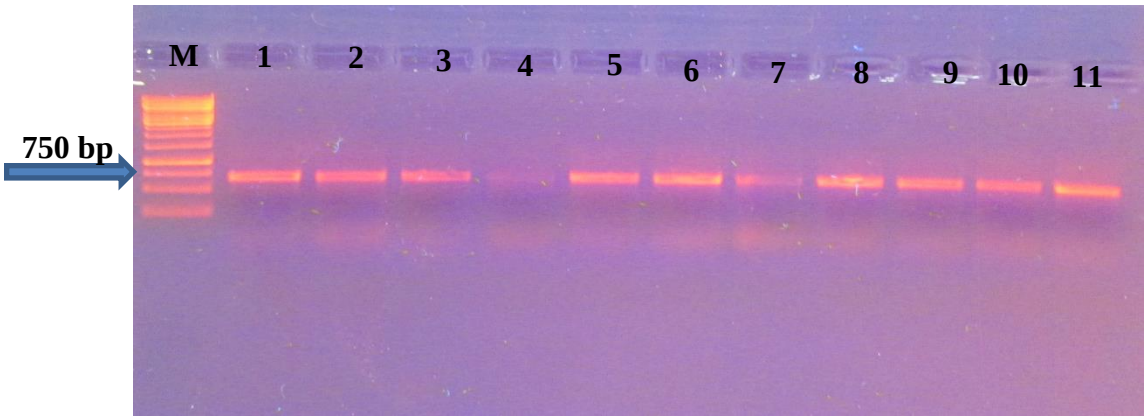


Figure 19: A 2% illustrative agarose gel electrophoresis showing 11 from the 96 individuals of *P. glaucum* amplified by primers ITS1 and ITS4. M indicates the DNA molecular size marker (O' gene Ruler 1Kb DNA ladder)

4.4.2 ITS phylogenetic analysis

4.4.2.1 Eighty nine sequence clusters

The eighty nine (89) *P. glaucum* genotypes combined with the seven NCBI retrieved *P. glaucum* ITS sequences formed three main groups labeled I, II and III. Group I had all the accessions under investigation plus six of the seven NCBI retrieved ITS sequences. Group II consisted of one of the seven NCBI retrieved ITS sequences while group III was formed by the out-group *P. purpureum*. Clusters 1, 44 and 47 were collapsed for suitable display of the data.

Group I was subdivided into two groups labeled A and B. Group A was composed of clusters designated 1-52. Cluster, 1 was formed by 31 accessions (bootstrap value of 95%). Furthermore, accessions Nam 85, Nam 2863 and Nam 890 formed a sub-cluster (bootstrap value of 80%). In addition, 2 accessions formed a smaller sub-cluster (bootstrap value of 81%). Cluster 2 consisted of only accession Nam 449 whereas cluster 3 was formed by 2 accessions (bootstrap value of 93%).

Clusters 4, 5, 6 and 7 were composed of single accessions while cluster 8 consisted of 2 accessions (77% bootstrap value). Clusters 9-13 were formed by single accessions whereas cluster 14 consisted of 2 accessions (bootstrap value of 89%). Clusters 15, 16, 17 and 18 consisted of single accessions. Cluster 19 was composed of 2 accessions (70% bootstrap value) while clusters 20, 21, 22, 23 and 24 were formed by single accessions. Cluster 25 was composed of 2 accessions (70% bootstrap value).

Clusters 26-43 had single accessions while cluster 44 was formed by 3 accessions (bootstrap value of 80%) and 2 accessions formed a sub-cluster (74% bootstrap value). Clusters 45 and 46 consisted of single accessions while cluster 47 was formed by 2 accessions (bootstrap value of 87%).

Clusters 48, 49, 50 and 51 consisted of single accessions and cluster 52 was composed of NCBI retrieved accessions AY628132.1, AY628130.1, AY628131.1 and AY628129.1 (100% bootstrap value). In addition, accessions AY628131.1 and AY628129.1 formed a sub-cluster (100% bootstrap value). Cluster B of group I was formed by the NCBI retrieved accessions AY628116.1 and AY628115.1 (100%

bootstrap value). Group II was formed by a single NCBI retrieved accession FJ766182.1 while group III consisted of the out-group *P. purpureum.*

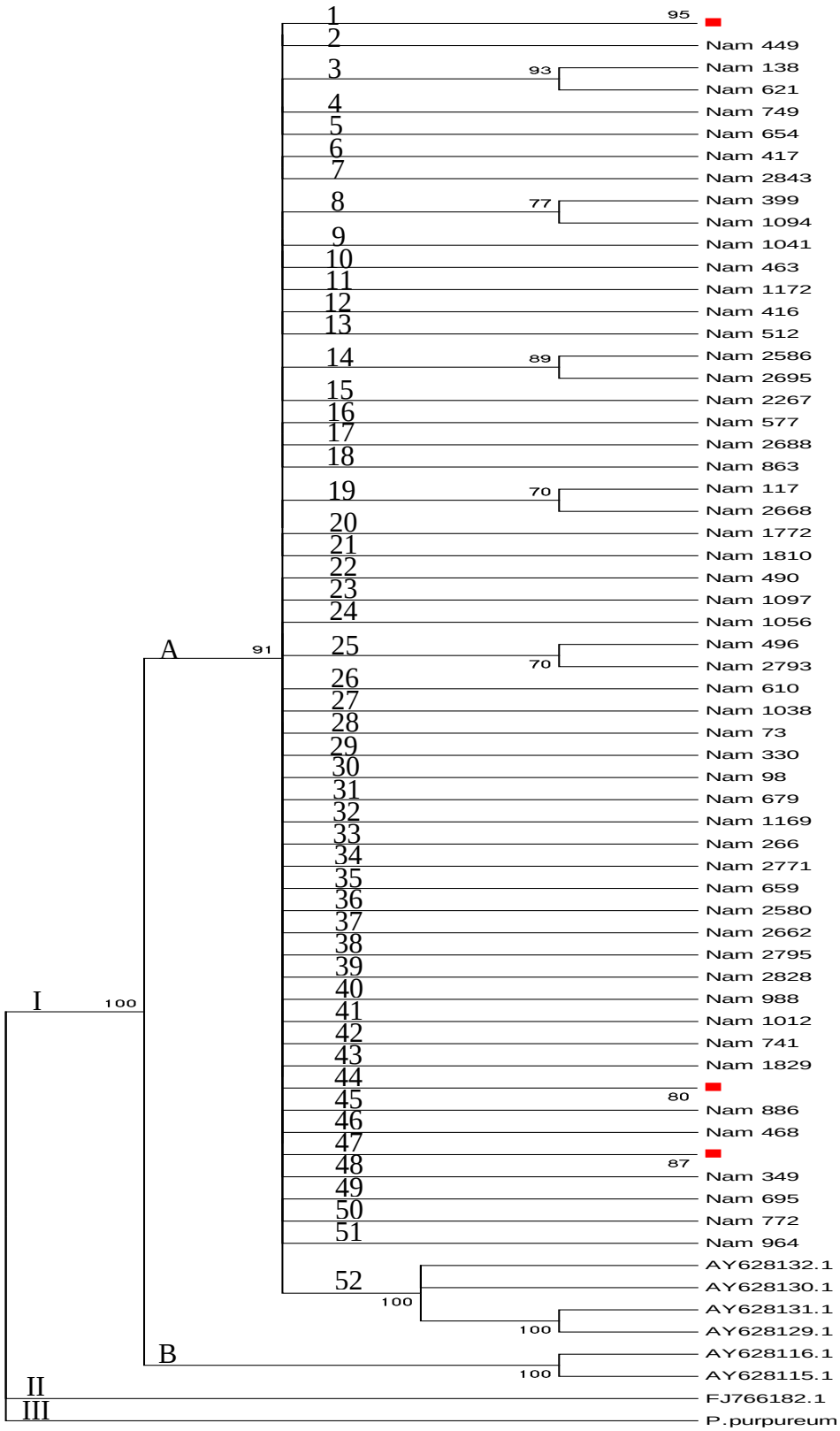


Figure 20: The phylogenetic tree for 89 plus seven NCBI retrieved *P. glaucum* genotypes based on ITS sequences inferred using the Maximum Likelihood method. *P. purpureum* was used as the out-group to root the tree generated with MEGA 6.0 software.

4.5 Morphology and molecular clusters

4.5.1.1 Morphology UPGMA clustering of 29 accessions

The morphological data-generated phenogram of the twenty nine *P. glaucum* landraces displayed the occurrence of two main clusters at 78% similarity level. These clusters were designated: I for the first main cluster and II for the second main cluster. The first main cluster (I) formed at similarity level of 78% and was formed by accession Nam 330. The second main cluster (II), was divided into 27 sub-clusters designated 1-27. All the sub-clusters were composed of one accession, except sub-cluster 16, which contained two accessions. Sub-cluster 1 formed at a similarity level of 80% and contained accession Nam 414.

The second sub-cluster (2) consisted at a 84% similarity level with only Nam 73 and the third sub-cluster (3) only by Nam 1163 also with a 84% similarity level . Sub-cluster 4 contained accession Nam 2580 (80% similarity level). Sub-cluster 5 contained 1 accession (86%), sub-cluster 6 accession Nam 1837 (86%), sub-cluster 7 accession Nam 2586 (96%), sub-cluster 8 accession Nam 2650 (96%) and sub-cluster 9 accession Nam 399 with a similarity level of 91%.

Accession Nam 2682 formed sub-cluster, 10 at a similarity level of 92%, sub-cluster 11 by accession Nam 58 (95%), sub-cluster 12 by accession Nam 212 (95%). Sub-cluster 13 was formed by accession Nam 85 (88%), sub-cluster 14 by accession Nam 349 (88%), sub-cluster 15 by accession Nam 2668 (94%), sub-cluster 16 was formed by 2 accessions (94%) , sub-cluster 17 by accession Nam 388 (92%), sub-

cluster 18 by accession Nam 463 (92%) and sub-cluster 19 by accession Nam 103(91%). Accession Nam 74 formed sub-cluster 20 (97%), sub-cluster 21 by accession Nam 117 (97%), sub-cluster 22 by accession Nam 112 (97%), sub-cluster 23 by accession Nam 120 (97%), sub-cluster 24 by accession Nam 1012(90%), sub-cluster 25 by accession Nam 416(94%),sub-cluster 26 by accession 417 (96%) and sub-cluster 27 by accession Nam 431 (96%).

Accessions in sub-clusters 5 and 6 possess an intermediate fodder potential while accessions in sub-clusters 7-12 possess a good fodder potential. Sub-clusters 13 and 14 had accessions with a low lodging susceptibility while accessions in sub-clusters 15 and 16 possess a conical spike shape and further accessions in sub-clusters 17, 18, 20 and 21 possess a candle spike shape. Sub-clusters 22 and 23 had accessions possessing a compact spike density while sub-clusters 25-27 had accessions with a high early vigour.

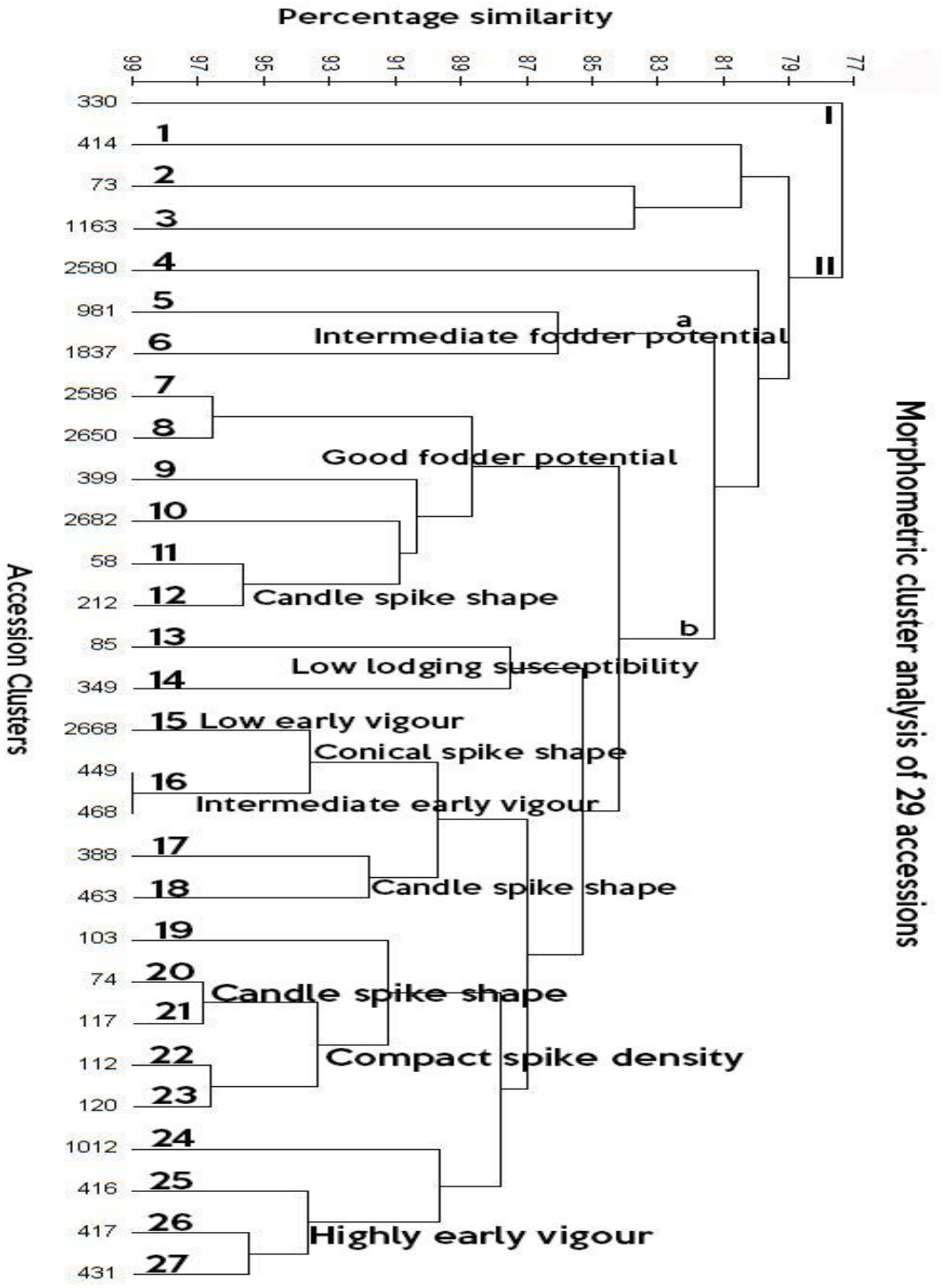


Figure 21: UPGMA phenogram constructed from morphometric data showing some phenotypic traits of 29 *P. glaucum* individuals. The scale represents Bray-Curtis similarity coefficients

4.5.1.2 Morphology data Principal Coordinate Analysis

The results from PCoA analysis based on morphological data of 29 *P. glaucum* accessions revealed diversity. The ordination plot revealed twenty eight clearly distinct groupings and confirmed consistency with the ones obtained from UPGMA cluster analysis.

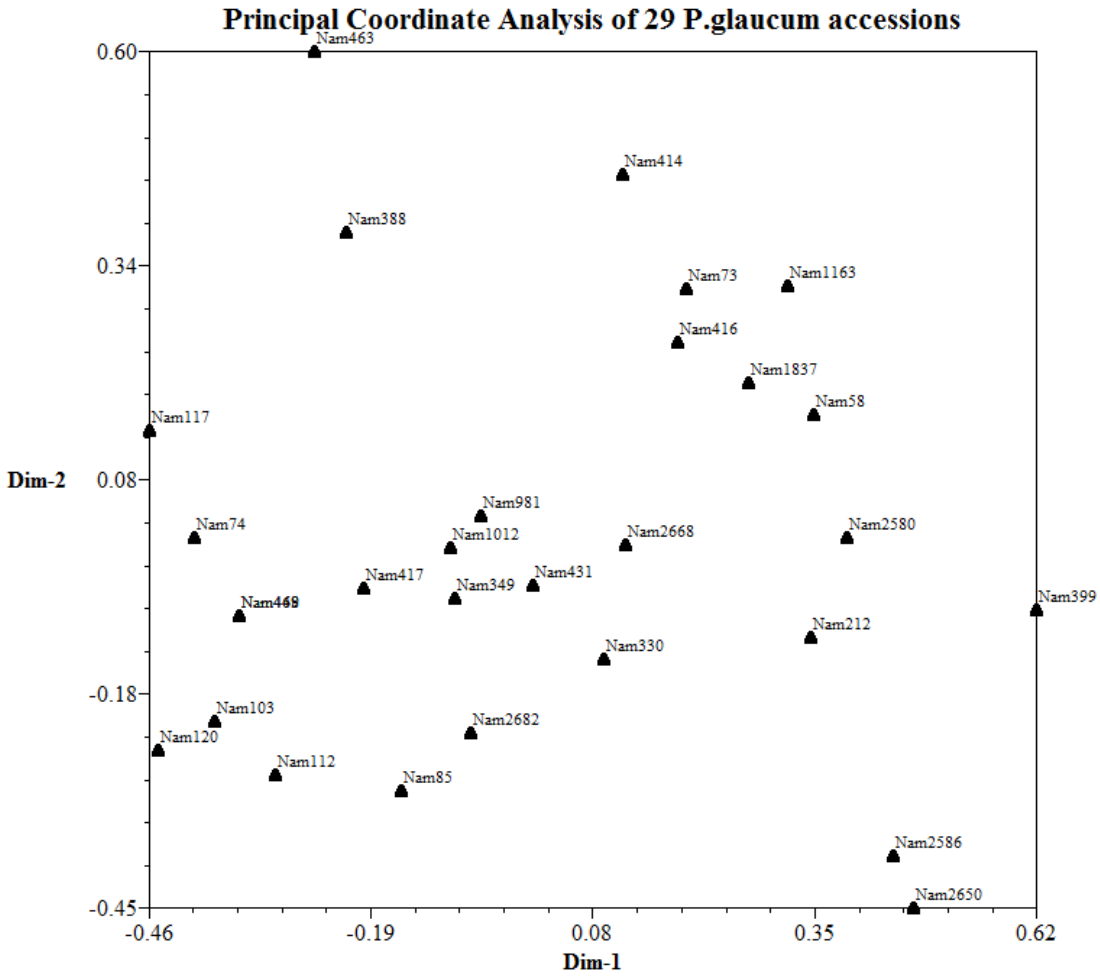


Figure 22: Principal Coordinate Analysis of the 29 *P. glaucum* genotypes based on morphological data.

4.5.2 Molecular (SSR, RAPD & ITS) clusters

4.5.2.1 SSR UPGMA clustering of 29 accessions

An SSR cluster was drawn based on the same twenty nine *P. glaucum* landraces used for morphology cluster construction. The results showed that the 29 accession could be grouped into two main groups using SSRs. The two main clusters formed at a similarity level of 69% and were designated: I for the first main cluster and II for the second main cluster. The first main cluster (I) (69%) consisted of 2 accessions. The first main cluster (I) was not divided into sub-clusters. However, the second main Cluster (II) was divided into 7 sub-clusters labeled 1-7. Sub-cluster 1 (73%) was formed by 3 accessions.

The second sub-cluster (2) was the largest cluster with 11 accessions (84% similarity level). The third sub-cluster (3) consisted of only accession Nam 103 (89% similarity level). Sub-cluster 4 consisted of 6 accessions (89% similarity level), sub-cluster 5 had only one accession Nam 2580 (83%), sub-cluster 6 also by one accession Nam 1163 (89%) and sub-cluster 7 contained 4 accessions (89%).

The first main cluster (I) had accessions with an intermediate spike density while sub-cluster 1 had accessions possessing a candle spike shape. Accessions in sub-cluster 2 possess a compact spike density whereas sub-cluster 3 had accessions with an intermediate tillering attitude. Sub-cluster 4 had accessions with an erect tillering attitude while sub-cluster 7 had accessions with a low lodging susceptibility.

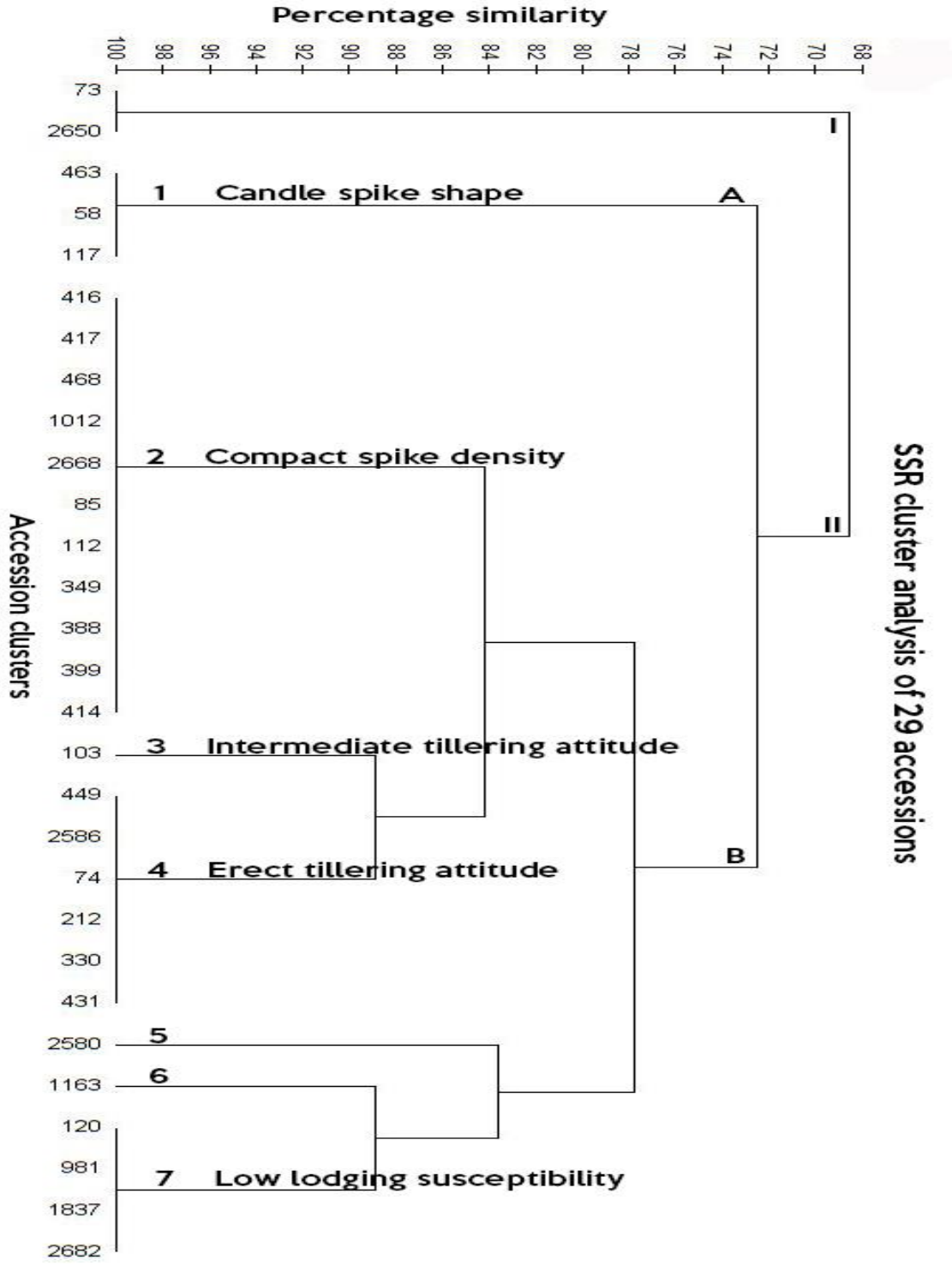


Figure 23: UPGMA phenogram constructed from 3005, 3035 and 3039 SSR profiles showing some phenotypic traits of 29 *P. glaucum* individuals. The scale represents Bray-Curtis similarity coefficients

4.5.2.2 SSR data Principal Coordinate Analysis of 29 *P. glaucum* accessions

The results from PCoA analysis of 29 *P. glaucum* accessions based on SSR markers disclosed diversity of the *P. glaucum* by assessing six alleles that were detected. The ordination plot revealed eight (8) clearly distinct groupings. The groupings of PcoA confirmed the clusters generated by UPGMA analysis.

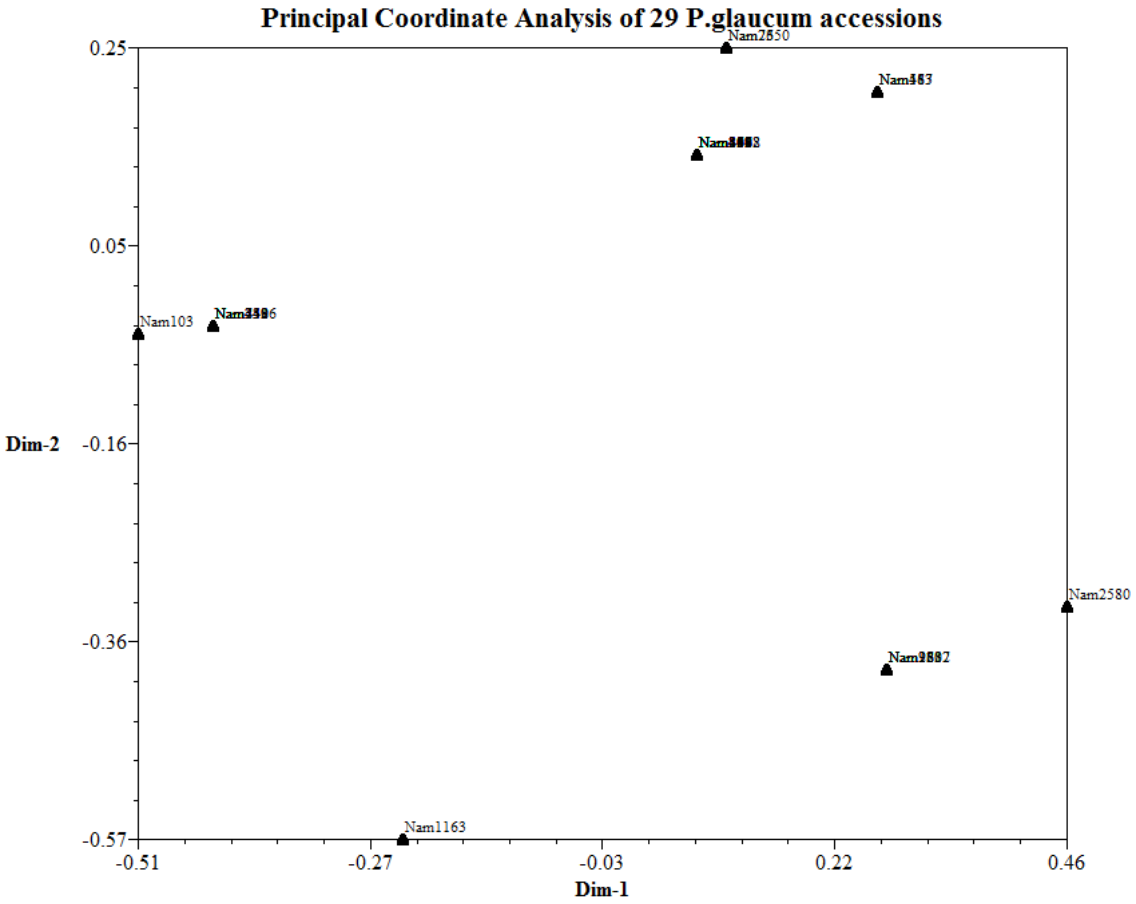


Figure 24: Principal Coordinate Analysis of the 29 *P. glaucum* genotypes based on SSR data.

4.5.2.3 RAPD UPGMA clustering of 29 accessions

RAPD clustering of the twenty nine *P. glaucum* landraces revealed that the accessions could be grouped into two main clusters at a similarity level of 60%. The first main cluster was termed I while the second was termed II. Both clusters I and II were divided into sub-clusters designated 1-9 for cluster I and A-D for cluster II. The first main sub-cluster 1 (80%) was formed by 2 accessions,sub-cluster, 2 (80%) composed of 3 accessions,sub-cluster 3(85%) by accession Nam 212 only,sub-cluster 4 (93%) by accession Nam 103 and sub-cluster 5 (93%) by accession 1163.

Sub-cluster 6 (79%) was formed by 6 accessions, sub-cluster 7 (82%) by 2 accessions,sub-cluster 8 (89%) by 2 accessions , sub-cluster 9 (89%) by 4 accessions. The second main cluster II was divided into four sub-clusters A-D. The first sub-cluster, A (80%) was formed by accession Nam 463, sub-cluster B (80%) by 2 accessions , sub-cluster C (80%) by accession Nam 2668 and sub-cluster D (80%) by 3 accessions.

Accessions in sub-cluster 1 possess a low lodging susceptibility, few numbers of nodal tillers and an erect tillering attitude while accessions in sub-cluster 2 possess a compact spike density and short bristle length. Sub-cluster 3 had a single accession with a good green fodder yield potential. Sub-clusters 4 and 5 had accessions possessing an intermediate green fodder yield potential whereas sub-cluster 6 had accessions with a compact spike density and low lodging susceptibility. Accessions in sub-clusters 7- 9 possess an erect tillering attitude.

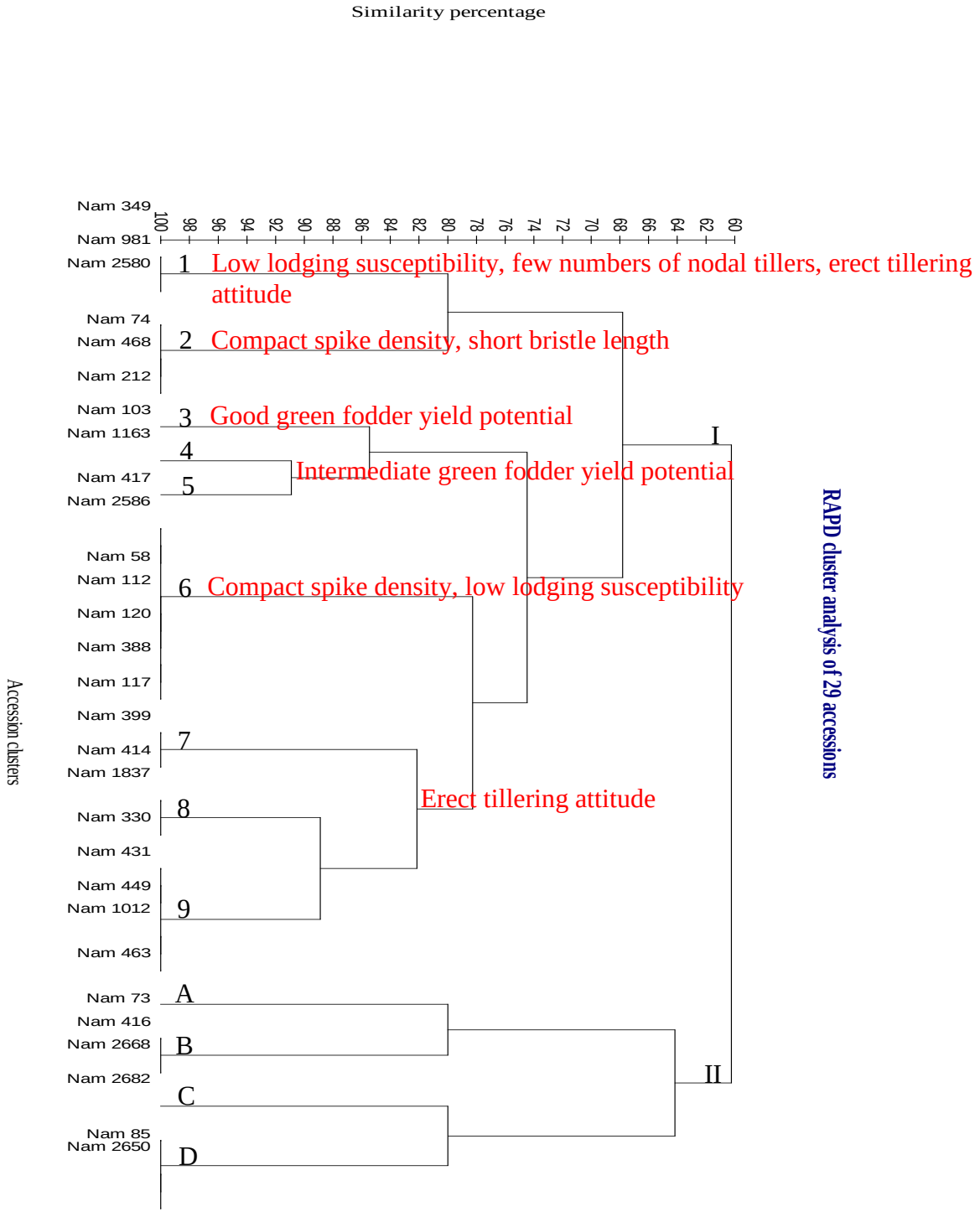


Figure 25: UPGMA phenogram generated from OPN-16, OPE-18 and OPE-09 RAPD profiles showing some phenotypic traits of 29 *P. glaucum* individuals. The scale represents Bray-Curtis similarity coefficients

4.5.2.4 RAPD data Principal Coordinate Analysis of 29 *P. glaucum* accessions

The results from PCoA analysis based on RAPD markers revealed diversity of the *P. glaucum* based on the six alleles that were detected. The ordination plot disclosed thirteen (13) clearly distinct groupings. Consistency was observed between the clusters generated by UPGMA and the groupings from PcoA.

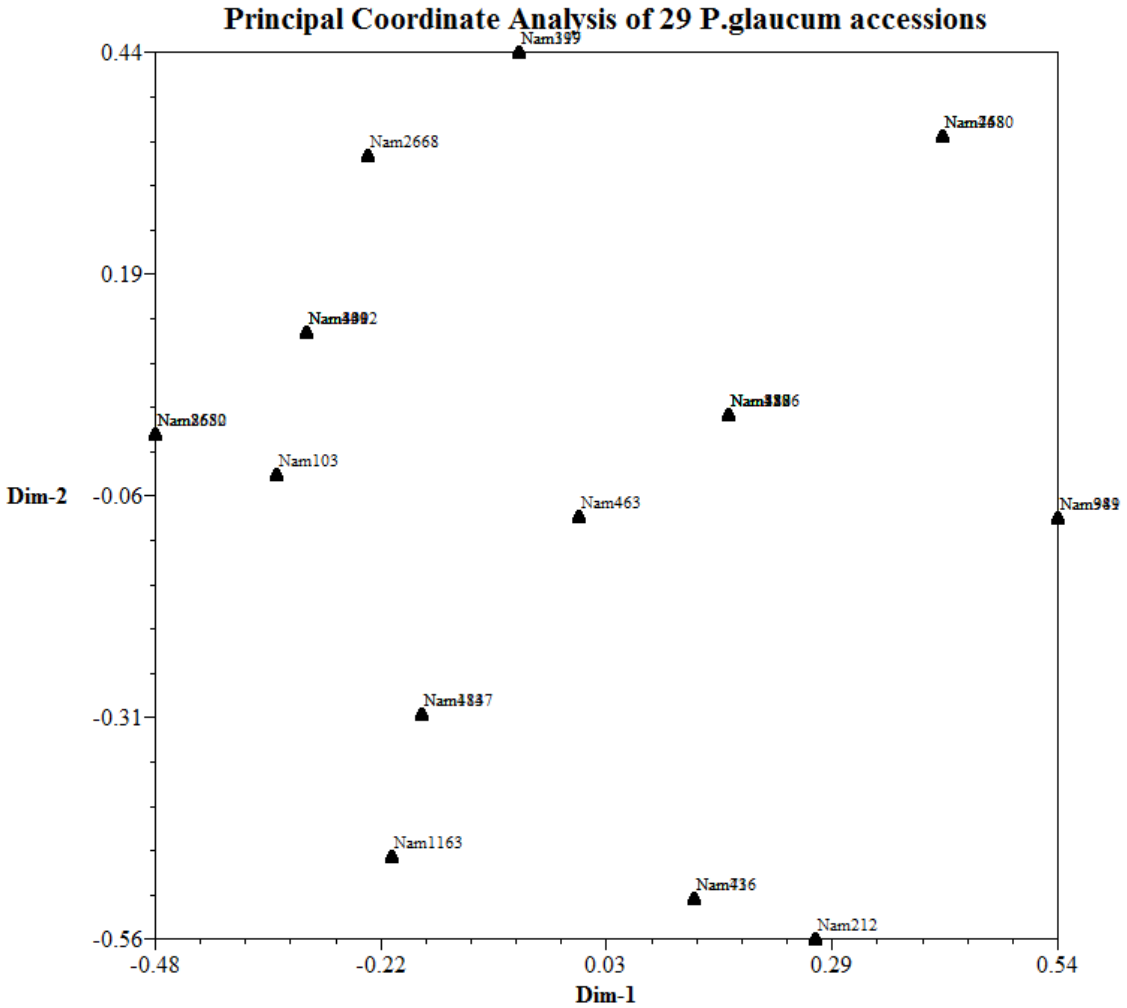


Figure 26: Principal Coordinate Analysis of the 29 *P. glaucum* genotypes based on RAPD data.

4.5.2.5 ITS phylogenetic analysis of 29 accessions

The twenty nine *P. glaucum* genotypes formed three main groups designated I, II and III. Group I had 8 sub-clusters designated 1-8 in which sub-cluster 1 of the eight sub-clusters was subdivided into two segments labeled A and B. Group II had a single accession while group III was composed of the out-group *P. purpureum*.

In sub-cluster 1, segment A (bootstrap value of 97%) was formed by 14 accessions, segment B (78% bootstrap value) was composed of 7 accessions. However, 2 accessions formed a small sub-cluster (77% bootstrap value), 4 accessions formed a small sub-cluster (71% bootstrap value). Furthermore, a smaller sub-cluster was formed by 3 accessions (bootstrap value of 88%). Another smaller sub-cluster was formed by 2 accessions (74% bootstrap value). Sub-clusters 2-8 consisted of single accessions. Group II was formed by a single accession Nam 349 while group III comprised of *P. purpureum* which was used as an out-group to root the tree. Accessions in segment A possess spindle shaped spikes and low lodging susceptibility whereas accessions in sub-clusters 2-8possess a compact spike density, short bristle and erect tillering attitude.

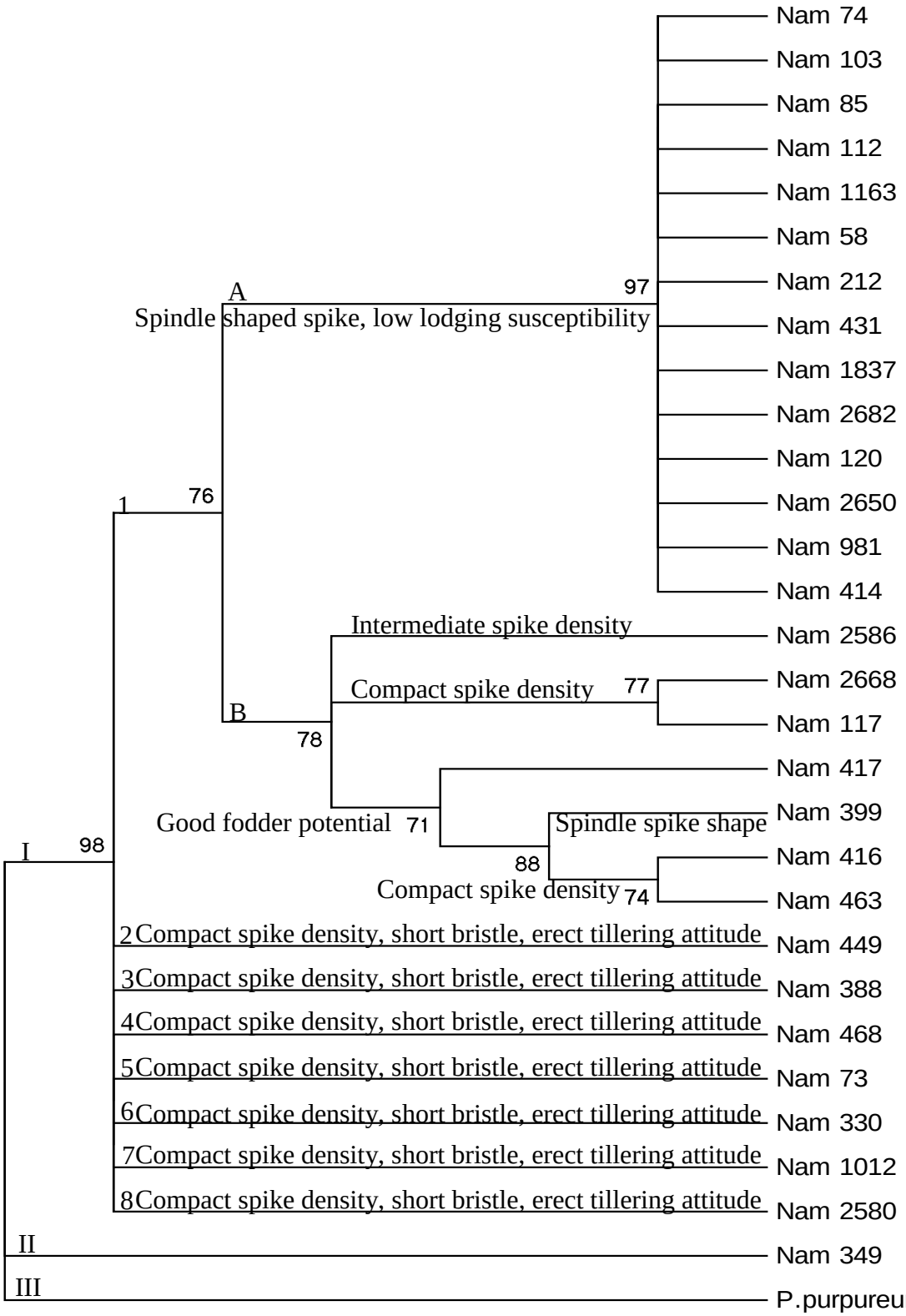


Figure 27: The phylogenetic tree for 29 *P. glaucum* genotypes showing some phenotypic traits based on ITS sequences inferred using the Maximum Likelihood method. *P. purpureum* was used as the out-group to root the tree generated with MEGA 6.0 software.

CHAPTER 5: DISCUSSION

Only reproducible bands were used for analysis and interpretation of data in the present study. To ensure reproducibility of the bands, the same template DNA was amplified in 3 different amplification reactions using the same primers. Only strong and reproducible bands were considered for further analyses. However, the unpredictable occurrence of null alleles may have possibly altered the observations on the gels; this could have had an implication on the low level of genetic diversity observed. Null alleles have been found to present problems in the interpretation of results in various studies (Varshney, Thiel, Stein, Langridge & Graner, 2002). Null alleles are alleles that do not give a polymerase chain reaction product. According to Varshney *et al.* (2002), null alleles were revealed by EST-SSR markers in studies on kiwifruit (Fraser, Harvey, Crowhurst & Desilva, 2004), rice (Cho *et al.*, 2000), spruce (Rungis *et al.*, 2004) and wheat (Eujayl, Sorrells, Baum, Wolters & Powell, 2001; Gupta *et al.*, 2003). Furthermore, Varshney *et al.* (2002) describes the incidence of null alleles pertaining to microsatellites as a consequence of the deletion of microsatellites at a specified locus and mutations (in-dels or substitutions) in the primer binding sites. This may result in the wrong interpretation of information because heterozygotes cannot be identified and reaction failures cannot be detected.

SSR primers (3005, 3035 and 3039) generated 318 bands while RAPD primers produced a total of 275 bands. These results showed that SSR markers were highly polymorphic and informative compared to RAPD primers which revealed less

polymorphism. This was observed through the high number and polymorphic appearance of SSR bands compared to RAPDs and complements the ability of SSR to effectively detect polymorphisms shown in other studies (Morgante & Olivieri, 1993; Powell, Machray & Provan, 1996) in which the observed frequence values were high than the expected in both plants (AT: observed-52, expected-12) and animals (AT: observed-140, expected-135). Microsatellites regions display high levels of polymorphism which is owed to unique replication slippage mechanism responsible for generating SSRs allelic diversity.

The present study reports PIC values ranging from 0.64 to 0.76 for SSRs and 0.66 to 0.82 for RAPDs. However other studies (Smith *et al.*, 1997) have reported higher values of 0.6 to 0.91 which could perhaps be attributed to the fact that they employed acryl amide gels for allele detection as opposed to agarose gels used in the present study. According to Agrama and Tuinstra (2004), acrylamide gels have greater resolving power than agarose gels. This makes acrylamide gels to have a higher preference due to their increased detection of large number of alleles per locus compared to agarose gels. Agrama and Tuinstra (2004) further argued that acrylamide gels are highly significant when employed in SSRs loci containing dinucleotide repeats whose amplification products are in the 130 to 200 bp range because agarose gels are unable to discriminate PCR products that vary by two base pairs. Primer pair 3005 is a dinucleotide repeat and is therefore subject to the limitation presented by the use of agarose gels for detection of alleles. Furthermore, other studies (Govindaraj *et al.*, 2009; Menkir, Goldsbrough & Ejeta, 1997; Tao, Manners, Ludlow & Henzell, 1993; Vierling,

Xiang, Joshi, Gilbert & Nguyen, 1994; Yang, De Oliveira, Godwin, Schertz & Bennetzen 1996) have reported a higher level of polymorphism with RAPDs compared to this study which could be attributed to the low number of RAPD primers employed in this study.

The results from the PIC of the primers used for analysis in present study indicated that the primers provided a significant amount of information regarding polymorphism. The PIC values for each primer used was greater than 0.5 which is the thresh hold for a primers ability to effectively detect polymorphism (Sharma *et al.*, 2009). On this basis, they were used for genetic diversity calculations using the Shannon diversity index. However, the Shannon diversity index revealed a low level of diversity. This low level of genetic diversity displayed by SSRs could perhaps be attributed to the type of primers used in that primer 3005 is a dinucleotide repeat, primer 3035 is a tetra-nucleotide repeat and primer 3039 is a trinucleotide repeat or there is simply no diversity at all and landraces are not really landraces but originated from a single parent. Using more primers that are specific to trinucleotide repeats might have possibly yielded a higher level of genetic diversity compared to the one achieved. Varshney *et al.* (2002) demonstrated that trinucleotide repeats (TNRs) are the most abundant followed either by dinucleotide repeats (DNRs) or tetra-nucleotide repeats (TTNRs). With cereal species, TNRs were the most frequent (54–78%) followed by DNRs (17.1–40.4%) and TTNRs (3–6%). Gao *et al.* (2004) as well as Kantety, La Rota, Matthews & Sorrells (2002) and also Morgante, Hanafey and Powell (2002) found that

wheat had a significant variation in the frequencies and distribution of different repeat motifs ranging from 49% to 83% for TNRs but DNRs and TTNR frequencies were low

The low level of polymorphism achieved might also have been caused by the primers used in the present study which are EST derived. Several research groups (Chabane, Ablett, Cordeiro, Valkoun & Henry, 2005; Cho *et al.*, 2000; Eujayl *et al.*, 2001; Rungis *et al.* 2004; Russell *et al.*, 2004; Scott *et al.*, 2000) have reported that EST-SSR primers are less polymorphic in crop plants compared with genomic derived SSRs due to better DNA sequence conservation in transcribed regions. In the present study, a low level of genetic diversity (0.7 and 0.45) was obtained from RAPDs and SSRs data which was revealed by the Shannon diversity index. The genetic diversity values (0.7 and 0.45) obtained are regarded as low in comparison to values found by Magurran (2004) and Pavoine and Bonsall (2011). They found Shannon diversity index values usually ranging between 1.5 and 3.5. The results of the present study therefore suggest that the selected *P. glaucum* genotypes have narrow genetic diversity, despite the high PIC value of primers used.

The results from the SSR and RAPD cluster analysis showed that the genotypes could only be grouped into two main clusters with a similarity level of 68% for SSR markers and 50% for RAPD markers. This level of similarity was relatively high indicating a low level of genetic diversity among the Namibian *P. glaucum* germplasm. However, it can be speculated that assessing genetic diversity of the current *P. glaucum* genotypes sampled directly from the fields of different regions would display a higher

level of genetic diversity accountable to possible mutation events that would have taken place from the time the accessions were collected (most accessions were collected in 1991). This is because the implications of mutation and migration play a huge role in the restoration of the genetic diversity and the influence of these two factors seem to be highly reduced in seed multiplication experiments that are routinely carried out at the NPGRC (National Gene Bank of Namibia).

Frankham *et al.* (2002) also argued that mutation and migration, and their interaction with selection have six important implications in conservation genetics: 1. the regeneration of genetic diversity due to mutation, 2. recovery of genetic diversity by migration, 3. migration often reverses inbreeding depression, 4. the impact of gene flow (migration) from related species (introgression), 5. maintenance of genetic diversity through mutation and migration, and 6. the load of deleterious alleles in population, this is caused by the balance between deleterious mutation and selection results in an ever present but changing pool of rare deleterious mutations (mutation load) in populations. It was further observed that the SSRs and RAPDs clusters did not form on the basis of the region of origin of the accessions but were mixed within the groupings. This could possibly be attributed to the wide germplasm exchange that occur between villages and could even extend among regions because some regions are the main source or producers of the crop compared to others. Kavango (East and West) region in Namibia is the largest producer of *P. glaucum.* Hence this region could also have been the source of germplasm for other regions through sharing between farmers. In addition, recurring drought and government's distribution of seed in Namibia is a contributing factor to the

low genetic diversity observed. When there is drought, the government distributes seeds to the community to compensate the farmer's losses. This leads to loss of genetic diversity because the landraces are lost to the drought and the seeds that are distributed then slowly replace the landraces.

This suggestion is supported by both Chakauya (2002) and Huvio (1999). They observed that famers acquired 80% of their seeds from their neighbors and approximately 60% of the farmers travel long distances in order to exchange or buy seeds in the quest to maintain the vigour of the crop. However, Chakauya (2002) reported that a high genetic diversity would be expected in landraces and open pollinated varieties of *P. glaucum* because pearl millet is naturally out crossing due to heterogeneity. In addition, Chakauya (2002) further suggested that genetic diversity is certainly lost in inbreeding depressions particularly through seed multiplication experiments that occur in a closed system.

Comparable results to the present study were also obtained by Mariac *et al.* (2006) who compared the amount of alleles between wild *P. glaucum* accessions and cultivated *P. glaucum* accessions. Their findings indicate a lower allelic richness in the cultivated *P. glaucum* compared to the wild *P. glaucum*. Their study also revealed that the cultivated accessions displayed the presence of less allele's (23%) in relation to the wild plants. In addition, they found an average gene diversity of 0.49 in the cultivated compared with 0.67 in the wild plants hence showing that cultivars have a lower diversity index. This gave an overall gene diversity of cultivated *P. glaucum* that is 28% lower than in wild *P. glaucum*. These findings seem to corroborate with the results from

the present study in that the *P. glaucum* landraces revealed a low genetic diversity of 0.45 and 0.7 that was demonstrated using both RAPDs and SSRs markers by the Shannon diversity index.

Previous studies have used Principal Component Analysis (PCA) to assess genetic diversity mostly between wild and cultivated accessions (Mariac *et al.*, 2006; Tostain, 1994). Tostain (1994) found that wild and cultivated accessions demonstrated a comparable distribution in the PCA analysis. Mariac *et al.* (2006) also found a noticeable variation between accessions of wild *P. glaucum* which displayed unique allele combinations and cultivated *P. glaucum* which revealed homogeneity. Similarly, a clear variation between accession groupings was achieved by PCoA in the present study. However the study revealed a low level of genetic diversity based on RAPDs, SSRs and ITS sequences whereas previous results by Tostain *et al.* (1987) and Tostain & Marchaise (1989) were based on isozymes analysis. The discrepancy could be therefore due to the different evolutionary properties of SSR and the ITS region, and isozyme loci.

The SSR, RAPD and ITS trees all showed a consistency above 70% with the morphology data phenogram. The percentage correlation of all the three techniques is significant and suggests that their clusters may well be used to identify *P. glaucum* genotypes and predict the phenotype of uncharacterized genotypes. However, both SSR and RAPD techniques can be preferably used over ITS sequence analysis to achieve

more information on the diversity of *P. glaucum* genotypes. This information is useful in resource management for germplasm conservation and breeding programs. The NPGRC will only store the accessions that are diverse; this will reduce the cost of storage incurred because duplicated accessions will be discarded and only genetically diverse accessions will be stored. Furthermore, there was no noticeable diversity pertaining to the regions of origin of the accessions for both the 29 analyzed accessions and the total sample size, this indicated a low level of genetic diversity among the Namibian germplasm.

The ITS phylogenetic tree showed that the 89 accessions of *P. glaucum*, the seven NCBI retrieved ITS sequences and the out-group *P. purpureum* formed three main groups. . The *P. purpureum* belongs to the secondary gene-pool of pearl millet while *P. glaucum* is part of the primary gene pool. Harlan and De Wet (1971) and Martel *et al.* (2004) argued that primary, secondary and tertiary gene pools are classified dependent on the ease with which the wild germplasm could be introgressed into the cultivated species. However, despite the aforesaid variation, *P. purpureum* and *P. glaucum* belong to the same morphological section of *Penicillaria*. On this basis; *P. purpureum* was used as the out-group in order to resolve the genetic diversity within the accessions. However, the ITS tree with the 89 accessions did not give any phylogenetic signal. In addition, the ITS phylogenetic tree produced less evolutionary information due to many polytomies. This suggests that the ITS region in *P. glaucum* diversity studies is not phylogenetically informative.

The 54 clusters (without the out-group) formed revealed the existence of a fairly low amount of genetic diversity within the accessions with 32% of the accessions belonging to one cluster highly supported by a bootstrap value of 95%. The analysis showed that certain accessions (a total of 31 accessions) have a closer genetic association compared to the rest. This is probably a result of the occurrence of concerted evolution among these accessions. Concerted evolution is the tendency of the different genes in a gene family or cluster to evolve in concert (Zhang *et al.*, 2002). Concerted evolution has been found to be common in the *Poaceae* family as described by Hsiao, Chatterton, Asay and Jensen (1995).

Several accessions (a total of 43 accessions) further formed individual groups signifying the existence of variation. The minor variations are possibly a consequence of insertions, substitutions and deletions in the bases which were evident by the multiple alignments (see Appendix D, for alignments). These results are similar to the findings of Ainouche and Bayer (1997) reporting the occurrence of diverse types of ITS repeats within individuals of some polyploid species in *Bromus* (*Poaceae*). Furthermore, an increased level of polymorphism in the 5S RNA gene of various diploid species from the *Triticeae* and *Poaceae,* such as cotton, has been revealed by Kellogg and Appels (1995), and Cronn, Zhao, Paterson and Wendel (1996).

Small clades were also observed which comprised of two to three accessions each. These branches were highly supported and disclosed that the accessions in the foresaid clusters share a high degree of similarity. This was also supported by

performing some pairwise alignment similarity calculations for some of the sequences within the clusters which gave identities of 0.94 for Nam 138 and Nam 621, 0.99 for Nam 399 and Nam 1094, and 0.97 for Nam 388 and Nam 2722.

In general, high similarities among clusters can also be attributed to the occurrence of concerted evolution within the nucleotide variations of the Internal Transcribed Spacer regions as well as the occurrence of intense gene flow. Gene flow is effective among biological species of the same kind as shown by cytogenetic and molecular studies (Zhang *et al.*, 2002). In addition, gene flow has been observed between species such as the triploid sterile hybrids of *P. purpureum* and the domesticated form *P. glaucum* (Hanna, 1986). The inclusion of the seven NCBI retrieved ITS sequences revealed that these sequences were fairly similar to the Namibian germplasm because most of them were under the main group I and some under the sub-group A. However, four (AY628132.1, 628130.1, 628131.1 and 628129.1) of the seven retrieved sequences grouped under a 100% bootstrap value suggesting that they are highly similar to each other compared to the accessions under investigation. In addition accessions AY628131.1 and AY628129.1 formed a sub-cluster with a 100% bootstrap which revealed that they had identical ITS sequences. Accessions AY628116.1 and AY628115.1 were grouped under a separate group B supported by a 100% bootstrap which also indicated that they have identical sequences whereas accession FJ766182.1 formed a separate group II which showed that it was more dissimilar compared to the other six NCBI retrieved sequences. A pairwise alignment also supported the resemblance of AY628131.1 and AY628129.1 by a 0.99

identity value whereas accessions AY628116.1 and AY628115.1 were supported by an identity value of 0.98 which suggests that the pairs of sequences are identical.

A comparison between the morphological data derived tree and the RAPD data derived tree showed less correspondence compared to ITS and SSR. Morphological data UPGMA clustering of 29 accessions revealed the formation of: two main clusters at 78% similarity level while the UPGMA clustering of SSR data displayed two main clusters at a similarity level of 69%. The UPGMA clustering of RAPD data showed two main clusters at a similarity level of 60%. In addition, the similarity level (69%) and cluster B of the SSR data phenogram was closely related to that of morphological data with a 78% resemblance while the similarity level (60%) and cluster I of the RAPD data phenogram resembled that of morphological data by 72%.

The minor variations in the particular accessions found in clusters of the ITS phylogenetic tree, and SSR and RAPD phenograms in comparison to the observed morphology clustering based on the ten selected discrete traits are possibly a consequence of the few number of selected traits. Environmental factors, such as soil composition, temperature and rainfall, influence the resulting morphology (appearance) of plants.

The morphology data phenogram (see Figure 21) disclosed that clustering of the accessions was as a result of possessing similar traits of; fodder potential, spike shattering, lodging susceptibility, early vigour, spike density, bristle length, number of nodal tillers and plant aspects. There were some overlaps in the observed traits in the

tree which is probably because some genetic characteristics occur along a continuum. That is a trait can be expressed over a range of phenotypes such as strongly, moderate or weakly in the same genotypes. Klug and Cumming (2003) also reported that variable expressivity occurs when variability in the level of phenotypic expression is observed from the same genotype.

The RAPD phenogram (see Figure 25) showed that some accessions were clustered based on a low lodging susceptibility, number of nodal tillers, tillering attitude, spike density, bristle length and green fodder yield potential.

From the SSR phenogram of the 29 morphologically characterized genotypes, four traits were clearly identified upon which some clusters were formed (see Figure 23). This suggests that these traits are strongly expressed in the genotypes under these clusters. Lodging susceptibility, tillering attitude, spike density and spike shape were identified as the strongly expressed characters. The cluster in group 7 was formed on the basis that accessions possessed a low lodging susceptibility. Lodging is the permanent dislocation of stems from the erect position. Lodging is mostly problematic in times of heavy rain and strong winds.

According to Verma *et al.* (2005), lodging has been a problem in cereals for a long time especially after storms that leave the whole fields of cereals flattened. However, Karim and Jahan (2013) argued that lodging resistance can be improved by selecting amongst the progeny plants for shorter and solid stems. Group 4 revealed that erect tillering attitude was intensely expressed by 6 accessions whereas group 3 was formed by accession Nam 103 on the basis of its possession of an intermediate tillering

attitude. Group 2 individuals strongly expressed their possession of a compact spike density.

The ITS phylogenetic tree (see Figure 27) clusters were mostly formed based on the spike density, bristle length, tillering altitude, folder potential, lodging susceptibility and spikelet shattering. The ITS sequence clusters displayed more resolution as compared to the SSR clusters.

CHAPTER 6: CONCLUSIONS

The genetic diversity of Namibian *P. glaucum* germplasm was determined by analyzing the SSRs and RAPDs molecular markers, and the ITS sequence analysis. The three techniques confirmed that the Namibian *P. glaucum* germplasm have a low genetic diversity. Furthermore, it was shown that microsatellite analysis was the most informative followed by RAPDs and lastly ITS sequences analysis. The study also showed that ITS sequences were less informative in terms of phylogeny and the ITS region does not give resolution compared to SSR and RAPD because of concerted evolution. However, on the basis of clusters corresponding to morphological data clustering, the ITS sequence clusters were the most consistent followed by SSR markers which were in turn more revealing than RAPD markers.

CHAPTER 7: RECOMMENDATIONS

The findings of this study support the recommendation that; efforts must be made to widen the available gene pool and germplasm in order to increase the genetic diversity of *P. glaucum.* In addition, the germplasm must be evaluated from time to time to investigate the level of genetic diversity. The present study demonstrates the importance of preferably using SSR and RAPD techniques in order to efficiently investigate the genetic diversity of *P. glaucum*. It is further recommended that a similar study should be carried out using recent collected accessions and the number of accessions in the analysis must be increased due to some inherent limitations that occur in the sampling methods and other regions of the DNA should be investigated for genetic diversity.

There is a need for molecular characterization of all the accession in the Namibian germplasm collection in order to select the most divergent genotypes with favorable phenotypes which can be used in hybridization programs in order to improve the crop genetically. Since crop improvement programs are ongoing initiatives, it is imperative that information regarding genetic diversity of *P. glaucum* is continuously assimilated on recent and old accessions not included in this study. It is also imperative to include farmers in an attempt to conserve and improve the crop (*P. glaucum*) because they also play a role in the maintenance of the gene pool in the country. It appears that offering agricultural trainings or workshops to farmers on the conservation of genetic resources would help solidify the advancement in the crop. This training will allow

participatory plant breeding in which farmers and plant breeder's work together to develop plant varieties.

CHAPTER 8: REFERENCES

Agrama, H. A., & Tuinstra, M. R. (2004). Phylogenetic diversity and relationships among sorghum accessions using SSRs and RAPDs. *African Journal of Biotechnology, 2* (10), 334-340.

Ainouche, M. L., & Bayer, R. J. (1997). On the origins of the tetraploid *Bromus* species (section *Bromus, Poaceae*): insights from Internal Transcribed Spacer sequences of nuclear ribosomal DNA. *Genome, 40* (5), 730-743.

Akkaya, M. S., Bhagwat, A. A. & Cregan, P. B. (1992). Length polymorphism of simple sequence Repeat DNA in Soybean. *Journal of Genetics*. *132* (4), 1131-1139.

Anjanadevi, I. P., John, N. S., John, K. S., Jeeva, M. L., & Misra, R. S. (2013). Isolation and Characterization of Nitrogen Fixing Bacteria from Elephant Foot Yam (*Amorphophallus paeoniifolius*) growing regions of South India. *Journal of Root Crops*, *39*(2), 154-162.

Arens, P., Coops, H., Jansen, J., & Vosman, B. (1998). Molecular genetic analysis of black poplar (*Populus nigra L.*) along Dutch rivers. *Molecular Ecology, 7* (1), 11-18.

Arkhipova, T. N., Veselov, S. U., Melentiev, A. I., Martynenko, E. V., & Kudoyarova, G. R. (2005). Ability of bacterium *Bacillus subtilis* to produce cytokinins and to influence the

growth and endogenous hormone content of lettuce plants. *Plant and Soil, 272* (1-2), 201-209.

Arnheim, N. (1983). Concerted evolution of multigene families. In M. Nei and R.K. Koehn (Eds.), *Evolution of genes and proteins* (38–61). Sinauer, Sunderland, Massachusetts, USA.

Asay, K. H., Hsiao, C., Chatterton, N. J., Jensen, K. B., & Wang, R. R. (1992). Nucleotide sequence of the internal transcribed spacer (ITS) region of rDNA in mountain rye, Secale montanum Guss.(Gramineae). *Plant Molecular Biology, 20* (1), 161-162.

Asghar, H., Zahir, Z., Arshad, M., & Khaliq, A. (2002). Relationship between in vitro production of auxins by rhizobacteria and their growth-promoting activities in *Brassica juncea* L. *Biology and Fertility of Soils, 35* (4), 231-237.

Baes, P., & Cutsem, P. V. (1993). Electrophoretic analysis of eleven isozyme systems and their possible use as biochemical markers in breeding of chicory (Cichorium intybus L.). *Plant breeding, 110* (1), 16-23.

Bailey, C.D., Carr, T.G., Harris, S.A., & Hughes, C.E. (2003). Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Molecular Phylogenetics and Evolution, 29* (3), 435–455.

Barcaccia, G. (2009). Molecular markers for characterizing and conserving crop plant germplasm. In *Molecular Techniques in Crop Improvement* (231-254). Springer Netherlands.

Bidinger, F. R., Mahalakshmi, V., & Rao, G. D. P. (1987). Assessment of drought resistance in pearl millet [*Pennisetum americanum (L.) Leeke*]. I. Factors affecting yields under stress. *Crop and Pasture Science*, *38* (1), 37-48.

Board on Science and Technology for International Development. (1996). *Lost Crops of Africa*. Vol 1. Washington , D.C: National Academy Press.

Bowman, L. H., Goldman, W. E., Goldberg, G. I., Hebert, M. B., & Schlessinger, D. (1983). Location of the initial cleavage sites in mouse pre-rRNA. *Molecular and Cellular Biology*, *3* (8), 1501-1510.

Brown, A. H. (2008). Indicators of genetic diversity, genetic erosion and genetic vulnerability for plant genetic resources for food and agriculture. *FAO, Rome*.

Bruford, M. W., & Wayne, R. K. (1993). Microsatellites and their application to population genetic studies. *Current Opinion in Genetics & Development*, *3* (6), 939-943.

Brunken, J. N. (1977). A systematic study of *Pennisetum* sect. *Pennisetum* (Gramineae). *American Journal of Botany*, 161-176.

Buth, D. G., & Murphy, R. W. (1999). The use of isozyme characters in systematic studies. *Biochemical Systematics and Ecology*, *27* (2), 117-129.

Chabane, K., Ablett, G. A., Cordeiro, G. M., Valkoun, J., & Henry, R. J. (2005). EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley. *Genetic Resources and Crop Evolution*, *52* (7), 903-909.

Chakauya, E. (2002). Genetic diversity assessment of sorghum *[Sorghum bicolor (L.) Moench]* and Pearl millet *(Pennisetum glaucum L. R. Br.)* landraces collected from Zimbabwe's communal areas, based on microsatellites and isozymes (Master's thesis, University of Zimbabwe).

Chandra-Shekara, A. C., Prasanna, B. M., Bhat, S. R., & Singh, B. B. (2007). Genetic diversity analysis of elite pearl millet inbred lines using RAPD and SSR markers. *Journal of Plant Biochemistry and Biotechnology*, *16* (1), 23-28.

Chase, M. R., Moller, C., Kesseli, R., & Bawa, K. S. (1996). Distant gene flow in tropical trees. *Journal of Nature*, *383* (6599), 398-399.

Chatterton, N. J., Hsiao, C., Asay, K. H., Jensen, K. B., & Wang, R. R. (1992). Nucleotide sequence of the internal transcribed spacer region of rDNA in barley, *Hordeum vulgare* L.(Gramineae). *Plant Molecular Biology*, *20* (1), 165-166.

Chatterton, N.J., Hsiao, C., Asay, K.H., Wang, R.R-C. & Jensen, K.B. (1992). Nucleotide sequence of the internal transcribed spacer region of rDNA in diploid wheat, *Triticum speltoides* L. (*Tausch*) Gren. ex Richter (Gramineae). *Plant Molecular Biology, 20* (1), 157-158.

Cho, Y. G., Ishii, T., Temnykh, S., Chen, X., Lipovich, L., McCOUCH, S. R.,... & Cartinhour, S. (2000). Diversity of microsatellites derived from genomic libraries and Gene Bank sequences in rice (*Oryza sativa L.*). *Theoretical and Applied Genetics*, *100* (5), 713-722.

Chowdari, K. V., Davierwala, A. P., Gupta, V. S., Ranjekar, P. K., & Govila, O. P. (1998). Genotype identification and assessment of genetic relationships in pearl millet [*Pennisetum glaucum (L.) R. Br*] using microsatellites and RAPDs. *Theoretical and Applied Genetics*, *97* (1-2), 154-162.

Chowdhury, M. K. V., & Smith, R. L. (1988). Mitochondrial DNA variation in pearl millet and related species. *Theoretical and Applied Genetics*. *76* (1), 25–32.

Ciofi, C., Funk, S. M., Coote, T., Cheesman, D. J., Hammond, R. L., Saccheri, I. J., & Bruford, M. W. (1998). Genotyping with microsatellite markers. In *Molecular Tools for Screening Biodiversity* (195-201). Springer, Netherlands.

Clarke, K. R., & Gorley, R. N. (2001). PRIMER v5. *PRIMER-E Ltd*.

Clegg, M. T., Rawson, J. R., & Thomas, K. (1984). Chloroplast DNA variation in pearl millet and related species. *Genetics*, *106* (3), 449-461.

Corriveau, J. L., & Coleman, A. W. (1988). Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany*, 1443-1458.

Cronn, R.C., Zhao, X.P., Paterson, A.H., & Wendel, J.F. (1996). Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *Journal of Molecular Evolution*, *42*, 685–705.

Dayanandan, S., Rajora, O. P., & Bawa, K. S. (1998). Isolation and characterization of microsatellites in trembling aspen (*Populus tremuloides*). *Theoretical and Applied Genetics*, *96* (6-7), 950-956.

De Freitas, J. R., Banerjee, M. R., & Germida, J. J. (1997). Phosphate-solubilizing rhizobacteria enhance the growth and yield but not phosphorus uptake of canola (*Brassica napus L.*). *Biology and Fertility of Soils*, *24* (4), 358-364.

De Vicente, M. C., & Fulton, T. (2003). Using molecular marker technology in studies on plant genetic diversity. *Illus. Nelly Giraldo. IPGRI, Rome, Italy and Institute for Genetic Diversity, Ithaca, New York, USA. ISBN*, 92-9043.

Dillon, S. L., Lawrence, P. K., & Henry, R. J. (2001). The use of ribosomal ITS to determine phylogenetic relationships within sorghum. *Plant Systematics and Evolution*, *230* (1-2), 97-110.

Dillon, S. L., Lawrence, P. K., Henry, R. J., & Price, H. J. (2007). Sorghum resolved as a distinct genus based on combined ITS1, *ndhF* and *Adh1* analyses. *Plant Systematics and Evolution*, *268* (1-4), 29-43.

Dillon, S. L., Lawrence, P. K., Henry, R. J., Ross, L., Price, H. J., & Johnston, J. S. (2004). *Sorghum laxiflorum* and *Sorghum macrospermum*, the Australian native species most closely related to the cultivated *Sorghum bicolor* based on ITS1 and *ndhF* sequence analysis of 25 sorghum species. *Plant Systematics and Evolution*, *249* (3-4), 233-246.

Doust, A. N., & Kellogg, E. A. (2002). Inflorescence diversification in the panicoid "bristle grass" clade (*Paniceae, Poaceae*): evidence from molecular phylogenies and developmental morphology. *American Journal of Botany*, *89* (8), 1203-1222.

Doyle, J. (1991). DNA protocols for plants. In *Molecular techniques in taxonomy* (283-293). Springer, Berlin Heidelberg.

Erskine, W., & Muehlbauer, F. J. (1991). Allozyme and morphological variability, outcrossing rate and core collection formation in lentil germplasm. *Theoretical and Applied Genetics*, *83* (1), 119-125.

Eujayl, I., Sorrells, M., Baum, M., Wolters, P., & Powell, W. (2001). Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. *Euphytica*, *119* (1-2), 39-43.

Falkenhagen, E. R. (1985). Isozyme studies in provenance research of forest trees. *Theoretical and Applied Genetics*, *69* (4), 335-347.

Frankham, R., Briscoe, D. A., & Ballou, J. D. (2002). *Introduction to conservation genetics*. Cambridge University Press.

Fraser, L. G., Harvey, C. F., Crowhurst, R. N., & De Silva, H. N. (2004). EST-derived microsatellites from *Actinidia* species and their potential for mapping. *Theoretical and Applied Genetics*, *108* (6), 1010-1016.

Fussell, L. K., Pearson, C. J., & Norman, M. J. T. (1980). Effect of temperature during various growth stages on grain development and yield of *Pennisetum americanum*. *Journal of Experimental Botany*, *31* (2), 621-633.

Gao, L. F., Jing, R. L., Huo, N. X., Li, Y., Li, X. P., Zhou, R. H., ... & Jia, J. Z. (2004). One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. *Theoretical and Applied Genetics*, *108* (7), 1392-1400.

Gao, L., Tang, J., Li, H., & Jia, J. (2003). Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Molecular Breeding*, *12* (3), 245-261.

Gaskins, M. H., Albrecht, S. L., & Hubbell, D. H. (1985). Rhizosphere bacteria and their use to increase plant productivity: a review. *Agriculture, Ecosystems & Environment*, *12* (2), 99-116.

Gepts, P., & Clegg, M. T. (1989). Genetic diversity in pearl millet (*Pennisetum glaucum [L.] R. Br.*) at the DNA sequence level. *Journal of Heredity*, *80* (3), 203-208.

Giussani, L. M., Cota-Sánchez, J. H., Zuloaga, F. O., & Kellogg, E. A. (2001). A molecular phylogeny of the grass subfamily Panicoideae (Poaceae) shows multiple origins of C4 photosynthesis. *American Journal of Botany*, *88* (11), 1993-2012.

Glaszmann, J. C., Fautret, A., Noyer, J. L., Feldmann, P., & Lanaud, C. (1989). Biochemical genetic markers in sugarcane. *Theoretical and Applied Genetics*, *78* (4), 537-543.

González, M., Rodríguez, R., Zavala, M. E., Jacobo, J. L., Hernández, F., Acosta, J., ... & Simpson, J. (1998). Characterization of Mexican isolates of *Colletotrichum lindemuthianum* by using differential cultivars and molecular markers. *Phytopathology*, *88*(4), 292-299.

Goryunova, S. V., Chikida, N. N., Gori, M., & Kochieva, E. Z. (2005). Analysis of nucleotide sequence polymorphism of internal transcribed spacers (ITS) of ribosomal genes in diploid *Aegilops* (L.) species. *Molecular Biology*, *39* (2), 173-176.

Govindaraj, M., Selvi, B., Prabhu, D. A., & Rajarathinam, S. (2009). Genetic diversity analysis of pearl millet (*Pennisetum glaucum [L.] R. Br.*) accessions using molecular markers. *African Journal of Biotechnology*, *8* (22).

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, *53* (3-4), 325-338.

Grimaldi, M. C., & Crouau-Roy, B. (1997). Microsatellite allelic homoplasy due to variable flanking sequences. *Journal of Molecular Evolution*, *44* (3), 336-340.

Gupta, P. K., Rustgi, S., Sharma, S., Singh, R., Kumar, N., & Balyan, H. S. (2003). Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Molecular Genetics and Genomics*, *270* (4), 315-323.

Gutiérrez-Mañero, F. J., Ramos-Solano, B., Probanza, A., Mehouachi, J. R., Tadeo, F., & Talon, M. (2001). The plant-growth-promoting rhizobacteria *Bacillus pumilus* and *Bacillus licheniformis* produce high amounts of physiologically active gibberellins. *Physiologia Plantarum, 111* (2), 206-211.

Hadrys, H., Balick, M., & Schierwater, B. (1992). Applications of random amplified polymorphic DNA (RAPD) in Molecular Ecology. *Molecular Ecology, 1* (1), 55-63.

Hall, T. (1998). BioEdit. Biological sequence alignment editor for Windows. *North Carolina, USA: Carolina State University*.

Hamrick, J. L., & Godt, M. J. W. (1997). Allozyme diversity in cultivated crops. *Crop Science, 37* (1), 26-30.

Hanna, W. W. (1986, April). Utilization of wild relatives of pearl millet. In *Proceedings of the International Pearl Millet Workshop* (7-11).

Harlan, J. R., & De Wet, J. M. (1971). Toward a rational classification of cultivated plants. *Taxon*, 509-517.

Heip, C. H., Herman, P. M., & Soetaert, K. E. R. (1998). Indices of diversity and evenness. *Oceanis, 24*.

Hershkovitz, M. A., & Zimmer, E. A. (1996). Conservation patterns in angiosperm rDNA ITS2 sequences. *Nucleic Acids Research, 24* (15), 2857-2867.

Hirata, T., & Takamatsu, S. (1996). Nucleotide sequence diversity of rDNA internal transcribed spacers (ITS) extracted from conidia and cleistothecia of several powdery mildew fungi. *Mycoscience*, *37* (3), 283-288.

Hokanson, S. C., Szewc-McFadden, A. K., Lamboy, W. F., & McFerson, J. R. (1998). Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus domestica Borkh*. core subset collection. *Theoretical and Applied Genetics*, *97* (5-6), 671-683.

Holden, J., & Peacock, W. J. (1993). *Genes, crops and the environment*. Cambridge University Press.

Hsiao, C., Chatterton, N., Asay, K. & Jensen, K. (1995) Phylogenetic relationships of the monogenomic species of the wheat tribe Triticeae (Poaceae) inferred from nuclear rDNA (ITS) sequences. *Genome, 38*, 211–223.

Huvio, T. 1999. Women in Development Service. In: *Development of Strategies for in Situ Conservation of Plant Genetic Resources for Food and Agriculture in the Semi-arid areas of Zimbabwe*. *Annual review workshop*. Genebank of Zimbabwe, IPGRI and FAO Collaborative Project. DR & SS, Harare, Zimbabwe, 23-45.

Jauhar, P. P. (1981). *Cytogenetics and breeding of pearl millet and related species*. Alan R. Liss Inc. Progress and Topics in Cytogenetics, Volume 1.

Jobes, D. V., & Thien, L. B. (1997). A Conserved Motif in the 5.8S Ribosomal RNA (rRNA) Gene is a Useful Diagnostic Marker for plant internal transcribed spacer (ITS) sequences. *Plant Molecular Biology Reporter*, *15* (4), 326–334.

Joo, G. J., Kim, Y. M., Lee, I. J., Song, K. S., & Rhee, I. K. (2004). Growth promotion of red pepper plug seedlings and the production of gibberellins by *Bacillus cereus*, *Bacillus macroides* and *Bacillus pumilus*. *Biotechnology Letters*, *26* (6), 487-491.

Kantety, R. V., La Rota, M., Matthews, D. E., & Sorrells, M. E. (2002). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Molecular Biology*, *48* (5-6), 501-510.

Karim, M.D. H., & Jahan, M. A. (2013). Study of lodging resistance and its associated traits in bread wheat. *Journal of Agricultural and Biological Science, 8* (10). 683-687.

Karp, A., Seberg, O. L. E., & Buiatti, M. (1996). Molecular techniques in the assessment of botanical diversity. *Annals of Botany*, *78* (2), 143-149.

Karvonen, P., Szmidt, A. E., & Savolainen, O. (1994). Length variation in the internal transcribed spacers (ITS) of ribosomal DNA in *Picea abies* and related species. *Theoretical and Applied Genetics*, *89* (7-8), 969-974.

Kellogg, E. A., & Appels, R. (1995). Intaspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. *Genetics*, *140* (1), 325–343.

Khalfallah, N., Sarr, A., & Siljak-Yakovlev, S. (1993) Karyological study of some cultivated and wild stocks of pearl millet from Africa (*Pennisetum typhoides Stapf* and *Hubb.* and *Pennisetum violaceum Rich*). *Caryologia 46* (2-3):127–138.

Klug, W.S. & Cumming, M.R. (2003). *Concepts of Genetics* (No. Ed. 7). Pearson Education Inc.

Korabečná, M., Liška, V., & Fajfrlik, K. (2003). Primers ITS1, ITS2 and ITS4 detect the intraspecies variability in the internal transcribed spacers and 5.8 S rRNA gene regions in clinical isolates of fungi. *Folia Microbiologica, 48* (2), 233-238.

Krauss, S. L. (1999). Complete exclusion of nonsires in an analysis of paternity in a natural plant population using amplified fragment length polymorphism (AFLP). *Molecular Ecology, 8* (2), 217-226.

Kumar, P., Gupta, V. K., Misra, A. K., Modi, D. R., & Pandey, B. K. (2009). Potential of molecular markers in plant biotechnology. *Journal of Plant Omics, 2* (4), 141-162.

Lee, S. B., & Taylor, J. W. (1992). Phylogeny of five fungus-like protoctistan Phytophthora species, inferred from the internal transcribed spacers of ribosomal DNA. *Molecular Biology and Evolution, 9* (4), 636-653.

Li, Y. C., Korol, A. B., Fahima, T., Beiles, A., & Nevo, E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology, 11* (12), 2453-2465.

Liston, A., Robinson, W. A., Oliphant, J. M., & Alvarez Buylla, E. R. (1996). Length variation in the nuclear ribosomal DNA internal transcribed spacer (ITS) region of nonflowering plants. *Systemic Botany*. *21* (2), 109–120.

Liu, J. S., & Schardl, C. L. (1994). A conserved sequence in internal transcribed spacer (ITS) 1 of plant nuclear rRNA genes. *Plant Molecular Biology*, *26* (2), 775-778.

Mace, E. S., Gebhardt, C. G., & Lester, R. N. (1999). AFLP analysis of genetic relationships in the tribe Datureae (Solanaceae). *Theoretical and Applied Genetics*, *99* (3-4), 634-641.

Magurran, A. E. (2004). Measuring biological diversity. *African Journal of Aquatic Science*, *29* (2), 285-286.

Maheshwari, J. K., Council of Scientific & Industrial Research (Inde). (1963). *The flora of Delhi*. New Delhi: Council of Scientific & Industrial Research.

Majer, D., Lewis, B. G., & Mithen, R. (1998). Genetic variation among field isolates of *Pyrenopeziza brassicae*. *Plant Pathology*, *47* (1), 22-28.

Mariac, C., Luong, V., Kapran, I., Mamadou, A., Sagnard, F., Deu, M., ... & Vigouroux, Y. (2006). Diversity of wild and cultivated pearl millet accessions (*Pennisetum glaucum [L.] R. Br.*) in Niger assessed by microsatellite markers. *Theoretical and Applied Genetics*, *114* (1), 49-58.

Martel, E., Poncet, V., Lamy, F., Siljak-Yakovlev, S., Lejeune, B., & Sarr, A. (2004). Chromosome evolution of *Pennisetum* species (Poaceae): Implications of ITS phylogeny. *Plant Systematics and Evolution*, *249* (3-4), 139-149.

Martel, E., Ricroch, A., & Sarr, A. (1996) Assessment of genome organization among diploid species (2n¼2x¼14) belonging to primary and tertiary gene pools of pearl millet using fluorescent in situ hybridization with rDNA probes. *Genome, 39* (4), 680–687.

Maundu, P. M., Ngugi, G. W., & Kabuye, C. H. S. (1999). *Traditional Foods of Kenya*. Kenya Resource Centre for indigenous knowledge (KENRIK), National Museum of Kenya. *Nairobi, Kenya.*

Menkir, A., Goldsbrough, P., & Ejeta, G. (1997). RAPD based assessment of genetic diversity in cultivated races of sorghum. *Crop Science*, *37* (2), 564-569.

Mondini, L., Noorani, A., & Paguotta, M.A. (2009). Assessing plant genetic diversity by molecular tools. *Diversity, 1* (1), 19-35.

Morgante, M., & Olivieri, A. M. (1993). PCR-amplified microsatellites as markers in plant genetics. *The Plant Journal*, *3* (1), 175-182.

Morgante, M., Hanafey, M., & Powell, W. (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics*, *30* (2), 194-200.

Mort, M. E., & Crawford, D. J. (2004). The continuing search: low-copy nuclear sequences for lower-level plant molecular phylogenetic studies. *Taxon*, 257-261.

Mueller, U. G., & Wolfenbarger L. L. (1999). AFLP genotyping and fingerprinting. *Tree, 14* (10), 389-394.

Mueller, U. G., Lipari, S. E., & Milgroom, M. G. (1996). Amplified fragment length polymorphism (AFLP) fingerprinting of symbiotic fungi cultured by the fungus-growing ant *Cyphomyrmex minutus*. *Molecular Ecology*, *5* (1), 119-122.

Nei, M. (1987). *Molecular evolutionary genetics*. Columbia University Press.

Neto, R. M., Von Pinho, E. V. R., Carvalho, B. L., & Pereira, G. S. (2013). Identification of earl millet cultivars using both microsatellites and enzymatic markers. *Genetics and Molecular Research*, *12*(1), 1-14.

O'Connell, M., & Wright, J. M. (1997). Microsatellite DNA in fishes. *Reviews in Fish Biology and Fisheries*, *7* (3), 331-363.

O'Donnell, K. (1992). Ribosomal DNA internal transcribed spacers (ITS) are highly divergent in the phytopathogenic *ascomycete Fusarium sambucinum* (*Gibberella pulicaris*). *Current Genetics*, *22* (3), 213-220.

Pavoine, S., & Bonsall, M. B. (2011). Measuring biodiversity to explain community assembly: a unified approach. *Biological Reviews, 86*(4), 792-812.

Powell, W., Machray, G. C., & Provan, J. (1996). Polymorphism revealed by simple sequence repeats. *Trends in Plant Science*, *1* (7), 215-222.

Provan, J., Russell, J. R., Booth, A., & Powell, W. (1999). Polymorphic chloroplast simple sequence repeat primers for systematic and population studies in the genus *Hordeum*. *Molecular Ecology*, *8* (3), 505-511.

Pushpavathi, B. (2003). Pathogenic And Genetic Diversity In Populations Of *Sclerospora Graminicola* The Incitant Of Downy Mildew In Pearl Millet (Doctoral dissertation, Acharya NG Ranga Agriculture University, Hyderabad, India).

Reed, K.M., & Philips, R.B. (2000). Structure and organization of the rDNA intergenic spacer in lake trout (*Salvelinus Namaycush*). *Chromosome Research, 8* (1), 5–16.

Richard, I., & Beckmann, J. S. (1995). How neutral are synonymous codon mutations? *Nature Genetics*, *10* (3), 259-259.

Robert, T., Lamy, F., & Sarr, A. (1992). Evolutionary role of gametophytic selection in the domestication of *Pennisetum typhoides* (pearl millet): a two-locus asymmetrical model. *Heredity*, *69*, 372-381.

Robinson, J. P., & Harris, S. A. (1999). Amplified fragment length polymorphisms and microsatellites: a phylogenetic perspective. *Which DNA Marker for which Purpose*, *20*.

Rohlf, F. J. (1992). *NTSYS-pc: numerical taxonomy and multivariate analysis system*. Applied Biostatistics.

Rokhbakhsh-Zamin, F., Sachdev, D., Kazemi-Pour, N., Engineer, A., Pardesi, K. R., Zinjarde, S., ... & Chopade, B. A. (2011). Characterization of plant-growth-promoting traits of

Acinetobacter species isolated from rhizosphere of *Pennisetum glaucum*. *Journal of Microbiology & Biotechnology, 21* (6), 556-566.

Rosendahl, S., & Taylor, J. W. (1997). Development of multiple genetic markers for studies of genetic variation in arbuscular mycorrhizal fungi using AFLP. *Molecular Ecology, 6* (9), 821-829.

Rossetto, M., Slade, R. W., Baverstock, P. R., Henry, R. J., & Lee, L. S. (1999). Microsatellite variation and assessment of genetic structure in tea tree (*Melaleuca alternifolia*–Myrtaceae). *Molecular Ecology*, *8* (4), 633-643.

Rungis, D., Bérubé, Y., Zhang, J., Ralph, S., Ritland, C. E., Ellis, B. E., ... & Ritland, K. (2004). Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *Theoretical and Applied Genetics*, *109* (6), 1283-1294.

Russell, J., Booth, A., Fuller, J., Harrower, B., Hedley, P., Machray, G., & Powell, W. (2004). A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome. *Genome*, *47* (2), 389-398.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, 74*(12), 5463-5467.

Satyavathi, C. T., Tiwari, S., Bharadwaj, C., Rao, A. R., Bhat, J., & Singh, S. P. (2013). Genetic diversity analysis in a novel set of restorer lines of pearl millet [Pennisetum

glaucum (L.) R. Br] using SSR markers. *Vegetos-An International Journal of Plant Research, 26*(1), 72-82.

Saunders, G. W., & Druehl, L. D. (1993). Revision of the kelp family Alariaceae and the taxonomic affinities of Lessoniopsis Reinke (Laminariales, Phaeophyta). *Hydrobiologia, 260* (1), 689-697.

Schlötterer, C. (2004). The evolution of molecular markers—just a matter of fashion?. *Nature Reviews Genetics, 5* (1), 63-69.

Schlötterer, C., & Pemberton, J. (1994). The use of microsatellites for genetic analysis of natural populations. In *Molecular ecology and evolution: Approaches and Applications* (203-214). Birkhäuser Basel.

Schmidhalter, U., Evéquoz, M., & Oertli, J. J. (1988). Adaptive mechanisms of root growth under drought stress. *Plant Roots and Their Environment. Abstracts Int. Soc. Root Research*, 27-28.

Scott, K. D., Eggler, P., Seaton, G., Rossetto, M., Ablett, E. M., Lee, L. S., & Henry, R. J. (2000). Analysis of SSRs derived from grape ESTs. *Theoretical and Applied Genetics, 100* (5), 723-726.

Sharma, A. (2001). Marker-assisted Improvement of Pearl Millet [*Pennisetum glaucum (L.) R. Br.*] Downy mildew resistance in elite hybrid parental line (Doctoral thesis, Chaudhary Charan Singh Haryana Agricultural University, Hisar, India).

Sharma, M. V., Kantartzi, S. K., & Stewart, J. M. (2009). Molecular diversity and polymorphism information content of selected *Gossypium hirsutum* accessions. *Summaries of Arkansas Cotton Research*, 124-126.

Small, R. L., Cronn, R. C., & Wendel, J. F. (2004). LAS Johnson Review No. 2. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany*, *17* (2), 145-170.

Smith, J. S. C., Chin, E. C. L., Shu, H., Smith, O. S., Wall, S. J., Senior, M. L., ... & Ziegle, J. (1997). An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPs and pedigree. *Theoretical and Applied Genetics*, *95* (1-2), 163-173.

Sobrino, B., Brión, M., & Carracedo, A. (2005). SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Science International*, *154* (2), 181-194.

Staub, J. E., Serquen, F. C., & Gupta, M. (1996). Genetic markers, map construction, and their application in plant breeding. *HortScience*, *31* (5), 729-741.

Steinkellner, H., Lexer, C., Turetschek, E., & Glössl, J. (1997). Conservation of (GA)n microsatellite loci between *Quercus* species. *Molecular Ecology*, *6* (12), 1189-1194.

Streiff, R., Ducousso, A., Lexer, C., Steinkellner, H., Gloessl, J., & Kremer, A. N. D. A. (1999). Pollen dispersal inferred from paternity analysis in a mixed oak stand of

Quercus robur L. and *Quercus petraea* (Matt.) Liebl. *Molecular Ecology, 8* (5), 831-841.

Sun, Y., Skinner, D. Z., Liang, G. H., & Hulbert, S. H. (1994). Phylogenetic analysis of sorghum and related taxa using internal transcribed spacers (ITS) of nuclear ribosomal DNA. *Theoretical and Applied Genetics*, *89* (1), 26-32.

Sunyaev, S., Hanke, J., Aydin, A., Wirkner, U., Zastrow, I., Reich, J., & Bork, P. (1999). Prediction of nonsynonymous single nucleotide polymorphisms in human disease-associated genes. *Journal of Molecular Medicine*. *77* (11), 754-760.

Takamatsu, S., Hirata, T., & Sato, Y. (1998). Phylogenetic analysis and predicted secondary structures of the rDNA internal transcribed spacers (ITS) of the powdery mildew fungi (Erysiphaceae). *Mycoscience*, *39* (4), 441-453.

Takundwa, M., Nepolo, E., Chimwamurombe, P.M., Cullis, C.A., Kandawa-Schultz, M., & Kunert, K. (2010). Development and use of Microsatellites markers for genetic variation analysis, in the Namibian germplasm, both within and between populations of Marama bean (*Tylosema esculentum*). *Journal of Plant Breeding and Crop Science*, *2* (8), 233-242.

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, *10* (3), 512-526.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution, 30* (12), 2725-2729.

Tao, Y., Manners, J. M., Ludlow, M. M., & Henzell, R. G. (1993). DNA polymorphisms in grain sorghum (*Sorghum bicolor (L.) Moench*). *Theoretical and Applied Genetics*, *86* (6), 679-688.

Tautz, D., Tautz, C., Webb, D., & Dover, G. A. (1987). Evolutionary divergence of promoters and spacers in the rDNA family of four Drosophila species: Implications for molecular coevolution in multigene families. *Journal of Molecular Biology*, *195* (3), 525-542.

Tobolski, J. J., & Kemery, R. D. (1992). Identification of red maple cultivars by isozyme analysis. *HortScience*, *27* (2), 169-171.

Tostain S (1994) Isoenzymatic classification of pearl millet (*Pennisetum glaucum*, Poaceae) landraces from Niger (West Africa). *Systematic Evolution. 193*, 81–93.

Tostain, S., & Marchais, L. (1989). Enzyme diversity in pearl millet (*Pennisetum glaucum*). *Theoretical and Applied Genetics*, *77* (5), 634-640.

Tostain, S., Riandey, M. F., & Marchais, L. (1987). Enzyme diversity in pearl millet (*Pennisetum glaucum*). *Theoretical and Applied Genetics*, *74* (2), 188-193.

Travis, S., Maschinski, J., & Keim, P. (1996). An analysis of genetic variation in *Astragalus cremnophylax* var. *cremnophylax*, a critically endangered plant, using AFLP markers. *Molecular Ecolology*, *5* (6), 735–745.

Vallejos, C. E. (1983). Enzyme activity staining. *Isozymes in Plant Genetics and Breeding*, *1*, 496-516.

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., Thiel, T., Stein, N., Langridge, P., & Graner, A. (2002). In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cellular and Molecular Biology Letters*, *7* (2A), 537-546.

Velegraki, A., Logotheti, M., Pyrri, I., & Kapsanaki-gotsi, E. (2001). Evaluation of two PCR-based methods in the identification of potentially infectious *Aspergillus*. *Mycoses, 44* (Supp. 1), 45.

Verma, V., Worland, A. J., Sayers E. J., Fish L., Caligari, P.D.S., & Snape J.W. (2005). Identification and characterization of quantitative trait loci related to lodging resistance and associated traits in bread wheat. *Plant Breeding*, *124*, 234-241.

Vierling, R. A., Xiang, Z., Joshi, C. P., Gilbert, M. L., & Nguyen, H. T. (1994). Genetic diversity among elite sorghum lines revealed by restriction fragment length

polymorphisms and random amplified polymorphic DNAs. *Theoretical and Applied Genetics*, *87* (7), 816-820.

Vom Brocke, K., Weltzien, E., Christinck, A., Presterl, T., & Geiger, H. H. (2003). Effects of farmers' seed management on performance and adaptation of pearl millet in Rajasthan, India. *Euphytica*, *130* (2), 267-280.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., ... & Zabeau, M. (1995). AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Research*, *23* (21), 4407-4414.

Wendel, J., Schinabel, A., & Seelanan, T. (1995). Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (Gossypium). *Proceedings of the National Academy of Science- USA*, *92*, 280–284.

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

Winfield, M. O., Arnold, G. M., Cooper, F., Le Ray, M., White, J., Karp, A., & Edwards, K. J. (1998). A study of genetic diversity in *Populus nigra* subsp. *betulifolia* in the Upper Severn area of the UK using AFLP markers. *Molecular Ecology*, *7* (1), 3-10.

Wolfe, K. H., Li, W. H., & Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences*, *84* (24), 9054-9058.

Yadav, S., Jain, S., Jain, V., & Jain, R. K. (2007). Genetic analysis of CMS, restorer, hybrid and open-pollinated genotypes of Indian pearl millet [*Pennisetum glaucum (L.) R. Br.*(L.) R. Br.] using ISSR markers. *Indian Journal of Biotechnology*, *6*, 349-353.

Yamane, T. (1967). Statistics: *An Introductory Analysis*, 2nd Edition, New York: Harper and Row.

Yang, W., De Oliveira, A. C., Godwin, I., Schertz, K., & Bennetzen, J. L. (1996). Comparison of DNA marker technologies in characterizing plant genome diversity: variability in Chinese sorghums. *Crop Science*, *36* (6), 1669-1676.

Zhang, W., Qu, L., Gu, H., Gao, W., Liu, M., Chen, J., & Chen, Z. (2002). Studies on the origin and evolution of tetraploid wheats based on the internal transcribed spacer (ITS) sequences of nuclear ribosomal DNA. *Theoretical and Applied Genetics*, *104* (6-7), 1099-1106.

APPENDICES

APPENDIX A: Main ingredients for the CTAB protocol

Main ingredients for the CTAB protocol

CTAB: for 1L of CTAB buffer

100ml of 1 M Tris-HCl, pH 8.0

280 ml of 5 M NaC1

40 ml of 0.5 M EDTA

20 g of CTAB (Cetyl-trimethy1 ammonium bromide) bring to 1L with H_2O

1M Tris, pH 8.0: for 1 L

121.1 g Tris

700 ml ddH_2O

Preparation of CTAB Buffer

Dissolve Tris and bring to 900 ml.

PH to 8.0 with concentrated HC1 (will need – 50ml)

Bring to 1 L

0.5 M EDTA pH 8.0: for 1 L

186.12 g of EDTA

750 ml ddH_2O

Add about 20 g of NaOH pellets

Slowly add more NaOH until pH 8.0; EDTA will not dissolve until the pH is near 8.0.

5 M NaC1: for 1 L

292.2 g of NaC1

700 ml ddH_2O

Dissolve and bring to 1 L

7.5 M ammonium acetate: for 250 ml

144.5 g ammonium acetate

Bring to volume with ddH_2O

APPENDIX B: List of primers used in the present study

Table11: Details of SSR primers developed from pearl millet EST sequences that were used in present study.

Name[1]	Forward primer sequence (5' – 3')	Reverse primer sequence (5' – 3')	Repeat	P[2]	T[3] (°C)	S[4] (bp)
3002	AAAGTTACCGGGAGGGTAAAAA	TCGCCTAAAAACTGGAGGAA	AAAC(3)	P	61	205
3005	CGCGGTGTTCTCACACAC	TGTGAATTCCGCGGGTATAG	AC(14)	M	48	140
3006	AAATCGGTCGTGGTGAAGTT	GAGAATGTGGGAGACACACG	AC(16)	M	52	180
3009	CTGTACCATGTGCGCTGATT	GCGCATATATGTGGGTGTGT	AC(16)	P	48	320
3011	CACGCCCTTTTTACCTTGAC	CGCGACACGTCCTACACTAA	AC(21)	P	48	150
3013	TGTGGGAGAGAGGAGAGTCC	CGCGAGATGATGTGTGGT	AC(33)	P	61	370
3014	TGCTTCACAGCCTCTCCATA	CCACCATGCAACAGCAATAA	ACC(8)	M	55	280
3016	TTGTGGCTGAAGAAGAGATCC	AATGTGGGGAGAGACACACG	CA(17)	P	45	450
3017	CACCAAACAGCATCAAGCAG	AGGTAGCCGAGGAAGGTGAG	CAG(7)	P	52	200
3018	CGATGACACCTGTGCGTATT	ATCGAACTGCACGTTAGCAA	CATG(4)	M	61	215
3019	GCGCACCACCTGTGTCTAT	CATGCAGAGAAAAATCAAGCA	CGTA(4)	M	48	210
3020	GTTCCATGGAGCTGGAAGTC	GCTAGAACAGGGCCGTTACA	CGTG(5)	M	58	180
3021	GCCGACAGGAAGATTACGAT	AGCAAAACGCAGAACAACAG	CGTG(5)	M	58	175
3022	CTGGAAGTCCTTCTCGGTT	CTGCTCCGCTCTGAATCT	CGTG(	M	58	19

	G	G	5)			0
3025	GTTGCAGATGAGCGATCGTA	AGCGCAAAGAGTGTAACTTGG	CTC(6)	M	58	180
3026	GTGAGGCCTCGAACAAACAC	GCCGACCAAGAACTTCATACA	CTC(6)	M	58	130
3027	ACACCATCACCGACAACAAA	AGTGACCTGGGGTACAGACG	GAT(6)	P	61	210
3028	ACGATTCTTCGTCGTTCCAG	ATACGATACGCGCGAGCTAC	GATC(4)	M	58	170
3029	ACCAGCAACAGCAGCAGAG	ACACACTGCGACAAGTGGAG	GCA(6)	P	48	260
3032	AGGTAGCCGAGGAAGGTGAG	CAACAGCATCAAGCAGGAGA	GCT(8)	P	61	190
3033	GAGGGCCAGCTCTCCTAGAT	CCCTAACCACAGAGGGACAC	TGCC(4)	P	58	220
3035	GCCAAGGAGGTCAAGATCG	ACACGACTCGACTCAGACCA	TGCC(4)	P	58	280
3037	CGTCGCTGCTCTTTCTTCTT	ATTTCAGAAACGGCAACCAA	TGGA(4)	M	58	200
3038	CTCTCGGTTTGACGGTTTGT	GGGGAAAACAAAGTTGCTCA	TGT(6)	P	58	180
3039	GGCACGAGGGGCTAAGTAA	GGAACGCCGAGTACACAGAT	TGT(6)	P	61	180

KEY: [1] Each SSR primer number in this column is preceded by the prefix 'ICMP' for ICRISAT millet primer. [2] Polymorphism: P = ICMB 841 and 863B are polymorphic; M = ICMB 841 and 863B are monomorphic [3] Annealing temperature [4] Expected PCR product size

Table 12: Details of RAPD primer sequences that were used in present study**.**

Name	**Primer sequence (5' to 3')**	**Annealing Temperature (^{O}C)**
OPL-11	ACGATGAGCC	39.5
OPAL-20	AGGAGTCGGA	39.5
OPAL-08	GTCGCCCTCA	43.6
OPAL-15	AGGGGACACC	43.6
OPM-16	GTAACCAGCC	39.5
OPN-16	AAGCGACCTG	39.5
OPF-10	GGAAGCTTGG	39.5

OPF-01	ACGGATCCTG	39.5
OPE-04	GTGACATGCC	39.5
OPE-09	CTTCACCCGA	39.5
OPE-18	GGACTGCAGA	39.5
OPAG-13	GGCTTGGCGA	43.6

APPENDIX C: Morphological characterization data for 29 *P. glaucum* accessions

Table 13: Morphological characteristics for 29 accessions of *P. glaucum.*

A.N	E.V	T.A	N N T	P.A	L.S	F.P	S.S	Sp.S	B.L	S.D
Nam 58	3	3	1	7	3	7	5	3	3	7
Nam 73	3	3	3	3	3	3	5	2	3	5
Nam 74	5	5	3	5	3	5	5	3	3	7
Nam 85	5	3	3	3	3	5	2	3	7	7
Nam 103	5	5	3	5	3	5	2	3	3	3
Nam 112	5	3	3	5	3	5	2	3	3	7
Nam 117	5	5	3	5	5	5	5	3	3	7
Nam 120	5	5	3	5	3	5	2	3	3	7
Nam 212	5	3	0	7	3	7	5	3	3	7
Nam 330	5	3	0	3	0	3	2	3	3	7
Nam 349	1	3	3	3	3	6	2	3	3	7
Nam 388	3	3	3	5	5	6	5	3	3	7
Nam 399	7	3	0	7	3	7	7	3	3	5
Nam 414	7	3	7	3	7	5	5	3	3	3
Nam 416	7	3	0	6	4	6	1	3	3	7
Nam 417	7	3	3	5	3	6	2	3	3	7
Nam 431	7	3	0	5	3	6	2	3	3	7
Nam 449	5	3	3	5	7	6	2	3	3	7
Nam 463	3	5	5	5	7	6	5	4	3	7
Nam 468	5	3	3	5	7	6	2	3	3	7
Nam 981	7	3	3	5	3	5	1	3	5	3
Nam 1012	3	3	1	5	3	5	2	3	3	7
Nam 1163	7	3	0	3	2	5	5	3	3	5
Nam 1837	3	3	5	7	3	5	1	3	3	3
Nam 2580	3	3	10	7	3	7	7	3	3	7
Nam 2586	5	3	0	7	3	7	2	3	3	5
Nam 2650	5	3	0	7	3	7	2	3	5	5
Nam 2668	3	3	3	7	7	7	2	3	3	7
Nam 2682	5	3	3	5	3	7	7	3	3	7

KEY:

A.N = Accession number, E.V = Early Vigour,T.A = Tillering Altitude, NNT = number of NodalTillers, P.A = Plant Aspects, L.S = Lodging Susceptility, F.P = Fodder Potential, S.S = Spike Shape, SP.S = Spikelet Shattering, B.L = Bristle Length, S.D = Spike Density.

Morphological classification

Early vigour

Recorded after thinning to avoid the effect of the plant number. Scored 18 days after emergence.

Tillering altitude (measured at head emergence)

3 = Erect

5 = Intermediate

7= Spreading

Number nodal tillers (measured at dough stage)

3 = Few

5 = Intermediate

7 = Many

Plant aspect (Overall agronomic desirability of accession at dough stage)

3 = Poor

5 = Intermediate

7 = Good

Lodging susceptibility (measured at dough stage)

3 = Low

5 = Intermediate

7 = High

Fodder potential (considers tillering, leafiness and bulk at flowering)

3 = Poor

5 = Intermediate

7 = Good

Spike shape (observed at dough stage)

1 = Cylindrical

2 = Conical

3 = spindle

4 = Club

5 = Candle

6 = Dumb-bell

7 = Lanceolate

8 = Oblanceolate

9 = Globose

10 = Other

Spikelet shattering (measured at maturity)

1 = Spontaneous shattering

2 = Shattering at touch

3 = Non-Shattering and free threshing

4 = Non-Shattering and difficult to thresh

Bristle length (measured at dough stage)

3 = Short (bristles below the level of the apex of the seed)

5 = Medium (Bristle Length from 0 and 2 cm above the seed

7 = Long (Bristles longer than 2 cm above the seed)

Spike density (measured at maturity)

3 = Loose

5 = intermediate

7 = Compact

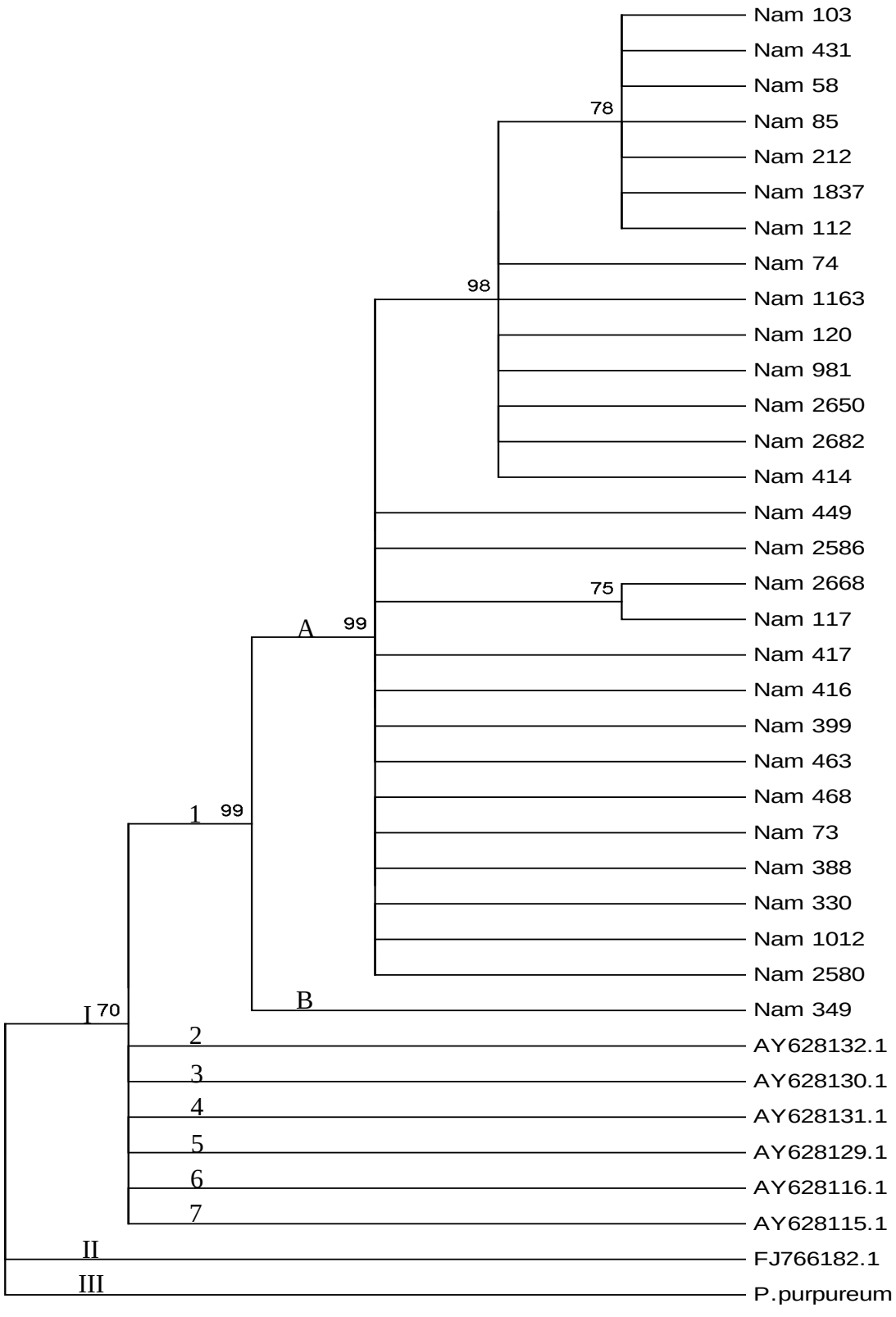


Figure 28: The phylogenetic tree for 29 plus seven NCBI retrieved *P. glaucum* genotypes based on ITS sequences inferred using the Maximum Likelihood method. Branches corresponding to partitions reproduced in less than 70% bootstrap replicates are collapsed. The bootstrap values show the confidence in the groupings as a percentage. *P. purpureum* was used as the out-group to root the tree generated with MEGA 6.0 software.

APPENDIX D: *P. GLAUCUM* ITS sequence alignments

```
964       ------------------------------------------------------------
868       -CAGACTGGACCCCAGCGCCAGGCTCGCTGTTTCGCGATGCCGCAGTACCCAGTAAATCG
1837      -------------------------GGTCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
414       --------------------------TCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
212       ---------------------------CGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
58        --------------------------TCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
2868      --------------------------TCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
1802      --------------------------TCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
431       --------------------------TCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
2682      -------------------------GTCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
981       ------------------------GGTCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
2863      ----------------------------------CCGAGCGA-CAAAGCGCTACGGTCG
85        ----------------------------------CCGAGCGA-CAAAGCGCTACGGTCG
890       ----------------------------------CCGAGCGA-CAAAGCGCTACGGTCG
2780      --------------------------TCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
2829      ------------------------GGGTCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
103       ---------------------------CGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
472       -------------------------GTCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
1163      -------------------------GTCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
487       ---------------------------CGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
2810      --------------------------TCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
582       -------------------------GGTCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
517       --------------------------TCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
112       ------------------------GGGTCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
129       -----------------------GGGGTCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
542       -------------------------GTCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
2497      --------------------------TCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
120       ------------------------GGGTCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
191       ---------------------------TCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
2650      -------------------------GTCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
2662      -------------------------GTCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
74        --------------------------TCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
1854      ---------------------------CGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
2819      --------------------------TCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
449       ----------------------------------GTCCGAGCGA-CAAAGCGCTACGGTCG
138       ----------------------GCAGTCGCGGTCCGAGCGACA-G-AAGCGCTACGGTCG
621       ----------------------CAGACTGGATCCGAGCGACA-G-AAGCGCTACGGTCG
749       ----------------------GGTCGCGGTCCGAGCGACA-G-AAGCGCTACGGTCG
654       ----------------------GGTCGCGGTCCGAGCGACA-G-AAGCGCTACGGTCG
416       --------------------GGCAGACGCGGTCCGAG-CGAC-AGAAGCGCTACGGTCG
463       -------------------CTGGGGTCGCGGTCCAAGC-GAC-AGAAGCGCTACGGTCG
1172      ----------------------GGGTCGCGGTCCGAGC-GAC-AGAAGCGCTACGGTCG
496       ----------------------GGGGTCGCGGTCCGAGCGA-CAAAGCGCTACGGTCG
2586      --------------------------ACTGGGTCCGAGCGAC-AGAAGCGCTACGGTCG
2688      ------------------------GTCGCGGTCC-GAGCGAC-AGAAGCGCTACGGTCG
1094      --------------------TGGGGTCGCGGTCC-GAGCGAC-A-AAGCGCTACGGTCG
399       -------------------CTGGGGTCGCGGTCCGAGCGACA-G-AAGCGCTACGGTCG
1041      --------------------TGGGGTCGCGGTCCGAGCGACA-G-AAGCGCTACGGTCG
117       ----------------------GCAGTCGCGGTCCGAGCGACA-AAGCGCTACGGTCGTT
1097      ----------------------CAGTCGCGGTCCGAGCGACA-GAAGCGCT-ACGGTCG
1810      ----------------------GGCAGTCGCGGTCCGAGCGACA-GAAGCGCT-ACGGTCG
```

```
1056      ----------------------GGGTCGCGGTCCGAGCGACA-GAAGCGCT-ACGGTCG
417       -----------------------GACGCGGTCCGAGCGACA-GAAGCGCT-ACGGTCG
2843      ----------------------GGTCGCGGTCCGAGCGACA-GAAGCGCT-ACGGTCG
2695      -----------------------GACTGGACCCGAGCGACA-GAAGCGCT-ACGGTCG
2793      ----------------------GGTCGCGGTCCGAGCGACA-GAAGCGCT-ACGGTCG
490       --------------------GGCAGTCGCGGTCCGAGCGACA-GAAGCGCT-ACGGTCG
512       ----------------------CAGTCGCGGTCCGAGCGACA-GAAGCGCT-ACGGTCG
2267      ----------------------AGACGCGGTCCGAGCGACA-GAAGCGCT-ACGGTCG
2668      ---------------------TCGCGGTCCGAGCGACA-GAAGCGCT-ACGGTCG
863       ----------------------GGGTCGCGGTCCGAGCGACA-GAAGCGCT-ACGGTCG
577       ----------------------AGTCGCGGTCCGAGCGACA-GAAGCGCT-ACGGTCG
1772      ----------------------AGTCGCGGTCCGAGCGACA-GAAGCGCT-ACGGTCG
988       ----------------------AGAGCGGGTCCGAGTCGACAAAGCTGCTACGGTCGTC
695       GCAGACTGGACCCCAGCGCCAGGGTCGCGGTCCG-AGCGATCAAAGCGCTACCCGGTACG
349       ---------------------TCGCGGTCCGA--GCGACAAAGC-GCTACGGTCG
892       -CAAACTGGACCCCAGCGCCAGGGTCGCTGGTTTCCGATGCGACAAAGCGCTACGGTACG
1829      -----------------ACCTGGGGTCGCGGTCCGAGCGACAAAGCGCTACCCGGTACG
388       -------------------------------GTCCGA--GCGACAAAGCGCT-ACGGTCG
2722      ---------------------TCGCGGTCCGA--GCGACAAAGCGCT-ACGGTCG
772       ----------------------GGGGTCGCGGTCCG---AGCGACAAAAGCT-ACGGTCG
468       ---------------------TCGCTGGCCGA--GCGACACAGCGCT-ACGGTCG
204       ----------------------CAGACGCGGTCCGA--GCGACAGAGCGCT-ACGGTCG
2875      ---------------------ACTGGGTCCGA--GCGACAGAGCGCT-ACGGTCG
741       ---------------------CGCGGTCCGA--GCGACAAAGCGCT-ACGGTCG
1169      ---------------------TCGCGGTCCGA--GCGACAAAGCGCT-ACGGTCG
1012      ---------------------GTCGCGGTCCGA--GCGACAAAGCGCT-ACGGTCG
330       ----------------------GGGGTCGAGGACCGA--GCGACAAAGCGCT-ACGGTCG
2828      ----------------------GGTCGCGGTCCGA--GCGACAAAGCGCT-ACGGTCG
2795      ----------------------GGGTCGCGGTCCGA--GCGACAAAGCGCT-ACGGTCG
2580      --------------------TGGGGTCGCGGTCCGA--GCGACAAAGCGCT-ACGGTCG
1038      ----------------------GGGTCGCGGACCGA--GCGACAAAGCGCT-ACGGTCG
679       ----------------------GGGTCGCGGTCCGA--GCGACAAAGCGCT-ACGGTCG
610       --------------------TGGGGTCGCGGACCGA--GCGACAAAGCGCT-ACGGTCG
266       ----------------------GGGGTCGCGGTCCGA--GCGACAAAGCGCT-ACGGTCG
659       ----------------------GGGGTCGCGGTCCGA--GCGACAAAGCGCT-ACGGTCG
2771      ----------------------GGGGTCGCGGTCCGA--GCGACAAAGCGCT-ACGGTCG
73        ----------------------GGGGTCGCGGACCGA--GCGACAAAGCGCT-ACGGTCG
98        ----------------------GGGGTCGCGGTCCGA--GCGACAAAGCGCT-ACGGTCG


964       --------------------CCAGGGCAGACTGGTTCCGAGCGAC-GAGCTCTACGGTCT
868       GGTTGCTCGGCTGAGGATCGGCTCTGTGACGTGGGTCCGAGCGACAAAGCGCTACGGTCG
1837      TCATGGGTCATTTAGGGCC----AACTA-------GCCC----------------------
414       TCATGGGTCATTTAGGCCC----AACGA-------GACC----------------------
212       TCATGGGTCATTTAGGGCC----AACGA-------GACC----------------------
58        TCATGGGTCATTTAGGGCA--------A-------GAGC----------------------
2868      TCATGGGTCATTTAGGGCC----AACGA-------GACC----------------------
1802      TCATGGGTCATTTAGGCCC----AACTA-------GACC----------------------
431       TCATGGGTCATTTAGGGCC----AACGA-------GACC----------------------
2682      TCATGGGTCATTTAGGCCC----AACTA-------GACC----------------------
981       TCATGGGTCATTTAGGCCC-------------AACGAGC----------------------
2863      TCATGGGTCATTTAGGGCC-------------AACGAGC----------------------
85        TCATGGGTCATTTAGGGCC-------------AACGAGC----------------------
890       TCATGGGTCATTTAGGGCC-------------AACGAGC----------------------
```

```
2780      TCATGGGTCATTTGAGTCG----GCCAACA-GAGCCGAC--------------------
2829      TCATGGGTCATTTAGG--------GCCAAC-GAGCCGAC--------------------
103       TCATGGGTCATTTAGG--------GCCAAC-GAGCCGAC--------------------
472       TCATGGGTCATTTAGG--------GCCAAC-GAGCCGAC--------------------
1163      TCATGGGTCATTTAGG--------GCCACC-GAGCCGAC--------------------
487       TCATGGGTCATTTAGG--------CCCAAC-GAGCCGAC--------------------
2810      TCATGGGTCATTTAGG--------CCCAAC-GAGCCGAC--------------------
582       TCATGGGTCATTTAGG--------CCCAAC-TAT-C-----------------------
517       TCATGGGTCATTTAGG--------CCCAAC-GAG-C-----------------------
112       TCATGGGTCATTTAGG--------GCCAAC-GAG-C-----------------------
129       TCATGGGTCATTTAGG--------CCCAAC-GAGCT-----------------------
542       TCATGGGTCATTTAGG--------CCCAAC-GAGCT-----------------------
2497      TCATGGGTCATTTAGG--------CCCAAC-TA-GC-----------------------
120       TCATGGGTCATTTAGG--------CCCAAC-TA-GC-----------------------
191       TCATGGGTCATTTAGG--------GCCAAC-TA-GC-----------------------
2650      TCATGGGTCATTTAGG--------CCCAAC-TA-GC-----------------------
2662      TCATGGGTCATTTAGG--------CCCAAC-TA-GC-----------------------
74        TCATGGGTCATTTAGG--------GCCAAC-GA-GC-----------------------
1854      TCATGGGTCATTTAGG--------GCCAAC-GA-GC-----------------------
2819      TCATGGGTCATTTAGG--------GCCAAC-GA-GC-----------------------
449       TCATGGGTCATTTAGGCCC----AACTAGC-CGACTCCT--------------------
138       TCATGGGTCATTTAGTGCC----CAGACGA-GCCGGACT--------------------
621       TCATGGGTCATTTAGTGGC----CAATCGA-GCCGGACT--------------------
749       TCATGGGTCATTTAGTGCC----CAACTAA-GCCGGACT--------------------
654       TCATGCGG-TCATTTAGGC----CAACTAA-TCCGGACT--------------------
416       TCATGGGTCATTTAGTGGC----CAATCGA-TGCCGACT--------------------
463       TCATGGGTCATTTAGTGGC----CAATCGA-TGCCGACT--------------------
1172      TCATGGGTCATTTAGGGCC----AACTAAA-TGCCGACT--------------------
496       TCATGGGTCATTTAGTGCC----CAACTAA-GCCGGACT--------------------
2586      TCATGGGTCATTTAGGGCC----AACTAAA-TCCAGACT--------------------
2688      TCATGGGTCATTTAGGGCC----AACTAAA-TGCCGACT--------------------
1094      TCATGGGTCATTTAGTGGC----CAATCGA-TGCCGACT--------------------
399       TCATGGGTCATTTAGTGGC----CAATCGA-TGCCGACT--------------------
1041      TCATGGGTCATTTAGTGGC----CAATCGA-TGCCGACT--------------------
117       CATGAGGTCATTTAGGCCC----CACTAAA-TGCCGACT--------------------
1097      TCATGGGTCATTTAGGCCC----CACTAAA-GCCGGACT--------------------
1810      TCATGGGTCATTTAGGGCC----CACTAAA-TGCCGACT--------------------
1056      TCATGGGTCATTTAGGCCC----CACTAAA-GCCGGACT--------------------
417       TCATGGGTCATTTAGTGCC----CAACTAA-TGCCGACT--------------------
2843      TCATGGGTCATTTAGTGCC----CAACTAA-TCCCGACT--------------------
2695      TCATGGGTCATTTAGGGCC----AACTAAA-TGCCGACT--------------------
2793      TCATGGGTCATTTAGTGGC----CACCGAA-GCCCGACT--------------------
490       TCATGGGTCATTTAGGCCC----CACTAAA-GCCGGACT--------------------
512       TCATGGGTCATTTAGTGGC----CAATCGA-TGCCGACT--------------------
2267      TCATGGGTCATTTAGGGCC----AACTAAA-TGCCGACT--------------------
2668      TCATGGGTCATTTAGGCCC----CACTAAA-TGCCGACT--------------------
863       TCATGGGTCATTTAGGGCC----AACTAAA-TGCCGACT--------------------
577       TCATGGGTCATTTAGGGCC----AACTAAA-TGCCGACT--------------------
1772      TCATGGGTCATTTAGGGCC----CACTAAA-TGCCGACT--------------------
988       TGATGGGTCATTTAGGCCC-----CCCTAGCCGACTCCT--------------------
695       TCATGGGTCATTTAGGGCC-----AACGAACCGACTCCT--------------------
349       TCATGGGTCATTTAGGGCC-----AACGAGCCGACTCCT--------------------
892       TCATGGGTCATTTAGGGCC-----AACGAGCCGACTCCT--------------------
1829      TCATGGGTCATTTAGGGCC-----AACGAGCCGACTCCT--------------------
```

```
388       TCATGGGTCATTTAGGCCC----A-CCGAGCCGACTCCT--------------------
2722      TCATGGGTCATTTAGGGCC----A-CCGAGCCGACTCCT--------------------
772       TCATGGGTCATTTAGGGCA------ACAAGCCAACTCCT--------------------
468       TCATGGGTCATTTAGCGCC----A-ACGAGCCGACTCCT--------------------
204       TCATGGGTCATTTAGGGCC----AA--AATCCGACTCCT--------------------
2875      TCATGGGTCATTTAGGGCC----AACTTAGCCGACTCCT--------------------
741       TCATGGGTCATTTAGGGCC----A-ACGAGCCGACTCCT--------------------
1169      TCATGGGTCATTTAGGCCC----A-ACGAGCCGACTCCT--------------------
1012      TCATGGGTCATTTAGGGCC----C-ACGAGCCGACTCCT--------------------
330       TCATGGGTCATTTAGGGCC----A-ACGAGCCGACTCCT--------------------
2828      TCATGGGTCATTTAGGGCC----A-ACGAGCCGACTCCT--------------------
2795      TCATGGGTCATTTAGGGCC----A-ACGAGCCGACTCCT--------------------
2580      TCATGGGTCATTTAGGGCC----A-ACGAGCCGACTCCT--------------------
1038      TCATGGGTCATTTAGGCCC----A-ACGAGCCGACTCCT--------------------
679       TCATGGGTCATTTAGGCCC----A-ACGAGCCGACTCCT--------------------
610       TCATGGGTCATTTAGGCCC----A-ACGAGCCGACTCCT--------------------
266       TCATGGGTCATTTAGGGCC----A-ACGAGCCGACTCCT--------------------
659       TCATGGGTCATTTAGGGCC----A-ACGAGCCGACTCCT--------------------
2771      TCATGGGTCATTTAGGGCC----A-ACGAGCCGACTCCT--------------------
73        TCATGGGTCATTTAGGCCC----A-ACGAGCCGACTCCT--------------------
98        TCATGGGTCATTTAGGCCC----A-ACGAGCCGACTCCT--------------------


964       CTCGATGCG--AGCAATTACCAGGCCATC----GAGCCTACTCCTCTGACGATACACTAC
868       TGCATGCCACGGTCATTCTATGGCCCAACGAGACCGACTCCTAGCCG--AGACACTGGAG
1837      ----------------------------------GGACTCCTAGCGCGAGGCACACTGCA
414       ----------------------------------GGACTCCTAGCGCGAGGCACACTGCA
212       ----------------------------------GGACTCCTAGCGCGAGGCACACTGCA
58        ----------------------------------CGACTCCTAGCGCGAGGCACACTGCA
2868      ----------------------------------GGACTCCTAGCGCGAGGCACACTGCA
1802      ----------------------------------GGACTCCTAGCGCGAGGCACACTGCA
431       ----------------------------------GGACTCCTAGCGCGAGGCACACTGCA
2682      ----------------------------------GGACTCCTAGCGCGAGGCACACTGCA
981       ----------------------------------CGACTCCTAGCGCGAGGCACACTGCA
2863      ----------------------------------CGACTCCTAGCGCGAGGCACACTGCA
85        ----------------------------------CGACTCCTAGCGCGAGGCACACTGCA
890       ----------------------------------CGACTCCTAGCGCGAGGCACACTGCA
2780      ----------------------------------TGGACCCTAGCGCGAGGCACACTGCA
2829      ----------------------------------TGGACCCTAGCGCGAGGCACACTGCA
103       ----------------------------------TGGACCCTAGCGCGAGGCACACTGCA
472       ----------------------------------TGGACCCTAGCGCGAGGCACACTGCA
1163      ----------------------------------TGGACCCTAGCGCGAGGCACACTGCA
487       ----------------------------------TGGACCCTAGCGCGAGGCACACTGCA
2810      ----------------------------------TGGACCCTAGCGCGAGGCACACTGCA
582       ----------------------------------CGACTCCTAGCGCGAGGCACACTGCA
517       ----------------------------------CGACTCCTAGCGCGAGGCACACTGCA
112       ----------------------------------CGACTCCTAGCGCGAGGCACACTGCA
129       ----------------------------------CGACTCCTAGCGCGAGGCACACTGCA
542       ----------------------------------CGACTCCTAGCGCGAGGCACACTGCA
2497      ----------------------------------CGACTCCTAGCGCGAGGCACACTGCA
120       ----------------------------------CGACTCCTAGCGCGAGGCACACTGCA
191       ----------------------------------CGACTCCTAGCGCGAGGCACACTGCA
2650      ----------------------------------CGACTCCTAGCGCGAGGCACACTGCA
2662      ----------------------------------CGACTCCTAGCGCGAGGCACACTGCA
```

```
74      ---------------------------------CGACTCCTAGCGCGAGGCACACTGCA
1854    ---------------------------------CGACTCCTAGCGCGAGGCACACTGCA
2819    ---------------------------------CGACTCCTAGCGCGAGGCACACTGCA
449     ---------------------------------AG--------CGCGAGGCACACTGCA
138     ---------------------------------CC--------TCAGCCGAGACACTGC
621     ---------------------------------CC--------TCAGCCGAGACACTGC
749     ---------------------------------CC--------TCAGCCGAGACACTGC
654     ---------------------------------CC--------TCAGCCGAGACACTGC
416     ---------------------------------CC--------TCAGCCGAGACACTGC
463     ---------------------------------CC--------TCAGCCGAGACACTGC
1172    ---------------------------------CC--------TCAGCCGAGACACTGC
496     ---------------------------------CC--------TCAGCCGAGACACTGC
2586    ---------------------------------CC--------TCAGCCGAGACACTGC
2688    ---------------------------------CC--------TCAGCCGAGACACTGC
1094    ---------------------------------CC--------TCAGCCGAGACACTGC
399     ---------------------------------CC--------TCAGCCGAGACACTGC
1041    ---------------------------------CC--------TCAGCCGAGACACTGC
117     ---------------------------------CC--------TCAGCCGAGACACTGC
1097    ---------------------------------CC--------TCAGCCGAGACACTGC
1810    ---------------------------------CC--------TCAGCCGAGACACTGC
1056    ---------------------------------CC--------TCAGCCGAGACACTGC
417     ---------------------------------CC--------TCAGCCGAGACACTGC
2843    ---------------------------------CC--------TCAGCCGAGACACTGC
2695    ---------------------------------CC--------TCAGCCGAGACACTGC
2793    ---------------------------------CC--------TCAGCCGAGACACTGC
490     ---------------------------------CC--------TCAGCCGAGACACTGC
512     ---------------------------------CC--------TCAGCCGAGACACTGC
2267    ---------------------------------CC--------TCAGCCGAGACACTGC
2668    ---------------------------------CC--------TCAGCCGAGACACTGC
863     ---------------------------------CC--------TCAGCCGAGACACTGC
577     ---------------------------------CC--------TCAGCCGAGACACTGC
1772    ---------------------------------CC--------TCAGCCGAGACACTGC
988     ---------------------------------AG--------ACC--GAAGACACTGC
695     ---------------------------------AG--------CCG----AGACACTGC
349     ---------------------------------AG--------CCG----AGACACTGC
892     ---------------------------------AG--------CCG----AGACACTGC
1829    ---------------------------------AG--------CCG----AGACACTGC
388     ---------------------------------AG--------CCG----AGACACTGC
2722    ---------------------------------AG--------CCG----AGACACTGC
772     ---------------------------------AG--------CCG----AGACACTGC
468     ---------------------------------GG--------CCG----AGACACTGC
204     ---------------------------------TG--------CCG----AGACACTGC
2875    ---------------------------------GG--------CCG----AGACACTGC
741     ---------------------------------AG--------CCG----AGACACTGC
1169    ---------------------------------AG--------CCG----AGACACTGC
1012    ---------------------------------AG--------CCG----AGACACTGC
330     ---------------------------------AG--------CCG----AGACACTGC
2828    ---------------------------------AG--------CCG----AGACACTGC
2795    ---------------------------------AG--------CCG----AGACACTGC
2580    ---------------------------------AG--------CCG----AGACACTGC
1038    ---------------------------------AG--------CCG----AGACACTGC
679     ---------------------------------AG--------CCG----AGACACTGC
610     ---------------------------------AG--------CCG----AGACACTGC
266     ---------------------------------AG--------CCG----AGACACTGC
```

```
659       ----------------------------------AG--------CCG----AGACACTGC
2771      ----------------------------------AG--------CCG----AGACACTGC
73        ----------------------------------AG--------CCG----AGACACTGC
98        ----------------------------------AG--------CCG----AGACACTGC


964       AC-CCAAAACAACTGTATGTCACCCTCCACAAC--CCGAGC-TCACTCACTTACCCCAA-
868       CACCGAGAACAACTGTATGTCGCCCACCACGTG--CGGTGCCTCGGCAAGTTACGCCGG-
1837      CCGAGATGACAACTGTATGTCGCCCACCACGTCGCGGTTGC-TCGGCAAGTTACGACCGG
414       CCGAGATGACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
212       CCGAGATGACAACTGTATGTCGCCCACCACGTCGCGGTTGC-TCGGCAAGTTACGACCGG
58        CCGAGATGACAACTGTATGTCGCCCACCACGTCGCGGTTGC-TCGGCAAGTTACGACCGG
2868      CCGAGATGACAACTGTATGTCGCCCACCACGTCGC-GGTGC-TCGGCAAGTTACGACCGG
1802      CCGAGATGACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
431       CCGAGATGACAACTGTATGTCGCCCACCACGTCGCGGTTGC-TCGGCAAGTTACGACCGG
2682      CCGAGATGACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
981       CCGAGATGACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
2863      CCGAGATGACAACTGTATGTCGCCCACCACGTCGCGGTTGC-TCGGCAAGTTACGACCGG
85        CCGAGATGACAACTGTATGTCGCCCACCACGTCGCGGTTGC-TCGGCAAGTTACGACCGG
890       CCGAGATGACAACTGTATGTCGCCCACCACGTCGCGGTTGC-TCGGCAAGTTACGACCGG
2780      CCGAGATGACAACTGTATGTCGCCCACCACGTCGCGGTTGC-TCGGCAAGTTACGACCGG
2829      CCGAGATGACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
103       CCGAGATGACAACTGTATGTCGCCCACCACGTCGCGGTTGC-TCGGCAAGTTACGACCGG
472       CCGAGATGACAACTGTATGTCGCCCACCACGTCGCGGTTGC-TCGGCAAGTTACGACCGG
1163      CCGAGATGACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
487       CCGAGATGACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
2810      CCGAGATGACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
582       CCGAGATGACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
517       CCGAGATGACAACTGTATGTCGCCCACCACGTCGCGGTTGC-TCGGCAAGTTACGACCGG
112       CCGAGATGACAACTGTATGTCGCCCACCACGTCGCGGTTGC-TCGGCAAGTTACGACCGG
129       CCGAGATGACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
542       CCGAGATGACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
2497      CCGAGATGACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
120       CCGAGATGACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
191       CCGAGATGACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
2650      CCGAGATGACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
2662      CCGAGATGACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
74        CCGAGATGACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
1854      CCGAGATAACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
2819      CCGAGATGACAACTGTATGTCGCCCACCACGTGCGGGTTGC-TCGGCAAGTTACGACCGG
449       CCGAGATGACAACTGTATGTCGCCCACCACGTC-GCGGTGC-TCGGCAAGTTACGACCGG
138       ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
621       ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
749       ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
654       ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
416       ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
463       ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
1172      ACCGAGACACAACTGTATGTCGCCCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
496       ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
2586      ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
2688      ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
1094      ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
399       ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
1041      ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
```

```
117       ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
1097      ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
1810      ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
1056      ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
417       ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
2843      ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
2695      ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
2793      ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
490       ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
512       ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
2267      ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
2668      ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
863       ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
577       ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
1772      ACCGAGACACAACTGTATGTCGACCCACCACGT-GCGGTGC-TCGGCAAGTTACGCCGGC
988       ACCGAGAACCAACGGTGCCTATGTCGTCCAGGA--ACACTC-ACAGTGCGGTG-CTCGGC
695       ACCGACAAACAACTGTATGTCGCCCACCACGTG--CGGTGC-AATGCGGCCGTCGAAGTT
349       ACCGAGAA-CAACTGTATGTCGCCCACCACGTG--CGGTGC-TCGGCAAGTTACGCCGGC
892       ACCGAGATACAACTGTATGTCGCCCACCACGTG--CGGTGC-AATGCGGCCGTCGAAGTT
1829      ACCGAGATACAACTGTATGTCGCCCACCACGTG--CGGTGC-AATGCGGCC---GAAGTT
388       ACCGAGAA-CAACTGTATGTCGCCCACCACGTG--CGGTGC-TCGGCAAGTTAC-G----
2722      ACCGAGAA-CAACTGTATGTCGCCCACCACGTG--CGGTGC-CTCGGCAAGTTACG----
772       ACCGAGAA-CAACTGTATG-CGCCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
468       ACCGAGAT-CAACTGTATGTCGCCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
204       ACCGAGAC-CAACTGTATGTCGTCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
2875      ACCGAGAG-CAACTGTATGTCGTCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
741       ACCGAGAA-CAACTGTATGTCGCCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
1169      ACCGAGAA-CAACTGTATGTCGCCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
1012      ACCGAGAA-CAACTGTATGTCGCCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
330       ACCGAGAA-CAACTGTATGTCGCCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
2828      ACCGAGAA-CAACTGTATGTCGCCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
2795      ACCGAGAA-CAACTGTATGTCGCCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
2580      ACCGAGAA-CAACTGTATGTCGCCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
1038      ACCGAGAA-CAACTGTATGTCGCCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
679       ACCGAGAA-CAACTGTATGTCGCCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
610       ACCGAGAA-CAACTGTATGTCGCCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
266       ACCGAGAA-CAACTGTATGTCGCCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
659       ACCGAGAA-CAACTGTATGTCGCCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
2771      ACCGAGAA-CAACTGTATGTCGCCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
73        ACCGAGAA-CAACTGTATGTCGCCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
98        ACCGAGAA-CAACTGTATGTCGCCCACCACGTG--CGGTGC-TCGGCAAGTTACGC----
                  **** **        *  *            *

964       ---------------------C------TACCCCCC-TCTTGCGAGG---CTTCCCACCT
868       ------------CAGCCCCAACTTCGGGTCCCACCGCACGGGACTCGCGGCT----TACG
1837      CAG------------CCCCAACTTCGGCTCCACCGCACCTCGCGG-----------CACG
414       CAG------------CCCCAACTTCGGCTCCACCGCACCTCGCGG-----------CACG
212       CAG------------CCCCAACTTCGGCTCCACCGCACCTCGCGG-----------CACG
58        CAG------------CCCCAACTTCGGCTCCACCGCACCTCGCGG-----------CACG
2868      CAG------------CCCCAACTTCGGCTCCACCGCACCTCGCGG-----------CACG
1802      CAG------------CCCCAACTTCGGCTCCACCGCACCTCGCGG-----------CACG
431       CAG------------CCCCAACTTCGGCTCCACCGCACCTCGCGG-----------CACG
2682      CAG------------CCCCAACTTCGGCTCCACCGCACCTCGCGG-----------CACG
981       CAG------------CCCCAACTTCGGCTCCACCGCACCTCGCGG-----------CACG
```

```
2863      CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
85        CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
890       CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
2780      CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
2829      CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
103       CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
472       CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
1163      CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
487       CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
2810      CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
582       CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
517       CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
112       CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
129       CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
542       CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
2497      CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
120       CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
191       CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
2650      CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
2662      CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
74        CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
1854      CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
2819      CAG-----------CCCCAACTTCGGCTCCACCGCACCTCGCGG----------CACG
449       CAG-----------CCCCAACTTCGGCTCCTGTTA-CCGCACCTCGAGTCGTGCACGCC
138       AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
621       AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
749       AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
654       AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
416       AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
463       AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
1172      AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
496       AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
2586      AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
2688      AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
1094      AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
399       AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
1041      AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
117       AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
1097      AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
1810      AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
1056      AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
417       AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
2843      AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
2695      AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
2793      AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
490       AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
512       AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
2267      AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
2668      AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
863       AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
577       AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
1772      AGC-----------CCCA-ACTTCGGCCCACCGCA-CCTCGCGG----------CACG
988       AAGTTACGCCGGCAGCCCCAACTTCGGCCCACCGCA-CCTCGCGGCA---------GCGA
695       GACGCCGGGCAGCCTCCAACTTCGTGCCCCACCGCA-CCTGGCGCGG---------CACG
```

```
349       TAGCCCCA-----------ACTTCGGCCCACCGCA-CCTCGCGGCA---------CGAG
892       GACGCCGGGCAGCCTCCAACTTCGTGCCCCACCGCA-CCTGGCGCGG---------CACG
1829      GACGCCGGGCAGCCGCCAACTTCGCGGCCCACCGCA-CCTTGCGCGG---------CACG
388       -----CCGGCAGCCCCAACCCTCTCGGCCCACCGCA-CCTCGCG--G---------CACG
2722      -----CCGGCAGCCCC---AACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
772       -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
468       -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
204       -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
2875      -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
741       -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
1169      -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
1012      -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
330       -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
2828      -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
2795      -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
2580      -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
1038      -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
679       -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
610       -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
266       -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
659       -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
2771      -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
73        -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
98        -----CGGCAG----CCCCAACTTCGGCCCACCGCA-CCTCGCG--G---------CACG
                                               *

964       CAAATCACGCCA-----CA-ACAAGCCTGTC------------CCGCCGTCCTTGCCGCC
868       GGGAGCGCGTTAAA----CACCACGTCCTTGAGCTACCCAGCGCGATAGG----------
1837      GGGAGCCAAACACC------ACG-TCCGTTCCCCACGGCCATCGGGT-------------
414       GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
212       GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATG-GGT-------------
58        GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
2868      GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATG-GGT-------------
1802      GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
431       GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
2682      GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
981       GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
2863      GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
85        GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
890       GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
2780      GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
2829      GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
103       GGGAGCCAAACACC------ACGATCCT-TCCCCACGGCCATCGGGT-------------
472       GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
1163      GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
487       GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
2810      GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
582       GGGAGCCAAACACC------ACG--TCCTTCCCCACGGCCATCGGGT-------------
517       GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
112       GGGAGCCAAACACC------ACG-TCCGTTCCCCACGGCCATCGGGT-------------
129       GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
542       GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
2497      GGGAGCCAAACACC------ACGATCCT-TCCCCACGGCCATCGGGT-------------
120       GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT-------------
```

```
191       GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT------------
2650      GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT------------
2662      GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT------------
74        GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT------------
1854      GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT------------
2819      GGGAGCCAAACACC------ACGATCCGTTCCCCACGGCCATCGGGT------------
449       GGGAGCCAAACAATCCGGACGTCGGGCTTCCCCACGGATGGCGTGGT------------
138       GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
621       GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
749       GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
654       GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
416       GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
463       GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
1172      GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
496       GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
2586      GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
2688      GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
1094      GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
399       GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
1041      GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
117       GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
1097      GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
1810      GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
1056      GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
417       GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
2843      GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
2695      GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
2793      GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
490       GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
512       GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
2267      GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
2668      GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
863       GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
577       GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
1772      GGGAGCCAA-----------ACACCACGTCCTTCCCCACGGATGGGT------------
988       TGGGAGCCAAA--T---CACCACGTCTCTTCCC---CACGGATGGGTT------------
695       GGGAGCCAAACACC---A-----CGTCCTTCCC---CACGGATGGGTTGG----------
349       GGAGC--C--AAAG---CACCCACGTCCTTCCC---CACGGATGGGTT------------
892       GGGAGCCTACCCAA---CACTCACGTCCTTCCCCACTGCGGATGGGTCCGCATGGGAGTC
1829      GGGAGCCGACTAAA---CACCCACGTCCTTCCCCAGCACGGATGGGTCGTCCTGGGAGTA
388       GGGAGCCAAAC------ACCAGTCGTCCTTCCC---CACGGATGGGTT------------
2722      GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------
772       GGGAGCCA--A------ACACCAGTCCTTCC-----CCAAGATGGGTT------------
468       GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------
204       GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------
2875      GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------
741       GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------
1169      GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------
1012      GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------
330       GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------
2828      GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------
2795      GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------
2580      GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------
1038      GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------
```

```
679       GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------
610       GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------
266       GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------
659       GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------
2771      GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------
73        GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------
98        GGGAGCCA--A------ACACCACGTCCTTCCC---CACGGATGGGTT------------


964       ACGGATCAATGGGGAGTGTCTTTTCGATCTCAATCAGCCCAAATTCGGGTCCGTGCCCAC
868       ------CGTTGGGAGTGGTCTTTTGGCGTAAGACGAC-CCAGGCAGTAGCTGTGCC-CTC
1837      ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
414       ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
212       ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
58        ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
2868      ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
1802      ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
431       ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
2682      ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
981       ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
2863      ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
85        ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
890       ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
2780      ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
2829      ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
103       ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
472       ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
1163      ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
487       ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
2810      ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
582       ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
517       ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
112       ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
129       ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
542       ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
2497      ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
120       ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
191       ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
2650      ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
2662      ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
74        ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
1854      ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
2819      ----------TGGGAGTGTCTTTTGGCGTGGAGCGCCCAGCCGCTAGGC--GTGCC-CTC
449       ----------GCGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCCTACTCAA-CCC
138       ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
621       ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
749       ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
654       ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
416       ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
463       ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
1172      ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
496       ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
2586      ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
2688      ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
```

```
1094      ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
399       ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
1041      ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
117       ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
1097      ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
1810      ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
1056      ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
417       ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
2843      ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
2695      ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
2793      ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
490       ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
512       ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
2267      ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
2668      ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
863       ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
577       ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
1772      ----------TGGGAGTGTCTTTTGGCGTGACGCCCAGGCAGGCGTGCC--------CTC
988       ----------GGGAGTGTCTTTTGGCGTGACGAACCCAGCAGCAGGCGTGCCCTCAGCC
695       ------------GAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCA
349       ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGGTCGCCCTCAG
892       GGTCTTCCGTTGCCCGCGTTGAACGCAAGTTCCAGAG----CAAGGCAGTGCCCTAACAC
1829      GGTCTTCTTGCGGGCGCTGC---ACGCAACGCCAGTG----CTAGGCCGTGCCCTCAGCC
388       ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGAC
2722      ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGAC
772       ----------GGGAGTGTCTTTTGGCATGACACCCA----AGCAAGCGGCCCTCAGCAA
468       ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCC
204       ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCC
2875      ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCC
741       ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCC
1169      ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCC
1012      ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCC
330       ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCC
2828      ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCC
2795      ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCC
2580      ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCC
1038      ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCC
679       ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCC
610       ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCC
266       ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCC
659       ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCC
2771      ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCC
73        ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCC
98        ----------GGGAGTGTCTTTTGGCGTGACGCCCA----GGCAGGCGTGCCCTCAGCC


964       GGGAAT-TAGTCCTTAGGAATGCCATTTTGCCTGCC--TTCTTCTTCAA-----------
868       AGCTCAGAAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGATGG-------T
1837      AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
414       AGCCAGAAGGCCTCGGGCGCAA---CTT-GCGTTCA----AAAACTCGAT---------G
212       AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
58        AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATGGGT
2868      AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
1802      AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATG-GT
```

```
431       AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
2682      AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
981       AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
2863      AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
85        AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATG-GT
890       AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
2780      AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
2829      AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
103       AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
472       AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
1163      AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
487       AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
2810      AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
582       AGCCAGAAGGCCGGCGTGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
517       AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
112       AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
129       AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
542       AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
2497      AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
120       AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
191       AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
2650      AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
2662      AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
74        AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
1854      AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
2819      AGCCAGAAGGCCTGCGGGCGCA---ACTTGCGTTCA----AAACCGCACTGCGGATCGGT
449       GCACAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCAA-------TCGG
138       AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
621       AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
749       AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
654       AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
416       AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA-------GTCG
463       AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA-------GTCG
1172      AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
496       AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
2586      AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
2688      AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
1094      AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
399       AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
1041      AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
117       AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
1097      AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
1810      AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
1056      AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
417       AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
2843      AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
2695      AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
2793      AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
490       AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
512       AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
2267      AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
2668      AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
863       AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
577       AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
```

```
1772      AGCCAGAA-GGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA---------TG
988       AGAATT-GCAGCCTCGGGCGCA--ACATTGCGTTCA----AAAACTCGA--------T-G
695       AGA-----AGGCCTCGGGCGCAAGTTTTTGCGTTACCAAAAGCACTCGA--------T-G
349       CCAGAACGGACCTCGGGCAGCA---ACTTGCGTTCA----AAAACTCGA--------T-G
892       GCCAAGTAAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------TAG
1829      CAAGAGGAAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------TGG
388       CTTAAGTAAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------TAG
2722      CTTAAGTAAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------TAG
772       AA-----GGCCT-----------------------------------------------
468       AG-----AAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------T-G
204       AG-----AAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------T-G
2875      AG-----AAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------A-C
741       AG-----AAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------T-G
1169      AG-----AAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------T-G
1012      AG-----AAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------T-G
330       AG-----AAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------T-G
2828      AG-----AAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------T-G
2795      AG-----AAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------T-G
2580      AG-----AAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------T-G
1038      AG-----AAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------T-G
679       AG-----AAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------T-G
610       AG-----AAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------T-G
266       AG-----AAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------T-G
659       AG-----AAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------T-G
2771      AG-----AAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------T-G
73        AG-----AAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------T-G
98        AG-----AAGGCCTCGGGCGCA---ACTTGCGTTCA----AAAACTCGA--------T-G


964       TGTCGAACCATCACATCCTTCTTT-------GAAAGTTTTGATTTATT--T-ATTTTTTA
868       ATCGACGGGATTCATGCAATTCGTACAACTAGCGTTATGGCGCATTTTGCTACGTTCTTC
1837      TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
414       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATT-TTGCTACGTTCTTC
212       TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
58        TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
2868      TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
1802      TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
431       TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
2682      TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
981       TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
2863      TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
85        TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
890       TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
2780      TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
2829      TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
103       TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
472       TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
1163      TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
487       TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
2810      TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
582       TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
517       TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
112       TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
129       TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
```

```
542       TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
2497      TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
120       TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
191       TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
2650      TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
2662      TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
74        TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
1854      TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
2819      TCGCACGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
449       TTCGCGGGATTCTGCAATTCATCA-------CCAGGTATCGCATTTTGCATACGTTCTTC
138       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
621       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
749       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
654       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
416       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCTAGCGTTCTTC
463       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCTAGCGTTCTTC
1172      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
496       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
2586      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTG-CTACGTTCTTC
2688      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
1094      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
399       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
1041      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
117       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
1097      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
1810      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
1056      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
417       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
2843      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
2695      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
2793      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
490       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
512       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
2267      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
2668      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
863       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
577       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
1772      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCGTACGTTCTTC
988       GTTCGCGGGATTCTGCAATTCACA-------CTCAGGTATCGCATTTTGCTACGTTCTTC
695       GTTCGGGGAATTCATAGCATTTCACGATCCAGAAGTTTCTACTTTTTTGCTTCTTCCTTC
349       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGCATTTTGCTACGTATCTTC
892       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCAGCATTTTGCTACGTTCTTC
1829      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGGCATTTTGCTACGTTCTTC
388       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCAGCATTTTGCTACGTTCTTC
2722      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCAGCATTTTGCTACGTTCTTC
772       ------------------------------------------------------------
468       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGC-ATTTTGCTACGTTCTTC
204       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGC-ATTTTGCCACGTTCTTC
2875      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGC-ATTTTGCTAGGTTCTTC
741       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGC-ATTTTGCTACGTTCTTC
1169      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGC-ATTTTGCTACGTTCTTC
1012      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGC-ATTTTGCTACGTTCTTC
330       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGC-ATTTTGCTACGTTCTTC
2828      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGC-ATTTTGCTACGTTCTTC
```

```
2795      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGC-ATTTTGCTACGTTCTTC
2580      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGC-ATTTTGCTACGTTCTTC
1038      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGC-ATTTTGCTACGTTCTTC
679       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGC-ATTTTGCTACGTTCTTC
610       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGC-ATTTTGCTACGTTCTTC
266       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGC-ATTTTGCTACGTTCTTC
659       GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGC-ATTTTGCTACGTTCTTC
2771      GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGC-ATTTTGCTACGTTCTTC
73        GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGC-ATTTTGCTACGTTCTTC
98        GTTCGCGGGATTCTGCAATTCACA-------CCAGGTATCGC-ATTTTGCTACGTTCTTC


964       CTCTCTTTTCATATAGAAACCAATTTTCCCGGTC-CTCTGGGCGTAACCGCCCTGCATTT
868       ATCGAACTGCGAGAGCCGAGATATCCAAGAATTGCCGAGAGTCGT-------GTAGATTT
1837      ATCGATGCGAGAGCCCTA-AG-AGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
414       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
212       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
58        ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
2868      ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
1802      ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
431       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
2682      ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
981       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
2863      ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
85        ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
890       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
2780      ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
2829      ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
103       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
472       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
1163      ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
487       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
2810      ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
582       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
517       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
112       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
129       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
542       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
2497      ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
120       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
191       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
2650      ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
2662      ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
74        ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
1854      ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
2819      ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
449       ATCGATGCGAGAGCCGAG-ATATCCGT-----TGCCGACAGT----------CGTGTGGA
138       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
621       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
749       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
654       ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
416       ATCGATGCGAGAGCCGAG-ATATCCGT-----TGCCGAGAGTCG----------TGTGGA
463       ATCGATGCGAGAGCCGAG-ATATCCG-----TTGCCGAGAGTCG----------TGTGGA
1172      ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
```

```
496     ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
2586    ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
2688    ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
1094    ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
399     ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
1041    ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
117     ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
1097    ATCGATGCGAGAGCCCTA-AGAGCATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
1810    ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
1056    ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
417     ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
2843    ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
2695    ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
2793    ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
490     ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
512     ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
2267    ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
2668    ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
863     ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
577     ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
1772    ATCGATGCGAGAGCCCTA-AGACGATATCCGTTGCCGAGAGTCGTA-----AGTGGATGT
988     ACTCGATGCGAGAC-TG--CCGAGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
695     ATCGGATCGCAAAAAAC--CCAATATTTCCGTTTGCCGGAGGGGTG-----TGTGGTTTT
349     ATGCGAT-GCGAGA-GC--CGAGATATCCGATTGCCGAGAGTCGT-------GTGGATTT
892     ATCGATGCGACGAG-CT--CGACGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
1829    ATCGATGCGATGAG-CA--CGAGGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
388     ATCGATGCGACGAG-CT--CGACGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
2722    ATCGATGCGACGAG-CT--CGACGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
772     ------------------------------------------------------------
468     ATCGATGCGAGA---GA--CCCAGATATCCGTTGCCGAGAGTCGT-------GGGGATTT
204     ATCGATGCGAGA---GA--CCCAGATATCCGTTGCCGAGAGTCGT-------GGGGATTT
2875    ATCGATGCGAGA---GC---CGAGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
741     ATCGATGCGAGA---GC---CGAGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
1169    ATCGATGCGAGA---GC---CGAGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
1012    ATCGATGCGAGA---GC---CGAGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
330     ATCGATGCGAGA---GC---CGAGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
2828    ATCGATGCGAGA---GC---CGAGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
2795    ATCGATGCGAGA---GC---CGAGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
2580    ATCGATGCGAGA---GC---CGAGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
1038    ATCGATGCGAGA---GC---CGAGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
679     ATCGATGCGAGA---GC---CGAGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
610     ATCGATGCGAGA---GC---CGAGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
266     ATCGATGCGAGA---GC---CGAGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
659     ATCGATGCGAGA---GC---CGAGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
2771    ATCGATGCGAGA---GC---CGAGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
73      ATCGATGCGAGA---GC---CGAGATATCCGTTGCCGAGAGTCGT-------GTGGATTT
98      ATCGATGCGAGA---GC---CGAGATATCCGTTGCCGAGAGTCGT-------GTGGATTT


964     TTAAACCATA---GCTGCCTGATC------------CGCCTACGCAACAATTACCAT---
868     CAACAGCTCCGG------ATC---------------------------------------
1837    TAGAAGTAGCATCGCTGCACGAGG------------TGCGTTCGGCAAGGCCGACC----
414     TAGAAGTAGCATCGCTGCACGAGG------------TGCGTTCGGCAACGGCCGACCG--
212     TAGAAGTAGCATCGCTGCACGAGG------------TGCGTTCGGCAACGGCCGACCG--
```

```
58      TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
2868    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
1802    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
431     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
2682    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
981     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAAG-GCCGACCG--
2863    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
85      TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
890     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
2780    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
2829    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
103     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
472     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
1163    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
487     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
2810    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
582     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
517     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
112     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
129     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
542     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
2497    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
120     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
191     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
2650    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
2662    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
74      TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
1854    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
2819    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
449     TTTAAATAGCATCGCTGCACGAGG-----------TGCATTCGGCAACGGCCGACCG--
138     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
621     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
749     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
654     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
416     TTTAGATAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
463     TTTAGATAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
1172    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
496     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
2586    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
2688    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
1094    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
399     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
1041    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
117     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
1097    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAAGCGCCGACCG--
1810    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
1056    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
417     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
2843    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
2695    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
2793    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
490     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
512     TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
2267    TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
```

```
2668      TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
863       TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
577       TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
1772      TAGAAGTAGCATCGCTGCACGAGG-----------TGCGTTCGGCAACGGCCGACCG--
988       AGATAGCATC-G--CTGCACGAGGTGCGTTCGGCCGA-AGTTGGGGCCGACGGCGCAACT
695       AAAAAAAGCC-CCCCGGCTGGCGTTGGGGCTCGTTCTGCGATATTCACAAGGAACCCCCA
349       AGATAGCAAG-TCGCTGCACGAGG-----------TGCGTTCTGGCTGAGGGCACGACC
892       AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A
1829      AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A
388       AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A
2722      AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A
772       ------------------------------------------------------------
468       AGCGAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCATGCCGACCGC--A
204       AGCGAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCATGCCGACCGC--A
2875      AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCATGCCGACCGC--A
741       AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A
1169      AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A
1012      AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A
330       AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A
2828      AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A
2795      AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A
2580      AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A
1038      AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A
679       AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A
610       AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A
266       AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A
659       AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A
2771      AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A
73        AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A
98        AGATAGCATC-G--CTGCACGAGG-----------TGCGTTCGGCAGGCCGACCGC--A


964       ------GTCCCCAAGCATTAGAAAATTGTAAACTCTTGAATGATCCCTCCCCAT------
868       ------------------------------------------------------------
1837      -------GCAACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
414       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
212       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
58        -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
2868      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
1802      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
431       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
2682      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
981       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
2863      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
85        -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
890       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
2780      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
2829      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
103       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
472       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
1163      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
487       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
2810      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
582       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
```

```
517       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
112       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
129       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
542       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
2497      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
120       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
191       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
2650      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
2662      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
74        -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
1854      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
2819      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
449       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
138       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
621       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
749       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
654       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
416       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
463       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
1172      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
496       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
2586      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
2688      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
1094      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
399       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
1041      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
117       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
1097      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTG-ACGCCGTTCG
1810      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
1056      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
417       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
2843      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
2695      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
2793      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
490       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
512       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
2267      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
2668      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
863       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
577       -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
1772      -------CACACCCCGAGCAAGACAATAT--------AAGTGTTCCTTGCACGCCGTTCG
988       TCCCGAGCACCGCACGTGGTGGACAATATACA----GTTGTTCTCGTTGACGC-CTCTCG
695       AC----CCGCCACCCAAGAAAAAGCAAAAAATTTTTTTTTTGTTCCTTGCCCCCCTCCCG
349       G------CAACCCCGAGTCAAGACAATAT--------AAGTGTTCCTTGACAGCCTTCGG
892       A------CCCCGAGCAACGACAA--TATA--------AGGTGTTCCTTGACGCCTTCGGC
1829      A------CCCCGCCCAAGCAAGACAATAT--------AAGTGTTCCTTGACGCCTTCGGC
388       A------CCCCGAAAACGCAAGACAATAT--------AAGTGTTCCTTGACGCCTTCGGC
2722      A------CCCCGAAAACGCAAGACAATAT--------AAGTGTTCCTTGACGCCTTCGGC
772       ------------------------------------------------------------
468       T------CCCCGAG----CAAGACAATAT--------AAGTGTTCCTTGTCGCCGTCGGC
204       T------CCCCGAG----CAAGACAATAT--------AAGTGTTCCTTGTCGCCGTCGGC
2875      G------CCCCGAG----CAAGACAATAT--------AAGTGTTCCTTGCCGCCTTCGGC
741       A------CCCCGAG----CAAGACAATAT--------AAGTGTTCCTTGACGCCTTCGGC
1169      A------CCCCGAG----CAAGACAATAT--------AAGTGTTCCTTGACGCCTTCGGC
```

```
1012      A------CCCCGAG----CAAGACAATAT--------AAGTGTTCCTTGACGCCTTCGGC
330       A------CCCCGAG----CAAGACAATAT--------AAGTGTTCCTTGACGCCTTCGGC
2828      A------CCCCGAG----CAAGACAATAT--------AAGTGTTCCTTGACGCCTTCGGC
2795      A------CCCCGAG----CAAGACAATAT--------AAGTGTTCCTTGACGCCTTCGGC
2580      A------CCCCGAG----CAAGACAATAT--------AAGTGTTCCTTGACGCCTTCGGC
1038      A------CCCCGAG----CAAGACAATAT--------AAGTGTTCCTTGACGCCTTCGGC
679       A------CCCCGAG----CAAGACAATAT--------AAGTGTTCCTTGACGCCTTCGGC
610       A------CCCCGAG----CAAGACAATAT--------AAGTGTTCCTTGACGCCTTCGGC
266       A------CCCCGAG----CAAGACAATAT--------AAGTGTTCCTTGACGCCTTCGGC
659       A------CCCCGAG----CAAGACAATAT--------AAGTGTTCCTTGACGCCTTCGGC
2771      A------CCCCGAG----CAAGACAATAT--------AAGTGTTCCTTGACGCCTTCGGC
73        A------CCCCGAG----CAAGACAATAT--------AAGTGTTCCTTGACGCCTTCGGC
98        A------CCCCGAG----CAAGACAATAT--------AAGTGTTCCTTGACGCCTTCGGC


964       ---------GTTCACCTATGCAGGCTCTTTCA--AAATCCTCGAGAATAAAGTTCTCCAG
868       ------------------------------------------------------------
1837      GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
414       GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
212       GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
58        GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
2868      GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
1802      GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
431       GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
2682      GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
981       GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGAAGGTCGAGG----
2863      GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
85        GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
890       GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
2780      GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
2829      GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
103       GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
472       GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
1163      GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
487       GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
2810      GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
582       GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
517       GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAAA--AAAAGGTCGAGGCC
112       GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
129       GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
542       GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
2497      GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
120       GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
191       GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
2650      GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
2662      GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
74        GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
1854      GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
2819      GCTGCCGTGGGTTCTTTTG--CGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
449       GCTGCCGTTGGGGGTTCTTTTGAGGCCTCT-----TCACCAGGATCTAGAAAAAGGTCGAG
138       GCTGCCGTTGGGGGTTCTTTTGCGGCCTCT-----TCCCTCTAGAACTGCGAAGGT----C
621       GCTGCCGTTGGGGGTTCTTTTGCGGCCTCT-----TCCCTCTAGAACTGCGAAGGT----C
749       GCTGCCGTTGGGGCTCGTTCTGTTATATT-----CACAAGGAA------CTGCGGAAGCC
654       GCTGCCGTTGGGGGTTCTTTTGCGGCCTCT-----TCCCTCTAGAACT-GCGAAGGTCGAG
```

```
416       GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAAAA--AAAGGTCGAGGCC
463       GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
1172      GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAAAAA--AAGGTCGAGGCC
496       GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TACCCTCTAAAAGAACGGTCGAGGCC
2586      GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
2688      GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
1094      GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
399       GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
1041      GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
117       GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
1097      GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
1810      GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAAAA--AAAGGTCGAGGCC
1056      GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
417       GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
2843      GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
2695      GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
2793      GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
490       GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
512       GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
2267      GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
2668      GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
863       GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
577       GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
1772      GCTGCCGTGGGTTCTTTT--GCGGCCTCT-----TCCCTCTAGAAGGAACGGTCGAGGCC
988       GCTAGCCGTGGGTTCTTTTGCGGCC-------TACTTCCCTCTAGAAGAAGGTAGCGAGG
695       GCACC--GTGGTTTTTTTTTCCCC-----CCCTCCTCCCCAAAAAAAAAAGGGGGGAGG-
349       GAACGCCGTGGGTTCTTTTGCGGCCTCTTC--CCTCTAGGTGAAGAAGGTCGAAGTTGGG
892       GCCGT----GGGTTCGTTTTGCGGCCCGTC----GTTCCCTCTAGAAGAAGGTCGAGGCC
1829      GCCGT----GGGTTCGTTTTGCGGTCCGTC----CTTCCCTCTAGAAGAAGGTCGAGGCC
388       GCCGT----GGGTTCGTTTTGCGG----CCCGTCGTTCCCTCTAGAAGAAGGTCGAGG--
2722      GCCGT----GGGTTCGTTTTGCGG----CCCGTCGTTCCCTCTAGAAGAAGGTCGAGG--
772       ------------------------------------------------------------
468       TCCGC----CACGCCAGCCGTCCGGCTGGCCTCTTCCCTCTAGCTGTTATGGTCGAGG--
204       TCCGT----GGGTTCTTTTGC--------GGCCTCTTCCCTCTAGAAGAACGTCGAGG--
2875      CCCGT----GGGTTCTTTTGC--------GGCCTCTTCCCTCTAGAATGAGGTCGAGG--
741       GCCGT----GGGTTCTTTTGC--------GGCCTCTTCCCTCTAGAAGAAGGTCGAGG--
1169      GCCGT----GGGTTCTTTTGC--------GGCCTCTTCCCTCTAGAAGAAGGTCGAGG--
1012      GCCGT----GGGTTCTTTTGC--------GGCCTCTTCCCTCTAGAAGAAGGTCGAGG--
330       GCCGT----GGGTTCTTTTGC--------GGCCTCTTCCCTCTAGAAGAAGGTCGAGG--
2828      GCCGT----GGGTTCTTTTGC--------GGCCTCTTCCCTCTAGAAGAAGGTCGAGG--
2795      GCCGT----GGGTTCTTTTGC--------GGCCTCTTCCCTCTAGAAGAAGGTCGAGG--
2580      GCCGT----GGGTTCTTTTGC--------GGCCTCTTCCCTCTAGAAGAAGGTCGAGG--
1038      GCCGT----GGGTTCTTTTGC--------GGCCTCTTCCCTCTAGAAGAAGGTCGAGG--
679       GCCGT----GGGTTCTTTTGC--------GGCCTCTTCCCTCTAGAAGAAGGTCGAGG--
610       GCCGT----GGGTTCTTTTGC--------GGCCTCTTCCCTCTAGAAGAAGGTCGAGG--
266       GCCGT----GGGTTCTTTTGC--------GGCCTCTTCCCTCTAGAAGAAGGTCGAGG--
659       GCCGT----GGGTTCTTTTGC--------GGCCTCTTCCCTCTAGAAGAAGGTCGAGG--
2771      GCCGT----GGGTTCTTTTGC--------GGCCTCTTCCCTCTAGAAGAAGGTCGAGG--
73        GCCGT----GGGTTCTTTTGC--------GGCCTCTTCCCTCTAGAAGAAGGTCGAGG--
98        GCCGT----GGGTTCTTTTGC--------GGCCTCTTCCCTCTAGAAGAAGGTCGAGG--


964       GGCCTACCCGTATCAAAAACCCTTGTAGCATGGGTGACTTTGGTCAGCTCCTTGGTTT--
868       ------------------------------------------------------------
```

```
1837      GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
414       GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
212       GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCCGGTTC
58        GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
2868      GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
1802      GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
431       GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
2682      GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
981       --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGATCTTCATCAGTT--CACGGATCT
2863      GTAGCAACCCGAGCAGAAGCCCGGGCAGCATGGGTGACCATGTTCCACT--CCCGGTTCT
85        GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
890       GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
2780      GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
2829      GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
103       GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
472       GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
1163      GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
487       GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
2810      GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
582       GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCA-GC--TCCGGATCT
517       GTAGCACC-CGAGCAAAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
112       GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
129       GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
542       GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
2497      GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTATCAGC--TCCGGATCT
120       GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCA-GC--TCCGGATCT
191       GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
2650      GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
2662      GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
74        GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
1854      GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
2819      GTAGCACC-CGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC--TCCGGATCT
449       GCCTAG-CCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTAGGCCAGCTCCGGATCT
138       GAGGCA-CCCAAGAATACGCCCGAGCAGGAAGCCCGGGTT----CAGC--TCCGGATCGG
621       GAGGCA-CCCAAGAATACGCCCGAGCAGGGAGAGCTTCTT----CCCG--GGCAGCATGG
749       TCTTCC-CTCTAGAAAACGCCGGAGAAGGTCGAGGCCTAG----CCCGAGCAGAAGCCCG
654       GCCTAC-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGCTCCGGATTT--
416       GTAGCA-CCCGAGCAAAAGCCCGGGCAGCATGGGTGATCATGTTCCAGCTCCGGATCT--
463       GTAGCA-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCACGGTCTGTT--TTG
1172      GTAGCA-CCCGAGAAGAAGCCCGGGCAGCATGGGGGACATGTTCAAAACGGTCTGTTTTT
496       GTAGCA-CCCGAGCAGAAGCCCGGGCAGGGACATGGGTCGACATGT----TCACGGATCT
2586      GTAGCA-CCCGAGCAGAAGCCCGGGCAGGCATGGGTGACATGTTCACGG-ATCT------
2688      GTAGCA-CCCGAGCAGAAGCCCGGGCAGGGACATGGGTGACATGTT-----CACGGATCT
1094      GTAGCA-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCACGG-TCTGTTTTGG
399       GTAGCA-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCACGG-TCTGTTTTGG
1041      GTAGCA-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCACGG-TCTGTTTTGG
117       GTAGCA-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC-TCCGGATCTG
1097      GTAGCA-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTATCAGC-TCCGGATCTG
1810      GTAGCA-CCCGAGCAAAAGCCCGGGCAGCATGGGTGATCATGTTCAGCT-CCGGATC--T
1056      GTAGCA-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCTTCATCCAGT-TCACGGATCT
417       GTAGCA-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCAGCT-C-CGGATC-T
2843      GTAGCA-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCAGCT-C-CGGATC-T
2695      GTAGCA-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC-TCCGGATC-T
2793      GTAGCA-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC-TCCGGATC-T
```

```
490       GTAGCA-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC-TCCGGATC-T
512       GTAGCA-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC-TCCGGATC-T
2267      GTAGCA-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC-TCCGGATC-T
2668      GTAGCA-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC-TCCGGATC-T
863       GTAGCA-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC-TCCGGATC-T
577       GTAGCA-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTATCAGC-TCCGGATC-T
1772      GTAGCA-CCCGAGCAGAAGCCCGGGCAGCATGGGTGATCATGTTCCAGC-TCCGGATC-T
988       CCTAGCTCCGAGCGAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTGTTT
695       --CCCCCCCCCAAAACAAACCCCGCGGGGGGGGGGGAGAG-------------------
349       --CCTAGCCCGAGCAGAAGCCCGGGCACCGCAAGTGGTGGGTGACATACAGTTGTTCACG
892       ----TAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTGTTT
1829      --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTCCCCAG
388       --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTGTTT
2722      --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTGTTT
772       ------------------------------------------------------------
468       --CCTAGCCCGAGCAGAAGCCCGGGCAGGATGGGTGACATGTTCACGGTCTGTTTTGTTT
204       --CCTAGTCCGAGCAGAAGCCCGGGCAGCATGGGTGAAATGTTAACCGGGCTGTTTTGTC
2875      --CCAAGAACGAGCAAAAGCCCGGGCAGCATGGGTGACATGTTGAAAGGGCTGTTTTGTT
741       --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTGTTT
1169      --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTGTTT
1012      --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTGTTT
330       --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTGT--
2828      --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTGTTT
2795      --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTGTT-
2580      --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTG---
1038      --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTGTT-
679       --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTGTT-
610       --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTG---
266       --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTGT--
659       --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTGT--
2771      --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTGT--
73        --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTGT--
98        --CCTAGCCCGAGCAGAAGCCCGGGCAGCATGGGTGACATGTTCACGGTCTGTTTTGT--


964       ------------GTTGAAAGGTCACGACAAT----------------------------
868       ------GGGCTCGTCTTCGGGCCCGTCGTACCCACTGTACAAGTGCCAG-----------
1837      G-------TTTTGTCTTCGGGCCCGTCGTACCCACTGTACAAGT-G----CCAG------
414       G-------TTTTGTCTTAAGGCCCGTCGTACCCACTGTACAAGT-G----CCAG------
212       T-------GTTTTGTTTTAAGCCCGTCGTACCCACTGTACAAGT-G----CCAG------
58        G-------TTTTGTCTTAAGGCCCGTCGTACCCACTGTACAAGT-G----CCA-------
2868      G-------TTTTGTCTTAAGGCCCGTCGTACCCACTGTACAAGT-G----CCAG------
1802      G-------TTTTGTCTTAAGGCCCGTCGTACCCACTGTACAAGT-G----CCAG------
431       G-------TTTTGTCTTAAGGCCCGTCGTACCCACTGTACAAGT-G----CCAG------
2682      G-------TTTTGTCTTAGGGCCCGTCGTACCCACTGTACAAGT-G----CCAG------
981       G-------TTTTGTCTTCGGGCCCGTCGTACCCACTGTACAAGT-G----CCAGACA---
2863      G-------TTTTGTTTAAGGGTCACTCGTACC-CACTGTACAAG-T----GCCA------
85        G-------TTTTGTTTAAGGGTCACTCGTA-CCCACTGTACAAG-T----GC--------
890       G-------TTTTGTTTAAGGGTCACGTCGTACCCACTGTACAAG-T----GCCA------
2780      G-------TTTTGTCTTAAGGCCCGTCGTACCCAC-TGTACAGT-G----CCA-------
2829      G-------TTTTGTCTTCGGGCCCGTCGTACCCAC-TGTACAGT-G----CCA-------
103       G-------TTTTGTCTTAAGGCCCGTCGTACCCAC-TGTACAAG-T----GCC-------
472       G-------TTTTGTCTTAGGGCCCGTCGTACCCAC-TGTACAAG-T----GCC-------
1163      G-------TTTTGTCTTAGGGCCCGTCGTACCCAC-TGTACAAG-T----GCC-------
```

```
487       G-------TTTTGTCTTAAGGCCCGTCGTACCCAC-TGTACAAG-T----GCC-------
2810      G-------TTTTGTCTTAAGGCCCGTCGTACCCAC-TGTACAAG-T----GCC-------
582       G-------TTTTGTCTTCGGGCCCGTCGTACCCAC-TGTACAGT-G----CC--------
517       G-------TTTTGTCTTAAGGCCCGTCGTACCCAC-TGTACAAG-T----GC--------
112       G-------TTTTGTCTTCGGGCCCGTCGTACCCAC-TGTACAGT-G----CCAG------
129       G-------TTTTGTCTTCGGGCCCGTCGTACCCAC-TGTACAGT-G----CCAGCAAC--
542       G-------TTTTGTCTTAGGGCCCGTCGTACCCAC-TGTACAGT-G----CCAGACAA--
2497      G-------TTTTGTCTTAAGGCCCGTCGTACCCAC-TGTACAAG-T----GCCAGA----
120       G-------TTTTGTCTTCGGGCCCGTCGTACCCAC-TGTACAAG-T----GCCAGA----
191       G-------TTTTGTCTTAAGGCCCGTCGTACCCAC-TGTACAGT-G----CCA-------
2650      G-------TTTTGTCTTAGGGCCCGTCGTACCCAC-TGTACAAG-T----GCCAGAC---
2662      G-------TTTTGTCTTAGGGCCCGTCGTACCCAC-TGTACAAG-T----GCCAGAC---
74        G-------TTTTGTATTAAGGCCCGTCGTACCCAC-TGTACAAG-T----G---------
1854      G-------TTTTGTCTTAAGGCCCGTCGTACCCAC-TGTACAAG-T----GCCA------
2819      G-------TTTTGTCTTAAGGCCCGTCGTACCCAC-TGTACAAG-T----GCCAGA----
449       G-------TTTTGTTTAAGGGTCCGTCGTACCCACTGTACAAGT-G----CC--------
138       G-------CTCGACATCGT-TCACGTCGTACCCACTGTACATTT-G----TTT-------
621       G-------TGACATGTTCGGGACGGTCGTGTCCTTTGTTTAAGT-G----CCA-------
749       G-------GCTAGTCATCGGGCCTGACATGTTCACGGTCTGTTT-T----GTTT------
654       G-------TTTTGTCTTAGGGCCCGTCGTACCCACTGTACAAGT-G----CCAGA-----
416       G-------TTTTGTCTTAAGGCCCGTCGTACCCACTGTACAGTG-C----C---------
463       G-------GCTCGTCTTCGGGCCCGTCGTACCCACTGTACAGTG----------------
1172      G-------TCTCGTCTTCGGGCCCGTCGTACCCACTGTACAAGT-G----CCAGA-----
496       G-------TTTTGTCTTCGGGCCCGTCGTACCCACTGTACAAGT-G----CCAGA-----
2586      G-------TTTTGTTTAAGGGTCCGTCGTACCCACTGTCACGTG-C----CAGAA-----
2688      G-------TTTTGTCTTAGGGCCCGTCGTACCCACTGTACAAAG-T----GCCAGACA--
1094      G-------CT-CGTCTTCGGGCCCGTCGTACCCACTGT-ACAAG-T----GCCAGACAAC
399       G-------CT-CGTCTTCGGGCCCGTCGTACCCACTGT-ACAGT-G----CCAGACAACC
1041      G-------CT-CGTCTTCGGGCCCGTCGTACCCACTGT-ACAAG-T----GCCAGA----
117       T-------TTTG-TCTTAAGGCCCGTCGTACCCACTGT-ACAGT-G----CCA-------
1097      T-------TT-TGTCTTAAGGCCCGTCGTACCCACTGTACAAGT-G----CCA-------
1810      G-------TTTTGTCTTAGGGCCCGTCGTAC-CCACTGTACAAG-T----GCC-------
1056      G-------TTTTGTCTTCGGGCCCGTCG-TACCCACTGTACAAG-T----GCCAGAC---
417       G-------TTTTGTCTTAAGGCCCGTCG-TACCCACTGTACAAG-T----GCCAGAC---
2843      G-------TTTTGTCTTCGGGCCCGTCG-TACCCACTGTACAAG-T----GCCAGACA--
2695      G-------TTTTGTTTAAGGGTCACGTCGTACCCACTGTACAAG-T----GCCAGA----
2793      G-------TTTTGTCTTCGGGCCCGTCG-TACCCACTGTACAGT-G----CCAGACAAAC
490       G-------TTTTGTCTTAGGGCCCGTCG-TACCCACTGTACAGT-G----CC--------
512       G-------TTTTGTCTTAGGGCCCGTCG-TACCCACTGTACAAG-T----GCCA------
2267      G-------TTTTGTCTTAAG----------------------------------------
2668      G-------TTTTGTCTTAAGGCCCGTCG-TACCCACTGTACAAG-T----GCCAGAC---
863       G-------TTTTGTCTTCGGGCCCGTCG-TACCCACTGTACAAG-T----GCCA------
577       G-------TTTTGTCTTAGGGCCCGTCG-TACCCACTGTACAAG-T----GCCAG-----
1772      G-------TTTTGTCTTAGGGCCCGTCG-TACCCACTGTACAAG-T----GCCAG-----
988       AAG-------------------------------------------------------
695       ----------------------------------------------------------
349       GTGCAGTGTTTTGTTTAAGAGTCGGCTCGTTGGCC--CTAAATGACC---CATGACGACC
892       ----------------------------------------------------------
1829      C---------------------------------------------------------
388       AA----GGGT------------------------------------------------
2722      AA--------------------------------------------------------
772       ----------------------------------------------------------
468       AA----CTCCGGAC--------------GGGCTCGTCTTCGGGCCCGTCTCGTACCCACT
204       TA----AGTCTTCTTCCAGCTCCGGATCGGGCTCGTCTTCGGG--CCCGTCGTACCCACT
```

```
2875      TA----AGGG--------TCTCCGGATCGGGCTCGTCTTCGGG--CCCGTCGTACCCACT
741       AA----G-----------------------------------------------------
1169      AA----------------------------------------------------------
1012      A-----------------------------------------------------------
330       ----------------------------------------------------------
2828      ----------------------------------------------------------
2795      ----------------------------------------------------------
2580      ----------------------------------------------------------
1038      ----------------------------------------------------------
679       ----------------------------------------------------------
610       ----------------------------------------------------------
266       ----------------------------------------------------------
659       ----------------------------------------------------------
2771      ----------------------------------------------------------
73        ----------------------------------------------------------
98        ----------------------------------------------------------


964       ---------------------------------------
868       ---------------------------------------
1837      ---------------------------------------
414       ---------------------------------------
212       ---------------------------------------
58        ---------------------------------------
2868      ---------------------------------------
1802      ---------------------------------------
431       ---------------------------------------
2682      ---------------------------------------
981       ---------------------------------------
2863      ---------------------------------------
85        ---------------------------------------
890       ---------------------------------------
2780      ---------------------------------------
2829      ---------------------------------------
103       ---------------------------------------
472       ---------------------------------------
1163      ---------------------------------------
487       ---------------------------------------
2810      ---------------------------------------
582       ---------------------------------------
517       ---------------------------------------
112       ---------------------------------------
129       ---------------------------------------
542       ---------------------------------------
2497      ---------------------------------------
120       ---------------------------------------
191       ---------------------------------------
2650      ---------------------------------------
2662      ---------------------------------------
74        ---------------------------------------
1854      ---------------------------------------
2819      ---------------------------------------
449       ---------------------------------------
138       ---------------------------------------
```

```
621         ----------------------------------------
749         ----------------------------------------
654         ----------------------------------------
416         ----------------------------------------
463         ----------------------------------------
1172        ----------------------------------------
496         ----------------------------------------
2586        ----------------------------------------
2688        ----------------------------------------
1094        C---------------------------------------
399         TT--------------------------------------
1041        ----------------------------------------
117         ----------------------------------------
1097        ----------------------------------------
1810        ----------------------------------------
1056        ----------------------------------------
417         ----------------------------------------
2843        ----------------------------------------
2695        ----------------------------------------
2793        ATT-------------------------------------
490         ----------------------------------------
512         ----------------------------------------
2267        ----------------------------------------
2668        ----------------------------------------
863         ----------------------------------------
577         ----------------------------------------
1772        ----------------------------------------
988         ----------------------------------------
695         ----------------------------------------
349         GTAGCGCTTTGTCGCTCGGACCGCGACCCCAGGTCAGAC
892         ----------------------------------------
1829        ----------------------------------------
388         ----------------------------------------
2722        ----------------------------------------
772         ----------------------------------------
468         GTACAAGTGCC-----------AG----------------
204         GTACAAGTGCC-----------A-----------------
2875        GTACAAGTGCC-----------AGAC--------------
741         ----------------------------------------
1169        ----------------------------------------
1012        ----------------------------------------
330         ----------------------------------------
2828        ----------------------------------------
2795        ----------------------------------------
2580        ----------------------------------------
1038        ----------------------------------------
679         ----------------------------------------
610         ----------------------------------------
266         ----------------------------------------
659         ----------------------------------------
2771        ----------------------------------------
73          ----------------------------------------
98          ----------------------------------------
```