DISTRIBUTION, DIVERSITY AND SALINITY INDUCED TRANSCRIPTOMICS IN DOUM PALM (*Hyphaene compressa*)

AGNES AWUOR OMIRE

DOCTOR OF PHILOSOPHY

(Genetics)

JOMO KENYATTA UNIVERSITY

OF

AGRICULTURE AND TECHNOLOGY

2023

Distribution, Diversity and Salinity Induced Transcriptomics in Doum palm *(Hyphaene compressa)*

Agnes Awuor Omire

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Genetics of the Jomo Kenyatta University of Agriculture and Technology

2023

DECLARATION

This thesis is my original work and has not been presented for a degree in any other university.

Signature: …… Date: ……

Agnes Awuor Omire

This thesis has been submitted for examination with our approval as university supervisors.

Signature: …… Date: ……

Dr. Cecilia M. Mweu, PhD

JKUAT, Kenya

Signature: …… Date: …….

Prof. Nancy Budambula, PhD

University of Embu, Kenya

Signature: …… Date: ……

Dr. Johnstone O. Neondo, PhD

JKUAT, Kenya

Signature: …… Date: ……

Prof. Robert W. Gituru, PhD

JKUAT, Kenya

DEDICATION

I dedicate this thesis to my family; my beloved husband Boniface in appreciation of his constant encouragement and support. To our daughters, Abigael, Matchel and Hawi, for giving me reasons to smile every time things became difficult. To my beloved parents Clementina Omire and the late Julius Omire who instilled in me the value of hard work; my siblings for their prayers, support and encouragement.

ACKNOWLEDGEMENT

I am most thankful to the Almighty God from whom my strength and endurance came from during this pursuit. I am forever indebted to my supervisors; Dr. Cecilia Mweu, Prof. Nancy Budambula, Dr. Johnstone Neondo and Prof. Robert Gituru for their valuable advice, insightful reviews and support throughout my studies. I gladly recognize the Kenya National Research Fund (NRF), for funding this research.

I really appreciate the invaluable knowledge that the people of Tharaka Nithi, Kwale, Tana River and Turkana contributed to this work. I would like to thank the Nuts and Oil Crops Directorate (Agriculture and Food authority) in Mombasa that guided the survey and sample collection in Tana River and Kwale counties. I am grateful to the Kenya Wildlife Service (Tharaka Nithi), The Anglican Development Services in Lodwar and The National Museums of Kenya, for their logistic support during sample collection. I am thankful for the assistance provided by Ms. Josephine Kyaa (National Museums of Kenya) during the field work and greenhouse set up.

I would also like to thank Institute for Biotechnology Research (IBR) for giving me access to IBR molecular laboratory. Special thanks go to Mr. Rotich for his immense knowledge on molecular work, Catherine Eyinda, Dr. Mercy Kidaha and Dr. Grace Mungai for their support at IBR. I appreciate Prof. Robert Gituru through the Sino-Africa Joint Research Centre (SAJOREC) for granting access to the molecular and laboratory. I acknowledge the assistance provided by my colleagues during data analysis; Mr. Stephen Ogada during GBS analysis, Dr. Samwel Were, Dr. Josiah Kuja, Dr. Joshua Muli, Dr. Margaret Odeph, Mr. Tofick Wekesa, Ms. June Fridah and Mr. Christopher Wesonga.

Finally, I would like to thank my family; My husband Boniface for being my number one cheerleader. I appreciate you proofreading the numerous pages of science material and to my children who had to endure so much at a tender age for me to achieve this. I also appreciate the support from my parents, the entire Omire family, Odhiambo family and friends who prayed for me, prayed with me and helped me in one way or the other, in the course of my research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND ACRONYMS

ABA	Abscisic acid
accD	Acetyl COA dehydrogenase
AFLP	Amplified length Polymorphism
ASALs	Arid and semi-Arid lands
AMOVA	Analysis of molecular variance
ANOVA	Analysis of variance
APX	Ascorbate peroxidases
BOI	Botrytis susceptible interactor
BR	Brassinosteroids
BRGs	BOI related genes
BUSCO`	Benchmarking universal single copy orthologs
CAT	Catalases
DAPC	Discriminant analysis of principal components
DArT	Diversity Arrays Technology
DARwin	Dissimilarity analysis and representation for windows
DEGs	Differentially expressed genes
DMSO	Dimethyl sulphoxide
EC	Electrical conductivity
ESTs	Expressed sequence tags
EST-SSRs	Expressed Sequence-Simple Sequence repeats
F_{IS}	Inbreeding coefficient
FL	Fidelity level
FPKM	Fragments per kilo base of exon per million mapped reads
F_{ST}	Fixation Index
g-SSR	Genomic SSRs
GBS	Genotyping by sequencing
GO	Gene ontology
H_e	Expected heterozygosity
H_o	Observed heterozygosity
HSPs	Heat Shock Proteins

IPGR	International plant genetics resources institute
ISSR	Inter simple sequence repeats
KEGG	Kyoto Encyclopedia of Genes and Genomes
KOG	Eukaryotic orthologous groups
KW	Kwale county
LD	Linear discriminant
LEA	Late embryogenesis abundant proteins
Log_2FC	Log 2v fold change
MAPK	Mitogen-activated protein kinase
matK	Maturase K
MCMC	Markov chain monte carlo
mRNA	Messenger RNA
ORF	Open reading frame
PAM	Partitioning around medoids
PC	Principal components
PCoA	Principal coordinate analysis
PPV	Plant Part Value
ndhJ	NADH dehydrogenase subunit
NGS	Next generation sequencing
NR	Non-redundant database of NCBI
NT	Nucleotide database
QC	Quality control
QTL	Quantitative trait loci
RAPDs	Random amplified polymorphic DNA
rbCL	Ribulose-1,5 biphosphate carboxylase/oxygenase large subunit
RCI	Relative cultural importance indices
RFLP	Restriction fragment length polymorphism
RGAP	Resistance Gene Analog polymorphism
RINe	RNA Integrity Number equivalent
RING	Really interesting new gene
RNAseq	RNA sequencing

ROS	Reactive oxygen species
rpoB	RNA Polymerase beta subunit
RU	Reported use
$RU_{Plant\ Part}$	Reported use per plant part
SCAR	Sequence characterized amplified regions
SOD	Superoxide dismutase
SOS	Salt Overly Sensitive
SNP	Single nucleotide polymorphisms
SSR	Simple sequence repeats
STR	Short tandem repeats
SSLP	Simple sequence length polymorphisms
TF	Transcription factor
TFDB	Transcription factor database
THK	Tharaka Nithi county
TR	Tana River county
TUR	Turkana county
STI	Salt tolerance index
SII	Salt injury index

ABSTRACT

Doum palm (*Hyphaene compressa*) is one of the few perennial evergreen plants that grow in Arid and Semi-Arid Lands (ASALs) of Africa. However, ethnobotanical knowledge about the plant is scanty and limited genetic data has been documented. Despite the doum palm's widespread presence in ASALs characterized by high saline soils, little is known about its salinity tolerance mechanisms. This study aimed to determine the ethnobotanical knowledge, the morphological and genetic diversity, develop molecular markers for diversity studies and to characterize the genes involved in salinity tolerance in *H. compressa* accessions from Kenya. Sampling of ethnobotanical data was done in four ASAL regions of Kenya; Turkana, Tharaka Nithi, Kwale and Tana River using the snowball technique. Interview schedules were administered to informants to determine the domestication status, utilization and the biotic and abiotic stresses impacting its growth: Responses from 79 respondents were evaluated. To ascertain the morphological variability of the vegetative and fruit features of *H. compressa* and to define its morphotypes, 90 *H. compressa* accessions were further analyzed. For morphological diversity, a total of 19 morphological characters, including seven quantitative and twelve qualitative aspects of fruit and vegetative attributes were used. Genotyping by sequencing (GBS) was done to determine the population structure and genetic diversity. Greenhouse experiments were conducted on accessions from Tharaka, Tana River and Turkana to evaluate salinity tolerance. Varying salinity levels of control, 100mM, 200mM and 300mM were imposed on the accessions for eight weeks. Morphological, physiological, proline content and ion content were determined. The RNA of the most tolerant accession was sequenced to identify differentially expressed genes (DEGs). Simple Sequence Repeats (SSR) markers were developed from unigenes obtained through RNA sequencing. Twenty of these primers were validated using 20 accessions from the four sampled regions. The results of this study showed that the sampled regions' levels of domestication varied, with the majority of the respondents exhibiting little interest in domesticating *H. compressa*. The study documented fourteen uses of *H. compressa* with food use (fruit) and soil erosion prevention (roots) scoring the greatest and least fidelity levels. Human intervention and pest infestation were the most prevalent biotic stresses, while salt and drought were the most prevalent abiotic stresses. All seven quantitative traits were highly effective at distinguishing doum palm phenotypes ($p < 0.001$). The 90 accessions belonged to five morphotypes, numbered 1, 2, 3, 4 and 5. Kwale accessions were heterogeneous. Genotyping by sequencing analysis on the other hand revealed two populations with high within-population diversity. Accessions from Turkana were grouped in one cluster while accessions from Tharaka, Tana River and Kwale accessions were grouped in another cluster. Moderate F_{ST} of 0.074 was obtained indicating moderate genetic differentiation in *H. compressa*. Tana River samples proved to be most tolerant to salinity stress followed by Turkana accessions. Tana River accessions also accumulated more proline and more biomass. Tana River control and salinity stressed samples at 300mM were sequenced. A total of 92,135 unigenes were obtained from *de_novo* assembly of the RNA data. In the current study, a total of 8611 DEGs were obtained with 3722 being up regulated and 4889 down regulated. A total of 25 gene ontology terms and 36 KEGG pathways were enriched. A total of 16,632 SSR markers

were mined from the 92,135 unigenes at the rate of 1SSR per 5.5kb of unigenes. The AG/CT SSR motifs were the most frequent trinucleotide motifs. The validated SSR markers amplified 55 alleles at the rate of 2.75 alleles per locus. The SSR markers revealed higher genetic diversity indices compared to Single Nucleotide Polymorphism (SNPs) data obtained through GBS approach. Further, the SSR markers clustered the accessions into three populations based on STRUCTURE analysis. The use of *H. compressa* as food can build resilience of pastoralist communities who are susceptible to famine during drought. Overall, the study designated *H. compressa* into five morphotypes and two genetic clusters. This study also demonstrated that *H. compressa* is moderately to highly tolerant to salinity. The most salinity tolerant accessions are those from Tana River. This study has described numerous salinity induced DEGs, enriched GO terms and KEGG pathways in *H. compressa* which have shed light on its tolerance mechanisms. The first ever functional SSR markers have been developed for this plant which are crucial for diversity studies, marker assisted breeding and should be validated in other members of the genus *Hyphaene* and related taxa.

CHAPTER ONE

INTRODUCTION

1.1 General background information

The Arecaceae (palm) family has over 183 genera and 2500 species (Govaerts & Dransfield, 2005). It is divided into five subfamilies; *Calamoideae, Nypoideae, Coryphoideae, Ceroxyloideae* and *Arecoideae* (Horn et al., 2009). In Africa, there are about 18 genera and 68 species of palms (Stauffer et al., 2017). Among the *Coryphoideae* subfamily, the genus *Hyphaene* is the most economically important genera albeit little is known about it (Stauffer et al., 2018). Additionally, the variety of the palm family in Africa is primarily due to this genus. *Hyphaene* palms belong to the tribe *Borasseae* and subtribe *Hyphaeninae* (Asmussen et al. 2006). One of the most common genera in Africa's Arid and Semi-arid Lands (ASALs) is the genus *Hyphaene*, often known as the "doum palms" (Stauffer et al., 2018). In comparison to other members of the palm family, the presence of stem branching in some species of this genus stands out as a distinctive characteristic (Stauffer et al., 2014). This genus has eight species; *Hyphaene compressa H. Wendl, H. coriacea Gaertn, H. guineensis Schumach. & Thonn., H. reptans Becc., H. macrosperma H. Wendl, H. thebaica (L.) Mart, H. dichotoma Furtado and H. petersiana Klotzsch ex Mart*. (Stauffer et al. 2014; Stauffer et al. 2018). The genus *Hyphaene* is of economic importance in Africa due to the variety of uses of the plants which include food, medicine, woven products and construction materials as demonstrated by several studies (Amwatta, 2004; Kahn & Luxereau, 2008; Aboshora et al., 2014, Martins & Shackleton, 2019; Nyambe et al., 2019; Omar et al., 2020).

Hyphaene compressa, is a palm typical in riverbeds and wadis in ASALs of East Africa (Vandenbeldt 1992; Orwa et al. 2009). In Swahili it is called Mkoma, Pokot (Tangayween), Kamba (Mukoma), Samburu (Iparwa), Turkana (Eeng'ol) and in Tharaka it is called Muruguju (Orwa et al., 2009). This palm is found at altitudes between 0-1400m with a mean annual temperature and rainfall of >28˚C and of 100-600mm respectively (Orwa et al. 2009). Even with the extensive geographical

distribution of *H. compressa* in Africa and Kenya, there is little validated documentation on its uses in ASALs. Further, doum palm is acclimatized to ASALs and would be a model plant to be studied to support the communities in these areas. Unfortunately, no known research efforts are in place. In some regions, farmers have not fully exploited the doum palm resource and lack awareness on the benefits of the palm. Other constraints of doum palm production include the lack of quality and quantity planting material and the long time (2 to 3 weeks) it takes for the plant to germinate (Orwa et al., 2009). The dioecious nature of the plant and the long time it takes to fruit is a major hindrance to its production. Notwithstanding the wide geographical coverage, economic and sustenance role of this genus to the ASALs of Africa, the genus is not well appreciated, recognized or delineated (Stauffer et al. 2017; Stauffer et al. 2018). Knowledge on the utilization of plants and related by products are often passed on progressively to successive generations (Gadgil et al. 1993). Nevertheless, loss of such knowledge is aggravated by disinterest of the younger generation (Nolan & Turner 2011). Consequently, it is paramount to document and safeguard traditional knowledge about *H. compressa.*

Human intervention, abiotic and biotic stresses threaten *H. compressa* populations. Overgrazing by pastoralist populations, especially in the riverine zones, pose a serious threat to this palm (Kigomo, 2001). The sedentarization of the pastoralists has increased the pressure on *H. compressa* resources (Amwatta, 2004). Sedentarization causes pastoralists to congregate around scarce resources, which ultimately results in land degradation (Johnson, 1993). Overharvesting and harvesting of immature sword leaves are two additional factors putting strain on *H. compressa*. These leaf pressures have been shown to cause a sister palm, *H. thebaica,* to alter its arborescent habit into a sub-terrenian crawling habit (Kahn & Luxereau, 2008). Additionally, *H. compressa* is subject to selection pressure from burning, logging and destructive tapping of wine from the apical meristems. The loss of particular genotypes caused by selection pressure results in genetic drift, which may potentially impact the *H. compressa* gene pool (Kigomo, 2001).

Plants are generally sensitive to stress particularly drought, elevated temperature and salinity stress which limit their growth (Das, 2013; Hernandez, 2019). Stress affects the productivity of plants (Dubouzet et al., 2003). Soil salinity is common in ASALs which often receive periodic rainfall and have elevated temperatures. Soil salinity is therefore a significant threat to food security. *Hyphaene compressa* grows in the ASALs of Kenya (Amwatta, 2004; Maundu & Tengnas, 2005) which have saline soils with high ion toxicity (Mugai, 2004). In most of the ASALs of Kenya, particularly Turkana, *H. compressa* is the predominant vegetation. This makes it the ideal plant for studying salinity tolerance in this region. In addition, wild plants and wild relatives of crop plants have been shown to harbor potential salt tolerance genes (Hernandez, 2019). Furthermore, salinity tolerance assays are requisite for breeding salt tolerant crops (Yaish et al., 2017).

Plants have adapted various ways of coping with salinity stress including production of enzymatic and non-enzymatic antioxidants (Das, 2013). Salt tolerant crops have the ability to avoid salt toxicity by having low rates of Na^+ and Cl^- transport to the leaves. Plants can also compartmentalize the toxic ions in the vacuole (Munns, 2002). Gene expression studies using transcriptomics enables the evaluation of differentially expressed genes (DEGs). Identification of salinity tolerance genes would provide essential markers for breeding salt tolerant crop plants, as well as provide valuable knowledge on the molecular and genetic pathways involved in tolerance mechanisms (Hernandez, 2019). Expressed Sequence Tags (EST) Simple Sequence Repeats (SSR) markers can be developed from the transcribed areas of the genome (Taheri et al., 2018). SSR markers created from these regions are called genic SSRs or EST-SSRs. They are excellent genetic markers since they are created from the expressed regions of the genome, are numerous and co dominant (Zhang et al., 2019). Additionally, due to their conservative nature, they are transferrable to related species (Ellis & Burke, 2007).

1.2 Statement of the problem

The diversity of food crops in arid and semi-arid regions is limited and largely seasonal. This is due to unsuitability of land for farming as a result of high salinity

which makes the community in these areas food insecure. Doum palm is one of the few perennial plants vastly growing in these areas. Doum palm has not been domesticated despite its important nutritional role and a source of income for the local community where it is growing. Furthermore, there is no structured propagation and conservation mechanisms for doum palm accessions in these areas. Doum palm populations are however declining due to habitat degradation, drought and overharvesting, with little to no material for further research on two *Hyphaene* species, *H. macrosperma* and *H. reptans* (Stauffer et al. 2018). *Hyphaene compressa* is essential to the ASAL populations' livelihoods, but its acreage is unknown and information on its benefits to the Kenyan population is limited. This predisposes it to diversity loss and loss of individual genes. *Hyphaene compressa* is a non-model plant that lacks model descriptors, making diversity and breeding programs difficult. The long maturation period, inability to distinguish between varieties, and inability to discriminate male from female trees are other limiting factors to improvement of agronomic traits in doum palm. In fact, many farmers find it difficult to identify cultivars outside of the fruiting season due to this genus's high adaptive flexibility. As a result, phenotype-based breeding would be exceedingly time-consuming in this plant

Doum palm has attracted very little research and hence scanty genotypic data exists for this plant. At the time of this investigation, there were no assembled genomes for *H. compressa* or any other *Hyphaene* palms. The absence of genetic information and widespread use, may exacerbate population decline and loss of genetic diversity. Moreover, breeding the plant for specific markers would be difficult. Soil salinity is a serious threat to food security. The insufficient rainfall in ASALs prevents the leaching out of soluble salts in the root zone which can lower crop output and even completely eliminate crops (Zaman et al., 2016). Doum palm is already adapted to ASALs characterized by such hot and dry environments and thrive in the harsher temperatures and in close proximity to the coast where the soils are extremely saline. In spite of this, it has received little focus. Instead, research has concentrated on alternative plant resources including cereals like sorghum, maize, millet and grain legumes. The salt tolerance mechanisms in this plant are lacking.

1.3 Justification of the study

Due to the scanty ethnobotanical information of *H. compressa*, a comprehensive *H. compressa* inventory of its utility is important which is a requisite for its improvement. It is also crucial to identify the unique phenotypic features for *H. compressa* that are necessary for differentiating this species. Marker assisted breeding would benefit from the creation of molecular markers, which are currently lacking in this plant.

Hyphaene compressa is already adapted to ASALs in Africa and Kenya. It therefore has the capacity to adapt to climate change and support Africa's ASALs' expanding population. The transcriptome sequencing obtained from this study will produce enormous amounts of doum palm transcript sequences for gene identification and genomic data resources, as well as understanding how the crucial doum palm products are made. Additionally, this will direct the choice of salt stress genes for overexpression and mutation. Breeding salt tolerant crops will alleviate food, energy and environmental problems facing the world today and in future. The SSR markers from transcriptome data will be useful for future diversity studies and fast track breeding programs in this species and related taxa.

The information gathered from this study will therefore serve as a crucial genomic resource for the palms in the *Hyphaene* genus and related palms. This is the first attempt to genotype *H. compressa* using genome wide SNPs. This study is anticipated to make important contributions to the conservation, management and exploitation of this species. The government and other key stakeholders will benefit greatly from being informed of the significance of the palm as one of the wild edible plants heavily relied upon in ASALs of Kenya. The objective of this study was to document the ethnobotanical knowledge, determine the morphological and genetic diversity, develop molecular markers for diversity studies and to characterize the genes involved in salinity tolerance in *H. compressa* accessions in Kenya.

1.4 Null hypotheses

1. Doum palm is not domesticated, has no uses and is unaffected by biotic and abiotic stresses in the ASALs of Kenya

2. Doum palm is not clustered into morphotypes
3. The *H. compressa* accessions obtained from the ASALs of Kenya are not structured into different populations and are not genetically diverse
4. Doum palm accessions from the ASALs of Kenya are not tolerant to salinity stress and there is no differential gene expression between salt stressed and non-salt stressed doum palm
5. Doum palm EST-SSR markers cannot be developed from doum palm RNA seq data and cannot be validated and used in diversity studies.

1.5 General objective

To determine the ethnobotany, diversity and salinity induced transcriptomics in doum palm (*H. compressa*)

1.5.1 Specific objectives

1. To document the ethnobotanical knowledge, domestication status, biotic and abiotic stress of *H. compressa* in Kenya.
2. To determine the morphological diversity of *H. compressa* in Kenya
3. To determine the genetic diversity and population structure of *H. compressa*.
4. To determine the salinity tolerance and to characterize genes that are down or up regulated due to salinity stress in *H. compressa*
5. To develop and validate EST-SSR markers for genetic studies in *H. compressa*.

CHAPTER TWO

LITERATURE REVIEW

2.1 Wild and semi wild edible plants

Plants that are locally accessible in their native habitat and are not domesticated or cultivated are considered wild food plants (Chakravarty et al., 2016). They are almost entirely found in natural woodlands and range lands (Kidane & Kejela, 2021). Wild and semi wild plants are consumed as vegetables, fruits, herbs, additives, juice or medicine (Laibuni et al., 2020; Nyero et al., 2021; Suwardi et al., 2020). They are also used for recreation purposes (Tebkew et al., 2014). The wild and semi wild plants are important in rural areas in many parts of the world as a source of food especially to the women and children who are the most vulnerable (Fentahun & Hager, 2009; Kidane & Kejela, 2021). Such plants are accessible all year round and hence are buffers during times of famine (Chakravarty et al., 2016; Fentahun & Hager, 2009; Tebkew et al., 2014). Besides food security, wild plants are sources of income to local populations in rural areas (Ngome et al., 2017; Suwardi et al., 2020). The importance of wild plants has been demonstrated by several studies in rural populations of Ethiopia (Fentahun & Hager, 2009; Melaku & Ebrahim, 2021; Tebkew et al., 2014), Indonesia (Suwardi et al., 2020), Congo (Ngome et al., 2017), Uganda (Nyero et al., 2021) and Kenya (Sarfo et al., 2020; Shumsky et al., 2014).

Nutrition and phytochemical analysis of some wild plants indicate that most of them have higher protein, fat content and antioxidants (Bvenura & Sivakumar, 2017; Chakravarty et al., 2016; Fentahun & Hager, 2009). Despite this, wild plants are poorly adopted in diets (Fentahun & Hager, 2009; Ngome et al., 2017). Evidence suggests that consumption of wild plants is declining which also threatens the indigenous knowledge connected to them (Kidane & Kejela, 2021). This decline is associated with several factors including economic, acceptability, neglect and favoritism for contemporary agricultural crops (Bvenura & Sivakumar, 2017; Fentahun & Hager, 2009; Ngome et al., 2017). Moreover, there is little knowledge on the harmful effects and nutritive value of wild plants compared to their exotic

counterparts that have been extensively studied (Bvenura & Sivakumar, 2017; Kidane & Kejela, 2021). It is crucial to quantify and standardize the uses of wild edible plants in order to promote their usage for diet diversification and enrichment. (Ngome et al., 2017).

Rural populations are often faced with scarcity of food due to crop failure. High crop failure in ASALs is a result of the increased frequency and intensity of droughts and floods (Ondiko & Karanja, 2021). Wild fruits provide room for diversification of food sources and improvement of nutrition (Chakravarty et al., 2016; Fentahun & Hager, 2009). Compared to their domesticated relatives, wild plants are more resistant to water stress (Bvenura & Sivakumar, 2017; Fentahun & Hager, 2009). Management of wild plants by local communities which has been demonstrated in Ethiopia ensures optimum harvests (Tebkew et al., 2014). Dependency on wild plants all year round has been shown to be common in resource poor households of Kenya (Shumsky et al., 2014). Wild plants have been shown to be nutrient rich therefore reduce the cost of food in resource poor households in Kenya (Laibuni et al., 2020), Turkana (Sarfo et al., 2020), Tharaka Nithi (Shumsky et al., 2014) and Western Kenya (Ekesa, 2009). In Kenya's ASALs, *H. compressa* is one of the wild edible plants with a variety of uses (Amwatta, 2004; Maundu et al., 1999).

2.2 *Hyphaene* genus

Hyphaene genus belongs to the coryphoid subfamily in the family arecaceae. This placement has been supported by phylogenetic studies (Stauffer et al., 2018). Molecular phylogenetic studies on the palm family have been done using plastid DNA (Asmussen et al., 2006; Asmussen & Chase, 2001). The relationship between and among species of *Hyphaene* however, remains sketchy. The *Hyphaene* genus has highly polymorphic fruits which have led to increase in synonyms for some species in this genus. For example, *H. compressa* has 33 synonyms which necessitates revision of the taxonomy of this genus (Stauffer et al., 2018). The *Hyphaene* project (www.hyphaene.org) initiated to define *Hyphaene* phylogenetics, is expected to enrich *Hyphaene* systematics and evolution (Stauffer et al., 2018).

The *Hyphaene* project (www.hyphaene.org) in the year 2022, released preliminary results on phylogenetic relationships of *Hyphaene* genus. Based on this, only five species out of the eight are recognized. Interestingly, these results also show that *H. dichotoma* which is endemic in Asia is not genetically distinct from *H. compressa*. The other four species are*; H. coriacea, H. thebaica, H. petersiana* and *H. guineensis* (www.hyphaene.org). Two *Hyphaene* species*; H. macrosperma and H. reptans* have minimal or non-existent material that can aid further studies (Stauffer et al., 2018). *Hyphaene reptans* is found in Kenya, Somalia and Yemen (Palmweb, 2023; Stauffer et al., 2014). *Hyphaene macrosperma* which is thought to be a lesser-known species of *Hyphaene*, has been recently regarded as a morphotype of *H. compressa* (Stauffer et al., 2017). Field studies carried out in the *Hyphaene* project revealed diverse distribution of the *Hyphaene* species in Africa. The species *H. thebaica* and *H. guineensis* are predominant in West Africa, *H. coriacea* and *H. petersiana* in South Africa while *H. compressa* and *H. coriacea* are common in East Africa (Stauffer et al., 2018).

2.2.1 Uses of palms in the genus *Hyphaene*

One of the three most widely used plant families is the Arecaceae, which includes palm trees (Balslev et al. 2016, Lee and Balick 2008). Non-timber items made from palms ensure that local communities thrive by providing food, medicine, building materials and a source of revenue through the sale of woven leaf products (Maca et al. 2011, Paniagua-Zambrana et al. 2007). Palm trees are significant both in rural settings and worldwide (Balslev et al. 2016). For instance, in Western Africa, non-timber products made from palms are substantially more valued (Stauffer et al. 2017).

Among the most economically important palm genera is the *Hyphaene* genus with almost every species having diverse uses (Stauffer et al., 2018). In Mozambique the most important plant part of *H. coriaceae* is the leaf which is used in making beverages, utensils, basketry, furniture, construction and as medicine (Martins & Shackleton, 2019). Ethno-veterinary uses of plants in the *Hyphaene* genus have been demonstrated. For example, kernels of *H. petersiana* have been used to treat dog lung disease in Namibia (Cheikhyoussef & Embashu, 2013). In Ethiopia, sheaths from *H.*

thebaica are used to treat eye problems in livestock (Mosissa et al., 2021). Ethno pharmacological uses of *Hyphaene* genus have also been documented. *Hyphaene petersiana* leaves and roots are used in Namibia for treatment of wounds and dry cough respectively (Cheikhyoussef & Embashu, 2013). Recently, leaf extracts of *H. compressa* were reported to inhibit α-Glucosidase enzyme that is associated with type 2 diabetes and ultimately reduce blood sugar (Khallaf et al., 2022). Zinc Oxide nanoparticles obtained from *H. thebaica* fruits in Egypt, showed therapeutic efficacy, antimicrobial potential and inhibition of *Leshmania tropica* (Mohamed et al., 2020). *Hyphaene thebaica* fruits have been confirmed to have antimicrobial properties (Omar et al., 2020).

Food uses and the nutritional value of the species in the *Hyphaene* genus has been reported across Sub Saharan Africa. Significant levels of ions, protein, fat and carbohydrates have been detected in *H. guineensis* in Congo (Ossoko et al., 2019). *Hyphaene thebaica* fruits were reported to have high radical scavenging activity hence they are not only nutritious but also healthier (Omar et al., 2020). *Hyphaene petersiana* fruits in Namibia were found to have high mineral concentrations and low anti-nutrient contents (Nyambe et al., 2019). The traditional usage of the *Hyphaene* species in the treatment of several ailments is supported by the presence of antioxidants (Omar et al., 2020). *Hyphaene* species are widely distributed in ASALs of Africa where they act as important forage for livestock (Stauffer et al., 2018). For example *H. thebaica* is used in Ethiopia as feed for cattle and donkeys (Mosissa et al., 2021). In the Congo, fruits of *H. guineensis* are eaten by apes, elephants and buffalos (Van Valkenburg & Dransfield, 2004). Communities in *Hyphaene* habitats use leaves to make baskets among other handiworks (Amwatta, 2004; DeMotts, 2017; Martins & Shackleton, 2019; Stauffer et al., 2018; Van Valkenburg & Dransfield, 2004). In some of these regions for example in South Africa, the products are heavily commercialized (Stauffer et al., 2018) making them important sources of revenue (Amwatta, 2004; DeMotts, 2017; Martins & Shackleton, 2019).

In Kenya, *H. compressa* is utilized for ornamental purposes, food and medicine (Amwatta 2004). In Turkana, *H. compressa* had the highest relative importance value

of 21.4 among all the other traditional wild plant species studied by Ejore et al. (2020). In the same study, *H. compressa* was widely used for thatching and repairing houses. Kenya's pastoralists and agro-pastoralists overly depend on the selling of woven handicrafts (Amwatta 2004). Basketry from *H. compressa* leaves has been shown to be the most income generating activity for the residents of Turkana (Akall, 2021) while the young leaves and shoots are utilized as vegetables and cattle feed during drought (Maundu & Tengnas 2005).

The Turkana pastoralists produce a concoction of powdered doum palm fruit and cow blood which they refer to as "Lokot". This is made when the nomads are preparing to move to migrate to a new grazing or pasture area. The pastoralists are sustained on this concoction during the journey (Maundu & Tengnass 2005). For animals like donkeys, camels and goats, doum palm fruits are an essential source of food. It has been established that doum palm fruits are rich in important minerals like potassium, sodium, calcium, magnesium, phosphorus and possess nutritional qualities that can benefit diabetics and persons living with hypertension (Aboshora, et al., 2014). Many native farmers of doum palms, particularly those in West Pokot, Kenya, use doum roots as a form of medicine to treat bilharzia (Orwa et al., 2009). Since the doum palm grows in nearly dry climates, food crops are not in competition with it there.

2.3 Hyphaene compressa

Hyphaene compressa is a wild edible palm common in East Africa and grows in sandy lowlands, open forests, dry areas and alluvial flats with a high water table (Kahn & Luxereau, 2008; Stauffer et al., 2014). In Kenya *H. compressa* is distributed in Tana River (Omari et al., 2019), Coastal and Northern Kenya (Amwatta, 2004; Maundu & Tengnas, 2005). *H. compressa* is an economic resource for the agro-pastoralists and pastoralists of Kenya. The pastoralist women utilize the leaves for various handicrafts for a sustainable source of income (Amwatta, 2004).

2.3.1 Conservation and domestication status of *H. compressa*

Hyphaene species are protected in five *ex situ* conservation sites and six protected areas worldwide (BGCI, 2020; Cosiaux et al., 2017). The largest *Hyphaene* collection

worldwide (92 collections) is held at the Royal Botanic Gardens, Kew (Stauffer et al., 2018). However, it is difficult to determine the *in situ* conservation status of doum palms in Africa (Johnson, 1998). The IUCN red list classifies the *H. compressa* as a species of Least concern (Cosiaux et al., 2017). This indicates that it still has a vast geographic range and is not currently under threat. This ranking may be a result of the weak and insufficient conservation efforts in Africa. Nonetheless, field studies of *Hyphaene* distribution in Africa have confirmed that there has been a gradual decline in population (Stauffer et al., 2018). The only Asian doum palm, *H. dichotoma* has been classified as near threatened by the IUCN. The population of this species has declined due to anthropogenic activities in India (Qureshimatva et al., 2018). In Namibia, *H. petersiana* is among the high priority species for conservation (Cheikhyoussef & Embashu, 2013). In Sudan, *H. thebaica* was declared threatened by the Sudan Ministry of Agriculture (Warrag et al., 2002). In Turkana, the invasive *Prosopis juliflora* has displaced indigenous woodland vegetation in the area including *H.compressa* (Akall, 2021). Ultimately support for further conservation efforts for this genus may be called for in future.

Efforts to domesticate wild plants is paramount in improving resilience of communities in marginal areas to food insecurity, providing access to wild edible plants and alleviating the pressure on the natural ecosystem (Bvenura & Sivakumar, 2017; Melaku & Ebrahim, 2021). Indeed, several species might become extinct if left uncultivated or domesticated (Melaku & Ebrahim, 2021). Even though it can be difficult, encouraging farmers to grow and protect wild plants in their farms is a significant step towards domestication of wild edible plants (Tebkew et al., 2018). It has been established that the doum palm was a significant holy fruit in ancient Egypt (Clement 1992; Janick 2014). It was discovered in pharaohs' tombs by archaeologists (Hamdy & Fahmy 2018). Despite this predynastic existence in Egypt, there have been little attempts to domesticate *Hyphaene* genera. Several studies on germination and seedling growth of *H. thebaica* (Idohou et al., 2015) and *H. compressa* (Stave et al., 2006) have been done. Such studies provide useful information for the development of domestication strategies, conservation and restoration initiatives. In Northern Africa, many crops are cultivated beneath *H. thebaica* and *P. dactylifera* canopies

thereby practicing some form of forest farming and ultimately conserving the palms (Pfadenhauer & Klotzli, 2020). Other palms such as date (*Phoenix dactylifera L.)*, coconut (*Cocos nucifera* L.), peach palm (*Bactris gasipaes Kunth*) and oil palm (*Elaeis guineensis Jacq*.) have been fully domesticated (Clement et al., 2017; Gros-Balthazard & Flowers, 2021; Nayar, 2017; Wales & Blackman, 2017). The restoration of deteriorated ecosystems and prevention of dwindling biodiversity are aided by cultivation and management of indigenous wild food plants (Melaku & Ebrahim, 2021). For example, in Northern Kenya community groups have planted various wild plants including *Hyphaene spp, Azadirachta indica, Commiphora Africana, Suaeda monoica* among others to stabilize the mobile sand dunes in an effort to combat desertification. In addition to the environmental management provided by the wild plants, the communities stand to benefit economically from the wild plants (Olukoye & Kinyamario, 2009).

2.3.2 Ethnobotanical studies

Ethnobotanical knowledge on the utility of plants and their products particularly wild plants is often passed from generation to generation (Gadgil et al. 1993). Information on wild plants may be passed on to successive generations through folklore, songs or orally (Kidane & Kejela, 2021; Wanjohi et al., 2020). Wild plant exploitation and preservation are threatened by the informed but aging population and the youth's lack of interest in wild flora. The lack of interest in wild plants is attributed to modernization (Nolan & Turner 2011). People may give up their traditional knowledge because they believe it will not equip them to deal with the challenges of modern life (Reyes-garcía et al., 2013). For instance, in Kenya , which is undergoing rapid cultural changes, traditional knowledge on wild edible plants which is orally passed on may be lost (Wanjohi et al., 2020). Anthropogenic activities such as deforestation also leads to loss of wild plants and ultimately the loss of knowledge associated with such plants. Therefore, it is crucial to carefully record traditional knowledge on wild edible plants in order to conserve it.

Numerous ethnobotanical studies have been done to document various palm uses across the world. For instance, date palm, *Borassus aethiopium*, oil palm, *Phoenix*

reclinata, *H. coriaceae* and *Hyphaene spp* (Martins & Shackleton, 2019; Panigua-Zambrana et al., 2007; Reddy et al., 2019; Sadeghi & Kuhestani, 2014; Salako et al., 2018; Stauffer et al., 2018). Diverse uses of *H. compressa* leaves in Kenya have been documented (Amwatta, 2004). However, ethnobotanical uses of other plant parts have not been documented.

2.3.3 Indices used in ethnobotany studies

The accurate documentation of ethnobotany is essential for the *in situ* preservation of crops and wild plant genetic resources. A range of methodologies, including descriptive research, cultural consensus analysis, participatory methods and hypothesis testing, are used to examine the importance of plants in ethnobotany (Stepp 2005). Indicators have been developed, such as the Relative Cultural Importance Index (RCI), for measuring and assessing hypotheses regarding ethnobotanical data (Hoffman & Gallaher 2007) as shown in Appendix I. The RCI indices have been employed by numerous ethnobotanical studies to quantify ethnobotanical data. (Altaf et al., 2019; Bhattarai, 2018; Fathir et al., 2021; Jadid et al., 2020; Mudzengi et al., 2017; Salako et al., 2018; Tardío & Pardo-De-Santayana, 2008). The most common index used is the fidelity level (FL) which is the relative frequency of mention of a use, that is, the number of times a use-report is cited (Hoffman & Gallaher, 2007; Salako et al., 2018).

2.4 Morphological diversity of plants

The outward physical attributes of a plant constitute its morphology. Morphological studies entail a thorough examination of vegetative and reproductive traits of plants in order to create a profile of the plant that can be used to identify varieties or to make morphological comparisons within the species (Wyatt, 2016). Even with the availability of next generation sequencing (NGS) platforms, plant morphology still remains useful (Wiens, 2004). Nearly all areas of plant biology, including ecology, genetics, evolutionary biology, physiology and systematics, still find morphological investigations to be important (Simpson, 2019). It is crucial to recognize that there is still alot to be done before fully characterizing and even sequencing all of the living species on Earth. Therefore, morphological diversity studies and systematics are key

to understanding poorly circumscribed plant groups (Saraswati & Srinivasan, 2015). It is possible to assess and distinguish plants using the variability of morphological features (Govindaraj et al. 2015; Haider et al. 2015). The first step in classifying, utilizing and ultimately conserving genetic resources is phenotypic evaluation (Zou et al. 2020).

2.4.1 Morphological studies on the genus *Hyphaene*

Hyphaene species are dioecious (Stauffer et al., 2014). The sex ratio in *H. thebaica* has been shown to be 0.5 (Rodrigue Idohou et al., 2016). However, monoecious trees also exist, although they bear small, infertile fruits (Orwa et al., 2009). Monoecious *H. compressa* individuals have been reported in Djibouti (Stauffer et al., 2018). The genus *Hyphaene* contains species ranging in size from tiny to huge and tall, basally clustered or solitary, erect or creeping stems with costapalmate leaves (Stauffer et al., 2014). The fruits of the species in this genus are strikingly polymorphic (Orwa et al., 2009; Stauffer et al., 2014; Stauffer et al., 2018). In comparison to other members of Arecaceae, the dichotomous stem branching in some species of this genus stands out as a distinctive characteristic (Orwa et al., 2009; Stauffer et al., 2014; Tomlinson & Huggett, 2012).

Several species of *Hyphaene*: *H. guineensis, H. compressa and H. petersiana*, have sharply defined characters that are easily identified (Stauffer et al., 2018). Several studies have been done on the morphological diversity of the doum palms. One such study was done using fruit traits of *H. thebaica* in Egypt. The results revealed significant variation in the accessions (Khalil et al., 2020). Fruit traits in *H. thebaica* have also been studied in Benin, West Africa (Idohou et al., 2015). The *Hyphaene* project has also elucidated the morphological differences of the eight species in the genus *Hyphaene*. These differences are summarized in Table 2.1.

2.4.2 Morphology of *H. compressa*

The leaves of *H. compressa* are fan-shaped and costapalmate. The leaves have complete borders and the leaf stalk is covered with curled thorns (Orwa et al., 2009; Stauffer et al., 2014). Typically, the palm is 10–17 meters tall (Orwa et al., 2009).

Table 2.1: Morphological diversity of the species in the genus *Hyphaene*

	Stem			Fruit		Leaves
	height	**Width**	**Habit**	**size**	**color**	**Number**
H. compressa	20m	40cm	Forking	12 cm long 9 cm wide	Orange brown	15
H. coriaceae	5m	25cm	Cluster	6cm long 4cm wide	Dark brown	8-15
H. thebaica	20m	15-20 cm	Forking 2 or more	4-6cm long 4-5 cm wide	Brown to dark brown	8-12
H. guineensis	14m	30-35cm	Forking 2 or more	6-8cm long 6-7cm wide	Dark red-brown	14
H. petersiana	20m	35cm	solitary	5-8cm long 5-6cm wide	Rich red brown to chestnut	20-25
H. dichotoma	15-20m	25-30cm	Forking 2-3 stems	4.5cm long 4cm wide	Dull brown	14-15
H. macrosperma	No further data on this species			7cm long 6cm wide		
H. reptans	Imperfectly studied species, It appears to be unique among *Hyphaene* in its crawling behavior					

This information has been extracted from https://www.hyphaene.org/index.php/species-and-synonyms

This palm is characterized by dichotomizing trunks where the upper branches are supported by a large stem at the base (Tomlinson & Huggett, 2012). The genus *Hyphaene* has this distinctive trait and they can build up to 16 crowns (Orwa et al., 2009; Stauffer et al., 2014). The inflorescence of this dioecious palm is the same in both sexes. Fruits that range in color from brown to orange are produced by female plants which produce clusters of 1 to 200 irregularly shaped, glossy orange-brown fruits (Kahn & Luxereau, 2008; Orwa et al., 2009; Stauffer et al., 2014). Doum palms grow relatively slowly. After planting, it takes the seeds around three weeks to root (Kahn & Luxereau, 2008). The fan shaped leaves of the palm appear two to three years after germination and the trunk appears after 18 to 20 years (Orwa et al., 2009).

2.4.3 Use of morphological descriptors

Despite being simple to analyze, morphological markers can be challenging to score since they are unpredictable, unstable, sluggish and some take longer to manifest on

the plant than others (Andersen & Lubberstedt, 2003; Mokhtar et al., 2016). A descriptor is a collection of uniform traits used to assess and categorize certain group of plants (Rizk and El Sharabasy, 2006). Descriptors support genebanks, *in-situ* and *ex-situ* conservation efforts as well as diversity studies (Rizk & El Sharabasy, 2006). The IPGRI website has descriptors for various other palms, including sago palm, *C. nucifera,* peach palm and *P. dactylifera*. Since limited data exist on the morphological descriptors of *Hyphaene genus,* date palm descriptors as outlined by Rizk & El Sharabasy (2006) are useful for diversity studies in the members of this genus.

Due to the dioecious nature of *Hyphaene* species (Orwa et al., 2009; Tomlinson & Huggett, 2012), a farmer, for example, might not be able to distinguish a female palm from a male palm until they reach adulthood, mature, and flower. Fortunately, several genetic markers have been developed for similar dioecious plants that can discriminate male and female palms before flowering. Most of these markers are SSR markers and Sequence characterized amplified regions (SCAR) markers (Al-Mahmoud et al., 2012; Al-Qurainy et al., 2018; Elmeer & Mattat, 2012; Maryam et al., 2016; Wang et al., 2020; Yousif et al., 2020). All these markers have been validated in date palm.

2.5 Plant genetic diversity studies using molecular markers

It is insufficient to evaluate plant diversity using a single marker, such as morphology (Khan et al., 2012). Molecular markers are DNA sequences that detect variability between the nucleotide sequences of various individuals (Nadeem et al., 2018). Variation may result from chromosomal deletions, duplications, inversions and insertions. Molecular markers and the phenotypic expression of a genetic characteristic may or may not agree (Govindaraj et al., 2015). Molecular and morphological markers are usually independently discriminative enough to delineate accessions (Pocovi et al., 2020). However, when used together, they become a powerful tool for phylogenetic reconstruction (Wiens, 2004). Three categories of molecular markers; hybridization-based, PCR-based and DNA sequence-based have been established (Govindaraj et al., 2015).

2.5.1 Hybridization based markers

Restriction Fragment Length Polymorphisms (RFLPs) and minisatellites are hybridization based markers. They have the advantage of being codominant and can detect unlimited number of loci (Govindaraj et al., 2015; Schlötterer, 2004). The RFLPs involve the digestion of DNA with restriction enzymes and the variation in length of fragments produced after hybridization to specific markers is noted. However, they are expensive, time consuming and require high molecular weight genomic DNA (Grover & Sharma, 2016). The RFLP markers have been used previously to genotype oil palm using 40 RFLP probes and cDNA RFLP probes (Barcelos et al., 2002; Mayes et al., 2000; Singh et al., 2008). An RFLP genetic map has also been constructed for oil palm using RFLP probes (Mayes et al., 1997). Coconut has also been genotyped using RFLP markers (Lebrun et al., 1999). However, none of the species in the genus *Hyphaene* have been genotyped using RFLP technique.

2.5.2 PCR based markers

PCR based molecular markers are quick to perform and require much less starting material for DNA extraction (Govindaraj et al., 2015). Randomly amplified polymorphic DNA (RAPDs) were the first PCR based markers. They are quick to perform and simple but they are not reproducible (Govindaraj et al., 2015). This method employs the use of a single 10 base random primers for amplification (Grover & Sharma, 2016). The RAPD markers have been employed in several palm species diversity studies including oil palm (Sathish & Mohankumar, 2007; Thawaro & Te-Chato, 2008) and coconut (Rajesh et al., 2014). Amplified fragment Length polymorphism (AFLP) use both PCR and RFLP where PCR products are restricted at specific recognition sites (Govindaraj et al., 2015). They are highly reproducible. Some of the palms that have been genotyped using AFLP include; date palm, sago palm and oil palm (Nisar & Hussain, 2022; Sabir et al., 2014; Ying et al., 2007)

Other genetic markers include, SSR and inter simple sequence repeats (ISSR) and EST SSR. Simple sequence repeats have been widely used to genotype palms either as genomic SSRs (Okoye et al., 2020; Purwoko et al., 2019; Salomon-Torres et al., 2017)

or EST SSRs (Bazzo et al., 2018; Ting et al., 2010; Zhao et al., 2013). Inter simple sequence repeats are amplified by PCR using a single long microsatellite primer (Grover & Sharma, 2016). They are highly reproducible. The SCAR markers are created via cloning, sequencing and creating lengthy primers that are complementary to the ends of amplified DNA fragments. Sequence characterized amplified regions have been used for various studies in plants including aquatic pteridophyte *Azolla*, *Physalis* genus, *Moringa oleifera*, *Commiphora wightii* and *C. myrrha* (Abraham et al., 2013; Feng et al., 2018; Ravi et al., 2021; Sairkar et al., 2016). The SCAR markers have also been used for gender identification in date palms seedlings (Al-Qurainy et al., 2018; Dhawan et al., 2013). Resistance gene analogue polymorphism (RGAP) amplifies plant resistance genes and their analogs using degenerate primers (Grover & Sharma, 2016). Various studies in plants have utilized RGAP markers (Mutlu et al., 2006; Shan et al., 2010; Sharma & Tamta, 2017; Yan et al., 2003).

2.5.3 Sequence based markers

Sequence-based markers are molecular markers that depend on the discovery of a certain DNA sequence in a pool of unidentified DNA (Nadeem et al., 2018). Identification of novel genomic markers has been aided by the development of NGS techniques like genotyping by sequencing (GBS) and diversity arrays Technology (DArT) which are sequence targeted techniques (Grover & Sharma, 2016). Through NGS analysis, various sequence based markers can be identified including; single nucleotide polymorphisms (SNPs) and SSR markers (Nadeem et al., 2018). Single nucleotide polymorphisms refer to changes in a single base in DNA relative to the DNA base that is expected at that loci (Nadeem et al., 2018). They are important for diversity studies and generating linkage maps (Egan, et al., 2012). Studies that have used SNP markers for palm diversity include (Bai et al., 2018; Cros et al., 2017; Klimova et al., 2018; Ong et al., 2018; Pootakham et al., 2015; Santos et al., 2020). In the study by Bai et al. (2018), RAD-seq was used to identify genome wide SNPs that were then used for linkage map construction in *E. guineensis.* The RAD-seq derived SNPs have also been characterized in the dwarf coconut for determination of population structure and genetic diversity (Santos et al., 2020). Over 5000 SNPs derived from GBS analysis have been used for genomic selection of yield components

and other studies in *E. guineensis* (Cros et al., 2017). In another study, SNPs were used for phylogenetic analysis to resolve taxonomic ambiguities among two palm genera*; Washingtonia and Brahea*. This analysis further demonstrated the value of GBS in elucidating strikingly distinct patterns of genome wide variation originating from multiple effects (Klimova et al., 2018). Single nucleotide polymorphisms have been used for identification of quantitative trait loci associated with agronomic traits in *E. guineensis* (Pootakham et al., 2015). They were also used to identify linkage disequilibrium patterns in *Camelia sinensis* (Xia et al., 2019). DArT and SNP markers have been used to genotype *E. guineensis* populations in Malaysia and to develop a high density DArT genetic map (Gan et al., 2018). Diversity arrays technology generates whole genome sequences by scoring for presence or absence of fragment without relying on any DNA sequence information (Grover & Sharma, 2016).

2.5.4 Genetic diversity studies in the order coryphoideae and the genus *Hyphaene*

The genetic data in the order coryphoideae is scanty. The species *P. dactylifera* is the only palm with an assembled annotated genome in this order (Hazzouri et al., 2019). One of the most economically important plants in the arecaceae family is *P. dactylifera*. Numerous beneficial items are produced by date palms for human consumption. This could be the motivation for the genetic diversity studies in this palm. Moreover, Genome Wide Association Studies (GWAS) have been done on the fruit traits and the sex determining region (Hazzouri et al., 2019). Most of the coryphoid palms are not economically significant, hence, little effort has been made to sequence them or understand their diversity.

According to NCBI gene database, the order coryphoideae has approximately 40416 genes. Out of these, only 133 genes have been deposited in the sub tribe *Hyphaeninae*. These 133 gene sequences are all from one species *Bismarckia nobilis*. However, several palm molecular systematics studies have used *Hyphaene* species. For example, *H. petersiana* and *H. coriaceae* among other plant species were used in a DNA bar coding experiment to determine the most suitable regions for DNA bar coding. Among the barcodes used were acetyl-COA carboxylase dehydrogenase *(accD),* ribulose-1,5-biphosphate carboxylase/oxygenase large sub unit *(rbCL),* RNA polymerase beta sub

unit *(rpoB),* maturase K *(mat K)* and NADH dehydrogenase sub unit *(ndhJ)* (Lahaye et al., 2008). *Hyphaene coriaceae* has also been used for molecular phylogenetic analysis of the palm family using rbCL and 18S nr DNA sequences (Hahn, 2002). Coding and non-coding plastid DNA was used for phylogenetic studies of palms including *H. thebaica* (Asmussen & Chase, 2001). The nucleotide sequences of all the above and other chloroplast genes are available at the NCBI Nr nucleotide database. The species of *Hyphaene* with nucleotide sequences at the NCBI are; *H. coriaceae* (69), *H. dichotoma* (7), *H. guineensis* (4), *H. petersiana* (33) and *H. thebaica* (28). However, there are no nucleotide sequences for *H. compressa* accessions at the NCBI nucleotide sequences database.

The two species; *H. thebaica* and *H. coriaceae* have been sequenced using NGS approaches in two separate studies. Leaf sample of *H. thebaica* was sequenced using illumina Miseq target sequence capture of exonic genes. These sequences together with sequences obtained from other plants were used for phylogenomic relationships and historical biogeography of South African palms (Escobar et al., 2022). Elsewhere, *H. coriaceae* is one of the species that has been used in constructing the tree of life by KEW, Royal Botanical garden https://treeoflife.kew.org/. Here, *H. coriaceae* was sequenced using illumina MiSeq platform and the sequences are available at NCBI Short Read Archive (SRA) database.

Overall, studies that target *Hyphaene* species diversity or population structure are scanty. One such study has used 10 SSR markers to genotype *H. thebaica* accessions in Saudi Arabia. The results indicated low to moderate levels of diversity with low gene flow (Hassan Mansour, 2021). In another study, twelve *H. thebaica* accessions in Egypt were genotyped using Inter Simple Sequence Repeats (ISSR) and Start Codon Targeted (SCoT) polymorphism markers (Khalil et al., 2020). *Hyphaene compressa* is poorly studied both morphologically and genetically.

2.5.5 Genotyping by sequencing

Whole genome sequencing would be the most appropriate for diversity studies in *H. compressa* because it is a non-model plant with little genetic information and no reference genome at the time of this study. Whole genome sequencing is nevertheless

costly. There are alternative, more affordable sequencing techniques available that can capture large amounts of data on portions of the genome (Wallace & Mitchell, 2017). Such methods include Restriction-Site-Associated DNA Sequencing (RAD-seq) and GBS. RAD-seq is suitable for organisms with large genomes, it is quick, specific and highly reproducible (Elshire et al., 2011). Genotyping by Sequencing (GBS) on the other hand is a variation of the RAD-Seq approach (Wallace & Mitchell, 2017).

Genotyping by sequencing is a special application of NGS technique that can be used to find and genotype SNPs in genomes and populations. The GBS is a straightforward, highly multiplexed approach for building reduced representation libraries for the Illumina NGS platform (Elshire et al., 2011). This approach allows high throughput identification of plant molecular markers at extremely low costs (Scheben et al., 2017). Reduced libraries for NGS platforms are done by restriction enzymes which target the genome. A single restriction enzyme, a barcoded adaptor and a common adapter were all utilized in the initial GBS protocol (Elshire et al., 2011). Different approaches have been developed since the inception of GBS some of which use two restriction enzymes (Poland et al., 2012). Lower copy regions can be effectively targeted by using the right endonuclease (Burghardt et al., 2017). The reduced subsets can then be ligated with DNA barcoded adapters, amplified using PCR and then the genomic subsets can be sequenced using high-throughput NGS technology on a single lane of flow cells (Burghardt, et al., 2017; Elshire et al., 2011; He et al., 2014). Due to the restriction of the genome, the genome can only be partially sequenced (Wallace & Mitchell, 2017).

Genotyping by sequencing is quick, simple and very reproducible (Burghardt et al., 2017; Davey et al., 2011). These characteristics make GBS appealing for a variety of genetic applications, such as phylogeny, genomic wide association studies, genomic selection, genetic diversity, physical and linkage maps (Burghardt et al., 2017). With the ability to detect SNPs, deletions, insertions and microsatellites, GBS is the ideal instrument for genetic diversity studies (Elshire et al., 2011).

Genotyping by sequencing analysis has been done for several plant species to determine the population structure, genetic diversity or linkage disequilibrium by use

of SNPs. For example, GBS analysis has been done on both cultivated and landraces of wheat (Alemu et al., 2020; Alipour et al., 2017; Yang et al., 2020), cultivated *Vigna unguiculata* from Zambia and Malawi (Nkhoma et al., 2020) and *Camellia sinensis* species (Niu et al., 2019). Genetic diversity studies have also been done on wild and semi-domesticated plants using GBS; *Sempervivum tectorum* (Fabritzek et al., 2021), wild species of *Dioscorea dumetorum* (Siadjeu et al., 2018), *Rhododendron canescens* (Yadav et al., 2019) and wild common bean (Cortés & Blair, 2018). Several plant species in the palm family have also been genotyped using GBS. These include; *E. guineensis* (Babu et al., 2019; Cros et al., 2017; Osorio-Guarín et al., 2019; Pootakham et al., 2015), coconut (Santos et al., 2020), *Washingtonia* and *Brahea* genera (Klimova et al., 2018). The closest relative of doum palms that has been studied using GBS is date palm (Mathew et al., 2015; Thareja et al., 2018).

There are two methods for detecting SNPs in GBS data, that is: *de_novo* assembly for non-model plants and reference based assembly for plants with reference genomes. At the time of this study, however, neither the doum palm nor any other members of the genus *Hyphaene* had assembled genomes. There were no conspecific (of the same species) or congeneric (of the same genus) reference genomes for *H. compressa*. The *de_novo* assembly of GBS reads is the best choice because doum palm is not a model plant. Congeneric and confamilial reference genomes can be used for SNP discovery and estimation of genetic diversity (Galla et al., 2019). In the palm family the *P. dactylifera* genome has been made available (Al-dous et al., 2011) and may be used as a confamilial reference genome for *H. compressa*.

2.6 Salinity induced transcriptomics in doum palm

2.6.1 Salinity tolerance in plants

Plant salt tolerance is the ability of plants to endure the impact of excessive salt concentrations in the root zone or surface without suffering severe damage (Hussein et al., 2011). Soil is regarded as saline when it has an Electrical Conductivity (EC) of 4dS/m which is equivalent to 40mM NaCl (Blake & Munns, 2017). Plants that can withstand high salt concentrations are called halophytes (Parida & Jha, 2010). Abiotic

stresses such as salinity, drought, erosion, freezing and heat have adverse effects on the productivity and quality of plants due to reduced photosynthesis (Dubouzet et al., 2003). In response to these stresses, plants respond at molecular, cellular, biochemical and physiological levels (do Amaral et al., 2016; Dubouzet et al., 2003).

Dominant ions in saline water are Na^- and Cl^- (Hussein et al., 2011). High salinity causes ion toxicity and osmotic stress thereby limiting the ability of the plants to take up water (Das, 2013; do Amaral et al., 2016). This leads to loss of turgor pressure, closure of stomata, reduced transpiration and photosynthesis as well as the accumulation of reactive oxygen species (ROS). Consequently, this leads to increased energy requirements for osmoregulation (Hussein et al., 2011). Roots regulate their growth under high salinity while cells in elongation enter into a quiescent stage (Feng et al., 2018). Some plants have developed specialized organs to enable them thrive under high salinity (Das, 2013). Salinity has a direct impact on the cell wall properties. Cell wall integrity pathways sense changes in the cell wall properties due to abiotic stresses. These pathways include plasma membrane localized receptors like kinases (RLKs) such as FERONIA (FER)- related malectin-domain containing THESUS 1 and ANXUR ½ (Feng et al., 2018). Tolerance to salinity depends on the genetic characteristics of the plant, salinity levels, soil characteristics and climate (Hussein et al., 2011).

2.6.2 Salinity induced response in plants

Rapid, almost immediate changes in growth rates occur in leaves in response to sudden salinity changes. Because the first decline in growth happens so fast and only temporarily and recovery happens so quickly, it must only be caused by changes in cell water relations and not salt specific responses (Munns, 2002). Reduced leaf and root growth is seen days after salinity treatment, but older leaves begin to die weeks after salinity treatment (Munns, 2002). Upon salinity stress induction, stress responsive mechanisms are activated to protect and repair damaged parts of the cell (Wang et al., 2003). Salinity stress leads to formation of ROS in cells. To counter act the effects of ROS which are detrimental to the cell membranes, plants have evolved numerous anti-oxidation strategies (Wang et al., 2003). Examples of these anti-

oxidants involved in the process include super oxide dismutase (SOD), catalases (CAT), ascorbate peroxidase (APX), ascorbate, carotenoids and glutathione (Das, 2013; Wang et al., 2003). Super oxide dismutase (SOD), is the first antioxidant against stress in plants (Das, 2013). The total SOD or the different SOD isoforms before and after stress treatments are important during transcriptome studies. The CAT and APX enzymes and their isoforms are involved in the removal of hydrogen peroxide from cells (Das, 2013). Plants also accumulate late embryogenesis abundant (LEA) proteins and heat shock proteins (Hsps) during stress as molecular chaperones (Wang et al., 2003). Amino acid proline has a very positive impact on plants under diverse stress circumstances. Its accumulation in stressed plants is beneficial (Alhasnawi, 2019). Proline serves as a superior osmolyte and also performs three other crucial functions under stress, including metal chelation, antioxidant defense and signaling (Das & Roychoudhury, 2014; Hayat et al., 2012).

2.6.3 Salt tolerance in *Hyphaene*

Salinity tolerance in the date palm which is a close relative of *Hyphaene* has been extensively demonstrated by several studies. Previous studies indicate that date palm is both drought and salinity tolerant (Kharusi et al., 2017; Yaish & Kumar, 2015). In a recent study, date palm accessions were exposed to between 200 to 1000mM salinity level for 13 days. In this study the *P. dactylifera* seedlings exposed to between 200 and 400mM NaCl had no changes while yellowish leaves were observed in seedlings exposed to 1000mM salinity level (Xu et al., 2022). Tolerance levels in date palm has been reported to vary among the date palm varieties with the *Khalas* variety being the most tolerant (Kharusi et al., 2017; Yaish & Kumar, 2015). Limited studies on salinity tolerance of the genus *Hyphaene* have been documented. In Okavango delta of Botswana, *H. petersiana* was shown to be abundant at higher ground water salinity levels of up to a maximum of 16.48mS/cm (McCarthy et al., 1994). *Hyphaene thebaica* has also been shown to tolerate medium to high salinity during germination and growth (Ali, 2007; Bezona et al., 2009).

Kenyan saline soils are categorized into four zones, each with a different level of ion toxicity and salinity: Areas along the coast, in Turkana, Mandera-Wajir and the

Southern Rift Valley (Mugai, 2004). *Hyphaene compressa* is predominant in the coastal areas and Turkana area (Maundu & Tengnas, 2005) which have been reported to be extremely saline (Mugai, 2004). The presence of *H. compressa* in the Serengeti savannah of Kenya is an indication of high salt in the soils (Bui, 2013). *H. compressa* tolerates moderate to high salinity stress (Orwa et al., 2009).

2.6.4 Transcriptomic studies on salinity

Transcriptomics is the study of all the RNA transcripts (transcriptome). Transcriptomic studies reveal changes in expression levels of different genes during exposure to different environments. The main aim of transcriptomics is to determine all the transcripts in a sample which include messenger RNAs (mRNA), non-coding RNAs and small RNAs (Imadi et al., 2015). In transcriptome analysis, only the transcribed genes are studied (Brautigam & Gowik, 2010). There are two techniques used in transcriptomic studies; microarrays (hybridization based) and RNA-seq (Imadi et al., 2015; Lo et al., 2012; Mortazavi et al., 2008). Most studies on responses to salinity or drought have employed transcriptomic analyses (Deyholos, 2010).

Transcriptomics analyses on salinity have been carried out in many plants including; cotton, mangroves, bread wheat, *Arabidopsis spp* and barley (Amirbakhtiar et al., 2019; Krishnamurthy et al., 2017b; Matsui et al., 2008; Ouertani et al., 2021; Zhang et al., 2021). Several transcriptomic studies have been done to determine salinity induced DEGs in certain palm species. For example, salinity induced DEGs was determined for oil palm *(E. guineensis)* which led to the identification of seven highly expressed genes during salt stress (Ferreira et al., 2021). Salinity induced DEGs have been studied in *P. dactylifera* roots exposed to 150mM salinity for two to four hours (Radwan et al., 2015), leaves and roots exposed to 300mM salinity level for ten days (Yaish et al., 2017) and leaves exposed to up to 1000mM salinity level for 13 days (Xu et al., 2022). However, no documented transcriptomic research in the genus *Hyphaene* had been documented at the time of this study.

2.6.5 RNA-Seq and the merits of transcriptomics

RNA seq is the method of choice for most transcriptomic studies. In RNA-Seq, RNA is extracted from samples followed by cDNA synthesis and finally sequencing of the cDNA strands (Deyholos, 2010). RNA Seq sequences all RNA transcripts by showing the expressed sequences at a particular time (Egan et al., 2012). RNA-seq generates millions of short reads that can be assembled and annotated to provide useful information on transcription and gene expression without having to sequence the whole genome (Lim et al., 2017). Pure samples and high read number are desired for transcriptome analysis (Brautigam & Gowik, 2010; Mortazavi et al., 2008). The presence and amount of each RNA in the sample can be quantified using RNA seq (Mortazavi et al., 2008). Integration of metabolomics and transcriptomics can be used to draw inferences on stress related gene expression (Deyholos, 2010).

The strength of detecting any given rare transcript with RNA seq is higher compared to using microarrays since it would be difficult to distinguish fluorescence from a low abundance positive in microarrays (Mortazavi et al., 2008). RNA-Seq has greater sensitivity and increased resolution to discriminate alleles and isoforms. It can also detect any molecule present in the sample unlike microarrays which require specific probes (Deyholos, 2010). By using transcriptomics, the non-coding sections are removed thereby reducing the size of the library and avoiding nonfunctional genes unlike metagenomics where the total DNA has a huge library size that contains nonfunctional genes which have to be processed (Sturmberger et al., 2016). The spatio-temporal gene expression patterns can also be determined using transcriptomics (Sturmberger et al., 2016). Transcriptomics is ideal for non-model plants like doum palm since no reference genome is needed (Egan et al., 2012). Transcriptomics is able to discover genes that cannot be easily identified phenotypically (Deyholos, 2010).

2.7 Molecular markers used for genetic diversity studies in plants

2.7.1 SSR markers

Simple Sequence repeats (SSRs) are also known as Short Tandem Repeats (STRs), microsatellites or Simple Sequence Length Polymorphisms (SSLPs) (Egan et al.,

2012; Govindaraj et al., 2015). They are repeating DNA sequences of 1 to 6 nucleotides that occur in a genome (Egan et al., 2012; Grover & Sharma, 2016). Other literature define SSRs as 1 to 10 bp (Govindaraj et al., 2015). Short Tandem Repeats are highly variable and evenly distributed throughout the genome with simple repeats of 2 ,3 or 4 nucleotides (Govindaraj et al., 2015). Primers are designed to amplify a region flanking the microsatellite (Grover & Sharma, 2016). Among all the available markers microsatellites are useful for a number of applications in plant genetics including Quantitative Trait Loci (QTL) mapping, gene flow, linkage map development, genetic diversity studies, marker assisted selection (MAS), germplasm evaluation and evolutionary studies (Egan et al., 2012; Govindaraj et al., 2015).

This is because they are easy to automate, universal, reliable, ubiquitous, highly polymorphic, multiallelic, codominant and require little DNA (Ghislain et al., 2004; Grover & Sharma, 2016; Schlötterer, 2004). The polymorphism patterns produced by SSRs are much more than any other marker (Grover & Sharma, 2016). The SSR markers are cost effective especially in studies involving no model organisms due to the high accuracy of its genetic map compared to other markers like SNPs (Schlötterer, 2004).

2.7.2 Detection of SSRs markers

Microsatellites were difficult to detect in the past, but the development of NGS has simplified this process (Grover & Sharma, 2016). Next Generation Sequencing approaches that are capable of detecting thousands of markers in any genome in a single step have emerged (Davey et al., 2011). It has also enabled the isolation of microsatellites even in non-model plants (Grover & Sharma, 2016). There are two types of SSRs which are based on their source; EST-SSRs and genomic SSRs (g-SSRs). Genomic SSRs can be detected from genomic sequences obtained from illumina sequencing. Transcribed regions of RNA sequences through transcriptomic studies allow for the identification of EST SSRs (Taheri et al., 2018). To detect g-SSRs, prior knowledge of the genomic sequences is required but the case is different when detecting SSRs in expressed sequence Tags (ESTs) (Durand et al., 2010; Grover & Sharma, 2016). Various tools exist for mining SSR markers from transcriptome

data. These tools include SSR Locator (Maia et al., 2008) and MISA tools (Beier et al., 2017).

Previous studies have suggested that EST SSRs produce diversity patterns of higher quality than genomic SSRs (Taheri et al., 2018). EST SSRs have proven more advantageous than g-SSRs due to the low cost involved in their discovery, higher genetic patterns and their high transferability to related organisms (Ellis & Burke, 2007). Transcriptome sequencing increases the number of EST-SSRs and this is important particularly for plants with limited genetic resources (Chen et al., 2015) like *H. compressa.* Several EST-SSRs have been developed and utilized in several plants including Adzuki bean, date palm, macauba palm, oil palm, olive, wild shrub *Stephanandra incisa*, *Camellia japonica* and cow peas (Arbeiter et al., 2017; Bazzo et al., 2018; Chen et al., 2015; Gupta & Gopalakrishna, 2010; Li et al., 2021; Ting et al., 2010; Zhang et al., 2021; Zhao et al., 2013).

CHAPTER THREE

MATERIALS AND METHODS

3.1 Study site

This study was conducted in the ASALs of Kenya between the months of January through September 2018. Four sampling locations (Figure 3.1), were chosen in this study.

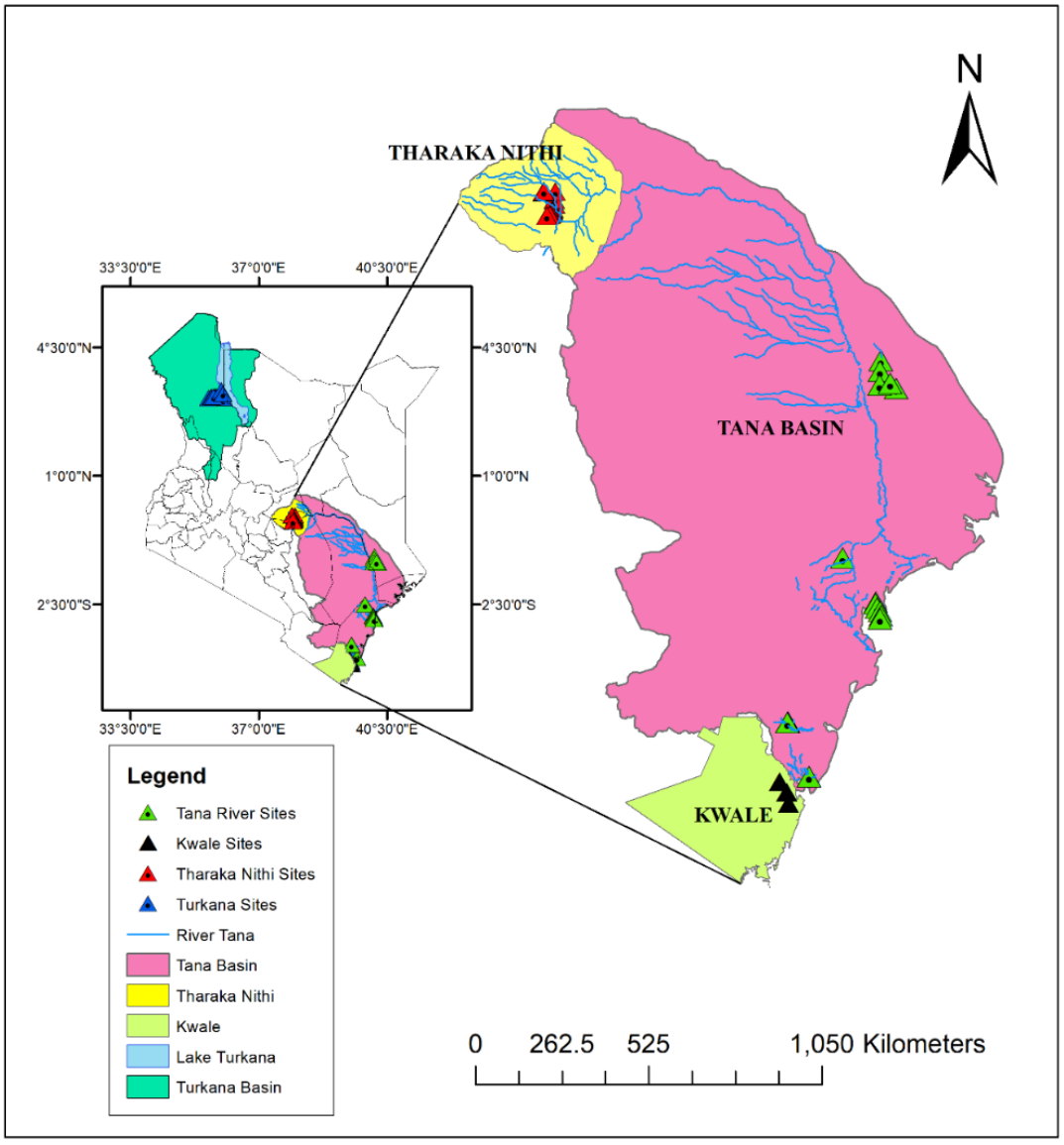


Figure 3.1: ASALs of Kenya where *H. compressa* accessions were sampled

The selection of the four sites was based on previous research which showed the distribution of *H. compressa* in these regions (Amwatta 2004, Maundu & Tengnas 2005). These sites were; Turkana (34°30′ and 36°40′ East and 1°30′ and 5°30′ North), Tharaka Nithi (37°19′ and 37°46′ East and 00°07′ and 00°26′ South), Kwale (38°31 and 39°31′ East and 30°3′ and 40°45′ South) and Tana River (38°25′ and 40°15′ East and 0°0′ and 2°0′ South). These regions are located between agro-ecological zones v and vii, which are distinguished by erratic rainfall, elevated temperatures and scanty vegetation (Jaetzold & Schmidt 2009). Prior to sample collection, pilot studies were carried out in these areas to confirm the presence of the plant in the areas and whether they were in the fruiting season.

There are 365,330 people living in the about 2610 km^2 Tharaka Nithi County in Kenya. It receives 500 to 800 mm of rain annually and experiences temperatures between 29°C and 36°C. The Aatharaka ethnic group is the main ethnic group living in Tharaka. The Tharaka Nithi people are farmers who focus on food crops such cowpeas, pigeon peas, green grams, millet and sorghum. Additionally, they raise dairy cows, goats, native zebu cattle, sheep and chickens (Icheria 2015). River Tana and its tributaries, which include the rivers Kithinu, Thingithu, Mutonga, Ura, Thanantu, Thangatha and Kathita, are the primary sources of water for irrigation (Integrated Plan, 2017). The abundance of River Tana tributaries encourages fishing activities. *Vachellia* spp. dominate the vegetative cover, as well as a few kinds of drought-tolerant grass (Gioto 2018). On the mainland and along the rivers in the area, there are isolated populations of the *H. compressa* species.

Turkana is an ASAL county in the Northern part of Kenya. With a surface area of 77,000 km^2 with Lake Turkana and the River Turkwell as its year-round water sources. High temperatures are prevalent in this county ranging from 20 to 41 degrees Celsius. This county also experiences erratic rainfall (200 millimeters on average) and has a population of 855,399 inhabitants. Turkana is home to the Pokot, Turkana, Samburu Rendile and Elmolo ethnic groups (*Turkana County Integrated Development Plan 2013-2017*, 2013). The Turkana are a pastoralist minority group in Kenya, which leaves them susceptible to enduring poverty (Ng'asike & Swadener, 2015). The

County is regarded as having a food shortage and its primary food sources are wild fruits, animals and emergency food (Kenya Inter Agency Rapid assessment, 2014). Zebu cattle, goats, donkeys and camels are kept by the residents for milk, meat, riches and dowry payments. Fishing is done by locals near Lake Turkana. Since *H. compressa* is the primary plant in this region, they are regarded as forests of Turkana (Turkana County Integrated Development Plan 2013-2017, 2013). Other plants growing include *Osyris lanceolata*, *Aloe* spp and *Vachellia* spp.

Amidst being home to the most significant coastal and deltaic ecosystem in East Africa and having a wide variety of animals and plants, Tana River County is categorized as an ASAL (https://www.tanariver.go.ke/environmental-wildlife-and-natural-resources/). The County experiences rainfall that varies from 400 mm to 750 mm, with an average temperature of 23° C and a maximum of 33° C (Kenya inter-Agency rapid Assessment 2014). It is only here can one find the unusual mangabey monkeys, whose primary food is the *H. compressa* fruits (Maundu & Tengnas 2005). There are 240,735 people living in this County, which covers approximately 35,375.8 km^2. The people of the Tana River are farmers who focus on food crops such maize, cashew nuts, mangoes, green grams, cassava and coconut. They also raise cattle and goats (MoALF 2016).

As a region with severe food insecurity, the County's citizens engage in meal skipping and commercial weaving for sale to manage their stress related to food (Kenya inter-Agency rapid Assessment 2014). The Pokomo, who are farmers, along with the Orma and the Wardey, who are pastoralists, make up the majority of the ethnic subgroups. Additional tribes include the hunter-gatherer Waata and Boni, Wailwana, Malakote and Bajuni (http://www.tanariver.go.ke/about-us-2/). Kwale has a size of 8270 km^2 and a population of 649,931 people. The three main economic areas are agriculture, tourism and fishing (Kenya inter-Agency rapid Assessment 2014). The Mijikenda are the primary settlers in this County. They cultivate mangoes, coconuts, cashew nuts, beans, cassava and maize. Additionally, the Mijikenda raise sheep, goats and zebu cattle (Kenya inter-Agency rapid Assessment 2014). Marshy grass, *C. nucifera, H.*

compressa and Offshore mangrove forests along the Indian Ocean make up the majority of the vegetative cover (Lang'at, 2008).

3.2 Ethnobotany, domestication, biotic and abiotic stress of *H. compressa*

The purpose of this study was to determine people's management methods and viewpoints regarding the domestication status, abiotic and biotic stresses, significant usage categories and the value of *H. compressa* plant parts. Indicators known as Relative Cultural Importance (RCI) indices were employed to quantify ethnobotanical data (Hoffman & Gallaher 2007). The following RC1 indices were used in this study (Hoffman and Gallaher 2007).

a) **Reported use (RU)** which is the total number of uses reported for the plant.
b) **Reported Use Value Per Plant Part** (**RU Plant Part)** which is the total number of uses for each plant part.
c) **Plant Part Value (PPV) in percentage (Percentage PPV)** which is the ratio between the total reported uses for each plant part and the total number of reported uses for the plant, that is

$$PPV\% = \frac{\Sigma \text{ RU per plant part}}{\Sigma \text{ RU}} x\ 100$$

Where,

RU is reported use

d) **Fidelity Level (FL) in percentage** which is the frequency with which a use report was mentioned (Salako et al. 2018).

$$FL = \frac{x}{n} x100$$

Where,

x is the total number of informants who have mentioned a specific use

n is the total number of informants.

3.2.1 Sampling procedure

Research permit and approval was granted by the Kenya National Commission for Science, Technology and innovation (NACOSTI). A prior informed consent was presented to the respondents stating clearly the objectives of the research and the potential impacts or output of the research. Those who consented to the research were the only ones allowed to participate. They were also informed that they had the right to refuse to be interviewed at any stage of the interview. Signed Informed consent for publication was also obtained for individuals whose images were captured during the survey.

The criteria used to determine the lead informants varied by region. Field staff from various organizations assisted with the identification of the lead informants within their jurisdiction and interview administration particularly when the informants were not able to independently fill out the form, were illiterate or needed language translation. The non-profit Organization (NGO) Anglican Development Services assisted in Turkana, officers from the Kenya Nuts and Oil Crops Directorate assisted in Tana River and Kwale while personnel from the Kenya Forestry Research Institute (KEFRI) handled it in Tharaka Nithi. The snowball method was used to choose the informants. Only established *H. compressa* product users were chosen, interviewed and provided lead to the next interviewee. Doum palm awareness was the lead question.

3.2.2 Data collection

To gather information on the four areas under investigation, oral and semi-structured interviews were conducted. The reliability and validity of the prepared questionnaire was checked by experts who reviewed the questionnaire for readability, clarity and comprehensiveness prior to being administered to the respondents. The interview schedule (Appendix II) utilized in this study was divided into four sections and included both closed and open-ended type of questions as shown in Table 3.1. A few open-ended questions were included in order to elicit unique and unexpected information regarding *H. compressa* ecology from the viewpoint of the respondents. Several closed-ended questions (dichotomous, Likert scale and fixed response) were

incorporated into the same interview schedule to help elicit more data regarding the study's goals. The researcher's perception of the evaluation and distribution of biotic and abiotic stress, as well as the domestication status of the palm, was also generated by observational questions (interviewer's viewpoint).

Table 3.1: Structure of the interview schedule used for sample collection

	Part of Questionnaire	Aspects interviewed
1	Demographic information	Name, gender, county, ethnic group of the informant
2	Distribution, domestication and maintenance	Number, distribution and seasonality, management practices, cropping systems and plants intercropped with *H. compressa*
3	Uses of doum palm	Plant parts uses (Leaves, stem, roots and fruits), special uses (rituals, medicinal, religious)
4	Biotic/ abiotic stress	Pests and diseases affecting doum palm production, abiotic stresses on *H. compressa*

3.2.3 Data analysis

All responses were coded and entered in SPSS. To facilitate data analysis, all weaving and related products were categorized as handicrafts for quantitative data analysis, while construction and related products were categorized as construction/building (Sadeghi & Kuhestani, 2014). Feeding was categorized as gathering doum palm products and feeding animals, as well as animals consuming doum palm in the field. For relevant categories, a univariate analysis was conducted. A data analysis plan based on the goals of the study was used. This method entailed first summarizing and outlining each response to each question. The analysis and presentation of data on category type and ordinal questions used percentages to show the frequency of responses. The relationship between two category types of questions was established using cross tabulation. Chi square was employed to test for independence of use levels across the various regions sampled.

3.3 Morphological diversity of doum palm

3.3.1 Sampling

Sampling was carried out in 2018, between January and July, the fruiting season for doum palm plants. The National Museums of Kenya's taxonomist assisted in identifying the doum palm in the field. The gender of the plant, tree maturity, general palm health and fruit quality were among the selection factors. For this objective, only fruiting palms were chosen. This is due to the difficulty in separating non-flowering males from non-fruiting females in natural populations. Furthermore, *H. compressa* contains limited descriptors that can be useful in studies of diversity; as a result, fruit features, which are absent in the male, are significant. Purposeful sampling was used to sample 30 doum palm trees from both Tana River and Kwale which were grouped as Coastal samples. Another 30 female trees from each of Turkana and Tharaka were chosen using purposeful sampling. This is because a minimum of 30 individual samples is needed in a population to achieve statistically significant estimates of genetic diversity (Miyamoto *et al.*, 2008; Sinclair & Hobbs, 2009). In the field, morphological evaluation was carried out using a set of descriptors as provided in Appendix III.

Accessions that were as far apart as possible from one another were sampled to lessen the likelihood of sampling close relatives. Ten fruits were chosen at random from each female tree that was sampled. The fruits were collected, labeled, put in bags and taken to the lab for morphological evaluation. Each tree's fruit harvest was combined and placed in a single bag for storage. Every fruit that was harvested was washed in sterile water and then spread out to dry in the sun (Okello et al. 2018). Assessment of fruit morphological descriptors followed afterwards (Table 3.2).

Some of the morphological descriptions were modified from a list of date palm descriptors (Rizk and El Sharabasy, 2006). Vernier calipers were used to measure the width and length of *H. compressa* fruits (Salako et al. 2019). An electronic weighing scale (Sartorius Entris 64-1S) was used to determine the weight of the fruit.

Table 3.2: Descriptors used for the morphological study of *H. compressa*

Plant Part	Quantitative descriptors	Qualitative descriptors
Whole plant*	Height (in m)	Trunk branching
Leaves	Length (cm), Breadth (cm), petiole length (cm),	Color (Dark green, green, light green), Mid rib color (Green, Yellow green) Pinnae density (very dense, dense, lax) Petiole color (Green with black stripes, green, light green, yellow with black stripes, brown with black stripes
Fruit	Length (cm), width (cm), weight (gm),	Shape (Round Oblong, obviate, ovate) Fruit apex (Truncate, Depressed), Fruit base (Truncate, Acute), Unripe fruit color (Green, maroon), Mature fruit color (reddish brown, brown, orange brown, orange

Reference (Rizk & El Sharabasy, 2006)

During field sampling, the stem and leaf morphology were evaluated. Typically, five fully formed *H. compressa* leaves were used to evaluate leaf morphological features (Rabei et al., 2012). Vegetative features that were both quantitative and qualitative were noted (Table 3.2). To describe the morphology of the plant's leaves, stem and fruits and to identify any differences, photographs of these parts were taken.

3.3.2 Data Analysis

For quantitative characteristics for each sampled region, the coefficient of variation, range and mean were computed. Additionally, the qualitative data frequencies were noted. To ascertain the variance among means of categories, analysis of variance (ANOVA) was calculated. The Games Howell Post Hoc Test was to uncover specific differences between group means. Data standardization was done in Excel where *Z*-scores were computed from raw scores using the mean and standard deviation. A linear mixed effects model was fitted to predict the fruit weight while the regions were used as random effects. All the quantitative variables were taken into account while fitting the model. The main axes of variation and the key variables in the data were determined using the principal component analysis (PCA). To estimate and characterize each population, discriminant analysis was utilized. Prior to doing the discriminant analysis, all of the quantitative data were standardized. Gower distance

was used for clustering using the PAM (Partitioning around Medoids) algorithm. The cluster analysis included both qualitative and quantitative data. The number of clusters was calculated using the silhouette coefficient. R software version 4.0.2 was used for statistical analysis of the morphological data.

3.4 Genetic diversity of *H. compressa* using GBS

3.4.1 Sample collection

The GBS approach was used to determine the genetic diversity of *H. compressa* accessions from the four different ASAL regions in Kenya. A total of 120 samples (30 samples from each of the four regions) were collected for GBS analysis. For morphological diversity, Coastal samples included both Tana River and Kwale accessions with each region having 15 samples. However, morphological analysis of the data showed that the two Coastal populations were different from one another. As a result, they were classified as two distinct populations, and an additional 15 samples from each of the two locations were included for the purpose of this objective.

Using sterile blades, leaf samples from the chosen plants were collected and put in sterile falcon tubes having 10g of silica gel (Chase & Hills, 1991; Moussoun et al, 2017). Accessions that were spread out as much as possible from one another were sampled. For additional processing, the samples were subsequently brought to the Institute for Biotechnology Research (IBR) at Jomo Kenyatta University of Agriculture and Technology.

3.4.2 DNA extraction

The DNeasy® Plant Mini Kit (Qiagen, Germany) was used for isolation of DNA as per the manufacturer's instructions. All centrifugation steps were done at room temperature (15–25°C). Buffer AW1 and Buffer AW2 concentrates were reconstituted by addition of ethanol. About 20 mg of silica dried doum palm leaves were ground in liquid nitrogen using a mortar and pestle to a fine powder. The powder was immediately placed in a 2ml tube. Immediately, 400 µl of Buffer AP1 and 4 µl RNase A were added. The mixture was then vortexed for 5 minutes and incubated for 15

minutes at 65°C in a water bath. The tubes were inverted every 5 minutes during incubation. After incubation, 130 µl of Buffer P3 was added. This was vortexed briefly to mix and then incubated for 5 minutes on ice. The lysate was centrifuged for 5 minutes at 11337 xg. The lysate was then pipetted into the QIAshredder spin column placed in a 2 ml collection tube. This was centrifuged at 11337 xg for two minutes. The flow through was transferred into a new tube. In case of pellet formation, the transfer of the flow through was done carefully not to interfere with the pellet. On the new tube with the lysate, 1.5 volumes of Buffer AW1 was added and mixed by pipetting and centrifuged at 4293 xg for 2 minutes.

The DNA spin column was placed in a 2 ml collection tube where 650 µl of the mixture was added and centrifuged at 4293 xg for 2 minutes. The flow through was discarded. The remaining mixture was also added to the DNeasy Mini spin column used in the previous step. A new collection tube was then added to the DNeasy Mini Spin column and 500 µl Buffer AW2 added and centrifuged at 4293 xg for 1 minute. The flow through was then discarded and another 500 µl Buffer AW2 added and centrifuged at 11337 xg for 2 minutes. The spin column was removed from the collection tube making sure that the spin column does not come into contact with the flow through to a new 1.5ml microcentrifuge tube. In the spin columns, 50 µl of Buffer AE was added incubated for 5 min at room temperature. This was then centrifuged for 1 minute at 4293 xg. Another 50 µl of Buffer AE was added to the spin column, incubated for 5 minutes at room temperature and centrifuged at 4293xg for 1 minute.

The integrity of the DNA was checked using 1% agarose gel electrophoresis at 150V for 40 minutes. The purity and quantity of the DNA were determined using Qubit fluorometer (Invitrogen) or microplate reader (DR-200B, Diatek).

3.4.3 GBS library preparation and sequencing

The DNA was sent to the Beijing Genomics Institute (BGI, Hong Kong). At the BGI institute, sample preparation, construction of the library and GBS sequencing was performed. Out of the 120 samples collected for GBS analysis, only 96 samples passed the sample QC and proceeded to sequencing. These included Kwale (28 accessions), Tharaka (27 accessions), Tana River (20 accessions) and Turkana (21 accessions).

Library preparation followed the protocol by Elshire et al. (2011). In summary, the DNA samples were barcoded before being plated with common adapter pairs and dried. After that, ApeK1 restriction enzyme was used to digest the samples (GCWGC as the recognition site). The ends of the genomic DNA fragments were then ligated with adapters. The next step was purification and pooling. To enrich the fragment pool, PCR was run using the suitable primers that had binding sites on the ligated adapters. This was followed by cleaning up of the PCR products. Using a Hiseq X10 platform, the DNA was sequenced as paired end 100 bp (Illumina PE 100) after the fragment sizes of the resulting library were verified on a DNA analyzer. The raw reads were purged of adapter sequences, sequences with poor read quality and sequences without barcodes. A total of 23.35Gb of clean data was obtained.

3.4.4 Alignment and variant calling

Fastqc and multiqc were used to check for the composition and quality of raw individual and multiple Fastq files respectively. Both the *de_novo* and reference-based techniques were used for assembly of the obtained sequences. In the *de_novo* assembly, ipyrad version 0.9.74 (Eaton & Overcast, 2020) was used to assemble sequences without a reference genome using the following parameters; assembly method *de_novo*, datatype pairgbs, mindepth_statistical 6, mindepth_majrule 6, min_samples_locus 4 and other parameters set to default. In the reference-based approach, paired read ends were mapped to the *Phoenix dactylifera* (date palm) genome (Hazzouri et al., 2019). SNP calling was performed using a confamilial (same family) reference genome (Galla et al., 2019) since *H. compressa* genome hadn't been fully assembled when this study was conducted. The genome of *P. dactylifera* was the only one that was accessible and closely related to *H. compressa*. These two palms are members of the Coryphoideae subfamily. Using the parameters for the Burrows-Wheeler Aligner (BWA) 'mem –t4 –k32 –M' (H. Li & Durbin, 2009), the sequence reads were aligned against the date palm reference genome. SAMTOOLS (mpileup and bcf tools) was used for SNP calling. SAMTOOLS mpileup command calculates the genotype likelihoods and the bcf tools does the SNP calling based on these likelihoods. The output of the bcf tools was a VCF file with sites and genotypes.

VCFtools version 0.1.16 (Danecek et al., 2011) was used for SNP filtering with the following parameters; biallelic SNPs, min meanDp 2, removing indels, Minor Allele Frequency (MAF) 0.05, minDP 2, max-missing 0.8.

3.4.5 Data analysis

Using the R program tidyverse (Wickham et al., 2019), the filtered VCF files' quality was evaluated. Using this program, the heterozygosity, read quality, read depth per site and read depth per accession were calculated.

3.4.6 Population structure analysis and genetic diversity

STRUCTURE software version 2.3.4 was used to determine the population structure using the admixture model (Pritchard et al., 2000). Populations of *K* (*K*=1 to 10) were run with three replications using a burn-in of 100000 generations and 100000 Markov Chain Monte Carlo (MCMC) iterations. The optimal *K* value using the *adhoc delta K* (Evanno et al., 2005) was determined using the software STRUCTURE HARVESTER http://taylor0.biology.ucla.edu/structureHarvester/ web Version 0.6.94 (Earl & VonHoldt, 2012). The R software POPHELPER version 2.3.1 (Francis, 2017) with a web interface was used to plot the structure results. Genotypes were assigned to Pure populations if they had membership proportions of ≥0.80 while those with less than this membership value were considered as admixed populations (Nkhoma et al., 2020).

The software adegenet version 2.1.3 (Jombart, 2008) in R was also utilized to assess the population structure of *H. compressa* using Discriminant Analysis of Principal Components (DAPC). A composite stacked bar plot with the likelihood of population membership on the Y-axis was created to show each sample's assignment. To ascertain the genetic linkages of *H. compressa* accessions, PCA was created using the R software package SNPrelate (Zheng et al., 2012).

Observed heterozygosity (H_o), Expected heterozygosity (H_e) fixation index (F_{ST}), inbreeding coefficient (F_{IS}), Analysis of Molecular Variance (AMOVA) and pairwise F_{ST} values of the population were determined using Arlequin version 3.5.2.2 (Excoffier & Lischer, 2010).

3.4.7 Phylogenetic analysis

The filtered VCF file was converted to a nexus file using the vcf2phylip.py script (Ortiz, 2019). The SplitsTree software Version 4.17.0 (Huson & Bryant, 2006) was used to create an unrooted splitstree using the neighbor net approach. The genetic clustering of *H. compressa* accessions was also represented by an UPGMA distance tree built with R software.

3.4.8 Migration rates of accessions along the River Tana basin

Using MIGRATE-n software version 3.6.11, gene flow was assessed to ascertain whether the population structure seen along the River Tana basin is influenced by seed dispersal along the River Tana. Constant mutation rates were employed across all loci in a Bayesian inference approach. At each locus, burn in was set at 5,000 iterations. The MCMC searches were enhanced by static heating at 4 distinct temperatures (1, 1.5, 3 and 6). One gene flow model, straight migration from Tharaka to Tana River and Kwale was assumed. This model was created based on River Tana basin's drainage system. Turkana accessions were not included in this model since PCA and structure analysis revealed minimal historical gene flow. Histograms and the effective population sizes were examined in order to determine whether the runs had converged on good conclusions.

3.5 Salinity induced transcriptomics

3.5.1 Plant material collection and germination

Seeds of doum palm were collected from three regions of Kenya: Tharaka (Eastern Kenya), Turkana (Northern Kenya) and Tana River (Coastal area). The seeds were sorted and only medium sized seeds were selected for germination (Negrão et al., 2017). Doum palm seeds were washed with running water. Mechanical scarification using a scalpel was done to remove dormancy (Plate 3.1A, B). This is because doum palms have impermeable hard pericarps that hinder their germination (Maundu et al., 1999; Moussa et al., 1998). The seeds were then soaked for three days in sterile distilled water (Moussa et al., 1998), then planted in sand nurseries (Maundu et al., 1999) in a greenhouse (Plate 3.1C). Doum palm plumules emerge only when the

radicle (Plate 3.1D) has attained an average depth of 60cm (Angaine, 2005). They were later transplanted in pots (Plate 3.1E).

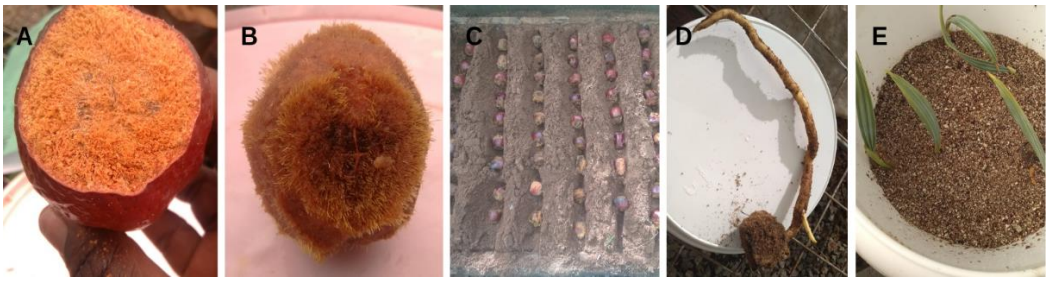


Plate 3.1: Planting doum palm seeds in the greenhouse

A. Mechanical scarification of doum palm seed, **B**. Doum palm scarified and ready for planting, **C**. doum seeds planted on a row on a sand nursery, **D**. Elongated Doum palm radicle, **E**. Doum palm seedlings transplanted in pots

3.5.2 Growth of *H. compressa*

After 3 weeks, the germinated seedlings were carefully transferred into 20 liter pots measuring 35.5cm in diameter and 32.5cm in height containing 15kgs of sterilized vermiculite. Due to doum palm germination habit of developing very long tap roots, 20 liter pots were used. The growth of *H. compressa* seedlings was done in the greenhouse under natural photoperiod with minimum and maximum temperatures of 31.5°C and of 39.2°C respectively. The potted plants were maintained by irrigation twice a week to field capacity with Hoagland's solution (Appendix IV).

Field capacity was determined using the method described by Vieira et al. (2020) where briefly, 10 samples containing 15kg of vermiculite (this is the quantity used for planting the seedlings) were placed in 20L pots with 2 draining holes at the bottom. This was weighed to obtain the fresh weight (FW). Water was then added slowly at the top until it started to drain through the holes. The pots were then covered with aluminium foil to prevent surface evaporation. The pots were kept under drainage until no more water came out to allow the total exhaustion of gravitational water. The vermiculite was then weighed to obtain weight at filled capacity (WFC). The amount of water needed for vermiculite to reach the maximum water holding capacity was the difference between WFC and FW. Watering to field capacity using the calculated

volume was maintained throughout the experimental set up in order to avoid draining which is a potential source of contamination.

3.5.3 Experimental conditions of *H. compressa*

The plants were maintained in the greenhouse until they had attained an average of 4 leaves (2-and-a-half-year-old doum palm). At this stage the plants were subjected to salinity stress. This study used the vegetative and non-reproductive stage of doum palm plants (Al-Abdoulhadi et al., 2011). A completely randomized block design was used blocked by region (Tharaka, Turkana and Tana River), with four treatments per region (0mM, 100mM, 200mM and 300mM NaCl) and a treatment size of 5 (Al-Abdoulhadi et al., 2011). Kwale accessions were excluded from the salinity assays due to the high heterogeneity (particularly the polymorphic nature of fruits) observed with accessions from this region as well as the informants' observations of varietal differences. For the salinity stress, full strength hoaglands solution was salinized with 0mM, 100mM, 200mM and 300mM NaCl which represent the control, medium, high and very high salinity levels (Al-Abdoulhadi et al., 2011; Cai & Gao, 2020). Sodium chloride was gradually increased by 50mM every 2 days until the desired concentration was attained to prevent osmotic shock (Negrão et al., 2017; Zhang et al., 2020). The desired concentrations per treatment were achieved on day 11 and gradual increments of 50mM was stopped (Table 3.3).

Table 3.3: Salinity treatment schedule showing gradual increments of NaCl

Week	Day	TREATMENTS			
		Control (5reps)	100mM (5reps)	200Mm (5 reps)	300Mm (5 reps)
1	1	0	50	50	50
	3	0	100	100	100
	5	0	100	150	150
2	7	0	100	200	200
	9	0	100	200	250
	11	0	100	200	300
3-8	13-56	0	100	200	300

After Day 11, the desired concentration of NaCl per treatment was achieved with the respective salinity levels.

Salt stress was imposed for 8 weeks. Longer days of salinity treatment was used based on a study that stated that salt specific differentially expressed genes are induced several days or weeks after treatment (Munns, 2002).

3.5.4 Morphological measurements

The morphological measurements documented during the course of the experiment included; chlorotic score, shoot growth, length of leaves and biomass. Morphological assessment was recorded twice (at the beginning of the experiment and at the end of the experiment) for all morphological measurements except biomass. Chlorotic score was done by visual assessment of the plants using a modified scale by Al-Abdoulhadi et al., (2011) of 1 to 5. Where 1=Healthy plant (No visible symptoms) as shown in Plate 3.2A, 2 slight damage to the plant of 1 to 2 % (Plate 3.2B), 3=moderate damage of 5-15% (Plate 3.2C), 4=more than 15% of the plant is affected (Plate 3.2D) and 5=plant death.

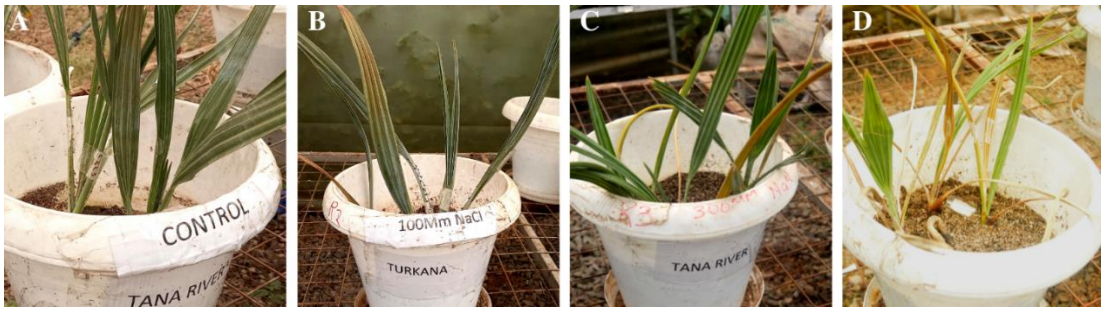


Plate 3.2: Chlorotic effects of salinity stress on *H. compressa* accessions

A. score of 1 meaning healthy plant**, B.** score of 2 meaning slight damage to the plant, **C.** score of 3 meaning moderate damage to the plant and **D.** score of 4 meaning more than 15% damage to the plant.

Shoot length and length of leaves was determined using a ruler. The shoot length was represented as the distance from the crown to the leaf tip (Tao et al., 2021). The length of leaves was measured for three categories of leaves as described by Al-Abdoulhadi et al. (2011) whereby the base leaf, the middle leaf and the upper leaf were measured.

At the end of the experiment, destructive harvesting was used to capture the biomass. Where briefly, whole plant, root and shoot fresh weight (FW) was immediately measured using an electronic balance (Cai & Gao, 2020). To obtain the dry weight (DW) of the plants, the plants were oven dried at 70°C for 48 hours until constant weight. The dry weight was determined for the roots and shoots.

3.5.5 Physiological measurements

The following physiological measurements were recorded; plant water content, chlorophyll content, SPAD readings and stomatal density. The plant water content was determined for each plant using the following formulae by Cai & Gao (2020):

$$Water\ content = \frac{FW - DW}{FW} X100$$

Fully expanded young leaves were used for leaf chlorophyll extraction. The chlorophylls were extracted using dimethyl sulphoxide (DMSO) according to the method described by Parry et al. (2014). Briefly, chlorophyll measurements were taken at midday to avoid effects of light intensity. Leaf disks were obtained from the leaves using a cork borer (Number 4, 8.9 mm diameter) and immediately placed in Eppendorf tubes containing 2ml of DMSO. The Eppendorfs were then incubated at 65^{o}C for an hour in an oven. The Eppendorfs were then kept in the dark for 3 days (Time it took for the disks to be transparent). After this, 200 µl aliquot was used for spectrophotometric readings at 665nm (chlorophyll a), 649 nm (chlorophyll b) and 480nm for carotenoids (Wellburn, 1994) using the following equations:

$C_a = 12.47\ A_{665.1} - 3.62A_{649.1}$
$C_b = 25.06A_{649.1} - 6.5A_{665.1}$
$C_{x+c} = (1000A_{480} - 1.29C_a - 53.78C_b)/220$

Where,

A is absorption at the specific wavelength

C_a is concentration of chlorophyll a,

C_b is concentration of chlorophyll b and

C_{x+c} is the concentration of total carotenoids.

The SPAD value of fresh leaves were determined twice during the course of the experiment (At week 6 and 8) using a SPAD chlorophyll meter (SPAD-502; Konica, Minolta sensing, Inc., Japan). Three point readings were taken evenly on the fully expanded leaf and averaged.

Leaf stomatal density (the number of stomata per unit area) was determined using the impression approach (Radoglou & Jarvis, 1990). The abaxial part of the leaf was used for estimation of stomatal density. The leaf surface was coated with clear nail vanish and allowed to dry for approximately 20 minutes. The thin layer was carefully peeled off using clear cellotape and then mounted on a slide. The impressions were then examined under a microscope. The number of stomata per unit area for each mount was counted.

3.5.6 Ion content in *H. compressa* leaves and roots

The leaves and roots of both salt treated and control plants previously oven dried were used for determination of sodium and potassium content according to Kalra (1998) using dry ashing. Whereby, the oven dried samples were crushed into powder using an electric blender to homogenise the sample. Empty crucibles were then weighed. Approximately 2g of the crushed oven dried sample was weighed into the crucible (Plate 3.3A). This was followed by slow charring on a hotplate (Plate 3.3B) for an hour. The crucible and its contents were then placed in a furnace (Plate 3.3C) and the temperature increased gradually to 500°C for 4 to 6 hours for complete ashing of the sample. Compete ashing was characterised by greyish white appearance (Plate 3.3D).

The crucible and the contents were then allowed to cool and dissolved in 50ml of 5N nitric acid. The solution was then filtered through a whatman filter paper and the resulting solution was made to 100ml. This solution was then assayed on a flame photometer. Sodium and potassium occur in ionic form and their organic salts are usually water soluble. Therefore, they can be pre-treated with a simple solution (Lambert, 1980).

The percentage ash content was determined as follows;

$$\%\ ash\ content = \frac{Ashed\ weight - crucible\ weight}{(crucible + sample) - crucible\ weight)}\ x100$$

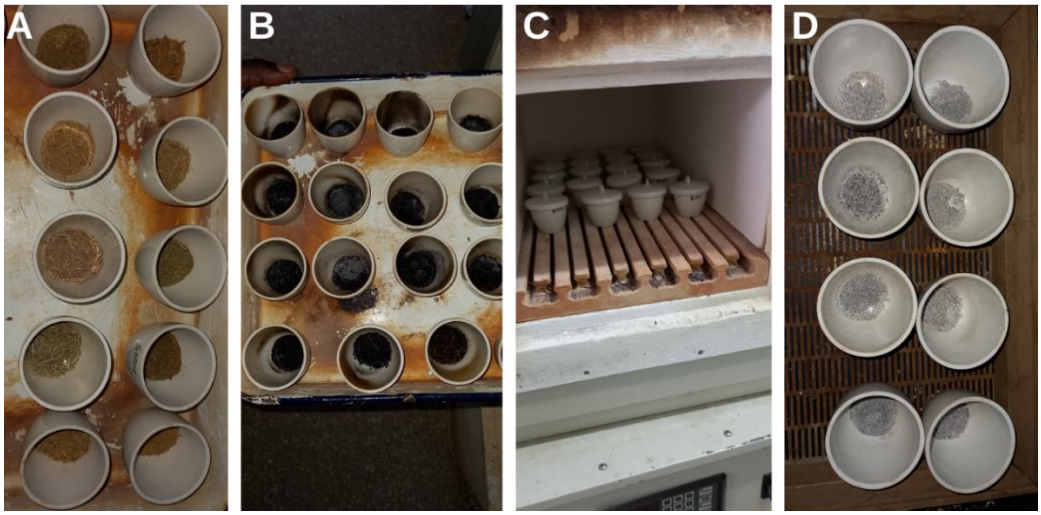


Plate 3.3: Na^+ and K^+ determination in shoot and root tissues of *H. compressa*

A. weighed, oven dried, ground samples ready for charring, **B.** Charred samples, **C.** Charred samples placed in a furnace, **D.** completely ashed samples characterised by greyish white appearance.

3.5.7 Determination of proline content in leaves of *H. compressa*

Proline content in leaves was determined using a modified protocol by Bates et al. (1973). Briefly, 0.25g of *H. compressa* fresh leaves were homogenized in 10ml of 3% sulfosalicylic acid. After a 3-hour incubation, the homogenate was centrifuged at 604 xg for 10 minutes. Two ml of the supernatant was then added to 2ml glacial acetic acid and 2ml acidic ninhydrin. This was then boiled at 100^{o}C for an hour and the reaction was terminated by placing the tube on ice. To the mixture, 4 ml of toluene was added and mixed vigorously and the mixture was allowed to warm to room temperature. Spectrophotometer readings at 520nm was then done using toluene as blank. A standard curve was prepared using the following concentrations; 2, 4, 6 and 8 µg per ml of L-Proline (Figure 3.2). The standard curve equation $Y = 0.2199*X + 0.07833$ was obtained showing a goodness of fit with a coefficient of determination of 0.97.

8 µg per ml of L-Proline (Figure 3.2). The standard curve equation $Y = 0.2199*X + 0.07833$ was obtained showing a goodness of fit with a coefficient of determination of 0.97. 8 µg per ml of L-Proline (Figure 3.2). The standard curve equation $Y = 0.2199*X + 0.07833$ was obtained showing a goodness of fit with a coefficient of determination of 0.97.

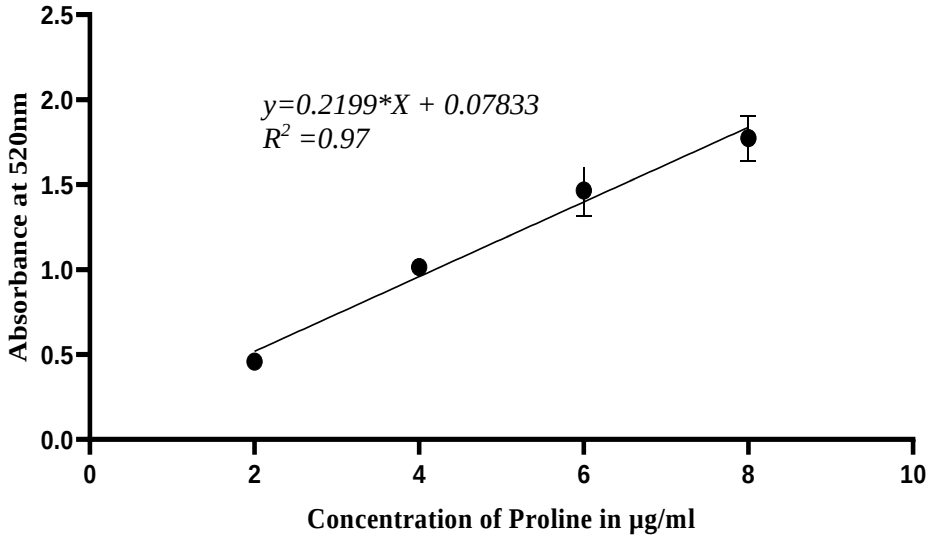


Figure 3.2: Graph of standard curve for proline.

3.5.8 Data analysis of morphological, physiological and biochemical traits

Two-way ANOVA followed by Tukey's multiple comparisons test was done using GraphPad Prism Version 8.0.2 for windows to compare the morphological, physiological and biochemical traits between the regions sampled and the salinity levels. In cases where there were repeated measures, two-way ANOVA was done followed by sidak test for multiple comparisons. Salt tolerance index (STI) was determined from the shoot dry biomass for *H. compressa* accessions from the three regions. Two-way ANOVA and Tukeys multiple comparison tests were done using GraphPad Prism Version 8.0.2 to determine the most tolerant accession based on STI.

3.5.9 Determination of salt tolerant accession

Salt tolerance index (STI) which is the ratio of total dry biomass under stress compared to the total biomass of their respective controls was used to distinguish the most tolerant accession (Tao et al., 2021). Reduction in biomass is considered under different saline treatments in relation to the controls where the lowest reductions compared between the controls and the stressed is indicative of high salinity tolerance (Lima et al., 2017; Zhou et al., 2014). According to Tao et al. (2021), STI can be categorized into four groups; sensitive (with STI lower than 0.35), moderately

sensitive (STI between 0.35 and 0.50), moderately tolerant (STI between 0.50 and 0.65) and tolerant (STI above 0.65). Salt Injury Index (SII) was also determined as follows; SII=1-STI. The most tolerant accessions to salinity stress were selected for transcriptomics analysis. Three days after the last salt treatment, leaf tissues were carefully collected for both salt stressed and salinity stressed at 300mM NaCl salinity level. For each of the control and salt stressed treatments, three replicates were sampled for transcriptomic studies. They were rinsed with distilled water and flash frozen in liquid nitrogen as described by Yaish et al. (2017).

3.5.10 RNA isolation and sequencing

Three biological replicates from each control and 300mM salt treatment of the most tolerant accessions were used. Total RNA was extracted from leaf samples of control and treated plants using Qiagen RNeasy plant mini kit. The integrity and quantity of RNA was determined using electrophoresis and NanoDrop respectively. The RNA was shipped to Beijing Genomics Institute in Hong Kong for sequencing. At BGI genomics, samples were thawed, fully mixed and centrifuged and 1ul of the sample was used for sample QC using the Agilent 4200.

3.5.11 Library construction and RNA sequencing

Strand specific transcriptome library construction was done at BGI where briefly, the total RNA was purified by oligo (dT)-attached magnetic beads followed by mRNA fragmentation. Complementary DNA (cDNA) synthesis was then done followed by end repair and adapter ligation. Several rounds of PCR were then performed in order to enrich the cDNA fragments. The quality and quantity of the library was then assessed followed by sequencing as paired end (PE 100) reads on Hiseq 4000 SBS Kit, Illumina.

3.5.12 Data Filtering

The obtained reads were filtered as follows; low quality reads (percentage of bases whose quality is less than 20 and is greater than 40% in a read), reads with adapters and reads with unknown bases (N bases more than 5%) were removed to get clean reads. The clean reads were stored in FASTQ format. Fastqc was used to check the

quality of the individual reads while multiqc was used to generate a single quality report of the fastqc results that included all the reads.

3.5.13 *De_novo* assembly

The high quality clean reads were used for *de_novo* transcriptome assembly using Trinity software Version 2.0.6 (Grabherr et al., 2013) with the following parameters; --min_contig_length 150 --CPU 8 --min_kmer_cov 3 --min_glue 3 --bfly_opts '-V 5 --edge-thr=0.1 --stderr' --SS_lib_type RF. The quality of the final assembly was assessed using the N50, contig length distribution and completeness assessment using BUSCO (Benchmarking Universal Single Copy Orthologs) version 5.2.2 (Manni et al., 2021). The BUSCO was run using eukaryota_odb10 lineage (Creation date: 2020-09-10). The software Tgicl version 2.0.6 (Pertea et al., 2003) was used to cluster transcripts, remove redundant transcripts and get unique genes (unigenes). The Tgicl was done for individual samples to get unigenes per sample which was then subjected to Tgicl to obtain final unigenes for downstream analyses.

3.5.14 Functional annotation of unigenes

The assembled unigenes were annotated using the following databases; NCBI non-redundant nucleotide sequences (NR), NCBI nucleotide sequences (NT), UniprotKB/ SwissProt, Kyoto Encyclopedia of Genes and Genomes (KEGG), Eukaryotic Ortholog Groups (KOG), Interpro and Gene Ontology (GO). The software diamond Version 2.0.14 (Buchfink et al., 2021) with an expectation value cut off of $<10^{-5}$ was used to align unigenes to NR, NT, KOG, KEGG and SwissProt databases. Blast2GO software (Conesa et al., 2005) was used for Gene Ontology (GO) annotation using NR annotations previously obtained. Interproscan was used to annotate the unigenes to the interpro database and also assign GO terms.

3.5.15 Prediction of coding regions in doum palm transcriptome unigenes

Transdecoder software (Haas & Papanicolaou, 2019) under the omics box (OmicsBox, 2019) was used to predict coding regions. The longest Open Reading Frame (ORFs) were first extracted, followed by pfam search to identify ORFs with homology to known proteins and finally, the coding regions were predicted.

3.5.16 Unigene expression

The clean reads obtained after filtering were mapped back to the 92135 unigenes that were *de_novo* assembled using Bowtie2 version 2.2.5 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml). Expression levels were calculated using RSEM version 1.2.12 (Li & Dewey, 2011). Box plots were drawn to show the distribution of gene expression. Fragments per Kilo base of transcripts per million mapped reads (FPKM) was used to show the different expression patterns per sample. The level of gene expression at different FPKM intervals (FPKM ≤:1, FPKM:1-10, FPKM≥10) was also estimated.

3.5.17 Differential gene expression due to salinity stress in *H. compressa*

Differentially expressed genes (DEGs) were calculated from the raw gene expression counts obtained using RSEM. The R software DEseq2 version 1.32.0 (Love et al., 2014) was used to determine DEGs with the following threshold; Alpha cut off (FDR) ≤ 0.05, $\log_2$FC > 1 (were upregulated) and $\log_2$FC< 1 (were downregulated).

3.5.18 Annotations of the DEGs

Annotations of DEGs were carried out with the following databases: GO and KEGG. The DEGs were classified based on the GO and KEGG annotation results previously obtained. The R software Phyper was used for functional enrichment of GO and KEGG annotations. The False Discovery Rate (FDR) for each p value was then calculated. Significant enrichments were defined when the FDR < 0.01.

3.5.19 Transcription factors (TF) in salinity stress

The protein sequences for transcription factors of *Phoenix dactylifera* were downloaded from plant Transcription Factor Database (TFDB) available at http://planttfdb.gao-lab.org/index.php?sp=Pda. Diamond program Version 2.0.14 (Buchfink et al., 2021) was used to align the sequences of the differentially expressed unigenes to the TF domains.

3.6 Development and validation of markers from RNA-seq data for genotyping studies in *H. compressa*

3.6.1 Development of SSR markers from unigene

The 92,135 unigenes obtained from *de_novo* RNAseq assembly were used for mining SSR markers. Perfect and compound SSR detection was performed using MISA (MicroSAtellite Identification tool) Version 2.1 (Beier et al., 2017) available at https://webblast.ipk-gatersleben.de/misa/index.php?. The minimum number of repeats for selecting mono-, di-, tri-, quad-, penta- and hexa- nucleotide motifs were set as 1, 2, 6, 5, 5 and 4 respectively. The maximum length of sequences between two SSR's in order to classify them as compound SSR's was set to 100 bp. The SSRs were also classified into three groups based on the number of repeat units as follows; Class I (>20bp), Class II (11 to 20bp) and Class III (<11bp) (Srivastava et al., 2019).

3.6.2 Primer design of the markers

Primer3 software at default parameters, was used for designing the forward and reverse primers that flank the obtained SSRs (Untergasser et al., 2012). The criteria utilized to choose prospective SSR markers using Primer3 were primer length between 18 and 20 bp, primer melting temperature (Tm) between 50 and 65°C, GC content of 40–60% and PCR product size between 100 and 300 bp. The *SSR repeat filter.py* python script was used to filter out loci that had the same priming site. The *get_orfs_or cdss.py* python script was used to explore the microsatellite marker loci in order to locate the longest ORFs and identify SSRs that are on the coding and non-coding regions.

3.6.3 PCR amplification

A total of 20 SSR markers both on coding and non-coding regions were randomly selected for validation with 5 accessions from each of the four regions (Tharaka, Tana River, Kwale and Turkana). Polymerase Chain Reaction conditions were optimized for each of the 20 SSR markers. The volume of each PCR reaction was 10 µl containing 2 µl of 5X reaction buffer (Bioline), 0.2 µl of Taq polymerase (Bioline), 0.2 µl of each primer (20 µM), 1 µl of template DNA and 6.4 µl of nuclease free water. All PCR reactions were conducted in Applied Biosystems Gradient PCR. The

amplification reactions consisted of initial denaturation of 3 minutes at 94°C followed by 30 cycles of 30 seconds at 94°C, 30 seconds at the primers' optimized annealing temperature and an extension of 45 seconds at 72°C. A final extension at 72°C was done for 10 minutes. The PCR products were viewed on a 1.8 % agarose gel electrophoresis system. A binary qualitative data matrix was prepared whereby bands on the gel were scored for presence of an amplicon (scored as 1) or absence of an amplicon (scored as 0) for each of the markers.

3.6.4 Cross genus transferability

In order to determine the cross genus transferability, two coconut samples were used for validation of the 20 SSR markers. They were also scored as other *H. compressa* accessions.

3.6.5 Genetic diversity

The number of observed alleles (*Na*), genetic diversity (*He*), Number of effective alleles (*Ne*), Shannon's Information Index (*I*), number of private alleles per population, observed heterozygosity *(Ho)*, expected heterozygosity (*He)*, unbiased expected heterozygosity *(uHe)*, F Statistics (F_{IS}, F_{ST}, F_{IT}), Analysis of Molecular Variance (AMOVA) and Principal Coordinate Analysis (PCoA) were calculated using GenALEx software version 6.5 (Peakall & Smouse, 2012). Polymorphic Information Content (PIC) which is a measure of the capacity of a marker to assess polymorphism (Pagnotta, 2018) was calculated for each of the 20 markers. The PIC value was calculated based on the following equation.

$$PIC = 1 - \sum_{i=1}^{n} P_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2P_i^2 P_j^2$$

where

n is the number of alleles,

pi and pj are the population frequency of the ith and jth alleles.

A PIC value of 0.00 indicates that all sample DNA exhibit the same banding pattern and the marker is said to be monomorphic while markers with PIC values greater than 0.00 are said to be polymorphic (Dalimunthe et al., 2020).

3.6.6 Population structure

The population structure of the 20 samples was inferred using the Bayesian clustering method in STRUCTURE software version 2.3.4 (Pritchard et al., 2000). Ten independent runs were carried out for each value of *k* from *k*=1 to *k*=10 to determine the optimal clusters. The runs were conducted with a burn-in of 100,000 iterations with 100,000 MCMC after burn-in while assuming an admixture model. The optimal number of populations (k) was determined using STRUCTURE HARVESTER (Evanno et al., 2005) available at http://taylor0.biology.ucla.edu/structureHarvester/. Accessions that had a membership probability less than 0.8 were categorized as admixture. An unweighted neighbor joining tree and Principal Coordinate Analysis (PCoA) were performed using DARwin (Dissimilarity Analysis and representation for windows) software version 6.0.21. The PCoA and the tree were drawn based on a dissimilarity matrix obtained from DARwin.

CHAPTER FOUR

RESULTS

4.1 Ethnobotany, domestication and effect of stress on *H. compressa*

4.1.1 Ethnobotany of *H. compressa*

Overall, 79 respondents from the four sampling locations, including 48.1% and 51.9% of women and men respectively, were surveyed for this study. In Turkana and Tana River, women were highly rated, whereas men were more frequently questioned in Kwale and Tharaka Nithi (Table 4.1). The lack of reliable informants and the sparsely populated regions in ASALs contributed to the low number of respondents.

Table 4. 1: *Hyphaene compressa* informants across the four ASAL counties

County	Ethnic group	Local Name	Number of respondents	Percentage of Women (%)	Percentage of Men (%)
Turkana	Rendile	*eng'ol*	27	51.9	48.1
Tana river	Pokomo	*Mkoma*	27	55.6	44.4
Kwale	Mijikenda	*Mkoma*	10	30	70
Tharaka Nithi	Aatharaka	*muruguju*	15	40	60

In Turkana, 27 respondents were surveyed and 26 (96.3%) of them reported that their farms had more than 15 doum palm trees. Turkana had the highest *H. compressa* density in one farm compared to all the other regions. This was followed by Kwale where 90% of the respondents reported more than 15 doum palm trees in one farm. In Tana River, between one and five doum plants were present in the farms of most respondents (51.9%). The other 48.1% had more than 5 doum plants per farm. Only 4 respondents in Tharaka Nithi reported more than 15 palm trees in one farm with the other respondents reporting less than this per farm.

The people of Kwale are knowledgeable about this plant. On the basis of variations in morphological characteristics, they were able to classify *H. compressa* into three types. They were known as Mkoma, Mbiye and Mkoko in their community. In Tharaka Nithi, just a few (20%) informants were able to distinguish between the varieties, although they did not give them names (Table 4.2). Turkana and Tana River informants were unable to distinguish between the varieties.

Table 4.2: Informant responses on ethnobotanical aspects of *H. compressa*

County	Ethnobotanical Aspects								
	Worthy to domesticate		Allow intercropping			Maintenance in the farm		Differentiate varieties	
	Yes	No	Yes	No	When Mature	Yes	No	Yes	No
Kwale	80.0	20.0	0.0	100.0	0	0	100	100	0
Tana River	55.6	44.4	0.0	100.0	0	0	100	0	100
Tharaka	53.3	46.7	86.7	0.0	13.3	100	0	20	80
Turkana	11.1	88.9	3.7	96.3	0	0	100	0	100

Values are in percentages

More than half of the people surveyed in Tharaka Nithi (53.3%), Tana River (55.6%) and Kwale (80%) agreed that *H. compressa* should be domesticated (Table 4.2). The least amount of respondents (11.1%) in Turkana supported domestication. Tharaka Nithi had the highest rate of intercropping doum palm and other crops, with 86.7% of informants doing so at any stage of the palm's development and just 13.3% doing so once it is mature (Plate 4.1A). According to respondents from Kwale and Tana River, *H. compressa* is not intercropped with other plants. However, only 3.7% of Turkana residents were open to intercropping. While none of the respondents from the other three counties performed any maintenance, all of the respondents from Tharaka Nithi engaged in some type of maintenance, such as pruning and weeding (Table 4.2).

Hyphaene compressa was intercropped with a variety of crops in Tharaka Nithi, including mangoes, cassava, cowpeas, green grams, sorghum, millet, pawpaws, pigeon peas, oranges and bananas (Plate. 4.1A, B). However, it is found in the wild in the counties of Kwale, Tana River and Turkana (Plate 4.1 C, D). It is the predominant type of vegetation in Turkana, while *Prosopis juliflora* and various *Vachellia* species are also present nearby. At the Kenyan coast, mangroves, cashew nuts, *Vachellia* species and coconuts were all seen to share the same ecology as *H. compressa* (Kwale and Tana River).

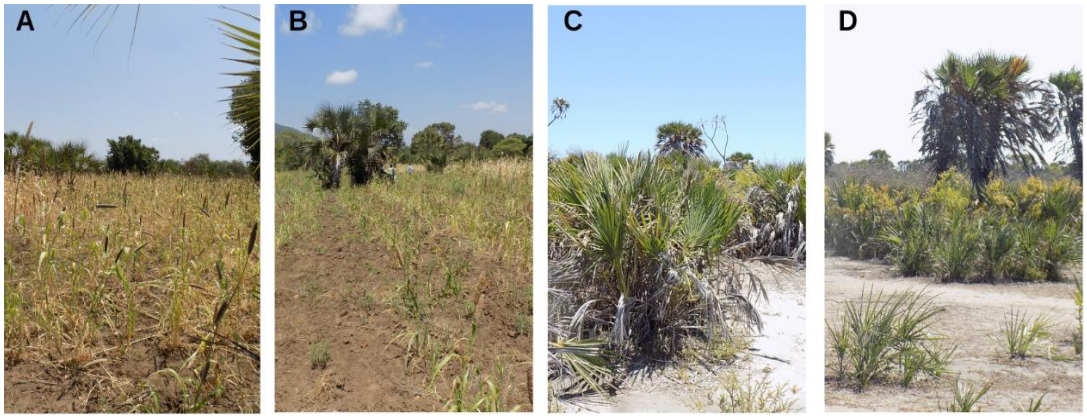


Plate 4.1: Doum palm maintenance

A and **B**. Doum palm intercropped with millet in Tharaka Nithi, **C.** Doum palm growing in the wild in Tana River**, D.** A bush of doum palm growing in Kwale

4.1.2 Uses of doum palm

4.1.2.1 *H. compressa* fruit as food

The utilization of doum palm fruit for food in each of the investigated areas was highly cited as evidenced by the high fidelity scores. The majority of respondents who reported the fruit as food were from Kwale (90%), followed by Turkana, Tana River and Tharaka Nithi (Table 4.3). People eat the fruit by smashing it and eating the flesh, or they can drink the water within to quench their thirst during drought (Plate 4.2A and B). The fruit can also be powdered and added to meals as a condiment. *Hyphaene compressa* fruits as feed was recorded in each of the regions. Animals in the ASALs consume leaves and ripe fruits that have fallen from doum palm trees (Plate 4.2C).

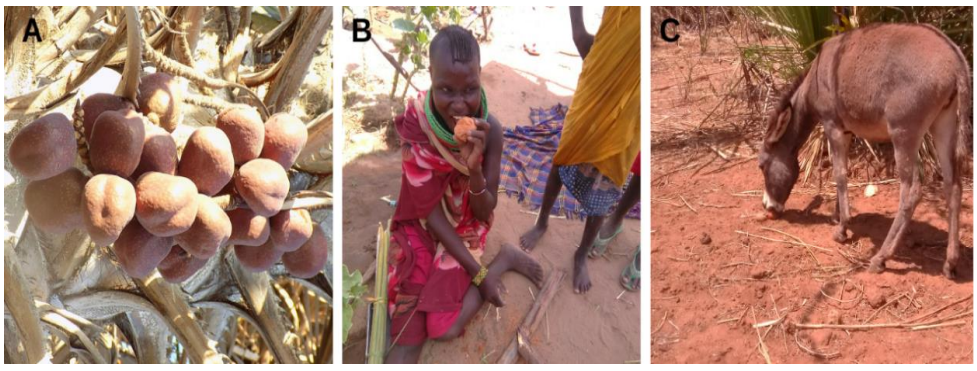


Plate 4.2: *H. compressa* fruits as food and feed

A. *H. compressa* fruits, **B**. A Turkana woman eating *H. compressa* fruit, **C**. A donkey eating *H. compressa* fruits that have fallen from the tree

The highest fidelity score in Tharaka Nithi was for the usage of *H. compressa* as food (80%), while the lowest score was for fastening farm tools (6.7%). Food utilization of the doum palm scored the highest in Tana River County (81.5%), while medicinal uses and fodder scored the lowest (7.4%). Fruit was most frequently used as food in Kwale (90%) and feed (leaves) was least frequently used (10%). Food usage of doum palms was also highly scored in Turkana (88.9%) whereas soil erosion control was least common (3.7%) as shown in Table 4.3.

It was less usual to use *H. compressa* fruits and leaves as animal feed than it was to use them as sustenance for humans. The evidence for this is the low fidelity scores for feed uses in comparison to food uses (Table 4.3).

4.1.2.2 Medicinal uses of *H. compressa*

In this study, three medicinal uses were noted. In Kwale, *H. compressa* crushed inflorescence is used to treat miscarriages and the fruit powder is used as a pain reliever for headaches. Crushed leaves are used to cure burns, however only 7.4% of respondents from Tana River reported this (Table 4.3). Tharaka and Turkana informants did not record any medicinal uses of this plant.

4.1.2.3 Doum palm leaves for weaving

Collecting the leaves for weaving is done on young palms, ideally those with longer midribs (Plate 4.3A and B). Residents of Tharaka Nithi weave mats, baskets and brooms out of *H. compressa* leaves (Plate 4.3C, F, G and H). Men, women and children were observed weaving in all the regions sampled. Nevertheless, the majority of weavers are women (Plate 4.3C). Compared to Tharaka Nithi weavers, local Turkana mat weavers have modernized their baskets and mats by adding color and more complex motifs (Plate 4.3 D and E). Doum palm leaves are used to create mats, baskets, ropes, fans and sieves in Kwale County (Plate 4.3 I, J, K and L).

Table 4.3: Plant parts used and fidelity scores of doum palm in Kenya

Plant Part	Use Category	Uses	Fidelity Level				P Value
			THK	TR	KW	TUR	
Fruit	Food	Fruit is crushed Water inside is taken, fruit powder is also used as food additive	80	81.5	90	88.9	0.568
	Feed	Donkeys, Camels	26.7	18.5	10	48.1	0.046
	Fuel	Dry Husks are used	40	55.6	-	-	0.001
	Medicine	Powder as painkiller for headaches Inflorescence crushed and given to expectant mothers to prevent miscarriage	-	-	20	-	0.267
	Other uses	Fasten Farm tools -The husk is put at the fulcrum for support.	6.7	-	-	-	0.229
Leaves	Handicrafts	Weaving mats, brooms, ropes, fishing nets, sieves, hats	53.3	70.4	60	74.1	0.589
	Medicine	Leaf ashes for treating burns	-	7.4	-		0.267
	Roofing	Thatching houses	33.3	74.1	70		0.033
	Feed	Donkeys, Camels, goats	26.7	7.4	10	22.2	0.396
Stem	Construction	Furniture, houses, boats	73.3	20	80	66.7	0.834
	Wine making	Apical meristem is cut to tap wine	-	37.03	40	-	0.000
	Ornamental	As hedge	20	-	-	-	0.004
	Other uses	Building Pet Houses	26.7	-	20	-	0.013
Roots	Prevention of Soil erosion	Prevent soil erosion	-	-	-	3.7	0.583

THK -Tharaka, **TR** -Tana River, **KW**-Kwale, **TUR**- Turkana

Plate 4.3: Handcraft made from *H. compressa* leaves

A. Woman from Tharaka Nithi collecting *H. compressa* leaves for weaving, **B**. Leaves ready for weaving in Tharaka Nithi, **C**. woman weaving baskets in Tharaka Nithi, **D** and **E**. Modernized baskets made from *H. compressa* by women in Turkana**, F, G** and **H**. mats, baskets and brooms made by people from Tharaka Nithi, **I, J, K** and **L**. baskets, fans, ropes and mats by women from Kwale.

4.1.2.4 Doum palm as construction materials

Logs from doum palms are cut and used in construction (Plate 4.4A). The petioles and trunk of the doum palm are used to build homes, granaries and animal kennels in Tharaka Nithi (Plate 4.4 C, E). Petioles are used in the fabrication of furniture (Plate 4.4 D). Doum palm is the most common building material in Tana River (Plate 4.4B). The indigenous populations thatch their mud homes with these palm leaves. Doum palm trunks are used by Turkana locals to build fishing boats. By chopping three to five medium-sized trunks into boats and connecting them together with ropes braided from doum palm leaves.

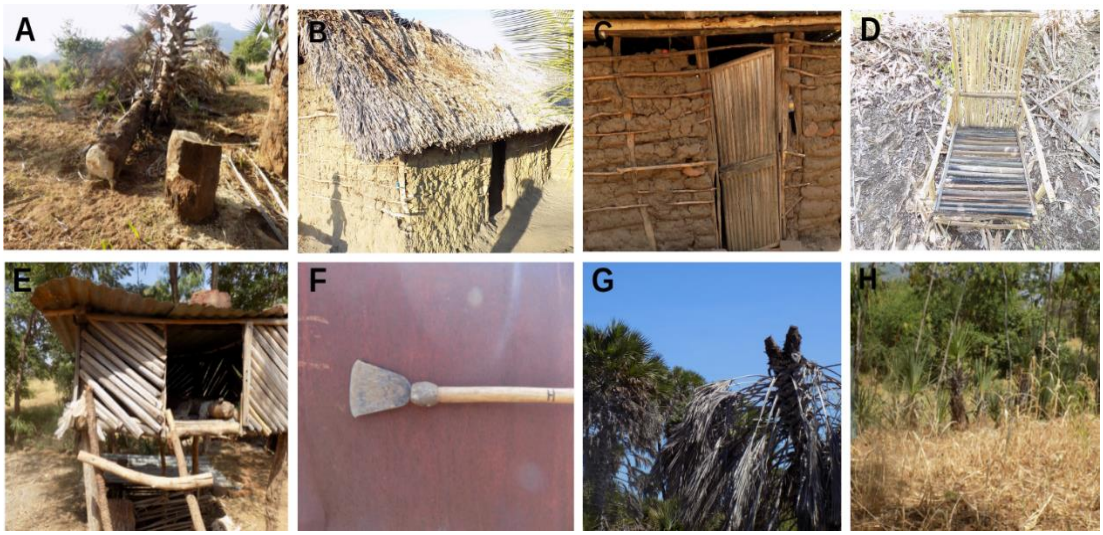


Plate 4.4: Construction and other *H. compressa* uses.

A. log of *H. compressa* ready for use in Tharaka Nithi, **B.** Mud hut roofed with *H. compressa* leaves in Tana River, **C**. door made of *H. compressa* in Tharaka Nithi, **D**. Chair in Kwale, **E**. pet house in Tharaka Nithi **F**. hoe arm fastened using *H. compressa* fruit in Tharaka Nithi, **G.** Stem cut for tapping wine by the coastal communities in Tana River, **H.** doum palm used as a hedge.

4.1.2.5 Other uses of doum palm

The fruit is also used to secure equipment for farming. The husk is inserted into the hoe's fulcrum to do this (Plate 4.4 F). Tana River and Kwale both appeared to be tapping wine from *H. compressa* (Plate 4.4 G). To make wine, the locals harvest the apical meristem's sap. Wine tapping was found to be damaging to the palms whereby in some cases, the plant's entire branch structure is removed during wine tapping, leaving it without any leaves. It then becomes very difficult for the plants to resume their vigorous growth.

Overall, there were no significant differences between the several regions sampled in the usage of doum palm as food (p=0.568), feed (leaf) (p=0.396), medicinal (leaf and fruit) (0.267), construction (p=0.834), weaving handicrafts (0.589), preventing soil erosion (p=0.583) and other applications of the fruit (p=0.229) as shown in Table 4.3. However, there were differences in the utilization of the fruit as feed (p =0.046), fuel (p=0.001) and roofing (p=0.033) among the different areas sampled (Table 4.3).

4.1.3 *H. compressa* plant part value

There were 14 reported uses for *H. compressa* in total. The stem had three functions, the leaves had four uses and the fruits had five uses. The only known use of *H. compressa* roots was reported in Turkana. With a percentage PPV Value of 35.7%, fruits were the most utilized plant part, followed by leaves and stem. The roots were used the least. The leaves were the most utilized *H. compressa* plant part in Tana River and Turkana with PPV values of 44.4% and 42.8% respectively compared to other plant parts. No utilization of roots was documented by Tharaka Nithi, Tana River, or Kwale (Table 4.4).

Table 4.4: RU_{Ppt} and PPV of *H. compressa* among four counties in Kenya

Plant Part	Overall		Tharaka		Tana River		Kwale		Turkana	
	RU_{Ppt}	PPV	RU_{Ppt}	PPV	RU_{Ppt}	PPV	RU_{Ppt}	PPV	RU_{Ppt}	PPV
Fruit	5	35.7	4	40	3	33.3	3	33.3	2	28.6
Leaves	4	28.6	3	30	4	44.4	3	33.3	3	42.8
Stem	4	28.6	3	30	2	22.2	3	33.3	1	14.3
Roots	1	7.1	-	-	-	-	-	-	1	14.3
RU	14		10		9		9		7	

RUPpt is Reported use per plant part, PPV: Plant Part Value in percentage

4.1.4 Biotic and abiotic stress in *H. compressa*

Human intervention and pest infestation were the most prevalent biotic stresses, whereas drought and salinity were the most prevalent abiotic stresses (Plate 4.5, Table 4.5). The highest levels of biotic stress, including pest infestation (Plate 4.5 D and E) and human interference on *H. compressa*, were recorded in Tharaka Nithi. The pests included aphids, scale insects and the palm butterfly. The harvesting practice in Tharaka Nithi stunts young palm trees (Plate 4.5A and B). Human involvement accounted for the majority of the observed biotic stress in Turkana, Kwale and Tana River (30%, 25% and 20%, respectively). Doum palm plants in Turkana and Tana River County suffered severe drought effects, whilst those in Tharaka Nithi and Kwale only experienced moderate damage (Table 4.5). Doum palms in Turkana, Tana River and Kwale showed evidence of saline damage (Plate 4.5C, Table 4.5).

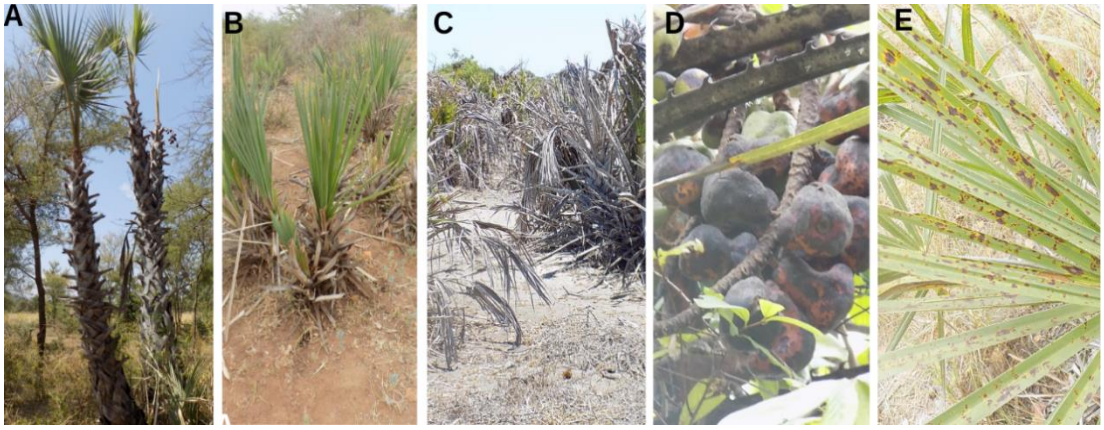


Plate 4.5: Biotic and abiotic stress of *H. compressa*

A and **B.** Overharvesting of leaves in Tharaka Nithi, **C**. Effects of high salinity in Tana River, **D** fruit infection in Tharaka and Kwale, **E**. leaf rust in Turkana

Table 4.5: Stresses on *H. compressa* accessions per sampled regions

Sample Area	**Biotic stress**		**Abiotic stress**	
	Morphological descriptor	**Percentage (%)**	**Descriptor**	**Range**
Tharaka Nithi	Leaf rust/pest infestation	73	Drought	Moderate
	Human interference (Overharvesting, logging, burning and clearing for human settlement)	60		
Turkana	Leaf rust/pest infestation	11	salinity	Severe
	Human interference (logging and burning)	30	Drought	
Kwale	Human interference (leaf overharvesting, wine tapping, logging and burning)	25	Salinity Drought	Moderate
Tana River	Human interference (leaf overharvesting, wine tapping, logging and burning)	20	Salinity Drought	Severe

4.2 Morphological diversity of doum palm

4.2.1 Morphological diversity of fruit and vegetative quantitative traits

Table 4.6 provides an overview of the quantitative trait frequencies. Doum palm height showed significant variation (cv=38.3%). Fruit weight (cv=21.5) ranged from 48.2 g to 148.8 g. *Hyphaene compressa* fruit length variation was minimal (cv=11.8).

Table 4.6: Frequency of *H. compressa* quantitative morphological traits

Descriptor	Range	Mean	CV(%)
Height	1.8-20	10.15	38.4
Leaf length	61-161	106.6	22.1
Leaf breadth	30-124	74.57	25.7
Petiole length	52-153.1	97.64	20.1
Fruit Length	4.7-8.4	7.05	11.8
Fruit Breadth	4.4-7.1	6.1	8.1
Fruit weight	48.3-148.8	107.21	21.5

The fruit and vegetative quantitative traits of *H. compressa* varied by region. All seven quantitative features were successful in phenotypically differentiating the doum palm ($p \leq 0.001$; Table 4.7). Kwale and Turkana *H. compressa* palms did not significantly differ from one another in terms of leaf length, leaf breadth, fruit length, or fruit weight. The average height was highest in Tharaka Nithi (13.5m) and the lowest in Kwale (5.65m). Tana River's leaf breadth (55.87 cm) was substantially less than that of the other sampling sites ($p \leq 0.001$).

Table 4.7: Mean of quantitative traits of doum palm from four regions in Kenya

Trait	Tharaka	Tana River	Kwale	Turkana	P value
	Mean±se	Mean±se	Mean±se	Mean±se	
H	13.5±3.73^{a}	9.93±2.58^{b}	5.65±2.28^{c}	9.16±1.94^{b}	8.24e-13***
LL	114.73±26.25^{a}	120.2±22.78^{a}	96.2±11.72^{b}	96.93± ^{b}	0.000473***
LB	89.47±13.12^{a}	55.87±7.73^{c}	73.33±12.53^{b}	69.63±20.3^{b}	3.24e-09***
PL	92.47±12.06^{b}	93.27±13.7^{ab}	109±24.94^{a}	99.33±23.38^{ab}	0.0426*
FL	7.59±0.34^{a}	7.64±0.3^{a}	6.33±1.36^{b}	6.56±0.25^{b}	6.32e-12***
FB	6.031±0.227^{a}	6.27±0.228^{a}	5.6±0.798^{b}	6.32±0.40^{a}	6.36e-06***
FWGT	127.6±11.24^{a}	111.53±9.3^{b}	91.73±34.94^{c}	92.37±9.11^{c}	2.09e-12***

Same letters within the row indicate no significant difference between the means while different letters indicate significant difference between the means at α=5% significance codes *=0.01, ***=0.000. **H** (Height), **LL** (Leaf length), **LB** (Leaf breadth), **PL** (Petiole length), **FL** (Fruit length), **FB** (Fruit breadth), **FWGT** (Fruit weight)

In addition, Tana River had the greatest overall leaf length (120.2cm) and fruit length (9.64cm) with a p value of 0.0005 and <6.32e-12, respectively. Figure 4.1 and 4.2 illustrates the variations in these physical characteristics between the various regions.

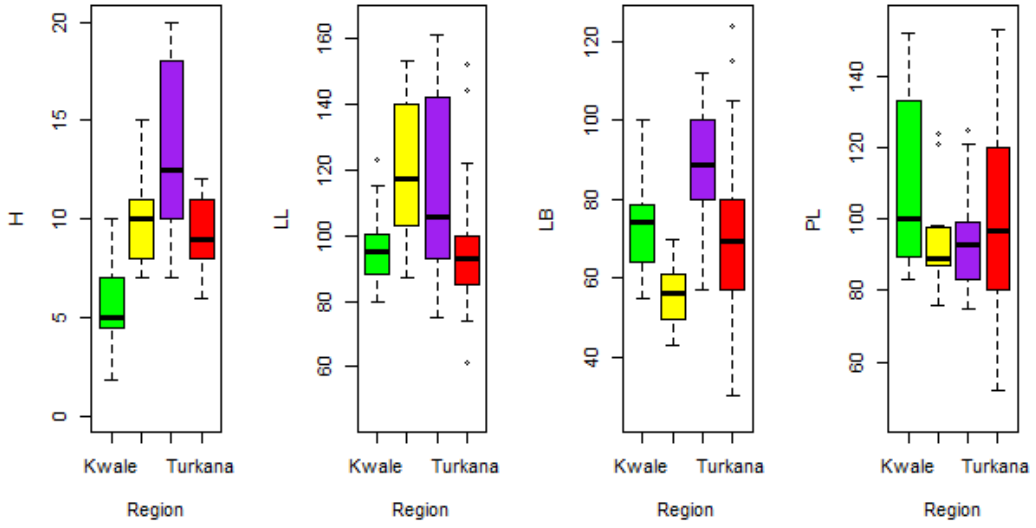


Figure 4.1: Box plots of *H. compressa* vegetative quantitative traits

H Height, **LL** Leaf Length, **LB** leaf breadth, **PL** Petiole Length, Green (Kwale), yellow (Tana River), Purple (Tharaka) and red (Turkana).

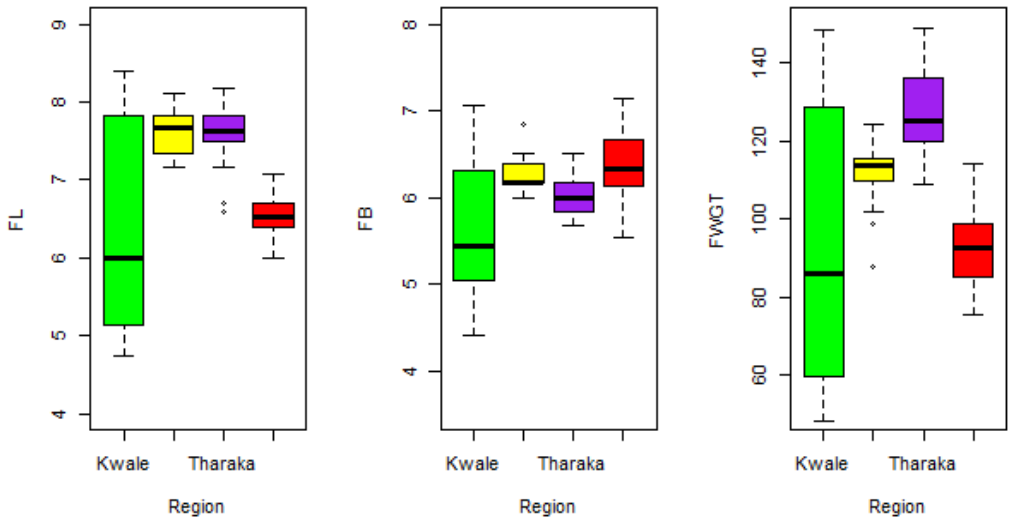


Figure 4.2: Box plots of *H. compressa* fruit quantitative traits

FL Fruit length, **FB** fruit breadth, **FWGT** Fruit weight. Green (Kwale), yellow (Tana River), Purple (Tharaka) and red (Turkana).

Doum palm height exhibited a significant positive correlation with leaf length ($p \leq 0.001$), fruit breadth ($p=0.029$), leaf breadth ($p= 0.006$), fruit weight ($p \leq 0.001$) and fruit length ($p \leq 0.001$). Negative correlation was observed between the length of the petiole and all three quantitative fruit traits: fruit weight ($p \leq 0.001$), fruit breadth ($P=0.004$) and fruit length ($p \leq 0.001$). Fruit weight was positively correlated with height ($r^2= 0.521$), leaf length ($r^2= 0.345$), leaf breadth ($r^2= 0.346$), fruit length ($r^2=0.861$), fruit breadth ($r^2=0.386$) and negatively correlated to petiole length ($r^2= -0.378$) as shown in Table 4.8.

Table 4.8: Correlation between quantitative traits of doum palm from Kenya

	H	**LL**	**LB**	**PL**	**FL**	**FB**
LL	0.425**					
LB	0.288**	0. 270**				
PL	-0.082	0.031	-0.073			
FL	0.501**	0.415**	0.247*	-0.425**		
FB	0.231*	0.121	-0.056	-0.300**	0.476	
FWGT	0.521**	0.345**	0.346**	-0.378**	0.861**	0.386**

**. Correlation is significant at the 0.01 level

*. Correlation is significant at the 0.05 level

H (Height), **LL** (Leaf length), **LB** (Leaf breadth), **PL** (Petiole length), **FL** (Fruit length), **FB** (Fruit breadth), **FWGT** (Fruit weight)

Leaf length, height, leaf breath, fruit length, petiole length and fruit breath were all taken into account while fitting a linear mixed effects model to predict *H. compressa* fruit weight. The four sampling zones were incorporated into the model as random effects. The model's overall explanatory power was strong (conditional $R^2 = 0.80$) and the portion that was due only to the fixed effects (marginal R^2) =0.66. The intercept of the model was -70.25. The effect of fruit length was positive and can be regarded as substantial and significant within this model (beta = 19.01, std. beta = 0.68, $p \leq 0$ .001) while the effect of fruit breadth was positive and can be regarded as very little and significant (beta = 7.35, std. beta = 0.16, $p < 0$ .05). Height, leaf length, leaf breath and petiole length were seen to have no significant effect on fruit weight.

4.2.2 Morphological diversity of fruit and vegetative qualitative traits

While there was little variation in qualitative traits like fruit apex shape, fruit shape, mid rib color, fruit base shape, petiole color and unripe fruit color, other qualitative traits like trunk branching, mature fruit color, trunk color, leaf color and trunk diameter

exhibited significant variations. The fruits of the doum palm that were sampled all had lustrous skin that had melded with the flesh. The mesocarp had an orange color, a fibrous texture and a distinct potent scent (Plate 4.6 A, F, G and H).

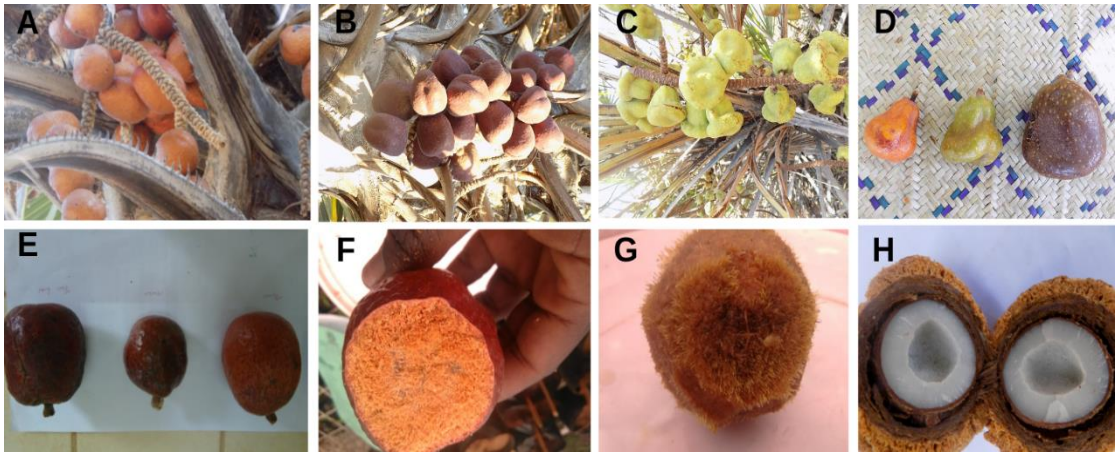


Plate 4.6: *H. compressa* fruit morphology in Kenya

A. Orange Mature fruits in Tharaka, **B**. Immature maroon fruits in Turkana, **C**. Immature green fruits in Kwale, **D**. Fruit morphology in Kwale, Left to right; obovate shape (orange), ovate shape (green) and round oblong shape (Brown in color), **E**. From left to right; Fruits from Tana River, Turkana and Tharaka Nithi, **F**. Orange mesocarp of doum palm fruit, **G.** Hairy mesocarp of doum palm fruit, **H**. Cross section of doum palm fruit

All of the fruits from Turkana and Tharaka Nithi that were sampled had oblong shapes and truncate bases (Table 4.9). The fruits from Kwale displayed the widest variety of characteristics, with various forms, bases and apices (Plate 4.6D). Unripe doum palm fruits were green in Turkana, Tana River, Tharaka Nithi and Kwale (Plate 4.6C). Immature fruits from Turkana were maroon (43.3%) in color (Plate 4.6B). All four sampling sites had different mature doum palm fruit colors, with most of the fruits being reddish brown. As indicated in Table 4.9, the fruits from Tharaka Nithi were either Orange brown (63.3%), brown (30%) or orange (6.7%) when fully ripe, as opposed to the fruits obtained from Tana River, which were all reddish brown. All leaf petioles had different petiole colors and were thicker at the base than the top (Table 4.9, Plate 4.7). Some doum palms did not branch at all. The bulk of the palms possessed dichotomous trunks, nevertheless. A total of 46.7% of the palms in Kwale had no trunk branching (Plate 4.8A).

Table 4.9: Fruit and vegetative qualitative traits of Kenyan doum palm accessions

Trait	Category	Tharaka %	Tana River %	Kwale %	Turkana %
Trunk color	Dark brown color	16.7	-	-	10
	Pale color	26.7	100	-	6.7
	Ashy color	56.7	-	100	88.3
Trunk diameter	Thick	53.3	-	26.7	16.7
	Medium	46.7	80	40	88.3
	Thin	-	20	33.3	-
Trunk branching	No branching	-	13.3	46.7	6.7
	2 trunk branching	66.7	80	40	60
	More than 2 branching	33.3	6.7	13.3	10
Leaf color	Dark green	36.7	-	-	6.7
	Green	63.3	-	100	46.7
	Light green	-	100	-	46.7
Mid rib color	Green	-	-	-	23.3
	Yellow green	100	100	100	76.7
Petiole color	Green with black stripes	33.3	-	6.7	80
	Green	16.7	-	26.7	16.7
	Yellow with black stripes	20	100	-	-
	Brown with black stripes	30	-	66.7	3.3
Pinnae density	Very dense	100	-	-	3.3
	Dense	-	100	53.3	96.7
	Lax		-	46.7	-
Fruit shape	Round Oblong	100	93.3	33.3	100
	Obviate	-	6.7	33.3	-
	Ovate	-	-	33.3	-
Fruit Apex	Truncate	100	100	66.7	100
	Depressed	-	-	33.3	-
Fruit Base	Truncate	100	100	33.3	100
	Acute	-	-	66.7	-
Fruit color-Unripe	Green	100	100	100	56.7
	Maroon	-	-	-	43.3
Fruit color when mature	Reddish brown	-	100	6.7	56.7
	Brown	30	-	60	43.3
	Orange brown	63.3	-	-	-
	Orange	6.7	-	33.3	-

In all of the study sites, two trunk branching was typical, with Tana River having the largest percentage of these palm trees (80%) as shown in Table 4.9. On the contrary, of the sampled doum, Turkana and Tharaka had 10% and 33.3%, palms with more than two trunk branches respectively (Table 4.9). *Hyphaene compressa* dichotomous branching was observed either mid-section or at the base (Plate 4.8 B, C and D).

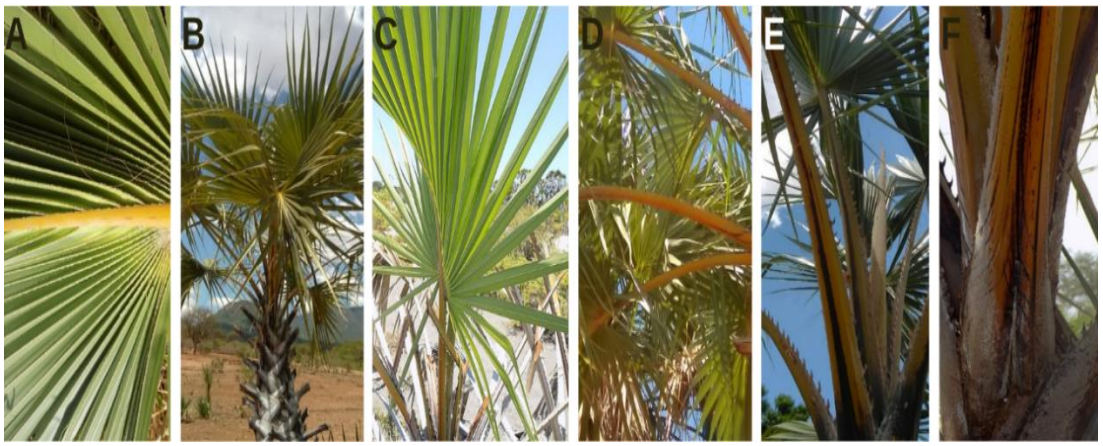


Plate 4.7: *H. compressa* leaf and petiole morphology

A, B, C. curved costa of *H. compressa*, **D**. Long thin petioles observed in Kwale, **E**. Petiole color (yellow with black stripe) in Tana River, **F**. Stout leaf base morphology

Plate 4.8: Branching morphology of *H. compressa*

A. Single trunk morphology showing low fruiting height in Kwale, **B**. Two trunk branching above the ground, **C**. Two trunk branching on the ground, middle and on top forming 8 crowns, **D**. More than two trunk branching at the ground level

4.2.3 Relationships between discriminant morphological descriptors

The following discriminant models were derived:

$$LD1 = 0.55Ht - 0.0009LL + 0.47LB - 0.47PL + 1.55FL - 0.52FB + 1.19FWGT$$

$$LD2 = 0.04Ht - 0.12LL - 0.24LB + 0.20PL + 0.92FL + 1.29FB - 0.79FWGT$$

$$LD3 = 0.90Ht - 0.57LL + 0.80LB - 0.02PL - 1.42FL + 0.66FB + 0.40FWGT$$

Where,

LD1, LD2 and ***LD3*** are discriminant functions,

Ht=Height

LL=Leaf Length

LB=Leaf Breadth

PL=Petiole length

FL=Fruit length

B=Fruit Breadth

FWGT=Fruit weight

LD1 accounted for 76.2% of the variation, while LD2 and LD3 each accounted for 15.03% and 8.8% respectively. The second and third elements have a minimal impact on group discrimination. Densities of the discriminant scores for each region are displayed on the LD1 separation plot. Overlaps on the separation plots show that the LD1 discriminant qualities do not distinguish the groups. Therefore, the absence of overlap suggests that the accessions from that area are morphologically unique from the others. Some Kwale samples revealed no overlap with any of the groups from Tharaka, Turkana, or Tana River. There was overlap (no separation) between Turkana and Kwale and between Tana River and Tharaka (Figure 4.3).

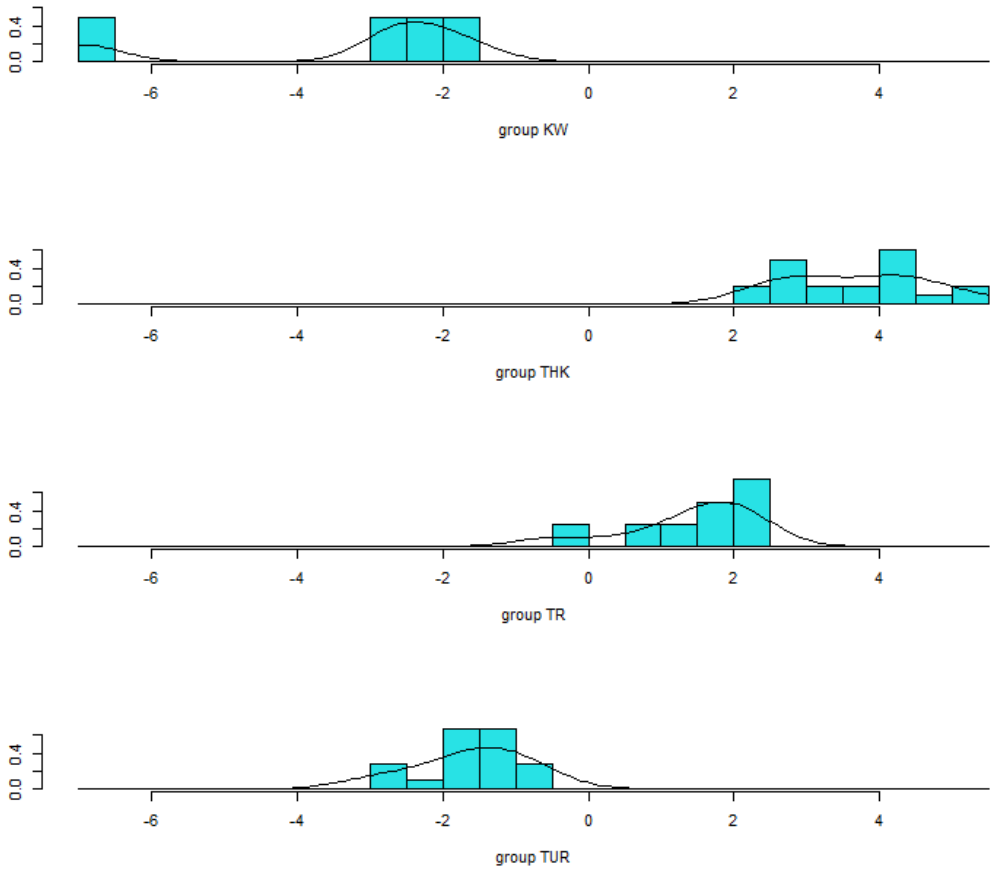


Figure 4.3: Separation between groups and overlapping areas that predict classes in doum palm.

KW- Kwale, **THK** – Tharaka, **TR** - Tana River, **TUR** – Turkana

4.2.4 Principal component analysis

Up to 59% of the diversity in doum palm qualitative features was explained by the first, second and third components (Table 4.10). Component one, which had positive correlations with fruit form, fruit apex, fruit base, trunk diameter and pinnae density but negative correlations with unripe fruit color and trunk branching, explained 25% of the variability (Figure 4.4a). The second component, which was positively connected to leaf color, accounted for 18% of the variability and the third component, which was positively correlated to trunk color and trunk diameter, accounted for 16%.

Table 4.10: PCA of qualitative traits of doum palm in ASALs of Kenya

	Principal components											
	1	2	3	4	5	6	7	8	9	10	11	12
FS	0.8	-0.2	0.4	-0.2	-0.2	0.1	0.0	0.3	0.0	0.1	-0.2	0
FA	0.8	-0.2	-0.4	0.3	0.1	0.1	-0.1	0.0	0.1	0.0	0.0	0
FBAS	0.8	-0.2	-0.4	0.3	0.1	0.1	-0.1	0.0	0.1	0.0	0.0	0
FC.UN	-0.1	0.4	0.4	0.4	0.6	0.1	-0.1	-0.2	0.0	0.2	-0.1	0
FC.M	0.1	-0.7	0.6	0.0	0.0	0.0	0.0	0.1	0.1	0.3	0.2	0
TC	0.3	0.1	0.4	0.6	-0.2	-0.6	0.3	-0.1	0.1	-0.1	0.0	0
TD	0.5	0.3	0.4	-0.4	0.0	0.4	0.3	-0.3	0.2	-0.2	0.0	0
TB	-0.5	-0.1	-0.4	0.4	0.1	0.4	0.4	0.2	0.1	0.0	0.0	0
LC	0.1	0.8	-0.3	-0.3	0.2	-0.2	-0.1	0.3	0.4	0.2	0.0	0
MRC	0.1	-0.4	0.1	-0.3	0.8	-0.2	0.1	0.2	0.0	-0.3	0.0	0
PC	0.3	-0.4	-0.6	-0.4	0.1	-0.2	0.3	-0.3	0.0	0.3	0.0	0
PD	0.6	0.6	-0.1	0.0	0.1	0.1	0.2	0.2	-0.4	0.1	0.1	0
Proportion of variance												
	0.3	0.2	0.2	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0
Cumulative percentage												
	0.25	0.43	0.59	0.69	0.79	0.85	0.89	0.94	0.94	0.97	0.99	1

FS (fruit shape), FA (fruit apex), FBAS (fruit base), FC.UN (Fruit colour-Unripe), FC.M (Fruit colour- mature), TC (trunk colour), TD(Trunk diameter), TB (trunk branching), LC (Leaf colour), MRC (Mid rib colour), PC (petiole colour), PD (Pinnae density).

The total quantitative traits variability explained by the first, second and third component was 75% (Table 4.11). The first component, which was associated to fruit length, fruit breadth, fruit weight, height and leaf length, accounted for 44% of the variability. The second component on the other hand accounted for 19% of the variation in petiole length and leaf breadth (Table 4.11).

Component one showed a negative relationship with petiole length and a positive relationship with all fruit characteristics, leaf length, width and tree height. This means that the petiole length decreased with increasing fruit size. Comparatively, component two was positively connected with vegetative data and negatively correlated with fruit attributes (Figure 4.4b).

Table 4.11: PCA of quantitative traits of doum palm in ASALs of Kenya

Trait	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7
H	0.69	0.33	0.21	0.23	-0.43	-0.36	0.01
LL	0.55	0.52	0.33	-0.49	0.26	-0.10	-0.04
LB	0.42	0.53	-0.60	0.26	0.32	-0.08	0.03
PL	-0.46	0.60	0.46	0.33	0.06	0.32	0.02
FL	0.91	-0.14	0.02	-0.07	-0.06	0.28	0.25
FB	0.54	-0.52	0.39	0.32	0.41	-0.15	-0.02
FWGT	0.89	-0.06	-0.10	0.05	-0.13	0.34	-0.23
Proportion of variance							
	0.44	0.19	0.13	0.08	0.08	0.07	0.02
Cumulative variance							
	0.44	0.63	0.75	0.84	0.92	0.98	1.00

PC (Principal components), H (Height), LL (Leaf length), LB (Leaf breadth), PL (Petiole length), FL (Fruit length), FB (Fruit breadth), FWGT (Fruit weight)

The doum palm was clustered using individual PCA based on qualitative and quantitative features into three and two major clusters, respectively (Figure 4.5a and b). Five Kwale samples formed their own cluster based on both qualitative and quantitative characteristics. Additionally, the same five samples formed their own cluster, which is represented as morphotype 4 (Table 4.12, Figure 4.6).

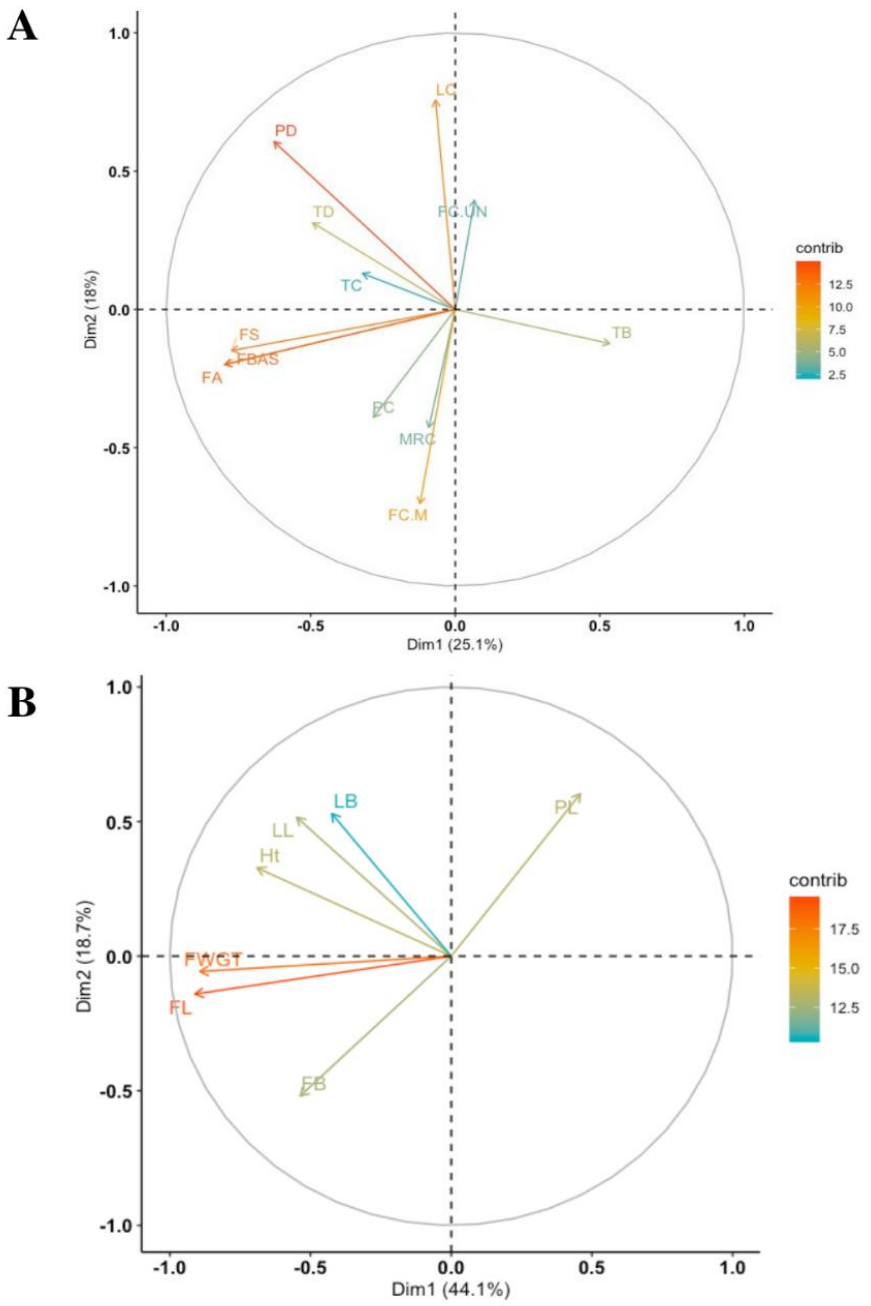


Figure 4.4: Variables PCA plot for *H. compressa* traits in Kenya

A. Variables PCA plot of qualitative traits
FS (fruit shape), **FA** (fruit apex), **FBAS** (fruit base), **FC.UN** (fruit colour-unripe), **FC.M** (fruit colour- mature), **TC** (trunk colour), **TD** (trunk diameter), **TB** (trunk branching), **LC** (leaf colour), **MRC** (mid rib colour), **PC** (petiole colour), **PD** (pinnae density)
B. Variables PCA of quantitative traits
H (height), **LL** (leaf length), **LB** (leaf breadth), **PL** (petiole length), **FL** (fruit length), **FB** (fruit breadth), **FWGT** (fruit weight).

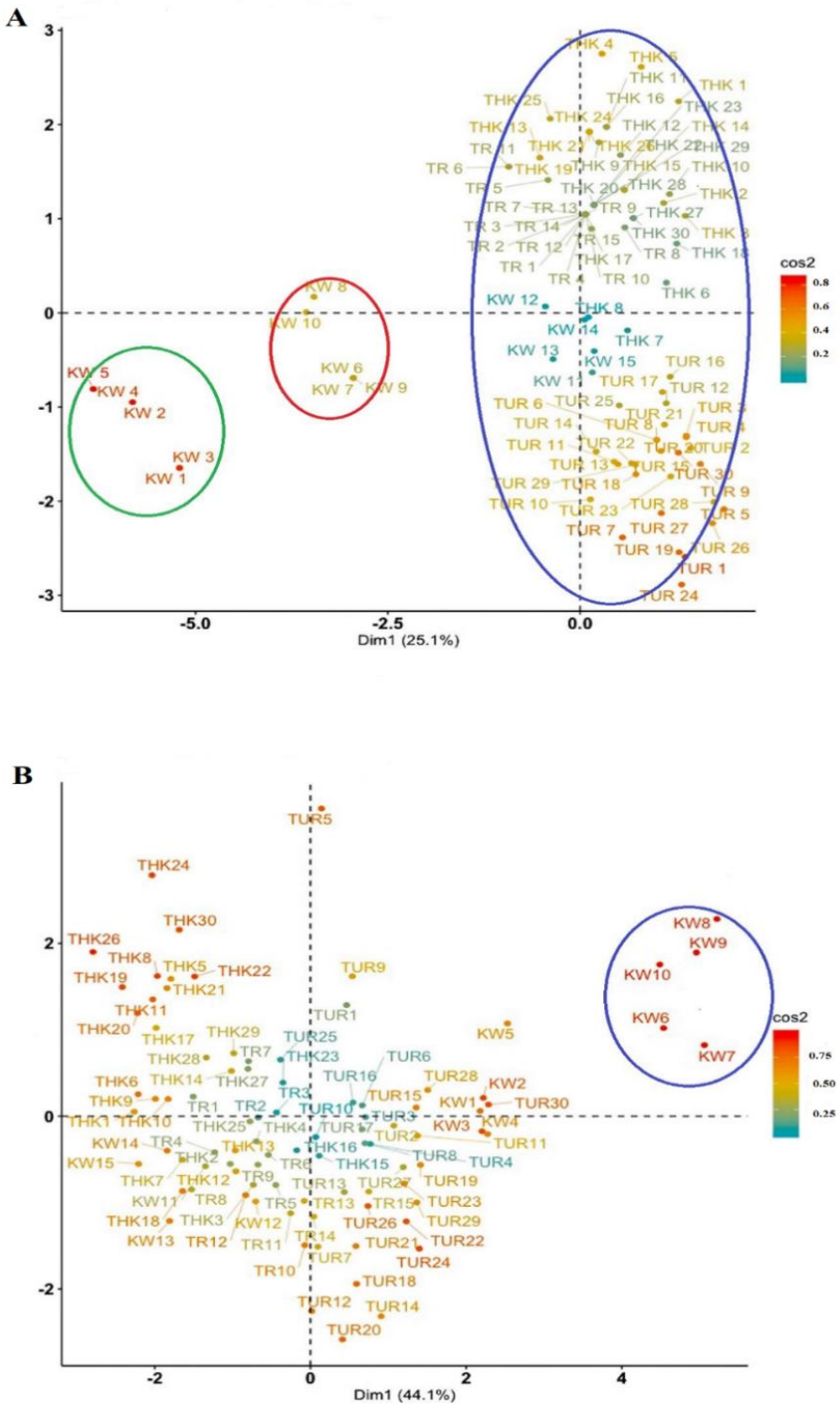


Figure 4.5: Individual accessions PCA for *H. compressa* traits in Kenya
A. Individuals PCA using qualitative traits showing three clusters,
B. Individuals PCA using quantitative traits showing two major clusters.
KW (Kwale), **THK** (Tharaka Nithi), **TR** (Tana River), **TUR** (Turkana)

4.2.5 Cluster analysis

The 90 samples of doum palms were grouped into 5 morphotypes by hierarchical clustering of quantitative traits (Table 4.12; Figure 4.6). A total of 77.3% of accessions in morphotype 1 were from Turkana. Morphotype 3 contained representative palms from the four Kenyan regions that were sampled. Morphotype 4 accessions were exclusively found in Kwale. Overall, 90.5% of the palms that belonged to morphotype 5 came from Tharaka. Some of the Kwale accessions that were sampled clustered with morphotypes 1, 3 and 5, demonstrating the heterogeneity of these palms.

4.2.6 Identification of elite doum palm

The minimum, maximum and mean of the morphological traits are shown in Table 4.12. The tallest trees (mean=14) and largest fruits (mean=129.4) were found in morphotype 5 accessions. This cluster contains palms from Kwale (9.5%) and Tharaka (90.5%). Tharaka samples that grouped together displayed a high degree of homogeneity. Fruit sizes and attributes were intermediate for accessions in morphotypes 2 and 3. Additionally, the longest leaves were recorded in morphotype 2. The shortest palms (mean=3.96), smallest fruits (mean=53.62) and longest petioles (mean=141.6) were observed in morphotype 4. Because of its fruit traits, morphotype 5 should be chosen for improvement.

Table 4.12: Quantitative traits in *H. compressa* morphotypes from ASALs of Kenya

Number	Morphotype 1 22(24.4%)		Morphotype 2 19 (21.1%)		Morphotype 3 23 (25.6 %)		Morphotype 4 5 (5.6%)		Morphotype 5 21 (23.3%)	
Trait	**Mean**	**Range**	**Mean**	**Range**	**Mean**	**Range**	**Mean**	**Range**	**Mean**	**Range**
Height	8.7	3-12	10	7-18	9.5	5-15	3.96	1.8-6	14	7-20
Leaf Breadth	69	30-124	61.7	40-93	77.7	47-108	61.2	55-67	92.62	73-112
Leaf length	98.6	61-152	116.7	79-153	100.6	74-132	87.6	80-95	117	85-161
Petiole length	93.9	52-153	91.8	76-132	98.9	75-124	141.6	127-152	94.95	76-125
Fruit length	6.7	6-7.8	7.4	6.6-8.1	6.98	5.7-8.4	4.97	4.7-5.3	7.6	6.7-8.2
Fruit breadth	6.2	5.3-6.9	6.3	5.7-7.1	6.1	4.7-8.9	4	4.4-5.2	6.1	5.7-7.1
Fruit weight	95.7	80.8-125.6	112.9	97.1-136	104.95	75.4-148.6	53.62	48-63	129.4	109-149
Number (%) per Sampling points										
Tharaka	4.5		21.1		26.1		0		90.5	
Tana River	4.5		57.9		13.0		0		0	
Kwale	13.6		0		21.7		100		9.5	
Turkana	77.3		21.1		39.1		0		0	

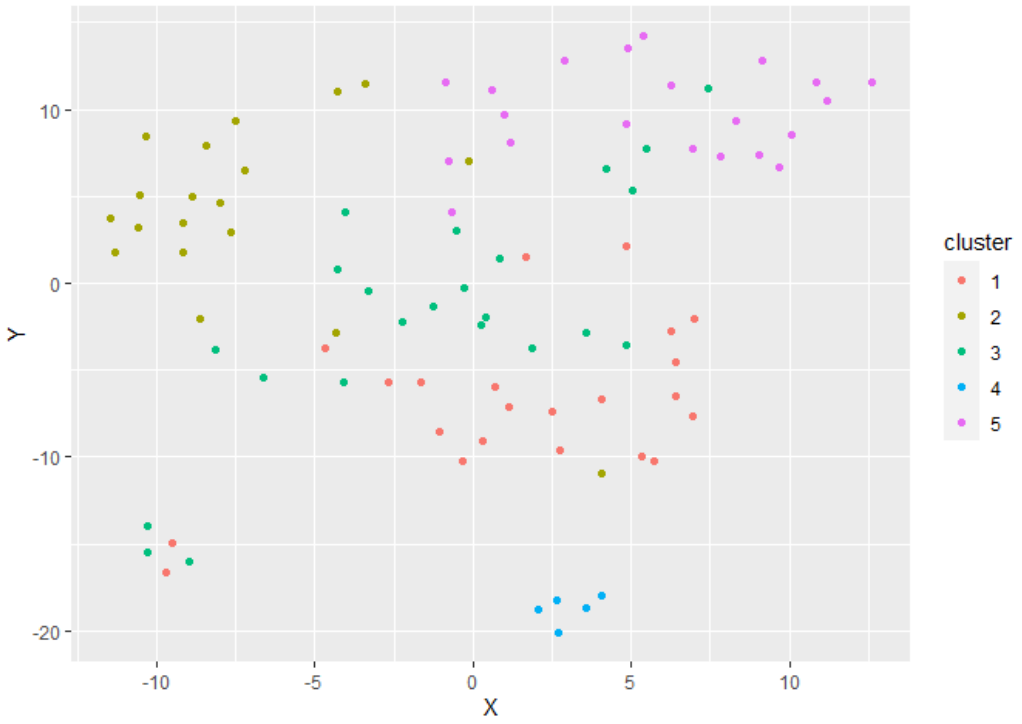


Figure 4.6: Cluster analysis of *H. compressa* accessions from Kenya

The morphotypes are color coded

4.3 Genetic diversity of *H. compressa*

4.3.1 Sample quality control (QC)

The integrity of 93 out of the 96 *H. compressa* DNA samples on a 1% agarose gel was as shown in Plate 4.9. Out of the 96 samples sent for sequencing, six were moderately degraded (Appendix V). However, they all qualified for library preparation and sequencing. The concentration of DNA ranged from 32.3 -302.5ng/µl (Appendix VI).

4.3.2 Genotyping by sequencing data

An average of 2.4 million reads were obtained from paired-end sequencing of 96 *H. compressa* accessions. The GC content ranged from 47.15 to 52.61 percent (Appendix VII). All of the samples received high phred ratings on multiqc reports, indicating good quality (Appendix VIII).

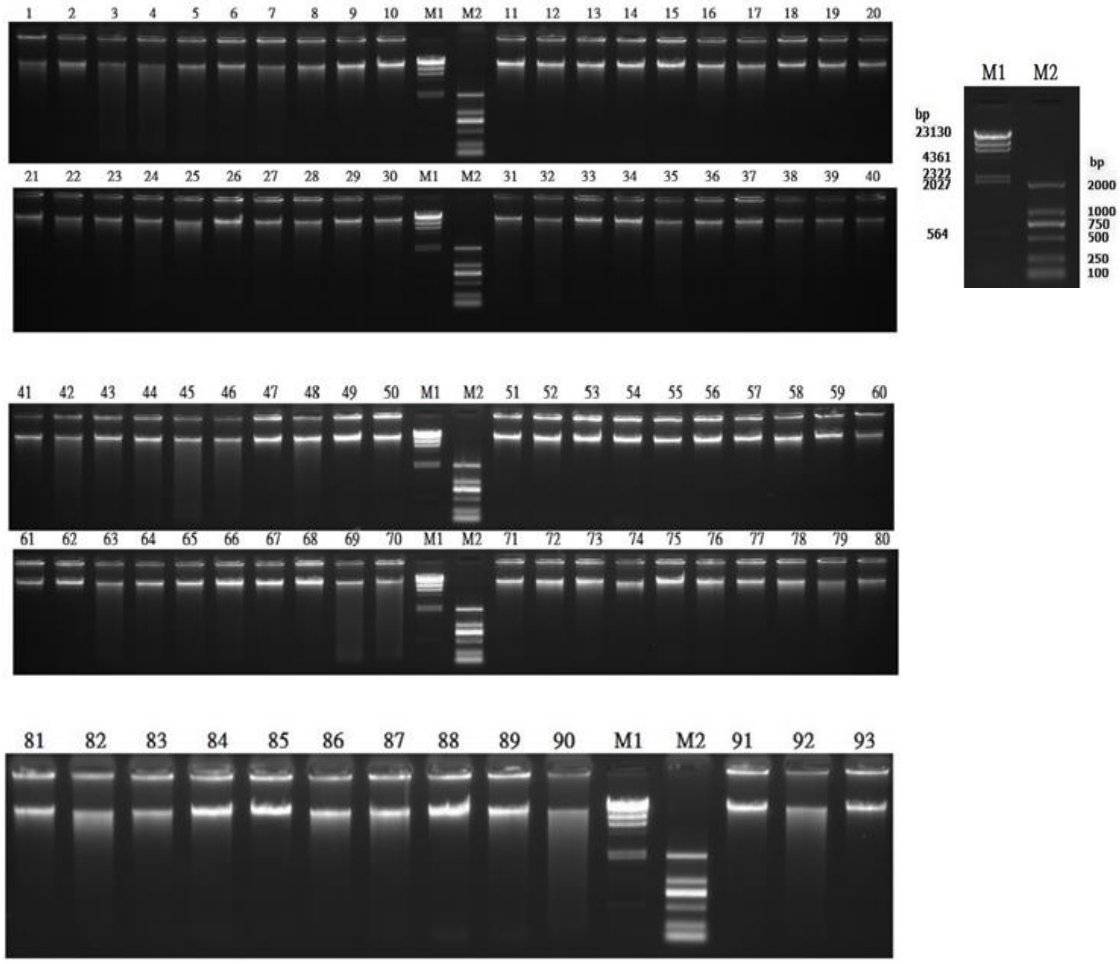


Plate 4.9: Gel electrophoretogram of doum palm DNA for GBS

M1. 1kb Plus molecular marker, **M2**. 100bp molecular marker

4.3.3 Single nucleotide polymorphism genotyping

A total of 3941 raw loci were produced by the *de_novo*-based assembly using Ipyrad software. Out of these, 2096 SNPs with a mean depth of 35.7 (minimum 10.47, maximum 217.45) were retained after filtering. On the other hand, 3.4 million loci were produced using reference-based assembly. Out of these, 23416 biallelic SNPs were obtained after filtering, with a mean depth of 3.5 (minimum 2, maximum 47.49).

Individual sequencing depth and mean depth indicate that the SNPs obtained from *de_novo* based assembly have greater depths than those obtained from reference based assembly. Both the *de_novo* based and reference based assemblies had minimal missing data percentages, with a maximum of 0.04 and 0.4 respectively. These VCF

quality statistics are shown for both the *de_novo* based assembly (Figure 4.7) and the reference based assembly (Figure 4.8). They include the mean depth, observed heterozygosity, depth per individual and missing data per individual

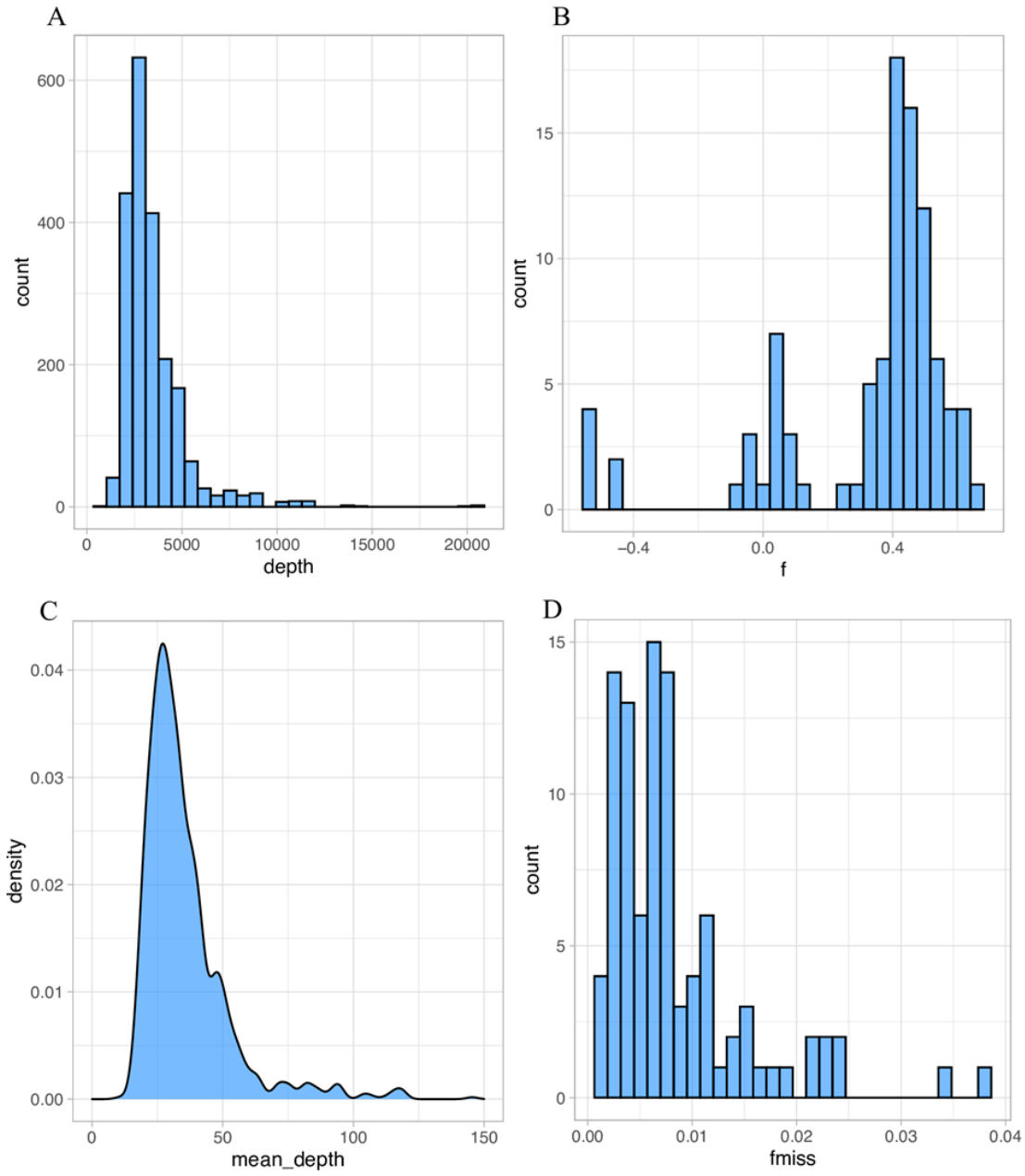


Figure 4.7: VCF SNP qualities of the *de_novo*-based assembly of GBS data

A. Depth per individual, **B.** observed heterozygosity, **C.** Mean depth and **D.** frequency of missing data per individual for *H. compressa* accessions from Kenya.

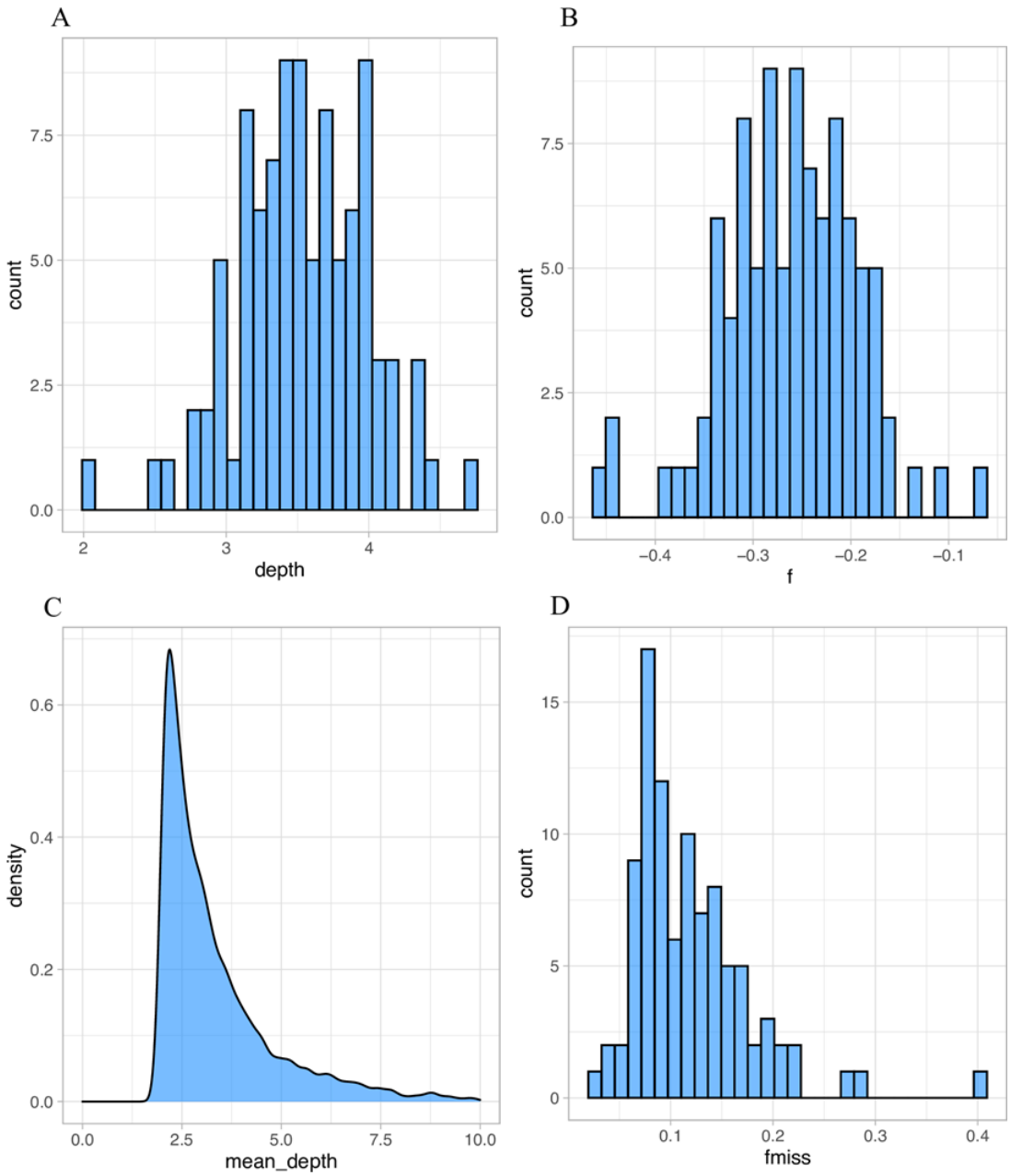


Figure 4.8: VCF SNP qualities of the reference-based assembly of GBS data

A. Depth per individual, **B.** observed heterozygosity, **C.** Mean depth and **D.** frequency of missing data per individual for *H. compressa* accessions from Kenya.

There were 1283 (61.2%) transition SNPs and 813 (38.8%) transversion SNPs using the *de_novo* based assembly with the following categories: A↔T type (174, 8.3%), A↔G type (651, 31.1%), C↔G type (222, 10.6%), C↔T type (632, 30.2%), G↔T type (225, 10.7%) and A↔C type (192, 9.2%). In the reference based assembly, there were 16598 (70.9%) transition SNPs and 6818 (29.1%) transversion SNPs with the

following categories: A↔C type (1684, 7.2%), A↔G type (8332, 35.6%), G↔T type (1673, 7.1%), C↔G type (1636, 7%), A↔T type (1825, 7.8%) and C↔T type (8266, 35.3%). For both assemblies, the A↔G and C↔T transition SNPs were the most common (Table 4.13). The ratio of transition SNPs to transversion SNPs (Ts/Tv) was 2.4 and 1.6 in the reference based assembly and in the *de_novo* based assembly respectively.

Table 4.13: Transition and transversion events of GBS analysis of *H compressa.*

SNP Type	*de_novo* assembly Total (Percentage)	Reference-based assembly Total (percentage)
Transitions	1283 (61.2)	16598 (70.9)
A↔G	651 (31.1)	8332 (35.6)
C↔T	632 (30.2)	8266 (35.3)
Transversions	813 (38.8)	6818 (29.1)
A↔C	192 (9.2)	1684 (7.2)
A↔T	174 (8.3)	1825 (7.8)
C↔G	222 (10.6)	1636 (7)
G↔T	225 (10.7)	1673 (7.1)

4.3.4 Population structure

Following STRUCTURE analysis, two main clusters were observed (optimal *delta K* at *K=2)* based on both the *de_novo*-based assembly (Figure 4.9A) and the reference-based assembly (Figure 4.9 B).

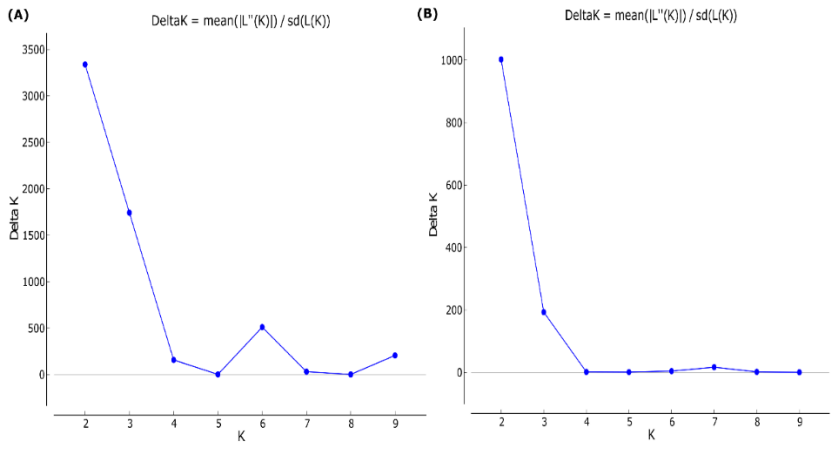


Figure 4.9: Optimal Delta k values inferred during STRUCTURE analysis of Kenyan *H. compressa* accessions
A delta k at *k*=2 for the reference-based assembly, **B.** *de_novo* based assembly

These two genetic structures (Cluster 1 and Cluster 2) are depicted in the STRUCTURE bar plot for the *de_novo*-based assembly (Figure 4.10A) and for the reference based assembly (Figure 4.10B). Cluster 1 contained all the Turkana accessions, which clustered together. While cluster 2 comprised accessions from Tharaka, Kwale and Tana River. Cluster 2 samples were all sampled along the Tana Basin (Figure 3.1) and Kwale county.

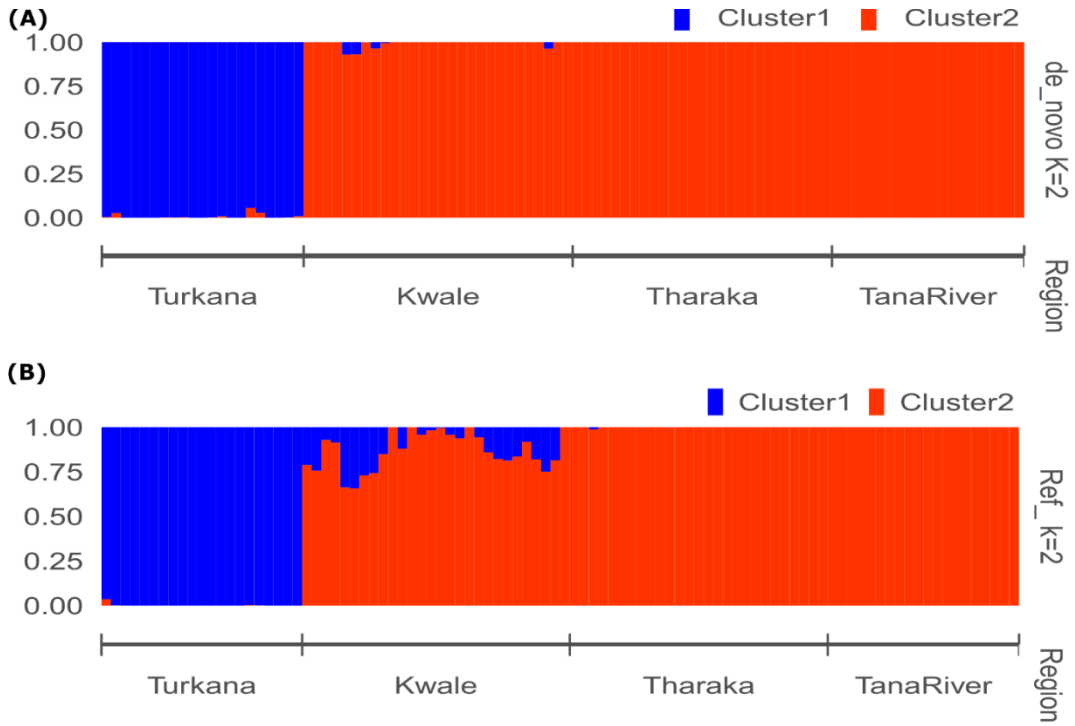


Figure 4.10: STRUCTURE bar plot of admixture model of 96 *H. compressa* accessions

A. STRUCTURE plot based on 2096 SNPs for the *de_novo* assembly, **B.** STRUCTURE plot based on 23416 SNPs using reference assembly. The accessions are divided into two clusters. A combination of different colors represents admixed populations.

The expected heterozygosity for cluster 1 (H_e=0.14) was lower than for cluster 2 (H_e=0.23) for the *de_novo*-based assembly. However, in the reference-based assembly, the two clusters had similar expected heterozygosity values (H_e=0.30). Cluster 1 had more genetic variation (*de_novo* F_{ST} =0.68 and reference-based F_{ST} =0.17) than Cluster 2 (*de_novo* F_{ST}=0.3 and reference-based F_{ST}=0.06). Using the

reference-based assembly, seven Kwale accessions had admixed ancestry (Table 4.14).

Table 4.14: Cluster assignment of *H. compressa* accessions based on STRUCTURE analysis.

Assembly method	Tharaka	Tana River	Kwale	Turkana	H_e	F_{ST}
Reference assembly						
Cluster 1	-	-	-	21	0.29	0.17
Cluster 2	27	20	21	-	0.30	0.06
Admixed	-	-	7	-	-	-
***De_novo* assembly**						
Cluster 1	-	-	-	21	0.14	0.68
Cluster 2	27	20	28	-	0.23	0.30

In the *de_novo* assembly, there were no admixed populations. According to the structure results, two gene pools best characterize the population structure of *H. compressa*. A minor peak identified at *K*=3 for both *de_novo* and reference-based assemblies (Figure 4.9) may indicate another informative *H. compressa* population grouping. STRUCTURE bar plot which depicts these three population groupings indicate that Tana River and Tharaka are clustered together (Figure 4.11).

STRUCTURE bar plots of *K*=2 to *K*=7 for the reference based and *de_novo* based assemblies are presented as appendix IX and X respectively. These STRUCTURE plots consistently show Turkana samples clustering on their own while Kwale samples consistently show a lot of admixture throughout the *K* values.

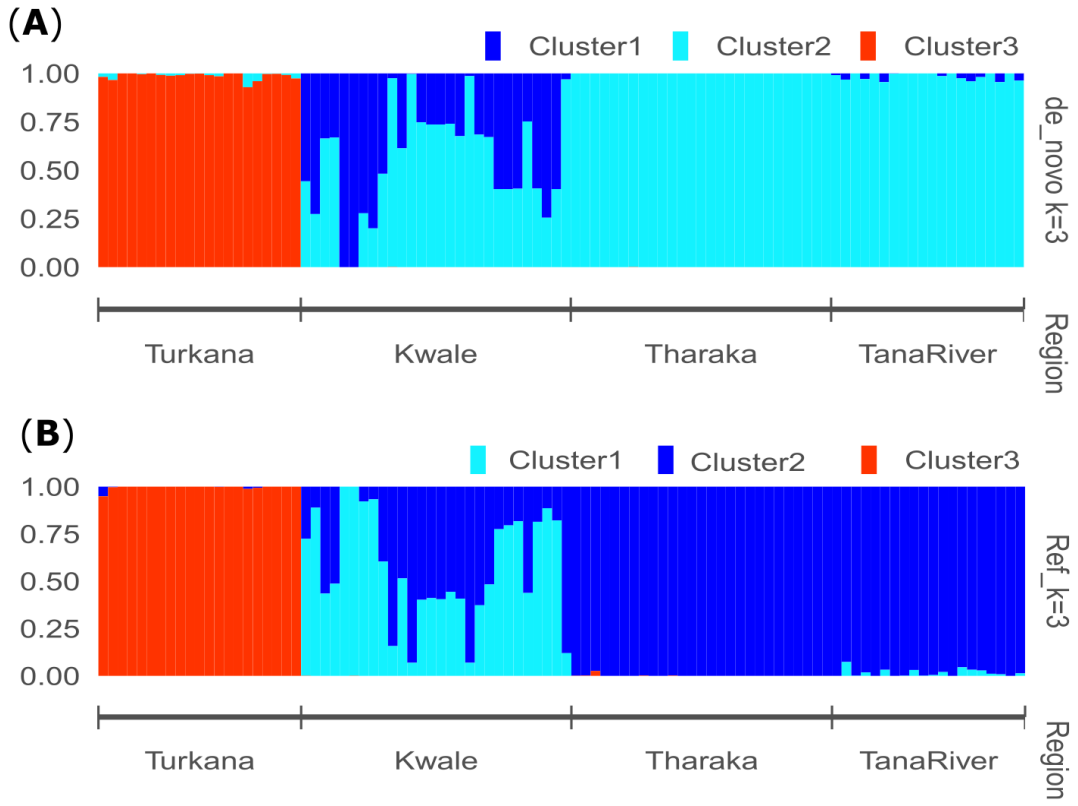


Figure 4.11: STRUCTURE bar plot of admixture model of 96 *H. compressa* accessions showing three clusters

A. based on 2096 Single Nucleotide Polymorphisms (SNPs) for the de-novo assembly and **B.** based on 23416 SNPs for the reference assembly. A combination of different colors represents admixed populations.

4.3.5 PCA and DAPC analysis

PCA plots were generated using both *de_novo* and reference-based assemblies, and both plots showed Tharaka Nithi, Kwale, and Tana River accessions clustering together (Figure 4.12).

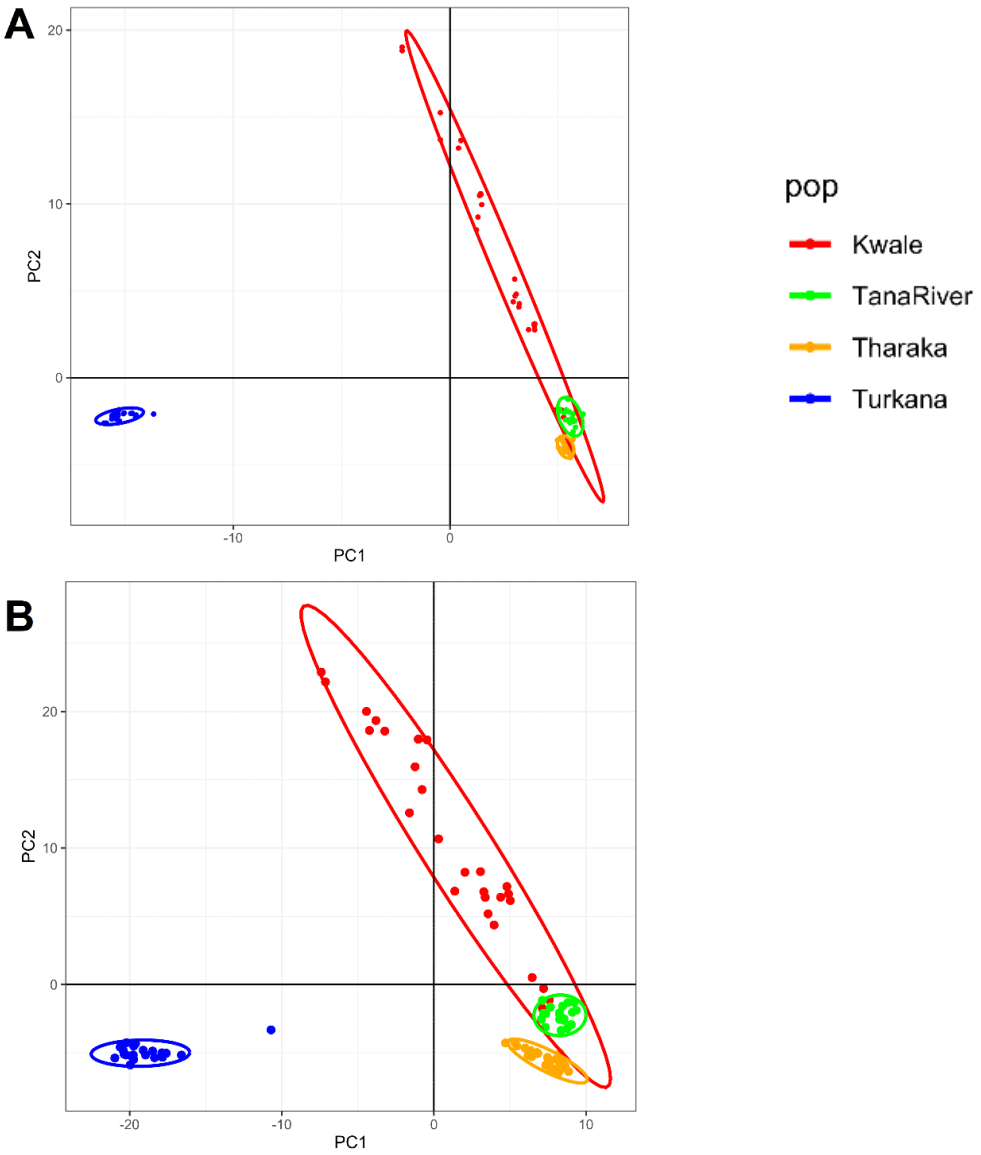


Figure 4.12: PCA of *H. compressa* Kenyan accessions based on GBS

A. using 2096 SNPs obtained from the *de_novo* assembly and **B.** using 23416 SNPs obtained from the date palm reference based assembly.

Similarly, DAPC analysis separated *H. compressa* accessions into two clusters, with Turkana samples on the right side of the DAPC vertical axis and the rest on the left. There was some overlap between Tana River and Kwale accessions, while Tharaka accessions were clearly distinct (Figure 4.13).

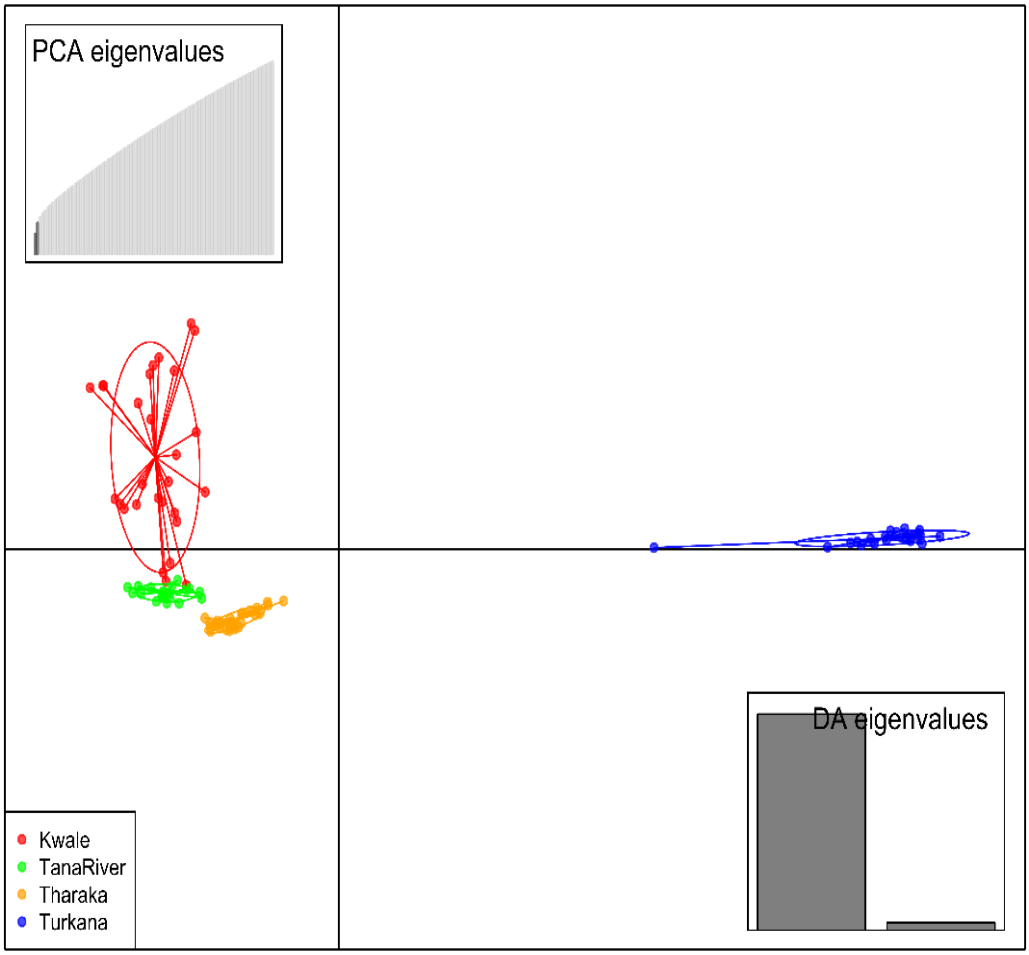


Figure 4.13: DAPC analysis of 23416 *H. compressa* SNPs derived from GBS analysis

Inertia ellipses represent clusters in different colors. Each dot on the circle represents an accession

Because both assembly approaches produced consistent findings in structure and PCA analysis, the composite plot, DAPC analysis and genetic diversity results are provided solely for the reference-based assembly SNP data. The assignment of population membership using the DAPC composite plot corroborated structure and PCA results. All of the accessions along the River Tana basin indicated admixture between Tharaka and Tana River accessions. Kwale had the most admixture (Figure 4.14). The DAPC data also revealed that there was no admixture of Turkana accessions and accessions from other regions.

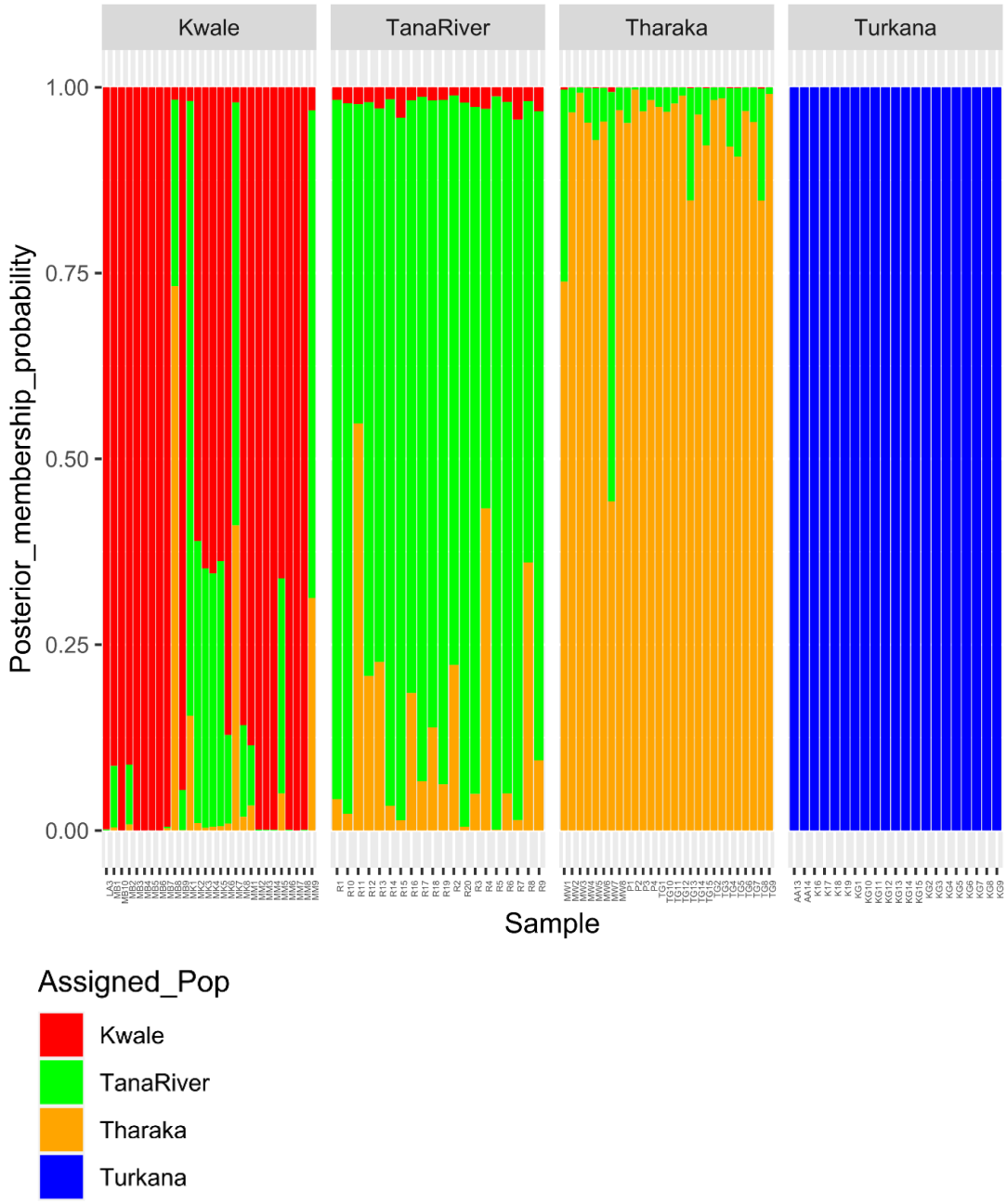


Figure 4.14: Composite plot of *H. compressa* Kenyan accessions

This composite analysis was done using 23416 SNPs derived from reference-based assembly and shows mixed ancestry between Kwale, Tana River and Tharaka. Each accession is a stacked bar chart with populations being shown in different colors.

4.3.6 Genetic diversity

The number of polymorphic sites, expected heterozygosity (H_e) or gene diversity, observed heterozygosity (H_o), F_{IS} and F_{ST} were all calculated for the four sampled

regions of Tharaka, Tana River, Kwale and Turkana. *Hyphaene compressa* accessions had moderate genetic variation (F_{ST} = 0.074, P ≤ 0.001). In all populations, the observed heterozygosity was greater than the expected heterozygosity (Table 4.15). All populations had negative F_{IS} value with Turkana having the lowest (-0.45). As indicated in Table 4.15, Kwale had the most polymorphic sites (11932), followed by Turkana (10,698). Tana River has the least diversity (H_e = 0.23, Polymorphic sites = 8370) of all of the regions studied (Table 4.15).

Table 4.15: Mean values of genetic diversity indices for *H. compressa* accessions

Genetic index	Region				Overall
	Tharaka	Turkana	Tana River	Kwale	
Number of polymorphic sites	9277	10698	8370	11932	23416
Observed heterozygosity (H_o)	0.45	0.47	0.46	0.44	0.404
Expected heterozygosity (H_e)	0.32	0.33	0.23	0.33	0.31
F_{IS}	-0.40	-0.45	-0.42	-0.37	-0.040
F_{ST}					0.074

F_{IS}-Inbreeding coefficient (detects inbreeding individuals relative to a sub population), F_{ST}- Fixation index (Measure of population differentiation).

The pairwise F_{ST} values varied from 0.025 (Tharaka and Tana River) to 0.105 (Turkana and Tana River). The F_{ST} levels were higher in Turkana and Tana River samples (Table 4.16). The pairwise F_{ST} between Tharaka and Tana River was the lowest, indicating gene flow between the two areas. AMOVA revealed that populations from Turkana, Tharaka, Tana River and Kwale differed slightly (P ≤ 0.001, Table 4.17). Within-population variation was higher (92.7%) than among population variation (7.3%).

Table 4.16: Pairwise F_{ST} values of Kenyan populations of *H. compressa*

Population	Turkana	Kwale	Tharaka	Tana River
Turkana				
Kwale	0.07952			
Tharaka	0.09795	0.03629		
Tana River	0.10541	0.03329	0.02505	0.00

Table 4.17: Analysis of molecular variance among *H. compressa* accessions

Source of variation	Sum of squares	Variance components	Percentage variation	P-value
Among populations	42728.129	266.91865	7.32597	0.00
Within populations	558905.557	3376.54026	92.67403	
Total	601633.686	3643.45891	100	

4.3.7 Phylogenetic analysis

H. compressa accessions were grouped by region in the neighbor net network. Tana River, Kwale and Tharaka accessions clustered together, whereas Turkana samples were isolated from the rest (Figure 4.15). Some Kwale accessions grouped closely with Tana River accessions, whereas others clustered with Tharaka accessions. The UPGMA phylogenetic tree revealed two primary clusters with Turkana accessions in one and the remainder of the accessions in the other (Figure 4.16).

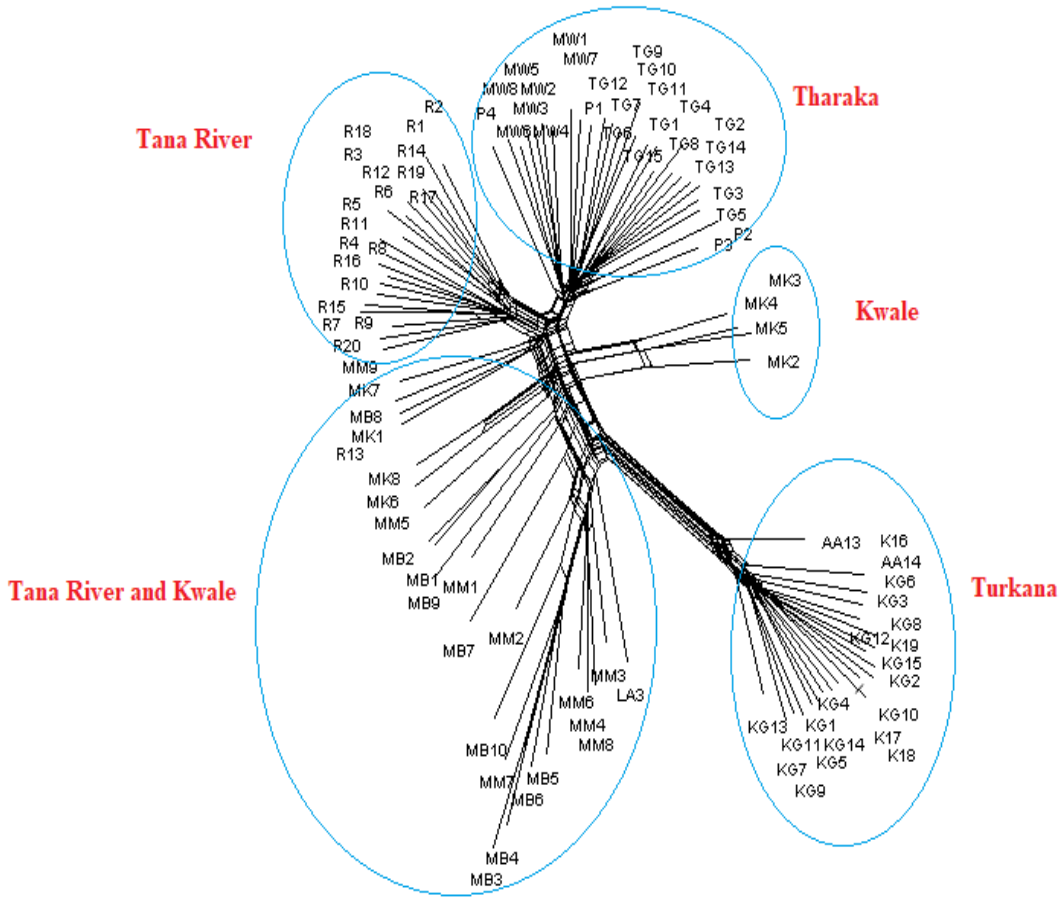


Figure 4.15: Splitstree generated from GBS analysis of Kenyan *H. compressa*
The tree was generated using 23416 SNPs using SplitsTree version 4.17.0.

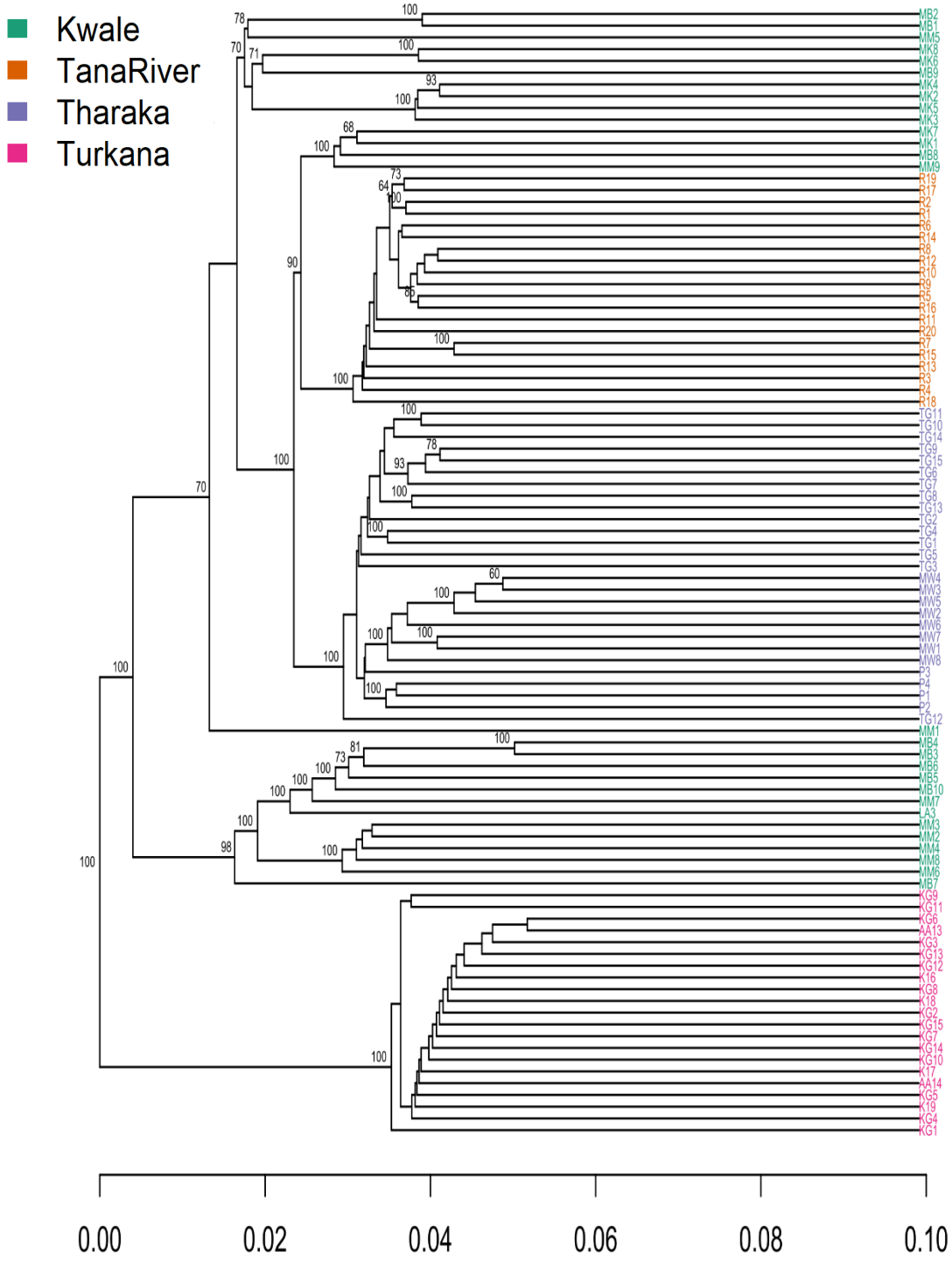


Figure 4.16: Unrooted UPGMA distance tree inferred from *H. compressa* GBS data

The UPGMA tree was constructed using 23416 SNPs based on the reference based assembly of *H. compressa* accessions

4.3.8 Migration rates among *H. compressa* accessions along the River Tana basin

The largest gene flow (m=139.1) was seen between Tharaka and Tana River accessions, followed by Kwale to Tharaka (m=102.7), Tana River to Kwale (m=63.1), Tana River to Tharaka (m=59.9) and Kwale to Tana River (m=57.7). These findings suggest that gene flow along the River Tana basin was mostly asymmetric (Figure 4.17).

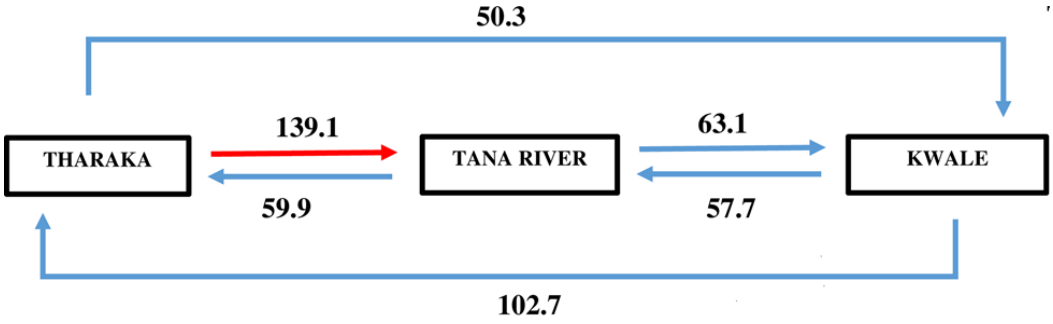


Figure 4.17: Migration rates between Tharaka, Tana River and Kwale

The migration distances were calculated using MIGRATE-n software. The highest migration rate from Tharaka to Tana River is highlighted in red.

4.4 Salinity induced transcriptomics

4.4.1 Morphological measurements of *H. compressa*

The chlorotic score recorded on a scale of 1-5 revealed that with increase in salinity stress, there was increase in chlorosis across all accessions from the three regions. At the end of salinity stress experiments, Tana River accessions showed the least chlorotic score (Plate 4.10; Figure 4.18).

There was no difference in chlorotic scores between salt imposition and week 3 of salinity stress levels of 100mM and 200mM in all accessions. However, at 300 mM salinity, there was a significant increase in chlorosis as early as week 3 ($P<0.001$). There was no observed necrosis or complete plant death in any of the accessions exposed to salinity. At 300mM, there was significantly increased chlorotic damage to the plant between salt imposition and the end of salinity stress in all accessions ($p<0.001$) as shown in Figure 4.18.

Plate 4. 10: Salinity symptoms on *H. compressa* accessions from different ASALs

A Control plants from Tana River, **B** 100mM stressed plants from Tana River**, C** 200mM salinity stressed plants from Tana River and **D**, 300Mm salinity stressed plants from Tana River, **E** Control plants from Turkana, **F** 100mM stressed plants from Turkana**, G** 200mM salinity stressed plants from Turkana**, H**, 300mM salinity stressed plants from Turkana, **I** Control plants from Tharaka, **J** 100mM stressed plants from Tharaka**, K** 200mM salinity stressed plants from Tharaka and **L**, 300mM salinity stressed plants from Tharaka.

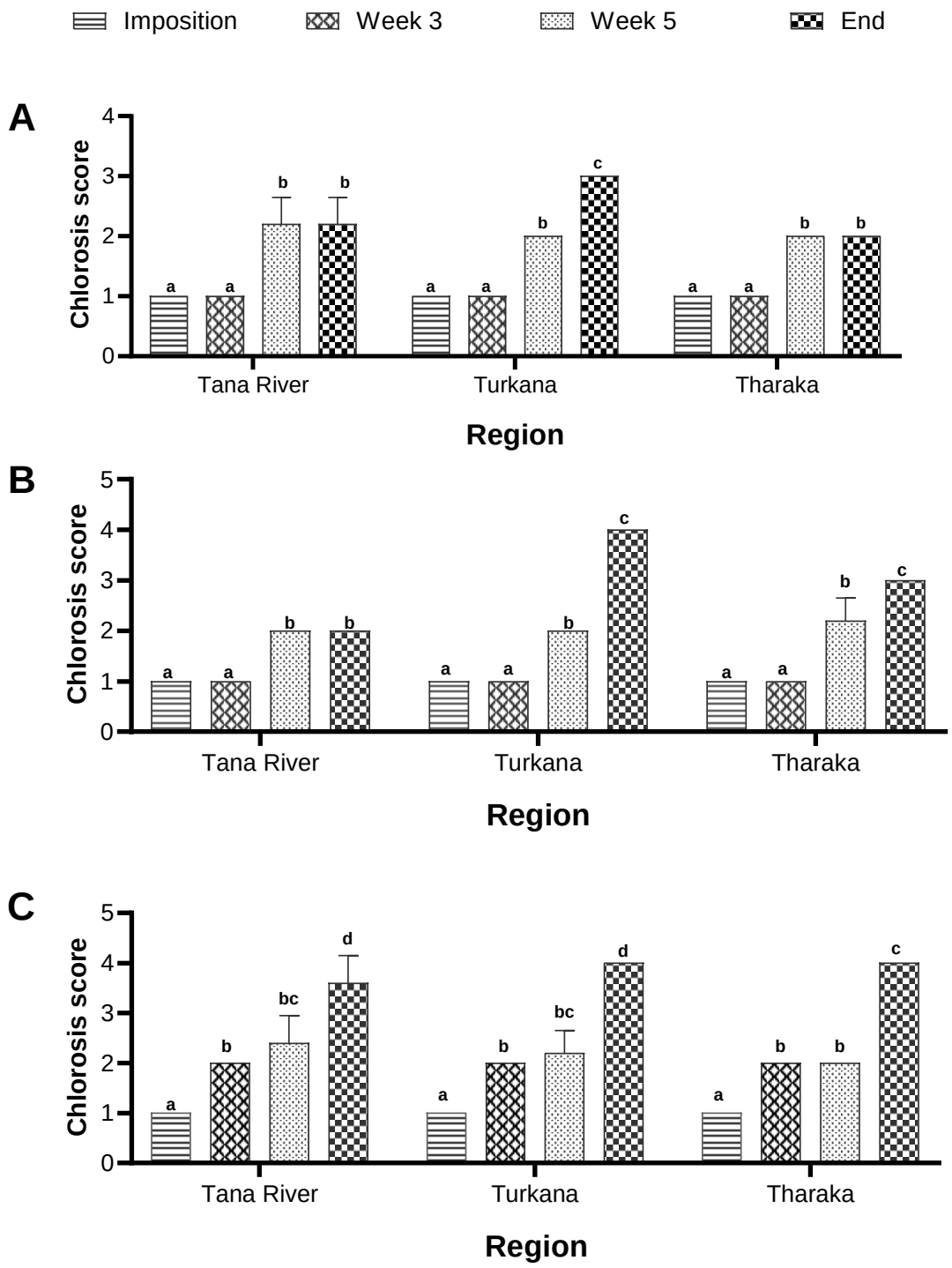


Figure 4.18: Chlorotic score on doum palm accessions exposed to salinity

A. Accessions exposed to 100mM salinity level, **B.** accessions exposed to 200mM salinity level, **C.** accessions exposed to 300mM salinity level. Bars with different letters indicate means with significant differences using Tukey's post hoc test ($p<0.05$).

Tana River showed the greatest shoot growth at the control and also at 100mM salinity levels. These accessions also showed minimal negative growth due to salinity stress. Accessions from Tharaka showed the greatest reduction in shoot length due to increment of salinity stress (Figure 4.19). However, there was no statistical difference in shoot length between the time of salt imposition and at the end of the salinity stress in all the accessions.

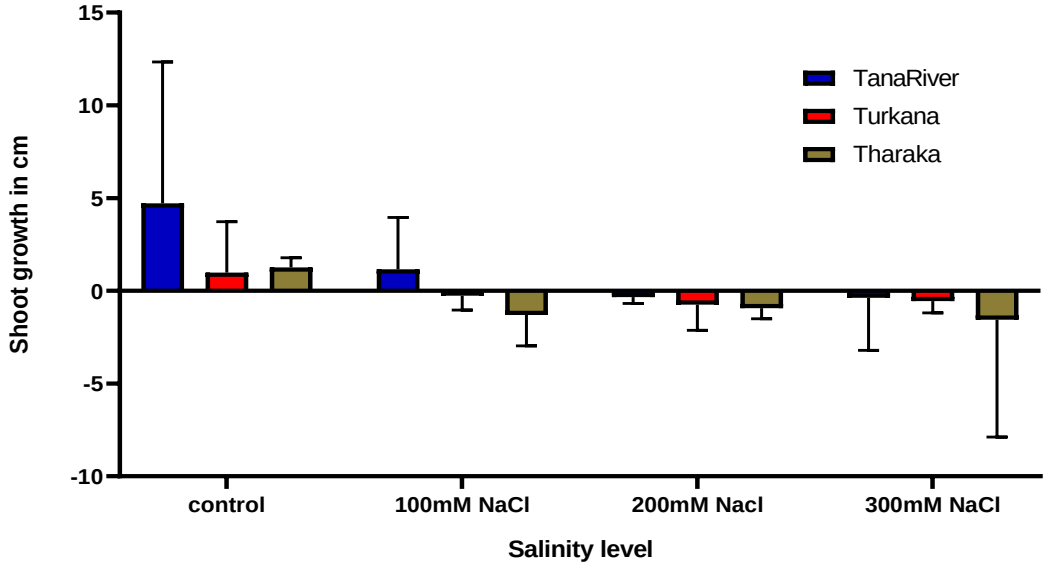


Figure 4.19: Effect of salinity on shoot growth of *H. compressa* accessions.

Decreased leaf length was observed with increase in salinity stress (Figure 4.20). Leaf 3 being the youngest upper leaves, little differences were observed in leaf length between controls and salinity stressed accessions. However, leaves 1 being the oldest base leaves, slightly large differences in leaf lengths between controls and salinity stress were observed. However, all the leaf length differences observed in this study were not statistically significant based on ANOVA test.

H. compressa shoot dry biomass was shown to significantly decrease with increase in salinity levels ($p < 0.0001$). Tana River accessions accumulated the most biomass at all levels of salinity while Tharaka accessions accumulated the least (Table 4.18). Significant reductions in biomass at 300mM salinity level compared to the control were significant in Tana River ($p = 0.0001$), Turkana ($p = 0.0008$) and Tharaka ($p =$

0.010). At 100mM salinity level reductions in dry shoot biomass was not significant in all the accessions (Table 4.18). Reductions in root biomass was observed with increment in salinity stress across accessions from all regions. However, these reductions in root biomass were not statistically significant except for Tharaka accessions where significant reduction in root biomass was recorded at the 300mM salinity level (Table 4.18).

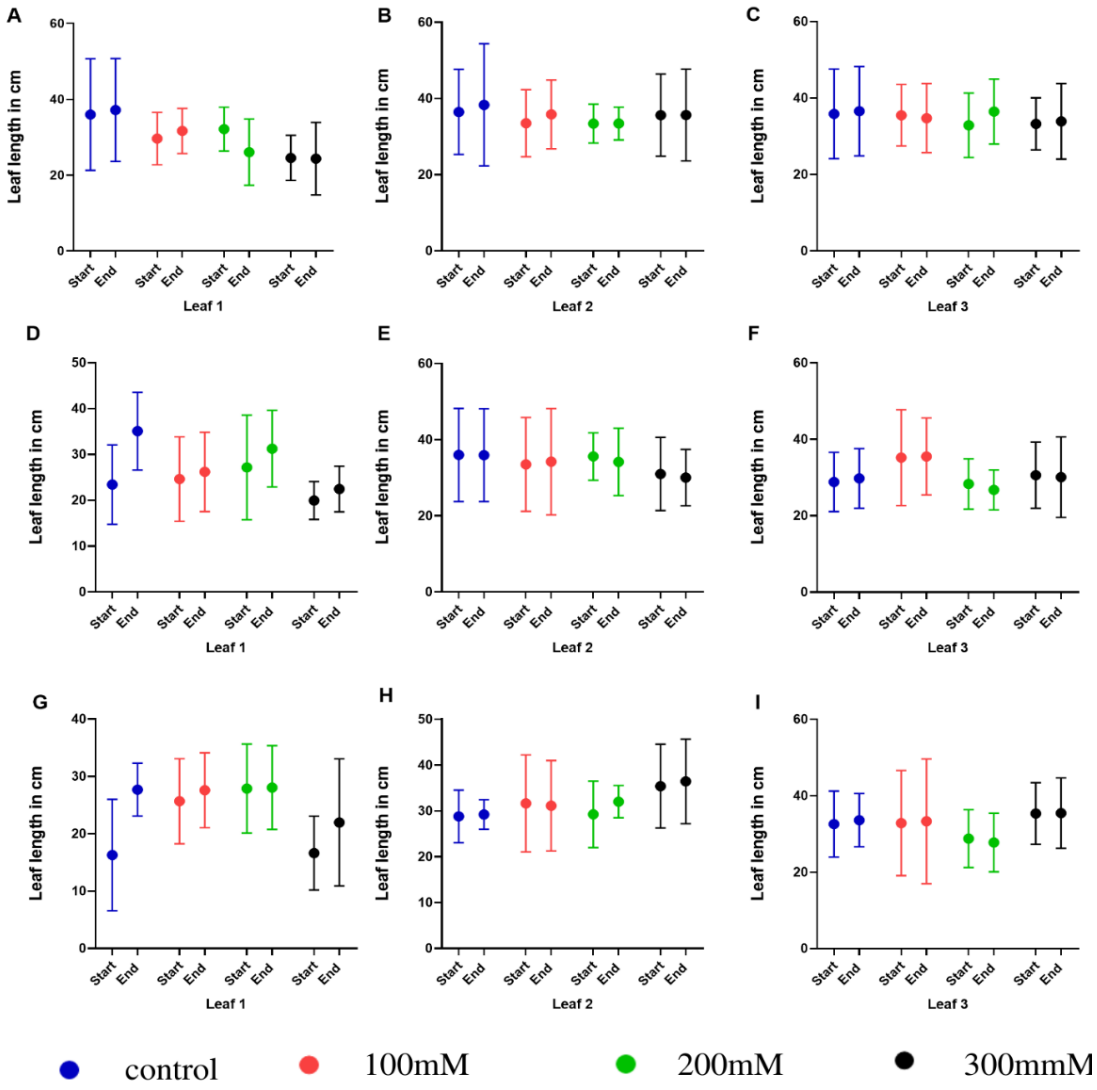


Figure 4.20: Leaf length of *H compressa* leaves at start and at the end of salinity stress.

A, B. and **C.** are leaf lengths for Tana River samples, D, E and F are leaf lengths for Turkana while G, H and I are leaf lengths of Tharaka accessions

Table 4.18: Effect of salinity on the mean dry biomass in *H. compressa*

Salinity level	Tana River		Turkana		Tharaka	
	Shoot Mean ±sd	Root Mean ±sd	Shoot Mean ± sd	Root Mean ±sd	Shoot Mean ±sd	Root Mean ±sd
Control	23.1 ±1.1[a]	94.9±21.8[a]	18.3±6.21[a]	69.3±17.1[a]	16.4 ±2.9[a]	78.2±6.6[a]
100mM	16.7 ±2.9[a]	97.1±46.1[a]	12.8±2.67[a]	76.1±1.8[a]	10.5±1.6[a]	79.4±3.4[a]
200mM	14.9±1.9[b]	84.3±20.8[a]	9.7 ±2.32[b]	76.4±29.1[a]	9.3 ± 3.5[b]	78.2±19.7[a]
300mM	12.5± 4.2[c]	86±4.7[a]	9.2 ±1.31[c]	45.7±1.6[a]	7.4 ± 0.6[c]	42.0±7.7[c]

Different letters denote significance difference at p<0.05 based on tukey post hoc test.

4.4.2 Physiological parameters of *H. compressa*

With increase in salinity levels, reduction in water content in *H. compressa* accessions was observed in all the regions sampled. All the accessions had higher water content levels in non-saline conditions. There were significant effects of salinity on water content (Figure 4.21). All the accessions had significant reductions in water content at the 300mM salinity stress level.

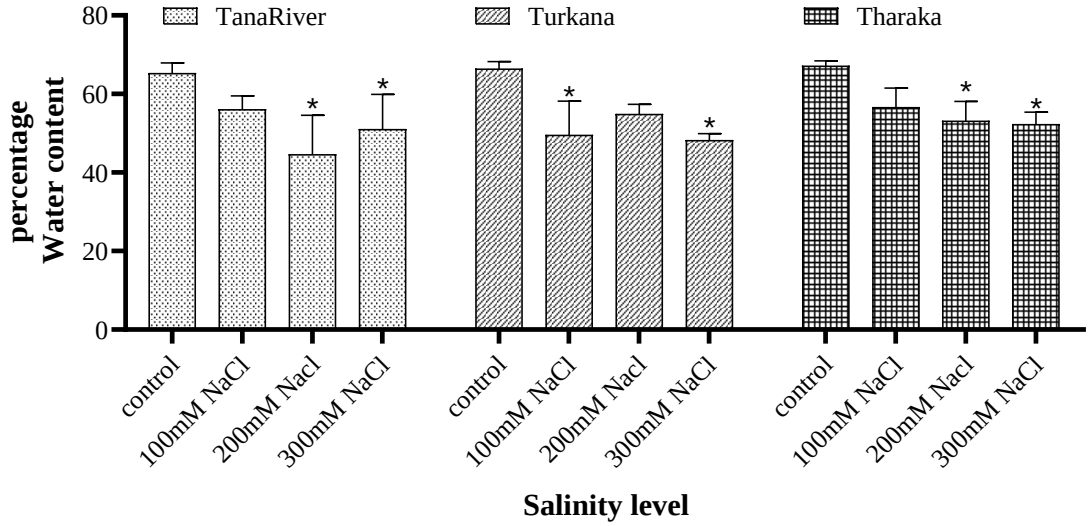


Figure 4.21: Water content in shoots of *H. compressa* exposed to salinity.
Significant levels at each group at p < 0.05 are shown using asterisks.

Salinity was shown to significantly reduce chlorophyll a, (p value is <0.0001), chlorophyll, b (p value = 0.029) and carotenoids (p value <0.0001) as shown in Figure 4.22A, B and C respectively.

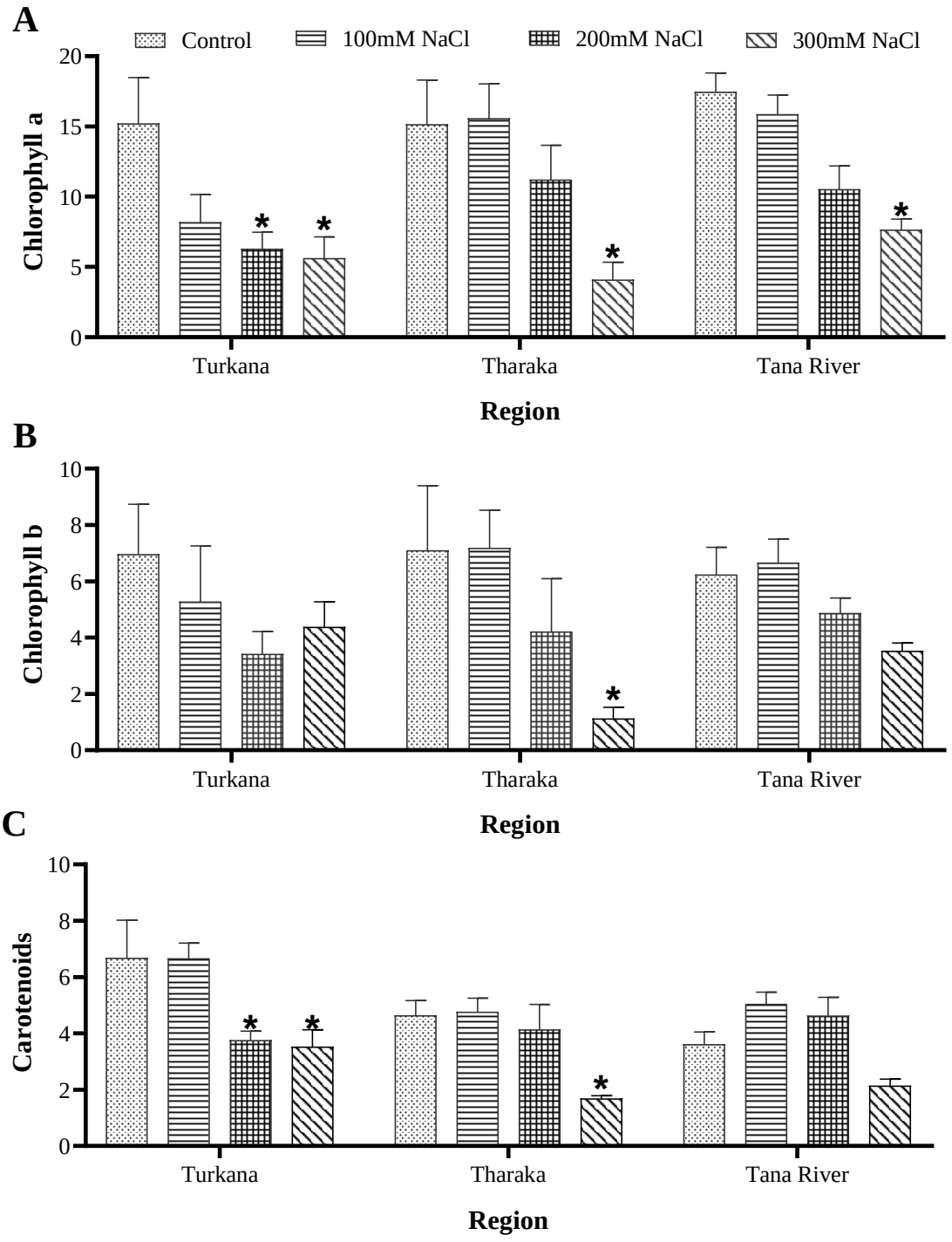


Figure 4.22*: H. compressa* contents of Chlorophyll a, b and carotenoids in leaves after 8 weeks of salinity treatment.

Significant levels at each group at $p < 0.05$ are shown with asterisks.

The levels of chlorophyll a between the control and the 300mM salinity levels was significantly reduced in all the three regions. Chlorophyll b on the other hand was only significantly reduced between the control and 300mM salinity treated accessions from Tharaka Nithi (P value = 0.014). Carotenoid levels between control and salinity stressed doum palms were not significantly reduced in Tana River accessions (P value =0.39, 0.67 and 0.36 for the control vs 100mM, 200mM and 300 mM NaCl treated respectively). At 100 and 200mM salinity levels, carotenoid levels were observed to be more than the control palms in Tana River accessions. Even though these differences were not statistically significant,

SPAD readings were higher in the control groups for all regions (Figure 4.23). There was gradual decrease in SPAD readings at 300mM salinity level from week 4 to 6 in all the regions. However, these differences were not significant in accessions from Turkana and Tana River. Tharaka accessions had significant differences in SPAD readings at 100mM (p = 0.0103), 200mM (p =0.0018) and 300mM (P=0.0298).

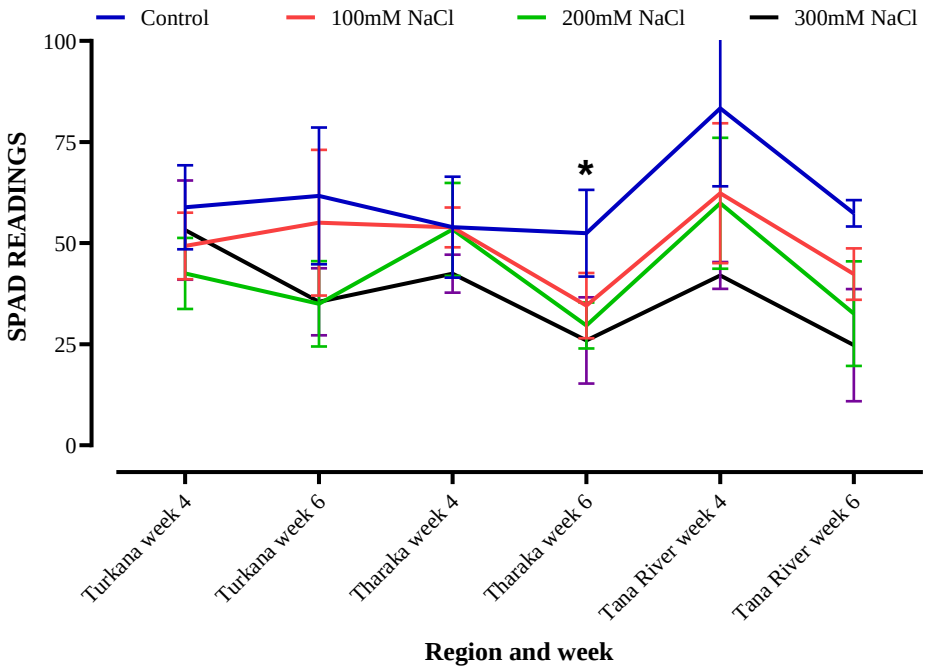


Figure 4.23: SPAD Readings in leaves of *H. compressa* at week 4 and week 8 of salinity treatment

Significant levels at each group at $p < 0.05$ are shown using asterisks

The number of stomata on abaxial leaves for the control (Figure 4.24A), 100mM, 200mM (Figure 4.24B) and 300mM (Figure 4.24C) were obtained. There were more stomata on the control samples of Tana River and Turkana with gradual decrease in number of stomata at higher salt concentrations (Figure 4.24D). Despite the recorded differences in stomatal density, they were not statistically significant.

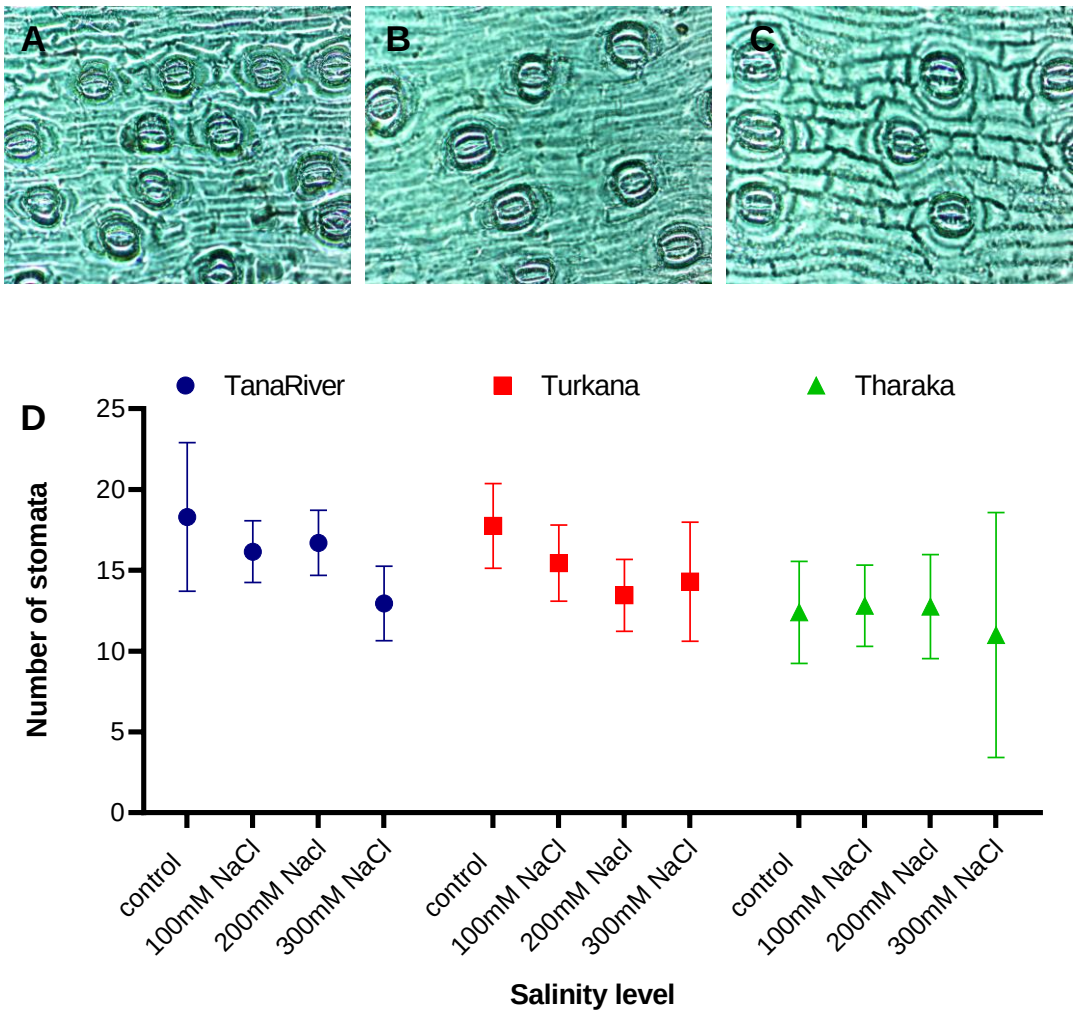


Figure 4.24: Stomatal density in abaxial *H. compressa* leaves exposed to salinity

A. Abaxial number of stomata in control, **B.** Abaxial number of stomata in200mM salinity, **C.** Abaxial number of stomata in 300mM salinity, **D.** stomatal density distribution per region and per salinity treatment

4.4.3 Na^+ and K^+ content in shoots and roots

The ash content determined on dry biomass in *H. compressa* accessions exposed to salinity stress indicated that with increase in salinity stress there was increase in percentage ash content in both the shoots (Figure 4.25A) and roots (Figure 4.25B) across all regions.

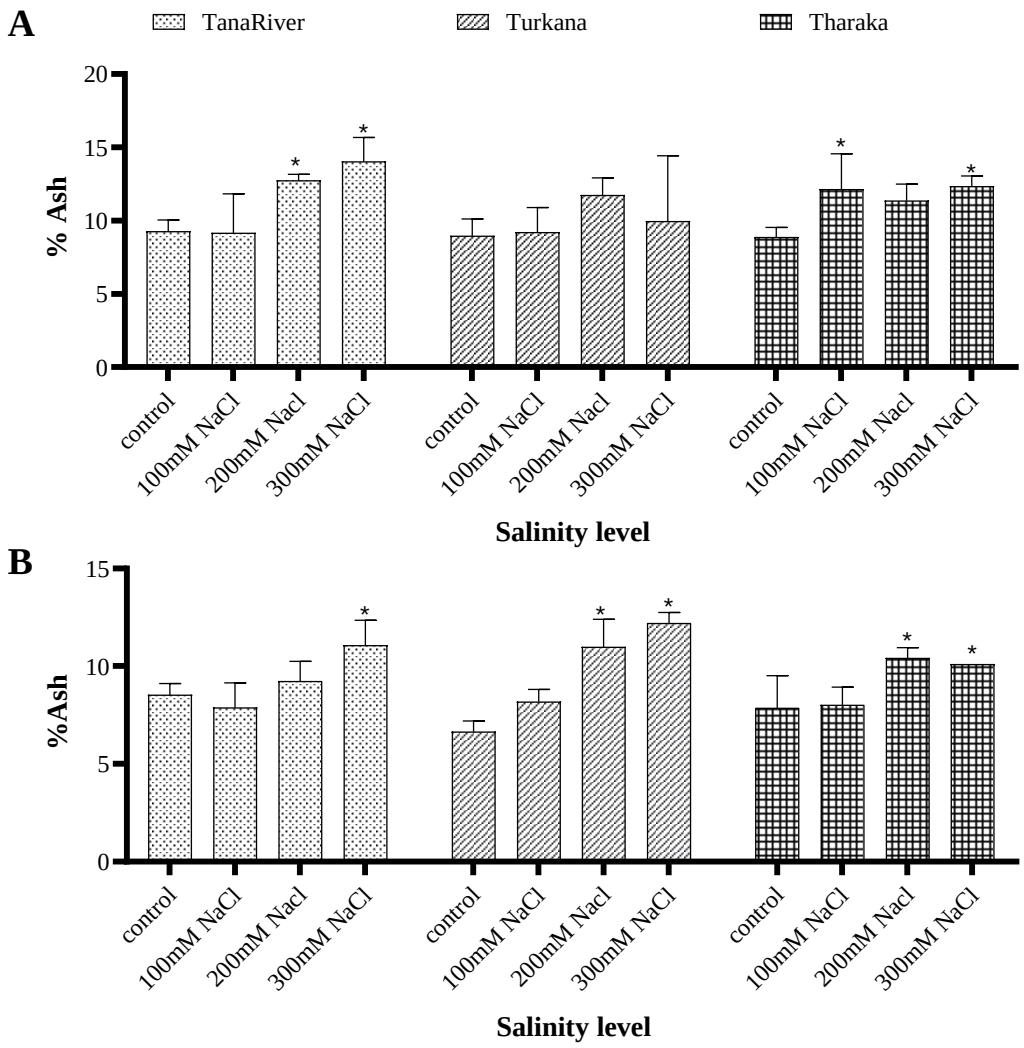


Figure 4.25: Ash content in *H. compressa* dry biomass exposed to salinity stress

A Percentage ash content in shoots, **B** percentage ash content in roots. Significant levels at each group at $p < 0.05$ are shown using asterisks

The differences in percent shoot and root ash per region were not statistically significant despite Tana River seeming to have accumulated the most ash in the shoots. There was statistically significant increase in percent shoot ash content between the

controls and 300mM salinity levels in Tana River and Tharaka accessions. Likewise, root ash content was seen to significantly increase between the controls and 300mM salinity level in accessions from all regions.

Na^+ accumulation was seen to increase with increase in salinity stress in both shoot and root samples. Tana River accessions had higher Na^+ shoot accumulation compared to Tharaka and Tana River. Despite this, there were no significant statistical differences in Na^+ accumulation in shoots among the different regions. There were statistically significant differences in accumulation of Na^+ between controls and the salinity stressed plants across accessions from all regions. (Figure 4.26).

Similarly, root Na^+ accumulation was seen with increase in salinity stress. Tana River accumulated less root Na^+ compared to accessions from Turkana (P value=0.048) and Tharaka (p=0.011). The differences in Na^+ accumulation between accessions from Turkana and Tharaka were not statistically significant (p= 0.35) as shown in Figure 4.26.

At 300mM salinity level, Tana River accessions increased their K^+ shoot accumulation (Figure 4.27). On the other hand, Turkana and Tharaka K^+ shoot accumulation declined with increase in salinity stress. The differences in K^+ shoot accumulation were not statistically significant between regions (p=0.98) or salinity levels (p=0.13). Tana River and Turkana accessions increased their K^+ root accumulation with increasing salinity while Tharaka steadily reduced their root K^+ accumulation with increasing salinity. Root K^+ accumulation was significantly different between regions (p=0.023) but not between salinity levels (p=0.58). Tana River root K^+ accumulation was significantly different from Tharaka (p=0.03) and not Turkana (p=0.07). There was no significant statistical difference in root K^+ accumulation between Turkana and Tharaka (p=0.94).

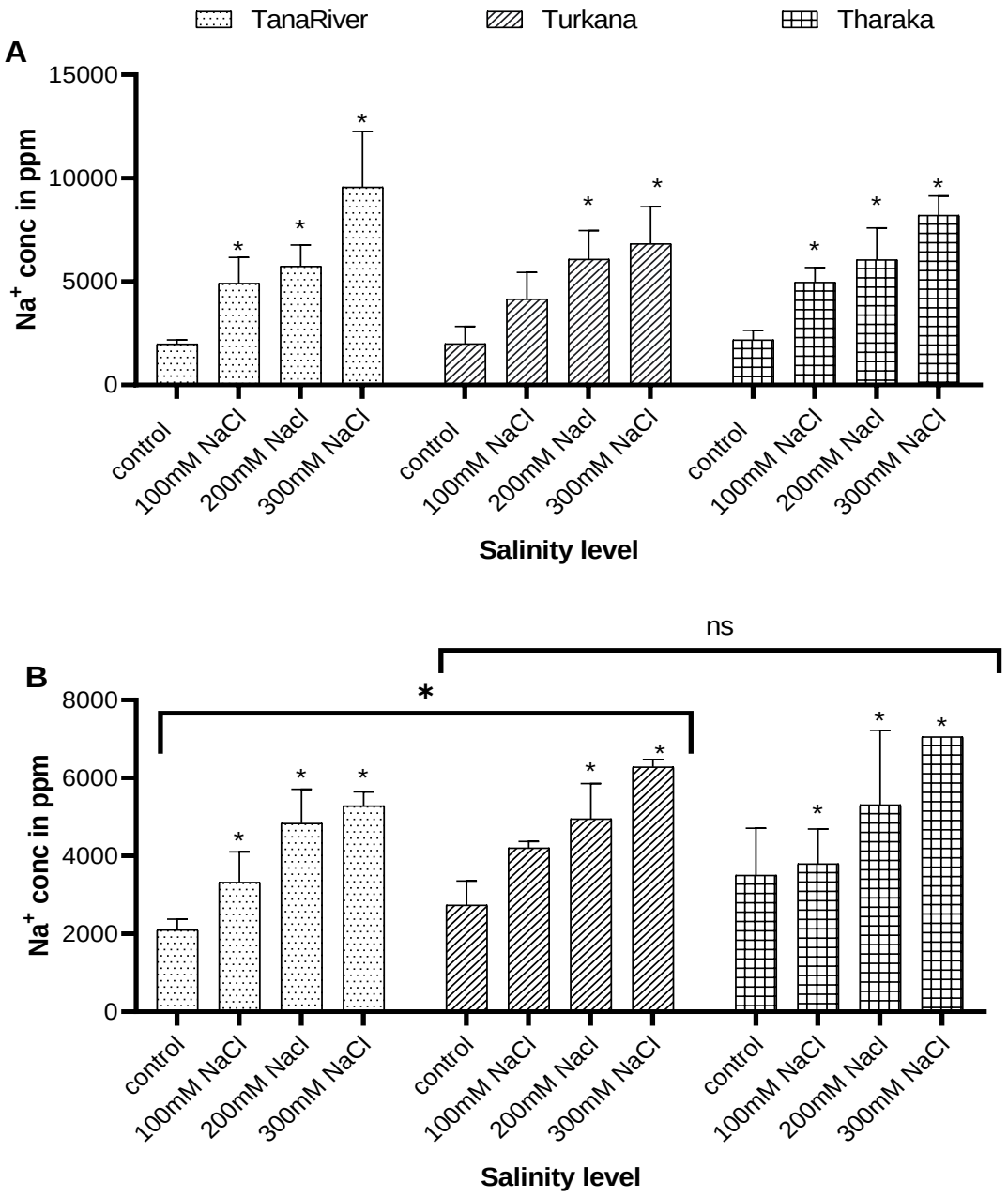


Figure 4.26: Effect of salinity on sodium ion accumulation in *H. compressa*

A. Effect of salinity on sodium ion accumulation on shoots, **B**. Effect of salinity on sodium ion accumulation on roots. Significant levels at each group at $p < 0.05$ are shown using asterisks.

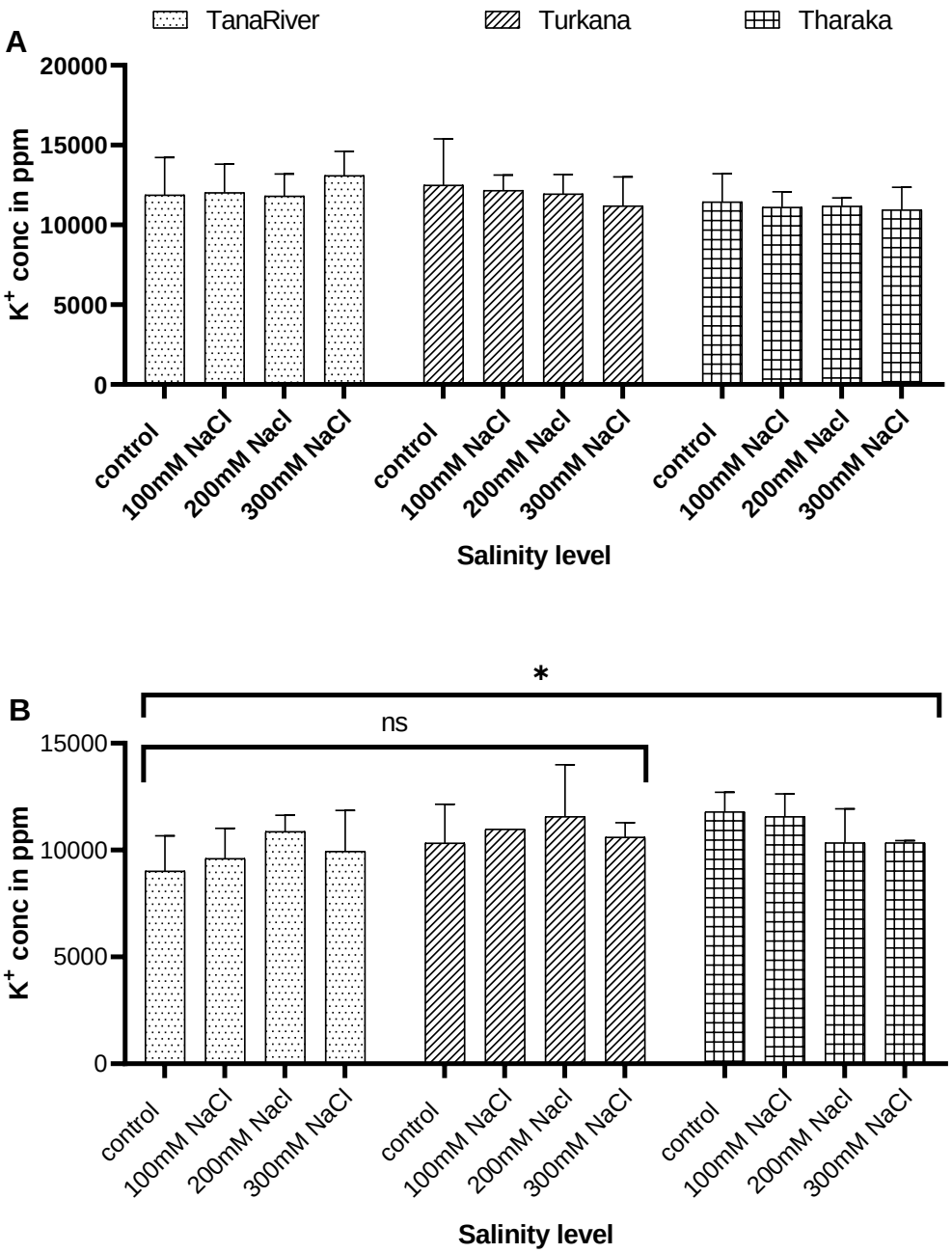


Figure 4.27: Effect of salinity on Potassium ion accumulation in *H. compressa*

A. Accumulation of potassium ion in shoots, **B.** Accumulation of potassium ion in roots. Significant levels at each group at $p < 0.05$ are shown using asterisks

4.4.4 Proline accumulation in leaves at different salinity levels

H. compressa from the three different regions had varied proline accumulation rates at different salinity levels as shown in Figure 4.28. Proline accumulation in the controls was significantly lower than any salt treatment group in all the samples from the 3 regions ($p < 0.0001$). Samples from Turkana showed the highest accumulation of proline at 100mM salinity stress level compared with the same salinity level in Tharaka and Tana River samples. Tana River samples on the other hand, showed high accumulation of proline at 300mM salinity compared with proline accumulation at the same salinity level in Turkana ($p = 0.17$), Tharaka ($p=0.13$) and Tana river control ($p < 0.0001$). The difference in proline accumulation within the different regions was not significant ($P = 0.34$). However, the differences in the accumulation of proline within the different salinity levels was significant $p < 0.0001$.

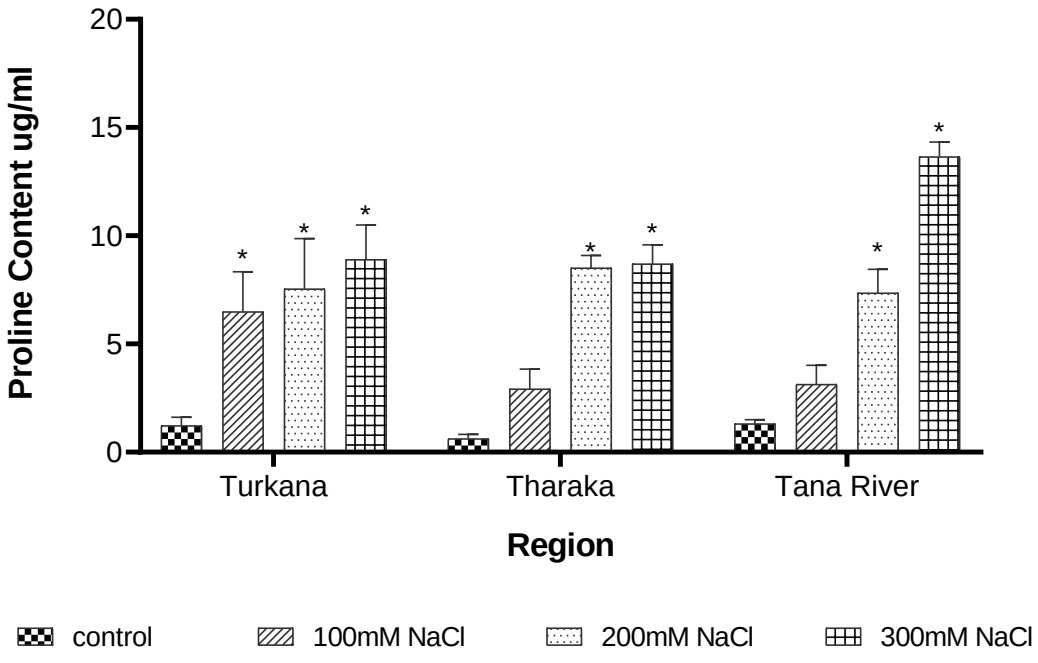


Figure 4.28: Proline accumulation in *H. compressa* leaves exposed to salinity stress

4.4.5 Salinity tolerance by *H. compressa* from Tana River, Turkana and Tharaka

Salt tolerance Index and SII for the three accessions are shown in Table 4.19 and Figure 4.29. Tana River accessions had the highest STI at all levels of salinity that

was significantly different from Tharaka (P= 0.0242). There was no significant difference in STI between Tana River and Turkana (p= 0.0541) and STI between Turkana and Tharaka (p= 0.6228) accessions. Based on STI, Tana River (STI 0.56) and Turkana (STI 0.56) accessions were moderately tolerant to salinity at 300mM, whereas Tharaka (STI 0.45) accessions were moderately sensitive (Table 4.19). These results confirm that *H. compressa* is moderately to highly tolerant to salinity stress.

Table 4.19: STI and SII in three *H. compressa* accessions at various salinity levels.

Salinity level	Tana River		Turkana		Tharaka	
	STI	SII	STI	SII	STI	SII
100mM NaCl	0.75	0.25	0.70	0.30	0.64	0.36
200mM NaCl	0.67	0.33	0.53	0.47	0.57	0.43
300mM NaCl	0.56	0.44	0.50	0.50	0.45	0.55

STI (Salt Tolerance Index), SII (Salt Injury Index)

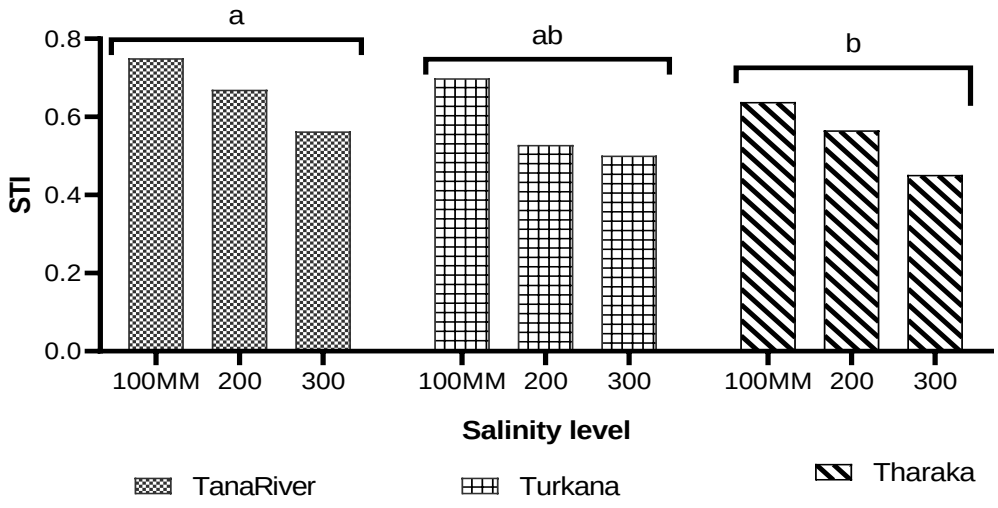


Figure 4.29: Salt Tolerance Index at various salinity levels in *H. compressa* accessions.
Different letters denote significant difference among the accessions.

Overall, based on the morphological, physiological, ion content and proline results coupled with STI, the most tolerant accessions to salinity stress were from Tana River. Tana River accessions were further processed for RNA extraction and sequencing to identify differentially expressed genes due to salinity stress.

4.4.6 RNA Sample quality control (QC)

RNA was isolated from 6 samples (3 replicates of control and 3 replicates of 300mM stressed plants) from Tana River. These RNA samples passed sample QC with RNA Integrity Number equivalent (RINe) above 7.1 and proceeded to RNA sequencing (Table 4.20). Sample QC results of the RNA samples using the Agilent 4200 is shown in Appendix XI.

Table 4.20: *H. compressa* RNA sample QC results

NO	Sample Name	Concentration (ng/µl)	Volume (µl)	Total Mass (µg)	RINe	28S/ 18S	Test result
1	TRC11	446	20	8.92	8.2	1.6	Qualified
2	TRC12	568	22	12.5	8.1	1.8	Qualified
3	TRC21	682	21	14.32	7.5	1.5	Qualified
4	TR311	227	24	5.45	7.1	1.6	Qualified
5	TR322	268	19	5.09	7.5	1.4	Qualified
6	TR323	356	22	7.83	7.3	1.5	Qualified

TR311, TR322, TR323 (Salinity stressed Tana River samples), TRC11, TRC12, TRC21 Tana River Control samples

4.4.7 RNA Sequencing

A total of 44.69, 45.07 and 44.94 million reads per control plants (replicate 1, 2 and 3) respectively and 44.89, 45.1 and 45.01 million reads for the salt- treated samples (replicate 1, 2 and 3) respectively were obtained (Table 4.21). The minimum clean reads ratio was 94.14%. Multiqc results are summarized in Appendix XII. Majority of the sequences had quality scores higher than the Q30 level.

Table 4.21: Quality metrics of RNA sequencing results of *H. compressa*

Sample	Treatment	Total Clean Reads (M)	Total Clean Bases (Gb)	Clean Reads Q20(%)	Clean Reads Q30(%)	Clean Reads Ratio(%)
TRC11	Control	44.69	4.47	97.94	94.39	94.14
TRC12	Control	45.07	4.51	98.34	95.36	94.93
TRC21	control	44.94	4.49	97.86	94.14	94.67
TR322	Salt- treated	44.89	4.49	97.93	94.41	94.55
TR311	Salt-treated	45.1	4.51	97.97	94.41	94.99
TR323	Salt treated	45.01	4.5	97.84	94.13	94.8

TR311, TR322, TR323 (Salinity stressed Tana River samples), TRC11, TRC12, TRC21 Tana River Control samples

4.4.8 *de_novo* assembly

The *de_novo* assembly using clean reads resulted in 498, 082 transcripts with 90754, 75241 and 75354 transcripts for the control replicates 1, 2 and 3 respectively while the salt stressed resulted in 90265, 81733 and 84730 transcripts for replicate 1, 2 and 3 respectively. The lowest average transcript length and the lowest N50 value was 699bp and 1205bp (Table 4.22). The transcript length distribution per sample is shown in Appendix XIII

Table 4.22: Transcript metrics after RNA sequencing of *H. compressa* accessions

Sample	Total Number of Transcripts	Total Length of transcripts	Mean Length of transcripts (bp)	N50	N70	N90	GC(%)
TRC11	90759	64687216	712	1253	657	270	45.33
TRC12	75241	52644665	699	1205	639	269	46.16
TRC21	75354	54441653	722	1257	674	275	46.1
TR311	90265	65409614	724	1308	673	271	45.33
TR322	81733	57647306	705	1234	646	267	46.1
TR323	84730	60759144	717	1260	667	272	45.81
TOTAL	498,082						

TR311, TR322, TR323 (Salinity stressed Tana River samples), TRC11, TRC12, TRC21 Tana River Control samples

4.4.9 Assembly metrics using BUSCO

BUSCO analysis showed completeness score of 89.4% [Single:3.5%, Duplicated:85.9%], Fragmented:7.5%, Missing:3.1% for all transcripts. TR311 had the highest completeness percentage (85.1% ;228) while TRC12 had the least (75.7%; 197) as shown in Figure 4.30.

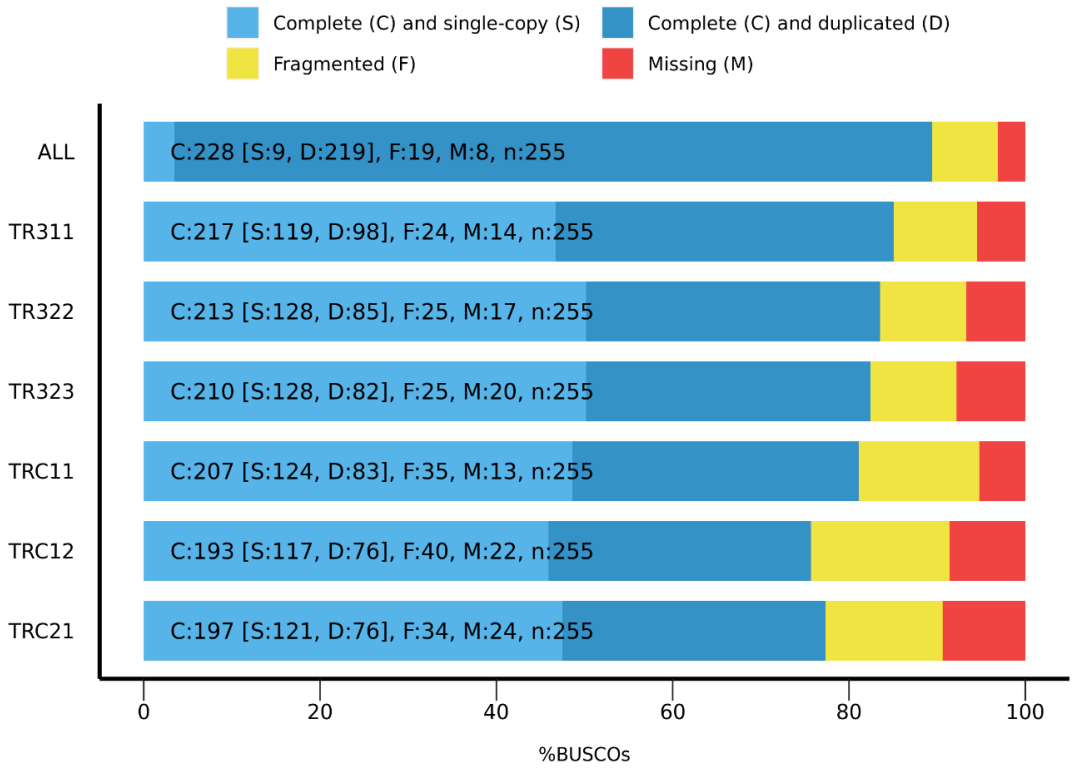


Figure 4.30: BUSCO assessment results for all transcripts of *H. compressa* accessions.

BUSCO results are shown for individual samples of *H. compressa* and the number of complete, Single, fragmented and missing BUSCOs.

4.4.10 Transcript clustering

After clustering of the assembled transcripts with Tgicl, a total of 92,135 unigenes were obtained with an N50 value of 1695bp, mean length of 988bp and GC content of 45 (Table 4.23). The highest number of unigenes (63,612) was obtained in the first replicate of the control treatment. The unigene length varied from 300 nucleotides (24,042 unigenes) to over 3,000 nucleotides long (4619 unigenes) as shown in Figure 4.31.

Table 4.23: Metrics of 92,135 unigenes obtained from *de_novo* assembly of RNA-Seq of *H. compressa* accessions

Sample	Total Number of unigenes	Total Length of unigenes	Mean Length of unigenes	N50	N70	N90	GC(%)
TR311	63,102	52,273,819	828	1397	790	320	45.33
TR322	56,614	46,056,691	813	1341	769	319	46.08
TR323	59,069	48,565,242	822	1357	778	322	45.81
TRC11	63,612	51,802,260	814	1348	767	320	45.34
TRC12	52,948	42,335,373	799	1295	748	317	46.12
TRC21	53,063	43,482,944	819	1334	778	325	46.06
All-Unigene	92,135	91,088,117	988	1695	1049	384	45

TR311, TR322, TR323 (Salinity stressed Tana River samples), TRC11, TRC12, TRC21 Tana River Control samples

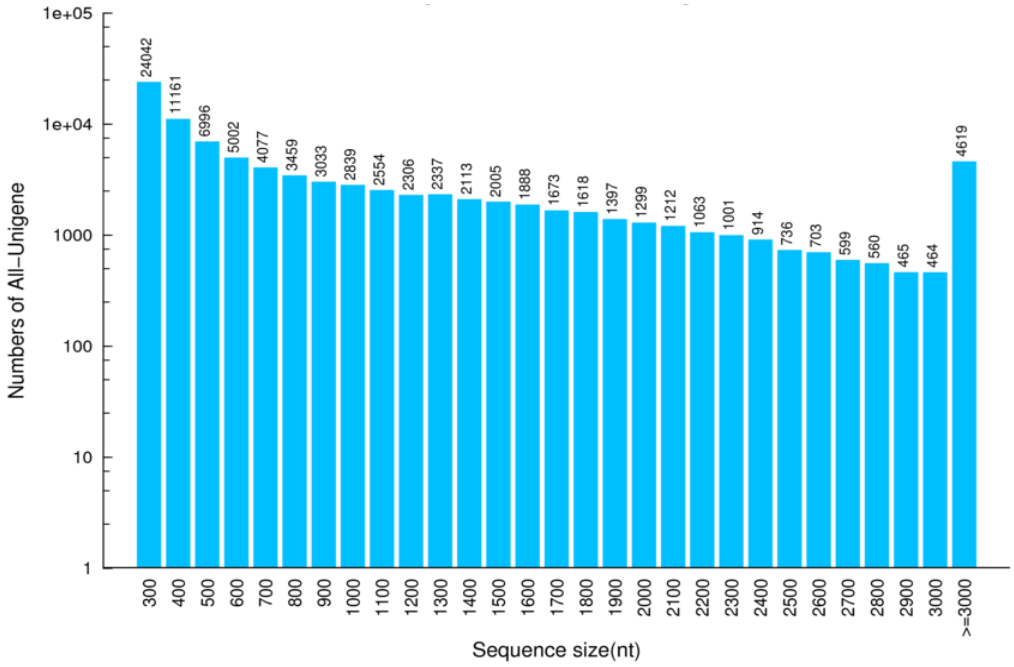


Figure 4.31: Length distribution of unigenes after clustering of *H. compressa* transcripts

4.4.11 Functional annotation of unigenes

The NT which is the official nucleic acid database of NCBI annotated the most unigenes (58,698, 63.71%) followed by NR (60.8%), Interpro (48.3%), KOG (46.5%), SwissProt (40.2%) and GO (12.5%). A total of 6434 were the only unigenes annotated by all seven databases (Figure 4.32).

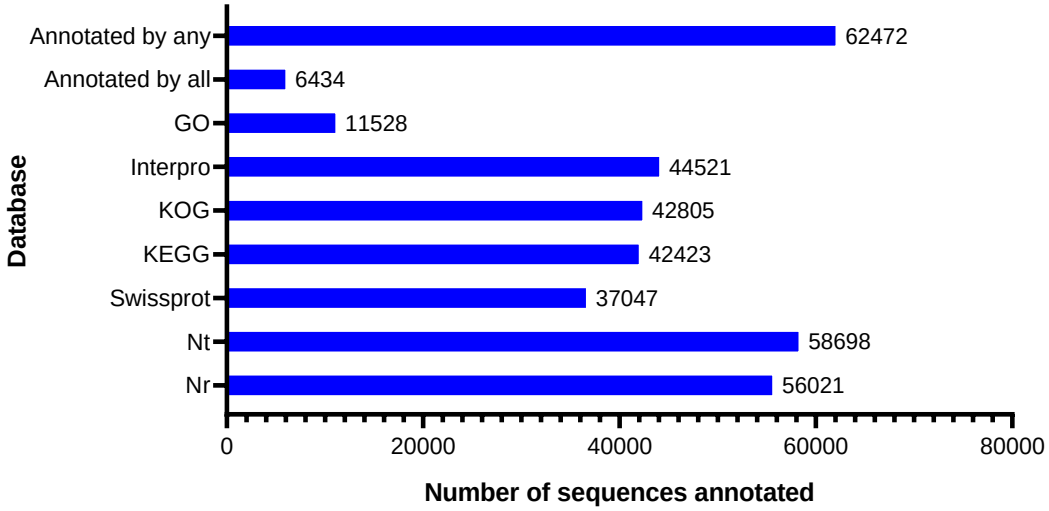


Figure 4.32: The number of *H. compressa* unigenes annotated by the seven databases

Phoenix dactylifera (date palm) sequences accounted for 50.5% of all the annotated unigenes obtained from *H. compressa* transcriptome based on the NR annotation (Figure 4.33). This was closely followed by *E. guineensis* (African oil palm) sequences with (39.6%), *Ananas comosus* (1.1%) while the sequences from the other organisms constituted 8.90% of the annotated *H. compressa* unigenes.

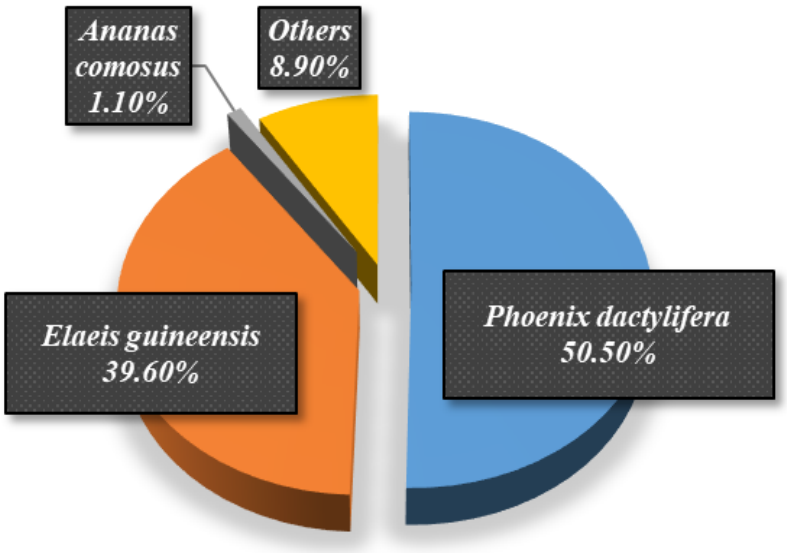


Figure 4.33: Species distribution of *H. compressa* transcripts based on Nr database.

Five databases including InterPro, SwissProt, KEGG, KOG and NR annotated a total of 27,920 genes (Figure 4.34).

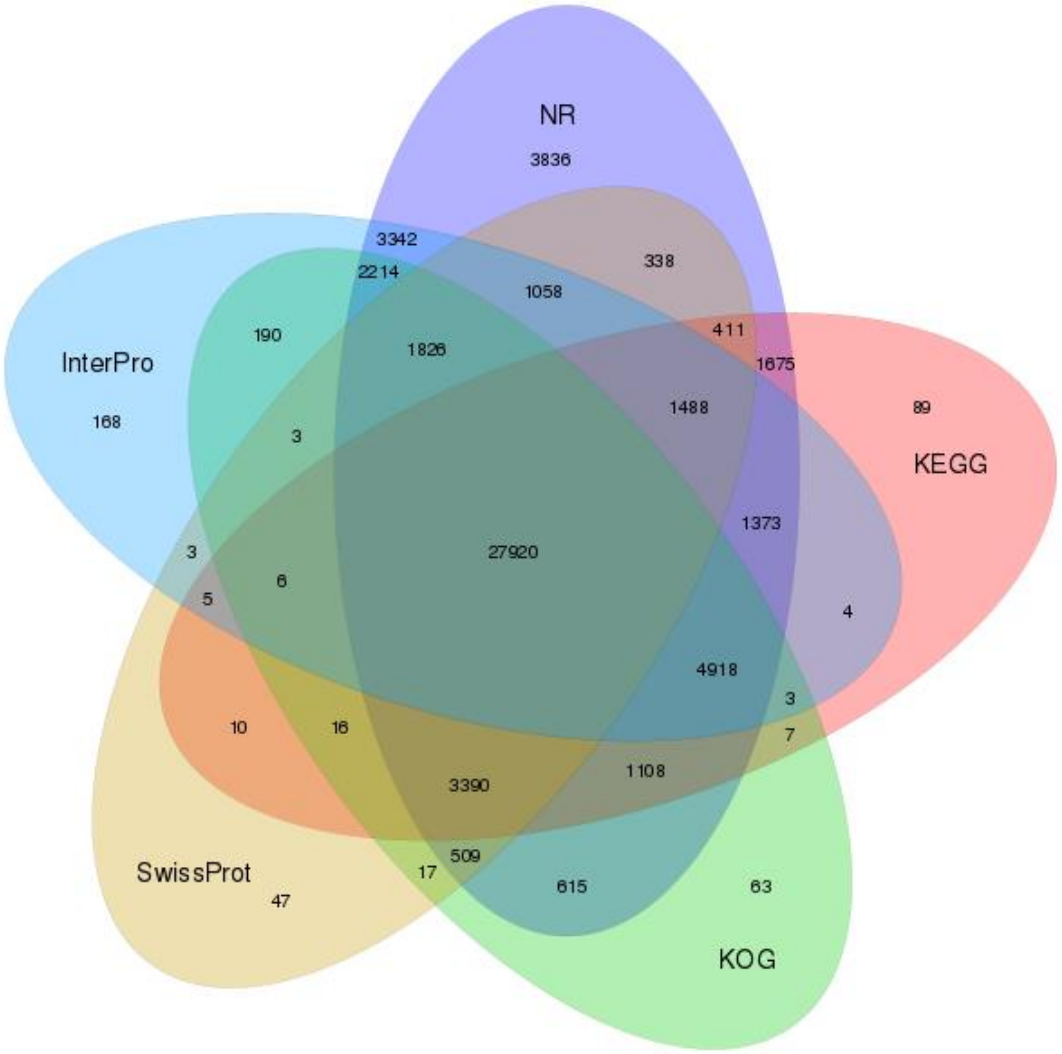


Figure 4.34: Venn diagram showing the annotated unigenes by five databases

The five databases include; **KOG** (Eukaryotic Clusters of orthologous Groups), **Swissprot database**, **KEGG** (Kyoto Encyclopedia of Genes and Genomes), **InterPro** database and **NR (**Non-redundant database of NCBI).

4.4.11.1 KOG Annotation of all transcripts

Based on KOG functional annotation, the General functional prediction category had the highest annotated genes (11,680) followed by signal transduction (8460) while cell motility had the lowest (92) annotated unigenes (Figure 4.35).

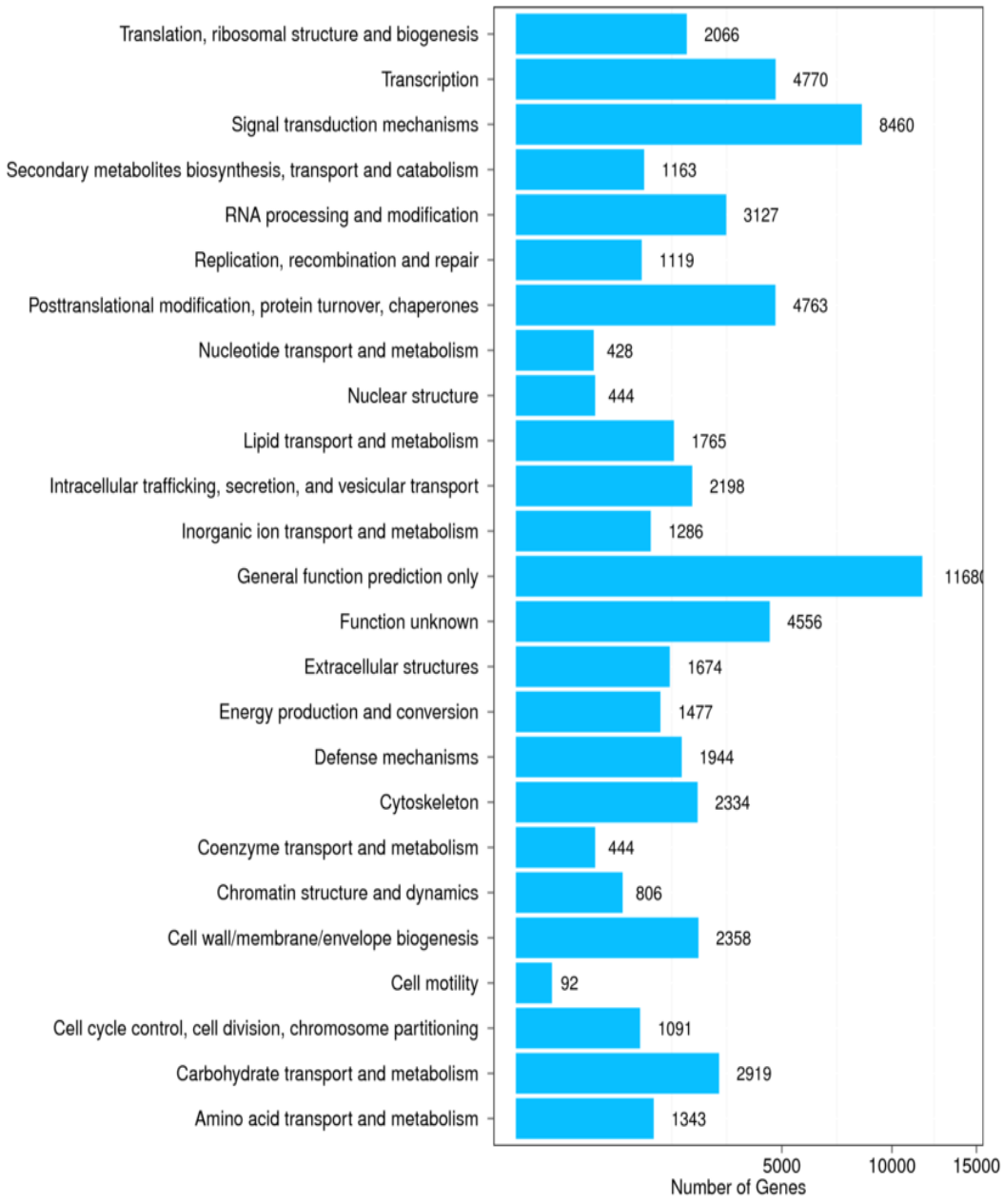


Figure 4.35: KOG annotation distribution of all *H. compressa* unigenes

4.4.11.2 KEGG Annotation of all transcripts

The KEGG annotation was done for both level 1 and level 2 categories. The annotated unigenes fell into 5 level 1 categories including cellular processes, environmental information processing, genetic information processing, metabolism and organismal systems. These categories were also divided into 20 level 2 categories (Figure 4.36). The largest level 1 category was metabolism with 11 level 2 categories. Some of the level 2 metabolism categories included; Global and overview map (8948), carbohydrate metabolism (3205), biosynthesis of secondary metabolites (1141), biosynthesis of cofactors and vitamins (1080), Glycan biosynthesis (787), metabolism of terpenoids and polyketides (625).

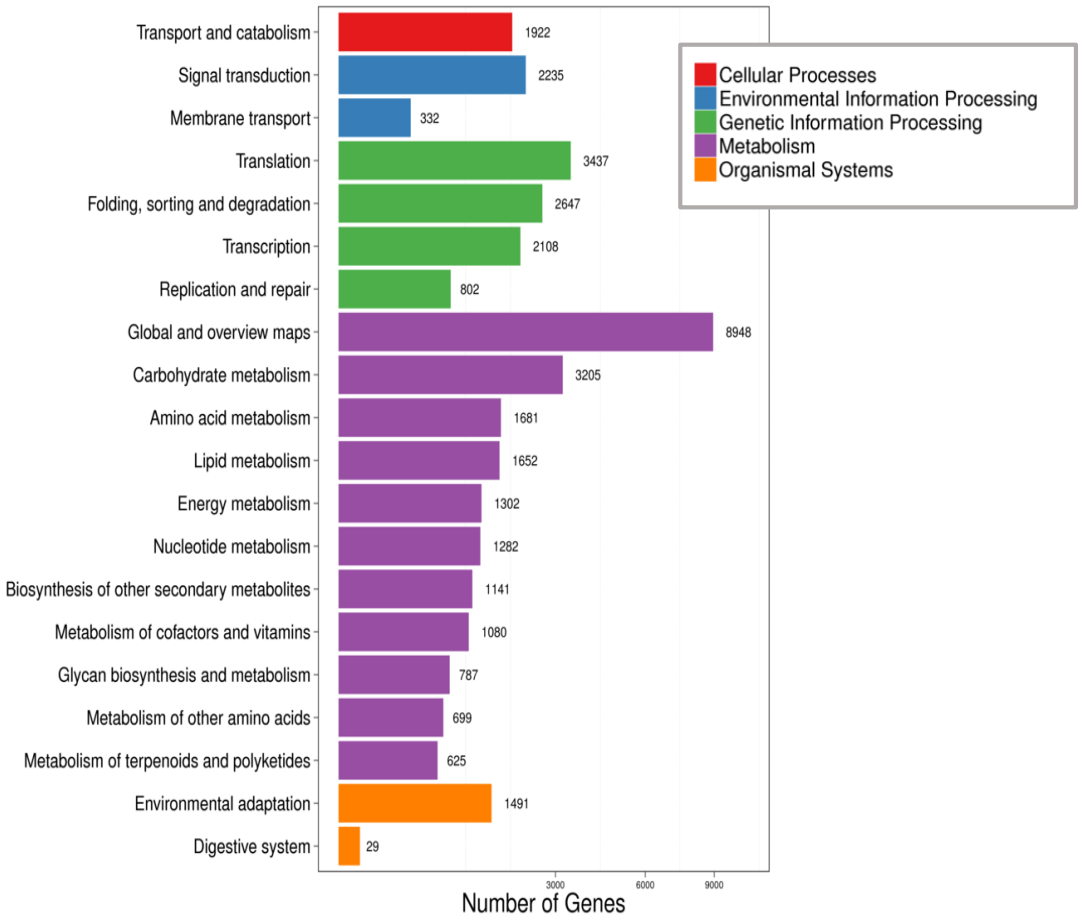


Figure 4.36: KEGG annotation distribution of all *H. compressa* unigenes

This annotation is based on level 1 and 2 categories

4.4.11.3 GO annotation of all transcripts

Hyphaene compressa transcripts were annotated using interPro. The GO terms were then assigned to interpro entries. *H. compressa* unigenes were classified into 53 GO terms. The GO terms were classified into cellular (24,615 genes), biological (15,244 genes) and molecular (11,917 genes) processes. Some unigenes were assigned to more than one GO term. The proportion of *H. compressa* unigenes assigned to each GO term is shown in Figure 4.37. Cellular and metabolic processes were the most highly represented in the biological processes group.

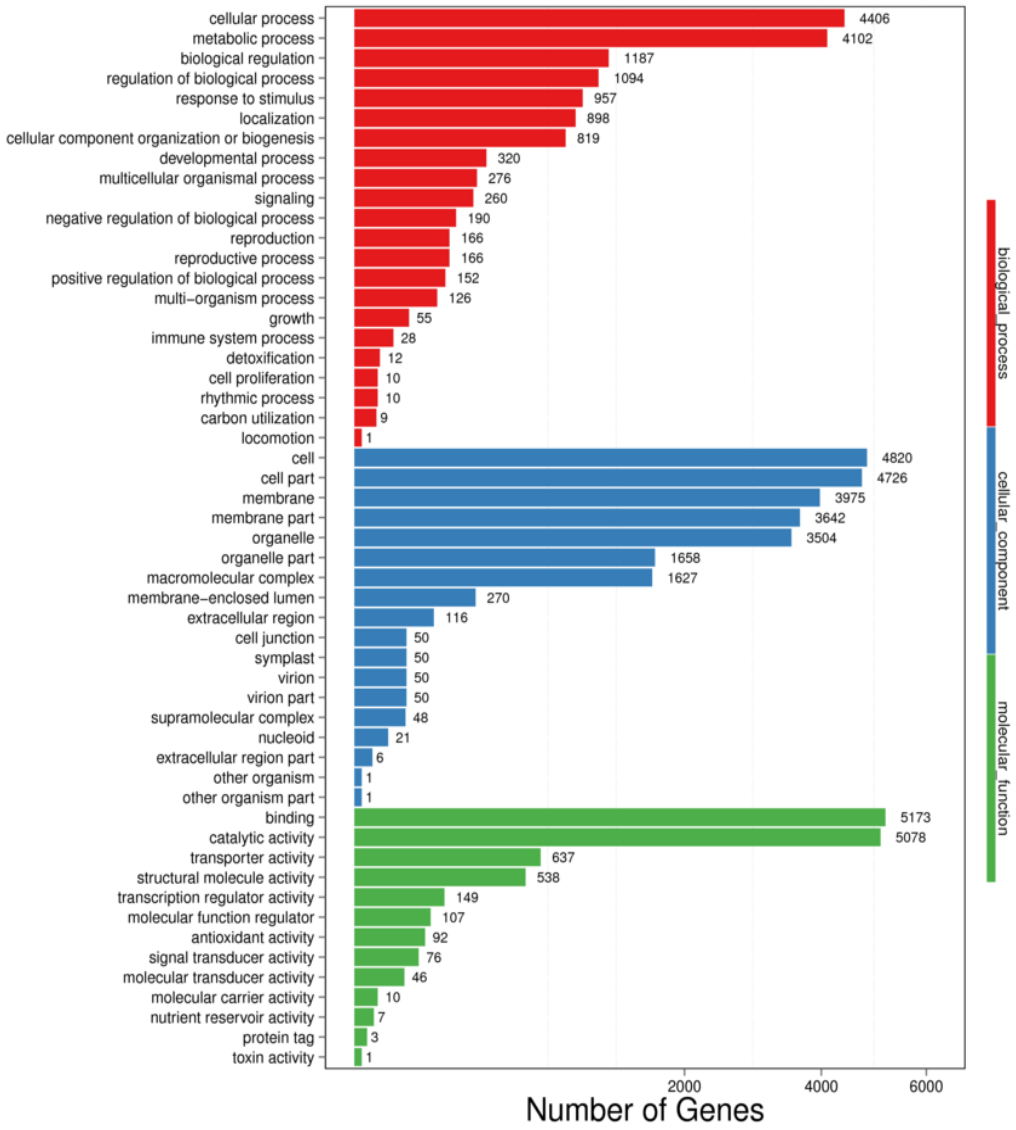


Figure 4.37: *H. compressa* unigene annotation based on GO terms.

4.4.12 Prediction of coding regions in doum palm transcriptome unigenes

Transdecoder software (Haas & Papanicolaou, 2019) was used to predict the coding regions using the omics box (OmicsBox, 2019). Out of the 92,135 unigene sequences, a total of 43,598 sequences had coding regions with the following types of ORF; compete (26,109, 60%), 5' partial (8859, 20%), 3' partial (3763, 9%) and internal (4767, 11%) as shown in Figure 4.38.

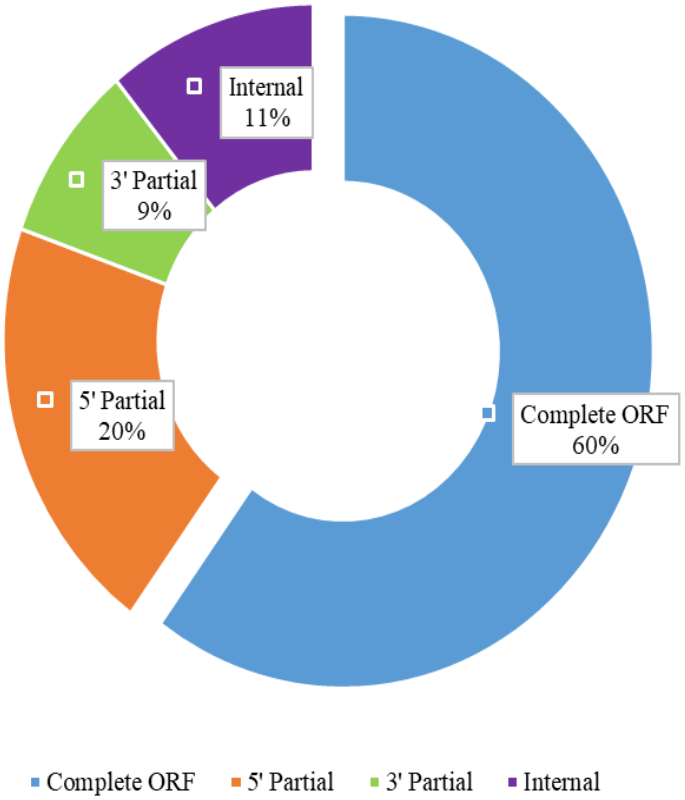


Figure 4.38: Coding regions obtained in 92,135 unigenes of *H. compressa*.

4.4.13 Unigene expression

After mapping clean reads to the unigenes, the lowest mapping percentage was 88.1% (TR311) and the highest was 90.5% (TRC12). The uniquely mapped reads were in the range of 18.5 to 19.3 million (Table 4.24). The gene expression levels per sample are shown in Figure 4.39. High transcript expression was found at expression values 1-10 FPKM (Figure 4.40).

Table 4.24: Summary of alignment results of clean data to unigenes

Sample	Total Bases	Total Reads	Total Mapped Reads	% mapping	Uniquely Mapped Reads
TR311	4509903000	45099030	39724562	88.1	18849308
TR322	4488966600	44889666	39848538	88.8	19193064
TR323	4500677200	45006772	40006674	88.9	19331068
TRC11	4469216200	44692162	39579186	88.6	18450572
TRC12	4507078600	45070786	40802920	90.5	19209842
TRC21	4494395000	44943950	40592056	90.3	19169556

TR311, TR322, TR323 (Salinity stressed Tana River samples), TRC11, TRC12, TRC21 Tana River Control samples

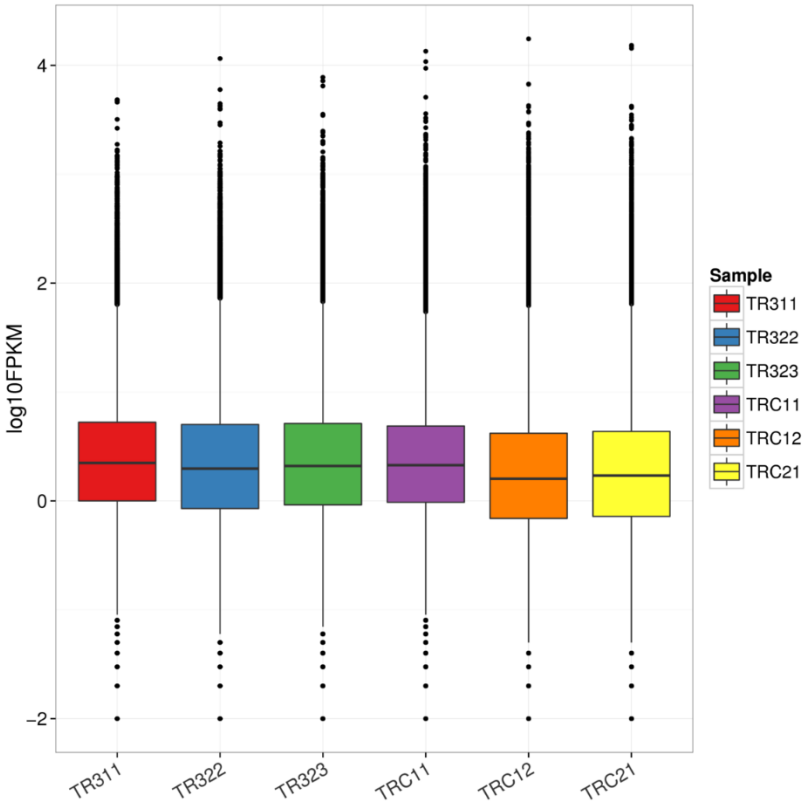


Figure 4.39: Gene expression levels of *H. compressa* leaf transcriptome samples

TR311, TR322, TR323 (Salinity stressed Tana River samples), **TRC11, TRC12, TRC21** Tana River Control samples

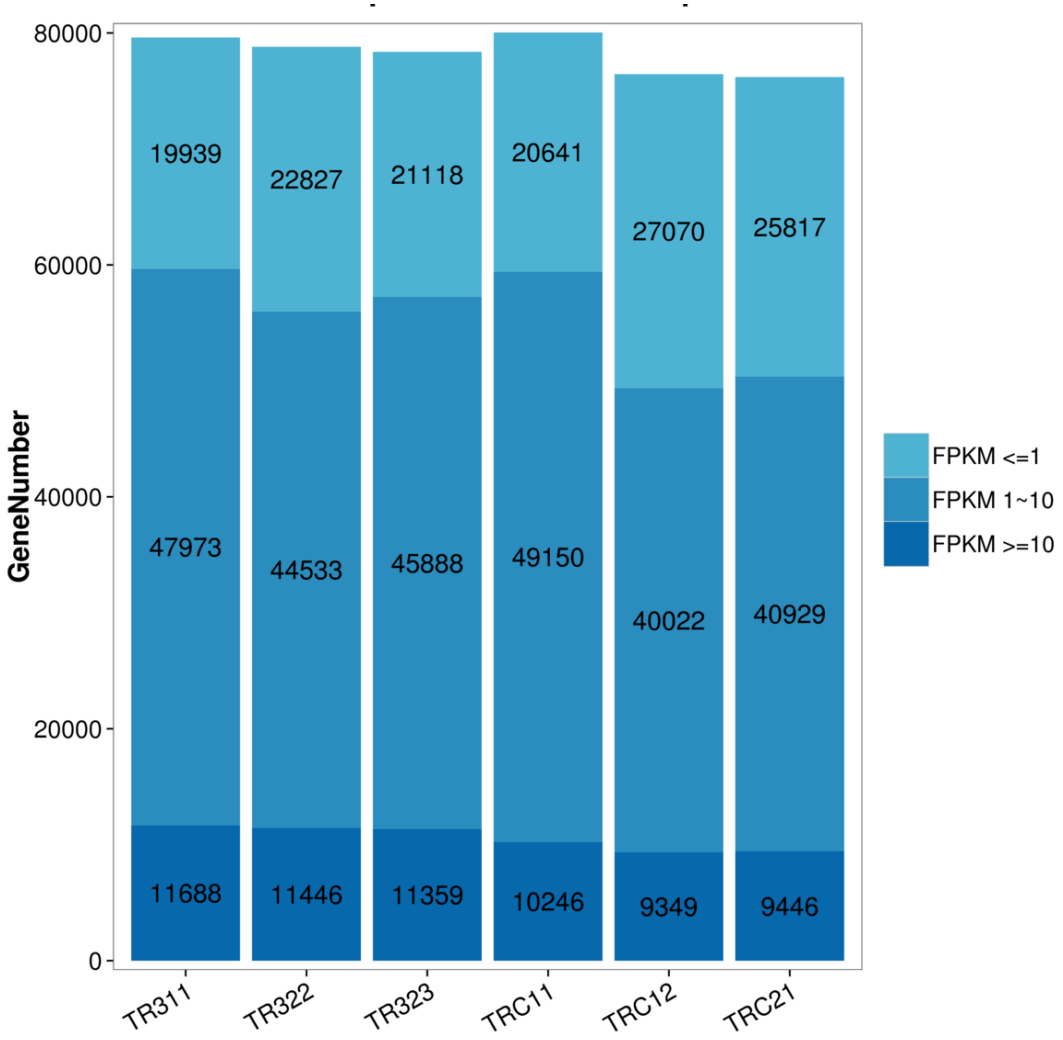


Figure 4.40: ***Hyphaene compressa*** **transcripts at three FPKM expression levels**

TR311, TR322, TR323 (Salinity stressed Tana River samples), **TRC11, TRC12, TRC21** Tana River Control samples

4.4.13.1 Gene expression levels between samples and groups

In the control group, TRC11 had the highest amount of specific genes (3732) while in the treatment group, TR322 had the highest (2755) as shown in Figure 4.41A and B respectively. A total of 66269 and 67428 genes were shared among the three control and treatment groups respectively.

4.4.14 Detection of differentially expressed genes (DEGs)

Using DESeq2, a total of 8611 DEGs were identified between salt stressed and control *H. compressa* palms. Out of the 8611, 3722 were upregulated while 4889 genes were

down regulated at the significance level *padj* < 0.05 through the Benjamini-Hochberg method (Figure 4.42).

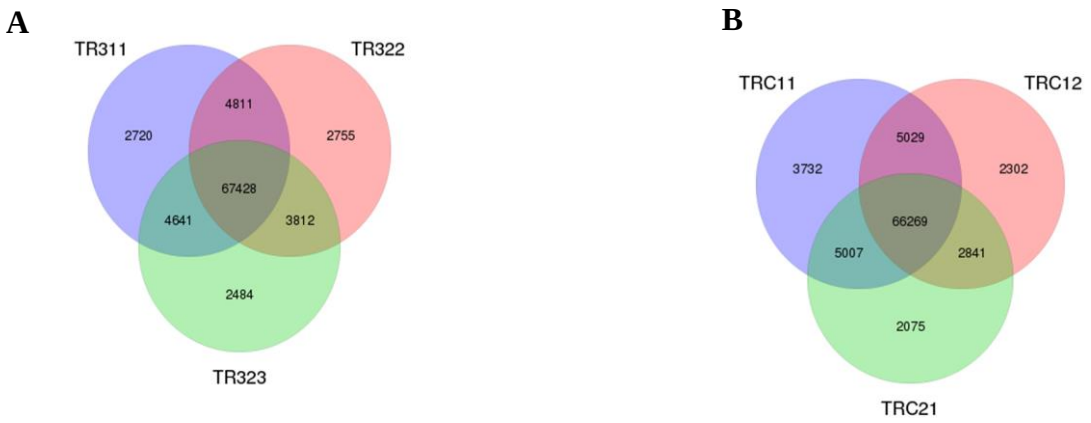


Figure 4.41: Gene expression levels among *H. compressa* samples

A. Tana River control samples (**TRC11, TRC12, TRC21**). **B.** Salinity stressed Tana River samples **(TR311, TR322, TR323**).

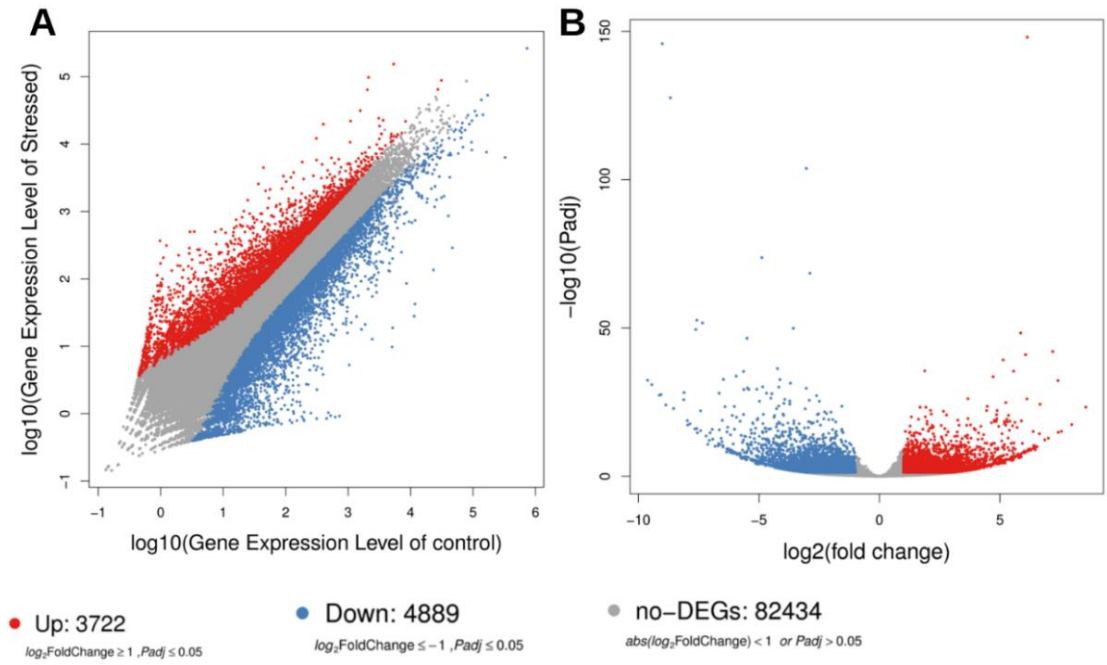


Figure 4.42: Salinity induced DEGs in *H. compressa* accessions
A. Scatter plot, **B.** volcano plot

The highest upregulation of genes (14986 genes) was observed between sample TRC12 (control) and TR311 (Salt treated). On the other hand, sample TRC21

(control) and TR322 (salt treated) had the least down regulation of genes (6593 genes) as shown in Figure 4.43.

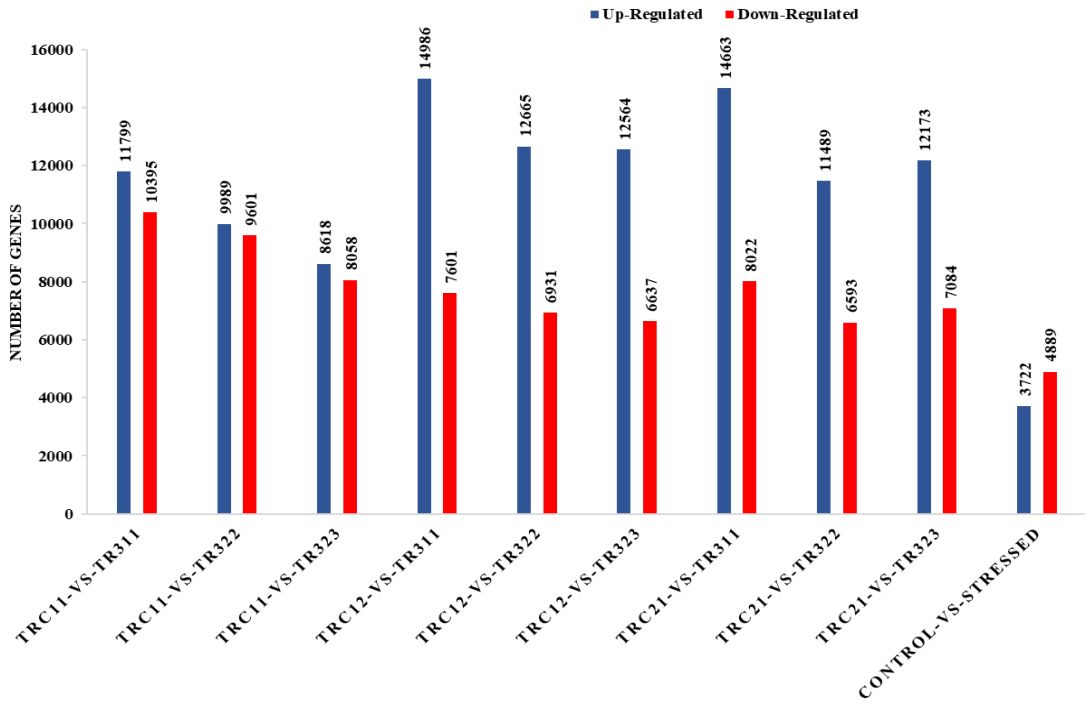


Figure 4.43: DEGs in *H. compressa* control and salinity stressed plants

TR311, TR322, TR323 (Salinity stressed Tana River samples), **TRC11, TRC12, TRC21** Tana River Control samples

The most significantly up regulated genes were NPL4-Like protein (L_2fc 8.56), Caotamer sub unit (L_2fc 7.98) and Phototropin (L_2fc 7.98). While the most down regulated genes were Cyclin H (L_2fc -9.62), Clathrin assembly protein (L_2fc -9.4) and Arabinose-5-phosphate Isomerase (L_2fc -8.95) as shown in Table 4.25.

Table 4. 25: The most significant up and down regulated genes in *H. compressa* exposed to salinity stress

Gene	Regulation	Log2foldchange	p value
NPL4-Like protein	UP	8.56	4.74E-24
Coatomer subunit beta	UP	7.98	3.90E-18
Phototropin	UP	7.58	7.06E-16
beta glucosidase	UP	7.42	1.56E-15
BOI Related E3 ubiquitin protein ligase	UP	7.4	5.59E-33
Poly (rC) binding protein	UP	7.19	8.84E-43
MYB Transcription factor	UP	6.85	5.21E-13
Serine/threonine protein Kinase	UP	6.66	5.28E-25
Polyamine oxidase	UP	6.44	2.39E-10
Expansin	UP	6.38	8.04E-16
Phosphate transporter	UP	6.29	1.24E-09
Splicing factor	UP	6.16	9.55E-10
Cyclin H	down	-9.62	4.26E-33
Clathrin assembly protein	down	-9.4	1.28E-31
Arabinose-5-phospate isomerase	down	-8.95	1.54E-26
Glucomannan 4 -beta-mannosyltransferase	down	-7.95	4.97E-18
Protein RIK	down	-7.58	8.29E-17

4.4.15 Functional annotation of DEGs

4.4.15.1 GO Functional annotation and enrichment of DEGs

Of the 8611 DEGs, only 1518 were annotated to GO terms. Out of the 4889 down regulated DEGs, 859 were annotated to GO terms whereas 659 out of the 3722 up regulated genes were annotated to GO terms (Figure 4.44). They were classified into three main categories; biological, cellular and molecular functions (Figure 4.45). In the biological processes category, Cellular (561) and metabolic (542) processes were predominant. Membrane (567), cell (564), cell part (556) and membrane part (523) were the most common in the cellular category whereas catalytic (718) and binding (709) were the most common molecular processes.

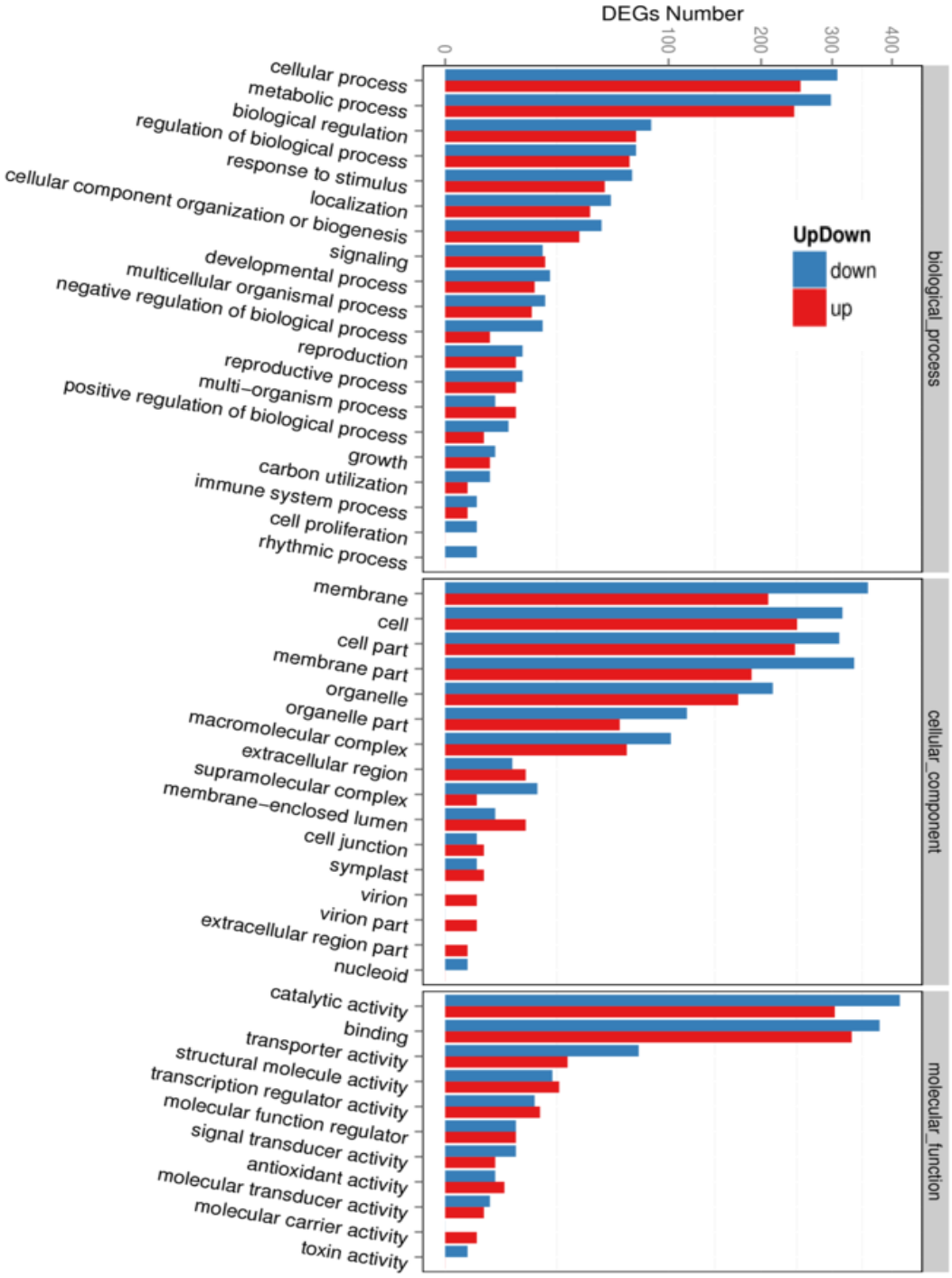


Figure 4.44: GO annotation of Up and down regulated *H. compressa* accessions exposed to salinity stress.

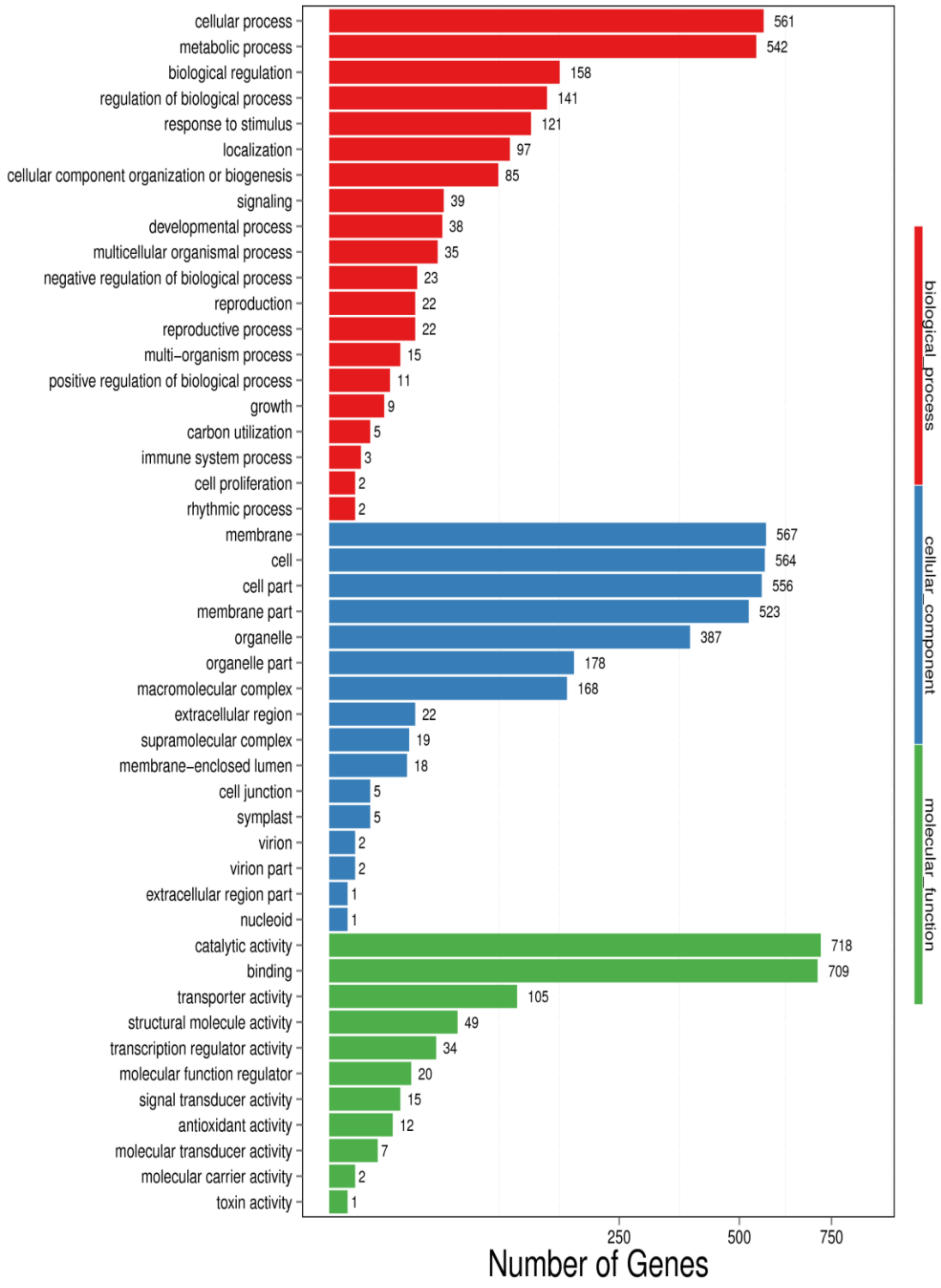


Figure 4.45: GO annotation of salinity induced DEGs obtained from *H. compressa*

A total of 25 DEGs were significantly enriched (FDR< 0.05) and were classified into three main categories; biological, cellular and molecular (Table 4.26). Three biological processes that were enriched were; photosynthesis, DNA replication and initiation and DNA dependent DNA replication. A total of 19 cellular processes were significantly enriched with photosystem (FDR 2.56e-07) being the top most enriched. In the molecular category, three processes were enriched; tetrapyrrole binding (FDR 0.00194), structural constituent of the cytoskeleton (FDR 0.00656) and DNA binding (FDR 0.00912).

Table 4.26: Gene ontology enrichment of salinity induced DEGs in *H. compressa*.

GO ID	Term	hits	FDR <0.05)
Biological Processes			
1.	Photosynthesis	63	8.06e-07
2.	DNA replication initiation	6	0.00470
3.	DNA dependent DNA replication	7	0.05129
Cellular Processes			
1.	Photosystem	46	2.56e-07
2.	Thylakoid part	69	4.29e-06
3.	Photosynthetic membrane	66	5.64e-06
4.	Photosystem II	38	7.48e-06
5.	Thylakoid	76	5.10e-05
6.	microtubule	19	0.00022
7.	microtubule cytoskeleton	23	0.00043
8.	Supramolecular complex	19	0.00048
9.	Supramolecular polymer	19	0.00048
10.	Supramolecular fiber	19	0.00048
11.	Polymeric cytoskeletal fiber	19	0.00048
12.	MCM complex	6	0.00080
13.	Photosystem I	24	0.00106
14.	Thylakoid membrane	57	0.00106
15.	Membrane	567	0.00243
16.	Membrane Part	523	0.00434
17.	Intrinsic component of membrane	489	0.00528
18.	Integral component of membrane	487	0.00787
19.	THO complex	6	0.04814
Molecular Processes			
1.	Tetrapyrrole binding	51	0.00194
2.	Structural constituent of cytoskeleton	14	0.00656
3.	DNA Binding	118	0.00912

This table only shows the genes that are significantly enriched (with FDR<0.05)

4.4.15.2 KEGG Functional annotation and enrichment of salinity induced DEGs

A total of 5411 DEGs were assigned to KEGG pathways. Pathways related to metabolic and synthesis of secondary metabolites were the most predominant (Figure 4.46).

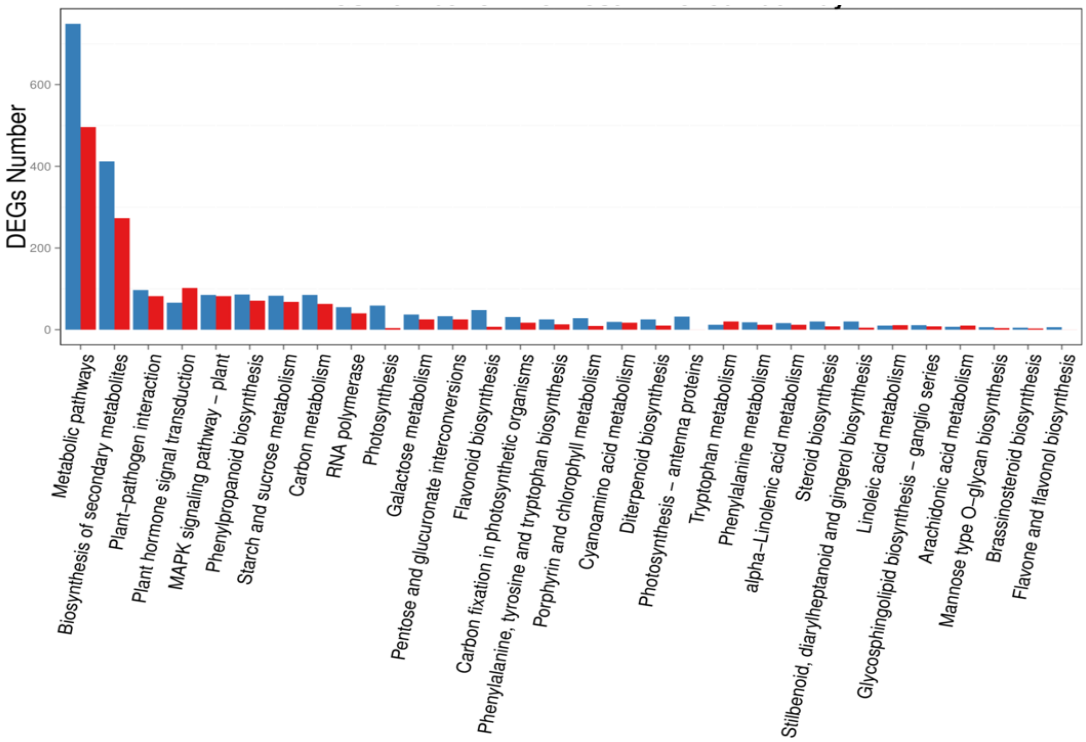


Figure 4.46: KEGG annotation of *H. compressa* salinity induced DEGs

Transcripts in red are upregulated while those in blue are down regulated

Carbohydrate metabolism and Global and overview maps (512) were the most predominant metabolism processes (Figure 4.47). Signal transduction pathway was the most predominant in the environmental information processing category. The 5411 unigenes assigned, were divided into 133 KEGG pathways. KEGG pathways with p value < 0.05) were considered significantly enriched in *H. compressa* accessions under salinity stress. Of the 133 pathways, 36 were significantly enriched (Figure 4.48). Flavonoid biosynthesis was the most enriched (P=9.04e-13) followed

by photosynthesis (p= 1.06e-10), biosynthesis of secondary metabolites (P=1.3e-10) and photosynthesis antennae proteins (P=1.9e-10).

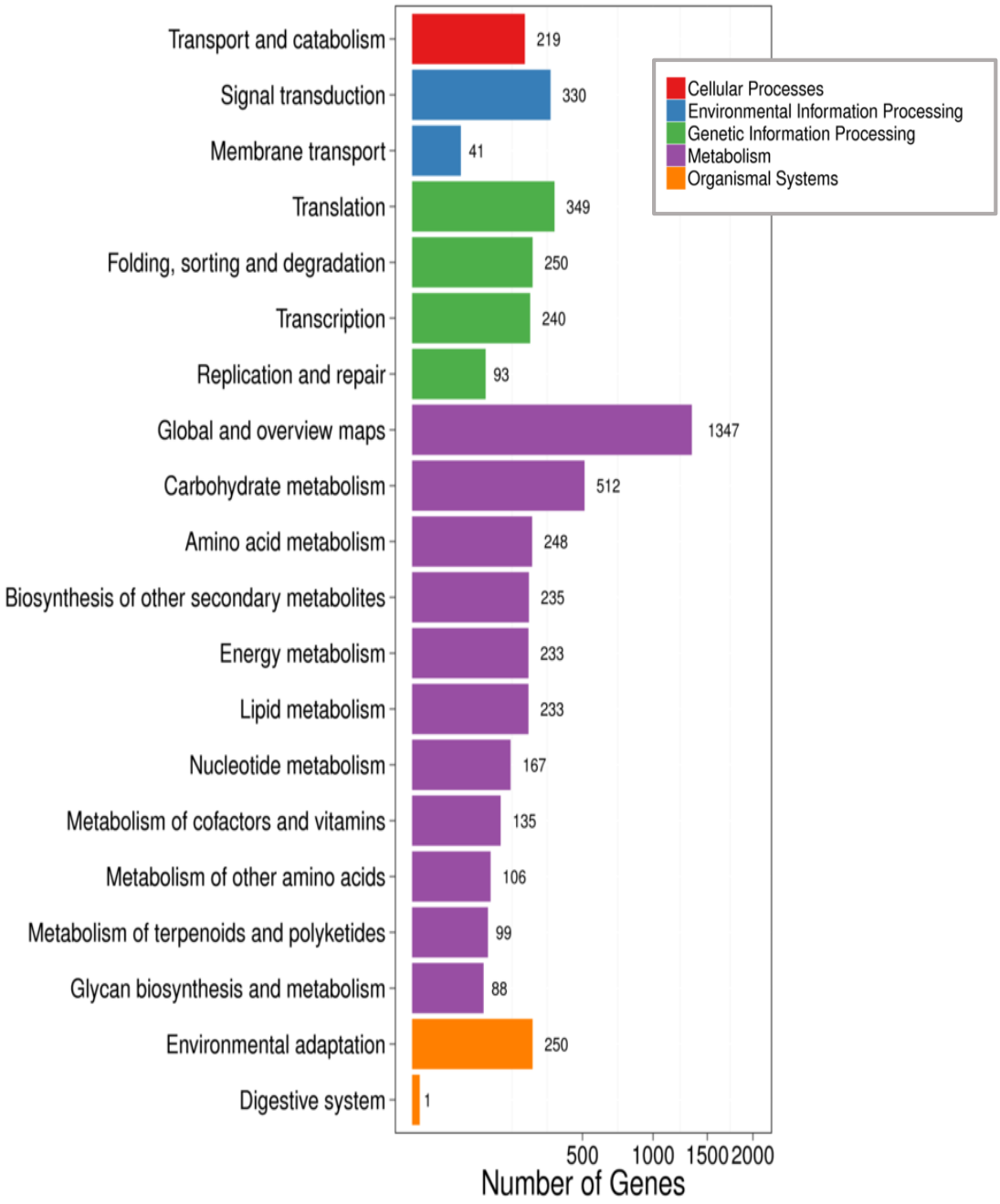


Figure 4.47: KEGG Annotation of salinity induced DEGs in *H. compressa*

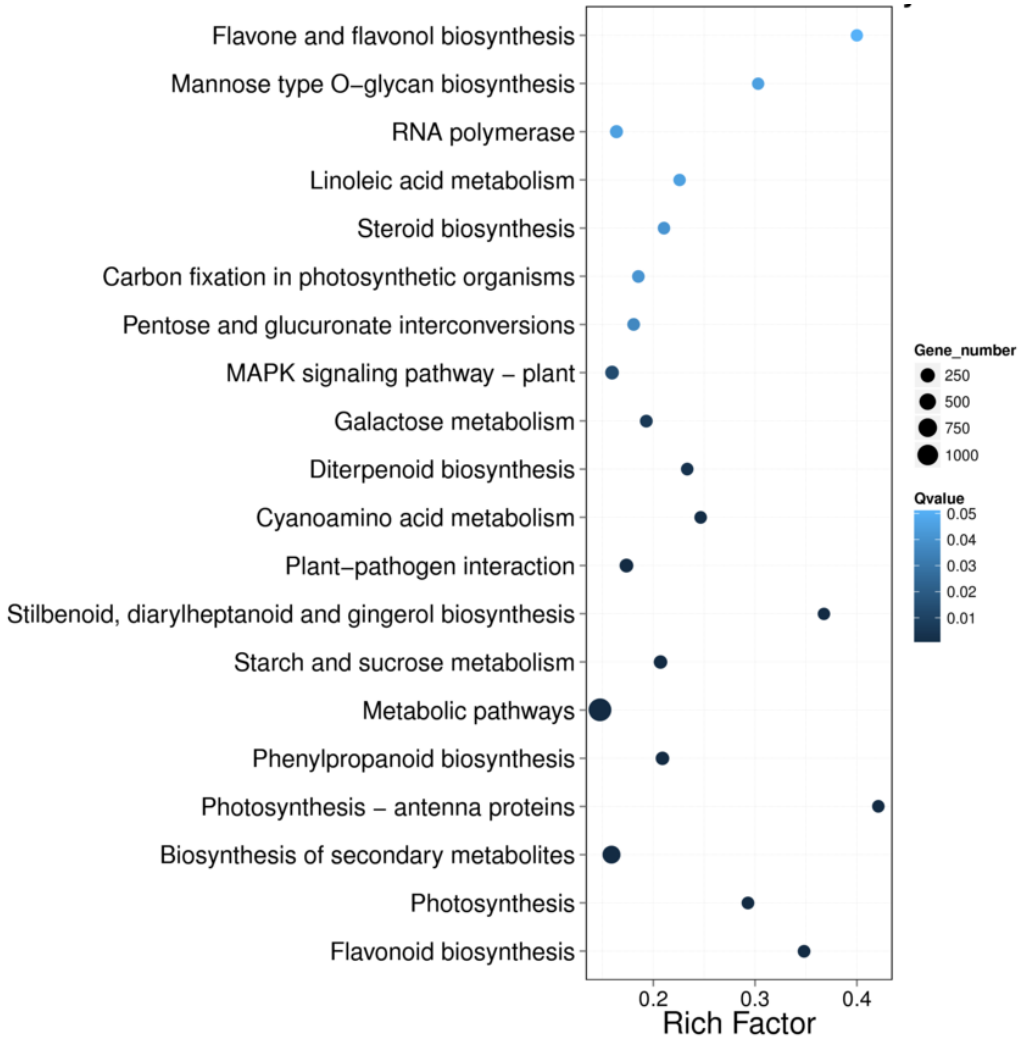


Figure 4.48: The 20 most enriched salinity induced KEGG pathways in *H. compressa*

Q value is the corrected P value. The darker the blue color, the more the significance value.

Photosynthetic pathway contained a total of 33 DEGs. In Photosystem I (PSI), there was a total of nine down regulated genes and one upregulated gene (PsaA). The cytochrome b6/f complex had two down regulated genes (PetA and PetN). The photosynthetic electron transport on the other hand had 3 down regulated genes. These differentially expressed genes are shown in Figure 4.49. The photosynthesis antenna proteins pathway contained a total of 10 DEGs all of which were down regulated. These DEGs were LHCa1, LHCa2, LHCa3, LHCa4, LHCb1, LHCb2, LHCb3,

LHCb4, LHCb5 and LHCb6 which are involved in light harvesting chlorophyll protein complex (LHC).

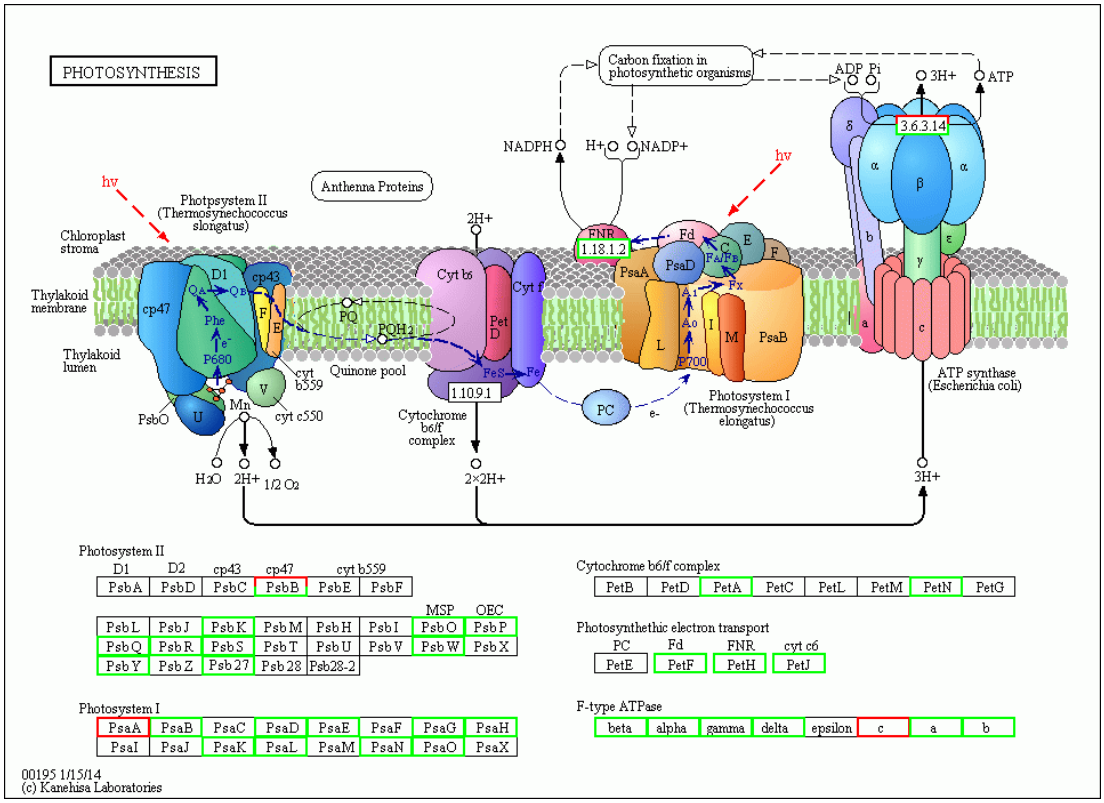


Figure 4.49: Photosynthesis KEGG map showing salinity induced DEGs in *H. compressa*

Genes within green boxes are down regulated while the genes within red boxes are up upregulated

4.4.16 Differentially expressed transcription factors

A total of 755 transcription factors were identified in differentially expressed unigenes after comparison to the TF domains of *P. dactylifera*. These transcription factors were divided into 46 TF families. MYB and MYB related TF were the most common (131) followed by NAC (106), bHLB (59) and ERF (54) as shown in Figure 4.50.

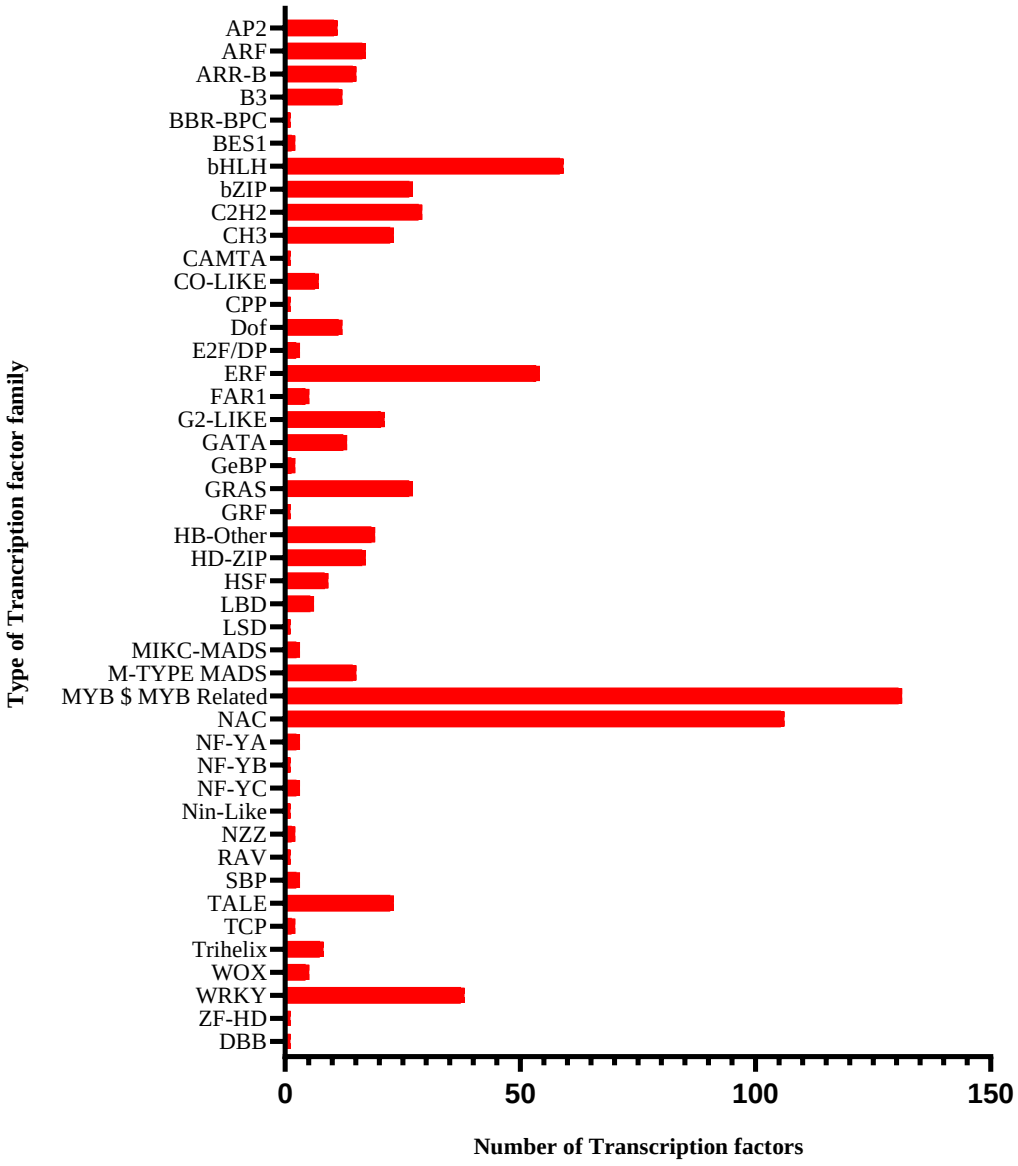


Figure 4.50: Salinity induced transcription factors in *H. compressa*

4.5 Development and validation of markers from RNA-seq data for diversity studies

4.5.1 Development of SSR markers using unigenes obtained from RNA Seq data

The number of unigenes examined for SSR detection was 92,135 sequences. A total of 16, 632 perfect SSR's were identified using MISA with a frequency of 0.183SSR per kb. A total of 2625 sequences contained more than 1 SSR while 1488 SSR's were present in compound formation (Table 4.27). The most frequent SSR's were

dinucleotide (7428, 44.7%) while the least common SSR type was pentanucleotides accounting for only 0.9% (146) of total SSRs (Appendix XIV). Class II SSR type (10,749) was the most predominant type. The CT/AG dinucleotide SSRs were the most frequent (5928), followed by AT/AT (727), AC/GT (721) and CG/CG (52). The most abundant trinucleotide repeats were AGG/CCT (1278) as shown in Figure 4.51.

Table 4.27: SSR detection using *de_novo* assembled *H. compressa* transcripts

Characteristic	Frequency (%)
Number of unigene sequences examined	92,135
Total size of examined sequences	91088117bp
Total number of identified SSRs	16,632
Frequency of SSR per kb	0.183 SSR per kb
Number of SSRs containing sequences	13,114
Number of sequences containing more than 1 SSR	2625
Number of SSRs present in compound formation	1488
Number of SSR Class I type	3086
Number of SSR class II type	10,749
Mononucleotides	3902 (23.5)
Dinucleotides	7428 (44.7)
Trinucleotides	4588 (27.6)
Tetranucleotides	306 (1.8)
Pentanucleotides	146 (0.9)
Hexanucleotides	262 (1.6)

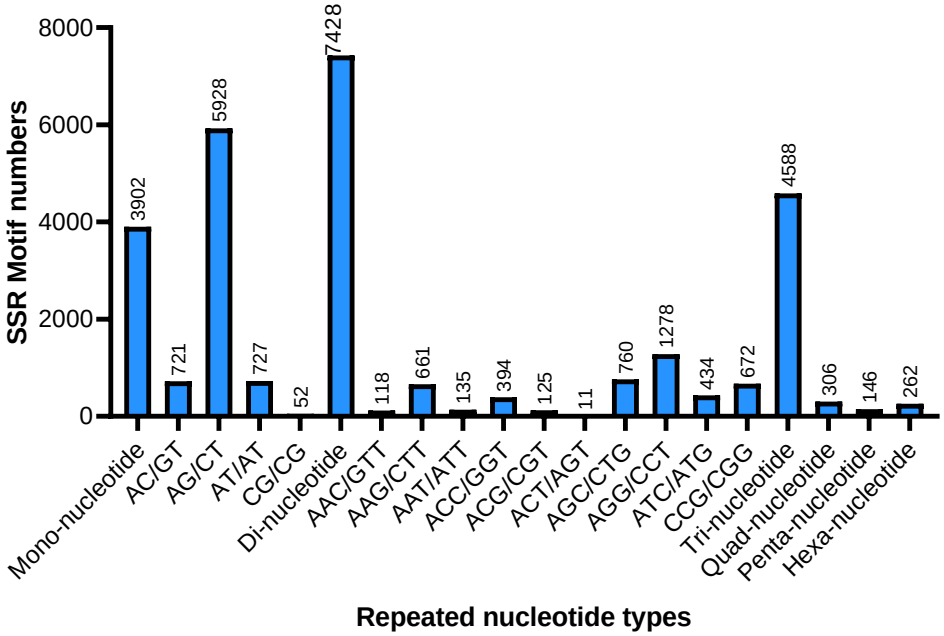


Figure 4.51: SSR repeat motifs of *de_novo* assembled *H. compressa* transcripts

4.5.2 SSR Primer design

Primer modelling using primer3 software (Untergasser et al., 2012) was successful for 11,437 microsatellite loci. However, some of these primers might be from the same locus because they may have been designed from inferred isoforms. Filtering was therefore done to remove primers with same priming sites. After filtering, 7318 SSR loci were retained. The script *get_orfs_or cdss.py* was used to get the location of the SSRs relative to ORFs. A total of 3078 SSR loci were found to be in non-coding regions with no overlap with an ORF while 4240 were found in coding regions (Table 4.28). Some of the SSR markers in non-coding and coding regions have been presented in Appendix XV and XVI.

Table 4.28: The number of ORFs and location of SSR markers from *de_novo* assembled *H. compressa* transcripts

Item	Type of SSR						Total
	p1	p2	p3	p4	p5	p6	
Number of ORFs							10,140**
Number of loci of each type	1836	2576	2119	91	58	111	7318
Number of loci of each type that do not overlap the ORF	1163	1251	332	54	32	26	3078
Number of loci of each type that have any overlap with an ORF	673	1325	1787	37	26	85	4240
Number of loci of each type in an ORF as percentage of total loci of each type (%)	36	51.4	84.3	40	44.8	76.6	
Number of loci of each type that have more than the designated (15bp) overlap with the ORF	95	470	741	30	15	79	1706

p (Perfect SSRs), 1-6 (mono, di, tri, tetra, penta and hexa nucleotide SSRs), ** some sequences have more than one ORF hence the high number of ORFs compared to the total SSRs.

A total of 20 pairs of primers (Table 4.29) were randomly selected from both coding and non- coding regions. These primers were then synthesized by Macrogen BV. Europe. Five primers were designed to amplify regions flanking the SSR markers on the noncoding sequences while majority of the primers amplified regions flanking SSR markers on coding sequences (15 primers) as shown in Table 4.29.

Table 4.29: List of primers, SSR markers and their locations relative to the ORFs from *de_novo* assembled *H. compressa* transcripts.

Primer	Unigene_ID	SSR Type	Bp start of SSR	Bp end of SSR	Bp start of ORF	Bp end of ORF	*overlap of SSR & ORF
DPSSR001	Unigene29361_All_7154_2	p2	192	203	124	549	11
DPSSR002	Unigene10862_All_5994_2	p2	184	203	3	431	19
DPSSR003	CL8783.Contig5_All_4544_1	p2	223	242	25	1920	19
DPSSR004	CL5948.Contig2_All_3615_3	p2	178	195	2	529	17
DPSSR005	Unigene22190_All_6708_1	p2	399	422	22	390	0
DPSSR006	CL5367.Contig1_All_3407_4	p3	691	708	107	727	17
DPSSR007	CL3171.Contig2_All_2263_4	p3	1243	1257	110	658	0
DPSSR008	Unigene26191_All_6969_2	p3	728	748	3	668	0
DPSSR009	CL5416.Contig3_All_3419_3	p3	1722	1736	2	1990	14
DPSSR010	Unigene23699_All_6781_4	p3	209	223	1	1026	14
DPSSR011	Unigene3909_All_5541_3	p4	402	421	238	480	19
DPSSR012	Unigene16704_All_6379_1	p4	623	642	2	661	19
DPSSR013	CL1661.Contig18_All_1220_4	p4	172	191	3	317	19
DPSSR014	Unigene28081_All_7072_2	p4	886	905	2	346	0
DPSSR015	CL5793.Contig2_All_3574_3	p5	1650	1669	303	1523	0
DPSSR016	CL3520.Contig2_All_2515_3	p5	2207	2226	501	2213	6
DPSSR017	CL3263.Contig1_All_2342_1	p5	355	374	1	429	19
DPSSR018	CL6531.Contig2_All_3859_1	p5	163	182	95	335	19
DPSSR019	Unigene26861_All_6985_5	p6	233	256	2	415	23
DPSSR020	CL1023.Contig5_All_812_4	p6	1133	1168	329	1615	35

*overlap-The number of bp of SSR markers that overlap the ORF.

The 5 primers amplifying non-coding regions are bolded

Primer information including the forward, reverse and melting temperatures of the designed primers are described in Table 4.30. The expected PCR product sizes were between 105bp (DPSSR016, AGCTT) and 160bp (DPSSR018, CGAGG). The product sizes, SSR sequences and the number of SSR repeats for each primer are listed in Appendix XVII.

Table 4.30: SSR primers designed from *de_novo* assembled *H. compressa* transcripts

Primer Name	Forward primer(5'-3')	Tm	Reverse primer(5'-3')	Tm
DPSSR001	TTCAAGAATATGCATGTCAGCAC	60.2	ACAAGAATTTCTGCTTCCCAAAT	60.3
DPSSR002	GTTTTCCACGTTCGTTCTTGTAG	60.1	AATTCAAACCGACAATCTCAAAA	59.9
DPSSR003	CCATCTCCCACCTCTTTTCTC	60.1	CTGCTTTTCCTGTTATAAGGGCT	60.2
DPSSR004	GGGAGTAGAGACGGTAAAGCAAG	60.6	CCCAACGTAATAAAAAGACGGAT	60.4
DPSSR005	ATCAAAAGCAAGTTCACCAACAT	59.9	ATAAACCCCTTCGATCAAGAGAC	59.9
DPSSR006	CGATCCAGCACCAGCTCTA	60.1	TTCTCGACGGCTACTGGTAGTT	60.3
DPSSR007	AGGTCTCACTGAGCTGGACATT	60.3	ATGCATCTTGTAATCTTGTTGCC	60.4
DPSSR008	TTCAATTTGGGAGGAGCTTCTAT	60.4	GTCGGAAGATGAAATCGAGAAG	60.2
DPSSR009	CTGTTCTGGTGCACCTTTGTAAT	60.5	CCTTCCTTTTACCCCTGTCTGT	60.7
DPSSR010	GAAGCTCCTCCAAAGGGTTATC	60.4	TCATAAGCTGGAGCTTGGTATGT	60.2
DPSSR011	CCTCCCCACTCTTCAAATAAAAA	60.7	TAGATCAAATCTGTCATCGAGCC	60.6
DPSSR012	TGGTGAGATAATTGACAGTGCAG	60.2	GCAATGGAATCAACAAGAACATT	60.2
DPSSR013	AGGTAAAATACCACCGAGCTTGT	60.3	CAGAAGAACGGATTAGAGAACGA	59.9
DPSSR014	AAGTAGAAGGATTGCTGGGAAAT	59.5	AGCATATTCACTCAGACAGGAGG	59.8
DPSSR015	TTATTCTGGTTTGGTTAGTGGGA	59.8	TAACTGATCACCCACAACATCTG	59.9
DPSSR016	ATGTGCTCCTCTCCTTTACCAA	60.1	GTGATGGGAGAGCACAGTAATTT	59.5
DPSSR017	TTCTTACTTTCCTCCGATTCCTC	60.1	CTCTGAATGATGCTACTTCGTCC	60.3
DPSSR018	CTACCTGCGCTATAGGTCTTCC	59.4	CGAGAACTCTTAAGTGATCAGACG	59.6
DPSSR019	GAAAAGCCCAGGTTCATCAGT	60.5	AGAGAGTGGAGAGAGGAGAAACG	60.5
DPSSR020	CGAGATCATTCAGACCGTGAT	60.1	AGAGTCCCTTTCTCTACCGTCAC	60.2

Tm is the meting temperature of the primer

4.5.3 Genetic diversity analysis using SSR markers

The 20 designed PCR primers were used for PCR screening with five accessions collected from each of the following; Tharaka (TG1, TG5, TG9, TG11 and T19), Tana River (R2, R4, R6, R16 and TR22), Turkana (KG2, K6, KG15, K17 and K21) and Kwale (MB4, MK2, MK6, MM3 and MM6) giving a total of 20 accessions (Table 4.31). Two primer pairs (DPSSR008 and DPSSR010) did not amplify any product. The 18 doum palm SSR markers amplified a total of 55 alleles at an average of 2.75 alleles per locus (Table 4.31).

Table 4.31: Primer information for 20 SSR markers used for genotyping *H. compressa*.

Primer	SSR	Expected product size	No. of alleles per sample		NBS Total
			Minimum	Maximum	
DPSSR001	TG(2*6)	129	1	1	20
DPSSR002	GA(2*10)	115	1	3	16
DPSSR003	CT(2*10)	142	1	3	12
DPSSR004	TA(2*9)	151	1	3	15
DPSSR005	TC(2*12)	121	1	3	14
DPSSR006	CGG(3*6)	117	1	3	9
DPSSR007	ATG(3*5)	158	1	6	17
DPSSR008	AAG(3*7)	154	0	0	0
DPSSR009	AGC(3*5)	126	1	2	18
DPSSR010	CAC(3*5)	146	0	0	0
DPSSR011	AAGA(4*5)	110	1	3	12
DPSSR012	AAAG(4*5)	132	1	3	11
DPSSR013	CCCT(4*5)	140	1	3	11
DPSSR014	TTTC(4*5)	134	1	3	19
DPSSR015	AAAAT(5*4)	159	1	3	20
DPSSR016	AGCTT(5*4)	105	1	3	20
DPSSR017	AACAG(5*4)	152	1	1	20
DPSSR018	CGAGG(5*4)	160	1	3	20
DPSSR019	GCCGTT(6*4)	141	1	3	18
DPSSR020	CGGGAC(6*6)	153	1	3	14
TOTAL ALLELES				55	28

∗NBS Number of bands

When the 20 SSR markers were used to genotype 20 *H. compressa* accessions, the number of effective alleles (*Ne*) varied from 1 to 2.7 with a mean of 1.65. Observed heterozygosity (*Ho*) ranged from 0.0 to 0.35 with a mean of 0.119. Gene diversity (*He)* ranged from 0.125 to 0.58 with a mean of 0.319. Shanon Information Index (*I*) ranged from 0.0 to 1.03 with a mean of 0.51 as shown in Table 4.32. The PIC values ranged from 0 to 0.795 with an average of 0.39. The result of the PIC value show that two primers (DPSSR01 and DPSSR17) are monomorphic (PIC of 0) while the rest are polymorphic (PIC >0.0). Primer DPSSR07 was the most polymorphic (Shannon Information index 1.03, PIC 0.795). Another highly informative marker was DPSSR12 (PIC 0.52, Shannon Information Index 0.67).

F statistics like F_{IS}, F_{IT} and F_{ST} for the entire population was 0.63, 0.72 and 0.28 respectively. Turkana accessions had the highest number of private alleles (0.17) while Tana River and Kwale had similar high percentages of polymorphic loci (77.8%) as shown in Table 4.33.

Table 4.32: Mean of diversity indices for primers used for genotyping *H. compressa*

	Na	Ne	I	Ho	He	uHe	PIC
DPSSR01	1.000	1.000	0.000	0.000	0.000	0.000	0.0
DPSSR02	2.500	2.039	0.760	0.350	0.465	0.517	0.481
DPSSR03	2.500	1.822	0.689	0.200	0.430	0.478	0.492
DPSSR04	2.250	1.811	0.632	0.200	0.400	0.444	0.442
DPSSR05	2.750	2.023	0.801	0.250	0.490	0.544	0.455
DPSSR06	2.500	1.771	0.678	0.100	0.420	0.467	0.436
DPSSR07	3.250	2.722	1.029	0.000	0.580	0.644	0.795
DPSSR09	1.500	1.235	0.250	0.000	0.160	0.178	0.164
DPSSR11	2.500	2.224	0.838	0.200	0.540	0.600	0.492
DPSSR12	2.250	1.821	0.665	0.250	0.445	0.494	0.516
DPSSR13	1.750	1.579	0.462	0.000	0.320	0.356	0.495
DPSSR14	1.500	1.368	0.298	0.250	0.205	0.228	0.276
DPSSR15	1.500	1.348	0.293	0.000	0.200	0.222	0.436
DPSSR16	1.500	1.173	0.206	0.050	0.125	0.139	0.310
DPSSR17	1.000	1.000	0.000	0.000	0.000	0.000	0.00
DPSSR18	1.500	1.348	0.293	0.000	0.200	0.222	0.460
DPSSR19	2.000	1.617	0.515	0.250	0.325	0.361	0.343
DPSSR20	2.250	1.874	0.676	0.050	0.435	0.483	0.416
OVERALL	2.000	1.654	0.505	0.119	0.319	0.354	0.390

∗Na- Number of different alleles, Ne-Number of effective alleles, I- Shannon Information Index, Ho-observed heterozygosity, He, expected heterozygosity, uHe –Unbiased expected heterozygosity, PIC –Polymorphic Information content

Table 4.33: F statistics estimates for different populations of *H. compressa*

Population	Tharaka	Tana River	Kwale	Turkana	Mean
No. of Private Alleles	0.1	0.1	0.1	0.17	
% Polymorphic loci	61.1	77.8	77.8	72.2	
F_{IS}					0.63
F_{IT}					0.72
F_{ST}					0.28

*F_{IS}-Inbreeding coefficient (detects inbreeding individuals relative to a sub population), F_{IT}, (mean deficiency of observed heterozygotes among individuals with respect to that expected for the total population), F_{ST}- Fixation index (Measure of population differentiation).

A total of seven markers revealed private alleles in some populations (Table 4.34). Turkana had three markers (DPSSR014), DPSSR018 and DPSSR020) showing private alleles. The DPSSRO16 SSR marker showed private alleles in Tana River and Kwale accessions while DPSSR014 revealed private alleles in Tana River and Turkana accessions.

Table 4.34: Markers detecting private alleles in different populations of *H. compressa.*

Population	Markers that detected private alleles		
Tharaka	DPSSR07	DPSSR15	
Tana River	DPSSR014	DPSSR016	
Kwale	DPSSR13	DPSSR016	
Turkana	DPSSR014	DPSSR018	DPSSR020

The analysis of molecular variance revealed that individual doum palm accessions drive diversity, with 54% more genetic variation among individuals than within individuals (24%) or among populations (22%).as shown in Table 4.35.

Table 4.35: AMOVA within and between populations of *H. compressa* from ASALs of Kenya.

Source	df	SS	MS	Est. Var.	%
Among Pops	3	45.875	15.292	0.946	22%
Among Individuals	16	93.300	5.831	2.378	54%
Within Individuals	20	21.500	1.075	1.075	24%
Total	39	160.675		4.399	100%

AMOVA-Analysis of molecular variance df- degrees of freedom, SS-sum of squares, MS-mean squares, Est.Var- Estimate of variance

4.5.4 Population structure

The optimal delta k value was K=3 (Figure 4.52). Therefore, STRUCTURE analysis based on the 20 SSR markers grouped *H. compressa* accessions into three clusters. STRUCTURE bar plots of *K*=2 to *K*=7 are presented as Figure 4.53. These STRUCTURE plots consistently show a lot of admixture in Kwale accessions throughout the *K* values.

Based on STRUCTURE analysis, six accessions were grouped in cluster 1 (comprising all accessions from Tharaka and one from Tana River) while eight accessions were grouped in cluster 2 (consisting all accessions from Turkana and three accessions from Kwale). Cluster 3 on the other hand comprised six accessions (four from Tana River and two from Kwale) as shown in Table 4.36

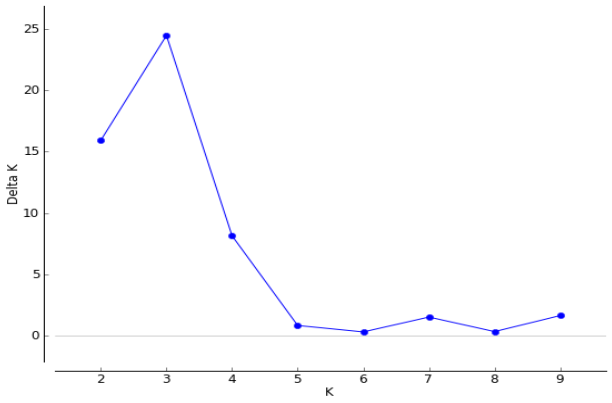


Figure 4.52: Optimal delta *k* for different *k* values among 20 *H. compressa* accessions based on 20 SSR markers

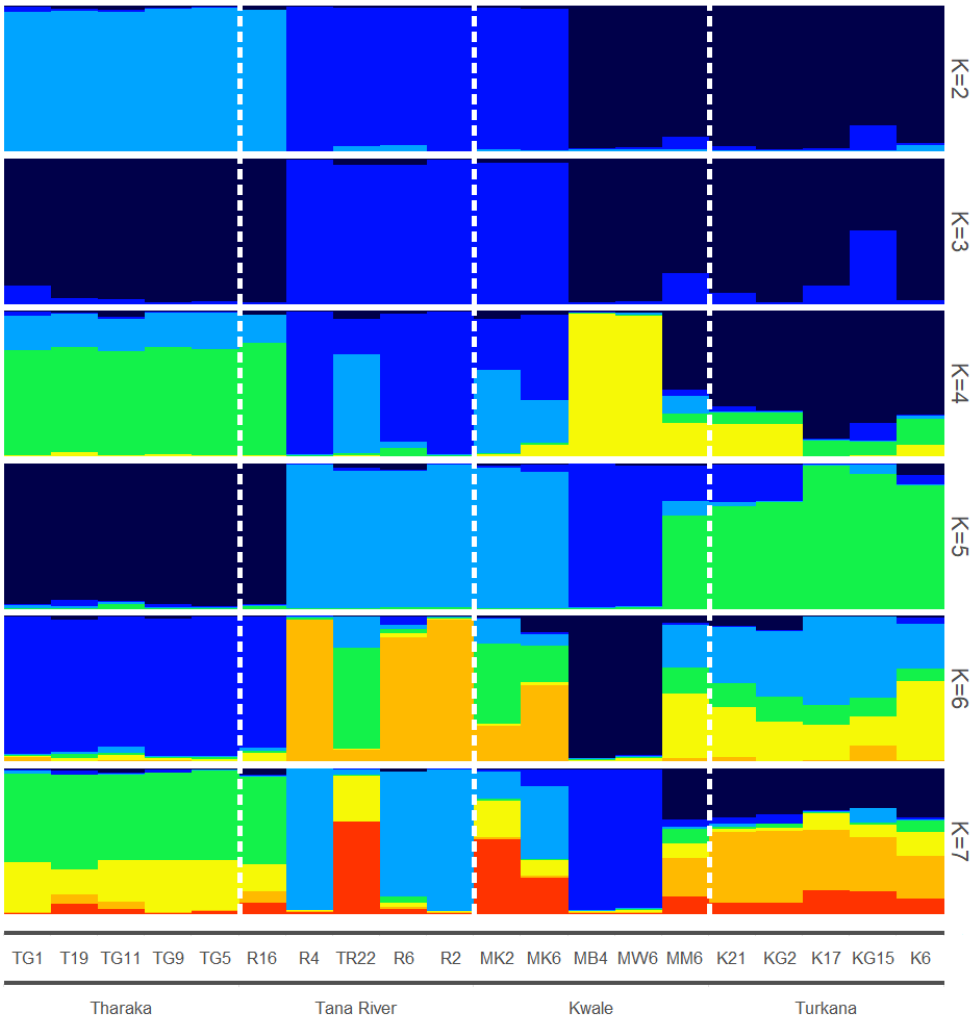


Figure 4.53: Population Structure bar plot showing *k*=2 to *k*=7 of 20 *H. compressa* accessions based on 20 SSR markers

Table 4.36: Inferred ancestry of 20 *H. compressa* accessions based on 20 SSR markers in each of three clusters using STRUCTURE software

	Sample	Population	Inferred cluster			Cluster assignment
			1	2	3	
1	TG1	Tharaka	0.96	0.01	0.03	1
2	T19	Tharaka	0.97	0.019	0.012	1
3	TG11	Tharaka	0.96	0.03	0.015	1
4	TG9	Tharaka	0.98	0.01	0.006	1
5	TG5	Tharaka	0.99	0.01	0.006	1
6	R16	Tana River	0.97	0.02	0.004	1
7	R4	Tana River	0.00	0.01	0.991	3
8	TR22	Tana River	0.04	0.01	0.948	3
9	R6	Tana River	0.05	0.01	0.937	3
10	R2	Tana River	0.00	0.01	0.986	3
11	MK2	Kwale	0.02	0.01	0.968	3
12	MK6	Kwale	0.01	0.02	0.973	3
13	MB4	Kwale	0.02	0.97	0.007	2
14	MM3	Kwale	0.02	0.97	0.011	2
15	MM6	Kwale	0.02	0.89	0.091	2
16	K21	Turkana	0.01	0.96	0.03	2
17	KG2	Turkana	0.01	0.99	0.004	2
18	K17	Turkana	0.01	0.98	0.015	2
19	KG15	Turkana	0.01	0.82	0.173	2
20	K6	Turkana	0.05	0.94	0.009	2

In addition, based on the proportion of membership of the four populations on the three clusters (Table 4.37), Tharaka and Turkana populations formed lone clusters 1 and 2 respectively while Tana River and Kwale populations were admixed populations (membership score less than 0.8). Lower levels of F_{ST} (0.15) were observed in cluster 2 accessions whilst higher levels of F_{ST} (0.45) were obtained for cluster 3 accessions. Cluster 2 had the highest expected heterozygosity values (0.43) followed by cluster 1 and lastly cluster 3 (Table 4.37).

Table 4.37: Membership for each population *H. compressa* populations in each cluster.

Given population	Inferred clusters			Number of individuals	Cluster Assignment
	1	2	3		
Tharaka	0.971	0.015	0.014	5	1
Tana River	0.215	0.012	0.773	5	Admixed
Kwale	0.017	0.572	0.41	5	Admixed
Turkana	0.016	0.938	0.046	5	2
F_{ST}	0.30	0.15	0.42		
H_e	0.34	0.43	0.32		

Population structure results were corroborated with the cluster analysis (Figure 4.54) and PCoA (Figure 4.55). The unweighted neighbor joining tree clustered the 20 accessions into three clusters with individual accessions being clustered in the same way they were clustered using STRUCTURE software.

Cluster 1 includes all accessions from Tharaka (TG1, TG5, TG9, TG11, T19) and one Tana River accession (R16), Cluster 2 includes all accessions from Turkana (KG2, K6, KG15, K17 and K21) and three accessions from Kwale (MB4, MW6 and MM6) while Cluster 3 includes four accessions from Tana River (R2, R4, R6 and TR22) and two accessions from Kwale (MK2 and MK6).

The PCoA analysis revealed that the *H. compressa* accessions are divided into three major clusters with Cluster 2 having accessions from Kwale and Turkana even though the accessions from Turkana and Kwale were more scattered on the PCoA (Figure 4.55). The first two components explained 45% of the genetic variance. The first and second principal components explaining 25.5% and 19.5% of the variation respectively.

Cluster 1 includes all accessions from Tharaka (TG1, TG5, TG9, TG11, T19) and one Tana River sample (R16), Cluster 2 includes all accessions from Turkana (KG2, K6, KG15, K17 and K21) and three accessions from Kwale (MB4, MW6 and MM6) while Cluster 3 includes four accessions from Tana River (R2, R4, R6 and TR22) and two accessions from Kwale (MK2 and MK6) as shown in Figure 4.55.

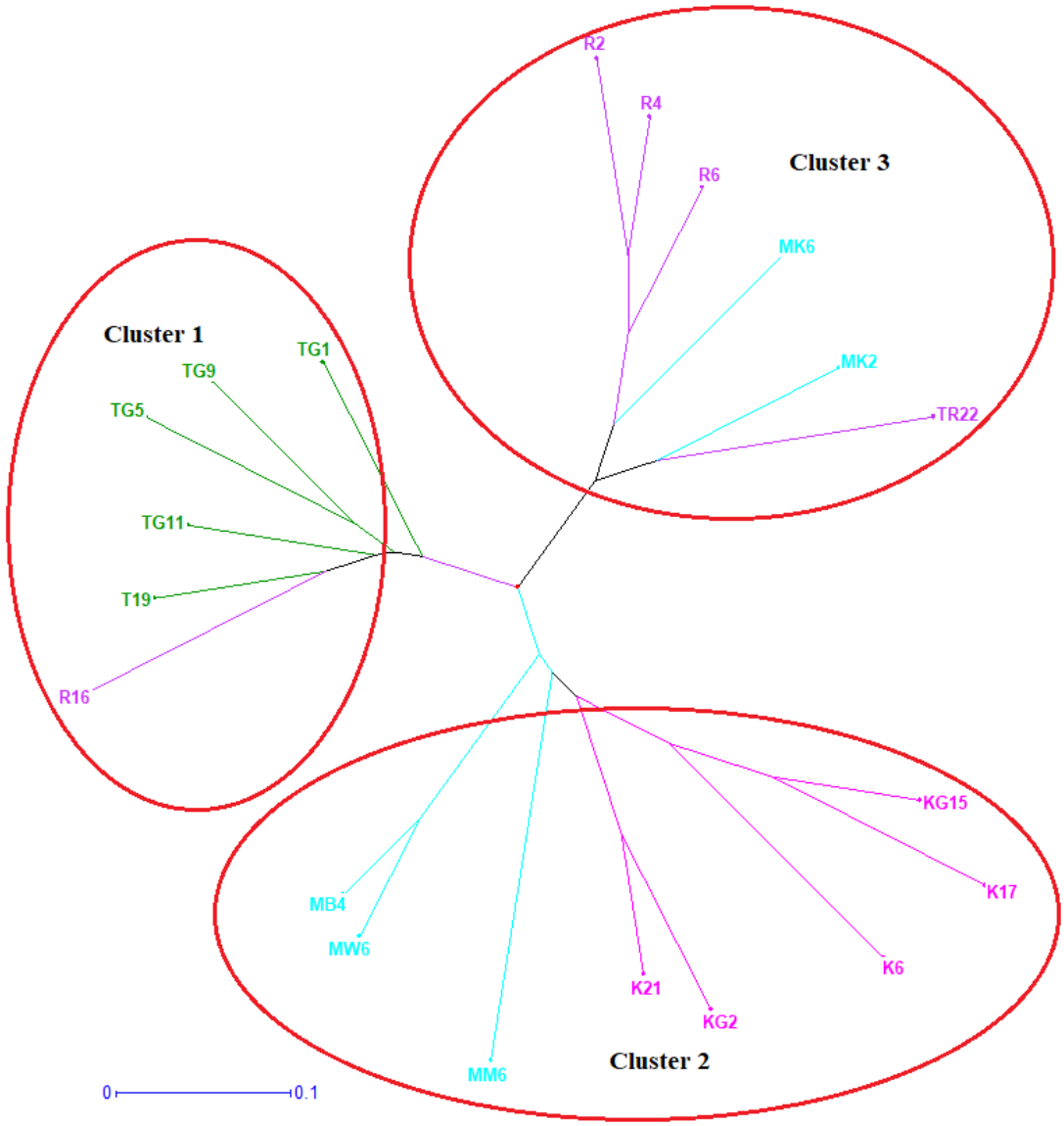


Figure 4.54: Neighbor joining dendrogram of 20 *H. compressa* accessions using 20 SSR markers derived from de_novo assembly of RNA transcripts

Cluster 1-TG1, TG5, TG9, TG11 and TG19 are Tharaka accessions while R16 is from Tana River
Cluster 2 -KG2, K6, K17, KG15 and K21 are Turkana accessions while MB4, MW6 and MM6 are Kwale accessions
Cluster 3 -R2, R4, R6 and TR22 are Tana River accessions while MK2 and MK6 are Kwale accessions

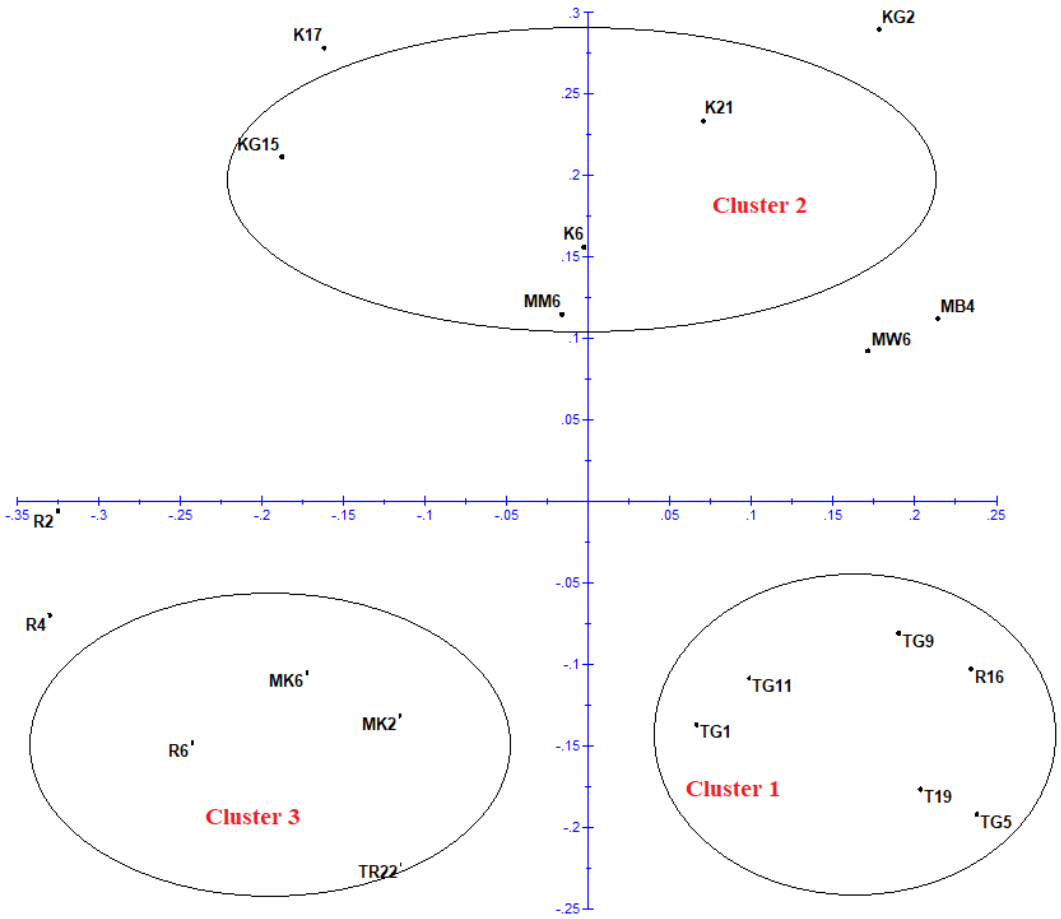


Figure 4.55: Factorial analysis of 20 *H. compressa* accessions based on 20 SSR markers

Cluster 1-TG1, TG5, TG9, TG11 and TG19 are Tharaka accessions while R16 is from Tana River
Cluster 2 -KG2, K6, K17, KG15 and K21 are Turkana accessions while MB4, MW6 and MM6 are Kwale accessions
Cluster 3 -R2, R4, R6 and TR22 are Tana River accessions while MK2 and MK6 are Kwale accessions

CHAPTER FIVE

DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

5.1 Ethnobotanical knowledge of doum palm

This study emphasizes the significance of *H. compressa* to the indigenous inhabitants in Kenya's ASAL regions. It is one of the wild edible perennial plants naturally occurring along riverine areas of ASALs (Amwatta, 2004; Stauffer et al., 2017).

In the present study, only farmers in Tharaka Nithi practiced some type of domestication of *H. compressa* accessions. When preparing the ground, they manage, weed, trim and intercrop it rather than clearing it from the field. The first phase in domesticating wild plants, according to Vodouhè et al. (2011), is field care due to established usage and sustenance. Farmers then weed and protect the plant during the second phase due to their interest in the plant. Plants that have become fully reliant on man for sustenance are domesticated (Charles, 1992). Therefore, based on this definition, *H. compressa* is thus classified as a semi-domesticated plant. The rationale for its preservation in Tharaka Nithi has been identified as leaf production, the high population distribution and the diligent nature of its citizens (Amwatta 2004; Icheria 2015). The majority of weavers buy bundles of leaves from farmers who maintain *H. compressa* on their farms. This income motivates the farmers to continue caring for the plant in their farms. Turkana, on the other hand, exhibited the highest *H. compressa* density and negligible intercropping with no management measures in place. The huge population of *H. compressa* contributes to the Turkana's lack of management practices on the plant. The most prevalent vegetation along rivers of Turkana are the doum palms and are hence dubbed Turkana's forests (*Turkana County Integrated Development Plan 2013-2017*, 2013). The Turkana people also rely on food aid and food relief from various government and non-governmental organizations (NGOs), which may have affected their perspective towards the conservation of *H. compressa* (Ng'asike & Swadener 2015). The presence of numerous NGOs in Turkana that assist agricultural activities is a likely explanation for the intercropping efforts seen in this county which is aimed at augmenting livestock products, wild fruits and relief food

(Ng'asike & Swadener 2015). Due to Kwale's growing tourism industry, there is a considerable demand for doum palm items thus the desire to domesticate *H. compressa* (Kenya inter-Agency rapid Assessment, 2014). Due to the damaging procedure of wine tapping, Tana River residents expressed interest in domesticating the plant. Overall, there is little evidence of domestication of *H. compressa* despite prior findings on the importance of doum palms in ancient Egypt (Clement 1992, Venugopal et al., 2017; Janick 2014).

The lack of interest in domesticating wild plants is due to ignorance of their potential (Onen & Oryem-Origa, 2017), inadequate government assistance and lack of value addition (Bvenura & Sivakumar, 2017). Domestication increases the productivity and preservation of wild edible species (Melaku & Ebrahim, 2021). To prevent dwindling of wild food plants, local communities should be involved in conservation efforts (Onen & Oryem-Origa, 2017). Indigenous communities should be urged to cultivate multipurpose wild and semi-wild edible plants on their own property in backyard gardens that they can combine with crops in farmlands and live fences, as well as to support the creation of local botanical gardens (Fentahun & Hager, 2009; Kidane & Kejela, 2021). *Hyphaene compressa* accessions in Tharaka Nithi could be losing its biodiversity albeit some domestication efforts are in place.

This study recorded 14 different applications of doum palm. The most commonly used plant materials were fruits, followed by leaves, stems and roots. A previous study on *H. compressa,* reported that the leaves were the most frequently utilized plant part (Amwatta, 2004). Across all the sampled regions, doum palm fruits for human food was the most recorded use. The fruits are crushed before consumption. Kwale residents also use the crushed mesocarp as a condiment. The importance of palms in food security has been consistently demonstrated in *Borassus aethiopum*, coconut, oil palm and date palm (Salako et al., 2018; Sadeghi and Kuhestani, 2014; Stauffer et al., 2014). Wild edible plants have been shown to be important sources of nutrients for vulnerable communities in rural areas particularly women and children who are often malnourished (Fentahun & Hager, 2009; Kidane & Kejela, 2021). Another benefit of wild edible plants is their year-round natural availability with notable overlap in

periods of acute food and nutrient scarcity (Fentahun & Hager, 2009; Ngome et al., 2017). In Turkana, for instance, the residents consume fruit sap as a refreshment and also blend crushed mesocarp with cattle blood to make a food called *'lokot'*. Maundu et al. (1999) previously described the use of *lokot* among Turkana inhabitants. The Turkana are a pastoralist minority community in Kenya, making them susceptible to persistent poverty (Ng'asike & Swadener 2015). In addition, the area undergoes long periods of drought, prompting supplementary nutrition. Under such vulnerable conditions, the Turkana often survive only on their cattle and plenty of doum palm fruits which are often in season all year round. According to Lokuruka (2008), doum palm fruit has a high amount of unsaturated fatty acids and is regarded as a healthier fruit than those derived from coconut or palm kernel oils. According to nutritional and phytochemical analyses of select wild plants done elsewhere, the majority of them have protein, fat and antioxidants (Bvenura & Sivakumar, 2017; Chakravarty et al., 2016; Fentahun & Hager, 2009). Therefore, wild edible plants are not only more nutrient-dense but also healthy. For example, *H. thebaica* fruits have been shown to be nutritious and have high radical scavenging action (Omar et al., 2020). Lack of interest especially by governments on the importance of wild edible plants to food security results in the loss of such plants' capacity to produce food, especially in ASALs (Shumsky et al., 2014). The widening disparity between the supply of food and the population increase is of concern. Wild edible plant species can fill this gap and provide additional nourishment (Melaku & Ebrahim, 2021). However, this is only possible if local communities are sensitized on the nutritional importance of wild plants with the goal of transferring them from the forests to the table (Bvenura & Sivakumar, 2017).

The benefits of wild edible plants to local communities have been elucidated by several studies as sources of food, medicine, non-timber products and income (Fentahun & Hager, 2009; Kidane & Kejela, 2021; Laibuni et al., 2020; Nyero et al., 2021; Suwardi et al., 2020). However, wild plants are often neglected and underutilized thereby face the risk of genetic erosion and decreased abundance (Onen & Oryem-Origa, 2017; Schunko et al., 2022). In addition, there is no motivation to cultivate the plants and no

management practices are in place to guarantee the longevity of the products derived from wild plants (Onen & Oryem-Origa, 2017).

In the current study, the Kwale community used immature inflorescence as a remedy for expectant mothers to avoid miscarriage and the crushed mesocarp as a pain killer. While in Tana River, leaf ash is used to treat burns. Other *Hyphaene* species have been utilized for ethno-veterinary purposes. For example, treatment of dog lung disease using *H. petersiana* (Cheikhyoussef & Embashu, 2013) and eye problems in cattle using *H. thebaica* sheaths (Mosissa et al., 2021). It has been demonstrated that wild agroforestry plants contain bioactive substances that function as antimicrobials and reduce intestinal parasites in livestock thereby enhancing their health, welfare and productivity (Salem et al., 2020). However, in this study, no ethno-veterinary uses of *H. compressa* were recorded. Earlier ethnobotanical investigations on the palm family reported a range of therapeutic properties, including anti-inflammatory, anti-diabetic, antibiotic and anti-neoplastic compound agents (Gruca et al., 2014; Nagata et al., 2011; Venugopal et al., 2017). Fruits of *H. thebaica* are known to be rich in antioxidants which diffuse free radicals and reduce the risk of oxidative stress and related diseases (Bvenura & Sivakumar, 2017; Hsu et al., 2006). Numerous edible wild plants have been shown to contain a variety of natural antioxidants, thus offering significant health advantages (Alabdallat & Bilto, 2015; Romojaro et al., 2013; Sanchez-Bel et al., 2015; Savo et al., 2019). Doum palm is used to treat a variety of diseases, including hypertension (El-Rashad & Hassan 2005) and diabetes (Khallaf et al., 2022). Doum palm's therapeutic properties can thus be investigated further.

In all four examined sites, the doum palm fruit and leaves provided major food sources for animals such as donkeys, camels and goats. However, the least recorded utility of leaves as animal feed was in Tana River and Kwale. This could be linked to the broad diversity of plants on the Kenyan Coast due to the humid agroecological zone, as opposed to the dry ASALs of Tharaka Nithi and Turkana. Previous results have shown that *H. compressa* fruits are supplementary feed for other mammals including elephants, baboons, monkeys and Mangabeys (Maundu & Tengnas 2005). The *H. compressa* leaves are evergreen despite the climatic conditions, providing enough

fodder for animals all year round. Oil palm fronds, fruit bunch and palm kernel oil have been used as poultry, fishes and livestock feeds (Bayão et al., 2020; Ishida & Hassan, 1997; Yusriani et al., 2021). Palm oil and palm kernel oil may replace butterfat when feeding young mammals (Tomkins & Drackley, 2010). Overall, it is beneficial to use wild plants for feed in ASAL regions and resource-limited areas. In doing so, the biodiversity of such wild plants will be preserved while also increasing the productivity of livestock (Quansah & Makkar, 2012). Most rural households depend on crop farming and livestock (Quansah & Makkar, 2012). However, some rural areas may not support agriculture due to unfavorable climatic conditions especially in the ASALs. This necessitates the use of alternative animal feed. Many wild plants come in handy in such environments (Quansah & Makkar, 2012; Salem et al., 2020).

In this study, utilization of handicraft varied from region to region. In Kwale, the leaves were used to make hats, fans, ropes, sieves and fishing nets. The leaves were used to produce baby baskets, mats, baskets, food warmers and fishing nets in Turkana. The Turkana who are largely pastoralists create makeshift homes called manyattas out of leaf petioles. In Tharaka Nithi, Kwale and Tana River, on the other hand, leaves were used to thatch mud houses. In all regions sampled, common leaf applications included thatching homes, weaving mats, trays and brooms. Amwatta (2004) reported comparable doum palm leaf applications. Sale of handicrafts and other products derived from doum palm is an important source of income especially in the ASALs (Abdullah et al., 2020). Other studies have encouraged weavers to be mindful of the needs and expectations of their clients. In that regard, consumers with higher income are attracted to handicrafts with aesthetic appeal, while those with lower income are attracted to products that are practical to use (Krishnaraj et al., 2022). For example, in this study, by using color and more intricate designs, native Turkana mat weavers have modernized their handicrafts. These modernized baskets are sold in Lodwar which is the largest town in North Western Kenya that houses many church organizations, voluntary and NGO’s (Chemelil, 2016). The modernization of Turkana's handicrafts has been driven by the ready market provided by employees of the many organizations in Lodwar. While the current study identified human food as the most important

application of doum palm, a recent evaluation identified leaves as the most important part of *H. compressa* plants. This is because the various handwoven products such as brooms, mats, baskets and ropes translates into an important source of income (El-Beltagi 2018).

Doum palm stem is used to build houses, granaries and fences. In some areas, wood prepared from male *H. compressa* trees is superior to wood prepared from female *H. compressa* trees (El-Beltagi 2018). In Sub-Saharan Africa, gathering non-timber forest products for use in building homes is common. This practice is motivated by poverty and a lack of funds to invest in higher-quality buildings and non-wood alternatives (Schaafsma et al., 2014). However, in order to preserve biodiversity and reap long-term benefits, harvesting or logging of non-timber items should be regulated (Talukdar et al., 2021). The long maturation time in *Hyphaene Spp* (Orwa et al., 2009), predisposes it to genetic erosion if logging is not regulated especially in Tharaka Nithi where the populations are declining.

Only Turkana communities recognized the significance of doum palm roots in preventing soil erosion. The cultivation and management of native wild forests contributes to the restoration of degraded ecosystems and the conservation of biodiversity (Melaku & Ebrahim, 2021). To combat desertification, community organizations in Northern Kenya have planted a variety of wild plants, such as *Hyphaene spp., Azadirachta indica, Commiphora Africana* and *Suaeda monoica,* among others, in an effort to stabilise mobile sand dunes (Olukoye & Kinyamario, 2009).

Existing doum palm plants in the study areas had biotic and abiotic stress indicators. In Tharaka Nithi, for example, there was overharvesting of doum palm leaves and insect damage. Stunting in *H. thebaica* has been demonstrated to result from overharvesting, which also causes biodiversity loss (Kahn & Luxereau, 2008). The abiotic stresses observed in Tana River and Kwale were drought and salinity. *Hyphaene compressa* is one of the rare persistent wild plant species growing in saline soils that can tolerate irregular unfavorable environmental conditions (Orwa et al. 2009; Venugopal et al. 2017). Both Turkana and Coastal saline soils have a high

concentration of sodium ions (Mugai, 2004). Wild plants are known to be more resilient to water stress compared to their domesticated relatives (Bvenura & Sivakumar, 2017; Fentahun & Hager, 2009). Thus, stress-tolerant wild plants are a potential source of new, more effective biotechnological tools for genetically enhancing stress tolerance in crop plants (Boscaiu et al., 2012). The primary biotic stress observed in Kwale and Tana River was the extraction of wine from this plant which involves stem trimming. It has been demonstrated that wine tapping in *H. petersiana* negatively affects the population structure and palm regeneration. As a result, remarkably fewer mature trees were observed, a situation that is made worse by the destructive wine tapping methods (Babitseng & Teketay, 2013). Some of these destructive methods comprise cutting down the stems, burning and pruning which prevents other uses like fruit production from being realized (Babitseng & Teketay, 2013; Mba et al., 2019).

5.2 Morphological diversity of *H. compressa*

In the current study, phenotypic heterogeneity in 90 *H. compressa* accessions from the ASALs of Kenya was examined using morphological parameters. A total of seven quantitative and nine qualitative descriptors were used. Quantitative features showed high variation within the individual doum palm trees studied as well as among the several ASAL regions sampled. This study identified five morphotypes of *H. compressa* based on cluster analysis. Whilst morphotype 4 was region specific (Kwale), all the other morphotypes were not region specific. This study revealed a broad range of variability in the traits examined with *H. compressa* height (cv=38.4) and leaf breadth (cv=25.7) being the most variable. The tallest accessions were those from Tharaka Nithi (mean 13.5m) while Kwale (mean 5.65m) had the shortest accessions. According to Stauffer et al. (2014), *H. compressa* plants can grow up to a maximum height of 20m thus the species is highly polymorphic. Morphological traits such as fruit weight, fruit length and fruit breadth have been previously used for the evaluation of morphological diversity of *H. thebaica* accessions in Egypt (Khalil et al., 2020)

The fruit weight ranged from 48.2 g to 148.8 g in this study. A study on African palms reported that fruits were bigger in regions where there was no water stress (Stauffer et al. 2017). Phenotypic plasticity as a result of resource scarcity (Sultan, 2003) could explain the tiny size of *H. compressa* fruits in Turkana. However, fruit sizes in Kwale ranged from very large (morphotype 5) to very small (morphotype 4), which may be attributed to the varietal differences. The fruit quality features were likewise highly variable, with Kwale having the most diverse fruits. Mature fruits ranged in color from reddish brown to brown to orange. When immature, Doum palm fruits are predominantly green, but as they ripen, they turn orange, brown, red, or yellow (Stauffer et al. 2017). Other studies, however, report that mature fruits tend to be orange brown in color (Maundu and Tengnas, 2005).

This research showed that several quantitative qualities were strongly correlated with one another. Correlations between ecologically significant plant trait combinations may be an adaptive result of natural selection favoring certain trait combinations over all the others (Westoby et al., 2002). In the current study, height was positively correlated with all the three fruit traits examined. That is the taller the plant, the bigger, wider and heavier the fruit. Taller species or growth forms are known to have larger fruits (Moles et al., 2004; Westoby et al., 2002; Wright et al., 2007). Different explanations have been put forth among them is that very large fruits can only be physically supported by bigger growth forms (Wright et al., 2007). With the exception of petiole length, fruit weight positively correlated with all other quantitative variables investigated in this study. Thus, selection of fruit weight in *H. compressa* through breeding would automatically improve all the other positively associated traits. The positive correlations of fruit traits with other quantitative traits has been demonstrated in prior studies (Zou et al., 2020). However, a linear mixed effects model fitted to predict fruit weight with all the other quantitative variables did not establish any effect of height, leaf length, leaf breadth and petiole length on the fruit weight.

Fruit traits are important to the farmers since they are able to assign varietal names to palms based on fruit features. During the sampling exercise in Kwale, farmers differentiated *H. compressa* into three varieties known as Mkoko, Mbiye and Mkoma

based on the fruit traits. In Tharaka Nithi, some farmers distinguished the varieties, although they did not assign them varietal names. Turkana and Tana River farmers were unable to distinguish the varieties. The importance of palm fruit features to farmer varietal assignment has been previously reported (Simozrag et al., 2016). The heterogeneity of fruit traits in this genus is well documented (Stauffer et al., 2014). In fact, the polymorphic nature of the fruit has led to over 33 synonyms of the species *H. compressa* (Stauffer et al., 2018). The variability in fruit traits observed particularly in Kwale indicates that the accessions in this region are highly heterogeneous.

The accessions had significant differences in all seven quantitative features per region which is attributable to the diversity of the examined traits. This is consistent with a study that analyzed the morphological and molecular diversity in *H. thebaica*, which reported significant variation in all morphological parameters (Khalil et al. 2019). The same study also reported that morphological and genetic analyses were useful in evaluating *H. thebaica,* despite the fact that they revealed various connections among the samples analyzed.

Using both quantitative and qualitative features for PCA clustering revealed two large clusters, with a subset of samples from Kwale clearly forming their own cluster, apart from the rest of the accessions which appear to be unrelated to the others. Based on linear discriminant analysis, these samples also appeared to have no overlap with any samples from the other regions. This, however, cannot be utilized to discriminate this group because advanced markers would be necessary to genotype them (Haider et al. 2015). According to the results of the linear discriminant analysis, some Kwale accessions clustered with Turkana accessions while Tharaka accessions only clustered with Tana River accessions. This implies that some of the varietal variations observed in certain accessions in Kwale may also be present in some accessions in Turkana.

The stems of most palm species are cylindrical, elongated and unbranched. *Hyphaene compressa,* on the other hand, possesses dichotomizing trunks, a *Hyphaene*-specific trait in which the base stem is overbuilt to accommodate later dichotomous branches (Tomlinson and Huggett, 2012). However, 46.7% of the Kwale accessions lacked dichotomizing trunks and were somewhat shorter than accessions in Tharaka Nithi.

This shows that the *H. compressa* accessions at the Kenyan Coast, particularly Kwale, are more variable than those from the other regions. A previous study noted that tall plants have high maintenance costs for stems and face a drawback of transport of water to the maximum height (Westoby et al., 2002). Light interception and reproductive dispersal are other benefits of height in plants. Some accessions from Kwale in the current study had unusually long petioles, substantially smaller fruits and very short stature. The long petioles could be an evolutionary advantage for the leaves to intercept light due to their short stature.

Doum palm is genetically heterogeneous, as evidenced by the distribution of various samples in distinct clusters. Therefore, the diverse morphotypes described in the current study may not be directly influenced by their surroundings. This is supported by Euclidean cluster analysis and PCA clustering, which demonstrated a high level of variability. In this study, cluster analysis demonstrated phenotypic variation and heterogeneity within samples from the same location. Accessions from Kwale, for example, were assigned to morphotypes 1, 3 and 5. Kwale also had some accessions that formed a single cluster, morphotype 4. As a result, they were the most diverse, with some accessions having very tall stems with very huge fruits and others having very short trunks with smaller fruits. This heterogeneity was also observed among Tharaka (four morphotypes), Tana River (three morphotypes) and Turkana accessions (three morphotypes). This variation among accessions from the same region has previously been noted (Haider et al., 2015).

There are no recognized doum palm regeneration and breeding programs available. The identification of existing morphotypes will assist farmers and breeders in identifying accessions for improvement and conservation. For weaving, Tharaka farmers prefer doum palm with longer and wider leaves. Such data can assist breeders in selecting traits for improvement and bulk production. According to the current study, Tharaka (90.5%) and Kwale (9.5%) had the longest and widest leaves that can be used for weaving (morphotype 5). The fact that some accessions from Turkana, the most arid location of all the examined areas, has accessions with long and wide leaves (morphotype 2; Mean leaf length 116.7), just like Tharaka, which receives somewhat

more rainfall than Turkana, suggests that the variance in leaf lengths and breadths in *H. compressa* may have a higher genetic component of the total variation. In addition, farmers might have selected these accessions for leaf length and hence improved it over time. In fact, large-scale weaving with doum palm leaves takes place in these areas (Tharaka and Turkana). In the current study, accessions from Tharaka Nithi produced the biggest fruits making them excellent candidates for improvement. Fruit traits typically have significant effects on plant breeding, with fruit weight being a key target when breeding fruit crops (Zou et al., 2020). However, the positive correlation between fruit weight and height (r^2 =0.521) demonstrated in this study may be a challenge since farmers would want accessible trees for fruit picking.

Costapalmate, fan-shaped leaves with entire edges, curving costa and curved thorns on the leaf stem characterize *H. compressa* palms (Orwa et al., 2009). The petiole length appeared to be a crucial feature in distinguishing *H. compressa*. Morphotype 4 had substantially longer petiole length. There was a negative relationship between petiole length and fruit characteristics. That is, the longer the petiole, the smaller the fruits and vice versa. Petioles are valuable resources for communities, particularly pastoralists in Kenya who utilize them for furniture and house construction (Amwatta, 2004). Despite having longer petioles, morphotype 4 is a short accession in the current study. For this reason, local communities would prefer this accession because of ease of harvesting the petioles.

Morphological diversity is advantageous in the preliminary assessment of doum palm genetic resources. The superior features described, however, cannot adequately resolve the variations in diversity and should be proven if they are genetically driven. This is the first study to the best of my knowledge to phenotypically characterize *H. compressa* accessions from the ASALs of Kenya.

5.3 Genetic diversity of *H. compressa*

To determine the population structure and genetic diversity of *H. compressa* accessions, two comparison methodologies, that is, reference-based and *de_novo*-based approaches were applied. *Phoenix dactylifera* was utilized as a reference

genome in the reference-based assembly. *Hyphaene compressa* did not have an assembled genome at the time of the investigation. Therefore, *P. dactylifera* was used as a confamilial genome. A confamilial reference genome can be used to offer equivalent estimates of diversity in the absence of a reference genome of the same species (conspecific) or genus (congeneric) (Brandies et al., 2019; Galla et al., 2019).

There were variations between the two techniques in terms of SNP abundance, quality scores and TS/TV ratios. For example, the reference-based assembly had more SNPs (23416) when compared to the *de_novo*-based assembly (2096). In determining the number of SNPs in olive cultivars, reference-based assembly had previously been shown to outcompete *de_novo* assembly (D′Agostino et al., 2018). It has also been noted that the settings used during sequence assembly, as well as the type of assembly, influence the quantity and depth of SNPs detected (Bohling, 2020). Furthermore, stricter parameters are typically employed for *de_novo* assembly. The GBS analysis of *H. compressa* accessions demonstrated significant SNP diversity, with transition SNPs (purine-purine or pyrimidine-pyrimidine) being the most abundant. This high frequency of transition SNPs has been documented in numerous plants including *Capsicum annuum, Vigna unguiculata, E. guineensis and Camelia sativa* (Hyun et al., 2020; Z. Luo et al., 2019; Pootakham et al., 2015; Taranto et al., 2016; Xiong et al., 2016). The C-T transition SNPs are the most common (Edwards et al., 2007). However, in *de_novo*-based assembly, a low TS/TV ratio was reported. This pattern has been previously reported and has been related to variation in SNP calling algorithms (Shafer et al., 2017). Despite their differences, structure and PCA generated comparable results. A comparable study, which employed a *de_novo* technique, a confamilial reference and a congeneric reference to assess the phylogenetic connection of the *Amaranthus* genus, yielded different SNP counts but similar phylogenetic trees (Shafer et al., 2017). Other studies have demonstrated that SNP abundance differs between reference-based and *de_novo* assemblies, while population clustering remains consistent (DAgostino et al., 2018; Shafer et al., 2017; Stetter & Schmid, 2017).

STRUCTURE analysis separated the *H. compressa* accessions into two gene pools. The PCA and DAPC analyses concurred with the STRUCTURE results. Turkana (Northern Kenya) accessions were grouped into cluster 1. Cluster 2 had the most genotypes, which included accessions collected along the River Tana basin locations (Tharaka Nithi and Tana River) and Kwale. Furthermore, an admixture of accessions from Kwale with mixed ancestry of Tharaka Nithi and Tana River was reported. This gene flow could be the result of genetic exchange between Kwale, Tharaka Nithi and Tana River accessions. This supports the morphology results obtained earlier in this study, which found that accessions from Kwale have the most morphological variations when compared to other areas. Genetic diversity analysis also revealed that Kwale accessions had the most polymorphic sites with a lot of admixture between Tharaka and TanaRiver populations. In this study, all four populations of *H. compressa*, had greater observed heterozygosity than expected heterozygosity indicating high genetic diversity within the populations (Sharma et al., 2016).

In this study, the overall F_{ST} value for *H. compressa* populations was 0.074. The fixation index (F_{ST}) is a useful tool for determining population divergence between populations (Nassiry et al., 2009). An F_{ST} of 0 - 0.05 is regarded as minor, 0.05 - 0.15 is considered moderate while 0.15 and above is considered very high (Nassiry et al., 2009). Based on the obtained F_{ST}, the genetic differentiation amongst *H. compressa* accessions was moderate. F_{ST} values obtained from STRUCTURE population clustering using both *de_novo* and referenced based approach, revealed higher genetic differentiation within cluster 1 (accessions from Turkana) than within cluster 2 accessions from the River Tana Basin. Cluster 2, on the contrary showed higher expected heterozygosity *(He)* than cluster 1. This suggests that accessions from the River Tana basin are highly diverse. The high expected heterozygosity and little genetic differentiation among *H. compressa* accessions along the River Tana Basin (cluster 2) could be attributed to gene flow-induced genetic exchange. Despite the fact that Tharaka Nithi is located around 262 and 630 kilometers from Tana River and Kwale, respectively, there appears to be a lot of genetic exchange between accessions from these three counties. This could be attributed to the flow of the River Tana, which runs through both Tharaka and Tana River counties and may act as a route of

germplasm dispersal. This could account for the high mixed ancestry shown in DAPC and STRUCTURE analysis as well as the Tharaka samples' proximity to Tana River accessions on the PCA. The River Tana is Kenya's longest and most important drainage basin. The river originates in Kenya's highlands and flows to the Eastern ASAL plateaus and Kenya's coast (Kitheka & Ongwenyi, 2002). Being a riverine plant, *H. compressa* population structure in the Kenyan Coast is significantly influenced by seed dispersal through the rivers. Dispersal of germplasm is crucial for biodiversity conservation since it promotes population dynamics, gene flow and functional connection between locations (Traveset & Rodríguez-Pérez, 2018). Effective seed dispersal encourages gene flow, promotes genetic diversity and reduces population genetic differentiation (da Paschoa et al., 2018). According to migration rates calculated by MIGRATE-n, there is asymmetric gene flow throughout the River Tana basin. This is consistent with the hypothesis that the population structure of *H. compressa* along the Coast is driven by germplasm dispersal via the River Tana. High migration rates between Kwale and Tharaka were also noted. Some Kwale accessions clustered with Tharaka accessions based on phylogenetic analysis, hence validating the gene flow findings.

Turkana has limited gene flow into or out of the region, which could cause its *H. compressa* population to differ from the other populations. This was confirmed by STRUCTURE analysis, PCA, DAPC and Neighbor-Net network, which distinguished Turkana accessions from the others by clustering them separately. The geographic seclusion of Turkana from the other populations may be the cause of this difference. Populations of Turkana that are isolated are unable to reproduce with other populations due to the physical barrier. Turkana is thought to be entirely a dryland with little rain-fed agriculture thus experiences different selection pressure (Barrow & Mogaka, 2007)

The *H. compressa* populations' negative F_{IS} values show low levels of inbreeding, and excess of heterozygotes possibly due to the mating mechanism. *Hyphaene compressa* is a dioecious plant (Stauffer et al., 2014), which supports obligate cross-pollination, which in turn enhances intrapopulation genetic variation (da Paschoa et al., 2018; Muyle et al., 2020). Dioecy has been shown to encourage outcrossing (Charlesworth,

2006). The AMOVA test, which indicated higher (92.7%) within population diversity than among population diversity (7.3%), further confirmed the high genetic diversity and little inbreeding in *H. compressa*.

Understanding *H. compressa* genetic diversity and population dynamics can help with future selection, breeding and effective conservation mechanisms. The preservation of all groups with high genetic variation should be considered. The two detected clusters must be taken into account when selecting *H. compressa* accessions for conservation in order to maintain the high level of population variability. This can be accomplished through extensive germplasm collection and *ex situ* conservation, particularly for cluster 2, which had the most diversity. This is the first study to describe the use of SNPs to characterize *H. compressa* accessions using the GBS technique. The SNP markers are relatively stable, frequent and specific to genomic areas, making them excellent for use in marker assisted selection (MAS) and diversity studies to promote germplasm conservation.

5.4 Differential gene expression due to salinity stress

In the current study, *H. compressa* accessions from Tana River, Tharaka and Turkana were exposed to low, medium, high and very high salinity levels for eight weeks. This duration was expected to increase the possibility of identifying salinity induced changes and not osmotic effects (Munns, 2002). Previous works determined that within minutes to hours of salt exposure, osmotic effects are observed but it takes longer for salt induced effects to manifest in plants (Blake & Munns, 2017; Yaish et al., 2017). The salinity effects also become pronounced over weeks of salinity exposure therefore salt sensitive accessions can be clearly distinguished (Munns, 2002). *Hyphaene compressa* accessions from three regions showed diverse response to varying salinity levels. Overall, *H. compressa* was moderately to highly tolerant to salinity stress consistent with its natural habitat. Morphological changes were induced by salinity stress in *H. compressa*. Reduced shoot growth was observed with increase in salinity stress. Tharaka accessions had the highest negative shoot growth while Tana River accessions had the least. Leaf length of all accessions was similarly reduced with increase in salinity stress. Tana River accessions had the least chlorotic

scores after eight weeks of salinity stress followed by Turkana and lastly Tharaka. Salt injury in plants is manifested as yellowing and or browning after weeks of salinity stress (Munns, 2002). Salt tolerant plants have the ability to compartmentalize salt in the vacuoles or may have low uptake levels and thus have few or little injury scores (Munns, 2002; Negrão et al., 2017). It has also been reported that some salt tolerant species are able to maintain the green nature of the leaf for longer durations even with increased salinity (Negrão et al., 2017).

Reduction in water content was observed across all accessions with increase in salinity stress. This is attributed to dehydration at the cellular level as salinity levels increase (Qin et al., 2010). When plants are salinity stressed, they are immediately affected by osmotic stress which limits the water uptake by the roots (Polash et al., 2018). More stomata on abaxial leaves of control samples was obtained compared to salinity stressed accessions. However, these differences were not statistically significant. Stomatal closure is mediated by abscisic acid (ABA) during salinity stress. This in turn limits the amount of water uptake and of CO_2 which ultimately lowers photosynthesis and biomass (Polash et al., 2018). Salinity stress was seen to reduce chla, chlb and carotenoids. Higher carotenoid levels were obtained in Tana River accessions across all salinity levels compared with Turkana and Tharaka. Carotenoids are composed of carotenes and xanthophylls. Carotenoids are important in light harvesting of photosynthesis (Zakar et al., 2016). Carotenoids scavenge free radicals generated during photosynthesis (Chaves et al., 2009; Das, 2013; Zakar et al., 2016). Higher carotenoid levels have also been obtained in tomatoes exposed to salinity stress (Borghesi et al., 2011).

With increase in salinity stress, there was increased ash content. Tana River accumulated the most Na^+ in the shoots than roots despite having less chlorotic scores. This suggests that Tana River accessions have a unique ability to tolerate the high internal Na^+ levels in the shoots/leaves (Munns & James, 2003). Tolerance may be attributed to the accumulation of Na^+ ions in the vacuole or other specialized cells such as epidermal cells (Munns & James, 2003). Root Na^+ accumulation was observed to increase with increase in salinity stress. Tana River accessions on the other hand

accumulated the least Na^+ in the roots. Tana River accessions also accumulated more K^+ in the shoots and roots while Tharaka and Turkana had reduced K^+ levels in roots and shoots with increase in salinity. Maintaining N^+/K^+ ratio is important for plant survival in saline environments (Gupta & Huang, 2014). In this study, proline content was seen to accumulate with increase in salinity stress. Tana River accessions accumulated the most proline at 300mM salinity level. In a similar study to assess the differential gene expression due to salinity stress in *P. dactylifera*, proline content was increased up to 15.2 fold due to salinity stress (Xu et al., 2022). Amino acids such as proline have been observed to accumulate in plants due to stress as they act as ROS scavengers, antioxidants and alleviate the effects of abiotic damage (Alhasnawi, 2019; Das & Roychoudhury, 2014; Rejeb et al., 2014).

Reduction in shoot and root dry biomass was evident with increase in salinity stress. Tana River accessions accumulated the most shoot and root biomass which was significantly different from Tharaka accessions. Previous studies on various plants have documented reduction in root and/or shoot biomass due to salinity stress (Kumar et al., 2021; Tao et al., 2021; Zhou et al., 2014). Overall, the reduction in biomass and growth parameters could be attributed to effects of salinity stress on cell division and elongation (Kumar et al., 2021). Plants accumulate more biomass in roots to increase water uptake following long term exposure to salinity stress (Sabino et al., 2021). This might have been the case for Tana River accessions which had higher root biomass accumulation. The Salt Tolerance Index (STI) is reduction in biomass relative to the control and is indicative of salinity tolerance in plants. Highly tolerant plants have minimal reduction in this ratio (Lima et al., 2017). Based on sensitivity to biomass reduction, Tana River accessions were the most tolerant (STI = 0.56) followed by Turkana (STI = 0.5) and Tharaka (STI = 0.45) at 300mM salinity level which is considered very high salinity. It is important to note that none of the *H. compressa* accessions died during the course of the eight weeks which affirms the tolerance of this palm to salinity.

Transcriptomics analysis using RNA-Seq was done for Tana River accessions. This study reports the transcriptome of *H. compressa* due to salinity stress. The leaf

transcriptome was used since it is the center of many metabolic processes in plants (Yaish et al., 2017). RNA-seq of *H. compressa* and *de_novo* assembly of transcripts resulted in 498,082 transcripts. BUSCO assembly metrics showed completeness score of 89.4% suggesting successful sequencing, assembly and reliable representation of the genes present (Feron & Waterhouse, 2022). A total of 92,135 unigenes were obtained in this study. In total, 62,472 unigenes were annotated by any of the seven databases (NR, NT, KEGG, KOG, Interpro, SwissProt and GO). Based on NR annotation, *P. dactylifera* accounted for 50.5% of all the annotated unigenes. This also supports the closeness of *P dactylifera* to *H. compressa*. From this study, 92,135 unigenes with 60% complete ORFs, were obtained. In total, 8611 DEGs were identified between salt stressed and control plants which shows the pathways and genes related to salinity tolerance in *H. compressa*. Previous studies reported salinity induced DEGs in date palm (Al-Harrasi et al., 2018; Radwan et al., 2015; Xu et al., 2022; Yaish et al., 2017) and oil palm (Ferreira et al., 2021). In a study of salinity induced transcriptomics in *P. dactylifera* exposed to 1000mM salinity stress for 13 days identified 15,151 DEGs (Xu et al., 2022). In this study, a total of 3722 of the DEGs were upregulated while 4889 were downregulated. Differentially expressed genes have been shown to increase with the increase in the number of days of salinity stress (Xu et al., 2022).

Among the most significantly up regulated genes were NPL4-Like protein, Coatamer sub unit beta, beta glucosidase, Boi related E3 ubiquitin protein. In the current study high levels of NPL4–Like protein (CL5122.Contig7_All, L_2fc =8.56) were expressed. The NPL4 has been reported to be a salinity tolerance gene (Jogawat et al., 2016) and a cofactor of Cell Division Cycle 48 (CDC 48). The CDC48 complex regulates the ubiquitin dependent degradation of proteins due to ROS (Li et al., 2022). The CDC48 is normally induced due to oxidative stress and has been shown to be linked to antioxidants such as ascorbate peroxidase (APX) and catalase (CAT) in response to ROS (Bègue et al., 2019). In the present study several antioxidants were upregulated including ascorbate peroxidase (Unigene25198_All, L_2fc = 1.994) and glutathione S-transferase (L_2fc = 5.54). Beta-glucosidase was highly up regulated (CL1015.Contig5_All, L_2fc = 7.4). The accumulation of beta-glucosidase has been

demonstrated to induce antioxidants like flavonols which ultimately leads to improved tolerance to abiotic stress by reducing ROS accumulation (Baba et al., 2017). In fact, flavonol biosynthesis was the most enriched KEGG pathway in the present study. Botrytis Susceptible Interactor (BOI) related genes (BRGs) are a group of Really Interesting New Gene (RING) E3 ligases which have been shown to confer stress tolerance in plants (Luo et al., 2010). The E3 ligases are reported to be involved in the degradation of abiotic stress related proteins (Al-Saharin et al., 2022). It has been established that protein ubiquitination regulates SOS (Salt Overly Sensitive) and MAPK (Mitogen-activated Protein Kinase) pathways whereby, E3 ligases are involved in SOS pathways and MAPK cascade, ABA signaling and ROS homeostasis (Wang et al., 2022). Up regulation of E3 ligases (CL6945.Contig4_All, L_2fc = 7.40) in *H. compressa* was observed due to salinity stress.

The current study recorded a significant up regulation of Chloroplast Enhancing Stress Tolerance (CEST) gene (CL4064.Contig6_All, L_2fc = 7.0). The CSET is a chloroplast protein with the functions of chloroplast development and tolerance to stress (Yokotani et al., 2011). Overexpression of this gene in *Arabidopsis thaliana* conferred resistance of the transgenics to various stresses (Li et al., 2021). In addition, this gene has been shown to reduce photo oxidative stress in the same plant (Yokotani et al., 2011).

A total of 25 DEGs were significantly enriched (FDR< 0.05) based on GO annotation of the DEGs. Photosystem, thylakoid part and photosynthetic membrane were the three top most enriched GO terms in the cellular processes. Photosystem II (PSII) is among the four major components of photosynthesis electron transport and the most sensitive to salt stress which causes disintegration of the complex (Jajoo, 2014). Photosynthesis was the most enriched in the biological process in concurrence with previous studies (Yaish et al., 2017;. Zhang et al., 2016). In the current study, almost all genes involved in the light reaction were down regulated including PSI genes (PSA), PSII genes (PSB) and photosynthesis antenna proteins (LHC). Photosynthesis antenna proteins were the most significantly down regulated pathways in *P. dactylifera* exposed to salinity stress (Xu et al., 2022). The down regulation of these genes has also been reported in cotton (Zhang et al., 2016). Down regulation of these genes

indicates disablement of the PSI light harvesting machinery in *H. compressa* as a result of salinity stress. This is corroborated by reduced chlorophyll level that was detected prior to transcriptomics experiments. The PsaA was upregulated in this study. The PsaA and PsaB dimers have been associated with binding of cofactors like chlorophyll a and β carotenes for light harvesting and electron transfer reactions (Caffarri et al., 2014). Tetrapyrrole binding was the most enriched molecular process based on GO annotation of the DEGs. Tetrapyrrole enrichment has been previously reported in *P. dactylifera* accessions exposed to 300mM salinity stress (Yaish et al., 2017). Tetrapyrrole binding seems to play an important role in *H. compressa* salinity tolerance mechanisms. Tetrapyrroles such as heme and chlorophyll are important in photosynthesis and respiration. They have also been implicated in abiotic stress tolerance mechanisms by activation of the ROS detoxification (Nagahatenna et al., 2015; Terry & Smith, 2013).

Plants respond to salinity stress through well-defined pathways that include stress sensing, signal transduction as well as activation of stress responsive genes and metabolites (Ryu & Cho, 2015). The KEGG pathway enrichment revealed complex molecular responses to salt stress including signal transduction, metabolism of secondary metabolites and stress response. Most unigenes were categorized under metabolism using the KEGG annotation. Some of these metabolism processes included global and overview maps, carbohydrate metabolism, biosynthesis of secondary metabolites, cofactors and vitamins. It has been demonstrated that during salinity stress, carbohydrates accumulate in plants. These carbohydrates function as osmoprotectants, carbon storage and scavenge ROS during salinity stress (Gupta & Huang, 2014). Metabolic processes and synthesis of secondary metabolites (P=1.3e-10) were the most enriched KEGG pathways. Other enriched KEGG pathways included flavonoid biosynthesis, phenylpropanoid biosynthesis, starch and sucrose metabolism, cyamino acid metabolism, galactose metabolism, diterpenoid metabolism, MAPK signaling pathway, steroid biosynthesis, linoleic acid metabolism, RNA polymerase, carbon fixation in photosynthetic organisms, flavone and flavonol biosynthesis.

Biosynthesis of secondary metabolites pathways like flavonoid biosynthesis and phenylpropanoid biosynthesis were enriched in the present study. Exposure of wheat to salinity stress has been shown to enhance phenylpropanoid biosynthesis (Cuong et al., 2020). When phenylpropanoid pathways are enhanced during stress they result in the production of antioxidative phenolic compounds (Dobariya et al., 2020; Sharma et al., 2019). Phenylpropanoids scavenge ROS and also protect the plant from excessive visible light using anthocyanins and UV such as flavonoids (Sharma et al., 2019). Various hormone signaling pathways were induced due to salinity stress including ABA, ethylene, brassinosteroid (BR), gibberelins and salicylic acid. The ABA is known to help plants adjust to the reduced water levels by closing the stomata and inducing the accumulation of various proteins that act as osmoprotectants (Ryu & Cho, 2015). Ethylene has been recently linked to salinity tolerance by enhancing ROS scavenging (Wang & Huang, 2019). The BR have been shown to reduce stress responses and improve homeostasis related to K^+/Na^+ ratios (de Oliveira et al., 2019; Ryu & Cho, 2015). Exogenous application of BR has been proven to alleviate salinity stress in Eucalyptus (de Oliveira et al., 2019).

A total of 755 TF were identified in differentially expressed unigenes. The MYB and MYB related were the most common TFs, followed by NAC, bHLH and ERF. Transcription factors are considered the most important switches that up regulate or down regulate gene expression (Gupta & Huang, 2014). The MYB TF have been shown to play a key role in salt stress (Ambawat et al., 2013). The TF are considered the most important gene expression regulators (Gupta & Huang, 2014). Gene expression is largely affected by chromatin structure that might be regulated epigenetically (Das, 2013). Epigenetic processes like methylation, acetylation and phosphorylation have been shown to have a modulating effect on gene specific salt response expression. It has been demonstrated that salt stress and salinity resistance phenotypes in plants are related to over methylation (Al-Harrasi et al., 2018). Processes like ubiquitination which has been demonstrated in this study is an example of posttranslational modification (Das, 2013). It is also possible to epigenetically transfer past salt stress responses to future generations (Das, 2013). This study not only identified DEGs due to salinity stress in doum palm but also provides numerous

genes in this plant which are now available and can be harnessed for various uses. This work also provides valuable understanding of the genetic salt tolerance mechanisms in this plant.

5.5 Development and validation of SSR markers

A total of 16,632 novel SSRs were identified at a frequency of 18% of the unigenes in the present study. This is comparable to 16% obtained in *P. dactylifera* (Zhao et al., 2013). The frequency of SSR was 1 per 5.5kb which is lower (1 per 2.4kb) than that of date palm and higher (1 per 9.8kb) than *V. unguiculata* (Gupta & Gopalakrishna, 2010; Zhao et al., 2013). This frequency varies based on the method used for SSR mining and the size of the data analyzed (Gupta & Gopalakrishna, 2010; Zhao et al., 2013). The abundance of the SSRs also varies depending on the plant species. The most frequent SSR motifs were dinucleotide (44.7%). The CT/AG dinucleotide SSRs were the most frequent in this study. Similar results were previously reported in date palm (Zhao et al., 2013). The AG motifs have been demonstrated to be rich dinucleotide marker motifs in *V.unguiculata* and *V. angularis* (Chen et al., 2015; Gupta & Gopalakrishna, 2010). Other studies have observed high trinucleotide repeats in cowpea and date palm (Gupta & Gopalakrishna, 2010; Zhao et al., 2013). The least SSR type was pentanucleotides accounting for only 0.9% of total SSRs compared to hexanucleotides (1.6%). Hexanucleotide motifs have been shown to be more frequent than pentanucleotides in several plants (Huang et al., 2014; Wu et al., 2014). Studies have also demonstrated that hexanucleotides and trinucleotides are abundant in coding regions (Qi et al., 2020; Wu et al., 2014). This might be the product of selection and evolution since tri- and hexa-nucleotide SSRs are assumed to be necessary in the coding region to maintain the reading frame, whereas di-, tetra-, and penta-nucleotide SSRs are easily able to shift the reading frame within coding sections and result in a negative mutation when SSR length variation happens (Wu et al., 2014). Variations in the repeat units of SSRs in coding areas can impact gene expression or function, such as gene transcription and/or translation, inactivate or activate genes, truncate proteins, silence genes and code for different or shorter proteins due to frameshift mutations (Filiz et al., 2012; Qi et al., 2018).

The EST-SSR have been previously developed and validated in many plants due to their transferability to related taxa and the ease of obtaining them from expression data (Chen et al., 2015). In the present study, 20 SSR markers were used to genotype 20 *H. compressa accessions* from four different ASAL regions of Kenya. Two primer pairs (DPSSR008 and DPSSR010) did not amplify any product. The 18 doum palm SSR markers amplified a total of 55 alleles at an average of 2.75 alleles per locus. This is comparable to EST-SSR markers from cowpea which yielded an average of 3.9 alleles per locus (Gupta & Gopalakrishna, 2010). The number of effective alleles (*Ne*) varied from 1 to 2.7 with a mean of 1.65. Observed heterozygosity (H_o) ranged from 0.0 to 0.35 with a mean of 0.12. This is slightly lower than the H_o obtained from SNP markers of 0.4 obtained previously in this study. High gene diversity (H_e) was obtained for the accessions using the SSR markers which ranged from 0.125 to 0.58 with a mean of 0.32. This genetic diversity is similar to the one obtained using SNP markers (*He* = 0.31). This indicates that SSR markers and GBS analysis produced comparable gene diversity estimates. Higher gene diversity compared to observed heterozygosity implies a lack of heterozygotes (Bernard et al., 2018).

F statistics like F_{IS}, F_{IT} and F_{ST} for the entire population was 0.63, 0.72 and 0.28 respectively. A high positive F_{IS} was obtained using SSR markers compared to F_{IS} of – 0.04 obtained using SNP markers. Therefore, SSR markers indicated deficiency of heterozygotes while SNP markers showed excess of heterozygotes. This could be due to various loci being subjected to various types and degrees of selection pressure, as well as other evolutionary forces (Tsehay et al., 2020). A lower F_{ST} of 0.074 was obtained using SNP markers compared to the F_{ST} obtained from SSR markers. These results show that mean values of genetic diversity indices differed between SSR markers and SNP markers with SSR markers having higher values. This phenomenon has been previously reported (García et al., 2018). The SNP markers have been shown to be more reliable for inferring diversity than SSR markers (Fischer et al., 2017). Further, SNPs are recommended for demographic inferences in non-model plants such as *H. compressa* (García et al., 2018). The differences in genetic diversity indices between SSRs and SNPs could also be attributed to the few number of SSR markers used for validation in this study. However, the resolving power of both markers is not

dependent on the number of SNPs and alleles but on the features of the markers themselves (Chen et al., 2017). Turkana accessions had the highest number of private alleles (0.17) while Tana River and Kwale had similar high percentages of polymorphic loci (77.8%). SNP markers also revealed high polymorphism in Kwale compared to the other regions.

In the present study, the primer DPSSR07 had the highest Shannon information index (1.03) and the highest PIC value (0.795). When a primer has a higher Shannon Index value than the others, it is suitable for investigating the genetic diversity in that population (Nassiry et al., 2009). Additionally, markers with a PIC value greater than 0.5 indicate that the marker is highly informative (Dalimunthe et al., 2020).

In the current study, longer repeats were not always associated with a higher number of alleles for example, the SSR marker DPSS07 with the repeat ATG 3*5 was highly polymorphic while the SSR marker DPSSRO20 with the repeat AACAG 5*4 was monomorphic. This phenomenon has been observed in cowpeas (Gupta & Gopalakrishna, 2010). In the present study, seven SSR markers; DPSSR07, DPSSR013, DPSSR014, DPSSR015, DPSSR016, DPSSR018 and DPSSR020 were sufficient to distinguish accessions from the four populations. Three SSR markers; DPSSR014, DPSSR018 and DPSSR020 detected private alleles in Turkana accessions. The DPSSRO16 SSR marker detected private alleles in Tana River and Kwale accessions, DPSSR07 and DPSSR015 detected private alleles in Tharaka accessions while DPSSR014 revealed private alleles in Tana River and Turkana accessions. This confirms the presence of private alleles in the populations of *H. compressa.* Previous studies have indicated that the presence of private alleles are indicators of gene flow and provide information on unique genetic patterns in certain loci (Eltaher et al., 2018). For example, in this study, private alleles were noted in Tana River and Kwale accessions using marker DPSSR016. This lends credence to the hypothesis of gene flow along the Tana River basin. SNP data revealed limited gene flow between Turkana and accessions along the Tana River basin. However, marker DPSSR014 revealed private alleles in Tana River and Turkana accessions. This may be explained by the multiallelic nature of SSR markers as opposed to SNP

markers. The unigene SSR markers developed in this study are thus highly polymorphic and informative. They can therefore be used to infer genetic diversity and MAS and population/variety identification of *H. compressa*. The transferability rate of 10% to coconut was inferred. The transferability of unigene SSR markers has also been demonstrated in *Vigna unguiculata* (Gupta & Gopalakrishna, 2010). The SSR markers obtained in this study could therefore be useful for other non-model palm species with scanty genetic information. Particularly palms in the genus *Hyphaene*.

The SSR markers grouped the 20 accessions into 3 clusters. Cluster 1 consisted of accessions from Tharaka and one Tana River sample, Cluster 2 had accessions from Turkana while Tana River and Kwale were admixed and polymorphic populations. This is also consistent with the morphological and SNP data analyses earlier obtained in this study. The SNP data clustered *H. compressa* accessions into two clusters with Turkana forming a lone cluster while populations from Kwale, Tana River and Tharaka forming another cluster. STRUCTURE results obtained from GBS analysis also revealed that Kwale accessions had a lot of admixture between Cluster 1 and Cluster 2 which could explain why some Kwale accessions clustered with Turkana accessions. Possibly, some of the accessions found in Turkana are also found in Kwale. In this study, Kwale respondents were able to distinguish three different varieties of doum palm. One variety could be the one found in Turkana as well. Furthermore, separation plots (Linkage disequilibrium) generated from doum palm morphological data and cluster analysis (morphotype 3) confirm that some accessions from Kwale are morphologically similar to some Turkana accessions. However, as previously stated, SNP and SSR markers have shown differences in genetic diversity estimates and population structure analyses in several other studies (Chen et al., 2017; Fischer et al., 2017; García et al., 2018). Some studies have alluded that SSRs are more efficient at determining the diversity studies while SNPs are better at population structure inferences (Singh et al., 2013). In contrast, other studies indicate that SSRs have higher resolution for population structure analyses than SNPs (Chen et al., 2017). In other studies, SNP markers have been shown to reflect ancient population divergence over long distances (Tsykun et al., 2017).

The SSR population structure results were also corroborated with the SSR cluster analysis and PCoA in this study. The SSR markers produced consistent results with the GBS data. The SSR markers developed in the present study have therefore proven to be reliable markers for *Hyphaene* species.

This study successfully demonstrated how transcriptome analysis is useful for marker development especially in non-model plants like *H. compressa*. To date no SSR markers had been described for *H. compressa*. These new set of SSR markers are suitable for genetic diversity estimates, population assignment studies and marker assisted selection. The novel *H. compressa* markers identified in this study in addition to enriching the genetic resources for the entire *Hyphaene* genus and related taxa will enhance diversity studies of doum palm.

5.6 Conclusions

The present study successfully evaluated the ethnobotany, diversity and salinity induced transcriptomics in doum palm. Based on the findings from this study, the following conclusions are drawn.

1. *H. compressa* is semi-domesticated in Tharaka Nithi County where farmers intercrop, prune and maintain it in their farms whereas in Turkana it is still abundant and naturally occurring in the wild. A total of 14 uses of doum palm are described in this study. Among the uses, food utility was the most important in all the regions sampled albeit there was no statistical significance in food utility between the regions sampled (P=0.568). Human intervention, salinity and drought are the main pressures on *H. compressa* accessions in the ASAL regions of Kenya. Therefore, the stated null hypothesis on objective one is rejected.
2. Seven quantitative and nine qualitative descriptors were used for morphological diversity of doum palm which effectively identified five morphotypes of *H. compressa.* Therefore, the stated null hypothesis on objective 2 is rejected. These morphotypes exhibited different superior traits for fruits, leaves and petioles that can be used for future breeding of *H.*

compressa. The morphological assessment also showed that Kwale accessions are the most diverse with some of the accessions from Kwale forming a lone cluster (Morphotype 4).

3. Based on population STRUCTURE analysis, the *H. compressa* accessions belong to two gene pools, one contains accessions from northern Kenya and the other contains accessions from the River Tana Basin. The *H. compressa* accessions in Cluster 2 (those found along the River Tana Basin) are the most diverse. Genotyping by sequencing analysis revealed that *H. compressa* accessions are interconnected with high gene flow and moderate genetic differentiation, evidenced by high within-population variation than among population variation. The findings of this objective support the rejection of the stated null hypothesis on objective 3.
4. Based on greenhouse salinity assays, *H. compressa* was proven to be moderately to highly tolerant to salinity stress with Tana River accessions being the most resistant to salinity stress, followed by Turkana and finally, Tharaka accessions. A total of 8611 DEGs were obtained with 3722 genes being upregulated and 4889 down regulated due to salinity stress. The most enriched pathways were flavonoid biosynthesis, photosystem, photosynthesis and synthesis of secondary metabolites among others and therefore the stated null hypothesis on objective 4 is rejected.
5. A total of 16,632 SSRs markers were identified from the transcriptomics work. Validation of 20 of these markers confirmed that 18 of them are polymorphic. The study identified seven out of the 20 SSR markers; DPSSR07, DPSSR013, DPSSR014, DPSSR015, DPSSR016, DPSSR018 and DPSSR020 that are able to distinguish accessions from the four populations. Two of the polymorphic SSR markers identified in this study showed cross genus transferability to *C. nucifera.* The stated null hypothesis on objective 5 is therefore rejected.

This is the first study to my knowledge to:

i) Genotype *H. compressa* using the GBS approach
ii) Determine salinity induced DEGs in doum palm

iii) Provide a set of functional EST-SSR markers for the genus *Hyphaene* derived from *H. compressa* transcriptome

5.7 Recommendations

Based on the findings of this study, the following recommendations are made;

1. Various *H. compressa* morphotypes demonstrated superior features in fruits, leaves and petioles, allowing superior accessions to be chosen for domestication and genetic improvement. Tharaka Nithi accessions are excellent candidates for breeding for fruit traits because of their superior fruit traits. Kwale accessions on the other hand especially the mkoko variety (short accession) should be selected because of their long petioles which are important for ease of harvesting and use for various construction purposes.
2. The two newly described genetic clusters be adopted for conservation purposes in order to preserve the high level of population variability. *Ex situ* conservation techniques can be used to achieve this, especially for cluster 2 which had the most diversity.
3. Breeders can now develop salt tolerance crops based on the improved understanding of salt tolerance mechanisms in this plant.
4. This study demonstrated that doum palm has economic uses in Kenya despite having little species/ varietal information. Consolidated *in situ* and *ex situ* conservation efforts are required due to the plant's increasing exploitation, particularly in Tharaka Nithi County.
5. More domestication efforts are advocated for especially in Tharaka and other ASAL regions
6. This study validated 20 SSR markers and recommends validation of the rest of the markers on *H. compressa*, other species of the *Hyphaene* genus and other palm species.
7. Future studies should evaluate additional descriptors so that the male doum palm diversity can also be determined

REFERENCES

Abdullah, Khan, S. M., Pieroni, A., Haq, Z. U., & Ahmad, Z. (2020). Mazri (Nannorrhops ritchiana (Griff) Aitch.): A remarkable source of manufacturing traditional handicrafts, goods and utensils in Pakistan. *Journal of Ethnobiology and Ethnomedicine*, *16*(1), 1–13.

Aboshora, W., Lianfu, Z., Dahir, M., Gasmalla, M. A. A., & Musa, A. (2014). *Physicochemical , Nutritional and Functional Properties of the Epicarp , Flesh and Pitted Sample of Doum Fruit (Hyphaene Thebaica)*. *2*(4), 180–186.

Aboshora, W., Lianfu, Z., Dahir, M., Gasmalla, M. A. A., Musa, A., Omer, E., & Thapa, M. (2014). Physicochemical , Nutritional and Functional Properties of the Epicarp , Flesh and Pitted Sample of Doum Fruit (Hyphaene Thebaica). *Journal of Food and Nutrition Research*, *2*(4), 180–186.

Abraham, G., Pandey, N., Mishra, V., Chaudhary, A. A., Ahmad, A., Singh, R., & Singh, P. K. (2013). Development of SCAR based molecular markers for identification of different species of Azolla. *Indian Journal of Biotechnology*, *12*(4), 489–492.

Akall, G. (2021). Effects of development interventions on pastoral livelihoods in Turkana County, Kenya. *Pastoralism*, *11*(1).

Al-Abdoulhadi, I. A., Dinar, H. A., Ebert, G., & Büttner, C. (2011). Effect of salinity on leaf growth, leaf injury and biomass production in date palm (Phoenix dactylifera L.) cultivars. *Indian Journal of Science and Technology*, *4*(11), 1542–1546.

Al-dous, E. K., George, B., Al-mahmoud, M. E., Al-jaber, M. Y., Wang, H., Salameh, Y. M., Al-azwani, E. K., Chaluvadi, S., Pontaroli, A. C., Debarry, J., Arondel, V., Ohlrogge, J., Saie, I. J., Suliman-elmeer, K. M., Bennetzen, J. L., Kruegger, R. R., & Malek, J. A. (2011). *De_novo* genome sequencing and comparative genomics of date palm (Phoenix dactylifera). *Nature Biotechnology*, *29*(6), 521–527.

Al-Harrasi, I., Al-Yahyai, R., & Yaish, M. W. (2018). Differential DNA methylation and transcription profiles in date palm roots exposed to salinity. *PLoS ONE*, *13*(1), 1–29.

Al-Mahmoud, M. E., Al-Dous, E. K., Al-Azwani, E. K., & Malek, J. A. (2012). DNA-based assays to distinguish date palm (arecaceae) gender. *American Journal of Botany*, *99*(1), 7–10.

Al-Qurainy, F., Al-Ameri, A. A., Khan, S., Nadeem, M., Gaafar, A. R. Z., & Tarroum, M. (2018). SCAR marker for gender identification in date palm (Phoenix dactylifera L.) at the seedling stage. *International Journal of Genomics*, *2018*.

Al-Saharin, R., Hellmann, H., & Mooney, S. (2022). Plant E3 Ligases and Their Role in Abiotic Stress Response. *Cells*, *11*(5), 1–29.

Alabdallat, N. G., & Bilto, Y. (2015). Antioxidant Properties of Selected Wild Plants. *International Journal of Science and Research (IJSR)*, *4*(2), 197–200.

Alemu, A., Feyissa, T., Letta, T., & Abeyo, B. (2020). Genetic diversity and population structure analysis based on the high density SNP markers in Ethiopian durum wheat (Triticum turgidum ssp. durum). *BMC Genetics*, *21*(1), 1–12.

Alhasnawi, A. N. (2019). Role of proline in plant stress tolerance: A mini review. *Research on Crops*, *20*(1), 223–229.

Ali, Y. S. S. (2007). Interaction of salinity and temperature on the germination of doum (Hyphaene thebaica) seed in Saudi Arabia. *Asian Journal of Plant Science*, *6*(6), 962–966.

Alipour, H., Bihamta, M. R., Mohammadi, V., Peyghambari, S. A., Bai, G., & Zhang, G. (2017). Genotyping-by-sequencing (GBS) revealed molecular genetic diversity of Iranian wheat landraces and cultivars. *Frontiers in Plant Science*, *8*(1293), 1–14.

Altaf, R., Bhatti, K. H., Mirza, S. A., Ajaib, M., & Ishtiaq, M. (2019). Ethnomedicinal

study of Tehsil Wazirabad Gujranwala Punjab Pakistan. *Pakistan Journal of Science*, *71*(4), 260–271.

Ambawat, S., Sharma, P., Yadav, N. R., & Yadav, R. C. (2013). MYB transcription factor genes as regulators for plant responses: An overview. *Physiology and Molecular Biology of Plants*, *19*(3), 307–321.

Amirbakhtiar, N., Ismaili, A., Ghaffari, M. R., Firouzabadi, F. N., & Shobbar, Z. S. (2019). Transcriptome response of roots to salt stress in a salinity-tolerant bread wheat cultivar. *PLoS ONE*, *14*(3).

Amwatta, C. J. M. (2004). Diversity of Use of Doum (Hyphaene compressa) in Kenya. *Palms*, *48*(4), 184–190.

Andersen, J. R., & Lubberstedt, T. (2003). Functional markers in plants. *Trends in Plant Science*, *8*(11), 554–560.

Angaine, M. P. (2005). *The Biology and Ecology of doum palm (Hyphaene compressa H. Wendl.): A Case study of Meru National Park*. Addis Ababa University.

Arbeiter, A. B., Hladnik, M., Jakše, J., & Bandelj, D. (2017). Identification and validation of novel EST-SSR markers in olives. *Scientia Agricola*, *74*(3), 215–225.

Asmussen, C. B., & Chase, M. W. (2001). Coding and Non coding Plastid DNA in Palm Systematics. *American Journal of Botany*, *88*(6), 1103–1117.

Asmussen, C. B., Dransfield, J., Deickmann, V., Barfod, A. S., Pintaud, J., & Baker, W. J. (2006). A new subfamily classification of the palm family (Arecaceae): evidence from plastid DNA phylogeny. *Botanical Journal of the Linnean Society*, *151*, 15–38.

Assessment, Kenya InterAgency rapid. (2014). *Kwale Secondary Data Review* (Vol. 1, Issue December).

Baba, S. A., Vishwakarma, R. A., & Ashraf, N. (2017). Functional characterization of

CsBGlu12, a β-glucosidase from crocus sativus, provides insights into its role in abiotic stress through accumulation of antioxidant flavonols. *Journal of Biological Chemistry*, *292*(11), 4700–4713.

Babitseng, T. M., & Teketay, D. (2013). Impact of wine tapping on the population structure and regeneration of Hyphaene petersiana Klotzsch ex Mart. In northern Botswana. *Ethnobotany Research and Applications*, *11*, 9–28.

Babu, B. K., Mathur, R. K., Ravichandran, G., Anitha, P., & Venu, M. V. B. (2019). Genome-wide association study for leaf area, rachis length and total dry weight in oil palm (Eleaeisguineensis) using genotyping by sequencing. *PLoS ONE*, *14*(8), 1–10.

Bai, B., Wang, L., Zhang, Y. J., Lee, M., Rahmadsyah, R., Alfiko, Y., Ye, B. Q., Purwantomo, S., Suwanto, A., Chua, N.-H., & Yue, G. H. (2018). Developing genome-wide SNPs and constructing an ultrahigh-density linkage map in oil palm. *Scientific Reports*, *8*(691), 1–7.

Barcelos, E., Amblard, P., Berthaud, J., & Seguin, M. (2002). Genetic diversity and relationship in American and African oil palm as revealed by RFLP and AFLP molecular markers. *Pesquisa Agropecuaria Brasileira*, *37*(8), 1105–1114.

Barrow, E., & Mogaka, H. (2007). *Kenya's Drylands - Wastelands or an undervalued National Economic Resource*. https://www.iucn.org/sites/dev/files/import/downloads/kenya_dryland_value_2007.pdf

Bates, L., Waldren, R., & Teare, I. (1973). Rapid determination of free proline for water-stress studies. *Plant and Soil*, *39*, 205–207.

Bayão, G. F. V., Cunha, C. S., de Queiroz, A. C., Pimentel, R. M., Cardoso, L. L., da Silva, T. É., Cardoso, A. J. da S., Sousa, K. R. S., & Marcondes, M. I. (2020). Heart-of-palm byproduct for lactating cows. *Journal of Applied Animal Research*, *48*(1), 1–6.

Bazzo, B. R., de Carvalho, L. M., Carazzolle, M. F., Pereira, G. A. G., & Colombo, C. A. (2018). Development of novel EST-SSR markers in the macaúba palm (Acrocomia aculeata) using transcriptome sequencing and cross-species transferability in Arecaceae species. *BMC Plant Biology*, *18*(1), 1–10.

Bègue, H., Besson-Bard, A., Blanchard, C., Winckler, P., Bourque, S., Nicolas, V., Wendehenne, D., & Rosnoblet, C. (2019). The chaperone-like protein CDC48 regulates ascorbate peroxidase in tobacco. *Journal of Experimental Botany*, *70*(10), 2665–2681.

Beier, S., Thiel, T., Münch, T., Scholz, U., & Mascher, M. (2017). MISA-web: A web server for microsatellite prediction. *Bioinformatics*, *33*(16), 2583–2585.

Bernard, A., Barreneche, T., Lheureux, F., & Dirlewanger, E. (2018). Analysis of genetic diversity and structure in a worldwide walnut (Juglans regia L.) germplasm using SSR markers. *PLoS ONE*, *13*(11), 1–19.

Bezona, N., Hensley, D., Yogi, J., Tavares, J., Rauch, F., Iwata, R., Kellison, M., & Wong, M. (2009). Salt and Wind Tolerance of Landscape Plants for Hawaii. *Landscape*, *L-13 (revi*(March), 1–9.

BGCI. (2020). *Botanic Gardens Conservation International. Richmond, U.K.*

Bhattarai, K. R. (2018). Ethnobotanical study of plants used by Thami community in Ilam District, eastern Nepal. *Our Nature*, *16*(1), 55–67.

Blake, C., & Munns, R. (2017). Screening for Salt Tolerance Using Supported Hydroponics. In *PromethiusWiki, Protocolsa in ecological and environmental Sciences*.

Bohling, J. (2020). Evaluating the effect of reference genome divergence on the analysis of empirical RADseq datasets. *Ecology and Evolution*, *10*(May), 7585–7601.

Borghesi, E., González-Miret, M. L., Escudero-Gilete, M. L., Malorgio, F., Heredia,

F. J., & Meléndez-Martínez, A. J. (2011). Effects of salinity stress on carotenoids, anthocyanins, and color of diverse tomato genotypes. *Journal of Agricultural and Food Chemistry*, *59*(21), 11676–11682.

Boscaiu, M., Donat, P., Llinares, J., & Vicente, O. (2012). Stress-tolerant wild plants: A source of knowledge and biotechnological tools for the genetic improvement of stress tolerance in crop plants. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, *40*(2), 323–327.

Brandies, P., Peel, E., Hogg, C. J., & Belov, K. (2019). The Value of Reference Genomes in the Conservation of Threatened Species. *Genes*, *10*(846).

Brautigam, A., & Gowik, U. (2010). What can next generation sequencing do for you ? Next generation sequencing as a valuable tool in plant research. *Plant Biology*, *12*, 831–841.

Buchfink, B., Reuter, K., & Drost, H. G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, *18*(4), 366–368.

Bui, E. . (2013). Soil salinity : A neglected factor in plant ecology and biogeography. *Journal of Arid Environments*, *92*, 14–25.

Burghardt, L. T., Young, N. D., & Tiffin, P. (2017). A Guide to Genome-Wide Association Mapping in Plants. *Curr. Protoc. Plant Biol*, *2*, 22–38.

Bvenura, C., & Sivakumar, D. (2017). The role of wild fruits and vegetables in delivering a balanced and healthy diet. *Food Research International*, *99*, 15–30.

Caffarri, S., Tibiletti, T., Jennings, R. C., & Santabarbara, S. (2014). A Comparison Between Plant Photosystem I and Photosystem II Architec-ture and Functioning. *Current Protein and Peptide Science*, *15*, 296–331.

Cai, Z. Q., & Gao, Q. (2020). Comparative physiological and biochemical mechanisms of salt tolerance in five contrasting highland quinoa cultivars. *BMC Plant Biology*, *20*(70), 1–15.

Chakravarty, S., Bhutia, K. D., Suresh, C. P., Shukla, G., & Pala, N. A. (2016). A review on diversity, conservation and nutrition of wild edible fruits. *Journal of Applied and Natural Science*, *8*(4), 2346–2353.

Charles, R. (1992). Domesticated Palms. *Principes*, *36*(2), 70–78.

Charlesworth, D. (2006). Evolution of Plant Breeding Systems. *Current Biology*, *16*(17), 726–735.

Chase, M. W., & Hills, H. H. (1991). Silica Gel : An Ideal Material for Field Preservation of Leaf Samples for DNA Studies Published by : International Association for Plant Taxonomy (IAPT) Stable URL : http://www.jstor.org/stable/1222975 REFERENCES Linked references are available on JSTOR. *JSTOR*, *40*(2), 215–220.

Chaves, M. M., Flexas, J., & Pinheiro, C. (2009). Photosynthesis under drought and salt stress: Regulation mechanisms from whole plant to cell. *Annals of Botany*, *103*(4), 551–560.

Cheikhyoussef, A., & Embashu, W. (2013). Ethnobotanical knowledge on indigenous fruits in Ohangwena and Oshikoto regions in Northern Namibia. *Journal of Ethnobiology and Ethnomedicine*, *9*(1).

Chemelil, P. (2016). The Contribution of Voluntary, Faith Based and Non Governmental Organizations To Socio-Economic Development in Turkana County, Kenya 1986-2006. *European Journal of Social Sciences Studies*, *3*(3), 200–220.

Chen, H., Liu, L., Wang, L., Wang, S., Somta, P., & Cheng, X. (2015). Development and validation of EST-SSR markers from the transcriptome of adzuki bean (Vigna angularis). *PLoS ONE*, *10*(7), 1–14.

Chen, W., Hou, L., Zhang, Z., Pang, X., & Li, Y. (2017). Genetic diversity, population structure, and linkage disequilibrium of a core collection of ziziphus jujuba assessed with genome-wide snps developed by genotyping-by-sequencing and

SSR Markers. *Frontiers in Plant Science*, *8*(April), 1–14.

Clement, C. R., de Cristo-Araújo, M., D'Eeckenbrugge, C., dos Reis, V. M., Lehnebach, R., & Picanço-Rodrigues, D. (2017). Origin and dispersal of domesticated peach palm. *Frontiers in Ecology and Evolution*, *5*, 1–19.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*(18), 3674–3676.

Cortés, A. J., & Blair, M. W. (2018). Genotyping by sequencing and genome–environment associations in wild common bean predict widespread divergent adaptation to drought. *Frontiers in Plant Science*, *9*(February), 1–13.

Cosiaux, A., Gardiner, L. ., & Couvreur, T. L. . (2017). Hyphaene compressa, The IUCN Red List of Threatened Species. *IUCN Red List*.

Cros, D., Bocs, S., Riou, V., Ortega-Abboud, E., Tisné, S., Argout, X., Pomiès, V., Nodichao, L., Lubis, Z., Cochard, B., & Durand-Gasselin, T. (2017). Genomic preselection with genotyping-by-sequencing increases performance of commercial oil palm hybrid crosses. *BMC Genomics*, *18*(1), 1–17.

Cuong, D. M., Kwon, S. J., Van Nguyen, B., Chun, S. W., Kim, J. K., & Park, S. U. (2020). Effect of salinity stress on phenylpropanoid genes expression and related gene expression in wheat sprout. *Agronomy*, *10*(3).

D′Agostino, N. D., Taranto, F., Camposeo, S., Mangini, G., Fanelli, V., Gadaleta, S., Miazzi, M. M., Pavan, S., Rienzo, V., Sabetta, W., Lombardo, L., Zelasco, S., Perri, E., Lotti, C., Ciani, E., & Montemurro, C. (2018). GBS-derived SNP catalogue unveiled wide genetic variability and geographical relationships of Italian olive cultivars. *Scientific Reports*, *10*(October), 1–13.

da Paschoa, R. P., Christ, J. A., Valente, C. S., Ferreira, M. F. da S., de Miranda, F. D., Garbin, M. L., & Carrijo, T. T. (2018). Genetic diversity of populations of the dioecious myrsine coriacea (Primulaceae) in the atlantic forest. *Acta Botanica*

Brasilica, *32*.

Dalimunthe, S. R., Siregar, L. A. M., Putri, L. A. P., Chairunnisa, T., & Hairmansis, A. (2020). Polymorphism levels of some SSR markers (Simple Sequence Repeat) for parental line identification on low temperature tolerance. *IOP Conference Series: Earth and Environmental Science*, *454*(1).

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.

Das, B. A. (2013). Bioprospecting and Genetic Engineering of Mangrove Genes to Enhance Salinity Tolerance in Crop plants. In S. . Jain & G. D.S (Eds.), *Biotechnology of Neglected and Underutilized Crops* (pp. 385–455). Springer.

Das, K., & Roychoudhury, A. (2014). Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants. *Frontiers in Environmental Science*, *2*(DEC), 1–13.

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, *12*(7), 499–510.

de Oliveira, V. P., Lima, M. D. R., da Silva, B. R. S., Batista, B. L., & da Silva Lobato, A. K. (2019). Brassinosteroids Confer Tolerance to Salt Stress in Eucalyptus urophylla Plants Enhancing Homeostasis, Antioxidant Metabolism and Leaf Anatomy. *Journal of Plant Growth Regulation*, *38*(2), 557–573.

DeMotts, R. B. (2017). Weaving a living: Gender, craft, and sustainable resource use in Botswana. *Journal of Political Ecology*, *24*(1), 368–385.

Deyholos, M. K. (2010). Making the most of drought and salinity transcriptomics. *Plant, Cell and Environment*, *33*(4), 648–654.

Dhawan, C., Kharb, P., Sharma, R., Uppal, S., & Aggarwal, R. K. (2013). Development of male-specific SCAR marker in date palm (Phoenix dactylifera L.). *Tree Genetics and Genomes*, *9*(5), 1143–1150.

do Amaral, M. N., Arge, L. W. P., Benitez, L. C., Danielowski, R., Silveira, S. F. da S., Farias, D. da R., de Oliveira, A. C., da Maia, L. C., & Braga, E. J. B. (2016). Comparative transcriptomics of rice plants under cold, iron, and salt stresses. *Functional and Integrative Genomics*, *16*(5), 567–579.

Dobariya, B. R., Singh, D., Suthar, K. P., & Parekh, V. (2020). Phenylpropanoid pathway leads to salinity stress tolerance in castor (Ricinus Phenylpropanoid Pathway Leads to Salinity Stress Tolerance in Castor (Ricinus communis L .). *International Journal of Tropical Agriculture*, *December*.

Dubouzet, J. G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E. G., Miura, S., Seki, M., Shinozaki, K., & Yamaguchi Shinozaki, K. (2003). OsDREB genes in rice, Oryza sativa L., encode transcription activators that function in drought , high salt and cold responsive gene expression. *The Plant Journal*, *33*(4), 751–763.

Durand, J., Bodénès, C., Chancerel, E., Frigerio, J., Vendramin, G., Sebastiani, F., Buonamici, A., Gailing, O., Koelewijn, H., Villani, F., Mattioni, C., Cherubini, M., Goicoechea, P. G., Herrán, A., Ikaran, Z., Cabané, C., Ueno, S., Alberto, F., Dumoulin, P., … Plomion, C. (2010). A fast and cost-effective approach to develop and map EST-SSR markers : oak as a case study. *BMC Genomics*, *11*(570), 1–13.

Earl, D. A., & VonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*, 359–361.

Eaton, D. A. R., & Overcast, I. (2020). Ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, *36*(8), 2592–2594.

Egan, A. N., Schlueter, J., & Spooner, D. M. (2012). Applications Of Next -

Generation Sequencing In Plant Biology. *American Journal of Botany*, *99*(2), 175–185.

Ejore, E., Ongugo, R., Kemboi, J., Ojunga, S., Mwenja, P., & Owino, J. (2020). Plant species and their importance to housing in the Turkana community, Kenya. *Journal of Horticulture and Forestry*, *12*(3), 101–108.

Ekesa, B. (2009). Accesibility To and Consumption of Indigenous Vegetables and Fruits by Rural Households in Matungu Division, Western Kenya. *African Journal of Food Agriculture Nutrition and Development*, *9*(8).

El-Beltagi, H., Mohamed, H., Yousef, H., & Fawz, E. (2018). Biological Activities of the Doum Palm (Hyphaene thebaica L .) and Its Its Bioactive Components. In E. Shalaby & G. M. Azzam (Eds.), *Antioxidants in food and its applications* (pp. 49–65). Intechopen.

El-Rashad, A., & Hassan, Z. M. . (2005). Potential Utilization And Healthy Effects Of Doum Fruits In Ice Cream And Sesame Butter (TEHENA). *Alex.Journal of Food Science and Technology*, *2*(1), 29–39.

Ellis, J. R., & Burke, J. M. (2007). EST-SSRs as a resource for population genetic analyses. *Heredity*, *99*(2), 125–132.

Elmeer, K., & Mattat, I. (2012). Marker-assisted sex differentiation in date palm using simple sequence repeats. *3 Biotech*, *2*(241), 241–247.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust , Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *Plos One*, *6*(5), 1–10.

Eltaher, S., Sallam, A., Belamkar, V., Emara, H. A., Nower, A. A., Salem, K. F. M., Poland, J., & Baenziger, P. S. (2018). Genetic diversity and population structure of F3:6 Nebraska Winter wheat genotypes using genotyping-by-sequencing. *Frontiers in Genetics*, *9*(MAR), 1–9.

Escobar, S., Helmstetter, A. J., Montufar, R., Couvreur, T. L. ., & Balslev, H. (2022). Phylogenomic relationships and historical biogeography in the South American vegetable ivory palms (Phytelepheae). *Molecular Phylogenetics and Evolution*, *21*(1), 1–9.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620.

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*, 564–567.

Fabritzek, A. G., Griebeler, E. M., & Kadereit, J. W. (2021). Hybridization, ecogeographical displacement and the emergence of new lineages – A genotyping-by-sequencing and ecological niche and species distribution modelling study of Sempervivum tectorum L. (Houseleek). *Journal of Evolutionary Biology*, *34*(5), 830–844.

Fathir, A., Haikal, M., & Wahyudi, D. (2021). Ethnobotanical study of medicinal plants used for maintaining stamina in madura ethnic, East Java, Indonesia. *Biodiversitas*, *22*(1), 386–392.

Feng, S., Zhu, Y., Yu, C., Jiao, K., Jiang, M., Lu, J., Shen, C., Ying, Q., & Wang, H. (2018). Development of species-specific SCAR markers, based on a SCoT analysis, to authenticate Physalis (Solanaceae) species. *Frontiers in Genetics*, *9*(MAY), 1–10.

Feng, W., Kita, D., Peaucelle, A., Cartwright, H. N., Doan, V., Duan, Q., Liu, M. C., Maman, J., Steinhorst, L., Schmitz-Thom, I., Yvon, R., Kudla, J., Wu, H. M., Cheung, A. Y., & Dinneny, J. R. (2018). The FERONIA Receptor Kinase Maintains Cell-Wall Integrity during Salt Stress through Ca2+Signaling. *Current Biology*, *28*(5), 1–10.

Fentahun, M. T., & Hager, H. (2009). Exploiting locally available resources for food and nutritional security enhancement: wild fruits diversity, potential and state of exploitation in the Amhara region of Ethiopia. *Food Security*, *1*(2), 207–219.

Feron, R., & Waterhouse, R. M. (2022). Assessing species coverage and assembly quality of rapidly accumulating sequenced genomes. *GigaScience*, *11*, 1–24.

Ferreira, T. M. M., Leão, A. P., de Sousa, C. A. F., & Souza Júnior, M. T. (2021). Genes highly overexpressed in salt-stressed young oil palm (Elaeis guineensis) plants. *Revista Brasileira de Engenharia Agricola e Ambiental*, *25*(12), 813–818.

Filiz, E., Koc, I., & Sakinoglu, F. C. (2012). In silico EST-SSRs Analysis in UniGene of Quercus robur L . *Research in Plant Biology*, *2*(6), 1–9.

Fischer, M. C., Rellstab, C., Leuzinger, M., Roumet, M., Gugerli, F., Shimizu, K. K., Holderegger, R., & Widmer, A. (2017). Estimating genomic diversity and population differentiation - an empirical comparison of microsatellite and SNP variation in Arabidopsis halleri. *BMC Genomics*, *18*(1), 1–15.

Francis, R. M. (2017). pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*(1), 27–32.

Gadgil, M., Berkes, F., & C, F. (1993). indigenous knowledge for biodiversity conservation. *Ambio*, *22*(3), 151–156.

Galla, S. J., Forsdick, N. J., Brown, L., Hoeppner, M. P., Knapp, M., Maloney, R. F., Moraga, R., Santure, A. W., & Steeves, T. E. (2019). Reference genomes from distantly related species can be used for discovery of single nucleotide polymorphisms to inform conservation management. *Genes*, *10*(9).

Gan, S. T., Wong, W. C., Wong, C. K., Soh, A. C., Kilian, A., Low, E. L., Massawe, F., & Mayes, S. (2018). High Density and DArT-based Genetic Linkage Maps of Two Closely Related Oil Palm Populations. *Journal of Applied Genetics*, *59*, 23–34.

García, C., Guichoux, E., & Hampe, A. (2018). A comparative analysis between SNPs and SSRs to investigate genetic variation in a juniper species (Juniperus phoenicea ssp. turbinata). *Tree Genetics and Genomes*, *14*(6).

Ghislain, M., Spooner, D. ., Rodriguez, F., Villamon, F., Nunez, J., Vasquez, C., Waugh, R., & Bonierbale, M. (2004). Selection of highly informative and user-friendly microsatellites (SSRs) for genotyping of cultivated potato. *Theor Appl Genet*, *108*, 881–890.

Govaerts, R., & Dransfield, J. (2005). *world checklist of palms*. Royal Botanical Garden.

Govindaraj, M., Vetriventhan, M., & Srinivasan, M. (2015). Importance of Genetic Diversity Assessment in Crop Plants and Its Recent Advances : An Overview of Its Analytical Perspectives. *Genetics Research International*, *2015*(1–14).

Grabherr, M. G. ., Brian J. Haas, Moran Yassour Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, Lin Fan, Raktima Raychowdhury, Qiandong Zeng, Zehua Chen, Evan Mauceli, Nir Hacohen, Andreas Gnirke, Nicholas Rhind, Federica di Palma, Bruce W., N., & Friedman, and A. R. (2013). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, *29*(7), 644–652.

Gros-Balthazard, M., & Flowers, J. M. (2021). A Brief History of the Origin of Domesticated Date Palms. In J. M. Al-Khayri, S. M. Jain, & D. . Johnson (Eds.), *The Date Palm Genome* (pp. 55–74). Compedium of Plant Genomes, Sringer.

Grover, A., & Sharma, P. C. (2016). Development and use of molecular markers : past and present. *Critical Reviews in Biotechnology*, *36*(2), 290–302.

Gruca, M., van Andel, T. R., & Balslev, H. (2014). Ritual uses of palms in traditional medicine in sub-Saharan Africa: A review. *Journal of Ethnobiology and Ethnomedicine*, *10*(60), 1–24.

Gupta, B., & Huang, B. (2014). Mechanism of salinity tolerance in

plants:Physiological, Biochemical and Molecular Characterisation. *International Journal of Genomics*, *2014*.

Gupta, S. K., & Gopalakrishna, T. (2010). Development of unigene-derived SSR markers in cowpea (Vigna unguiculata) and their transferability to other Vigna species. *Genome*, *53*(7), 508–523.

Haas, B. ., & Papanicolaou, A. (2019). *TransDecoder (Find Coding Regions Within Transcripts)*. https://github.com/TransDecoder/TransDecoder/wiki

Hahn, W. J. (2002). A molecular phylogenetic study of the palmae (Arecaceae) based on atpB, rbcL, and 18S nrDNA sequences. *Systematic Biology*, *51*(1), 92–112.

Haider, S. M., Khan, I. A., Jaskani, M., Naqvi, S. A., Hameed, M., Azam, M., Khan, A. A., & Pintaud, J. C. (2015). Assessment of morphological attributes of date palm accessions of diverse agro- ecological origin. *Pakistan Journal of Botany*, *47*(3), 1143–1151.

Hamdy, R., & Fahmy, A. G. (2018). Study of Plant Remains from the Embalming Cache KV63 at Luxor , Egypt : Progress in African Archaeobotany. In A. Mercuri, C. D'Andrea, F. Rita, & H. Alexa (Eds.), *Plants and people in the African past* (Issue February, pp. 40–56). Springer International Publishing.

Hassan Mansour. (2021). Population genetics of Hyphaene thebaica Mart., in Rabigh province, Saudi Arabia: Implications for conservation. *GSC Biological and Pharmaceutical Sciences*, *17*(3), 153–159.

Hayat, S., Hayat, Q., Alyemeni, M. N., Wani, A. S., Pichtel, J., & Ahmad, A. (2012). Role of proline under changing environments: A review. *Plant Signaling and Behavior*, *7*(11), 1456–1466.

Hazzouri, K. M., Gros-Balthazard, M., Flowers, J. M., Copetti, D., Lemansour, A., Lebrun, M., Masmoudi, K., Ferrand, S., Dhar, M. I., Fresquez, Z. A., Rosas, U., Zhang, J., Talag, J., Lee, S., Kudrna, D., Powell, R. F., Leitch, I. J., Krueger, R. R., Wing, R. A., … Purugganan, M. D. (2019). Genome-wide association

mapping of date palm fruit traits. *Nature Communications*, *10*, 1–14.

He, J., Zhao, X., Laroche, A., Lu, Z., Liu, H., & Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science*, *5*(484), 1–8.

Hernandez, A. J. (2019). Salinity Tolerance in Plants : Trends and Perspectives. *International Journal of Molecular Sciences*, *20*(2408).

Hoffman, B., & Gallaher, T. (2007). Importance indices in ethnobotany. *Ethnobotany Research and Applications*, *5*, 201–218.

Horn, J. W., Fisher, J. B., Barry Tomlinson, P., Lewis, C. E., & Laubengayer, K. (2009). Evolution of lamina anatomy in the palm family (Arecaceae). *American Journal of Botany*, *96*(8), 1462–1486.

Hsu, B., Coupar, I. M., & Ng, K. (2006). Antioxidant activity of hot water extract from the fruit of the doum palm, Hyphaene thebaica. *Food Chemistry*, *98*, 317–328.

Huang, D., Zhang, Y., Jin, M., Li, H., Song, Z., Wang, Y., & Chen, J. (2014). Characterization and high cross-species transferability of microsatellite markers from the floral transcriptome of Aspidistra saxicola (Asparagaceae). *Molecular Ecology Resources*, *14*(3), 569–577.

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

Hussein, N., Al-Rasbi, S., Al-Wahaibi, S. N., Abdul, G., Al-hanum, R., & Abdalla, E.-S. O. (2011). Salinity problems and their management in date palm production. In *Salinity problems and their management in date palm production* (pp. 87–123).

Hyun, D. Y., Sebastin, R., Lee, K. J., Lee, G. A., Shin, M. J., Kim, S. H., Lee, J. R., & Cho, G. T. (2020). Genotyping-by-Sequencing Derived Single Nucleotide Polymorphisms Provide the First Well-Resolved Phylogeny for the Genus Triticum (Poaceae). *Frontiers in Plant Science*, *11*(688), 1–15.

Icheria, B. K. (2015). A Social Narrative on Tharaka People in Kenya , Africa. *International Journal of Humanities Social Sciences and Education*, *2*(2), 49–55.

Idohou, R., Assogbadjo, A. E., Houehanou, T., Glèlè Kakaï, R., & Agbangla, C. (2015). Variation in Hyphaene thebaica mart. Fruit: Physical characteristics and factors affecting seed germination and seedling growth in Benin (West Africa). *Journal of Horticultural Science and Biotechnology*, *90*(3), 291–296.

Idohou, Rodrigue, Assogbadjo, A. E., Azihou, F., Kakaï, R. G., & Adomou, A. (2016). Influence of the landscape context on stand structure and spatial patterns of the doum palm (Hyphaene thebaica Mart.) in the Republic of Benin (West Africa). *Agroforestry Systems*, *90*(4), 591–605.

Imadi, S. R., Kazi, A. G., Ahanger, M. A., Gucel, S., & Ahmad, P. (2015). Plant transcriptomics and responses to environmental stress: an overview. *Journal of Genetics*, *94*(3), 525–537.

Ishida, M., & Hassan, A. O. (1997). Utilization of oil palm frond as cattle feed. *Japan Agricultural Research Quarterly*, *31*(1), 41–47.

Jadid, N., Kurniawan, E., Himayani, C. E. S., Andriyani, Prasetyowati, I., Purwani, K. I., Muslihatin, W., Hidayati, D., & Tjahjaningrum, I. T. D. (2020). An ethnobotanical study of medicinal plants used by the Tengger tribe in Ngadisari village, Indonesia. *PLoS ONE*, *15*(7 July), 1–16.

Jaetzold, R., & Schmidt, H. (2009). Farm Management Handbook of Kenya. In *Natural Condition and Farm Management Information*. Ministry of Agriculture and German Agricultural Team (GTZ), Nairobi.

Jain, M. S., Al-Khayri, J. M., & Johnson, D. V. (2011). Date Palm Biotechnology. In *Springer*.

Jajoo, A. (2014). Changes in Photosystem II Heterogeneity in Response to High Salt Stress. In S. . Allakhverdiev, A. . Rubin, & V. . Shuvalov (Eds.), *Contemporary Problems of Photosynthesis* (Vol. 2, Issue June, pp. 397–413).

Janick, J. (2014). Agriculture : Origins of Agriculture in Egypt. *Springer Reference*, 1–18.

Jogawat, A., Vadassery, J., Verma, N., Oelmüller, R., Dua, M., Nevo, E., & Johri, A. K. (2016). PiHOG1, a stress regulator MAP kinase from the root endophyte fungus Piriformospora indica, confers salinity stress tolerance in rice plants. *Scientific Reports*, *6*(October), 1–15.

Johnson, D. V. (1998). 7 AFRICAN AND THE WESTERN INDIAN OCEAN REGION Africa. In *Non Timber Forest products, Tropical palms* (p. 147). Food and Agriculture Organisation of the United Nations.

Jombart, T. (2008). adegenet : a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405.

Kahn, F., & Luxereau, A. (2008). Doum palm habit and leaf collecting practices in Niger. *Palms*, *52*(1), 23–29.

Kalra, Y. P. (1998). *Handbook of Reference Methods for Plant Analysis* (Vol. 38, Issue 6). Taylor & Francis.

Kenya Inter Agency Rapid assessment. (2014). *Turkana Secondary Data Review As At March 2014* (Issue March).

Khalil, O. A., Ibrahim, R. A., & Youssef, M. (2020). A comparative assessment of phenotypic and molecular diversity in Doum (Hyphaene thebaica L .). *Molecular Biology Reports*, *47*, 275–284.

Khallaf, I., Wahman, R., Farghaly, H., & Bayoumi, S. (2022). The Inhibitory Effect of Doum Palm (Hyphaene thebaica) Leaves Extract on α-Glucosidase. *Bulletin of Pharmaceutical Sciences*, *45*(1), 163–176.

Khan, S., Al-qurainy, F., & Nadeem, M. (2012). Biotechnological approaches for conservation and improvement of rare and endangered plants of Saudi Arabia. *Saudi Journal of Biological Sciences*, *19*(1), 1–11.

Kharusi, L. Al, Assaha, D. V. M., Al-Yahyai, R., & Yaish, M. W. (2017). Screening of date palm (Phoenix dactylifera L.) cultivars for salinity tolerance. *Forests*, *8*(4).

Kidane, L., & Kejela, A. (2021). Food security and environment conservation through sustainable use of wild and semi-wild edible plants: a case study in Berek Natural Forest, Oromia special zone, Ethiopia. *Agriculture and Food Security*, *10*(1), 1–16.

Kitheka, J. U., & Ongwenyi, G. S. (2002). The Tana River Basin and the opportunity for research on the land-ocean interaction in the Tana Delta. In *Aquadocs*.

Klimova, A., Ortega-Rubio, A., Vendrami, D. L. J., & Hoffman, J. I. (2018). Genotyping by sequencing reveals contrasting patterns of population structure, ecologically mediated divergence, and long-distance dispersal in North American palms. *Ecology and Evolution*, *8*(11), 5873–5890.

Krishnamurthy, P., Mohanty, B., Wijaya, E., Lee, D. Y., Lim, T. M., Lin, Q., Xu, J., Loh, C. S., & Kumar, P. P. (2017). Transcriptomics analysis of salt stress tolerance in the roots of the mangrove Avicennia officinalis. *Scientific Reports*, *7*(1), 1–19.

Krishnaraj, G., Elagovan, N., & Prakash, C. (2022). Factors Influencing Consumer Purchasing Behaviour Towards Purchase of Palm Leaf Craft. *Journal of Natural Fibers*, 1–11.

Kumar, S., Li, G., Yang, J., Huang, X., Ji, Q., Liu, Z., Ke, W., & Hou, H. (2021). Effect of Salt Stress on Growth, Physiological Parameters, and Ionic Concentration of Water Dropwort (Oenanthe javanica) Cultivars. *Frontiers in Plant Science*, *12*(June).

Lahaye, R., Van Der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Maurin, O., Duthoit, S., Barraclough, T. G., & Savolainen, V. (2008). DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences*

of the United States of America, *105*(8), 2923–2928.

Laibuni, N., Losenge, T., & Bokelmann, W. (2020). Can African Indigenous Vegetables Contribute To Nutrition Security? a Policy Perspective. *International Journal of Food and Agricultural Economics (IJFAEC)*, *8*(2), 111–124.

Lambert, M. J. (1980). Preparation of plant material for estimating a wide range of elements. In *Research Note No. 29*.

Lang'at, J. K. S. (2008). *Variability of mangrove forests along the Kenyan coast*.

Lebrun, P., Grivet, L., & Baudouin, L. (1999). Use of RFLP markers to study the diversity of the coconut palm. *Current Advances in Coconut Biotechnology*, 73–87.

Li, B., & Dewey, C. . (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a refrenece genome. *BMC Bioinformatics*, *12*(33).

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760.

Li, J., Yuan, J., Li, Y., Sun, H., Ma, T., Huai, J., Yang, W., Zhang, W., & Lin, R. (2022). The CDC48 complex mediates ubiquitin-dependent degradation of intra-chloroplast proteins in plants. *Cell Reports*, *39*(2), 110664.

Li, Q., Su, X., Ma, H., Du, K., Yang, M., Chen, B., Fu, S., Fu, T., Xiang, C., Zhao, Q., & Xu, L. (2021). Development of genic SSR marker resources from RNA-seq data in Camellia japonica and their application in the genus Camellia. *Scientific Reports*, *11*(1), 1–11.

Li, X., Yang, W., Jia, J., Zhao, P., Qi, D., Chen, S., Cheng, L., Cheng, L., & Liu, G. (2021). Ectopic expression of a salt-inducible gene, LcSAIN3, from sheepgrass improves seed germination and seedling growth under salt stress in arabidopsis. *Genes*, *12*(12).

Lim, S. L., D'Agui, H. M., Enright, N. J., & He, T. (2017). Characterization of Leaf

Transcriptome in Banksia hookeriana. *Genomics, Proteomics and Bioinformatics*, *15*(1), 49–56.

Lima, B. L. D. C., Lacerda, C. F. De, Neto, M. F., Ferreira, J. F. S., Bezerra, A. M. E., & Marques, E. C. (2017). Physiological and ionic changes in dwarf coconut seedlings irrigated with saline water. *Revista Brasileira de Engenharia Agrícola e Ambiental*, *21*(2), 122–127.

Lo, H. Y., Li, C. C., Huang, H. C., Lin, L. J., Hsiang, C. Y., & Ho, T. Y. (2012). Application of transcriptomics in chinese herbal medicine studies. *Journal of Traditional and Complementary Medicine*, *2*(2), 105–114.

Lokuruka, M. N. I. (2008). Fatty acids in the nut of the Turkana doum palm (Hyphaene coriaceae). *African Journal of Food Agriculture Nutrition and Development*, *8*(2), 118–132.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(550), 1–21.

Luo, H., Laluk, K., Lai, Z., Veronese, P., Song, F., & Mengiste, T. (2010). The arabidopsis botrytis susceptible1 interactor defines a subclass of RING E3 ligases that regulate pathogen and stress responses. *Plant Physiology*, *154*(4), 1766–1782.

Luo, Z., Brock, J., Dyer, J. M., Kutchan, T., Schachtman, D., Augustin, M., Ge, Y., Fahlgren, N., & Abdel-Haleem, H. (2019). Genetic diversity and population structure of a Camelina sativa spring panel. *Frontiers in Plant Science*, *10*(February), 1–12.

Maia, L. C. Da, Palmieri, D. A., Souza, V. Q. De, Kopp, M. M., Carvalho, F. I. F. De, & Costa De Oliveira, A. (2008). SSR locator: Tool for simple sequence repeat discovery integrated with primer design and PCR simulation. *International Journal of Plant Genomics*, *2008*.

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021).

BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, *38*(10), 4647–4654.

Martins, D. O. R. A., & Shackleton, M. C. (2019). Local use and knowledge of Hyphaene coriaceae and Phoenix reclinata in South African Journal of Botany. *South African Journal of Botany*, *138*, 65–75.

Maryam, Jaskani, M. J., Awan, F. S., Ahmad, S., & Khan, I. A. (2016). Development of molecular method for sex identification in date palm (Phoenix dactylifera L.) plantlets using novel sex-linked microsatellite markers. *3 Biotech*, *6*(22), 1–7.

Mathew, L. S., Seidel, M. A., George, B., Mathew, S., Spannagl, M., Haberer, G., Torres, M. F., Al-Dous, E. K., Al-Azwani, E. K., Diboun, I., Krueger, R. R., Mayer, K. F. X., Mohamoud, Y. A., Suhre, K., & Malek, J. A. (2015). A genome-wide survey of date palm cultivars supports two major subpopulations in Phoenix dactylifera. *G3 Genes, Genomes, Genetics*, *5*(7), 1429–1438.

Matsui, A., Ishida, J., Morosawa, T., Mochizuki, Y., Kaminuma, E., Endo, T. A., Okamoto, M., Nambara, E., Nakajima, M., Kawashima, M., Satou, M., Kim, J. M., Kobayashi, N., Toyoda, T., Shinozaki, K., & Seki, M. (2008). Arabidopsis transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. *Plant and Cell Physiology*, *49*(8), 1135–1149.

Maundu, P. ., Ngugi, G. ., & Kabuye, H. . (1999). *Traditional Food plants of Kenya*. National Museums of Kenya.

Maundu, P., & Tengnas, B. (2005). Useful trees and shrubs for Kenya. In *Technical Handbook Number 35*. World Agroforestry Centre- Eastern and Central Africa Regional Programme.

Mayes, S., Jack, P. L., & Corley, R. H. V. (2000). The use of molecular markers to investigate the genetic structure of an oil palm breeding programme. *Heredity*, *85*(3), 288–293.

Mayes, S., Jack, P. L., Marshall, D. F., & Corley, R. H. V. (1997). Construction of a RFLP genetic linkage map for oil palm (Elaeis guineensis Jacq.). *Genome*, *40*(1), 116–122.

Mba, E. H., Ekpo, A. S., Ozim, E. C., & Oladeinde, S. O. (2019). Assessment of Traditional Palm Wine Tapping Practice Effect on Vegetation in Nasarawa State, Nigeria. *International Journal of Environment and Climate Change*, *9*(12), 841–851.

McCarthy, T. S., Ellery, W. N., & Gieske, A. (1994). Possible Ground Water Pollution By Sewage Effluent At Camps In The Okavango Delta :Suggestions For Its Prevention. In *Botswana Notes and Records* (Vol. 26, pp. 129–138). The Botswana Society Gaberone.

Melaku, A., & Ebrahim, M. A. (2021). Critical Review on Wild-Edible Fruit Species in Ethiopia. *International Journal of Forestry Research*, *2021*.

Miyamoto, N., Fernández-Manjarrés, J. F., Morand-prieur, M.-E., Bertolino, P., & Frascaria-Lacoste, N. (2008). What sampling is needed for reliable estimations of genetic diversity in Fraxinus excelsior L. (Oleaceae)? *Annals of Forest Science*, *65*(4), 1–7.

Mohamed, H. E. A., Afridi, S., Khalil, A. T., Zia, D., Shinwari, Z. K., Dhlamini, M. S., & Maaza, M. (2020). Structural, Morphological and Biological Features of ZnO Nanoparticles Using Hyphaene thebaica (L.) Mart. Fruits. *Journal of Inorganic and Organometallic Polymers and Materials*, *30*(8), 3241–3254.

Mokhtar, M. M., Adawy, S. S., El-assal, S. E. S., & Ebtissam, H. (2016). Genic and Intergenic SSR Database Generation , SNPs Determination and Pathway Annotations , in Date Palm (Phoenix dactylifera L .). *Plos One*, *11*(7), 1–19.

Moles, A. T., Falster, D. S., Leishman, M. R., & Westoby, M. (2004). Small-seeded species produce more seeds per square metre of canopy per year, but not per individual per lifetime. *Journal of Ecology*, *92*(3), 384–396.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, *5*(7), 621–628.

Mosissa, Dereje, Agajie, M., & Reda, D. (2021). An Assessment of the Diversity of Indigenous Forage Plant Species of the Dry Season in Dry lands of North Western Ethiopia: Implication for their Conservation and Sustainable Use. *South Asian Research Journal of Biology and Applied Biosciences*, *3*(1), 1–10.

Moussa, H., Margolis, H. A., Dubé, P. A., & Odongo, J. (1998). Factors affecting the germination of doum palm (Hyphaene thebaica Mart.) seeds from the semi-arid zone of Niger, West Africa. *Forest Ecology and Management*, *104*(1–3), 27–41.

Moussouni, S., Pintaud, J., Vigouroux, Y., & Bouguedoura, N. (2017). Diversity of Algerian oases date palm (Phoenix dactylifera L ., Arecaceae): Heterozygote excess and cryptic structure suggest farmer management had a major impact on diversity. *Plos One*, *12*(4), 1–14.

Mudzengi, C. P., Murwira, A., Zengeya, F. M., & Murungweni, C. (2017). Screening key browse species in a semi-arid rangeland. *Cogent Food and Agriculture*, *3*.

Mugai, E. N. (2004). Salinity Characterization of the Kenyan Saline Soils. *Soil Science and Plant Nutrition*, *50*(2), 181–188.

Munns, R. (2002). Comparative physiology of salt and water stress. *Plant, Cell and Environment*, *25*(2), 239–250.

Munns, R., & James, R. A. (2003). Screening methods for salinity tolerance: A case study with tetraploid wheat. *Plant and Soil*, *253*(1), 201–218.

Mutlu, N., Miklas, P. N., & Coyne, D. P. (2006). Resistance gene analog polymorphism (RGAP) markers co-localize with disease resistance genes and QTL in common bean. *Molecular Breeding*, *17*(2), 127–135.

Muyle, A., Martin, H., Zemp, N., Mollion, M., Gallina, S., Tavares, R., Silva, A.,

Bataillon, T., Widmer, A., Glémin, S., Touzet, P., & Marais, G. A. B. (2020). Dioecy Is Associated with High Genetic Diversity and Adaptation Rates in the Plant Genus Silene. *Molecular Biology and Evolution*, *38*(3), 805–818.

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., Hatipoğlu, R., Ahmad, F., Alsaleh, A., Labhane, N., Özkan, H., Chung, G., & Baloch, F. S. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnology and Biotechnological Equipment*, *32*(2), 261–285.

Nagahatenna, D. S. K., Langridge, P., & Whitford, R. (2015). Tetrapyrrole-based drought stress signalling. *Plant Biotechnology Journal*, *13*(4), 447–459.

Nagata, J. M., Jew, A. R., Kimeu, J. M., Salmen, C. R., Bukusi, elizabeth A., & Cohen, C. R. (2011). Medical Pluralism on Mfangano Island:Use of medicinal plants among persons living with HIV/AIDS in Suba District, Kenya. *Journal of Ethnopharmacology*, *135*(2), 501–509.

Nassiry, M. R., Javanmard, A., & Tohidi, R. (2009). Application of Statistical Procedures for Analysis of Genetic Diversity in Domestic Animal Populations. *American Journal of Animal and Veterinary Sciences*, *4*(4), 136–141.

Nayar, N. . (2017). Origin and domestication. In P. Chowdappa, V. Niral, B. . Jerard, & K. Samsudeen (Eds.), *Coconut* (pp. 39–72). Daya Publishinh House.

Negrão, S., Schmöckel, S. M., & Tester, M. (2017). Evaluating physiological responses of plants to salinity stress. *Annals of Botany*, *119*(1), 1–11.

Ng'asike, T. J., & Swadener, B. B. (2015). Turkana Indegenous Knowledge, EnvironmentalSustainablity and Pastoralist Lifestyle. In S. E. Huaman & B. Sriraman (Eds.), *Advances in Innovation Education* (pp. 107–127). Sense Publishers.

Ngome, T. P. I., Shackleton, C., Degrande, A., & Tieguhong, J. C. (2017). Addressing constraints in promoting wild edible plants' utilization in household nutrition:

Case of the Congo Basin forest area. *Agriculture and Food Security*, *6*(20), 1–10.

Nisar, M., & Hussain, H. (2022). Assessment of the genetic variations of sago palm Metroxylon sagu in three regions of Sarawak , Malaysia using amplified fragment length polymorphism (AFLP) marker. *Chemical and Biological Technologies in Agriculture*, *1*, 1–13.

Niu, S., Song, Q., Koiwa, H., Qiao, D., Zhao, D., Chen, Z., Liu, X., & Wen, X. (2019). Genetic diversity, linkage disequilibrium, and population structure analysis of the tea plant (Camellia sinensis) from an origin center, Guizhou plateau, using genome-wide SNPs developed by genotyping-by-sequencing. *BMC Plant Biology*, *19*(1), 1–12.

Nkhoma, N., Shimelis, H., Laing, M. D., Shayanowako, A., & Mathew, I. (2020). Assessing the genetic diversity of cowpea [Vigna unguiculata (L.) Walp.] germplasm collections using phenotypic traits and SNP markers. *BMC Genetics*, *21*(110), 1–16.

Nolan, J. M., & Turner, N. J. (2011). Ethnobotany: The Study of People-Plant Relationships. In *Ethnobiology* (Issue Chapter 9, pp. 133–147).

Nyambe, M. M., Hakwenye, H., & Benyamen, M. S. (2019). Nutritional and anti-nutritional composition of Diospyros mespiliformis and Hyphaene petersiana fruits from Namibia. *Int.Sci.Technol.J.Namibia*, *13*, 1–10.

Nyero, A., Achaye, I., Odongo, W., Anywar, G., & Malinga, G. M. (2021). Wild and semi-wild edible plants used by the communities of acholi sub-region, northern uganda. *Ethnobotany Research and Applications*, *21*, 1–12.

Okello, J., Okullo, L. B. J., Eilu, G., Nyeko, P., & Obua, J. (2018). Morphological Variations in Tamarindus indica LINN . Fruits and Seed Traits in the Different Agroecological Zones of Uganda. *International Journal of Ecology*, *2018*, 1–12.

Okoye, M. ., Singh, R., Uguru, M. ., & Bakoumé, C. (2020). Application of microsatellite markers for hybrid verification and genetic analysis of oil palm

(Elaeis guineensis Jacq.). *Nigerian Journal of Biotechnology*, *37*(2), 1–12.

Olukoye, G. ., & Kinyamario, J. . (2009). Community Participation in the Rehabilitation of a Sand Dune Environment in Kenya. *Land Degradation & Development*, *20*, 397–409.

Omar, A., Abdealsiede, S., Alrasheid, A. A., & Elbashir, A. A. (2020). Antioxidant, Antimicrobial Activities and HPLC Quantitative Analysis of Some Sudanese Medicinal Plants. *International Journal of Food and Nutritional Science*.

Omari, J. K., Mworia, J. K., Gichuki, N., & Mligo, C. (2019). Woody Species Composition in Upper Tana River Floodplain of Kenya : Potential Effects of Change in Flood Regimes. *Journal of Sustainability , Environment and Peace*, *1*(3), 91–97.

OmicsBox. (2019). *Bioinformatics Made Easy, BioBam Bioinformatics*. https://www.biobam.com/omicsbox

Ondiko, J. H., & Karanja, A. M. (2021). *Spatial and Temporal Occurrence and Effects of Droughts on Crop Yields in Kenya*. OALib.

Onen, C., & Oryem-Origa, H. (2017). Ethno-Botanical Uses of Ugandan edible wild fruit species. *Environment, Earth and Ecology*, *1*(2), 62–74.

Ong, P. W., Maizura, I., Marhalil, M., Rajanaidu, N., Abdullah, N. A. P., Rafii, M. Y., Ooi, L. C. L., Low, E. T. L., & Singh, R. (2018). Association of SNP markers with height increment in mpob-angolan natural oil palm populations. *Journal of Oil Palm Research*, *30*(1), 61–70.

Ortiz, E. M. (2019). convert a VCF matrix into several matrix formats for phylogenetic analysis. In *vcf2phylipV.2.0* (No. 2).

Orwa, C., Mutua, A., Kindt, R., Jamnadass, R., & Anthony, S. (2009). Hyphaene thebaica (L .) Mart. In *Agroforestree database:a tree reference and selection guide* (pp. 1–5).

Osorio-Guarín, J. A., Garzón-Martínez, G. A., Delgadillo-Duran, P., Bastidas, S., Moreno, L. P., Enciso-Rodríguez, F. E., Cornejo, O. E., & Barrero, L. S. (2019). Genome-wide association study (GWAS) for morphological and yield-related traits in an oil palm hybrid (Elaeis oleifera x Elaeis guineensis) population. *BMC Plant Biology*, *19*(533), 1–11.

Ossoko, J. P. L., Yoca, J., Okandza, Y., Dzondo, M. G., & Tsieri, M. D. (2019). Biochemical characterization of the almonds of hyphaene guineensis of prefecture of Pointe noire in the Republic of Congo. *International Journal of Advanced Research in Engineering and Technology*, *10*(4), 32–39.

Ouertani, R. N., Arasappan, D., Abid, G., Chikha, M. Ben, Jardak, R., Mahmoudi, H., Mejri, S., Ghorbel, A., Ruhlman, T. A., & Jansen, R. K. (2021). Transcriptomic analysis of salt-stress-responsive genes in barley roots and leaves. *International Journal of Molecular Sciences*, *22*(15).

Pagnotta, M. (2018). Comparison among Methods and Statistical Software Packages to Analyze Germplasm Genetic Diversity by Means of Codominant Markers. *Multidisciplinary Scientific Journal*, *1*(1), 197–215.

Palmweb. (2023). *Hypahene reptans Becc.*,. https://palmweb.org/cdm_dataportal/taxon/c161ba3f-14a7-4c58-9ed8-ea92f7d3b77a#distribution

Panigua-Zambrana, N. Y., Byg, A., Svenning, J. C., Moraes, M., Grandez, C., & Balslev, H. (2007). Diversity of palm uses in the western Amazon. *Biodiversity and Conservation*, *16*, 2771–2787.

Parida, A. K., & Jha, B. (2010). Salt tolerance mechanisms in mangroves: A review. *Trees*, *24*(2), 199–217.

Parry, C., Blonquist, J. M., & Bugbee, B. (2014). In situ measurement of leaf chlorophyll concentration: Analysis of the optical/absolute relationship. *Plant Cell and Environment*, *37*(.), 2508–2520.

Peakall, R., & Smouse, P. E. (2012). GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, *28*(19), 2537–2539.

Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J., & Quackenbush, J. (2003). TIGR gene indices clustering tools (TGICL): A software system for fast clustering of large EST datasets. *Bioinformatics*, *19*(5), 651–652.

Pfadenhauer, J. ., & Klotzli, F. (2020). Zonal Vegetation of the Tropical-Subtropical Dry Zone. In *Global vegetation* (Vol. 43, Issue 4). Springer Nature.

Pocovi, M. I., Collavino, N. G., Gutiérrez, Á., Taboada, G., Castillo, V., Delgado, R., & Mariotti, J. A. (2020). Molecular versus morphological markers to describe variability in sugar cane (Saccharum officinarum) for germplasm management and conservation. *Revista de La Facultad de Ciencias Agrarias*, *52*(1), 40–60.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H., Sorrells, M., & Jannink, J. (2012). Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. *The Plant Genome*, *5*(3).

Polash, M., Sakil, M., Tahjib-Ul-Arif, M., & M, H. (2018). Effect of salinity on osmolytes and relative water content of selected rice genotypes. *Tropical Plant Research*, *5*(2), 227–232.

Pootakham, W., Jomchai, N., Ruang-areerate, P., Shearman, J. R., Sonthirod, C., Sangsrakru, D., Tragoonrung, S., & Tangphatsornruang, S. (2015). Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). *Genomics*, *105*, 288–295.

Pritchard, J., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Purwoko, D., Cartealy, I. C., Tajuddin, T., Dinarti, D., & Sudarsono, S. (2019). SSR

identification and marker development for sago palm based on NGS genome data. *Breeding Science*, *69*(1), 1–10.

Qi, W. H., Jiang, X. M., Yan, C. C., Zhang, W. Q., Xiao, G. S., Yue, B. S., & Zhou, C. Q. (2018). Distribution patterns and variation analysis of simple sequence repeats in different genomic regions of bovid genomes. *Scientific Reports*, *8*(1), 1–13.

Qi, W. H., Lu, T., Zheng, C. L., Jiang, X. M., Jie, H., Zhang, X. Y., Yue, B. S., & Zhao, G. J. (2020). Distribution patterns of microsatellites and development of its marker in different genomic regions of forest musk deer genome based on high throughput sequencing. *Aging*, *12*(5), 4445–4462.

Qin, J., Dong, W. Y., He, K. N., Yu, Y., Tan, G. D., Han, L., Dong, M., Zhang, Y. Y., Zhang, D., Li, A. Z., & Wang, Z. L. (2010). NaCI salinity-induced changes in water status, ion contents and photosynthetic properties of Shepherdia argentea (Pursh) Nutt. Seedlings. *Plant, Soil and Environment*, *56*(7), 325–332.

Quansah, E. ., & Makkar, H. P. . (2012). *Use of lesser-known plants and plant parts as animal feed resources in tropical regions* (Animal Production and Health Working Paper No. 8).

Qureshimatva, U., D'Cruz, L., Solanki, H., & Yadav, S. (2018). Distribution of near threatened palm Hyphaene dichotoma (White) Furtado in Gujarat State and Union Territories. *International Journal of Research in Advent Technology*, *6*(11), 3014–3017.

Rabei, S., Said, W. M., Rizk, R. M., & El sharabasy, S. F. (2012). Morphometric Taxonomy of Date Palm Diversity Growing in Egypt. *Egypt. J. Bot*, *189*, 175–189.

Radoglou, K. M., & Jarvis, P. G. (1990). Effects of CO2 enrichment on four poplar clones. II. Leaf surface properties. *Annals of Botany*, *65*(6), 627–632.

Radwan, O., Arro, J., Keller, C., & Korban, S. S. (2015). RNA-Seq Transcriptome

Analysis in Date Palm Suggests Multi-Dimensional Responses to Salinity Stress. *Tropical Plant Biology*, *8*(3–4), 74–86.

Rajesh, M. K., Jerard, B. A., Preethi, P., Thomas, R. J., & Karun, A. (2014). Application of RAPD markers in hybrid verification in coconut. *Crop Breeding and Applied Biotechnology*, *14*(1), 36–41.

Ravi, D., Siril, E. A., & Nair, B. R. (2021). SCAR Marker Development for the Identification of Elite Germplasm of Moringa Oleifera Lam.-A Never Die Plant. *Plant Molecular Biology Reporter*, *39*(4), 850–861.

Reddy, M. T., Kalpana, M., Sivaraj, N., Kamala, V., Pandravada, S. R., & Sunil, N. (2019). Indigenous Traditional Knowledge on Health and Equitable Benefits of Oil Palm (Elaeis spp.). *OALib*, *06*(01), 1–25.

Rejeb, K., Abdelly, C., & Savouré, A. (2014). How reactive oxygen species and proline face stress together. *Plant Physiology and Biochemistry*, *80*(April), 278–284.

Reyes-garcía, V., Guèze, M., Luz, A. C., Paneque-gálvez, J., Macia, J. M., Pino, J., & Rubio-Campillo, X. (2013). Evidence of traditional knowledge loss among a contemporary indigenous society. *Evol Hum Behav*, *34*(4), 249–257.

Rizk, R. M., & El Sharabasy, S. F. (2006). A Descriptor for Date palm (Phoenix dactylifera L.) Characterisation and evaluation in Gene Banks. *American-Eurasian J.Agric. & Environ. Sci*, *1*(2), 133–145.

Romojaro, A., Botella, M. Á., Obón, C., & Pretel, M. T. (2013). Nutritional and antioxidant properties of wild edible plants and their use as potential ingredients in the modern diet. *International Journal of Food Sciences and Nutrition*, *64*(8), 944–952.

Ryu, H., & Cho, Y. G. (2015). Plant hormones in salt stress tolerance. *Journal of Plant Biology*, *58*(3), 147–155.

Sabino, F. da C., de Souza, L. S. B., de Souza, M. A. G., de Barros, J. P. A., de Lucena, L. R. R., Jardim, A. M. da R. F., Rocha, A. K. P., & da Silva, T. G. F. (2021). Morphological characteristics, biomass accumulation and gas exchange of an important species native for restoration in Semi-arid Brazilian areas affected by salt and water stress. *Plant Stress*, *2*.

Sabir, J. S. M., Abo-Aba, S., Bafeel, S., Zari, T. A., Edris, S., Shokry, A. M., Atef, A., Gadalla, N. O., Ramadan, A. M., Al-Kordy, M. A., El-Domyati, F. M., Jansen, R. K., & Bahieldin, A. (2014). Characterization of ten date palm (Phoenix dactylifera L.) cultivars from Saudi Arabia using AFLP and ISSR markers. *Comptes Rendus - Biologies*, *337*(1), 6–18.

Sadeghi, Z., & Kuhestani, K. (2014). Ethnobotany of Date Palm (Phoenix Dactylifera) in Baluch Tribe of Saravan region, Baluchistan, Iran. *International Journal of Agricultural Technology*, *10*(6), 1563–1571.

Sairkar, P. K., Sharma, A., & Shukla, N. P. (2016). SCAR Marker for Identification and Discrimination of Commiphora wightii and C. myrrha. *Molecular Biology International*, *2016*, 1–10.

Salako, K. V., Moreira, F., Gbedomon, R. C., Tovissodé, F., Assogbadjo, A. E., & Glèlè Kakaï, R. L. (2018). Traditional knowledge and cultural importance of Borassus aethiopum Mart. in Benin: Interacting effects of socio-demographic attributes and multi-scale abundance. *Journal of Ethnobiology and Ethnomedicine*, *14*(1), 1–16.

Salako, V., Kegbe, A. M., Chadare, F. J., Kafoutchoni, K. M., Amagnide, A., Gbedomon, R. C., Assogbadjo, A. E., Agbangla, C., & Glele Kakai, R. (2019). Potential for domestication of Borassus aethiopum Mart ., a wild multipurpose palm species in Sub-Saharan Africa. *Genet Resour Crop Evol*, *7*, 1–16.

Salem, A. Z. M., Kunst, C. R., & Jose, S. (2020). Alternative animal feeds from agroforestry plants. *Agroforestry Systems*, *94*(4), 1133–1138.

Salomon-Torres, R., Ortiz-Uribe, N., Villa-Angulo, C., Villa-Angulo, R., & Yaurima-Basaldúa, V. H. (2017). Assessment SSR markers used in analysis of genetic diversity of date palm (Phoenix dactylifera L.). *Plant Cell Biotechnology and Molecular Biology*, *18*(5–6), 269–280.

Sanchez-Bel, P., Romojaro, A., Egea, I., & Pretel, M. T. (2015). Wild edible plants as potential antioxidant or nutritional supplements for beverages minimally processed. *Lwt - Food Science and Technology*, *62*(1), 830–837.

Santos, P. H. A. D., Venâncio, T. M., dos Santos, P. H. D., Ramos, H. C. C., Arêdes, F. A. S., Azevedo, A. O. N., Boechat, M. S. B., de Souza Filho, G. A., Ramos, S. R. R., Mirisola, L. A., Aragão, W. M., & Pereira, M. G. (2020). Genotyping-by-sequencing technology reveals directions for coconut (Cocos nucifera L.) breeding strategies for water production. *Euphytica*, *216*(3).

Saraswati, K. P., & Srinivasan, M. . (2015). Morphology, Taxonomy and Concepts of Species. In *Micropaleontology* (pp. 53–65). Springer.

Sarfo, J., Keding, G. B., Boedecker, J., Pawelzik, E., & Termote, C. (2020). The Impact of Local Agrobiodiversity and Food Interventions on Cost, Nutritional Adequacy, and Affordability of Women and Children's Diet in Northern Kenya: A Modeling Exercise. *Frontiers in Nutrition*, *7*(August), 1–12.

Sathish, D. K., & Mohankumar, C. (2007). RAPD markers for identifying oil palm (Elaeis guineensis Jacq.) parental varieties (dura & pisifera) and the hybrid tenera. *Indian Journal of Biotechnology*, *6*(3), 354–358.

Savo, V., Salomone, F., Mattoni, E., Tofani, D., & Caneva, G. (2019). Traditional Salads and Soups with Wild Plants as a Source of Antioxidants: A Comparative Chemical Analysis of Five Species Growing in Central Italy. *Evidence-Based Complementary and Alternative Medicine*, *2019*.

Schaafsma, M., Morse-Jones, S., Posen, P., Swetnam, R. D., Balmford, A., Bateman, I. J., Burgess, N. D., Chamshama, S. A. O., Fisher, B., Freeman, T., Geofrey, V.,

Green, R. E., Hepelwa, A. S., Hernández-Sirvent, A., Hess, S., Kajembe, G. C., Kayharara, G., Kilonzo, M., Kulindwa, K., … Turner, R. K. (2014). The importance of local forest benefits: Economic valuation of non-timber forest products in the eastern Arc mountains in Tanzania. *Global Environmental Change*, *24*(1), 295–305.

Scheben, A., Batley, J., & Edwards, D. (2017). Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnology Journal*, *15*(2), 149–161.

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

Schunko, C., Li, X., Klappoth, B., Lesi, F., Porcher, V., Porcuna-Ferrer, A., & Reyes-García, V. (2022). Local communities' perceptions of wild edible plant and mushroom change: A systematic review. *Global Food Security*, *32*(December 2021).

Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., & Wolf, J. B. W. (2017). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, 907–917.

Shan, F., Seaton, K., & Yang, H. (2010). Early identification of waxflower (Chamelaucium) hybrids using RGAP markers. *New Zealand Journal of Crop and Horticultural Science*, *38*(3), 217–224.

Sharma, A., Shahzad, B., Rehman, A., Bhardwaj, R., Landi, M., & Zheng, B. (2019). Response of phenylpropanoid pathway and the role of polyphenols in plants under abiotic stress. *Molecules*, *24*(13), 1–22.

Sharma, Rekha, Kumar, B., Arora, R., Ahlawat, S., Mishra, A. K., & Tantia, M. S. (2016). Genetic diversity estimates point to immediate efforts for conserving the endangered Tibetan sheep of India. *Meta Gene*, *8*, 14–20.

Sharma, Ruchika, & Tamta, S. (2017). Red rot resistant gene characterization using RGAP markers among sugarcane cultivars resistant and susceptible to the red rot disease. *3 Biotech*, *7*(5), 1–9.

Shumsky, S. A., Hickey, G. M., Pelletier, B., & Johns, T. (2014). Understanding the contribution of wild edible plants to rural Social-ecological resilience in semi-arid Kenya. *Ecology and Society*, *19*(4).

Siadjeu, C., Mayland-Quellhorst, E., & Albach, D. C. (2018). Genetic diversity and population structure of trifoliate yam (Dioscorea dumetorum Kunth) in Cameroon revealed by genotyping-by-sequencing (GBS). *BMC Plant Biology*, *18*(1), 1–14.

Simozrag, A., Chala, A., Djerouni, A., & Bentchikou, M. E. (2016). Phenotypic diversity of date palm cultivars (Phoenix dactylifera L .) from Algeria. *Gayana Bot*, *73*(1), 42–53.

Simpson, M. G. (2019). *Plant systemstics* (3rd editio). Elsevier Inc.

Sinclair, E. A., & Hobbs, R. J. (2009). Sample size effects on estimates of population genetic structure: Implications for ecological restoration. *Restoration Ecology*, *17*(6), 837–844.

Singh, N., Choudhury, D. R., Singh, A. K., Kumar, S., Srinivasan, K., Tyagi, R. K., Singh, N. K., & Singh, R. (2013). Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PLoS ONE*, *8*(12), 1–14.

Singh, R., Soon-Guan, T., Panandam, J., Rahman, R. A., & Cheah, S. C. (2008). Identifcation of cDNA-RFLP markers and their use for molecular mapping in oil palm (elaeis guineensis). *Asia-Pacific Journal of Molecular Biology and Biotechnology*, *16*(3), 53–63.

Srivastava, S., Avvaru, A. K., Sowpati, D. T., & Mishra, R. K. (2019). Patterns of microsatellite distribution across eukaryotic genomes. *BMC Genomics*, *20*(1), 1–14.

Stauffer, F.W, Ouattara, D., & Stork, A. . (2014). Monocotyledons 2 (Anthericaceae - Palmae). In J. Lebrun & A. . Stork (Eds.), *Tropical African Flowering Plants:ecology and distribution* (Vol. 8, Issue 1, pp. 326–354). Conservatoire et Jardin botaniques de la Ville de Genève.

Stauffer, F W, Roguet, D. J., Christe, C., Naciri, Y., Perret, M., & ... (2018). A multidisciplinary study of the doum palms (Hyphaene Gaertn.): origin of the project, current advances and future perspectives. *Societe Botanique de Geneva*, *July*, 1–24.

Stauffer, Fred W., Ouattara, D. N., Roguet, D., da Giau, S., Michon, L., Bakayoko, A., & Ekpe, P. (2017). An update to the African palms (Arecaceae) floristic and taxonomic knowledge, with emphasis on the West African region. *Journal of Plant Taxonomy and Geography*, *72*(1), 17–30.

Stauffer, Fred W, Roguet, D. J., Christe, C., Naciri, Y., Perret, M., & Quattara, N. D. (2018). A multidisciplinary study of the doum palms (Hyphaene Gaertn .): origin of the project , current advances and future perspectives. *Saussurea*, *47*(July), 97–115.

Stave, J., Oba, G., Nordal, I., & Stenseth, N. C. (2006). Seedling establishment of Acacia tortilis and Hyphaene compressa in the Turkwel riverine forest, Kenya. *African Journal of Ecology*, *44*(2), 178–185.

Sturmberger, L., Wallace, P. W., Glieder, A., & Birner-Gruenberger, R. (2016). Synergism of proteomics and mRNA sequencing for enzyme discovery. *Journal of Biotechnology*, *235*, 132–138.

Sultan, S. E. (2003). Phenotypic plasticity in plants : a case study in ecological development. *Evolution and Development*, *5*(1), 25–33.

Suwardi, A. B., Navia, Z. I., Harmawan, T., Syamsuardi, & Mukhtar, E. (2020). Wild edible fruits generate substantial income for local people of the gunung leuser national park, aceh tamiang region. *Ethnobotany Research and Applications*,

20(11), 1–13.

Taheri, S., Abdullah, T. L., Yusop, M. R., Hanafi, M. M., Sahebi, M., Azizi, P., & Shamshiri, R. R. (2018). Mining and development of novel SSR markers using Next Generation Sequencing (NGS) data in plants. *Molecules*, *23*(2).

Talukdar, N. R., Choudhury, P., Barbhuiya, R. A., & Singh, B. (2021). Importance of Non-Timber Forest Products (NTFPs) in rural livelihood: A study in Patharia Hills Reserve Forest, northeast India. *Trees, Forests and People*, *3*(100042).

Tao, R., Ding, J., Li, C., Zhu, X., Guo, W., & Zhu, M. (2021). Evaluating and Screening of Agro-Physiological Indices for Salinity Stress Tolerance in Wheat at the Seedling Stage. *Frontiers in Plant Science*, *12*(646175), 1–12.

Taranto, F., D'Agostino, N., Greco, B., Cardi, T., & Tripodi, P. (2016). Genome-wide SNP discovery and population structure analysis in pepper (Capsicum annuum) using genotyping by sequencing. *BMC Genomics*, *17*(1), 1–13.

Tardío, J., & Pardo-De-Santayana, M. (2008). Cultural importance indices: A comparative analysis based on the useful wild plants of southern Cantabria (northern Spain). *Economic Botany*, *62*(1), 24–39.

Tebkew, M., Asfaw, Z., & Zewudie, S. (2014). Underutilized wild edible plants in the Chilga District, northwestern Ethiopia:focus on wild woody plants. *Agriculture and Food Security*, *3*(12), 1–16.

Tebkew, M., Gebremariam, Y., Mucheye, T., Alemu, A., Abich, A., & Fikir, D. (2018). Uses of wild edible plants in Quara district, northwest Ethiopia: Implication for forest management. *Agriculture and Food Security*, *7*(12), 1–15.

Terry, M. J., & Smith, A. G. (2013). A model for tetrapyrrole synthesis as the primary mechanism for plastid-to-nucleus signaling during chloroplast biogenesis. *Frontiers in Plant Science*, *4*(FEB), 1–14.

Thareja, G., Mathew, S., Mathew, L. S., Mohamoud, Y. A., Suhre, K., & Malek, J. A.

(2018). Genotyping-by-sequencing identifies date palm clone preference in agronomics of the State of Qatar. *PLoS ONE*, *13*(12), 1–11.

Thawaro, S., & Te-Chato, S. (2008). RAPD (random amplified polymorphic DNA) marker as a tool for hybrid oil palm verification from half mature zygotic embryo culture. *Journal of Agricultural Technology*, *4*(2), 165–176.

Ting, N. C., Zaki, N. M., Rosli, R., Low, E. T. L., Ithnin, M., Cheah, S. C., Tan, S. G., & Singh, R. (2010). SSR mining in oil palm EST database: Application in oil palm germplasm diversity studies. *Journal of Genetics*, *89*(2), 135–145.

Tomkins, T., & Drackley, J. K. (2010). Applications of palm oil in animal nutrition. *Journal of Oil Palm Research*, *22*(DECEMBER), 835–845.

Tomlinson, B. P., & Huggett, B. A. (2012). Cell longevity and sustained primary growth in palm stems. *American Journal of Botany*, *99*(12), 1891–1902.

Tomlinson, H. ., & Huggett, B. A. (2012). Cell longevity and sustained primary growth in palm stems. *American Journal of Botany*, *99*(12), 1891–1902.

Traveset, A., & Rodríguez-Pérez, J. (2018). Seed dispersal. In B. Fath (Ed.), *Encyclopedia of Ecology* (2nd ed., Issue October 2017, pp. 592–599). Elsevier Inc.

Tsehay, S., Ortiz, R., Johansson, E., Bekele, E., Tesfaye, K., Hammenhag, C., & Gutela, M. (2020). New Transcriptome-Based SNP Markers for Noug (Guizotia abyssinica) and Their Conversion to KASP Markers for Population Genetics Analyses. *Genes*, *11*(1373), 1–23.

Tsykun, T., Rellstab, C., Dutech, C., Sipos, G., & Prospero, S. (2017). Comparative assessment of SSR and SNP markers for inferring the population genetic structure of the common fungus Armillaria cepistipes. *Heredity*, *119*(5), 371–380.

Turkana County Integrated Development Plan 2013-2017, (2013). http://www.nafis.go.ke

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*, *40*(15), 1–12.

Van Valkenburg, J., & Dransfield, J. (2004). Hyphaene guineensis. *Palms*, *48*(1), 10–16.

Vandenbeldt, R. (1992). Problems with range-wide provenance trials of Faidherbia albida on sandy soils in Niger. In Vandenbeldt RJ (Ed.), *Proceedings of a Workshop, ICRISAT-ICRAF* (pp. 83–86).

Venugopal, A., Rinu, K. A., & Joseph, D. (2017). Cocos Nucifera : It ' s Pharmacological Activities. *World Journal of Pharmaceutical Sciences*, *5*(8), 195–200.

Vieira, L. R., Silva, V. N. B., Casari, R. A. das C. N., Carmona, P. A. O., Sousa, C. A. F. de, & Souza Junior, M. T. S. (2020). Morphophysiological responses of young oil palm plants to salinity stress. *Pesquisa Agropecuaria Brasileira*, *55*, 1–15.

Vodouhè, R., Dansi, A., Avohou, H. T., Kpèki, B., & Azihou, F. (2011). Plant domestication and its contributions to in situ conservation of genetic resources in Benin. *International Journal of Biodiversity and Conservation*, *3*(2), 40–56.

Wales, N., & Blackman, B. K. (2017). Plant Domestication: Wild Date Palms Illuminate a Crop's Sticky Origins. *Current Biology*, *27*(14), R702–R704.

Wallace, J. G., & Mitchell, S. E. (2017). Genotyping-by-Sequencing. *Curr. Protoc. Plant Biol*, *2*(March), 64–77. https://doi.org/10.1002/cppb.20042

Wang, J., & Huang, R. (2019). Modulation of ethylene and ascorbic acid on reactive oxygen species scavenging in plant salt response. *Frontiers in Plant Science*, *10*(March), 2014–2019.

Wang, S., Lv, X., Zhang, J., Chen, D., Chen, S., Fan, G., Ma, C., & Wang, Y. (2022). Roles of E3 Ubiquitin Ligases in Plant Responses to Abiotic Stresses.

International Journal of Molecular Sciences, *23*(4).

Wang, W., Vinocur, B., & Altman, A. (2003). Plant responses to drought, salinity and extreme temperatures: Towards genetic engineering for stress tolerance. *Planta*, *218*(1), 1–14.

Wang, Y., Ihase, L. O., Htwe, Y. M., Shi, P., Zhang, D., Li, D., Emoghene, B. O., & Iserhienrhien, A. (2020). Development of sex-linked SSR marker in the genus Phoenix and validation in P. dactylifera. *Crop Science*, *60*(5), 2452–2466.

Wanjohi, B. K., Sudoi, V., Njenga, E. W., & Kipkore, W. K. (2020). An Ethnobotanical Study of Traditional Knowledge and Uses of Medicinal Wild Plants among the Marakwet Community in Kenya. *Evidence-Based Complementary and Alternative Medicine*, *2020*.

Warrag, E. ., Elsheiksh, E. ., & Elfeel, A. . (2002). Forest Genetic Resources Conservation in Sudan. In *Forest Genetic Resources*.

Wellburn, A. R. (1994). The Spectral Determination of Chlorophylls a and b, as well as Total Carotenoids, Using Various Solvents with Spectrophotometers of Different Resolution. *Journal of Plant Physiology*, *144*, 307–313.

Westoby, M., Falster, D. S., Moles, A. T., Vesk, P. A., & Wright, I. J. (2002). Plant ecological strategies: Some leading dimensions of variation between species. *Annual Review of Ecology and Systematics*, *33*, 125–159.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43).

Wiens, J. J. (2004). The role of morphological data in phylogeny reconstruction. *Systematic Biology*, *53*(4), 653–661.

Wright, I. J., Ackerly, D. D., Bongers, F., Harms, K. E., Ibarra-Manriquez, G., Martinez-Ramos, M., Mazer, S. J., Muller-Landau, H. C., Paz, H., Pitman, N. C. A., Poorter, L., Silman, M. R., Vriesendorp, C. F., Webb, C. O., Westoby, M., & Wright, S. J. (2007). Relationships among ecologically important dimensions of plant trait variation in seven neotropical forests. *Annals of Botany*, *99*(5), 1003–1015.

Wu, J., Cai, C., Cheng, F., Cui, H., & Zhou, H. (2014). Characterisation and development of EST-SSR markers in tree peony using transcriptome sequences. *Molecular Breeding*, *34*(4), 1853–1866.

Wyatt, J. (2016). Grain and Plant Morphology of Cereals and how Characters can be used to Identify Varieties. In C. Wrigley, H. Corke, K. Seetharaman, & J. Faubion (Eds.), *Encyclopedia of Food Grains* (Second, Vol. 13, Issue 4, pp. 51–72). Academic Press.

Xia, W., Luo, T., Zhang, W., Mason, A. S., Huang, D., Huang, X., Tang, W., Dou, Y., Zhang, C., & Xiao, Y. (2019). Development of high-density snp markers and their application in evaluating genetic diversity and population structure in elaeis guineensis. *Frontiers in Plant Science*, *10*(February), 1–11.

Xiong, H., Shi, A., Mou, B., Qin, J., Motes, D., Lu, W., Ma, J., Weng, Y., Yang, W., & Wu, D. (2016). Genetic diversity and population structure of cowpea (vigna unguiculata l. walp). *PLoS ONE*, *11*(8), 1–15.

Xu, Z., Zhang, N., Fu, H., Wang, F., Wen, M., Chang, H., Wu, J., Abdelaala, W. B., Luo, Q., Li, Y., Li, C., Wang, Q., & Wang, Z. Y. (2022). Salt Stress Modulates the Landscape of Transcriptome and Alternative Splicing in Date Palm (Phoenix dactylifera L.). *Frontiers in Plant Science*, *12*(January), 1–13.

Yadav, L. K., McAssey, E. V., & Wilde, H. D. (2019). Genetic diversity and population structure of Rhododendron canescens, a native azalea for urban landscaping. *HortScience*, *54*(4), 647–651.

Yaish, M. W., & Kumar, P. P. (2015). Salt tolerance research in date palm tree (Phoenix dactylifera L.), past, present, and future perspectives. *Frontiers in Plant Science*, *6*(MAY), 1–5.

Yaish, M. W., Patankar, H. V., Assaha, D. V. M., Zheng, Y., Al-Yahyai, R., & Sunkar, R. (2017). Genome-wide expression profiling in leaves and roots of date palm (Phoenix dactylifera L.) exposed to salinity. *BMC Genomics*, *18*(1), 1–17.

Yan, G. P., Chen, X. M., Line, R. F., & Wellings, C. R. (2003). Resistance gene-analog polymorphism markers co-segregating with the Yr5 gene for resistance to wheat stripe rust. *Theoretical and Applied Genetics*, *106*(4), 636–643.

Yang, X., Tan, B., Liu, H., Zhu, W., Xu, L., Wang, Y., Fan, X., Sha, L., Zhang, H., Zeng, J., Wu, D., Jiang, Y., Hu, X., Chen, G., Zhou, Y., & Kang, H. (2020). Genetic Diversity and Population Structure of Asian and European Common Wheat Accessions Based on Genotyping-By-Sequencing. *Frontiers in Genetics*, *11*, 1–14.

Ying, S. T., Zaman, F. Q., Ling, H. O. C., Ithnin, M., & Rao, V. (2007). Flanking AFLP markers for the Virescens Trait in Oil Palm. *Journal of Oil Palm*, *19*(December), 381–392.

Yokotani, N., Higuchi, M., Kondou, Y., Ichikawa, T., Iwabuchi, M., Hirochika, H., Matsui, M., & Oda, K. (2011). A novel chloroplast protein, CEST induces tolerance to multiple environmental stresses and reduces photooxidative damage in transgenic Arabidopsis. *Journal of Experimental Botany*, *62*(2), 557–569.

Yousif, S., Alwan, A., Hatem, L., & Ibraheem, H. A. (2020). Sex Determination of Iraqi Date Palms Based on DNA Markers Agricultural Research Directorate , Ministry of Science & * Correspondence author : yousifshatha@yahoo.com Abstract. *Basrah Journal of Date Palm Research*, *19*(1), 18–30.

Yusriani, Y., Rahmah, F. F., Ratnawaty, S., Hilmiati, N., Tirajoh, S., & Hasyim, A. R. (2021). The Potential Side Product of the Oil Palm Plant as Animal Feed in Aceh

Tamiang District, the Province of Aceh. *Advances in Biological Sciences Research*, *13*(ISPLRSAD 2020), 157–162.

Zakar, T., Laczko-Dobos, H., Toth, T. N., & Gombos, Z. (2016). Carotenoids assist in cyanobacterial photosystem II assembly and function. *Frontiers in Plant Science*, *7*(MAR2016), 1–7.

Zaman, M., Shahid, S. A., & Pharis, R. P. (2016). Salinity a Serious Threat to Food Security – Where Do We Stand? *Soils Newsletter Vol. 39, No. 1, July 2016*, *39*(1), 9–10.

Zhang, C., Wu, Z., Jiang, X., Li, W., Lu, Y., & Wang, K. (2021). *De_novo* transcriptomic analysis and identification of EST-SSR markers in Stephanandra incisa. *Scientific Reports*, *11*(1), 1–10.

Zhang, F., Zhu, G., Du, L., Shang, X., Cheng, C., Yang, B., Hu, Y., Cai, C., & Guo, W. (2016). Genetic regulation of salt stress tolerance revealed by RNA-Seq in cotton diploid wild species , Gossypium davidsonii. *Scientific Reports*, *February*, 1–15.

Zhang, J., Zhang, P., Huo, X., Gao, Y., Chen, Y., Song, Z., Wang, F., & Zhang, J. (2021). Comparative Phenotypic and Transcriptomic Analysis Reveals Key Responses of Upland Cotton to Salinity Stress During Postgermination. *Frontiers in Plant Science*, *12*(April), 1–12.

Zhang, X., Yao, Y., Li, X., Zhang, L., & Fan, S. (2020). Transcriptomic analysis identifies novel genes and pathways for salt stress responses in Suaeda salsa leaves. *Scientific Reports*, *10*(4236), 1–12.

Zhang, Z., Xie, W., Zhao, Y., Zhang, J., Wang, N., Ntakirutimana, F., Yan, J., & Wang, Y. (2019). EST-SSR marker development based on RNA-sequencing of E. sibiricus and its application for phylogenetic relationships analysis of seventeen Elymus species. *BMC Plant Biology*, *19*(235), 1–18.

Zhao, Y., Williams, R., Prakash, C. S., & He, G. (2013). Identification and

characterization of gene-based SSR markers in date palm (Phoenix dactylifera L.). *BMC Plant Biology*, *12*(237).

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328.

Zhou, L., Wang, C., Liu, R., Han, Q., Vandeleur, R. K., Du, J., Tyerman, S., & Shou, H. (2014). Constitutive overexpression of soybean plasma membrane intrinsic protein GmPIP1;6 confers salt tolerance. *BMC Plant Biology*, *14*(181), 1–13.

Zou, S., Yao, X., Zhong, C., Gao, P., Wang, Z., & Huang, H. (2020). Phenotypic characterization of Stauntonia obovatifoliola Hayata subsp . urophylla germplasm : a potential new fruit crop. *Genetic Resources and Crop Evolution*, 14.

APPENDICES

Appendix I: Relative Cultural Importance indices adopted from Hoffman and Gallaher 2007

Relative Cultural Importance (RCI) index		
Data Source	Formula	Calculation/Explanation
1) Uses Totaled (Researcher-Tally)		
	$= \sum Uses_{Species\ (i)}$	A simple sum of all known uses for each species. The uses can be categorized by utility, plant taxon or vegetation type.
2) Subjective Allocation (Researcher-Score)		
Use Value (Prance et al. 1987)	$UV_S = \sum_{i}^{n} Value_{UseCategory(i)}$	The species Use Value is a sum of the researcher-generated scores for each of its uses. "Major" uses are scored 1 while "minor" uses are scored 0.5. Uses refer to use-categories (such as construction or food), not specific uses.
Index of Cultural Significance (Turner 1988)	$ICS = \sum_{i=1}^{n} (q * i * e)$	For each species, scores for all uses cited (from 1 to n uses) are added together. The score for each use is determined from the multiplied scores derived from three ordinal scales of significance. q = quality of use [critical resource (5) to little noticed (0)]. i = intensity of use [high (5), low (0)]. e = exclusivity of use: [sub- stitutions available?, (2)-(1)-(0.5)]
Ethnic Index of Cultural Significance (Lajones & Lemas 2001, Stoffle 1990)	$EICS = \sum_{i=1}^{n} (p/u * i * e * c)$	Modified from Turner (1988) to be less subjective. Calculated as the sum of the total number of uses and/or plant parts used for a specific purpose (p/u) multiplied by: i = intensity of use [same as Turner 1988) e = exclusivity of use [preferred by at least one informant (2), not mentioned as preferred (1)]. c = contemporary usage [contemporary (2) or not (1)]
Cultural Significance Index (Silva et al. 2006)	$CSI = \sum_{i=1}^{n} (i * e * c) * CF$	Designed to combine elements from former indices with consensus methodology and binary use classes to reduce subjectivity. i = species mgmt [non-man-aged (1) or managed (2)] e = Use Preference [not preferred (1) or preferred (2)] c = Use Frequency [rarely used (1) or used frequently (2)] CF = Correc-tion factor [number of citations for a given species divided by the number of citations for the most-mentioned species].
3) Informant Consensus (Informant Tally)		
Corrected Fidelity Level (Rank Order Priority) (Friedman 1986)	$FL = I_p / I_u * 100\%$ $ROP = FL * RPL$	The FL quantifies the importance of a species for a given purpose. Ip = number of informants who cited the species for the particular use. Iu = Total number of informants that mentioned the plant for any use. RPL or Relative Popularity Level is a number be- tween 0-1.

Use Values (Phillips & Gentry 1993)		
Species Use-Value for one informant	$UV_{is} = (\sum U_{is})/(n_{is})$	U_{is} = number of uses mentioned for species s by in- formant i and n_{is} = the number of 'events' in which in- formant i cites a use for species s. Tally the number of plant uses mentioned for a given species (all uses equal) and divide by the number of 'events' (all use citations over time of the study for a species by one informant).
Species Use Value (For one species across all informants)	$UV_s = (\sum UV_{is})/(n_i)$	n_i = total number of informants interviewed for spe- cies s. Sum the informant use values for a species and divide by the total number of informants
Family Use Value	$FUV = \sum UVs/(n_s)$	n_S = total number of species within a given family Sum the use values for all the species within a given family and divide by n_S.
Relative Use Value	$RUV_i = \sum (UV_{is}/UV_s)/n_i$	n_i = the number of study species with data from two or more other informants. This gives a standardized measure of how many plant uses an informant knows relative to the average knowledge among all infor- mants.
Overall Use Value (and Plant Part Value) (Gómez-Beloz 2003)		
Reported Use Value	$RU = \sum_{i}^{n} Species_i$	The total number of uses reported for each plant. This is the same value as UV_{iS} (Phillips et al. 1993) except that the number of species citation 'events' per informant is always one (interviews were not re- peated).
Reported Use Value (per plant part)	$RU_{PlantPart}$	The number of uses cited for each plant part (e.g. outer bark, inner bark, root, leaf, flower, fruit).
Plant Part Value	$PPV = \sum RU_{(plantpart)}/\sum RU$	The ratio between the total reported uses for each plant part and the total number of reported uses for a given plant.
Specific Reported Use	SU	The number of times a specific use is reported by the informant (used for partitioning the data into use categories).
Intra-specific Use Value	$IUV = \sum SU_{(plantpart)}/RU_{(plantpart)}$	The ratio of the number of specific uses and reported uses for a given plant part.
Overall Use Value	$OUV = PPV * IUV$	Allows for ranking and comparison of uses within a group of plants. May be calculated in various ways.

Appendix II: Questionnaire for collection of ethnobotany data

GPS Coordinates of sampling sites……

1. DEMOGRAPHIC INFORMATION

a) Description of the farmer
   a. Name of the farmer……
   b. Gender……
b) County……
c) Date……
d) Ethnic group……..

2. DISTRIBUTION AND DOMESTICATION OF DOUM PALM

a) Do you know doum palm?
   - Yes
   - No
b) Seasonality
   - Available only in season
   - Available throughout the year
c) How many trees of Doum palm are in in your farm?
   - 1-5
   - 5-10
   - 10-15
   - More than 15
d) Have you observed any varietal differences in doum palm?
   - Yes
   - No
e) If yes, how many varieties of doum palm can you identify?----------
f) Name them……
g) What is the distribution of doum palm?
   - widely distributed
   - Limited
   - Endangered
h) Is it worthy to domesticate doum palm?
   - Yes
   - No
i) Cropping System
   1. Monoculture
   2. Intercropping
j) Does it allow intercropping?
   - Yes
   - No
   - Only when mature
k) Plants intercropped with doum palm……..
l) Do you do any form of maintenance of doum palm on the farm?
   - Yes
   - No
m) How do you maintain doum palm in your farm?.....

3. USES OF THE PLANT

a) How do you use the doum palm fruit? ……..….
b) How do you use doum palm leaves……..
c) How do you use doum palm stem……..……..
d) How do you use doum palm roots……..
e) Are there special uses of doum palm

1. Feasts
2. Religious purpose
3. Medicine

If yes, describe the use……..……..

f) Economic uses
   1. Export
   2. Local consumers
   3. Traditional industries
   4. Handicrafts

4. BIOTIC AND ABIOTIC STRESS

(a) Is doum palm affected by pests
   If yes describe the pest……..….

(b) Is doum palm affected by diseases
   If yes describe the disease……

(c) Human interference (Harvesting by farmers, logging, clearing)
   1. High
   2. Moderate
   3. No/ low

(d) browsing by animals
   1. High
   2. moderate
   3. Low

e) Signs of Salinity stress…….……..……..

f) Signs of Drought Stress……

Appendix III: Questionnaire for collection of morphological data

Adapted from descriptors for Date palm (Rizk & Sharabasy 2006)

Vegetative Data

(a) Trunk colour
   1. Dark Color
   2. Pale color
   3. Ashy color

(b) Trunk Diameter
1. Thick (>70 cm)
2. Medium (between 50 to 69)
3. Thinner (Less than 50 cm)

(c) Trunk Branching
1. No branching
2. 2 branches
3. More than 2 branches

(d) Leaf length in cm……
(e) Leaf Width at the middle in cm …….
(f) Color of the leaf
   1. Dark green
   2. Green
   3. light green
   4. Ashy green

(g) Mid rib color
   1. Dark green
   2. Glossy green
   3. Light green

(h) Petiole length(cm)……..

(i) Petiole Shape
   1. Slender
   2. Base stout than above

(j) Pinnae Density
1. Very dense
2. dense
3. lax
4. very lax

Fruit Data

(a) Fruit length in cm……..
(b) Fruit width in cm……...
(c) Fruit weight (gm)……..
(d) Fruit volume (cm^3)……...

(e) Fruit shape
   1. Cylindrical
   2. ovate-elongate
   3. obviate elongate
   4. ovate

5. obviate

(f) Fruit apex
1. Obtuse
2. Blunt
3. Truncate
4. Retuse

(g) Fruit base
1. Obtuse
2. Blunt
3. Truncate
4. Retuse

(h) Fruit colour unripe
1. Yellow
2. Yellow brown
3. yellow orange
4. orange
5. green
6. black

(i) Fruit color- mature (Ripe)
1. yellow
2. yellow brown
3. Yellow orange
4. green
5. Black
6. orange

(j) Perianth color
1. Yellow
2. orange yellow
3. orange

(k) Skin nature
1. Smooth
2. Smooth and loose from flesh
3. wrinkled and united with flesh

(l) Skin appearance
1. Shiny
2. Not shiny

(m) Flesh color
1. white
2. yellow
3. whitish creamy
4. brown

(n) Flesh texture
1. soft
2. firm
3. Fibrous
4. dry

(o) Fruit Aroma
1. strong
2. Intermediate
3. Poor

Appendix IV: Procedure for the preparation of Hoagland's nutrient solution for hydroponic studies

Stock Solution
1. Six clean containers are used, 1 to 6 as indicated in column 1 of the table below.
2. Molar solutions for macronutrients and mM solutions are prepared as per the table.
3.Each nutrient is put in a separate bottle except the micronutrients (container 5) which are all combined in one container.
4. These are stock solutions. To prepare working solutions, the **"volume of stock solution per liter"** column is used

Table of concentration of stock and working solutions of Hoaglands nutrient Media.

Container	Macronutrients	M.wgt	M of SS	Amount required for SS	Volume of stock solution per liter of final solution	Element
1	KNO3	101.1	1	101.1	6 mL/L of water	K, N
2	Ca(NO3)2.4H2O	236.16	1	236.16	4 mL/L of water	Ca, N
3	(NH4)2PO4	115.08	1	115.08	2 mL/L of water	N, P
4	MgSO4.7H2O	246.49	1	246.49	1 mL/L of water	Mg, S
5	Micronutrients	(all in one)			1 mL/L of water	
	KCl	74.55	50	3.7275		K, Cl
	H3BO3	61.84	25	1.546		B
	MnSO4.H2O	169.01	2	0.33802		Mn, S
	ZnSO4.7H2O	289.55	2	0.5791		Zn, S
	CuSO4.5H2O	249.71	0.5	0.124855		Cu, S
	H2MoO4 (85% MoO3)	161.97	0.5	0.088985		Mo
6.	Fe-EDTA	346.08	20	6.9216	1 mL/L of water	Fe

Appendix V: DNA integrity results for GBS sequencing

No.	Sample Name	Concentration (ng/μL)	Volume (μL)	Total Mass(μg)	sample Integrity	Test result
1	MW1	261.9	30	7.86	Degraded slightly	Qualified
2	MW2	98	30	2.94	Degraded slightly	Qualified
3	MW3	237.3	21	4.98	Degraded slightly	Qualified
4	MW4	95	30	2.85	Degraded Moderate	Risky
5	MW5	298	28	8.34	Degraded slightly	Qualified
6	MW6	190.5	30	5.72	Degraded slightly	Qualified
7	MW7	90.4	30	2.71	Degraded slightly	Qualified
8	MW8	83.7	30	2.51	Degraded slightly	Qualified
9	KG1	221.4	21	4.65	Degraded slightly	Qualified
10	KG2	147.7	22	3.25	Degraded slightly	Qualified
11	KG3	166.5	33	5.49	Degraded slightly	Qualified
12	KG4	257	22	5.65	Degraded slightly	Qualified
13	KG5	115.8	30	3.47	Degraded slightly	Qualified
14	KG6	135	32	4.32	Degraded slightly	Qualified
15	KG7	142.5	30	4.28	Degraded slightly	Qualified
16	KG8	162.8	47	7.65	Degraded slightly	Qualified
17	KG9	221.2	30	6.64	Degraded slightly	Qualified
18	KG10	49	45	2.21	Degraded slightly	Qualified
19	KG11	148.1	47	6.96	Degraded slightly	Qualified
20	KG12	203.8	30	6.11	Degraded slightly	Qualified
21	KG13	197.8	35	6.92	Degraded slightly	Qualified
22	KG14	110.6	35	3.87	Degraded slightly	Qualified
23	KG15	97.2	36	3.5	Degraded slightly	Qualified
24	K16	164.8	31	5.11	Degraded slightly	Qualified
25	K17	128.8	30	3.86	Moderate degraded	Risky
26	K18	242	32	7.74	Degraded slightly	Qualified
27	K19	162	30	4.86	Degraded slightly	Qualified
28	MM1	57	18	1.03	Degraded slightly	Qualified
29	MM2	123.1	32	3.94	Degraded slightly	Qualified
30	MM3	50.8	24	1.22	Degraded slightly	Qualified
31	MM4	71.8	32	2.3	Degraded slightly	Qualified
32	MM5	45.7	32	1.46	Degraded slightly	Qualified
33	MM6	89.8	21	1.89	Degrade slightly	Qualified
34	MM7	93.6	24	2.25	Degraded slightly	Qualified
35	MM8	119.1	26	3.1	Degraded slightly	Qualified
36	MM9	62.5	31	1.94	Degraded slightly	Qualified
37	R18	142.9	31	4.43	Degraded slightly	Qualified
38	MB1	181.8	31	5.64	Degraded slightly	Qualified
39	MB2	73.5	20	1.47	Degraded slightly	Qualified
40	MB3	87.2	25	2.18	Degraded slightly	Qualified
41	MB4	80.3	27	2.17	Degraded slightly	Qualified
42	MB5	127.9	24	3.07	Degraded slightly	Qualified
43	MB6	199.6	27	5.39	Degraded slightly	Qualified
44	MB7	62	27	1.67	Degraded slightly	Qualified
45	MB8	117.1	22	2.58	Degraded slightly	Qualified
46	MB9	80.7	22	1.78	Degraded slightly	Qualified
47	TG1	129.8	26	3.37	Degraded slightly	Qualified
48	MB10	84	21	1.76	Degraded slightly	Qualified
49	TG2	122	46	5.61	Degraded slightly	Qualified
50	TG3	41.3	46	1.9	Degraded slightly	Qualified
51	TG4	94.1	46	4.33	Degraded slightly	Qualified

52	TG5	161.2	46	7.42	Degraded slightly	Qualified
53	TG6	194.2	50	9.71	Degraded slightly	Qualified
54	TG7	86.8	50	4.34	Degraded slightly	Qualified
55	TG8	74	52	3.85	Degraded slightly	Qualified
56	TG9	152.1	40	6.08	Degraded slightly	Qualified
57	TG10	250.7	30	7.52	Degraded slightly	Qualified
58	TG11	90	40	3.6	Degraded slightly	Qualified
59	TG12	130.4	47	6.13	Degraded slightly	Qualified
60	TG13	122.2	40	4.89	Degraded slightly	Qualified
61	TG14	112.8	47	5.3	Degraded slightly	Qualified
62	TG15	201.4	47	9.47	Degraded slightly	Qualified
63	MK1	59.9	32	1.92	Degraded slightly	Qualified
64	MK2	103.3	32	3.31	Degraded slightly	Qualified
65	MK3	108.7	32	3.48	Degraded slightly	Qualified
66	MK4	136.2	27	3.68	Degraded slightly	Qualified
67	MK5	85.5	27	2.31	Degraded slightly	Qualified
68	MK6	106.5	35	3.73	Degraded slightly	Qualified
69	MK7	53.4	24	1.28	Degraded slightly	Qualified
70	MK8	275.7	34	9.37	Degraded slightly	Qualified
71	R1	208.1	32	6.66	Degraded slightly	Qualified
72	R2	208.9	29	6.06	Degraded slightly	Qualified
73	R3	302.5	29	8.77	Degraded slightly	Qualified
74	R4	275.2	29	7.98	Degraded slightly	Qualified
75	R5	252	29	7.31	Degraded slightly	Qualified
76	R6	279.5	28	7.83	Degraded slightly	Qualified
77	R7	77.5	19	1.47	Degraded slightly	Qualified
78	R8	55.9	29	1.62	Degraded slightly	Qualified
79	R9	63	25	1.58	Degraded Moderate	Risky
80	R10	99.9	28	2.8	Degraded slightly	Qualified
81	R11	70	28	1.96	Degraded slightly	Qualified
82	R12	32.3	28	0.9	Degraded moderately	Risky
83	R13	76.1	28	2.13	Degraded slightly	Qualified
84	R14	272	28	7.62	Degraded slightly	Qualified
85	R15	238	26	6.19	Degraded slightly	Qualified
86	R16	192	28	5.38	Degraded slightly	Qualified
87	R17	250	26	6.5	Degraded slightly	Qualified
88	R19	386	28	10.81	Degraded slightly	Qualified
89	R20	126.4	26	3.29	Degraded slightly	Qualified
90	P1	86.1	26	2.24	Degraded Moderate	Risky
91	P2	136.5	26	3.55	Degraded slightly	Qualified
92	P3	80	26	2.08	Degraded Moderately	Risky
93	P4	123	21	2.58	Degraded slightly	Qualified
94	AA13	190	28	3.4	Degraded slightly	Qualified
95	AA14	200	26	3.8	Degraded slightly	Qualified

Appendix VI: DNA concentration results of Qubit Fluorometer or Microplate Reader

	Sample Name	Test Instrument	Dilution Ratio(×)	Test Volume (µL)	Test Concentration (ng/µL)	Concentration of original sample(ng/µL)
1.	MW1	Microplate reader	1	1	261.886	261.9
2.	MW2	Qubit	1	1	98	98
3.	MW3	Microplate reader	1	1	237.273	237.3
4.	MW4	Microplate reader	1	1	94.967	95
5.	MW5	Qubit	1	1	298	298
6.	MW6	Microplate reader	1	1	190.48	190.5
7.	MW7	Microplate reader	1	1	90.358	90.4
8.	MW8	Microplate reader	1	1	83.748	83.7
9.	KG1	Microplate reader	1	1	221.449	221.4
10.	KG2	Microplate reader	1	1	147.698	147.7
11.	KG3	Microplate reader	1	1	166.536	166.5
12.	KG4	Microplate reader	1	1	257.046	257
13.	KG5	Microplate reader	1	1	115.824	115.8
14.	KG6	Qubit	1	1	135	135
15.	KG7	Microplate reader	1	1	142.532	142.5
16.	KG8	Microplate reader	1	1	162.769	162.8
17.	KG9	Microplate reader	1	1	221.232	221.2
18.	KG10	Microplate reader	1	1	49.013	49
19.	KG11	Microplate reader	1	1	148.064	148.1
20.	KG12	Microplate reader	1	1	203.822	203.8
21.	KG13	Microplate reader	1	1	197.799	197.8
22.	KG14	Microplate reader	1	1	110.591	110.6
23.	KG15	Microplate reader	1	1	97.19	97.2
24.	K16	Microplate reader	1	1	164.776	164.8
25.	K17	Microplate reader	1	1	128.846	128.8
26.	K18	Qubit	1	1	242	242
27.	K19	Microplate reader	1	1	162.043	162
28.	MM1	Microplate reader	1	1	57.014	57
29.	MM2	Microplate reader	1	1	123.115	123.1
30.	MM3	Microplate reader	1	1	50.792	50.8
31.	MM4	Microplate reader	1	1	71.835	71.8
32.	MM5	Microplate reader	1	1	45.663	45.7
33.	MM6	Microplate reader	1	1	89.768	89.8
34.	MM7	Microplate reader	1	1	93.623	93.6
35.	MM8	Microplate reader	1	1	119.058	119.1
36.	MM9	Microplate reader	1	1	62.546	62.5
37.	R18	Microplate reader	1	1	142.948	142.9
38.	MB1	Microplate reader	1	1	181.83	181.8
39.	MB2	Microplate reader	1	1	73.451	73.5
40.	MB3	Microplate reader	1	1	87.192	87.2
41.	MB4	Microplate reader	1	1	80.267	80.3
42.	MB5	Microplate reader	1	1	127.91	127.9
43.	MB6	Microplate reader	1	1	199.617	199.6
44.	MB7	Microplate reader	1	1	61.951	62
45.	MB8	Microplate reader	1	1	117.121	117.1
46.	MB9	Microplate reader	1	1	80.656	80.7
47.	TG1	Microplate reader	1	1	129.757	129.8
48.	MB10	Qubit	1	1	84	84
49.	TG2	Microplate reader	1	1	122.023	122
50.	TG3	Microplate reader	1	1	41.325	41.3
51.	TG4	Microplate reader	1	1	94.084	94.1
52.	TG5	Microplate reader	1	1	161.157	161.2
53.	TG6	Microplate reader	1	1	194.18	194.2
54.	TG7	Microplate reader	1	1	86.784	86.8

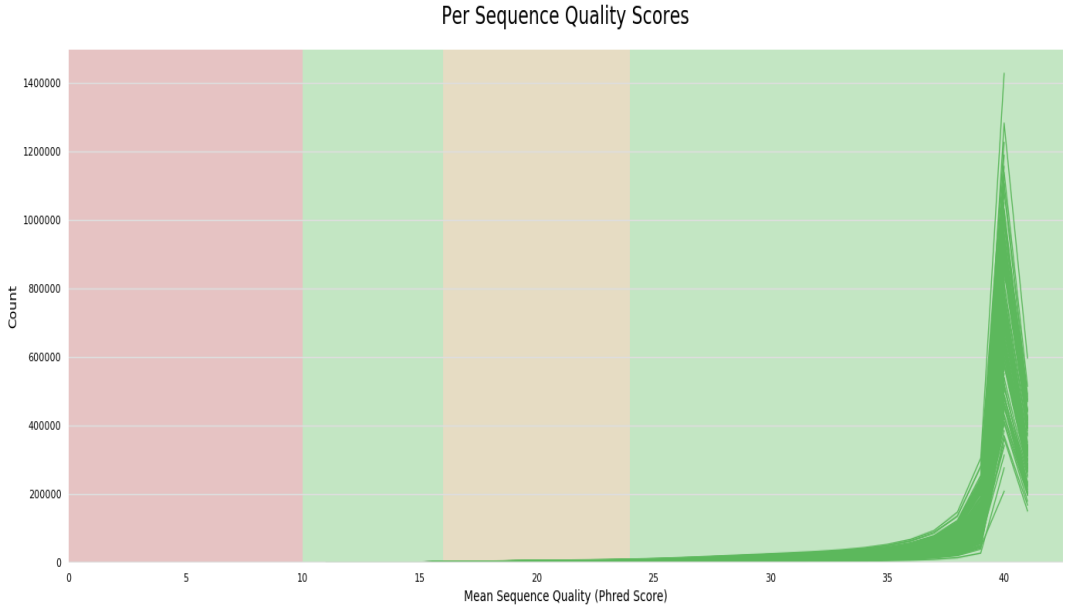
55.	TG8	Microplate reader	1	1	74.043	74
56.	TG9	Microplate reader	1	1	152.111	152.1
57.	TG10	Microplate reader	1	1	250.68	250.7
58.	TG11	Microplate reader	1	1	90.042	90
59.	TG12	Microplate reader	1	1	130.434	130.4
60.	TG13	Microplate reader	1	1	122.173	122.2
61.	TG14	Microplate reader	1	1	112.794	112.8
62.	TG15	Microplate reader	1	1	201.447	201.4
63.	MK1	Microplate reader	1	1	59.855	59.9
64.	MK2	Microplate reader	1	1	103.271	103.3
65.	MK3	Microplate reader	1	1	108.683	108.7
66.	MK4	Microplate reader	1	1	136.191	136.2
67.	MK5	Microplate reader	1	1	85.485	85.5
68.	MK6	Microplate reader	1	1	106.476	106.5
69.	MK7	Microplate reader	1	1	53.433	53.4
70.	MK8	Microplate reader	1	1	275.695	275.7
71.	R1	Microplate reader	1	1	208.138	208.1
72.	R2	Microplate reader	1	1	208.888	208.9
73.	R3	Microplate reader	1	1	302.491	302.5
74.	R4	Microplate reader	1	1	275.187	275.2
75.	R5	Qubit	1	1	252	252
76.	R6	Microplate reader	1	1	279.453	279.5
77.	R7	Microplate reader	1	1	77.524	77.5
78.	R8	Microplate reader	1	1	55.876	55.9
79.	R9	Microplate reader	1	1	63.002	63
80.	R10	Microplate reader	1	1	99.896	99.9
81.	R11	Microplate reader	1	1	69.974	70
82.	R12	Microplate reader	1	1	32.336	32.3
83.	R13	Microplate reader	1	1	76.132	76.1
84.	R14	Qubit	1	1	272	272
85.	R15	Qubit	1	1	238	238
86.	R16	Microplate reader	1	1	192.029	192
87.	R17	Qubit	1	1	250	250
88.	R19	Qubit	1	1	386	386
89.	R20	Microplate reader	1	1	126.39	126.4
90.	P1	Microplate reader	1	1	86.052	86.1
91.	P2	Microplate reader	1	1	136.486	136.5
92.	P3	Microplate reader	1	1	79.985	80
93.	P4	Qubit	1	1	123	123
94.	AA13	Qubit	1	1	134	122
95	AA14	Qubit.	1	1	160	134
96	R18	Qubit	1	1	150.3	145

Appendix VII: Data statistics of clean GBS data

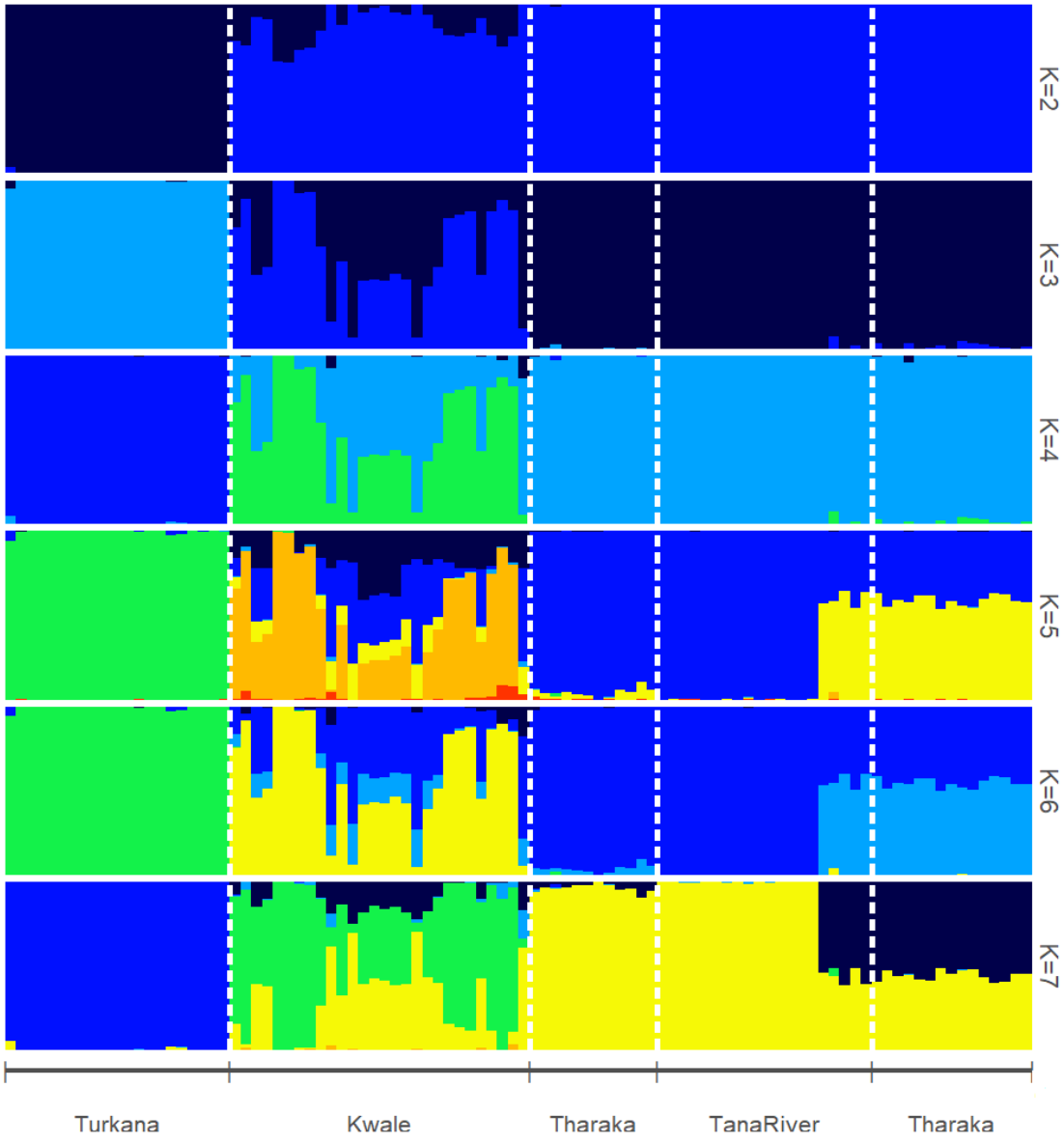
Sample name	Read number (M)	Base number (Mb)	GC (%)	Q20 (%)	Q30 (%)
AA13	0.7	68.67	51.34	98.04	95.2
AA14	1.74	170.61	52.48	98.05	95.2
K16	2.32	225.88	49.53	98.01	95.21
K17	3.96	386.1	49.76	98.07	95.32
K18	2.65	257.95	49.31	98.02	95.22
K19	3.32	324.14	49.7	98.08	95.34
KG1	3.79	369.15	48.6	98.09	95.39
KG10	2.73	265.96	49.02	98.04	95.27
KG11	3.31	323.05	48.47	98.08	95.37
KG12	2.56	249.86	48.9	98.11	95.41
KG13	1.88	182.84	48.08	97.97	95.12
KG14	3.25	316.76	49.38	98.09	95.37
KG15	3.3	321.72	49.46	98.08	95.35
KG2	2.97	289.22	48.37	98.05	95.31
KG3	1.78	173.08	48.19	98.05	95.3
KG4	2.42	235.88	48.36	98.02	95.24
KG5	2.55	249.01	48.02	98.03	95.29
KG6	1.66	161.69	48.62	98.05	95.3
KG7	2.2	214.07	48.84	98.01	95.2
KG8	2.29	222.84	48.61	98.08	95.37
KG9	3.6	351.38	49.14	98.07	95.33
MM10	2.2	215.61	50.41	98.07	95.31
MB1	1.78	172.77	50.69	98.02	95.19
MB10	3.43	332.63	51.4	98.04	95.22
MB2	3.4	330.22	53.01	98.03	95.15
MB3	4	387.68	51.95	97.96	95.02
MB4	3.25	315.18	51.16	98.02	95.2
MB5	2.65	257.22	49.62	98.11	95.39
MB6	3.7	359.27	51.74	98.04	95.22
MB7	2.76	267.79	50.93	98.07	95.29
MB8	3.35	324.83	52.22	98.01	95.13
MB9	1.68	162.69	50.6	98.01	95.17
MK1	2.33	225.27	51.02	98.06	95.26
MK2	1.6	154.44	49.04	98.05	95.29
MK3	2.57	247.79	50.67	98.1	95.35
MK4	2.26	217.98	49.86	98.04	95.24
MK5	2.82	271.78	51.22	98.01	95.16
MK6	2.89	278.73	51.21	97.79	94.61
MK7	2.28	220.03	50.65	98.05	95.25
MK8	1.76	169.97	51.56	98	95.13
MM1	2.74	267.34	50.68	98.03	95.22
MM2	1.79	174.35	50	98.05	95.26
MM3	2.73	266.23	50.37	98.1	95.37
MM4	3.72	361.31	51.88	98.08	95.29
MM5	3.51	340.94	52.28	97.99	95.11
MM6	4.74	460.17	49.93	98.06	95.29
MM7	4.18	405.94	52.61	98.02	95.16
MM8	3.4	329.44	51.37	98.03	95.2
MM9	3.73	361.33	50.06	97.99	95.14
MW1	1.91	187.24	50.75	98.03	95.23
MW2	2.08	204.11	48.2	98.1	95.42
MW3	1.04	101.51	51.12	98.12	95.39
MW4	1.15	112.72	50.31	98.06	95.26
MW5	1.28	125.18	49.55	98.06	95.3
MW6	1.99	194.69	50.25	97.99	95.14
MW7	1.53	149.97	49.45	98.07	95.34
MW8	1.95	190.95	48.96	98.08	95.34

P1	2.12	203.39	49.01	98.12	95.42
P2	1.18	108.51	49.22	99.22	97.45
P3	2.16	207.5	48.52	98.13	95.45
P4	1.81	174.15	48.64	98.06	95.32
R1	2.58	249.16	49.59	98.05	95.26
R10	1.53	146.52	49.09	98.06	95.31
R11	1.9	182.36	49.29	98.04	95.25
R12	1.49	142.7	50.65	97.99	95.12
R13	2.66	255.05	50.75	98.04	95.23
R14	2.41	231.82	49.52	98.01	95.2
R15	2.65	254.19	49.05	98.07	95.31
R16	1.35	130.02	49.05	98.08	95.33
R17	2.55	244.33	51.16	98.02	95.17
R18	3.17	307.54	49.41	98.06	95.29
R19	2.82	271.03	49.22	97.75	94.57
R2	2.34	226.04	49.33	98.07	95.34
R20	2.71	259.84	49.73	98.06	95.3
R3	2.2	211.9	49.49	98.05	95.27
R4	2.49	240.15	48.7	98.05	95.3
R5	2	192.24	49.68	97.98	95.14
R6	2.39	229.83	50.03	97.91	94.97
R7	2.12	203.27	50.05	98.05	95.27
R8	1.58	151.52	49.3	98.01	95.21
R9	1.71	164.24	50.13	98.05	95.25
TG1	2.67	258.83	49.5	97.95	95.06
TG10	1.73	166.59	48.18	98.05	95.3
TG11	2.35	227.24	48.67	97.9	94.96
TG12	3.19	307.61	49.5	98.04	95.26
TG13	3.35	323.31	50.02	98.1	95.37
TG14	2.57	247.78	48.82	98.03	95.26
TG15	1.3	125.46	47.67	97.74	94.61
TG2	3.42	331.87	48.96	98.03	95.24
TG3	3.12	303.1	49.2	98.08	95.37
TG4	3.53	342.55	49.5	98.01	95.19
TG5	3.01	291.79	49.79	98.06	95.29
TG6	1.68	161.91	46.88	98.07	95.35
TG7	3.73	359.74	49.56	98.02	95.2
TG8	2.35	226.96	48.44	98.04	95.26
TG9	1.91	183.89	47.15	98.08	95.4

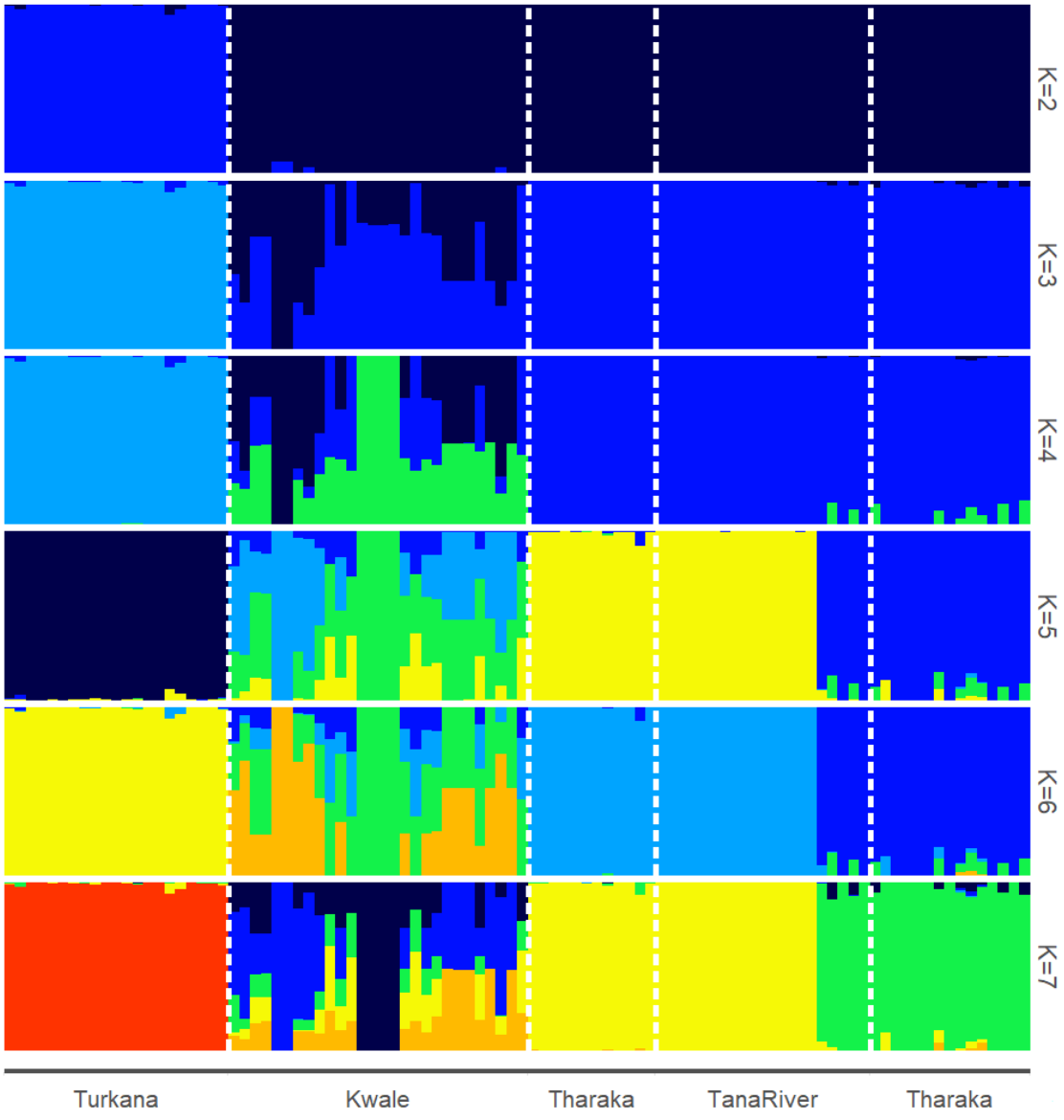
Appendix VIII: GBS sequence quality scores



Appendix IX: STRUCTURE bar plots of K=2 to K=7 for reference based assembly of GBS data

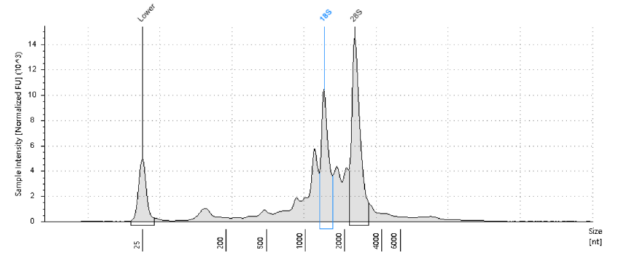


Appendix X: STRUCTURE bar plot of K=2 to K=7 for *de_novo* based assembly of GBS data



Appendix XI: RNA Quality check

A1: TRC11 [1:1 dilution]



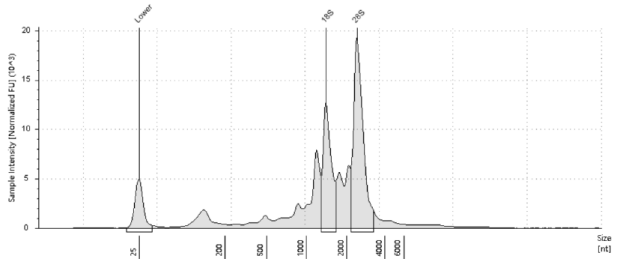
Sample Table

Well	RINe	28S/18S (Area)	Conc. [ng/µl]	Sample Description	Alert	Observations
A1	8.2	1.6	223	TRC11 [1:1 dilution]		

Peak Table

Size [nt]	Calibrated Conc. [ng/µl]	Assigned Conc. [ng/µl]	Peak Molarity [nmol/l]	%Integrated Area	Peak Comment	Observations
25	40.0	40.0	4710	-		Lower Marker
1402	43.4	-	91.1	38.33		18S
2423	69.9	-	84.8	61.67		28S

B1: TRC12 [1:1 dilution]



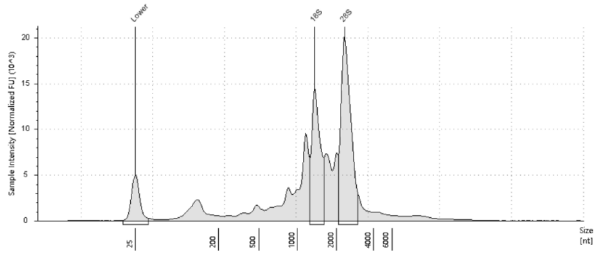
Sample Table

Well	RINe	28S/18S (Area)	Conc. [ng/µl]	Sample Description	Alert	Observations
B1	8.1	1.8	284	TRC12 [1:1 dilution]		

Peak Table

Size [nt]	Calibrated Conc. [ng/µl]	Assigned Conc. [ng/µl]	Peak Molarity [nmol/l]	%Integrated Area	Peak Comment	Observations
25	40.0	40.0	4710	-		Lower Marker
1400	54.5	-	114	36.06		18S
2405	96.6	-	118	63.94		28S

C1: TRC21 [1:1 dilution]



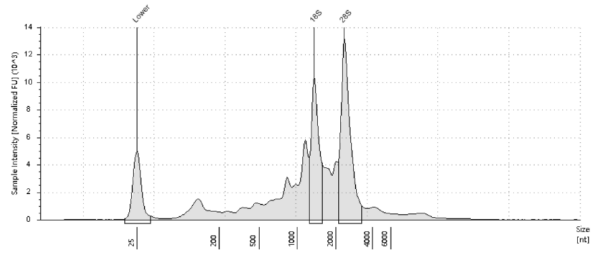
Sample Table

Well	RINe	28S/18S (Area)	Conc. [ng/µl]	Sample Description	Alert	Observations
C1	7.5	1.5	341	TRC21 [1:1 dilution]		RINe edited

Peak Table

Size [nt]	Calibrated Conc. [ng/µl]	Assigned Conc. [ng/µl]	Peak Molarity [nmol/l]	%Integrated Area	Peak Comment	Observations
25	40.0	40.0	4710	-		Lower Marker
1356	65.2	-	141	39.22		18Sedited
2312	101	-	129	60.78		28Sedited

D1: TR311



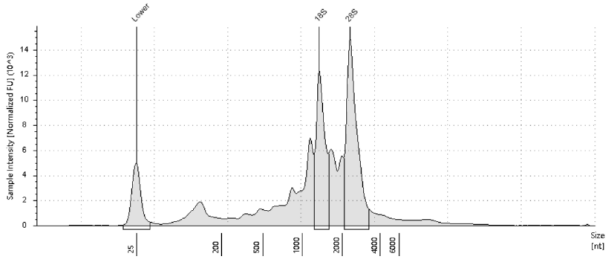
Sample Table

Well	RINe	28S/18S (Area)	Conc. [ng/µl]	Sample Description	Alert	Observations
D1	7.1	1.6	227	TR311		RINe edited

Peak Table

Size [nt]	Calibrated Conc. [ng/µl]	Assigned Conc. [ng/µl]	Peak Molarity [nmol/l]	%Integrated Area	Peak Comment	Observations
25	40.0	40.0	4710	-		Lower Marker
1357	40.1	-	87.0	39.03		18Sedited
2326	62.7	-	79.2	60.97		28Sedited

E1: TR322



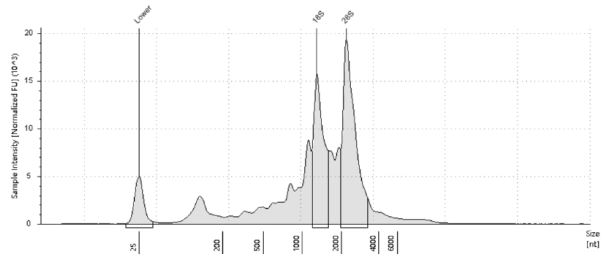
Sample Table

Well	RINe	28S/18S (Area)	Conc. [ng/µl]	Sample Description	Alert	Observations
E1	7.5	1.4	268	TR322		RINe edited

Peak Table

Size [nt]	Calibrated Conc. [ng/µl]	Assigned Conc. [ng/µl]	Peak Molarity [nmol/l]	%Integrated Area	Peak Comment	Observations
25	40.0	40.0	4710	-		Lower Marker
1344	52.0	-	114	40.95		18Sedited
2306	75.0	-	95.7	59.05		28Sedited

F1: TR323



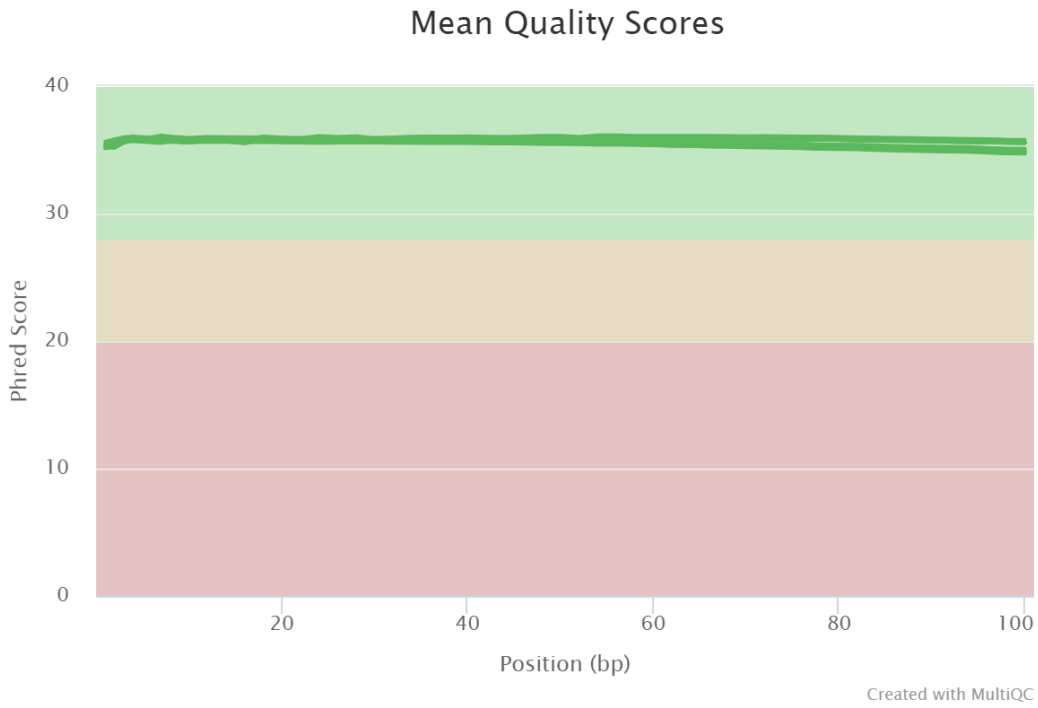
Sample Table

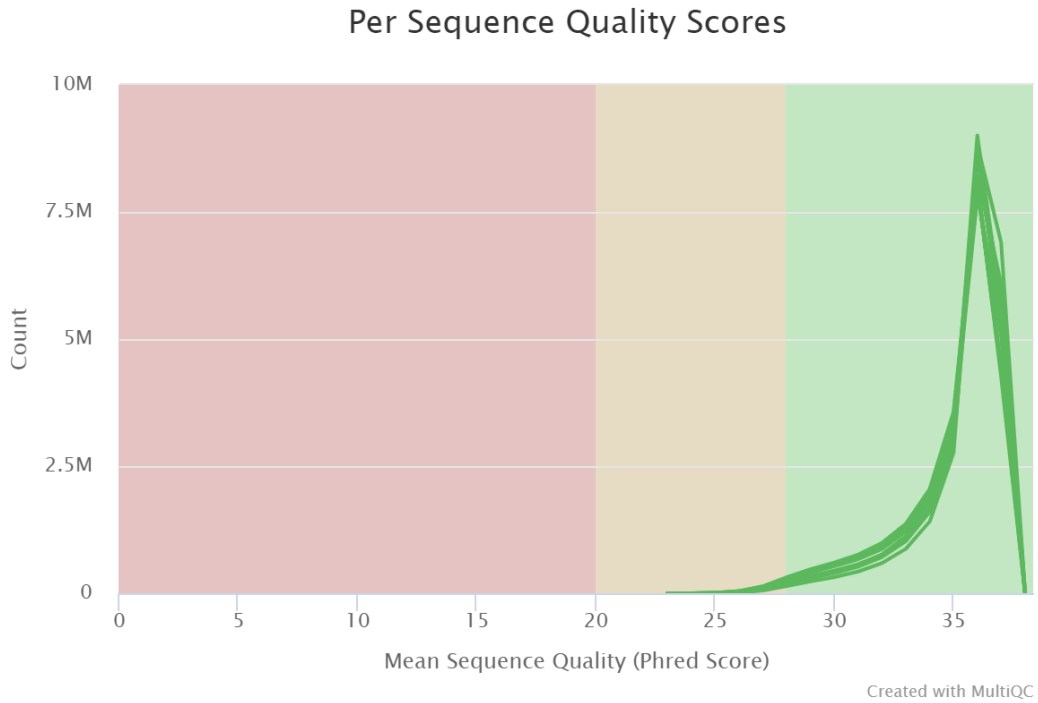
Well	RINe	28S/18S (Area)	Conc. [ng/µl]	Sample Description	Alert	Observations
F1	7.3	1.5	356	TR323		RINe edited

Peak Table

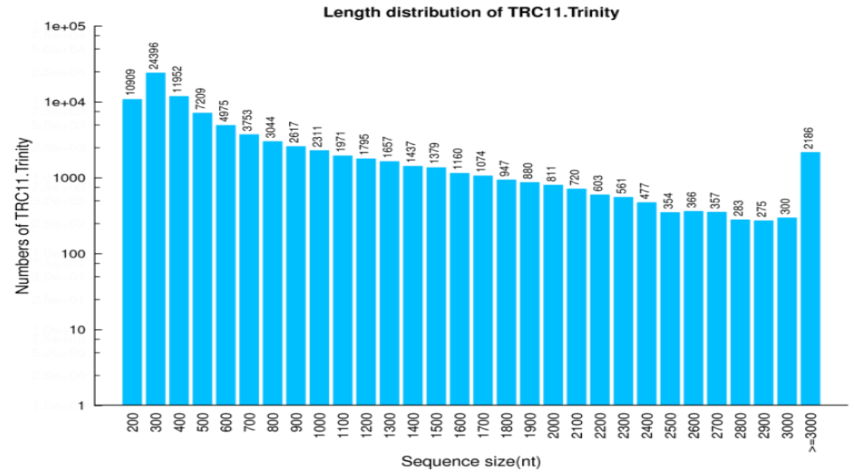
Size [nt]	Calibrated Conc. [ng/µl]	Assigned Conc. [ng/µl]	Peak Molarity [nmol/l]	%Integrated Area	Peak Comment	Observations
25	40.0	40.0	4710	-		Lower Marker
1302	74.1	-	167	40.09		18Sedited
2193	111	-	148	59.91		28Sedited

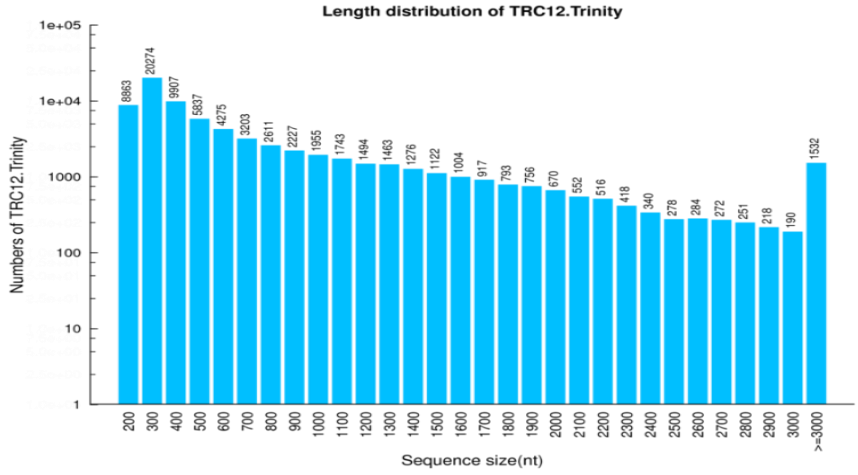
Appendix XII: MultiQC results of RNA sequence data

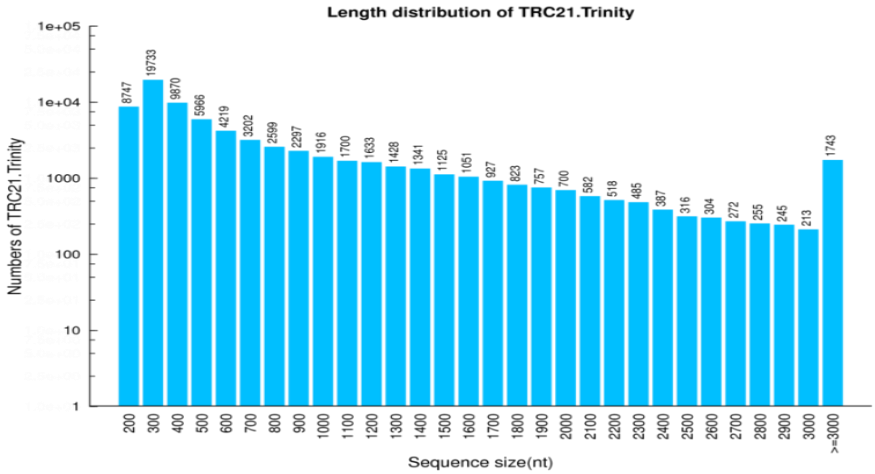


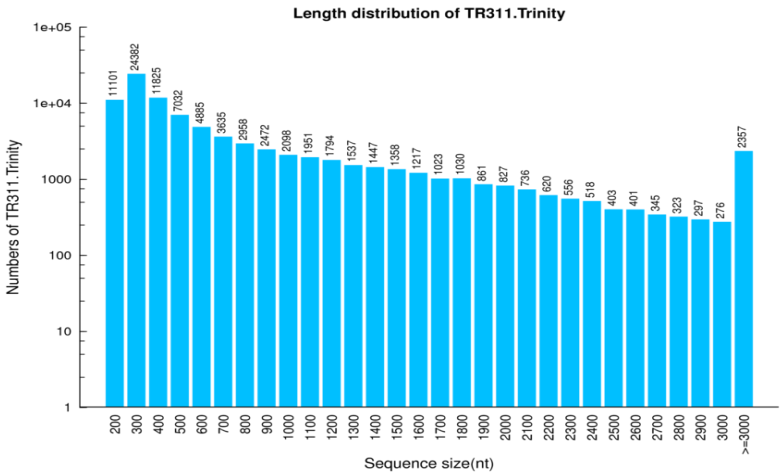


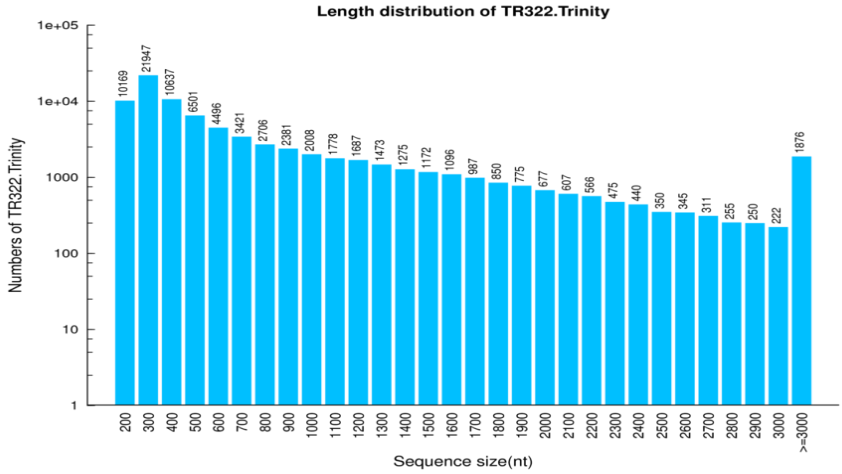
Appendix XIII: Transcript length distribution after RNA-Seq assembly

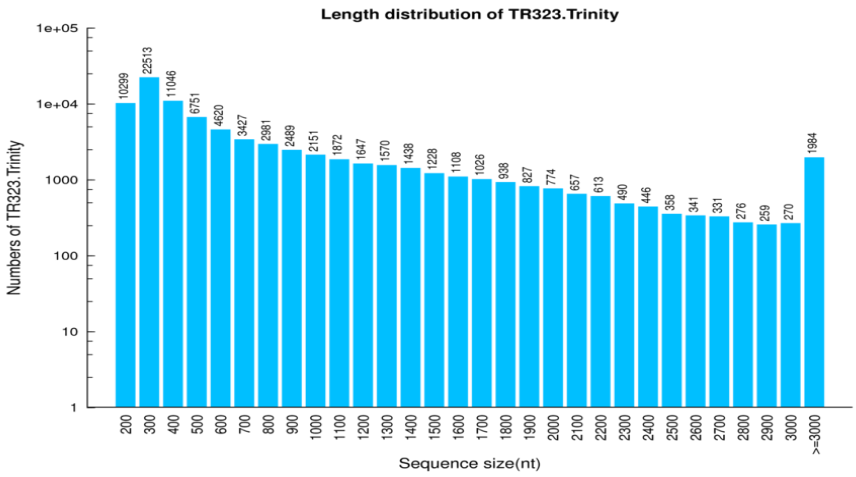







Length distribution of TR311.Trinity
Numbers of TR311.Trinity
Sequence size(nt)
1e+05
1e+04
1000
100
10
1
11101 24382 11825 7032 4885 3635 2958 2472 2098 1951 1794 1537 1447 1358 1217 1023 1030 861 827 736 620 556 518 403 401 345 323 297 276 2357
200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000 2100 2200 2300 2400 2500 2600 2700 2800 2900 3000 >=3000


Length distribution of TR322.Trinity
Numbers of TR322.Trinity
Sequence size(nt)
1e+05
1e+04
1000
100
10
1
10169 21947 10637 6501 4496 3421 2706 2381 2008 1778 1687 1473 1275 1172 1096 987 850 775 677 607 566 475 440 350 345 311 255 250 222 1876
200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000 2100 2200 2300 2400 2500 2600 2700 2800 2900 3000 >=3000


Length distribution of TR323.Trinity
Numbers of TR323.Trinity
Sequence size(nt)
1e+05
1e+04
1000
100
10
1
10299 22513 11046 6751 4620 3427 2981 2489 2151 1872 1647 1570 1438 1228 1108 1026 938 827 774 657 613 490 446 358 341 331 276 259 270 1984
200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000 2100 2200 2300 2400 2500 2600 2700 2800 2900 3000 >=3000

Appendix XIV: SSR repeat motifs and their copy number obtained from de_novo assembled *H. compressa* transcripts

Copy Number	Mono-nucleotide	Di-nucleotide	Tri-nucleotide	Quad-nucleotide	Penta-nucleotide	Hexa-nucleotide
4	0	0	0	0	101	212
5	0	0	2,474	191	39	25
6	0	1,848	1,156	78	3	9
7	0	1,225	495	11	1	1
8	0	985	300	8	0	12
9	0	797	46	10	1	0
10	0	649	48	3	0	1
11	0	467	13	2	0	0
12	1,060	429	15	2	0	2
13	714	61	12	0	0	0
14	592	157	8	1	0	0
15	424	116	4	0	0	0
16	293	124	2	0	1	0
17	204	88	2	0	0	0
18	129	92	1	0	0	0
19	83	95	1	0	0	0
20	66	78	1	0	0	0
21	53	56	0	0	0	0
22	37	38	1	0	0	0
23	73	36	1	0	0	0
24	24	24	1	0	0	0
25	9	16	2	0	0	0
26	6	16	2	0	0	0
27	17	9	0	0	0	0
28	18	3	0	0	0	0
29	10	9	3	0	0	0
30	16	4	0	0	0	0
31	1	2	0	0	0	0
32	11	2	0	0	0	0
33	7	0	0	0	0	0
34	8	1	0	0	0	0
35	2	0	0	0	0	0
36	3	0	0	0	0	0
37	7	0	0	0	0	0
38	3	0	0	0	0	0
39	2	0	0	0	0	0
40	1	0	0	0	0	0
41	3	1	0	0	0	0

42	2	0	0	0	0	0
44	1	0	0	0	0	0
45	2	0	0	0	0	0
46	1	0	0	0	0	0
47	1	0	0	0	0	0
48	1	0	0	0	0	0
49	5	0	0	0	0	0
51	4	0	0	0	0	0
54	2	0	0	0	0	0
57	1	0	0	0	0	0
59	3	0	0	0	0	0
65	1	0	0	0	0	0
78	1	0	0	0	0	0
95	1	0	0	0	0	0
SubTotal	**3,902**	**7,428**	**4,588**	**306**	**146**	**262**

Appendix XV: Perfect and compound SSR markers in non-coding regions with no overlap between SSR and ORF.

scaffold no.	SSR Type	Repeat sequence	Start of SSR	End of SSR	Start of ORF	End of ORF
CL4706_Contig3_All	p1	(T)14	1223	1236	3	764
CL4890_Contig1_All	p1	(T)12	2154	2165	237	2015
CL3291_Contig1_All	p3	(AGA)5	899	913	168	785
Unigene22342_All	p1	(T)13	504	516	2	301
Unigene28081_All	p4	(TTTC)5	886	905	2	346
Unigene22190_All	p2	(TC)12	399	422	22	390
CL7369_Contig2_All	p1	(A)19	761	779	91	294
Unigene16829_All	p1	(T)14	59	72	10	159
Unigene10074_All	p2	(GA)7	582	595	270	470
Unigene11078_All	p1	(A)13	467	479	5	421
CL4552_Contig1_All	p1	(T)14	1388	1401	1	1254
Unigene198_All	p3	(GAG)6	469	486	1	351
CL7355_Contig1_All	p1	(T)16	1294	1309	2	1111
Unigene6750_All	p2	(AT)10	1488	1507	501	1328
Unigene25844_All	p2	(AG)7	59	72	2	130
Unigene329_All	p1	(T)16	2040	2055	1147	1608
CL1662_Contig1_All	p2	(GA)11	3134	3155	342	1625
Unigene19221_All	p2	(TA)6	409	420	57	290
Unigene19221_All	p1	(T)12	550	561	57	290
CL2934_Contig2_All	p1	(A)14	2281	2294	20	1639
CL10390_Contig1_All	p1	(T)12	859	870	45	626
Unigene3425_All	p1	(T)12	1164	1175	246	506
Unigene7062_All	p2	(CT)24	578	625	45	347
Unigene7062_All	p3	(ACC)5	818	832	45	347
Unigene15997_All	p2	(CT)20	939	978	381	632
CL5765_Contig2_All	p1	(C)21	1813	1833	2	1672
CL8536_Contig3_All	p2	(AG)11	2108	2129	57	464
CL4592_Contig1_All	p1	(T)14	1461	1474	120	1355
CL6942_Contig2_All	p1	(T)13	1089	1101	36	941
Unigene11136_All	c	(TAC)10(AAC)6	35	82	2	112
Unigene10446_All	p3	(TCT)5	952	966	13	759
Unigene21657_All	p2	(TA)10	797	816	214	762
Unigene21657_All	p1	(T)13	1437	1449	214	762
CL2138_Contig2_All	p2	(TG)9	2838	2855	1	2562
Unigene91_All	p1	(T)12	1609	1620	1	1512
CL3227_Contig1_All	p1	(T)14	1355	1368	156	1160
Unigene3909_All	p1	(T)18	635	652	238	480

CL8367_Contig1_All	p3	(GAG)5	991	1005	3	959
CL514_Contig2_All	p3	(GAG)6	4013	4030	3	3743
CL7618_Contig1_All	p3	(GCT)5	667	681	3	575
CL9942_Contig2_All	p2	(TC)8	1441	1456	113	1426
CL5094_Contig1_All	p3	(AGA)5	836	850	2	640
Unigene26191_All	p3	(AAG)7	728	748	3	668
CL4394_Contig3_All	p5	(TTTTC)4	2827	2846	187	2619
Unigene19171_All	p2	(AG)6	1542	1553	1	933
CL5663_Contig1_All	p1	(T)32	2123	2154	517	2082
Unigene13114_All	p1	(T)12	1218	1229	3	1151
CL5451_Contig1_All	p1	(T)12	1423	1434	1	1086
CL9005_Contig3_All	p3	(GCA)6	57	74	43	300
CL293_Contig2_All	p3	(GGC)7	97	117	1	498
Unigene24416_All	p1	(C)15	244	258	89	289
CL8404_Contig2_All	p2	(CT)9	1085	1102	106	435
CL4917_Contig2_All	p1	(A)13	612	624	9	533
Unigene3800_All	p2	(GA)9	344	361	172	345
Unigene9522_All	p2	(TA)16	647	678	2	427
CL5793_Contig2_All	p5	(AAAAT)4	1650	1669	303	1523
CL10189_Contig5_All	p4	(AAAT)5	1610	1629	3	1595
Unigene17668_All	p1	(T)13	717	729	200	700
CL3416_Contig1_All	p2	(AG)18	1616	1651	2	1303
Unigene13005_All	p1	(A)13	706	718	1	600
CL4717_Contig1_All	p1	(T)13	1181	1193	69	899
Unigene10812_All	c	(AG)8catcgagtcatgtgaagtc taaaaAggtgaaggaatagtggcta aaaggaaagtgagatttaccctctcgc tcgcttgctctctctctt(TC)6	64	180	125	424
CL1103_Contig1_All	c	(A)13gggtttttttatttgaatttttgg ctgTttggtagggagcattggatatc taccggaaaggg(A)12	2006	2091	374	1774
CL7761_Contig1_All	p2	(CT)6	2510	2521	62	2230
Unigene30507_All	p2	(AG)8	74	89	72	182
CL343_Contig1_All	p1	(T)13	926	938	70	756
Unigene4416_All	p1	(T)13	125	137	2	199
Unigene25110_All	c	(CT)6gatttgctgttgttattgtttt tacttctgtGcaataggtgcgccta tttatgtgcctaaaccttaatcagattt ggctggctacgcc(T)13	1964	2076	424	1539
Unigene44647_All	p1	(A)15	317	331	241	456
Unigene22012_All	p1	(T)17	2102	2118	78	1778
Unigene32_All	p1	(G)14	1655	1668	447	875
CL41_Contig2_All	p2	(TC)6	896	907	202	420
CL1451_Contig3_All	p1	(A)15	1242	1256	1	1080
CL3632_Contig1_All	p2	(TC)12	725	748	86	700
CL5187_Contig1_All	p1	(C)13	1332	1344	287	1225
Unigene13600_All	p2	(AT)7	1012	1025	36	350

Unigene28917_All	p2	(GA)6	974	985	15	878
Unigene22674_All	p3	(CTT)7	1297	1317	12	1250
CL10370_Contig2_All	p2	(CT)9	36	53	2	184
CL1312_Contig5_All	p1	(A)13	1288	1300	28	759
CL3171_Contig2_All	p3	(ATG)5	1243	1257	110	658
CL5483_Contig1_All	p1	(A)16	118	133	2	151
CL4189_Contig3_All	p2	(GA)14	395	422	17	259
Unigene18958_All	p1	(T)14	2626	2639	2	1768
Unigene1123_All	p2	(CT)6	1425	1436	1	1068
CL4434_Contig6_All	p2	(AT)9	1066	1083	527	892
CL10308_Contig1_All	c	(AGC)5atcattc(GAT)7	1243	1285	1	714
CL8659_Contig2_All	p1	(T)12	1639	1650	298	1599
Unigene20118_All	c	(TC)6tttattgtgagtataaaaa cttgtattTgtgtgtttatgttttctc ccattttaatgtctcgcttc(CT)8	530	623	1	435
CL1023_Contig1_All	p2	(TC)10	1237	1256	704	1126
Unigene2327_All	p1	(T)12	332	343	23	220
CL7287_Contig1_All	p3	(TTC)5	1847	1861	530	1795
CL7243_Contig1_All	c	(TC)7ttttttttt(TG)6ttcatt Cccataactaatgcttcagcagct atattac(CT)15	3024	3126	122	835
CL4757_Contig1_All	p2	(TC)7	1482	1495	328	1308
CL2752_Contig1_All	p1	(A)12	931	942	179	928
Unigene13417_All	p1	(T)19	3846	3864	312	3755
Unigene6765_All	p1	(T)14	1792	1805	124	1782
CL7073_Contig1_All	p1	(A)12	1742	1753	2	775
Unigene3813_All	p2	(TA)6	2411	2422	287	2323
CL8080_Contig1_All	p3	(AGA)7	417	437	1	342
Unigene2533_All	c	(T)12aaatcactgctctacttc ttat(TC)6	557	602	152	433
CL6433_Contig1_All	p2	(AG)8	969	984	3	560
CL24_Contig4_All	p5	(TTATT)4	4198	4217	196	3780
CL2949_Contig2_All	p2	(TC)10	1008	1027	290	808
Unigene19057_All	p2	(GA)12	1450	1473	128	1441
CL5056_Contig1_All	p2	(CT)10	1750	1769	68	1699
Unigene19105_All	p2	(TC)6	585	596	64	432
CL5375_Contig1_All	p1	(T)16	2470	2485	98	1951
CL4416_Contig2_All	p1	(A)15	929	943	3	785
CL1183_Contig1_All	p1	(A)12	1043	1054	1	879
CL1183_Contig1_All	p1	(A)13	1263	1275	1	879
CL7072_Contig1_All	p1	(A)12	1583	1594	120	971
Unigene18919_All	p3	(TTG)5	1263	1277	295	1239
CL8637_Contig2_All	p1	(A)18	992	1009	42	941
Unigene25305_All	p1	(T)15	1110	1124	208	996
Unigene6656_All	p1	(A)12	834	845	121	612
Unigene27803_All	p2	(TG)9	906	923	496	813

Unigene9627_All	p1	(T)16	1407	1422	107	1369
CL1953_Contig1_All	p1	(T)12	1022	1033	3	893
CL10131_Contig1_All	p1	(A)19	921	939	1	885
CL2440_Contig4_All	c	(CT)10ccacccgcacccca (C)12	3432	3477	404	2842
CL4155_Contig2_All	p1	(T)14	1822	1835	1	1617
Unigene961_All	p1	(T)13	289	301	3	230
Unigene22714_All	p1	(T)13	1271	1283	21	941
Unigene13118_All	p1	(T)12	1586	1597	68	1330
Unigene6635_All	p1	(A)15	1694	1708	279	1394
CL7452_Contig3_All	p1	(A)15	895	909	1	588
CL2805_Contig2_All	p2	(CT)9	484	501	219	422
CL6329_Contig1_All	p2	(CT)7	1173	1186	109	1086
CL10394_Contig1_All	p3	(TTC)5	2248	2262	561	2117
CL9216_Contig1_All	p3	(TGT)5	864	878	1	546
CL1073_Contig2_All	p3	(AAG)5	2187	2201	2	2122
CL951_Contig1_All	p2	(AT)16	4977	5008	403	4854
Unigene10862_All	p4	(AGAT)5	444	463	3	431
Unigene10516_All	p1	(A)17	534	550	3	434
CL6599_Contig4_All	p2	(GA)11	2501	2522	632	2197
CL10102_Contig1_All	p2	(TA)9	1229	1246	129	1130

Appendix XVI: Perfect SSR Markers (excluding mononucleotides) in coding regions that have at least 30bp overlap between SSR and ORF.

Scaffold name/ORF NO.	SSR	Repeat sequence	BP start of SSR	Bp end of SSR	Bp start of ORF	Bp end of ORF	Length of overlap between ORF and SSR in bp
Unigene13547_All	p2	(GA)16	46	77	2	1183	31
Unigene21622_All	p3	(TCC)12	250	285	190	480	35
CL7766_Contig2_All	p6	(ACGGCA)6	1075	1110	1	1518	35
CL7774_Contig3_All	p3	(GGC)11	1130	1162	710	2167	32
CL8644_Contig2_All	p2	(AG)18	380	415	349	1821	35
CL5065_Contig1_All	p2	(TC)27	285	338	191	490	53
Unigene6864_All	p2	(TC)24	373	420	171	461	47
Unigene2041_All	p2	(TC)17	226	259	58	294	33
Unigene13076_All	p2	(CT)23	185	230	79	858	45
CL8196_Contig1_All	p2	(CT)19	545	582	2	586	37
CL7355_Contig1_All	p2	(AG)23	44	89	2	1111	45
CL721_Contig1_All	p2	(TA)21	31	72	1	207	41
CL8747_Contig2_All	p3	(AAC)13	205	243	31	1209	38
Unigene581_All	p2	(GA)21	311	352	250	2187	41
CL4196_Contig5_All	p2	(CT)20	37	76	30	2639	39
CL7962_Contig1_All	p2	(GA)22	38	81	40	624	41
Unigene9717_All	p2	(CT)18	261	296	22	345	35
Unigene28154_All	p2	(TC)21	33	74	2	235	41
CL3255_Contig5_All	p2	(CT)20	161	200	115	801	39
CL1023_Contig5_All	p6	(CGGGAC)6	1133	1168	329	1615	35
CL10039_Contig1_All	p2	(TC)16	193	224	158	436	31
CL3533_Contig4_All	p2	(CT)25	222	271	2	1810	49
Unigene10690_All	p2	(TC)16	41	72	1	231	31
Unigene18971_All	p2	(CT)20	589	628	488	673	39
CL1732_Contig4_All	p2	(GA)17	1506	1539	858	1628	33
CL9942_Contig2_All	p2	(TA)16	316	347	113	1426	31
CL4542_Contig3_All	p2	(CT)21	47	88	3	458	41
Unigene4422_All	p2	(TC)20	53	92	44	742	39
Unigene16620_All	p2	(GA)19	155	192	2	388	37
Unigene24283_All	p2	(AG)18	29	64	25	291	35
Unigene28929_All	p2	(CT)21	337	378	186	404	41
CL2983_Contig2_All	p2	(GA)21	46	87	3	590	41
CL9463_Contig1_All	p2	(CT)17	440	473	419	2110	33
CL4525_Contig1_All	p2	(AG)21	350	391	74	580	41

CL712_Contig2_All	p2	(GA)17	194	227	61	1854	33
Unigene9768_All	p2	(CT)19	184	221	35	1279	37
CL453_Contig1_All	p6	(CCGTCG)7	106	147	1	585	41
CL495_Contig2_All	p3	(CTC)17	925	975	11	1225	50
Unigene8438_All	p2	(AG)18	45	80	1	300	35
CL2260_Contig8_All	p2	(TC)22	269	312	1	1110	43
CL1627_Contig1_All	p2	(AG)19	37	74	2	661	37
CL1007_Contig3_All	p2	(CT)20	32	71	2	298	39
CL2680_Contig7_All	p2	(CA)17	434	467	345	701	33
Unigene22100_All	p2	(AC)17	371	404	260	499	33
CL10108_Contig1_All	p2	(GA)19	37	74	3	860	37
CL4342_Contig1_All	p2	(AG)18	607	642	563	3160	35
Unigene9740_All	p2	(CT)16	329	360	213	2237	31
CL7576_Contig1_All	p2	(CT)21	23	64	1	606	41
Unigene13086_All	p2	(CT)16	232	263	77	1174	31
CL7287_Contig1_All	p6	(GCATCG)8	626	673	530	1795	47
Unigene31089_All	p2	(GA)20	163	202	2	535	39
CL4979_Contig1_All	p2	(AG)17	51	84	1	483	33
Unigene13631_All	p2	(CT)16	49	80	1	741	31
Unigene19848_All	p2	(AG)17	459	492	3	515	33
CL5219_Contig1_All	p2	(GA)16	60	91	1	972	31
CL624_Contig1_All	p2	(CT)23	306	351	29	349	43
Unigene13969_All	p3	(CAT)11	211	243	80	277	32
Unigene15959_All	p2	(CT)20	44	83	2	892	39
Unigene10712_All	p2	(CT)18	103	138	2	277	35
Unigene9784_All	p3	(ATC)12	642	677	603	851	35
Unigene40996_All	p6	(CCTCAA)8	880	927	1	1143	47
CL5204_Contig1_All	p2	(TC)26	40	91	20	352	51
CL1246_Contig5_All	p2	(TC)17	52	85	3	1316	33
Unigene27058_All	p2	(CT)18	46	81	1	252	35
Unigene3155_All	p2	(TC)18	177	212	2	346	35
CL2672_Contig2_All	p3	(GCA)11	1511	1543	216	1673	32
Unigene39743_All	p2	(TC)18	153	188	53	220	35
Unigene20685_All	p2	(AG)16	575	606	346	639	31
CL5593_Contig2_All	p3	(CTC)13	611	649	3	1631	38
CL6081_Contig2_All	p3	(GAT)13	935	973	440	1258	38
CL515_Contig3_All	p3	(AGA)11	1706	1738	28	2883	32
CL10088_Contig1_All	p2	(CT)19	218	255	93	299	37
CL8538_Contig1_All	p2	(AG)16	54	85	2	364	31
Unigene28016_All	p3	(CTC)12	337	372	128	1333	35
Unigene1174_All	p2	(AG)17	47	80	2	343	33
Unigene19423_All	p2	(AG)16	302	333	286	1569	31

Appendix XVII: Expected PCR amplification product sizes of the SSR markers designed from unigenes of *H. compressa* transcriptome

Primer	SSR_seq	SSR	Product	Size
DPSSR001	TGTGTG TGTGTG	TG(2*6)	TTCAAGAATATGCATGTCAGCACCTACTA ACACACATAAACATGTATGCACCCGCTCA ATGTGTATT[**TGTGTGTGTGTG**]CATCAG AGAGAGCATTTTGACTTGCCAATTTGGGA AGCAGAAATTCTTGT	129
DPSSR00	GAGAG AGAGA GAGAG AGAGA	GA(2*10)	GTTTTCCACGTTCGTTCTTGTAGGACTTTG ATGTG**[GAGAGAGAGAGAGAGAGAGA**]T TTGAGTGCCTGATGGAAACATTCATGTGT GAAGCCTTTTGAGATTGTCGGTTTGAATT	115
DPSSR003	CTCTCT CTCTCT CTCTCT CT	CT(2*10)	CCATCTCCCACCTCTTTTCTCGCCTCCAC[**CTCTCTCTCTCTCTCTCTCT**]CCCCCTCC ACATGGCTGCCTCTCTACCTGCCTCCTTCC CCCGCCGGCGATCGCCTCCCCCAAATCCG AAGCCCTTATAACAGGAAAAGCAG	142
DPSSR004	TATATA TATATA TATATA	TA(2*9)	GGGAGTAGAGACGGTAAAGCAAGTGTCG CAGCCGGAAGTTCGTCCCAATCTCTCGGT TTTTTTGGACTCCAATACTTGAAGAACAA GTTCTTCTTCCTTTGC[**TAATATATATATA TATA**]TTATATCTATCCGTCTTTTTATTACG TTGGG	151
DPSSR005	TCTCTC TCTCTC TCTCTC TCTCTC	TC(2*12)	ATCAAAAGCAAGTTCACCAACATCCGCTC CGGCCTCTGAACTCCTAT[**TCTCTCTCTCT CTCTCTCTCTCTC**]ATCTCTTAAAATCTA GCTATGGTGATTGTC TCTTGATCGAAGGGGTTTAT	121
DPSSR006	CGGCGG CGGCGG CGGCGG	CGG(3*6)	CGATCCAGCACCAGCTCTAGCGCCAGCGC [**CGGCGGCGGCGGCGGCGG**]CGACTGGT CCTCGTCGTAAAAAGACGACGATATGTAG GAGGCAGCGGCAACTACAGTAGCCGTCG AGA	117
DPSSR007	ATGATG ATGATG ATG	ATG(3*5)	AGGTCTCACTGAGCTGGACATTGGATAGC ACTTCTACAGGA[**ATGATGATGATGATG**] TGGCTGCAAATTTTAAACGGCGTGCAGGA CAGTGGTCTTCTTCCGGACGTCAAAGATC CCTGCAATGAAGTCCCCTGGTGGCAACAA GATTACAAGATGCAT	158
DPSSR008	AAGAA GAAGA AGAAG AAGAA G	AAG(3*7)	TTCAATTTGGGAGGAGCTTCTATGGACAT GAGTAAATTCTTGTCAGTCCACCGCAAGG ATAAATGGACCTACGCTTTAAAAACGTTA AAA[**AAGAAGAAGAAGAAGAAGAAG**]GG GTGCGTGGCGGCGAAGAAGCTTCTCGATT TCATCTTCCGAC	154
DPSSR009	AGCAGC AGCAGC AGC	AGC(3*5)	CTGTTCTGGTGCACCTTTGTAATGGGG[**AG CAGCAGCAGCAGC**]ATTTGGCTGAGGCC CAGGAGCGTGGGGTGATGGGTCCATTGGA GGAACTGGCACAGAGTGGCACAGACAGG GGTAAAAGGAAGG	126

DPSSR010	CACCAC CACCAC CAC	CAC(3*5)	GAAGCTCCTCCAAAGGGTTATCCCGCTTT ACCACCTCCACCAGCCTATGCTCCACCTG CATACGAAGCTCCTGTCTATGCACCACCA GCATACG[**CACCACCACCACCAC**]CAGCT TACGAGCCAACATACCAAGCTCCAG CTTATGA	146
DPSSR011	AAGAA AGAAA GAAAG AAAGA	AAGA(4*5)	CCTCCCCACTCTTCAAATAAAAAAAAAGA AGG[**AAGAAAGAAAGAAAGAAAGA**]CTG ACAGACAGACGCAAAAAAATAAAACTCA GCTTGGCTCGATGACAGATTTGATCTA	110
DPSSR012	AAAGA AAGAA AGAAA GAAAG	AAAG(4*5)	TGGTGAGATAATTGACAGTGCAGAGAACT GGGAAAGGGTGATT[**AAAGAAAGAAAGA AAGAAAG**]AGAAGAAAAATCCACTTAATG TGAAGCAGCTGAAGTTCTATTCATGAATG TTCTTGTTGATTCCATTGC	132
DPSSR013	CCCTCC CTCCCT CCCTCC CT	CCCT(4*5)	AGGTAAAATACCACCGAGCTTGTTCCTCC AAACCCCACCTCTCTCTCCTCTCCTCC[**CC CTCCCTCCCTCCCTCCCT)**CCCGTTCGAA ACCCTAGATTTCGGTTTCGTATCTTGTTTT CTCGTTCTCTAATCCGTTCTTCTG	140
DPSSR014	TTTCTT TCTTTC TTTCTT TC	TTTC(4*5)	AAGTAGAAGGATTGCTGGGAAATTGTTAC AGATTTACTATAAATGTG[**TTTCTTTCTTT CTTTCTTTC**]CCCTCTAAAAGCACAGGTG ATGATCGGATTGACAAATCCTTGCGCCTC CTGTCTGAGTGAATATGCT	134
DPSSR015	AAAATA AAATAA AATAAA AT	AAAAT(5*4)	TTATTCTGGTTTGGTTAGTGGGAGTTTTGG TGTGGTCCGAAAGTTACCAGCAAAAAAA AACCAAAA[**AAAATAAAATAAAATAAAA T**]AAAAAGAAGAAACTTGACCGAAAAAA AAAGGTGGGGAACAAAACAAAGATCAGA TGTTGTGGGTGATCAGTTA	159
DPSSR016	AGCTTA GCTTAG CTTAGC TT	AGCTT(5*4)	ATGTGCTCCTCTCCTTTACCA[**AGCTTAGC TTAGCTTAGCTT**]AGCTAGCTCCTCTTTCT CCTTTTGTTAATA ATTAGTAAAATAAATTACTGTGCT CTCCCATCAC	105
DPSSR017	AACAGA ACAGAA CAGAAC AG	AACAG(5*4)	TTCTTACTTTCCTCCGATTCCTCTACTCCA CCAGGTTTGCTCCACTTTTCTTTTATCAGA AGAAA[**AACAGAACAGAACAGAACAG**]A AATTTCTCTCAAAGAAAAAAAAATTGCAT GATTCTTTGCTTTCGGACGAAGTAGCATC ATTCAGAG	152
DPSSR018	CGAGGC GAGGCG AGGCGA GG	CGAGG(5*4)	CTACCTGCGCTATAGGTCTTCCCCCGCTTT GTTACAGAGTGCGGCAAGGCGAGGCGTC GGCGAGTCTTCTGGGAGGTTCCGAT**[CGA GGCGAGGCGAGGCGAGG**] TGAAAGTGAGGAATCCGGGACTGTGCGTT TTTTCGTCTGATCACTTAAGAGTTCTCG	160
DPSSR019	GCCGTT GCCGTT GCCGTT GCCGTT	GCCGTT(6*4)	GAAAAGCCCAGGTTCATCAGGTTCGGTGC GCCGTC[**GCCGTTGCCGTTGCCGTTGCC GTT**]GCTGGCTTCTGCATAGTGCTTCCTCT CGCCGCCATTTATTTTAATTCTATTTTGCT TGCGTTTCTCCTCTCTCCACTCTCT	141

DPSSR020	CGGGAC CGGGAC CGGGAC CGGGAC CGGGAC CGGGAC	CGGGAC(6* 6)	CGAGATCATTCAGACCGTGAT[**CGGGACC GGGACCGGGACCGGGACCGGGACCGG GAC**]CGGAATCGGGACCTGGATCGGGATC GTGACAGGGAACGAGACAGGAGGCATGA GCATGATAGACGTGGGGATCGTGACGGTA GAGAAAGGGACTCT	153

The SSR sequence is bolded on the product sequence