DISSERTATION

MOLECULAR INVESTIGATIONS IN DATE PALM

GENETIC STRUCTURE AND DIVERSITY AMONG COMMERCIALLY IMPORTANT DATE PALM CULTIVARS (*Phoenix dactylifera* L.)

Submitted by

Salah E. Zaid

Department of Horticulture and Landscape Architecture

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2019

Doctoral Committee:

Advisor: Harrison Hughes

Sarada Krishnan
Mark Brick
Robert Davidson

Copyright by Salah E. Zaid 2019

All Rights Reserved

ABSTRACT

MOLECULAR INVESTIGATIONS IN DATE PALM
GENETIC STRUCTURE AND DIVERSITY AMONG COMMERCIALLY IMPORTANT
DATE PALM CULTIVARS (*Phoenix dactylifera* L.)

The date palm, *Phoenix dactylifera* L. is the notable palm which produces a nutrient-rich edible fruit (the date), well known for its unique attributes of medicine and healthy energy. It is a species that has been cultivated since early civilizations in the fertile crescent and later in the Middle East. It is typically cloned with many cultivars (over 3000). A means of accurately identifying specific clones and an understanding of the relationships among major commercial cultivars would provide valuable information for the maintenance, potentially an improvement and continued conservation of superior genotypes. Phylogenetic relationships amid commercial date cultivars are poorly understood, despite their importance. This research aimed at providing applicable knowledge through an expedient technique, by developing an exclusively tailored Simple Sequence Repeat (SSR) panel, custom-made for date palm fingerprinting and molecular identification also named as 'Dates PalmàPrinting'. This assembled modified genotyping by microsatellite markers provides a standardized approach to cultivar identification and a quality control application in date palm micropropagation production. A deeper understanding and relationship of today's major commercial cultivars is incomplete. Improving the development and productivity of this tree species is restricted due to few genetic resources. Only regionally narrowed studies have been conducted but it is more important to have a broader base of such knowledge.

The present research reports on 20 selected, commercially important date palm cultivars, consisting of 18 females and 2 males, which are grown throughout the world. The knowledge of relationships among these cultivars is needed, although the date palm genome has been mostly sequenced (90.2 %) with 41,660 gene models representing an 82,354 scaffold. The relationships among the major cultivars remain unclear. Presently, the information on the characterization of these cultivars requires an assessment to better understand the relationships among the superior genotypes. The use of microsatellites, due to their accuracy and high polymorphic capability, have led to fine scaled phylogenies. The phylogenetic relationships were determined using neighbor joining un-rooted trees correlated with genetic structure clustering. Primer selections were achieved from evaluation of 14 nuclear SSR loci isolated from *P. dactylifera*. Results revealed a high degree of polymorphism observed in the 20 cultivars with fewer common alleles than anticipated. Within the cultivars studied, a broad heterozygosity across base pair (bp) amplification data has led to an understanding of limited inbreeding, accounting for possible adaptation to environmental changes and revealing conserved extensive array of genomic structure. Population structure analysis suggests a large genetic boundary between Northwest African and Middle Eastern cultivars with 6 subpopulations that represent divergences and fragments of admixture in cultivars present in these regions. The possible selection of potential and good quality parents is achievable for improving cultivars by generating population and structure maps. This analysis documents patterns of relationship and provides genetic structure and diversity of gene pool specificity complexes of date palm cultivars.

This study provides insights about the relationships that exist among cultivars of interest through genetic sequence analysis using SSRs, facilitating the development of a standard approach to identification and enhancements to the micropropagation process.

ACKNOWLEDGEMENTS

The present Ph.D. work has been realized within the framework of Colorado State University jointly undertaken by UAE University Khalifa Center for Genetic Engineering & Biotechnology and University of California Davis Plant Reproductive Biology.

I first wish to express my heartfelt appreciations to Professor Harrison G. Hughes for his unbroken belief in my potential, his continuous support, and invaluable encouragement since my venture in higher-level education.

Dr. Sarada Krishnan deserves my gratitude and indebtedness for her consistent valuable assistance and unremitting reinforcement. Dr. Hassan Shabana and Dr. Samir Al Shaker equally are recognized for their extensive knowledge and confirmation of the work, while assisting without reservation.

Special thanks also go to Professor Eduardo Blumwald, to accept my appreciations and my revered recognition for his guidance and certainly adjusting my bearings. It was an honor to have worked in his lab and enhance my molecular skills, for they have been invaluable. He was always willing and available to help.

I wish to recognize all colleagues and staff of every part of the three institutions that my work had me investigate in, especially the Blumees. They were all very helpful and they will always be remembered. The department of Horticulture and LA, Colorado State University in Fort Collins, Colorado, USA, Dr. Robert Davidson | The Khalifa Center for Genetic Engineering and Biotechnology in Al Ain, UAE | The Eduardo Blumwald Laboratory, University of California Davis, California, USA.

Finally, I would like to distinguish Dr. Abdelouahhab Zaid, Prof., my Father, for standing by me at all times, for providing the most ideal conditions in achieving my goals, and for my treasured education, the most essential gift.

Experiments were conducted at 3 research facilities, Colorado State University's Hughes Laboratory, in Fort Collins, Colorado, USA, under supervision of *Prof. Harrison Hughes*; United Arab Emirates University's Date Palm Tissue Culture Laboratory and Khalifa Center for Genetic Engineering and Biotechnology, in Al Ain, UAE, under supervision of *Dr. Mounir El Bellaj* and *Dr. Ajay Garg*; and University of California Davis's Blumwald's Laboratory, in Davis, California, USA, under the supervision of *Prof. Eduardo Blumwald* and team.

I would also recognize and thank *Dr. Sarada Krishnan* (Denver Botanical Gardens), *Dr. Fabienne Moreau* (ADNid), and *Julie Orjuela-Bouniol* (Qualtech-groupe), for their assistance in verification of results and statistical analysis. The confirmation of overall information was conducted by *Dr. Hassan Shabana, Dr. Samir Al Shaker* (both | Iraq), Dr. *Helal Al Kaabi* (UAE), and Dr. *Abdellah Meddich* (Morocco), we thank them for their efforts and service to this investigation.

DEDICATION

I am greatly honored to dedicate this thesis to:

Humanity & my Family,

to whom we strive to make gratified and represent our ancestry,
to my Father, for the seed besides the offshoot surely did not fall too far from the tree,
to my Mother where I still feel all the kisses she bestowed,
to humanity for we hope they benefit from my exploration.

In the name of Jesus Christ, the most illustrious of Abrahamic Monotheistic faiths

"The righteous will flourish like a palm tree, they will grow like a cedar of Lebanon (12); planted in the house of the Lord, they will flourish in the courts of our God (13). They will still bear fruit in old age, they will stay fresh and green (14); proclaiming, "The Lord is upright; he is my Rock, and there is no wickedness in him (15)". (The Holy Bible, Psalm 92 niv.).

In the name of Ancient Atlantis, Sumer, and Egypt

Thoth God of knowledge sewed the cultivation, selection by refinement and prominence of date palms. Such a tree of knowledge is the Palm-tree, carrying a palm-branch as his distinctive note or mark; the date palm was the tree of Thoth and the palm branch his book.

In the name of Moses '*Moshe Rabbenu'*, Prophet of God Almighty

The Palm is referred to the backbone, uprightness. The seventy date palm trees further epitomize the seventy "faces" or aspects of Torah that are revealed to those who eat of its fruit, an "orchard" of spiritual and intellectual delights, symbolizing the tzaddik, the righteous person, of whom it is said, "The righteous like the date palm will flourish." (Psalms 92:13), through which man is to serve God (cf. Sefer ha-Hinukh, #285).

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

أَفَلَمْ يَنظُرُوا إِلَى السَّمَاء فَوْقَهُمْ كَيْفَ بَنَيْنَاهَا وَزَيَّنَّاهَا وَمَا لَهَا مِن فُرُوجٍ (6) وَنَزَّلْنَا مِنَ السَّمَاء مَاء مُّبَارَكًا فَأَنبَتْنَا بِهِ جَنَّاتٍ وَحَبَّ الْحَصِيدِ (9) وَالنَّخْلَ بَاسِقَاتٍ لَّهَا طَلْعٌ نَّضِيدٌ (10) رِزْقًا لِّلْعِبَادِ وَأَحْيَيْنَا بِهِ بَلْدَةً مَّيْتًا كَذَلِكَ الْخُرُوجُ (11)

In the name of Allah, the most Beneficent, the most Merciful

"Have they not looked at the heaven above them - how we structured it and adorned it and it has no rifts? (6); And we send down from the sky rain charged with blessing, and we produce therewith gardens and grain for harvests (9); and stately palm trees, with shoots of fruit stalks, piled one over another (10); As sustenance for Allah's (God's) servants; and We give new life therewith to land that is dead: thus, will be the resurrection (11)". (The Holly Quran, Surrah, Qaf. Verses: 6,9-11).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SUPPLEMENTAL MATERIALS | APPENDICES

KEYWORDS

Microsatellites, Phoenix dactylifera L., Simple sequence repeats, Phylogenetics, Genetic structure, Date palm, Genetic diversity

DECLARATION

I declare that this dissertation is my own work and has not been submitted in any previous application for a degree. The results presented here are from research analyses that were carried out by myself, other contributions have been acknowledged respectively.

Salah E. Zaid

Chapter I

Introduction and Literature Review

1. General Introduction

1.1. Historical Perspective

The historic record on date palm (*Phoenix dactylifera Linnaeus | Linné, 1753*) has long symbolized it as 'the blessed tree'. The date palm is part of the Arecaceae or palm family, which was formerly referred to as Palmae. This is a unique palm family of 202 genera with 1500 species belonging to the Angiosperms-Monocotyledons (Dowson, 1982; Rizk and El Sharabasy, 2007). Dates are in the genus *Phoenix* and are generally thought of as one of the oldest cultivated trees in the world. The botanical name of the date palm '*Phoenix dactylifera* L.' is derived from a Phoenician name "*Phoenix*", which means date palm, and "*dactylifera*" derived from a Greek word "daktulos" meaning a finger, illustrating the fruit's general form (Linné, 1734). The tree is dioecious with large single crowns of pinnately encircling leaves. The trunk is covered with old-leaf foundations while the leaves are stiff leaflets specialized for moisture retention, sharply pointed and bifurcated in arrangement on either side of the rachis. The tree produces highly sought fruit with elongated furrowed monocot seed with a thick sweet mesocarp. This tree plays an essential role in the bionetwork, economy, and sociology of the Saharan ecosystem, representing a potential vital function in future arable and modified environments.

There is evidence of its use as a food source and as a medicine long ago. The discovery of dates in the tombs of several Egyptian Pharaohs and in Neolithic and Upper Paleolithic sites places its use back around 9,000 - 36,000 years ago and demonstrates the historical significance of the species to human nutrition and wellbeing (Chazan, 2009; Reader, 2011; Guatelli-Steinber, 2016). The history of its diffusion and cultural significance by the human population remains unclear, since the domestication of this species still holds uncertainty due to its prehistoric presence. The recent progression of date palms spans the garden cultivations of the Middle Eastern sub-continent, since its beginnings until the early Islamic era (Tengberg, 2012).

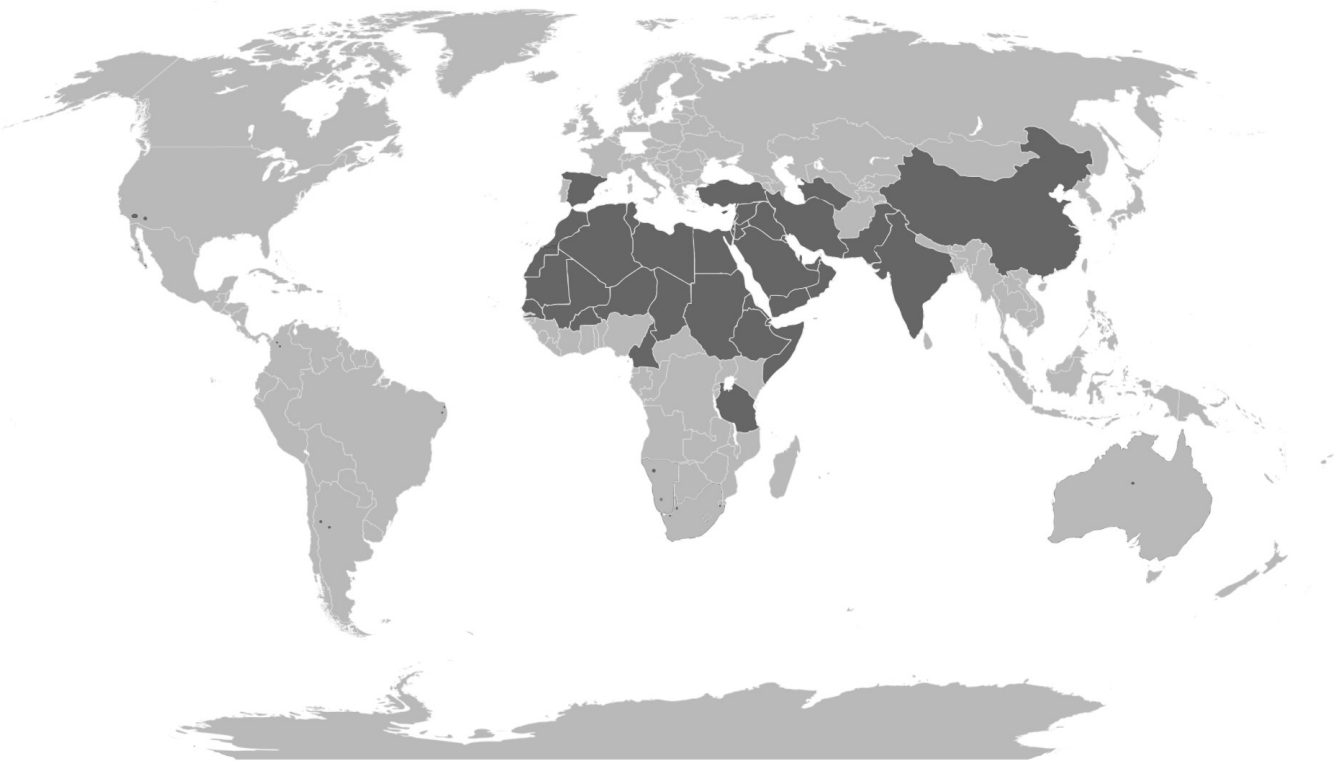
Date palm plantations are crucial to agriculture in many hot and arid regions. Date palm production constitutes an important economical pillar of the agricultural sector in most of the Middle East and North African countries, i.e. the MENA region. It represents a major source of income for a considerable number of farmers and inhabitants. The excellent adaptation to harsh climates and environmental conditions puts the date palm in a far more advantageous spot when dealing with the choice of what to grow in a commercial, private or small-scale plantation. The ability of the date palm to withstand adverse environmental changes such as drought, flooding, and extreme temperature fluctuations as well as its use in prevention and control of desertification, makes this crop a unique plant in the horticultural world. The market for dates is large and is considered to have even greater potential.

1.1.1. Geographical distribution

Global hot deserts, many of which support date palm (*Phoenix dactylifera)* groves and orchards, occur from the South Western borders of Asia to the coasts of North Africa and from the

northern lands of Syria to the Indian Ocean. These regions that are used for the cultivation of dates fall in zones that lay between the 15th and 35th parallels of latitude of the northern hemisphere. This unique tree is a desert species found mainly in arid hot lands (Figure 1.1).

In the commercial setting, the geographic distribution of date production is limited to certain areas known for their semi-arid to arid weather where they are used primarily for the export market as acknowledged by the United Nations trade commission (UNCTAD).



Geographic distribution of Date Palm (*Phoenix*), Selected places west to east: USA : California-Indio, Mecca, Coachella Valley, Tempe; Arizona : Phoenix ; Mexico : San Ignacio, Todos Santos, Loreto ; Colombia : Soatà, Bucaramanga ; Brazil : São Gonçalo, Arcoverde, Castro Alves (Bahia) ; Argentina : Cruz del Eje (Cardoba), Patolea ; Namibia : Naute, Eersbegin, Aussenkehr ; Swaziland : Pasture valley, Nhlangano ; Australia : Alice Springs (Soussa, 1969).

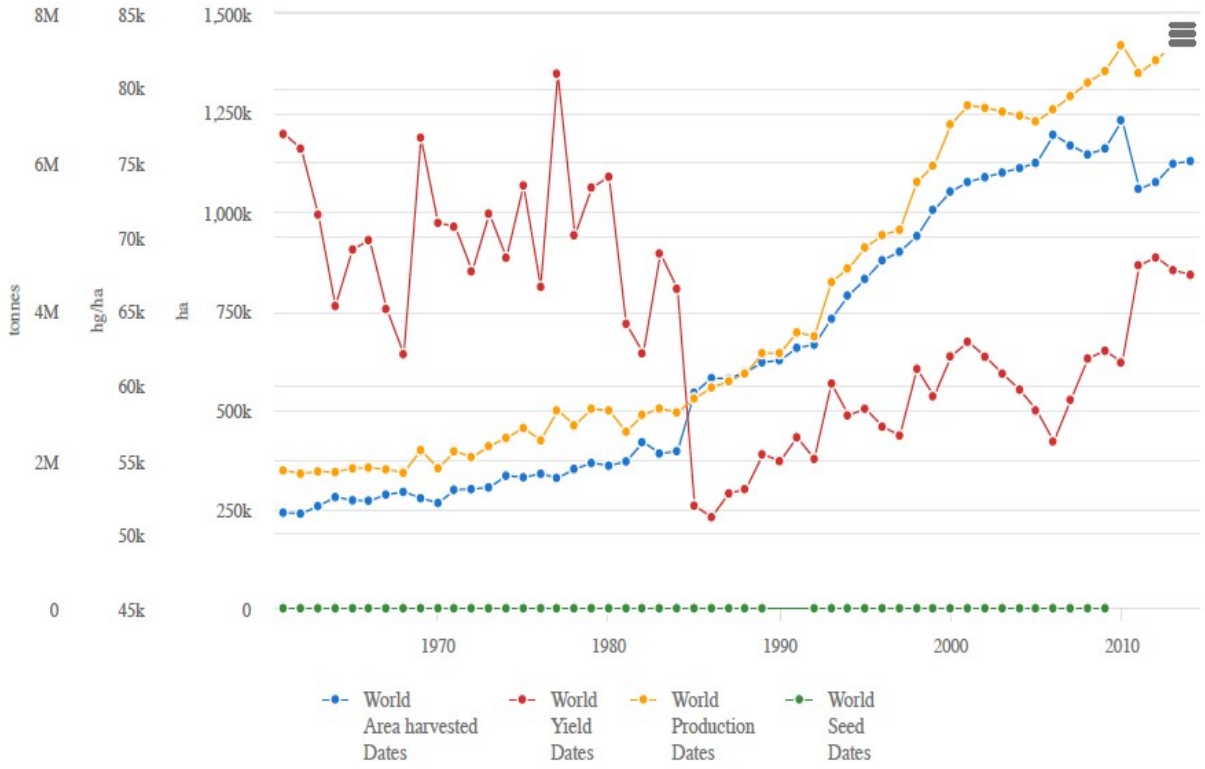
Figure 1.1. Map of the world showing areas of date cultivation and selected sites.

Historically though, the type species for the *Phoenix* genera is generally agreed to have sprouted from the boundaries of the fertile crescent; also known as the Mesopotamian peninsula or the modern-day landmass area of Iraq. Minimal rainfall and long hot dry summers are characteristic of this area. From this plausible center of origin, date palms spread by means of trade caravans and peoples that traversed the land following an equatorial distribution across the Middle East and into North Africa and Southern Asia. While the first evidence of the cultivation of date palm in southern Mesopotamia dates back to the 5th millennium B.C., oasis agriculture seems to have developed mainly during the early Bronze Age (from around 3000 B.C.) (Tengberg, 2012).

Krueger (2001) stated that date palm has been cultivated and subjected to selection by man since ancient times and the distinction between 'wild' and cultivated date palms is quite blurred. This may not be the realistic case as recent genome sequencing of various cultivars and wild types sourced globally have led to the understanding that there exists a great deal of diversity in both cultivated and wild populations. Conservation of the wild population of date palm has sadly experienced a decline since the main interest in date palm phyla has been solely based on high quality fruit production. Subsequently, the demand for increasing quantities of fruit has resulted in methods of large-scale propagation (micropropagation) that are driving the date palm industry today.

Within *Phoenix,* it is agreed that between 10 - 13 species exist as noted by Moore (1963), Bailey Hortorium (1976), and Barrow (1998). The dioecious nature of the genus allows for considerable interbreeding, producing hybrids that are interspecific in nature. Wrigley (1995) has indicated that the species in the genus should all be considered under one species umbrella, as morphologically the cultivars and wild varieties are indistinguishable. The use of molecular

markers has become essential in describing such traits and has been shown to be highly useful in phylogenic relationship determination and cultivar characterization among genotypes. A modern generation of molecular markers has been developed based on polymorphic exploration and establishing steady intracultivar stability in further characterizing date palm germplasm. Diversity investigations were previously limited in scope and in sequencing fine scale relationship elaboration. This led to the resolve of improving of the crop and is of great interest in genome comprehension, permitting cultivar genotype identification, which has been greatly restrictive in large scale dates production.



Source: *FAOSTATS2017*.

Figure 1.2. Relative global date production, areas harvested, and yield from 1961 - 2015.

1.1.2. Date palm cultivation

Date palm plantations (orchards) may produce high quality fruit and useful bi-product material for around 150 years. Some date palm trees are known to have survived for a few hundred years. Nixon and Carpenter (1978) found that cultural practices used in date production do vary across the globe. In wild populations, pollination is by means of natural wind and in certain areas by insects. In the commercial world, a mechanical approach for collection of pollen and application is adopted which assures more than 75 % of the female flowers to produce adequate (commercial level) fruit sets (Zaid and De Wet, 2002). Commercially, relatively few male lines are used for pollinations. These include the Al Ain City Male Palm (MP), a cultivar that is believed to have originated in North Africa, which is used as a main quality pollinizer. This male cultivar is reputed to be closely related to the Medjhool cultivar due to geographic proximity of place of origin. Incompatibility has been a known phenomenon when considering pollinizer selections, as well as the 'Xenia' effect (Reuveni, 1986) which influences fruit structure. No known observations on the xenia effect have been referenced for the modern commercial pollinators, as favorable environments have been carefully selected and tailored (Chao and Krueger, 2007).

Large-scale production of dates found in arid areas with high temperature require large amounts of water, found in oasis', for good growth and yield of good quality fruit. The oasis environment is perfect for the support of multiple types of plant life, where the date palm plays a vital structural role. The requirements of water vary with those cultivars which are larger fruiting types requiring a much greater supply of water than smaller fruiting types under conditions that are similar. A supply of excess water did not always increase rate of growth and fruit yields. However, during the summer periods of high heat the withholding of water suppressed growth and lowered accumulated moisture content of fruits by time of harvest (Reuther and Crawford, 1945).

Modern watering applications involve drip-irrigation as a main method with fertigation on a commercial scale. Conventional methods of furrow irrigation are still used in many parts of the world due to lack of development, technology, and training in those areas.

The data set includes data on total relative production in tonnes around the world. Source; *FAOSTATS 2017-actual 2014-2015*. Source: CartoDB.

Figure 1.3. Production quantities of dates by country, showing total annual relative production in 2014 - 2015.

Fertilization of date palms is usually applied with irrigation (fertigation). Nitrogen is the primary nutrient applied as noted by Nixon and Carpenter (1978). Zaid et al. (2002) suggested the use of fish emulsion as an alternative fertilizer, which supplies nitrogen as well as a diverse group of micro and macronutrients, enhancing the overall growth and period of juvenility. There is no quantifiable link of fertilizer application in terms of quantity and nutrient level quality on overall

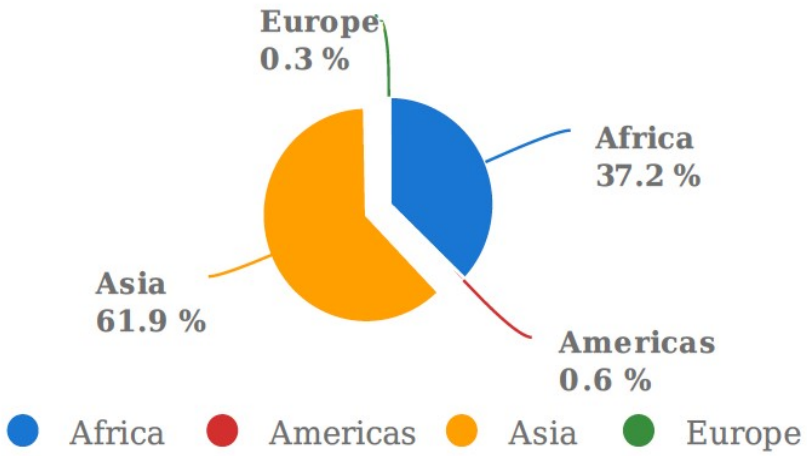
fruit production. Conventional practices use manure in date production while modern approaches depend on many inorganic nutrient fertilizers as a source. There is no evidence of occurrence of variation in nutrient levels in relation to source of nutrients.

1.1.3. World Production and Trade

The area of origin of a domesticated crop is typically the region where the crop is popular and has had a deeper cultural and intrinsic impact. The Arab countries, where dates play more of a livelihood enhancement role than a simple food source, possess roughly 70% of the 125 million date palms and produce about 67% of the global date production (Figures 1.2 and 1.3). During the past 50 years, date palm plantations were degraded due to extensive exploitation, resulting from the increase in human population and the feeding of domestic animals. The Bayoud disease (*Fusarium oxysporum* f.sp. *albedinis*) has also impacted yields significantly, especially in the North African region. Land loss associated with land degradation and various forms of desertification have impacted the overall production as well. Furthermore, urban development and the mismanagement of rural acreages have reduced production as well as negatively impacted wild populations. Currently, date production levels are about 8.89 million metric tons of fruit annually. Approximately 1.4 million hectares are used to produce date fruit, which is found in 41 nations (Figures 1.2 and 1.3). A more concentrated level of production of diverse populations is needed to meet future demand and mitigate potential forthcoming losses.

The leading date producers are Egypt, Iran, Saudi Arabia, UAE, Pakistan, and Algeria, respectively. These nations alone account for 70 % of global production. The recent destabilization of certain regions where dates were formerly grown such as Iraq, Syria, and Libya have led to increased production in neighboring nations (Figures 1.3 and 1.4). Note that due to

lack of data availability the consideration for Sudan and South Sudan reflects the situation up to July 2009 and Iraq up to 2003-2004 (Figure 1.2). Average production has dramatically increased in China, Egypt, India, Morocco, Namibia, Saudi Arabia and UAE by an approximately an average of 3-7 % during the last 5 years. In a span of 40 years the world export of dates has increased by approximately 1.76%, while production has improved by 2.9 times. During this time, the world population has doubled (Figure 1.2). Asia and Africa are by far dominating production by accounting for 61.9 % (2 520 139.94 tons) and 37.2 % (1 512 918.72 tons), respectively. An expedited non-linear transformation is likely as technology transfer and knowledge dissemination become more readily available.



Actual range (1961-2015), Source (FAOSTAT / Nov. 2016). Aggregate may include official, semi-official, estimated or calculated data. Asia incudes Middle East region signifying its dominance.

Figure 1.4. Global date palm fruit regional production share averages of dates.

Egypt is still, in 2018, considered the top producer in terms of quantity but has effectively no export potential. This is due to the low level of fruit quality inherent to the forests of the northern provinces. Poor fruit quality is associated with the lack of proper selection and

propagation of well adapted clones. There is a great interest in revolutionizing the Egyptian date industry so as to become a leading competitor on a global scale when compared to the U.S and Israel. This is especially true in the case of organic dates. In the Siwa region, Egypt North West of the capital city of Cairo, considerable development has recently taken place with the establishment of next generation facilities equipped to produce close to a million of elite propagules annually. On the other hand, a new player is Morocco, even though it is considered an old leader in the date industry, it is slowly but gradually becoming one of the best quality producers. The main focus in the future is that Morocco serve internationally as a source of tissue culture-derived planting and tissue culture initiation phase material. Note the most praised and widely distributed cultivar, Medjhool, originated there.

1.1.4. Importance of the Species

The most important tree crop of arid to semi-arid regions is the date palm due to its adaptive qualities. More than a fruit tree, the religious monotheistic relevance is significant in the Christian and Islamic faiths. Cited many times in the Holy Quran, it is used to break one's fast with a glass of milk during the holy month of Ramadan when fasting takes place annually. A multitude of ancient religions and faiths acclaim the benefit of dates.

In terms of large-scale cultivation as well as in smaller farm units, the advantages of this tree's elasticity to biotic and abiotic stresses is unmatched. The tremendous advantage is its resilience, its requirement for limited inputs, its long-term productivity and its multiple purpose attributes (Dowson, 1982). The inherent ability to tolerate environments that induce extreme stresses such as due to salt and drought has been well noted among producers. Micro ecosystems

may also be created using date palms and in support of such constructs palms in general that can reach great heights as pillars forming a greenhouse environment for other fruit trees and legumes.

The vast applications of the tree structure and byproduct materials are very important and may be used to construct reliable shelter structures. It not only produces a fruit which is a unique and important source of nutrition on its own, but it is also used for structural support, shading, thread and rope (basket making) and bark-based materials as well as products that include sugar, honey, oil, resin, and wine. Date fruit harvested are largely consumed, 91%, within the nation of origin.

The nutritional richness of dates is moreover well documented, and they have additionally been used for medicinal routines throughout human history. Rahmani et al. (2014) reported the valuable vitamins and minerals as well as fiber present in date fruits. The nutrients readily available are oils, calcium, iron, copper, manganese, magnesium, potassium, and phosphorous, found to be relatively present. Dates are different from most fruits as nutrient sources because of their major effects on brain health, due to reasonable quantities of B6 vitamin levels which have been found to be associated with improved student performance and scores (Balk et al., 2007). The inclusion of dates in the diet is ultimately prominently beneficial, although those that may have a sweet tooth should refrain from over consumption, resulting in adverse health conditions owing to the relatively high sugar content. Some health concerns have been reported to be treatable by date utilization and diet implementation. These vary from inflammation relief, constipation laxative, bone health and strength promotion, control of intestinal disorders, improvement of those with anemia; nervous system health; night blindness; diarrhea; to abdominal cancer (Al-Shahib and Marshall, 2003; Al Kuran et al., 2011).

1.2. Structure and Systematic Description

The Date Palm structure is unique to the *Arecaceae* family, formerly known as the *Palmae*, with broad leaves sprawling from the center of the apical dome and falling harmoniously to form a 360° floweret structure. Uhl and Dransfield (1986) have classified the date palm as: Group (*Spadiciflora*), Order (*Arecales*), Family (*Palmae / Arecaceae*), Sub-family (*Coryphyoideae*), Genus (*Phoenix),* and Species (*dactylifera*). As the adult tree sprouts out of the ground, it creates a solid dominant meristem, which rarely forms branches except for offshoots at the base of former leaves above ground. Typically, all date palm cultivars will produce offshoots as a form of specialized self-multiplication. However, in the commercial settings where date fruit are the specific crop, these are averted as they limit the potential of the mother plant's ability to produce a good quality product.

The usual tree consists of a main trunk that occasionally supports a few to several offshoots spawning from either the soil level or continue rising as the tree matures (Figure 1.5). The trunk is referred to as the stipe, apically covered by fronds that are spatulate or elongate leaves, which are supported by a vast network of fibrous tissues. The female, male or hermaphroditic reproductive flowers are produced in clusters, which arise from the axils of the leaves (Zaid and De Vet, 2002). Leaf clustering are numbered according to parastichies by numeral sequence pattern with a left-hand chronological curve, the relative position of the leaves will be exactly symmetric: n, n+3 at its left, n+5 at its right, n+8 at its left, of 5, 8, and 13. This is typical for systems designed in respect to Fibonacci arrangements. The helices of 3 and 8 wind round in the same direction as the chronological curve and helix 5 in the opposite direction (Ferry, 1998;

Elhoumaizi et al., 2002). The female flowers have 3 carpels narrowly pushed together. Only one of the carpels is fertilized and develops into a one-seeded fruit berry. The other two carpels abort. The seed is grooved ventrally and has a small embryo with a solid endosperm. The seed is long lived as demonstrated by Sallon et al. (2008) who germinated 2 000 year old date seed that was excavated in the Judean area of Masada, the Dead Sea.

1.2.1. Genome Relationships

In the *Arecaceae* family as classified by Dransfield and Uhl (1986, 1998, 2008) and Uhl and Dransfield (1987) and initially described by Moore (1973), there are 6 subfamilies, 14 tribes and 38 subtribes. *Phoenix* is morphologically distinguished as a pinnate leaved genus with split leaf lamina of induplicately folded segments. Among *Arecaceae* studies, there is little genomic assessment with the exception for the economic crops of date palm and oil palm. A total of 7 265 conserved microsatellites have been identified in *Elaeis guineensis* and *P. dactylifera* by Xiao et al. (2016). This level of shared relationship is observed in closely related genera belonging to *Arecaceae*, a point that potentially offers the possibility of a bridge for the transfer of genes, cDNA cloning, and flanking sequences within the same genus. For example, Singh et al. (2013) have shown a 65 Mya divergence between *P. dactylifera* and *E. guineensis*. There are still many understudied species of *Arecaceae*, which, with further assessments of genetic relation and full sequencing, is needed to aid in the conservation of this family.

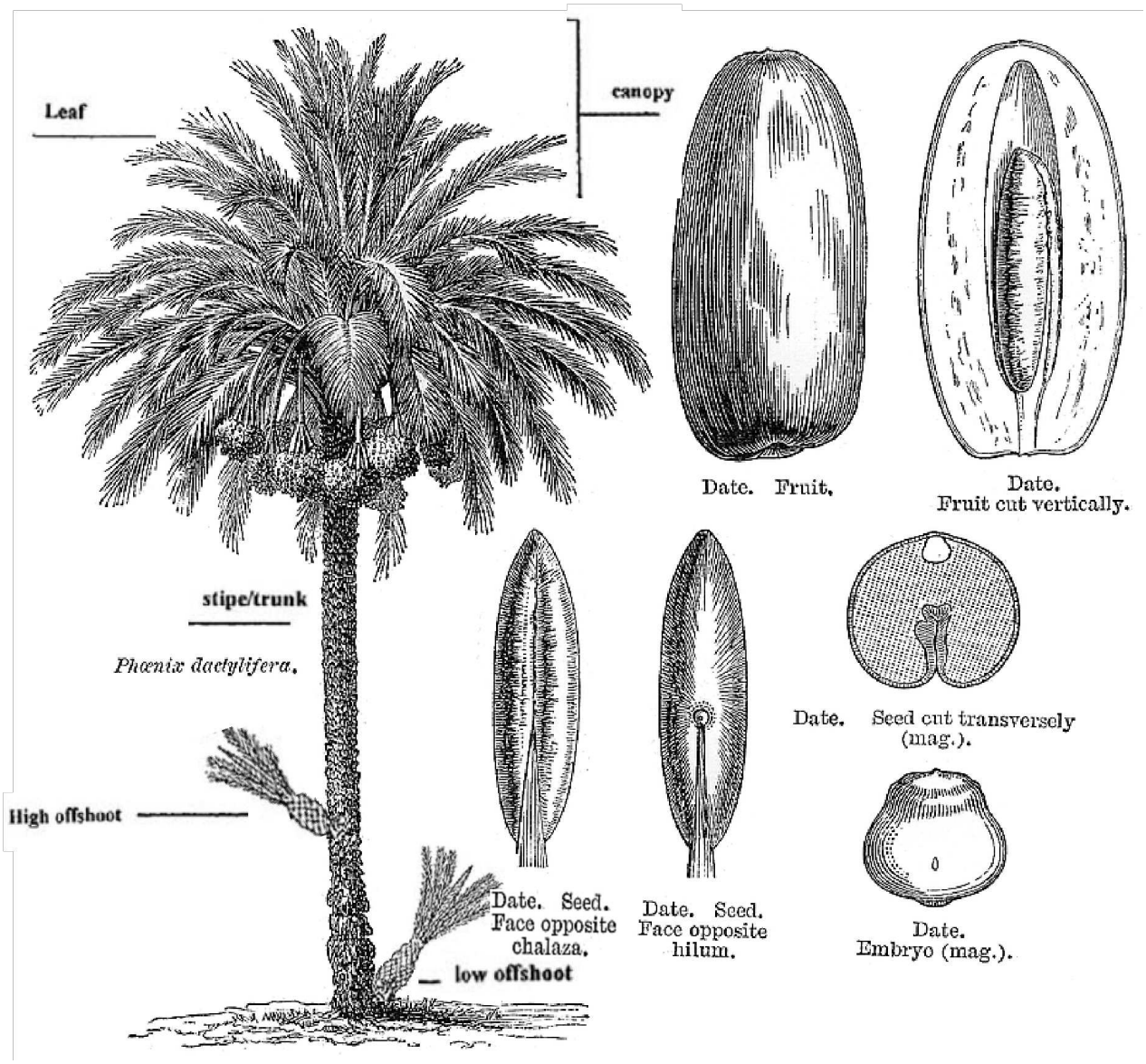


Figure 1.5. Representation of the date palm structure, illustrating components of morphological features and specialized offshoot structures. (Watson and Dallwitz, 1992).

The transferability of utility markers is another way to gauge the genome relationships that may exist. Billotte et al. (2001) completed PCR amplification tests on 16 other *Arecaceae* species and sequences of cloned alleles that indicated that *E. guineensis* microsatellites are potentially transferable markers in other palms. These were reconfirmed by Billotte et al. (2004) who

developed nuclear microsatellite markers of 16 nuclear simple sequence repeat (SSR) loci characterized in *P. dactylifera* and across taxa. This demonstrated the utility of most SSR markers which were observed in 11 other *Phoenix* species and the transferability of some of them to *Elaeis guineensis*, 11 species of *Pritchardia*, *Pritchardiopsis jeanneneyi* and six species of *Astrocaryum*. Using Plastid genomes, Barrett et al. (2016) generated 39 plastomes representing palms and the related family *Dasypogonaceae* for phylogenetic and molecular evolution annals. They found 'deep' relationships among the commelinid orders, among the five palm subfamilies, and among tribes of the subfamily *Coryphoideae*. These represent ways of providing genome-scale support for phylogenetic and genomic relationship studies. A recent partial sequencing of the date palm genome has solidified the distribution of shared gene families among principal angiosperms.

Al-Mssallem et al. (2013) reported proteome comparisons of *P. dactylifera* to *Arabidopsis thaliana*, *Oryza sativa*, *Sorghum bicolor* and *Vitis vinifera,* revealing 8 093 gene families that were shared among all five plant genomes and 1 127 gene families that were unique to *P. dactylifera*. Such differences, similarities, and genetic structure notions will be valuable in the future to aid in understanding the potential of date palms to eradicate chronic hunger.

1.2.2. Conservation and Genetics

Dates have a long history of domestication. However, wild populations are poorly documented relative to diversity and conservation. Conservation projects have focused on preserving the various commercialized varieties. There is an implicit higher level of diversity in wild populations since all commercial genotypes are clonally propagated which would likely lead to lower diversity. There are exceptions such as in the case of wild relatives of dates such as *P.*

theophrasti. These are mainly landlocked variants, which are considered in a vulnerable state and found only in a limited number of isolated locations (González-Pérez et al., 2004; Pintaud et al., 2010).

The modern approach to selecting and establishing varieties to be produced has led to a significant erosion of genetic diversity. There is a tendency to have a goal to replace long-established and local genotypes with ones considered 'elite' such as 'Medjhool' and 'Fard Abyad' (white). This leads to fewer genotypes and thus potential genetic erosion. The banana (*Musa* sp.) is propagated via cloning which has resulted in a narrow genetic base and lack of variation within commercial populations. Seriously threatened by the re-emergence of a *Fusarium* Wilt. The disease, caused by the soil-borne fungi *Fusarium oxysporum* f. sp. *cubense* (Foc) and also known as "Panama disease", that was resolved via a rapid replacement of the major cultivar, resistant to the disease via tissue culture production (Pérez-Vicente et al., 2014). In date palm tissue culture, this has become a challenge, as a very small percentage of the genotypes of offshoots meristematic tissues respond to the media formulae. Techniques based on the culture of meristematic apical growth points (shoot apical dome) and their culture on a tailored medium (each date cultivar varies in its requirements for growth), is where plant development takes place. This in part relates to its fibrous structure, which makes it difficult to work with. This threat will continue to worsen while the micropropagation approach remains limited to those genotypes that respond to this rapid multiplication technique and thus potentially leading to a weakening of future genetic diversity.

The largest collections of date palm cultivars and species of *Phoenix* are located in Algeria, Brazil, India, Iraq, UAE and the US. Various approaches of *in vitro*, cryo-storage, and *ex situ* methods are being used for the conservation and preservation of the germplasm (Bekheet, 2011).

The genetic diversity of natural or wild populations of date palms has had little study but remains a potential resource that can be relied on in future genetic manipulations (Jain, 2011). These preservation centers appear to consist of mainly commercial cultivars and hybrid lines; therefore, the diversity is expected to be rather low as clonal propagation is the means of maintenance. It is important to not only preserve as many cultivars as possible but also a good representation of wild / natural populations. As wild populations are predominantly based on seed propagation, they are considered more highly diverse since individuals could be cataloged as a unique genotype. The preservation of natural genotypes remains hindered by political instabilities and wars in various date growing regions (i.e. Iraq, Syria, Iran, Libya, Yemen). This has led to losses of germplasm and variations in populations, which is necessary to preserve and protect the great diversity of the species.

The unreported genetic resources and collections found in areas such as China, represent important sources for use in the future and therefore need protection. There have also been reports of small collections and academic preservation projects across the globe (Bajaj, 1995; Dickie et al., 1992; Johnson, 1996; González-Pérez et al., 2004). These remain a small percentage of the recent decade efforts at major research and development centers, which are using innovative and superior methods to preserve genetic diversity. Little information is available on natural / wild populations, and their genetic diversity present (MoRrcr, 1998).

As ornamentals, *Phoenix* species do quite well since their selection is mainly limited to visual appreciation and does not entail the characteristics of commercial fruits. Numerous *Phoenix* spp. are considered valuable, although those that are widely distributed such as *P. canariensis, P. roebelenii, and P. loureiroi* are considered well preserved as are *P. sylvestris, P. reclina, and P.*

paludosa. As conservation has mainly been focused on species or crops of monetary profit the status of *P. acuaulis, P. atalantica, P. caespitosa, and P. pusilla,* to name a few, along with most wild populations of *Phoenix* are still unknown (Johnson, 1996).

Arable dry to semidry land is of major concern to the conservation of the *Phoenix* spp. The loss of habitat is more of a concern for palms, especially in their prevalent regions as these provinces lack the assessments that lands of the western world entail. Where genetic erosion is of concern, extensive losses have been reported although we do not have enough information and accumulated data among collections to conclude the level of losses. Hence, the knowledge of genetic vulnerability of date palm, with respect to understanding how well it is distributed and its genetic diversity both of cultivated and wild populations, as influenced by desertification, man-made land degradation, and climate change are not well understood. There have been conservations by means of *ex situ* and *in situ* maintenance although they have not had a significant effect relative to an understanding of genetic attrition (Gebauer et al., 2007; Al-Yahyai and Al-Khanjari, 2008).

The genetic organization of date palm, with the recent resequencing of the entire genome across 100 cultivated varieties at NY University, Abu Dhabi (Purugganan, unpublished), has elaborated on previously reported incomplete numbers. Hazzouri et al. (2015) determined and confirmed two assemblies of the 690 Mb genome of the elite Khalas cultivar by two independent groups, covering 90.2 % of the genome with 41,660 gene models. The most complete assembly consists of 82,354 scaffolds, with an N 50 of 329.9 kb and a maximum scaffold size of ± 4.5 Mb. It appears that the date palm genome is much smaller than its counterparts of perennial

monocotyledons. The date palm is diploid although previous publications have indicated various chromosome ($2n = 2x = 36$) numbers (26-36) (El Hadrami et al., 2011), which have been associated with the multiplication process by tissue culture and the inherent cultivar type.

1.3. Date Palm Propagation

The propagation of date palms today is driven by, and dependent upon, the high demand for planting material of elite cultivars. Date palms are a dioecious species, which require a system of cloning to preserve genotypes and specifically fruit phenotypes of good commercial quality. The adopted method in the past was through offshoot harvesting, as seeds were not used to propagate dates since each seed could be considered a new genotype and be different in fruiting characteristics when compared to the mother plant's expression.

1.3.1. Propagation Systems

To propagate the date palm three techniques are available; seed propagation, offshoot harvesting, and tissue culture methods. Seed propagation is the most straightforward approach; however, each genotype is unique with one half being males, which are of limited value commercially. Although recently there are some molecular markers that have been identified that could be used to separate male and female plants this is both time consuming and costly (Tisserat, 1979a). Compared to established cloned trees, seed propagation leads to late maturing fruits that are variable in quality. The production potential, overall quality, and harvest time are limitations opposed by commercial operations (Ben, 1990). The ornamental industry does use seed propagation as the main approach because fruit are not of interest.

Offshoots are the other nonspecialized method less used today. These are axillary buds protruding from the lower trunk area, which are considered true clones capable of producing identical fruit to the mother plant. The harvesting of offshoots typically occurs after 5 plus years from axillary bud initiation. This time frame discourages the use of offshoots when compared to tissue culture techniques (Chao and Krueger, 2007). The use of offshoots for commercial propagation is impractical since offshoot production is limited by a short vegetative phase (roughly 10 - 15 years), a restricted number of offshoots are produced per tree (20-30 across the lifespan which is cultivar dependent), low offshoot transferability and survival rates, and they have higher frequencies of the spreading of diseases and pests (Zaid and Arias-Jimenez, 1999). The final method of propagation, and most widely used is tissue culture or *in vitro* techniques (micropropagation) (Sinnott, 1950). In some countries, such as the Kingdom of Morocco and the UAE, date palms are largely propagated by tissue culture for both ornamental and fruit bearing trees. Without a cambium cylinder, date palms as a monocot are impossible to graft or bud.

1.3.2. Micropropagation: Date Palm in vitro tissue culture

At the beginning of the 20th century there was a need for a system of propagation that resulted in a rapid increase in crop numbers including wild populations. German scientist Gottlieb Haberlandt (1969) was the first to propose the idea of 'Tissue Culture' and is acknowledged as the father of plant tissue culture. This is the aseptic culture of cells and various tissues grown in a controlled environment *in vitro* with chemicals (Thorpe, 1990). Plant propagation by tissue culture is beneficial since once initiated the culture can be indefinitely increased. The first true cultures were from cambial tissues of *Salix capraea, Acer pseudoplatanus, and Robinia pseudoacacia* that were produced by Gautheret (1934, 1935, 1939) and White (1934). Most of the early investigations

and techniques were developed used meristematic tissues. Following their early efforts indole acetic acid and other growth regulators (hormones), as well as additional media formulae constituents such as vitamins and various micro and macronutrients became available. The stage was then set for later scientists such as Skoog (1948), Tsui (1948), Nobécourt (1955), Steward et al. (1958), and Murashige and Skoog (1962) to standardize the method. By the 1970's, the modern groundwork and commercial process was well established by Cheng Tsai-Ying and today we can multiply numbers of most plant species many fold (Murashige, 1974).

Tissue culture and specialized vegetative propagation are *in vitro* procedures used for research and mass propagation. The requirements are explicit nutrient media, controlled temperature, lighting, and aseptic conditions. The exploitation of tissue culture technology by large-scale facilities has mainly been by the ornamental sectors which produce herbaceous plants with more limited usage in woody plants (Jain et al., 2013). The advantages of using this system instead of others is that mass propagation of selected cultivars (elite or endangered) may be accomplished and mother plant clone conformity of product (true-to-type / genetically identical). This approach may only be successful, when effective means for establishment of aseptic cultures, multiplication of initiated cultures through the establishment of organogenesis or embryogenesis and hardening of tissue culture derived plants prior to their field planting are met.

Two methods have been used successfully for date palm. These are organogenesis (division of shoot apical tips and lateral buds) and somatic (asexual) embryogenesis (involving embryo germination and initiation from proliferated callus) (Jazinizadeh et al., 2015). There are 4 stages of embryogenesis: callus accumulation, multiplication, followed by germination and elongation

(Jain et al., 2013). The organogenesis method is used most commonly for commercial production because it has higher survival rates when plantlets are transferred to fields. It is of note that the spread of good quality date cultivars is dependent on the way they are propagated.

1.3.2.1. Organogenesis

The use of meristematic tissues has allowed for micropropagation while avoiding callus formation, reducing he length of the *in vitro* period, limiting the frequencies of somaclonal variation, and lessened media hormone requirements. The use of 2, 4-dichlorophenoxy acetic acid (2,4-D) is not required thereby reducing the growth substances applied to the media. Furthermore, the elimination of 2, 4-D reduces somatic variation as this growth regulator has been shown to be associated with increased mutations (somaclonal variation). The meristematic bud initiation stage is regarded as the most critical for success. It is this step that has been associated with the majority of assorted problems observed in the *in vitro* process. Once initiation of the meristematic explant is accomplished in the dark, the 5 remaining steps consist of multiplication, elongation, swelling, rooting, and acclimatization.

There exists a lack of clear information on date palm organogenesis tissue culture. This is because most facilities that practice date palm tissue culture have a low efficacy with organogenesis. However, it remains superior to embryogenesis as it avoids most issues such as high somaclonal variation. The organogenesis method requires further improvements because of differing media requirements by the numerous cultivars that are grown. Many facilities have not had the success and production numbers for organogenesis to be considered an established method in date palm propagation. The one facility that has had greater success is the 'UAE tissue culture laboratory' in Al Ain (UAE university).

This facility has broken production numbers in the region by nearly doubling them in recent years. There still exists somaclonal variation when used. Although there is low occurrence it does happen and appears to be associated with the use of high levels of growth regulators and farmer cultivation practices.

1.3.3. Affiliate Problems in Date Palm Tissue Culture

1.3.3.1. Commercial Facility Restrictions

The commercial facility has its own level of management protocols. Several laboratories around the world have initiated *in vitro* micropropagation of date palm but only a few have been successful using either somatic (asexual) embryogenesis or by the better method of organogenesis. The requirements for sizeable production involve the establishment of essential production goals, which ultimately determine the factors and method used.

A major weakness of palm tissue culture is the undesired off-types associated with somaclonal variation. This is less frequent when buds are prematurely present rather than undifferentiated meristematic tissues. Only a few commercial tissue culture facilities have the knowledge to genetically insure true to type clones. The most reliable method is by Simple Sequence Repeats Palmàprinting (identification fingerprinting) (Zaid S. E., unpublished) although certifying mother-plant conformity can be achieved by; Iso-enzyme, RAPD, RFLP, and RAFLP approaches. Ultimately, due to various factors such as random mutation, environment by gene interaction, epigenetic cursors, and farmer cultivation practices, the only reliable way to confirm true to type is by field responses and comparisons with the mother source.

1.3.3.2. Contamination

A major issue is contaminants and contaminated vessels and tissues, which are frequently observed. Most contaminants are brought in by technicians on clothing or unwashed hands and less so by relative design of a laboratory. Endemic infections also occur in some clones from which cultures are initiated. Certain levels of infectivity may be treated by replacing the vessel with a media formula encompassing anti - bacterial and fungal components, although only a relatively small amount of initially contaminated vessels and tissues are saved. With fungal-based infections, these are generally thrown away since eradication has proven to be difficult.

1.3.3.3. Cultivar Mixing

The mixing of cultivars within the laboratory and nursery setting is a major concern that is faced throughout the tissue culture industry. As the cost of a system to control cultivar mixing can become overwhelming, wherever a cheaper alternative is feasible it is adopted. The use of color-coded stickers or numbering remains at the discrepancy of the technicians and assistants. Human error is common and verification protocols are not widely used. The availability of other means such as a barcoding system can alleviate this hindrance in practice, although these applications still remain costly and in an established facility have shown to be very difficult to implement.

1.4. Simple Sequence Repeats: A Review

1.4.1. Molecular markers

Within the last decade, genetic molecular characterization has seen tremendous advances with the development of numerous marker systems including; Random Amplified Polymorphic DNA (RAPD) (Soliman et al. 2003; Sakr et al. 2012), Amplified Fragment Length Polymorphism

(AFLP) (Cao & Chao 2002; El-Assar et al. 2005) and microsatellites or Simple Sequence Repeats (SSR) (Billotte et al. 2004; Elmeer et al. 2011). In the goal to assembling draft genomes, phylogeny relationship studies based at a molecular level and cultivar identifications represent the practical approach for use. The possibility to produce an abundance of molecular markers by use of next generation sequencing (NGS) technology has revolutionized the molecular capacity to carry out further relative research. Kumpatla and Mukhopadhyay (2005) demonstrated the utility of molecular markers expressed sequence tags, while Xu et al. (2010) used in silico analysis of existing bacterial artificial chromosomes (BAC) sequences. Bohra et al. (2011) then achieved this by means of BAC-end sequences and conclusively was followed with the application of whole genome shotgun sequences (WGSs) of Davey et al. (2011). This approach is still considered unsurpassed in using NGS technology.

While new techniques such as micro-arrays and DNA chips provide cutting-edge, high-throughput solutions to distance determination of clades, as such, they are costly, involve highly specialized equipment and require the further characterization of Single Nucleotide Polymorphisms (SNPs). Within the family of *Arecaceae*, the utility of transcriptome libraries and cDNA tag archives has led to the development of EST-SSRs in *P. dactylifera* (Zhao et al., 2012). Others have achieved similar results in *Cocos nucifera* and *E. guineensis (*Xiao et al., 2008*)*. Although most of these approaches are impractical due to cost of application, they are also dated as the utility of Simple Sequence Repeats / microsatellites have developed. Their general expression is mainly in codominant form with very high polymorphisms. This is used heavily in diversity analysis, Quantitative Trait Locus (QTLs) mapping, constructing maps characterizing genomes and their relationships, and have been successfully used in the delimitation of closely related species *(*Xiao and Fayer, 2008*;* Xiao et al., 2012; Zhang et al., 2013). Li et al. (2004)

mention that SSRs may provide an evolutionary advantage of fast adaptation to new environments as evolutionary tuning knobs (Trifonov and Berezovsky, 2003; Kashi et al., 1997), whereby allowing unparalleled investigation in understanding relationships and structure in general.

1.4.2. Microsatellites: A Succeeding Category of DNA Polymorphism

Microsatellites are simple sequence tandem repeats (SSTRs) occurring mainly in noncoding regions of eukaryotic genomes due to their long di-nucleotide repeats as well as in tri- or tetranucleotides (McDonald and Potts, 1997). Based on repeats of short 2-6 base pairs, they are considered a class of markers on their own. Their development was established with Akkak et al. (2009) who developed 17 microsatellite primers in *P. dactylifera* from two enriched libraries using $(GA)_n$ and $(GT)_n$ repeats. Earlier, Billotte et al. (2004) developed 16 microsatellite primers from a $(GA)_n$ microsatellite enriched library and reported their transferability across 11 other *Phoenix* species. In plant studies, SSRs have been demonstrated to be more variable, up to 5 times greater than other markers, and their utility has been well documented across many staple species such as maize (Senior and Heun, 1993), rice (Wu and Tanksley, 1993), wheat (Roder et al., 1998), and *Arabidopsis* (Bell and Ecker, 1994). In recent years, they have been shown to have a wide application with their use in genomic mapping and marker assisted breeding, population genetic analysis, and evaluations of polymorphism of germplasm (Fahima et al., 1998; Li et al., 2000; Song et al., 2002; Liu et al., 2003; Alamerew et al., 2004; Teklu et at., 2006; Liu et al. 2007). The detection of polymorphic microsatellites is carried out when an array is identified and flanked via both sides by unique sequences. Due to the differences in microsatellite repeat numbers a southern blot and probing with probe Q can be done. The autoradiogram band size is then based on set repeats in an array and their allelic calls. The popularity of these markers through numerous studies can be summarized by their high levels of allelic diversity at dissimilar loci, the ease of

amplification at PCR level, and as already mentioned their codominant nature. In date palm, SSR efficiency has been extensively used to study the genome expression of only a few of the over 2,500 date palm cultivars (Al-Khalifah and Askari, 2003; Askari et al., 2003), for assessing germplasm and their affiliate relationships (Billotte et al., 2004 ; Zehdi et al., 2004 ; Akkak et al., 2009 ; Zehdi et al., 2012 ; Arabnezhad et al., 2012 ; Haider et at., 2012 ; Baker et al., 2013).

SSR genetic markers used in molecular phylogeny, have been shown to be locus specific in contrast to multi-locus markers such as RAPDs. Experiencing a lack of selection pressure and generally are regarded as neutral 'genetic markers', however microsatellites are generally used as codominant homozygotes and operated under a binary 1 / 0 bases. Their DNA can be degraded or prehistoric but because they are highly polymorphic, and PCR based, their hyper-variability provides considerable patterns that are useful in phylogenetic studies. Therefore, SSRs can be useful for individual identification directed at fine-scale phylogenies. This requires only minute amounts of good quality DNA as well as no restriction enzymes and no unique probing. They can be exceptionally useful for addressing degree of relatedness of individuals or groups while also able to assess magnitudes of gene flow between populations (Li et al. 2002; Al-Khalifah and Askari, 2003; Adawy et al., 2006; Adawy, 2007; Elshibli and Korpelainen, 2008). This is because SSRs are easily detectable and carry high discrimination powers allowing for various genetic structure investigation.

The classifying of phylogenetic relationships among plant phyla is a very important and powerful part of understanding the genome associations at many levels. Therefore, more basically equipped molecular labs still rely on SSR analysis in application for the purposes of cultivar identification, distance determination, closeness, relationship establishment, true-to-type

determination, and genetic variation. The relatively high cost of microarray development, and its associated application effectively places that methodology beyond the reach of ordinary laboratories or facilities that are equipped for only basic molecular biology.

Currently, microsatellites still represent an easy, robust and cost-effective way of performing cultivar and 'cultivar phenodistinct substitute' (variation within the fruit inflorescence & phenotypic phytomorphology amid a cultivar type due to factors which are not necessarily gene driven) determination by phylogenetic correlations for such institutions. Hence, microsatellites can provide a realistic practical use for understanding genomes; the available sequence of species is ultimately the key to identifying microsatellite markers providing more concrete opportunities in understanding the relationships amid and within plant families.

1.4.3. Palmàprinting | Cultivar ID Fingerprinting

The use of phenotypic traits for absolute cultivar identification is less reliable to nearly impossible when compared to the DNA-based characterization for identification. The environmental interactions do not affect the basic genome sequence or molecular identity and thus are clearly a more reliable approach. In general, molecular marker techniques have been well developed using inherent genetic polymorphisms to properly distinguish between and among date palm cultivars. Corniquel and Mercier (1994), with the use of RFLPs and cDNA probes such as cDNA 1, conceived the reliable cultivar identification and screening of cultivated populations. These methods remain limited compared to the utility of SSRs, although they may be used as forms of verification with speciation uncertainties, and feasibility of cost. SSRs are easily applied, provide a high degree of polymorphisms and with accurate methodology may precisely distinguish

amid families of various species. As such any cultivar or wild type in the *Palmae* order and Sub-family *Coryphyoideae,* may benefit from microsatellite usage.

1.4.4. Software Review

The analysis of microsatellite data (allele frequency amplification bp calls) depends heavily on software and programs that manage the complexities of matrix computations across large populations by running various close and distant relationship statistics. For analysis of population structures, genetic differentiation and gene flows, gDNA and the utility of microsatellites has proven reliably routine, allowing for various genetic distance measures and constructing relationship trees of populations or closely related species. Understanding genomes of sub-sequenced species like the date palm using one of the many software programs have proven useful. However, with increasing sophistication of computational statistics more software packages are being developed.

The computation of genetic distance measures and dendrogram trees of populations are made possible by program POPTREE2. This software can handle closely related phylogenies among varietal types or cultivar phenodistinct substitutes, from gene allelic frequency data by using Neighbor-Joining structure. This was released in 2009 by Naoko Takezaki and Koichiro Tamura from Japan. This was used to facilitate Nei's and Da genetic distance as demonstrated in the 2nd chapter of this dissertation. It can also perform bootstrapping, and compute heterozygosity and Gst measures of the extent of genetic variation in a population and genetic differentiation among subdivided populations (Takezaki et al., 2009).

The computation of population genetics, such as estimation of gene frequencies, linkage disequilibrium tests, and diversity indices among and between populations was enabled by Laurent Excoffier, from the University of Bern, Switzerland, along with Stephan Schneider, and David Roessli released Arlequin (Schneider et al., 2000). This program was used for initial testing the final testing. The final data was confirmed and presented using DARwin (Data Analysis and Retrieval with Indexed Nucleotide / Peptide Sequences) created by Gaston Gonnet and Chantal Korostensky in Zürich, Switzerland. DARwin was able to carry out most of Arlequin analysis, including phylogeny, sequence methods, construction of Neighbor-Joining (NJ) trees and the assembly of Multiple Factor Analysis (MFA) (Gonnet et al., 2000).

Primary assessments of raw data from the capillary sequencers is done by using Data Collection Software along with Gene Mapper for assessing the genetic capillary analysis. This software combination generates electropherograms (frequency intensity and richness plots) on DNA fragment migration, producing the initial analysis results needed for potential relationship development. These results are then able to be transferred to secondary analysis software's for further processing.

Genetic structure can be assessed by the use of genetic markers such as AFLPs, RFLPs SNPs and SSRs (Microsatellites). The program STRUCTURE is a package for multi-locus genotype data to investigate structure of populations and individuals that represent varietal or sub genus populations. It is used in identifying migrants and admixed individuals as well as estimating population allelic frequencies that have many individuals or groups admixed, clustered, and further sub - clustering (Porras-Hurtado et al., 2013). The program CLUMPAK (Clustering Markov Packager Across K) (Kopelman et al., 2015), processes STRUCTURE Q-matrices, summarizing

and graphically representing the results along with the aid of software such as Distruct and Inkscape to precisely adjust, manipulate, and visualize the estimated memberships and relationships while adjusting color and structural coding respectively (Rosenberg, 2004).

The statistical analysis of the data and information produced by programs that yield diversity and genetic structures may be done by use of the program GenAIEx. Genetic marker analysis in a wide range of population genomic studies are produced with rich graphical outputs for data exploration and data manipulation. This program runs with Microsoft Excel, offering analysis of codominant, haploid and binary genetic loci and DNA sequences. This allows for distant based analysis consisting of AMOVA, PCoA, multivariate spatial autocorrelation (Blyton & Flanagan, 2012) and for allelic frequency correlation; F-statistics, heterozygosity, population assignment, and relatedness (Peakall and Smouse, 2001).

“Microsatellite marker data was used in this study to determine relationships among commercially significant date palm cultivars as well as their genetic structure and phylogenetic associations. SSRs (microsatellites) for specific cultivar identification or Palmàprinting was also evaluated as part of this dissertation.”

Chapter II

Genetic Structure and Diversity of Commercially Important Date Palm Cultivars (*Phoenix dactylifera* L.) Using Phylogenetic Relationships and Simple Sequence Repeats (*microsatellites*)

2.1. Introduction

Date palm, *Phoenix dactylifera* L. plays a particularly central role in the ecology, economy, and sociology of the North African (MENA) environments. Dates are well recognized for their nutritive significance and are successfully grown in irrigable desert lands. The dependability of this tree is crucial to agriculture, particularly high temperature (ranging around in July 73°F to 89° F, while across the Middle East and similar environments, summer temperatures are regularly around 85 F, but frequently rise above 100° F) regions.

Production of this fruit is a vital economical part of the agricultural sector of countries across the MENA region and provides a large portion of income for local populations. The excellent adaptation of the numerous cultivars to various climate conditions allows it to be grown over a wide area of desert climates where it aids in preventing and controlling desertification. The date palm market is composed of many genetically discrete clones, representing thousands of types that are commercially viable in specific areas. The product application for date palms is vast while it is a highly nutritious fruit with a long shelf life make it favorably valuable. The micropropagation of elite cultivars delivers a dependable source of plants for supplying market demand.

Knowledge of relationships between cultivars is slim even though the date palm genome of a handful of cultivars has been sequenced by Hazzouri et al. (2015). The knowledge summation regarding phylogenetic and genetic structure relationships are not well understood but would add much to the comprehensive understanding of the origin and germplasm of existing date palms. Presently, the accessibility of information on genetic grouping of these cultivars requires a review to better understand the connections amid these important collections. Given the limited genomic information for *P. dactylifera,* simple sequence repeats (SSRs) are ideal to determine relationships within and among the important date palm market commercial varieties.

Phylogenetic associations by microsatellites are useful due to their accuracy and ability to determine high polymorphic capabilities. This enables genetic relationships to be determined and fine scale phylogenies to be determined. Many groups of organisms have been investigated within and between closely related species through use of microsatellites markers and genetic structures (Kretzer et al. 2003; Billotte et al., 2004; Duminil et al., 2006; Zhao et al., 2012; Arabnezhad et al., 2012). These have been used in elaborating on sub cultivar populations and species complexes. The agro-biodiversity of the date palm has been investigated using various molecular markers specifying important alleles and systematically detecting distinguishing factors by means of SSRs (Salem et al. 2001; Rhouma et al. 2008; Elhoumaizi et al. 2006; Ahmed et al., 2009).

The genotypes and their relationships are resolvable by using marker based Capillary Sequencing producing specific base pair amplifications of alleles that can be applied to distance matrix assessments, allowing better understanding of correlations among cultivars. The primer selections were also achieved from the synopsis of characteristics of 16 nuclear sample SSR loci

isolated from *P. dactylifera as* described by Billotte *et al*. (2004). Selections of 20 commercial date palm cultivars representing those most commonly grown throughout the world were used for this study. Hence with this approach of relatedness among commercial date palm cultivars one may associate the phylogeny to divergences and frequency of allele amplification. Gene pool complexes and genetic diversity of date palm have been greatly shaped by human intervention and natural occurrences, through clonal propagation and early trade route movement of germplasm.

2.2. Materials and Methods

2.2.1. Date Palm Cultivars

A total of 20 major cultivars of date palm (*Phoenix dactylifera* L.) were selected for this work. The commercial importance of these cultivars was the basis of their selection with specific priority given to consumer preferences, fruit market demands, and offshoot production relative to cultivar establishment. The original names of the date palm cultivars are in Arabic and the various translations have led to variations in spelling and pronunciation (in this study, the mostly widely used spelling were adopted). These selected fruiting cultivars, listed alphabetically, were Aboumaan, Ajwa, Barhee, Chichi, Dabbass, Fard Abyad (white), Helwa, Khadri, Khadraoui, Khlass, Khenezi, Lulu, Medjhool, Maktoumi, Nabtat Seif, Sakaii, Sukkari, and Sultana. The two pollinators used in this study were Al Ain City Male Palm (MP) and Mdasry (males) (Table 2.1).

2.2.2. Plant Material

Fresh leaflet samples were harvested from each of the 20 selected cultivars from numerous sources. Samples from 7 countries and 23 locations were used. A total of 2 leaves of 10 - 15 cm

leaflets for each of 5 individuals were randomly selected per cultivar type from germplasm repositories, government collections, and private preservation orchards. These samples were sourced, in vivo, only from healthy, established and well-characterized trees that were 15 - 25 + years. Determination of cultivar identity of these selected trees was verified in conjunction with specialists from the Al Ain Dates Factory and UAE University, Al Ain, UAE. Genotyping and confirmation were accomplished using microsatellite sequences (Simple Sequence Repeats - SSR's) according to Billotte et al. (2004). Sample collections were made during the 2013 summer growing season from April to mid-September at the sources of germplasm listed in Table 2.1. The cultivar and some cases new varietal nomenclature was originally written in Arabic, hence slightly changes depending on the area of cultivation. The Trilateral Cultivar Confirmation System (TCCS) was developed and adopted. This system in this process of 'dates Palmàprinting' was based on three forms of verification used to identify most of the genotypes indicated in this study. Initially, the genotypes were verified according to annotated reference from the collection site caretaker along with any history of the individual tree or plantation from which the sourced material (explant) was collected. This verification was also reflected in the material, whether a meristematic tissue from an offshoot during testing or leaf tissue from field collections were designated as either known, a new cultivar or as cultivar phenodistinct substitute. The 2nd level of this system was the molecular endorsement using allelic data through SSR technology. The final confirmation step verifies the cultivar by the fruit physiognomies by the drupelet phenotype, as an early (4-5 years) adult or towards the end of the juvenility period revealed according to published descriptions. This trilateral approach achieves the confidence required to claim and classify a cultivar or cultivar phenodistinct substitute. Whereby the classification of such cultivars would be assessed through a trilateral consensus confirmation (TCC), as described in this work. The

youngest leaves, which were at the softest state, were collected from the tree crowns and used as sample sources for subsequent DNA extraction and SSR genotyping. Several leaflets averaging 10 - 15 cm in length, were harvested from each tree, and were bagged and placed in liquid nitrogen. These samples were then transported at –196 °C (–321 °F) and held at 4 °C at the Khalifa Center for Genetic Engineering and Biotechnology (KCGEB) and the Date Palm Tissue Culture Laboratory (DPTCL) of UAE university until the initiation of gDNA extraction.

Microsatellite profiles were developed post phylogenetic analysis and confirmed using samples from certified sources at the UAE Ministry of Agriculture, Water & Environment. Origin and geographical distribution of cultivars were sourced from several references (Cook, 1901 ; Costantini, 1985 ; Willcox, 1990 ; Beech, 2003 ; Ishida et al., 2003 ; Zohary et al., 2012).

2.2.3. Phenotypic Cultivar Association

The confirmation of cultivars was based on phenotypic characteristics fulfilling the trilateral cultivar confirmation system of date fruits from the fruiting mother plant accessions collected during the period 15 - 31 July 2013. Most were recollected from two major locations; Al Ain city region in the UAE and Al Hassa Date Palm Research Station in the Kingdom of Saudi Arabia, since these locations belonged to highly developed preservation collections and were considered national germplasm sites. This selection was based on known established collections from specific germplasm sources and with the technical assistance from date palm specialists of Al Ain / UAE Dates Factory. The leaflet tissues were stored in liquid nitrogen and the phenotypes were validated visually based on the fruits collected from the same trees during the fruiting season in late June to early August of 2013.

2.2.4. Phenotypic - Genotypic Context

The cultivar genotypes were verified using SSR microsatellite primers 15, 25, 32, and 85. As indicated previously, fruit characteristics were also evaluated along with control samples with assistance of a date palm specialist from the Al Ain Dates Factory. This served as an added validation of cultivar identification for use in the development of the trilateral cultivar confirmation system. The selected cultivars represent those available that were in high public demand due to their high quality and overall fruit characteristics. Their ability to be easily micropropagated on a large scale was a final criterion for their use in this study.

2.2.5. Genomic DNA Isolation

DNA was extracted for this study according to the methodology described by Doyle and Doyle, (1990) from freshly thawed date palm plant tissue. Modifications (noted below under 2.2.7.) to the protocols for DNA extraction were vital to optimize high yields of good quality DNA. Sample disruption was achieved using a TissueLyser II (QIAGEN, Hilden, Germany) with 5 mm stainless steel pellet beads. Quantification of the DNA was done on a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific, Walham, Massachusetts, USA) where the concentration and quality of each sample of nucleic acid was assessed. The DNA was then checked on a 1.5 % agarose gel for signs of denaturation. The DNA samples were also checked for signs of protein contamination using a 260/180 optical density (OD) ratio reading. A reading between 1.8 and 2.0 is regarded as being of sufficient purity. Readings were taken to provide an estimation of purity where the ratio of 260/280 nm was used. Samples with an OD ratio equaling or higher than the 260 / 280 ratio, and in the range of 2.0 to 2.2 was considered pure and of appropriate quality for this study.

Table 2.1. Locality information and mother plant germplasm sources of 20 commercially important date palm used to determine genetic relationships.

Cultivar Population	CP Code	Region	Collection sites | date	Latitude	Longitude	Origins
Aboumaan	ABM	United Arab Emirates	Al Deed | 2013	24°12'53.3"N	55°46'15.3"E	GCC
			Ras el Khaïmah | 2006	25°53'17.3"N	56°03'26.2"E	
			Al Ain Date Factory | 2006	24°13'16.0"N	55°47'34.4"E	
		Kingdom of Saudi Arabia	Al Ehsa | Ash-Sharqiyah | 2004	22°17'43.4"N	50°40'45.8"E	
			Al kharj | 2004	24°15'40.2"N	47°11'33.1"E	
Ajwa	AJW	United Arab Emirates	Ras el Khaïmah | 2006	25°48'02.5"N	55°58'34.3"E	KSA?
			Jabel Hafeet Palace | 2013	23°58'56.1"N	55°50'02.5"E	
		Kingdom of Saudi Arabia	Al Madina | 2004	24°35'59.0"N	39°33'08.5"E	
			Al Qassiem | 2006	26°02'11.6"N	44°00'13.2"E	
Barhee	BRH	United Arab Emirates	Al Ain Date Factory | 2006	24°13'16.0"N	55°47'34.4"E	Iraq
		Iraq	Basrah | 2004	30°32'16.9"N	47°50'18.9"E	
Chichi / Shishi	CHI	United Arab Emirates	Al Rawdah Palace | 2013	24°06'18.6"N	55°30'05.5"E	KSA
		Kingdom of Saudi Arabia	Al Qassiem | 2006	26°02'11.6"N	44°00'13.2"E	
Dabbass	DBS	United Arab Emirates	Liwa Private Collection | 2013	23°07'09.6"N	53°40'08.9"E	UAE
Fard Abyad (White)	FDW	United Arab Emirates	Al Ain Date Factory | 2013	24°13'16.0"N	55°47'34.4"E	UAE
		Kingdom of Saudi Arabia	Al Qassiem | 2006	26°02'11.6"N	44°00'13.2"E	
		Oman	Rustak | (Unknown)	23°26'10.8"N	57°26'26.5"E	
Helwa	HLW	United Arab Emirates	Al Kuweitat | Al Ain | 2013	24°12'59.4"N	55°46'06.9"E	KSA
		Kingdom of Saudi Arabia	Al Hassa Date Palm Research Station | 2006	25°33'46.5"N	49°36'23.6"E	
Khadri	KAD	United Arab Emirates	Al Ain Date Factory | 2013	24°13'16.0"N	55°47'34.4"E	KSA
		Kingdom of Saudi Arabia	Al Hassa Date Palm Research Station | 2006	25°33'46.5"N	49°36'23.6"E	
			Al Qassiem | 2004	26°02'11.6"N	44°00'13.2"E	
			Al Houfoof | 2004	25°17'28.9"N	49°38'1.107"E	
Khadraoui	KDR	United Arab Emirates	Al Rawdah Palace	24°06'18.6"N	55°30'05.5"E	Iraq
		Iraq	Basrah | 2004	30°32'16.9"N	47°50'18.9"E	
Khlass	KHL	United Arab Emirates	Al Ain Date Factory | 2013	24°13'16.0"N	55°47'34.4"E	KSA
		Kingdom of Saudi Arabia	Al Qassiem | 2006	26°02'11.6"N	44°00'13.2"E	
Khenezi	KNZ	United Arab Emirates	Al Ain Date Factory | 2013	24°13'16.0"N	55°47'34.4"E	KSA
		Kingdom of Saudi Arabia	Al Ehsa | Ash-Sharqiyah | 2004	22°17'43.4"N	50°40'45.8"E	
		Bahrain	Location unknown (central) ? | 2004	26°07'11.4"N	50°31'33.3"E	

Lulu	LUL	United Arab Emirates	Al Kuweitat | Al Ain | 2013	24°12'59.4"N	55°46'06.9"E	Qatar
			Ras el khaima | 2013	25°48'02.5"N	55°58'34.3"E	
		Bahrain	Location unknown ? | 2004	26°7'47.9''N	50°33'18.0''E	
Medjhool	MDJ	United Arab Emirates	Al Kuweitat Presidential Villa | Al Ain Oasis | 2013	24°12'59.4"N	55°46'06.9"E	Morocco
		Kingdom of Saudi Arabia	Al Hassa Date Palm Research Station | 2006	25°33'46.5"N	49°36'23.6"E	
		Morocco	Erfoud city (site a) | 2013	31°45'39.8"N	4°12'58.266"E	
			Erfoud city (site b) | 2013	31°37'0.807"N	4°12'57.98"E	
Maktoumi	MKT	United Arab Emirates	Al Kuweitat Presidential Villa | Al Ain Oasis | 2013	24°12'59.4"N	55°46'06.9"E	Iraq
		Iraq	Basrah | 2004	30°32'16.9"N	47°50'18.9"E	
Al Ain city male (Male)	MP	United Arab Emirates	Al Deed, Al Ain | 2013	24°12'53.3"N	55°46'15.3"E	Tunisia?
			Al Kuweitat | Al Ain | 2013	24°12'59.4"N	55°46'06.9"E	
Mdasry (Male)	MSR	United Arab Emirates	Liwa Private Collection | 2013	23°07'09.6"N	53°40'08.9"E	Oman
		Oman	Rustak | 2006	23°26'10.8"N	57°26'26.5"E	
			Samaiel | 2004	?	?	
Nabtat Seif	NSF	United Arab Emirates	Al Kuweitat | Al Ain | 2013	24°12'59.4"N	55°46'06.9"E	KSA
		Kingdom of Saudi Arabia	Al kharj | 2006	24°15'38.2"N	47°11'18.8"E	
			Al Qassiem | 2006	26°02'11.6"N	44°00'13.2"E	
			Al Houfoof | 2004	25°17'28.9"N	49°38'1.107"E	
Sakaii	SKA	Kingdom of Saudi Arabia	Al Hassa Date Palm Research Station | 2006	25°33'46.5"N	49°36'23.6"E	KSA
			Al Qassiem | 2006	26°02'11.6"N	44°00'13.2"E	
Sukkari	SKR	United Arab Emirates	Al Kuweitat | Al Ain | 2013	24°12'59.4"N	55°46'06.9"E	KSA
			Al Qassiem | 2006	26°02'11.6"N	44°00'13.2"E	
		Egypt	Specific location unsure (SW) | 2004	?	?	
Sultana	SUL	United Arab Emirates	Al Kuweitat | Al Ain | 2013	24°12'59.4"N	55°46'06.9"E	KSA
		Kingdom of Saudi Arabia	Al Hassa Date Palm Research Station | 2006	25°33'46.5"N	49°36'23.6"E	
			Al Qassiem | 2006	26°02'11.6"N	44°00'13.2"E	
			Al Ehsa | Ash-Sharqiyah | 2004	22°17'43.4"N	50°40'45.8"E	

For each of the 5 individuals chosen at random, 2 accessions were gathered. CP Code : Cultivar Population Code. ? : Origin uncertain (or conflicting info.). HEL / HLW. Information was verified by Dr. *Hassan Shabana*, Dr. *Samir Al Shaker* (both | Iraq), Dr. *Helal Al Kaabi* (UAE), and Dr. *Abdellah Meddich* (Morocco).

Few methods are available for isolating genomic DNA (gDNA) (Paterson et al., 1993; Kim et al., 1997; Allen et al., 2006) from the date palm; hence, a selection process took place to determine a reliable technique of isolating high quality gDNA from adult date palm tree samples.

The younger leaf samples were found to produce better quantities and quality of gDNA in contrast to older leaflets. A manual technique was used based on the high-quality DNA obtained as noted in the Doyle & Doyle (1990) modified protocol. This was a crucial step, as date palms leaflets are known to have an extremely fibrous structure, which in turn makes DNA extraction from leaves challenging.

2.2.6. Genotypic Evaluation

Genotyping was verified using SSR's developed by Billotte et al. (2004) with 5' modifications developed for this study (Table 2.2). This technique enabled one to screen unknown date palm samples (18 total), using specific subsets or panels of SSR primers, and accurately identifying up to 56 discernible cultivars, based on intraspecific polymorphisms observed in the following paper by Zaid S. E. (unpublished). Fourteen SSR primer pairs were assigned to 5 panels, as shown in Table 2.2. Each of the 20 selected cultivars mentioned in Table 2.1 was screened using the 5 panels. These panels were established to encompass and benefit from multiplexing and by avoiding overlapping 5' modifications while using 3rd generation fluorescent dyes; 6-FAM Blue, VIC Green, NED Yellow (Table 2.2).

2.2.7. Procedural Protocol for Date Palm DNA Extraction

Utensils used to cut leaf tissue were initially sterilized in an oven at 121 °C and with 70 % Ethanol routinely during handling of samples. The leaves, which had been initially placed in micro centrifuge tubes and kept in liquid nitrogen at -196 °C until arrival at the laboratory were then transferred and held at 4 °C. Small pieces of leaves, approximately 25 - 50 mg wet weight, were individually placed in 2.2 ml micro centrifuge tubes with two 5 mm stainless steel beads per tube along with 800 µl CTAB extraction buffer (2% CTAB; 1.4 M NaCl2; 0.02 M EDTA; 0.1 M TRIS

HCl; pH 8.0; sterilized by autoclaving) and 2.4 µl 2 - Mercaptoethanol. Micro centrifuge tubes were then placed in the TissueLyzer for 4 min 30 s (3 x 90 s cycles) at settings of 30 Hz, for sample disruption, followed by incubation in a water-bath at 60 °C for 60 min. During the incubation period 3 sets of micro centrifuge tubes of 2.2 ml, 1.5 ml, and 1.5 ml for each sample were prepared for the following steps.

After incubation, 800 µl 24:1 Chloroform: Isoamyl alcohol (CI) solution was added to each sample which was then subjected to gentle mixing by inversion and immediately centrifuged at 12,000 rpm for 8 min at room temperature. The DNA - containing upper supernatant phase, approximately 600 µl, was transferred to new micro - centrifuge tubes (avoiding any white precipitation or cell debris during separation) with 400 µl of 25:24:1 Phenol : Chloroform : Isoamyl Alcohol solution (PCI) added to each sample. These were then subjected to brief mixing by inversion and immediately centrifuged at 12,000 rpm for 2 min at room temperature. Following this, 500 µl of the upper supernatant phase was transferred to new micro centrifuge tubes and 500 µl CI solution was added to each sample followed by thorough mixing by inversion and immediate centrifugation at 12,000 rpm for 5 min at room temperature. A volume of 300 µl of the upper supernatant phase was transferred to new micro centrifuge tubes and an equal volume of Isopropanol was added to each sample and thoroughly mixed by inversion. The samples were then incubated at-20 °C for a minimum period of 1 hour to allow precipitation of the DNA. Following incubation, samples were centrifuged at 12,000 rpm for 10 min at 4 °C to pellet the DNA in a refrigerated centrifuge.

The next few steps required careful handling of each micro centrifuge tube to reduce loss of DNA. The aqueous supernatant was discarded carefully, making sure not to lose the DNA

pellet, and 1 ml of cold 70 % Ethanol was added to each sample followed by centrifugation at 12,000 rpm for 5 minutes at 4 °C. The fluid was discarded once again, taking care not to discard the DNA pellet, and each micro centrifuge tube was carefully blotted on a paper towel to allow the DNA pellets to dry. DNA pellets were then dissolved in 50 µl 1× TE Buffer (10 mM TRIS-HCl; 1 mM EDTA; pH 8.0) containing 40 µg / mL RNase A (DNase-free) and incubated for 30 min at 37 °C. After incubation, 5 µl (1 / 5 volume) of 3 M sodium acetate (pH 5.0), followed by 110 µl of 100 % ethanol, was added to each sample and mixed well by inversion. Samples were again centrifuged at 12,000 rpm for 10 min at room temperature.

At this stage of the protocol, it was vital that each micro-centrifuge tube sample was handled with care when discarding the aqueous solution. The supernatant was poured off carefully as some pellets were still loose. The DNA samples were further purified by adding 1 ml 70 % Ethanol and centrifuged at 12,000 rpm for 5 min at room temperature. This purification step was then repeated. Finally, the supernatant fluid was carefully discarded and the samples blotted on paper towels and placed in an oven for 5 min at 50 °C to dry the pellets. The DNA pellets were redissolved in 50 µl sterile MilliQ Water (sterilized triple distilled H2O). The genomic DNA samples were then kept at 4 °C for immediate use or placed at - 20 °C for long term storage.

2.2.8. Primer Selection and Testing

According to the Oligonucleotide primer sequences sourced by Billotte et al. (2004), 14 specific primers were selected with respect to non-overlapping allelic ranges, number of alleles detected on certain loci, and heterozygosity. Compatibility of primers was important due to the use of dyes and the amplification success, as certain primers with specific cultivars were

overlapping. These primers were custom ordered (Alpha DNA, Montreal, Quebec, Canada) and grouped into 5 panels with respect to their 5' modification labeling.

Forward primers for all microsatellite markers mPdCIR0 (10, 15, 16, 25, 35, 44, 48, 57, 90, and 93) were 5'- labeled with 3rd generation fluorescent dyes; 6 - carboxyfluorescein (6 - FAM) is the most commonly used fluorescent dye attachment for oligonucleotides and is compatible with most fluorescence detection equipment. At a pH below 7 it becomes protonated and has decreased fluorescence. Therefore, it is typically used in the pH range of 7.5 - 8.5. FAM can be attached to the 5' or 3' end of oligonucleotides. With the exception of forward primers for mPdCIR032 and mPdCIR085, the primers were 5'- labeled with 6 - carboxyrhodamine (VIC) while mPdCIR070 and mPdCIR078 were 5'- labeled with benzofluorotrichlorocarboxy - fluorescein (NED) (Applied Biosystems) (Table 2.2). The primer mixture for each panel for the PCR reaction totaled a volume of 100 µl. Each forward and reverse primer for each panel was added (2 µl for a total of 4 µl per microsatellite) to a 1.5 ml micro - centrifuge. Panel 1 and 2 consisted of eight primers of four microsatellites, which required 84 µl of MilliQ water to be placed in the prepared tubes. Panels 3, 4, and 5 consisted of four primers of two microsatellites, which required 92 µl MilliQ water per tube. While the ordered primers were defrosting, five 1.5 ml micro centrifuge tubes where prepared and dispensed with the appropriate amount of MilliQ water as described above. Once ready the forward and reverse primers were transferred to the appropriate five micro centrifuge tubes and subjected to mixing by inversion and labeled, accordingly. The testing of panels 1 through 5, took place in a sterile setting by which in pre-study verification, many tissue culture leaflet samples were readily examined across the 14 microsatellites applied in the study.

Table 2.2. The allelic ranges, optimal annealing temperatures, oligonucleotide sequences, panel assignments, 5' florescent dye used and the characteristics in *P. dactylifera* for 14 SSR primers.

		Characteristics in P. dactylifera					
Panel #	SSR Locus	Allelic Range (bp)	Optimal Ta (°C)	Ho	No. of alleles	Primer Sequences (5' - 3')	Modification (5')
1	mPdCIR015	120–156	51.6	0.68	12	F: AGCTGGCTCCTCCCTTCTTA R: GCTCGGTTGGACTTGTTCT	6-FAM
	mPdCIR025	199–231	49.3	0.57	6	F: GCACGAGAAGGCTTATAGT R: CCCCTCATTAGGATTCTAC	6-FAM
	mPdCIR032	284–305	51.5	0.72	12	F: CAAATCTTTGCCGTGAG R: GGTGTGGAGTAATCATGTAGTAG	VIC
	mPdCIR085	152–183	50.4	0.56	18	F: GAGAGAGGGTGGTGTTATT R: TTCATCCAGAACCACAGTA	VIC
2	mPdCIR044	281–332	51.7	0.16	6	F:ATGCGGACTACACTATTCTAC R: GGTGATTGACTTTCTTTGAG	6-FAM
	mPdCIR048	156–192	51.4	0.51	12	F: CGAGACCTACCTTCAACAAA R: CCACCAACCAAATCAAACAC	6-FAM
	mPdCIR070	182–208	48.7	0.20	12	F: CAAGACCCAAGGCTAAC R: GGAGGTGGCTTTGTAGTAT	NED
	mPdCIR078	117–152	49.6	0.78	14	F: TGGATTTCCATTGTGAG R: CCCGAAGAGACGCTATT	NED
3	mPdCIR010	118–161	55.9	0.72	13	F: ACCCCGGACGTGAGGTG R: CGTCGATCTCCTCCTTTGTCTC	6-FAM
	mPdCIR035	175–221	53.9	0.37	11	F: ACAAACGGCGATGGGATTAC R: CCGCAGCTCACCTCTTCTAT	6-FAM
4	mPdCIR057	251–278	55.4	0.59	7	F: AAGCAGCAGCCCTTCCGTAG R: GTTCTCACTCGCCCAAAAATAC	6-FAM
	mPdCIR093	153–184	51.8	0.71	15	F: CCATTTATCATTCCCTCTCTTG R: CTTGGTAGCTGCGTTTCTTG	6-FAM
5	mPdCIR016	130–138	51.7	0.60	5	F: AGCGGGAAATGAAAAGGTAT R: ATGAAAACGTGCCAAATGTC	6-FAM
	mPdCIR090	142–175	48.6	0.59	10	F: GCAGTCAGTCCCTCATA R: TGCTTGTAGCCCTTCAG	6-FAM

The panel numbers illustrate date palm SSR loci categorized for the reactions. According to CIRAD's standard (mPdCIRxxx) the simple sequence repeat (SSR) loci are represented as primer sequences with the appropriate 5' modification, synthesized by Alpha DNA (Montreal, Canada); *m* corresponds to microsatellite, *Pd* to *P. dactylifera* L. and *CIR* to CIRAD. Modification (5') indicate the 3rd generation fluorescent dyes are; 6-FAM Blue, VIC Green, NED Yellow. Source; (Billotte *et al*. (2004)).

2.2.9. Polymerase chain reaction cycling

All PCR amplifications were performed in an Applied Biosystems 2720 96 - well thermal cycler (Life Technologies, Carlsbad, California, USA) under the adjusted optimized conditions as per Billotte et al. (2004). For each reaction, 50 ng of DNA was used totaling a volume of 10 μL. A homogenous master mix used for all reactions consisted of 1 × NH_4 PCR Buffer (BIOLINE, London, UK); 200 μM dNTP's (BIOLINE, London, UK); 0.2 μM of forward (labeled) and reverse (unlabeled) primer for each microsatellite mPdCIR0 (10, 15, 16, 25, 32, 35, 44, 48, 57, 70, 78, 85, 90, and 93); 3 mM of $MgCl_2$ (BIOLINE, London, UK); and 0.5 U of BIOTAQ™ DNA Polymerase (BIOLINE, London, UK). The thermal cycler settings used: an initial denaturation at 95 °C for 5 minutes followed by 35 cycles of denaturation at 94 °C for 30 seconds with primer annealing at 52 °C for 90 seconds and elongation at 72 °C for 60 seconds which was followed by a final elongation step at 72 °C for 30 minutes. The final soak temperature was set at 4 °C since a low soaking temperature assures high quality DNA while reducing risk of degradation.

2.2.10. SSR Capillary Sequencing

An Applied Biosystems 3500 Genetic Analyzer with 8 Capillaries, a Dell workstation including Data Collection and GeneMapper software, as well as accompanying installation kit instructions were used in collecting data and for initial analysis. The components used for analysis were all Applied Biosystems consumables & reagents for the 3500 Genetic Analyzer (for SSRs / microsatellites / fragment analysis applications)-Including: POP - 7 polymer, 8 × 50 cm capillary arrays; anode, cathode & septa buffer containers; conditioning reagent; 96 - well retainers & base; 96 - well septa; Microamp 96 well reaction plates; Hi-Di formamide; DS-33 (Dye set G5) Matrix Standard Kit; GeneScan 600LIZ Size Standard; and sequencing with installed standard Big Dye Terminator v. 3.1. The data was verified by re-sequencing at a separate institution for accurate

authentication; 'University of Stellenbosch, Central Analytical Facility: DNA Sequencing Facility, Department of Genetics - Lab B702, Matieland, Stellenbosch, 7602, R.S.A (South Africa)'.

2.2.11. Microsatellite Data Analysis Technique

The allele sizes were assigned into their appropriate allele "bins" using overall standards from Billotte et al. (2004). High - throughput fluorescent genotyping requires a considerable amount of automation for accurate and efficient processing of genetic markers. We observed variation in allelic ranges and calls associated with some microsatellites by some cultivars when compared to Billotte et al. (2004). This was resolved by re-running in triplicate and going through all amplifications and manually adjusting those that were miscalled or outside the allelic range.

2.2.12. Cultivar Phylogeny Relatedness Analysis

The data was primarily assessed by Applied Biosystems' 3500 Series Data Collection Software v. 3.1 to compile the output from the genetic capillary analyzer with examination by Gene Mapper v. 4.1, during the control sampling evaluations and initial experimental stages to infer richness and feasibility of data pool / samples. As part of the genetic diversity evaluation, STRUCTURE v. 2.3 software (Pritchard et al., 2000; Rosenberg et al., 2001; Rosenberg et al., 2002; Porras-Hurtado et al., 2013) was used to determine population arrangement among samples besides genotype assignments to subpopulations | sub - clusters. With increasing K (population number) by multi locus allele frequency data values from 2 to 20 running with 100,000 burn-ins, 300,000 iterations and 10 repeats, to identify K, without using information of geographic background (collection sites; no loc Prior (Hubisz, et al., 2009)). The K value with the highest ln

P (D) values were calculated following Evanno et al., (2005). Evanno's method allows identification of single K values, out of a range of K values, capturing the uppermost level of structure. Testing phases were performed by trying various combinations to determine the number of gene pools represented in the complete data set of 285 individuals. All figures were correlated with a color-coded system, uniform throughout and by key specification, for ease of relationship association. Furthermore, ln (Pr (X | K) values were used to identify the K value for which Pr (K = k) is the highest (as described in STRUCTURE Manual 'Pritchard') which was calculated by CLUMPAK (Kopelman et al., 2015). The graphical displays by STRUCTURE were generated using DISTRUCT v. 1.1 (NA Rosenberg, 2004), and color altered by Inkscape (xquartz 2.711/ xorg-server 1.18.4).

For the establishment of phylogeny and relationship by distance in population sets, the software DARwin v. 5.0 (Perrier and Flori, 2003) was used to construct two Neighbor-Joining (NJ) trees and the assembly of a Multiple Factor Analysis (MFA). K = 1 to k = 20 were observed with the best K produced and chosen by the Evanno test (Evanno et al., 2005). In addition, the best K chosen Prob (K) by the Pritchard test was also produced.

2.2.13. Genetic structure and diversity analysis

Genetic diversity indices were compared among the 20 populations identified by STRUCTURE and CLUMPAK (http://clumpak.tau.ac.il) and cross referenced by POPTREE 2's phylogenetic tree initially produced for data verification and testing of primers (Figure 2.4). Observed heterozygosity (Ho), expected heterozygosity (He), genetic differentiation between populations (FST), and molecular variance analysis (AMOVA) (Michalakis and Excoffier, 1996)

including fixation indices such as FIS and FIT were performed using the program GenAIEx v. 6.5 (Peakall and Smouse, 2012). The population's genetic structure from various geographic backgrounds and across selected cultivar types (the most common in the plant material production by tissue culture facilities and of highly demanded markets) was explored using analysis of molecular variance (AMOVA). The statistical significance of each variance component was assessed based on 9,999 permutations of the data sets assessing the significance. The genetic diversity indices were calculated among the clusters identified by STRUCTURE.

2.2.14. Factorial Analysis

The Multiple Factor Analysis (MFA) results were produced; inertia of each axis (axe) was calculated and correlated with the Eigenvalues for each separate axis. The different relationships among samples were shown by MFA results. The MFA was done using the set of 285 palm samples with accessions representing 20 cultivar collection populations (Axis 1, 2, and 5 were used). The separation of populations and distribution by Principle Component (PCA) was observable in axis 1 by 2, whereby values that were closest to x-axis represent higher force and are better reported. Axis PCA combinations were produced and are available in Appendices (App 8.1.2).

2.2.15. Statistical Analysis

Using software GenAIEx, the following calculations were performed: allele frequencies heterozygosity, F statistics, Nei's Genetic Distance and Identity, and analysis of molecular variance (AMOVA), additional results can be found in 'Supplemental Materials' section (Appendices; 4.1, 4.4, 4.5, 4.6, 4.7, 4.8).

2.2.16. Phylogenetic Tree Construction

Genetic distances between each pair of accessions were calculated by measuring the shared allele frequencies. A dissimilarity matrix was built in order to calculate the genetic distance between accessions by using the SSR set of each analysis. Two Neighbor Joining trees (Da & dmyu) were constructed using the bootstrap algorithm. Color coding used in NJ trees correspond to samples from clustered populations found by STRUCTURE analysis, uniform in all results figures and tables and concurrent among figure keys. The outputs of POPTREE 2 and DARwin were correlated and confirmed matching results.

The neighbor-joining tree was constructed using the allele frequency data by using the neighbor-joining (NJ) method (Saitou and Nei 1987) and Bootstrap (Felsenstein 1985) algorithm. To better visualize populations, a bootstrap higher than 50% was used. In the NJ method (Saitou and Nei 1987), starting from a star-tree (all branches are connected to one node | non-rooted), a pair of taxa (populations) which gives the smallest sum of branch lengths are combined into a cluster and form a composite group or 1 of 6 sub clusters / clades. This process is repeated until an unrooted tree is produced. The branch lengths are computed by the least-squares method in each step (Takezaki and Nei, 1996). The measures of genetic distance (D_A distance) used among cultivars and their populations were calculated as defined by Nei et al. (1983):

$$D_A = 1 - \frac{1}{r}\sum_j^r \sum_i^{m_j} \sqrt{x_{ij} y_{ij}}$$

Where x_{ij} and y_{ij} are the frequencies of the i-th allele at the j-th locus in populations X and Y, respectively, mj is the number of alleles at the j-th locus, and r is the number of loci used.

The angle (qj) of the two or n populations is given by:

$$\cos\theta_j = \sum_i^{m_j} \sqrt{x_{ij} y_{ij}}\,.$$

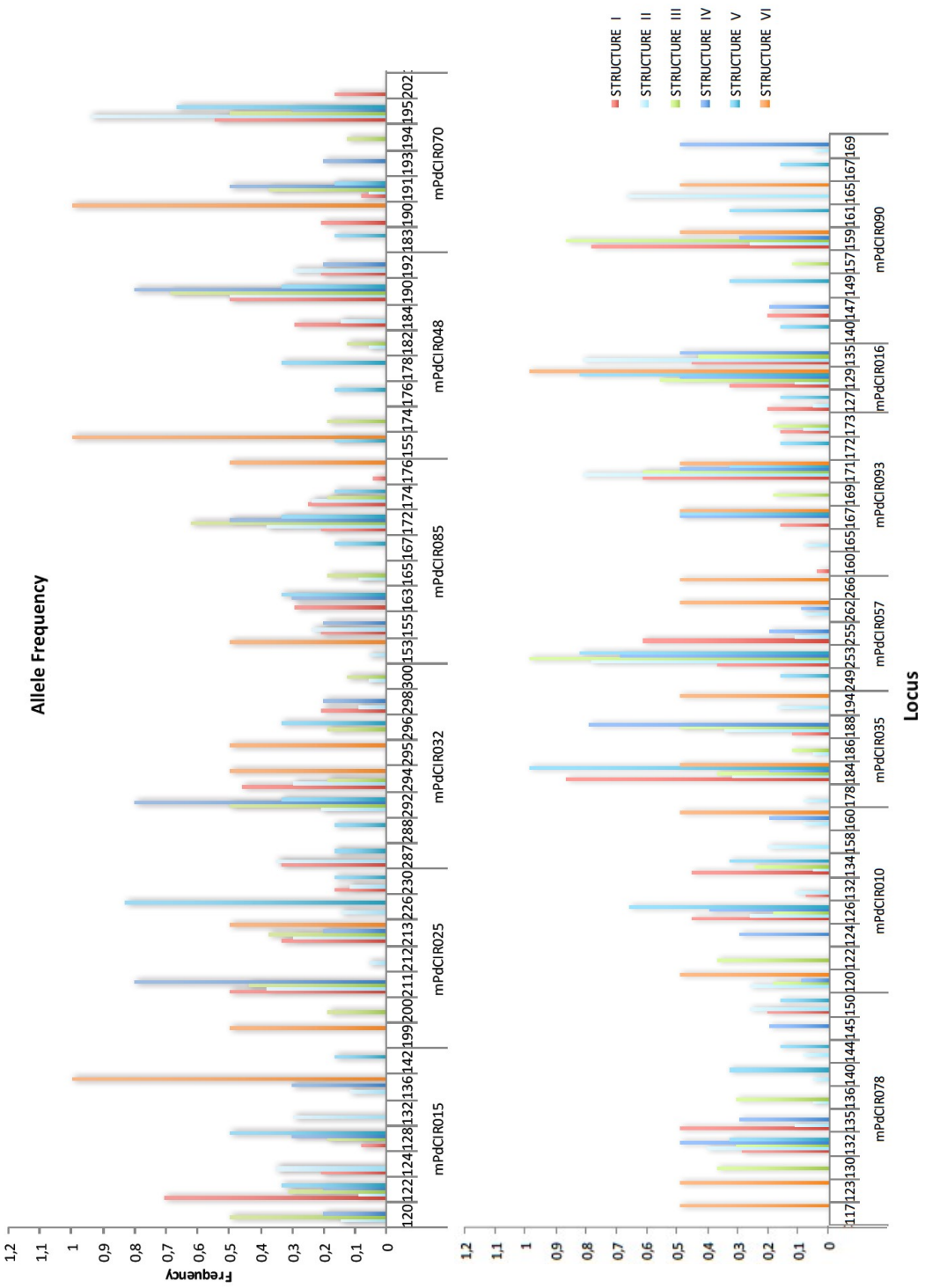
In addition, using STRUCTURE analysis, populations are separated by scaling color, respectively in reference to the NJ tree & PCA results. As a means to verify data from STRUCTURE, the program POPTREE 2 software was used along with data collection software v. 3.1 and analysis carried out by Gene Mapper v. 4.1. DARwin software was also used in reproducing results for phylogenetic relationship and factorial coordinates (PCA). This was done initially as a verification of cultivar candidates and whether the SSR markers worked, followed by 3 repeats to produce allelic calls of each cultivar genotype individual collected.

2.3. Results and discussion

Morphological characteristics that are vegetative are similar in most cultivars which makes it difficult to differentiate one cultivar from another in date palms. The specific date palm individuals used in this study were from a limited area in the Middle East and shared similar stable growing environments. Relatively few cultivars made their way East and West when considering the approximately 3000 cultivars and these maintained their phenotypic etiology. Commercial cultivar sample groups demonstrated close relationships while high polymorphism was observed among populations. (This was shown through a degree of genetic admixture shared gene pool lineages into a cultivar group).

There was considerable polymorphism observed with the 13 microsatellite markers used in this research. The relationships revealed genotypes of fifteen arrays among the cultivars (Figures 2.1 and 2.3). The primers that were developed from Billotte et al. (2004) did not always fall between the allelic ranges established by his team. Hence, an assessment of amplification quality and congruency in verifying the output of Gene Mapper v 4.1 was required. By re-running in triplicate and going through all amplifications and manually adjusting those that were miscalled or outside the allelic range. The difference could have been the result of the limitations of Nuclear Microsatellite Markers or Simple Sequence Repeats (SSRs) vs. Single-Nucleotide Polymorphisms (SNPs) as well as the presence of additional microsatellite alleles. This could be the result of the inherent susceptibility of the affected areas of such alterations and from mutations in the DNA mismatch repair mechanism, that would normally be resolved if gene chips or Microsatellite Instability linkages (MSI) were used. The limitations of SSRs do exist although genotyping by inferred amplification of a locus flanked by PCR primers could be a contributing factor as the primers were multiplexed, causing such mismatched results in the bp calls.

The analysis of allele frequency per locus in structure populations II, III, and IV, showed considerable variation (Figure 2.1), that it provided suitable insight among the cultivar types for further clarity of the gene pool distribution. Within bp call amplification comparisons the SSR primers among individuals displayed less dissimilarity, there was some discrepancy when it came to some of the sample group individuals. This was attributed to cultivar distinct substitutes or intra varietal diversity, within the cultivar. The need for additional SSR development is important to broaden the spectrum of distinction, STRUCTURE and DARwin programs, to differentiate better among individuals, loci, alleles, in reporting their assigned clustering.



For each Locus, the microsatellite allelic call amplification and the frequency observed across the collected samples is shown. Data are frequencies of allele counts and total count by population. Color-coding of genetic pool population clustering STRUCTURE I-VI is uniform in all figures and tables and concurrent among figure keys.

Figure 2.1. Allelic frequency and amplification via resultant STRUCTURE populations over loci by number of alleles for 285 palms in 6 clusters across 20 cultivars.

2.3.1. Genetic diversity estimates

All 285 collected accessions were included in this study of commercial varieties of *P. dactylifera* (Table 2.1). We initially began with 14 microsatellite markers. During the initial evaluations we determined sequence marker allele 'mPdCIR044' and was unable to amplify sufficiently across most of the varietal types and it was then removed from the investigation. Using 13 SSRs, a total of 92 alleles were detected across all accessions (Figure 2.1). We desisted from allocating into two major groups, from North Africa and Middle East regions. While the commercial context is present but understudied in previous assessments using AFLP's, it was incapable of discerning admixture presence among cultivars and allelic frequencies. Allelic frequency by population and locus from all clusters of STRUCTURE I - IV was observed in loci mPdCIR085 and mPdCIR078. The greatest frequencies observed for loci mPdCIR0; 70, 35, 16, 57, 48, and 15 under STRUCTURE were for populations III, V, and VI. There was a significant disparity in allelic frequency, on average 0.60, of STRUCTURE as observed in clusters assigned to population VI across alleles 136, 155, 190, and 129. This further confirmed the separation of the North African and Middle Eastern accessions due to the strong electropherogram calls obtained by the allelic frequency correlated with specific loci that fell within either of the two populations. Results from AGF calculations of allele frequencies with graphs by population and locus for codominant data showed alleles across all six STRUCTURE cluster populations. Exhibited substantial locus expression at 122, 211, 213, 292, 172, 190, 195, 132, 126, 184, 253, 171, 129, 135, and 159 on average generating a 0.653 allelic frequency the diversity within a locus, i.e. number of alleles per locus which will allow you to interpret the number of genotypes per locus (Figure 2.1). These alleles representing bp amplification measures are to be considered important in future breeding by alleles found in some cultivars and identifying genetic diversity of a particular group. Alleles found in all 6 populations were 184, 171, and 129 found in loci

mPdCIR035, mPdCIR093, and mPdCIR016 respectively. These are of great interest for further relationship determination studies and display a common relationship of sequences. The allelic richness of markers loci mPdCIR078 and mPdCIR090 produced the greatest number of alleles with 10 and 9 respectively. The remaining eleven ranged from 8 to 3 with the fewest found in mPdCIR016. This locus is found only in a few cultivars (App 8.1.1). The limited sampling and number of SSRs used in this study indicated a need for a broader population of accessions per site and more than 5 trees per cultivar from various areas throughout the growing world of date palm. This can also be explained by the ability of STRCUTURE to detect certain locus clusters of individuals at different levels of dispersal among the group clusters. A robust genetic differentiation made by AMOVA, partitioning the structure and producing gene pool populations by molecular variance (overall fixation indices F'ST of 0.581 among cultivars and molecular variance of 77% within all individuals (285) while a lower 23% among 6 gene pool populations) (*Table 2.3)*. The AMOVA framework is theoretically related to K-means clustering of the results by the Evanno method we produced.

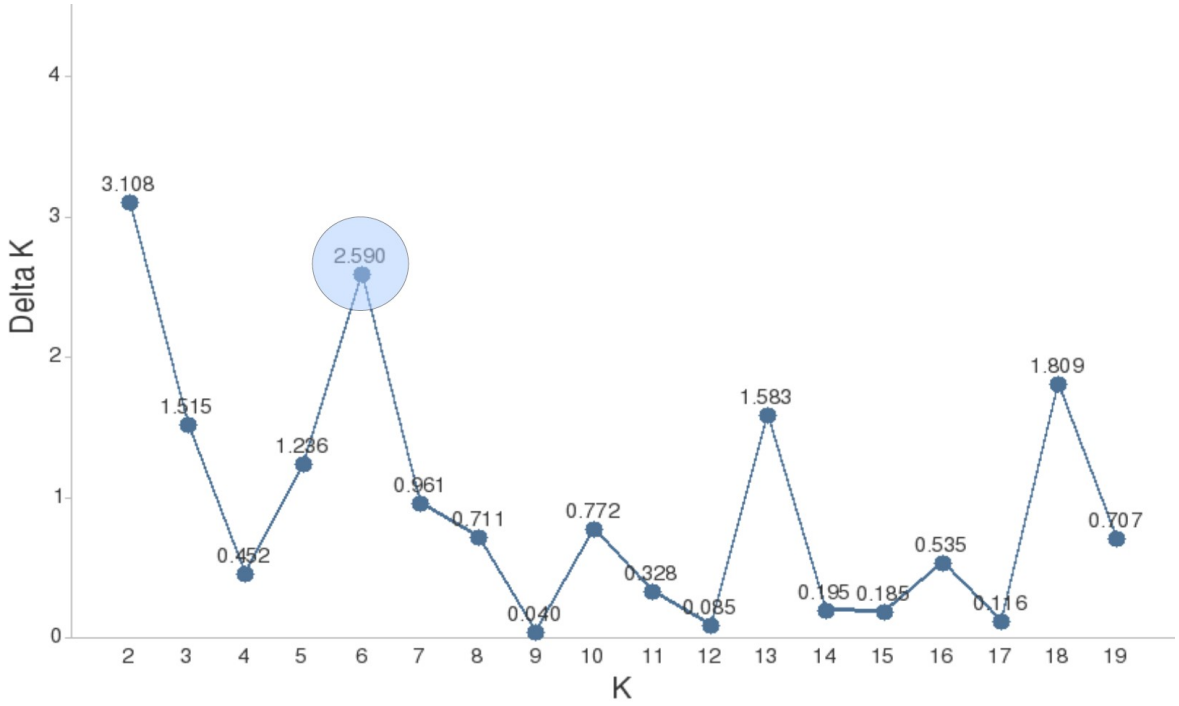
2.3.2. Bayesian clustering analysis

In deducing population structure among the 285 date palm samples, two methods were useful to detect and examine the true K estimated posterior probability of the SSR data. These were the maximum value of ln P (D) by the Pritchard method (Pritchard et al., 2000) and ΔK by Evanno (Evanno et al., 2005). This is assuming no specific geographic origin, the multi-locus allele frequency data can be used to identify K and thus aid in determining the number of gene pool clusters that may exist in the SSR data.

The detection of the number of clusters of individuals using STRUCTURE software required the estimation of K. The true number of populations returned by STRUCTURE was then identified by the maximum value of L (K). The best optimal delta K obtained was K = 6 at 2.590 (Figure 2. 1). Although K = 2 at 3.108 was produced by way of general best K, it did not fully resonate with the population and geographical background data for the genetic diversity inquiry and was confirmed as reasonable in the substructuring genetic pool of the populations and known historical records of the dates. Only K = 1 to K = 10 were cross checked and observed in vectorial format. However, the K = 1 to K = 20 were observed in vectorial format, K = 2 to K = 10 are shown in 'supplemental material' section (App 8.1.9). The maximum value of ln P (D) in structure output was also obtained. However, the calculations of the best *K* based on variation in the likelihood across multiple runs did not indicate a single probability (*K)* value except at K = 18 (App 8.1.10).

The realistic number of populations (K) is generally understood to be identifiable using the maximal value of L (K) returned by the STRUCTURE program and Pritchard method (Ciofi et al. 2002). Although as observed in common cases using Pritchard's method, once the true K is reached, the Prob (K) in this study had an initial base plateaued at 0 until K = 18, yielding a 0.999. This phenomenon is not mentioned in STRUCTURE's manual (Evanno et al., 2005). Using the Delta K Evanno's method, the rational surrounding these two approaches was to make salient the break in slope of distribution of L (K) as the true K. This in turn confirmed that the Evanno test best K results were deemed more informative for assessing population structure across the cultivar populations, indicating the strength of the signal detected by STRUCTURE.

As Evanno et al. (2005) stated; the distribution of delta K almost always showed a mode at the real K. The clustering algorithm by STRUCTURE categorized clusters set 1 - 4 (Structure I - IV) inferring the gene pools on the basis of membership coefficient (q), that were collected from geographically one main region, the central middle East (inclusive of the southern region of the Fertile Crescent) (Table 1) spanning from most northern 34°42'27.8"N 41°20'19.4"E to southern 14°11'21.1"N 44°47'30.0"E and share a good level of admixture from Structure I (red) mostly among the sub varietal groups (Figure 2.3).



Computed as ΔK = mean(|L″(K)|)/stdev[L(K)]. There appears to be six clusters based on geographical correlation of the collected samples. The modal value of this distribution is the true K or the uppermost level of structure.

Figure 2.2. Delta K plot output implemented in STRUCTURE program over 20 runs detecting the true K using all 285 individuals and 13 SSR loci.

The six gene pools produced by STRUCTURE were assigned by estimating relationships among individuals based on estimated membership coefficient (q) of common relations from each gene pool based on allelic data (Figure 2.3).

The majority of individuals from the 20 cultivars. conveyed admixture from Structure clusters I and III with various subcluster assignments in trees belonging to; Al Ain City Male Palm (MP), Barhee (BRH), Dabbass (DBS), and Helwa (HEL, HLW) cultivar populations (Figure 2.3). The exceptions were Chichi / Shishi (CHI), Sukkari (SKR), and Khadraoui (KDR) which demonstrated multiple admixtures but comparatively minimal as the gene pool disparity in varietal populations from geographical migration across time existed among these early populations.

All sampled trees that originated in Northwest Africa, were assigned to the gene pool cluster STRUCTURE VI (Figure 2.3, pink). These cultivars were confirmed as Medjhool, Ajwa, and MP (2 females and 1 male respectively) in the sub-cluster by distance (Da) neighbor-joining matrix analysis (Figure 2.5 and 2.6).

2.3.3. Marker function among Phoenix *genera*

The 14 microsatellites initially tested revealed confirmable amplifications as observed by Billotte et al. (2004) in the screening phase and in the development of the 5 panels in multiplexing for genotyping the cultivar samples collected. Some of the amplified products observed were dissimilar than Billotte et al.'s (2004) observations. The expected allele sizes did not always match those previously observed with some falling slightly outside the allelic range. The one SSR primer that failed to generate amplifications across the cultivar sample populations was mPdCIR044 as

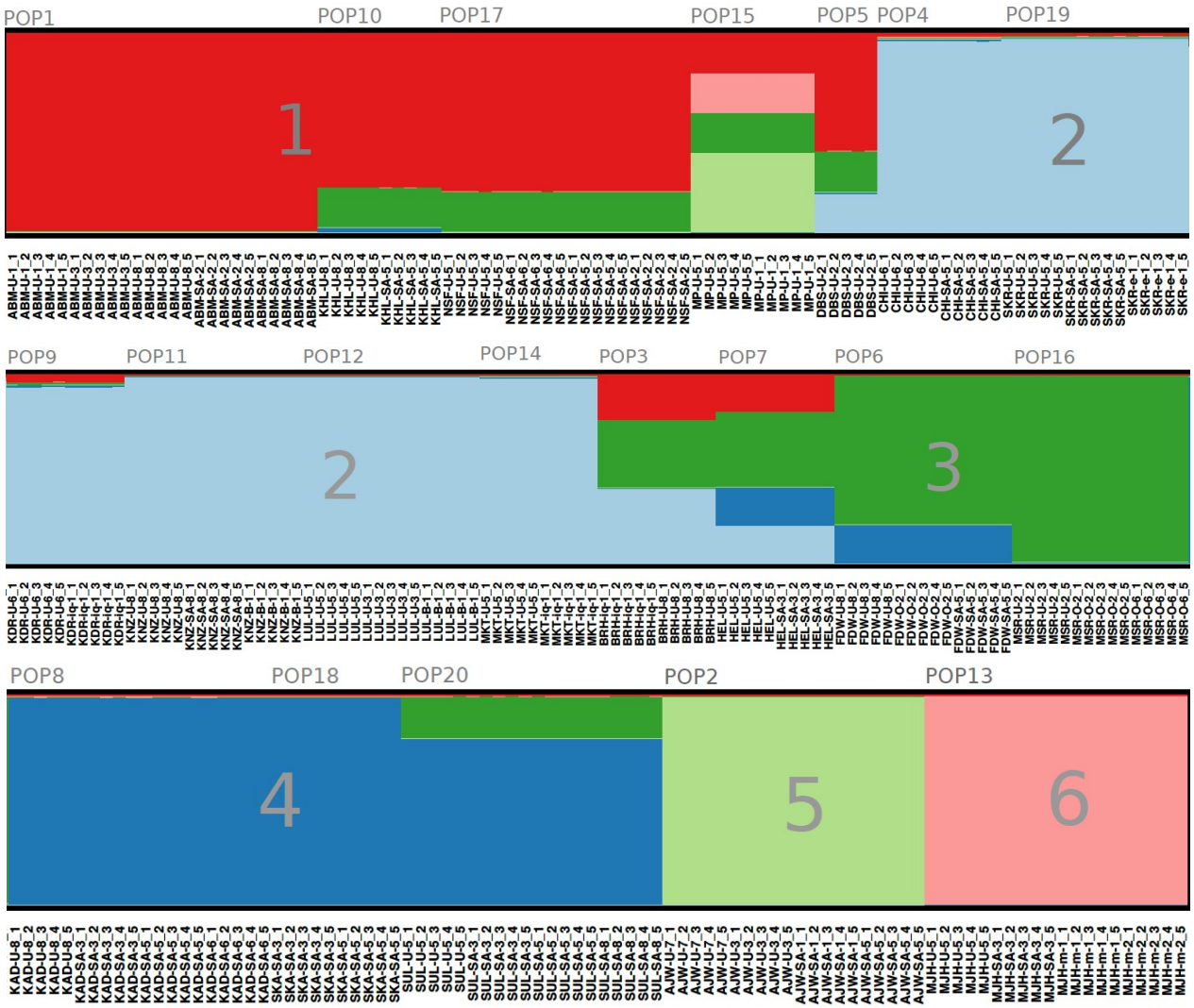


Figure 2.3. Results of Bayesian cluster analysis of 20 cultivar genotypes at 13 microsatellite loci implemented in STRUCTURE for K = 6 population stratification, representing genetic relatedness between palm accessions. Each bar represents a date palm individual; each color represents a distinct gene pool cluster inferred from the analysis. The 1 - 20 POP, denote cultivar population samples originally collected.

previously mentioned. This may have been due to the marker development and testing across several (var. substitute) types as well as the inability of Gene Mapper software to properly distinguish the amplifications, since overlapping was not always observed but did exist in certain cultivar sample accessions. There were false calls produced with marker mPdCIR044's evaluation. This marker was thereby removed from subsequent statistical analysis and phylogenetic

correlations. These results indicate a similar pattern seen in microsatellite development whereby the spectrum of limited trials and the number of markers produced is unaccounted in the available span of the possible coverage of genotype data and software compatibility that may be used reliably. The markers were multiplexed resulting in 5 group panels since there was a relatively large sample count of 285 palms. Since this work was completed, there have been emergent advanced markers using SNPs and NGS approaches. These approaches had a broader ability to distinguish further when compared to the SSRs used in this study. This is by no means a hindrance since SSR gDNA applied sequences are composed of repeated iterations of 1-6 motifs of oligonucleotides and allowed for the improved functioning or structural gene pool clustering and discrimination from other type of molecular markers.

2.3.4. Partitioning of genetic structure and diversity

The geographic areas and population boundaries of each date palm cultivar group was collected and were used with their derived polymorphisms from SSR amplification data in calculating variation using GenAlEx v. 6.3 and 6.5 software (as a repeat confirmation). At 9 999 permutations by allelic distance matrix, F-statistics was used in analysis of molecular variance (AMOVA), sufficiently testing for significance. A fitting differentiation is revealed among the 6 gene pool group populations (22.57% molecular variance) as these 6 subclusters reveal a greater geographic isolation of cultivars sampled. A majority of the differentiation of 77.43 % was within the 285 individuals was observed with respect to the allocated cultivar groups (Table 2.3). The estimated variation coupled with the phylogenetic association (Table 2.4 and Figure 2.5) by neighbor joining distances (Da) indicates that the date palms may have only one origin, as indicated historically. Assuming, with respect to the admixtures present at delta K = 6, the

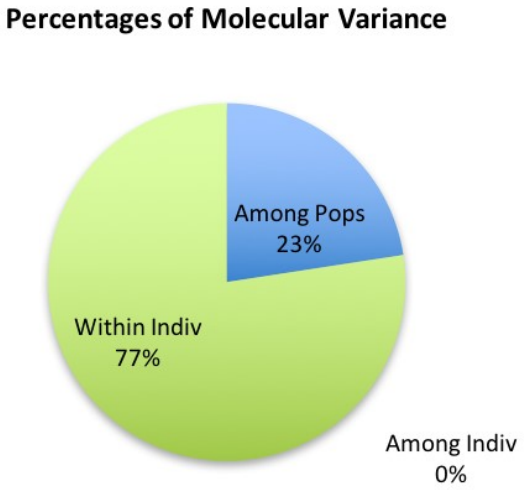
observation of some overlap among the 6 sub clusters would indicate that there exist some feral escapes or hybridization among the clusters, particularly in the two major sub clusters of *Phoenix*. The North African sub group gene pools (structure | cluster 5 and 6, Figure 2.3) are perhaps hybrids consisting of small fragments of admixtures of the dates from the Middle Eastern sub-continent (Figure 2.3). The delta K = 2 result is indicative of the two major subpopulations of the species which is reported in other published works (Figure 2.2). Medjhool, MP, and Ajwa were separated long ago from the Middle East grouping.

Molecular variance among individuals of the same cultivar type produced 0% since initial, subsequent analysis, and final results showed that the cultivar accession populations were genetically identical while the phenotypic traits noted at the collection stage directed towards a form of cultivar phenodistinct substitute phenomenon at the fruitlet level (*Table* 2.3). These produce various subtle characteristics which are most likely genotype by environmental (G X E) interaction; although some may be known effects of cultivation practices or even possibly retrotransposons (Wei and Cao, 2016; Xiao et al., 2008; Goulet et. al., 2012; Tam et al., 2005). These results suggest that differentiation among cultivars, isolated to large regions is smaller than thought across the types presently understood. The Da matrix was rerun using DARwin by dissimilarity of allelic data. The re-run bootstrapping replicates within the phylogenetic analysis reported the majority at 100 as well-supported.

Table 2.3. AMOVA, molecular variance separation by genetic partitioning of 20 cultivars *of P. dactylifera* L. segregated across 6 gene pool populations among 285 palms from the Middle Eastern and Norwest African areas.

		Variance components				
Source of variation	df	SS	MS	Estimated variation	% variation	P - value
Among 6 gene pool Populations	5	606.96	121.39	1.30	23	0.001
Among 285 (per) cultivar individuals	279	669.58	2.40	0.00	0	1.000
Within 285 individuals	285	1275.00	4.47	4.47	77	0.002

Processed through 9 999 permutations. The degrees of freedom (df), variance components SS (sum of squares) MS (SS, MS), estimated variation, percent variation, and P-values for the AMOVA screening are specified. The est. variation for among cultivar individuals of -1,03688734737434 was converted to zero for pie chart.



2.3.5. DA (distance) - grouping cluster evaluation

The 6 subclusters produced through the STRUCTURE program and Clustering Markov Packager Across K (CLUMPAK) were confirmed by neighbor-joining dendrograms ("Da": DA distance and "Dmyu": *(δμ)^2* distance) (Figure 2.6), with the cultivar populations segregating with respect to their perceived origins (reference Table 2.1). The clusters of closely related clones or those sharing the closest ancestral parent in distance and divergence made up each of the 6 subclusters. Generally, most cultivars are divided into 2 broad subpopulations with one from western North African consisting of 2 subclusters (STRUCTURE Cluster V and VI). Medjhool, Ajwa, and MP (Figures 2.4 and 2.6) made up cluster VI. Within the larger subpopulation consisting of 85 % of the cultivars studied, 4 clusters were determined which are generally thought to be from the Middle Eastern sub population. Only 15 % of the cultivars studied were observed

to be related to the western subpopulation. These are generally thought to be from geographic localities in the west which are historically believed to have moved eastward. A second larger group showed a successive distribution observed by isolating Structure clusters I and II (red and light blue, Figure 2.5) as the most distant and divergently resultant group 1 (AJW, MJH, and MP) localized from the Western North African major subpopulation. Overall, 6 gene pool clusters from 20 populations were resolved and mostly correlated with the specific collection sites or zones of sample collection (Figure 2.5).

The Ajwa cultivar demonstrates some discrepancy from the others in terms of isolation from commonly accepted historical knowledge. This could be related to the marker frequencies, and number of markers used was insufficient to clearly separate it. Therefore, Ajwa's relationship with the closest cultivar types is relatively distributed and would require more developed SSR marker sequences or more advanced markers and genome wide analysis methods (SNPs) to further associate it with a cluster of related cultivars. Ajwa is an ancient cultivar which falls into both subclusters II and IV, based on admixture (Figure 2.3), which indicates that Ajwa comprises common classification in Structure gene pool II and IV, while sharing inherently homologous features.

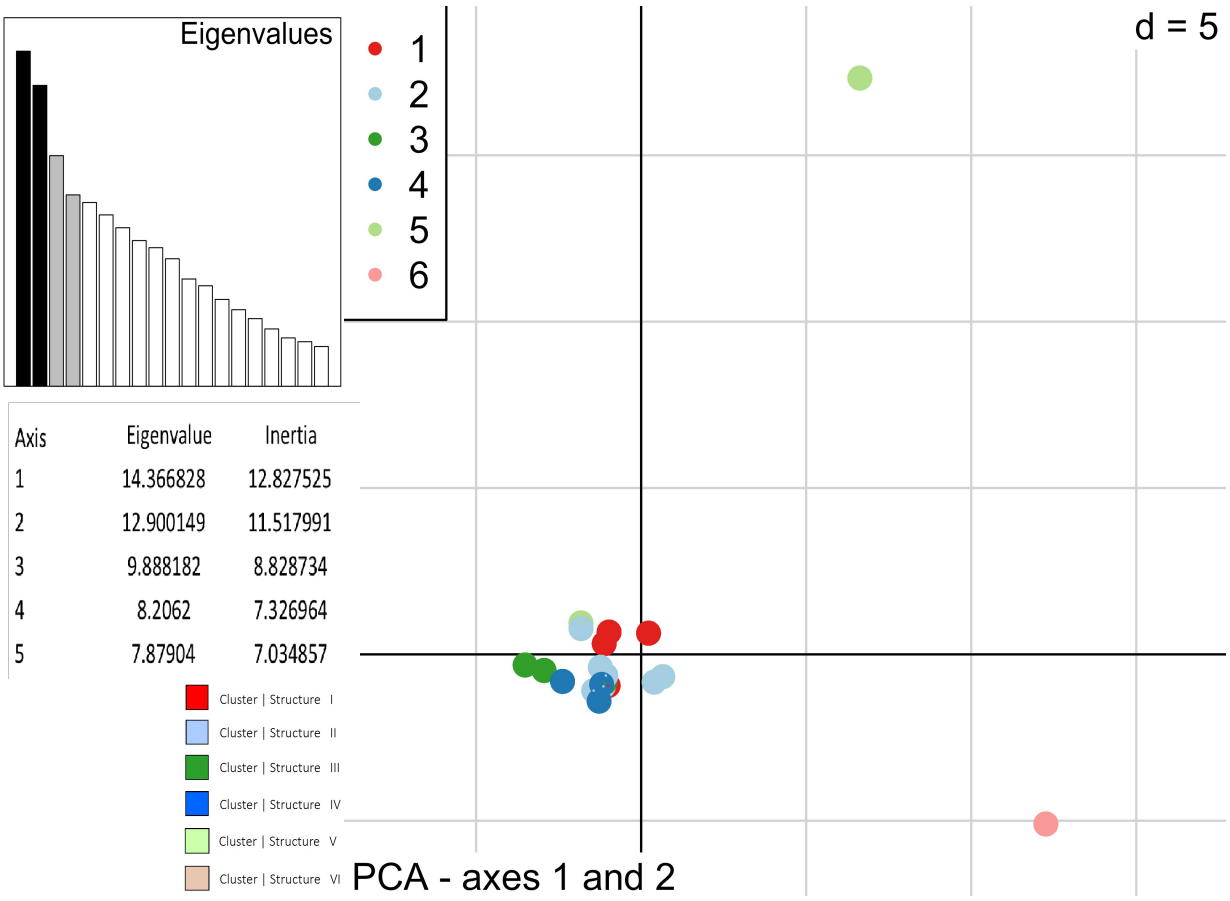
Ajwa has a genetic close relationship with the MP and Medjhool (MJH). The apparent background understanding on how cultivar Ajwa made it to the Middle Eastern region, was through a gifting of Deglet Noor (not part of study cultivars) cultivars during the 1980's. One of the offshoots collected and sent was a male (issue from a seed) now known as the Al Ain City Male Palm (MP). There is a close genetic relationship between MP and Ajwa, although MP and Medjhool are even more closely related (Figures 2.4 and 2.6).

Table 2.4. Summary of distance matrix produced using POPTREE2, generating the phylogenetic dendrogram Neighbor - Joining Unrooted tree chart (Figure 2.5).

Distance matrix

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.536	0.491	0.484	0.443	0.536	0.606	0.513	0.452	0.276	0.529	0.545	0.667	0.622	0.676	0.558	0.491	0.606	0.420	0.538
2		0.645	0.568	0.529	0.481	0.606	0.443	0.590	0.481	0.699	0.660	0.792	0.513	0.683	0.529	0.606	0.497	0.523	0.536
3			0.491	0.452	0.452	0.452	0.584	0.536	0.523	0.629	0.468	0.814	0.561	0.737	0.545	0.660	0.629	0.475	0.599
4				0.414	0.520	0.574	0.436	0.337	0.391	0.552	0.491	0.629	0.414	0.769	0.613	0.452	0.436	0.359	0.599
5					0.452	0.427	0.436	0.382	0.282	0.645	0.507	0.737	0.436	0.814	0.436	0.414	0.443	0.337	0.475
6						0.420	0.359	0.443	0.375	0.574	0.629	0.776	0.574	0.776	0.430	0.552	0.420	0.443	0.513
7							0.436	0.375	0.459	0.568	0.561	0.907	0.497	0.869	0.529	0.536	0.427	0.475	0.497
8								0.366	0.321	0.574	0.660	0.760	0.475	0.776	0.475	0.574	0.145	0.497	0.459
9									0.375	0.436	0.414	0.715	0.350	0.907	0.481	0.459	0.334	0.375	0.513
10										0.552	0.529	0.721	0.513	0.660	0.398	0.321	0.382	0.327	0.484
11											0.391	0.753	0.436	0.846	0.645	0.622	0.536	0.398	0.513
12												0.814	0.260	0.792	0.529	0.715	0.597	0.468	0.513
13													0.744	0.798	0.699	0.814	0.699	0.731	0.814
14														0.792	0.481	0.721	0.404	0.513	0.529
15															0.760	0.753	0.808	0.814	0.830
16																0.622	0.459	0.529	0.497
17																	0.590	0.420	0.513
18																		0.574	0.398
19																			0.552

((KAD:0.06964600,SKA:0.07486800)0.9120:0.11475900,SUL:0.24129400,((((((ABM:0.18601600,KHL:0.08981400)0.320
0:0.02719700,NSF:0.24066700)0.1150:0.02282100,DBS:0.16248700)0.0270:0.01707400,((CHI:0.19498900,KDR:0.141
83200)0.1110:0.01761700,((KNZ:0.22413400,(LUL:0.14515400,MKT:0.11474400)0.6720:0.05966100)0.4830:0.0619600
0,SKR:0.17540800)0.0490:0.00326300)0.1130:0.03024200)0.0100:0.02761400,((BRH:0.24730300,HEL:0.20490300)0.2
330:0.03062100,FDW:0.17955100)0.2210:0.02595700)0.0050:0.01038200,((AJW:0.25967900,(MJH:0.38663800,MP:0.4
1172200)0.4980:0.07850900)0.1820:0.02940700,MSR:0.23117500)0.0630:0.01360600)0.1550:0.02219600);



The inertia of each axes (axis) indicating percent variance in population separation along the Eigenvalue, PCA - 1 and 2 axis(es). These account for 12.83 and 11.52 variation respectively displaying proportion of variability among date palm cultivars and distinctive relationships in spatial representation. Calls closer to x axis: Force better reported. Color coding (1 – 6) respective to allelic frequency and Bayesian cluster analysis (Figure 2.1 and 2.3 Resp.).

Figure 2.4. Factorial coordinates - principal components analysis (PCA) separation of population by six gene pools constructed with 13 SSRs for 20 commercial cultivars representing 285 palms.

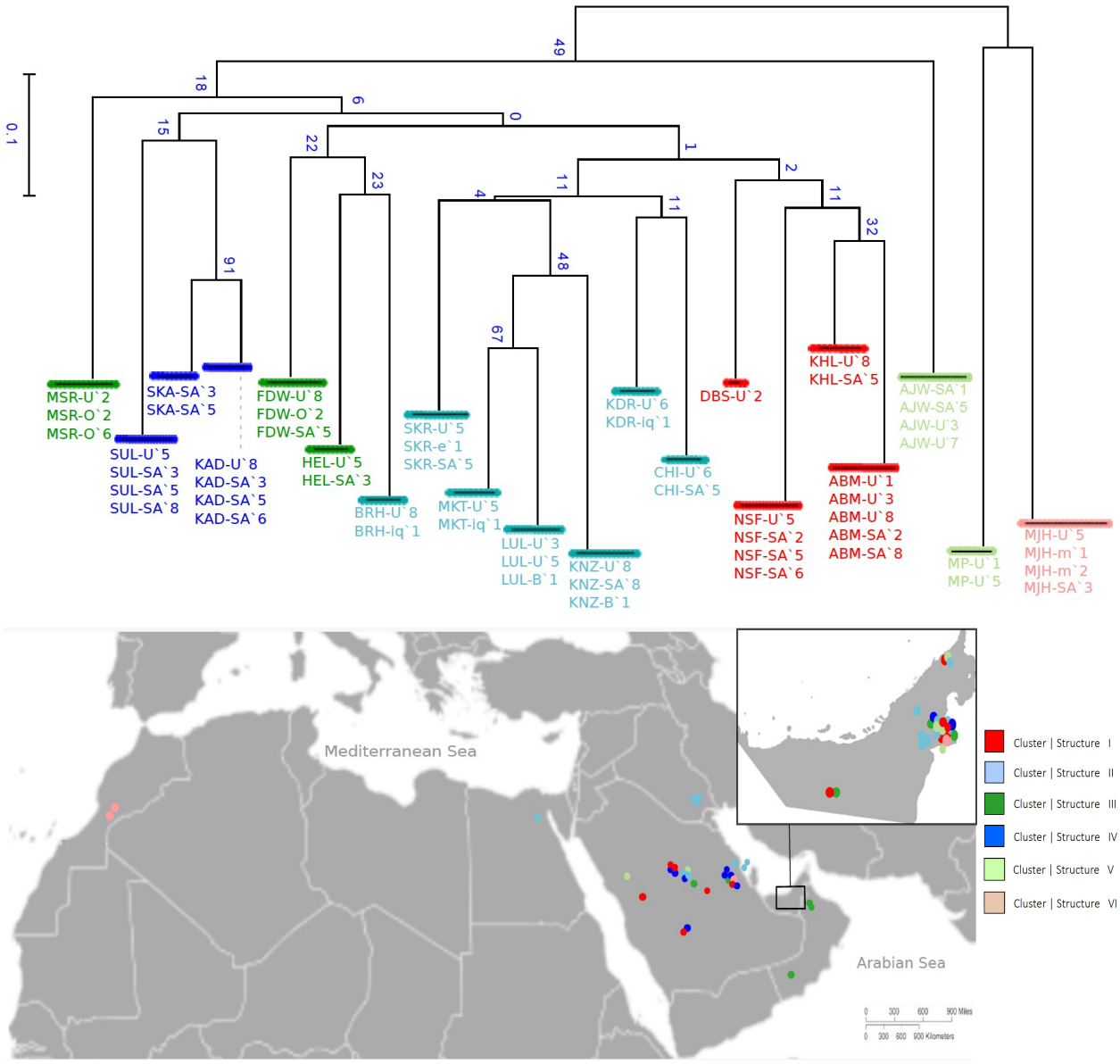
Ajwa's origin is non-representative of its true genetic background and geographical liaison, since as with cultivar Deglet Noor (a prized international cultivar dominating most date markets) a seedling was transferred, accidently, from Tunisia to the middle east before the establishment of appropriate cultivar characterization. The MP and Ajwa remain closely related (the MP and Ajwa

story was confirmed by Dr. *Hassan Shabana*, Dr. *Samir Al Shaker* (both | Iraq), Dr. *Helal Al Kaabi* (UAE), and Dr. *Abdellah Meddich* (Morocco)). The only cultivar population that separated equidistant between clusters was Barhee (BRH) which exhibited a level of uncertainty between sub-clusters 2 (light blue), 4 (green), and 1 (red) at 45 %, 35 %, and 20% respectively (Figure 2.3 and 2.4). Further molecular markers are needed to resolve this ambiguity of which gene pool cluster the cultivar group it should be more appropriately assigned.

2.3.6. Relatedness and genetic population differentiation of 285 palm

Population differentiation among domesticated cultivars was based on geographically narrow cultivars of date palm in earlier investigations (Mathew et al., 2015). These results relied on limited markers and lacked current computational capability. These tools were limited and lacked depth in gene pool grouping and thus for differentiating among and within cultivars populations. SSRs have proven to be useful in determining subclusters and auxiliary clades, allowing further separation of the two major sub-populations (North Africa and the Middle East). The results herein suggest a further distinction of genotypes with gene pools that diverged early enough to permit divergent fruit quality as illustrated in MP, Ajwa and Medjhool (Figures 2.4 and 2.5).

The separation of populations was apparent in the first and second axis of the Principle Component Analysis (PCA) (Figure 2.4) which was a factorial variance examination using eigenvalues and axis factors. The inertia of each axis was reached, and the distance matrices chosen used for MFA and neighbor-joining tree formation.



The numbers along the nodes indicate bootstrap values. Overlapping of geographical position signifies relative corresponding location / geographic locality. Map color coding signifies STRUCTURE assigned cluster | structure I - VI of gene pool composition associated with PCA Factorial Analysis of each ancestral subpopulation bundle and are placed on regional plot in the collected location, GPS-accurate (*Table 1*) to countries of source of each cultivar. Representing 285 date palm accessions.

Figure 2.5. A Phylogenetic dendrogram Neighbor-joining unrooted tree representing 6 subpopulation clusters, on distances (Da) by relationship and geographic distribution by collection locality of 20 comprehensively confirmed cultivar populations.

These indicated the highest variance in the data, observed in the direction of its eigenvector of 14.367 and 12.900 of axis 1 | 2 (| = by) and inertia or variation percentage of 12.828 and 11.518 (Figure 2.4), we produced 1 | (3 - 5), 2 | (3 - 5), 3 | (4 and 5), and 4 | 5 (App 8.1.2). A good correlation was observed with PCA and the genetic distance based NJ dendrogram (Figure 2.5) results. Populations Structure | clusters 5 and 6 (V and VI) from factors axis 1 and 2 are relatively distant to the remaining populations of clusters 1 through 4 since they are approximately orthogonal with respect to the main grouping (at right angles in factorial coordination) (Figure 2.4). This fits the model of population stratification whereby the genotypes of each individual were comprised of their understood localities of predicted origins except for the cultivar Ajwa, an outlier previously thought to be nested among the Middle Eastern cultivar populations. Yet it turns out that Ajwa is much more closely related to the MP which we know to have originated from the North African region. These results do not undermine the main accepted notion based on K = 2, that the date palm population is separated into two major sub populations as understood from their geographical support.

2.4. Conclusion

Electropherogram amplification was tested using the 14 - microsatellite loci across all samples of the 20 collected cultivars. They all showed clear SSR patterns and polymorphisms within the expected allelic ranges as noted by Billotte et al. (2004). Locus *mPdCIR044* produced erratic to no amplifications across many samples and hence it was dropped from further studies. This was possibly due to a polymorphism that is mutational in nature at several annealing sites. Some cultivar accessions fell slightly outside the microsatellites allelic ranges which may be

Table 2.5. Origin and fruit characteristics of 20 Cultivars used in the determination of relatedness.

Common Name	Origin	Fruit Quality	Fruit Ripening	Fruit Flavour	Fruit Colour	Fruit Shape	Fruit Texture	Fruit Size	Desirable Characters
Aboumaan	KSA | UAE	Good	Early | Middle	Good	Yellow	Oval | Long	Soft | Dry	Large	Black at tamar stage
Ajwa | Ajwaat Al Madina	KSA | Almadyna | NW Africa	Very good | Excellent | Old Var.	Middle	V. Good	Red | Black	Oblong | Ovoid	Chewy | Semi dry	Medium	Fleshy | Highly relevant to ancient times | Islam
Barhee	Iraq | Basrah	V. Good	Middle | Late	V. Good	Light brown | Yellow	Ovoid | Circular	Soft	Small | Medium	Flavor & taste | One of the Best
Chichi | Shishi	KSA | Al Hassa	Average | Good	Middle	Sweet	Yellow | Green	Oval | Long	Soft | Semi dry	Medium	Aka Abu Taouik
Dabbass	UAE | Liwa	Medium	Early | Middle	Good	Red | Brown | Yellow	Ovoid | Elongated	Tough	Small | Medium	Tasty flesh | Early tamar stage
Fard Abyad (White)	Oman | Aldhahra	Good	Late	Good	Dark | Brown | Red	Oblong | Elongated	Soft | Semi Dry	Small | Medium	Shiny | Smooth | Good for export
Helwa	Iraq | Basrah	Good | V. Good	Middle	Excellent	Dark | Golden | Red	Ovoid | Elongated	Soft | Semi Dry	Medium	Matures mid season & consumed at all stages of maturaty
Khadri	KSA | Al Qaseem	Good	Late	Good	Dark | Brown | Red	Cylindrical | Elongated	Hard | Semi dry	V. Large	Large stuffing | Highly demanded
Khadraoui	Iraq | Basrah	Good	Early | Middle	Rich	Yellow | Green	Oval	Soft | Dry	Medium	Flavor & Taste
Khlass	KSA | Al Hassa	V. Good	Middle	Good	Light brown	Oval | Oblong	Soft	Medium | Large	Light texture
Khenezi	KSA | Al Katif	Average | Good	Middle	Sweet | Close to Barhee	Red	Oval	Soft	Medium	Tolerates high humidity
Lulu	UAE | Ras Al Khaima	Good	Middle | Late	Good	Black | Yellow	Spherical | Inverted	Chewy | Soft	Medium | Small	Tasty texture | Consumed at rutab & tamar stages
Medjhool	MOR | Wadyzeez	Excellent	Early | Middle	V. Good	Brown | Red	Eliptical | Ovoid	V. Textured | Semi soft	V. Large	Fleshy | Large | One of the best
Maktoumi	Iraq | Baghdad	Very good | Excellent	Late	V. Good	Yellow	Cylindrical	soft	Medium | Large	Light | Large
Al Ain city male (Male)	UAE | Al Ain	Excellent	V. Late	V. Good	Flourense white	Male (no Fruit)	na	Long	Excllent Pollen | Best in Industry
Mdasry (Male)	No i.	Excellent	Late	V. Good	White flower	Male (no Fruit)	na	na	Good Pollinator
Nabtat Seif	KSA | Al Hassa & Qaseem	Excellent	Middle	Excellent	Yellow	Oval | Round	Soft	Medium	Highly prized var. but lower than Khlass
Sakaii	Oman | Al Dakghla	Excellent	Early | Middle	V. Good	White	Cylindrical	Soft	Medium	Early & high activity
Sukkari	Iraq | Basrah	Very good | Excellent	Middle	Sweet	Yellow	Heart shaped	Soft	Small	A small seed
Sultana	KSA	Excellent	Middle | Late	Slightly sweet	Yellow	Oval | Round	Soft	Medium	A very rare variety

na - not applicable, v. - very, i. - information.

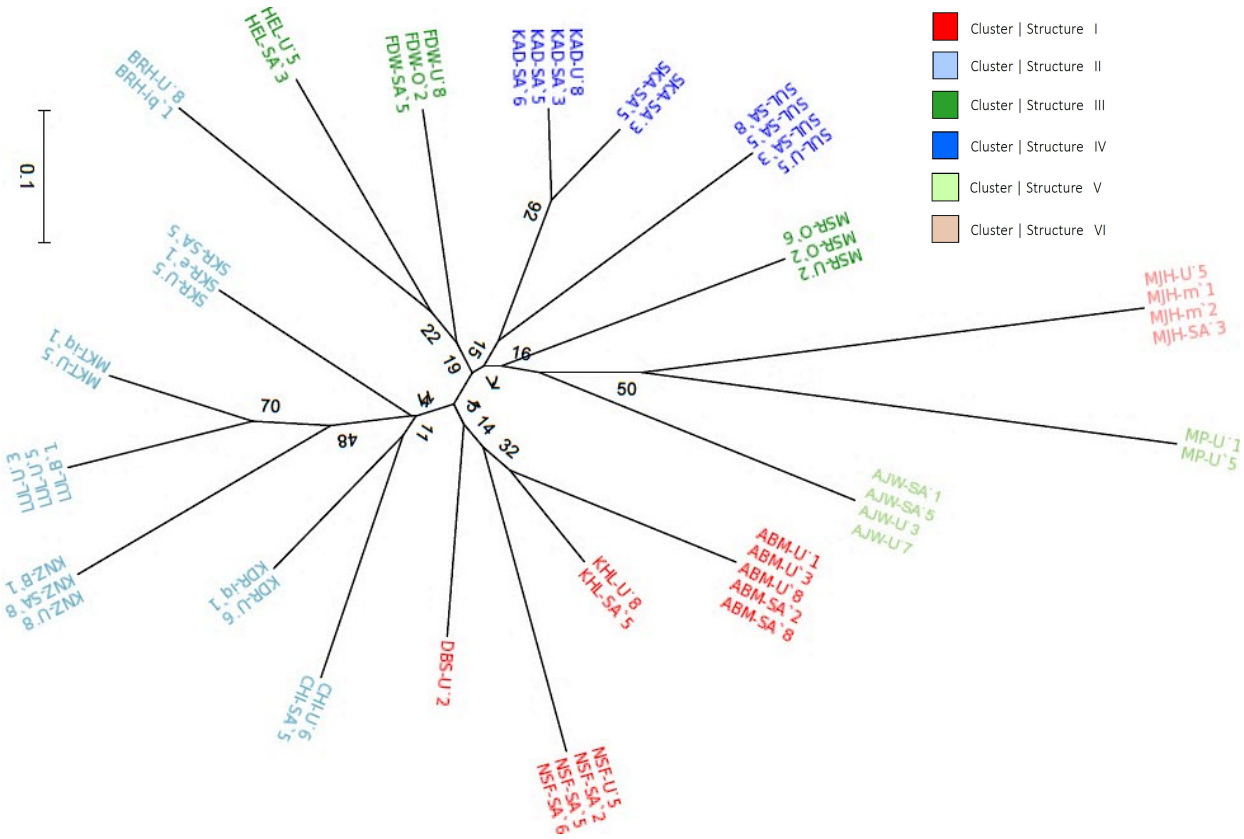


Figure 2.6. Alternative representation of distance link-relationships of 285 palms with respect to STRUCTURE results by Neighbor-Joining tree distance (Da) matrix, inclusive of relevant AFLP results of previous geographically narrow studies.

attributed to variation existing among cultivars, perhaps as a form of phenotypic plasticity since multigenerational successions may have adapted to different environments giving rise to certain mutations or silencing of genes due to G x E interaction. This was resolved by re-running in triplicate and going through all amplification electropherograms and manually adjusting those that were miscalled or outside the allelic range. The basis for this disparity may be due to environmental or farmer management stimulus leading to the observed fruit phenotypic variation among cultivar types. As the results indicated, the Middle Eastern sub-group showed successive branching by distance phylogeny where Structure I clusters I and II are derived and closely related from cluster

III and IV (Figure 2.5). We would expect and prefer to see a monosimilarity across the fruit phenotype among cultivars studied, with the understanding that miniscule variation is expected from cultural practices and is understood to be nongene associated but rather driven by environment interaction.

The trilateral confirmation of cultivars from all collection sites of the 27 localities, has fixed and associated into gene pool assemblages these 20 cultivar populations, that are regarded as globally commercially important. The importance of these cultivars not only of interest due to demand but also draw a better understood relationship, through genetic structure and phylogeny, between one another globally and why these are commercially important.

The fruiting phenotypes or the differences in fruit character from the studied cultivars has been reviewed by Sabir et al. (2014) on other cultivars. These differences are present in the occurrences of various subtle expressions conveyed in the molecular data and demonstrated by correlation with the SSR results. This is relevant when observing fruit color polymorphisms as the phenotypic resultant expression and commercial quality discrepancies related to the allelic call deviations we observed as rational for their commercial selection (Table 2.5) (Hamza et al., 2012 and Sabir et al., 2014). This work serves as a molecular confirmation and explanation to these rich polymorphisms that are not limited to the entirety of the fruit complex characteristics, as retrotransposon elements could be a factor of these expressive variations. As Hazzouri et al. (2015), stated the yellow fruiting color within the date palm and oil palm is caused by a truncated allelic dominance of negative mutations, further giving indication of variation of fruit physiognomies within genetically identical cultivars.

The Al Ain City Male Palm (MP), has an interestingly unique story behind its high pollen quality and elevated commercial interest. In these results it is represented to be closely related to the Medjhool cultivar as observed in the phylogenetic dendrogram trees (neighbor-joining) (Figures 2.5 and 2.6), while STRUCTURE results show a further clustering of the MP and Ajwa, both closely related to Medjhool. This is due to the inherent background association of these cultivars and the strict clonal propagation practices they experienced as preferred varieties. The phylogenetic closeness to Medjhool and its known origins as confirmed by these results is presumed to be from the Deglet Noor varietal relatives (gene pool) originally from the North West African - Tunisian region.

MP represents a date palm selected and bred male cultivar that has a history and is significantly different when compared to other male pollinator types. Also known as "Fahl / dokkars" that are different from females, with typically less stability since they are mainly cultivated by seeds and rarely multiplied through offshoots or other clonal means. Only a few males are known that have been propagated clonally. Commercially, a good adult date palm male produces on average of 20 inflorescences that contain 60 to 280 stands during a 30-day period with around 500 g of pollen (Marbeen et al., 2005). At the commercial level, the production of high quality fruit requires good pollination with typically one male to 35 females being sufficient. The Medjhool female cultivar and MP are very closely related and share a narrow genetic relationship, representing their leading role in commercial date palm cultivation.

Domesticated selection in cultivated dates overall has led to the selection of the many cultivars of today. The cultivars in this study were less diverse than anticipated due to their geographical concentration since they were domesticated a long time ago and through careful

clonal propagation by offshoots. They have persisted through use of tissue culture which has led to large monoculture populations (Meyer et al, 2012). The date palm species retains a high diversity of known and wild types due to their dioecious nature and manner of propagation. The highest diversity is found in the Middle Eastern subpopulation as we demonstrated in an admixture within cultivars from Structure | cluster populations I-IV (Figure 2.3*)* that belonged to populations 3, 4, 5, 6, 7, 9, 10, 15, and 20.

The use of isoenzyme analysis and SSRs for molecular marker investigations and genetic relationships as done by Ghislain et al. (2004), Garris et al. (2005), and D'Ambrosio et al., (2013), recognized the system of microsatellites. The study proved through genotype identification that *P. dactylifera* are genetically identical among each cultivar type. The western accessions of Medjhool, MP, and Ajwa belong to a major sub-population of *P. dactylifera*, however they separated into two clusters whereby MP and Ajwa are more closely linked to the Middle Eastern sub-group, demonstrated additionally by collection locality distances.

In conclusion, this analysis demonstrated that molecular markers of this category are ideal in determining relationships between and among cultivars of the date palm commercial complex. From this research, 13 specific SSR primers were applied across 20 cultivars in phylogenetic analysis and genetic structure expansion of the commercial cultivars of *P. dactylifera* and can be used in future gDNA identification and genomic organization analysis with a completely sequenced date palm genome. The genetic differentiation and structure by phylogenetic studies based on capillary sequenced data has shown an appropriate level of variation evaluating intraspecific relationships of commercial cultivars of date palm.

Chapter III

Through Simple Sequence Repeats (SSRs): A Practical Application for Next - Generation Commercial Micropropagation Production - Cultivar Identification - of Date Palm (*Phoenix dactylifera* L.)

3.1. Introduction

The date palm (*Phoenix dactylifera* L.) is a dioecious, perennial monocot that is well known for its edible, sugar-rich fruit. It has known a lengthy history of cultivation in countries surrounding the Persian Gulf, believed to be its region of origin, and stretching from northwest Africa to China (Zohary & Spiegel-Roy, 1975; Morton, 1988). An immense number of the global population depends on the nutritive date fruits and its by-products for use in generating biofuels and as structural materials to benefit their livelihood in those regions where date palms flourish. Notably a crucial species in arid regions and areas where naturally grown.

Elite cultivars of date palm are propagated clonally for commercial purposes. It is therefore critical that all planting material be true-to-type to the original genotype / mother plant. Historically, the date palm was only propagated by means of clonal offshoots (Chao & Krueger, 2007). However, offshoot production is limited and has been a limiting factor in commercial production. With the advent of tissue culture micropropagation of date palm (Tisserat, 1979b), the limited availability of suitable planting material was solved. Presently, several institutions, nurseries and companies produce large numbers of date palm plantlets via tissue culture micropropagation.

Numerous advances to characterize the date palm genome on a molecular level have been made within the last decade. Various molecular marker systems such as Amplified Fragment Length Polymorphism (AFLP) (Cao & Chao, 2002; El-Assar et al., 2005), Random Amplification Polymorphic DNA (RAPD) (Soliman et al., 2003; Sakr et al., 2012) and Simple Sequence Repeats (SSR's or microsatellites) (Billotte et al., 2004; Elmeer et al., 2011) have been employed to characterize specific genotypes. This offers a real practical system for clearly distinguishing specific date palm cultivars. Several studies have demonstrated the application of specific microsatellites in effectively identifying some date palm cultivars (Al-Ruqaishi et al., 2008; Elshibli and Korpelainen, 2008). Molecular characterization of the date palm has successfully culminated in the sequencing and publication of a draft sequence for the date palm genome by AlDous et al. (2011). Supplementary studies of the genome sequence will likely facilitate the development of a microarray chip system employing newly discovered Single Nucleotide Polymorphisms (SNP), which will further broaden the range of date palm cultivars that can be characterized at a molecular level. While such a genotyping system can provide reliable information for functional genomics laboratories with fairly large budgets, the associated cost of microarrays and their respective chip readers are still relatively high.

The relatively high cost of microarray development, and its associated applications effectively places the methodology beyond the reach of ordinary laboratories that are equipped for only basic molecular biology. Currently, microsatellites still represent an easy, robust and economic means of performing cultivar determination for these facilities. A microsatellite-based genotyping technique was directly developed and applied to a set of commercially important date palm cultivars chosen because of their relative importance. This research revealed underlying intraspecific variation which suggests that considerable intra-varietal diversity may exist within

male cultivars. This could be used for inferring genetic relatedness and for constructing phylogenetic trees.

This work demonstrates the practical application of characterized microsatellites in date palm cultivar determination for use in quality control. In addition, genotypes of important commercial cultivars, some with similar backgrounds, may be molecularly determined by a panel of 2 SSRs for routine sampling and a panel of 4 SSRs for trilateral verification and confirmation by 14 markers used as noted by Zaid S. E. (unpublished).

3.2. Materials and methods

3.2.1. Sampling

Leaf tissue samples for 38 (33 female; 5 male) different date palm genotypes were collected from the Al Ain Date Factory plantation, Kuweitat Presidential Villa and Al Rawda Palace in Al Ain, UAE, during the fruiting season. Samples were collected only from well-established trees (15+ years) whose cultivar identity had been determined and confirmed by date palm experts. Each plant sample source was comprised of 5 up to 10 accessions. All 38 confirmed date palm genotypes were also sampled from other regional sources within the Abu Dhabi Emirate and sourced from 13 locations in triplicate, that were also used in the genetic structure and phylogeny study (ch.2) (Table 2.1). The intent was to determine consensus genotypes while establishing whether intra-varietal variation or marginal divergence was present and to determine if additional microsatellite development were necessary. A further 18 samples from unconfirmed, allegedly new potential cultivars, were also collected from various farms and government collections (i.e. universities, government preservation center assemblies and others) in the MENA region. These still required

the 3 level (trilateral) confirmation. Although discernable, they have not been confirmed under the trilateral cultivar confirmation system as used for the 38 cultivars. The trilateral cultivar confirmation system can also be termed trilateral confirmation system when introducing a new variety for consideration of commercialization. These new cultivars require more SSR testing of loci compatibility to confirm them, along with sourcing plant material for testing from a broader collection area.

3.2.2. DNA extraction

Fresh plant tissue was homogenized in a TissueLyser II (QIAGEN, Hilden, Germany) blender in plant DNA extraction buffer from a BIOLINE ISOLATE Plant DNA Kit (BIOLINE, London, UK). Further extraction was done according to the protocol provided with the kit. The DNA was quantified using a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific, Waltham, Massachusetts, USA) while checking for protein contamination.

3.2.3. Primers and polymerase chain reaction

Oligonucleotide primers were custom ordered (Alpha DNA, Montreal, Quebec, Canada) according to the sequences used by Billotte et al. (2004); the forward primer for microsatellites mPdCIR015 and mPdCIR025 were 5'-labeled with 6-Carboxyfluorescein (6-FAM) while the forward primers for mPdCIR032 and mPdCIR085 were 5'-labeled with 6-Carboxyrhodamine (VIC) to prevent overlapping during data collection. The use of these florescent dyes enabled visualization of results and permitted a shortened analysis period, using primer master mixes that comprised of up to two primer sequences (*Table 3.1*). This table also illustrates sequences for the forward and reverse primers, with their respective 3rd generation 5' fluorescent dyes.

Amplification was done in 0.1 mL MicroAmp® Fast 96-well reaction plates (Applied Biosystems by Life Technologies, Carlsbad, California, USA). For each reaction, 50 ng of template DNA was used in a total volume of 10 μL, containing: 1× NH4 PCR Buffer; 200 μM dNTP's (BIOLINE, London, UK); 0.2 μM of forward (labeled) and reverse (unlabeled) primer for each microsatellite (mPdCIR015; mPdCIR025; mPdCIR032; mPdCIR085); 3 mM of $MgCl_2$; and 0.5 U of BIOTAQ™ DNA Polymerase (BIOLINE, London, UK). Thermal cycling was performed in an Applied Biosystems 2720 Thermal Cycler (Life Technologies, Carlsbad, California, USA) as follows: initial denaturation at 95°C for 5 minutes; 35 cycles of denaturation at 94°C for 30 seconds, primer annealing at 52°C for 90 seconds, elongation at 72°C for 60 seconds; followed by a final elongation step at 72°C for 30 minutes; the final soak temperature was set at 4°C until recovery. A low soaking temperature assures high quality DNA while preventing risk of degradation, since extraction of useful quality DNA material from date palms is challenging due to the tree's fibrous leaf tissue framework.

3.2.4. Trilateral cultivar confirmation system

A system based on three forms of verification was used to identify most of the genotypes in this study. Initially, the genotypes were verified according to verbal or annotated reference from the collection site caretaker along with any history of the individual or plantation from which the sourced material (explant) was collected. This reflected the material whether meristematic tissue or leaf tissue designated as a known, new cultivar or cultivar phenodistinct substitute. The 2nd level of this system was the molecular confirmation using allelic calls and multiplex matching through SSR technology. The allelic bp calls are assigned by a task performed by comparing allele length estimates to a database of fragments of known sizes. This can be allele sizes observed in

Table 3.1. Primer sequences (forward and reverse) used in the characterization of the date palm cultivars.

SSR locus	Primer sequence (5' → 3')	5' Modification
mPdCIR015	F: AGC TGG CTC CTC CCT TCT TA	6 – FAM
	R: GCT CGG TTG GAC TTG TTC T	
mPdCIR025	F: GCA CGA GAA GGC TTA TAG T	6 – FAM
	R: CCC CTC ATT AGG ATT CTA C	
mPdCIR032	F: CAA ATC TTT GCC GTG AG	VIC
	R: GGT GTG GAG TAA TCA TGT AGT AG	
mPdCIR085	F: GAG AGA GGG TGG TGT TAT T	VIC
	R: TTC ATC CAG AAC CAC AGT A	

Loci are represented as primer sequences with appropriate 5' fluorescent dyes from Alpha DNA (Montreal, Canada); *m* corresponds to microsatellite, *Pd* to *P. dactylifera* and *CIR* to CIRAD; 3rd generation dyes / fluorescent dyes; 6FAM Blue, and VIC Green). F: forward primer sequence, R: reverse primer sequence. Primer sequences were according to Billotte *et al*. (2004).

newly derived data sets. Alternatively, by visually defining bin boundaries or allelic ranges based on observed results across several samples (Idury & Cardon, 1997), it is possible to manually define by vision bin boundaries based on allele sizes observed in a newly acquired dataset (Amos et al., 2007).

The final confirmation step in this method of 'dates Palmàprinting', verifies the cultivar by the fruit physiognomies where the drupelet phenotype, as an early (4-5 years) adult or towards the

end of the juvenility period. This trilateral approach achieves the confidence required to claim and classify as a cultivar or cultivar phenodistinct substitute.

3.2.5. PCR Product Analysis

The PCR products were then purified using Nucleofast® 96 PCR Plates (Macherey Nagel, Düren, Germany) with post-PCR cleanup on a Tecan EVO robotic platform (Tecan, Männedorf, Switzerland). They were subjected to capillary electrophoresis on an Applied Biosystems ABI 3730xl DNA Analyzer (Life Technologies, Carlsbad, California, USA) according to the manufacturer's instructions using 50 cm capillary-arrays containing POP-7™ polymer (Applied Biosystems by Life Technologies, Carlsbad, California, USA).

The raw data was collected using Data Collection v.3.0 software, and fragment analysis was performed using GeneMapper® v.4.1 software (Applied Biosystems by Life Technologies, Carlsbad, California, USA). Fragments were normalized to correct allelic amplifications. The alleles were sized into their appropriate allele "bins" using least-squares minimization protocols. High-throughput fluorescent genotyping requires a considerable amount of automation for accurate and efficient processing of genetic markers, which is a type-limiting factor.

3.3. Results and Discussion

Amplification products revealed that the amount of genotypic polymorphism detected was sufficient to distinguish 54 of the 56 samples (cultivars). The tailored system developed here for identification demonstrated its accuracy in distinguishing individual propagules for commercial applications. The genotyping data produced by the microsatellite panel for the confirmed

genotypes is presented in *Table 3.2* and the unconfirmed cultivars in *Table 3.3* revealed that each were genetically distinct and unique. This illustrates the potential of SSRs for the routine cultivar genotyping of in-vitro cultures, greenhouse and nursery date plants.

During fragment analysis, visual confirmation of the allelic sizes and ranges on the electropherogram outputs proved essential. This was because GeneMapper® software occasionally missed certain allele calls since the microsatellites sequence amplifications were sometimes outside the sequence range or slightly shifted off-microsatellite range. An example of such electropherograms is shown in Figure 3.1. The process described herein was developed to provide commercial and private tissue culture laboratories a means of *Dates Palmàprinting* their lines and products from introduction into the facility to subsequent point of sale of date palm *P. dactylifera*. It also provides the date palm industry with a standardized means to resolve one critical issue, which is cultivar mixing and true-to-type.

The miscalled *mPdCIR025* allele (199 bp) for cultivar "Medjhool" occurred due to the fragment amplifying marginally outside the defined allelic range. This is an illustration that the software programs such as GeneMapper used here still have difficulties in standardizing data output. This could also be due to the phenomenon of 'cultivar phenodistinct substitutes' which represent slight variation, yet they are genetically identical. It is thus necessary to normalize the data for the purpose of distinction, with visual confirmation until more accurate and reliable means are developed. Human based normalizing and verification are necessary for repeatable results at this stage, by assigning allele sizes into their appropriate allele "bins" using least-squares minimization protocols and confirming through references.

Table 3.2. Genotype data obtained for the 38 confirmed date palm cultivars using 4 multiplexed primers.

Cultivar / Var. name	SSR locus – allele sizes (bp)			
	mPdCIR015	*mPdCIR025*	*mPdCIR032*	*mPdCIR085*
Aboumaan	122 124	211 213	287 294	155 163
Ajwa	122 128	226 226	292 296	163 172
Ashal Hassa	128 128	211 213	287 298	155 174
Baqlat Bent Manii	122 128	211 230	287 298	153 172
Barhee	120 132	213 230	287 300	155 172
Chichi	132 136	211 230	287 294	153 172
Dibbas	122 122	211 213	287 298	172 176
Fard White	120 128	200 211	294 296	172 172
Ghannami*†	124 124	211 230	300 304	172 176
	122 136	211 226	292 294	163 163
Ghareef*†	122 122	211 213	292 300	149 165
	124 124	211 213	287 294	155 163

Hilali	122	132	211	230	292	294	172	172
Jabri	132	136	211	213	292	294	172	172
Jech Fatima	120	128	213	230	294	300	155	161
Jech Ramli	122	134	200	211	292	292	153	172
Kadri‡	128	136	211	211	292	292	163	172
Khadraoui	124	124	211	213	292	294	172	172
Khenezi	124	132	211	211	292	294	155	174
Khenezi White	120	124	211	211	292	303	155	174
Khisab	120	132	211	213	294	298	155	174
Khlass	122	128	211	213	287	294	163	174
Lulu	120	124	213	226	287	298	155	174
Madayan	120	136	213	213	287	298	163	163
Maktoumi	124	136	212	226	287	292	172	174
Mdasry*	120	122	213	213	292	292	165	176
Medjhool	136	136	199	213	294	295	153	176
Hilali Red	128	132	200	230	294	294	172	174

MP (Al Ain Male)*	128	142	226	230	287	288	167	174
Mumtaz	128	132	211	213	294	294	155	174
Nabtat Breem	122	132	213	230	294	298	153	172
Nabtat Mazroui	128	132	211	212	287	300	155	173
Nabtat Seif	122	122	211	230	294	298	172	174
Nadira	128	136	211	230	294	298	163	174
Nawader	122	124	211	211	287	294	153	174
Sakaii‡	128	136	211	211	292	292	163	172
Sekka*†	122	128	211	213	292	300	149	149
	122	124	211	211	292	292	151	176
Sukkarri	122	132	211	213	287	294	165	172
Sultana	120	122	211	213	292	298	155	172
Zamli	122	128	211	230	292	294	155	161

* Male date palm cultivars. Cultivar / Var. denotes that farmers tend to readily use the term variety, facilitating relation.
† Date palm cultivars which showed variation among samples; a consensus genotype could not be established.
‡ Microsatellite profiles using the above primers could not distinguish between Kadri and Sakaii cultivars. Each column illustrates the SSR locus allele size calls amid the derived allelic range MP / Al Ain City Male Palm - still has not been classified under specific cultivar.

Only two cultivars, Sakaii and Kadri, were not distinguishable from one another using the above microsatellite panel. Further investigation with additional primers revealed that they differ

at SSR loci *mPdCIR010*, *mPdCIR035* and *mPdCIR093* (data not shown). Additional testing of new microsatellite primers will be required to further broaden the coverage in date palm cultivar *Palmàprinting*. Consequently, 6-FAM labeled primers for *mPdCIR010* and *mPdCIR035* are sufficient for the routine screening / testing of cultures and plantlets. Further tailoring will be needed to facilitate identification of additional commercially important candidates.

For three of the male cultivars, Ghannami, Ghareef and Sekka, a consensus genotype could not be established. Control samples sourced over the several locations of collection resulted in two distinctly different genotypes for each cultivar. These were experimental repeats of genotyping on various male date palm offshoots (including Ghannami, Ghareef, Mdasry and Sekka) designated for potential explant initiation material.

When additional microsatellite sequences were used, many different and unique genotypes were revealed within a collection of offshoots of cultivar accessions. None of the accessions appeared to match. This suggests that much intra-varietal diversity exists within male cultivars of date palm also determined as intra-cultivar type phenodistinct substitutes. This is likely due to historically less stringent selection on male cultivars which have fewer identifiable morphological characteristics. The female cultivars are traditionally more strictly clonally propagated as well as more easily identifiable by their fruit characteristics and other morphological traits.

Table 3.3. Genotype data obtained using 4 primers for 18 unconfirmed cultivars, showing microsatellite loci allele call amplifications.

Variety/cultivar name	SSR locus – allele sizes (bp)			
	mPdCIR015	*mPdCIR025*	*mPdCIR032*	*mPdCIR085*
Abu Badia	122 132	211 213	292 294	172 174
Abu Zabd	120 122	213 230	292 294	165 172
Ashal Khass	124 128	211 213	287 287	163 163
Ayassha	128 136	211 230	292 298	174 174
Diyala Sukkarri	124 124	211 212	287 298	172 172
Ganda	128 132	211 213	287 294	155 172
Hamdiya	122 124	211 213	294 298	165 174
Hilali Senee	132 132	211 213	294 294	161 172
Kuweitat	124 132	212 230	292 294	155 161
Maymona	132 132	211 230	287 303	172 172
Nabtat Dakhil	122 136	211 211	292 294	155 163
Nabtat Moneef	128 132	211 213	294 303	155 163
Nagal Hilali	120 132	211 213	294 294	172 172

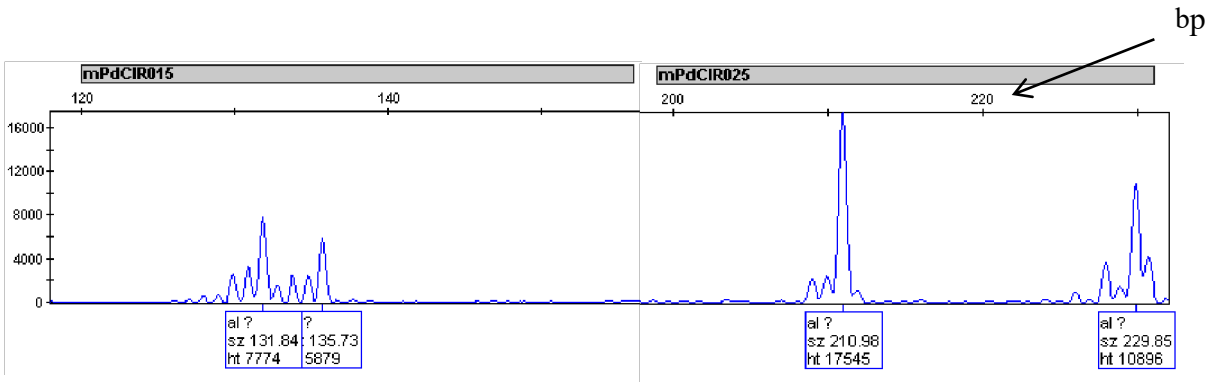
Rotana Masfoot	122	136	211	230	287	292	161	172
Salmia	128	132	211	213	296	298	165	172
Sheikha	120	132	211	213	287	294	163	176
Thinal	122	132	211	211	287	294	163	163
Umm Thaq	132	136	211	211	287	292	172	172

Each column illustrates the SSR locus allele size calls amid the derived allelic range. A consensus trilateral confirmation of genotype has not been completed on these cultivars

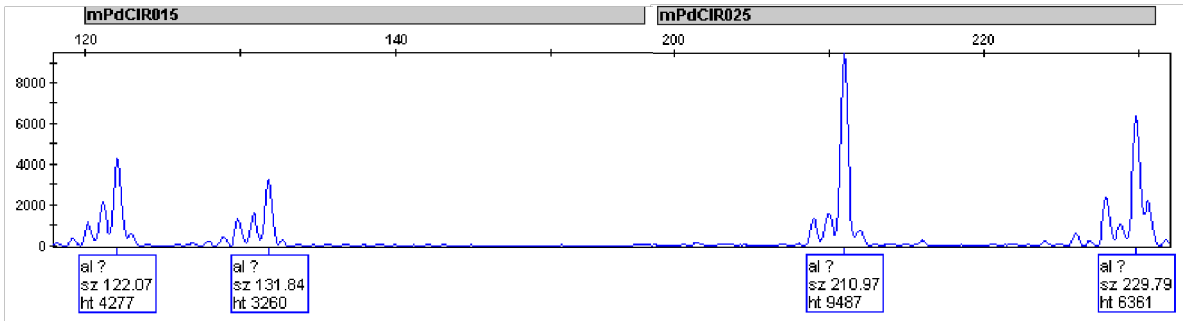
Furthermore, males could likely have been propagated by seed which may be a factor leading to this variability since one is not directly concerned with the resulting fruit. The diversity may likewise be the result of cultivars originating in or occupying the same geographic area caused by assortive mating, one gene, with linkage disequilibrium between genes or of traits subjected to disruptive selection leading to a form of divergence (Ortiz-Barrientos & Rieseberg, 2006). This may be better understood through the use of additional SSRs. The unconfirmed 18 cultivars fall under Ortiz-Barrientos & Rieseberg (2006) theory where some of the cultivars are not different but require different marker sets, allelic range sequences. It also supports the Trilateral Consensus Confirmation for accurate determination of genotypes.

The genotyping of the remaining 18 cultivars since they are new and potentially commercially viable requires additional microsatellites or SNPs to be developed and tested across most plants (offshoots) potentially found and collected from multiple sites for verification according to the trilateral consensus confirmation system. In regard to morphological differences,

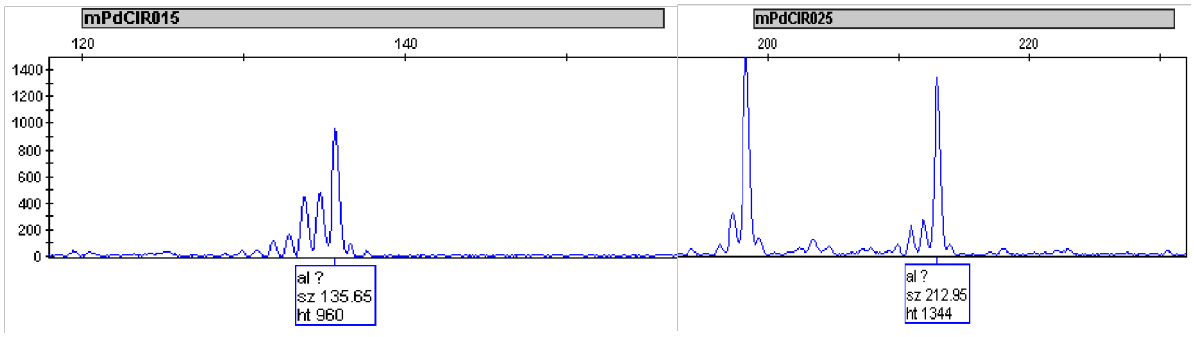
Chichi :



Hilali :



Medjhool :



bp - base pair.

Figure 3.1. Allele differences in electropherogram (SSR loci *mPdCIR015* and *mPdCIR025*) examples of date palm cultivars (Chichi; Hilali; Medjhool) using illustration produced via GeneMapper® software.

they are quite apparent in the fruit, although not as understood in the remaining organs such as leaf, stems, and main stalk.

3.4. Conclusion

Employment of the microsatellite panel adopted in the present study presented as noted here would enable and facilitate quality control monitoring of micropropagated date palms. These Short Tandem Repeat microsatellites provide a viable alternative for cultivar identification in date palm at a time when focus is shifting towards functional genomics and absolute next generation sequencing, microchip genotyping, SNPs, and whole genome characterization. The multiplexing approach easily facilitated the genotyping of a large number of accessions. This method of microsatellite application remains effective, robust, cost efficient and highly suitable to automation for a high sample throughput, enabling and facilitating quality control monitoring of micropropagated date palms. Thus, necessary for micropropagation of true-to-type plants and germplasm conservation. A panel of only four microsatellites was able to distinguish among 54 date palm cultivars of which 38 are of commercial production, thus demonstrating its potency to the industry.

Bibliography

Adawy SS (2007). An evaluation of the utility of simple sequence repeat loci (SSR), expressed sequence tags (ESTs) and expressed sequence tag microsatellites (EST-SSR) as molecular markers in cotton. *Journal of Applied Sciences Research* 3(11): 1581-1588.

Adawy SS, Hussein EHA, Ismail SEME, El-Itriby HA (2006). Genomic diversity in date palm (*Phoenix dactylifera* L.) as revealed by AFLPs in comparison to RAPDs and ISSRs. *Abstracts of 3rd international date palm conference, Abu Dhabi, United Arab Emirates* : 19-21 February.

Ahmed T, Al-Qaradaw A (2009). Molecular phylogeny of Qatari date palm genotypes using simple sequence repeats markers. *Biotechnology (Faisalabad)* 8(1): 126-131.

Akkak A, Scariot V, Marinoni DT, Boccacci PAOLO, Beltramo CHIARA, Botta R (2009). Development and evaluation of microsatellite markers in Phoenix dactylifera L. and their transferability to other Phoenix species. *Biologia Plantarum* 53(1): 164-166.

Al-Dous EK, George B, Al-Mahmoud ME, Al-Jaber MY, Wang H, Salameh YM, Al-Azwani EK, Chaluvadi S, Pontaroli AC, DeBarry J, Arondel V, Ohlrogge J, Saie IJ, Suliman-Elmeer KM, Bennetzen JL, Kreugger RR, Malek JA (2001). De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nature Biotechnology* 29(6): 521-527.

Al-Khalifah NS, Askari E (2003). Molecular phylogeny of date palm (*Phoenix dactylifera* L.) cultivars from Saudi Arabia by DNA fingerprinting. *Theor. Appl. Genet* 107: 1266-1270.

Al-Kuran O, Al-Mehaisen L, Bawadi H, Beitawi S, Amarin Z (2011). The effect of late pregnancy consumption of date fruit on labor and delivery. *Journal of Obstetrics & Gynecology* 31(1): 29-31.

Al-Mssallem IS, Hu S, Zhang X, Lin Q, Liu W, Tan J, Yin Y (2013). Genome sequence of the date palm *Phoenix dactylifera* L. *Nature communications 4*: 3274.

Al-Ruqaishi IA, Davey M, Alderson P, Mayes S (2008). Genetic relationships and genotype tracing in date palm (*Phoenix dactylifera* L.) in Oman, based on microsatellite markers. *Plant Genetic Resources: Characterization and Utilization* 6(1): 70-72.

Al-Shahib W, Marshall RJ (2003). The fruit of the date palm: Its possible use as the best food for the future? *International journal of food sciences and nutrition 54*(4): 247-259.

Al-Yahyai R, Al-Khanjari S (2008). Biodiversity of date palm in the Sultanate of Oman. *African Journal of Agricultural Research 3*(6): 389-395.

Alamerew S, Chebotar S, Huang X, Röder M, Börner A (2004). Genetic diversity in Ethiopian hexaploid and tetraploid wheat germplasm assessed by microsatellite markers. *Genetic Resources and Crop Evolution 51*(5): 559-567.

Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF (2006). A modified protocol for rapid DNA isolation from plant tissues using *cetyltrimethylammonium bromide*. *Nature protocols* 1(5): 2320-2325.

Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS (2007). Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Resources 7*(1): 10-14.

Arabnezhad H, Bahar M, Mohammadi HR, Latifian M (2012). Development, characterization and use of microsatellite markers for germplasm analysis in date palm (*Phoenix dactylifera* L.). *Scientia Horticulturae 134*: 150-156.

Askari EJAZ, Al-Khalifah NS, Ohmura T, Al-Hafidh YS, Khan FA, Al-Hindi A, Okawara RYOJI (2003). Molecular phylogeny of seven date palm (*Phoenix dactylifera* L.) cultivars by DNA fingerprinting. *Pak. J. Bot 35*(3): 323-330.

Bailey Hortatorium LH (1976). Hortus Third. Macmillan, New York. (CF Krueger RR 1998).

Proceedings the First International Conference on Date Palms. Al Ain, United Arab Emirates 8-10 March.

Bajaj YPS (1995). Cryopreservation of plant cell, tissue, and organ culture for the conservation of germplasm and biodiversity. *Cryopreservation of plant germplasm I,* Springer Berlin Heidelberg: 3-28.

Balk EM, Raman G, Tatsioni A, Chung M, Lau J, Rosenberg I H (2007). Vitamin B6, B12, and folic acid supplementation and cognitive function: a systematic review of randomized trials. *Archives of Internal Medicine 167*(1): 21-30.

Baker WJ, Couvreur T L (2013). Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. I. Historical biogeography. *Journal of Biogeography 40*(2): 274-285.

Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, Medina J (2016). Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytologist 209*(2): 855-870.

Barrow SC (1998). A monograph of *Phoenix* L. (*palmae: Coryphoideae*). *Kew bulletin*: 513-575.

Beech M (2003). Archaeobotanical evidence for early date consumption in the Arabian Gulf, in: The date palm from traditional resource to green wealth. *The Emirates Center for Strategic Studies and Research,* Abu Dhabi 11: 31-39.

Bekheet SA (2011). In vitro conservation of date palm germplasm. In *Date Palm Biotechnology*, Springer Netherlands: 337-360.

Bell CJ, Ecker JR (1994). Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics 19*(1): 137-144.

Ben A (1990). Date palm growing. *Options Mediterraneennes Serie A Seminaires Mediterraneens* 11: 105-120.

Billotte N, Risterucci AM, Barcelos E, Noyer JL, Amblard P, Baurens FC (2001). Development, characterisation, and across-taxa utility of oil palm (*Elaeis guineensis Jacq*.) microsatellite markers. *Genome 44*(3): 413-425.

Billotte N, Marseillac N, Brottier P, Noyer JL, Jacquemoud-Collet JP, Moreau C, Couvreur T, Chevallier MH, Pintaud JC, Risterucci AM (2004). Nuclear microsatellite markers for the date palm (*Phoenix dactylifera* L.*):* characterization and utility across the genus *Phoenix* and in other palm genera. *Molecular Ecology Resources / Notes* 4(2): 256-258.

Blyton MDJ, Flanagan NS (2012). A comprehensive guide to: GenAIEx genetic analysis in Excel. *Canberra: Australian National University*.

Bohra A, Dubey A, Saxena RK, Penmetsa RV, Poornima KN, Kumar N, Ramesh S (2011). Analysis of BAC-end sequences (BESs) and development of BES-SSR markers for genetic mapping and hybrid purity assessment in pigeonpea (*Cajanus* spp.). *BMC plant biology 11*(1): 56.

Cao BR, Chao CT (2002). Identification of date cultivars in California using AFLP markers. *HortScience* 37(6): 966-968.

Carpenter JB (1981). Improvement of traditional date culture. *Date Palm J*. 1:1–16.

Cavalli-Sforza LL, Edwards AWF. (1967). Phylogenetic analysis: models and estimation procedures. *Am J Hum Genet* 19: 233-257.

Chakraborty R (1977). Estimation of time of divergence from phylogenetic studies. *Can J Genet Cytol* 19: 217-223.

Chambers GK, MacAvoy ES (2000). Microsatellites: consensus and controversy. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol* 126: 455-476.

Chao CT, Krueger RR (2007). The date palm (*Phoenix dactylifera* L.*)*: overview of biology, uses, and cultivation. *HortScience 42*(5): 1077-1082.

Chazan M (2009). World Prehistory and Archaeology: Pathways through Time. *Canadian Edition. Pearson*.

Ciofi C, Milinkovitch MC, Gibbs JP, Caccone A, Powell JR (2002). Microsatellite analysis of genetic divergence among populations of giant Galápagos tortoises. *Molecular Ecology* 11(11): 2265-2283.

Cook OF (1901). Origin and distribution of the cocoa palm (Vol. 7, No. 2). *US Government Printing Office*.

Corniquel B, Mercier L (1994). Date palm (*Phoenix dactylifera* L.*)* cultivar identification by RFLP and RAPD. *Plant Science 101*(2): 163-172.

Costantini L (1985). Considerazioni su alcuni reperti di palma da dattero e sul centro di origine e l'area di coltivazione della *Phoenix dactylifera* L. Rome In: Gnoli, G., Lanciotti, S. (Eds.). *Orientalia Josephi Tucci Memoriae Dicata* : 209-218.

Crow JF (2004). Assessing population subdivision. In: Wasser SP, editor. Evolutionary Theory and Processes. Modern Horizons, Netherlands. *Kluwer Academic Publishers*: 35-42.

D'Ambrosio C, Arena S, Rocco M, Verrillo F, Novi G, Viscosi V, Scaloni A (2013). Proteomic analysis of apricot fruit during ripening. *Journal of proteomics 78*: 39-57.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics 12*(7): 499.

Deka R, Shriver MD, Yu LM, Ferrell RE, Chakraborty R (1995). Intra- and inter-population diversity at short tandem repeat loci in diverse populations of the world. *Electrophoresis* 16: 1659-1664.

Dickie JB, Balick MJ, Linington IM (1992). Experimental investigations into the feasibility of ex situ preservation of palm seeds; an alternative strategy for biological conservation of this economically important plant family. *Biodiversity and Conservation 1*(2): 112-119.

Dowson VHW (1982). Date production and protection with special reference to North Africa and the Near East. *FAO Technical Bulletin No. 35, Rome*: 294.

Doyle JJ, Doyle JL (1990). Isolation of DNA from small amounts of plant tissues. Also known as; Total DNA preparation procedure for fresh plant tissue. *BRL focus 12*: 13-15.

Dransfield J, Uhl NW, Asmussen CB et al. (2008). Genera palmarum: the evolution and classification of palms. *Kew Publishing, Kew*

Dransfield J, Uhl NW (1986). An outline of a classification of palms. *Principes* 30(1): 3-11.

Dransfield J, Uhl NW (1998). Palmae. In Flowering Plants Monocotyledons. *Springer Berlin Heidelberg*: 306-389.

Duminil, J. (2006). Etudes comparatives de la structure génétique des plantes. Doctoral dissertation, *Nancy*.

El-Assar AM, Kreuger RR, Devanand PS, Chao CCT (2005). Genetic analysis of Egyptian date (*Phoenix dactylifera* L.) accessions using AFLP markers. *Genetic Resources and Crop Evolution* 52(5): 601-607.

El Hadrami A, Daayf F, El Hadrami I (2011). Date palm genetics and breeding. In *Date Palm Biotechnology, Springer Netherlands*: 479-512.

Elhoumaizi MA, Lecoustre R, Oihabi A (2002). Phyllotaxis and handedness in date palm (*Phoenix dactylifera* L.). *Fruits 57*(5-6): 297-303.

Elhoumaizi MA, Devanand PS, Fang J, Chao CCT (2006). Confirmation of Medjool date as a landrace variety through genetic analysis of Medjhool accessions in Morocco. *Journal of the American Society for Horticultural Science 131*(3): 403-407.

Ellegren H (2004). Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* 5: 435–445.

Elmeer K, Sarwath H, Malek J, Baum M, Hamwieh A (2011). New microsatellite markers for assessment of genetic diversity in date palm (*Phoenix dactylifera* L.). *3 Biotech* 1(2): 91-97.

Elshibli S, Korpelainen H (2008). Microsatellite markers reveal high genetic diversity in date palm (*Phoenix dactylifera* L.) germplasm from Sudan. *Genetica* 134: 251-260.

Estoup A, Jarne P, Cornuet J (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol Ecol* 11: 1591-1604.

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.

Fahima T, Röder MS, Grama A, Nevo E (1998). Microsatellite DNA polymorphism divergence in Triticum dicoccoides accessions highly resistant to yellow rust. *TAG Theoretical and Applied Genetics 96*(2): 187-195.

Felsenstein J (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783-791.

Ferry M (1998). The phyllotaxis of the date palm (*Phoenix dactylifera* L.). In *Proc. Inter. Conf. on Date Palms, Al-Ain, UAE*: 559-571.

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005). Genetic structure and diversity in Oryza sativa L. *Genetics 169*(3): 1631-1638.

Gautheret RJ (1934). Culture du tissus cambial. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences 198*: 2195–2196.

Gautheret RJ (1935). Recherches sur la culture des tissus végétaux. *Ph.D. Thesis, Paris*.

Gautheret RJ (1939). Sur la possibilité de réaliser la culture indéfinie des tissus de tubercules de carotte. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences 208*: 118–120.

Gebauer J, Luedeling E, Hammer K, Nagieb M, Buerkert A (2007). Mountain oases in northern Oman: an environment for evolution and in situ conservation of plant genetic resources. *Genetic resources and crop evolution 54*(3): 465-481.

Ghislain M, Spooner DM, Rodriguez F, Villamón F, Nunez J, Vásquez C, Bonierbale M (2004). Selection of highly informative and user-friendly microsatellites (SSRs) for genotyping of cultivated potato. *Theoretical and applied Genetics 108*(5): 881-890.

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995). Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc Natl Acad Sci USA* 92: 67236727.

Gonnet GH, Hallett MT, Korostensky C, Bernardin L. (2000). Darwin v. 2.0: an interpreted computer language for the biosciences. *Bioinformatics* 16: 101-103.

González-Pérez MA, Caujapé-Castells J, Sosa PA (2004). Allozyme variation and structure of the Canarian endemic palm tree *Phoenix canariensis* (*Arecaceae*): implications for conservation. *Heredity 93*(3): 307-315.

Goulet C, Mageroy MH, Lam NB, Floystad A, Tieman DM, Klee HJ (2012). Role of an esterase in flavor volatile variation within the tomato clade. *Proceedings of the National Academy of Sciences 109*(46): 19009-19014.

Guatelli-Steinberg D (2016). What Teeth Reveal about Human Evolution. *Cambridge University Press*.

Haider N, Nabulsi I, MirAli N (2012). Phylogenetic relationships among date palm (*Phoenix dactylifera* L.) cultivars in Syria using RAPD and ISSR markers. *J. Plant Biol. Res 1*(2): 12-24.

Haberlandt G (1969). Experiments on the culture of isolated plant cells. *The Botanical Review 35*(1): 68-88.

Hamza H, Benabderrahim MA, Elbekkay M, Ferdaous G, Triki T, Ferchichi A (2012). Investigation of genetic variation in Tunisian date palm (*Phoenix dactylifera* L.) cultivars using ISSR marker systems and their relationship with fruit characteristics. *Turkish Journal of Biology 36*(4): 449-458.

Hazzouri KM, Flowers JM, Khierallah HS, Rosas U, Pham GM, Haider N (2015). Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. *Nature communications 6*: 8824.

Hedrick PW (2005). A standardized genetic differentiation measure. *Evolution* 59: 1633-1638.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009). Inferring weak population structure with the assistance of sample group information. *Molecular ecology resources 9*(5): 1322-1332.

Idury RM, Cardon LR (1997). A simple method for automated allele binning in microsatellite markers. *Genome Research 7*(11): 1104-1109.

Ishida S, Parker AG, Kennet D, Hodson MJ (2003). *Phytoliths* from the archaeological site of Kush, Ras' al-Khaimah, United Arab Emirates. *Quaternary Research* 59: 310-321.

Jain SM (2011). Prospects of in vitro conservation of date palm genetic diversity for sustainable production. *Emirates Journal of Food and Agriculture 23*(2): 110.

Jain SM, Gupta PK, Newton RJ (Eds.) (2013). Somatic embryogenesis in woody plants (Vol. 6). *Springer Science & Business Media*.

Jazinizadeh E, Zarghami R, Majd A, Iranbakhsh A, Tajaddod G (2015). In vitro production of date palm (*Phoenix dactylifera* L.) cv.'Barhee'plantlets through direct organogenesis. In *Biological Forum Research Trend* (Vol. 7) 2: 566.

Johnson DV (Ed.) (1996). Palms: Their conservation and sustained utilization: Status survey and conservation action plan (Vol. 31). *IUCN*.

Jost L (2008). GST and its relatives do not measure differentiation. *Mol Ecol* 17: 4015-4026.

Kashi Y, King D, Soller M (1997). Simple sequence repeats as a source of quantitative genetic variation. *Trends in genetics 13*(2): 74-78.

Kim CS, Lee CH, Shin JS, Chung YS, Hyung NI (1997). A simple and rapid method for isolation of high-quality genomic DNA from fruit trees and conifers using PVP. *Nucleic acids research 25*(5): 1085-1086.

Kimura M, Crow JF (1964). The number of alleles that can be maintained in a finite population. *Genetics* 49:725-738.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular ecology resources 15*(5): 1179-1191.

Kretzer AM, Luoma DL, Molina R, Spatafora JW (2003). Taxonomy of the Rhizopogon vinicolor species complex based on analysis of ITS sequences and microsatellite loci. *Mycologia* 95(3): 480-487.

Krueger RR (2001). Date palm germplasm: overview and utilization in USA. In: Proceedings: The first international conference on date palms. *Al-Ain*, March 1998: 2-37.

Kumpatla SP, Mukhopadhyay S (2005). Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome 48*(6): 985-998.

Latter BDH (1972). Selection in finite populations with multiple alleles. III. Genetic divergence with centripetal selection and mutation. *Genetics* 70: 475-490.

Li X, Fu J, Zhang S, Yuan L, Li M (2000). Genetic variation of inbred lines of *maize* detected by SSR markers. *Scientia Agricultura Sinica 33*(2): 1-9.

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002). Microsatellites: genomic distribution, putative functions, and mutational mechanisms: a review. *Mol. Ecol* 11: 2453-246.

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2004). Microsatellites within genes: structure, function, and evolution. *Molecular biology and evolution 21*(6): 991-1007.

Linné (1734) cited in KEANEY TH (1906): Date varieties and Date Culture in Tunis. Washington, U.S.D.A; *Bureau of Plant Industry, Bulletin No*. 92.

Liu K, Goodman M, Muse S, Smith JS, Buckler E, Doebley J (2003). Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics 165*(4): 2117-2128.

Liu L, Liu G, Gong Y, Dai W, Wang Y, Yu F, Ren Y (2007). Evaluation of genetic purity of F1 hybrid seeds in cabbage with RAPD, ISSR, SRAP, and SSR markers. *Hortscience 42*(3): 724727.

Marbeen MI, Al-Snafi AE, Marbut MM, Allahwerdy IY (2005). The probable therapeutic effects of date palm pollen in the treatment of male infertility. *Tikrit Journal of Pharmaceutical Sciences 1*(1): 30-35.

Mathew LS, Seidel MA, George B, Mathew S, Spannagl M, Haberer G, Krueger RR (2015). A genome-wide survey of date palm cultivars supports two major subpopulations in *Phoenix dactylifera*. *G3: Genes, Genomes, Genetics 5*(7): 1429-1438.

McDonald DB, Potts WK (1997). Microsatellite DNA as a genetic marker at several scales. In Avian Molecular Evolution and Systematics (D. Mindell, ed.). *Academic Press, New York*: 29-49.

Meyer RS, DuVal AE, Jensen HR (2012). Patterns and processes in crop domestication: a historical review and quantitative analysis of 203 global food crops. *New Phytologist 196*(1): 29-48.

Michalakis Y, Excoffier L (1996). A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics 142*(3): 1061-1064.

Michel-Dansac F, Caubet A (2013). L'iconographie et le symbolisme du palmier dattier dans l'Antiquité (Proche-Orient, Égypte, Méditerranée orientale). *Revue d'ethnoécologie* : 4.

Moore HE (1963). An annotated checklist of cultivated palms. *Principes* 7: 119-189.

Moore HE (1973). The major groups of palms and their distribution. *Gentes Herb 11*: 27-141.

Morton JF (1988). Notes on distribution, propagation, and products of Borassus Palms (Arecaceae). *Economic Botany 42*(3): 420-441.

MoRrcr C (1998). *Phoenix canariensis* in the wild.

Murashige T, Skoog F (1962). A revised medium for rapid growth and bioassays with *tobacco* tissue cultures. *Physiologia Plantarum 15*: 473–497.

Murashige T (1974). Plant propagation through tissue cultures. *Annual review of plant physiology 25*(1): 135-166.

Nei M, Chesser RK (1983). Estimation of fixation indices and gene diversities. *Ann Hum Genet* 47: 253-259.

Nei M, Kumar S (2000). Molecular Evolution and Phylogenetics. *Oxford University Press, New York*.

Nei M, Roychoudhury AK (1974). Sampling variances of heterozygosity and genetic distance. *Genetics* 76: 379-390.

Nei M, Tajima F, Tateno Y (1983). Accuracy of estimated phylogenetic trees from molecular data. *J Mol Evol* 19: 153-170.

Nei M (1972). Genetic distance between populations. *Am Nat* 106: 283-291.

Nei M (1973). Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* 70: 3321-3323.

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

Nei M (1987). Molecular Evolutionary Genetics. *Columbia University Press, New York*.

Nixon RW, Carpenter JB (1978). Growing dates in the United States. United States Department of Agriculture Bulletin no. 207, *U.S. Department of Agriculture, Washington, DC*.

Nobécourt P (1955). Variations de la morphologie et de la structure de cultures de tissues végétaux. *Berichte der Schweizerische Botanischen Gesellschaft* 65: 475-480.

Ohta T, Kimura M (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet Res* 22: 201-204.

Ortiz-Barrientos D, Rieseberg LH (2006). Splitting when together. *Divergence* 97: 2-3.

Paterson AH, Brubaker CL, Wendel JF (1993). A rapid method for extraction of cotton (*Gossypium spp.)* genomic DNA suitable for RFLP or PCR analysis. *Plant Molecular Biology Reporter 11*(2): 122-127.

Peakall R, Smouse PE (2001). GenAIEx v5: genetic analysis in excel. Population genetic software for teaching and research. *Canbera (Australia): Australian National University*.

Peakall R, Smouse P (2012). GenAlEx v6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28: 2537–2539.

Perrier X, Flori A, Bonnot F (2003). Methods for data analysis. Genetic diversity of cultivated tropical plants. *Science Publishers, Inc. and CIRAD, Montpellier*: 31-63.

Pérez-Vicente L, Dita MA, Martínez-de la Parte E (2014). Technical Manual Prevention and diagnostic of Fusarium Wilt (Panama disease) of banana caused by *Fusarium oxysporum* f. sp. Cubense Tropical Race 4 (TR4), Food and Agriculture Organization of the United Nations. http://www.fao.org/fileadmin/templates/banana/documents/Docs_Resources_2015/TR4/13Ma n ualFusarium.pdf. (Accessed August 2016).

Pintaud JC, Zehdi S, Couvreur T, Barrow S, Henderson S, Aberlenc-Bertossi F, Billotte N (2010). Species delimitation in the genus *Phoenix (Arecaceae*) based on SSR markers, with emphasis on the identity of the date palm (*Phoenix dactylifera* L.). *Diversity, phylogeny, and evolution in the monocotyledons. Arhus University Press, Denmark*: 267-286.

POPTREE2: *Molecular Biology and Evolution,* http://www.med.kagawa-u.ac.jp/~genomelb/takezaki/poptree2/index.html.&http://iubio.bio.indiana.edu/soft/molbio/evol-ve 27: 747-752.

Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo Á, Lareu MV (2013). An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in genetics* : *4*.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics 155*(2): 945-959.

Purugganan MD (unpublished). Whole genome sequencing of 100 date varieties. http://purugganan lab.bio.nyu.edu/index.php?action=100Dates. (Accessed February, 2016).

Rahmani AH, Aly SM, Ali H, Babiker AY, Srikar S (2014). Therapeutic effects of date fruits (*Phoenix dactylifera*) in the prevention of diseases via modulation of anti-inflammatory, antioxidant and anti-tumor activity. *International journal of clinical and experimental medicine* 7(3): 483.

Reader J (2011). Missing links: in search of human origins. *OUP Oxford*.

Reuther W, Crawford CL (1945). Irrigation experiments with dates. *Date Growers Inst. Rep*. 22: 11–14.

Reuveni O (1986). Date. In: S.P. Monselise (ed.). *CRC handbook of fruit set and development*. *CRC Press, Boca Raton, Fl*: 119–144.

Rhouma S, Dakhlaoui-Dkhil S, Salem AOM, Zehdi-Azouzi S, Rhouma A, Marrakchi M, Trifi M (2008). Genetic diversity and phylogenic relationships in date-palms (*Phoenix dactylifera* L.) as assessed by random amplified microsatellite polymorphism markers (RAMPOs). *Scientia horticulturae 117*(1): 53-57.

Rizk RM, El Sharabasy SF (2007). Descriptors for date palm (*Phoenix dactylifera* L.) characterization and evaluation in gene banks. *Plant Genetic Resources Newsletter, Bioversity International/FAO*.

Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier MH, Leroy P, Ganal MW (1998). A microsatellite map of wheat. *Genetics 149*(4): 2007-2023.

Rosenberg NA (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes 4*(1): 137-138.

Rosenberg NA, Burke T, Elo K, Feldman MW, Freidlin PJ, Groenen MA, Wimmers K (2001). Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics 159*(2): 699-713.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002). Genetic structure of human populations. *science 298*(5602): 2381-2385.

Rousset F (2008). GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol Ecol Resour* 8: 103-106.

Rougeulle A (1982). Des «étuves» à dattes à Bahrain et en Oman: le problème de l'apparition des techniques de transformation de la datte. *Paléorient* : 67-77.

Sabir JS, Arasappan D, Bahieldin A, Abo-Aba S, Bafeel S, Zari TA, Atef A (2014). Whole mitochondrial and plastid genome SNP analysis of nine date palm cultivars reveals plastid heteroplasmy and close phylogenetic relationships among cultivars. *PloS one 9*(4): 94158.

Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425.

Sakr MM, Hassan WM, Abu Zeid IM, Hassan ARE, Baz AEE (2012). The application of Random Amplified Polymorphic DNA (RAPD) in the classification of Egyptian date palm (*Phoenix dactylifera* L.) cultivars. *Journal of American Science* 8(1): 152-159.

Salem AOM, Trifi M, Sakka H, Rhouma A, Marrakchi M (2001). Genetic inheritance analysis of four enzymes in date-palm (*Phoenix dactylifera* L.). *Genetic Resources and Crop Evolution 48*(4): 361-368.

Sallon S, Solowey E, Cohen Y, Korchinsky R, Egli M, Woodhatch I, Kislev M (2008). Germination, genetics, and growth of an ancient date seed. *Science 320*(5882): 1464-1464.

Schneider S, Roessli D, Excoffier L (2000). Arlequin: a software for population genetics data analysis. *User manual 2*: 2496-2497.

Senior ML, Heun M (1993). Mapping maize microsatellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. *Genome 36*(5): 884-889.

Shriver MD, Jin L, Boerwinkle E, Deka R, Ferrell RE, Chakraborty R (1995). A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol Biol Evol* 12: 914-920.

Singh R, Ong-Abdullah M, Low ET, Manaf MA, Rosli R, Nookiah R et al. (2013). Oil palm genome sequence reveals divergence of interferfertile species in Old and New worlds. Nature 500: 335–339.

Sinnott EW (1950). Cell and Psyche the Biology of Purpose.

Skoog F, Tsui C (1948). Chemical control of growth and bud formation in tobacco stem segments and callus cultured *in vitro*. *American Journal of Botany 35*: 782–787.

Sneath PHA, Sokal RR (1973). Numerical Taxonomy. W. H. Freeman, San Francisco.

Soussa A (1969). Irrigation and Civilization in Oued Rafi din. First Part. Adib (ed.). Baghdad, Iraq. In: Zaid, A. and Arias-Jimenez, E.J. (eds.). *Date Palm Cultivation. FAO, Rome*.

Soliman SS, Ali BA, Ahmed MMM (2003). Genetic comparisons of Egyptian date palm cultivars (*Phoenix dactylifera* L.) by RAPD-PCR. *African Journal of Biotechnology* 2(4): 86-87.

Song QJ, Fickus EW, Cregan PB (2002). Characterization of trinucleotide SSR motifs in wheat. *TAG Theoretical and Applied Genetics 104*(2): 286-293.

Steward FC, Mapes MO, Mears K (1958). Growth and organized development of cultured cells. II. Organization in cultures grown from freely suspended cells. *American Journal of Botany 45*: 705–708.

Takezaki N, Nei M (1996). Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144: 389-399.

Takezaki N, Nei M (2008). Empirical tests of the reliability of phylogenetic trees constructed with microsatellite DNA. *Genetics* 178: 385-392.

Takezaki N, Nei M, Tamura K (2009). POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. *Molecular biology and evolution 27*(4): 747-752.

Tam SM, Mhiri C, Vogelaar A, Kerkveld M, Pearce SR, Grandbastien MA (2005). Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposonbased SSAP, AFLP and SSR. *Theoretical and Applied Genetics 110*(5): 819-831.

Tamura K, Stecher G, Peterson D, Filipsli A, Kumar S (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol* 30: 2725-2729.

Teklu Y, Hammer K, Huang XQ, Röder MS (2006). Analysis of microsatellite diversity in Ethiopian tetraploid *wheat* landraces. *Genetic resources and crop evolution 53*(6): 1115-1126.

Tengberg M (2012). Beginnings and early history of date palm garden cultivation in the Middle East. *Journal of Arid Environments* 86: 139-147.

Thorpe TA (1990). The current status of plant tissue culture. In Bhojwani SS (Ed.), *Plant tissue culture: Applications and limitations, Amsterdam: Elsevier*: 1-33.

Tisserat B (1979a). Propagation of date palm (*Phoenix dactylifera* L.) in vitro. *Journal of Experimental Botany 30*(6): 1275-1283.

Tisserat B (1979b). Tissue culture of the date palm. *Journal of Heredity* 70(3): 221-222.

Trifonov EN, Berezovsky IN (2003). Evolutionary aspects of protein structure and folding. *Current opinion in structural biology 13*(1): 110-114.

Tsui C (1948). The role of zinc in auxin synthesis in the tomato plan. *American Journal of Botany* : 172-179.

Uhl NW, Dransfield J (1987). Genera palmarum: a classification of palms based upon the work of Harold E Moore, Jr. *Allen Press, Lawrence*.

UNCTAD. Statistics. Palais des Nations, Av. de la Paix, 1211 Genève 10, Suisse. http://unctad.org/fr/Pages/statistics.aspx : 8-14. (Accessed December 2016)

UNSC. Statistics of International Trade in Services and Tourism. International Merchandise Trade

Statistics (IMTS) Section. https://unstats.un.org/Unsd/tradeserv/default.htm. (Accessed December 2016)

Vrydaghs L, Doutrelpont H, Beeckman H, Haerinck E (2001). Identifications of a morphotype association of *Phoenix dactylifera* L. lignified tissues origin at ed- Dur (1st AD), Umm al Qaiwain (U.A.E.). In Meunier, J.D., Colin, F. (Eds.), *Phytoliths in Earth Sciences and Human History*. *AA Balkema, Rotterdam*: 239-250.

Watson L, Dallwitz MJ (1992) onwards. The families of flowering plants: descriptions, illustrations, identification, and information retrieval. *Version*: 1st June 2007. delta-intkey.com.

Wei L, Cao X (2016). The effect of transposable elements on phenotypic variation: insights from plants to humans. *Science China Life Sciences 59*(1): 24-37.

White PR (1934). Potentially unlimited growth of excised tomato root tips in a liquid medium. *Plant Physiology 9*(3): 585.

Willcox G (1990). The plant remains from Hellenistic and Bronze age levels at Failaka, Kuwait. A preliminary report. In Calvet Y, Gachet J (Eds.), Failaka. *Fouilles Françaises* : 1986-1988. *Travaux de la Maison de l'Orient 18*, Lyon: 43-50.

Wright S (1951). The general structure of populations. *Ann Eugen* 15: 323-352.

Wrigley G (1995). Date palm (*Phoenix dactylifera*). In: Smartt J, Simmonds NW (eds.) The evolution of crop plants, 2nd edn. Longman, Essex.

Wu KS, Tanksley SD (1993). Genetic and physical mapping of telomeres and macrosatellites of rice. *Plant molecular biology 22*(5): 861-872.

Xiao H, Jiang N, Schaffner E, Stockinger EJ, Van Der Knaap E (2008). A retrotransposon mediated gene duplication underlies morphological variation of tomato fruit. *science 319*(5869): 1527-1530.

Xiao L, Fayer R (2008). Molecular characterization of species and genotypes of Cryptosporidium and Giardia and assessment of zoonotic transmission. *International journal for parasitology 38*(11): 1239-1255.

Xiao L, Zhao Z, Du D, Yao Y, Xu L, Tang G (2012). Genetic characterization and fine mapping of a yellow-seeded gene in Dahuang (a Brassica rapa landrace). *Theoretical and applied genetics 124*(5): 903-909.

Xiao Y, Xia W, Ma J, Mason AS, Fan H, Shi P, Peng M (2016). Genome-wide identification and transferability of microsatellite markers between Palmae species. *Frontiers in plant science 7*: 1578.

Xu J, Qian X, Wang X, Li R, Cheng X, Yang Y, Liu K (2010). Construction of an integrated genetic linkage map for the A genome of Brassica napus using SSR markers derived from sequenced BACs in B. rapa. *BMC genomics 11*(1): 594.

Zaid A, Arias-Jimenez EJ (1999). Date palm cultivation (No. 156). *Food and Agriculture Organization* (FAO).

Zaid A, De Wet PF (2002). Date Palm Cultivation. Plant Production and Protection Paper, No 156. Rome. FAO.

Zaid A, De Wet PF (2002). Pollination and bunch management. In A. Zaid (ed.). Date palm cultivation. Food and Agriculture Organization Plant Production and Protection Paper no. 156.

Food and Agriculture Organization of the United Nations, Rome, Italy: 145-175.

Zaid A, De Wet PF, Djerbi M, Oihabi A (2002). Diseases and pests of date palm. In A. Zaid (ed.). Date palm cultivation. Food and Agriculture Organization Plant Production and Protection Paper no. 156. Food and Agriculture Organization of the United Nations, Rome, Italy: 227-281.

Zehdi S, Trifi M, Billotte N, Marrakchi M, Christophe Pintaud J (2004). Genetic diversity of Tunisian date palms (*Phoenix dactylifera* L.) revealed by nuclear microsatellite polymorphism. *Hereditas 141*(3): 278-287.

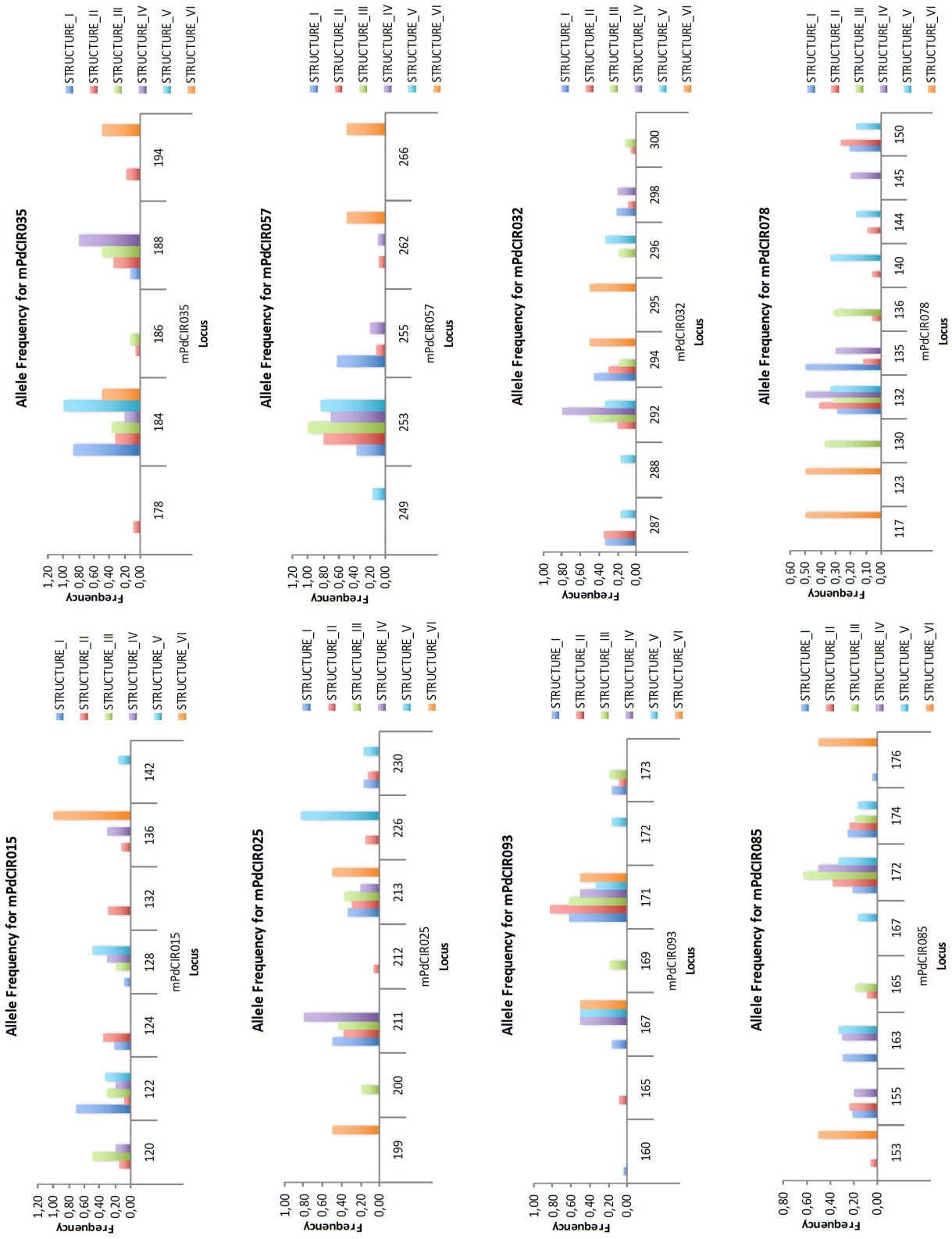
Zehdi S, Cherif E, Rhouma S, Santoni S, Hannachi AS, Pintaud JC (2012). Molecular polymorphism and genetic relationships in date palm (*Phoenix dactylifera* L.): The utility of nuclear microsatellite markers. *Scientia horticulturae 148*: 255-263.

Zhang J, Kapli P, Pavlidis P, Stamatakis A (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29(22): 2869-2876.

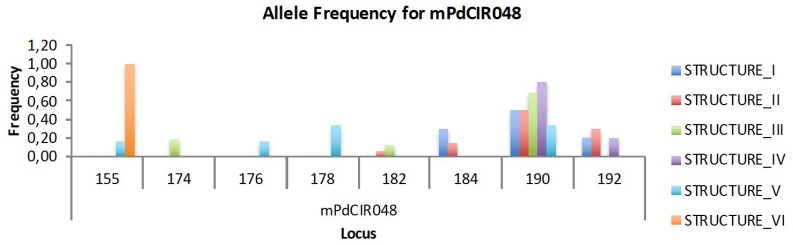
Zhao Y, Williams R, Prakash CS, He G (2012). Identification and characterization of gene-based SSR markers in date palm (*Phoenix dactylifera* L.). *BMC plant biology* 12(1): 237.

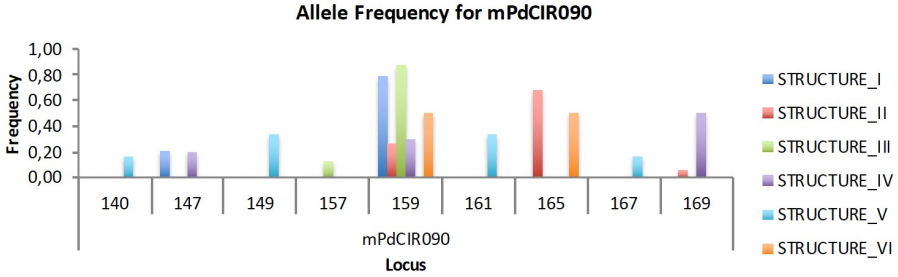
Zohary D, Spiegel-Roy P (1975). Beginnings of fruit growing in the old world. *Science* 187: 319327.

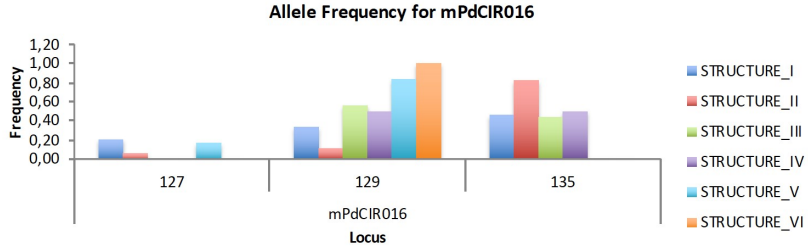
Zohary D, Hopf M, Weiss E (2012). Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin. *Oxford University Press on Demand*.

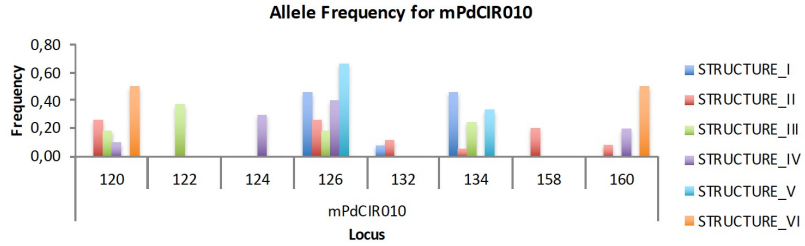


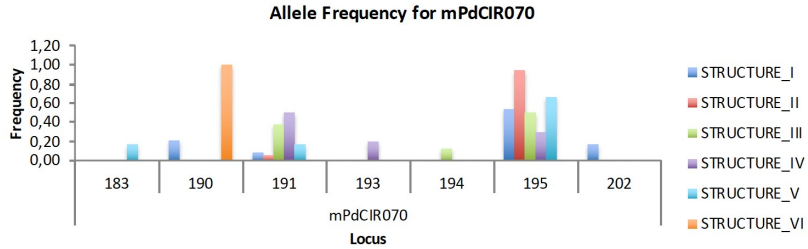
4.1. Allele frequencies with graphs by population and locus for codominant data.

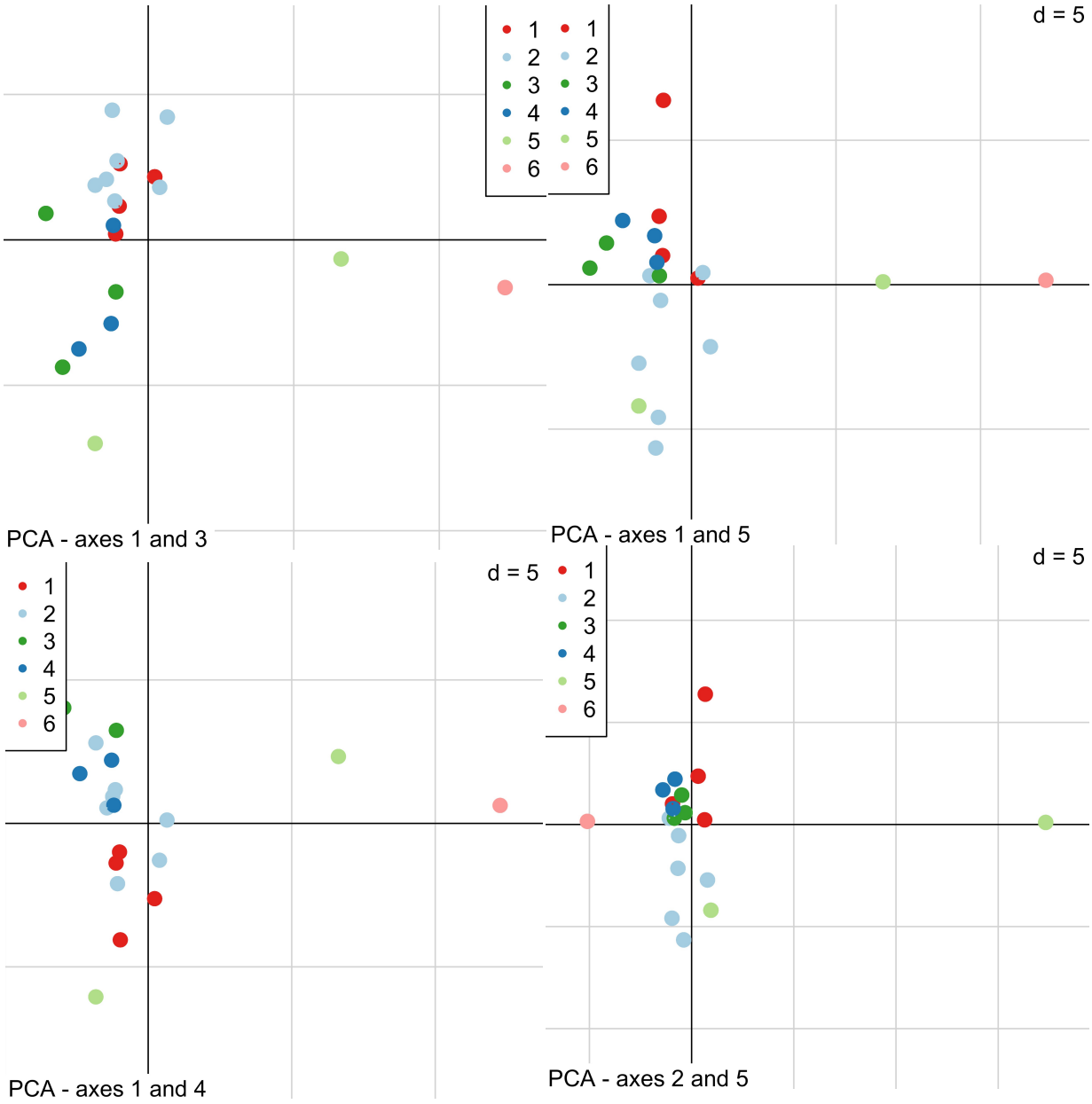

Allele Frequency for mPdCIR048
Frequency
1,20
1,00
0,80
0,60
0,40
0,20
0,00
155
174
176
178
182
184
190
192
mPdCIR048
Locus
STRUCTURE_I
STRUCTURE_II
STRUCTURE_III
STRUCTURE_IV
STRUCTURE_V
STRUCTURE_VI


Allele Frequency for mPdCIR090
Frequency
1,00
0,80
0,60
0,40
0,20
0,00
140
147
149
157
159
161
165
167
169
mPdCIR090
Locus
STRUCTURE_I
STRUCTURE_II
STRUCTURE_III
STRUCTURE_IV
STRUCTURE_V
STRUCTURE_VI

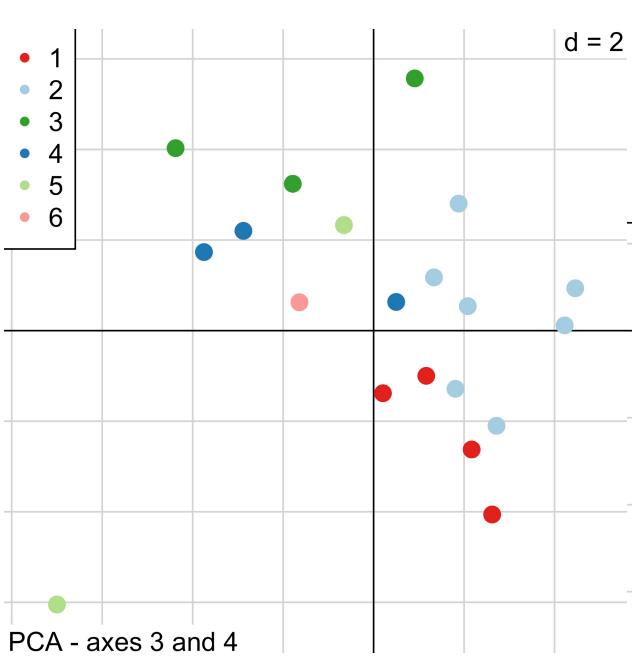

Allele Frequency for mPdCIR016
Frequency
1,20
1,00
0,80
0,60
0,40
0,20
0,00
127
129
135
mPdCIR016
Locus
STRUCTURE_I
STRUCTURE_II
STRUCTURE_III
STRUCTURE_IV
STRUCTURE_V
STRUCTURE_VI


Allele Frequency for mPdCIR010
Frequency
0,80
0,60
0,40
0,20
0,00
120
122
124
126
132
134
158
160
mPdCIR010
Locus
STRUCTURE_I
STRUCTURE_II
STRUCTURE_III
STRUCTURE_IV
STRUCTURE_V
STRUCTURE_VI

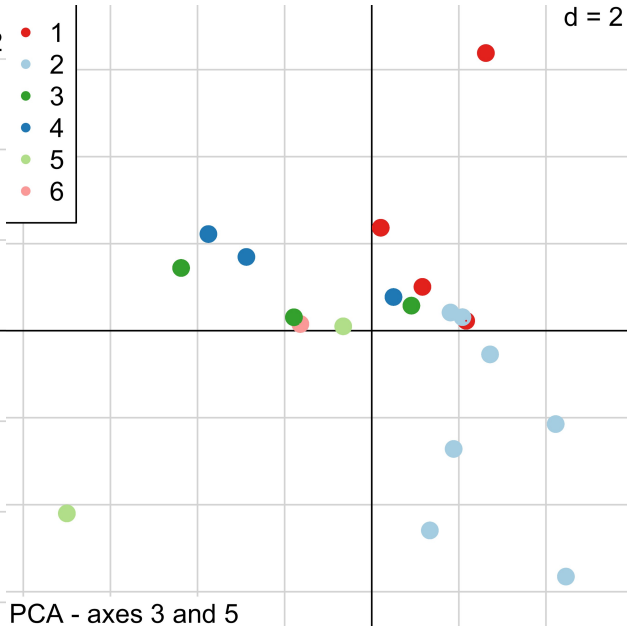

Allele Frequency for mPdCIR070
Frequency
1,20
1,00
0,80
0,60
0,40
0,20
0,00
183
190
191
193
194
195
202
mPdCIR070
Locus
STRUCTURE_I
STRUCTURE_II
STRUCTURE_III
STRUCTURE_IV
STRUCTURE_V
STRUCTURE_VI



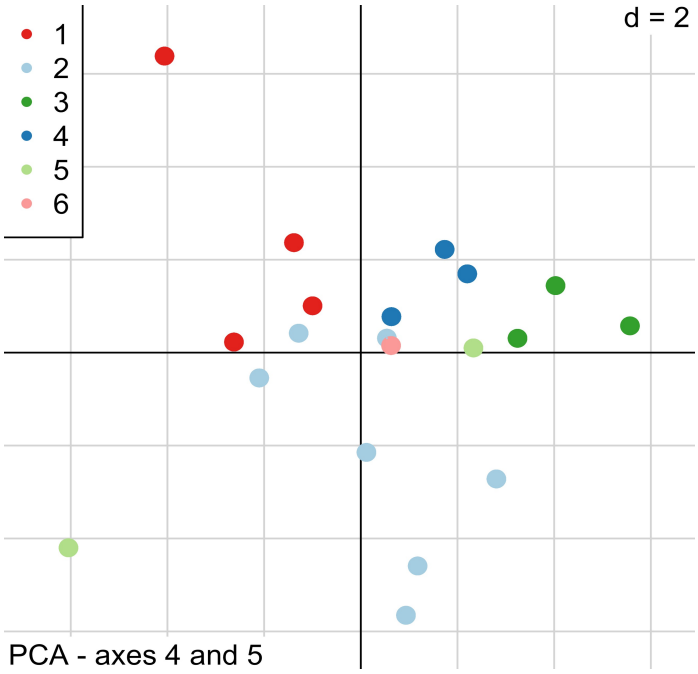
4.2. Axis(es) PCA combinations - multiple factor analysis detecting subpopulations.



PCA - axes 3 and 4



PCA - axes 3 and 5



PCA - axes 4 and 5

4.3. Heterozygosity, Fstatistics by population for codominant data.

		Sample Size	No. Alleles	No. Effective Alleles	Info. Index	Observed Heterozygosity	Expected Heterozygosity	Unbiased Expect Heterozygosity	Fixation Index
Pop		**N**	**Na**	**Ne**	**I**	**Ho**	**He**	**uHe**	**F**
STRUCTURE_I	**Mean**	60.000	3.077	2.408	0.931	0.750	0.547	0.551	-0.350
	SE	0.000	0.239	0.205	0.081	0.076	0.041	0.041	0.074
STRUCTURE_II	**Mean**	85.000	4.231	2.887	1.103	0.670	0.567	0.570	-0.167
	SE	0.000	0.361	0.344	0.131	0.082	0.063	0.064	0.046
STRUCTURE_III	**Mean**	40.000	2.846	2.336	0.879	0.625	0.520	0.526	-0.197
	SE	0.000	0.222	0.198	0.099	0.092	0.056	0.056	0.109
STRUCTURE_IV	**Mean**	50.000	2.692	2.265	0.843	0.723	0.512	0.517	-0.380
	SE	0.000	0.208	0.212	0.085	0.089	0.043	0.043	0.102
STRUCTURE_V	**Mean**	30.000	2.923	2.469	0.887	0.641	0.509	0.517	-0.205
	SE	0.000	0.288	0.286	0.125	0.116	0.066	0.067	0.126
STRUCTURE_VI	**Mean**	20.000	1.692	1.692	0.480	0.692	0.346	0.355	-1.000
	SE	0.000	0.133	0.133	0.092	0.133	0.067	0.068	0.000
Grand Mean and SE over Loci and Pops									
		N	**Na**	**Ne**	**I**	**Ho**	**He**	**uHe**	**F**
Total	**Mean**	47.500	2.910	2.343	0.854	0.684	0.500	0.506	-0.354
	SE	2.412	0.131	0.103	0.046	0.040	0.024	0.024	0.046

4.4. Fstatistics and estimates of Nm over populations for each locus.

All Pops.	**Locus**	**Ht**	**Mean He**	**Mean Ho**		**Fis**	**Fit**	**Fst**		**Nm**
	mPdCIR015	0.806	0.527	0.744		-0.412	0.076	0.346		0.473
	mPdCIR025	0.754	0.512	0.655		-0.281	0.131	0.322		0.527
	mPdCIR032	0.805	0.596	0.838		-0.405	-0.040	0.260		0.713
	mPdCIR085	0.809	0.646	0.876		-0.356	-0.083	0.201		0.993
	mPdCIR048	0.716	0.463	0.671		-0.449	0.064	0.353		0.457
	mPdCIR070	0.676	0.409	0.561		-0.373	0.169	0.395		0.383
	mPdCIR078	0.841	0.643	0.980		-0.524	-0.165	0.236		0.812
	mPdCIR010	0.801	0.623	0.748		-0.202	0.066	0.223		0.873
	mPdCIR035	0.601	0.393	0.569		-0.446	0.053	0.345		0.475
	mPdCIR057	0.574	0.342	0.488		-0.426	0.149	0.403		0.370
	mPdCIR093	0.593	0.501	0.620		-0.236	-0.045	0.154		1.369
	mPdCIR016	0.547	0.368	0.455		-0.234	0.169	0.326		0.516
	mPdCIR090	0.733	0.477	0.680		-0.426	0.072	0.350		0.465
					Mean	-0.367	0.047	0.301		0.648
					SE	0.027	0.029	0.022		0.081

Nm = ((1 / Fst) - 1) / 4, Ht = Total He = 1-Sum tpi^2 (tpi is the frequency of the ith allele for the total & sum tpi^2 is the sum of the squared total allele frequencies).

4.5. Percentages of polymorphic Loci.

Population	%P
STRUCTURE_I	100.00%
STRUCTURE_II	100.00%
STRUCTURE_III	92.31%
STRUCTURE_IV	100.00%
STRUCTURE_V	92.31%
STRUCTURE_VI	69.23%
Mean	92.31%
SE	4.87%

4.6. Pairwise population matrix of Nei genetics distance.

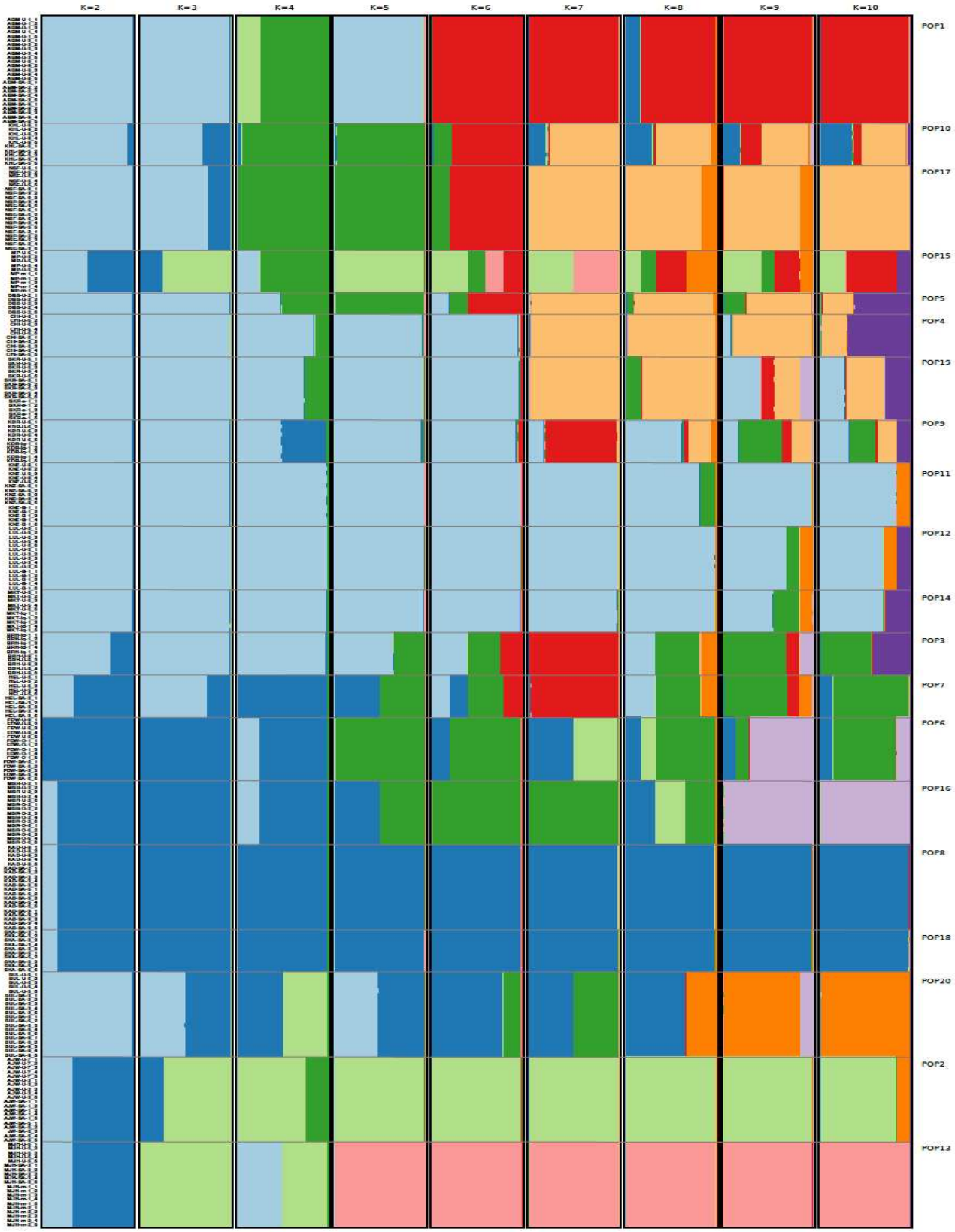
STRUCTURE_I	STRUCTURE_II	STRUCTURE_III	STRUCTURE_IV	STRUCTURE_V	STRUCTURE_VI	
0.000						STRUCTURE_I
0.357	0.000					STRUCTURE_II
0.404	0.336	0.000				STRUCTURE_III
0.484	0.412	0.282	0.000			STRUCTURE_IV
0.563	0.676	0.593	0.642	0.000		STRUCTURE_V
1.174	1.286	1.357	1.378	1.347	0.000	STRUCTURE_VI

4.7. Pairwise population matrix of Nei genetic identity.

STRUCTURE_I	STRUCTURE_II	STRUCTURE_III	STRUCTURE_IV	STRUCTURE_V	STRUCTURE_VI	
1.000						STRUCTURE_I
0.700	1.000					STRUCTURE_II
0.668	0.715	1.000				STRUCTURE_III
0.616	0.662	0.754	1.000			STRUCTURE_IV
0.569	0.508	0.553	0.526	1.000		STRUCTURE_V
0.309	0.276	0.257	0.252	0.260	1.000	STRUCTURE_VI

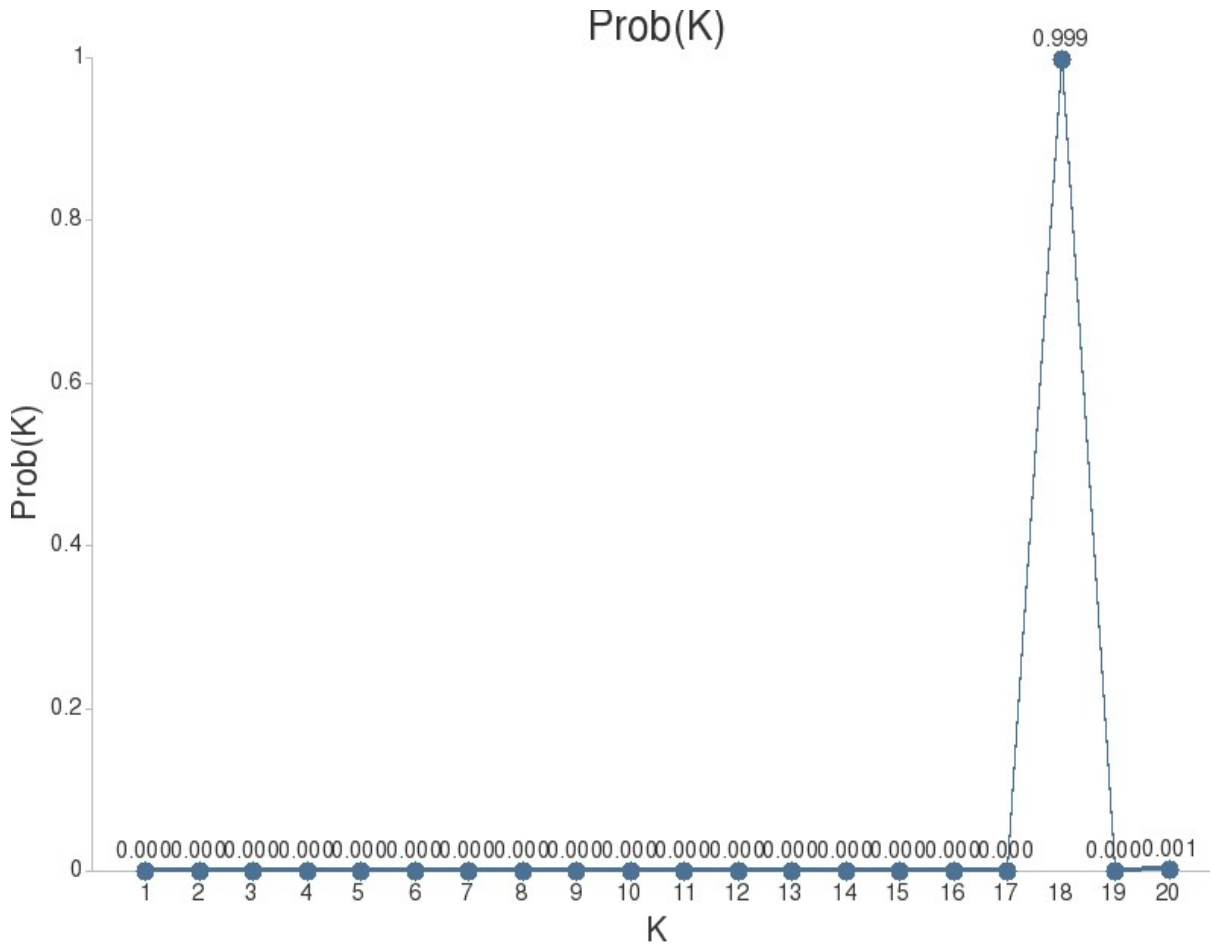
4.8. Pairwise population Fst values.

STRUCTURE_I	STRUCTURE_II	STRUCTURE_III	STRUCTURE_IV	STRUCTURE_V	STRUCTURE_VI		
0.000						STRUCTURE_	I
0.111	0.000					STRUCTURE_	II
0.131	0.112	0.000				STRUCTURE_	III
0.147	0.127	0.102	0.000			STRUCTURE_	IV
0.155	0.175	0.172	0.192	0.000		STRUCTURE_	V
0.301	0.319	0.330	0.333	0.322	0.000	STRUCTURE_	VI



4.9. Bayesian clustering of 20 cultivar genotypes at 13 SSR loci of 285 individuals implemented in STRUCTURE, K = 2 to K = 10 population stratification.

4.10. The maximum value of ln P (D) by the Pritchard method (Pritchard et al., 2000), identifying K.



List of abbreviations

Molecular

2,4-D	- 2,4-dichlorophenoxyacetic acid
2,4,5-T	- 2,4,5-trichlorophenoxyacetic acid
2iP	- N6-Δ2 isopentenyl adenine
6-FAM	- 6-carboxyfluorescein
AC	- activated charcoal
AFLP	- amplified fragment length polymorphism
ATP	- adenosine-5'-triphosphate
B5	- gamborg's B5 medium
BAP	- benzylaminopurine
BGT	- bud generative tissue
bp	- base pair
BSA	- bovine serum albumin
CAPS	- cleaved Amplified Polymorphic Sequence
cDNA	- complementary DNA
conc	- conclusion
CTAB	- hexadecyltrimethyl-ammonium bromide
Cvs. (Cv.)	- cultivar
dATP	- 2'-deoxyadenosine-5'-triphosphate
DNA	- deoxyribonucleic acid

DNase - deoxyribonuclease

dNTP - 2'-deoxynucleotide-5'-triphosphate

DP - date palm

E. coli - escherichia coli

EDTA - ethylenediamine tetra-acetate

EtBr - ethidium bromide

Foc - *fusarium oxysporum* f. sp. *cubense*

G - guanine

GA3 - gibberellic acid (3)

GA7 - gibberellic acid (7)

gDNA - genomic deoxyribonucleic acid

IAA - indole acetic acid

IBA - indole-3-butyric acid

ISSR - inter simple sequence conformation polymorphism

kb - kilobase

KIN - kinetin, 6-furfurylaminopurine

MS - murashige and skoog

MSI - microsatellite instability linkages

Mya - million years ago

Mwt - molecular weight

NAA - α-naphthaleneacetic acid

NaOCl - sodium hypochlorite

NOA - naphthoxy acetic acid

OD	- optical density
PAGE	- polyacrylamide gel electrophoresis
PCR	- polymerase chain reaction
PVP	- polyvinyl pyrrolidone
RAPD	- random amplified polymorphic DNA
RDA	- representational difference analysis
RAFLP	- restriction amplification fragment length polymorphisms.
RFLP	- restriction fragment length polymorphism
RNA	- ribonucleic acid
RNase	- ribonuclease
SCAR	- sequence characterization amplified region
SDS	- sodium dodecyl sulphate
SNP	- single nucleotide polymorphism
SSR	- simple sequence repeats
STR	- short tandem repeat
STS	- sequence tagged site
T	- thymine
Taq	- thermus aquaticus
TBE	- tris-borate EDTA
TC	- tissue culture
TCS	- trilateral confirmation system
TEMED	- N,N,N',N'-tetramethylethylenediamine
Tris	- trishydroxymethylaminomethane

UV - ultraviolet (light)

VIC - 6-carboxyrhodamine

Miscellaneous

DPTCL - Date Palm Tissue Culture Laboratory

EC - European Community

FAO - Food and Agriculture Organization of the United Nations

GDP - Gross Domestic Product

GNP - Gross National Product

KCBGE - Khalifa Center for Biotechnology and Genetic Engineering

KSA - Kingdom of Saudi Arabia

MT - Metric Tons

RSA - Republic of South Africa

UAE - United Arab Emirates

UAEU - United Arab Emirates University

UK - United Kingdom

UNDP - United Nations Development Programme

UNOPS - United Nations Office for Project Services

$US - United State Dollar

USA - United States of America