Annual Plant Reviews (2021) **4**, 57–86
doi: 10.1002/9781119312994.apr0748

http://onlinelibrary.wiley.com

THE GENOMICS OF DATE PALMS: CURRENT PERSPECTIVES

Tiago Capote[1], Jonathan M. Flowers[1], Khaled M. Hazzouri[2], Michael Purugganan[1] and Khaled M.A. Amiri[2,3]

[1] *Center for Genomics and Systems Biology, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi United Arab Emirates*
[2] *Khalifa Center for Genetic Engineering and Biotechnology, Al Ain, Abu Dhabi, United Arab Emirates*
[3] *Department of Biology, United Arab Emirates University, Al Ain, Abu Dhabi, United Arab Emirates*

Abstract: The date palm (*Phoenix dactylifera*) is one of the most important fruit crops in the world and is especially important in the arid regions of the Middle East and North Africa. Although both economically and culturally important, the biology of date palms remains obscure. In the last decade, however, there have been major advances in genomic studies of this important crop species. Multiple reference genomes are now available, which has facilitated identification of genes underlying major traits such as sex, fruit colour and sugar composition. Moreover, whole-genome re-sequencing from some of the approximately 3000 date palm varieties have provided crucial insights into the diversity and evolution of this perennial tree crop. The overview of the field highlights major advances and also provides a guidepost for future challenges.

Keywords: plant genomics, fruit trees, domestication, crop genomics

1 Introduction

The date palm (*Phoenix dactylifera*) is a dioecious, perennial diploid tree species in the Arecaceae family and is the most important fruit-bearing crop of the Middle East and North Africa (Barrow, 1998; Food and Agriculture Organization of the United Nations, 2020). Date palms grow primarily in hot, arid habitats including desert oases, river valleys, and today in well-irrigated

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

farms. Individual varieties are valued primarily for fruit-related traits including moisture and sugar content, and date palms not only play a critical role in the agricultural economies of the arid regions of the Middle East and North Africa, but they are also a major cultural icon in many regions.

The study of date palms provide key information on the genetics of the domestication and diversification of fruit tree crops (Gros-Balthazard et al., 2018), and as a mainstay of arid land agriculture, they can also provide clues on adaptation of crops to desert regions. There is interest in developing new varieties with even more abiotic stress tolerance, particularly to salinity (Hazzouri et al., 2020a), to biotic challenges such as the red palm weevil (Hazzouri et al., 2020b) and to greater yield and fruit quality. Breeding and/or genetic modification of date palms, however, remain challenging, especially given its prolonged juvenile phase (4–5 years), and despite strong interest, there has been little recent development of new varieties.

Recently, however, there has been concerted effort to study the genomics of date palms, which has provided key genomic resources and enabled new genetic analyses that can serve as the foundation for future efforts at agronomic improvement as well provide clues to the domestication and diversification of perennial tree crops. In this article, we discuss some of these emerging resources and some of the areas that new studies have begun to illuminate.

2 Date Palms and the Genus *Phoenix*

2.1 The Genus *Phoenix*

The domesticated date palm is one of 13 palm species in the genus *Phoenix* L., a group that is found in the Canary and Cape Verde Islands in the west, across northern and central Africa, the extreme southeast of Europe (Crete), and southern Asia from Turkey to southern China and Malaysia, including all of the Arabian Peninsula (Barrow, 1998; Dransfield et al., 2008; Gros-Balthazard et al., 2017; Henderson et al., 2006). Members of the genus consist of palm trees that are mostly medium to large in size, but also includes dwarf species. These plants have pinnate leaves, 1–6 m long, which all share the common feature of having metamorphosed the lower-leaf segments into long, sharp spines. The leaves have short or absent petioles and possess the rare feature among pinnate palms of induplicate (V-shaped) leaflets (Barrow, 1998; Dransfield et al., 2008; Krueger, 2011).

The members of the genus *Phoenix* are dioecious, with male and female flowers on separate plants, and pollination is facilitated by wind and insect vectors (Barrow, 1998; Dransfield et al., 2008). The flowers are inconspicuous yellowish-brown about 1 cm wide, and the inflorescence emerges from boat-shaped leathery bracts forming large pendent clusters. *Phoenix* fruits

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

develop from one carpel as a drupe, 1–7 cm long, and range in colours from yellow to red-brown or dark purple when mature, and contain one elongated, deeply grooved seed (Popenoe, 1922). Flowering and fruiting are regular and annual, with reproduction occurring via seeds as well as by vegetative multiplication (Chao and Krueger, 2007). Many species of *Phoenix* produce vegetative offshoots called bulbils from the basal portions of their stems which, on rooting, develop new saplings (Chao and Krueger, 2007; Krueger, 2011).

The taxonomy of the genus *Phoenix* has not been well established in the literature until recently, and there has been disagreement among taxonomic treatments which lead to a lack of consensus regarding species name and validity. Although 19 species of *Phoenix* have been named, most taxonomic treatments of *Phoenix* list about 13 species as valid, although there is not necessarily agreement on species delimitation (Barrow, 1998; Jaradat, 2015; Krueger, 2011; Pintaud et al., 2010).

The monograph of Barrow (1998) clarified the relationship within this genus, reporting 13 species, including one new species from the Andaman Islands and two varieties within *Phoenix loureiroi* and *Phoenix humilis*. Moreover, Barrow (1998) defined species limits and distributions as well as performed systematic analysis of the genus, combining data from studies of morphology and lamina anatomy with sequence analysis of nuclear ribosomal deoxyribonucleic acid (DNA) data. The Barrow (1998) taxonomic analysis has become widely accepted and describes *Phoenix acaulis*, *Phoenix andamanensis*, *Phoenix caespitosa*, *Phoenix canariensis*, *P. dactylifera*, *Phoenix loureiroi* var. *loureiroi*, *Phoenix loureiroi* var. *humilis*, *Phoenix paludosa*, *Phoenix pusilla*, *Phoenix reclinata*, *Phoenix roebelenii*, *Phoenix rupicola*, *Phoenix sylvestris*, *Phoenix theophrasti* as species (Table 1). More recently, González-Pérez (2004) found genetic evidence that suggests that *Phoenix atlantica* should be recognized as a distinct species (González-Pérez et al., 2004), although population genomic analysis indicates it may be a feral *P. dactylifera* (Flowers et al., 2019; Gros-Balthazard et al., 2017). It should be noted that the 13 species of the *Phoenix* genus are closely related, as evidenced by easy cross-pollination and hybridization (Gros-Balthazard, 2013) and low levels of divergence, at least among the closest relatives of cultivated date palm (Flowers et al., 2019).

2.2 *P. dactylifera*, the Domesticated Date Palm

Phoenix dactylifera L. is a unique species cultivated worldwide primarily for its fruit, the date, which represents the largest fruit among *Phoenix* species (Barrow, 1998; Hazzouri et al., 2019). Date palms grow well in arid conditions where water is close to the surface, and its leaves are adapted to hot and dry conditions. Despite its adaptation to arid conditions, it is not a xerophyte and requires a source of abundant water for growth (Krueger, 2011);

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

Table 1 *Phoenix* species taxonomic, common names and world geographical distribution.

Species	Common name	Geographical distribution
P. acaulis	Stemless date palm	Bhutan, Nepal, North India
P. anadamensis	Andaman Island date palm	Bay of Bengal, Andaman Islands
P. atlantica[a]	Cape Verde Island date palm	Cape Verde Islands, Morocco
P. caespitosa	Date palm	Gulf of Aden shores, Arabia Peninsula
P. canariensis	Canary Island date palm	Canary Islands, Italy, Spain
P. dactylifera	Date palm	Middle East, North Africa, Pakistan, India, Australia, California
P. loureiroi	Mountain date palm	China, Himalayas mountains, Indochina, Malaysia, Philippines
P. paludosa	Mangrove date palm	Andaman Islands, India, Indochina, Malaysia, Sumatra
P. pusilla	Ceylon date palm	South India, Sri Lanka
P. reclinata	Senegal date palm	South Arabian Peninsula, Africa, Comoros, Madagascar
P. roebelenii	Pygmy date palm	South China to North Indo-China
P. rupicola	Cliff date palm	India, Bhutan, Andaman Islands
P. sylvestris	Rain palm/Indian date palm	India, Pakistan, Southern China
P. theophrasti	Cretan palm	Crete, Greek Islands, coastal Turkey

[a]Not fully recognized as a distinct species.
Source: Adapted from Al-Alawi et al. (2017), Barrow (1998), Khan et al. (2018), Krueger (2011), Pintaud et al. (2010).

thus, although the date palm is found in geographic areas with long, very hot seasons with little or no rain, it is cultivation relies on abundant underground water. Typically, these conditions are found in oases and river valleys in the arid sub-tropical deserts of the Middle East and North Africa, and the traditional range of date palms spans Morocco in the west to Pakistan and India in the east. More recently, date palm has been introduced and cultivated in the southwestern USA, Mexico, South America, and Australia (Jaradat, 2015; Johnson et al., 2015; Krueger, 2015).

It is estimated that there may be as many as 3000 named varieties of dates although some are expected to be synonymous and homonymous (Gros-Balthazard et al., 2020; Krueger, 2011). Population genetic studies using microsatellites (Arabnezhad et al., 2012; Cherif et al., 2013; Gros-Balthazard et al., 2017; Zehdi-Azouzi et al., 2015; Zehdi et al., 2012) and whole-genome sequences (Gros-Balthazard et al., 2017; Hazzouri et al., 2015; Mathew et al., 2015) suggest two genetically distinct populations, one in the Middle East and the other in the North Africa with hybrids found in a contact zone in and around Egypt and Sudan (Gros-Balthazard et al., 2017; Hazzouri et al., 2015; Mathew et al., 2015). Additional fine-scale population structure has also been reported within these two divergent and well-defined geographic populations (Mohamoud et al., 2019; Zehdi-Azouzi et al., 2015).

The exact origin of cultivated date palm remains unclear. Although *P. sylvestris* found in India is the closest species to *P. dactylifera*, the ancestor of date palms remains a controversial subject (Barrow, 1998; Chaluvadi et al., 2019; Pintaud et al., 2010; Zehdi-Azouzi et al., 2015). Archaeological evidence for the earliest exploitation and consumption of date palms have been found in Dalma Island, United Arab Emirates and As-Sabiyah, Kuwait, dating to the Arabian Neolithic at ~5100 BCE (Beech, 2003; Beech and Shepherd, 2001). Evidence of widespread cultivation in Mesopotamia dates back to 4700–4000 BCE and supports the hypothesis that the region around the Arabian Gulf was the centre of domestication (Zohary et al., 2012; Zohary and Spiegel-Roy, 1975). Interestingly, population genomic studies have identified wild date palms in Oman as an old lineage of *P. dactylifera* that may represent a relictual population of a wild progenitor of domesticated date palms (Gros-Balthazard et al., 2017). The ancient distribution of this species is poorly understood, but several reports have suggested that North African date palms were domesticated independently from Middle Eastern *P. dactylifera* (Gros-Balthazard et al., 2017; Hazzouri et al., 2015; Zehdi-Azouzi et al., 2015). It has been shown, however, that introgressive hybridization of Middle Eastern *P. dactylifera* by a *P. theophrasti*-like population, which today occurs largely in Crete in the eastern Mediterranean, may account for much (but possibly not all) of the genetic diversity and genomic distinctiveness of North African date palms (Flowers et al., 2019) [see below].

3 Genomic and Genetic Resources

3.1 Draft Genome Sequences: A Key Resource

Date palms have been reported to have different ploidy levels (Al-Ani et al., 2010) and genome size estimates that range from 680 to 1491 Mb (reviewed in Al-Mssallem et al., 2013). This raises the possibility of genome size variation with date palms, but a recent study that utilized the same flow cytometry technique on multiple varieties concluded that the species is a diploid ($2n = 36$) with a genome size of 870–899 Mb (Hazzouri et al., 2019). At the chromosome level, the date palm genome has long-range synteny with the oil palm *Elaeis guineensis* despite the fact that the oil palm genome is almost double in size (1.54 Gb, RefSeq accession: GCF_000442705.1) although it has only 16 chromosomes ($2n = 32$) (Singh et al., 2013); this species diverged from *P. dactylifera* ~65–85 million years ago (Baker and Couvreur, 2013; Singh et al., 2013). Synteny analysis between the date palm genetic map and oil palm chromosomes suggests that oil palm chromosome 2 is the result of a Robertsonian fusion involving date palm linkage groups (LGs) 1 and 10, while oil

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

palm chromosomes 1, 6, and 10 is syntenic to date palm LG11 (Mathew et al., 2014).

The first publicly available date palm draft genome sequence was released in 2011 (Al-Dous et al., 2011) by a team from Cornell-Qatar, and constituted a major advance in date palm genomics and genetics. The ~380 Mb draft genome of the female variety Khalas (NCBI accession number ACYX02, Table 2), with an N50 of 30.5 Kb, spanned primarily gene-rich regions and identified >25 000 genes, although it covered only ~43% of the date palm genome. This study found >3.5 million polymorphic sites, including >10 000 gene copy number variants among varieties, including in the three elite date palm varieties Khalas, Deglet Noor, and Medjool (Al-Dous et al., 2011). The annotation of this genome constituted an important breakthrough in the study of gene content and allelic variation in date palms, not least being the identification of potential sex-linked regions in the genome that could underlie dioecy. This assembly has been updated by incorporating long-read and linked read sequencing data, which resulted in a new assembly (NCBI accession number ACYX03, Table 2) with 11 170 contigs and a contig N50 of ~114 Kb [unpublished data]. This as yet unpublished assembly allowed the Cornell-Qatar group to scaffold the ACYX03 draft genome into 906 scaffolds with a scaffold N50 of 1.45 Mb.

In 2013, a collaborative Saudi-China effort (Al-Mssallem et al., 2013) published a second draft genome sequence, also from the Khalas variety, based on complementary data from different sequencing platforms. This genome (NCBI accession number ATBV01, Table 2) covered 605.4 Mb of genome sequence with an N50 of 330 Kb and contained an estimated >96% of its genes (~41 660). This genome assembly, an improvement both in precision and contiguity, provided evidence for ancient genome-wide duplication (GWD) events and/or massive consecutive segmental duplications. This study also reported that stress-resistance and sugar metabolism-related gene regions were enriched in chromosomal regions with single nucleotide polymorphism (SNP) deserts, where SNP density is relatively low. It was suggested that these regions reflect trait selection under cultivation and thus may be a signature of domestication in date palm (Al-Mssallem et al., 2013).

In 2019, a collaborative effort between NYU Abu Dhabi, the Khalifa Center for Genetic Engineering and Biotechnology, and the University of Arizona (NYUAD/KCGEB/UoA) took advantage of SMRT long-read sequencing and released a new date palm draft genome (Hazzouri et al., 2019). Obtained from a male plant that resulted from 4 generations of backcrossing to the variety Barhee (the 'BC4 male' assembly), this long-read based draft genome sequence spanned 772 Mb assembled into 2390 scaffolds and 2706 contig sequences, with half of the contigs assigned to LGs based on a previously published genetic map (Mathew et al., 2014). This genome assembly was annotated as having 36 162 non-transposable elements (TEs) protein-coding

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

Table 2 Date palm genome sequence assemblies in public databases.

NCBI[a] WGS project	ACYX02	ATBV01	ACYX03	PEFZ01	NBZB01
Release year/publication	2011 (Al-Dous et al., 2011)	2013 (Al-Mssallem et al., 2013)	2020 (no publication associated)	2019 (no publication associated)	2019 (Hazzouri et al., 2019)
Sex	Female	Female	Female	Female	Male
Cultivar	Khalas	Khalas	Khalas	Khanezi	Barhee
Total sequence length (Kb)	381 563	556 480	854 663	454 367	772 474
Total ungapped length (Kb)	366 606	507 544	819 260	414 902	772 442
Gaps between scaffolds	0	0	706	0	0
Number of scaffolds	57 277	80 317	906	252 335	2 390
Scaffold N50	30 480	335 289	1 449 885	4881	4 728 343
Scaffold L50	3179	325	161	20 980	19
Number of contigs	142 304	143 382	11 170	840 310	2707
Contig N50	6460	10 936	114 381	1255	897 195
Contig L50	14 087	12 229	1591	70 619	196
Total number of chromosomes and plasmids	0	2	18	0	19
Number of sequences (WGS or clone)	142 304	143 382	11 170	252 335	2390
Sequencing technology	Illumina GAIIx	Roche 454; SOLiD; ABI3730	PacBio RSII; Illumina HiSeq; 10X Genomics	Illumina HiSeq	PacBio RSII; Illumina HiSeq

[a]National Center for Biotechnology Information accession number.

genes and 51 395 gene models (including isoforms), with a high-confidence set (i.e. only genes containing a Pfam domain) consisting of 43 815 isoforms encoded by 28 595 genes. Median gene size in the high-confidence set was ~4.2 Kb (Hazzouri et al., 2019).

An important feature of the BC4 male assembly was the high contiguity of this genome assembly, with a contig N50 of ~897 Kb (Table 2). Globally this assembly (NCBI accession number NBZB01, Table 2) represented an increase in assembly size of ~18%, a ~2.7-fold increase in N50, and ~20-fold fewer assembly fragments compared to the previous Cornell-Qatar and Saudi/China draft genome assemblies (Table 2). In addition, the NYUAD/KCGEB/UoA assembly allowed phasing of ~71% of the genome in haploid contigs (haplotigs) with an N50 of ~70.9 Kb. The high contiguity of this draft assembly allowed investigators for the first time to undertake genome-wide association studies (GWAS) to map genes for sex, fruit colour, and sugar composition (Hazzouri et al., 2019) (see below). The Date Palm Genome Hub website (https://datepalmgenomehub.abudhabi.nyu.edu) hosts an interactive genome browser and access to additional tools and downloadable resources related to the BC4 male assembly.

Finally, also in 2019, a draft genome assembly was made available by the University of Nizwa in Oman for the female variety Khanizi (= Khenizi) [NCBI accession number PEFZ01, Table 2]. The details of the assembly are as yet unpublished, although the NCBI data reports a low scaffold N50 (~4.9 Kb).

The three female and BC4 male draft genome assemblies are invaluable resources for the date palm and plant genomics research communities. Nevertheless, despite these efforts, the date palm genome requires significant improvement to reach chromosome-level sequence assemblies achieved in oil palm (Singh et al., 2013) and other fruit crops (Daccord et al., 2017). One of the challenges to assembly is the high degree of heterozygosity of date palms (Hazzouri et al., 2015). A similar problem in the early drafts of the apple genome was overcome through generation and sequencing of a double haploid (Daccord et al., 2017). While date palm might also benefit from a similar effort, application of additional sequencing technologies to existing draft assemblies should also lead to clear improvements in reference genome sequence quality. Recent reductions in the cost of long-read sequencing (Amarasinghe et al., 2020), linked-read sequencing (Wang et al., 2019), Hi-C (Ghurye and Pop, 2019), and optical mapping (Levy-Sakin and Ebenstein, 2013; Shelton et al., 2015; You et al., 2018) approaches, opens the possibility of further improvement of existing date palm genome assemblies. Development of more and better genome sequence for the cultivated *P. dactylifera* as well as other members of the genus *Phoenix* will allow for more in-depth comparative and evolutionary genomic analyses in this key palm group.

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

3.2 Date Palm Genetic Map

A high-resolution genetic map is a powerful tool for successful breeding programs, since it allows for marker-assisted selection and quantitative trait locus (QTL) analysis; it is also helpful in several types of population genomic analyses that need a recombination map (An et al., 2019; Apuli et al., 2020; Bai et al., 2018; Ma et al., 2019; O'Leary et al., 2018; Spence and Song, 2019; Zhang et al., 2017). The first genetic map of date palm was developed in 2014 (Mathew et al., 2014) using a mapping population of 85 F1 siblings from a Khalas × unknown male cross. The small size of the mapping population limits the number of recombination events that can be detected but appears sufficient for coarse-grained mapping. The map utilized ~4000 genotype-by-sequencing (GBS) SNP markers and 1999 framework markers spanning a total of 1293 cM. With this map, around 19% of the draft Cornell-Qatar genome ACYX02 scaffolds were placed in LGs for the first time. Assuming a genome size of 890 Mb, the Cornell-Qatar genome has an average genome-wide recombination distance of 0.69 Mb/cM (or 1.45 cM/Mb).

This date palm genetic map has 18 LGs, and there is consistency between the centimorgan lengths of LGs and the physical length of chromosomes measured by microscopy (Al-Salih and Al-Rawi, 1987). The genetic map has evenly distributed inter-markers, although it still has large gaps in LG5, 6, 9, and 10. In the case of LG5 and LG10, both have several gaps greater the 3 cM, reaching as high as 9 and 7 cM respectively, while LG6 has a large genetic marker gap of 9.7 cM (Mathew et al., 2014). Long regions in the map without markers are in part a consequence of the non-detection of heterozygous sites by the GBS approach used in the study. Despite the relatively low density of genetic markers in some regions of the map, on average there is a unique marker every ~210 Kb of the genome sequence, considering these markers are derived from 1823 unique sequence scaffolds of the Cornell-Qatar draft genome (Mathew et al., 2014).

3.3 Genetic Markers: RFLP, RAPD, AFLP, SSR, and SNPs

Molecular markers were the first genomic resources developed in date palm by numerous groups, initially used for genome fingerprinting, and measures of genetic diversity and relatedness. These included early work with restriction fragment length polymorphism (RFLP) (Corniquel and Mercier, 1994, 1997; Khanam et al., 2012; Sakka et al., 2003; Samuelson and Larsson, 1994), random amplified polymorphic deoxyribonucleic acid (RAPD) (Al-Khalifah and Shanavaskhan, 2017; Arunachalam, 2012; Corniquel and Mercier, 1994, 1997; González-Pérez et al., 2004; Saker et al., 2006; Samuelson and Larsson, 1994), amplified fragment length polymorphism (AFLP) (El-Assar et al., 2005; Sabir et al., 2014a; Saker et al., 2006) and

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

simple sequence repeat/microsatellite (SSR) markers (Akkak et al., 2009; Arabnezhad et al., 2012; Billotte et al., 2004; Elmeer and Mattat, 2015; Elshibli and Korpelainen, 2008; Hamza et al., 2013; Johnson, et al., 2009; Pintaud et al., 2010; Zehdi-Azouzi et al., 2015; Zhao et al., 2012), which provided some of the first estimates of genome-based genetic diversity (Cullis, 2011; Elshibli and Korpelainen, 2011; Zehdi-Azouzi et al., 2011). For example, the determination that Middle Eastern and North African date palms are genetically differentiated came about from studies using SSR markers (Arabnezhad et al., 2012; Zehdi-Azouzi et al., 2015). These genetic markers have also been utilized to provide molecular tags for key agronomic traits, most notably sex (Awan et al., 2017; Cherif et al., 2013, 2016; Elmeer and Mattat, 2012; Maryam et al., 2016). Because of the great impact of molecular marker technology on date palm biotechnology and biodiversity management, these studies have been extensively reviewed elsewhere (Al-Khayri et al., 2015a,b; Bekheet and Hanafy, 2011; Cullis, 2011; Elshibli and Korpelainen, 2011; Zehdi-Azouzi et al., 2011).

The rise of high-throughput short-read sequencing technology has enabled the identification of SNPs across the date palm genome. For example, GBS on 70 date palm varieties produced 67 496 SNPs for population genetic analysis (Mathew et al., 2015). Whole-genome re-sequencing studies on 62 (Hazzouri et al., 2015) and 157 (Hazzouri et al., 2019) varieties identified ~seven million SNPs or ~nine SNPs per Kb. These SNP data provided greater resolution in population genetic analyses (Hazzouri et al., 2015; Mathew et al., 2015), and facilitated genome-wide association studies (Hazzouri et al., 2019) [see below].

Both SSR and SNP markers have been collected in an online Date Palm Genomic Resource Database (DRDB) (He et al., 2017), which catalogues 246 445 SSR markers and 6 375 806 SNPs that were obtained from both previously published draft genome assemblies (Al-Mssallem et al., 2013) as well as whole-genome re-sequencing analysis (Hazzouri et al., 2015). As indicated above, a Date Palm Genome Hub that is a repository for reference genome information is also available (Hazzouri et al., 2019). Both the Date Palm Genome Hub and the DRDB database constitutes a public resource available to breeders and scientists to analyse genetic variation among date palm cultivars and incorporates bioinformatic tools to mine molecular marker data.

4 Genome-enabled Studies: Evolution of Date Palms

4.1 Population Genomics and the Evolution of Domesticated Date Palms

The availability of whole-genome sequences for date palms provides the basis for using genome-wide data to examine the population genomics of

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

this crop species. The first comprehensive analysis of date palm variation used whole-genome re-sequencing data on 62 varieties from around the world to develop the first genome-wide SNP map for this tree species (Hazzouri et al., 2015). Estimates of nucleotide diversity indicate that it is higher in date palms compared to other domesticated fruit tree species (Hazzouri et al., 2015), with Watterson's theta (θ_W) and nucleotide diversity (π) to be 0.010 and 0.009, respectively. Of the ~7 million SNPs catalogued in this set of varieties, ~5.2 million SNPs (or ~73%) are located in intergenic regions. SNPs in genic regions (excluding TEs) include ~201 000 synonymous and ~235 000 nonsynonymous polymorphisms, and ~1.48 million SNPs are in introns. The site-frequency spectrum indicates that both nonsynonymous and nonsense polymorphisms are skewed towards lower frequencies compared to synonymous and intron SNPs, suggesting the former are enriched for slightly deleterious mutations (Hazzouri et al., 2015).

Linkage disequilibrium (LD) decays relatively rapidly in date palms; the r^2 between SNPs reaches ~50% of its maximum at about 6 Kb and 90% of its maximum at about 40 Kb (Hazzouri et al., 2015). It should be noted that another study using a larger set of mostly Middle Eastern cultivars suggest slightly stronger LD, decaying to ~50% of its maximum at about 40 Kb (Hazzouri et al., 2019) [see below]. This relatively rapid decay of LD suggests that GWASs should enable high-resolution mapping of genes associated with traits of agricultural significance (see below).

The initial use of GBS data on 70 date palm varieties demonstrated that Middle Eastern and North African date palms were genetically differentiated (Mathew et al., 2015), confirming similar conclusions from previous studies using microsatellite loci (Arabnezhad et al., 2012; Zehdi-Azouzi et al., 2015). This was further explored using whole-genome re-sequencing data, which used STRUCTURE analysis (Hazzouri et al., 2015) to show that Middle Eastern and South Asian cultivars of date palms form a group, while North African varieties have a genomic composition separate from that of varieties from the Middle East. Samples from Egypt and Sudan show admixture between North African and the Middle Eastern populations, with 55–65% of their genomes being derived from the Middle Eastern population in the STRUCTURE analysis, as well as located in an intermediate position on SNP principal component analysis charts and neighbour-joining trees (Hazzouri et al., 2015). Interestingly, the North African date palm population has ~33% higher levels of nucleotide diversity ($\pi = 0.0108$) compared with the Middle Eastern/South Asian population ($\pi = 0.0081$). This suggests that North African varieties have a larger effective size (N_e) than the Middle Eastern population.

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

4.2 Introgressive Hybridization and the Origin of North African Dates

Hybridization between species is an evolutionary process that facilitates the integration of alleles from one genetic background into another (Harrison and Larson, 2014). This process results in the genomes of separate species becoming partially fused to generate mosaic genomes through the incorporation of alleles from one species into the gene pool of another (Runemark et al., 2019). Evidence of this process of introgressive hybridization in date palm was recently reported in a genomic study of North African date palm (Flowers et al., 2019) that reported introgression of alleles from wild *P. theophrasti*, or a *theophrasti*-like population, into Middle East *P. dactylifera* cultivars. *P. theophrasti* is a distinct but closely related species to domesticated date palm found primarily on the island of Crete in the eastern Mediterranean, with some individuals found in other Aegean islands and southern Turkey.

At the genome level, 1281 introgressed segments of ≥10 Kb, which spans 24.6 Mb of the nuclear genome, were identified in North African date palms as arising from *P. theophrasti* (Flowers et al., 2019). The introgressed regions are located throughout the genome, but no *P. theophrasti* genomic segment is fixed in North African dates. Based on nucleotide divergence, the North African population is less diverged from *P. theophrasti* than the Middle Eastern population, and genome regions with a higher introgression fraction are negatively correlated with population subdivision (Flowers et al., 2019). Globally, the results support the idea that an interspecific hybridization event occurred that gave rise to the North African date palms and introduced new genetic material into *P. dactylifera*.

4.3 The Wild Date Palms of Oman

In 2017, a study reported evidence of a distinct and ancient lineage of *P. dactylifera* that persists as a wild population in the Hajar mountains of Oman (Gros-Balthazard et al., 2017). Microsatellite analysis found significantly greater allelic diversity in this population compared with cultivated individuals from the Middle East, that these wild individuals formed a genetically unique cluster in both admixture analysis and principal component analysis (PCA), and that they are ancestral to the Middle Eastern clade (Gros-Balthazard et al., 2017). Whole-genome re-sequencing analysis confirmed that the Omani wild date palms formed a distinct genetic population compared to date palms from the Middle East and North Africa (Flowers et al., 2019; Hazzouri et al., 2015). Together, these results suggest that this is not a feral population whose ancestry traces to domesticated date palms, but instead represent an ancient and relictual lineage of wild *P. dactylifera* that may provide clues to the origins of domesticated date palms.

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

5 Genome-enabled Studies: Date Palm Genes

5.1 Sex-determining Region

Date palms are dioecious with separate male and female individuals, and there has been concerted effort to try to identify the date palm sex-determination genes. Identifying these loci can help establish the mechanism for sex determination in this perennial fruit tree crop and also provide molecular markers that can differentiate male vs. female individuals at the seedling or juvenile phases. The latter is critical since only females are fruit-bearing, and thus identifying females early in the life cycle ensures maximal productivity of farms.

Date palms appear to have an XY sex-determination system (AbdAlla and Abd El-Kawy, 2010; Al-Dous et al., 2011; Siljak-Yakovlev et al., 1996). Male-linked molecular markers were identified in 2012 (Al-Mahmoud et al., 2012), but these markers were validated only in a small sample with an accuracy of 90%. Later, Cherif et al. (2013), using a sample of 52 males and 55 female genotypes identified three genetically linked SSR loci – mPdIRD50, mPdIRD52, and mPdIRD80 – that are heterozygous only in male individuals. These male-specific alleles allowed the sex identification in all tested individuals, and the low diversity observed in what was defined as Y haplotypes was consistent with paternal transmission of a nonrecombining male-determining region. These results provided further evidence of an XY chromosome system, with males being the heterogametic sex in date palm (Al-Dous et al., 2011; Cherif et al., 2013).

The three sex-linked SSR loci identified in date palm were also analysed in eight additional species from the genus *Phoenix* (Cherif et al., 2016). All 59 males analysed were heterozygous (X_{allele}/Y_{allele}) for the three microsatellite loci. The Y-linked alleles could be assigned to 11 Y haplotypes across the other *Phoenix* species and, more importantly, none were found in X haplotypes. These findings indicate that recombination does not seem to occur between Y and X mPdIRDP80, mPdIRDP50, and mPdIRDP52 loci in any of the studied *Phoenix species*, which strongly suggest a common sex-linked nonrecombining region in the genus (Cherif et al., 2016).

In order to verify that the conservation of the sex-linked SSR across *Phoenix* species was not due to homoplasy, the mPdIRDP80 locus neighbourhood was further investigated and found to contain a MYB gene previously reported in an analysis of the draft genome sequence (Al-Dous et al., 2011) (see below). The gene was named *PdMYB1* and the mPdIRDP80 microsatellite is localized within the gene promoter sequence 176 bps upstream of the transcription start site. After cloning *PdMYB1* in seven related *Phoenix* species, the authors identified X/Y pairs based on sizes of the mPdIRDP80 X and Y alleles (Cherif et al., 2016). Phylogenetic reconstruction of *PdMYB1* found that distinct Y and X alleles clustered separately instead of by species

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

(Cherif et al., 2016). This finding is correlated with the allelic distribution of mPdIRDP80, mPdIRDP50, and mPdIRDP52 SSRs loci and together suggest that they are all contained in an XY non-recombing region that had diverged in the ancestor of these *Phoenix* species (Al-Dous et al., 2011; Al-Mahmoud et al., 2012; Cherif et al., 2013). Cytological studies suggest that the date palm Y chromosome is smaller than the X (AbdAlla and Abd El-Kawy, 2010). Together, these results provide evidence of a smaller and partially degenerated Y chromosome in date palms, and are consistent with the idea of an ancient origin of dioecy in the genus *Phoenix* (Cherif et al., 2013, 2016).

The availability of draft genome assemblies provided the ability to identify sex-linked regions in the date palm genome. In the first genome sequence released, analysis identified SNPs that segregate with sex and also pointed to an XY sex-determination model with males as the heterogametic sex (Al-Dous et al., 2011). Predicted genes in these regions include (i) the MYB transcription factor gene, now referred to as *PdMYB1* discussed above, (ii) a homolog of yeast *rcd-1*, a gene that encodes for a differentiation control factor essential for the onset of sexual development, and (iii) a gene encoding a putative prenyltransferase of the rab geranylgeranyl transferase family (Al-Dous et al., 2011). Interestingly, *rcd-1* is crucial in yeast sexual development (Okazaki et al., 1998) and interacts with a c-MYB protein (Haas et al., 2004). Moreover, prenylation, which is catalysed by prenyltransferases, is fundamental for the correct activation of MADS box genes which orchestrate floral initiation (Kaufmann et al., 2010) and control flower development (Yalovsky et al., 2000). Altogether the described genes may have an important role in cell differentiation control in date palm floral development, although further investigation is needed to support this functional hypothesis.

To further identify sex-determining genes, a k-mer approach using whole-genome short-read sequencing data was utilized to find male-specific loci found across multiple *Phoenix* species (Torres et al., 2018). This study found that the divergence between X and Y extends over ~13 Mb, similar in length to sex-determination regions in other angiosperms such as papaya (Wang et al., 2012; Zerpa-Catanho and Ming, 2019) or poplars (Geraldes et al., 2015; Hou et al., 2015). This approach identified at least three potential genes suggested to be involved in sex determination in *Phoenix*: the cytochrome P450 gene *CYP703, GPAT3* which is a glycerol-3-phosphate acyltransferase 6-like gene expressed only in male flowers, and the *Lonely-Guy* (*LOG*) gene that shows higher expression in male vs. female flowers (Torres et al., 2018).

Differential expression of the *LOG* gene between male and female flowers suggests a role in suppressing female flower formation in males, although more studies are needed to understand its function in date palm. The other two genes, *CYP703* and *GPAT3*, have functions in pollen formation and another development, mainly because they are crucial in lipid synthesis

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

pathways (Aya et al., 2009; Men et al., 2017; Nelson et al., 2004; Somaratne et al., 2017; Yang et al., 2014). *CYP703* is a single member gene subfamily in land plants (Nelson et al., 2004), which suggests that it encodes a unique function, and knockouts of *CYP703* homologs in Arabidopsis (Morant et al., 2007), rice (Yang et al., 2014), and maize (Somaratne et al., 2017) all resulted in male sterility with no other visible effect on vegetative development. Deletion of *GPAT3* in rice also resulted in male sterility without affecting normal growth (Men et al., 2017). Accordingly, this suggests that *CYP703* may not have a paralog that could compensate for its deletion and that a single copy of the Y-linked *GPAT3* gene in *Phoenix* can sustain normal male flower development (Torres et al., 2018).

These studies suggest that the ancestor of the genus *Phoenix* may have evolved from hermaphroditism to dioecy through a gynodioecious intermediate, which occurred by loss of male function, either involving or following deletion of *CYP703* and/or *GPAT3* from a proto-X chromosome, and leaving both genes intact on the proto-Y chromosome (Torres et al., 2018). The model has three main steps: (i) deletion of *GPAT3* and/or *CYP703* in the proto-X chromosome, (ii) an inversion of the corresponding region in the proto-Y chromosome which led to Y chromosome recombination arrest, followed by (iii) a duplication and translocation of the *LOG* gene into the region creating the final Y chromosome, which suppresses the female flower development in male individuals (Figure 1). This model requires further investigation in order to verify whether the *LOG* gene does function in female flower development in *Phoenix* species and in date palm specifically. It will also be interesting to examine when the duplication/translocation evolutionary step that led to this sex-determination system occur. In addition, a fourth gene encoding for a cytidine deaminase was found to be present at the sex-determining region (Torres et al., 2018). Further study is needed to determine whether cytidine deaminase has a function in sex determination, or if its genome localization at the boundary of the sex-determining region is just a consequence of the same duplication/inversion event that brought *GPAT3* to its current location of the Y chromosome.

5.2 Fruit Colour in Date Palms

Date palm fruit come generally in red, yellow, and a range of intermediate colours that become apparent at advanced stages of fruit development. The *VIRESCENS* (*VIR)* gene was first isolated in the oil palm, *E. guineensis*, which also has fruits that are either violet/black (nigrescens) or orange (virescens) in colour (Singh et al., 2014; Ying et al., 2007). Molecular characterization of the oil palm *VIR* gene showed that orange-coloured fruits were determined by naturally occurring nonsense alleles at this locus, which encodes an R2R3-myb transcription factor (Singh et al., 2014). The molecular genetic

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

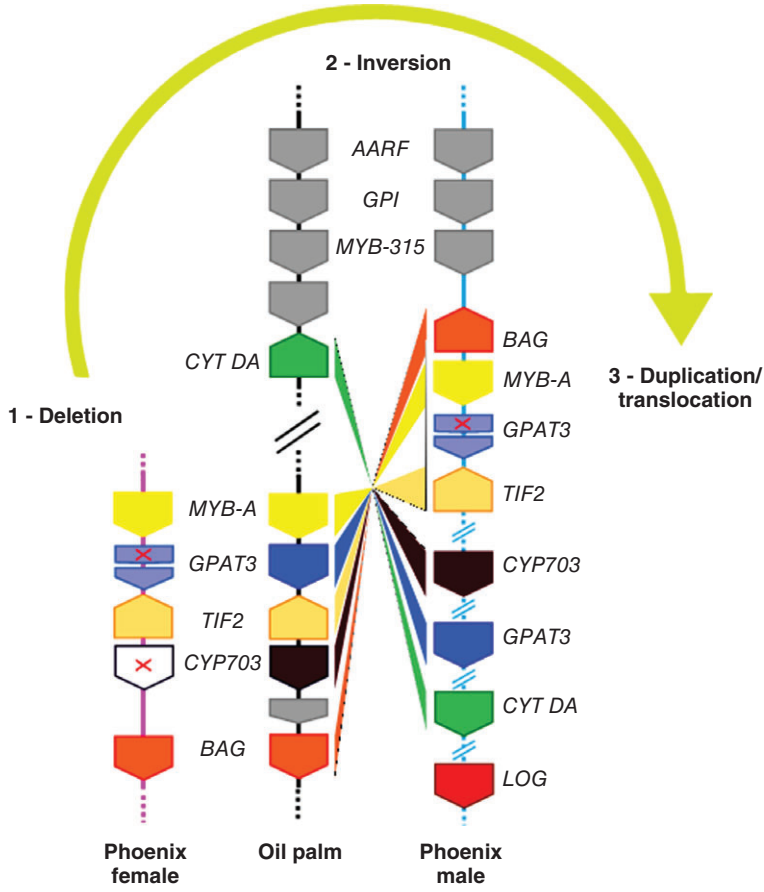

Figure 1 Proposed model for dioecy in *Phoenix*. (Step 1) A deletion of *GPAT*3 and/or *CYP703* (red arrows) occurred leading to gynodioecy. (Step 2) Gene inversion of different genes (coloured triangles and pentagons) would create a proto-Y chromosome with recombination arrest which was followed by (step 3) duplication and translocation events into the adjacent region of the *LOG* gene Source: Adapted from Torres et al. (2018).

basis of fruit colour was first examined in date palm in 2015, when a candidate gene analysis revealed that ortholog of oil palm *VIR* also controls colour variation in date palm and that varieties with yellow fruits also contained a *copia*-like retrotransposon insertion in exon 3 of the gene (Hazzouri et al., 2015).

The role of date palm *VIR* was further confirmed through a genome-wide association study, which examined the genetic architecture of 21 fruit traits, including fruit colour in a mapping population of 142 female varieties (Hazzouri et al., 2019). The LD in this mapping population decays fairly

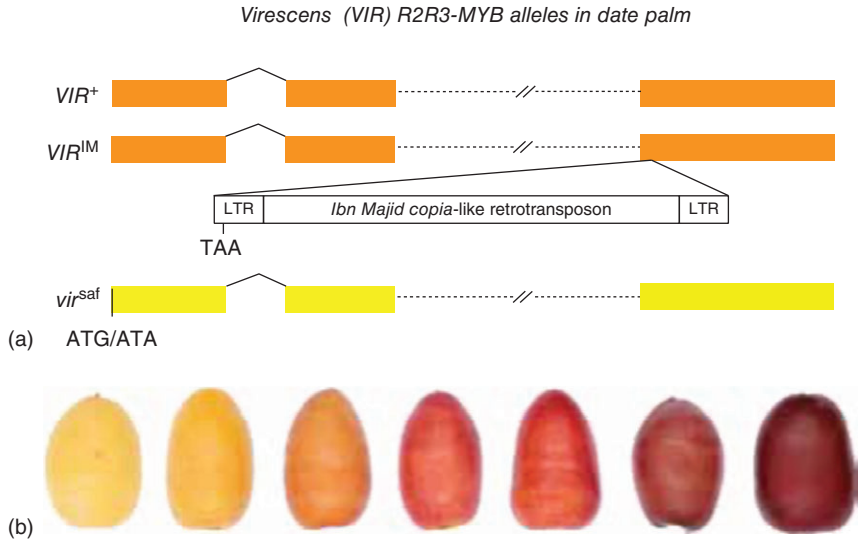

Figure 2 The date palm *VIRESCENS* gene. Allele genomic structure of date palm *VIR* gene (a) responsible for dates colour palette (b). The *VIR* gene has three exons (boxes) and three alleles being VIR^{+} the wild-type allele. The VIR^{IM} allele has the *Ibn Majid* retrotransposon insertion which introduces an earlier stop codon (TAA) in exon 3. A polymorphism at the start codon (AT**G**/AT**A**) is represented for VIR^{saf} allele. Source: Adapted from Hazzouri et al. (2019).

quickly, achieving a half-maximum level at ~22.9 Kb, indicating that GWAS mapping could have fairly high resolution. GWAS analysis revealed a QTL for fruit colour at LG4 in a ~20 Kb region that included the *VIR* gene, and together with the long-read BC4 genome assembly, the full-length date palm *VIR* gene was identified and designated as the VIR^{+} allele (Figure 2). VIR^{+} homozygotes produce relatively high amounts of fruit anthocyanin and have a predominantly red phenotype (Hazzouri et al., 2019).

The analysis also characterized the *VIR* allele with a complete copy of a *copia*-like retrotransposon insertion in the coding region. The low copy number retrotransposon in this allele is ~11.7 Kb in length, has 469-bp long terminal repeats, has a 5-bp target site duplication at its insertion site, and was named *Ibn Majid (IM)* in tribute of a medieval Arabic navigator. The allele with the *IM* insertion is designated as VIR^{IM} and both homozygotes of this allele as well as heterozygous VIR^{+}/VIR^{IM} plants produce trace amounts of anthocyanin and are yellow in colour (Hazzouri et al., 2019).

A third allele at *VIR* was observed when it was shown that some varieties did not have the retrotransposon insertion allele but nevertheless produced yellow fruit (Hazzouri et al., 2019). Analysis of the *VIR* gene in these varieties

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

revealed that the gene carried a SNP mutation that abolished that translation initiation site, and alleles with this polymorphism are designated as vir^{saf}. The red colour observed in VIR^{+}/vir^{saf} fruits and the yellow fruit colour of vir^{saf} homozygotes implies that VIR^{+} is a functional wild-type allele and is haplosufficient, while vir^{saf} is recessive and is a loss-of-function allele. In contrast, the yellow or orange colour of VIR^{+}/VIR^{IM} genotypes suggests that VIR^{IM} is dominant or semi-dominant, and a negative inhibitor of VIR^{+} allele expression (Hazzouri et al., 2019).

5.3 Sugar Composition in Date Palm Fruits

Sugar composition is an important trait of date fruits and contributes in part to the distinctive flavour profile of individual varieties. The relative concentration of sugars is used to classify date palm varieties as either sucrose- or reducing-sugar (fructose, glucose) type, the latter characterized by relatively high levels of glucose and fructose compared to sucrose (Hazzouri et al., 2019). Analysis of sugar composition in fully ripened fruit found that the concentrations of these three sugars are strongly correlated, and GWAS mapped a major sugar composition QTL found on LG14 spanning about ~1.1 Mb containing 162 annotated protein-coding genes (Hazzouri et al., 2019).

The GWAS study identified three genes in this genomic region encoding β-fructofuranosidase (invertase) enzymes, which have been shown to be responsible in part for fruit sugar accumulation in a range of plant species (Ruan et al., 2010). These genes include an alkaline/neutral invertase (*A/N-INV1*; gene ID chr14G0028200, genome assembly NBZB01) located at the centre, and two cell wall invertase genes (*CWINV1* and *CWINV3*, gene IDs chr14G0022900 and chr13G0023100, respectively) located at the 5′ region of the GWAS peak (Figure 3). *CWINV1* and *CWINV3* seem to be tandem duplicates of each other with 98.58% amino acid identity. In addition, an unannotated invertase gene (*CWINV2)* was also identified, but the sequence has multiple frame-shit mutations in some varieties and was considered a pseudogene (Figure 3) (Hazzouri et al., 2019).

Examination of the genomic region spanning the sugar GWAS peak found a deletion of ~40 Kb that includes the promoter and 2 exons of *CWINV1*, the *CWINV2* pseudogene and a region ~5 Kb downstream *CWINV3* (Hazzouri et al., 2019; Malek et al., 2020). The presence of this 40-Kb deletion is associated with reduced transcript levels of *CWINV1* and *CWINV3* and preferentially occurs in sucrose type varieties, where the frequency of a homozygous deletion is higher compared with reducing-sugar-type varieties (Hazzouri et al., 2019). In addition, another ~5-Kb deletion was found just upstream the *A/N-INV1* gene, which also appeared to be found more often in sucrose-type varieties (Hazzouri et al., 2019).

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

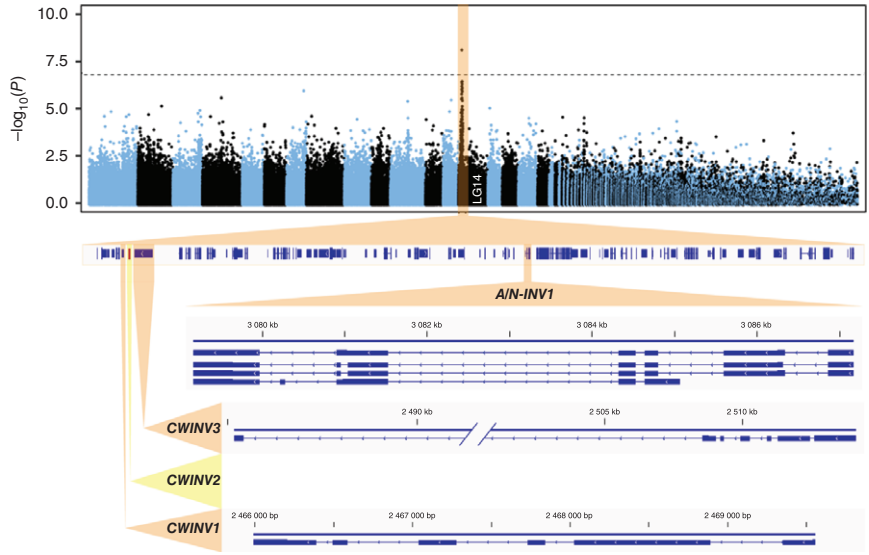

Figure 3 GWAS mapping of sugar composition in date palm fruit. Close-up for gene models located at the GWAS peak is shown (orange rectangle at LG14). *CWINV1*, *CWINV2,* and *CWINV3* appear as tandem duplicated genes. *CWINV1* and *CWINV2* have 98.58% identity in protein primary sequence. *CWINV2* is unannotated in the BC4 male genome (NCBI accession number NBZB01) and was considered a pseudogene. *A/N-INV1* is annotated with four splicing variants. Dotted lines represent the Bonferroni significance threshold in the GWAS analysis. Source: Adapted from Hazzouri et al. (2019).

6 Date Palm Organellar Genomes

6.1 Mitochondrial Genome

The date palm mitochondrial genome is one of the largest angiosperm mitochondrial genomes, comprising 715 001 bps with an average GC content of 45.1% (Fang et al., 2012). Only ~6.5% of the genome are coding sequences, with 38 protein-coding genes that encode, among others, for nine subunits of nicotinamide adenine dinucleotide dehydrogenase (complex I), apocytochrome b (complex III), three subunits of cytochrome c oxidase (complex IV), five subunits of ATP synthase F1 (complex V), and four cytoplasmic membrane proteins required for cytochrome c maturation. The genetic complement of the date palm mitochondrial genome also includes 30 tRNAs and 3 ribosomal RNAs. Interestingly, ~10% of the mitochondrial genome is populated with chloroplast derived DNA, where several typical chloroplast genes such as *petA*, *petG*, *petL*, *psaJ*, *psbT*, *rpl20*, *rpl33*, and *rps8* have been identified (Fang et al., 2012). This constitutes one of the highest proportions

of chloroplast-derived sequences in mitochondria observed in plants (Fang et al., 2012; Gandini and Sanchez-Puerta, 2017).

Indeed, the date palm mitochondrial genome is highly diverged from other angiosperms, sharing only ~21% of mitochondrial genome over 70% identity with *Vitis*, ~15% with *Oryza*, and ~11% with *Zea* (Clifton et al., 2004; Fang et al., 2012; Goremykin et al., 2009; Notsu et al., 2002; Tian et al., 2006). A phylogenetic tree inferred from 22 genes common to 15 plant mitochondrial genomes indicates that *P. dactylifera* is rooted at the basal position of the sampled monocots (Fang et al., 2012). In addition, the date palm mitochondrial genome appeared to have less repetitive sequences when compared with other angiosperms, with long repeats only contributing 2.3% of the genome, and a similarly low level of tandem repeats (Fang et al., 2012). The low repeat content is unusual and an intriguing feature of the date palm mitochondrial genome.

6.2 Chloroplast Genome

The chloroplast genome sequence has been reported in the Khalas variety of date palm from Saudi Arabia (Al-Mssallem et al., 2013; Yang et al., 2010), the Pakistani Assel (Khan et al., 2012), and in the two Omani varieties Naghal and Khanezi (Khan et al., 2018), and the recently sequenced BC4 male (Hazzouri et al., 2019). Genome size varies from 158 210 bp in Naghal to 158 462 bp in Khalas, which represents a variation in size of ~0.16% (Khan et al., 2018). The original Khalas sequence is AT-rich (62.77%), has a pair of inverted repeat sequences (IR) comprising 27 276 bps, separated by a long single-copy sequence (LSC) at 86 198 bps, and a short single-copy sequence (SSC) at 17 712 bps (Yang et al., 2010). The genome encodes 131 predicted genes, of which 112 are unique and 19 are duplicated in the IR regions. Among the unique genes, there are 79 protein-coding, 29 tRNA and 4 rRNA genes, respectively. Outside of the coding regions, 41.57% of the genome are non-coding and contain introns, intergenic spacers, and pseudogenes. Of the 112 unique genes in the date palm chloroplast genome, 18 possess introns (Khan et al., 2012; Khan et al., 2018; Yang et al., 2010).

The date palm genome seems to evolve slowly, with the rate of nucleotide substitutions approximately eightfold lower than the rates observed in annual plants (Yang et al., 2010). However, an analysis of whole-chloroplast genome sequences among four varieties showed comparatively low sequence identity (Khan et al., 2018). There are highly divergent non-coding sequences, including psbK-trnG, trnT-trnL, rbcL-accD, petA-psbJ, and psaC-ndhE gene spacers, which can serve as potential genetic markers. In addition, a number of chloroplast indels and SNPs (Table 3) have been identified and used in genetic diversity analysis, including markers that have demonstrated the difference between Middle East and North African *P. dactylifera* (Flowers et al., 2019; Gros-Balthazard et al., 2017; Mohamoud

Table 3 Number of Indels and SNP in chloroplast genomes from Khanizi an Naghal vs. Aseel and Khalas date palm varieties.

Khanizi			Naghal		
	INDEL	SNP		INDEL	SNP
Naghal	35	23	**Khanizi**	35	23
Aseel	293	18	**Aseel**	292	10
Khalas	299	16	**Khalas**	296	12

Source: Adapted from Khan et al. (2018)

et al., 2019; Moussouni et al., 2017). Recently, ~200 mitochondrial and chloroplast genomes from different date palm populations were analysed (Mohamoud et al., 2019), revealing four haplotypes associated with different cultivars. From these, three haplotypes predominate in the current date palm population worldwide, with a unique haplotype found mostly in Tunisia, Algeria, and Egypt, in agreement with a distinct haplotype found at high frequency in the organellar genomes of North African populations observed in a separate study (Flowers et al., 2019). Finally, one should note that there is evidence for heteroplasmy in date palms, suggesting that multiple chloroplast genotypes can exist in a single plant (Sabir et al., 2014b).

7 Summary and Perspectives

Genomic studies in date palms have accelerated over the last decade, spurring advances in identifying key genes and our understanding of genetic diversity in this fruit tree crop. Looking forward, advances in the field will focus on developing more genomic resources, including high-quality reference genomes for all members of the genus *Phoenix*. Studies of genomic diversity should also aim at a comprehensive sampling of the approximately 3000 varieties, coupled with information on phenotypes and environmental and microbiome associations to better link genome information with plant phenotypes.

Functional genomic information remains sparse, and examining transcriptomic and metabolomic diversity and how it changes with development and environment can help us better understand key aspects of date palm biology. Mapping of genes, especially those associated with critical features of date palm biology, will benefit from developing new genetic mapping populations, including those for linkage mapping and genome-wide association studies. Finally, translating this information into concerted breeding programs and genetic engineering interventions can help advance date palm agriculture, particularly in the face of challenges brought about by climate change as well as continued abiotic and biotic pressures. In all this,

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

the foundation of genomics of date palms has been secured over the last decade, setting the stage for research in the next few years.

Acknowledgements

This work was supported in part by grants from Ministry of Presidential Affairs (UAE) to KCGEB (K.A.) and the NYU Abu Dhabi Research Institute to M.D.P.

References

AbdAlla, M.M. and Abd El-Kawy, A.M. (2010). Karyotype analysis for date palm (*Phoenix dactylifera* L) compared with tissue culture derived plants. *New York Science Journal* **3** (**11**): 165–170.

Akkak, A., Scariot, V., Torello Marinoni, D. et al. (2009). Development and evaluation of microsatellite markers in *Phoenix dactylifera* L. and their transferability to other *Phoenix* species. *Biologia Plantarum* **53**: 164–166.

Al-Alawi, R., Al-Mashiqri, J.H., Al-Nadabi, J.S.M. et al. (2017). Date palm tree (*Phoenix dactylifera* L.): natural products and therapeutic options. *Frontiers in Plant Science* **23** (**8**): 845.

Al-Ani, B., Zaid, A., and Shabana, H. (2010). On the status of chromosomes of the date palm (*Phoenix dactylifera* L.). In: *Acta Horticulturae,* vol. 882. Leuven, Belgium: International Society for Horticultural Science (ISHS), pp. 253–268. doi: 10.17660/ActaHortic.2010.882.28

Al-Dous, E.K., George, B., Al-Mahmoud, M.E. et al. (2011). De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nature Biotechnology* **29**: 521–527.

Al-Khalifah, N.S. and Shanavaskhan, A.E. (2017). Molecular identification of date palm cultivars using random amplified polymorphic DNA (RAPD) markers. In: *Methods in Molecular Biology*, 1e, vol. **1638** (ed. J. Al-Khayri, S. Jain and D. Johnson). New York: Humana Press.

Al-Khayri, J.M., Jain, S.M., and Johnson, D.V. (2015a). *Date Palm Genetic Resources and Utilization: Volume 1: Africa and the Americas* (ed. J.M. Al-Khayri, S.M. Jain and D.V. Johnson). Springer Netherlands.

Al-Khayri, J.M., Jain, S.M., and Johnson, D.V. (2015b). *Date Palm Genetic Resources and Utilization: Volume 2: Asia and Europe* (ed. J.M. Al-Khayri, S.M. Jain and D.V. Johnson). Springer Netherlands.

Al-Mahmoud, M.E., Al-Dous, E.K., Al-Azwani, E.K., and Malek, J.A. (2012). DNA-based assays to distinguish date palm (Arecaceae) gender. *American Journal of Botany* **99** (**1**): e7–e10.

Al-Mssallem, I.S., Hu, S., Zhang, X. et al. (2013). Genome sequence of the date palm *Phoenix dactylifera* L. *Nature Communications* **4**: 2274.

Al-Salih, A.A. and Al-Rawi, A.M.A. (1987). A study of the cytology of two female cultivars of date palm. *Date Palm Journal (FAO/NENADATES)* **5**(**2**): 123–142.

Amarasinghe, S.L., Su, S., Dong, X. et al. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology* **21** (**30**): 1–16.

An, Z., Zhao, Y., Zhang, X. et al. (2019). A high-density genetic map and QTL mapping on growth and latex yield-related traits in *Hevea brasiliensis* Müll Arg. *Industrial Crops and Products* **132**: 440–448.

Apuli, R.P., Bernhardsson, C., Schiffthaler, B. et al. (2020). Inferring the genomic landscape of recombination rate variation in European Aspen (*Populus tremula*). *G3: Genes, Genomes, Genetics* **10** (**1**): 299–309.

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

Arunachalam, V. (ed.) (2012). Coconut. In: *Genomics of Cultivated Palms*, 2e. Elsevier, pp. 13–17. http://www.sciencedirect.com/science/article/pii/B9780123877369000029

Awan, F., Maryam, S., Jaskani, M.J., and Sadia, B. (2017, 1638). Gender identification in date palm using molecular markers. *Methods in Molecular Biology* 209–225.

Aya, K., Tanaka, M.U., Kondo, M. et al. (2009). Gibberellin modulates anther development in rice via the transcriptional regulation of GAMYB. *Plant Cell* **21** (**5**): 1453–1472.

Bai, B., Wang, L., Zhang, Y.J. et al. (2018). Developing genome-wide SNPs and constructing an ultrahigh-density linkage map in oil palm. *Scientific Reports* **8** (**1**): 1–7.

Baker, W.J. and Couvreur, T.L.P. (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.

Barrow, S.C. (1998). A monograph of *Phoenix* L. (Palmae: Coryphoideae). *Kew Bulletin* **53** (**3**): 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 (ECSSR)*. Abu Dhabi, Emirates Center for Strategic Studies and Research.

Beech, M. and Shepherd, E. (2001). Archaeobotanical evidence for early date consumption on Dalma Island, United Arab Emirates. *Antiquity* **75** (**287**): 83–89.

Bekheet, S.A. and Hanafy, M.S. (2011). Towards sex determination of date palm. In: *Date Palm Biotechnology* (ed. S. Jain, J. Al-Khayri and D. Johnson). Dordrecht: Springer.

Billotte, N., Marseillac, N., Brottier, P. et al. (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 Notes* **4** (**2**): 256–258.

Chaluvadi, S.R., Young, P., Thompson, K. et al. (2019). Phoenix phylogeny, and analysis of genetic variation in a diverse collection of date palm (*Phoenix dactylifera*) and related species. *Plant Diversity* **41** (**5**): 330–339.

Chao, C.C.T. and Krueger, R.R. (2007). The date palm (*Phoenix dactylifera* L.): overview of biology, uses, and cultivation. *HortScience* **42** (**5**): 1077–1082.

Cherif, E., Zehdi, S., Castillo, K. et al. (2013). Male-specific DNA markers provide genetic evidence of an XY chromosome system, a recombination arrest and allow the tracing of paternal lineages in date palm. *New Phytologist* **197** (**2**): 409–415.

Cherif, E., Zehdi-Azouzi, S., Crabos, A. et al. (2016). Evolution of sex chromosomes prior to speciation in the dioecious *Phoenix* species. *Journal of Evolutionary Biology* **29** (**8**): 1513–1522.

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

Clifton, S.W., Minx, P., Fauron, C.M.R. et al. (2004). Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiology* **136** (**3**): 3486–3503.
Corniquel, B. and Mercier, L. (1994). Date palm (*Phoenix dactylifera* L.) cultivar identification by RFLP and RAPD. *Plant Science* **101** (**2**): 163–172.
Corniquel, B. and Mercier, L. (1997). Identification of date palm (*Phoenix dactylifera* L.) cultivars by RLFP: partial characterization of a cDNA probe that contains a sequence encoding a zinc finger motif. *International Journal of Plant Sciences* **158** (**2**): 152–156.
Cullis, C. (2011). Molecular markers in date palm. In: *Date Palm Biotechnology* (ed. S.M. Jain, J.M. Al-Khayri and D.V. Johnson). Dordrecht: Springer Netherlands.
Daccord, N., Celton, J.M., Linsmith, G. et al. (2017). High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics* **49** (**7**): 1099–1106.
Dransfield, J., Uhl, N.W., Lange, C.B.A. et al. (2008). *Genera Palmarum: The Evolution and Classification of Palms*. Kew: Royal Botanic Gardens.
El-Assar, A.M., Krueger, R.R., Devanand, P.S., and Chao, C.C.T. (2005). Genetic analysis of Egyptian date (*Phoenix dactylifera* L.) accessions using AFLP markers. *Genetic Resources and Crop Evolution* **52**: 601–607.
Elmeer, K. and Mattat, I. (2012). Marker-assisted sex differentiation in date palm using simple sequence repeats. *3 Biotech* **2** (**3**): 241–247.
Elmeer, K. and Mattat, I. (2015). Genetic diversity of Qatari date palm using SSR markers. *Genetics and Molecular Research* **14** (**1**): 1624–1635.
Elshibli, S. and Korpelainen, H. (2008). Microsatellite markers reveal high genetic diversity in date palm (*Phoenix dactylifera* L.) germplasm from Sudan. *Genetica* **134** (**2**): 251–260.
Elshibli, S. and Korpelainen, H. (2011). Biodiversity in date palm: molecular markers as indicators. In: *Date Palm Biotechnology* (ed. S.M. Jain, J.M. Al-Khayri and D.V. Johnson). Dordrecht: Springer Netherlands.
Fang, Y., Wu, H., Zhang, T. et al. (2012). A complete sequence and transcriptomic analyses of date palm (*Phoenix dactylifera* L.) mitochondrial genome. *PLoS ONE* **7** (**5**): e37164.
Flowers, J.M., Hazzouri, K.M., Gros-Balthazard, M. et al. (2019). Cross-species hybridization and the origin of North African date palms. *Proceedings of the National Academy of Sciences of the United States of America* **116** (**5**): 1651–1658.
Food and Agriculture Organization of the United Nations. (2020). FAOSTAT Statistical Database. Rome: FAO. http://www.fao.org/faostat/en/#home
Gandini, C.L. and Sanchez-Puerta, M.V. (2017). Foreign plastid sequences in plant mitochondria are frequently acquired via mitochondrion-to-mitochondrion horizontal transfer. *Scientific Reports* **6** (**7**): 43402.
Geraldes, A., Hefer, C.A., Capron, A. et al. (2015). Recent Y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). *Molecular Ecology* **24** (**13**): 3243–3256.
Ghurye, J. and Pop, M. (2019). Modern technologies and algorithms for scaffolding assembled genomes. *PLoS Computational Biology* **15** (**6**): 1–20.
González-Pérez, M.A., Caujapé-Castells, J., and Sosa, P.A. (2004). Molecular evidence of hybridisation between the endemic *Phoenix canariensis* and the widespread *P. dactylifera* with random amplified polymorphic DNA (RAPD) markers. *Plant Systematics and Evolution* **247** (**3–4**): 165–175.

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

Goremykin, V.V., Salamini, F., Velasco, R., and Viola, R. (2009). Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. *Molecular Biology and Evolution* **26** (**1**): 99–110.

Gros-Balthazard, M. (2013). Hybridization in the genus *Phoenix*: a review. *Emirates Journal of Food and Agriculture* **25** (**11**): 831–842.

Gros-Balthazard, M., Galimberti, M., Kousathanas, A. et al. (2017). The discovery of wild date palms in Oman reveals a complex domestication history involving centers in the Middle East and Africa. *Current Biology* **27** (**14**): 2211–2218.e8.

Gros-Balthazard, M., Hazzouri, K.M., and Flowers, J.M. (2018). Genomic insights into date palm origins. *Genes* **9** (**10**): 502.

Gros-Balthazard, M., Battesti, V., Ivorra, S. et al. (2020). On the necessity of combining ethnobotany and genetics to assess agrobiodiversity and its evolution in crops: a case study on date palms (*Phoenix dactylifera* L.) in Siwa Oasis, Egypt. *Evolutionary Applications* **13**: 1818–1840.

Haas, M., Siegert, M., Schürmann, A. et al. (2004). c-Myb protein interacts with Rcd-1, a component of the CCR4 transcription mediator complex. *Biochemistry* **43** (**25**): 8152–8159.

Hamza, H., Ben Abederrahim, M.A., Elbekkay, M., and Ferchichi, A. (2013). Comparison of the effectiveness of ISSR and SSR markers in determination of date palm (*Phoenix dactylifera* L.) agronomic traits. *Australian Journal of Crop Science* **4** (**6**): 763–769.

Harrison, R.G. and Larson, E.L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity* **105** (**S1**): 795–809.

Hazzouri, K.M., Flowers, J.M., Visser, H.J. et al. (2015). Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. *Nature Communications* **6**: 8824.

Hazzouri, K.M., Gros-Balthazard, M., Flowers, J.M. et al. (2019). Genome-wide association mapping of date palm fruit traits. *Nature Communications* **10**: 4680.

Hazzouri, K.M., Flowers, J.M., Nelson, D. et al. (2020a). Prospects for the study and improvement of abiotic stress tolerance in date palms in the post-genomics era. *Frontiers in Plant Science* **11**: 293.

Hazzouri, K.M., Sudalaimuthuasari, N., Kundu, B. et al. (2020b). The genome of pest *Rhynchophorus ferrugineus* reveals gene families important at the plant-beetle interface. *Communications Biology* **3** (**1**): 323.

He, Z., Zhang, C., Liu, W. et al. (2017). DRDB: an online date palm genomic resource database. *Frontiers in Plant Science* **8**: 1889.

Henderson, S.A., Billotte, N., and Pintaud, J.C. (2006). Genetic isolation of Cape Verde Island Phoenix atlantica (Arecaceae) revealed by microsatellite markers. *Conservation Genetics* **7**: 213–223.

Hou, J., Ye, N., Zhang, D. et al. (2015). Different autosomes evolved into sex chromosomes in the sister genera of *Salix* and *Populus*. *Scientific Reports* **5**: 9076.

Jaradat, A.A. (2015). Biodiversity, genetic diversity, and genetic resources of date palm. In: *Date Palm Genetic Resources and Utilization: Volume 1: Africa and the Americas* (ed. J.M. Al-Khayri, S.M. Jain and D.V. Johnson). Dordrecht: Springer Netherlands.

Johnson, C., Cullis, T.A., Cullis, M.A., and Cullis, C.A. (2009). DNA markers for variety identification in date palm (*Phoenix dactylifera* L.). *Journal of Horticultural Science and Biotechnology* **84** (**6**): 591–594.

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

Johnson, D.V., Al-Khayri, J.M., and Jain, S.M. (2015). Introduction: date production status and prospects in Africa and the Americas. In: *Date Palm Genetic Resources and Utilization: Volume 1: Africa and the Americas* (ed. J.M. Al-Khayri, S.M. Jain and D.V. Johnson). Dordrecht: Springer Netherlands.

Kaufmann, K., Wellmer, F., Muiñ, J.M. et al. (2010). Orchestration of floral initiation by APETALA1. *Science* **328** (**5974**): 85–89.

Khan, A., Khan, I.A., Heinze, B., and Azim, M.K. (2012). The chloroplast genome sequence of date palm (*Phoenix dactylifera* L. cv. 'Aseel'). *Plant Molecular Biology Reporter* **30**: 666–678.

Khan, A.L., Asaf, S., Lee, I.J. et al. (2018). First chloroplast genomics study of *Phoenix dactylifera* (var. Naghal and Khanezi): a comparative analysis. *PLoS ONE* **13** (**7**): e0200104.

Khanam, S., Sham, A., Bennetzen, J.L., and Aly, M.A.M. (2012). Analysis of molecular marker-based characterization and genetic variation in date palm (*Phoenix dactylifera* L.). *Australian Journal of Crop Science* **6** (**8**): 1236–1244.

Krueger, R.R. (2011). Date palm germplasm. In: *Date Palm Biotechnology* (ed. S.M. Jain, J.M. Al-Khayri and D.V. Johnson). Dordrecht: Springer Netherlands.

Krueger, R.R. (2015). Date palm status and perspective in the United States. In: *Date Palm Genetic Resources and Utilization: Volume 1: Africa and the Americas* (ed. J.M. Al-Khayri, S.M. Jain and D.V. Johnson). Dordrecht: Springer Netherlands.

Levy-Sakin, M. and Ebenstein, Y. (2013). Beyond sequencing: optical mapping of DNA in the age of nanotechnology and nanoscopy. *Current Opinion in Biotechnology* **24** (**4**): 690–698.

Ma, X., Chen, X., Zhao, J. et al. (2019). Identification of QTLs related to cadmium tolerance from wild rice (Oryza nivara) using a high-density genetic map for a set of introgression lines. *Euphytica* **215**: 205.

Malek, J.A., Mathew, S., Mathew, L.S. et al. (2020). Deletion of beta-fructofuranosidase (invertase) genes is associated with sucrose content in date palm fruit. *Plant Direct* **4**: 1–7.

Maryam, Jaskani, M.J., Awan, F.S. et al. (2016). Development of molecular method for sex identification in date palm (*Phoenix dactylifera* L.) plantlets using novel sex-linked microsatellite markers. *3 Biotech* **6** (**1**): 22.

Mathew, L.S., Spannagl, M., Al-Malki, A. et al. (2014). A first genetic map of date palm (*Phoenix dactylifera*) reveals long-range genome structure conservation in the palms. *BMC Genomics* **15**: 285.

Mathew, L.S., Seidel, M.A., George, B. et al. (2015). A genome-wide survey of date palm cultivars supports two major subpopulations in *Phoenix dactylifera*. *G3: Genes, Genomes, Genetics* **5** (**7**): 1429–1438.

Men, X., Shi, J., Liang, W. et al. (2017). Glycerol-3-phosphate acyltransferase 3 (OsGPAT3) is required for another development and male fertility in rice. *Journal of Experimental Botany* **68** (**3**): 513–526.

Mohamoud, Y.A., Mathew, L.S., Torres, M.F. et al. (2019). Novel subpopulations in date palm (*Phoenix dactylifera*) identified by population-wide organellar genome sequencing. *BMC Genomics* **20**: 498.

Morant, M., Jørgensen, K., Schaller, H. et al. (2007). CYP703 is an ancient cytochrome P450 in land plants catalyzing in-chain hydroxylation of lauric acid to provide building blocks for sporopollenin synthesis in pollen. *Plant Cell* **19** (**5**): 1473–1487.

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

Moussouni, S., Pintaud, J.C., Vigouroux, Y., and Bouguedoura, N. (2017). Diversity of Algerian oases date palm (*Phoenix dactylifera* L., Arecaceae): heterozygote excess and cryptic structure suggest farmer management had a major impact on diversity. *PLoS ONE* **12** (**4**): e0175232.

Nelson, D.R., Schuler, M.A., Paquette, S.M. et al. (2004). Comparative genomics of rice and arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. *Plant Physiology* **135** (**2**): 756–772.

Notsu, Y., Masood, S., Nishikawa, T. et al. (2002). The complete sequence of the rice (Oryza sativa L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Molecular Genetics and Genomics* **268**: 434–445.

O'Leary, S.J., Hollenbeck, C.M., Vega, R.R. et al. (2018). Genetic mapping and comparative genomics to inform restoration enhancement and culture of southern flounder, Paralichthys lethostigma. *BMC Genomics* **19** (**1**): 1–11.

Okazaki, N., Okazaki, K., Watanabe, Y. et al. (1998). Novel factor highly conserved among Eukaryotes controls sexual development in fission yeast. *Molecular and Cellular Biology* **18** (**2**): 887–895.

Pintaud, J.-C., Zehdi, S., Couvreur, T. et al. (2010). Species delimitation in the genus Phoenix (Arecaceae) based on SSR markers, with emphasis on the identity of the date palm (*Phoenix dactylifera* L.). In: *Diversity, Phylogeny, and Evolution in the Monocotyledons* (ed. O. Seberg, G. Petersen, A. Barfod and J. Dennis). Denmark, Aarhus University Press. http://www.couvreurlab.org/uploads/6/4/8/8/64882181/pintaud_et_al._2010_monocots_iv_phoenix.pdf

Popenoe, P. (1922). The pollination of the date palm. *Journal of the American Oriental Society* **42**: 343–354.

Ruan, Y.L., Jin, Y., Yang, Y.J. et al. (2010). Sugar input, metabolism, and signaling mediated by invertase: roles in development, yield potential, and response to drought and heat. *Molecular Plant* **3** (**6**): 942–955.

Runemark, A., Vallejo-Marin, M., and Meier, J.I. (2019). Eukaryote hybrid genomes. *PLoS Genetics* **15** (**11**): e1008404.

Sabir, J.S.M., Abo-Aba, S., Bafeel, S. et al. (2014a). Characterization of ten date palm (*Phoenix dactylifera* L.) cultivars from Saudi Arabia using AFLP and ISSR markers. *Comptes Rendus - Biologies* **337** (**1**): 6–18.

Sabir, J.S.M., Arasappan, D., Bahieldin, A. et al. (2014b). 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**): e94158.

Saker, M.M., Adawy, S.S., Mohamed, A.A., and El-Itriby, H.A. (2006). Monitoring of cultivar identity in tissue culture-derived date palms using RAPD and AFLP analysis. *Biologia Plantarum* **50**: 198–204.

Sakka, H., Zehdi, S., Ould Mohamed Salem, A. et al. (2003). Tunisian date-palm (*Phoenix dactylifera* L.) genotypes identification mediated by plastid PCR/RFLP based DNA. *Journal of Genetics and Breeding* **57** (**3**): 259–264.

Samuelson, M.E. and Larsson, C. (1994). Date palm (*Phoenix dactylifera* L.) cultivar identification by RFLP and RAPD. *Plant Science* **101**: 163–172.

Shelton, J.M., Coleman, M.C., Herndon, N. et al. (2015). Tools and pipelines for BioNano data: molecule assembly pipeline and FASTA super scaffolding tool. *BMC Genomics* **16**: 734.

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

Siljak-Yakovlev, S., Cerbah, M., Sarr, A. et al. (1996). Chromosomal sex determination and heterochromatin structure in date palm. *Sexual Plant Reproduction* **9** (**3**): 127–132.

Singh, R., Ong-Abdullah, M., Low, E.T.L. et al. (2013). Oil palm genome sequence reveals divergence of interfertile species in old and new worlds. *Nature* **500** (**7462**): 335–339.

Singh, R., Low, E.T.L., Ooi, L.C.L. et al. (2014). The oil palm VIRESCENS gene controls fruit colour and encodes a R2R3-MYB. *Nature Communications* **5**: 4106.

Somaratne, Y., Tian, Y., Zhang, H. et al. (2017). ABNORMAL POLLEN VACUOLATION1 (APV1) is required for male fertility by contributing to anther cuticle and pollen exine formation in maize. *Plant Journal* **90** (**1**): 96–110.

Spence, J.P. and Song, Y.S. (2019). Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. *Science Advances* **5** (**10**): eaaw9206.

Tian, X., Zheng, J., Hu, S., and Yu, J. (2006). The rice mitochondrial genomes and their variations. *Plant Physiology* **140** (**2**): 401–410.

Torres, M.F., Mathew, L.S., Ahmed, I. et al. (2018). Genus-wide sequencing supports a two-locus model for sex-determination in Phoenix. *Nature Communications* **9**: 3969.

Wang, J., Na, J.K., Yu, Q. et al. (2012). Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proceedings of the National Academy of Sciences of the United States of America* **109** (**34**): 13710–13715.

Wang, O., Chin, R., Cheng, X. et al. (2019). Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome Research* **538** (**7624**): 243–247.

Yalovsky, S., Rodriguez-Concepcion, M., Bracha, K. et al. (2000). Prenylation of the floral transcription factor APETALA1 modulates its function. *Plant Cell* **12** (**8**): 1257–1266.

Yang, M., Zhang, X., Liu, G. et al. (2010). The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS ONE* **5** (**9**): 1–14.

Yang, X., Wu, D., Shi, J. et al. (2014). Rice CYP703A3, a cytochrome P450 hydroxylase, is essential for development of anther cuticle and pollen exine. *Journal of Integrative Plant Biology* **56** (**10**): 979–994.

Ying, S.T., Zaman, F.Q., Ling, H.O.C. et al. (2007). Flanking AFLP markers for the virescens trait in oil palm. *Journal of Oil Palm Research* **19**: 381–392.

You, F.M., Xiao, J., Li, P. et al. (2018). Chromosome-scale pseudomolecules refined by optical, physical and genetic maps in flax. *Plant Journal* **95** (**2**): 371–384.

Zehdi, S., Cherif, E., Rhouma, S. et al. (2012). Molecular polymorphism and genetic relationships in date palm (*Phoenix dactylifera* L.): the utility of nuclear microsatellite markers. *Scientia Horticulturae* **148**: 255–263.

Zehdi-Azouzi, S., Rhouma, S., Dkkhil-Dakhlaoui, S. et al. (2011). Polymorphism and genetic relationship in date palm using molecular markers. In: *Date Palm Biotechnology* (ed. S.M. Jain, J.M. Al-Khayri and D.V. Johnson). Dordrecht, Springer Netherlands.

Zehdi-Azouzi, S., Cherif, E., Moussouni, S. et al. (2015). Genetic structure of the date palm (*Phoenix dactylifera*) in the Old World reveals a strong differentiation between eastern and western populations. *Annals of Botany* **116** (**1**): 101–112.

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.

Zerpa-Catanho, D.P. and Ming, R. (2019). Papaya sex chromosome evolution and domestication. *Acta Horticulturae* **1250**: 29–138.

Zhang, X., Huang, C., Wu, D. et al. (2017). High-throughput phenotyping and QTL mapping reveals the genetic architecture of maize plant growth. *Plant Physiology* **173** (**3**): 1554–1564.

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

Zohary, D. and Spiegel-Roy, P. (1975). Beginnings of fruit growing in the Old World. *Science* **187** (**4174**): 319–327.

Zohary, D., Hopf, M., and 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: Oxford University Press.

Annual Plant Reviews Online, Volume 4. Edited by Jeremy Roberts.