3. Applications of molecular markers in date palm genome analysis and breeding

Hussam S.M. Khierallah

Date Palm Research Unit, College of Agriculture, University of Baghdad
P.O. Box 47054, Baghdad, Iraq

Abstract. Recently, date palm (Phoenix dactylifera L.) polymorphism and genetic diversity have been studied extensively, bringing about a revolution in this area of research. Date palm population structures and lineage relationships can be identified with the DNA markers since they have been successfully designed and data have proved their efficiency in the genome assessment of this crop. RFLP, AFLP, ISSR and SSR markers took the approach to a new level. In fact, these now represent an important tool for many date palm investigations. Several hundreds of RAPD and ISSR primers have allowed identifying several marker candidates that can distinguish partially or totally between Bayoud disease resistant and susceptible cultivars of date palm. Meanwhile, Microsatellites and AFLPs appear to be an efficient technique, simple and fast DNA marker for the early detection of genetic variations in palms propagated by tissue culture. The problem of sex determination can be solved by employing molecular tools. The most profound effect on the development of the DNA marker resource for date palm is the newly available shotgun sequence. Mining this sequence database
and the steady lowering of the costs of high throughput sequencing will speed up the molecular marker resources and their application to date palm over the next few years. Moreover, efforts are continuing to produce draft linkage and physical maps taking advantage of the Expressed Sequence Tags (ESTs) and the traditional whole-genome mapping methods.

1. Introduction

Date palm (Phoenix dactylifera L.) \((2n = 2x = 36)\) is among the most important species in the Palm family (Arecaceae), which comprise about 200 genera and more than 2500 species (El Hadrami and El Hadrami, 2009, Jain et al., 2011) including \(P.\) canariensis (Canary Islands date palm), \(P.\) reclinata (Senegal date palm) and \(P.\) sylvestris (Indian sugar palm).

The date palm is one of the oldest cultivated fruit trees grown for at least 7000 years. It is believed to have originated in Mesopotamia (Wrigley, 1995). It was well-known in Babylon, Iraq 4000 B.C. The palms were celebrated during that period for their strength and majesty. People were heavily dependent on this tree to provide food and wood for making tools, furniture and baskets. Fig. 1 shows the blessed date palm during the Sumerian era. Date palm is an arborescent, dioecious, highly heterozygous, monocotyledonous plant, with a very slow growth rate and a late reproductive phase. Palm is one of the tallest domesticated trees and is an excellent candidate for cultivation in arid and semi-arid regions of the world due to its high tolerance to environmental stresses. It is of great economic importance to oasis agriculture and creates favorable conditions for improving secondary crops. Accordingly, date palm represents a cornerstone of the economy in many producing countries, especially in North Africa and the Middle East. Over 100 million trees are currently grown worldwide on an estimated area of 1 million ha. Date palm provides fruit, fuel, fiber and shade for other essential crops. The annual world production of dates has reached 6-8 million mt, representing a market exchange value of over 1 billion USD. (El Hadrami and El Hadrami, 2009; El Hadrami et al., 2011).

The top three producing countries are Egypt, Iran and Saudi Arabia; the largest importer of dates is India. The number of known cultivars distributed all over the world was reported to be approximately 5000 (Bashah, 1996), out of which 650 are cultivated in Iraq (Ibrahim, 2008), 340 in Saudi Arabia (Al-Mssallem, 1996) and 135 in United Arab Emirates (Ghaleb, 2008).
Figure 1 A, B: The blessed date palm in the Sumerian era, C, D: General morphological features, E: The tree structure
http://www.fao.org/docrep/006/y4360e/y4360e00.jpg
2. Genetic diversity in date palm

The success of any plant genetic conservation or breeding program depends on understanding the amount and distribution of the genetic variation present in the gene pool. This understanding will be helpful in:

(a) Gathering date palm cultivars in newly-established orchards using genetically diverse cultivars to avoid genetic vulnerability to various biotic and abiotic stresses.
(b) Selection of the diverse parents in combination with the aim of segregating progenies with genetic variability.
(c) Providing further gain for selection and molecular mapping.

2.1. Genetic diversity as inferred from morphological traits

Morphological traits have been used to describe genetic variation in date palm cultivars, which are mainly related to the fruit, leaf, trunk and other parts of the tree. An early survey of Iraqi cultivars using morphological description was done by Al-Baker (1962); he distinguished 531 cultivars and their origin, availability and fruit quality. Another morphological identification was conducted on 110 Iraqi cultivars by Al-Jboory (1971). In 1972, Al-Baker again described 627 Iraqi cultivars in his celebrated book in Arabic, The Date Palm, adding some fruit traits to the earlier description. Three decades later, Husien (2002) and Husien and Graib (2004) also gave a brief description of 36 and 50 cultivars, respectively. Al-Saleh and Al-Ansary (2005) adopted 12 phenotypic traits to describe 110 cultivars, documented with color photos. Hamza et al. (2011) evaluated nine morphological traits as markers for estimating the diversity and genetic structure of Tunisian date palm subpopulations. Genetic identification of date palm cultivars using morphological markers is usually not possible until fruits are produced and frequently requires a large set of phenotypic data that is often difficult to assess and is sometimes variable due to environmental influences.

2.2. Genetic diversity as inferred from biochemical markers

The development of molecular tools has changed the way in which individual cultivars can be identified and useful information concerning the genetic control of many agronomic characteristics can be analyzed. The ability to apply these molecular tools depends to some extent on the amount of other genomic information available for the specific plant species.
Molecular marker technologies involve the use of isozymes or followed by a series of DNA marker technologies and, most recently, by possibilities to compare complete genomes (Cullis, 2011). Isozymes originate through amino acid alterations, which cause changes in net charge, or the spatial structure (conformation) of the enzyme molecules and also, therefore, their electrophoretic mobility. After specific staining the isozyme profile of individual samples can be observed (Hadacová and Ondrej, 1972, Vallejos, 1983, Soltis and Soltis, 1989). Allozymes are allelic variants of enzymes encoded by structural genes. In addition, biochemical studies, including isozymes and activity analyses of peroxidases have been used to characterize date palms in Morocco and Tunisia (Baaziz, 1988, Baaziz and Saaidi, 1988, Bendiab et al., 1998, Ould Salem Mohamed et al., 2001, Majourhat et al., 2002). Such analyses do not reflect precisely the polymorphisms which have occurred (Al-Jibouri and Adham 1990). Date palm genomic research is one the efficient way to complement breeding practices including diversity investigation, and mapping proteins involved in important biochemical pathways relevant to different and important traits. The main weakness of allozymes is their relatively low abundance and low level of polymorphism. Moreover, proteins with identical electrophoretic mobility (co-migration) may not be homologous for distantly related germplasm. In addition, their selective neutrality may be in question (Berry and Kreitman, 1993, Hudson et al., 1994, Krieger and Ross, 2002). Lastly, often allozymes are considered molecular markers since they represent enzyme variants, and enzymes are molecules. However, allozymes are in fact phenotypic markers, and as such they may be affected by environmental conditions. For example, the banding profile obtained for a particular allozyme marker may change depending on the type of tissue used for the analysis (e.g. root vs. leaf). This is because a gene that is being expressed in one tissue might not be expressed in other tissues.

2.3. Genetic diversity as inferred from DNA molecular markers

Development of suitable DNA molecular markers for date palm may allow researchers to estimate genetic diversity, which will ultimately lead to the genetic conservation of date palm. The success of particular genetic conservation or breeding program is dependent on understanding the amount and distribution of the genetic variation already present in the gene pool (Jubrael et al., 2005). The development of DNA (or molecular) markers has irreversibly change the disciplines of plant genetics and breeding (Collard and Mackill, 2008). Joshi et al (2011) reported an ideal DNA makers should however possess the following properties.
1- High polymorphism, which is the simultaneous occurrence of a trait at the same population of two or more discontinues variants or genotypes.

2- Co dominant inheritance- different form of marker should be detected in a diploid organism to allow discrimination of homozygote and heterozygote.

3- Frequent occurrence in genome.

4- Selective neutral behavior (the DNA sequences of any organism which are neutral to environmental conditions or management practices).

5- Easy access (availability).

6- Easy and fast assay.

7- Reproducible – high level of reproducibility.

8- Easy exchange of data between laboratories.

It is extremely difficult for a single genetic marker to possess all properties above. Depending on the type of study to be undertaken a marker system can be identified that would fulfill at least a few of the above characteristics. Among the various kinds of genotyping techniques characterized so far, some types and description of DNA markers used in date palm fingerprinting are illustrated as follows:

### 2.3.1. Restriction Fragment Length Polymorphism (RFLP)

Non-PCR based genetic markers (Restriction Fragment Length Polymorphism) are the first and forest molecular markers system. It was developed in early 1980. RFLP is the most widely used hybridization-based molecular marker, initially used to identify DNA sequence polymorphisms for genetic mapping of a temperature-sensitive mutation of adenovirus serotypes. It was initially used for human genome mapping (Botstein et al., 1980) but later on was adopted for plant genomes.

In RFLP analysis, restriction enzyme-digested genomic DNA is obtained by gel electrophoresis and then blotted on nitrocellulose membrane. Specific banding patterns are then visualized by hybridization with labeled probe. These probes are mostly species-specific single- or multi-locus probes of about 0.5-3.0 kb in size, obtained from a cDNA library or a genomic library.

RFLP markers were used for the first time in the construction of genetic maps. Being co-dominant markers, RFLP can detect coupling phase of DNA molecules, as DNA fragments from all homologous chromosomes are detected (Fig. 2).
Figure 2. The various outcomes of hybridization with different probes to identify the basis of various RFLPs. I – An RFLP caused by a loss of a restriction site. II – The banding patterns after Southern blotting and hybridization with probes a and b. III – RFLPs caused by an insertion (ii) or a deletion (iii) in the original allele (i). IV – The banding patterns after Southern blotting and hybridization with the probe for these three alleles (Cullis 2011).

The major strengths of RFLP markers are their high reproducibility, co-dominant inheritance, and good transferability between laboratories which provides locus-specific markers that allow synteny (conserved order of genes between related organisms) studies. For this, no sequence information is required and they are relatively easy to score due to large size differences between fragments (Kesawat and Das, 2009). Still, there are several limitations for RFLP analysis: it requires the presence of high quantity and quality of DNA (Young et al., 1992), requirement of radioactive isotope makes the analysis relatively expensive and hazardous. The assay is time-consuming and labor-intensive. RFLPs can be applied in diversity and phylogenetic studies ranging from individuals within populations or species, to closely related species. It is widely used in gene mapping studies because of their high genomic abundance due to the
ample availability of different restriction enzymes and random distribution throughout the genome. Basically, it was used to investigate relationships of closely related taxa (Miller and Tanksley, 1990), fingerprinting tools for diversity studies, and for studies of hybridization and introgression, as well as studies of gene flow between crops and weeds (Desplanque et al., 1999).

The development of RFLPs in date palm has been reported by Corniquel and Mercier (1994, 1996). These molecular markers were designed for Tunisian date palms in which homologous and heterologous labeled probes were described to discriminate date palm cultivars (Trifi et al., 2000). However, the number of genotypes studied was limited to a small set of cultivars that are widely cultivated in oases and/or characterized by fleshy dates, and they are not representative of the germplasm. Moreover, taking into account the use of radio-labeled probes, the large amounts of DNA used in these studies, the development of additional molecular markers has become imperative to overcome these inconveniences (Zehdi et al., 2011).

2.3.2. Random Amplified Polymorphic DNA (RAPD)

These are DNA fragments amplified by the PCR using short synthetic primers (generally 10 bp long) of random sequence. The oligonucleotides serve as both forward- and reverse-primers and are usually able to amplify fragments from one to ten genomic sites simultaneously (Fig 3). Size of amplified fragments, usually ranges from 0.5 to 5 kb in size, are separated by agarose gel electrophoresis. After ethidium bromide staining polymorphisms can be detected, as the presence or absence of bands of particular sizes. Polymorphisms are considered to be primarily due to variation in the primer annealing sites, but it can also be generated by length differences in the amplified sequence between primer annealing sites (Williams et al., 1990). The main advantages of RAPDs are: they are less time consuming, easy to assay, and low quantities of template DNA are required, usually 5-50 ng per reaction. Due to the commercial availability of random primers, no sequence data for primer construction is needed. Moreover, RAPDs have a very high genomic abundance and are randomly distributed throughout the genome. The main disadvantage of RAPDs is low reproducibility hence highly standardized experimental procedures are needed because of its sensitivity to the reaction conditions. RAPD analyses generally require purified, high molecular weight DNA (Kesawat and Das, 2009).
At the end of last century it was necessary to find out techniques for characterizing date palm germplasm and to survey relationships among cultivars. Taking into account the advantages of RAPD (i.e. small amounts of DNA, fast, no knowledge of the genome is needed and low costs), work has focused on development of the mentioned PCR-based methods to examine genetic diversity as well to characterize genotypes. Studies have described the use of universal random primers to generate molecular markers suitable to more precise genetic diversity at the DNA level. In Morocco, Sedra et al. (1998) reported evidence of RAPDs correlated with the resistance to bayoud disease from a set of date palm cultivars. Saker and Moursy (1999) performed DNA fingerprinting for five Egyptian date palm cultivars (Zaghloul, Amhat, Samany, Haiany and Siwi) using four RAPD primers. Moreover, Jubrael et al. (2001) used RAPD markers to identify nine Iraqi date palm female cultivars. A genetic comparison of Egyptian date palm cultivars by RAPD-PCR technique was carried out for four

**Figure 3.** The principle of RAPD-PCR technique. Arrows indicate primer annealing sites (Arif et al., 2010).
well-known female cultivars and four unknown males (Soliman et al., 2003). Other studies were also conducted to characterize RAPD profile for date palm cultivars from Saudi Arabia (Askari et al., 2003, Motawei et al., 2003) and Gaza Strip (Palestinian Authority) (El Kichaoui1 et al., 2013).

2.3.3. Amplified Fragment Length Polymorphism (AFLP)

Amplified Fragment Length Polymorphism (AFLP) is where DNA is amplified by a selective PCR of a completely digested genomic DNA. This technique utilizes the power of RFLP with PCR-based technology by ligation of primer recognition sequences (adaptors) to the restricted DNA. The key feature of AFLP is its capacity for “genomerepresentation” and the simultaneous screening of representative DNA regions distributed randomly throughout the genome. AFLPs are DNA fragments (80-500 bp) obtained from digestion with restriction enzymes, followed by ligation of oligonucleotide adapters to the digestion products and selective amplification by the PCR. The PCR primers consist of a core sequence (part of the adapter) and a restriction enzyme specific sequence, and one to five selective nucleotides (the higher the number of selective nucleotides, the lower the number of bands obtained per profile) (Fig. 4). The banding profiles are the result of variations in the restriction sites or in the intervening regions. The technique simultaneously generates fragments from many genomic sites (usually 50-100 fragments per reaction) that are separated by polyacrylamide gel electrophoresis and generally scored as dominant markers (Vos et al., 1995). AFLP markers were used to generate a genetic map in a segregating mapping progenies based on the pattern of the segregation of AFLP generated bands as well as mapping some fruit quality traits, salinity and humidity tolerance (El-Kharbotly et al., 1998). They were also used for the identification of Californian cultivars (Cao and Chao, 2002).

PCR-based DNA profiling of five Egyptian date cultivars was conducted using AFLPs (El-Khishin et al., 2003). A total of 433 amplification products were generated using six primer pair combinations (EcoR1 and Mse1) with a mean of 72.17 amplicons per assay. The dendrogram generated by the UPGMA (unweighted pair group method using arithmetic averages) formed two major clusters with Siwi and Hayany being the most genetically similar cultivars, and in the second cluster Amhat and Samany being next, while Zaghloul was the most distinct cultivar. Jubrael et al. (2005) reported that high level of intervartietal AFLP polymorphism using five primer combinations among 18 Iraqi date palm cultivars could be partly due to the strong out-crossing
mechanism in this species, which is likely to increase the degree of polymorphism. El Assar et al. (2005) or Rhouma et al. (2007) in Tunisia or Egypt respectively, tested different sets of primer pairs for their ability to generate AFLP banding patterns using total cellular DNAs as templates. As a result large numbers of AFLPs were easily produced suggesting that the primers tested in the studies are a powerful means to provide evidence of DNA polymorphisms in this crop. This assumption is strongly supported in regard to the high percentage values of polymorphic bands scored using each primer set. Therefore, these data suggest that AFLP constitutes a very attractive and informative procedure for providing evidence of the genetic diversity among date palm ecotypes (Fig. 5). Khierallah et al. (2011a) carried out AFLP analysis for another 18 Iraqi cultivars. A total of 83 polymorphic AFLP fragments were detected.

Figure 4. A schematic flow chart showing the principle of the AFLP method (Arif et al., 2010).
All primer combinations contributed to the discrimination of date palm cultivars, suggesting the efficiency of AFLP method in assessing genetic diversity in date palm (Tables 1, 2). Although, Iraqi date palm cultivars have been grown in Iraq for many years, Jaccard’s similarity index and Principle Component Analysis (PCA) revealed diverse relationships among them (Fig. 6). In fact, the cultivars were clustered independently of their geographic origin in spite of their phenotypic distinctiveness.
Table 1. Number of fragments amplified, polymorphic bands, primer efficiency and discrimination power of six primer combinations used for AFLP analysis (Khierallah et al., 2011a).

<table>
<thead>
<tr>
<th>Primer combination</th>
<th>Total fragment (No.)</th>
<th>% Primer efficiency</th>
<th>No. polymorphic fragment</th>
<th>% Polymorphism</th>
<th>% Discrimination power</th>
</tr>
</thead>
<tbody>
<tr>
<td>P11-aagc/M88-tgc</td>
<td>63</td>
<td>25</td>
<td>12</td>
<td>19</td>
<td>14</td>
</tr>
<tr>
<td>P104-aagc/M95-aaaa</td>
<td>33</td>
<td>13</td>
<td>7</td>
<td>21</td>
<td>8</td>
</tr>
<tr>
<td>P74-ggt/M95-aaaaa</td>
<td>46</td>
<td>18</td>
<td>26</td>
<td>57</td>
<td>31</td>
</tr>
<tr>
<td>P11-aagc/M95-aaaaa</td>
<td>31</td>
<td>12</td>
<td>5</td>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td>P293-taca/M62-tgc</td>
<td>45</td>
<td>18</td>
<td>21</td>
<td>47</td>
<td>25</td>
</tr>
<tr>
<td>P101-aagc/M95-aaaaa</td>
<td>34</td>
<td>13</td>
<td>12</td>
<td>35</td>
<td>14</td>
</tr>
<tr>
<td>Total</td>
<td>252</td>
<td></td>
<td>83 (33 %)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Major allele frequency, gene diversity and polymorphism information content (PIC) estimated by AFLP markers in 18 date palm cultivars of Iraq (Khierallah et al., 2011a).

<table>
<thead>
<tr>
<th>Primer combination</th>
<th>Major allele frequency</th>
<th>Range of gene diversity</th>
<th>Average of gene diversity</th>
<th>Range of PIC</th>
<th>Average of PIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>P11-aa/M88-tgc</td>
<td>0.75</td>
<td>0.10–0.48</td>
<td>0.29</td>
<td>0.10–0.36</td>
<td>0.23</td>
</tr>
<tr>
<td>P104-aagc/M95-aaaa</td>
<td>0.78</td>
<td>0.10–0.48</td>
<td>0.29</td>
<td>0.10–0.36</td>
<td>0.23</td>
</tr>
<tr>
<td>P74-ggt/M95-aaaaa</td>
<td>0.72</td>
<td>0.10–0.50</td>
<td>0.30</td>
<td>0.10–0.38</td>
<td>0.24</td>
</tr>
<tr>
<td>P11-aa/M95-aaaaa</td>
<td>0.75</td>
<td>0.10–0.49</td>
<td>0.30</td>
<td>0.10–0.37</td>
<td>0.24</td>
</tr>
<tr>
<td>P293-taca/M62-ctt</td>
<td>0.72</td>
<td>0.10–0.50</td>
<td>0.30</td>
<td>0.10–0.38</td>
<td>0.24</td>
</tr>
<tr>
<td>P101-aagc/M95-aaaaa</td>
<td>0.70</td>
<td>0.20–0.50</td>
<td>0.35</td>
<td>0.18–0.38</td>
<td>0.28</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td>0.31</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Figure 6. Principle component analysis of the 18 date palm cultivars estimated by AFLP markers. (Khierallah et al., 2011a).
2.3.4. Microsatellites

The microsatellites (Litt and Lutty, 1989), also known as Simple Sequence Repeats (SSRs), are sections of DNA, consisting of tandemly repeated mono-, di-, tri-, tetra- or penta-nucleotide units that are arranged throughout the genomes of most eukaryotic species (Powell et al., 1996). Microsatellite markers, developed from genomic libraries, belong to either the transcribed region or the non-transcribed region of the genome, and rarely is there information available regarding their functions. Microsatellite sequences are especially suited to distinguish closely related genotypes; because of their high degree of variability, they are, therefore, favored in population studies (Smith and Devey, 1994) and for the identification of closely related cultivars. Microsatellite polymorphism can be detected by Southern hybridization or PCR (Fig. 7).

Microsatellites, like minisatellites, represent tandem repeats, but their repeat motifs are shorter (1–6 base pairs). If nucleotide sequences in the flanking regions of the microsatellite are known, specific primers (generally 20–25 bp long) can be designed to amplify the microsatellite by PCR. Microsatellites and their flanking sequences can be identified by constructing a small-insert genomic library, screening the library with a synthetically labeled oligonucleotide repeat and sequencing the positive

---

Figure 7. Microsatellite (SSR) markers (modified from www.takween.com).
clones (Kumar et al. 2009). They are ideal DNA markers for population studies and genetic mapping due to their abundance, high level of polymorphism, dispersion throughout diverse genomes, ease to assay by PCR, and ease to disseminate among laboratories (Udupa and Baum, 2001). Although the frequency of microsatellites varies among species, they are abundant, dispersed throughout the genome and show higher levels of polymorphism than other genetic markers. These features, coupled with their ease of detection, suggesting their use as molecular markers. Their potential for automation and inheritance in a co-dominant manner are additional advantages when compared with other types of molecular tools (Holton et al. 2001).

SSR markers have been used for investigating genetic diversity in date palm (Billotte et al., 2004). Zehdi et al. (2004) applied these markers to characterize Tunisian cultivars, while Al-Ruqaishi et al. (2008) utilized these primers to screen and analyze the genetic diversity among clonal genotypes of Omani cultivars. The technique spread to other date palm producing countries. Elshibli and Korpelainen (2008) investigated genetic diversity in Sudan germplasm representing 37 female and 23 male accessions using 16 SSR primers. In Qatar, Ahmed and Al-Qaradawi (2009) employed 10 primers to analyze genetic diversity among 15 cultivars. Khierallah et al. (2011b) tested genetic diversity of Iraqi date palm using 33 SSR primer pairs developed by Billotte et al. (2004) and Akkak et al. (2009). A total of 22 primers successfully showed polymorphic bands among the 30 cultivars (24 female and 6 male) as displayed in Table 3. Obtained electropherogram data were analyzed by using Foundation Data Collection software (Genetic Analyser Data Collection Version 2.0© Applied Biosystem 3100). Allele size scoring was performed by Gene Mapper Software (GeneMapper® Software Version 3.7, Applied Biosystem, Carlsbad, California). In order to investigate the genetic relationships among the cultivars, the Jaccard similarity matrix was used for cluster analysis using the unweighted pair group method arithmetic average (UPGMA). Jaccard similarity index, major allele frequency, heterozygosity, gene diversity and polymorphism information content (PIC) estimation were done using a software package (PowerMarker version 1.31) (Liu and Muse, 2005). The phylogenetic diagram was drawn by PAST software version 1.91 (Hammer et al., 2001) on the basis of the Hamming similarity index with 100 bootstrap (Fig. 8).
The microsatellites examined in this study showed highly polymorphic patterns with a great number of alleles (188) distributed in 30 date palm cultivars. The number of detected alleles per locus in this study (8.54) was higher than 7.6 alleles per locus scored by Zehdi et al. (2004) in 46 date palm cultivars cultivated in Tunisia, and those studied by Ahmed and Al-Qaradawi (2009) representing 15 cultivars grown in Qatar. The mean heterozygosity value detected in the Iraqi cultivars was 0.503, indicating the presence of high genetic diversity. The long history of date palm

<table>
<thead>
<tr>
<th>Marker</th>
<th>Major allele frequency</th>
<th>Genotypes (no.)</th>
<th>Alleles (no.)</th>
<th>Heterozygosity</th>
<th>Gene Diversity</th>
<th>PIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>mPdClR010</td>
<td>0.204</td>
<td>19</td>
<td>14</td>
<td>0.852</td>
<td>0.890</td>
<td>0.880</td>
</tr>
<tr>
<td>mPdClR016</td>
<td>0.500</td>
<td>8</td>
<td>5</td>
<td>0.529</td>
<td>0.654</td>
<td>0.602</td>
</tr>
<tr>
<td>mPdClR025</td>
<td>0.224</td>
<td>15</td>
<td>8</td>
<td>0.690</td>
<td>0.822</td>
<td>0.798</td>
</tr>
<tr>
<td>mPdClR032</td>
<td>0.333</td>
<td>13</td>
<td>5</td>
<td>0.815</td>
<td>0.772</td>
<td>0.736</td>
</tr>
<tr>
<td>mPdClR035</td>
<td>0.481</td>
<td>10</td>
<td>7</td>
<td>0.519</td>
<td>0.695</td>
<td>0.659</td>
</tr>
<tr>
<td>mPdClR050</td>
<td>0.370</td>
<td>17</td>
<td>12</td>
<td>0.926</td>
<td>0.813</td>
<td>0.797</td>
</tr>
<tr>
<td>mPdClR057</td>
<td>0.909</td>
<td>3</td>
<td>3</td>
<td>0.000</td>
<td>0.604</td>
<td>0.536</td>
</tr>
<tr>
<td>PdClR070</td>
<td>0.533</td>
<td>3</td>
<td>3</td>
<td>0.000</td>
<td>0.604</td>
<td>0.536</td>
</tr>
<tr>
<td>mPdClR078</td>
<td>0.320</td>
<td>16</td>
<td>15</td>
<td>0.760</td>
<td>0.832</td>
<td>0.816</td>
</tr>
<tr>
<td>mPdClR085</td>
<td>0.313</td>
<td>8</td>
<td>7</td>
<td>0.250</td>
<td>0.779</td>
<td>0.746</td>
</tr>
<tr>
<td>mPdClR090</td>
<td>0.391</td>
<td>10</td>
<td>9</td>
<td>0.261</td>
<td>0.781</td>
<td>0.757</td>
</tr>
<tr>
<td>mPdClR093</td>
<td>0.741</td>
<td>6</td>
<td>6</td>
<td>0.296</td>
<td>0.434</td>
<td>0.415</td>
</tr>
<tr>
<td>PDCAT4</td>
<td>0.407</td>
<td>13</td>
<td>7</td>
<td>0.556</td>
<td>0.738</td>
<td>0.701</td>
</tr>
<tr>
<td>PDCAT 5</td>
<td>0.276</td>
<td>13</td>
<td>7</td>
<td>0.862</td>
<td>0.816</td>
<td>0.791</td>
</tr>
<tr>
<td>PDCAT 6</td>
<td>0.231</td>
<td>21</td>
<td>21</td>
<td>0.654</td>
<td>0.903</td>
<td>0.897</td>
</tr>
<tr>
<td>PDCAT 11</td>
<td>0.600</td>
<td>7</td>
<td>7</td>
<td>0.100</td>
<td>0.598</td>
<td>0.566</td>
</tr>
<tr>
<td>PDCAT 12</td>
<td>0.850</td>
<td>5</td>
<td>4</td>
<td>0.100</td>
<td>0.269</td>
<td>0.256</td>
</tr>
<tr>
<td>PDCAT 14</td>
<td>0.220</td>
<td>19</td>
<td>10</td>
<td>0.960</td>
<td>0.865</td>
<td>0.850</td>
</tr>
<tr>
<td>PDCAT 15</td>
<td>0.440</td>
<td>8</td>
<td>6</td>
<td>0.480</td>
<td>0.732</td>
<td>0.699</td>
</tr>
<tr>
<td>PDCAT 17</td>
<td>0.385</td>
<td>11</td>
<td>9</td>
<td>0.500</td>
<td>0.774</td>
<td>0.746</td>
</tr>
<tr>
<td>PDCAT 18</td>
<td>0.140</td>
<td>19</td>
<td>19</td>
<td>0.560</td>
<td>0.923</td>
<td>0.918</td>
</tr>
<tr>
<td>PDCAT 21</td>
<td>0.731</td>
<td>6</td>
<td>4</td>
<td>0.269</td>
<td>0.431</td>
<td>0.394</td>
</tr>
<tr>
<td>Mean</td>
<td>0.436</td>
<td>11.364</td>
<td>8.545</td>
<td>0.503</td>
<td>0.695</td>
<td>0.669</td>
</tr>
</tbody>
</table>

Table 3. Major allele frequency, number of genotypes showed polymorphic bands, number of alleles generated, heterozygosity, gene diversity and polymorphism information content (PIC) estimated by 22 SSR markers in 30 Iraqi date palm cultivars. (Khierallah et al., 2011b).
domestication of an unknown origin (Wrigley 1995) and the nature of date palm culture may have played an important role in the composition of date palm genome (Elshibli and Korpelainen, 2008). New cultivars may appear as a result of the continuous selection carried out by farmers following sexual reproduction. Exchange of propagules, which are a mixture of vegetative and seed-propagated materials, is conducted between farmers. All these processes together may result in a mixed genome within the same country (Elshibli and Korpelainen, 2008).

**Figure 8.** Genetic relationships among 30 Iraqi date palm cultivars based on 22 microsatellite loci with 100 bootstrap (Khierallah et al. 2011b).
A small number of developed SSR markers for date palm cultivars discrimination up to 2009 has encouraged some researchers to develop new primers. Hamwieh et al. (2010) developed over 1000 Simple Sequence Repeat (SSR) primer pair by mining genome sequencing data derived from an assembly draft of the date palm genome generated by whole genome shotgun next generation DNA sequencing issued by researchers in the Weill Cornell Medical College in Qatar (WCMC-Q) (Malek et al., 2010). Analysis of the microsatellite motifs across the date palm genome was done (Table 4). This work was made in Integrated DNA Technologies (IDT), Inc and awarded the Khalifa International Date Palm Award for research in 2012.

Elmeer et al. (2011) found that these new co-dominant markers would be a starting point for researchers making use of such markers for genetic mapping and diversity analysis of date palm. They tested 30 primer pairs of genomic DNA microsatellite markers in order to assess the genetic diversity of eleven date palm genotypes. The results indicated that out of thirty primers, only seven (23.3%) failed to amplify the expected PCR fragments, while thirteen primers (43.3%) amplified monomorphic banding patterns and the remaining ten primers (33.4%) generated polymorphic banding patterns.

In Iran, Arabnezhad et al. (2011) assessed genetic relationships among date palm genotypes grown in different geographical regions by using newly developed simple sequence repeat (SSR) markers. Two SSR-enriched genomic libraries including repeat motifs (AG)n and (AAG)n were constructed in date palm. Based on DNA sequences of positive clones, 25 primer pairs were designed of which 22 pairs were able to detect polymorphism in 16 date palm cultivars from Iran, Iraq and Africa. The selected SSR primers amplified a total of 106 alleles with an average of 4.82 alleles per locus among the cultivars and the average values of He and PIC were 0.719 and 0.668, respectively. Neighbor-Joining cluster analysis based on Nei’s genetic distance divided date palm accessions into three major clusters in agreement with their geographical origin. Cluster analysis significantly

Table 4. Frequency of various types of microsatellite motifs observed across the date palm genome (Hamwieh et al., 2010).

<table>
<thead>
<tr>
<th></th>
<th>Di-nucleotide</th>
<th>Tri-nucleotide</th>
<th>Tetra-nucleotide</th>
<th>Penta-nucleotide</th>
<th>Hexa-nucleotide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perfect</td>
<td>24256</td>
<td>13182</td>
<td>3729</td>
<td>5449</td>
<td>22</td>
</tr>
<tr>
<td>Imperfect</td>
<td>27484</td>
<td>15180</td>
<td>1807</td>
<td>7409</td>
<td>3545</td>
</tr>
<tr>
<td>Total</td>
<td>51740</td>
<td>28362</td>
<td>5536</td>
<td>12858</td>
<td>5797</td>
</tr>
<tr>
<td>Compound</td>
<td>702</td>
<td>141</td>
<td>19</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>52442</td>
<td>28503</td>
<td>5555</td>
<td>12873</td>
<td>5810</td>
</tr>
</tbody>
</table>
distinguished African cultivars from Iranian and Iraqi ones, suggesting that the domestication of African date palms has followed a different route than those grown in the Middle-East, an assumption which is supported by Mantel test and Bayesian analysis. The set of date palm SSR loci developed in this study could be an informative marker system for geographic partitioning and genotyping of date palm germplasm.

2.3.5. Inter-Short Sequence Repeat (ISSR)

This marker comprises the amplification of DNA segments present at the region between two identical microsatellite repeats oriented in opposite directions. The technique uses microsatellites as primers in a single primer PCR reaction targeting multiple genomic loci to amplify mainly ISSR of different sizes. The microsatellite repeats used as primers for ISSRs can be di-nucleotide, tri-nucleotide, tetra-nucleotide, or penta-nucleotide. The primers can be either unanchored (Meyer et al., 1993, Gupta et al., 1994) or more usually anchored at 3’ or 5’ end with one to four degenerate bases extended into the flanking sequences (Zietkiewicz et al., 1994). ISSRs use longer primers (15-30 mers) as compared to RAPD primers (10 mers), which permit the subsequent use of a high annealing temperature leading to higher stringency. The annealing temperature depends on the GC content of the primer used and ranges from 45° to 65° C. The amplified products are usually 200-2000 bp long and can be detected by both agarose and polyacrylamide gel electrophoresis. ISSRs exhibit the specificity of microsatellite markers, but do not need sequence information for primer synthesis utilizing the advantage of random markers (Joshi et al., 2000). ISSR markers are randomly distributed throughout the genome and usually show high polymorphism although the level of polymorphism has been shown to vary with the detection method used. Disadvantages include the possibility of nonhomology of similar-sized fragments. Moreover, ISSRs, like RAPDs, may have reproducibility problems. ISSR analysis can be applied in studies involving genetic identity, parentage, clone and strain identification, and taxonomic studies of closely related species as well as in gene mapping studies (Gupta et al., 1994; Zietkiewicz et al., 1994).

Hamza et al. (2012) investigated the structure and the distribution of genetic variability in 26 cultivars of Tunisian continental oases using seven ISSR primers and to assess the efficiency of ISSR in determining fruit characteristics. Genetic diversity and relationships among cultivars were assessed and a total of 43 amplified bands were obtained. Principal Component Analyses based on Nei genetic distances showed no geographical separation with the exception of the 'Dhahbi' cultivar, which
has a limited geographical distribution. A group of cultivars that is also phonetically clustered was distinguished. These cultivars have a common maturity period and a common fruit consistency. The Mantel test emphasizes a significant correlation between genetic distance and fruit consistency \( (r = -0.120; P = 0.026) \). A significant differentiation was observed between the soft and dry subpopulations \( (\Phi_{PT} = 0.126; P = 0.007) \). Discriminate analyses highlight the association of markers with fruit consistency groups. Talaat et al. (2013) determined the genetic similarity or diversity among and within the well-known Qatari date palm cultivars using 18 different primers of inter simple sequence repeat (ISSR). Five common and the most cultivated date palm cultivars in Qatar were selected including 'Khalas', 'Sheshy', 'Rezezy', 'Barhee' and 'Khanezy' from three different locations (Al-Shamal, Al-Khour and Al-Rayan). All primers had amplified polymorphic bands in the studied cultivars either among the cultivars or within each cultivar in different cultivated areas. These results reveal the existence of genetic variations among the studied cultivars as well as within each cultivar, supporting the observed variation in some morphological and quality characters for different trees that are grown in different environments and derived from the same cultivar.

### 2.3.6. Multiple markers analysis

Many researchers combined two or more DNA markers in studying genetic diversity in date palm. Different marker systems differ in the mechanism of detecting polymorphism, genome coverage and the ease of application. Therefore, they could complement each other to draw more accurate conclusions. Trifi et al. (2000) and Zehdi et al. (2004) reported the use of RAPD and ISSR to generate discrete markers correlated with fruit parameters such as fruit size and color of Tunisian cultivars. On the other hand, starting from a set of local and introduced cultivars, together with male genotypes, either RAPDs or ISSRs derived genotypic clustering was achieved, independent of the gender of the trees and/or the geographical origin of the cultivars. This suggests a common genetic basis which characterizes date palm cultivars and concurs with the Mesopotamian domestication origin (i.e. the Fertile Crescent) of this fruit crop as proposed by Munier (1973) and Wrigley (1995).

In Egypt, Adawy et al. (2004) combined data of three types of DNA markers AFLP, RAPD and ISSR to assay relationships among fourteen date palm accessions collected from different locations representing six Egyptian cultivars. The dendrogram based on the combined data from the different types of markers (RAPD, ISSR and AFLP) was closest to the AFLP-based
Applications of molecular markers in date palm genome analysis and breeding

The AFLP exhibited considerably high sum effective number of alleles (205.7) compared to RAPD and ISSR (45.1 and 17.8, respectively). The average heterozygosity was also higher in AFLP (0.39) than in RAPD and ISSR (0.36 and 0.35, respectively). The MI was 117.3 in AFLP while it was 95.9 and 10.4 in RAPD and ISSR respectively. Thus, the results indicated that AFLP is more effective in detecting high level of polymorphism. The correlation coefficient was considerably high between RAPD and ISSR (0.68), and it was lower between RAPD and AFLP (0.23) than that between AFLP and ISSR (0.34).

RAPD and ISSR markers were used to assess genetic polymorphism in four date palm cultivars from Egypt (Adawy et al., 2002), four cultivars from Saudi Arabia (Munshi and Gamal, 2010), 23 date palm cultivars grown in Syria representing 18 female and five male cultivars (Haider et al., 2012) and 17 Iraqi cultivars (Khierallah et al., 2014).

2.3.7. Gene-based markers

Recent trends in plant research towards the use of gene-targeted rather than random DNA markers as inexpensive and faster estimation of genome sequence, lately offer enormous potential for the development of such gene-based markers (Andersen and Lubberstedt, 2003). Gene-based markers are more useful in mapping of quantitative trait loci (QTL), molecular breeding and gene cloning.

2.3.7.1. Expressed Sequence Tags derived SSRs (EST-SSRs)

The Expressed Sequence Tags derived SSRs (EST-SSRs) are also known as genic SSRs. EST-derived SSRs form a valuable genetic marker type, a class of functional markers as a putative function, in mapping candidate genes. Distribution of genic SSRs on the genetic map shows the distribution of genes in the genome. Thus, EST-SSRs have been widely used to construct high-density linkage maps in recent years (Chen et al., 2008, Durand et al., 2010, Ramchiary et al., 2011) and some EST-SSRs associated with phenotype are useful in marker assisted breeding programs (Qi et al., 2010, Zhang et al., 2011). Another important feature of the genic SSR markers is that, unlike genomic SSRs, they are transferable among related species and genera (Varshney et al., 2005). To date, no gene-based markers have been identified in date palm. The challenge remains not only to identify genes that are responsible for the traits of agronomic interest, but also to identify the gene-related markers that could be used in a breeding program. Therefore, the need is urgent to expand the density and availability of DNA
markers, particular gene-based markers for the possibility of molecular breeding in date palm, and to survey the status of molecular characterization of traits of interest across date palm germplasm. A large number of EST sequences in date palm identified by Al-Dous et al. (2011) using de novo next-generation sequencing have provided a useful resource to develop gene-based markers. Zhao et al. (2013) characterized genic markers, EST-SSRs, in date palm, to evaluate and compare the frequency and distribution of various types of EST-SSRs in genic sequences, and to develop EST-SSR markers as genetic and genomic tools for date palm. In that study, 28,889 EST sequences from the date palm genome database, were analyzed to identify simple-sequence repeats (SSRs) and develop gene-based markers, i.e. expressed sequence tag-SSRs (EST-SSRs). About 4,600 ESTs as containing SSRs were identified, among which, trinucleotide motifs (69.7%) were the most common, followed by tetranucleotide (10.4%) and dinucleotide motifs (9.6%). The motif AG (85.7%) was most abundant in dinucleotides, while motifs AGG (26.8%), AAG (19.3%), and AGC (16.1%) were most common among trinucleotides. A total of 4,967 primer pairs were designed for EST-SSR markers derived from the computational data.

2.3.7.2. Chloroplast gene based markers

Generally, the chloroplast genome of plants shows maternal inheritance but in a few cases paternal or biparental mode of inheritance has been reported (Dong et al., 1992, Mason et al., 1994). In date palm, maternal inheritance has also been reported while hybridizing the species of genera Orbignya and Phoenix. Due to putative maternal inheritance of chloroplast DNA, it may generate erroneous results in cladistic studies (Al-Qurainy et al. 2011). The rapidly and slowly evolving loci including psbA-trnH and rpoB differentiate at genera and species level, respectively. To overcome this confusion due to considerable variation in DNA sequence alignments, the same two locus barcodes were used as alignment associated with the corresponding single-locus barcodes. Moreover, the main reason of choosing these two loci for date palm was that these exhibit considerable genetic variability and divergence, ease of amplification, short sequence length, conserved flanking sites for developing universal primers and ease of alignment and analysis (Kress et al., 2005, 2007). Al-Qurainy et al. (2011) assessed molecular signature of some of the economically and medicinally important date cultivars from Saudi Arabia based on rpoB and psbA-trnH chloroplast DNA sequences data. Eight different cultivars of dates 'Khodry', 'Khalas', 'Ruthana', 'Sukkari', 'Sefri', 'Segae', 'Ajwa' and 'Hilali' were sequenced for rpoB and psbA-trnH genes and analyzed using bioinformatic
tools to establish a cultivar-specific molecular signature. The analysis clearly reveals three major groups of these cultivars: (i) 'Khodary', 'Sefri', 'Ajwa', 'Ruthana' and 'Hilali' (58% BS); (ii) 'Sukkari' and 'Khalas' (64% BS); and (iii) 'Segae'. The economically most important cultivar 'Ajwa' showed similarity with 'Khodary' and 'Sefri' (67% BS). The sequences of the date cultivars generated in this study showed bootstrap values between 38% and 70%, so these sequences could be carefully used as a molecular signature for potential date cultivars under trading and selection of genuine cultivars at the seedling stage for farming.

3. Date palm genome sequencing

Even a comprehensive study of germplasm accession of all the different date palm cultivars growing in the different geographic locations using a universal set of RAPD, AFLP, ISSR and SSR loci were not suffice to answer the biodiversity questions. What is needed is a whole genome sequencing of all those indigenous cultivars from the different geographic locations of the species range in order to answer where the biodiversity spot is. The road to a whole genome sequencing projects start with a draft nuclear genome for the cultivar 'Khalas' female by Cornell Qatar (El-Dous et al., 2011), which results in a assembly of 56000 scaffolds representing 380 Mb in size, annotating 25000 genes of the genome. Another group in Saudi Arabia at the University of King Abdul Aziz for Science and Technology at the same time started a similar project and published a second draft genome (Al-Mssallem et al., 2013) for the same cultivar and assembled around 88000 scaffolds, which represent around 605 Mb in size and annotating 41660 genes. El-Dous et al. in 2011 sequenced other eight cultivars (5 females and 3 males) than Khalas and found 1605 Single Nucleotide Polymorphism (SNP) that segregated with gender and were useful in discriminating between male and female. The complete genome of chloroplast of date palm is sequenced and available (Yang et al., 2010). The draft nuclear and chloroplast genome of date palm is the first and important step in the genomic research of date palm. A complete genome assembly (18 chromosomes) of date palm will offer a great opportunity for further comparative genomics of other cultivated palms, where the genome is also available, oil palm (Singh et al., 2013). As a part of the road map mentioned above, mapping of important proteins involved in biochemical pathways for important traits is required. A couple of reports dealt with proteomes available for date palm (Gomez et al., 2009; Sghaier et al., 2009). For instance, plant defense or photosynthesis associated proteins are expressed during the invasion or colonization of endophytic, entomopathogenic fungi in date palm grown in field (Gomez et al.,
2009). A myosin heavy chain-like polypeptide will be expressed and accumulated during the injection in vitro with the same fungi. Another twenty-one proteins were accumulated during the development and germination of zygotic date palm embryos. Enzymes responsible for the glycolysis pathway, carbohydrate synthesis, storage and stress related proteins were found accumulating during the development and germination of date palm zygotic embryos (Sghaier et al., 2009). An understanding of such basic process will help to shed light on the importance of storage and stress related proteins including resistance to diseases and abiotic stress (salinity, heat, drought) or those involved in biochemical pathways at different developmental stages, treatments and in different tissues. For instance, the increased tolerance to biotic and abiotic stresses is a polygenic complex trait and thus it is hard to identify. The best approach is to scan the present genetic variation by screening different cultivars, wild and landraces. Forward genetics is very useful in this context, however the accurate and reliable phenotypic quantification methods are crucial in assisting the identification of these traits. Additionally, recent advances in high-throughput functional genomics such as next generation sequencing (NGS), transcriptomics, genomics, proteomics and epigenomics will enable researchers to use a system biology approach for dissecting biochemical pathways. Most functional studies have used the model system Arabidopsis or other systems and to date there is no full system biology study conducted on the genetically complex fruit tree date palm. To fully investigate that, a study is needed to find out the difference of the whole genome between good and poor date palm yield, cultivars as well as tissues resistant and susceptible to a disease. Therefore a whole genome association mapping experiment (GWAS) bridging “omics” will help understanding the different components as a unit involved in such processes. This would be essential to enhance the knowledge and later to develop appropriate markers for biotechnology application including genetic engineering in date palm improvement.

4. Molecular markers and genomic stability of in vitro-derived plantlets

Plant tissue culture is considered as a mean for vegetative propagation in which phenotypically and genetically identical clones are produced rapidly. However, somaclonal variation is known to result from changes in the nuclear, mitochondrial or chloroplast genomes of regenerants. Accordingly, detection of genetic stability of date palm plantlets is necessary to confirm genetic fidelity. Ali et al. (2007) used RAPD-PCR technique for detection of
genetic stability in regenerated plantlets in Barhee cv. Reproducible RAPD patterns were obtained using 30 primers. Three (OPC.16, OPG.08 and OPN.16) produced polymorphic bands in some of tested samples when compared with the DNA fingerprints of the mother offshoots suggesting the possibility of genetic variation among the resultant plants. Bader et al. (2007) used RAPD markers and tested 25 universal primers performed on DNA extracted from fresh leaves of the mother tree and samples randomly taken from plantlets derived from tissue culture for cvs. Barhee and Maktoom. Reproducible RAPD patterns were obtained with 20 primers; 17 primers showed completely monomorphic bands in all tested samples of the progeny (Fig. 9). Only 3 primers showed some polymorphic bands for both cultivars in some of the tested samples compared with the DNA banding pattern for the intact trees; these were OPD.01 primer for Barhee cv., and OPB.07 and OPC.08 for Maktoom.

![Figure 9](image)

**Figure 9.** (a) Monomorphic banding patterns of samples tested revealed by OPD.06 primer for Barhee cv., (b, c, and d) polymorphic banding patterns using the primers (OPB.07, OPC.08 for Maktoom and OPD.01 for Barhee cv.). Numbers on the left indicate the fragment size of molecular weight markers (lane M) in kb. The lanes 0 are the banding pattern of the intact trees, while the lanes 1 to 10 are the banding pattern of the samples selected randomly from tissue culture derived plantlets (Bader et al., 2007).
Meanwhile Khierallah et al. (2008) employed ALFP markers successfully to trace genetic fidelity of date palm in vitro plantlets for cvs. Barhee and Maktoom derived from inflorescence explants (Fig. 10). Somaclonal variation in plantlets which were induced by 2,4-D during recurrent somatic embryogenesis were tested by random amplified polymorphic DNA (RAPD) profiles (Othmani, 2009). Nine arbitrary 10-mer primers were used to amplify DNA extracted from 180 plantlets. RAPD patterns of the plantlets were identical with the original mother plant, indicating that 2,4-D did not cause somaclonal variation. In Egypt, Abd Alla (2010) utilized ISSR-PCR technique to assess the genetic stability of the micropropagated date palm plantlets var. Karama. The detected 49 amplicons from the five ISSR anchored primers showed that there were a high similarity within the micropropagated date palm plantlets and the

Figure 10. AFLP banding pattern of the two date palm cvs. Barhee and Maktoom as revealed by primer combinations P74/M95 and P104/M88 respectively. Numbers on the left indicate the fragment size of molecular weight markers (lane M) in base pairs (bp). The lanes 0 are the banding pattern of the mother trees, while the lanes 1 to 10 are the banding pattern of the samples selected randomly from direct organogenesis (Barhee) and indirect organogenesis (Maktoom) derived plantlets (Khierallah et al., 2008).
mother plant and some dissimilarity was detected among some samples. Genetic similarity between the mother plants of the cultivar namely 'Unknown' and 'Ferhi' and several plants regenerated in vitro from both cultivars were examined by RAPD analysis using 10 random primers (Moghaieb et al, 2011). The data indicate that the regenerated plants from the 'Unknown' and 'Ferhi' cultivars showed 36.2 and 37.8 % polymorphism and sharing 63.8 and 62.2% of similarities, respectively with their mother plants. Talaat et al. (2012) used ISSR markers to detect genetic variation between the regenerated plantlets and their corresponding callus. Somaclonal variation between callus and more than one regenerated plant was observed in three cultivars: 'Kubkub', 'Merziban' and 'Dharlag'. Recently, Khierallah and Hussien (2013) generated RAPD pattern with 12 primers to identify genetic fidelity of date palm in vitro plantlets produced using Picloram instead of 2,4-D for callus induction and some natural organic extracts (coconut water or casein hydrolysate) for somatic embryogenesis. Genetic variations may occur in plantlets derived from callus proliferated from shoot tips as well as in those derived from inflorescences. RAPD and AFLP appear to be an efficient technique, simple and fast DNA marker for the early detection of genetic variations in plants propagated by tissue culture.

5. Date palm breeding and development

Date palm is a dioecious species having male and female individuals on separate trees. During fertilization, recombination and other linkage disequilibrium events occur, leading to coupling and repulsion arrangements between loci that can be either syntenic or non-syntenic. Syntheny represents the way loci physically co-localize on the same chromosome of a given cultivar. It is related to the genetic linkage between loci, representing lower-than-expected recombination frequencies in panmictic populations. Theoretically, all linked loci are inevitably syntenic, but not all syntenic loci are linked. In genomics for instance, the genetic loci on a given chromosome are syntenic regardless of how they are brought together (i.e. sequencing and assembly, physical localization or mapping) (El Hadrami et al., 2011). Syntheny is an important notion in breeding because it allows prediction of the co-localization of loci on chromosomes from related species used in crossing. Genetic events such as translocation that occurs during genome re-arrangements may result in a loss of syntheny among syntenic loci or in a gain of syntheny between nonsyntenic ones. When syntheny exceeds expectancy, a selection is possible for functional relationships between syntenic genes. This is often explored to increase inheritance of alleles that would provide a substantial advantage when
inherited in association rather than individually (i.e. improving the yield) or when the two alleles share the same regulatory mechanisms (El Hadrami et al., 2011). The inheritance in date palm in not fully understood so far due to the unavailability of enough segregate populations with sufficient time-depth following their establishment by a series of crosses and backcrosses. Salem et al. (2001) reported on the genetic inheritance of a selected set of isozymes in four date palm cultivars and described five polymorphic loci involved with 12 alleles. The genetic differentiation and structure of date palm is governed by many factors. Three of these appear to greatly affect the genetic differentiation of date palm populations and/or cultivar groups. These factors include (i) the geographic isolation of the species, (ii) the long biological life-cycle of the trees and (iii) the impact of the regional environment, where the plant thrives, especially with respect to aridity that often affects the organoleptic, textural and nutritional characteristics of the dates. Cultural practices are very difficult parameter to trace and judge but have certainly played an important role in making the current genetic structure of date palm. This also includes offshoots trade or exchange that occurred among regions or countries. In addition, it is noteworthy to point out that growers’ practices regarding the choice of well appreciated and/or adapted cultivars, along with male pollinators for their date palm trees, may have contributed to the genetic structure of date palms on a global scale. A few studies have already shown using highly repetitive microsatellite markers, those date palm populations are structured into groups on the basis of their geographic location and the type of fruits they produce (Zehdi et al., 2002, Elshibli and Korpelainen, 2008, 2009).

5.1. Genetic map of date palm

The date palm draft genome released by Al-Dous et al. (2009) has recently led to the establishment of the first genetic map for Khalas cultivar by the scientists of Weill Cornell Medical College in Qatar (Mathew et al., 2014). Based on a modified genotyping-by-sequencing approach (GBS), they presented the first genetic map for date palm and identify the putative date palm sex chromosome. About 4000 markers were placed on the map using nearly 1200 framework markers spanning a total of 1293 cM. Also they integrated the genetic map, derived from the Khalas cultivar, with the draft genome and placed up to 19% of the draft genome sequence scaffolds onto linkage groups for the first time. This analysis revealed approximately ~1.9 cM/Mb on the map. Comparison of the date palm linkage groups revealed significant long-range synteny to oil palm. Analysis of the date palm sex-determination region suggests telomeric position on the linkage
group 12 and recombination is not suppressed in the full chromosome. Combined with the recent draft genome sequence of the same cultivar, this resource offers a critical new tool for date palm biotechnology, palm comparative genomics and a better understanding of sex chromosome development in the palms.

5.2. Breeding for resistance to bayoud disease

Date palm is a host to a number of diseases and pests, some of which have a destructive economic impact and threaten the sustainability of the crop. Others are still not well characterized and little is known about their interaction with date palm (El Hadrami et al., 2011). Breeding of resistant cultivars through biotechnology is also still challenging due to the peculiarities of the species. However, efforts have been made, over the years, in terms of studying plant x pathogen interactions and preparing the ground for future breeding for resistance. Several clones with interesting traits have been obtained through targeted and nontargeted genetic crosses (El Hadrami et al., 1998, El Hadrami and El Hadrami, 2009). However, most of the attempts to establish date palm crosses and progenies with a good level of resistance to pathogens and pests have failed or been abandoned for several reasons. First, often introducing resistance is dependent on maintaining date fruit quality. Second, a high number of backcrosses to stabilize the selected resistant traits and to recurrently restore fruit quality is needed, which is incompatible with the lengthy life-cycle of the species. Third, sex determination is not possible until the plant reaches maturity (5–8 years), reducing the final number of producing trees at the end of a selection program. Last, continuity of research programs on this horticultural crop is poorly funded and often pursued on restricted budgets (El Hadrami et al., 2011).

Bayoud is a devastating vascular disease in date palm that originated in Morocco back in the 1870s. It is a fusariosis caused by *Fusarium oxysporum* f. sp. *Albedinis* that led to the near extinction of the highly-valued cv. Medjool in North Africa. Most breeding progress for resistance against this disease has been conducted in the country of origin, Morocco, and to some extent in Algeria and Tunisia. Sedra (2011) tested ISSR and RAPD markers with aim to identify markers related to response to bayoud disease of date palm cultivars. In this respect, at least seven resistant and seven susceptible cultivars (Table 5) and numerous susceptible and resistant young plant-hybrids derived from controlled crosses Black Bousthammi (resistant female) × INRA-A18 (susceptible male) and Jihel (susceptible female) × male INRA-NP4-Boufegopus (resistant male) were used.
Table 5. Moroccan date palm cultivars used in investigation of RAPD and ISSR markers associated with resistance to bayoud disease (Sedra 2011).

<table>
<thead>
<tr>
<th>Cultivars</th>
<th>Main geographical area</th>
<th>Fruit quality</th>
<th>Phenotype to bayoud disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boufeggous ou Moussa Black</td>
<td>Bani, Draa valley, Tafilalet, Sagho, Anti-Atlas</td>
<td>Fair</td>
<td>Resistant</td>
</tr>
<tr>
<td>Bousthammi</td>
<td>Draa valley</td>
<td>Moderate</td>
<td>Resistant</td>
</tr>
<tr>
<td>Boukhanni</td>
<td>Anti-Atlas, Bani, Draa valley, Sagho</td>
<td>Fair</td>
<td>Resistant</td>
</tr>
<tr>
<td>Iklane</td>
<td>Anti-Atlas, Bani, Draa valley, Sagho</td>
<td>Fair</td>
<td>Resistant</td>
</tr>
<tr>
<td>Najda</td>
<td>Draa valley</td>
<td>Good</td>
<td>Resistant</td>
</tr>
<tr>
<td>Sairlayalate</td>
<td>Bani</td>
<td>Moderate</td>
<td>Resistant</td>
</tr>
<tr>
<td>Tademainte</td>
<td>Anti-Atlas, Bani, Draa valley, Oriental, Sagho</td>
<td>Fair</td>
<td>Resistant</td>
</tr>
<tr>
<td>White Bousthammi Ahardane</td>
<td>Anti-Atlas, Bani</td>
<td>Fair</td>
<td>Susceptible</td>
</tr>
<tr>
<td>Boufeggous</td>
<td>Draa valley, Tafilalet, Ziz, Anti-Atlas, Sagho, Bani, Ferkala, Gheris, Guir, Todra, Oriental, between Sagho and High-Atlas</td>
<td>Good</td>
<td>Susceptible</td>
</tr>
<tr>
<td>Bourar</td>
<td>Draa valley, Bani, Sagho, Tafilalet</td>
<td>Fair</td>
<td>Susceptible</td>
</tr>
<tr>
<td>Bouskri</td>
<td>Bani, Draa valley, Sagho, Todra, Oriental, Tafilalet, between Sagho and High-Atlas, Anti-Atlas</td>
<td>Moderate</td>
<td>Susceptible</td>
</tr>
<tr>
<td>Deglet Noor</td>
<td>Oriental</td>
<td>Good</td>
<td>Susceptible</td>
</tr>
<tr>
<td>Jihel</td>
<td>Draa valley, Bani, Anti-Atlas, Tafilalet, Sagho, between Sagho and High-Atlas</td>
<td>Good</td>
<td>Susceptible</td>
</tr>
<tr>
<td>Medjool</td>
<td>Tafilalet, Ziz valley, Oriental, Draa valley, Sagho, between Sagho and High-Atlas</td>
<td>Excellent</td>
<td>Susceptible</td>
</tr>
</tbody>
</table>

The RAPD analysis and PCR were achieved according to the protocol used by Sedra et al. (1998) with some modifications. The RAPD marker UBC-145-1.22 was present in five out of seven resistant cultivars (Table 6). The RAPD-UBC-578-1.50 was present in five out of six resistant cultivars. The ISSR marker Mic19-1.37 was detected in all six resistant cultivars. Other markers were only detected in one or a few resistant cultivars. It was
confirmed that these markers have been revealed in the majority of resistant hybrids (young plantlets) derived from crossing a resistant parent with a susceptible one, which proved that they can be transmitted to the progeny. The presence of multiple candidate markers suggests that the resistance could be encoded by different genes (Sedra, 2011).

**Table 6.** RAPD and ISSR candidate markers for detection of the resistance to bayoud disease in the date palm (Sedra 2011).

<table>
<thead>
<tr>
<th>Primer</th>
<th>Total number of bands</th>
<th>Size of marker candidates (Kb)</th>
<th>No. markers detected/resistant cultivars</th>
<th>Cultivars</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAPD-OP-D16</td>
<td>6</td>
<td>1.06</td>
<td>3/7</td>
<td>Black Bousthammi, Iklane, Tadmainte</td>
</tr>
<tr>
<td>RAPD-OP-D19</td>
<td>6</td>
<td>0.10</td>
<td>4/7</td>
<td>Black Bousthammi, White Bousthammi, Boukhani, Tadmainte</td>
</tr>
<tr>
<td>RAPD-UBC-145</td>
<td>5</td>
<td>1.22</td>
<td>5/7</td>
<td>Black Bousthammi, White Bousthammi, Boufeggous ou Moussa, Boukhani, Tadmainte</td>
</tr>
<tr>
<td>RAPD-UBC-578</td>
<td>4</td>
<td>1.50</td>
<td>5/6</td>
<td>Black Bousthammi, White Bousthammi, Iklane, Sairlayalate, Tadmainte</td>
</tr>
<tr>
<td>RAPD-UBC-594</td>
<td>14</td>
<td>0.64</td>
<td>3/6</td>
<td>Black Bousthammi, White Bousthammi, Iklane, Sairlayalate, Tadmainte</td>
</tr>
<tr>
<td>ISSR-Mic 19</td>
<td>11</td>
<td>1.37</td>
<td>6/6</td>
<td>Black Bousthammi, White Bousthammi, Iklane, Sairlayalate, Tadmainte</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.01</td>
<td>2/6</td>
<td>Black Bousthammi, Iklane</td>
</tr>
</tbody>
</table>

RAPD sequences have been used to construct a random genomic library of Tunisian date palm cultivars (Hela et al., 2000). The recombinant DNA included fragments ranging from 200 to 1,600 bp inserts. This library could potentially be used to anonymously probe southern hybridizations and in the development of markers associated with bayoud resistance and sex determination.

### 5.3. Marker-assisted selection

Marker assisted selection (MAS) is breeding approach that combines traditional genetics and molecular biology. MAS allows the selection of
genes that control traits of interest. Combined with traditional selection techniques, MAS has become a valuable tool in selecting organisms for traits of interest, such as color, fruit quality, or disease resistance. Few studies have been published on MAS employment in date palm breeding and improvement. El- Kharbotly et al. (1998) reported the use of AFLP markers to study two F1 and F4 backcrossed populations to tentatively map traits related to the fruit and seed weights, seed volume and the total soluble solids (TSS), as well as other measured morphological characters. In that study, AFLP markers on parental clones and tested 35 primer combinations were used to detect polymorphic markers for further construction of a linkage map.

6. Determination of date palm sex using molecular markers

The date palm is a dioecious plant whose sex cannot be determined until it reaches a reproductive stage with an age between 5 and 8 years. An early selection and differentiation of young seedlings into males and females could enhance breeding and assist research programs for genetic improvements. In dioecious plants where sex chromosomes have not been identified, markers for maleness indicate either the presence of sex chromosomes which have not been distinguished by cytological methods or that the marker is tightly linked to a gene involved in sex determination. Numbers of research groups have used subtraction techniques of either cDNA (Differential screening of a subtracted cDNA library) or genomic DNA in attempts to isolate sex determining genes from Silene latifolia (Scutt et al. 1997; Scutt and Gilmartin 1998). There are various molecular markers linked to sex, such as RAPD, RFLP, AFLP include Randomly Amplified Polymorphic DNA (RAPDs), Restriction Fragment Length Polymorphisms (RFLPs), Amplified Fragment Length Polymorphism (AFLPs) and microsatellites (Bekheet and Hanafy 2011). In this area, ALFP has been used to initiate a genetic map of date palm (El-Kharbotly et al., 1998). Five cultivars: Bunarenga, Damous, Fardh, Khalas and Khenizi and pollen of male palm coded: BN-96, DN-96, Fr-96, KL-96 and KN-96 amended by the cultivars from Rumais Agriculture Research Station, Sultanate of Oman, were used to establish F1 populations. Data taken from 3-week old seedlings showed that the populations were segregating with 1:0, 1:1 or 3:1 ratios for erect and slanting, respectively. The erect and slanting leaf was controlled with a simple genetic factor following Mendelian inheritance. The AFLP fingerprinting for both parental clones (mentioned above) was obtained with 32 primer combinations. Some primer combinations (E-AAC, M-CAG; E-AAG, M-CAT and E-ACG, M-CAA) showed few unique bands linked
with males while few unique bands were observed in the females using the combinations of E-ACG, M-CAA. The polymorphism between the two parents ranged from 4 to 55% depending on the primer combination. This study showed the possibility of using the AFLP technique to characterize the date palm genome. In an attempt to determine the genetic difference between male and female date palms, genome DNA was extracted from leaves of four female cultivars 'Deglet Noor', 'Allig', 'Kentich', 'Menakher', a male genotype pollinator T23 and F1 hybrid. The results of RAPD gave reproducible polymorphic bands with 11 primers out of 53 primers used. Accordingly, RAPD was successfully used to differentiate between female cultivars, male and F1 hybrid (Ben-Abdallah et al., 2000). Soliman et al. (2003) used RAPD technique to compare genetic material from four female date palms and four unknown male Egyptian trees. The banding profiles obtained suggested that two male clones are genetically related to the four female date palm cultivars 'Zaghloul', 'Amhat', 'Samany' and 'Siwi' ranged from 87.5 to 98.9%. In this respect, identification of some Egyptian date palm males from female cultivars using molecular markers was reported (Ahmed et al., 2006). Genomic DNA and RNA were extracted, measured and used as a template to detect genetic relationship and similarities among four known females 'Sakkoty', 'Malkabi', 'Bartamoda' and 'Dagana' and three unknown males of Egyptian date palm based on DNA and RNA technology. Results showed that differential display and RAPD analysis provided a rapid and effective method to detect the genetic relationship and similarities between four males and females of Egyptian date palms. For early identification of cultivars and tracing genetic diversity among date palm genotypes of different origin, offshoot-derived, male and female plants of cultivars 'Barhee' and 'Sukkary', seed-derived plants, and two in vitro cultures of both of these cultivars were subjected to RAPD analysis (Al-Khalifa et al., 2006). Similarity matrixes show that offshoot-derived male plant of 'Barhee' was 73.6% genetically similar to its female counterpart, while similarity between male and female plants of 'Sukkary' was 43.1%. In the case of seedlings, male and female plants of Barhee were 87.2% similar and those of Sukkary were 62.3% genetically alike. Sexual embryos of date palm were in vitro cultured and molecular analysis was used for early identification of sex type (Bekheet et al., 2008). In that study, the potential of isozymes and RAPD markers in sex identification of in vivo grown and in vitro differentiated cultures of date palm was investigated. In vitro zygotic lines were proliferated from mature and immature zygotic embryos of date palm. Early estimation of sex type of in vitro differentiated lines has been realized via the activity levels of two enzymes. A high level of peroxidase activity was observed in adult and offshoot females. Acid phosphatase and glutamate oxaloacetate enzymes
gave a strong difference between male and female date palms. Otherwise, the RAPD technique was used to compare genetic material from male, female and unknown lines of date palm. RAPD analysis showed a relatively close relation between the two females (adult and offshoot) cultures, since they have large number of homologous bands. Although, there was a low relationship between male and female, results of similarity could not confirm a link to sex or estimate the sex type of unknown clones. Moreover, an attempt to identify sex-specific DNA markers for date palm using molecular technique (RAPD and ISSR) was achieved by Younis et al. (2008). Four dry date cultivars 'Sakoty', 'Bertmoda', 'Malkabi', 'Dagana' and three males 'Dagana', 'Malkabi', 'Sakoty' recognized as superior date pollinators were used in the study. RAPD analyses gave three positive specific markers for females and two for males in addition to five positive specific markers for males in ISSR analysis. Al-Mahmoud et al. (2012) identified regions in the date palm genome that were linked to gender in seven males and seven females collected from Qatar and from California, USA.

Investigation of these regions revealed that the date palm employs a XX/XY sex determination system with the male being the heterogametic sex. The regions also showed significant polymorphism between the male and female alleles. This polymorphism can be used in the development of assays to distinguish the two sexes at an early stage. The authors employed two approaches in the development of DNA-based assays for sex differentiation in date palm. The first was PCR-based restriction fragment length polymorphism (PCR-RFLP) approaches that require amplification followed by restriction digestion and gel electrophoresis. In the second approach, they attempted to design a PCR-only method that would take advantage of the high heterogeneity in the sex-linked region to remove the need for the restriction digestion step. PCR primers were designed to span multiple polymorphisms (Fig. 11) that were unique to either the male or female haplotype in the hope that would allow specific amplification of each haplotype. The male is heterozygous containing both the male and female alleles and should yield two distinct PCR products. The female is homozygous, which should yield a single band representing both copies of the female allele. By designing primers on sex-linked polymorphisms, it is possible to simplify the process. Meanwhile, Elmeer and Mattat (2012) identified a promising candidate microsatellite marker that could be used to differentiate between male and female date palm. 14 microsatellite primer pairs with 129 date palm leaves and tissue culture samples from 34 cultivars which represent the major date palm diversity of Qatar were used. About 254 microsatellite loci were detected, of these, 22 microsatellite loci could be used to identify 9 out of 12 male date palm samples (75%). The data also
indicated that the heterozygous allele with the size 160/190 produced by the primer mPdCIR048 reoccurred 4 times exclusively in the 12 individual male samples but not in any of the 117 female date palm samples tested.

7. Conclusion and prospective

Preliminary studies have demonstrated that population structures and lineage relationships can be identified with different types of DNA markers. As noted, the availability of the complete genome sequence will facilitate the development for suiting of different marker types to be applied appropriately. The development of a series of sequenced tagged sites will supply resources needed for the screening of collections to reduce the number of samples kept in germplasm banks. They will also add impetus to identifying markers linked to the various disease-resistant genes. With the steady increase in the sequencing resources, SNPs will also become more useful but the relative costs of SNP and SSR analyses may well determine which of the two-marker systems become most widely used. Prospects for the application of molecular markers to date palms are still very minimal. The availability of the shotgun sequence and the steady lowering of the costs of high throughput sequencing will increase the resources and their application rapidly over the next few years. As with many plant species, decisions have to be made concerning the level of whole genome sequencing compared to target re-sequencing as the most efficient method for useful applications.

Acknowledgment

I gratefully acknowledge and thank Prof. Dr. Kadhim M. Ibrahim, Dept. of Biotechnology, College of Science, Al-Nahrain University, Iraq and
Dr. Khaled Hazzouri, Center for Genomics and Systems Biology, NYU Abu Dhabi, UAE, for their assistance.

References

Applications of molecular markers in date palm genome analysis and breeding


40. Cullis C. Molecular Markers in Date Palm chapter 18 In: Date Palm Using Molecular Markers chapter 20 S.M. Jain et al. (eds.), Date Palm Biotechnology, Springer p 407-426, 2011.


69. Ibrahim A.O. Date Palm, the Tree of Life. Arab Centre for the Study of Arid Zones and Dry Lands. Damascus, Syria, 2008.
Applications of molecular markers in date palm genome analysis and breeding

77. Khierallah H.S.M., Bader S.M., Baum M., Choumane W. Assessment of AFLP variations in date palm in vitro plantlets derived from inflorescence explant. Proceeding of the 4th Symposium of date palm, KSA, 2008, 552-564.


