Chapter 32

Date Palm

☆ *Arvind K. Yadav, Rhitu Rai, Prasanta K. Dash and S.V. Ramesh*

1. Introduction

Date palm (*Phoenix dactylifera* L.) is a multipurpose tree that provides fibre, carbohydrates, minerals and vitamins besides being source of an array of medicinal compounds. The tree is a long-living monocotyledonous plant belonging to the family of Arecaceae. Cultivated for their edible sweet fruit. Because of their long cultivation history their place of origin is unknown but it is speculated to have originated from lands around Iraq (Morton, 1987). However, the plant is widely cultivated and is naturalized in many tropical and subtropical regions worldwide. The date palm trees typically reach about 75-80 feet (23-25 meter) height and the leaves are approximately 20 feet (6 meter) long. The full span of the crown of plants ranges from 6–10 meter (Robinson *et al.*, 2012). It is one of the oldest cultivated plants in the world and is the most important subsistence crop in Northern Africa and the Middle East countries. Nearly 2000 cultivars of date palm are known in the world. Among them, only genotypes some have been evaluated for their performance and fruit quality (Parvin *et al.*, 2015). Because of its high nutritional value and long-life the date palm has been mentioned as the 'tree of life' (Augstburger *et al.*, 2002). Date pulp hold easily digestible sugars (70%), mostly glucose, sucrose and fructose; dietary fibres and less proteins and fats. It also comprises vitamins like riboflavin, biotin, thiamine, ascorbic and folic acid that are essential for the body. The pulp of the fruit is rich in calcium, iron, copper, cobalt, magnesium, fluorine, manganese, phosphorus, potassium, copper, sodium, boron, sulfur, zinc and selenium (Al-Farsi and Lee, 2008). Dates are rich in nutrients and are a good source of rapid energy due to high carbohydrate content (70–80%). Biochemically, date fruits contain fat (0.20–0.50%), protein (2.30–5.60%), dietary fibre (6.40–11.50%), minerals (0.10–916 mg/100 g dry weight) and vitamins with very little or no starch (Al-Shahib and Marshal, 2003). The fruit is also a good source of important phytochemicals, including carotenoids, phenolics and flavonoids. Date fruits not only provide antioxidant, anti-mutagenic and immune-modulatory benefits to health but also have diverse medicinal values, including anti-hyperlipidemic, anticancer, gastroprotective, hepatoprotective and nephroprotective properties (Tang *et al.*, 2013).

2. Genetic Map of Date Palm

Date palm is a diploid crop with estimated size of 685Mb. The exact number of date palm chromosomes has been uncertain with some publications reporting 14, 16, or 18 pairs with most evidence pointing to 18 pairs (Al-Salih and Al-Rawi, 1987). Despite its importance from antiquity, few genetic resources are available for improving the productivity and development of the dioecious date palm. Analysis of the genetic maps combined with genomic data will assist in the improvements of this commercially important palm. In order to overcome the challenges Mathew *et al.* (2014), presented the first genetic map of date palm and identified the putative date palm sex chromosome based on a modified gentoyping-by-sequencing (GBS) approach. The total genetic map length was 1293 cM. Assuming a genome size of approximately 670 Mb, genetic to physical distance conversion in date palm is approximately 1.93 cM/Mb of sequence or 520 kb/cM. Inter-marker distances were evenly distributed though some biases were observed. The largest gap between any two markers was 9.7 cM on linkage group (LG) 6. Both LG5 and LG10 had an excess of gaps greater than three cM with a total of nine and seven respectively. Distribution of marker types among the linkage groups showed pronounced marker type bias in LG5, 6, 9 and 10. Analysis of the date palm sex-determination region suggests it is telomeric on linkage group 12 and recombination is not suppressed in the full chromosome.

3. Date Palm Genome

A major initiative in the genomics of date palm was published by Zhang *et al.*, 2012, who generated large scale gene sequences from 14 cDNA libraries representing various tissues derived from the cultivar Khalas. Large scale annotation of gene sequences based on plant database and functional domain annotation based on Pfam database were performed (Zhang *et al.*, 2012). Interestingly, comparative analysis revealed that 70.6 %, 69.4 %, 68.4 % and 69.3 % of date palm gene models were shown to be shared with rice, sorghum (monocots), Arabidopsis and grapevine (dicots) respectively.

The genome sequence of date palm has been made available by sequencing an elite cultivar Khalas (Al-Mssallem *et al.*, 2013). The size of genome is 605.4 Mb and it covers >90% of the genome (~671 Mb) and >96% of its genes (~41,660 genes). The sequencing effort yielded 41,660 gene models (42,957 isoforms) in 10,363 scaffolds (472,329,057 bp in length; 84.6% of the total length). Proteome comparison of date palm to Arabidopsis, Rice, Sorghum and Grape revealed that 8,093 gene families are shared among all five plant genomes and 1,127 gene families are unique to date palm. These unique gene families are mostly related to DNA/RNA metabolic processes and ion binding. In date palm the abundant biologically-defined repeats that accounted for 21.99% of the genome, of which 14.03% and 4.17% are Ty1/Copia and Ty3/Gypsy, respectively. The genome harbours 38.41% repetitive sequences and was observed that Ty1/Copia elements have a much higher copy number than Ty3/Gypsy elements. The most abundant Ty1/Copia elements in *P. dactylifera* show the highest homology to the rice retrotransposon element 1 in the conserved region of the reverse transcriptase genes.

Genome-wide duplication (GWD) provides essential genetic material for the creation of novel functions for adapting new environment and tolerating biotic and abiotic stresses (Freeling, 2009). Using 4,215 paralogous gene pairs in 411 collinear regions of the date palm genome assembly, the distribution of Ks or 4DTv was assessed and it showed two distinct peaks: Ks ~0.314 (4DTv~0.107) and Ks ~0.833 (4DTv~0.332). The first peak corresponds to a GWD event shared among all angiosperms and the second peak derives from either a single more ancient GWD or massive consecutive segmental duplications when the slow substitution rate of the palm family. Significant macro-synteny between date palm and other monocotyledons were also found but macro-synteny is non-existent between date palm and any dicotyledons. The biggest scaffold, pdS00001 (~4.5 Mb in length), appears highly conserved and is part of the 'concentric circles' of monocotyledons.

Whole genome re-sequencing of 62 cultivars of data palm has paved way for generation of a comprehensive registry of approximately seven million single nucleotide polymorphisms (SNPs) that has a greater potential for date palm crop improvement programmes. Population and genetic diversity analysis of date palm cultivars using those SNPs again confirmed a genetic dichotomy between North Africa and the Middle East/South Asian date palms. Furthermore, selection pressure analysis among the population of date palms based on SNP genotyping revealed that geographic adaptation of these cultivars was correlated to preferential selection in few genomic regions. Thus, the analysis is a comprehensive genomics resource for this important crop that captures diversity of date palm post domestication (Hazzouri *et al.*, 2015).

Genotyping-by-sequencing (GBS) is a potential strategy to rapidly describe genetic composition of plant cultivars. A collection of 13,000-65,000 SNPs were utilized in genotyping 70 date palm cultivars using leaf and fruit samples to gain knowledge on the origin of date palm cultivation. SNP based genotyping also provided further credence to the earlier findings that North African and Arabian Gulf regions are the two main earliest centres of date palm domestication. Furthermore, the study found genomic regions of date palm that display high degree of geographic segregation and higher proportion of allele fixation in sex chromosomes (Mathew *et al.*, 2014).

4. Chloroplast Genome

A complete sequence of the date palm chloroplast (cp) genome based on pyrosequencing was reported by Yang *et al.*, (2010). The date palm chloroplast genome is 158,462 bp in length and has a typical circular double-stranded DNA molecule. It shares a common quadripartite structure with the vast majority of other angiosperms: a pair of inverted repeats (IRs, 27,276 bp) separated by the large single-copy (LSC, 86,198 bp) and small single-copy (SSC, 17,712 bp) regions. It encodes 131 predicted functional genes; 112 are unique and 19 are duplicated in the IR regions. Among the 112 unique genes, 79 protein-coding, 29 transfer RNA and 4 ribosomal RNA genes were identified. 50.93%, 1.79%, and 5.71% of the genome sequence encode proteins, tRNAs and rRNAs, respectively, whereas the remaining 41.57% are non coding and filled with introns, intergenic spacers, or pseudogenes. The date palm chloroplast genome has 18 intron-containing genes among the 112 unique genes. Almost all are single-intron except two genes, *ycf3* and *clpP*, whose exons are separated by two introns. A total of 22,950 codons represent the coding capacity of all protein-coding genes of date palm chloroplast genome. Among these codons, 2001 (8.72%) encode for isoleucine and 271 (1.18%) for cysteine, which were the most and the least amino acids, respectively. Similar to other chloroplast genomes (Raubeson *et al.*, 2007), the date palm cp genome is also AT-rich (62.77%), and the values vary slightly among defined sequences of non-coding, protein-coding, tRNA and rRNA, where their A+T contents are 66.60%, 61.03%, 57.94% and 52.19%, respectively. 78 SNPs were identified as major intra-varietal polymorphisms within the population of a specific chloroplast genome, most of which were located in genes with vital functions. Based on RNA-sequencing data, 18 polycistronic transcription units and three highly expression-biased genes—*atpF*, *trnA-UGC* and *rrn23* were also found (Yang *et al.*, 2010) in date palm.

5. Mitochondrial Genome

Based on next-generation sequencing strategy,data from pyrosequencing and ligation-based sequencing, date palm mitochondrial genome (cultivar Khalas, Al-Hasa Oasis, Saudi Arabia) was deciphered (Fang *et al.*, 2012). Analysis of the mitochondrial genome sequence and transcriptomic data are of importance in revealing mechanisms underlying mitochondrial genome evolution and the unique evolutionary status of date palm among angiosperms. The mitochondrial genome chromosome is made of 715,001 bp circular molecule with an average GC content of 45.1%. Date palm mitochondrial genome represents the fourth largest mitochondrial genome sequenced after cucumber with 1,555,935 bp, melon with 982,833 bp and grape with 773,279 bp. The complete sequence of the date palm mitochondrial genome for analysis is available at GenBank (accession number JN375330). Its protein coding sequence is composed of only 6.5% of the mitochondrial genome (46,770 bp) and this gene content is similar to other angiosperm genomes. The mitochondrial genome contains at least 38 protein-coding genes, 30 tRNAs, three ribosomal RNAs and five complete ORFs. Most of these genes encode proteins of the electron transport chain. The majority (93.5%) of the genome sequence is comprised of chloroplast-derived (10.3% with respect to the whole genome length) sequences and are non-coding which harbours 0.33% tandem and 2.3% long repeats. This is the second highest proportion (10.3%) of chloroplast-derived sequences among the sequenced mitochondrial genomes to date, of which several intact genes, such as petA, petG, petL, psaJ, psbT, rpl20, rpl33 and rps8 are identified (Fang *et al.*, 2012).

Chloroplast and mitochondrial genomes are known to share sequences due to frequent gene transfer events (Stern and Lonsdale, 1982). Frequent DNA transfer from chloroplast DNA to mitochondrial DNA occurs as far back as the common ancestor of the extant gymnosperms and angiosperms, about 300 MYA (million-years-ago) (Wang *et al.* 2007). Fang *et al.* (2012) reported that mitochondrial genome of date palm contains more than 100 fragments of chloroplast origin (over 80% identity) ranges from 50 to 6,521 bp in length. The total fraction of chloroplast DNA sequences of date palm present in its genome is 73,691 bp, corresponding to 10.3% of the whole mitochondrial genome, and 46.5% of date palm chloroplast genome. These finding suggest that chloroplast DNA sequence insertion is an important mechanism for plant mitochondrial genome size expansion and sequence diversity.

6. Transcriptomics

Transcriptome refers to the complete set of all RNA (coding and non-coding), that are transcribed from the genome. Transcriptomics studies using next generation sequencing technologies has widely been adapted to unravel gene expression mechanism in many crops. In date palm transcriptional changes during development of fruit has been studied. Using RNA-Seq data 4,134 differentially expressed genes (DEGs) were identified in date palm whose expressions significantly vary among seven fruit developmental stages. When DEGs were clustered into different groups (up-regulated, down-regulated and not-regulated) it was found that different enrichments of DEGs such as gluconeogenesis, cellular carbohydrate metabolism and small molecule biosynthesis were up-regulated and biological regulation, transcription and regulation of RNA metabolic process were found in the down-regulated group. This large-scale genomic data is a basis for further genomic studies not only on date palm but also in other *Arecaceae* plants (Xin *et al.*, 2015).

Among the non-coding RNAs, microRNAs (miRNAs) are class of small RNAs that play major role in gene regulation thereby decide the growth and development of an organism. Expression profiling of date palm identified 276 novel fruit-development associated (FDA) miRNAs and their targets. Expectedly, most of the miRNA target genes were encoding transcripts involved in starch/ sucrose metabolism (Xin *et al.*, 2015). Deep sequencing of leaves and roots of date palm treated with NaCl stress identified majority of conserved miRNA homologs (153) and novel miRNAs (180). Among the salt stress responsive miRNAs it was demonstrated that 54 and 25 miRNAs derived from leaves and roots respectively were upregulated. The target transcripts for miRNAs were identified to be potassium channel AKT2-like proteins, vacuolar protein sorting-associated protein, calcium-dependent and mitogen-activated proteins (Yaish and Kumar, 2015).

7. Proteomics

Date palm (cultivar 'Barhi') proteome was studied to understand the fruit development biology at molecular level and provide the first detailed comparative proteome of the date fruit (Marondedze *et al.*, 2014). Altogether 189 significantly altered proteins were observed of which 171 were positively identified by using LC MS/MS. Of the 193 identified unique proteins that were significantly differentially expressed 82 proteins and 96 proteins were up or down regulated at different times. Further 36 proteins showed differential accumulation throughout development. It was noted that 29 proteins were specific to date fruit and belong to ten functional categories and these proteins have not been identified in any other fleshy fruits (Bevan *et al.*, 1998). 64 differentially expressed date proteins showed contrasting expression patterns in other fruits. Additionally, five proteins, proline iminopeptidase, aspartyl tRNA synthetase, GDP dissociation inhibitor, disproportionating enzyme and ornithine carbamoyl transferase have been detected in other fleshy fruits, but not identified in this fruit. The identified proteins were classified into 14 functional categories as, 'disease and defense' (16.5%), 'metabolism' (15.4%), 'unclassified' (15.1%), 'protein destination and storage' (10.7%), 'energy' (9.9%), 'cellular structure' (7.0%), 'secondary metabolism' (5.5%), 'signal transduction' (5.1%), 'protein synthesis' (5.1%), 'unclear classification' (2.6%), 'transporters' (2.6%), 'transcription' (2.2%), 'cell growth/division' (1.8%) and 'intracellular traffic' (0.4%). In response to abiotic stress, some proteins have dual roles. In this category, nine protein were identified as date hypanthium-specific, 29 proteins showed contrasting accumulation patterns and seven had similar patterns. Two date-specific proteins involved in protein synthesis, group antigen polymerase (Gag-pol) polyprotein (spot 249) and aspartyl tRNA synthetase were upregulated during ripening. Further, five stress-responsive proteins classified in the category 'protein destination and storage' showed differential accumulation during development. Three proteins were detected as upregulated and two other proteins decreased at all stages. Three heat shock proteins (HSP82) were identified as down regulated at NTR and RIPE, and these were date response specific proteins. Other stress-related proteins were detected as differentially expressed and included universal-stress protein (USP) and E3 ubiquitin ligase. USPs are involved in ethylene-mediated stress adaptation (Sauter *et al.*, 2002) and Ubiquitination plays a crucial role in abiotic and biotic stress responses (Mazzucotelli *et al.*, 2006). A total of 20 proteins involved in energy generation were identified. This proteomics analysis provides insights into physiological processes during date fruit development and ripening, and offers a reference proteome for the study of regulatory mechanisms that can help to improvements of horticultural traits including fruit quality and yield.

In addition, comparative proteome analysis of date palm subjected to salt stress, drought (PEG induced and non-irrigated) stress using protein 2D gel electrophoresis technique showed differential protein expression pattern. Under all the three conditions of stress, ATP synthase CF1 alpha chains were found to be upregulated. Abundance in Rubisco activase and oxygen-evolving enhancer protein 2 was observed in salt and drought stress induced under no-irrigation conditions whereas expression levels of transketolase was significantly altered in PEG-induced drought stress (El-Rabey *et al.*, 2016). Thus proteome analysis has provided a comprehensive insight into drought and salt induced stress in date palm. Most importantly, protein biomarker linked to gender of date palm received a fillip with a proteome study by Dakhlaoui-Dkhil *et al.* (2013). Comparative proteome maps of male and female date palms (cultivar Deglet Nour) identified ABC superfamily ATP binding cassette transporter as male specific protein. The relevance of this protein spot was further confirmed by analysing proteomes of other cultivars such as Aligue, Khouet Aligue, Kentichi and Kenta (Dakhlaoui-Dkhil *et al.*, 2013).

8. Conclusion

Date palm, due to its invaluable nutritious fruit, has invited attention of the biotechnologists worldwide. Developments in the field of sequencing technologies, generation of SNP based markers, has played a greater role in refining genetic and linkage map of date palm. The advent of next generation sequencing (NGS) technologies enabled deciphering transcriptional changes during fruit development and helped identifying gene regulatory networks under the control of small non-coding RNAs such as miRNAs. Thus, it is certain that developments in the field of date palm genomics and transcriptomics would lead to deeper understanding of molecular phenomenon that underlie various important aspects of date palm cultivation, and aid in crop improvement programmes.

References

Al Farsi, M.A. and Lee, C.Y. (2008). Nutritional and functional properties of dates: a review. *Critical Reviews in Food Science and Nutrition*, **48**: 877–887.

Al-Mssallem, I.S., Hu, S., Zhang, X., Lin, Q., Liu, W., Tan, J., Yu, X., Liu, J., Pan, L., Zhang, T. and Yin, Y. (2013). Genome sequence of the date palm *Phoenix dactylifera* L. *Nature Communications*, **4**: 2274. doi:10.1038/ncomms3274.

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*, **5(2)**: 123–142.

Al-Shahib, W. and Marshal, R.J. (2003). The fruit of the date palm: it's possible use as the best food for the future. *Food Science and Nutrition*, **54**: 247–259.

Augstburger, F., Berger, J., Censkowsky, U., Heid, P., Milz, J. and Streit, C. (2002). Date Palm. Naturland, Germany.

Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., Bergkamp, R., Dirske, W., Van Staveren, M. and Stiekema, W. (1998). Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. *Nature*, **391**: 485- 488.

Dakhlaoui-Dkhil, S.S., Coquet, L.L., Cosette, P., Elkahoui, S., Song, P.C.T., Vaudry, D., Hannachi, A.S., Trifi, N., Limam, F. and Jouenne, T. (2013). The date palm (*Phoenix dactylifera* L.) leaf proteome: identification of a gender biomarker to screen male parents. *Plant Omics*, **6(1)**: 18-23.

El-Rabey, H.A., Al-Malki, A.L. and Abulnaja, K.O. (2016). Proteome analysis of date palm (*Phoenix dactylifera* L.) under severe drought and salt stress. *International Journal of Genomics*, **2016**: 7840759. doi:10.1155/2016/7840759.

Fang, Y., Wu, H., Zhang, T., Yang, M., Yin, Y., Pan, L., Yu, X., Zhang, X., Hu, S., Al-Mssallem, I.S. and Yu, J. (2012). A Complete sequence and transcriptomic analyses of date palm (*Phoenix dactylifera* L.) mitochondrial genome. *PLoS ONE*, **7(5)**: e37164. doi:10.1371/journal.pone.0037164.

Freeling, M. (2009). Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annual Review of Plant Biology*, **60**: 433–453.

Hazzouri, K.M., Flowers, J.M., Visser, H.J., Khierallah, H.S., Rosas, U., Pham, G.M. and Haider, N. (2015). Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. *Nature Communications*, 6. doi:10.1038/ncomms9824.

Marondedze, C., Gehring, C. and Thomas, L. (2014). Dynamic changes in the date palm fruit proteome during development and ripening. *Horticulture Research*, **1**: 1–15.

Mathew, L.S., Spannagl, M., Al-Malki, A., George, B., Torres, M.F., Al-Dous, E.K., Al-Azwani, E.K., Hussein, E., Mathew, S., Mayer, K.F., Mohamoud, Y.A., Suhre, K. and Malek, J.A. (2014). A first genetic map of date palm (*Phoenix dactylifera*) reveals long-range genome structure conservation in the palms. *BMC Genomics*, **15**: 285. doi:10.1186/1471-2164-15-285.

Mazzucotelli, E., Belloni, S., Marone, D., De Leonardis, A.M., Guerra, D., Di Fonzo, N., Cattivelli, L. and Mastrangelo, A.M. (2006). The E3 ubiquitin ligase gene family in plants: regulation by degradation. *Current Genomics*, **7**: 509–522.

Morton, J.F. (1987). Date. *Fruits of warm climates*. JF Morton. pp. 5-11.

Parvin, S., Easmin, D., Sheikh, A., Biswas, M., Sharma, S.C.D., Jahan, M.G.S., Islam, M.A., Roy, N. and Shovon, M.S. (2015). Nutritional analysis of date fruits (*Phoenix dactylifera* L.) in perspective of Bangladesh. *American Journal of Life Sciences*, **3(4)**: 274-278.

Raubeson, L.A., Peery, R., Chumley, T.W., Dziubek, C., Fourcade, H.M., Boore, J.L. and Jansen, R.K. (2007). Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics*, **8**: 174. doi:10.1186/1471-2164-8-174.

Robinson, M.L., Brown, B. and Williams, C. (2012). The date palm in Southern Nevada. *Nevada:* The University of Nevada, 23p.

Sauter, M., Rzewuski, G., Marwedel, T. and Lorbiecke, R. (2002). The novel ethylene-regulated gene OsUsp1 from rice encodes a member of a plant protein family related to prokaryotic universal stress proteins. *Journal of Experimental Botany*, **53**: 2325-2331.

Stern, D.B. and Lonsdale, D.M. (1982). Mitochondrial and chloroplast genomes of maize have a 12-kilobase DNA sequence in common. *Nature*, **299**: 698–702.

Tang, Z.X., Shi, L.E. and Aleid, S.M. (2013). Date fruit: chemical composition, nutritional and medicinal values, products. *Journal of the Science of Food and Agriculture*, **93**: 2351–2361.

Wang, D., Wu, Y.W., Shih, A.C., Wu, C.S., Wang, Y.N. and Chaw, S.M. (2007). Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 MYA. *Molecular Biology and Evolution*, **24**: 2040–2048.

Xin, C., Liu, W., Lin, Q., Zhang, X., Cui, P., Li, F., Zhang, G., Pan, L., Al-Amer, A., Mei, H. and Al-Mssallem, I.S. (2015). Profiling microRNA expression during multi-staged date palm (*Phoenix dactylifera* L.) fruit development. *Genomics*, **105(4)**: 242-251.

Yaish, M.W. and Kumar, P.P. (2015). Salt tolerance research in date palm tree (*Phoenix dactylifera* L.), past, present, and future perspectives. *Frontiers in Plant Science*, **6**: 348. doi:10.3389/fpls.2015.00348.

Yang, M., Zhang, X., Liu, G., Yin, Y., Chen, K., Yun, Q., Zhao, D., Al-Mssallem, I.S. and Yu, J. (2010). The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS ONE*, **5(9)**: e12762. doi:10.1371/journal.pone.0012762.

Zhang, G., Pan, L., Yin, Y., Liu, W., Huang, D., Zhang, T., Wang, L., Xin, C., Lin, Q., Sun, G. and Abdullah, M.M. (2012). Large-scale collection and annotation of gene models for date palm (*Phoenix dactylifera* L.). *Plant Molecular Biology*, **79(6)**: 521-536.