Date palms of Al Jufrah oasis: genetic fingerprinting of local cultivars and pollinators

Milvia Luisa Racchi & Alessandro Camussi

Department of Agri-Food Production and Environmental Sciences - Genetics Section
University of Florence - Via Maragliano 77, 50144 Firenze

Date is the most important fruit crop of arid climate region in North African and Middle East. Unlike other North African countries, in which the predominance of elite cultivars determined severe genetic erosion, date palm germplasm in Libya still preserves an enormous richness. More than 400 different date varieties are still grown in the country, out of which 95 are of commercial interest. This incredible richness has served as a highly effective natural defence for the Libyan plantations, which have remained safe from pathogens as *Fusarium oxysporum f. sp. albedinis* (Bayoud disease) and deserves to be preserved and evaluated in term of genetic diversity. Dates can be classified based on the morphological and fruit features at harvest. Libya's date varieties can be divided into three major groups: the fleshy-fruited coastal varieties, the semi-soft varieties from the central zone and the less succulent varieties, from the southern oases. The Al Jufrah oasis because it represents one of the most interesting Libyan regions for date palm cultivation (Figure 1a) was chosen for sampling. It is located on 29th parallel and consists of five localities, Hun, Waddan, Sokna, Zellah and Al Fugha.

Libya's date palm genetic resources deserve to be evaluated with the aim both to organize their preservation, to transmit a significant genetic richness and also to exploit it.

In Libya, each palm grove is typified by a distinct cultivar composition that results from local selection within the oases. Date palms have been mainly clonally propagated by offshoots, in just a few cases seed propagation is performed using the pollen available from male trees of undefined origin. In general each cultivar derives from an individual seed, cloned thereafter by vegetative multiplication to ensure the identity and uniformity of the cultivar. However intra-cultivar variation could potentially cause problems in cultivar identification. The demonstration of the true-to-type character of the plants is an important part of quality assurance and it requires the use of markers effective in distinguishing the cultivars.

Molecular markers, based on polymorphisms at the DNA level, are currently used and have proved effective to assess genetic diversity. Microsatellites, or Simple-Sequence Repeats (SSR), represent a suitable tool for genotyping because of their particular features such as their co-dominant nature and their typically high levels of allelic diversity at different loci (Kalia et al. 2011). Genetic fingerprinting by means of molecular markers of Libyan date palm cultivars has been performed with the aim both to identify the cultivars and to investigate the genetic diversity in Libya to improve production of this crop. The identification of male plants before flowering, because of the long juvenile phase of the date palm, represents a major constraint to breeding programs. So far, farmers chose the pollinators according to the availability during the female flowering time without any knowledge about their genotype. For that purpose the parentage analysis of pollinator plants was also attempted to contribute to fruit quality breeding.

Eighteen cultivars, representing common genotypes in Al Jufrah oasis were selected for their good fruit quality and were analyzed using 16 highly polymorphic microsatellite loci. Plant materials for DNA extraction consisted of young leaves of adult female trees of cultivars randomly sampled in the localities of Sokna, Hun, Waddan, Zellah and Al Fugha: the number of plants sampled is presented in Figure 1b and the fruits of the cultivars are shown in Figure 2.

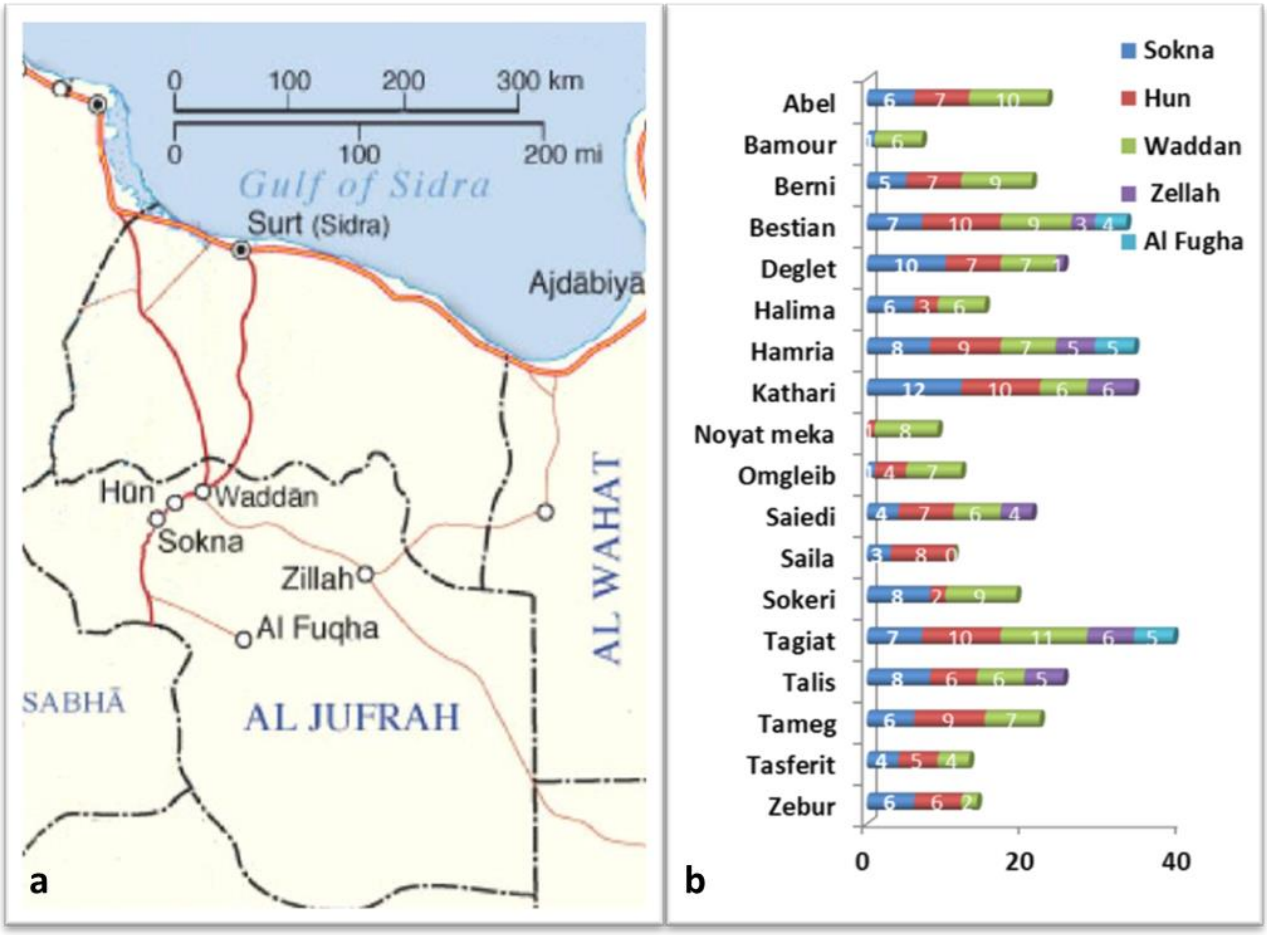

Figure 1. a) Localities of Al Jufrah oasis chosen for sampling; b) number of female plants of the different cultivars sampled

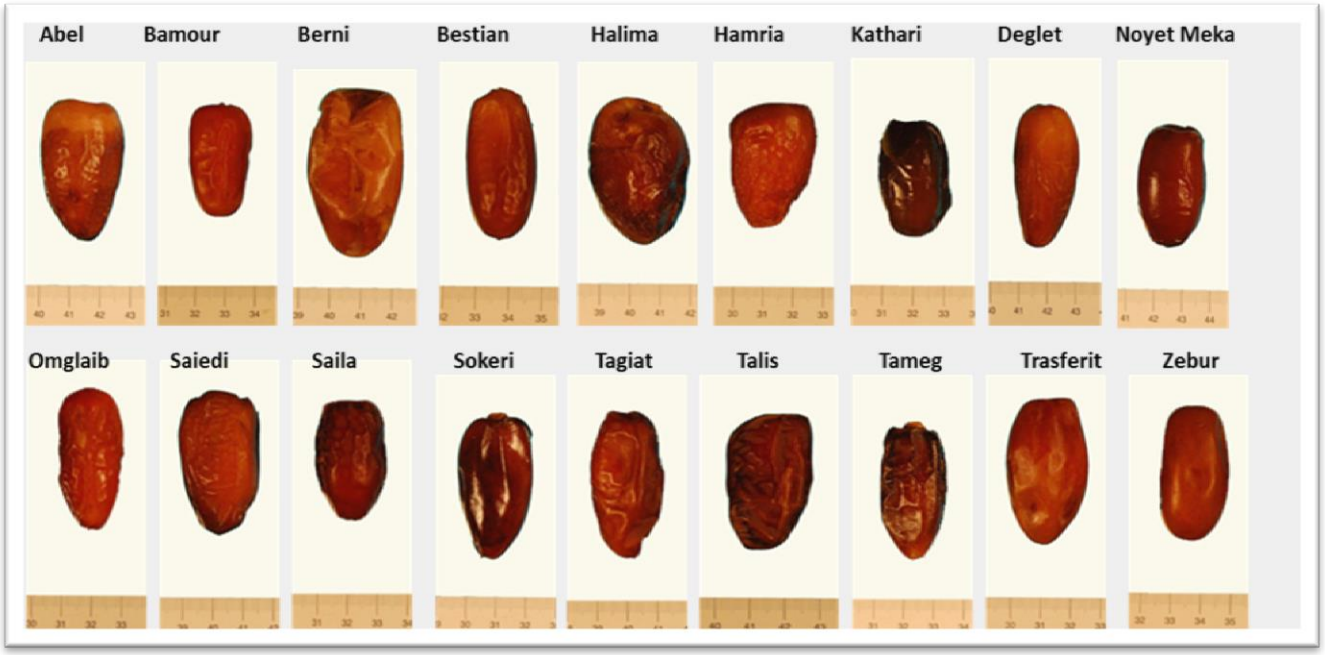

Figure 2. Fruits of cultivars sampled at Al Jufra oasis.

DNA amplifications were performed and PCR products sequenced (details in Racchi et al. 2014).

A large number of SSR alleles were revealed with a mean of 6.88 alleles per locus and a relatively high degree of genetic variability was detected (Table 1). A high level of polymorphism was detected among Al Jufra cultivars as previously reported for cultivars in Algeria, Morocco, Tunisian and Sudan using both isoenzymes and SSR markers (Bennaceur et al. 1991; Elhoumaizi et al. 2006; Elshibli and Korpelainen 2008, 2009; Zehdi et al. 2004a, b). More recently, the presence of higher polymorphism within the date palm genome was evidenced by the results obtained from parallel sequencing (Al-Dous et al. 2011).

Locus code	Allelic range (bp)	Total Alleles	Number of Genotypes
PdCIR10	138–176	6	13
mPdCIR15	142–157	6	15
mPdCIR25	219–257	6	17
mPdCIR32	306–321	5	13
mPdCIR70	205–227	9	32
mPdCIR78	126–173	11	36
mPdCIR85	175–199	8	39
mPdCIR93	181–197	7	17
PDCAT1	103–123	4	10
PDCAT2	186–209	7	20
PDCAT6	142–172	7	17
PDCAT8	222–258	6	14
PDCAT11	154–177	6	20
PDCAT14	141–163	9	20
PDCAT17	131–157	6	14
PDCAT18	123–149	8	29

Table 1. Microsatellite allelic data as revealed by 16 SSR loci in female trees of 18 Libyan date palm cultivars (modified from Racchi et al. 2014).

A total of 110 alleles with an average of 6.88 alleles per locus were scored. The number of alleles per locus ranged from four for locus PDCAT1 to eleven for locus mPDCIR78; expected heterozygosity values ranged from 0.46 (mPdCIR10) to 0.85 (mPdCIR78 and mPdCIR85) indicating that the Libyan date palm germplasm is characterized by a high degree of genetic diversity. Each cultivar results from an empirical selection carried out by the farmers in the oases based on morphological characters and fruit quality; this fact justifies the presence, at the same time, of fixed alleles, 28 out of 110, due to random drift and the high level of heterozygosity due to a clonal breeding procedure for heterosis. Interestingly, all the alleles at PDCAT1 locus were fixed even in different cultivars. All 120 pair-wise comparisons among the 16 SSR loci did not show significant linkage disequilibrium (Racchi et al. 2014).

Both number and frequencies of alleles vary among the localities due to a different presence of the cultivars in the localities. A good example is represented in Figure 3 by locus mPdCIR10, which exhibits six alleles: the allele 154 is fixed in 9 out of 18 cultivars, while alleles at locus CAT11 are greatly polymorphic. These

loci well exemplify the different distribution among the oases; in fact while CAT11 alleles are present in all the oases, some CIR10 alleles are not equally distributed.

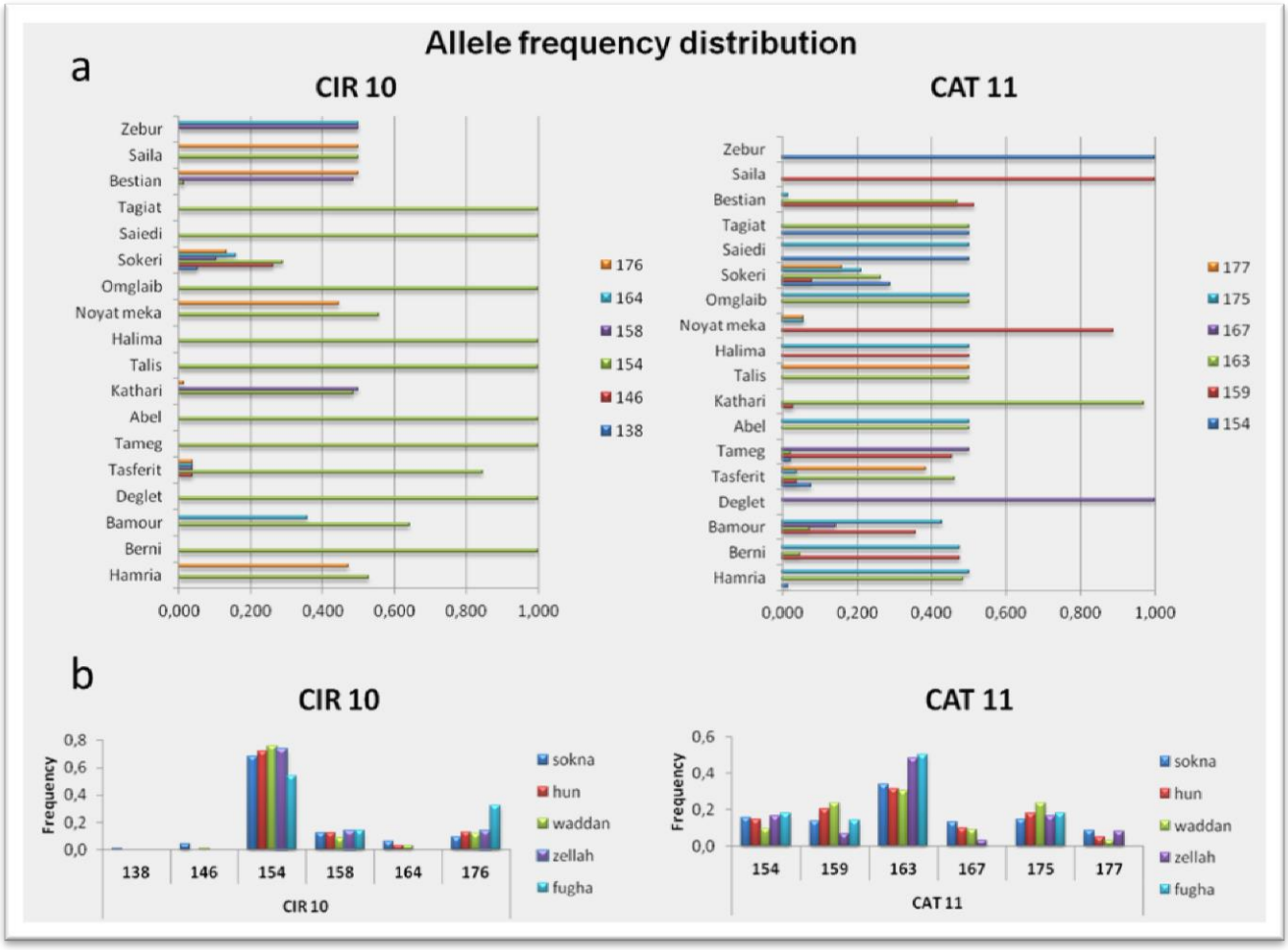

Figure 3. Frequency distribution of alleles of CIR10 and CAT11marker that varies both among: **a** Cultivars and **b** Oases.

Analysis of the genetic structure of the cultivars reveals that all are characterized by negative values of the fixation index (F) due to an excess of heterozygotes respect to HW equilibrium, though at different level. In particular cvs Talis, Halima, Omglaib, Saiedi, Tagiat, Saila and Zebur present F= -1, which indicates a strong heterotic selection at the base of the clonal breeding of these cultivars. On the other hand an F value close to 0 is expected under random mating, as observed in Sokeri that is traditionally seed propagated. Genetic diversity among cultivars was estimated as *Codominant Genotypic Distances* (Smouse and Peakall, 1999) obtained from SSR profiles. The genetic distance matrix was then submitted to analysis of Principal Coordinates, a multivariate technique that allows to find and to plot the major patterns within it in two or more dimensions. The plot is presented in Figure 4. The first two coordinates account for 45 % of total variability and the third explains a further 18 % (not shown in Figure). The first two coordinates reveal the separation of cultivars in two main groups. The use of codominat genotypic distances allows estimating the average similarity internal to each cultivar.

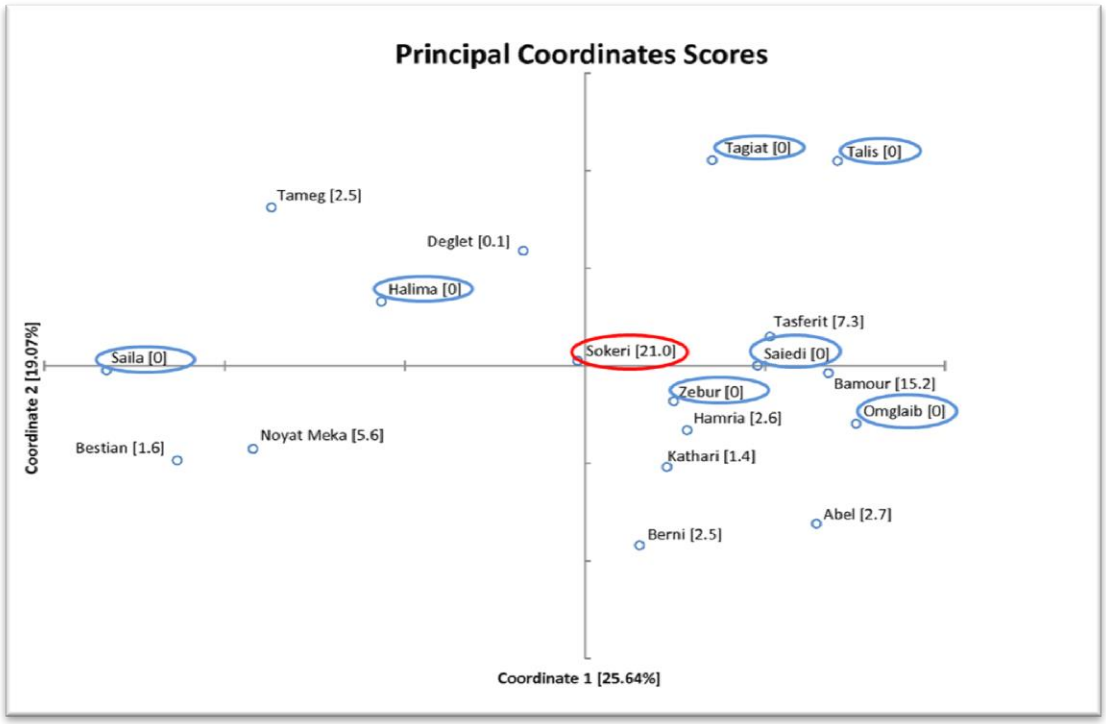

Figure 4. Analysis of principal coordinates of SSR loci

Consequently, genetic heterogeneity of each cultivar is indicated by the coefficient of genetic dissimilarity (number in bracket). The coefficient of genetic dissimilarity values range from 0 to 20.98. Cvs Talis, Halima, Omglaib, Saiedi, Tagiat, Saila and Zebur showed value 0, indicating absence of genetic difference within cultivar in agreement with their fixation index. This result evidences that farmers have good skills in propagation by offshoot. In contrast, the high value (20.98) shown by cv. Sokeri is explained by the practice of seed propagation of this cultivar. Clonal propagation, beyond to guarantee genetic uniformity of the cultivars, also limits the negative effect of inbreeding. In fact, it provides the maintenance within cultivars of high level of heterozygosity achieved by assorting heterotic positive characteristics resulting from the empirical selection of plants with good pomology features and fruit quality. However in Libya as in Sudan (Elshibli and Korpelainen 2008), seed propagation still occurs because of the ease and rapidity of seed reproduction coupled with their large availability. Consequently, date palm plantations are a mixture of plants both clonally or seed propagated with a high genetic variability within cultivars. Nevertheless cases of misclassification can occur during propagation because of the difficulty in identifying some cultivars on the base of morphology.

The effectiveness of SSR in discriminating among all the accessions and cultivars evidences the usefulness of these markers for clonal fingerprinting and cultivar identification. Accordingly, because of the strong cultivar genetic identity observed, it was possible to design an identification key that allowed the labelling of all the cultivars studied. For each date palm cultivar, the detected genotypes for mPdCIR78, mPdCIR93, mPdCIR25 microsatellite loci were scored. A total of 23 alleles were identified in these loci: ten alleles labeled (a1–a11) for mPdCIR78 locus, seven alleles (b1–b7) for mPdCIR93 locus and six alleles (c1–c6) for mPdCIR25 locus. Since each cultivar was identified by a unique profile, it is possible to generate an individual barcode using the multi-locus genotype (an example is shown in Figure 5). Similar result was previously obtained by Zehdi et al. (2006) in the analysis of 49 Tunisian accessions with three SSR loci. DNA barcoding so far proposed as an ideal supplementary tool for palm systematic (Jeanson et al. 2011, Ballardini, 2013) could become useful also in the certification and the control of origin labels of date palm products.

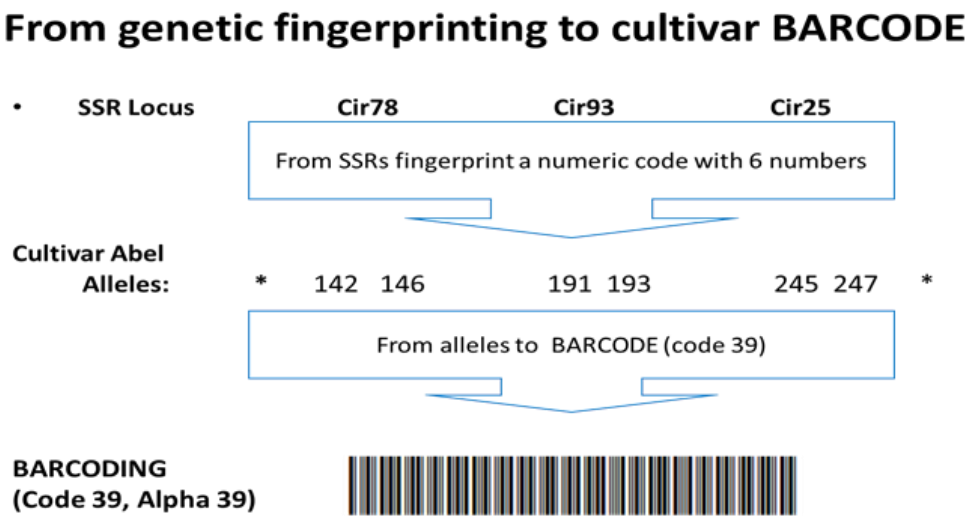

Figure 5. Procedure for generating a barcode based on data of genetic fingerprinting

The *Structure* software, based on a Bayesian procedure (Pritchard et al., 2000), represents a different approach for genetic characterization of cultivars. The program implements a model-based clustering method for inferring population structure using genotype data consisting of unlinked markers, as in our case. Applications of the method include identifying distinct genetic populations and assigning individuals to populations. We assumed a model in which the number of populations (the parameter K) is 18 (the number of cultivars), each of which is characterized by a set of allele frequencies at each locus. The Figure 6 shows how the single plants are assigned (probabilistically) to the cultivars represented by different colors. The assignment was very efficient, evidencing, as in Hamria, Tameg, Tasferit and Katari, some sampling mistakes.

The genetic heterogeneity of Sokeri is also confirmed. The Bayesian approach was effective even if the program assumes that, within populations, the loci are at Hardy-Weinberg equilibrium and linkage equilibrium. While our SSR loci are in linkage equilibrium, the cultivars are generally far from HW equilibrium.

The improvement of fruit yield and quality is based on the possibility to identify the genotype of male plants used as pollinators. Pollination of female plants is generally carried out mixing pollen from the few male plants present in the farm. Most of the time the cultivar identity of male plant is unknown because of seed propagation and the exchanges that often occur among farmers. Considering that genetic variability observed among cultivars is higher than that within cultivars, a full sib marker assisted selection procedure could be proposed starting from the cross of cultivars with different positive traits. For that purpose to attribute an unknown male tree to a cultivar becomes important. Consequently, a parentage analysis was performed through comparing genotypes of male plants and cultivars.

Considering the SSR effectiveness in fingerprinting genotypes, we used them to assign male plants, sampled in each farm of the different localities within the Al Jufrah oasis. Two approaches were used: the method of *Maximum-likelihood Paternity Assignment* (*Cervus* software, Kalinowski et al., 2007) and the Structure procedure, as above.

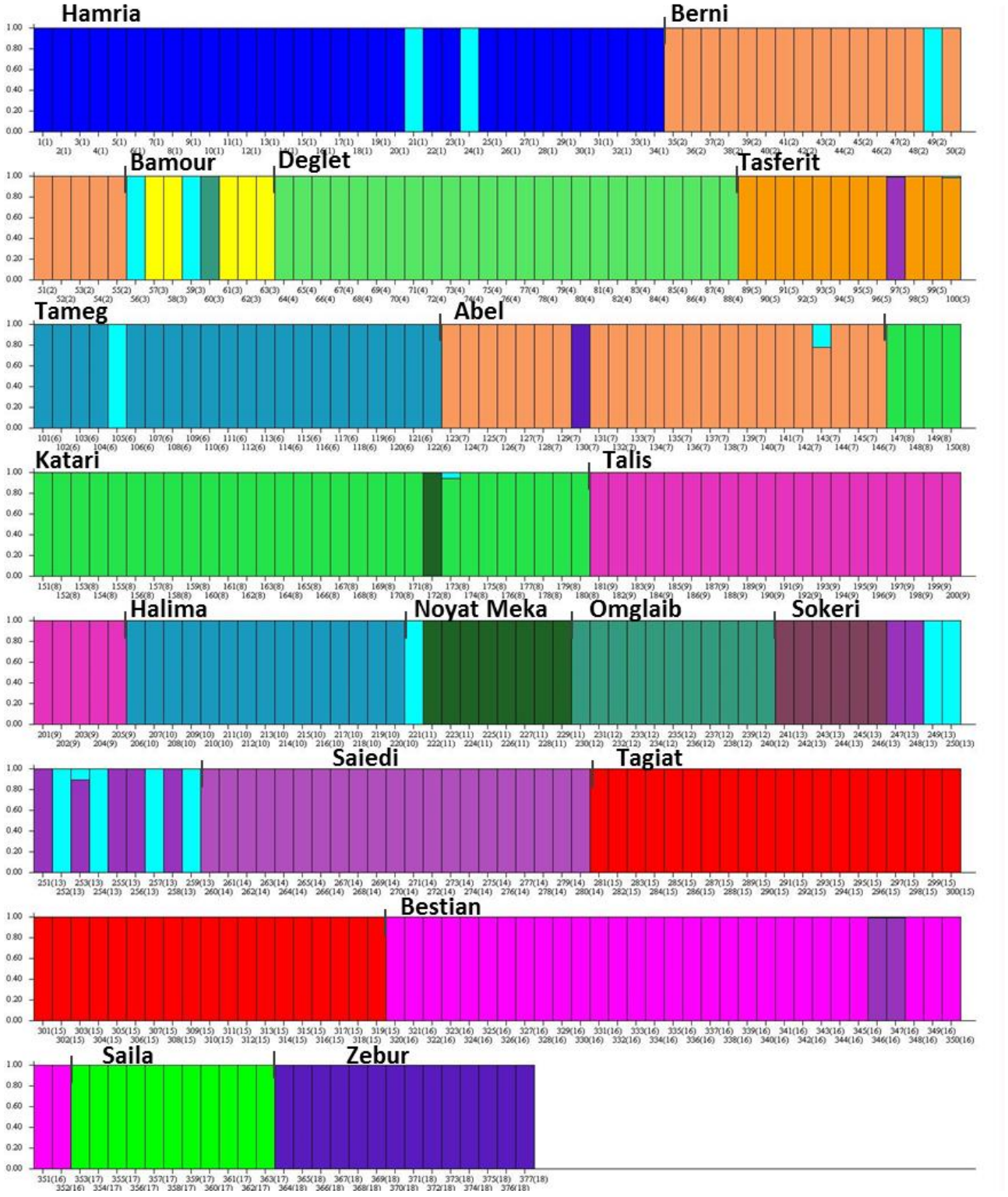

Figure 6. Plant assignment to 18 cultivars according to STRUCTURE Parameters: no admission model; K = 18; 5000 Burn-in period; 50000 Reps

The first approach allowed assigning males to a single cultivar: 55 out of the 63 male plants were assigned to 9 cultivars with strict confidence. The 19 male plants presenting positive LOD score evidenced that each of them has at least one allele of the loci included in the identification key in common with the cultivar assigned by the parentage analysis

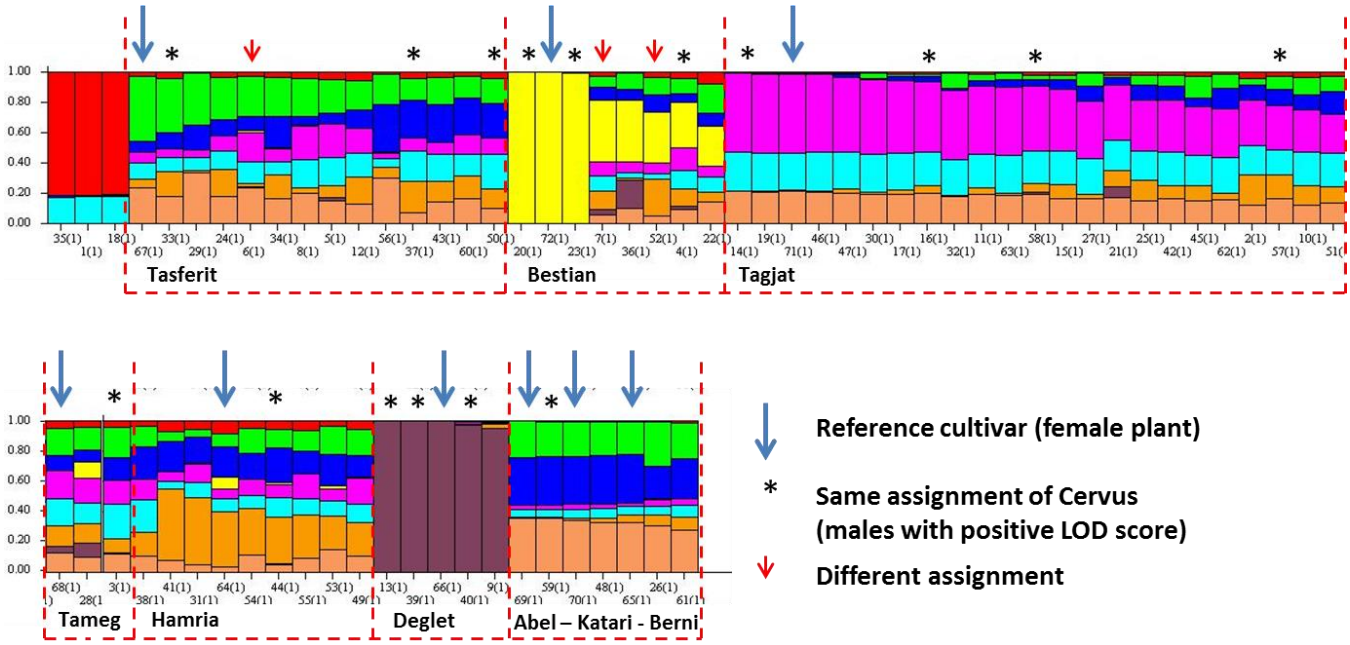

Figure 7. Male date plant assignment to cultivars according to Structure.
Parameters: no admission model ; K = 9; 5000 Burn-in period; 50000 Reps.

The profiles of the 63 male plants together with those of the nine reference cultivars were submitted to the Structure procedure. The Structure analysis confirmed the results previously obtained (Figure 7). In fact, 16 out of the 19 male plants with a positive LOD score in the Cervus analysis showed the same assignment. These positive results confirm the effectiveness of the SSR markers also in the assignment of unknown plants to their putative cultivar and further highlight the suitability of SSR for genotyping, opening new prospects for date palm breeding.

REFERENCES

Al-Dous EK, George B, Al-Mahmud ME, Al-Jabber MY, Wang H, Salameh YM, Al-Azwani EK, Chaluvadi S, Pontaroli AC, DeBarry J, Arondel V, Ohlrogge J, Saie IJ, Suliman-Elmeer KM, Bennetzen JL, Kruegger RR, Malek JA (2011) De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). Nat Biotechnol 29(6):521–527

Ballardini M (2013) Barcoding in the *Phoenix* genus *Dies Palmarum* VII Biennale Europea delle Palme 5-7 December 2013

Benaceur M, Lanaud C, Chevallier MH, Bounaga N (1991) Genetic diversity of the date palm (*Phoenix dactylifera* L.) from Algeria revealed by enzyme markers. Plant Breed 107:56-57

Elhoumaizi MA, Devanand PS, Fang J, Chao CT (2006) Confirmation of 'Medjool' date as a landrace variety through genetic analysis of 'Medjool' accessions in Morocco. J Amer Soc Hort Sci 131:403-407

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

Elshibli S, Korpelainen H (2009) Biodiversity of date palms (*Phoenix dactylifera* L.) in Sudan: chemical, morphological and DNA polymorphism of selected cultivars. Plant Genet Resour. doi: 10.1017/s1479262108197489

Jeanson MJ, Labat JN, Little DP (2011) DNA barcoding: a new tool for palm taxonomist? Annals of Botany 108:1445-1451,

Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2011) Microsatellite markers: an overview of the recent progress in plants. Euphytica 177:309-334

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

Racchi ML, Bove A, Turchi A, Bashir G, Battaglia M, Camussi A (2014) Genetic characterization of Libyan date palm resources by microsatellite markers. 3 BIOTECH 4: 21-32, ISSN:2190-572X DOI 10.1007/s13205-013-0116-6

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity 82: 561–573

Steven T, Kalinowski M, Taper L, Tristan C, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology 16: 1099–1106

Zehdi S, Sakka H, Rhouma A, Salem AOM, Marrakchi M, Trifi M (2004a) Analysis of Tunisian date palm germplasm using simple sequence repeat primers. Afri J Biotech 3: 215-219

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

Zehdi S, Pintaud JC, Billotte N, Salem AOM, Sakka H, Rhouma A, Marrakchi M, Trifi M (2006) Etablissement d'une clé d'identification variétale chez le palmier dattier (*Phoenix dactylifera* L.) par les marqueurs microsatellites. Plant Genet Resour Newsl 145: 11-18