Original Article

Insertion/deletion markers for assessing the genetic variation and the spatial genetic structure of Tunisian *Brachypodium hybridum* populations

Mohamed Neji[1,2]*, Filippo Geuna[2], Sean P. Gordon[3], Wael Taamalli[1], John P. Vogel[3], Yosra Ibrahim[1], Abd Errazak Smaoui[1], Chedly Abdelly[1], Mhemmed Gandour[1,4]

[1]Laboratory of Extremophile Plants, Centre of Biotechnology of Borj-Cédria, BP 901 Hammam Lif 2050 Tunisia, [2]Department of Agricultural and Environmental Sciences DISAA, Laboratory of Molecular Genetics, University of Milan, Via Celoria 2, 20133 Milan, Italy, [3]USDA-ARS Western Regional Research Center, 800 Buchanan St., Albany, CA 94710, USA, [4]Faculty of Sciences of Sidi Bouzid, University of Kairouan, Tunisia

ABSTRACT

The wild annual grass *Brachypodium hybridum,* an allotetraploid species derived from the natural hybridization between the diploid species *Brachypodium distachyon* ($2n = 10$) and *Brachypodium stacei* ($2n = 20$). This trio of species has been suggested as a model system for polyploidy. *B. hybridum* is the most widespread *Brachypodium* species in Tunisia. Natural diversity can be used as a powerful tool to uncover gene function and, in the case of *B. hybridum*, to understand the functional consequences of polyploidy. Here, we examined the spatial distribution of genetic variation of *B. hybridum* across its entire range in Tunisia and tested underlying factors that shaped its genetic variation. Population genetic analyses were conducted on 145 individuals from 9 populations using 8 insertion/deletion markers. Results indicated a relatively high level of within population genetic diversity ($H_e = 0.35$) and limited among population differentiation ($\Phi_{PT} = 0.20$) for this predominantly self-pollinating grass. Unweighted pair group method with. Arithmetic averaging cluster analyses, principal coordinate analysis, and Bayesian clustering supported the demarcation of the populations into three groups that were not correlated with location or altitude, suggesting a loose genetic affinity of *B. hybridum* populations in relation to their geographical locations, and no obvious genetic structure among populations across the study area. This pattern was associated with a considerable amount of an asymmetric gene flow between populations. Overall, the obtained results suggest that the long-distance seed dispersal is the most important factor in shaping the spatial genetic structure of *B. hybridum* in Tunisia. They also provide key guidelines for on-going and future work including breeding programs and genome-wide association studies.

KEY WORDS: *Brachypodium hybridum*, genetic diversity, insertion/deletion, long-distance seed dispersal, spatial genetic structure

INTRODUCTION

A long-standing goal in evolutionary and conservation biology is to understand how the fundamental evolutionary forces such as genetic drift, gene flow, and natural selection interact with each other in shaping the pattern of the spatial genetic structure of natural populations (Hartl and Clark, 2007). The low cost of next generation sequencing and the availability of reference genome sequences for many plant species have dramatically lowered the cost and increased the utility of high-throughput genotyping. Simple sequence repeat (SSR) markers and single nucleotide polymorphisms (SNP) are the most commonly used markers for population studies. However, they both have some technical limitations. While SSRs are highly polymorphic and can be scored on simple agarose gels, differences in polymerase chain reaction (PCR) product size estimation between experiments and operators hinder accurate comparison of many accessions or experiments. In contrast, SNPs are much more reproducible but accurate genotyping requires the creation of custom chips or sequencing individual amplicons both very expensive for the large populations. In this context insertion/deletion (InDel) markers have several advantages including being inexpensive and easy

to score; highly polymorphic, co-dominant inherence, and a high degree of reproducibility. Thus, they have proven to be efficient and cost-effective tools and are commonly used to study genetic diversity of many plant species (Pacurar *et al.*, 2012).

The genus *Brachypodium* includes several species endowed with widespread implications for forage, bioenergy production, turf and ornamental purposes and noticeable ecological potential for revegetation, soil stabilization and characterized by many valuable qualities such as wide adaptation range, stress tolerance, and disease resistance (Brkljacic *et al.*, 2011; Betekhtin *et al*, 2014). Supported by its small, recently sequenced genome (IBI, 2010), *Brachypodium distachyon* has gained a lot of interest and has been adopted as a model for cereals, forage and bioenergy grasses. *B. distachyon* was initially described as a polyploid series with cytotypes of $2n = 10$, 20 and 30 chromosomes. However, recent investigations have demonstrated that the $2n = 10$ and $2n = 20$ cytotypes are two diploid species *B. distachyon* and *Brachypodium stacei*, respectively, whereas the $2n = 30$ cytotype is an allotetraploid species deriving from a spontaneous hybridization between *B. distachyon* and *B. stacei* (Catalan *et al.*, 2012). These three mainly self-pollinating species are morphologically similar to each other but express different traits. *B. stacei* and *Brachypodium hybridum* have higher seed yield, more biomass, and grow faster than *B. distachyon* (Catalan *et al.*, 2012). This trio of species is being developed as a model for polyploidy, and the genomes of *B. stacei* and *B. hybridum* have been sequenced (Catalan *et al.*, 2014). According to Lopez-Alvarez *et al.* (2012), *B. hybridum* exhibited wider geographic distribution and greater ecological adaptation than its progenitors in the Mediterranean region. However, remarkably little is known about the factors shaping its spatial genetic structure.

In Tunisia, *B. hybridum* is characterized by a wide geographic distribution; it grows on open hillsides, mixed grass steppes and abandoned lands and covers ecological environments varying from arid to sub-humid (Neji *et al.*,

2014). However, investigations on the genetic variation of its natural germplasm remain scarce; it has lagged behind with only two studies up to date. Our preliminary surveys based on phenotypic traits and SSR markers demonstrated that the Tunisian *B. hybridum* germplasm is characterized by a substantial phenotypic and genetic variation (Neji *et al.*, 2014; Neji *et al.*, 2015).

A broad genetic base is important for genetic resources and plant breeding applications which include but are not limited to the generation of core collections, parental lines selection, quantitative trait loci identification, association mapping or genomic selection, all of which require broad information of genetic variation patterns. In this study, InDel markers were used to provide a broader picture about the genetic diversity and the population structure of Tunisian *B. hybridum* populations to contribute knowledge on the germplasm genetic base for breeding strategies.

MATERIALS AND METHODS

Plant Material and DNA Extraction

In this study, we characterize 145 *B. hybridum* lines representative of nine populations that span the geographical range of *B. hybridum* in Tunisia (Table 1 and Figure 1). This panel of lines was previously characterized using morpho-phenotypic traits and SSR markers (Neji *et al.*, 2014; Neji *et al.*, 2015). Plants were grown under greenhouse conditions as previously described by Neji *et al.* (2014). Genomic DNA was isolated from about 60 mg of young leaves as described by Geuna *et al.* (2003). The relative purity and concentration of the extracted DNA were estimated spectrophotometrically at 260 nm and 280 nm and by 1% agarose gel, and the final concentration of each DNA sample was adjusted to 20-30 ng/μL.

Marker Selection and Amplification

From a set of 287 InDel markers developed for *B. distachyon*, we selected 24 markers evenly distributed across the 5 chromosomes of *B. distachyon*. We tested the

Table 1: Characteristics of the origin sites of the studied *Brachypodium hybridum* populations (altitudinal classes: Class 1: 0-80 m, class 2: 80-400 m and class 3: >400 m)

ID	Location	Latitude	Longitude	Altitude (m)	Altitudinal class	Bioclimatic area (eco-region)
F	Fayedh	35°4′4.1″N	9°40′32.1″E	83	2	Arid
E	Enfidha	36°7′8.04″N	10°27′44.90″E	7	1	Lower semi arid
Z	Jbel Zaghouan	36°22′6.80″N	10° 5′20.00″E	450	3	Upper semi arid
H	Hawaria	37°2′21.00″N	11° 0′53.20″E	23	1	Sub-humid
R	Raouad	36°57′0.00″N	10°14′14.00″E	3	1	Upper semi arid
S	Sejnen	37°4′38.90″N	9° 9′51.10″E	115	2	Humid
A	Ain Drahem	36°48′10.33″N	8°41′29.02″E	134	2	Humid
K	El Kef	36° 6′36.80″N	8°37′39.00″E	482	3	Lower semi arid
D	Douar El Hej Wniss	35°40′56.40″N	8°53′49.00″E	880	3	Upper semi arid

primer pairs for amplification efficiency with 2 high-quality DNA samples from each population. Amplification was optimized by varying the concentration of the template DNA, annealing temperature, and the Mg^{2+} concentration. Optimized PCRs were set up in a total volume of 20 µL containing 30 ng of genomic DNA, 2 µL of ×10 reaction buffer, 10 mmol/L Tris-HCl at pH 8.3 and 50 mmol/LKCl, 0.5 µL of each dNTP 10 mM, 0.8 µL of $MgCl_2$ 50 mM, 0.5 µL forward primer 10 mm, 0.5 µL reverse primer 10 mm and 1 U Taq polymerase (EURx). PCR reactions were performed in PTC-100 MJ thermocycler. Thermocycling parameters were as follows: 94°C for 2 min followed by 35 cycles of 94°C for 30 s, 57°C for 30 s, 72°C for 30 s, and 72°C for 10 min. Eight primer pairs reproducibly amplified and were chosen to genotype all the 145 genotypes (Table 2). The amplification products were electrophoretically separated in 3% agarose gels buffered with ×0.5 TBE (89.15 mM Tris base, 88.95 mM boric acid, and 2.23 mM EDTA) and stained with ethidium bromide at a constant voltage 100 V for 3 h. DNA fragments were identified by image analysis software for a gel documentation. The sizes of the amplified fragments were estimated by comparison to the 100-bp DNA ladder (Invitrogen, Milan, Italy) molecular marker. Stutter and background bands were excluded.

Data Processing

Within population genetic diversity

InDel bands were scored in the form of single individual genotypes and arranged in a genotype matrix for analysis of allelic variation. For the nine populations, a set of genetic diversity parameters including observed number of alleles per locus (N_a), effective number of alleles (N_e), Shannon's

Figure 1: Tunisian populations of *Brachypodium hybridum* and their bioclimatic regions: (1) Fayedh (arid); (2) Enfidha (lower mi-arid); (3) Jbel Zaghouan (upper mi-arid); (4) Haouaria (sub-humid); (5) Raouad (upper mi-arid); (6) Jnen (humid); (7) Ain Drahem (humid); (8) El Kef (lower mi-arid); (9) Douar El Hej Wniss (mi-arid)

Table 2: Primers used to analyze the genetic diversity of the 145 *Brachypodium hybridum* lines

Primer name	tm	Forward primer	Reverse primer	N_a	N_e
Bd1:53541858-53542539	58	CACCGGAGAAATCAAGCAAC	GCGATTCCACCTAGGATTGT	4	2,746
Bd2:4941937-4942694	58	GATGACGAAGCAAACACACG	CATCCGTGTTCATGGCCTA	2	1,000
Bd2:26038441-26039244	58	GAAGGAAATGGGAACCACAA	TTCGGTGTGTTGTCCAAGAA	2	2,000
Bd3:15078416-15079063	58	TGAACGGAGTTACACCCATTT	CATTTGGGCTGAGGAATTTT	2	1,246
Bd3:48848974-48849645	58	CGTGTGTTAAGCCCAAAAGC	AGGAGACATTTGCCCCCTAT	2	2,000
Bd3:59861087-59861706	58	TGAAAACATGGCTGTGAGGA	ATCGTTGCCTGGTCCGTAT	2	1,528
Bd4:30688612-30689218	58	AAGCAACAGGCACACCATCT	GGCAAAAGAAGAAGAAAGTTTATAGT	3	2,793
Bd5:19053980-19054773	58	CACCCTTATAGCGTGCAACA	TTGGTTTGTGCCAACAACTG	2	1,385

Table 3: Number of different alleles (N_a), number of effective alleles (N_e), shanon index (I), observed heterozygosity (H_0), expected heterozygosity (H_e), fixtation index (Fis) and allelic richness (Ar) for each population (Value±SE*)

ID	N_a	N_e	I	H_0	H_e	Fis	Ar
Fayedh	2.255±0.31	1.828±0.23	0.611±0.13	0.454±0.16	0.382±0.07	0.112±0.28	2.134
Enfidha	2.375±0.18	1.668±0.2	0.562±0.10	0.247±0.128	0.343±0.07	0.415±0.22	1.625
Zaghouan	2.125±0.22	1.731±0.13	0.579±0.09	0.228±0.12	0.390±0.06	0.457±0.24	1.877
Haouria	2.250±0.25	1.572±0.12	0.512±0.09	0.314±0.11	0.329±0.06	0.187±0.23	2.155
Raouad	2.125±0.22	1.697±0.13	0.564±0.09	0.315±0.1	0.378±0.06	0.137±0.21	2.042
Sejnen	2.000±0.18	1.633±0.16	0.515±0.1	0.295±0.12	0.339±0.07	0.085±0.25	1.867
Ain Drahem	1.625±0.26	1.481±0.2	0.354±0.14	0.375±0.18	0.238±0.09	0.08±0.34	1.885
El Kef	2.250±0.16	1.856±0.16	0.664±0.07	0.366±0.16	0.432±0.04	0.319±0.28	1.927
Douar Hej Wniss	2.000±0.26	1.673±0.21	0.505±0.13	0.266±0.12	0.327±0.08	0.247±0.25	1.946
Mean	2.111±0.078	1.683±0.05	0.541±0.03	0.312±0.04	0.351±0.02	0.205±0.08	1.93

*Standard error

information index (*I*), polymorphism percentage (*P*), observed heterozygosity (*H*$_0$), and expected heterozygosity (*H*$_e$) were quantified using GenAlEx 6.5 software (Peakall and Smouse, 2012). Furthermore, the allelic richness (*Ar*) and inbreeding coefficient were calculated using FSTAT version 2.9.3 (http://www.unilch/izea/softwares/fstat). (Goudet, 2002).

Population differentiation and genetic structure analyses

To assess the population subdivision, the distribution of variation among regions/altitudinal classes (Φ$_{RT}$), populations within regions/altitudinal classes (Φ$_{PR}$), and populations (Φ$_{PT}$), respectively, an analysis of the molecular variance (AMOVA), with the populations nested within regions or altitudinal classes, was performed using GenAlEx 6.5 and 9999 permutations. To infer levels of population genetic differentiation, the same program was used to calculate Φ$_{PT}$ among populations (an analog of Fst, i.e., genetic diversity among populations).

To define the genetic relatedness between populations, a principal coordinate analysis principal coordinate analysis (PCoA) and an unweighted pair group method with arithmetic averaging (UPGMA) cluster analysis based on the between-populations Φ$_{PT}$ matrix were carried out using GenAlEx 6.5 (Peakall and Smouse, 2012) and MEGA Version 5 (Tamura *et al.*, 2011). For the UPGMA dendrogram, bootstrap analysis was carried out using POPTREE2 (Takezaki *et al.*, 2010) to assess statistical support for each cluster.

Furthermore, a Bayesian analysis of genetic structure of the sampled populations was carried out in the program STRUCTURE Version 2.2 using admixture model with correlated allele frequencies and no prior information on the sampling locations (Pritchard *et al.*, 2000). We performed 10 independent runs for each value of K (number of genetically distinct clusters) ranging from 1 to 9. Each run included 10^5 burn-in iterations and 20^5 Markov chain Monte Carlo replicates. The best value of K was determined by computing ΔK following Evanno *et al.* (2005) using STRUCTURE HARVESTER (Earl and von Holdt, 2011). Each individual was assigned to the inferred clusters using a threshold proportion of inferred ancestry (*Q*), i.e. *Q* ≥ 0.80%.

To evaluate spatial genetic structure, isolation by distance and by elevation was tested by comparing regressed pairwise Φ$_{PT}$ values against paired population matrices of geographic distances, and altitudinal differences between populations using the 14th version XLSTAT (https://www.xlstat.com/fr/) with 10,000 permutations for significance tests.

Dispersal and population connectivity

Migration rates between the studied populations were assessed using a range of genetic estimators. First, for the long-term effective gene flow, rates were calculated using the indirect Φ$_{PT}$-based method (Φ$_{PT}$ = 1/4N_m + 1) to estimate an effective number of migrants (*N*$_m$) (Wright, 1969). *N*$_m$ values were also calculated using the private allele method as implemented in GENEPOP 1.2 (Raymond and Rousset, 1995). These two estimates were then compared and correlated with each other.

To evaluate the contemporary dispersal patterns between sample sites and first generation migrants, we applied two Bayesian approaches implemented in GeneClass 2.0 (Piry *et al.*, 2004) that assign or exclude populations as the origin of sampled individuals using multilocus genotype data. First, we used self-assignment tests for detecting the most likely source populations of sampled individuals using the assignment criterion of Rannala and Mountain (1997). This test enables the detection of individuals with immigration ancestry of up to two generations earlier, that is, it allows identifying the descendants of immigrants. Individuals were assigned to the population for which they had the highest assignment likelihood.

In addition, we applied the first-generation migrant test to detect migrants directly and to reject the null hypothesis that each individual originated from its home population (Paetkau *et al.*, 2004; Boessenkool *et al.*, 2009). The statistical criterion for likelihood estimation was L_home/L_max, where L_home is the likelihood of an individual genotype within the population where the individual was sampled and where L_max is the highest likelihood value among all available population samples including the population where the individual was sampled. The probability calculation with a threshold value of 0.01 was based on Monte Carlo re-sampling using 10,000 simulated individuals (Paetkau *et al.*, 2004).

RESULTS

Within population Genetic Diversity

A total of 24 InDel primer pairs were screened on four selected individuals from each population. 13 loci were excluded because they generated multiple bands, and another three were excluded because they were monomorphic. Only eight primers that produced clear

and interpretable bands and showed polymorphism were validated to be effective for genotyping the entire collection. The selected primers yielded a total of 20 alleles in the range of 250-670 bp; all of them were polymorphic (100%). The number of observed alleles per locus (N_a) ranged from 2 to 4 with an average of 2.37 alleles across all populations. However, the number of effective alleles (N_e) was much lower than the number of observed alleles (N_a), averaging 1.83 alleles per locus. This suggests a large number of rare alleles in the studied germplasm. None of the eight loci showed significant deviation from Hardy–Weinberg equilibrium across the nine populations.

Our results revealed a relatively high genetic diversity homogeneously distributed across the analyzed populations as suggested by the low standard errors of all the genetic diversity estimators used in the analysis (Table 3). All the analyzed populations recorded narrow allelic diversity with N_a, N_e and *Ar* averaging 2.1, 1.68 and 1.95, respectively. The polymorphism rate at the population level ranged from 50% in Ain Drahem to 100% in Enfidha and El Kef, with 84.72% on average (Table 3). Observed heterozygosity and genetic diversity expressed by H_e were relatively high and varied slightly among the nine populations. The observed heterozygosity H_0 ranged from 0.228 to 0.454 with an average of 0.312. The genetic diversity, averaging in 0.35, had slightly higher values than H_0 for the majority of populations. The highest level was recorded in the population of El Kef ($H_e = 0.43$), whereas the least was observed in the population originated from Ain Drahem ($H_e = 0.24$). Moreover, the mean value

of fixation indices *Fis* was 0.2, which points toward an overall deficiency of heterozygosity across populations and confirms the predominance of self-pollination in *B. hybridum*.

Dispersal and Connectivity between Populations

Pair-wise migration data (Table 4) showed a considerably high level of gene flow among different *B. hybrdium* populations. The overall long-term gene-flow rates between populations N_m, estimated using the Φ_{PT}-based method and the private allele method were relatively high ($N_m = 1.08$, $N_m = 2.75$, respectively). Although the overall N_m estimated by the second approach was about three times higher than that obtained by first one, the N_m between pairwise populations were highly correlated ($r^2 = 0.65$; $P = 0.01$). Interestingly, the most of the migration rates recorded between pairs of populations exceeded the unity, which suggests that *B. hybridum* populations are strongly interconnected through a multidirectional gene flow.

The results from GeneClass2 indicated a considerable dispersal between sample sites. The likelihood estimation from the Rannala and Mountain (1997) method indicated 83 individuals (57.3%) with probability (i.e. probability to belong to the sample site) below the threshold $P < 0.05$. The simulation following the (Paetkau *et al.*, 2004) method, explained those individuals belong genetically to other populations rather than those where they were sampled from, rejecting the null hypothesis that they originated from their home populations.

Table 4: Effective number of migrants (N_m) calculated between each population pair using the private-allele method (above diagonal) and Φ_{PT} -bad method (below diagonal)

Populations	Enfidha	Haouria	Raouad	Fayedh	Ain Drahem	jnen	El Kef	Daouar Hej Wniss	Jbel Zaghouan
Enfidha	0.000	2.680	2.420	12.360	1.450	1.540	1.860	1.230	2.580
Haouria	0.755	0.000	9.010	5.760	0.860	2.590	1.920	1.020	3.630
Raouad	1.155	0.808	0.000	3.080	0.230	2.360	0.036	0.260	2.690
Fayedh	64.677	1.423	1.493	0.000	1.940	1.340	2.350	2.680	3.660
Ain Drahem	4.273	0.925	0.650	5.151	0.000	0.040	1.790	1.350	0.060
Jnen	0.496	0.535	1.406	0.776	0.437	0.000	0.060	0.093	7.310
El Kef	1.459	2.334	0.959	3.118	2.243	0.588	0.000	0.090	0.040
Daouar El Hej Wniss	1.172	1.113	1.443	7.600	1.459	1.427	1.097	0.000	0.130
Jbel Zaghouan	0.793	0.849	5.743	0.994	0.490	1.592	0.816	1.060	0.000

Table 5: Nei's genetic distances (below diagonal) and Φ_{PT} (above diagonal) between pairwi *Brachypodium hybridum* populations

Populations	Fayedh	Ain Drahem	jnen	El Kef	Enfidha	Daouar El Hej Wniss	Haouria	Raouad	Jbel Zaghouan
Fayedh	0	0.043	0.156	0.08	0.031	0.041	0.084	0.126	0.172
Ain Drahem	0.046	0	0.22	0.092	0.055	0.09	0.112	0.221	0.302
Jnen	0.244	0.364	0	0.276	0.252	0.084	0.221	0.111	0.092
El Kef	0.074	0.1	0.298	0	0.12	0.149	0.062	0.239	0.271
Enfidha	0.004	0.055	0.335	0.146	0	0.116	0.156	0.148	0.22
Daouar El Hej Wniss	0.032	0.146	0.149	0.186	0.176	0	0.087	0.111	0.132
Haouria	0.149	0.213	0.318	0.097	0.249	0.183	0	0.191	0.167
Raouad	0.143	0.278	0.151	0.207	0.178	0.148	0.236	0	0.07
Jbel Zaghouan	0.201	0.338	0.136	0.235	0.24	0.191	0.227	0.042	0.038

Population Differentiation and Genetic Structure

AMOVA analysis, with populations nested within regions, revealed that 80% of the genetic variation was attributed to the differences within populations ($P < 0.0001$). The total differentiation among populations was about 20% ($\Phi_{PT} = 0.2$), 15% of which was due to the among population, within-regions component ($\Phi_{PR} = 0.15$). By dividing populations into three altitudinal classes, AMOVA yielded similar patterns of within- and among-populations genetic variation partitioning but without any evidence of differentiation between altitudinal classes. Meanwhile, pairwise Φ_{PT} and Nei genetic distances (D) comparisons subsequently revealed low to moderate genetic differentiation between the studied populations. The Fayedh and Enfidha populations were found to be the most genetically close ($\Phi_{PT} = 0.04$; D $= 0.031$), whereas Ain Drahem and Jbel Zaghouan were the most divergent ($\Phi_{P} = 0.33$; D $= 0.302$) (Table 4).

The genetic relationships among populations were examined by cluster analysis, PCoA analysis, and a Bayesian clustering. Three major groups of populations were identified by the UPGMA dendrogram generated by the cluster analysis (Figure 2a). The first group included the populations from Fayedh, El Kef, Ain Drahem and Douar El Hej Wniss; the second group comprised the populations from Haouaria and Enfidha; while, the third group was formed by Sejnen, Raouad and Jbel Zaghouan
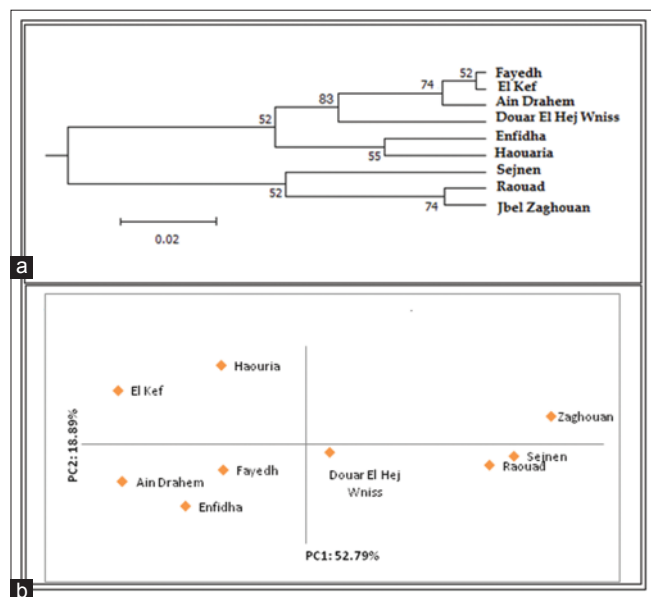
Figure 2: Distance-bad clustering of the nine populations of *Brachypodium hybridum,* (a) Unweighted pair group method with arithmetic averaging dendrogram bad on Nei's genetic distances, bootstrap values bad on 1000 permutations are indicated in each node. (b) Principal coordinates analysis of the studied *B. hybridum* populations

populations. The resulting UPGMA dendrogram was supported with relatively high bootstrap values (52-83%), suggesting a high reliability of the pattern observed. Consequently, the bi-dimensional scatter plot formed by the first two axes of the PCoA analysis (Figure 2b), showing the majority of genetic variance (81.68%), revealed similar clustering pattern. Both analyses showed that populations from distinct geographic area and altitudinal classes could be clustered together. The Mantel test revealed a nonsignificant association between genetic distances and geographic distances ($r = 0.023$; $P = 0.39$) suggesting that isolation by distance did not have a limiting effect on gene flow. Furthermore, no significant correlation was found between genetic distances and elevation ($r = -0.017$; $P = 0.48$). Hence, elevation is unlikely to be a physical barrier to gene flow. All these results indicated the possible long-distance and frequent gene flow among the *B. hybridum* populations from different geographical locations and suggested the lack of an obvious genetic structure among the *B. hybridum* populations across the country.

The Bayesian clustering showed a first relevant peak of ΔK at $K = 3$ and a small peak at $K = 7$, (Figure 3a) suggesting that the 145 *B. hybridum* individuals were most likely genetically partitioned into three main clusters, which confirms the results of cluster analysis and PCoA. As shown in Figure 3b and according to the inferred ancestry value of each genotype ($Q = 80\%$), the whole set of genotypes were distributed across three easily distinguished clusters S1, S2, and S3 although there were a few admixed genotypes. In fact, A total of 124 genotypes (85.5%) displayed a $Q > 75\%$ and showed a clear membership to one of the three clusters $Q > 80\%$, whereas only 21 genotypes (20.5%) were categorized as admixture. Noticeably, strict clustering based on geographical origin was not detected with the Bayesian-based clustering; almost all populations contained genotypes with varied memberships to the three clusters. However, some tendencies of population assignment could be observed: Cluster S1 (red color) encompassed genotypes from all the populations with a clear predominance of Ain Drahem, El Kef and Fayedh populations; cluster 2 (green color) was mainly formed by Haouria and Enfidha genotypes with some samples from Douar El Hej Wniss, Jbel Zaghouan and Raouad, whereas cluster S3 (blue color) was mainly represented by Raouad, Sjenen, Daouar El Hej Wniss and Jbel Zaghouan genotypes. Such pattern of clustering strongly supports the high gene flow detected between populations and suggests the lack of an obvious genetic structure of *B. hybridum* in Tunisia.
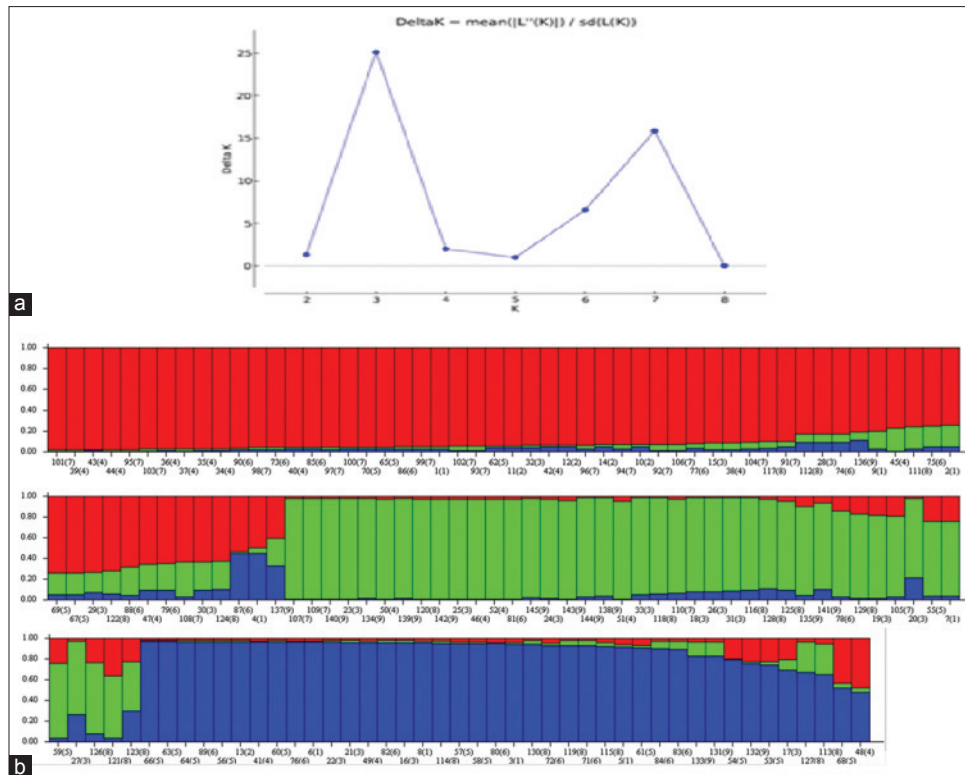
Figure 3: (a) ∆K calculated as ∆K = *m* |*L*" *K*|/*s* [*LK*] (Evanno *et al.*, 2005). (b) Estimated genetic structure of 145 *Brachypodium hybridum* individuals assd by STRUCTURE *K* = 3 each individual is reprented by a thin vertical bar partitioned into up to *K* colored segments

DISCUSSION

Choosing the most appropriate marker depends on the purpose of the research, reproductive strategy of the species and its genetic structure (Wellington *et al.*, 2013). So far, InDel markers have been successfully used and found to be highly efficient in investigating genetic relationships among cultivars and natural populations of various biologically and economically important plants species including *Solanum lycopersicum* L., *Sesamum indicum* L., *Brassica rapa*, *Arabidopsis*, *Helianthus annuus*, and *Citrus* (Liu *et al.*, 2013; Pacurar *et al.*, 2012; Hou *et al.*, 2010; Heesacker *et al.*, 2008; García-Lor *et al.*, 2012).

The model grass *B. distachyon* is a diploid species with numerous experimental tools including a high-quality reference genome sequence (IBI, 2010). Recently, a large set of InDel have been identified in this grass based on resequencing of six diverse *B. distachyon* accessions (Gordon *et al.*, 2015). In this study, we report, for the first time, on the application of InDel markers developed from *B. distachyon* for assessing the genetic variation in Tunisian *B. hybridum* populations. Since *B. hybridm* genome is the combination two subgenomes deriving from the two parental species *B. distachyon* and *B. stacei* (Catalan *et al.*, 2012), the 13 primers that generated an unintrepretable multiple bands pattern have most likely amplified the *B. stacei* subgenome. However, the eight retained InDel were found to be highly informative as they revealed an overall level of polymorphism of 84% across the studied populations. Such level is comparable to that detected in the same panel of *B. hybridum* lines using SSR markers (100%) (Neji *et al.*, 2015) and other *Brachypodium* species using various molecular markers (Bąba *et al.*, 2012a; Xinchun *et al.*, 2013; Jaroszewicz *et al.*, 2012). In addition, results showed much lower expected heteozygosity ($H_e = 0.35$) in comparison with that detected by 15 SSR markers in our previous study ($H_e = 0.79$) (Neji *et al.*, 2015). While consistent with some previous findings suggesting that InDel markers are less polymorphic than SSR markers (García-Lor *et al.*, 2012; Kladmook *et al.*, 2012), this result could be attributed to the modest number of makers used in this study. As such, the level genetic diversity revealed in this study remains relatively high and confirms the substantial phenotypic variation the Tunisian *B. hybridum* germplasm (Neji *et al.*, 2014) as it was similar to that observed in *B. sylvaticum* using SSR markers ($H_e = 0.4$) (Xinchun *et al.*, 2013) and slightly higher than that reported by Jaroszewicz *et al.*, (2012) in *B. distachyon* using random amplified polymorphic DNA markers ($H_e = 0.22$).

AMOVA analysis revealed that the molecular variance was greatest within populations (i.e., among individuals;

80%) compared with 20% among populations. These results indicate "moderate" genetic differentiation among Tunisian *B. hybridum* populations. Undoubtedly, this finding contradicts the common pattern of genetic structure for selfing species, in which major genetic diversity should occur among populations (Semagn *et al.*, 2000). However, it is in perfect agreement with what have recorded in our previous investigations on the same germplasm using morpho-phenological traits and SSR markers (Neji *et al.*, 2014; Neji *et al.*, 2015). Similar results were also detected in *B. distachyon* (Filiz *et al.*, 2009a), *Brachypodium pinnatum* (Bąba *et al.*, 2012b) and other selfing poaceae, such as *Elymus nutans* and *Elymus burchan-buddae* (Yan *et al.*, 2010) and *Oryza sativa f. spontanea* (He *et al.*, 2014).

In addition, our results showed that distance-based clustering method PCoA and cluster analysis and the Baysien inference implemented in STRUCTURE uncover similar pattern of three distinct groups without any geographical pattern. This finding was supported by the Mantel test, which showed that the genetic divergence was not attributed neither to the geographic distances nor the altitudinal differences between populations.

Seed-mediated gene flow or seed dispersal plays a key role in determining the level of spatial genetic structure within and among plant populations and may promote the genetic homogeneity among populations occurring under a heterogeneous environment (Ellstrand, 1999). Therefore, its events have often been used to explain the wide distribution range of plants. *Brachypodium* species are typically characterized by small hairy seeds and easily dispersed by animals and humans (Vogel *et al.*, 2009). Hence, a lot of investigations have highlighted the mojor role played by the long seed-mediated gene flow in determining the level of their spatial genetic structure (Vogel *et al.*, 2009; Filiz *et al.*, 2009b; Manzaneda *et al.*, 2012; Bąba *et al.*, 2012b).

It important to point out that the studied populations are geographically distant from each other and differ in a large number of variables such as macro- and micro-climate, topography, and soil type (Neji *et al.*, 2014). Moreover, the sampling sites are surrounded by pastoral lands heavily grazed by cattle and sheep. The movement of animals may promote considerable gene flow by seed dispersion. Consequently, the lack of strong genetic structure among Tunisian *B. hybridum* populations can easily be interpreted by the excess gene flow from grazing activities that promoted the seed dispersal over long distances, in addition to geographical conditions (e.g., strong winds) at the sampling sites. This was aptly translated into correspondingly high rate of gene flow revealed by Φ_{PT} and private alleles-based approaches ($N_m = 1.04$ and $N_m = 2.75$, respectively) and the lack of the environmental effects on the distribution of genetic variation (AMOVA analysis). Such findings prove that the long distance seed dispersal is a key driver of the current genetic structure of the Tunisian *B. hybridum* populations and confirm the wide range adaptation of this species and its high capability to occur under different environmental conditions (Lopez-Alvarez *et al.*, 2012).

Overall, the results of this study are consistent with our previous investigations using morpho-phenological traits and SSR markers (Neji *et al.*, 2014; Neji *et al.*, 2015) as they confirm the major role played by the long seed dispersal in shaping the spatial genetic structure of the Tunisian *B. hybridum* populations. However, the comparison between the overall level of the genetic divergence obtained in this study ($\Phi_{PT} = 0.2$) and the level of quantitative differentiation (Qst = 0.4) (Neji *et al.*, 2014) suggest that, along with the long seed dispersal, the genetic structure of Tunisian *B. hybridum* heavily draws from diversifying selection (Qst > Φ_{PT}). The findings are reminiscent of those obtained by (Huang *et al.*, 2014) in *Miscanthus floridulus/sinensis* complex and in wild sorghum species (Magomere *et al.*, 2015).

CONCLUSION

To sum up, this study reveals that InDel markers are fully adequate for characterizing genetic variation and the population structure of *B. hybridum*. Along with our early findings using morpho-phenotypic and SSR analyses, the results we provide an in-depth understanding of the pattern of genetic diversity of this new model species on a broad geographic scale as they confirm the high genetic variation of Tunisian *B. hybridum* and the role played by the long seed dispersal in its spatial genetic structure. Along with the results of SSR analysis, the results of this study will be applied for quantitative trait locus mapping and genome-wide association studies to identify gene regions as well as genes that are responsible for morpho-phenological variation.

ACKNOWLEDGMENTS

REFERENCES

Available from: http://www.2.unil.ch/popgen/softwares/fstat.htm. [Last accessed on 2015 Mar 17].

Available from: https://www.xlstat.com/fr. [Last accessed on 2015 Apr 20].

Bąba W, Kurowska M, Kompała-Bąba A, Wilczek A, Długosz A, Szarejko I. Genetic diversity of the expansive grass *Brachypodium pinnatum* in a changing landscape: Effect of habitat age. Flora 2012a;207:346-53.

Bąba W, Kurowska M, Kompała-Bąba A, Wilczek A, Długosz J, Szarejko I. Genetic diversity of populations of *Brachypodium pinnatum* LP Beauv: Expansive grass in a fragmented landscape. Pol J Ecol 2012b;6:3140.

Betekhtin A, Jenkins G, Hasterok R. Reconstructing the evolution of *Brachypodium* genomes using comparative chromosome painting. PLoS One 2014;9:e115108.

Boessenkool S, Austin JJ, Worthy TH, Scofield P, Cooper P, Seddon PJ, *et al*. Relict or colonizer? Extinction and range expansion of penguins in southern New Zealand. Proc R Soc B Biol Sci 2009;276:815-21.

Brkljacic J, Grotewold E, Scholl R, Mockler T, Garvin DF, Vain P, *et al*. *Brachypodium* as a model for the grasses: Today and the future. Plant Physiol 2011;157:3-13.

Catalan P, Müller J, Hasterok R, Jenkins G, Mur LA, Langdon T, *et al*. Evolution and taxonomic split of the model grass *Brachypodium distachyon*. Ann Bot 2012;109:385-405.

Catalan P, Chalhoub B, Chochois V, Garvin DF, Hasterok R, Manzaneda AJ, Mur LA, Pecchioni N, Rasmussen SK, Vogel JP, Voxeur A. Update on the genomics and basic biology of Brachypodium: International Brachypodium Initiative (IBI). Trends in Plant Sciences 2014;19:414-8.

Earl DA, von Holdt BM. Structure harvester: A website and program for visualizing structure output and implementing the Evanno method. Conserv Genet Resour 2011;4:359-361.

Ellstrand NC, Prentice HC, Hancock JF. Gene flow and introgression from domesticated plants into their wild relatives. Ann Rev Ecol Syst 1999;30:539-63.

Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: A simulation study. Mol Ecol 2005;14:2611-20.

Filiz E, Ozdemir BS, Budak F, Vogel JP, Tuna M, Budak H. Molecular morphological and cytological analysis of diverse *Brachypodium distachyon* inbred line. Genome 2009a;52:876-90.

Filiz E, Ozdemir BS, Tuna M, Budak H. Diploid *Brachypodium distachyon* of Turkey: Molecular and morphologic analysis. In: The Proceedings of the 5th International Symposium on the Molecular Breeding of Forage and Turf; 2009b. p. 83-9.

García-Lor AF, Navarro LL, Ollitrault P. Comparative use of indel and SSR markers in deciphering the interspecific structure of cultivated citrus genetic diversity; a perspective for genetic association studies. Mol Genet Genom 2012;287:77-94.

Geuna F, Toschi M, Bassi D. The use of AFLP markers for cultivar identification in apricot. Plant Breed 2003;122:526-31.

Goudet J. FSTAT, a Program to Estimate and Test Gene Diversities and Fixation Indices (Version 2932). 2002. Available from: http://www.unilch/izea/softwares/fstat.

Gordon SP, Priest H, Des Marais DL, Schackwitz W, Figueroa M., Martin J, *et al*. Genome diversity in Brachypodium distachyon: deep sequencing of highly diverse inbred lines. The Plant Journal 2014;79:361-74.

Hartl DL, Clark AG. Principles of Population Genetics Sunderland. Massachusetts: Sinauer Associates Inc.; 2007. p. 652.

He Z, Jiang X, Ratnasekera D, Grassi F, Perera U, Lu BR. Seed mediated gene flow promotes within- population genetic variation of weedy rice: Implication for weed management. PLoS One 2014;9:e112778.

Heesacker A, Kishore VK, Gao WX, Tang SX, Kolkman JM, Gingle A, *et al*. SSRs and INDELs mined from the sunflower EST database: Abundance polymorphisms and cross-taxa utility. Theor Appl Genet 2008;1177:1021-9.

Hou XH, Li LC, Peng ZY, Wei BY, Tang SJ, Ding MY, *et al*. A platform of high-density INDEL/CAPS markers for map-based cloning in *Arabidopsis.* Plant 2010;63:880-8.

Huang CL, Ho CW, Chiang YC, Shigemoto Y, Hsu TW, Hwang CC, *et al*. Adaptive divergence with gene flow in incipient speciation of *Miscanthus floridulus/sinensis* complex (Poaceae). Plant J 2014;80:834-47.

International *Brachypodium* Initiative. Genome sequencing and analysis of the model grass *Brachypodium* distachyon. Nature 2010;463:763-8.

Jaroszewicz M, Kosina R, Stankiewicz PR. RAPD karyology and selected morphological variation in a model grass *Brachypodium distachyon*. Weed Res 2012;52:204-16.

Kladmook M, Kumchoo T, Hongtrakul V. Genetic diversity analysis and subspecies classification of Thailand rice landraces using DNA markers. Afr J Biotechnol 2012;11:14044-53.

Liu B, Wang Y, Zhai W, Deng J, Wang H, Cui Y, *et al*. Development of InDel markers for *Brassica rapa* based on whole-genome re-sequencing. Theor Appl Genet 2013;1261:231-9.

Lopez-Alvarez D, Lopez-Herranz ML, Betekhtin A, Catalan P. A DNA barcoding method to discriminate between the model plant *Brachypodium distachyon* and its close relatives *B. stacei* and *B. hybridum* (Poaceae). PLoS One 2012;7:e51058.

Magomere TO, Eliud K, Ngugi S, Obukosia D, Eunice M, Shibairo OI. The population structure of wild sorghum species in agro-ecological zones of Western Kenya. Afr J Biotechnol 2015;14:1475-92.

Mantel N. The detection of disease clustering and a generalized regression approach. Cancer Res 1967;27:209-20.

Manzaneda AJ, Rey PJ, Bastida JM, Weiss-Lehman T, Raskin E, Mitchell-Olds T. Environmental aridity is associated with cytotype segregation and polyploidy occurrence in *Brachypodium distachyon* (Poaceae*)*. N Phytol 2012;193:797-805.

Neji M, Geuna F, Taamalli W, Ibrahim Y, Chiozzotto R, Abdelly C, *et al*. Assessment of genetic diversity and population structure of Tunisian populations of *Brachypodium hybridum* by SSR markers. Flora 2015;216:42-9.

Neji M, Geuna F, Taamalli W, Ibrahim Y, Smida M, Abdelly C, *et al*. Morphophenological diversity among Tunisian natural populations of *Brachypodium hybridum.* J Agric Sci 2014; 153:1006-16.

Pacurar DI, Pacurar ML, Street N, John DB, Tiberia IP, Laurent G, *et al*. A collection of INDEL markers for map-based cloning in seven *Arabidopsis* accessions. J Exp Bot 2012;63:2491-501.

Paetkau D, Slade R, Burden M, Estoup A. Genetic assignment methods for the direct real-time estimation of migration rate: A simulation-based exploration of accuracy and power. Mol Ecol 2004;3:55-65.

Peakall R, Smouse PE. GenAlEx 65: Genetic analysis in excel, population genetic software for teaching and research: An update. Bioinformatics 2012;2819:2537-9.

Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A. GeneClass2: A Software for genetic assignment and first-generation migrant detection. J Hydro-Environ Res 2004;95:536-9.

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;1552:945-59.

Rannala B, Mountain JL. Detecting immigration using multilocus genotypes. Proc Natl Acad Sci USA 1997;94:9197-201.

Raymond M, Rousset R. GENEPOP (version 12): Population genetics software for exact tests and ecumenicism. J Hydro-Environ Res 1995;86:248-9.

Semagn K, Bjørnstad A, Stedje B, Bekele E. Comparison of multivariate methods for the analysis of genetic resources and adaptation in *Phytolacca dodecandra* using RAPD. Theor Appl Genet 2000;101:1145-54.

Takezaki N, Nei N, Tamura K. POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. Mol Biol Evol 2010;274:747-52.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA 5: Molecular evolutionary genetics analysis using maximum likelihood evolutionary distance and maximum parsimony methods. Mol Biol Evol 2011;28:2731-9.

Vogel JP, Tuna M, Budak H, Huo N, Gu YQ, Steinwand MA. Development of SSR markers and analysis of diversity in Turkish populations of *Brachypodium distachyon.* BMC Plant Biol 2009;11:9-88.

Wellington F, Nascimento J, Rodrigues F, Samantha K, Paul G, Elizabeth A, *et al*. Spatially structured genetic diversity of the *Amerindian yam Dioscoreatrifida* L assessed by SSR and ISSR markers in Southern Brazil. Genet Resour Crop Evol 2013;12:123-36.

Wright S. Evolution and the Genetics of Populations the Theory of Gene Frequencies. Vol. 2. Chicago: University of Chicago Press; 1969.

Xinchun M, Ju G, Lizhi G. Characterization of microsatellite markers and their application to genetic diversity analysis of *Brachypodium sylvaticum* var *breviglume* from Yunnan China. Am J P Sci 2013;4:1427-34.

Yan XB, Guo YX, Zhang C, Liu FY, Lu BR. Population structure affected by excess gene flow in self-pollinating *Elymusnutans* and *E. burchan-buddae* Triticeaee: Poaceae. Pop Ecol 2010;52:233-41.