Genetic diversity and population structure of elite drought tolerant bread wheat (*Triticum aestivum* L.) genotypes

Sibusisanganye Y. Mduli1, Hussein Shimelis12 and Assefa B. Amelework1,2

1University of KwaZulu-Natal, African Centre for Crop Improvement, P/Bag X01, Scottsville 3209, Pietermaritzburg, South Africa
2Agricultural Research Council, Vegetable and Ornamental Plants, Private Bag X293, Pretoria, 0001, South Africa

*Corresponding author: Shimelish@ukzn.ac.za

Abstract

The objective of this study was to assess genetic diversity and population structure of 47 bread wheat genotypes obtained from the International Maize and Wheat Improvement Centre (CIMMYT) using 10 polymorphic simple sequence repeat (SSR) markers. Data was subjected to analysis for generating a dissimilarity matrix by the Jaccard index for clustering by the Neighbour-joining algorithm on DARwin 6.5 software. GenAlex Software was used to analyse the number of detected alleles (Na), number of effective alleles (Ne), observed heterozygosity (Ho), expected heterozygosity (He), genetic distance (GD), genetic identity (GI), gene flow (Nm), fixation index (F), Shannon’s Information Index (I), Analysis of molecular variance (AMOVA) and polymorphic information content (PIC). Results revealed that, at the locus level, marker Xgwm 132 had the highest Na (21), Ne (14.5), Ho (1.0) and He (0.94), while at the population level, Population III had the highest Na (21), Ne (5.59), He (0.83), and I (1.78). The mean PIC recorded was 0.80, and ranged from 0.63 (Wmc 78) to 0.93 (Xgwm 132). AMOVA revealed significant differences in genetic variation allocated within individuals (60%), across different individuals (37%) and across populations (3%) (P < 0.001). Four populations were distinguished based on pedigrees with GD ranging from 0.01 (Populations III and IV) to 0.31 (Populations II and III), while GI ranged from 0.74 (Populations II and III) to 0.99 (Populations III and IV). The selected markers successfully distinguished test genotypes with the most informative marker being Xgwm 132. Populations II and III were most distinct, thus suitable for parental selection and further drought tolerance breeding.

Keywords: bread wheat; cluster analysis; genetic variation; microsatellite markers; population structure; polymorphism.

Abbreviations: AMOVA_ Analysis of Molecular Variance; Fc_ Fixation index; GD_ Genetic distance; GI_ Genetic identity; He_ Expected heterozygosity; Hs_ observed heterozygosity; I_Shannon Information Index; N_d_ Number of detected alleles; Ne_ Number of effective alleles; Nm_ Gene flow; PIC_ Polymorphic Information Content.

Introduction

A vital foundation for all plant breeding programmes is an exceptional degree of genetic diversity, especially for future parental selection (Nielsen et al., 2014; Salem et al., 2015). Screening a set of germplasm for favourably broad variability is faster and more convenient with the help of DNA or molecular markers compared with morphological markers. Molecular markers are not reliant upon crop growth stages and they are independent to environmental variation (Govindaraj et al., 2015). Among the popular molecular markers are the simple sequence repeat (SSR) markers (Henry, 2001). SSRs are typically detected in eukaryotes (Roder et al., 1998). These markers are randomly scattered throughout the genome (Roder et al., 1998), flanking vital regions associated with a particular trait, depending on whether these regions are coding or non-coding regions (Henry 2001). SSR markers are favoured for their high polymorphism, co-dominant nature (Semagn et al., 2006), their basis on polymerase chain reaction (PCR) (Roder et al., 1998) and their relatively small quantity of DNA (50 nano-grams) requirement (Nadeem et al., 2018; Semagn et al., 2006). The varying number of repeats and repeat lengths is valuable for detecting polymorphism among individuals (Falconer and Mackay, 1996). The Hardy-Weinberg equilibrium states that in an infinitely large population of randomly mating individuals, the gene and genotype frequencies will remain constant from generation to generation (Falconer and Mackay, 1996). This law remains true in the absence of linkage and evolutionary change (Hart and Andrew, 1997), thus the Hardy-Weinberg equilibrium serves as a "null model" (Hamilton, 2009; Andrew, 2010). The Hardy-
Weinberg law is only functional in sexually reproducing species of non-overlapping generations (Falconer and Mackay, 1996; Hartl and Andrew, 1997; Andrew, 2010). In natural environments, the prerequisites of the Hardy-Weinberg may be violated, thus giving rise to unequal gene or genotype frequencies, causing significant population stratification and impacting the genetic diversity (Hamrick, 1982; Morjan and Rieseberg, 2004). According to Nielsen et al. (2014), the top factors that contribute to genotype subgrouping are differences in geographical origin, genetic drift, as well as human and environmental influences. In wheat, which is predominantly a self-fertilizing crop, the exchange of seed from breeder to breeder as well as farmer to farmer majorly influences the genetic diversity among other causes of genetic variation (Mwadzingeni et al., 2016). Bread wheat (Trityum aestivum L.; 2n=6x=42; AABBDD) has a large hexaploid genome of approximately 80% repetitive DNA sequences (Roder et al., 1998; Nielsen et al., 2014). SSR markers are well suited for genetic diversity studies in bread wheat for their genome specificity (Roder et al., 1998). In a study conducted by Tekeu et al. (2017), the genetic diversity and population structure of 17 Cameroonian accessions of bread wheat was investigated using 11 microsatellite markers. The set of markers were determined to be highly descriptive, with an average polymorphic information content (PIC) of 0.69, which ranged from 0.46 (Xgwm 125) to 0.90 (Xwmc 177) (Tekeu et al., 2017). In another study by Nezafoui et al. (2014), 16 Tunisian durum wheat accessions were examined for genetic diversity using 9 SSR markers distributed across the A and B genomes of bread wheat. The markers were found to be efficiently descriptive, particularly marker Xgwm 493 with the highest PIC value (0.55) and genetic diversity (0.63) (Nezafoui et al., 2014). Nezafoui et al. (2014) also reported that a small collection of markers can be used, provided that they are efficiently diagnostic in nature. According to Botstein et al. (1980), markers with PIC values greater than 0.5 are considered polymorphic and informative. Therefore, the objective of the current study was to assess the genetic diversity and population structure of 47 bread wheat genotypes sourced from CIMMYT using 10 selected diagnostic polymorphic SSRs for drought tolerance breeding. This study may provide vital information on the genetic composition of the test wheat genotypes for targeted hybridisation and drought tolerance breeding for the wheat industry in South Africa.

Results

Polymorphism and allele diversity of the SSR markers

The results of the genetic parameter estimates calculated using GenAlex software, based on the 10 SSR markers, are shown in Table 3. The $\chi^2$ analysis revealed highly significant variation of the allele frequencies of the amplified fragments, for which the major allele frequency detected per locus ranged from 0.10 to 0.55, with a mean of 0.29. The 10 SSR markers produced a total of 109 putative fragments from the 47 wheat genotypes. The amplified fragment size ranged from 118 to 397 base pairs (bp). The greatest size range among the alleles was observed on marker Wms 179, for which 181 bp difference was observed between the longest and shortest alleles. The mean Na was 10.9, while minimum and maximum values detected were 4 (Wmc 78) and 21 (Xgwm 132), respectively. On the other hand, Ne values ranged from 2.7 (Wmc 78) to 14.5 (Xgwm 132), and mean Ne was 6.3. The observed heterozygosity ranged from 0.02 (Wmc 78) and 1.00 (Wms 153, Xgwm 132, Wms 179, Wms 30), with a mean value of 0.50. Comparatively, expected heterozygosity was quite high, with a mean of 0.81, and minimum and maximum values of 0.64 (Wmc 78) and 0.94 (Xgwm 132), respectively. The marker Wms 153 (-0.22) revealed an excess of heterozygote alleles, whereas the highest $F_s$ value was 0.96 (Wmc 78) and mean $F_s$ was 0.41. The high PIC values recorded are indicative of the highly informative nature of the selected markers, with a mean of 0.8 and minimum and maximum values of 0.63 (Wmc 78) and 0.93 (Xgwm 132), respectively.

Genetic variation within and among populations

The results of the estimated genetic parameters for the 4 populations identified are shown in Table 4. The minimum and maximum Na values recorded were 5.9 (Population I) and 7.6 (Population III), respectively, whereas mean Na was 6.78. The mean Ne value was 4.94, and values ranged from 4.31 (Population II) to 5.59 (Population III). On the other hand, Shannon’s information index ranged from 1.46 (Population II) to 1.78 (Population III), while the mean was 1.62. The Ho was noticeably higher than He, with means of 0.50 and 0.78, respectively. Values for Ho ranged from 0.47 (Population IV) to 0.54 (Population II), whereas values for He ranged from 0.72 (Population II) to 0.83 (Population III). The number of private alleles detected was greatest for Population III (10), while the least number was recorded for Population I (5). The greatest genetic distance recorded was 0.31 (between Population II and Population III), while the least was 0.01 (between Populations III and IV), as shown in Table 5. Similarly, genetic identity was greatest between Populations III and IV (0.99) and the least between Populations II and III (0.74). Thus, revealing the strong relation between Population III and Population IV. The $F_{ST}$ ranged from 0.02 (Populations III and IV) to 0.07 (Populations II and III), revealing a narrow range and moderate differentiation (Wright, 1978) between Populations II and III. The gene flow was greatest between Populations III and IV (12.3) and the least between Populations II and III (4.7) (Table 5).

Cluster analysis

The dendrogram constructed for the 47 wheat genotypes based on the 10 SSR markers using Jaccard’s coefficient of dissimilarity is presented in Fig 1. The mean genetic distance detected between genotypes was 0.57, which was considerably high. This indicated a significant genetic variability between the different genotypes. The greatest dissimilarity values recorded were 0.86 (between SYM2016-037 and SYM2016-002) and 0.85 (between SYM2016-037 and
SYM2016-029), while comparatively low dissimilarities recorded were 0.02 (between SYM2016-026 and SYM2016-027, SYM2016-010 and SYM2016-009 and SYM2016-002 and SYM2016-029). Based on the dendrogram constructed, 3 major clusters were observed, namely A (denoted by black scheme), B (blue) and C (red) (Fig 1). These clusters were closely related as indicated by the proximity of clustering. Cluster A, B and C consisted of 21, 19 and 7 genotypes (Table 6).

Analysis of molecular variance (AMOVA)

The results of the analysis of molecular variance for the 4 populations based on parentage are presented in Table 7. Significant differences were detected for genetic variation allocated within individuals, across different individuals and different populations (P < 0.001) (Table 7). The greatest variation was assigned to genetic variation within the different genotypes (60%), while the rest was allocated to variation across the different individuals (37%) and variation across the different populations (3%).

Discussion

In the present study, genetic diversity and population structure of 47 elite CIMMYT bread wheat genotypes was investigated using 10 SSR markers to select the most diverse genotypes for downstream breeding. The 10 SSR markers amplified a total of 109 bands, with an average number of 10.9 alleles per locus. These values are much higher than those found in a genetic diversity study of different wheat populations (3%).

Analysis of molecular variance (AMOVA)

The results of the analysis of molecular variance for the 4 populations based on parentage are presented in Table 7. Significant differences were detected for genetic variation allocated within individuals, across different individuals and different populations (P < 0.001) (Table 7). The greatest variation was assigned to genetic variation within the different genotypes (60%), while the rest was allocated to variation across the different individuals (37%) and variation across the different populations (3%).

Discussion

In the present study, genetic diversity and population structure of 47 elite CIMMYT bread wheat genotypes was investigated using 10 SSR markers to select the most diverse genotypes for downstream breeding. The 10 SSR markers amplified a total of 109 bands, with an average number of 10.9 alleles per locus. These values are much higher than those found in a genetic diversity study of different wheat populations (3%).

Analysis of molecular variance (AMOVA)

The results of the analysis of molecular variance for the 4 populations based on parentage are presented in Table 7. Significant differences were detected for genetic variation allocated within individuals, across different individuals and different populations (P < 0.001) (Table 7). The greatest variation was assigned to genetic variation within the different genotypes (60%), while the rest was allocated to variation across the different individuals (37%) and variation across the different populations (3%).

Discussion

In the present study, genetic diversity and population structure of 47 elite CIMMYT bread wheat genotypes was investigated using 10 SSR markers to select the most diverse genotypes for downstream breeding. The 10 SSR markers amplified a total of 109 bands, with an average number of 10.9 alleles per locus. These values are much higher than those found in a genetic diversity study of different wheat populations (3%).
| Genotype          | Pedigree                                                                 |
|-------------------|--------------------------------------------------------------------------|
| SYM2016-001       | 1447/PASTOR                                                             |
| SYM2016-002       | ALTAR 84/AEGILOPS SQUARROSA (TAUS)//OPATA/3/ATTILA                    |
| SYM2016-003       | BABAX/3/PRL/SARA/TSI/VEE#5/4/CROC_1/AE.SQUARROSA (224)//2*OPATA        |
| SYM2016-004       | BABAX/3/PRL/SARA/TSI/VEE#5/4/WBL1                                      |
| SYM2016-005       | BAU/KAUZ//PASTOR                                                         |
| SYM2016-006       | BUC/MN72253/PASTOR                                                      |
| SYM2016-007       | CHIBIA/WEAVER                                                            |
| SYM2016-008       | CNDO/R143/ENTE/MEXI_2/3/AEGILOPS SQUARROSA (TAUS) /4'/ WEAVER/5/ 2*FRAME |
| SYM2016-009       | CROC_1/AE.SQUARROSA [213]//PGO/3/NG8319//SHA4/URA                        |
| SYM2016-010       | CROC_1/AE.SQUARROSA (205)//*BORL95/3/KENNEDY                            |
| SYM2016-011       | CROC_1/AE.SQUARROSA (224)//OPATA/3/RAC655                              |
| SYM2016-012       | CROC_1/AE.SQUARROSA (205)//KAUZ/3/SLVS                                  |
| SYM2016-013       | CROC_1/AE.SQUARROSA (224)//2*OPATA/3//2*RAC655                          |
| SYM2016-014       | D67.2/P66.270/AE.SQUARROSA (320)/L/CUNNINGHAM                          |
| SYM2016-015       | HD30/5/CNDO/R143/ENTE/MEXI75/3/AE.SQ/4/2*OCI                           |
| SYM2016-016       | JNRB.5/PIFED                                                            |
| SYM2016-018       | PASTOR/3/VEE#5//DOVE/BUC                                                  |
| SYM2016-019       | SLVS /6/ FILIN/ IRENA/5/ CNDO/ R143//ENTE/MEXI_2/3/AEGILOPS SQUARROSA (TAUS)/4'/ WEAVER |
| SYM2016-020       | SRN/AE.SQUARROSA (358)//MILAN/SHA7                                      |
| SYM2016-021       | SW89.5277/BORL95//SKAUZ                                                  |
| SYM2016-023       | SW94.6002/4/KAUZ*2//DOVE/BUC/3/KAUZ/5/SW91-12331                         |
| SYM2016-025       | TIE CHUAN 1*2/3/HE1/3*CND79/2*SERI                                     |
| SYM2016-026       | VEE#8//JUP/BY/3/F3.71/2*WEAVER/5/CNDO/R143//ENTE/MEXI_2/3/AEGILOPS SQUARROSA (TAUS)/4'/ WEAVER/6/WEAVER |
| SYM2016-027       | WORRAKATTA/2*PASTOR                                                      |
| SYM2016-028       | LOCAL CHECK                                                              |
| SYM2016-029       | CHAM 6                                                                   |
| SYM2016-030       | KLEIN CHAMACO                                                           |
| SYM2016-031       | HIDHAB                                                                   |
| SYM2016-032       | DHARWAR DRY                                                              |
| SYM2016-033       | FRTL/CMH83.2S17                                                          |
| SYM2016-034       | SARA/THB/VEE/3//BY/COG//PRL/BOW                                         |
| SYM2016-035       | PASTOR/FLORKWA.1/ PASTOR                                                 |
| SYM2016-036       | CHAM6/ATTILA/PASTOR                                                      |
| SYM2016-037       | CROC_1/AE.SQUARROSA (224)//OPATA/3/PASTOR/4/PASTOR*2//OPATA             |
| SYM2016-038       | CROC_1/AE.SQUARROSA (224)//OPATA/3/ALTAR 84/AEGILOPS SQUARROSA (TAUS)//OPATA/4/PASTOR |
| SYM2016-039       | CROC_1/AE.SQUARROSA (224)//OPATA/3/ALTAR 84/AEGILOPS SQUARROSA (TAUS)//OPATA/4/PASTOR |
| SYM2016-040       | CROC_1/AE.SQUARROSA (224)//OPATA/3/ALTAR 84/AEGILOPS SQUARROSA (TAUS)//OPATA/4/PASTOR |
| SYM2016-041       | ATTILA/PASTOR/PASTOR                                                     |
| SYM2016-042       | ATTILA//PGO/SERI/3/PASTOR                                                |
| SYM2016-043       | PASTOR/TODY/BAU/3/PASTOR                                                 |
| SYM2016-044       | ALTAR 84/AE.SQ/2*OPATA/3//PASTOR                                         |
| SYM2016-045       | KRICAUFF/2*PASTOR                                                        |
| SYM2016-046       | KABY/2*ALUBUC/BAYA                                                       |
| SYM2016-047       | ALTAR 84/AEGILOPS SQUARROSA (TAUS)//OCI/3/VEE/MII/2*TUI                 |
| SYM2016-048       | ALTAR 84/AEGILOPS SQUARROSA (TAUS)//OCI/3/VEE/MII/2*TUI                 |
| SYM2016-049       | ALTAR 84/AEGILOPS SQUARROSA (TAUS)//OCI/3/VEE/MII/2*TUI                 |
| SYM2016-050       | MILAN/KAUZ/PRINIA/3/BABAX                                                |
Fig 1. Dendrogram for the 47 wheat genotypes based on Jaccard’s coefficient of dissimilarity.
### Table 2. List of 10 wheat SSR markers used in the current study.

| Markers | Marker sequences | AT(°C) | PIC  | References               |
|---------|-----------------|--------|------|--------------------------|
| Wmc 177| F: AGGGCTCTCTTTAATTTGCT  
          R: GGCTCATGTAATACACCTGTA | 51     | 0.94 | Somers and Isaac (2004)  |
| Wmc 78 | F: AGTAAATCCTCCCTCGGCTTC  
          R: AGCTTCTTTGCTAATCCTGCTG | 61     | 0.93 | Roder et al. (1998)      |
| Wms 30 | F: ATCTTAGCATAGAAGGAGTGGG  
          R: TTCTGACCGCTGGTGAT       | 94     | 0.92 | Roder et al. (1998)      |
| Wms 169| F: ACCACTGCAAGAAACACATACG  
          R: CATGATCTTGCTTGATGCTG | 94     | 0.90 | Roder et al. (1998)      |
| Wms 304| F: AGGAAACAGAAATATCGCGG  
          R: AGGACTGTGGGGGAATGAATG | 94     | 0.91 | Roder et al. (1998)      |
| Wmc 179| F: CATGGTGGCCATGAGTGGAGGT  
          R: CATGATCTTGCTTGATGCTG | 61     | 0.87 | Somers and Isaac (2004)  |
| Xgwm 132| F: TACCAAATCGAAACATCAGG  
           R: CATATCAAGGTCTTCCCC     | 94     | 0.99 | Roder et al. (1998)      |
| Wmc 532| F: GATACATCAAGATCGTGCCAAA  
          R: GGGAGAAATCATTAACGAAGGG | 61     | 0.96 | Somers and Isaac (2004)  |
| Xgmw 484| F: ACATCGCTCTTCACAAACCC  
           R: AGTTCCGGTCATGGCTAGG    | 94     | 0.89 | Roder et al. (1998)      |
| Wmc 153| F: ATGAGGACTCGAAGGCTTGGG  
          R: CTGAGCCTTTTGCGTGTTGAC  | 61     | 0.87 | Somers and Isaac (2004)  |

Key: F=forward primer; R=reverse primer; AT=annealing temperature (°C); PIC=polymorphic information content.

### Table 3. Genetic parameters for the 47 wheat genotypes in the current study.

| Markers | Na  | Ne  | Ho  | He  | FIS | PIC  | A   | Asr  |
|---------|-----|-----|-----|-----|-----|------|-----|------|
| Wmc 177| 6.00| 3.99| 0.04| 0.76| 0.94| 0.75 | 0.34| 199-212|
| Wmc 78 | 4.00| 2.70| 0.02| 0.64| 0.96| 0.63 | 0.55| 248-279|
| Wms 30 | 12.00| 8.93| 1.00| 0.90| -0.13| 0.89 | 0.16| 233-255|
| Wms 169| 13.00| 5.49| 0.30| 0.83| 0.64| 0.82 | 0.30| 207-245|
| Wms 304| 9.00| 3.22| 0.15| 0.70| 0.78| 0.69 | 0.40| 216-238|
| Wms 179| 13.00| 9.11| 1.00| 0.90| 0.72| 0.89 | 0.13| 216-397|
| Xgwm 132| 21.00| 14.53| 1.00| 0.94| -0.07| 0.93 | 0.10| 118-159|
| Wmc 532| 7.00| 4.41| 0.04| 0.78| 0.94| 0.77 | 0.33| 176-199|
| Xgwm 484| 12.00| 5.13| 0.47| 0.81| 0.42| 0.80 | 0.39| 164-197|
| Wms 153| 12.00| 5.55| 1.00| 0.83| -0.22| 0.82 | 0.23| 155-202|
| Mean   | 10.90| 6.31| 0.50| 0.81| 0.41| 0.80 | 0.29|       |
| SE     | 1.50 | 1.14| 0.14| 0.03| 0.16| 0.03 | 0.04|       |

Key: Na = number of alleles per locus; Ne = number of effective alleles per locus; Ho = observed gene diversity within genotypes; He = expected gene diversity within genotypes; FIS = genetic differentiation; PIC = polymorphic information content; A = major allele frequency; Asr = allele size range (base pairs); SE = standard error.
Table 4. Genetic parameters for the 4 wheat populations.

| Populations     | N   | Na  | Ne  | I   | Ho  | He  | Pa  |
|-----------------|-----|-----|-----|-----|-----|-----|-----|
| Population I    | 9.00| 5.90| 4.59| 1.53| 0.51| 0.77| 5.00|
| Population II   | 12.00| 6.20| 4.31| 1.46| 0.54| 0.72| 8.00|
| Population III  | 12.00| 7.60| 5.59| 1.78| 0.49| 0.83| 10.00|
| Population IV   | 14.00| 7.40| 5.28| 1.72| 0.47| 0.80| 9.00|
| Mean            | 11.50| 6.78| 4.94| 1.62| 0.50| 0.78| -   |

SE

Key: N = number of genotypes per population; Na = mean number of alleles per locus per population; Ne = number of effective alleles per locus per population; Ho = mean observed gene diversity within genotypes per population; He = mean expected gene diversity within genotypes per population; I = Shannon’s information index; Pa = number of private alleles; SE = standard error.

Table 5. Pairwise estimates of genetic differentiation, gene flow, genetic distance and genetic identity for 4 wheat populations.

| Populations | Population I | Population II | Population III | Population IV |
|-------------|--------------|---------------|----------------|---------------|
| Population I| 0.06 (4.7)   | 0.04 (4.8)    | 0.05 (4.8)     |               |
| Population II| 0.21 (0.81) | 0.07 (4.7)    | 0.05 (4.8)     |               |
| Population III| 0.13 (0.88) | 0.31 (0.74)   | 0.02 (12.3)    |               |
| Population IV| 0.17 (0.84) | 0.18 (0.83)   | 0.01 (0.99)    |               |

Key: GD = genetic distance (bottom diagonal outside brackets); GI = genetic identity (bottom diagonal within brackets); F<sub>ST</sub> = genetic differentiation (top diagonal outside brackets); N<sub>m</sub> = gene flow (top diagonal within brackets).

Table 6. Clustering patterns of 47 wheat genotypes based on the Jaccard’s coefficient of dissimilarity.

| Cluster (number of genotypes) | Genotype names |
|------------------------------|----------------|
| A (21)                       | SYM2016-043, SYM2016-005, SYM2016-037, SYM2016-023, SYM2016-038, SYM2016-040, SYM2016-049, SYM2016-029, SYM2016-025, SYM2016-032, SYM2016-010, SYM2016-050, SYM2016-031, SYM2016-002, SYM2016-012, SYM2016-011, SYM2016-020, SYM2016-009, SYM2016-030, SYM2016-046, SYM2016-018 |
| B (19)                       | SYM2016-044, SYM2016-004, SYM2016-028, SYM2016-016, SYM2016-019, SYM2016-015, SYM2016-035, SYM2016-014, SYM2016-036, SYM2016-008, SYM2016-006, SYM2016-033, SYM2016-007, SYM2016-021, SYM2016-013, SYM2016-034, SYM2016-048, SYM2016-047, SYM2016-039 |
| C (7)                        | SYM2016-042, SYM2016-045, SYM2016-041, SYM2016-003, SYM2016-001, SYM2016-026, SYM2016-027 |

Table 7. Results of the analysis of molecular variance of the 4 populations of wheat genotypes.

| Sources of variation | Df | SS  | MS  | Estimated Variation | Percentage variance | Significance levels |
|----------------------|----|-----|-----|---------------------|---------------------|---------------------|
| Among populations    | 3  | 24.106 | 8.035 | 0.106               | 3%                  | F<sub>ST</sub> ≤ 0.001 |
| Among individuals    | 43 | 239.319 | 5.566 | 1.533               | 37%                 | F<sub>ST</sub> ≤ 0.001 |
| Within individuals   | 47 | 117.500 | 2.500 | 2.500               | 60%                 | F<sub>IT</sub> ≤ 0.001 |
| Total                | 93 | 380.926 | 4.139 | 100%                |                      |                     |

Key: Df = degrees of freedom; SS = sum of squares; MS = mean squares.
Therefore, genetic diversity among the wheat populations in the present study was low. The population stratification can be caused by geographical isolation of a group of individuals, artificial and natural selection as well as genetic drift (Nielsen et al., 2014). The pedigree stratification used in this study might not be sufficient to classify genotypes into sub-populations. The commonly shared parents or progenitors in the current study were Pastor, Altar 84, Aegilops squarrosa and Piﬁed, which contributed to limited genetic variability between the populations. Nefzaoui et al. (2014) and Henkar et al. (2016) reported genetically related progenitors yielding a relatively low genetic distance among populations. Hence, Populations II and III, in the current study, are the most genetically divergent, having expressed the highest genetic distance and the least genetic identity. Populations II and III are prime candidates to retrieve potential crossing parents. In contrast, Populations III and IV expressed the least genetic diversity. Soriano et al. (2016) indicated that linkage disequilibrium can be a consequence of an uneven frequency of alleles within different populations of genotypes. Therefore, further investigation of linkage disequilibrium in the current collection of bread wheat genotypes should be investigated.

Based on the dendrogram constructed, 3 major clusters were observed: A, B and C (Fig 1). These clusters were closely related as indicated by the proximity of clustering. Clusters A, B and C consisted of 21, 19 and 7 genotypes (Table 5). From these clusters, 15 bread wheat genotypes were selected as part of downstream breeding. These best accounted for the existing variation and would be potentially suitable parents in hybridisation. The genotypes selected were SYM2016-037, SYM2016-038, SYM2016-029, SYM2016-010, and SYM2016-012 from Cluster A, SYM2016-044, SYM2016-004, SYM2016-016, SYM2016-019, SYM2016-014, SYM2016-008, SYM2016-006, and SYM2016-047 from Cluster B and SYM2016-042 and SYM2016-027 from Cluster C (Table 6).

**Materials and methods**

**Plant materials and study sites**

The study used 47 elite bread wheat lines (Table 1) that were selected from 100 accessions based on their agronomic performance and adaptation under South African growing conditions. All the genotypes were breeding lines developed for drought tolerance by the International Maize and Wheat Improvement Centre (CIMMYT). The genotypes were introduced in to South Africa for selection of genetically unique parents for further breeding. Genotypes were grouped into 4 populations according to their pedigree relationships: Population I were single crosses involving Pastor in their parentage, Population II were genotypes resulted from crosses with Aegilops squarrosa as common parent, Population III composed of crosses derived with different parentages and Population IV comprised of a mixture of lines.

**DNA extraction, purification and quantification**

Seeds of the tested wheat genotypes were planted in 5L plastic pots at the Controlled Environmental Facility (CEF), University of KwaZulu-Natal, Pietermaritzburg, South Africa. Young fresh leaves were harvested from 20 plants per genotype two weeks after planting. The leaf samples were sent to INCOTEC PROTEIOS Laboratory (INCOTECH South Africa Pty Ltd, Pietermaritzburg, South Africa) for SSR analysis. DNA extraction was done following cetyl trimethylammonium bromide (CTAB) method as described by (Pask et al., 2005). The concentration of the extracted DNA was determined using 0.7% Tris-borate EDTA agarose gel. The extracted DNA was standardized using a working concentration of 10 ng·µL⁻¹ (Arora et al., 2014). The samples were bulked and used in SSR amplification.

**Polymerase chain reaction and SSR analysis**

All samples were used in bulked amplification using DNA extracted from 20 individual leaf samples. Ten SSR markers were used in this study, which were selected based on their high PIC (Table 2). These SSR markers are detailed and listed in the Grain Genes Database for Triticeae and Avena species as a tool for genetic studies in wheat (GrainGenes, 2018, http://wheat.pw.usda.gov; Roder et al., 1998; Somers and Isaac, 2004). PCR was done using 12 µL of reaction mixture containing 1X PCR buffer, 2.5 mM Mg²⁺, 0.2 µL each of dNTPs (Bioline) 1 U of Taq polymerase (Bioline) and 5-10 ng of genomic DNA. Four fluorescent dyes were used to label the primers. The initial denaturation step was performed at 94°C for 2 minutes, followed by 33 cycles at 94°C for 30 seconds. In addition, annealing was done at 63°C for 30 seconds and 72°C for 45 seconds with a final extension for 20 minutes. Polymerase chain reactions products were fluorescently labelled and separated by capillary electrophoresis using an ABI 3130xl automatic sequencer (Applied Biosystems, Johannesburg, South Africa) and the analysis was performed using GeneMapper 4.1 (Applied Biosystems, Johannesburg, South Africa).

**Data analysis**

Two approaches were adopted to investigate the genetic structure and diversity among the wheat genotypes. In the first approach, polymorphisms were treated as binary data (presence or absence). In this case, each amplified fragment was considered as one locus. However, to determine the genetic structure within and among genotypes, a second approach based on the codominant nature of the marker was adopted. Genotypic data were subjected to analyses with various measures of genetic diversity within and among genotypes using GenAlex software version 6.5 (Peakall and Smouse, 2012). Chi-square (χ²) test was performed to determine the different allele frequencies among the SSR markers.
Genetic parameters such as number of detected alleles (N_a), effective alleles (N_e), observed (H_o) and expected heterozygosity (H_e), genetic distance (GD), genetic identity (GI), gene flow (N_m), fixation index (F), Shannon’s Information Index (I), analysis of molecular variance (AMOVA) and polymorphic information content (PIC) were used to describe the genetic structure of the wheat genotypes. The number of polymorphic loci detected was analysed according to genotypes’ parental origin using their respective pedigrees. The marker PIC was calculated using the following formula:

\[ PIC = 1 - \frac{1}{\Sigma p_i j^2} - \left[ \Sigma (p_i j^2) \right]^2 + \Sigma \left\{ (p_i j^2) \right\}^2 \]

Where, the \( p_i j \) represents the frequency of the \( j \)th alleles upon the \( i \)th locus. Nei’s unbiased genetic distance was computed by employing GenAlEx software. The genetic diversity formula is given as: \( GD = 1 - \Sigma p_i j^2 \) (Nei, 1973). The method described by Merimans (2006) was employed to generate the FST, genetic differentiation. Also, the analysis of molecular variance (AMOVA) was done for establishing total genetic variation partitioning using GenAlEx software.

For the cluster analysis, a dissimilarity matrix was constructed from a binary data using the Jaccard’s dissimilarity index. The generated matrix was used to form genetic relationships based on neighbour-joining algorithm using the unweighted pair group mean arithmetic (UPGMA). Bootstrap analysis was done for accurate node construction whereby the bootstrap value was set at 10,000 bootstrap values. The software utilised for cluster analysis was DARwin 6.5 (Perrier and Jacquemoud-Collet, 2006).

Conclusion

The current study attempted to assess the genetic diversity and population structure of 47 elite bread wheat genotypes using 10 SSR markers. The selected markers in the current study exhibited high polymorphism and were effective in discriminating the test genotypes, of which the most polymorphic marker was Xgwm 132. The major sources of genetic diversity were private alleles, especially of Population III, also heterozygosity within individuals than among populations. Hence, it can be concluded the tested accessions in the present study exhibit a high potential for segregation. The reserved differences between populations can be related to genotypes being introduced from the same source and related pedigree. In the current study, Populations II and III were considered to be genetically distinct, thus favourable for selection of desirable parents for breeding. As a result, 15 bread wheat genotypes were selected, as part of downstream breeding, from the pool of 47 bread wheat genotypes. The genotypes selected were SYM2016-037, SYM2016-038, SYM2016-029, SYM2016-010, and SYM2016-012 from Cluster A, SYM2016-044, SYM2016-004, SYM2016-016, SYM2016-019, SYM2016-014, SYM2016-008, SYM2016-006, and SYM2016-047 from Cluster B and SYM2016-042 and SYM2016-027 from Cluster C.

Acknowledgements

The authors thank the African Centre for Crop Improvement (ACCI), National Research Foundation (NRF) of South Africa and the University of KwaZulu-Natal for financial support of the study.

References

Abdellatif KF, Abouzeid HM (2011) Assessment of genetic diversity of Mediterranean bread wheat using Randomly Amplified Polymorphic DNA (RAPD) markers. JGEB. 9:157-163.
Amelework B, Shimelis H, Tongoona P, Laing M, Mengistu F (2016) Genetic diversity of lowland sorghum landraces assessed by morphological and microsatellite markers. Aust J Crop Sci. 3:291-298.
Andrews CA (2010) The Hardy-Weinberg Principle. Nature. 3:65.
Arora A, Kundu S, Dilbaghi N, Sharma I, Tiwari R (2014) Population structure and genetic diversity among Indian wheat varieties using microsatellite (SSR) marker. Aust J Crop Sci. 8:1281-1289.
Botstein D, White RL, Skolnick M, Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet. 32:314-331.
Desta Z, Orabi J, Jahoor A, Backes G (2014) Genetic diversity and structure found in samples of Eritrean bread wheat. Plant Genet Resour-C. 12:151-155.
Dresigacker S, Zhang P, Warburton ML, Skovmand B, Hosington D, Bohn M, Melchinger AE (2004) SSR and pedigree analyses of genetic diversity among CIMMYT wheat lines targeted to different mega environments. Crop Sci. 44:381-388.
Falconer DS, Mackay TF (1996) Introduction to quantitative genetics. Pearson/Prentice Hall New York.
Govindaraj M, Vetrivenththan M, Srinivasan M (2015) Importance of genetic diversity assessment in crop plants and its recent advances: an overview of analytical perspectives. Genet Res Int. 2015: 1-14.
GrainGenes (2018) A database for triticeae and avena species. Access date: 2017-01-12 http://wheat.pw.usda.gov.
Hamilton MB (2009) Population genetics. John Wiley and Sons, United Kingdom.
Hamrick JL (1982) Plant population genetics and evolution. Am J Bot. 69:1685-1693.
Hartl DL, Clark AG, Clark AG (1997) Principles of population genetics. Sunderland: Sinauer Associates, United States of America.
Henkrar F, El-Haddouy J, Ouabhou H, Nsarellah N, Iraqi D, Bendaou N, Udupa SM (2016) Genetic diversity reduction in improved durum wheat cultivars of Morocco as revealed by microsatellite markers. Sci Agri. 73: 134-141.
Henry RJ (2001) Plant genotyping: the DNA fingerprinting of plants. CAB.
Triticum aestivum

Laido G, Mangini G, Taranto F, Gadaleta A, Blanco A, Cattivelli L, Marone D, Mastrangelo AM, Papa R, De Vita P (2013) Genetic diversity and population structure of tetraploid wheats (Triticum turgidum L.) estimated by SSR, DARt and pedigree data. Plos One. 8:1-17.

Liu J, Liu L, Hou N, Zhang A, Liu C (2007) Genetic diversity of wheat gene pool of recurrent selection assessed by microsatellite markers and morphological traits. Euphytica. 155: 249-258.

Merimas PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. Evolution. 60:2399-2402.

Morjan CL, Rieseberg LH (2004) How species evolve collectively: implications of geneflow and selection for the spread of advantageous alleles. Mol Ecol. 13:1341-1356.

Mwadzingeni L, Shimelis H, Laing MD, Tsilo TJ (2016) Breeding for wheat drought tolerance: progress and technologies. J Integr Agr. 15:935-943.

Mwale VM, Tang X, Chilembwe E (2016) Assessment of genetic diversity among sixty bread wheat (Triticum aestivum) cultivars using microsatellite markers. Afr J Biotechnol. 15:960-973.

Nadeem MA, Nawaz MA, Shahid MQ, Dogan Y, Comertpay G, Yildiz M, Hatipoglu R, Ahmad F, Alsaleh A, Labhane N, Ozkan H, Chung G, Baloch FS (2018) DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnol Biotech Eq. 32:261-285.

Nezfaou M, Udupa SM, Ghabi MS, Bouhadida M, Iraqi D (2014) Molecular diversity of Tunisian durum wheat accessions based on microsatellite marker analysis. Rom Agric Res. 31: 33-39.

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA. 70:3321-3323.

Nielsen NH, Backes G, Stouggaard J, Anderson SU, Jahoor A (2014) Genetic diversity and population structure analysis of European hexaploid bread wheat (Triticum aestivum L.) varieties. Plos One. 9:1-13.

Pask AID, Pietragalla J, Mullan DM, Reynolds MP (2005) Laboratory protocols: Applied molecular genetics laboratory. International Maize and Wheat Improvement Center, Mexico.

Peakall R, Smouse PE (2012) GenAlex 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. Bioinformatics: Applications note 28:2537-2539.

Perrier X, Jacquemoud-Collet JP (2006) DARWin software. Access date: 2016-03-13 http://www.darwin.cirad.fr/darwin.html.

Plaschke J, Ganal MW, Roder MS (1995) Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor Appl Genet. 91: 1001-1007.