

Research Article

Open Access



¹ Biotechnology Engineering Dept/Technological Engineering Faculty/University of Aleppo, Aleppo, Syria.

² General Commission for Scientific Agricultural Research (GCSAR)/Ministry of Agriculture, Aleppo, Syria.

³ Department of Genome Mapping, Molecular Genetics and Genome Mapping Laboratory, Agricultural Genetic Engineering Research Institute, Giza, Egypt.

⁴ International Center for Agricultural Research in the Dry Areas (ICARDA), Cairo, Egypt.

Contacts of Authors



* To whom correspondence should be addressed: Laila D. Nahas.

Citation:

Received: 7 April, 2020

Accepted: 3 May, 2020

Published: 10 May, 2020

Copyright: © 2020 Nahas et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and supplementary materials

Funding: The authors have no support or funding to report.

Competing interests: The authors declare that they have no competing interests.

Characterization of EST-SSR markers in bread wheat EST related to drought tolerance and functional analysis of SSR-containing unigenes

Laila Dabab Nahas^{1,2*}, Alsamman M. Alsamman³, Aladdin Hamwieh⁴, Naim Al-Husein², Ghinwa Lababidi¹

Abstract

Bread wheat (*Triticum aestivum*) is an important staple food around the world. The enormous volume of the genome of wheat makes it quite slow to progress in traditional scientific research. On the other hand, incessant databases and suitable tools on web sites make progress in wheat research quicker and easier. Drought is a major abiotic stress in accordance with weather changes and accelerated increase in drylands. In this study, 9077 ESTs related to drought tolerance in hexaploid wheat were downloaded from NCBI and assembled into 12062 contigs and 4141 singletons. It was found that trinucleotide had the highest frequency 64.71%. Moreover, 53.80% of SSRs found in coding regions in respect of ORFs. The highest amino acids found for tri- and hexanucleotides were Arginine. In addition, 81% of SSR-containing unigenes had one chromosome location and the highest number of loci was found in chromosomes 1B (69). The distribution of genic SSR loci among the 21 wheat chromosomes, the three subgenomes, and the seven homoeologous groups of wheat chromosomes was significant, with $P < 0.01$ indicating a non-random distribution. Functional annotation and characterization of SSR-containing unigenes have been performed. Eighty-six sequences were identified and sorted into 25 putative TF families and establish 166 pathways using KEGG. Primer-BLAST was used to predict the polymorphism, which was 39% of the 63 primer pairs of SSR markers. Our current study attempts to help farmers in wheat breeding programs to have drought-tolerant accessions, particularly in developing countries.

Keywords: Drought, wheat, EST-SSR, *in silico*, computational.

Introduction

Bread wheat (*Triticum aestivum*) is the most widely grown cereal crop around the world and is cultivated extensively in dry regions. It resulted from hybridization between cultivated tetraploid wheat (AABB, *T. dicoccoides*) and diploid goatgrass (DD, *Aegilops tauschii*) about 8,000 years ago [1], so, it is composed of three subgenomes, A, B, and D, every subgenome has 7 chromosomes, $n = 21$. Therefore, seven homoeologous groups are organized; each homoeologous set has three similarly associated chromosomes from each subgenome. Moreover, the A, B, and D subgenomes have been determined their progenitors [2]. Bread wheat is the enormous volume of the wheat genome about 17 GB [3] which is a consequence of polyploidy and the existence of areas with repeat motifs, make it hard to sequence the complete genome until recently, so presently there is an opportunity for the research to make real advances in this field [4].

Drought is a serious environmental stress that threatens the productivity of wheat around the world. With frequent episodes of drought, global climate models predict changing patterns of precipitation. Although drought reduces wheat production at all stages of growth, it is more critical during the phases of flowering and grain filling (terminal drought) and results in significant yield losses [5]. However, drought tolerance is a complicated trait that is managed by multiple genes, each with tiny effects and disclosure of genes during water stress is significant to recognize their responses [6]. Drought induces many kinds of responses including modifications in gene expression, the synthesis of specific proteins (e.g., chaperone proteins), and the accumulation of osmotically compounds [7].

Plant breeding has been very effective in the production of improved varieties using traditional tools and methodologies. The availability of databases and bioinformatics tools is now contributing to a new revolution in plant breeding, as they are improving genotype research and its relationship with the phenotype, especially for complex traits. Next Generation Sequencing (NGS) technology generates a wide range of genomic data. Through analyzing NGS data via bioinformatics advances, new genes and regulatory sequences and their locations are discovered and large sets of molecular markers are made available. All tools and databases promote studying and enable markers associated with genes and QTLs to be identified using a variety of techniques such as bulked segregant analysis (BSA), association mapping, or fine genetic mapping [8].

The RNA sequencing is a valuable way that could be used to examine genome coding regions in particular conditions like biotic and abiotic stresses which facilitates and focus on the genes expressed in that time. The SSRs generated from EST have some substantial gains over genomic SSRs. Their development is cheap, they are adjacent or in the transcribed genes, by a homology search, the putative function can often be detected, the ability to be transferred to close species and the ESTs have an important higher frequency of microsatellites than in genomic DNA over species [9].

A number of studies have been conducted on ESTs or RNaseq expressed in wheat drought stress e.g. [10–12]. Recently, the interest has altered from ordinary SSRs (genomic) to EST-SSRs because of the release of huge databases of ESTs from all of the plants especially crops. There are many studies in EST-SSR such as wheat [13–15], hordeum [16], chickpea [17], olives [18], mustard [19], and rice [20].

In this study, an in-depth survey, using computational approaches for searching SSR markers related to drought tolerant in hexaploid wheat, detecting their positions with respect to ORFs (open reading frames), identifying tri- and hexanucleotide amino acids, and predicting polymorphism, was conducted after downloading and assembling ESTs related to drought tolerance. The function of SSR-containing unigenes was characterized and chromosome localization was detected and the random distribution of EST-SSR mapped loci was measured using statistical tests. Finally, primers were designed and the best of them were chosen using in-silico PCR. To predicate SSR markers polymorphism Primer-BLAST was used.

Materials and Methods

Retrieving data, cleaning, and assemblage

A total of 9077 TSEs were downloaded from the GeneBank database, linked to *Triticum aestivum* and some wild relatives, expressed in dry conditions from the diverging tissue (leaf, root and seed) of the plant. Then, the data was pre-processed to reduce the errors. The sequences were cleaned and trimmed for contamination, poor quality and short sequences using the EGAssembler web tool [21]. Next, in order to reduce redundancy, the processed ESTs were clustered and assembled into contigs using the Cap3 [22] program. The remaining ESTs without assembly resulted in singletons.

Mining of SSR, the position of SSR in respect to ORF, and amino acids

All contigs and singletons engendered in the study were used to search for SSR using the MISA Perl script (<http://pgrc.ipk-gatersleben.de/misa/>). The parameters used were 5, 4, 4, 3, 3 repeat units for di-, tri-, tetra-, penta-, and hexanucleotide, respectively. The ORF Finder in the NCBI detected the position of the SSRs corresponding to the coding regions. In addition, all amino acids for tri- and hexanucleotide SSR have been identified.

Functional characterization of unigenes containing SSR

All SSR-containing unigenes were analyzed using the PlantTFcat online tool (<http://plantgrn.noble.org/PlantTFcat>) for TFs. Gene ontology (GO) was analyzed using Blast2GO v 2.5 [23]. Whereas, these unigenes were first translated in all reading frames. Then, compared with the protein sequence database (NCBI nr) to identify potential translation products using BLASTX. All unigenic sequences containing SSR homologous with annotated proteins in the nr database have been selected for functional characterization. According to the GO vocabulary, the EST-contigs sequences were then categorized into three groups, i.e. molecular function, cellular component and biological process. While, KEGG pathways were identified by the online KAAS [24], (<http://www.genome.jp/kegg/kaas>) using SBH (single-directional best hit) method.

Chromosome localization

The positions of all unigenes containing SSR were determined using the downloaded IWGSC Reference Sequence v1.0 assembly and a stand-alone blat [25]. Blat produces alignment at the DNA level between two sequences that were of 90% or higher similarity, but which may contain large inserts. The criteria used here for the blat were $-\text{minidentity}=97\%$ and $-\text{minscore}=400$. In addition, it was excluded all alignments that match less than 100 bp. For the primers design, we chose, at first, the unigenes related to functions as “response to stimulation, enzymes, transcription factors, and transport”.

Random Distribution of EST-SSR mapped loci using Statistical Tests

The predictable number of genic SSR loci on the wheat chromosomes was evaluated on the knowledge of the size of genome assembly Ref v1. The equation has been used

$\chi^2 = \sum [(O_i - E_i)^2 / E_i]$, where O_i is the observed number of SSR loci and E_i is the expected SSR loci.

Primers design and validation

Then, primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/>) with product size 150-350 bp, primer size 18-24 bp, optimal primer T_m 55 °C, and primer GC (30-70%). After that, In-silico PCR, part of the blat suite [25] and KaspSpoon [26] were used as in-silico PCR. These tools were used here to predict the size and position of the designed primers on the hexaploid wheat genome (IWGSC RefSeq v1.0). In the case of KaspSpoon "3' mismatch" and "product size", 5 and 100-350 bp were fixed, respectively.

To predict primers polymorphism, the online Primer-BLAST tool was used (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>). The parameters was sated as follows: primer melting temperature ranged from 50 to 60 and 55 as an optimal t_m with a difference of 2 max t_m , 'nr' as a database and *Triticum* (taxid:4564), *Triticum aestivum* (taxid:4565) and *Triticum aestivum subsp. aestivum* (taxid:4565) as an organisms. The mismatch at the 3' end was set as 3 in the last bps, and the maximum target size was detected as 350 bp. Primer Size was set from 18 to 24 and 20 as optimal size, GC primer content as (30-70 %) and max self-complementarity as 6. The rest of the parameters remained as standard. By this step in the validation of polymorphism, the expected results of the proposed drought-related primers have been computationally improved.

Results

Assembling of EST sequences

A total of 9077 EST sequences, related to drought stress tolerant of wheat, were downloaded from GeneBank. The average length of these ESTs was 552 bp. The initial sequences analysis showed a total length of 5010591 bp and a total GC count of non-redundant EST set was 55%. Out of these 9077 EST sequences, 18 sequences discarded through the sequence cleaning process; the total repetitive elements were masked were 163114 bp i.e. about 3.26 % of the total size of the query sequence. The remaining ESTs sequences for *Triticum aestivum* were assembled into 12062 contigs and 4141 singletons. Most of the contigs consist of two or three ESTs. The assembled ESTs account for 56.17% of total ESTs. The singletons representing expressed transcripts could not be assembled into larger contigs. These singletons may act as genes where only one mRNA represented one gene.

SSR results

Nine hundred and fifty nine unigenes have SSRs consisting of 301 (31.39 %) contigs and 658 (68.61%) singletons. Out of these unigenes, 1159 SSRs were found. The SSRs with one locus or perfect SSRs were the most frequent, with a frequency of (81.54 %), followed by compound (16.91%), while imperfect SSRs was only (1.55 %). The percentages of dinucleotides, trinucleotides, tetranucleotides, pentanucleotides, and hexanucleotides in EST-SSRs were 17.69, 64.71, 4.23, 5.78 and 7.59 %, respectively (Figure 1). The SSRs with one locus or perfect

SSRs were the most abundant, with a frequency of (81.54 %), followed by SSRs with two loci SSRs (16.05 %), three loci (1.29 %), four loci (0.76 %), and five loci (0.48 %).

Mainly, sixty-one types of EST-SSR motifs were recognized in the study (Table 1). In general, the SSRs were found to be unequally distributed across motif types. The motifs CGC, GGC, CCG, and GCC have the highest frequency of (5.35), (4.66), (4.49), and (3.88) % in trinucleotide respectively, followed by the motifs GC (2.50 %), CG (2.33), and CA (2.07 %) in dinucleotides. The other types of motifs own a frequency of < 3 percentage (Table 1). In addition, the motifs with 4 repeat copy number were the most abundant (53.06%) followed by 5 repeat copy number (26.14%) (Figure 2). In others studies, SSRs based on the repeat copy number were also usually commenced at 4 [13].

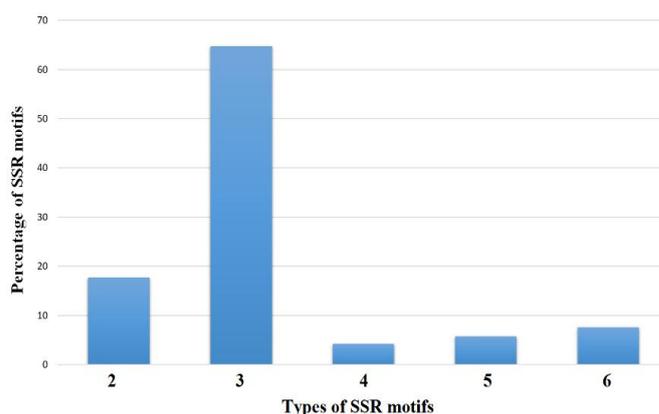


Figure 1. The percentages of SSR motifs

Distribution of SSRs in open reading frames and the amino acids for tri- and hexanucleotide

Of all identified microsatellites, 53.80% were distributed across the middle region or coding regions, while about 34.87% were located near the 5' end, and 11.34 % were located near the 3' end. In addition, hexanucleotide (68.42%) and tri-nucleotide (61.25%) SSRs were the most frequent in the coding regions, whereas, tetra and SSRs were the least frequent there. In contrast, pentanucleotide (60.71%) and tetranucleotide (50 %) SSRs were the most frequent in UTRs 5'. Each tri-nucleotide or hexanucleotide motif codes for amino acids that have a potential role in the biological activity of protein molecules. Of the 838 tri- and hexanucleotides identified in the present study, 17.45% encode Arginine, 16.78 % encode Alanine, 11.95 % Proline, and 8.99 % Glycine (Figure 3).

Functional characterization of unigenes containing SSR

Out of 959 unigenes, 854 have homologous with known proteins. The outcome showed that 6.46 % is exclusive to *Triticum aestivum* and more than 90 % was conserved between the other species like *Zea mays*, *Aegilops tauschii*, *Oryza sativa*, *Brachypodium distachyon*, and *Hordeum vulgare* (Figure 4). The percentages of the conserved regions for *Zea mays*, *Aegilops tauschii*, and *Oryza sativa* were 12.13 %, 7.78 %, and 7.34% respectively. Additionally, these 854 unigenes were submitted to additional annotation using Gene Ontology vocabulary.

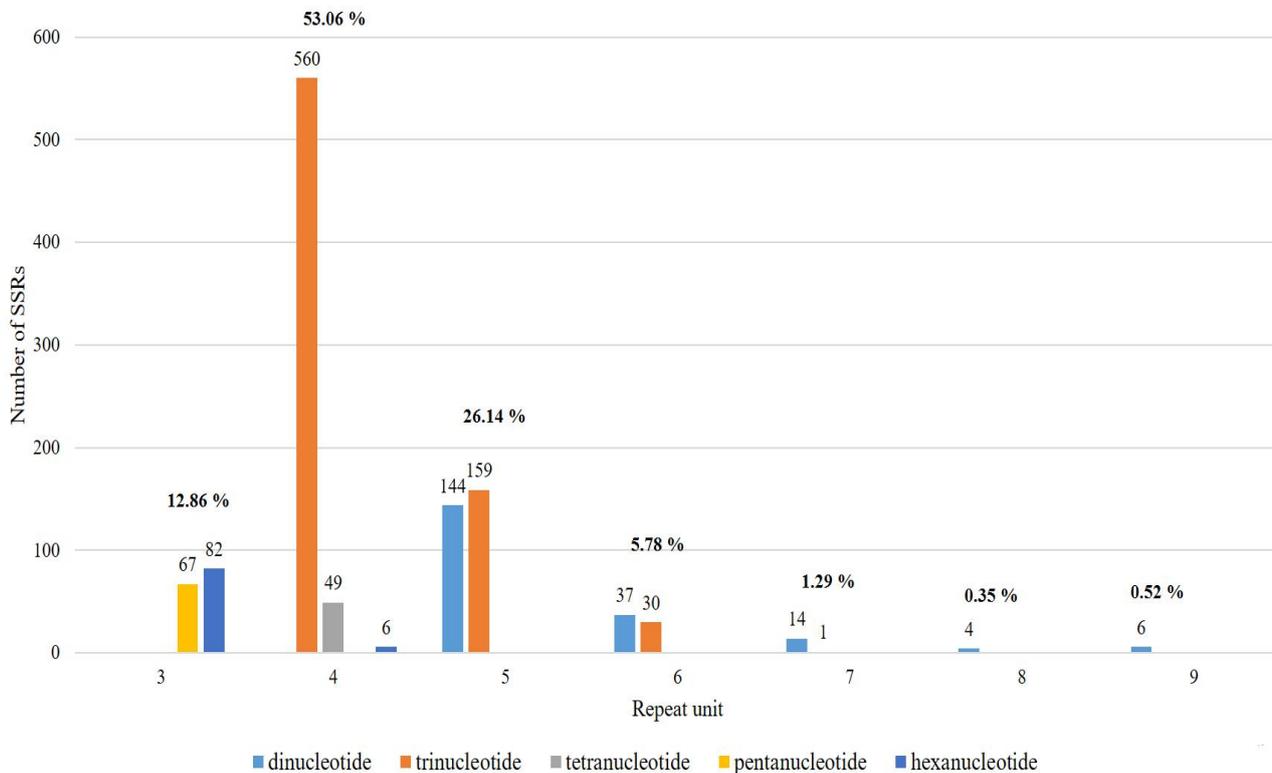


Figure 2. The motifs with the repeat copy number.

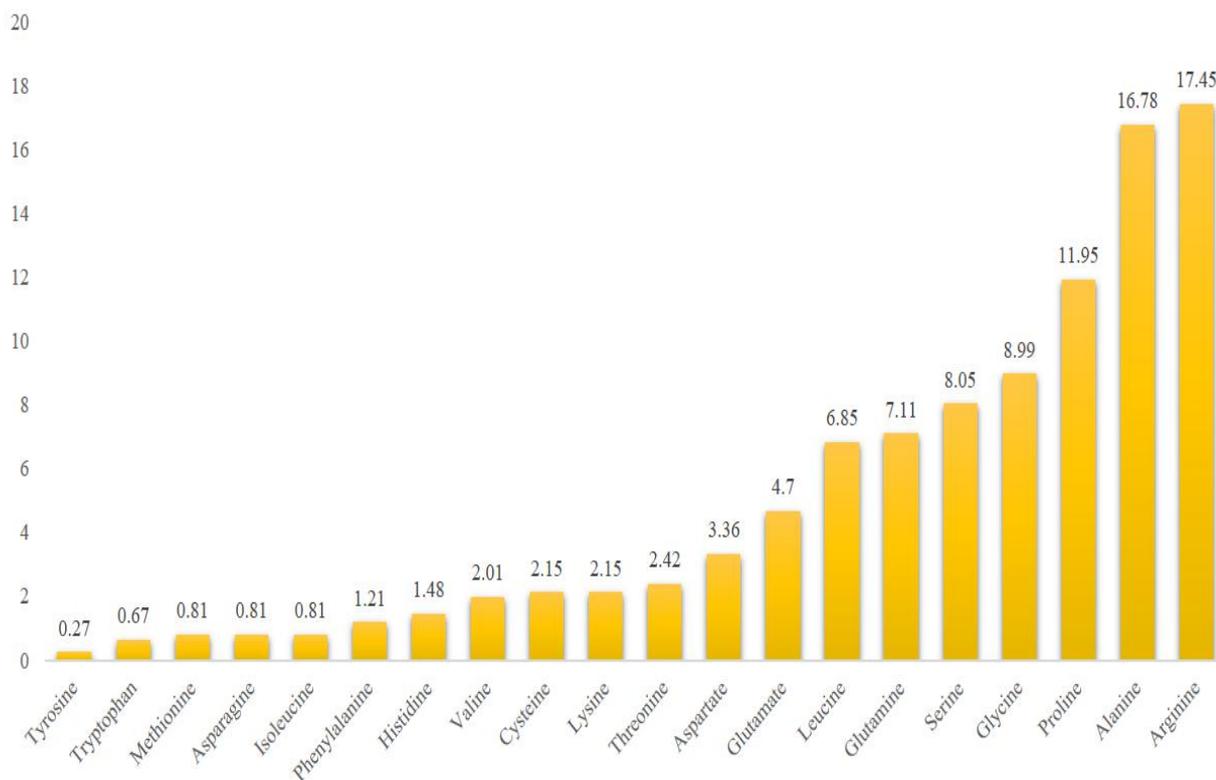


Figure 3. Distribution of amino acids in tri- and hexanucleotide SSR.

Table 1. Frequency of main types of SSRs found in this study.

Motif type	Frequency %						
ATA	0.26	TC	1.04	TGG	0.52	CAA	2.16
CTT	0.26	TG	1.21	ATG	0.6	GCG	2.16
GTG	0.26	GAC	1.21	TGA	0.6	GCA	2.24
CATC	0.26	GCT	1.21	AGA	0.69	CG	2.33
CGGC	0.26	GT	1.29	TTC	0.69	CCT	2.33
GCAGCG	0.26	GAA	1.29	CAGCTC	0.69	CTC	2.42
AAT	0.35	AGG	1.38	TA	0.78	GC	2.5
ACC	0.35	TGC	1.38	CTG	0.78	AGC	2.5
CTA	0.35	CCA	1.55	AC	0.86	CGG	2.59
GGT	0.35	GGA	1.55	AT	0.86	GGC	3.88
TCA	0.35	AAG	1.64	GAT	0.86	CCG	4.49
TCG	0.35	TCC	1.73	AG	0.95	GCC	4.66
CCTC	0.35	CT	1.81	ACA	0.95	CGC	5.35
ACG	0.43	CAG	1.81	CAC	0.95		
CGT	0.43	GA	1.98	CGA	0.95		
GTC	0.52	GAG	1.98	CA	2.07		

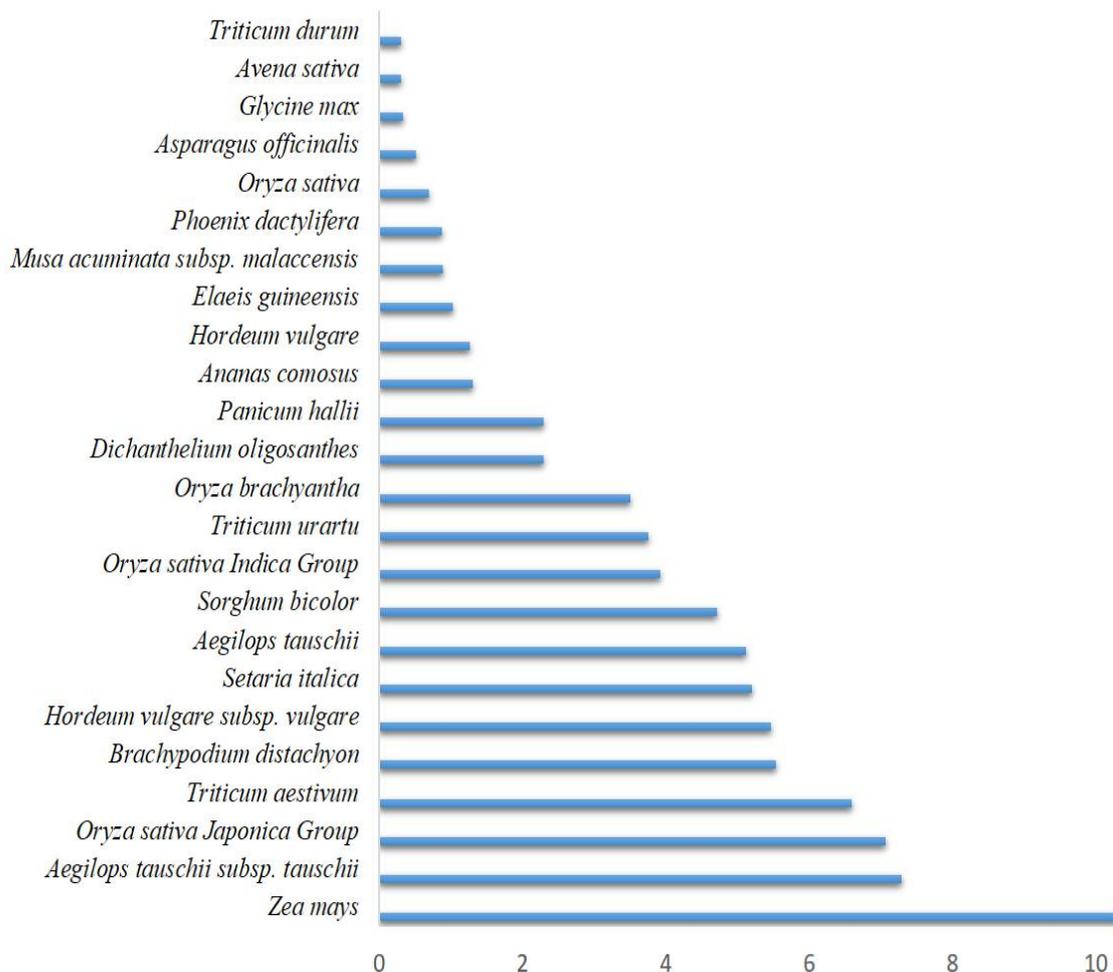


Figure 4. Species distribution of *T. aestivum* SSR-containing unigenes.

Out of 854 unigenes, GO terms were available for 706. It has been observed that 2719 GO term obtained, that means on average 4 for a unigene. Maximum 30 GO terms for one and minimum one GO for 100 unigenes. So, unigenes classified consistent with the GO Annotation into three groups: Cellular Component, Molecular Function, and Biological Process (Figure 5) according to the level 2.

Then, as mentioned above, putative functions were available to most sequences 73.61%, which have homology with known proteins, while 26.38 % were homologs with hypothetical or unknown proteins. Eighty-sixSSR-containing unigenes were identified and sorted into 25 putative TF families. Among the 25 TF families (Figure 6), C2H2 is the most abundant category (18.6%), followed at 15.1% by Hap3/NF-YB, and 9.3% for MYB-HB-like TFs. In addition, 714 SSR-containing unigenes have been categorized into 166 KEGG pathways in order to assess the biological roles and associations between genes. The 714 KEGG annotated SSR-containing unigenes have been classified into five various functional categories (Table 2). Of these, the "metabolic" pathway contained the highest number of unigenes (40.32%), followed by "Chemical Processes" (21.24%) and "Environmental Information Processing" (13.53%). Overall, "signal

transduction" (13.53 per cent) was the most frequent pathway supplemented by " Carbohydrate metabolism" (10.08%). Most ESTs encoded the following enzymes: hydrolases (32%), transferases (28%), and oxidoreductases (19.5%) (Figure 7).

Table 2. KEGG classification results and Number of SSR-containing unigenes (NSU).

KEGG categories		NSU
Metabolism	Carbohydrate metabolism	72
	Energy metabolism	35
	Lipid metabolism	23
	Nucleotide metabolism	3
	Amino acid metabolism	34
	Metabolism of other amino acids	24
	Metabolism of cofactors and vitamins	5
	Metabolism of terpenoids and polyketides	6
	Biosynthesis of other secondary metabolites	17
Xenobiotics biodegradation and metabolism	25	
Genetic Information Processing	Transcription	20
	Translation	55
	Folding, sorting and degradation	44
	Replication and repair	5
	Environmental Information Processing	109
	Signal transduction	109
Organismal Systems	Immune system	39
	Endocrine system	26
	Circulatory system	4
	Digestive system	7
	Excretory system	3
	Nervous system	25
	Development and regeneration	12
	Aging	17
	Environmental adaptation	13
	Transport and catabolism	31
Cellular Processes	Cell growth and death	45
	Cellular community	11
	Cell motility	4
	Total	714

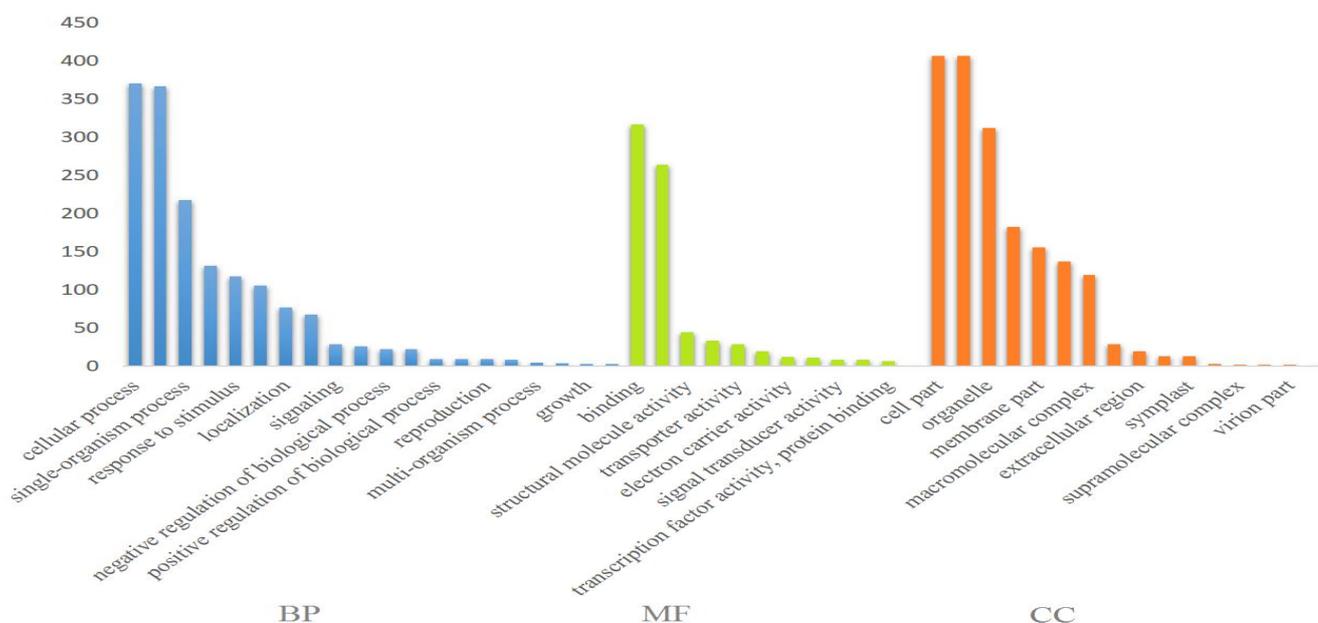


Figure 5. Distribution of GO terms of unigene sequences containing SSR in three main groups: biological process, cellular component, and molecular function.

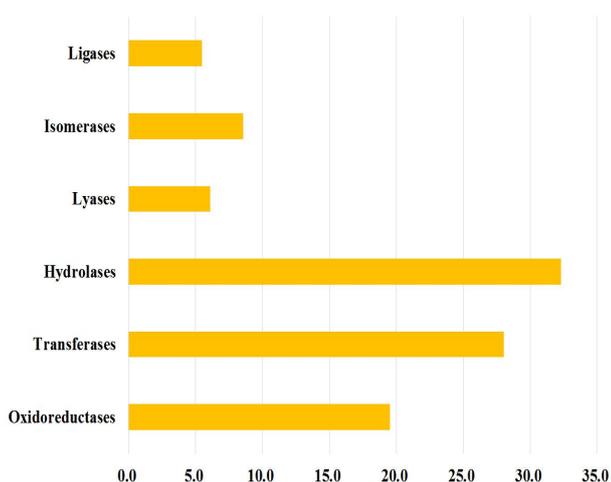


Figure 7. Distribution of Enzymes in the SSR-containing unigenes.

Chromosome localization

It was found that 929 unigenes had positions on the assembly out of 959 unigenes, and 1242 loci were located on all the 21 wheat chromosomes (in addition on the Un-named sequences) (Table 3; Figure 8). Of the mapped unigenes, 754 (81%) sequences have one location on a chromosome, 88 (9.5%) sequences have two homo-loci, and 40 (4.3%) sequences have three homo-loci pointed on three homologous chromosomes, while 47 (5%) sequences have multiple loci on nonhomologous chromosomes, which may expose duplications. The chromosomes 1B, 1D, 2A, 2D and 5A had the highest number of loci. In another hand, the 401, 392 and 376 loci, were localized on subgenome A, D and B respectively (Table 3).

Table 3. Distribution of unigenes containing SSR across the chromosomes, the three subgenomes, and the seven homologous of bread wheat.

Homologous group	Genome			Total
	A	B	D	
1	57	69	65	191
2	62	56	62	180
3	50	57	51	158
4	56	43	50	149
5	63	49	57	169
6	59	59	51	169
7	54	43	56	153

Random Distribution test of EST-SSR Loci on Individual Chromosomes, Subgenomes, and Homoeologous Groups

The distribution of genic SSR loci among the 21 wheat chromosomes, in the three subgenomes, and among the seven homoeologous groups of wheat chromosomes were (df=3-1=2, X²=19.08, P < 0.001), (df=7-1=6, X²= 12.4, P < 0.001) and (df=21-1, X²= 45.54, P < 0.01) respectively. Statistically, they were significant for each of them, suggesting a non-random distribution. However, the distribution of genic SSR loci among the three subgenomes vas the seven homoeologous groups was independent (df=(3-1)*(7-1), x²=7.49, P >0.05).

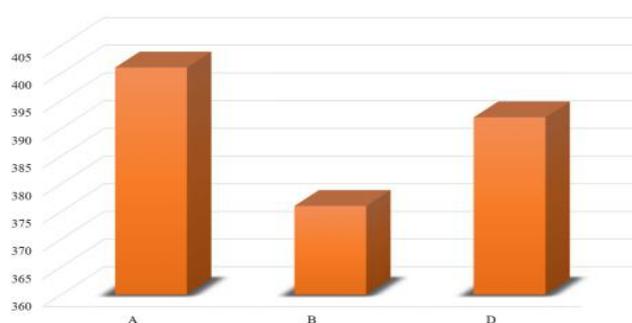


Figure 8. Distribution of unigenes containing SSR in the three subgenomes of hexaploid wheat.

Primers design & validation

Using primer3, 301 primers from 328 SSR-containing unigenes were designed to be related to enzymes, transport, transcription factors, and proteins responsible to stimulations that appear to be more related to drought stress genes. Then, a set of 63 designed drought-related primers were chosen from all 301 designed primers depending on the in-silico PCR results (Table 3). The selecting process of primers in this study was with respect to gain the expected product size on the predictable position of the related chromosomes and making sure there were no other undesirable sequences could give the same product size, so this EST-SSR primers were more related to genes expressed in the drought stress. Of the 63 primer pairs, 39% expected to produce polymorphic bands. Polymorphism was the difference in length of the predicted size of each expected band from the designed primers (Table 4, Figure 9).

Discussion

Drought is one of the most common environmental stresses affecting plant growth and development. It continues to be a critical challenge for breeders and researchers. Knowledge of functional classification, placement on chromosomes, and the use of appropriate tools to predicate and select unique specific SSR markers may have a positive impact on breeding programs and may dramatically improve the effectiveness of a specific investigation strategy or the inference of markers linked to favorable agronomic features such as drought tolerance. In this study, the dehydration-related SSR markers were mined and characterized using a computational method using the WGSC Reference Sequence v1.0 hexaploid wheat assembly. According to the Go terms, SSR-containing unigenes have been categorized as cellular components biological processes and molecular function categories (Figure 5). Among the biological process group, cellular process (72.6 %), metabolic process (71.82 %) and single-organism process (42.6 %) were the most strongly represented categories. The cell (85.5%), cell part (85.5%), organelle (65.55%), membrane (38.45%), membrane part (32.77%), organelle part (28.78%), and macromolecular complex (25.21%) categories within the cellular function categories participated the biggest portion of all annotations. In the molecular function group, binding (59.25%) and catalytic activity (49.35%) were the only two major groups. These results provide a good picture of the fact that the unigenes examined, expressed during drought stress, participate in the various biological interactions of the organization, growth and development process, metabolism and normal death of wheat cells, and are in agreement with a similar wheat study [15].

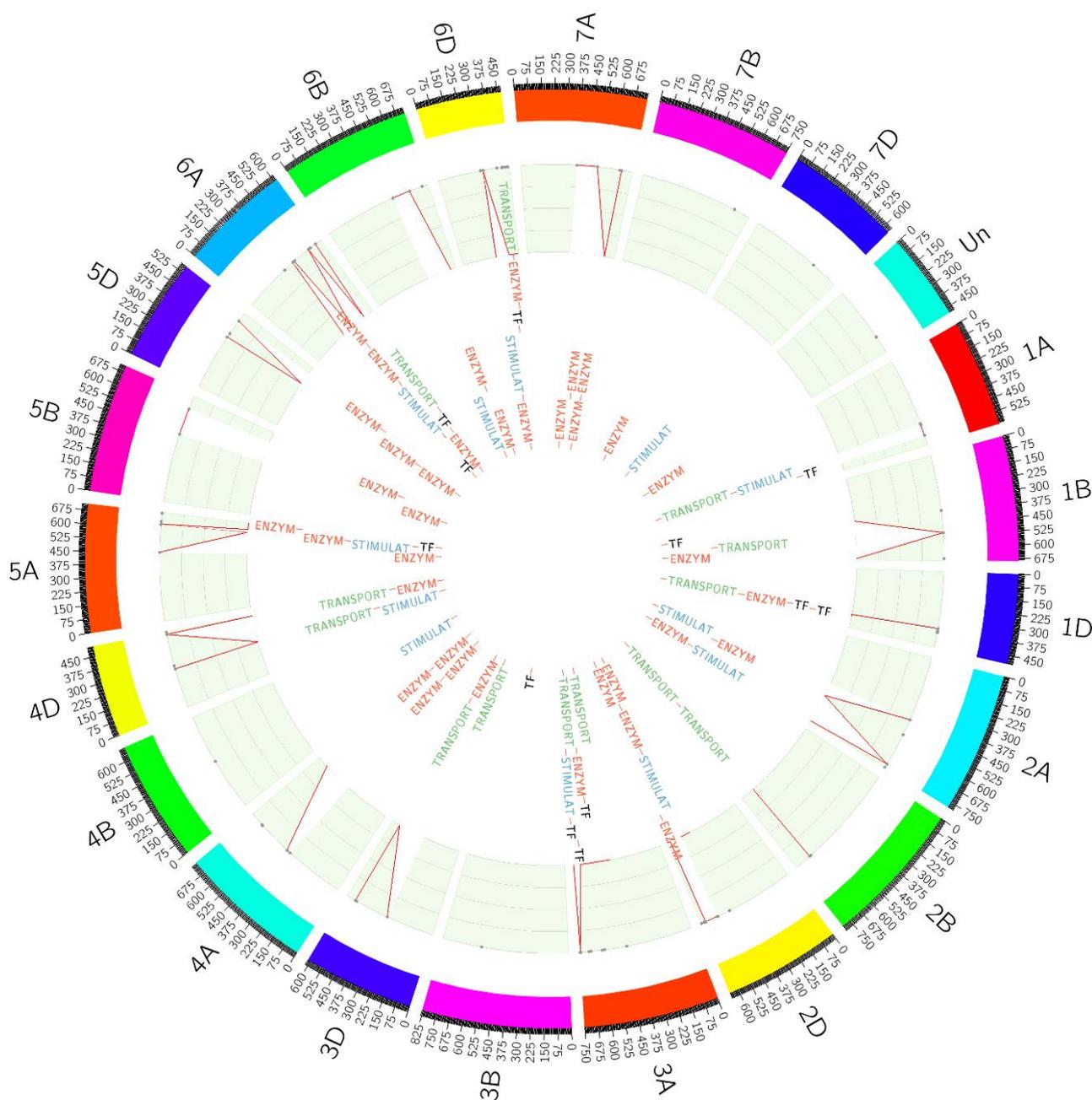


Figure 9. Chromosome wise SSR-containing unigenes distribution over 21 chromosomes of wheat by the circular plot. The context inside of the circus is the names of the primers' category in the SSR-containing unigenes. The red lines are represented the polymorphic primers that were discovered in silico.

Drought is known to be regulated at the transcriptional level and TFs have been at the center of interest for improved yields of cultivars since targeting one TF will affect other regulatory aspects of drought tolerant [7]. In this study, the most abundant was C2H2, followed by Hap3/NF-YB and MYB-HB-like. A number of studies showed the importance of these TF families in abiotic stress-tolerant. C2H2 has been shown to play a significant role in biological processes and in reacting to various abiotic stresses including oxidative and drought stress in rice [27]. In the other hand, they found NF-YB family genes are responsive to drought stress in a study on Poplar [28]. Over-expression of OsMYB3R-2 in Arabidopsis OsMYB48-1 and OsMYB48-1 in rice increases resistance to drought stress [29]. In addition, depending on the

similarity search (BLASTx), First, the enzymes had the highest frequency in our study were hydrolase and transferase. In wheat, glutathione S-transferase (GST) has been found to be high in drought stress [30]. In a study on wheat endosperm cells that were affected by drought stress, nucleic acid hydrolase activity was higher in dry treatment compared to control [31].

Another approach to identifying gene functions with a concentration on biochemical pathways was used in our study by conducting a single-directional BLAST scan against the KEGG protein database to better clarify the relationship between the putative proteins and the biological functions derived from 714 SSR-containing unigenes.

Table 4. Primers designed and polymorphism for enzymes, transport, transcription factors, and proteins responsible to stimulations for SSR-containing unigenes.

Target	Repeat motifs	Left primer sequence	Tm (°C)	Right primer sequence	Tm (°C)	Chromosome	Polymorphism
Enzymes	(CCG)5	ACTCTCCCCCTTCTCCT	57.59	CAGAGTTCTCGATCTCGTC	57.51	2D	poly
	(CGC)4	ACTCTTTGGAGTTCACG	50.61	GGATCGCTCATGTAGTTCT	52.6	7A	poly
	(GCC)4	CCTCCAGTCATCTGACATC	55.46	GTCTGGATCTTGCTCCACT	55.15	5A/5B/5D	poly
	(GCC)4	CTCCACGGCTTCAACTTCT	58.01	CCTAGCACACGGGCAGAT	59.81	3D	na
	(GC)5	CCTGCTTGTCAGTTGTCAC	55.05	TGTTAGCGGTGATCTCGT	55.09	5D	na
	(CTC)5	TCTTCTTCGGCTCTCCTG	56.62	GCAGCTGAGGATCCACTC	56.79	6A/6B	poly
	(CGC)5	CAGCAAATTCGGGTACTG	55.13	AGTGGTCTCGTGGGAAGT	55.36	Un	na
	(GCC)4	CAGCTCCCACATCTCAAC	59.13	CCCAGAACCATACCCTCTG	58	2A/2D	na
	(CA)6	GCAAATCACAACAGCTAAA	54.17	GTATTGCACGATGATCAGG	54.85	6D	na
	(AAT)5	GTGGTCAAGGTGCTCAAC	54.73	CTCCAGAACTCACAAGC	55.16	7A/7B	na
	(AAG)5	CCTTGACTTGAGGTGAGT	55.43	GTTCCGCTCTGTGTTCTC	54.02	5A/5B/5D	poly
	(CCA)5	AGGATCGAACCCAGAATATC	55.58	GCTCGTTGAGCTCGTACA	56.02	6D	na
	(GCC)5	AGATCTTCATGTGGACCATC	54.73	CATGGGTAATTAAGACTTGA	55.94	4A	na
	(CG)5	ACTCCCACACACTACAGCTC	55.2	GGCAGTCGTTGAGGTACA	55.47	7A/7B/7D	poly
	(CAGG)4	CACGATGAAGTGAAGTCA	55.45	CCTTCTCTGAATCTGTTT	54.91	2D	na
	(GCC)4	GTGATCAGCTCGTCCAAC	54.82	AGAATCGGACCTCTCT	54.66	6A/6B/6D	na
	(GCC)4	GCATTTCTCGAACACATC	56.55	AGAGCACGACCTCTTCC	55.32	6A/6B/6D	poly
	(CCG)4	ATCTCTCCCTCACCTC	56.38	GGAGGAATTTAGGGAGAAAAA	56.52	4D/4B	poly
	(CAG)4	CATACTGGAAGAACCATCA	55.95	CTGTACACACGTACGGATA	54.85	2A	poly
	(CCG)4	CTCACACTCAGAGGAAGGAG	54.96	ACAGGGACAGAGGAGACAT	54.3	1D/1B	na
	(GAC)4	GACTCCACCTACCGTCTC	57.03	CACGATGGCGATCTTCAG	58.84	7A	poly
	(ACA)4	CAAGATGAGGGAGAACGTC	55.62	CGGAACTTGGTGACATA	55.21	6D	na
	(CCG)4	TCGACATGGGTATTCTCTC	55.12	AGCTAGGCTTTCATTAGA	54.42	6D/6D	na
	(GCC)4	CAGCGAAGAAAGATTGAACC	57.55	GCAGGTTGATCTCGAACG	57.82	2D/2A	na
	(CCT)4	TCTCACTTCATCGGTATTCTC	54.41	GACTGGTTCGCTTCTTCAG	54.23	3A/175	na
	(GGA)4	GTTGTTGAGCAGGGAGTG	54.94	AGGTGGTCTGACACGATG	55.07	4A	poly
	(GAG)5	CTCGTTGATAGCTTAGAG	50.4	GTTGATGTGGGATACGTC	51.5	1B	na
	(TGC)4	GCAGCTCTTTCACACCAC	55.12	ACTCTTGGATGAGGGAGAAG	55.88	4A/4B	poly
	(CGGC)4	GTCTACAAGTTTGGCTACGG	55.08	GAGGTGCTGTCTCTGAAC	55.7	4A/4B/4D	na
	(CGT)4	CAAGTCCATCGGCTAACTT	55.4	GTCGATCATGTCTGCCAC	55.46	5A	na
	(GCA)4	GACAGATCTTGGCTCACTTC	54.86	CTGAAGAGCAGGAACACG	55.33	4B	na
	(GC)5	CCGGACTACGTGGAGGAG	59.2	ACGTACGAAGGCACTAGCAT	57.54	4D	poly
(GCG)5	AGCTAGCAGGTTGTGACTGA	56.18	GTCACCTGGAGGCACTTG	57.61	3A	poly	
(CGC)4	GACGGACGTGAAGAAGAAG	55.88	ACTTAGTAACGCGAGCATGT	55.26	6A/6B/6D	poly	
(CCT)4	TAGCTAGGAACCATGTCCAA	55.84	GTGCTTGGGTGGTTCGAG	57.79	1A	poly	
(CCA)4	CAGCTTGCTCTGTTCCAT	54.76	GTTCTTGGCGAAGGTCTC	55.78	7D/7A/7B	poly	
(TCC)4	CAAGACCAGTCTCCAGTCAT	55.08	AGGACTGCTCGTAGACCTC	54.43	2A	poly	
(TCCA)4	TCTACAAGAAGTTCAGGGTGA	55.01	AGTAAGCTAGATGGGGAAAC	55	2D	poly	
(CCA)4	GCTCATCTACGAGGAGACC	54.79	CGATAAAGTCTCCAATTTC	55.06	5A	na	
(GCT)4	CACACATTCCTCGAATCAC	54.69	AGACGATGAGGATGATGAAG	55.1	2A	na	
(CG)5	CTGACCGAGGAGAGGAAG	55.8	TCGATGTACTGGATCAGCTC	56.3	3B	poly	
(TCC)4	GTTTCTGGGTCTTACTACTAC	55.16	TCCGTACGCATTTATTCTT	55.09	3A	na	
(AAG)5	CCCAAGCAACCACACTCT	57.09	AGAGCATGAGGAGGCAGA	56.98	5A/5B/5D	na	
(CGC)4	TCGCTCTGACGATTGATAC	56.44	GTGTGGAACGTGCTCTCC	57.61	6A	na	
(CATG)4	GTGTTACATCACATCTCCA	55.07	GTTAGCACACAAAGGACCAT	55.14	3A	na	
(GCACCA)3	CTCCTCCGAACCAACTC	58.13	CTTCAAAGGTCTACCAACAGG	56.48	1D	na	
(TCC)4	TCTCTTCTCCCTCTCCTG	56.21	CGTAAGCGACGTGGAGTC	57.91	1B/1D	na	
(TCC)4	CACATCAAGCGCAAGTCTC	59.23	CTACCGACAGTCCAGGTT	57.56	1A	poly	
(GGC)4	GATTCCCAAGGTGATGTTT	55.18	AGAACTCGGTCATGGAGAG	55.16	6D/6A	na	
(AAG)4	GAAACCATCGTCTTCGTTTA	55.38	AAGGACTACACGCTCACAGT	54.94	2D	na	
(TAAT)4	CTGTTGGTAACTCGGAGAAC	54.85	TTGTAATAAACACAAAGCAAGG	54.89	3A/3B	na	
(TG)5	CTCCGATCAAACATCGAG	54.97	CCATTCATGGAGAACAAGAT	54.95	6A/6D	na	
(CGC)4	ATCTTCCAGACCAACGATAG	54.27	GAGAGACTGGGGTCTTCC	54.33	1D	na	
(GT)5	ATGGCCTTCTCTCATGTA	54.85	TAGACGCATCGCTAGTAACA	54.82	4D	na	
(TGC)4	ACCAGAGTTTTCCCATC	54.7	CTTCATGATGGACTTCTGGT	55.01	6A/6D/6B	poly	
(CA)5	CAAGTATCCCTCCAAGTG	55.5	TACAACAACTCCCGTCTCT	54.85	3D	poly	
(GTAC)4	AGGACCACTGCTACTACTTG	57.19	CTTCCCAGGGAGAGGAT	59.13	1A	poly	
(TCT)4	GCAACCTGATGATCTACCTC	54.66	GTCGTTCTCTGTCGAAGT	55.13	3A/3D	poly	
(CTG)5	CAGAACACTCATCCGAATCT	55.19	CTTAACCTACCGGTGTTTTC	54.03	3A	na	
(CGG)5	AAAATGCACCCACGTAGC	54.73	CATGAGCCTGAGCTTGTG	56.33	2B/2D	poly	
(GGC)4	TGAATTTACCTTGGGAGTC	54.28	GATCTCCTTTGGAGGCTAT	55.05	3A	na	
(ACC)4	AGGAGTCCACCCAATCT	57.34	GAATCCACGAGGTCCAG	56.51	2B/2A	na	
(GGC)6	GTGCTGTCAAGGACACTTCT	55.34	TTGGTAAGCACCTTCTCAAT	54.96	4D/4A	na	

According to KEGG findings, 714 unigenes were mapped to 166 expected metabolic pathways (Table 2); abundance unigenes of "metabolism" were in "carbohydrate metabolism," "energy metabolism" and "amino acid metabolism". This suggests that carbohydrates, energy metabolism and amino acids have been involved under abiotic stress and that are in line with other research [32]. KEGG pathway is a valuable method for predicting potential genes and their functions at the overall transcriptome level.

In our study, the abundance of SSRs (perfect, compound and imperfect) in unigenes was 1 in 4.7 unigenes and 1 in 2.7 kb, compared to 1 in 3.1 Mb in wheat assembly. It is vital to note that the total frequency and frequency of different lengths of SSRs and repeat motifs are reverting to the difference in SSR search tools with different SSR search principles and different datasets used for EST database mining, which change significantly in different studies. For instance, in wheat, the frequency of SSRs in ESTs varied from 1 in 0.74 kb to 1 in 9.2 kb [33].

The trinucleotides had the highest frequency of SSR motifs followed by dinucleotides (Figure 1). In a study of SSR analysis using the database of the whole hexaploid wheat genome, it was observed that di- followed by hexanucleotide were the most abundant. However, many of the previous studies have agreed that trinucleotide are the most abundant motif [15]. The repeated copy number at the locus is unsteady. In fact, the increase or decrease in the number of repeated copies in microsatellites in non-coding regions believed to modify gene regulation, while in coding regions it usually results in alterations in the reading frames resulting in changes in protein products [13].

The relative location of the di-, tri-, tetra-, penta- and hexanucleotide repeats was identified and labelled in regards to the open frame reading (ORF). From this analysis, it was observed that more than half of all microsatellites were distributed along the coding regions, while the remainder were unevenly separated between the 5' end and the 3' end. This agreed with earlier studies of the genus *Prunus* [34] and *Gentianaceae* [35]. In addition, trinucleotides and hexanucleotides have been the most abundant in coding regions. Similar studies have parallel outcomes [9,13,15]. This showed a significant bias in the frequency of SSRs in coding regions and UTRs. The increased rate of SSRs in coding regions indicating their role in protein structure and therefore their function [34]. In addition, many trimeric SSRs in ESTs were correlated with the non-existence of frameshift mutations in coding regions when the length of these SSRs changed [9].

Arginine was the most common amino acid followed by alanine, proline and glycine (Figure 3). In a recent study, the metabolome analysis showed amino acids, organic acids, and sugars as the main metabolites varied in multitude upon to water shortages. In particular, a spring wheat cultivar (Bahar cv) display increased levels of proline, arginine and lysine in drought stress [36]. In addition, a peanut study found that drought increased the amount of arginine in all genotypes [37]. In a study of *Triticeae* species to compare metabolite profiles of drought stress, alanine was found to increase exposure to drought stress [38]. In addition, proline amino acid is commonly known to occur in higher plants and usually accumulates in large quantities in response to environmental pressures; therefore, proline has advanced positive effects as the root/shoot proportion of wheat

germplasm has improved under water-limited stress [39]. The arginine code sequences were CGG, AGA, CGC, CGT, and CGA, both tri- and hexanucleotide. It has been reported that in coding regions, this triple could possibly encode amino acids while in coding regions or UTRs. Such forms of motifs are considered to be amino acid coding sequences that could be used as a source of several functional markers [13].

The average number of mapped loci per genic SSR in the present study on the genome assembly was 1.3 and ranged from 1 to 4; these results are consistent with previous studies [40]. From the mapped genic SSR, most sequences have a single location and are consistent with the knowledge that ESTs usually appear as a single locus [15,40]. While the presence of EST-SSRs with multiple loci is essential and may be referred to the conserved nature of EST-SSRs that tend to occur as homoeologous in two or three genetically similar regions in subgenomes [40]. The presence of multiple positions on non-homoeologous chromosomes (5%) may be represented by repetition. This concept is consistent with the gene duplication reported for the wheat genome [40], but differs from the expected 25 to 30%; and this may return to many purposes, one of which used a different alignment tool (e.g. *blat*) and used EST return to a specific trait, drought.

In the present study, in drought conditions, the distribution of genic SSR markers to all 21 chromosomes was non-random ($P < 0.001$) with a maximum of 69 genic SSR loci mapped to chromosome 1B and a minimum of 43 genic SSR loci to chromosome 4B and 7B and 49 genic SSR loci to chromosome 5B. These results are consistent with previous studies in which chromosome 1B was observed to have the maximum number of genic SSR loci [40], while 4B had the minimum number of SSR loci with the lowest genic intensity [40]. In addition, 2D and 5A were found to have high loci numbers in this study, and this is consistent with a previous study on bread wheat under drought where QTLs on chromosomes 5A and 2D were involved in controlling stomatal density and size traits and were useful for molecular breeding [41]. In addition, 5A was found to carry many more QTL for a number of traits within short distances for agronomic traits under drought stress in wheat [42]. While 4B found to have no QTLs for stay-green and agronomic traits in wheat under drought stress [43]. The distribution of the genic SSR loci in the three subgenomes (Table 3.) was also a non-random distribution ($P < 0.001$) with the maximum number of genic SSR mapped on the subgenome A while the minimum number was on the genome B, although the subgenome B was larger than the subgenome A or the subgenome D [44]. This is not consistent with earlier studies that recorded more EST-SSR loci mapped to the B subgenome compared to those mapped to the A or D subgenome [33]. It can be returned to the disparity between the methods used for mapping sequences found in the SSR, the genome assemblies used, the type of EST datasets, and the reliance on the computational method instead of the wet-lab as in the previous studies. The distribution of genic SSR loci among the seven homoeologous groups of wheat chromosomes was also non-random ($P < 0.001$). The most loci were found to be in the homologous groups 1, 2, and 5, while the least was in the homologous groups 3, 4 and 7, which tend to be close to the QTL mapping study of traits such as stay-green and agronomic in wheat with multiple water shortage levels [43]. Nevertheless, the distribution of

genic SSR loci within the three subgenomes was the seven homoeologous assemblies was independent ($P > 0.05$) and is consistent with the other wheat study [40].

The rate of successfully designed primers ranges from 60-90 % for both genomic and genetic SSRs [9]. In our research, variations between the predicted size of the primer3 and that of the wheat genome were identified using in-silico PCR analysis. In these cases, the PCR product was absent, smaller or greater than predicted. These agreements with many similar studies, in which similar results are obtained by testing EST-SSR primers on the wheat genome in the wet lab [9]. Such drought stress-related SSR-markers are known to be useful markers and must be tested to be inserted afterwards in breeding programs for cultivated wheat. The rate of predicted polymorphism in our study was low and that consists with many studies on ginec SSR e.g. [11].

Due to greater conservation of DNA sequences in transcribed regions, EST-SSR primers in crop plants were found to be less polymorphic than genomic SSRs [9]. However, our findings were comparable to other studies [14,45,46]. Here we have tried to overcome one of the common issues with primers designed when performed in a web lab, which is 'primers polymorphism' (Table 4).

Usually, studies give a picture of the SSR-marker and the relative maps dedicating their positions on the chromosomes using real PCR. However, in addition to the positions of high-quality predicted SSR-markers and their polymorphism on one of the newest wheat assemblies, our research provided a comprehensive view of the SSR-containing drought-related sequences and their distribution and the most accurate chromosomal positioning using a wide range of tools.

Conclusion

Due to the size and complexity of the wheat genome, the bioinformatics tools are needed to facilitate and accelerate research, particularly with regard to the availability of databases and resources. A comprehensive survey using computational approaches was performed in this study, using several tools for characterization, detection of random distribution of EST-SSR mapped loci, designed and validation of SSR marker polymorphism related to drought-tolerant. The methods and findings used in the present study may help researchers and breeders accelerate breeding programs by inserting computational way to laboratory work subsequently reducing the time and labor intensive needed to select suitable markers associated with a particular trait.

References

- Brenchley R, Spannagl M, Pfeifer M, Barker GLA, D'Amore R, Allen AM, et al. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* [Internet]. 2012;491(7426):705–10. Available from: <http://dx.doi.org/10.1038/nature11650>
- Gill BS, Appels R, Botha-Oberholster AM, Buell CR, Bennetzen JL, Chalhoub B, et al. A workshop report on wheat genome sequencing: International genome research on wheat consortium. In: *Genetics*. 2004.
- Rudd S. Expressed sequence tags: Alternative or complement to whole genome sequences? *Trends in Plant Science*. 2003;
- Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., Pozniak, C.J., Choulet, F., Distelfeld, A., Poland, J. and Ronen G. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *RESEARCH*. 2018;7191.
- Farooq M, Hussain M, Siddique KHM. Drought Stress in Wheat during Flowering and Grain-filling Periods. *Critical Reviews in Plant Sciences*. 2014.
- Bernardo R. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Science*. 2008.
- Budak H, Kantar M, Yucebilgili Kurtoglu K. Drought tolerance in modern and wild wheat. *The Scientific World Journal*. 2013;2013.
- M. Perez-de-Castro A, Vilanova S, Canizares J, Pascual L, M. Blanca J, J. Diez M, et al. Application of Genomic Tools in Plant Breeding. *Current Genomics*. 2012;
- Varshney RK, Graner A, Sorrells ME. Genic microsatellite markers in plants: features and applications. 2005;23(1).
- Nahas LD, Al-husein N, Lababidi G, Hamwieh A. In silico prediction of novel genes responsive to drought and salinity in bread wheat (*Triticum aestivum*). *PLoS ONE*. 2019;1–13.
- Iqbal MA, Sharma P, Jasrotia RS, Jaiswal S, Kaur A. RNAseq analysis reveals drought- responsive molecular pathways with candidate genes and putative molecular markers in root tissue of wheat. *Scientific Reports*. 2019;(October 2018):1–18.
- SHI J feng, MAO X guo, JING R lian, PANG X bin, WANG Y guo, CHANG X ping. Gene Expression Profiles of Response to Water Stress at the Jointing Stage in Wheat. *Agricultural Sciences in China*. 2010;
- Asadi AA, Rashidi Monfared S. Characterization of EST-SSR markers in durum wheat EST library and functional analysis of SSR-containing EST fragments. *Molecular Genetics and Genomics*. 2014;289(4):625–40.
- Singh AK, Chaurasia S, Kumar S, Singh R, Kumari J, Yadav MC, et al. Identification , analysis and development of salt responsive candidate gene based SSR markers in wheat. *BMC Plant Biology*. 2018;1–15.
- Yang ZJ, Peng ZS, Yang H. Identification of novel and useful EST-SSR markers from de novo transcriptome sequence of wheat (*Triticum aestivum* L .). *GMR*. 2016;177(1):1–14.
- Thiel T, Michalek W, Varshney K, Graner A.

- Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). 2003;411–22.
17. Varshney RK, Hiremath PJ, Lekha P, Kashiwagi J, Balaji J, Deokar AA, et al. A comprehensive resource of drought- and salinity- responsive ESTs for gene discovery and marker development in chickpea (*Cicer arietinum* L.). *BMC Genomics*. 2009;10.
 18. Arbeiter AB, Hladnik M, Jakše J, Bandelj D. Identification and validation of novel EST-SSR markers in olives. *Scientia Agricola*. 2017;2(June):215–25.
 19. Saini P, Yadav DKRC, Yadav NR. SRAPs and EST-SSRs provide useful molecular diversity for targeting drought and salinity tolerance in Indian mustard. *Molecular Biology Reports* [Internet]. 2019;0(0):0. Available from: <http://dx.doi.org/10.1007/s11033-019-04590-4>
 20. Xia H, Zheng X, Chen L, Gao H, Yang H, Long P, et al. Genetic Differentiation Revealed by Selective Loci of Drought-Responding EST-SSRs between Upland and Lowland Rice in China. 2014;9(10).
 21. Masoudi-Nejad A, Tonomura K, Kawashima S, Moriya Y, Suzuki M, Itoh M, et al. EGassembler: Online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments. *Nucleic Acids Research*. 2006;
 22. Huang X, Madan A. CAP3: A DNA sequence assembly program. *Genome Research*. 1999;
 23. Conesa A, Götz S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics*. 2008;2008.
 24. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*. 2007;
 25. Kent WJ. BLAT — The BLAST -Like Alignment Tool. *Genome Research*. 2002;12:656–64.
 26. Alsamman AM, Ibrahim SD, Hamwieh A. KASPs spoon: An in vitro and in silico PCR analysis tool for high-throughput SNP genotyping. *Bioinformatics*. 2019;
 27. Sun SJ, Guo SQ, Yang X, Bao YM, Tang HJ, Sun H, et al. Functional analysis of a novel Cys2/His2-type zinc finger protein involved in salt tolerance in rice. *Journal of Experimental Botany*. 2010;
 28. Yan DH, Xia X, Yin W. NF-YB Family Genes Identified in a Poplar Genome-wide Analysis and Expressed in *Populus euphratica* Are Responsive to Drought Stress. *Plant Molecular Biology Reporter*. 2013;
 29. Xiong H, Li J, Liu P, Duan J, Zhao Y, Guo X, et al. Overexpression of OsMYB48-1, a novel MYB-related transcription factor, enhances drought and salinity tolerance in rice. *PLoS ONE*. 2014;
 30. Nezhadahmadi A, Prodhhan ZH, Faruq G. Drought Tolerance in Wheat. 2013;2013.
 31. Li C, Li C, Wang B, Zhang R, Fu K, Gale WJ, et al. Programmed cell death in wheat (*Triticum aestivum* L.) endosperm cells is affected by drought stress. 2018;1039–52.
 32. Goyal E, Amit SK, Singh RS, Mahato AK, Chand S. Transcriptome profiling of the salt- stress response in *Triticum aestivum* cv. Kharchia Local. *Nature Publishing Group* [Internet]. 2016;(June):1–14. Available from: <http://dx.doi.org/10.1038/srep27752>
 33. Peng JH, Lapitan NLV. Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers. *Functional and Integrative Genomics*. 2005;
 34. Sorkkeh K, Prudencio AS, Ghebinejad A, Dehkordi MK, Erogul D, Rubio M, et al. In silico search, characterization and validation of new EST-SSR markers in the genus *Prunus*. *BMC Research Notes*. 2016;9(1):1–11.
 35. Sathishkumar R, Lakshmi PTV, Annamalai A, Arunachalam V. Mining of simple sequence repeats in the Genome of Gentianaceae. *Pharmacognosy Research*. 2011;
 36. Michaletti A, Naghavi MR, Toorchi M, Zolla L, Rinalducci S. Metabolomics and proteomics reveal drought-stress responses of leaf tissues from spring-wheat. *Scientific Reports*. 2018;
 37. Aninbon C, Jogloy S, Vorasoot N, Nuchadomrong S. Change of arginine content and some physiological traits under midseason drought in peanut genotypes with different levels of drought resistance. 2017;285–93.
 38. Ullah N, Yüce M, Neslihan Öztürk Gökçe Z, Budak H. Comparative metabolite profiling of drought stress in roots and leaves of seven Triticeae species. *BMC Genomics*. 2017;
 39. Iqbal MJ, Maqsood Y, Abdin ZU, Manzoor A, Hassan M, Jamil A. SSR Markers Associated with Proline in Drought Tolerant Wheat Germplasm. *Applied Biochemistry and Biotechnology*. 2016;
 40. Mohan A, Goyal A, Singh R, Balyan HS, Gupta PK. Physical mapping of wheat and rye expressed sequence tag-simple sequence repeats on wheat chromosomes. *Crop Science*. 2007;
 41. WANG S guang, JIA S shan, SUN D zhen, FAN H, CHANG X ping, JING R lian. Mapping QTLs for

- stomatal density and size under drought stress in wheat (*Triticum aestivum* L.). *Journal of Integrative Agriculture* [Internet]. 2016;15(9):1955–67. Available from:
[http://dx.doi.org/10.1016/S2095-3119\(15\)61264-3](http://dx.doi.org/10.1016/S2095-3119(15)61264-3)
42. Gahlaut V, Jaiswal V, Tyagi BS, Singh G, Sareen S, Balyan HS, et al. QTL mapping for nine drought-responsive agronomic traits in bread wheat under irrigated and rain-fed environments. *PloS one*. 2017;12(8):e0182857.
 43. Shi S, Azam FI, Li H, Chang X, Li B, Jing R. Mapping QTL for stay-green and agronomic traits in wheat under diverse water regimes. *Euphytica*. 2017;213(11):1–19.
 44. Gill BS, Friebe B, Do TREN. Standard karyotype and nomenclature system for description of chromosome bands and structural aberrations in wheat (*Triticum aestivum*)'. 1991;1986(Gill 1986).
 45. Zhuang X, McPhee KE, Coram TE, Peever TL, Chilvers MI. Development and Characterization of 37 Novel EST-SSR Markers in *Pisum sativum* (Fabaceae) . *Applications in Plant Sciences*. 2013;
 46. Qiu L, Yang C, Tian B, Yang JB, Liu A. Exploiting EST databases for the development and characterization of EST-SSR markers in castor bean (*Ricinus communis* L.). *BMC Plant Biology*. 2010;