

RESEARCH

Open Access

DNA barcoding of different *Triticum* species



Samira A. Osman* and Walaa A. Ramadan

Abstract

Background: The genus *Triticum* L. includes diploid, tetraploid, and hexaploid species. DNA barcoding is a new method to identify plant taxa by using short sequences of DNA and within a short time. In this investigation, we determined a phylogenetic analysis of 20 different *Triticum* species by partial chloroplast Maturase encoding gene (*matK*).

Materials and methods: Twenty accessions of different *Triticum* species diploid, tetraploid, and hexaploid were obtained from different countries. Genomic DNA was isolated from young leaves of studied samples and then used as a template for PCR reaction. PCR products were checked by electrophoresis, purified, sequenced, and submitted in the GenBank nucleotide sequence database, the nucleotide sequence was translated into an amino acid sequence. The nucleotide and amino acid sequences were aligned with Clustal W multiple sequence alignment programs to obtain the phylogenetic tree depending on two statistical data analysis such as bootstrapping and pairwise distance from both nucleotide and amino acid sequences.

Results: The phylogenetic tree obtained from both nucleotide and amino acid sequences divided the 20 *Triticum* species into two groups, A and B. Group A represented the diploid *Triticum* species. Group B was divided into two subgroup, I and II. Subgroup I represented the hexaploid *Triticum* species and subgroup II represented the tetraploid species.

Conclusion: The *matK* gene sequence has a critical role in discriminating the closely related *Triticum* species. So these sequences could be used as a DNA barcode for detecting the evolutionary history of *Triticum* species.

Keywords: *Triticum* species, Genetic relationships, *matK* gene, DNA barcode

Introduction

Poaceae is a large family that includes the Triticeae tribe, this tribe has around 400–500 species. Genus *Triticum* L. exists as a polyploid species such as a diploid $2n = 2x = 14$, a tetraploid $2n = 4x = 28$, and a hexaploid $2n = 6x = 42$ species, many of these species have economic importance as a food crop (Doebley et al. 2006). *Triticum* species genomes were designated as A, B, D, and G contributes to the genome constitution. Several types of analysis gave critical knowledge about the ancestry of the definite genomes in allopolyploid species (Zhang et al. 2002; Gu et al. 2004). It has been generally accepted that diploid wheat and *Aegilops squarrosa* L. (= syn. *Ae. tauschii*) (Goat grass) are donors of the A and D genomes, respectively (McFadden and Sears 1946). Many different species have been reported as the original

donor of B and G genomes but it is now largely believed that the progenitor was a member of the Sitopsis section of the genus *Aegilops*, namely, *Ae. bicornis*, *Ae. longissima*, *Ae. searsii* or, most likely, *Ae. Speltoides* (Provan et al. 2004). Also *Ae. speltoides* was considered as the maternal donor of *Triticum* species (Dizkirici et al. 2016). There is a hypothesis that the B genome of polyploidy wheat is from a polyphyletic origin, i.e., it is a recombined genome derived from two or more diploid.

DNA sequence analysis techniques are considered as a modern approach in studying evolutionary relationship and biodiversity (Stoeckle 2003; Ferri et al. 2009). DNA barcoding technique has an essential role in the identification of species (Hebert et al. 2003) due to the small size of the DNA sequence with a high discriminatory power between the organisms. So, they play an important role in the identification of the plants having a problematic taxonomic identity for the biodiversity investigation and the identification of polymorphic plant species (Ajmal et al. 2014; Skuza et al. 2015).

* Correspondence: s_nrc82@yahoo.com

Genetics and Cytology Department, Genetic Engineering and Biotechnology Research Division, National Research Centre, Giza P.O. 12622, Egypt

There are many plant DNA barcodes such as *rbcl*, *matK*, *trnH-psbA*, and *ITS* (CBOL Plant Working Group 2009; China Plant BOL Group 2011; Li et al. 2015). The group of Consortium for the Barcode of Life (CBOL) recommended using a combination of two chloroplastic barcodes (*matK* and *rbcl*) as the standard plant DNA barcode supplemented with an additional barcode as required (CBOL Plant Working Group 2009).

The chloroplastic *matK* gene region (coding sequence) has a complete size with about 1500 bp that is translated into around 500 amino acid sequences for protein (maturase-like protein). The *matK* gene is one of the useful regions because it is the most rapidly evolving plastid gene, which provides sufficient information to identify the phylogenetic relationships at the intergeneric level (Young and dePamphilis 2000). *MatK* gene has a high rate of substitution compared with other genes used in grass systematics, also this gene has a large proportion of variation at the nucleic acid level at first and second codon position, low transition/transversion ratio and is characterized by the presence of mutationally conserved sectors. All these features of the *matK* gene are useful to determine the relationships of family and species (Liang and Hilu 1996).

The aim of this research is to investigate the genetic relationship among the following 20 different *Triticum* species: 3 diploid *Triticum monococcum* L. (einkorn wheat), 11 tetraploid species (one *T. dicoccon* subsp. *dicoccon* (emmer), 2 *T. turgidum* subsp. *dicocoides* (wild emmer) and 8 *T. turgidum* subsp. *durum* (Desf.) (durum or macaroni wheat)) and 6 hexaploid *T. aestivum* (common wheat) were collected from different countries by using one type of DNA barcodes like *matK* gene and its translated amino acid sequence (151 amino acid) that form maturase K like protein.

Materials and methods

Plant materials

Twenty different *Triticum* species such as a diploid (*Triticum monococcum* L. A^mA^m), a tetraploid (*Triticum turgidum* subsp. *dicocoides*, *Triticum dicoccon* subsp. *dicoccon*, and *Triticum turgidum* subsp. *durum* (Desf.) BBA⁴A⁴), and a hexaploid (*Triticum aestivum* BBA⁴A⁴DD) were obtained from International Center for Agricultural Research in the Dry Areas (ICARDA, Aleppo, Syria), Leibniz Institute of Plant Genetics and Crop Plant Research (IPK, Gatersleben, Germany), Agricultural Research Center (ARC, Giza, Egypt) and Egyptian National Gene Bank (Agricultural Research Center, Giza, Egypt) as mentioned in Table 1.

Genomic DNA isolation

Genomic DNA was isolated from 100 mg young leaves samples using Gene Jet Plant Genomic DNA purification

Mini Kits (Thermo scientific K0791). The extracted DNA was assessed by agarose gel electrophoresis and spectrophotometry (NanoDrop 2000; Thermo Scientific) and diluted to 50 ng/μl, then used as a template for PCR reaction (Golovnina et al. 2007).

MatK primer design

Seven *matK* gene sequences of different *Triticum* species were retrievable from the National Center for Biotechnology Information (NCBI) database (GenBank). The used sequences have accession numbers DQ420054.1 (*T. monococcum*, partial sequence), KC608185.1 (*T. monococcum* subsp. *aegilopoides*, partial sequence), KC608186.1 (*T. monococcum* subsp. *aegilopoides*, partial sequence), KC608208.1 (*T. turgidum* subsp. *dicoccon*, partial sequence), KC608210.1 (*T. turgidum* subsp. *dicoccon*, partial sequence), DQ420019.1 (*T. aestivum*, partial sequence), DQ420050.1 (*T. aestivum*, partial sequence), and AF164405.1 (*T. aestivum*, complete sequence). Then, the downloaded *matK* gene sequences with these accession numbers were saved in fasta files then aligned by mega program version 6. The sequence from base 123 to base 644 bp was commonly present in all aligned sequences with length about 521 bp, this part of the sequence was used for designing the *matK* primer using online program Primer 3 (version 4) (<http://bioinfo.ut.ee/primer-3-0.4.0>). The *matK* primer forward 5'-ACCTGTGGAAATAGTTGTTAGTTGT-3' and reverse 5'-CCAATTCGAATAGTAGTTGAGAAAG-5' was designed to amplify 454 bp only from *matK* sequence. After that, the designed primer was tested in silico by aligning the retrievable complete *matK* gene sequence of *Triticum* with the designed *matK* primer to ensure that this primer was already a specific *matK* primer and attached with *Triticum matK* gene by 100%.

PCR amplification

The PCR reaction was carried out in duplicate in a T100™ Thermal Cycler (Bio-Rad) in the final volume of 25 μl. The single PCR reaction mixture contained: 5× Taq Buffer, MgCl, 0.2 mM dNTP, 10 pM of each primer, 50 ng genomic DNA, and 1 U Go Taq DNA Polymerase (Promega, USA). The thermal profile used was 95 °C for 4 min followed by 35 cycles of 95 °C for 30 s, 57 °C for 1 min, and 72 °C for 1 min, and a final extension at 72 °C for 5 min. PCR products were checked by running on 1.5% agarose gel containing ethidium bromide in 1X TAE buffer (pH 8.0). The gel was analyzed and archived using the Molecular Imager® GelDoc™XR software. Bands were scored and analyzed with the Quantity One software (Bio-Rad). The size of the products was determined by comparison with 100 bp DNA Ladder H3 RTU (GeneDirex, cat no. DM003-R500). The sequences isolated in this paper have been deposited in the GenBank

Table 1 The scientific and common name of 20 different *Triticum* species from a different country with code name and its GenBank accession numbers

Sample	Scientific name	Common name	Genome(s) and chromosome no.	Code name	Origin	GenBank accession numbers.	Collect from
Sample (1)	<i>T. monococcum</i> L.	Einkorn	A ^m A ^m (2n = 14)	IG 109083	Iraq	MN047218	ICARDA
Sample (2)	<i>T. monococcum</i> L.	Einkorn	A ^m A ^m (2n = 14)	IG 113259	Iran	MN047219	
Sample (3)	<i>T. monococcum</i> L.	Einkorn	A ^m A ^m (2n = 14)	IG 44936	Syrian	MN047220	
Sample (4)	<i>T. turgidum</i> subsp. dicoccoides	Wild emmer	BBA ^u A ^u (2n = 4x = 28)	IG 46467	Syrian	MN047221	
Sample (5)	<i>T. turgidum</i> subsp. dicoccoides	Wild emmer	BBA ^u A ^u (2n = 4x = 28)	IG 46447	Syrian	MN047222	
Sample (6)	<i>T. dicoccon</i> subsp. dicoccon	Emmer	BBA ^u A ^u (2n = 4x = 28)	TRI 28920	Ursprungsland	MN062364	IPK
Sample (7)	<i>T. turgidum</i> subsp. durum (Desf.)	Durum or macaroni wheat	BBA ^u A ^u (2n = 4x = 28)	TRI 28834	Turkey	MN062365	
Sample (8)	<i>T. turgidum</i> subsp. durum (Desf.)	Durum or macaroni wheat	BBA ^u A ^u (2n = 4x = 28)	TRI 27360	Italien	MN062366	
Sample (9)	<i>T. turgidum</i> subsp. durum (Desf.)	Durum or macaroni wheat	BBA ^u A ^u (2n = 4x = 28)	TRI 27284	Italien	MN062367	
Sample (10)	<i>T. turgidum</i> subsp. durum (Desf.)	Durum or macaroni wheat	BBA ^u A ^u (2n = 4x = 28)	TRI 19242	Iran	MN062368	
Sample (11)	<i>T. turgidum</i> subsp. durum (Desf.)	Durum or macaroni wheat	BBA ^u A ^u (2n = 4x = 28)	TRI 19223	Egypt	MN062369	
Sample (12)	<i>T. turgidum</i> subsp. durum (Desf.)	Durum or macaroni wheat	BBA ^u A ^u (2n = 4x = 28)	Cultivar, Sohag 4	Egypt	MN062370	ARC
Sample (13)	<i>T. turgidum</i> subsp. durum (Desf.)	Durum or macaroni wheat	BBA ^u A ^u (2n = 4x = 28)	Landraces, Sohag, Almonshaah 34	Egypt	MN062371	Egyptian National Gene Bank
Sample (14)	<i>T. turgidum</i> subsp. durum (Desf.)	Durum or macaroni wheat	BBA ^u A ^u (2n = 4x = 28)	Landraces, Sohag, Almonshaah 41	Egypt	MN062372	
Sample (15)	<i>T. aestivum</i>	Common wheat	BBA ^u A ^u DD (2n = 6x = 42)	TRI 28936	Indian	MN062373	IPK
Sample (16)	<i>T. aestivum</i>	Common wheat	BBA ^u A ^u DD (2n = 6x = 42)	TRI 13955	Libyan	MN062374	
Sample (17)	<i>T. aestivum</i>	Common wheat	BBA ^u A ^u DD (2n = 6x = 42)	Cultivar, Giza 168 (MIL/BUC//Seri CM93046-8 M-OY-0 M-2Y-0B)	Egypt	MN062375	ARC
Sample (18)	<i>T. aestivum</i>	Common wheat	BBA ^u A ^u DD (2n = 6x = 42)	Cultivar, Sids 4 (Maya'S'/Mon'S'// CMH74A-592/3/ Giza157*2SD-3SD-2SD-OSD)	Egypt	MN062376	
Sample (19)	<i>T. aestivum</i>	Common wheat.	BBA ^u A ^u DD (2n = 6x = 42)	Landraces, New Valley, Dakhla 7	Egypt	MN062377	Egyptian National Gene Bank
Sample (20)	<i>T. aestivum</i>	Common wheat	BBA ^u A ^u DD (2n = 6x = 42)	Landraces, Qena, Nag Hamad 27	Egypt	MN062378	

nucleotide sequence database (National Center for Biotechnology Information (NCBI)) under accession numbers MN047218, MN047219, MN047220, MN047221, MN047222, MN062364, MN062365, MN062366, MN062367, MN062368, MN062369, MN062370, MN062371, MN062372, MN062373, MN062374, MN062375, MN062376, MN062377, and MN062378 (Table 1).

Phylogenetic analysis

The chromatogram data were visualized by using the Bio-Edit program version 3 (Hall 1999). The nucleotide sequences were aligned with the Clustal W multiple sequence alignment program. The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura 3-parameter model (Tamura

1992). Statistical support for each constructed tree was provided by two statistical data analysis as bootstrapping (1000 replications) and pairwise distance. Total nucleotide length (bp), estimates of evolutionary divergence between sequences, percentage of nucleotide composition and polymorphism estimation, maximum likelihood of substitution matrix and maximum likelihood of transition/transversion bias were calculated by MEGA 6.06 program (Tamura et al. 2013).

Amino acid sequence

The nucleotide sequences were translated into amino acid sequences by the ExPASy online program (<https://web.expasy.org/translate>) for each studied *Triticum* species. The amino acid sequences were aligned with the Clustal W multiple sequence alignment program to construct the phylogenetic tree. The evolutionary history was inferred by using the maximum likelihood method based on the Tamura 3-parameter model (Tamura 1992). Statistical support for each constructed tree was provided by two statistical data analysis as bootstrapping (1000 replications) and pairwise distance.

Results

The selected portion of the *matK* gene was successfully amplified, then amplicons were sequenced and deposited in the GenBank under accession numbers MN047218, MN047219, MN047220, MN047221, MN047222, MN062364, MN062365, MN062366, MN062367, MN062368, MN062369, MN062370, MN062371, MN062372, MN062373, MN062374, MN062375, MN062376, MN062377, and MN062378 for all studied *Triticum* species. (Fig. 1 and Table 1). The length of the amplified *matK* gene was about 454 bp in all studied samples (partial gene) with 189 monomorphic nucleotide positions and 265 polymorphic sites (58.37% polymorphism). The GC% content average was found to be around 35.3% in all tested samples.

The nucleotide sequences of all studied species were analyzed by Tamura (1992) model to estimate the rates of different transitional and transversional substitutions as shown in Table 2. Base substitution mutation is the base of single-nucleotide polymorphism (SNP) which is either involves a transition (pyrimidines/pyrimidines or purines/purines) or transversions (pyrimidines against purines or vice versa) exchange. The estimated transition/transversion bias (R) is 0.99. Substitution pattern and rates were estimated, the nucleotide frequencies are $A = 32.34\%$, $T/U = 32.34\%$, $C = 17.66\%$, and $G = 17.66\%$.

Molecular phylogenetic analysis based on DNA sequence of partial *matK* gene

The sequence of the chloroplast *matK* gene was deciphered to verify the phylogenetic relationships of studied *Triticum* species. The evolutionary history was conducted using the maximum likelihood method depending on the Tamura 3-parameter model by two statistical data analysis bootstrapping and pairwise distance. The two types of data analysis gave the same phylogenetic tree result (Figs. 2 and 3). The phylogenetic tree divided all studied sample (20 *Triticum* species) into two groups A and B. Group A (green color) represented the diploid *Triticum* species $2n = 2x = 14$ (*T. monococcum* L. $A^m A^m$) with common name Einkorn collected from three different countries Iraq (IG 109083), Iran (IG 113259), and Syrian (IG 44936). This group was split into two subgroups: the first subgroup contained *T. monococcum* L. from Iran and Syrian while the second sub-group contained *T. monococcum* L. from Iraq only. Group B was split into two subgroups, I and II. Subgroup I represented the hexaploid *Triticum* species (red color) $2n = 6x = 42$ (*T. aestivum* ($BBA^u A^u DD$)) and subgroup II represented the tetraploid species (blue color) $2n = 4x = 28$ (*T. turgidum* subsp. *dicoccoides* $BBA^u A^u$ (Wild emmer), *T. dicoccon* subsp. *dicoccon* $BBA^u A^u$ (emmer) and *T. turgidum* subsp. *durum* $BBA^u A^u$ (macaroni wheat)). It

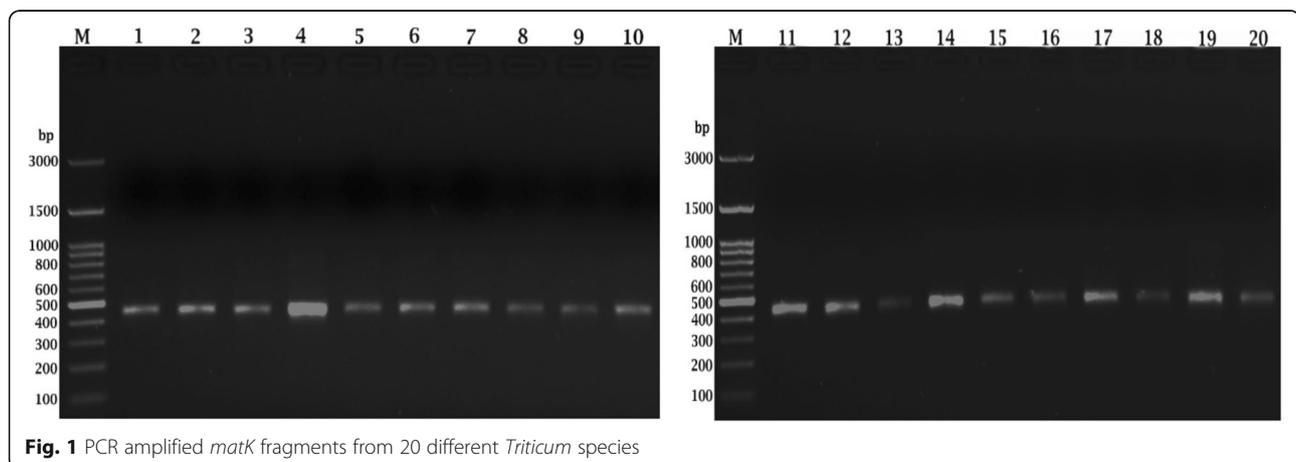


Fig. 1 PCR amplified *matK* fragments from 20 different *Triticum* species

Table 2 Maximum likelihood estimate of substitution matrix

	A	T/U	C	G
A	–	7.74	4.23	9.20
T/U	7.74	–	9.20	4.23
C	7.74	16.86	–	4.23
G	16.86	7.74	4.23	–

The likelihood of substitution (*r*) from one base (row) to another base (column). Rates of different transitional substitutions are shown in bold and those of transversal substitutions are shown in italics

was observed from the subgroup I (red color) that *T. aestivum* from Indian (TRI 28936) and Libyan (TRI 13955) were closely related to each other while *T. aestivum* accessions (Egyptian cultivar, sides 4 and Egyptian landraces, Qena, Nag Hamad 27) were different from each other and from *T. aestivum* accessions collected from Indian and Libyan, also *T. aestivum* accessions (Egyptian cultivar, Giza 168 and Egyptian landraces, New Valley, Dakhla 7) were different from each other and from all other *T. aestivum* accessions. The subgroup

II (blue color) was divided into two clusters. The first cluster was split into two subclusters, the first subcluster contained *T. turgidum* subsp. dicoccoides (wild emmer) from Syrian with code number IG 46467 and IG 46447, this indicated that these two species were closely related to each other while the second subcluster contained *T. dicoccon* subsp. dicoccon (emmer) from Ursprungsland (TRI 28920) only. The second cluster contained *T. turgidum* subsp. durum (Desf), this cluster was divided into two sub-cluster. The first subcluster contained *T. turgidum* subsp. durum from Turkey (TRI 28834) and Iran (TRI 19242), these two species were closely related to each other. The second subcluster was divided into two sections, the first section contained *T. turgidum* subsp. durum from Italian (TRI 27360 and 27284) which were closely related to each other. The second section was divided into two subsections; the first subsection contained *T. turgidum* subsp. durum from Egypt (TRI 19223) and Egyptian cultivar Sohag 4 while the second subsection contained *T. turgidum* subsp. durum Egyptian landraces from Sohag, Almonshaah 34 and Sohag, Almonshaah 41.

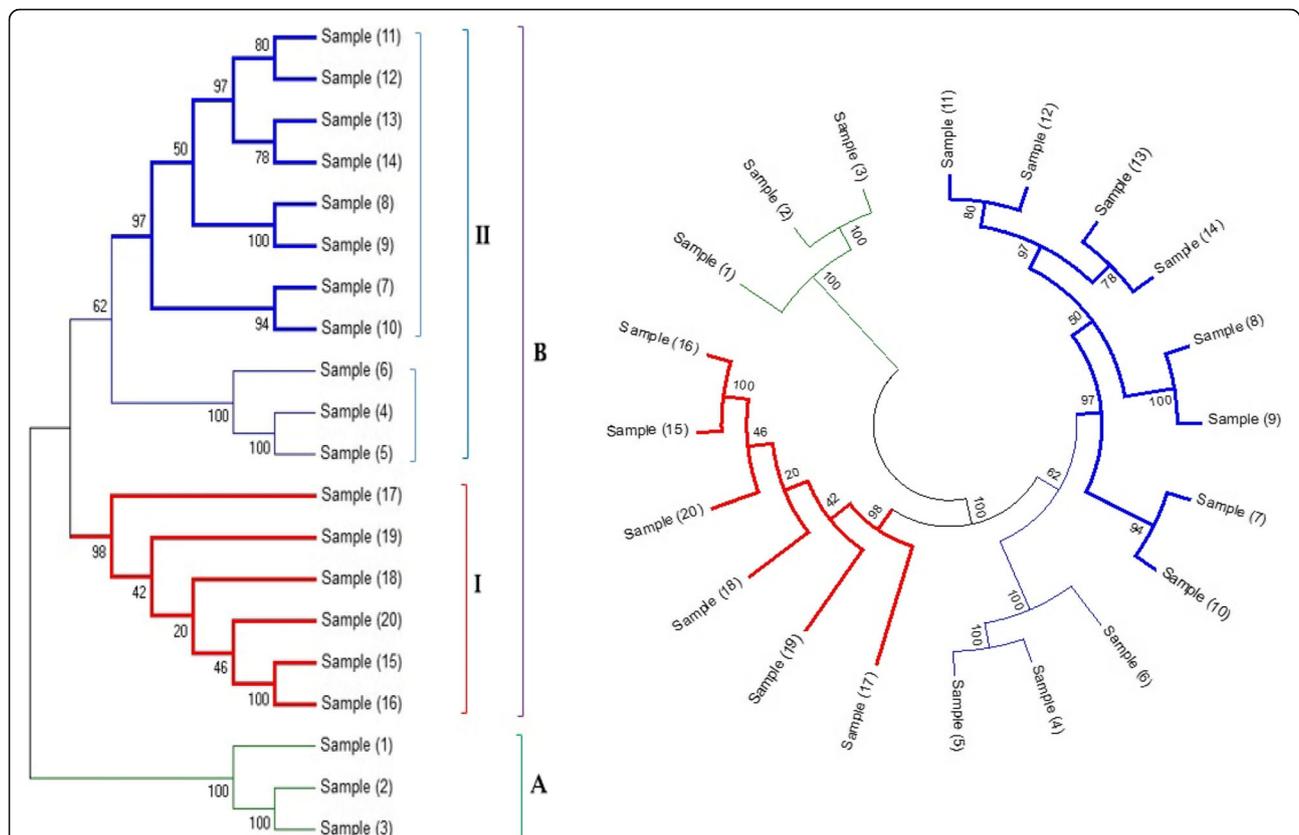
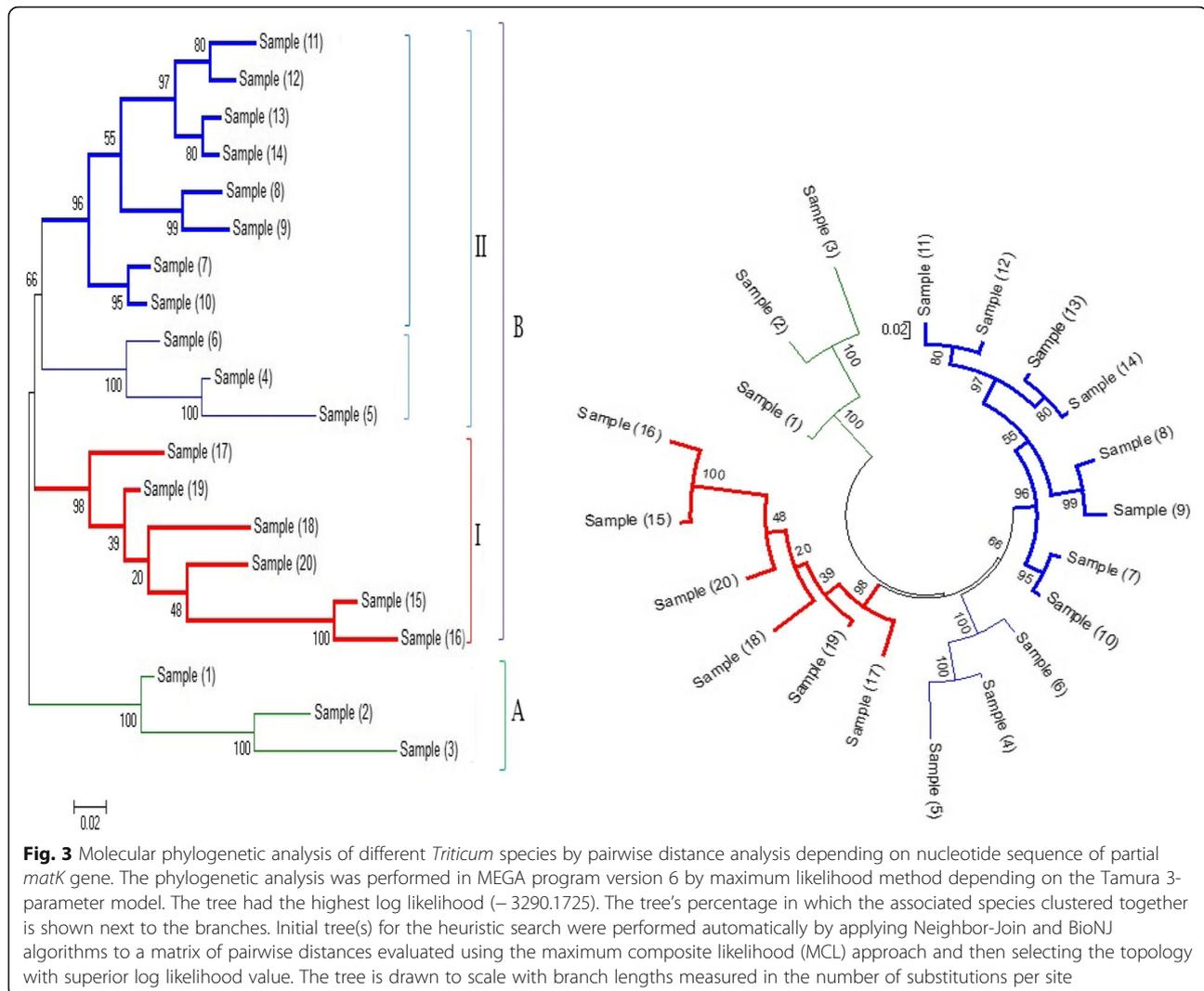


Fig. 2 Molecular phylogenetic analysis of different *Triticum* species by bootstrapping analysis depending on nucleotide sequence of partial matK gene. The phylogenetic analysis was performed in MEGA program version 6 by maximum likelihood method depending on the Tamura 3-parameter model. The bootstrap consensus tree deduced from 1000 replicates is obtained to determine the evolutionary history of the species analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are excluded. The percentage of replicate trees in which the associated species clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985)



Estimates of evolutionary divergence between sequences

The number of base substitutions per site was estimated between nucleotide sequences (454 bp) of all studied *Triticum* species as shown in Table 3. All ambiguous positions were removed for each sequence pair. Analyses were performed using the Tamura 3-parameter model with MEGA program version 6. It was observed from Table 3 and Figs. 2 and 3 that the highest evolutionary divergence of studied species was found between *T. monococcum* L. from Syrian IG 44936 and *T. turgidum* subsp. dicoccoides from Syrian IG 46447 was 0.37, this indicated that these two species were highly different. While the least evolutionary divergence between *T. turgidum* subsp. durum (Turkey TRI 28834 and Iran TRI 19242) and also between *T. turgidum* subsp. durum Egyptian landraces from Sohag, Almonshaah 34 and Sohag, Almonshaah 41 were 0.02, this indicated that every two species with high similarity. The evolutionary divergence between Egyptian *Triticum aestivum*

cultivars (Giza 168 and sids 4) was 0.12 while The evolutionary divergence between Egyptian *Triticum aestivum* landraces (New Valley, Dakhla 7 and Qena, Nag Hamad 27) was 0.08, this indicated that the difference that was found between both Egyptian cultivars and Egyptian landraces was relatively low.

Estimates of base composition bias difference between sequences

From the analysis of all nucleotide sequences, the difference in base composition bias per site was compute recorded in Table 4 (Kumar and Gadagkar 2001). Even when the substitution patterns are homogeneous among lineages, the compositional distance will correlate with the number of differences between sequences. It was observed from Table 4 and Figs. 2 and 3 that the highest compositional distance found between *T. monococcum* L. (Iran, IG 113259) and *T. turgidum* subsp. dicoccoides (Syrian, IG 46447) was 0.83. While *T. turgidum* subsp.

Table 3 Estimates of evolutionary divergence between sequences of 20 different *Triticum* species

Sample (1)	0.00																			
Sample (2)	0.11	0.00																		
Sample (3)	0.17	0.12	0.00																	
Sample (4)	0.19	0.27	0.28	0.00																
Sample (5)	0.26	0.34	0.37	0.08	0.00															
Sample (6)	0.15	0.25	0.29	0.07	0.13	0.00														
Sample (7)	0.14	0.22	0.26	0.17	0.25	0.14	0.00													
Sample (8)	0.18	0.26	0.30	0.21	0.29	0.18	0.10	0.00												
Sample (9)	0.18	0.26	0.30	0.20	0.29	0.17	0.10	0.05	0.00											
Sample (10)	0.14	0.22	0.25	0.17	0.25	0.14	0.02	0.09	0.10	0.00										
Sample (11)	0.17	0.25	0.29	0.19	0.28	0.16	0.12	0.12	0.12	0.11	0.00									
Sample (12)	0.16	0.24	0.29	0.18	0.26	0.15	0.10	0.12	0.12	0.11	0.04	0.00								
Sample (13)	0.17	0.25	0.29	0.19	0.27	0.16	0.11	0.11	0.12	0.11	0.06	0.06	0.00							
Sample (14)	0.16	0.24	0.27	0.19	0.27	0.16	0.10	0.11	0.12	0.10	0.07	0.06	0.02	0.00						
Sample (15)	0.22	0.30	0.34	0.26	0.34	0.23	0.21	0.24	0.24	0.20	0.22	0.23	0.23	0.22	0.00					
Sample (16)	0.22	0.31	0.35	0.28	0.36	0.23	0.21	0.24	0.24	0.20	0.22	0.23	0.24	0.22	0.05	0.00				
Sample (17)	0.17	0.25	0.29	0.19	0.27	0.16	0.15	0.20	0.19	0.15	0.17	0.17	0.17	0.16	0.17	0.18	0.00			
Sample (18)	0.19	0.28	0.31	0.22	0.29	0.19	0.17	0.22	0.22	0.17	0.20	0.19	0.19	0.17	0.17	0.19	0.12	0.00		
Sample (19)	0.15	0.24	0.28	0.17	0.24	0.14	0.13	0.18	0.17	0.13	0.16	0.15	0.15	0.14	0.14	0.16	0.07	0.08	0.00	
Sample (20)	0.18	0.27	0.29	0.21	0.29	0.19	0.17	0.21	0.21	0.17	0.19	0.19	0.19	0.17	0.14	0.17	0.11	0.11	0.08	0.00
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20

1 *T. monococcum* L. (Iraq, IG109083), 2 *T. monococcum* L. (Iran, IG113259), 3 *T. monococcum* L. (Syrian, IG44936), 4 *T. turgidum* subsp. *dicoccoides* (Syrian, IG46467), 5 *T. turgidum* subsp. *dicoccoides* (Syrian, IG46447), 6 *T. dicoccon* subsp. *dicoccon* (Ursprungsland, TRI 28920), 7 *T. turgidum* subsp. *durum* (Turkey, TRI28834), 8 *T. turgidum* subsp. *durum* (Italien, TRI27360), 9 *T. turgidum* subsp. *durum* (Italien, TRI 27284), 10 *T. turgidum* subsp. *durum* (Iran, TRI19242), 11 *T. turgidum* subsp. *durum* (Egypt, TRI19223), 12 *T. turgidum* subsp. *durum* (Egyptian cultivar, Sohag 4), 13 *T. turgidum* subsp. *durum* (Egyptian landraces, Sohag, Almonshaah 34), 14 *T. turgidum* subsp. *durum* (Egyptian landraces, Sohag, Almonshaah 41), 15 *T. aestivum* (Indian, TRI28936), 16 *T. aestivum* (Libyan, TRI13955), 17 *T. aestivum* (Egyptian cultivar, Giza 168), 18 *T. aestivum* (Egyptian cultivar, Sids 4), 19 *T. aestivum* (Egyptian landraces, New Valley, Dakhla 7), 20 *T. aestivum* (Egyptian landraces, Qena, Nag Hamad 27)

durum from Italien (TRI 127360 and TRI 127284) and *T. turgidum* subsp. *durum* Egyptian landraces (Sohag, Almonshaah 34 and Sohag, Almonshaah 41) had not a compositional distance. The compositional distances between Egyptian *Triticum aestivum* cultivar sids 4 and two Egyptian *Triticum aestivum* landraces (New Valley, Dakhla 7 and Qena, Nag Hamad 27) were 0.07 and 0.09, respectively; this indicated that the composition distance between these two landraces and cultivar sids 4 was a very low value.

Molecular phylogenetic analysis based on amino acid sequence from partial *matK* gene translation

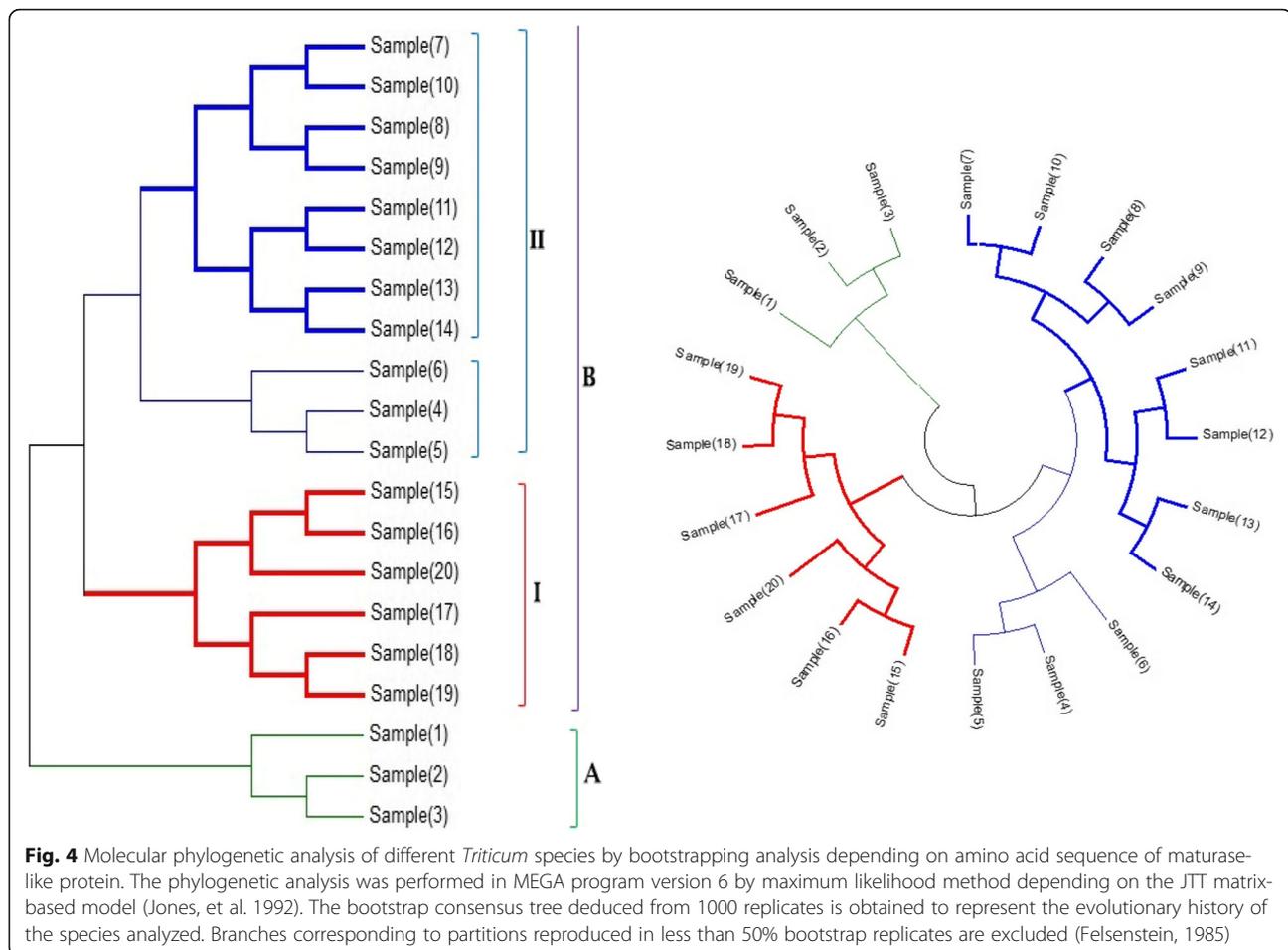
The translated amino acid sequences were used to detect the phylogenetic relationships between all studied species. The amino acid sequences of all studied species consist of 20 types of amino acid such as Alanine, cysteine, aspartic acid, glutamic acid, phenylalanine, glycine, histidine, isoleucine, lysine, leucine, methionine, asparagine, proline, glutamine, arginine, serine, threonine, valine, tryptophan, and tyrosine with a percentage average of its frequencies 1.06, 1.49, 4.77, 6.03, 8.44, 1.42, 4.11, 4.11, 4.44, 14.14, 1.26, 5.79, 6.62, 5.03, 4.74, 8.31, 0.73, 6.75, 1.16, and 6.19, respectively, as shown in Table 5.

The evolutionary history was conducted by using the Maximum Likelihood method based on the Tamura 3-parameter model by two statistical data analysis bootstrapping and pairwise distance analysis. The two types of data analysis gave different phylogenetic tree result (Figs. 4 and 5). The phylogenetic tree using bootstrapping analysis gave the same result based on the nucleotide sequences except for some differences that will be mentioned in the following context: group B was divided into two subgroups, I and II. Subgroup I (red color) was split into two clusters. The first cluster was divided into two subclusters, the first subcluster contained *T. aestivum* (Egyptian cultivar, sids 4 and Egyptian landraces, New Valley, Dakhla 7) while the second subcluster contained *T. aestivum* Egyptian cultivar, Giza 168 only. The second cluster was divided into two subclusters, and the first subcluster contained *T. aestivum* from Indian (TRI 28936) and Libyan (TRI 13955); this indicated that these two species were closely related to each other while the second subcluster contained *T. aestivum* Egyptian landraces, Qena, Nag Hamad 27 only. Subgroup II (blue color) was divided into two clusters. The first cluster was split into two subclusters, and the first subcluster contained *T. turgidum* subsp. *dicoccoides* (wild emmer) from Syrian with code number IG 46467 and IG 46447;

Table 5 Types of amino acid and its percentage frequencies in each sample and the average of all 20 different *Triticum* species in amino acid sequence

	Alanine	Cysteine	Aspartic acid	Glutamic acid	Phenylalanine	Glycine	Histidine	Isoleucine	Lysine	Leucine	Methionine	Asparagine	Proline	Glutamine	Arginine	Serine	Threonine	Valine	Tryptophan	Tyrosine
Sample (1)	1.32	2.65	4.64	4.64	7.28	1.32	3.97	8.61	4.64	13.91	1.32	6.62	6.62	5.30	5.30	7.28	0.00	7.28	1.32	5.96
Sample (2)	0.66	0.66	1.99	5.96	6.62	0.66	3.31	8.61	6.62	14.57	1.32	7.28	5.96	5.30	5.96	7.28	0.00	7.95	1.32	7.95
Sample (3)	1.32	1.32	2.65	5.96	7.95	0.66	1.99	9.27	3.97	15.23	1.32	7.28	5.96	4.64	5.30	6.62	1.32	9.27	0.66	7.28
Sample (4)	1.99	1.32	3.31	5.96	8.61	1.99	4.64	9.27	3.31	15.23	0.66	4.64	6.62	5.30	5.96	9.27	1.32	5.30	0.66	4.64
Sample (5)	3.31	1.99	3.97	5.96	7.95	2.65	5.30	7.95	2.65	13.91	0.66	5.96	6.62	3.97	5.96	8.61	1.32	5.30	0.66	5.30
Sample (6)	1.32	1.32	4.64	5.30	8.61	1.99	3.97	8.61	4.64	13.25	0.66	5.30	8.61	5.96	3.97	7.95	1.32	7.28	0.66	4.64
Sample (7)	0.66	1.32	7.95	3.97	8.61	1.32	3.31	7.28	5.30	13.25	1.32	4.64	8.61	5.30	4.64	8.61	0.00	7.28	0.66	5.96
Sample (8)	0.66	1.99	4.64	7.28	9.27	1.99	2.65	7.28	3.97	13.91	1.32	5.30	5.96	3.97	5.30	9.27	1.32	6.62	0.66	6.62
Sample (9)	0.66	1.32	5.96	5.30	9.27	1.99	3.97	5.96	3.97	13.91	1.32	6.62	5.96	4.64	5.96	8.61	0.00	8.61	0.66	5.30
Sample (10)	0.66	1.32	6.62	4.64	8.61	1.32	3.31	7.28	5.30	14.57	1.32	5.30	6.62	5.30	4.64	8.61	0.00	7.28	1.32	5.96
Sample(11)	0.66	1.99	5.30	6.62	7.95	1.32	3.31	6.62	4.64	14.57	1.32	5.96	7.95	3.97	4.64	5.30	0.66	8.61	1.32	7.28
Sample (12)	0.66	0.66	5.30	7.28	8.61	1.32	3.31	7.95	5.30	14.57	1.99	5.30	7.28	4.64	3.97	7.28	0.00	6.62	0.66	7.28
Sample (13)	0.66	0.66	3.97	7.28	8.61	0.66	3.31	6.62	4.64	13.91	1.99	5.96	5.30	4.64	5.30	8.61	0.66	8.61	1.32	7.28
Sample (14)	0.66	0.66	5.30	7.28	8.61	0.66	3.31	7.28	4.64	14.57	1.99	5.30	4.64	5.30	5.30	8.61	0.00	7.28	1.32	7.28
Sample(15)	0.66	1.99	4.64	6.62	9.27	0.66	5.30	5.30	3.97	12.58	1.32	5.30	5.96	5.30	3.31	9.93	2.65	5.96	3.31	5.96
Sample (16)	1.32	1.99	3.97	6.62	10.60	1.32	3.97	5.96	4.64	12.58	0.66	5.96	5.96	5.96	3.31	7.95	1.99	5.30	3.31	6.62
Sample (17)	1.32	1.32	4.64	6.62	7.28	1.99	5.30	5.96	3.97	15.23	1.32	6.62	6.62	5.96	5.30	8.61	0.66	5.30	0.66	5.30
Sample (18)	0.66	1.99	5.30	5.30	8.61	1.32	7.28	9.27	3.97	13.91	0.66	5.30	6.62	5.30	3.31	9.93	0.66	4.64	0.66	5.30
Sample (19)	1.32	1.99	5.30	5.96	7.28	1.32	4.64	9.27	4.64	14.57	0.66	5.96	7.28	4.64	3.31	8.61	0.00	5.96	0.66	6.62
Sample (20)	0.66	1.32	5.30	5.96	9.27	1.99	5.96	5.96	3.97	14.57	1.99	5.30	7.28	5.30	3.97	9.27	0.66	4.64	1.32	5.30
Average	1.06	1.49	4.77	6.03	8.44	1.42	4.11	4.11	4.44	14.14	1.26	5.79	6.62	5.03	4.74	8.31	0.73	6.75	1.16	6.19

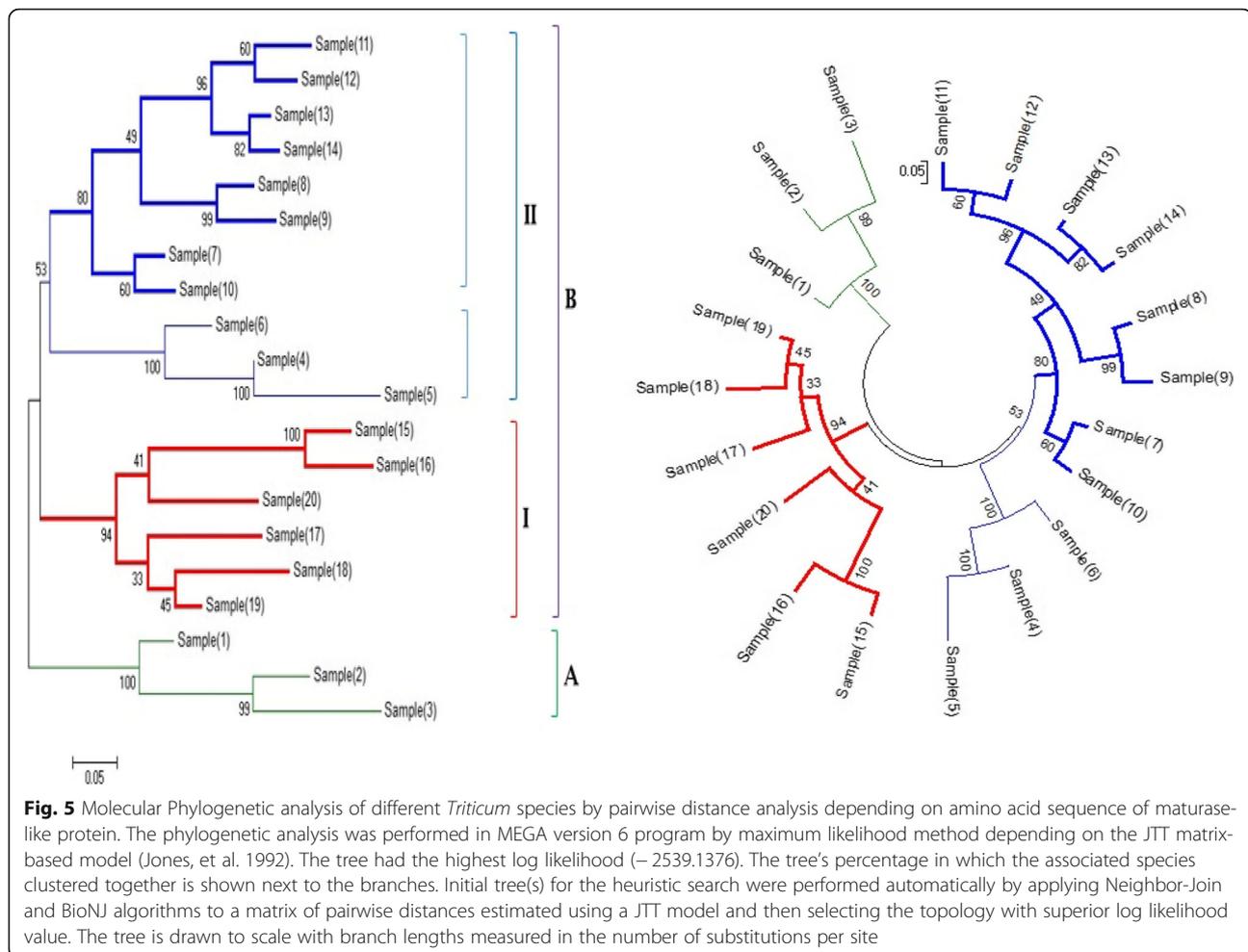
1 *T. monococcum* L. (Iraq, IG109083), 2 *T. monococcum* L. (Iran, IG113259), 3 *T. monococcum* L. (Syrian, IG44936), 4 *T. turgidum* subsp. *dicoccoides* (Syrian, IG46467), 5 *T. turgidum* subsp. *dicoccoides* (Syrian, IG46447), 6 *T. dicoccum* subsp. *dicoccum* (Ursprungland, TRI 28920), 7 *T. turgidum* subsp. *durum* (Turkey, TRI28834), 8 *T. turgidum* subsp. *durum* (Italy, TRI27360), 9 *T. turgidum* subsp. *durum* (Italy, TRI 27284), 10 *T. turgidum* subsp. *durum* (Iran, TRI19242), 11 *T. turgidum* subsp. *durum* (Egypt, TRI19223), 12 *T. turgidum* subsp. *durum* (Egyptian cultivar, Sohag 4), 13 *T. turgidum* subsp. *durum* (Egyptian landraces, Sohag, Almonshaah 4), 14 *T. turgidum* subsp. *durum* (Egyptian landraces, Sohag, Almonshaah 34), 15 *T. aestivum* (Indian, TRI28936), 16 *T. aestivum* (Libyan, TRI13955), 17 *T. aestivum* (Egyptian cultivar, Giza 168), 18 *T. aestivum* (Egyptian landraces, New Valley, Dakkla 7), 20 *T. aestivum* (Egyptian landraces, Qena, Nag Hamad 27)



obtained from combining *rbcL* and *atpB* sequences. Many studies indicated that Chloroplast *matK* barcode is an essential marker for discrimination of species or taxa (Newmaster and Ragupathy 2009; DeMattia et al. 2011), also this gene is used to resolve intergeneric or interspecific relations among flowering plants, such as *Malpighiaceae*, *Poaceae*, *Nicotiana*, *Orchidaceae* (Liang and Hilu 1996; Cameron et al. 2001; Salazar et al. 2003). The Plant Working Group (PWG) of the Consortium for the Barcoding of Life (CBOL) recommended that two regions of genes, *rbcL* and *matK*, could be adopted as the plant DNA barcode standard, and nuclear gene *ITS* as the supplement barcodes (CBOL Plant Working Group 2009) while Dizkirici et al. (2016) investigated the phylogenetic relationships between different *Triticum* and *Aegilops* species by nuclear *ITS* and chloroplast *matK* genes, they found that the relationships between different polyploid wheat and *Ae. speltoides* species that obtained from both chloroplast *matK* and nuclear *ITS* sequences were the same, this ensured the idea of co-inheritance of nuclear and chloroplast genomes where *Ae. speltoides* was the maternal donor.

Our results showed that the partial region of the *matK* gene amplified and sequenced gave high polymorphism between all studied species (58.37%). This nearly agreed with Skuza et al. (2019) who observed that nucleotide sequences had a high variability within *matK* and *rbcL* regions. Polymorphism of the sequences was 2.2% in the *rbcL* region, while in the *matK* region was 6.5%. The most variable *trnH-psbA* (15.6%) intergenic region was the most useful for rye barcoding so different DNA barcodes should be used. This indicates that the *matK* region is suitable for differentiation and discrimination between the studied species.

Awad et al. (2017) performed DNA barcoding using *matK* and *rbcL* barcodes to discriminate 18 different Egyptian *Triticum* accessions. They used a universal *matK* primer from previously published literature to amplify the *matK* gene and this primer gave 100% PCR amplification for 18 samples while DNA sequencing was successfully performed for 6 *matK* sequences only from 18 fragments. Also, the analysis of their results demonstrated a limited ability of *matK* gene in discrimination between six Egyptian *Triticum* accessions (Sinai-AlGora-



AlArish (114), Sinai-AlGora-AlArish (113), Northern coast-Raas ElHekma (117), Northern coast-Matroh (115), Bani Sweif 1, and Seds12). Their results showed the importance of in silico primer testing in the case of studies the closely related species. Their results conflicted with our results; this may be due to our using a specific primer that was designed from partial *matK* sequences using primer 3 version 4 online program. After that, we in silico tested the designed primer as mentioned in the primer design section. Also, we used *Triticum* species collected from different countries including Egypt. The Egyptian accessions that we used in our work differ from the Egyptian accession which used in their work; this may explain the expected reasons that distinguish Egyptian *Triticum aestivum* landraces and cultivars. Bafeel et al. (2011) found that using the universal *matK* primer leads to the inconsistent success rate of *matK* as a barcode so the universal primer needs further improvements.

The phylogenetic analysis considered the most effective method to determine the suitability of a DNA region for using as a barcode, because it should detect species-

specific clusters. From our results, we documented the relation between 20 different *Triticum* species diploid $2n = 2x = 14$ (*Triticum monococcum* L. $A^m A^m$ (einkorn)), tetraploidy $2n = 4x = 28$ (*Triticum turgidum* subsp. *dicoccoides* (wild emmer), *Triticum dicoccon* subsp. *dicoccon* (emmer) and *Triticum turgidum* subsp. Durum $BBA^u A^u$ (Durum or macaroni wheat)) and hexaploid $2n = 6x = 42$ (*Triticum aestivum* $BBA^u A^u DD$ (common wheat)) based on partial chloroplast *matK* gene sequence and its translated amino acid sequence, our phylogenetic tree that discriminated all studied species was consistent with Sourdille et al. 2001 and Feuillet et al. 2008, who reported that the genome allohexaploid species (*T. aestivum*) is composed of genomes A, B, and D (AABBDD; $2n = 42$) which is derived from three different diploid species. Whereas, *T. turgidum* subsp. durum is a tetraploid having A^u and B genomes (AABB; $2n = 28$), A^u genome is originated from *T. urartu* while B genome is originated from *Aegilops speltoides* commonly known as a wild or weedy goatgrass. The A^m genome is derived from *T. monococcum* L. (einkorn) which represents both the wild and cultivated varieties and is

generally known as *Triticum boeoticum* Bosis. Emend. Schiem. The D genome is derived from the wild or weedy grass *Aegilops tauschii* L.

It was found from the cytoplasmic studies that the Sitopsis diploid species (*Ae. Speltoides*) was considered a maternal donor in the original cross that resulted in the tetraploid *T. turgidum* (Vedel et al. 1978). Several other investigations by Bowman et al. 1983 and Dizkirci et al. 2016 also confirmed that *Ae. Speltoides* was considered as the maternal donor and the source of the B genome of *T. turgidum* and *T. aestivum*.

The current investigation suggests the effectiveness of *matK* gene sequence data to resolve the phylogenetic problem in *Triticum* species. Also, the sequence variation, mean evolutionary rates, patterns, and transition/transversion rate in the nucleotide sequence, nucleotide diversity of *matK* gene can be used for the interpretation of evolutionary relationship within interspecies level of *Tritium* species. Finally, the *matK* sequence can discriminate the closely related *Tritium* species. So these sequences can be used as a DNA barcode for *Triticum* species.

Conclusion

The *matK* sequence has an important role in discriminating the closely related *Triticum* species. So these sequences can be used as a DNA barcode for detecting the evolutionary history of *Triticum* species. It was found that there is a relation between hexaploid and tetraploid species because they are in the same group while Diploid species are in another group.

Abbreviations

bp: Base pair; C: Cytosine; DNA: Deoxyribonucleic acid; G: Guanine; *MatK* gene: Maturase encoding gene; MEGA software: Molecular Evolutionary Genetic Analysis software; NCBI: National Center for Biotechnology Information; PCR: Polymerase chain reaction; T: Thymine; U: Uracil

Acknowledgements

We would like to thank the National Research Centre, Dokki, Giza, Egypt, for supporting this research through the project No AR111105.

Authors' contributions

This work was carried out in collaboration between the two authors. Author SAO designed the study, wrote the protocol, and analyzed the data of this study. SAO and WAR managed the lab work and manage the literature searches and wrote this manuscript together. Both authors read and approved the final manuscript.

Funding

This research was funded by the project No AR111105.

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 23 June 2019 Accepted: 12 September 2019

Published online: 05 December 2019

References

- Ajmal AM, Gyulai G, Hidwegi N, Kerti B, Al Hemaid F, Pandey AK, Lee J (2014) The changing epitome of species identification—DNA barcoding. *Saudi J Biol Sci* 21:204–231
- Awad M, Fahmy RM, Mosa KA, Helmy M, El-Feky FA (2017) Identification of effective DNA barcodes for *Triticum* plants through chloroplast genome-wide analysis. *Comput Biol Chem* 71:20–31
- Bafeel SO, Arif IA, Bakir MA, Khan HA, Al Farhan AH, Al Homaidan AA, Ahamed A, Thomas J (2011) Comparative evaluation of PCR success with universal primers of maturase K (*matK*) and ribulose-1, 5-bisphosphate carboxylase oxygenase large subunit (*rbcl*) for barcoding of some arid plants. *Plant Omics* 4:195
- Bowman CM, Bonnard G, Dyer TA (1983) Chloroplast DNA variation between species of *Triticum* and *Aegilops*—location of the variation on the chloroplast genome and its relevance to the inheritance and classification of the cytoplasm. *Theor Appl Genet* 65:247–262
- Cameron KM, Chase MW, Anderson WR, Hills HG (2001) Molecular systematics of Malpighiaceae: evidence from plastid *rbcl* and *matK* sequences. *Am J Bot* 88:1847–1862
- CBOL Plant Working Group (2009) A DNA barcode for land plants. *Proc Natl Acad Sci U S A* 106:12794–12797
- China Plant BOL Group (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (*ITS*) should be incorporated into the core barcode for seed plants. *Proc Natl Acad Sci U S A* 108:19641–19646
- Cowan RS, Chase MW, Kress WJ, Savolainen V (2006) 300,000 species to identify: problems, progress and prospects in DNA barcoding of land plants. *Taxon* 55:611–616
- DeMattia F, Bruni I, Galimberti A, Cattaneo F, Casiraghi M, Labra M (2011) A comparative study of different DNA barcoding markers for the identification of some members of *Lamiaceae*. *Food Res Int* 44:693–702
- Dizkirci A, Kansu C, Onde S (2016) Molecular phylogeny of *Triticum* and *Aegilops* genera based on *ITS* and *MatK* sequence data. *Pak J Bot* 48(1):143–153
- Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. *Cell* 127:1309–1321
- Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783–791.
- Ferri E, Barbuto M, Bain O, Galimberti A, Uni S, Guerrero R, Ferté H, Bandi C, Martin C, Casiraghi M (2009) Integrated taxonomy: traditional approach and DNA barcoding for the identification of filarioid worms and related parasites (Nematoda). *Front Zool* 6:1
- Feuillet C, Langridge P, Waugh R (2008) Cereal breeding takes a walk on the wild side. *Trends Genet* 24:24–32
- Golovnina KA, Glushkov SA, Blinov AG, Mayorov VI, Adkison LR, Goncharov NP (2007) Molecular phylogeny of the genus *Triticum* L. *Plant Syst Evol* 264:195–216
- Gu YQ, Coleman-Derr D, Kong X, Anderson OD (2004) Rapid genome evolution revealed by comparative sequence analysis of orthologous regions from four *Triticeae* genomes. *Plant Physiol* 135:459–470
- Hall TA (1999) Bio-Edit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98
- Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003) Biological identification through DNA barcodes. *Proc R Soc Lond* 270:313–322
- Kumar S, Gadagkar SR (2001) Disparity Index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. *Genetics* 158:1321–1327
- Li XW, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S (2015) Plant DNA barcoding: from gene to genome. *Biol Rev Camb Philos Soc* 90:157–166
- Liang HP, Hilu KW (1996) Application of the *matK* gene sequences to grass systematics. *Can J Bot* 74:125–134
- McFadden ES, Sears ER (1946) The origin of *Triticum spelta* and its free-threshing hexaploid relatives. *J Hered* 37:81–89
- Newmaster SG, Ragupathy S (2009) Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, *Fabaceae*). *Mol Ecol Resour* 9: 172–180

- Notredame C, Higgins DG, Heringa J (2000) *T-Coffee*: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302:205–217
- Provan J, Wolters P, Caldwell KH, Powell W (2004) High-resolution organellar genome analysis of *Triticum* and *Aegilops* sheds new light on cytoplasm evolution in wheat. *Theor Appl Genet* 108:1182–1190
- Salazar GA, Chase MW, Arenas MAS, Ingrouille M (2003) Phylogenetics of Cranichideae with emphasis on Spiranthinae (*Orchidaceae*, *Orchidoideae*): evidence from plastid and nuclear DNA sequences. *Am J Bot* 90:777–795
- Savolainen V, Chase MW, Hoot SB, Morton CM, Soltis DE, Bayer C, Fay MF, De Buijn AY, Sullivan S, Qiu YL (2000) Phylogenetics of flowering plants based upon a combined analysis of plastid *atpB* and *rbcl* gene sequences. *Syst Biol* 49:306–362
- Skuzaj L, Filip E, Szućko I (2015) Intergenic spacer length variability in cultivated, weedy and wild rye species. *Open Life Sci* 10(1):175–181
- Skuzaj L, Szućko I, Filip E, Adamczyk A (2019) DNA barcoding in selected species and subspecies of Rye (*Secale*) using three chloroplast loci (*matK*, *rbcl*, *trnH-psbA*). *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* 47(1):54–62 DOI:47.15835
- Sourdille P, Tavaud M, Charmet G, Bernard M (2001) Transferability of wheat microsatellites to diploid Triticeae species carrying the a, B and D genomes. *Theor Appl Genet* 103:346–352
- Stoeckle M (2003) Taxonomy, DNA, and the bar code of life. *BioScience* 53:796–797
- Tamura K (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. *Mol Biol Evol* 9:678–687
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729
- Vedel F, Quetier F, Dosba F (1978) Study of wheat phylogeny by EcoRI analysis of chloroplastic and mitochondrial DNAs. *Plant Sci Lett* 13:97–102
- Young ND, dePamphilis CW (2000) Purifying selection detected in the plastid gene *matK* and flanking ribozyme regions within a group II intron of nonphotosynthetic plants. *Mol Biol Evol* 17:1933–1941
- Zhang W, Qu LJ, Gu H, Gao W, Liu M, Chen J, Chen Z (2002) Studies on the origin and evolution of the tetraploid wheats based on the internal transcribed spacer (*ITS*) sequences of nuclear ribosomal DNA. *Theor Appl Genet* 104:1099–1106

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
