Iraqi Journal of Agricultural Sciences –2026:57(3):676-688
P-ISSN: 0075- 0530 /E-ISSN: 2410-0862
DOI: https://doi.org/10.36103/4c5g0q37

SEQUENCE DIVERSITY AND MOLECULAR EVOLUTION ANALYSIS OF INTERNATIONAL TOMATO BASED ON THE ENTIRE ITS REGION

*[1]Ayyad W. Al-Shahwany, [2]Mohammed M. Hawash, [2]Thaer Hamid A. Hajeej
*[1]Dept. Biol., Coll. Sci., University of Baghdad, Iraq
[2]Anbar Educ. Direct., Ministry of Education ,Anbar, Iraq

ABSTRACT

This study aimed to assess the genetic diversity and molecular evolution of tomato (*Solanum lycopersicum*) populations from various countries, which hold significant potential for future breeding strategies and germplasm conservation. To achieve this, a total of 15 sequences deposited in GenBank were analyzed using the complete internal transcribed spacer (ITS) region of the nuclear ribosomal DNA (nrDNA). The spacer sequence lengths ranged from 156 base pairs (bp) in the Swedish tomato to 713 bp in the Palestinian tomato. A notable variation in GC content was observed, with the Thai tomato exhibiting the highest value (67.48%) and the South Korean tomato the lowest (49.56%). Phylogenetic trees were constructed using both the distance-based Neighbor-Joining (NJ) and Maximum Parsimony (MP) methods. Sequence analysis revealed 38 monomorphic (invariable) sites and 15 polymorphic sites, of which 14 were singleton variable sites and one was parsimony-informative. Alignment of the 15 sequences enabled the identification of five haplotypes. The estimated transition/transversion bias (R) was 17.339, indicating a greater frequency of transitions over transversions in this region. Neutrality tests, including Tajima's D and Fu and Li's statistics, produced statistically significant results. The highest levels of genetic diversity were observed in South Korean, Iraqi, and Indian tomato samples.

Keywords: GC content, Haplotypes, Mismatch, Nucleotide composition, PCA, Selective neutrality, *Solanum lycopersicum*

Received: 12/5/2025, Accepted: 17/8/2025, Published: 31/3/2026

INTRODUCTION

Tomatoes (*Solanum lycopersicum*) are a major dietary source, rich in a wide range of nutrients that promote a healthy life. The tomato is the most popular vegetable in the world and makes up 14% of global vegetable production. This is because its great supplier of micronutrients for the human diet. A comprehensive understanding and effective management of tomato genetic resources are essential for breeding programs. Tomato, a member of the Solanaceae family, belongs to *Solanum* section *Lycopersicon*, which includes the cultivated tomato and twelve wild relatives native to western South America (Ramírez Ojeda *et al*., 2021). The cultivated tomato has faced many bottlenecks over the millennia. The genetic diversity of tomatoes is estimated to have dropped dramatically over time. Thus, research on their origins variations could allow for significant advancements in describing the diversity of tomatoes. Plant conservation initiatives grew at the same time, ensuring the collection and conservation of wild species and landraces. This resulted in the identification of tomatoes wild cousins' genetic potential for breeding. At the same time, the ecological and taxonomic diversity exhibited by tomato has established it as a model species in evolutionary research. Since the middle of the 20th century, new techniques such as controlled hybridization have made it possible to cross cultivated and wild tomatoes. Such advanced molecular approaches have resulted in a more comprehensive grasp of plant breeding genetic control (Bauchet and Causse, 2012). As tomato's genome has been entirely sequenced a significant advancement in our

Iraqi Journal of Agricultural Sciences –2026:57(3):676-688
P-ISSN: 0075- 0530 /E-ISSN: 2410-0862
DOI: https://doi.org/10.36103/4c5g0q37

understanding of tomato diversity using sequencing techniques has been reached. Genomes and associated levels of expression can be read comprehensively in variety of plants. The use of bioinformatics approaches in genome-wide association studies or population studies will make it easier to manage germplasm and characterise complicated traits genetically. A common source of evolutionary information spanning the whole spectrum of life is the ribosomal RNA (rRNA) genes and their spacer regions (Dixon and Hillis, 1993). The fact that the rDNA locus serves the same purpose in all free-living species may be the reason for its prominence in phylogenetics. They are structurally the same or nearly identical across many different taxa. The plant genome contains tandem arrays of the nuclear ribosomal locus that codes for the large subunit. The rRNA's subunit-coding regions have evolved at the slowest rate. It serves as a marker for phylogenetic reconstruction on several levels. For phylogeny reconstruction, it has gained popularity since Porter and Collins (1991) employed it for the first time. Almost all organisms include the ITS, which is a component of the transcriptional unit of rDNA (Calonje *et al*., 2008). Moreover, this high rate of divergence represents a valuable resource for investigating population expansion and growth variation (Strasburg *et al*., 2012). The spacer regions are also commonly used in phylogenetics, such as the external transcribed spacers (ETS) and internal transcribed spacers (ITS). That offers the excellent opportunity to investigate intricate evolutionary processes, including array duplications or reticulate occurrences. These mechanisms are understood in light of the fact that many copies of rDNA regions have been homogenised through coordinated evolution . The ITS region lies between the 18S and 28S rRNA genes within the rDNA locus. The ITS region consists of three parts: the ITS1, ITS2, and the highly conserved 5.8S rDNA exon located between them (Baldwin *et al*., 1995). In plants, ribosomal RNA copy numbers and individual rDNA repeat lengths vary considerably. The length polymorphisms come from changes in the quantities of repeated sequences inside the intergenic spacer. In contrast to maternally inherited mitochondrial and chloroplast markers, this region has the following advantages: biparental inheritance, easy PCR amplification with multiple universal primers available for different types of organisms, multicopy structure, moderate size allowing easy sequencing and, according to published studies, exhibiting a level of variation that makes it suitable for evolutionary studies at the species or generic level (Baldwin *et al*., 1995; Liston *et al*., 1996). A careful assessment of polymorphisms is essential, as their presence may limit the accuracy of phylogenetic interpretations. Over the past decade, nuclear ribosomal DNA sequences from the internal transcribed spacer region have served as a primary source for investigating low-level evolutionary relationships among plant species. Baldwin *et al*., (1996) in addition to Ainouche and Bayer (1997), reported homogeneous nrDNA arrays within individuals, likely due to concerted evolution mechanisms such as gene conversion and unequal crossing-over. However, other studies have demonstrated intra-individual nrDNA variation across different taxa. Since the 1980s, regions such as ETS, NTS, ITS1, and ITS2 have been shown to exhibit significant intra- and interspecific variability, making them valuable markers in phylogenetic and molecular evolution research. Nevertheless, the genetic diversity and molecular characteristics of global tomato cultivars remain insufficiently explored. This article will first demonstrate how tomato evolution worldwide has progressed from its early domestication to the present. We will investigate the importance of natural variation in tomato genetic resources, as it not only emphasises variety for cultivated tomato improvement but also provides insights into the development and genetic basis of complex traits. In the final section, we demonstrate how molecular markers have complemented our perspective.

MATERIALS AND METHODS

Source of sequences: This analysis used all sequences from various nations that were previously deposited in GenBank, National Center for Biotechnology Information (NCBI).

Iraqi Journal of Agricultural Sciences –2026:57(3):676-688
P-ISSN: 0075- 0530 /E-ISSN: 2410-0862
DOI: https://doi.org/10.36103/4c5g0q37

Sequences were extracted from the entire ITS region (ITS1, 5.8S, and ITS2) and were preserved and listed under the accession numbers shown in Table (1).

Sequence analysis: The nucleotide sequences were download, aligned and then assessed with MEGA version 12.0.7 (Kumar *et al*., 2024). A Maximum Composite Likelihood (MCL) was employed to assess the pairwise sequence divergence of tomato sequences in the ITS region. GC content was calculated with DAMBE v.6 (Xia, 2017), and sequence-specific GC content charts were generated using the online tool provided by Biologics Corp. (https://www.com/tools/GCcontent/). The resulting distance matrix was then calculated to created phylogenetic trees using the "Neighbour-Joining" (NJ) and "Maximum Parsimony" (MP) approach (Felsenstein, 1985) with 1000 bootstrap replications. Phylogenetic trees were inferred based on the computed genetic distance matrix using the Neighbor-Joining algorithm, as originally proposed by Saitou and Nei, (1987). Calculated the consistency indexes (CI), retention indexes (RI), and homoplasy index (HI). The transition/transversion ratio (ti/tv) was calculated according to the equation $R = [(A \cdot G \cdot k_1) + (T \cdot C \cdot k_2)] / [(A + G)(T + C)]$, in which A, G, C, and T denote the nucleotide frequency distributions (Kumar *et al*., 2024).The study revealed the number of substitutions per site nucleotide event over the sequences under study. The aligned MEGA data sequences were analyzed with DnaSP software (version 5.10.01) to estimate indices of genetic polymorphism (Librado and Rozas, 2009). To assess genetic diversity among sequences in the ITS region, utilise the nucleotide diversity indices {Pi} and the diversity of haplotype indices {Hd} (Nei and Tajima, 1983) along with the standard deviation. The average (pairwise nucleotide differences) (K) for selective neutrality was assessed using the Tajima D test (Tajima, 1989), as well as Fu and Li's for the D* and F* statistic (Fu and Li, 1993). However, to assess the demographics employing the distribution of the allelic frequency, along with the distribution of pairwise sequencing differences {mismatch distribution} (Rogers and Harpending, 1992). Thus, the chart curve clearly displayed the expected number of time generations to reach in equilibrium. The principal component analysis (PCA) scatter chart plot has been incorporated using PAST version 5.0.2 (Hammer *et al*., 2001), as well as the computed eigenvalue and variance. Further, the (NETWORK) application 4.6.1.0 (Bandelt et al., 1999) was used to generate an illustration of the genetic relationship between worldwide tomatoes determined with haplotype detection.

Table 1. Show ID accession numbers of the tomato in GenBank sequences, GC contents, and sequence lengths.

Accession sequences	GenBank accession number	GC content%	Sequence Length bp
American tomato	GQ221566	65.22	624
Bulgarian tomato	PP060550	64.21	447
Chinese tomato	MW018152	64.28	699
Indian tomato	MK975256	57.23	636
Iraqi tomato	ON167515	61.53	668
Italian tomato	MZ489691	64.21	447
Japanese tomato	AB373816	64.23	671
Nigerian tomato	OR809190	64.55	660
Palestinian tomato	JN713146	63.11	713
Saudi tomato	OQ152615	63.68	669
South Korean tomato	KC213747	49.56	684
Swedish tomato	OK073664	55.15	165
Swiss tomato	OQ910047	65.29	635
Thai tomato	MH718333	67.48	326
UK tomato	KY700408	63.09	259
Average		62.18	466.6

Iraqi Journal of Agricultural Sciences –2026:57(3):676-688
P-ISSN: 0075- 0530 /E-ISSN: 2410-0862
DOI: https://doi.org/10.36103/4c5g0q37

RESULTS AND DISCUSSION

Analysis of the genetic diversity

Sequence length: Analysis of the obtained sequences revealed variations in length and nucleotide composition (Table 1). ITS region ranged in length from 165 bp in Swedish tomato to 713 bp in Palestinian tomato. All of the variation results could be due to variations in the conditions and the methodology that researchers employed during the amplification process. In an analysis of 15 tomato ITS sequences from around the world, the average length of the ITS region was 466.6 bp. This is comparable to ITS lengths reported for other plant species, including Persian clover (695 bp) (Ansari *et al*., 2018), wheat (597–605 bp), and barley (595–598 bp) (Sharma *et al*., 2002). In the Asteraceae family, the length of the entire ITS region varies between 650 and 750 bp (Amar *et al*., 2012), whereas in *Ficus carica* the mean length of the complete ITS region is 697.5 bp (Baraket *et al*., 2013). Kehie *et al.* (2016) reported that the ITS region of Naga King Chilli has an average length of 620 bp. Among angiosperms, the ITS region typically varies from 565 to 700 bp. In contrast, other plant lineages, including Coniferales, Ginkgoales, Cycadales, and Gnetales, exhibit considerably larger ITS regions, ranging from 975 to 3125 bp (Liston *et al*., 1996). By contrast, it is larger than those found in Tunisian date palms, which range from 441 bp to 445 bp (Maina *et al*., 2019).

Nucleotide frequencies of nrDNA ITS region: In fact, for A, T, C, and G, the nucleotide composition was 19.30%, 17.53%, 33.86%, and 29.31%, respectively, closely matching the composition reported in Naga King Chilli (18.85%, 17.56%, 33.95%, and 29.64%, respectively) (Kehie *et al*., 2016). Likewise, the Tunisian figs (Baraket *et al*., 2013) showed that the complete ITS region exhibits a base composition of A (19.7%), T (18.6%), C (31.4%), and G (30.2%). In contrast, other species show different nucleotide compositions; in the Asteraceae family, A, T, C, and G constituted 25%, 24%, 26%, and 25%, respectively (Amar *et al*., 2012). In *Phoenix dactylifera* (Tunisian date palm), the internal transcribed spacer (ITS) displays nucleotide frequencies of 24.76% A, 25.67% T, 27.22% C, and 21.95% G (Maina *et al*., 2019). In *Trifolium resupinatum* (Persian clover), the internal transcribed spacer (ITS) shows average nucleotide frequencies of 23.7% A, 26.8% T, 22.5% C, and 26.9% G (Ansari *et al*., 2018).

GC content of the entire ITS sequence

The GC content of the amplified sequences varied for the entire ITS. Indeed, the percentage of GC varied from 49.56% in South Korean tomatoes to 67.48% in Thai tomatoes, with an average of 62.18% for all ITS sequences studied (Table 2 and figs. 1, 2). It appears that the variation is entirely due to adaptations to different climatic conditions and further substitutions occurring in nitrogenous bases or due to the influence of DNA methylation. Notably, the GC content obtained is highly consistent with that reported in other plant species, such as Naga King Chilli (63.59%) (Kehie *et al*., 2016), Tunisian figs (61.6%) (Baraket *et al*., 2013), and Italian *Quercus* spp. (63.9%) (Bellarosa et al., 2005). Further, Tunisian date palm (Maina *et al*., 2019) and Persian clover (Ansari *et al*., 2018) possess lower GC content in the ITS region (49.5% and 49.4%, respectively). Moreover, despite considering all that both have separate evolutionary origins (Baldwin *et al*., 1995), a theory of a co-evolutionary mechanism could offer an explanation for this outcome (Van der Sande *et al*., 1992).

Table 2. Nucleotide substitution rates inferred from the nrDNA ITS region of the tomato tree

	A	T	C	G
A	-	**0.45**	**0.86**	**32.46**
T	**0.49**	-	**27.06**	**0.74**
C	**0.49**	**14.01**	-	**0.74**
G	**21.38**	**0.45**	**0.86**	-

Note: Each value corresponds to the substitution probability (r) from the nucleotide in the row to that in the column. Transitional substitutions are highlighted in bold, while transversional substitutions are displayed in *italics*.

Iraqi Journal of Agricultural Sciences –2026:57(3):676-688
P-ISSN: 0075- 0530 /E-ISSN: 2410-0862
DOI: https://doi.org/10.36103/4c5g0q37

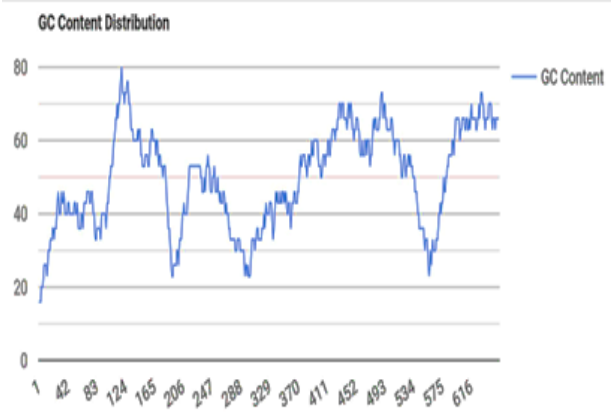

Fig. 1. The percentage and variation of GC content at 49.56% among rDNA sequences of South Korean tomatoes.

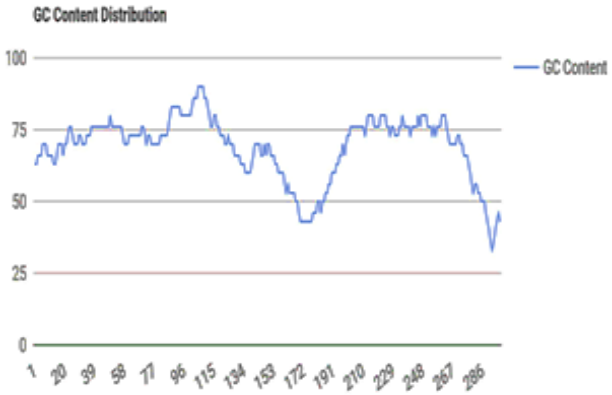

Fig. 2. The percentage and variation of GC content at 67.48% among rDNA sequences of Thai tomatoes.

Variability of Nucleotide composition and mutations: ITS sequences analysis revealed that the transition/transversion ratio was calculated as K1 equal to 43.573 for purine and K2 equal to 31.442 for pyrimidine bases. The R ratio for all bases was equal to 17.339 (Table 2). It is worth noting that, except for sites with ITS sequence gaps, 53 alignment positions were considered. The high ratio of R rate indicates that the rate of transitions was much greater than the transversions within the ITS region of tomato sequences. A total transition/transversion ratio (R) of 17.339 was observed in the entire ITS region, markedly higher than the ratio reported for *Hordeum spontaneum* (7.4) and *Triticum aestivum* (6.90) (Sharma *et al*., 2002), the Tunisian date palm R of 4.375 (Maina *et al*., 2019), *Capsicum* sp. (ti/tv) of 3.746 (Kehie *et al*., 2016), and the Asteraceae family (ti/tv) of 1.43 (Amar *et al*., 2012). However, it was much higher than the transition/transversion (ti/tv) ratio recorded in Tunisian figs (0.7) (Baraket *et al*., 2013) and in Iraqi date palm based solely on the ITS1 region, with values of 0.800 and 0.751 (Hawash and Al-Shamma, 2025a; Hawash and Al-Shamma, 2025b), respectively. Moreover, The different substitutions detected are given in Table 2 also and show that in tomato plants, the transitions of G/A and C/T are more frequent than the A/G and T/C transitions in the entire ITS region. Notably, the composition variation of nucleotide sequence alignment of ITS within a matrix of 838 characters has revealed the presence of 53 conserved sites, 38 monomorphic sites, 15 polymorphic sites, consisting of 14 singleton variable sites and one parsimony informative site, were detected. The singleton variables were at positions of 403, 407, 411, 412, 420, 427, 432, 433, 436, 437, 442, 446, 452, and 454, while the parsimony informative was at position 428. The ITS sequences of the nr DNA indicated signs of genetic diversity. The analysis of fifteen cultivars revealed the existence of five haplotypes. Table 3 shows the variety of haplotypes (Hd) and nucleotides (p), as well as their standard deviations. Calculations on all the tomato plant sequences under study revealed values of 0.476 ± SD 0.155 and 0.04223 ± SD 0.01849, respectively. *Solanum lycopersicum* seems to have lower nucleotide and haplotype diversities than other species like Tunisian figs (Hd = 0.996, Pi = 0.072) and *Capsicum* sp. (Hd = 1, Pi = 0.01499) (Baraket *et al*., 2013; Kehie *et al*., 2016). According to Maina *et al.* (2019), Tunisian palm displays a higher haplotype diversity (Hd = 0.552) while maintaining a low nucleotide diversity (π = 0.00155).

Iraqi Journal of Agricultural Sciences –2026:57(3):676-688
P-ISSN: 0075- 0530 /E-ISSN: 2410-0862
DOI: https://doi.org/10.36103/4c5g0q37

Table 3. Tests for neutrality and sequence polymorphism performed on (nrDNA).

Tomato sequences	The value
No. of sequences	15
Monomorphic characters (invariable)	38
Polymorphic characters (variable)	15
Singleton variable	14
Parsimony informative	1
Haplotype number	5
Variance of Haplotype diversity	0.02387
Nucleotide diversity (π) ± SD	0.04223 ± 0.01849
Haplotype Diversity (Hd) ± SD	0.476 ± 0.155
(Pi) Jukes-Cantor (JC)	0.0459
Average no. of nucleotide differences K	2.2381
(R2)	0.1291
Raggedness statistic (r)	0.2292
Tajima's D	-2.19018 (s)** (P < 0.01)
Fu and Li's D	-2.80381(s)** (P < 0.02)
Fu and Li's F	-3.02852(s)** (P < 0.02)
Fu's Fs statistic	0.487

Moreover, the calculation of pairwise nucleotide differences (K) revealed a value of 2.23810, suggesting a relatively high polymorphism level as compared to those values reported by previous researchers. The ITS region showed a limited extent of genetic diversity but was higher than observed in Tunisian palm with K 0.035 (Maina *et al*., 2019). By comparison, Tunisian figs and *Capsicum* sp. exhibit average nucleotide differences (K) of 35.34 and 9.267, respectively, reflecting a high degree of polymorphism in their ITS regions (Baraket *et al*., 2013; Kehie *et al*., 2016).

Genetic relationships of ITS sequences

Genetic distances among ITS sequences were estimated using the Maximum Likelihood Composite (MCL) method. The distances ranged from 0.00 to 0.21, with a mean value of 0.10, indicating moderate sequence divergence. This outcome suggests the presence of a relatively high genetic diversity. No distance (0.00) was established between Nigerian tomato, Japanese tomato, Palestinian tomato, and American tomato and between Thai tomato, Italian tomato, Swiss tomato, Swedish tomato, and Bulgarian tomato, suggesting that these sequences have the highest degree of homology in this region. By contrast, the greatest genetic divergence was observed among tomato samples from South Korea, Iraq, and India, with genetic distance values of 0.21, 0.20, and 0.11, respectively. Both the "Maximum Parsimony (MP)" and "Neighbor-Joining (NJ)" approaches are used to build phylogenetic trees. The parsimony analysis generated five equally parsimonious trees, among which the shortest tree had a length of 357 steps. This tree was then evaluated, yielding a consistency index value of 0.952381, a retention index value of 0.679245, and a composite index value of 0.646900, thus signifying the existence of homoplasic characters within the analysed sequences. The Neighbor-Joining (NJ) dendrogram is shown in Fig. 3. It grouped plant tomatoes into two main groups, according to the number of mutations in their sequences. The first is composed of South Korean tomatoes and Indian tomatoes. The second is further subdivided into two groups; where the first subgroup consists solely of Iraqi tomatoes. Moreover, all remaining cultivars clustered into a second subgroup, although these groupings reflect a continuum of variability rather than discrete divisions. The tomato sequences under study have been grouped regardless of the plant's origin (country) or geographical location (continent).

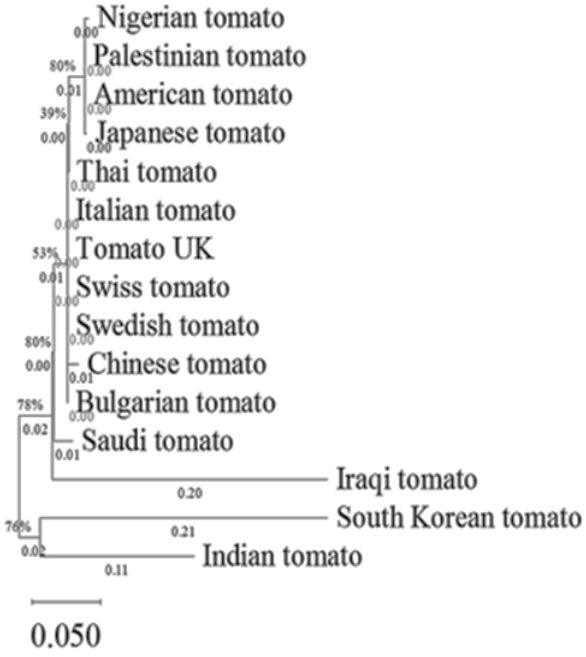

Fig. 3. Neighbour Joining tree and distance among world tomato sequences based on the ITS region of nrDNA.

Iraqi Journal of Agricultural Sciences –2026:57(3):676-688
P-ISSN: 0075- 0530 /E-ISSN: 2410-0862
DOI: https://doi.org/10.36103/4c5g0q37

Molecular evolution

Selective neutrality: The Tajima (Tajima, 1989) and Fu and Li (Fu and Li, 1993) tests are often employed to assess selective neutrality. The neutrality tests yielded significantly negative values, with Tajima's D of −2.19018 ($P < 0.01$) and Fu and Li's D (−2.80381, $P < 0.02$) and F (−3.02852, $P < 0.02$) (Table 3). The obtained values are consistent with the action of selection and suggest a recent demographic expansion in tomato inferred from the nrDNA ITS region. Additionally, an excess of rare mutations was observed in the analyzed singleton sequences. It explains the deviation from selection neutrality shown by the Tajima and Fu & Li tests. While the statistic of Fu is assessed to elucidate the reason for a departure from neutrality. This statistic is considered sufficiently powerful to detect deviations from neutrality and to test population growth and recent expansion in tomato. As shown in Table 3 and Fig. 4, this parameter exhibited a low, negative, and statistically significant value. In contrast, Fu's Fs (Fu, 1997) was positive (0.487) for the ribosomal DNA internal transcribed spacer (ITS) region. The Fu and Li neutrality tests showed significant deviations from neutrality in the analyzed regions, suggesting a recent demographic expansion in the global tomato dataset. Mismatch distribution: Analysis of the mismatch distribution of rDNA ITS sequences indicated predominantly neutral variation. The allele frequency pattern among 15 worldwide tomato sequences suggested a recent demographic expansion rather than population equilibrium (Fig. 5). On the other hand, Harpending's raggedness (Rogers and Harpending, 1992) and the Ramos-Onsins statistic (Ramos-Onsins and Rosas, 2002) (r = 0.2292 and R2 = 0.1291) have not supported these findings. Also, the low positive worth of Fu's Fs does not show that the tomato worldwide has recently had a population expansion (Table 3). Antithesis: as seen in Fig. (4), the curves of mismatch in the figure do seem to support the notion of a demographic expansion. This outcome is consistent with the results of Baraket *et al.* (2013) and Maina *et al.* (2019) in Tunisian figs and date palms, respectively.

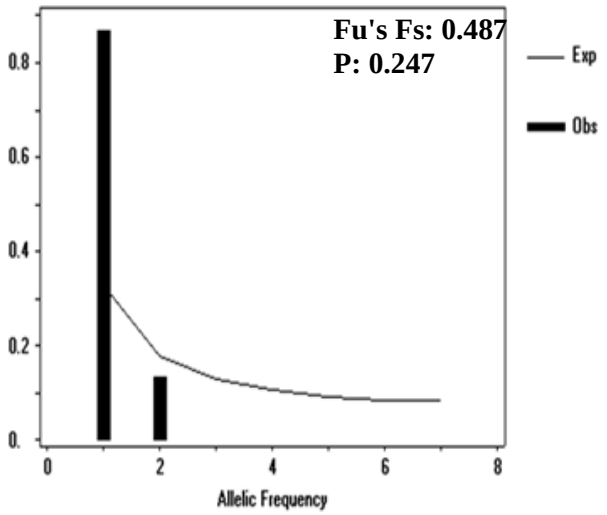

Fig. 4. The frequency spectrum of rDNA sequences discovered in tomato at ITS region. The distributions determined for neutrality and balance (mutation drift) appear in the spectrum as solid lines.

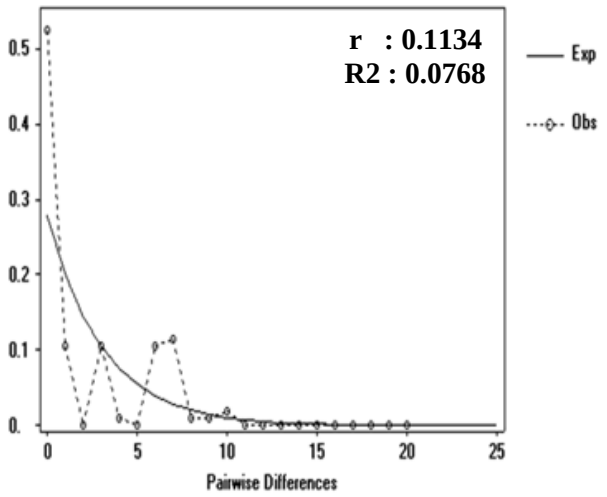

Fig. 5. An illustration to the mismatch distribution of a tomato population with an initial theta of 0.000, a final theta of 1000, and a final tau of 3.889.

Distribution of haplotypes according to ITS sequencing: A median joining network was created to identify genetic relationships between haplotypes in *S. lycopersicum*. Sequences were linked hierarchically based on mutational alterations. The haplotype network constructed from nuclear ribosomal DNA ITS sequences (Fig. 6) suggests that the tomato population originated from an ancestral haplotype and subsequently underwent demographic expansion. This is emphasised by the network structure at the founder haplotype H3 turn, which is represented by the cultivars of Nigerian tomato, Thai tomato, Palestinian tomato, American tomato, Chinese tomato, Swiss tomato, Saudi tomato, UK tomato,

Iraqi Journal of Agricultural Sciences –2026:57(3):676-688
P-ISSN: 0075- 0530 /E-ISSN: 2410-0862
DOI: https://doi.org/10.36103/4c5g0q37

Bulgarian tomato, Swedish tomato, and Italian tomato. The H3 haplotype may be the ancestor of other sequences that have changed over time. The study reveals a network of haplotypes: H1 represented by Japanese tomato, H2 represented by South Korean tomato, H4 represented by Indian tomato, and H5 represented by Iraqi tomato, with extensive branches indicating substitution numbers. Haplotype H1 (Japanese tomato) and H4 (Indian tomato) are both connected to the founder haplotype H1, represented by Nigerian tomato. Interestingly, H2 (South Korean tomato) diverges from H5 (Iraqi tomato). The large number of mutational events observed suggests that tomato has an ancient evolutionary history worldwide. The haplotype network (Fig. 6) corroborates the phylogenetic relationships depicted in Fig. 3. In the network, sequences with few mutational events appear to have diverged recently, while those with numerous mutations reflect older lineages.

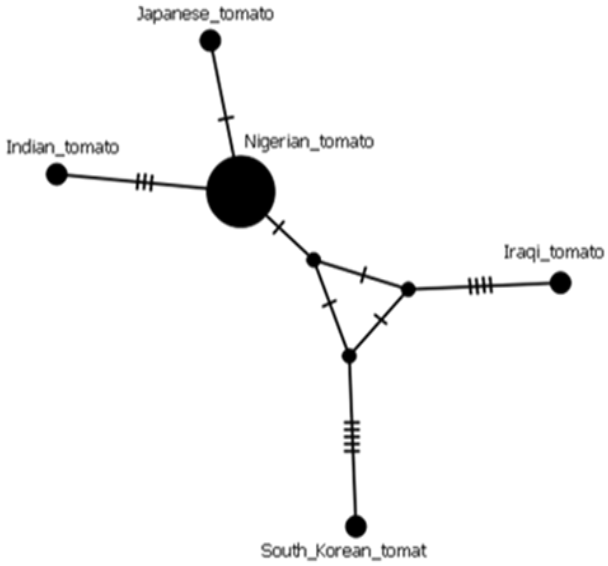

Fig. 6. The haplotype network derived from ITS sequences, the hatch marks on connecting lines indicate the number of mutations. H1: Japanese tomato. H2: South Korean tomato. H3: Nigerian tomato, Thai tomato, Palestinian tomato, American tomato, Chinese tomato, Swiss tomato, Saudi tomato, Tomato UK, Bulgarian tomato, Swedish tomato, Italian tomato. H4: Indian tomato. H5: Iraqi tomato. The circles represent distinct haplotypes, with diameters indicating sequence frequency, while inferred ancestral nodes are shown by dark, unlabelled circles.

Principal Component Analysis (PCA)

The PCA of the ITS region revealed that the first three components together accounted for 88.63% of the total genetic variance. PC1 contributed 36.067% of the variance (eigenvalue = 33.2766), PC2 contributed 29.035% (eigenvalue = 26.7885), and PC3 contributed 23.546% (eigenvalue = 21.7238). It seems that the percentage of variance is rather high when compared to the percentages derived from tomato plants by other researchers, or it is almost equal if only the first two components are considered using genetic or phenotypic traits. The majority of earlier studies reported the following values for the average variance and number of components: The first component explained 74.6% and 24.1% of the total variance (Akhter *et al*., 2021, Kıymacı *et al*., 2024); 2 components displayed the variance, registering around 77.41% and 39% of the total variance (Al-Khayri *et al*., 2023, Williams and Anbuselvam, 2023); 5 components explained 78.67% of the total variance (Vitelleschi and Pratta, 2024); 9 components contributed cumulatively to 61.076% of the total variance (Chen *et al*., 2025). Furthermore, in *Solanum lycopersicum*, the first six components collectively explained 79.59%, 78.73%, and 84% of the total variance, in accordance with studies (Binbir *et al*., 2020; Mukul *et al*., 2022; Yamaji *et al*., 2007), respectively. Similar to the previous studies, there was no evidence of grouping according to the geographical location (Fig. 7). Likewise, the PCA scatter diagram result is consistent with the distribution of haplotypes in Fig. 6 and the phylogenetic tree in Fig. 3. Principal component analysis (PCA) has shown that, despite the fact that tomatoes have been grown extensively around the world for a long time, the tomato plants have different relationships with one another. Likewise, the topology of the dendrogram and the distribution of worldwide tomatoes in the PCA analysis have demonstrated that the tomato plant germplasm is typically characterised by continuous genetic diversity.

Iraqi Journal of Agricultural Sciences –2026:57(3):676-688
P-ISSN: 0075- 0530 /E-ISSN: 2410-0862
DOI: https://doi.org/10.36103/4c5g0q37

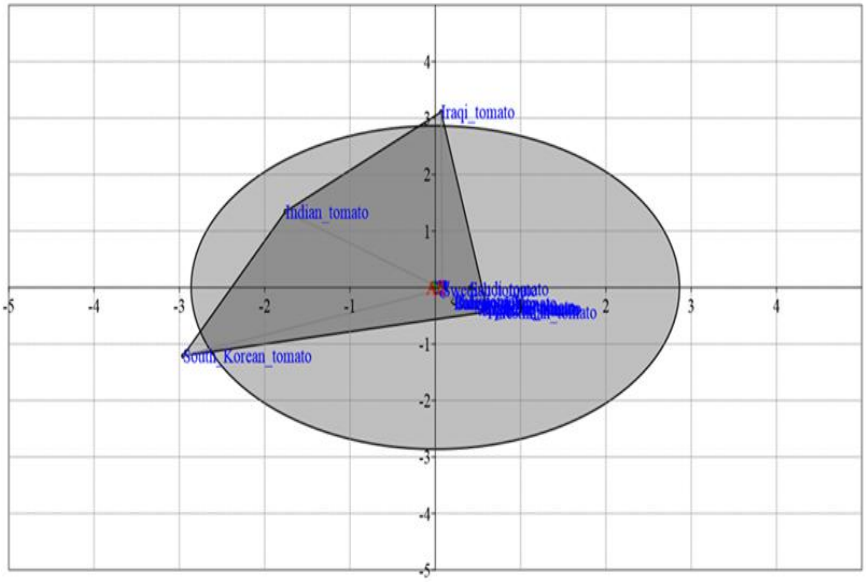

Fig. 7. Analysing the 15 worldwide tomato sequences based on ITS region using PCA

CONCLUSION

The ITS analysed in this study revealed relatively high levels of polymorphism patterns and persistent genetic variation, which is consistent with other researchers results that have employed other molecular techniques on tomato plants throughout the world. Notably, all our testing revealed that the South Korean tomato, the Iraqi tomato, and the Indian tomato, all belong to the Asian continent and have greater genetic variety than the tomatoes of other countries under study. Despite the genetic diversity, all of the conclusions derived from our investigation suggest that the tomato sequences under study are all descended from a common progenitor. This underscores the critical importance of preserving genetic diversity to safeguard species with valuable traits and to facilitate understanding of their genetic characteristics and variation, with the aim of improving and perpetuating them. Moreover, the ITS technique, particularly in plants, is a helpful tool for detecting genetic variations early on because it uses simple and quick DNA markers. Further, it is critical to collect information on each cultivar grown in Iraq using different genetic markers. As a result, evaluation of both the magnitude and distribution of genetic variation plays a crucial role in elucidating a cultivar's genetic predisposition. Finally, it is essential to conduct a more in-depth and detailed examination of the available genetic diversity.

CONFLICT OF INTEREST

The authors declare that no conflict of interest exists with respect to the publication of this manuscript.

ETHICAL APPROVAL AND ANIMAL WELFARE

The present study did not involve human subjects or the use of experimental animals; therefore, no ethical or animal welfare approval was required**.**

FUNDING

No external financial support was received for this research.

AUTHORS' DECLARATION

The authors declare that this manuscript is original, has not been published previously, and is not currently under consideration by any other journal. All figures and tables are original and prepared by the authors. Any material obtained from third parties has been included with the required permissions. All authors have read and approved the final manuscript.

AUTHORS' CONTRIBUTION STATEMENT

All authors made equal contributions to the study design, methodology, experimental work, data analysis, and manuscript writing. All authors reviewed and approved the final version of the manuscript.

REFERENCES

1.Ainouche, M.L. and R.J. Bayer, 1997. On the origins of the tetraploid *Bromus* species (section *Bromus*, Poaceae): insights from internal transcribed spacer sequences of

Iraqi Journal of Agricultural Sciences –2026:57(3):676-688
P-ISSN: 0075- 0530 /E-ISSN: 2410-0862
DOI: https://doi.org/10.36103/4c5g0q37

nuclear ribosomal DNA. Genome, 40(5): 730–743. https://doi.org/10.1139/g97-796

2.Al-Khayri, J.M., S.M. Alshamrani, A.A., Rezk, W.F. Shehata, *et al.* 2023. Pre-breeding genetic diversity assessment of tomato (*Solanum lycopersicum* L.) cultivars based on molecular, morphological and physicochemical parameters, Phyton-International Journal of Experimental Botany, 92(5): 1493-1512. https://doi.org/10.36103/ijas.v50i1.292

3.Amar, M.H., A.H.M. Hassan, and E.A.M. El Sherbeny. 2012. Assessment of genetic diversity in some wild plants of Asteraceae family by ribosomal DNA sequence. Egypt. J. Genet. Cytol. 41: 195–208. http://dx.doi.org/10.21608/ejgc.2012.10534

4.Akhter, M., F.N. Apon, M.M.R., Bhuiyan, AB., Siddique, A. Husna, and N. Zeba, 2021. Genetic variability, correlation coefficient, path coefficient and principal component analysis in tomato (*Solanum lycopersicum* L.) genotypes. Plant Cell Biotechnology and Molecular Biology, 22(25&26): 46-59. ikprress.org/index.php/PCBMB/article/view/6160

5.Ansari, S., M. Solouki, B. Fakheri, *et al.* 2018. Assessment of molecular diversity of internal transcribed spacer region in some lines and landrace of Persian clover (*Trifolium resupiantum* L.). Potravinarstvo Slovak Journal of Food Sciences, 12(1): 657-666. https://doi.org/10.5219/960

6.Baldwin, B.G., M.J., Sanderson, M.J. Porter, *et al.* 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Annals of the Missouri Botanical Garden, 82(2): 247–277. https://doi.org/10.2307/2399880

7.Bandelt, H.J., P. Forster, and A. Rohl, 1999. Median-Joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution, 16(1): 37-48. doi.org/10.1093/oxfordjournals.molbev.a026036

8.Baraket, G., A. Ben Abdelkrim, M. Mars, and A. Salhi-Hannachi, 2013. Genetic diversity and molecular evolution of the internal transcribed spacer (ITSs) of nuclear ribosomal DNA in the Tunisian fig cultivars (*Ficus carica* L.; Moracea). Biochem. Syst. Ecol. 48, 20–33. http://dx.doi.org/10.1016/j.bse.2012.11.017

9.Bauchet G. and M. Causse, 2012. Genetic Diversity in Tomato (*Solanum lycopersicum*) and Its Wild Relatives. Genetic Diversity in Plants. In Tech. Available at: http://dx.doi.org/10.5772/33073.

10.Bellarosa, R., M.C. Simeone, A. Papini, and B. Schirone, 2005. Utility of ITS sequence data for phylogenetic reconstruction of Italian *Quercus* spp. Molecular Phylogenetics and Evolution34:355–370. https://doi.org/10.1016/j.ympev.2004.10.014

11.Binbir, S., A. Kahraman, S. Mutlu, and M.A. Haytaoğlu, 2020. Genetic diversity in tomato (*Solanum lycopersicum* L.) genetic resources collected from the Aegean Region as revealed by agromorphological traits. Acta Horticulturae 1297, 23: 167-174. doi.org/10.17660/ActaHortic.2020.1297.23

12.Calonje M., S. Martin-Bravo, C. Dobeš, W. Gong, *et al.* 2008. Non-coding nuclear DNA markers in phylogenetic reconstruction. Plant Syst Evol. 282(3): 257-280 doi: 10.1007/s00606-008-0031-1

13.Chen, X., Y. Liu, F. Zheng, G. Cheng, *et al.* 2025. Construction of a core collection of tomato (*Solanum lycopersicum)* germplasm based on phenotypic traits and SNP markers. Scientia Horticulturae, 399: 113855. https://doi.org/10.1016/j.scienta.2024.113855

14.Dixon, M.T. and D.M. Hillis, 1993. Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. Molecular Biology and Evolution, 10(1):256-267. doi.org/10.1093/oxfordjournals.molbev.a039998

15.Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution, 39: 783–791. https://doi.org/10.2307/2408678.

16.Fu, YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics, 147(2): 915-925. https://doi.org/10.1093/genetics/147.2.915

17.Fu, YX. and WH. Li, 1993. Statistical tests of neutrality of mutations. Genetics, 133 (3): 693–709. https://doi.org/10.1093/genetics/133.3.693

Iraqi Journal of Agricultural Sciences –2026:57(3):676-688
P-ISSN: 0075- 0530 /E-ISSN: 2410-0862
DOI: https://doi.org/10.36103/4c5g0q37

18.Hammer, Ø.; D.A.T. Harper, and P.D. Ryan, 2001. PAST: Paleontological statistics software package for education and data analysis. Palaeontologia Electronica, 4(1): 9-18.
palaeo-electronica.org/2001_1/past/issue1_01.htm

19.Hawash, M.M. and L.M.J. Al-Shamma, 2025a. An internal transcribed spacer1 (ITS1) region to assess genetic variety and molecular evolution for the Iraqi date palm. Iraqi Journal of Biotechnology, 24(1): 86-102. jige.uobaghdad.edu.iq/index.php/IJB/article/view/799/585

20.Hawash, M.M. and L.M.J. Al-Shamma, 2025b. Analysis of sequences and molecular evolution in the Iraqi date palm cultivars (*Phoenix dactylifera* L.) propagated by tissue culture, based on ITS1 region. Iraqi Journal of Science, 66(5): 1824-1840. https://doi.org/10.24996/ijs.2025.66.5.5

21.Kehie, M., S. Kumaria, K. Sangeeta Devi, and P. Tandon, 2016. Genetic diversity and molecular evolution of Naga King Chili inferred from internal transcribed spacer sequence of nuclear ribosomal DNA. Meta Gene 7: 56–63. doi.org/10.1016%2Fj.mgene.2015.11.006

22.Kıymacı G., A.Ö. Uncu, and Ö. Türkmen, 2024. Description of the phenotypic characteristics of some tomato genotypes. Selcuk Journal of Agriculture and Food Sciences, 38(1): 9–26. 10.15316/SJAFS.2024.002

23.Kumar, S., G. Stecher, M. Suleski, M. Sanderford, *et al.* 2024. Molecular evolutionary genetics analysis version 12 for adaptive and green computing. Molecular Biology and Evolution 41(12):1-9. https://doi.org/10.1093/molbev/msae263

24.Librado, P. and J. Rozas, 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 25(11): 1451-1452. https://doi.org/10.1093/bioinformatics/btp187

25.Liston, A., W.A. Robinson, J.M. Oliphant, and E.R. Alvarez-Buylla, 1996. Length variation in the nuclear ribosomal DNA internal transcribed spacer region of non-flowering seed plants. Syst. Bot. 21(2): 109–120. https://doi.org/10.2307/2419742.

26.Maina, N., G. Baraket, A. Salhi-Hannachia and H.B. Sakkaa, 2019. Sequence analysis and molecular evolution of Tunisian date palm cultivars (*Phoenix dactylifera L.*) based on the internal transcribed spacers (ITSs) region of the nuclear ribosomal DNA. Scientia Horticulturae. 247: 373–379. doi.org/10.1016/J.SCIENTA.2018.12.045

27.Mukul, Sandhya, M. Kumar, and R.K. Agarwal, 2022. Principal component analysis based on yield and its attributing traits in tomato (*Solanum lycopersicum* L.) genotypes. pharma Innovation Journal, 11(1): 1836-1841.

28.Nei, M. and F. Tajima, 1983. Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics, 105: 207–217. https://doi.org/10.1093/genetics/105.1.207

29.Porter, C.H., and F.H. Collins, 1991. Species-diagnostic differences in the ribosomal DNA internal transcribed spacer from the sibling species Anopheles freeborni and Anopheles hermsi (Diptera: Culicidae). Am J Trop Med Hyg, 45: 271–279. https://doi.org/10.4269/ajtmh.1991.45.271

30.Ramírez-Ojeda, G., I. E. Peralta, E. Rodríguez-Guzmán, J. L. Chávez-Servia, J. Sahagún-Castellanos, and J. E. Rodríguez-Pérez, 2021. Climatic diversity and ecological descriptors of wild tomato species (*Solanum* sect. *Lycopersicon*) and closely related species (*Solanum* sect. *Juglandifolia* and *Lycopersicoides*) in Latin America. *Plants*, 10(5), 855. https://doi.org/10.3390/plants10050855

31.Ramos-Onsins, S.E. and J. Rosas, 2002. Statistical properties of new neutrality tests against population growth. Mol. Biol. Evol. 19(12): 2092–2100. doi.org/10.1093/oxfordjournals.molbev.a004034

32.Rogers, A. R., and H. Harpending, 1992. Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution, 9(3): 552–569. doi.org/10.1093/oxfordjournals.molbev.a040727

33.Sharma, S., S. Rustgi, H.S. Balyan, and P.K. Gupta, 2002. Internal transcribed spacer (ITS) sequences of ribosomal DNA of wild

Iraqi Journal of Agricultural Sciences –2026:57(3):676-688
P-ISSN: 0075- 0530 /E-ISSN: 2410-0862
DOI: https://doi.org/10.36103/4c5g0q37

barley and their comparison with ITS sequences in common wheat. Barley Genetics Newsletter, 32: 38–45.

34.Strasburg J.L., N.A. Sherman, K.M. Wright, L.C. Moyle, J.H. Willis and LH. Rieseberg, 2012. What can patterns of differentiation across plant genomes tell us about adaptation and speciation? Philos Trans R Soc Lond B Biol Sci., 5;367(1587): 364-73. https://doi.org/10.1098/rstb.2011.0199

35.Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, 123(3): 585-595 https://doi.org/10.1093/genetics/123.3.585

36.Saitou, N., and M. Nei, 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 4(4), 406–425. doi.org/10.1093/oxfordjournals.molbev.a040454

37.Van der Sande, C.A., Kwa, M., Van Nues, R.W., Van Heerikhuizen, H., Raué, H.A. and Planta, R.J. 1992. Functional analysis of internal transcribed spacer 2 of Saccharomyces cerevisiae ribosomal DNA, Journal of Molecular Biology, 223(4): 899-910. https://doi.org/10.1016/0022-2836(92)90251-e

38.Vitelleschi, M.S. and G.R. Pratta, 2024. Evaluación de la estructura genética de dos poblaciones de mejoramiento diferentes de tomate mediante Análisis de Componentes Principales. FAVE Sección Ciencias Agrarias, (23), e0024. https://doi.org/10.14409/fa.2024.23.e0024

39.Williams, G. and Y. Anbuselvam, 2023. Assessment of genetic divergence through principal component analysis and clustering in tomato germplasm accessions. Environment and Ecology, 41(4D) : 3060-3065. https://doi.org/10.60151/envec/YLSS4838

40.Xia, X., 2017. DAMBE 6: New tools for microbial genomics, phylogenetics and molecular evolution. Journal of Heredity 108(4): 431-437. https://doi.org/10.1093/jhered/esx033

41.Yamaji, H., T. Fukuda, J. Yokoyama, J. Pak, *et al.* 2007. Reticulate evolution and phylogeography in Asarum sect. Asiasarum (Aristolochiaceae) documented in internal transcribed specer sequences (ITS) of nuclear ribosomal DNA. Mol Phylogenet Evol 44:863–884. https://doi.org/10.1016/j.ympev.2007.01.011

Iraqi Journal of Agricultural Sciences –2026:57(3):676-688
P-ISSN: 0075- 0530 /E-ISSN: 2410-0862
DOI: https://doi.org/10.36103/4c5g0q37

تحليل تنوع التسلسل والتطور الجزيئي في الطماطم الدولية استناداً الى مطقة الـ ITS بالكامل

[1*]اياد وجيه الشهواني, [2]محمد مخلف هواش, [2]ثائر حامد عبد هجيج

[1*]قسم علوم الحياة, كلية العلوم, جامعة بغداد, بغداد, العراق

[2]مديرية تربية الانبار، وزارة التربية، الانبار، العراق

المستخلص

هدفت هذه الدراسة إلى تقييم التنوع الوراثي والتطور الجزيئي لمجتمعات الطماطم (*Solanum lycopersicum*) من دول مختلفة، لما لها من أهمية كبيرة في تطوير استراتيجيات التربية المستقبلية وحفظ الموارد الوراثية, و لذلك، قد تم تحليل 15 تسلسلاً وراثيًا مودعًا في قاعدة بيانات GenBank باستخدام منطقة فاصل النسخ الداخلي (ITS) بالكامل من الحمض النووي الريبوسومي (nrDNA), تراوح طول تسلسلات الفاصل بين 156 زوجًا قاعديًا في الطماطم السويدية و713 زوجًا قاعديًا في الطماطم الفلسطينية, كما لوحظ تباين واضح في محتوى GC، حيث سجلت الطماطم التايلاندية أعلى نسبة (67.48%)، بينما سجلت الطماطم الكورية الجنوبية أدنى نسبة (49.56%)وقد تم بناء الأشجار الوراثية باستخدام طريقتي ربط الجوار الأقرب (Neighbor–Joining) والحد الأدنى للتغير (Maximum Parsimony), وكشف تحليل التسلسلات عن وجود 38 موقعًا ثابتًا (أحادِي الشكل) و15 موقعًا متغيرًا، منها 14 موقعًا متغيرًا منفردًا وموقع واحد معلوماتياً (parsimony–informative), وقد أتاحت محاذاة التسلسلات التعرف على خمسة أنماط فردية (haplotypes), وبلغ معدل انحياز التحولات/التبدالات (R) نحو 17.339، مما يشير إلى أن التحولات تحدث بمعدل أكبر من التبدالات في هذه المنطقة, كما أظهرت اختبارات الحيادية، بما في ذلك اختبار تاجما (Tajima's D) واختبار فو ولي (Fu and Li)، نتائج ذات دلالة إحصائية.

الكلمات المفتاحية: محتوى الكوانين– سايتوسين، الأنماط الفردانية، عدم التطابق, تكوين النيوكليوتيدات، تحليل المكونات الأساسية, الحياد الانتقائي، الطماطم