Sequence analysis and molecular evolutionary assessment of *Eucalyptus spp*. across different environmental regions within Baghdad city based on the complete ITS region of nuclear ribosomal DNA.

Thaer Hamid A. Hajeej

Directorate of Anbar Educational, Ministry of Education ,Anbar, Iraq
Email:thaeer.hamid@gmail.com

Abstract

To explore the genetic diversity and molecular evolution of *Eucalyptus spp*., 20 reference ITS sequences from various areas were analysed. The GC content varies from 60.51% to 64.07%, with an average of 61.99%. The data revealed that transition substitutions were more common than transversions in the ITS region, with an overall transition/transversion bias (R) of 2.229. Eighteen different haplotypes were found after analysing the 20 *Eucalyptus* species DNA. The aligned ITS dataset includes 56 polymorphic sites, which included 25 singleton variable sites and 31 parsimony-informative sites. A non-geographically structured clustering pattern was identified among the overall *Eucalyptus* spp. reference sequences in the Neighbor-Joining (NJ) phylogenetic tree, indicating a lack of grouping according to geographic basis. The *E. erythrocorys* (YRM) accession had the highest degree of genetic divergence, with a genetic distance of 10.250, as confirmed by both principal component analysis (PCA) and haplotype network analysis. The first three principal components accounted for 70.748% of the overall genetic variation, showing a significant amount of the dataset's genetic structure. Furthermore, molecular evolutionary and neutrality test results demonstrate that the whole dataset supports a neutral evolutionary model in *Eucalyptus* communities, with no significant divergence from neutrality or evidence of recent demographic expansion.

Keywords. GC content, Nucleotide composition, Mismatch, Haplotypes, PCA

1. Introduction

The genus *Eucalyptus* was officially Discovered by a French botanist in 1788 and covers roughly 700 species, predominantly native to Australia. *Eucalypts* are widely grown around the world, covering an appraised 20 million hectares. The genus is known for its vast ecological amplitude, which allows it to adapt to a variety of soil types and agroclimatic

ISSN 2072-3857

conditions. *Eucalyptus* species belongs to the Myrtaceae family and is classified as an aromatic, evergreen, blooming tree that is widely planted in many nations, including Iraq [1]. The flora of Iraq is reported to have an estimated 37 species of the genus *Eucalyptus*, with a high abundance in Baghdad; however, several of these taxa are taxonomically undetermined [2].

Despite the widespread belief that natural products and herbal medicines produced from *Eucalyptus* are inherently safe [3], their use necessitates thorough scientific testing. *Eucalyptus* is commonly farmed for high-quality lumber and a variety of industrial applications due to its significant economic worth and excellent biological and technological properties [4]. Pharmacological studies have established that *Eucalyptus* possesses a broad spectrum of biological activities, including anti-inflammatory, gastrointestinal, analgesic, antioxidant, antidiabetic, antimicrobial, anticancer, antiparasitic, insecticidal, and repellent properties, in addition to important roles in oral and dental care, dermatological, and nasal treatments [5].

Toxicological studies have found that large amounts of *Eucalyptus* oil can harm key organs, including the lungs, liver, kidneys, and heart. The lungs appear to be the most susceptible organ, followed by the liver and kidneys, with the heart showing relatively modest vulnerability. All of these past use results indicate that while topical application of *Eucalyptus* oil is widely thought to be safe under the right circumstances, oral intake of the oil in high quantities is linked to obvious harmful effects [6].

Clarifying the genetic linkages across germplasm resources requires an evaluation of population structure and genetic diversity. In plant improvement initiatives, genetic variability both within and between populations is essential because it makes it easier to find and choose advantageous alleles linked to desired qualities [7]. High genetic variety populations are important resources for expanding the genetic base. A variety of molecular techniques can be used to assess genetic diversity in *Eucalyptus* species in order to efficiently characterise and make use of gene bank materials [8].

The characterisation of genetic diversity underlying many phenotypic features in organisms is made possible by genome survey sequencing, which is a quick and economical method for producing complete genomic information [9]. Moreover, it makes it easier to generate accurate and dependable molecular indicators, which are crucial for contemporary plant breeding initiatives [10]. High-throughput sequencing technologies have greatly increased the scope,

ISSN 2072-3857

depth, and speed of genomic research. Nonetheless, there is still much to learn about the complete genomic architecture of *Eucalyptus* species. The majority of genomic studies in this genus to date have mostly concentrated on a small number of economically important species, both *Eucalyptus grandis* and *camaldulensis* [11].

The internal transcribed spacer (ITS) region of nuclear ribosomal DNA (nrDNA) is commonly used as a molecular marker to measure genetic diversity and understand the evolutionary history of *Eucalyptus species* cultivars across different geographical locations. Collectively, ITS1, 5.8S, and ITS2 constitute the complete ITS region, which has become one of the most informative nuclear markers in flowering plants for molecular systematics and phylogenetic analyses [12]. A distinctive feature of nrDNA sequences, including the ITS region, is the process of concerted evolution, which promotes sequence homogenization among multiple paralogous copies within an individual genome. This molecular mechanism results in high intra-individual similarity of ITS sequences, enhancing their reliability in comparative studies. Moreover, the ITS region is more efficient for resolving relationships between closely related and recently divergent taxa due to its relatively high rate of nucleotide alteration, which gives adequate phylogenetic evidence at lower taxonomic levels [13]. Accordingly, the present study evaluated patterns of genetic structure among *Eucalyptus* species studied from different areas of Baghdad in order to determine whether any spatial variation could be associated with local environmental conditions [14].

The objective of this study was to assess genetic diversity and establish evolutionary relationships between twenty sequences of *Eucalyptus* spp.. GenBank sequences are used to reference accessions with a variety of geographic origins. By grouping accessions according to evolutionary divergence and genetic variation patterns, population structure was further assessed.

2. Materials and methods

2.1.Resource of *Eucalyptus spp.* reference sequences

Sequences previously recorded at the National Center for Biotechnology Information (NCBI), U.S. National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA, and designated under ID accession (OP696596-OP6966015; see Al-Tememmi *et al*. [1]) were downloaded, as instructed in Table 1. The ITS (ITS1, ITS2, and 5.8 gene), an internal transcription spacer region found in eukaryotic rRNA genes between the (18S) small and (28S) big subunits, was the source of all of them.

Table 1: Displayed *Eucalyptus* species, origin, accession number (ID), GC content and length.

Eucalyptus species	Region	GenBank accession number	GC conten %	Sequence Length bp
E. curtisii (ZAF)	Al Zafaraniya	OP696596.1	63.15	540
E. globoidea (ZAF)	Al Zafaraniya	OP696597.1	62.92	480
E. leucoxylon (HAR)	Al Harthiya	OP696598.1	61.67	540
E. erythrocorys (YRM)	Al Yarmouk	OP696599.1	60.93	540
E. macarthurii (BU)	Baghdad	OP696600.1	62.41	540
E. botryoides (ZAF)	University	OP696601.1	62.29	480
E. nicholii (UTF)	Al Zafaraniya	OP696602.1	62.38	540
E. pauciflora (UTF)	Al Utaifiya	OP696603.1	64.07	540
E. delegatensis (SHB)	Al Utaifiya	OP696604.1	63.56	568
E. siderophloia (HUS)	Al Shaeb	OP696605.1	60.74	540
E. alba (HAR)	Al Hussainiya	OP696606.1	61.67	540
E. camaldulensis (ZRP)	Al Harthiya	OP696607.1	62.22	540
E. vicina (KAR)	Al Zawraa Park	OP696608.1	61.60	540
E. camaldulensis (SHB)	Karrada	OP696609.1	61.30	540
E. macarthurii (RUS)	Al Shaeb	OP696610.1	62.57	545
E. tereticornis (BIP)	Rustamiya	OP696611.1	61.34	551
E. siderophloia (JAD)	Baghdad-Island	OP696612.1	60.51	552
E. vicina (RUS)	Park	OP696613.1	61.41	540
E. alba (ZAF)	Al Jadriya	OP696614.1	61.67	540
E. tereticornis (JAD)	Rustamiya	OP696615.1	61.55	541
	Al Zafaraniya			
	Al Jadriya			

2.2.Sequence analysis

ISSN 2072-3857

Reference sequences previously deposited in the NCBI GenBank were retrieved, comprising 20 accessions of *Eucalyptus* spp. collected from various regions of Baghdad and representing the complete ITS region, as listed in Table 1. Sequence alignment and subsequent molecular analyses were performed using MEGA software version 12.0.7 [15]. In addition, the aligned datasets were analyzed using DnaSP version 5.10.01 [16] to evaluate nucleotide polymorphism and infer historical demographic dynamics. To assess genetic variation among cultivars based on ITS reference sequences, indices of haplotype diversity (Hd) [17] and nucleotide diversity (π) [18], together with their corresponding standard deviations (± SD), were calculated. Demographic parameters were further evaluated using DnaSP version 5.10.01 [13], based on the site-frequency spectrum approach of Tajima [19] and the mismatch distribution analysis of Rogers and Harpending [20]. Selective neutrality was assessed using Fu and Li's D* and F* statistical tests [21]. On the other hands, the GC content was evaluated across the entire ITS reference sequences was assessed using DAMBE version 6 [22]. In addition, estimates were established for the minimum number of recombination events (Rm) and average number of pairwise nucleotide differences (K). Graphical representations for every reference sequence were created using the online tool provided by Biologics Corp. https://www.com/tools/GCcontent/.

Sequence divergence between *Eucalyptus* species, pairwise. To estimate ITS reference sequences, the Maximum Composite Likelihood (MCL) technique was applied. Using MEGA version 12 and the Neighbor-Joining (NJ) method [23], phylogenetic relationships were rebuilt. Tree robustness was estimated through bootstrap analysis with 1,000 replicates [24], and the consistency (CI), retention (RI), and composite indices were calculated to assess tree reliability. The transition/transversion (ti/tv) ratio was calculated using the equation $R = \frac{A*G*k1+T*C*k2}{(A+G)(T+C)}$, where A, G, C, and T denote nucleotide composition. The PAST program was applied to construct the Principal Component Analysis (PCA) scatter plot, which included eigenvalues and variance estimates [25]. NETWORK also served to depict haplotype-based genetic relationships [26].

3. Results and Discussions

3.1. Analysis of genetic diversity

3.1.1. Sequence length of ITS region and nucleotide composition

The reference sequences displayed variation in both sequence length and

nucleotide composition across all studied *Eucalyptus* species. In particular, the entire ITS region diverse from 568 bp in *E. delegatensis* (SHB) to 480 bp in *E. botryoides* (ZAF) and *E. globoidea* (ZAF) (Table 2). The internal transcribed spacer region diverse in length, with an average size of roughly 634 (bp) across the majority of Australian *Eucalyptus* sequence studied, according to Steane et al. [27]. The 5S rDNA region in *Eucalyptus* spp. exhibitions length difference ranging from 310 to 415 (bp) [28]. An ITS length of 636 bp was detected in some *Eucalyptus* spp. and nitrogen base composition (A: 20.91%, T: 17.02%, C: 32.24%, G: 29.84%) demonstrations no considerable deviation from the values described by Steane *et al*. [29] for *Eucalyptus* species (A: 19%, T: 20%, C: 40%, G: 21%), indicating general consistency in entire the ITS nucleotide composition patterns between international and Iraqi datasets (Table 1).

3.1.2 **Guanine Cytosine content (GC)**

Indeed, from 60.51% in *E. siderophloia* (JAD) to 64.07% in *E. pauciflora* (UTF), the average GC content of the entire ITS reference sequence has fluctuated between 61.99% (Table 1 & Figures 1 and 2).In addition, in contrast to Bayly *et al*. [30] and Gibbs *et al*. [31], where divergent sequences were characterized by reduced GC content, lower structural stability, and increased methylation-associated C→T substitutions relative to typical sequences, the present local *Eucalyptus* dataset exhibited higher GC content compared with external references, indicating a different genomic pattern with reduced evidence of methylation related sequence degradation. In this context, that methylation-associated processes contribute to sequence degradation through elevated substitution rates and mutation accumulation. In Iraqi date palm, Hawash and Al-Shamma [32, 33, 34] reported GC content ranging from 52% to 53% using ITS1 molecular markers. This observation suggests the possibility of distinct evolutionary origins of the ITS region [12]; however, a co-evolutionary mechanism may also be proposed to explain the observed pattern.

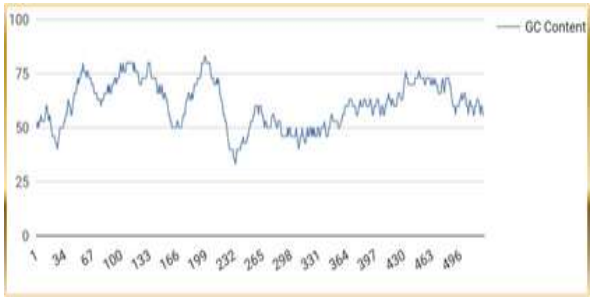

Figure 1: GC content variation in *E. siderophloia* (JAD), showing a GC percentage of 60.51%.

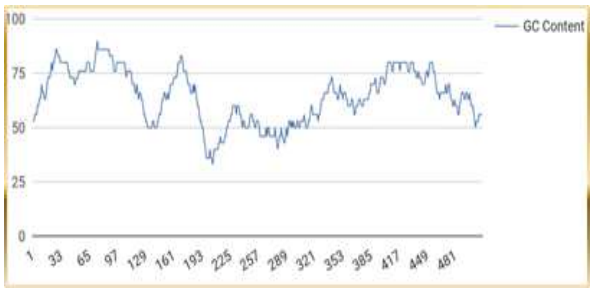

Figure 2: GC content variation in *E. pauciflora* (UTF), showing a GC percentage of 64.07%

Remarkably, the limited fluctuation in GC content across ITS sequences suggests a high degree of homogeneity between different reference sequences., despite their occurrence in geographically and environmentally diverse regions. This is consistent with the conserved nature of ribosomal genes in eukaryotes [35]. The observed pattern further suggests that substitution processes may be biased, thereby increasing the difficulty of detecting homoplasy in these reference sequences compared with those evolving under equal substitution probabilities.

3.1.3 Nucleotide variance, mutational event

An examination of the complete ITS reference sequences revealed that purine bases had a transition/transversion (Ti/Tv) ratio of K1 = 2.562, whereas pyrimidine bases had K2 = 7.241. The transition/transversion (Ti/Tv) ratio (R) for all bases was calculated using MEGA version 12.0.7 [15] to be 2.229.The result could indicate that reference sequences with lower Ti/Tv ratios (R) have evolved more extensively than those with higher ratios. However, lower transition/transversion ratios were observed in the analyzed samples compared with Iraqi date palm propagated by tissue culture (R = 0.751), offshoots (R = 0.661), and combined propagation methods (R = 0.800) based on the ITS1 region [32, 33, 34]. In contrast, a markedly higher transition/transversion ratio was reported in international tomato by Al-Shahwany *et al*. [36], reaching 7.339, indicating a substantially greater frequency of transitions than transversions in this region. The present study demonstrates that, within the ITS intergenic spacer region of *Eucalyptus* spp. reference sequences from selected Baghdad regions, transitions occurred more commonly than transversions. As revealed in Table 2, the C↔T transitions occurred more frequently than the G↔A transitions, suggesting a transition-biased substitution pattern across the sequences studied.

Table 2: Calculated the relative frequency of nucleotide variations across the entire ITS region.

	A	T	C	G
A	-	*2.48*	*4.69*	**11.14**
T	*3.04*	-	**33.99**	*4.35*
C	*3.04*	**17.94**	-	*4.35*
G	**7.8**	*2.48*	*4.69*	-

ISSN 2072-3857

Eucalyptus sequences Values	
Number of sequences	20
Invariable (monomorphic) sites	392
Variable (polymorphic) sites	56
Singleton variable sites	25
Parsimony informative sites	31
Number of Haplotypes (H)	18
Variance of Haplotype diversity	0.00037
Pi±SD	0.03308±0.00297
Haplotype diversity (Hd)±SD	0.989±0.019
Jukes and Cantor Pi(JC)	0.03400
K	14.82105
Minimum number of recombination	7
Ramos-Onsins and Rozas (R2)	0.1125
Raggedness statistic (r)	0.0206
Probability	0.007
Tajima's D	-0.55949, (NO S.), P > 0.10
Fu and Li's D	-0.85668, (NO S.), P > 0.10
Fu and Li's F	-0.89488, (NO S.), P > 0.10
Fu's Fs statistic	-4.801

Note 1*. With rows denoting the nitrogen base that is being created and columns denoting the nitrogen base that has been substituted, each lockup displays the estimated substitution probability (r) between nucleotides. Transversional substitutions are revealed in italicised designs, while transitional substitutes are bolded.

The nucleotide sequence alignment of the entire ITS region exhibited considerable genetic difference, including 448 bp without gaps and missing data. Of these, 392 sites were conserved and 61 were variable (Table 3). The variable sites included 25 singleton positions (71, 77, 78, 109, 111, 120, 123, 138, 168, 229, 245, 258, 277, 296, 321, 336, 352, 355, 369, 386, 426, 427, 454, 461 and 484) and 31 parsimony-informative sites (including positions 57, 6,2 86, 87, 97, 104, 107, 113, 114, 117, 130, 132, 133, 139, 147, 183, 253, 382, 387, 404, 433, 436, 441, 442, 449, 451, 455, 468, 470, 471 and 481). Indicating a moderate level of sequence polymorphism within the analyzed dataset.

Table 3: The tests for neutrality and sequence polymorphism are based on the entire ITS region of rDNA of *Eucalyptus* species.

Moreover, the ITS region of ribosomal DNA exhibited substantial sequence diversity. Actually, 18 pinpoint haplotypes were identified among the 20 reference sequences.

ISSN 2072-3857

The assessed nucleotide diversity (Pi±SD) and haplotype diversity (Hd±SD) plus their standard deviation for all ITS sequences were 0.03308±0.00297 and 0.989±0.019, respectively. Hawash and Al-Shamma [33] identified 8 haplotypes among 9 Iraqi date palm sequences based on the ITS1 region, with nucleotide diversity (Pi±SD) and haplotype diversity (Hd±SD) values estimated at 0.0061 ± 0.0023 and 0.972 ± 0.064, respectively. In a separate study, Hawash and Al-Shamma [34] reported complete haplotypic differentiation among 10 Iraqi date palm sequences, with corresponding Pi±SD and Hd±SD values of 0.00614 ± 0.00066 and 1.000 ± 0.045, respectively. Additionally, point mutations, translocation mechanisms, and insertion/deletion (indel) events may be responsible for sequence diversity within the analysed ITS region. The positions of gaps across most indel regions were readily resolved owing to the conserved nature of adjacent sequence segments. In the other hands, point alterations and translocations likely reveal underlying evolutionary mechanisms and contribute to the genetic differentiation detected among the analysed reference sequences.

An analysis of the entire ITS region of nr DNA showed an elevated degree of genetic variety among the *Eucalyptus spp.* accessions studied. The accumulation of nucleotide changes over the ITS region is indicated by the average number of pairwise nucleotide differences (K = 14.82105), which indicates a considerable degree of polymorphism. This pattern suggests substantial genetic variance among the sequence of studied species (Table 3). Overall, the results display a high degree of ITS variability, designating minimal sequence conservation within the dataset. This variability may be a result of tandem repeat dynamics, that increase ITS sequence divergence by permitting non-homologous recombination and unequal crossing-over events. Comparative evidence shows greater K values reported in Tunisian *Ficus carica* (35.34) and *Capsicum spp*. (9.267), reflecting pronounced ITS polymorphism in those accessions' sequences studied [13, 37]. In contrast, lower levels of genetic diversity have been reported in Tunisian and Iraqi date palms, with K values of 0.686, 3.733, 3.889, and 3.911, as documented by Maina *et al*. [38] and Hawash and Al-Shamma [32, 33, 34], respectively.

3.2 Genetic relations of entire ITS reference sequences

The genetic distances among *Eucalyptus* spp. ITS sequences ranged from near-zero values (0.000–0.500) (Figure 3), indicating close genetic relatedness among certain intraspecific or closely related accessions, to a maximum value of

10.250 observed in *Eucalyptus* erythrocorys (YRM), which represented the most genetically divergent lineage within the dataset. This wide range of genetic distances reflects heterogeneous evolutionary divergence among the analyzed taxa, with both highly conserved and strongly differentiated lineages present, indicating that these sequences accession have the highest degree of homology. The reference sequences that are being examined show that the four most parsimonious trees, each with 92 steps, exist. The homoplasic properties can be observed by the consistency index (CI) of 0.762712, the retention index (RI) of 0.909677, and the composite index (CI) of 0.693822. Steane *et al.* [27], based on an analysis of 54 samples, reported that one of the 144 most parsimonious trees comprised 291 steps, with a consistency index (CI) of 0.551 and a retention index (RI) of 0.821. The strict consensus tree of nine equally parsimonious trees (tree length = 553; CI = 0.605; RI = 0.851) was obtained from the maximum parsimony analysis of the combined dataset (ETS + ITS) [39]. In addition, studies on the *Eucalyptus* subgenus based on combined nrDNA regions [31] identified 72 most parsimonious trees, each with a tree length of 161 steps, a consistency index (CI) of 0.77, and a retention index (RI) of 0.76.

ISSN 2072-3857

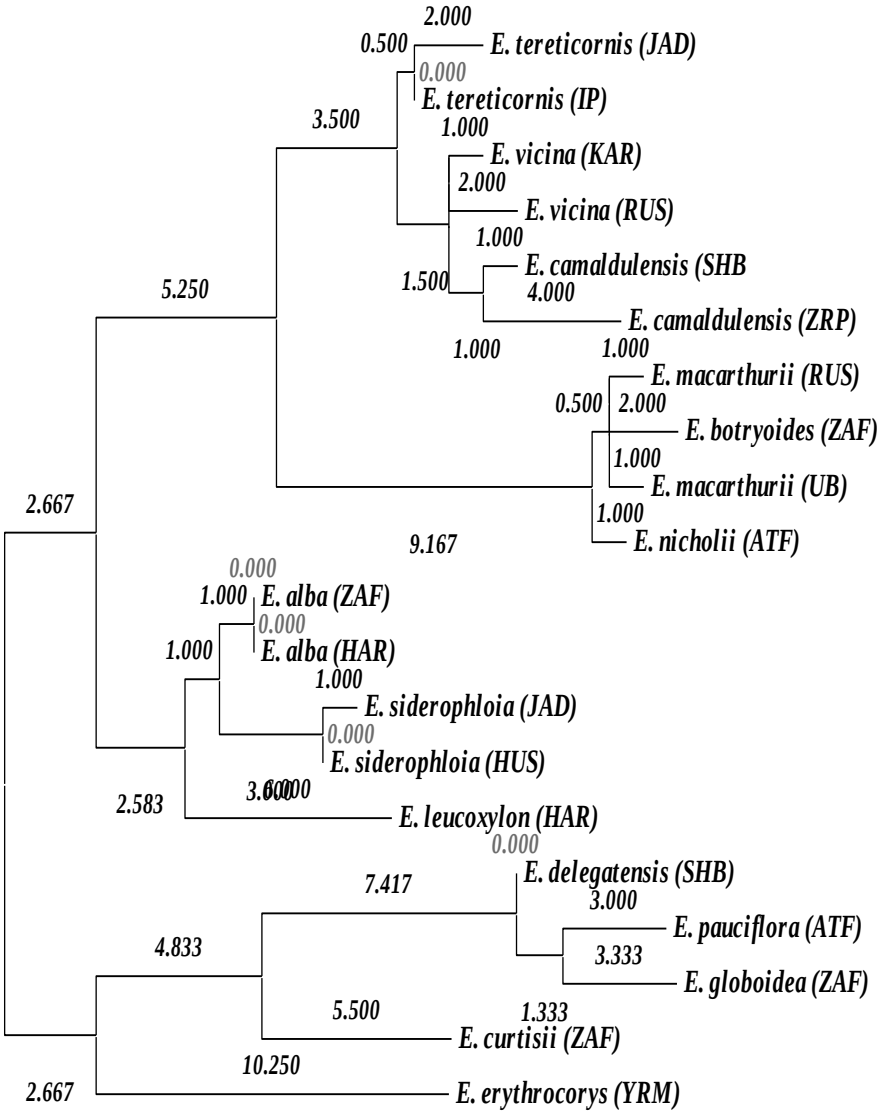

Figure 3: The neighbour-joining tree based on the entire ITS region of the nrDNA from various locations in Baghdad city explains the relationships and distances between various *Eucalyptus* species.

The Neighbor-Joining (NJ) dendrogram of *Eucalyptus* species based on ITS sequences indicated a hierarchical clustering structure, with major groups further fragmented into separate subclades, as shown in the phylogenetic tree (Figure 3). The upper main clade was initially divided into two well-defined subclusters: the first included *E. tereticornis* [JAD, IP], *E. vicina* [KAR, RUS], *and E. camaldulensis* [SHB, ZRP], while the second subcluster comprised *E. macarthurii* [RUS, UB], *E. botryoides* [ZAF], and *E. nicholii* [ATF]. These subclades further discovered internal bifurcation, reflecting fine-scale genetic variety among closely related taxa. Equally, the second major clade

also displayed a clear branching construction, splitting into two principal subgroups. One subgroup included *E. alba* [ZAF, HAR], *E. siderophloia* [JAD, HUS] and *E. leucorylon* [HAR], whereas the second subgroup consisted of *E. delegatensis* [SHB], *E. pauciflora* [ATF], *E. globodea* [ZAF], *E. curtisii* [ZAF] and *E. erythrocorys* [YRM]. The *E. erythrocorys* (YRM) revealed the highest genetic deviation, occupying a further isolated branch site within the dendrogram. However, the observed hierarchical clustering and sub-clustering patterns weren't associated with the geographic origin of the accessions from various regions of Baghdad, indicating a lack of apparent spatial or biogeographic structuring. Instead, the branching topology indicates that evolutionary divergence, rather than geographic separation, established the genetic associations between the sequences studied. This pattern may suggest differences in ITS sequence composition among the *Eucalyptus* accessions investigated, resulting in varying levels of genetic similarity and divergence. Clustering multiple taxa within the same clade suggests stronger genetic links, but separating others into different branches indicates greater genetic diversity.

3.3 Molecular evolution

3.3.1 The recombination occurrence

Eucalyptus' recombination pattern offers valuable insight into its intraspecific genetic structure by shaping DNA polymorphism across the nr ITS region. The present collection contains the complete reference sequences of *Eucalyptus* species and unveiled a total of seven recombination events. These events occurred at the following frequency intervals: (87-97), (97-117), (130-132), (147-183), (183-433), (433-449), and (449-471). This level of recombination is relatively moderate when compared with other plants. For instance, it is bigger than the pattern seen in Tunisian palm trees, which showed only one recombination event [38], but it is far lower than that described in Tunisian fig populations, which showed 20 recombination events [13]. Overall, these results suggest that *Eucalyptus* maintains a relatively conserved recombination landscape, which may have been responsible for the structured, intraspecific variation and genetic stability patterns exhibited in the ITS region.

3.3.2 Test selective neutrality

Several neutrality tests, including Tajima's D [19] and Fu and Li's D* and F* statistics [21], were used to assess whether the observed patterns

of genetic diversity in the nuclear ribosomal DNA ITS reference sequences of the *Eucalyptus* species differ from a neutral equilibrium model. Applying these techniques, the null hypothesis that the mutations acquired are selectively neutral was tested. The results showed consistently negative but non-significant values for Tajima's D across all *Eucalyptus* reference sequences ($D = -0.55949$; $P > 0.10$), indicating an excess of low-frequency polymorphisms; however, this deviation was not sufficient to reject the neutrality hypothesis. The neutrality was further supported by Fu and Li's tests over the entire ITS region, which likewise provided negative and non-significant results ($D^* = -0.85668$; $P > 0.10$ and $F^* = -0.89488$; $P > 0.10$). Further, Fu's Fs [40] statistic produced a relatively high negative value (Fu Fs = −4.801) (Table 3). Although the deviations from neutrality were not statistically significant, the neutrality indices revealed a minor tendency toward demographic expansion or an excess of rare haplotypes. However, given the absence of statistical significance, this pattern should be understood cautiously as a weak and non-definitive indication of population expansion. In general, the combined neutrality tests suggest that the ITS region of *Eucalyptus* species reference sequences is mainly consistent (equilibrium) with a neutral model of evolution, where mutation and genetic drift are the predominant forces affecting genetic variation, although the negative trend in the statistics may reflect subtle historical demographic procedures.

3.3.3 Mismatch distribution

These outcomes are further supported by Harpending's raggedness test and the Ramos-Onsins and Rozas (R2) test in the *Eucalyptus* reference sequences. The acquired values ($r = 0.0206$; $P = 0.149$; $R2 = 0.1125$ see Table 3) indicate a smooth and non-significant mismatch distribution pattern, suggesting no deviation from a demographic equilibrium model. The non-significant R2 statistic and the low Harpending's raggedness index value indicate a steady population structure devoid of signs of recent demographic expansion or abrupt population growth. These outcomes are consistent with a model of constant population size, where the observed genetic deviation is primarily formed by long-term mutation–drift equilibrium rather than recent demographic expansion. In general, *Eucalyptus* species exhibit a mismatch distribution pattern. Neutrality-based interpretations are supported by ITS sequences, which also show that there hasn't been a recent expansion event in the population.

In Figures 4 and 5, the mismatch distribution analysis of ITS rDNA sequences in *Eucalyptus* spp. revealed a trend essentially compatible with demographic

equilibrium, with no evident evidence of recent population growth throughout the examined accessions from various regions of Baghdad. This inference is further supported by neutrality test statistics, including (Tajima's D) and (Fu and Li's D and F), all of which showed non-significant deviations from neutral expectations, indicating an absence of strong signals of selection or demographic expansion. Overall, these findings indicate that observable genetic variation is mostly shaped by mutation-drift equilibrium rather than recent demographic or selection processes. Overall, the allele frequency pattern suggests a steady evolutionary history, with only slight and insignificant aberrations that cannot be used to infer recent demographic expansion.

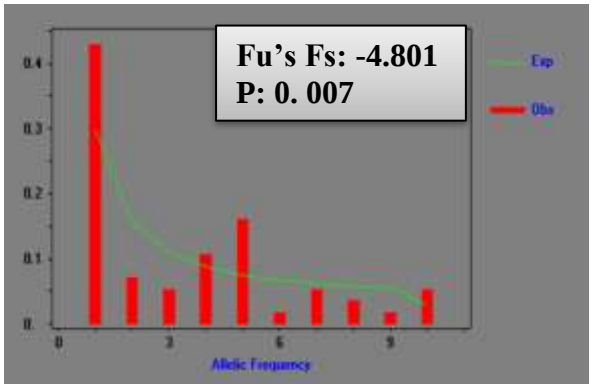

Figure 4: The frequency spectrum of *Eucalyptus* species ITS rDNA sequences from various regions of Baghdad. The apparent distribution of sequence diversity within the ITS region is presented, with solid lines reflecting the expected distribution under a neutral evolution with mutation-drift equilibrium.

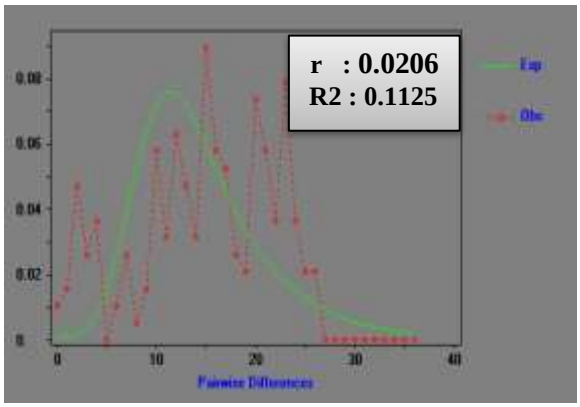

Figure 5: Mismatch distribution of the *Eucalyptus* population, depicting observed and predicted values under a model of population size fluctuation. The analysis used an initial theta (θ_0 = 5.699) and a final tau (τ = 1000.000) to evaluate past demographics and future demographic expansion tendencies.

3.3.4 Haplotypes distribution

The haplotype network presented in Figure 6 showed significant genetic variety among the detected haplotypes. *E. alba* (HAR) and *E. alba (ZAF),* which represent Hap3, are located semi-centrally in the network, indicating that they may play intermediate or ancestral haplotype roles in population diversification. Several haplotypes, including Hap1, which is represented *by E. tereticornis* (JAD), Hap4, which is represented by *E. vicina* (KAR), Hap5, which is represented by *E. vicina* (RUS), and Hap9, which

is represented by *E. camaldulensis* (SHB), differed from the founder haplotype by only one to three mutational events, indicating close genetic relationships and recent divergence events. With roughly 10 mutational events, Hap15, which is represented by *E. erythrocorys* (YRM), revealed the largest mutational distance from the founder haplotype, suggesting substantial genetic differentiation and a possibly unique evolutionary history. Likewise, the mutational distances to the center haplotype cluster of Hap12, represented by *E. pauciflora* (UTF), Hap16, represented by *E. leucoxylon* (HAR), and Hap18, represented by *E. curtisii* (ZAF), are notably considerable. The combination of hypothesised or unsampled intermediary haplotypes that might have influenced the evolutionary relationships among the tested accessions is further suggested by the prevalence of median vectors within the network. All things reflected, the network construction encourages extensive genetic structuring and gradual evolutionary diversification within the studied *Eucalyptus* population, enabling the coexistence of highly divergent and closely related haplotypes. Although Hap16 showed several mutational steps within the haplotype network, Hap15 appeared to be the most divergent branch in the phylogenetic tree (Figure 3), most likely due to its isolated evolutionary position and accumulation of unique nucleotide changes over the analysed ITS region.

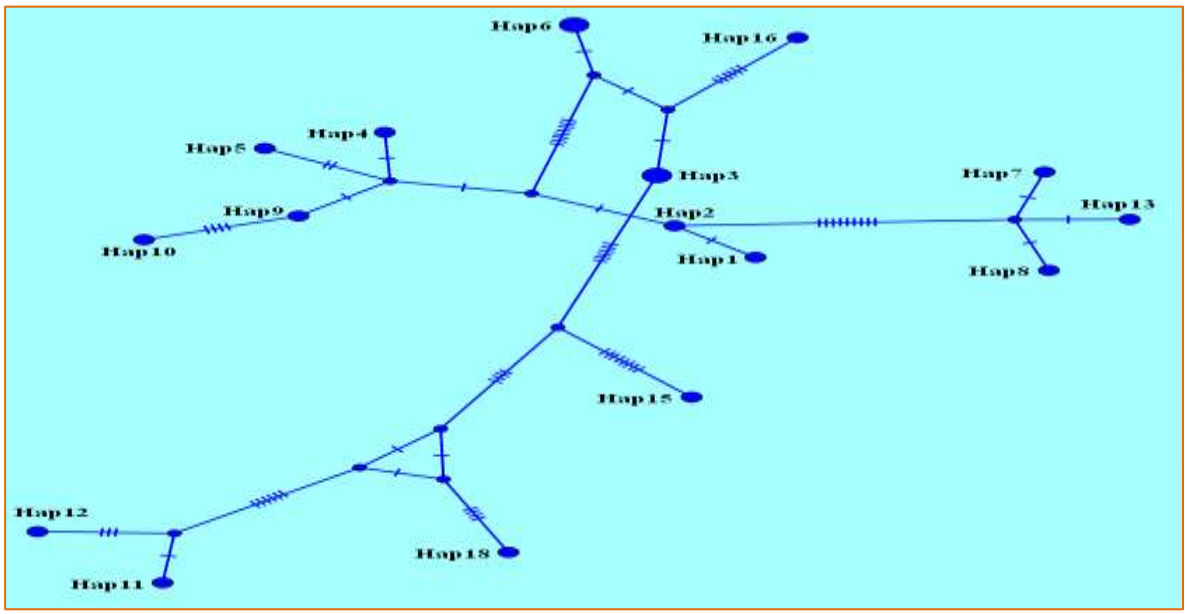

Figure 6: Haplotype network derived from *Eucalyptus* species' ITS

reference sequences. The number of mutational steps between haplotypes is shown by hatch marks on the connecting lines. Small, unlabelled blue circles indicate inferred ancestral or intermediate haplotypes, while blue circles, whose size is proportionate to haplotype frequency, reflect observed haplotypes. Hap 1 (*E. tereticornis*, JAD), Hap 2 (*E. tereticornis*, BIP), Hap 3 (*E. alba*, HAR and ZAF), Hap 4 (*E. vicina*, KAR), Hap 5 (*E. vicina*, RUS), Hap 6 (*E. siderophloia*, JAD and HUS), Hap 7 (*E. macarthurii*, RUS), Hap 8 (*E. macarthurii*, BU), Hap 9 (*E. camaldulensis*, SHB), Hap 10 (*E. camaldulensis*, ZRP), Hap 11 (*E. delegatensis*, SHB), Hap 12 (*E. pauciflora*, UTF), Hap 13 (*E. nicholii*, UTF), Hap 14 (*E. botryoides*, ZAF), Hap 15 (*E. erythrocorys*, YRM), Hap 16 (*E. leucoxylon*, HAR), Hap 17 (*E. globoidea*, ZAF), and Hap 18 (*E. curtisii*, ZAF).

3.4 The principal component analysis

The first principal component (PC1), with an eigenvalue of 9.46904, explained 33.978% of the overall variation in ITS region sequences, according to Principal Component Analysis (PCA). With eigenvalues of 6.31419 and 3.93312, respectively, the second (PC2) and third (PC3) components accounted for 22.657% and 14.113% of the variation. 70.748% of the overall genetic variation across the examined Iraqi *Eucalyptus spp*. accessions was explained by the first three main components taken together. Different levels of genetic divergence among the examined taxa were indicated by the PCA scatter plot, which displayed a diverse distribution of accessions with numerous genotypes placed apart from the major clusters (Figure 7). Overall, the PCA results support the presence of moderate to high genetic variability and are consistent with the phylogenetic and haplotype network analyses. Nasr *et al.* [41] reported that the first five principal components (PCs) explained 43.38% of total variation in immature markers (IMs) and 45.52% in mature markers (MGs). Moreover, PC1 and PC2 accounted for 30.4% and 12.0% under adequate P and 47.4% and 13.7% under low P circumstances, respectively, under various phosphorus treatments [42]. Furthermore, it was demonstrated that PCA-derived latent variables enhanced ANN prediction accuracy by explaining 83% of the total variance [43]. Additionally, it has been revealed that PC1, PC2, and PC3 together account for 92% of the total variation, explaining 48.15%, 38.09%, and 5.75% of total variability, respectively [44].

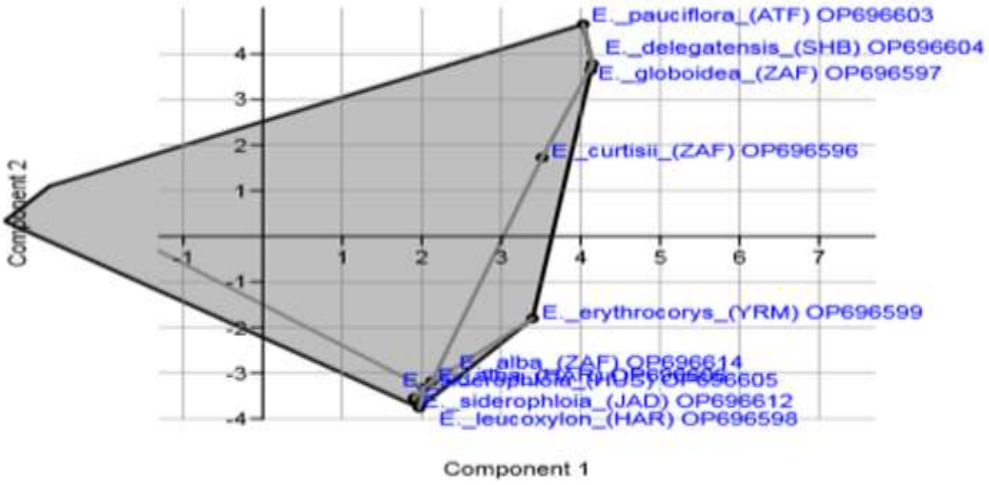

Figure 7: The relative distribution of 20 *Eucalyptus* reference sequences based on the entirety of the ITS region appears in a scatter plot diagram for principal component analysis.

4. Conclusion

The extensive molecular and multivariate studies revealed that the *Eucalyptus* accessions under study have a significant degree of genetic variability. Analyses of population structure, haplotype networks, and phylogenetic reconstruction consistently showed a generally unstructured genetic pattern with no discernible genetic clumping based on geographic origin. This conclusion is consistent with dendrogram, haplotype, and population structure analyses, all of which indicated no clear grouping based on geographic origin. A alike pattern has been reported in earlier studies on *Eucalyptus*, where genetic or phenotypic modification was distributed without strong spatial structuring. This pattern was more supported by nucleotide frequency and genetic distance analyses, which are long-established measurable but widely distributed sequence variants among accessions. Deviations from neutrality were revealed by neutrality tests, such as Tajima's D and Fu and Li's statistics, suggesting possible historical demographic processes like expansion or selection. On the other hand, Haplotype analysis supported genetic heterogeneity by revealing several unique haplotypes with modest differentiation. Furthermore, these outcomes were supported by PCA, which clarified a significant amount of overall genetic variety in a small number of components. The vast and non-geographically structured genetic variety that is often highlighted by the agreement between all analytical approaches reflects a complex evolutionary background that may be relevant for future breeding, conservation, and genetic enhancement activities. These results reveal an understanding of how genetic resources are organised within the Baghdad *Eucalyptus*

species and create a useful framework for future breeding initiatives and conservation. This finding may encourage more efficient use of *Eucalyptus* species genetic resources and help ongoing efforts in genetic improvement.

Acknowledgements

I am incredibly grateful to Noor J. Al-Tememmi, Neamat J. Al-Judy, and Labeeb A. Al-Zubaidi for making their sequencing data publicly accessible via GenBank, which has made it possible to conduct additional research and analysis.

References

1- Al-tememmi, N. J., Al-judy, N. J., & Al-zubaidi, L. A. (2024). Genetic Diversity and Polymorphism of *Eucalyptus* L' Hėr Species by ITS Sequence. Iraqi Journal of Science, 65(6), 3040–3046.

2- Lobab G.A., & Al.Musawi, A.H.A. (2025). Thirty. three new record dicot species for Baghdad city and four new records to Iraqi flora at Iraq. The Iraqi Journal of Agricultural Sciences, 56(6), 2109–2121.

3- Salvatori, E.S., Morgan, L.V., Ferrarini, S., Zilli, G.A.L., Rosina, A., Almeida, M.O.P., Hackbart, H.C.S., Rezende, R.S., Albeny-Simões, D., Oliveira, J.V., Gasparetto, A., Müller L.G., & Dal Magro, J. (2023). Anti-Inflammatory and Antimicrobial Effects of *Eucalyptus* spp. Essential Oils: A Potential Valuable Use for an Industry Byproduct. Evidence-based complementary and alternative medicine. eCAM, 2582698.

4- Kaur A., & Monga, R. (2017). *Eucalyptus* Trees Plantation: A Review on Suitability and their Beneficial Role. IJBSM, 12(1) pp. 016–025.

5- Al-Snafi A.E. (2017). The pharmacological and therapeutic importance of *Eucalyptus* species grown in Iraq. IOSR Journal of Pharmacy, 7(3), 72-91.

6- Muhsan, Z., & Alsodany, A. (2022). Histological Study for Median lethal Dose (LD50) of *Eucalyptus* Oil Administrated Orally in (Mice mus musculus). Bionatura. 7(2), 1-7.

7- Aesomnuk, W., Ruengphayak, S., Ruanjaichon, V., Sreewongchai, T., Malumpong, C., Vanavichit, A., Toojinda, T., Wanchana S., & Arikit S. (2021). Estimation of the Genetic Diversity and Population Structure of Thailand's Rice Landraces Using SNP Markers. Agronomy, 11(5), 995.

8- Wang, C., Lan, J., Wang, J., He, W., Lu, W., Lin, Y., & Luo, J. (2023). Population structure and genetic diversity in *Eucalyptus pellita* based on SNP markers. Front. Plant Sci. 14, 1278427.

9- Dasgupta, M.G., Dharanishanthi, V., Agarwal, I., & Krutovsky, K.V. (2015). Development of genetic markers in *Eucalyptus* species by target enrichment and exome

sequencing. PloS one, 10(1), e0116528.

10- Temas, A. (2017). Development of molecular markers for the *Eucalyptus* species identification. Temas Agrarios, 22(2), 32-41.

11- de Oliveira, D.A., da Silva, P.H.M., Novaes, E., & Grattapaglia, D. (2023). Genome-wide analysis highlights genetic admixture in exotic germplasm resources of *Eucalyptus* and unexpected ancestral genomic composition of interspecific hybrids. PloS one, 18(8) e0289536.

12- Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S., & Donoghue, M.J. (1995). The ITS of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Annals of the Missouri Botanical Garden, 822, 247–277.

13- Ghada, B., Ben Abdelkrim, A., Mars, M., & 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). Biochemical Systematics and Ecology, 48, 20–33.

14- Teixeira, G.C., Konzen, E.R., Faria, J.C.T., Gonçalves, D.S., de Carvalho, D., & Brondani, G.E. (2020). Diversidade genética de duas espécies de *Eucalyptus* usando marcadores ISSR. Ciênc. Florest., 30(1), 270–278.

15- Kumar, S., Stecher, G., Suleski, M., Sanderford, M., Sharma, S., & Tamura, K. (2024). Molecular evolutionary genetics analysis version 12 for adaptive and green computing. Molecular Biology and Evolution, 41(12), 1-9.

16- Librado, P. & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 25, 1451-1452.

17- Nei, M., & Tajima, F. (1983). Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics, 105, 207–217.

18- Jukes, T.H., & Cantor, C.R. (1969). Evolution of protein molecules. In: Munroled, H.N. (Ed.), "Mammalian Protein Metabolism". Academy Press, New York, 31–132.

19- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, 123, 585–595.

20- Rogers, A.R., & Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol., 9, 552–569.

21- Fu, Y.X., &. Li, W.H. (1993). Statistical tests of neutrality of mutations. Genetics, 133, 693–709.

22- Xia, X. DAMBE 6: (2017). New tools for microbial genomics, phylogenetics and molecular evolution. Journal of Heredity, 108, 431-437.

23- Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing

phylogenetic trees. Molecular Biology and Evolution, 4(4), 406–425.

24- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. Evolution, 39, 783–791.

25- Bandelt, H.J., Forster, P., & Rohl, A. (1999). Median-Joining Networks for Inferring Intraspecific Phylogenies. Molecular Biology and Evolution, 16, 37-48.

26- Hammer, Ø., Harper, D.A.T., & Ryan, P.D. (2001). PAST: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica, 4(1), 9-18.

27- Steane, D.A., McKinnon G.E., Vaillancourt R.E., & Potts B.M. (1999). ITS sequence data resolve higher level relationships among the eucalypts. Molecular Phylogenetics and Evolution 12(2), 215–223.

28- Udovicic, F., Mcfadden G.I., & Ladiges, P.Y. (1995). Phylogeny of *Eucalyptus* and Angophora based on 5S rDNA spacer sequence data. Mol. Phylogen. Evol., 4(3), 247–256.

29- Steane, D.A., Nicolle, D., McKinnon, G.E., Vaillancourt, R.E., & Potts B.M. (2002). Higher-level relationships among the eucalypts are resolved by ITS sequence data. Australian Systematic Botany, 12(2), 49–62.

30- Bayly, M.J., Pauline, Y., & Ladiges, P.Y. (2007). Divergent paralogues of ribosomal DNA in *Eucalypts* (Myrtaceae). Molecular Phylogenetics and Evolution, 44(1), 346–356.

31- Gibbs, A.K., Udovicic, F., Drinnan, A.N., & Ladiges, P.Y. (2009). Phylogeny and classification of *Eucalyptus* subgenus *Eudesmia* (Myrtaceae) based on nuclear ribosomal DNA, chloroplast DNA and morphology. Australian Systematic Botany, 22(3), 158–179.

32- Hawash, M.M. & Al-Shamma, L.M.J. (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.

33- Hawash, M.M. & Al-Shamma, L.M.J. (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.

34- Hawash, M.M. & Al-Shamma, L.M.J. (2026). Genetic variability and molecular evolution of the Iraqi date palm cultivars using internal transcribed spacer 1 (ITS1) region of nuclear ribosomal DNA. The Iraqi Journal of Agricultural Sciences, 57(2), 440-453.

35- Torres, R.A., Ganal, M. & Hemleben, V. (1990). GC balance in the internal transcribed spacers ITS1 and ITS2 of nuclear ribosomal RNA genes. Journal of Molecular Evolution, 30, 170–181.

36- Al-Shahwany, A.W., Hawash, M.M., & Hajeej, T.H.A. (2026).

Sequence diversity and molecular evolution analysis of international Tomato based on the entire ITS region. The Iraqi Journal of Agricultural Sciences, *57*(3), 676-688.

37- Kehie, M., Kumaria, S., Sangeeta Devi, K., & Tandon, P. (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.

38- Maina, N., Baraket, G., Salhi-Hannachia, A., & Sakkaa, H.B. (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.

39- Parra-O, C., Bayly, M., Udovicic F. & Ladiges, P. (2006). ETS sequences support the monophyly of the *eucalypt* genus *Corymbia* (Myrtaceae). TAXON, 55(3), 653–663.

40- Fu, Y.X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics, 147, 915–925.

41- Nasr, A., Bian, M-J., Chen, X-F., Wen, B., Wang, H., Shao J-W. & Zhu, G-P. (2024). Metabolites profiling of five *Eucalyptus* species by gas chromatography-mass spectrometry and multivariate analysis. Biologia Plantarum, 68(1), 97-106.

42- Bulgarelli, R.G.,. de Oliveira Silva, F.M., Bichara, S., Andrade, S.A.L., & P. Mazzafera (1019). *Eucalypts* and low phosphorus availability: between responsiveness and efficiency. Plant Soil, **445**, 349–368.

43- Dolácio, C.J.F., Oliveira, R.S., Nakajima, N.Y., Júnior, . I.daS.T., da Rocha, J.E.C., Ebling, Â.A. & Gama, M.A.P. (2020). Integration of principal component analysis and artificial neural network to modeling productive capacity of eucalypt stands from biophysical attributes. Forest Ecology and Management, 460, 117862.

44- Sahoo, H. & Kumar A.(2022). Genetics and Tree Improvement division, Institute of Forest Productivity, Ranchi-835303 (Jharkhand), India. International Journal of Science and Research Archive, 6(1), 063–067.