Iraqi Journal of Biotechnology, 2024, Vol. 23, No. 3, 260-269

Characterization of Repetitive DNA Sequences in Iraqi *Gerbera jamesonii*

[1]Noora J. Alsaedi, [1]Lamiaa Kh. Jawad

[1]Dept. Horti. Landscape Gard, College of Agricultural Engineering Science,University of Baghdad

Received: November 19, 2023 / Accepted: February 14, 2024 / Published: December 30, 2024

Abstract: The genome of the Transvaal daisy (*Gerbera Jamesonii*) contains several transposable elements, including long-term repeat retrotransposons. Nevertheless, little is known about the species-specific variability linked to retrotransposons. The study aimed to conduct a comprehensive analysis of the retrotransposon section of the Iraqi *Gerbera jamesonii's* genome in terms of quantity and quality of intraspecific variation relying on next-generation sequencing (NGS) technologies. Iraqi *Gerbera jamesonii's* genome was commercially sequenced using Illumina Novaseq6000 reads by DNA LINK Co. in Korea. The data have been utilized to recognize the tandem repeat clusters within the genome sequences using Geneious Prime software and the Repeat Explorer platform. Iraqi *Gerbera jamesonii* genome of 78485920 input reads of the random sample comprised approximately 80-82% repetitive elements, providing a total coverage of 2.37X. Out of 3928601 reads, 3209438, representing 72% of the genome, were clustered in 243 clusters. The majority of the top clusters were unannotated by Repeat Explorer. It was concluded that the annotated cluster proportion was 8.6124 of the LTR-RTs of Gypsy (2.4561) and Copia (6.1563).

Key words: *Gerbera jamesonii*, NGS technology, DNA marker, tandem repeats, clusters.

Corresponding author: (E-mail: noraa.jabor@coagri.uobaghdad.edu.iq).

Introduction

Tandem repeats are sequences of one or more nucleotides that are repeated in DNA in a certain manner and next to each other. (1), low copy sequences, coding sections, and regulatory units make up a very minor portion of a plant's genome, whereas repetitive DNA makes up the majority of it. Many thousands of superfamilies', families, and subfamilies can be found in repetitive DNA sequences, which are diversified. The length of the motifs, copy numbers, and genomic arrangements of these sequences differ. Changes in these repetitive DNA sequences are known to happen swiftly and to occur in tandem with increasing plant speciation (2).

A considerable amount of repetitive DNA can be found in eukaryotic genomes; these sequences can range in length from di-nucleotide motifs to more than 10 kb. The structure of repetitive DNA in the genome (either scattered or tandem arrays), the location of its chromosomes, and its purpose can all be used to classify it. The study of repeated DNA sequences in the genome has been made easier by recent developments in next-generation sequencing (NGS), which have made it feasible to sequence multiple eukaryotic genomic DNA sequences quickly and cheaply. Our understanding of repetitive

DNA sequences has evolved as a result of these developments (3, 4).

Gerbera jamesonii is one of the most important decorative plants; it is a member of the Compositae family. *Gerbera jamesonii* and *Gerbera viridifolia*, two wild African species, crossed to create this extremely heterozygous species (5, 6). utilizing the chromosomal count 2n = 2x = 50. Molecular markers are widely employed in the assessment of genetic diversity, genotype selection, cultivar identification, genetic connections, and the effective mapping and tagging of target genes. Next-generation sequencing makes it feasible to generate tandem repeats in batches. In the past, 40 gerbera accessions' genetic diversity was examined using 55 markers that were found in the transcriptome of gerbera (7, 8, 9).

Transposons are classified into three subclasses according on the number of broken DNA strands that occur during transposition. These subclasses are as follows: Cut-and-paste DNA transposons are found in Subclass 1, rolling-circle (Helitrons) DNA transposons are found in Subclass 2, and self-synthesizing (Polintons) DNA transposons are found in Subclass 3. (10, 11).

The objective of the research was to identify repetitive elements in the Iraqi *Gerbera Jamesonii* DNA based on sequencing of leaf tissues by using these markers for the development of molecular markers and genetic maps.

Material and methods

Plant material

The goal of the study was identify repetitive elements of Adapted Iraqi *Gerbera* were obtained from the agricultural nurseries (Abbas Al-Khafaji) in Mansour Street on October 2021, Baghdad. *Gerberas* abounded in it for more than ten years. These plants were distinguished by the color of white flowers. The leaves and roots were selected and chopped into small pieces, about one cm^2 in size, and placed in eppendorf tube. Fifteen replicates were divided into five replicates from the leaves to extract DNA, and five to extract RNA, as well as five replicates from the roots to extract RNA.

DNA link company

The samples were immersed in the RNA later in eppendorf tubes and sent to DNA link Company in South Korea, to extract and sequence DNA and RNA.

The results as compressed files

The zipped files were collected from the company as R1 and R2 for each of DNA and RNA. Then, the data were approached to the process of bioinformatics platforms.

The DNA sequencing details

The Illumina Novaseq6000 platform was used to sequence the best sample. The read length was 2x151 and the TruSeqDNA Nano 350bp kit was used with an index sequence of CGTTAGAA-GACCTGAA for Iraqi *Gerbera jamesonii* DNA as showed in (Table 1).

Table (1): Iraqi *Gerbera jamesonii* leaf DNA Sequence Information.

Sample ID	Index	Kit	Machine	Library	Read_Length
***Gerbera* leaf DNA**	F: CGTTAGAA R: GACCTGAA	TruSeqDNA Nano 350bp	Illumina Novaseq6000	PE	2X151

DNA quantification

The purity, quality, and quantity of genomic DNA were evaluated using a Nanodrop ND-1000 spectrophotometer, measuring at a wavelength of 260/280 nm with 1 µl of

each DNA sample. A high-quality DNA sample was considered when the OD 260/280 nm value was approximately 1.8-2.2.

Graph-based Clustering of raw reads (RepeatExplorer pipeline) of Iraqi *Gerbera jamesonii* :

Genomic DNA of Iraqi *Gerbera jamesonii* was sequenced commercially by DNA LINK, Korea using Illumina Novaseq6000reads.These data have been used in a basic analysis of tandem repeat clusters within genome sequences using Geneious (http://www.geneious.com/) (12). The platforms RepeatExplorer (13), and Tandem Repeat Analyser (TAREAN) (14), were used for graph-based clustering of repeated sequences in the raw reads were used for checking, finding and characterising repeat sequences, and the database of conserved protein motifs of retroelements (15). Copy numbers in raw reads were calculated by counting number of reads mapping to reference sequences.

To identify repetitive DNA families, graph-based clustering was performed on a random subset of *Gerbera jamesonii* genomic DNA, Then, the paired-end whole genome raw sequences were uploaded to the RE pipeline as FASTA files, and the clustering was feasible on the representative data set. A graph-based clustering approach was implemented Iraqi *Gerbera jamesonii* datasets in a default mode except checking the option of paired end.

The main results presented by RepeatExplorer are in HTML format and contain a table listing all clusters, genome proportions of each cluster and similarity hits. Graph clustering resulted in thousands of clusters. Most of the cluster graphs, with cluster annotation and their build up unassembled reads from the archive files were investigated manually. Clusters that represented high proportions of genomic DNA, clusters with unique pattern, high similarity hits to specific repeats, unclassified clusters, "Low complexity" and "Simple repeats" clusters were among the most interested clusters that have been investigated here. Each cluster was consists of a number of contigs, the contigs chosen randomly and sometimes the longer contigs have been chosen.

Data analysis

Genome proportions of each repetitive sequence from the annotated clusters were calculated by taking their summation to investigate the total genome proportions of each repetitive DNA family. The contig sequences were imported to the Geneious program to analyse, and search for any tandem repetitive DNA using dot plot and Tandem Repeats Finder (16), by using the default parameters.

De novo assembly

During the next generation sequencing procedures, the whole genomic DNAs have been broken into thousands or millions of DNA fragments (150-300 bp in length in current study). Assembling procedure make all these fragments reassemble into a continuous sequence again. The DNA fragments overlaps after assembly, however, repetitive sequence characteristic can complicate this process. The result of sequence assembly is multiple contigs (1000 contigs) with different lengths, and the contigs consensus sequences extracted as the reconstructed sequence from numerous of overlapped DNA sequence reads.

Dot-plot analysis

Geneious program dot-plot tool was used to compare two sequences against each other or one sequence against itself. This tool allows

identifying and detecting sequence homologous or polymorphism and localizing the region of similarity in the sequence whether it is present from start to end or in a particular region of the sequence, identifying the tandemly repeated array in the same sequence, or observation of the reverse complement for nucleotide comparisons.

Map to reference

In the Geneious, the procedure of assembling a target sequence and makes it as a reference to the whole genomic sequences to find either the genomic copy number for that reference sequence or find similarity hits. To assemble complete genes or chloroplast genome, it is easier to assemble the whole genome sequences against known sequence, this tool known as the map to reference. The result of the map to reference is producing just one contig per reference.

Copy number

It is a measure of the number of copies of any desired sequence from the genome and is calculated within an equation based on the results of the analysis of the desired sequence.

Results and discussion

The findings analysis of quality control of the Iraqi *gerbera jamesonii* DNA are shown in Table (2). Two wavelengths were measured with a nanodrop spectrophotometer to assess the sample's quality control. The sample concentration (2.86) µg and volume (40) µl were recorded at wavelength 260/280 (1.78), and at wavelength 260/230 (1.90).

Table (2): Iraqi *Gerbera jamesonii* DNA Sample QC Details

Sample request info			DNA sample info.					
Order ID (List)		Type	Nanodrop Conc.(ng/µl)	260/280	260/230	Conc. (ng/µl)	vol. (µl)	total (µg)
1	*Gerbera* leaf DNA	Tissue	72.4	1.78	1.90	71.5	40	2.86

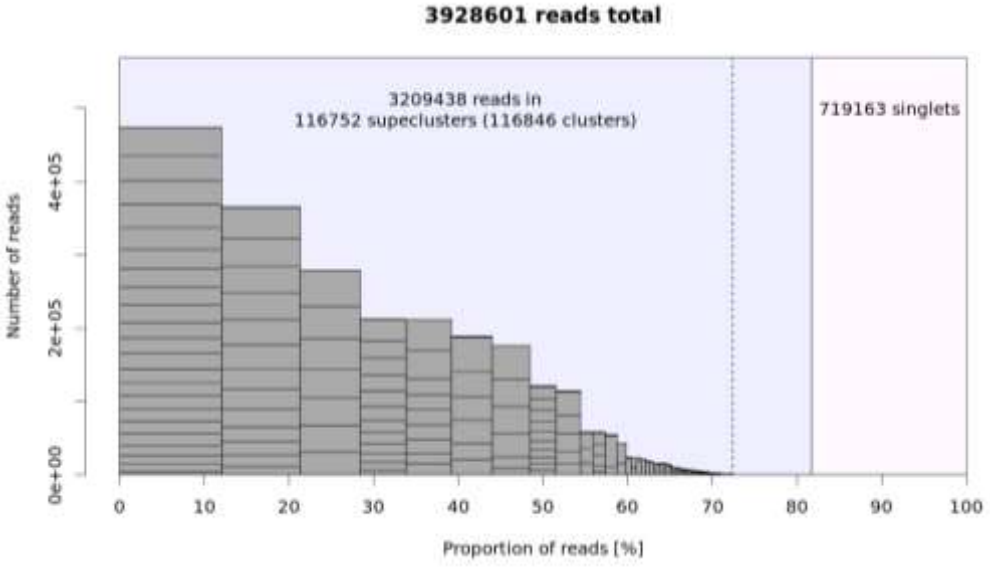

Figure (1): Graphical summary of the clustering results.

Superclusters are represented by bars, whose heights and widths, respectively, correlate to the percentage of each read in all the investigated reads (x-axis) and the number of reads in the superclusters (y-axis). Individual clusters are represented by rectangles inside the supercluster bars. The

clusters that were impacted by the filtering of abundant satellites are indicated in green, and their sizes match the modified values. The percentage of reads that stayed single and clustered are displayed in the pink and blue backdrop panels, respectively. On the left side of the dotted line are the top clusters.

Identification of tandem repeats by graph-based clustering

Approximately 80-82% of the Iraqi *Gerbera jamesonii* genomes consisted of repetitive elements, as shown in (Figure 1). The repetitive component of the Iraqi *gerbera jamesonii* genome was analyzed using RepeatExplorer (13) on a random sample of 78485920 input reads, providing a total coverage of 2.37X according to equation below: number of reads X reads length / genome size.

This tool identifies repetitive sequences in eukaryotic genomes, allowing *de novo* repeat identification by finding and quantifying similarities between individual sequences reads. This approach generated distinct, automatically labeled clusters of frequently linked reads that indicated distinct families of repetitive elements.

RepeatExplorer results showed that the proportion of repetitive DNA in Iraqi *Gerbera jamesonii* had about 243 clusters, 72% of the genome was represented by 243 clusters from 3209438 out of 3928601 reads. During our experiment, for the explanation of TAREAN, Putative satellites (high confidence) have two cluster (208, 241) in Proportion [%] (0.018, 0.010) and number of reads (719,409) respectively. About Putative satellites (low confidence) the results showed four clusters in this group (7, 111, 135, and 169) proportion (%) (1.100, 0.230, 0.110, 1.041) number of reads (41733, 9227, 4136, 1607) respectively (Tables 3, 4). It is possible that differences in low-copy retrotransposons can impact the physical traits of individuals. However, a reference genome sequence and substantial genotype resequencing are necessary for assessing these changes, especially in loci that contain these components (17).

Table (3): The explanation of TAREAN output in Iraqi *gerbera jamesonii*.

Putative satellites	Cluster no.	number of reads	Assembled reads	%	Galaxy %
high confidence	208	719	14,731	0.0188	0.018
	241	409	8,539	0.0109	0.01
low confidence	111	9227	213,122	0.2715	0.23
	135	4136	72,385	0.0922	0.11
	169	1607	154,224	0.1965	0.041
	7	41733	378,200	0.4819	1.1

Analyzing the Iraqi *gerbera* genome, involving 78485920 reads, with the repeat explorer platform, revealed 243 clusters comprising 3928601 reads constituting 5% of the total genome. Results in (Table 4, 5) demonstrate the cytogenetic and genome features of the Iraqi *gerbera* consisting of 50 chromosomes (2n=2x). The uploaded data showed that the GC percentage was 37.7%, the read length was 151 bp. and the Illumina coverage was 2.37x. The detailed analysis of repetitive elements from genomes of Iraqi *Gerbera jamesonii* used complementary tools (graph-based repeat clustering using unassembled raw reads, raw read mapping).

Even though there may have been biases for particular sequence types, we assumed that Illumina readings in our tests were sampled

equally. We totaled the reads (per million) for each genotype to determine the frequency of each cluster associated with LTR-retrotransposons. Retrotransposons become more frequent by placing retrotranscribed copies in various locations throughout the genome. The mobility of TEs can have a significant impact on phenotypic effects due to the insertion of elements near or within genes, resulting in a loss or alteration of their function (18, 19, 20).

Table (4): Repeat Explorer results out-come

Iraqi *gerbera jamesonii*	No. of reads uploaded	number of reads that were used in the clustering	percentage of total readings used for clustering	Cluster number (genome Proportion >= 0.01)
	78485920	3928601	5%	243

The development of DNA sequencing technology has substantially enhanced our understanding of genome size variation and the mechanisms underlying it. In plants, the majority of variation is found in the retrotransposon portion of the genome, which can change quickly (21). Retrotransposons can increase in number in a short amount of time because some of them are able to avoid control by the host genome's epigenetics (22). Furthermore, TEs can be swiftly eliminated by illegitimate and uneven homologous recombination (23, 24).

Table (5): Cytogenetics, genomic and sequencing features

Iraqi *gerbera jamesonii*	Cytogenetics, genomic and sequencing features
Chromosome number	2n=2x=50
Haploid genome size (kb)	2,5/1000
Whole genome sequences (GC %)	37.7%
Sequences read length (bp)	151
Illumine coverage	2,37X
Chloroplast read count throughout the entire genome	2,384944 0,03%
The genome's size (Gb) that was uploaded to RepeatExplorer	3928601

Novel approaches utilizing NGS and bioinformatics analysis can be employed to investigate the repetitive region of the genome even in animals lacking a reference genome sequence (25). Regardless of the genotypes within a given species, these novel approaches enable large-scale, genome-wide comparative study and profiling of DNA repeats. We used NGS techniques to create a retrotransposon sequence library for sunflower. The makeup of this library is consistent with the structure of the sunflower genome as previously studied and reported by (17, 26). This collection of LTR-RTs' intraspecific variation in redundancy is analyzed quantitatively and qualitatively using the library of sequences linked to retrotransposons. The mobility of retrotransposons in plants can impact the plant's phenotype based on their insertion location near genes, which may alter the gene's expression rate (18, 19). Additionally, retrotransposon mobility can cause inactivation of the region into which it is inserted through methylation. If a gene is present in that region, it may become inactivated (27). (28) Have verified that LTR-RTs are a collection of diverse creatures arranged into superfamilies, which are akin to species, within the genome.

Table (6): List of repetitive elements of Iraqi *Gerbera Jamesonii*

	Repeat Explorer outcome		Proprotion	No. of clusters
	Repeatitive elements		1.1690	6
LTR	**copia**	**Angela**	2.7003	22
		TAR	0.0787	10
		TORK	0.1015	29
		IKEROS	0.0681	21
		IVANA	0.4364	46
		SIRE	2.5599	43
		LYCO	0.0021	2
		BIANCA	0.0022	2
		Ale	0.2070	38
	Total copia		6.1563	
	gypsy	**Pararetrovirese**	0.0109	3
		chromovirus	1.6034	22
		non-chromovirus	0.8527	11
	Total gypsy		2.4561	
organelle/mitochondria			2.6490	40
organelle/plastid			2.7120	32
rDNA			0.1500	1
Other			28.6100	70
	Class_II/Subclass_1/TIR/	**Hat**	0.1270	6
		EnSpm	0.0860	5
		MuDR_Mutator	0.0456	6
		PIF_Harbinger	0.0134	7
		Tc1_Mariner	0.0015	1
	Subclass II Total		0.2735	
	Class_I/LINE:LINE	**RT**	0.0869	6
		ENDO	0.0437	4
		RH	0.0008	3
	Class I Total		0.1315	
	Helitron	**Helitron**	0.0069	1
	Grand Total		38.9533	

The functional and structural roles of repetitive elements

This study utilized bioinformatics analysis and cytogenetic tools to examine the diversity of tandemly repeated sequences. These elements, which are present in *Gerbera* genomes in low amounts, were identified in eight different types of clusters. These clusters are located in various regions of *Gerbera* chromosomes, including the centromeric, peri-centromeric, sub-telomeric, telomeric, and intercalary regions (Table 6).

The presence or absence of long terminal repeats (LTRs), which are direct repeats located at the extremities of the element, is used to categorize retrotransposons in research. There are two types: LTR-retrotransposons and non-LTR retrotransposons. Repeat Explorer identified that most of the top clusters were not annotated. However, out of the annotated clusters, (8.6124) were found to be similar to Gypsy (2.4561) and Copia (6.1563) LTR-RTs. Protein-encoding domains in LTR-RT clusters have been used to identify which families these clusters belong to within the Gypsy and Copia superfamilies. Thus, clusters that belong to the same LTR-RT family can be considered as LTR-RT subfamilies. Nine families were identified in the Copia superfamily (as shown in Table

6) (Figure 2), with Maximus/Angela (22 clusters, 2.7003% of the genome) and SIRE elements (43 clusters, 2.5599%) being the most prevalent. In the Gypsy superfamily (Figure 3), only three families were identified, with Chromovirus being the most abundant (22 clusters, over 1.6034% of the genome). A library was produced using all sequences from 243clusters annotated as LTR-RTs to serve as a reference for analyzing intraspecific variability of Iraqi *gerbera jamesonii* related to LTR-RTs (Table 6).

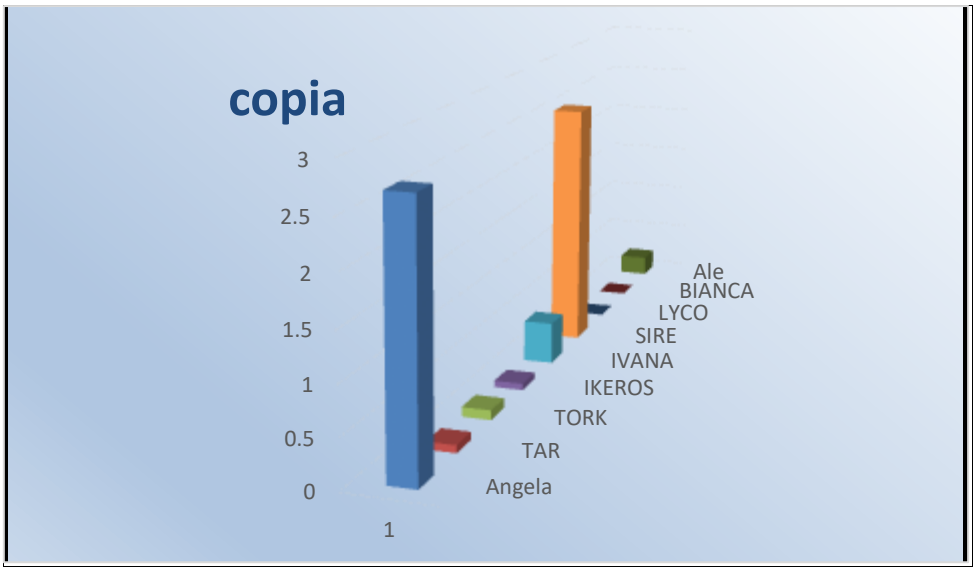

Figure (2): Copia superfamilies'.

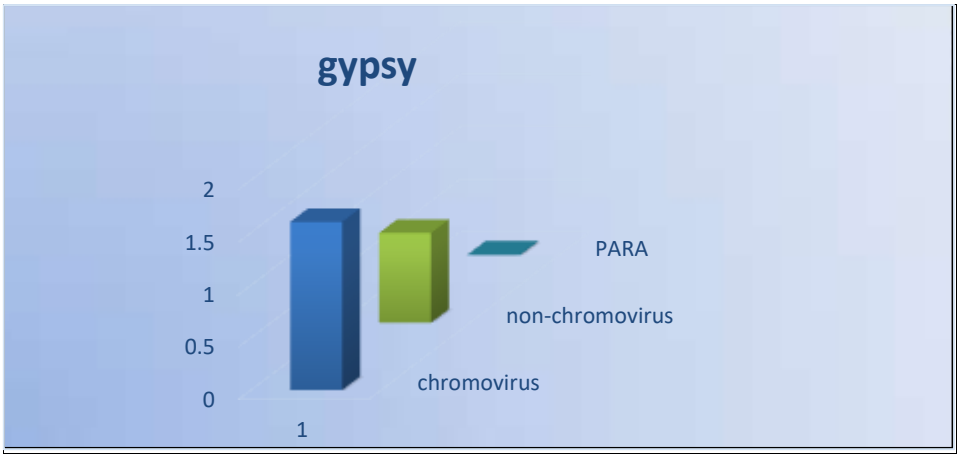

Figure (3): Gypsy superfamilies'.

There are "subspecies," or families, which differ in terms of their evolutionary background, activity, and protein sequences. According to a study by Lu *et al*, (29), some of the genetic changes that occur during the process of domestication can actually be harmful. Specifically, mutations that involve the amplification or loss of transposons may play a significant role. In fact, there have been a few reported instances where transposons have been found to be involved in the domestication of certain plant species (30). It is possible that the impact of varying redundancy of specific LTRRTs on the phenotype would be linked to the areas in the genome where these variations occur.

Typically, retrotransposon insertion leads to changes in both the structural and functional chromatin modifications, which can further as it is believed that repetitive DNA tandem repeats play a protective role in coding DNA by providing stability and protecting it from shock caused by stress conditions. These repeats have been utilized in nuclear architecture studies, identification of chromosomes and genomes, and in phylogenetic analysis. This information has been supported by studies conducted by Heslop-Harrison (1, 31) effect the functioning of nearby genes.

Conclusion

The research revealed that 80–82% of the repetitive elements in Iraqi *Gerbera jamesonii* were discovered to be present. Additionally, the annotated cluster fraction of the LTR-RTs of Gypsy (2.4561) and Copia (6.1563) was determined to be 8.6124. All sequences from 243 clusters designated as LTR-RTs were assembled into a library, which was then used as a reference to analyze intraspecific variability of Iraqi *gerbera jamesonii* associated with LTR-RTs.

References

1. Heslop-Harrison, J. S. (2000). Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. The Plant Cell, 12(5): 617-635.
2. Bennett, M. D. and Leitch, I. J. (2011). Nuclear DNA amounts in angiosperms: targets, trends and tomorrow. Annals of Botany, 107(3): 467-590.
3. Heslop-Harrison, J. S. and Schwarzacher, T. (2011). Organisation of the plant genome in chromosomes. The Plant Journal, 66(1): 18-33.
4. López-Flores, I. and Garrido-Ramos, M. A. (2012). The repetitive DNA content of eukaryotic genomes. Repetitive DNA, 7, 1-28.
5. Jasim, B. N.; Al-Salihy, A. A. and Moner, A. M. (2020). The Partial DNA Sequencing and Phylogenic Analysis of Tomato yellow leaf curl virus Isolated from Iraqi tomato. Iraqi Journal of Biotechnology, 19(1).
6. Al-Amery, L. K. J.; Abdul-Qader, Z. M. and Husni, H. S. (2023). Improving propagation of Echinacea purpurea and its content of some active compounds by using Tyrosine and Salicylic acid in vitro. Baghdad Science Journal, 20(3) 0919-0926.
7. Al-khalidy, R. M. S. and Aburesha, R. A. (2023). Molecular detection of some virulence gene in proteus mirabilis isolated from urinary tract infection in Iraq. Iraqi Journal of Agricultural Sciences, 54(3): 709-715.
8. Annon, A. H. and Abdulrasool, I. J. (2020). Effect of gamma radiation and ethyl methanesulfonate (EMS) on potato salt stress tolerance in vitro. The Iraqi Journal of Agricultural Science, 51(4): 982-990.
9. Hashoosh, Q. H. and AL-Araji, A. M. (2023). Molecular Identification of Tinea spp. Causing Tinea Disease using ITS Sequencing Analysis. Iraqi Journal of Biotechnology, 22(1): 220-228.
10. Morgante, M.; Hanafey, M. and Powell, W. (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nature Genetics, 30(2): 194-200.
11. Arnaot, M. R. T. (2015). Epigenetics, Biotechnology, and Other Scientific Advances: The Role of Iraqi Researchers. Iraqi Journal of Biotechnology, 14(1) 1-8.
12. Mehmood, F.; Shahzadi, I.; Waseem, S.; Mirza, B.; Ahmed, I. and Waheed, M. T. (2020). Chloroplast genome of Hibiscus rosa-sinensis (Malvaceae): comparative analyses and identification of mutational hotspots. Genomics, 112(1): 581-591.
13. Salih, M. and Hussein, R. (2017). Nuclear and chloroplast genome diversity in apomictic microspecies of Taraxacum (Doctoral dissertation, University of Leicester).
14. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S., *et al*. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28(12): 1647-1649.
15. Moner, R. A. S. A. M. (2022). Mapping to Reference: An Efficient Approach to Achieve Sufficient Consensus for Phylogenomic Study of Oryza Chloroplast Genome. Iraqi Journal of Biotechnology, 21(2) 439-445.

16. Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research, 27(2): 573-580.
17. Hansen, C. and Heslop-Harrison, J. S. (2004). Sequences and phylogenies of plant pararetroviruses, viruses, and transposable elements. Advances in Botanical Research. 41, 165-193.
18. Butelli, E.; Licciardello, C.; Zhang, Y.; Liu, J.; Mackay, S.; Bailey, P., *et al*. (2012). Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. The Plant Cell, 24(3): 1242-1255.
19. Falchi, R.; Vendramin, E.; Zanon, L.; Scalabrin, S.; Cipriani, G.; Verde, I., *et al*. (2013). Three distinct mutational mechanisms acting on a single gene underpin the origin of yellow flesh in peach. The Plant Journal, 76(2): 175-187.
20. Alisawi, O. N. (2019). Virus integration and tandem repeats in the genomes of Petunia (Doctoral dissertation, University of Leicester).
21. Wang, Q. and Dooner, H. K. (2006). Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus. Proceedings of the National Academy of Sciences, 103(47): 17644-17649.
22. Vitte, C. and Bennetzen, J. L. (2006). Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. Proceedings of the National Academy of Sciences, 103(47): 17638-17643.
23. Vitte, C. and Panaud, O. (2003). Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice Oryza sativa L. Molecular Biology and Evolution, 20(4): 528-540.
24. Vitte, C.; Fustier, M. A.; Alix, K. and Tenaillon, M. I. (2014). The bright side of transposons in crop evolution. Briefings in Functional Genomics, 13(4): 276-295.
25. Giordani, T.; Cavallini, A. and Natali, L. (2014). The repetitive component of the sunflower genome. Current Plant Biology, 1, 45-54.
26. Novák, P.; Ávila Robledillo, L.; Koblížková, A.; Vrbová, I.; Neumann, P. and Macas, J. (2017). TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. Nucleic Acids Research, 45(12): 111.
27. Hollister, J. D. and Gaut, B. S. (2009). Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Research, 19(8): 1419-1428.
28. Venner, S.; Feschotte, C. and Biemont, C. (2009). Dynamics of transposable elements: towards a community ecology of the genome. Trends in Genetics, 25(7): 317-323.
29. Lu, J.; Tang, T.; Tang, H.; Huang, J.; Shi, S. and Wu, C. I. (2006). The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. Trends in Genetics, 22(3): 126-131.
30. Naito, K.; Cho, E.; Yang, G.; Campbell, M. A.; Yano, K.; Okumoto, Y., *et al*. (2006). Dramatic amplification of a rice transposable element during recent domestication. Proceedings of the National Academy of Sciences, 103(47): 17620-17625.
31. Pluhar, S. A.; Erickson, L. and Pauls, K. P. (2001). Effects of tissue culture on a highly repetitive DNA sequence (E180 satellite) in Medicago sativa. Plant cell, Tissue and Organ Culture, 67: 195-199.