DNA-barcoding and Identification for Some Common Bean and Teparry Bean Genotypes using *rbcL* Gene

By

Mayada ,Y. E.[1,2], El-Akkad T. A.[1,2], Agez, A.[3], Refaat M. H.[1,2] and Abd-El-Sabour M. S.[1]

[1] Department of Genetics, Faculty of Agriculture, Benha University, Qalubia, Egypt.
[2] Moshtohor Research Park, Molecular Biology Lab., Benha University, Egypt.
[3] Agricultural Genetic Engineering Research Institute (AGERI), ARC, Giza, Egypt.

Abstract

Among the different biological sources, the common bean, and the wild type (Phaseolus spp.) Have many biotechnological applications. The DNA barcode of the use of the *rbcL* gene has proven its usefulness in the study of Phaseolus spp. Phylogenetic diversity of genotypes, more than a cryptic introduction, environmental modulation, geographic distribution, and species identification in exclusive bean species. Eleven samples of common bean had been obtained from unique locations in Egypt. The gene of rbcl has proven successful in PCR reactions. Seven samples of complete array with size ٤٠٣١ base pair have been obtained. The spot ranged from ٥٧٤ to ٥٧٥ bp with a common size of ٥٧٦ bp. The identification of the species published that the samples of the specimen River Garow, Bronco, Cerdon, Goru, Giza٣ three belong to *Phaseolus vulgaris*, and two samples of the specimen Teparry١١٣, Teparry١١٦ belong to *Phaseolus acutifolius*, respectively. The *rbcL*-based DNA barcode was once nearly profitable to discover distinctive specimens of Phaseolus spp according to species and genus. Some samples of the rbcL gene were no longer sufficient to perceive the stage of the genus and could not differentiate between relatively comparable species.

Key words: Barcoding; *rbcL* gene; phylogenetic diversity*;* Common Bean; Phaseolus spp.

Introduction

DNA barcoding relies on the discovery of unique conserved areas in different speciemans to collect huge library of standard genomes. DNA bar-coding is a universal way to know the species . Sequences of Nuclear and mitochondrial are lined up in rare speciemans to plan fully DNA-based barcode using to analyze (phylogenetic, genetic diversity, and species distinctions) in single creatures. The model classifies plant germplasm and shows genetic data of a species' ancestral origin inheritance behind it tags new types [Son, *et al*. ٢٠٠٣ and Hebert, et al., ٢٠٠٤].

Chloroplasts are an energetic metabolic team in inexperienced flowers for converting soft energy into Starches. High sequencing has resulted in the sequencing of nearly ٨٠٠ distinctive flower genomes of chloroplast [Kelchner, et al., ٢٠٠٠]. Two conserved areas of the plastid

(chloroplast) genome (*matk + rbcl*) had been submitted as barcode regions to characterize a massive collection of angiosperms. The molecular barcode picks out of revealing deception in plants. [Glossman, ٢٠٠٧ and Sikdar and Dutta ٢٠٠٨].

For profitable bar-coding experiment, the sequence distinction between ٢ species should be more suitable for distinguishing between the two. On the other hand , the intra and interspecies difference, the sequence version must be greatly reduced to discriminate the differences ٧ chloroplast genomic regions were evaluated via the Plant Biogenic Barcode Working Group Consortium [Jinbo, *et al.,* ٢٠١١].

The rbcl daily primer has excessive universality but low decision, while matK gives low universality but excessive decision between specific species. A mixture of each *matK + rbcl* marker can help discriminate most species. However, to achieve the highest rate of discrimination between two strongly associated species, the China Plant BOL group has already recommended a mixture of internal transcription spacer (*ITS + matK + rbcl*) [Li, *et al.,* ٢٠١١].

DNA molecule is a safer than RNA and decided in all plant tissues. Therefore, thoroughly DNA-based markers are favored for the identification of legume plants.

Objectives of the present topic were ٤ in number: (١) amplification and sequencing of rbcl regions in ١١ common bean cultivars of Phaseolus vulgaris and two wild accessions of Phaseolus acutifolius (٢) to study the homology model of the rbcl gene between two species using the Basic Local Alignment Research Tool (BLAST), (٣) comparative genomic analysis, to study these sequences in elements between different genotypes of common bean and tepary bean (٤) to explore the relationship between Phaseolus ssp. The development of the study would also be useful for the identification, authentication, and detection of alteration of the genetic material of legumes in national and international trade.

Materials and Methods
١. Plant Material:
A set of ١١ common bean cultivars of Phaseolus vulgaris and two wild accessions of *Phaseolus acutifolius* (٢n=٢٢) were evaluated under Egyptian condition. The ١١ *Phaseolus vulgaris* genotypes included belonged to the gene pools of common bean (River grow, GORU, Cerdon, Bronco, Giza٣, Scala, Giza ٦, Matilda, Paulista, Nambale and Samantha) and two wild accession Teparry١١٣(G٤٠٠٨٣) and Teparry١١٦(G٤٠٠٨٤),Seeds of these cultivars were obtained from the National Germplasm Resources Laboratory, Beltsville, USA and Germplasm Preservation Laboratory, Faculty of Agriculture at Moshtohor, Benha University and Horticulture Research Institute, Agricultural Research Center, Dokki, Egypt.

۲.DNA Barcoding and Sequencing Analysis:

For most of the seven samples of wild and common beans (Phaseolus spp.), DNA was once extracted from (۱۰۰ mg) of leaves the usage of DNeasy Plant Mini Kit (Qiagen, Santa Clarita, CA) and DNA template with specific bands in For Polymerase chain reaction amplification and rbcL sequencing regions , with specific primers were respectively *rbcLF* (۵'ATGTCACCACAAACAGAGACTAAAGC۳') and *rbcLR*(۵'TCGCATGTACCTGCAGTAGC۳') The product size predicted by PCR was ۶۰۰ bp.

DNA amplification was performed by a thermal cycler(PCR) TProfessional (Biometra, Germany) using specific protocol and PCR products were posted on the UV Gel Documentation System (Fire Reader XSD۵۵۲۰Mand) with DNA Marker ladder (۱۰۰ bp); Amplified products for all PCR had been purified the use of Qiagen® PCR Purification Kit(Qiagen, Santa Clarita, CA).

The purified PCR product and incubated at room temperature for two minutes and hold DNA at (-۲۰°C). The rbcL PCR merchandise DNA sequencing by usage of Sanger DNA sequencing method (Macrogen® Inc., Seoul, Rep. of Korea).

The similarity search of the PCR product was performed with a Basic local alignment searching tool (BLAST), MEGA and Neighbor-Joining(NJ) (Thompson, *et al.* ۱۹۹٤; Altschul, *et al.*۱۹۹۷ and Kumar *et al*. ۲۰۰۱), using all the regions of the two different species of Phaseolus were used as seven request sequences (LC۵۷۸۸۳۵٫۱, LC۵۷۸۸۳۷٫۱, LC۵۷۸۸۳۸٫۱, LC۵۷۸۸۳۹٫۱, LC۵۷۸٤۰٫۱, LC۵۷۸٤۰٫۱ and LC۵۷۸۸٤۵٫۱).

Results

primers of PCR rbcL were successful in amplifying and producing a PCR product with expected band size (≠ ۶۰۰ bp) (**Fig. ۱**).

۲.Species identification:

All species sequences had been efficiently recognized and most of the sequences had an identity level higher than ۹۶٪. However , the coverage of queries representing the size of comparable nucleotides between the query (our PCR product sequences) and the challenge (rbcL genes in the NCBI database) domain from ۹۹٫۶۵٪ to ۱۰۰٪, indicating a similarity to all common bean queries. , could be the low variety of sequences of this species in the database or the high effectivity of the rbcL gene to pick out these genotypes.

The species identification revealed that the specimen's samples River Garow(LC۵۷۸۸۳۵٫۱), Bronco(LC۵۷۸۸۳۸٫۱), Cerdon (LC۵۷۸۸۳۷٫۱), GORU(LC۵۷۸۸۳۶٫۱), Giza ۳(LC۵۷۸۸۳۹٫۱) belongs

to *Phaseolus vulgaris*, and two specimen's samples Teparry١١٣ (LC٥٧٨٨٤٠,١) , Teparry١١٦ (LC٥٧٨٨٤٥,١) belongs to *Phaseolus acutifolius* , respectively.

٣.Phylogenetic analysis:

The *rbcL* regions sequences together with immoderate similar GenBank genes were used to generate the dataset of nucleotide for phylogenetic evaluation (Fig. ٣). The phylogenetic tree advanced for LC٥٧٨٨٣٥,١ (River Garow) divided specimens from into four groups. The common branch duration changed into ٠,٠٠١٦, with a version of (٠,٠), domain from (-٠,٠١١ to ٠,٠٠٠٩). The question (River Garow) with shaped a monophyletic clade together with exceptional species belongs to Phaseolus genus. While the Species (*Phaseolus lunatus* and *Phaseolus coccineus*) formed a particular clade (Fig. ٣A).

In LC٥٧٨٨٣٨,١ (Bronco) the phylogenetic tree endorse size is (٠,٠٠١٤٨) with a variance of (٦١,٤٤). Where the tree dimension domains from (٠,٠٠١١ to ٠,٠٠٣٢). The phylogenetic tree consists of ٦ clades, the area Bronco genotype shaped one clade with any different Phaseolus lunatus. The Phaseolus genus shaped separate clades, which encouraged that rbcL gene can effectively differentiate between these genera (Fig. ٣B).In query LC٥٧٨٨٣٧,١ (Cerdon), the phylogenetic tree suggests that the dimension is ٠,٠٠١٥٤ with a variance of ٦١,٥٤. Where the tree dimension ranged from (٠,٠٠١١ to ٠,٠٠٢٧).

The phylogenetic tree is composed of ٥ clades, the location Cerdon genotype formed one clade with each other *Phaseolus lunatus and Phaseolus coccineus*. The Phaseolus genus formed separate clades, which endorsed that rbcL gene can successfully differentiate amongst those genera (Fig. ٣C).

The query's LC٥٧٨٨٣٩,١ and LC٥٧٨٨٣٨,١ (Giza٣ and Bronco) genotypes fashioned with sequences belongs to the identical genus (*Phaseolus vulgaris*) a clade, whilst sample ٦ shaped with rbcL sequence belongs to the identical species (Fig. ٣D-E).Clearly clustered all rbcL sequences of Phaseolus lunatus in one clade and *Phaseolus coccineus* in another, indicating that even though sequences more than a few inside the species, variations are minimal pattern to cluster the same species in one clade indicating the DNA barcode conceivable of the rbcL gene.

Moreover, the phylogenetic tree built with LC٥٧٨٨٤٠,١ (Teparry-١١٣) and different GenBank sequences divided into ٦ clades, every with a recommend dimension and variance of (٠,٠٠٦٩ and ٩٨,٥٥), respectively.

The branch sizes on the phylogenetic tree ranged from ٠,٠٠١٦ to ٠,٠٢٤. Different Phaseolus ssp. *rbcL* sequences formed a clade with the LC٥٧٨٨٤٠,١ (*Phaseolus acutifolius*) (Fig. ٣F).

For the LC٥٧٨٨٤٥٫١ (Teparry-١١٣) tree, the suggested department size and variance were (٠٫٠٠٦٩ and ٩٨٫٥٥), respectively. The branch size ranged from ٠٫٠١٣ to ٠٫٠٠٣ mm. *Phaseolus lunatus, Phaseolus coccineus, Phaseolus carteri,* and *Orbexilum pedunculatum* sequences were all found in the LC٥٧٨٨٤٥٫١ (*Phaseolus acutifolius*) clade (Fig. ٣G).

Discussion

Information on plastid genome sequences is of brilliant price importance for barcoding plants. Sequence data performs a key position in figuring out and featured species, thus allowing the certificates of industrial and particularly products.

We used the chloroplast trnL (Taberlet, et al., ٢٠٠٧) and the rpoC١ zones and the ITS٢ nuclear vicinity to barcode the necessary Mediterranean legume species. Using DNA barcoding sequences, (Gao and Chen, ٢٠٠٩) evaluated the opportunity of ٤ coding chloroplast areas (rpoB, rpoC١, rbcL, and matK) and two noncoding nuclear areas (ITS, ITS٢) serving as barcodes for the plants in the Fabaceae family.

The sequencing information for rbcL turned into applied to create a dendrography depiction of the variations through the species studied. Although it isn't always a real phylogenetic tree, it could be used to differentiate among *Phaseolus coccineus* and *Phaseolus lunatus*, in addition to *Vigna unguiculata* and *Vigna radiata*. Because of the technique hired to assemble the tree, the position of the various species covered on this examine withinside the phylogenetic tree can't be immediately as compared to placements mounted in different studies. We received ١٠٠٪ identity with *Phaseolus lunatus, Phaseolus coccineus, and Phaseolus carteri* from the National Center for Biotechnology Information database with *Phaseolus vulgaris*. DNA barcodes were approach in local species which could be a step to installation DNA totally definitely tracking protocols for economic plant adulteration below national and global trading. However, there are lot of legumes genomes lacking series information.

A genus totally based identity tool could be a technique of preference in species identity in indigenous plants. Resequencing of additional loci for target-primarily based totally upgrades could as an alternative be beneficial to study greater conserved genomic areas among notable plant species. The findings of contemporary-day find out about propose that using frequent primers (rbcL) for DNA barcoding is worthwhile for amplification, identity, and discrimination of above-cited indigenous plant species. The amplification fulfillment fees have been ١٠٠٪ for all the ٣ species. Both the primer pairs resulted sharp bands that have been required for reliable DNA sequencing. Thus, using DNA barcodes primers systems is reliable, fast, and inexpensive tool identity of legumes flora at genus and species degree of indigenous land species.

Generating a large DNA database while focusing additional conserved areas would be effective in identifying legume flowers [Padial, et al., ٢٠١٠; Schindel and Miller, ٢٠٠٥]. These data would also be useful in discovering the taxonomy, ecology, phylogeny, and morphology of distinctive species [Hajibabaei, et al., ٢٠٠٧]. However, improving new amplification protocols and techniques with novel primer cocktails would dramatically revolutionize the discipline of DNA barcodes through the representation of additional information about the genome of unique species.

Conclusion

Finally , Chloroplast genomes' sequences had been used for phylogenetic evaluation via ML based totally on unique Phaseolus species. *Phaseolus vulgaris* and *Phaseolus lunatus* are sister species, *Phaseolus lunatus* is extra carefully associated to *Phaseolus vulgaris, Vigna unguiculata* and *Vigna radiata*. Consistent with the gene order results, they're all of subtribe phaseolinae.

According to species and genus, *rbcL* based DNA barcoding was almost effective in identifying distinct seven common and wild bean genotypes specimens. Some *rbcL* specimens were unable to distinguish between extremely similar species and were unable to identify genus level. We propose that, in addition to *rbcL*, further DNA barcoding approaches be used to improve species identification resolution.

These findings are extremely helpful in the development of molecular markers for barcoding. Teparry bean has a unique genome when compared to other legume species (*Phaseolus acutifolius*).

References

Altschul SF *et al.*, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, ١٩٩٧; ٢٥(١٧): ٣٣٨٩–٣٤٠٢. https://doi.org/١٠,١٠٩٣/nar/٢٥,١٧,٣٣٨٩

Bakker, F.T, Culham A, Gomez-Martinez R, Carvalho J, et al. (٢٠٠٠). Patterns of nucleotide substitution in angiosperm cpDNA *trnL* (UAA)-*trnF* (GAA) regions. *Mol. Biol. Evol.* ١٧: ١١٤٦-١١٥٥.

Chase, M.W., Cowan RS, Hollingsworth PM and van den Berg C (٢٠٠٧). A proposal for a standardised protocol to barcode all land plants. *Taxon* ٥٦: ٢٩٥-٢٩٩.

Deng CY, Xin GL, Zhang JQ, Zhao DM. Characterization of the complete chloroplast genome of Dalbergia hainanensis (Leguminosae), a vulnerably endangered legume endemic to China. Conserv Genet Resour. ٢٠١٩;١١:١٠٥–٨.

Gargiulo, G.M, Morabito M, Manghisi A. A re-assessment of reproductive anatomy and postfertilization development in the systematics of Grateloupia (Halymeniales, Rhodophyta). *Cryptogam. Algol.,* ٢٠١٣; ٣٤(١): ٣–٣٦.

Glossman, M.D. CHIH-DFT determination of the molecular structure and infrared and ultraviolet spectra of-solanine. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, (٢٠٠٧); ٦٦(١): ٢٠٨-٢١١.

González, T.E., Tapia E., López-Meraz L., Navarrete A., Reyes-Ramírez A, et al. Anticonvulsant Effect of Annona diversifolia Saff. and Palmitone on Penicillin-induced Convulsive Activity. A Behavioral and EEG Study in Rats. Epilepsia, (٢٠٠٦); ٤٧(١١): ١٨١٠-١٨١٧.

Hajibabaei M, Singer GA, Hebert PD, Hickey DA. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends in Genetics, (٢٠٠٧); ٢٣(٤): ١٦٧-١٧٢.

Hebert, P.D, Cywinska A, Ball SL. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London B: Biological Sciences, (٢٠٠٣); ٢٧٠(١٥١٢): ٣١٣-٣٢١.

Hebert, P.D, Penton EH, Burns JM, Janzen DH, Hallwachs W. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. Proceedings of the National Academy of Sciences USA, (٢٠٠٤); ١٠١(٤١): ١٤٨١٢-١٤٨١٧.

Hollingsworth PM, Graham SW and Little DP (٢٠١١). Choosing and using a plant DNA barcode. *PLoS One* ٦: e١٩٢٥٤.

Hollingsworth, P.M, Forrest LL, Spouge JL and Hajibabaei M (٢٠٠٩). A DNA barcode for land plants. *Proc. Nat. Acad. Sci. U. S. A.* ١٠٦: ١٢٧٩٤-١٢٧٩٧.

Jinbo U, Kato T, Ito M. Current progress in DNA barcoding and future implications for entomology. Entomological Science, (٢٠١١); ١٤(٢): ١٠٧-١٢٤.

Kaushik D, Jogpal V, Kaushik P, Lal S, Saneja A, et al. Evaluation of activities of Solanum nigrum fruit extract. Archives of Applied Science Research, (٢٠٠٩); ١(١): ٤٣-٥٠.

Kelchner SA. The evolution of non-coding chloroplast DNA and its application in plant systematics. Annals of the Missouri Botanical Garden, (٢٠٠٠); ٨٧(٤): ٤٨٢-٤٩٨.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, et al. (٢٠٠٥). Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* ١٠٢: ٨٣٦٩-٨٣٧٤.

Kumar S, Tamura K, Jakobsen IB, Nei M (٢٠٠١) MEGA٢: molecular evolutionary genetics analysis software. Bioinformatics ١٧(١٢):١٢٤٤–١٢٤٥

Li FW, Kuo LY, Rothfels CJ, Ebihara A, Chiou WL, et al. rbcL and matK earn two thumbs up as the core DNA barcode for ferns. PLoS One, (٢٠١١); ٦(١٠): e٢٦٥٩٧.

Montes M., Rico J.M., Garcia-Vazquez E., Pichs Y.J.B. Molecular barcoding confirms the presence of exotic Asian seaweeds (*Pachymeniopsis gargiuli* and *Grateloupia turuturu*) in the Cantabrian Sea, Bay of Biscay. *Peer J.*, ٢٠١٧; (٥): e٣١١٦.

Padial JM, Miralles A, De la Riva I, Vences M. The integrative future of taxonomy. Frontiers in Zoology, (٢٠١٠); ٧(١): ١.

Ravi V, Saleem TM, Patel S, Raamamurthy J, Gauthaman K. Anti-inflammatory effect of methanolic extract of Solanum nigrum Linn berries. International Journal of Applied Research in Natural Products, (٢٠٠٩); ٢(٢): ٣٣-٣٦.

Schindel DE, Miller SE. DNA barcoding a useful tool for taxonomists. Nature, (٢٠٠٥); ٤٣٥(٧٠٣٨): ١٧-١٧.

Sharma Y, Singh H, Mehra B. Hepatoprotective effect of few Ayurvedic herbs in patients receiving antituberculus treatment. Indian Journal of Traditional Knowledge, (٢٠٠٤); ٣٣٩١-٣٩٦.

Sikdar M, Dutta U. Traditional phytotherapy among the Nath people of Assam. Ethno-med, (٢٠٠٨); ٢(١): ٣٩-٤٥.

Son YO, Kim J, Lim JC, Chung Y, Chung GH, et al. Ripe fruits of Solanum nigrum L. inhibits cell growth and induces apoptosis in MCF-٧ cells. Food and Chemical Toxicology, (٢٠٠٣); ٤١(١٠): ١٤٢١-١٤٢٨.

Taberlet P, Coissac E, Pompanon F and Gielly L (٢٠٠٧). Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* ٣٥: e١٤

Thompson JD, Higgins DG, Gibson TJ cluster W. Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, ١٩٩٤; ٢٢(٢٢): ٤٦٧٣–٤٦٨٠. https://doi.org/١٠,١٠٩٣/nar/٢٢,٢٢,٤٦٧٣.

Yang M.Y., Kim M.S. Taxonomy of Grateloupia (Hymeneals', Rhodophyta) by DNA barcode marker analysis and a description of Pachymeniopsis volvita sp. Nov. *J. Appl. Phycol.*, ٢٠١٥; ٢٧(٣): ١٣٧٣–١٣٨٤.

Zha X, Wang X, Li J, Gao F, Zhou Y. Complete chloroplast genome of Sophora alopecuroides (Papilionoideae): molecular structures, comparative genome analysis and phylogenetic analysis. J Genet. ٢٠٢٠;٩٩:١٣.

Zhang W., Wang H.W., Zhao D., Sheng Y.W., Luan R. X. Morphological observation and rbcL gene sequence analysis of *Grateloupia orientalis* (Rhodophyta), a new record of the Hainan province. Mar. Sci., ٢٠١٢; ٣٦(٧): ١٠٩–١١٦

Table (١): The NCBI blast results details for the Seven Phaseolus spp. genotypes *rbcL* Region sequences.

Genotypes	Taxonomic Level		Sequence Length	Max Score	Total Score	Sequence Cover	E - value Blast	Blast Similarity %	Similar Sequences No.
	Genus	Species							
River garow (LC٥٧٨٨٣٥,١)	*Phaseolus*	*vulgaris*	576	1061	1061	100%	0.0	100	
GORU (LC٥٧٨٨٣٦,١)	*Phaseolus*	*vulgaris*	577	1061	1061	100%	0.0	100	**LT٥٧٦٨٥١,١** **LT٥٧٦٨٥٣,١** **MK٣٤٨٦٢٦,١** **MK٣٤٨٦٢٥,١** **LT٥٧٦٨٥٠,١** **KJ٧٧٣٧٣٤,١**
Cerdon (LC٥٧٨٨٣٧,١)	*Phaseolus*	*vulgaris*	576	1061	1061	100%	0.0	100	
Bronco (LC٥٧٨٨٣٨,١)	*Phaseolus*	*vulgaris*	574	1061	1061	100%	0.0	100	
Giza٣ (LC٥٧٨٨٣٩,١)	*Phaseolus*	*vulgaris*	576	1050	1050	99.65%	0.0	100	
Teparry-١١٣ (LC٥٧٨٨٤٠,١)	*Phaseolus*	*acutifolius*	576	1050	1050	99.65%	0.0	100	

Teparry-١١٦ (LC٥٧٨٨٤٥,١)	Phaseolus	acutifolius	576	1050	1050	99.65%	0.0	100	

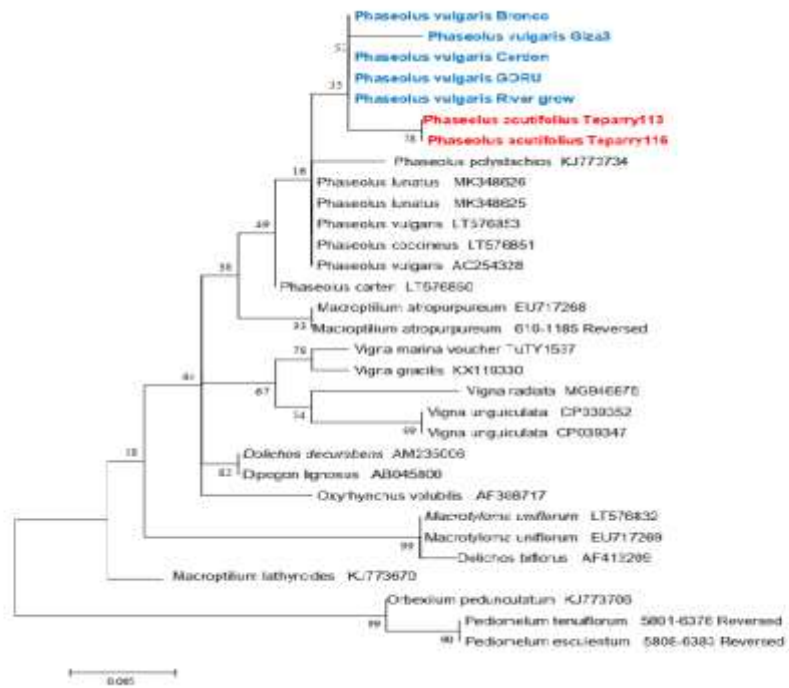**Fig. ١.** The gel electrophoresis result for the PCR product of *rbcL* primers.
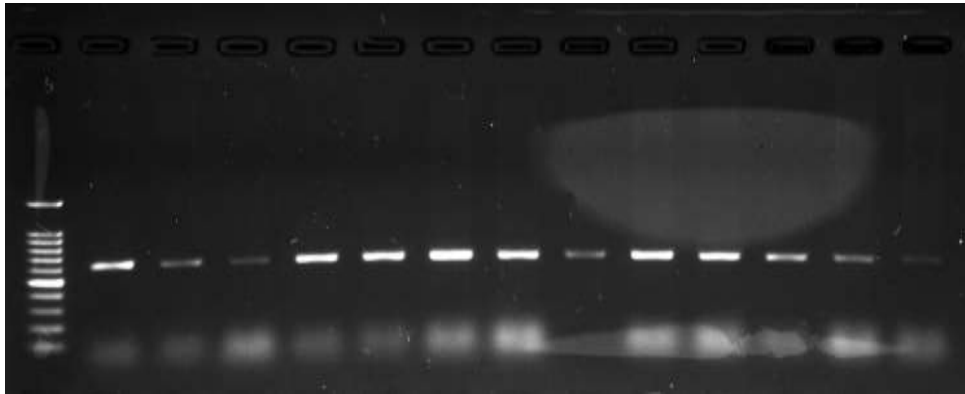

Fig. ٢. The phylogenetic trees depicting the relationship between different Phaseolus spp.
Genotypes specimens and GenBank sequences.

Fig. ٣. The phylogenetic trees depicting the relationship between different Phaseolus spp.
Genotypes specimens and GenBank sequences.