JJBS

Volume 10, Number 2,June 2017
ISSN 1995-6673
Pages 79 - 84

Jordan Journal of Biological Sciences

# Molecular Cloning and in *Silico* Analysis of *rps7* Gene from the *Lactuca sativa*

Mahdieh Gholipour , Bahram Baghban Kohnehrouz *

*Department of Plant Breeding and Biotechnology, Faculty of Agriculture, University of Tabriz, Tabriz, Iran*

*Received: October 10, 2016 Revised: March 2, 2017 Accepted:March 12, 2017*

**Abstract**

The *rps7* genes encode 30S ribosomal protein S7 that bind directly to the parts of the 3'end of 16S ribosomal RNA. The *rps7* gene is one of the most important plastidial genes that have not been studied so far in lettuce (*Lactuca sativa*). The molecular analysis of *rps7* gene provides an opportunity to make phylogenetic studies and a conserved homologous recombination site for designing chloroplastic expression vectors with high-efficiency lettuce cultivars and its related species. Here, we have cloned plastid *rps7* gene from lettuce. The full length of lettuce *rps7* gene is 2579 bp, composed of the 289 bp promoter region, an Exon 468 bp long encoding 155 amino acids. The *rps7* gene contains a conserved domain (1-155). The promoter region analysis showed that the cloned fragment contains motifs, like TATA box, CAAT-box, MBS, AAGAA-motif and GA-motif. The alignment analysis *rps7* showed the high homology with identity of 96% compared with *Asteraceae* family at the nucleotide level, while the amino acid sequence of the gene showed a high similarity value of 98.1-90% compared with hairy sunflower, wormwood, ragwort and soybean. The phylogenetic analysis result shows the *rps7* gene from dicots is close to *rps7* gene from lettuce varieties.

**Keywords:** Plastid Genome, Lettuce, *rps7* Gene, PCR

## 1. Introduction

Chloroplasts are heterogeneous plastidial organelles responsible for many metabolic processes such as photosynthesis, synthesis of amino acids, fatty acids and secondary metabolites (Marín *et al*., 2007). Chloroplasts contain circular double-stranded DNA with 76-217kbp long includes approximately 120-130 genes present in 1,000–10,000 copies per cell, and maternally inherited in the most angiosperm-plant species. Although they have prokaryotic origin distinct from the nucleus, their transcription factors with prokaryotic characteristics can be matched with the transcription factors of eukaryotic mRNAs. Furthermore, they are polyploid in higher plants and they have an extremely conserved organization (Raubeson and Jansen, 2005), (Hinsinger and Strijk, 2015) and (Zhang and Gao, 2016). The proteins specifically Multi-subunit functional protein complexes that are involved in photosynthesis are encoded by plastidial genome and directly synthesized within the chloroplast (Daniell *et al.*, 2016). Because of its abundance in plant cells and ease of sequencing, chloroplast DNA (cpDNA) has been widely utilized in studies of plant taxonomy and evolution (Kress and Erickson, 2007) and (Taberlet *et al*., 2007).

The *Lactuca sativa* chloroplast genome (DQ383816) with 152772 bp-long contain an inverted repeats (IRs) (25034 bp) that are separated by a large (84105bp) and small single-copy (18599bp) (LSC and SSC) regions. Moreover, the cpDNA from *L.sativa* contain several genes including encoding genes of tRNAs, rRNAs, RNA polymers, ribosomal proteins, ATP synthesis, cytochrome b6/f, NADH dehydrogenases. In higher plants, a pair of *rps7* genes encoded ribosomal protein S7 protein from the small ribosomal subunit was located in Inverted Repeats (IRs) regions of the plastid genome. Therefore, *rps7* genes can be used as a homologous recombination site to achieve high yield expression level of a gene of interest. Following, a gene of *rps7* that is located in the IRa, is transcribed in the clockwise direction and the other that is located in the IRb, is transcribed in the counter clockwise direction (Timme *et al*., 2007). In *Escherichia coli*, S7 is known as the protein that binds directly to the parts of the 3'end of 16S ribosomal RNA. It belongs to a family of ribosomal proteins, which have been grouped on the basis of sequence similarities (Klussmann *et al*., 1993) and (Ignatovich *et al*., 1995). This entry represents the S7

* Corresponding author. e-mail: bahramrouz@yahoo.com.

structural domain, which consists of a bundle of six helices and an extended beta hairpin between helices 3 and 4 with two or more RNA-binding sites on its surface (Wimberly *et al*., 1997). This entry also represents 30S ribosomal proteins S7 (bacterial, archaeal, plastid, mitochondrial), and eukaryotic 40S ribosomal proteins S5 (cytoplasmic). The 30S ribosomal protein S7 make an interaction between ribosomal proteins of S9 and S11. It is also one of the primary rRNA binding proteins that bind directly to 16S rRNA, where it nucleates assembly of the head domain of the 30S subunit S7 is located at the subunit interface close to the decoding center shown to contact mRNA (Robert and Brakier-Gingras, 2001). It has also been shown to contact tRNA in both the P and E sites, it probably blocks the exit of the E site tRNA.

At higher taxonomic levels (family level), protein-coding regions and conserved sequences of the chloroplast genome can be used to phylogenetic analysis and domestication studies (Jansen *et al*., 2007). Since, the *rps7* gene is one of the most important genes of chloroplast genome with a conserved structure, the molecular analysis of this gene can be used for phylogenetic studies. Understanding the genetic relationships between crops and their close relatives would provide an attempt to introduce specific advantageous traits into the related crops.

## 2. Materials and Methods

### *2.1. Plant material and DNA Extraction*

Lettuce (L.*sativa.var*. *Salinas*) seeds were planted in plastic trays, and let to grow to 2-3 leaf stage before extracting their DNA. Total genomic DNA was extracted from fresh leaves using a modified CTAB method (Doyle, 1987; Yang, Li & Li, 2014). Quality, quantity and concentration of the extracted DNA were determined by 0.8% agarose gel electrophoresis.

### *2.2. Designing of Primers*

To obtain the complete sequence of the lettuce *rps7* gene a pair of specific primer was designed by using Primer3 online software. Sense and anti-sense primers for this fragment were (F: 5'-AGTTTTCATTCTGTACATGCCAG -3') (R: 5'-CTCACTAAGCCGGGATCACT -3').

### *2.3. PCR Amplification and Bacterial Transformation*

Total genomic DNA of lettuce was used as a template for amplification of the target fragment in the concentration of 5 ng/µl for the fragment. The amplification profile was 94°C for 5min, followed by 35 cycles of 94°C for 30 Sec, 50°C for 40 Sec, and 72°C for 1min and 20 Sec, and a final extension step at 72°C for 10min. The amplified DNA product was purified by using a gel extraction kit from Bioneer Company. The ligation reaction of the related fragment with Vivantis company *p*TG19-T vector was done through T/A cloning by using *T4* DNA ligase enzyme (produced by Termo Company). In order to transform, first *E.coli* strain DH5α bacterial competent cells were prepared by TSS method, then thermal shock method was used for transformation of competent cells. Transformants were selected on white-blue test medium containing antibiotic ampicillin, IPTG and X-gal.

Colony PCR was performed to confirm the presence of the cloned fragment in the white colonies bacteria. The plasmid DNA was extracted from the liquid culture of PCR positive colonies using a plasmid extraction kit from Bioneer Company. Finally, the recombinant plasmids DNA was reconfirmed with *EcoR*I (produced by Termo Company) digestion (because the restriction site of this enzyme is located on the *p*TG19-T) and was sent for sequencing to Bioneer Co.

### *2.4. Bioinformatics Studies*

Alignment of plant *rps7* At first, DNA-to-protein translation of DNA sequences was performed using the EXPASY translate tool (http://au.expasy.org/tools/dna.html) then the amino acid sequence of the lettuce *rps7* gene was compared with other *rps7* genes by ClustalW2 (www.ebi.ac.uk/Tools/msa/clustalw2/) software. Sequence analysis of rps7 promoter to find regulatory elements in promoter sequences, was done by PLANTCARE (http://Bioinformatics.psb.ugent.be/webtols/plantcare/html) software. Domain detection was done for finding all known motifs by using the EXPASY Motif Scan tool (http://myhits.isb-sib.ch/cgi-bin/motif_scan).The phylogenetic analysis was conducted by NCBI.

## 3. Result and Discussion

Using primers generated from the plastid DNA sequence *rps7* of IRa, DNA fragment of 2629 bp was synthesized by PCR from lettuce plastid DNA as a template (Figure 1). This fragment was cloned into *p*TG19-T. (Fragment length=2579 bp and *p*TG19-T length=2880 bp).

The full length of lettuce *rps7*gene (L.*rps7*) is 2579 bp, composed of the 289 bp promoter region, an exon and 1822 bp3'terminus. Exon is 468 bp long, beginning with nucleotide 1, extending to nucleotide 468 encoding 155 amino acids (Figure 2).

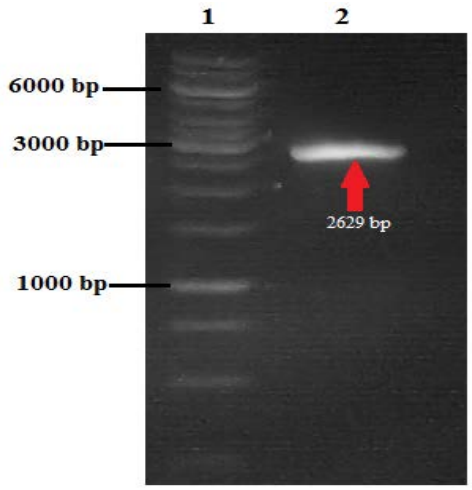


**Figure 1.** 1: DNA marker 2: PCR product

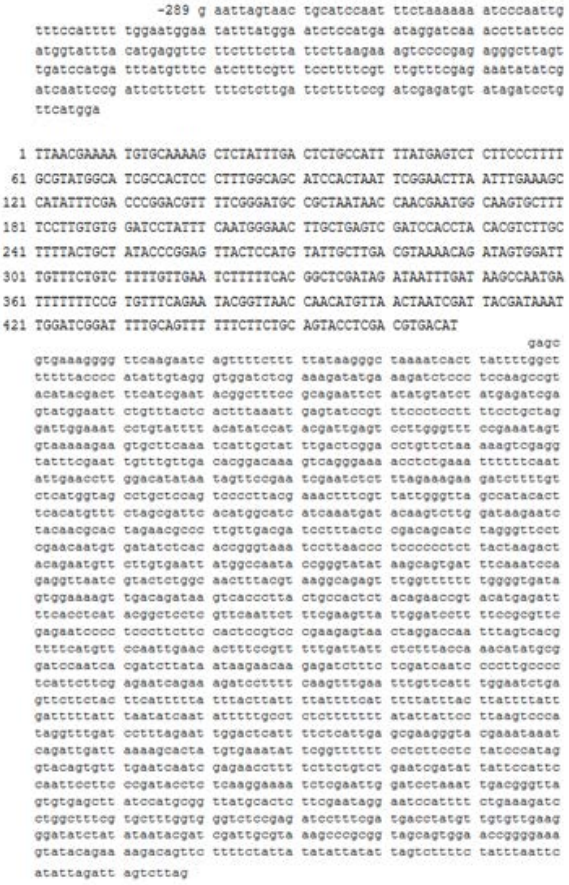

**Figure 2.** The nucleotide sequence of the *rps7* gene. From -1 to -289 is Promoter region. Exon is 468 bp long beginning with nucleotide 1, extending to nucleotide 468

An inserted fragment in the vector *p*TG19-T was further corroborated by PCR and digestion by restriction enzymes of *Eco*RI (Figure 3).

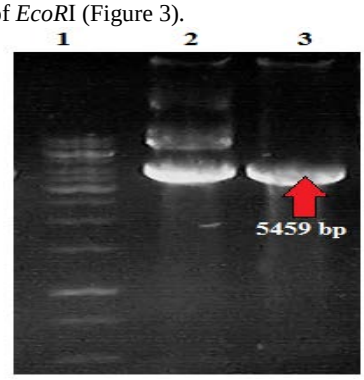


**Figure 3.** Gel agarose description of digested fragments. The extracted plasmids were digested by enzyme *Eco*RI 1: DNA marker 2: uncut 3: cut

### *3.1. Alignment of Amino Acid of rps7*

DNA-to-protein translation of DNA sequences was performed using the EXPASY translate tool using the standard genetic code. The deduced amino acid sequence of the lettuce *rps7* gene was compared with seven other *rps7* genes (Figure 4). The amino acid sequence showed a high level of identity with the genes of *Lasthenia burkei*, *Helianthus hirsutus* and *Artemisia montana* (98.1%), *Jacobaea vulgaris* (96.8%), *Elaeis oleifera* (95.5%). There was 50% identity with the genes of *Spinacia oleracea*, *Solanum lycopersicum* and *Lotus japonicas*. The identity dropped to 56.9%, 56.1%, 54.8% and 54.8% with *Microcoleus sp*, *Raphidiopsis brookii*, *Nostoc sp* and *Anabaena variabilis*, respectively. Nevertheless, 74 amino acids were conserved in all eight of the considered *rps7* genes, but most of the amino acids of this protein are conserved in plants. These conserved amino acids are probably involved in the important biological functions of polypeptides.

The nucleotide sequence of *rps7* was also compared with other *rps7* genes. This sequence showed a high level of identity with the *rps7* gene of other plants especially with *Asteraceae* family. This observation is consistent with the view that the sequence of the plastid genome is more conserved than the same sequence of the nuclear genome. This may also allow the use of the *rps7* gene as a homologous recombination site of different plants for plastid transformation.

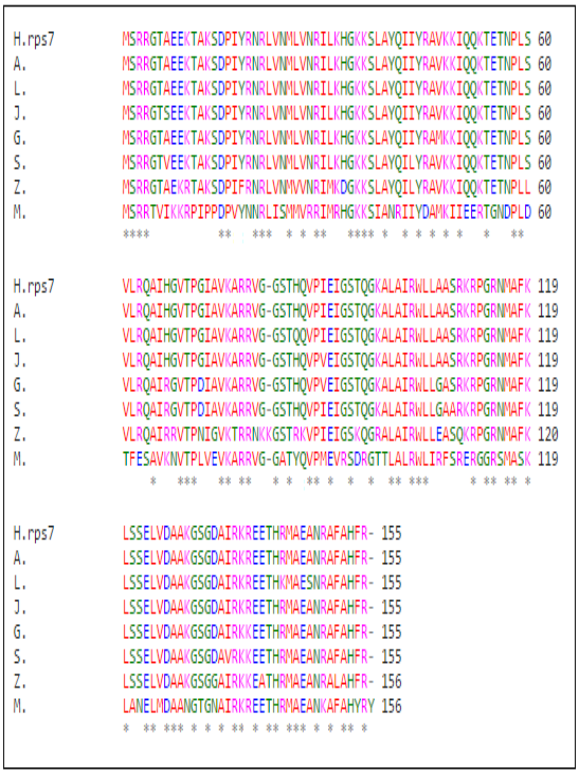

**Figure 4.** Comparison between the deduced amino acid sequence of *Lactuca sativa* plastid *rps7* (L.*rps7*) and the corresponding sequences from *Helianthus hirsutus* (H.*rps7*), *Artemisia montana* (A.*rps7*), *Jacobaea vulgaris* (J.*rps7*), *Glycine max* (G.*rps7*), *Spinacia oleracea* (S.*rps7*), *Zea maize* (Z.*rps7*) and *Microcoleus sp* (M.*rps7*) Identical amino acids are marked by asterisks in all 8 sequences

### *3.2. Sequence Analysis of rps7 Promoter*

To find regulatory elements in promoter sequences, the PLANTCARE software was applied. Analysis of the sequences of the promoter fragment revealed several kinds of specific promoter motifs in 289 bp of these cloned promoter fragments as shown in (Figure 5) and description of each one shown in Table 1.

The CAAT-box motif with common cis-acting element in the promoter and enhancer regions is often present at -80 to-150 bp upstream of TIS (Transcription Initiation Site) and may operate cooperatively with other putative conserved motifs (Gelinas, 1985). However, no unifying expression pattern of plant genes containing putative CCAAT elements (Rieping and Schoffl, 1992).

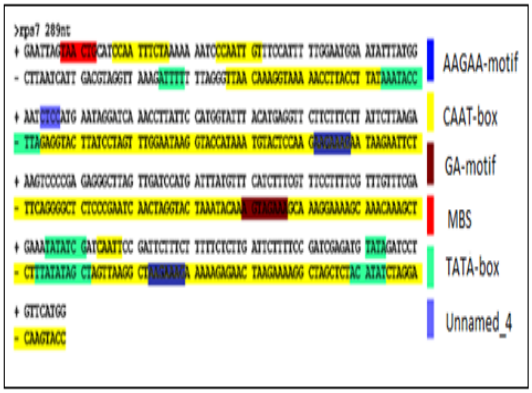


**Figure 5.** Sequence analysis of *rps7* promoter fragments

**Table 1.** Regulatory elements in *rps7* promoter sequence of lettuce

| Site Name | Position-Strand | Sequence | Function |
|---|---|---|---|
| CAAT-box | 38- | CAAT | Common cis-acting element in promoter and enhancer regions |
| | 224+,18+, 36+,37- | CAATT | |
| | 17+,35+ | CAAAT | |
| GA-motif | 180- | AAAGATGAA | Part of a light responsive element |
| MBS | 8+ | TAACTG | myB binding site involved in drought inducibility |
| TATA-box | 214+ | ATATAT | Core promoter element around -30 of transcription start |
| | 215+, | TATA | |
| | 271+ 213- | tcTATATAtt | |
| | 25- | TTTTA | |
| | 62- | tcTATAAAta | |
| | 269- | TATACA | |
| AAGAA-motif | 558- | GAAAGAA | |
| Unnamed_4 | 74+ | CTCC | |

Moreover, multiple copies of the genes coding for the subunits of the CCAAT, -binding protein exists in *Arabidopsis thaliana, Glycine max and Hordeum vulgare*, suggesting the potential for multiple alternative forms of these complexes in plants (Shirsat *et al*., 1989) and (Edwards *et al*., 1998). The first element described as regulating this process was a classical TATA box, TATA (A/T) A, located -25 to -30 base positions upstream of TIS (Davison *et al*., 1983). However, subsequent studies suggested that AT-rich sequences completely unrelated to the TATA-box stimulate transcription with equal or increased efficiency (Singer *et al*., 1990). Furthermore, although the first step of transcription initiation is highly specific, TFIID also binds with high affinity to several TATA elements that do not match the consensus sequence and is active in promoting transcription in vitro from these elements (Hahn *et al*., 1989). In eukaryotic promoters, between 10 and 20% of all genes (Gershenzon and Ioshikhes, 2005) contain a TATA box (sequence ATATAT, TATACA, TATAA, TTTTA, TATA, tcTATATAtt, tcTATAAAta), which provides for a TATA binding protein and assists the formation of the RNA polymerase transcriptional complex (Smale and Kadonaga, 2003). The TATA box typically lies very close to the transcription initiation site (often within 50 bases), and tends to be surrounded by GC rich sequences. The GA-motif is part of a light, responsive element that is similar to this motif in *Helianthus annuus* - motif (Eghtedary, 2014). MBS (myB binding site) has also been localized in the upstream promoter region that involved in drought and it is similar to this motif in *Arabidopsis thaliana* (Van Moerkercke *et al*., 2011). AAGAA-motif element is present in 19.23% of the genes. This element contains sequences GAAAGAA and GTAAAGAAA, and its function is unknown (Karimzadeh *et al*., 2013). We suggest designing the 5'UTR in genetic engineering and gene transfer issues to make the plants more stress tolerant, use above cis-elements especially CAAT-box, TATA-box and MBS.

### *3.3. Domain Detection*

Domain detection was done for finding all known motifs that occur in a sequence using the EXPASY Motif Scan tool. As a result, the domain of 30S ribosomal protein S7 (1-155) whit N-score=30.261 and E-value= 1.2e-23 shown in (Figure 6).

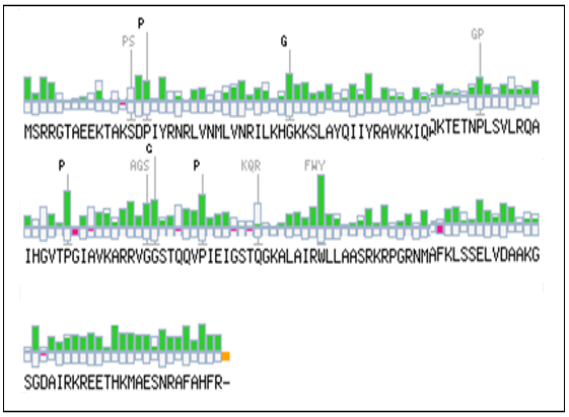


**Figure 6.** The green parts of the figure represent the amino acid of domains is in its proper place and red parts show the amino acid is not in its proper place compared with a protein domain which can be compared

### *3.4. Relationship Analyses of the rps7 Gene*

The phylogenetic analysis was conducted by NCBI to investigate the closely related species to L.*sativa*. The resulted phylogenetic tree was divided into two clusters. The first cluster contains species of monocots and the second cluster includes species of dicotyledons. According to the results of the phylogenetic tree, the *rps7* gene sequence of dicotyledonous species is close to *rps7* gene sequence of lettuce varieties (Figure 7).

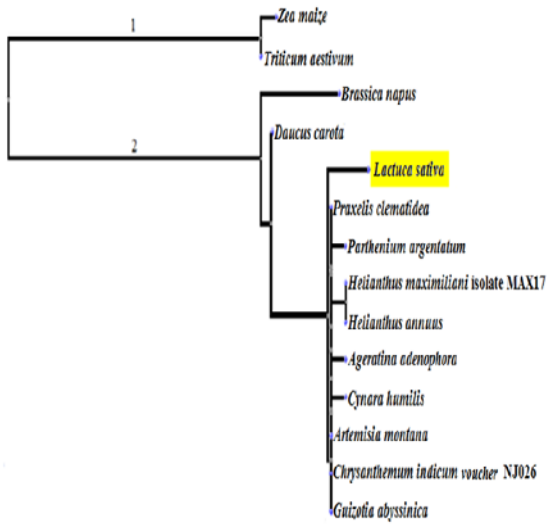


**Figure 7.** Phylogenetic relationship of *rps7* from *Lactuca sativa* and the other close species

## 4. Conclusion

The aim of the present work was to clone and characterize *rps7* gene from *Lactuca sativa* plastid which has not been studied so far in lettuce. The PCR fragment was cloned into *p*TG19-T cloning vector. The fragment was validated by *Eco*RI restriction digestion analysis and verified by sequencing. The alignment analysis results were analyzed showed the high homology with identity of 96% compared with *rps7* gene from lettuce. The genetic and phylogenetic studies of this gene can be used to identify varieties and in the conservation of breeding resources. Success in breeding is determined by genetic compatibility and chloroplast genomes serve as a valuable tool for identifying plants that are likely to be closely related and, therefore, genetically compatible. On the other hand, the *rps7* gene can be used as a secure homologous recombination site for plastid transformation with high-efficiency for lettuce cultivars and the other close species as well.

## Acknowledgements

The authors would like to express their gratitude to the University of Tabriz, Iran for their supports.

## References

Daniell H, Lin Ch.S, Yu M and Chang W.J. 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biology*, 17:134

Davison B.L, Egly J.M, Mulvihill ER and Chambon P P. 1983. Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences. *Nature*, **301:** 680-686.

Doyle J.J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue.

*Phytochemical Bulletin*, **19:**11–15.

Edwards D, Murray J.A.H and Smith A.G. 1998. Multiple Genes Encoding the Conserved CCAAT-Box Transcription Factor Complex Are Expressed in Arabidopsis. *Plant Phys*, **117:** 1015-1022.

Eghtedary Naeini A, Etedali F, Mortazavi S.E, Sobhani Najafabadi A, Golabadi M, Gharyazi B and M.A Malboobi. 2014. Structures of Phosphoenolpyruvate carboxylase (PEPC) Gene Promoter from C4 and C3 Flaveria species Using Sequence Analysis by Bioinformatics Tools. *Annual Res & Review in Biology*, **4:** 2779-2794.

Gelinas R, Endlich B, Pfeiffer C, Yagi M and Stamatoyannopoulos G. 1985. G to A substitution in the distal CCAAT box of the A2-globin gene in Greek hereditary persistence of fetal hemoglobin. *Nature*, **313:** 323- 325.

Gershenzon NI and Ioshikhes IP. 2005. Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics*, **21:** 1295-1300.

Hahn S, Buratowski S, Sharp P.A and Guarente L. 1989. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. Proc. Natl. Acad.Sci. *U.S.A*, **86:** 5718-5722.

Hinsinger DD, Strijk JS. 2015. Complete chloroplast genome sequence of Castanopsis concinna (Fagaceae), a threatened species from Hong Kong and South-Eastern China. *Mitochondrial DNA*, **18:**1–2.

Ignatovich O, Cooper M, Kulesza HM and Beggs JD. 1995. Cloning and characterization of the gene encoding the ribosomal protein S5 (also known as rp14, S2, YS8) of Saccharomyces cerevisiae. *Nucleic Acids Res*, **23:** 4616-4619.

Jansen RK, Cai Z, Raubeson LA, Daniell H, Leebens-Mack J, Müller KF, *et al*. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc Natl Acad Sci U S A*, **104:**19369–74.

Karimzadeh M, Nematzadeh G and Hashemi S.H. 2013. Study of 5'UTR's cis-elements region of some Plants'P5CS gene. *IJACS Journal*, **6-11:** 623-626.

Klussmann S, Franke P, Bergmann U, Kostka S and Wittmann-Liebold B. 1993. N-terminal modification and amino-acid sequence of the ribosomal protein HmaS7 from Haloarcula marismortui and homology studies to other ribosomal proteins. *Biol Chem Hoppe Seyler*, **374:** 305-312.

Kress WJ, Erickson DL. 2007. A two-locus global DNA barcode for land plants: the coding *rbcL*

gene complements the non-coding trnH-psbA spacer region. *PLoS ONE*, **2 (6):** e508.

Marín-Navarrol J, Manuell A.L, Wu J and Mayfield S.P. 2007. Chloroplast translation regulation. *Photosynthesis Research*, **94:** 359-374.

Raubeson L.A and Jansen R.K. 2005. Chloroplast genome of plants. In: Henry R.J, editor, Diversity and Evolution of Plants-Genotypic and Phenotypic Variation in Higher Plants. *CABI Publishers, Cambridge*, MA**:** 45-68.

Rieping M and Schoffl F. 1992. Common sequence found in the 5'-non-coding regions of eukaryotic genes; "CCAAT box" found in the promoter of heat shock protein genes; Located immediately upstream from the most distal HSE of the promoter; "CCAAT box" act cooperatively with HSEs to increase the hs promoter activity. *Mol. Gen. Genet*, **231:** 226- 232.

Robert. F and Brakier-Gingras L. 2001. Ribosomal protein S7 from Escherichia coli uses the same determinants to bind 16S ribosomal RNA and its messenger RNA. *Nucleic Acids Res*, **29:** 677-682.

Shirsat A, Wilford N, Croy R and Boulter D. 1989. Sequences responsible for the tissue specific promoter activity of a pea legumin gene in Tobacco. *Mol Gen Genet*, **215:** 326-331.

Singer V.L, Wobbe C.R and Struhl K K. 1990. A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. *Genes Dev*, **4:** 636-645.

Smale T and Kadonaga T. 2003. The RNA polymerase II core promoter. *Annual Review of Biochemistry*, **72:** 449-479.

Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, Willerslev E. 2007. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, **35 (3):** e14.

Timme R.E, Kuehl J.V, Boor J.L and Jansen R.K. 2007. A comparative analysis of the *Lactuca* and *helianthus* (asteracea) plastids genomes identification of divergent regions and categorization of shared repeats. American Journal of Botany, **94:** 302-312.

Van Moerkercke A1, Haring MA and Schuurink RC. 2011. The transcription factor EMISSION OF BENZENOIDS II activates the MYB ODORANT1 promoter at a MYB binding site specific for fragrant petunias. *Plant J*, **67:** 917-28.

Wimberly BT, White SW and Ramakrishnan V. 1997. The structure of ribosomal protein S7 at 1.9 A resolution reveals a beta-hairpin motif that binds double-stranded nucleic acids. *Structure*, **5:** 1187-1198.

Yang J.B, Li D.Z and Li H.T. 2014. Highly effective sequencing whole chloroplast genomes of

angiosperms by nine novel universal primer pairs. *Molecular Ecology Resources*, **14 (5):** 1024–1031.

Zhang Q.J, Gao L.Z. 2016. The complete chloroplast genome sequence of desert poplar (Populus euphratica). *Mitochondrial DNA Part A*, **27 (1):** 721–723.