Journal of Plant Molecular Breeding (JPMB)
Vol. 2/No. 1/June 2014/ 12-20

# Codon bias patterns in photosynthetic genes of halophytic grass *Aeluropus littoralis*

**E. Younesi-Melerdi[1], Gh. Nematzadeh[2*], E. Shokri[3]**

1. M.Sc in Plant Biotechnology, University of Agriculture Sciences and Natural Resources, Sari, Iran.
2. Prof. in Plant Genetics, University of Agriculture Sciences and Natural Resources, Sari, Iran.
3. Faculty member of Genetic & Agricultural Biotechnology Institute of Tabarestan (GABIT), Sari, Iran.

*Corresponding Author, Email: gh.nematzadeh@sanru.ac.ir

Received: May 2014 Accepted: June 2014

**Abstract:**
Codon bias refers to the differences in the frequency of occurrence of synonymous codons in coding DNA. Pattern of codon and optimum codon utilization is significantly different between the lives. This difference is due to the long term function of natural selection and evolution process. Genetics drift, mutation and regulation of gene expression are the main reasons for codon bias. In this study, the codon bias analysis was done on photosynthesis and respiratory related genes of phosphoenolpyruvate carboxylase (PEPC), NADP-malic enzyme (NADP-ME), pyruvate orthophosphate dikinase (PPDK), glycerate kinase (GK) (nuclear genes), rubisco, NADH-dehydrogenase subunit F and cytochrome-C (chloroplast genes) from *Aeluropus littoralis* plant. Nuclear gene sequences were obtained after partial isolation and for chloroplast genes obtained from nucleotide database. Calculation of codon adaptation index (CAI) showed that studied genes with direct or indirect association with photosynthesis, had high level of gene expression and had also a tendency to optimum codon utilization. The results also showed the difference in codon bias between genes encoded in nucleus and chloroplast for some amino acids.
**Key words:** Codon bias, photosynthesis genes, codon adaptation index, *Aeluropus littoralis*.

## Introduction

Genetic code (codon) is the combination of three consecutive nucleotide sequence of a gene that codes for a specific amino acid. From 20 amino acids in the genetic code, nine are coded by two synonymous codons, one is coded by three (as is the stop signal), five are coded by four, three are coded by six, and only two amino acids are coded by one codon (Salim and Cavalcanti, 2008). Some amino acids are encoded by more than one codon. Because of such redundancy it is said that the genetic code is degenerate. Synonymous codons are genomic control method for reducing and counteracting the effects of point

mutations and controlling gene expression. Synonymous codons, in the genomes of different species and among genes of a genome can be used with different frequency, which refers to the phenomenon of codon bias (Hershberg and Petrov, 2008). On the other hand, codon bias or codon usage bias is the probability of a codon being used for coding an amino acid over a different codon which codes for the same amino acid. Pattern of codon and optimized codon usage differs significantly between the genomes of organisms. The dispute stems from the long-term functioning of the processes of natural selection and development (Palidwor *et al*., 2010; Sharp *et al*., 2010). At the same time, choosing the frequent and rare codons is generally consistent across genes within each genome (Chen *et al*., 2004). Genetic drift, mutation and regulation of gene expression are the main reasons of biased codon (Palidwor *et al*., 2010; Sharp *et al*., 2010). Strength of codon bias also varies among organisms. In some organisms codon bias is very strong, while in some others different synonymous codons with similar frequencies are used (Andersson and Sharp, 1996; Duret, 2002; Sharp *et al*., 2005). Likewise, the strength of codon bias varies across genes within each genome, with some genes using a highly biased set of codons and with others using different synonymous codons with similar frequencies (Gouy and Gautier, 1982; Ikemura, 1985; Sharp *et al*., 1988).

Codon bias is more prevalent in highly expressed genes. In these genes some certain codons are preferred more than other. For example, in a group of high expressed protein in yeast, over 96% of amino acids are coded by only 25 genetic codons out of the 61 avalable ones. Results have shown that replacing the optimal codon with minimal codon significantly reduce the translation rate of these genes (Harrison and Charlesworth, 2011). Codon bias not only is limited to the coding regions of the genome, but it is also observable in adjacent intron region of genes. From this perspective, the stability of codon composition and mRNA molecules affect the level of transcription. There is a positive correlation between codon bias and tRNA gene copy number and expression level (cellular tRNA pool), which can interfere in the process of natural selection for more and better efficiency in translation of genes. In addition, several other factors such as the level of amino acids conservation, the amount and duration of protein hydrophobicity, GC content, and the optimum temperature for growth and adaptation have been reported in association with codon bias (Rao *et al*., 2011; Paul *et al*., 2008). In the present study, after partial gene (coding sequence) isolation by polymerase chain reaction (PCR) and sequencing, codon bias of phosphoenolpyruvate carboxylase (PEPC), NADP-malic enzyme (NADP-ME), pyruvate orthophosphate dikinase (PPDK) glycerate kinase (GK) (from nucleus), rubisco, NADH-dehydrogenase subunit F and cytochrome-C genes (from chloroplast) which were (directly/indirectly) related to photosynthesis and respiration were

investigated in order to the recognize codon usage pattern in *Aeluropus littoralis*. *Aeluropus littoralis* is a grass from a wild relative of wheat with C4 photosynthetic system which has high tolerance to high salt concentrations. Study of codon bias in *Aeluropus littoralis* can be important from halophyte view and association of codon bias with salt tolerance.

## Materials and methods

### *Isolating and sequencing the photosynthesis genes of C4 pathway*

After total RNA extraction from plant leaves, the complementary DNA (cDNA) was prepared using oligo dt primers and reverse transcriptase enzyme (Fermentas, Thermo Scientific). After aligning the ortholog gene sequences in Gramineae species and other plants, specific primer pairs were designed for partial isolation of PEPC, NADP-ME, PPDK and GK coding sequence from *Aeluropus littoralis* plant. Sequence aligning and primer designing were done using BioEdit version 7.2.9 (Tom Hall Ibis Biosciences) and Oligo 5 (Molecular Biology Insights, Inc.) softwares, respectively. The PCR products were purified from agarose gel (Agarose Gel DNA Extraction Kit- Roche) and sequenced (Bioneer Inc.). The sequences of genes encoded by chloroplast were obtained from database (GenBank: JN681717, EF125095 and JQ345047).

### *Analyzing codon bias and calculating CAI*

After determining open reading frame and homology searches associated with the desired gene sequences in the nucleotide and protein database of NCBI, the codon bias index (CAI) were analyzed using CAI calculator tool (*http://genomes.urv.es/CAIcal/E-CAI/*) according to standard codon usage tables of *Aeluropus littoralis* (*http://www.kazusa.or.jp/codon/cgi-bin/spsearch.cgi?species=Aeluropus+littoralis+&c=s*).

## Results and Discussion

Polymerase chain reaction of degenerate primers and cDNA samples from *Aeluropus littoralis* as template, resulted in amplified bands in expected sizes of 855, 625, 503 and 467 bp for PEPC, NADP-ME, PPDK and GK, respectively (Fig. 1).

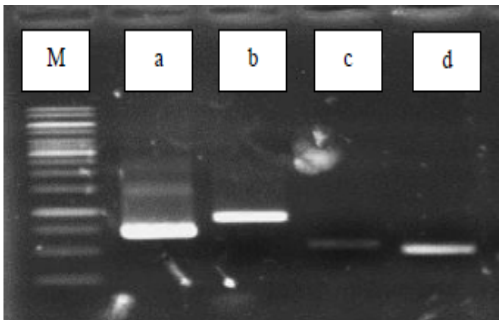


**Fig. 1.** PCR products of photosynthetic genes of *Aeluropus littoralis*. M: DNA molecular marker (SM0311, Fermentas), a, b, c and d: Amplified fragments for NADP-ME, PEPC, PPDK and GK genes.

After sequencing and confirming the open reading frame, the ORF sequence was translated to protein and then analyzed using the BLASTP in NCBI database. All isolated genes showed significant homology (over 90%) with other genes in the orthologous protein

levels (results are not presented). Determinig the nucleotide composition of the genes showed significant differences in terms of GC (G + C %) content between nuclear and chloroplast genes. Accordingly, PEPC gene with 66% cytosine and guanine nucleotides was identified as the GC-richest gene. Furthermore, the average nucleotide composition (A + T %) was 63% for the chloroplast genes which was higher than nuclear genes (45%) (Fig. 2). The finding was in concordance with previous findings in which GC content of nuclear genes was more than mitochondrial and chloroplast genes (Shanker, 2012).

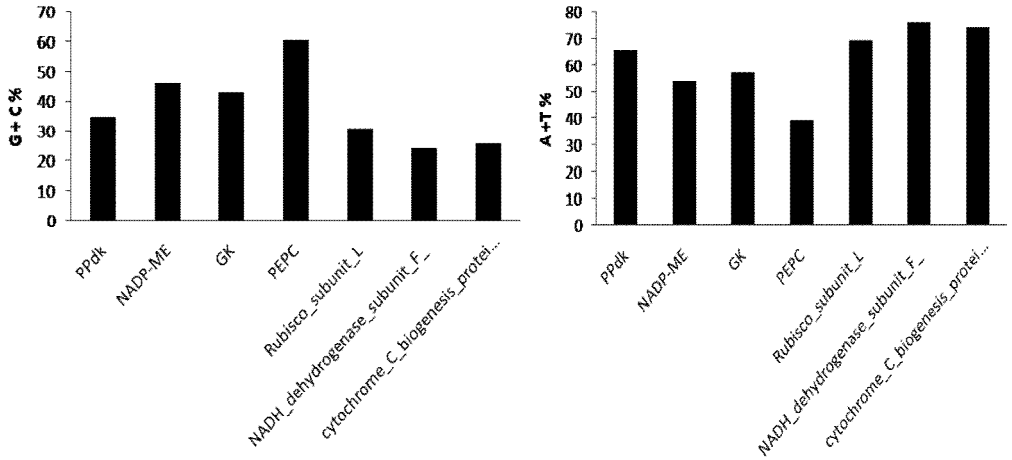


**Fig. 2.** Codon bias comparison in degenerate region (third nucleotide) of genetic code according to use of G+C (left) and A+T (right) nucleotides in studied genes.

Codon bias can be assessed in two classes of selection explanation and mutation explanation. The former explains the contribution of codon bias to an effective and accurate protein translation; therefore, codon bias has been chosen and maintained by selection. The latter explains the existence of codon bias due to non-random mutation patterns. Different studies showed that the GC content is the most important factor determining codon bias between organisms (Chen *et al.*, 2004; Kanaya *et al.*, 2001; Knight *et al.*, 2001). It is generally acknowledged that codon preferences reflect a balance between mutational biases and natural selection for translational optimization.

In this study, the comparison of nucleotide composition at the third position of the genetic code (degenerate site) in studied loci showed a strong tendency for chloroplast genesto use A or T nucleotides, while the nucleus genes prefer G and C nucleotides. This can be due to high A+T content in chloroplast genes comparing with nuclear genes. Furthermore, the results showed no significant difference between nuclear and chloroplast genes at the second position of genetic codeHowever, as the third position,

there was a heavy bias between the nucleus and the chloroplast photosynthesis gene in the selection of G and C nucleotides for the first position in the genetic code (Fig. 3). Interestingly, the use of high G and C nucleotides (91%) in the first position of the genetic code was seen in PEPC gene which was associated with a high GC content of this gene. Sharp *et al*. (1995) found that GC content were correlated with codon usage bias. Proposed hypotheses for this correlation included the ideas that nucleotide patterns might be determined by selection, mutational bias or recombination, since there was an association between recombination and GC-rich chromosomal regions (Salim and Cavalcanti, 2008).

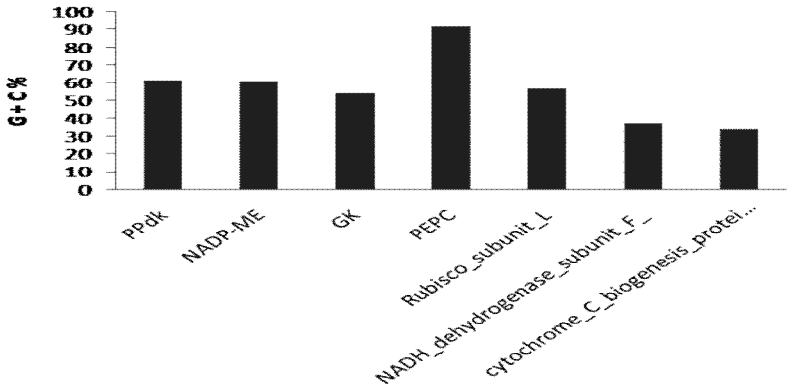


**Fig. 3.** Codon bias according to G+C% in first position of genetic code in studied genes.

Calculating CAI indicated more than 60% codon bias in all studied genes This indicates the high levels of gene expression and requires the use of optimal codons of these genes (Table 1). A strong correlation has been determined between codon bias and gene expression level. This finding has been shown using large-scale gene expression data, in organisms as diverse as *E. coli*, *S. cerevisiae, C. elegans, Arabidopsis thaliana,* and *D. melanogaster* (Castillo-Davis and Hartle, 2002; Duret and Mouchiroud, 1999; Ghaemmaghami *et al*., 2003; Goetz and Fuglsang, 2005). In the present study, except for the amino acids of glutamate and histidin in which no codon bias is seen, there is a tendency to use one or more certain codons in the rest of the amino acids. There were also differences of codon bias between the nucleus and the chloroplast photosynthetic genes for the amino acids of arginine, aspargine, glutamine, cysteine, glycine, lysine, phenylalanine and tyrosine.

**Table 1.** CAI in photosynthetic genes of *Aeluropus littoralis*.

| Gene | Source | Length (bp[a]) | CAI[b] |
|---|---|---|---|
| PEPC[c] | Nucleus | 837 | 0.752 |
| NADP-ME[d] | Nucleus | 300 | 0.746 |
| PPDK[e] | Nucleus | 486 | 0.661 |
| GK[f] | Nucleus | 393 | 0.752 |
| Rubisco | Chloroplast | 1347 | 0.724 |
| NADH-dehydrogenase subunit F | Chloroplast | 2064 | 0.790 |
| cytochrome-C | Chloroplast | 858 | 0.769 |

a:base pair, b: codon adaptation index, c: phosphoenolpyruvate carboxylase, d: NADP-malic enzyme, e: pyruvate orthophosphate dikinase, f: glycerate kinase.

**Table 2.** Similarity in codon bias between nucleus and chloroplast genes.

| Amino acid | CB[a] (nucleus and chloroplast) | Amino acid | CB (nucleus and chloroplast) |
|---|---|---|---|
| Alanine | GCT | leuscine | CTT |
| Aspartic acid | GAT | proline | CCA,CCT |
| Glutamic acis | Non-bias | serine | TCA |
| Histidine | Non-bias | threonin | ACT,ACA,ACC |
| Isoleucine | ATT | Valine | GTT |

a: codon bia

**Table 3.** Differences in codon bias between nucleus and chloroplast genes.

| Amino acid | Nucleus | Chloroplast |
|---|---|---|
| Argenine | Non-bias | AGA,AGG,CGA |
| Aspargine | Non-bias | AAT |
| Cysteine | TGT | Non-bias |
| Glutamine | CAG | Non-bias |
| Glysine | GGA,GGC,GGT | GGA,GGT |
| Lysine | Non-bias | AAA |
| Phenylalanine | Non-bias | TTT |
| Tyrosine | TAC | TAT |

The patterns of similarities and differences in the codon usage were shown in Tables 2 and 3. having analyzed 558 genes, Ingvarsson (2007) showed direct but difference effects of gene expression on both codon usage and the level of selective constraint of proteins in *Populus tremula*. Camiolo *et al.* (2012) also reported systematic differences in the usage of synonymous codons among Arabidopsis thaliana genes that were expressed specifically in distinct tissues. Their analysis showed that in some cases, codon usage

in genes that were expressed in a broad range of tissues was influenced primarily by the tissue in which the gene was expressed maximally. On the basis of their finding, it was proposed that genes that were expressed in certain tissues might show a tissue-specific compositional signature in relation to codon usage.

These findings might have implications for the design of transgenes in relation to optimizing their expression. Barozai *et al*. (2012) studied the relation between synonymous codon usage and salt tolerance in five salt resistant and three housekeeping genes in *Arabidopsis thaliana* and *Oryza sativa*. From It was concluded that there was a straight correlation between codon usage bias and salt stress among the studied species. They suggested that plant salt stress resistance could be improved by optimizing the codon usage. Association between codon usage and salt tolerant would help us to engineer the salt resistant crop by adjusting the codon usage.

## Conclusion

The codon bias analysis was done on some photosynthesis and respiration related genes of nucleus and chloroplast of *Aeluropus littoralis*. As expected, the GC content was significantly different tha analysis of codon bias in some photosynthesis and respiratory related genes of nucleus and chloroplast of Aeluropus littoralis revealed, as it was expected, a significant difference in GC content between nucleus and chloroplast genes the nucleus originated genes showed more GC content than chloroplast genes. The CIA was also more than 60% for all studied genes. Obtained results showed a difference in the codon bias between genes encoded in nucleus and chloroplast for some amino acids. The results can be used in future studies on the association of codon bias with salt tolerance.

## References

Andersson, S. G. and Sharp, P. M. 1996. Codon usage and base composition in *Rickettsia prowazekii*. *J. Mol. Evo,* 42: 525–36.

Barozai, M.Y. K., Kakar, A. G. and Din, M. 2012. The relationship between codon usage bias and salt resistant genes in Arabidopsis thaliana and Oryza sativa. Pure Appl. Bio, 1 (2): 48-51.

Camiolo, S., Farina, L. and Porceddu, A. 2012. The Relation of Codon Bias to Tissue-Specific Gene Expression in Arabidopsis thaliana. *Genetics, 192: 641–649.*

Castillo-Davis, C. I. and Hartl, D. L. 2002. Genome evolution and developmental constraint in *Caenorhabditis elegans*. *Mol. Biol. Evol,* 19: 728–35.

Chen, SL., Lee, W., Hottes, A. K., Shapiro, L. and McAdams, H. H. 2004. Codon usage between genomes is constrained by genome-wide mutational processes. *Proc. Natl. Acad. Sci. USA,* 101: 3480–85.

Duret, L. and Mouchiroud, D. 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA,* 96: 4482–87.

Duret, L. 2002. Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev,* 12: 640–49.

Harrison, R.J. and Charlesworth, B. 2011. Biased gene conversion affects patterns

of codon usage and amino acid usage in the Saccharomyces sensu stricto group of yeasts. *Molecular biology and evolution,* 28: 117-129.

Hershberg, R. and Petrov, D.A. 2008. Selection on codon bias. Annu. Rev. Genet, 42: 287–99.

Ikemura, T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol,* 2: 13–34.

Ingvarsson, P.K. 2007. Gene Expression and Protein Length Influence Sequence Evolution in Populus tremula. *Mol. Biol. Evol,* 24 (3): 836–844.

Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W. and Belle, A. 2003. Global analysis of protein expression in yeast. *Nature,* 425: 737–41.

Gouy, M. and Gautier, C. 1982. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res,* 10: 7055–7074.

Goetz, R. M. and Fuglsang, A. 2005. Correlation of codon bias measures with mRNA levels: analysis of transcriptome data from *Escherichia coli*. *Biochem. Biophys. Res. Commun,* 327: 4–7.

Kanaya, S., Kinouchi, M., Abe, T., Kudo, Y. and Yamada, Y. 2001. Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the *E. coli* O157 genome. *Gene,* 276: 89–99.

Knight, R. D., Freeland, S. J. and Landweber, L. F. 2001. A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biol,* 2 (4): RESEARCH0010. doi:10.1186/gb-2001-2-4-research0010.

Palidwor, G.A., Perkins, T.J. and Xia, X. 2010. A general model of codon bias due to GC mutational bias. *PLoS One, 5 (*10): e13431.

Paul, S., Bag, SK., Da,s S., Harvill, E. T. and Dutta, C. 2008. Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. *Genome Biol*, 9 (4): R70. doi:10.1186/gb-2008-9-4-r70.

Rao, Y., Wu, G., Wang, Z., Chai, X., Nie, Q. and Zhang, X. 2011. Mutation Bias is the Driving Force of Codon Usage in the Gallus gallus genome. *DNA research,* 18: 499-512.

Salim, H. M. W. and Cavalcanti, A. R. O. 2008. Factors Influencing Codon Usage Bias in Genom. J. Braz. Chem. Soc, 19 (2): 257-262.

Shanker, A. 2012. Chloroplast Genomes of Bryophytes: A Review. Archive for Bryology, 143: 1-5.

Sharp, P. M., Averof, M., Lloyd, A. T., Matassi, G. and Peden, J. F. 1995. DNA sequence evolution: the sound of silence. *Philos. Trans. R. Soc. Lond. B, Biol. Sci, 349*: 241-247.

Sharp, P.M., Cowe, E., Higgins, D.G., Shields, D.C., Wolfe, K.H. and Wright, F. 1988. Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. *Nucleic Acids Res,* 16: 8207–8211.

Sharp, P. M., Bailes, E., Grocock, R. J., Peden, J. F. and Sockett, R. E. 2005. Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res,* 33: 1141–1153.

Sharp, P.M., Emery, L.R. and Zeng, K. 2010. Forces that influence the evolution of codon bias. *Phil. Trans. R. Soc. B,* 365: 1203-1212.

# الگوهای تمایل کدونی در ژن‌های فتوسنتزی علف شورزی، آلوروپوس لیتورالیس

**الهام یونسی ملردی[۱]، قربانعلی نعمت‌زاده[۲*]، احسان شکری[۳]**

۱. دانشجوی کارشناسی ارشد، دانشگاه علوم کشاورزی و منابع طبیعی ساری

۲. استاد ژنتیک گیاهی، دانشگاه علوم کشاورزی و منابع طبیعی ساری

۳. عضو هیئت علمی پژوهشکده ژنتیک و زیست فناوری کشاورزی طبرستان

[*]نویسنده مسئول: gh.nematzadeh@sanru.ac.ir

**چکیده:**

تمایل کدنی به تفاوت در فراوانی وقوع کدون‌های مترادف در DNA کد شونده اطلاق می‌شود. الگوی استفاده از کدون و کدون‌های بهینه، به طور معنی‌داری بین موجودات متفاوت است. این اختلاف ناشی از کارکرد بلند مدت فرآیندهای انتخاب طبیعی و تکامل می‌باشد. رانش ژنتیکی، موتاسیون و تنظیم بیان ژن از دلایل اصلی وجود اریب کدونی است. در این مطالعه، تجزیه و تحلیل تمایل کدنی روی ژن‌های فسفو انول پیروات کربوکسیلاز (PEPC)، NADP- مالیک آنزیم (NADP-ME)، پیرووات ارتوفسفات دی‌کیناز (PEPDK) و گلیسرات کیناز (GK) (ژن‌های هسته‌ای)، روبیسکو، NADH دهیدروژناز زیر واحد F و سیتوکروم C مرتبط با فتوسنتز و تنفس از گیاه آلوروپوس لیتورالیس انجام شد. توالی ژن‌های هسته‌ای بعد از جداسازی جزیی و برای ژن‌های کلروپلاست از بانک داده‌های نوکلئوتیدی به دست آمدند. محاسبه شاخص سازگاری کدن (CAI) نشان داد که ژن‌های مورد مطالعه که به طور مستقیم یا غیر مستقیم با فتوسنتز در ارتباط هستند، از سطح بیان بالایی برخوردار بوده و همچنین تمایل به استفاده از کدون بهینه در آن‌ها وجود دارد. همچنین نتایج بدست آمده نشان دادند که تمایل کدونی بین ژن‌های کد شونده در هسته و کلروپلاست برای برخی از اسیدهای آمینه متفاوت است.

**کلمات کلیدی:** تمایل کدونی، ژن‌های فتوسنتز، شاخص سازگاری کدون، آلوروپوس لیتورالیس.