

Contents lists available at SciVerse ScienceDirect

Journal of Theoretical Biology



journal homepage: www.elsevier.com/locate/yjtbi

Analysis of codon use features of stearoyl-acyl carrier protein desaturase gene in *Camellia sinensis*



Lu-Lu Pan, Yu Wang, Jian-Hui Hu, Zhao-Tang Ding*, Chen Li

Tea Research Institute, Qingdao Agricultural University, Changcheng Road 700#, Chengyang District, Qingdao, Shandong 266109, China

HIGHLIGHTS

• The CsSAD gene has similar codon usage bias with C. sinensis genome genes.

• The CsSAD gene is biased toward codons ending with A/T.

• The E. coli expression system is superior to yeast expression system for the CsSAD gene.

• Clustering model was validated to be the optimal one with the jackknife test method.

ARTICLE INFO

Article history: Received 27 March 2013 Received in revised form 3 June 2013 Accepted 6 June 2013 Available online 14 June 2013

Keywords: Camellia sinensis SAD gene Codon usage bias Clustering analysis

ABSTRACT

The stearoyl-acyl carrier protein desaturase (*SAD*) gene widely exists in all kinds of plants. In this paper, the *Camellia sinensis SAD* gene (*CsSAD*) sequence was firstly analyzed by Codon W, CHIPS, and CUSP programs online, and then compared with genomes of the tea plant, other species and *SAD* genes from 11 plant species. The results show that the *CsSAD* gene and the selected 73 of *C. sinensis* genes have similar codon usage bias. The *CsSAD* gene has a bias toward the synonymous codons with A and T at the third codon position, the same as the 73 of *C. sinensis* genes. Compared with monocotyledons such as *Triticum aestivum* and *Zea mays*, the differences in codon usage frequency between the *CsSAD* gene and dicotyledons such as *Arabidopsis thaliana* and *Nicotiana tobacum* are less. Therefore, *A. thaliana* and *N. tobacum* expression systems may be more suitable for the expression of the *CsSAD* gene. The analysis result of *SAD* genes from 12 plant species also shows that most of the *SAD* genes are biased toward the synonymous codons with G and C at the third codon position. We believe that the codon usage bias analysis presented in this study will be essential for providing a theoretical basis for discussing the structure and function of the *CsSAD* gene.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

As we all know, nucleic acids are templates of protein synthesis and the genetic code chooses 61 codons to represent 20 standard amino acids. Each amino acid can be coded by 1–6 codons and these alternative codons for the same amino acid are termed as synonymous codons (Prabha et al., 2012). Research showed that genes prefer to use some codons in synonymous codon usage, namely existing as codon usage biased (Zhou et al., 2007). Codon usage is highly variable among different species and principally related to gene function (Ma et al., 2002; Fuglsang, 2003), and these variations are also found within the related species and genes (Ma et al., 2009). It indicates these genomes and genes bear different pressure in evolution process. Codon usage bias analysis has important significance in many aspects. It helps to reveal the genetic evolution law in some pertinent species or between gene families of a certain species (Sorimachi, 2009, 2010a, 2010b) and contributes to understanding the regulatory mechanism in the process of transcription and translation. It also can improve the gene expression level by predicting the optimum host of the exogenous gene and ameliorating the exogenous gene (Wu et al., 2007).

At present, there are two kinds of theories accounting for codon bias phenomenon, namely the Neutral theory and the "Selection–Mutation–Drift" theory (Bulmer, 1991). The Neutral theory holds that the mutations occurring in the third base position of a codon is the result of neutral selection, and synonymous mutations do not affect survival fitness; this kind of selection on codons only related to mutation, which are not affected by natural selection pressure. The "Selection–Mutation– Drift" theory considers that the occurrence of mutation has direction; synonymous codon usage bias reflects effects on both

^{*} Corresponding author. Tel.: +86 532 88030231.

E-mail addresses: panlulutea@163.com (L.-L. Pan), dzttea@163.com (Z.-T. Ding).

^{0022-5193/\$ -} see front matter @ 2013 Elsevier Ltd. All rights reserved. http://dx.doi.org/10.1016/j.jtbi.2013.06.006

selection of the optimal codon and mutation-drift of the nonsynonymous codon (Romero et al., 2000). It has been noted that codon usage is influenced by not only selection and mutation but also other factors (Zhou et al., 2010) such as base composition (Zhang et al., 2011), the level of gene expression (Hiraoka et al., 2009), tRNA abundance (Duret, 2000), gene length (Duret and Mouchiroud, 1999), mRNA secondary structure (Gu et al., 2003; Gu et al., 2004), protein hydrophobicity (Chou, 2001), amino acids conservatism and bonding strength between codons and anticodons (Liu et al., 2000). Most research has studied gene codon usage in species such as bacteria (Sharp et al., 1988), veast (Freire-Picos et al., 1994), Escherichia coli (Zhang and Chou, 1994), and Arabidopsis thaliana (Mathé et al., 1999), but little research has focused on species such as humans (Chou and Zhang, 1992; Chou and Zhang, 1993; Zhang and Chou, 1993), primates (McWeeney and Valdes, 1999) and xylophyta (Ingvarsson, 2008).

The stearoyl-acyl carrier protein (ACP) desaturase (SAD, EC 1.14. 99.6) is a key enzyme in fatty acid synthesis. It determines the ratio of saturated to unsaturated fatty acids in higher plants, and this ratio is closely associated with many functions in plants (Luo et al., 2006), particularly in plants acclimated to low temperatures (Byfield and Upchurch, 2007). Since codon usage of the SAD gene in C. sinensis has not been investigated in any detail, it is not clear how codons should vary in the SAD gene. In this paper, we studied SAD gene codon usage bias and compared it with other species, which is significant to choose an appropriate expression system and improve expression levels of the SAD gene in heterogenous expression systems. It also lays a foundation for transforming the CsSAD gene into a model plant to conduct functional verification. All the indices of codon usage bias, which are used in this study, are popular indices and extensively utilized in research on codon usage in many organisms (Liu et al., 2004; Wang and Hickey, 2007).

2. Materials and methods

2.1. Sequence data

The *C. sinensis SAD* gene used in this study was cloned by our laboratory, the length of cDNA of *SAD* is 1191 bp, coding for 396 amino acids, and its GenBank accession number is KC242133. *SAD* sequences of other 11 plants (*Glycine max:* L34346, *Arabidopsis thaliana:* AY128883, *Vernicia Montana:* EU072353, *Vernicia fordii:* GU363502, *Cucumis sativus:* M59858, *Ginkgo biloba:* HQ694561, *Helianthus annuus:* U70374, *Kosteletzkya virginica:* FJ750952, *Pinellia ternate:* JQ390410, *Brassica napus:* AY642537, and *Linum usitatissimum:* X70962) were obtained from GenBank (http://www.ncbi.nlm.nih.gov). Genomes sequences of six model species were obtained from Codon Usage Database (http://www. Kazusa.or.jp/codon), and a total of 73 publicly available *C. sinensis* cDNA sequences that contain complete coding sequences were screened from the GenBank nucleotide database randomly in October, 2012.

2.2. Indices of codon usage bias

In order to investigate the characteristics of synonymous codon usage of differing amino acid compositions in the gene sample, relative synonymous codon usage (RSCU) of 59 informative codons (excluding Met, Trp, and the three termination codons) was computed. The RSCU value was calculated by dividing the observed codon usage by that expected when all codons for the same amino acid are used equally (Liu et al., 2004; Sharp and Li, 1986). When the codon has a RSCU value close to 1.0, it means that this codon is chosen equally and randomly (Sau et al., 2006). The effective number of codons (ENC) is the best estimator of absolute synonymous codon usage bias, which is often used to measure the magnitude of codon bias for an individual gene (Wright, 1990). Values of ENC range from 20 to 61, while the larger the codon preference extent in a gene, the smaller the ENC. The GC3s value, which is the frequency of G+C at the third codon position, was computed as a good indicator of the extent of base composition bias (Sharp and Li, 1987).

2.3. Clustering analysis

According to a series of recent publications (Chou, 2011, 2013; Chen et al., 2012, 2013; Chou et al., 2012; Lin et al., 2013; Xiao et al., 2013; Xu et al., 2013), it is necessary to establish a really useful analysis model, simulation method, or statistical predictor for a biological system. We need to consider the following procedures: (i) construct or select a valid benchmark dataset to train and test the predictor; (ii) formulate the biological samples with an effective mathematical expression that can truly reflect their intrinsic correlation with the target to be analyzed or predicted; (iii) introduce or develop a powerful algorithm (or engine) to operate the analysis or prediction; (iv) properly perform cross-validation tests to objectively evaluate the anticipated accuracy of the model; and (v) establish a user-friendly webserver for the new method that is accessible to the public. Below, let us describe how to deal with these steps.

We used the Euclidean square distance to conduct clustering analysis based on RSCU values of 12 *SAD* genes after data standardization. In the clustering process, variables consisting of different RSCU values in a single mRNA sequence are regarded as points of multidimensional space (excluding Met, Trp, and the



Fig. 1. RSCU distribution of 59 codons in CsSAD and C. sinensis genes (+: C. sinensis SAD gene, +: C. sinensis genes).

Download English Version:

https://daneshyari.com/en/article/6370859

Download Persian Version:

https://daneshyari.com/article/6370859

Daneshyari.com