http://dx.doi.org/10.22037/afb.v4i3.14396

Research Article

APPLIED FOOD BIOTECHNOLOGY, 2017, 4 (3):167-178
Journal homepage: www.journals.sbmu.ac.ir/afb

pISSN: 2345-5357
eISSN: 2423-4214

Bioinformatics Evaluation of Plant Chlorophyllase, the Key Enzyme in Chlorophyll Degradation

Ebrahim Sharafi[1], Ali Dehestani[1], Jamshid Farmani[1,2*], Ali Pakdin Parizi[1]

1- Genetics and Agricultural Biotechnology Institute of Tabarestan, Sari Agricultural Sciences and Natural Resources University, Sari, Iran.

2- Department of Food Science and Technology, Faculty of Agricultural Engineering, Sari Agricultural Sciences and Natural Resources University, Sari, Iran.

Abstract

Background and Objective: Chlorophyllase catalyzes the hydrolysis of chlorophylls to chlorophyllide and phytol. Recently, several applications including removal of chlorophylls from vegetable oils, use in laundry detergents and production of chlorophyllides have been described for chlorophyllase. However, there is little information about the biochemical characteristics of chlorophyllases.

Material and Methods: 35 chlorophyllase protein sequences were obtained from the National Centre for Biotechnology Information database. All of the sequences were analyzed using bioinformatics tools for their conserved domain, phylogenetic relationships and biochemical characteristics.

Results and Conclusion: The overall domain architecture of chlorophyllases consisted of the esterases/lipases superfamily domain over their full length and the alpha/beta hydrolase family domain over the middle part of their sequences. Plant chlorophyllases could be classified into 4 clades. Molecular weight and pI of the chlorophyllases ranged 32.65-37.77 kDa and 4.80-8.97, respectively. The most stable chlorophyllase is probably obtained from *Malus domestica*. Chlorophyllases form *Solanum pennellii*, *Triticum aestivum*, *Triticum urartu*, *Arabidopsis lyrata*, *Pachira macrocarpa*, *Prunus mume* and *Malus domestica* were predicted to be soluble upon overexpression in *Escherichia coli*, *Beta vulgaris* and *Chenopodium album* chlorophyllases were predicted to form no disulfide bond. Chlorophyllases from *Jatropha curcas*, *Amborella trichopod*, *Setaria italica*, *Piper betle*, *Triticum urartu* and *Arabidopsis thaliana* were predicted to be in non-N-glycosylated form.

Conflict of interest: The authors declare no conflict of interest.

Article Information

Article history:

Received	11 Oct 2016
Revised	2 Apr 2017
Accepted	14 May 2017

Keywords:

- Bioinformatics
- Biochemical characteristics
- Chlorophyllase
- Phylogenetic

***Corresponding author:**

Jamshid Farmani,
Genetics and Agricultural Biotechnology Institute of Tabarestan, Sari Agricultural Sciences and Natural Resources University, P.O.Box 578, Sari, Iran.

Tel/Fax: +98-113-3687718
E-mail:
jamshid_farmani@yahoo.com

1. Introduction

Chlorophyll is a green spectrum pigment found in cyanobacteria, algae and plants and chlorophyllase or chlorophyll chlorophyllidohydrolase (EC 3.1.1.14) is the key enzyme in chlorophyll metabolism. It hydrolyzes chlorophyll to chlorophyllide and phytol. Chlorophyllases can be found in plants, algae or cyanobacteria. In plant physiology, it plays an important role in leaf senescence and fruit ripening [1]. The catalytic activity of chlorophyllase has potential industrial and agricultural applications, as well. In this context, the application of chlorophyllase could be useful in removal of green pigments from edible oils to improve their oxidative stability [2]. Other application of the enzyme is the use in laundry detergents [3]. In addition, the products of chlorophyllase, chlorophyllides and their derivatives (such as pheophytin) have been demonstrated to have antiviral, antioxidant, antimutagenic, and anticarcinogenic activities in vitro [4]. Accordingly, in recent years, trends for purification of natural chlorophyllases or recombinant production of the enzyme have been increased [4-6,9].

Nucleotide sequences of several plant chlorophyllases have been determined, and some of them are characterized biochemically. Recombinant production of some chlorophyllases including those from *Chenopodium* (*C.*) *album* (Lamb's Quarters) [7], *Triticum* (*T.*) *aestivum* (wheat) [3], *Pachira* (*P.*) *macrocarpa* (Guiana Chestnut) [8], *Cyanothece* spp. (cyanobacteria) [4] and *Chlamydomonas* (*C.*) *reinhardtii*, (algae) [9] has been studied previously. Moreover, Khalyfa et al. and Tsuchiya et al. isolated cholorophyllase from *Phaeodactylum* (*P.*) *tricornutum* and *C. album*, respectively. Talat and Wang performed a comparative bioinformatics analysis of the chloroplast genomes of a wild diploid Gossypium and two cultivated allotetraploid species [10-12]. However, to our knowledge, little (or in most cases no) information exists about the biochemical properties, phylogenetic relationships and

heterologous expression of plant chlorophyllases. In this work, we characterized plant chlorophyllases (as one of the main sources of chlorophyllase) using different bio-informatics tools. Obviously, this information could be useful in screening chlorophyllases and development of heterologous expression systems for overproduction of plant chlorophyllases.

2. Materials and Methods

2.1. Chlorophyllase sequences

Amino acid sequences of plant chlorophyllases were acquired by a search of the word "chlorophyllase" in the National Centre for Biotechnology Information (NCBI) database (National Institutes of Health, Maryland, USA) via the Entrez cross-database search system (http://www-.ncbi.nlm.nih.gov/).

2.2. Putative conserved domains

Recognition of putative conserved domains within the chlorophyllases was carried out using the NCBI conserved domain database (CDD, National Institutes of Health, Maryland, USA) [13].

2.3. Multiple alignments and distance trees

Chlorophyllase protein sequences were multiple aligned by Clustalw (European Molecular Biology Laboratory, Heidelberg, Germany) [14]. Distance trees, based on the neighbor-joining method and bootstrapping analysis, were drawn by using Mega 5 (The Biodesign Institute, Arizona, USA) [15].

2.4. In silico biochemical characteristics

Number of cystein residues, aliphatic index, molecular weight, pI, instability index, and grand average of hydropathicity (GRAVY) were predicted by ProtParam tool (Swiss Institute of Bioinformatics, Geneva, Switzerland) [16]. Prediction of solubility upon overexpression in *Escherichia* (*E.*) *coli* was performed using SOLpro (Institute for Genomics and Bioinformatics, University of California, CA, USA) [17]. Formation of disulfide bond was predicted using DIpro (Institute for Genomics and Bioinformatics, University of California, CA, USA) [18]. NetNGlyc 1.0 (Bio-Centrum, Technical University of Denmark, Lyngby, Denmarkwas used to predict potential N-glycosylation sites in chlorophyllase sequences) [19].

3. Results and Discussion

3.1. Chlorophyllases

Our search on the Gene Bank resulted in 35 plant chlorophyllase sequences, which are shown in Table 1.

Table 1. Plant chlorophyllases and their properties

Source	Accession no.	Substrate	Localization	Recombinant production
Amborella tauschii	EMT19932	Chlorophyll a	Chloroplast	No
Amborella trichopoda	XP_006838958	chlorophyll a	Chloroplast	No
Arabidopsis lyrata	EFH69319	chlorophyll a	Chloroplast	No
Arabidopsis thaliana	NP_564094	chlorophyll a	Outside the Chloroplast	No
Beta vulgaris	XP_010669258	chlorophyll a	Chloroplast membrane	No
Brassica napus	CDX96924	chlorophyll a	Chloroplast	No
Brassica oleracea	AAN51934	chlorophyll a	Chloroplast	No
Chenopodium album	BAA93635	chlorophyll a	Vacuole	*E. coli*
Citrus limon	ACI06105	chlorophyll a	Chloroplast	No
Citrus sinensis	NP_001275819	chlorophyll a	Chloroplastmembrane	No
Citrus unshiu	BAB47176	chlorophyll a	Plastid	No
Cucumis sativus	XP_004145391	chlorophyll a	Plastid	No
Elaeis guineensis	XP_010934773	chlorophyll a	Chloroplast	No
Ginkgo biloba	AAP44978	chlorophyll a	Thylakoid membrane	No
Jatropha curcas	KDP40227	chlorophyll a	Chloroplast	No
Malus domestica	XP_008355440	chlorophyll a	Chloroplast	No
Medicago truncatula	KEH41056	chlorophyll a	Chloroplast	No
Oryza sativa	AAP53795	chlorophyll a	Chloroplast	No
Pachira macrocarpa	ACO50429	chlorophyll a	Chloroplast	*E. coli*
Phoenix dactylifera	XP_008775648	chlorophyll a	Chloroplast	No
Picea sitchensis	ACN40275	chlorophyll a	Chloroplast	No
Picrorhiza kurrooa	AHZ35334	chlorophyll a	Chloroplast	No
Piper betle	AAP92160	chlorophyll a	Chloroplast	No
Populus trichocarpa	EEF01923	chlorophyll a	Chloroplast	No
Prunus mume	XP_008235366	chlorophyll a	Chloroplast	No
Pyrus x bretschneideri	AEO19902	chlorophyll a	Chloroplast	No
Sesamum indicum	XP_011090675	chlorophyll a	Chloroplast	No
Setaria italica	XP_004983205	chlorophyll a	Chloroplast	No
Solanum lycopersicum	XP_010326690	chlorophyll a	Chloroplast	No
Solanum pennellii	ADZ24715	chlorophyll a	Chloroplast	No
Solanum tuberosum	XP_006363484	chlorophyll a	Chloroplast	No
Triticum aestivum	AHJ14565, BT009214	chlorophyll a	Chloroplast	*E. coli*
Triticum urartu	EMS62705	chlorophyll a	Chloroplast	No
Vitis vinifera	XP_002271936	chlorophyll a	Chloroplast	No
Zea mays	NP_001130783	chlorophyll a	Chloroplast	No

Chlorophyllase gene is located in the nucleus and the expressed enzyme is localized in the chloroplast (chloroplast membrane, thylakoid membrane), outside the chloroplastor vacuole (Table 1) [20]. Chlorophyllases accept chlorophyll a as the main substrate; however, it canhydrolyze pheophytins, as well. During the last decade, chlorophyllase (as a crude extract, purified or recombinant protein) from several plant and algae has been investigated in enzymatic hydrolysis of chlorophyll or its derivatives [3,7,10]. Evidently, in most of the studied flowering plants, more than one isoform of chlorophyllase existed Tsuchiya et al. and Chen et al. isolated two chlorophyllase isozymes from *C. album* and *P. macrocarpa*, respectively. Based on the Table 1, only three types of plant chlorophyllases (from *T. aestivum*, *C. album* and *P. macrocarpa*) have been recombinantly expressed [7,8].

3.2. Conserved domain search

To find the conserved domains within the chlorophyllases, the sequences were analyzed using CDD suite. All the chlorophyllase sequences shared similar conserved domains. Figure 1A shows the domain architecture of the *C. album* chlorophyllase as a representative. The overall domain architecture of plant chlorophyllases consisted of the esterases/lipases superfamily domain (cl21494) over the full length of sequences (with E-values ranging from 0 to $9.81823e^{-162}$), and the alpha/beta hydrolase family domain (pfam12695) over the middle part of the sequences (with E-values ranging from 0.00438209 to $8.87331e^{-09}$). The alignment of *C. album* chlorophyllase with esterases/lipases superfamily domain is shown in Figure 1B.

A

B

```
                10       20       30       40       50       60       70       80
          ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
gi 7415999    30 NSEKITDVFHKGNFQVTNNPI-RVKRYEFSAPEPLIIISPKEAGVYPVLLFIHGTMLSNEDYSLFFNYIASHGFIVVAPK 108
Cdd:pfam07224  1 SSEKTTDVFHKGNYQTTNITVdRSSRYSPPPPKPLIIITPKEAGTYPVVLFLHGTMLSNEFYSLFFNHIASHGFIVVAPQ 80
                90       100      110      120      130      140      150      160
          ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
gi 7415999   109 LFRLFPPklPSQQDEIDMAASVANWMPLYLQVVLQryvTGVEGDLEKLAISGHSRGGKSAFALALGFSnikLDVTFSALI 188
Cdd:pfam07224  81 LYRLFPP--PSQQDEIDSAAEVANWLPLGLQVVLP---TGVEANLSKLALSGHSRGGKTAFALALGYS---LDVTFSALI 152
                170      180      190      200      210      220      230      240
          ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
gi 7415999   189 GVDPVAGRSVDDRTLPHVLTYKPNSFNLSIPVTVIGSGLG---NHTISCAPNHVSHQQFYDECKENSSHFVITKYGHMDM 265
Cdd:pfam07224 153 GVDPVAGTSKDDRTDPHVLTYKPNSFNLSIPVTVIGSGLGarnNHTPPCAPTGVNHEEFFDECKENSAHFVATDYGHMDM 232
                250      260      270      280      290      300      310
          ....*....|....*....|....*....|....*....|....*....|....*....|....*
gi 7415999   266 LNEFRLSPIAVTMSLMCAQSFRPKATMRRTLGGIMVAFLNAYFRDDGRQYYAIIANRSLAPTNLFAEKKGFNFGF 340
Cdd:pfam07224 233 LDDDRLGPIGVTSSCMCKNGFEPRAPMRRFVGGIVVAFLKAYFRDDGREYRAIIKDPSLAPTKLSAEKEGASFGF 307
```

C

```
                10       20       30       40       50       60       70       80
          ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
gi 7415999    76 VLLFIHGTMLSNEDYSLFFNYIASHGFIVVAPKLfrlfppKLPSQQDEIDMAASVANWMPLYLqvvlqryvtgveGDLEK 155
Cdd:pfam12695  1 LVVLLHGAGGDPEAYAPLARALASRGYNVLAPDY------PGHGASLGAPDAEAVLADAALLL----------IDPER 62
                90       100      110      120      130      140      150      160
          ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
gi 7415999   156 LAISGHSRGGKSAFALALGFSNIKldvtfsALIGVDPVagrsVDDRTLPHVltykpnsfnlSIPVTVIGSGLGNHTisca 235
Cdd:pfam12695  63 IVLVGHSLGGAVALLLAARDPRIK------GVVLLAPY----PDTDALAKL----------KVPVLIIHGTRDGVV---- 118
                170      180
          ....*....|....*....|....*..
gi 7415999   236 pNHVSHQQFYDECKENSSHFVITKYGH 262
Cdd:pfam12695 119 -PPEEAEALYAALPGPAELVVIEGAGH 144
```

Figure 1. Domain architecture of *Chenopodium album* (Accession no. BAA93635) chlorophyllase. A) The overall domain organization of plant chlorophyllases consisted of the esterases/lipases superfamily domain (cl21494) over the full length of sequences, and the alpha/beta hydrolase family domain (pfam12695) over the middle part of the sequences. B) Alignment of *Chenopodium album* chlorophyllase sequence with the consensus sequence of esterases/lipases superfamily (pfam07224, a member of the superfamily cl21494; domain consensus length: 307, E-value: $0e^{+00}$, Bitscore: 529.41). C) Alignment of *Chenopodium album* chlorophyllase sequence with the consensus sequence of alpha/beta hydrolase family (domain consensus length: 145, E-value: $3.30e^{-09}$, Bitscore: 53.88).

Esterases/lipases superfamily enzymes (which include fungal lipases, cholinesterases, etc.) act on carboxylic esters. The catalytic apparatus involves three residues (catalytic triad): a serine, a glutamate or aspartate, and a histidine. These catalytic residues are responsible for the nucleophilic attack on the carbonyl carbon atom of the ester bond. In contrast with the other members of alpha/beta hydrolase fold family, p-nitrobenzyl esterase and acetylcholine esterase have a glutamate instead of aspartate at their active site [13]. The esterases/lipases super family domain of chlorophyllases may propose a lipolytic activity for them. However, no report has been published on the lipolytic activity of chlorophyllases.

The pfam12695, alpha/beta hydrolase superfamily, contains a diverse range of alpha/beta hydrolase enzymes. The alignment of *C. album* chlorophyllase with alpha/beta hydrolase family domain is shown in Figure 1C. These enzymes all have a nucleophile histidine acid catalytic triad evolved to operate efficiently on substrates with different chemical compositions or physicochemical properties and in various biological contexts [21].

3.3. Multiple alignments and distance trees

All the chlorophyllase sequences were aligned over their full length using Clustalw to verify sequence conservation among them. Amino acid conservation between 35 plant chlorophyllase sequences was about 50%. In addition, the domain conservation between all the plant chlorophyllase sequences was about 90%. Multiple sequence alignment of plant chlorophyllases showed that all the sequences shared the lipase motif GXSXG (Figure 2). More exactly, the consensus sequence for this motif was found to be GHSRG in plant chlorophyllases (Figure 2). Multiple aligned sequences were also used for bootstrap neighbor-joining tree construction. The phylogenetic tree revealed that chlorophyllase sequences were divided into four clades (Figure 3). The chlorophyllases grouped in clade I were from the plants belonging to Fabaceae, Brassicaceae, Solanaceae and Poaceae families. The chlorophyllase from *Piper* (*p.*) *betle* (Piperaceae family) was grouped in clade II. The chlorophyllases from *Citrus* (*C.*) *sinensis*, *C. unshiu*, and *C. limon* were grouped in clade III. These plants belonged to Rutaceae family. Finally, the chlorophyllases from Brassicaceae, Pinaceae, Cocoeae, Rosaceae, Vitaceae, Poaceae and Amaranthaceae families were grouped in clade IV.

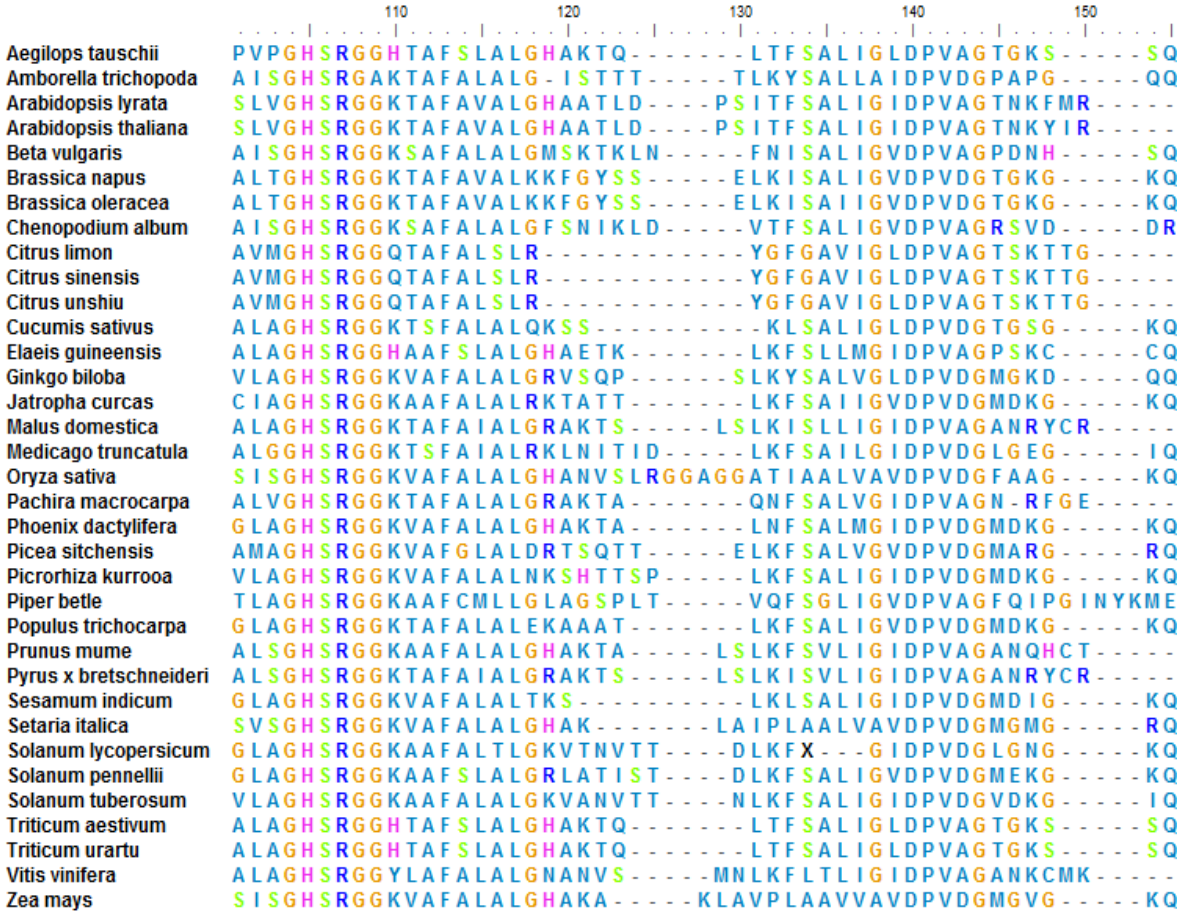

Figure 2. Multiple alignment of plant chlorophyllases. The red line represents the conserved GHSRG lipase motif.

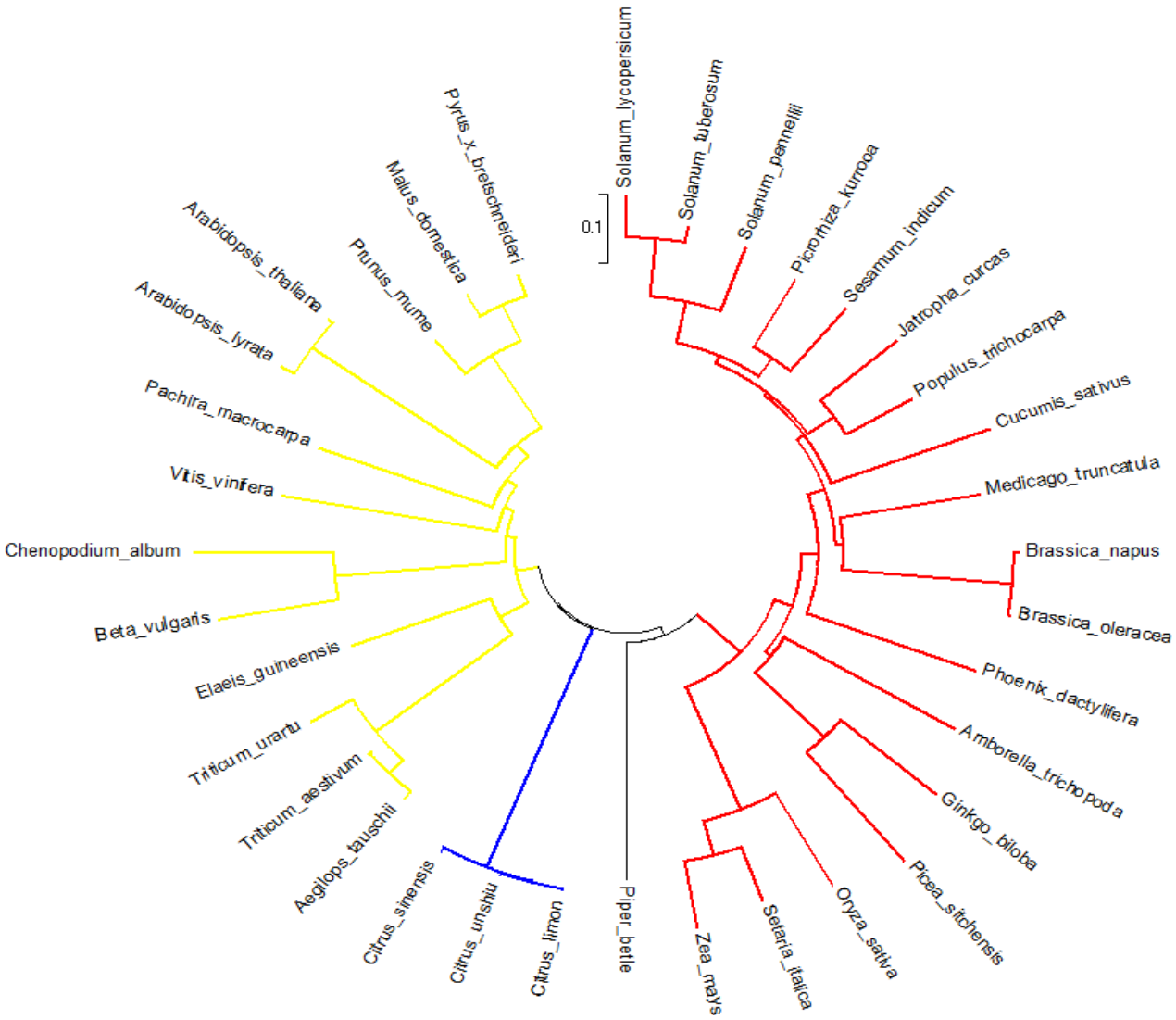

Figure 3. Phylogenetic analysis of plant chlorophyllases. The un-rooted phylogenetic tree was constructed in MEGA5 using the Neighbor-Joining method. Each color represents a clade: Red clade, I; Black line, clade II; Blueclade III; and Yellow, clade IV

As determined by CDD analysis, all the chlorophyllase sequences shared similarity to the esterases/lipases superfamily (cl21494) and alpha/beta hydrolase family (pfam12695). To further analyze the similarity between chlorophyllases, the sequences were multiple-aligned with the consensus sequence of esterases/lipases superfamily (over their full length) or alpha/beta hydrolase family (over their partial sequences having similarity to pfam12695), separately and subjected to tree construction (Figures 4 and 5). Alignment of the chlorophyllases with the cl21494 domain resulted in the formation of four distinct clades. Accordingly, the cl21494 domain was grouped in clade I together with the chlorophyllases from *Solanum* (*S.*) *ycopersicum*, *S. tuberosum*, *S. pennellii*, *Phoenix* (*P.*) *dactylifera*, *G. biloba*, *Amborella* (*A.*) *trichopoda*, *Picrorhiza (P) kurrooa*, *Sesamum* (*S.*) *indicum*, *Jatropha (J.) curcas*, *Populus (P.) trichocarpa*, *Medicago* (*M.*) *truncatula*, *Brassica* (*B.*) *napus*, *B. oleracea*, *Cucumis* (*C.*) *Nsativus*, *Picea* (*P.*) *sitchensis*, *Oryza* (*O.*) *sativa*, *Setaria* (*S.*) *italic* and *Zea* (*Z.*) *mays* (Figure 4). This means that these sequences had higher similarities to the consensus sequence of esterases/lipases superfamily (cl21494).

As shown in Fig. 5, alignment of the chlorophyllases' partial sequences with the pfam12695 domain resulted in the formation of a tree with seven distinct clades, and the pfam12695 domain was grouped in cladeIII together with the chlorophyllase from *B. napus*, *B. oleracea*, *Prunus* (*P.*) *mume*, *A. tauschii*, *T. aestivum* and *T. urartu*. The similarity between the members of each clade may reflect their evolutionary relatedness and the similarity between their functions.

3.4. Physicochemical properties prediction

Biochemical characteristics of plant chlorophyllases predicted using ProtParam are shown in Table 2. Amino acid count of plant chlorophyllases lied between 296 and 367. The enzymes from *B. napus* and *O. sativa* had the

shortest and the longest sequences, respectively. Accordingly, the lowest molecular weight (32.65 kDa) was found to be for *Brassica napus* chlorophyllase, and the highest molecular weight (37.77 kDa) was for the chlorophyllase of *O. sativa* (Table 2).

Vitis (*V.*) *vinifera* chlorophyllase was predicted to have the lowest pI (4.80), and *Chenopodium album* had the highest pI (8.97, Table 2). Based on their pI value, plant chlorophyllases can be divided into three groups: acidic chlorophyllases, having pI≤6, including chlorophyllase from *P. sitchensis, P. betle, A. tauschii, T. aestivum, C. unshiu, C. limon, V. vinifera, Arabidopsis* (*A.*) *lyrata, A. thaliana,* and *P. macrocarpa*; neutral chlorophyllases, having pI=6-8, including those from *S. lycopersicum, S. tuberosum, S. pennellii, P. kurrooa, Sesamum (S.) indicum, Cucumis* (*C.*) *sativus, Phoenix* (*P.*) *dactylifera, Medicago* (*M.*) *truncatula, B. oleracea, A. trichopoda, O. sativa, Setaria* (*S.*) *italic, Z. mays, G. biloba, T. urartu, Elaeis* (*E.*) *guineensis, C. sinensis, P. mume, Malus* (*M.*) *domestica* and *P. bretschneideri*; and alkaline chlorophyllases, having pI>8, including those from *C. album, Beta vulgaris, J. curcas, P. trichocarpa* and *S. lycopersicum.*

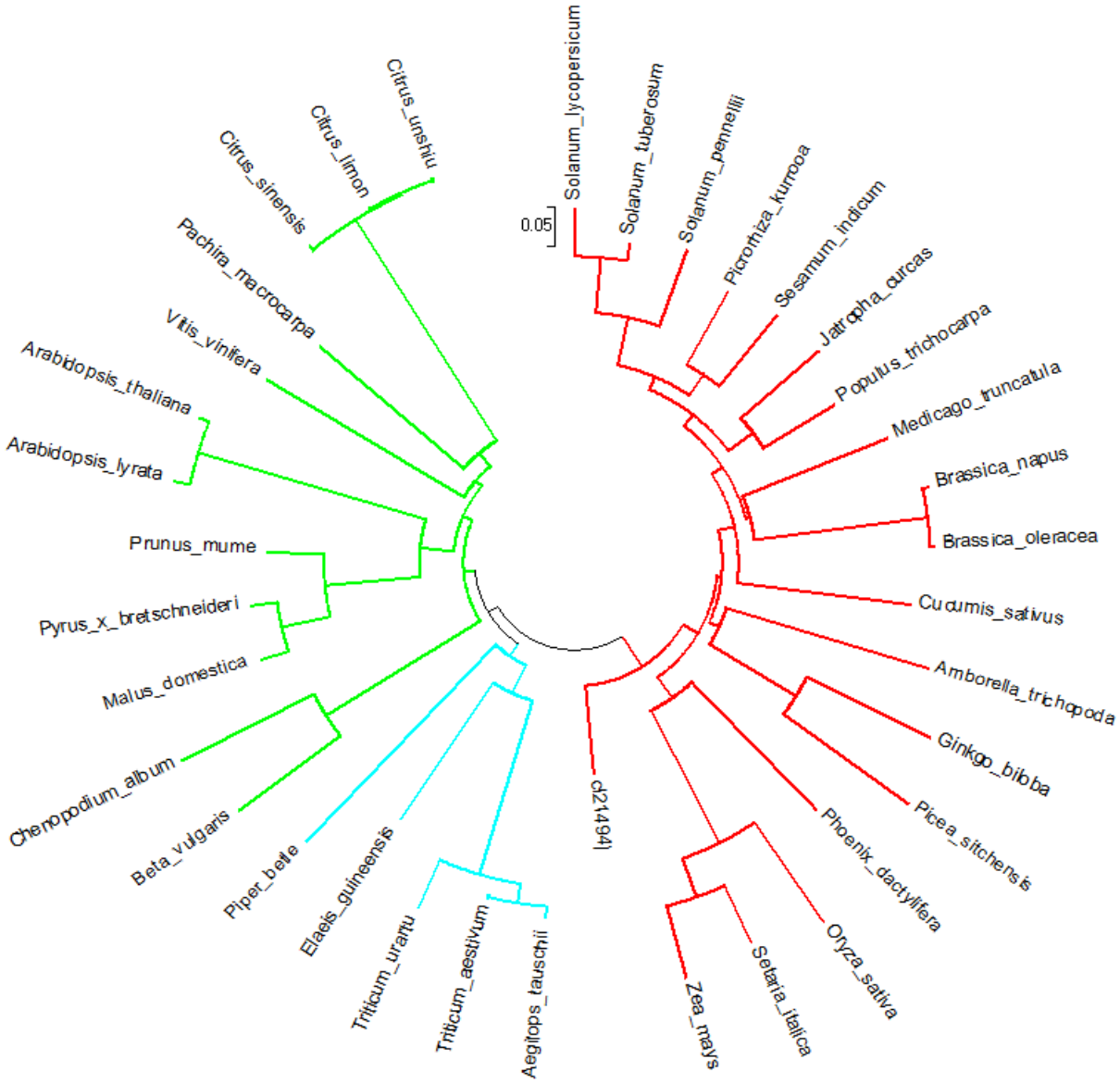

Figure 4. Phylogenetic analysis of plant chlorophyllases sequences aligned with esterases/lipases superfamily (cl21494) domain consensus sequence. Each color represents a clade: Red, clade I; Blue, clade II; and Green, clade III.

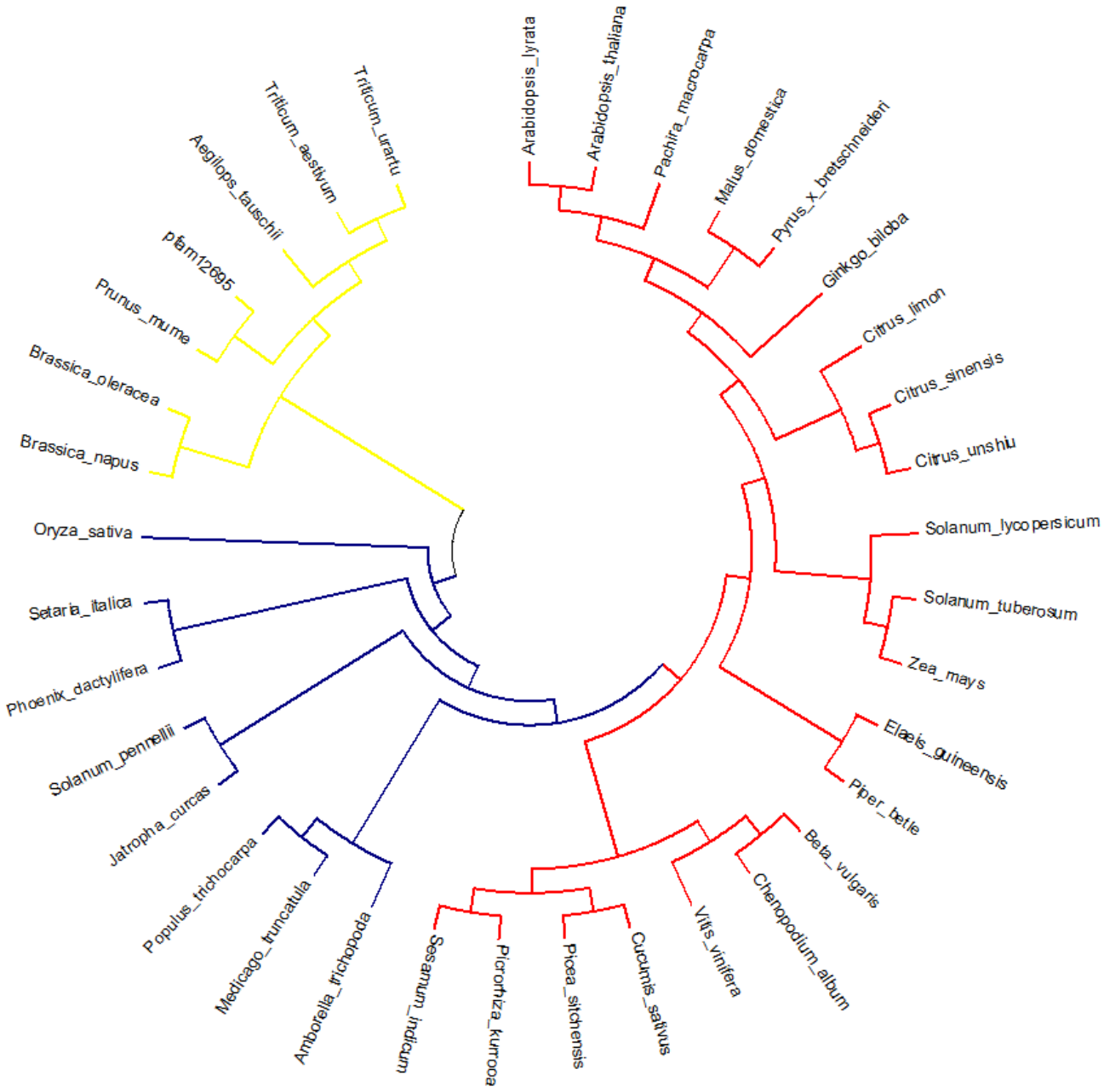

Figure 5. Phylogenetic analysis of partial plant chlorophyllase sequences aligned with alpha/beta hydrolase family (pfam12695) domain consensus sequence. Each color represents a clade: Red, clade I; Blue, clade II; and Yellow, clade III.

Chlorophyllases from *E. guineensis* and *C. limon* had the highest and the lowest aliphatic indices among the chlorophyllases, respectively. The aliphatic index of protein is the relative volume occupied by aliphatic side chains (e.g. alanine, valine, isoleucine, and leucine) [22]. Total number of amino acids alanine, valine, isoleucine, and leucine from *E guineensis* and *C. limon* chlorophyllases were 99 and 88, respectively. It has been found that a direct correlation exists between the aliphatic index and thermostability of proteins [22]. Enzymes with higher thermostability could be used in higher reaction temperatures, which leads to the acceleration of reaction (by decrease of diffusional limitations). In addition, half-life of thermostable enzymes is, generally, greater than that of thermolabile enzymes [23]. Table 2 shows the instability index of the proteins. As indicated, *P. x bretschneideri* and *M. domestica* provide the most unstable and stable plant chlorophyllases, respectively. Stable enzymes are of interest as they can be used for a longer time in biocatalysis [23]. Tsuchiya et al. [7] and Chen et al. found that *C. album* and *P. macrocarpa* chloro-phyllases overexpressed in *E. coli* were stable which con-firms our prediction [8].

Table 2. Predicted physicochemical properties of plant chlorophyllases

Chlorophyllase	Accession no.	Length	MW (KDa)	pI	Instability index	Aliphatic index	GRAVY
Aegilops tauschii	EMT19932	319	33.91	5.71	40.06 (unstable)	319	33914.1
Amborella trichopoda	XP_006838958	314	33.71	7.13	37.07 (stable)	314	33713.7
Arabidopsis lyrata	EFH69319	324	34.84	5.33	42.95 (unstable)	82.78	0.028
Arabidopsis thaliana	NP_564094	324	34.85	5.44	44.02 (unstable)	85.77	0.027
Beta vulgaris	XP_010669258	337	37.20	8.86	37.77 (stable)	337	37200.9
Brassica napus	CDX96924	296	32.65	5.64	36.96 (stable)	85.91	-0.121
Brassica oleracea	AAN51934	321	35.24	6.25	39.10 (stable)	321	35236.5
Chenopodium album	BAA93635	347	38.71	8.97	38.22 (stable)	347	38706.7
Citrus limon	ACI06105	329	35.29	5.90	37.96 (stable)	329	35286.0
Citrus sinensis	NP_001275819	329	35.25	6.06	36.13 (stable)	83.92	0.018
Citrus unshiu	BAB47176	329	35.25	5.75	36.62 (stable)	83.62	-0.006
Cucumis sativus	XP_004145391	316	33.91	6.86	47.56 (unstable)	316	33908.1
Elaeis guineensis	XP_010934773	307	33.23	6.06	37.36 (stable)	307	33232.4
Ginkgo biloba	AAP44978	342	37.18	6.66	40.79 (unstable)	342	37181.6
Jatropha curcas	KDP40227	320	34.36	8.08	44.74 (unstable)	84.12	-0.014
Malus domestica	XP_008355440	323	34.73	7.52	39.94 (stable)	87.86	0.107
Medicago truncatula	KEH41056	306	33.91	6.16	32.39 (stable)	94.93	0.108
Oryza sativa	AAP53795	367	37.77	6.59	38.76 (stable)	367	37768.4
Pachira macrocarpa	ACO50429	313	33.78	5.66	36.62 (stable)	313	33780.5
Phoenix dactylifera	XP_008775648	316	34.54	6.56	35.86 (stable)	316	34539.1
Picea sitchensis	ACN40275	329	35.72	5.81	40.29 (unstable)	329	35719.7
Picrorhiza kurrooa	AHZ35334	315	34.56	6.79	32.55 (stable)	315	34559.6
Piper betle	AAP92160	306	33.01	5.89	37.34 (stable)	306	33005.9
Populus trichocarpa	EEF01923	334	36.06	8.35	36.60 (stable)	334	36058.9
Prunus mume	XP_008235366	319	34.16	6.05	44.17 (unstable)	90.47	0.120
Pyrus x bretschneideri	AEO19902	302	32.31	6.32	47.97 (unstable)	302	32315.3
Sesamum indicum	XP_011090675	315	34.06	7.60	36.29 (stable)	315	34065.2
Setaria italica	XP_004983205	347	35.38	7.19	40.34 (unstable)	347	35377.8
Solanum lycopersicum	XP_010326690	314	34.36	8.17	41.23 (unstable)	88.41	-0.077
Solanum pennellii	ADZ24715	315	34.39	6.82	39.24 (stable)	315	34386.7
Solanum tuberosum	XP_006363484	318	34.77	6.70	39.57 (stable)	94.06	0.022
Triticum aestivum	AHJ14565	319	33.84	5.71	38.25 (stable)	90.94	0.117
Triticum urartu	EMS62705	303	32.44	6.94	37.65 (stable)	303	32437.6
Vitis vinifera	XP_002271936	315	33.26	4.80	30.17 (stable)	315	33257.6
Zea mays	NP_001130783	346	35.99	7.77	36.19 (stable)	346	35988.3

MW, molecular weight; pI, isoelectric point; GRAVY, grand average of hydropathicity

The GRAVY value for a protein or a peptide is calculated by adding the hydropathy values for each residue and dividing by the length of the sequence. As shown in Table 2, the GRAVY of plant chlorophyllase from *O. sativa* is more hydrophilic than the other chlorophyllases. In addition, *C. unshiu* chlorophyllase is less hydrophilic than the other chlorophyllases.

3.5. Solubility upon overexpression

The success of modern biotechnology results from the ability to express foreign or heterologous genes in a host organism. However, transcription and translation of a recombinant gene do not always lead to the accumulation of a correctly folded fully active protein [24], especially when overexpressed in *E. coli* [25]. Idicula-Thomas and Balaji and Bertone et al. suggest that the primary structure of a protein and host/vector-dependent factors may be the determining factors in the inherent propensity of a protein to inclusion body formation [26,27]. The most important primary structure-dependent determinants of propensity to inclusion body formation include hydrophobicity [28], aliphatic index, instability index, molecular weight, acidic residues count, serine composition, Asn, Thr, and Tyr content, and tripeptide composition [26,27,29].

In this study, the SOLpro tool was used to predict the solubility of plant chlorophyllases upon overexpression in *E. coli* [17]. The chlorophyllases from *M. domestica*and *C. unshiu* were predicted to have the highest probabilities of solubility and insolubility, respectively (Table 3). Of 35 plant chlorophyllases, only eight sequences (*S. pennellii, T. aestivum, T. urartu, A. lyrata, P. macrocarpa, P. mume* and *M. domestica*) were predicted to be soluble while the other 28 plant chlorophyllases were predicted to be insoluble upon overexpression in *E. coli*. In fact, as discussed before, chlorophyllases localize in the chloroplast, in some cases, they are known to be membrane-associated proteins. This type of localization requires a more hydrophobic protein sequence, which though helping the protein's travel to the membrane, leads to the lower protein solubility. Insoluble proteins are more prone to inclusion body formation; however, the activity of insoluble proteins can be recovered from inclusion bodies by re-folding techniques [25,26].

Table 3. Prediction of solubility upon overexpression, disulfide bond formation and N-glycosylation of plant chlorophyllase upon expression in *E. coli*

Chlorophyllase	Accession no.	Solubility on overexpresssion[a)]	Cys	Disulfide bonds [b)]	N-glycosylation[c)]
Aegilops tauschii	EMT19932	Insoluble (0.670372)	7	3 (220-234,262-264,270-277)	1 (11)
Amborella trichopoda	XP_006838958	Insoluble (0.500000)	6	2 (75-97,212-226)	0
Arabidopsis lyrata	EFH69319	Soluble (0.530532)	6	2 (49-92,215-229)	1(36)
Arabidopsis thaliana	NP_564094	Insoluble (0.551266)	6	2 (49-92,215-229)	0
Beta vulgaris	XP_010669258	Insoluble (0.700899)	3	1 (235-253)	4 (198,213,257,311)
Brassica napus	CDX96924	Insoluble (0.637278)	6	2 (194-208,272-281)	1 (106)
Brassica oleracea	AAN51934	Insoluble (0.748598)	8	3 (28-30,219-233,297-306)	1 (131)
Chenopodium album	BAA93635	Insoluble (0.788478)	3	1 (234-248)	4 (215,229,251,321)
Citrus limon	ACI06105	Insoluble (0.823975)	5	2 (209-213,263-276)	5 (77,137,229,249,264)
Citrus sinensis	NP_001275819	Insoluble (0.811271)	5	2 (209-213,263-276)	5 (77,137,229,249,267)
Citrus unshiu	BAB47176	Insoluble (0.850124)	5	2 (209-213,263-276)	5 (77,137,229,249,267)
Cucumis sativus	XP_004145391	Insoluble (0.510869)	6	2 (38-39,215-229)	2 (109,103)
Elaeis guineensis	XP_010934773	Insoluble (0.587407)	8	3 (168-169,208-226,252-258)	2
Ginkgo biloba	AAP44978	Insoluble (0.729898)	8	3 (28-44,214-223,271-341)	3 (24,35,103)
Jatropha curcas	KDP40227	Insoluble (0.769768)	7	2 (38-140,234-239)	0
Malus domestica	XP_008355440	Soluble (0.861276)	12	5 (25-29,96-102,180-215,220-234,238-266)	2 (7,270)
Medicago truncatula	KEH41056	Insoluble (0.742705)	4	2 (206-220,252-254)	2 (115,144)
Oryza sativa	AAP53795	Insoluble (0.793297)	8	3 (56-86,272-277,306-340)	2 (189,364)
Pachira macrocarpa	ACO50429	Soluble (0.707594)	6	2 (206-209,259-272)	3 (34,156,246)
Phoenix dactylifera	XP_008775648	Insoluble (0.753297)	6	2 (45-124,207-221)	2 (12,151)
Picea sitchensis	ACN40275	Insoluble (0.572077)	8	3 (90-107,212-226,260-284)	0
Picrorhiza kurrooa	AHZ35334	Insoluble (0.677170)	7	3 (24-29,213-227,228-261)	4 (4,16,125,147)
Piper betle	AAP92160	Insoluble (0.572077)	8	3 (90-107,212-226,260-284)	4 (20,23,111,173)
Populus trichocarpa	EEF01923	Insoluble (0.623300)	7	2 (50-72,234-248)	1 (149)
Prunus mume	XP_008235366	Soluble (0.685624)	10	4 (32-35,176-211,216-230,234-262)	2 (191,266)
Pyrus x bretschneideri	AEO19902	Insoluble (0.614554)	11	4 (8-18,75-159,194-199,213-217)	1 (249)
Sesamum indicum	XP_011090675	Insoluble (0.696239)	9	3 (34-39,214-228,262-267)	3 (2,21,131)
Setaria italica	XP_004983205	Insoluble (0.729457)	7	3 (14-20,235-246,257-286)	0
Solanum lycopersicum	XP_010326690	Insoluble (0.654519)	5	2 (213-227,232-261)	4 (20,35,154,289)
Solanum pennellii	ADZ24715	Soluble (0.646704)	8	3 (11-32,214-228,229-233)	1(290)
Solanum tuberosum	XP_006363484	Insoluble (0.634907)	7	3 (217-231,232-236,263-265)	4
Triticum aestivum	AHJ14565	Soluble (0.581306)	7	3 (220-234,262-264,270-277)	1 (11)
Triticum urartu	EMS62705	Soluble (0.563515)	7	3 (218-222,246-248,254-261)	0
Vitis vinifera	XP_002271936	Insoluble (0.776868)	16	6 (30-59,166-170,205-207,221-225,244-247,260-273)	1 (144)
Zea mays	NP_001130783	Insoluble (0.812731)	4	2 (51-238,259-288)	1 (151)

a) Probability values are presented within parentheses. b) The position of disulfide bonds is written within the parentheses. c) The position of N-glycosylation sites is written within the parentheses.

3.6. Disulfide bond formation

The result of connecting the thiol groups of two cysteine amino acids is formation of disulfides bond on the polypeptide chains. These bonds are responsible for stabilizing the globular structure and correct conformation of the protein [30]. In addition, extracellular proteins often have several disulfide bonds, whereas intracellular proteins usually lack them. Disulfide bonds' prediction can be useful in selecting recombinant expression system [30].

Table 3 shows the prediction by DIpro of disulfide bond formation for different chlorophyllases. Chlorophyllases from *Beta vulgaris* and *C. album* were predicted to have the lowest number (only one) of disulfide bonds (Table 3). Enzymes from *V. vinifera* and *M. domestica* contained the highest amount of cysteine residues, 16 and 12, respectively, and the highest disulfide bonds, 6 and 5, respectively. This means that according to our prediction, recombinant production of these chlorophyllases might be performed using expression systems that provide a suitable redox medium for disulfide bond formation. In this context, yeast extracellular or *E. coli* preplasmic expression systems may be of interest.

3.7. Potential glycosylation sites

Glycosylation is one of the most common and structurally diverse forms of post-translational alteration of proteins. The bacterial host lacks post-translational modification systems such as glycosylation, but species from animals, plants and fungi have the ability to glycosylate proteins [31]. Glycosylation may increase the stability of the protein, and in some cases, decrease the biological activity because of masking of the active site [31]. There are two basic types of protein glycosylation: N-glycosylation and O-glycosylation [27]. N-glycans are produced from a 14-mer precursor structure that is added to asparagine residues in the consensus sequence Asn-X-Ser/Thr, where X can be any amino acid except proline. O-glycans are present on Serine or Threonine residues and do not present a linear consensus sequence from which their position in a protein can be detected. Comparing to O-

glycosylation, N-glycosylation is more effective in molecular weight increase [32].

Table 3 shows the potential N-glycosylation sites of plant chlorophyllases predicted by NetNglyc. Chlorophyllases from *J. curcas, A. trichopoda, S. italica, P. betle, T. urartu* and *A. thaliana* have no potential N-glycosylation site. The highest numbers of N-glycosylation site were found for chlorophyllases from *C. sinensis* (5 potential sites), *Citrus unshiu* (5 potential sites), *C. limon* (5 potential sites), *S. lycopersicum* (4 potential sites), *S. tuberosum* (4 potential sites), *P. kurrooa* (4 potential sites), *P. betle*, *C. album* (4 potential sites), and *B. vulgaris* (4 potential sites). These enzymes together with the other chlorophyllases, which were predicted to contain one to three N-glycosylation sites, may be N-glycosylated if secreted by yeast hosts.

Shemer et al. and Jacob-Wilk et al. suggested that the chlorophyllase from citrus fruit (*C. sinensis*) may be a glycoprotein [20,33]. Terpstra showed that the chlorophyllase from *P. tricomutum* (diatom) contains at least three different glycopeptides that play a role in the enzymatic activity [34]. However, Jacob-Wilk et al. reported that glycosylation is not a prerequisite for the hydrolytic activity of *E. coli*-produced citrus chlorophyllase [33]. There is still little information on the dependence of different plant chlorophyllases on posttranslational glycosylation. Laboratory experiments are suggested to address the issue.

4. Conclusion

In this study, 35plant chlorophyllase sequences were evaluated using different bioinformatics tools. Generally, plant chlorophyllases were similar in terms of their domain structure (the esterases/lipases superfamily domain (cl214 94) over their full length, and the alpha/beta hydrolase family domain (pfam12695) over their middle part sequences). Plant chlorophyllases could be classified into four clades based on the similarity among their sequences. Based on their pI value, plant chlorophyllases could be divided into acidic, neutral and basic enzymes. *V. vinifera* chlorophyllase was predicted to have the lowest pI (4.80) and the chlorophyllase from *C. album* was predicted to have the highest pI (8.97). The chlorophyllases form *S. pennellii, T. aestivum, T. urartu, A. lyrata, P. macrocarpa, P.mume* and *M. domestica* were predicted to be soluble upon overexpression in *E. coli*. The enzyme from *V. vinifera* was found to have the highest number of disulfide bond. The chlorophyllases from *J. curcas, A. trichopoda, S. italica, P. betle, T. urartu* and *A. thaliana* were predicted to be in non-N-glycosylated form. However, due to the possible for-mation of more than two disulfide bonds in these proteins, expression using *E. coli* preplasmic system is suggested for these enzymes. Also due to the prediction of some glycosylation sites or some disulfide bonds in the other enzymes, yeast extracellular expression system is recommended for these enzymes.

5. Acknowledgements

This research was funded by and conducted at the Genetics and Agricultural Biotechnology Institute of Tabarestan (Sari, Mazandaran, Iran).

6. Conflict of Interest

The authors report no conflicts of interest.

References

1. Hörtensteiner S, Kräutler B. Chlorophyll breakdown in higher plants. Biochim Biophys Acta. 2011; 1807(8): 977-988. doi: 10.1016/j.bbabio.2010.12.007
2. Yi Y, Kermasha S, Hocine L, Neufeld R. Encapsulation of chlorophyllase in hydrophobically modified hydrogel. J Mol Catal B-Enzym. 2002; 20: 319-325. doi: 10.1016/S13811177-(02)00182-0
3. Arkus K, Cahoon EB, Jez JM. Mechanistic analysis of wheat chlorophyllase. Arch Biochem Biophys. 2005; 438(2): 146-155. doi: 10.1016/j.abb.2005.04.019
4. Chou YL, Ko CY, Yen CC, Chen LC, Shaw LF. A novel recombinant chlorophyllase1 from *Chlamydomonas reinhardtii* for the production of chlorophyllide derivatives. J Agric Food Chem. 2015; 63(43): 9496-9503. doi: 10.1021-acs.jafc.5b02787
5. Yang CH, Yen CC, Jheng JJ, Wang CY, Chen SS, Huang PY, Huang KS, Shaw JF. Immobilization of *Brassica oleracea* chlorophyllase 1 (BoCLH1) and *Candida rugosa* Lipase (CRL) in magnetic alginate beads: An enzymatic evaluation in the corresponding proteins. Molecul. 2014; 19(8): 11800-11815. doi: 10.3390/molecules190811800
6. Chou YL, Ko CY, Chen LFO, Yen CC, Shaw JF. Purification and immobilization of the recombinant *Brassica oleracea* chlorophyllase 1 (BoCLH1) on DIAION®CR11 as potential biocatalyst for the production of chlorophyllide and phytol. Molecul. 2015, 20(3): 3744-3757. doi: 10.3390/molecules20-033744
7. Tsuchiya T, Ohta H, Okawa K, Iwamatsu A, Shimada H, Masuda T, Takamiya K. Cloning of chlorophyllase, the key enzyme in chlorophyll degradation: Finding of a lipase motif and the induction by methyl jasmonate. Proc Nat Acad Sci. 1999; 96(26): 15362-15367.
8. Chen M, Yang J, Liu C, Lin K, Yang, C. Molecular, structural, and phylogenetic characterization of two chlorophyllase isoforms in *Pachira macrocarpa.* Plant Syst Evol. 2014; 300(4): 633-643. doi: 10.1007/s00606-013-0908-5
9. Chou YL, Lee CC, Yen LF, Chen LC, Shaw LF, A novel recombinant chlorophyllase from *Cyanobacterium cyanothece* sp. ATCC 51142 for the production of bacteriochlorophyllide a. Biotechnol Appl Biochem. 2016; 63(3): 371-377. doi: 10.1002/bab.1380

10. Khalyfa A, Kermasha S, Khamessan A, Marsota P, Alli I. Purification and characterization of chlorophyllase from alga (*Phaeodactylum tricornutum*) by preparative isoelectric focusing. Biosci Biotechnol Biochem. 1993; 57(3): 433-437. doi: 10.1271/bbb.57.433

11. Tsuchiya T, Oht H, Masuda T, Mikami B, Kita N, Shioi Y, Takamiya K. Purification and characterization of two isozymes of chlorophyllase from mature leaves of *Chenopodium album*. Plant Cell Physiol. 1997; 38(9): 1026-1031.

12. Talat F, Wang K. Comparative bioinformatics analysis of the chloroplast genomes of a wild diploid Gossypium and two cultivated allotetraploid. Iran J Biotech. 2015; 13(3): 1-10. doi: 10.15171/ijb.1231

13. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. CDD NCBI's conserved domain database. Nucleic Acids Res. 2015; 43: 222-226. doi: 10.1093/nar/gku1221

14. Thompson JD, Higgins DG, Gibson TJ. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22(22): 4673-4680.

15. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol. 1978; 4(4): 406-425.

16. Gasteiger E, Hoogland C, Gattiker A, Duvaud S. Protein identification and analysis tools on the ExPASyserver. In: Walker JM. The proteomics protocols handbook, Humana Press, New Jersey (USA). 2005: 571-607.

17. Magnan CN, Randall A, Baldi P. SOLpro: Accurate sequence-based prediction of protein solubility. Bioinformatics. 2009; 25(17): 2200-2207. doi:10.1093/bioinformatics-btp386

18. Cheng J, Saigo H, Baldi P. Large-scale prediction of disulphide bridges using kernel methods, two-dimensional recursive neural networks, and weighted graph matching. Proteins. 2006; 62(3): 617-629. doi: 10.1002/prot.20787

19. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. Pac Symp Biocomput. 2002; 7: 310-322.

20. Shemer TA, Smadar HS, Belausov E, Lovat N, Krokhin O, Spicer V, Standing KG, Goldschmidt EE, Eyal Y. Citrus chlorophyllase dynamics at ethylene-induced fruit color-break: A study of chlorophyllase expression posttranslational processing kinetics, and in situ intracellular localization. Plant Physiol. 2008; 148(1): 108-118. doi: 10.1104/pp.108.124933

21. Holmquist M. Alpha/Beta- hydrolase fold enzymes: structures, functions and mechanisms. Curr Protein Pept Sci. 2000; 1(2): 209-235.

22. Ikai A. Thermostability and aliphatic index of globular proteins. J Biochem. 1980; 88(6): 1895-1898.

23. Bommarius S, Riebel BR. Biocatalysis: Fundamentals and applications. Wiley-VCH, Weinheim. 2004: 10-36.

24. Manzur M, Muñoz R, Lucero A, Ayub M, Alvarez S, Ciuffo G. Production of recombinant enzymes of wide use for research. Electron J Biotechnol. 2006; 9(3): 3-16. doi:10.222-5/vol9-issue3-fulltext-16

25. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982; 157(1): 105-132.

26. Idicula-Thomas S, Balaji PV. Understanding the relationship between the primary structure of proteins and its propensity to be soluble on over expression in *Escherichia coli*. Protein Sci. 2005; 14(3): 582-592. doi: 10.1110/ps.041009005

27. Bertone P, Kluger Y, Lan N, Zheng D, Goh C, Echols N, Shawn M, Milburn D, Xiao R, Ma L, Wunderlich Z, Acton T, Gaetano T, Gerstein M. SPINE: An integrated tracking database and data mining approach for identifying feasible targets in high-throughput structural proteomics. Nucleic Acids Res. 2001; 29(13): 2884-2898.

28. Luan CH, Qiu S, Finley JB, Carson M, Gray RJ, Huang W, Johnson D, Tsao J, Reboul J, Vaglio P, Hill DE, Vidal M, Delucas LJ, Luo M. High-throughput expression of *C. elegans* proteins. Genome Res. 2004; 14(10B): 2102-2110. doi: 10.1101/gr.2520504

29. Goh CS, Lan N, Douglas SM, Wu B, Echols N, Smith A, Milburn D, Montelione GT, Zhao H, Gerstein M. Mining the structural genomics pipeline: identification of protein properties that affect high throughput experimental analysis. J Mol Biol; 2004: 336(1): 115-130.

30. Bulleid N, Ellgaard L. Multiple ways to make disulfides. Trends Biochem Sci. 2011; 36(9): 485-492. doi: 10.1016/j-tibs.2011.05.004

31. Easton R. Glycosylation of proteins structure, function and analysis. Life Sci-Technic Bulletin. 2011; 48:1-6.

32. Daly R, Hearn MTW. Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. J Mol Recogn. 2005: 18(2): 119-138. doi: 10.1002/jmr.687

33. Jacob-Wilk D, Holland D, Goldschmidt EE, Riov J, Eyal Y. Chlorophyll breakdown by chlorophyllase: isolation and functional expression of the chlase1 gene from ethylene-treated citrus fruit and its regulation during development. Plant J. 1999; 20(6): 653-661.

34. Terpstra w. Identification of chlorophyllase as a glycoprotein. FEBS let. 1981; 126(2): 231-235. doi: 10.1016/00145793-(81)80249-9

http://dx.doi.org/10.22037/afb.v4i3.14396

Research Article

APPLIED FOOD BIOTECHNOLOGY, 2017, 4 (3): 167-178
Journal homepage: www.journals.sbmu.ac.ir/afb

pISSN: 2345-5357
eISSN: 2423-4214

بررسی زیست‌داده‌ورزی کلروفیلاز گیاهی، آنزیم کلیدی در تجزیه کلروفیل

ابراهیم شرفی[1]، علی دهستانی[1]، جمشید فرمانی[1و2*]، علی پاکدین پاریزی[1]

1- پژوهشکده ژنتیک و زیست فناوری طبرستان، دانشگاه علوم کشاورزی و منابع طبیعی ساری، ساری، ایران.

2- گروه علوم و صنایع غذایی، دانشکده مهندسی کشاورزی، دانشگاه علوم کشاورزی و منابع طبیعی ساری، ساری، ایران.

چکیده

سابقه و هدف: کلروفیلاز هیدرولیز کلروفیل به کلروفیلید و فیتول را کاتالیز می‌کند. اخیرا، چندین کاربرد شامل حذف کلروفیل از روغن‌های گیاهی، استفاده در شوینده های لباس و تولید کلروفیلید برای کلروفیلاز توصیف شده‌است. با این حال، اطلاعات کمی درباره ویژگی‌های بیوشیمیایی کلروفیلازها وجود دارد.

مواد و روش‌ها: 35 توالی پروتئینی کلروفیلاز از پایگاه داده مرکز ملی اطلاعات زیست فناوری به‌دست آمد. همه توالی‌ها از نظر دامنه‌های حفظ شده، روابط تبارژنگانی (phylogenetic) و ویژگی‌های بیوشیمیایی با استفاده از ابزارهای زیست‌داده‌ورزی (Bioinformatic) آنالیز شدند.

یافته‌ها و نتیجه‌گیری: ساختار کلی دمین کلروفیلازها متشکل از دمین ابرخانواده استرازها/لیپازها در طول توالی کامل آنها و دمین خانواده آلفا/بتا هیدرولازها در طول قسمت میانی توالی آنها بود. کلروفیلازهای گیاهی قابل طبقه بندی در 4 گروه تبارژنگانی بودند. وزن ملکولی و pH کلروفیلازها به ترتیب در محدوده 32/65-37/77 کیلودالتون و 4/80-8/97 بود. پایدارترین کلروفیلاز احتمالا مربوط به *Malus domestica* می‌باشد. پیش بینی شد کلروفیلازهای *Solanum pennellii*، *Triticum aestivum*، *Triticum urartu*، *Arabidopsis lyrata*، *Pachira macrocarpa*، *Prunus mume* و *Malus domestica* در هنگام بیش بیان در *E. coli* محلول باشند. و همچنین کلروفیلازهای *Beta vulgaris* و *Chenopodium album* پیوند دی سولفیدی تشکیل ندهند. پیش بینی شد کلروفیلازهای *Jatropha curcas*، *Amborella trichopod*، *Setaria italica*، *Piper betle*، *Triticum urartu* و *Arabidopsis thaliana* به شکل غیر گلیکوزیله تولید شوند.

تعارض منافع: نویسندگان اعلام می‌کنند که هیچ تعارض منافعی وجود ندارد.

تاریخچه مقاله

دریافت 11 اکتبر 2016
داوری 2 آوریل 2017
پذیرش 14 می 2017

واژگان کلیدی

- زیست‌داده‌ورزی
- کلروفیلاز
- تبارژنگانی
- ویژگی‌های بیوشیمیایی

***نویسنده مسئول**

جمشید فرمانی، پژوهشکده ژنتیک و زیست فناوری طبرستان، دانشگاه علوم کشاورزی و منابع طبیعی ساری، صندوق پستی 578، ساری، ایران.

تلفن/دورنگار: 1133687718-98+

پست الکترونیک:
jamshid_farmani@yahoo.com