UNIVERSITÀ
DEGLI STUDI
DI MILANO

ORIGINAL ARTICLE   **Open Access**

Comparative analysis of structural and functional aspects of phytoene synthase from *Meiothermus taiwanensis* strain RP

Trinetra Mukherjee and Subhra Kanti Mukhopadhyay[*] (iD)

Abstract

Purpose: The aim of this study is the in silico characterization of the structure and function of the phytoene synthase (PSY) of a red carotenoid producing thermophile *Meiothermus taiwanensis* strain RP with a comparative approach.

Methods: PSYs from *M. taiwanensis* strain RP and other groups of thermophilic, mesophilic and psychrophilic bacteria, plants, protozoa, and algae were analyzed by ExPASy ProtParam, NCBI Conserved Domain Search, SOPMA, PSIPRED, Robetta server, ProQ, and QMEAN, with the superposition of 3-D structures in PyMOL.

Results: RP PSY shows the highest (97.5 %) similarity with *M. ruber* and the lowest with the psychrophile *Gelidibacter algens* (36.7 %). The amino acid sequence of RP PSY is one of the shortest, with 275 residues. The instability index of RP PSY is much lower compared with plant sequences. Alanine, arginine, glycine, and leucine residues are the highest in *Meiothermus* sp., and they have a high amount of alpha-helix. Most of the 32 active site residues are conserved in all the sequences. However, some residues are more prone to substitutions in other PSYs except *M. ruber*. The three-dimensional structures of *M. taiwanensis* strain RP, *Gelidibacter algens*, *Thermus thermophiles*, *Meiothermus ruber*, and *Brassica napus* PSYs were homology modeled, validated, and submitted to Protein Model Databank. The superposition of the 3-D structures shows that their active site region structure is identical.

Conclusion: RP PSY is one of the most stable PSYs and knowledge of its individual properties, similarities, and dissimilarities with other PSYs may be useful for genetic engineering and purification of the protein for improved carotenoid production.

Keywords: Thermophilic, Alpha-helix, Aspartate rich region, Homology modeling, Superposition

Introduction

Carotenoids are a subfamily of isoprenoids with an extensive conjugated double bond system, acting as chromophores. They are red, yellow, or orange colored and are found in plants, fungi, bacteria, protozoa, algae, and archaea. Higher order animals do not synthesize carotenoids but may obtain them through diet. Carotenoids have a wide variety of functions such as coloring property, anti-oxidant property, conferring membrane stability, and protection against heat and light (Britton et al. 2004; Britton 2008a). The carotenoid beta-carotene acts

as the precursor for vitamin A, which is a vital nutrient component of human health (Britton 2008b). Phytoene, the first 40 carbon intermediate carotenoid formed in the carotenoid biosynthetic pathway, is colorless in nature (Paniagua-Michel et al. 2012). However, phytoene is the precursor for the synthesis of colorful carotenoids through structural modifications and the addition of functional groups. Phytoene synthase (PSY) is the enzyme which catalyzes the synthesis of phytoene from the condensation of two molecules of geranylgeranyl diphosphate using manganese as co-factor. This reaction is the rate limiting and first committed step of carotenoid biosynthesis. Thus, it is an essential regulatory enzyme controlling carotenoid pigment formation (López-Emparán

* Correspondence: microskm2015@gmail.com
Department of Microbiology, The University of Burdwan, Burdwan, West Bengal, India

et al. 2014). This enzyme is so important that when suppressed it leads to a decreased production of carotenoids (Kato et al. 2017). An in-depth study of this enzyme is crucial to understand, regulate, and engineer carotenoid biosynthesis for increased production of carotenoids. Phytoene synthases have been studied in silico, directly from natural sources and by cloning and expression in host cells (Iwata-Reuyl et al. 2003; Fu et al. 2014; Agarwal et al. 2015). Most of the previous studies on phytoene synthases have focused on plant origin enzymes; studies on phytoene synthases of bacterial origin are few (Iwata-Reuyl et al. 2003). Clearly, we need to focus more on bacterial origin PSYs, especially thermostable ones.

Meiothermus taiwanensis strain RP, belonging to the Deinococcus-Thermus phylum, produces red carotenoid pigments having a high anti-oxidative capacity (Mukherjee et al. 2017). We isolated this thermophilic organism from the Paniphala hot spring of India (Mukherjee et al. 2016). Detailed analysis of phytoene synthases from the Deinococcus-Thermus group of bacteria have not been reported until now. *Meiothermus* sp. forms red and yellow pigments related to carotenoid glycoside esters. One carotenoid pigment has been identified from *M. ruber* as 1′-β-glucopyranosyl-3,4,3′,4′-tetradehydro-1′,2′-dihydro-β,ψ-caroten-2-one glucose acetylated at position 6 (Burgess et al. 1999). In our study, we took the sequence of phytoene synthase of *Meiothermus taiwanensis* strain RP from its whole genome sequence database, which was previously sequenced by us (Mukherjee et al. 2016) and analyzed in silico. We analyzed the protein structurally on three levels—primary, secondary, and tertiary, along with phylogenetic analysis and study of active site residues. Simultaneously, we also carried out the structural, functional, and phylogenetic analysis of phytoene synthases from other organisms for comparison. This comparative analysis revealed crucial facts regarding structural, functional, and phylogenetic similarity and dissimilarity among different phytoene synthases.

Materials and methods

Retrieval of phytoene synthase (PSY) sequences

The FASTA format of the phytoene synthase protein sequence of *Meiothermus taiwanensis* strain RP (KZK15231.1) (Mukherjee et al. 2016) was retrieved from its whole genome database. Other phytoene synthase sequences used in different sections of this study were retrieved from the NCBI protein database (https://www.ncbi.nlm.nih.gov/protein/). After downloading a large number of PSY sequences from NCBI, they were initially checked in NCBI conserved domain search (Marchler-Bauer et al. 2011) for removal of sequences which lacked some or all of the PSY functional domains, or contained multiple unrelated domains indicating unspliced sequences from translated genomic data (observed in case of some eukaryotic sequences). Details of all

the phytoene synthase protein sequences finally selected for comparison are given in Table 1. SI 1–6 contain sequences from 6 different species of *Meiothermus* sp. for intra-genus comparison. SI 7–13 contain sequences from Deinococcus-Thermus phylum for intra-phylum analysis. SI 1–13 formed a comparative entity for analyzing any patterns arising in thermophilic PSYs. SI 14–22 show some sequences selected from mesophilic bacteria, psychrophilic bacteria, and cyanobacteria and were selected for intra-domain comparison among various taxonomic groups of bacteria. SI 20 contains the PSY sequence from the psychrophilic bacteria *Gelidibacter algens* for thermo-psychro PSY analysis. SI 21–22 contain sequences from cyanobacteria, SI 23 contains sequence from protozoa, and SI 24–28 contain sequences of plant origin. The percentage of similarity and identity of the sequence with other phytoene synthase sequences was calculated using Water Pairwise Sequence Alignment (EMBOSS) (Rice et al. 2000) (http://www.ebi.ac.uk/Tools/psa/emboss_water/) following the Smith-Waterman algorithm (Smith and Waterman 1981).

Phylogenetic analysis of phytoene synthase

Multiple sequence alignment (MSA) of RP PSY with all the sequences given in Table 1 was done using MUSCLE (Edgar 2004) in MEGA7 (Kumar et al. 2016). Using this MSA, the phylogenetic tree was also constructed in MEGA 7 (Kumar et al. 2016) by the maximum likelihood method (Kishino and Hasegawa 2001) based on the Jones, Taylor, and Thornton (JTT) model (Jones et al. 1992) with a gamma distribution (Nei and Zhang 2006). The scale of the branch lengths of the tree corresponds to the number of substitutions per site.

Primary structure analysis and physicochemical characterization

For primary structure analysis, the ExPASy ProtParam tool (Gasteiger et al. 2005) was used. The molecular weight, instability index, pI value (isoelectric point), aliphatic index, and grand average of hydropathicity (GRAVY) of the phytoene synthase of *Meiothermus taiwanensis* strain RP and all the proteins of Table 1 were calculated and compared. Proteins with low pI values were considered as acidic, and those with high pI values were considered as basic (Moldoveanu et al. 2017). The instability index predicted the stability of the protein in a test tube and values above 40 indicated that the protein might be unstable (Guruprasad et al. 1990). The aliphatic index was the relative volume occupancy of the protein by side chains, which are aliphatic and was considered to be greater in the case of thermophilic proteins (Ikai 1980). Positive GRAVY values indicated the hydrophobic nature of the protein (Kyte and Doolittle 1982). The percentage amino-acid composition of the phytoene synthase sequences of *Meiothermus taiwanensis*, *M.*

Table 1 Details of the phytoene synthase sequences used in this study along with the percentage identity and similarity with the phytoene synthase of *Meiothermus taiwanensis* strain RP

SI no.	Organism	Accession number	Sequence identity (%) with RP	Sequence similarity (%) with RP
1	*Meiothermus ruber* DSM 1279	ADD27155.1	94	97.5
2	*Meiothermus cerbereus*	WP_027876284.1	89	94.5
3	*Meiothermus rufus*	WP_027882295.1	81	87.6
4	*Meiothermus timidus*	WP_018467035.1	76	84.7
5	*Meiothermus chliarophilus*	WP_027891330.1	74	83.9
6	*Meiothermus silvanus* DSM 9946	ADH63377.1	70	80.3
7	*Thermus thermophilus*	WP_014511254.1	64.4	76.7
8	*Thermus aquaticus*	WP_053768023.1	65	75.99
9	*Thermus filiformis*	WP_045245896.1	63	74.3
10	*Thermus arciformis*	SDE98932.1	64	76.5
11	*Thermus scotoductus*	WP_054392157.1	58	72.1
12	*Deinococcus murrayi*	WP_034407839.1	52	62.99
13	*Deinococcus radiodurans*	WP_010887508.1	49	60.7
14	*Streptomyces coelicolor* A3(2)	NP 630832.1	30.6	43.5
15	*Bradyrhizobium diazoefficiens* USDA 110	NP_771372.1	22.4	43.8
16	*Methylobacterium extorquens* DSM 13060	EHP92507.1	36	48.9
17	*Staphylococcus aureus* subsp. aureus Tager 104	AMO53214.1	30.5	48.7
18	*Lactobacillus plantarum* IPLA88	EPD23201.1	33.1	47.5
19	*Bacillus stratosphericus* LAMA 585	EMI13846.1	38.2	55.7
20	*Gelidibacter algens* ACAM 536	LZRN00000000	23.5	36.7
21	*Cyanobacterium aponinum* PCC 10605	AFZ53003.1	37.9	52.6
22	*Trichormus azollae*	WP_013191723.1	36	52.4
23	*Euglena gracilis*	BAU24764.1	36.9	51
24	*Brassica napus*	AHE79749.1	36.9	53.5
25	*Arabidopsis thaliana*	AED92399.1	36.9	54.2
26	*Rhododendron japonicum* f. flavum	BAS69427.1	37.3	55.4
27	*Ricinus communis*	AIU48728.1	36.6	53.8
28	*Solanum tuberosum*	AIU48722.1	36.9	56.5

ruber, *Cyanobacterium aponinum*, *Euglena gracilis*, *Gelidibacter algens* and *Arabidopsis thaliana* was determined and compared.

Analysis of the secondary structure of the proteins
PSIPRED v3.3(McGuffin et al. 2000) was used to determine the secondary structure of the phytoene synthase of *M. taiwanensis* strain RP. The percentages of the different secondary structures found in the phytoene synthase of RP and all the sequences from Table 1 were

determined with the help of SOPMA (Geourjon and Deléage 1995) and compared with RP.

Active site analysis of the phytoene synthases
The functional analysis of the phytoene synthase protein from *Meiothermus taiwanensis* strain RP was performed and compared with other phytoene synthases. For this study, the substrate binding pocket, substrate Mg^{2+} binding site, active lid-residues, catalytic residues, and aspartate-rich regions of the phytoene synthases from *Meiothermus*

taiwanensis strain RP, *M. ruber*, *M. silvanus*, *Thermus thermophilus*, *Deinococcus radiodurans*, *Gelidibacter algens*, *Brassica napus*, *Staphylococcus aureus*, and *Cyanobacterium aponinum* were determined with the help of NCBI Conserved Domain Search (Marchler-Bauer et al. 2011) and compared. We detected significant variations and similarities between the active sites of this protein from different organisms.

Comparative analysis of three-dimensional structures of phytoene synthases

The three-dimensional structure of the phytoene synthase sequences of *Meiothermus taiwanensis* strain RP, *M. ruber*, *Thermus thermophilus*, *Brassica napus*, and *Gelidibacter algens* was elucidated by homology modeling in Robetta server (Kim et al. 2004) using multiple sequence alignment of related models. The PDB files of the models were visualized using the PyMOL Molecular Graphics System, version 2.0 Schrödinger, LLC. The models were energy minimized using the Chiron server (Ramachandran et al. 2011). The servers which were used for validation and model quality interpretation were QMEAN server (Benkert et al. 2009), ProQ (Wallner and Elofsson 2003), PROCHECK (Laskowski et al. 1993), PROVE (Pontius et al. 1996), ERRAT (Colovos and Yeates 1993), and Verify3D (Bowie et al. 1991). The structures were compared with each other by superposition studies using the align method of alignment in PyMOL. The PDB files of the predicted structures were submitted to Protein Model Databank (Castrignanò et al. 2006).

Results and discussion

Retrieval of phytoene synthase sequences and phylogenetic analysis

The percentage similarity and identity of *Meiothermus taiwanensis* strain RP PSY with other phytoene sequences is shown in Table 1. Even with its closest relative *M. ruber*, its percentage similarity is not more than 97.5% percent. It shows the lowest intra-genus similarity with *M. silvanus* (83.9 %). *Meiothermus* belongs to the Deinococcus-Thermus phylum, comprising mostly of microbes living under thermophilic conditions (Griffiths and Gupta 2007). Strain RP shows minimum intra-phylum similarity with *Deinococcus radiodurans* (60. 7 %). The similarity of strain RP PSY varies between 55.7 and 43.5% with other mesophilic or thermotolerant bacteria and 55.4 to 51% with phytoene synthase sequences from cyanobacteria, protozoa, and plants given in the table. The phytoene synthase sequence from *Gelidibacter algens* (Bowman et al. 1997) is of psychrophilic origin in contrast to the thermophilic RP PSY. Interestingly, this sequence has the least similarity with RP, with a percentage of 36.7%. The phylogenetic analysis of the sequence of RP is shown in Fig. 1. Here the phytoene synthase sequence

of strain RP forms a phylogenetic group with other members of the Deinococcus-Thermus family, initially with *Meiothermus* sp. and then with *Thermus* sp. and *Deinococcus* sp. The previously reported phylogenetic trees constructed from analysis of phytoene synthase genes (Phadwal 2005; Klassen 2010) also showed phytoene synthase genes of Deinococcus-Thermus phylum forming a cluster. The PSY genes from eukaryotes, actinobacteria, proteobacteria, cyanobacteria, and bacteroidetes formed separate clusters in previous studies (Phadwal 2005; Klassen 2010). The phytoene synthase amino-acid sequences used in this study also show similar topologies.

Primary structure analysis and physicochemical characterization

The number of amino acids, pI value (isoelectric point), aliphatic index, and grand average of hydropathicity (GRAVY) of all the PSY proteins from Table 1 are given in Additional file 1: Table S1. The molecular weight and instability index of the proteins are represented in Fig. 2.

Number of amino acids and molecular weight

The amino acid sequence of *M. taiwanensis* strain RP is 275 residues long. All the sequences of *Meiothermus* sp. are between 275 and 280 amino acid (a.a.) residues long. The molecular weight of strain RP is 30.6 kDa, and the *Meiothermus* sequences vary between 30.5 and 31.5 kDa. The phytoene synthases from *Meiothermus* sp. are some of the smallest PSY proteins among the tested sequences used in this study. From the Brenda database (https://www.brenda-enzymes.org/enzyme.php?ecno=2.5.1.32&onlyTable=Sequence), it can be observed that phytoene synthases from bacteria have a wide range of size and sequence length, but most of them are found towards the lower range of the molecular weight of phytoene synthases (< 36 kDa), except cyanobacteria. Phytoene synthase of the lowest sequence length having all the necessary conserved domains has been derived from previously reported whole genome sequence of the bacteria *Cronobacter sakazakii* 696 (240 amino acid residues and 26.5 kDa molecular weight) (https://www.brenda-enzymes.org/sequences.php?ID=4201498&fasta=1) (Joseph et al. 2012). The PSY sequences from plant origin have shown the highest molecular weights in this study. The results are similar to previous reports of other phytoene synthases of plant origin such as *Oryza sativa* subsp. *japonica* (444 amino acid residues and 49.3 kDa molecular weight) (Welsch et al. 2008), *Triticum aestivum* (425 amino acid residues and 47.5 kDa molecular weight) (Crawford et al. 2011) and *Zea mays* (426 amino acid residues and 47.3 kDa molecular weight) (Li et al. 2008). Phytoene synthases have directly evolved
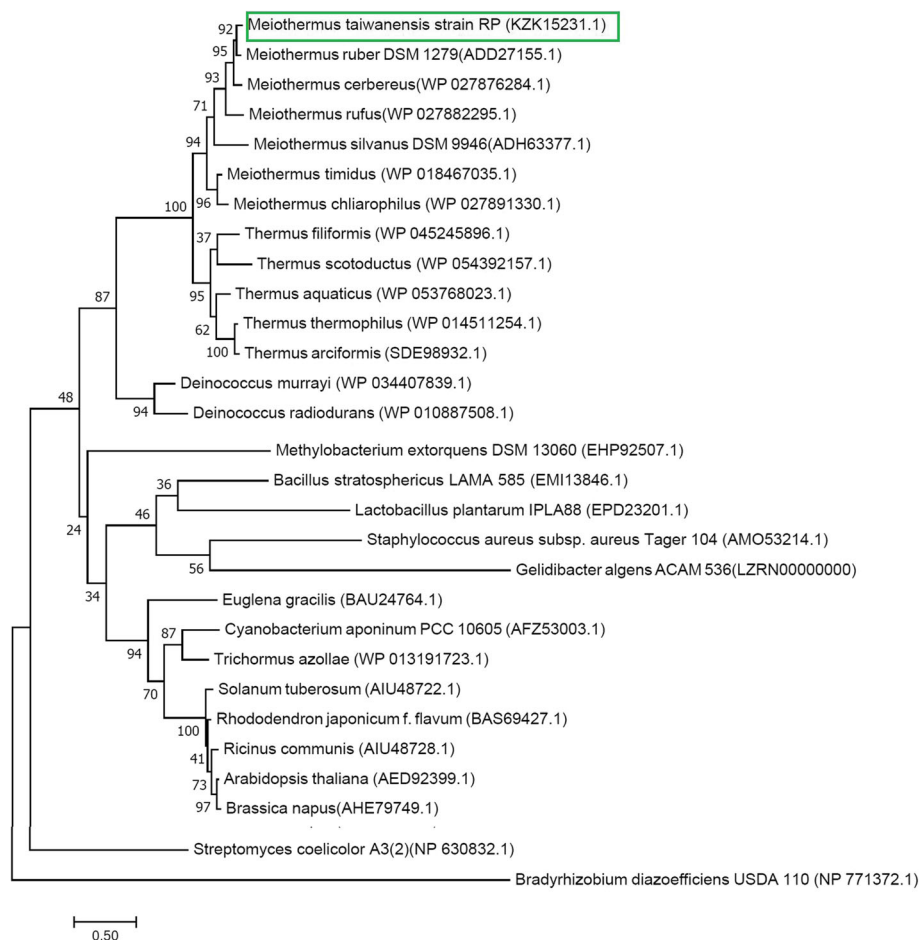
Fig. 1 Phylogenetic analysis of the phytoene synthase of *Meiothermus taiwanensis* strain RP showing its relationship with other phytoene synthase sequences using the maximum-likelihood method in MEGA 7

from bacteria to plants through various bacterial taxonomic groups and algae (Sandmann 2002). The phytoene synthases of plants are located in plastids (Shumskaya et al. 2012) showing tissue-specific expression (Dibari et al. 2012), whereas bacteria do not have such compartmentalization. The differences in localization, internal environmental conditions, the presence/absence of introns (Dibari et al. 2012; Ahrazem et al. 2019), and the presence/absence of signal peptides (Kadono et al. 2015) might be the reason for the variation in sequence length among the different PSY proteins.

Stability of the proteins

The instability index of *Meiothermus taiwanensis* strain RP sequence is 31.09, which is well below the value of 40, above which proteins are deemed to be unstable. Thus, the RP PSY protein is stable in nature. *M. rufus*, *M. ruber*, *M. timidus*, and *M. cerbereus* also have values below 40. However, *M. silvanus* and *M. chliarophilus* show values of 39.59 and 44.41; thus, they are unstable

proteins. Forty-four percent of the tested sequences have instability index values below 40, and the rest have above 40. The psychrophilic *Gelidibacter* sp. protein is also stable. The plant, cyanobacterial, and protozoal PSY sequences used in this study are unstable. PSY sequences from plants have been reported to be unstable previously (Zhao et al. 2011). A chaperone protein named Orange protein has been reported to help in the stability of phytoene synthases in many plants (Ahrazem et al. 2019); no PSY stabilizing protein in bacteria has been reported so far.

Isoelectric point (pI)

The PSY proteins used in this study have a wide range of pI values from 5.1 to 9.53. Thus, they may be acidic or basic. The phytoene synthase from *M. taiwanensis* strain RP shows a pI value of 8.36 and is basic similar to *M. rufus*. The other sequences from *Meiothermus* sp. are neutral (*M. chliarophilus*) or acidic (*M. ruber*, *M. silvanus*, *M. timidus*, and *M. cerbereus*). The tested sequences do not show any intra-genus or inter-genus or
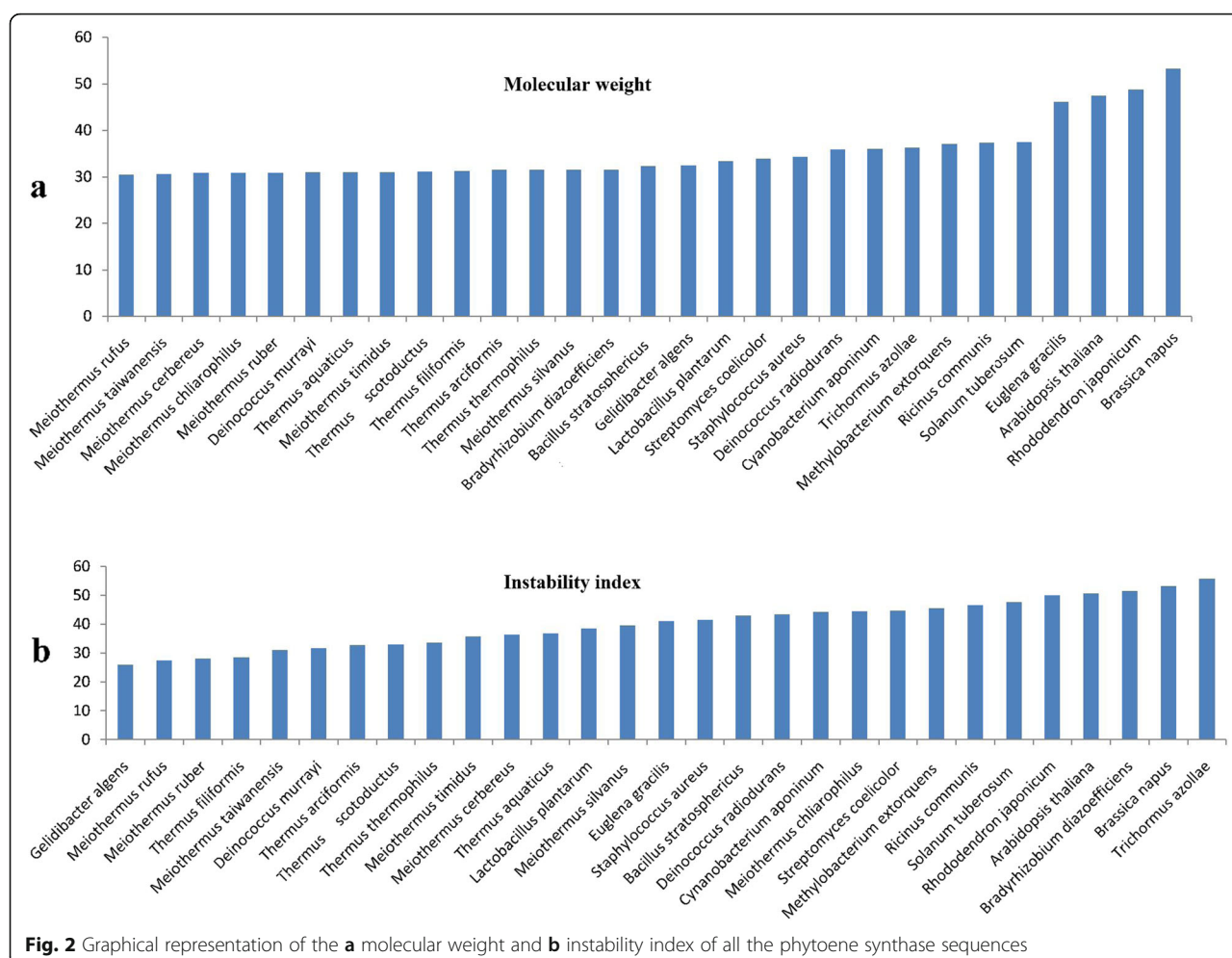
Fig. 2 Graphical representation of the **a** molecular weight and **b** instability index of all the phytoene synthase sequences

inter-domain similarity concerning pI values which is similar to previous reports of plant origin phytoene synthases showing a wide range of pI values (Iwata-Reuyl et al. 2003; Lao et al. 2011; Han et al. 2015).

Aliphatic index
The aliphatic index of all the *Meiothermus* sp. sequences is greater than 90. The aliphatic index of *M. taiwanensis* is 96 and that of *M. cerbereus* is 98.07. The high value of the aliphatic index shows the thermophilic nature of the proteins from *Meiothermus*. The aliphatic index of the psychrophilic *Gelidibacter* sp. is 88.46, lower than the thermophilic proteins, though higher than many of the other sequences. All the phytoene synthases have aliphatic values of greater than 80, which is similar to some previously reported phytoene synthases of plant origin (Leng et al. 2017).

GRAVY value
The GRAVY value of all the sequences lies between − 0.82 and − 0.069. *M. taiwanensis* has a GRAVY value of − 0.133, which lies on the hydrophobic side of the index.

M. rufus is the most hydrophobic one with a value of − 0.069. *M. ruber* and *M. cerbereus* also lie towards the hydrophobic side of the index, whereas *M. timidus*, *M. silvanus*, and *M. chliarophilus* have slightly smaller values. The other organisms have values throughout the range of the GRAVY index.

Amino acid composition
The comparative analysis of the percentage amino-acid composition of the phytoene synthase sequences of *Meiothermus taiwanensis*, *M. ruber*, *Cyanobacterium aponinum*, *Euglena gracilis*, *Gelidibacter algens*, and *Arabidopsis thaliana* is given in Fig. 3. *Meiothermus taiwanensis* strain RP and *M. ruber* show an abundance of alanine, arginine, glutamic acid, tryptophan, glycine, and leucine residues with percentages higher than all the other phytoene synthase sequences including *Gelidibacter*. Cysteine and histidine are present in low numbers in all the sequences. The psychrophilic *Gelidibacter* sp. sequence has higher percentages of asparagine, aspartic acid, lysine, and phenylalanine as compared with other sequences and higher percentages of histidine,
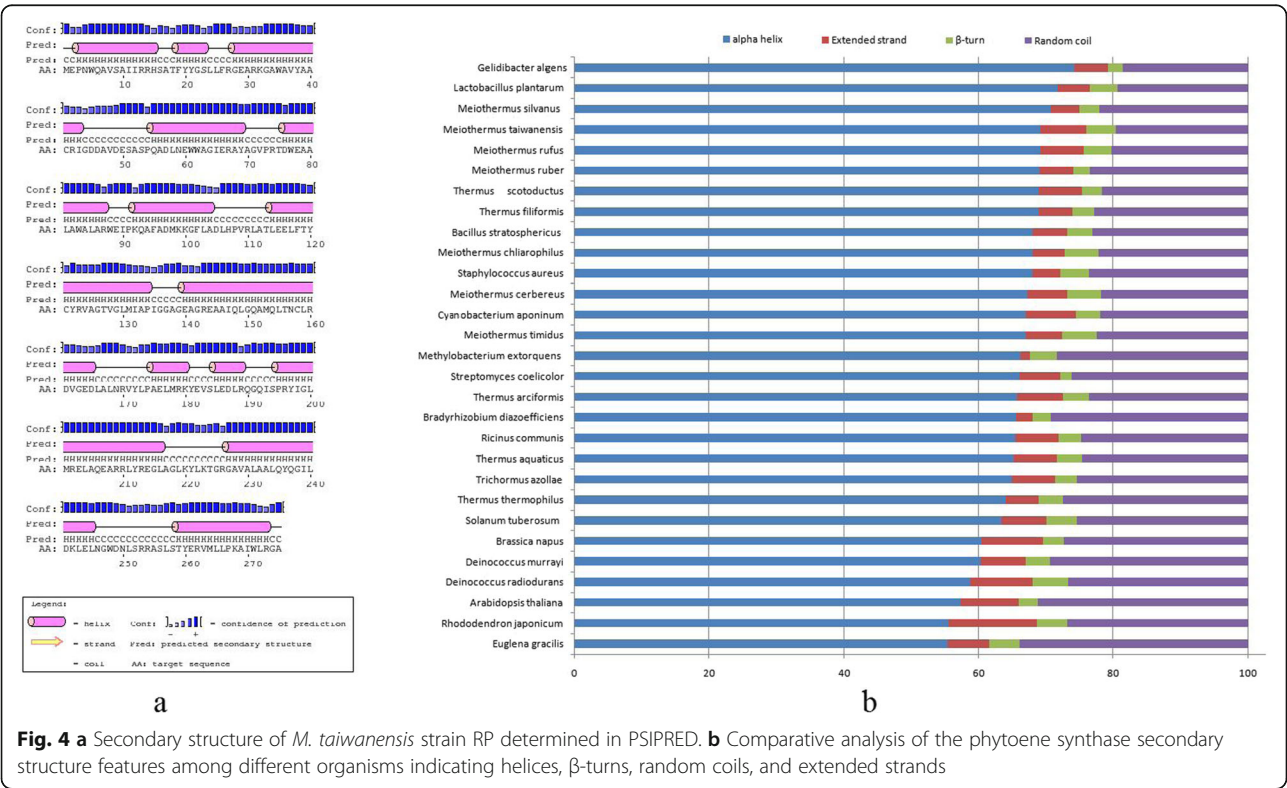
Fig. 3 Comparative analysis of the percent amino-acid composition of the phytoene synthase sequences of *Meiothermus taiwanensis* strain RP, *M. ruber*, *Cyanobacterium aponinum*, *Euglena gracilis*, *Gelidibacter algens*, and *Arabidopsis thaliana*
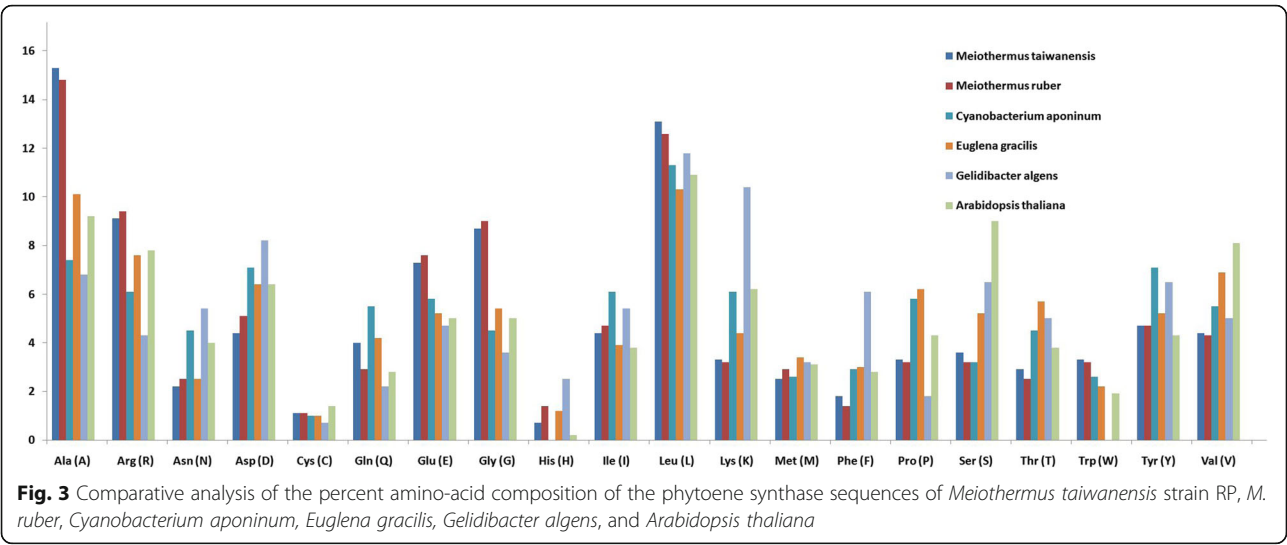
isoleucine, methionine, serine, threonine, tyrosine, and valine over the thermophilic *M. taiwanensis* and *M. ruber*. Some amino acids like threonine, histidine, asparagine, serine, and glutamine are present in lower numbers in the thermophilic RP PSY and have been reported to be found in lower numbers in other thermophilic proteins too on many instances (Kreil and Ouzounis 2001; Metpally and Reddy 2009; Sælensminde et al. 2009).

Analysis of the secondary structure of the proteins

The secondary structure of *M. taiwanensis* strain RP PSY and its comparative analysis with secondary structures of the other PSY sequences are shown in Fig. 4. The major secondary structure present in the tested phytoene synthases is alpha-helix and no β-strands are present. Some of the previously reported protozoan and algal phytoene synthases do not have β-strands (Agarwal et al. 2015; Li et al. 2015), while some have a very short

Fig. 4 a Secondary structure of *M. taiwanensis* strain RP determined in PSIPRED. **b** Comparative analysis of the phytoene synthase secondary structure features among different organisms indicating helices, β-turns, random coils, and extended strands

single stretch of β-strand (Lao et al. 2011; Agarwal et al. 2015). Extended strands are present in the phytoene synthases used in our study, which are not part of β-strands. These extended strands contain irregular hydrogen bonding and might form loops (Eswar et al. 2003). β-turns are present at the ends. Random coils are present in the secondary structure of all the sequences, though the proportions vary. Previously reported phytoene synthases also contained extended strands, β-turns, and random coils (Lao et al. 2011; Agarwal et al. 2015; Li et al. 2015). Random coils are non-native protein states which have no structure except some inherent local interactions (Smith et al. 1996). The psychrophile *Gelidibacter algens* contains the greatest alpha-helix percentage. The thermophilic *Meiothermus taiwanensis*, *M. silvanus*, *M.rufus*, and *M. ruber* also have higher alpha-helix percentages compared with other sequences except *Lactobacillus plantarum* and also lower percentages of extended strands and random coils. The bacterial sequences of *Deinococcus* sp., the protozoa *Euglena*, and the plant sequences of *Arabidopsis thaliana*, *Rhododendron japonicum*, *Brassica napus*, and *Solanum tuberosum* have lower percentages of alpha-helix and higher percentages of extended strands and random coils than the other PSY structures. Thus, *Gelidibacter algens*, *Meiothermus taiwanensis*, *M. silvanus*, *M.rufus*, and *M. ruber* have more well-defined secondary structures with proper folding than the other phytoene synthases.

Active site analysis of the phytoene synthases

The active site residues (substrate binding pocket, substrate Mg^{2+} binding site, active lid-residues, catalytic residues, and aspartate-rich regions) of the PSY sequences from *Meiothermus taiwanensis* strain RP, *M. ruber*, *M. silvanus*, *Thermus thermophilus*, *Deinococcus radiodurans*, *Gelidibacter algens*, *Brassica napus*, *Arabidopsis thaliana*, *Staphylococcus aureus*, *Lactobacillus plantarum*, and *Cyanobacterium aponinum* are given in Table 2. The active sites contain a total of 32 residues. Most of the active site residues are completely identical and conserved across all the sequences. Substitutions of minimum one amino acid and maximum of six amino acids occur across the nine PSY sequences, which are from organisms of various domains of life and grow at different growth temperatures. The number of residues involved in substrate binding, substrate Mg^{2+} binding, active lid-formation, and catalysis of reaction is 18, 10, 9, and 15, respectively. The residues involved in these four functional roles are not mutually exclusive. Many of the residues are involved in more than one role. Two aspartate-rich regions DXXXD are conserved in all the sequences. The total number of active site residues involved in different functions with the presence of two

aspartate rich DXXXD regions have been previously reported to be constant in *Brassica napus* (López-Emparán et al. 2014), *Plasmodium* sp. and other protozoa (Agarwal et al. 2015), *Dunaliella* (Lao et al. 2011), and *Triticum aestivum* (Flowerika et al. 2016). However, there were some differences in the amino acid residues present. The only exception was found in the case of *Chlorella* PSY which contained three aspartate rich DXXXD regions instead of two (Li et al. 2015). All the active site residues of *Meiothermus taiwanensis* are 100% similar in *Meiothermus ruber*. In *M. silvanus*, an alanine residue replaces a glycine residue in RP PSY (227 RP) which is a similar amino acid. This residue is involved in substrate binding and catalysis. In *Thermus thermophilus*, a total of 5 active site residues are different. The methionine (153), glycine (227), aspartic acid (46), serine (256), and isoleucine (132) residues in RP PSY are replaced by leucine, alanine, glutamate, histidine, and methionine residues respectively in *T. thermophiles* PSY. A glutamate residue replaces the aspartic acid (46) residue in RP PSY in *Gelidibacter algens*, *Cyanobacterium aponinum*, and *Brassica napus*. This particular position (second amino acid of the first aspartate-rich region) can be occupied by any of the two negatively charged amino acids. An unusual feature of the substrate Mg2+ binding residues (the second aspartate region) in *Gelidibacter algens* is the replacement of valine (162 RP), glycine (163 RP), and glutamate (164 RP) by leucine (175), lysine (176), and alanine (177). Few residues in RP like alanine (17), methionine (153), glycine (227), alanine (228), alanine (47), histidine (15), threonine (18), alanine (255), serine (256), leucine (257), isoleucine (132), methionine (153), alanine (228), and alanine (232) are prone to replacement in all the sequences except species of the genus *Meiothermus*.

Comparative analysis of the three-dimensional structures of phytoene synthases

For homology modeling of the three-dimensional (3-D) structures of phytoene synthases, PDB files given in Additional file 1: Table S2 containing crystallographic data of 3-D structures of proteins from the Protein Data Bank (Parasuraman 2012) were used for multiple sequence alignment. The three-dimensional structure of *Meiothermus taiwanensis* strain RP is given in Fig. 5 and that of *Gelidibacter algens*, *Thermus thermophiles*, *Meiothermus ruber*, and *Brassica napus* are given in Additional file 1: Fig. S1. The parameters used for quality analysis of the predicted 3-D structures of the phytoene synthases were Z score analysis from QMEAN4, error analysis and confidence of the models from Robetta, Ramachandran plot analysis from PROCHECK, QMEANDisCo Global values, LGscore, MaxSub values, Verify3D, and Errat values; and the Z value mean

Table 2 The substrate binding pocket, substrate Mg^{2+} binding site, active lid-residues, catalytic residues, and aspartate-rich regions of the phytoene synthases from *Meiothermus taiwanensis* strain RP, *M. ruber*, *M. silvanus*, *Thermus thermophilus*, *Deinococcus radiodurans*, *Gelidibacter algens*, *Brassica napus*, *Staphylococcus aureus*, and *Cyanobacterium aponinum*. The residues which are different from strain RP are written in boldface

Organisms	Substrate binding pocket residues	Substrate Mg2+ binding site residues	Active site lid-residues	Catalytic residues	Aspartate rich regions
Meiothermus taiwanensis strain RP	Ala(17), Phe (19), Tyr (38), Arg(42), Asp(45), Asp (49), Tyr(120), Ala (125), Gly (129), Gly (150), Met (153), Arg (160), Asp (161), Glu (164), Asp (165), Arg (170), Gly (227), Ala (228)	Asp (45), Asp (46), Ala (47), Val (48), Asp (49), Asp (161), Val (162), Gly (163), Glu (164), Asp (165)	His (15), Ser (16), Ala (17), Thr (18) , Phe (19), Arg (254), Ala (255), Ser (256), Leu (257)	Phe (19), Tyr (38), Asp (45), Asp (49), Tyr (120), Val (128), Ile (132), Met (153), Arg (160), Asp (161), Asp (165), Arg (170), Gly (227), Ala (228), Ala (232)	Asp (45), Asp (46), Ala (47), Val (48), Asp (49), Asp (161), Val (162), Gly (163), Glu (164), Asp (165)
Meiothermus ruber	Ala(19), Phe (21), Tyr (40), Arg(44), Asp(47), Asp (51), Tyr(122), Ala (127), Gly (131), Gly (152), Met (155), Arg (162), Asp (163), Glu (166), Asp (167), Arg (172), Gly (229), Ala (230)	Asp (47), Asp (48), Ala (49), Val (50), Asp (51), Asp (163), Val (164), Gly (165), Glu (166), Asp (167)	His (17), Ser (18), Ala (19), Thr (20) , Phe (21), Arg (256), Ala (257), Ser (258), Leu (259)	Phe (21), Tyr (40), Asp (47), Asp (51), Tyr (123), Val (131), Ile (134), Met (155), Arg (162), Asp (163), Asp (167), Arg (172), Gly (229), Ala (230), Ala (234)	Asp (47), Asp (48), Ala (49), Val (50), Asp (51), Asp (163), Val (164), Gly (165), Glu (166), Asp (167)
Meiothermus silvanus	Ala(17), Phe (19), Tyr (38), Arg(42), Asp(45), Asp (49), Tyr(120), Ala (125), Gly (129), Gly (153), Met (156), Arg (163), Asp (164), Glu (167), Asp (168), Arg (173), **Ala(230)**, Ala (231)	Asp (45), Asp (46), Ala (47), Val (48), Asp (49), Asp (164), Val (165), Gly (166), Glu (167), Asp (168)	His (15), Ser (16), Ala (17), Thr (18) , Phe (19), Arg (257), Ala (258), Ser (259), Leu (260)	Phe (19), Tyr (38), Asp (45), Asp (49), Tyr (120), Val (128), Ile (132), Met (156), Arg (163), Asp (164), Asp (168), Arg (173), **Ala (230)**, Ala (231), Ala (235)	Asp (45), Asp (46), Ala (47), Val (48), Asp (49), Asp (164), Val (165), Gly (166), Glu (167), Asp (168)
Thermus thermophilus	Ala(17), Phe (19), Tyr (38), Arg(42), Asp(45), Asp (49), Tyr(120), Ala (125), Gly (129), Gly (150), **Leu (153)**, Arg (160), Asp (161), Glu (164), Asp (165), Arg (170), **Ala(227)**, Ala (228)	Asp (45), **Glu (46)**, Ala (47), Val (48), Asp (49), Asp (161), Val (162), Gly (163), Glu (164), Asp (165)	His (15), Ser (16), Ala (17), Thr (18) , Phe (19), Arg (254), Ala (255), **His(256)**, Leu (257)	Phe (19), Tyr (38), Asp (45), Asp (49), Tyr (120), Val (128), **Met (132)**, **Leu (153)**, Arg (160), Asp (161), Asp (165), Arg (170), **Ala (227)**, Ala (228), Ala (232)	Asp (45), **Glu(46)**, Ala (47), Val (48), Asp (49), Asp (161), Val (162), Gly (163), Glu (164), Asp (165)
Deinococcus radiodurans	**Lys (51)**, Phe (53), Tyr (72), Arg(76), Asp(79), Asp(83), Tyr(157), Ala (162), Gly (166), Gly (189), Met (192), Arg (199), Asp (200), Glu (203), Asp (204), Arg (209), **Leu (271)**, Ala (272)	Asp (79), Asp (80), **Ile (81)**, Val (82), Asp (83), Asp (200), Val (201), Gly (202), Glu (203), Asp (204)	His (49), Ser (50), **Lys (51)**, Thr (52) , Phe (53), Arg (298), Ala (299), **Tyr (300)**, **Val (301)**	Phe (53), Tyr (72), Asp (79), Asp (83), Tyr (157), Val (165), **Val (169)**, Met (192), Arg (199), Asp (200), Asp (204), Arg (209), **Lys (271)**, Ala (272), Ala (276)	Asp (79), Asp (80), **Ile (81)**, Val (82), Asp (83), Asp (200), Val (201), Gly (202), Glu (203), Asp (204)
Gelidibacter algens	**Thr (22)**, Phe (24), Tyr (43), Arg(47), Asp(50), Asp (54), Tyr(128), Ala (133), Gly (137), Gly (163), **Phe (166)**, Arg (173), Asp (174), **Ala (177)**, Asp (178), Arg (184), **Phe (227)**, **Gly (228)**	Asp (50), **Glu (51)**, **Ile (52)**, Val (53), Asp (54), Asp (174), **Leu (175)**, **Lys (176)**, **Ala (177)**, Asp (178)	**Tyr (20)**, Ser (21), **Thr (22)**, **Ser (23)** , Phe (24), Arg (255), **Ile (256)**, **Arg (257)**, **Val (258)**	Phe (24), Tyr (43), Asp (50), Asp (54), Tyr (128), Val (136), **Cys (140)**, Phe (166), Arg (173), Asp (174), Asp (178), Arg (184), **Phe(227)**, **Gly (228)**, Ala (232)	Asp (50), **Glu (51)**, **Ile (52)**, Val (53), Asp (54), Asp (174), **Leu (175)**, **Lys (176)**, **Ala (177)**, Asp (178)
Staphylococcus aureus	**Lys(20)**, Phe (22), Tyr (41), Arg(45), Asp(48), Asp (52), Tyr(129), Ala (134), Gly (138), Gly (161), **Lys(164)**, Arg (171), Asp (172), Glu (175), Asp (176), Arg (181), **Pro (239)**, **Ile (240)**	Asp (48), Asp (49), **Ser (50)**, **Ile (51)**, Asp (52), Asp (172), Val (173), Gly (174), Glu (175), Asp (176)	His (18), Ser (19), **Lys (20)**, **Ser (21)** , Phe (22), Arg (265), **Val (266)**, **Phe (267)**, **Val(268)**	Phe (22), Tyr (41), Asp (48), Asp (52), Tyr (129), Val (137), **Leu (141)**, **Leu (164)**, Arg (170), Asp (171), Asp (175), Arg (181), **Pro (239)**, **Ile(240)**, Ala (244)	Asp (48), Asp (49), **Ser (50)**, **Ile (51)**, Asp (52), Asp (172), Val (173), Gly (174), Glu (175), Asp (176)
Cyanobacterium aponinum	**Lys(33)**, Phe (35), Tyr (54), Arg(58), Asp(61), Asp (65), Tyr(139), Ala (144), Gly (148), Gly (182), **Asn (185)**, Arg (192), Asp (193), Glu (196), Asp (197), Arg (202), **Trp (260)**, **Pro (261)**	Asp (61), **Glu (62)**, **Leu (63)**, Val (64), Asp (65), Asp (193), Val (194), Gly (195), Glu (196), Asp (197)	**Tyr(31)**, Ser (32), **Lys (33)**, Thr(34) , Phe (35), Arg (287), Ala(288), **Phe (289)**, **Val(290)**	Phe (35), Tyr (54), Asp (61), Asp (65), Tyr (139), Val (147), **Ser(151)**, Asn(185), Arg (192), Asp (193), Asp (197), Asp (202), Gly (227), **Trp (260)**, **Pro (261)**	Asp (61), **Glu (62)**, **Leu (63)**, Val (64), Asp (65), Asp (193), Val (194), Gly (195), Glu (196), Asp (197)
Brassica napus	**Lys(148)**, Phe (150), Tyr (169), Arg(173), Asp(176), Asp (180), Tyr(254), Ala (259), Gly (263), Gly (291), **Asn (294)**, Arg (301), Asp (302), Glu (305), Asp (306), Arg (311), **Trp(369)**, **Pro (370)**	Asp (176), **Glu (177)**, **Leu (178)**, Val (179), Asp (180), Asp (302), Val (303), Gly (304), Glu (305), Asp (306)	**Tyr(146)**, **Ala (147)**, **Lys (148)**, Thr (149) , Phe (150), Arg (396), Ala (397), **Tyr (398)**, **Val (399)**	Phe (150), Tyr (169), Asp (176), Asp (180), Tyr (254), Val (262), **Ser (266)**, **Asn (294)**, Arg (301), Asp (302), Asp (306), Arg (311), **Trp (369)**, **Pro (370)**, **Ser (374)**	Asp (176), **Glu (177)**, **Leu (178)**, Val (179), Asp (180), Asp (302), Val (303), Gly (304), Glu (305), Asp (306)
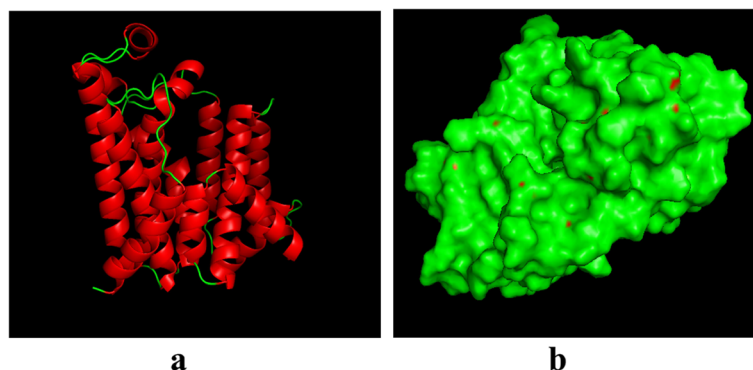
Fig. 5 The three-dimensional structure of the phytoene synthase of *Meiothermus taiwanensis* strain RP through two different view modes. **a** Cartoon view showing the secondary structures. **b** Surface view. The helix chains are indicated by red color, and the coils are indicated by green color

obtained from PROVE. The values obtained for each parameter prove the validity of the models and indicate their high quality (data given in Additional file 1: Figs. S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, S12, S13, S14, S 15, and S16, and Table S3). The 3-D structures of the phytoene synthases of *Meiothermus taiwanensis* strain RP, *M. ruber*, *Thermus thermophilus*, *Brassica napus*, and *Gelidibacter algens* were submitted to the Protein Model Databank with the accession numbers PM0082187, PM0082192, PM0082193, PM0082188, and

PM0082190 respectively. The three-dimensional structures of the phytoene synthases from *Meiothermus ruber*, *Thermus thermophilus*, *Gelidibacter algens*, and *Brassica napus* were superimposed on *Meiothermus taiwanensis* strain RP PSY with RMSD (Root Mean Square Deviation) values of 0.596, 0.859, 1.562, and 1.026 respectively (Fig. 6). The phytoene synthase of the thermophiles *Meiothermus ruber* and *Thermus thermophilus* are structurally most similar to RP. The superposition image with *Meiothermus ruber* (Fig. 6a) is the most
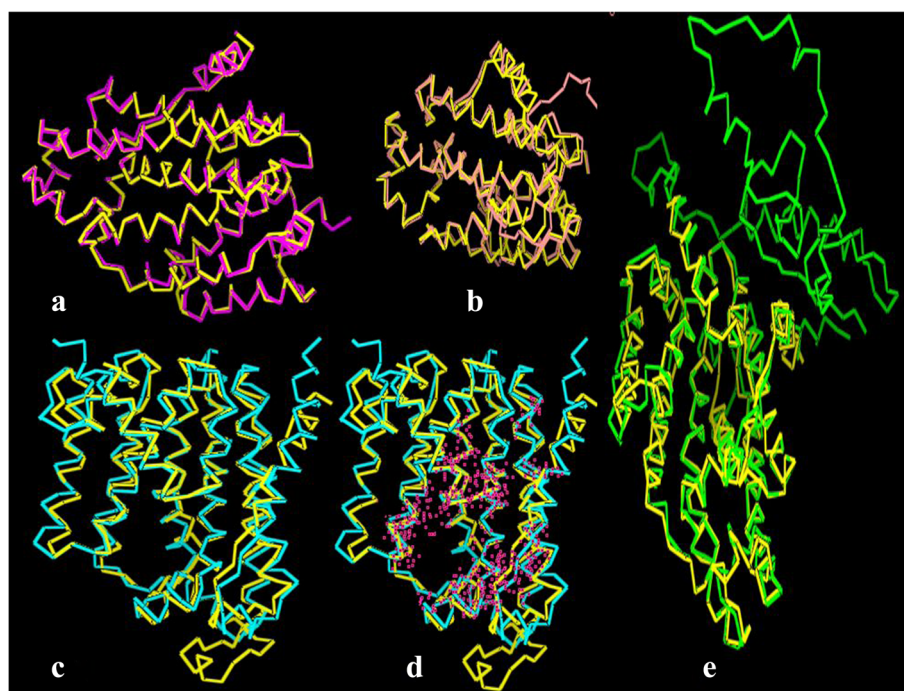
Fig. 6 Superposition images of the three-dimensional structures of the phytoene synthases of different organisms with that of *Meiothermus taiwanensis* strain RP (yellow). **a** *Meiothermus ruber* (pink), **b** *Thermus thermophilus* (tint), **e** *Brassica napus* (green) **c**, **d** *Gelidibacter algens* (blue). Both **c** and **d** are the same superpositions; however, in d, the active site region is denoted with pink dots

perfect. In the case of *Thermus thermophilus* (Fig. 6b), a part of the structure was not superposed as the sequence is greater in length than RP. However, the non-superimposed part does not have any active site residues. Huge areas of the phytoene synthase of *Brassica napus* (Fig. 6e) could not be superimposed; those areas did not contain any active site residues. The sequence length and molecular weight of *B. napus* phytoene synthase are much higher than that of RP. However, the superimposed areas are almost similar in structure. The psychrophilic phytoene synthase of *Gelidibacter algens* (Fig. 6c, d) is most dissimilar to RP structurally. In Fig. 6d, the area covered by the active site residues is marked with pink spots. Though there are a lot of structural differences between *G. algens* and *M. taiwanensis* strain RP, most of them occur outside the active site region. Overall, despite having differences in structures, the enzyme phytoene synthase tries to retain its active site region structure, irrespective of its taxonomic affiliation or the growth temperature range of its source organism. Superposition studies of the three-dimensional structures of *Plasmodium falciparum* and *Thermotoga maritima* PSY (Agarwal et al. 2015) showed that amino acid differences affected 3-D structures but did not affect active sites. The regions with differences in structure are regions outside the catalytic region, and they may play a role in maintaining the properly folded state under various internal and external environmental conditions, such as differences in growth temperature of the organisms, or localization within the cell.

Conclusion

Meiothermus taiwanensis RP PSY was found to be one of the smallest PSY among the tested sequences with low instability index and high aliphatic index, which indicate its stability and thermophilicity. In some cases, intragenus patterns were found, for example, in sequence size and molecular weight, active sites, and structural aspects, while in some cases like pI values and GRAVY values, no patterns were found. There is a marked difference in the molecular weight and instability index of RP PSY and plant-origin PSYs, which lie at two extremities of the graph with the forming having one of the smallest values and the latter having the highest value. There is a significant difference in the percentages of some amino acid residues between the psychrophilic *G. algens* PSY and the thermophilic *M. taiwanensis*. Amino acids like alanine, arginine, glutamine, and glycine may be required in higher numbers for the stability of phytoene synthases at high temperatures, whereas, amino acids like asparagine, aspartic acid, histidine, lysine, and phenyl alanine may be required for maintenance of structural flexibility at low temperatures. Among the two temperature groups (thermophilic and psychrophilic) of PSY, the

differences in some of the properties like amino acid percentage composition, aliphatic index, and 3-D structures may be clearly attributed to the differences of their growth temperatures, while other properties like sequence dissimilarity and differences in active site composition can also be due to factors other than their temperature differences. Interestingly, the differences among structures at the secondary structure level are negligible, as all of them contain high amounts of alpha-helix. Most of the PSY active site residues are conserved across all the sequences with slightly more differences between *G. algens* and RP PSY. In the folded state, the active site regions of the PSYs are almost similar, though there are dissimilarities outside the region. Overall, this study provides structural and functional information which can be used in purification, genetic engineering, and expression of RP PSY with improved stability under different temperature conditions and in different organisms for carotenoid production without compromising the activity of the enzyme.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s13213-020-01558-9.

Additional file 1: Fig. S1 The three-dimensional structure of the phytoene synthase of a) *Gelidibacter algens,* b) *Thermus thermophilus,* c) *Meiothermus ruber* and d) *Brassica napus.* The helix chains are indicated by red color and the coils are indicated by green color. Fig. S2 The error estimates of the three dimensional model of *Meiothermus taiwanensis* strain RP phytoene synthase. Fig. S3 Z score value (<1) of the predicted 3D model of the phytoene synthase of *Meiothermus taiwanensis* strain RP shown as a red star. (If Z score value is >5, the quality analysis fails, if the value is <1 the quality of the structure is good.) Graph is obtained from Qmean server. Fig. S4 Plot statistics of the Ramachandran plot of the predicted 3D model of the phytoene synthase of *Meiothermus taiwanensis* strain RP as obtained from PROCHECK, based on an analysis of 118 structures of resolution of at least 2.0 Angstroms (residues in most favoured regions [A,B,L]: 94.2%, residues in additional allowed regions: [a, b, l, p] 5.4%, residues in generously allowed regions [~a, ~b, ~l, ~p]: 0.4% and residues in disallowed regions: 0.0%). Glycine residues are shown as triangles. A good quality model has more than 90 % residues in favoured regions. Fig. S5 The error estimates of the three dimensional model of *Meiothermus ruber* phytoene synthase. Fig. S6 Z score value (<1) of the predicted 3D model of the phytoene synthase of *Meiothermus ruber* shown as a red star. (If Z score value is >5, the quality analysis fails, if the value is <1 the quality of the structure is good.) Graph is obtained from Qmean server. Fig. S7 Plot statistics of the Ramachandran plot of the predicted 3D model of the phytoene synthase of *Meiothermus ruber* as obtained from PROCHECK, based on an analysis of 118 structures of resolution of at least 2.0 Angstroms (residues in most favoured regions [A, B, L]: 95.0%, residues in additional allowed regions [a, b, l, p]: 4.1 %, residues in generously allowed regions [~a, ~b, ~l, ~p] and 0.4% residues in disallowed regions). Glycine residues are shown as triangles. A good quality model has more than 90 % residues in favoured regions. Fig. S8 The error estimates of the three dimensional model of *Thermus thermophilus* phytoene synthase. Fig. S9 Z score value (<1) of the predicted 3D model of the phytoene synthase of *Thermus thermophilus* shown as a red star. (If Z score value is >5, the quality analysis fails, if the value is <1 the quality of the structure is good.) Graph is obtained from Qmean server. Fig. S10 Plot statistics of the Ramachandran plot of the predicted 3D model of the phytoene synthase of *Thermus thermophilus* as obtained from PROCHECK, based on an analysis of 118 structures of resolution of

at least 2.0 Angstroms (residues in most favoured regions [A,B,L] 94.6 %, residues in additional allowed regions [a, b, l, p]: 4.6 %, residues in generously allowed regions [~a, ~b, ~l, ~p]: 0.4% and residues in disallowed regions: 0.4 %). Glycine residues are shown as triangles. A good quality model has more than 90 % residues in favoured regions. Fig. S11 The error estimates of the three dimensional model of *Gelidibacter algens* phytoene synthase. Fig. S12 Z score value (<1) of the predicted 3D model of the phytoene synthase of *Gelidibacter algens* shown as a red star. (If Z score value is >5, the quality analysis fails, if the value is <1 the quality of the structure is good.) Graph is obtained from Qmean server. Fig. S13 Plot statistics of the Ramachandran plot of the predicted 3D model of the phytoene synthase of *Gelidibacter algens* as obtained from PROCHECK ,based on an analysis of 118 structures of resolution of at least 2.0 Angstroms (residues in most favoured regions [A,B,L]: 93.9 %, residues in additional allowed regions [a, b, l, p]: 5.7 %, residues in generously allowed regions [~a, ~b, ~l, ~p]: 0.0% and residues in disallowed regions: 0.4 %). Glycine residues are shown as triangles. A good quality model has more than 90 % residues in favoured regions. Fig. S14 The error estimates of the three dimensional model of *Brassica napus* phytoene synthase. Fig. S15 Z score value (<1) of the predicted 3D model of the phytoene synthase of *Brassica napus* shown as a red star. (If Z score value is >5, the quality analysis fails, if the value is <1 the quality of the structure is good.) Graph is obtained from Qmean server. Fig. S16 Plot statistics of the Ramachandran plot of the the predicted 3D model of the phytoene synthase of *Brassica napus* as obtained from PROCHECK based on an analysis of 118 structures of resolution of at least 2.0 Angstroms (residues in most favoured regions [A, B, L]: 91.4 %, residues in additional allowed regions [a, b, l, p]: 8.1 %, residues in generously allowed regions [~a, ~b, ~l, ~p]: 0.0 % and residues in disallowed regions: 0.5 %). Glycine residues are shown as triangles. A good quality model has more than 90 % residues in favoured regions. Table S1 The molecular weight, instability index, pI value, aliphatic index, and grand average of hydropathicity (GRAVY) of the all the phytoene synthase sequences. Table S2 Details of PDB files containing crystallographic data of 3-D structures of proteins from the Protein Data Bank used for multiple sequence alignment during 3-D structure modelling. Table S3 The evaluation of the quality of the three dimensional models of the phytoene synthases from different organisms. The quality of the models is considered as worst when the value in the first two columns (confidence of the models from Robetta and QMEAN-DisCo Global values) is 0.0 and becomes better with increase in value up to 1.0. The LGscore of the models is obtained from ProQ (>1.5 fairly good model, >2.5 very good model, >4 extremely good model). The fourth column contains the MaxSub values of the models obtained from ProQ (>0.1: fairly good model, >0.5: very good model, >0.8: extremely good model). The fifth column contains the Verify3D values of the models, where the percentage of amino acids scoring >=0.2 in their 3D/1D profile is determined. The model is valid if this percentage is >=80 %. The Errat quality factor is considered as good when the values are >=95 %. The Z value mean is obtained from PROVE (>5: the quality analysis fails, <1: the quality of the structure is good). (DOCX 1865 kb)

Research involving human participants and/or animals
N/A

Authors' contributions
TM and SKM made substantial contributions to design of research, experiments, acquisition of data and analysis, interpretation of data and drafting the manuscript. The authors read and approved the final manuscript.

Ethics approval and consent to participate
All prevailing local, national and international regulations and conventions, and normal ethical scientific practices have been respected.

Competing interests
The authors declare that they have no competing interests.

References
Agarwal S, Sharma V, Phulera S et al (2015) Structural insights into a key carotenogenesis related enzyme phytoene synthase of *P. falciparum*: a novel drug target for malaria. Syst Synth Biol. https://doi.org/10.1007/s11693-015-9168-8

Ahrazem O, Diretto G, Argandoña Picazo J et al (2019) The specialized roles in carotenogenesis and apocarotenogenesis of the phytoene synthase gene family in saffron. Front Plant Sci 10:249. https://doi.org/10.3389/fpls.2019.00249

Benkert P, Künzli M, Schwede T (2009) QMEAN server for protein model quality estimation. Nucleic Acids Res 37:W510–W514. https://doi.org/10.1093/nar/gkp322

Bowie JU, Lüthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional stucture. Science 53:164–170. https://doi.org/10.1126/science.1853201

Bowman JP, McCammon SA, Brown JL et al (1997) *Psychroserpens burtonensis* gen. nov., sp. nov., and *Gelidibacter algens* gen. nov., sp. nov., psychrophilic bacteria isolated from Antarctic lacustrine and sea ice habitats. Int J Syst Bacteriol 47:670–677. https://doi.org/10.1099/00207713-47-3-670

Britton G (2008a) Functions of intact carotenoids. In: Britton G, Liaaen-Jensen S, Pfander H (eds) Carotenoids vol. 4: natural functions. Springer Science & Business Media, pp 189–212

Britton G (2008b) Functions of carotenoid metabolites and breakdown products. In: Britton G, Liaaen-Jensen S, Pfander H (eds) Carotenoids vol. 4: natural functions. Springer Science & Business Media, pp 309–324

Britton G, Liaaen-Jensen S, Pfander H (2004) Natural carotenoids. In: Carotenoids: Handbook, pp 42–45

Burgess ML, Barrow KD, Gao C et al (1999) Carotenoid glycoside esters from the thermophilic bacterium *Meiothermus ruber*. J Nat Prod 62:859–863. https://doi.org/10.1021/np980573d

Castrignanò T, De Meo PD, Cozzetto D et al (2006) The PMDB protein model database. Nucleic Acids Res 34:D306–D309. https://doi.org/10.1093/nar/gkj105

Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2:1511–1519. https://doi.org/10.1002/pro.5560020916

Crawford AC, Stefanova K, Lambe W et al (2011) Functional relationships of phytoene synthase 1 alleles on chromosome 7A controlling flour colour variation in selected Australian wheat genotypes. Theor Appl Genet 123:95–108. https://doi.org/10.1007/s00122-011-1569-9

Dibari B, Murat F, Chosson A et al (2012) Deciphering the genomic structure, function and evolution of carotenogenesis related phytoene synthases in grasses. BMC Genomics 13:221. https://doi.org/10.1186/1471-2164-13-221

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340

Eswar N, Ramakrishnan C, Srinivasan N (2003) Stranded in isolation: structural role of isolated extended strands in proteins. Protein Eng Des Sel 16:331–339. https://doi.org/10.1093/protein/gzg046

Flowerika AA, Kumar J et al (2016) Characterization and expression analysis of phytoene synthase from bread wheat (*Triticum aestivum* L.). PLoS One 11: e0162443. https://doi.org/10.1371/journal.pone.0162443

Fu X, Feng C, Wang C et al (2014) Involvement of multiple phytoene synthase genes in tissue- and cultivar-specific accumulation of carotenoids in loquat. J Exp Bot 65:4679–4689. https://doi.org/10.1093/jxb/eru257

Gasteiger E, Hoogland C, Gattiker A et al (2005) Protein identification and analysis tools on the ExPASy server. In: The proteomics protocols handbook. Humana Press, pp 571–607

Geourjon C, Deléage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 11:681–684

Griffiths E, Gupta RS (2007) Identification of signature proteins that are distinctive of the Deinococcus-Thermus phylum. Int Microbiol 10:201–208

Guruprasad K, Reddy BV, Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng 4:155–161

Han Y, Zheng QS, Wei YP et al (2015) In silico identification and analysis of phytoene synthase genes in plants. Genet Mol Res 14:9412–9422. https://doi.org/10.4238/2015.August.14.5

Ikai A (1980) Thermostability and aliphatic index of globular proteins. J Biochem 88:1895–1898

Iwata-Reuyl D, Math SK, Desai SB, Poulter CD (2003) Bacterial phytoene synthase: molecular cloning, expression, and characterization of *Erwinia herbicola* phytoene synthase. Biochemistry 42:3359–3365. https://doi.org/10.1021/bi0206614

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Bioinformatics 8:275–282. https://doi.org/10.1093/bioinformatics/8.3.275

Joseph S, Desai P, Ji Y et al (2012) Comparative analysis of genome sequences covering the seven *Cronobacter* species. PLoS One 7:e49455. https://doi.org/10.1371/journal.pone.0049455

Kadono T, Kira N, Suzuki K et al (2015) Effect of an introduced phytoene synthase gene expression on carotenoid biosynthesis in the marine diatom *Phaeodactylum tricornutum*. Mar Drugs 13:5334–5357. https://doi.org/10.3390/md13085334

Kato S, Soshino M, Takaichi S et al (2017) Suppression of the phytoene synthase gene (EgcrtB) alters carotenoid content and intracellular structure of *Euglena gracilis*. BMC Plant Biol 17:125. https://doi.org/10.1186/s12870-017-1066-7

Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 32:W526–W531. https://doi.org/10.1093/nar/gkh468

Kishino H, Hasegawa M (2001) Maximum Likelihood. Encycl Genet:1157–1160. https://doi.org/10.1006/RWGN.2001.0803

Klassen JL (2010) Phylogenetic and evolutionary patterns in microbial carotenoid biosynthesis are revealed by comparative genomics. PLoS One 5:e11257. https://doi.org/10.1371/journal.pone.0011257

Kreil DP, Ouzounis CA (2001) Identification of thermophilic species by the amino acid compositions deduced from their genomes. Nucleic Acids Res 29:1608–1615. https://doi.org/10.1093/nar/29.7.1608

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054

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

Lao Y-M, Xiao L, Ye Z-W et al (2011) In silico analysis of phytoene synthase and its promoter reveals hints for regulation mechanisms of carotenogenesis in *Duanliella bardawil*. Bioinformatics 27:2201–2208. https://doi.org/10.1093/bioinformatics/btr371

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291. https://doi.org/10.1107/S0021889892009944

Leng X, Wang P, Wang C et al (2017) Genome-wide identification and characterization of genes involved in carotenoid metabolic in three stages of grapevine fruit development. Sci Rep 7:4216. https://doi.org/10.1038/s41598-017-04004-0

Li F, Vallabhaneni R, Wurtzel ET (2008) PSY3, a new member of the phytoene synthase gene family conserved in the *Poaceae* and regulator of abiotic stress-induced root carotenogenesis. Plant Physiol 146:1333–1345. https://doi.org/10.1104/pp.107.111120

Li M, Cui Y, Gan Z et al (2015) Isolation and analysis of the cppsy gene and promoter from *Chlorella protothecoides* CS-41. Mar Drugs 13:6620–6635. https://doi.org/10.3390/md13116620

López-Emparán A, Quezada-Martinez D, Zúñiga-Bustos M et al (2014) Functional analysis of the *Brassica napus* L. phytoene synthase (PSY) gene family. PLoS One 9:e114878. https://doi.org/10.1371/journal.pone.0114878

Marchler-Bauer A, Lu S, Anderson JB et al (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39:D225–D229. https://doi.org/10.1093/nar/gkq1189

McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16:404–405. https://doi.org/10.1093/bioinformatics/16.4.404

Metpally RPR, Reddy BVB (2009) Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: insights into the molecular basis of cold adaptation of proteins. BMC Genomics 10:11. https://doi.org/10.1186/1471-2164-10-11

Moldoveanu SC, David V, Moldoveanu SC, David V (2017) Properties of analytes and matrices determining HPLC Selection. Sel HPLC Method Chem Anal:189–230. https://doi.org/10.1016/B978-0-12-803684-6.00005-6

Mukherjee T, Bose S, Mukhopadhyay SK (2017) Antioxidant properties of the carotenoid extracts of three Deinococcus–Thermus phylum bacteria, *Meiothermus* sp. strains RP and TP and *Thermus* sp. strain YY from Paniphala hot spring, India. Nutrire 42:7. https://doi.org/10.1186/s41110-017-0032-3

Mukherjee T, Bose S, Sen U et al (2016) Genome sequence of the red pigment-forming *Meiothermus taiwanensis* strain RP isolated from paniphala hot spring, India. Genome Announc 4:e00629–e00616. https://doi.org/10.1128/genomeA.00629-16

Nei M, Zhang J (2006) Evolutionary distance: estimation. In: Encyclopedia of Life Sciences. Wiley, Chichester

Paniagua-Michel J, Olmos-Soto J, Ruiz M (2012) Pathways of carotenoid biosynthesis in bacteria and microalgae. In: Microbial carotenoids from bacteria and microalgae. Humana Press, pp 1–12

Parasuraman S (2012) Protein data bank. J Pharmacol Pharmacother 3:351. https://doi.org/10.4103/0976-500X.103704

Phadwal K (2005) Carotenoid biosynthetic pathway: molecular phylogenies and evolutionary behavior of crt genes in eubacteria. Gene 345:35–43. https://doi.org/10.1016/J.GENE.2004.11.038

Pontius J, Richelle J, Wodak SJ (1996) Deviations from standard atomic volumes as a quality measure for protein crystal structures. J Mol Biol 264:121–136. https://doi.org/10.1006/jmbi.1996.0628

Ramachandran S, Kota P, Ding F, Dokholyan NV (2011) Automated minimization of steric clashes in protein structures. Proteins Struct Funct Bioinforma 79:261–270. https://doi.org/10.1002/prot.22879

Rice P, Longden I, Bleasby A (2000) EMBOSS: the European molecular biology open software suite. Trends Genet 16:276–277. https://doi.org/10.1016/S0168-9525(00)02024-2

Sælensminde G, Halskau Ø, Jonassen I (2009) Amino acid contacts in proteins adapted to different temperatures: hydrophobic interactions and surface charges play a key role. Extremophiles 13:11–20. https://doi.org/10.1007/s00792-008-0192-4

Sandmann G (2002) Molecular evolution of carotenoid biosynthesis from bacteria to plants. Physiol Plant 116:431–440. https://doi.org/10.1034/j.1399-3054.2002.1160401.x

Shumskaya M, Bradbury LMT, Monaco RR, Wurtzel ET (2012) Plastid localization of the key carotenoid enzyme phytoene synthase is altered by isozyme, allelic variation, and activity. Plant Cell 24:3725–3741. https://doi.org/10.1105/tpc.112.104174

Smith LJ, Fiebig KM, Schwalbe H, Dobson CM (1996) The concept of a random coil: residual structure in peptides and denatured proteins. Fold Des 1:R95–R106. https://doi.org/10.1016/S1359-0278(96)00046-6

Smith TF, Waterman MS (1981) Identification of common molecular subsequences. J Mol Biol 147:195–197. https://doi.org/10.1016/0022-2836(81)90087-5

Wallner B, Elofsson A (2003) Can correct protein models be identified? Protein Sci 12:1073–1086. https://doi.org/10.1110/ps.0236803

Welsch R, Wüst F, Bär C et al (2008) A third phytoene synthase is devoted to abiotic stress-induced abscisic acid formation in rice and defines functional diversification of phytoene synthase genes. Plant Physiol 147:367–380. https://doi.org/10.1104/pp.108.117028

Zhao D, Zhou C, Sheng Y et al (2011) Molecular cloning and expression of phytoene synthase, lycopene beta-cyclase, and beta-carotene hydroxylase genes in persimmon (*Diospyros kaki L.*) Fruits. Plant Mol Biol Report 29:345–351. https://doi.org/10.1007/s11105-010-0238-5

Publisher's Note