cDNA cloning and expression analysis of the chalcone synthase gene (CHS) from *Polygonum minus*

Nur Diyana Roslan[1,2], Cheng-Seng Tan[1,2], Ismanizan Ismail[1,2] and Zamri Zainal[1,2]*

[1]**School of Biosciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia**
[2]**Institute of Systems Biology (INBIOSIS), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia**

*Corresponding author: zz@ukm.my

Abstract

Flavonoids are secondary metabolites that are present at high levels in the *Polugonum minus*. In recent years, flavonoids have attracted the interest of researchers because of their antioxidant properties .Flavonoids are synthesised via the condensation of 4 coumaryl-CoA and 3 malonyl-CoA by the action of Chalcone *synthase*. A cDNA encoding a chalcone synthase was isolated from the leaves of *Polygonum minus* by rapid amplification of cDNA ends (RACE) and designated *pmCHS* (GenBank accession no. JQ801338). The full-length cDNA of *P. minus pmCHS* was 1472 bp with a 1179 bp open reading frame (ORF) that corresponded to a predicted protein of 392 amino acid deduced protein. *In silico* analysis showed that the calculated molecular weight and theoretical isoelectric point (pI) of *pm*CHS were 43.1 kDa and 5.78, respectively. Several important motifs, such as the product binding site, active site and dimer interface, were also successfully identified from the deduced amino acid sequence. Multiple sequence alignment indicated that the *pm*CHS sequence was highly conserved and shared high sequence identity (>90%) with chalcone synthases from other plants. Gene expression analysis via qRT-PCR showed that *pmCHS* was most highly expressed in the roots, showing a 10-fold increase compared to leaves and a 15-fold increase compared to stems . The specific mechanism underlying the high expression of *pmCHS* in the roots requires further investigation.

Keywords: chalcone synthase, flavonoid, cDNA cloning, *in silico*, *Polygonum minus*.
Abbreviations: RACE, rapid amplification of cDNA ends; PKS, polyketide synthase; CHS, chalcone synthase.

Introduction

Flavonoids are of particular interest in plant physiology as they perform a variety of roles, including pigmentation, pollination, protection of cells against UV-light and other defense mechanism.. Due to their biological properties, flavonoid are believed to have antioxidant, anti-inflamatory, antiviral and antibacterial activities, and they may have a direct protective effect on the coronary and vascular system (Cazarolli et al., 2008). These characteristics place them among the most attractive natural substances that can be isolated from *P.minus.* The enzyme chalcone synthase (CHS, EC 2.3.1.74), catalyses the first committed step of the flavonoid biosynthesis pathway thus regulating flavonoid biosynthesis. Structurally and mechanistically, CHS is the simplest member of the type III PKS superfamily and, functions as a homodimer (monomer $M_w \approx$ 42-45 kDa) (Tropf et al., 1995; Schröder, 1997). This enzyme catalyses the production of 2',4',6',4-tetrahydroxychalcone, also known as naringenin chalcone through the condensation of 4-coumaroyl-CoA and three acetate residues from malonyl-CoA (Stefan and Axel, 2005) (Figure 1). Naringenin chalcone then serves as the precursor molecule for the biosynthesis of anthocyanin pigments, proanthocyanidins, flavones and antimicrobial phytoalexins (Jez et al., 2001; Dao et al., 2011). Over the past few decades

Chalcone synthases have been studied in higher plants. Currently the isolation and characterisation of new *CHSs* from other plant sources is on-going. Some extensively studied *CHS* genes include those from *Oryza sativa*, *Medicago sativa* and *Zea mays* (Lei et al., 2010). Studies have indicated that *CHS* gene expression can be induced by exposure to visible orUV light, elicitors, wounding, low temperature and pathogen infection (Dalkin et al., 1990; Shvarts et al., 1997; Christensen et al., 1998; Schenk et al., 2000). Other studies have shown that CHS genes are differentially expressed in different plant tissues (Li et al., 2010) and are also controlled by the plant circadian clock (Thain et al., 2002). *Polygonum minus* Hud, locally known as kesum in Malaysia, has been reported to have a high total flavonoid content (Maizura et al., 2011). This finding implies that the enzymes involved in flavonoid biosynthesis are highly active in this species. Because CHS catalyses the first reaction in the flavonoid biosynthesis pathway, it was chosen to be the first gene isolated from *P. minus*. This effort is expected to broaden our understanding of *CHS* gene structure and the catalytic mechanism of CHS, which underlies flavonoid synthesis. The isolated *CHS* gene may be used to increase the *in vivo* yield of biologically and pharmacologically active compounds *in vivo*, such as anthocyanins and flavonol which are important for both agricultural and medicinal purposes. In

this paper, we report the isolation of a cDNA encoding chalcone synthase from *P. minus*, designated *pmCHS* and the expression profiling of *pmCHS* in different tissues of *P. minus*.

Result and Discussion

cDNA isolation and sequence analysis

A unigene transcript Cn206 putatively annotated as a CHS was retrieved from the previously established *P. minus* EST database of INBIOSIS, UKM. Blast analysis revealed that the Cn206 partial sequence which has a poly-A tail at its 3' end, exhibited high percentage of sequences identity with *Polygonum cuspidatum* (97%), *Fallopia multiflora* (97%), *Rheum palmatum* (96%) and *Citrus sinensis* (97%). However, this partial sequence lacked approximately 400 bp of sequence from the upstream region of the full-length sequence. A gene-specific primer, 5'GSP, was designed to obtain the missing *pmCHS* sequence (Fig 2). A ~1000 bp 5'RACE product of approximately 1000 bp was isolated and sequenced, and the resulting sequence was aligned and assembled with the Cn206 sequence to obtain the full-length *pmCHS* gene (GenBank accession no. JQ801338). The validity of the *pmCHS* sequence was confirmed by RT-PCR using the primer pairs FullCHS-f and Full-CHS-r. Sequence analysis indicated that the full-length cDNA of *pm*CHS was 1472 bp long, containing a 1179 bp open reading frame (ORF) that corresponded to a deduced protein sequence of 392 amino acid residues. The ORF region was bordered by a 46 bp 5'-untranslated region (UTR) and a 253 bp 3'-UTR. Additionally a putative poly-adenylation signal (AATAA) was identified 211 bp downstream from the stop codon (TAG) in the 3'-UTR (Figure 3). The calculated molecular weight and theoretical isoelectric point (pI) of pmCHS were 43.1 kDa and 5.78, respectively. This result was identical to previous reports in which with CHS was isolated from tree peony (Zhou et al., 2011) and *Physcomitrella patens* (Jiang et al., 2006). The pm*CHS* sequence was highly conserved and shared a maximum sequence identity of 94% with CHSs from *Fallopia multiflora, Rheum palmatum* and *Polygonum cuspidatum*. Together, these data confirm that the *pmCHS* gene encoding a chalcone synthase was successfully isolated from *P. minus.*

Prediction of the structural and physicochemical properties of pmCHS

Protein sequence analysis indicated that *pm*CHS exhibited extremely high sequence identity (>90%) to other plant CHSs, particularly those from the Polygonaceae family. The high conservation of CHS sequences across species was exploited to identify catalytically important residues in the *pm*CHS sequence. Multiple sequence alignment revealed that the catalytic triad (Cys[160], His[303] and Asn[341]) and a highly conserved Phe residue (Phe[216]) were found in the pmCHS sequence (Figure 4). These residues are important for the formation of the CHS active site as well as for the substrate specificity of the CHS enzyme. According to Jez et al. (2000), the Cys residue in the catalytic triad functions as a nucleophile in polyketide formation, whereas the His and Asn residues assist in the decarboxylation of malonyl-CoA. The Phe residue may play a role in substrate orientation at the active site when the polyketide intermediate undergoes elongation. Additionally
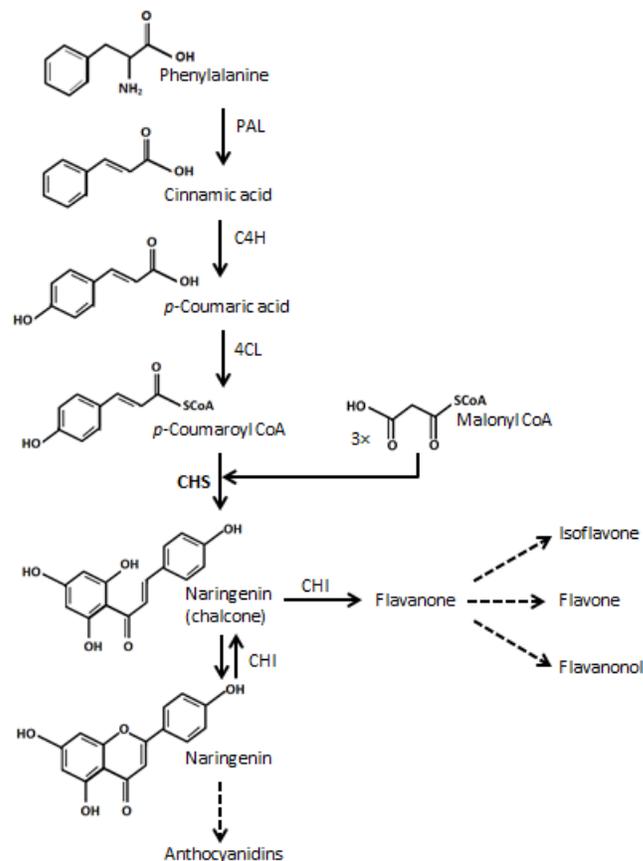
Fig 1. The flavonoid biosynthesis pathway.   C4H, cinnamate-4-hydroxylase; 4CL, 4-coumarate-CoA ligase; CHI, chalcone isomerase; CHS, chalcone synthase; PAL, phenylalanine ammonia lyase

seven amino acid residues (Met[138], Thr[132], Phe[216], Ile[255], Gly[257], Phe[266] and Pro [376]) involved in the formation of the cyclisation pocket, a malonyl-CoA binding motif (V[314]EAKLGLK-EEKLKATRQ) and a highly conserved CHS signature sequence (G[373]FGPG) (Suh et al., 2000) were also identified in the *pmCHS* sequence. All of these amino acids are important for the catalytic activity of chalcone synthase. Subcellular localisation analysis of pmCHS was performed using the TargetP 1.1 server. The results suggested that *pm*CHS is located in the cytoplasm. This finding is consistent with the absence of a signal peptide in the pmCHS protein sequence, as determined by analysis with the online SignalP 4.0 Server by Hrazdina (1992) which showed that flavonoids are synthesised in the cytoplasm, further supporting this result. In addition, analysis using the TMHMM Server v.2.0 and ProtScale predicted that *pm*CHS was a hydrophilic protein with no transmembrane structures; this observation implies that *pm*CHS catalyses the biosynthesis of naringenin chalcone directly in the cytoplasm.

Phylogenetic analysis

A phylogenetic tree was created by the Neighbour-Joining method using *pm*CHS and other CHS coding sequences. CHS sequences from bacterial sources (i.e., *Actinoplane* CHS,

Mycobacterium CHS, *Rhizobium* CHS and *Agrobacterium* CHS) were used as outgroups. In the phylogenetic analysis based on 1000 replicates, the phylogeny tree *was* diverged into three major groups, and the classification of each monophyletic group was supported with a high bootstrap value. Bacterial CHSs were first clustered into a group separate from the plant CHSs. As CHS expected, angiosperms and gymnosperms clustered together in the plant group. CHS proteins from gymnosperms were grouped immediately before the angiosperm cluster, while angiosperm proteins were further divided into two distinct clades, specifically dicots and monocots. As expected *P. minus* CHS was located in the dicot group and was closely related to the CHSfrom other members of the Polygonaceae family, (Figure 5). This result was supported by Mallika et al., (2011) in which plant and bacterial CHS diverged into different groups. Similar results were obtained using the Minimum Evolution and UPGMA methods. These results suggest that CHS are well conserved among plants of different groups and have distinct phylum specificity.

Three-dimensional model of pmCHS

A homology-based 3-D protein structure of *pm*CHS was constructed using SWISS-MODEL based on the 1.56 Å-resolution crystal structure of alfalfa CHS2 (1bq6), which shared an 81.14% identity with the query sequence (Fig 6). Protein structure modeling revealed that *pm*CHS monomer consisted of two structural domains. In the upper domain, an N-terminal α-helix and a protruding tight loop important for protein dimerisation were observed (Jez et al., 2001). Most residues that made up the dimer interface (purple) were located within the upper domain, which produced a relatively flat surface in the pmCHS monomer. In the lower domain, a large substrate binding site required for chalcone formation was connected to the CoA-binding tunnel and the active site (green) was buried in the cleft formed between the two domains. The quality of the pmCHS protein model was confirmed using Ramachandran Plot statistics (Morris et al., 1992). The pmCHS protein model exhibited a good fit with the reference geometry with 92.3% of non-glycine and non-proline residues having the most favourable combinations of phi-psi values. None of the non-glycine residues were situated in the disallowed regions. These results suggest that the *pm*CHS protein model represents a valid stereochemical conformation and that the protein adopts a molecular structure typical of other chalcone synthases.

Expression level of pmCHS in different tissues

To understand the spatial regulation of the *pmCHS* gene in *P. minus, pmCHS* gene expression in different tissues of *P. minus* was measured using relative quantitative real time PCR (qRT-PCR). qRT-PCR analysis showed that the expression of *pmCHS* was significantly higher in the roots than in other tissues, and it was approximately 15-fold and 10-fold more highly expressed in the roots than in the leaves and stems, respectively (Fig 8). This result is consistent with the findings of Li et al. (2010) who reported that most of the flavonoid biosynthesis genes they studied, including *CHS*, were highly expressed in the lower parts of the plant, especially the roots. Examination of the subcellular localisation of *CHS* in the *Arabidopsis* root demonstrated that high levels of *CHS* transcripts were present
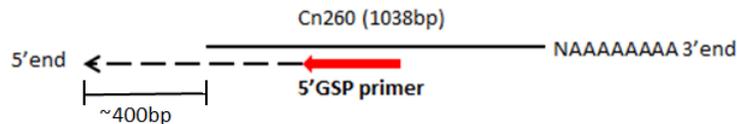
Fig 2. A schematic representation of the EST containing a fragment of *pmCHS* (Cn260) with a polyadenylated tail. The red arrow indicates the 5'GSP primer, which was used to obtain the remaining ~400bp of sequence in the upstream region to complete the full-length sequence of *pmCHS*.

in the epidermal and cortex cells of the elongation zone of the root tip (Saslowsky and Winkel, 2001). The expression of *CHS* in plants can be stimulated by wounding, light, nutrient supply, plant-microbe interactions and plant hormones (Dixon and Harrison, 1990). In particular, flavonoid production associated with plant-microbe interactions has been widely studied. It is well established that plants can release a wide range of secondary metabolites into the rhizosphere, or the root-soil interface (Bais et al., 2006). These root exudates are ecologically and physiologically important to the plant because they support beneficial symbiosis between the plant and the microbial community in the vicinity of the root, affect plant resistance to pests and provide advantages over competing plants (Bertin et al., 2003). Several lines of evidence indicate that flavonoids present in the root exudates function as key signalling molecules for several symbiotic and pathogenic plant-microbe interactions (Steinkellner et al., 2007). For example, the mycelial growth, hyphal differentiation and root colonisation of arbuscular mycorrhizal fungi (AMF), which are beneficial root endophytes, are influenced by root flavonoids in several plants (Poulin et al., 1993; Scervino et al., 2005). However, further investigations will be required to verify whether the same interactions occur in and contribute to *pmCHS* expression in the roots of *Polygonum minus*.

Material and methods

Plant material and RNA isolation

P. minus plants originating from Ulu Yam, Selangor, Malaysia were harvested from the UKM *Polygonum minus* plot. Leaves were plucked from the collected plants and frozen in liquid nitrogen for RNA extraction. A total of 5 g of *P. minus* leaves were used for each RNA extraction. Total RNA was isolated as previously described by Roslan et al. (2012) with some modifications. Extracted total RNA was dissolved in diethylpyrocarbonate (DEPC)-treated water. RNA was quantified using Nanodrop ND-1000 spectrophotometer (NanoDrop Technologies, USA) and its integrity was assessed by agarose gel electrophoresis. The RNA was stored at -80˚C until use.

Cloning of full-length pmCHS cDNA by RACE

mRNA was isolated from 1 mg of total RNA using PolyATtract® mRNA Isolation System III (Promega, Catalogue no. Z5300) according to the manufacturer's instructions. Primer3 Input (version 0.4.0) was used to design gene-specific primers (GSP) based on the previously identified *pmCHS* EST sequence Cn260 retrieved from the *P. minus* EST

```
  1                                         acatatacaacattgatcaatcacattccaacatatagatttaggg
 47 atggctccgtcggtcgaacagatcgggaaggcccagcgtgccgagggccccgccaccgtccttgccatcggcaccgccacccctcccaac
     M  A  P  S  V  E  Q  I  G  K  A  Q  R  A  E  G  P  A  T  V  L  A  I  G  T  A  T  P  P  N
137 tgcgtctcccaagccgactaccccgactactacttccgcgtcaccaatagcgagcacatgaccgacctcaaggagaagttccgccgcatg
     C  V  S  Q  A  D  Y  P  D  Y  Y  F  R  V  T  N  S  E  H  M  T  D  L  K  E  K  F  R  R  M
227 tgcgataagtcgatgatcgagaagcgttacatgtacttgacggaggagatcctcaaggagaacccaaacatgtgtgcctacatggagaag
     C  D  K  S  M  I  E  K  R  Y  M  Y  L  T  E  E  I  L  K  E  N  P  N  M  C  A  Y  M  E  K
317 tcactagactcccgccaggacatcgtcgtgacggaggtgccgcgactcggcaaggaggccgcccagaaggcgatcaaggagtggggccag
     S  L  D  S  R  Q  D  I  V  V  T  E  V  P  R  L  G  K  E  A  A  Q  K  A  I  K  E  W  G  Q
407 gccaagtccaagatcaccacgtcatcatgtgcaccacctccggagtcgacatgcccggggcagactaccagctcaccaagctcctcggc
     A  K  S  K  I  T  H  V  I  M  C  T  T  S  G  V  D  M  P  G  A  D  Y  Q  L  T  K  L  L  G
497 ctccggccttccgtcaagcgcttcatgatgtaccagcaaggctgcttcgccggtggcaccgtcctccgtctcgcccaaggatttggcggag
     L  R  P  S  V  K  R  F  M  M  Y  Q  Q  G  C  F  A  G  G  T  V  L  R  L  A  K  D  L  A  E
587 aacaacaagggcgcccgagtcctcgtggtgtgctccgagatcaccgcggtttgcttccgtgggcccaccgacaccccacttggactccatg
     N  N  K  G  A  R  V  L  V  V  C  S  E  I  T  A  V  C  F  R  G  P  T  D  T  H  L  D  S  M
677 gtgggccaggccttgttcggcgatggagctggagcggtcatcattggggctgaccccgacttgtccatcgaaaagcccatcttcgagctc
     V  G  Q  A  L  F  G  D  G  A  G  A  V  I  I  G  A  D  P  D  L  S  I  E  K  P  I  F  E  L
767 gtgtggaccgcccaaaccatcctgcctgactccgaaggcgccatcgatggccatttgcgcgaagttgggctcaccttccatccccttaag
     V  W  T  A  Q  T  I  L  P  D  S  E  G  A  I  D  G  H  L  R  E  V  G  L  T  F  H  P  L  K
857 gacgttcccgggttgatctctaagaacattgacaagtctcttaaggaggctttctctcctcttaacatcgacgattggaactccctcttc
     D  V  P  G  L  I  S  K  N  I  D  K  S  L  K  E  A  F  S  P  L  N  I  D  D  W  N  S  L  F
947 tggattgcacaccccggtggtccagcgatcctagacaagtggaggccgaagctggggctgaagcaggagaagctcaaggcgacgaggcaa
     W  I  A  H  P  G  G  P  A  I  L  D  Q  V  E  A  K  L  G  L  K  Q  E  K  L  K  A  T  R  Q
1037 gtgttgaacgactatggaaacatgtcgagtgcgtgtgtgctcttcattatggatgagatgaggaagaagtcggtcgagaatgggcatgca
      V  L  N  D  Y  G  N  M  S  S  A  C  V  L  F  I  M  D  E  M  R  K  K  S  V  E  N  G  H  A
1127 accaccgggagaaggggttggattggggagtcttgttcgggttcggcccagggcttacagttgagactgtggttctacacagtgtaccggtt
      T  T  G  E  G  L  D  W  G  V  L  F  G  F  G  P  G  L  T  V  E  T  V  V  L  H  S  V  P  V
1217 gctcactagtagagatcgcgatctttgtgagaggagatcgatcgatattacttacttaaggtgtgatcggactttgtgtgtgtcaagttc
      A  H  *
1307 ggccttaagagattaatgatgtcttcttcttttttcttttaaatttggggatttgttttacatggatgtgtgattccctatttgaagatca
1397 ctagatagcaagatattgtaatttggtgttgaatgaattgaataaaactcaaagcttggtttaaaaaaaaaaaaaaaaaa
```

Fig 3. The nucleotide and deduced amino acid sequences of *pm*CHS. The coding sequence shown in capital letters is indicated below the nucleotide sequence (lowercase letters). The start codon (ATG) is coloured in red, and the stop codon (TAG) is marked with an asterisk. The poly-A tail and polyadenylation signal (AATAA) are shown in bold.

```
F.multiflora   MAPSVQEIRKAQRAEGPATVLAIGTATPPNCIYQADYPDYYFRVTNSEHMTDLKEKFRRMCDKSMIEKRYMHLTEEILKENQNMCAYMASSLDSRQDMVVSEVPRLGKEAAQKAIKEWGQ 120
P.cuspidatum   MAPSVQEIRKAQRAEGPATVLAIGTATPPNCIYQADYPDYYFRVTNSEHMTDLKEKFRRMCDKSMIEKRYMHLTEEILKENQNMCAYMASSLDSRQDMVVSEVPRLGKEAAQKAIKEWGQ 120
R.palmatum     MAPTVQEIRKAQRAEGPATVLAIGTATPANCIYQADYPDYYFRVTNSEHMTDLKEKFRRMCDKSMIEKRYMHLTEEILKENQNMCEYMAPSLDSRQDMVVSEVPRLGKEAAQKAIKEWGQ 120
P.minus        MAPSVEQIGKAQRAEGPATVLAIGTATPPNCVSQADYPDYYFRVTNSEHMTDLKEKFRRMCDKSMIEKRYMYLTEEILKENPNMCAYMEKSLDSRQDIVVTEVPRLGKEAAQKAIKEWGQ 120
P.alba         MAPSIEEIRKAQRASGPATILAIGKATPANCVQADYPDYYFRITNSEHMTELKEKFKRMCDKSMIKKRYMHITEEILKENPSMCEYMAPSLDARQDMVVVEVPKLGKEAATKAIKEWGQ 120
V.vinifera     -MVSVAEIRKAQRAEGPATVLAIGTATPANCVYQADYPDYYFRITNSEHMTELKEKFKRMCEKSMINKRYMHLTEEILKENPNVCAYMAPSLDARQDMVVVEVPKLGKEAAVKAIKEWGQ 119
G.max          -MVSVEEIRKAQRAEGPATVMAIGTATPDVDQSTYPDYYFRITNSEHMTELKEKFKRMCDKSMIKKRYMYLNEEILKENPSVCAYMAPSLDARQDMVVMEVPKLGKEAATKAIKEWGQ 119
               ::  :* *****.****::***.***.**:  *:  *******:*******:*****:***:****:****::  :*  **  ***:****:** ***:***** ***:******  ********

F.multiflora   PKSKITHVIMCTTSGVDMPGADYQLTKLLGLRPSVKRFMMYQQGCFAGGTVLRLAKDLAENTKGARVLVVCSEITAICFRGPTDTHLDSMVGQALFGDGAGAVIIGADPDLSVERPIFEM 240
P.cuspidatum   PKSKITHVIMCTTSGVDMPGADYQLTKLLGLRPSVKRFMMYQQGCFAGGTALRLAKDLAENTKGARVLVVCSEITAICFRGPTDTHLDSMVGQALFGDGAGAVIIGADPDLSIERPIFEL 240
R.palmatum     PKSKITHVIMCTTSGVDMPGADYQLTKLLGLRPSVKRFMMYQQGCFAGGTVLRLAKDLAENTRGARVLVVCSEITAICFRGPTDTHLDSMVGQALFGDGAGAVIIGSDPDLSIERPIFEL 240
P.minus        AKSKITHVIMCTTSGVDMPGADYQLTKLLGLRPSVKRFMMYQQGCFAGGTVLRLAKDLAENNKGARVLVVCSEITAVCFRGPTDTHLDSMVGQALFGDGAGAVIIGADPDLSIEKPIFEL 240
P.alba         PKSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRLAKDLAENNKGSRVLVVCSEITAVTFRGPSDTHLDSMVGQALFGDGAAAVIVGADPDTSIERPLFQL 240
V.vinifera     PKSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRLAKDLAENNAGSRVLVVCSEITAVTFRGPSDTHLDSLVGQALFGDGAAAVIIGADPDTKIERPLFEL 239
G.max          PKSKITHLIFCTTSGVDMPGADYQLTKLLGLRPSVKRYMMYQQGCFAGGTVLRLAKDLAENNKGARVLVVCSEITAVTFRGPTDTHLDSLVGQALFGDGAAAVIVGSDP-LPVEKPLFQL 238
               .******::***************.*****:*****:*****:*****************.    ::****:**::****:  **** ***********.:* **:**:**:** .

F.multiflora   VWTAQTILPDSEGAIDGHLREVGLTFHLLKDVPGLISKNIEKSLTEAFSPLNISDWNSLFWIAHPGGPAILDQVEAKLGLKEEKLKATRQVLNDYGNMSSACVLFIMDEMRKKSLENGHA 360
P.cuspidatum   VWTAQTILPDSEGAIDGHLREVGLTFHLLKDVPGLISKNIEKSLTEAFSPLNISDWNSLFWIAHPGGPAILDQVEAKLGLKEEKLKATRQVLNDYGNMSSACVLFIMDEMRKKSLENGHA 360
R.palmatum     VWTAQTILPDSEGAIDGHLREVGLTFHLLKDVPGLISKNIHKSLAEAFSPLNITDWNSLFWIAHPGGPAILDQVEAKLGLKEEKLKATRQVLNDYGNMSSACVLFIMDEMRKKSLENGHA 360
P.minus        VWTAQTILPDSEGAIDGHLREVGLTFHPLKDVPGLISKNIDKSLKEAFSPLNIDDWNSLFWIAHPGGPAILDQVEAKLGLKQEKLKATRQVLNDYGNMSSACVLFIMDEMRKKSVENGHA 360
P.alba         VSAAQTILPDSDGAIDGHLREVGLTFHLLKDVPGLISKNIEKSLVEAFAPIGINDWNSIFWIAHPGGPAILDQVEIKLDLKEEKLRATRNVLSDYGNMSSACVLFILDEMRNKSLEEGKS 360
V.vinifera     VSAAQTILPDSEGAIDGHLREVGLTFHLLKDVPGLISKNIEKSLVEAFTPIGISDWNSLFWIAHPGGPAILDQVELKLGLKEEKLRATRHVLSEYGNMSSACVLFILDEMRKKSIEEGKA 359
G.max          VWTAQTILPDSEGAIDGHLREVGLTFHLLKDVPGLISKNIEKALVEAFQPLGISDYNSIFWIAHPGGPAILDQVEAKLGLKPEKMEATRHVLSEYGNMSSACVLFILDQMRKKSIENGLG 358
               * :********.**************.*:** *:.* *:** :.:****:*****    **.** **:.***:**.:*.:*********:*:**:**:* .

F.multiflora   TTGEGQDWGVLFGFGPGLTVETVVLHSVPVAH---- 392
P.cuspidatum   TTGEGLDWGVLFGFGPGLAVETVVLHSVPVAHH-- 393
R.palmatum     TTGEGLDWGVLFGFGPGLTVETVVLHSVPVAH---- 392
P.minus        TTGEGLDWGVLFGFGPGLTVETVVLHSVPVAH---- 392
P.alba         TTGEGLEWGVLFGFGPGLTVETVVLHSVPVEQTIYS 396
V.vinifera     STGEGLEWGVLFGFGPGLTVETVVLHSVSAPPAH-- 393
G.max          TTGEGLDWGVLFGFGPGLTVETVVLRSVTV------ 388
               :**** :*******:***:*****:**..
```

Fig 4. Multiple sequence alignment of pmCHS and other plant CHS amino acid sequences. Amino acid position is given on the right. Identical, conserved and semi-conserved amino acids in the same column are indicated with the symbols "*", ":" and ".", respectively. Amino acid residues involved in the catalytic Cys-His-Asn motif (yellow), formation of the cyclisation pocket (purple), the malonyl-CoA binding motif, veaklglkeeklkatrq (blue) and highly conserved CHS signature sequence, gfgpg (red) are highlighted. The primary accession numbers of the protein sequences used in the alignment are as follows: *Fallopia multiflora* (PKS, GenBank: ADK45325), *Polygonum cuspidatum* (CHS, GenBank: ABK92282), *Rheum palmatum* (CHS2, GenBank: ABB13608), *Fagopyrum esculentum* (CHS, GenBank: ACZ51476) and *Citrus sinensis* (CHS2, GenBank: Q9XJ57).

database of INBIOSIS, UKM. Because the 3' sequence of *pmCHS* was readily available in this EST database, only 5'-RACE was required to obtain the full-length sequence of *pmCHS*. The GSP primer used to obtain the 5' *pmCHS* cDNA sequence was 5'GSP: 5'-CAG CTT CGC CTC CAC TTG GTC TAG G-3'. The SMARTer™ RACE cDNA amplification kit (Clontech, USA, Catalogue no. 634923) was used to synthesise first strand cDNA from *P. minus* for 5'-RACE, according to the manufacturer's protocol. The 5'GSP primer, a universal primer A mix (UPM, provided in the kit) and 5'-RACE-ready cDNA were used for the 5'-RACE amplification. The PCR amplification was carried out under the following conditions: 1 min at 94°C, followed by 35 cycles of 30s at 94°C, 30s at 62°C and 2 min at 72°C, and a final extension for 10 min at 72°C. The PCR products were purified and then sequenced by First BASE Laboratories Sdn. Bhd. (Selangor, Malaysia). The resulting partial sequences were aligned and assembled in a BioEdit Sequence Alignment Editor (Version 7.0.9.0) to obtain the full-length cDNA sequence of *pmCHS in silico*. The full-length sequence was used as a template to design a pair of specific primers (pmCHS_full-fwrd, 5'-GAT GGC TCC GTC GGT CGA ACA GAT C-3'; pmCHS_full-rev, 5'-CTA GTG AGC AAC CGG TAC ACT GTG-3'). The full-length *pmCHS* cDNA sequence was amplified with the specific primer pair Full CHS-f and Full CHS-r under the following PCR conditions: 2 min at 94°C, followed by 35 cycles of 30s at 94°C, 30s at 64°C and 2 min at 72°C, followed by 10 min at 72°C. The PCR products were purified, cloned into the pGEM®-T Easy Vector System I (Promega, Catalogue no. A1360) and transformed into *E. coli* JM109 for sequencing. Two rounds of PCR amplification and sequencing were used to confirm the sequencing results.

Bioinformatics analysis

The nucleotide and amino acid sequences of *P. minus* chalcone synthase were analysed using bioinformatics software tools. Comparative sequence analysis of *pmCHS* was performed online using NCBI BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The open reading frame of *pmCHS* was predicted by ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) Conserved domains were identified with NCBI Batch CD-Search Tool (Marchler-Bauer et al., 2011) and protein classification was performed with Superfamily 1.75 (http://supfam.cs.bris.ac.uk/SUPERFAMILY/). Subcellular loc-alisation analysis was performed using the TargetP 1.1 server. The presence of signal peptides was evaluated using the SignalP 4.0 Server (http://www.cbs.dtu.dk/services/SignalP/) . The TMHMM Server v.2.0 (Ikeda et al., 2002) and ProtScale (Kyte and Doolittle, 1982) were used to predict the cellular function, transmembrane helices and hydrophobicity of the pmCHS protein. Multiple sequence alignment of protein sequences was conducted with ClustalW2 using default parameters (http://www.ebi.ac.uk/Tools/msa/clustalw2/). A phylogenetic tree was drawn by the Neighbour-Joining method with 1000 replicates; bootstrap values for each node calculated using MEGA 4 software. All sequences used in the phylogenetic analysis were retrieved from the NCBI GenBank database (http://www.ncbi.nlm.nih.gov/genbank/) The three-dimensional (3D) protein structure homology-modelling of the pmCHS was performed with SWISS-MODEL (Arnold et al.,
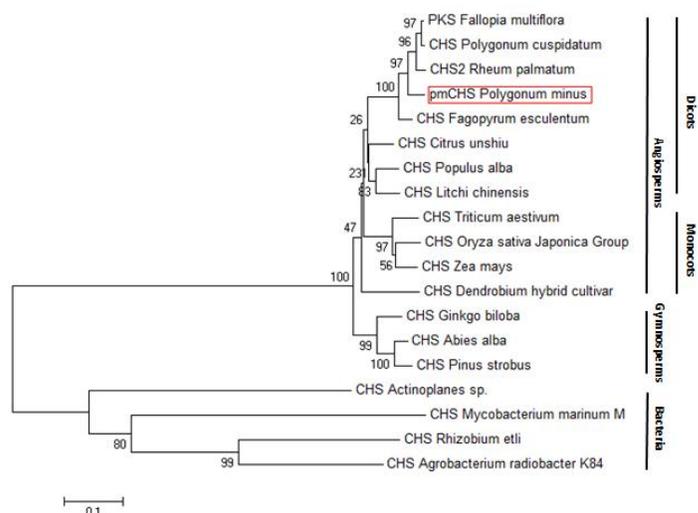
Fig 5. The phylogenetic tree of plant and bacterial CHSs. The tree was constructed by MEGA 4 using Neighbour-Joining method with 1000 bootstrap replicates. The bar indicates an evolutionary distance of 0.1%. *P.minus* CHS is boxed. The database accession numbers of the CHS sequences used are as follows: *Fallopia multiflora* (PKS, GenBank: ADK45325), *Polygonum cuspidatum* (CHS, GenBank: ABK92282), *Rheum palmatum* (CHS2, GenBank: ABB13608), *Fagopyrum esculentum* (CHS, GenBank: ACZ51476), *Citrus unshiu* (CHS, GenBank: ACQ41890), *Populus alba* (CHS, GenBank: ABD24222), *Litchi chinensis* (CHS, GenBank: ADB44077), *Triticum aestivum* (CHS, GenBank: AAQ19321), *Oryza sativa* (CHS, GenBank: BAB39764), *Zea mays* (CHS, GenBank: NP_001149022), *Dendrobium* sp. (CHS, GenBank: CAR64528), *Ginkgo biloba* (CHS, GenBank: AAT68477), *Abies alba* (CHS, GenBank: ABD38610), *Pinus strobus* (CHS, GenBank: CAA06077), *Actinoplanes* sp. (CHS, GenBank: AEV88919), *Mycobacterium marinum* M. (CHS, GenBank: YP_001852574), *Rhizobium etli* (CHS, GenBank: YP_001984314) and *Agrobacterium radiobacter* (CHS, GenBank: YP_002543476).
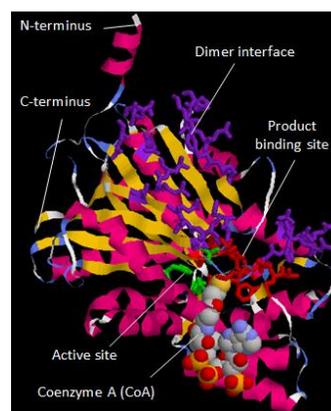
Fig 6. The 3D structure of the *P. minus* CHS monomer. The N- and C- termini, coenzyme A (space-filling molecule) and other important motifs are colored and labelled, active site, green; product binding site, red; and dimer interface, purple. Alpha helixes and beta-strands are indicated in pink and yellow, respectively.
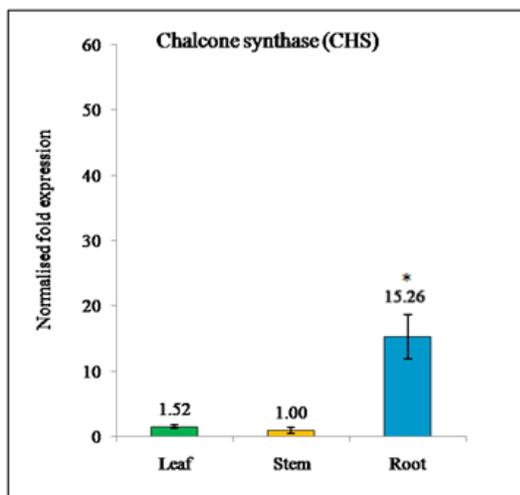
Fig 7. The relative expression of *pmCHS* in different tissues of *P. minus*. The CHS mRNA levels are expressed relative to the amount of *Tubulin* and *β-actin* mRNA. The data are presented as the mean ± SE (n = 3), in arbitrary units. *, *P<0.05* indicates a significant difference, as analysed by Tukey's test.

2006). The stereochemical quality of the protein structure was checked with PROCHECK (http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/) . RasMol Version 2.7.3 was used to visualise the 3D protein structure.

Expression profile analysis

Total RNA was extracted from *P. minus* leaves, stems and roots asdescribed above and treated with DNase using the RQ1 RNase-Free DNase kit (Promega, Catalogue no. M6101). DNase-treated RNA (2 μg) was reverse transcribed with the Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Fermentas, Catalogue no. K1641) in a total volume of 20 μl. qRT-PCR was performed in a 96-well optical plate using a Bio-Rad iQ5 instrument (Bio-Rad) and universal cycling conditions (3 min at 95°C, 40 cycles of 10 s at 95°C and 30 s at 62°C A melting curve was generated at the end of each run to check the specificity of amplification. The reaction mixture contained iQ^{TM} SYBR® Green Supermix (Bio-Rad, Catalogue no. 170-8880), 100 μM of the specific primers and 100 ng of cDNA in 15 μl reactions. Primer efficiencies and standard deviations were calculated based on a standard curve generated using different concentrations of in triplicates. *Tubulin* and *β-actin* were selected as reference genes.

Statistical analysis

The data were expressed as the mean ± SE. Tukey's tests (Minitab Inc., USA) was performed to assess the statistical significance of differences between tissues. Statistical significance was set at *P < 0.05*.

Conclusion

The cDNA of the *pmCHS* gene has been successfully cloned and characterised from the aromatic herb *P. minus*. Important motifs identified from the highly conserved pmCHS protein sequence imply that this protein is functionally involved in the flavonoid biosynthesis pathway. The high expression of *pmCHS* in the root provides the basis for the further investigation into *pmCHS* gene expression and regulation *in vivo*. These data also contribute to the literature on plant CHSs and may provide useful information for studying of CHS gene polymorphism.

References

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics. 22: 195-201.

Bais HP, Weir TL, Perry LG, Gilroy S, Vivanco JM (2006) The Role of root exudates in rhizophere interactions with plants and other organisms. Annu Rev Plant Biol. 57: 233-66.

Bertin C, Yang XH, Weston LA (2003) The role of root exudates and allelochemicals in the rhizophare. Plant Soil. 256: 67-83.

Cazarolli LH, Zanatta L, Alberton EH, Figueiredo MS, Folador P, Damazio RG, Pizzolatti MG, Silva FR 2008.Flavonoids: prospective drug candidates. Mini Rev Med Chem 8(13): 1429-1440.

Christensen AB, Gregersen PL, Schröder J, Collinge DB (1998) A chalcone synthase with an unusual substrate preference is expressed in barley leaves in response to UV light and pathogen attack. Plant Mol Biol. 37: 849-857.

Dalkin K, Edwards R, Edington B, Dixon RA (1990) Stress responses in alfalfa (*Medicago sativa* L.). I. Induction of phenylpropanoid biosynthesis and hydrolytic enzymes in elicitor-treated cell suspension culture. Plant Physiol. 92: 440-446.

Dao TTH, Linthorst HJM and Verpoorte R (2011) Chalcone synthase and its function in plant resistance. Phytochem Rev. 10: 397-412.

Dixon RA, Harrison MJ (1990) Activation, structure and organization of genes involved in microbial defense in plants. Adv Genet. 28: 165-234.

Hrazdina G (1992) Compartmentation in aromatic metabolism. In: Stafford HA and Ibrahim RK (ed) Recent Advances in Phytochemistry. Plenum Press, New York.

Ikeda M, Arai M, Lao DM (2002) Transmembrane topology prediction methods: A reassessment and improvement by a consensus method using a dataset of experimentally characterized transmembrane topologies. *In Silico* Biol. 2: 1-33.

Jez JM, Austin MB, Ferrer JL, Bowman ME, Shröder J, Noel JP (2000) Structural control of polyketide formation in plant-specific polyketide synthases. Chem Biol. 40: 1-12.

Jez JM, Ferrer J-L, Bowman ME, Austin MB, Schröder J, Dixon RA, Noel JP (2001) Structure and mechanism of chalcone synthase-like polyketide synthases. J Ind Microbiol Biot. 27: 393-398.

Jiang CG, Schommer CK, Kim SY, Suh DY (2006) Cloning and characterization of chalcone synthase from the moss, *Physcomitrella patens*. Phytochem. 67: 2531-2540.

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

Lei W, Tang S-H, Luo K-M, Sun M (2010) Molecular cloning and expression profiling of a chalcone synthase gene from hairy root cultures of *Scutellariaviscidula* Bunge. Genet Mol Biol. 33(2): 285-291.

Li XH, Park NI, Xu H, Woo S-H, Park CH, Park SU (2010) Differential expression of flavonoid biosynthesis genes and accumulation of phenolic compounds in common buckwheat (*Fagopyrum esculentum*). J Agr Food Chem. 58: 12176-12181.

Mallika V, Sivakumar KC, Soniya EV (2011) Evolutionary implications and physicochemical analyses of selected proteins of type iii polyketide synthase family. EvolBioinforms 7: 41-53.

Maizura M, Aminah A, Wan Aida WM (2011) Total phenolic content and antioxidant activity of kesum (*Polygonum minus*), ginger (*Zingberofficinale*) and turmeric (*Curcuma longa*) extract. Int Food Res J. 18: 529-534.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH (2011) CDD: a Conserved domain database for the functional annotation of proteins. Nucleic Acids Res. 39(D): 225-9.

Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structure coordinates. Proteins. 12: 345-364.

Poulin M-J, Bel-Rhlid R, Piché Y, Chênevert R (1993) Flavonoids released by carrot (*Daucus carota*) seedlings stimulate hyphal development of vesicular-arbuscular mycorrhizal fungi in the presence of optimal CO_2 enrichment. J Chem Ecol. 19: 2317-2327.

Roslan ND, Yusof JM, Baharum SM, Othman R, Mohamed-Hussein Z-A, Ismail I, Noor MN, Zainal Z (2012) Flavonoid biosynthesis genes putatively identified in the aromatic plant polygonum minus via expressed sequences tag (EST) analysis. Int J Mol Sci. 13(3): 2692-2706.

Saslowsky D, Winkel SB (2001) Localization of flavonoid enzymes in Arabidopsis roots. Plant J. 27: 37–48.

Scervino JM, Ponce MA, Erra-Bassels R, Vierheilig H, Ocampo JA, Godeas A (2005) Arbuscular mycorrhizal colonization of tomato by *Gigaspora* and *Glomus* species in presence of roots flavonoids. J Plant Physiol. 162: 625-633.

Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Somerville SC, Manners JM (2000) Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. Proc Natl Acad Sci USA. 97: 11655-11660.

Schröder J (1997) A family of plant–specific polyketide synthases: facts and predictions. Trends Plant Sci. 2:373-378.

Shvarts M, Borochov A, Weiss D (1997) Low temperature enhances petunia flower pigmentation and induces chalcone synthase gene expression. Physiol Plantarum. 99: 67-72.

Stefan M, Axel M (2005) Flavones and flavone synthases. Phytochem. 66: 2399-2407.

Steinkellner S, Lendzemo V, Langer I, Schweiger P, Khaosaad T, Toussaint J-P, Vierheilig (2007) Flavonoids and strigolactones in root exudates as signals in symbiotic and pathogenic plant-fungus interactions. Molecules. 12: 1290-1306.

Suh D-Y, Fukuma K, Kagami J, Yamazaki Y, Shibuya M, Ebizuka Y, Sankawam U (2000) Identification of amino acid residues important in the cyclization reactions of chalcone and stilbene synthases. Biochem J. 350: 229-235.

Thain SC, Murtas G, Lynn JR, McGrath RB, Millar AJ (2002) The circadian clock that controls gene expression in *Arabidopsis* is tissue specific. Plant Physiol. 130: 102-110.

Tropf S, Kärcher B, Schröder G, Schröder J (1995) Reaction mechanisms of homodimeric plant polyketide synthases (stilbene and chalcone synthase): a single active site for the condensing reactions is sufficient for synthesis of stilbene, chalcone and 6′-deoxychalcones. J Biol Chem. 270: 7922-7928.

Zhou L, Wang Y, Peng ZH (2011) Molecular characterization and expression analysis of chalcone synthase gene during flower development in tree peony (*Paeonia suffruticosa*). Afr J Biotechnol. 10(8): 1275-1284.