

Preliminary investigations on the pathogenesis-related protein expression profile of the medicinal herb *Macleaya cordata* and anti-bacterial properties of recombinant proteins

Jinjun Chen^{a,*}, Zihao Liu^a, Yisong Liu^b, Xuewen Zhang^a, Jianguo Zeng^{b,**}

^a College of Bioscience and Biotechnology, Hunan Agricultural University, Changsha, Hunan, 410128, China

^b Hunan Key Laboratory of Traditional Chinese Veterinary Medicine, Hunan Agricultural University, Changsha, 410128, China

ARTICLE INFO

Keywords:

Macleaya cordata

Papaveraceae

Expressed sequence tags

Pathogenesis-related protein

Defensin

ABSTRACT

The plant pathogenesis-related (PR) proteins play a crucial role in the defense of plants against pathogens and orchestrate the innate immune system of plants. In this paper, a non-normalized cDNA library of the leaf was constructed to obtain a comprehensive view of PR proteins of *Macleaya cordata*. Specifically, 511 expressed sequence tags (ESTs) were generated using Sanger sequencing. All ESTs were assembled into 364 non-redundancy sequences, including 78 clusters and 286 singlets. The PR protein expression profile of the medicinal herb *M. cordata* has been investigated and is represented by defensin, lipid-transfer protein, (S)-norco-claurine synthase, and major allergen protein, suggesting that the herb contains rich active proteins against pathogens. Furthermore, two defensins were selected for recombinant expression in yeast, and the antimicrobial activities were explored. Since they both present a broad antimicrobial spectrum, they are of particular importance for agricultural and medicinal applications. Our study describes defensins in Papaveraceae for the first time and provides novel insights into the effective components. In addition to the alkaloids, PR proteins (such as defensins, lipid transfer proteins, (S) - norco-claurine synthase, major allergen protein, and Class IV chitinases) are involved in the antibacterial and anti-inflammatory activities of *M. cordata*.

1. Introduction

The plant *Macleaya cordata* (Willd.) R. Br. is a species in the family Papaveraceae and acts as a toxic folk plant for its anti-inflammatory and insecticidal activities (Kosina et al., 2010). It was recorded in “Ben Cao Shi Yi” and “Compendium of Materia Medica”, which are the illustrious Chinese medical material books (Liu et al., 2013; Zeng et al., 2013). Nowadays, *M. cordata* is an important economic crop and has been grown on a large scale. It is the major raw material of benzyloisoquinoline alkaloids as an animal food additive. This is recorded on the European Food Safety Authority (EFSA) list of plants (EFSA, 2010). A close genus in America called *Bocconia* spp. also contains alkaloids and is applied to control mange, lice, and intestinal worms, and treat ulcers, wounds, and even Leishmaniasis (Lin et al., 2018; Zeng et al., 2013). The bioactivities and pharmacological activities of the alkaloids are remarkable and intensively investigated in various aspects, such as anticancer (Liu et al., 2013), molluscicidal (Ming et al., 2011) and cytotoxic activities

(Hammerová et al., 2011), which have been well-reviewed (Lin et al., 2018).

Pavel Kosina and colleagues have demonstrated that the mixtures of seeds, seed capsules, and the soluble proteins of leaves have a stronger antimicrobial effect than pure alkaloids, inferring that there are synergistic effects of other components in these soluble proteins (Kosina et al., 2010). Besides, ample evidence indicates that plants develop immune responses to protect themselves from pathogen invasion through the expression of various pathogenesis-related (PR) proteins, which are highly diverse in primary structure and exhibit different antimicrobial activity (Sels et al., 2008). However, no cysteine-rich PR protein was identified in our previous high-throughput next-generation sequencing (NGS) data (Zeng et al., 2013) due to the difficulty in assembling of a multigene family. In the present study, the pathogenesis-related proteins of *M. cordata* were discussed and especially regarding defensins rather than alkaloids, and the undescribed anti-microbial substances were predicted, contributing to the development and application of the product from *M. cordata*.

* Corresponding author.

** Corresponding author.

E-mail addresses: jinjunchen@hunau.edu.cn (J. Chen), zengjianguo@hunau.edu.cn (J. Zeng).

<https://doi.org/10.1016/j.phytochem.2021.112667>

Received 23 February 2020; Received in revised form 7 January 2021; Accepted 9 January 2021

Available online 4 February 2021

0031-9422/© 2021 Elsevier Ltd. All rights reserved.

Abbreviations

Amino acid aa
 Cluster of Orthologous Groups COG
 Expressed Sequence Tags ESTs
 Gene Ontology GO
 4-hydroxyphenylacetaldehyde 4-HPAA
 Kyoto Encyclopedia of Genes and Genomes KEGG
Macleaya cordata defensin 1-5 McDef1-5
Macleaya cordata lipid transfer protein 2 McLTP 2
Macleaya cordata (S)-norcoclaurine synthase 1 McNCS1
Macleaya cordata Major Allergen 1 Mac-co 1
 Millimeter mm
 (S)-norcoclaurine synthase NCS
 Pathogenesis-Related PR
 Three-dimension 3D

2. Results and discussion

2.1. An overview of the analysis of *M. cordata* (willd.) R. Br. leaf ESTs

The cDNA library from *M. cordata* leaf was non-normalized, and the titer was 1.0×10^6 cfu/mL with recombinant clones. After the poor-quality sequences were discarded, 511 high-quality ESTs were used to investigate the gene expression profile of *M. cordata* leaf. The initial sequences were grouped into 364 non-redundancy sequences, of which 78 clusters exhibited more than one EST and 286 singlets (Supplementary Table 1). The average length of the non-redundant sequence was 513 nucleotides (ranging from 105 to 1013 bp) (Supplementary Figure 1). Additionally, the newly discovered EST sequences of the cDNA library were submitted to Genbank with the accession number from JK752513 to JK752907.

With the development of high-throughput next-generation sequencing technologies, sequencing the transcriptomes of medicinal plants with immense pace has become a hot pursuit, generating more generalized insights. However, de novo assembly is a big challenge, especially for the assembly of cysteine-rich defensins, because these peptides are hypervariable sequences using conserved scaffolds. The cDNA library with ESTs Sanger sequencing approach has been verified to be a rapid and reliable method for uncovering undescribed genes, and the interesting candidates provide the gene templates for subsequent recombinant expression.

2.2. Annotation and classification of the ESTs

2.2.1. Searching against non-redundant protein sequences (NR) database

The non-redundant sequences were used to search against the non-redundant protein sequences (NR) database in NCBI with an E-value threshold of 10^{-5} by performing the BLASTX algorithm, so as to annotate the non-redundant sequences. The similarity distribution of the BLAST hits for each sequence ranging from 36% to 100%, and the three top-hits were *Vitis vinifera* (11.25%), *Populus trichocarpa* (10.69%), and *Glycine max* (10.42%).

2.2.2. Gene Ontology (GO) annotation

The Gene Ontology (GO) annotation was used to classify the non-redundant sequences into three ontology domains: Cellular Component, Molecular Function, and Biological Process. There are 244 non-redundant sequences annotated and 1672 GO terms obtained. Among them, 855 terms (51.14%) belong to Biological Process, 355 terms (19.44%) belong to Molecular Function, and 492 terms (29.42%) belong to Cellular Components. An overall view of the distribution of the sequences in the three ontologies is presented in Supplementary Figure 2.

Besides, the three most represented functional classes in Biological Processes are 'cellular process', 'metabolic processes', and 'response to stimulus'. The classes 'binding' and 'catalytic' are particularly well represented in the Molecular Function. The predominant components in Cellular Component are 'cell', 'cell part', and 'organelle part'.

2.2.3. Cluster of Orthologous Groups (COGs) annotation

The database of Clusters of Orthologous Groups of proteins (COGs) is a useful approach for a phylogenetic classification of the proteins (Tatusov et al., 2001). There were 83 non-redundant sequences (22.8% of total unigenes) assigned to 25 COG classifications. Among them, the cluster for "Translation, ribosomal structure, and biogenesis" represents the largest group, followed by the clusters for "Posttranslational modification, protein turnover, chaperones" and "General function prediction only" (Supplementary Figure 3).

2.2.4. Kyoto encyclopedia of genes and genomes (KEGG) annotation

By conducting GO-EnzymeCode Mapping of the BLAST2GO program (Conesa et al., 2005), the result of mapping exhibited that 48 non-redundant sequences acquired annotation and enzymatic characters, including EC numbers and relevant probable pathways. According to the function of the total enzymes, the enzymes can be divided into six classes: oxidoreductases (EC 1), transferases (EC 2), hydrolases (EC 3), lyases (EC 4), isomerases (EC 5), ligases (EC 6). The statistical results indicate that the transferases (EC 2) are the most significant class with 18 non-redundant sequences (37.50%) in the present transcriptional data. The transferring phosphorus-containing group (EC 2.7) is a sub-class of the transferase (EC 2) class and contains an array of enzymes involved in all kinds of kinase activities. There are 14 items (29.17%) in the oxidoreductase (EC 1) group, which is the second large class. The lyase (EC 4) group contains six items (12.50%). The group of ligase (EC 6) and hydrolase (EC 3) contain five items (10.42%) and four items (8.30%), respectively. The isomerase (EC 5) is the fewest group with only one non-redundant sequence identified, suggesting that the distribution of isomerase in the leaf of *M. cordata* is relatively low. The KEGG profiles of transcripts from *M. cordata* leaf are presented in Supplementary Table 2. Most of the non-redundant sequences were associated with metabolism, including energy metabolism (17 non-redundant sequences), amino acid metabolism (15 non-redundant sequences), and carbohydrate metabolism (12 non-redundant sequences). Notably, 2 non-redundant sequences were predicted as the cDNA of LRR receptor-like serine/threonine-protein kinase FLS2 involved in the plant-pathogen interaction pathway, and one sequence was annotated as (S)-norcoclaurine synthase of isoquinoline alkaloid biosynthesis pathway.

2.3. ESTs related to pathogenesis-related (PR) proteins

Pathogenesis-related (PR) proteins are assumed to be ubiquitous in the plant kingdom and triggered by pathogen attack (biotic), abiotic stress, hypersensitive response (HR), and systemic acquired resistance (SAR). PR proteins are multifarious and have been classified into 17 different families based on primary structure, serological activity, and biological activity. Remarkably, most of them possess antimicrobial activity in vitro (Van Loon et al., 2006). In the present study, five groups in 4 families of PR proteins-like genes (PR-3, 2 types of PR-10, PR-12 and PR-14) are discovered in 511 random ESTs from *M. cordata* healthy leaf, demonstrating that the traditional medical herb *M. cordata* has high activity against the plant pathogens.

PR-12: Plant defensin. Using the Blastx program, 5 defensin precursors can be deduced from 8 ESTs, namely, *Macleaya cordata* defensin 1-5 (McDef1-5). Homologous analysis of the deduced amino acid sequences of McDef1-5 presents that McDef1 is much different from McDef2-5. McDef1 contains an N-terminal signal peptide with 23 amino acids, and the sequences of McDef2-5 all contain a longer signal peptide with 31 amino acids. The eight cysteine residues in each mature domain

of the 5 McDefs are highly conserved, contributing to the formation of four disulfide bridges (Fig. 1). The theoretical pI/MW of each part of the precursors of McDef1-5 was calculated with the ExPASy-Compute pI/MW tool (Supplementary Table 3).

The mature domain of McDef1 shares high homology to the defensins from different plant species, and the highest identity sequence gb|AAL14240.1| is from *Musa acuminata*. In contrast, the mature peptides of McDef2-5 are most similar to DmAMP1 ([sp|POC8Y4.1|]) from *Dahlia merckii* with fewer homologs. Notably, each of the 5 McDefs contains a γ -core (GXC (X3–9)C), an α -core (GXC (X3–5)C), and a typical cysteine-stabilized α/β motif (CS α/β) composed of three antiparallel β -strands and one α -helix (Sagaram et al., 2011). The γ -core motif of McDef1 and McDef2-5 is GKCQGVRRRC and GACHVRGGKHMC, respectively. Besides, the α -core motif of McDef1 and McDef2-5 is GLCVRKSNC and GNCGNTQHC, respectively (Fig. 2). The tertiary structure of the mature peptide of McDef1 was predicted based on the template 4aazA, X-ray structure of *Nicotiana glauca* Defensin 1 (Nad1). The tertiary structures of McDef2-5 were predicted according to 1bk8A, the three-dimensional solution structure of *Aesculus hippocastanum* antimicrobial protein 1 (Ah-Amp1) (Fig. 3A & B).

Plant defensins exist throughout the plant kingdom, belong to the plant innate immune system, and provide broad-spectrum resistance against fungi, bacteria, and viruses (Munoz et al., 2014; Sathoff et al., 2019). Although they share a conserved cysteine spacing pattern and three-dimensional structure, they have highly diverse sequences and display a range of activities including antifungal and antibacterial activities, inhibitory enzyme activities, and heavy metal tolerance and development (Van der Weerden and Anderson, 2013). To date, no defensin has been described in *M. cordata*, even in the Papaveraceae family. In the present work, the transcriptional information of defensins was obtained from the cDNA library of *M. cordata* leaf. This is the first time to report defensin in Papaveraceae. McDefs consist of the N-terminal signal peptide fragment and the mature peptide fragment. Intriguingly, the isoelectric points of the McDefs signal peptides are more alkaline than those of mature peptides (Supplementary Table 3), indicating that the McDefs are different from general plant defensins containing a signal peptide with net negative charges (Carvalho and Gomes, 2009). The acidic signal peptides of defensins were reported to not only target the peptide for secretion but also mitigate the biological activity of the mature peptide until it is released to function (Michaelson et al., 1992). McDef1 and the homologs share high similarity within the γ -core but a low similarity in the α -core. The residues in the γ -core are the primary determinants of the antifungal activity and morphogenesis of these defensins. The more the alkaline of the γ -core, the higher the potent antifungal activity of the defensin (Sagaram et al., 2011). The γ -core peptide of McDef1 (GKCQGVRRRC) shares a high identity to that of MtDef4 (GRCRGFRRRC) (Kaur et al., 2011). Interestingly, the full mature sequence of McDef1 was close to Defensin D2 from spinach (sp|P81571.1|); this was referred to as the activity against Gram-positive and Gram-negative bacterial pathogens and fungi (Segura et al., 1998). In contrast, there are fewer defensins similar to McDef2-5, and DmAMP1 (POC8Y4.1) from *Dahlia merckii* is the closest to McDef2-5 sequences. DmAMP1 has antifungal action, involving membrane

interaction with sphingolipids and ergosterol, as well as membrane permeabilization (Thevisen et al., 2003). Since McDef2-5 share 80% identities with the DmAMP1, they may own a similar function and action mechanism.

PR-14: Lipid transfer protein. One EST was matched to the type 2 non-specific lipid transfer protein, which was named *Macleaya cordata* lipid transfer protein 2 (McLTP 2) with 100 amino acid residues in the precursor. McLTP 2 precursor contains an N-terminal signal peptide (1–32 aa) and a mature peptide (33–100 aa). The theoretical molecular mass of the mature peptide is 7356.68 Da with pI of 9.47. The four disulfide bridges of McLTP 2 were predicted to be C_I-C_V, C_{II}-C_{III}, C_{IV}-C_{VII}, and C_{VI}-C_{VIII} (C means cysteine) (Fig. 4) according to the homologs analysis. The three-dimensional (3D) protein structure of the mature domain of McLTP 2 is similar to type 2 non-specific lipid transfer protein from wheat (Parent PDB: 1tuk) (Hoh et al., 2005). According to the 3D template 1tuk, McLTP 2 contains an N-terminal helix (Ala 5-Cys11), five α -helices (α 1 (Tyr 12-Ser 17), α 2 (Ala 23-Gln 32), α 3 (Leu 36-Tyr 39), α 4 (Lys 46-Val 49), α 5 (Ala 52-Thr 60)), and a left-handed polyproline type 2 helical conformation C-terminal chain (Val 63-Ile 67) (Fig. 3C). It was predicted by the homologous analysis that the residues of Leu 8, Leu 36, Tyr 39, Leu 45, Phe 48, and Val 49 in McLTP 2 played a vital role in structural stability and lipid-binding ability (N.-J. Wang et al., 2012).

In the plant, the family of lipid transfer proteins is the PR-14 family of pathogenesis-related proteins. Since the first plant lipid transfer protein was isolated from potato tuber in 1975 by Kader (Kader, 1975), numerous gene sequences and protein sequences have been identified and characterized (D'Agostino et al., 2019; Salminen et al., 2016). The large family of lipid transfer proteins can transfer phospholipids, fatty acids, and steroids between membranes (Hoh et al., 2005), named as “non-specific lipid-transfer protein” (nsLTP) (Kader, 1996). They are small and basic peptides with four conserved disulfide bridges from eight-Cys motif (C-Xn-C-Xn-CC-Xn-CXC-Xn-C-Xn-C) and ubiquitous in the plant (Ng et al., 2012). The nsLTPs from wheat, rice, and arabidopsis can be classified into nine types (from I to IX) (Ng et al., 2012). McLTP 2 belongs to the type II Prosite-styled pattern (CX4LX2CX9-11 P [S,T]X2CCX5QX2-4C [L,F]CX2 [A,L,I]X [D,N]PX10-12 [K,R]X4-5CX3-4PX0-2C). Based on the result of deduced three-dimensional structure, McLTP 2 is similar to 1tuk (Hoh et al., 2005) and contains a main central hydrophobic cavity and a smaller cavity, providing the structural basis for the capacity of McLTP 2 with lipid-transfer. nsLTPs have been identified with a wide variety of biological functions, such as participating in cutin and wax assembly (Pyee and Kolattukudy, 1995), enhancing the plant tolerance to biotic and abiotic stress including antibacterial (S. Y. Wang et al., 2004), antifungal (Isaac Kirubakaran et al., 2008; Zhu et al., 2012), anti-freezing (Hincha, 2002), resistance to drought stress (Jung et al., 2005), and salt stress (Gonorazky et al., 2005).

PR-10: (S)-norcoclaurine synthase. In the present study, two identical sequences encode a 193-aa protein similar to (S)-norcoclaurine synthase (NCS) with a theoretic molecular mass of 21289.31 Da, pI of 5.30 and no signal peptides. The deduced protein was named as McNCS1. The enzyme NCS catalyzes the stereospecific Pictet-Spengler cyclization between dopamine and 4-hydroxyphenylacetaldehyde,



Fig. 1. The alignment analysis of McDef1-5. The underlined amino acids represent the signal sequence predicted with SignalP. The cysteines are marked in yellow background and the putative disulfide bonds are also shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

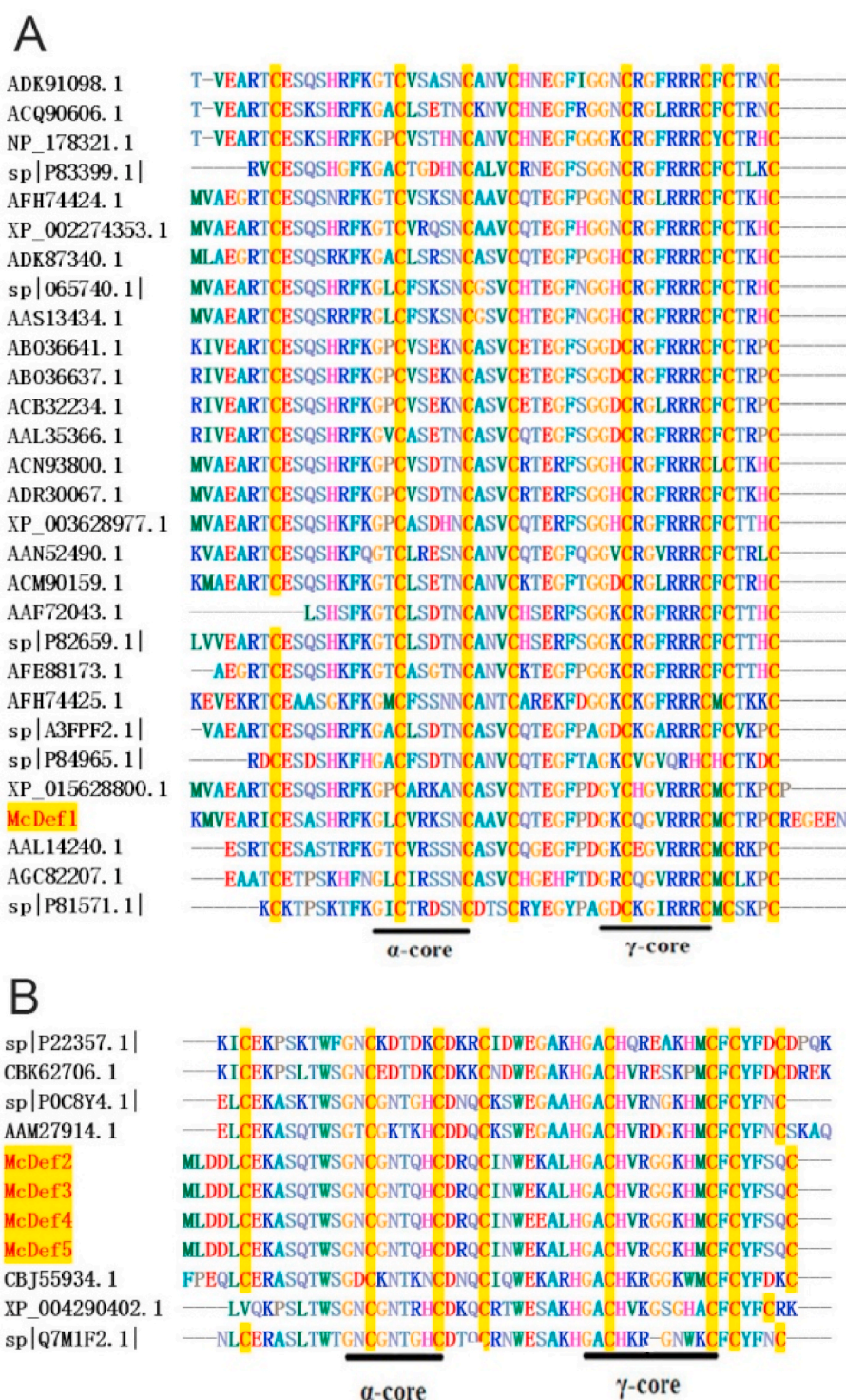


Fig. 2. Multiple sequence alignments of the McDef1 (A) and McDef2-5 (B) homologs from different plants with ClustalW2. The conserved domains α -core and γ -core are marked under the homologous sequences respectively. The target proteins McDef1 in (A) and McDef2-5 in (B) of this study are highlighted with yellow fluorescent background. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

becoming the only known member of the PR-10/Bet v1 family with unambiguously catalytic activity in plants (Fernandes et al., 2013; Lee and Facchini, 2010). Notably, McNCS1 exhibits a high level of identity with NCS from Papaveraceae and a medium level of similarity to other PR-10 proteins from various plant kingdoms (Supplementary Figure 4). Besides, the sequence motif (AGDGGVGT) of McNCS1 is a highly conserved glycine-rich loop and a most rigid structural element in the PR-10 fold, presenting the symbol domain of PR-10 proteins (Fernandes

et al., 2013). In the NCS sequence, the Lys, Asp, Glu, and Tyr residues dominate the geometry of the NCS active site and shape the binding site of dopamine and 4-hydroxyphenylacetaldehyde (4-HPAA) (Ilari et al., 2009). The Tyr69, Glu71, Lys83, and Asp102 residues in McNCS1 are also conserved. Moreover, the primary structure of McNCS1 is very identical to the NCS from *P. somniferum*, *C. saxicola*, and *A. mexicana*. A predicted tertiary structure of McNCS1 is close to the template 2vq5A, X-ray structure of norcoclaurine synthase from *Thalictrum flavum*.

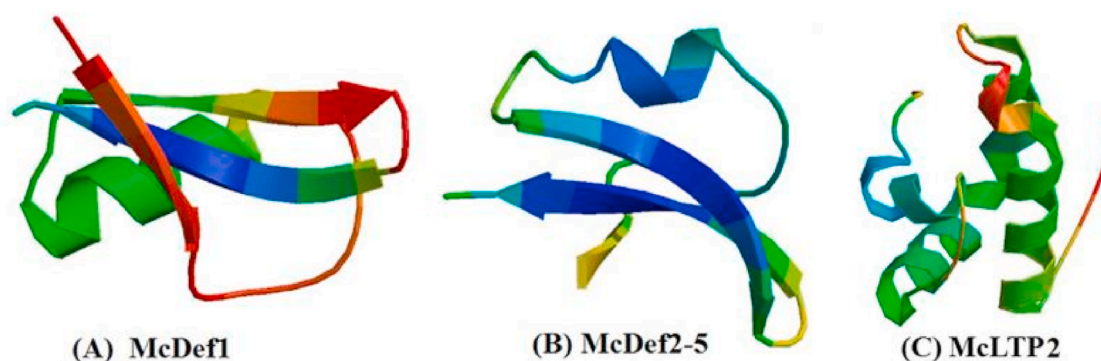


Fig. 3. The tertiary structures of McDef1 (A), McDef2-5 (B), McLTP 2 (C) were obtained by SWISS-MODEL.

McLTP2	
CBI33621. 3	ATCKATELASGYSAITSSSPPSALCCRLKEQRGLCKYLDPNLKFFVNSANARRVCKTGGVPFPIC
XP_003620406. 1	VTCSPLLELSSGFAAITSSAPPSSMCCSKLREQRPCLCGYIRDPNLSQYVNSANARRVASTCGVPFPNC
ACM78621. 1	VTCSPVQLSSGVSAITSSTPPSNLCCSKIKEQRPCLCQYMKNPYLKFFVNTPNARKVANTCGTPFPRC
BAM05641. 1	VTCSPAQLSPCVGAITSSSPPSGLCCSKVRQKPCCLCQYMKNPYLKFFVNTPNARKVASQCRTPFPRC
BAM05643. 1	VACRPIELSACVNIAITSATPPTTLCCSKIKEQKPCCLCQYLQNPYLKFFVSSPNARKVANTCGTPFPKC
BAM05644. 1	VACRPIELSACVNIAITSATPPTTLCCSKIKEQKPCCLCQYLQNPYLKFFVSSPNARKVASTCGTPFPKC
NP_188456. 1	VTCSPMQLSPCATAITSSSPPSALCCAKLKEQRPCLCGYMRNPRLRFVSTPNARKVSKSCKLPFRC
NP_564532. 1	VTCSPMQLASCAAMTSSSPPSAECCTKLREQKPCCLCGYMRNPRTLQYVSSPNARKVNSCKIPSPSC
AAA33933. 1	ATCSVTLELMPCCSAFTSSAAPTACCTKLKEQSPCLCGYLNPTLQYITNPNAKKVTSTCGVIPNC
sp Q43681. 1	VTCNPTLELSSGVPATGGSKPSSCCSKLVQEPCLCNYIKNPQLKQYVNSPCAKKVLNCGVTPNC
ABQ53930. 1	APCNPVLELSPCLAAISGQPPSAECCSKLKEQVPCFCDYVKNPVFKPYVDSNARKVAAATCGVAFPTC
KYP39148. 1	VTCSPVQLSSGVSAITSSTPPSNLCCSKIKEQKPCCLCQYLNPNLKKFVDSNARRVARTCGTPFPRC
sp P82353. 1	VTCSPVQLSPCLGPIGNSCAPSPTTCCQKLREQRPCLCGYLNPNLQYVNSPNARKLASNCGVPVPQC
BAJ25798. 1	VTCSVVELSPCAGAISSQPSSAACAKLKEQKPCCLCGYLNPNLQYVNSPNARKVAKSCGVPTPSC
AFP49331. 1	VTCNPLQLSPCLGAIITSGQPPSAAACCKLKEQKPCCLCGYIKDPNLCKYVNSPNARKVASSCGVAVPNC
ACM78620. 1	VTCSPLQLSPCAGAITSGGPPSATCCSRLREQTPCLCGYYRDPNLRQYVNSPNARKVARTCGVSVR-C
ACM78623. 1	ATCNPVGLSPCAGAMSGGPPSKLCCSKLREQKPCCLCGYYRDPNLRQYVDSNARKVATACDVHVT-C
ACG48207. 1	TACDATQLTPCAGAIITGRSPSAAACCSRLKEQKPCCLCTYARDPNLRQYVNSPNARKVAAACKVPVPSC
ACG48207. 1	TACDATQLTPCAGAIITGRSPSAAACCSRLKEQKPCCLCTYARDPNLRQYVNSPNARKVAAACKVPVPSC
AAB47967. 1	T-CNALQLTPCAGAIMNAAPTASCCSKMREQKPCCLCQYARDPNLRQYVDSNARKVAAACKVPVPSC

Fig. 4. Multiple sequence alignments of the mature peptides of McLTP1 homologs from different plants with ClustalW2. The target proteins McLTP1 of this study are highlighted with yellow fluorescent background. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

PR-10: Major allergen protein. One putative amino acid sequence from the EST is similar to major allergen protein which belongs to the family PR-10 proteins. The open reading frame (ORF) of the EST encodes a protein of 164 amino acids with a calculated molecular mass of 17696.35 Da and a predicted pI of 8.31. The deduced amino acid sequence named Major Allergen Mac-co 1 presents 30–45% identity to PR-10 proteins from various plant species. The most homolog of Mac-co 1 is JcPR-10a protein with RNase and antifungal activity from *Jatropha curcas* (Agarwal et al., 2013). The motif (SGAPGEPGC) in Mac-co 1 is not a typical glycine-rich loop with the sequence EG (D/N)GG (V/P)G (T/S) which constitutes a signature motif of PR-10 protein (Supplementary Figure 5). According to homologous analysis, the three amino acids Asp100, Glu142, and Tyr144, along with the glycine-rich loop, play a role in RNase activity. Using the SWISS-MODEL, the structure of Mac-co 1 shows similarity to the 3D template 1h2oA which belongs to the PR-10 proteins (Neudecker et al., 2003). To explore the evolutionary

relationship, a phylogenetic tree of Mac-co 1 amino acid sequences, other NCSs, and PR-10 proteins were constructed by a maximum likelihood method, which illustrated that Major allergen proteins and NCSs belong to two distinct blades (Supplementary Figure 6).

Two types ((S)-norcoclaurine synthase and Major allergen protein) of PR-10 proteins, McNCS1 and Major allergen Mac-co 1, were obtained in the present study. In the biosynthesis pathway of benzylisoquinoline alkaloids (BIAs), the norcoclaurine synthase (NCS; EC 4.2.1.78) catalyzes the stereoselective Pictet-Spengler condensation of the L-Tyr derivatives dopamine and 4-hydroxyphenylacetaldehyde (4-HPAA), yielding (S)-norcoclaurine (Lee and Facchini, 2010). The alkaloid sanguinarine is toxic to insects, vertebrates, and microorganisms (Schmeller et al., 1997) through intercalating double-stranded DNA (dsDNA), penetrating membranes, and inhibiting various enzymes containing SH-groups (Weiss et al., 2006). Compared with other known proteins with NCS activity, McNCS1 is so far the only one lacking an extended

N-terminal domain (Ps NCS1, Ps NCS2, Tf NCS, and Cj PR10A) (Lee and Facchini, 2010). Besides, the ubiquitous PR-10 proteins are usually described as small (154–163 amino acids), slightly acidic, and resistant to proteases (Fernandes et al., 2013). However, Major allergen Mac-co 1 is predicted as an essential protein with theoretical pI 8.31. Major allergen Mac-co 1 is a distant homolog in the PR-10 protein superfamily with an untypical glycine-rich loop.

PR-3: Class IV chitinase. We identified one EST as a class IV chitinase gene. The 5'-terminal was truncated. The ORF of the EST is not in full-length. The sequence encodes a partial of mature protein with 73 amino acid residues matching significantly (80% positives) with the class IV chitinase (CAA87074.1) from *Sambucus nigra* that belongs to pathogenesis-related protein, PR-3 type.

2.4. Expression and purification of defensins

With the purpose of constructing the recombinant expression vectors, the coding sequences corresponding to mature domains of McDef1 and McDef2 were produced by PCR, digested with the restriction endonucleases *Xba* I and *Hind* III, and ligated with the yeast expression plasmid pVT102U/ α double-digested with the same endonucleases. The target genes were expressed by *Saccharomyces cerevisiae* 78 carrying the constructs, respectively. The signal peptide of the α -mating factor in pVT102U/ α is considered to induce the sequential amino acid sequences into the culture medium and then be digested with Kex2. One liter of fermentation supernatant was initially separated by chromatography on a CM32 cation-exchange column with gradient concentrations of NaCl elution. The eluent with the solution in 0.3 M of NaCl was further purified on a C18 column. The elution peaks were collected and analyzed using Tricine/SDS-PAGE (Fig. 5), demonstrating the right molecular weight sizes of the two target peptides. The purified defensins were lyophilized for subsequent study.

2.5. Antibacterial activity of recombinant McDef1 and McDef2

The supernatants obtained by the recombinant yeast were used for preliminary assay antibacterial activities against *Escherichia coli*, *Aeromonas hydrophila*, *Staphylococcus aureus*, and *Salmonella pullorum* with the conventional agar-diffusion method. As illustrated in Fig. 6, McDef1

and McDef2 produced by recombinant yeast have significant antibacterial activities, though they are just secreted in the supernatant without concentration and purification against both Gram-negative and Gram-positive bacteria, including all of the four bacteria. Then, the solutions with a concentration gradient (30 μ g/ml, 40 μ g/ml, 50 μ g/ml, 60 μ g/ml, 80 μ g/ml, 100 μ g/ml) of purified recombinant McDef1 and McDef2 were employed to determine the minimum inhibitory concentration (MIC) (Figs. 7 and 8). McDef1 and McDef2 both exhibited the highest activity against *A. hydrophila* with a MIC value of 40 μ g/ml and *S. aureus* with a MIC value of 50 μ g/ml and 60 μ g/ml, respectively. Besides, McDef2 presented antibacterial activities against *S. pullorum* with a MIC value of 60 μ g/ml. In comparison, the MIC of McDef1 against *S. pullorum* was beyond the concentration range tested. The MIC values of McDef1 and McDef2 against *E. coli* neither fell in the concentrations tested, though the clones were less with higher concentration.

Isolating defensin from plant tissues is usually time-consuming and labor-intensive with a low output. In this study, a strategy for large-scale production of purified active defensins has been employed using heterologous expression systems, providing high purity products for further study on the function and structure of the valuable peptides. As Figs. 6–8 show, the antimicrobial activities of McDef1 and McDef2 both act in a concentration-dependent manner against *E. coli*, *A. hydrophila*, *S. aureus*, and *S. pullorum* with different MIC values. Except for the eight cysteines essential for defensin structural folding, the amino acid sequences of McDef1 and McDef2 are highly diverse. This is considered to be responsible for the specificity and diversity of their biological activities (Sagaram et al., 2011). In the present study, these findings provide the opportunity for further investigation on the function and mechanisms and the development of the plant defensins as potential effective antimicrobial agents.

3. Conclusion

In this study, the ESTs from the transcriptome of *M. cordata* leaf were generated and analyzed. Five groups of pathogenesis-related (PR) proteins were described for the first time in the genus of *Macleaya*, including defensins, lipid-transfer proteins, (S)-norococlaurine synthase, major allergen protein, and Class IV chitinase. Furthermore, two previously undescribed plant defensins were produced using heterologous expression systems. Their antimicrobial activities were studied primarily, confirming their antibacterial activity against both Gram-positive and Gram-negative pathogens. This study provides a primary global view of the genetic programs for the leaf of *M. cordata* described so far and sheds new light on the molecular mechanism of antimicrobial and pesticidal function of *M. cordata*. Inevitably, this study only represents a snapshot of PR gene activity in healthy plant leaves that have not been significantly attacked by pathogens or pests. Thus, further work involving infected material is required in future.

4. Materials and methods

4.1. cDNA library construction and EST sequencing

M. cordata leaves were collected from the field for scientific research of Hunan Agricultural University, Changsha, China (28°18'33" N, 113°07'69" E) and instantly frozen in liquid nitrogen for later RNA extraction and EST library construction. TRIzol reagent was purchased from Invitrogen. The mRNA Purification Kit was from the purchased Promega. The Creator SMART cDNA Library Construction Kit and restrictive endonuclease were purchased from TaKaRa (Japan). *Escherichia coli* strain DH5 α was used for cDNA library construction.

The total RNA was prepared according to the previous methods (Chen et al., 2008). Two grams of frozen material was ground into a fine powder in liquid nitrogen, followed by cell lysis in the presence of TRIzol reagent. PolyA (+) RNA was purified from the total RNA on oligo (dT)-cellulose affinity column using the mRNA Purification Kit

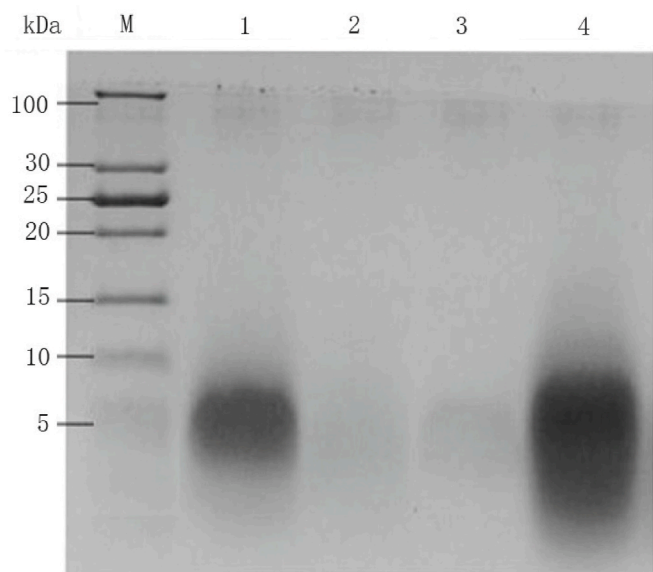


Fig. 5. Tricine-SDS-PAGE analysis of recombinant McDef1 and McDef2 purified from fermentation supernatant. M, Protein standard MW marker; Lane 1, recombinant McDef1; Lane 2, negative control of recombinant McDef1; Lane 3, negative control of recombinant McDef2; Lane 4, recombinant McDef2.

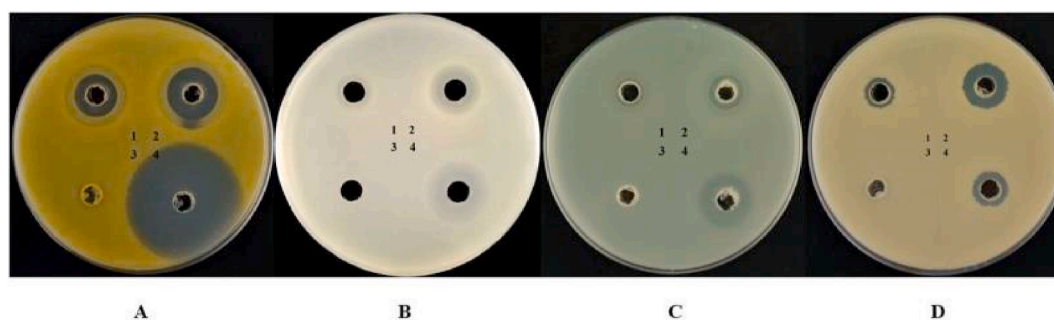


Fig. 6. Antibacterial effects of recombinant yeasts with McDef1 or McDef2 fermentation supernatant on *S. aureus* (A), *S. pullorum* (B), *E. coli* (C) and *A. hydrophila* (D). 1: Fermentation supernatant of the recombinant McDef1; 2: Fermentation supernatant of the recombinant McDef2; 3: Negative control, empty vector expression supernatant; 4: Positive control, A, B, C: 0.2 mg/mL Ampicillin; D: 0.5 mg/mL Kanamycin.

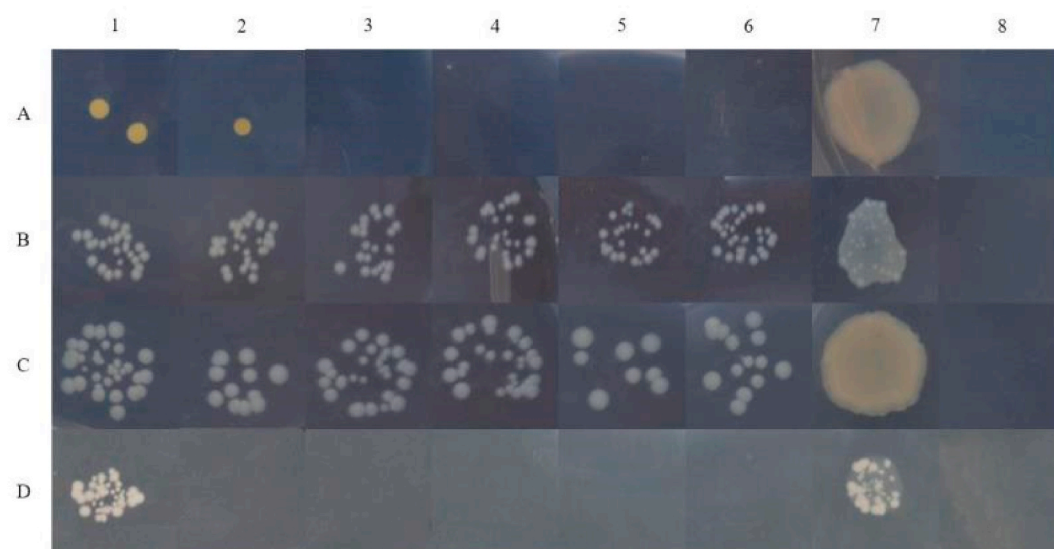


Fig. 7. Antibacterial effects of a series of purified McDef1 concentration against *S. aureus* (A), *E. coli* (B), *S. pullorum* (C) and *A. hydrophila* (D). 1: 30 µg/ml, 2: 40 µg/ml, 3: 50 µg/ml, 4: 60 µg/ml, 5: 80 µg/ml, 6: 100 µg/ml, 7: Negative control, empty vector expression supernatant; 8: Positive control, for A, B, C: 0.2 mg/mL Ampicillin, for D: 0.5 mg/mL Kanamycin.

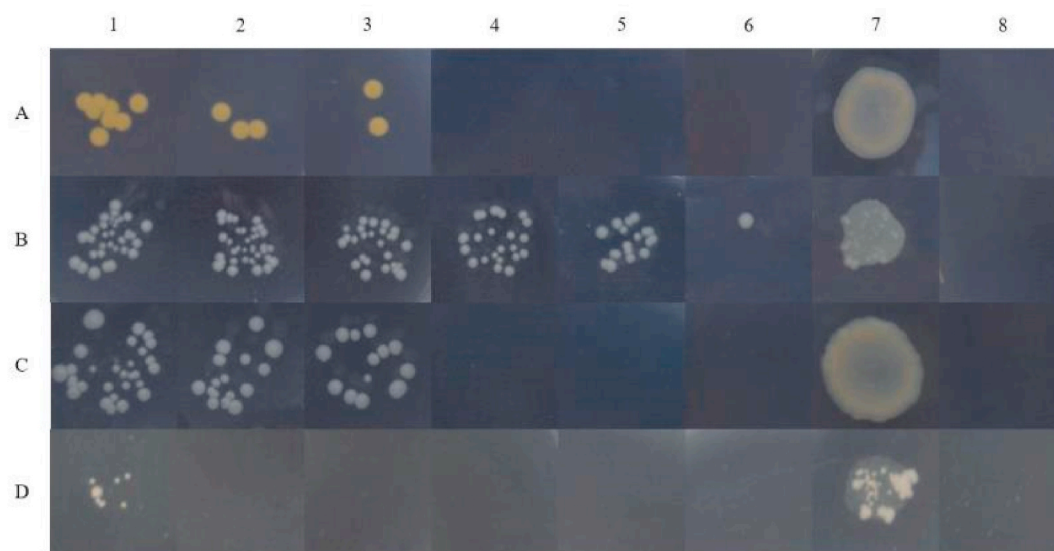


Fig. 8. Antibacterial effects of a series of purified McDef2 concentration against *S. aureus* (A), *E. coli* (B), *S. pullorum* (C) and *A. hydrophila* (D). 1: 30 µg/ml, 2: 40 µg/ml, 3: 50 µg/ml, 4: 60 µg/ml, 5: 80 µg/ml, 6: 100 µg/ml, 7: Negative control, empty vector expression supernatant; 8: Positive control, for A, B, C: 0.2 mg/mL Ampicillin, for D: 0.5 mg/mL Kanamycin.

depending on the manufacturer's protocol. The full-length cDNA library was constructed based on the instructions provided with the Creator SMART cDNA Library Construction Kit. The randomly selected clones were grown in LB medium containing chloramphenicol (30 mg/mL) for 16 h. The plasmids were extracted by alkaline lysis and single-pass sequenced with an automated ABI PRISM 3700 sequencer (PerkinElmer).

4.2. Data processing and bioinformatics analysis

Prior to further analysis, cDNA sequencing outputs were trimmed by removal of vector, primer sequences, and polyA tails with ABI PRISM DNA Sequencing Analysis Software V.3.3 (Chou and Holmes, 2001). High-quality ESTs were assembled into contigs using Phrap software. Default settings were adopted except for 40 bp minimum overlap and 99% identity. Besides, the overall data of the ESTs were analyzed using Blast2GO (Conesa et al., 2005). The database of Clusters of Orthologous Groups of proteins (COGs) (Tatusov et al., 2001) and GO-EnzymeCode Mapping (Kanehisa and Goto, 2000) was identified to provide more useful sequence annotations for comparative studies. Furthermore, ESTs were submitted to GOBlet (Groth et al., 2004) and AmiGO (Ashburner et al., 2000) web servers to improve sequence annotation and acquire insights into the cellular processes that each sequence could be involved in. This procedure aimed to identify relevant Gene Ontology (GO) terms. The result of Gene Ontology (GO) annotation was provided by WEGO (Ye et al., 2006). The signal peptides were predicted with the SignalP 4.0 program (Petersen et al., 2011) (<http://www.cbs.dtu.dk/services/SignalP/>). The sequence similarity was computed with the program EMBOSS (<http://www.ebi.ac.uk/Tools/emboss/align/>). The three-dimensional structures of proteins were predicted using the SWISS-MODEL workspace (Arnold et al., 2006).

4.3. Recombinant expression of the defensins

S. cerevisiae S78 was the strain for recombinant expression of the gene inserted into plasmid pVT102U/α. T4 DNA ligase and Taq DNA polymerase were obtained from Fermentas (Lithuania). Tricine-SDS-PAGE gel preparation kit was purchased from Kangwei (Beijing, China). Primers were synthesized by Sangon (Shanghai, China) as follows: McDef1-F, 5'-CGTCTAGATAAGAGAGAAGAAATGGGACCT-3'; McDef1-R, 5'-CCGAAGCTTGGGTAAATTCTCTTCGC-3'; McDef2-F, 5'-CGTCTAGATAAGAGAGCTATTAAGCCGG-3'; McDef2-R, 5'-CGAAGCTTGGGCTAACATTGGG-3' (The underlined sequences indicate the enzyme cutting site of XbaI and HindIII respectively.)

The recombinant defensins were expressed and purified using the methods described before (Chen et al., 2009). Briefly, the cDNA sequences coding mature peptides of the two defensins, McDef1 and McDef2, were sub-cloned into the vector pVT102U/α with double restriction enzyme sites, XbaI and HindIII, respectively. The recombinant plasmids were subsequently transformed into *S. cerevisiae* S-78 (Leu2, Ura3, Rep4) for recombinant expression. The recombinant expression products were purified by ion-exchange, followed by reverse-phase HPLC. After fermentation, the supernatant of the culture was adjusted to pH 4.2 with acetic acid, centrifuged (15,000 g, 4 °C, 10 min), and filtered through 0.45 μm microfilters. Then, the sample was applied to a CM32-cellulose cation-exchange column, followed by equilibration with 0.1 M sodium acetate. Upon reaching a steady baseline, stepwise elution with 0.1, 0.2, 0.3, and 0.5 M NaCl equilibration buffer was used to wash the column. The fractions were directly applied to a reverse-phase C18 column. Buffer A contained 0.1% trifluoroacetic acid in water; buffer B contained 0.1% trifluoroacetic acid in acetonitrile. Reversed-phase chromatography was conducted using an AKTA purifier chromatography system (Amersham Biosciences AB). The peptide elution was monitored at 215 nm and 280 nm. The fractions were manually collected and analyzed using Tricine/SDS-PAGE.

4.4. Antimicrobial assays

Standard reference bacterial strains (*Staphylococcus aureus* CCTCC AB910393, *Escherichia coli* CCTCC AB91112, *Salmonella pullorum* CCTCC SP7220) from China Center for Type Culture Collection and *Aeromonas hydrophila* provided by the Aquatic Microbiology Laboratory, Hunan Agricultural University were used for assaying antibacterial activities.

The antibacterial activity of the two recombinant defensins against the growth of *E. coli*, *S. aureus*, *A. hydrophila* and *S. pullorum* were determined by the conventional agar-diffusion method. Before agar coagulation, each strain culture containing 10⁶ CFU was mixed with melted LB agar medium and then aseptically poured into a Petri dish. After solidification at room temperature, 4 holes of 6-mm diameter were punched on each plate using a sterilized pipette tip. The fermentation supernatant of the recombination expression was applied into the hole. The triplicate plates were incubated at 37 °C for 24 h. Afterward, the inhibition zones were measured.

4.5. Minimal bactericidal concentration

The four strains, *E. coli*, *S. aureus*, *A. hydrophila*, and *S. pullorum* were cultured to logarithmic growth period (OD_{600 nm} = 0.4–0.6) and diluted to 10⁴ cfu/mL. Besides, 5 μL of the solution of the purified recombinant expression product (30 μg/mL, 40 μg/mL, 50 μg/mL, 60 μg/mL, 80 μg/mL, 100 μg/mL) was mixed with 5 μL of bacterial liquid in a PCR tube. Then, 10 μL of each mixture was dropped on LB medium plate and incubated at 37 °C. The growth of bacteria was observed after 16 h incubation. Each experiment was performed at least three times for each species of bacteria.

Declaration of competing interest

There are no conflicts to declare.

Acknowledgments

This work was supported by the Educational Department Foundation of Hunan Province [grant number 16A098], Hunan Provincial Science and Technology Department [grant number 2017NK2311], and National Natural Science Foundation of China [grant number 30900242]. We are grateful to Dr. Emre Bektik at Boston Children's Hospital and Ms. Phraew Sakon at Brigham and Women's Hospital, Harvard Medical School for the professional revision of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.phytochem.2021.112667>.

References

- Agarwal, P., et al., 2013. Pathogenesis-related gene, JcPR-10a from *Jatropha curcas* exhibit RNase and antifungal activity. *Mol. Biotechnol.* 54 (2), 412–425. <https://doi.org/10.1007/s12033-012-9579-7>.
- Arnold, K., et al., 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22 (2), 195–201. <https://doi.org/10.1093/bioinformatics/bti770>.
- Ashburner, M., et al., 2000. Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium. Nat. Genet.* 25 (1), 25–29. <https://doi.org/10.1038/75556>.
- Carvalho, A.D.O., Gomes, V.M., 2009. Plant defensins—prospects for the biological functions and biotechnological properties. *Peptides* 30 (5), 1007–1020. <https://doi.org/10.1016/j.peptides.2009.01.018>.
- Chen, J., et al., 2008. Transcriptome analysis revealed novel possible venom components and cellular processes of the tarantula *Chilobrachys jingzhao* venom gland. *Toxicon* 52 (7), 794–806. <https://doi.org/10.1016/j.toxicon.2008.08.003>.
- Chen, J., et al., 2009. Expression and characterization of jingzhaotoxin-34, a novel neurotoxin from the venom of the tarantula *Chilobrachys jingzhao*. *Peptides* 30 (6), 1042–1048. <https://doi.org/10.1016/j.peptides.2009.02.018>.
- Chou, H.H., Holmes, M.H., 2001. DNA sequence quality trimming and vector removal. *Bioinformatics* 17 (12), 1093–1104.

- Conesa, A., et al., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21 (18), 3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>.
- D'Agostino, N., et al., 2019. Identification of non-specific Lipid Transfer Protein gene family members in *Solanum lycopersicum* and insights into the features of Sola 13 protein. *Sci. Rep.* 9 (1), 1607. <https://doi.org/10.1038/s41598-018-38301-z>.
- EFSA, E.F.S.A., 2010. Statement on the use of feed additives authorised/applied for use in feed when supplied via water. *EFSA Journal* 8 (12), 1956. <https://doi.org/10.2903/j.efsa.2010.1956>.
- Fernandes, H., et al., 2013. Structural and functional aspects of PR-10 proteins. *FEBS J.* 280 (5), 1169–1199. <https://doi.org/10.1111/febs.12114>.
- Gonorazky, A.G., et al., 2005. Stress induction and antimicrobial properties of a lipid transfer protein in germinating sunflower seeds. *J. Plant Physiol.* 162 (6), 618–624. <https://doi.org/10.1016/j.jplph.2004.10.006>.
- Groth, D., et al., 2004. GOblet: a platform for Gene Ontology annotation of anonymous sequence data. *Nucleic Acids Res.* 32, W313–W317. <https://doi.org/10.1093/nar/gkh406>. Web Server issue).
- Hammerová, J., et al., 2011. Benzo[c]phenanthridine alkaloids exhibit strong anti-proliferative activity in malignant melanoma cells regardless of their p53 status. *J. Dermatol. Sci.* 62 (1), 22–35. <https://doi.org/10.1016/j.jdermsci.2011.01.006>.
- Hincha, D.K., 2002. Cryoprotectin: a plant lipid-transfer protein homologue that stabilizes membranes during freezing. *Phil. Trans. Roy. Soc. Lond. B Biol. Sci.* 357 (1423), 909–916. <https://doi.org/10.1098/rstb.2002.1079>.
- Hoh, F., et al., 2005. Structure of a liganded type 2 non-specific lipid-transfer protein from wheat and the molecular basis of lipid binding. *Acta Crystallogr D Biol. Crystallogr* 61 (Pt 4), 397–406. <https://doi.org/10.1107/S0907444905000417>.
- Ilari, A., et al., 2009. Structural basis of enzymatic (S)-norcoclaurine biosynthesis. *J. Biol. Chem.* 284 (2), 897–904. <https://doi.org/10.1074/jbc.M803738200>.
- Isaac Kirubakaran, S., et al., 2008. Characterization of a new antifungal lipid transfer protein from wheat. *Plant Physiol. Biochem.* 46 (10), 918–927. <https://doi.org/10.1016/j.plaphy.2008.05.007>.
- Jung, H., et al., 2005. Identification of pathogen-responsive regions in the promoter of a pepper lipid transfer protein gene (CALTPI) and the enhanced resistance of the CALTPI transgenic *Arabidopsis* against pathogen and environmental stresses. *Planta* 221 (3), 361–373. <https://doi.org/10.1007/s00425-004-1461-9>.
- Kader, J.C., 1975. Proteins and the intracellular exchange of lipids. I. Stimulation of phospholipid exchange between mitochondria and microsomal fractions by proteins isolated from potato tuber. *Biochim. Biophys. Acta* 380 (1), 31–44.
- Kader, J.C., 1996. Lipid-transfer proteins in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47, 627–654. <https://doi.org/10.1146/annurev.arplant.47.1.627>.
- Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28 (1), 27–30. <https://doi.org/10.1093/nar/28.1.27>.
- Kaur, J., et al., 2011. Can plant defensins be used to engineer durable commercially useful fungal resistance in crop plants? *Fungal Biology Reviews* 25 (3), 128–135. <https://doi.org/10.1016/j.fbr.2011.07.004>.
- Kosina, P., et al., 2010. Phytochemical and antimicrobial characterization of *Macleaya cordata* herb. *Fitoterapia* 81 (8), 1006–1012. <https://doi.org/10.1016/j.fitote.2010.06.020>.
- Lee, E.J., Faccini, P., 2010. Norcoclaurine synthase is a member of the pathogenesis-related 10/Bet v1 protein family. *Plant Cell* 22 (10), 3489–3503. <https://doi.org/10.1105/tpc.110.077958>.
- Lin, L., et al., 2018. Medicinal plants of the genus *Macleaya* (*Macleaya cordata*, *Macleaya microcarpa*): a review of their phytochemistry, pharmacology, and toxicology. *Phytother. Res.* 32 (1), 19–48. <https://doi.org/10.1002/ptr.5952>.
- Liu, M., et al., 2013. In vitro assessment of *Macleaya cordata* crude extract bioactivity and anticancer properties in normal and cancerous human lung cells. *Exp. Toxicol. Pathol.* 65 (6), 775–787. <https://doi.org/10.1016/j.etp.2012.11.004>.
- Michaelson, D., et al., 1992. Cationic defensins arise from charge-neutralized propeptides: a mechanism for avoiding leukocyte autotoxicity? *J. Leukoc. Biol.* 51 (6), 634–639.
- Ming, Z., et al., 2011. Evaluation of molluscicidal activities of benzo [c] phenanthridine alkaloids from *Macleaya cordata* (Wild) R. Br. on snail hosts of *Schistosoma japonicum*. *J. Med. Plants Res.* 5 (4), 521–526.
- Munoz, A., et al., 2014. Specific domains of plant defensins differentially disrupt colony initiation, cell fusion and calcium homeostasis in *Neurospora crassa*. *Mol. Microbiol.* 92 (6), 1357–1374. <https://doi.org/10.1111/mmi.12634>.
- Neudecker, P., et al., 2003. Mutational epitope analysis of Pru av 1 and Api g 1, the major allergens of cherry (*Prunus avium*) and celery (*Apium graveolens*): correlating IgE reactivity with three-dimensional structure. *Biochem. J.* 376 (1), 97–107. <https://doi.org/10.1042/bj20031057>.
- Ng, T.B., et al., 2012. Lipid-transfer proteins. *Peptide Science* 98 (4), 268–279. <https://doi.org/10.1002/bip.22098>.
- Petersen, T.N., et al., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8 (10), 785–786. <https://doi.org/10.1038/nmeth.1701>.
- Pyee, J., Kolattukudy, P.E., 1995. The gene for the major cuticular wax-associated protein and three homologous genes from broccoli (*Brassica oleracea*) and their expression patterns. *Plant J.* 7 (1), 49–59. <https://doi.org/10.1046/j.1365-313X.1995.07010049.x>.
- Sagaram, U.S., et al., 2011. Structure-activity determinants in antifungal plant defensins MsDef1 and MtDef4 with different modes of action against *Fusarium graminearum*. *PLoS One* 6 (4). <https://doi.org/10.1371/journal.pone.0018550> e18550.
- Salminen, T.A., et al., 2016. Lipid transfer proteins: classification, nomenclature, structure, and function. *Planta* 244 (5), 971–997. <https://doi.org/10.1007/s00425-016-2585-4>.
- Sathoff, A.E., et al., 2019. Plant defensin peptides have antifungal and antibacterial activity against human and plant pathogens. *Phytopathology* 109 (3), 402–408. <https://doi.org/10.1094/PHYTO-09-18-0331-R>.
- Schmeller, T., et al., 1997. Biochemical activities of berberine, palmatine and sanguinarine mediating chemical defence against microorganisms and herbivores. *Phytochemistry* 44 (2), 257–266. [https://doi.org/10.1016/S0031-9422\(96\)00545-6](https://doi.org/10.1016/S0031-9422(96)00545-6).
- Segura, A., et al., 1998. Novel defensin subfamily from spinach (*Spinacia oleracea*). *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 435 (2–3), 159–162. [https://doi.org/10.1016/S0014-5793\(98\)01060-6](https://doi.org/10.1016/S0014-5793(98)01060-6).
- Sels, J., et al., 2008. Plant pathogenesis-related (PR) proteins: a focus on PR peptides. *Plant Physiol. Biochem.* 46 (11), 941–950. <https://doi.org/10.1016/j.plaphy.2008.06.011>.
- Tatusov, R.L., et al., 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* 29 (1), 22–28. <https://doi.org/10.1093/nar/29.1.22>.
- Thevissen, K., et al., 2003. DmAMP1, an antifungal plant defensin from dahlia (*Dahlia merckii*), interacts with sphingolipids from *Saccharomyces cerevisiae*. *FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Lett.* 226 (1), 169–173. [https://doi.org/10.1016/S0378-1097\(03\)00590-1](https://doi.org/10.1016/S0378-1097(03)00590-1).
- Van der Weerden, N.L., Anderson, M.A., 2013. Plant defensins: common fold, multiple functions. *Fungal Biology Reviews* 26 (4), 121–131. <https://doi.org/10.1016/j.fbr.2012.08.004>.
- Van Loon, L.C., et al., 2006. Significance of inducible defense-related proteins in infected plants. *Annu. Rev. Phytopathol.* 44, 135–162. <https://doi.org/10.1146/annurev.phyto.44.070505.143425>.
- Wang, S.Y., et al., 2004. A non-specific lipid transfer protein with antifungal and antibacterial activities from the mung bean. *Peptides* 25 (8), 1235–1242. <https://doi.org/10.1016/j.peptides.2004.06.004>.
- Wang, N.-J., et al., 2012. Construction and analysis of a plant non-specific lipid transfer protein database (nsLTPDB). *BMC Genom.* 13 (Suppl. 1), S9.
- Weiss, D., et al., 2006. Sanguinarine reductase, a key enzyme of benzophenanthridine detoxification. *Plant Cell Environ.* 29 (2), 291–302. <https://doi.org/10.1111/j.1365-3040.2005.01421.x>.
- Ye, J., et al., 2006. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 34 (Suppl. 2), W293–W297. <https://doi.org/10.1093/nar/gkl031>.
- Zeng, J., et al., 2013. Integration of transcriptome, proteome and metabolism data reveals the alkaloids biosynthesis in *Macleaya cordata* and *Macleaya microcarpa*. *PLoS One* 8 (1), e53409. <https://doi.org/10.1371/journal.pone.0053409>.
- Zhu, X., et al., 2012. Overexpression of wheat lipid transfer protein gene TaLTP5 increases resistances to *Cochliobolus sativus* and *Fusarium graminearum* in transgenic wheat. *Funct. Integr. Genom.* 12 (3), 481–488. <https://doi.org/10.1007/s10142-012-0286-z>.

Supplementary Tables

Supplementary Table.1 Summary of *Cordata(Willd.) R. Br.* Leaf cDNA library sequencing results

Result Name	Number	Length (bp)
Sequenced ESTs	883	/
High quality ESTs	511	/
Non-redundant sequences (singltons+clusters)	364	/
Number of clusters	78	/
Number of ESTs in clusters	223	/
Number of singltons	286	/
Average sequences/clusters	3.62	/
Average non-redundant sequence size	/	513
Shortest non-redundant sequence size	/	105
Longest non-redundant sequence size	/	1013

Supplementary Table.2 The KEGG profiles of transcripts from *Macleaya cordata*

Pathway	Enzyme	Enzyme ID	Seqs
Glycolysis / Gluconeogenesis	alcohol dehydrogenase	ec:1.1.1.1	1
Fatty acid degradation	alcohol dehydrogenase	ec:1.1.1.1	1
Glycine, serine and threonine metabolism	alcohol dehydrogenase	ec:1.1.1.1	1
Tyrosine metabolism	alcohol dehydrogenase	ec:1.1.1.1	1
Retinol metabolism	alcohol dehydrogenase	ec:1.1.1.1	1
Chloroalkane and chloroalkene degradation	alcohol dehydrogenase	ec:1.1.1.1	1
Drug metabolism - cytochrome P450	alcohol dehydrogenase	ec:1.1.1.1	1
Metabolism of xenobiotics by cytochrome P450	alcohol dehydrogenase	ec:1.1.1.1	1
Naphthalene degradation	alcohol dehydrogenase	ec:1.1.1.1	1
Butanoate metabolism	hydroxybutyrate dehydrogenase	ec:1.1.1.30	1
Synthesis and degradation of ketone bodies	hydroxybutyrate dehydrogenase	ec:1.1.1.30	1
Valine, leucine and isoleucine degradation	3-hydroxyisobutyrate dehydrogenase	ec:1.1.1.31	1

Pentose phosphate pathway	phosphogluconate dehydrogenase (NADP+-dependent, decarboxylating)	ec:1.1.1.44	2
Glutathione metabolism	phosphogluconate dehydrogenase (NADP+-dependent, decarboxylating)	ec:1.1.1.44	2
Glyoxylate and dicarboxylate metabolism	2-hydroxy-3-oxopropionate reductase	ec:1.1.1.60	1
Oxidative phosphorylation	ubiquinol---cytochrome-c reductase	ec:1.10.2.2	2
Arachidonic acid metabolism	monooxygenase	ec:1.14.14.1	1
Fatty acid degradation	monooxygenase	ec:1.14.14.1	1
Linoleic acid metabolism	monooxygenase	ec:1.14.14.1	1
Steroid hormone biosynthesis	monooxygenase	ec:1.14.14.1	1
Tryptophan metabolism	monooxygenase	ec:1.14.14.1	1
Retinol metabolism	monooxygenase	ec:1.14.14.1	1
Caffeine metabolism	monooxygenase	ec:1.14.14.1	1
Aminobenzoate degradation	monooxygenase	ec:1.14.14.1	1
Drug metabolism - cytochrome P450	monooxygenase	ec:1.14.14.1	1
Metabolism of xenobiotics by cytochrome P450	monooxygenase	ec:1.14.14.1	1
Steroid biosynthesis	7-dehydrocholesterol reductase	ec:1.3.1.21	2
Vitamin B6 metabolism	pyridoxal 5'-phosphate synthase	ec:1.4.3.5	1
Oxidative phosphorylation	NADH:ubiquinone reductase (H ⁺ -translocating)	ec:1.6.5.3	2
Oxidative phosphorylation	NADH dehydrogenase	ec:1.6.99.3	1
Sulfur metabolism	phosphoadenylyl-sulfate reductase (thioredoxin)	ec:1.8.4.8	1
Sulfur metabolism	adenylyl-sulfate reductase (glutathione)	ec:1.8.4.9	1
Sulfur metabolism	adenosine 5'-phosphosulfate	ec:1.8.99.2	1

	reductase		
Oxidative phosphorylation	cytochrome-c oxidase	ec:1.9.3.1	1
Glycerophospholipid metabolism	phosphoethanolamine N-methyltransferase	ec:2.1.1.103	1
Lysine degradation	histone-lysine N-methyltransferase	ec:2.1.1.43	1
Pentose phosphate pathway	transketolase	ec:2.2.1.1	1
Carbon fixation in photosynthetic organisms	transketolase	ec:2.2.1.1	1
Biosynthesis of ansamycins	transketolase	ec:2.2.1.1	1
Sulfur metabolism	serine O-acetyltransferase	ec:2.3.1.30	1
Cysteine and methionine metabolism	serine O-acetyltransferase	ec:2.3.1.30	1
Galactose metabolism	galactinol synthase	ec:2.4.1.123	1
Purine metabolism	adenine phosphoribosyltransferase	ec:2.4.2.7	1
N-Glycan biosynthesis	dolichyl-diphosphooligosaccharid e---protein glycotransferase	ec:2.4.99.18	1
Glutathione metabolism	glutathione transferase	ec:2.5.1.18	2
Drug metabolism - cytochrome P450	glutathione transferase	ec:2.5.1.18	2
Metabolism of xenobiotics by cytochrome P450	glutathione transferase	ec:2.5.1.18	2
Cysteine and methionine metabolism	S-adenosylmethionine synthetase	ec:2.5.1.6	2
Plant-pathogen interaction	LRR receptor-like serine/threonine-protein kinase FLS2	ec:2.7.11.1	2
MAPK signaling pathway	mitogen-activated protein kinase	ec:2.7.11.24	2
Purine metabolism	nucleoside-diphosphate kinase	ec:2.7.4.6	1
Pyrimidine metabolism	nucleoside-diphosphate kinase	ec:2.7.4.6	1
Sulfur metabolism	adenylyltransferase	ec:2.7.7.4	1

Purine metabolism	adenylyltransferase	ec:2.7.7.4	1
Selenocompound metabolism	adenylyltransferase	ec:2.7.7.4	1
Methane metabolism	phosphoserine phosphatase	ec:3.1.3.3	1
Glycine, serine and threonine metabolism	phosphoserine phosphatase	ec:3.1.3.3	1
Amino sugar and nucleotide sugar metabolism	chitinase	ec:3.2.1.14	1
Arginine and proline metabolism	N-carbamoylputrescine amidase	ec:3.5.1.53	1
Purine metabolism	microtubule-severing ATPase	ec:3.6.4.3	1
Glyoxylate and dicarboxylate metabolism	ribulose 1,5-bisphosphate carboxylase/oxygenase	ec:4.1.1.39	2
Carbon fixation in photosynthetic organisms	ribulose 1,5-bisphosphate carboxylase/oxygenase	ec:4.1.1.39	2
Cysteine and methionine metabolism	S-adenosylmethionine decarboxylase	ec:4.1.1.50	1
Propanoate metabolism	malonyl-CoA decarboxylase	ec:4.1.1.9	1
beta-Alanine metabolism	malonyl-CoA decarboxylase	ec:4.1.1.9	1
Nitrogen metabolism	carbonic anhydrase	ec:4.2.1.1	1
Isoquinoline alkaloid biosynthesis	(S)-norcoclaurine synthase	ec:4.2.1.78	1
Ubiquitin mediated proteolysis	peptidyl-prolyl isomerase	cis-trans ec:5.2.1.8	1
Aminoacyl-tRNA biosynthesis	glycyl-tRNA synthetase / glycine--tRNA ligase	ec:6.1.1.14	1
Glyoxylate and dicarboxylate metabolism	glutamine synthetase	ec:6.3.1.2	1
Nitrogen metabolism	glutamine synthetase	ec:6.3.1.2	1
Alanine, aspartate and glutamate metabolism	glutamine synthetase	ec:6.3.1.2	1
Arginine and proline metabolism	glutamine synthetase	ec:6.3.1.2	1
Circadian rhythm – plant	ubiquitin---protein ligase	ec:6.3.2.19	2

Alanine, aspartate and glutamate metabolism	argininosuccinate synthase	ec:6.3.4.5	1
Arginine and proline metabolism	argininosuccinate synthase	ec:6.3.4.5	1

Supplementary Table 3 Theoretical pI/Mw of McDef1-5

Protein	MASS(Da)			pI		
	precursor	signal	mature	precursor	signal	mature
McDef1	9645.31	2596.18	7067.15	8.73	9.35	8.52
McDef2	10286.92	3404.16	6900.78	8.15	10	7.05
McDef3	10276.88	3394.12	6900.78	8.15	10	7.05
McDef4	10303.91	3408.15	6913.77	6.86	10	6.02
McDef5	10302.97	3408.15	6912.83	8.15	10	7.05

Supplementary Figure legends

Supplementary Figure 1 The length distribution of ESTs.

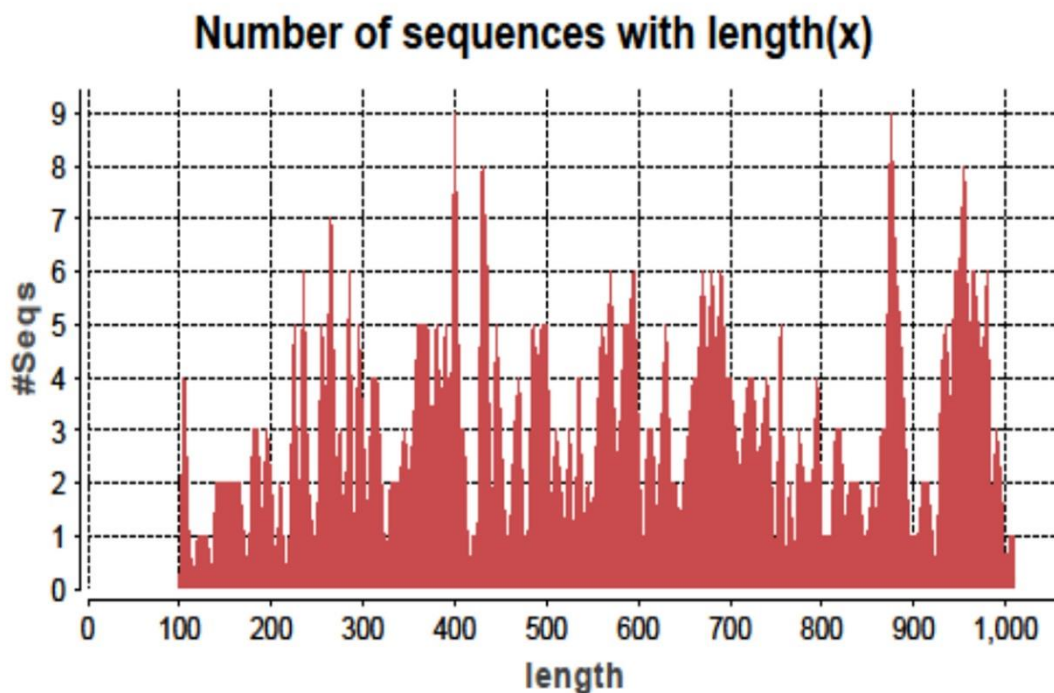
Supplementary Figure 2 The results of analysis of the Gene Ontology (GO)

Supplementary Figure 3 Analysis of the Cluster of Orthologous Groups (A: RNA processing and modification B: Chromatin Structure and dynamics C: Energy production and conversion D: Cell cycle control and mitosis E Amino Acid metabolism and transport F Nucleotide metabolism and transport G: Carbohydrate metabolism and transport H: Coenzyme metabolism I: Lipid metabolism J: Translation K: Transcription L: Replication and repair M: Cell wall/membrane/envelop biogenesis N: Cell motility O: Post-translational modification, protein turnover, chaperone functions P: Inorganic ion transport and metabolism Q: Secondary Structure T: Signal Transduction U: Intracellular trafficking and secretion Y: Nuclear structure Z: Cytoskeleton R General Functional Prediction only S: Function Unknown).

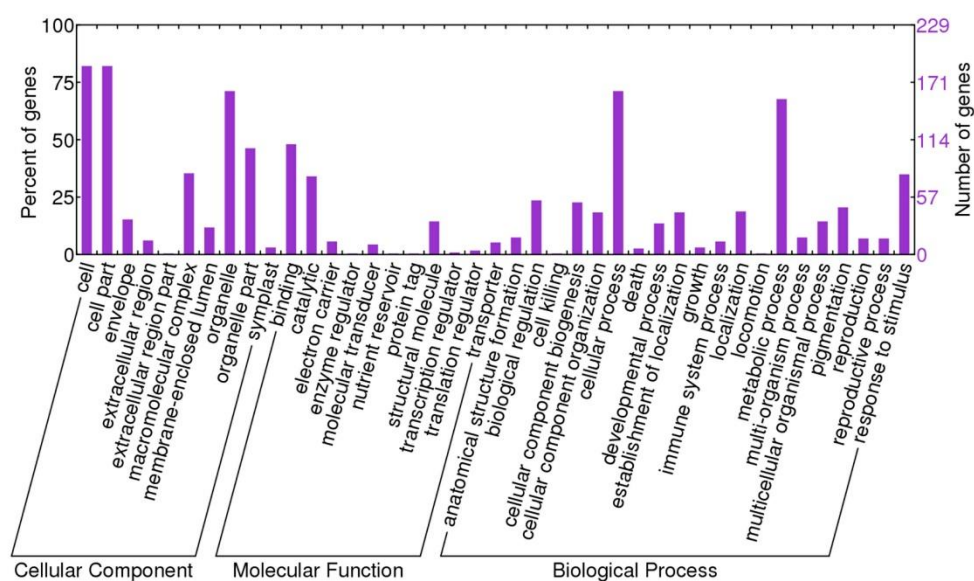
Supplementary Figure 4. Sequence alignment of McNCS 1

Supplementary Figure 5. Sequence alignment of Mac-co 1

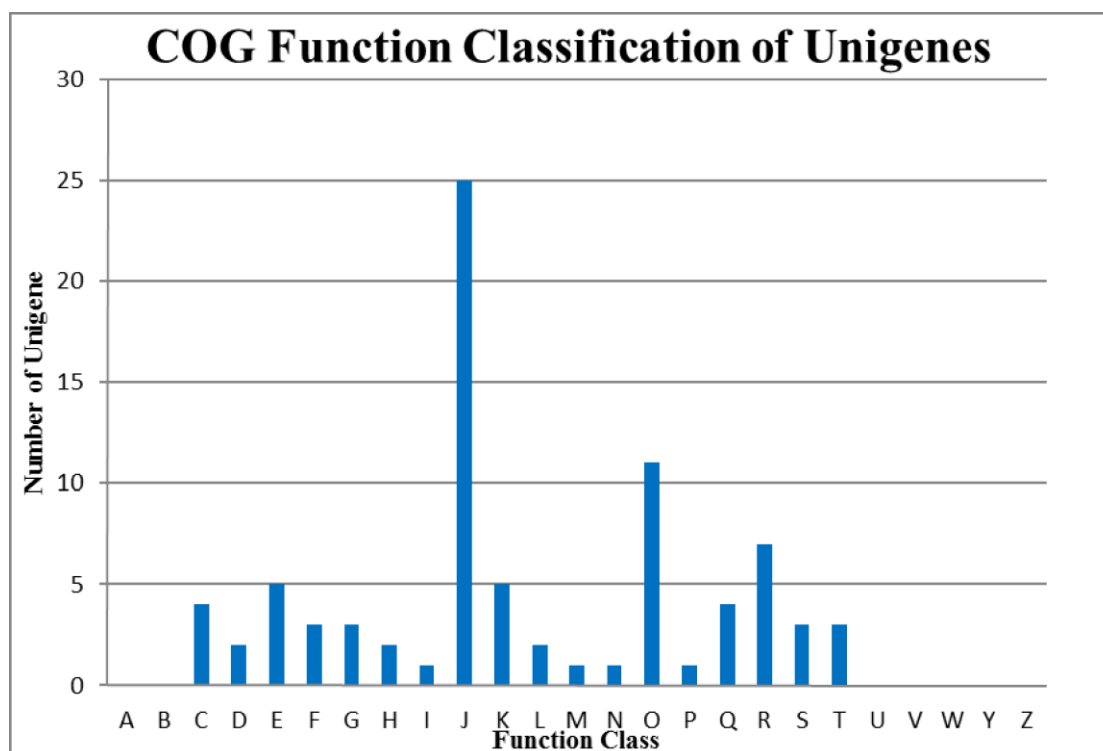
Supplementary Figure 6. The phylogenetic tree of (S)-norcoclaurine synthase and major allergen protein



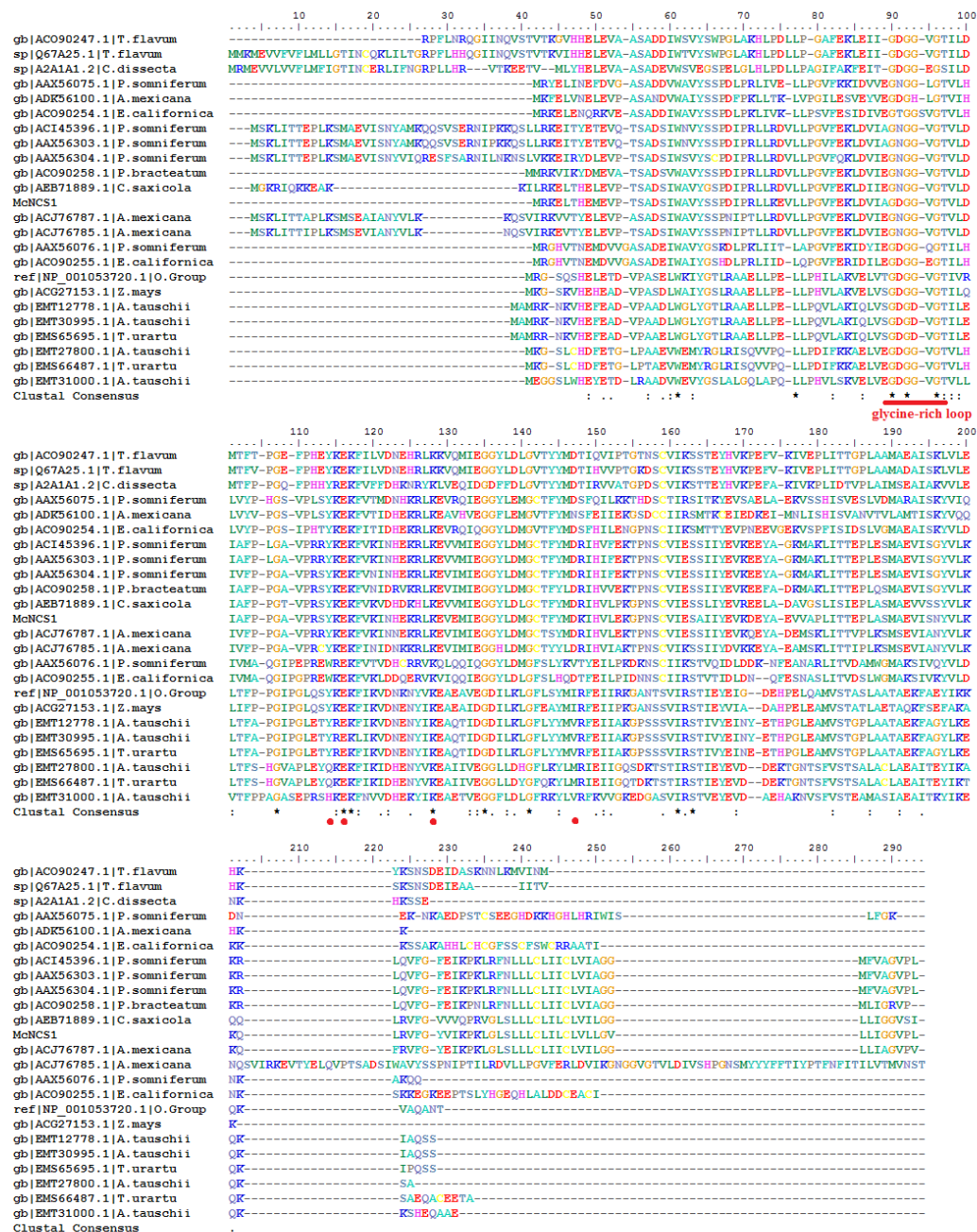
Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4

