

# Discovery and Characterization of Novel Cyclotides Originated from Chimeric Precursors Consisting of Albumin-1 Chain a and Cyclotide Domains in the Fabaceae Family<sup>1,2</sup>

Giang Kien Truc Nguyen, Sen Zhang, Ngan Thi Kim Nguyen, Phuong Quoc Thuc Nguyen, Ming Sheau Chiu, Antony Hardjojo, and James P. Tam<sup>a</sup>

From the School of Biological Sciences, Nanyang Technological University, Singapore 637551

<sup>a</sup> To whom correspondence should be addressed.  
Fax: 65-6515-1632; E-mail: jptam@ntu.edu.sg.

The tropical plant *Clitoria ternatea* is a member of the Fabaceae family well known for its medicinal values. Heat extraction of *C. ternatea* revealed that the bioactive fractions contained heat-stable cysteine-rich peptides (CRPs). The CRP family of A1b (Albumin-1 chain b/leginsulins), which is a linear cystine knot CRP, has been shown to present abundantly in the Fabaceae. In contrast, the cyclotide family, which also belongs to the cystine knot CRPs but with a cyclic structure, is commonly found in the Rubiaceae, Violaceae, and Cucurbitaceae families. In this study, we report the discovery of a panel of 15 heat-stable CRPs, of which 12 sequences (clotide T1–T12) are novel. We show unambiguously that the clotides are cyclotides and not A1bs, as determined by their sequence homology, disulfide connectivity, and membrane active properties indicated by their antimicrobial activities against *Escherichia coli* and cytotoxicities to HeLa cells. We also show that clotides are prevalent in *C. ternatea* and are found in every plant tissue examined, including flowers, seeds, and nodules. In addition, we demonstrate that their precursors are chimeras, half from cyclotide and the other half from Albumin-1, with the cyclotide domain displacing the A1b domain in the precursor. Their chimeric structures likely originate from either horizontal gene transfer or convergent evolution in plant nuclear genomes, which are exceedingly rare events. Such atypical genetic arrangement also implies a different mechanism of biosynthetic processing of cyclotides in the Fabaceae and provides new understanding of their evolution in plants.

---

<sup>1</sup> This work was supported by A\*STAR Biomedical Research Council Grant 09/1/22/19/612 and Academic Research Fund Grant ARC21/08 from the Ministry of Education in Singapore.

<sup>2</sup> The on-line version of this article (available at <http://www.jbc.org>) contains supplemental text and Tables S1–S6. The nucleotide sequence(s) reported in this paper has been submitted to the DDBJ/GenBank™/EBI Data Bank with accession number(s) JF931988, JF931989, JF931990, JF931991, JF931992, JF931993, JF931994, JF931995, JF931996, JF931997, and JF931998.

*Clitoria ternatea* is a perennial climber well known for its butterfly-shaped vivid blue flowers. It belongs to the Fabaceae family and displays a broad spectrum of medicinal usages in almost every plant part (1). It is often cited in the Ayurvedic system of Indian medicines as an effective antidepressant, antimicrobial, antipyretic, and nerve tonic for enhancing memory (1). It is also used as an alternative medicine in America and other tropical Asian countries (1). In Cuba, decoctions of roots and flowers are reported to have emmenagogue properties that promote menstruation and uterine contraction. Studies on animals have shown that the aqueous extracts of the flowers and leaves display antihyperglycemic effects in rats (2). Decoctions of roots and leaves elicit a wide spectrum of activities on the central nervous system and have been shown to enhance acetylcholine content in rat hippocampus (3–5).

Preliminary phytochemical screenings of *C. ternatea* extracts have shown that the biologically active fractions are rich in peptides and proteins, while showing negative tests for alkaloids, saponins, flavonoids, coumarins, and lignans (3, 6). In these studies, plant extracts were prepared by boiling the pulverized plants in hot water. Although the exact chemical components have not been identified, it is plausible to speculate that the active principles are heat-stable proteins. The view of proteins as viable bioactive herbal components, however, is conceptually contradictory to our current knowledge of traditional medicines, which has been biased toward small molecules with molecular masses less than 500 Da. Peptides and proteins whose molecular masses are considerably larger than 500 Da have generally not been considered as active principles with the common perception that they are unstable and unavailable as a source of active principles in decoctions. This bias is partly attributed to the intrinsic instability of peptides and proteins against heat during decoction preparation or their susceptibility to enzymatic and acidic hydrolysis during digestion. However, recent literature precedents suggest otherwise.

Cumulative evidence shows that several classes of cysteine-rich peptides (CRPs)<sup>2</sup> in plants such as defensins, A1bs (also known as leginsulins), knottins, and cyclotides are heat-stable (7–11). Although their primary sequences, biochemical properties, and functions may differ greatly, CRPs possess multiple disulfide bridges that cross-brace their structures, often conferring thermal, chemical, and enzymatic stability (8, 12). Of these, A1bs are well documented CRPs characteristic of the Fabaceae family and have been identified in several legume species (13, 14). They consist of 35–40 amino acid residues in length and contain three disulfide bridges (13). A1bs are highly stable and able to survive the acids and digestive enzymes in the porcine stomach and intestine (15, 16). They have been shown to possess several biological activities such as insecticidal and hormonal functions in plants (17, 18). They also affect mammalian physiological functions such as regulation of glucose metabolism in mice (19).

Cyclotides are another class of CRPs that has recently gained interest because of their extraordinary stability and diverse biological functions (20, 21). They are macrocyclic peptides from plants of the Rubiaceae, Violaceae, and Cucurbitaceae families (22, 23). They contain 28–37 residues and are structurally distinguished from the conventional linear CRPs such as A1bs by being cyclic. They possess a circular peptide backbone fortified by a cystine knot motif (20, 25). Such structural elements produce their stability

to heat, chemical, and enzymatic degradation (26). These advantages enhance the potential of cyclotides as active principles in traditional medicines.

Very recently, Poth *et al.* (27) reported the isolation of 12 novel cyclotides from the seeds of *C. ternatea*, which marked the first discovery of cyclotides in the Fabaceae family. Their finding prompts us to report our independent work on the discovery of cyclotides in this new plant family. Here, we describe and report the isolation and characterization of novel cyclotides as heat-stable CRPs in every plant part of *C. ternatea* including leaves, stems, roots, shoots, nodules, flowers, pods, and seeds. Fifteen cyclotides were isolated, of which 12 sequences are novel and designated as cliotide T1–T12. Our results show unequivocally that cliotides, at the protein level, belong to the cyclotide and not the A1b family. They were found to preserve the cystine knot motif, active against *Escherichia coli* and cytotoxic to HeLa cells. These findings suggest that cliotides may be the active ingredients responsible for the indications of *C. ternatea* in traditional medicines as anti-infectives and antitumorogenic agents (28). Interestingly, molecular cloning of novel cliotide genes revealed that they are derived from an entirely new genetic arrangement differing from the known cyclotides of the Rubiaceae and Violaceae families. Cliotides, at the gene level, have chimeric structures containing genetic elements from both Albumin-1 (A1) of the Fabaceae and cyclotides of the Rubiaceae and Violaceae families. This unusual arrangement suggests a novel mode of biosynthetic processing of cyclotides in the Fabaceae family and provides hints about their evolution in plants.

## EXPERIMENTAL PROCEDURES

*Screening for Heat-stable Biologics in C. ternatea*—Fresh flowers (~50 mg) were macerated with 500  $\mu$ l of water and incubated at 100 °C for 30 min. The heat-stable fraction was separated from the denatured protein precipitates and other plant debris by centrifugation. For comparison, a control experiment was performed without the heat treatment. The aqueous extracts with or without heating were monitored by MS and HPLC.

*Peptides Isolation and Purification*—Fresh plant materials (~1 kg, whole plant) were collected and homogenized with 5 liters of water and incubated at 100 °C for 1 h. After filtration, the aqueous extract was subjected to a flash chromatography column packed with 150 g of C18 material (Grace Davison). The column was washed with 20% ethanol and eluted with 80% ethanol to obtain the CRP-enriched fraction. This fraction was concentrated and subjected to preparative HPLC on Shimadzu system with a C18 Vydac column (250  $\times$  21 mm) at a flow rate of 8 ml/min. A linear gradient of 1% min<sup>-1</sup> of 0 – 80% buffer B (100% acetonitrile, 0.05% trifluoroacetic acid) was applied, and the eluants were monitored by UV detection at 220, 254, and 280 nm. The fractions obtained were repurified by semi-preparative HPLC with a C18 Vydac column (250  $\times$  10 mm) at a flow rate of 3 ml/min with the same gradient. The approximate yields obtained for cliotides T1, T2, T3, T4, T7, T8, T9, and T10 and Cter B were 5, 10, 70, 6, 50, 80, 80, 80, and 70 mg, respectively.

*S-Reduction and S-Alkylation*—Approximately 20 µg of each peptide was dissolved in 100 µl of NH<sub>4</sub>HCO<sub>3</sub> buffer (100 mM, pH 7.8) containing 10 mM DTT, and incubated for 1 h at 37 °C. 2-fold excess of iodoacetamide (IAA) over the total thiol was added and incubated for 1 h at 37 °C. S-Alkylated peptides were purified by HPLC.

*Enzymatic Digestion and Sequence Determination*—Lyophilized S-alkylated peptides were dissolved in 5 µl of NH<sub>4</sub>HCO<sub>3</sub> buffer (20 mM, pH 7.8) and incubated with endoproteinase Glu-C, trypsin, or chymotrypsin (Roche Applied Science) at a final peptide-to-enzyme ratio of 50:1. The digestions were allowed to proceed for 1 h. Peptide fragments resulting from the digestions were examined by MALDI-MS followed by MALDI-MS/MS analysis using a 4800 Proteomics Analyzer MALDI-TOF/TOF mass spectrometer. Primary peptide sequences were obtained by interpreting the b- and y-ions formed during the MS/MS fragmentation. Assignments of isobaric residues Ile/Leu and Lys/Gln were based on enzymatic digestion patterns, quantitative amino acid analysis, cDNA sequences, and homology to known cyclotides. Amino acid analysis was performed for five cliotides without complete gene sequences including cliotides T1, T5, T10, and T11 and Cter D (supplemental Tables S2-S6). Because of sample limitation, the Ile/Leu assignment of cliotide T6 was based solely on chymotryptic digestion and sequence homology.

*Disulfide Mapping*—Clotide T2 (0.2 mg) was partially reduced in 500 µl of 100 mM citrate buffer, pH 3.0, 20 mM Tris(2-carboxyethyl)phosphine at 37 °C for 35 min. Subsequently, *N*-ethylmaleimide (NEM) powder was added directly to a final concentration of 50 mM and incubated at 37 °C for another 15 min. The reaction was quenched by immediate injection of samples into a Vydac C18 column (250 × 4.6 mm) at a flow rate of 1 ml/min. Intermediate species were separated with a linear gradient of 0.3% min<sup>-1</sup> of 10 – 60% buffer B and analyzed with MALDI-TOF-MS to verify the number of NEM-alkylated cysteines. NEM-alkylated intermediate species were then fully reduced with 20 mM DTT, incubated at 37 °C for 60 min. The reduced peptides were S-alkylated with 40 mM IAA, incubated at 37 °C for 30 min before stopping the reaction by injection into HPLC. S-Alkylated peptides were lyophilized and redissolved in 20 µl of 10% acetonitrile, 20 mM NH<sub>4</sub>HCO<sub>3</sub>, pH 7.8, for trypsin digestion. Fragments obtained were sequenced by MS/MS.

*Antibacterial Assay*—Five bacterial strains from the ATCC were used. Gram-positive bacteria included *Staphylococcus aureus* ATCC12600 and *Enterococcus faecalis* ATCC 47077. Gram-negative bacteria included *E. coli* ATCC 700926, *Pseudomonas aeruginosa* ATCC 39018, and *Klebsiella pneumonia* ATCC 13883. All of the strains were cultured in trypticase soy broth. The antimicrobial activities of cliotides T1 to T4 were examined by using the radial diffusion assay described by Lehrer *et al.* (29). D4R, a synthetic antimicrobial peptide, was used as the positive control (30).

*Cytotoxicity Assay*—HeLa cells were diluted in Dulbecco's modified Eagle's medium to a density of 20,000 cells/ml. 100 µl of the cell suspension was added into each well of a 96-well plate (Nunc) and incubated for 24 h. The cyclotides were serially diluted in PBS, and 5 µl was dispensed into each well. After incubation for 72 h, 20 µl of 5 mg/ml thiazolyl blue tetrazolium bromide (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) solution was added to each well. The plate was incubated for 3.5 h at 37 °C. The

medium was removed, and 150  $\mu$ l of dimethyl sulfoxide was added for cell lysis. The absorbance at 590 nm was measured with a reference filter at 620 nm. The survival index was calculated as a percentage of growth inhibition in comparison with the control wells (added PBS only). Cyclotide cytotoxicities were indicated by IC<sub>50</sub> values (concentration that gives an survival index of 50%).

*Hemolytic Assay*—Blood type A was taken from a healthy volunteer. Red blood cells were washed three times with PBS and resuspended in PBS to give a final 1% suspension. 95  $\mu$ l of the suspension was added to each well of a 96-well plate. Cyclotides were serially diluted in PBS, and 5  $\mu$ l was dispensed into each well. Each cyclotide concentration was tested in triplicate. The plate was incubated at 37 °C for 1 h and centrifuged at 1,000 rpm for 6 min. Aliquots of 60  $\mu$ l were transferred to a new 96-well plate. The absorbance was measured at 415 nm. The level of hemolysis was calculated as the percentage of maximum lysis (1% Triton X-100 control) after adjusting for minimum lysis (PBS control). The cyclotide concentration causing 50% hemolysis (HD<sub>50</sub>) was calculated.

*Cloning of Clotide Genes*—Total RNA was extracted from fresh plant materials (leaves, flowers, and roots) using the PureLink™ RNA purification system (Invitrogen). cDNA libraries were prepared from total RNA using the Smarter™ RACE cDNA amplification kit (Takara). First, 3' RACE PCR was conducted using degenerate primers designed against clotide T2 (5' -GGIGARTTYATHAARTGYGGIGA-3' encoding GEFIKCGE). The PCR products were cloned into pGEM-T Easy vector (Promega) and sent for sequencing. The 5'-end partial gene was obtained from the same procedure using the Universal Primer A Mix primer from the kit and a specific primer against the 3'-untranslated region of the newly found gene. The full transcript was then assembled from the two partial genes. This strategy was also applied to other clotides. In addition, we also used the specific primers against loop 1 of some newly obtained sequences in 3' RACE PCR to target random cyclotides, *i.e.* 5'-TGCGGCGAGAGTTGC-3' encoding CGESC and 5'-ATCCCATGTGGGGAAAGTTGT-3' encoding IPCGESC.

To identify intron locations, genomic DNA from *C. ternatea* flowers was extracted. PCRs on DNA templates were then conducted with specific primers designed against 5'- and 3'-untranslated regions of each clotide. PCR products were purified and subjected to the procedure described above to obtain the DNA gene sequences.

*Phylogenetic Tree Construction*—Phylogenetic tree was constructed using translated precursors of A1bs/leginsulins, cyclotides, and clotides. A1b/leginsulin sequences used included four from *Pisum sativum* (GenBank: CAB82859, CAE00465, CAE00467, and CAE00468), two from *Medicago truncatula* (AC233561, CAE00461, CAE00462, ACJ85909), one from *Glycine max* (BAA04219), and one from *Glycine soja* (CAA09880). Cyclotide sequences included two from *Viola baoshanensis* (DQ851860, ABI48957), two from *Viola odorata* (P58440, AY630565), one from *Gloeospermum blakeanum* (GQ438777), and one from *Hedyotis centranthoides* (FJ211188). Clotide precursors of clotide T1, T2, T3, T7, and T11 were used. The precursor protein sequences were aligned using COBALT multiple alignment (31). Phylogenetic tree was generated using a fast minimum evolution method and visualized by Tree View software (32).

## RESULTS

*Screening of Heat-stable Biologics in C. ternatea Flowers*—Heat-stable CRPs were extracted from *C. ternatea* flowers by incubation with boiling water at 100 °C for 30 min. The aqueous soluble fraction was separated from the denatured protein precipitates and other plant debris by centrifugation. A control experiment without the heat treatment was performed in parallel. The aqueous extracts were subjected to MALDI-TOF-MS to detect the presence of putative biologics within the 1–10-kDa mass range. The mass spectrum revealed a group of peptides with strong  $m/z$  intensity at ~3 kDa (Fig. 1A). Comparison of the reverse phase HPLC profiles before and after the heat treatment revealed that >80% of these peptides survived the boiling water incubation (Fig. 1, B and C). Their disulfide contents were subsequently analyzed by *S*-reduction with DTT and *S*-alkylation with IAA followed by MS analysis. The number of disulfide bonds was deduced by comparing the mass difference before and after reductive *S*-alkylation. Each *S*-alkylated half-cystine residue caused a mass shift of 58 Da. Most of these compounds displayed a mass shift of 348 Da, suggesting the presence of three disulfide bridges, a common structural feature found in many plant biologics.

*Peptide Isolation and Sequencing: Cyclotides as Heat-stable Biologics in C. ternatea*—To isolate sufficient CRP samples for *de novo* sequencing and biological activities studies, 1 kg of fresh plant materials (whole plant) were collected and extracted with boiling water (100 °C). Using reverse phase HPLC, 15 heat-stable CRPs were isolated.

To determine their primary sequences, each HPLC-purified peptide was *S*-reduced, *S*-carbamidomethylated, and digested with trypsin, chymotrypsin, or endoproteinase Glu-C. Generated fragments were sequenced by tandem mass spectrometry. As an example, MS/MS sequencing of cliotide T1 is shown in Fig. 2. Cliotide T1 and its *S*-alkylated form had  $m/z$  values of 3084 and 3432 Da, respectively. Enzymatic digestion of *S*-alkylated cliotide T1 by endoproteinase Glu-C gave a single fragment with a mass increase of 18 Da, suggesting the linearization of a cyclic amide backbone by hydrolysis. Its circular structure was confirmed by tryptic and chymotryptic digestions that yielded overlapping fragments in their entirety. Sequence analysis of cliotide T1 revealed 84% identity to circulin A, a prototypic cyclotide, further suggesting that it is a novel member of cyclotide family. Using this approach, we successfully elucidated the primary structures of 15 peptides, all of which belong to the cyclotide family (Table 1). Twelve sequences (cliotide T1–T12) are novel, and three sequences (Cter A, Cter B, and Cter D) have been identified previously (27).

*cDNA Cloning Revealed Chimeric Structures of Cliotide Encoding Genes*—To determine their encoding genes and to provide insight about their biosynthesis, we designed degenerate primers for 3' RACE PCR based on MS/MS-determined sequences. A total of 11 partial clones were obtained. Afterward, a series of specific primers based on the newly obtained genetic sequences were used for 5' RACE PCR to complete the full-length genes. Nine clones were named *ctc1*, *ctc2*, *ctc3*, *ctc4*, *ctc5*, *ctc7*, *ctc8*, *ctc9*, and *ctc12* according to their cliotide inserts. Two clones encoding Cter A and Cter B were named *ctc13* and *ctc14*, respectively.

The translated sequences of cliotide clones are shown in Fig. 3. They feature an entirely different arrangement from previously published cyclotide precursors. In the Rubiaceae and Violaceae families, cyclotide genes such as *oak1* and *voc1* (33) contain an endoplasmic reticulum (ER) signal sequence, an N-terminal pro-domain (NTPP), an N-terminal repeat region (NTR), a cyclotide domain, and a short C-terminal tail. Sequence analysis revealed that cliotide precursors contained no NTPP or NTRs. Instead, the ER signal sequence was followed directly by the cyclotide domain. In addition, a novel cysteine-rich domain was found at the C-terminal end of the precursor proteins separated from the cyclotide domain by a short putative linker region. A BLAST search showed that this novel domain was highly homologous to the A1a domain (Albumin-1 chain a) of A1 genes found in many legume species (11, 34). Surprisingly, the gene architecture of cliotides displayed striking similarity to those of A1 genes (Fig. 4), which contain a signal peptide followed by an A1b domain, a linker peptide, and an A1a domain (35). The A1b domain in pea seeds encodes for a 37-residue insecticidal peptide containing three disulfide bonds (11). The cliotide genes shared exactly the same arrangement with A1 genes except that the A1b domain was replaced by the cyclotide domain. The homology of cliotide genes to both A1 genes (Fabaceae) and cyclotide genes (Rubiaceae, Violaceae) makes them naturally occurring chimeric genes.

*Cliotide Genes Contain a Single Intron at the ER Region*— Known families of CRP genes are generally characterized by the presence of a single intron located within the ER domain (36–39). To compare the genetic structure of cliotides at both DNA and mRNA levels, five cliotide precursor genes including cliotides T2, T3, T7, T8, and T10 were cloned from the leaf DNA using cDNA-derived sequences as primers. The DNA clones revealed a single intron located in the signal peptide region of the cliotide genes, which is similar to A1 genes from the Fabaceae and cyclotide genes from the Rubiaceae family (Fig. 4). Thus, the single-intron architecture of cliotide DNA genes is consistent with genes from CRP families, which include not only cyclotides and A1s but also plant defensins and thionins (36, 40). It is interesting to note that the cliotide T7 gene contains no intron, which is similar to those characterized from the Violaceae family. This suggests an intron loss event in some of the cliotide genes, possibly mediated through gene conversion with a reverse transcriptase product of a spliced mRNA.

*Disulfide Mapping of Cliotide T2: New Precursor, Same Connectivity*—The disulfide connectivity of novel cliotides was sought to determine whether they possessed a different arrangement from the cystine knot motif found in cyclotides. Cliotide T2, present in high abundance in the floral tissue, was selected as a representative. The connectivity of cliotide T2 was determined by a differential *S*-tagging strategy through a sequential *S*-reduction and *S*-alkylation (41). The S–S bonds of native cliotide T2 were first partially ruptured with Tris(2-carboxyethyl)phosphine to generate a series of isoforms with one or two disulfide bonds being reduced. The released thiols were immediately *S*-alkylated with an excess of NEM. The whole process was performed under acidic conditions at pH 3.5 to avoid scrambling of disulfide linkages. The partially *S*-reduced and *S*-alkylated peptides were purified by reverse phase HPLC.

Four chromatography-separated peaks were collected (Fig. 5A) and analyzed by MALDI-TOF-MS to determine the number of NEM-tagged cysteines. Each NEM-tagged

cysteine caused a mass increase of 126 Da. The mass gain after *S*-alkylation was then used to deduce the number of reduced disulfide bonds. Peak 1 had a  $m/z$  of 3512 Da, which was 252 Da larger than the original peptide (3260 Da). Based on the mass gain, we could assign the intermediates eluted in peak 1 as a 2 disulfide bond intermediate, having two NEM-labeled cysteine residues and two intact disulfide bonds, peak 2 as the native peptide with an intact cystine core, peak 3 as a 1SS species with four NEM-tagged cysteine residues and one remaining disulfide bond ( $m/z$  of 3764 Da), and peak 4 as a product with all six cysteines tagged by NEM. It is noteworthy that the 2SS species was more hydrophilic than the native peptide, whereas the 1SS species and fully NEM-tagged peptides became more hydrophobic upon alkylation.

To obtain the connectivity, the remaining disulfide bonds of 1SS and 2SS species were again *S*-reduced with DTT and then *S*-tagged with IAA, a second alkylation reagent. *S*-Alkylated peptides were digested with trypsin, and the resulting fragments were analyzed by MS/MS. Tryptic digestion generated three proteolytic products with two fragments containing no cysteine residue (K and NGEFLK) and thus were universal for both IAA derivatives of 1SS and 2SS species (Fig. 5B). The third fragment was unique for each of the isoforms and contained all six cysteine residues. MS/MS analysis of this fragment provided unambiguous assignment of disulfide bridges based on the differential *S*-tagging of NEM and IAA on the cysteine residues.

The connectivity of the 2SS species was established as Cys II–V with two NEM-tagged groups on Cys II and Cys V and four IAA-tagged groups on the remaining cysteine residues (Fig. 5C). Similarly, the connectivity of the 1SS species was identified as Cys III–VI with four NEM-tagged groups on Cys I, II, IV, and V and two IAA-tagged groups on Cys III and Cys VI (Fig. 5D). The third disulfide bond of Cys I–IV was obtained by deduction. Taken together, these results provided evidence for the knotted cysteine arrangement of cliotide T2, a disulfide connectivity common to both cyclotides and A1bs.

*Antimicrobial Activity*—To assess their antimicrobial activity, four cliotides (cliotides T1–T4) were tested against a panel of five different bacterial strains using a radial diffusion assay. Their selection was primarily based on abundance and varied structural characteristics, consisting of two Möbius and two bracelet cyclotides. Table 2 summarizes their antimicrobial activities. D4R, a synthetic antimicrobial peptide, was used as the positive control (30). The bracelet subgroup including cliotide T1 and cliotide T4 showed strong antimicrobial activity against three species of Gram-negative bacteria with the MIC values in the low micromolar range. Among these strains, *E. coli* was the most susceptible to cliotide bactericidal activities. Cliotide T1 and T4, however, were relatively ineffective against both tested Gram-positive bacterial strains *S. aureus* and *E. faecalis*. The Möbius subgroup comprising cliotides T2 and T3 was inactive against all five tested bacterial strains at concentrations up to 100  $\mu\text{M}$ .

*Hemolytic Activity*—The hemolytic effects of cliotides T1–T4 were conducted using human type A erythrocytes. Melittin, a known peptide from bee venom with strong hemolytic activity, was used as positive control. Cliotide T2 was relatively nonhemolytic and only lysed ~18% of the red blood cells at 40  $\mu\text{M}$ . The  $\text{HD}_{50}$  of the other three cliotides

ranged from 7.1 to 13.1  $\mu\text{M}$  (Table 2), which were significantly less potent than melittin (3.2  $\mu\text{M}$ ).

*Cytotoxicity against HeLa Cells*—The cytotoxic properties of cliotides T1–T4 were assessed by 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide assay, which monitored mitochondrial activity in living cells and hence indicated cell viability. HeLa cells were selected as a model for the assay. Five concentrations were tested for each cliotide, and  $\text{IC}_{50}$  values were determined from the survival curves. All of the four tested cliotides displayed sharp dose-response curves with  $\text{IC}_{50}$  values ranging from 0.6 to 8.0  $\mu\text{M}$  (Table 2). Cliotide T1 and T4 were the most potent with  $\text{IC}_{50}$  values of 0.6  $\mu\text{M}$ . Cliotide T3 had moderate cytotoxic effect, with  $\text{IC}_{50}$  at 2.0  $\mu\text{M}$ , and cliotide T2 was the least cytotoxic, with  $\text{IC}_{50}$  at 8.0  $\mu\text{M}$ .

*Tissue-specific Distribution*—To determine whether the novel cliotides are expressed universally or in a tissue-specific manner, MALDI-TOF-MS was utilized to profile their expression in different plant parts. A total of eight different tissues were collected including leaves, stems, roots, flowers, pods, seeds, nodules, and shoots. They were extracted separately using boiling water and analyzed by mass spectrometry. The mass spectra revealed that each tissue had a unique profile distinguished by the number and amount of expressed cliotides (Fig. 6). Each tissue produced at least 10 different cliotides, but only a few specific cliotides dominated the expression profiles. For instance, cliotide T2 and cliotide T3 were highly expressed in the flowers and pods, accounting for more than 70% of total cyclotides in these tissues (supplemental Table S1). Similarly, cliotides T8, T9, and T10 were the major constituents in seeds. However, not all cliotides were tissue-specific. Certain cliotides, such as cliotides T1 and T4, were found to be expressed in all examined tissues, suggesting their housekeeping role in plant defense.

## DISCUSSION

In this study, we report the discovery of cyclotides as heat-stable biologics in *C. ternatea*. Fifteen cyclotides have been characterized from plant extracts, of which 12 sequences are novel. Three cliotide sequences are identical to those isolated from the seeds of *C. ternatea* previously reported by Poth *et al.* (27), who identified a suite of 12 novel cyclotides. The lack of sequence overlaps could be attributed to several factors including methods of extraction, tissues examined, and geographical variations. Here, we show that cyclotides are not only present in seeds but in every plant part, including flowers, leaves, roots, nodules, pods, stems, shoots, and seeds. The presence of cyclotides in nodules is noteworthy because of their high abundance and possible roles in controlling the symbiosis of their host plant and nitrogen-fixing bacteria.

The identification of cliotides as cyclotides and not A1bs is unequivocal. A1bs previously discovered in the legumes are CRPs of similar sizes as cyclotides but contain a linear structure. Cyclotides have a cyclic structure with a head-to-tail cyclized peptide backbone. Using a combination of enzymatic fragmentations and MS sequencing, all 15 cliotides are shown to be cyclic proteins. Disulfide mapping of cliotide T2 confirms that

it retains a cystine knot motif similar to cyclotides. In addition, sequence and cysteine-spacing comparisons also firmly place cliotides in the cyclotide family (Table 1).

Disulfide mapping of cliotide T2 provides useful information about its unfolding mechanism based on the intermediate species isolated. The abundance of 2SS species formed by the breaking of the Cys II–V bond suggests that this cystine linkage is most susceptible to reducing reagents. 1SS species was subsequently generated by reducing of the Cys I–IV bond with the remainder of the Cys III–VI bond as the most stable disulfide linkage. This is consistent with our understanding of cyclotide cystine knot structure and with the idea that the penetrating disulfide bond of cystine knot motif (Cys III–VI) is shielded and buried deep inside the cystine core to account for its resistance against reducing reagents (42).

Functionally, cyclotides are characterized by their membrane active properties for a wide variety of biological actions (23, 43). Among these, antimicrobial, hemolytic, and cytotoxic activities have been studied to dissect the membrane binding ability of cyclotides on three different membrane types of bacteria, erythrocytes, and cancer cells (23). These bioactivities were thus selected for functional evaluations of novel cliotides. Our results show that cliotides retain the characteristic functions of cyclotides with similar potency. Cliotides T1 and T4 are active against *E. coli* with MIC value of  $\sim 1 \mu\text{M}$  as compared with  $1.5 \mu\text{M}$  for cyclopsychotride (21). They are cytotoxic to HeLa cells with an  $\text{IC}_{50}$  within the  $0.6 - 8.0 \mu\text{M}$  range, which is comparable with the cytotoxic effects of several cyclotides such as varv A ( $\text{IC}_{50} = 2.7 - 6.3 \mu\text{M}$ ), varv F ( $\text{IC}_{50} = 2.6 - 7.4 \mu\text{M}$ ), and cycloviolacin O<sub>2</sub> ( $\text{IC}_{50} = 0.1 - 0.3 \mu\text{M}$ ) (44). They also have similar hemolytic activities ( $\text{HD}_{50} = 7.9 - 13.4 \mu\text{M}$ ) as compared with kalata B1 ( $11.7 \mu\text{M}$ ) (45).

Extraction of cyclotides in our study was achieved by using boiling water, demonstrating their heat stability. Although numerous extraction techniques, often involving organic solvents, have been used to isolate active principles, decoction preparation by soaking the herbs in boiling water is the most relevant method in traditional medicines. Heating of the plant extract at  $100 \text{ }^{\circ}\text{C}$  for 30 min showed nearly identical cliotide profiles before and after the treatments. Our results also demonstrate the thermal stability of cyclotides not only as pure compounds dissolved in water (26) but also in the complex mixture of the plant extract, which contains many herbal-derived, small molecules. Thus, our findings provide experimental support that cyclotides can survive heat treatment during the decoction preparation process and may be one of the active principles attributed to the medicinal values of *C. ternatea*.

Similar to many CRPs, the cliotide genes contain a single intron in the ER domain (40). Intriguingly, our results show unexpectedly that cliotide precursors display an entirely new arrangement differing from the known cyclotides from the Rubiaceae and Violaceae families. Cliotide genes contain a hybrid coding region consisting of peptide domains from two unrelated genes. In most legume species, the A1 genes encode three major domains: ER, A1b, and A1a (11). Cliotide genes share essentially the same arrangement except for the exchange of the A1b for the cyclotide domain. This raises the question of whether A1b peptides are still expressed in *C. ternatea* or whether they have been totally replaced by cyclotides. Our attempts to identify A1b peptides in the *C.*

*ternatea* extract have failed to identify similar peaks in the known mass range of A1bs. It is possible that they are present in low abundance that escapes our detection.

Characterization of cliotide genes provides new understanding about the biosynthetic mechanism of cyclotides in the Fabaceae family. Cliotide precursors, unlike their cyclotide counterparts, are devoid of the NTPP and NTRs typically separating the ER signal from the mature cyclotide domain. This suggests that cliotides are N-terminally processed by signal peptidases (SPase I) (Fig. 7). The absence of the NTPP and NTRs also implies that these protein domains may not be essential for the biosynthesis of cyclotides in the Fabaceae family. They instead may have other endogenous functions in plants. At the C terminus, the cyclotide domains are probably cleaved by asparaginyl endopeptidase (AEP) as in Rubiaceae and Violaceae species indicated by the presence of the highly conserved Asn at the processing site. AEP is a cysteine proteinase and forms with the C-terminal Asn a reactive thioester bond that leads a head-to-tail ligation to afford the cyclized structure of a cyclotide. Interestingly, the conserved Leu residue located at two residues downstream of the C-terminal Asn processing site is replaced by Val or Ile in cliotide genes. This reflects a difference in specificity at the P2' position of AEP in *C. ternatea*.

In addition to providing new understanding about the biosynthetic mechanism of cyclotides, the cliotide genes also support the natural occurrence of cliotide T1 and T12. These two peptides share identical primary sequences, except for the single amino acid difference at the ligation site, the Asn and Asp variation in loop 6. Five similar cyclotide pairs with such a variation have been identified in the previous study by Poth *et al.* (27) from the extract of *C. ternatea* seeds. The origin of the Asn- and Asp-cyclotide variants has been a topic of debate. It is uncertain whether the Asp variants are gene-encoded or formed as the result of the conversion of the Asn to Asp because of deamidation. The isolation of both cliotide T1 and T12 encoding genes (*ctc1* and *ctc12*) in this study provides firm evidence that the Asp variants in *C. ternatea* are gene-encoded. However, it remains to be determined whether the same molecular mechanism is responsible for the Asn and Asp variation observed in other species such as *Oldenlandia affinis* and *V. odorata* (46, 47).

The genetic replacement of A1bs by cyclotides raises another question about the functional overlap between these two peptide families. The peptides are similar in size, both possessing a cystine knot scaffold and exhibiting insecticidal activity (48). A1bs are known to exert physiological functions in plants such as growth, differentiation, and cell proliferation by binding to a 43-kDa receptor located at the plant cell wall (34, 49, 50). In contrast, little is known about the physiological functions, intracellular location, or receptor binding capabilities of cyclotides in plants. The substitution of A1bs for cyclotides in *C. ternatea* suggests that they may share certain roles and provides hints about the physiological functions of cyclotides.

How does the chimeric structure of cliotide genes occur? The cyclotide-encoding domains in *C. ternatea* may arise by vertical heredity, horizontal gene transfer, or convergent evolution. In vertical heredity, cyclotide genes would have to exist in the ancestral legume lineages and are inherited in *C. ternatea* by means of sexual

reproduction. This hypothesis, however, cannot explain why only A1bs are detected in all the Fabaceae hitherto examined (11).

In a horizontal gene transfer event, *C. ternatea* acquires the cyclotide domains horizontally from other species. Such a scenario is supported by three lines of evidence. First, there is an unexpected clustering of *C. ternatea* with several Rubiaceae and Violaceae species instead of Fabaceae species using phylogenetic tree constructed with the cyclotide, A1, and cliotide genes (Fig. 8). It should be noted that three different gene types were used for constructing the phylogenetic tree because of the chimeric nature of the cliotide precursors. Second, a DNA database search of 22 Fabaceae species, including two plants with full genome sequences *G. max* and *M. truncatula*, identified orthologs of A1 genes, but none of them contains the cyclotide domain. Finally, our preliminary screening of two related *Clitoria* species, *Clitoria fairchildiana* and *Clitoria laurifolia*, did not detect the presence of cyclotides, but peaks at 3.7 kDa resembling those of A1b peptides were observed (data not shown). It is thus possible that *C. ternatea* may have received the cyclotide genes from donor species of the Rubiaceae or Violaceae families. If this hypothesis is shown to be valid, it will provide important evidence of horizontal gene transfers between plant nuclear genomes. Horizontal gene transfers are known to occur mostly in plant mitochondria but rarely in nuclear and chloroplast DNA (51). Intriguingly, the displacement of A1bs for cyclotide domains in cliotide genes appears to interrupt their normal functions, an event that often leads to negative selection (52, 53). However, in this instance, the cyclotide domains were not only transferred but also correctly processed, folded, and fully functional. Furthermore, they also display differential distribution in different plant parts, which suggests that they have been positively selected and fine-tuned for tissue-specific functions.

The third possibility is convergent evolution from A1bs to cyclotides. Both are similar in size, share a cystine knot structure, and overlap in certain functions. However, for this to happen, it would require massive mutations and substitutions for complete conversion from A1bs to cyclotides. An interesting example of convergent evolution has recently been reported for the platypus (54). The toxin genes identified from this animal show striking homology to those found in other venomous species such as snakes, starfish, and sea anemones. Genetic analysis reveals that these venoms have evolved independently from different origins but ended up developing similar families of molecules (55).

Genetic characterization of cyclotides in the Fabaceae may open up new avenues of research allowing genetic manipulation and production of cyclotide-based therapeutics in agriculturally important Fabaceae plants. There has been an increasing interest in using cyclotides as molecular scaffolds for grafting of bioactive epitopes. The current bottlenecks for chemical synthesis of these engineering peptides are high production cost and uncertainty in obtaining the correct folding. The use of plants as bioreactors can provide a cost-efficient solution and rapid production of cyclotides for medical and industrial applications.

In conclusion, our data provide the first genetic characterization of cyclotides in the Fabaceae. We show that their precursors are (Violaceae) and the other half from Albumin-1 (Fabaceae), with the cyclotide domain replacing the A1b domain in the precursor. Their

chimeric structures suggest that the occurrence of cyclotides in the Fabaceae could be the result of a horizontal gene transfer between plant nuclear genomes or a convergent evolution from A1bs to cyclotides. Collectively, our findings provide new understanding about the biosynthetic mechanism of cyclotides in the Fabaceae and their evolution in plants.

## REFERENCES

1. Mukherjee, P. K., Kumar, V., Kumar, N. S., and Heinrich, M. (2008) *J. Ethnopharmacol.* **120**, 291–301
2. Daisy, P., Santosh, K., and Rajathi, M. (2009) *Afr. J. Microbiol. Res.* **3**, 287–291
3. Rai, K. S., Murthy, K. D., Karanth, K. S., Nalini, K., Rao, M. S., and Srinivasan, K. K. (2002) *Fitoterapia* **73**, 685–689
4. Jain, N. N., Ohal, C. C., Shroff, S. K., Bhutada, R. H., Somani, R. S., Kasture, V. S., and Kasture, S. B. (2003) *Pharmacol. Biochem. Behav.* **75**, 529–536
5. Kulkarni, C., Pattanshetty, J. R., and Amruthraj, G. (1988) *Indian J. Exp. Biol.* **26**, 957–960
6. Rai, K. S., Murthy, K. D., Karanth, K. S., and Rao, M. S. (2001) *Indian J. Physiol. Pharmacol.* **45**, 305–313
7. Lay, F. T., and Anderson, M. A. (2005) *Curr. Protein Pept. Sci.* **6**, 85–101
8. Heitz, A., Avrutina, O., Le-Nguyen, D., Diederichsen, U., Hernandez, J. F., Gracy, J., Kolmar, H., and Chiche, L. (2008) *BMC Struct. Biol.* **8**, 54–71
9. Broekaert, W. F., Cammue, B. P., DeBolle, M. F., Thevissen, K., DeSamblanx, G. W., and Osborn, R. W. (1997) *Crit. Rev. Plant Sci.* **16**, 297–323
10. Broekaert, W. F., Mariën, W., Terras, F. R., De Bolle, M. F., Proost, P., Van Damme, J., Dillen, L., Claeys, M., Rees, S. B., Vanderleyden, J., and Cammue, B. P. (1992) *Biochemistry* **31**, 4308–4314
11. Louis, S., Delobel, B., Gressent, F., Dupont, G., Diol, O., Rahioui, I., Charles, H., and Rahbe, Y. (2007) *Phytochemistry* **68**, 521–535
12. Kolmar, H. (2009) *Curr. Opin. Pharmacol.* **9**, 608–614
13. Louis, S., Delobel, B., Gressent, F., Rahioui, I., Quillien, L., Vallier, A., and Rahbé Y. (2004) *Plant Sci.* **167**, 705–714
14. Da Silva, P., Strzepa, A., Jouvensal, L., Rahioui, I., Gressent, F., and Delmas, A. F. (2009) *Biopolymers* **92**, 436–444
15. Dun, X. P., Wang, J. H., Chen, L., Lu, J., Li, F. F., Zhao, Y. Y., Cederlund, E., Bryzgalova, G., Efendic, S., Jörnvall, H., Chen, Z. W., and Bergman, T. (2007) *FEBS J.* **274**, 751–759
16. Le Gall, M., Quillien, L., Guéguen, J., Rogniaux, H., and Sève, B. (2005) *J. Nutr.* **135**, 1215–1222
17. Hanada, K., Nishiuchi, Y., and Hirano, H. (2003) *Eur. J. Biochem.* **270**, 2583–2592

18. Rahioui, I., Laugier, C., Balmand, S., Da Silva, P., Rahbe, Y., and Gressent, F. (2007) *Biochimie* **89**, 1539–1543
19. Dun, X. P., Li, F. F., Wang, J. H., and Chen, Z. W. (2008) *Peptides* **29**, 891–897
20. Craik, D. J., Daly, N. L., Bond, T., and Waine, C. (1999) *J. Mol. Biol.* **294**, 1327–1336
21. Tam, J. P., Lu, Y. A., Yang, J. L., and Chiu, K. W. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8913–8918
22. Hernandez, J. F., Gagnon, J., Chiche, L., Nguyen, T. M., Andrieu, J. P., Heitz, A., Trinh Hong, T., Pham, T. T., and Le Nguyen, D. (2000) *Biochemistry* **39**, 5722–5730
23. Daly, N. L., Rosengren, K. J., and Craik, D. J. (2009) *Adv. Drug Deliv. Rev.* **61**, 918–930
24. Crooks, G.E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) *Genome Res.* **14**, 1188–1190
25. Sze, S. K., Wang, W., Meng, W., Yuan, R., Guo, T., Zhu, Y., and Tam, J. P. (2009) *Anal. Chem.* **81**, 1079–1088
26. Colgrave, M. L., and Craik, D. J. (2004) *Biochemistry* **43**, 5965–5975
27. Poth, A. G., Colgrave, M. L., Philip, R., Kerenga, B., Daly, N. L., Anderson, M. A., and Craik, D. J. (2011) *ACS Chem. Biol.* **6**, 345–355
28. Devi, B. P., Boominathan, R., and Mandal, S. C. (2003) *Fitoterapia* **74**, 345–349
29. Lehrer, R. I., Rosenman, M., Harwig, S. S., Jackson, R., and Eisenhauer, P. (1991) *J. Immunol. Methods* **137**, 167–173
30. Tam, J. P., Lu, Y. A., and Yang, J. L. (2002) *Eur. J. Biochem.* **269**, 923–932
31. Papadopoulos, J. S., and Agarwala, R. (2007) *Bioinformatics* **23**, 1073–1079
32. Page, R. D. (1996) *Comput. Appl. Biosci.* **12**, 357–358
33. Dutton, J. L., Renda, R. F., Waine, C., Clark, R. J., Daly, N. L., Jennings, C. V., Anderson, M. A., and Craik, D. J. (2004) *J. Biol. Chem.* **279**, 46858–46867
34. Watanabe, Y., Barbashov, S. F., Komatsu, S., Hemmings, A. M., Miyagi, M., Tsunasawa, S., and Hirano, H. (1994) *Eur. J. Biochem.* **224**, 167–172
35. Tan, J. Z., Lou, C. F., Zhang, G. Y., Sun, B. Y., and Ping, Y. J. (2003) *Yi Chuan* **25**, 168–172
36. Meyer, B., Houlne G., Pozueta-Romero, J., Schantz, M. L., and Schantz, R. (1996) *Plant Physiol.* **112**, 615–622

37. Chen, J. J., Chen, G. H., Hsu, H. C., Li, S. S., and Chen, C. S. (2004) *J. Agric. Food Chem.* **52**, 2256–2261
38. Silverstein, K. A., Moskal, W. A., Jr., Wu, H. C., Underwood, B. A., Graham, M. A., Town, C. D., and VandenBosch, K. A. (2007) *Plant J.* **51**, 262–280
39. Graham, M. A., Silverstein, K. A., and VandenBosch, K. A. (2008) *Crop Sci.* **48**, S3–S11
40. Carvalho Ade, O., and Gomes, V. M. (2009) *Peptides* **30**, 1007–1020
41. Gray, W. R. (1993) *Protein Sci.* **2**, 1732–1748
42. Göransson, U., and Craik, D. J. (2003) *J. Biol. Chem.* **278**, 48188–48196
43. Huang, Y. H., Colgrave, M. L., Daly, N. L., Keleshian, A., Martinac, B., and Craik, D. J. (2009) *J. Biol. Chem.* **284**, 20699–20707
44. Lindholm, P., Göransson, U., Johansson, S., Claeson, P., Gullbo, J., Larsson, R., Bohlin, L., and Backlund, A. (2002) *Mol. Cancer Ther.* **1**, 365–369
45. Wang, C. K., Colgrave, M. L., Gustafson, K. R., Ireland, D. C., Goransson, U., and Craik, D. J. (2008) *J. Nat. Prod.* **71**, 47–52
46. Plan, M. R., Göransson, U., Clark, R. J., Daly, N. L., Colgrave, M. L., and Craik, D. J. (2007) *ChemBiochem* **8**, 1001–1011
47. Ireland, D.C., Colgrave, M. L., and Craik, D. J. (2006) *Biochem. J.* **400**,1–12
48. Gruber, C. W., Cemazar, M., Anderson, M. A., and Craik, D. J. (2007) *Toxicon* **49**, 561–575
49. Yamazaki, T., Takaoka, M., Katoh, E., Hanada, K., Sakita, M., Sakata, K., Nishiuchi, Y., and Hirano, H. (2003) *Eur. J. Biochem.* **270**, 1269–1276
50. Nishizawa, N. K., Mori, S., Watanabe, Y., and Hirano, H. (1994) *Plant Cell Physiol.* **35**, 1079–1085
51. Bock, R. (2010) *Trends Plant Sci.* **15**, 11–22
52. Graham, T., and Boissinot, S. (2006) *J. Biomed. Biotechnol.* **2006**, 75327
53. Cordaux, R., and Batzer, M. A. (2009) *Nat. Rev. Genet.* **10**, 691–703
54. Whittington, C. M., Papenfuss, A. T., Bansal, P., Torres, A. M., Wong, E. S., Deakin, J. E., Graves, T., Alsop, A., Schatzkamer, K., Kremitzki, C., Ponting, C. P., Temple-Smith, P., Warren, W. C., Kuchel, P. W., and Belov, K. (2008) *Genome Res.* **18**, 986–994

55. Whittington, C. M., Papenfuss, A. T., Locke, D. P., Mardis, E. R., Wilson, R. K., Abubucker, S., Mitreva, M., Wong, E. S., Hsu, A. L., Kuchel, P. W., Belov, K., and Warren, W. C. (2010) *Genome Biol.* **11**, R95
56. Trabi, M., Mylne, J. S., Sando, L., and Craik, D. J. (2009) *Org. Biomol. Chem.* **7**, 2378–2388
57. Higgins, T. J., Chandler, P. M., Randall, P. J., Spencer, D., Beach, L. R., Blagrove, R. J., Kortt, A. A., and Inglis, A. S. (1986) *J. Biol. Chem.* **261**, 11124–11130
58. Schneider, T. D., and Stephens, R. M. (1990) *Nucleic Acids Res.* **18**, 6097–6100

## List of Tables

**TABLE 1** Novel cyclotides from *Clitoria ternatea*

**TABLE 2** Antibacteria, hemolytic, and cytotoxic activities of selected cliotides

## List of Figures

- FIGURE 1. **Thermal stability of biologics in the *C. ternatea* flower.** *A*, MS spectrum of the heat-stable aqueous extract of the *C. ternatea* flower. Peaks labeled  $K^+$  indicate the *potassium* adduct (+38 Da). *B* and *C*, HPLC profile of the flower extract before (*B*) and after (*C*) heat treatment at 100 °C for 30 min. The main peaks were labeled according to their cliotide contents.
- FIGURE 2. **MS/MS sequencing of cliotide T1.** *A*, enzymatic digestion of *S*-alkylated cliotide T1 by trypsin yielded two fragments with *m/z* values of 812 and 2657 Da, and digestion by chymotrypsin yielded 1395- and 2074-Da fragments. *B–E*, MS/MS spectra of digested fragments. Analysis of these spectra gave overlapping sequences indicative of a circular amide backbone of cliotide T1.
- FIGURE 3. **Deduced amino acid sequences of cliotide precursors.** The sequences display a novel gene arrangement as compared with previously known cyclotide precursors from the Rubiaceae and Violaceae families. The precursors contain an ER signal region directly followed by a mature cyclotide domain, a short linker region, an A1a domain, and a short C-terminal tail. This gene organization resembles that of pea Albumin-1 precursor except that the A1b domain is replaced by the cyclotide domain. The *inverted triangles* indicate the locations of introns. The *pal* gene can be found in GenBank™ with accession number CAB82859.
- FIGURE 4. **Schematic comparison of the genetic arrangements of cyclotide (Rubiaceae and Violaceae), cliotide, and A1 (Fabaceae) genes (56, 57).** Signal peptides are highlighted in *black*, N-terminal pro-domains are in *gray*, N-terminal repeats are in *purple*, A1b and cyclotide domains are in *green*, and A1a domains are in *orange*. The *inverted triangles* indicate the introns located in the signal peptide regions.
- FIGURE 5. **Disulfide mapping of cliotide T2.** *A*, reverse phase HPLC profile of partially *S*-reduced and *S*-NEM-labeled cliotide T2. *P1–P4* indicate peaks 1–4, and the labels in *parentheses* indicate the corresponding peptide species. Peaks 1 and 3 are 2SS and 1SS intermediate species with two and one intact disulfide bonds, respectively. Peak 2 (N) is the native peptide; Peak 4 (RA) is the fully *S*-NEM-labeled peptide. *B*, schematic presentation of disulfide mapping of cliotide T2. 1SS and 2SS species were isolated, reduced with DTT, and subsequently *S*-alkylated with IAA. *S*-Alkylated peptides were treated with trypsin and fragments were sequenced by

MS/MS. *C* and *D*, MALDI-TOF-MS/MS spectra of trypsin-digested peptide fragments. Their interpreted sequences are shown at the *top* of the figures.

FIGURE 6. **Tissue-specific distribution of cliotides in *C. ternatea*.** Eight different plant parts were collected including flower, stem, shoot, root, leaf, seed, pod, and nodule. The extracts were profiled by mass spectrometry. The peaks are labeled according to the masses of the corresponding cliotides. Peaks labeled  $K^+$  indicate the *potassium* adducts (+38 Da).

FIGURE 7. **Putative enzymes involved in the cyclotide bioprocessing.** In *C. ternates* (Fabaceae), the signal peptidase I (SPase I) responsible for ER excision is also the N-terminal processing enzyme of cliotides. In the Rubiaceae and Violaceae families, unknown enzyme(s) are required for the N-terminal cleavage of the cyclotide domain. The C termini of cyclotides in these three families are processed by AEPs because of the highly conserved Asn/Asp residue at the processing site. The sequence logos were built using WebLogo server (24, 58).

FIGURE 8. **Phylogenetic tree showing the relationship among the cyclotide, cliotide, and A1 genes from different Rubiaceae, Violaceae, and Fabaceae species.** The precursor protein sequences were aligned using COBALT multiple alignment. The phylogenetic tree shows that the cliotide genes cluster with cyclotide genes from the Rubiaceae and Violaceae but not the A1 genes from the Fabaceae.

Peptide	Sequence						MW <sup>a</sup>	AA <sup>b</sup>	Subfamily <sup>c</sup>	
Loop	6	1	2	3	4	5	6			
cT1	GI-P-C	GESC	VFIPC	-ITGAIG	CSC	KSKV	CYRN	3083	30	B
cT2	GEFLK	C	GESC	VQGE	C	-YT--PG	CSCDWPI	3259	30	M
cT3	GL-PT	C	GETC	TLGTC	-YV--PD	CSC	SWPICMKN	3057	29	M
cT4	GI-P-C	GESC	VFIPC	-ITAAIG	CSC	KSKV	CYRN	3097	30	B
cT5	GI-P-C	GESC	VFIPC	-ISTVIG	CSC	KNKV	CYRN	3168	30	B
cT6	SI-P-C	GESC	VYIPC	-ITTIVG	CSC	KNSV	CYSN	3118	30	B
cT7	GI-P-C	GESC	VFIPC	TVTALLG	CSC	KDKV	CYKN	3226	31	B
cT8	GI-P-C	GESC	VFIPC	-ISSVVG	CSC	KSKV	CYNN	3071	30	B
cT9	GI-P-C	GESC	VFIPC	-ITTVVG	CSC	KNKV	CYNN	3126	30	B
cT10	GI-P-C	GESC	VYIPC	TVTALLG	CSC	KDKV	CYKN	3242	31	B
cT11	GI-P-C	GESC	VFIPC	TITALLG	CSC	KDKV	CYKN	3240	31	B
cT12	GI-P-C	GESC	VFIPC	-ITGAIG	CSC	KSKV	CYRD	3084	30	B
Cter A <sup>d</sup>	GVIP-C	GESC	VFIPC	-ISTVIG	CSC	KNKV	CYRN	3267	31	B
Cter B <sup>d</sup>	GV-P-C	AESC	VWIPC	TVTALLG	CSC	KDKV	CYLN	3250	31	B
Cter D <sup>d</sup>	GI-P-C	AESC	VWIPC	TVTALLG	CSC	KDKV	CYLN	3264	31	B
A1b <sup>e</sup>	ASC	NGV	CSPF	EMPP	CGT	SAC	RRCIPVGLVIGY	3738	37	A1b

<sup>a</sup> Molecular weights are reported as monoisotopic mass.

<sup>b</sup> Number of amino acids (AA).

<sup>c</sup> Subfamily: Bracelet (B), Möbius (M). Cysteine residues are highlighted in yellow, negatively charged residues are in red, and positively charged residues are in blue.

<sup>d</sup> Cter A, Cter B, and Cter D have been previously reported by Poth *et al.* (27).

<sup>e</sup> A1b is shown for comparison (57).

**TABLE 1**

Peptide	MIC <sup>a</sup>			HD <sub>50</sub> <sup>b</sup>	IC <sub>50</sub> <sup>c</sup>
	<i>E. coli</i>	<i>K. pneumonia</i>	<i>P. aeruginosa</i>		
	$\mu\text{M}$				
cT1	1.1	2.7	4.7	7.1	0.6
cT2	>100	>100	>100	>100	8.0
cT3	>100	>100	>100	13.1	2.0
cT4	1.0	5.5	7.5	8.4	0.6
D4R <sup>d</sup>	1.3	1.9	1.5		
Melittin <sup>e</sup>				3.2	

<sup>a</sup> MIC, minimal inhibitory concentration.

<sup>b</sup> HD<sub>50</sub> refers to the peptide concentration that causes 50% lysis of red blood cells.

<sup>c</sup> IC<sub>50</sub> refers to the peptide concentration that causes 50% death of HeLa cells.

<sup>d</sup> D4R is a synthetic antibacterial peptide that was used as the positive control for antibacterial assay.

<sup>e</sup> Melittin was used as the positive control for hemolytic assay.

**TABLE 2**

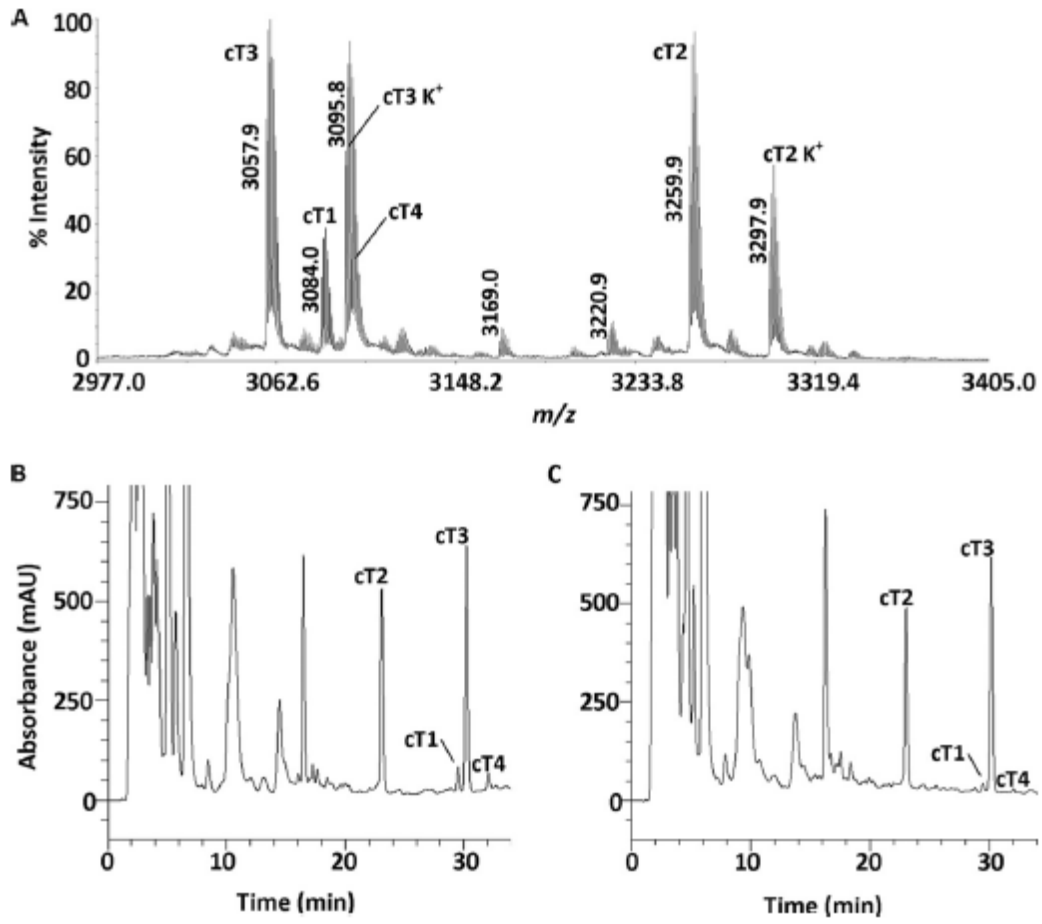


FIGURE 1

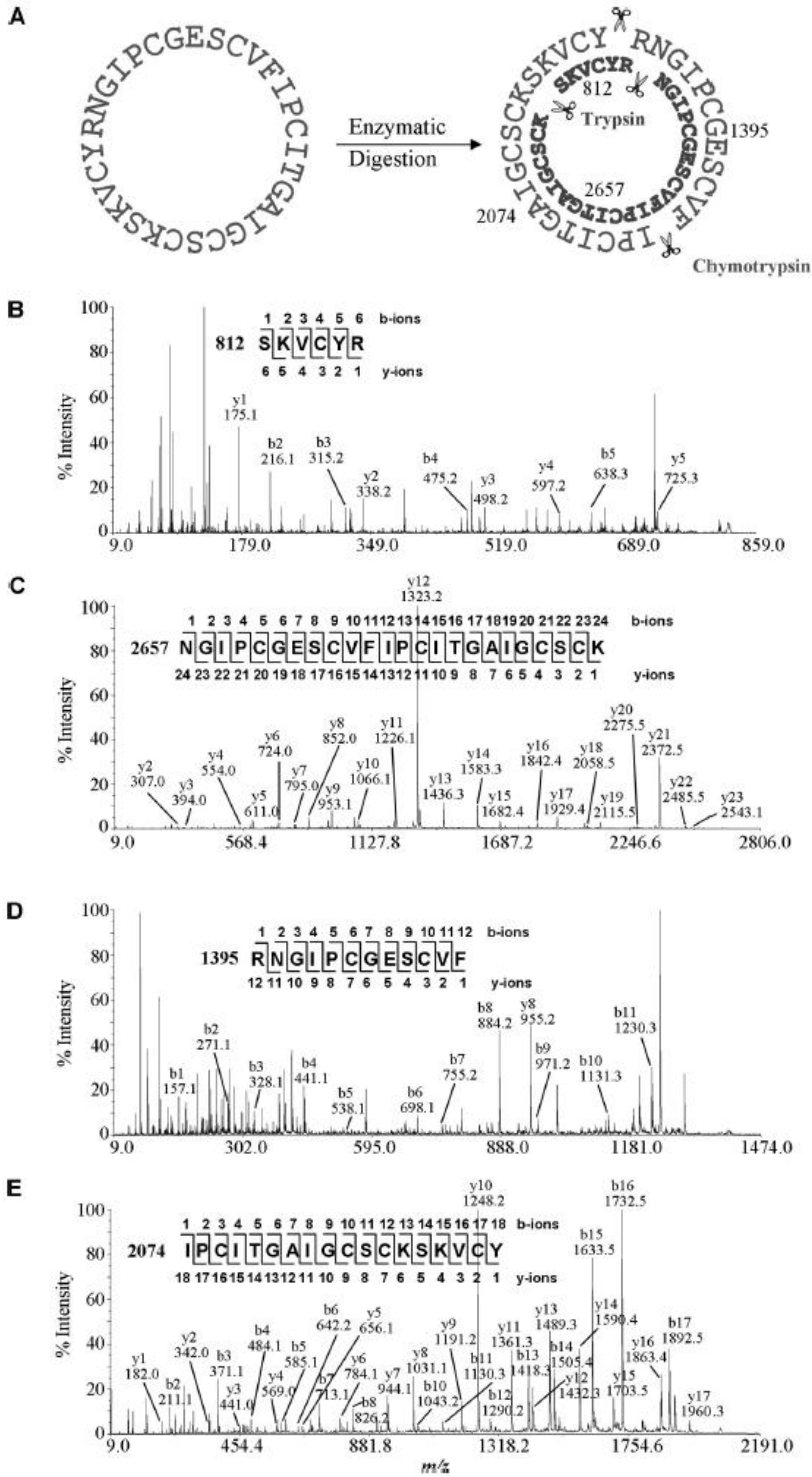


FIGURE 2

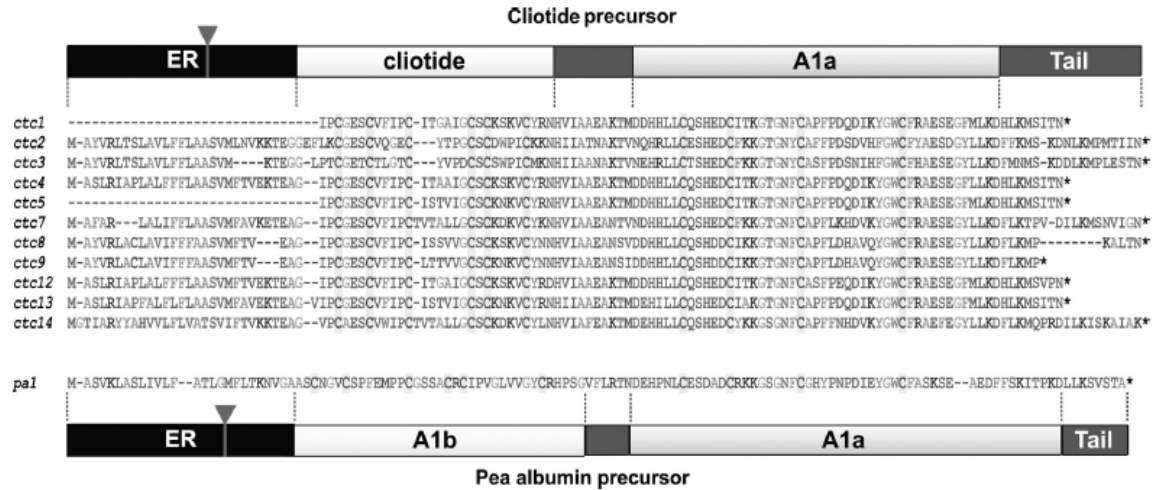


FIGURE 3

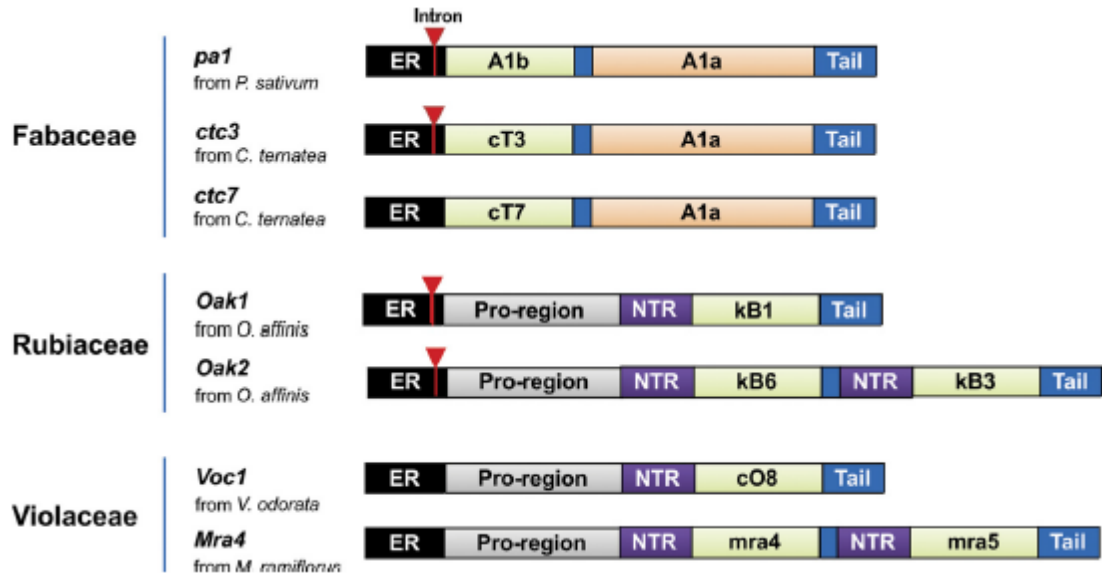


FIGURE 4



FIGURE 5

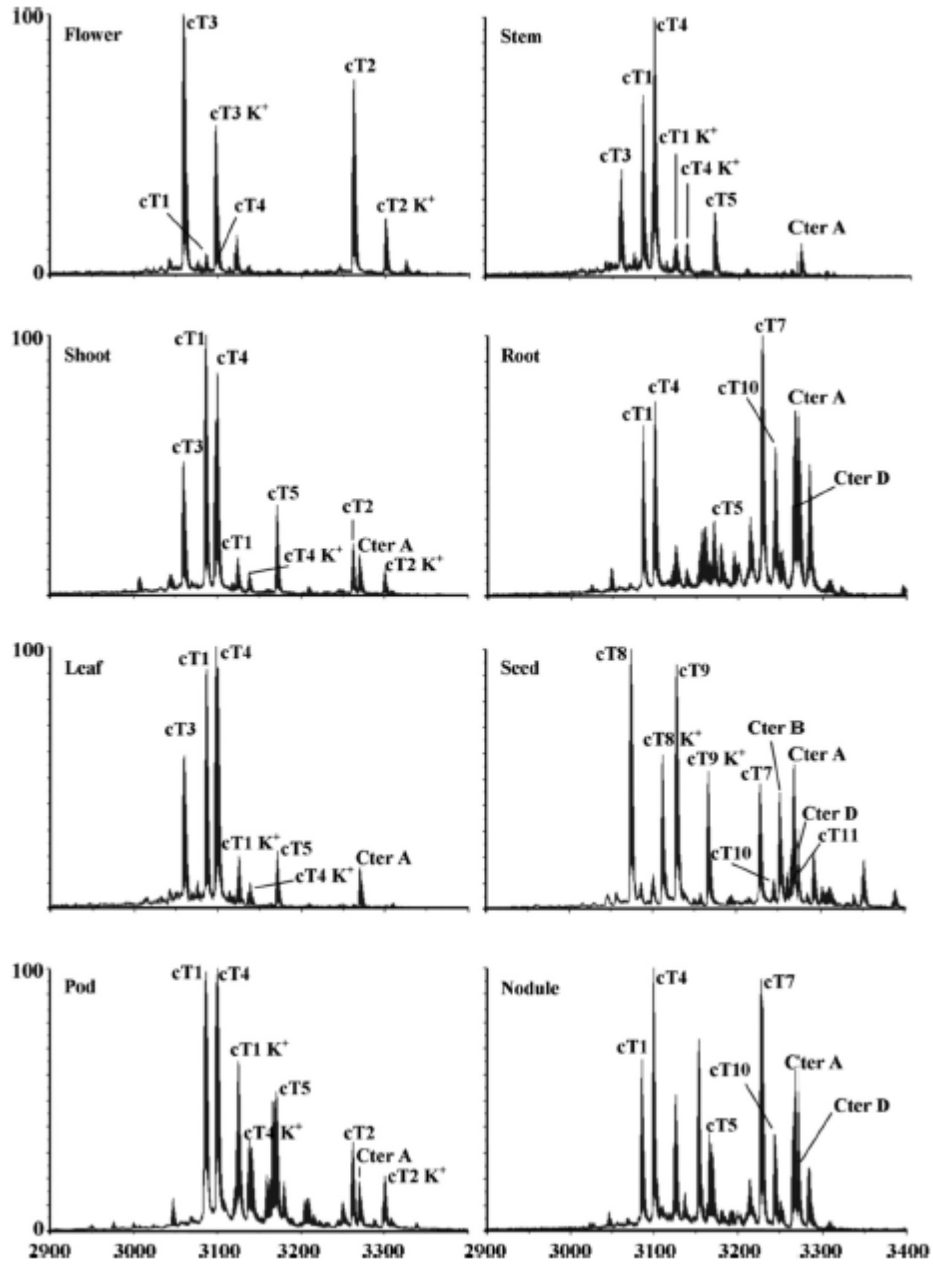


FIGURE 6

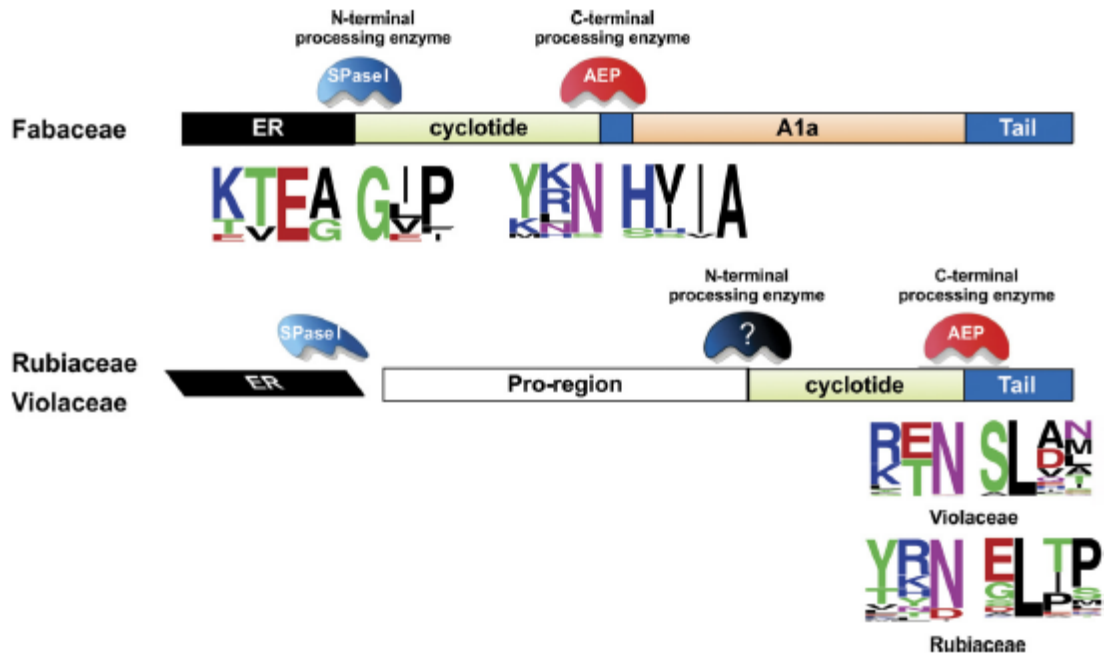


FIGURE 7

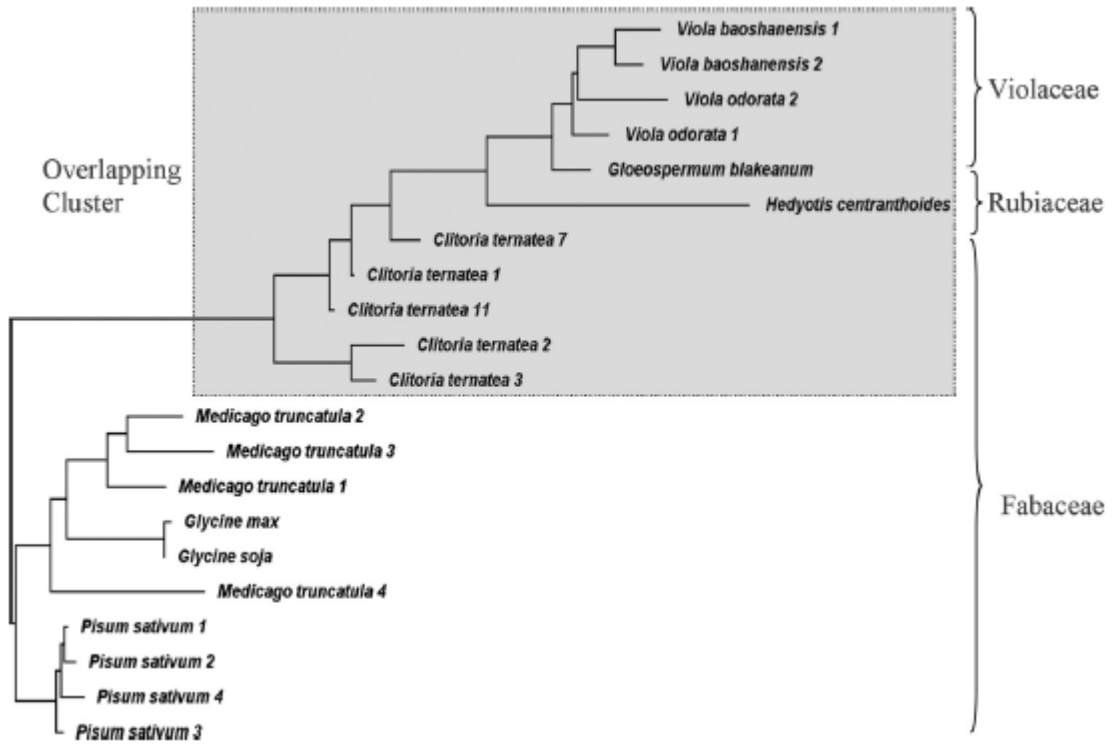


FIGURE 8