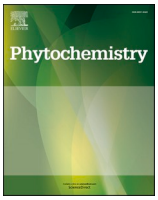
Phytochemistry 187 (2021) 112749

Contents lists available at ScienceDirect

Phytochemistry

journal homepage: www.elsevier.com/locate/phytochem



Tropical vibes from Sri Lanka - cyclotides from *Viola betonicifolia* by transcriptome and mass spectrometry analysis

Sanjeevan Rajendran [a,b], Blazej Slazak [a,c], Supun Mohotti [b], Adam A. Strömstedt [a], Ulf Göransson [a], Chamari M. Hettiarachchi [b], Sunithi Gunasekera [a,*]

[a] *Pharmacognosy, Department of Pharmaceutical Biosciences, Uppsala University, 751 23, Uppsala, Sweden*
[b] *Department of Chemistry, Faculty of Science, University of Colombo, Thurston Rd, Colombo 03, Sri Lanka*
[c] *W. Szafer Institute of Botany, Polish Academy of Science, 46 Lubicz St., 31-512, Cracow, Poland*

ARTICLE INFO

Keywords:
Viola betonicifolia
Violaceae
Cyclotides
LC-MS
Transcriptome *de novo* sequencing
Cyclisation enzyme
Protein disulfide isomerase

ABSTRACT

Cyclotides are an extremely stable class of peptides, ubiquitously distributed in Violaceae. The aim of the present study was to investigate the presence of cyclotides in Sri Lankan Violaceae plants, using combined tools of transcriptomics and mass spectrometry. New cyclotides were discovered for the first time in the wild flora of Sri Lanka, within *Viola betonicifolia*, a plant used in traditional medicine as an antimicrobial. Plant extracts prepared in small scale from *Viola betonicifolia* were first subjected to LC-MS analysis. Subsequent transcriptome *de novo* sequencing of *Viola betonicifolia* uncovered 25 new (vibe 1–25) and three known (varv A/kalata S, viba 17, viba 11) peptide sequences from Möbius and bracelet cyclotide subfamilies as well as hybrid cyclotides. Among the transcripts, putative linear acyclotide sequences (vibe 4, vibe 10, vibe 11 and vibe 22) that lack a conserved asparagine or aspartic acid vital for cyclisation were also present. Four asparagine endopeptidases (AEPs), *Vb*AEP1-4 were found within the *Viola betonicifolia* transcriptome, including a peptide asparaginyl ligase (PAL), potentially involved in cyclotide backbone cyclisation, showing >93% sequence homology to *Viola yedoensis* peptide asparaginyl ligases, *Vy*PALs. In addition, we identified two protein disulfide isomerases (PDIs), *Vb*PDI1-2, likely involved in cyclotide oxidative folding, having high sequence homology (>74%) with previously reported Rubiaceae and Violaceae PDIs. The current study highlights the ubiquity of cyclotides in Violaceae as well as the utility of transcriptomic analysis for cyclotides and their putative processing enzyme discovery. The high variability of cyclotide sequences in terms of loop sizes and residues in *V. betonicifolia* showcase the cyclotide structure as an adaptable scaffold as well as their importance as a combinatorial library, implicated in plant defense.

1. Introduction

Cyclotides are an exceptionally stable family of gene-encoded plant miniproteins. Their N and C termini are joined by a peptide bond to form a continuous circular backbone, which together with three intertwined disulfide crosslinks, result in their signature cyclic cystine knot (CCK) motif (Craik et al., 1999) as shown in Fig. 1. Among the plant families investigated so far, cyclotides are found within the families Rubiaceae (Lorents Gran, 1973), Solanaceae (Poth et al., 2012), Fabaceae (Colgrave et al., 2011), Cucurbitaceae (Hernandez et al., 2000) and Violaceae (Göransson et al., 1999; Schoepke et al., 1993). Cyclotides show a diverse range of bioactivities including antimicrobial (Pranting et al., 2010; Tam et al., 1999), cytotoxic (Lindholm et al., 2002; Schöpke et al., 1993), anti-HIV (Gustafson et al, 1994, 2000), insecticidal (Jennings et al., 2001) and uterotonic activities (Lorents Gran, 1973; Sletten and Gran, 1973). Their ecological role in plants points to host-defense (Jennings et al., 2001; Slazak et al., 2016). Cyclotides are ribosomally synthesised and post-translationally modified peptides (RiPPs) (Van der Donk., 2013; Dutton et al., 2004; Jennings et al., 2001). This allows for their identification by genetic tools, presenting complementing tools for peptide discovery. As shown in Fig. 1, cyclotides grouped into the Möbius subfamily contain a Pro reside in loop 5 usually preceded by a Trp residue, facilitating a *cis*-Pro bond and resulting in a conceptual 180° twist in the backbone (Craik et al., 1999). Cyclotides, lacking this *cis*-Pro

* Corresponding author. Pharmacognosy, Department of Pharmaceutical Biosciences, Biomedical Center, Uppsala University, Box 591, SE, 751 24, Uppsala, Sweden.
E-mail address: sunithi.gunasekera@farmbio.uu.se (S. Gunasekera).

https://doi.org/10.1016/j.phytochem.2021.112749
Received 10 January 2021; Received in revised form 19 March 2021; Accepted 20 March 2021
Available online 28 April 2021

residue, instead having a *trans*-amide bond at the corresponding position, have a topology that mimics an amino acid bracelet, thus termed bracelet cyclotides - the second subfamily. However, the differentiation of cyclotides goes beyond the *cis*-Pro conformation, as hybrid cyclotides having loops similar to both Möbius and bracelet groups have also been identified (Daly et al., 2006). The third subfamily, confined to Cucurbitaceae and differentiated by trypsin inhibition, is atypical in sequence and loop size to other cyclotides and is more similar to linear trypsin inhibitor knottins from squash plants (Hernandez et al., 2000). Variants of naturally occurring cyclotides with an open backbone were also identified in plants and termed acyclotides (Nguyen et al., 2012; Poth et al., 2012).

Violaceae consists of 23–31 genera and approximately 1000–1100 species worldwide (Burman et al., 2015; Wahlert et al., 2014). Most genera contain a small number of species and the three largest genera, *Viola*, *Hybanthus* and *Rinorea* account for more than 90% of the family. The largest of these three genera is *Viola*, commonly known as violets, a genus comprising herbs with spurred flowers. The other genera in the Violaceae family have radially symmetrical flowers and are lianas, shrubs, or either large or small trees (Wahlert et al., 2014). Eight species from the Violaceae are found in Sri Lanka: three from genus *Rinorea*, three from genus *Viola,* and two from genus *Hybanthus*. *Rinorea* species are thought to be extinct in Sri Lanka and endangered plants from the other two *Viola* and *Hybanthus* species are found only in wild forests (Ministry of Mahaweli Development and Environment, 2016) (Wijesundara et al., 2012).

Cyclotide discovery efforts on Violaceae outnumber any other plant family and almost every Violaceae plant screened has shown a positive outcome, either at peptide or genetic level, for the presence of cyclotides (Burman et al., 2015). To our knowledge, no examples for Violaceae plants devoid of cyclotides are reported in literature. The number of cyclotide variants among Violaceae plants is estimated to above 150,000 different sequences, or on average 160 per species (Hellinger et al., 2015). Notably, neither cyclotides nor any other secondary metabolites from Sri Lankan *Viola* plants have been investigated, making them good candidates for discovery of new chemistry. The presence of cyclotides in *Viola betonicifolia* collected in Australia, was hinted in previous studies but their peptide sequences and/or structural information were not probed in detail (Craik et al., 1999; Gruber, 2010; Gruber et al., 2008; Koehbach et al., 2013).

Geographical location and environmental conditions are factors that influence the plant cyclotide profile (Trabi et al., 2004). For example, *Viola odorata* found within India expresses a suite of different cyclotides, depending on its geographical location (Narayani et al., 2017). This is attributed to large variations in environmental factors such as the type of soil, altitude, water, temperature, humidity, sunlight and related stress factors associated with geographical location (Seydel et al., 2007). Thus, it was also of interest to investigate the cyclotide expression at transcriptome and peptide level within Violaceae originating from Sri Lanka.

Water extract or decoction of whole herb or part of the plant such as the flowers of *V. betonicifolia* is used by local people and tribes in India and Pakistan in traditional medicine as a diuretic and to treat blood disorders, as well as for a variety of therapeutic conditions including nervous system diseases like epilepsy, cough, pneumonia, pharyngitis and boils (Chandra et al., 2015; Bhatt and Negi, 2006; Zahoor Husain et al., 2008). This traditional practice also motivated the current study because cyclotides were speculated as a prime group of candidate molecules responsible for the plant's bioactivities. This was supported by the pre-existing knowledge that cyclotides commonly occur within Violaceae and are naturally endowed with a wide variety of inherent bioactivities (Srivastava et al., 2020).

Herein we discovered a new set of cyclotides and acyclotides in *V. betonicifolia* using transcriptomic analysis, small scale peptide extraction and mass spectrometric analysis. The sequence diversity of new cyclotides indicate their high bioactive potential. The precursor sequences of cylotides give further insight into potential processing sites

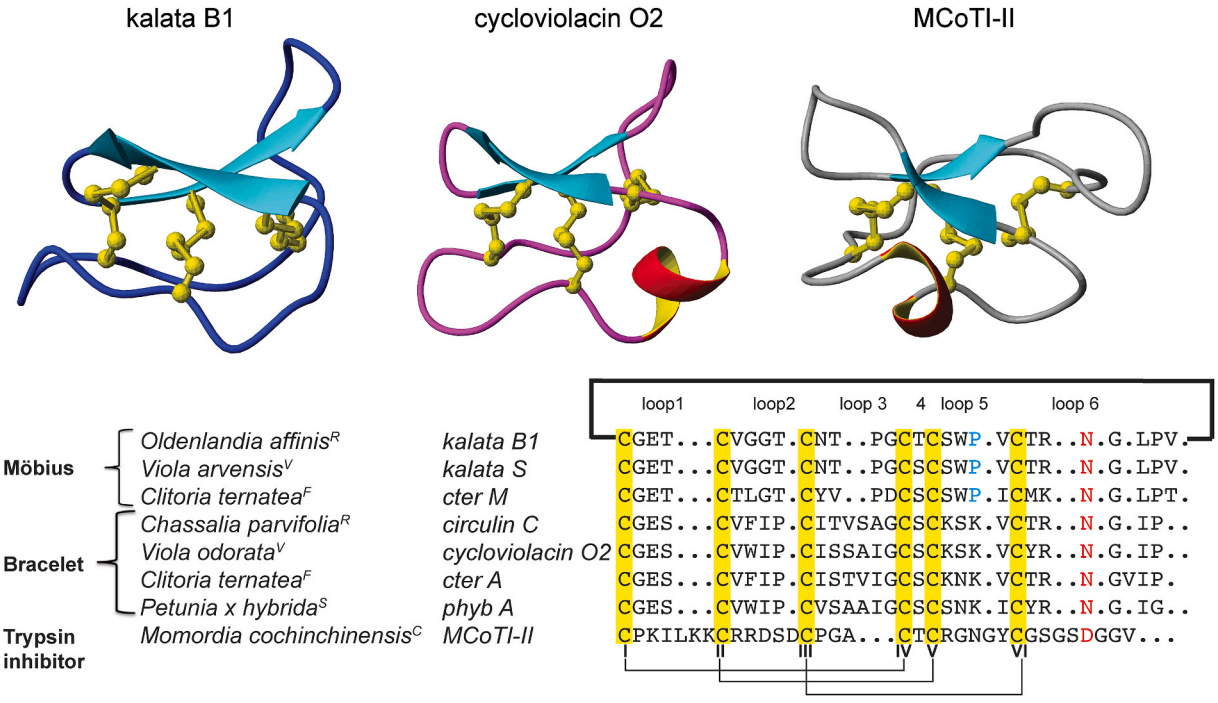


Fig. 1. Example structures and sequences from the three main cyclotide subfamilies. The structures are based on the PDB files for kalata B1 (1nb1), cycloviolacin O1 (1nbj) and *Momordica cochinchinensis* II (MCoTI-II) (1ib9). The unique cyclic cystine knot (CCK) topology of cyclotides arises when the ring formed by Cys^{I}-Cys^{IV} and Cys^{II}-Cys^{V} together with the backbone loops 1 and 4 are penetrated by the third disulfide between Cys^{III} and Cys^{VI}. The cyclotide producing plant families are denoted by ^{R}Rubiaceae ^{V}Violaceae, ^{F}Fabaceae, ^{S}Solanaceae and ^{C}Cucurbitaceae. The conserved Cys residues are highlighted in yellow and the *cis*-pro in loop 5 that defines the Möbius subfamily is highlighted in blue. The conserved Asn/Asp residue at which cyclisation occurs is highlighted in red. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

accessible to processing enzymes during cyclotide maturation. The transcriptome also formed the basis for identifying a putative enzyme *Vb*AEP1 that could presumably be involved in cyclotide biosynthesis. Our results highlight that combination of techniques, using transcriptomic and peptidomic tools can guide and accelerate cyclotide discovery.

2. Results and discussion

2.1. Identification of novel cyclotide sequences in the Viola betonicifolia transcriptome

A high-quality transcriptome assembly is a prerequisite for studying biosynthesis and regulation of specialised plant metabolites. To capture active transcripts of *V. betonicifolia*, cDNA libraries were prepared from the combined tissues obtained from leaves, stems, seeds and roots. RNA-sequencing was performed using an Illumina Hi-Seq 3000 platform. Summary of detailed information on the *de novo* transcriptome assembly for *V. betonicifolia* is given in Supplementary Table S1.

A total of 28 cyclotide precursor sequences were prepared as DNA sequences for the sequence alignment (Fig. 2, Supplementary Data S1), comprising selected domains of the precursor, *i.e.*, N-terminal propeptide (NTPP), N-terminal repeat (NTR), and mature cyclotide domains. To guide the alignment of DNA sequences of the precursors, they were translated into protein sequences and aligned independently for each precursor domains using ClustalX in Ugene software package. The resulting protein-guided alignment of nucleotide sequences was then subjected to manual adjustments within reading frames or precursor domains.

Analysis of precursor sequences and putative processing sites provided further insight into cyclotide biosynthesis in *V. betonicifolia* and assisted to distinguish cyclotides from other cysteine-rich plant proteins. A total of 25 new peptide sequences (vibe 1–25) and three known sequences (viba 11, viba 17, and varv A/kalata S) were identified from transcriptomic analysis (Table 1). New cyclotides (vibe 1–25) were named according to the nomenclature proposed by Broussalis et al. (2001), with the plant binomial name followed by the number indicating the order (vibe for *Viola betonicifolia*).

As highlighted in Table 1, these 25 new cyclotides/acyclotides were detected in *V. betonicifolia* transcriptome and 12 of them were only found as transcripts but not at peptide level. Two cleavage events, one at the N-terminal and the other at the C-terminal in the cyclotide precursors, lead to the formation of a linear peptide that is subsequently ligated to form the head-to-tail cyclized backbone. In demarcating loop 6 where potential cyclisation occurs, we propose that the vibe precursors are most likely cleaved after an Asn/Asp residue at the C-terminal processing site and N-terminal of a Gly residue, at the N-terminal processing site. This notion was esablished based on cyclotide precursors identified in previous studies to date (Dutton et al., 2004), Based on these most probable cleavage sites, the expected mature vibe cyclotide masses were calculated and compared with observed peptide masses (Table 1). Based on the presence or absence of the Pro residue in loop 5, members of both Möbius and bracelet cyclotides were found at transcriptome level. Some cyclotides are hybrids between the two subfamilies (vibe 18–19, vibe 23–24), in particular they have a loop 3 more aligned with typical Möbius-type. Vibe 22 likewise lacks the *cis*-Pro but its loops lack similarity with either subfamily besides carrying the conserved PG dipeptide sequence in loop 3 (Möbius characteristic), together making it a distinct hybrid cyclotide.

All mature cyclotide products begin with a Gly residue at the N-terminal except vibe 16 that contains an Ala at the corresponding position (Fig. 2 and Table 1). This indicates that either a Gly or Ala at N-terminal are capable of undergoing a transpeptidase reaction with the conserved Asn/Asp at C-terminal during cyclisation. In 21 of the new vibe precursor sequences identified, an Asn or Asp occupies the C-terminus of the cyclotide domain which is followed by a highly conserved SL and rarely AL or SI dipeptide sequence. As it is known that the conserved Asn/Asp facilitates head-to-tail cyclisation, these vibe sequences were identified as 'cyclotides'. In contrast, four of the precursors differ significantly in the C-terminal processing sites and were identified as acyclotides (vibe 4, vibe 10, vibe 11, vibe 22) (Table 1, Fig. 2, supplementary Data S1). Vibe 4 and 11 precursor sequences terminate at a C-terminal Asn and Asp residue, respectively, due to a stop codon. However, despite the presence the Asn and Asp critical for cyclisation, the absence of the C terminal signal peptide, prevents the cyclotide precursor affinity for the putative cyclizing enzyme (Dutton et al., 2004; Gillon et al., 2008). Vibe 10 and 22 likewise contain stop codons in their mature cyclotide domain and lack the conserved Asn/Asp residue, vital for cyclisation. In previous reports, the occurrence of acyclotides is infrequently reported for species of the Rubiaceae (Plan et al., 2007), Violaceae (Hellinger et al., 2015), Solanaceae (Poth et al., 2012) and Poaceae (Nguyen et al., 2013) plant families. Cyclotide precursors with repeating mature domains were also observed. Two different mature domains are embedded in vibe 24 precursor, both vibe 24 and vibe 14 (Fig. 2). In contrast, varv A/kalata S transcript, contains two repeating mature varv A/kalata S domains.

Analysis of cyclotide precursors identified so far reveals that the specific tail sequence of the cyclotide precursor protein, recognized by the cyclisation enzyme, starts with a sequence of two amino acids, GL/SL in Rubiaceae (Dutton et al., 2004; Gillon et al., 2008), SL/AL in Violaceae (Dutton et al., 2004) and HV in Fabaceae (Nguyen et al., 2011). In vibe precursors, we also observe an SI combination of residues in the C-terminal propeptide sequence (vibe 19). Thus, the prerequisites for cyclisation in vibes is most likely the conserved Asn/Asp at the S1 binding site, followed by the presence of a residue with a small side chain (G, S, A or H), and subsequently followed by a hydrophobic residue (L, I or V) in the C-terminal recognition sequence.

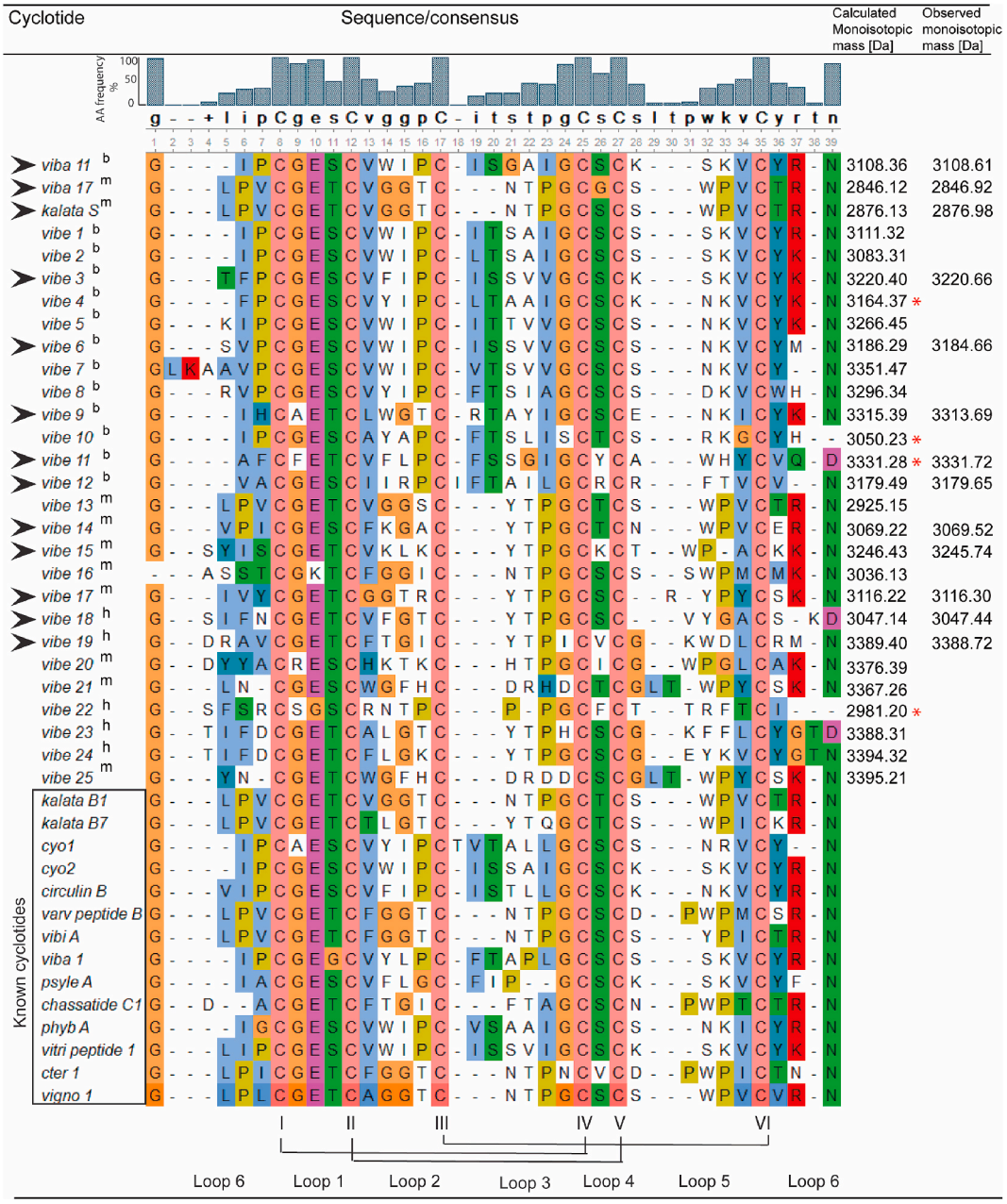
To investigate the plasticity of vibe cylotides, we analysed them in detail, loop by loop. Cyclotide loop 1 is reported to contain a highly conserved GES/GET triad (Burman et al., 2015). However, among the new sequences identified, loop 1 is occupied by AET (vibe 9) and FET (vibe 11) sequences. Notably, positively charged residues, Lys in vibe 16 (GKT) and Arg in vibe 20 (RES) in loop 1 are present. Another peculiarity is observed in two sequences; vibe 16 (GKT in loop1) and vibe 22 (SGS in loop1) lack the conserved E, that is typically present in loop 1. In previous studies, the conserved Glu in loop 1 of cyclotides was shown to be important for overall structural stability as well as bioactivity. In cycloviolacin O2, methylation of the Glu or substitution with Ala abolishes the membrane disrupting activity that is characteristic of the family (Göransson et al., 2009). Wang et al. also showed that in kalata B12 which has an Asp in place of Glu in loop 1, the overall cyclotide fold is maintained, but due to a weaker H bond network, the three-dimensional structure is more flexible. Kalata B12 is devoid of the prototypical hemolytic activity of cyclotides and the structure is less resistant to treatment with acids (Wang et al., 2011). Cyclotide sequences where Glu in loop 1 is substituted with Ser or Gly have also been also observed at transcript level in *Carapichea ipecacuanha* (Brot.) L. Andersson of Rubiaceae family - sequences SST and TGT occupy loop 1 in cyclotide caripe 4 and caripe 6 respectively, but the impact on biological activity due to the absence of Glu has not been studied in detail (Koehbach et al., 2013).

Loop 2 of bracelet cyclotides appear rather conserved in *V. betonicifolia* with VX (I/L)P (X = W, F, Y) residue pattern except, vibe 9 (LWGT), 10 (AYAP) and 12 (IIRP). In comparison, loop 2 of Möbius vibe cyclotides is much more variable. In particular, the presence of Gly-rich sequences (VGGT, VGGS, FGGI) and Lys-rich sequences (VKLK, HKTK) are apparent. Vibe 15, which contains five Lys with a net charge of +4 projected as the most basic peptide of all vibe cyclotide (Fig. 4). Loop 2 and 5 of vibe 15 and vibe 20 are markedly very similar to corresponding loops of previously reported lysine-rich cyclotides - cycloviolacin O14 *Viola odorata* (Ireland et al., 2006), mela 1–7 from *Melicytus latifolius* (Ravipati et al., 2015) and globa E from *Gleospermum blakeanum*

Fig. 2. Schematic representation of the general organisation of cyclotide precursors and multiple sequence alignment of 28 cyclotide precursor proteins obtained from the *de novo* transcriptome assembly of *V. betonicifolia*. A. ER signal domain at the start of each precursor sequence is highlighted in blue where cleavage of the signal domain is predicted to occur between blue and pink residues (Dutton et al., 2004). The precursor is organised with an N-terminal propeptide domain, NTPP (black), N-terminal repeat, NTR (green), mature cyclotide domain (red) and C-terminal propeptide domain, CTPP (purple). An AEP mediated cleavage potentially occurs at the conserved Asn/Asp adjacent to the CTPP of all cyclotide sequences; B. New cyclotides are named vibe 1–25. In varv A/kalata S and vibe 24 transcripts, two repeating mature domains are present. In acyclotides, either the conserved Asn/Asp or CTPP sequence is absent. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1

ClustalX alignment of amino acid sequences of *Viola betonicifolia* cyclotides/acyclotides in comparison with previously characterised cyclotides. Vibe cyclotides were identified based on transcriptome analysis and the monoisotopic masses were calculated using mMass software tool (Niedermeyer and Strohalm, 2012). Vibe cyclotides/acyclotides observed in the transcriptome, coinciding with corresponding candidate cyclotide masses present at peptide level are highlighted with an arrow. The expected and observed masses for $(M + H)^+$ are shown. Jalview coloring scheme was used to efficiently display sequence homology. Acyclotides are highlighted by a "*" next to the calculated mass.



[b]Bracelet cyclotides, [m]Möbius cyclotides, [h]Hybrid cyclotides. Previously reported known cyclotides are highlighted in the box.

(Burman et al., 2010). In previous reports, lysine-rich cyclotides have been suggested as favorable scaffolds to design new anticancer drugs due to their lower hemolytic activity and their tendency to specifically target membranes rich in anionic-lipids found on cancer cells (Ravipati et al., 2015).

Loop 3 of vibe cyclotides shows similar features as previously described cyclotides (Ravipati et al., 2017). Möbius subfamily cyclotides typically comprise four residues in their loop 3. Bracelet cyclotides, on the other hand, show greater diversity in sequence and size (four to six residues) in this loop. Both subfamilies possess a highly conserved Gly residue in the last position of loop 3 for most sequences. However, some exceptions are observed in vibe cyclotides: vibe 10, 19, 21, and 25 sequences lack a Gly as the last residue in loop 3. The unusual vibe 22 contains only three residues its loop 3 (PPG), which likely affects the conformation of this loop significantly. Vibe 12 consists of 7 residues in loop 3, which is the longest loop 3 among all cyclotides identified in *V. betonicifolia*.

To identify unique features, we compared vibe cyclotides identified herein with the cyclotides reported in cybase, a database for all reported cyclotides and acyclotides (Wang et al., 2008) (www.cybase.org.au). Loop 4, the smallest of all loops, contains only a single residue, Ser or Thr in majority of the vibe cyclotides. This is in line with other cyclotides

reported to date. However, variable residues are observed in loop 4, including Lys, Arg, Val, Gly, Tyr, Ile, which is unusual. In general, 5–8 residues are typically present in cyclotide loop 6. Some exceptions are reported, such as vitri peptide 24a, vitri peptide 36 and Tricylon where up to ten residues in loop 6 have been noted (Hellinger et al., 2015). Along similar lines, vibe 7 appears to contain the longest loop 6 comprising of nine residues.

2.1.1. Small scale LC-MS screening of cyclotides in Violaceae

Previously, *V. betonicifolia* was described as a cyclotide containing species using LC-MS (Craik et al., 1999). However, cyclotide sequences from *V. bentonicifolia* have not been previously reported. In our search to identify these molecules, we first used LC-MS to detect cyclotide masses at peptide level. LC-MS analysis identified peaks in the range of 2.8–3.4 kDa for $(M+1)^{1+}$, characteristic of potential cyclotides in crude *V. betonicifolia* extract as shown in Fig. 3 and Supplementary Table S2. Of 25 transcripts apparent in the transcriptome, 13 cyclotide candidate masses were observed by LC-MS (Table 1, Fig. 3). These include cyclotides, vibe 3, 6, 9, 11, 12, 14, 15, 17, 18, 19, viba 11, viba 17 and varv A/kalata S. Some of the most abundant cyclotide-like masses in the extract were partially purified by RP-HPLC and subjected a reduction/alkylation procedure (Yeshak et al., 2011). The presence of six cysteines was confirmed by a mass increase of 348 Da for each cyclotide candidate (6 × 58 Da), following the reduction of the six cysteines and their *S*-carbamidomethylation (Fig. 3, Supplementary Figure S1 and Supplementary Table S3). Upon opening of the cyclic backbone by endoproteinase GluC, due to cleavage after a single conserved Glu present in most cyclotides, the $(M+1)^{1+}$ mass further increased by 18 Da (Fig. 3, Supplementary Table S4). However, due to limited plant material at hand, vibe cyclotides were not fully sequenced using MSMS, thus the masses identified are 'candidate cyclotides' in *V. betonicifolia* that matched in mass with the *V. betonicifolia* cyclotides predicted in the transcripts.

Of vibe cyclotides identified, vibe 12 stands out as the most hydrophobic and positively-charged cyclotide with a calculated hydropathicity index of 1.1 (hydrophobicity 43.3%) and a net charge of 1.7 at

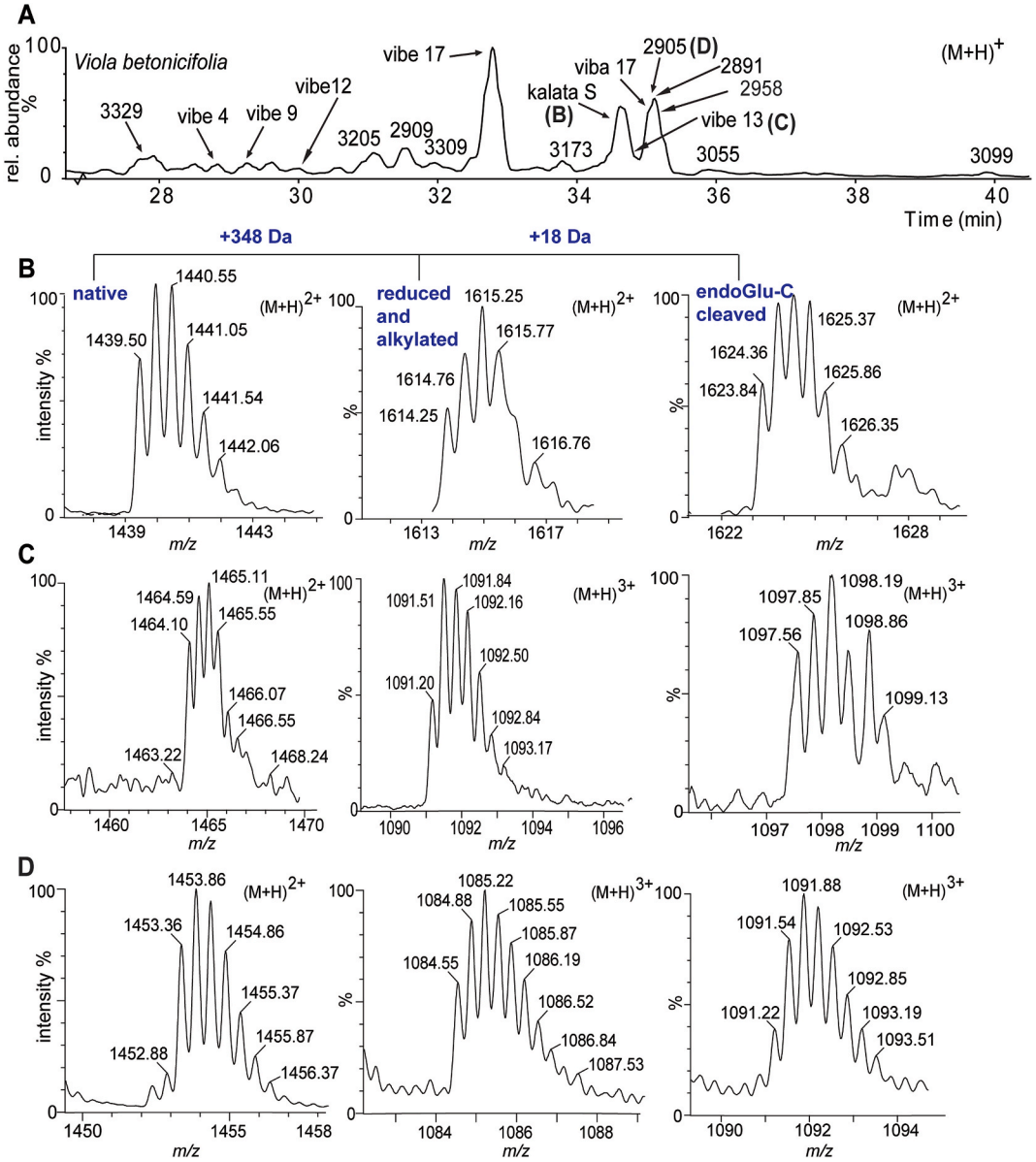


Fig. 3. Vibe cyclotides identified from small scale extraction of *V. betonicifolia*. A. Base peak ion (BPI) chromatogram from *V. betonicifolia* containing deconvoluted masses for $(M + H)^{+}$ of candidate cyclotides. B. Isotopic mass pattern of native, reduced/alkylated, endoproteinase GluC cleaved cyclotides exemplified by kalata S/varv A (B), vibe 13 (C) and a new cyclotide present in the extract but absent in the transcriptome (D).

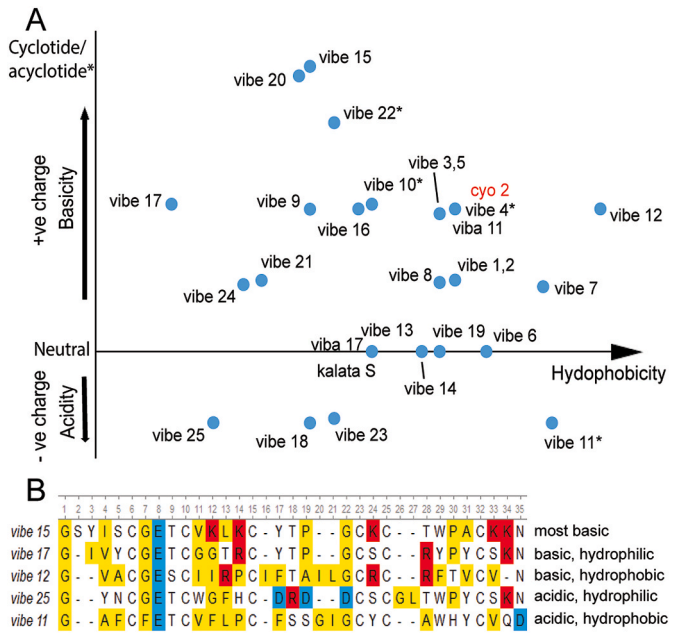


Fig. 4. Vibe cyclotide/acyclotide properties A. Graphical representation of net acidity, basicity and hydrophobicity of cyclotide/acyclotide* from *V. betonicifolia,* calculated using a peptide property calculation tool, https://www.peptide2.com/N_peptide_hydrophobicity_hydrophilicity.php (Kyte and Doolittle, 1982; Sims, 2010). B. Sequence alignment highlighting physicochemical similarity. Hydrophobic residues are highlighted in yellow, basic in red and acidic in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

pH 7 as shown in Fig. 4. In both cationicity and hydrophobicity, vibe 12 appears to be superior to well-characterised membrane disrupting cyclotide, cycloviolacin O2 (Strömstedt et al., 2017). In contrast, vibe 25 is the most hydrophilic and acidic sequence with hydropathicity index of −0.72 (hydrophobicity 16.7%) and net charge of −2.2 at pH 7. Aside from vibe 25, several prominent acidic sequences including vibe 11, vibe 18 and vibe 23 were also identified. Among the other cyclotides identified, vibe 17 and 11 are at two extremes with regard to their physiochemical properties - vibe 17, a positively charged but extremely hydrophilic cyclotide and vibe 11, an acidic but extremely hydrophobic cyclotide.

The presence of varv A/kalata S in *V. betonicifolia* validates the previous observations that it is a cyclotide hosted in many plant species with a high tendency to be expressed in Violaceae. Some of the transcripts were absent at peptide level, presumably due to production as minor components or because these peptides were not expressed in the plant. On the contrary, several cyclotide masses were observed at peptide level (Fig. 3, Supplementary Figure S1), however their corresponding transcripts were absent, similar to what has been observed in other cyclotide discovery work (Hellinger et al., 2015; Slazak et al., 2015). This discrepancy can occur due to several reasons. The sample for transcriptomic analysis was not collected at the same time as the sample used for peptide analysis, in fact a quality RNA sample suitable for analysis from *V. betonicifolia* was only obtained after our fourth attempt. Thus, the physiological and seasonal conditions may have affected the cyclotide expression. This also highlights that the full repertoire of cyclotides produced by the plants can be detected only by a combination of methods and cannot be restricted to one point in time. It is also a reminder that analysis at peptide level is irreplaceable by transcriptomic tools applied in discovery. Full genome sequencing is the only means to get the complete picture of the full repertoire of cyclotides that a plant species can potentially make. However, techniques including transcriptomic and MS analyses have the advantage of highlighting those cyclotides that are expressed and functionally important to the plant.

Aside from the three known cyclotides, all 25 cyclotides identified in *V. betonicifolia* are new. In line with the previous observation that less than ~1/3 of the known cyclotides occur in multiple plants (Burman et al., 2015), most of the cyclotides either expressed in the transcriptome or peptide level in *V. betonicifolia* currently appear unique. The presence of a suite of cyclotides varying in loop size, sequence, net charge and hydrophobicity in *V. betonicifolia*, would be valuable in countering emerging resistance in plant pests and pathogens and supports the natural defense role of cyclotides (Slazak et al., 2018). It is known that cytotoxicity of cyclotides derives from their ability to disrupt lipid bilayers where residues of their surface 'bioactive patch' bind to phospholipid head groups and insert into the lipid bilayers via a hydrophobic patch (Burman et al., 2011). Thus, a mixture of cyclotides that are rich in cationic residues as well hydrophobicity would be favorable in targeting different types of organisms, organelles or cell membranes, associated with different insect herbivores and pathogens. A receptor modulated mechanism for cyclotides, involving G protein-coupled receptors (GPCRs), possibly linked to insects or more generally invertebrate neuropeptide receptors, affecting invertebrate development, reproductive behavior and water homeostasis has also been proposed (Gruber, 2014; Keov et al., 2018). Whether different cyclotides play a complementing role by supporting each other in their defense is yet to be shown. Among the transcripts of vibe cyclotides, highly cationic peptides that are ideally suited to interact with the negatively charged bacterial and fungal membranes are present. Bacterial infections potentially underpin some of the traditional medicinal indications of *V. betonicifolia*, such as pneumonia, boils and cough (Chandra et al., 2015), thus drawing a connection to cyclotides as potential drug candidates in the herbal extracts seems like a plausible choice that requires validation by bioactivity studies.

2.2. Transcriptome mining resulted in a sequence similar to known AEPs involved in cyclisation

Although deciphering how unique cyclotide topology arises from linear precursors has been challenging, the involvement of AEP in cyclotide maturation is established for several AEPs from cyclotide bearing plant families to date. First of these enzymes is butelase 1 isolated from *Clitoria ternatea* of the Fabaceae family (Nguyen et al., 2014). Subsequently, *Oa*AEP1b from *Olenlandia affinis* of the Rubiaceae family (Harris et al., 2015) and PxAEP3b from *Petunia X hybrida* E. Vilm. of the Solanaceae (Jackson et al., 2018) were discovered. Within Violaceae, several ligases, *He*AEP3 from *Hybanthus enneaspermus* (Jackson et al., 2018) and *Vy*PAL1-4 from *Viola yedoensis* (Hemu et al., 2019) were found. Mostly recently, *MCo*AEP2 from *Momordica cochinchinensis* of the Cucurbitacease family involved in both N-terminal excision and C terminal cyclisation of the precursors *in vitro* was described (Du et al., 2020). These AEPs not only perform cleavage, but also have the ability to ligate linear peptides.

BLAST search in the *V. betonicifolia* transcriptome for AEP enzymes similar to verified cyclizing enzymes from cyclic peptide bearing plants yielded four main hits (Supplementary Table S5). The sequences of the enzyme catalytic region were aligned and a similarity grid was generated (Fig. 5). If the vibe enzymes show high sequence similarity to well-characterised ligases identified in other cyclotide bearing species, and more importantly in other *Viola* species, we speculated that those enzymes will have high likelihood to be potential ligases involved in vibe cyclotide processing. First of the AEPs identified, *V. betonicifolia* AEP1 (VbAEP1) showed 95% and 93% sequence similarity to *Vy*PAL1 and *Vy*PAL2 from *V. yedoensis* respectively, two verified ligases with high *in vitro* backbone cyclisation efficiency at pH 5–8 (Hemu et al., 2019) (Supplementary Table S2). Several 'ligase activity determining' regions are proposed for AEPs from cyclotide bearing plants to distinguish whether the enzymes are predominantly potential proteases or ligases. As shown in Fig. 5, these regions include, marker of ligase region (MLA), ligase activity determinants 1 and 2 (LAD1 and LAD2), 'gate keeper'

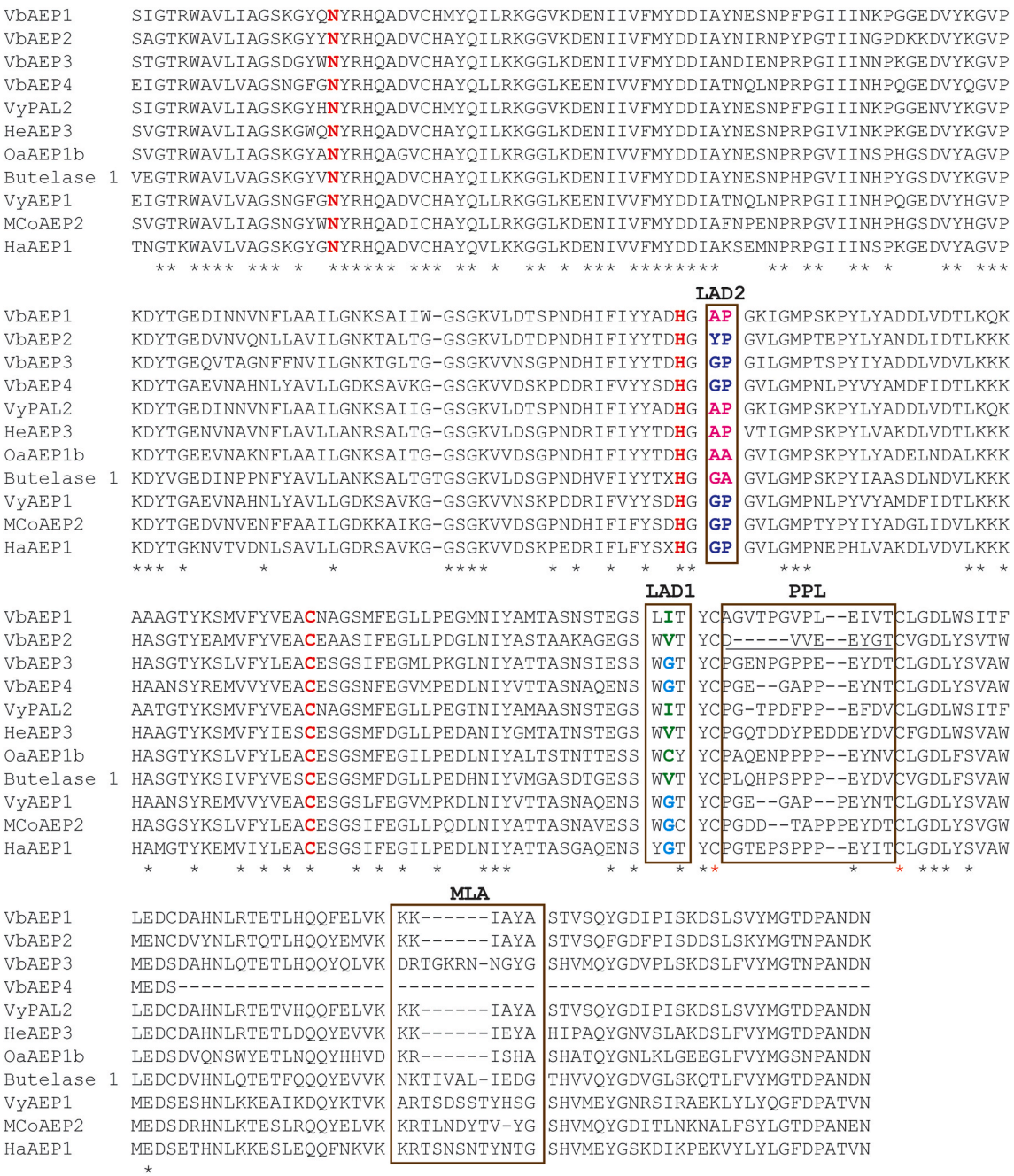


Fig. 5. Multiple sequence alignment of vibe AEPs and other functionally verified ligase- and protease-type AEPs identified in cyclic peptide producing plants. The AEP conserved catalytic triad residues Asn, Cys and His are highlighted in red. In the aligned AEPs, ligase activity determinant 1 (LAD1) containing gate keeper residue (highlighted in green/cyan, position orthologous to CYS247 in *Oa*AEP1b (Harris et al., 2015) and ligase activity determinant 2 (LAD2) regions (highlighted in pink/blue) (Hemu et al., 2019), poly-proline region (Jackson et al., 2018) and marker of ligase activity region (MLA) (Jackson et al., 2018) are highlighted in boxes. *Vy*PAL2 from *V. yedoensis* (Hemu et al., 2019), *He*AEP3 from *H. enneaspermus* (Jackson et al., 2018), *Oa*AEP1b from *O. affinis* (Harris et al., 2015) and butelase 1 from *C. ternatea* (Nguyen et al., 2014) are functionally verified ligases. *MCo*AEP2 has also shown efficient *in vitro* ligase activity, despite the presence of protease-type ligase activity determinant regions (Du et al., 2020). *Vy*AEP1 from *V. yedoensis* and *Ha*AEP1 from *H. annus* are protease-type AEPs with weaker ligase activity at high/neural pH (Haywood et al., 2018; Hemu et al., 2019). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

residue within LAD1 and poly-proline region. *Vb*AEP1 has a *i)* 'Ile' as the gate keeper residue within LAD1, *ii)* 'AP' dipeptide sequence in LAD2 and an *iii)* truncated MLA region, features indicative of a potential ligase (Hemu et al., 2019; Jackson et al., 2018). Because the latter regions correspond well with 'ligase-type' determinants identified previously for *Vy*PAL2 (Hemu et al., 2019), we propose that *Vb*AEP1 is most likely involved in vibe cyclotide backbone cyclisation.

The second enzyme, *Vb*AEP2 most closely matched *Vy*PAL3 (with 89% sequence homology), a ligase with low catalytic efficiency identified in *Viola yedoensis* (Hemu et al., 2019). A bulky residue such as Tyr as the first residue in 'YP' dipeptide within LAD2 has been shown to contribute to protease activity and concurrent low ligase efficiency of *Vy*PAL3 (Hemu et al., 2019). Despite the presence of ligase favorable residues - *i)* 'Val' as the gate keeper residue within LAD1 and *ii)* a truncated MLA region, *Vb*AEP2 also contains Tyr as the first residue within 'YP' dipeptide sequence in LAD2 similar to *Vy*PAL3, skewing it

more towards a ligase with low catalytic efficiency. Notably, *Vb*AEP2 also has a truncated poly-proline region different from what is observed in previously established plant proteases and ligases (Fig. 5).

*Vb*AEP3 and *Vb*AEP4 show high sequence homology (>98%) to more protease-type *Vy*AEP2 and *Vy*AEP3 identified in *Viola yedoensis* (Supplementary Table S5). In *Vb*AEP3 and *Vb*AEP4, the gate keeper residue is 'Gly' in LAD1 and 'GP' dipeptide sequence is found within LAD2. In addition, an untruncated MLA region is present. These features correspond well with verified protease-type AEPs (Fig. 5). It must be noted that our proposal for the vibe AEPs to be either ligase or protease-type are subject to subsequent isolation or expression of the enzymes to facilitate concluding functional studies. Based on proposed ligase predictive residues, *Mco*AEP2 appears to be a protease, but, functional studies have demonstrated to the contrary, that it is in fact a ligase (Du et al., 2020). This indicates the presence of sequence elements and interactions other than known ligase determinants, such as substrate affinity that potentially influence an enzyme's function (Du et al., 2020). In addition, factors such as favorable pH may alter the behavior of an AEP to display either ligase or protease activity. For example, *Vy*AEP1 identified in *Viola yedoensis* (Hemu et al., 2019) and *Ha*AEP2 from in *Helianthus annus* (Jackson et al., 2018), act as hydrolases at low pH, but their ligase activity becomes prominent at neutral/high pH. Although in the current context of the study, the functional roles of *Vb*AEPs were not evaluated, their discovery is important and add to the growing number of AEPs identified in cyclotide bearing plants. Under the conditions *Vb*AEPs act as ligase or proteases, whether they are involved in cyclotide biosynthesis and whether they have potential biotechnological use still remains to be shown in future studies.

Transcriptomic analysis is practical in the context of studying endangered plant species when access to large-scale plant material is not feasible. The cost of transcriptomic analysis has also significantly reduced in the past decade making this technique affordable and accessible. Another technical advantage of transcriptomic analysis is, it removes the ambiguity of isobaric residues, Ile and Leu in the translated peptide sequences, which is usually not distinguishable in peptide sequencing approaches based on MS/MS. A future development from this work will be to synthesise selected cyclotides identified in the transcriptome for bioactivity screening studies. In addition, studies aimed at investigating the functional role of potential cyclisation enzyme identified in the transcriptome, will be useful to gain insight into the cyclotide biogenesis in *V. betonicifolia*. We encountered many difficulties with obtaining quality RNA due to problems associated with storage and transportation of the material under heat stable conditions. In hindsite, the contribution of peptide level analysis cannot be discounted and it remains an important part of cyclotide discovery work.

2.3. Transcriptome mining resulted in two sequences similar to known PDIs involved in oxidative folding of cyclotides

Enzymes presumably involved in introducing disulfides (oxidative folding) and providing quality control by catalysing the rearrangement of incorrect disulfides (isomerase activity) of cyclotides, known as protein disulfide isomerases (PDIs) are reported from cyclotide producing Rubiaceae and Violaceae families. Blast search of *V. betonicifolia* transcriptome for sequences with high homology to *Oldenladia affinis* PDI full sequence (*Oa*PDI) (Gruber et al., 2007) and *Gleospermum blakeanum* PDI partial sequence (*Gb*PDI) (Burman et al., 2010) resulted in two closely matching hits (Fig. 6 and Supplementary Fig. S2). *Vb*PDI1 matched *Oa*PDI with 74.6% sequence homology (356 matching residues of total 477 residues) and *Vb*PDI2 matched *Gb*PDI with 82.1% sequence homology (69 matching residues of total 84 residues). These results show that the two enzymes *Vb*PDI1-2 are candidate enzymes likely involved in oxidative folding of vibe cyclotides.

3. Conclusions

This work highlights that transcriptomic analysis, in combination with analysis at peptide level, is an efficient and powerful tool to accelerate the rate of new cyclotide discovery. Using this method, 21 new cyclotides, four new acyclotides and three known cyclotide sequences (varv A/kalata S, viba 17, viba 11) were discovered from *V. betonicifolia*. The presence of cyclotides in Viola species from Sri Lanka supports the notion of ubiquitous occurrence of cyclotides in Violaceae. Discovery of new AEPs with presumably ligase activity are important for potential biotechnological applications ranging from *in vitro* peptide engineering to plant molecular farming. These AEP enzymes, in combination with PDI enzymes and vibe cyclotide precursor sequences are important tools to gain insight into how cyclotides are potentially processed from linear precursors. The novel cyclotide sequences identified in the current study validates that Violaceae is a rich source of diverse cyclotides, fortifying the plasticity of the scaffold and possible sequence variability that can be accommodated within the cyclotide scaffold.

4. Experimental

4.1. General experimental procedures

The initial mass spectrometry was carried out on a LCQ mass spectrometer (Finnigan LCQ, San Jose, CA, USA), equipped with an electrospray ionization source operated in positive mode. Cyclotides were analysed on a UPLC-QTof mass spectrometer (Waters, Milford, MA) along with a nano LC Waters C_{18} analytical column (150 × 0.075 mm column, 5 μm) for Mass analysis. MeOH and ACN (VWR, Sweden) were used as extraction media and for chromatographic techniques. The Unipro-Ugene software package was used for all sequence alignments and the mMass software tool was used for cyclotide mass calculations.

```
OaPDI     YAPWCGHCQMLEPTYNKLGKHLRGIDSLVIAKMDGTTNEHHR--AKPDGFPTILFFPAGNKSF
VbPDI_1   YAPWCGHCQSLEPTYNKLAKHLRDVDSLVIAKMDGTTNEHPR--AKSDGFPTLLFFPAGNKSF
GbPDI     YAPWCGHCKKLAPILDEVASSYQSDADVVIAKLDATANDYPTDTFEVQGYPTMYFRSASG---
VbPDI_2   YAPWCGHCKKLAPILDEVASSFLKEADVVIAKLDATTNDYPTDIFDVQGYPTLYFISSSG---
          ********: * *  :::::      .:****:*.*:*::     . :*:**: * .:..

OaPDI     DPIAFDGDRTVVELYKFLKKHATHPFKIQKPATSS-----PQTKGSGVSQDES-STSKDLKDEL
VbPDI_1   DPITVDTDRTVVAFYKFLKKNAAIPFKLQKPASATKPETKPETSESKGSDEKSGSSSSEAKDEL
GbPDI     NLVQYDGDRTKEAIIEFIEKNR--------------------DKV-------AQQEQEPAKDEL
VbPDI_2   ALVQYDGERTKEHIIEFIENNR--------------------NKAPEAVEQEAEAEAETAKDEL
            :  * :**   : :*::::                        .       :    .  ****
```

Fig. 6. Multiple sequence alignment for the catalytic domain of *V. betonicifolia* protein disulfide isomerases, *Vb*PDI1-2 with previously reported PDIs from Rubiaceae (*Oa*PDI) and Violaceae (*Gb*PDI). The active site residues CGHC are highlighted.

4.2. Plant material

Viola betonicifolia Sm. from Violaceae family was collected from Horton Plains national park (6°49′5″N 80°50′33″E) in Sri Lanka. The plant specimens were authenticated (by N. P. T. Gunawardena of the National Herbarium of the Royal Botanical Garden, Peradeniya) and voucher specimens were deposited at the National Herbarium of the Royal Botanical Garden, Peradeniya with a collection number HP16. Permission for plant material collection was obtained from the Department of Wildlife Conservation (permit no-WL/3/2/48/15).

4.3. Small scale extraction and isolation of crude cyclotides

Dried samples (leaves, stem and root) were crushed into powder using mortar and pestle. About 250 mg of pulverised sample was extracted with 4 mL of 30% acetonitrile (ACN) containing 0.1% formic acid (FA) in water (2 h shaking). Following centrifugation, the supernatant was collected and the remaining plant material was extracted further with 4 mL of 60% ACN containing 0.1% FA in water, and then 4 mL of 100% ACN containing 0.1% FA in water.

The total extract was collected and diluted to 10% ACN with 0.1% FA in water. The diluted extract was loaded on SPE cartridges (ISOLUTE C_{18}, 500 mg, 10 mL) and the hydrophilic compounds were washed with 10 mL of 10% ACN containing 0.1% FA solution. Hydrophobic compounds were eluted with 5 mL of the 60% ACN containing 0.1% FA in water. The collected fractions were freeze dried and redissolved with 30% ACN in water and then subjected to gel filtration using PD-10 column (GE Health care). The high molecular weight fractions were collected and then diluted with 0.05% FA in water for LC-MS analysis.

4.3.1. Reduction, alkylation and enzymatic cleavage of cyclotides

Partially-purified, freeze dried cyclotide fractions were reduced with dithiothreitol in 0.25 M Tris-HCl (pH 8.5) containing 4 mM EDTA and 8 M guanidine-HCl, incubated at 37 °C in the dark and under N_2 for 2 h. Iodoacetamide (50 mg, in 0.5 M Tris-HCl, 2 mM EDTA) was added to alkylate free thiols. The S-alkylation was quenched after 10 min by adding 250 μL of 0.5 M citric acid. The reduced and alkylated peptides were purified by size exclusion chromatography (PD-10) and cleaved with endoproteinase GluC (37 °C, 4 h) in 50 mM NH_4HCO_3 buffer (pH 7.8).

4.4. Liquid chromatography/mass spectrometry (LC-MS) analysis

The fractions identified to contain cyclotides were analysed using ultra performance liquid chromatography coupled to quadrupole time-of-flight mass spectrometry (nanoAcquity UPLC/QTof Waters, Milford, MA). Samples were eluted using a gradient of ACN in FA (1–90% ACN over 50 min). A nanoLC column (Waters BEH, 75 μm (i.d) × 150 mm) operated at 0.3 μL/min flow rate was used. The capillary temperature was set at 220 °C and the nanoelectrospray operated at 4 kV. The mass spectrum was obtained in positive ion mode using a mass window of 300–2000 *m/z*. LC-MS chromatogram and MS spectral data were processed through MassLynx V4.1 (Waters, Milford, MA).

4.5. Transcriptome de novo sequencing of Viola betonicifolia

Whole RNA extraction, sequencing and analysis were performed according to the method described in the literature (Park et al., 2017). The plant tissues were pooled from all of the plant's major organs, *i.e.*, the roots, stems, flowers, and leaves. The collected tissues were immediately frozen with liquid nitrogen and powdered using an autoclaved mortar and pestle. For RNA isolation, a small plant piece was stored in RNA later solution (Ambion, Thermo fisher) at −20 °C. The sequencing was performed by an external service provider (Macrogen Inc., Seoul, South Korea). The tissues were directly extracted by RNeasy Plant Mini Kit, (Qiagen, Hilden, Germany) according to the manufacturers' protocols. Quality and quantity of RNA were measured using an Agilent 2100 Bioanalyser (Agilent Technologies, Santa Clara, CA, USA) with an RNA Integrity Number (RIN) and rRNA ratio. The cDNA library was prepared using a TruSeq preparation kit according to the manufacturer's instructions (Illumina, San Diego, U.S.A). The sequencing was performed using Illumina HiSeq 3000 Sequencing System (Illumina, USA).

4.6. Analysis and assembling of sequencing data

FASTQC (version 0.11.3) was used to determine the quality of RNA sequencing data. *De novo* assembly was performed using the Trinity platform (Grabherr et al., 2011). Summary of the detailed information on the *de novo* transcriptome assembly for *V. betonicifolia* given in Supplementary Table S1.

The assembled transcriptome was searched for sequences similar to cyclotides from Cybase (cut-off date: March 03, 2020; http://www.cybase.org.au/) (Mulvenna et al., 2006) by the standalone NCBI-blast + service (tblastn, E-value cutoff: 20) in the Ugene software package (v.1.31.0) (Okonechnikov et al., 2012). An additional search was done using a motif search (Sigrist et al., 2002) in Fuzzpro application of EMBOSS (v. 5.0.0) (Rice et al., 2000). Sequences containing cyclotides were extracted and the result file was further processed by manual inspection after ClustalX alignment. The sequence was assumed to be a cyclotide precursor if the sequence contained the six conserved cysteines aligned with previously known cyclotides, and if the conserved N-terminal domain showed sequence similarity to known cyclotide precursors.

4.6.1. Transcriptome mining for cyclisation and oxidative folding enzymes

The assembled *V. betonicifolia* transcriptome was searched for asparagine endopeptidase (AEP) enzymes similar to known AEPs involved in cyclic peptide maturation – *Vy*PAL2, *He*AEP3, butelase 1, OaAEP1, VyAEP1, MCoAEP2 and HaAEP1 (sequences reported in GenBank with respective ID numbers: QCW05335, AWD84474, KF918345, KR259377, QCW05330, QFR54168, 6AZT_A). Similarly, *V. betonicifolia* transcriptome was searched for oxidative folding enzymes belonging to protein-disulfide isomerase (PDI) class, *Oldenlandia affinis* PDI (*Oa*PDI) and *Gleospermum blakeanum* PDI (*Gb*PDI) (GenBank accession numbers: EF611425 and GQ443304 respectively). The enzyme sequences obtained from NCBI protein database were searched against *V. betonicifolia* transcriptome using Unipro Ugene tool. The gene that matched in highest percentage in *V. betonicifolia* transcriptome was then translated to protein sequences. The enzyme sequences (*Vb*AEP1-4) and *Vb*PDI1-2 were identified and aligned using ClustalW multiple sequence alignment in the Unipro Ugene software. In all searches *Vb*AEP1-4 were the only hits that matched in highest percentages with the reported AEPs (Supplementary Table S5). Similarly, *Vb*PDI1-2 were identified as oxidative folding enzymes showing highest homology to the previously reported PDIs (Supplementary Figure S2).

4.7. Accession numbers

The accession numbers of Sequence Read Archive (SRA) and Transcriptome Shotgun Assembly (TSA) for the transcriptome of *V. betonicifolia* project deposited at DDBJ/EMBL/GenBank are *i)* TSA: GIYB00000000 and *ii)* SRA: PRJNA687650/SRR13316929).

Following accession numbers have been obtained, for nucleotide and translated sequences of *V. betonicifolia* AEPs: *Vb*AEP1- MW429035, VbAEP2- MW429036, VbAEP3- MW429037, VbAEP4- MW429038. The accession numbers for *V. betonicifolia* PDIs are, *Vb*PDI1-MW792069 and *Vb*PDI2-MW792070.

4.8. Sequence alignment for phylogenetic analysis

A total of 28 precursor sequences were prepared as DNA sequence for the sequence alignment, and the sequences included only three domains

of the precursor, i.e., NTPP, NTR, and cyclotide domains. To guide the alignment of DNA sequences of the precursors, they were translated into protein sequences and aligned independently for each molecular species using ClustalW in the Ugene software package. Those alignments were in turn combined and realigned (*i.e.*, keeping the indel positions from the alignment of each of the molecular species). The resulting protein-guided alignment of nucleotide sequences was manually adjusted within reading frames.

Author contributions

CH, UG and SG designed the study. SR and BS performed all the experiments and contributed to analysis and interpretation of data. SR, BS, AS and SG drafted the manuscript. All authors participated in revising and finalising of the manuscript.

Declaration of competing interest

The authors declare no competing financial interest.

Acknowledgments

This work was supported by the Swedish Research Council Linkage grant (2013–06672), led by SG. Participation of BS in the research study was made possible through funding from W. Szafer Institute of Botany, Polish Academy of Sciences; the Polish National Science Center (2017/26/D/NZ8/00658 and 2017/01/X/NZ8/00606). We would like to thank Dr. Scott Jarmusch for proof-reading the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.phytochem.2021.112749.

References

Bhatt, V.P., Negi, G.C.S., 2006. Ethnomedicinal plant resources of jaunsari tribe of garhwal himalaya, Uttaranchal. Indian J. Tradit. Knowl. 5, 331–335.

Broussalis, A.M., Göransson, U., Coussio, J.D., Ferraro, G., Martino, V., Claeson, P., 2001. First cyclotide from Hybanthus (Violaceae). Phytochemistry 58, 47–51. https://doi.org/10.1016/S0031-9422(01)00173-X.

Burman, R., Gruber, C.W., Rizzardi, K., Herrmann, A., Craik, D.J., Gupta, M.P., Göransson, U., 2010. Cyclotide proteins and precursors from the genus Gloeospermum: filling a blank spot in the cyclotide map of Violaceae. Phytochemistry 71, 13–20. https://doi.org/10.1016/j.phytochem.2009.09.023.

Burman, R., Strömstedt, A.A., Malmsten, M., Göransson, U., 2011. Cyclotide-membrane interactions: defining factors of membrane binding, depletion and disruption. Biochim. Biophys. Acta Biomembr. 1808, 2665–2673. https://doi.org/10.1016/j.bbamem.2011.07.004.

Burman, R., Yeshak, M.Y., Larsson, S., Craik, D.J., Rosengren, K.J., Göransson, U., 2015. Distribution of circular proteins in plants: large-scale mapping of cyclotides in the Violaceae. Front. Plant Sci. 6 https://doi.org/10.3389/fpls.2015.00855.

Chandra, D., Kohli, G., Prasad, K., Bisht, G., Punetha, V.D., Khetwal, K.S., Devrani, M.K., Pandey, H.K., 2015. Phytochemical and ethnomedicinal uses of family Violaceae. Curr. Res. Chem. 7, 44–52. https://doi.org/10.3923/crc.2015.44.52.

Craik, D.J., Daly, N.L., Bond, T., Waine, C., 1999. Plant cyclotides: a unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. J. Mol. Biol. 294, 1327–1336.

Daly, N.L., Clark, R.J., Plan, M.R., Craik, D.J., 2006. Kalata B8, a novel antiviral circular protein, exhibits conformational flexibility in the cystine knot motif. Biochem. J. 393, 619–626. https://doi.org/10.1042/BJ20051371.

Du, J., Yap, K., Chan, L.Y., Rehm, F.B.H., Looi, F.Y., Poth, A.G., Gilding, E.K., Kaas, Q., Durek, T., Craik, D.J., 2020. A bifunctional asparaginyl endopeptidase efficiently catalyzes both cleavage and cyclization of cyclic trypsin inhibitors. Nat. Commun. 11 https://doi.org/10.1038/s41467-020-15418-2.

Dutton, J.L., Renda, R.F., Waine, C., Clark, R.J., Daly, N.L., Jennings, C.V., Anderson, M. A., Craik, D.J., 2004. Conserved structural and sequence elements implicated in the processing of gene-encoded circular proteine. J. Biol. Chem. 279, 46858–46867. https://doi.org/10.1074/jbc.M407421200.

Gillon, A.D., Saska, I., Jennings, C.V., Guarino, R.F., Craik, D.J., Anderson, M.A., 2008. Biosynthesis of circular proteins in plants. Plant J. 53, 505–515. https://doi.org/10.1111/j.1365-313X.2007.03357.x.

Göransson, U., Luijendijk, T., Johansson, S., Bohlin, L., Claeson, P., 1999. Seven novel macrocyclic polypeptides from viola arvensis. J. Nat. Prod. 62, 283–286.

Göransson, U., Herrmann, A., Burman, R., Haugaard-Jönsson, L.M., Rosengren, K.J., 2009. The conserved glu in the cyclotide cycloviolacin O2 has a key structural role. Chembiochem 10, 2354–2360. https://doi.org/10.1002/cbic.200900342.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883.

Gran, Lorents, 1973. On the effect of a polypeptide isolated from "kalata-kalata" (Oldenlandia affinis DC) on the oestrogen dominated uterus. Basic Clin. Pharmacol. Toxicol. 33, 400–408.

Gruber, C.W., 2010. Global cyclotide adventure: a journey dedicated to the discovery of circular peptides from flowering plants. Biopolymers 94, 565–572. https://doi.org/10.1002/bip.21414.

Gruber, C.W., 2014. Physiology of invertebrate oxytocin and vasopressin neuropeptides. Exp. Physiol. 99, 56–61. https://doi.org/10.1113/expphysiol.2013.072561.

Gruber, C.W., Čemažar, M., Clark, R.J., Horibe, T., Renda, R.F., Anderson, M.A., Craik, D. J., 2007. A novel plant protein-disulfide isomerase involved in the oxidative folding of cystine knot defense proteins. J. Biol. Chem. 282, 20435–20446. https://doi.org/10.1074/jbc.M700018200.

Gruber, C.W., Elliott, A.G., Ireland, D.C., Delprete, P.G., Dessein, S., Göransson, U., Trabi, M., Wang, C.K., Kinghorn, A.B., Robbrecht, E., Craik, D.J., 2008. Distribution and evolution of circular miniproteins in flowering plants. Plant Cell 20, 2471–2483. https://doi.org/10.1105/tpc.108.062331.

Gustafson, K.R., Sowder, R.C., Henderson, L.E., Parsons, I.C., Kashman, Y., Cardellina, J. H., McMahon, J.B., Robert W, B.J., Pannell, L.K., Boyd, M.R., 1994. Circulins A and B. Novel human immunodeficiency virus (HIV)-inhibitory macrocyclic peptides from the tropical tree Chassalia parvifolia. J. Am. Chem. Soc. 116, 9337–9338.

Gustafson, K.R., Walton, L.K., Sowder, R.C., Johnson, D.G., Pannell, L.K., Cardellina, J. H., Boyd, M.R., 2000. New circulin macrocyclic polypeptides from Chassalia parvifolia. J. Nat. Prod. 63, 176–178. https://doi.org/10.1021/np990432r.

Harris, K.S., Durek, T., Kaas, Q., Poth, A.G., Gilding, E.K., Conlan, B.F., Saska, I., Daly, N. L., van der Weerden, N.L., Craik, D.J., Anderson, M.A., 2015. Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. Nat. Commun. 6, 10199. https://doi.org/10.1038/ncomms10199.

Haywood, J., Schmidberger, J.W., James, A.M., Nonis, S.G., Sukhoverkov, K.V., Elias, M., Bond, C.S., Mylne, J.S., 2018. Structural basis of ribosomal peptide macrocyclization in plants. Elife 7. https://doi.org/10.7554/eLife.32955.

Hellinger, R., Koehbach, J., Soltis, D.E., Carpenter, E.J., Wong, G.K.S., Gruber, C.W., 2015. Peptidomics of circular cysteine-rich plant peptides: analysis of the diversity of cyclotides from viola tricolor by transcriptome and proteome mining. J. Proteome Res. 14, 4851–4862. https://doi.org/10.1021/acs.jproteome.5b00681.

Hemu, X., Sahili, A. El, Hu, S., Wong, K., Chen, Y., Wong, Y.H., Zhang, X., Serra, A., Goh, B.C., Darwis, D.A., Chen, M.W., Sze, S.K., Liu, C.F., Lescar, J., Tam, J.P., 2019. Structural determinants for peptide-bond formation by asparaginyl ligases. Proc. Natl. Acad. Sci. U.S.A. 116, 11737–11746. https://doi.org/10.1073/pnas.1818568116.

Hernandez, J.F., Gagnon, J., Chiche, L., Nguyen, T.M., Andrieu, J.P., Heitz, A., Hong, T. T., Pham, T.T.C., Le Nguyen, D., 2000. Squash trypsin inhibitors from Momordica cochinchinensis exhibit an atypical macrocyclic structure. Biochemistry 39, 5722–5730. https://doi.org/10.1021/bi9929756.

Ireland, D.C., Colgrave, M.L., Craik, D.J., 2006. A novel suite of cyclotides from Viola odorata: sequence variation and the implications for structure, function and stability. Biochem. J. 400, 1–12. https://doi.org/10.1042/BJ20060627.

Jackson, M.A., Gilding, E.K., Shafee, T., Harris, K.S., Kaas, Q., Poon, S., Yap, K., Jia, H., Guarino, R., Chan, L.Y., Durek, T., Anderson, M.A., Craik, D.J., 2018. Molecular basis for the production of cyclic peptides by plant asparaginyl endopeptidases. Nat. Commun. 2411 https://doi.org/10.1038/s41467-018-04669-9.

Jennings, C., West, J., Waine, C., Craik, D., Anderson, M., 2001. Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from Oldenlandia affinis. Proc. Natl. Acad. Sci. U.S.A. 98, 10614–10619. https://doi.org/10.1073/pnas.191366898.

Keov, P., Liutkevičiute, Z., Hellinger, R., Clark, R.J., Gruber, C.W., 2018. Discovery of peptide probes to modulate oxytocin-type receptors of insects. Sci. Rep. 8 https://doi.org/10.1038/s41598-018-28380-3.

Koehbach, J., Attah, A.F., Berger, A., Hellinger, R., Kutchan, T.M., Carpenter, E.J., Rolf, M., Sonibare, M.A., Moody, J.O., Wong, G.K.-S., Dessein, S., Greger, H., Gruber, C.W., 2013. Cyclotide discovery in Gentianales revisited-identification and characterization of cyclic cystine-knot peptides and their phylogenetic distribution in Rubiaceae plants. Biopolymers 100, 438–452. https://doi.org/10.1002/bip.22328.

Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132. https://doi.org/10.1016/0022-2836(82)90515-0.

Lindholm, P., Göransson, U., Johansson, S., Claeson, P., Gullbo, J., Larsson, R., Bohlin, L., Backlund, A., 2002. Cyclotides: a novel type of cytotoxic agents. Mol. Canc. Therapeut. 1, 365–369.

Ministry of Mahaweli Development and Environment, 2016. Sri Lanka species data base [WWW Document]. Clear. house Mech. Sri Lanka (accessed 4.18.19). http://lk.chm-cbd.net.

Mulvenna, J.P., Wang, C., Craik, D.J., 2006. CyBase: a database of cyclic protein sequence and structure. Nucleic Acids Res. 34, D192–D194. https://doi.org/10.1093/nar/gkj005.

Narayani, M., Chadha, A., Srivastava, S., 2017. Cyclotides from the Indian medicinal plant *Viola odorata* (banafsha): identification and characterization. J. Nat. Prod. 80, 1972–1980. https://doi.org/10.1021/acs.jnatprod.6b01004.

Nguyen, G.K.T., Zhang, S., Nguyen, N.T.K., Nguyen, P.Q.T., Chiu, M.S., Hardjojo, A., Tam, J.P., 2011. Discovery and characterization of novel cyclotides originated from chimeric precursors consisting of albumin-1 chain a and cyclotide domains in the fabaceae family. J. Biol. Chem. 286, 24275–24287. https://doi.org/10.1074/jbc.M111.229922.

Nguyen, G.K.T., Lim, W.H., Nguyen, P.Q.T., Tam, J.P., 2012. Novel cyclotides and uncyclotides with highly shortened precursors from chassalia chartacea and effects of methionine oxidation on bioactivities. J. Biol. Chem. 287, 17598–17607. https://doi.org/10.1074/jbc.M111.338970.

Nguyen, G.K.T., Lian, Y., Pang, E.W.H., Nguyen, P.Q.T., Tran, T.D., Tam, J.P., 2013. Discovery of linear cyclotides in monocot plant Panicum laxum of Poaceae family provides new insights into evolution and distribution of cyclotides in plants. J. Biol. Chem. 288, 3370–3380. https://doi.org/10.1074/jbc.M112.415356.

Nguyen, G.K.T., Wang, S., Qiu, Y., Hemu, X., Lian, Y., Tam, J.P., 2014. Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. Nat. Chem. Biol. 10, 732–738. https://doi.org/10.1038/nchembio.1586.

Niedermeyer, T.H.J., Strohalm, M., 2012. mMass as a software tool for the annotation of cyclic peptide tandem mass spectra. PloS One 7, e44913. https://doi.org/10.1371/journal.pone.0044913.

Okonechnikov, K., Golosova, O., Fursov, M., Varlamov, A., Vaskin, Y., Efremov, I., German Grehov, O.G., Kandrov, D., Rasputin, K., Syabro, M., Tleukenov, T., 2012. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics. https://doi.org/10.1093/bioinformatics/bts091.

Park, S., Yoo, K.-O., Marcussen, T., Backlund, A., Jacobsson, E., Rosengren, K.J., Doo, I., Göransson, U., 2017. Cyclotide evolution: insights from the analyses of their precursor sequences, structures and distribution in violets (viola). Front. Plant Sci. 8, 1–19. https://doi.org/10.3389/fpls.2017.02058.

Plan, M.R.R., Göransson, U., Clark, R.J., Daly, N.L., Colgrave, M.L., Craik, D.J., 2007. The cyclotide fingerprint in Oldenlandia affinis: elucidation of chemically modified, linear and novel macrocyclic peptides. Chembiochem 8, 1001–1011. https://doi.org/10.1002/cbic.200700097.

Poth, A.G., Mylne, J.S., Grassl, J., Lyons, R.E., Millar, A.H., Colgrave, M.L., Craik, D.J., 2012. Cyclotides associate with leaf vasculature and are the products of a novel precursor in petunia (Solanaceae). J. Biol. Chem. 287, 27033–27046. https://doi.org/10.1074/jbc.M112.370841.

Pranting, M., Loov, C., Burman, R., Goransson, U., Andersson, D.I., 2010. The cyclotide cycloviolacin O2 from Viola odorata has potent bactericidal activity against Gram-negative bacteria. J. Antimicrob. Chemother. 65, 1964–1971. https://doi.org/10.1093/jac/dkq220.

Ravipati, A.S., Henriques, S.T., Poth, A.G., Kaas, Q., Wang, C.K., Colgrave, M.L., Craik, D. J., 2015. Lysine-rich cyclotides: a new subclass of circular knotted proteins from Violaceae. ACS Chem. Biol. 10, 2491–2500. https://doi.org/10.1021/acschembio.5b00454.

Ravipati, A.S., Poth, A.G., Troeira Henriques, S., Bhandari, M., Huang, Y.H., Nino, J., Colgrave, M.L., Craik, D.J., 2017. Understanding the diversity and distribution of cyclotides from plants of varied genetic origin. J. Nat. Prod. 80, 1522–1530. https://doi.org/10.1021/acs.jnatprod.7b00061.

Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. 16, 276–277.

Schoepke, T., Hasan Agha, M.I., Kraft, R., Otto, A., Hiller, K., 1993. Haemolytisch aktive Komponenten aus Viola tricolor L. und Viola arvensis murray. Sci. Pharm. 61, 145–153.

Schöpke, T., Hasan, A.M., Kraft, R., Otto, A., Hiller, K., 1993. Hämolytisch aktive komponenten aus Viola tricolor L. und Viola arvensis Murray. Sci. Pharm. 61, 145–153.

Seydel, P., Gruber, C.W., Craik, D.J., Dörnenburg, H., 2007. Formation of cyclotides and variations in cyclotide expression in Oldenlandia affinis suspension cultures. Appl. Microbiol. Biotechnol. 77, 275–284. https://doi.org/10.1007/s00253-007-1159-6.

Sigrist, C.J. a, Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., Bairoch, A., Bucher, P., 2002. PROSITE: a documented database using patterns and profiles as motif descriptors. Briefings Bioinf. 3, 265–274. https://doi.org/10.1093/bib/3.3.265.

Sims, P.A., 2010. Use of a spreadsheet to calculate the net charge of peptides and proteins as a function of pH: an alternative to using "canned" programs to estimate the isoelectric point of these important biomolecules. J. Chem. Educ. 87, 803–808. https://doi.org/10.1021/ed100232j.

Slazak, B., Jacobsson, E., Kuta, E., Göransson, U., 2015. Exogenous plant hormones and cyclotide expression in Viola uliginosa (Violaceae). Phytochemistry 117, 527–536. https://doi.org/10.1016/J.PHYTOCHEM.2015.07.016.

Slazak, B., Kapusta, M., Malik, S., Bohdanowicz, J., Kuta, E., Malec, P., Göransson, U., 2016. Immunolocalization of cyclotides in plant cells, tissues and organ supports their role in host defense. Planta 244, 1029–1040. https://doi.org/10.1007/s00425-016-2562-y.

Slazak, B., Kapusta, M., Strömstedt, A.A., Słomka, A., Krychowiak, M., Shariatgorji, M., Andrén, P.E., Bohdanowicz, J., Kuta, E., Göransson, U., 2018. How does the sweet violet (viola odorata L.) fight pathogens and pests – cyclotides as a comprehensive plant host defense system. Front. Plant Sci. 9, 1296. https://doi.org/10.3389/fpls.2018.01296.

Sletten, K., Gran, L., 1973. Some molecular properties of kalatapeptide B1. Medd. Nor. Farm. Selsk. 35, 69–82.

Strömstedt, A.A., Park, S., Burman, R., Göransson, U., 2017. Bactericidal activity of cyclotides where phosphatidylethanolamine-lipid selectivity determines antimicrobial spectra. Biochim. Biophys. Acta Biomembr. 1859, 1986–2000. https://doi.org/10.1016/J.BBAMEM.2017.06.018.

Tam, J.P., Lu, Y.-A., Yang, J.-L., Chiu, K.-W., 1999. An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides. Proc. Natl. Acad. Sci. Unit. States Am. 96, 8913–8918. https://doi.org/10.1073/pnas.96.16.8913.

Trabi, M., Svangård, E., Herrmann, A., Göransson, U., Claeson, P., Craik, D.J., Bohlin, L., 2004. Variations in cyclotide expression in Viola species. J. Nat. Prod. 67, 806–810. https://doi.org/10.1021/np034068e.

Wahlert, G.A., Marcussen, T., de Paula-Souza, J., Feng, M., Ballard, H.E., 2014. A phylogeny of the Violaceae (malpighiales) inferred from plastid DNA sequences: implications for generic diversity and intrafamilial classification. Syst. Bot. 39, 239–252. https://doi.org/10.1600/036364414X678008.

Wang, C.K.L., Kaas, Q., Chiche, L., Craik, D.J., 2008. CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering. Nucleic Acids Res. 36, D206–D210. https://doi.org/10.1093/nar/gkm953.

Wang, C.K.L., Clark, R.J., Harvey, P.J., Johan Rosengren, K., Cemazar, M., Craik, D.J., 2011. The role of conserved glu residue on cyclotide stability and activity: a structural and functional study of kalata B12, a naturally occurring glu to asp mutant. Biochemistry 50, 4077–4086. https://doi.org/10.1021/bi2004153.

Wijesundara, S., Kathriarachchi, H.S., Ranasinghe, S.W., Hapuarachchi, G., 2012. The national red list 2012 of Sri Lanka. In: Weerakoon, D.K., Wijesundara, S. (Eds.), Conservation Status of the Fauna and Flora. Ministry of Environment, Colombo., pp. 340–345

Yeshak, M.Y., Burman, R., Asres, K., Göransson, U., 2011. Cyclotides from an extreme habitat: characterization of cyclic peptides from viola abyssinica of the ethiopian highlands. J. Nat. Prod. 74, 727–731.

Zahoor Husain, S., Naseem Malik, R., Javaid, M., Bibi, S., 2008. Ethonobotanical properties and uses of medicinal plants of Morgah biodiversity park. Rawalpindi. Pak. J. Bot 40, 1897–1911.