Identification of cysteine knot peptides in four Philippine plant species using mass spectrometry and transcriptome analysis

Bea A. Crisostomo*[1], Arnelson Arwin G. Atis[2], Denise Noelle R. Mirano-Bascos[1], and Christine C. Hernandez[2]

[1]National Institute of Molecular Biology and Biotechnology, College of Science, University of the Philippines Diliman, Quezon City, Philippines
[2]Institute of Chemistry, College of Science, University of the Philippines Diliman, Quezon City, Philippines

Knottins are small proteins with at least three disulfide bridges that form a cysteine knot, giving them exceptional stability. These molecules are candidates for developing peptide-based drugs that are envisioned to fill the gap between small-molecule drugs and biologics as they are bioavailable, like small-molecule drugs, and target-specific, like biologics. Plant-derived knottins are able to preserve their stable conformation in the presence of mutations and peptide grafts, making them potential drug scaffolds. Through LC/MS screening of 185 plant extracts from 180 plant species, we identified 6 Philippine plants that produce disulfide-rich peptides structurally related to knottins. Four of these plants were further analyzed using transcriptome screening and we confirmed that *Alstonia scholaris, Wrightia pubescens* and *Tabernaemontana pandacaqui* from the Apocynaceae family produce knottin-like peptides similar to previously reported Apocynaceae alpha amylase inhibitors. Given the novelty of their sequences, it is possible that these peptides possess novel functions as well. This is the first report of cysteine knot peptides from *W. pubescens and T. pandacaqui*. Finally, we showed that LC/MS screening followed by transcriptome analysis is a more efficient workflow for identifying cysteine knot peptides as opposed to LC/MS screening alone.

*Corresponding author
Email Address: beacrisostomo@gmail.com
Date received: January 21, 2019
Date revised: April 2, 2019
Date accepted: April 10, 2019

KEYWORDS

cyclotide, cysteine knot peptide, mass spectrometry, transcriptome

INTRODUCTION

The knottin family of proteins are small cysteine-rich peptides that include conotoxins, spider toxins, scorpion toxins, and plant enzyme inhibitors. They are characterized by having at least three disulfide bridges that form a cysteine knot (CK; (Pallaghy et al. 1994). This compact structure gives them exceptional stability against thermal, chemical, and enzymatic degradation (Colgrave and Craik 2004, Ireland et al. 2006, Nguyen et al. 2014, 2015b,Loo et al. 2016). In plants, knottins are composed of about 30 amino acids with six conserved cysteines that form the CK.

Plant-derived knottins are presumed to function primarily for plant defense, as evidenced by their insecticidal (Jennings et al. 2001, 2005, Poth et al. 2011a, Pinto et al. 2012)and molluscicidal (Plan et al. 2008)activities. Beyond these, however, they exhibit a range of activities including anti-bacterial (Ovesen et al. 2011, Fensterseifer et al. 2015), anthelmintic (Colgrave et al. 2008, 2009), anti-tumor (Lindholm et al. 2002) and antiviral(Gustafson et al. 1994, Daly et al. 2004, Chen et al. 2005, Nguyen et al. 2015b) activities. While most proteins are easily degraded in the body before reaching their targets, several knottins have been shown to be orally bioavailable(Gran 1970, 1973 in Gran, Sandberg, & Sletten,

2000; Thell et al., 2016). These properties make knottins potential drug candidates and grafting scaffolds for other drugs (Gunasekera et al. 2008).

Plant knottins can be categorized based on their backbone cyclization and cysteine motif. Linear knottins have the motif C-6X-C-8X-CC-3XC-10X-C(Tam et al. 2015), similar to that of the conotoxin O superfamily (Kaas et al. 2010). On the other hand, cyclic knottins, or cyclotides, have the motif C-3X-C-4X-C-4X-C-X-C-4X-C(Tam et al. 2015), similar to some conotoxins in the P superfamily (Kaas et al. 2010). In addition to the CK motif, cyclotides have a cyclic backbone, which makes them even more stable compared to linear knottins.

The cyclotide subfamily gained a lot of attention over the past decade, leading to the development of cyclotide screening methods. The standard protein-based cyclotide screening procedure was established by Gruber and colleagues (2008). In this method, liquid chromatography-mass spectroscopy (LC/MS) is used to separate cyclotides, which are expected to have a mass range of 2500-4000 Da and elute between 25-55% acetonitrile. These potential cyclotide fractions are then subjected to reduction-alkylation reactions to confirm the presence of the 6 conserved cysteines. For each Cys, a mass shift of 58 Da is expected due to the addition of an alkyl group; a mass-shift of 348 Da is expected for a 6-Cys molecule. Finally, in order to confirm the cyclic backbone of the peptide, the reduced-alkylated cyclotides are digested with endoproteinase GluC, which targets the C-terminus of Glu residue, a conserved residue in Loop 1 of the cyclotide. Digestion with GluC results in the linearization of a cyclic peptide, which is detected as a single product with a mass-shift of 18 Da.

This protein-based method was used to screen more than 300 plant species from Rubiaceae and related families, confirming the presence of cyclotides in 22 Rubiaceae species and 12 Apocynaceae species (Gruber et al. 2008). However, a major limitation of protein-based screening is its bias towards detecting only the well-expressed cyclotides. Moreover, the isolation of individual cyclotides using this method is challenging especially for cyclotides with single amino acid differences. Alternatively, genomic or transcriptomic analyses may be used to detect poorly-expressed genes thus offering a more comprehensive characterization of the potential cyclotide and knottin repertoires of a plant species.

Genome analysis has been effectively utilized by research groups for gene mining. Zhang et al. (2015) developed a pipeline that can be used to efficiently screen genomes for cyclotide motifs. The script, CyPerl, first searches for open reading frames that are at least 50 amino acids (or 150 nucleotides) long to account for the size of the precursor protein. Next, it searches for motifs matching ''C[^C]{3,6}C[^C]{3,8}C[^C]{3,7}C[^C]C[^C]{4,7}C'' where [^C] and {n1,n2}represent the non-cysteine residues and their residue ranges, respectively. Using this method, Zhang et al. (2015) predicted 145 cyclotide analogues from 10 families, which included Brassicaceae, Fabaceae, Malvaceae, Myrtaceae, Poaceae, Rosaceae, Rutaceae, Salicaceae, Solanaceae, Vitaceae. Six of these families had not been previously reported as cyclotide-producing families.

While over 300 cyclotides have already been characterized, majority of these have been isolated from plants in the continents of Africa, Europe, the Americas, and Oceania, as curated in CyBase (Mulvenna 2006, Wang et al. 2007). Few cyclotide-producing plants have been reported from Asia and most of these were from Singapore, China, and Iran (Xiang et al. 2010, He et al. 2011, Nguyen et al. 2011a, 2011b, 2013, Hashempour et al. 2013, Zarrabi et al. 2013). The Philippines is a mega-biodiverse country and ranks fifth in the number of plant species, maintaining about 5% of the world's flora (Convention on Biological Diversity n.d.). With a wide variety of plant species, the use of herbal medicine has persisted in Filipino culture. While most studies on bioactive plant components focus on small molecule compounds and secondary metabolites, little is known about their ribosomally synthesized proteins. It is of great interest therefore to discover what unique collection of cyclotides, as well as linear knottins, the Philippine flora holds.

In this study, we used a combination of LC/MS and transcriptome-based analysis to screen for CK peptides from 180 plants. Based on the mass shifts in the LC/MS data after reduction-alkylation, six species that potentially contain cyclotides were identified. Three of these species were further analyzed using transcriptome analysis. However, the analysis revealed no expression of cyclotides in these samples. Instead, potentially novel knottin sequences were identified, suggesting that a combination of LC/MS and transcriptome analysis is a more expedient method for discovering cyclotides and the wider knottin family of peptides. These results therefore show that Philippine flora contains previously uncharacterized cysteine-rich peptides. In addition, a combined protein-based and transcriptome-based analysis should be used in future studies to quickly characterize the knottin family-content of other indigenous Philippine species.

MATERIALS AND METHODS

Sample collection

A total of 185 different plant samples belonging to 180 plant species from 55 families (see Supplementary table 1 for complete list of plant extracts) were collected. For crude extraction, fresh leaves were immediately washed and air-dried after being received. For RNA extraction, fresh leaves were immediately stored in RNAlater™ Stabilization Solution (Invitrogen™) then frozen at -20°C. Plant samples (Supplementary table 1) were obtained from the following locations in the country: UP Diliman campus, Quezon City; Mt. Kinapitpitan, Masinloc, Zambales; Mt. Samat, Bataan; Brooke's Point, Palawan. All collection activities were covered by the Wildlife Gratuitous Permit (WGP-NCR 2016-05) issued by the Department of Environment and Natural Resources.

Crude extraction

Plant samples were air-dried for 3 to 7 days, then homogenized using a blender and subsequently soaked in a 60% acetonitrile (ACN) /0.1% formic acid (FA) solution for at least 1 day. The soaked samples were then filtered and concentrated *in vacuo* until only the aqueous portion remained. The aqueous layer was further concentrated to remove any excess organic solvent, after which the aqueous layer was freeze-dried to finally obtain the crude plant extract.

The crude extracts were subjected to further SPE purification using Waters Seppak C18 cartridges using a method adapted from Poth et al. (2012). Briefly, the cartridges were solvated by rinsing with at least six (6) bed volumes of methanol, and then equilibrated by rinsing with at least six (6) bed volumes of 10% acetonitrile/0.1% formic acid. The crude extracts were then loaded onto the cartridges and washed with at least six volumes of 10% acetonitrile/0.1% formic acid solution. The peptides were then eluted out from the cartridges by using acetonitrile/0.1% formic acid solutions with the acetonitrile concentrations increasing by 10% each wash until 60%

Table 1: Top BLAST hit refers to the sequence in the database that has the highest identity (%ID) with the cyclotide mined from the transcriptome. Transcriptome refers to the plant transcriptome from which the sequence was identified. RT-PCR refers to the sample from which the sequence was validated by RT-PCR.

Code	Transcriptome*	RT-PCR*	Top BLAST hit		
			GenBank ID	Record name	% ID
CT1	CT, AS, TP, WP	CT	ALL96773.1	albumin 1 [C. ternatea]	98
CT2	CT, AS, TP, WP	CT	ALL96758.1/ AML32979.1	cliotide T32 (partial)/ cliotide T19, partial [C. ternatea]	100
CT3	CT	CT	AML33003.1	cliotide T44, partial [C. ternatea]	100
CT4	CT	CT	AML32981.1	cliotide T22 (partial)	100
CT5	CT, AS	CT	P86899.1	cyclotide cter-M	100
CT6	CT, AS, TP, WP	CT	ALL96759.1	albumin 1 [C. ternatea]	100
CT7	CT, AS, TP	CT	ALL96763.1/ AML33004.1	albumin 1 [C. ternatea]/cliotide T45(partial)	100
CT8	CT	nv	AML32975.1	cliotide T13 (partial)	100
CT9	CT	CT	ALL96751.1	albumin 1 [C. ternatea]	99
CT10	CT, TP	CT	AML32991.1	cliotide T32, partial [C. ternatea]	100
CT11	CT	CT	ALL96774.1/ AML32988.1	albumin 1, partial [C. ternatea]/cliotide T29	100
CT12	CT	CT	AML32992.1	cliotide T33, partial	98
CT13	CT	CT	P86841.2	cyclotide cter-A	96
CT14	CT	CT	G1CWH1.1	cliotide T2	100
CT15	CT	CT	ALL96776.1	albumin 1 [C. ternatea]	100
CT16	CT	CT	AML32997.1	cliotide T38, partial	99
CT17	CT	nv	AML32995.1	cliotide T36, partial [C. ternatea]	100
CT18	CT, TP, WP	CT	G1CWH0.2	cliotide T1	100
CT19	CT, TP, WP	CT	G1CWH4.2	cliotide T5	100

acetonitrile/0.1% formic acid is used as the final washing solvent. Not all extracts were purified by SPE since the peptides from some extracts eluted out of the cartridges at 10% acetonitrile, which was supposed to elute only the highly polar contaminants.

Cyclotide screening using LC-MS

The cyclotide screening procedure was adapted from the protocol of Gruber et al. (2008) and Poth etal.(2012). Cyclotides usually elute between 25 to 55% acetonitrile in RP-HPLC and have a mass-to-charge (m/z) of approximately 1000-3000 depending on the overall charge of the molecule, which corresponds to a mass of between 2500-4000 Da. Crude extracts were introduced into a UPLC Acquity Class-H (Waters) UPLC system, directly connected to a Xevo G2-XS QToF mass spectrometer (Waters). Reverse phase separation of analytes was done using a linear gradient comprising of Solvent A (water + 0.1% formic acid) and B (acetonitrile + 0.1% formic acid) at a flow rate of 0.35 mL/min, which was applied to an Acquity UPLC HSS T3 column (Waters) with a particle size of 1.8 µm, and dimensions of 2.1 mm x 100 mm. Before every run, the instrument was calibrated using a NaI solution. Leu-enkephalin was used as an internal standard (Lockmass for Waters instruments) to ensure measurement accuracy.

LC/MS data were obtained and analyzed using Masslynx 4.1 software (Waters). The data-dependent acquisition (Fast DDA) mode of the Masslynx 4.1 console was also used to collect spectra over 500-3500 m/z for MS and 50-3500 m/z for the MS/MS for all crude extracts. The Fast DDA mode allows us to focus the MS/MS analyses on a single peak from the LC/MS spectra that fits certain criteria that are set before the run, essentially eliminating the need for a pure sample to be analyzed using MS/MS. The following parameters were used for fast DDA: polarity: positive mode; analyzer: sensitivity mode; capillary voltage: 3.00 kV; mass range (MS): 500-3500 m/z; peak detection: by charge state (+2, +3, +4, +5); collision energy: ramp mode (15-30 eV for low mass and 70-90 eV for high mass).

To confirm the presence of the six cysteine (Cys) residues conserved in all cyclotides, the crude extracts which exhibited potential cyclotide presence based on LC/MS data were reduced and alkylated. Briefly, approximately 1 mg of the freeze-dried extract was re-dissolved in 150 µLof 100 mM ammonium bicarbonate (Sigma-Aldrich) then reduced by incubating with15µL of 100 mM dithiothreitol (DTT; Sigma-Aldrich) at 60°C for 30 min under nitrogen (N_2) gas. The reduced cyclotide was then alkylated (specifically, carbamidomethylated) by incubating with15µL of 250 mM iodoacetamide (Sigma-Aldrich) for 60 min at room temperature. Cyclotides or peptides with six oxidized Cys residues show a distinct mass shift of 348 Da after reduction and alkylation, corresponding to the addition of a carbamidomethyl residue (58 Da) on each of the 6 Cys residues (which is conserved in all cyclotides).

RNA extraction

Total RNA was extracted from the leaves of the target plants using TRIzol™ reagent (Invitrogen™) at a ratio of 50 mg sample:1 mLTRIzol™. Briefly, leaf samples were removed from the RNAlater™ solution, weighed in sterile tubes, snap-frozen in liquid nitrogen and ground using a sterile micropestle, and finally mixed with the TRIzol™ reagent. The next steps were then carried out as per the manufacturer's instructions. RNA extracts were treated with DNaseI (Zymo Research) at a concentration of 0.1 U/µL then cleaned and concentrated using ZR RNA Clean and Concentrate™ (Zymo Research). The

Table 2: Top BLAST hit refers to the sequence in the database that has the highest identity (%ID) with the knottin-like sequencemined from the transcriptome. Transcriptome refers to the plant transcriptome from which the sequence was identified. RT-PCR refers to the sample from which the sequence was validated by RT-PCR.

Code	Transcriptome*	RT-PCR*	Top BLAST hit		
			GenBank ID	Record Name	% ID
AS1	AS	AS	ALI96627.1	alpha amylase precursor Alstotide S4 [A. scholaris]	65
AS17	AS	nv	ALI96625.1	alpha amylase precursor Alstotide S1 [A. scholaris]	66
AS18	AS	AS	ALI96625.1	alpha amylase precursor Alstotide S1 [A. scholaris]	98
TP3	AS, TP	TP	XP_011706800.1	PREDICTED: omega-conotoxin-like protein 1 [Wasmannia auropunctata]	37
TP4	TP	nv	AGN03148.1	alpha amylase inhibitor precursor allatide C1 [Allamanda cathartica]	43
TP7	TP	nv	AGN03148.1	alpha amylase inhibitor precursor allatide C1 [Allamanda cathartica]	36
TP8	TP	nv	ALI96626.1	alpha amylase precursor Alstotide S2 [A. scholaris]	43
TP9	TP	nv	AHC03342.1	alpha amylase inhibitor precursor Wrightide R1 [W. religiosa]	36
TP10	TP	nv	ALI96627.1	alpha amylase precursor Alstotide S4 [A. scholaris]	42
TP11	TP	nv	ALI96627.1	alpha amylase precursor Alstotide S4 [A. scholaris]	41
TP16	TP	nv	ALI96626.1	alpha amylase precursor Alstotide S2 [A. scholaris]	50
TP19	TP	nv	ALI96626.1	alpha amylase precursor Alstotide S2 [A. scholaris]	52
TP20	AS, TP, WP	nv	AHC03344.1	alpha amylase inhibitor precursor Wrightide R3 [W. religiosa]	72
TP23	TP	nv	AHC03342.1	alpha amylase inhibitor precursor Wrightide R1 [W. religiosa]	44
TP25	AS, TP, WP	nv	ALI96627.1	alpha amylase precursor Alstotide S4 [A. scholaris]	70
TP30	AS, TP, WP	nv	ALI96626.1	alpha amylase precursor Alstotide [A. scholaris]	82
TP31	AS, TP	TP	AGN03148.1	alpha amylase inhibitor precursor allatide C1 [Allamanda cathartica]	45
WP2	WP	WP	ALI96627.1	alpha amylase precursor Alstotide S4 [A. scholaris]	48
WP10	WP	nv	AHC03342.1	alpha amylase inhibitor precursor Wrightide R1 [W. religiosa]	63
WP11	AS, TP, WP	WP	AHC03343.1	alpha amylase inhibitor precursor Wrightide R2 [W. religiosa]	62
WP12	AS, TP, WP	WP	CDP10890.1	unnamed protein product [Coffea canephora]	37
WP17	WP	nv	AHC03342.1	alpha amylase inhibitor precursor Wrightide R1 [W. religiosa]	49
WP18	WP	nv	AHC03343.1	alpha amylase inhibitor precursor Wrightide R2 [W. religiosa]	77

quality of the purity and integrity of RNA extracts were assessed using the NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific), MCE®-202 MultiNA (Shimadzu Corporation) and agarose gel electrophoresis. The RNA extracts were quantified using the Qubit 2.0 fluorometer (Thermo Fisher Scientific).

Transcriptome sequencing and first-pass assembly

RNA extracts with quality (Supplementary table 2, Supplementary figure 3) as required by the sequencing facility were sent for sequencing to BGI Tech Solutions Co., Limited (Tai Po, Hong Kong). TruSeq® RNA Prep Kit (llumina, Inc.) was used to construct mRNA-focused libraries from the total RNA samples. The HiSeq® 4000 System (llumina, Inc.) at 150bp paired-end reads was used as the sequencing platform. The sequence reads were then assembled by BGI Tech Solutions Co., Ltd. using the Trinity *de novo* assembler (Grabherr et al. 2011).

Gene mining from the transcriptome

Open reading frames (ORFs) of at least 90 nucleotides were predicted from the assembled transcripts using TransDecoder (Haas and Papanicolaou 2017). Transcripts bearing cysteine knot motifs were screened using the CyPerl script (Zhang et al. 2015). For cyclotide screening, the script was modified to search for all sequences with the motif C[^C]{3,6}C[^C]{3,8}C[^C]{3,7}C[^C]C[^C]{4,7}C, where [^C] and {n1,n2} represent the non-cysteine residues and their residue ranges, respectively. For linear knottin screening, the motif C[^C]{6}C[^C]{4,8}CC[^C]{3,6}C[^C]{4,8}C was searched. Sequences similar to the predicted genes were identified using the Basic Local Alignment Search Tool (BLAST; Altschul et al., 1990; Gish and States, 1993)search in the databases curated by the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/).

Validation of transcripts

Primers were designed based on the mined sequences. Reverse transcription PCR (RT-PCR) was performed to validate the presence of the mined genes. Each of the amplified sequences was sent to either Macrogen or the Philippine Genome Center for capillary sequencing. The resulting sequence was aligned with the mined cyclotide sequences using MultAlin (http://multalin.toulouse.inra.fr/; Corpet, 1988). These were then translated into amino acid sequences using the ExpASy Translate tool (https://web.expasy.org/translate/; Gasteiger et al., 2003)to determine if the amino acid sequence in the cyclotide domain resulting from RT-PCR is identical with the sequence mined from the transcriptome.

RESULTS AND DISCUSSION

Cyclotide screening via LC/MS

We identified 6 out of 180 plant species as potential cyclotide producers through LC/MS screening. These were (1) *Alstonia scholaris* (Apocynaceae family), (2) *Wrightia pubescens* (Apocynaceae family), (3) *Momordica charantia* (Cucurbitaceae family), (4) *Tabernaemontana pandacaqui* (Apocynaceae family), (5) *Talinum sp*. (Talinaceae family) and (6) *Clitoria racemosa* (Fabaceae family). The low occurrence of potential cyclotide-producers in the sample pool screened is an evidence of the non-ubiquity of cyclotides in the plant kingdom. Currently, it is believed that cyclotides are only ubiquitous in the Violaceae family and occur in a small number of plants belonging to the Rubiaceae, Cucurbitaceae, Solanaceae, and Fabaceae families. Cyclotide distribution in all other plant families is yet to be determined (Ravipati et al. 2017).

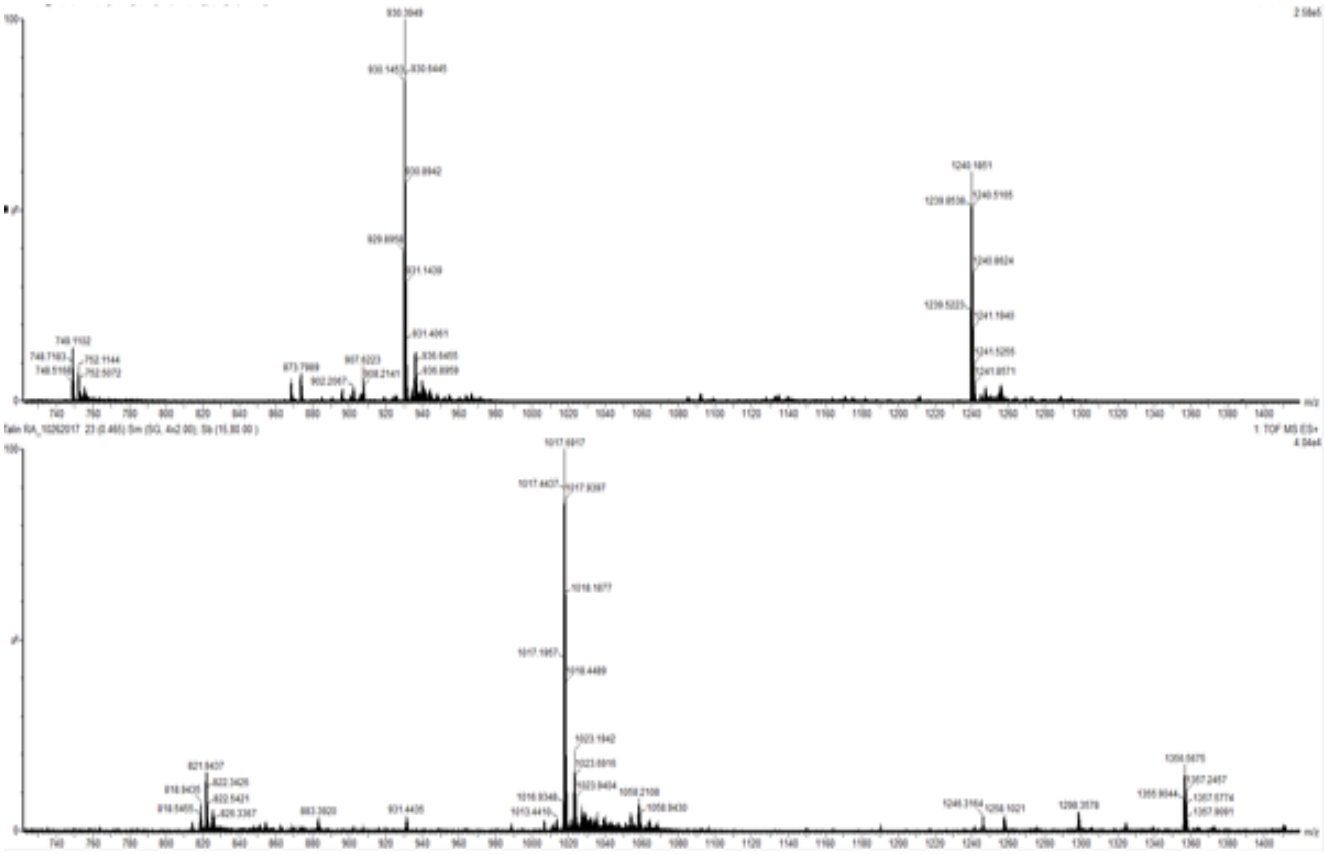


Figure 1: Mass spectra of crude (top) and alkylated (bottom) *Talinum* sp. extract showing mass shift of + 348 Da.

Since *M. charantia* and *A. scholaris* have already been found to contain knottins (He et al. 2013, Nguyen et al. 2015b),we focused on analyzing the four remaining plant extracts that tested positive for cyclotide-like peaks in their LC/MS profiles. To test the presence of 6 conserved cysteines, we alkylated the extracts and then analyzed their LC-MS profiles to see if they exhibit the + 348 Da mass shift. The leaf extract from *Clitoria ternatea* (butterfly pea plant; Fabaceae family) was used as a positive control because it is already a well-studied cyclotide producer (Poth et al. 2011a, 2011b). As expected, the alkylation of the *C. ternatea* extract was successful because of the observed +348 Da mass shift of the cyclotide peaks in the mass spectrum (Supplementary figure 1).

The spectra of *Talinum sp*., *C. racemosa* and *T. pandacaqui* all exhibit positive mass shifts of +348 Da in the cyclotide-like peaks, which makes them highly likely to contain cyclotides, or at the very least, peptides closely related to cyclotides (Figures 1, 2, and 3). On the other hand, the mass spectra for *W. pubescens* did not show +348 Da mass shift (Fig. 4). Thus, we focused on the purification and structure elucidation of the cyclotide-like peaks that we have seen in *T. pandacaqui*, *Talinum* sp., and *C. racemosa.* The crude and alkylated extracts of these three samples were subjected to further LC/MS and MS/MS analyses.

For *Talinum sp*., the cyclotide-like precursor ion from the crude extract, which corresponds to the native peptide (m/z = 1240.31), shows little to no fragmentation in its MS/MS spectrum (Fig. 5A, B). Likewise, the cyclotide-like precursor ion from the alkylated extract, which corresponds to the reduced and alkylated peptide (m/z = 1356.48), shows minimal fragmentation (Fig. 5C, D). The same was observed for *C. racemosa* extracts: the precursor ions from both the crude extract (m/z = 1034.21) the alkylated extract (m/z = 1150.30) show little to no fragmentation in its MS/MS spectrum (Fig. 6). For *T. pandacaqui*, the cyclotide-like precursor ion from the crude extract (m/z = 615.48) shows some fragmentation in its MS/MS spectrum (Figure 7A, B). On the other hand, the cyclotide-like precursor ion from the alkylated extract (m/z = 684.93) is absent from its MS/MS spectrum, and in its place, several fragment ions are observed (Figure 7C, D).

Poth and colleagues (2011b) proposed that the lack of fragmentation observed in tandem MS analyses of reduced and alkylated cyclotides is indicative of their cyclic nature and may be used as a marker for cyclotide presence in plants. In contrast, a linearized, reduced and alkylated cyclotide shows numerous fragment ions in its MS/ MS spectrum(Poth et al. 2011b). Since the precursor ions from both the crude and the alkylated extracts of *Talinum sp*. and *C. racemosa* show minimal fragmentation, we postulate that the detected peptides are either cyclotides or backbone-cyclized disulfide-rich peptides structurally related to cyclotides. In contrast, the precursor ion from the alkylated extract of *T. pandacaqui* fragments easily in MS/MS, indicating that these peptides are possibly acyclotides.

Acyclotides, or uncyclotides, are linear-backbone cyclotides having the same CK arrangement as backbone-cyclized cyclotides (Nguyen et al. 2012, Poth et al. 2012). These include hedyotides from *Hedyotis biflora* and panitides from *Panicum laxum*(Nguyen et al. 2011b, 2013). We postulate that linear-backbone cyclotides fragment more easily in MS/MS than head-to-tail cyclized cyclotides after alkylation because the CK is already destroyed, making the peptide essentially linear. It is also possible that the detected peptides are disulfide-rich pseudocyclic peptides structurally similar to conventional cyclotides.

Since the MS/MS data of the alkylated extract of *T. pandacaqui* fragmented easily after alkylation, we ran this through PEAKS Studio software (PEAKS Studio 8.5 Build 20180105), using its de-novo sequencing function to predict the most plausible peptide sequence for the raw MS/MS data of the precursor ion with m/z = 684.93 (m/z = 615.48 from the crude extract).

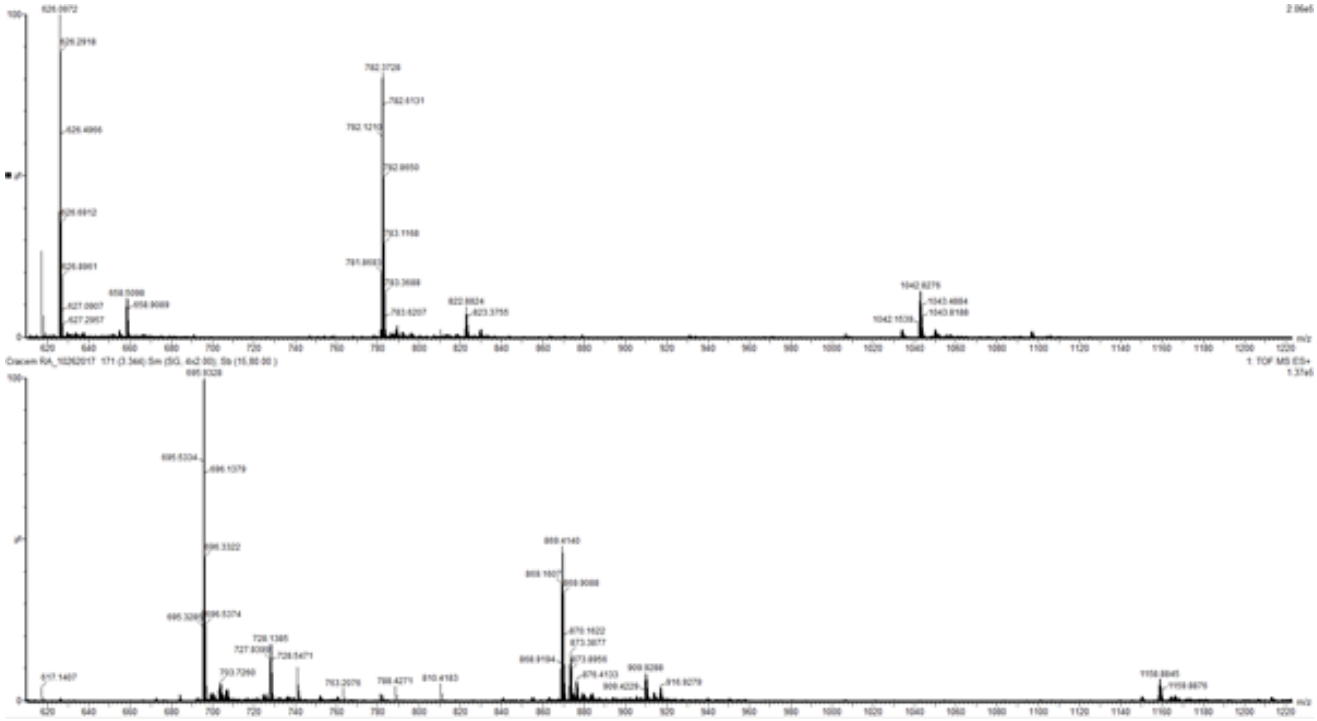


Figure 2: Mass spectra of crude (top) and alkylated (bottom) *C. racemosa* extract showing mass shift of + 348 Da

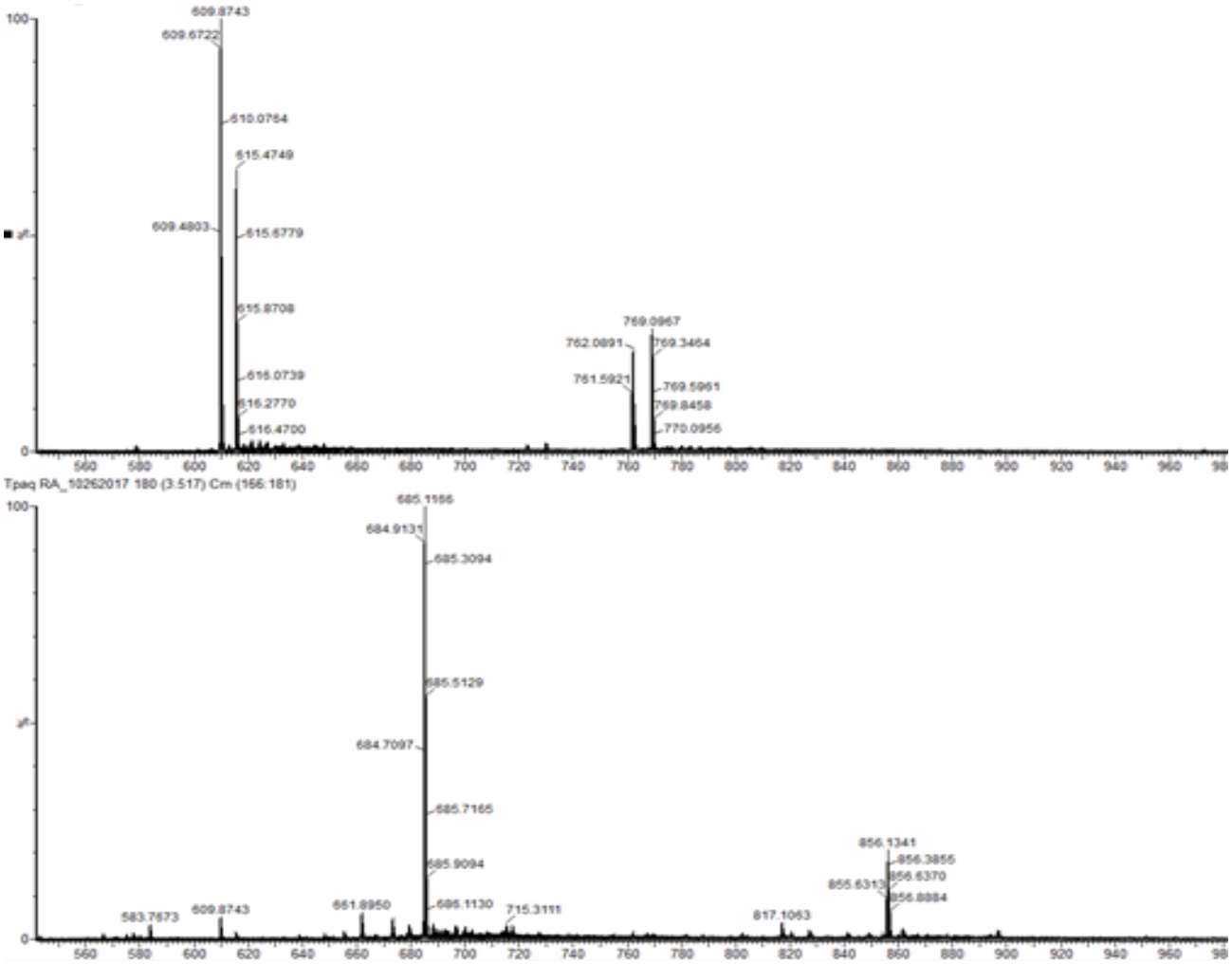


Figure 3: Mass spectra of crude (top) and alkylated (bottom) *T. pandacaqui* extract showing mass shift of + 348 Da

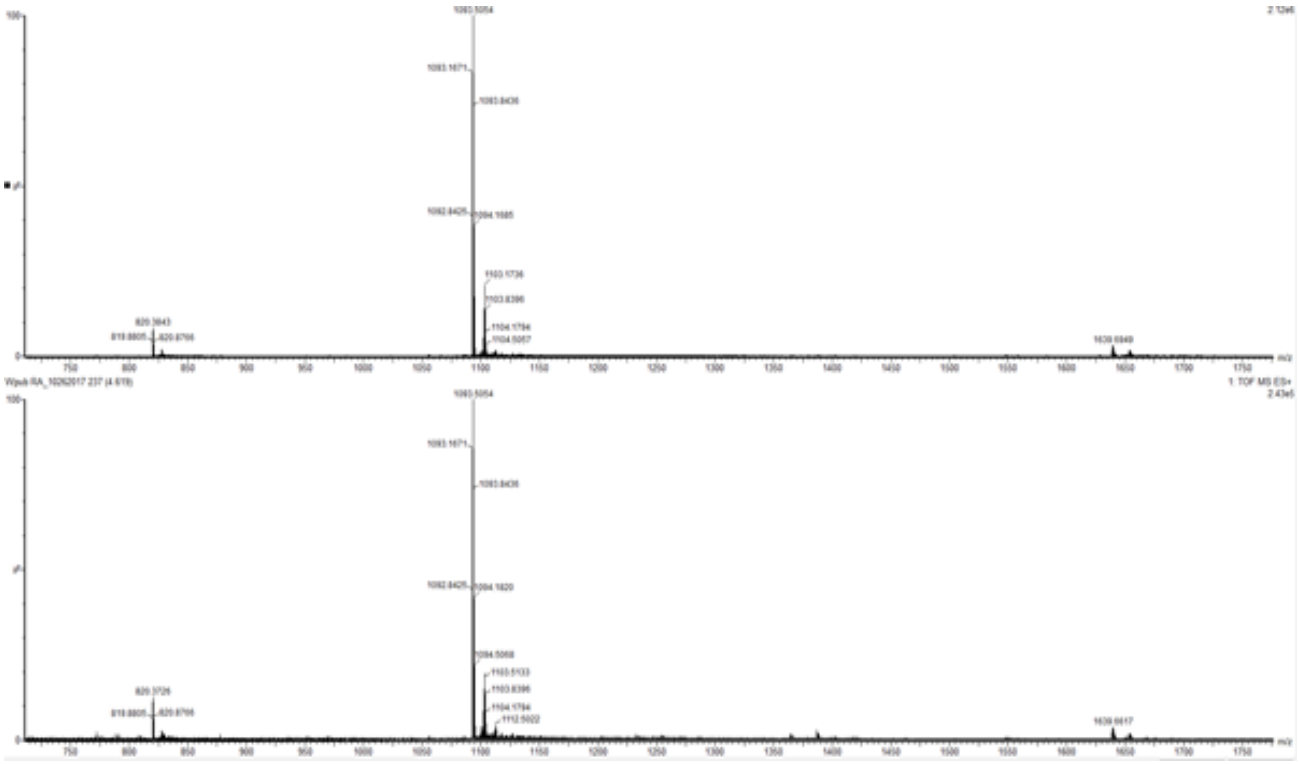


Figure 4: Mass spectra of crude (top) and alkylated (bottom) *W. pubescens* extract showing no mass shift

However, the top two sequences predicted lack the 6 conserved cysteine residues of cyclotides/acyclotides, or other disulfide-rich cyclic and pseudocylic peptides (Supplementary fig. 2). This may have been caused by the presence of contaminating peptides in the unpurified sample. As such, more structural studies (i.e. MS/MS and NMR) need to be performed on the pure peptide from *T. pandacaqui* for the sequence to be more accurate. This, however, would require a large amount of the raw material, which we did not have on hand.

Identification of cyclotide-like sequences from the transcriptome

Since we were unable to isolate and identify the peptides using the LC/MS approach, we analyzed the transcriptome of *T. pandacaqui*. We also decided to re-analyze *W. pubescens* using this approach because it was possible that the absence of the mass shift in the LC/MS analysis was due to the low amount of peptide in the sample pool. Likewise, other underrepresented CK peptides could be present in *A. scholaris* aside from the four alpha-amylase inhibitors isolated by Nguyen et al.(2015b) using a peptide-based approach. Thus, we screened the transcriptomes of these three plants. Again, we used *C. ternatea* as our positive control.

Cyclotides are ribosomally synthesized as linear precursors (Fig. 8) typically composed of an endoplasmic reticulum (ER) signal region, a pro-domain, an N-terminal repeat (NTR), a cyclotide domain and finally a short C-terminal repeat (CTR; Fig. 8A). Multiple cyclotide domains may be present in a single precursor, as in the case of Kalata B2 from *O. affinis* (Fig. 8B; Jennings et al., 2001). However, the *C. ternatea* cyclotide precursor lacks the pro-domain and the NTR of typical cyclotide precursors (Fig. 8C). Instead, the ER signal is immediately followed by the cyclotide sequence, and is connected by a short spacer to an Albumin-1 chain a (A1a) domain (Nguyen et al. 2011a)

Since a cyclotide is approximately 30 amino acids long, we used a cutoff of 90 nucleotides for ORF prediction in order to screen all potential cyclotide sequences, regardless of the presence or absence of their precursors. Potential cyclotide sequences were then identified from the predicted ORFs using a modified CyPerl script, which uses pattern-matching to search for transcripts having the conserved cyclotide motif. This script was originally written by J. Zhang et al. (2015) to search only the last 45 amino acid residues of a translated ORF, as the cyclotide domain of the typical cyclotide precursor is within this region. However, this excludes non-conventional cyclotide precursors such as those from *C. ternatea* so we removed this constraint in order to include such precursors in our search.

Nineteen (19) unique cyclotide sequences were mined from the assembled transcriptomes (Fig. 9). Seven (7) of these were screened from at least two transcriptomes (Table 1), which could either indicate the presence of identical cyclotide genes in two distantly related species or the cross-contamination of signals during sequencing caused by a phenomenon known as "index switching". The latter is a more likely explanation since all the distinct features of the *C. ternatea* cyclotide precursor is present in all of the mined sequences. Indeed, search results show that all of the putative cyclotide domains were identical to the cyclotides of *C. ternatea* (Table 1) that were identified in samples from Singapore and Australia (Nguyen et al. 2011a, Gilding et al. 2016). Index switching is a known limitation of multiplexed next generation sequencing (NGS) technologies and impacts the accuracy of NGS results (Sinha et al. 2017, Costello et al. 2018, Illumina,Inc. 2018). Despite this, the ability of NGS to generate a wealth of useful information cannot be discounted. Since we suspected the occurrence of index switching, we validated the presence of the mined genes by amplification from the cDNA of the plant samples.

Of the 19 sequences, only 17 were successfully confirmed to be present in the *C. ternatea* sample via RT-PCR. The resulting sequences of the RT-PCR products had high sequence similarities with that of the mined sequences after translation into amino acid sequences (Table 1; Supplementary fig. 4). Problems were encountered in the amplification and sequencing of CT8 and CT17, which resulted to sequences with low similarity in the cyclotide domain. Meanwhile, amplification from the three Apocynaceae plants yielded no result, confirming

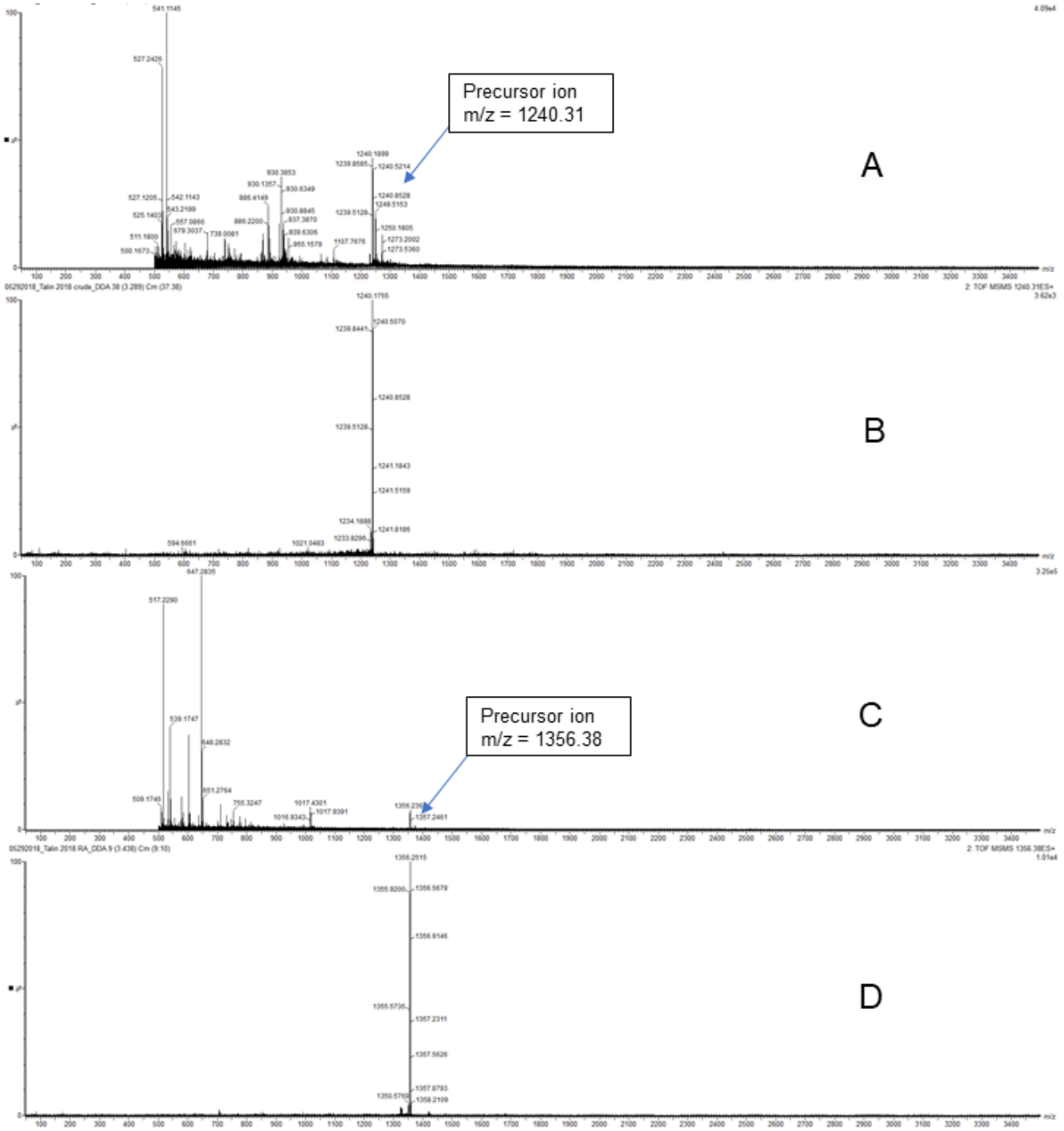


Figure 5: (A) MS and (B) MS/MS spectra of cyclotide-like precursor ion (m/z = 1240.31; charge = +3; M_{mono} = 3,717.906) from alkylated extract of *Talinum* sp.; (C) MS and (D) MS/MS spectra of cyclotide-like precursor ion (m/z = 1356.38; charge = +3; M_{mono} = 4,066.11) from alkylated extract of *Talinum* sp.

that the mined sequences from their transcriptomes were products of index switching.

Since cyclotide sequences were absent in the three Apocynaceae plants, we hypothesized that the signals detected from the LC/MS screen are peptides with properties similar to cyclotides, meaning they are similar in size and have 3 pairs of disulfide bonds. Linear knottins, such as alpha-amylase inhibitors from the Apocynaceae plants, share these properties with cyclotides (Nguyen et al. 2014, 2015b, 2015a).

Pattern search using the linear knottin motif yielded 23 unique sequences (Fig. 10). Seven (7) of these were found in at least two of the transcriptomes, which again might have been caused by index switching (Table 2). Of these seven, we were able to validate three (3) via RT-PCR and confirmed that each of these three transcripts are present in a single sample (Table 2; Supplementary fig. 5). Most of the mined knottin-like transcripts contain the three conserved domains characteristic of Apocynaceae alpha-amylase inhibitor precursors (Fig. 10). The transcripts of AS1, AS18, TP8, TP19, TP23 and WP10 were not fully sequenced; they lack the ER signal in the N-terminus. On the other hand, the pro-domain of TP3 share little similarity with that of the Apocynaceae alpha-amylase inhibitors, as well the rest of the mined sequences. BLAST search for similar sequences returned Apocynaceae alpha-amylase inhibitors as top hits for most of the sequences (Table 2). The CK domain of AS18 was found to be identical to alstotide S1 (Genbank

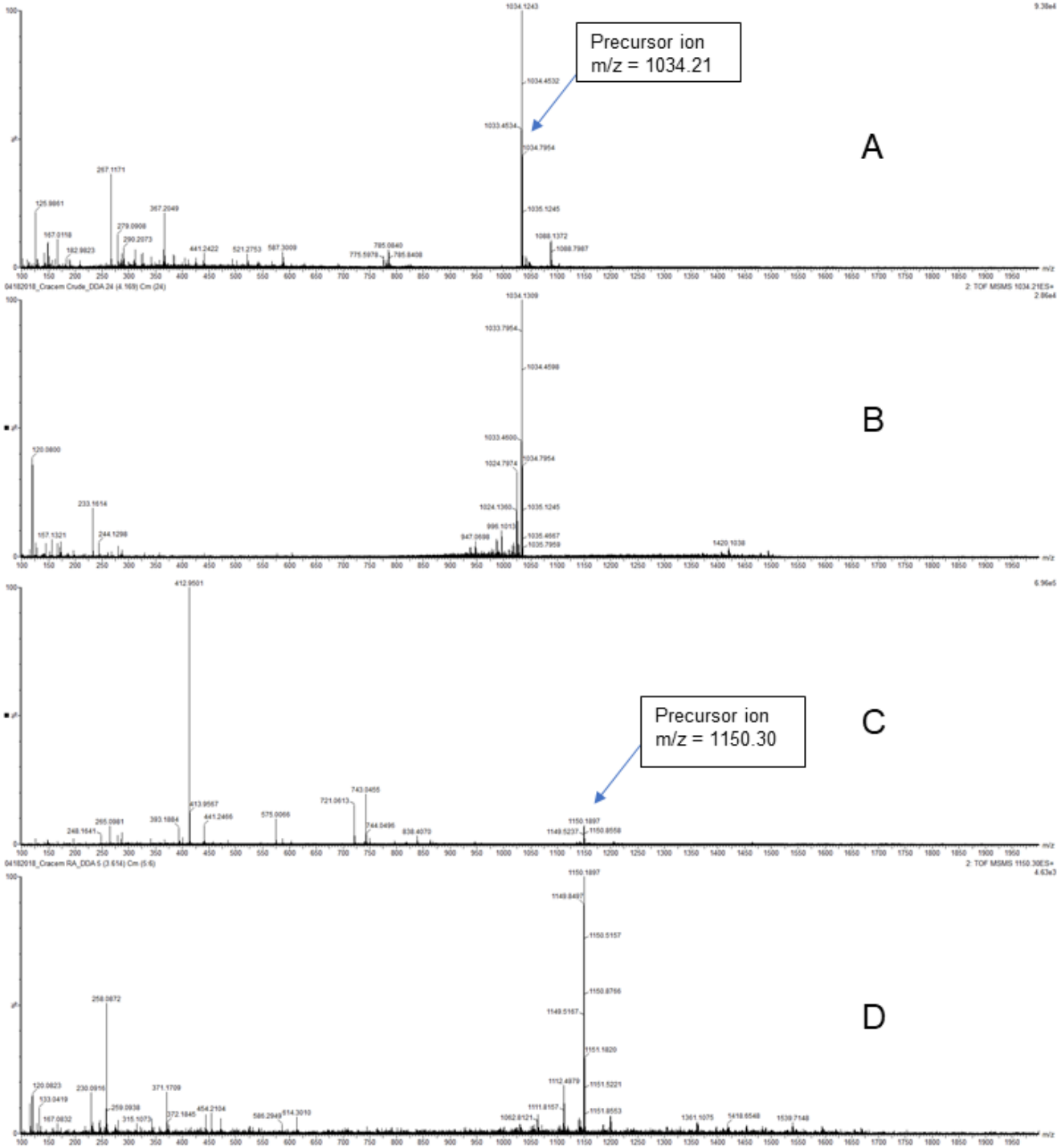


Figure 6: (A)MS and (B) MS/MS spectra of cyclotide-like precursor ion (m/z = 1034.21; charge = +3; M_{mono} = 3,099.63) from crude extract of *C. racemosa*; (C)MS and (D) MS/MS spectra of cyclotide-like precursor ion (m/z = 1150.30; charge = +3; M_{mono} = 3,447.9) from alkylated extract of *C. racemosa*

accession no. ALI96625.1; Nguyen et al. 2015b). The rest of the mined sequences, however, do not share identical CK domains with their top BLAST hits, implying that they are possibly novel uncharacterized proteins.

Despite the lack of a cyclic backbone, linear knottins are still good candidates for drug development. Like cyclotides, linear knottins exhibit peptide promiscuity, in which a single peptide structure could have multiple functions(Franco 2011). For instance, alstotide S1, which functions primarily as an insect alpha-amylase inhibitor, was shown to have antiviral activities against infectious bronchitis virus and Dengue virus type 2 (Nguyen et al. 2015b). Similarly, the peptide sequences we mined may have other functions aside from plant defense. Thus, it is valuable to pursue functional studies on these peptides in the future.

While naturally-derived knottins already exhibit a wide range of bioactivity, novel functions may be engineered into these peptides. It has been shown that their CK residues are essential and sufficient to sustain the stable knotted conformation while the rest of the residues are amenable to hypermutation, which could be used to introduce novel functions (Clark et al. 2006, Thell et al. 2016). Several knottins have already been modified to target various proteins for both therapeutic and molecular imaging purposes (Kintzing and Cochran 2016). However, despite the structural similarities among knottins, the effectivity of the grafted peptide for a specific application still varies

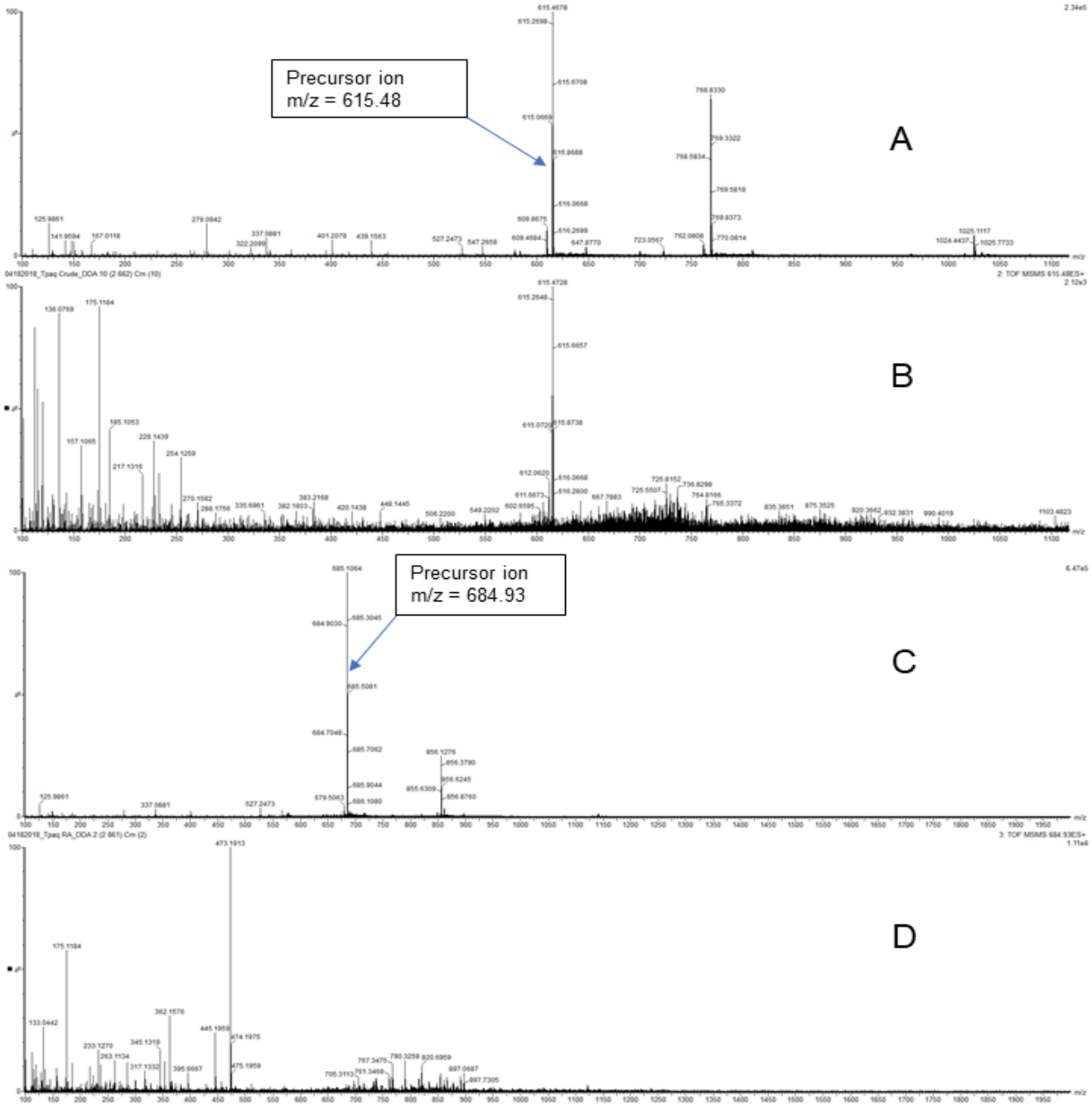


Figure 7: (A) MS and (B) MS/MS spectra of cyclotide-like precursor ion (m/z = 615.48; charge = +5; M_{mono} = 3,072.36) from crude extract of *T. pandacaqui*; (C) MS and (D) MS/MS spectra of cyclotide-like precursor ion (m/z = 684.93; charge = +5; M_{mono} = 3,419.61) from alkylated extract of *T. pandacaqui*

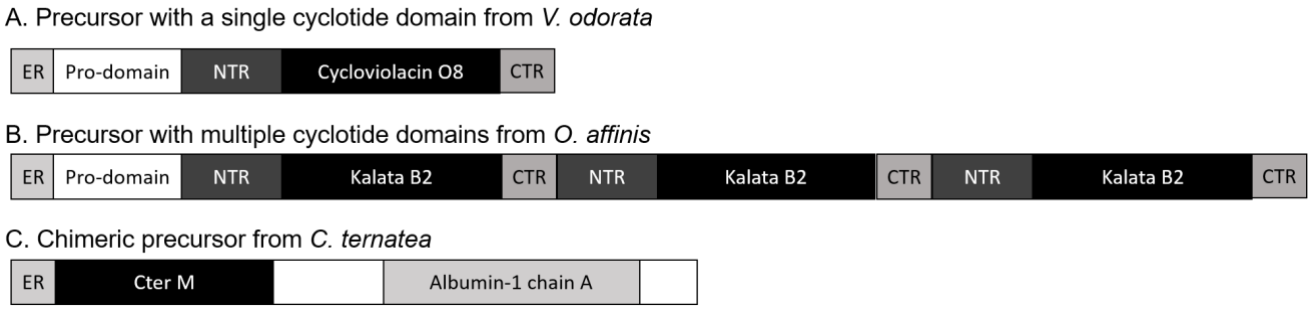


Figure 8: Linear cyclotide precursors. Most cyclotide precursors may have a single cyclotide domain (A) or multiple cyclotide domains (B), which are preceded by the NTR and followed by the CTR. The CTR contains the recognition site required for cyclization. (C) In *C. ternatea*, cyclotides are co-expressed with the albumin-1 chain A.

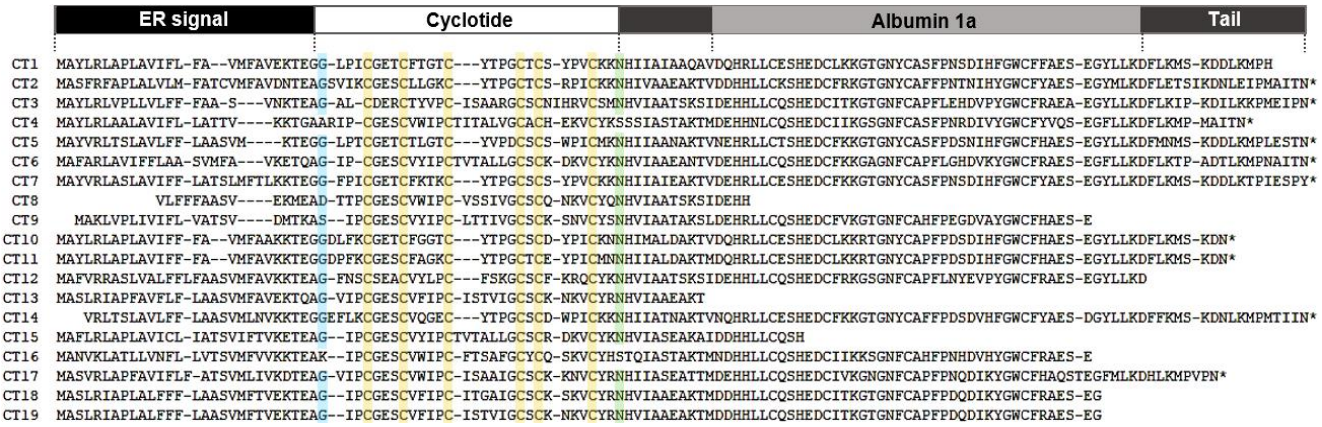


Figure 9: Cyclotides mined from the transcriptome. The mined cyclotide transcripts contain the three conserved domains characteristic of *C. ternatea* cyclotide precursors namely, the ER signal, the cyclotide domain and the Albumin 1a domain. The cyclotide domains contain the conserved cysteine motif (yellow). The Asn (green) required for cyclization is present in all the mined cyclotides except for CT4 and CT16. This Asn will form a peptide bond with the N-terminal residue in blue. Alignment was done using the MultAlin server (Corpet, 1988).

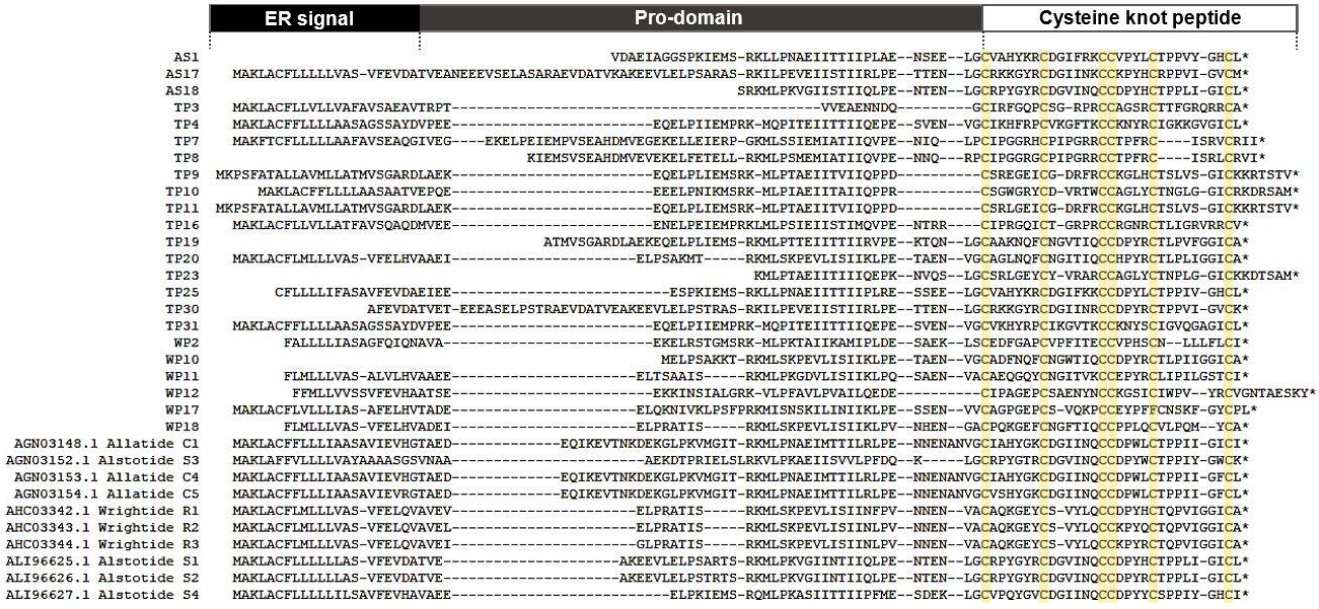


Figure 10: Putative knottin peptides mined from the transcriptome. Most of the mined knottin-like transcripts contain the three conserved domains characteristic of Apocynaceae knottin precursors namely, the ER signal, the pro-domain and the cysteine knot domain. The cysteine knot domains contain the conserved cysteine motif of knottins (yellow). Initial alignment was done using the MultAlin server (Corpet, 1988) then refined manually.

depending on the knottin scaffold used. For instance, thrombopoietin-antagonists derived from EETI were more potent than those derived from the human Agouti-related protein (Krause et al. 2007). Moreover, while several knottins were shown to be non-toxic to human cells (Nguyen et al. 2015a, 2015b, Loo et al. 2016) some showed potent hemolytic and cytotoxic activities (Tam et al. 1999, Chen et al. 2006, Wang et al. 2008). Thus, it is important to consider the peptide's purpose in selecting a scaffold. The discovery of new knottin sequences not only expands the database of naturally bioactive peptides but also enriches the pool of potential molecular scaffolds for various applications.

CONCLUSION

We identified 6 Philippine plants that produce disulfide-rich peptides structurally related to cyclotides. These were *M. charantia*, *A. scholaris*, *W. pubescens*, *T. pandacaqui*, *Talinum* sp., and *Clitoria racemosa*. While CK peptides from *M. charantia* and *A. scholaris* have already been identified, no reports of such peptides have been made for the rest of these plants. Through transcriptome analysis, we confirmed that the leaves sampled from *A. scholaris*, *W. pubescens*, and *T. pandacaqui* do not contain cyclotides. Instead, they contain knottin-like peptides that are structurally related to cyclotides and from their transcriptomes, we were able to identify 22 putative knottins similar to the alpha-amylase inhibitors from the Apocynaceae family. However, future experiments must be done to validate the presence of these genes in the plant. These genes are novel and should be studied further to determine their function and potential applications. This is the first report of knottins found in *W. pubescens* and *T. pandacaqui*, which are indigenous to the Philippines. In addition, we showed that *Talinum* sp. and *Clitoria racemosa* are potential producers of cyclic CK peptides based on the LC-MS screening and it is worthwhile to pursue transcriptome analysis to identify these peptides.

In this study we showed that protein-based screening can be effectively used for CK peptide screening of large libraries of plant extracts. However, a transcript-based approach offers a more comprehensive analysis of a plant's CK peptide pool. As demonstrated, we were able to identify both cyclotides and linear knottins by pattern search. By simply modifying the pattern to be searched, it is possible to discover new CK peptides that may be useful in different fields of applications.

ACKNOWLEDGMENT

This work was supported by the Emerging Interdisciplinary Research Program (EIDR) of the Office of the Vice-President for Academic Affairs (OVPAA), University of the Philippines (UP) under grant number EIDR C06-021, and the Department of Science and Technology (DOST).

The authors would also like to thank the following: the Philippine Council on Agriculture and Aquatic Resources Research and Development (PCAARRD) for monitoring the research program; Ms. Giselle, Ms. Lina, and Dr. Lourdes Wagan from the Department of Environment and Natural Resources – National Capital Region (DENR-NCR) for the issuance of the Wildlife Gratuitous Permit used for this study; Dr. Maria Lea H. Villavicencio and Mr. John Christopher Polinar from the National Plant Genetic Resources Laboratory (NPGRL) in UP Los Baños for providing samples of some of the 30 medicinal plants endorsed by the Department of Health; Mr. Ramon Bandong of the Institute of Biology, UP Diliman, for assisting with the collection of plant samples within UP Diliman and Bataan; Dr. Rogel A. Santos of MacroAsia Mining Corporation and their staff for assisting with sample collection within their property in Brooke's Point, Palawan; and Ms. Ma. Ysabella Elaine D. Conde of the National Institute of Molecular Biology & Biotechnology, UP DIliman, for amplifying six of the cyclotide sequences.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.

Chen B, Colgrave ML, Daly NL, Rosengren KJ, Gustafson KR, Craik DJ. Isolation and Characterization of Novel Cyclotides from *Viola hederaceae*: SOLUTION STRUCTURE AND ANTI-HIV ACTIVITY OF vhl-1, A LEAF-SPECIFIC EXPRESSED CYCLOTIDE. J Biol Chem. 2005;280(23):22395–405.

Chen B, Colgrave ML, Wang C, Craik DJ. Cycloviolacin H4, a Hydrophobic Cyclotide from *Viola hederaceae*. J Nat Prod. 2006;69(1):23–8.

Clark RJ, Daly NL, Craik DJ. Structural plasticity of the cyclic-cystine-knot framework: implications for biological activity and drug design. Biochem J. 2006;394(1):85–93.

Colgrave ML, Craik DJ. Thermal, Chemical, and Enzymatic Stability of the Cyclotide Kalata B1: The Importance of the Cyclic Cystine Knot †. Biochemistry (Mosc). 2004;43(20):5965–75.

Colgrave ML, Kotze AC, Huang Y-H, O'Grady J, Simonsen SM, Craik DJ. Cyclotides: Natural, Circular Plant Peptides that Possess Significant Activity against Gastrointestinal Nematode Parasites of Sheep †. Biochemistry (Mosc). 2008;47(20):5581–9.

Colgrave ML, Kotze AC, Kopp S, McCarthy JS, Coleman GT, Craik DJ. Anthelmintic activity of cyclotides: In vitro studies with canine and human hookworms. Acta Trop. 2009;109(2):163–6.

Corpet F. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. 1988;16(22):10881–90.

Costello M, Fleharty M, Abreu J, Farjoun Y, Ferriera S, Holmes L, Granger B, Green L, Howd T, Mason T, Vicente G, Dasilva M, Brodeur W, DeSmet T, Dodge S, Lennon NJ, Gabriel S. Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. BMC Genomics. 2018;19(1).

Daly NL, Gustafson KR, Craik DJ. The role of the cyclic peptide backbone in the anti-HIV activity of the cyclotide kalata B1. FEBS Lett. 2004;574(1–3):69–72.

Fensterseifer ICM, Silva ON, Malik U, Ravipati AS, Novaes NRF, Miranda PRR, Rodrigues EA, Moreno SE, Craik DJ, Franco OL. Effects of cyclotides against cutaneous infections caused by Staphylococcus aureus. Peptides. 2015;63:38–42.

Franco OL. Peptide promiscuity: An evolutionary concept for plant defense. FEBS Lett. 2011;585(7):995–1000.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003;31(13):3784–8.

Gilding EK, Jackson MA, Poth AG, Henriques ST, Prentis PJ, Mahatmanto T, Craik DJ. Gene coevolution and regulation lock cyclic plant defence peptides to their targets. New Phytol. 2016;210(2):717–30.

Gish W, States DJ. Identification of protein coding regions by database similarity search. Nat Genet. 1993;3(3):266–72.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644–52.

Gran L. Oxitocic principles found in Oldenlandia affinis. Meddelelser Nor Farmasøytisk Selsk. 1970;32:173 – 180.

Gran L. Oxytocic principles of Oldenlandia affinis. Lloydia. 1973;36(2):174–8.

Gran L, Sandberg F, Sletten K. Oldenlandia affinis (R&S) DC. A plant containing uteroactive peptides used in African traditional medicine. J Ethnopharmacol. 2000;70(3):197–203.

Gruber CW, Elliott AG, Ireland DC, Delprete PG, Dessein S, Goransson U, Trabi M, Wang CK, Kinghorn AB, Robbrecht E, Craik DJ. Distribution and Evolution of Circular Miniproteins in Flowering Plants. PLANT CELL ONLINE. 2008;20(9):2471–83.

Gunasekera S, Foley FM, Clark RJ, Sando L, Fabri LJ, Craik DJ, Daly NL. Engineering Stabilized Vascular Endothelial Growth Factor-A Antagonists: Synthesis, Structural Characterization, and Bioactivity of Grafted Analogues of Cyclotides. J Med Chem. 2008;51(24):7697–704.

Gustafson KR, Sowder RC, Henderson LE, Parsons IC, Kashman Y, Cardellina JH, McMahon JB, Buckheit RW, Pannell LK, Boyd MR. Circulins A and B. Novel human immunodeficiency virus (HIV)-inhibitory macrocyclic peptides from the tropical tree Chassalia parvifolia. J Am Chem Soc. 1994;116(20):9337–8.

Haas, BJ, Papanicolaou, A. Transdecoder (Find Coding Regions Within Transcripts). 2017.

Hashempour H, Koehbach J, Daly NL, Ghassempour A, Gruber CW. Characterizing circular peptides in mixtures: sequence fragment assembly of cyclotides from a violet plant by MALDI-TOF/TOF mass spectrometry. Amino Acids. 2013;44(2):581–95.

He W, Chan LY, Zeng G, Daly NL, Craik DJ, Tan N. Isolation and characterization of cytotoxic cyclotides from Viola philippica. Peptides. 2011;32(8):1719–23.

He W-J, Chan LY, Clark RJ, Tang J, Zeng G-Z, Franco OL, Cantacessi C, Craik DJ, Daly NL, Tan N-H. Novel Inhibitor Cystine Knot Peptides from Momordica charantia. Driscoll PC, editor. PLoS ONE. 2013;8(10):e75334.

Illumina,Inc. Effects of Index Misassignment on Multiplexing and Downstream Analysis [White Paper]. 2018;(Pub. No. 770-2017-004-D). Accessed on 21 March 2019 at https://www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf?linkId=36607862

Ireland DC, Colgrave ML, Nguyencong P, Daly NL, Craik DJ. Discovery and Characterization of a Linear Cyclotide from Viola odorata: Implications for the Processing of Circular Proteins. J Mol Biol. 2006;357(5):1522–35.

Jennings C, West J, Waine C, Craik D, Anderson M. Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from Oldenlandia affinis. Proc Natl Acad Sci U S A. 2001;98(19):10614–9.

Jennings CV, Rosengren KJ, Daly NL, Plan M, Stevens J, Scanlon MJ, Waine C, Norman DG, Anderson MA, Craik DJ. Isolation, Solution Structure, and Insecticidal Activity of Kalata B2, a Circular Protein with a Twist: Do Möbius Strips Exist in Nature? † , ‡. Biochemistry (Mosc). 2005;44(3):851–60.

Kaas Q, Westermann J-C, Craik DJ. Conopeptide characterization and classifications: An analysis using ConoServer. Toxicon. 2010;55(8):1491–509.

Kintzing JR, Cochran JR. Engineered knottin peptides as diagnostics, therapeutics, and drug delivery vehicles. Curr Opin Chem Biol. 2016;34:143–50.

Krause S, Schmoldt H-U, Wentzel A, Ballmaier M, Friedrich K, Kolmar H. Grafting of thrombopoietin-mimetic peptides into cystine knot miniproteins yields high-affinity thrombopoietin antagonists and agonists: Cystine knot miniprotein TPO agonists. FEBS J. 2007;274(1):86–95.

Lindholm P, Göransson U, Johansson S, Claeson P, Gullbo J, Larsson R, Bohlin L, Backlund A. Cyclotides: a novel type of cytotoxic agents. Mol Cancer Ther. 2002;1(6):365–9.

Loo S, Kam A, Xiao T, Nguyen GKT, Liu CF, Tam JP. Identification and Characterization of Roseltide, a Knottin-type Neutrophil Elastase Inhibitor Derived from Hibiscus sabdariffa. Sci Rep. 2016;6(1).

Mulvenna JP. CyBase: a database of cyclic protein sequence and structure. Nucleic Acids Res. 2006;34(90001):D192–4.

Nguyen GKT, Lian Y, Pang EWH, Nguyen PQT, Tran TD, Tam JP. 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. 2013;288(5):3370–80.

Nguyen GKT, Lim WH, Nguyen PQT, Tam JP. Novel Cyclotides and Uncyclotides with Highly Shortened Precursors from *Chassalia chartacea* and Effects of Methionine Oxidation on Bioactivities. J Biol Chem. 2012;287(21):17598–607.

Nguyen GKT, Zhang S, Nguyen NTK, Nguyen PQT, Chiu MS, Hardjojo A, Tam JP. 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. 2011a;286(27):24275–87.

Nguyen GKT, Zhang S, Wang W, Wong CTT, Nguyen NTK, Tam JP. Discovery of a Linear Cyclotide from the Bracelet Subfamily and Its Disulfide Mapping by Top-down Mass Spectrometry. J Biol Chem. 2011b;286(52):44833–44.

Nguyen PQT, Luu TT, Bai Y, Nguyen GKT, Pervushin K, Tam JP. Allotides: Proline-Rich Cystine Knot α-Amylase Inhibitors from *Allamanda cathartica*. J Nat Prod. 2015a;78(4):695–704.

Nguyen PQT, Ooi JSG, Nguyen NTK, Wang S, Huang M, Liu DX, Tam JP. Antiviral Cystine Knot α-Amylase Inhibitors from *Alstonia scholaris*. J Biol Chem. 2015b;290(52):31138–50.

Nguyen PQT, Wang S, Kumar A, Yap LJ, Luu TT, Lescar J, Tam JP. Discovery and characterization of pseudocyclic cystine-knot α-amylase inhibitors with high resistance to heat and proteolytic degradation. FEBS J. 2014;281(19):4351–66.

Ovesen RG, Brandt KK, Göransson U, Nielsen J, Hansen HCB, Cedergreen N. Biomedicine in the environment: Cyclotides constitute potent natural toxins in plants and soil bacteria. Environ Toxicol Chem. 2011;30(5):1190–6.

Pallaghy PK, Norton RS, Nielsen KJ, Craik DJ. A common structural motif incorporating a cystine knot and a triple-stranded β-sheet in toxic and inhibitory polypeptides. Protein Sci. 1994;3(10):1833–9.

Pinto MFS, Fensterseifer ICM, Migliolo L, Sousa DA, de Capdville G, Arboleda-Valencia JW, Colgrave ML, Craik DJ, Magalhães BS, Dias SC, Franco OL. Identification and Structural Characterization of Novel Cyclotide with Activity against an Insect Pest of Sugar Cane. J Biol Chem. 2012;287(1):134–47.

Plan MRR, Saska I, Cagauan AG, Craik DJ. Backbone Cyclised Peptides from Plants Show Molluscicidal Activity against the Rice Pest *Pomacea canaliculata* (Golden Apple Snail). J Agric Food Chem. 2008;56(13):5237–41.

Poth AG, Colgrave ML, Lyons RE, Daly NL, Craik DJ. Discovery of an unusual biosynthetic origin for circular proteins in legumes. Proc Natl Acad Sci. 2011a;108(25):10127–32.

Poth AG, Colgrave ML, Philip R, Kerenga B, Daly NL, Anderson MA, Craik DJ. Discovery of Cyclotides in the Fabaceae Plant Family Provides New Insights into the

Cyclization, Evolution, and Distribution of Circular Proteins. ACS Chem Biol. 2011b;6(4):345–55.

Poth AG, Mylne JS, Grassl J, Lyons RE, Millar AH, Colgrave ML, Craik DJ. Cyclotides Associate with Leaf Vasculature and Are the Products of a Novel Precursor in *Petunia* (Solanaceae). J Biol Chem. 2012;287(32):27033–46.

Ravipati AS, Poth AG, Troeira Henriques S, Bhandari M, Huang Y-H, Nino J, Colgrave ML, Craik DJ. Understanding the Diversity and Distribution of Cyclotides from Plants of Varied Genetic Origin. J Nat Prod. 2017;80(5):1522–30.

Sinha R, Stanley G, Gulati GS, Ezran C, Travaglini KJ, Wei E, Chan CKF, Nabhan AN, Su T, Morganti RM, Conley SD, Chaib H, Red-Horse K, Longaker MT, Snyder MP, Krasnow MA, Weissman IL. Index Switching Causes "Spreading-Of-Signal" Among Multiplexed Samples In Illumina HiSeq 4000 DNA Sequencing. bioRxiv. 2017;125724.

Tam J, Wang S, Wong K, Tan W. Antimicrobial Peptides from Plants. Pharmaceuticals. 2015;8(4):711–57.

Tam JP, Lu YA, Yang JL, Chiu KW. An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides. Proc Natl Acad Sci U S A. 1999;96(16):8913–8.

Thell K, Hellinger R, Sahin E, Michenthaler P, Gold-Binder M, Haider T, Kuttke M, Liutkevičiūtė Z, Göransson U, Gründemann C, Schabbauer G, Gruber CW. Oral activity of a nature-derived cyclic peptide for the treatment of multiple sclerosis. Proc Natl Acad Sci. 2016;113(15):3960–5.

Wang CKL, Colgrave ML, Gustafson KR, Ireland DC, Goransson U, Craik DJ. Anti-HIV Cyclotides from the Chinese Medicinal Herb *Viola yedoensis*. J Nat Prod. 2008;71(1):47–52.

Wang CKL, Kaas Q, Chiche L, Craik DJ. CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering. Nucleic Acids Res. 2007;36(Database):D206–10.

Xiang B, Du G-H, Wang X-C, Zhang S-X, Qin X-Y, Kong J-Q, Cheng K-D, Li Y-J, Wang W. [Elucidating the structure of two cyclotides of Viola tianshanica maxim by MALDI TOF/TOF MS analysis]. Yao Xue Xue Bao. 2010;45(11):1402–9.

Zarrabi M, Dalirfardouei R, Sepehrizade Z, Kermanshahi RK. Comparison of the antimicrobial effects of semipurified cyclotides from Iranian *Viola odorata* against some of plant and human pathogenic bacteria. J Appl Microbiol. 2013;115(2):367–75.

Zhang J, Hua Z, Huang Z, Chen Q, Long Q, Craik DJ, Baker AJM, Shu W, Liao B. Two Blast-independent tools, CyPerl and CyExcel, for harvesting hundreds of novel cyclotides and analogues from plant genomes and protein databases. Planta. 2015;241(4):929–40.

Philippines - Country Profile, Biodiversity Facts - Status and trends of biodiversity, including benefits from biodiversity and ecosystem services. Conv. Biol. Divers.

SUPPLEMENTARY INFORMATION

Supplementary Table 1: List of plant extracts screened in this study. Tag – Tagalog; Pal – Palawano

Code	Scientific Name	Common/Filipino Name	Plant Family	Plant part sampled	Sampling Location
CYC 15-001	*Adenanthera pavovina*	Malatanglin	Fabaceae	Leaves	Metro Manila
CYC 15-002	*Albizzia lebbeck*	Langil	Fabaceae	Leaves	Metro Manila
CYC 15-003	*Caesalpinia pulcherrima*	Bulaklak ng paraiso	Fabaceae	Leaves	Metro Manila
CYC 15-004	*Cassia alata*	Akapulko	Fabaceae	Leaves	Metro Manila
CYC 15-005	*Cassia fistula*	Golden shower/ Kanya pistula	Fabaceae	Leaves	Metro Manila
CYC 15-006	*Cynometra ramiflora*	Balitbitan	Fabaceae	Leaves	Metro Manila
CYC 15-007	*Intsia bijuga*	Ipil	Fabaceae	Leaves	Metro Manila
CYC 15-008	*Leucaena leucocephala*	Ipil-ipil	Fabaceae	Leaves	Metro Manila
CYC 15-009	*Pithecellobium dulce*	Kamachile	Fabaceae	Leaves	Metro Manila
CYC 15-010	*Clitoria ternatea*	Butterfly pea	Fabaceae	Leaves	Metro Manila
CYC 16-001	*Alstonia scholaris*	Dita	Apocynaceae	Leaves	Metro Manila
CYC 16-002	*Wrightia pubescens*	Lanete	Apocynaceae	Leaves	Metro Manila
CYC 16-003	*Tabernaemontana pandacaqui*	Kampupot	Apocynaceae	Leaves	Metro Manila
CYC 16-004	*Gliricidia sepium*	Kakawate	Fabaceae	Leaves	Metro Manila
CYC 16-005	*Erythrina orientalis*	Dapdap	Fabaceae	Leaves	Metro Manila
CYC 16-006	*Breynia* sp.	--	Euphorbiaceae	Leaves	Zambales
CYC 16-007	*Bridelia* sp.	--	Phyllanthaceae	Leaves	Zambales
CYC 16-008	*Duranta repens*	Duranta/Golden dew drop	Verbenaceae	Leaves	Zambales
CYC 16-009	*Clerodendrum quadriloculare*	Bagauak na pula	Lamiaceae	Leaves	Zambales
CYC 16-010	*Callicarpa* sp.	--	Lamiaceae	Leaves	Zambales
CYC 16-011	*Flemingia strobilifera*	Payang-payang	Fabaceae	Leaves	Zambales
CYC 16-012	*Ficus oleifolia*	--	Moraceae	Leaves	Zambales
CYC 16-013	*Nauclea* sp.	--	Rubiaceae	Leaves	Zambales
CYC 16-014	*Crotalaria* sp.	--	Fabaceae	Leaves	Zambales
CYC 16-015	*Melastoma malabathricum*	Malatungaw	Melastomataceae	Leaves	Zambales
CYC 16-016	Plant # 30 (Unidentified)	--	--	Leaves	Zambales
CYC 16-017	*Tabernaemontana pandacaqui (Zam)*	Pandakaki-puti/Kampupot	Apocynaceae	Leaves	Zambales
CYC 16-018	Plant # 35 (Unidentified)	--	--	Leaves	Zambales
CYC 16-019	*Elephantopus tomentosus*	--	Compositae	Leaves	Zambales
CYC 16-020	*Morinda citrifolia*		Rubiaceae	Leaves	Zambales
CYC 16-021	Plant # 41 (Unidentified)	--	--	Leaves	Zambales
CYC 17-001	*Hydnocarpus sumatrana*		Achariaceae	Leaves	Bataan
CYC 17-002	*Albizia procera*		Fabaceae	Leaves	Bataan
CYC 17-003	*Cratoxylum formosum*		Hypericaceae	Leaves	Bataan
CYC 17-004	*Tetracera scandens*	Malakatmon	Dilleniaceae	Leaves	Bataan
CYC 17-005	*Pongamia pinnata*	Bani	Fabaceae	Leaves	Bataan
CYC 17-006	*Tectona philippinensis*	Philippine teak	Lamiaceae	Leaves	Bataan
CYC 17-007	*Sterculia oblongata*		Sterculiceae	Leaves	Bataan
CYC 17-008	*Litsea glutinosa*	Puso-puso	Lauraceae	Leaves	Bataan
CYC 17-009	*Hopea plagata*		Dipterocarpaceae	Leaves	Bataan

CYC 17-010	*Streblus asper*	Sandpaper tree	Moraceae	Leaves	Bataan
CYC 17-011	*Diplodiscus paniculatus*		Malvaceae/ Tiliceae	Leaves	Bataan
CYC 17-012	*Terminalia microcarpa*	Kalumpit	Combretaceae	Leaves	Bataan
CYC 17-013	*Rauvolfia sumatrana*		Apocynaceae	Leaves	Bataan
CYC 17-014	*Pterospermum obliquum*	Kulatingan	Sterculiceae	Leaves	Bataan
CYC 17-015	*Melochia umbellata*		Malvaceae	Leaves	Bataan
CYC 17-016	*Mallotus cumingii*		Euphorbiaceae	Leaves	Bataan
CYC 17-017	*Bauhinia malabarica*	Alibangbang	Fabaceae	Leaves	Bataan
CYC 17-018	*Bauhinia malabarica subsp. cumingiana*		Fabaceae	Leaves	Bataan
CYC 17-019	*Derris elliptica*	Tubli/Poison vine	Fabaceae	Leaves	Bataan
CYC 17-020	*Ficus pseudopalma*	Niyog-niyogan	Moraceae	Leaves	Bataan
CYC 17-021	*Cinnamonum mercadoi*	Kalingag	Lauraceae	Leaves	Bataan
CYC 17-022	*Citrus hystrix*	Kabuyau	Rutaceae	Leaves	Bataan
CYC 17-023	*Anaxagora luzonensis*	Dalaira	Annonaceae	Leaves	Bataan
CYC 17-024	*Diospyros blancoi*	Mabolo	Ebenaceae	Leaves	Bataan
CYC 17-025	*Phaeanthus ebtracteolatus*	Kalimatas	Annonaceae	Leaves	Bataan
CYC 17-026	*Dracontomelon dao*	Dao	Anacardiaceae	Leaves	Bataan
CYC 17-027	*Ficus septica*	Hauili	Moraceae	Leaves	Bataan
CYC 17-028	*Syzygium cuminii*	Duhat	Myrtaceae	Leaves	Bataan
CYC 17-029	*Melicope triphylla*		Rutaceae	Leaves	Bataan
CYC 17-030	*Diospyros pilosanthera*	Bolongeta	Ebenaceae	Leaves	Bataan
CYC 17-031	*Diospyros maritima*		Ebenaceae	Leaves	Bataan
CYC 17-032	*Clausina anisum-olens*	Kayumanis	Rutaceae	Leaves	Bataan
CYC 17-033	*Aleurites moluccana*	Lumbang	Euphorbiaceae	Leaves	Bataan
CYC 17-034	*Syzygium tripinnatum*		Myrtaceae	Leaves	Bataan
CYC 17-035	*Hibiscus tiliaceus*	Malabago	Malvaceae	Leaves	Bataan
CYC 17-036	*Piper betle*	ikmo	Piperaceae	Leaves	Bataan
CYC 17-037	*Donax canniformis*	Bamban	Marantaceae	Leaves	Bataan
CYC 17-038	*Callicarpa erioclona*	Tubang-dalag	Lamiaceae	Leaves	Bataan
CYC 17-039	*Artocarpus blancoi*	Antipolo	Moraceae	Leaves	Bataan
CYC 17-040	*Ixora philippinensis*		Rubiaceae	Leaves	Bataan
CYC 17-041	*Dillenia philippinensis*	Katmon	Dilleniaceae	Leaves	Bataan
CYC 17-042	*Blumea balsamifera*	Sambong	Compositae	Leaves	Bataan
CYC 17-043	*Parartocarpus venenosus subsp. papuanus*		Moraceae	Leaves	Bataan
CYC 17-044	*Cassia fistula*	Kanya pistula	Fabaceae	Seeds	Laguna
CYC 17-045	*Euphorbia hirta*	Tawa-tawa	Euphorbiaceae	Seeds	Laguna
CYC 17-046	*Lagerstroemia speciosa*	Banaba	Lythraceae	Seeds	Laguna
CYC 17-047	*Centella asiatica*	Takip kuhol	Umbelliferae	Seeds	Laguna
CYC 17-048	*Vitex negundo*	Lagundi	Verbenaceae	Seeds	Laguna
CYC 17-049	*Momordica charantia*	Ampalaya	Cucurbitaceae	Seeds	Laguna
CYC 17-050	*Leucaena glauca*	Ipil-ipil	Fabaceae	Seeds	Laguna
CYC 17-051	*Cassia alata*	Akapulko	Fabaceae	Seeds	Laguna
CYC 17-052	*Curcuma longa*	Luyang dilaw rhizome	Zingiberaceae	Rhizome	Laguna

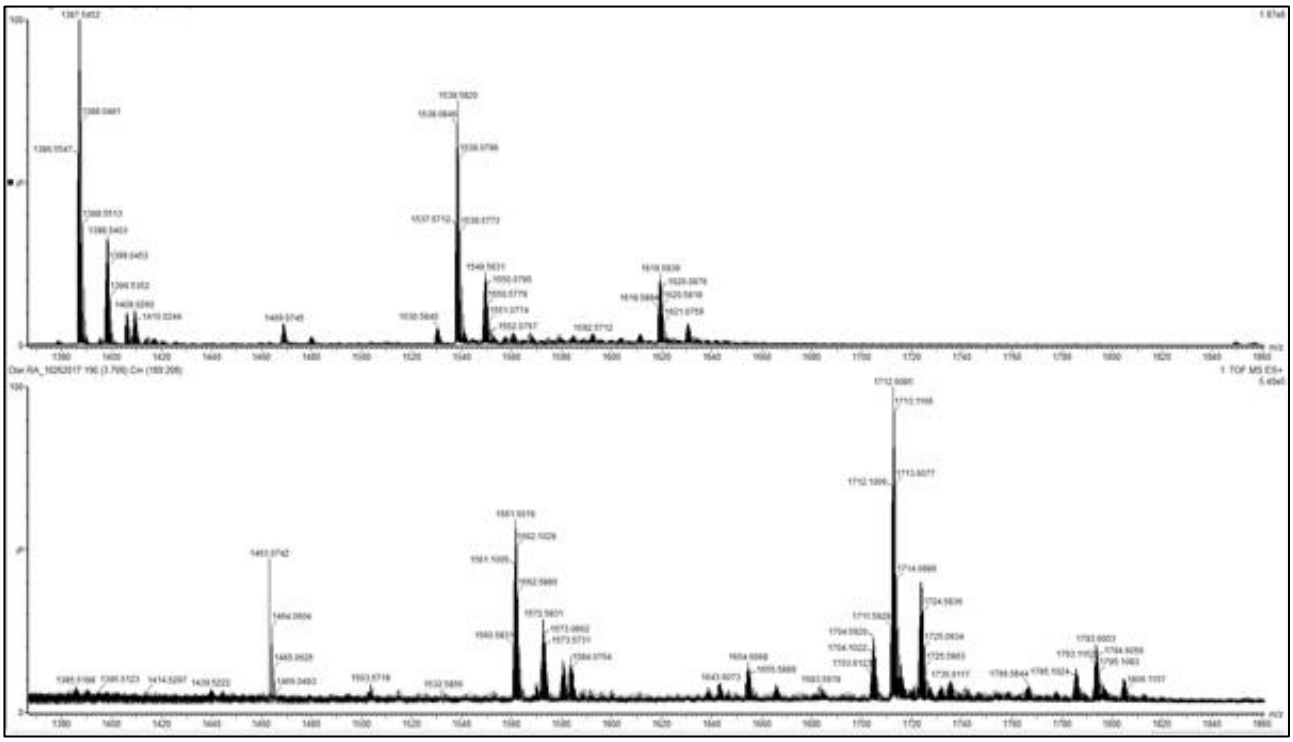
CYC 17-053	*Tinospora rumphii*	Makabuhay stem	Menispermaceae	Stem	Laguna
CYC 17-054	*Vitex negundo* (sample 1)	Lagundi	Verbenaceae	Leaves	Laguna
CYC 17-055	unidentified				Laguna
CYC 17-056	*Aloe barbadensis*	Aloe vera	Liliaceae	Leaves	Laguna
CYC 17-057	*Carmona retusa* (sample 1)	Tsaang gubat	Boraginaceae	Leaves	Laguna
CYC 17-058	*Vitex negundo* (sample 2)	Lagundi	Verbenaceae	Leaves	Laguna
CYC 17-059	*Vitex negundo* (sample 3)	Lagundi	Verbenaceae	Leaves	Laguna
CYC 17-060	*Mentha x cordifolia*	Yerba buena	Lamiaceae	Leaves	Laguna
CYC 17-061	*Gliricidia sepium*	Kakawate	Fabaceae	Leaves	Laguna
CYC 17-062	*Annona muricata*	Guyabano	Annonaceae	Leaves	Laguna
CYC 17-063	*Moringa oleifera*	Malunggay	Moringaceae	Leaves	Laguna
CYC 17-064	*Psidium guajava*	Bayabas	Myrtaceae	Leaves	Laguna
CYC 17-065	*Garcinia mangosteen*	Mangosteen	Guttiferae	Leaves	Laguna
CYC 17-066	*Quisqualis indica*	Niyog-niyugan	Combretaceae	Leaves	Laguna
CYC 17-067	*Annona squamosa*	Atis	Annonaceae	Leaves	Laguna
CYC 17-068	*Peperomia pellucida*	Ulasimang bato	Piperaceae	Leaves	Laguna
CYC 17-069	*Andropogon citratus*	Salay	Gramineae	Leaves	Laguna
CYC 17-070	*Carmona retusa* (sample 2)	Tsaang gubat	Boraginaceae	Leaves	Laguna
CYC 17-071	*Wrightia pubescens*	Lanete	Apocynaceae	Leaves	Metro Manila
CYC 17-072	*Alstonia scholaris*	Dita	Apocynaceae	Leaves	Metro Manila
CYC 17-073	*Tabernaemontana pandacaqui*	Pandakaki-puti	Apocynaceae	Leaves	Metro Manila
CYC 17-074	*Talinum* sp.		Talinaceae	Leaves	Metro Manila
CYC 17-075	*Afzelia rhomboidia*		Fabaceae	Leaves	Metro Manila
CYC 17-076	*Bauhinia purpurea*		Fabaceae	Leaves	Metro Manila
CYC 17-077	*Cassia javanica* ssp. *nodosa*		Fabaceae	Leaves	Metro Manila
CYC 17-078	*Delonix regia*		Fabaceae	Leaves	Metro Manila
CYC 17-079	*Peltophorum petrocarpum*		Fabaceae	Leaves	Metro Manila
CYC 17-080	*Saraca indica*		Fabaceae	Leaves	Metro Manila
CYC 17-081	*Saraca thaipingiensis*		Fabaceae	Leaves	Metro Manila
CYC 17-082	*Senna siamea*		Fabaceae	Leaves	Metro Manila
CYC 17-083	*Senna spectabilis*		Fabaceae	Leaves	Metro Manila
CYC 17-084	*Sindora supa*		Fabaceae	Leaves	Metro Manila
CYC 17-085	*Acacia auriculiformis*		Fabaceae	Leaves	Metro Manila
CYC 17-086	*Acacia mangium*		Fabaceae	Leaves	Metro Manila
CYC 17-087	*Albizia lebekkoides*		Fabaceae	Leaves	Metro Manila
CYC 17-088	*Enterolobium cyclocarpum*		Fabaceae	Leaves	Metro Manila
CYC 17-089	*Parkia timoriana*		Fabaceae	Leaves	Metro Manila
CYC 17-090	*Samanea saman*		Fabaceae	Leaves	Metro Manila
CYC 17-091	*Wallaceodendron celebicum*		Fabaceae	Leaves	Metro Manila
CYC 17-092	*Erythrina* sp.		Fabaceae	Leaves	Metro Manila
CYC 17-093	*Clitoria racemosa*		Fabaceae	Leaves	Metro Manila
CYC 18-001	*Eleusine indica*		Poaceae	Leaves	Metro Manila
CYC 18-002	*Costus speciosus*	Tabubungiaw	Zingiberaceae	Leaves	Metro Manila
CYC 18-003	*Clitoria racemosa*		Fabaceae	Leaves	Metro Manila
CYC 18-004	*Andrographis paniculata*		Acanthaceae	Leaves	Metro Manila

CYC 18-005	*Syzygium* sp.	Mala-bayabas (Tag); Kayo-palawan (Pal)	Myrtaceae	Leaves	Palawan
CYC 18-006	*Macaranga bicolor*	Miligan (Pal)	Euphorbiaceae	Leaves	Palawan
CYC 18-007	*Dillenia* sp.	Mala-katmon (Tag); Tiguyansang (Pal)	Dilleniaceae	Leaves	Palawan
CYC 18-008	*Canarium asperum*	Salong	Burseraceae	Leaves	Palawan
CYC 18-009	*Memecylon* sp.	Sokdan (Pal)	Melastomaceae	Leaves	Palawan
CYC 18-010	*Syzygium* sp. (2)	Lumboy-lumboy (Pal)	Myrtaceae	Leaves	Palawan
CYC 18-011	*Garcinia rubra*	Kandis-kandis (Pal)	Clusiaceae	Leaves	Palawan
CYC 18-012	*Cantium* sp.	Sayuso (Pal)	Rubiaceae	Leaves	Palawan
CYC 18-013	*Mangifera* sp.	Rimaraw (Pal)	Anacardiaceae	Leaves	Palawan
CYC 18-014	*Flemingia* sp.	Dangan-dangan (Pal)	Fabaceae	Leaves	Palawan
CYC 18-015	*Schefflera blancoi*	Pito-pito (Pal)	Araliaceae	Leaves	Palawan
CYC 18-016	*Litsea* sp.	Abuka-abukado (Pal)	Lauraceae	Leaves	Palawan
CYC 18-017	*Dimonorhops molis*	Gatasan (Pal)	Palmae	Leaves	Palawan
CYC 18-018	*Orania* sp.	Banga	Palmae	Leaves	Palawan
CYC 18-019		Amugis (Tag); Sambulawan (Pal)	Anacardiaceae	Leaves	Palawan
CYC 18-020	*Cinnamomum* sp.	Piranod (Pla)	Lauraceae	Leaves	Palawan
CYC 18-021	*Garcinia* sp.	Bunog (Pal)	Clusiaceae	Leaves	Palawan
CYC 18-022	*Actinodaphne* sp.	Puso-puso (Tag)	Lauraceae	Leaves	Palawan
CYC 18-023	*Harpullia* sp.	Buntsikag (Pal)	Sapindaceae	Leaves	Palawan
CYC 18-024	*Hyptis capitata*		Labiatae	Leaves	Palawan
CYC 18-025	*Bauhinia* sp.		Fabaceae	Leaves	Palawan
CYC 18-026	*Ovaria* sp.		Annonaceae	Leaves	Palawan
CYC 18-027	*Casuarina* sp.		Casuarinaceae	Leaves	Palawan
CYC 18-028	*Garcinia binucao*	Kandis	Clusiaceae	Leaves	Palawan
CYC 18-029	*Wallaceodendron celebicum*		Fabaceae	Leaves	Palawan
CYC 18-030	*Pipturus* sp.		Urticaceae	Leaves	Palawan
CYC 18-031	*Sciatea contaminans*	Masok-masok (Pal)	Sciateaceae	Leaves	Palawan
CYC 18-032	*Nepenthes alata*	Pitsel-pitsel	Nepenthaceae	Leaves	Palawan
CYC 18-033	*Melastoma malabathricum*	Antutupo (Pal)	Melastomaceae	Leaves	Palawan
CYC 18-034	*Smilax* sp.	Banag (Pal)	Smilacaceae	Leaves	Palawan
CYC 18-035	*Cratoxylum sumatranum*	Paguringon	Clusiaceae	Leaves	Palawan
CYC 18-036	*Donax canniformis*	Bamban	Marantaceae	Leaves	Palawan
CYC 18-037	*Dioscorea* sp.		Dioscoreaceae	Leaves	Palawan
CYC 18-038	*Calophyllum blancoi*	Bitanghol	Clusiaceae	Leaves	Palawan
CYC 18-039		Buta-buta (Pal)	Rubiaceae	Leaves	Palawan
CYC 18-040	*Calamos* sp.	Maruwa (Pal)	Palmae	Leaves	Palawan
CYC 18-041	*Polyscias nodosa*		Araliaceae	Leaves	Palawan
CYC 18-042	*Sida acuta*		Malvaceae	Leaves	Palawan
CYC 18-043	*Embelia phillipinensis*		Primulaceae	Leaves	Palawan
CYC 18-044	*Lygodium flexuum*	Nito	Schizaeaceae	Leaves	Palawan
CYC 18-045	Unidentified		Apocynaceae	Leaves	Palawan

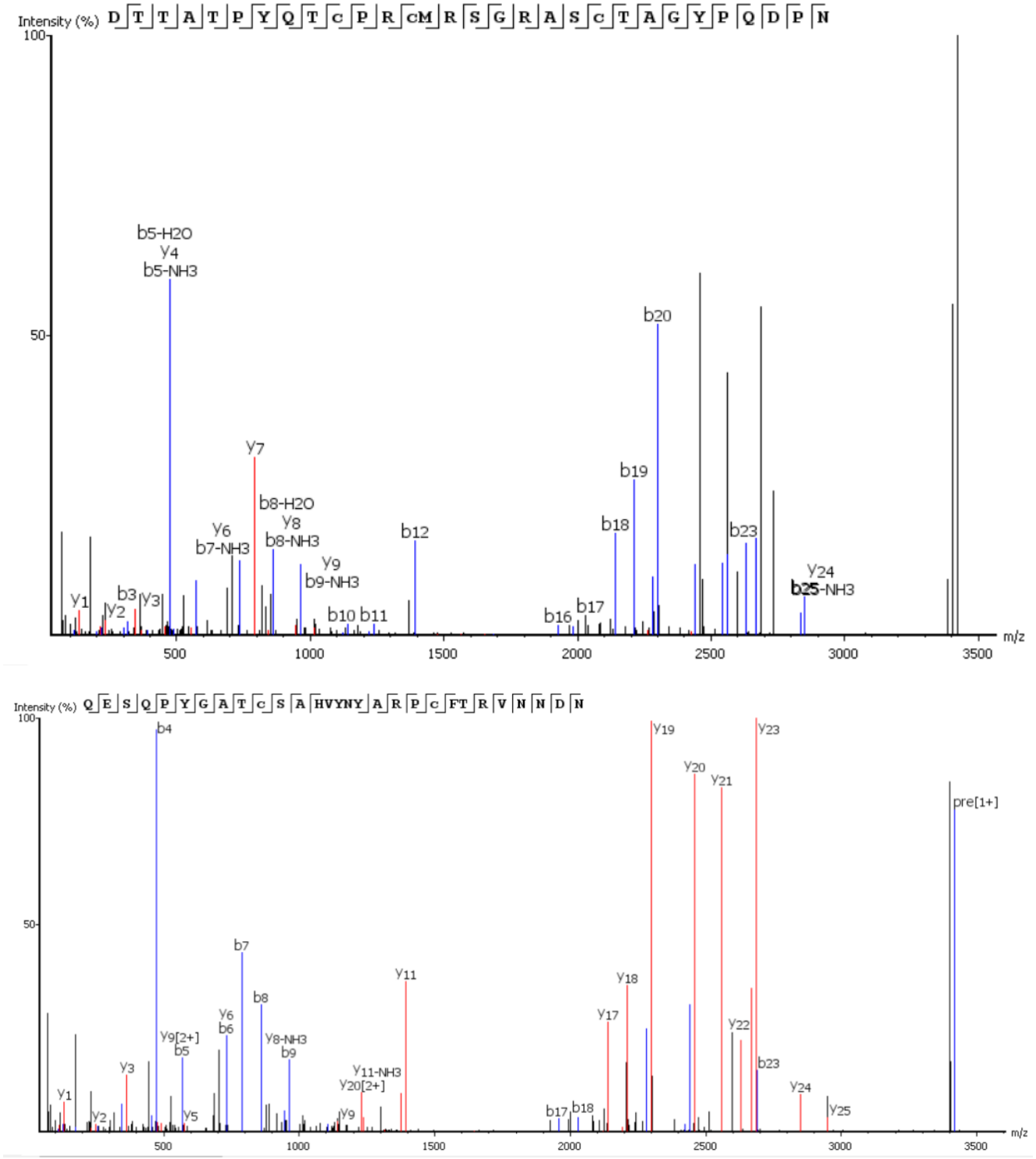
CYC 18-046	*Cratoxylum formosum*		Clusiaceae	Leaves	Palawan
CYC 18-047	Unidentified		--	Leaves	Palawan
CYC 18-048	*Agathis philippinensis*	Almaciga	Araucariaceae	Leaves	Palawan
CYC 18-049	*Viticipremma philippinensis*	Ablas	Lamiaceae	Leaves	Palawan
CYC 18-050	*Micromelum minutum*		Rutaceae	Leaves	Palawan
CYC 18-051	*Barringtonia* sp.	Ulam	Lecythidaceae	Leaves	Palawan
CYC 18-052		Ulos	Phyllanthaceae	Leaves	Palawan
CYC 18-053	*Breynia rhamnoides*	Duruan namog (Pal)	Phyllanthaceae	Leaves	Palawan
CYC 18-054	*Buchanania arborea*	Mangangabo (Pal)	Anacardiaceae	Leaves	Palawan
CYC 18-055	*Dillenia fruticosa/ sibuyanensis/palawanensis*	Labning (Pal)	Dilleniaceae	Leaves	Palawan
CYC 18-056	*Intsia bijuga*	Ipil	Fabaceae	Leaves	Palawan
CYC 18-057	*Antidesma ghaesembilla*	Inyam	Phyllanthaceae	Leaves	Palawan
CYC 18-058	Unidentified	Palpal (Pal)	Fabaceae	Leaves	Palawan
CYC 18-059	*Barringtonia asiatica*	Botong	Lecythidaceae	Leaves	Metro Manila
CYC 18-060	*Phragmites vulgaris*	Tambo	Poaceae	Leaves	Metro Manila
CYC 18-061	*Ardisia elliptica*	Tagpo	Myrsinaceae	Leaves	Metro Manila

Supplementary Table 2: Absorbance ratios of total RNA extracts. Ratios of ~2 are generally accepted as pure RNA.

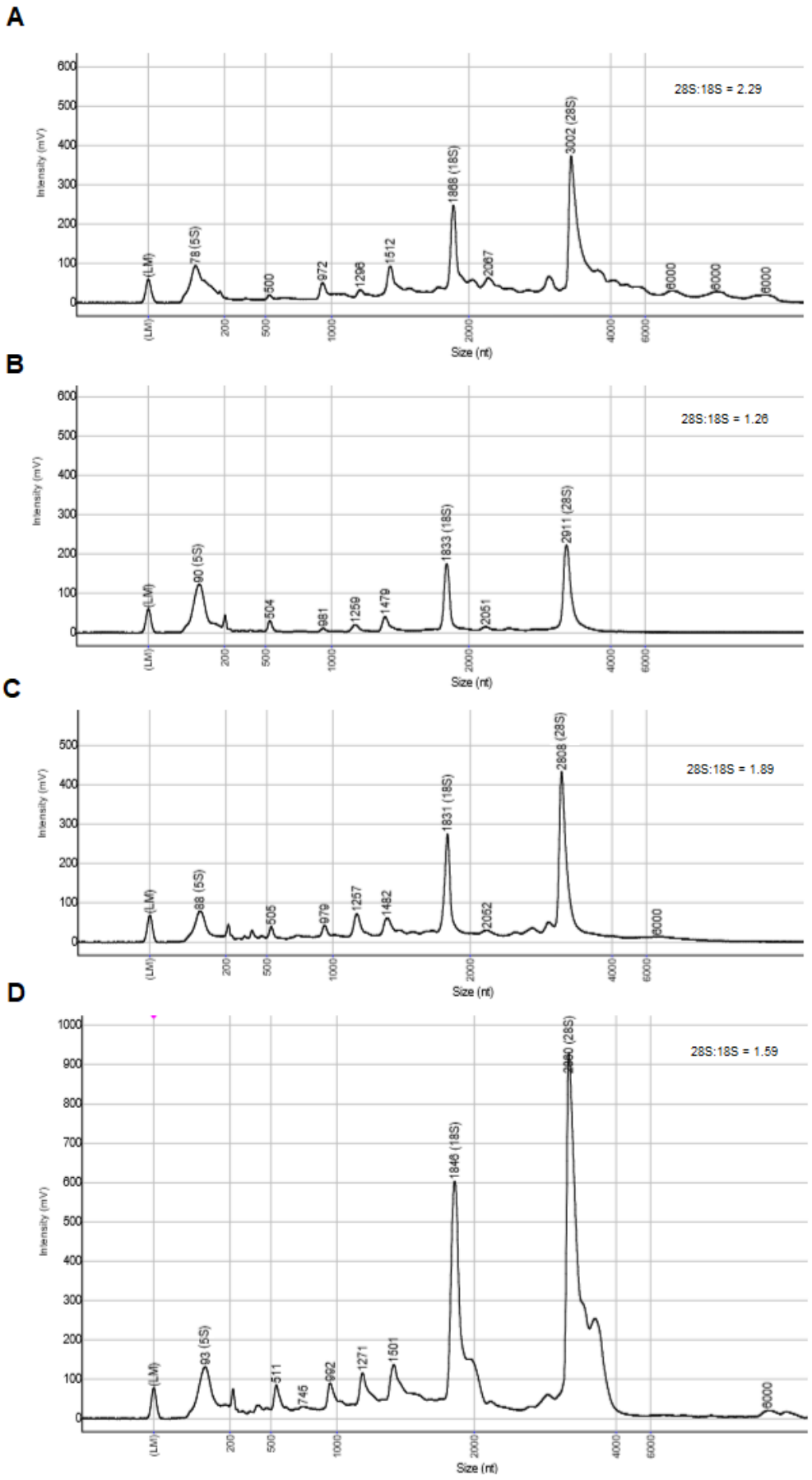
Sample	OD 260/280	OD 260/230
C. ternatea	2.08	2.3325
A. scholaris	1.9925	2.0825
T. pandacaqui	2.0625	2.27
W. pubescens	2.0825	2.35



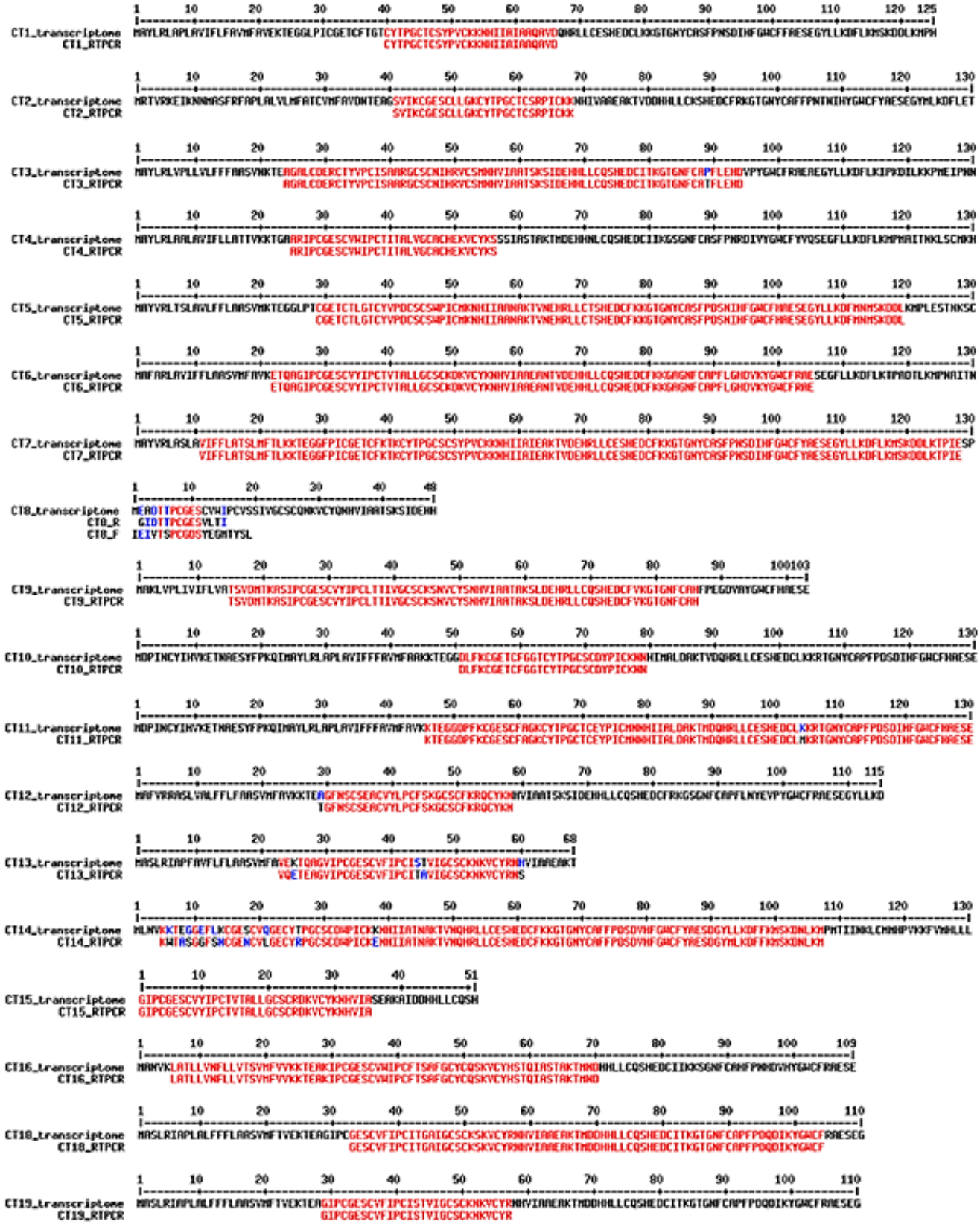
Supplementary Figure 1: Mass spectra of crude (top) and alkylated (bottom) *C. ternatea* extract showing mass shift = + 348 Da.



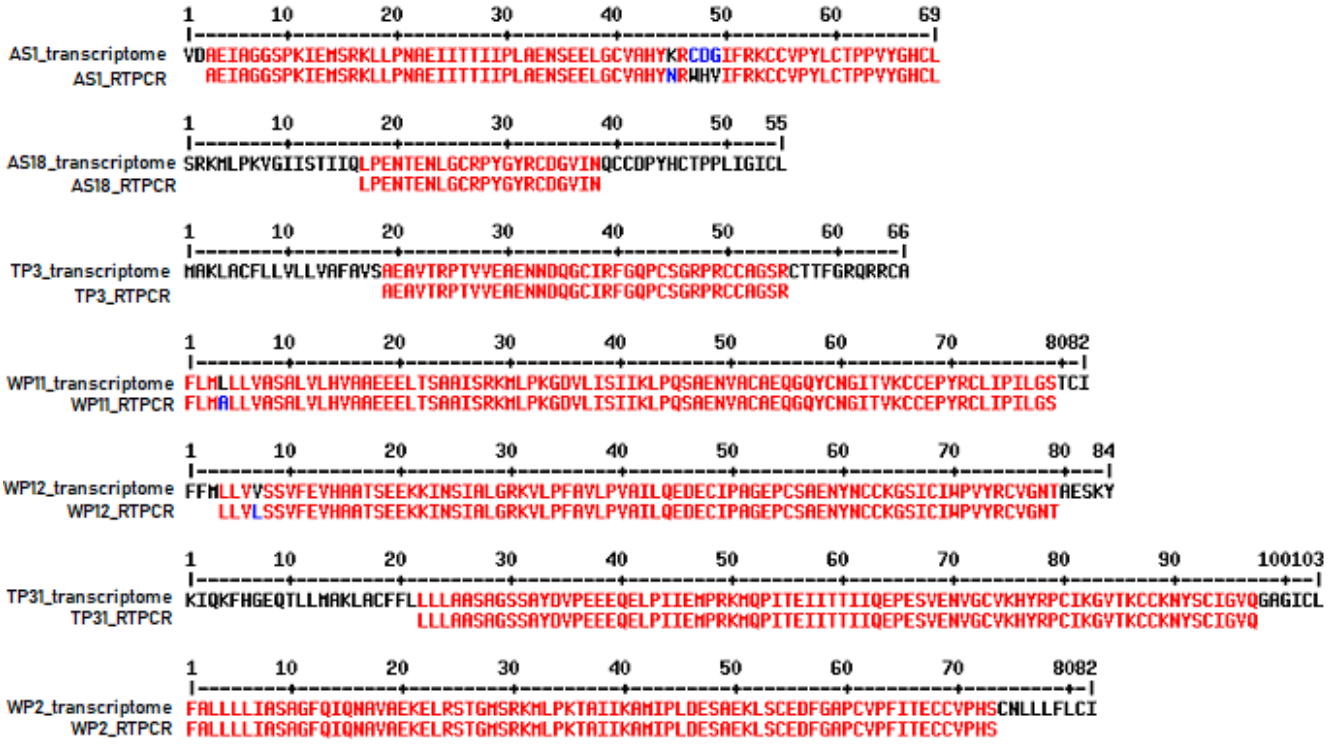
Supplementary Figure 2: Top one (top) and top two (bottom) predicted sequences of cyclotide-like precursor ion (m/z = 684.93; charge = +5; M_{mono} = 3,419.61) from alkylated extract of *T. pandacaqui*. The *denovo* sequencing function of PEAKS Studio software (PEAKS Studio 8.5 Build 20180105) was used to predict the most plausible peptide sequence. The mass tolerance was set to 15 ppm and S-carboxymethylation was considered as a result of Cys alkylation.



Supplementary Figure 3: Electropherogram of total RNA extracts. (A) *C. ternatea*, (B) *A. scholaris*, (C) *T. pandacaqui* and (D) *W. pubescens* extracts have 28S:18S rRNA ratios of at least 1 and moderate 5S rRNA peaks indicating intact RNA suitable for transcriptome sequencing.



Supplementary Figure 4: Alignment of the cyclotide sequences mined from the transcriptome and the sequences from RT-PCR. The translated cyclotide domain sequences amplified from the cDNA (RT-PCR) have high similarities with those mined from the transcriptome. For CT8, the sequences labeled CT8_F and CT8_R correspond to the forward and reverse sequence results from capillary sequencing. Sequences in red indicate 100% identity while sequences in blue show a low consensus value of 50%. Sequence alignment was performed using Multalin (Corpet, 1988).



Supplementary Figure 5: Alignment of the knottin sequences mined from the transcriptome and the sequences from RT-PCR. The translated knottin domain sequences amplified from the cDNA (RT-PCR) have high similarities with those mined from the transcriptome. Sequences in red indicate 100% identity while sequences in blue show a low consensus value of 50%. Sequence alignment was performed using Multalin (Corpet, 1988).