Mining raw plant transcriptomic data for new cyclopeptide alkaloids

Draco Kriger‡, Michael A. Pasquale‡, Brigitte G. Ampolini‡ and Jonathan R. Chekan*

Full Research Paper

Address:
Department of Chemistry and Biochemistry, University of North Carolina at Greensboro, Greensboro, NC, USA

Email:
Jonathan R. Chekan* - jrchekan@uncg.edu

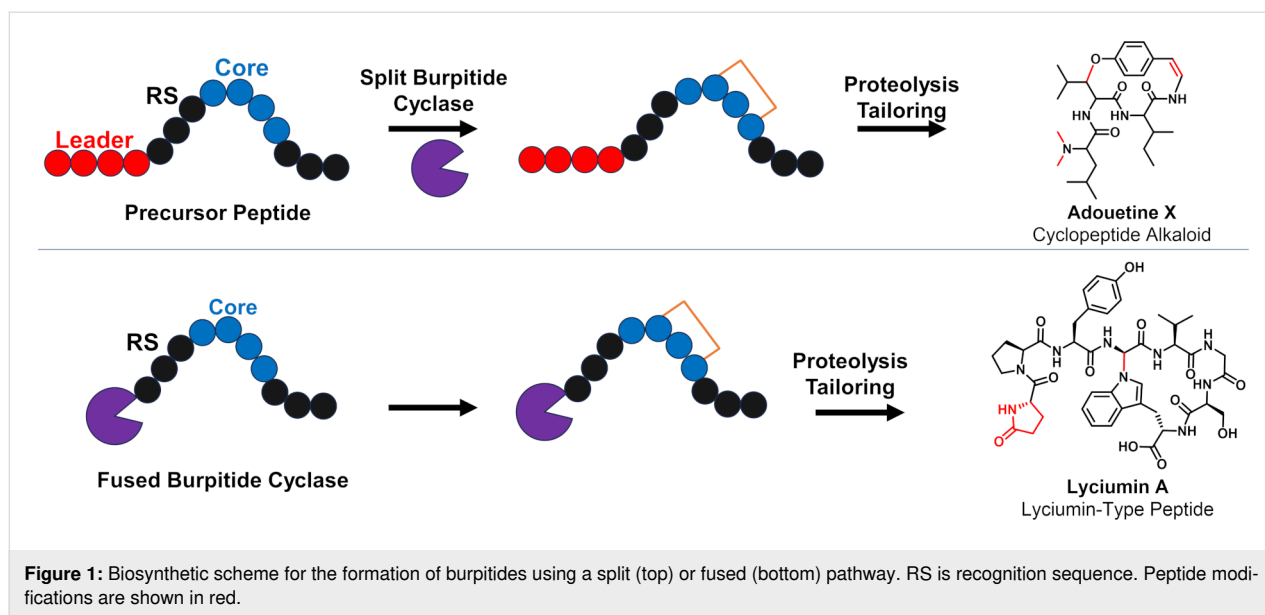* Corresponding author   ‡ Equal contributors

Abstract

In recent years, genome and transcriptome mining have dramatically expanded the rate of discovering diverse natural products from bacteria and fungi. In plants, this approach is often more limited due to the lack of available annotated genomes and transcriptomes combined with a less consistent clustering of biosynthetic genes. The recently identified burpitide class of ribosomally synthesized and post-translationally modified peptide (RiPP) natural products offer a valuable opportunity for bioinformatics-guided discovery in plants due to their short biosynthetic pathways and gene encoded substrates. Using a high-throughput approach to assemble and analyze 700 publicly available raw transcriptomic data sets, we uncover the potential distribution of split burpitide precursor peptides in Streptophyta. Metabolomic analysis of target plants confirms our bioinformatic predictions of new cyclopeptide alkaloids from both known and new sources.

Introduction

Plants are prolific producers of cyclic peptide natural products, making 1000s of different molecules [1]. While the orbitide [2] and cyclotide [3] classes of peptides are well known, it has been recently discovered that a new class of molecules called burpitides are also prevalent in plants [4]. Like all known plant peptides, burpitides fall under the ribosomally synthesized and post-translationally modified peptides (RiPPs) superclass of natural products (Figure 1). The typical pathway for a RiPP starts with a precursor peptide made by the ribosome that undergoes post-translational modifications by specific tailoring en-

zymes [5,6]. This precursor peptide substrate can be subdivided into multiple segments including 1) an N-terminal leader or recognition sequence used for binding by the tailoring enzymes and 2) a core peptide that is targeted for modification by the biosynthetic enzymes. Ultimately proteolysis releases the modified core peptide as the mature RiPP natural product [5,6].

In the case of the newly described burpitide family of RiPPs, the defining feature is the presence of amino acid side-chain crosslinks installed by a copper-dependent burpitide cyclase [4].

Figure 1: Biosynthetic scheme for the formation of burpitides using a split (top) or fused (bottom) pathway. RS is recognition sequence. Peptide modifications are shown in red.

These RiPP natural products encompass a wide range of structural scaffolds including cyclopeptide alkaloids that contain a phenolic ether-linkage and the lyciumin-type peptides that are composed of a crosslink between the Trp-indole-N and the carbon in another amino acid side chain or peptide backbone (Figure 1). Recent enzymatic reconstitution has demonstrated that the burpitide cyclases can function either autocatalytically (fused) or as traditional stand-alone proteins with separate free peptide substrates (split) [4,7-11]. When fully matured, burpitides possess a wide range of bioactivities including analgesic, sedative, and cytotoxic [12-14].

In order to identify new potential burpitides, we searched the NCBI Sequence Read Archive (SRA) for Streptophyta (all land plants and most green algae) transcriptomes that could be assembled and examined with a custom hidden Markov model (HMM) for split burpitide precursor peptides. Using this approach, we assembled 700 transcriptomes and used predicted precursor peptide diversity for target discovery of new cyclopeptide alkaloids. Additionally, metabolomic analysis revealed the tissue distribution of these molecules across select members of the Rubiaceae and Rhamnaceae families. Ultimately, our results lay the groundwork for the rapid discovery of new burpitides across diverse plant species.

Results and Discussion
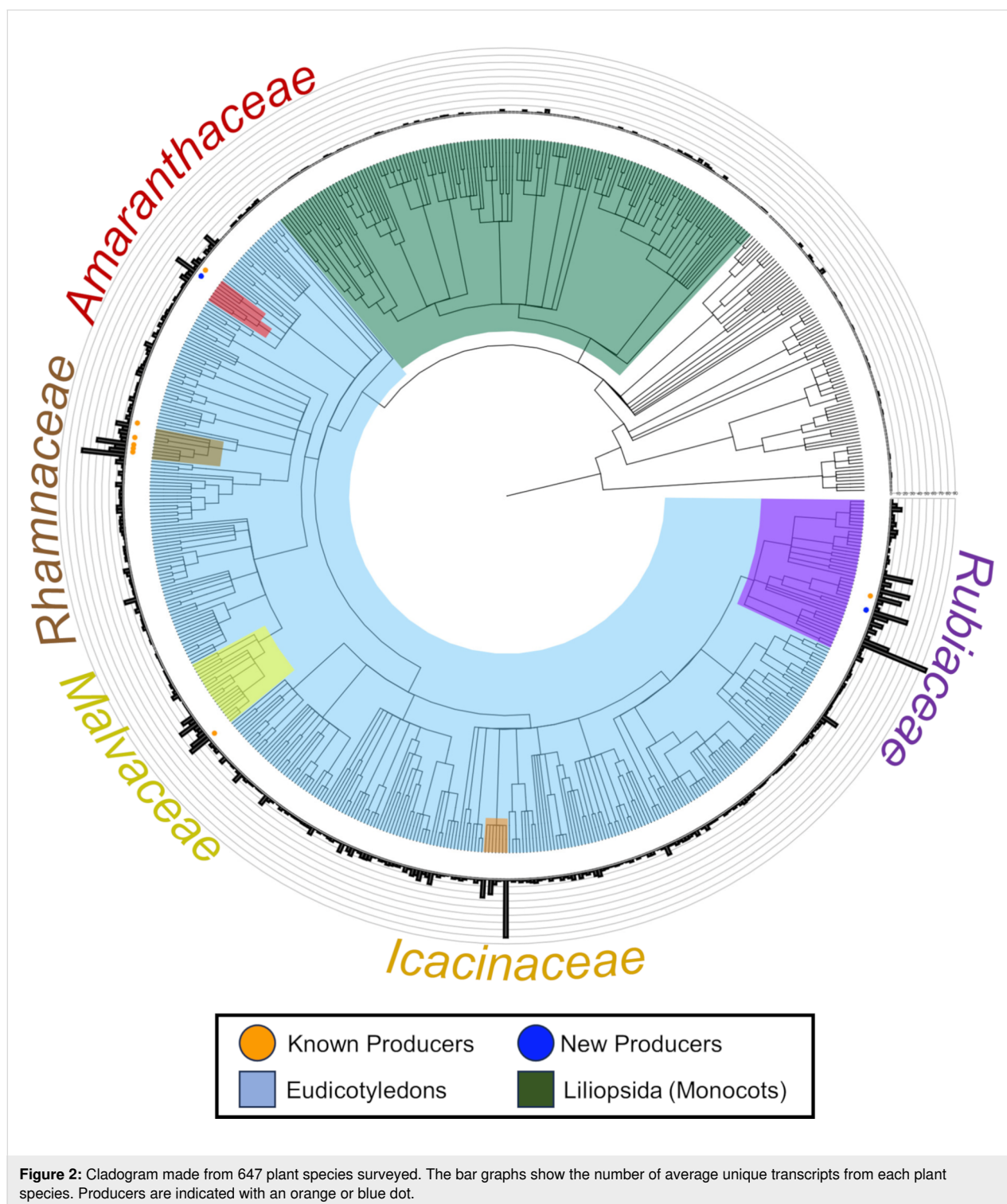## Assembly and analysis of 700 public transcriptomes

In order to efficiently download, assemble, and analyze raw SRA data in a high throughput manner, we created an automated pipeline. Raw Illumina sequencing data were downloaded using the NCBI SRA Toolkit, the transcriptome was then assembled with rnaSPAdes [15], and the open reading frames were detected with TransDecoder [16,17]. To search for putative precursor peptides for split burpitide pathways, we developed a custom HMM (see Supporting Information File 1, Figure S1). This approach gave us the speed, sensitivity, and customization needed to identify diverse sequences. We targeted a diversity of plants from every available order in Streptophyta for a total of 700 transcriptomes from 647 species (Table in Supporting Information File 2).

Potential split burpitide precursor peptide transcripts were found in 66% of the assemblies using the custom HMM. However, due to the only recent discovery of split burpitides, the rules that dictate the presence of a candidate precursor peptide and observation of a corresponding small molecule are unclear. For example, precursors rich in S/TxY core motifs are prevalent in nature, but do not correspond to a known molecule [7]. Therefore, we sought to rapidly identify the most promising candidates. This presented a challenge as the publicly available data was generated using different protocols, from different plant tissues, and with different levels of sequencing depth. Ultimately, we reasoned that the best candidate plants for prolific production of burpitides should have numerous unique precursor peptide transcripts which may arise from multiple gene copies, multiple core sequences, or multiple RNA splicing events. Therefore, we counted the number of translated transcripts from each assembly that scored better than an E-value of 0.1 using our custom precursor peptide HMM. To select for only the split precursor peptides, we removed any transcript that had contained a BURP domain as determined by the public HMM model (PFAM 03181) with an E-value of 0.1 inclusion

threshold. The BURP domain itself is named after the four founding members BNM2, USP-like, RD22, and PG1β, which were bioinformatically described in 1998 [18]. This domain is typically around 300 amino acids in length and has a conserved $CHX_{10}CHX_{25\text{-}27}CHX_{25\text{-}26}CH$ motif which comprises the active site of the burpitide cyclase. The resulting number of

stand-alone transcripts from this filtering step were mapped onto a cladogram of the 647 species we surveyed (Figure 2 and Supporting Information File 3). For species with multiple transcriptomes, the number of precursor peptides was averaged (Table in Supporting Information File 2). We further mapped known producers with a corresponding precursor peptide onto

Figure 2: Cladogram made from 647 plant species surveyed. The bar graphs show the number of average unique transcripts from each plant species. Producers are indicated with an orange or blue dot.

the cladogram (orange dots). Using a cut-off of 30 unique transcripts, there were five clear phylogenetic hot spots for the production of burpitides. These families were examined in detail for their composition of core peptide sequences.
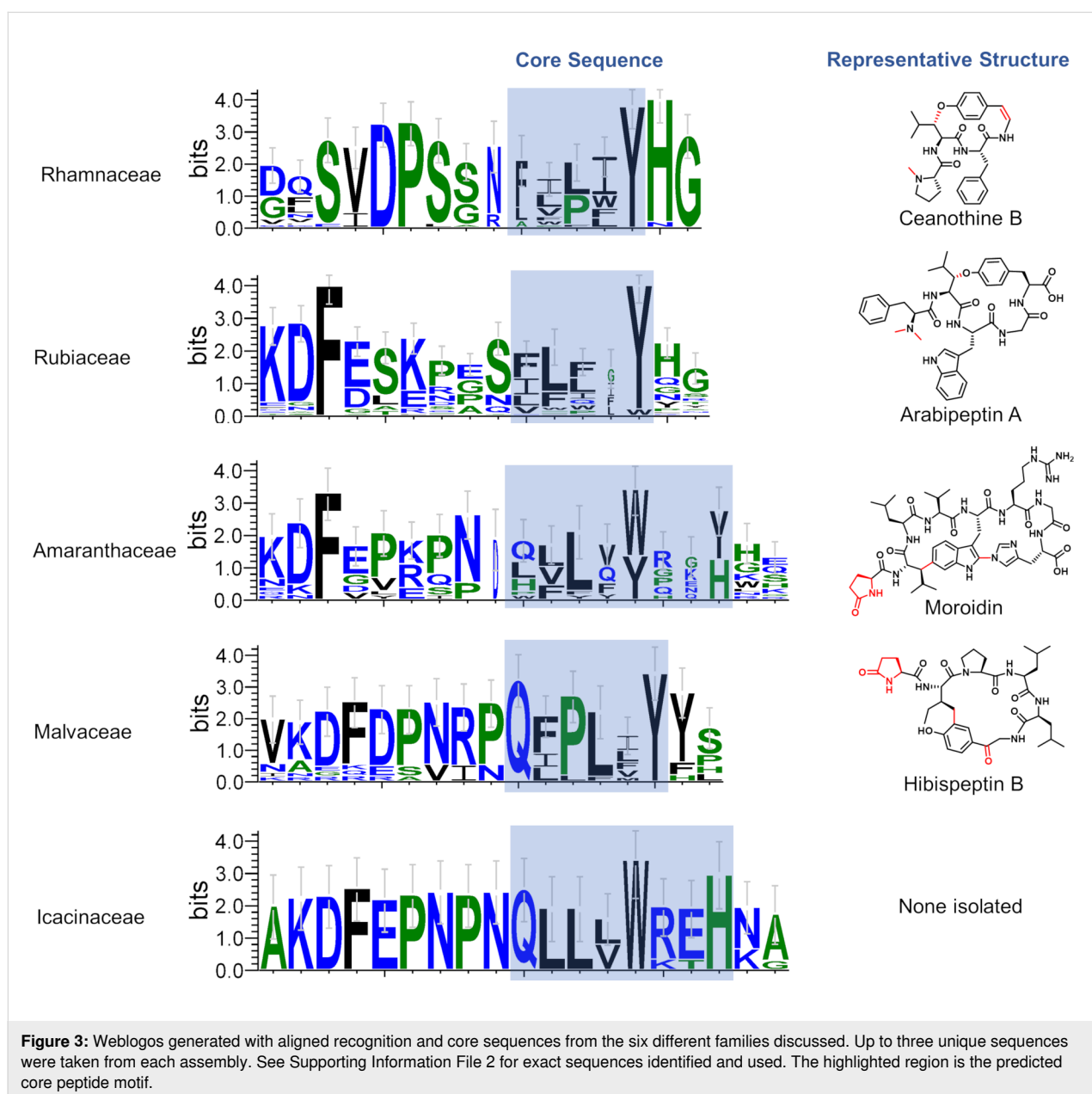
Rhamnaceae family

The Rhamnaceae family is amongst the most iconic producers of cyclopeptide alkaloids, containing both *Ziziphus* and *Ceanothus* genera. In addition to the subclass defining Tyr-phenol-O to carbon linkage, the cyclopeptide alkaloids of this family are typically oxidatively decarboxylated and *N*-methylated (Figure 1, adouetine X and Figure 3, ceanothine B) [19-21]. Our transcriptome analysis supports the prevalence of cyclopeptide

alkaloids in Rhamnaceae with multiple species transcribing 20+ potential split precursor peptides. Alignment of the predicted recognition and core sequences shows the core peptides are 4–5 amino acids in length and appear to exclusively encode for cyclopeptide alkaloids due to the invariant C-terminal Tyr residue (Figure 3).

Rubiaceae family

Our previous results revealed *Coffea arabica* (Arabica coffee plant), to be a cyclopeptide alkaloid producer [7]. In contrast to those from the Rhamnaceae family, these cyclopeptide alkaloids were only cyclized and lack an oxidative decarboxylation of the C-terminus (Figure 3, arabipeptin A). Our transcriptome

Figure 3: Weblogos generated with aligned recognition and core sequences from the six different families discussed. Up to three unique sequences were taken from each assembly. See Supporting Information File 2 for exact sequences identified and used. The highlighted region is the predicted core peptide motif.

mining results further support the prevalence of cyclopeptide alkaloids in the Rubiaceae family. *Chiococca alba* (Snowberry), *Cinchona calisaya* (Yellow Cinchona), *Gardenia jasminoides* (Cape Jasmine), *C. arabica*, and *Coffea eugenioides* (Eugenioides coffee plant) all contained >30 unique transcripts for potential precursor peptides. The core sequences themselves were typically 4 or 5 amino acids in length and most often contained a tyrosine for cyclization. However, a few sequences appeared to have a tryptophan at the position for cyclization, indicating they may correspond to the stephanotic acid-type burpitides like moroidin (Trp-indole-C to carbon crosslink, Figure 3) [4].

Amaranthaceae family

The Amaranthaceae family is home to the known moroidin producer, *Celosia argentea* var. *cristata* (Cockscomb and previously *Celosia cristata*) (Figure 3) [22]. Transcriptome analysis revealed that many other species in this family and the recently merged Chenopodioideae subfamily [23] have the capacity to produce burpitides as well. For example, *Alternanthera bettzickiana* (Calico Plant) was rich in 31 potential unique precursor peptide sequences. Examination of possible cores suggested that the producers closely related to *C. argentea* (*A. bettzickiana* and *Amaranthus tricolor* [edible amaranth] are likely to make bicyclic burpitides but may vary in the cycle size and identity of the cyclized residues (Figure 3 and Supporting Information File 2).

Malvaceae family

The Malvaceae family contains *Hibiscus syriacus* (Rose of Sharon), the only known producer of hibispeptin-type burpitides (Figure 3, hibispeptin B) [24,25]. These molecules contain a C–C linkage between the phenol derived from tyrosine and the γ-carbon of isoleucine. As expected from previous genomic analysis [7], the transcriptome of *H. syriacus* contained precursor peptides corresponding to both hibispeptin A and B (Figure 3 and Supporting Information File 2). The closely related *Hibiscus cannabinus* possessed similar transcripts, while *Hibiscus mutabilis* (Confederate rose) appeared to code for core peptides that were only four amino acids in length and not clearly hibispeptins. Outside of the *Hibiscus* genera, *Quararibea asterolepisn, Kosteletzkya pentacarpos* (Coastal Mallow)*,* and *Talipariti hamabo* all contained core peptides indicating hibispeptin-type burpitide production.

Icacinaceae family

In addition to these known families of producers, our transcriptomic results suggest production of burpitides from the Icacinaceae family. In particular, *Merrilliodendron megacarpum* produced 81 unique transcripts, while *Pyrenacantha malvifolia* (Monkey chair) produced more than 20. Analysis of the pre-

dicted core sequence seemed to suggest moroidin-like molecules (Figure 3), but no cyclic peptides have been isolated from any member of this family. Unfortunately, sourcing members of this small family proved challenging and we were unable to evaluate these hypotheses.

Iconic moroidin producers contain split precursor peptides

The moroidin family of plant peptides has previously been shown to be biosynthesized from a fused burpitide pathway using a transcriptome mining approach [9]. However, precursor peptides for the well-known producers of moroidin (*Dendrocnide moroides* [Gympie stinger] and *C. argentea*) [22,26] and the structurally related celogentin molecules (*C. argentea* var. *cristata*) [27,28] remained elusive. Using the publicly available sequencing data from *D. moroides* and *C. argentea* var. *cristana*, we could identify core sequences that perfectly matched the structure of moroidin (Figure 4). Furthermore, manually searching for the "QLLVWRGH" core sequence in the full transcriptome only revealed the stand-alone precursor peptides, further suggesting the biosynthetic pathways contain split burpitide cyclases in these plants. This seems to suggest that both split and fused burpitide biosynthetic pathways may be responsible for producing the stephanotic acid sub-family of RiPPs that is characterized by the Trp-indole-C to carbon linkage [4].
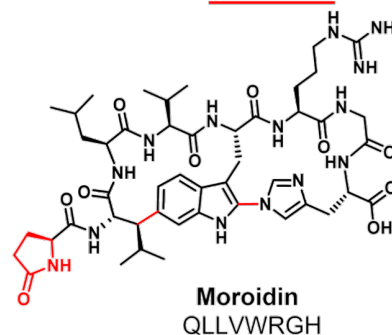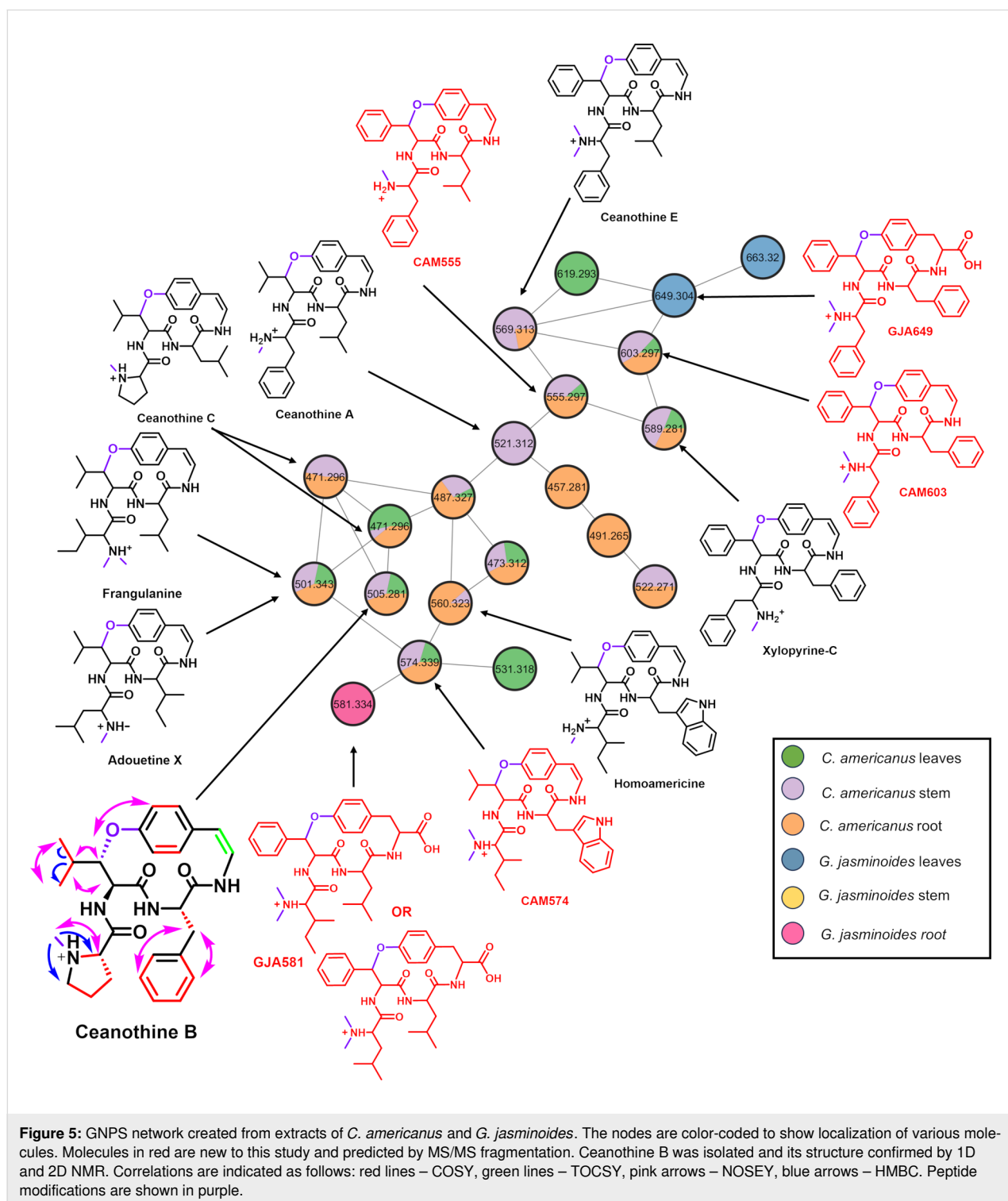

Figure 4: Core peptide sequences from the putative precursor peptides map to moroidin.

Molecular networking of known and potential producers

The transcriptome analysis indicated that many additional plant species may be responsible for the production of burpitides. To evaluate this, we examined *G. jasminoides* and *A. bettzickiana* from the Rubiaceae and Amaranthaceae families, respectively.

We analyzed methanol extracts of these plants along with the known cyclopeptide alkaloid producer *C. ameranicus* using UHPLC-HRMS/MS. Finally, we generated a global natural product social (GNPS) network to show the correlation between metabolites from these different plant species (Figure 5) [29].

Figure 5: GNPS network created from extracts of *C. americanus* and *G. jasminoides*. The nodes are color-coded to show localization of various molecules. Molecules in red are new to this study and predicted by MS/MS fragmentation. Ceanothine B was isolated and its structure confirmed by 1D and 2D NMR. Correlations are indicated as follows: red lines – COSY, green lines – TOCSY, pink arrows – NOSEY, blue arrows – HMBC. Peptide modifications are shown in purple.

Cyclopeptides in *Ceanothus americanus*

The GNPS network showed the presence of multiple features from *C. americanus* that appeared to match known molecules such as ceanothine B (505.2809 *m/z*), ceanothine C (471.2966 *m/z*), ceanothine A (521.3122 *m/z*), ceanothine E (569.3128 *m/z*) fragulanine/adouetine X (501.3435 *m/z*), and homo-americine (560.3231 *m/z*). Despite being first isolated over 50 years ago [30-32], the majority of the proposed structures for these known cyclopeptide alkaloids are simply predicted by MS/MS fragmentation. Therefore, we isolated ceanothine B and completed full structural characterization using NMR, Marfey's analysis, and MS/MS fragmentation (Supporting Information File 1, Figure S2, Figures S16–S26, Table S1). Indeed, our results support the proposed structure of ceanothine B while also assigning absolute stereochemistry consistent with all ʟ-amino acids (Figure 5).
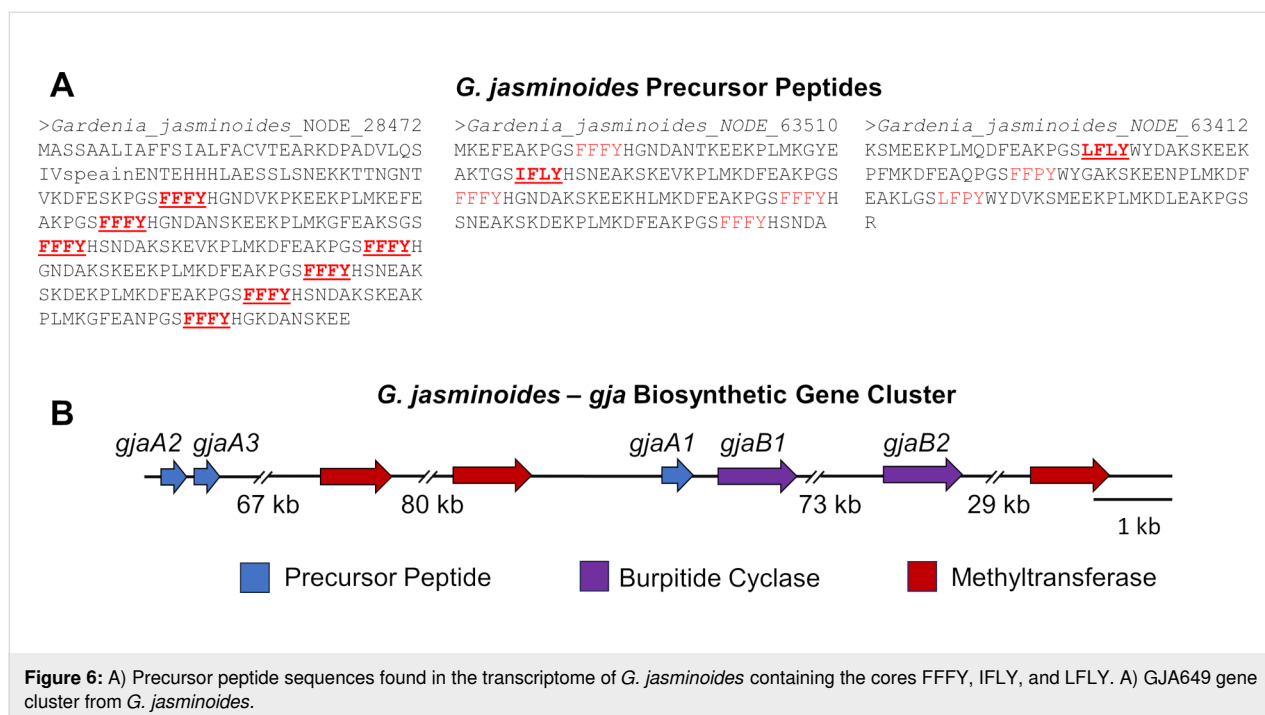
We next explored the GNPS network for new cyclopeptide alkaloids from *C. americanus*. We previously noted the presence of a feature as CAM603 (603.29 *m/z*) [7]. Fragmentation indicated a core sequence of FFFY that was observed in precursor peptides found in the transcriptome (Supporting Information File 1, Figure S3). To further support this assignment, we partially purified the compound and completed Marfey's analysis, which confirmed the presence of all ʟ-Phe in the structure (Supporting Information File 1, Figure S27). Beyond this, multiple other features in the GNPS network appeared to represent new molecules for *C. americanus* (Supporting Information File 1, Figures S4–S11). Xylopyrine-C was previously isolated

from *Zizyphus xylopyra* [33] and matches the exact mass and fragmentation of the 589.2820 *m/z* feature. The 555.2953 *m/z* feature (CAM555) appears to be a mono-methylated version of ceanothine E. Likewise, the 574.3396 *m/z* (CAM574) feature is consistent with a methylated analogue of homoamericine.

New cyclopeptide alkaloids in *Gardenia jasminoides*

The GNPS network also suggested the presence of new cyclopeptide alkaloids from *G. jasminoides*, a previously unknown producer (Figure 5). This data corroborates the transcriptome mining data where *G. jasminoides* was shown to possess an average of 50 unique precursor peptide transcripts. The 649.3047 *m/z* feature, which we named GJA649, had an exact mass and fragmentation suggested a molecule with an FFFY core (Supporting Information File 1, Figure S12). Indeed, the transcript data assembled as part of this study identified multiple transcripts from *G. jasminoides* that directly match this FFFY sequence (Figure 6A). Like other members of the Rubiaceae family [7,34], the tyrosine that forms the ether linkage remains intact, and not oxidatively decarboxylated as seen in members of the Rhamnaceae family. It is also predicted to contain a dimethylation at the N-terminal Phe residue. Additional 663.3192 and 581.3345 *m/z* features were also present in the GNPS network, but MS/MS fragmentation could not definitively support a specific structure.

For a better understanding of the biosynthesis of burpitides in *G. jasminoides*, we explored the genomic context by using the

Figure 6: A) Precursor peptide sequences found in the transcriptome of *G. jasminoides* containing the cores FFFY, IFLY, and LFLY. A) GJA649 gene cluster from *G. jasminoides*.

burpitide cyclase sequence and precursor peptides found during transcriptome assembly by performing a tblastn search against the unannotated *G. jasminoides* genome (NCBI GenBank assembly: GCA_013103745.1). Gene matches were primarily found on chromosome 4 (NCBI GenBank assembly: CM023098.1) with the location of the biosynthetic gene cluster (BGC) on the chromosome between base pairs 8,210,000 and 8,480,000. The BGC contained precursor peptides with the core FFFY (*gjaA1*), FLFY (*gjaA2*), and FLLY (*gjaA3),* three potential methyltransferases, and two complete burpitide cyclases all within 266,286 base pairs (Figure 6B). While the proposed gene cluster is missing a potential peptidase, it contains all the other proteins necessary for the biosynthesis of GJA649 and the other putative *G. jasminoides* cyclopeptide alkaloids.

New cyclopeptide alkaloids in *Alternanthera bettzickiana*

Based on the assembled transcriptomic data sets, it appeared that *A. bettzickiana* had the capacity to produce new cyclopeptide alkaloids. To evaluate this, a methanol extract of *A. bettzickiana* was prepared and analyzed by UHPLC-HRMS/MS. Subsequent GNPS networks did not show any clustering with cyclopeptide alkaloids from either *C. americanus* or *G. jasminoides*. Instead, the MS/MS data was searched for fragmentation patterns representing peptides, namely iminium ions [8]. Both the exact mass and fragment ions show evidence for three potential compounds with alpha-keto groups that matched putative core sequences (Figure 7 and Supporting Information File 1, Figures S13–S15). While unusual, this motif could be biosynthetically related to the alpha-hydroxy modification seen in vignatic acid [35]. These potential compounds are bicyclic which resemble the bicyclic motif found in moroidin from the closely related *C. argentea* (Figure 4). Unlike moroidins which are part of the stephanotic acid subclass of burpitides defined by the Trp-indole-C to carbon crosslink, the molecules from the *A. bettzickiana* seems to be cyclopeptide alkaloids with two tyrosine derived ether linkages that resemble selanine A isolated from African clubmoss (*Selaginella kraussiana*) [8]. To confirm these proposed structures from *A. bettzickiana*, full elucidation by NMR will be required.

Localization of cyclopeptide alkaloids in plants

Traditionally, cyclopeptide alkaloids have been isolated from the roots and bark of the producing plants [1,19-21], likely due to the abundance of these materials. With more sensitive metabolomics techniques, we can better identify where these molecules are located in the plant. For the described GNPS network, live samples of *C. americanus* and *G. jasminoides* were processed by separating the leaves, stems, and roots. With
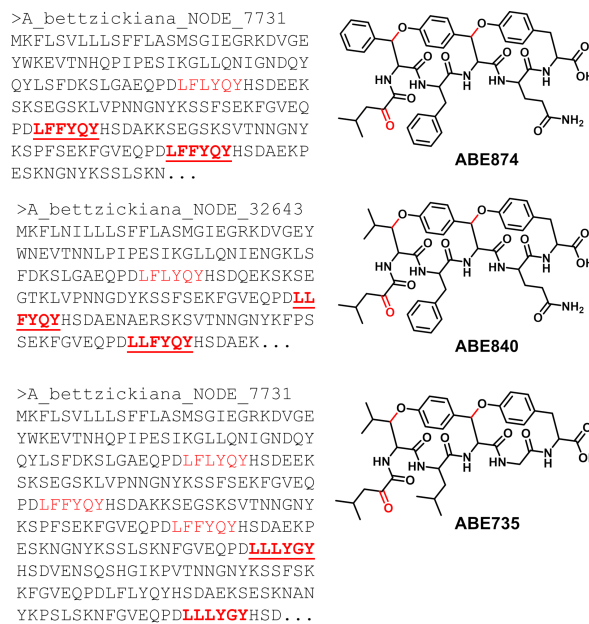


```
>A_bettzickiana_NODE_7731
MKFLSVLLLSFFLASMSGIEGRKDVGE
YWKEVTNHQPIPESIKGLLQNIGNDQY
QYLSFDKSLGAEQPDLFLYQYHSDEEK
SKSEGSKLVPNNGNYKSSFSEKFGVEQ
PDLFFYQYHSDAKKSEGSKSVTNNGNY
KSPFSEKFGVEQPDLFFYQYHSDAEKP
ESKNGNYKSSLSKN...

>A_bettzickiana_NODE_32643
MKFLNILLLSFFLASMGIEGRKDVGEY
WNEVTNNLPIPESIKGLLQNIENGKLS
FDKSLGAEQPDLFLYQYHSDQEKSKSE
GTKLVPNNGDYKSSFSEKFGVEQPDLL
FYQYHSDAENAERSKSVTNNGNYKFPS
SEKFGVEQPDLLFYQYHSDAEK...

>A_bettzickiana_NODE_7731
MKFLSVLLLSFFLASMSGIEGRKDVGE
YWKEVTNHQPIPESIKGLLQNIGNDQY
QYLSFDKSLGAEQPDLFLYQYHSDEEK
SKSEGSKLVPNNGNYKSSFSEKFGVEQ
PDLFFYQYHSDAKKSEGSKSVTNNGNY
KSPFSEKFGVEQPDLFFYQYHSDAEKP
ESKNGNYKSSLSKNFGVEQPDLLLYGY
HSDVENSQSHGIKPVTNNGNYKSSFSK
KFGVEQPDLFLYQYHSDAEKSESKNAN
YKPSLSKNFGVEQPDLLLYGYHSD...
```

ABE874

ABE840

ABE735

Figure 7: Core sequences from the putative precursor peptides map to three predicted products from *A. bettizickiana* observed by LC–MS/MS.

few exceptions, cyclopeptides alkaloids from *C. americanus* were found distributed throughout multiple tissues in the plant and not only localized to the roots (Figure 5). In contrast, the cyclopeptide alkaloids in G. *jasminoides* appear to be exclusively present in either the leaves or root. As the biological functions of these molecules are currently unknown, their localization may give insights into their physiological purpose.

Conclusion

Overall, this study illustrates the ability to rapidly assemble and analyze public transcriptomic data for the targeted discovery of new burpitide natural products from plants. Phylogenetic visualization of the potential precursor peptides validated families known to make these cyclic peptides, but also suggested new producers. Finally, metabolomic analyses supported the predicted new sources of burpitides, aiding in our understanding of the precursor rulesets for the biosynthesis of this emerging class of plant natural products.

Experimental
Construction of precursor peptide HMM

A custom hidden Markov model designed to target the precursor peptides from split burpitide pathways was created to favor producers of known burpitides. Sequences containing cores matching known molecules such as hibispeptins, moroidins,

jubanines, and lyciumins were selected for the new model (see Supporting Information). The sequences were aligned and the hmm build function of HMMER v3.3.2 was used to generate the model [36]. The HMM is available in Supporting Information File 4 compressed archive.

Transcriptome mining and cladogram construction

Representatives of every available order of Streptophyta were selected from the NCBI Sequence Read Archive. The full list of accessions and species is available in the Supporting Table (Supporting Information File 2). rnaSPAdes was used to assemble the raw reads using the default settings [15]. This software package was chosen primarily for its speed ($\approx$1.5 h per transcriptome), which was essential for this large analysis effort. Our custom HMM was used to search for potential precursor peptides using an E-value inclusion threshold of 0.1. These results are present in Supporting Information File 4 compressed archive file. To remove fused burpitide cyclases, the HMM for the BURP domain (PFAM 03181) was also run against the assembled transcriptomes using the same 0.1 E-value inclusion threshold. Transcriptomes that indicated a BURP domain were removed from the putative precursor peptide list. The resulting precursor peptides were counted and mapped as bar graphs onto a cladogram of the species phylogeny. phyloT was used to generate the cladogram and R-studio and ggplot2 were used to visualize it with the number of transcripts [37,38]. For our analysis, the previously distinct Chenopodioideae family was viewed as members of the Amaranthaceae family based on the most up to date Angiosperm Phylogeny Group classification (APG IV) [23].

Isolation of burpitides from *C. americanus*

Three separate extractions were performed overnight with methanol using 200 g of ground root powder (Mountain Rose Herb Lot# 26356) and 1.5 L of methanol. After overnight extraction, vacuum filtration was used to separate the solids from the liquids. The methanol extract was dried and saved. The solids were extracted with methanol two more times. After the extractions were finished and dried by rotary evaporation, the extract was resuspended in basic water (pH 10 using NaOH) and vacuum filtered to remove insoluble particulates. Liquid–liquid extraction was performed three times with basic water and DCM. The DCM layer was collected and dried. Afterward, the extract was resuspended in 0.1% formic acid in water.

A Combiflash EZ prep system was used to purify the burpitides from *C. americanus*. Water with 0.1% formic acid (A) and methanol (B) were used as solvents for a reversed phase column (C18 21 × 250 mm Kinetix 5 μm C18 100 Å (00G-4601-P0-AX5) with a 30–80% B gradient over 50 min at 10 mL/min. Roughly 100 mg of sample was loaded using 5 mL of a 25% MeOH/H$_2$O solution. The fractions were tested using UHPLC-HRMS/MS for purity and presence of desired burptidies. Ceanothine B was successfully isolated using this approach. For CAM603, an additional purification step using an HPLC system (Agilent) with a C18 column (C18 250 × 4.6 mm Luna Omega 5 5 μm C18 100 Å) and gradient of 55–70% B was employed.

Structure elucidation of ceanothine B

Ceanothine B was analyzed by 1D and 2D NMR in methanol-d_4 in 5 mm NMR tubes (Wilmad LabGlass). NMR instrumentation was a JEOL ECA-500 MHz NMR spectrometer (JEOL Ltd.). ^{1}H NMR (500 MHz, methanol-d_4) δ 7.17 (m, 2H), 7.12 (m, 2H), 7.09 (m, 2H), 7.04 (m, 1H), 7.01 (m, 1H), 6.98 (m, 1H), 6.96 (m, 1H), 6.73 (d, *J* = 7.5 Hz, 1H), 6.19 (d, *J* = 7.0 Hz, 1H), 4.80 (dd, *J* = 2.0 Hz, *J* = 8.0 Hz, 1H), 4.42 (d, *J* = 8.0 Hz, 1H), 4.42 (m, 1H), 3.05 (m, 1H), 2.85 (d, *J* = 13.5 Hz, 1H), 2.74 (dd, *J* = 4.0 Hz, *J* = 10.0 Hz, 1H), 2.59 (t, *J* = 12.5 Hz, 1H), 2.33 (m, 1H), 2.16 (m, 1H), 1.97 (s, 1H), 1.95 (m, 1H), 1.79 (m, 1H), 1.68 (m, 2H), 1.15 (d, *J* = 7.0 Hz, 1H), 0.926 (d, *J* = 7.0 Hz, 1H)

Ceanothine B macrocyclization between Tyr-3 and Leu-β is supported by HMBC correlation between Tyr-C3 (δ 156.3 ppm) and Tyr-C4 (δ120.5 ppm) and NOESY correlation between Tyr4-H (δ 7.01 ppm, 1H) and Leuβ-H (δ 4.80 ppm, 1H). Similar to other characterized cyclopeptide alkaloids, *J*-coupling constant ($J_{\alpha\text{-}\beta}$ = 8.0 Hz) supports *S*-stereochemistry at the β-carbon of leucine ether linkage. Structurally related cyclopeptide alkaloids have similar *J*-values ($J_{\alpha\text{-}\beta} \approx$ 8.0 Hz) and have been assigned as L-erythro (*anti*) conformation [20,39].

Metabolomic analysis and GNPS networking

Live samples of *C. americanus* and *G. jasminoides* were separated into leaves, stems, and roots and lyophilized for extraction. *G. jasminoides* samples were prepared by extracting 89.24 g of dried leaf material over three days with 1 L of methanol each day yielding 19.10 g of crude extract. The root and stem extracts were prepared by extracting 180.6 mg and 145.2 mg of plant material, respectively, in methanol overnight. *A. bettzickiana* was lyophilized as the whole plant. Each sample was extracted separately in methanol overnight and the extract was dried by rotary evaporation. All the residues were diluted in methanol to 2 mg/mL, except for the *G. jasminoides* leaves (1.5 mg/mL). UHPLC-HRMS/MS data was collected using an Q Exactive Plus (Thermo Scientific) coupled to a Acquity UPLC (Waters), and Acquity UPLC BEH 1.7 μm C18

reversed phase 130 Å 2.1 × 50 mm column. The following gradient was used at 1 mL/min: 0 min, 5% B; 1 min, 5% B; 11 min 100% B; 12 min, 100% B; 12.1 min, 5% B; 13 min, 5% B where solvent A was water with 0.1% of formic acid and solvent B was acetonitrile with 0.1% of formic acid. A top 5 HCD method at 25 NCE (200–2000 *m/z*) was used on each plant tissue extract. See Supporting Information File 1, Table S2 for complete acquisition parameters. The mass spectra were used to generate a GNPS network using the following parameters: 0.2 Da precursor ion mass tolerance, 0.5 Da fragment ion tolerance, and a minimum of 8 matched fragment ions.

Marfey's analysis

To determine the stereochemistry of the new molecules, Marfey's analysis was used. Standards were prepared to match the core amino acids of the predicted ceanothine B and CAM603 cyclopeptide alkaloids. To make each amino acid standard, 0.2 mg of the L- and D-amino acid was aliquoted into separate reaction vials. To each vial, 50 μL of water, 20 μL of 1 M NaHCO₃, and 100 μL 1% Marfey's reagent (sodium (2,4-dinitro-5-fluorophenyl)-L-alaninamide) in acetone were added. The reaction mixture was incubated at 40 ℃ for 1 h with periodic agitation. The reactions were quenched by adding 10 μL of 2 M HCl and dried under a nitrogen stream. Dried samples were dissolved in 1.7 mL of MeOH and injected both individually and as a combined mixture (5 μL) onto UPLC-MS (Thermo Orbitrap Elite mass spectrometer coupled to a Dionex Ultimate 3000 UPLC), with a gradient of 15–50% B over 10 min on a Kinetex 1.7 μm C18 100 Å 50 × 2.1 mm LC column using solvents of A: 0.1% formic acid in water and B: 0.1% formic acid in acetonitrile. Products were observed at 340 nm and monitored by their expected [M + H]⁺ ion.

To hydrolyze the isolated burpitides, 0.2 mg of sample was aliquoted into reaction vial. To each vial, 0.5 mL of 6 M HCl was added. The sample was incubating for 24 h at 90 ℃ and dried under a nitrogen stream. Afterward, 25 μL of water, 10 μL of 1 M NaHCO₃ and 50 μL of 1% Marfey's reagent in acetone were added to the reaction vial. The reaction was incubated for 1 h at 40 ℃ with periodic agitation. The reaction was quenched by adding 5 μL of 2 M HCl and dried under a nitrogen stream. The dried peptides were resuspended in 200 μL of MeOH. The sample was injected (6 μL) onto UPLC-MS using the same conditions and instrument as the amino acids standards. To normalize the retention time, 1 μM of fluorescein was added to the standard and peptide samples before injection onto the UPLC-MS. The retention time for the fluorescein sample was adjusted to 6.30 min to correct for run to run drifts in retention time.

Supporting Information

The sequencing data used in this study is openly available in the NCBI SRA.

Supporting Information File 1
Additional details and figures including NMR spectra and MS/MS fragmentation.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-138-S1.pdf]

Supporting Information File 2
Specific dataset used are listed in this Supporting Table File along with core peptide sequence alignments.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-138-S2.xlsx]

Supporting Information File 3
The full cladogram with species names as a high-resolution pdf.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-138-S3.pdf]

Supporting Information File 4
The split burpitide precursor peptide HMM and output from all 700 transcriptomes as a compressed archive.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-138-S4.zip]

Author Contributions

Draco Kriger: data curation; investigation; writing – original draft; writing – review & editing. Michael A. Pasquale: conceptualization; investigation; writing – original draft; writing – review & editing. Brigitte G. Ampolini: conceptualization; software; writing – original draft; writing – review & editing. Jonathan R. Chekan: conceptualization; data curation; funding acquisition; software; supervision; visualization; writing – original draft; writing – review & editing.

ORCID® iDs

Michael A. Pasquale - https://orcid.org/0000-0002-5480-6355
Brigitte G. Ampolini - https://orcid.org/0000-0002-6328-3306
Jonathan R. Chekan - https://orcid.org/0000-0003-4651-0594

Data Availability Statement

All data that supports the findings of this study is available in the published article and/or the supporting information to this article. The data used for/in this study is openly available in NCBI SRA. Specific dataset used are listed in Supporting Table.

References

1. Tan, N.-H.; Zhou, J. *Chem. Rev.* **2006,** *106,* 840–895. doi:10.1021/cr040699h
2. Shim, Y. Y.; Young, L. W.; Arnison, P. G.; Gilding, E.; Reaney, M. J. T. *J. Nat. Prod.* **2015,** *78,* 645–652. doi:10.1021/np500802p
3. de Veer, S. J.; Kan, M.-W.; Craik, D. J. *Chem. Rev.* **2019,** *119,* 12375–12421. doi:10.1021/acs.chemrev.9b00402
4. Chekan, J. R.; Mydy, L. S.; Pasquale, M. A.; Kersten, R. D. *Nat. Prod. Rep.* **2024,** in press. doi:10.1039/d3np00042g
5. Montalbán-López, M.; Scott, T. A.; Ramesh, S.; Rahman, I. R.; van Heel, A. J.; Viel, J. H.; Bandarian, V.; Dittmann, E.; Genilloud, O.; Goto, Y.; Grande Burgos, M. J.; Hill, C.; Kim, S.; Koehnke, J.; Latham, J. A.; Link, A. J.; Martínez, B.; Nair, S. K.; Nicolet, Y.; Rebuffat, S.; Sahl, H.-G.; Sareen, D.; Schmidt, E. W.; Schmitt, L.; Severinov, K.; Süssmuth, R. D.; Truman, A. W.; Wang, H.; Weng, J.-K.; van Wezel, G. P.; Zhang, Q.; Zhong, J.; Piel, J.; Mitchell, D. A.; Kuipers, O. P.; van der Donk, W. A. *Nat. Prod. Rep.* **2021,** *38,* 130–239. doi:10.1039/d0np00027b
6. Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K.-D.; Fischbach, M. A.; Garavelli, J. S.; Göransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Müller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J. T.; Rebuffat, S.; Ross, R. P.; Sahl, H.-G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Süssmuth, R. D.; Tagg, J. R.; Tang, G.-L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. A. *Nat. Prod. Rep.* **2013,** *30,* 108–160. doi:10.1039/c2np20085f
7. Lima, S. T.; Ampolini, B. G.; Underwood, E. B.; Graf, T. N.; Earp, C. E.; Khedi, I. C.; Pasquale, M. A.; Chekan, J. R. *Angew. Chem., Int. Ed.* **2023,** *62,* e202218082. doi:10.1002/anie.202218082
8. Chigumba, D. N.; Mydy, L. S.; de Waal, F.; Li, W.; Shafiq, K.; Wotring, J. W.; Mohamed, O. G.; Mladenovic, T.; Tripathi, A.; Sexton, J. Z.; Kautsar, S.; Medema, M. H.; Kersten, R. D. *Nat. Chem. Biol.* **2022,** *18,* 18–28. doi:10.1038/s41589-021-00892-6
9. Kersten, R. D.; Mydy, L. S.; Fallon, T. R.; de Waal, F.; Shafiq, K.; Wotring, J. W.; Sexton, J. Z.; Weng, J.-K. *J. Am. Chem. Soc.* **2022,** *144,* 7686–7692. doi:10.1021/jacs.2c00014
10. Lee, Y.-Y.; Guler, M.; Chigumba, D. N.; Wang, S.; Mittal, N.; Miller, C.; Krummenacher, B.; Liu, H.; Cao, L.; Kannan, A.; Narayan, K.; Slocum, S. T.; Roth, B. L.; Gurevich, A.; Behsaz, B.; Kersten, R. D.; Mohimani, H. *Nat. Commun.* **2023,** *14,* 4219. doi:10.1038/s41467-023-39905-4
11. Mydy, L. S.; Hungerford, J.; Chigumba, D. N.; Konwerski, J. R.; Jantzi, S. C.; Wang, D.; Smith, J. L.; Kersten, R. D. *Nat. Chem. Biol.* **2024,** *20,* 530–540. doi:10.1038/s41589-024-01552-1
12. Trevisan, G.; Maldaner, G.; Velloso, N. A.; da Silva Sant'Anna, G.; Ilha, V.; de Campos Velho Gewehr, C.; Rubin, M. A.; Morel, A. F.; Ferreira, J. *J. Nat. Prod.* **2009,** *72,* 608–612. doi:10.1021/np800377y
13. Kaleem, W. A.; Muhammad, N.; Qayum, M.; Khan, H.; Khan, A.; Aliberti, L.; De Feo, V. *Fitoterapia* **2013,** *91,* 154–158. doi:10.1016/j.fitote.2013.08.024
14. Xu, X.; Jiang, N.; Liu, S.; Jin, Y.; Cheng, Y.; Xu, T.; Wang, X.; Liu, Y.; Zhang, M.; Du, S.; Fan, J.; Zhang, A. *J. Nat. Prod.* **2022,** *85,* 1918–1927. doi:10.1021/acs.jnatprod.1c01215
15. Bushmanova, E.; Antipov, D.; Lapidus, A.; Prjibelski, A. D. *GigaScience* **2019,** *8,* giz100. doi:10.1093/gigascience/giz100
16. Prjibelski, A.; Antipov, D.; Meleshko, D.; Lapidus, A.; Korobeynikov, A. *Curr. Protoc. Bioinf.* **2020,** *70,* e102. doi:10.1002/cpbi.102
17. Haas, B. J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P. D.; Bowden, J.; Couger, M. B.; Eccles, D.; Li, B.; Lieber, M.; MacManes, M. D.; Ott, M.; Orvis, J.; Pochet, N.; Strozzi, F.; Weeks, N.; Westerman, R.; William, T.; Dewey, C. N.; Henschel, R.; LeDuc, R. D.; Friedman, N.; Regev, A. *Nat. Protoc.* **2013,** *8,* 1494–1512. doi:10.1038/nprot.2013.084
18. Hattori, J.; Boutilier, K. A.; Lookeren Campagne, M. M.; Miki, B. L. *Mol. Gen. Genet.* **1998,** *259,* 424–428. doi:10.1007/s004380050832
19. Gournelis, D. C.; Laskaris, G. G.; Verpoorte, R. Cyclopeptide Alkaloids. *Fortschritte der Chemie organischer Naturstoffe/Progress in the chemistry of organic natural products.;* Springer: Vienna, Austria, 1998; Vol. 75, pp 1–179. doi:10.1007/978-3-7091-6507-2_1
20. Gournelis, D. C.; Laskaris, G. G.; Verpoorte, R. *Nat. Prod. Rep.* **1997,** *14,* 75–82. doi:10.1039/np9971400075
21. Tuenter, E.; Exarchou, V.; Apers, S.; Pieters, L. *Phytochem. Rev.* **2017,** *16,* 623–637. doi:10.1007/s11101-016-9484-y
22. Morita, H.; Shimbo, K.; Shigemori, H.; Kobayashi, J. *Bioorg. Med. Chem. Lett.* **2000,** *10,* 469–471. doi:10.1016/s0960-894x(00)00029-9
23. The Angiosperm Phylogeny Group; Chase, M. W.; Christenhusz, M. J. M.; Fay, M. F.; Byng, J. W.; Judd, W. S.; Soltis, D. E.; Mabberlay, D. J.; Sennikov, A. N.; Soltis, P. S.; Stevens, P. F. *Bot. J. Linn. Soc.* **2016,** *181,* 1–20. doi:10.1111/boj.12385
24. Yun, B.-S.; Ryoo, I.-J.; Lee, I.-K.; Yoo, I.-D. *Tetrahedron Lett.* **1998,** *39,* 993–996. doi:10.1016/s0040-4039(97)10797-3
25. Yun, B.-S.; Ryoo, I.-J.; Lee, I.-K.; Yoo, I.-D. *Tetrahedron* **1998,** *54,* 15155–15160. doi:10.1016/s0040-4020(98)00967-3
26. Christina Leung, T.-W.; Williams, D. H.; Barna, C. J.; Foti, S.; Oelrichs, P. B. *Tetrahedron* **1986,** *42,* 3333–3348. doi:10.1016/s0040-4020(01)87397-x
27. Kobayashi, J.; Suzuki, H.; Shimbo, K.; Takeya, K.; Morita, H. *J. Org. Chem.* **2001,** *66,* 6626–6633. doi:10.1021/jo0103423
28. Suzuki, H.; Morita, H.; Iwasaki, S.; Kobayashi, J. *Tetrahedron* **2003,** *59,* 5307–5315. doi:10.1016/s0040-4020(03)00762-2

29. Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.;
    Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.;
    Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.;
    Meehan, M. J.; Liu, W.-T.; Crüsemann, M.; Boudreau, P. D.;
    Esquenazi, E.; Sandoval-Calderón, M.; Kersten, R. D.; Pace, L. A.;
    Quinn, R. A.; Duncan, K. R.; Hsu, C.-C.; Floros, D. J.; Gavilan, R. G.;
    Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.;
    Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.;
    Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C.-C.;
    Yang, Y.-L.; Humpf, H.-U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.;
    Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.;
    Larson, C. B.; Boya P, C. A.; Torres-Mendoza, D.; Gonzalez, D. J.;
    Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.;
    O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.;
    Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.;
    Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.;
    Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.;
    Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.;
    Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.;
    Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodríguez, A. M. C.;
    Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.;
    Allard, P.-M.; Phapale, P.; Nothias, L.-F.; Alexandrov, T.; Litaudon, M.;
    Wolfender, J.-L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D.-T.;
    VanLeer, D.; Shinn, P.; Jadhav, A.; Müller, R.; Waters, K. M.; Shi, W.;
    Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. Ø.;
    Pogliano, K.; Linington, R. G.; Gutiérrez, M.; Lopes, N. P.;
    Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N.
    Nat. Biotechnol. **2016,** *34,* 828–837. doi:10.1038/nbt.3597
30. Servis, R. E.; Kosak, A. I.; Tschesche, R.; Frohberg, E.;
    Fehlhaber, H. W. *J. Am. Chem. Soc.* **1969,** *91,* 5619–5624.
    doi:10.1021/ja01048a032
31. Warnhoff, E. W.; Pradhan, S. K.; Ma, J. C. N. *Can. J. Chem.* **1965,** *43,*
    2594–2602. doi:10.1139/v65-356
32. Klein, F. K.; Rapoport, H. *J. Am. Chem. Soc.* **1968,** *90,* 2398–2404.
    doi:10.1021/ja01011a033
33. Singh, A. K.; Pandey, M. B.; Singh, V. P.; Pandey, V. B.
    J. Asian Nat. Prod. Res. **2008,** *10,* 715–718.
    doi:10.1080/10286020802015828
34. Onozawa, T.; Kitajima, M.; Kogure, N.; Peerakam, N.; Santiarworn, D.;
    Takayama, H. *J. Nat. Prod.* **2017,** *80,* 2156–2160.
    doi:10.1021/acs.jnatprod.7b00290
35. Sugawara, F.; Ishimoto, M.; Le-Van, N.; Koshino, H.; Uzawa, J.;
    Yoshida, S.; Kitamura, K. *J. Agric. Food Chem.* **1996,** *44,* 3360–3364.
    doi:10.1021/jf960166c
36. Finn, R. D.; Clements, J.; Arndt, W.; Miller, B. L.; Wheeler, T. J.;
    Schreiber, F.; Bateman, A.; Eddy, S. R. *Nucleic Acids Res.* **2015,** *43,*
    W30–W38. doi:10.1093/nar/gkv397
37. Letunic, I. phyloT : a phylogenetic tree generator.
    https://phylot.biobyte.de/ (accessed June 23, 2022).
38. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis;* Springer:
    Berlin, Germany, 2016.
39. González Sierra, M.; Mascaretti, O. A.; Diaz, F. J.; Rúveda, E. A.;
    Chang, C.-J.; Hagaman, E. W.; Wenkert, E.
    J. Chem. Soc., Chem. Commun. **1972,** 915–916.
    doi:10.1039/c39720000915

License and Terms