

Non-targeted Screening of Natural Products from 288 Fungal Endophytes from Canadian Fruit Crops

Natasha DesRochers, Justin B. Renaud, Joey B. Tanney, Ashraf Ibrahim, Ken K.-C. Yeung, and Mark W. Sumarah*



Cite This: *ACS Omega* 2023, 8, 24561–24572



Read Online

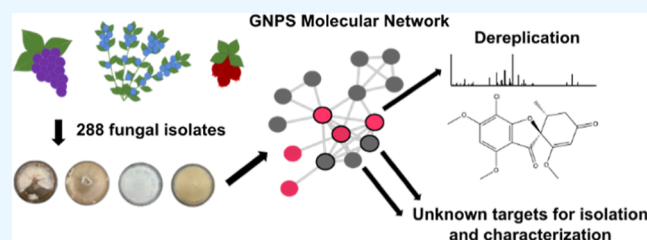
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Many diverse species of fungi naturally occur as endophytes in plants. The majority of these fungi produce secondary metabolites of diverse structures and biological activities. Culture extracts from 288 fungi isolated from surface-sterilized blueberries, cranberries, raspberries, and grapes were analyzed by LC–HRMS/MS. Global Natural Products Social (GNPS) Molecular Networking modeling was used to investigate the secondary metabolites in the extracts. This technique increased the speed and simplicity of dereplicating the extracts, targeting new compounds that are structurally related. In total, 60 known compounds were dereplicated from this collection and seven new compounds were identified. These previously unknown compounds are targets for purification, characterization, and bioactivity testing in future studies. The fungal endophytes characterized in this study are potential candidates for providing bio-protection to the host plant with a reduced reliance on chemical pesticides.



INTRODUCTION

Fungi that live in plant tissue without causing disease are called “endophytes”. Many plants from grasses to conifers to some seaweeds have formed mutualisms with fungi. In some cases, the benefit to the plant is clear in terms of the fungal metabolites reducing herbivory or fungal disease.¹ Natural products from fungal endophytes are of particular interest to study as they exhibit a wide range of bioactivities. Some endophyte metabolites have been used in medical applications in their pure form, such as topical emodepside, for the treatment of nematode infections in cats.^{2,3} They may also be used to inoculate cool season fescues or white spruce with their respective endophytes that produce anti-insectan compounds that deter harmful pests.⁴

To better understand the endophyte species diversity that exists in Canadian fruit crops, we isolated fungal endophytes from highbush and lowbush blueberries (248 isolates from *Vaccinium angustifolium* and *Vaccinium corymbosum*), grapes (14 isolates from *Vitis vinifera*), cranberries (18 isolates from *Vaccinium macrocarpon*), pear (1 isolate from *Pyrus communis*), and raspberries (7 isolates from *Rubus idaeus*). Between 2011 and 2015, nearly 300 strains of fungi were isolated from the leaves and stems of these plant samples and were identified at the species level, where possible. Building on this work, there is a need to explore the natural products produced by these strains to better understand their role in the host-endophyte relationship.

To date, our preliminary investigations within this collection resulted in the discovery of several novel and bioactive

secondary metabolites including the antifungal polyketides trienylfuranones from the raspberry endophyte *Hypomontagnella submonticulosa*,⁵ the antibacterial non-ribosomal peptides ellisiamides from the blueberry-*Pinus* endophyte *Xylaria ellisii*,^{6,7} and the antimicrobial polyketides nemanilactones and nemanifuranones from the grape endophyte tentatively identified as *Nemania serpens*.⁸ However, only a fraction of the species within the collection have been investigated for their ability to produce novel compounds. In a classical natural product discovery approach, fungal isolates are grown in a variety of culture conditions. When there is sufficient growth, the metabolites are extracted, purified, and characterized when possible. This process is laborious and, in the aforementioned discoveries, further complicated by the typically slow growth rate of endophyte species.

Another aspect of the classical discovery approach is to screen crude extracts in one or more bioassays and identify the compounds responsible for any assay hits. However, these approaches often lead to the identification of the most abundant bioactive compounds, which are generally known compounds. More importantly this can miss minor novel compounds. A metabolomics-guided approach may also be

Received: April 23, 2023

Accepted: June 9, 2023

Published: June 27, 2023



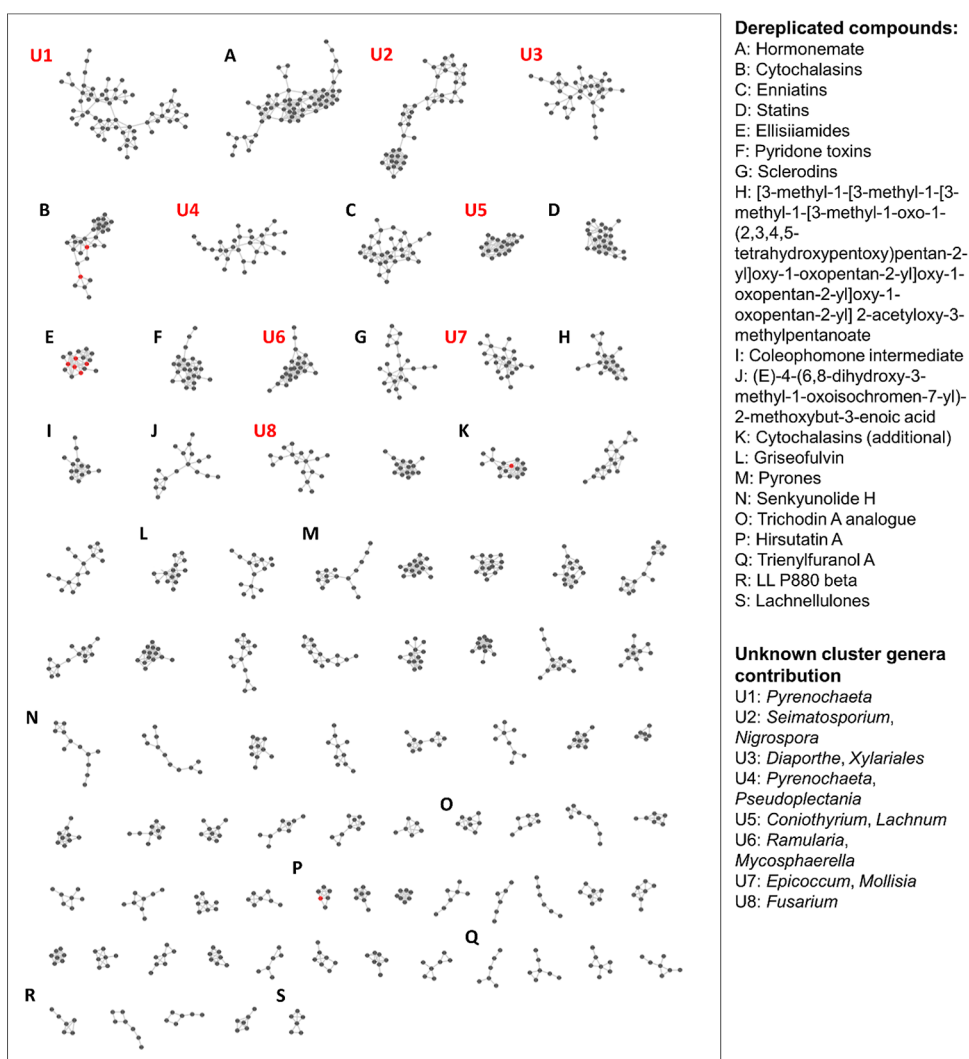


Figure 1. Molecular network of LC-HRMS/MS features from methanol extracts of Canadian fungal endophytes, generated with GNPS. Nodes indicate LC-MS/MS features, while lines indicate that the features share a cosine similarity score above the cutoff of 0.75. Red nodes indicate compounds from seed spectra, while gray nodes indicate compounds arising from endophyte extracts.

taken, where culture extracts are first screened by LC-HRMS and are grouped statistically using multivariate analysis, such as orthogonal projections to latent structures discriminant analysis (OPLS-DA) or principal component analysis (PCA), based on their secondary metabolite profiles.⁶ This approach allows strains with common metabolites to be grouped together, while unique metabolite profiles can be readily identified as divergent outliers. Furthermore, the datasets can be rapidly dereplicated or, in other words, have known compounds identified, saving time and costly efforts in purification and structural characterization.⁹

The complex biosynthetic pathways that create natural products often result in the existence of chemical classes, rather than unique individual compounds.¹⁰ Non-ribosomal peptides and polyketides are synthesized by large proteins with modular domains; each module executes a step in growing the chemical structure, therefore variation can occur if a module has relaxed specificity for its function.¹⁰ For example, the tyrocidines are a large class of cyclic decapeptides with antibiotic activity arising from three peptide synthetases. Tyrocidines exhibit structural variance due to the reduced specificity of these synthetases for aromatic amino acids at select locations along the peptide

chain.¹¹ This results in at least 28 tyrocidines with a common core sequence of amino acids that differ by the interchange of select aromatic amino acids.¹² A similar biosynthetic mechanism is demonstrated by the prochlorosins—a group of cyclic, lanthionine-containing antimicrobial peptides derived from marine cyanobacteria. The nearly thirty prochlorosins are formed by the activity of one enzyme with exceptionally low substrate specificity.¹³

Regardless of compound class, tandem mass spectrometry (MS/MS) is a critical tool for natural product discovery. The MS/MS fragmentation patterns of natural products are highly dependent on their molecular structure, meaning structurally similar compounds often follow common fragmentation pathways leading to diagnostic product ions and/or neutral loss products. For some compound classes, dereplication methods are advanced, such as for peptidic natural products via iSNAP—informatic search algorithm for natural products—which dereplicates non-ribosomal peptides against a database of in-silico tandem MS spectra.¹⁴ For other compound classes, the identification of structurally related unknowns can be a challenge, especially for complex classes such as polyketides. In recent years molecular networking has become a popular tool

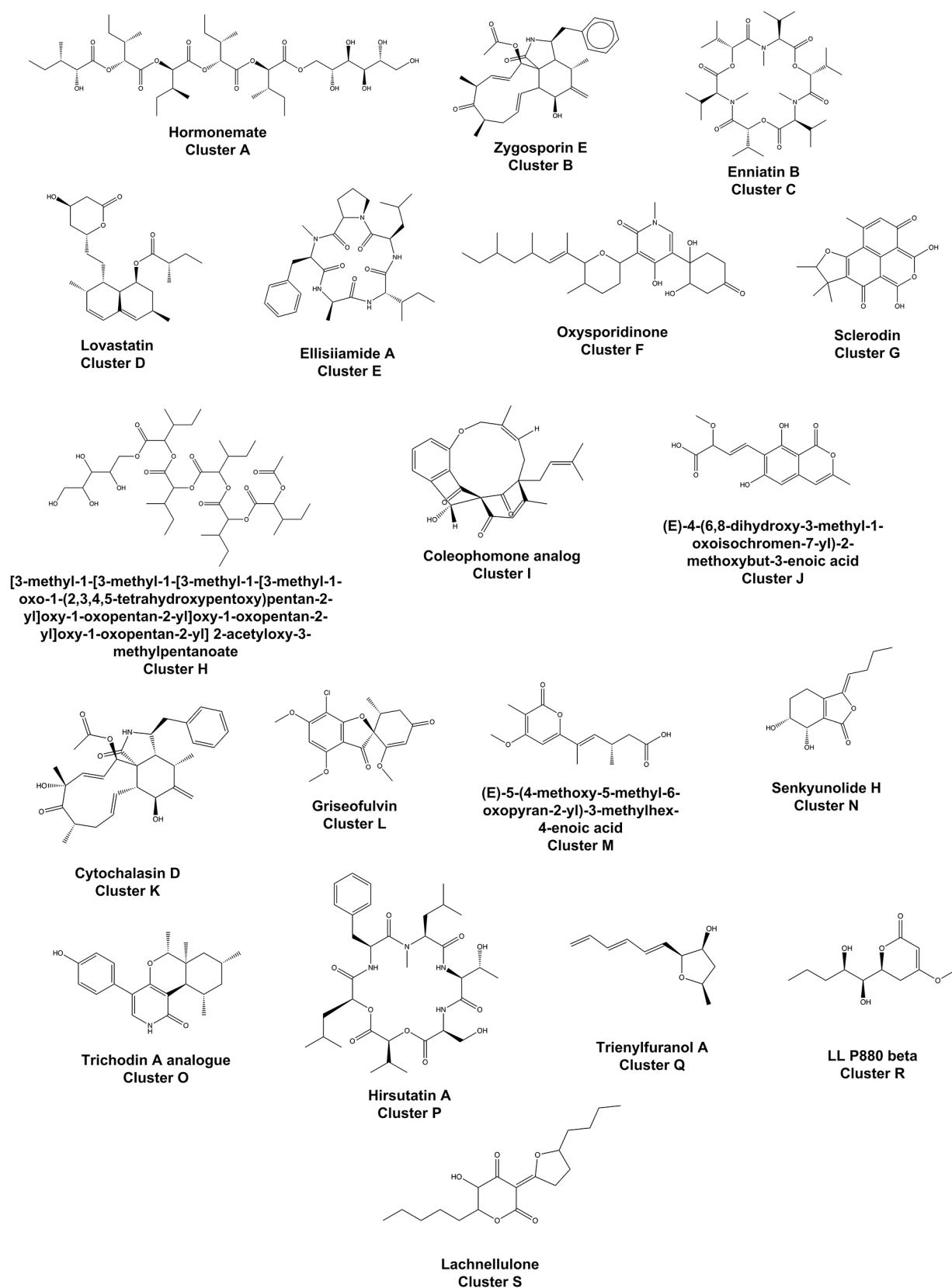


Figure 2. Representative structures for dereplicated compounds in the molecular network shown in Figure 1.

for addressing this gap in rapid compound identification.¹⁵ It takes advantage of the fragmentation similarities among compounds within a class to map chemicals visually according

to their degree of relatedness. Each compound is represented by a node, connected to one another by lines called edges, if they share similar fragmentation sequences or transitions

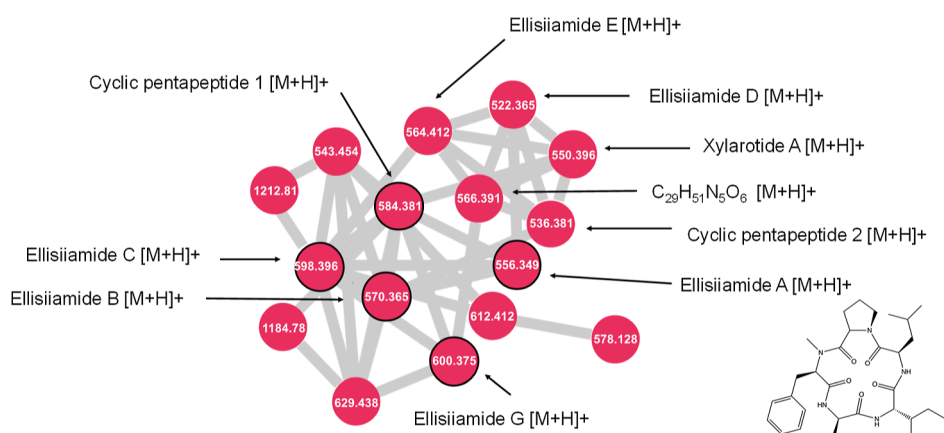


Figure 3. Close-up of cluster E from Figure 1. Ellisiamides and related compounds are labeled. Nodes outlined in black were included as seed spectra. All nodes were detected in extracts of *X. ellisii*.

within their tandem MS spectra. The Global Natural Products Social (GNPS) Molecular Network takes this a step further with the ability to also compare compounds to a large publicly shared database of known fragmentation spectra and to other user-generated spectra for community identifications.

Unlike the OPLS-DA and PCA statistical approaches previously employed on our collection of endophyte extracts, which grouped strains based on metabolite correlations and abundance, GNPS groups the metabolites themselves according to their MS/MS spectral similarity. This additional information about structural relatedness can guide us toward unknown compounds that are related to known bioactive compounds or new chemotypes, giving us a more streamlined approach to prioritize the discovery of novel natural products.

In this work, we examine a collection of 288 fungal endophyte extracts using non-targeted analysis by LC–HRMS/MS and the GNPS molecular networking platform to explore the chemical diversity of fungi endogenous to agriculturally important fruiting plants to target novel natural products for further isolation and characterization. This approach has led to a more comprehensive understanding of endophytic fungal diversity seen across these plants and of the varied chemical profiles present among fungal species. It has also allowed for a focused effort to identify unique fungal isolates that produce novel compounds, while also identifying natural product targets with potential utility that warrant purification, characterization, and bioactivity testing.

RESULTS AND DISCUSSION

Based on ITS sequence similarity, the 288 endophyte strains represented 87 unique species across 71 genera. There are caveats with using only ITS to identify taxa, for example, it provides insufficient resolution for delineating species in common genera, such as *Aspergillus*, *Cladosporium*, and *Fusarium*, and species complexes comprising distinct species with identical or very similar ITS sequences.¹⁶ Furthermore, alpha diversity may be overestimated due to high intragenomic rDNA variation or conversely underestimated due to low or absent interspecific ITS variation.¹⁷ The species identifications provided here are considered tentative or first diagnoses pending sequencing with additional secondary barcodes.

Endophyte extracts were analyzed by LC–MS/MS and data were subjected to molecular networking and dereplication using GNPS. Samples were dereplicated by comparison with the suite of libraries found within the GNPS search function,

which include nearly 600,000 entries ranging from plant and microbial natural products, to human metabolites, to known drugs, and other reference standards. MS/MS spectra of the unknown samples were compared to the databases and were considered a match if their cosine score is above the user-assigned cut-off value of 0.75 for this study. Clusters that contain at least one known compound are labeled alphabetically in Figure 1. Unknown compounds within the dereplicated clusters are putatively structurally similar to the known compounds; their proposed chemical formulas are listed in Table S3. A subset of dereplicated clusters is examined in detail below, and the full list of dereplicated compounds is detailed in Table S2.

PCA and GNPS Molecular Network. The performance of the LC–HRMS/MS system was evaluated by injecting quality control (QC) samples to assess instrument drift over the course of analysis. QC samples were made from pooled sample aliquots and were injected once every 10 test sample injections. Test and QC samples were subjected to PCA. While test samples are spread across the entire plot over a range of PC scores—as we would expect from a large group of diverse extracts—the QC samples group closely together due to their similar PC scores, indicating little variation among samples and, therefore, minimal instrumental drift (Figure S1).¹⁸ After assessing the reproducibility of the LC–HRMS/MS analysis, molecular networking with GNPS was performed. Seed spectra of known compounds were included in the network and belonged to compounds previously isolated from our fungal endophyte collection, including *Coniochaeta tetraspora*, *Nigrospora sphaerica*, *Sphaerulina vaccinii*, *Xylaria castorea*, and *X. ellisii*.^{6,19} All but *X. ellisii* were previously identified tentatively based on ITS data. Processing the LC–MS data by XCMS yielded ~8000 molecular features in the PCA plot, therefore in the resulting molecular network, the cosine score cutoff was set at 0.75—slightly higher than the default setting of 0.6—to reduce the number of connections that warrant investigation. Following the removal of single-node clusters and features detected in the blank media controls, the molecular network contained 2804 nodes with 4119 connections (most of the resulting network is shown in Figure 1, with smaller clusters after S omitted to provide higher image resolution).

Sixty known compounds were dereplicated across 19 clusters in the network by comparison to the databases available through GNPS and to an in-house library of MS/MS spectra. These compounds are outlined in Figure 1. Dereplicated

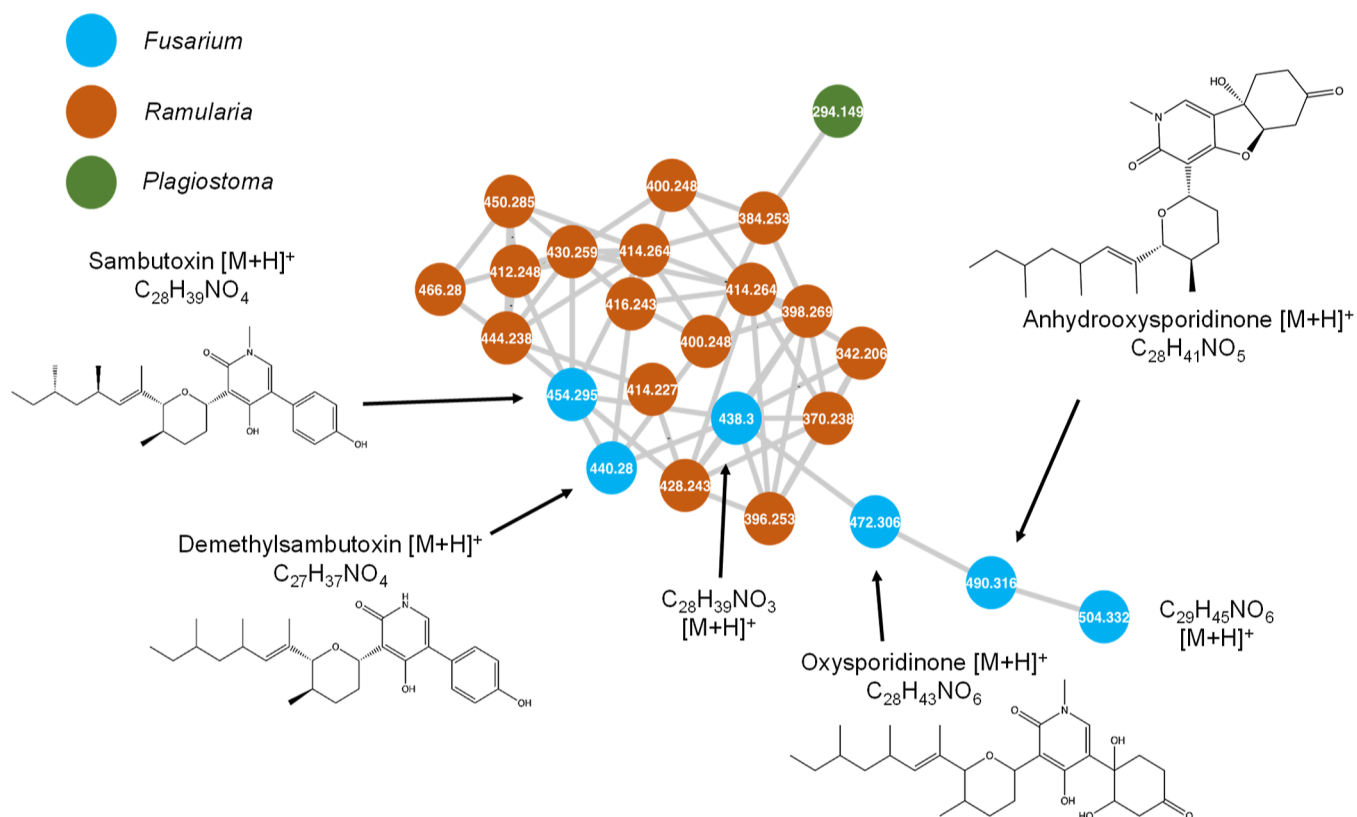


Figure 4. Close-up of cluster F from Figure 1. Oxysporidinone and related compounds are labeled. Blue nodes were produced exclusively by *Fusarium*.

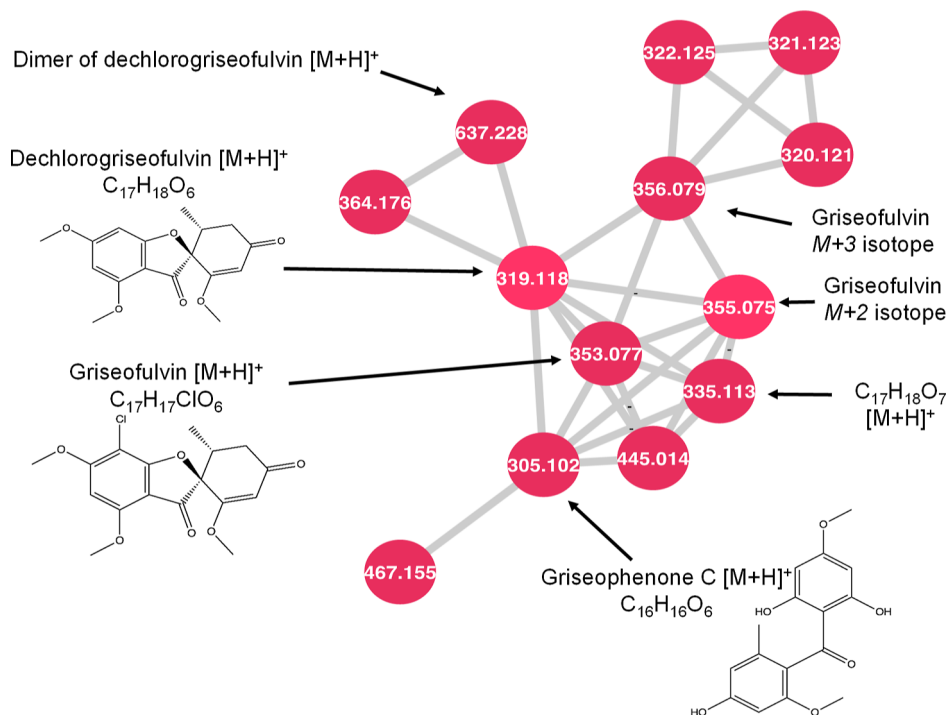


Figure 5. Close-up of cluster L from Figure 1. Griseofulvin and related compounds are labeled. All nodes were detected in multiple isolates of *X. ellisii*.

compounds listed in Supporting Information Table S2 cover a broad range of structures (Figure 2). We designated clusters containing one or more dereplicated compounds as “dereplicated clusters” (Figure 1; Table S3). Clusters that

contain no dereplicated compounds represent the highest probability of being novel and are, therefore, ideal targets for future purification and characterization efforts. We classified these clusters as “unknown clusters” (Figure 1; Table S3). Of

note, within the network, the positioning of clusters in relation to one another does not convey information, they are simply organized from largest to smallest (Figures 3–5).

Cluster E—Ellisiamides. Ellisiamides A, B, C, and G, along with cyclic pentapeptide 1, were included as seed spectra in the network and, as expected, were identified among extracts of *X. ellisii*, the organism from which they were first identified.⁶ These grouped together in cluster E, along with ellisiamide D, cyclic pentapeptide 2, and an unidentified congener with an m/z value of 566.3914 $[M + H]^+$ and chemical formula of $C_{29}H_{51}N_5O_6$, tentatively identified as a new cyclic pentapeptide. It has a mass difference of 15.995 amu from xylarotide A, representing the addition of an oxygen atom and does not share a formula with any previously published members of this class. It also has a distinct retention time from xylarotide A, indicating that it is not an in-source fragment of xylarotide A. It is produced by two isolates of *X. ellisii* from the collection, namely E-206 and E-244. This new cyclic pentapeptide also shares high cosine similarity scores ranging from 0.86 to 0.93 with four dereplicated cyclic pentapeptides, lending further confidence to the identification.

Ellisiamides A–C have previously been tested for bioactivity against *Candida albicans*, *Escherichia coli*, and *Saccharomyces cerevisiae*. Ellisiamide A has modest activity against *E. coli* with an MIC of 100 $\mu\text{g}/\text{mL}$, while ellisiamides B and C did not show any activity against the organisms tested. The remaining ellisiamides have yet to be tested in bioassays, along with cyclic pentapeptide 1. Xylarotide A has been tested against *Bacillus pumilus*, *C. albicans*, *E. coli*, and *Staphylococcus aureus* and does not exhibit activity against any of them. However, other closely related cyclic pentapeptides from *Xylaria* spp. do exhibit bioactivity, such as cyclo(*N*-methyl-*L*-Phe-*L*-Val-*D*-Ile-*L*-Leu-*L*-Pro) from an endolichenic species of *Xylaria*, which shows synergistic antifungal activity with ketoconazole against *C. albicans*.²⁰ It differs from ellisiamide A only by a substitution of the alanine residue for a valine, and was only tested against *C. albicans*, indicating that more ellisiamides may have additional bioactivities if tested against a broader range of organisms.

Cluster F—Oxysporidinone. Oxysporidinone is an antifungal compound originally identified from the fungus *Fusarium oxysporum* and, here, was detected in extracts of *Fusarium* cf. *tricinctum*.²¹ It is cytotoxic to several plant pathogenic fungi, including *Alternaria alternata*, *Aspergillus niger*, *Botrytis cinerea*, and *Venturia inaequalis*.²¹ Oxysporidinone is present within cluster F as a protonated ion (m/z 490.3162, $[M + H]^+$) and was dereplicated with the GNPS library search. Several compounds within the cluster have chemical formulas similar to that of oxysporidinone, including formulas matching those of several known structural relatives of oxysporidinone. Sambutoxin was observed in *F. cf. tricinctum* extracts with an m/z value of 454.2952 $[M + H]^+$. It is a mycotoxin produced by *Fusarium* spp. and has demonstrated toxicity in rats. The derivatives demethylsambutoxin and anhydrooxysporidinone were also present with m/z values of 440.2793 $[M + H]^+$ and 472.3059 $[M + H]^+$, respectively. Both are known products of *Fusarium* spp., but neither have recorded bioactivity with the small range of organisms they have been tested against. All dereplicated compounds were produced by all three isolates of *F. cf. tricinctum* in the collection.

Cluster F also contains two nodes representing chemical formulas of $C_{28}H_{39}NO_3$ and $C_{29}H_{45}NO_6$, given by m/z values of 438.3001 $[M + H]^+$ and 504.3321 $[M + H]^+$, respectively.

The compound with the molecular formula $C_{28}H_{39}NO_3$ resembles sambutoxin, with one less oxygen atom, and connected to sambutoxin by a high cosine similarity score of 0.98, indicating nearly identical fragmentation patterns. The formula of compound $C_{29}H_{45}NO$ is one CH_2 group larger than oxysporidinone, and the two are connected by a cosine similarity score of 0.85. These unidentified formulas did not match any published structural relatives within this compound class and are likely oxysporidinone relatives due to their high spectral similarity to known pyridine alkaloids. Both unknown compounds were produced by only two isolates of *F. cf. tricinctum*, namely E-178 and E-259. Other known structural relatives produced by *Fusarium* spp., such as the antibacterial fusapyridones, were not present among crude extracts of *F. cf. tricinctum*,²² but the diversity of bioactivity already documented in this class make these unknown compounds targets to pursue further. Although there were several additional nodes in the cluster contributed by *Ramularia* and *Plagiostoma* spp., none were dereplicated with the methods used.

Cluster L—Griseofulvin. Within cluster L, griseofulvin and dechlorogriseofulvin were dereplicated by the GNPS library search function. The $M + 2$ node is present for griseofulvin and represents its ³⁷Cl isotope, which lends further confidence to the identification. Two other nodes connected with griseofulvin have m/z values of 305.1019 $[M + H]^+$ and 335.1125 $[M + H]^+$, which correspond to formulas of $C_{16}H_{16}O_6$ and $C_{17}H_{18}O_7$, respectively. Griseophenone C is tentatively identified from the molecular formula $C_{16}H_{16}O_6$, which shares a prominent product ion of m/z 165.0546 with griseofulvin and griseophenone B. The compound $C_{17}H_{18}O_7$ shares this same product ion, but the formula does not match any previously published griseofulvin-related compounds. $C_{17}H_{18}O_7$ is also connected to griseofulvin and dechlorogriseofulvin by high cosine scores of 0.94 and 0.93, respectively. Griseofulvin and dechlorogriseofulvin were present in most extracts of *X. ellisii* in the collection. Griseophenone C was present in five isolates of *X. ellisii*.

Griseofulvin is a known antifungal drug that is a common ingredient in topical antifungal creams and is used to treat a broad variety of human fungal infections. It also shows promise as a treatment for other conditions like gout and ischemic heart disease.²³ Dechlorogriseofulvin also exhibits some antifungal activity, but to a lesser degree than its chlorinated analogue.^{24,25} Both compounds are known metabolites of *X. ellisii*.¹⁹ Griseophenone C has not been previously identified from *X. ellisii*, but it is a known precursor to griseofulvin within the biosynthetic pathway, so it is not unexpected to detect it along with other griseofulvin compounds. It has antibacterial activity against several bacterial species, including methicillin-resistant *S. aureus* and *E. coli*. No other griseophenones were identified in the network, although two unpaired nodes with m/z values matching those of griseophenone B (m/z 339.0629, $C_{16}H_{15}ClO_6$) and griseophenone D (m/z 291.0863, $C_{15}H_{14}O_6$) were present. The tandem mass spectrum collected at the m/z of griseophenone B also possessed the prominent product ion at m/z 165.0546 observed in the other griseofulvin-related compounds, and a ³⁷Cl isotope pattern, suggesting a chlorine-containing griseofulvin relative. However, the tandem mass spectra for both features had fewer than ten fragment ions under these conditions and as such they did not form any connections with other nodes. Therefore, they were removed, and could not be conclusively dereplicated from the overall network.

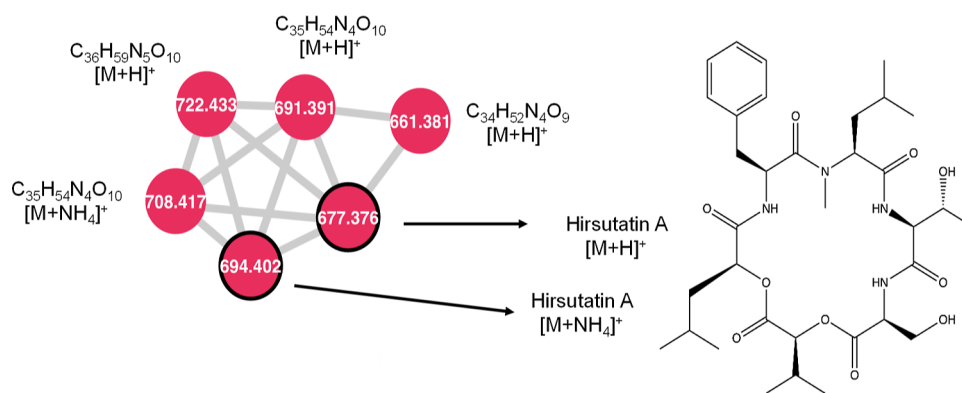


Figure 6. Close-up of cluster P from Figure 1. Hirsutatin A and related compounds are labeled above. All nodes were produced by isolates of *Xylaria*. Nodes outlined in black were included as seed spectra.

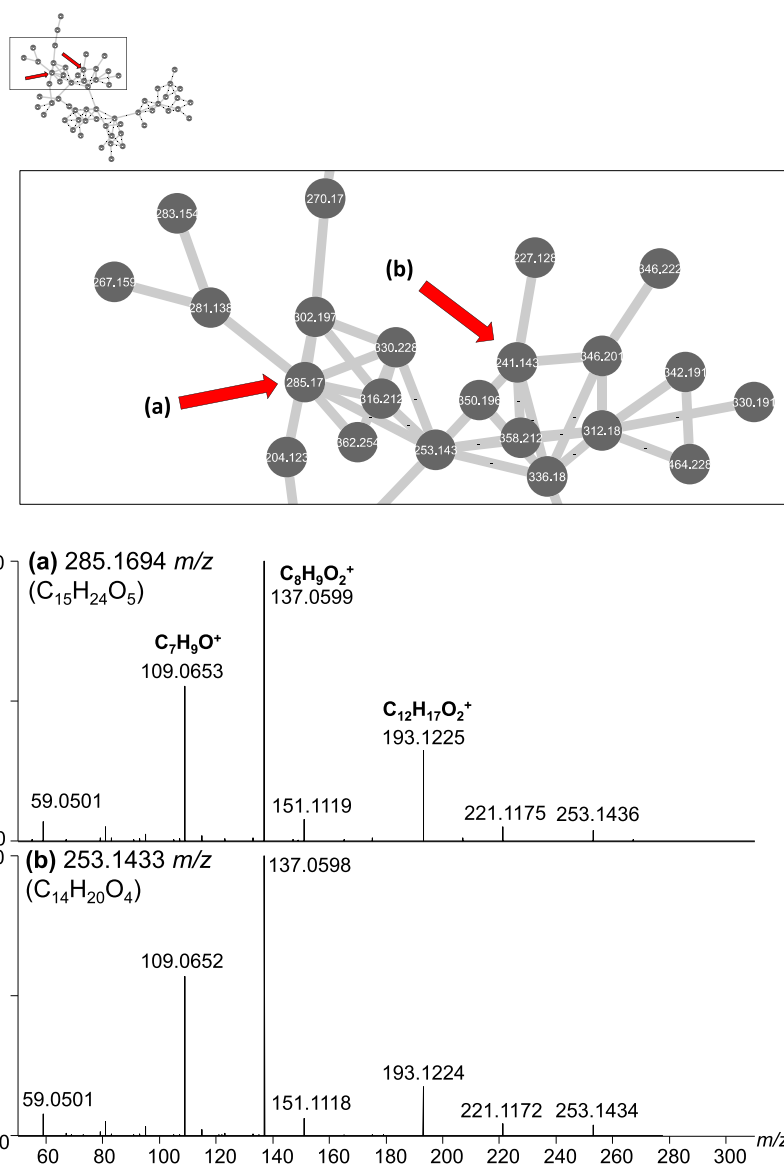


Figure 7. Close-up of cluster U1 from Figure 1. U1 is the largest cluster. Common molecular features include product ions at m/z $C_7H_9O^+$, $C_8H_9O_2^+$, and $C_{12}H_{17}O_2^+$. The major producer of these compounds are *Neocucurbitaria* spp.

Cluster P—Hirsutatin A. Hirsutatin A was dereplicated in extracts by comparing with the included seed spectrum of hirsutatin A. It was additionally present in one extract each of

Xylaria cubensis and of *Godronia cassandrae*. Hirsutatin A is a cyclohexadepsipeptide originally identified from an insect pathogenic fungus and was later also observed as a natural

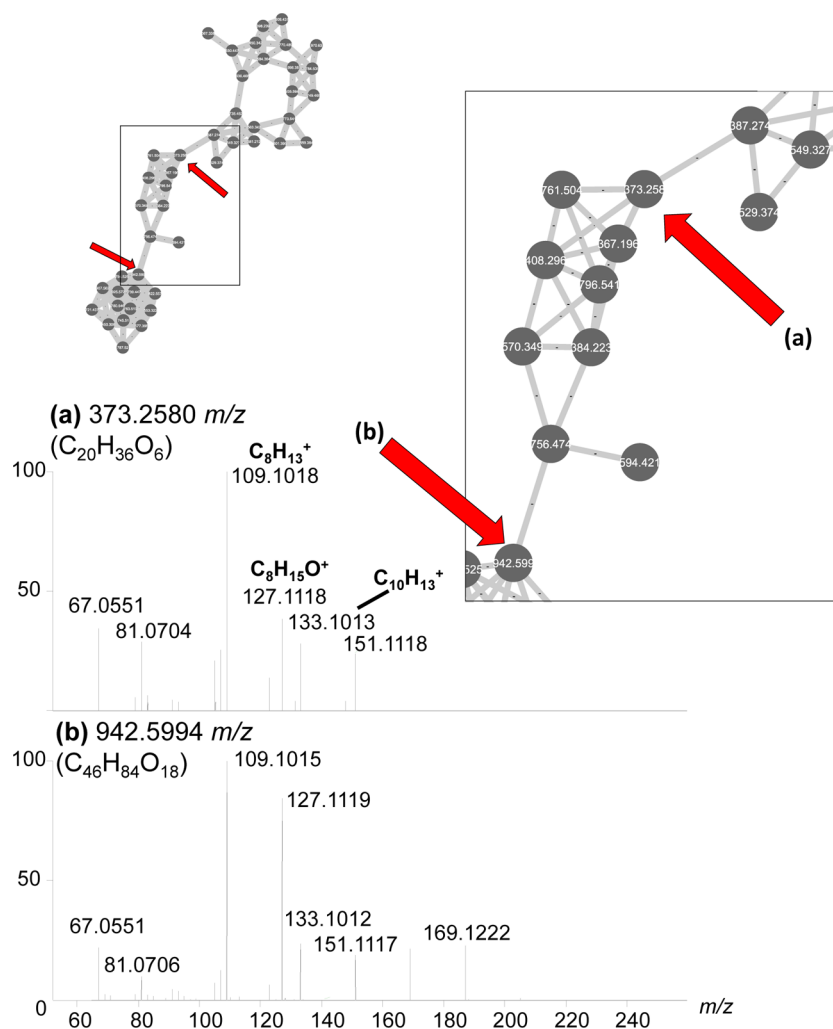


Figure 8. Close-up of cluster U2 from Figure 1. Molecular features within cluster U2 had common product ions at $C_8H_{13}^+$, $C_8H_{15}O^+$, and $C_{10}H_{15}O^+$. At this collision energy, all major product ions are below m/z 200. The major producers of these compounds are *Seimatosporium* and *Nigrospora* spp.

product of the endophytic fungus *X. ellisii*. Within cluster P, it is present in both protonated and ammoniated forms. Hirsutatin A was connected with three unknown compounds with m/z values of 661.3804 [$M + H$] $^+$, 691.3909 [$M + H$] $^+$, and 722.4332 [$M + H$] $^+$, corresponding to chemical formulas of $C_{34}H_{52}N_4O_9$, $C_{35}H_{54}N_4O_{10}$, and $C_{36}H_{59}N_5O_{10}$, respectively (Figure 6). These compounds are each connected to the hirsutatin A node with cosine similarity scores of 0.84, 0.88, and 0.95, respectively, indicating a high likelihood of structural similarity and did not match the formulas of any published depsipeptides.

Hirsutatin A has biological activity against *Mycobacterium tuberculosis*—the causative agent of tuberculosis in humans. Hirsutatin B was not present in the molecular network or the raw files, but it also has activity against *M. tuberculosis*, and strong activity against a multi-drug resistant strain of *Plasmodium falciparum*, which is the parasite responsible for malaria.²⁶ Other cyclic depsipeptides also show a range of bioactivities, such as the enniatins produced by *Fusarium*, which are antibacterial, antihelminthic, insecticidal, antifungal, and herbicidal. Because of the bioactivities demonstrated by this compound class, the unidentified compounds noted in this cluster are worthwhile to investigate further.

Unknown Clusters. As shown in Figure 1, many of the largest clusters did not contain any known compounds based on comparison to our seed spectra and the GNPS library. This does not necessarily indicate that they are a novel series of natural products, only that these spectra should be investigated in detail to determine if they represent tangible targets for purification and characterization. For example, by screening the strains represented in these clusters through bioactivity assays or by analyzing the existing extracts with negative mode ionization. In Figure 1, the eight largest unknown clusters (based on number of nodes) are listed numerically from U1–U8. Details about the detected analytes are listed in Table S3 and the three largest unknown clusters are discussed in more detail below.

Cluster U1. U1 is the largest cluster detected based on number of member nodes (Figure 1). A common spectral feature for many of the high intensity compounds of this cluster are the product ions of m/z 137.060 ($C_8H_9O_2^+$) and 109.0652 ($C_7H_9O^+$). The major genus that produces compounds within this node is *Neocurcurbitaria*.

Two sample nodes from Cluster U1 are shown above in Figure 7. Chemical formulas were calculated from HRMS m/z values and were determined to be $C_{15}H_{24}O_5$ and $C_{14}H_{20}O_4$, neither of which yielded matches when compared to the GNPS

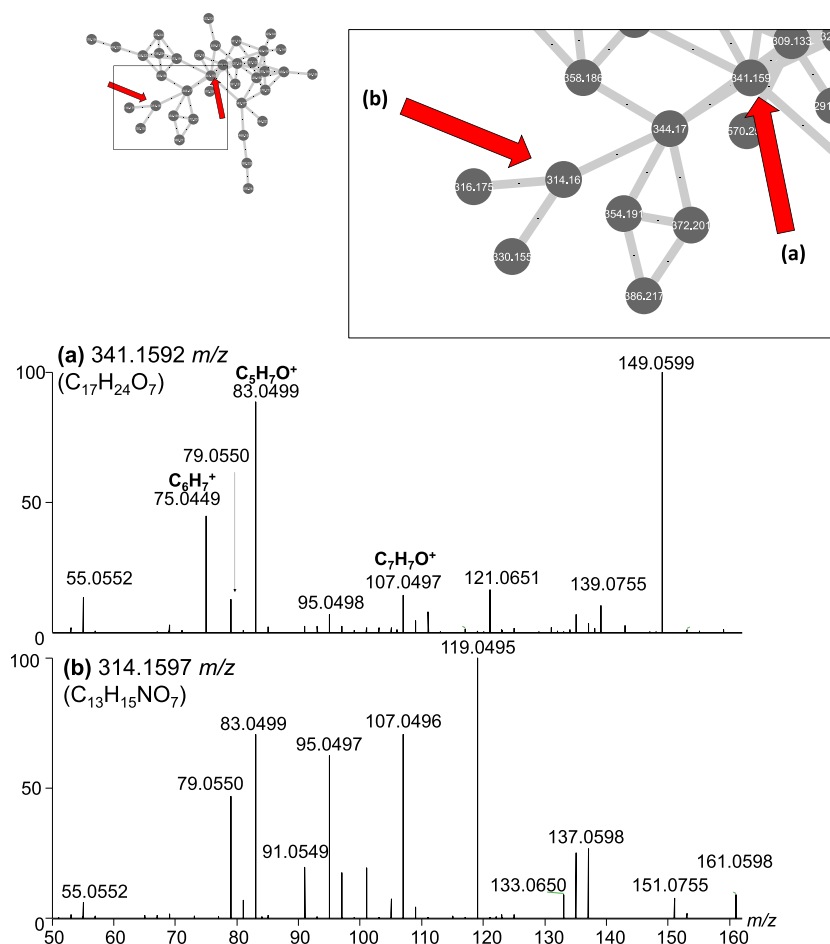


Figure 9. Close-up of cluster U3 from Figure 1. Molecular features within cluster U3 had common product ions at $C_6H_7^+$, $C_5H_7O^+$, and $C_7H_7O^+$. The major producers of these compounds are *Diaporthe* spp. and a species from the order *Xylariales*.

Table 1. Unknown Compounds Identified from Canadian Fungal Endophytes as Targets for Isolation and Characterization

formula	$[M + H]^+$ m/z	RT (min)	mass error (Δ ppm)	structurally related to	produced by
$C_{17}H_{18}O_7$	335.1125	3.41	-0.117	griseofulvin	<i>X. ellisii</i>
$C_{28}H_{39}NO_3$	438.3001	5.79	-0.298	oxysporidinone	<i>Fusarium</i> cf. <i>tricinctum</i>
$C_{29}H_{45}NO_6$	504.3321	5.07	0.189	oxysporidinone	<i>Fusarium</i> cf. <i>tricinctum</i>
$C_{29}H_{51}N_5O_6$	566.3914	4.08	0.369	ellisiimides	<i>X. ellisii</i>
$C_{34}H_{52}N_4O_9$	661.3804	4.60	-0.507	hirsutatin A	<i>X. ellisii</i>
$C_{35}H_{54}N_4O_{10}$	691.3909	4.48	-0.044	hirsutatin A	<i>X. ellisii</i>
$C_{36}H_{59}N_5O_{10}$	722.4332	4.37	-0.414	hirsutatin A	<i>X. ellisii</i>

library or seed spectra. Compounds within the cluster ranged in m/z from 170 to 527. There are no previously reported metabolites from *Neocucurbitaria* detectable within this cluster. Cluster U4 contains two nodes that match m/z values of neocucurbinol A and neocucurbin G; however, there was no way to confirm the identities. Overall, there are few bioactive compounds reported from any *Neocucurbitaria*, making them interesting to investigate for new and novel compounds in the future.

Cluster U2. Nodes in Cluster U2 are predominantly seen in the genera *Seimatosporium* and *Nigrospora*. Common product ions identified among several compounds in the cluster include m/z 109.1015 ($C_8H_{13}^+$), 127.1119 ($C_8H_{15}O^+$), and 133.1012 ($C_{10}H_{13}^+$) (Figure 8).

Compounds in this cluster have m/z values ranging from 367 to 970 and all have formulas that contain only carbon, hydrogen, and oxygen. Both $[M + H]^+$ and $[M + NH_4]^+$

adducts of compounds were present in many instances, as well as $[M + H - H_2O]^+$ ions. There were no other compounds present in the network previously reported from *Seimatosporium* or *Nigrospora* spp.; however, these are both understudied fungal species with few reported compounds to begin with, giving them a higher likelihood of producing new compounds.

Cluster U3. The major contributor to nodes within cluster U3 are species from the genus *Diaporthe*, as well as an undefined species from the order *Xylariales* (Figure 9).

Compounds in this cluster ranged from m/z 281 to 570 and are composed of carbon, hydrogen, and oxygen, while some also contain nitrogen. There were no other identifiable compounds previously reported from *Diaporthe* spp. in the network.

Seed Spectra. Of the previously isolated compounds that were included as seed spectra, several were not identifiable within the final network, namely coriloxin, abscisic acid,

Table 2. Seed Spectra Used in Molecular Network of Canadian Fungal Endophytes^a

name	calc <i>m/z</i> [M + H] ⁺	formula	fungal source
coriloxin	171.0652	C ₈ H ₁₀ O ₄	<i>X. castorea</i>
abscisic acid	265.1434	C ₁₅ H ₂₀ O ₄	<i>N. sphaerica</i>
ascochitine	277.1071	C ₁₅ H ₁₆ O ₅	<i>Coniochaeta cf. marina</i>
7-hydroxy-3-(hydroxymethyl)-2-(2-hydroxypropyl)-6-methoxy-4 <i>H</i> -chromen-4-one (fulvic acid derivative)	281.1019	C ₁₄ H ₁₆ O ₆	<i>S. vaccinii</i>
4,10-Dihydro-3,7,8-trihydroxy-3-methyl-10-oxo-1 <i>H</i> ,3 <i>H</i> -pyrano[4,3- <i>b</i>][1]benzopyran-9-carboxylic acid (fulvic acid analogue)	309.0605	C ₁₄ H ₁₂ O ₈	<i>S. vaccinii</i>
zygosporin E	492.2744	C ₃₀ H ₃₇ NO ₅	<i>X. ellisii</i>
cytochalasin D	508.2693	C ₃₀ H ₃₇ NO ₆	<i>X. ellisii</i>
epoxycytochalasin D	524.2642	C ₃₀ H ₃₇ NO ₇	<i>X. ellisii</i>
ellisiamide A	556.3493	C ₃₀ H ₄₅ N ₅ O ₅	<i>X. ellisii</i>
ellisiamide B	570.3650	C ₃₁ H ₄₇ N ₅ O ₅	<i>X. ellisii</i>
cyclic pentapeptide 1	584.3806	C ₃₂ H ₄₉ N ₅ O ₅	<i>X. ellisii</i>
ellisiamide C	598.3963	C ₃₃ H ₅₁ N ₅ O ₅	<i>X. ellisii</i>
ellisiamide G	600.3755	C ₃₂ H ₄₉ N ₅ O ₆	<i>X. ellisii</i>
hirsutatin A	677.3756	C ₃₄ H ₅₂ N ₄ O ₁₀	<i>X. ellisii</i>

^aSeed spectra included in the network are bioactive compounds previously isolated from fungal endophytes.

aschochitine, 4,10-dihydro-3,7,8-trihydroxy-3-methyl-10-oxo-1*H*,3*H*-pyrano[4,3-*b*][1]benzopyran-9-carboxylic acid, and 7-hydroxy-3-(hydroxymethyl)-2-(2-hydroxypropyl)-6-methoxy-4*H*-chromen-4-one. These were present in the mass spectra of fungi that are known to produce them, however upon closer inspection, either they had too few fragments to form connections to other nodes or there were no additional compounds similar enough to connect with.

Other seed spectra were present within the network but did not connect to any novel compounds. Cytochalasin D was dereplicated within *X. ellisii* extracts and was present within cluster K but was not connected with any new compounds. Epoxycytochalasin D and zygosporin E were both dereplicated within cluster B from extracts of *X. ellisii*, along with several other cytochalasins, however, no new compounds were present (Table 1).

CONCLUSIONS

Ultimately, the use of molecular networking as a dereplication and processing strategy with a large dataset of 288 fungal endophyte extracts was successful in identifying seven unknown compounds as targets for isolation that are strong candidates for bioactivity due to their relatedness to known potent bioactive compounds. These include relatives from diverse chemical classes, namely, one griseofulvin relative, three hirsutatin-related compounds, a new ellisiamide, and two new sambutoxin- and oxysporidinone-related compounds.

A challenge of molecular networking as a dataset mining strategy is that it is more successful when samples contain multiple compounds with similar chemical structures. If a dataset contains few structurally related compounds, this approach will overlook potentially valuable compounds that have no neighbors within the molecular network. Luckily, most biosynthetic pathways generate mixtures of structurally related compounds. While molecular networking may not be able to assign relationships to all compounds within a sample, it is becoming a more powerful dereplication tool, especially as databases improve, thus rapidly speeding up the analysis of large tandem mass spectra datasets.

Future works will focus on purifying the targeted new compounds for structural characterization and biological

activity assessments. The fungal endophytes will ideally be employed as part of an integrated crop management strategy in their host plants. The use of endophyte-enhanced plants in agriculture will help meet the Government of Canada's mandate to reduce our reliance on chemical pesticides.

MATERIALS AND METHODS

Endophyte Sampling and Extraction. Leaves and stems from Canadian fruit bearing crops, including blueberries, cranberries, raspberries, and grapes, were sampled in Ontario, New Brunswick, and Nova Scotia, Canada, between 2011 and 2015 (Supporting Information Table S1). Endophytes were isolated from these plants following the method provided by Ginn (1998).²⁷ In short, leaves and stems were surface-sterilized with bleach and ethanol, then were sliced into 1 cm pieces and placed on potato dextrose agar (PDA) media (Sigma-Aldrich, MO, USA). Fungal samples that grew from plant material were extracted with an UltraClean Microbial DNA Isolation Kit (MoBio, Carlsbad, CA). Polymerase chain reactions (PCR) were performed in an Eppendorf Mastercycler Nexus Gradient Thermal Cycler. The total volume of the reaction was 25 μ L, consisting of 22.5 μ L of Platinum Blue SuperMix (Invitrogen, Carlsbad, CA, USA), 1 μ L of genomic DNA (15 ng/ μ L), and 0.75 μ L each of forward and reverse primers (ITS1 and ITS4, respectively) for a final concentration of 0.3 μ M. PCR conditions include an initial denaturation at 94 $^{\circ}$ C for 2 min, then 35 cycles of 1 min at 94 $^{\circ}$ C, 30 s at 60 $^{\circ}$ C, and 30 s at 72 $^{\circ}$ C, then a final elongation of 5 min at 72 $^{\circ}$ C. DNA sequences were interpreted using NCBI BLAST blastn suite (<https://blast.ncbi.nlm.nih.gov/>). Isolated endophytes were grown on 20 mL of PDA at 23 $^{\circ}$ C for 2–6 weeks or until cells reached confluence. Cultures were extracted by homogenizing the agar and cells with 20 mL of methanol (Fisher Scientific, NJ, USA) and then gravity filtering the homogenate with a no. 1 Whatman filter. One-milliliter aliquots of filtrate were taken, then dried at room temperature under nitrogen. Dried samples were stored at -20° C until analysis.

Seed Spectra. Seed spectra included in analysis are presented in Table 2. These were previously isolated by HPLC and characterized by NMR as described in Ibrahim

(2017). Seed spectra compounds were analyzed by LC–MS/MS alongside endophyte extracts.

Analysis by LC–HRMS/MS. In preparation for analysis, samples and seed spectra were reconstituted in 1 mL of methanol (Fisher Scientific, Fair Lawn, NJ, USA) and analyzed by LC–HESI–HRMS/MS on a Thermo Q-Exactive Orbitrap mass spectrometer paired with an Agilent 1290 UHPLC system. Additionally, pooled QC samples were prepared by combining 10 μ L aliquots from each sample.

Chromatographic separation was accomplished using a dual-solvent system with acetonitrile + 0.1% formic acid (solvent A, Fisher Scientific, Fair Lawn, NJ, USA) and water + 0.1% formic acid (solvent B, Fisher Scientific, Fair Lawn, NJ, USA) at a rate of 0.3 mL/min. The gradient was held at 0% B for 0.5 min, increased to 100% B over 3 min, held at 100% B for 2.5 min, then decreased to 0% B over 0.5 min and held at 0% B for 1 min. All samples were injected in 5 μ L portions on an EclipsePlus RRHD C-18 column (2.1 \times 50 mm, 1.8 μ m; Agilent) that was maintained at 35 °C. Heated electrospray ionization (HESI) conditions were as follows: capillary temperature, 400 °C; sheath gas, 17 units; auxiliary gas, 8 units; probe heater temperature, 450 °C; S-Lens RF level, 50; and capillary voltage, 3.9 kV.

Data were acquired in positive ionization mode with data-dependent acquisition with the following settings: resolution, 70,000; automatic gain control (AGC) target, 1×10^6 ; max IT, 256 ms; scan range, 100–1500 m/z . The 10 ions with highest intensity from each MS scan were selected to be fragmented by MS/MS at resolution 17,500; AGC target 1×10^6 ; max IT, 64 ms; stepped NCE, 28/50; isolation window, 1.2 m/z ; intensity threshold, 1.3×10^5 ; dynamic exclusion, 10.0 s. QC samples were injected at the beginning, end, and throughout the run to assess instrument drift.

Data Processing and Principal Component Analysis. Thermo Raw files were converted to mzML format using MSConvert (v. 3) with the following settings: 32 bit binary encoding precision, no file compression, and peak picking from levels 1 to 2.²⁸ Converted files were brought into R (3.5.3) to perform principal component analysis (PCA) with the packages xcms (3.2.0), FactoMineR (2.3), and MetabolAnalyze (1.3.1).^{28–32} Settings for processing in R were as follows: method, centWave; prefilter, (5, 5000); ppm, 5; snthresh, 10; peakwidth, (5, 20); noise, 500,000; bw, 5; minfrac, 0.001; and mzwid, 0.015. Zero values were imputed with two-thirds of the lowest value measured for each metabolite. Peak area values were log-10 transformed and pareto scaled. PCA was performed with only QC samples to assess instrument drift over the course of analysis. The first and second principal components were plotted against one another.

GNPS Parameters and Processing. Converted mzML files were also uploaded to the GNPS molecular networking site with FileZilla (3.9.0.5) and analyzed with the following settings: precursor ion mass tolerance, 0.02 Da; fragment ion mass tolerance, 0.02 Da; min pairs cosine, 0.75; network topK, 10; maximum connected component size, 100; minimum matched fragment ions, 5; minimum cluster size, 2; and MScluster, on. Network output was downloaded as a GRAPHML file and was imported into Cytoscape (3.6.1) for visualization. To simplify network analysis, all unconnected nodes were removed, along with nodes attributed to blank media, clusters formed solely from seed spectra, and background ions.

Files were also assessed with the Library Search function of GNPS to dereplicate compounds. The following settings were used: precursor ion mass tolerance, 0.02 Da; fragment ion mass tolerance, 0.02 Da; min matched fragment ions, 5; and cut-off score, 0.75. Additional compounds were dereplicated by comparing to an in-house database of MS/MS spectra.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.3c02786>.

PCA plot performed to assess instrument performance, additional views of unknown clusters from the molecular network, all endophyte isolates examined in this study, and identities of all dereplicated compounds present in the network (PDF)

Precursor mass, retention times, calculated formulas, unique file sources, and putative identifications of all features from their corresponding clusters (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Mark W. Sumarah – London Research and Development Centre, Agriculture and Agri-Food Canada, London, Ontario NSV 4T3, Canada; Department of Chemistry, University of Western Ontario, London, Ontario N6A 3K7, Canada; orcid.org/0000-0003-0977-1656; Phone: +01-226-688-6187; Email: mark.sumarah@agr.gc.ca

Authors

Natasha DesRochers – London Research and Development Centre, Agriculture and Agri-Food Canada, London, Ontario NSV 4T3, Canada; Department of Chemistry, University of Western Ontario, London, Ontario N6A 3K7, Canada

Justin B. Renaud – London Research and Development Centre, Agriculture and Agri-Food Canada, London, Ontario NSV 4T3, Canada

Joey B. Tanney – Pacific Forestry Centre, Canadian Forest Service, Natural Resources Canada, Victoria, British Columbia V8Z 1M5, Canada

Ashraf Ibrahim – McMaster University, Hamilton, Ontario L8S 4L8, Canada

Ken K.-C. Yeung – Department of Chemistry and Department of Biochemistry, University of Western Ontario, London, Ontario N6A 3K7, Canada; orcid.org/0000-0001-6933-2941

Complete contact information is available at: <https://pubs.acs.org/10.1021/acsomega.3c02786>

Author Contributions

Conceptualization, M.W.S.; Methodology, N.D. and J.B.R.; Formal Analysis, N.D. and J.B.R.; Data Curation, N.D., J.B.R., J.T., and A.I.; Writing-Original Draft Preparation, N.D.; Writing-Review & Editing, all authors; Supervision, M.W.S. and K.K.-C.Y.; Project Administration, M.W.S.; Funding Acquisition, M.W.S.

Funding

This work was supported by an AAFC grant to MWS. Natasha DesRochers was partially supported by Western University.

Notes

The authors declare no competing financial interest.

Data available at the Metabolomics Repository with Study ID number ST002482.

ACKNOWLEDGMENTS

We would like to thank Megan Kelman for technical support.

REFERENCES

- (1) Kusari, S.; Hertweck, C.; Spiteller, M. Chemical Ecology of Endophytic Fungi: Origins of Secondary Metabolites. *Chem. Biol.* **2012**, *19*, 792–798.
- (2) Traversa, D.; Milillo, P.; Di Cesare, A.; Lohr, B.; Iorio, R.; Pampurini, F.; Schaper, R.; Bartolini, R.; Heine, J. Efficacy and Safety of Emodepside 2.1 %/Praziquantel 8.6% Spot-on Formulation in the Treatment of Feline Aelurostrongylosis. *Parasitol. Res.* **2009**, *105*, 83–90.
- (3) Sasaki, T.; Takagi, M.; Yaguchi, T.; Miyadoh, S.; Okada, T.; Koyama, M. A new anthelmintic cyclodepsipeptide, PF1022A. *J. Antibiot.* **1992**, *45*, 692–697.
- (4) Miller, J. D.; Adams, G. W. Endophyte enhanced seedlings with increased pest tolerance. U.S. Patent 8,455,395 B2, 2017.
- (5) Burgess, K. M. N.; Ibrahim, A.; Sørensen, D.; Sumarah, M. W. Trienylfuranol A and trienylfuranone A–B: metabolites isolated from an endophytic fungus, Hypoxylon submoniticulosum, in the raspberry *Rubus idaeus*. *J. Antibiot.* **2017**, *70*, 721–725.
- (6) Ibrahim, A.; Tanney, J. B.; Fei, F.; Seifert, K. A.; Cutler, G. C.; Capretta, A.; Miller, J. D.; Sumarah, M. W. Metabolomic-guided discovery of cyclic nonribosomal peptides from *Xylaria ellisii* sp. nov., a leaf and stem endophyte of *Vaccinium angustifolium*. *Sci. Rep.* **2020**, *10*, 4599–4617.
- (7) Richardson, S. N.; Walker, A. K.; Nsiama, T. K.; McFarlane, J.; Sumarah, M. W.; Ibrahim, A.; Miller, J. D. Griseofulvin-producing *Xylaria* endophytes of *Pinus strobus* and *Vaccinium angustifolium*: evidence for a conifer-understorey species endophyte ecology. *Fungal Ecol.* **2014**, *11*, 107–113.
- (8) Ibrahim, A.; Sørensen, D.; Jenkins, H. A.; Ejim, L.; Capretta, A.; Sumarah, M. W. Epoxynemanione A, nemanifuranones A–F, and nemanilactones A–C, from *Nemania serpens*, an endophytic fungus isolated from Riesling grapevines. *Phytochemistry* **2017**, *140*, 16–26.
- (9) Nielsen, K. F.; Larsen, T. O. The importance of mass spectrometric dereplication in fungal secondary metabolite analysis. *Front. Microbiol.* **2015**, *6*, 71.
- (10) Cane, D. E.; Walsh, C. T.; Khosla, C. Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* **1998**, *282*, 63–68.
- (11) Mootz, H. D.; Marahiel, M. A. The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol.* **1997**, *179*, 6843–6850.
- (12) Tang, X.-J.; Thibault, P.; Boyd, R. K. Characterisation of the tyrocidine and gramicidin fractions of the tyrothricin complex from *Bacillus brevis* using liquid chromatography and mass spectrometry. *Int. J. Mass Spectrom. Ion Processes* **1992**, *122*, 153–179.
- (13) Li, B.; Sher, D.; Kelly, L.; Shi, Y.; Huang, K.; Knerr, P. J.; Joewono, I.; Rusch, D.; Chisholm, S. W.; Van Der Donk, W. A. Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. *Proc. Natl. Acad. Sci.* **2010**, *107*, 10430–10435.
- (14) Ibrahim, A.; Yang, L.; Johnston, C.; Liu, X.; Ma, B.; Magarvey, N. A. Dereplicating nonribosomal peptides using an informatic search algorithm for natural products (iSNAP) discovery. *Proc. Natl. Acad. Sci.* **2012**, *109*, 19196–19201.
- (15) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kaponov, C. A.; Luzzatto-Knaan, T.; et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837.
- (16) Lücking, R.; Aime, M. C.; Robbertse, B.; Miller, A. N.; Ariyawansa, H. A.; Aoki, T.; Cardinali, G.; Crous, P. W.; Druzhinina, I. S.; Geiser, D. M.; et al. Unambiguous identification of fungi: where do we stand and how accurate and precise is fungal DNA barcoding? *IMA Fungus* **2020**, *11*, 14.
- (17) Paloi, S.; Luangsa-ard, J. J.; Mhuantong, W.; Stadler, M.; Kobmoo, N. Intragenomic variation in nuclear ribosomal markers and its implication in species delimitation, identification and barcoding in fungi. *Fungal Biol. Rev.* **2022**, *42*, 1–33.
- (18) Dunn, W. B.; Wilson, I. D.; Nicholls, A. W.; Broadhurst, D. The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. *Bioanalysis* **2012**, *4*, 2249–2264.
- (19) Ibrahim, A. *Microbial Secondary Metabolomics for Natural Product Discovery*; McMaster University: Hamilton, Ontario, Canada, 2017.
- (20) Wu, W.; Dai, H.; Bao, L.; Ren, B.; Lu, J.; Luo, Y.; Guo, L.; Zhang, L.; Liu, H. Isolation and structural elucidation of proline-containing cyclopentapeptides from an endolichenic *Xylaria* sp. *J. Nat. Prod.* **2011**, *74*, 1303–1308.
- (21) Breinholt, J.; Ludvigsen, S.; Rassing, B. R.; Rosendahl, C. N.; Nielsen, S. E.; Olsen, C. E. Oxysporidinone: A Novel, Antifungal N-Methyl-4-hydroxy-2-pyridone from *Fusarium oxysporum*. *J. Nat. Prod.* **1997**, *60*, 33–35.
- (22) Tsuchinari, M.; Shimanuki, K.; Hiramatsu, F.; Murayama, T.; Koseki, T.; Shiono, Y. Fusapyridones A and B, Novel Pyridone Alkaloids from an Endophytic Fungus, *Fusarium* sp. YG-45. *Z. Naturforsch., B* **2007**, *62*, 1203–1207.
- (23) Finkelstein, E.; Amichai, B.; Grunwald, M. H. Griseofulvin and its uses. *Int. J. Antimicrob. Agents* **1996**, *6*, 189–194.
- (24) Park, J.-H.; Choi, G. J.; Lee, H. B.; Kim, K. M.; Jung, H. S.; Lee, S. W.; Jang, K. S.; Cho, K. Y.; Kim, J. C. Griseofulvin from *Xylaria* sp. strain F0010, an endophytic fungus of *Abies holophylla* and its antifungal activity against plant pathogenic fungi. *J. Microbiol. Biotechnol.* **2005**, *15*, 112–117.
- (25) Brian, P. W.; Curtis, P. J.; Hemming, H. G. A substance causing abnormal development of fungal hyphae produced by *Penicillium janczewskii* Zal.: I. Biological assay, production and isolation of ‘curling factor’. *Trans. Br. Mycol. Soc.* **1946**, *29*, 173–187.
- (26) Isaka, M.; Palasarn, S.; Sriklung, K.; Kocharin, K. Cyclohexadepsipeptides from the Insect Pathogenic Fungus *Hirsutella nivea* BCC 2594. *Journal of Natural Products* **2005**, *68*, 1680–1682.
- (27) Ginn, F. M. *Endophytic fungi in Vaccinium macrocarpon (cranberry) and Vaccinium angustifolium (blueberry)*. University of New Brunswick, 1998.
- (28) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology* **2012**, *30*, 918.
- (29) Smith, C. A.; Want, E. J.; O’Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry* **2006**, *78*, 779–787.
- (30) Lê, S.; Josse, J.; Husson, F. FactoMineR: an R package for multivariate analysis. *Journal of Statistical Software* **2008**, *25*, 1–18.
- (31) Nyamundanda, G.; Brennan, L.; Gormley, I. C. Probabilistic principal component analysis for metabolomic data. *BMC Bioinf.* **2010**, *11*, 571.
- (32) Scrucca, L.; Fop, M.; Murphy, T. B.; Raftery, A. E. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *R J.* **2016**, *8*, 289.