nature microbiology

ARTICLES

PUBLISHED: 31 OCTOBER 2016 | VOLUME: 2 | ARTICLE NUMBER: 16197

# Indexing the *Pseudomonas* specialized metabolome enabled the discovery of poaeamide B and the bananamides

Don D. Nguyen[1†], Alexey V. Melnik[2†], Nobuhiro Koyama[2,3], Xiaowen Lu[4], Michelle Schorn[5], Jinshu Fang[1], Kristen Aguinaldo[6], Tommie L. Lincecum Jr[6], Maarten G. K. Ghequire[7], Victor J. Carrion[8], Tina L. Cheng[9], Brendan M. Duggan[10], Jacob G. Malone[11,12], Tim H. Mauchline[13], Laura M. Sanchez[10], A. Marm Kilpatrick[9], Jos M. Raaijmakers[8], René De Mot[7], Bradley S. Moore[5,10], Marnix H. Medema[4*] and Pieter C. Dorrestein[2,5,10*]

**Pseudomonads are cosmopolitan microorganisms able to produce a wide array of specialized metabolites. These molecules allow *Pseudomonas* to scavenge nutrients, sense population density and enhance or inhibit growth of competing microorganisms. However, these valuable metabolites are typically characterized one-molecule–one-microbe at a time, instead of being inventoried in large numbers. To index and map the diversity of molecules detected from these organisms, 260 strains of ecologically diverse origins were subjected to mass-spectrometry-based molecular networking. Molecular networking not only enables dereplication of molecules, but also sheds light on their structural relationships. Moreover, it accelerates the discovery of new molecules. Here, by indexing the *Pseudomonas* specialized metabolome, we report the molecular-networking-based discovery of four molecules and their evolutionary relationships: a poaeamide analogue and a molecular subfamily of cyclic lipopeptides, bananamides 1, 2 and 3. Analysis of their biosynthetic gene cluster shows that it constitutes a distinct evolutionary branch of the *Pseudomonas* cyclic lipopeptides. Through analysis of an additional 370 extracts of wheat-associated *Pseudomonas*, we demonstrate how the detailed knowledge from our reference index can be efficiently propagated to annotate complex metabolomic data from other studies, akin to the way in which newly generated genomic information can be compared to data from public databases.**

The production of specialized metabolites by *Pseudomonas* species leads to strain-specific biological activities[1,2–5]. For instance, siderophores and cyclic lipopeptides produced by strains of *Pseudomonas putida* and *Pseudomonas fluorescens* act as bioactive agents against susceptible plant and animal pathogens, thereby conveying protection and promoting plant growth, while toxins and glycolipids produced by *Pseudomonas syringae* and *Pseudomonas aeruginosa* contribute to virulence and pathogenicity[2,6–10]. A key challenge when analysing isolates is the effort required to identify known molecules. The effort involved in determining activity, biochemical characterization, and structure elucidation represents a huge expense in terms of time and money and produces a wealth of information that is not easily accessible. Other fields such as sequencing have seen a large increase in the value of expensive data and have made the data searchable for the rest of the scientific public. Unlike the situation with sequencing data, we cannot take natural product data and compare and contrast this information with other previously collected data sets. However, with increasing computational power, the annotation of known molecules can be facilitated by creating a reference index.

Mass spectrometry (MS) has become an invaluable tool for natural product discovery due to its sensitivity and throughput. One challenge in specialized metabolite research is identifying known versus unknown metabolites detected by MS (ref. 11). Dereplication (the identification of known molecules) of natural products can be performed with the Dictionary of Natural Products, AntiBase and MarinLit[11–14]. However, these databases are behind paywalls, not searchable with raw data, and certainly not searchable with millions of data points at once. Global Natural Products Social Molecular Networking (GNPS), which uses tandem mass spectrometry (MS/MS) as a proxy for molecular structure, enables dereplication, visualization of molecular space as a network and enables propagation of chemical features to unidentified molecules[11,15–17].

[1]Department of Chemistry and Biochemistry, University of California San Diego, California 92093, USA. [2]Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, California 92093, USA. [3]Graduate School of Pharmaceutical Sciences, Kitasato University, Tokyo 108-8641, Japan. [4]Bioinformatics Group, Wageningen University, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands. [5]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, California 92093, USA. [6]Ion Torrent by Thermo Fisher, 5781 Van Allen Way, Carlsbad, California 92008, USA. [7]Centre of Microbial and Plant Genetics, KU Leuven, 3001 Heverlee, Belgium. [8]Department of Microbial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), 6708PB Wageningen, The Netherlands. [9]Department of Ecology and Evolutionary Biology, University of California, 1156 High Street, Santa Cruz, California 95064, USA. [10]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, California 92093, USA. [11]Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK. [12]School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK. [13]Department of AgroEcology, Rothamsted Research, West Common, Harpenden AL5 2JQ, UK. [†]These authors contributed equally to this work.
*e-mail: marnix.medema@wur.nl; pdorrestein@ucsd.edu

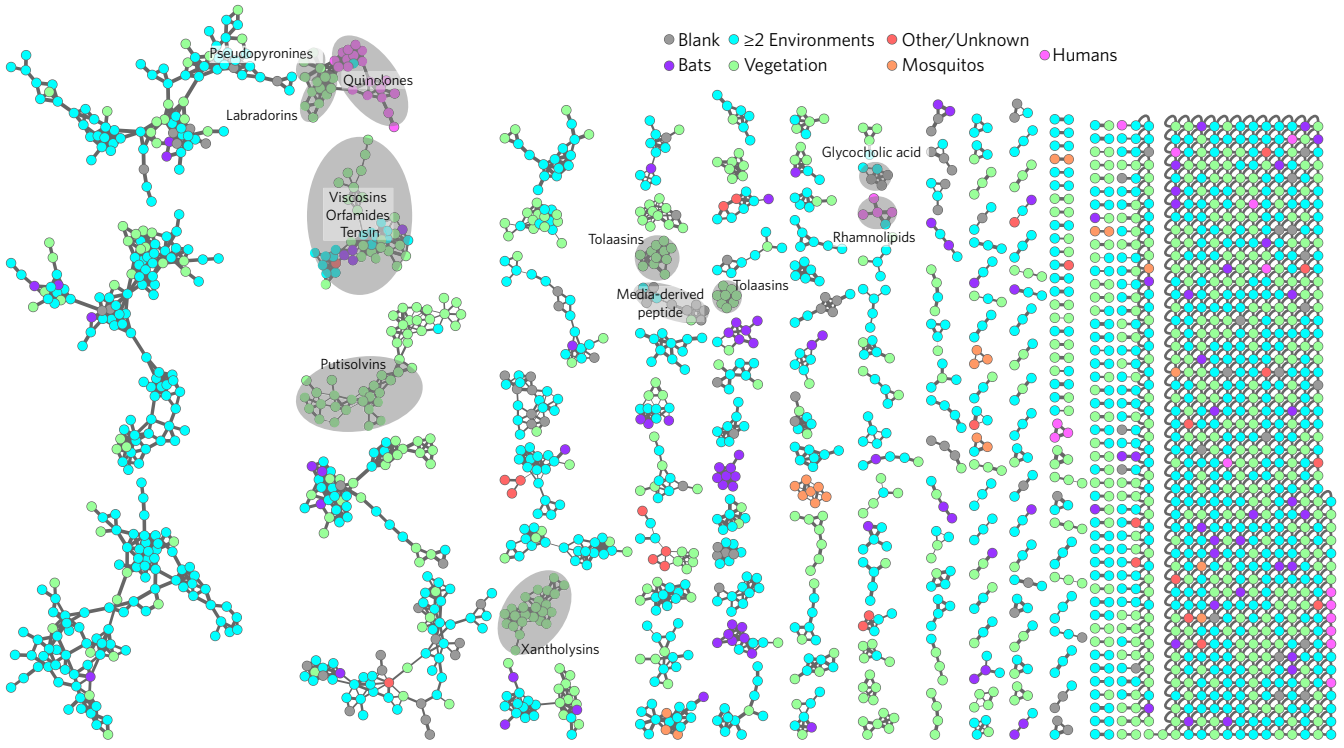


**Figure 1 | *Pseudomonas* molecular network: environmental isolation visualization.** Known *Pseudomonas* compounds that were observed in our data were dereplicated and highlighted. Node colours represent environments from which the strains were isolated. Grey nodes (blank) include blank injections, media controls, and internal standard (glycocholic acid).

## Curating a *Pseudomonas* specialized metabolite index

To create a reference metabolite index for *Pseudomonas*, multiple laboratories contributed to a collection of 260 pseudomonads, isolated from locations around the globe and from a range of environmental niches. Bacteria were cultured and extracts subjected to liquid chromatography tandem mass spectrometry (LC–MS/MS) (Supplementary Fig. 1 and Supplementary Tables 1 and 2). To generate a molecular map of the detectable metabolites, the LC–MS/MS data were subjected to molecular networking on GNPS and visualized in Cytoscape (Fig. 1; see Methods and Supplementary Fig. 2)[17,18]. Characteristics of the samples, such as environmental niche, molecular weight, geographic isolation and species-specific molecule production, were visualized in the network (Fig. 1 and Supplementary Fig. 2). We observe distinct molecules from environments even where fewer individual strains were analysed. For example, in the less well-represented *Pseudomonas*, the strain library contains 37 bat-, 3 mosquito- and 3 human-associated *Pseudomonas* isolates, and these environments show molecules not observed in other *Pseudomonas*[19]. This observation shows a correlation between a strain's environment and the molecules that are produced and begs the question of whether the environment dictates molecule production and whether certain molecules are required to thrive in specific niches[20,21]. The strain library consists of 21 different species of *Pseudomonas*, but is primarily composed of *P. putida* and *P. fluorescens* (68%, Supplementary Fig. 2 and Supplementary Table 1). A total of 5% of molecules are uniquely produced by *P. putida*, 10% by *P. fluorescens*, and 65% by two or more species, indicating that most are produced by multiple species (Supplementary Fig. 3 and Supplementary Tables 1 and 2)[4,8,22].

One of the challenges associated with specialized metabolite discovery is the rediscovery of previously characterized metabolites. Previous data are often spread between multiple databases and primary literature, or lost among laboratory notebooks. Although we aimed to use the dereplication feature of GNPS, where experimentally derived MS/MS are matched to the spectra of annotated and curated MS/MS spectra within the GNPS database, at the start of this project GNPS and other public MS/MS libraries did not contain many of the *Pseudomonas* specialized metabolites present in the literature, with the exception of lipid annotations[1,17]. We therefore manually dereplicated MS and MS/MS spectra against the literature. Using the 2009 review by Gross and Loper as a reference, there were found to be 119 natural products from pseudomonads belonging to 30 molecular families, after including the xantholysin, rhamnolipid, labradorin and pseudopyronine molecular families currently in the literature[1]. We observed nine of these families, or 30% of the *Pseudomonas* molecular families described (Fig. 1, Supplementary Figs 4 and 5 and Supplementary Tables 1–4). However, lack of molecular observation may occur for several reasons: (1) the strain(s) responsible for the production of a compound is not in our *Pseudomonas* collection; (2) the compounds are not produced in high enough titre to be observed; (3) the extraction conditions used, while broad, may not be suitable for some compounds; and (4) the current chromatography conditions select for more hydrophobic molecules.

For many observed molecules, examining MS/MS spectra for characteristic mass shifts can shed light on structural information. Mass shifts of 162 or 176 Da suggest sugar moieties, while mass shifts of 14, 28 or 42 Da suggest lipid or alkyl side chains[23]. Characteristic mass shifts were combined with accurate mass measurements and information about the samples (for example, bacterial genus and species), and were compared to literature values. Based on the Metabolomics Standards Initiative's reporting standards, manual dereplication of non-peptidic molecules resulted in level 2 (putatively annotated compounds) of known human-associated *Pseudomonas* metabolites for the rhamnolipid and quinolone molecular families, as well as the labradorin and pseudopyronine molecular families from vegetation-associated *Pseudomonas* (Fig. 1, Supplementary Fig. 3 and Supplementary Tables 1–4)[24–28]. The rhamnolipids are structurally distinct molecules produced by human-associated strains. Rhamnolipids behave as biosurfactants,

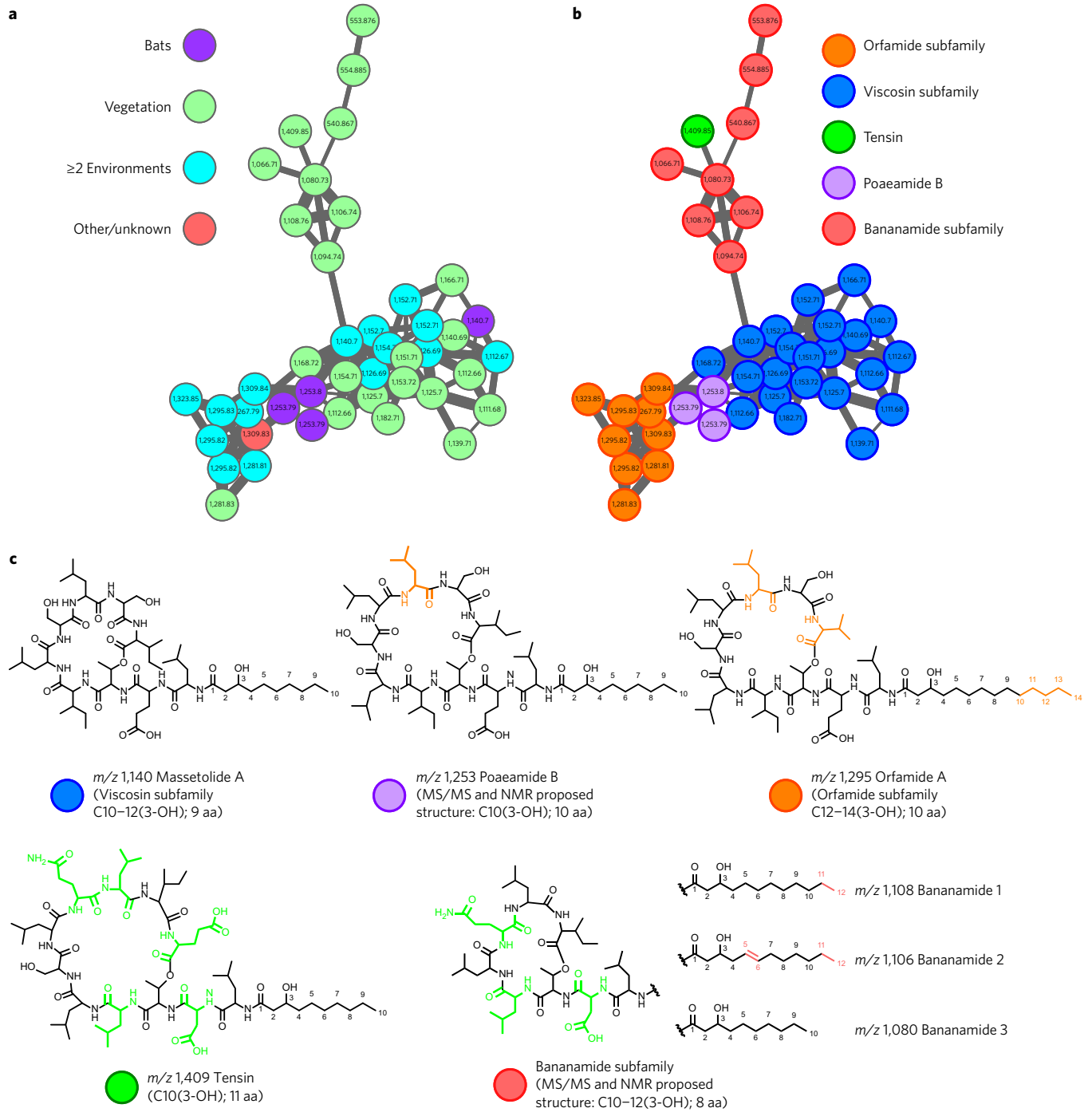


**Figure 2 | A molecular family comprising related peptides including viscosin, WLIP, viscosinamide, massetolides A–F, orfamides A–C and tensin.** **a**, The nonribosomal cyclic lipopeptides molecular family with environmental isolation mapping. Node colours delineate the environments from which the specific molecules of the cyclic lipopeptide family originate. **b**, The nonribosomal cyclic lipopeptide molecular family with node colours representing the subfamilies of the different molecules. **c**, Molecular representatives from each subfamily showing structural similarities. Massetolide A (from the viscosin subfamily) was chosen as a starting point and drawn in black. All other molecules were drawn in black where structural similarities to massetolide A existed. Portions of molecules highlighted in colours correspond to the subfamily and illustrate the structural differences in comparison to massetolide A.

promote the uptake and biodegradation of substrates, and act as immune modulators and virulence factors[29]. Similar to the rhamnolipids, the quinolones are produced by human-associated strains and behave as quorum signals that coordinate biofilm formation, virulence and antibiotic resistance[30]. Conversely, the labradorins and pseudopyronines are produced by vegetation-associated *Pseudomonas*, where the original characterizations are from a phytopathogen and plant-derived pseudomonads but can also be retrieved from a marine sponge-derived *Pseudomonas*. Both have antimicrobial properties[25–27,31,32].

For peptidic molecules, MS/MS spectra yield fragment ions with mass differences corresponding to amino acid monomers, where consecutive mass differences represent a *de novo* peptide sequence tag[33]. As with non-peptidic molecules, accurate masses and amino acid sequence tags can be compared to the literature. We were able to identify a number of peptide molecular

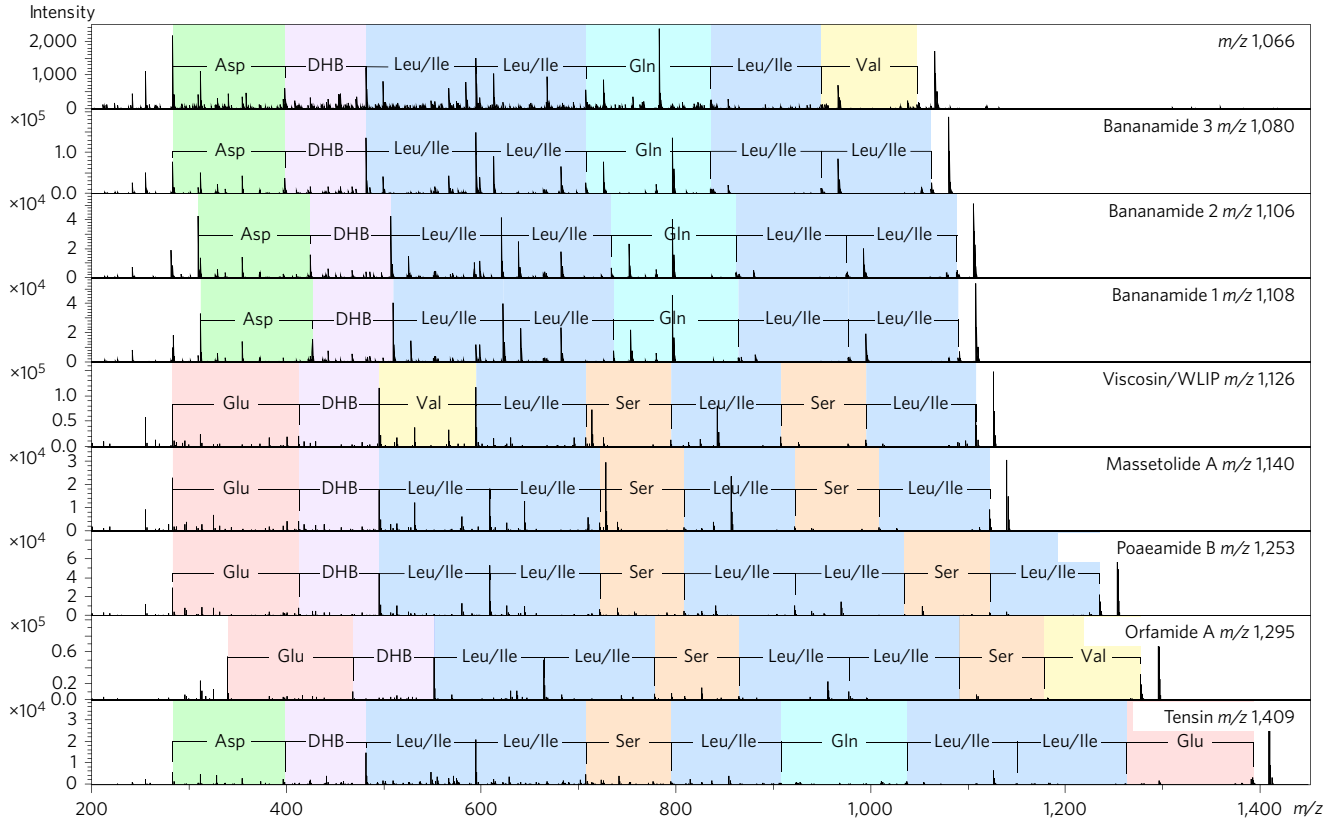


**Figure 3 | MS/MS comparison of select molecules from the cyclic lipopeptide molecular family.** Molecules are arranged by increasing *m/z*. Sequence tags for each molecule are shown and colour-coded according to amino acid monomer. All molecules form a lactone linkage with the C-terminus of isoleucine (massetolide A and poaeamide B), leucine (bananamides 1–3), valine (orfamide A), or glutamic acid (tensin) to threonine (observed in the MS/MS spectra as dehydrobutyrine, DHB).

families, including viscosin/white-line-inducing principle (WLIP)/massetolides, orfamides, putisolvins, xantholysins and tolaasins (Fig. 1 and Supplementary Fig. 3)[3–5,24,34–37]. All of these compounds are involved in motility, behave as biosurfactants and have antimicrobial, antiparasitic and antibiofilm activities[3–5,24,34–39]. All the MS/MS spectra associated with these annotations are publicly available at http://gnps.ucsd.edu (ref. 17).

Because molecular networking clusters molecules based on structural similarity, a single match to the GNPS *Pseudomonas* library allows for propagation of that structure through an entire molecular family. Of the metabolites we dereplicated, three seemingly separate molecular families (quinolones, labradorins and pseudopyronines) cluster together due to similarities in alkyl side chain fragmentation. More specifically, the alkyl side chains are adjacent to an olefin and attached to a heterocyclic moiety that does not readily fragment, thereby resulting in clustering, primarily due to alkyl fragmentation. Even though these molecules share a molecular family due to the inherent gas-phase behaviour defined by their chemical structure, the families are separated into subfamilies based on the subtleties of their structural differences.

## Discovery of specialized metabolites

Subfamilies are also observed in other molecular families. Figure 2 demonstrates a molecular family comprising related peptides including viscosin, WLIP, viscosinamide, massetolides A–F, orfamides A–C and tensin[4,34,40]. Differences due to amino acid substitution and varying fatty acid chains lead to the formation of subfamilies. Further analysis of the viscosin molecular subfamilies led to the identification of two uncharacterized members: a molecule at *m/z* 1,253 and the subfamily at *m/z* 1,108, 1,106, 1,094, 1,080 and 1,066, a subfamily most similar to tensin and massetolide A.

*m/z* 1,253, produced solely by *Pseudomonas synxantha* CR32, was isolated from the bat species *Myotis mystacinus* in the Hranice Abyss of the Czech Republic. MS/MS analysis yielded an amino acid sequence tag of Glu-Dhb-Ile/Leu-Ile/Leu-Ser-Ile/Leu-Ile/Leu-Ser-Ile/Leu, a tag similar to orfamide A (*m/z* 1,295) and massetolide A (*m/z* 1,140) (Fig. 3). Compared to orfamide A, *m/z* 1,253 substitutes an Ile/Leu for Val in the 10th position. Compared to massetolide A, *m/z* 1,253 contains an additional Ile/Leu (Figs 2–4). *m/z* 1,253 was isolated and NMR confirmed the identity of the amino acid residues predicted from MS/MS and revealed a C10 3-hydroxy fatty acid tail (Supplementary Fig. 6 and Supplementary Table 5). *m/z* 1,253 is similar to poaeamide A from *Pseudomonas poae* and we therefore call *m/z* 1,253 poaeamide B (Fig. 4)[22].

*m/z* 1,108, 1,106, 1,094, 1,080 and 1,066, which we now call the bananamides, could not be dereplicated. The bananamides are named as such because they are only found to be produced by *P. fluorescens* collected from the banana rhizoplane in the wetlands of Galagedara, Sri Lanka[41]. Analysis of the MS/MS data of *m/z* 1,108, 1,106 and 1,080 yielded the amino acid tag Asp-Dhb-Ile/Leu-Ile/Leu-Gln-Ile/Leu-Ile/Leu. The molecule at *m/z* 1,066 yielded a sequence tag Asp-Dhb-Ile/Leu-Ile/Leu-Gln-Ile/Leu-Val, where the 14 Da difference between *m/z* 1,066 and 1,080 is due to substitution of Ile/Leu for Val. Bananamides 1, 2 and 3 (*m/z* 1,108, 1,106 and 1,080) were purified, and NMR validated the sequence tag observed by MS/MS (Fig. 3, Supplementary Figs 7–9 and Supplementary Table 6). MS and integrated proton values provide evidence for a C12 3-hydroxy fatty acid in *m/z* 1,108, while *m/z* 1,080 contains a C10 3-hydroxy fatty acid (Supplementary Figs 7 and 9 and Supplementary Table 6). *m/z* 1,106 shows two olefinic protons with COSY correlations to two methylene protons that come from a C12 3-hydroxy unsaturated

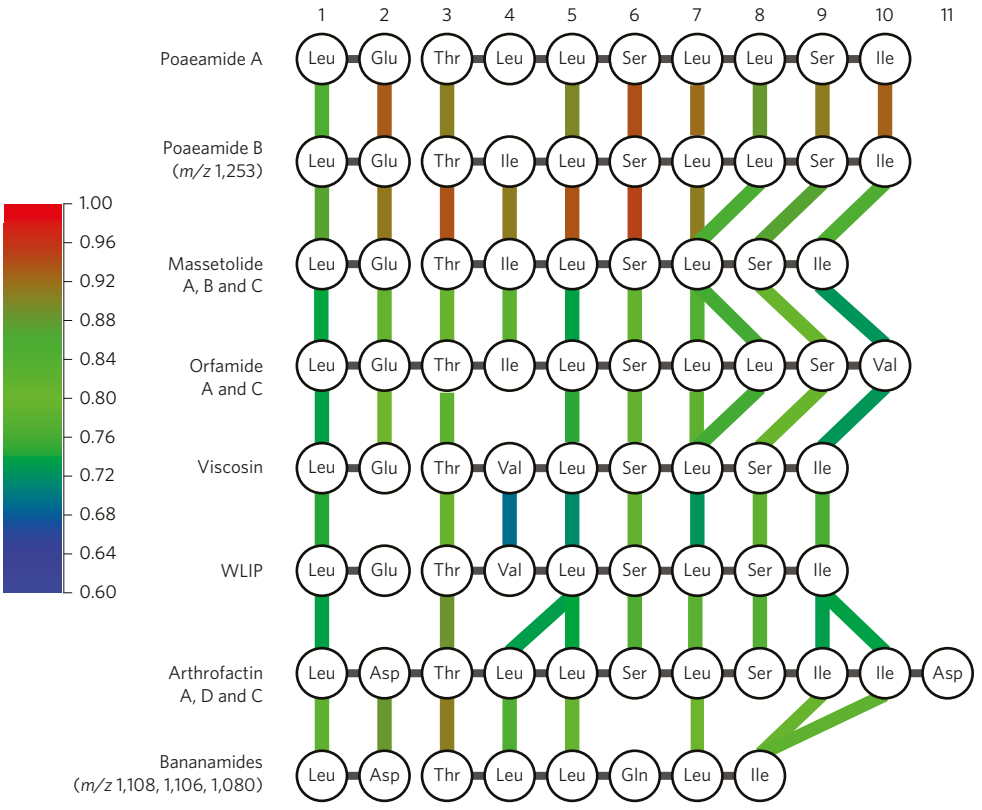


**Figure 4 | Evolutionary relationships between the assembly lines of the newly identified cyclic lipopeptides and structurally related molecules, based on their constituent phylogenetic groups of adenylation domains (Supplementary Fig. 12).** Each NRPS module is represented by its A-domain, together with its substrate specificity. Coloured lines connect A-domains in adjacently depicted assembly lines that belong to the same phylogenetic group; note that some clades contain multiple substrates due to recent evolution of the substrate specificity of an individual member (for example, Val-10 in the orfamide BGC) and that some substrate specificities are split across two clades (for example, Glu-2 of the WLIP BGC originates from a different Glu-specific clade). The colour gradient represents a cosine similarity score between two given A-domains.

fatty acid at the fifth position (Fig. 2, Supplementary Fig. 8 and Supplementary Table 6). Such unsaturations have only been observed in a few *Pseudomonas* cyclic lipopeptides[36,42–47]. Compared to massetolide A, the bananamides substitute the Glu with an Asp, the first Ser residue with a Gln and an Ile with a Leu. An equivalent of the second Ser residue is absent. Compared to tensin, the bananamides lack the Ser, Glu and one of the Leu residues (Figs 2 and 3)[4].

The automated peptidogenomics platform Pep2Path was used to attempt to match the MS/MS sequence tag data from poaeamide B and the bananamides to gene cluster families of public genome sequences[16,48]. None were found. The genomes of *P. synxantha* CR32 and *P. fluorescens* BW11P2 were therefore sequenced and subjected to antiSMASH analysis[49]. antiSMASH revealed a nonribosomal peptide synthetase (NRPS) gene cluster predicted to make the poaeamide B (Fig. 3, Supplementary Fig. 10 and Supplementary Table 7)[22], while the BW11P2 genome revealed a single NRPS gene cluster predicted to incorporate the amino acids consistent with the bananamide core peptide (Fig. 3, Supplementary Fig. 11 and Supplementary Table 8). Details of the biosynthetic gene cluster (BGC) for poeaemide B and the bananamides are summarized in the following MiBIG[50] links: http://mibig.secondarymetabolites.org/repository/BGC0001346/index.html#cluster-1 and http://mibig.secondarymetabolites.org/repository/BGC0001347/index.html#cluster-2 (Fig. 4, see Methods and Supplementary Figs 12–15). The Methods, Supplementary Figs 12–15 and Fig. 4 outline the evolutionary relationships among the BGCs and reveal that the structural relations observed in the metabolite index are mirrored in the genetic relationships of the BGCs of these molecules.

## Applying the *Pseudomonas* specialized metabolite index

The *Pseudomonas* metabolite index was then used to examine an alternative *Pseudomonas* data set. We compared the metabolite index curated from the original 260 strains with an additional 370 wheat-associated *Pseudomonas* extracts obtained from the UK to determine if our index aided molecular annotations (Supplementary Figs 4, 5 and 15 and Supplementary Tables 1 and 2). Twenty eight percent of detectable features are unique to our original collection, 39% are unique to the additional samples, and 33% of molecules overlap between the two collections (Supplementary Fig. 16). Our current *Pseudomonas* index contains nine annotated molecular families. By adding the additional 370 samples, seven of the nine molecular families increase in the number of contributing samples and are automatically annotated. The lipopeptide molecular family focused on here (Fig. 2) was produced by 34 strains out of the original 260 strains and increased to 97 contributing strains on addition of the 370 additional UK samples (Fig. 5). The same subfamilies from Fig. 2 are observed and highlighted in Fig. 5, but the addition of the UK samples increases the size of the overall molecular family and reveals many uncharacterized analogues (Fig. 5). Poaeamide B, which was only produced by a single strain from the original 260, is now produced by a total of 45 strains. Conversely, the bananamides are still only identified in a single strain. Indexing specialized metabolites enabled us to determine the frequency of molecular detection in large bacterial collections. The molecular family, on addition of the 370 extracts, reveals that many uncharacterized analogues are associated with the known subfamilies and even provides the insight that additional subfamilies remain to be discovered.

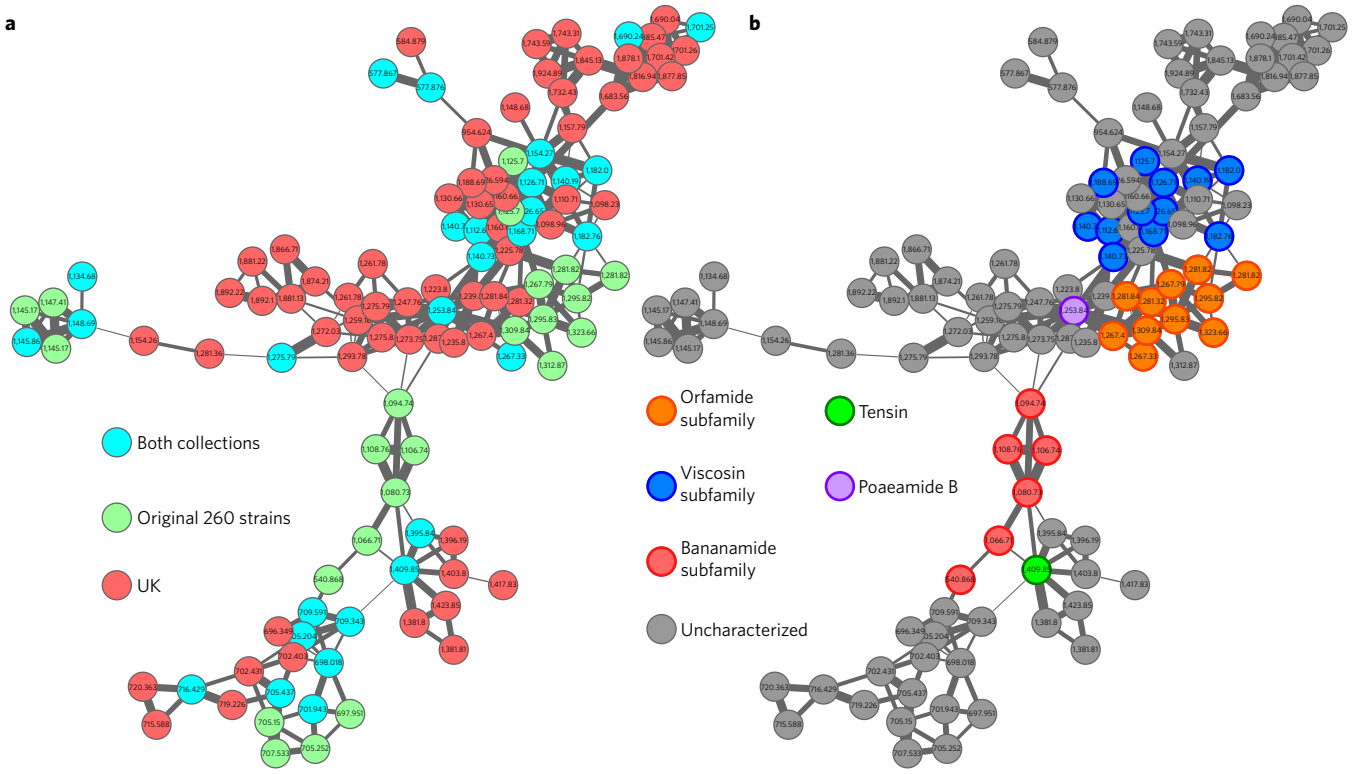


**Figure 5 | The cyclic lipopeptide molecular family upon addition of the 370 wheat-associated samples from the UK. a**, Node colours delineate which molecules are found from which *Pseudomonas* collection. The additional 370 wheat-associated UK samples are shown in red. **b**, The nonribosomal cyclic lipopeptide molecular family with node colours representing the subfamilies of the different molecules.

Ultimately, indexing known *Pseudomonas* compounds into GNPS allows for quick matching of these molecules when analysing alternate data sets, thereby increasing the speed at which molecular characterization can take place from large culture collections. The effectiveness of the index will only increase as molecular knowledge continues to be added to MS/MS spectra in GNPS.

## Conclusion

(Re)-discovery and (re)-characterization of molecules and their evolutionary relationships is a time-consuming and costly process, sometimes taking person-years to characterize a single molecule. The cost of dereplication, molecular annotation and structure elucidation is not often disclosed in manuscripts, but when we do consider these costs, it is clear that the scientific community must organize this type of information for efficient reutilization and make the data searchable. The structural prediction of poaeamide B and bananamides 1–3 took 79 days and cost roughly US$38,000 for all four molecules; this includes mass spectrometry costs and personnel salaries. The structural validation of poaeamide B after structural prediction based on MS/MS patterns and molecular family relationships to other known cyclic lipopeptides that were already established took 355 days and cost US$86,000, while structural validation for bananamides 1–3 took 90 days and cost US$25,000. These costs are small compared to the discovery of other molecules. In the past, discovery of these molecules would be published, but the data, and knowledge of the data, would not be searchable in the same way as gene sequences are searchable, thereby always making annotation of metabolomics data from microorganisms a time-consuming process. In comparison, if we had four genes of interest we could search these genes in public databases and know which organisms contain these sequences, which of these sequences are most similar, and whether or not the sequence/sequence products have been characterized experimentally. All of this analysis could be accomplished in an afternoon. Indexing reference metabolomes and molecules provides similar capabilities, which are currently the norm in sequence comparisons. For this reason, we believe in the importance of developing searchable indexes that are publicly accessible, allowing researchers to ask probing questions associated with evolutionary relatedness, uniqueness of molecules and chemical diversity. Such capabilities will also lead to new ways to look at metabolomics data.

## Methods

**Pseudomonad culture and extract conditions.** Frozen stocks of *Pseudomonas* spp. were inoculated into 600 µl of liquid tryptic soy broth (TSB, Bacto soybean-casein digest medium, 30 g $l^{-1}$) in 2.0 ml 96 deep well plates (Thermo Scientific, Nunc 2.0 ml DeepWell Plate). Cultures were grown overnight at 30 °C and 200 r.p.m. and then diluted 500× into a second 2.0 ml 96 deep well plate containing fresh TSB liquid. 5 µl of the 500× dilution was inoculated into a third 2.0 ml 96 deep well plate containing 600 µl TSB agar (15 g agar per l), sealed with 96 well-cap mats (Thermo Scientific, Nunc 96 well-cap mats) and incubated at 30 °C for 72 h. The cultures were extracted with 300 µl 50/50 vol/vol ethyl acetate (Fisher Scientific, HPLC grade)/methanol (Fisher Scientific, HPLC grade). The plates were resealed with the same 96 well-cap mats, sonicated for 10 min and extracted for an additional 50 min. A total of 250 µl of these crude extracts were transferred into a pre-washed 96-well plate (Agilent Technologies, 96-well plates, 0.5 ml, polypropylene) and lyophilized to dryness. The extract protocol was repeated once more for a total extract volume of 500 µl.

**LC–MS/MS analysis.** Dried samples were redissolved in 200 µl methanol and centrifuged for 5 min at 1,000 r.p.m. A 150 µl volume of material was transferred into a new 96-well plate containing 50 µl of 400 µM glycocholic acid (Calbiochem, sodium salt) to serve as an injection standard and quality control for the chromatography (final concentration 100 µM) and then sealed with Zone-Free Sealing Film (Excel Scientific). MS analysis was performed on a micrOTOF-Q II (Bruker Daltonics) mass spectrometer with an electrospray ionization (ESI) source, controlled by OTOF control and Hystar. MS spectra were acquired in positive ion mode over a mass range of 100–2,000 *m/z*. An external calibration with ESI-L Low Concentration Tuning Mix (Agilent Technologies) was performed prior to data acquisition and hexakis(1*H*,1*H*,3*H*-tetrafluoropropoxy)phosphazene (Synquest Laboratories) *m/z* 922.009798 was used as a lock mass internal calibrant during data acquisition. The following instrument settings were used for data acquisition: capillary voltage of 4,500 V, nebulizer gas (nitrogen) pressure of 3 bar, ion source temperature of 200 °C, dry gas flow of 9 l $min^{-1}$, source temperature and spectra

acquisition rate of 3 Hz for MS1 and MS2. Minutes 0–0.5 were sent to waste. Minutes 0.5–10 were recorded with auto MS/MS turned on. The ten most intense ions per MS1 scan were selected and subjected to collision-induced dissociation according to the following fragmentation and isolation list (values are *m/z*, isolation width and collision energy, respectively): 100, 4, 16; 300, 5, 24; 500, 6, 30; 1,000, 8, 40; 1,500, 10, 50; 2,000, 12, 70. In addition, the basic stepping function was used to fragment ions at 100 and 160% of the collision-induced dissociation (CID) energy calculated for each *m/z* from the above fragmentation and isolation list with a timing of 50% for each step. The basic stepping had a collision radiofrequency (RF) of 198 and 480 peak-to-peak voltage, $V_{PP}$, with a timing of 50% for each step and transfer time stepping of 75 and 92 µs with a timing of 50% for each step. The MS/MS active exclusion parameter was set to 5 and released after 0.5 min. The injected samples were chromatographically separated using an Agilent 1290 Infinity Binary LC System (Agilent Technologies) controlled by Hystar software (Bruker Daltonics), using a 50 × 2.1 mm Kinetex 1.7 µM, C18, 100 Å chromatography column (Phenomenex), 30 °C column temperature, 0.5 ml $min^{-1}$ flow rate, mobile phase A 99.9% water (J.T.Baker, LC-MS grade)/0.1% formic acid (Fisher Scientific, Optima LC/MS), mobile phase B 99.9% acetonitrile (J.T.Baker, LC-MS grade)/0.1% formic acid (Fisher Scientific, Optima LC/MS), with the following gradient: 0–0.5 min 10% B, 0.5–1 min 50% B, 1–6 min 100% B, 6–9 min 100% B, 9–9.5 min 10% B, 9.5–10 min 10% B. Blank (0 µl) injections—methanol injections containing glycocholic acid and agar treated and extracted under the same conditions as the culture conditions—were used as controls.

**Molecular networking.** All LC–MS/MS data were converted to mzXML format using Compass Data Analysis (Bruker Daltonics) and uploaded to the Global Natural Products Social Molecular Networking webserver (http://gnps.ucsd.edu). The LC–MS/MS data for the 260 *Pseudomonas* isolates were analysed using the Molecular Networking workflow with the following settings: Parent Mass Tolerance 0.9 Da, Ion Tolerance 0.45 Da, Min Pairs Cos 0.6, Min Matched Peaks 6, Network TopK 10, Minimum Cluster Size 2 and Maximum Connected Component Size 100. Molecular networking will merge all identical MS and MS/MS spectra, including identical MS/MS spectra of isomers. The molecular network was visualized using Cytoscape version 2.8.3 and displayed using an unweighted force directed layout. The data are publicly accessible at http://gnps.ucsd.edu under MassIVE accession no. MSV000079450 and the networking results and parameters can be found at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=c50e46ab24bc4e31a914c69e1df63b4e. On addition of the 370 wheat-associated *Pseudomonas* samples, Parent Mass Tolerance and Ion Tolerance were set to 1 Da and 0.5 Da, respectively, while Maximum Connected Component Size was increased to 100, to accommodate the additional 569,800 MS/MS scans used for the network. The remaining network settings were unchanged: Min Pairs Cos 0.6, Min Matched Peaks 6, Network TopK 10 and Minimum Cluster Size 2. At these settings, the GNPS community has determined that 1% of the annotations are incorrect, 4% not enough information to tell, 4% could be isomers or correct and 91% was determined to be correct[17]. The data are publicly accessible under MassIVE accession no. MSV000079619 and the networking results and parameters can be found at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=b28463fdb3ce4d6cbc4bc6ea0129fdf3. The index itself can be composed of MS/MS spectra from intact molecules, but also products of in-source fragments, different adducts ($Na^+$, $K^+$, $NH_4^+$, $Al^{3+}$, $Fe^{3+}$ and so on)[28], biosynthetic intermediates, biosynthetic diversification, isotopes ($^{13}C$, $^{34}S$, halogenated compounds) or shunt products and system impurities[51,52]. Because molecular networking is a map of the diversity of MS/MS spectra, the index contains all such possibilities. System impurities can be readily identified by ensuring one has the proper blanks, and are colour-coded grey in the network (Fig. 1). The majority of the time, in-source fragments are spotted in the following ways within the molecular network. (1) The MS/MS outputs of in-source fragments have a pattern similar to the parent, so such artefacts usually become a subcluster of a molecular family. (2) Retention times (RTs) of in-source fragments are always identical to the parent because are generated from the parent ion eluting from the column at a certain point in time. We export all the RTs for all ions in GNPS so they can be verified during analysis. For the peptides analysed in the networks, no in-source fragment ions were observed.

**Isolation of poaeamide B.** An overnight culture of *P. synxantha* CR32, isolated from the bat species *M. mystacinus* found in the Hranice Abyss of the Czech Republic, was prepared from frozen stock in 7 ml of TSB liquid medium in a 14 ml round-bottomed culture tube (Corning) and shaken at 200 r.p.m. and 30 °C. Overnight culture (20 µl) was used to inoculate 25 lawns of *P. synxantha* CR32 on 10 ml TSB agar plates. The lawns were incubated for 72 h at 30 °C, transferred into a 500 ml Erlenmeyer flask and extracted three times with 200 ml of 50/50 vol/vol ethyl acetate/methanol (Fischer). The extract supernatant was filtered away from the media and dried *in vacuo*. Dried crude material was dissolved in 2 ml of methanol and separated on an Agilent 1260 HPLC equipped with a 250 × 10 mm Discovery 5 µM, C18, 180 Å chromatography column (Supelco). LC conditions were as follows: 30 °C column temperature, 2.0 ml $min^{-1}$ flow rate, mobile phase A 99.9% water and 0.1% formic acid, mobile phase B 99.9% acetonitrile and 0.1% formic acid with the following isocratic gradient: 0–30 min 80% B, 31–33 min 100% B, 34–35 min 80% B. Poaeamide B was collected at 26–28 min on an MS-based fraction collection. Molecules were verified simultaneously by MS/MS fragmentation.

**Isolation of bananamides 1, 2 and 3.** An overnight culture of *P. fluorescens* BW11P2, isolated from the banana rhizoplane in Galgadera, Sri Lanka, was prepared from frozen stock in 7 ml of TSB liquid medium in a 14 ml round-bottomed culture tube (Corning) and shaken at 200 r.p.m. and 30 °C. Overnight culture (5 ml) was added to 50 ml of liquid TSB containing 1 g of sterile glass beads (3 mm diameter, Kimble Chase) and 1 g of sterile Amberlite XAD-16 resin (Sigma) and incubated for 10 days at 30 °C and 200 r.p.m. in a 250 ml Erlenmeyer flask. XAD-16 resin and cells were collected by vacuum filtration and extracted three times with 25 ml of 50/50 vol/vol ethyl acetate/methanol (Fischer) by shaking for 1 h at 200 r.p.m. The resin and glass beads were filtered and the crude extract supernatant dried *in vacuo*. To separate bananamides 1, 2 and 3, the dried crude material was dissolved in 5 ml methanol and separated on an Agilent 1260 HPLC equipped with a 250 × 10 mm Discovery 5 µM, C18, 180 Å chromatography column (Supelco). LC conditions were as follows: 30 °C column temperature, 2.0 ml $min^{-1}$ flow rate, mobile phase A 99.9% water and 0.1% formic acid, mobile phase B 99.9% acetonitrile and 0.1% formic acid, with the following isocratic gradient: 0–30 min 85% B, 31–33 min 100% B, 34–35 min 85% B. Bananamides 1, 2 and 3 were collected at 24.6–25.2, 20.6–21.2 and 16.1–16.7 min, respectively, by MS-based fraction collection. Molecules were verified simultaneously by MS/MS fragmentation.

**NMR measurements of poaeamide B and bananamides 1, 2 and 3.** 1D $^1H$-NMR, 2D $^1H$–$^1H$ double quantum filtered correlation spectroscopy (DQF-COSY), 2D $^1H$–$^{13}C$ heteronuclear single quantum coherence (HSQC) and 2D $^1H$–$^{13}C$ heteronuclear multiple bond correlation (HMBC) spectra of purified poaeamide B and bananamides 1, 2 and 3 were acquired at 25 °C using a 600 MHz NMR (Magnex superconducting magnet, 14.1T) fitted with a 1.7 mm cryoprobe and Bruker Avance II console operated using Bruker TopSpin 2.1 software. For NMR acquisition, 10–100 µg of poaeamide B and bananamides 1, 2 and 3 were dissolved in 50 µl $CD_3OD$ (Cambridge Isotope Laboratories).

**Genome sequencing, assembly and analysis.** Genomic DNA from *P. synxantha* CR32 (poaeamide B producer [*m/z* 1,253], accession nos. KU936045 and KU936046) and *P. fluorescens* BW11P2 (bananamides producer [*m/z* 1,108, 1,106 and 1,080], accession nos. LRUN00000000 and KX437753 for the bananamide BGC) was isolated using a Wizard Genomic DNA Purification Kit (Promega) in $n = 3$ biological replicates. Sequencing libraries were constructed from 1 µg of genomic DNA using the Ion Xpress Plus Fragment Library Kit (ThermoFisher). DNA was sheared using the Covaris S2 (Covaris) to an average of 400 bp. After nick-repair and adapter ligation, the Pippin Prep instrument (Sage Science) was used to size select for 475 bp fragments using a 2% agarose gel DF cassette with Marker L, following the standard protocol. The library was quantified using a DNA High Sensitivity kit on the BioAnalyzer 2100 system (Agilent). The Ion PGM Template OT2 Kit (ThermoFisher) was used for sample preparation with the Ion OneTouchTM 2 System with a modified thermoprofile. Changes to the thermoprofile included an increase in melting temperature to 97 °C and extended cycling parameters. Sequencing was performed using an Ion Torrent Personal Genome Machine (ThermoFisher) with an Ion PGM Hi-Q Sequencing Kit (ThermoFisher), according to the standard protocol, on a 318v2 sequencing chip (ThermoFisher). *De novo* genome assembly was performed using CLC Genomics Workbench software v5.01 (CLC bio), and the full bananamide BGC was reconstructed by combining this with a second assembly using SPAdes (ref. 53). The *P. synxantha* CR32 poaeamide B gene cluster was separated across two chromosomal regions, spanning 9.9 and 31.3 kb, respectively. The *P. fluorescens* BW11P2 genome assembled into 6.0 Mb of 130 contigs with an N50 of 87 kb. BGCs in the genomes were analysed with antiSMASH and processed with custom Python scripts. BGC annotations were submitted to MIBiG (ref. 50) under accession nos. BGC0001346 (bananamides) and BGC0001347 (poaeamide B). Phylogenetic analysis was performed using MEGA 7.0 (ref. 54).

To obtain an overview of the evolutionary relationships of BGCs of different types of *Pseudomonas* cyclic lipopeptides, we compiled a list of 18 different biosynthetic gene clusters of cyclic lipopeptides. The phylogenetic tree of the adenylation domains contained distinct functional clades in which the adenylation domains share the same amino acid substrate specificity (Supplementary Fig. 12)[36,52]. Several subgroups of cyclic lipopeptides were identified that are more distantly related to poaeamide B and the bananamides. The BGCs encoding larger assembly line structures, such as the syringopeptin BGC, have lower overall sequence and architectural similarity. Poaeamide B and bananamide BGCs are closely related to six other BGCs (arthrofactin, orfamide, massetolide, poaeamide A, viscosin and WLIP). Using this tree, we constructed pseudo-sequences of adenylation domain clades for all BGCs that represent the functional architectures of the encoded assembly lines to estimate the evolutionary distances between the gene clusters (Supplementary Fig. 12). We used the distance metric with domain types defined as the adenylation domain clades from Supplementary Fig. 12 and with weights of the Jaccard index, Goodman–Kruskal gamma index and domain duplication index at 0.5, 0.25 and 0.25. This analysis revealed many interesting aspects of evolutionary relationships that manifest themselves in the MS/MS data of the molecules found in the index. Overall, four other subfamilies of cyclic lipopeptide BGCs can be distinguished: sessilin/tolaasin, syringopeptin/nunapeptin/chicopeptin, putisolvin/entolysin/xantholysin and cichofactin/syringafactin. Some

pathways are encoded on two separate genomic loci, while others are encoded in a single BGC configuration. The distribution of these two architectural configurations is notably discontinuous, also when plotted onto a phylogeny of C-starter domains, which constitute the most conserved part of the assembly lines. This suggests that multiple independent split/join events might have taken place during the evolution of this BGC family (Supplementary Fig. 13). To understand specific evolutionary events on the domain level, such as duplications, deletions and insertions, a two-dimensional clustered heatmap was constructed (Supplementary Fig. 14). The BGC of poaeamide B is related to BGCs that encode the production of poaeamide A, massetolides and orfamides (Fig. 4 and Supplementary Figs 14 and 15). Although almost all A-domain sequences of poaeamide A and poaeamide B show similarity, the poaeamide B BGC distinguishes itself from the poaeamide A BGC by the presence of a distinct A-domain substituting an Ile for a Leu residue in the 4th position (Supplementary Fig. 15). In addition, poaeamide B biosynthesis shows similarity to the massetolide gene cluster with a duplication of the seventh A-domain that activates leucine. The bananamides, however, are more of a molecular and evolutionary outlier. As reflected in the comparisons of the biosynthetic machineries, the first five modules of the bananamide NRPS assembly line are similar and co-linear with those of the arthrofactin gene cluster. The observed evolutionary relationships between these BGCs corroborate the structural relationships of the molecules visualized by molecular networking (Figs 2 and 4 and Supplementary Figs 14 and 15). Conservation on the domain level, however, allows for the identification of the evolutionary modularity underlying their structures[55]. For example, the substructure synthesized by modules 5-6-7 in the poaeamide A, poaeamide B, massetolide, orfamide, WLIP, viscosin and arthrofactin assembly lines is shared among all of these molecules (Fig. 4). The module conservation indicates a possible key role of this Leu-Ser-Leu substructure in mediating the biological activity of this group.

**Code availability.** All data and Python scripts used for generating Fig. 4 and Supplementary Figs 12–14 are available at https://git.wageningenur.nl/Xiaowen/pseudomonas/tree/master.

**Data availability.** LC–MS/MS data are publicly accessible under MassIVE accession no. MSV000079450 or can be accessed at https://gnps.ucsd.edu/ProteoSAFe/result.jsp?task=5728ca4b0dfd4c058e0ef6151a31f9c4&view=advanced_view.

Molecular networking results and parameters can be found at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=c50e46ab24bc4e31a914c69e1df63b4e. The full genome sequencing data for *Pseudomonas fluorescens* strain BW11P2 (bananamides producer) can be found under NCBI accession no. LRUN00000000. The genome sequencing data for the bananamide BGC can be found under NCBI accession no. KX437753 and MiBIG accession no. BGC0001346. The genome sequencing data for the poaeamide B BGC from *Pseudomonas synxantha* strain CR32 (poaeamide B producer) can be found under NCBI accession nos. KU936045 and KU936046. In addition, the poaeamide B BGC can be found under MiBIG accession no. BGC0001347.

Received 28 April 2016; accepted 9 September 2016; published 31 October 2016; corrected 23 January 2017

## References


1. Gross, H. & Loper, J. E. Genomics of secondary metabolite production by *Pseudomonas* spp. *Nat. Prod. Rep.* **26,** 1408–1446 (2009).
2. Kloepper, J. W., Leong, J., Teintze, M. & Schroth, M. N. *Pseudomonas* siderophores: a mechanism explaining disease-suppressive soils. *Curr. Microbiol.* **4,** 317–320 (1980).
3. Ron, E. Z. & Rosenberg, E. Natural roles of biosurfactants. *Environ. Microbiol.* **3,** 229–236 (2001).
4. Raaijmakers, J. M., de Bruijn, I. & de Kock, M. J. D. Cyclic lipopeptide production by plant-associated *Pseudomonas* spp.: diversity, activity, biosynthesis, and regulation. *Mol. Plant. Microbe Interact.* **19,** 699–710 (2006).
5. Raaijmakers, J. M., De Bruijn, I., Nybroe, O. & Ongena, M. Natural functions of lipopeptides from *Bacillus* and *Pseudomonas*: more than surfactants and antibiotics. *FEMS Microbiol. Rev.* **34,** 1037–1062 (2010).
6. Davies, J. Specialized microbial metabolites: functions and origins. *J. Antibiot.* **66,** 361–364 (2013).
7. Ichinose, Y., Taguchi, F. & Mukaihara, T. Pathogenicity and virulence factors of *Pseudomonas syringae*. *J. Gen. Plant Pathol.* **79,** 285–296 (2013).
8. Raaijmakers, J. M., Vlami, M. & de Souza, J. T. Antibiotic production by bacterial biocontrol agents. *Antonie Van Leeuwenhoek* **81,** 537–547 (2002).
9. Mascuch, S. J. *et al.* Direct detection of fungal siderophores on bats with white-nose syndrome via fluorescence microscopy-guided ambient ionization mass spectrometry. *PLoS ONE* **10,** e0119668 (2015).
10. Hoyt, J. R. *et al.* Bacteria isolated from bats inhibit the growth of *Pseudogymnoascus destructans*, the causative agent of white-nose syndrome. *PLoS ONE* **10,** e0121329 (2015).
11. Yang, J. Y. *et al.* Molecular networking as a dereplication strategy. *J. Nat. Prod.* **76,** 1686–1699 (2013).
12. Laatsch, H. *AntiBase 2014: The Natural Compound Identifier* (Wiley, 2014).
13. Buckingham, J. *Dictionary of Natural Products* Supplement 4 (Taylor & Francis, 1997).
14. Blunt, J. W. & Munro, M. *MarinLit. A Database of the Literature on Marine Natural Products* (Royal Society of Chemistry, 2003); http://pubs.rsc.org/marinlit/
15. Watrous, J. *et al.* Mass spectral molecular networking of living microbial colonies. *Proc. Natl Acad. Sci. USA* **109,** E1743–E1752 (2012).
16. Nguyen, D. D. *et al.* MS/MS networking guided analysis of molecule and gene cluster families. *Proc. Natl Acad. Sci. USA* **110,** E2611–E2620 (2013).
17. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34,** 828–837 (2016).
18. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27,** 431–432 (2011).
19. Dong, Y., Manfredini, F. & Dimopoulos, G. Implication of the mosquito midgut microbiota in the defense against malaria parasites. *PLoS Pathog.* **5,** e1000423 (2009).
20. Jošić, D. *et al.* Antifungal activities of indigenous plant growth promoting *Pseudomonas* spp. from alfalfa and clover rhizosphere. *Front. Life Sci.* **8,** 131–138 (2015).
21. Pauwelyn, E. *et al.* New linear lipopeptides produced by *Pseudomonas cichorii* SF1-54 are involved in virulence, swarming motility, and biofilm formation. *Mol. Plant Microbe Interact.* **26,** 585–598 (2013).
22. Zachow, C. *et al.* The novel lipopeptide poaeamide of the endophyte *Pseudomonas poae* RE*1-1-14 is involved in pathogen suppression and root colonization. *Mol. Plant Microbe Interact.* **28,** 800–810 (2015).
23. Kersten, R. D. *et al.* Glycogenomics as a mass spectrometry-guided genome-mining method for microbial glycosylated molecules. *Proc. Natl Acad. Sci. USA* **110,** E4407–E4416 (2013).
24. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis chemical analysis working group (CAWG) metabolomics standards initiative (MSI). *Metabolomics* **3,** 211–221 (2007).
25. Pettit, G. R. *et al.* Isolation of labradorins 1 and 2 from *Pseudomonas syringae* pv. *coronafaciens*. *J. Nat. Prod.* **65,** 1793–1797 (2002).
26. Kong, F., Singh, M. P. & Carter, G. T. Pseudopyronines A and B, alpha-pyrones produced by a marine *Pseudomonas* sp. F92S91, and evidence for the conversion of 4-hydroxy-α-pyrone to 3-furanone. *J. Nat. Prod.* **68,** 920–923 (2005).
27. Chu, M. *et al.* Structure of sch 419560, a novel α-pyrone antibiotic produced by *Pseudomonas fluorescens*. *J. Antibiot.* **55,** 215–218 (2002).
28. Moree, W. J. *et al.* Interkingdom metabolic transformations captured by microbial imaging mass spectrometry. *Proc. Natl Acad. Sci. USA* **109,** 13811–13816 (2012).
29. Abdel-Mawgoud, A. M., Lépine, F. & Déziel, E. Rhamnolipids: diversity of structures, microbial origins and roles. *Appl. Microbiol. Biotechnol.* **86,** 1323–1336 (2010).
30. Gonçalves-de-Albuquerque, C. F. *et al.* Possible mechanisms of *Pseudomonas aeruginosa*-associated lung disease. *Int. J. Med. Microbiol.* **306,** 20–28 (2016).
31. Grundmann, F. *et al.* Identification and isolation of insecticidal oxazoles from *Pseudomonas* spp. *Beilstein J. Org. Chem.* **8,** 749–752 (2012).
32. Bauer, J. S. *et al.* Biosynthetic origin of the antibiotic pseudopyronines A and B in *Pseudomonas putida* BW11M1. *ChemBioChem* **16,** 2491–2497 (2015).
33. Kersten, R. D. *et al.* A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat. Chem. Biol.* **7,** 794–802 (2011).
34. Gross, H. *et al.* The genomisotopic approach: a systematic method to isolate products of orphan biosynthetic gene clusters. *Chem. Biol.* **14,** 53–63 (2007).
35. Kuiper, I. *et al.* Characterization of two *Pseudomonas putida* lipopeptide biosurfactants, putisolvin I and II, which inhibit biofilm formation and break down existing biofilms. *Mol. Microbiol.* **51,** 97–113 (2004).
36. Li, W. *et al.* The antimicrobial compound xantholysin defines a new group of *Pseudomonas* cyclic lipopeptides. *PLoS ONE* **8,** e62946 (2013).
37. Rainey, P. B., Brodey, C. L. & Johnstone, K. Biological properties and spectrum of activity of tolaasin, a lipodepsipeptide toxin produced by the mushroom pathogen *Pseudomonas tolaasii*. *Physiol. Mol. Plant Pathol.* **39,** 57–70 (1991).
38. Tran, H., Kruijt, M. & Raaijmakers, J. M. Diversity and activity of biosurfactant-producing *Pseudomonas* in the rhizosphere of black pepper in Vietnam. *J. Appl. Microbiol.* **104,** 839–851 (2008).
39. Kruijt, M., Tran, H. & Raaijmakers, J. M. Functional, genetic and chemical characterization of biosurfactants produced by plant growth-promoting *Pseudomonas putida* 267. *J. Appl. Microbiol.* **107,** 546–556 (2009).
40. Rokni-Zadeh, H. *et al.* Genetic and functional characterization of cyclic lipopeptide white-line-inducing principle (WLIP) production by rice rhizosphere isolate *Pseudomonas putida* RW10S2. *Appl. Environ. Microbiol.* **78,** 4826–4834 (2012).
41. Vlassak, K., Van Holm, L., Duchateau, L., Vanderleyden, J. & De Mot, R. Isolation and characterization of fluorescent *Pseudomonas* associated with the roots of rice and banana grown in Sri Lanka. *Plant Soil* **145,** 51–63 (1992).
42. Emanuele, M. C. *et al.* Corpeptins, new bioactive lipodepsipeptides from cultures of *Pseudomonas corrugata*. *FEBS Lett.* **433,** 317–320 (1998).

43. Pauwelyn, E. *Epidemiology and Pathogenicity Mechanisms of Pseudomonas cichorii, the Causal Agent of Midrib Rot in Greenhouse-Grown Butterhead Lettuce (Lactuca sativa L.)* (Ghent University, 2012).
44. Morikawa, M. *et al.* A new lipopeptide biosurfactant produced by *Arthrobacter* sp. strain MIS38. *J. Bacteriol.* **175,** 6459–6466 (1993).
45. D'aes, J. *et al.* To settle or to move? The interplay between two classes of cyclic lipopeptides in the biocontrol strain *Pseudomonas* CMR12a. *Environ. Microbiol.* **16,** 2282–2300 (2014).
46. Huang, C.-J. *et al.* Characterization of cichopeptins, new phytotoxic cyclic lipodepsipeptides produced by *Pseudomonas cichorii* SF1-54 and their role in bacterial midrib rot disease of lettuce. *Mol. Plant. Microbe. Interact.* **28,** 1009–1022 (2015).
47. Lange, A., Sun, H., Pilger, J., Reinscheid, U. M. & Gross, H. Predicting the structure of cyclic lipopeptides by bioinformatics: structure revision of arthrofactin. *ChemBioChem* **13,** 2671–2675 (2012).
48. Medema, M. H. *et al.* Pep2Path: automated mass spectrometry-guided genome mining of peptidic natural products. *PLoS Comput. Biol.* **10,** e1003822 (2014).
49. Weber, T. *et al.* antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43,** W237–W243 (2015).
50. Medema, M. H. *et al.* Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* **11,** 625–631 (2015).
51. Vizcaino, M. I. & Crawford, J. M. The colibactin warhead crosslinks DNA. *Nat. Chem.* **7,** 411–417 (2015).
52. Medema, M. H., Cimermancic, P., Sali, A., Takano, E. & Fischbach, M. A. A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis. *PLoS Comput. Biol.* **10,** e1004016 (2014).
53. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19,** 455–477 (2012).
54. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33,** 1870–1874 (2016).
55. Medema, M. H., Takano, E. & Breitling, R. Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol. Biol. Evol.* **30,** 1218–1223 (2013).


## Acknowledgements

Financial support was provided by the National Institutes of Health (NIH) grant GM097509 (to B.S.M. and P.C.D.). M.H.M. was supported by Rubicon (825.13.001) and Veni (863.15.002) grants from the Netherlands Organization for Scientific Research (NWO). J.R. and V.J.C. were supported by a grant from the Netherlands BEBasic Foundation (project F07.003.01) R.D.M. was supported by KU Leuven grant GOA/011/2008. M.G.K.G. is the recipient of a post-doctoral fellowship from FWO Vlaanderen (12M4615N). A.M.K. and T.L.C. were supported by NSF grants DEB-1115895 and DEB-1336290 and US FWS grant F12AP01081. L.M.S. was supported by a National Institutes of Health IRACDA K12 GM068524 grant award. J.G.M. was supported by BBSRC Institute Strategic Program (ISPG) grant BB/J004553/1 and University of East Anglia start-up funding. T.H.M. was supported by the BBSRC Institute Strategic Program (ISPG) 'Optimization of nutrients in soil-plant systems' (BBS/E/C/00005196). The authors acknowledge Bruker and NIH grants GMS10RR029121 and P41-GM103484 for support in the form of shared instrumentation and the computational infrastructure. NMR data were acquired at the University of California, San Diego Skaggs School of Pharmacy and Pharmaceutical Sciences NMR Facility. The authors thank V.V. Phelan for a review of this manuscript and M. Wang, A.T. Nelson and L.-F. Nothias-Scaglia for contributions.

## Author contributions

D.D.N., A.V.M., X.L., M.H.M. and P.C.D. designed the research. D.D.N., A.V.M., N.K., X.L., M.S., J.F., K.A., T.L.L., B.M.D., B.S.M., M.H.M. and P.C.D. performed research. D.D.N., A.V.M., N.K., X.L., M.S., M.G.K.G., J.F., B.M.D., R.D.M., M.H.M. and P.C.D. analysed data. M.G.K.G., V.J.C., T.L.C., J.G.M., T.H.M., L.M.S., A.M.K., J.M.R. and R.D.M. contributed microbial strains or extracts. D.D.N., A.V.M. and P.C.D. wrote the paper.

## Additional information

**Supplementary information** is available for this paper.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Correspondence and requests for materials** should be addressed to M.H.M. and P.C.D.

**How to cite this article:** Nguyen, D. D. *et al.* Indexing the *Pseudomonas* specialized metabolome enabled the discovery of poaeamide B and the bananamides. *Nat. Microbiol.* **2,** 16197 (2016).

## Competing interests

The authors declare no competing financial interests.

# Erratum: Indexing the *Pseudomonas* specialized metabolome enabled the discovery of poaeamide B and the bananamides

Don D. Nguyen, Alexey V. Melnik, Nobuhiro Koyama, Xiaowen Lu, Michelle Schorn, Jinshu Fang, Kristen Aguinaldo, Tommie L. Lincecum Jr, Maarten G. K. Ghequire, Victor J. Carrion, Tina L. Cheng, Brendan M. Duggan, Jacob G. Malone, Tim H. Mauchline, Laura M. Sanchez, A. Marm Kilpatrick, Jos M. Raaijmakers, René De Mot, Bradley S. Moore, Marnix H. Medema and Pieter C. Dorrestein

*Nature Microbiology* **2,** 16197 (2016); published online 31 October 2016; corrected 23 January 2017

In the original version of this Article, co-author René De Mot's name was coded wrongly resulting in it being incorrect when exported to citation databases. This has now been corrected, though no visible changes will be apparent.