Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Omics-based strategies to discover novel classes of RiPP natural products

Alexander M Kloosterman[1], Marnix H Medema[2] and Gilles P van Wezel[1]

Ribosomally synthesized and post-translationally modified peptides (RiPPs) form a highly diverse class of natural products, with various biotechnologically and clinically relevant activities. A recent increase in discoveries of novel RiPP classes suggests that currently known RiPPs constitute just the tip of the iceberg. Genome mining has been a driving force behind these discoveries, but remains challenging due to a lack of universal genetic markers for RiPP detection. In this review, we discuss how various genome mining methodologies contribute towards the discovery of novel RiPP classes. Some methods prioritize novel biosynthetic gene clusters (BGCs) based on shared modifications between RiPP classes. Other methods identify RiPP precursors using machine-learning classifiers. The integration of such methods as well as integration with other types of omics data in more comprehensive pipelines could help these tools reach their potential, and keep pushing the boundaries of the chemical diversity of this important class of molecules.

Addresses
[1] Institute of Biology, Leiden University, Sylviusweg 72, 2333BE Leiden, The Netherlands

[2] Bioinformatics Group, Wageningen University, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands

Corresponding authors: Medema, Marnix H (marnix.medema@wur.nl), van Wezel, Gilles P (g.wezel@biology.leidenuniv.nl)

Current Opinion in Biotechnology 2021, **69**:60–67

This review comes from a themed issue on **Pharmaceutical biotechnology**

Edited by **Blaine A Pfeifer** and **Guojian Zhang**

For a complete overview see the Issue and the Editorial

Available online 28th December 2020

https://doi.org/10.1016/j.copbio.2020.12.008

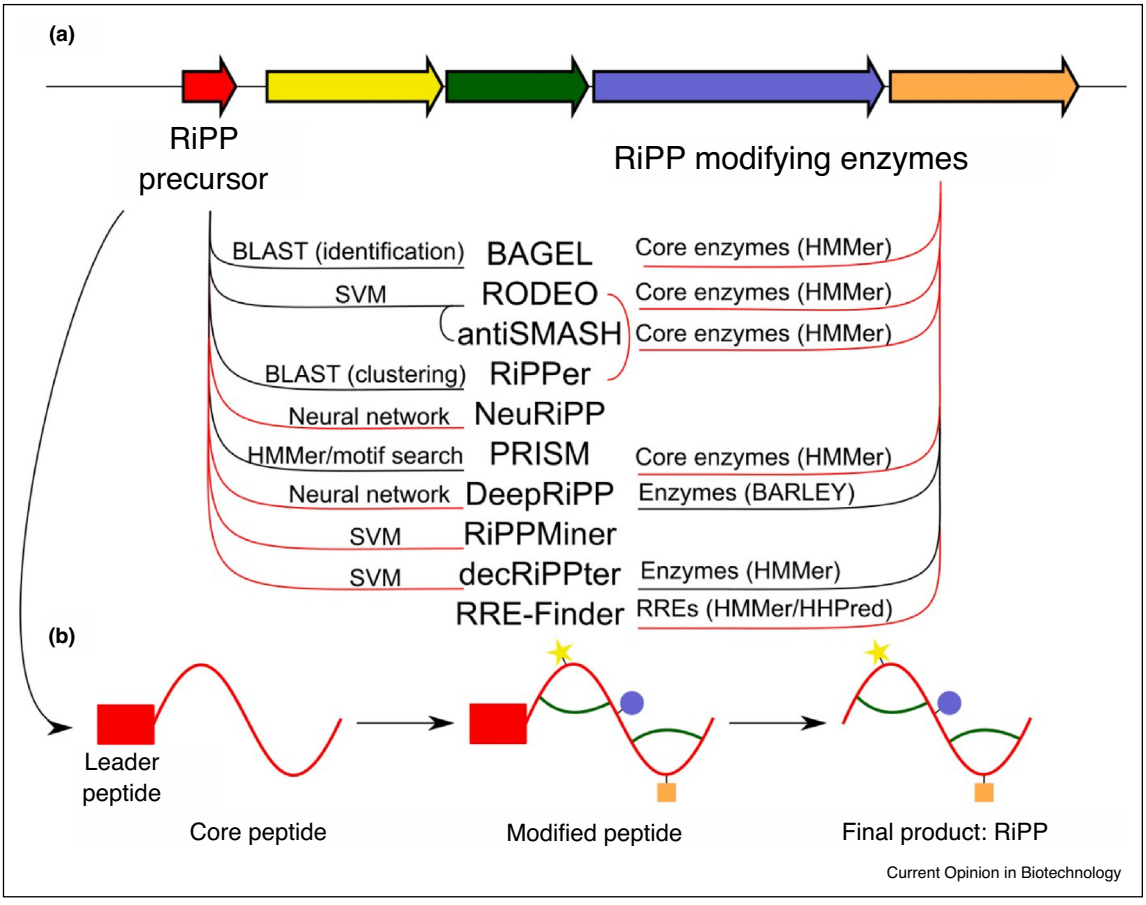
Introduction

Small organic molecules from a biological origin, collectively called natural products, comprise a dazzling array of diverse chemical structures [1]. These molecules play many different roles in nature, including interspecies signaling, resource competition and host defense. As such, bioactive natural products find their way into the clinic, as a result of their antimicrobial, antifungal, anticancer, antiviral or immunosuppressant activities. The discovery of antibiotics has had a huge impact on human life span, and in the mid-20th century bacterial infections seemed like a call from the past. However, multi-drug resistance and newly emerging infectious diseases once more impose serious health threats, and new drugs are now needed more than ever [2,3]. Meanwhile, traditional high-throughput screening (HTS) efforts lost their power, primarily due to the problem of replication, the rediscovery of known compounds [4,5]. Next-generation sequencing efforts surprisingly revealed that the capacity of bacteria to produce natural products had been grossly underestimated. This has led to a revolution in drug discovery based on the efficient mining of the rapidly growing genome sequence data [6]. Numerous tools and databases have been developed to explore, compare and catalogue biosynthetic gene clusters (BGCs) and their chemical products [7–11]. The discrepancy between the low return on investment of HTS and the apparent massive reservoir of biosynthetic potential is likely explained by the fact that the compounds are not produced under laboratory conditions. In other words, the BGCs that specify these natural products are only activated in response to specific environmental signals [12,13], and are therefore referred to as cryptic or silent BGCs.

The biosynthesis of natural products typically requires a distinct set of conserved enzymes that is responsible for their biosynthesis. Their encoding genes can be used as a basis for the detection of BGCs of that class [14,15]. Completely new biosynthetic pathways and natural product classes are much harder to discover with bioinformatics alone, however. A few methods that do not rely on the detection of known marker domains have been developed for that purpose, leading to the discovery of new chemical scaffolds [16–19].

One class of natural products which likely still comprises a large amount of hidden chemical diversity, is that of the ribosomally synthesized and post-translationally modified peptides (RiPPs) [20,21••]. RiPP biosynthesis always follows the same logic—a precursor peptide is ribosomally synthesized, biochemically modified and finally cleaved, resulting in the finished product (Figure 1)—but numerous different precursors and sets of modifying enzymes acting on them leads to large structural diversity (Figure 2). Their functions are equally diverse, and

Figure 1



Common biosynthetic logic as foundation for genome mining.
(a) RiPP BGCs contain a gene for a precursor peptide and genes encoding modifying enzymes, either of which can be used as a target. The red lines indicate a method's main way of identifying a BGC of interest. **(b)** Template for RiPP biosynthesis. A gene is translated into a precursor peptide, which has both a leader and a core region. The core is modified by modifying enzymes, and the leader is cleaved off, resulting in the final product.

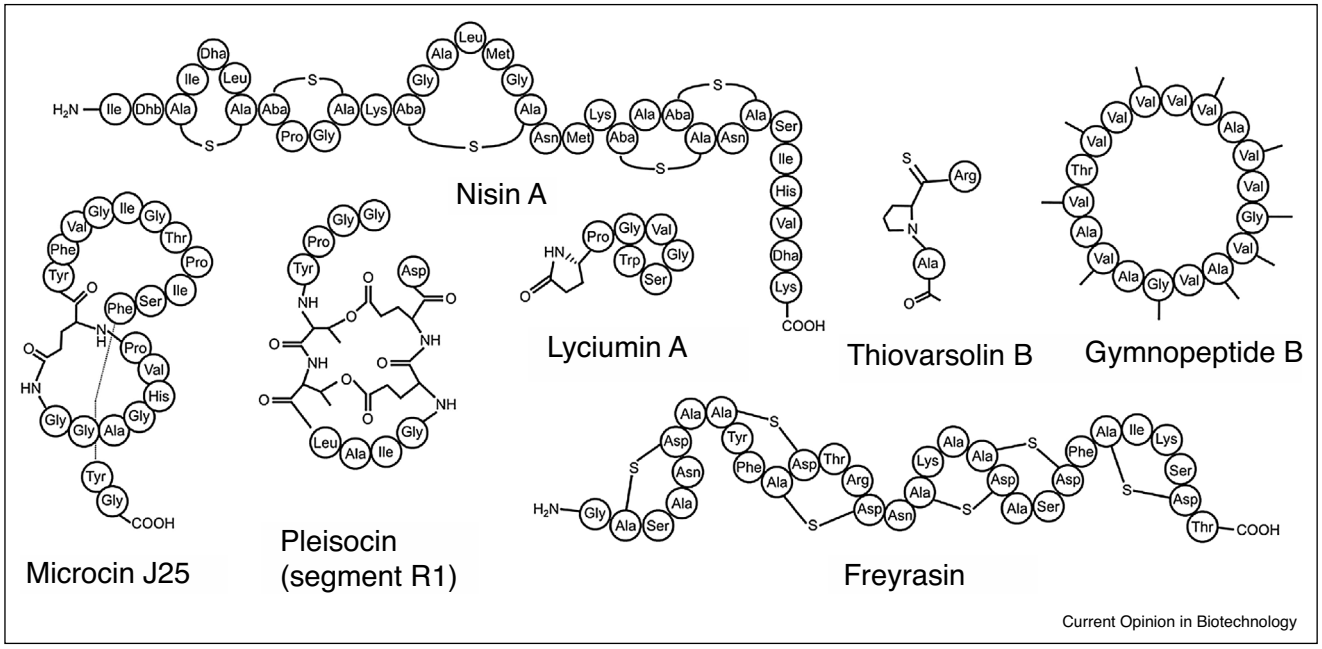
include quorum sensing, acting as enzyme co-factors, roles in cellular development, mediating host–microbe interactions, but also the much sought-after antibacterial and antifungal properties that would make them interesting for clinical applications [22].

RiPPs are classified into subclasses or families, each of which is defined by one or more characteristic modifications. For example, all lasso peptides have at least a single crosslink, forming a small loop through which the amino acid chain is threaded. Given the simple architecture of a RiPP BGC and the expected ease with which they might have evolved, it is likely that many more RiPP classes exist that do not contain the marker genes of known classes. This is exemplified by the rapid increase of known RiPP classes, which have nearly doubled between 2013 and 2020 [20,21••]. Discovery of novel RiPPs presents itself as a promising avenue to identify novel chemical scaffolds and drug leads. However, the lack of a single genetic marker among all RiPP classes requires novel strategies to leverage genomic data to this end. In this review, we discuss approaches and strategies for explorative RiPP genome mining aimed at the discovery of novel RiPP families (Table 1).

Taking the bait: how to prioritize regions of interest

Marker-based genome mining has been the main strategy for the high-confidence detection of BGCs from known classes of RiPPs. For many RiPP classes, modifying enzymes are split into core and accessory types, depending on the type of modification installed. RiPP BGCs are then detected by targeting the core modifying enzymes of one class, and many tools like antiSMASH [14], BAGEL [23], PRISM [24] and RODEO [25•,26–30] have implemented this strategy. The rules for each RiPP class result in the high-confidence detection of BGCs from known RiPP classes. For the detection of novel RiPP classes,

Figure 2



Examples of the rich chemical diversity of RiPPs.
RiPP precursors are highly diverse in sequence and can be modified in many ways. Several old and new examples are shown here: nisin A (lanthipeptide) [66], lyciumin A (lyciumin) [58], thiovarsolin B (thioamitide) [31••], gymnopeptide B (borosin) [58], microcin J25 (lasso peptide) [67], plesiocin segment R1 (omega-ester containing peptide/graspetide) [68] and freyrasin (ranthipeptide) [27]. Abu-S-Ala: beta-methyllanthionine. Dha: dehydroalanine. Dhb: dehydrobutyirine.

however, these rules need to be broken, even if this will result in more false positives.

A method exploring this principle, RiPPer, allows a user to find novel BGC architectures based on a single enzyme query [31••]. Candidate precursors are prioritized based on their conservation among the candidate BGCs of interest. The authors themselves use an enzyme involved in thioviridamide maturation to find the thiovarsolins, expanding the thioamitide class. Arguably, any enzyme can be used in such a strategy, as long as it is frequently associated with RiPP biosynthesis. For enzymes like YcaO and LanD, this has long been known to be the case [32,33]. The discovery of lanthinidins as intermediates between lanthipeptides and linaridins suggests that the exchange of domains between different RiPP classes may be more common than previously thought [34,35]. Another versatile enzyme is the radical *S*-adenosyl methionine (RaS) enzyme, involved in the maturation of RiPPs [36]. Exploration of RaS genes in novel contexts has recently led to the identification of new RiPP classes, including the spliceotides [37•], ryptides [38], rotapeptides [39] and WGK peptides [40•]. In rare cases, overlap in domains can even lead to the discovery of novel non-RiPP BGCs using RiPP modifying enzymes. LanB enzymes, vital for the maturation of type I lanthipeptides, have also been found in new genetic contexts, which led to the identification of the pearlins [41].

A key identifier for many different RiPP families is the RiPP Recognition Element (RRE). RREs can be encoded in small peptides or fused to modifying enzymes, and are required for the recognition of the precursor peptide. Although their secondary structure is conserved, RREs have highly diverse sequences, and could until recently only be effectively identified using HHPred [42,43], a highly sensitive but compute-intensive algorithm that compares families of sequences and their secondary structures to each other. RRE detection can be sped up using HMMer instead of HHPred, although novel RiPP classes will then only be detected if their RREs are similar on the sequence level. To detect an RRE-like secondary structure, HHPred can be used with a smaller database aimed specifically at RREs, rather than the large uniclust database which is used by default [44]. Both methods have been implemented in RRE-Finder [45•]. With this tool, several RREs were detected fused to enzymes not known to be involved in RiPP biosynthesis. RREs are not a high-confidence marker, as they are not contained in all RiPPs [46], and sometimes the domain is vestigial

Table 1

Tools currently available for RiPP genome mining

Name	BGC identification target	Method description	Possibilities of identifying novel classes	Reference
antiSMASH	Core enzymes	Identifies RiPP BGCs with core enzymes per class. Identifies precursor peptides with RODEO's SVMs.	Focuses on known classes. Can identify novel classes if they share core enzymes.	Blin *et al.* [14]
BAGEL	Core enzymes	Identifies RiPP BGCs with core enzymes per class. Identifies precursor peptides with BLAST and a known precursor database.	Focuses on known classes. Can identify novel classes if they share core enzymes.	Van Heel *et al.* [23]
RiPP-PRISM	Core enzymes	Identifies RiPP BGCs with core enzymes per class. Identifies precursor peptides with HMMer and a motif search.	Focuses on known classes. Can identify novel classes if they share core enzymes.	Skinnider *et al.* [15,24]
RODEO	Core enzymes	Identifies RiPP BGCs with core enzymes per class. Identification of precursor peptides with SVMs.	Focuses on known classes. Can identify novel classes if they share core enzymes (e.g. ranthipeptides). Custom queries possible.	Tietz *et al.* [29], Schwalen *et al.* [28], Hudson *et al.* [27], DiCaprio *et al.* [30], Walker *et al.* [25•], Georgiou *et al.* [26]
RiPPer	Any enzyme	Identifies RiPP BGCs with any query enzyme. Prioritizes candidate precursor peptides with prodigal-short and BLAST-based clustering.	Can identify novel classes if a query is selected that is associated with RiPP biosynthesis in a class-independent manner, with a higher chance of false positives depending on the query used.	Santos-Aberturas *et al.* [31••]
RRE-Finder	RiPP Recognition Elements (RREs)	Identifies RREs with HMMer or HHPred-like pipeline. Can identify novel RRE-fusions, depending on the cutoffs used.	Can identify novel RRE-fusions with exploratory mode or with precision mode and lower cutoffs, at the cost of more false positives.	Kloosterman *et al.* [45•]
RiPPMiner	Precursor peptides	Identifies and classifies precursors with a single SVM.	Can detect precursors of novel classes if they are similar from the perspective of the classifier.	Agrawal *et al.* [53]
NeuRiPP	Precursor peptides	Identifies precursors with a neural network.	Can detect precursors of novel classes if they are similar from the perspective of the classifier.	De Los Santos [55]
DeepRiPP	Precursor peptides	Identifies and classifies precursors and BGCs with a neural network (NLPPrecursor). Predicts products and estimates novelty based on genetic context and known modifications (BARLEY). Compares metabolomics and matches MS/MS spectra to predicted products (CLAMS).	Focuses on known classes. Can detect precursors of novel classes if they are similar from the perspective of the precursor classifier.	Merwin *et al.* [56••]
decRiPPter	Precursor peptides	Identifies and classifies precursors with a single SVM. Uses genetic context to prioritize novel RiPP BGCs. Forms candidate RiPP families based on precursor and BGC similarities.	Can detect precursors of novel classes if they are similar from the perspective of the classifier. Comparative genomics is used to prioritize hits.	Kloosterman *et al.* [54••]
DEREPLICATOR	NA	Clusters peptide natural products based on MS/MS spectra.	Focuses on finding spectra of known peptidic natural products. Can find novel RiPPs if part of the RiPP can be related to previously identified products.	Mohimani *et al.* [62]
VarQuest	NA	Matches peptide natural products to their variants with unknown modifications based on MS/MS spectra.	Focuses on finding spectra related to known peptidic natural products. Can find novel RiPPs if part of the RiPP can be related to previously identified products. More flexible with regards to unknown modifications than DEREPLICATOR.	Gurevich *et al.* [64]
MetaMiner	Core enzymes	Identifies RiPP BGCs with antiSMASH. Predicts products based on genetic context and known modifications. Matches predicted products to MS/MS spectra.	Can identify RiPP subclasses with novel modifications using a blind modification search.	Cao *et al.* [65•]

[47]. Still, these results show that RREs can serve as excellent beacons leading to the discovery of novel RiPP classes.

Prioritizing BGCs of interest without the use of a query domain is much more challenging, but has the potential to identify completely novel machinery. These approaches are particularly important for eukaryotic BGCs, for which few BGC markers are known [48,49]. Bacteria, fungi, plants and other eukaryotes each produce their own unique RiPP subclasses. The discovery of RiPPs across these branches of life is therefore mostly independent from one another, from a bioinformatic point of view. Most of the subclasses have been found in bacteria, although increasingly more subclasses are being uncovered for fungal and plant RiPPs, and likely many more exist with still unknown modifying enzymes [49]. The bioinformatics discovery of these could rely on query-independent strategies. The conserved genomic location and co-regulation of BGCs, for example, has been exploited to prioritize regions of interest in fungi, resulting in the discovery of a novel RiPP [19,50,51]. Similar methods can be exploited for other eukaryotes as well, when sequencing of many of their large genomes becomes viable. ClusterFinder [17], and more recently, DeepBGC [16], both detect BGCs without being confined to specific domains. Such methods have not been developed and trained with a specific focus on RiPP BGCs, but could be highly valuable when combined with precursor detection (see below).

Finding the needle: new methods for the detection of precursor peptides

The first step in RiPP biosynthesis is the translation of the precursor gene. Since these genes are small, they are notoriously hard to detect by gene-finding algorithms. Often, the precursor genes are found nearby their modifying enzymes, which limits the search space. BAGEL4, for example, identifies precursor peptides by BLASTing all small ORFs against a large database of previously characterized precursors peptides [23]. Precursor gene detection is even more important when they are not encoded near the genes encoding their modifiying enzymes, such as for animal RiPPs and cyanobactins [49,52].

Machine-learning based classification of precursor peptides is quickly gaining traction as a viable way for their detection with high accuracies and low false discovery rates (FDR). Support Vector Machines (SVMs) are especially popular. Rather than the peptide sequence itself, SVMs use features calculated from the sequence, such as charge, hydrophobicity, or abundance of amino acids or amino acid pairs.

Depending on the features selected, a wealth of information can be extracted from the peptide sequence, to precisely separate precursors and non-precursors. This method is used by RODEO [25•,26–30], which uses a different SVM per RiPP class to detect precursors in detected BGCs. RiPP-MINER [53] and decRiPPter [54••] are available as standalone tools and use a single SVM to identify precursor peptides regardless of class.

Neural networks bypass the need for feature selection by taking in the raw sequence of the peptides as a vector. Two tools have been developed that use neural networks to identify and classify precursors: NeuRiPP [55] and NLPPrecursor [56••]. The tools use different network architectures: NeuRiPP's most successful architecture is the parallel convoluted neural network (CNN), while NLPPrecursors uses a Universal Language Model Fine-Tuning (ULMFiT) neural network to detect encoded precursor peptides. The latter is a neural network architecture used for language processing that has shown to be highly effective in building models from training sets with low amount of data.

Both neural networks, as well as the SVMs from RiPP-MINER and decRiPPter, can detect precursors of many different classes using only a single model. Interestingly, both NeuRiPP and decRiPPter can identify RiPP precursors that are not included in the training data, suggesting that they should allow the detection of precursors of currently undiscovered families of RiPPs. Apparently, some properties are common to RiPP precursors regardless of their class. These are not directly obvious from their sequence, but are still picked up by the classifiers. How suitable precursor classifiers are for detection of novel RiPP classes likely depends on the selected features and model architecture. A standardized test with curated databases could be a valuable addition to benchmark and compare classifiers. By leaving out RiPP classes during the training process and testing how well they are still detected, the approximate 'explorativeness' of a classifier can be measured. This process could also give valuable insights into how RiPP precursors from different classes relate to one another, from the perspective of the classifier and the features it is trained on. Some classes of RiPP precursors may be more similar than others, which would mean that some classes are more easily discovered with our current training data than others.

Another interesting feature that has been reported in increasingly more studies is the presence of multiple core regions in a single precursor peptide. The encoding of multiple copies of the same core region allows for the efficient production of several RiPP variants, while only needing a single leader peptide. These repeats are found often in eukaryotic RiPPs [49,57,58], and could provide a handhold for their identification without prior knowledge of their primary sequence.

Assembling the whole: the integration of –omics

Exploiting the capabilities of a precursor classifier to detect novel RiPP classes is a promising route that has only been partially explored. However, any precursor-based approach faces a difficult challenge with regard to the ratio of false positives to true positives, as the number of small ORFs far exceeds the number of expected RiPP precursors. Integration of these tools into larger pipelines and combination of –omics datasets could help these tools reach their potential.

Purely genomic approaches can help prioritize regions of interest around predicted precursors. Any of the methods mentioned above could be combined, such as the requirement of a novel RRE-enzyme fusion and a predicted precursor. decRiPPter builds on its precursor predictions to prioritize regions of interest with a marker-independent strategy [54••]. Candidate RiPP BGCs are prioritized by filters, such as requiring the presence of a transporter gene or the lack of household genes that are common to the taxonomic group of organisms studied. Candidate RiPP families are formed by clustering identified BGCs based on their precursor sequences and encoded enzymatic domains, which resulted in the identification of 42 new candidate RiPP BGC families across 1295 *Streptomyces* genomes.

Transcriptomic and proteomic data can be used to identify co-regulated genes and link them to their cognate bioactivity [59,60]. This principle was previously used to identify ustiloxin B [51]. More recently, a pipeline that integrates the use of RNASeq data was used to identify novel RiPP BGCs in the fungus *Trichoderma* spp. [61]. The authors used ClusterFinder [17] to identify candidate BGCs, in which candidate BGCs were identified with RiPPMINER [53]. The results were further filtered by removing gene islands which were not activated, and clustering the predicted precursors, resulting in the prediction of several candidates.

The integration of metabolomics data could accelerate up the identification of novel classes. There is a large potential for the identification of RiPP-like compounds by automated detection from spectral data. In contrast to normal proteins, however, RiPPs contain modified amino acids and are rarely linear. For known RiPP classes, the modifications can be predicted based on genomic information. Predicted peptide fragments containing these modifications can be matched to the spectra with tools like DEREPLICATOR [62] (recently updated with NPS [63]) and CLAMS (available within the DeepRiPP pipeline [56••]). DeepRiPP is perhaps the most integrative pipeline for RiPP discovery. Besides structure prediction based on the identification of known modifications, it also combines comparative genomics with comparative metabolomics, to prioritize peaks whose presence/absence matches that of the BGCs of interest. While DeepRiPP mostly prioritizes RiPPs of known classes, a similar pipeline could be conceptualized aimed at the discovery of novel classes. All of this will depend on whether compounds are expressed at sufficiently high levels to facilitate their detection. Elicitors should therefore be added to activate the expression of cryptic BGCs, whereby comparative metabolomics combined with transcriptomics or proteomics will allow linkage of BGC expression profiles to changes in metabolites. This will allow scientists not only to observe more metabolites than under one specific growth condition, but also to predict which metabolites are produced by which BGCs.

A major challenge for automated MS/MS analysis that remains is dealing with new modifications. VarQuest [64], an extension of DEREPLICATOR, can identify peptide variants based on known peptides, even if these variants contain unknown modifications. MetaMiner [65•] combines genomics and metabolomics to predict precursor modifications and find associated spectra, which can contain unknown modifications. Completely *de novo* identification of novel RiPPs with only unknown modification has yet to be explored by tools like these, but represents a sizable computational challenge. Even so, just matching a small sequence of unmodified amino acids to part of a candidate novel RiPP precursor is a valuable addition to more explorative RiPP searches. Identified, novel precursors could then be fed back to the training data of the precursor classifiers, creating an iterative process in which the classifiers will become increasingly specific and tuned toward a larger variety of RiPP classes.

Conclusions and final perspectives

RiPPs represent a diverse class of natural products within which new subclasses are being quickly discovered. The lack of universal genetic markers makes genome-based mining for novel RiPP BGCs challenging, but the amount of sequence data and the level of computational power available allows for many highly interesting strategies. As more RiPP classes are being discovered, more modifications are found to be shared between different classes, which can lead the way to novel variants. Precursor classification is a powerful addition to the list of available tools, both for known and novel RiPP identification. Finally, integrative approaches combining comparative genomics, eliciting strategies, transcriptomics, proteomics and metabolomics will help us explore the vast and diverse chemical space of this promising class of natural products.

Conflict of interest statement

M.H.M. is on the scientific advisory board of Hexagon Bio and co-founder of Design Pharmaceuticals.

Acknowledgement

a.m.K. was supported by the Netherlands Organization for Scientific Research (NWO, grant 731.014.206 to G.P.vW.).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Berdy J: **Bioactive microbial metabolites**. *J Antibiot (Tokyo)* 2005, **58**:1-26.

2. O'Neill J: *Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations*. UK: The Review on Antimicrobial Resistance; 2014.

3. WHO: *Antimicrobial Resistance: Global Report on Surveillance*. 2014. Geneva, Switzerland.

4. Kolter R, van Wezel GP: **Goodbye to brute force in antibiotic discovery?** *Nat Microbiol* 2016, **1**:15020.

5. Silver LL: **Challenges of antibacterial discovery**. *Clin Microbiol Rev* 2011, **24**:71-109.

6. Nett M, Ikeda H, Moore BS: **Genomic basis for natural product biosynthetic diversity in the actinomycetes**. *Nat Prod Rep* 2009, **26**:1362-1384.

7. Blin K *et al.*: **Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters**. *Brief Bioinform* 2019, **20**:1103-1113.

8. van Santen JA *et al.*: **Microbial natural product databases: moving forward in the multi-omics era**. *Nat Prod Rep* 2020.

9. Kautsar SA *et al.*: **MIBiG 2.0: a repository for biosynthetic gene clusters of known function**. *Nucleic Acids Res* 2020, **48**:D454-D458.

10. Kautsar SA *et al.*: **BiG-FAM: the biosynthetic gene cluster families database**. *Nucleic Acids Res* 2020.

11. Navarro-Munoz JC *et al.*: **A computational framework to explore large-scale biosynthetic diversity**. *Nat Chem Biol* 2020, **16**:60-68.

12. Rutledge PJ, Challis GL: **Discovery of microbial natural products by activation of silent biosynthetic gene clusters**. *Nat Rev Microbiol* 2015, **13**:509-523.

13. van Bergeijk DA *et al.*: **Ecology and genomics of actinobacteria: new concepts for natural product discovery**. *Nat Rev Microbiol* 2020, **18**:546-558.

14. Blin K *et al.*: **antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline**. *Nucleic Acids Res* 2019, **47**: W81-W87.

15. Skinnider MA *et al.*: **PRISM 3: expanded prediction of natural product chemical structures from microbial genomes**. *Nucleic Acids Res* 2017, **45**:W49-W54.

16. Hannigan GD *et al.*: **A deep learning genome-mining strategy for biosynthetic gene cluster prediction**. *Nucleic Acids Res* 2019, **47**:e110.

17. Cimermancic P *et al.*: **Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters**. *Cell* 2014, **158**:412-421.

18. Selem-Mojica N *et al.*: **EvoMining reveals the origin and fate of natural product biosynthetic enzymes**. *Microb Genom* 2019, **5**.

19. Umemura M, Koike H, Machida M: **Motif-independent de novo detection of secondary metabolite gene clusters-toward identification from filamentous fungi**. *Front Microbiol* 2015, **6**:371.

20. Arnison PG *et al.*: **Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature**. *Nat Prod Rep* 2013, **30**:108-160.

21. •• Montalbán-López M: *New Developments in RiPP Discovery, Enzymology and Engineering*. 2020
An updated comprehensive overview of all currently known RiPP classes and recent advances in RiPP genome mining, enzymology and engineering.

22. Li Y, Rebuffat S: **The manifold roles of microbial ribosomal peptide-based natural products in physiology and ecology**. *J Biol Chem* 2020, **295**:34-54.

23. van Heel AJ *et al.*: **BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins**. *Nucleic Acids Res* 2018, **46**:W278-W281.

24. Skinnider MA *et al.*: **Genomic charting of ribosomally synthesized natural product chemical space facilitates targeted mining**. *Proc Natl Acad Sci U S A* 2016, **113**:E6343-E6351.

25. • Walker MC *et al.*: **Precursor peptide-targeted mining of more than one hundred thousand genomes expands the lanthipeptide natural product family**. *BMC Genomics* 2020, **21**:387
Large-scale genome mining of one of the best-studied RiPP classes expands it even further. Many BGCs with novel accessory modifying enzymes and their precursors are identified through the combination of SVMs with HMMer.

26. Georgiou MA *et al.*: **Bioinformatic and reactivity-based discovery of linaridins**. *ACS Chem Biol* 2020, **15**:2976-2985 2020.07.09.196543.

27. Hudson GA *et al.*: **Bioinformatic mapping of radical S-adenosylmethionine-dependent ribosomally synthesized and post-translationally modified peptides identifies new calpha, cbeta, and cgamma-linked thioether-containing peptides**. *J Am Chem Soc* 2019, **141**:8228-8238.

28. Schwalen CJ *et al.*: **Bioinformatic expansion and discovery of thiopeptide antibiotics**. *J Am Chem Soc* 2018, **140**:9494-9501.

29. Tietz JI *et al.*: **A new genome-mining tool redefines the lasso peptide biosynthetic landscape**. *Nat Chem Biol* 2017, **13**:470-478.

30. DiCaprio AJ *et al.*: **Enzymatic reconstitution and biosynthetic investigation of the lasso peptide fusilassin**. *J Am Chem Soc* 2019, **141**:290-297.

31. •• Santos-Aberturas J *et al.*: **Uncovering the unexplored diversity of thioamidated ribosomal peptides in actinobacteria using the RiPPER genome mining tool**. *Nucleic Acids Res* 2019, **47**:4624-4637
A tool for the discovery of novel RiPP classes based on a single query enzyme, allowing for user-friendly explorative RiPP genome mining. The tool was used to discover the thiovarsolins.

32. Burkhart BJ *et al.*: **YcaO-dependent posttranslational amide activation: biosynthesis, structure, and function**. *Chem Rev* 2017, **117**:5389-5456.

33. Sit CS, Yoganathan S, Vederas JC: **Biosynthesis of aminovinyl-cysteine-containing peptides and its application in the production of potential drug candidates**. *Acc Chem Res* 2011, **44**:261-268.

34. Ortiz-Lopez FJ *et al.*: **Cacaoidin, first member of the new lanthidin RiPP family**. *Angew Chem Int Ed Engl* 2020, **59**:12654-12658.

35. Xu M *et al.*: **Functional genome mining reveals a class V lanthipeptide containing a d-amino acid introduced by an F420 H2-dependent reductase**. *Angew Chem Int Ed Engl* 2020, **59**:18029-18035.

36. Benjdia A, Balty C, Berteau O: **Radical SAM enzymes in the biosynthesis of ribosomally synthesized and post-translationally modified peptides (RiPPs)**. *Front Chem* 2017, **5**:87.

37. • Morinaka BI *et al.*: **Natural noncanonical protein splicing yields products with diverse beta-amino acid residues**. *Science* 2018, **359**:779-782

A new RiPP is characterized with highly unusual beta-amino acids, called the spliceotides. The use of this modification as a precursor for drugs is also explored.

38. Caruso A *et al.*: **Macrocyclization via an arginine-tyrosine crosslink broadens the reaction scope of radical S-adenosylmethionine enzymes**. *J Am Chem Soc* 2019, **141**:16610-16614.

39. Clark KA, Bushin LB, Seyedsayamdost MR: **Aliphatic ether bond formation expands the scope of radical SAM enzymes in natural product biosynthesis**. *J Am Chem Soc* 2019, **141**:10610-10615.

40. • Bushin LB *et al.*: **Charting an unexplored streptococcal biosynthetic landscape reveals a unique peptide cyclization motif**. *J Am Chem Soc* 2018, **140**:17674-17684
A BGC organization consisting of a radical SAM and quorum sensing regulators is found throughout streptococcal genomes, leading to the discovery of several new RiPP classes.

41. Ting CP *et al.*: **Use of a scaffold peptide in the biosynthesis of amino acid-derived natural products**. *Science* 2019, **365**:280-284.

42. Burkhart BJ *et al.*: **A prevalent peptide-binding domain guides ribosomal natural product biosynthesis**. *Nat Chem Biol* 2015, **11**:564-570.

43. Soding J, Biegert A, Lupas AN: **The HHpred interactive server for protein homology detection and structure prediction**. *Nucleic Acids Res* 2005, **33**:W244-W248 (Web Server issue).

44. Mirdita M *et al.*: **Uniclust databases of clustered and deeply annotated protein sequences and alignments**. *Nucleic Acids Res* 2017, **45**:D170-D176.

45. • Kloosterman AM *et al.*: **RRE-finder: a genome-mining tool for class-independent RiPP discovery**. *mSystems* 2020, **5**
A tool for the discovery of RiPP Recognition Elements, class-independent RiPP markers that have the potential to be used for the discovery of novel RiPP classes.

46. Rahman IR *et al.*: **Substrate recognition by the class II lanthipeptide synthetase HalM2**. *ACS Chem Biol* 2020, **15**:1473-1486.

47. Zhang Z *et al.*: **Biosynthetic timing and substrate specificity for the thiopeptide thiomuracin**. *J Am Chem Soc* 2016, **138**:15511-15514.

48. van der Lee TAJ, Medema MH: **Computational strategies for genome-based natural product discovery and engineering in fungi**. *Fungal Genet Biol* 2016, **89**:29-36.

49. Luo S, Dong SH: **Recent advances in the discovery and biosynthetic study of eukaryotic RiPP natural products**. *Molecules* 2019, **24**.

50. Takeda I *et al.*: **Motif-independent prediction of a secondary metabolism gene cluster using comparative genomics: application to sequenced genomes of *Aspergillus* and ten other filamentous fungal species**. *DNA Res* 2014, **21**:447-457.

51. Umemura M *et al.*: **MIDDAS-M: motif-independent de novo detection of secondary metabolite gene clusters through the integration of genome sequencing and transcriptome data**. *PLoS One* 2013, **8**:e84028.

52. Gu W *et al.*: **The biochemistry and structural biology of cyanobactin pathways: enabling combinatorial biosynthesis**. *Methods Enzymol* 2018, **604**:113-163.

53. Agrawal P *et al.*: **RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links**. *Nucleic Acids Res* 2017, **45**:W80-W88.

54. •• Kloosterman AM *et al.*: **Integration of machine learning and pan-genomics expands the biosynthetic landscape of RiPP natural products**. *bioRxiv* 2020 http://dx.doi.org/10.1101/2020.05.19.104752
A genomics-based pipeline based on artifical intelligence that prioritizes BGCs of potentially novel RiPP classes. The tool was used to identify 42 candidate RiPP families, one of which was experimentally characterized as a new class of lanthipeptides.

55. de Los Santos ELC: **NeuRiPP: neural network identification of RiPP precursor peptides**. *Sci Rep* 2019, **9**:13406.

56. •• Merwin NJ *et al.*: **DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products**. *Proc Natl Acad Sci U S A* 2020, **117**:371-380
A comprehensive RiPP discovery pipeline combining precursor detection with genomic context interpretation and comparative metabolomics and MS/MS spectra matching. The tool was used to identify three new RiPPs of known classes.

57. Quijano MR *et al.*: **Distinct autocatalytic alpha-N-methylating precursors expand the borosin RiPP family of peptide natural products**. *J Am Chem Soc* 2019, **141**:9637-9644.

58. Kersten RD, Weng JK: **Gene-guided discovery and engineering of branched cyclic peptides in plants**. *Proc Natl Acad Sci U S A* 2018, **115**:E10961-E10969.

59. Du C, van Wezel GP: **Mining for microbial gems: integrating proteomics in the postgenomic natural product discovery pipeline**. *Proteomics* 2018, **18**:e1700332.

60. Gubbens J *et al.*: **Natural product proteomining, a quantitative proteomics platform, allows rapid discovery of biosynthetic gene clusters for different classes of natural products**. *Chem Biol* 2014, **21**:707-718.

61. Vignolle GA *et al.*: **Novel approach in whole genome mining and transcriptome analysis reveal conserved RiPPs in *Trichoderma* spp**. *BMC Genomics* 2020, **21**:258.

62. Mohimani H *et al.*: **Dereplication of peptidic natural products through database search of mass spectra**. *Nat Chem Biol* 2017, **13**:30-37.

63. Tagirdzhanov AM, Shlemov A, Gurevich A: **NPS: scoring and evaluating the statistical significance of peptidic natural product-spectrum matches**. *Bioinformatics* 2019, **35**:i315-i323.

64. Gurevich A *et al.*: **Increased diversity of peptidic natural products revealed by modification-tolerant database search of mass spectra**. *Nat Microbiol* 2018, **3**:319-327.

65. • Cao L *et al.*: **MetaMiner: a scalable peptidogenomics approach for discovery of ribosomal peptide natural products with blind modifications from microbial communities**. *Cell Syst* 2019, **9**:600-608 e4
A tool integrating the use of genomics and product prediction with peptide-spectra matching identifies 31 known and 7 novel RiPPs.

66. Berridge NJ, Newton GG, Abraham EP: **Purification and nature of the antibiotic nisin**. *Biochem J* 1952, **52**:529-535.

67. Salomon RA, Farias RN: **Microcin 25, a novel antimicrobial peptide produced by *Escherichia coli***. *J Bacteriol* 1992, **174**:7428-7435.

68. Lee H, Park Y, Kim S: **Enzymatic cross-linking of side chains generates a modified peptide with four hairpin-like bicyclic repeats**. *Biochemistry* 2017, **56**:4927-4930.