W46–W50 *Nucleic Acids Research, 2023, Vol. 51, Web Server issue* *Published online 4 May 2023*
https://doi.org/10.1093/nar/gkad344

antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation

Kai Blin[1,*], Simon Shaw[1], Hannah E. Augustijn[2,3], Zachary L. Reitz[3], Friederike Biermann[3,4,5], Mohammad Alanjary[3], Artem Fetter[3,6], Barbara R. Terlouw[3], William W. Metcalf[7,8], Eric J.N. Helfrich[4,5], Gilles P. van Wezel[2], Marnix H. Medema[3,*] and Tilmann Weber[1,*]

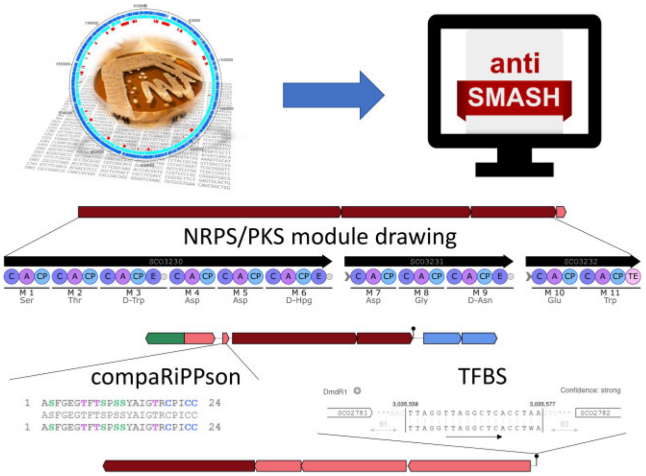
[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs.Lyngby, Denmark, [2]Molecular Biotechnology, Institute of Biology, Leiden University, Leiden, The Netherlands, [3]Bioinformatics Group, Wageningen University, Wageningen, The Netherlands, [4]Institute of Molecular Bio Science, Goethe-University Frankfurt, Frankfurt am Main, Germany, [5]LOEWE Center for Translational Biodiversity Genomics. Frankfurt am Main, Germany, [6]Institute of Technical Chemistry, Leibniz University Hannover, Hannover, Germany, [7]Department of Microbiology, University of Illinois Urbana–Champaign, Urbana, IL, USA and [8]Institute for Genomic Biology, University of Illinois Urbana–Champaign, Urbana, IL, USA

Received February 24, 2023; Revised March 31, 2023; Editorial Decision April 17, 2023; Accepted April 26, 2023

ABSTRACT

Microorganisms produce small bioactive compounds as part of their secondary or specialised metabolism. Often, such metabolites have antimicrobial, anticancer, antifungal, antiviral or other bio-activities and thus play an important role for applications in medicine and agriculture. In the past decade, genome mining has become a widely-used method to explore, access, and analyse the available biodiversity of these compounds. Since 2011, the 'antibiotics and secondary metabolite analysis shell—antiSMASH' (https://antismash.secondarymetabolites.org/) has supported researchers in their microbial genome mining tasks, both as a free to use web server and as a standalone tool under an OSI-approved open source licence. It is currently the most widely used tool for detecting and characterising biosynthetic gene clusters (BGCs) in archaea, bacteria, and fungi. Here, we present the updated version 7 of antiSMASH. antiSMASH 7 increases the number of supported cluster types from 71 to 81, as well as containing improvements in the areas of chemical structure prediction, enzymatic assembly-line visualisation and gene cluster regulation.

GRAPHICAL ABSTRACT



INTRODUCTION

Small bioactive compounds produced by microorganisms form the basis of many drugs (1) and crop protection agents (2). Traditionally, new compounds were discovered using a 'find and grind' workflow of extracting from natural sources, chemically isolating, purifying, and then testing compounds. This approach is now routinely complemented by sequencing and subsequent mining of genome and metagenome data to identify natural product biosyn-

[*]To whom correspondence should be addressed. Tel: +45 93511306; Email: kblin@biosustain.dtu.dk
Correspondence may also be addressed to Marnix H. Medema. Tel: +31 317482036; Email: marnix.medema@wur.nl
Correspondence may also be addressed to Tilmann Weber. Tel: +45 24896132; Email: tiwe@biosustain.dtu.dk

thetic pathways (3). Software tools for 'genome mining', i.e. searching genomes for secondary/specialised metabolite (SM) biosynthetic gene clusters (BGCs), have existed for over a decade (4–7).

Since its 2011 release, antiSMASH (8–13) has established itself as the most widely used tool for mining microbial genomes for SM BGCs. Around antiSMASH, an ecosystem of independent tools that incorporate or utilise antiSMASH results has developed, such as the antibiotics resistance target seeker (ARTS 2) (14), the mass-spectrometry-guided peptide mining tool Pep2Path (15), the sgRNA design tool CRISPY-web 2 (16), the BGC networking and clustering platform BiG-SCAPE (17), and the related big data BGC clustering tool BiG-SLiCE (18). In turn, antiSMASH can also incorporate and display BGC predictions from other tools such as DeepBGC (19) by using the side-loading mechanism introduced in antiSMASH 6 (13). antiSMASH BGC predictions are included in many genomic and BGC-oriented databases, like the Joint Genome Institute's Integrated Microbial Genomes database with its Atlas of Biosynthetic gene Clusters IMG-ABC (20), the MicroScope platform for microbial genome annotation and analysis (21), the MIBiG database of manually curated BGCs (22), the BGC family database BiG-FAM (23) and the antiSMASH database (24).

antiSMASH uses a rule-based approach to identify many different types of biosynthetic pathways involved in SM production. More in-depth analyses are performed for BGCs encoding non-ribosomal peptide synthetases (NRPSs), type I and type II polyketide synthases (PKSs), and the ribosomally synthesised and post-translationally modified peptide (RiPP) classes of lanthipeptides, lasso peptides, sactipeptides, and thiopeptides. For these, cluster-specific analyses can provide more information about the biosynthetic steps performed and thus also provide more detailed predictions on the compound(s) produced.

Here, we present version 7 of antiSMASH. It improves upon and further extends previous versions by adding and updating BGC detection rules, enhancing regulatory function detection by predicting transcription factor binding sites represented in the LogoMotif database (https://logomotif.bioinformatics.nl/), and adding new visualisations for NRPS and PKS clusters, PFAM (25) and TIGRFAM (26) domain hits, as well as tables listing all genes in a region with dynamic search and filter functions.

NEW FEATURES AND UPDATES

New cluster types and dynamic detection profiles

antiSMASH uses manually curated and validated 'rules' that define which core biosynthetic functions need to exist in a genomic region in order to constitute a BGC. To identify these biosynthetic functions, antiSMASH uses profile hidden Markov models (pHMMs) from PFAM (25), TIGRFAMs (26), SMART (27), BAGEL (28), Yadav *et al.* (29) and custom models. antiSMASH 6 contained rules for 71 BGC types (13). In antiSMASH 7, this number increases to 81, adding support for 2-deoxy-streptamine aminoglycosides, aminopolycarboxylic acid metallophores, arginine-containing cyclo-dipeptides (RCDPs), crocagins, methanobactins, mycosporines, NRP-metallophores (30), opine-like metallophores, and fungal-RiPP-likes. NRP metallophore BGCs were previously detected by the general NRPS detection rules, but are now recognised specifically on the basis of genes encoding the biosynthesis of functional groups involved in metal chelation (30). The phosphonates rule was updated, with the old rule retained under the name of 'phosphonate-likes'. In addition to an improved phosphoenolpyruvate (PEP) mutase detection model, supporting models (Supplementary Tables S1-S2) are leveraged to reduce false positives and improve delineation of cluster boundaries (Supplementary Figure S1).

Because not all features of a BGC can be captured with pHMMs, antiSMASH 7 adds the option of creating dynamic profiles that are defined by Python code instead. This is currently being used to detect cyanobactin precursors based on the M.KKN[IL].P....PV.R motif as described in (31), a conserved sequence motif too small to be picked up in a pHMM reliably.

NRPS & PKS improvements

To improve PKS annotations in fungal gene clusters, we have added detection profiles for carnitine-AT (cAT), product template (PT) and thiocysteine/beta-eliminating lyase (SH) domains. The ketosynthase (KS) domains of bacterial trans-acyltransferase polyketide synthases (trans-AT PKSs) are now also annotated using the subtype-specific pHMMs of transATor (32). PKS KS domains and NRPS condensation (C) domains can be submitted to the recently released version 2 of the Natural Product Domain Seeker (NapDoS2) (33) for phylogenetic analysis. The recent MIBiG 3 release (22) adds substrate specificities for over 2000 NRPS adenylation (A) and related domains. To allow our users to benefit from the additional information, we have replaced the NRPSPredictor2 A domain substrate prediction tool (34) we have been shipping since 2011 with the new 'NRPyS' library (https://github.com/kblin/nrpys/) that allowed us to update the Stachelhaus code (35) lookup table from previously 554 to now 2319 entries. As the 10 amino acid (AA) code used by Stachelhaus does not always resolve to a single substrate prediction in the new data set, most likely due to substrate promiscuity of the A domain involved, NRPyS reports all equal quality 10 AA code hits, ranked by the highest match to the 34 AAs predicted to be in an 8 Å radius around the A domain active site following the description of Rausch et al. used in NRPSPredictor 1 (36). To act as a full drop-in replacement, NRPyS still runs the original support vector machine (SVM) models from NRPSPredictor 2.

RiPP precursor comparisons

To help users with evaluating the novelty of RiPP precursor peptides, we have developed the CompaRiPPson analysis that compares the (predicted) core peptides of identified RiPP precursors to RiPP precursors in the antiSMASH-DB (24) and MIBiG 3.1 (22) databases. Hits for these databases are presented separately, with the antiSMASH-DB hits containing a much larger dataset of 10583 predicted precursors in version 3.0 versus 28 experimentally verified and annotated precursors from MIBiG. Precursor hits are labelled

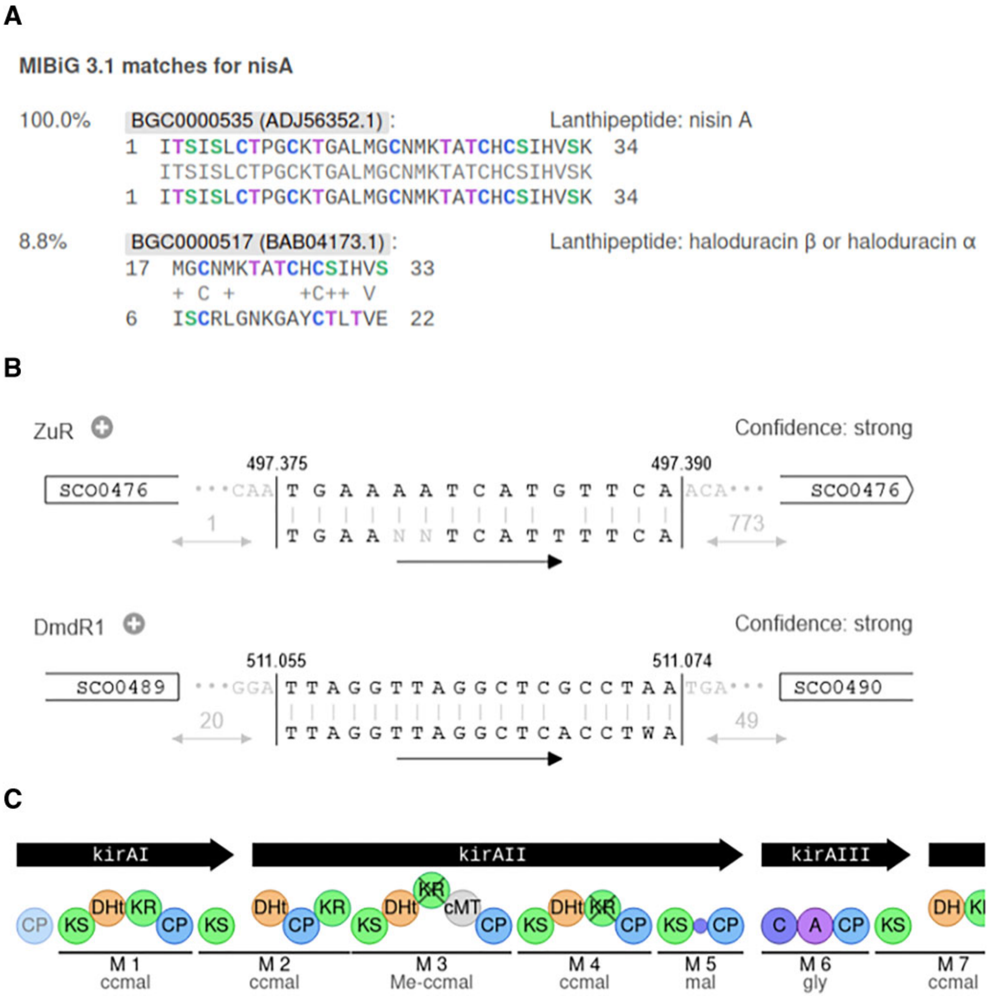

A
MIBiG 3.1 matches for nisA
100.0%
BGC0000535 (ADJ56352.1) :
Lanthipeptide: nisin A
1 ITSISLCTPGCKTGALMGCNMKTATCHCSIHVSK 34
ITSISLCTPGCKTGALMGCNMKTATCHCSIHVSK
1 ITSISLCTPGCKTGALMGCNMKTATCHCSIHVSK 34
8.8%
BGC0000517 (BAB04173.1) :
Lanthipeptide: haloduracin β or haloduracin α
17 MGCNMKTATCHCSIHVS 33
+ C + +C++ V
6 ISCRLGNKGAYCTLTVE 22
B
ZuR
Confidence: strong
497.375
497.390
SCO0476
•••CAA
T G A A A A T C A T G T T C A
ACA•••
SCO0476
1
T G A A N N T C A T T T T C A
773
DmdR1
Confidence: strong
511.055
511.074
SCO0489
•••GGA
T T A G G T T A G G C T C G C C T A A
TGA•••
SCO0490
20
T T A G G T T A G G C T C A C C T W A
49
C
kirAI
kirAII
kirAIII
CP
KS
DHt
KR
CP
KS
DHt
CP
KR
KS
DHt
KR
cMT
CP
KS
DHt
KR
CP
KS
CP
C
A
CP
KS
DH
KI
M 1
ccmal
M 2
ccmal
M 3
Me-ccmal
M 4
ccmal
M 5
mal
M 6
gly
M 7
ccmal

Figure 1. Examples of new antiSMASH visualisations. (**A**) shows the CompaRiPPSon MIBiG matches for the lanthipeptide class I nisin A input sequence with a 100% match for the self hit, and a much lower (8.8%) match with another lanthipeptide. (**B**) shows two high confidence TFBS-finder hits on *Streptomyces coelicolor* A3(2). The first hit, a putative ZurR binding site, is located right at the start of gene SCO0476, with the last two bases of the ATG start codon being the first two bases of the binding site. The DmdR1 hit is located between and upstream of both SCO0489 and SCO0490. (**C**) shows the first modules of the *Streptomyces colinus* Tü 365 hybrid trans-AT PKS/NRPS kirromycin gene cluster (MIBiG ID: BGC0001070).

by precursor gene locus tag for the antiSMASH-DB and labelled by compound name for MIBiG. Ordered by sequence identity, database hits with identical precursor sequences are grouped together. Query and hits are displayed with alignments (Figure 1A).

Transcription factor binding site predictions

The LogoMotif database (https://logomotif.bioinformatics.nl/) contains a curated collection of experimentally validated transcription factor binding site (TFBS) profiles and corresponding position weight matrices (PWMs), focusing on Actinobacteria. The antiSMASH TFBS-finder module uses these PWMs to annotate putative TFBSs. Depending on the hit score, TFBS-finder displays a confidence level of strong, medium, or weak, respectively. Binding sites are displayed in their genomic context, indicating the orientations and distances to surrounding genes (Figure 1B). All hits link to the LogoMotif website for more in-depth information about specific profiles.

Gene table

Every region now lists all contained gene features in a filterable, interactive table. Genes can be filtered by entering a search term in the search box (plain text and regex are both supported). Genes that match the filter will be shown in the region view and, if enabled, the view will automatically zoom to the selection. Information used for filtering currently includes the name of the gene, its biosynthetic type, and gene function annotations (e.g. smCOG hits).

Updated visualisations and other optimisations

A new visualisation for NRPS and PKS clusters draws enzymatic domains and modules in the predicted assembly order in conventional publication style, which allows researchers to use the antiSMASH vector graphics as a starting point for their publication-quality figures (Figure 1C). PFAM (25) and TIGRFAMs (26) domain hits in a region are now shown in a similar fashion to the existing NRPS/PKS domain visualisations.

Following the MIBiG 3.1 release (22), the KnownClusterBlast and ClusterCompare databases were updated.

CONCLUSIONS & FUTURE PERSPECTIVES

Genome mining for natural product BGCs with tools like antiSMASH forms a cornerstone of modern natural product discovery workflows. With the additions and updates presented in this manuscript, antiSMASH is being continuously updated to remain the go-to solution for microbial natural product genome mining. The Open Source antiSMASH software continues to contribute to the thriving ecosystem of computational tools in the natural products field. In addition to providing microbial natural product predictions directly, antiSMASH also serves as the technology platform for other tools, such as the plant natural products prediction tool plantiSMASH (37), the primary metabolism gene cluster prediction tool gutSMASH (38,39), and other tools currently in development. In future updates, we will continue our work on improving compound structure and subcomponent predictions, adding additional TFBS profiles for different taxa (e.g. fungal profiles from JASPAR (40)), as well as integrating with other tools in the ecosystem. We have also started providing a website to try out potential future antiSMASH features at https://experimentalsmash.secondarymetabolites.org/.

DATA AVAILABILITY

The bacterial and fungal versions of antiSMASH 7 can be freely accessed at https://antismash.secondarymetabolites.org and https://fungismash.secondarymetabolites.org, respectively.

The antiSMASH documentation is available at https://docs.antismash.secondarymetabolites.org/.

The antiSMASH source code is licensed under the GNU Affero General Public License (AGPL) v3.0. antiSMASH is also available via Docker. See the documentation website for details on how to download and install antiSMASH.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Novo Nordisk Foundation [NNF20CC0035580 to T.W., NNF16OC0021746 to T.W.]; Center for Microbial Secondary Metabolites (CeMiSt), Danish National Research Foundation [DNRF137 to T.W.]; ERC Starting Grant [948770-DECIPHER to M.H.M.]; Novel Antibacterial Compounds and Therapies Antagonising Resistance program (NACTAR) from the Dutch Research Council (NWO) [16440 to M.H.M.]; M.A is supported by the NWO Talent programme Veni science domain (VI.Veni.202.130). Funding for open access charge: Novo Nordisk Foundation.

Conflict of interest statement. M.H.M. is a co-founder of Design Pharmaceuticals and a member of the scientific advisory board of Hexagon Bio.

REFERENCES

1. Newman,D.J. and Cragg,G.M. (2020) Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.*, **83**, 770–803.
2. Sparks,T.C. and Bryant,R.J. (2022) Impact of natural products on discovery of, and innovation in, crop protection compounds. *Pest Manag. Sci.*, **78**, 399–408.
3. Ziemert,N., Alanjary,M. and Weber,T. (2016) The evolution of genome mining in microbes – a review. *Nat. Prod. Rep.*, **33**, 988–1005.
4. Weber,T. (2014) In silico tools for the analysis of antibiotic biosynthetic pathways. *Int. J. Med. Microbiol.*, **304**, 230–235.
5. Medema,M.H. and Fischbach,M.A. (2015) Computational approaches to natural product discovery. *Nat. Chem. Biol.*, **11**, 639–648.
6. Weber,T. and Kim,H.U. (2016) The secondary metabolite bioinformatics portal: computational tools to facilitate synthetic biology of secondary metabolite production. *Synth. Syst. Biotechnol.*, **1**, 69–79.
7. Blin,K., Kim,H.U., Medema,M.H. and Weber,T. (2019) Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. *Brief. Bioinform.*, **20**, 1103–1113.
8. Medema,M.H., Blin,K., Cimermancic,P., de Jager,V., Zakrzewski,P., Fischbach,M.A., Weber,T., Takano,E. and Breitling,R. (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res*, **39**, W339–W346.
9. Blin,K., Medema,M.H., Kazempour,D., Fischbach,M.A., Breitling,R., Takano,E. and Weber,T. (2013) antiSMASH 2.0—A versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res*, **41**, W204–W212.
10. Weber,T., Blin,K., Duddela,S., Krug,D., Kim,H.U., Bruccoleri,R., Lee,S.Y., Fischbach,M.A., Müller,R., Wohlleben,W. *et al.* (2015) antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res*, **43**, W237–W243.
11. Blin,K., Wolf,T., Chevrette,M.G., Lu,X., Schwalen,C.J., Kautsar,S.A., Suarez Duran,H.G., de los Santos,E.L.C., Kim,H.U., Nave,M. *et al.* (2017) antiSMASH 4.0—Improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res*, **45**, W36–W41.
12. Blin,K., Shaw,S., Steinke,K., Villebro,R., Ziemert,N., Lee,S.Y., Medema,M.H. and Weber,T. (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res*, **47**, W81–W87.
13. Blin,K., Shaw,S., Kloosterman,A.M., Charlop-Powers,Z., van Wezel,G.P., Medema,M.H. and Weber,T. (2021) antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res*, **49**, W29–W35.
14. Mungan,M.D., Alanjary,M., Blin,K., Weber,T., Medema,M.H. and Ziemert,N. (2020) ARTS 2.0: feature updates and expansion of the Antibiotic Resistant Target Seeker for comparative genome mining. *Nucleic Acids Res*, **48**, W546–W552.
15. Medema,M.H., Paalvast,Y., Nguyen,D.D., Melnik,A., Dorrestein,P.C., Takano,E. and Breitling,R. (2014) Pep2Path: automated mass spectrometry-guided genome mining of peptidic natural products. *PLoS Comput. Biol.*, **10**, e1003822.
16. Blin,K., Shaw,S., Tong,Y. and Weber,T. (2020) Designing sgRNAs for CRISPR-BEST base editing applications with CRISPy-web 2.0. *Synth. Syst. Biotechnol.*, **5**, 99–102.
17. Navarro-Muñoz,J.C., Selem-Mojica,N., Mullowney,M.W., Kautsar,S.A., Tryon,J.H., Parkinson,E.I., De Los Santos,E.L.C., Yeong,M., Cruz-Morales,P., Abubucker,S. *et al.* (2020) A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.*, **16**, 60–68.
18. Kautsar,S.A., van der Hooft,J.J.J., de Ridder,D. and Medema,M.H. (2021) BiG-SLiCE: a highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters. *GigaScience*, **10**, giaa154.
19. Hannigan,G.D., Prihoda,D., Palicka,A., Soukup,J., Klempir,O., Rampula,L., Durcak,J., Wurst,M., Kotowski,J., Chang,D. *et al.* (2019) A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res*, **47**, e110.
20. Palaniappan,K., Chen,I.-M.A., Chu,K., Ratner,A., Seshadri,R., Kyrpides,N.C., Ivanova,N.N. and Mouncey,N.J. (2020) IMG-ABC

v.5.0: an update to the IMG/Atlas of Biosynthetic Gene Clusters Knowledgebase. *Nucleic Acids Res*, **48**, D422–D430.

21. Vallenet,D., Calteau,A., Dubois,M., Amours,P., Bazin,A., Beuvin,M., Burlot,L., Bussell,X., Fouteau,S., Gautreau,G. *et al.* (2020) MicroScope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Res*, **48**, D579–D589.
22. Terlouw,B.R., Blin,K., Navarro-Muñoz,J.C., Avalon,N.E., Chevrette,M.G., Egbert,S., Lee,S., Meijer,D., Recchia,M.J.J., Reitz,Z.L. *et al.* (2023) MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res*, **51**, D603–D610.
23. Kautsar,S.A., Blin,K., Shaw,S., Weber,T. and Medema,M.H. (2021) BiG-FAM: the biosynthetic gene cluster families database. *Nucleic Acids Res*, **49**, D490–D497.
24. Blin,K., Shaw,S., Kautsar,S.A., Medema,M.H. and Weber,T. (2021) The antiSMASH database version 3: increased taxonomic coverage and new query features for modular enzymes. *Nucleic Acids Res*, **49**, D639–D643.
25. Mistry,J., Chuguransky,S., Williams,L., Qureshi,M., Salazar,G.A., Sonnhammer,E.L.L., Tosatto,S.C.E., Paladin,L., Raj,S., Richardson,L.J. *et al.* (2021) Pfam: the protein families database in 2021. *Nucleic Acids Res*, **49**, D412–D419.
26. Haft,D.H., Selengut,J.D., Richter,R.A., Harkins,D., Basu,M.K. and Beck,E. (2013) TIGRFAMs and genome properties in 2013. *Nucleic Acids Res*, **41**, D387–D395.
27. Letunic,I., Khedkar,S. and Bork,P. (2021) SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res*, **49**, D458–D460.
28. van Heel,A.J., de Jong,A., Song,C., Viel,J.H., Kok,J. and Kuipers,O.P. (2018) BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res*, **46**, W278–W281.
29. Yadav,G., Gokhale,R.S. and Mohanty,D. (2009) Towards prediction of metabolic products of polyketide synthases: an In silico analysis. *PLOS Comput. Biol.*, **5**, e1000351.
30. Reitz,Z.L., Butler,A. and Medema,M.H. (2022) Automated genome mining predicts combinatorial diversity and taxonomic distribution of peptide metallophore structures. bioRxiv doi: https://doi.org/10.1101/2022.12.14.519525, 16 December 2022, preprint: not peer reviewed.
31. Leikoski,N., Fewer,D.P., Jokela,J., Wahlsten,M., Rouhiainen,L. and Sivonen,K. (2010) Highly diverse cyanobactins in strains of the genus Anabaena. *Appl. Environ. Microbiol.*, **76**, 701–709.
32. Helfrich,E.J.N., Ueoka,R., Dolev,A., Rust,M., Meoded,R.A., Bhushan,A., Califano,G., Costa,R., Gugger,M., Steinbeck,C. *et al.* (2019) Automated structure prediction of trans-acyltransferase polyketide synthase products. *Nat. Chem. Biol.*, **15**, 813–821.
33. Klau,L.J., Podell,S., Creamer,K.E., Demko,A.M., Singh,H.W., Allen,E.E., Moore,B.S., Ziemert,N., Letzel,A.C. and Jensen,P.R. (2022) The Natural Product Domain Seeker version 2 (NaPDoS2) webtool relates ketosynthase phylogeny to biosynthetic function. *J. Biol. Chem.*, **298**, 102480.
34. Röttig,M., Medema,M.H., Blin,K., Weber,T., Rausch,C. and Kohlbacher,O. (2011) NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res*, **39**, W362–W367.
35. Stachelhaus,T., Mootz,H.D. and Marahiel,M.A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.*, **6**, 493–505.
36. Rausch,C., Weber,T., Kohlbacher,O., Wohlleben,W. and Huson,D.H. (2005) Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res*, **33**, 5799–5808.
37. Kautsar,S.A., Suarez Duran,H.G., Blin,K., Osbourn,A. and Medema,M.H. (2017) plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res*, **45**, W55–W63.
38. Pascal Andreu,V., Roel-Touris,J., Dodd,D., Fischbach,M.A. and Medema,M.H. (2021) The gutSMASH web server: automated identification of primary metabolic gene clusters from the gut microbiota. *Nucleic Acids Res*, **49**, W263–W270.
39. Pascal Andreu,V., Augustijn,H.E., Chen,L., Zhernakova,A., Fu,J., Fischbach,M.A., Dodd,D. and Medema,M.H. (2023) gutSMASH predicts specialized primary metabolic pathways from the human gut microbiota. *Nat. Biotechnol.*
40. Castro-Mondragon,J.A., Riudavets-Puig,R., Rauluseviciute,I., Berhanu Lemma,R., Turchi,L., Blanc-Mathieu,R., Lucas,J., Boddie,P., Khan,A., Manosalva Pérez,N. *et al.* (2022) JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*, **50**, D165–D173.