

University of Groningen

antiSMASH 2.0-a versatile platform for genome mining of secondary metabolite producers

Blin, Kai; Medema, Marnix H.; Kazempour, Daniyal; Fischbach, Michael A.; Breitling, Rainer; Takano, Eriko; Weber, Tilmann

Published in:
Nucleic Acids Research

DOI:
[10.1093/nar/gkt449](https://doi.org/10.1093/nar/gkt449)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2013

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Blin, K., Medema, M. H., Kazempour, D., Fischbach, M. A., Breitling, R., Takano, E., & Weber, T. (2013). antiSMASH 2.0-a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Research*, 41(W1), W204-W212. <https://doi.org/10.1093/nar/gkt449>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Supplementary material for

antiSMASH 2.0 – a versatile platform for genome mining of secondary metabolite producers

Kai Blin^{*}, Marnix H. Medema^{*}, Daniyal Kazempour, Michael A. Fischbach, Rainer Breitling, Eriko Takano, and Tilmann Weber

Content

- Table S1: Signature HMMs for detection of secondary metabolite biosynthesis gene clusters
- Table S2: Rules for detection of secondary metabolite biosynthesis gene clusters
- Table S3: Validation of the detection of new gene cluster classes
- Table S4: antiSMASH benchmark
- Supplementary references

Table S1: Signature HMMs for detection of secondary metabolite biosynthesis gene clusters (extended according to Medema et al. 2011)

Compound class	Description	HMM name	Source
NRPS	Condensation domain	Condensation	PFAM PF00668.13
NRPS	Adenylation domain	AMP-binding	PFAM PF00501.21
NRPS	Adenylation domain with integrated oxidase	A-OX	Medema et al. (2011)
NRPS/PKS	Thiolation domain	PP-binding	PFAM PF00550.18
PKS	Ketosynthase domain	PKS_KS	SMART
PKS	Acyltransferase domain	PKS_AT	SMART
PKS	Trans-acyltransferase docking domain	ATd	Medema et al. (2011)
PKS (neg.)	Bacterial type I fatty acid synthase	bt1fas	Medema et al. (2011)
PKS (neg.)	Fungal type I fatty acid synthase	ft1fas	Medema et al. (2011)
PKS (neg.)	Type II fatty acid synthase	t2fas	Medema et al. (2011)
PKS (neg.)	FabH fatty acid synthase	fabH	Medema et al. (2011)
PKS	Enediyne ketosynthase	ene_KS	Yadav et al. (2009)
PKS	Modular ketosynthase	mod_KS	Yadav et al. (2009)
PKS	Hybrid ketosynthase	hyb_KS	Yadav et al. (2009)
PKS	Iterative ketosynthase	itr_KS	Yadav et al. (2009)
PKS	Trans-AT ketosynthase	tra_KS	Yadav et al. (2009)
PKS	Unusual PKS HgID-like	hgID	Medema et al. (2011)
PKS	Unusual PKS HgIE-like	hgIE	Medema et al. (2011)
PKS	Type II PKS ketosynthase	t2ks	Medema et al. (2011)
PKS	Type II PKS ketosynthase, model 2	t2ks2	Medema et al. (2011)
PKS	Type II PKS Chain length factor	t2clf	Medema et al. (2011)
PKS	Type III PKS N-terminal	Chal_sti_synt_N	PFAM PF00195.12
PKS	Type III PKS C-terminal	Chal_sti_synt_C	PFAM PF00195.12
Terpene	Terpene synthase C terminal	Terpene_synth_C	PFAM PF03936.9
Terpene	Terpene synthase	Terpene_synth	PFAM PF01397.14
Terpene	Phytoene synthase	phytoene_synth	Medema et al. (2011)
Terpene	Lycopene cyclase	Lycopene_cycl	PFAM PF05834.5
Terpene	Terpene cyclase	terpene_cyclase	Medema et al. (2011)
Terpene	NapT7-like protein	NapT7	Medema et al. (2011)
Terpene	Fungal geranylgeranyl	fung_ggpp	Medema et al.

Compound class	Description	HMM name	Source
	pyrophosphate synthase		(2011)
Terpene	Fungal geranylgeranyl pyrophosphate synthase, model 2	fung_ggpp2	Medema et al. (2011)
Terpene	Dimethylallyl tryptophan synthase	dmat	Medema et al. (2011)
Terpene	Trichodiene synthase	trichodiene_synt	Medema et al. (2011)
Lantipeptides	LanC-like Lantibiotics biosynthesis protein	LANC_like	PFAM PF05147.6
Lantipeptides	Lantibiotic dehydratase, N-terminus	Lant_dehyd_N	PFAM PF04737.6
Lantipeptides	Lantibiotic dehydratase, C-terminus	Lant_dehyd_C	PFAM PF04738.6
Lantipeptides	Lantibiotic antimicrobial peptide 18	Antimicrobial18	PFAM PF08130.4
Lantipeptides	Gallidermin	Gallidermin	PFAM PF02052.8
Lantipeptides	Lantibiotic, type A	L_biotic_typeA	PFAM PF04604.6
Lantipeptides	Lantibiotic, gallidermin/nisin family	TIGR03731	TIGR03731
Lantipeptides	Lantibiotic leader lacticin 481 group	LE-LAC481	De Jong et al. (2010)
Lantipeptides	Lantibiotic leader mersacidin cinnamycin group	LE-MER+2PEP	De Jong et al. (2010)
Lantipeptides	Lantibiotic leader LanBC modified	LE-LanBC	De Jong et al. (2010)
Lantipeptides	Lantibiotic peptide lacticin 481 group	MA-LAC481	De Jong et al. (2010)
Lantipeptides	Lantibiotic peptide nisin epidermin group	MA-NIS+EPI	De Jong et al. (2010)
Lantipeptides	Lantibiotic peptide nisin group	MA-NIS	De Jong et al. (2010)
Lantipeptides	Lantibiotic peptide epidermin group	MA-EPI	De Jong et al. (2010)
Lantipeptides	Lantibiotic peptide two component alpha	MA-2PEPA	De Jong et al. (2010)
Lantipeptides	Lantibiotic peptide two component beta	MA-2PEPB	De Jong et al. (2010)
Lantipeptides	lantibiotic peptide lacticin 481 group (dufour et al)	LE-DUF	De Jong et al. (2010)
Lantipeptides	lantibiotic leader lacticin 481 group (dufour et al)	MA-DUF	De Jong et al. (2010)
Lantipeptides	FxLD family lantipeptide	TIGR04363	TIGR04363
Lantipeptides	<i>Streptomyces</i> PEQAXS motif lantipeptide	Strep_PEQAXS	This study
Bacteriocin	Putative <i>Streptomyces</i> bacteriocin	strepbact	Medema et al. (2011)
Bacteriocin	Antimicrobial peptide 14	Antimicrobial14	PFAM PF08109.4

Compound class	Description	HMM name	Source
Bacteriocin	Bacteriocin_IIC domain	Bacteriocin_IIC	PFAM PF10439.2
Bacteriocin	Bacteriocin_IID domain	Bacteriocin_IID	PFAM PF09221.3
Bacteriocin	Bacteriocin_IIDC_cydomain	Bacteriocin_IIDC_cy	PFAM PF12173.1
Bacteriocin	Bacteriocin_II domain	Bacteriocin_II	PFAM PF01721.11
Bacteriocin	Bacteriocin_III domain	Bacteriocin_III	PFAM PF11758.1
Bacteriocin	Lactococcin	Lactococcin	PFAM PF04369.6
Bacteriocin	Antimicrobial peptide 17	Antimicrobial17	PFAM PF08129.4
Bacteriocin	Lactococcin 972	Lactococcin_972	PFAM PF09683.3
Bacteriocin	Lactococcin G-beta	LcnG-beta	PFAM PF11632.1
Bacteriocin	Subtilosin A	Subtilosin_A	PFAM PF11420.1
Bacteriocin	Cloacin	Cloacin	PFAM PF03515.7
Bacteriocin	Linocin M18	Linocin_M18	PFAM PF04454.5
Bacteriocin	Bacteriocin biosynthesis cyclodehydratase	TIGR03603	TIGR03603
Bacteriocin	Bacteriocin biosynthesis docking scaffold	TIGR03604	TIGR03604
Bacteriocin	SagB-type dehydrogenase	TIGR03605	TIGR03605
Bacteriocin	Bacteriocin, circularin A/uberolysin famil	TIGR03651	TIGR03651
Bacteriocin	Bacteriocin, microcyclamide/patellamide family	TIGR03678	TIGR03678
Bacteriocin	Thiazole-containing bacteriocin maturation protein	TIGR03693	TIGR03693
Bacteriocin	Bacteriocin propeptide	TIGR03798	TIGR03798
Bacteriocin	Bacteriocin biosynthesis cyclodehydratase	TIGR03882	TIGR03882
Bacteriocin	Bacteriocin, BA_2677 family	TIGR03601	TIGR03601
Bacteriocin	Bacteriocin protoxin, streptolysin S family	TIGR03602	TIGR03602
Bacteriocin	Two-chain TOMM family	TIGR03795	TIGR03795
Bacteriocin	NHLP leader peptide domain	TIGR03793	TIGR03793
Bacteriocin	Bacteriocin maturation radical SAM protein 1	TIGR03975	TIGR03975
Bacteriocin	Microviridin A	mvnA	Medema et al. (2011)
Bacteriocin	Thiostrepton-like thiopeptides	thiostrepton	Medema et al. (2011)
Bacteriocin	Putative subtilosin biosynthesis enzyme ywiA	subtilosin	This study
Bacteriocin	Cypermycin prepeptide	cypermycin	This study
Bacteriocin	Marinostatin/microviridin prepeptide	mvd	This study
Bacteriocin	Lasso peptide modification enzyme	lasso	This study

Compound class	Description	HMM name	Source
Bacteriocin	Small prepeptide associated domain	DUF692	PFAM DUF692
Bacteriocin	Microcin J25 prepeptide	micJ25	This study
Bacteriocin	Microcin J25 processing protein McjC	mcjC	This study
Bacteriocin	Glycocin prepeptide	glycocin	This study
Bacteriocin	Bottromycin biosynthesis enzyme	botH	This study
Bacteriocin	SkfC biosynthesis enzyme	skfc	This study
Bacteriocin	Thuricin prepeptide	thuricin	This study
Bacteriocin	Sublancin prepeptide	sublancin	This study
Beta-lactams	Beta-lactam synthase	BLS	Medema et al. (2011)
Beta-lactams	Clavulanic acid synthase-like	CAS	Medema et al. (2011)
Beta-lactams	Tabtoxin synthase-like	Tabtoxin	Medema et al. (2011)
Aminoglycosides / aminocyclitols	2-deoxy-scyllo-inosose synthase	DOIS	Medema et al. (2011)
Aminoglycosides / aminocyclitols	NeoL-like deacetylase	neoL_like	Medema et al. (2011)
Aminoglycosides / aminocyclitols	SpcD-/SpcK-like thymidylyltransferas	spcDK_like_glyc	Medema et al. (2011)
Aminoglycosides / aminocyclitols	SpcF-/SpcG-like glycosyltransferase	spcFG_like	Medema et al. (2011)
Aminoglycosides / aminocyclitols	StrH-like glycosyltransferase	strH_like	Medema et al. (2011)
Aminoglycosides / aminocyclitols	StrK-like phosphatase	strK_like1	Medema et al. (2011)
Aminoglycosides / aminocyclitols	StrK-like phosphatase	strK_like2	Medema et al. (2011)
Aminoglycosides / aminocyclitols	ValA-like 2-epi-5-epi-valiolone synthase	valA_like	Medema et al. (2011)
Aminoglycosides / aminocyclitols	2-epi-5-epi-valiolone synthase, SalQ-like	salQ	Medema et al. (2011)
Aminocoumarins	NovK-like reductase	novK	Medema et al. (2011)
Aminocoumarins	NovJ-like reductase	novJ	Medema et al. (2011)
Aminocoumarins	NovI-like cytochrome P450	novl	Medema et al. (2011)
Aminocoumarins	NovH-like protein	novH	Medema et al. (2011)
Aminocoumarins	SpcD/SpcK-like thymidylyltransferase, aminocoumarins group	spcDK_like_cou	Medema et al. (2011)
Siderophores	Siderophore synthase	IucA_IucC	PFAM PF04183.5

Compound class	Description	HMM name	Source
Ectoines	Ectoine synthase	ectoine_synt	Medema et al. (2011)
Butyrolactones	AfsA-like butyrolactone synthase	AfsA	PFAM PF03756.6
Indoles	StaD-like chromopyrrolic acid synthase domain	indsynth	Medema et al. (2011)
Nucleosides	LipM-like nucleotidyltransferase	LipM	Medema et al. (2011)
Nucleosides	LipU-like protein	LipU	Medema et al. (2011)
Nucleosides	LipV-like dehydrogenase	LipV	Medema et al. (2011)
Nucleosides	ToyB-like synthase	ToyB	Medema et al. (2011)
Nucleosides	TunD-like putative N-acetylglucosamine transferase	TunD	Medema et al. (2011)
Nucleosides	Pur6-like synthetase	pur6	Medema et al. (2011)
Nucleosides	Pur10-like oxidoreductase	pur10	Medema et al. (2011)
Nucleosides	NikJ-like protein	nikJ	Medema et al. (2011)
Nucleosides	NikO-like enolpyruvyl transferase	nikO	Medema et al. (2011)
Phosphoglycolipids	MoeO5-like prenyl-3-phosphoglycerate synthase	MoeO5	Medema et al. (2011)
Phosphoglycolipids	Phosphoglycolipid glycosyltransferase	moeGT	This study
Melanins	MelC-like melanin synthase	melC	Medema et al. (2011)
Oligosaccharide	Secondary metabolite-related glycosyltransferase	Glycos_transf_1	PFAM PF00534.14
Oligosaccharide	Secondary metabolite-related glycosyltransferase	Glycos_transf_2	PFAM PF00535.20
Oligosaccharide	Secondary metabolite-related glycosyltransferase	Glyco_transf_28	PFAM PF03033.14
Oligosaccharide	Secondary metabolite-related glycosyltransferase	DUF1205	PFAM DUF1205
Oligosaccharide	Secondary metabolite-related glycosyltransferase	MGT	This study
Oligosaccharide	Secondary metabolite-related glycosyltransferase	MGT2	This study
Furan	MmyO-like protein	mmyO	This study
Homoserine lactone	Autoinducer synthetase	Autoind_synth	PFAM PF00765.12
Thiopeptide	YcaO-like	YcaO	PFAM PF02624.11
Phenazine	Phenazine biosynthesis	phzB	This study

Compound class	Description	HMM name	Source
	gene		
Phosphonate	Phosphonate biosynthesis	phosphonates	This study
Others	NAD-binding domain 4	NAD_binding_4	PFAM PF07993.5
Others	LmbU-like protein	LmbU	Medema et al. (2011)
Others	Goadsporin-like protein	goadsporin_like	Medema et al. (2011)
Others	Neocarzinostatin-like protein	Neocarzinostat	Medema et al. (2011)
Others	Cyanobactin protease	cyanobactin_synth	Medema et al. (2011)
Others	Cyclodipeptide synthase	cycdipepsynth	Medema et al. (2011)
Others	Fom1-like phosphomutase	fom1	Medema et al. (2011)
Others	BcpB-like phosphomutase	bcpB	Medema et al. (2011)
Others	FrbD-like phosphomutase	frbD	Medema et al. (2011)
Others	MitE-like CoA-ligase	mitE	Medema et al. (2011)
Others	Valanimycin biosynthesis VlmB domain	vlmB	Medema et al. (2011)
Others	Pyrrolnitrin biosynthesis PrnB domain	prnB	Medema et al. (2011)
Others	Nitrosynthase domain	CaiA	This study
Others	Bacilysin-related ligase	bacilysin	This study
Others	Cypermycin biosynthesis cypl enzyme	cypl	This study

Table S2: Rules for detection of secondary metabolite biosynthesis gene clusters (extended according to Medema et al., 2011)

Biosynthetic class	Rules
Type I PKS	<ul style="list-style-type: none"> - KS & AT HMM scores > 50 within one protein - KS score > bactTypeIFAS / fungTypeIFAS / HgID&E / FabH scores
Trans-AT type I PKS	<ul style="list-style-type: none"> - trans-AT docking domain HMM score > 65 - KS score > 50 - no match to rules for normal type I PKSs as above
Type II PKS	<ul style="list-style-type: none"> - type II KS or CLF HMM score > 50 - KS/CLF score > enedyneKS / modularKS / hybridKS / iterativeKS / transATKS / bactTypeIFAS / fungTypeIFAS / TypeIIIFAS / HgID&E / FabH HMM scores - no match to rules for normal/trans type I PKSs as above
Type III PKS	<ul style="list-style-type: none"> - Chal_sti_synt_C or Chal_sti_synt_N HMM scores > 35 - no match to rules for normal type I&II PKSs as above
Type IV PKS	<ul style="list-style-type: none"> - HgIE or HgID HMM scores > 50 - HgID/E HMM score > bactTypeIFAS / fungTypeIFAS / TypeIIIFAS / FabH HMM scores - no match to rules for normal type I&II&III PKSs as above
Non-ribosomal peptide synthetase	<ul style="list-style-type: none"> - C & A / A-OX domain HMM scores > 20 within one protein OR single domain C & A proteins scores > 20 within 20 kb distance
Terpene	<ul style="list-style-type: none"> - Terpene_Synth HMM score > 23 OR Terpene_Synth_C HMM score > 23 OR phytoene_synt HMM score > 20 OR Lycopene_cycl HMM score > 80 OR terpene_cyclase HMM score > 50 OR NapT7 HMM score > 250 OR fung_ggpps HMM score > 420 OR fung_ggpps2 HMM score > 312 OR dmat HMM score > 200 OR trichodiene_synth HMM score > 150
Lantipeptides	<ul style="list-style-type: none"> - LANC_like HMM score > 80 OR (Lant_dehydN and LantdehydC HMM scores > 20 within one protein) OR one of a range of lantibiotic prepeptide HMM scores > 20 OR TIGR03731 HMM score > 18
Bacteriocins	<ul style="list-style-type: none"> - Streptbact HMM score > 50 OR Antimicrobial14 HMM score > 90 OR Bacteriocin_IId HMM score > 23 OR Bacteriocillc_cy HMM score > 92 OR Bacteriocin_II HMM score > 40 OR Lactococcin HMM score > 24 OR Antimicrobial17 HMM score > 31 OR Lactococcin_972 HMM score > 25 OR Bacteriocin_Ilc HMM score > 27 OR LcnG-beta HMM score > 78 OR Bacteriocin_Iii HMM score > 56 OR Subtilosin_A HMM score > 98 OR Cloacin HMM score > 27 OR Linocin_M18 HMM score > 25 OR TIGR03603 HMM score > 150 OR TIGR03604 HMM score > 440 OR TIGR03605 HMM score > 200 OR TIGR03651 HMM score > 18 OR TIGR03678 HMM score > 35 OR TIGR03693 HMM score > 400 OR TIGR03798 HMM score > 16 OR TIGR03882 HMM score > 150 OR TIGR03601 HMM score > 50 OR TIGR03602 HMM score > 50 OR TIGR03795 HMM score > 41 OR TIGR03793 HMM score > 51 OR TIGR03975 HMM score > 282 OR mvnA HMM score > 20 OR thiostrepton HMM score > 20 OR subtilosin HMM score > 140 OR cypermycin HMM score > 10 OR mvd HMM score > 20 OR lasso HMM score > 400 OR DUF692 HMM score > 40 OR micJ25 HMM score > 21 OR mcjC HMM score > 60 OR glycocin HMM score > 30 OR both HMM score > 65 OR skfc HMM score > 70 OR thuricin HMM score > 30 OR sublancin HMM score > 30
Beta-lactams	<ul style="list-style-type: none"> - Beta-lactam synthase HMM score > 250 OR clavulanic acid synthase HMM score > 250 OR tabtoxin synthase score > 500
Aminoglycosides /	<ul style="list-style-type: none"> - strH HMM score > 50 OR strK1 HMM score > 800 OR strK2 HMM score >

Biosynthetic class	Rules
aminocyclitols	650 OR NeoL HMM score > 50 OR DOIS HMM score > 500 OR ValA HMM score > 600 OR SpcFG HMM score > 200 OR SpcDK_glyc HMM score > 600 OR salQ HMM score > 480
Aminocoumarins	- novK HMM score > 200 OR novJ HMM score > 350 OR novI HMM score > 600 OR novH HMM score > 750 OR spcDK_like_cou HMM score > 600
Siderophores	- lucA_lucC HMM score > 30
Ectoines	- Ectoine_synt HMM score > 35
Butyrolactones	- AfsA HMM score > 25
Indoles	- ind_synth HMM score > 100
Nucleosides	- LipM HMM score > 50 OR LipU HMM score > 30 OR LipV HMM score > 375 OR ToyB HMM score > 175 OR TunD HMM score > 200 OR pur6 HMM score > 200 OR pur10 HMM score > 600 OR nikJ HMM score > 200 OR nikO HMM score > 400
Phosphoglycolipids	- MoeO5 HMM score > 65 OR moeGT HMM score > 40
Melanins	- melC HMM score > 40
Oligosaccharide	- at least three out of (Glycos_transf_1 HMM score > 20, Glycos_transf_2 HMM score > 20, Glyco_transf_28 HMM score > 26, MGT HMM score > 100, MGT2 HMM score > 150, DUF1205 HMM score > 20)
Furan	- mmvO HMM score > 500
Homoserine lactone	- Autoind_synth HMM score > 20
Thiopeptide	- (Lant_dehyd_N HMM score > 20 OR Lant_dehyd_C HMM score > 20) and YcaO HMM score > 25 exist within 10 kilobases of each other
Phenazine	- phzB HMM score > 80
Phosphonate	- phosphonates HMM score > 275
Others	- PP-binding & AMP-binding HMM scores > 20 within one protein OR (PP-binding HMM score > 20 and A-OX HMM score > 50 within one protein) OR (NAD_binding_4 HMM score > 40 and A-OX HMM score > 50 within one protein) OR (NAD_binding_4 HMM score > 40 and AMP-binding HMM score > 20 within one protein) OR LmbU HMM score > 50 OR goadsporin_like HMM score > 500 OR Neocarzinostat HMM score > 28 OR cyanobactin_synth HMM score > 80 OR cycdipepsynth HMM score > 110 OR fom1 HMM score > 750 OR bcpB HMM score > 400 OR frbD HMM score > 350 OR mitE HMM score > 400 OR vlmB HMM score > 250 OR prnB HMM score > 200 or CaiA HMM score > 200 or bacilysin HMM score > 160 or cypl HMM score > 15

Table S3: Validation of the detection of new gene cluster classes

Compound	Compound class	GenBank Accession nr	Detected successfully?
everninomicin	oligosaccharide	GY241466	yes
avilamycin	oligosaccharide	AF333038	yes
thiocillin	thiopeptide	NC_004722	yes
cyclothiazomycin	thiopeptide	FJ472825	yes
thiostrepton	thiopeptide	FJ652572	yes
thiomuracin	thiopeptide	FJ461360	yes
siomycin	thiopeptide	FJ436355	yes
nosiheptide	thiopeptide	FJ438820	yes
nocathiacin	thiopeptide	GU564398	yes
GE37468	thiopeptide	GE37468	yes
GE2270	thiopeptide	GE2270	yes
TP-1161	thiopeptide	TP-1161	yes
phenazine	phenazine	JX843821	yes
phenazine	phenazine	JQ659263	yes
phenazine	phenazine	FN178498	yes
phenazine	phenazine	AM384985	yes
phenazine	phenazine	AY927995	yes
phenazine	phenazine	HM594285	yes
methyleneomycin	furan	AJ276673	yes
homoserine lactone	homoserine lactone	ECU17224	yes
homoserine lactone	homoserine lactone	ASU65741	yes
homoserine lactone	homoserine lactone	L48616	yes
homoserine lactone	homoserine lactone	AF079136	yes
homoserine lactone	homoserine lactone	AF057718	yes
fosfomycin	phosphonate	EU924263	yes
dehydropbos	phosphonate	GU199252	yes
FR900098	phosphonate	DQ267750	yes

Table S4: antiSMASH benchmark

antiSMASH 1.0

	Total annotated gene clusters in genome publication	Detected annotated gene clusters	Newly detected gene clusters, not annotated in genome publication	False positives	False negatives	Misannotated
<i>Pseudomonas fluorescens Pf-5</i>	10	9		0	0	1
<i>Streptomyces griseus IFO 13350</i>	34	33		8	0	1
<i>Kitasatospora setae NBRC 14216T</i>	24	24		12	0	0
<i>Salinispora tropica CNB-440</i>	17	16		5	0	1
<i>Aspergillus fumigatus Af293</i>	26	26		10	0	0
	111	108		35	0	3
						2

Percentage detected 97.3
 Percentage false positives 0.0
 Percentage false negatives 2.7
 Percentage found new 31.5

antiSMASH 2.0

	Total annotated gene clusters in genome publication	Detected annotated gene clusters	Newly detected gene clusters, not annotated in genome publication	False positives	False negatives	Misannotated
<i>Pseudomonas fluorescens Pf-5</i>	10	9		2	0	1
<i>Streptomyces griseus IFO 13350</i>	34	33		9	0	1
<i>Kitasatospora setae NBRC 14216T</i>	24	24		13	0	0
<i>Salinispora tropica CNB-440</i>	17	16		5	0	1
<i>Aspergillus fumigatus Af293</i>	26	26		10	0	0
	111	108		39	0	3
						2

Percentage detected 97.3
 Percentage false positives 0.0
 Percentage false negatives 2.7
 Percentage found new 35.1

Supplementary references

1. 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.
2. de Jong, A., van Heel, A.J., Kok, J. and Kuipers, O.P. (2010) BAGEL2: mining for bacteriocins in genomic data. *Nucleic Acids Res.*, **38**, W647-651.