W72–W79 *Nucleic Acids Research, 2017, Vol. 45, Web Server issue* *Published online 29 April 2017*
doi: 10.1093/nar/gkx344

SBSPKSv2: structure-based sequence analysis of polyketide synthases and non-ribosomal peptide synthetases

Shradha Khater, Money Gupta, Priyesh Agrawal, Neetu Sain, Jyoti Prava, Priya Gupta, Mansi Grover, Narendra Kumar and Debasisa Mohanty*

National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110067, India

Received February 24, 2017; Revised April 10, 2017; Editorial Decision April 16, 2017; Accepted April 25, 2017

ABSTRACT

Genome guided discovery of novel natural products has been a promising approach for identification of new bioactive compounds. SBSPKS web-server has been a valuable resource for analysis of polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) gene clusters. We have developed an updated version - SBSPKSv2 which is based on comprehensive analysis of sequence, structure and secondary metabolite chemical structure data from 311 experimentally characterized PKS/NRPS gene clusters with known biosynthetic products. A completely new feature of SBSPKSv2 is the inclusion of features for search in chemical space. It allows the user to compare the chemical structure of a given secondary metabolite to the chemical structures of biosynthetic intermediates and final products. For identification of catalytic domains, SBSPKS now uses profile based searches, which are computationally faster and have high sensitivity. HMM profiles have also been added for a number of new domains and motif information has been used for distinguishing condensation (C), epimerization (E) and cyclization (Cy) domains of NRPS. In summary, the new and updated SBSPKSv2 is a versatile tool for genome mining and analysis of polyketide and non-ribosomal peptide biosynthetic pathways in chemical space. The server is available at: http://www.nii.ac.in/sbspks2.html.

INTRODUCTION

Polyketides (PKs) and non-ribosomal peptides (NRPs) are two major classes of secondary metabolites with diverse chemical structures (1,2) and a valuable source of pharmaceutically important molecules. The enormous diversity in chemical structures and hence their bioactivities, stem from the thio-template mechanism used by polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs). The tailoring enzymes that act after biosynthesis of core polyketide or non-ribosomal peptide scaffold, are capable of adding a plethora of functional groups to further diversify the final metabolites (3). An in-depth understanding of the biosynthetic mechanism and ways to adapt it, might yield valuable results in the form of therapeutically important products (4,5). Given their pharmaceutical relevance, PKS and NRPS gene clusters and the metabolites have been extensively characterized (6). The pharmaceutical importance of these natural products and the role genome that mining has played in the discovery and characterization of new natural products, prompted us to develop SBSPKS (Structure based sequence analysis of PKS and NRPS)—a web-based tool for sequence and structural analysis of PKSs and NRPSs (7). SBSPKS is one of the user friendly web-servers for analysis of PKS and NRPS megasynthases, their substrate prediction and a variety of other sequence and structural analysis (8–12). Recent reviews on computational methods for natural product discovery, have compared various features of SBSPKS and other similar bioinformatics tools like AntiSMASH (13), ClusScan (11), NP.Searcher (14) and SMURF (15), and have provided overviews on utilities of such tools in genome mining studies (13,16). Since the first version of SBSPKS was released, advances in high throughput technologies have unveiled a large number of microorganisms containing putative natural product biosynthetic gene clusters with unknown biosynthetic products (17), and also large number of natural products for which biosynthetic gene clusters are unknown. Since SBSPKS uses a knowledge based approach for formulation of its prediction rules, it is essential that its backend databases are updated to include information on experimentally characterized PKS and NRPS gene clusters. It is also necessary that computational methods/tools are suitably updated for optimum execution time with increased data size and to facilitate new types of searches. In addition to robust genome mining tools, tools which aid in search of chemical space are also required. Therefore, we have devel-

*To whom correspondence should be addressed. Tel: +91 11 26703749; Fax: +91 11 26742125; Email: deb@nii.res.in

oped SBSPKSv2 which integrates genomic and chemical information, and helps not only in improved analysis of PKS and NRPS gene clusters, but also in analysis of the chemical space of these secondary metabolites. Table 1 provides a summary of the comparative analysis of various features of the major web-servers currently used in genome mining of secondary metabolites. Most of these software do not store chemical structures of starters, extenders, biosynthetic intermediates and final secondary metabolites in SMILES format. Hence, the feature for search in chemical space is hitherto unavailable in most other web servers available for analysis of PK and NRP biosynthetic pathways. Currently PRISM is the only other tool which allows comparison of predicted chemical structures of secondary metabolites with structures of known secondary metabolites (18). However, detailed analysis of biosynthetic PKS/NRPS pathways in chemical space cannot be carried out using PRISM.

The updated version of SBSPKS has been divided into chemical and genomic space. The chemical space of SBSPKSv2 can be probed using available tools like search for chemically similar compounds and search for potential tailoring reactions. These search tools are based on manually curated database of more than 200 biosynthetic pathways. The pathways can be visualized as interactive graphs. The utility of these tools has been described using an orphan PK-Albocycline. To the best of our knowledge there are no databases or tools which catalog the information on PKS and NRPS pathways in chemical space at such details and provide users with tools to analyze it (Table 1). Generic pathway databases like KEGG catalog a common pathway map for all PKs and NRPs (19). Recently, Khater *et al.* and Dejong *et al.* have independently developed bioinformatics pipelines for retro biosynthetic analysis of PKs and NRPs (20,21), but they lack a curated database and user friendly interfaces for analysis of characterized pathways (22). A well curated database of PK and NRP biosynthetic pathways in chemical space will also help in verification of the available tools for retro biosynthetic enumeration of biochemical transformations. The genomic space of SBSPKSv2 has also been updated and it now includes 311 manually curated gene clusters. Though extensive manual curation and restricting our database to only experimentally characterized clusters limit the number of entries in SBSPKSv2, it makes this web-server a valuable resource for accessing experimentally characterized PKS/NRPS gene clusters. The genome mining tool of SBSPKSv2 now uses faster and more sensitive profile based search to detect regular PKS/NRPS catalytic domains as well as other unusual domains which occur less frequently in PKS/NRPS biosynthetic gene clusters. In addition to modeling three dimensional (3D) structures of PKS modules, a new feature to model 3D structures of NRPS module has also been included. The interfaces, for analysis of PKS/NRPS biosynthetic pathways in genomic and chemical space have also been seamlessly interlinked with each other.

Combined with the new features and updates, SBSPKSv2 can potentially help in characterization of new secondary metabolites and in redesigning known biosynthetic pathways to produce novel compounds of therapeutic importance. In summary, SBSPKSv2 is an user-friendly, up-to-date and manually curated web server which has undergone several crucial improvements.

METHODS AND IMPLEMENTATION

New features

SBSPKS chemical space. Traditional methods like microbial isolation and culturing combined with newer methods like genetic engineering and metagenomics have yielded >11000 PKs and NRPs (20). Also, advances in sequencing technologies have exponentially increased the rate of discovery of new PKS and NRPS gene clusters. Of the 11000 PKs and NRPs discovered, a very small percentage has its biosynthetic gene cluster known. Gene cluster discovery of these secondary metabolites can be facilitated by comparing them to characterized PKs, NRPs and their biosynthetic intermediates. Two essential requirements for such searches are, a well curated database containing characterized biosynthetic pathways of PKs and NRPs and suitable tool(s) to search and analyze the chemical structures of secondary metabolites and their biosynthetic intermediates. Therefore, to assist in the discovery of gene clusters of orphan PKs and NRPs and help in rational design of novel engineered products, we have developed a completely new interface in SBSPKSv2–PKS/NRPS chemical space.

Similar chemical structure search. To understand the biosynthetic pathway of an orphan PK or NRP, user can search for chemically similar molecules using the 'Reaction Search' module (Figure 1). The search for chemically similar PKs and NRPs accepts chemical structure of query molecule in SMILES format. Chemical structures in SMILES format can be obtained from PUBCHEM for a large number of metabolites (18). If not available in PUBCHEM or other websites, user can generate it using PubChem Sketcher (23). 'Reaction Search' module allows users to restrict their search by defining the number of matches, Tanimoto score or sub-structural patterns in SMARTS format. The algorithm then compares the given molecule to ∼2000 biosynthetic intermediates and final products of experimentally characterized PKs and NRPs using the similarity search option of Open Babel which is based on substructure based fingerprints (24). Links to the biosynthetic pathway page of the hits provided by the tool can help in deciphering putative biosynthetic pathways of the query compound.

Tailoring reaction search. In addition to the variation in starter/extender molecules and length of PKs and NRPs, cyclization reactions and post PKS/NRPS modifications add to the complexity and diversity of PKs and NRPs. The tailoring enzymes are usually present in synteny of PKS and NRPS genes. Therefore, deciphering the cyclization modes and tailoring steps will not only help in understanding the pathway but will also help in narrowing down the biosynthetic gene cluster. Extensive analysis of the biosynthetic pathways of PKs and NRPs helped us in extracting close to 20 functional groups involved in tailoring reactions and cyclizations (Supplementary Table S1). These functional groups are stored in SMARTS format and form the basis of search for potential tailoring reactions (Figure 2). Open

Table 1. Comparison of various web servers for analysis of PKS and NRPS biosynthetic pathways

	Features								
Webserver	Identification of NRPS/PKS Domains	Identification of clusters having similar ORFs	Similar biosynthetic Cluster prediction	Specificity prediction (A/AT)	NRPS/PKS 3D Modeling	SMILES for st arter/extender /intermediates and final secondary metabolite	Comparison of pathways in chemical space	Tailoring reaction detection	Chemical structure similarity search
SBSPKSv2	+	+		+	+	+	+	+	+
AntiSMAH	+	+	+	+					
PRISM	+	+	+	+					+
SMURF	+								
CLUSEAN	+			+					
ClustScan	+			+					
NP.Searcher	+			+					
NRPSpredictor2				+					

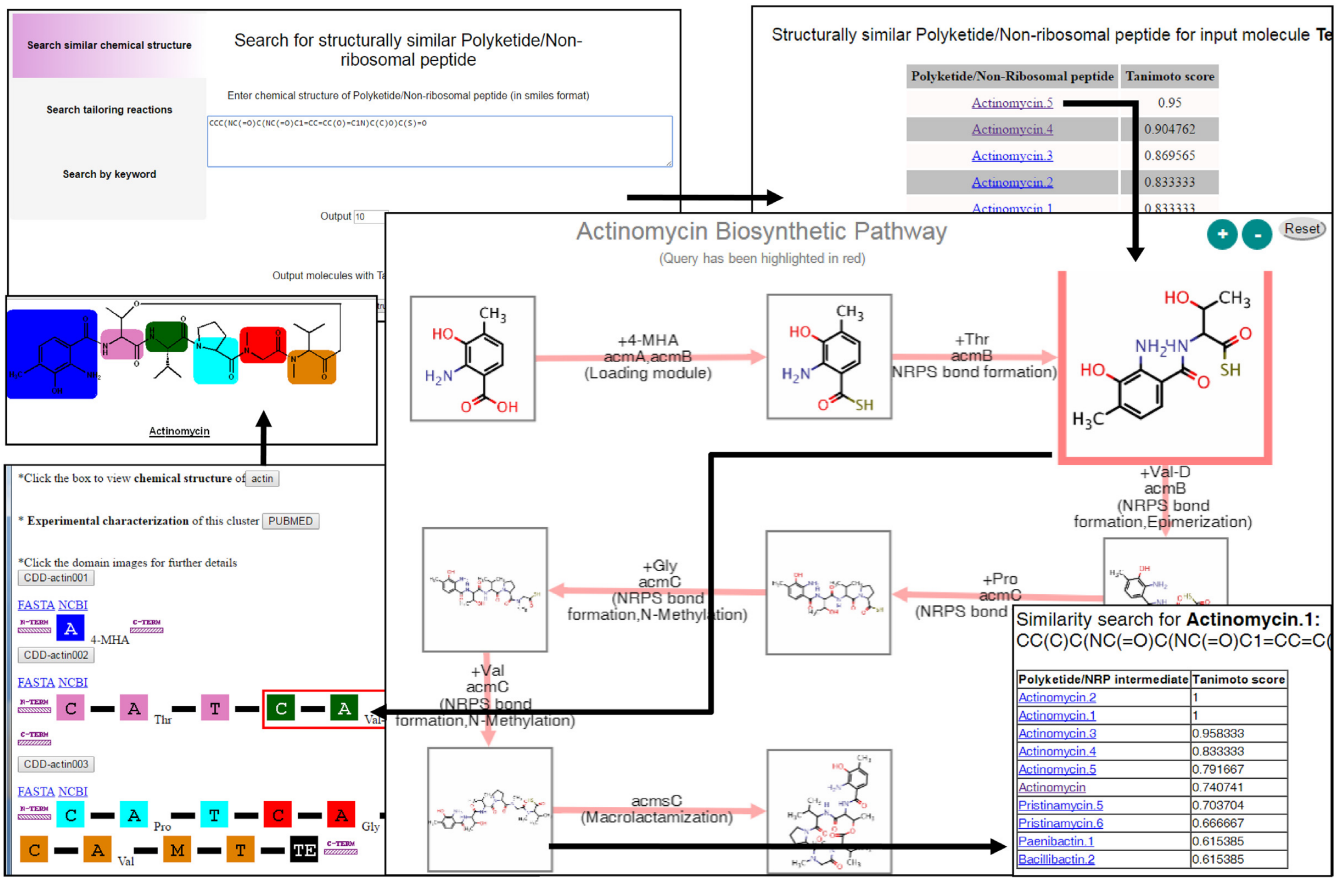


Figure 1. The figure depicts search for similar structures in chemical space. The search for structurally similar polyketide and non-ribosomal peptide allows users to match a query molecule to the biosynthetic intermediates of experimentally characterized polyketide and non-ribosomal peptide. The links on the result page can be used to navigate to the respective page in the biosynthetic pathway database. The database catalogs biosynthetic pathways of >200 polyketides and non-ribosomal peptides. Chemical structures of each step are stored in SMILES format, along with the reactions, monomer/extender unit and enzymes involved. Clicking on the reaction arrow links to the respective module/enzyme in the genomic space of SBSPKS. The genomic space also provides a cross link to the chemical space. Chemical structures similar to the biosynthetic intermediates can be searched by clicking the intermediates.

Babel is used to match the query molecule (SMILES format) with the stored functional groups. A hit indicates presence of the functional group and hence suggests that the respective reaction is potentially involved in the biosynthesis of query. The result page also provides an option to visualize the functional group added by the predicted reaction by highlighting it in chemical structure of the query.

Biosynthetic pathways database. The similarity search and search for potential tailoring reaction uses an elaborate database of biosynthetic pathways in chemical space at the backend. The database contains biosynthetic pathway of >200 experimentally characterized PKs and NRPs. Based on extensive manual curation of published literature, chemical structures of metabolites and sequences of biosynthetic enzymes, each step involved in the biosynthesis of PKs or NRPs have been cataloged in the database along

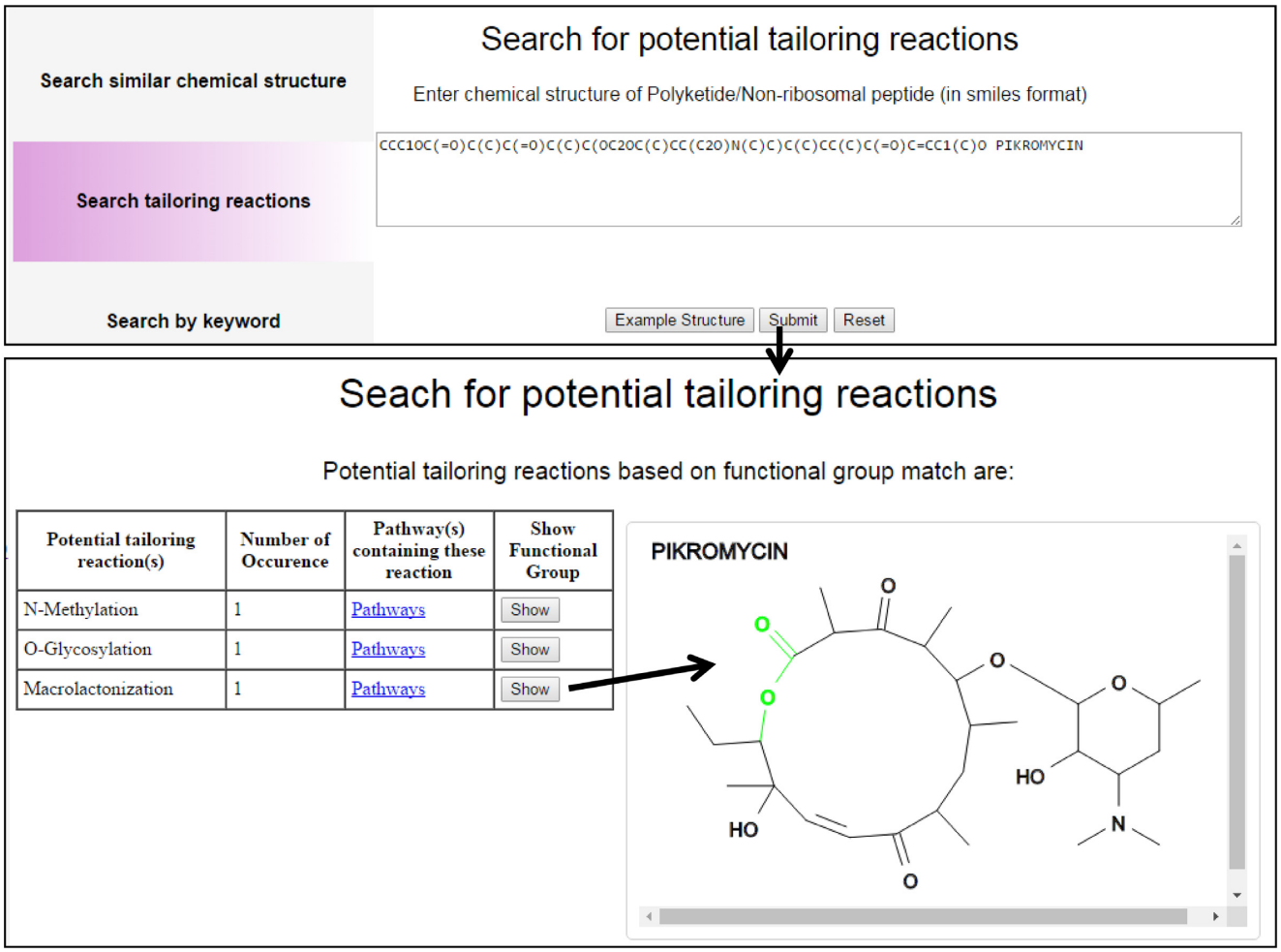


Figure 2. The reaction search part of SBSPKSv2 provides search based on chemical structures (Figure 1 lower panel), search for possible tailoring reactions and search for keywords. The search for potential tailoring reaction, lists the predicted reactions along with link to other biosynthetic pathways containing the same functional group and also provides a link to visualize the functional group by highlighting it in green.

with the reactions, enzyme names, accession numbers and monomers added. Approximately 2000 chemical structures of biosynthetic intermediates are stored in SMILES format and >1000 sequences of enzymes involved in the characterized PKs/NRPs pathway have been stored. The PK and NRP pathways have been represented as interactive graphs (Figure 1). The pathway pages use embedded JavaScript-based Cytoscape.js (25). Each graph starts with the starter moiety and catalogs the intermediate steps to terminate at the complete metabolite. The nodes of the graph represent the biosynthetic intermediates and the edges represent the reaction converting each intermediate. Images of chemical structure of intermediates have been used to depict the nodes. All nodes and edges in the graph based viewer can be dragged by the user to any desired position and can be clicked to show additional details. Individual nodes can be clicked to view a larger image of chemical structure, representation in SMILES format and link to structurally similar metabolites. Each edge label depicts the monomer being added (if applicable), gene name corresponding to the enzyme involved and reaction name. The web-server also allows user to download the pathway map of each metabolite as a flat file. Feature for searches in the text part of the database has been made available using the keyword search functionality. For example, it can help in search for all PKS/NRPS pathway where the monomer alanine or methyl malonate is added or all pathways where a particular reaction like methyl-transfer or epoxidation occurs. The identified pathways can then be visualized as interactive graphs.

Interlinking chemical and genomic space. The genomic and the chemical space of SBSPKSv2 have been interlinked by cross references between related features/records. Clicking on the edge of a reaction graph in chemical space allows the user to visualize the corresponding biosynthetic enzyme in genomic space of SBSPKS and carry out further analysis of its sequence or structural features. The link displays the complete biosynthetic gene cluster where the selected enzyme is highlighted (Figure 1). Similarly in the HTML pages which depict domain organizations for each biosynthetic gene cluster in genomic space, each domain has been interlinked to the chemical transformation it catalyzes in

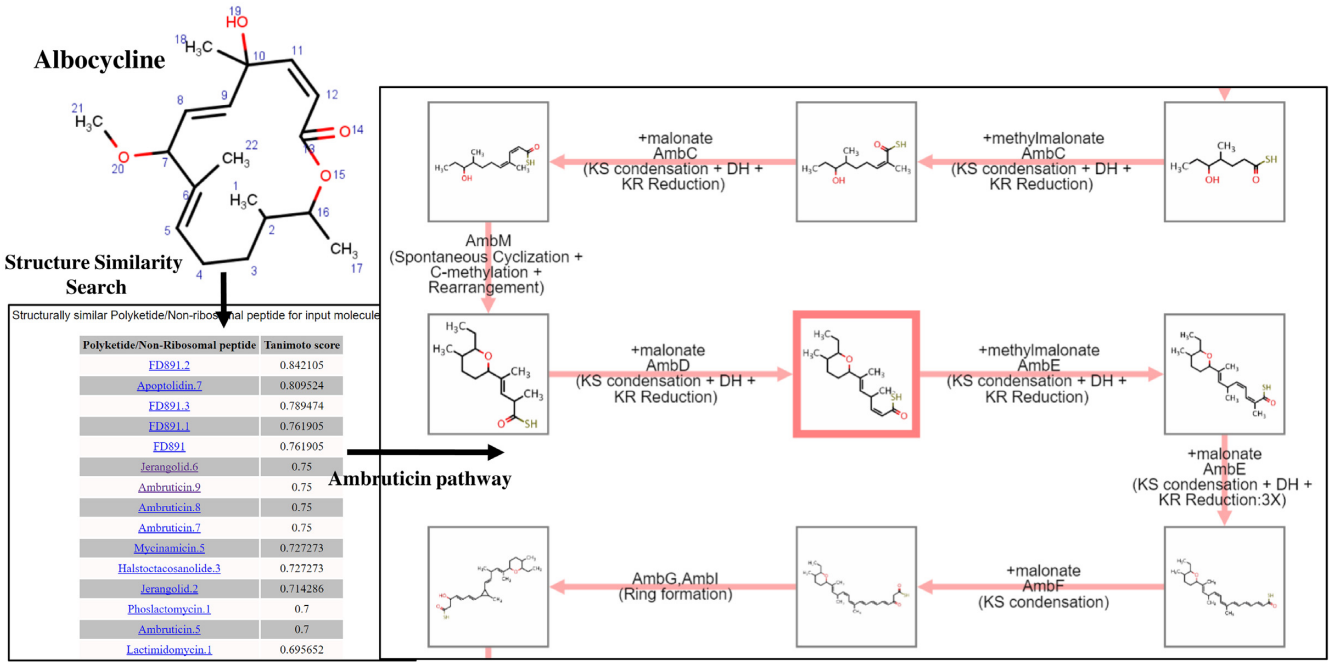


Figure 3. Understanding the origin of unusual double bond in orphan polyketide Albocycline. Search for chemical structures similar to albocycline showed similarity to jerangolid and ambruticin among others. Interestingly, these two polyketides contain the same unusual double bond. Study of the complete pathway revealed the origin of double bond through rearrangement.

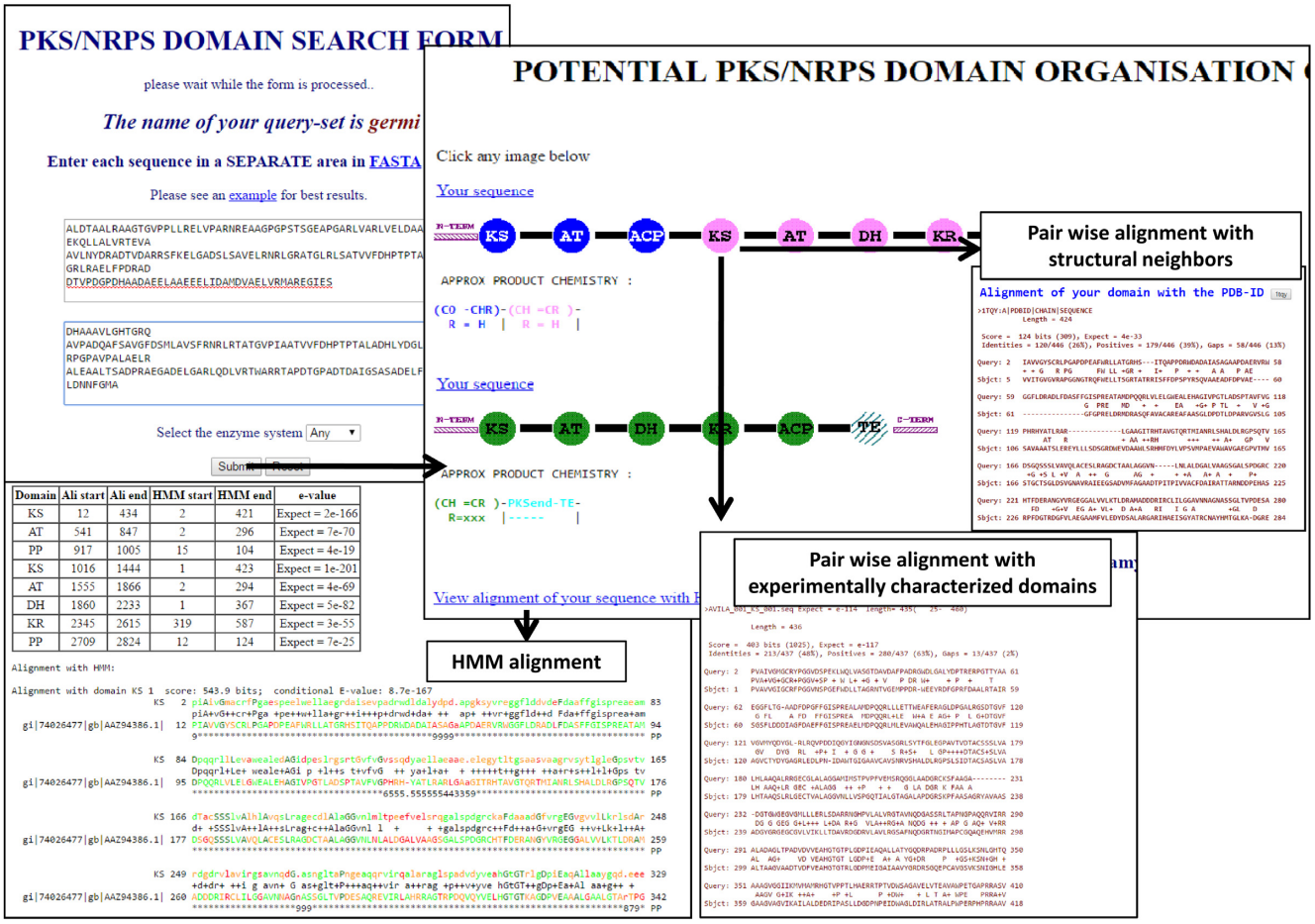
chemical space. Clicking on the domain leads to a page which not only provides interfaces for a variety of sequence as well as structural analysis, but also provides a link to the biosynthetic pathway database in the chemical space (Supplementary Figure S1). The reaction catalyzed by the selected domain is highlighted in red. Thus SBSPKSv2 provides interfaces for seamless transitions between genomic and chemical space and carry out various types of analysis.

Case study. The new chemical space interface of SBSPKSv2 can therefore help in the search for biosynthetic cluster of orphan PKs and NRPs. The utility of SBSPKSv2 chemical space can be demonstrated using an orphan antibiotic–albocycline (Figure 3). Albocycline has been shown to be effective against methicillin resistant *Staphylococcus aureus* but its biosynthetic gene cluster still remains unknown (26,27). Though an *in silico* analysis has predicted albocycline to be a product of PKS gene cluster comprised of six elongation modules (28), the origin of the unusual diene system (C8-C9 and C11–C12) remains obscure. Therefore to understand the biosynthesis of albocycline and origin of the unusual double bond we searched for closest structural match to albocycline in chemical space. Though the overall structure of albocycline looks similar to pikromycin and erythromycin the closest structural match came from biosynthetic intermediates of FD891, jerangolid and ambruticin. A closer look at the ambruticin and jerangolid intermediates revealed that they too share the skipped diene system of albocycline. As evident from the ambruticin and jerangolid pathways in chemical space, the skipped diene is a result of carbon excision and rearrangement. Therefore a similar carbon excision and rearrangement can be envisioned for albocycline. The methyl group at C10 might be the excised and rearranged from the main PK chain. Therefore the 'Reaction Search' module of SBSPKS was able to predict the biosynthetic origin of the unusual diene system of albocycline and hence aided in better understanding of the possible biosynthetic origin of this molecule.

Cluster search. The genomic space of SBSPKSv2 now has a new interface named 'Cluster Search', for searching ORFs in the experimentally characterized PKS/NRPS gene clusters having similarity to the query sequence and also for identifying biosynthetic reactions catalyzed by the domains/modules present in the matching ORFs (Supplementary Figure S2). This search interface uses the latest version of NCBI BLAST+ (29) at its backend and the search space of this interface includes sequences of the megasynthases as well as the tailoring enzymes in biosynthetic gene clusters (BGC) present in SBSPKSv2. It provides interlink between the chemical and the genomic space of SBSPKSv2. User can input multiple sequences to search in both genomic and chemical space and can predict the potential enzymatic reactions catalyzed by the input sequences. This interface is useful for identifying tailoring enzymes.

Updates

In the past decade, a large number of PKS and NRPS gene clusters have been identified and characterized. Resources like MIBig, IMG-ABC and antiSMASH database contain a large number of predicted secondary metabolite gene clusters (30–32). These databases are excellent resources containing a catalog of all predicted gene clusters and their domain annotations, but often it is difficult to distinguish information about experimentally characterized



Domain	Ali start	Ali end	HMM start	HMM end	e-value
KS	12	434	2	421	Expect = 2e-166
AT	541	847	2	296	Expect = 7e-70
PP	917	1005	15	104	Expect = 4e-19
KS	1016	1444	1	423	Expect = 1e-201
AT	1555	1866	2	294	Expect = 4e-69
DH	1860	2233	1	367	Expect = 5e-82
KR	2345	2615	319	587	Expect = 3e-55
PP	2709	2824	12	124	Expect = 7e-25

Figure 4. The figure depicts usage of PKS/NRPS domain search. The search identifies various catalytic domains present in PKS/NRPS gene clusters based on twenty profile HMMs. The similarity and alignment of each domain can be visualized by using the HMM alignment link. Each domain is further linked to its alignment with structural homologs and with experimentally characterized sequences.

biosynthetic gene clusters (BGC) from information which is predicted for uncharacterized BGCs. Therefore, there is a need to comprehensively annotate and store the information regarding experimentally characterized BGCs and make them easily accessible for analysis. The few databases that contain manually curated gene clusters of PKS and NRPS are DoBISCUIT and ClusterMine360 (33,34). They contain 135 and 245 gene clusters respectively, corresponding to unique compound families. But since their last update the number of characterized gene cluster has increased. Therefore to catalog the growing information comprehensively, NRPS_PKS—the genomic database of SBSPKS has been updated. NRPS_PKS now contains >300 gene clusters belonging to unique compound families (Supplementary Table S2). The database catalogs information about genes involved in the biosynthesis of PKs and NRPs, its modules and domains, specificity of acyltransferase (AT) and adenylation (A) domains and their active sites. Each domain is linked to the respective domain organization page which allows for various analyses like pairwise alignment with other characterized domains; search for nearest structural homolog, threading alignments, comparison of the active site with other characterized sequences. As a number of new 3D structures of PKS and NRPS domains have been elucidated since the last NRPS_PKS update, we have incorporated them into SBSPKSv2.

Earlier version of SBSPKS identified PKS/NRPS domains by pair wise alignment of query sequence to template sequences of various domains, and multiple template sequences were used for domains like ACP which had highly diverged sequences. Since profile based methods are more efficient for domain identification, other software like AntiSMASH, NRPSsp and NRPSpredictor (13,14,35) use Hidden Markov Models (HMMs) not only for domain identification, but also for prediction of substrate specificity of adenylation (A) domains of NRPS. We have now implemented HMM based method in SBSPKSv2 for quick and efficient domain identification. In the last few years, not only has the number of characterized gene clusters increased, but a number of new domains like product template (PT), starter unit:acyl-carrier protein transacylase (SAT), Formyl transferase (FT) have also been identified in these megasynthases (Supplementary Table S3). To detect these new domains and the canonical PKS/NRPS domains we have either developed HMM models or used HMM models from Pfam (22,36). Cut-off was determined for each domain

after extensive analysis of the characterized sequences with profile HMMs. The sensitivity, specificity and precision of all our HMM based models are >0.9 (Supplementary Table S4). As Condensation (C), Epimerization (E) and Cyclization (Cy) domains of NRPS shares high sequence similarity, we have used motif based methods to distinguish these domains. Though a number of tools exist for genome mining of PKS/NRPS gene clusters, detection of several unusual domains is exclusive to SBSPKSv2 (Supplementary Table S3). In addition to domain detection the genome mining tool of SBSPKSv2 also predicts substrate specificity, active site, closest structural homolog and experimentally characterized domain sequences (Figure 4). Updated SBSPKSv2 now uses specificity determining active site profile from 160 different A domain monomers and 15 AT domain substrates. This significantly enhances the performance of SBSPKS in predicting starter/extender substrates selected by PKS/NRPS modules in a newly identified sequence.

Since the last SBSPKS release, 3D structures of three NRPS module has been elucidated. (14,35,37). Given a NRPS module sequence, 'Model 3D-PKS/NRPS' interface of SBSPKSv2 builds its homology model using these structures as templates. SCWRL program (15) is used to build the side chain coordinates of these homology models.

Implementation

Open Babel was used to build database of biosynthetic intermediates (24). Chemaxon (http://www.chemaxon.com) was used for chemical structure drawing. The interactive pathway graphs are visualized using Cytoscape.js (25). HMM profiles were built using HMMER3 software (22). Pairwise alignments are performed using latest version of BLAST+ (29).

CONCLUSION AND FUTURE PROSPECT

An update of SBSPKS was planned due to three reasons: (i) since the last update the number of characterized PKS/NRPS gene cluster have increased, (ii) advances in high throughput technology has exponentially increased the number of orphan PKs and NRPs as well as the mega-synthases and (iii) The chemical space of PKs and NRPs' biosynthesis has not yet been curated and cataloged in any database and hence is not available for analysis. Therefore, to augment these three areas we have manually curated the chemical space of characterized PK and NRP biosynthetic pathways, developed tools to analyze and search in the chemical space, updated the genomic database of biosynthetic gene clusters and updated the genome mining tool to increase its efficiency. In summary, the new features and key improvements in SBSPKSv2 make it a comprehensive bioinformatics resource for search and analysis in the genomic as well as chemical space of polyketides and non-ribosomal peptides.

Though we have tried to create an updated and user friendly web-server, there are still some aspects which might need improvement. We are in the process of adding more number of PKS/NRPS pathways in chemical space. In the future, the download format of the pathway will be updated to XML formatted files like SBML to help user to use the pathways in simulation and modeling applications. The database of tailoring reaction will be increased so that the usability of the tool is further enhanced.

AVAILABILITY

http://www.nii.ac.in/sbspks2.html. This website is free and open to all users and there is no login requirement.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Department of Biotechnology, Government of India grant to National Institute of Immunology, New Delhi; Department of Biotechnology, India [BTIS (BT/BI/03/009/2002), COE (BT/COE/34/SP15138/2015) to D.M.]; Council of Scientific & Industrial Research, India (to N.S., M.G.). Funding for open access charge: NII, New Delhi (to D.M.).
Conflict of interest statement. None declared.

REFERENCES

1. Cane,D.E. and Walsh,C.T. (1999) The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. *Chem. Biol.*, **6**, R319–R325.
2. Nikolouli,K. and Mossialos,D. (2012) Bioactive compounds synthesized by non-ribosomal peptide synthetases and type-I polyketide synthases discovered through genome-mining and metagenomics. *Biotechnol. Lett.*, **34**, 1393–1403.
3. Winn,M., Fyans,J.K., Zhuo,Y. and Micklefield,J. (2016) Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.*, **33**, 317–347.
4. Cane,D.E., Walsh,C.T. and Khosla,C. (1998) Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science*, **282**, 63–68.
5. Marsden,A.F., Wilkinson,B., Cortes,J., Dunster,N.J., Staunton,J. and Leadlay,P.F. (1998) Engineering broader specificity into an antibiotic-producing polyketide synthase. *Science*, **279**, 199–202.
6. Meier,J.L. and Burkart,M.D. (2009) The chemical biology of modular biosynthetic enzymes. *Chem. Soc. Rev.*, **38**, 2012–2045.
7. Anand,S., Prasad,M.V., Yadav,G., Kumar,N., Shehara,J., Ansari,M.Z. and Mohanty,D. (2010) SBSPKS: structure based sequence analysis of polyketide synthases. *Nucleic Acids Res.*, **38**, W487–W496.
8. Rebets,Y., Tokovenko,B., Lushchyk,I., Ruckert,C., Zaburannyi,N., Bechthold,A., Kalinowski,J. and Luzhetskyy,A. (2014) Complete genome sequence of producer of the glycopeptide antibiotic Aculeximycin Kutzneria albida DSM 43870(T), a representative of minor genus of Pseudonocardiaceae. *Bmc Genomics*, **15**, 885.
9. Midha,S. and Patil,P.B. (2014) Genomic insights into the evolutionary origin of Xanthomonas axonopodis pv. citri and its ecological relatives. *Appl. Environ. Microb.*, **80**, 6266–6279.
10. Bhetariya,P.J., Prajapati,M., Bhaduri,A., Mandal,R.S., Varma,A., Madan,T., Singh,Y. and Sarma,P.U. (2016) Phylogenetic and structural analysis of polyketide synthases in Aspergilli. *Evol. Bioinform.*, **12**, 109–119.
11. Esmaeel,Q., Pupin,M., Kieu,N.P., Chataigné,G., Béchet,M., Deravel,J., Krier,F., Höfte,M., Jacques,P. and Leclère,V. (2016) Burkholderia genome mining for nonribosomal peptide synthetases reveals a great potential for novel siderophores and lipopeptides synthesis. *Microbiologyopen*, **5**, 512–526.
12. 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.

13. Amoutzias,G.D., Chaliotis,A. and Mossialos,D. (2016) Discovery strategies of bioactive compounds synthesized by nonribosomal peptide synthetases and type-I polyketide synthases derived from marine microbiomes. *Mar. Drugs*, **14**, 80.
14. Reimer,J.M., Aloise,M.N., Harrison,P.M. and Schmeing,T.M. (2016) Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. *Nature*, **529**, U239–U305.
15. Canutescu,A.A., Shelenkov,A.A. and Dunbrack,R.L. (2003) A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci.*, **12**, 2001–2014.
16. Medema,M.H. and Fischbach,M.A. (2015) Computational approaches to natural product discovery. *Nat. Chem. Biol.*, **11**, 639–648.
17. Walsh,C.T. and Fischbach,M.A. (2010) Natural products version 2.0: connecting genes to molecules. *J. Am. Chem. Soc.*, **132**, 2469–2493.
18. Kim,S., Thiessen,P.A., Bolton,E.E., Chen,J., Fu,G., Gindulyte,A., Han,L., He,J., He,S., Shoemaker,B.A. *et al.* (2016) PubChem substance and compound databases. *Nucleic Acids Res.*, **44**, D1202–D1213.
19. Kanehisa,M. (2002) The KEGG database. *Novartis Found. Symp.*, **247**, 91–101.
20. Dejong,C.A., Chen,G.M., Li,H., Johnston,C.W., Edwards,M.R., Rees,P.N., Skinnider,M.A., Webster,A.L. and Magarvey,N.A. (2016) Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. *Nat. Chem. Biol.*, **12**, 1007–1014.
21. Khater,S., Anand,S. and Mohanty,D. (2016) In silico methods for linking genes and secondary metabolites: the way forward. *Synth. Syst. Biotechnol.*, **1**, 80–88.
22. Eddy,S.R. (2011) Accelerated profile HMM searches. *PLoS Comput. Biol.*, **7**, e1002195.
23. Ihlenfeldt,W.D., Bolton,E.E. and Bryant,S.H. (2009) The PubChem chemical structure sketcher. *J. Cheminformatics*, **1**, 20.
24. O'Boyle,N.M., Banck,M., James,C.A., Morley,C., Vandermeersch,T. and Hutchison,G.R. (2011) Open babel: an open chemical toolbox. *J. Cheminformatics*, **3**, 33.
25. Franz,M., Lopes,C.T., Huck,G., Dong,Y., Sumer,O. and Bader,G.D. (2016) Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics*, **32**, 309–311.
26. Koyama,N., Yotsumoto,M., Onaka,H. and Tomoda,H. (2013) New structural scaffold 14-membered macrocyclic lactone ring for selective inhibitors of cell wall peptidoglycan biosynthesis in Staphylococcus aureus. *J. Antibiotics*, **66**, 303–304.
27. Nagahama,N., Suzuki,M., Awataguchi,S. and Okuda,T. (1967) Studies on a new antibiotic, albocycline. I. Isolation, purification and properties. *J. Antibiotics*, **20**, 261–266.
28. O'Brien,R.V., Davis,R.W., Khosla,C. and Hillenmeyer,M.E. (2014) Computational identification and analysis of orphan assembly-line polyketide synthases. *J. Antibiotics*, **67**, 89–97.
29. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
30. Blin,K., Medema,M.H., Kottmann,R., Lee,S.Y. and Weber,T. (2017) The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res.*, **45**, D555–D559.
31. Hadjithomas,M., Chen,I.M., Chu,K., Ratner,A., Palaniappan,K., Szeto,E., Huang,J., Reddy,T.B., Cimermancic,P., Fischbach,M.A. *et al.* (2015) IMG-ABC: a knowledge base to fuel discovery of biosynthetic gene clusters and novel secondary metabolites. *mBio*, **6**, e00932.
32. Li,Y.F., Tsai,K.J., Harvey,C.J., Li,J.J., Ary,B.E., Berlew,E.E., Boehman,B.L., Findley,D.M., Friant,A.G., Gardner,C.A. *et al.* (2016) Comprehensive curation and analysis of fungal biosynthetic gene clusters of published natural products. *Fungal Genet. Biol.: FG & B*, **89**, 18–28.
33. Conway,K.R. and Boddy,C.N. (2013) ClusterMine360: a database of microbial PKS/NRPS biosynthesis. *Nucleic Acids Res.*, **41**, D402–D407.
34. Ichikawa,N., Sasagawa,M., Yamamoto,M., Komaki,H., Yoshida,Y., Yamazaki,S. and Fujita,N. (2013) DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res.*, **41**, D408–D414.
35. Miller,B.R., Drake,E.J., Shi,C., Aldrich,C.C. and Gulick,A.M. (2016) Structures of a nonribosomal peptide synthetase module bound to MbtH-like proteins support a highly dynamic domain architecture. *J. Biol. Chem.*, **291**, 22559–22571.
36. Finn,R.D., Coggill,P., Eberhardt,R.Y., Eddy,S.R., Mistry,J., Mitchell,A.L., Potter,S.C., Punta,M., Qureshi,M., Sangrador-Vegas,A. *et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.*, **44**, D279–D285.
37. Drake,E.J., Miller,B.R., Shi,C., Tarrasch,J.T., Sundlov,J.A., Allen,C.L., Skiniotis,G., Aldrich,C.C. and Gulick,A.M. (2016) Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature*, **529**, U235–U289.