Supplementary Information

RadicalSAM.org: A Resource to Interpret Sequence-Function Space and Discover New Radical SAM Enzyme Chemistry

Nils Oberg[1], Timothy W. Precord[1, 2], Douglas A. Mitchell[1-3*], and John A. Gerlt[1, 2, 4*]

[1]Carl R. Woese Institute for Genomic Biology, [2]Department of Chemistry, [3]Department of Microbiology, and [4]Department of Biochemistry, University of Illinois at Urbana-Champaign, 1206 West Gregory Drive, Urbana, Illinois 61801, United States.

Table S1. The list of Pfam families and InterPro families/domains used to identify members of the RSS in UniProt Release 2020_05 and InterPro Release 82 in the release of RadicalSAM.org described in this Perspective is provided on the "Sequence Families" subtab of the "Current Release" tab on Home page of RadicalSAM.org. In recent unpublished updates of the RSS, the Pfam families and InterPro families/domains in this table were added to the list to provide a more complete inventory of the members of the RSS.

Identifier	**Short Name**	**Identifier**	**Short Name**
IPR017672	MA_4551-like	IPR027626	Pseudo_SAM_Halo
IPR023805	Uncharacterised_Spl-rel	IPR030894	Ahb_Proteobacteria
IPR023807	Peptide_mod_rSAM	IPR030950	rSAM_PoyD
IPR023820	rSAM_GDL-assoc	IPR031004	rSAM_YfkAB
IPR023858	RSAM_HmdB	IPR031010	rSAM_mob_pairA
IPR023862	CHP03960_rSAM	IPR031014	rSAM_BlsE
IPR023863	rSAM_PTO1314	IPR031019	rSAM_vs_C_rich
IPR023867	Sulphatase_maturase_rSAM	IPR034519	TunB-like
IPR023874	DNA_rSAM_put	IPR034532	OxsB-like
IPR023984	rSAM_ocin_1	IPR040087	MJ0021-like
IPR024021	FeFe-hyd_HydE_rSAM	IPR040088	MJ0103-like
IPR024032	rSAM_paired_HxsC	IPR041582	RimO_TRAM
IPR026351	rSAM_SeCys	PF04055	Radical_SAM
IPR026401	CXXX_matur	PF08497	Radical_SAM_N
IPR027559	B12_rSAM_oligo	PF12345	DUF3641
IPR027622	rSAM_Clo7bot		

Expanded images for seven text figures.

Figure S2. The SSN generated with a maximum e-value edge threshold of 1e-20 used by the SFLD to identify its 20 functionally characterized subgroups (colored/numbered clusters) and 22 uncharacterized subgroups. Large nodes represent experimentally characterized proteins: downward arrows indicate a structurally characterized protein; diamonds indicate no structural characterization. Reproduced with permission from reference 9; http://www.elsevier.com.

Figure S7. A representative Explore page: Megacluster-3-1, 7-carboxy-7-deazaguanine synthase-like, SFLD subgroup 1.

Figure S9. A representative GND Viewer page: Cluster-6 in the "diced" SSN for Megacluster-1-1 generated with a minimum edge alignment score threshold of 60.

Figure S12A. Navigation through diced megaclusters. "Previous" and "Next" navigation buttons (green arrows) direct the cluster selection backward and forward, respectively, in the selected diced SSN.

Figure S12B. Navigation through diced megaclusters. "Cluster" drop down menu allows the user to select any cluster in the currently viewed diced SSN.

Figure S12C. Navigation through diced megaclusters. Alignment score drop down menu allows the user to view the cluster in the selected diced SSN.

Figure S13. AS Walk-Through pop-up window showing identity of the progenitor cluster ("Previous Cluster) and progeny clusters ("Next Clusters").

Figure S14. Search functions: "Find by UniProt ID" identifies the cluster(s) containing the user-specified accession ID; "Find by Sequence" identifies the cluster(s) with the best HMM match to the user-specified sequence; "GND Lookup" provides the GND(s) for the user-specified UniProt

ID(s); “Find by Taxonomy” provides a list of the UniProt accession IDs and their clusters for the user-specified genus/species.

Figure S15. The Submit page for community submission of enzymatic activities and metabolic functions.

Figure S16. Contact page form for submitting feedback or questions.

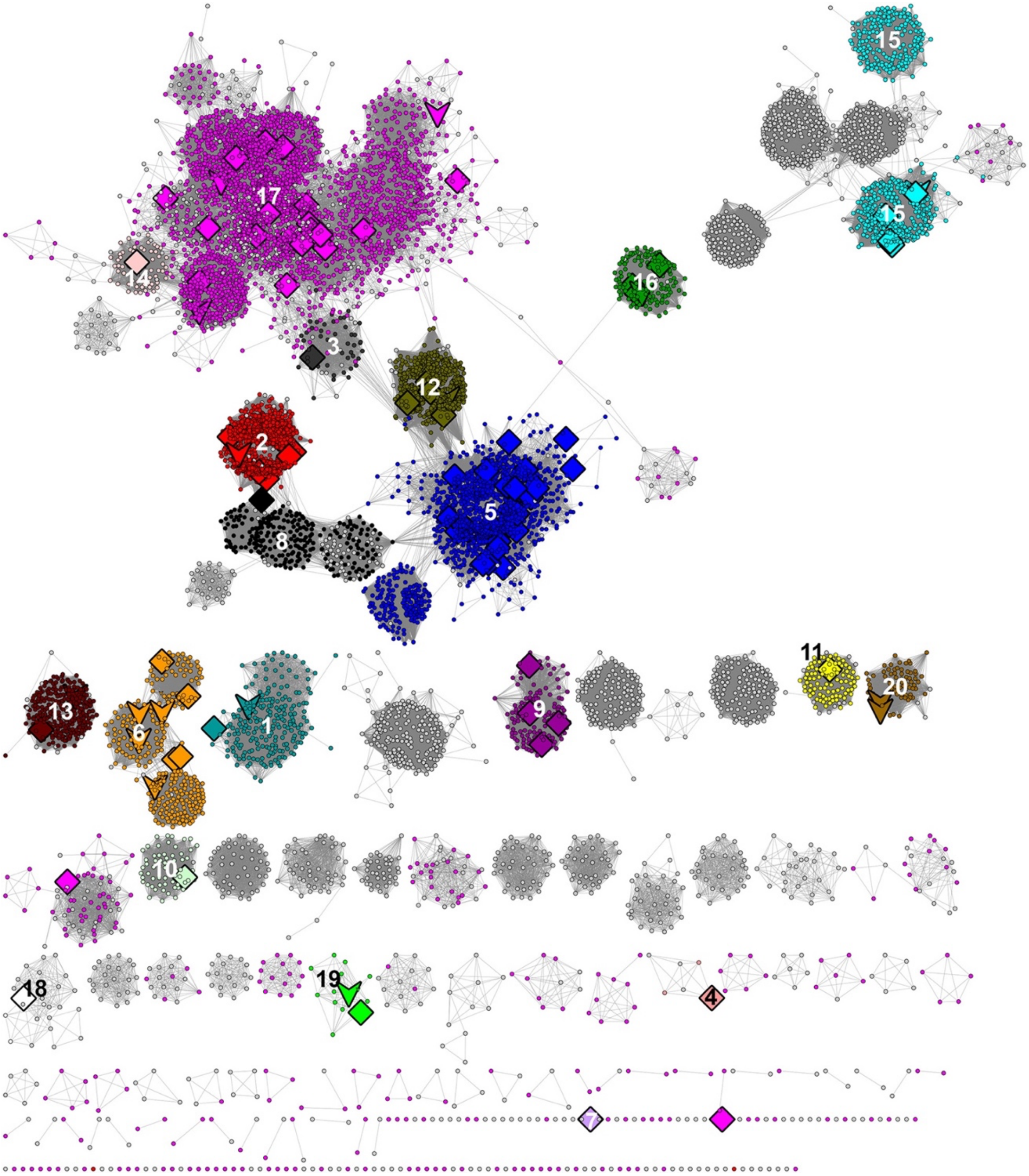
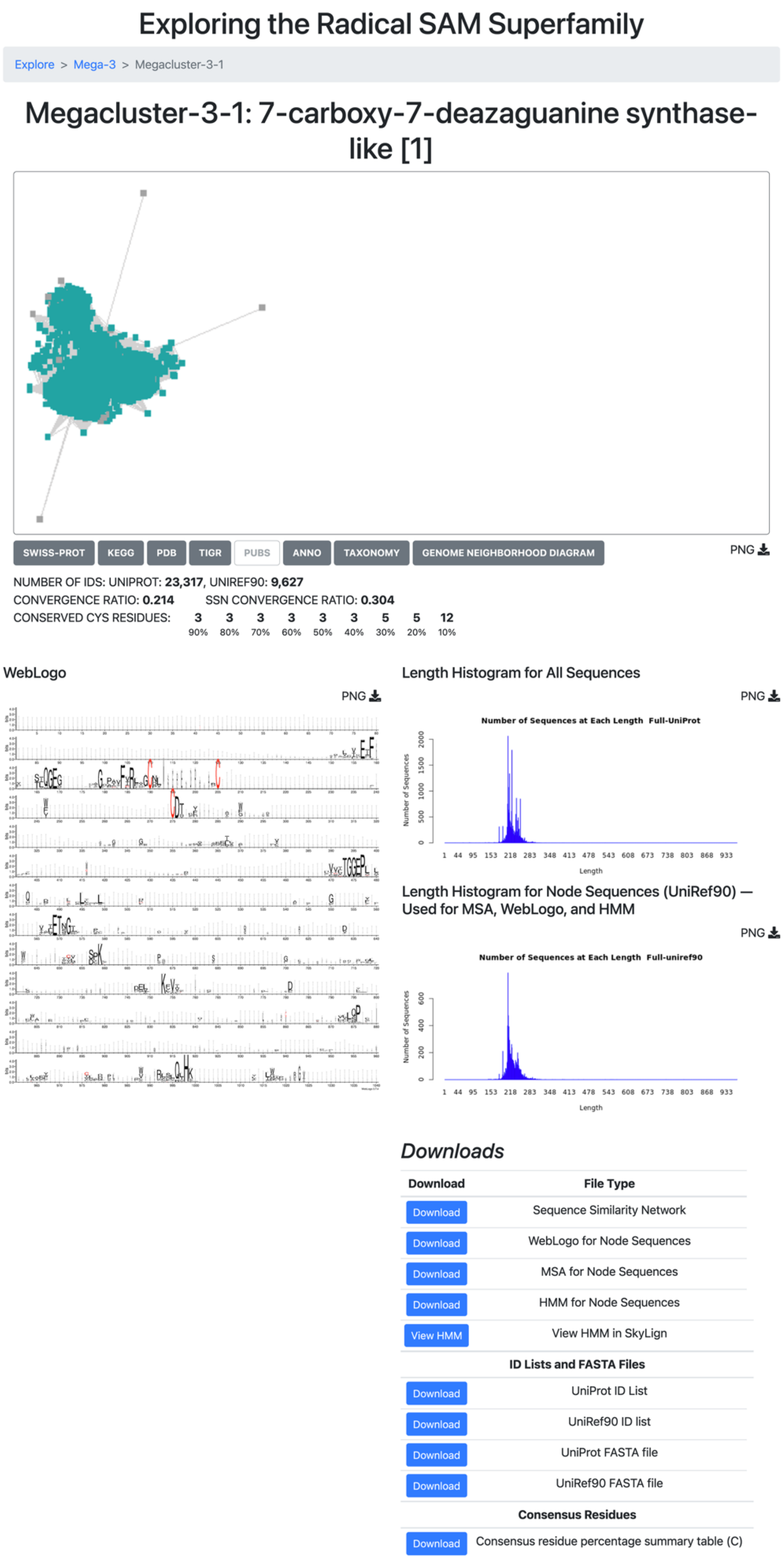


Figure 2. The SSN generated with a maximum e-value edge threshold of 1e-20 used by the SFLD to identify its 20 functionally characterized subgroups (colored/numbered clusters) and 22 uncharacterized subgroups. Large nodes represent experimentally characterized proteins: downward arrows indicate a structurally characterized protein; diamonds indicate no structural characterization. Reproduced with permission from reference 9. Copyright 2018 Elsevier.



Exploring the Radical SAM Superfamily

Explore > Mega-3 > Megacluster-3-1

Megacluster-3-1: 7-carboxy-7-deazaguanine synthase-like [1]

NUMBER OF IDS: UNIPROT: **23,317**, UNIREF90: **9,627**

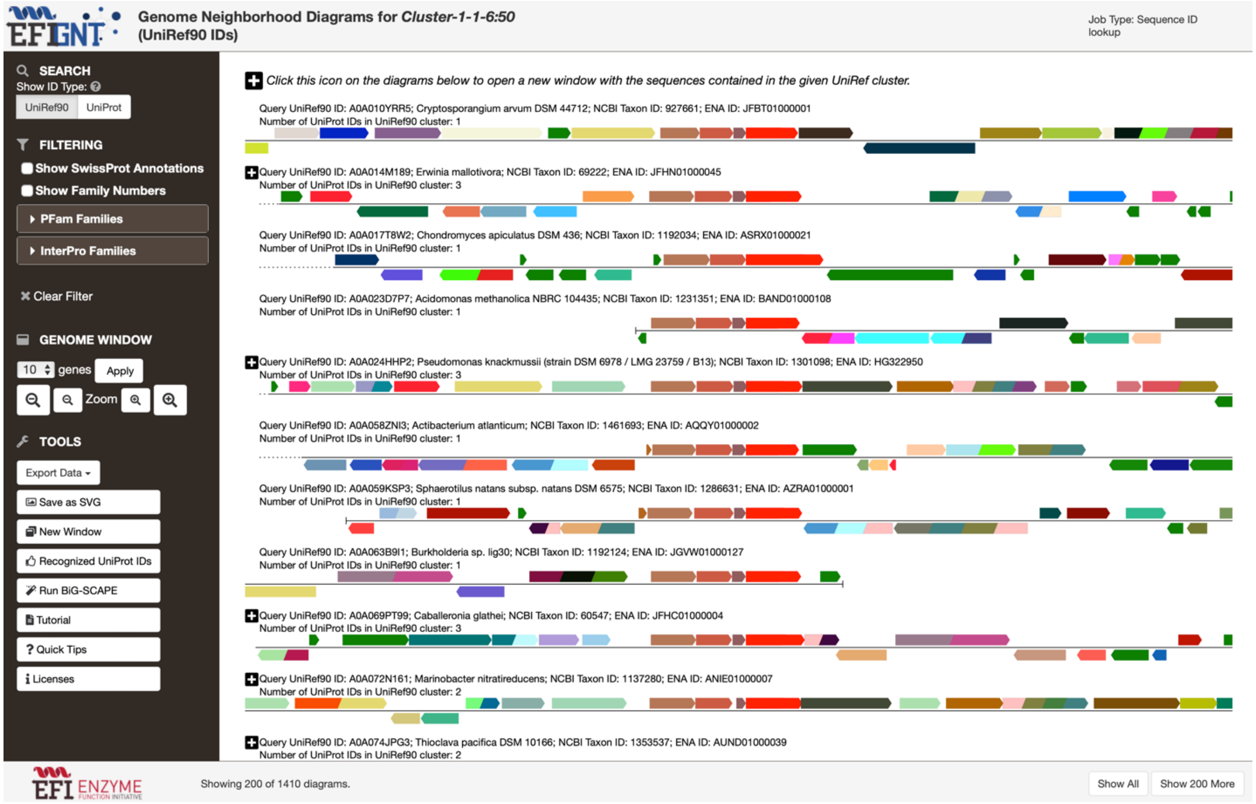
CONVERGENCE RATIO: **0.214** SSN CONVERGENCE RATIO: **0.304**

CONSERVED CYS RESIDUES:	3	3	3	3	3	3	5	5	12
	90%	80%	70%	60%	50%	40%	30%	20%	10%

Downloads

Download	File Type
Download	Sequence Similarity Network
Download	WebLogo for Node Sequences
Download	MSA for Node Sequences
Download	HMM for Node Sequences
View HMM	View HMM in SkyLign
ID Lists and FASTA Files	
Download	UniProt ID List
Download	UniRef90 ID list
Download	UniProt FASTA file
Download	UniRef90 FASTA file
Consensus Residues	
Download	Consensus residue percentage summary table (C)

Figure 7. Example of an Explore page. Shown is Megacluster-3-1: 7-carboxy-7-deazaguanine synthase-like (SFLD subgroup 1).



Click this icon on the diagrams below to open a new window with the sequences contained in the given UniRef cluster.

Query UniRef90 ID: A0A010YRR5; Cryptosporangium arvum DSM 44712; NCBI Taxon ID: 927661; ENA ID: JFBT01000001
Number of UniProt IDs in UniRef90 cluster: 1

Query UniRef90 ID: A0A014M189; Erwinia mallotivora; NCBI Taxon ID: 69222; ENA ID: JFHN01000045
Number of UniProt IDs in UniRef90 cluster: 3

Query UniRef90 ID: A0A017T8W2; Chondromyces apiculatus DSM 436; NCBI Taxon ID: 1192034; ENA ID: ASRX01000021
Number of UniProt IDs in UniRef90 cluster: 1

Query UniRef90 ID: A0A023D7P7; Acidomonas methanolica NBRC 104435; NCBI Taxon ID: 1231351; ENA ID: BAND01000108
Number of UniProt IDs in UniRef90 cluster: 1

Query UniRef90 ID: A0A024HHP2; Pseudomonas knackmussii (strain DSM 6978 / LMG 23759 / B13); NCBI Taxon ID: 1301098; ENA ID: HG322950
Number of UniProt IDs in UniRef90 cluster: 3

Query UniRef90 ID: A0A058ZNI3; Actibacterium atlanticum; NCBI Taxon ID: 1461693; ENA ID: AQQY01000002
Number of UniProt IDs in UniRef90 cluster: 1

Query UniRef90 ID: A0A059KSP3; Sphaerotilus natans subsp. natans DSM 6575; NCBI Taxon ID: 1286631; ENA ID: AZRA01000001
Number of UniProt IDs in UniRef90 cluster: 1

Query UniRef90 ID: A0A063B9I1; Burkholderia sp. lig30; NCBI Taxon ID: 1192124; ENA ID: JGVW01000127
Number of UniProt IDs in UniRef90 cluster: 1

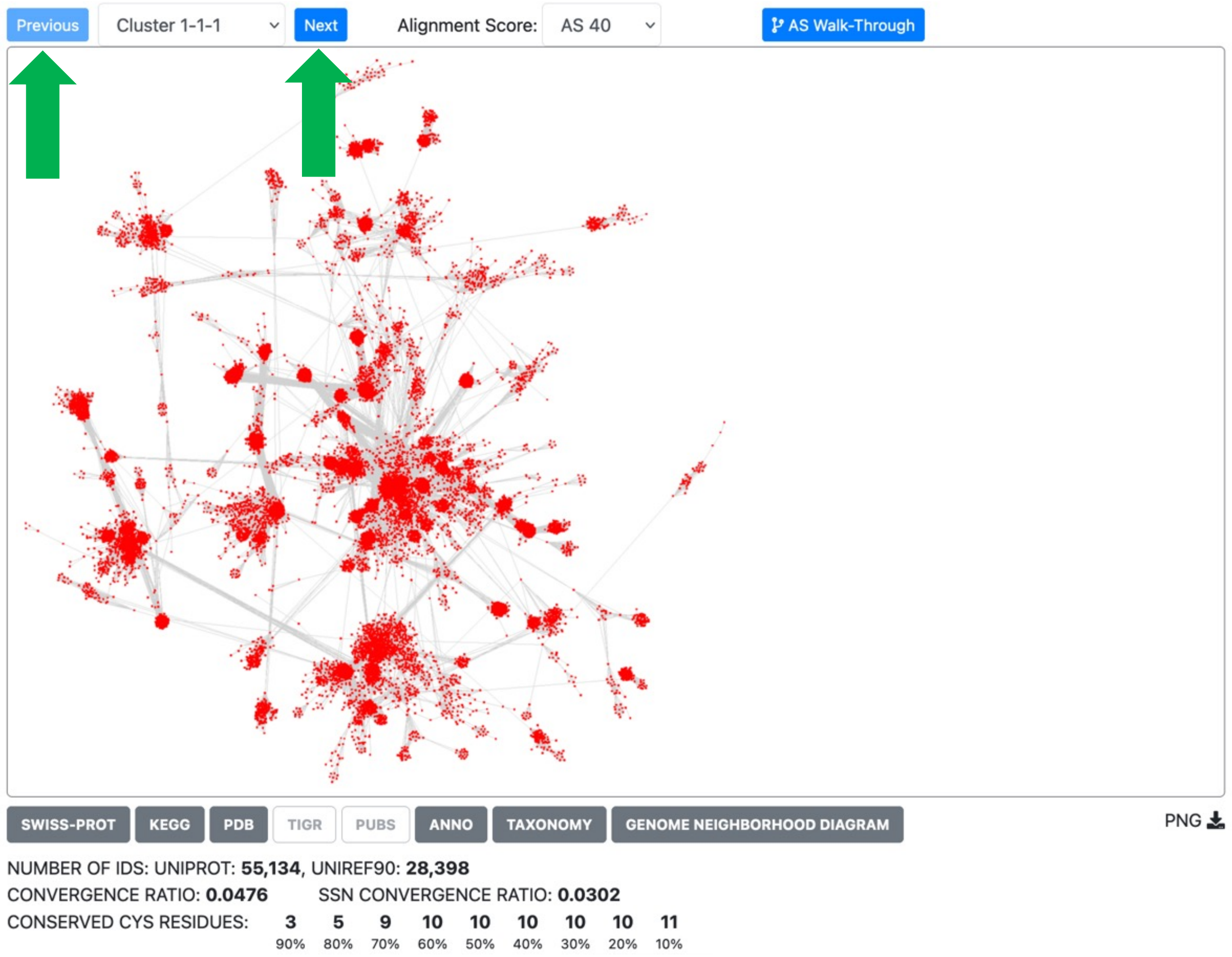
Query UniRef90 ID: A0A069PT99; Caballeronia glathei; NCBI Taxon ID: 60547; ENA ID: JFHC01000004
Number of UniProt IDs in UniRef90 cluster: 3

Query UniRef90 ID: A0A072N161; Marinobacter nitratireducens; NCBI Taxon ID: 1137280; ENA ID: ANIE01000007
Number of UniProt IDs in UniRef90 cluster: 2

Query UniRef90 ID: A0A074JPG3; Thioclava pacifica DSM 10166; NCBI Taxon ID: 1353537; ENA ID: AUND01000039
Number of UniProt IDs in UniRef90 cluster: 2

Showing 200 of 1410 diagrams.

Figure 9. A representative GND Viewer page: Cluster-6 in the “diced” SSN for Megacluster-1-1 generated using an alignment score of 60.



Cluster 1-1-1

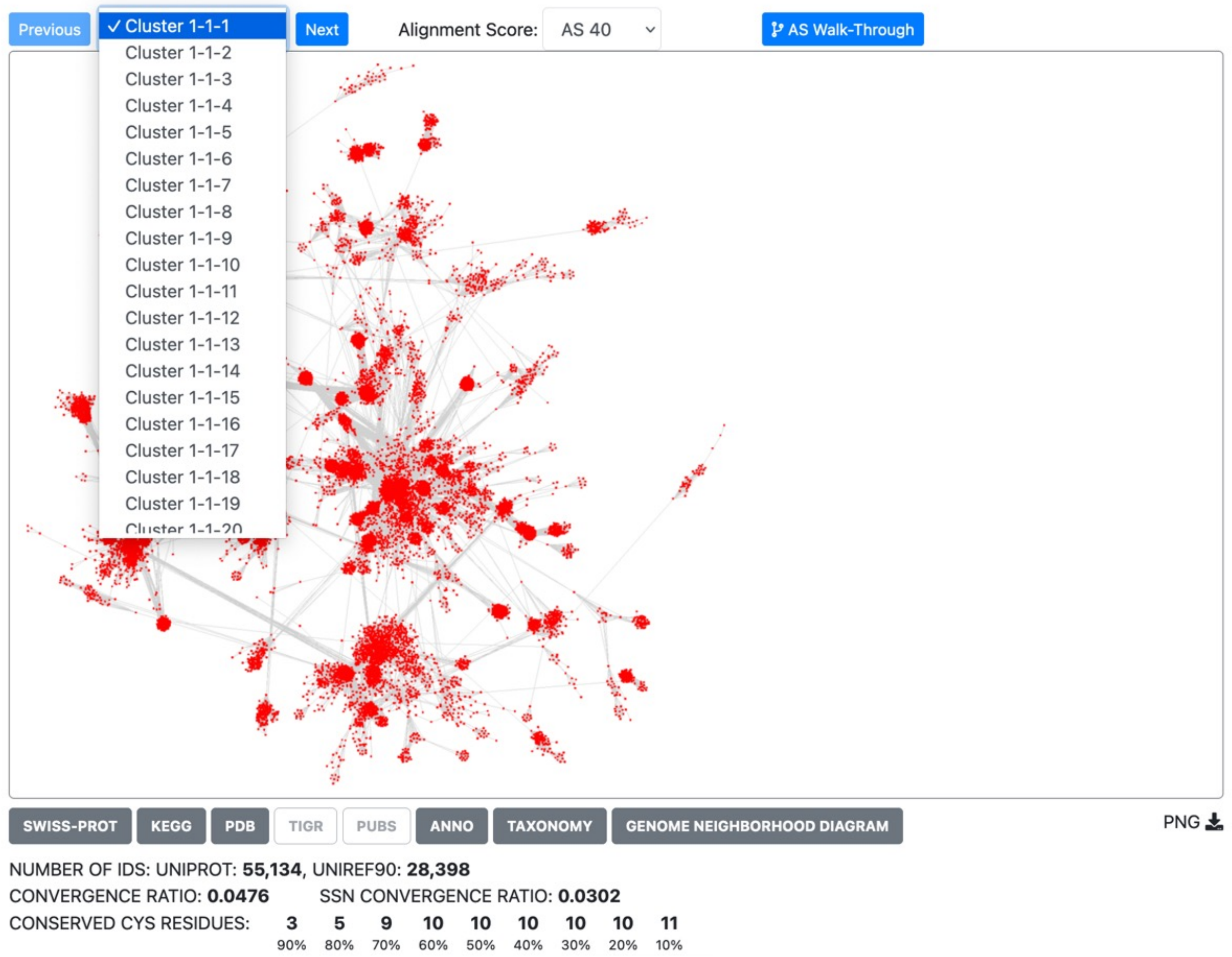
Alignment Score: AS 40

NUMBER OF IDS: UNIPROT: **55,134**, UNIREF90: **28,398**

CONVERSION RATIO: **0.0476** SSN CONVERGENCE RATIO: **0.0302**

CONSERVED CYS RESIDUES:	**3**	**5**	**9**	**10**	**10**	**10**	**10**	**10**	**11**
	90%	80%	70%	60%	50%	40%	30%	20%	10%

Figure 12A. Navigation through diced megaclusters. "Previous" and "Next" navigation buttons (green arrows) direct the cluster selection backward and forward, respectively, in the selected diced SSN.



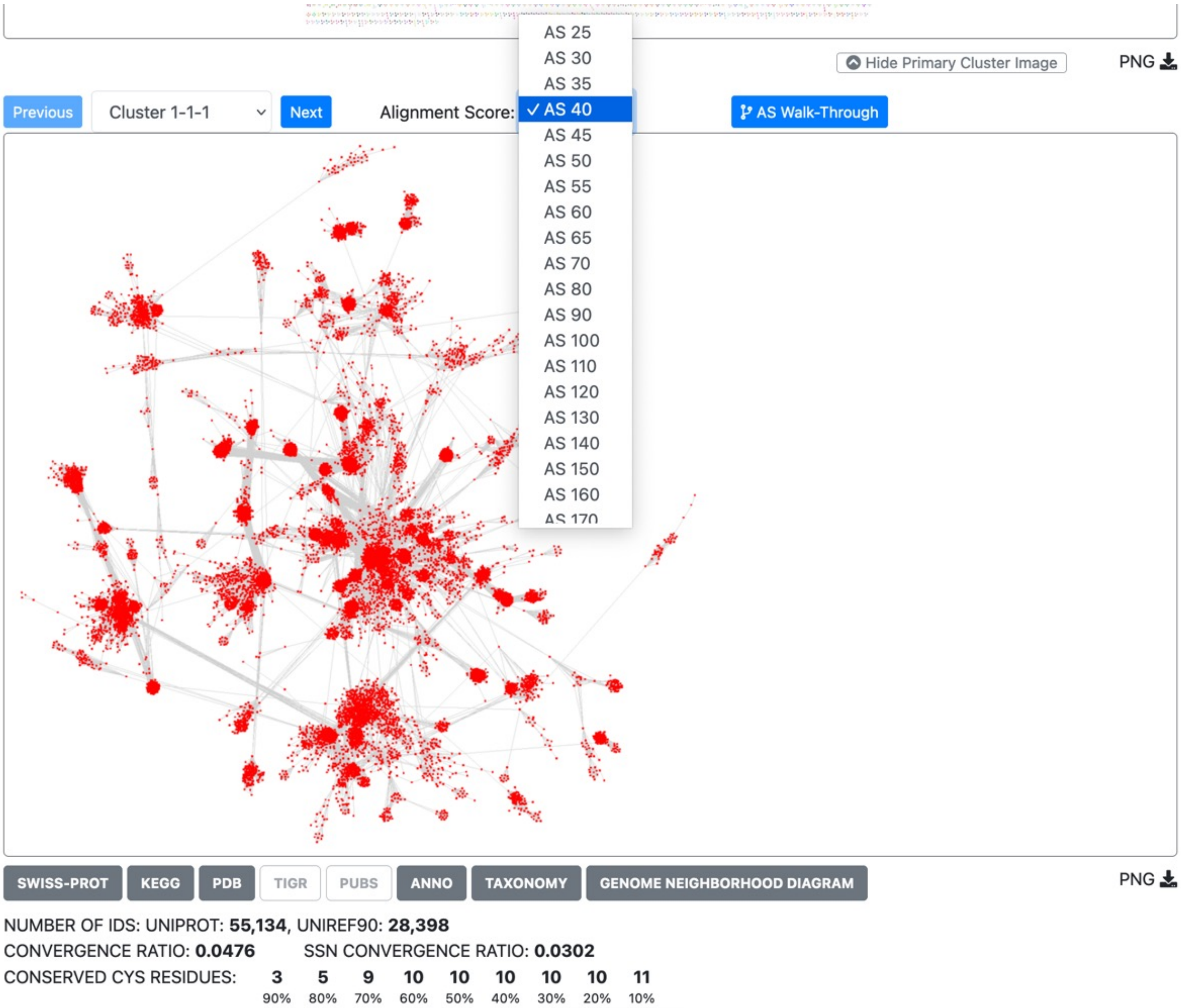
Alignment Score: AS 40

NUMBER OF IDS: UNIPROT: **55,134**, UNIREF90: **28,398**

CONVERGENCE RATIO: **0.0476** SSN CONVERGENCE RATIO: **0.0302**

CONSERVED CYS RESIDUES:	90%	80%	70%	60%	50%	40%	30%	20%	10%
	3	**5**	**9**	**10**	**10**	**10**	**10**	**10**	**11**

Figure 12B. Navigation through diced megaclusters. “Cluster” drop down menu allows the user to select any cluster in the currently viewed diced SSN.

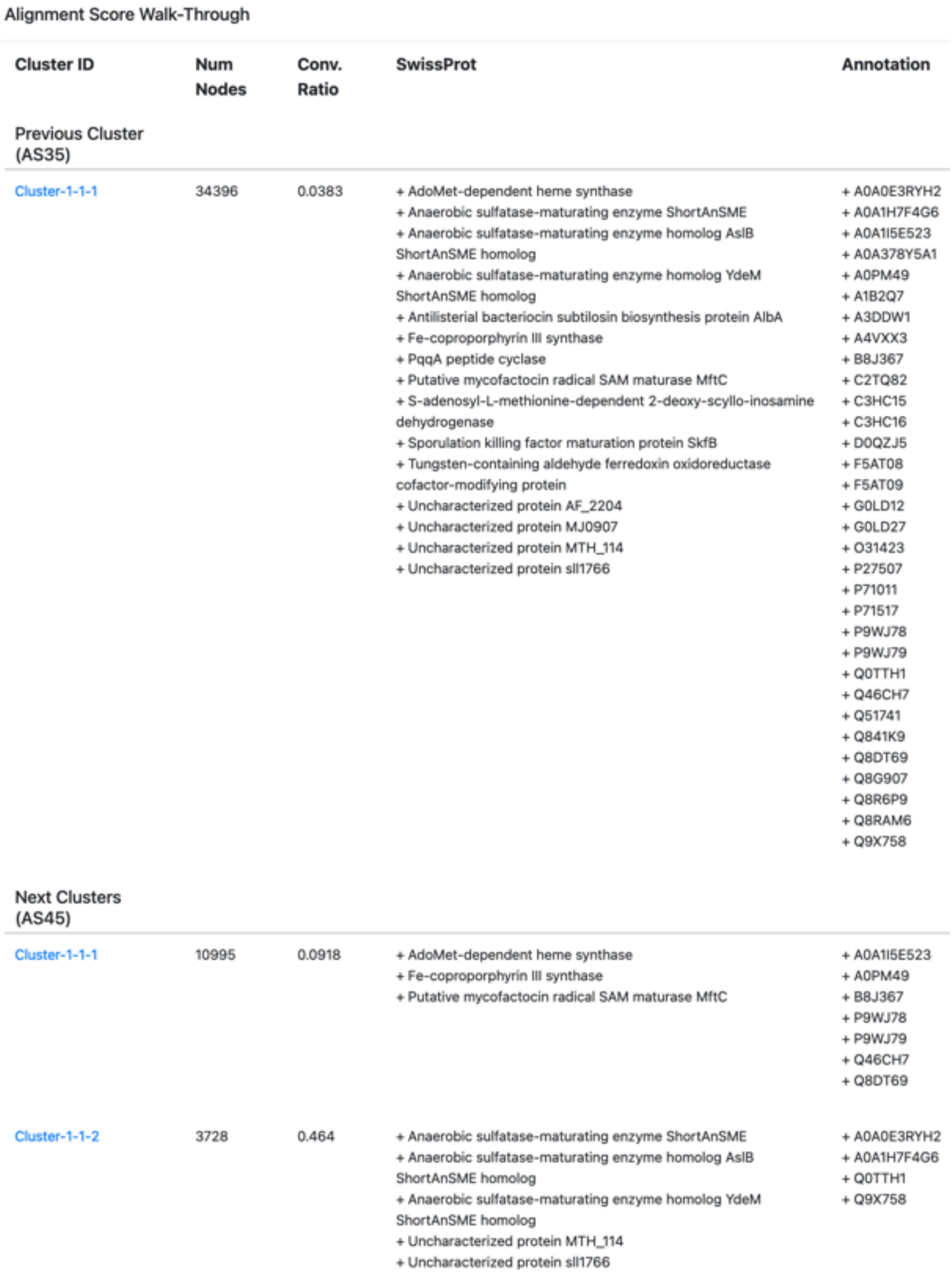


NUMBER OF IDS: UNIPROT: **55,134**, UNIREF90: **28,398**

CONVERGENCE RATIO: **0.0476** SSN CONVERGENCE RATIO: **0.0302**

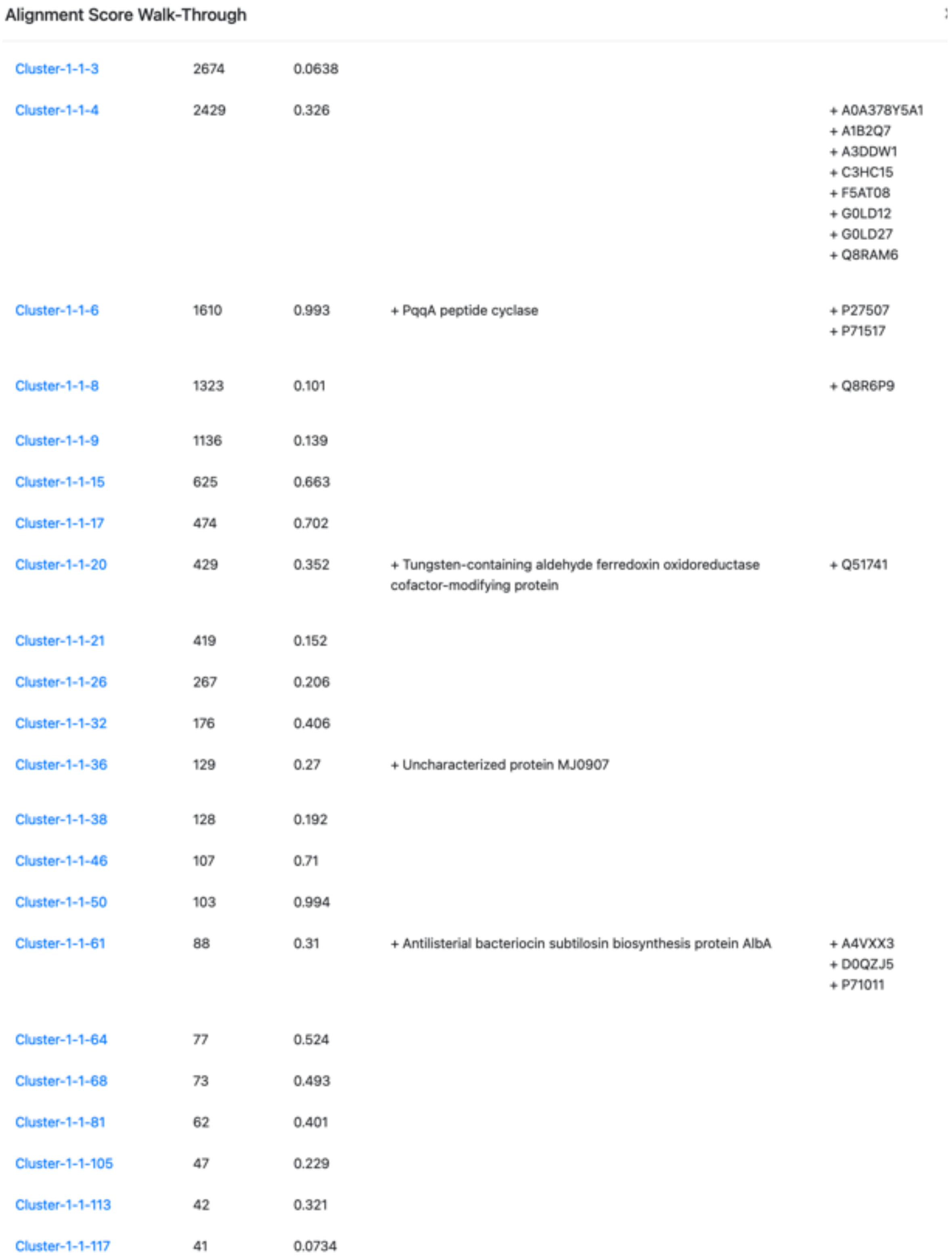
CONSERVED CYS RESIDUES:	**3**	**5**	**9**	**10**	**10**	**10**	**10**	**10**	**11**
	90%	80%	70%	60%	50%	40%	30%	20%	10%

Figure 12C. Navigation through diced megaclusters. Alignment score drop down menu allows the user to view the cluster in the selected diced SSN.



Alignment Score Walk-Through

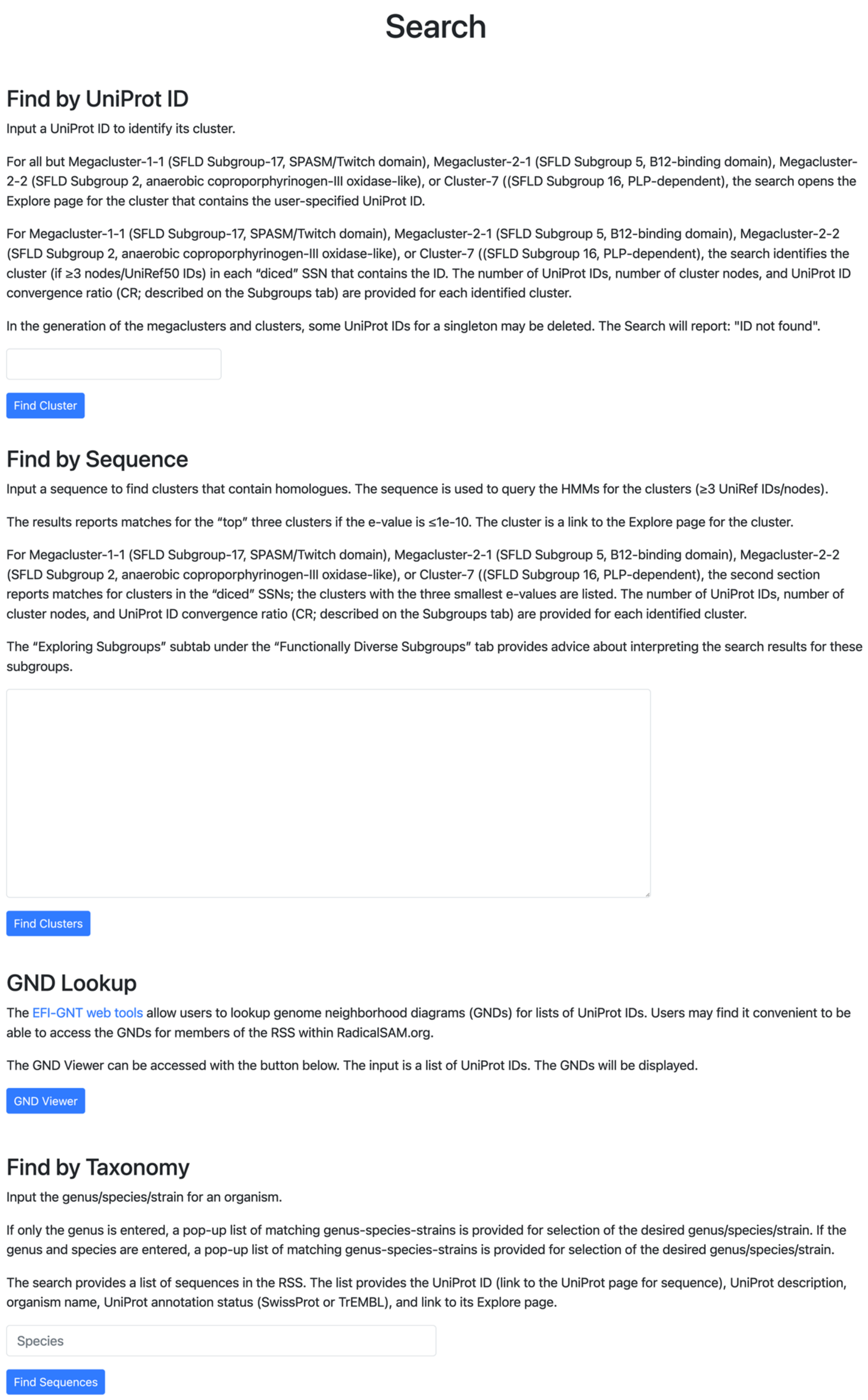
Cluster ID	Num Nodes	Conv. Ratio	SwissProt	Annotation
Previous Cluster (AS35)				
Cluster-1-1-1	34396	0.0383	+ AdoMet-dependent heme synthase + Anaerobic sulfatase-maturating enzyme ShortAnSME + Anaerobic sulfatase-maturating enzyme homolog AslB ShortAnSME homolog + Anaerobic sulfatase-maturating enzyme homolog YdeM ShortAnSME homolog + Antilisterial bacteriocin subtilosin biosynthesis protein AlbA + Fe-coproporphyrin III synthase + PqqA peptide cyclase + Putative mycofactocin radical SAM maturase MftC + S-adenosyl-L-methionine-dependent 2-deoxy-scyllo-inosamine dehydrogenase + Sporulation killing factor maturation protein SkfB + Tungsten-containing aldehyde ferredoxin oxidoreductase cofactor-modifying protein + Uncharacterized protein AF_2204 + Uncharacterized protein MJ0907 + Uncharacterized protein MTH_114 + Uncharacterized protein sll1766	+ A0A0E3RYH2 + A0A1H7F4G6 + A0A1I5E523 + A0A378Y5A1 + A0PM49 + A1B2Q7 + A3DDW1 + A4VXX3 + B8J367 + C2TQ82 + C3HC15 + C3HC16 + D0QZJ5 + F5AT08 + F5AT09 + G0LD12 + G0LD27 + O31423 + P27507 + P71011 + P71517 + P9WJ78 + P9WJ79 + Q0TTH1 + Q46CH7 + Q51741 + Q841K9 + Q8DT69 + Q8G907 + Q8R6P9 + Q8RAM6 + Q9X758
Next Clusters (AS45)				
Cluster-1-1-1	10995	0.0918	+ AdoMet-dependent heme synthase + Fe-coproporphyrin III synthase + Putative mycofactocin radical SAM maturase MftC	+ A0A1I5E523 + A0PM49 + B8J367 + P9WJ78 + P9WJ79 + Q46CH7 + Q8DT69
Cluster-1-1-2	3728	0.464	+ Anaerobic sulfatase-maturating enzyme ShortAnSME + Anaerobic sulfatase-maturating enzyme homolog AslB ShortAnSME homolog + Anaerobic sulfatase-maturating enzyme homolog YdeM ShortAnSME homolog + Uncharacterized protein MTH_114 + Uncharacterized protein sll1766	+ A0A0E3RYH2 + A0A1H7F4G6 + Q0TTH1 + Q9X758



Alignment Score Walk-Through

Cluster-1-1-3	2674	0.0638		
Cluster-1-1-4	2429	0.326		+ A0A378Y5A1 + A1B2Q7 + A3DDW1 + C3HC15 + F5AT08 + G0LD12 + G0LD27 + Q8RAM6
Cluster-1-1-6	1610	0.993	+ PqqA peptide cyclase	+ P27507 + P71517
Cluster-1-1-8	1323	0.101		+ Q8R6P9
Cluster-1-1-9	1136	0.139		
Cluster-1-1-15	625	0.663		
Cluster-1-1-17	474	0.702		
Cluster-1-1-20	429	0.352	+ Tungsten-containing aldehyde ferredoxin oxidoreductase cofactor-modifying protein	+ Q51741
Cluster-1-1-21	419	0.152		
Cluster-1-1-26	267	0.206		
Cluster-1-1-32	176	0.406		
Cluster-1-1-36	129	0.27	+ Uncharacterized protein MJ0907	
Cluster-1-1-38	128	0.192		
Cluster-1-1-46	107	0.71		
Cluster-1-1-50	103	0.994		
Cluster-1-1-61	88	0.31	+ Antilisterial bacteriocin subtilosin biosynthesis protein AlbA	+ A4VXX3 + D0QZJ5 + P71011
Cluster-1-1-64	77	0.524		
Cluster-1-1-68	73	0.493		
Cluster-1-1-81	62	0.401		
Cluster-1-1-105	47	0.229		
Cluster-1-1-113	42	0.321		
Cluster-1-1-117	41	0.0734		

Figure 13. AS Walk-Through pop-up window showing identity of the progenitor cluster (“Previous Cluster) and progeny clusters (“Next Clusters”).



Search

Find by UniProt ID

Input a UniProt ID to identify its cluster.

For all but Megacluster-1-1 (SFLD Subgroup-17, SPASM/Twitch domain), Megacluster-2-1 (SFLD Subgroup 5, B12-binding domain), Megacluster-2-2 (SFLD Subgroup 2, anaerobic coproporphyrinogen-III oxidase-like), or Cluster-7 ((SFLD Subgroup 16, PLP-dependent), the search opens the Explore page for the cluster that contains the user-specified UniProt ID.

For Megacluster-1-1 (SFLD Subgroup-17, SPASM/Twitch domain), Megacluster-2-1 (SFLD Subgroup 5, B12-binding domain), Megacluster-2-2 (SFLD Subgroup 2, anaerobic coproporphyrinogen-III oxidase-like), or Cluster-7 ((SFLD Subgroup 16, PLP-dependent), the search identifies the cluster (if ≥3 nodes/UniRef50 IDs) in each "diced" SSN that contains the ID. The number of UniProt IDs, number of cluster nodes, and UniProt ID convergence ratio (CR; described on the Subgroups tab) are provided for each identified cluster.

In the generation of the megaclusters and clusters, some UniProt IDs for a singleton may be deleted. The Search will report: "ID not found".

Find by Sequence

Input a sequence to find clusters that contain homologues. The sequence is used to query the HMMs for the clusters (≥3 UniRef IDs/nodes).

The results reports matches for the "top" three clusters if the e-value is ≤1e-10. The cluster is a link to the Explore page for the cluster.

For Megacluster-1-1 (SFLD Subgroup-17, SPASM/Twitch domain), Megacluster-2-1 (SFLD Subgroup 5, B12-binding domain), Megacluster-2-2 (SFLD Subgroup 2, anaerobic coproporphyrinogen-III oxidase-like), or Cluster-7 ((SFLD Subgroup 16, PLP-dependent), the second section reports matches for clusters in the "diced" SSNs; the clusters with the three smallest e-values are listed. The number of UniProt IDs, number of cluster nodes, and UniProt ID convergence ratio (CR; described on the Subgroups tab) are provided for each identified cluster.

The "Exploring Subgroups" subtab under the "Functionally Diverse Subgroups" tab provides advice about interpreting the search results for these subgroups.

GND Lookup

The EFI-GNT web tools allow users to lookup genome neighborhood diagrams (GNDs) for lists of UniProt IDs. Users may find it convenient to be able to access the GNDs for members of the RSS within RadicalSAM.org.

The GND Viewer can be accessed with the button below. The input is a list of UniProt IDs. The GNDs will be displayed.

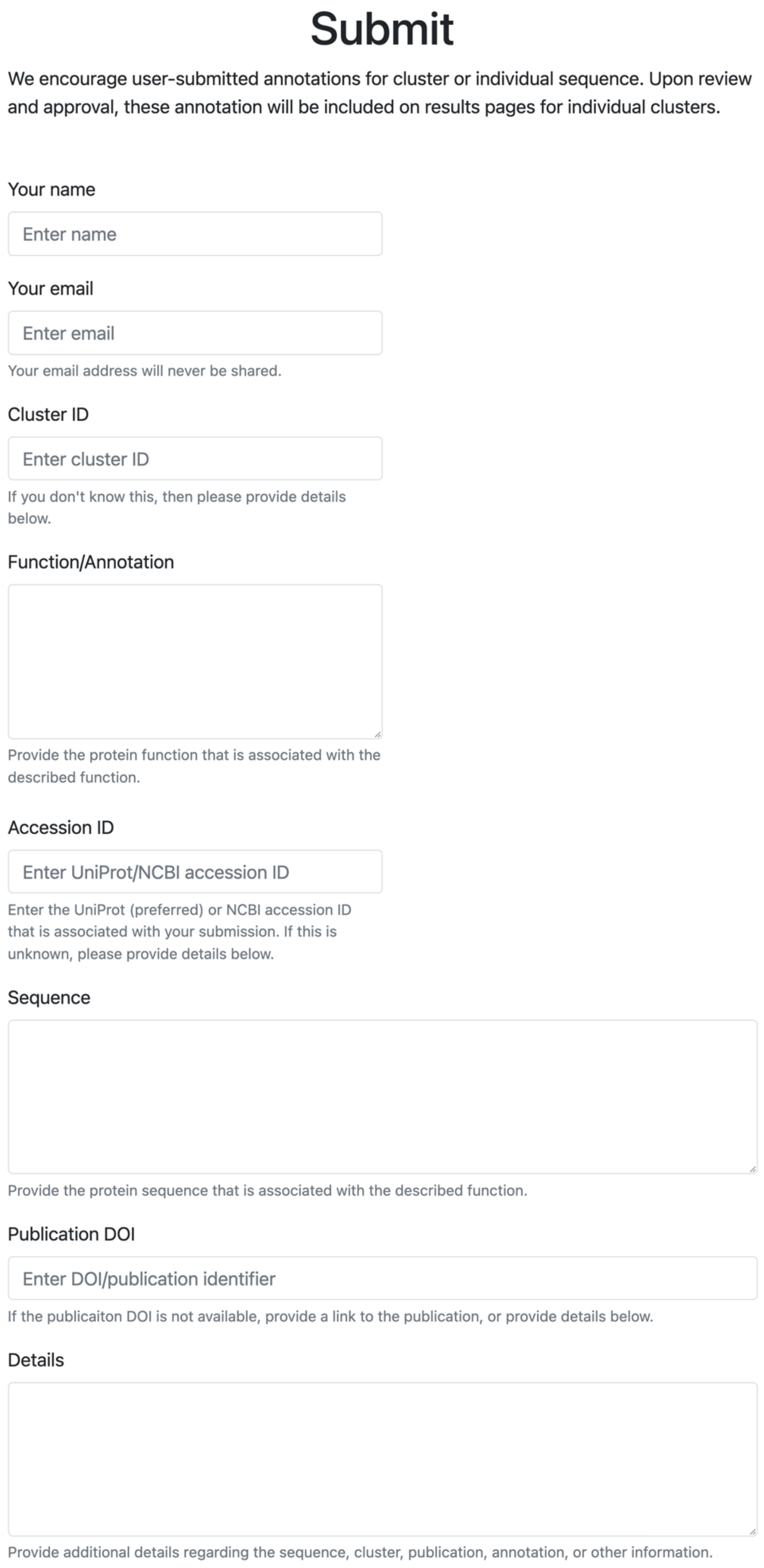
Find by Taxonomy

Input the genus/species/strain for an organism.

If only the genus is entered, a pop-up list of matching genus-species-strains is provided for selection of the desired genus/species/strain. If the genus and species are entered, a pop-up list of matching genus-species-strains is provided for selection of the desired genus/species/strain.

The search provides a list of sequences in the RSS. The list provides the UniProt ID (link to the UniProt page for sequence), UniProt description, organism name, UniProt annotation status (SwissProt or TrEMBL), and link to its Explore page.

Figure 14. Search functions: “Find by UniProt ID” identifies the cluster(s) containing a user-specified UniProt ID. “Find by Sequence” identifies the cluster(s) with the best HMM match to a user-specified sequence. “GND Lookup” provides the GND(s) for the user-specified UniProt ID(s). “Find by Taxonomy” provides a list of the UniProt IDs and their clusters for the user-specified genus/species.



Submit

We encourage user-submitted annotations for cluster or individual sequence. Upon review and approval, these annotation will be included on results pages for individual clusters.

Your name

Enter name

Your email

Enter email

Your email address will never be shared.

Cluster ID

Enter cluster ID

If you don't know this, then please provide details below.

Function/Annotation

Provide the protein function that is associated with the described function.

Accession ID

Enter UniProt/NCBI accession ID

Enter the UniProt (preferred) or NCBI accession ID that is associated with your submission. If this is unknown, please provide details below.

Sequence

Provide the protein sequence that is associated with the described function.

Publication DOI

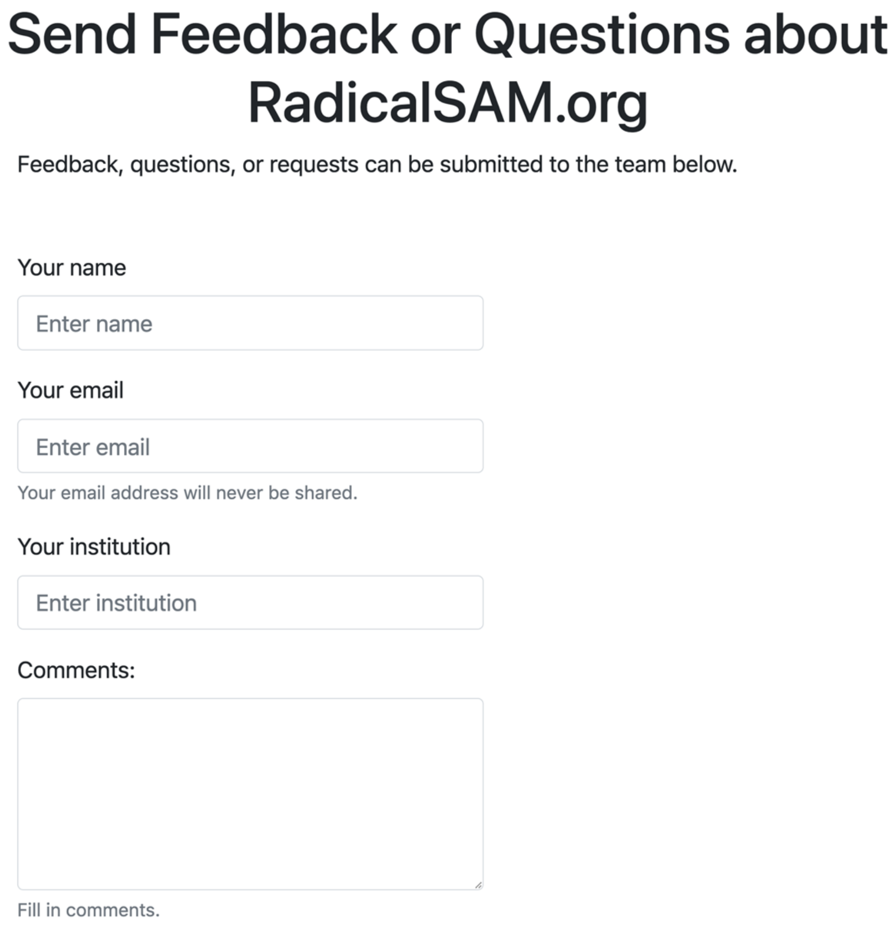
Enter DOI/publication identifier

If the publicaiton DOI is not available, provide a link to the publication, or provide details below.

Details

Provide additional details regarding the sequence, cluster, publication, annotation, or other information.

Figure 15. The Submit page for community submission of enzymatic activities and metabolic functions.



Send Feedback or Questions about RadicalSAM.org

Feedback, questions, or requests can be submitted to the team below.

Your name

Enter name

Your email

Enter email

Your email address will never be shared.

Your institution

Enter institution

Comments:

Fill in comments.

Figure 16. Contact page form for submitting feedback or questions.