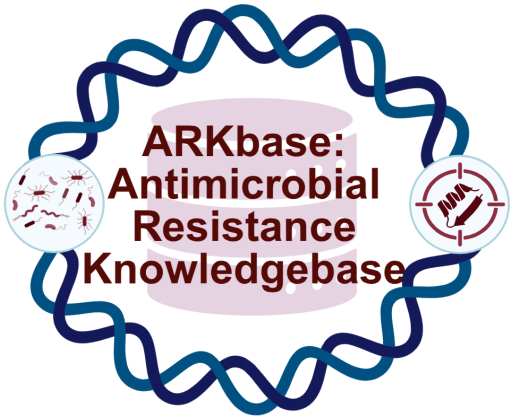
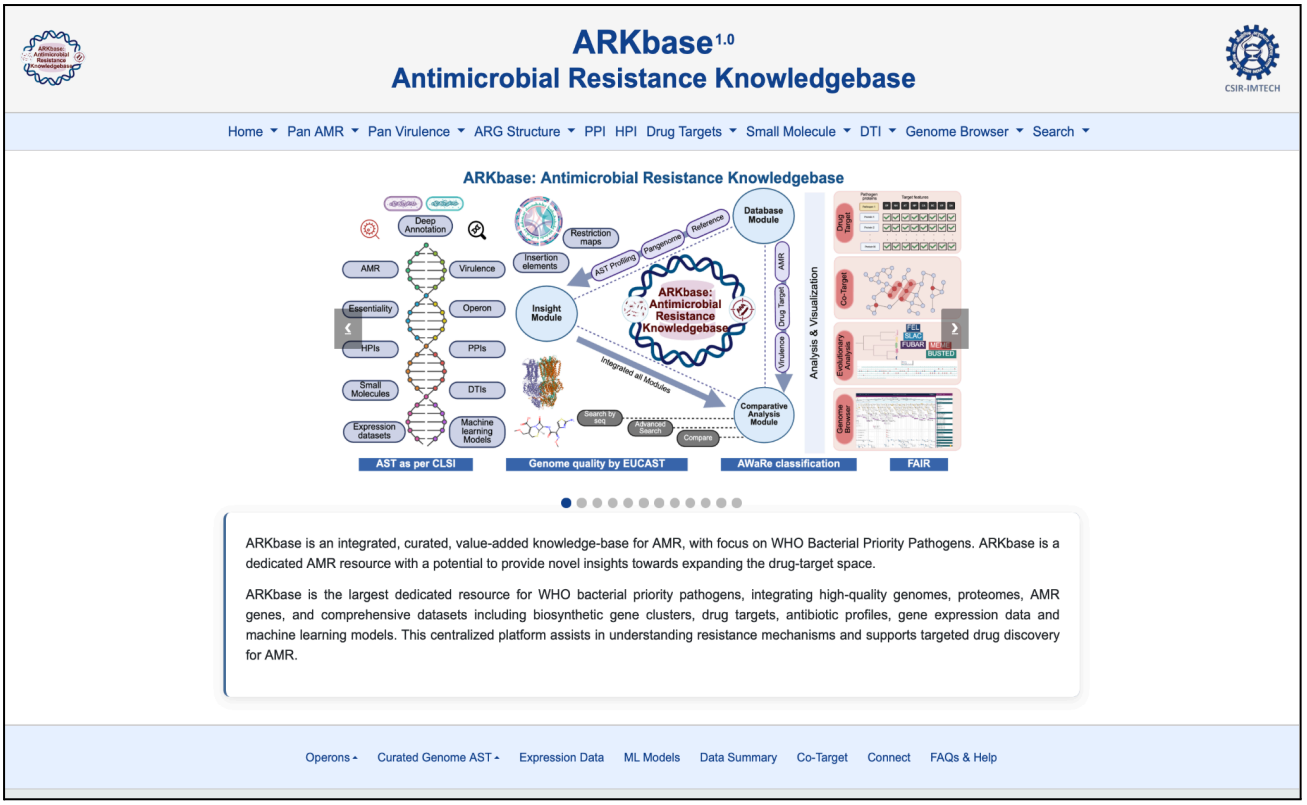


ARKbase HelpFile

https://datascience.imtech.res.in/anshu/arkbase/

ARKbase is an integrated, curated, value-added knowledgebase developed with the objective of providing a dedicated resource for deep annotation of bacterial priority pathogens (BPP) identified by WHO (**https://www.who.int/publications/i/item/9789240093461**).



Modules in ARKbase

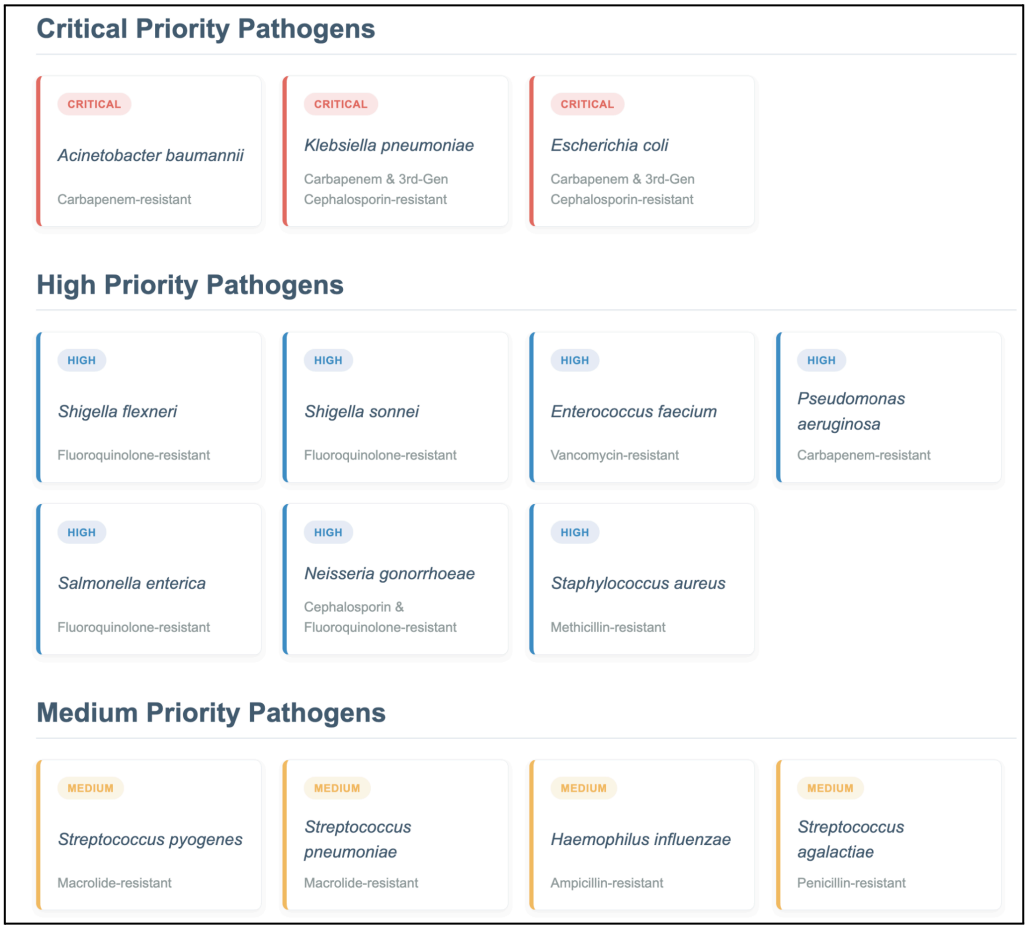
ARKbase offers three core modules—

- ☐ Database module
- ☐ Insight module
- ☐ Comparative Analysis module

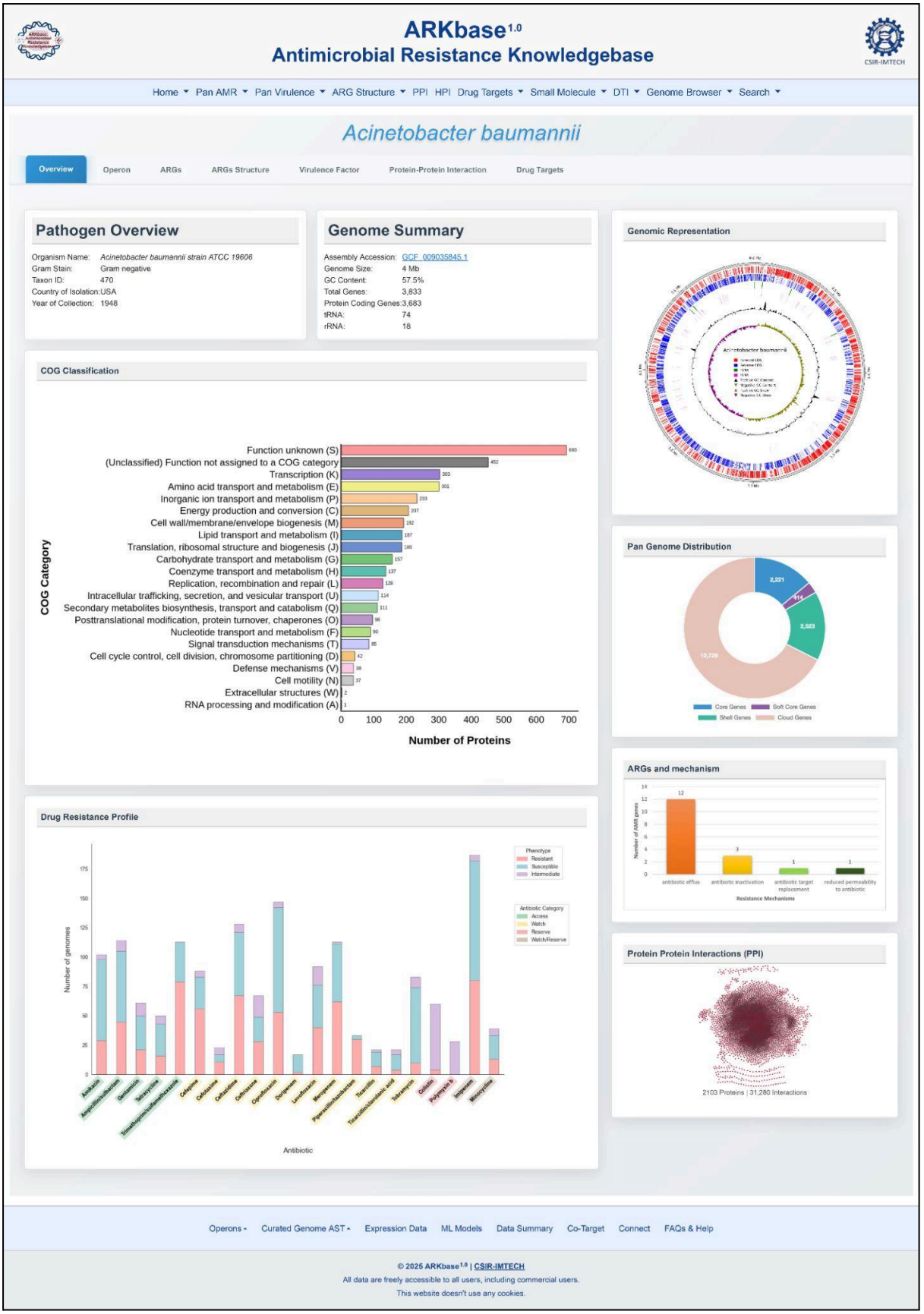
Database Module: This module is divided into the reference genome and the curated genome AST.

(i) Reference genomes

1. *Acinetobacter baumannii*
2. *Klebsiella pneumoniae*
3. *Escherichia coli*
4. *Shigella flexneri*
5. *Shigella sonnei*
6. *Enterococcus faecium*
7. *Pseudomonas aeruginosa*
8. *Salmonella enterica*
9. *Neisseria gonorrhoeae*
10. *Staphylococcus aureus*
11. *Streptococcus pyogenes*
12. *Streptococcus pneumoniae*
13. *Haemophilus influenzae*
14. *Streptococcus agalactiae*

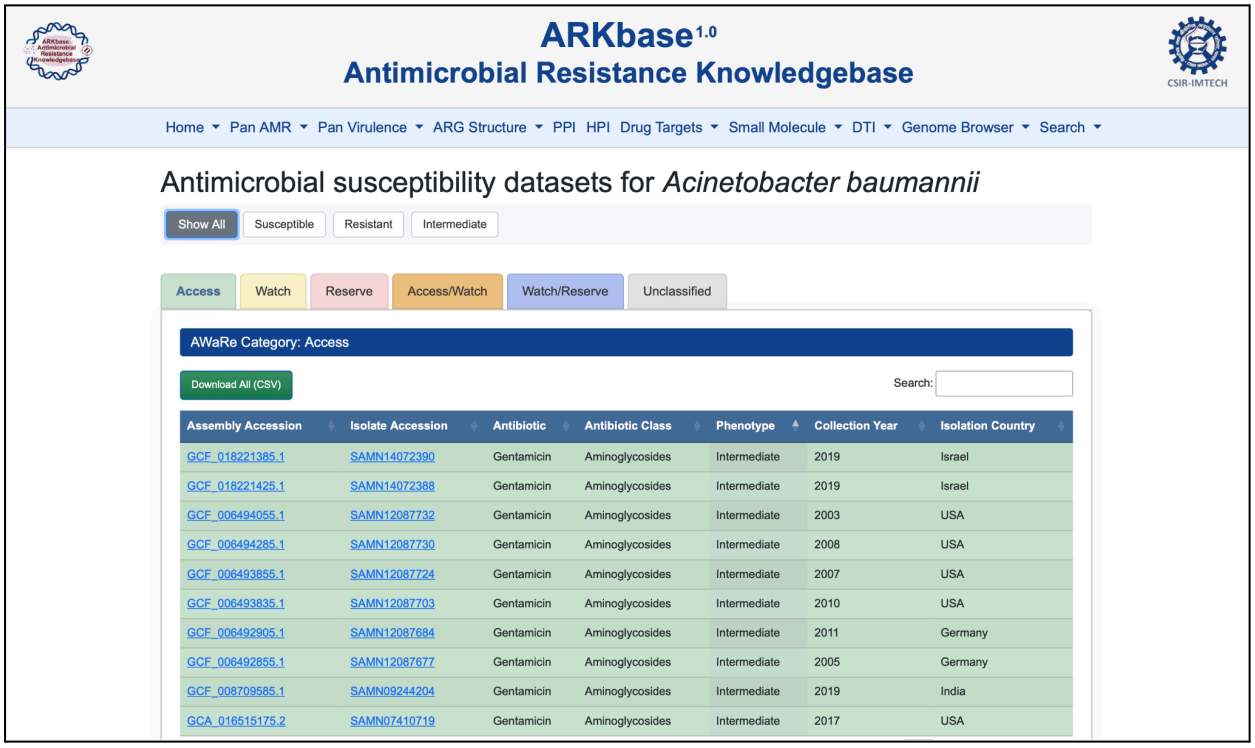


The reference genome page provides pathogen overview, genome overview, genomic representation, COG classification, pangenome distribution, drug resistance profile, ARGs and mechanisms, and protein-protein interactions (PPI).



(ii) Curated Genome with Antimicrobial susceptibility (AST)

Genome datasets are obtained from publicly available resources and then quality checked primarily on two criteria - The MICs reported for these strains are quality checked against CLSI breakpoints and the corresponding genome datasets were checked as per EUCAST criteria of good quality genomes. More details can be read at https://pubmed.ncbi.nlm.nih.gov/36817105/

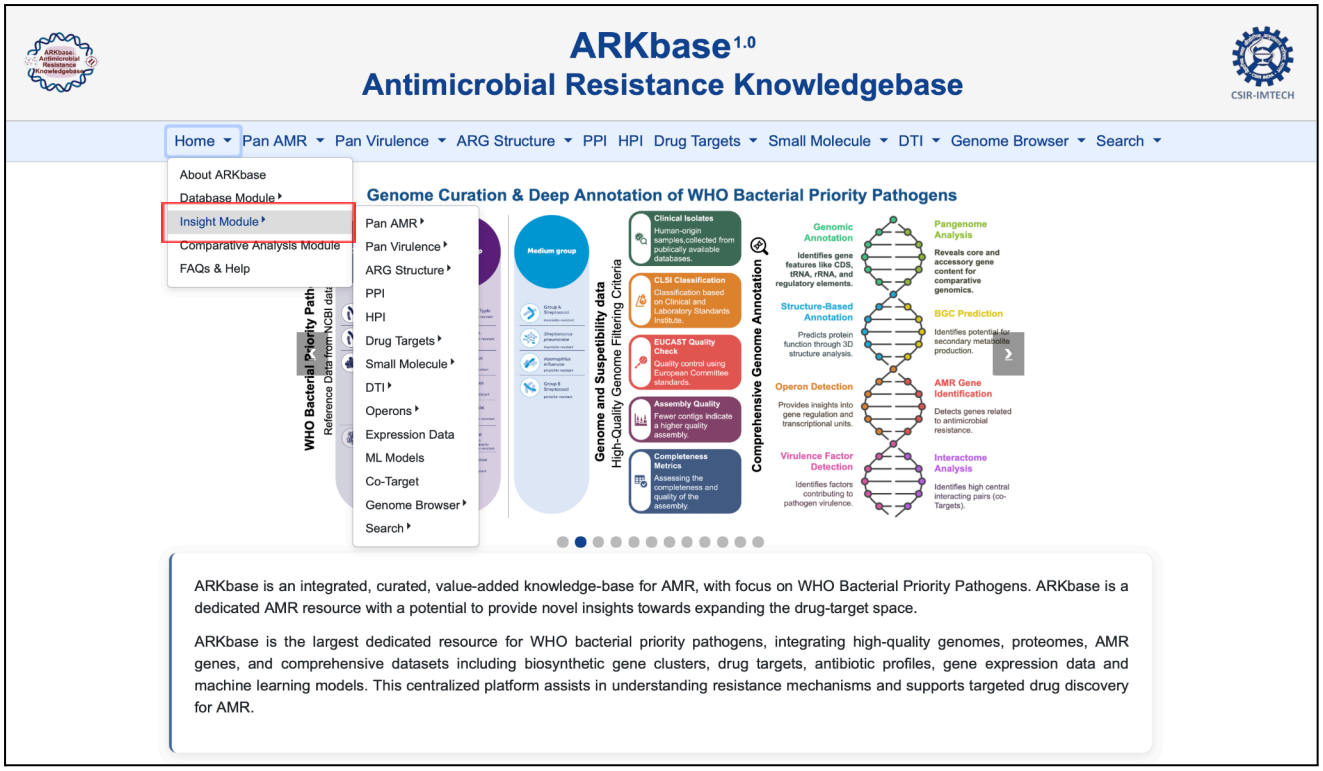


Assembly Accession	Isolate Accession	Antibiotic	Antibiotic Class	Phenotype	Collection Year	Isolation Country
GCF_018221385.1	SAMN14072390	Gentamicin	Aminoglycosides	Intermediate	2019	Israel
GCF_018221425.1	SAMN14072388	Gentamicin	Aminoglycosides	Intermediate	2019	Israel
GCF_006494055.1	SAMN12087732	Gentamicin	Aminoglycosides	Intermediate	2003	USA
GCF_006494285.1	SAMN12087730	Gentamicin	Aminoglycosides	Intermediate	2008	USA
GCF_006493855.1	SAMN12087724	Gentamicin	Aminoglycosides	Intermediate	2007	USA
GCF_006493835.1	SAMN12087703	Gentamicin	Aminoglycosides	Intermediate	2010	USA
GCF_006492905.1	SAMN12087684	Gentamicin	Aminoglycosides	Intermediate	2011	Germany
GCF_006492855.1	SAMN12087677	Gentamicin	Aminoglycosides	Intermediate	2005	Germany
GCF_008709585.1	SAMN09244204	Gentamicin	Aminoglycosides	Intermediate	2019	India
GCA_016515175.2	SAMN07410719	Gentamicin	Aminoglycosides	Intermediate	2017	USA

Insight Modules

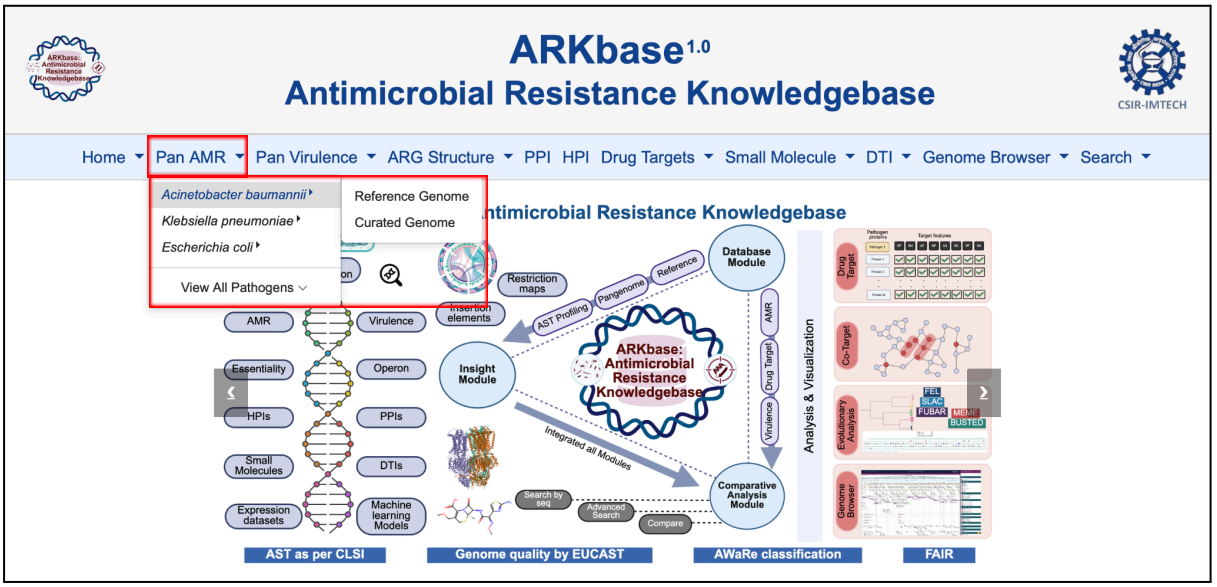
Insight module is a composite module and is further divided into 14 modules as follows:

1. Pan AMR
2. Pan Virulence
3. ARG structure
4. Protein-Protein Interactions
5. Host-Pathogen Interactions
6. Drug Targets
7. Small Molecules
8. Drug-Target Interactions
9. Operons
10. Expression data
11. Machine Learning Models
12. Co-Target
13. Genome Browser
14. Search

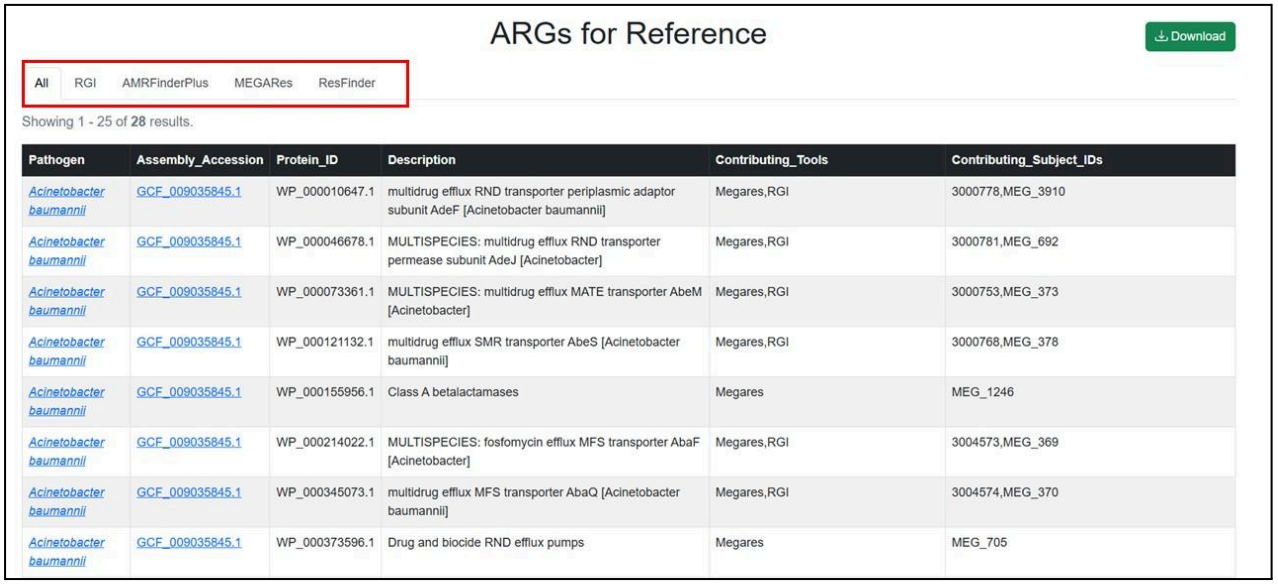


Pan AMR:

This module contains pathogen-specific ARGs, divided into two pages: reference and curated genomes.

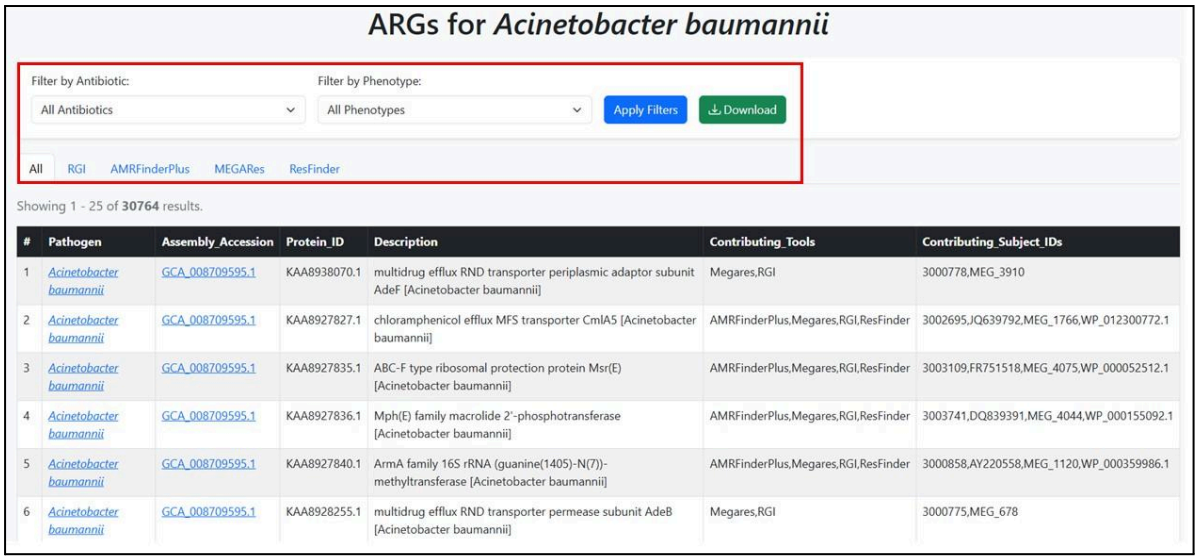


For the reference page, there are five sections: outputs from four different tools, RGI, AMRFinderPlus, MEGARes, and ResFinder, along with a combined tool output. The user can download the data.



Pathogen	Assembly_Accession	Protein_ID	Description	Contributing_Tools	Contributing_Subject_IDs
Acinetobacter baumannii	GCF_009035845.1	WP_000010647.1	multidrug efflux RND transporter periplasmic adaptor subunit AdeF [Acinetobacter baumannii]	Megares,RGI	3000778,MEG_3910
Acinetobacter baumannii	GCF_009035845.1	WP_000046678.1	MULTISPECIES: multidrug efflux RND transporter permease subunit AdeJ [Acinetobacter]	Megares,RGI	3000781,MEG_692
Acinetobacter baumannii	GCF_009035845.1	WP_000073361.1	MULTISPECIES: multidrug efflux MATE transporter AbeM [Acinetobacter]	Megares,RGI	3000753,MEG_373
Acinetobacter baumannii	GCF_009035845.1	WP_000121132.1	multidrug efflux SMR transporter AbeS [Acinetobacter baumannii]	Megares,RGI	3000768,MEG_378
Acinetobacter baumannii	GCF_009035845.1	WP_000155956.1	Class A betalactamases	Megares	MEG_1246
Acinetobacter baumannii	GCF_009035845.1	WP_000214022.1	MULTISPECIES: fosfomycin efflux MFS transporter AbaF [Acinetobacter]	Megares,RGI	3004573,MEG_369
Acinetobacter baumannii	GCF_009035845.1	WP_000345073.1	multidrug efflux MFS transporter AbaQ [Acinetobacter baumannii]	Megares,RGI	3004574,MEG_370
Acinetobacter baumannii	GCF_009035845.1	WP_000373596.1	Drug and biocide RND efflux pumps	Megares	MEG_705

For the curated page, there are five sections: outputs from four different tools—RGI, AMRFinderPlus, MEGARes, and ResFinder—along with a combined tool output, the same as the reference page. In addition, the user can filter the data using antibiotics and phenotypes.

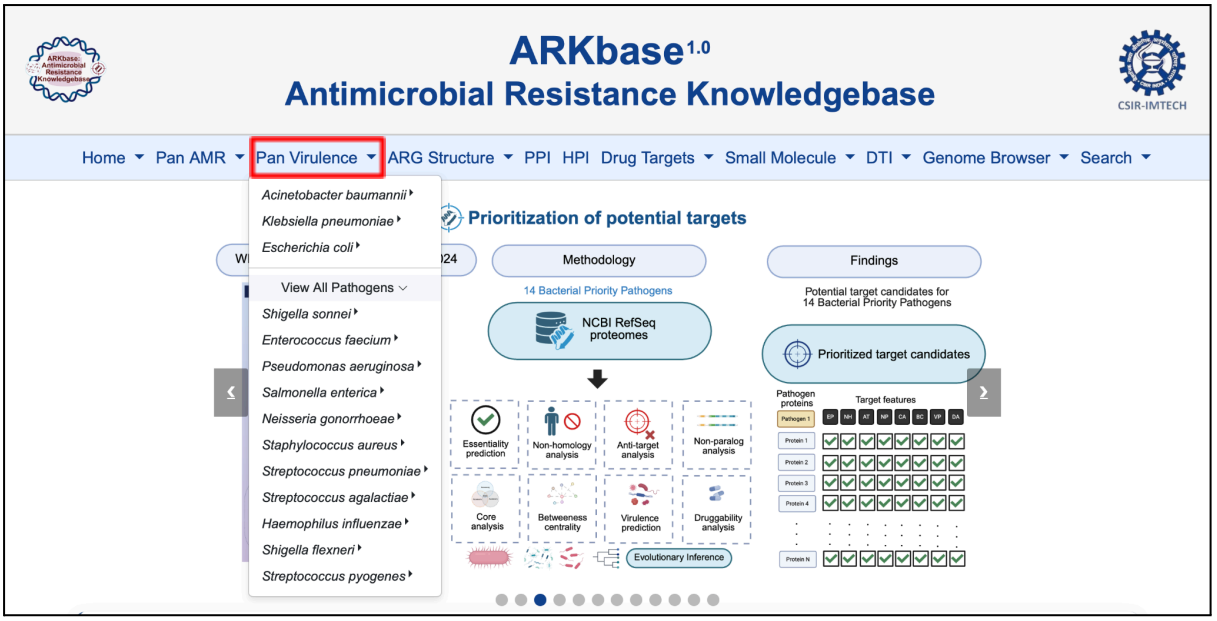


#	Pathogen	Assembly_Accession	Protein_ID	Description	Contributing_Tools	Contributing_Subject_IDs
1	*Acinetobacter baumannii*	GCA_008709595.1	KAA8938070.1	multidrug efflux RND transporter periplasmic adaptor subunit AdeF [Acinetobacter baumannii]	Megares,RGI	3000778,MEG_3910
2	*Acinetobacter baumannii*	GCA_008709595.1	KAA8927827.1	chloramphenicol efflux MFS transporter CmlA5 [Acinetobacter baumannii]	AMRFinderPlus,Megares,RGI,ResFinder	3002695,JQ639792,MEG_1766,WP_012300772.1
3	*Acinetobacter baumannii*	GCA_008709595.1	KAA8927835.1	ABC-F type ribosomal protection protein Msr(E) [Acinetobacter baumannii]	AMRFinderPlus,Megares,RGI,ResFinder	3003109,FR751518,MEG_4075,WP_000052512.1
4	*Acinetobacter baumannii*	GCA_008709595.1	KAA8927836.1	Mph(E) family macrolide 2'-phosphotransferase [Acinetobacter baumannii]	AMRFinderPlus,Megares,RGI,ResFinder	3003741,DQ839391,MEG_4044,WP_000155092.1
5	*Acinetobacter baumannii*	GCA_008709595.1	KAA8927840.1	ArmA family 16S rRNA (guanine(1405)-N(7))-methyltransferase [Acinetobacter baumannii]	AMRFinderPlus,Megares,RGI,ResFinder	3000858,AY220558,MEG_1120,WP_000359986.1
6	*Acinetobacter baumannii*	GCA_008709595.1	KAA8928255.1	multidrug efflux RND transporter permease subunit AdeB [Acinetobacter baumannii]	Megares,RGI	3000775,MEG_678

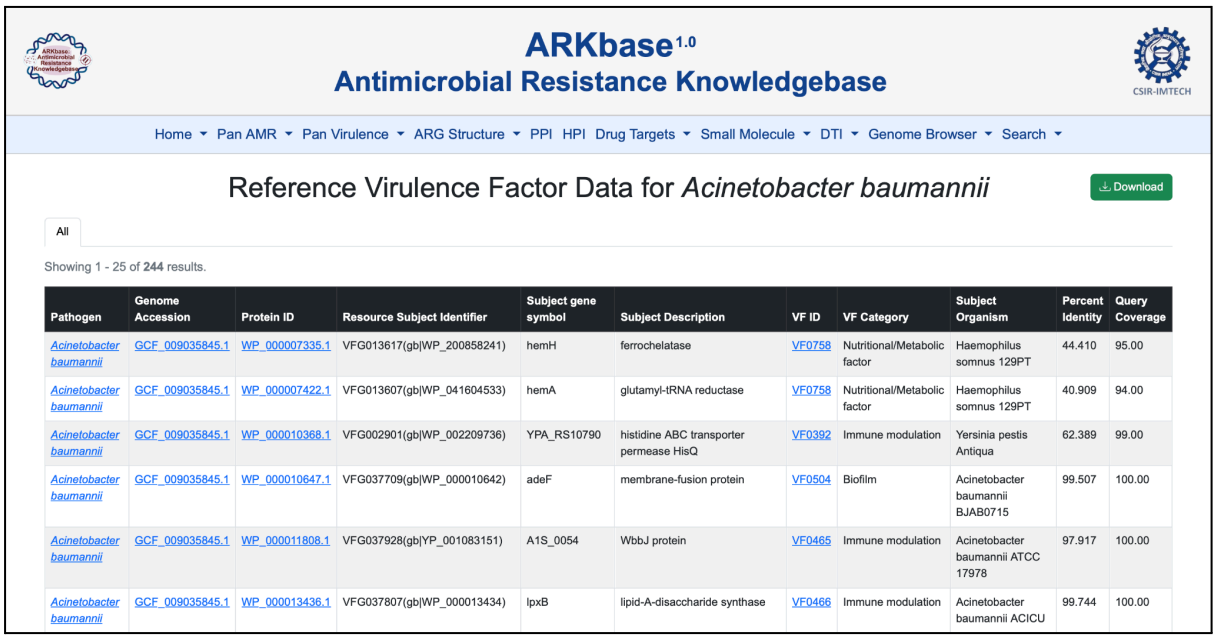
Pan Virulence:

The PanVirulence module contains pathogen-specific virulence, divided into two pages: reference and curated genomes.

This tab provides the user a consolidated table of virulence factors for each pathogen chosen from the drop-down menu. The user can explore the virulence genes with the corresponding protein id, gene name, virulence factor category, operon id, VF ID and the organism name where the particular gene has matched. The protein_ids are directly linked to the NCBI database. The page also allows the user to search specific fields and to download the tables.

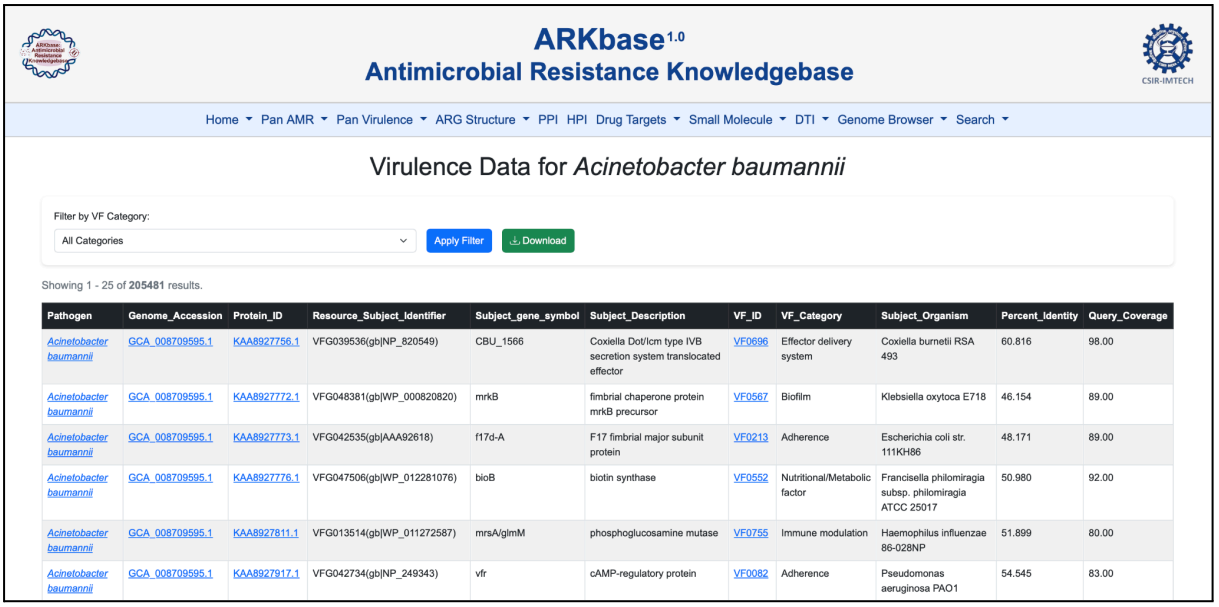


The module includes an interactive filter to search across fields by protein ID, protein name,VF ID and so on. The download tab allows the user to download the complete virulence data for that particular genome.



Pathogen	Genome Accession	Protein ID	Resource Subject Identifier	Subject gene symbol	Subject Description	VF ID	VF Category	Subject Organism	Percent Identity	Query Coverage
Acinetobacter baumannii	GCF_009035845.1	WP_000007335.1	VFG013617(gb|WP_200858241)	hemH	ferrochelatase	VF0758	Nutritional/Metabolic factor	Haemophilus somnus 129PT	44.410	95.00
Acinetobacter baumannii	GCF_009035845.1	WP_000007422.1	VFG013607(gb|WP_041604533)	hemA	glutamyl-tRNA reductase	VF0758	Nutritional/Metabolic factor	Haemophilus somnus 129PT	40.909	94.00
Acinetobacter baumannii	GCF_009035845.1	WP_000010368.1	VFG002901(gb|WP_002209736)	YPA_RS10790	histidine ABC transporter permease HisQ	VF0392	Immune modulation	Yersinia pestis Antiqua	62.389	99.00
Acinetobacter baumannii	GCF_009035845.1	WP_000010647.1	VFG037709(gb|WP_000010642)	adeF	membrane-fusion protein	VF0504	Biofilm	Acinetobacter baumannii BJAB0715	99.507	100.00
Acinetobacter baumannii	GCF_009035845.1	WP_000011808.1	VFG037928(gb|YP_001083151)	A1S_0054	WbbJ protein	VF0465	Immune modulation	Acinetobacter baumannii ATCC 17978	97.917	100.00
Acinetobacter baumannii	GCF_009035845.1	WP_000013436.1	VFG037807(gb|WP_000013434)	lpxB	lipid-A-disaccharide synthase	VF0466	Immune modulation	Acinetobacter baumannii ACICU	99.744	100.00

For the curated page, the module includes an interactive filter to search across fields by protein ID, protein name, VF ID and so on. The download tab allows the user to download the complete virulence data for that particular genome.



ARKbase 1.0
Antimicrobial Resistance Knowledgebase

CSIR-IMTECH

Home ▾ Pan AMR ▾ Pan Virulence ▾ ARG Structure ▾ PPI HPI Drug Targets ▾ Small Molecule ▾ DTI ▾ Genome Browser ▾ Search ▾

Virulence Data for *Acinetobacter baumannii*

Filter by VF Category:

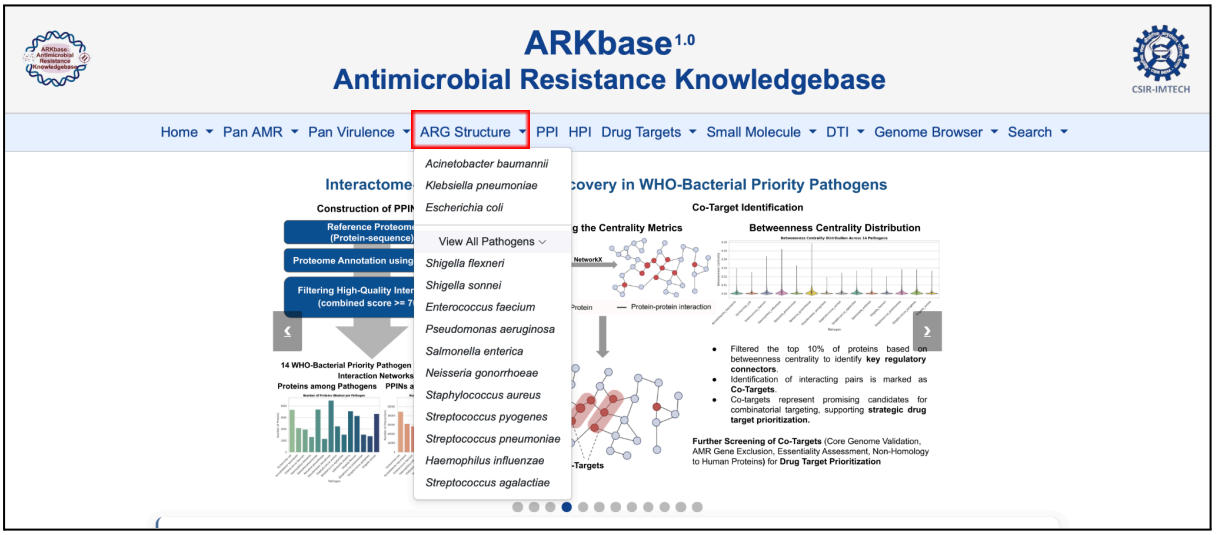
All Categories | Apply Filter | Download

Showing 1 - 25 of **205481** results.

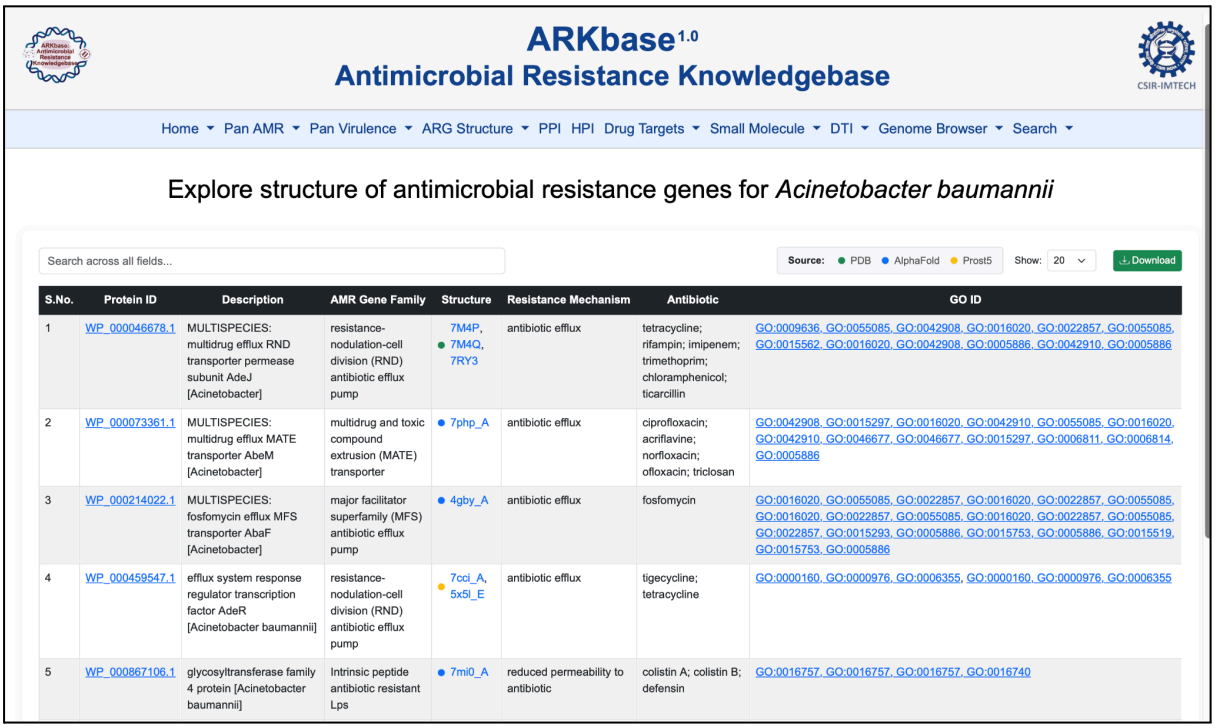
Pathogen	Genome_Accession	Protein_ID	Resource_Subject_Identifier	Subject_gene_symbol	Subject_Description	VF_ID	VF_Category	Subject_Organism	Percent_Identity	Query_Coverage
Acinetobacter baumannii	GCA_008709595.1	KAA8927756.1	VFG039536(gb|NP_820549)	CBU_1566	Coxiella Dot/Icm type IVB secretion system translocated effector	VF0696	Effector delivery system	Coxiella burnetii RSA 493	60.816	98.00
Acinetobacter baumannii	GCA_008709595.1	KAA8927772.1	VFG048381(gb|WP_000820820)	mrkB	fimbrial chaperone protein mrkB precursor	VF0567	Biofilm	Klebsiella oxytoca E718	46.154	89.00
Acinetobacter baumannii	GCA_008709595.1	KAA8927773.1	VFG042535(gb|AAA92618)	f17d-A	F17 fimbrial major subunit protein	VF0213	Adherence	Escherichia coli str. 111KH86	48.171	89.00
Acinetobacter baumannii	GCA_008709595.1	KAA8927776.1	VFG047506(gb|WP_012281076)	bioB	biotin synthase	VF0552	Nutritional/Metabolic factor	Francisella philomiragia subsp. philomiragia ATCC 25017	50.980	92.00
Acinetobacter baumannii	GCA_008709595.1	KAA8927811.1	VFG013514(gb|WP_011272587)	mrsA/glmM	phosphoglucosamine mutase	VF0755	Immune modulation	Haemophilus influenzae 86-028NP	51.899	80.00
Acinetobacter baumannii	GCA_008709595.1	KAA8927917.1	VFG042734(gb|NP_249343)	vfr	cAMP-regulatory protein	VF0082	Adherence	Pseudomonas aeruginosa PAO1	54.545	83.00

Antimicrobial Resistance Gene (ARG) Structure:

This module gives pathogen-wise structure of the antimicrobial resistance (AMR) gene of the WHO Bacterial Priority Pathogen List.

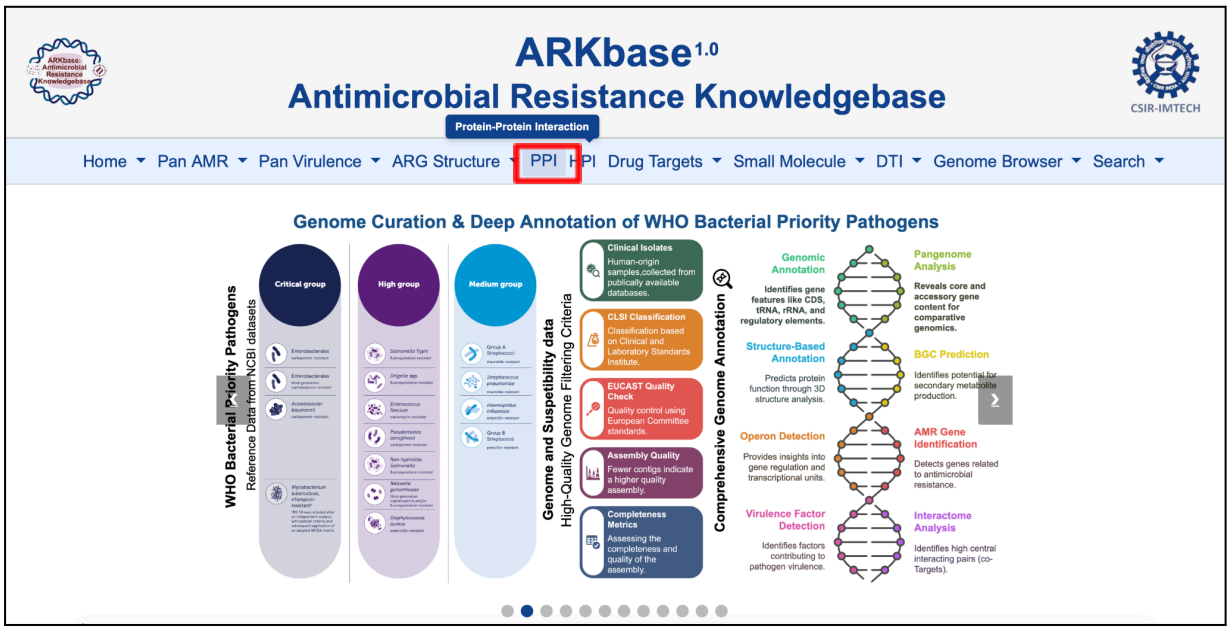


The user can search for any data field through the search interface. The user can view the consolidated table with AMR gene ID, description, AMR gene family, resistance mechanism, operon and their structure, antibiotic, and GO ID of that particular pathogen. The user can download the whole data in CSV/Excel format for analysis.

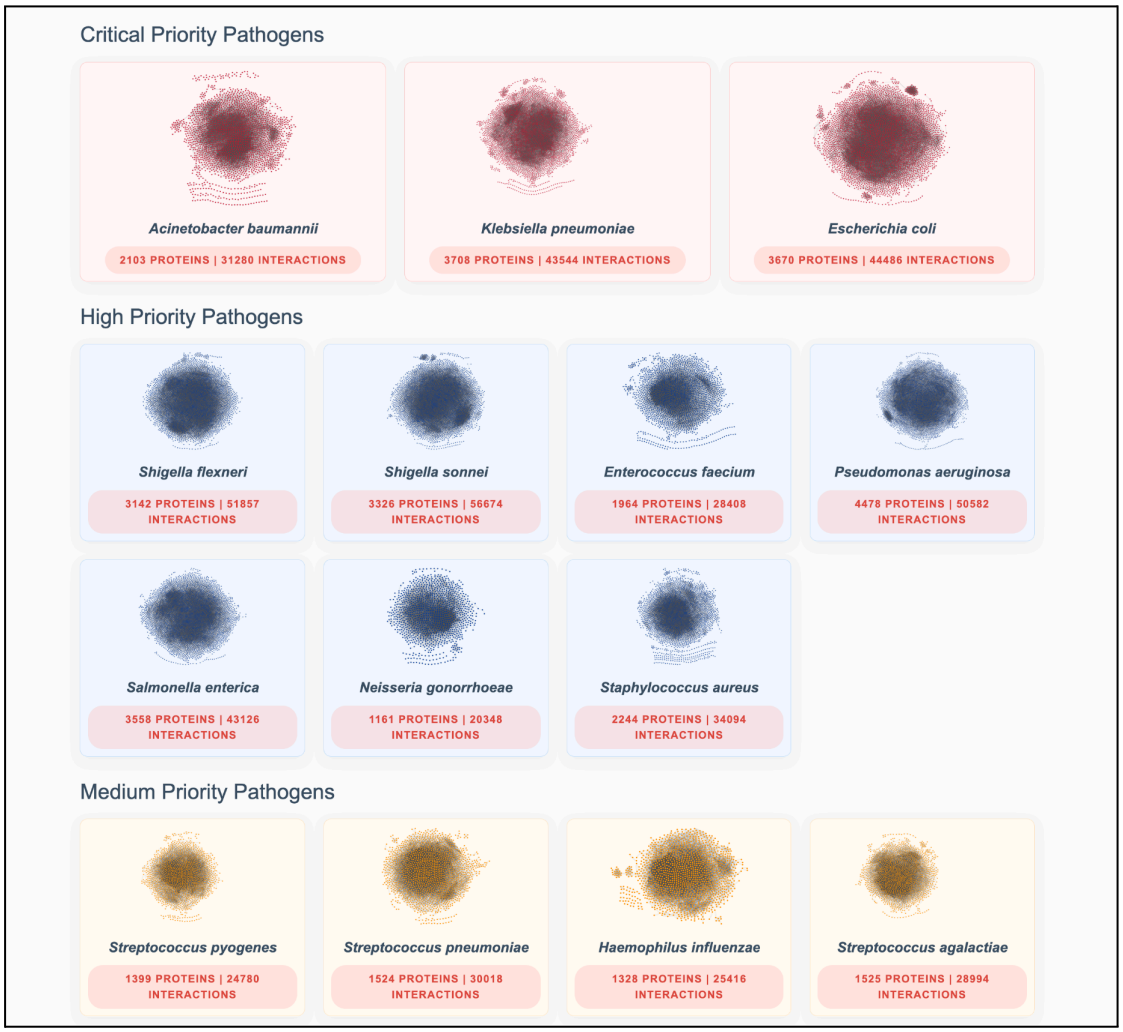


Protein-protein interactions (PPIs):

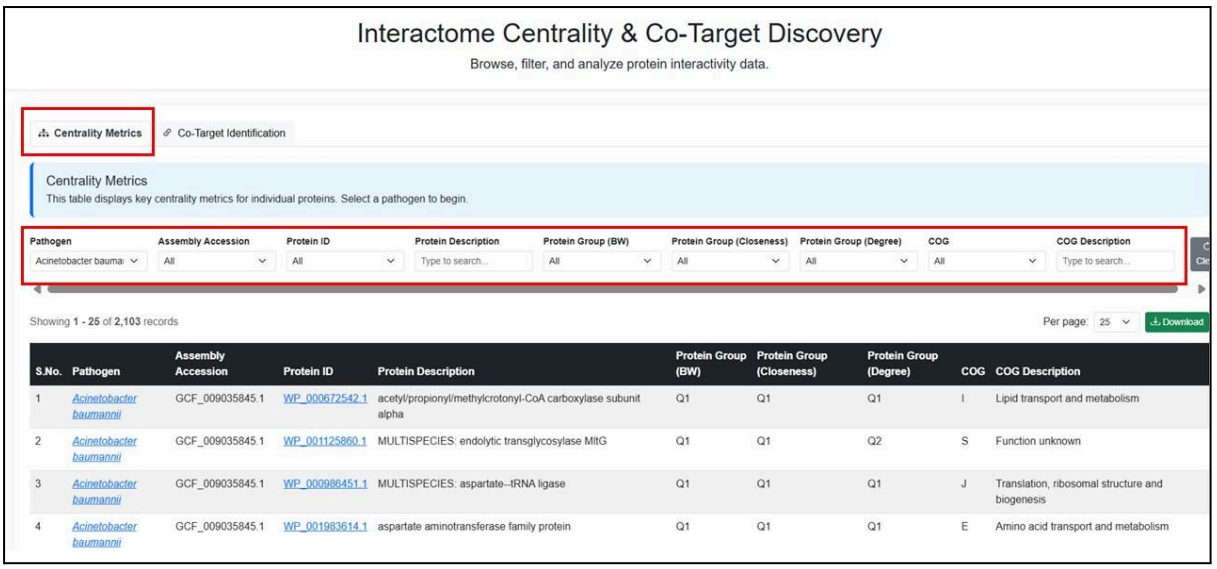
This module provides a protein-protein interactions landscape for BPPs.



This module consists of 14 PPIs of WHO BBP classified into three categories - Critical, High, and Medium Priority Pathogens, along with the number of proteins and interactions. The user can select any pathogen to explore centrality and co-targets.



This module is organised into two main sections: Centrality Metrics and Co-Target Interactions.



1. **Centrality Metrics**

   This section provides centrality metrics of each protein for fourteen pathogens. For each network, three centrality metrics are calculated: Betweenness centrality, Closeness centrality, and Degree centrality.

   Each centrality metric is classified into four groups:

   Q1 – Top 25% of nodes (highest centrality values)

   Q2 – 25–50% range

   Q3 – 50–75% range

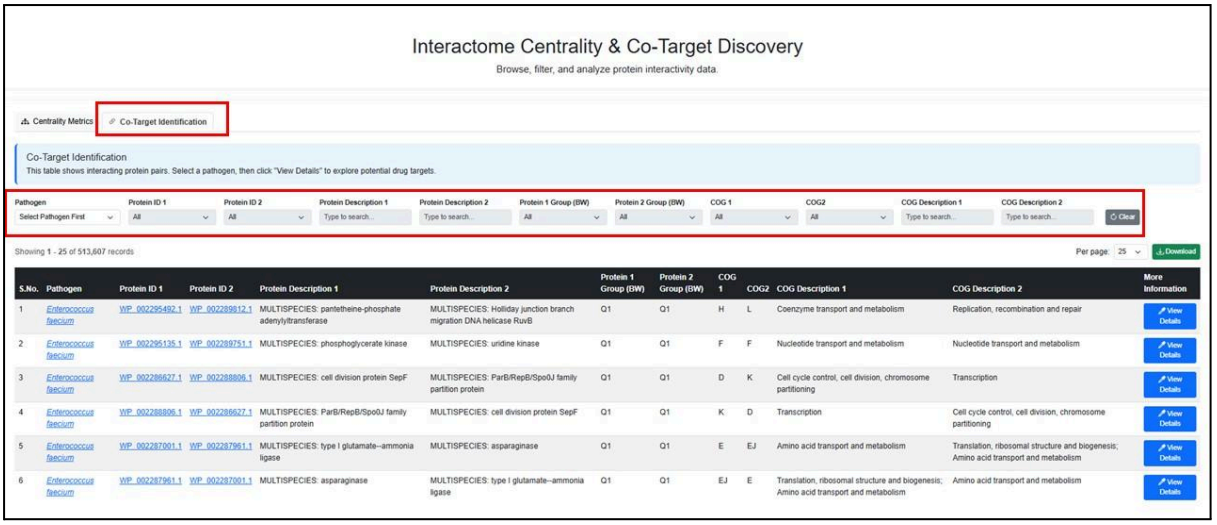
   Q4 – 75–100% range (lowest centrality values)

   This classification helps in identifying highly influential nodes (Q1) versus peripheral nodes (Q4) for each pathogen's interactome.

   The module includes nine interactive filters for targeted data exploration:

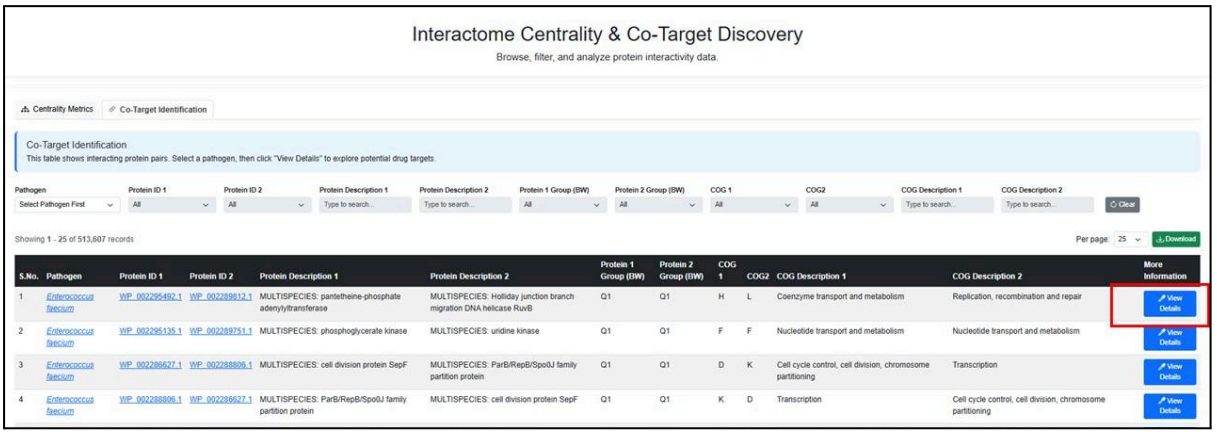
   - Pathogen
   - Assembly Accession
   - Protein ID
   - Protein Description
   - Protein Groups (Betweenness)
   - Protein Groups (Closeness)
   - Protein Groups (Degree)
   - COG (Clusters of Orthologous Groups)
   - COG Description

   The user can also download the data.



2. Co-Target Discovery

This section highlights the interacting pairs for each pathogen. The module includes nine interactive filters for targeted data exploration: Pathogen, Protein ID1, Protein ID2, Protein Description 1, Protein Description 2, Protein Group 1(BW), Protein Group 2(BW), COG 1, COG 2, COG Description 1, and COG Description 2. Co-Targets are the co-interactors with drug target features.



View details tab lists drug target features of each co-interactor as shown below.

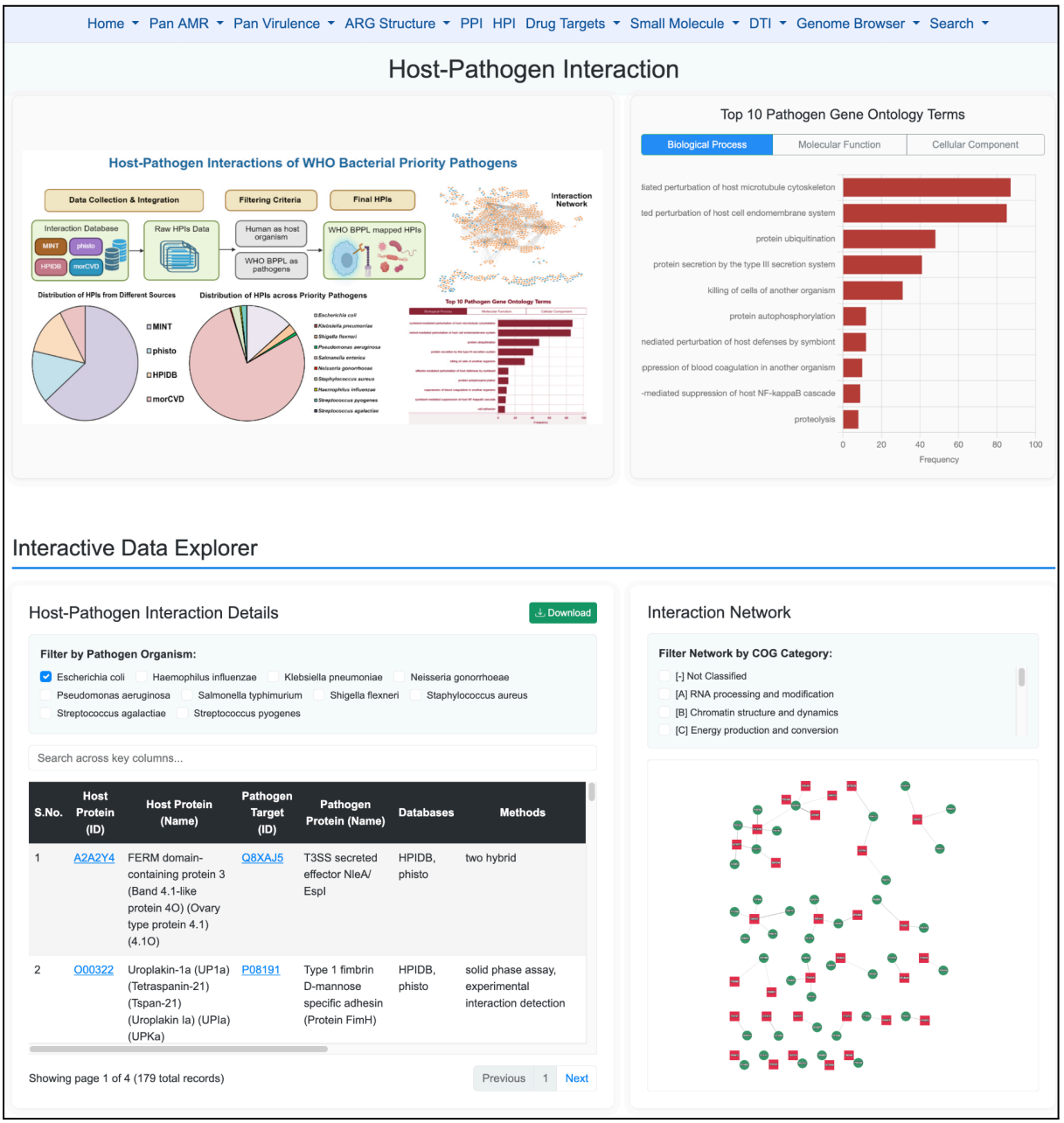
Details for Protein 1: WP_002295492.1

Pathogen Name	*Enterococcus faecium*
Prot Id	WP_002295492.1
Prot Desc	MULTISPECIES: pantetheine-phosphate adenylyltransferase
Non Paralog	Yes
Virulence	Yes
Essential	Yes
Ttd Novel	Yes
Drugbank Novel	Yes
Human Non Homologous	Yes
Anti Target	No
Top 10 Central	Yes
Core	Yes
Not Amr	Yes

Host-Pathogen Interactions:

It provides an interactive data explorer for host-pathogen interaction. It includes an interaction score and experimental information, as well as the top 10 Gene Ontology (GO) terms. The user can search by UniProt identifiers for pathogen and host proteins. This module is divided into four sections:

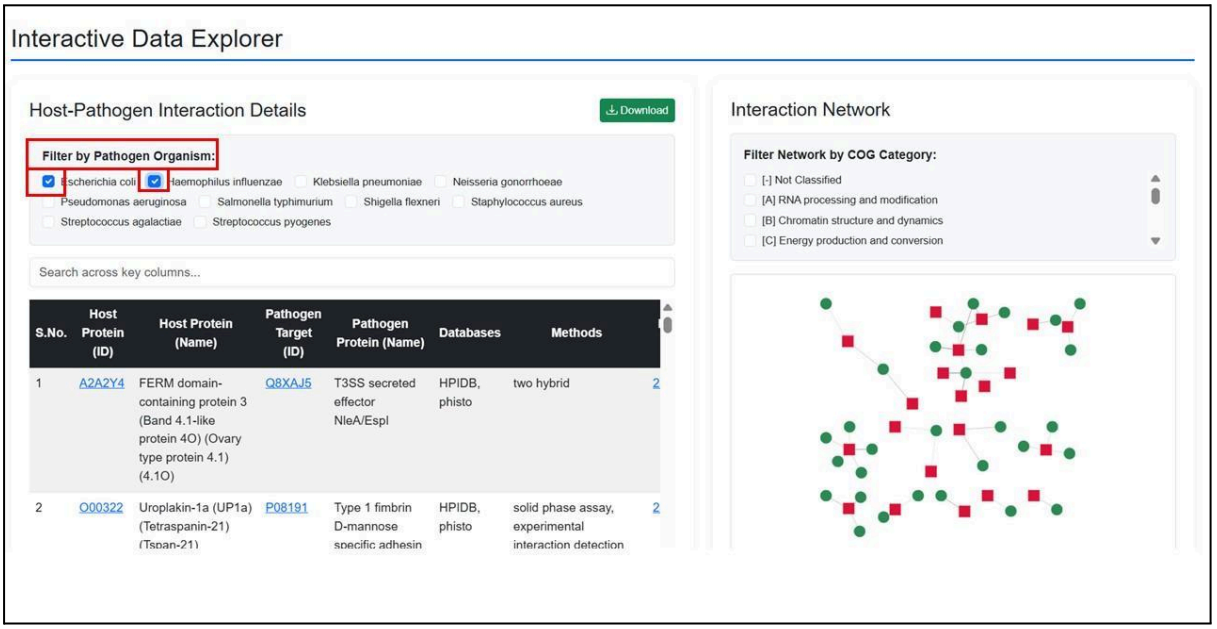
1. Graphical abstract for the overall summary
2. Top 10 Pathogen Gene Ontology Terms
3. Host-Pathogen Interaction Details
4. Interaction Network



1. Graphical abstract for the overall summary:
This section shows the host-pathogen interaction data sources, pathogen distribution, interaction network, and top functional annotations.

2. Top 10 Pathogen Gene Ontology Terms:
This section displays the bar graph for the top 10 Gene ontology terms. User can select the three gene ontology terms - Biological Process , Molecular Function and Cellular Component.

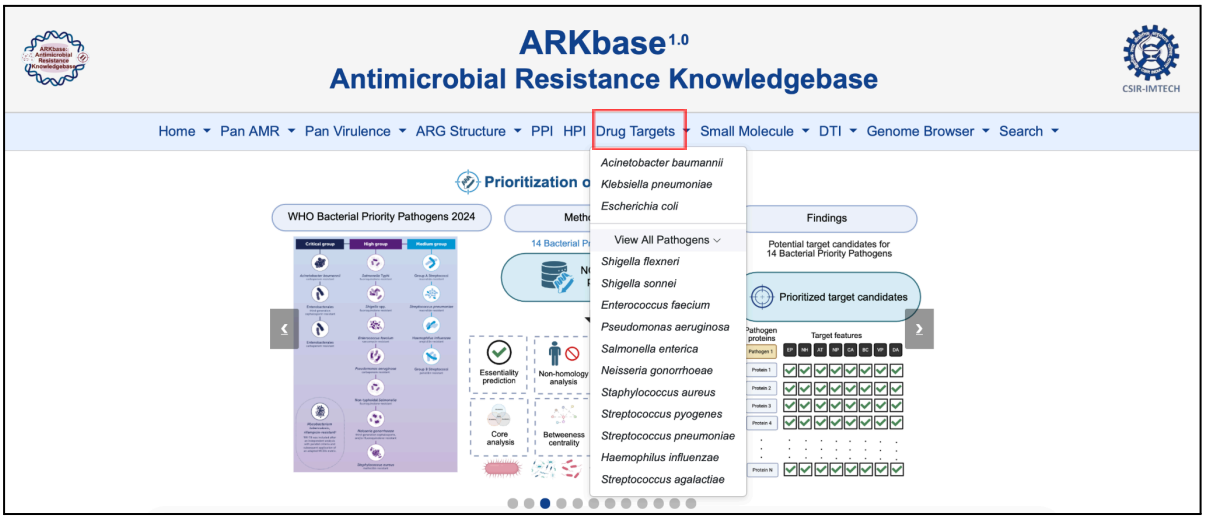
3. Host-Pathogen Interaction Details:
It provides the pathogen protein's ID and name, the corresponding host protein's ID and name, the source databases, the experimental methods, etc. The user can filter the host-pathogen interaction data by selecting one or multiple pathogen organism names. The result can be displayed in the form of network. The user can download the data.



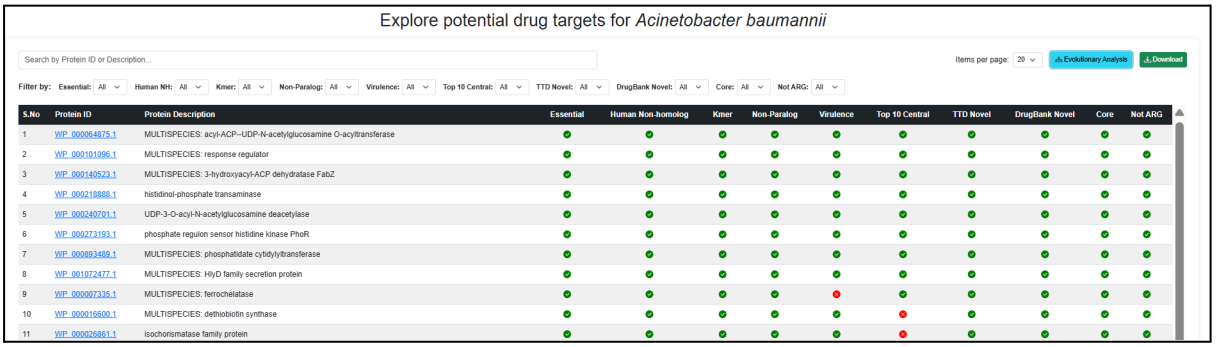
4. Interaction Network:
The user can filter interaction networks by selecting the host proteins according to their COG (Clusters of Orthologous Genes) functional categories.

Drug Targets:

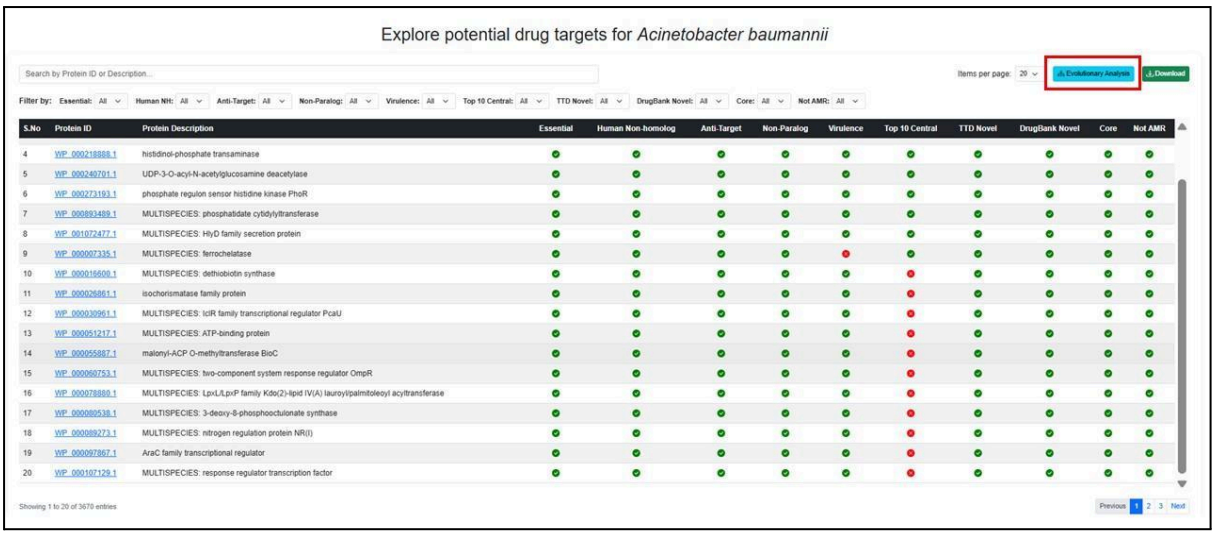
This module provides pathogen-wise potential predicted drug targets for the 14 WHO pathogens; any pathogen can be selected from the drop-down menu of the home page. This module allows users to search by pathogen protein ID and their name (description). The user has the option to filter the targets based upon the provided properties like essentiality, virulence, betweenness centrality metrics, human non-homolog, non-paralogous, and k-mer based non-homolog (anti-target), along with their novelty based upon the Therapeutic Target Database (TTD) and DrugBank.



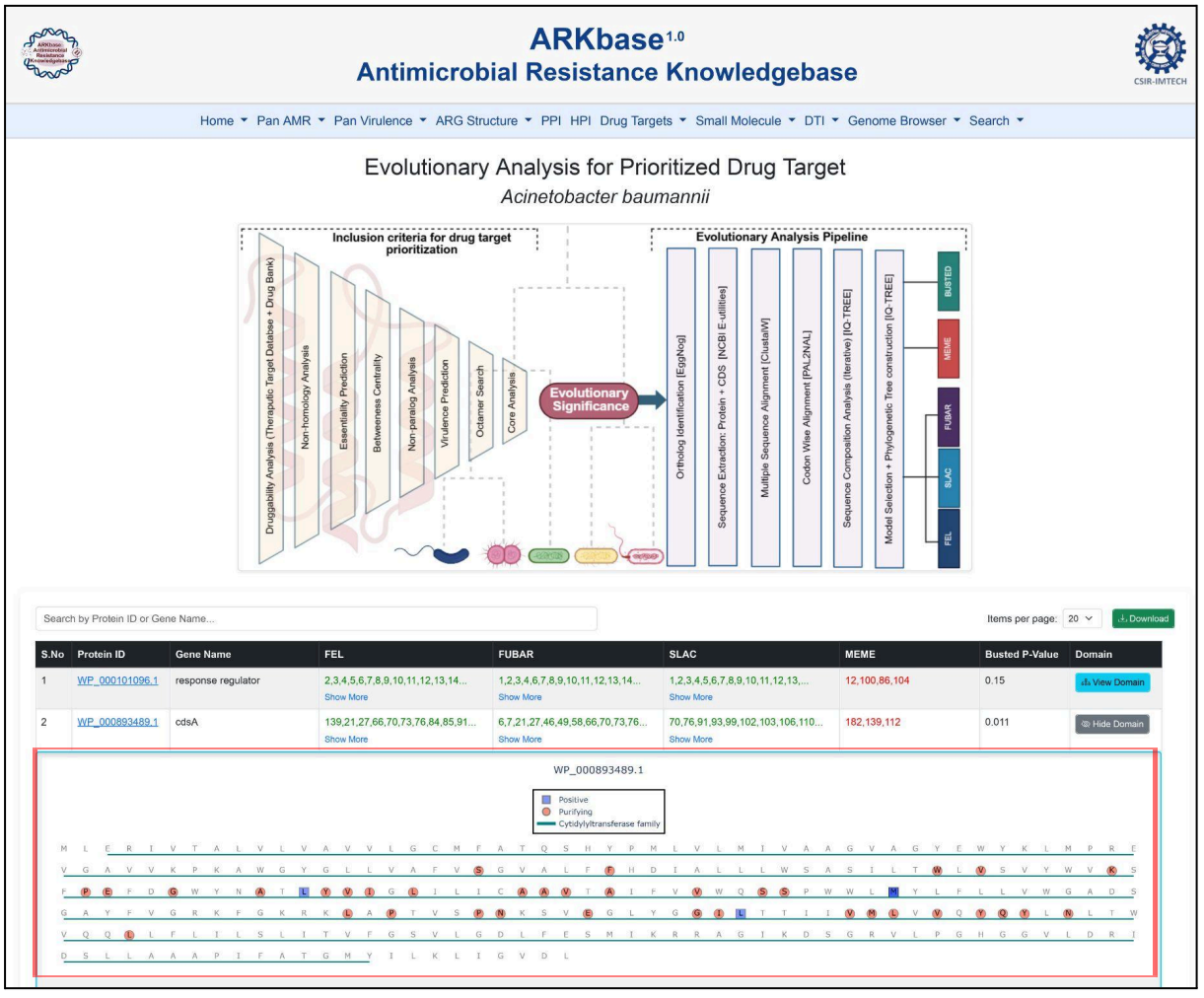
This module allows users to search by pathogen protein ID and their name (description). The user has the option to filter the targets based upon the provided properties like essentiality, virulence, centrality metrics, human non-homology, non-paralogous, and anti-target, along with their novelty based upon the Therapeutic Target Database (TTD) and DrugBank. The download tab allows the user to download the drug target datasets.



In this module, the user can also browse the evolutionary analysis of the prioritized drug target by selecting ‘Evolutionary analysis’.

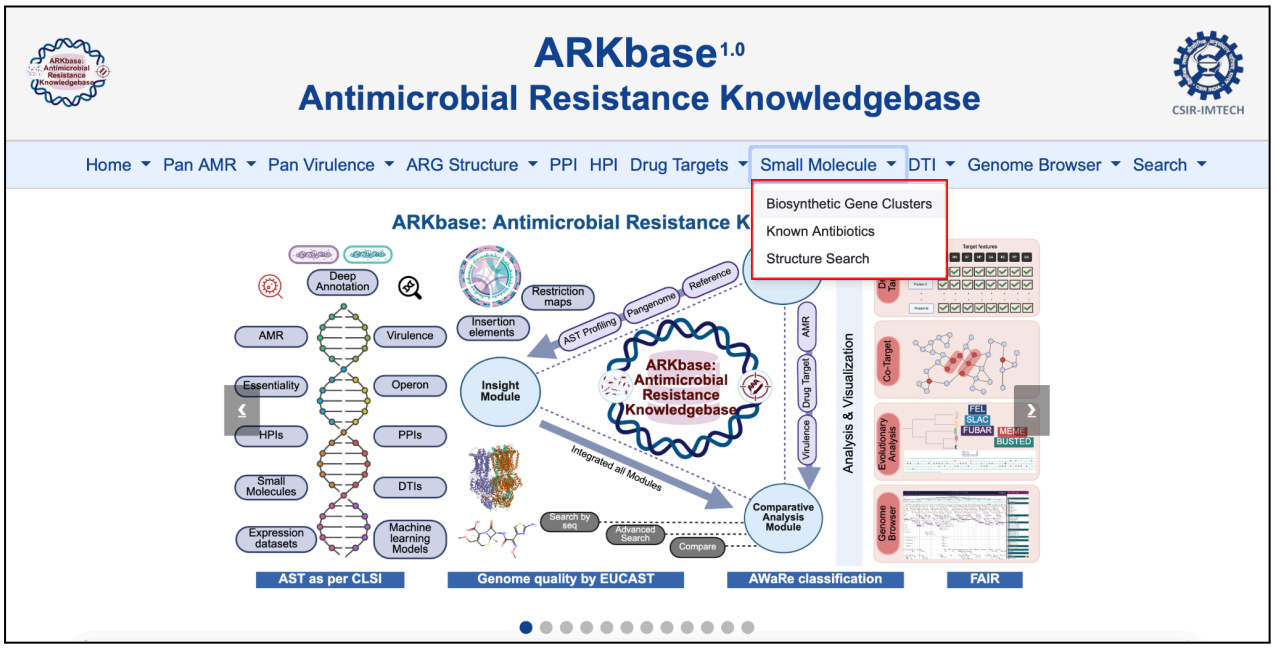


The user can search the data by search interface. In addition, by selecting the view domain, it provides an evolutionary constraint for each site per domain.



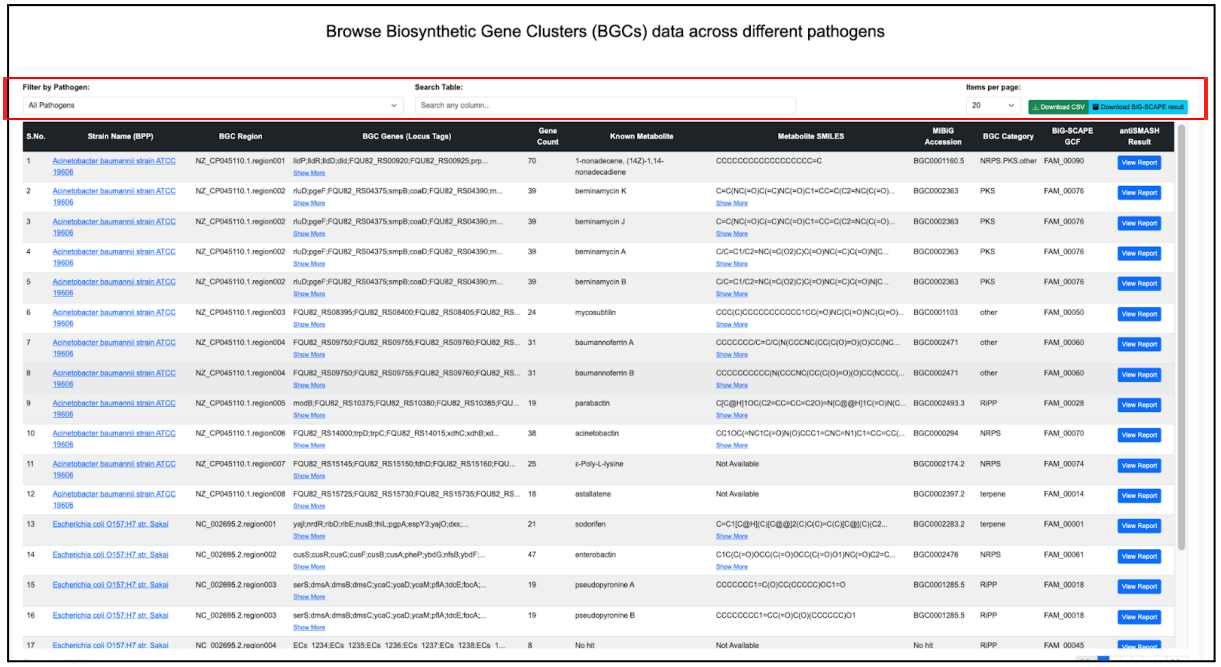
Small Molecule:

This section is divided into 3 submodules:
1. Biosynthetic Gene Clusters
2. Known Antibiotics
3. Structure Search

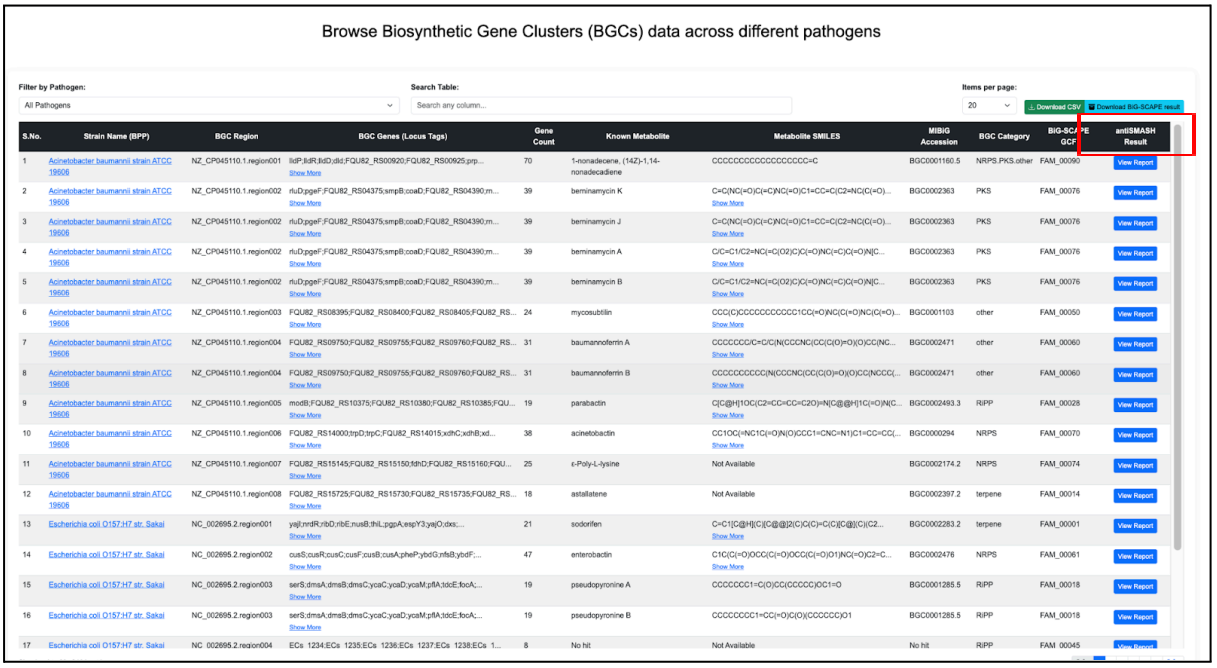


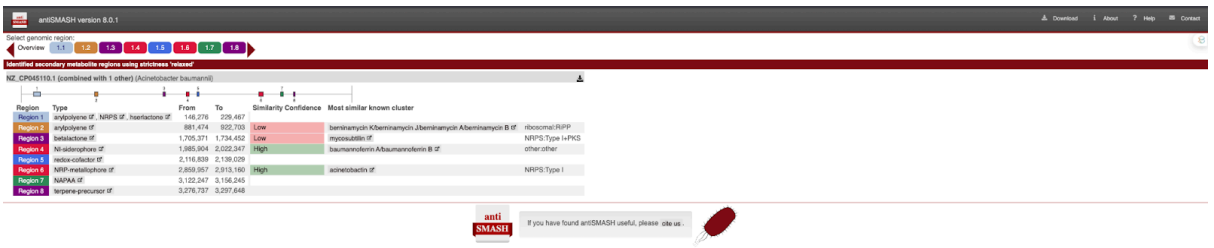
1. Biosynthetic Gene Clusters:

The module gives the user comprehensive data for BGCs—Strain Name, BGC region, BGC genes, Gene Count, Known Metabolites, SMILES, MIBiG Accession, BGC Category, BiG-SCAPE GCF, and antiSMASH Result. BGCs analysis identifies the genomic regions of each pathogen that produce metabolites conferring innate resistance. The user can filter data by pathogen or search any data field. There is also a download option, allowing users to download the entire table and the Bigscape clustering data.



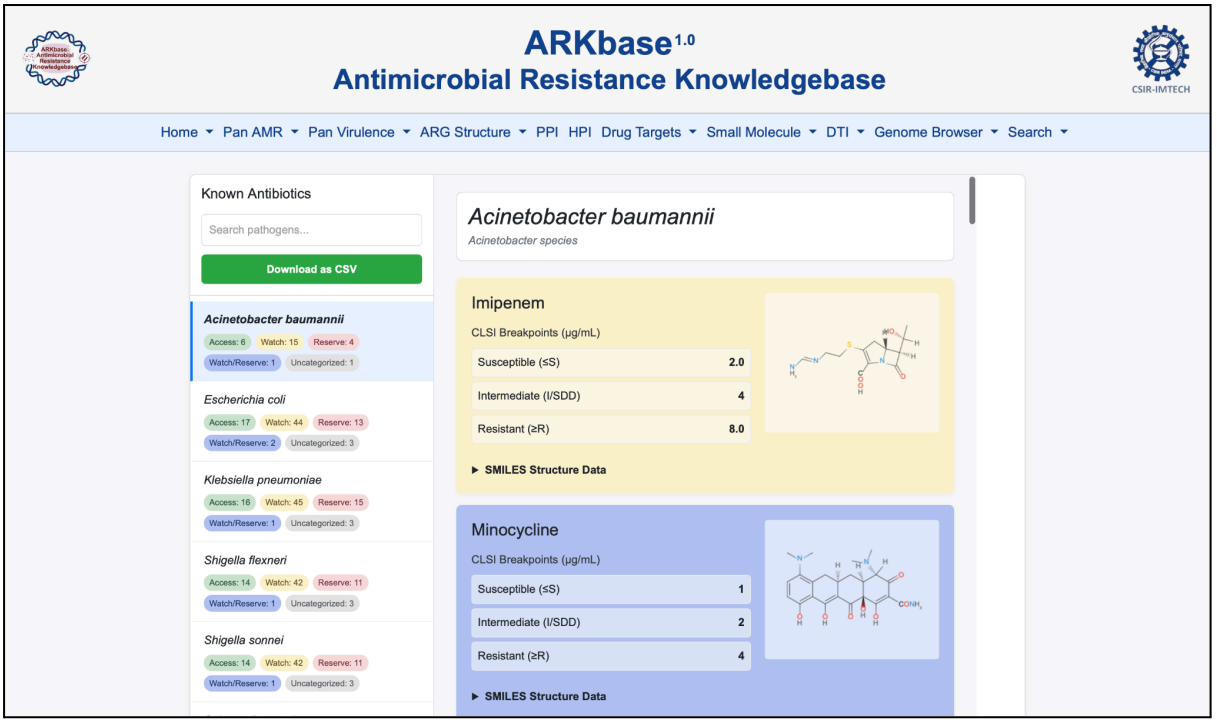
The user can view the antiSmash results by clicking View Report.





2. Known Antibiotics:

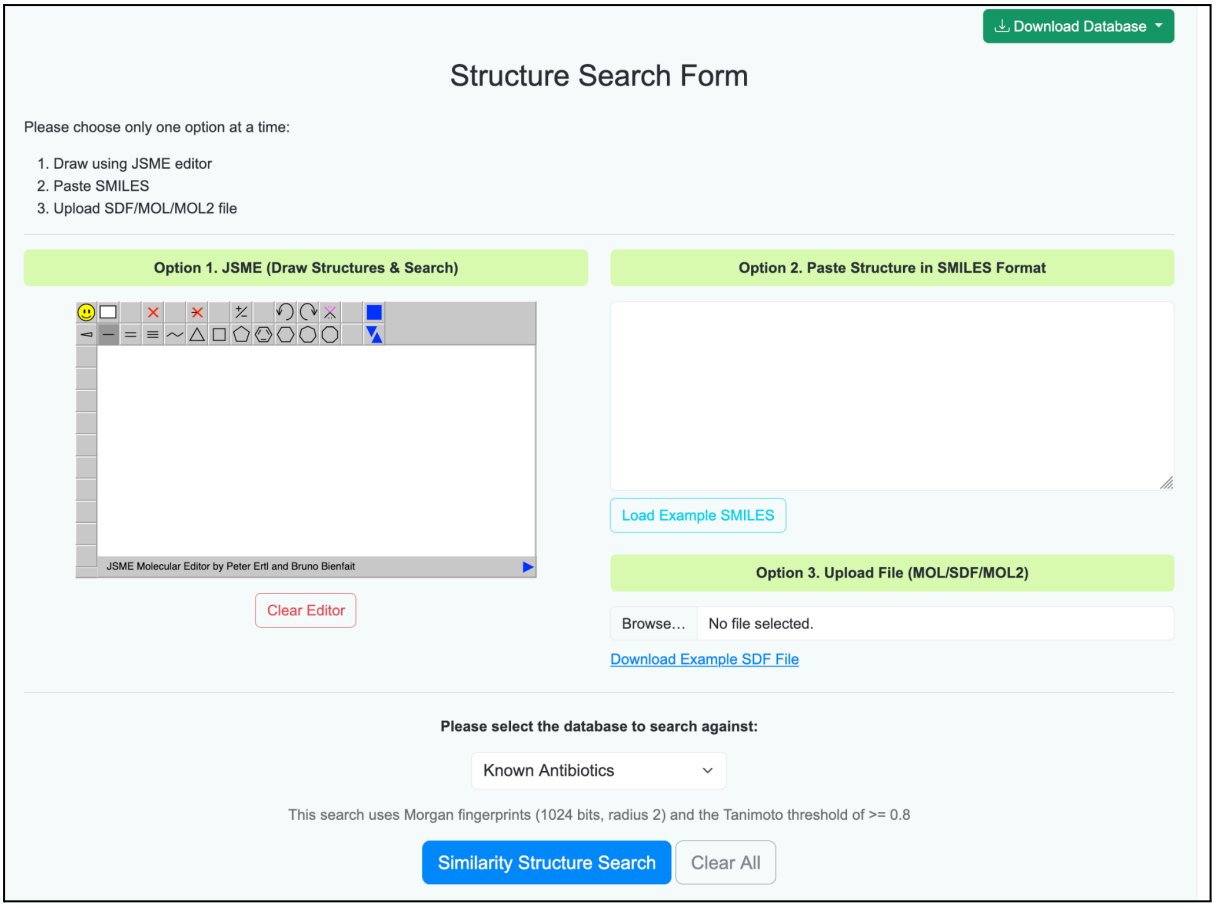
This page features a comprehensive repository of known antibiotics for 14 pathogens with MIC breakpoints, retrieved from the CLSI, with AWaRe annotation. The user can access the antibiotics associated with each pathogen from the left section. It features a search option that enables exploring all 14 pathogens. Additionally, the user can download the data in CSV format. The user can select any annotation, such as Reserve, to see the antibiotics listed under this category.



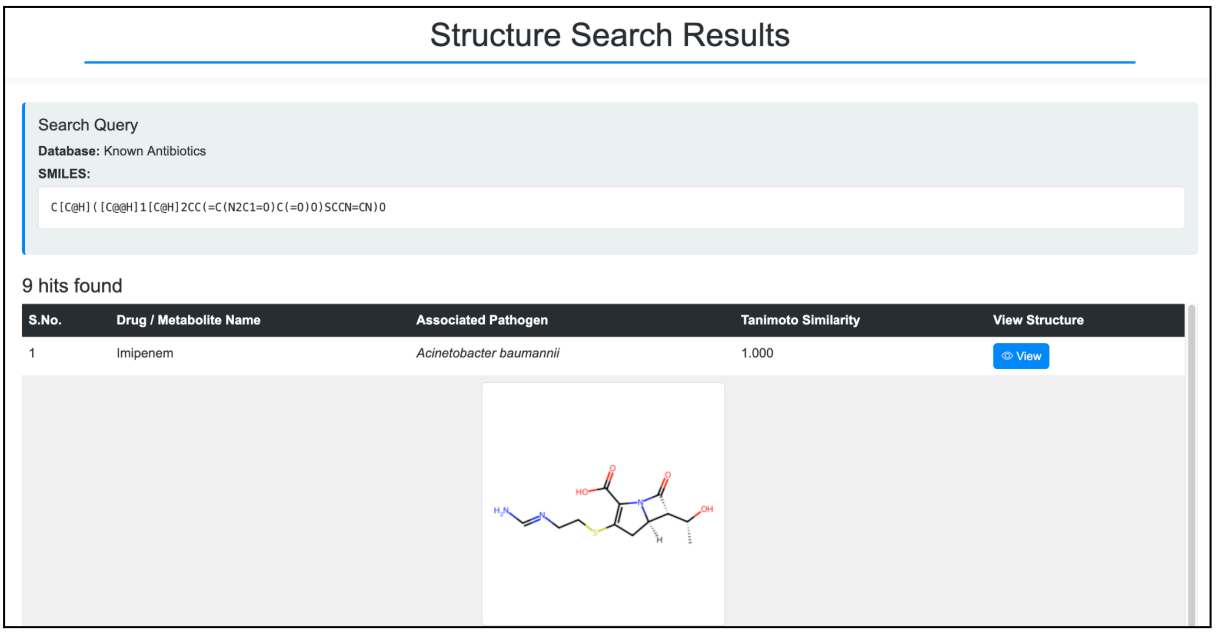
3. Structure Search:

This page assists users in identifying matches for their queried SMILES, MOL, or SDF against the structures of secondary metabolites predicted from BGC analysis and known antibiotics from the CLSI database. The user can perform a structure search using either the BGC library or the Known Antibiotic library. The user can also download the output result table as well as the whole library for BGCs and known antibiotics.

The search uses Morgan fingerprints (1024 bits, radius 2) and the Tanimoto threshold of >= 0.8



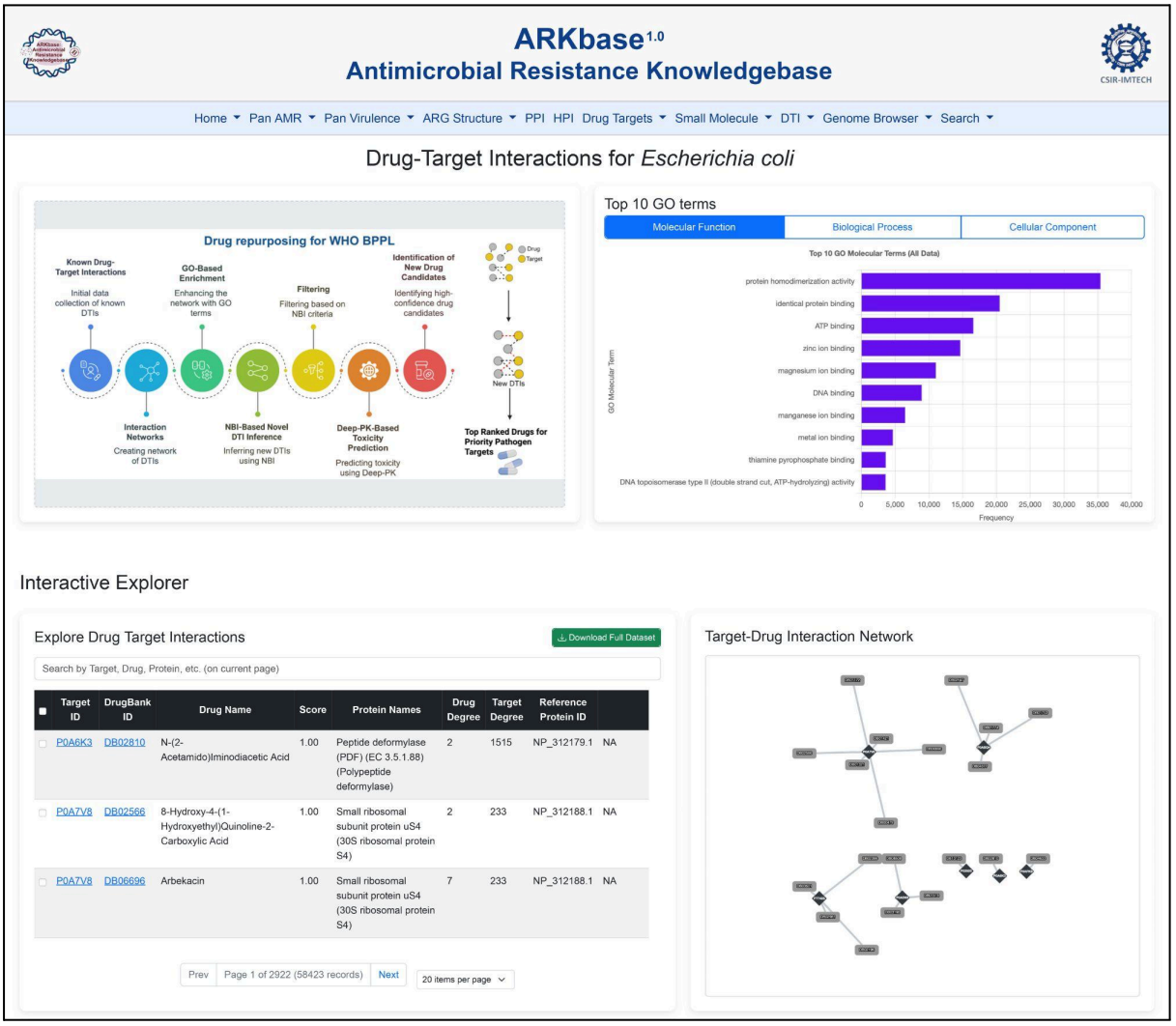
The result page shows all the possible hits along with their structure.



Drug-Target Interaction:

This pathogen-wise module provides an interactive, searchable data table for visualising a selected set of targets with their corresponding drugs. This module is divided into four sections:

1. Graphical abstract for the overall summary
2. Top 10 Gene Ontology Terms
3. Explore Drug Target Interactions
4. Target-Drug Interaction Network.

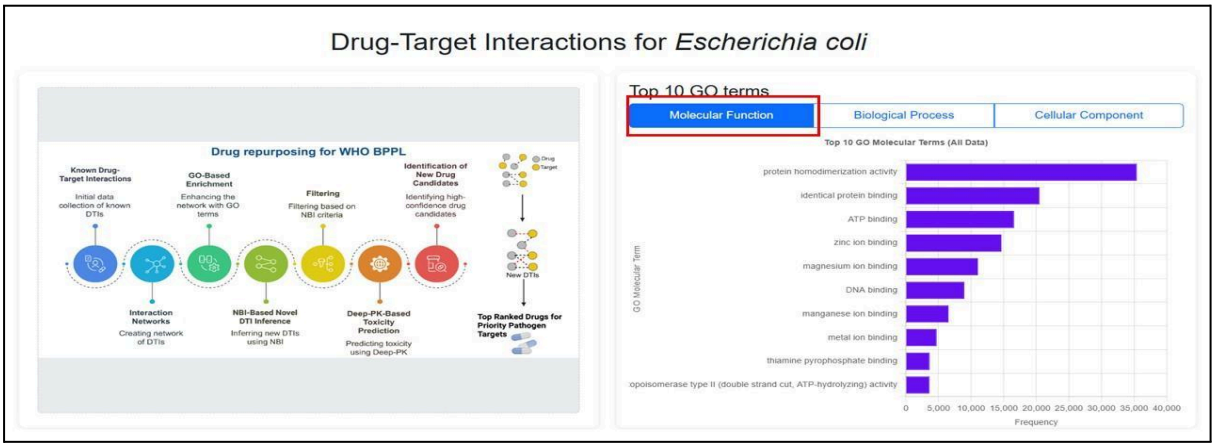


1. Graphical abstract for the overall summary:

This figure depicts the overall methodology and workflow of drug repurposing for the WHO Bacterial Priority Pathogen List (BPPL).

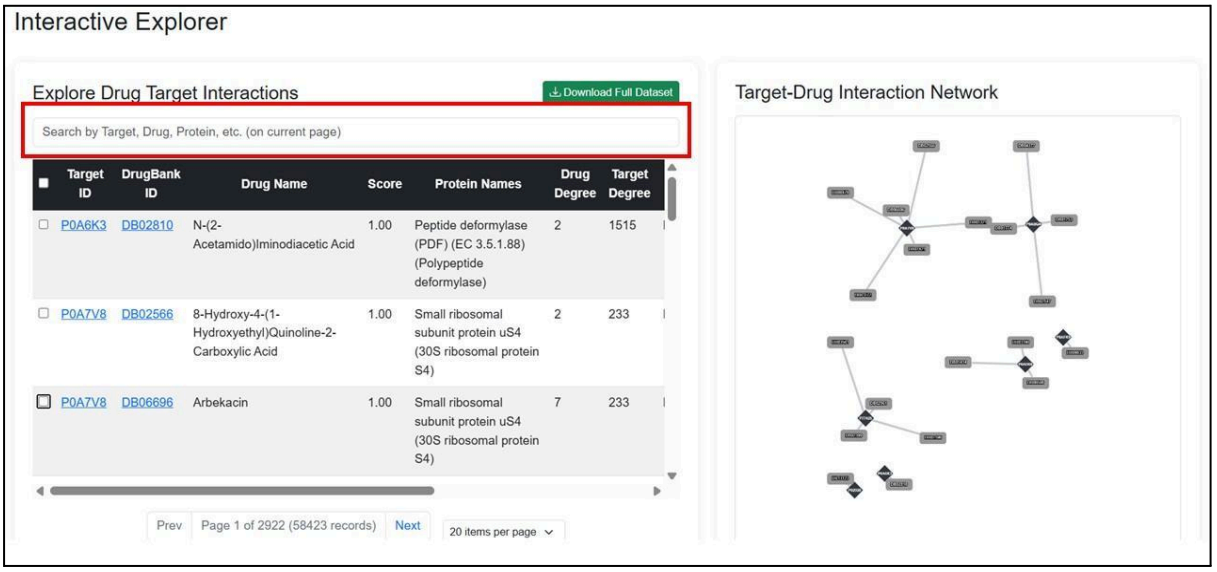
2. Top 10 Gene Ontology Terms:

This section displays the bar graph for the top 10 Gene Ontology terms. The user can select the three gene ontology terms - Molecular Function, Biological Process and Cellular Component.



3. Explore Drug Target Interactions:

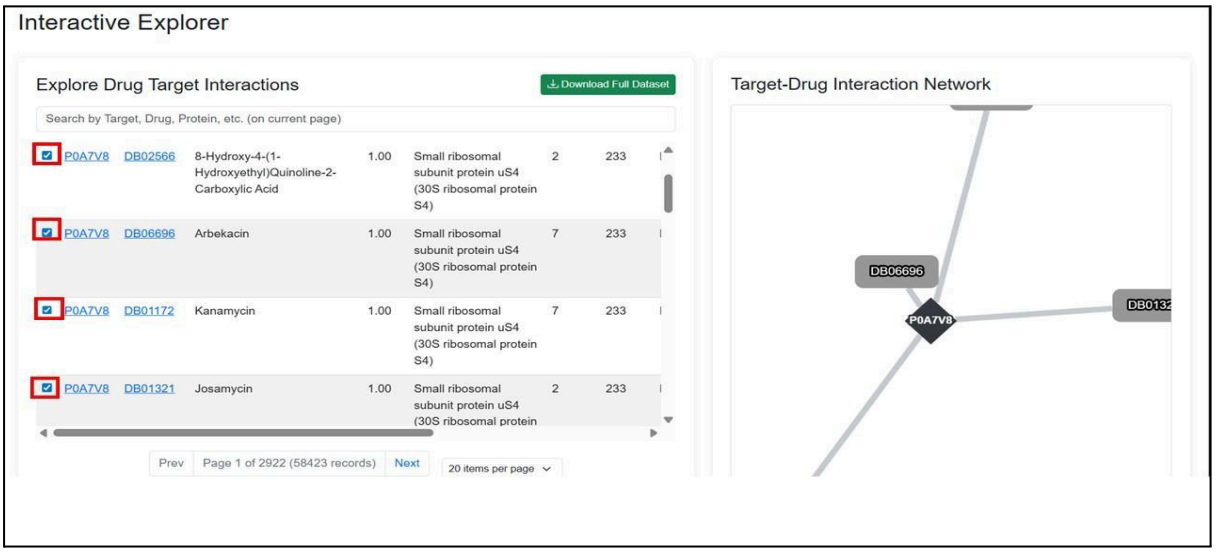
The user can select a pathogen from the drop-down menu in the Drug-Target Interaction module. The search interface is provided for exploring any data field. The results will display detailed information about each interaction, including the Target ID, DrugBank ID, drug name, score, protein names, drug and target degree, relevant pathways, and toxicity safety. Additionally, users can view the data in a network format.



The user can select one or multiple target IDs to explore drug-target interactions and the network of selected pathogens. The download tab allows the user to download the complete interaction datasets.

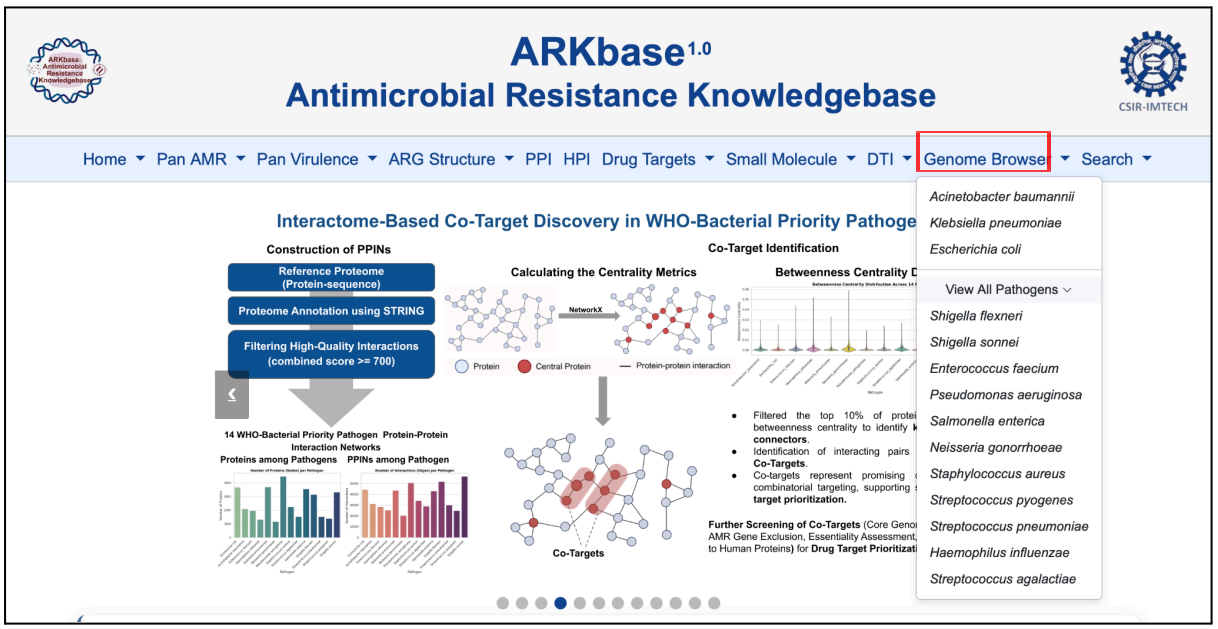
4.Target-Drug Interaction Network:

The interaction network visualizes connections between drug candidates (diamond nodes) and their bacterial protein targets (rectangular nodes), highlighting potential therapeutic relationships for WHO-priority pathogens.

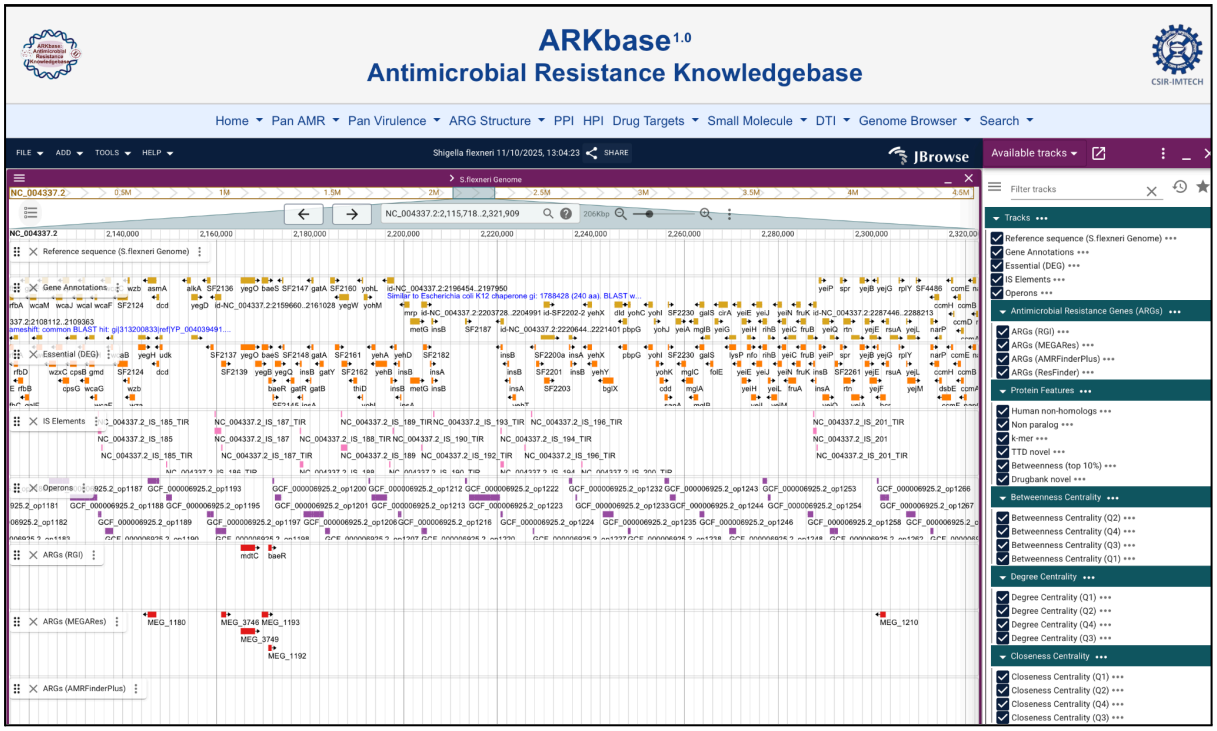


Genome Browser:

The Genome Browser module provides an interactive platform for visual exploration of the genomic features of 14 bacterial pathogens for the WHO Priority Pathogen List, 2024, included in this database. This tool enables users to visualise, navigate, and interpret genomic regions with ease, offering a detailed view of both reference sequences and associated annotation tracks.



The module covers genome data for 14 bacterial pathogens, curated from high-quality reference sources. 30+ annotation tracks are available for each pathogen for visualisation, representing a variety of data types such as gene annotation, operons, insertion elements, Antibiotic Resistance Genes (AMR), restriction sites, protein features, Centrality Metrics for protein-protein interactions (Degree, Closeness and Betweenness) and so on. The user can interactively navigate the browser for the chosen pathogen by zooming in/out to view genomic features at nucleotide or broader chromosomal scales, scrolling across genomic regions for comparative exploration and clicking on features to access detailed metadata.



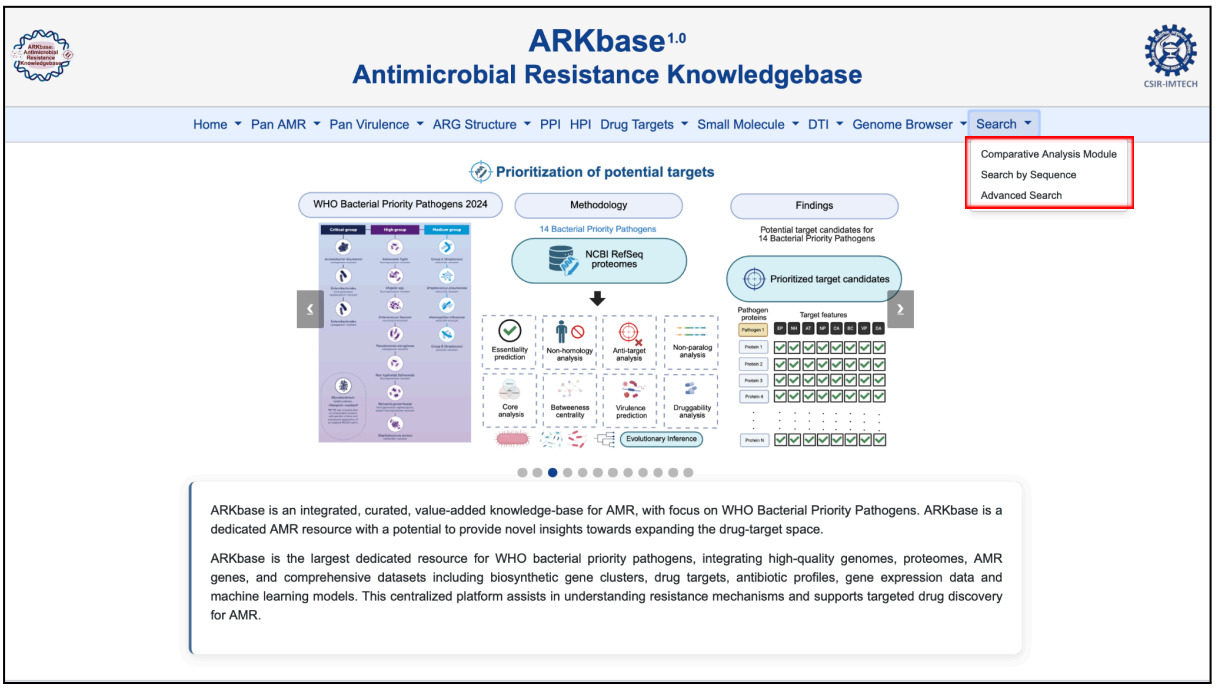
How to Use

1. Select an organism from the drop-down list or organism overview page.
2. Choose tracks of interest from the track selector to customize your view.
3. Use the zoom and pan controls or the search bar to locate genes, loci, or specific genomic coordinates.
4. Click any feature to open a pop-up with details, including functional annotation. Users can also hover over the features for their corresponding key information.
5. Download GFF files for the selected organism from the Download section for further analysis in genome analysis pipelines.

Data download of PanGenome annotation for each organism is available in GFF format for offline analysis. Download links are provided directly from the Genome Browser interface and annotation files are organized by pathogen for convenient bulk or individual retrieval.

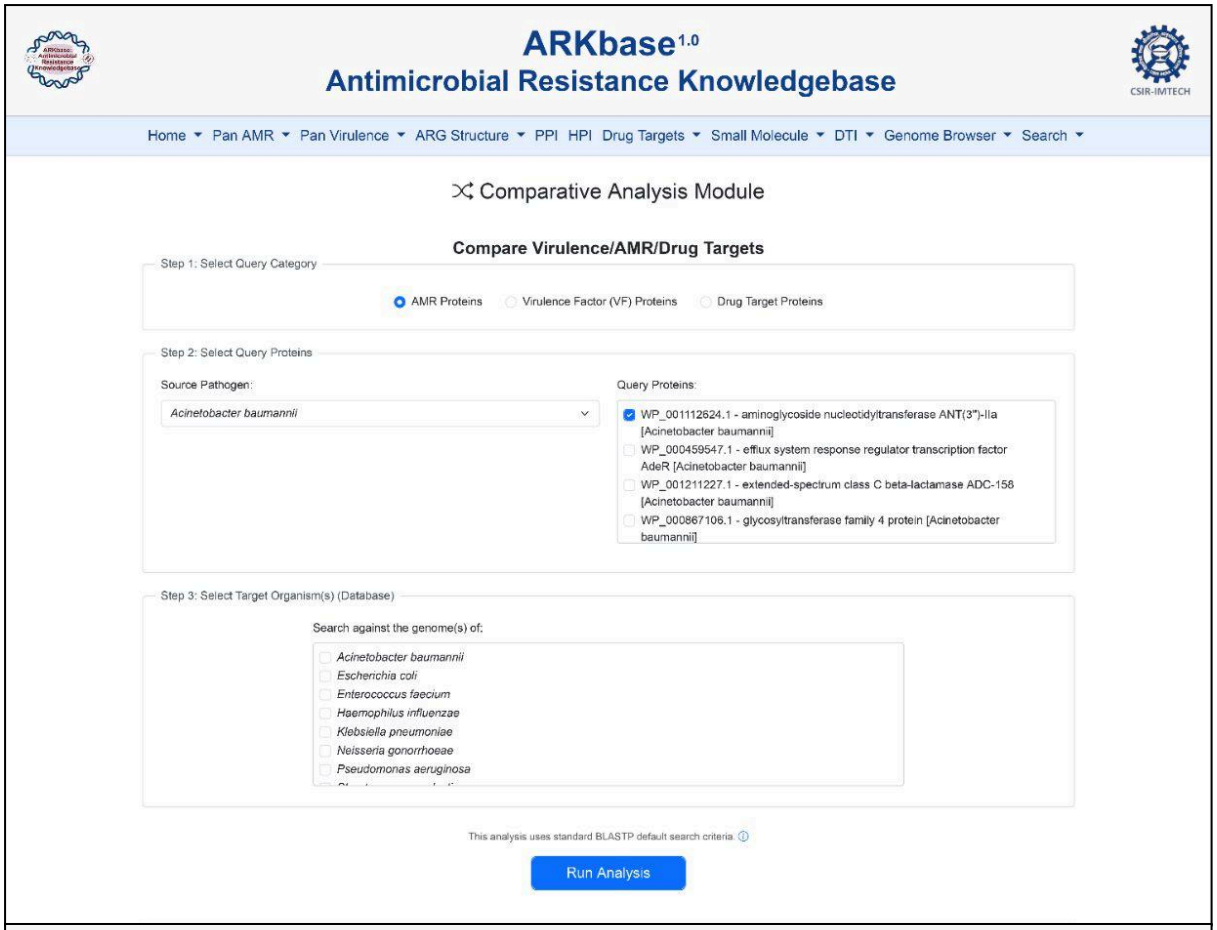
Search:

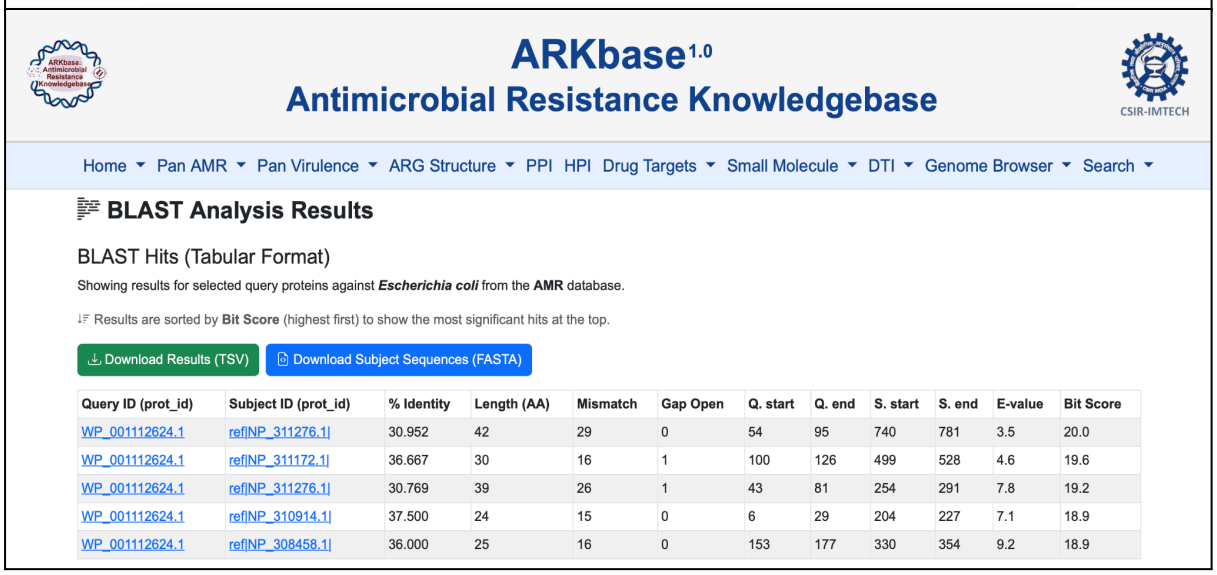
The search page is divided into 3 sections: Comparative Analysis Module, Search by Sequence and Advanced Search.



1. Comparative Analysis Module

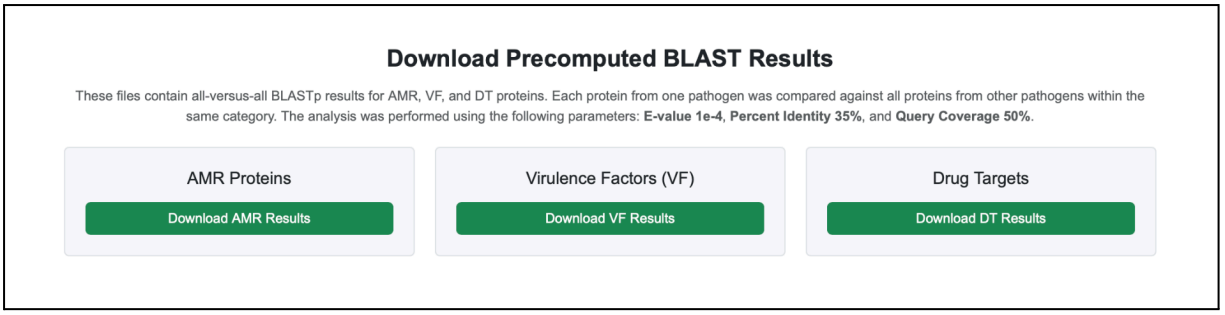
This module allows users to compare data related to virulence, antimicrobial resistance (AMR), and drug target proteins. Users can select a query category, choosing from AMR proteins, virulence factor (VF) proteins, or drug target proteins of any pathogen, and compare them against that same category for any other pathogen. For example, users can compare the AMR proteins of *Acinetobacter baumannii* with the AMR proteins of *Escherichia coli.* By selecting Run Analysis, the user can access the results in a new page. The result can be downloaded in tsv format.



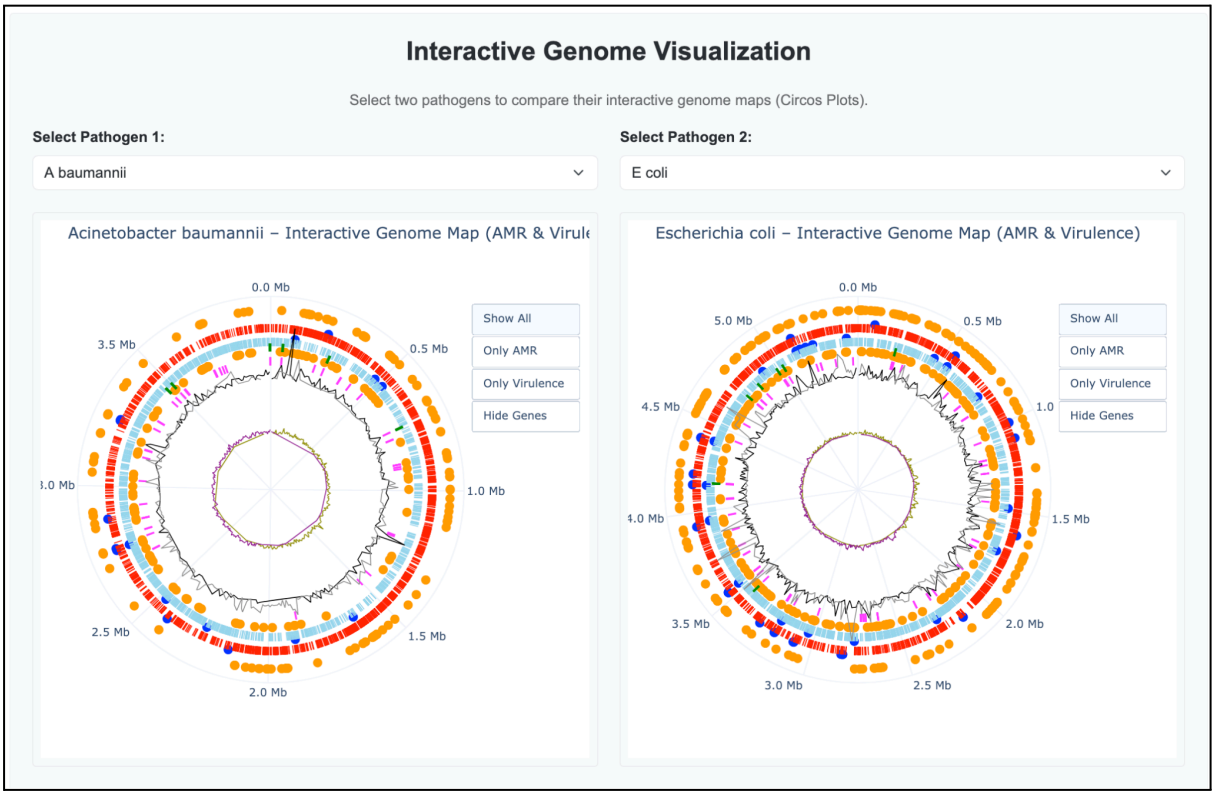


Query ID (prot_id)	Subject ID (prot_id)	% Identity	Length (AA)	Mismatch	Gap Open	Q. start	Q. end	S. start	S. end	E-value	Bit Score
WP_001112624.1	ref|NP_311276.1|	30.952	42	29	0	54	95	740	781	3.5	20.0
WP_001112624.1	ref|NP_311172.1|	36.667	30	16	1	100	126	499	528	4.6	19.6
WP_001112624.1	ref|NP_311276.1|	30.769	39	26	1	43	81	254	291	7.8	19.2
WP_001112624.1	ref|NP_310914.1|	37.500	24	15	0	6	29	204	227	7.1	18.9
WP_001112624.1	ref|NP_308458.1|	36.000	25	16	0	153	177	330	354	9.2	18.9

There is a download section with files containing the complete all-vs-all BLASTP analysis for each protein category, comparing every protein from one pathogen against all proteins from every other pathogen within the same category. The analysis was performed using the following parameters: E-value 1e-4, Percent Identity 35%, and Query Coverage 50%.

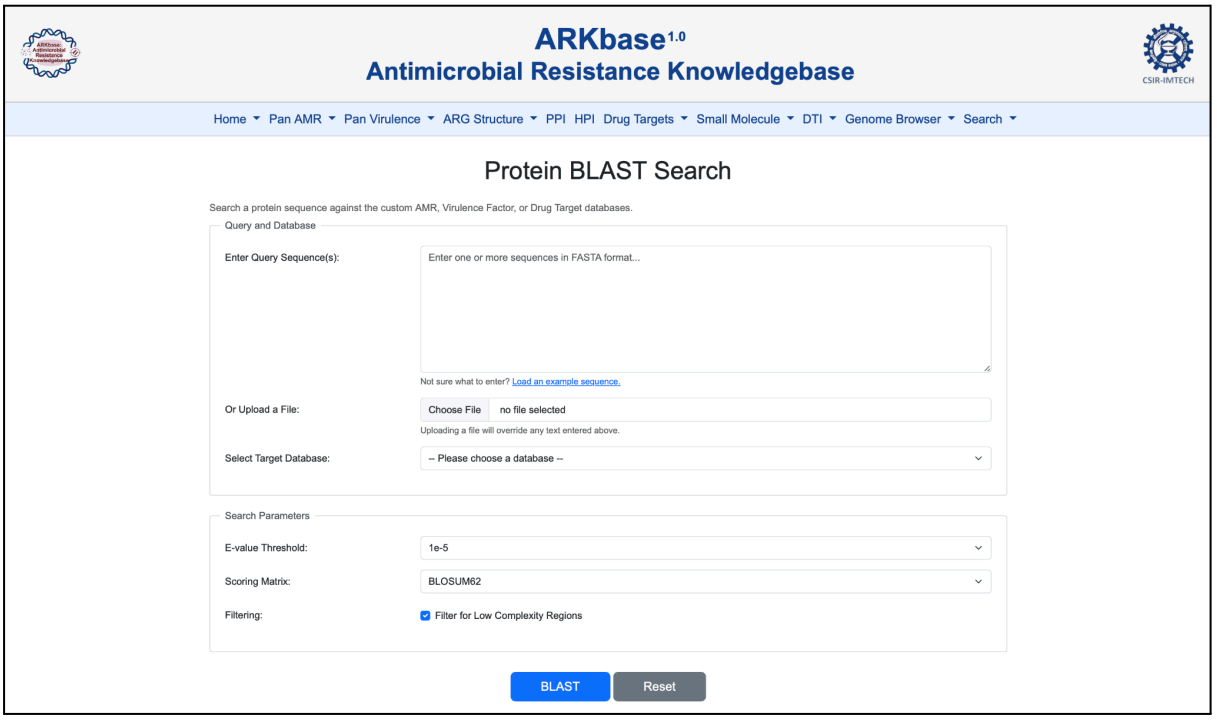


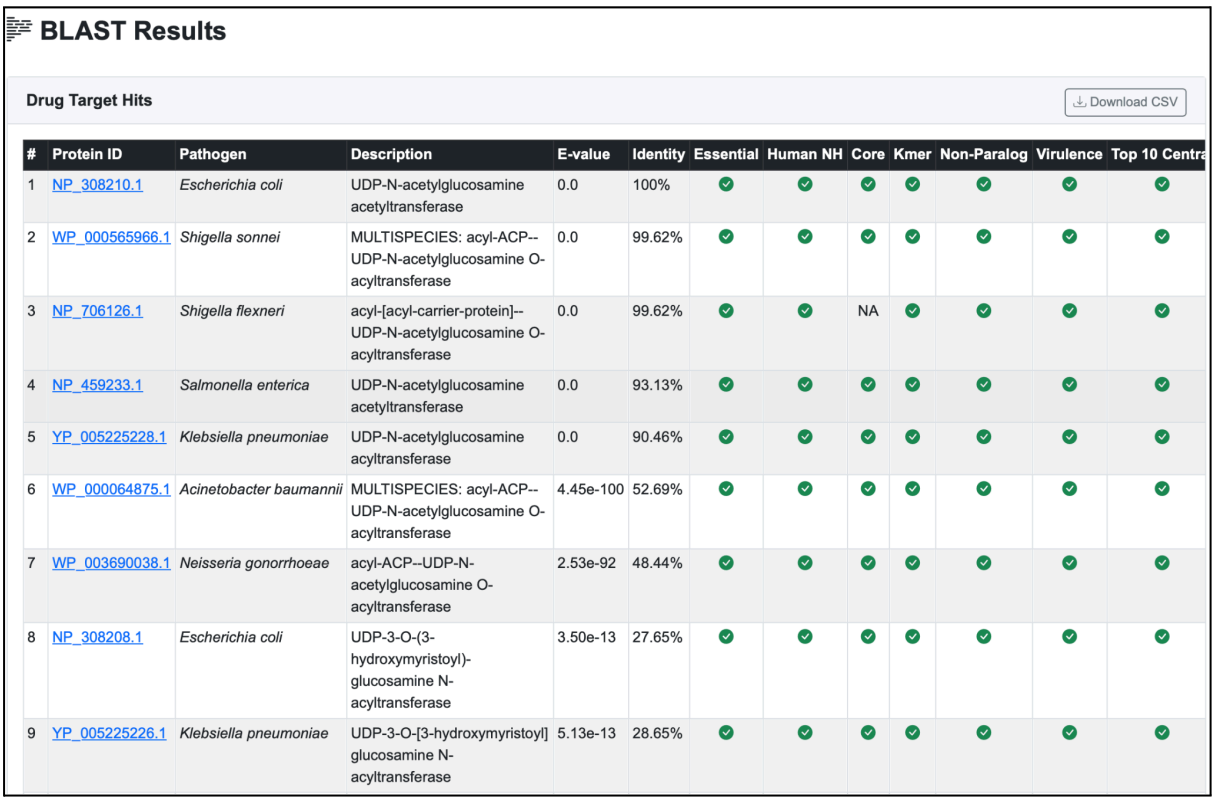
There is also an Interactive Genome Visualisation section featuring comparative genome maps (Circos plots). Users can interactively explore and select AMR genes, virulence genes, or all genes (AMR and Virulence together) for detailed viewing.



2. Search by Sequence :

This module allows users to search a query by entering or uploading a protein sequence, which is then searched against AMR, virulence, and drug target proteins.





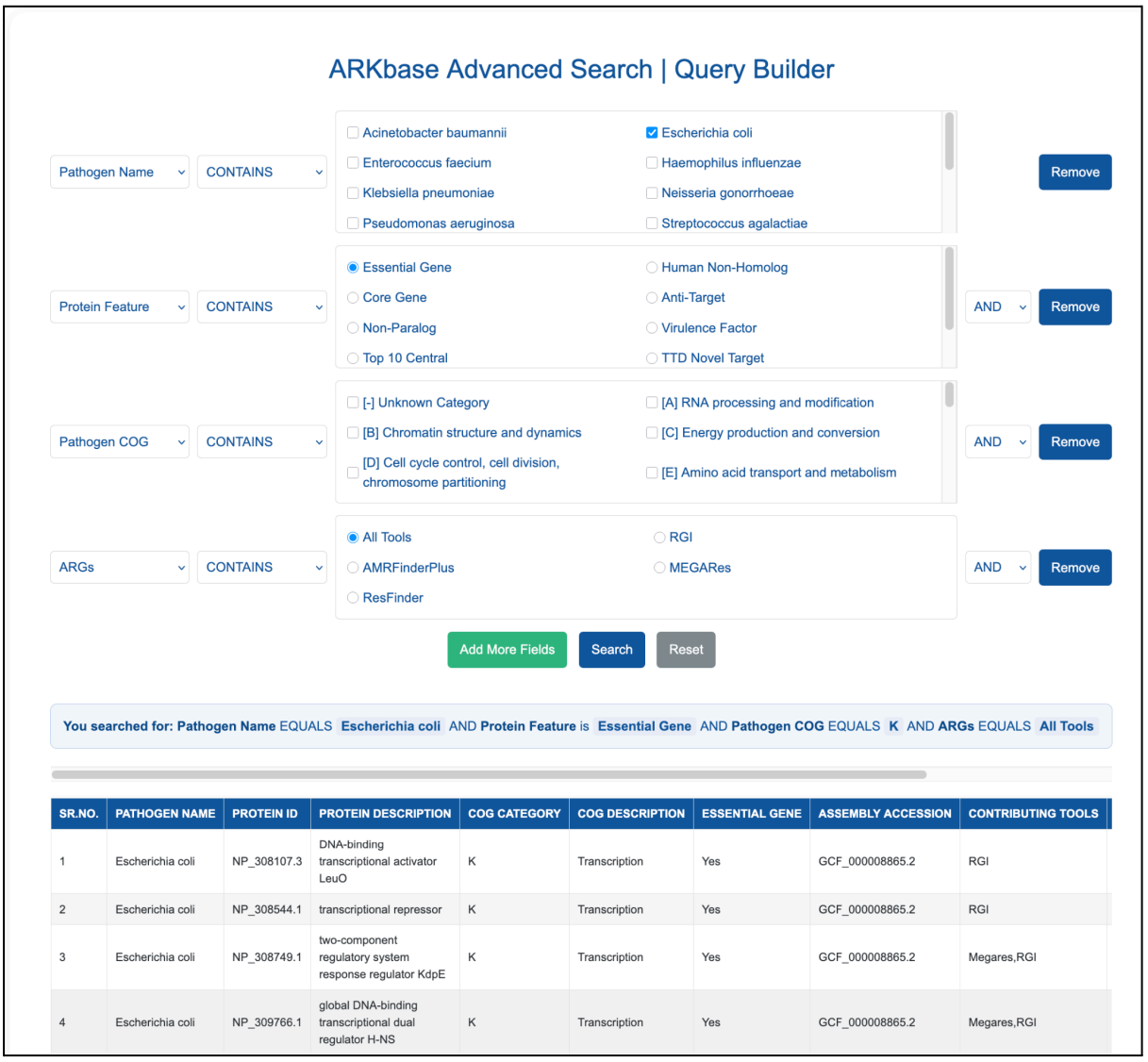
BLAST Results

Drug Target Hits — Download CSV

#	Protein ID	Pathogen	Description	E-value	Identity	Essential	Human NH	Core	Kmer	Non-Paralog	Virulence	Top 10 Centra
1	NP_308210.1	*Escherichia coli*	UDP-N-acetylglucosamine acetyltransferase	0.0	100%	✓	✓	✓	✓	✓	✓	✓
2	WP_000565966.1	*Shigella sonnei*	MULTISPECIES: acyl-ACP--UDP-N-acetylglucosamine O-acyltransferase	0.0	99.62%	✓	✓	✓	✓	✓	✓	✓
3	NP_706126.1	*Shigella flexneri*	acyl-[acyl-carrier-protein]--UDP-N-acetylglucosamine O-acyltransferase	0.0	99.62%	✓	✓	NA	✓	✓	✓	✓
4	NP_459233.1	*Salmonella enterica*	UDP-N-acetylglucosamine acetyltransferase	0.0	93.13%	✓	✓	✓	✓	✓	✓	✓
5	YP_005225228.1	*Klebsiella pneumoniae*	UDP-N-acetylglucosamine acyltransferase	0.0	90.46%	✓	✓	✓	✓	✓	✓	✓
6	WP_000064875.1	*Acinetobacter baumannii*	MULTISPECIES: acyl-ACP--UDP-N-acetylglucosamine O-acyltransferase	4.45e-100	52.69%	✓	✓	✓	✓	✓	✓	✓
7	WP_003690038.1	*Neisseria gonorrhoeae*	acyl-ACP--UDP-N-acetylglucosamine O-acyltransferase	2.53e-92	48.44%	✓	✓	✓	✓	✓	✓	✓
8	NP_308208.1	*Escherichia coli*	UDP-3-O-(3-hydroxymyristoyl)-glucosamine N-acyltransferase	3.50e-13	27.65%	✓	✓	✓	✓	✓	✓	✓
9	YP_005225226.1	*Klebsiella pneumoniae*	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	5.13e-13	28.65%	✓	✓	✓	✓	✓	✓	✓

3. Advanced Search

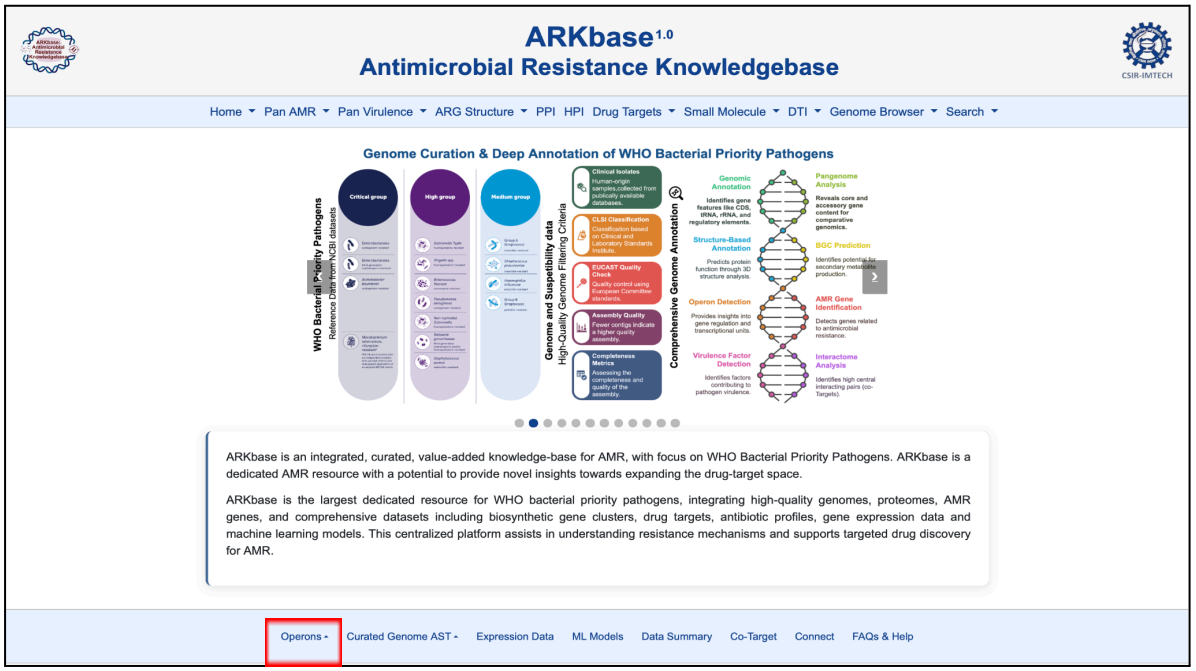
This module enables users to query based on 14 pathogens, protein features (e.g., essential, virulence, top 10 central, etc), COG categories, and ARGs. For example, users can identify proteins that are essential, virulence factors, or ARGs across all 14 pathogens.



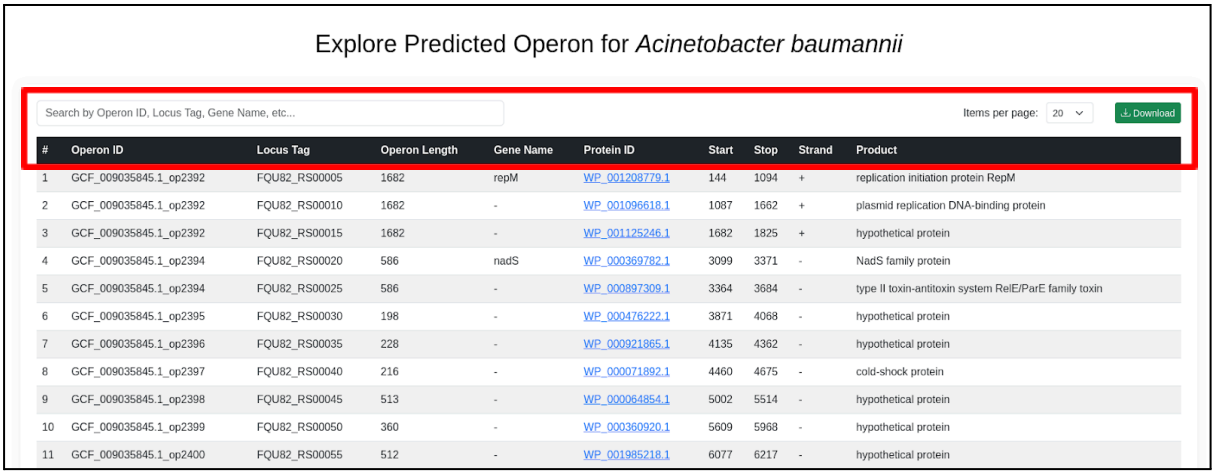
SR.NO.	PATHOGEN NAME	PROTEIN ID	PROTEIN DESCRIPTION	COG CATEGORY	COG DESCRIPTION	ESSENTIAL GENE	ASSEMBLY ACCESSION	CONTRIBUTING TOOLS
1	Escherichia coli	NP_308107.3	DNA-binding transcriptional activator LeuO	K	Transcription	Yes	GCF_000008865.2	RGI
2	Escherichia coli	NP_308544.1	transcriptional repressor	K	Transcription	Yes	GCF_000008865.2	RGI
3	Escherichia coli	NP_308749.1	two-component regulatory system response regulator KdpE	K	Transcription	Yes	GCF_000008865.2	Megares,RGI
4	Escherichia coli	NP_309766.1	global DNA-binding transcriptional dual regulator H-NS	K	Transcription	Yes	GCF_000008865.2	Megares,RGI

Operons:

This module allows the user to view the total number of operons present in the pathogen. The user can choose a pathogen from the drop-down menu to explore predicted operons. Features like the total length of the operon, locus tags of the gene belonging to the operon, operon_id, gene name with gene coordinates and strand info, protein ID, and the product have been provided. The protein_id is directly linked to the NCBI protein database.



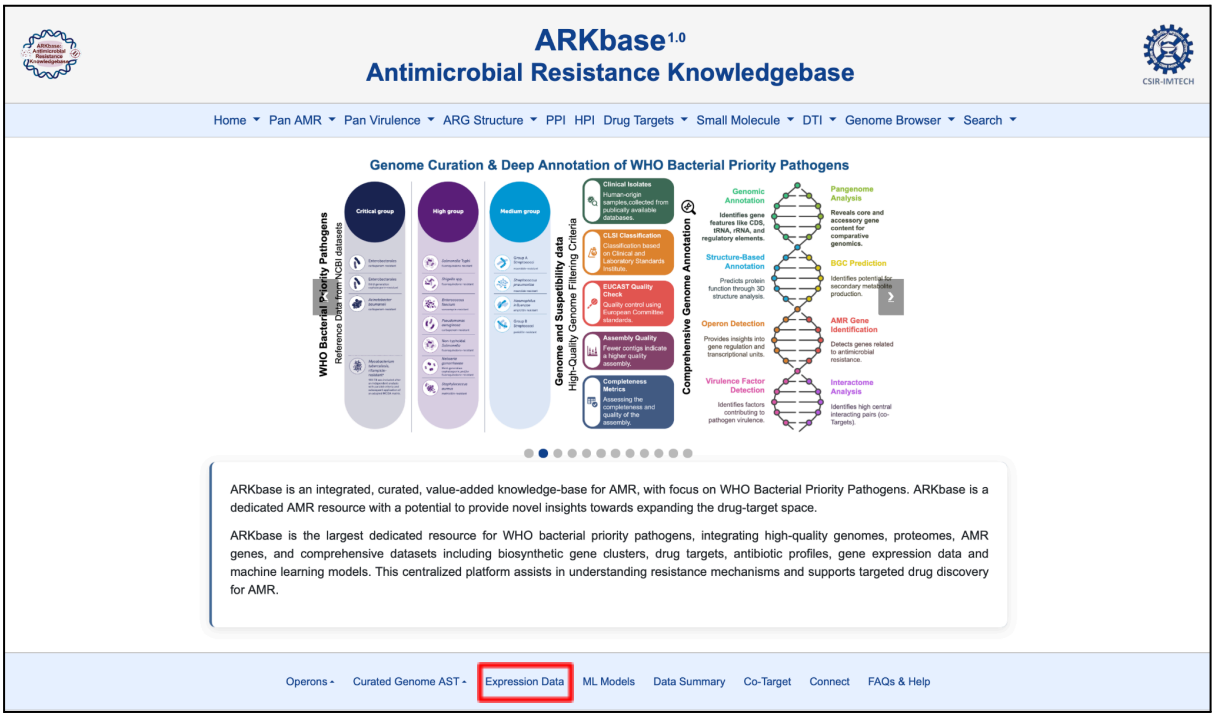
The module includes an interactive filter to search by Operon ID, Locus Tag, Gene name and so on. The download tab allows the user to download the complete operon summary for that particular genome.



#	Operon ID	Locus Tag	Operon Length	Gene Name	Protein ID	Start	Stop	Strand	Product
1	GCF_009035845.1_op2392	FQU82_RS00005	1682	repM	WP_001208779.1	144	1094	+	replication initiation protein RepM
2	GCF_009035845.1_op2392	FQU82_RS00010	1682	-	WP_001096618.1	1087	1662	+	plasmid replication DNA-binding protein
3	GCF_009035845.1_op2392	FQU82_RS00015	1682	-	WP_001125246.1	1682	1825	+	hypothetical protein
4	GCF_009035845.1_op2394	FQU82_RS00020	586	nadS	WP_000369782.1	3099	3371	-	NadS family protein
5	GCF_009035845.1_op2394	FQU82_RS00025	586	-	WP_000897309.1	3364	3684	-	type II toxin-antitoxin system RelE/ParE family toxin
6	GCF_009035845.1_op2395	FQU82_RS00030	198	-	WP_000476222.1	3871	4068	-	hypothetical protein
7	GCF_009035845.1_op2396	FQU82_RS00035	228	-	WP_000921865.1	4135	4362	-	hypothetical protein
8	GCF_009035845.1_op2397	FQU82_RS00040	216	-	WP_000071892.1	4460	4675	-	cold-shock protein
9	GCF_009035845.1_op2398	FQU82_RS00045	513	-	WP_000064854.1	5002	5514	-	hypothetical protein
10	GCF_009035845.1_op2399	FQU82_RS00050	360	-	WP_000360920.1	5609	5968	-	hypothetical protein
11	GCF_009035845.1_op2400	FQU82_RS00055	512	-	WP_001985218.1	6077	6217	-	hypothetical protein

Expression Data:

Transcriptomics encompasses the analysis of RNA sequences, their expression levels, and how they are regulated. This module is a compilation of all the datasets to date, involving pathogens exposed to their respective priority-resistant antibiotics. General information about the dataset, including project title, accession, release date, and sample size, has been provided.



The module includes two interactive filters for targeted data exploration:

- Priority antibiotic name
- Pathogen name

The user can also search pathogen and/or priority antibiotic combinations of their interest using the Browse tab.

ARKbase 1.0
Antimicrobial Resistance Knowledgebase

CSIR-IMTECH

Home | Pan AMR | Pan Virulence | ARG Structure | PPI | HPI | Drug Targets | Small Molecule | DTI | Genome Browser | Search

Transcriptomics Antibiotics Exposure

Antibiotic: -- All --

Organism: -- All --

Filter

Show 10 entries

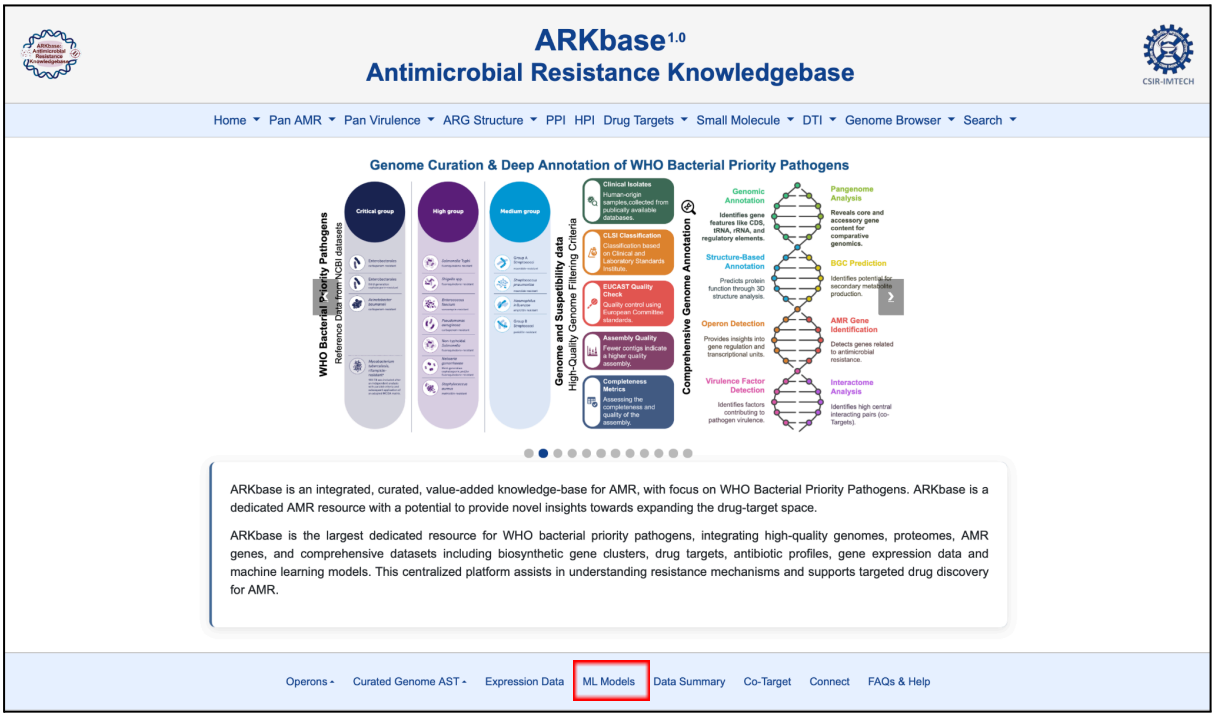
Browse:

Download CSV

#	Organism	Project Title	Project Database	Project Accession	Antibiotic Name	Release Date	Samples
1	*Acinetobacter baumannii*	The transcriptomic response of Acinetobacter baumannii to colistin and doripenem alone and in combination in an in vitro pharmacokinetics/pharmacodynamics model	GEO	PRJNA265162	Doripenem	Nov-30-2014	22
2	*Acinetobacter baumannii*	Transcriptional response to carbapenem treatment in Acinetobacter baumannii	GEO	PRJNA787205	Meropenem	May-20-2022	18
3	*Escherichia coli*	Prediction of antibiotic resistance by large-scale phenotypic and genotypic data	GEO	PRJNA255254	Cefoperazone, Cefixime	Jul-14-2014	42
4	*Escherichia coli*	Global gene expression profiling and antibiotic susceptibility after repeated exposure to the carbon monoxide-releasing molecule-2 (CORM-2) in multidrug-resistant ESBL-producing uropathogenic Escherichia coli	GEO	PRJNA345430	Cefotaxime	Oct-05-2016	16
5	*Escherichia coli*	Tellurite-mediated cefotaxime (CTX) potentiation	GEO	PRJNA171495	Cefotaxime	Jul-27-2012	9
6	*Escherichia coli*	Expression of Escherichia coli treated with cefsulodin and mecillinam	GEO	PRJNA108353	Cefsulodin	Jul-15-2008	61
7	*Escherichia coli*	Expression of Escherichia coli treated with cefsulodin and mecillinam, alone at the minimum inhibitory concentration	GEO	PRJNA109011	Cefsulodin	Jul-15-2008	18
8	*Escherichia coli*	Expression of Escherichia coli treated with cefsulodin and mecillinam, alone and in combination	GEO	PRJNA109009	Cefsulodin	Jul-15-2008	43
9	*Escherichia coli*	Comparative transcriptomic response analysis of Escherichia coli K-12 MG1655 to nine antibiotics	GEO	PRJNA910221	Ceftazidime, Imipenem	Dec-08-2022	33

Machine Learning (ML) Models:

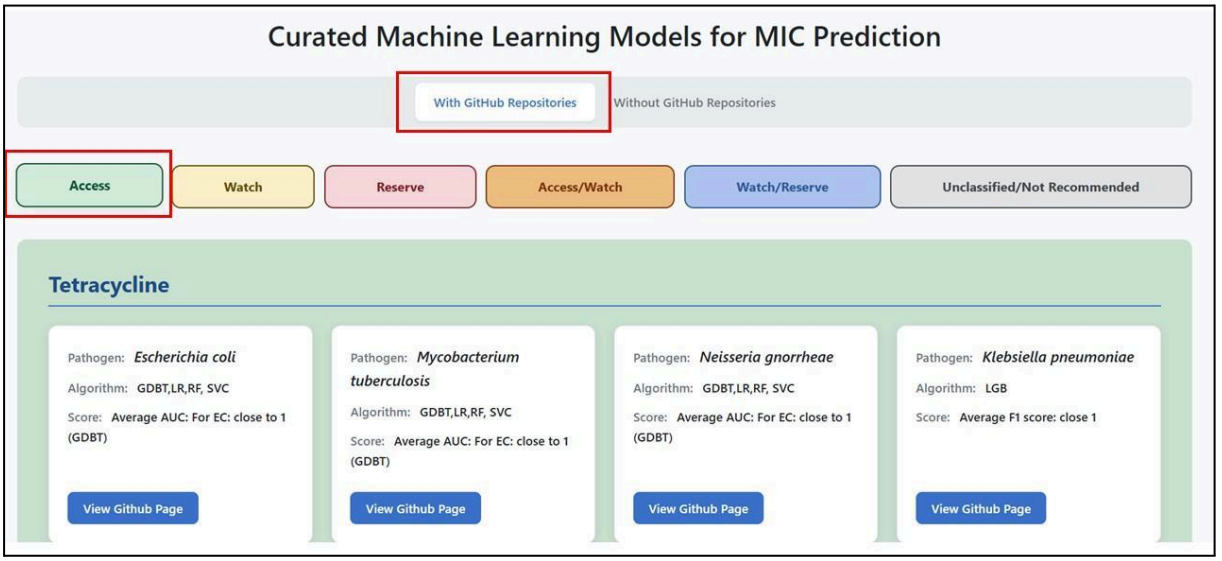
This module highlights machine learning models developed for MIC prediction targeting pathogens from the WHO Bacterial Priority Pathogen List.



The user can get an idea of what the different ML models reported for different antibiotics divided into AWaRe classes. Users can access the curated models with GitHub repositories or without GitHub repositories.

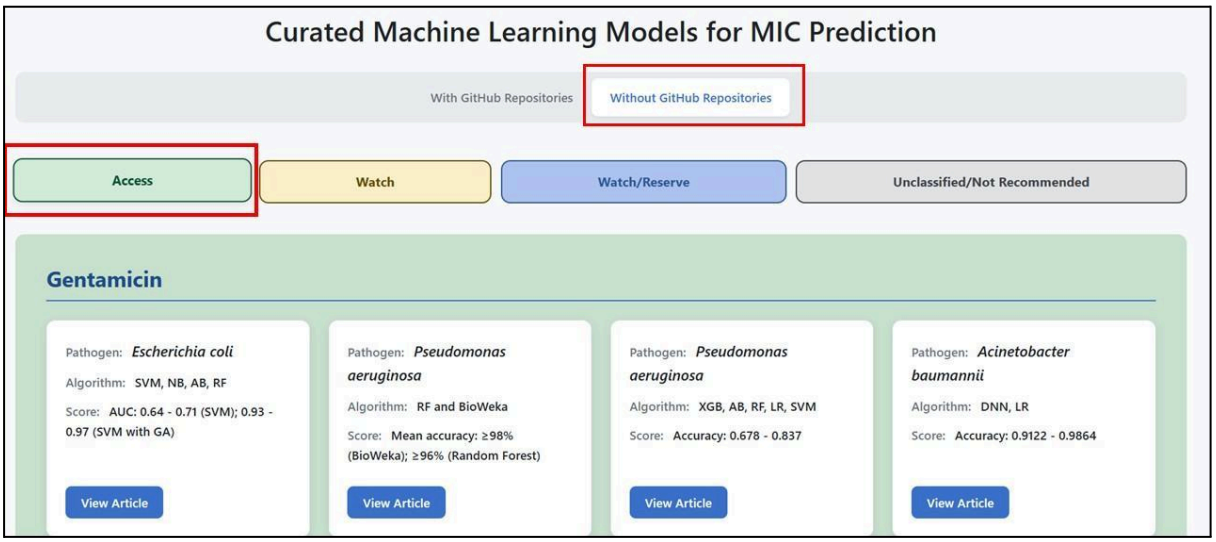
When user select **With GitHub repositories**, models are categorized into

- Access
- Watch
- Reserve
- Access/Watch
- Watch/Reserve
- Unclassified/Not recommended

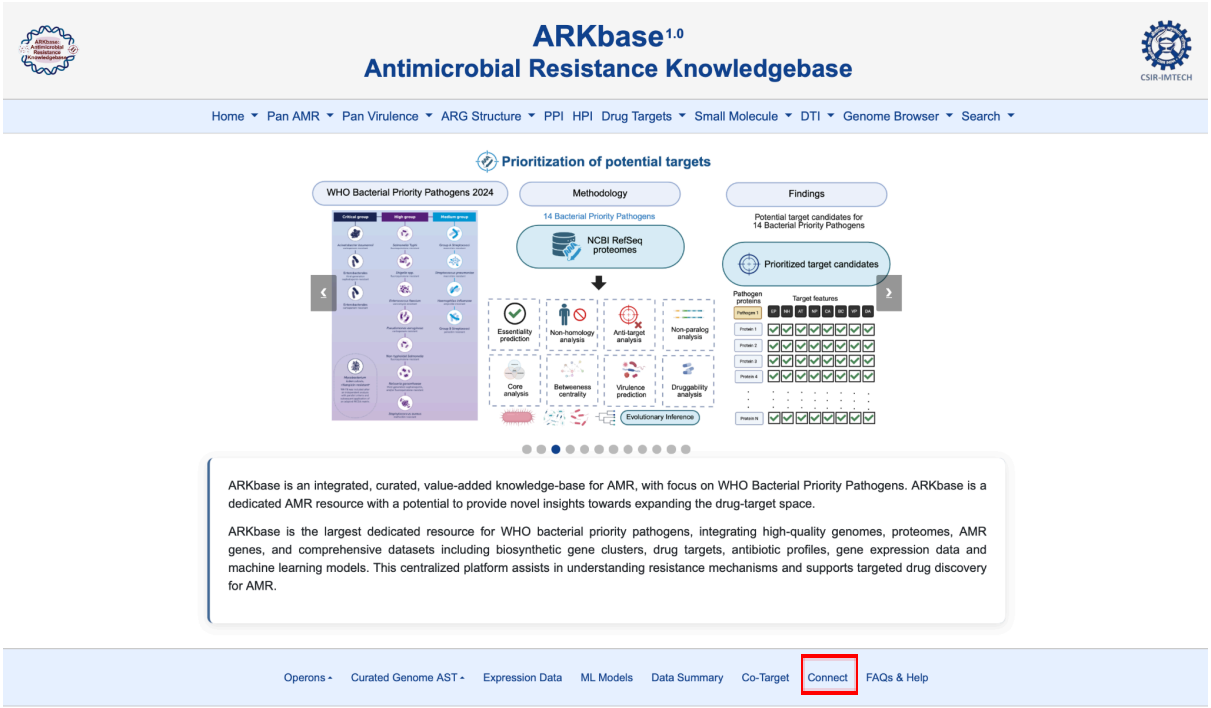


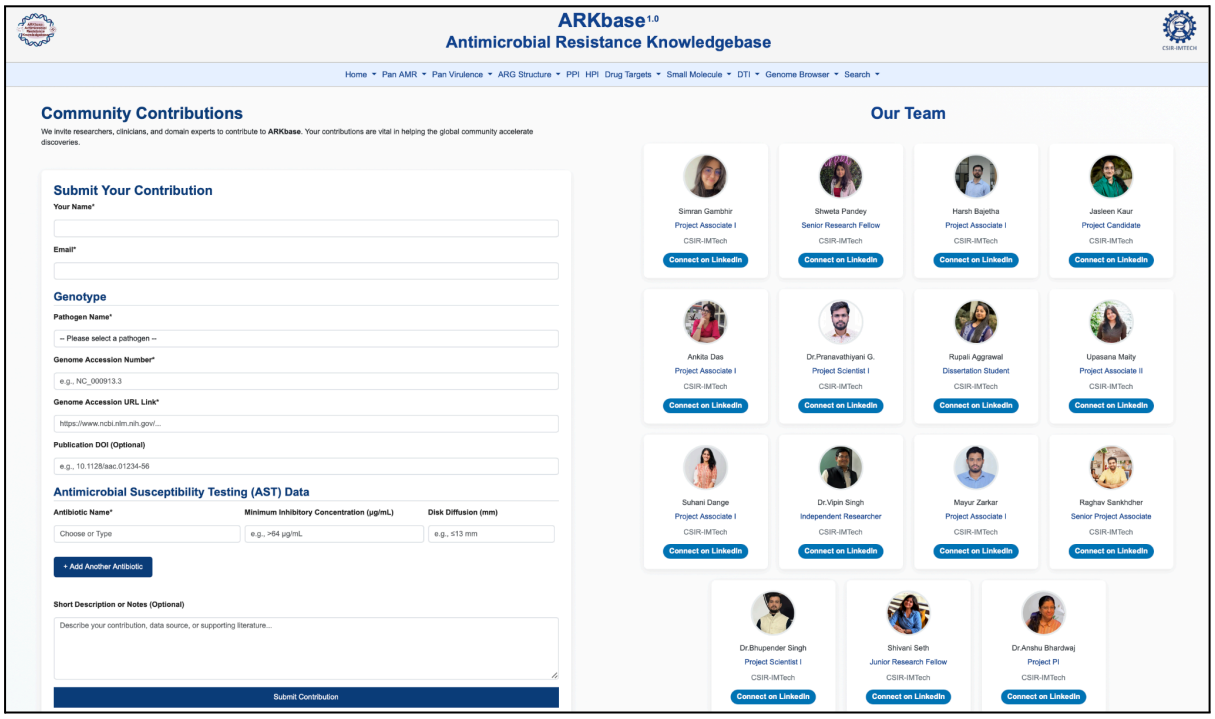
Users can view **Without GitHub repositories,** which are divided into

- Access
- Watch
- Watch/Reserve
- Unclassified/Not Recommended



Connect:





The Connect page is divided into two sections: the left section features community contributions, while the right section displays team members. The community contribution section invites researchers, clinicians, and domain experts to contribute to ARKbase, an open-access Antimicrobial Resistance Knowledgebase. The

submission allows contributions exclusively for genotype and phenotype data related to antimicrobial resistance in WHO Bacterial Priority Pathogens. For genotype data, users can select a specific pathogen from the BPPL list and provide a valid Genome Accession Number (e.g., from NCBI), and include the corresponding accession URL. For phenotype data, users must provide Antimicrobial Susceptibility Testing (AST) information for one or more antibiotics, including either Minimum Inhibitory Concentration (MIC) or Disk Diffusion (DD) values.

The unique features of ARKbase

ARKbase is a unique resource linking curated datasets with multi-omics, systems biology, AI/ML methods for predicting AMR determinants, and Foldseek/AlphaFold-based approaches towards understanding the molecular basis of resistance, identification of novel drug resistance determinants, and prioritizing potential therapeutic targets. It offers several unique features, like implementation of CLSI/EUCAST standards for reporting AST and genome datasets, identification of co-targets, providing a structured database for ARGs, offering a curated list of AI/ML models as per AWaRe classification, operon mapping of AMR and virulence genes, systems-level data for host-pathogen interactions, and PPIs for WHO priority pathogens.

ARKbase is an open database

Thank you