BV-BRC

Bacterial and Viral (BV) - Bioinformatics Resource Center (BRC)

Monthly Usage Metrics Report

Performance Period: December 1, 2023 – December 31, 2023

Issued to:

National Institute of Allergy and Infectious Diseases
National Institute of Health

Contract No.: 75N93019C00076
Contract Title: Bioinformatics Resource Centers for Infectious Diseases

Submission Date:
January 10, 2024

Submitted by:
Rick Stevens (contact PI)
Associate Laboratory Director
Argonne National Laboratory
Professor, Computer Science
University of Chicago
5801 South Ellis Avenue
Chicago, IL 60637-5418
630.252.3378 (phone)
630.252.6333 (fax)

BV-BRC Usage Metrics Report

Website Usage Metrics

Website usage is a key measure for evaluating use of the resource by the research communities. The number of website sessions unique users in a given period provide insights into trends, such as increased traffic resulting from outreach activities and prominent research topics and endeavors. Both the BRCs will use **AWStats** to monitor and track website usage by and report the number of unique visitors, visits, page views, pages/visit and visits/visitors for a given reporting period, aggregated across all constituent BRC websites, as summarized in the table below. In addition, we will also provide links to the live website usage statistics pages generated by AWStats from respective BRC websites, which will provide more detailed usage statistics by day of the week/month, country, browser / operating system, and more.

- **Total visits**
  - *Definition* - Number of visits made by all visitors. Think "session" here, say a unique IP accesses a page, and then requests three other pages within an hour. All of the "pages" are included in the visit, therefore you should expect multiple pages per visit and multiple visits per unique visitor (assuming that some of the unique IPs are logged with more than an hour between requests)
  - *Measurement mechanism* - AWStats.
  - *Measure* - Total number of visits per month.
- **Total unique visitors**
  - *Definition* - A unique visitor is a person or computer (host) that has made at least 1 hit on 1 page of your web site during the current period shown by the report. If this user makes several visits during this period, it is counted only once. Visitors are tracked by IP address, so if multiple users are accessing your site from the same IP (such as a home or office network), they will be counted as a single unique visitor.
  - *Measurement mechanism* - AWStats.
  - *Measure* - Total number of unique visitors per month.
- **Total page views**
  - *Definition* - The number of "pages" viewed by visitors. Pages are usually HTML, PHP or ASP files, not images or other files requested as a result of loading a "Page" (like js,css... files).
  - *Measurement mechanism* - AWStats.
  - *Measure* - Total pageviews per month.
- **Average pages per visit**
  - *Definition* - The average number of pages viewed during a visit. Repeated views of a single page are counted.
  - *Measurement mechanism* - AWStats.
  - *Measure* - Average number of pages per visit per month.
- **Average visits per visitor**
  - *Definition* - The average number of visits per visitor.
  - *Measurement mechanism* - AWStats.
  - *Measure* - Average number of visits per visitor per month.
- **Average visit duration**
  - *Definition* - The average time a visitor spent on the site for each visit, measured in seconds.
  - *Measurement mechanism* - AWStats.
  - *Measure* - Average visit duration per month.
- **Total bandwidth**

- *Definition* - Total number of bytes for pages, images and files downloaded by web browsing. This number includes traffic for web only (or mail only, or ftp only depending on value of LogType). This number does not include technical header data size used inside the HTTP or HTTPS protocol or by protocols at a lower level (TCP, IP...). Note that this number is often lower than the bandwidth usually reported by internet providers as it is counted at a lower level and includes all IP and UDP traffic.
- *Measurement mechanism* - AWStats.
- *Measure* - Total bandwidth per month.

- **Registered users that run a service**
  - *Definition* – Total number of unique registered users that run an analysis service (requiring login) during the month.
  - *Measurement mechanism* – Service logs.
  - *Measure* - Total unique registered users per month.

Table 1. BV-BRC Website Usage Metrics

Metric	BV-BRC
Total visits	153,961
Total unique visitors	23,608
Total pageviews	638,983
Avg. pages / visit	4.15
Avg. visits / visitor	6.52
Avg. visit duration (seconds)	574
Bandwidth (GB)	36.00
Registered users that run a service[1]	1346

Notes:

1. *Note: This measure only represents a fraction of the total usage by registered users because they may be doing other types of work on the site, either logged in or not.*

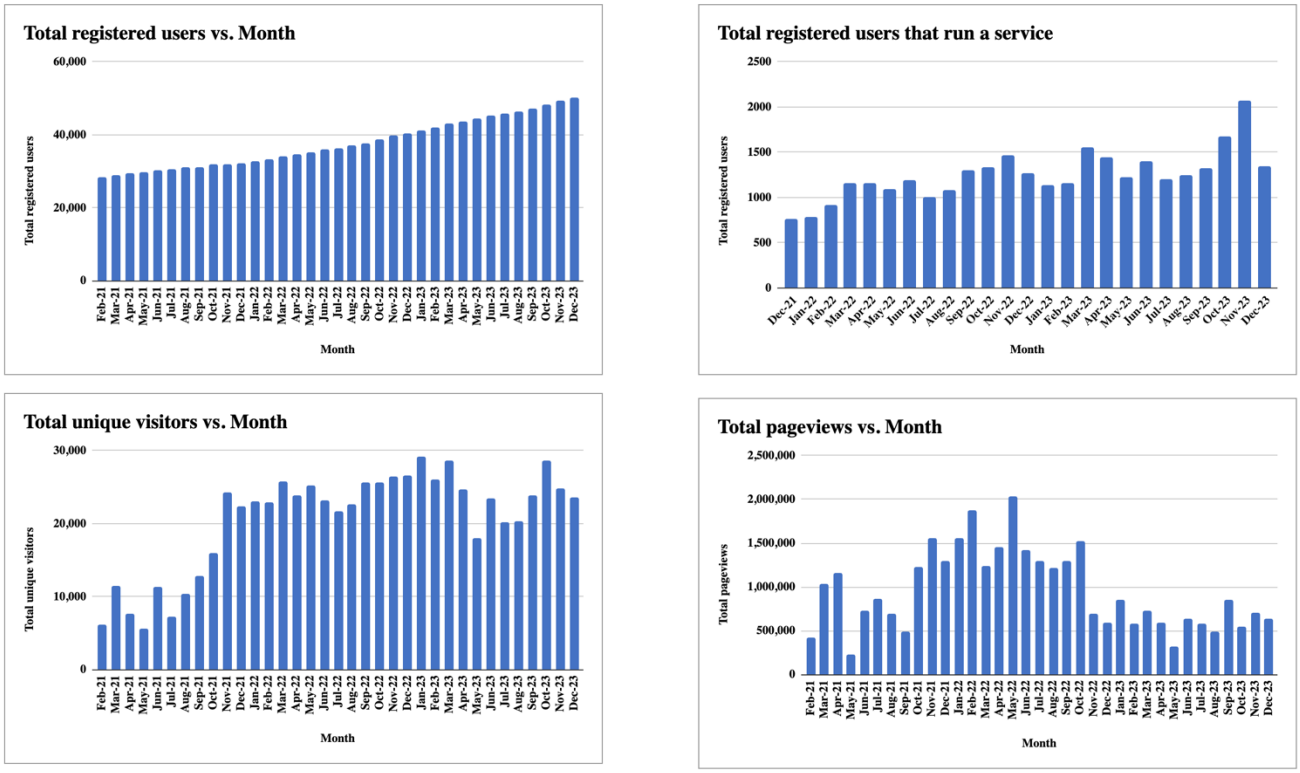

Figure 1. Selected BV-BRC website usage metrics.

Observations:

- We have crossed 50,000 registered users in BV-BRC. Through 2022 and 2023, an average of 745 new users registered per month in BV-BRC.
- Unique visitors per month remains stable at approximately 23,000
- Pageviews per month remains stable at approximately 640,000. The increased number of pageviews in the October 2021 through October 2022 is largely due to the transition of the legacy PATRIC, IRD, and ViPR users to BV-BRC. During that time we had all four resources in operation, and users were likely moving back and forth among the resources as we made new releases, announcements, and training materials.

Website Usage by Taxa

BRCs support a variety of organism taxa containing human pathogens and their vectors, along with related genomic and other omics data types. These taxa vary widely in the number of species and genomes they contain, availability of omics data, as well as the size of the research communities studying them. Measuring the BRC website usage by taxa allows us to understand how BRC resources are used by various organism communities. We will report the number of pageviews by taxa, which will be measured by querying the website usage statistics in Google Analytics by taxa name.

Table 2. BV-BRC Website Usage by Taxa

Taxa	Taxon ID	Domain	Species	Genomes	Page Views
Acinetobacter	469	Bacteria	763	17,223	642

Bacillus	1386	Bacteria	1,129	9,927	784
Bartonella	773	Bacteria	84	383	23
Borreliella	64895	Bacteria	24	5,880	8
Brucella	234	Bacteria	97	1,489	74
Burkholderia	32008	Bacteria	336	6,232	69
Campylobacter	194	Bacteria	341	10,323	191
Chlamydia	810	Bacteria	23	686	3
Clostridium	1485	Bacteria	495	7,265	147
Coxiella	776	Bacteria	15	213	6
Ehrlichia	943	Bacteria	7	46	5
Escherichia	561	Bacteria	198	55,630	1,605
Francisella	262	Bacteria	31	1,227	29
Helicobacter	209	Bacteria	95	4,567	60
Listeria	1637	Bacteria	34	6,725	21
Mycobacterium	1763	Bacteria	339	34,149	176
Pseudomonas	286	Bacteria	2,247	20,953	494
Rickettsia	780	Bacteria	78	443	20
Salmonella	590	Bacteria	399	34,522	419
Shigella	620	Bacteria	113	5,728	33
Staphylococcus	1279	Bacteria	582	29,547	525
Streptococcus	1301	Bacteria	464	41,682	211
Vibrio	662	Bacteria	488	8,201	162
Yersinia	629	Bacteria	35	1,761	83
Adenoviridae	10508	Virus	580	27,550	5
Asfarviridae	137992	Virus	4	12,290	6
Bunyaviridae	1980410	Virus	1,618	57,889	0
Caliciviridae	11974	Virus	258	69,701	22
Coronaviridae	11118	Virus	1,101	8,707,129	25
Filoviridae	11266	Virus	21	4,628	12

Flaviviridae	11050	Virus	544	390,479	51
Hepadnaviridae	10404	Virus	44	131,097	4
Hepeviridae	291484	Virus	96	25,724	43
Herpesviridae	10292	Virus	796	66,690	13
Orthomyxoviridae	11308	Virus	177	1,117,762	47
Paramyxoviridae	11158	Virus	716	68,361	35
Parvoviridae	10780	Virus	837	32,654	3
Picornaviridae	12058	Virus	1,192	170,352	19
Pneumoviridae	11244	Virus	15	61,167	18
Polyomaviridae	151341	Virus	277	13,786	2
Poxviridae	10240	Virus	283	20,735	22
Reoviridae	2732541	Virus	458	159,450	6
Rhabdoviridae	11270	Virus	725	40,436	1
Togaviridae	11018	Virus	70	14,059	6
SARS-CoV-2	2697049	Virus	1	8,655,025	157

Website Usage by Data Types

BRCs support genomic and a variety of other omics data types, providing an integrated view of these multi-omics data and related analysis tools. Tracking the website usage by primary data types allows us to understand how these data types are us. We will report the number of website page views by primary data types, which will be measured by querying the website usage statistics in Google Analytics by data type.

Table 3. BV-BRC Website Usage by Data Type

Data Type	BRC Domain	Page Views
Taxonomy	BV-BRC	5,210
Genome	BV-BRC	30,649
Feature (genes/proteins)	BV-BRC	8,723
Specialty (gene)	BV-BRC	597
Families (protein)	BV-BRC	18
Pathway	BV-BRC	1,306
Subsystem	BV-BRC	92

Transcriptomics	BV-BRC	15
Interactions	BV-BRC	21
Phylogeny	BV-BRC	96
Antibiotic	BV-BRC	5
Workspace (user data)	BV-BRC	88,038
Strain	BV-BRC	2,232
Epitope	BV-BRC	37
Ortholog	BV-BRC	0
Drug	BV-BRC	4
(Protein) structure	BV-BRC	41
Domain (/Motif)	BV-BRC	104
Plasmid	BV-BRC	17
SFVT	BV-BRC	0
Surveillance	BV-BRC	54
Serology	BV-BRC	12
Phenotype	BV-BRC	154
Primer	BV-BRC	118
Variant (SARS Variant Tracker)	BV-BRC	19

Service/Tool Usage

Both BRC analysis services and tools allow users to analyze data pulled from the respective BRC databases and their own private data, compare to other datasets, and save the results in their private workspaces. Since the types of tools vary across the BRCs, we will report aggregated usage of all tools in each BRC, and also a breakdown by service/tool. We will also report the total amount of storage used for user data. VIPR/IRD tools/services are combined (added together) that are common in both systems.

- **Total number of analysis tasks submitted and completed successfully by users**
  - *Definition* - The total number of analysis tasks submitted and completed successfully by users for a given month. An analysis task usually involves users providing input data/search terms and/or parameters to initiate a search or analysis task, which may perform one or more searches, data transformations, or data analysis steps, generate results that provide additional insights into the data and present it back to the user in structured view and/or file formats via web interface and/or user workspace.
  - *Measurement mechanism* - Analysis tasks are recorded via website and server logs, which are used to tally the number.
  - *Measure* - Analysis tasks submitted and completed successfully per month.
- **Analysis tasks submitted and successfully completed by service/tool**
  - *Definition* - A breakdown of total number of analysis tasks (see metric above), summarized by service/tool during the specified date range.

- *Measurement mechanism* - Analysis tasks submitted by users and successfully completed are captured via website and server logs, which are used to tally the number.
- *Measure* - Jobs per month, tallied by service/tool.

Table 4. BRC Tools/Services Usage Metrics

Tool/Service	BRC Domain	Jobs Submitted	Jobs Completed
Codon Tree	BV-BRC	357	346
Comparative Systems	BV-BRC	166	145
Comprehensive Genome Analysis	BV-BRC	2638	2299
Differential Expression	BV-BRC	3	3
FastqUtils	BV-BRC	1447	1304
Gene Tree	BV-BRC	105	77
Genome Alignment	BV-BRC	187	180
Genome Annotation	BV-BRC	3707	3338
Genome Assembly	BV-BRC	6050	5752
Genome Comparison	BV-BRC	166	139
HA Subtype Numbering	BV-BRC	100	55
Homology	BV-BRC	1522	1373
MSA	BV-BRC	370	321
MetaCATs	BV-BRC	41	36
Metagenome Binning	BV-BRC	435	378
Metagenomic Read Mapping	BV-BRC	450	446
Primer Design	BV-BRC	37	37
RNASeq Analysis	BV-BRC	1042	987
Sequence Submission	BV-BRC	36	36
Subspecies Classification	BV-BRC	10	5
Taxonomic Classification	BV-BRC	428	415
TnSeq Analysis	BV-BRC	1071	855
Variation Analysis	BV-BRC	19	16
SARS-2 Genome Assembly and Annotation	BV-BRC	437	389

Publications and Citations

Publications and citations provide insights into how the BRC is moving science and technology forward and how the resources are serving their respective research communities. Lists of BRC-generated publications (including publications supported by the BRC program in collaboration with various partners) are updated when new manuscripts are accepted and published. Citations to BRC resources are measured using Google Scholar and augmented using PubMed and custom queries as needed to identify citations to the resource that do not cite the official reference publication(s).

- **Citations to BRC publications**
  - *Definition* - Citations to the BRC as measured by citations to key BRC publications, which describe the overall BRC resources, new data and/or analysis tools, or novel use cases supported by them.
  - *Measurement mechanism* - Set up a common Google Scholar profile covering key BRC resource publications (grouped by BRC) and show aggregated citations for each group. The use of Google Scholar profile makes it easier to view the list of publications used to track citations, update the list with new publications, and provide citation counts for individual publications as well as aggregated counts for each resource. Below is the link to the common BRC Google Scholar Profile.
    - https://scholar.google.com/citations?user=kXLGwkYAAAAJ
  - *Measure* - Cumulative number of citations.
- **Citations to BRC resources**
  - *Definition* - Citations to the BRC resource as measured Google Scholar searches using predetermined set of keywords based on name and/or acronym of each of the BRC resources, and additional keywords to filter out any false positive or negative results to the extent possible. This is complementary to the citations to the BRC publications described above and necessary because, often, users cite BRC resources by mentioning the resource name or URL in the manuscript text, instead of citing relevant publications.
  - *Measurement mechanism* - Define set of keywords based on name and/or acronym of each of the BRC resources and additional keywords to filter out any false positive or negative results to the extent possible. Using these keywords as search terms, create Google Scholar URLs for each of the BRC resources, which will be checked every month to report a cumulative number of citations for each resource. Because of the limitations of the logical and advanced query operations supported by Google Scholar search interface, we are dividing BV-BRC query into three distinct sub queries as shown below.
    - VEuPathDB (merged DB, including legacy VectorBase, FungiDB & parasite resources): https://scholar.google.com/scholar?q=OrthoMCL+OR+PlasmoDB+OR+ToxoDB+OR+CryptoDB+OR+TrichDB+OR+GiardiaDB+OR+TriTrypDB+OR+AmoebaDB+OR+MicrosporidiaDB+OR+%22FungiDB%22+OR+PiroplasmaDB+OR+%22vectorbase%22+OR+veupathdb+OR+ApiDB+OR+EuPathDB+-encrypt+-cryptography+-hymenoptera
    - BV-BRC:
      - PATRIC BRC: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C39&q=%28PATRIC+AND+Wattam%29+OR+%E2%80%9Cpatricbrc%22+OR+%22pathosystems+resource+integration+center%22
      - RAST/RASTtk: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C39&q=%28RAST+AND+overbeek%29+OR+%22rast.nmpdr.org%22
      - IRD/ViPR: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C39&q=%22viprbrc%22+OR+

%22virus+pathogen+resource%22+OR+%E2%80%9Cfludb%22+OR+%22influenza+research+database%22

- *Measure* - Cumulative number of citations, cumulative.

Table 5. Citations to BRC Publications and Resources

	Number of Citations (YTD)	Number of Citations (Cumulative)
Citations to BV-BRC publications	2,123	19,762
Citations to BV-BRC resources	2,280	21,590

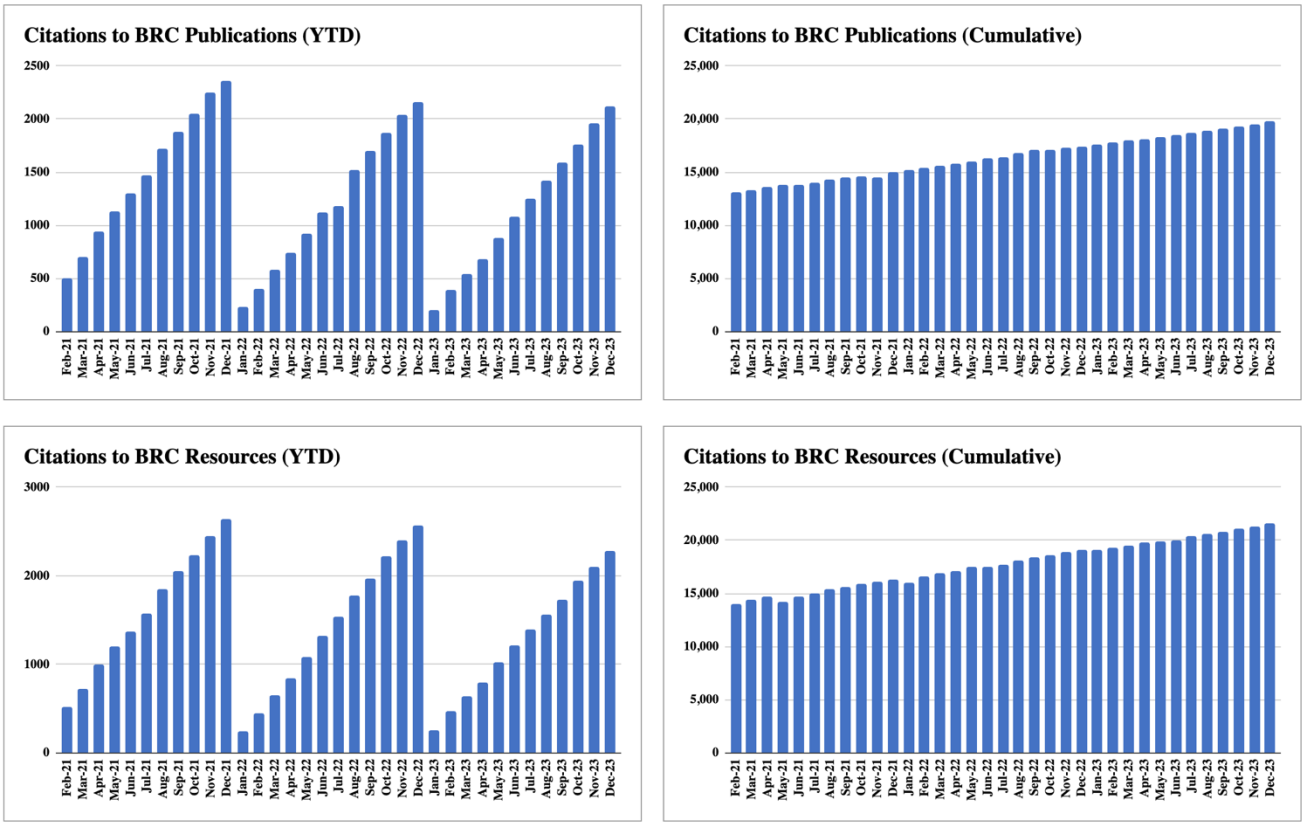

Figure 2. Citations to BV-BRC resources and publications.

Observations:

- Citations to BV-BRC resources remains steady, adding approximately 200 citations per month over the past 2 years.
- Cumulative citations to BV-BRC publications is nearing 20,000, and citations to BV-BRC resources has exceeded 21,000.

User Activities

Outreach activities provide additional channels to engage users. User requests for help typically come in through the help desk functionality available from both BRC websites and are tracked using ticketing software tools. Webinar and workshop participants are counted at the time of registration and participation at the event. Counts of access to recorded webinars may be used to augment the

total. Followers on social media (Twitter, Facebook, YouTube) are counted using the built-in mechanisms those platforms provide.

- **Total registered users**
  - *Definition* - Total cumulative number of users who have registered with the BRC via the website registration mechanism, from inception to the specified date.
  - *Measurement mechanism* - The registration process creates an entry in the registered user database for each BRC. Total number of registered users is queried from the database at the specified date.
  - *Measure* - Total number of registered users (cumulative).
- **Total storage used for user data**
  - *Definition* - Total amount of disk storage in use to host user data at the specified date. This metric provides an additional indication of resource usage that may not be reflected by website traffic or analysis jobs.
  - *Measurement mechanism* - Inspection of disk usage via query or automated script.
  - *Measure* - Total terabytes (TB) currently in use.
- **User requests for help**
  - *Definition* - Total number of user-initiated contacts to the BRC to request help or information during the specified date range. In addition to summarizing total user requests, we will also summarize them by the following categories: Requests for help, Bug reports, and New features / enhancements.
  - *Measurement mechanism* - Manual tally of the auto-generated helpdesk tickets triggered by user requests. Tallies may be augmented with manual counts of interactions where the user bypassed the helpdesk system, *e.g.* via direct email or messaging to BRC team members.
  - *Measure* - Requests per month.
- **Webinar/workshop events and participants**
  - *Definition* - Total number of outreach events (i.e. BRC webinars, workshops, and online courses) held per month and total number of participants who attended those events.
  - *Measurement mechanism* - Manual tally of participants in attendance at the time of the webinar or workshop, summed over all of the events held per month.
  - *Measure* - Cumulative number of participants per month
- **Followers on social media**
  - *Definition* - Total number of followers, by social media outlet, at the specified date. Current active BRC social media outlets are Twitter, Facebook, and YouTube.
  - *Measurement mechanism* - Inspection of the number of followers reported by the media outlet at the specified date.
  - *Measure* - Total number of followers, by media outlet.

Table 5. BV-BRC User Activities

	BV-BRC
Total registered users	50,062
Total storage used for user data (TB)	464.9
User requests: • Request for help • Report bug	66 66 (100%) 0 (0%)

• Suggest improvement	0 (0%)
Webinar/workshop events	1
Total webinar/workshop participants	17
Total MOOC registrants (cumulative)	13,335
Twitter (X) followers: PATRIC[1] IRD/ViPR[1] BV-BRC Total	 640 444 391 1,475
Facebook followers: PATRIC[1] IRD/ViPR[1] BV-BRC Total	 258 2,349 1,100 3,707
YouTube subscribers: PATRIC[1] IRD/ViPR[1] BV-BRC Total	 465 194 521 1,180
YouTube views: PATRIC[1] IRD/ViPR[1] BV-BRC Total	 212 27 1,404 1,643
BRC Subreddit members	104
BRC Subreddit views	33
LinkedIn Followers	93

Notes:

1. *The PATRIC, IRD, and ViPR resources have now been decommissioned and are no longer available. They will be removed from the next month's report. Statistics from the PATRIC, IRD, and ViPR social media channels (Twitter, Facebook, and YouTube) will be maintained until a sufficient number of users have migrated to the corresponding BV-BRC channels.*

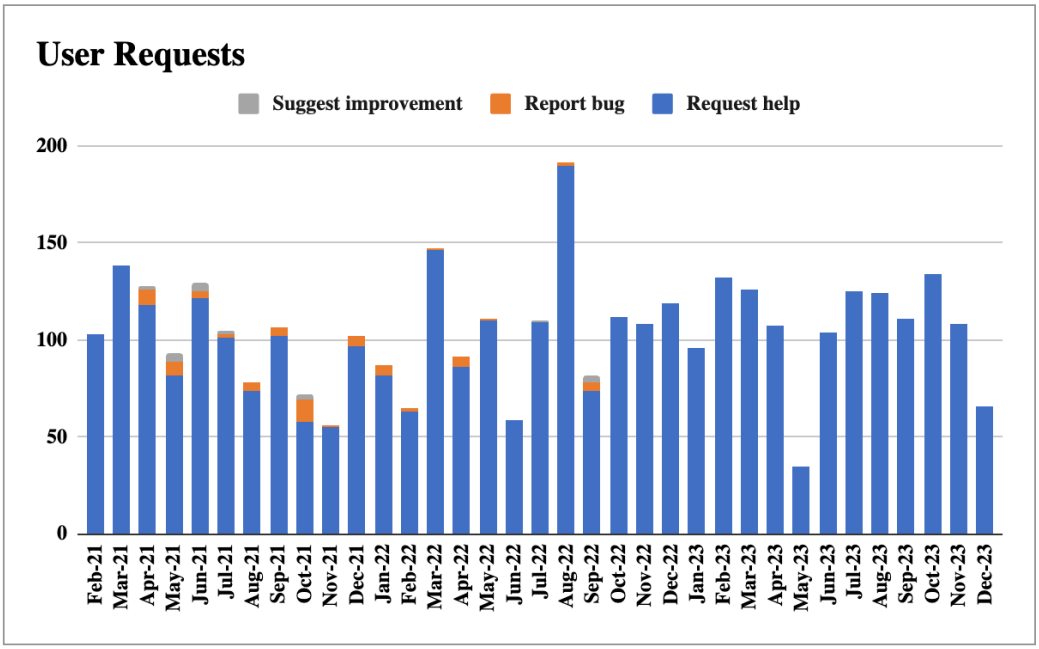

Figure 3. Requests by users, sorted by type.

Observations:

- The number of user requests in December (n=66) was significantly lower than the typical number of approximately 100. This is likely due to less site usage during the holidays.