BV-BRC

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

Monthly Usage Metrics Report

Performance Period: January 1, 2022 – January 31, 2022

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: February 10, 2022

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

Note: As per the recent request from NIAID, we are working with the other BRC to provide jointly agreed plots showing accumulative usage data over time. We will start including them in the monthly reports, starting with the next monthly report.

This monthly usage metrics report provides a summary of the BV-BRC usage for the current reporting period in accordance with the Joint-BRC Common Usage Metrics Plan developed by the BRCs and subsequently approved by NIAID.

As per the plan, each BRC will aggregate and report usage metrics for their constituent parts, *i.e.*, PATRIC and IRD/ViPR for BV-BRC. These metrics will serve as a basis for collecting quantitative measures of usage of the BRC resources to identify trends, areas that are performing well, and areas for improvement. Usage metrics will be reported to NIAID individually by each BRC monthly, and in combination on the BRC Gateway website once this is publicly available. In addition, annual summaries will be included in the Annual Progress Reports.

It is important to note that usage metrics across the two BRCs are highly dependent on the relative sizes of the respective research communities, the associated quantities, and types of available public data, and how each of the resources delivers the data and tools to the user. Thus, cross-BRC comparisons of individual metrics are not necessarily indicative of relative usage or performance.

Common usage metrics covering both BRCs (note that this list is subject to modification, based on feasibility of collection, changes in availability technologies, BRC website development, suggestions from NIAID program and other stakeholders, *etc.*):

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)
- o Measurement mechanism AWStats.
- o *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.

o *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.
- o 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.
- o Measurement mechanism AWStats.
- o *Measure* Average number of pages per visit per month.

Average visits per visitor

- o Definition The average number of visits per visitor.
- o Measurement mechanism AWStats.
- o Measure Average number of visits per visitor per month.

Average visit duration

- o Definition_- The average time a visitor spent on the site for each visit, measured in seconds.
- Measurement mechanism AWStats.
- o 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.
- o 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.
- o *Measure* Total unique registered users per month.

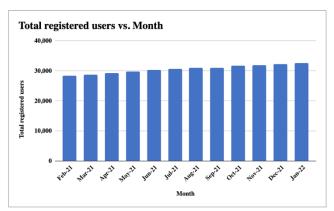
Table 1. BV-BRC Website Usage Metrics¹

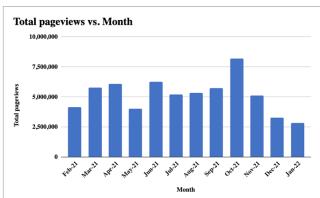
Metric	PATRIC	IRD	ViPR	BV-BRC Prod.	All Combined
Total visits	317,150	6,553	13,319	5,075	338,705
Total unique visitors	14,210	4,267	8,837	1,566	25,640
Total pageviews	2,100,933	499,636	226,420	15,469	2,843,818
Avg. pages / visit	6.62	76.24	16.99	3.04	8.39
Avg. visits / visitor	22.31	1.53	1.5	3.24	13.21
Avg. visit duration (seconds)	1,103	593	369	499	1,070

Bandwidth (GB)	216.50	7.24	89.97	2.25	315.99
Registered users that run a service ^{2,3}	732	57	57	732	789

Notes:

- A link to the BV-BRC summary AWStats page is available from the BV-BRC About page (https://www.bv-brc.org/about)
- 2. Note: This measure This will only be 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.
- 3. PATRIC and BV-BRC Production are the same because both resources use the same computational services infrastructure. Similarly, IRD and ViPR use the same computational infrastructure, so those numbers are the same as well.





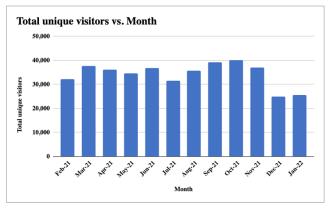


Figure 1. Selected BV-BRC website usage metrics.

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	Domain	Species	Genomes	Page Views
Acinetobacter	Bacteria	690	11,790	1,078
Bacillus	Bacteria	873	7,181	8,000
Bartonella	Bacteria	76	214	1,077
Borreliella	Bacteria	17	868	13
Brucella	Bacteria	86	1,180	2,100
Burkholderia	Bacteria	315	5,039	392
Campylobacter	Bacteria	268	7,027	871
Chlamydia	Bacteria	21	597	435
Clostridium	Bacteria	438	3,670	1,292
Coxiella	Bacteria	12	120	265
Ehrlichia	Bacteria	7	40	501
Escherichia	Bacteria	192	39,333	4,797
Francisella	Bacteria	29	1,078	103
Helicobacter	Bacteria	88	2,880	704
Listeria	Bacteria	43	5,760	561
Mycobacterium	Bacteria	313	31,105	1,417
Pseudomonas	Bacteria	1,846	15,394	3,480
Rickettsia	Bacteria	51	197	834
Salmonella	Bacteria	309	28,655	1,919
Shigella	Bacteria	111	4341	919
Staphylococcus	Bacteria	578	23,969	1,798
Streptococcus	Bacteria	426	37,204	2,922
Vibrio	Bacteria	385	6,239	1,320
Yersinia	Bacteria	28	1,508	355
Bunyavirales	Virus	611	16,648	1,930
Caliciviridae	Virus	241	63,994	254
Coronaviridae	Virus	1,194	3,507,661	5,094

Filoviridae	Virus	26	4,305	344
Flaviviridae	Virus	480	370,123	5,285
Hepeviridae	Virus	50	20,909	189
Herpesviridae	Virus	861	64,716	3,907
Influenza	Virus	4	5,193	28,946
Paramyxoviridae	Virus	742	86,243	856
Picornaviridae	Virus	1,149	150,722	1,077
Pneumoviridae	Virus	17	45,828	573
Poxviridae	Virus	300	11,482	671
Reoviridae	Virus	406	136,472	3,304
Rhabdoviridae	Virus	706	37,950	714
SARS-CoV-2	Virus	1	3,725,120	1,505
Togaviridae	Virus	68	14,716	770
SARS-CoV-2 (BV-BRC)	Virus	1	3,481,117	925

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 pageviews by primary data types, which will be measured by querying the website usage statistics in Google Analytics by data type. VIPR/IRD pages views are combined (added together) based on data type.

Table 3. BRC Website Usage by Data Type (BV-BRC)

Data Type	BRC Domain	Page Views
Taxonomy	PATRIC	28,301
Genome	PATRIC	74,241
Genome sequence	PATRIC	2,784
Feature (Genes/Proteins)	PATRIC	43,113
Specialty gene	PATRIC	6,983
Protein families	PATRIC	3,658
Pathway	PATRIC	8,880
Subsystems	PATRIC	2,986
Transcriptomics	PATRIC	1,530

Interactions	PATRIC	768
Phylogeny	PATRIC	1,650
Antibiotic	PATRIC	25
Workspace (User Data)	PATRIC	62,732
Genome	IRD/ViPR	21,599
Gene/Protein	IRD/ViPR	6,837
Strain	IRD/ViPR	6,914
Immune epitopes	IRD/ViPR	399
Ortholog groups	IRD/ViPR	59
Antiviral drugs	IRD/ViPR	342
Host factors	IRD/ViPR	198
Protein structures	IRD/ViPR	267
Protein domains and motifs	IRD/ViPR	35
Plasmids	IRD/ViPR	48
SFVT	IRD/ViPR	67
Surveillance	IRD/ViPR	178
Serology	IRD/ViPR	22
Phenotypes	IRD/ViPR	34
PCR Primers	IRD/ViPR	232
SARS-CoV-2 Variant Tracker	BV-BRC	619

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

Table 4. BRC Tools/Services Usage Metrics

Tool/Service	BRC Domain	Submitted	Completed
Codon Tree	PATRIC	459	443
Comprehensive Genome Analysis	PATRIC	1751	1647
Differential Expression	PATRIC	4	2
FastqUtils	PATRIC	156	147
Genome Alignment	PATRIC	144	131
Genome Annotation	PATRIC	2747	2644
Genome Assembly	PATRIC	3358	3155
Genome Comparison	PATRIC	175	146
Metagenome Binning	PATRIC	461	453
Metagenomic Read Mapping	PATRIC	129	128
RNASeq Analysis	PATRIC	33	24
Taxonomic Classification	PATRIC	350	347
TnSeq Analysis	PATRIC	59	33
Variation Analysis	PATRIC	239	227
Alignment Viewer	IRD/ViPR	44	44
Antiviral-Resistance-Risk	IRD/ViPR	28	28
BLAST	IRD/ViPR	365	364
Enrichment	IRD/ViPR	1	1
Genotype-Recombination	IRD/ViPR	8	8
H1-Clade Classifier	IRD only	90	90
H1N1-classifier	IRD only	15	15
H5N1-classifier	IRD only	104	92
Ha Numbering	IRD only	64	64
MGC	IRD/ViPR	36	33

MSA	IRD/ViPR	472	424
Mutation-analysis	IRD/ViPR	45	45
Primer3	IRD/ViPR	39	37
Read-seq	IRD/ViPR	46	45
Rva Genotyper	IRD/ViPR	1,058	1,052
Short-seqsearch	IRD/ViPR	10	8
SNP-analysis	IRD/ViPR	515	309
Surveillance-data-mapping	IRD/ViPR	12	10
Tbl-formatter	IRD/ViPR	1	0
Tree	IRD/ViPR	237	211
VIGOR Annotator	IRD/ViPR	60	59
SARS-2 Genome Assembly and Annotation	BV-BRC	29	22

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+Cry ptoDB+OR+TrichDB+OR+GiardiaDB+OR+TriTrypDB+OR+AmoebaDB+OR+Microsporidi aDB+OR+%22FungiDB%22+OR+PiroplasmaDB+OR+%22vectorbase%22+OR+veupath db+OR+ApiDB+OR+EuPathDB+-encrypt+-cryptography+-hymenoptera
- BV-BRC:
 - PATRIC BRC:

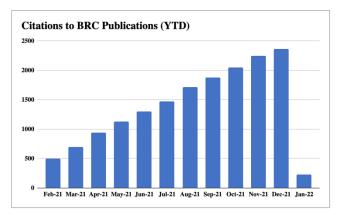
https://scholar.google.com/scholar?hl=en&as_sdt=0%2C39&q=%28PATRIC+AND+Wattam%29+OR+%E2%80%9Cpatricbrc%22+OR+%22pathosystems+resource+integration+center%22

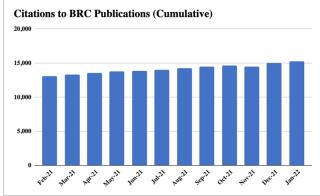
- o RAST/RASTtk:
 - https://scholar.google.com/scholar?hl=en&as_sdt=0%2C39&q=%28RAST+AND+overbeek%29+OR+%22rast.nmpdr.org%22
- IRD/ViPR: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C39&q=%22viprbrc%22+OR+ %22virus+pathogen+resource%22+OR+%E2%80%9Cfludb%22+OR+%22influenza+
- o Measure Cumulative number of citations, cumulative.

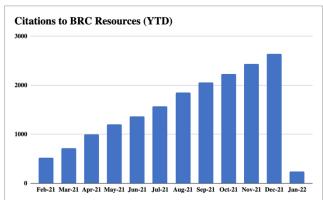
research+database%22

Table 5. Citations to BRC Publications and Resources

	Number of Citations (YTD)	Number of Citations (Cumulative)
Citations to BV-BRC publications	227	15,238
Citations to BV-BRC resources	246	15,970







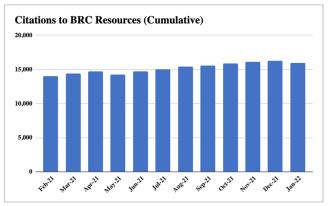


Figure 2. Citations to BV-BRC resources and publications. The January 2022 (Jan-22) YTD citations to BRC publications and resources reflect the start of the new year (2022), thus the drop in monthly citations.

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 guery or automated script.
- o Measure Total terabytes (TB) currently in use.

User requests for help

o 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.
- o *Measure* Requests per month.

• Webinar/workshop events and participants

- o *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.
- o Measure Cumulative number of participants per month

• Followers on social media

- o *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.
- o Measure Total number of followers, by media outlet.

	PATRIC	IRD/ViPR	Total
Total registered users	32,638 ¹	11,842	32,638
Total storage used for user data (TB)	170.8	0.58	171.38
User requests:	61	26	87
Request for helpReport bugSuggest improvement	100% 0% 0%	80% 19% 0%	94% 6% 0%
Webinar/workshop events	0	0	0
Total webinar/workshop participants	0	0	0
Total MOOC registrants (cumulative)	4,975	NA	4,975
Twitter followers	527	373	900
Facebook followers	244	1,735	1,979
YouTube subscribers	320	184	504
YouTube views	988	94	1082
BRC Subreddit members	NA	NA	78
BRC Subreddit viewers	NA	NA	631

^{1.} The number of total PATRIC registered users had an apparent large increase due to the merger of IRD/ViPR and PATRIC user databases. The Total (BV-BRC) is an accurate count of both resources combined

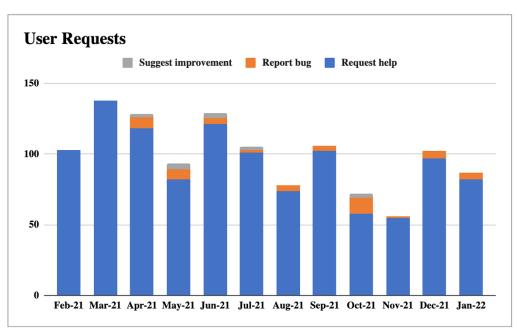


Figure 3. Requests by users, sorted by type.