VEuPathDB BRC contract HHSN75N93019C00077

Usage Metrics Report

Reporting Period: August 1-31, 2021
Submission Date: September 9, 2021

Notes & Change Log

Date	Version/release	Description & Notes
9/9/2021	1	VEuPathDB Performance Metrics for August 2021 ● In coordination with our sister BRC - added graphs of users and site usage, citations, and support requests over time ● In coordination with our sister BRC - Added metrics for the cross-BRC Reddit to Table 6

Joint-BRC Common Usage Metrics Plan

This report will be made available from all VEuPathDB sites, *e.g.*, https://veupathdb.org/, from the About menu.

This monthly usage metrics report provides a summary of the VEuPathDB 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 metrics for their constituent parts, *i.e.* FungiDB, PlasmoDB, OrthoMCL-DB, VectorBase, *etc.* for VEuPathDB. 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 on a monthly basis, and in combination on the BRC Gateway website once this is publicly available. Annual summaries will be included in the Annual Progress Reports.

It is important to note that 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. For VEuPathDB, live website usage statistics pages generated by AWStats from individual websites can be accessed at https://veupathdb.org/awstats/awstats.pl, https://plasmodb.org/awstats/awstats.pl, etc. by replacing individual site names in the URL. These links provide more detailed usage statistics by day of the week/month, country, browser / operating system, and more.

- **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 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.

Table 1 VEuPathDB Website Usage Metrics (August 1-31, 2021)

Metric	Result
Total registered users	23,683
Total visits	105,229
Total unique visitors	34,149
Total pageviews	10,299,866
Avg. pages / visit	97.88
Avg. visits / visitor	3.08
Avg. visit duration (seconds)	501

Bandwidth (GB)	366.28

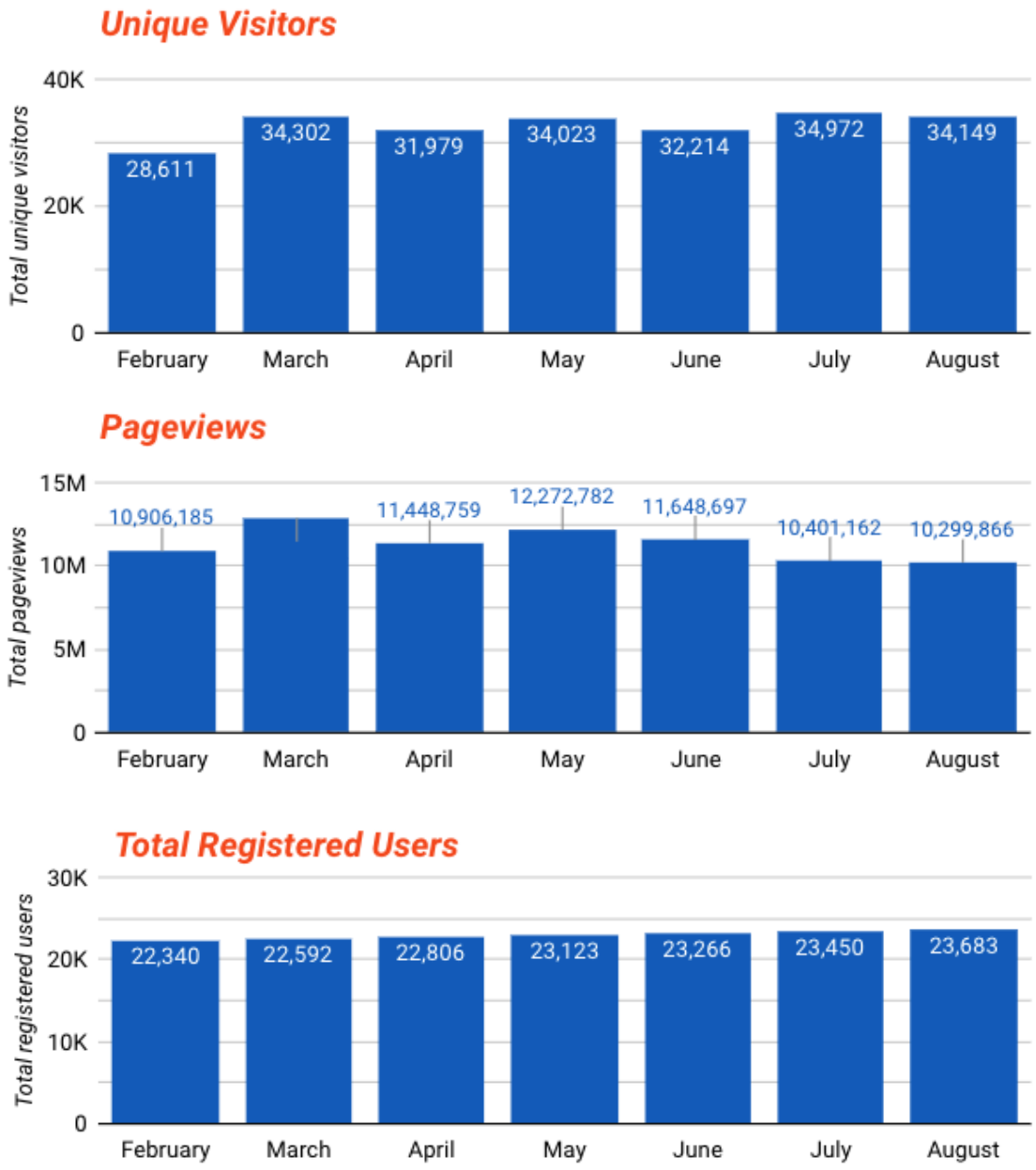


Figure 1 Unique visitors, Page views, and Total registered users over time

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 website page views by taxa, which will be measured by querying the website usage statistics in Google Analytics by taxa name.

Table 2 VEuPathDB Website Usage by Taxa (August 1-31, 2021)

Taxa	Domain	Page Views	# of Species	# of Genome Seqs
Plasmodium	Protozoa	142910	22	51
Toxoplasma	Protozoa	136971	1	15
Trypanosoma	Protozoa	39162	8	25
Leishmania	Protozoa	16403	15	24
Cryptococcus	Fungi	16339	5	10
Aedes	Vectors	13857	2	3
Aspergillus	Fungi	11271	23	28
Saccharomyces	Fungi	10972	1	1
Anopheles	Vectors	10677	19	24
Fusarium	Fungi	8369	7	13
Cryptosporidium	Protozoa	7185	7	11
Pyricularia	Fungi	6865	1	2
Neurospora	Fungi	6403	3	3
Giardia	Protozoa	5214	4	6
Entamoeba	Protozoa	3580	5	9
Ixodes	Vectors	1919	1	2
Candida	Fungi	1765	8	15
Babesia	Protozoa	1734	6	6
Trichomonas	Protozoa	1620	1	1
Drosophila	Vectors	1615	1	1
Culex	Vectors	1515	1	1
Ustilago	Fungi	1437	1	1
Phytophthora	Fungi	1387	7	7
Neospora	Protozoa	1370	1	2
Theileria	Protozoa	1344	4	4
Trichoderma	Fungi	1236	2	3
Naegleria	Protozoa	1223	2	3
Rhodnius	Vectors	1050	1	1
Crithidia	Protozoa	1011	1	1
Acanthamoeba	Protozoa	982	1	1

Eimeria	Protozoa	964	8	8
Glossina	Vectors	829	6	7
Coccidioides	Fungi	668	2	6
Bodo	Protozoa	638	1	1
Schizosaccharomyces	Fungi	527	3	3
Besnoitia	Protozoa	471	1	1
Histoplasma	Fungi	451	1	5
Encephalitozoon	Protozoa	401	4	9
Clavispora	Fungi	394	1	1
Leptomonas	Protozoa	383	2	2
Mucor	Fungi	352	2	2
Musca	Vectors	330	1	1
Lutzomyia	Vectors	304	1	1
Zymoseptoria	Fungi	303	1	2
Sarcocystis	Protozoa	296	1	2
Lomentospora	Fungi	271	1	1
Hepatocystis	Protozoa	267	1	1
Paratrypanosoma	Protozoa	266	1	1
Botrytis	Fungi	261	1	1
Angomonas	Protozoa	256	1	1
Malassezia	Fungi	252	3	4
Phlebotomus	Vectors	242	1	1
Endotrypanum	Protozoa	232	1	1
Blechomonas	Protozoa	232	1	1
Cimex	Vectors	231	1	1
Stomoxys	Vectors	230	1	1
Homo	Host	228	1	1
Chromera	Protozoa	203	1	1
Coprinopsis	Fungi	203	1	1
Hammondia	Protozoa	200	1	1
Puccinia	Fungi	195	4	5
Culicoides	Vectors	193	1	1

Allomyces	Fungi	191	1	1
Cyclospora	Protozoa	186	1	2
Nosema	Protozoa	178	2	3
Rhizopus	Fungi	166	1	1
Biomphalaria	Vectors	158	1	1
Cystoisospora	Protozoa	149	1	1
Mus	Host	145	1	1
Sclerotinia	Fungi	143	1	1
Sordaria	Fungi	138	1	1
Kwoniella	Fungi	133	4	4
Anncaliia	Protozoa	130	1	2
Pediculus	Vectors	126	1	1
Batrachochytrium	Fungi	125	1	1
Spizellomyces	Fungi	120	1	1
Phanerochaete	Fungi	117	1	1
Scedosporium	Fungi	103	1	1
Sporisorium	Fungi	94	1	1
Blastomyces	Fungi	92	3	4
Talaromyces	Fungi	92	2	2
Globisporangium	Fungi	91	3	4
Leptotrombidium	Vectors	86	1	1
Gregarina	Protozoa	76	1	1
Yarrowia	Fungi	68	1	2
Podospora	Fungi	66	1	1
Paracoccidioides	Fungi	65	2	3
Phycomyces	Fungi	62	1	1
Peronospora	Fungi	58	1	1
Enterocytozoon	Protozoa	56	2	2
Tremella	Fungi	55	1	1
Albugo	Fungi	54	2	2
Colletotrichum	Fungi	52	1	1
Monocercomonoides	Protozoa	51	1	1

Hyaloperonospora	Fungi	49	1	1
Phytopythium	Fungi	49	1	1
Spironucleus	Protozoa	48	1	1
Bremia	Fungi	48	1	1
Sarcoptes	Vectors	46	1	1
Rhizophagus	Fungi	45	1	2
Aphanomyces	Fungi	45	2	2
Cytauxzoon	Protozoa	44	1	1
Pleurotus	Fungi	41	1	1
Edhazardia	Protozoa	40	1	1
Cenococcum	Fungi	39	1	1
Penicilliopsis	Fungi	38	1	1
Vitrella	Protozoa	37	1	1
Melampsora	Fungi	37	1	1
Ophiostoma	Fungi	35	1	1
Cladophialophora	Fungi	34	2	2
Pythium	Fungi	34	2	2
Penicillium	Fungi	34	1	1
Macaca	Host	32	2	2
Trichosporon	Fungi	29	1	1
Pichia	Fungi	29	1	1
Saprolegnia	Fungi	29	2	2
Kluyveromyces	Fungi	28	1	1
Exophiala	Fungi	27	3	3
Bos	Host	24	1	1
Hanseniaspora	Fungi	24	2	2
Verruconis	Fungi	22	1	1
Sporothrix	Fungi	22	2	2
Thermothelomyces	Fungi	18	1	1
Nematocida	Protozoa	14	3	5
Fonsecaea	Fungi	14	1	1
Claviceps	Fungi	12	1	1

Ascosphaera	Fungi	12	1	1
Cyphellophora	Fungi	11	1	1
Blumeria	Fungi	10	1	1
Vittaforma	Protozoa	10	1	1
Rhinocladiella	Fungi	10	1	1
Pseudogymnoascus	Fungi	9	1	1
Hepatospora	Protozoa	9	1	2
Uncinocarpus	Fungi	7	1	1
Enterospora	Protozoa	6	1	1
Mitosporidium	Protozoa	5	1	1
Pseudoloma	Protozoa	4	1	1
Pneumocystis	Fungi	3	1	1
Spraguea	Protozoa	2	1	1
Vavraia	Protozoa	1	1	1

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 used. 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.

Table 3 VEuPathDB Website Usage by Data Type August 1-31, 2021)

Data Type	Domain	Page Views	Searches
Taxonomy	VEuPathDB	450427	541
Genomes	VEuPathDB	450427	1381
Genome sequences	VEuPathDB	450427	3277
Genes/Proteins	VEuPathDB	450427	56646
Transcriptomics	VEuPathDB	388372	34959
Proteomics	VEuPathDB	353752	1799
Variation data	VEuPathDB	307470	5467
Epigenomics	VEuPathDB	267201	6
Enzyme commission	VEuPathDB	153704	55

Gene Ontology	VEuPathDB	283434	144
Protein domains	VEuPathDB	450427	266
Immunology	VEuPathDB	389666	36
Gene Orthology	VEuPathDB	445947	782
Synteny	VEuPathDB	450427	NA
Metabolic pathways	VEuPathDB	1321	173
Phenotype	VEuPathDB	134521	3371
Isolate data	VEuPathDB	909	5095
Subcellular localization	VEuPathDB	321864	646
ESTs	VEuPathDB	425210	64
Compounds	VEuPathDB	216	1450

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.

- **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 are captured via website and server logs, which are used to tally the number.
  - *Measure* - Jobs per month, tallied by service/tool.

Table 4. VEuPathDB Tools/Services Usage Metrics (August 1-31, 2021)

Tool/Service	BRC Domain	Submitted
Sequence retrieval tool	VEuPathDB	10504
BLAST	VEuPathDB	10741

Enrichment Analyses	VEuPathDB	1467
Web services	VEuPathDB	8296
Boolean operations	VEuPathDB	2906
Apollo (Access)	VEuPathDB	459
Site Search	VEuPathDB	135597
Galaxy Jobs	VEuPathDB	768
Genome Browser	VEuPathDB	409497
User Comments	VEuPathDB	50
Multiple sequence alignment (isolates)	VEuPathDB	5301
Results downloads	VEuPathDB	4399
Data analysis searches (breakdown below)		
Annotation searches	VEuPathDB	4188
Epigenomics	VEuPathDB	6
Function prediction	VEuPathDB	199
Gene models	VEuPathDB	114
Genetic variation	VEuPathDB	107
Genomic Location	VEuPathDB	145
Immunology	VEuPathDB	36
Orthology and synteny	VEuPathDB	782
Pathways and interactions	VEuPathDB	97
Phenotype	VEuPathDB	3371
Protein features and properties	VEuPathDB	283
Protein targeting and localization	VEuPathDB	646
Proteomics	VEuPathDB	1799
Sequence analysis	VEuPathDB	7956
Structure analysis	VEuPathDB	46
Taxonomy	VEuPathDB	541
Text	VEuPathDB	1371

Transcriptomics	VEuPathDB	34959
Popset Isolate Sequences	VEuPathDB	5095
Genomic Sequences	VEuPathDB	3186
Genomic Segments	VEuPathDB	91
SNPs	VEuPathDB	5360
ESTs	VEuPathDB	64
Metabolic Pathways	VEuPathDB	173
Compounds	VEuPathDB	1450

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).

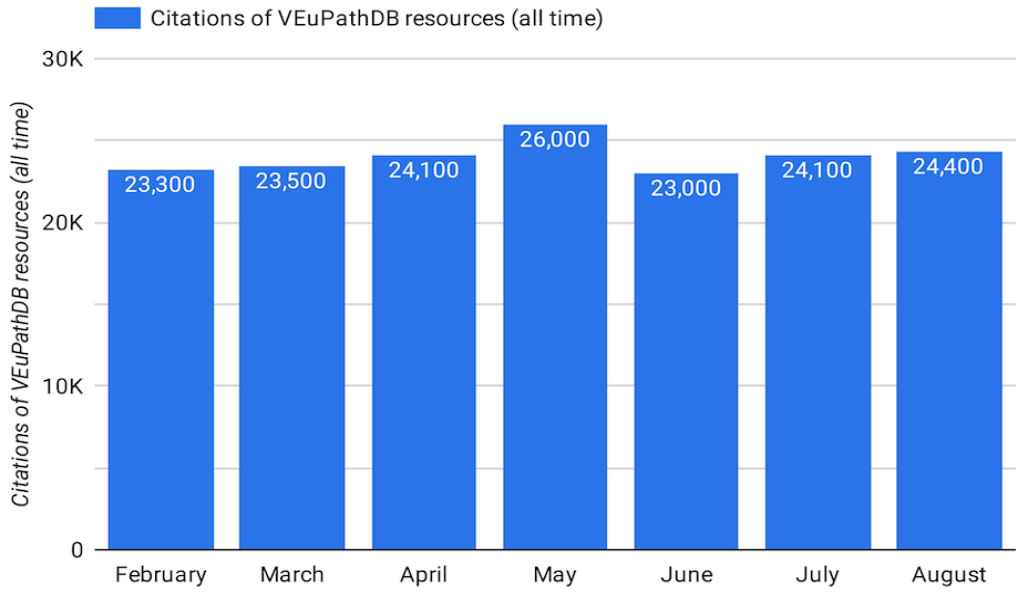
- **Citations to BRC publications**
  - *Definition* - Citations to the BRC as measured by citations to key BRC publications, which describe the overall BRC resources, new data and/or analysis tools, or novel use cases supported by them.
  - *Measurement mechanism* - Set up a common Google Scholar profile 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, year to date.
- **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+MicrosporidiaDB+OR+%22FungiDB%22+OR+PiroplasmaDB+OR+%22vectorbase%22+OR+veupathdb+OR+ApiDB+OR+EuPathDB+-encrypt+-cryptography+-hymenoptera

- *Measure* - Cumulative number of citations, year to date.

Table 5: Citations

Metric	Year to date	Cumulative
Citations of BRC Publications	842	10746
Citations of BRC Resources	1710	24400



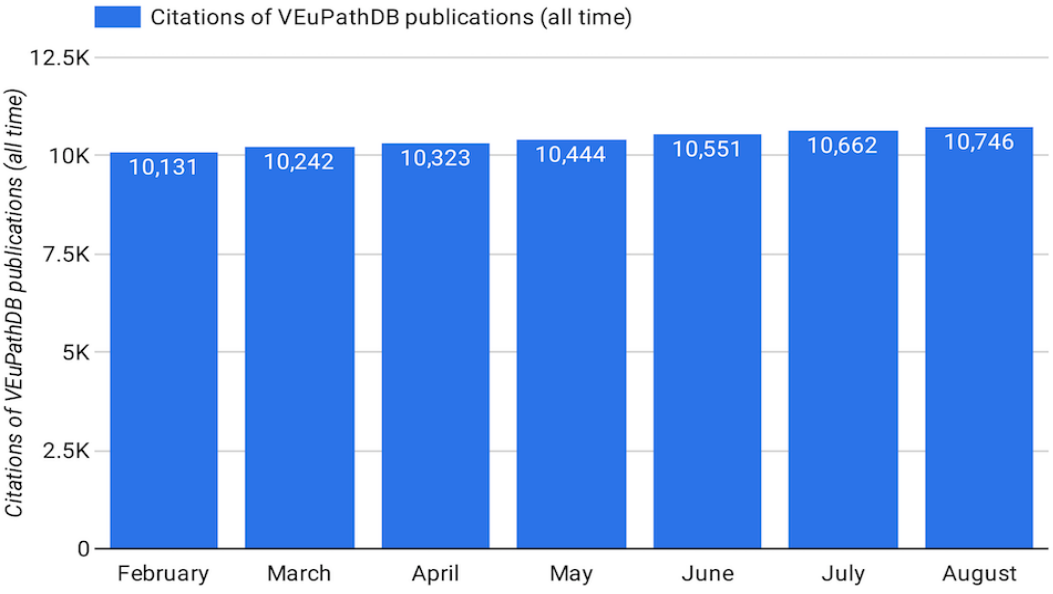


Figure 2 Cumulative citations of VEuPathDB resources and publications

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 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. Note that because some emails fit into multiple categories the total percent can exceed 100.
- **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, YouTube, and Reddit.
  - *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 6: VEuPathDB User Activities (**August 1-31)**

Metric	**Results (reporting period)**
VEuPathDB integrated user data	~53G
Galaxy user data	~9.6T
User requests for help (some fit multiple categories and total may be >100%)	50 (22% bugs, 46% help, 8% new data, 7% new feature, 7% other)
Webinar/workshop events and participants	Release 53 webinar, 27 attendees
Followers on social media: (reported as total)	

FaceBook @VEuPathDB	1843
FaceBook @FungiDB	566
FaceBook @VectorBase	2183
Twitter @VEuPathDB	2838
Twitter @FungiDB	3222
Twitter @VectorBase	1957
YouTube	544
BRC Reddit subscribers	39
BRC Reddit views	729

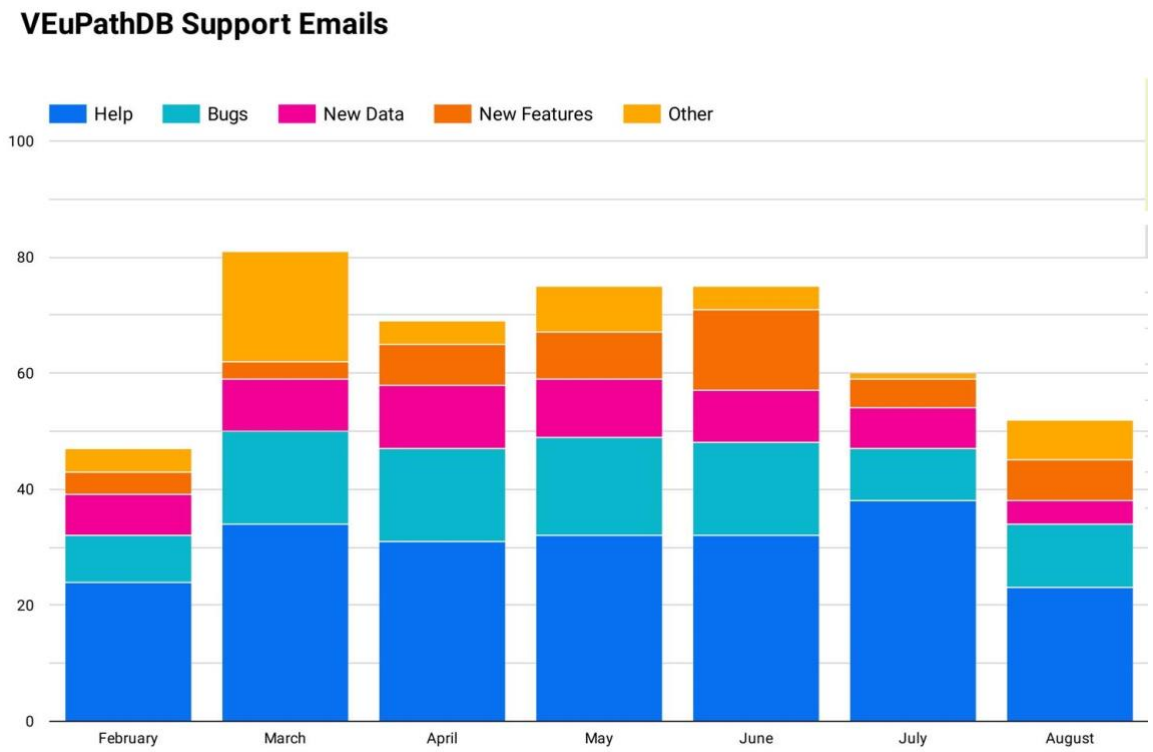


Figure 3 VEuPathDB Support emails over time