



VEuPathDB BRC contract HHSN75N93019C00077

Usage Metrics Report

Reporting Period: September 1-30, 2021

Submission Date: October 11, 2021

Notes & Change Log

Date	Version/release	Description & Notes
10/11/2021	1	<p>VEuPathDB Performance Metrics for September 2021</p> <ul style="list-style-type: none">• In the next iteration we will also report the number of registered users that log in each month• Based on COR feedback from the August report, both BRCs will seek guidance at the next monthly meeting on how to make the number of page views more useful

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**

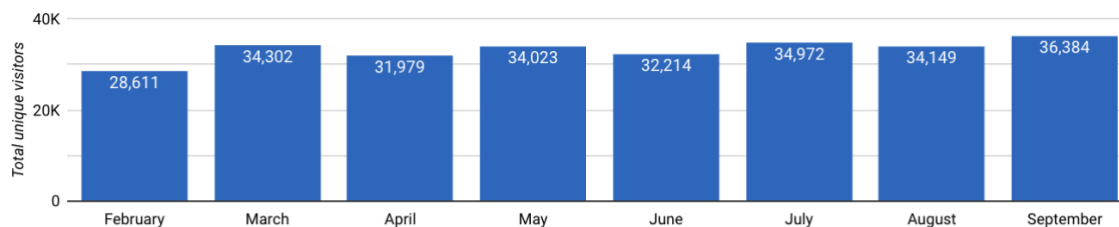
- *Definition* - Total number of bytes for pages, images and files downloaded by web browsing. This number includes traffic for web only (or mail only, or ftp only depending on value of LogType). This number does not include technical header data size used inside the HTTP or HTTPS protocol or by protocols 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 (September 1-30, 2021)

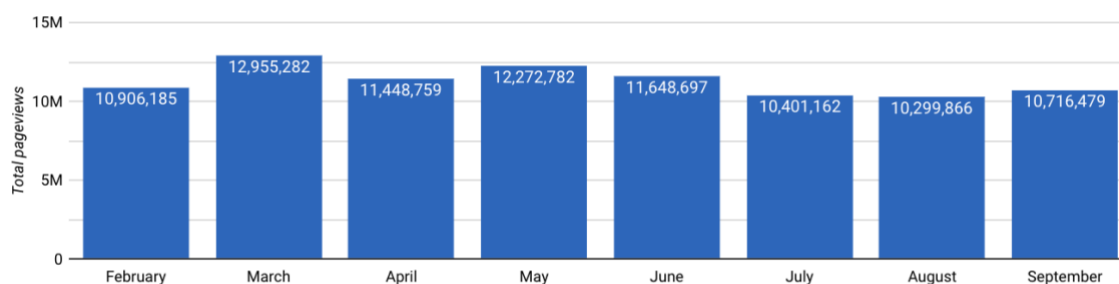
Metric	Result
Total registered users	23,895
Total visits	85,810.00
Total unique visitors	36,384.00
Total pageviews	10,716,479.00
Avg. pages / visit	124.88
Avg. visits / visitor	2.35

Avg. visit duration (seconds)	538.00
Bandwidth (GB)	450.41

Unique Visitors



Pageviews



Total Registered Users

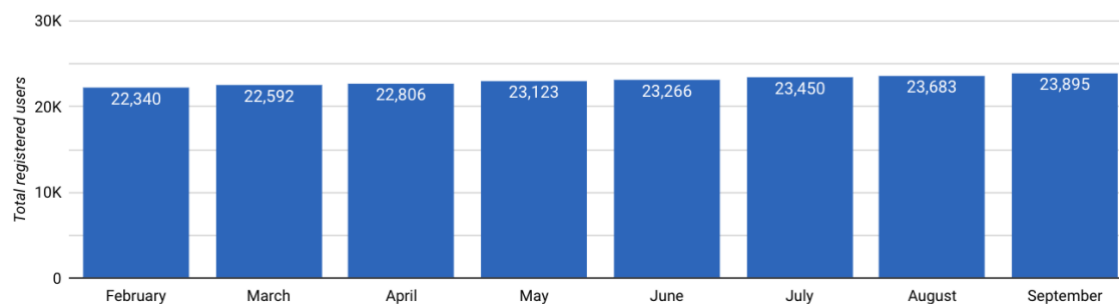


Figure 1 Unique visitors, Page views, and Total registered users over time

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 (September 1-30, 2021)

Taxa	Domain	Page Views	# of Species	# of Genome Seqs
<i>Plasmodium</i>	Protozoa	210197	22	51
<i>Toxoplasma</i>	Protozoa	102428	1	15
<i>Trypanosoma</i>	Protozoa	45136	8	25
<i>Leishmania</i>	Protozoa	22381	15	24
<i>Pyricularia</i>	Fungi	20080	1	2
<i>Anopheles</i>	Vectors	12627	19	24
<i>Aedes</i>	Vectors	11926	2	3
<i>Cryptococcus</i>	Fungi	11848	5	10
<i>Aspergillus</i>	Fungi	10396	23	28
<i>Fusarium</i>	Fungi	8850	7	13
<i>Neurospora</i>	Fungi	8066	3	3
<i>Cryptosporidium</i>	Protozoa	5859	7	11
<i>Saccharomyces</i>	Fungi	5225	1	1
<i>Giardia</i>	Protozoa	4307	4	6
<i>Entamoeba</i>	Protozoa	3792	5	9
<i>Neospora</i>	Protozoa	3624	1	2
<i>Eimeria</i>	Protozoa	2134	8	8
<i>Theileria</i>	Protozoa	1674	4	4
<i>Phytophthora</i>	Fungi	1604	7	7
<i>Trichomonas</i>	Protozoa	1463	1	1
<i>Naegleria</i>	Protozoa	1384	2	3
<i>Hammondia</i>	Protozoa	1379	1	1
<i>Ixodes</i>	Vectors	1299	1	2
<i>Babesia</i>	Protozoa	1205	6	6
<i>Cystoisospora</i>	Protozoa	1175	1	1
<i>Candida</i>	Fungi	1154	8	15

<i>Culex</i>	Vectors	1147	1	1
<i>Fonsecaea</i>	Fungi	955	1	1
<i>Drosophila</i>	Vectors	837	1	1
<i>Besnoitia</i>	Protozoa	826	1	1
<i>Rhodnius</i>	Vectors	825	1	1
<i>Crithidia</i>	Protozoa	756	1	1
<i>Ustilago</i>	Fungi	670	1	1
<i>Blastomyces</i>	Fungi	612	3	4
<i>Chromera</i>	Protozoa	572	1	1
<i>Glossina</i>	Vectors	557	6	7
<i>Homo</i>	Host	458	1	1
<i>Sarcocystis</i>	Protozoa	451	1	2
<i>Encephalitozoon</i>	Protozoa	432	4	9
<i>Acanthamoeba</i>	Protozoa	410	1	1
<i>Schizosaccharomyces</i>	Fungi	366	3	3
<i>Bodo</i>	Protozoa	359	1	1
<i>Mucor</i>	Fungi	349	2	2
<i>Leptomonas</i>	Protozoa	335	2	2
<i>Cimex</i>	Vectors	311	1	1
<i>Hepatocystis</i>	Protozoa	300	1	1
<i>Paratrypanosoma</i>	Protozoa	281	1	1
<i>Leptotrombidium</i>	Vectors	276	1	1
<i>Lutzomyia</i>	Vectors	274	1	1
<i>Nosema</i>	Protozoa	274	2	3
<i>Coccidioides</i>	Fungi	260	2	6
<i>Angomonas</i>	Protozoa	250	1	1
<i>Botrytis</i>	Fungi	243	1	1

<i>Trichoderma</i>	Fungi	218	2	3
<i>Yarrowia</i>	Fungi	218	1	2
<i>Blechnomonas</i>	Protozoa	209	1	1
<i>Cyclospora</i>	Protozoa	205	1	2
<i>Endotrypanum</i>	Protozoa	196	1	1
<i>Musca</i>	Vectors	194	1	1
<i>Phlebotomus</i>	Vectors	191	1	1
<i>Coprinopsis</i>	Fungi	186	1	1
<i>Pediculus</i>	Vectors	184	1	1
<i>Stomoxys</i>	Vectors	182	1	1
<i>Histoplasma</i>	Fungi	171	1	5
<i>Rhizopus</i>	Fungi	154	1	1
<i>Culicoides</i>	Vectors	151	1	1
<i>Phycomyces</i>	Fungi	134	1	1
<i>Malassezia</i>	Fungi	124	3	4
<i>Kwoniella</i>	Fungi	123	4	4
<i>Puccinia</i>	Fungi	119	4	5
<i>Anncalia</i>	Protozoa	116	1	2
<i>Spironucleus</i>	Protozoa	114	1	1
<i>Phanerochaete</i>	Fungi	111	1	1
<i>Sarcoptes</i>	Vectors	103	1	1
<i>Allomyces</i>	Fungi	98	1	1
<i>Mus</i>	Host	84	1	1
<i>Gregarina</i>	Protozoa	83	1	1
<i>Spizellomyces</i>	Fungi	81	1	1
<i>Zymoseptoria</i>	Fungi	78	1	2
<i>Nematocida</i>	Protozoa	76	3	5

<i>Biomphalaria</i>	Vectors	75	1	1
<i>Pleurotus</i>	Fungi	67	1	1
<i>Sporisorium</i>	Fungi	65	1	1
<i>Trichosporon</i>	Fungi	62	1	1
<i>Tremella</i>	Fungi	60	1	1
<i>Sordaria</i>	Fungi	58	1	1
<i>Mastigamoeba</i>	Protozoa	57	1	1
<i>Sclerotinia</i>	Fungi	55	1	1
<i>Penicillium</i>	Fungi	53	1	1
<i>Talaromyces</i>	Fungi	44	2	2
<i>Kluyveromyces</i>	Fungi	42	1	1
<i>Melampsora</i>	Fungi	42	1	1
<i>Pythium</i>	Fungi	41	2	2
<i>Lomentospora</i>	Fungi	41	1	1
<i>Aphanomyces</i>	Fungi	41	2	2
<i>Vitrella</i>	Protozoa	40	1	1
<i>Sporothrix</i>	Fungi	40	2	2
<i>Monocercomonoides</i>	Protozoa	39	1	1
<i>Rhizophagus</i>	Fungi	35	1	2
<i>Enterocytozoon</i>	Protozoa	35	2	2
<i>Hyaloperonospora</i>	Fungi	34	1	1
<i>Podospira</i>	Fungi	32	1	1
<i>Globisporangium</i>	Fungi	29	3	4
<i>Exophiala</i>	Fungi	24	3	3
<i>Vittaforma</i>	Protozoa	23	1	1
<i>Thermothelomyces</i>	Fungi	23	1	1
<i>Paracoccidioides</i>	Fungi	22	2	3

<i>Albugo</i>	Fungi	22	2	2
<i>Cladophialophora</i>	Fungi	20	2	2
<i>Saprolegnia</i>	Fungi	20	2	2
<i>Colletotrichum</i>	Fungi	19	1	1
<i>Pichia</i>	Fungi	16	1	1
<i>Bremia</i>	Fungi	16	1	1
<i>Macaca</i>	Host	16	2	2
<i>Ordospora</i>	Protozoa	16	1	1
<i>Clavispora</i>	Fungi	15	1	1
<i>Bos</i>	Host	15	1	1
<i>Batrachochytrium</i>	Fungi	15	1	1
<i>Peronospora</i>	Fungi	15	1	1
<i>Penicillioptosis</i>	Fungi	14	1	1
<i>Phytopythium</i>	Fungi	14	1	1
<i>Mitosporidium</i>	Protozoa	13	1	1
<i>Edhazardia</i>	Protozoa	13	1	1
<i>Spraguea</i>	Protozoa	13	1	1
<i>Ascospaera</i>	Fungi	12	1	1
<i>Cytauxzoon</i>	Protozoa	12	1	1
<i>Cenococcum</i>	Fungi	11	1	1
<i>Hanseniaspora</i>	Fungi	11	2	2
<i>Pseudoloma</i>	Protozoa	9	1	1
<i>Claviceps</i>	Fungi	9	1	1
<i>Uncinocarpus</i>	Fungi	8	1	1
<i>Enterosporea</i>	Protozoa	7	1	1
<i>Vavraia</i>	Protozoa	7	1	1
<i>Pneumocystis</i>	Fungi	7	1	1

<i>Scedosporium</i>	Fungi	7	1	1
<i>Pseudogymnoascus</i>	Fungi	6	1	1
<i>Cyphellophora</i>	Fungi	6	1	1
<i>Trachipleistophora</i>	Protozoa	5	1	1
<i>Hepatospora</i>	Protozoa	5	1	2
<i>Rhinoclatiella</i>	Fungi	2	1	1
<i>Ophiostoma</i>	Fungi	2	1	1
<i>Blumeria</i>	Fungi	1	1	1
<i>Amphiblastys</i>	Protozoa	1	1	1
<i>Verruconis</i>	Fungi	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 (September 1-30, 2021)

Data Type	Domain	Page Views	Searches
Taxonomy	VEuPathDB	494735	375
Genomes	VEuPathDB	494735	1493
Genome sequences	VEuPathDB	494735	2800
Genes/Proteins	VEuPathDB	494735	59218
Transcriptomics	VEuPathDB	434362	36028
Proteomics	VEuPathDB	389687	1972
Variation data	VEuPathDB	436773	5580
Epigenomics	VEuPathDB	304514	6
Enzyme commission	VEuPathDB	178264	62
Gene Ontology	VEuPathDB	330064	415
Protein domains	VEuPathDB	494735	445

Immunology	VEuPathDB	439062	68
Gene Orthology	VEuPathDB	489084	1301
Synteny	VEuPathDB	494735	NA
Metabolic pathways	VEuPathDB	792	152
Phenotype	VEuPathDB	104295	3461
Isolate data	VEuPathDB	494	5071
Subcellular localization	VEuPathDB	358035	932
ESTs	VEuPathDB	460873	43
Compounds	VEuPathDB	227	1484

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 (September 1-30, 2021)

Tool/Service	BRC Domain	Submitted
Sequence retrieval tool	VEuPathDB	18492
BLAST	VEuPathDB	10412
Enrichment Analyses	VEuPathDB	2348

Web services	VEuPathDB	10260
Boolean operations	VEuPathDB	3756
Apollo (Access)	VEuPathDB	664
Site Search	VEuPathDB	159996
Galaxy Jobs	VEuPathDB	876
Genome Browser	VEuPathDB-	433348
User Comments	VEuPathDB	11
Multiple sequence alignment (isolates)	VEuPathDB	5523
Results downloads	VEuPathDB	5631
<i>Data analysis searches (breakdown below)</i>		
Annotation searches	VEuPathDB	4519
Epigenomics	VEuPathDB	6
Function prediction	VEuPathDB	477
Gene models	VEuPathDB	83
Genetic variation	VEuPathDB	126
Genomic Location	VEuPathDB	207
Immunology	VEuPathDB	68
Orthology and synteny	VEuPathDB	1301
Pathways and interactions	VEuPathDB	40
Phenotype	VEuPathDB	3461
Protein features and properties	VEuPathDB	474
Protein targeting and localization	VEuPathDB	932
Proteomics	VEuPathDB	1972
Sequence analysis	VEuPathDB	8330
Structure analysis	VEuPathDB	22
Taxonomy	VEuPathDB	375
Text	VEuPathDB	797
Transcriptomics	VEuPathDB	36028

Popset Isolate Sequences	VEuPathDB	5071
Genomic Sequences	VEuPathDB	2666
Genomic Segments	VEuPathDB	134
SNPs	VEuPathDB	5454
ESTs	VEuPathDB	43
Metabolic Pathways	VEuPathDB	152
Compounds	VEuPathDB	1484

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+CryptosporidiumDB+OR+TrichDB+OR+GiardiaDB+OR+TriTrypDB+OR+AmoebaDB+OR+Microsporidi>

[aDB+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	936	10916
Citations of BRC Resources	1880	24500

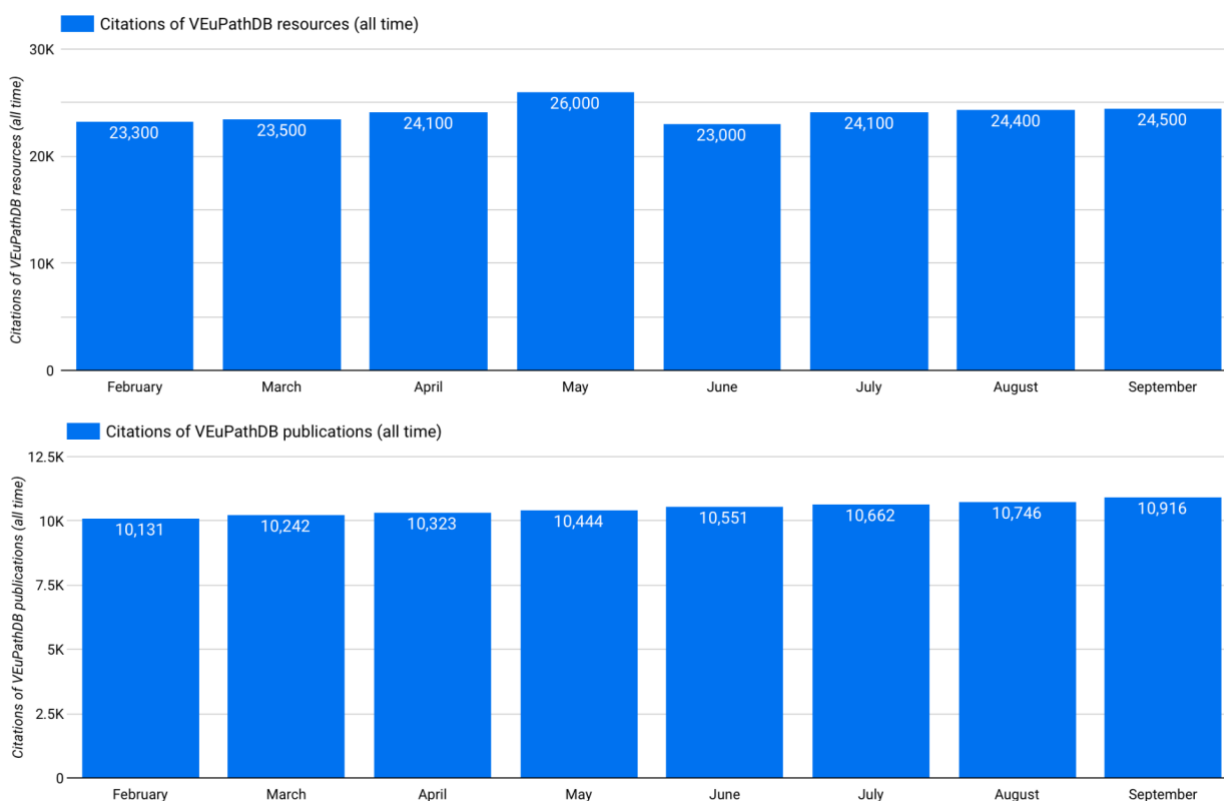


Figure 2 Cumulative citations of VEuPathDB resources and publications

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 (September 1-30)

Metric	Results (reporting period)
VEuPathDB integrated user data	~53G
Galaxy user data	~6.8T
User requests for help (some fit multiple categories and total may be >100%)	38(18% bugs, 55% help, 3% new data, 16% new feature, 11% other)
Webinar/workshop events and participants	Release 54 webinar, 28 attendees
Followers on social media: (reported as total)	
FaceBook @VEuPathDB	1,842
FaceBook @FungiDB	571
FaceBook @VectorBase	2,186
Twitter @VEuPathDB	2889
Twitter @FungiDB	3246
Twitter @VectorBase	1981
YouTube	550
BRC Reddit subscribers	55

BRC Reddit views	917
------------------	-----

VEuPathDB Support Emails

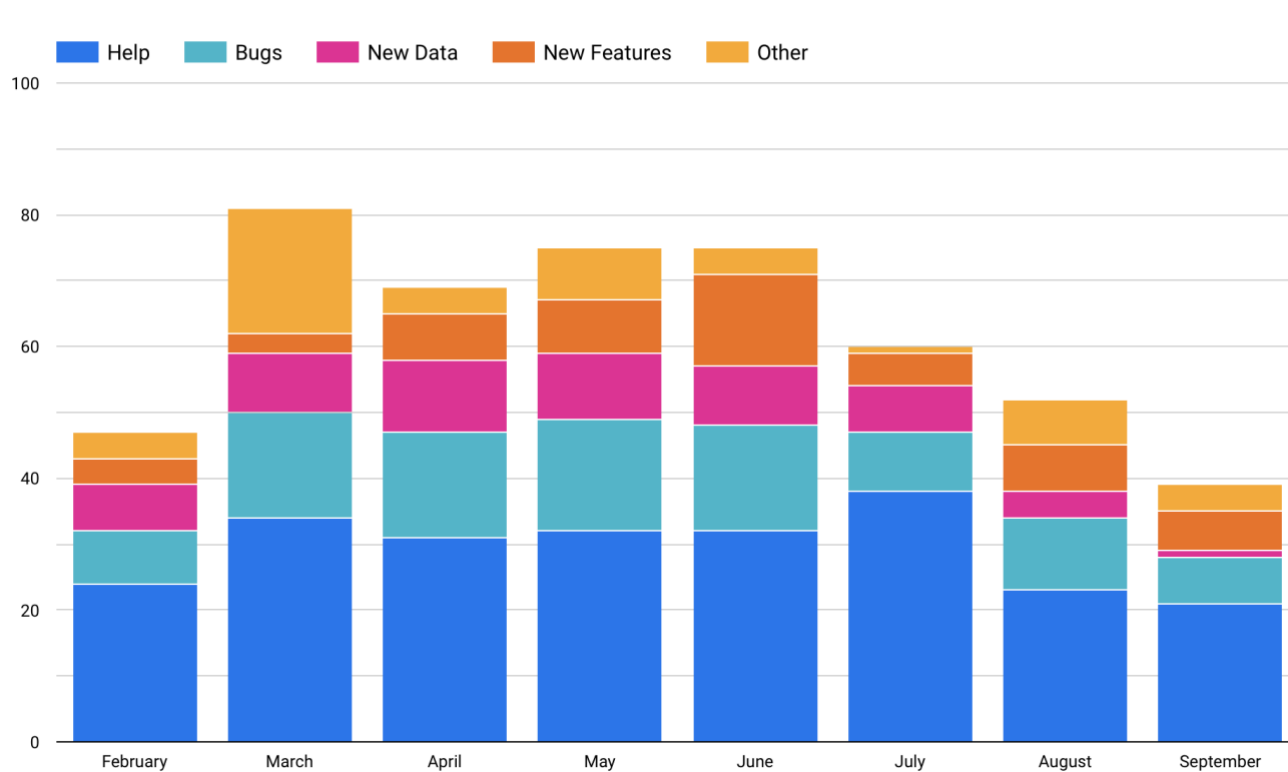


Figure 3 VEuPathDB Support emails over time