VEuPathDB BRC contract HHSN75N93019C00077

Usage Metrics Report

Reporting Period: October 1-31, 2021
Submission Date: November 10, 2021

Notes & Change Log

Date	Version/release	Description & Notes
11/10/2021	1	VEuPathDB Usage Metrics for October 2021- In response to COR feedback ● A high-level interpretation of trends has been added to figure legends. ● A count of Total Unique Logins has been added to Website Usage Metrics (Table 1).

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 unique logins**
  - *Definition* - A unique login is a registered user that has logged into at least one site at least one time in the reporting period. Each registered user is counted only once. This measure does not provide the total number of logins in a reporting period. This metric should not be used to measure site usage since login is not required to use the sites and users may not have chosen to log in (see 'Total unique visitors').
  - *Measurement mechanism* - User database query.
  - *Measure* - Total number of unique registered users that log in at least once 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 (October 1-31, 2021)

Metric	Result
Total registered users	24081
Total visits	77,073
Total unique visitors	33,423
Total unique logins	439
Total pageviews	10,464,041
Avg. pages / visit	135.76
Avg. visits / visitor	2.3
Avg. visit duration (seconds)	572
Bandwidth (GB)	704.43

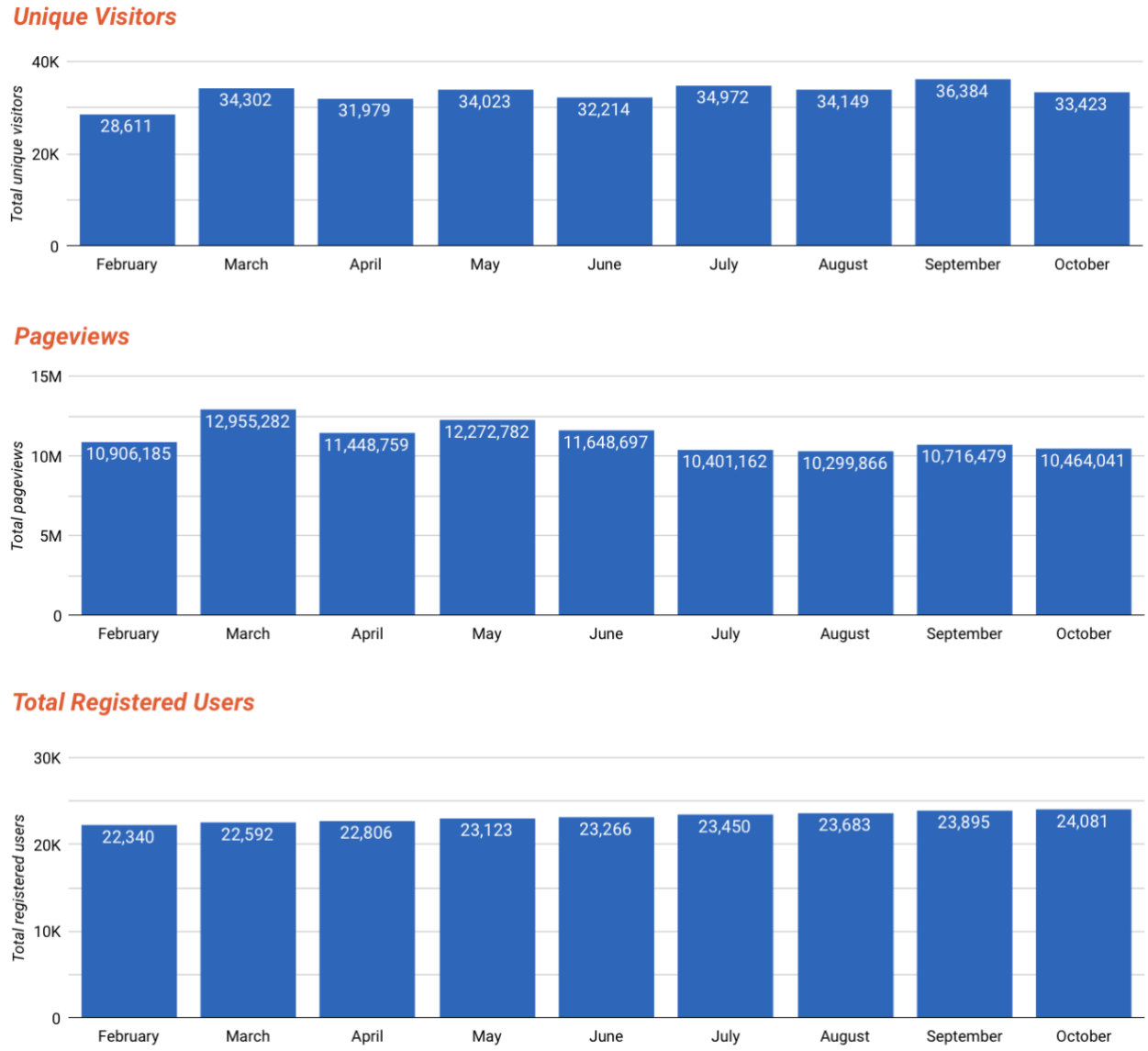

Figure 1 Unique visitors, Page views, and Total registered users over time. *Note: While page views do fluctuate over time, the number of pages viewed remains high, indicating high usage overall. This is consistent with observations pre-dating BRC4. We typically do not see dramatic changes in this metric unless measurement algorithms change or there are actual changes in user behavior. Both Total Unique Visitors and Total Registered Users remain high and slowly increasing over time, indicating continued use of the resource and an expansion of the number of users.*

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 (October 1-31, 2021)

Taxa	Domain	Page Views	# of Species	# of Genome Seqs
Plasmodium	Protozoa	175036	22	51
Toxoplasma	Protozoa	67094	1	15
Leishmania	Protozoa	45642	15	24
Trypanosoma	Protozoa	45028	8	25
Aedes	Vectors	14831	2	3
Anopheles	Vectors	11620	19	24
Cryptococcus	Fungi	10452	5	10
Aspergillus	Fungi	9824	23	28
Cryptosporidium	Protozoa	7075	7	11
Pyricularia	Fungi	6804	1	2
Neurospora	Fungi	6000	3	3
Fusarium	Fungi	5434	7	13
Saccharomyces	Fungi	5243	1	1
Entamoeba	Protozoa	3436	5	9
Giardia	Protozoa	2939	4	6
Trichomonas	Protozoa	2224	1	1
Glossina	Vectors	1847	6	7
Culex	Vectors	1828	1	1
Babesia	Protozoa	1654	6	6
Eimeria	Protozoa	1594	8	8
Phytophthora	Fungi	1506	7	7
Drosophila	Vectors	1389	1	1
Candida	Fungi	1345	8	15
Rhodnius	Vectors	1105	1	1
Ixodes	Vectors	1090	1	2
Theileria	Protozoa	1083	4	4

Crithidia	Protozoa	963	1	1
Lutzomyia	Vectors	930	1	1
Neospora	Protozoa	763	1	2
Naegleria	Protozoa	747	2	3
Pythium	Fungi	733	2	2
Homo	Host	629	1	1
Paracoccidioides	Fungi	542	2	3
Chromera	Protozoa	528	1	1
Ustilago	Fungi	450	1	1
Bodo	Protozoa	437	1	1
Cimex	Vectors	396	1	1
Sarcocystis	Protozoa	387	1	2
Phlebotomus	Vectors	381	1	1
Leptomonas	Protozoa	368	2	2
Schizosaccharomyces	Fungi	361	3	3
Angomonas	Protozoa	344	1	1
Leptotrombidium	Vectors	328	1	1
Musca	Vectors	328	1	1
Coccidioides	Fungi	317	2	6
Botrytis	Fungi	315	1	1
Hepatocystis	Protozoa	309	1	1
Hammondia	Protozoa	291	1	1
Encephalitozoon	Protozoa	283	4	9
Pediculus	Vectors	282	1	1
Malassezia	Fungi	279	3	4
Stomoxys	Vectors	278	1	1
Culicoides	Vectors	269	1	1

Zymoseptoria	Fungi	224	1	2
Histoplasma	Fungi	223	1	5
Acanthamoeba	Protozoa	220	1	1
Besnoitia	Protozoa	216	1	1
Paratrypanosoma	Protozoa	214	1	1
Nosema	Protozoa	201	2	3
Endotrypanum	Protozoa	199	1	1
Clavispora	Fungi	199	1	1
Blechomonas	Protozoa	197	1	1
Trichoderma	Fungi	188	2	3
Sarcoptes	Vectors	182	1	1
Coprinopsis	Fungi	162	1	1
Cystoisospora	Protozoa	157	1	1
Cyclospora	Protozoa	156	1	2
Globisporangium	Fungi	154	3	4
Biomphalaria	Vectors	146	1	1
Spizellomyces	Fungi	142	1	1
Kwoniella	Fungi	140	4	4
Pleurotus	Fungi	128	1	1
Puccinia	Fungi	123	4	5
Phanerochaete	Fungi	122	1	1
Mucor	Fungi	118	2	2
Anncaliia	Protozoa	99	1	2
Sclerotinia	Fungi	98	1	1
Blastomyces	Fungi	95	3	4
Mus	Host	91	1	1
Colletotrichum	Fungi	90	1	1

Gregarina	Protozoa	89	1	1
Sporisorium	Fungi	85	1	1
Hepatospora	Protozoa	82	1	2
Rhizophagus	Fungi	79	1	2
Talaromyces	Fungi	75	2	2
Penicillium	Fungi	72	1	1
Trichosporon	Fungi	68	1	1
Podospora	Fungi	65	1	1
Batrachochytrium	Fungi	64	1	1
Allomyces	Fungi	62	1	1
Claviceps	Fungi	62	1	1
Phycomyces	Fungi	60	1	1
Rhizopus	Fungi	59	1	1
Yarrowia	Fungi	57	1	2
Sordaria	Fungi	56	1	1
Fonsecaea	Fungi	56	1	1
Kluyveromyces	Fungi	54	1	1
Thermothelomyces	Fungi	52	1	1
Tremella	Fungi	51	1	1
Cladophialophora	Fungi	48	2	2
Melampsora	Fungi	47	1	1
Sporothrix	Fungi	47	2	2
Ascosphaera	Fungi	44	1	1
Pseudogymnoascus	Fungi	43	1	1
Monocercomonoides	Protozoa	41	1	1
Scedosporium	Fungi	41	1	1
Spironucleus	Protozoa	40	1	1

Saprolegnia	Fungi	39	2	2
Mitosporidium	Protozoa	37	1	1
Pseudoloma	Protozoa	36	1	1
Vittaforma	Protozoa	36	1	1
Trachipleistophora	Protozoa	35	1	1
Nematocida	Protozoa	35	3	5
Cytauxzoon	Protozoa	35	1	1
Mastigamoeba	Protozoa	34	1	1
Lomentospora	Fungi	34	1	1
Albugo	Fungi	34	2	2
Hyaloperonospora	Fungi	32	1	1
Vavraia	Protozoa	30	1	1
Enterocytozoon	Protozoa	30	2	2
Vitrella	Protozoa	29	1	1
Cenococcum	Fungi	29	1	1
Spraguea	Protozoa	28	1	1
Penicilliopsis	Fungi	28	1	1
Ophiostoma	Fungi	27	1	1
Rhinocladiella	Fungi	26	1	1
Bremia	Fungi	26	1	1
Bos	Host	24	1	1
Aphanomyces	Fungi	24	2	2
Exophiala	Fungi	23	3	3
Blumeria	Fungi	22	1	1
Pichia	Fungi	21	1	1
Pneumocystis	Fungi	21	1	1
Verruconis	Fungi	21	1	1

Hanseniaspora	Fungi	19	2	2
Uncinocarpus	Fungi	19	1	1
Enterospora	Protozoa	17	1	1
Cyphellophora	Fungi	15	1	1
Phytopythium	Fungi	11	1	1
Peronospora	Fungi	9	1	1
Ordospora	Protozoa	6	1	1
Macaca	Host	1	2	2

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 (October 1-31, 2021)

Data Type	Domain	Page Views	Searches
Taxonomy	VEuPathDB	423592	493
Genomes	VEuPathDB	423592	1464
Genome sequences	VEuPathDB	423592	3216
Genes/Proteins	VEuPathDB	423592	66117
Transcriptomics	VEuPathDB	340966	36664
Proteomics	VEuPathDB	334229	1888
Variation data	VEuPathDB	254153	5614
Epigenomics	VEuPathDB	219703	3
Enzyme commission	VEuPathDB	148722	129
Gene Ontology	VEuPathDB	290473	257
Protein domains	VEuPathDB	423592	230
Immunology	VEuPathDB	353409	80
Gene Orthology	VEuPathDB	419776	874

Synteny	VEuPathDB	423592	NA
Metabolic pathways	VEuPathDB	1016	170
Phenotype	VEuPathDB	57053	3531
Isolate data	VEuPathDB	626	5147
Subcellular localization	VEuPathDB	265739	4038
ESTs	VEuPathDB	381570	64
Compounds	VEuPathDB	258	1528

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 by users**
  - *Definition* - The total number of analysis tasks submitted 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 per month.
- **Analysis tasks submitted 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 (October 1-31, 2021)

Note: The total number of Galaxy jobs submitted was under-reported from April to September. This was a result of an error in the report generation SOP. The correct numbers are below. If desired, these reports can be resubmitted with the corrected metrics.

- *April 3,727*
- *May 3,710*
- *June 2,908*
- *July 4,406*
- *August 1,155*
- *September 1,232*

Note: The relatively lower number of Galaxy jobs in August and September is accurate and we attribute it to increased vacation time for users in this period. This trend is similar if we look at total number of jobs during the same period in previous years. We have contacted Navipoint and asked them to inspect the metrics to ensure this is correct.

Tool/Service	**BRC Domain**	**Submitted**
Sequence retrieval tool	VEuPathDB	22538
BLAST	VEuPathDB	10752
Enrichment Analyses	VEuPathDB	1670
Web services	VEuPathDB	14128
Boolean operations	VEuPathDB	2260
Apollo (Access)	VEuPathDB	546
Site Search	VEuPathDB	148710
Galaxy Jobs	VEuPathDB	3531
Genome Browser	VEuPathDB-	427269
User Comments	VEuPathDB	52
Multiple sequence alignment (isolates)	VEuPathDB	5135
Results downloads	VEuPathDB	4568
Data analysis searches (breakdown below)		
Annotation searches	VEuPathDB	4194
Epigenomics	VEuPathDB	3
Function prediction	VEuPathDB	386
Gene models	VEuPathDB	157
Genetic variation	VEuPathDB	70

Genomic Location	VEuPathDB	481
Immunology	VEuPathDB	80
Orthology and synteny	VEuPathDB	874
Pathways and interactions	VEuPathDB	52
Phenotype	VEuPathDB	3531
Protein features and properties	VEuPathDB	3804
Protein targeting and localization	VEuPathDB	4038
Proteomics	VEuPathDB	1888
Sequence analysis	VEuPathDB	8230
Structure analysis	VEuPathDB	27
Taxonomy	VEuPathDB	493
Text	VEuPathDB	1145
Transcriptomics	VEuPathDB	36664
Popset Isolate Sequences	VEuPathDB	5147
Genomic Sequences	VEuPathDB	3053
Genomic Segments	VEuPathDB	163
SNPs	VEuPathDB	5544
ESTs	VEuPathDB	64
Metabolic Pathways	VEuPathDB	170
Compounds	VEuPathDB	1528

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

Table 5: Citations

Metric	Year to date	Cumulative
Citations of BRC Publications	995	11080
Citations of BRC Resources	2040	24800

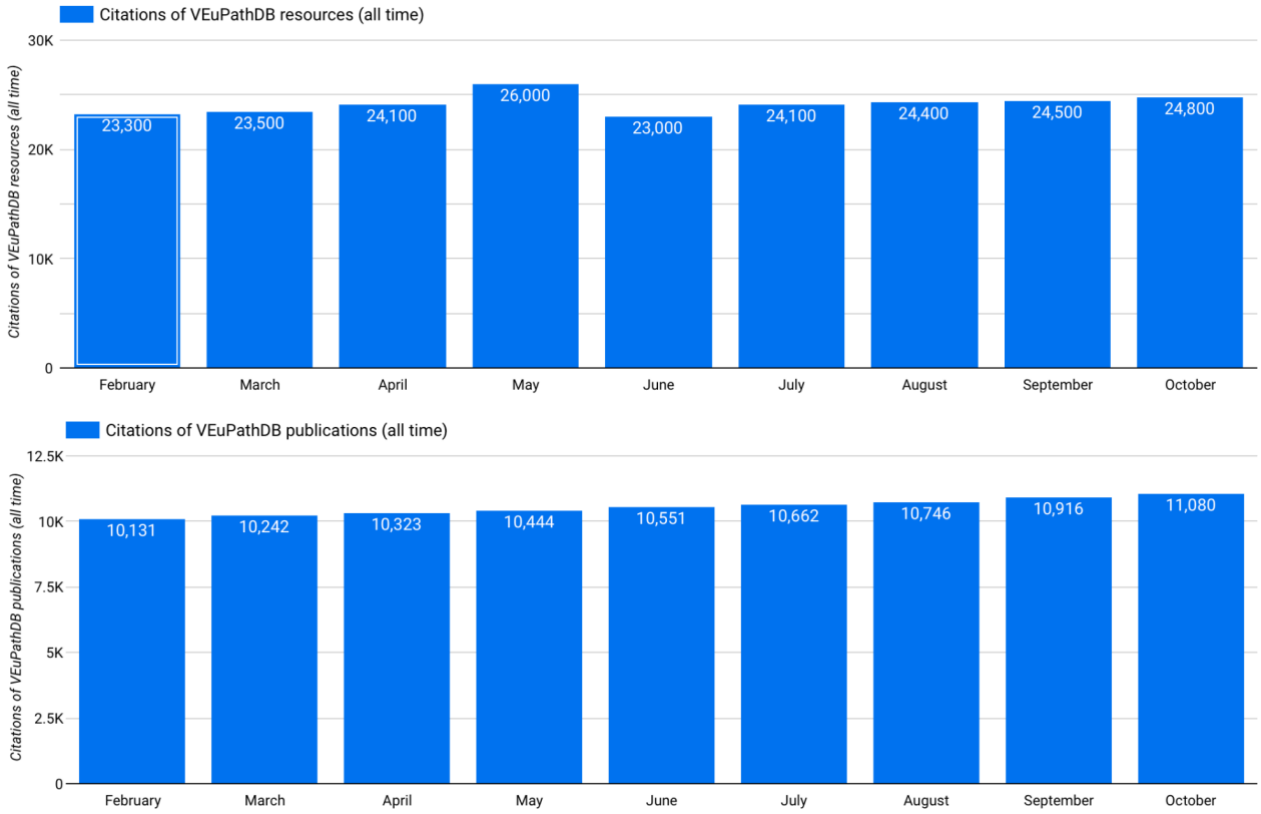

Figure 2 Cumulative citations of VEuPathDB resources and publications. *Note: The number of citations and publications remains high and increasing over time. This is consistent with observations pre-dating BRC4. Changes to this trend typically indicate changes in the algorithms used to count citations (e.g. April, May, and June for Citations of VEuPathDB resources).*

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 (**October 1-31, 2021)**

Metric	**Results (reporting period)**
VEuPathDB integrated user data	~53G
Galaxy user data	~12T
User requests for help (some fit multiple categories and total may be >100%)	36 (8% bugs, 61% help, 8% new data, 8% new feature, 14% other)
Webinar/workshop events and participants	None
Followers on social media: (reported as total)	
FaceBook @VEuPathDB	1,845
FaceBook @FungiDB	572
FaceBook @VectorBase	2,187
Twitter @VEuPathDB	2,910
Twitter @FungiDB	3,268
Twitter @VectorBase	1,992
YouTube	560
BRC Reddit subscribers	56
BRC Reddit views	515

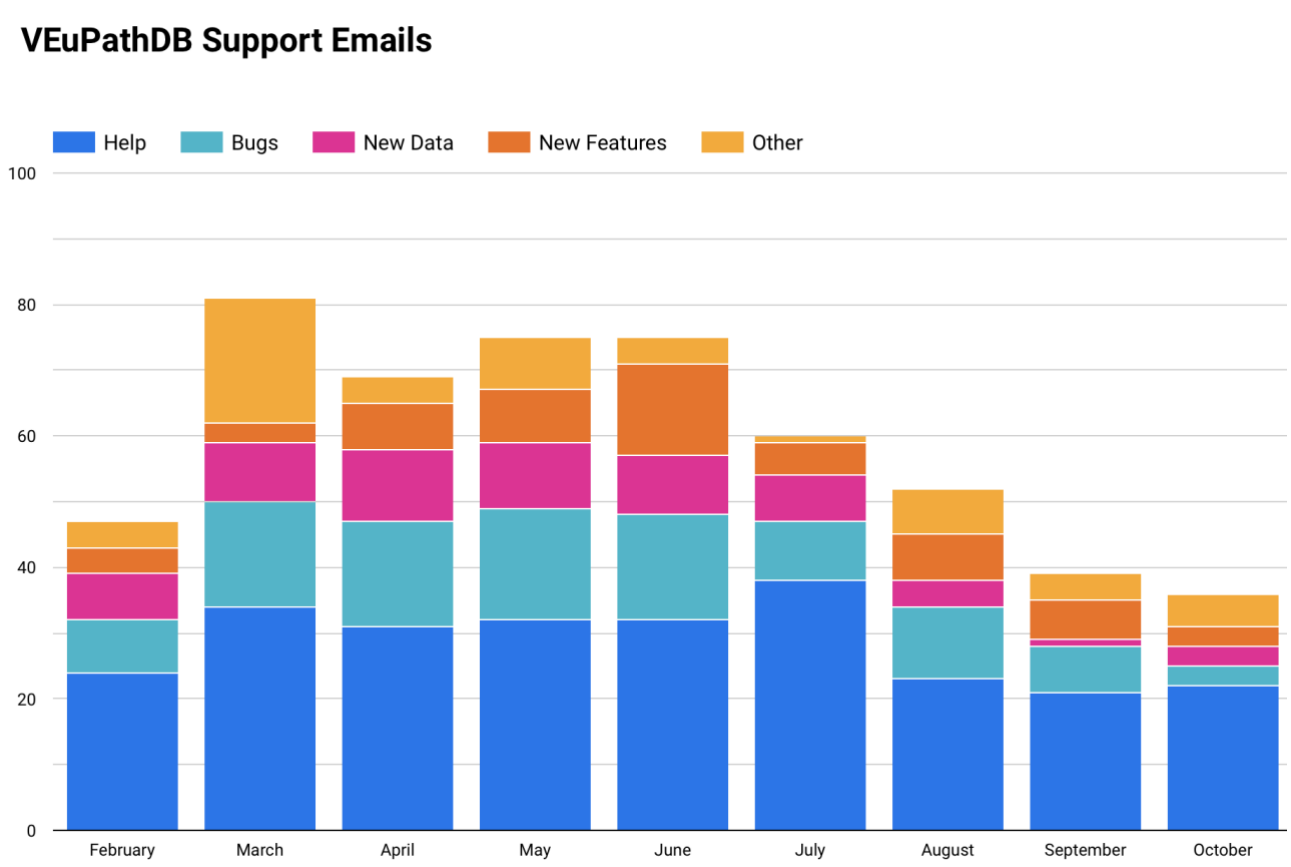

Figure 3 VEuPathDB Support emails over time. *Note: The relative proportion of different types of user requests remain generally consistent over time (e.g., help requests are the single most abundant types of support emails). Trends in the overall number of support emails likely represent annual vacation trends among users but may also indicate increased use of our FAQ resource released earlier this year (https://veupathdb.org/veupathdb/app/static-content/landing.html).*