



VEuPathDB BRC contract HHSN75N93019C00077

Performance Metrics Report

Reporting Period: February 1-28, 2021

Submission Date: March 10, 2021

Revision History

| Date | Version/release | Description |
|-----------|-----------------|--|
| 3/10/2021 | 1 | Initial VEuPathDB Performance Metrics Report |

Joint-BRC Common System Performance Metrics Plan

This report is available from all VEuPathDB sites, e.g., <https://veupathdb.org/> , from the About menu.

This monthly systems performance metrics report provides a summary of the VEuPathDB BRC performance for the current reporting period in accordance with the Joint-BRC Common System Performance Metrics Plan developed by the BRCs and subsequently approved by NIAID.

As per the plan, each BRC will report and aggregate performance 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 performance of the BRC resources to identify trends, areas that are performing well, and areas for improvement. Once the system performance plan is approved by NIAID, each BRC will submit a system performance report for their resource on a monthly basis. 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 data, complexity of various analysis tools, 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 system performance 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 Performance

Every month, each BRC will report the performance of the key web pages from their website, starting with the pages listed in the table below and adding new pages as they are released on the website. For each page, the average page load time will be computed based on a predefined set of pages and compared against the target page load time set as a target benchmark. This will help us ensure that the performance of the individual pages and the overall website is maintained as the amount of data and usage increase with time over the life of the project. If performance of any of the pages is below the set benchmark, we will address it by performing necessary hardware or software optimizations.

- **Target page load time**

- *Definition* - Target page load time measured in seconds, set as a benchmark. The target page load times may vary for various pages depending on their complexity and amount of data they present / visualize to the user.
- *Measurement mechanism* - Manual / custom performance measurement scripts.
- *Measure* - Page load time in seconds.

- **Average page load time**

- *Definition* - Average page load time measured in seconds, after N executions. The average page load times may vary for various pages depending on their complexity and amount of data they present / visualize to the user. Hence, average load time for a web page should be compared only to the benchmark set for that page.
- *Measurement mechanism* - Manual / custom performance measurement scripts.
- *Measure* - Average page load time measured seconds, after N executions.

Table 1 VEuPathDB Website Performance (Feb 1-28, 2021)

| Web Page | BRC Domain | Target Load Time (Seconds) | Avg. Load Time (Seconds) |
|---|------------|----------------------------|--------------------------|
| Home page | VEuPathDB | 5 | 3.01 |
| Gene search form with filterParam | VEuPathDB | 5 | 2.09 |
| Gene search result (default organism) | VEuPathDB | 5 | 4.67 |
| Gene record page | VEuPathDB | 5 | 1.97 |
| Site search result | VEuPathDB | 5 | 1.5 |
| Organism table (search result on strategy panel) | VEuPathDB | 5 | 1.52 |
| Data Sets table (search result on answerController) | VEuPathDB | 5 | 2.05 |
| Fasta SRT result (click submit) | VEuPathDB | 5 | 0.96 |

Service/Tool Performance

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. Both the BRCs will monitor and report the performance of all analysis services/tools available in their resource on a monthly basis. The performance reports will be generated based on the actual usage of these services/tools by BRC users in a given month. For each analysis service, we will compute the total number of jobs submitted by users, number of jobs completed successfully, failed, average wait time for the jobs queued in the system, and average run time. Monitoring the fraction of jobs that fail and/or reported by users will allow us to identify recurring problems and address them in a timely manner to make the services more robust and reliable. The job wait time depends on the variation in the usage patterns and system load, while the run time depends heavily on the size of the input data and the parameters selected. Monitoring these metrics will allow us to identify factors affecting the overall performance of the application services and tools and address them by performing necessary software and/or hardware scaling or optimization to meet the user expectations.

- **Analysis tasks submitted**

- *Definition* - A breakdown of total number of analysis tasks submitted by users, summarized by service/tool, during the specified date range.
- *Measurement mechanism* - Captured via website and server logs, which are used to tally the number.
- *Measure* - Jobs per month, tallied by service/tool.

- **Analysis tasks completed**
 - *Definition* - A breakdown of total number of analysis tasks submitted by users and completed successfully, summarized by service/tool, during the specified date range.
 - *Measurement mechanism* - Captured via website and server logs, which are used to tally the number.
 - *Measure* - Jobs per month, tallied by service/tool.
- **Analysis tasks failed**
 - *Definition* - A breakdown of total number of analysis tasks submitted by users and failed, summarized by service/tool, during the specified date range.
 - *Measurement mechanism* - Captured via website and server logs, which are used to tally the number.
 - *Measure* - Jobs per month, tallied by service/tool.
- **Average run time by service/tool**
 - *Definition* - A breakdown of average run time for all analysis tasks submitted by users, summarized by service/tool, during the specified date range.
 - *Measurement mechanism* - Captured via website and server logs, which are used to tally the number.
 - *Measure* - Average run time measured in seconds, tallied by service/tool.
- **Input limits**
 - *Definition* - Maximum size of the input supported by a service/tool, beyond which it may degrade the performance or fail to produce results.
 - *Measurement mechanism* - Defined by requirements, design and/or testing of a service/tool.
 - *Measure* - Input size defined as number or size of the input parameters. The units can vary depending on tool/service.

Table 2. VEuPathDB Tools/Services Performance Metrics (Feb 1-28, 2021)

| Tool/Service | BRC Domain | Jobs Submitted | Jobs Completed | Jobs Failed | Avg Run Time | Input limits |
|--------------------------------|------------|----------------|----------------|-------------|--------------|--------------|
| BLAST | VEuPathDB | 10570 | 10447 | N/A | 6.5 sec | 31kb |
| Galaxy Jobs - Details below: | VEuPathDB | | | | | |
| FastQC | VEuPathDB | 398 | 370 | 1 | 891 | N/A |
| Data Upload | VEuPathDB | 389 | 385 | 1 | 39 | N/A |
| FASTQ Groomer | VEuPathDB | 357 | 304 | 0 | 8649 | N/A |
| Trimmomatic | VEuPathDB | 193 | 173 | 0 | 2401 | N/A |
| HTSeqCountToTPM | VEuPathDB | 190 | 134 | 30 | 7499 | N/A |
| BAM to BigWig | VEuPathDB | 168 | 141 | 1 | 1923 | N/A |
| HISAT2 | VEuPathDB | 182 | 156 | 0 | 2279 | N/A |
| OrthoMCL Blast Parser | VEuPathDB | 224 | 134 | 60 | 89 | N/A |
| OrthoMCL Preprocess Fasta File | VEuPathDB | 151 | 139 | 8 | 400 | N/A |

| | | | | | | |
|------------------------------------|-----------|-----|----|----|------|-----|
| OrthoMCL Map Proteome To Groups | VEuPathDB | 93 | 82 | 0 | 49 | N/A |
| MCL Clustering | VEuPathDB | 71 | 63 | 0 | 59 | N/A |
| Tophat2 | VEuPathDB | 8 | 8 | 0 | 6929 | N/A |
| Deeptools BAM Coverage | VEuPathDB | 119 | 47 | 72 | 600 | N/A |

Database / Data API Performance

Both the BRCs will monitor database performance using predefined search and retrieval queries for various data types, measure average response time in seconds, and report it on a monthly basis. These database queries will capture the most common data queries used by various web pages and tools on the BRC websites as well as user queries used to download large amounts of data in batch mode using the data API, web services, or Command Line Interface (CLI). For each query, the average response time will be compared to the set benchmark. This will help us ensure that the performance of individual data queries as well as the overall database meets the performance benchmarks as well as user expectations. If the performance of any query does not meet the benchmark, we will address it by performing necessary database, query, or hardware optimizations.

- **Target response time**

- *Definition* - Target response time measured in seconds, set as a benchmark. The target response times may vary for various queries depending on the complexity of the query and amount of data retrieved.
- *Measurement mechanism* - Manual / custom performance measurement scripts.
- *Measure* - Page load time in seconds.

- **Average response time**

- *Definition* - Average response time measured in seconds, after N executions. The average response times may vary for various pages depending on the complexity of the query and amount of data retrieved. Hence, average load time for a web page should be compared only to the benchmark set for that page.
- *Measurement mechanism* - Manual / custom performance measurement scripts.
- *Measure* - Average response time measured seconds, after N executions.

Table 3 VEuPathDB Database / Data API Performance (Feb 1-28, 2021)

| Database Query | BRC Domain | Target Response Time (milliseconds) | Avg Response Time (milliseconds) |
|--|------------|-------------------------------------|----------------------------------|
| Data analysis searches (all, see below for breakdown): | VEuPathDB | NA | NA |
| Annotation searches | VEuPathDB | 1000 | 264 |
| Epigenomics | VEuPathDB | 1000 | 61 |
| Function prediction | VEuPathDB | 1000 | 38 |

| | | | |
|--|-----------|-------|-----|
| Gene models | VEuPathDB | 1000 | 61 |
| Genetic variation | VEuPathDB | 1000 | 70 |
| Genomic Location | VEuPathDB | 1000 | 42 |
| Immunology | VEuPathDB | 1000 | 38 |
| Orthology and synteny | VEuPathDB | 1000 | 39 |
| Pathways and interactions | VEuPathDB | 1000 | 37 |
| Phenotype | VEuPathDB | 1000 | 64 |
| Protein features and properties | VEuPathDB | 1000 | 46 |
| Protein targeting and localization | VEuPathDB | 1000 | 40 |
| Proteomics | VEuPathDB | 1000 | 37 |
| Sequence analysis | VEuPathDB | 1000 | 38 |
| Structure analysis | VEuPathDB | 1000 | 41 |
| Taxonomy | VEuPathDB | 1000 | 45 |
| Text | VEuPathDB | 1000 | 435 |
| Transcriptomics | VEuPathDB | 1000 | 38 |
| Popset Isolate Sequences | VEuPathDB | 1000 | 35 |
| Genomic Sequences | VEuPathDB | 1000 | 39 |
| Genomic Segments | VEuPathDB | 1000 | 41 |
| SNPs | VEuPathDB | 1000 | 98 |
| ESTs | VEuPathDB | 1000 | 37 |
| Metabolic Pathways | VEuPathDB | 1000 | 29 |
| Compounds | VEuPathDB | 1000 | 36 |
| Sequence retrieval tool | VEuPathDB | 1000 | 657 |
| Site Search | VEuPathDB | 1000 | 389 |
| User Comments | VEuPathDB | 1000 | 49 |
| Multiple sequence alignment (isolates) | VEuPathDB | 10000 | 161 |