

# BV-BRC

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

## Monthly System Performance Metrics Report

**Performance Period:** February 1, 2021 – February 28, 2021

*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:**

March 10, 2021

---

**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 System Performance Metrics Report

This monthly system performance metrics report provides a summary of the BV-BRC system 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.*, PATRIC and IRD/ViPR for BV-BRC. 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 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. BV-BRC Website Performance**

<b>Web Page</b>	<b>BRC Domain</b>	<b>Target Load Time (Seconds)</b>	<b>Avg. Load Time (Seconds)</b>
PATRIC home page	PATRIC	5	2.84
Global search – genome	PATRIC	5	2.47
Global search – gene	PATRIC	5	2.72
Taxonomy overview	PATRIC	5	3.29
Genome overview	PATRIC	5	12.36
Gene overview	PATRIC	5	4.66
Genome list	PATRIC	5	4.04
Feature list	PATRIC	5	3.63
Genome browser	PATRIC	5	3.34
Circular viewer	PATRIC	5	4.18
Home Page	IRD/ViPR	5	1.22
Genome Search	IRD/ViPR	10	8.64
Protein Search	IRD/ViPR	10	6.74
Strain Search	IRD/ViPR	10	8.96
Epitope Search	IRD/ViPR	10	8.24
Surveillance Search	IRD/ViPR	10	9.05
SFVT Search	IRD/ViPR	10	5.09
Strain Details Page	IRD/ViPR	5	4.96
Protein Details Page	IRD/ViPR	5	23.51
Protein Structure Search	IRD/ViPR	5	7.57

## 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. BV-BRC Tools/Services Performance Metrics**

<b>Tool/Service</b>	<b>BRC Domain</b>	<b>Jobs Submitted</b>	<b>Jobs Completed</b>	<b>Jobs Failed</b>	<b>Avg Run Time</b>	<b>Input Limits</b>
Codon Tree	PATRIC	532	521	11	12,533	100 genomes
Comprehensive Genome Analysis	PATRIC	1,541	1,385	156	9,756	-
Differential Expression	PATRIC	2	1	1	66	-
FastqUtils	PATRIC	369	348	21	523	-
Genome Alignment	PATRIC	178	168	10	698	20 genomes
Genome Annotation	PATRIC	172,683	160,024	12,659	164	-
Genome Annotation GenBank	PATRIC	38,130	37,788	342	246	-
Genome Assembly	PATRIC	9,615	6,291	3,324	2,496	-
Genome Comparison	PATRIC	177	169	8	578	10 genomes
Metagenome Binning	PATRIC	42	41	1	37,137	-
Metagenomic Read Mapping	PATRIC	413	407	6	344	-
Model Reconstruction	PATRIC	18	13	5	245	-
RNASeq Analysis	PATRIC	296	291	5	758	-
Taxonomic Classification	PATRIC	470	469	1	413	-
TnSeq Analysis	PATRIC	5	1	4	547	-
Variation Analysis	PATRIC	227	196	31	4,110	-
Alignment Viewer	IRD/ViPR	55	55	0	1,322	
Antiviral-Resistance-Risk	IRD/ViPR	15	15	0	1	-
BLAST	IRD/ViPR	418	413	5	533	-
Enrichment	IRD/ViPR	5	5	0	30	-
Genotype-Recombination	IRD/ViPR	31	8	23	17,126	-
H1-Clade Classifier	IRD/ViPR	80	79	1	2,140	-
H1N1-classifier	IRD/ViPR	15	15	0	28	-
H5N1-classifier	IRD/ViPR	62	60	2	2,056	-
Ha Numbering	IRD/ViPR	120	120	0	989	-

MGC	IRD/ViPR	70	65	5	3,590	-
MSA	IRD/ViPR	606	587	19	18,126	-
Muscle	IRD/ViPR	0	0	0	0	-
Mutation-analysis	IRD/ViPR	41	41	0	2,436	
Primer3	IRD/ViPR	29	29	0	845	-
RaxML hpc2_tgb	IRD/ViPR	0	0	0	0	-
Read-seq	IRD/ViPR	72	72	0	316	-
Rva Genotyper	IRD/ViPR	450	450	0	28	-
Short-seqsearch	IRD/ViPR	134	125	9	3,018	-
SNP-analysis	IRD/ViPR	341	340	1	5,278	-
Surveillance-data-mapping	IRD/ViPR	7	7	0	9	-
Tbl-formatter	IRD/ViPR	5	1	4	1,322	-
Tree	IRD/ViPR	927	911	16	4,372	-
VIGOR Annotator	IRD/ViPR	142	412	0	670	-
SARS-2 Genome Assembly and Annotation	BV-BRC	37	33	4	1,363	

## 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. BV-BRC Database / Data API Performance**

Database Query	BRC Domain	Target Response Time (Seconds)	Avg Response Time (Seconds)
Genome	PATRIC	5	0.06
Genome sequence	PATRIC	5	1.04
Genome feature	PATRIC	5	0.57
Genome AMR	PATRIC	5	0.04
Feature sequence	PATRIC	5	0.06
Specialty gene	PATRIC	5	0.05
Pathway	PATRIC	5	0.42
Subsystem	PATRIC	5	0.30
Taxonomy	PATRIC	5	0.05
Transcriptomics Experiment	PATRIC	5	0.05
Transcriptomics Sample	PATRIC	5	0.09
Transcriptomics Genes	PATRIC	5	0.78
PPI	PATRIC	5	0.21

Sequence & Strain	IRD/ViPR	5	5
Animal Surveillance	IRD/ViPR	5	1
3D Protein Structures	IRD/ViPR	5	1
Human Surveillance	IRD/ViPR	5	1
Serology Experiments	IRD/ViPR	5	2
Antiviral Drugs	IRD/ViPR	5	1
Host Factor Data	IRD/ViPR	5	2
Immune Epitopes	IRD/ViPR	5	2
Phenotypes	IRD/ViPR	5	2
PCR Primer Probe Data	IRD/ViPR	5	0.5
Sequence Feature Variant Types	IRD/ViPR	5	1
Human Clinical Studies and Lab	IRD/ViPR	5	1
Experiments	IRD/ViPR	5	1
ORFeome Plasmid Data	IRD/ViPR	5	1