StarBLAST

Release 1.0.0

Jun 02, 2021

Home

1	1 Welcome to StarBLAS	Г!		1
	1.1 What is Sequences	Server?		1
	1.2 The StarBLAST S	uite		2
	1.3 Contacts & Issues			2
	1.4 Official Publicatio	n & Citation		2
2	2 StarBLAST-VICE: We	b Deployment for Small Classes (<25)		3
	2.1 Launching StarBL	AST-VICE with Example Databases		3
	2.2 Adding Your Own	Databases to StarBLAST-VICE		6
	2.3 Launching StarBL	AST-VICE with Your Own Databases .		9
	2.4 Accessing your ru	nning Apps		9
3	3 StarBLAST-Docker: C	loud Deployment for Medium Classes (2	25-100)	13
3	3 StarBLAST-Docker: C 3.1 Launching Forema	loud Deployment for Medium Classes (2 an & Worker Instances	25-100)	13 13
3 4	 3 StarBLAST-Docker: Cl 3.1 Launching Forema 4 StarBLAST-HPC: HPC 	loud Deployment for Medium Classes (an & Worker Instances	25-100)	13 13 21
3 4	 3 StarBLAST-Docker: Cl 3.1 Launching Forema 4 StarBLAST-HPC: HPC 4.1 HPC Requirement 	loud Deployment for Medium Classes (2 an & Worker Instances C Deployment for Large Classes (>100) s and Setup	25-100)	13 13 21 21
3	 3 StarBLAST-Docker: Cl 3.1 Launching Forema 4 StarBLAST-HPC: HPC 4.1 HPC Requirement 4.2 Launching Worker 	loud Deployment for Medium Classes (: an & Worker Instances C Deployment for Large Classes (>100) s and Setup rs on the HPC	25-100)	13 13 21 21 23
3	 3 StarBLAST-Docker: Cl 3.1 Launching Forema 4 StarBLAST-HPC: HPC 4.1 HPC Requirement 4.2 Launching Worker 4.3 Setting Up the Ma 	loud Deployment for Medium Classes (an & Worker Instances (an & Classes (>100) C Deployment for Large Classes (>100) (an A Setup	25-100)	13 13 21 21 23 24
3 4 5	 3 StarBLAST-Docker: Cl 3.1 Launching Forema 4 StarBLAST-HPC: HPC 4.1 HPC Requirement 4.2 Launching Worker 4.3 Setting Up the Ma 5 Using SequenceServer 	loud Deployment for Medium Classes (an & Worker Instances C Deployment for Large Classes (>100) as and Setup rs on the HPC ster VM on the Cloud Service	25-100)	 13 21 21 23 24 27
3 4 5	 3 StarBLAST-Docker: Cl 3.1 Launching Forema 4 StarBLAST-HPC: HPC 4.1 HPC Requirement 4.2 Launching Worker 4.3 Setting Up the Ma 5 Using SequenceServer 5.1 SequenceServer's 	loud Deployment for Medium Classes (an & Worker Instances (an & Worker Instances C Deployment for Large Classes (>100) (as and Setup as and Setup (a) (as and Setup rs on the HPC (a) (as and Setup aster VM on the Cloud Service (a)	25-100)	 13 21 23 24 27 27
3 4 5	 3 StarBLAST-Docker: Cl 3.1 Launching Forema 4 StarBLAST-HPC: HPC 4.1 HPC Requirement 4.2 Launching Worker 4.3 Setting Up the Ma 5 Using SequenceServer 5.1 SequenceServer's 5.2 BLAST Loading & 	loud Deployment for Medium Classes (an & Worker Instances (an & Worker Instances C Deployment for Large Classes (>100) as and Setup (an A B B B B B B B B B B B B B B B B B B	25-100)	 13 21 23 24 27 27 29

CHAPTER 1

Welcome to StarBLAST!



StarBlast is a scalable extension of SequenceServer BLAST, making BLAST accessible to educators and researchers running classroom-scale searches concurrently.

StarBLAST utilizes cctools for faster, distributed computing and CyVerse's Visual Interactive Computing Environment (VICE).

1.1 What is SequenceServer?

SequenceServer is a front-end implementation of BLAST with improved GUI and customizable database input developed by the Wurmlab at Queen Mary University of London (Priyam et al., 2019). However, it is limited in its scaling capabilities and may be difficult to deploy for some people. StarBLAST extends SequenceServer for easier deployment and to scale for a larger number of simultaneous users (e.g., students).

1.2 The StarBLAST Suite

Navigate to each implementation for information on guided deployment:

- StarBLAST-VICE: Web Deployment for Small Classes (<25 students)
- StarBLAST-Docker: Cloud Deployment for Medium Classes (25-100 students)
- StarBLAST-HPC: HPC Deployment for Large Classes (>100 students)

1.3 Contacts & Issues

If you have questions, suggestions or have encountered a problem, please raise an issue on our GitHub Issues page.

1.4 Official Publication & Citation

The official publication at the Journal of Open Source Education (JOSE): https://doi.org/10.21105/jose.00102

Cite our work as: "Cosi, M., Forstedt, J., Gonzalez, E., Xu, Z., Peri, S., Tuteja, R., Blumberg, K., Campbell, T., Merchant, N. & Lyons E. (2021). StarBLAST: a scalable BLAST+ solution for the classroom. Journal Of Open Source Education, 4(38), 102. doi: 10.21105/jose.00102"

CHAPTER 2

StarBLAST-VICE: Web Deployment for Small Classes (<25)

StarBLAST-VICE is a customizable implementation of SequenceServer, deployed as a VICE (Visual and Interactive Computing Environment) web application and hosted on the CyVerse Discovery Environment (DE). StarBLAST-VICE is launchable with a maximum of 8 CPU cores, 16 GB RAM, and 512 GB disk space.

Note: Before proceeding, a CyVerse account is required. Click here to register or log in.

2.1 Launching StarBLAST-VICE with Example Databases

(1) Click on the following button to launch SequenceServer in CyVerse Discovery Environment with the SWISS-PROT protein database (**requires CyVerse account**). If you are already in the DE, you can navigate to the StarBLAST-VICE app through the Apps button and searching for "StarBLAST-VICE".

Note: SWISS-PROT is a curated protein sequence database, read more on the release or its orignal publication.

CyVerse Disco	overy Environment		₽ ± 0
Data Data Data: Genome: Upload ~ File ~ E Navigation	Apps Vorkflow Share Refresh Categories (My Apps Operation Topic HPC Apps under development Favorite Apps My public apps Shared with me My Communities	starBLAST-VICE Search results: 1 found for StarBLAST-VICE Name All StarBLAST-VICE-2.0 REETU TUTEJA de * * * * * (0)	Image Tools Image Tools Image Tools Image Tools

(2) Choose your own analysis name and the DE output folder. Click "Launch Analysis".

CyVerse Discov	ery Environment		₽	?
Data	Select output folder:			
Analyses	/iplant/home/cosimichele/analyses Retain Inputs? Enabling this flag will copy all the input files into the analysis result folder. Input	Browse		
	Create Quick Launch	Launch Analysis		

(3) Check the notifications Bell Icon for a link to access your SequenceServer instance. This might take a few minutes. Once the notification shows that the app is running, click on the link. This will open a loading screen

CyVerse Disco	overy Environment				r 🕹 👘
Data	⊊ Apps			A_thal_genome completed A_thal_genome running	0
Senomes	Apps 🗸 Workflow 🗸 Share 👻 Refresh	Create BLAST D	atabase	A_thal_genome submitted	
Upload - File - Eo	Categories « My Apps Operation Topic HPC	agave	* * * * * (0)	StarBLAST-VICE-2.0 running. <u>Ac</u>	cess your running analysis here.
▲ ∰ cosimichele	Apps under development			StarBLAST-VICE-2.0 submitted	
 ↓ ↓	 Favorite Apps My public apps Shared with me My Communities 	de	Create BLAST database- Upendra Kumar Devisett		
		de	Create BLAST database-: Upendra Kumar Devisett $\star \star \star \star \star (1)$	New Notifications (30)	Mark all as read
	09 A	thaliana	2020 May 2 17:30:12	:	¥
		ice	2020 May 2 17:40:09	:	

in a new tab; Once the app is loaded, you should be able to BLAST through the SequenceServer app.

Loading VICE analysis...

Kubernetes Deployment created.

Kubernetes Service created.

Kubernetes Ingress created.

Kubernetes pod phase: Pending

CYVERSE

SequenceServer 2.0.0.rc8

Paste query sequence(s) or drag file containing query sequence(s) in FASTA format here
L	· · · · · · · · · · · · · · · · · · ·
Protein databases	
Non-redundant UniProt	B/SwissProt sequences
Advanced parameters:	eg: -evalue 1.0e-5 -num_alignments 100 ? Open results in new tab BLAST
	Run BLAST locally with Sequenceserver. Y Tweet
	riease cite relevant data sources and, rityam et al. (2019) sequenceserver, a modern graphical user interface for custom bLAST databases.

- (4) To test, click here for a sample DNA sequence.
- (5) Paste the query sequence, select both the available databases and submit job.

2.2 Adding Your Own Databases to StarBLAST-VICE

To add your own BLAST databases you will need a .fasta (or .fa, .faa, .fna) file containing the reference sequences you'd like to use. These are easily aquirable from NCBI or other databases.

- (1) Within the CyVerse DE, click on the "Data" icon.
- (2) Select "Upload" and specify the import from your Desktop or its URL. This will be stored in your personal folder.

Control Contro Control Control				
Upload File * Edit * Download * Share * Metadata * @ Refreki Details Simple Upload from Desktop Cenomes Details Bulk Upload from Desktop Viewing: //plant/home/cosimichele/Genomes Details Import from URL Name Last Modified Size Import from CoGe Viewing: //plant/home/cosimichele/Genomes Details Import from CoGe 0.0_Taw 2020 May 2 17:12:06 File * Coll Import Genomes 0.0_Taw 2020 May 2 17:39:15 File * Gold * To view View View View View View View View V	Senomes			() 🗆 🖨 🖉
A simple typled from Destop A simple typled from Destop Import from URL Import from URL Import from Code C fenomes (n.p.	Upload 🗸 File 👻 Edit 👻 Download 🗸	Share 🗸 Metadata 🗸 🧬 Refresh		🔍 Trash 🔻
▲ Bulk Upload from Deskop Name Last Modified Size ▲ Import from URL Name Last Modified Size ● FUR.B.obicity_sub50 0.0_Trum 2020 May 2 17:42:04 Image: Size ● Genomes_n_p 0.0_LE_coli 2020 May 2 17:42:04 Image: Size ● Genomes_n_p 0.0_LE_coli 2020 May 2 17:42:04 Image: Size ● Genomes_n_p 0.0_LE_coli 2020 May 2 17:39:15 Image: Size ● Genomes_n_p 0.0_LE_coli 2020 May 2 17:39:15 Image: Size ● Genomes_n_p 0.0_LE_coli 2020 May 2 17:39:15 Image: Size ● Genomes_n_p 0.0_LE_coli 2020 May 2 17:39:15 Image: Size ● Genomes_n_p 0.0_LE_coli 2020 May 2 17:39:15 Image: Size ● Genomes_db 0.0_Lebrafish 2020 May 2 17:39:15 Image: Size ● Genomes_cott 0.0_Lablana 2020 May 2 17:30:12 Image: Size ● Genomes_cott 0.0_Lablana 2020 May 2 17:30:12 Image: Size ● Genomes_cott 0.0_Lablana 2020 May 2 17:12:05 8.0 KB Image: Size ● Genomes_cott Image: Size Image: Size Im	A Simple Upload from Desktop	* Genomes		Details
Import from URL Import from URL Import Genome from CoGe Import Genomes_n_p Genomes_n	A Bulk Upload from Desktop	 Viewing: /iplant/home/cosimichele/Genomes 		Select a file or folder to view its details
A Import Genome from CoGe 0 0,raw 2020 May 2 17:12:06 1 0 FLIX_plot(c)c_sub50 0 0,raw 2020 May 2 17:42:04 1 0 Genomes 0 0, 2, Vest 2020 May 2 17:42:04 1 0 Genomes_n_p 0 0, 2, Vest 2020 May 2 17:39:15 1 0 Genomes_n_p 0 0, 2, Vest 2020 May 2 17:39:15 1 0 Genomes_n 0 0, 2, Vest 2020 May 2 17:39:15 1 0 Genomes_n 0 0, 2, Vest 2020 May 2 17:39:15 1 0 Genomes_n 0 0, 2, Vest 2020 May 2 17:39:15 1 0 Genomes_n 0 0, 2, Vest 2020 May 2 17:39:15 1 0 Genomes_n 0 0, Mouse 2020 May 2 17:42:09 1 0 Genomes_not 0 0, A, thaliana 2020 May 2 17:12:05 8.0 KB 0 Genomes_db 2020 May 2 17:12:05 8.0 KB 1 0 Genomes_db 2020 May 2 17:12:05 8.0 KB 1 0 Genomes_dt 10, Rice 2020 May 2 17:12:05 8.0 KB 1 0 <td>🛓 Import from URL</td> <td>Name Name</td> <td>Last Modified Size</td> <td>details</td>	🛓 Import from URL	Name Name	Last Modified Size	details
Image: Control of the control of th	A Import Genome from CoGe	🔲 🚞 00_raw	2020 May 2 17:12:06	A
C Genomes C Genomes_n_p C Genomes_n_p C VABA_input C ABA_input C ABA_input	b C ELIR platicin sub50	🔲 🧰 01_E_coli	2020 May 2 17:40:05	
> @ Genomes_n_p > 003_C_elegans 2020 May 2 17:39:15 : > @ VABA_input > 004_Zebrafsh 2020 May 2 17:39:15 : > @ backup_EnvLog_S10 > 0.6_Drosophila 2020 May 2 18:13:26 : > @ backup_EnvLog_S10 > 0.6_Mouse 2020 May 2 17:56:59 : > @ blast_db_n > 0.9_A_thaliana 2020 May 2 17:42:09 : > @ fieldmosaic_out > 0.9_A_thaliana 2020 May 2 17:42:09 : > @ fieldmosaic_testset > 0.9_A_thaliana 2020 May 2 17:40:09 : > @ fieldmosaic_testset > 0.9_S_Store 2020 May 2 17:12:05 8.0 KB > @ fieldmosaic_testset > 0.9_S_Store 2020 May 2 17:12:05 8.0 KB > @ fieldmosaic_testset > 0.9_S_Store 2020 May 2 17:12:05 8.0 KB > @ soli_masLout > 0.9_S_Store 2020 May 2 17:12:05 8.0 KB : > @ storeoTop > .0 storeoTop > .0 updated_ortho_out	Finite protect prot	📄 🧰 02_Yeast	2020 May 2 17:42:04	
YABA_input analyses backup_EnvLog_S10 blast_db blast_db blast_db blast_db cope_data	Genomes_n_p	🔲 🧰 03_C_elegans	2020 May 2 17:39:15	
> Images Images Images Images Images Images Images Images Images Images Images Images Images Images Images Images Images Images	VABA_input	04 Zebrafish	2020 May 2 17:30:58	
> Code May 2 10:10:20 > Code May 2 10:10:	i i analyses	C 05 Drosophila	2020 May 2 18:13:26	
> bidgue_data > bidgue_data > bidgue_data > bidgue_data > bidgue_data > bidgue_data > bidgue_data > bidgue_data > bidgue_data > fiddmosaic_uette > fiddmosaic_testset > file/double > file/double </td <td>backup_EnvLog_S10</td> <td></td> <td>2020 May 2 10.10.20</td> <td></td>	backup_EnvLog_S10		2020 May 2 10.10.20	
> Disac_udu	bisque_data	00_Mouse	2020 May 2 18:14:18	
 Cached_betydb Coge_data Og_A_thaliana 2020 May 2 17:42:09 Image: Cached_betydb Og_A_thaliana 2020 May 2 17:30:12 Image: Cached_betydb I	⊳ 🚰 blast_db n	🔲 🧰 07_Human	2020 May 2 17:56:59	
D Coge_data D Gedmosaic_out D fieldmosaic_testset D fieldmosaic_testset D fill/cout D fieldmosaic_testset D Genomes_db 2020 May 2 17:12:05 8.0 KB D Sapefile D Sapefile D Sitched_ortho_out	Cached_betydb	🗐 🚞 08_Chimpanzee	2020 May 2 17:42:09	
 C fieldmosaic_out fieldmosaic_tetset C fin:2tf_outALL C fin:2tf_outALL C fin:0ut C meantemp_out C shapefile C sol_mask_out S stahed_ortho_out C updated_ortho_out 	⊳ 📁 coge_data	🔲 🚞 09_A_thaliana	2020 May 2 17:30:12	
 C fieldmosaic_testset C fieldmosaic_testset C filtrati-outALL C fi	Fieldmosaic_out	🔲 🧰 10_Rice	2020 May 2 17:40:09	
> Initizity_outALL Image: DS_Store 2020 May 2 17:12:05 8.0 KB > Initout Image: DS_Store 2020 May 2 17:12:05 8.0 KB > Image: Image: DS_Store 2020 May 2 17:12:05 8.0 KB > Image: Image: DS_Store 2020 May 2 17:12:05 8.0 KB > Image: Image: DS_Store 2020 May 2 17:12:05 8.0 KB > Image: Image: DS_Store Image: DS_Store Image: DS_Store > Image: Image: DS_Store Image: DS_Store Image: DS_Store > Image: DS_	iii fieldmosaic_testset	🔲 🧰 Genomes_db	2020 May 2 18:37:02	
> Intr Intr Intr Intr Intr Intr Intr Intr Intr Intr Intr Intr	▷ ☐ flir2tif_outALL	DS Store	2020 May 2 17:12:05 8.0 KB	
▷ intoloc	Firle		•	
Image: Solution of the soluti	b c meantemp out			
Image: Solution of the second of the seco	b is shapefile			
 ▷ ci stereoTop ▷ ci stitched_ortho_out ▷ ci updated_ortho_out 	i joint mask_out			
 ▷ G² stitched_ortho_out ▼ 	Image: StereoTop			
b 💋 updated_ortho_out	Image: provide the state of			
Diaphyling 1, 42 of 42	Updated_ortho_out	Displaying 1 42 of 42	0 item/a)	

(3) Click on the "Apps" icon and use the search bar to find "Create BLAST Database" or click here. As there is a possibility that there may be multiple apps with the same name, please locate the "Create BLAST Database" developed by developed by Upendra Kumar Devisetty (use the better reviewed one).

CyVerse Dis	covery Environment		
Dota	Apps Vorkflow - Share - 2 Refresh	Create BLAST Database	O □ ● ■ O Manage Tools hange Switch View
Apps Apps Analyses	Categories (My Apps Operation Topic HPC Apps under development Favorite Apps My public apps Shared with me My Communities	agave ★★★★(0) Create BLAST database-2 Upendra Kumar Devisetty ★★★★★(13) β	
		$\stackrel{\text{Upendra Kumar Devisetty}}{\bigstar \bigstar \bigstar \bigstar \bigstar (1)} \beta$	

- (4) Enter a name for your database under "Analysis Name"; this will become your database containing folder.
- 1. "Select output folder" should be your personal folder or any folder of your choice (default will be a foler named "analyses" within your personal folder).
- 2. In the "Inputs" tab, select "Browse" and choose the fasta file you uploaded. Select Nucleotide or Protein under "Input Sequence Format". Under "Prefix" choose a name to well reflect your database (e.g. a_thaliana).

C A D	Secreate BLAST database-2.6.0+	
	Analysis Name:A_thal_genome	
Data	Analysis Name:	
_	A_thal_genome	
	Comments:	
Apps		
	Select output folder:	
المالك	/iplant/home/cosimichele/analyses	Brows
Analyses	Retain Inputs? Enabling this flag will copy all the input files into the analysis result folder.	
	* Inputs	
	Parameters	

	0
	0
Analysis Name:A_thal_genome	
* Inputs	
If you need multiple input files for your database, please combine them using the "Concater Files" app	nate Multiple
* Input file:	(
P2 /iplant/home/cosimichele/Genomes/09_A_thaliana/GCF_000001735.4_TAIR10.1_gen	nomic.f Brow
* Input Sequence Format:	(
Nucleotide	
* Input type:	(
Fasta	
* Prefix to use for database:	(

(5) Click "Launch Analysis" and wait to be notified of its completion. Upon completion, navigate to the output folder specified in step 4.1; Inside you will find a directory with the name you specified in step 4 followed by a timestamp. Within this folder you will find logs and the newly generated database (if nucleotide .nhr,.nin, .nog, .nsd, .nsi, .nsq files will be found; if protein, .phr, .psq files will be found).

2.3 Launching StarBLAST-VICE with Your Own Databases

To launch StarBLAST-VICE with your own database:

- (1) Use the same button for Example Databases, but **do not click "Launch Analysis" just yet**.
- (2) In the "Input" tab, select the folder containing your database (if not specified, default is swissprot-db).

Note: The user will not be able to see the databases within the folder at this step. Ensure that databases files (as explained in step 4.2 of the previous section) are present beforehand.

(3) Click "Launch Analysis". This might take a few minutes.

2.4 Accessing your running Apps

The notification bell should show your currently running apps and jobs.

CyVerse Disc	overy Environment				r 🕹 🕹
Data	CApps	Create PLACE		A_thal_genome completed A_thal_genome running	3
Upload + File + Ec	Categories (My Apps Operation Topic HPC	agave	★★★★(0)	A_thal_genome submitted StarBLAST-VICE-2.0 running. <u>Acc</u> StarBLAST-VICE-2.0 submitted	ess your running analysis here.
	Apps under development Favorite Apps My public apps Shared with me My Communities	de	Create BLAST database- Upendra Kumar Devisett		
		de	Create BLAST database∹ Upendra Kumar Devisett ★ ★ ★ ★ ★ ★ (1)	New Notifications (30)	Mark all as read
	09_A	_thaliana	2020 May 2 17:30:12 2020 May 2 17:40:09	:	*

In case you want to see all your jobs and access your running apps (and app history), navigate to the Analyses button.

	Sequenceserver	
	Analysis Name:StarBLAST-VICE_tutorial	
Data	* Input	
	* database:	
	/iplant/home/shared/iplantcollaborative/example_data/GEA_Blast_dbs	Browse
Apps	* Number of CPUs:	
	8	~
<u> </u>	Resource Requirements	
Analyses		

se Discovery Environment		e 19
S Analyses		8 🖬 🖨 🖉 🔇
ata Analyses CRefresh	✓ All ✓ Q Search	
Name	Owner App Date	↓ End ↓ Status Date
StarBLAST-VICE_tutorial	cosimichelesequenceserver	Running
A_thal_genome	cosimicheleCreate BLAST database-2.6.0+	Completed
		•

CHAPTER 3

StarBLAST-Docker: Cloud Deployment for Medium Classes (25-100)

To deploy StarBLAST setup in a cloud provider, you will need accounts with those providers. This example uses XSEDE's JetStream Cloud service. You can access JetStream using an XSEDE account, a Globus account, or via institutional access to XSEDE (search for your institution name from the drop down menu in JetStream's login page).

This setup uses a "Foreman" instance for the front-end sequenceServer and one or more "Worker" instances to distribute the computational load of running blast. Docker containers are used to deploy the Foreman and Workers through deployment scripts. These deployment scripts are designed to:

- · Get appropriate Worker and Foreman containers
- Download specified BLAST databases (can customize)
- Launch sequenceServer front-end that can be accessed using the code: *<FOREMAN_IP_ADDRESS>* of the instance
- Connect factory of workers to the Foreman

3.1 Launching Foreman & Worker Instances

- 1. Login to JetStream Cloud.
- 2. From JetStream's top menu, navigate to "Projects" and select "Create New Project".

	Jetstream	II Dashboard	🖀 Projects	🗎 Images	O Help		cosi +
0	Projects					CREATE NEW	PROJECT
	Jetstrear	m Homepage - Jet	stream Partners	- Citing Jetstre	am - Jetstream is supported by NSF ACI-1445604	FEEDBACK & SUPPORT	

3. In the "Project Name" field, name your project and add a description.

jetstream di Dashbo	ard SProjects M images O Help	cosi +
0 Projects	Create Project × Project Name My BLAST Project Description	CREATE NEW PROJECT
	CANCEL CREATE	
Jetstream Homepa	ge - Jetstream Partners - Citing Jetstream - Jetstream is supported by NSF ACI-1445604	DBACK & SUPPORT

4. From JetStream's dashboard, select "Launch New Instance".

jetstream 🕼 Dashboard 🛎 Projec	cts 🍽 Images 🛛 Help	cosi +
Getting Started		
Launch New Instance Browse Atmosphere's list of available images and select one to launch a new instance.	Browse Help Resources View a video tutorial, read the how-to guides, or email the Atmosphere support team.	Change Your Settings Modify your account settings, view your resource quota, or request more resources.
Resources Used NEED MORE?		
Allocation Source	RN190014	0 Instances
Allocation	84.16% 50 60 70 80 90 100 t of Allocation Used	
Provider Resources Jetstream - Indiana Unive	ersity 📕 Jetstream - TACC	0 Volumes
CPU 0%		
Jetstream Homepage - Jetstream Par	tners - Citing Jetstream - Jetstream is supported by NSF ACI-1-	445604 FEEDBACK & SUPPORT

5. Be sure to change the default tab from "Show Featured" to "Show All", search for "Docker_starBLAST" and select the "Docker_starBLAST" image (or click here); click "Launch".

	Jetstream	📲 Dashboard	Projects 🗎 Images	s 🛛 Help		cosi 🗸
	Q SEARCH	FAVORITES (0)	MY IMAGES (0)	MY IMAGE REQUESTS	TAGS	
N	Image Se	arch				
	Docker_starBLAST	Docker starBLAST'				
V	All Images					
	Docker_Sta Apr 27th 20 0	rBlast 33:07 by admin	Imported Application - Dock	ker_StarBlast		
		Jetstream Homepage - Jet	tstream Partners - Citing Jets	stream - Jetstream is supported by NSF /	ACI-1445604 FEEDBACK & SUPPORT	

_je	etstream a	Dashboard 🗁 Projects 🎽 Imag	es 🛛 Help	cosi +	
Q SEAL	ARCH FAVOR	TES (0)	A MY IMAGE REQUESTS TAGS		
←	Docker_StarBla	st		ADD TO PROJECT	
954 1 95	Created: 4/ Created by: ad Description: Im Visibility: Pu Tags: Thi	27/2020 03:07 pm MST min ported Application - Docker_StarBlast blic resource has not been tagged.			
Versio	ons				
)H)	1.0 Apr 29th 20, 12:31 by adm	in		Available on Jetstream - TACC Jetstream - Indiana University	
	Jetstream	Homepage - Jetstream Partners - Citing Je	tstream - Jetstream is supported by NSF ACI-1445604	FEEDBACK & SUPPORT	

6. In the pop up menu you can customize your image (e.g. Instance Size. Use a minimum of m1.xlarge instance for Foreman, with at least 60GB disk space); select "Advanced Options".

Jetstream	II Dashboard 🖀 Projects 🗎 Images 🛛 Hel	p	cosi -
Q SEARCH	Launch an Instance / Basic Options	×	
	Basic Info	Resources	
← Docker_Sta	Instance Name	Allocation Source	TO PROJECT Launch
	Master	TG-IBN190014 🗘	
Created:	Base Image Version	Provider	
Description:	1.0 \$	Jetstream - TACC 🗘	
Visibility:	Project	Instance Size	
Tags:	My BLAST Project \$	m1.xlarge (CPU: 24, Mem: 60 GB, Disk: 60 GB)	
		Allocation Used	
		87% of 90000 SUs from TG-IBN190014	
Versions		Resources Instance will Use	
4 <u></u> 1.0		A total 68 of 132 alloted CPUs	
Apr 29th 20, 12:31		A total 180 of 360 alloted GBs of Memory) na University
	Advanced Options	CANCEL LAUNCH INSTANCE	
Jetstro	aam Homepage - Jetstream Partners - Citing Jetstream - <mark>Jets</mark>	tream is supported by NSF ACI-1445604 FEEDBACK & SUPP	ORT

7. Select "Create a New Script".

Jetstream	JI Dashboard 🐡 Projects 🎽 Images 🔍 Help	cosi +
Q SEARCH	Launch an Instance / Advanced Options	
	Deployment Scripts	
← Docker_Sta	Deployment scripts will be executed when a user has launched their instance. They will also be executed each time an instance is "Started", 'Resumed', or 'Restarted'. As such, these scripts should be able to handle being run multiple times without adverse effects.	IO PROJECT Launch
Description:	Add Scripts to Your Instance These Scripts will be Added	
Visibility: Tags:	Select scripts to add to your instance \$	
N	- OR -	
	CREATE A NEW SCRIPT	
Versions		
1.0 Apr 29th 20, 12:31		2 na University
	C Restore Default Options CONTINUE TO LAUNCH	
Jetstr	aam Homepage - Jetstream Partners - Citing Jetstream - Jetstream is supported by NSF ACI-1445604 FEEDBACK & SUP	PORT

8. Title the script "Foreman" or similar, select "Raw Text" and copy and paste the Foreman script, linked below. The scripts generate a password and username based on the user account, but these can be personalized if needed (*not suggested for new users*). Select "Save and Add Script" and then "Continue to Launch".

Deployment Scripts

- The deployment scripts for a *Foreman instance (atmo_deploy_master.sh)* can be found here.
- The deployment scripts for a *Worker instance (atmo_deploy_worker.sh)* can be found here.

Note: This step is required to be done **once** for the Foreman and **once for each Worker instance**. The deployment scripts are stored for future use.

Jetstream	Ju Dashboard Projects H Images 9 Help	×	cosi ~
Q SEARCH	Deployment Scripts		
← Docker_Sta	Create and Add a New Script		TO PROJECT Launch
Versions	Script Title Master script Input Type URL Raw Text Execution Strategy Type Run script on first boot Run script on each deployment Deployment Type Walk for script to complete	Raw Text #/bin/bash WORKQUEUE_PASSWORD= PROJECT_NAME= CPU_PER_WORKER=8	
1.0 Apr 29th 20, 12:31	C Restore Default Options	CANCEL CREATE SCRIPT	2 na University
Jetstr	eam Homepage - Jetstream Partners - Citing Jetstream - Jetstr	eam is supported by NSF ACI-1445604 FEEDBACK & SUPP	ORT

9. Repeat steps 4-8 for one or more Worker instance(s), using the Worker deployment script. Use large or extra large images (at least 60GB of disk space is required).

Note: JetStream cloud will take at least 10-20 minutes and the wait-time will increase with the size of the BLAST database.

Start BLASTING! Now anyone can enter the <FOREMAN_IP_ADDRESS> into their browser and access Sequence-Server.

Je	estream de Dasi	hboard 🖀 Projec	ts 💾 Image	es 😡 Help			cosi 🗸
III RES	SOURCES 🔲 DETAILS						OPTIONS -
Му	BLAST Proje	ct					
NE	w C O	•					
n:	stances						Sort By Name \$
	Name	Status	Activity	IP Address	Size	Provider	
	💥 <u>Master</u>	Active	N/A	129.114.104.47	M1.Xlarge	Jetstream - TACC	
	Master-HPC	 Active 	N/A	129.114.16.70	M1.Xlarge	Jetstream - TACC	
	Worker Worker	 Active 	N/A	149.165.169.13	M1.Xlarge	Jetstream - TACC	
 ■ Vo You hav ➡ Im You hav Q Lit You hav 	olumes ve not added any volumes to this nages ve not added any images to this nks ve not added any links to this pro	n project. project.					
	Jetstream Home	page - Jetstream Par	tners - Citing Jet	stream - Jetstream is supported	by NSF ACI-14456	D4 FEEDBACK & SUPPORT	

CHAPTER 4

StarBLAST-HPC: HPC Deployment for Large Classes (>100)

The StarBLAST-HPC Setup is designed to distribute BLAST searches across multiple nodes on a High-Performance Computer and uses a Master-Worker set-up similar to StarBLAST-Docker (an atmosphere instance as the Master, and the HPC as the Worker). It is suggested that the Worker is set up ahead of time.

Some command line knowledge is required for setup.

4.1 HPC Requirements and Setup

It is important that the following software are installed on the HPC:

- iRODS version 4.0 or newer
- ncbi-BLAST+ version 2.9.0 or newer
- CCTools version 7.0.21 or newer
- glibc version 2.14 or newer
- Support for CentOS7
- CyVerse user account

iRODS, ncbi-BLAST+ and CCTools should be available in your home directory, which can be found using

cd pwd

It should output something similar to

/home/<U_NUMBER>/<USER>/

4.1.1 iRODS Installation Guide

(1) From your home directory, obtain and install iRODS with the command

(2) Upon installation, set up the iCommands (requires a CyVerse account):

iinit

(3) You will be prompted to connect to the CyVerse with:

```
host name (DNS): data.cyverse.org
port #: 1247
username: <CyVerse_ID>
zone: iplant
password: <CyVerse_password>
```

iRODS should be installed and configured. If problems persists, a more in depth tutorial on iRODS and iCommands installation can be found here.

4.1.2 ncbi-BLAST+ Installation Guide

(1) From your home directory, obtain and decompress ncbi-BLAST+ with

(2) Add ncbi-BLAST+ to the path (change the path to reflect the correct location of the ncbi-BLAST+ bin files):

export PATH=\$HOME</PATH/TO/BLAST/BIN/>:\$PATH

At this point, ncbi-BLAST+ should be installed and accessible.

(3) BLAST databases need to be downloaded in a <DATABASE>/ directory in the home folder.

```
/home/<U_NUMBER>/<USER>/<DATABASE>/
```

Note: An example of BLAST databases can be downloaded with iRODS here: /iplant/home/cosimichele/ 200503_Genomes_n_p. Read more on installing iRODS and iCommands above.

4.1.3 CCTools Installation Guide

(1) From your home directory, obtain and decompress CCTools with

```
wget https://ccl.cse.nd.edu/software/files/cctools-7.1.6-source.tar.gz
tar -xvf cctools-7.1.6-source.tar.gz
```

(2) Add CCTools to the path (change the path to reflect the correct location of the CCTools bin files):

```
export PATH=$HOME</PATH/TO/CCTOOLS/BIN/>:$PATH
```

At this point, CCTools should be installed and accessible.

Note: CCTools only works if your HPC has glibc version 2.14 or newer. In the following examples, glibc and BLAST+ are loaded through module load. module load is not necessary if the HPC system already supports glibc 2.14 and if ncbi-BLAST+ has been added to the path as described above.

4.2 Launching Workers on the HPC

The HPC uses a .pbs and qsub system to submit jobs.

(1) Create a .pbs file that contains the following code and change the <VARIABLES> to preferred options:

```
#!/bin/bash
#PBS -W group_list=<GROUP_LIST>
#PBS -q windfall
#PBS -1 select=<N_OF_NODES>:ncpus=<N_OF_CPUS>:mem=<N_MEMORY>gb
#PBS -1 place=pack:shared
#PBS -l walltime=<MAX_TIME>
#PBS -1 cput=<MAX_TIME>
module load blast
module load unsupported
module load ferng/glibc
module load singularity
export CCTOOLS_HOME=/home/<U_NUMBER>/<USER>/<CCTOOLS_DIRECTORY>
export PATH=${CCTOOLS_HOME}/bin:$PATH
cd /home/<U NUMBER>/<USER>/<WORKERS DIRECTORY>
MASTER_IP=<MASTER_IP>
MASTER_PORT=<PORT_NUMBER>
TIME_OUT_TIME=<TIME_OUT_TIME>
PROJECT_NAME = < PROJECT_NAME >
/home/<U_NUMBER>/<USER>/<CCTOOLS_DIRECTORY>/bin/work_queue_factory -T local -M
→$PROJECT_NAME --cores <N_CORES> -w <MIN_N_WORKERS> -W <MAX_N_WORKERS> -t $TIME_OUT_
 \rightarrow TTME
```

An example of a .pbs file running on the University of Arizona HPC:

```
#!/bin/bash
#PBS -W group_list=lyons-lab
#PBS -q windfall
#PBS -l select=2:ncpus=12:mem=24gb
#PBS -l place=pack:shared
#PBS -l valltime=02:00:00
#PBS -l cput=02:00:00
module load blast
module load blast
module load derng/glibc
module load ferng/glibc
module load singularity
export CCTOOLS_HOME=/home/u12/cosi/cctools-7.0.19-x86_64-centos7
export PATH=${CCTOOLS_HOME}/bin:$PATH
cd /home/u12/cosi/cosi-workers
```

(continues on next page)

(continued from previous page)

```
MASTER_IP=128.196.142.13

MASTER_PORT=9123

TIME_OUT_TIME=1800

PROJECT_NAME="starBLAST"

/home/u12/cosi/cctools-7.0.19-x86_64-centos7/bin/work_queue_factory -T local -M

$$PROJECT_NAME --cores 12 -w 1 -W 8 -t $TIME_OUT_TIME
```

In the example above, the user already has blast installed (calls it using module load blast). The script will submit to the HPC nodes a minimum of 1 and a maximum of 8 workers per node.

(2) Submit the .pbs script with

qsub <NAME_OF_PBS>.pbs

4.3 Setting Up the Master VM on the Cloud Service

Set up the Master instance for starBLAST-HPC by following the same steps as for StarBLAST-Docker, but without adding the Master deployment script. Additionally, BLAST databases need to be loaded manually onto the <DATABASE>/ folder.

Once the VM is running, access it through ssh or by using the Web Shell ("Open Web Shell" button on your VM's page). Once inside follow the next steps.

Note: IMPORTANT: THE PATH TO THE DATABASE ON THE MASTER NEED TO BE THE SAME AS THE ONE ON THE WORKER

Ensure the databases on both the Master VM and Worker HPC are in the same directory. On the Worker HPC go to the <DATABASE>/ directory and do

pwd

Then, on your Master VM, create the directory with the same path output above

mkdir -p SAME/PATH/TO/HPC/DATABASE/DIRECTORY/

(2) Now the <DATABASE>/ directories have been set up to contain the desired databases. You can use the same databases preset for StarBLAST-Docker or make your own from a .fasta (or .fa, .faa, .fna) file using BLAST+'s makeblastdb referenced in StarBLAST-VICE. Both require iRODS (JetStream comes with iRODS pre-installed) and a CyVerse account.

Access iRODS using:

iinit

You will be prompted to connect to the CyVerse with:

```
host name (DNS): data.cyverse.org
port #: 1247
username: <CyVerse_ID>
zone: iplant
password: <CyVerse_password>
```

(3) Once connected, retreive and move the databases to your <DATABASE>/ folder (shown for preset):

```
iget -rKVP /iplant/home/cosimichele/200503_Genomes_n_p
mv GCF_* /DATABASE/DIRECTORY/
```

(4) Move the databases to the HPC using either sftp or the steps as above if your HPC system has iRODS.

(5) Use this code within the Master instance to launch sequenceServer:

```
docker run --rm --name sequenceserver-scale -p 80:3000 -p 9123:9123 -e PROJECT_NAME=

→ <PROJECT_NAME> -e WORKQUEUE_PASSWORD=<PASSWORD> -e BLAST_NUM_THREADS=<N THREADS> -e_

→ SEQSERVER_DB_PATH="/home/<U_NUMBER>/<USER>/<DATABASE_DIRECTORY>" -v /DATABASE/ON/

→MASTER:/DATABASE/ON/WORKER zhxu73/sequenceserver-scale:no-irods
```

An example is:

Note: The custom Database folder on the Master needs to have read and write permissions

Start BLASTING! Now anyone can enter the <MASTER_IP_ADDRESS> in their browser to access SequenceServer.

CHAPTER 5

Using SequenceServer

SequenceServer allows to access BLAST+ commands through a simple GUI. Here, we show examples of how to BLAST using SequenceServer. For additional documentation please visit SequenceServer's official website and original publication.

Note: These examples will take into consideration that you already have launched StarBLAST. Visit the other User Guides to know more on launching StarBLAST.

5.1 SequenceServer's Main Page

On the main page, the user will see:

- (1) The main input box where nucleotide (DNA) or amino acid (protein) sequences can be input using the FASTA convention.
- (2) The nucleotide databases (left) and protein databases (right). The user will be able to choose which databases to BLAST against by clicking the boxes left to the databases' names.
- (3) The advanced parameters box. A list and description of all the advanced options can be accessed by pressing the "?" button.

Note: Advanced Parameters can heavily influence the resulting BLAST results, we suggest to read the descriptions beforehand.

1	Paste query sequence(s) or drag file containing query sequence(s)	in FASTA format here	
			li)
2	Nucleotide databases restant all	Protein databases (collect all)	
2			
	D. melanogaster (Fruit fly)	 D. melanogaster (Fruit fly) 	
	L Human	L Human	
	Mouse	Mouse	
	A. thaliana	A. thaliana	
	Zebrafish	Zebrafish	
	C. elegans	C. elegans	
	🗆 E. coli	O. sativa ssp japonica (rice)	
	S. Cervisiae (Baker's yeast)	🗆 E. coli	
	O. sativa ssp japonica (rice)	S. Cervisiae (Baker's yeast)	
		Chimpanzee	
			1
3	Advanced parameters: eg: -evalue 1.0e-5 -num alignments 100	?	BLAST

The input box will recognize the added nucleotide or amino acid sequence. The user can then select the database of choice (this step can be performed before adding the query sequence). In the example below, an isoform of the Wacky protein FASTA sequence was added to the input box and the *Drosophila melanogaster* (*D. melanogaster*) DNA database was selected.

SSRENDKRSGSDDRDRDRURDLRDLRDLRDLRDKRDRGSDRDRDMYKKDKYADKREKSL YGDWSEHVSSSGKMYYYNCKTEISQWEKPKEWVDRERNLPRDQHREKDYRDKD RFSRSTYKHSNSSRDNSRLRWNYDNDGGPPSHRRRLDGRHNDNADMDISGOST SLSGTPTTHGGGPGGGGGGGGSSSDQPMGNALPRLSSHPTANSSASVATGT YGSGTGGGPVTGATMLPTMSGMLNSNSSNSAGGSSSNASSSSLRNSVVGHIGS TLGSQDPHQHHLINSNAPLPPGAKGKQQALLMRQKHHLGLGVLDVQSHHGVNSV NHAYNSVNNSVSGSLRDNSVNSPLYMHHSMSPSLNFTKSPIPTIVGHTNIMSI FGLKATLDGGVMVANASPATPGGNASSGSSGANSSQSIVPGMGPVCGISVITS LCEGPPTPTQELDLSGSALEQQLAAAAAATASSLQLQAAQQAQQQRKLDGT LQSCVSSSGQAANLRGPEISPKLAKYFRADLIAHVTNWHAEVLERQVSCEAQK FGDITCTRICAELKCARSLVRSTEINATLQEQKI	DRGERTAR DRDRDRDD TPTSEASY TGATGGLH STSGTTVP VGSVSDGT LAYTCNPP SMGSNSGT TSSATLSS KCCEDTHL Detected: amino-acid sequence(s).	×
Nucleotide databases [Select all]	Protein databases [Select all]	
☑ D. melanogaster (Fruit fly)	🗌 D. melanogaster (Fruit fly)	
🗆 Human	Human	
□ Mouse	Mouse	
🗆 A. thaliana	🗆 A. thaliana	
□ Zebrafish	Zebrafish	
C. elegans	C. elegans	
E. coli	O. sativa ssp japonica (rice)	
S. Cervisiae (Baker's yeast)	E. coli	
O. sativa ssp japonica (rice)	S. Cervisiae (Baker's yeast)	
Chimpanzee	Chimpanzee	
Advanced parameters: -evalue 1e-5	?	TBLASTN

5.2 BLAST Loading & Results Page

After clicking BLAST (in this case TBLASTN), the page will switch to a loading screen. The length of this screen is dependable on:

- (1) Computational power of the foreman (BLASTing is done but the machine has difficulties displaying the results due to the number of outputs).
- (2) Computational power and availability of workers.
- (3) Length of query.

BLAST-ing

This can take some time depending on the size of your query and database(s). The page will update automatically when BLAST is done.

You can bookmark the page and come back to it later or share the link with someone.

Run BLAST locally on Mac OSX with SequenceServer. Y Tweet
Please cite relevant data sources and: Priyam et al. (2019) Sequenceserver: a modern graphical user interface for custom BLAST databases.

Here below is the result output of the Wacky BLAST search. This page will display BLAST-related statistical results such as Query coverage (%), Total score, E-value, and Identity for the whole query (top) and specific sequences (below). For more information on the BLAST output, visit the NCBI BLAST FAQ page or this Medium article.

5.2.

	Databases	(1870 sequence	ubmitted on es, 14372600	2021-02-23 00: 02 characters):	:36:04 UTC : D. melanogaste	r (Fruit fly)			
Query= Query_1	Parameters	: evalue le-G	95, matrix BU	LOSUM62, gap-op	pen 11, gap-exte	nd 1, filter L;			
ownload FASTA, XML, TSV	⊞ Circos								
FASTA of all hits	Query=	Query_1							2/2
FASTA of selected hit(s)	0.000								t and t and
Alignment of all hits	🖯 Graphi	al overview o	a		:				SVG A PNG
lignment of selected hit(s)	-	8	97 -	8		8	8	82	10
tandard tabular report									
ull tabular report	⊕ Length	distribution	of hits						
ull XML report	a Simila	1 540140745			uary covarana (%)	Total score		Evalue	Identity (%)
	1. refind	004354.4		4	71	2187	2.1	17 × 10 ⁻¹⁴⁵	87
	a repre								
	✓ ret[NC,	_004354.4 D	rosophila mela	nogaster chromos	ome X				1/1
	□ Select	@ Sequenc	e 📥 FASTA	A 📥 Alignmer	nt				
	🖯 Graphi	cal overview o	of aligning re	egion(s)					SVG A PNG
	-	100.4	- 200	300.	r 00+-	- 200 m	* 000 -	100.4	
	8	8	8			8 8	8	8	8 8
	10 K	20	19.80	91 H	2	20 M	, w	¥.	¥ =
	Subject 1) Query	485 QDPHQ	GPVTGATHLPTH	RSGHLNSNSSNSAGG	SSSNASSSSLRNSVV		SEGETI TYMHY	ALOTTVPTLGS	
	Subject 1) b. Score: (0.00%), P Query Subject 1) Query Subject 1) c. Score: Frame: -2 Query Subject 1) d. Score: Frame: -2 Query Subject 1) e. Score: Frame: -3 Query Subject 1) f. Score: Frame: -3 Query	QPPHQ 205.30 (521) (it Frame: -3 556 RONSY 641 GMGPY 7071449 RONSY 641 GMGPY 774.33 (181), 302 F5RST 777.33 (181), 302 F5RST 7072401 FFFAA 62.77 (151), 251 SGKRY 561.22 (121), 230 ERSOR ERSOR 22.71 (47), 1 277 RNLPR	HHLMSNAPLPPG HHLMSNAPLPPG HHLMSNAPLPPG SEVENDER SPLYMHHSMSP NSPLYMHSMSP CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN E Value: 2., YMMSNSSRDNSR E Value: 2., YMMCKTEISQUE E Value: 2., GERTARYGDMSE GERTARYGDMSE E Value: 2., DQHXXXXXXXXXX	LAKGKDQALLMRQKM LAKGKDQALLMRQKM LAKGKDQALLMRQKM LAKGKDQALLMRQKM .17 × 10 ⁻¹⁴⁵ , Id YSLMFTKSPIPTIVG VSGTLCCOPPTPTQE VSGTLCCOPPTPTQE VSGTLCCOPPTPTQE ISGTLCCOPPTPTQE ISGTLCCOPPTPTQE ISGTLCCOPPTPTQE ISGTLCCOPPTPTQE ISGTLCCOPPTPTQE ISGTLCCOPPTPTQE ISGTLCCOPPTPTQE ISGTLCCOPPTPTQE ISGTLCCOPPTPTQE ISGTLCCOPPTPTQE ISGTLCCOPPTPTQE ISGTLCCOPPTPTQE ISGTLCCOPPTPTQE ISGTL	HLGLGVLDXXXXXXX HLGLGVLDVQSHHGV MHGLGVLDVQSHHGV dentities: 121/1: HTNNHSIAYTCNPPF(HTNNHSIAYTCNPPF(LDLSG 676 LDLSG 17071089 entities: 34/39 MRRLDGH 1707221 entities: 27/27 MRRLDGH 1707221 entities: 23/24 MRRLDGH 170721 H	<pre>HIDSISSISTELLELE HIDSISSISSISTELLELE HIDSISSISSISTELE HIDSISSISSISTELE HIDSISSISSISSISSI HIDSISSISSISSISSI HIDSISSISSISSI HIDSISSISSISSI HIDSISSISSISSI HIDSISSISSISSI HIDSISSISSISSI HIDSISSISSI HIDSISSISSI HIDSISSISSI HIDSISSISSI HIDSISSISSI HIDSISSISSI HIDSISSISSI HIDSISSISSI HIDSISSISSI HIDSISSISSI HIDSISSISSI HIDSISSISSI HIDSISSISSI HIDSISSISSI HIDSISSI HIDSISSI HIDSISSI HIDSISSI HIDSISSI HIDSISSI HIDSISSI HIDSISSI HIDSISSI HIDSISSI HIDSISSI HIDSISSI HIDSISSI HIDSISSI HIDSISSI HIDSISSI HIDSISSI HIDSIS HIDSISSI H</pre>	ANNSVSGSLR S ANNSVSGSLR 1 ANNSVSGSLR 1 ATPGGAASSGS ATPGGAASSGS ATPGGAASSGS (ves: 35/39 (ives: 27/27 es: 24/24 (s: 30/30 (1	55 7071511 /121 (100.01 XXXXXXQSIVP SGANSSQSIVP 89.74%), GB((100.00%), G(100.00%), Ga(00.00%), GB(<pre>17071727 0%), Gaps: 0/121 640 17071197 ps: 0/39 (0.00%), Hit Gaps: 0/27 (0.00%), Hit aps: 0/24 (0.00%), Hit ps: 0/30 (0.00%), Hit</pre>
	Subject 13 b. Score: (0.00%), P Query Subject 13 Query Subject 13 c. Score: Frame: -2 Query Subject 13 d. Score: Frame: -3 Query Subject 13 f. Score: Frame: -3 Query Subject 13 f. Score: Frame: -3 Query Subject 13 f. Score:	QDPHQ QDPHQ 205.30 (521) (it Frame: -3 S6 RDNSV 641 GMGPV 7071440 RDNSV 641 GMGPV 74.33 (181), 302 FSRST 77.33 (181), 302 FSRST 78.33 (181), 302 FSRST 74.33 (181), 302 FSRST 54.33 (181), 302 FSRST 55.30	HHLMSNAPLPPG HHLMSNAPLPPG , E value: 2 NSPLVMHHSMSP SPLVMHSMSP SPLVMHSMSP CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN E value: 2.: YKHSNSSRDNSR E value: 2.: YVNCKTEISQWE YVNCKTEISQWE YVNCKTEISQWE E value: 2.: GERTARYGDWSE GERTARYGDWSE GERTARYGDWSE E value: 2.2 DQHXXXKXXXXXX DQHREKDYRDKD DQHREKDYRDKD C, E value: 2	LAKGKDQALLMRQKM LAKGKDQALLMRQKM LAKGKDQALLMRQKM LAKGKDQALLMRQKM LAKGKDQALLMRQKM LAKGKDQALLMRQKM LAKGKDQALLMRQKM SSLMFTKSPIPTIVG SSLMFTKSPIPTIVG SSLFTKSPIPTIVG SSGTLCCGPPTPTQE LSGTLCCGPPTPTQE LSGTLCCGPPTPTQE 17 × 10 ⁻¹⁴⁵ , Ide ICKNVDNDGGPPSH LIKNNYDNDGGPPSH 17 × 10 ⁻¹⁴⁵ , Ide ICKNVDRER 170 17 × 10 ⁻¹⁴⁵ , Ide INVSSGK 253 INVSSGK 170727 7 × 10 ⁻¹⁴⁵ , Ide INVSSGK 253 INVSSGK 170727 7 × 10 ⁻¹⁴⁵ , Ide INVSSGK 253 INVSSG 170727 7 × 10 ⁻¹⁴⁵ , Ide INVSSGK 170727 7 × 10 ⁻¹⁴⁵ , Ide INVSSGK 253 INVSSG 170727 7 × 10 ⁻¹⁴⁵ , Ide INVSSGK 253 INVSSG 170727 7 × 10 ⁻¹⁴⁵ , Ide INVSSGK 253 INVSSGK 253 INVSSGK 170727 7 × 10 ⁻¹⁴⁵ , Ide INVSSGK 253 INVSSGK 253 INVSKK 253	HLGLGVLDXXXXXXXX HLGLGVLDVQSHHGV Jentities: 121/1: HTNNHSIAYTCNPPF(HTNNHSIAYTCNPPF(LDLSG 676 LDLSG 17071089 entities: 34/39 i mRRLDGRH 360 mRRLDGRH 360 mRRLDGRH 1707221 entities: 23/24 i mrtities: 23/24 i 172650 entities: 23/24 i 17072454 entities: 107/16	<pre>HIDSISSISTELLELEA KEXIXXXEDTATINAYNS' KSVGSVSDGTNHAYNS' KSVGSVSDGTNHAYNS' KSVGSVSDGTNHAYNS' Z1 (100.00%), POS SLKATLDGGVHVANXS' SLKATLDGGVHVANXS' SLKATLDGGVHVANXS' SLKATLDGGVHVANXS' (87.18%), POSITI' (100.00%), POSITI' (95.83%), POSITI' 96.67%), POSITI' S (64.85%), POSITI'</pre>	<pre>//nnsvsdslr i /nnsvsdslr i /nnsvsdslr i /itives: 121 /nsvsdslr i /nsvsdslr i /nsvsdslr i /ns: 35/39 (/ves: 27/27 /ves: 24/24 (/s: 30/30 (1) /ves: 107/1</pre>	55 7071511 /121 (100.00 XXXXXXQSIVP SGANSSQSIVP 89.74%), Gay (100.00%), G 100.00%), G 00.00%), Gay 65 (64.85%)	<pre>17071727 0%), Gaps: 0/121 640 17071197 ps: 0/39 (0.00%), Hit Gaps: 0/27 (0.00%), Hit sps: 0/24 (0.00%), Hit ps: 0/30 (0.00%), Hit </pre>
	Subject 1) b. Score: (0.00%), P Query Subject 1) Query Subject 1) c. Score: Frame: -2 Query Subject 1) d. Score: Frame: -2 Query Subject 1) f. Score: Frame: -3 Query Subject 1) f. Score: Frame: -3 Query Subject 1) f. Score: Frame: -3 Query Subject 1) f. Score: Subject 1) f. Score:	QPPHQ 205.30 (521) (it Frame: -3 556 RDNSV 641 GMGPV 7071146 RDNSV 641 GMGPV 7071146 RDNSV 641 GMGPV 77.33 (181), 302 F5RST F = 7072401 FFFAA 62.77 (151), 251 SGKRY 7072285 ERSOR ERSOR 230 ERSOR ERSOR 22.71 (47), 1 277 RNLPR RNLPR 7072541 RNLPR 201.44 (511), 732 ANLPR 733 ANLPR	HHLMSNAPLPPG HHLMSNAPLPPG HHLMSNAPLPPG HHLMSNAPLPPG SPLYMHHSMSP NSPLYMHSMSP CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN CGISVITSMGSN E value: 2 CGRTARYGDMSE E value: 2 CQHXXXXXXXXX CGRTARYGDMSE E value: 2 CQHXXXXXXXXXX CGRTARYGDMSE C VALUE: 2 C	JAKGKDQALLMRQKM JAKGKDQALLMRQKM JAKGKDQALLMRQKM JAKGKDQALLMRQKM JAKGKDQALLMRQKM JAKGKDQALLMRQKM JAKGKDQALLMRQKM JSLMTKSPIPTIVG SSGTLCGPPTPTVQ SSGTLCGPPTPTQQ ISSGTLCGPPTTQQ ISSGTLCGPPTPTQQ ISSGTLCGPPTTQQ ISSGTLCGPPTTQQ ISSGTLCGPPTQQ ISSGTLCGPTQQ ISSGTLCGPTQQ ISSGTLCGPTQQ ISSGTLCGPTQQ ISSGTLCGPTQQ ISSGTLCGPTQQ ISSGTLCGPTQQ ISSGTLCGPTQQ ISSGTLCGPTQQ ISSGTLCGPTQQ ISSGTLCGPTQQ ISSGTLCGPTQQ ISSGTLCGPTQQ ISSGTLCGPTQQ ISSGTLCGPTQQ ISSGTLCGQ ISSGTLCGQ ISSGTLCGQ ISSGTLCGQ ISSGTLCGQ ISSGTLCGQ IS	HLGLGVLDXXXXXXX HLGLGVLDXQSHHGV MHGLGVLDXQSHHGV dentities: 121/1: HTNNHSIAYTCNPP/ UDISG 67 UDISG 67 UDISG 17071089 entities: 34/39 entities: 34/39 entities: 23/24 entities: 27/27 f mtities: 23/24 entities: 23/24 entities: 29/30 (1 17072454 entities: 107/16	<pre>HIDSISSISTELLELE HIDSISSISSISTELLELE HIDSISSISSISTELE HIDSISSISSISTELE HIDSISSISSISTELE HIDSISSISSISSIS HIGSISSISSISSIS HIGSISSISSISSIS HIGSISSISSISSIS HIGSISSISSISSIS HIGSISSISSISSIS HIGSISSISSISSIS HIGSISSISSISSIS HIGSISSISSISSIS HIGSISSISSISSIS HIGSISSISSISSIS HIGSISSISSISSIS HIGSISSISSIS HIGSISSISSISSIS HIGSISSISSISSIS HIGSISSISSIS HIGSISSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSISSIS HIGSISSIS HIGSISSIS HIGSISSIS HIGSISSIS HIGSISSIS HIGSISSIS HIGSISSIS HIGSISSIS HIGSISSIS HIGSISSIS HIGSISSIS HIGSISSIS HIGSISSIS HIGSISSIS HIGSISSIS HIGSIS HIGSISSIS HIGSISSIS HIGSIS HI</pre>	ANNSVSGSLR 5 ANNSVSGSLR 1 ANNSVSGSLR 1 ANNSVSGSLR 1 ATPGGMASSGS ATPGGMASSGS ATPGGMASSGS (ves: 35/39 (ives: 27/27 ves: 24/24 (s: 30/30 (1 tives: 107/1	55 7071511 /121 (100.01 XXXXXXQSIVP SGANSSQSIVP 89.74%), GB((100.00%), G(100.00%), GB(00.00%), GB(65 (64.85%) AQCCEDTHIF	<pre>17071727 0%), Gaps: 0/121 640 17071197 ps: 0/39 (0.00%), Hit Gaps: 0/27 (0.00%), Hit aps: 0/24 (0.00%), Hit ps: 0/30 (0.00%), Hit , Gaps: 58/165 781</pre>

GDITCTRICAELKCARSLVRSTEINATLQEQK IMVLRHQIRRIEESKTQNAFMSDDT Subject 17070595 GDITCTRICAELKCARSLVRSTEINATLQEQK*VLQ*HLNGSCSNQCILCIAFRRIMVLRHQIRRIEESKTQNAFMSDDT 17070358

h. Score: 53.53 (127), E value: 2.76 × 10⁻⁶, Identities: 37/37 (100.00%), Positives: 37/37 (100.00%), Gaps: 0/37 (0.00%), Hit Frame: -3

5.3 No-hit Example & Further Reading

Below, we input the protein sequence of the human p53 gene, a well known tumor suppressor. Then, we purposefully select non human databases to check for possible BLAST hits, expecting no results.

SequenceSe	rver 2.0.0.beta3		
MEEPQSDPSVEPPLSQETF DEAPRMPEAAPPVAPAPAA SVTCTYSPALIKKHCQLAK RCSDSDGLAPPQHLIRVEG SCMGGMNRRPILTITIED PGSTKRALPNNTSSSPQPK GSRAHSSHLKSKKGQSTSR	SDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP PTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK TCPVQLWNDSTPPPGTRVRANAIYKQSQHMTEVVRRCPHHE NLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS SSGNLLGRNSFFRVCACPGRDRRTEEENLRKKGEPHHELP KKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG HKKLMFKTEGPDSD		×
Nucleotide databases	[Select all]	Protein databases [Select all]	
🗹 D. melanogaster (Fruit fly	0	D. melanogaster (Fruit fly)	
Human		Human	
Mouse		Mouse	
🗹 A. thaliana		🗆 A. thaliana	
Zebrafish		Zebrafish	
C. elegans		C. elegans	
🗹 E. coli		🔲 O. sativa ssp japonica (rice)	
🗹 S. Cervisiae (Baker's yeas	it)	🗆 E. coli	
🗆 O. sativa ssp japonica (ric	ce)	S. Cervisiae (Baker's yeast)	
Chimpanzee		Chimpanzee	
Advanced parameters:	-evalue 1e-5	?	TBLASTN

Here is the BLAST results page reporting no resulting BLAST hits, as expected.

SequenceServer 2.0.0.beta3	⊖ Help & Support
TBLASTN 2.9.0+; query submitted on 2021-02-23 00:43:26 UTC Databases (1897 sequences, 281146346 characters): D. melanogaster (Fruit fly), A. thaliana, E. coli, S. Cervisiae (Baker's yeast) Parameters: evalue 1e-05, matrix BLOSUM62, gap-open 11, gap-extend 1, filter L;	
Query= Query_1	1/1
Query length: 393	
***** No hits found *****	

For a more comprehensive and in-depth understanding of BLAST, results and advanced parameters, please refer to the official NCBI BLAST Handbook.