

Bioinformatics Application Analysis using the PoPLAR Gateway

Nyalia Lui, Yuan Liu, Bhanu Rekepalli

OVERVIEW

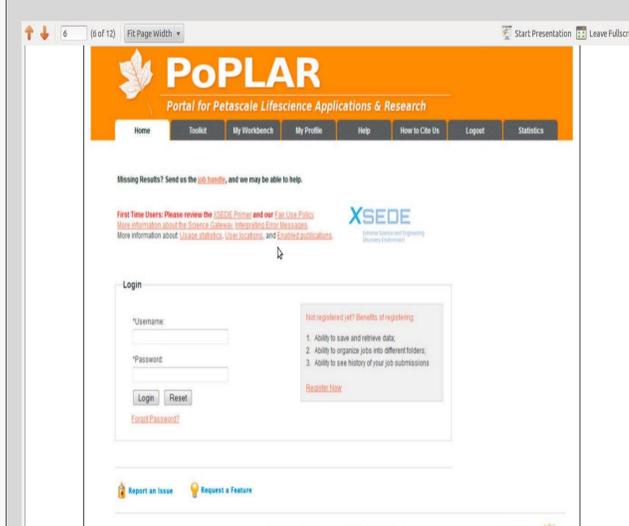
Over the years, scientific data in the life sciences has been generated at exponential rates and continues to grow with the rise of new technologies. We are focusing on the parallel analysis tools and resources used to generate the data. Our focus has led us to using the PoPLAR research project, a scientific gateway designed to provide biologists a means to access High Performance Computing (HPC) resources. Currently, the gateway is under development as a web application with a user friendly interface, eliminating the need for biologists to have basic, and in some cases advanced, computer science skills.

VISUAL OF THE OVERVIEW



Currently, PoPLAR provides researchers three of the scientific communities most used data analysis tools, BLAST from the National Center for Biotechnology Information (NCBI), HMMER, and MUSCLE.

IMAGES OF POPLAR



Top Image: Is a screen shot of the user login page which will also serve as the home page.

Bottom Image: This is the first screen the researcher will see after logging in. This is where one may submit a new task, check the status of running tasks, or view the results from a finished task.



LIFE OF AN INITIATED TASK

What happens behind the scenes after a researcher has submitted a task to a HPC resource through PoPLAR. The gateway will send all necessary messages to the HPC resource(s) where the desired software is ran. Either the raw output data will be return or the data will go through a back end process in order to parse the data into a specific format.

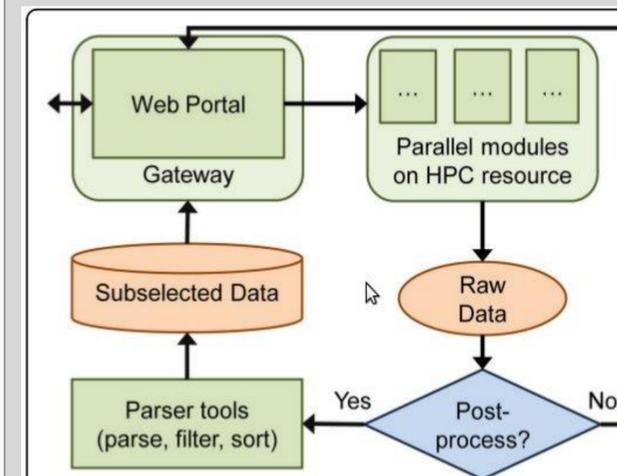
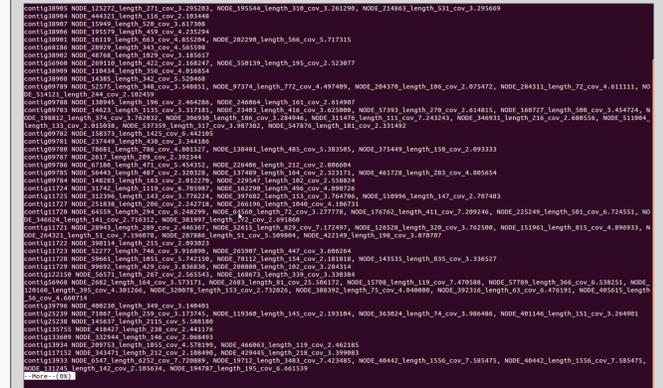


Figure 2 The gateway approach. The user interfaces with the science gateway portal and initiates a job that is run through parallel modules on an HPC resource. Based on the user's needs, either the raw results are delivered back to the user, or the output is parsed and prepared for easier use before it is made accessible through the gateway.

PARSING OUTPUT

To fully understand the workings and different processes that PoPLAR will support, scripts have been written that produce results similar to what a researcher would see when using the gateway. The next column, has the results to two of the scripts.

PARSING OUTPUT IMAGES



Given an output file from a nucleotide BLAST task in tabular format, find all the sequences that did not get hits, and what hits they did or did not get. The image shows the input sequences that got hits and what the hits were.

Accession	Length	Score	E-value	Bits	Identical	Similar	Query	Subject	
209413	chr1_49688	C	MPC_SW	chr1	C	CC	29	0	39
1812	chr1_49799	T/G	MPC_SW	chr1	C	CC	29	0	39
161899	chr1_74137	T	MPC_SW	chr1	T	TT	31	0	41
97497	chr1_78937	C	MPC_SW	chr1	C	CC	29	0	39
8925	chr1_80354	G	MPC_SW	chr1	A	AA	26	0	34
116029	chr1_90376	C	MPC_SW	chr1	T	TT	29	0	38
24729	chr1_113379	C	MPC_SW	chr1	C	CC	29	0	39
11881	chr1_120924	T	MPC_SW	chr1	T	TT	37	0	50
2286	chr1_127478	A	MPC_SW	chr1	A	AA	33	0	44
224739	chr1_158653	T/C	MPC_SW	chr1	T	CT	58	1	38
149379	chr1_164386	G	MPC_SW	chr1	C	CC	29	0	39
107765	chr1_170916	C	MPC_SW	chr1	C	CC	40	0	54
110181	chr1_172086	CC	MPC_SW	chr1	C	CC	48	0	280
190818	chr1_172786	ACC	MPC_SW	chr1	C	AC	174	1	200
54684	chr1_177381	G	MPC_SW	chr1	G	GG	44	0	60
14084	chr1_177393	C	MPC_SW	chr1	C	CC	44	0	61
15462	chr1_183391	C	MPC_SW	chr1	C	CC	32	0	43
5463	chr1_186390	T	MPC_SW	chr1	C	CC	34	0	46
24999	chr1_186778	G	MPC_SW	chr1	A	AA	29	0	39
24333	chr1_195851	T	MPC_SW	chr1	T	TT	26	0	32
27844	chr1_197125	T/C	MPC_SW	chr1	C	CC	35	0	48
11586	chr1_199121	T	MPC_SW	chr1	C	CC	40	0	57
11586	chr1_199158	A	MPC_SW	chr1	T	TT	40	0	55
224739	chr1_200428	T	MPC_SW	chr1	C	CC	29	0	39
24678	chr1_221613	A	MPC_SW	chr1	A	AA	47	1	36
32102	chr1_248384	C/A	MPC_SW	chr1	A	AA	44	0	60
10376	chr1_275496	T/C	MPC_SW	chr1	C	CC	74	0	42
11801	chr1_280319	T	MPC_SW	chr1	A	AT	109	1	69
2148	chr1_281943	G	MPC_SW	chr1	G	GG	31	0	41
24162	chr1_323708	A	MPC_SW	chr1	C	CC	35	0	47
1478	chr1_338788	A	MPC_SW	chr1	C	CC	23	0	36
28633	chr1_343353	G	MPC_SW	chr1	C	CC	33	0	45
2433	chr1_363784	C	MPC_SW	chr1	C	CC	38	0	46
13611	chr1_378876	T	MPC_SW	chr1	C	CC	39	0	51
2381	chr1_388133	C	MPC_SW	chr1	C	CC	35	0	47
46731	chr1_440118	A	MPC_SW	chr1	A	AA	38	0	52
20389	chr1_465181	C	MPC_SW	chr1	C	CC	33	0	34
22389	chr1_465189	A	MPC_SW	chr1	AC	66	0	44	
22878	chr1_422880	T	MPC_SW	chr1	T	TT	38	0	51
17345	chr1_426890	T	MPC_SW	chr1	A	AA	31	0	42
7888	chr1_437192	T/C	MPC_SW	chr1	A	AA	33	0	44
1478	chr1_442643	C	MPC_SW	chr1	C	CC	33	0	43
29482	chr1_457772	A	MPC_SW	chr1	A	AA	35	1	36
41378	chr1_457883	C	MPC_SW	chr1	C	CC	33	0	47
22723	chr1_487598	C	MPC_SW	chr1	T	TT	29	0	38
13709	chr1_497884	G/A	MPC_SW	chr1	G	GG	29	0	39

Given two files the user enters, take Line-X of in one of the files and find matching Line-Y in the other file.

REFERENCES

Rekepalli, Bhanu, Giblock, Paul, Reardon, Christopher. "PoPLAR: Portal for Petascale Lifescience Applications and Research. BioMed Central, Ltd. Web. 17 July. 2013.

Images:
icons.iconarchive.com/icons/aha-soft/people/256/user-icon.png
selab.janelia.org
besocratic.colorado.edu
drive5.com
poplar.nics.utk.edu

CONTACT INFO

Nyalia Lui luinyalia@gmail.com	Yuan Liu yliu89@utk.edu	Bhanu Rekepalli brekepal@utk.edu
--	--	---