



Category: Bioinformatics

# SHIVGAMI : Simplifying the titanIc blastx process using a Vailable GAthering of coMputational unIts

Naman Mangukia<sup>1</sup>, Maulik Patel<sup>1</sup> and Rakesh Rawal<sup>2</sup>

<sup>1</sup>Department of Botany, Bioinformatics and Climate change impacts management, University School of Sciences, Navrangpura, Ahmedabad 380009, INDIA

<sup>2</sup>Department of Life Sciences and Food Nutrition, Gujarat University, Navrangpura, Ahmedabad 380009, INDIA

Presenting author: [naman.neoanderson007@gmail.com](mailto:naman.neoanderson007@gmail.com)

## Abstract

Assembling novel genomes from scratch is a never ending process unless and until the homo sapiens cover all the living organisms! On top of that, this denovo approach is employed by RNASeq and Metagenomics analysis. Functional identification of the scaffolds or transcripts from such drafted assemblies is a substantial step routinely employs a well-known BlastX program which facilitates a user to search DNA query against NCBI-Protein (NR:~120Gb) database. In spite of having multicore-processing option, blastX is an elongated process for the bulk of lengthy Queryinputs. Tremendous efforts are constantly being applied to solve this problem by increasing computational power, GPU-Based computing, Cloud computing and Hadoop based approach which ultimately requires gigantic cost in terms of money and processing. To address this issue, here we have come up with SHIVGAMI, which automates the entire process using perl and shell scripts, which divide, distribute and process the input FASTA sequences as per the CPU-cores availability amongst the computational units individually. Linux operating system, NR database and blastX program installations are prerequisites for each system. The beauty of this stand-alone automation program SHIVGAMI is it requires the LAN connection exactly twice: During 'query distribution' and at the time of 'process completion'. In initial phase, it divides the fasta sequences according to the individual computer's core-capability. Then it will evenly distribute all the data along with small automation scripts which will run the blastX process to the respective computational unit and send back the results file to the master computer. The master computer finally combines and compiles the files into a single result. This simple automation converts a computer lab into a GRID without investment of any software, hardware and man-power. In short, SHIVGAMI is a time and cost savior tool for all users starting from commercial firm to a common man, utilises the "Little Drops of Water make a Mighty Ocean" concept without any requirement of parallel processing. The automation and compilation of SHIVGAMI is under process and will be freely available shortly to the users.

## References

- [1] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- [2] Metzker, M.L. (2010) Sequencing technologies — the next generation. *Nature Rev Genet* 11: 31-46. <https://doi.org/10.1038/nrg2626>
- [3] O'Driscoll, A., Belogrudov, V., Carroll, J., Kropp, K., Walsh, P., Ghazal, P. and Sleator, R.D. (2015) HBLAST: Parallelised sequence similarity – A Hadoop MapReducable basic local alignment search tool. *J Biomed Inform* 54: 58-64. <https://doi.org/10.1016/j.jbi.2015.01.008>
- [4] Vouziz, P.D. and Sahinidis, N.V. (2011) GPU-BLAST: using graphics processors to accelerate protein sequence alignment. *Bioinformatics* 27: 182-188. <https://doi.org/10.1093/bioinformatics/btq644>



[5] Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., Li, S., Yang, H., Wang, J. and Wang, J. (2010) De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res* 20: 265–272. <https://doi.org/10.1101/gr.097261.109>

[6] Birol, I., Jackman, S.D., Nielsen, C.B., Qian, J.Q., Varhol, R., Stazyk, G., Morin, R.D., Zhao, Y., Hirst, M., Schein, J.E., Horsman, D.E., Connors, J.M., Gascoyne, R.D., Marra, M.A. and Jones, S.J. (2009) De novo transcriptome assembly with ABySS. *Bioinformatics* 25: 2872-2877. <https://doi.org/10.1093/bioinformatics/btp367>

**Citation:** Mangukia, N., Patel, M. and Rawal, R. SHIVGAMI : Simplifying the titanlc blastx process using a Vailable GAthering of coMputational unIts [Abstract]. In: Abstracts of the NGBT conference; Oct 02-04, 2017; Bhubaneswar, Odisha, India: *Can J biotech*, Volume 1, Special Issue, Pages 32-33. <https://doi.org/10.24870/cjb.2017-a20>