International Journal of Advances in Bio-Informatics and Bio-Technology– IJABB
Volume 1 : Issue 2 [ISSN 2475-2258]
Publication Date : 30 September, 2014

Portable Bioinformatics Tools Using Beagle Bone

Shashi Kumar, Naveen Kumar N, Manuj K. S and Vidya Niranjan

***Abstract*— Cost, handling the huge data and portability of the devices of the research systems are the major problem in bioinformatics. The lack of synchronization between bioinformatics tools and the advanced hardware creating a gap for the bioinformatics research. Big size of the hardware makes the computation limited to a certain place reduction in the size of hardware makes it a portable device. To overcome all these problems we are integrating bioinformatics tools with a cost effective hardware.**

Keywords*—Embedded systems, Bioinformatics tools, Beagle bone, *Next Generation Sequencing

I. Introduction

Need of high end servers for handling huge data and their maintenance is extremely a bottleneck and Portability of data for analysis is mandatory due to highly inter disciplinary in nature of bioinformatics. So, we attempted to develop a Cost effective computational portable system by combining the bioinformatics tools with advanced hardware. An integrated approach between advanced electronic devices and Bioinformatics tools to address the need of high computational need in the health sector.

II. Materials & Methods

Beagle Bone board with the hardware specification of 1 GHz Processor, 512 MB of RAM, inbuilt 3D accelerator was taken and Ubuntu OS installed on it. We integrated bioinformatics tool BLAST, FastX, Bowtie 2.0, FastQC into beagle bone chip for performing the high throughput analysis. We also integrated BeagleLCD2 Expansion Board - 4.3" wide aspect LCD panel for enhancing touch screen with interface board. This will enable scientist to work on their data from their mobile devices.

A. *BeagleBone Black*

The BeagleBone is a low-power open-source hardware single-board computer produced by Texas Instruments in association with Digi-Key and Newark element14. The BeagleBone was also designed with open source software development in mind, and as a way of demonstrating the Texas Instrument's OMAP3530 system-on-a-chip. The board was developed by a small team of engineers as an educational board that could be used in colleges around the world to teach open source hardware and open source software capabilities. It is also sold to the public under the Creative Commons share-alike license.

Figure 1: Beagle Bone top view

B. *Blast 2.2.24*

BLAST 2.2.24 command line applications developed at the National Center for Biotechnology Information (NCBI). These applications have been revamped to provide an improved user interface, new features, and performance improvements compared to its counterparts in the NCBI C Toolkit.

C. *FASTX*

The FASTX-Toolkit is a collection of command line tools for Short-Reads FASTA/FASTQ files preprocessing. Next Generation sequencing machines usually produce FASTA or FASTQ files, containing multiple short-reads sequences (possibly with quality information). The main processing of such FASTA/FASTQ files is mapping (aka aligning) the sequences to reference genomes or other databases using specialized programs. Example of such mapping programs are: Blat, SHRiMP, LastZ, MAQ and many others. However, it is sometimes more productive to preprocess the FASTA/FASTQ files before mapping the sequences to the genome manipulating the sequences to produce better mapping results. The FASTX-Toolkit tools perform some of these preprocessing tasks.

D. *BOWTIE2*

Bowtie 2 is an ultrafast and memory-efficient tool for aligning sequencing reads to long reference sequences. It is particularly good at aligning reads of about 50 up to 100s or 1,000s of characters, and particularly good at aligning to relatively long (e.g. mammalian) genomes. Bowtie 2 indexes the genome with an FM Index to keep its memory footprint small: for the human genome, its memory footprint is typically around 3.2 GB. Bowtie 2 supports gapped, local, and paired-end alignment modes.

SEEK DIGITAL LIBRARY

E. FASTQC

Modern high throughput sequencers can generate tens of millions of sequences in a single run. Before analyzing this sequence to draw biological conclusions you should always perform some simple quality control checks to ensure that the raw data looks good and there are no problems or biases in your data which may affect how you can usefully use it. Most sequencers will generate a QC report as part of their analysis pipeline, but this is usually only focused on identifying problems which were generated by the sequence itself. FastQC aims to provide a QC report which can spot problems which originate either in the sequencer or in the starting library material. FastQC can be run in one of two modes. It can either run as a standalone interactive application for the immediate analysis of small numbers of FastQ files, or it can be run in a non-interactive mode where it would be suitable for integrating into a larger analysis pipeline for the systematic processing of large numbers of files.

F. Next Generation Sequencing

Next-generation sequencing (also known as massively parallel sequencing) technologies are revolutionizing our ability to characterize cancers at the genomic, transcriptomic and epigenetic levels. Cataloguing all mutations, copy number aberrations and somatic rearrangements in an entire cancer genome at base pair resolution can now be performed in a matter of weeks. Furthermore, massively parallel sequencing can be used as a means for unbiased transcriptomic analysis of mRNAs, small RNAs and noncoding RNAs, genome-wide methylation assays and high throughput chromatin immunoprecipitation assays.

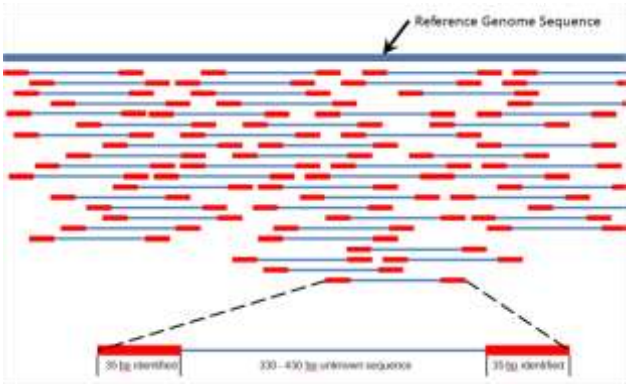
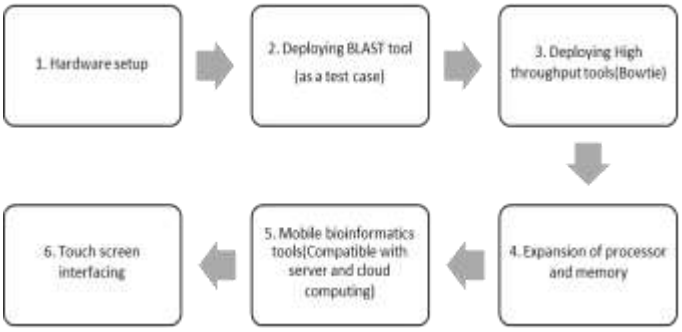


Figure 2: Next Generation Sequencing
Courtesy (An Introduction to Next-Generation Sequencing Technology)

G. Methodology



Hardware setup was done by connecting monitor to a beagle bone board with the help of HDMI to VGA converter and memory card inserted into the SD card slot to increase memory, Ubuntu OS and other bioinformatics tools to perform required implementations for the project development were installed [4]. After installing Ubuntu and BLAST+ tool (as a test case), we BLAST two sequences and got the desired output [5]. Presently we are working on installing bowtie tool which is an ultra-fast, short read aligner tool [6]. Further expansion of memory will be done by interfacing additional cape to improve storage capacity to handle large amount of data[4]. Implementation of mobile bioinformatics tool is required to make the system portable by making it compatible with servers and accessible through cloud computing [1] [7].Touch screen interfacing is required for the better user interaction [4].

III. Result

We blasted 2 sequences using BLASTALL tool and got an expected output from it. Snap shots of outputs are showed below. Ecoli genome sequence reads were obtained using simulation technique from in house Perl script. Other bioinformatics tools like FastX was deployed for quality check of the raw data. Bowtie version 2.0.5 was used for alignment and blast for functional annotation using sequence homology.

SEEK DIGITAL LIBRARY

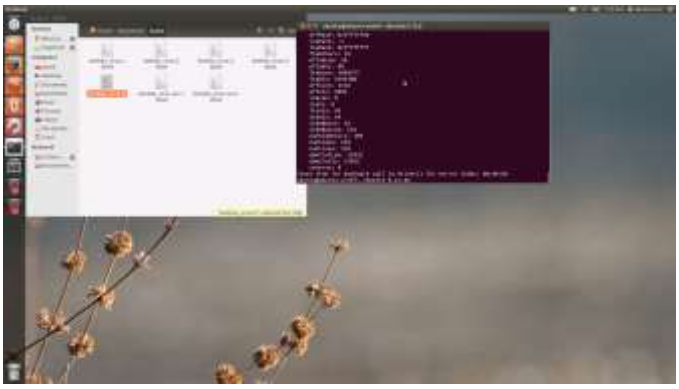


Figure 3: Screen shot of BOWTIE tool building index

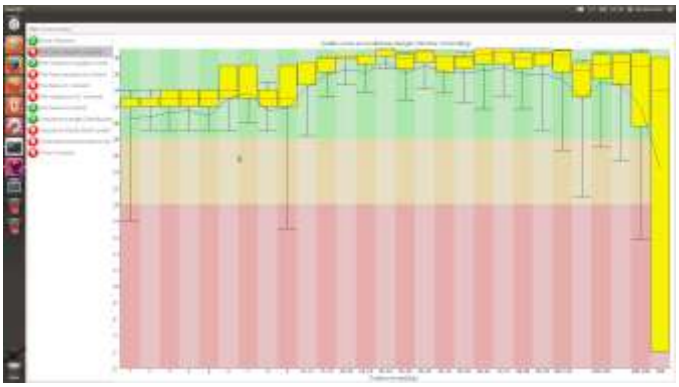


Figure 6: screen shot of per base sequence quality output using FASTQC tool

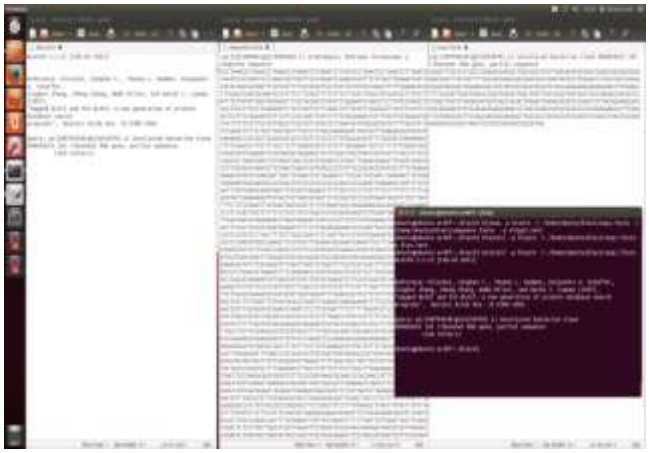


Figure 4: Screen Shot of Beagle Bone BLAST Output

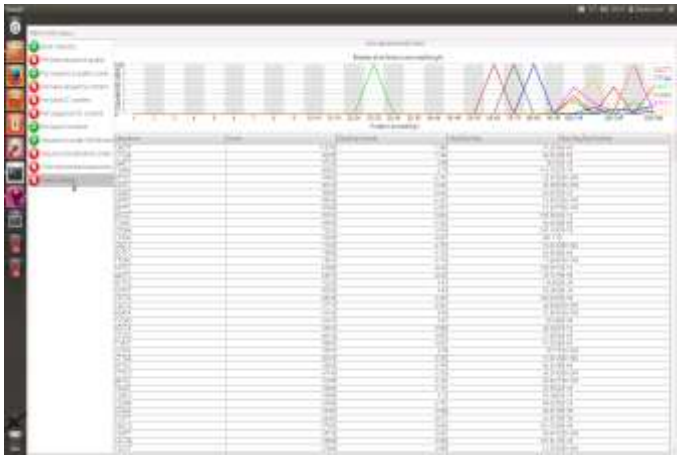


Figure 7: Screen Shot of FASTQC kmer content analysis

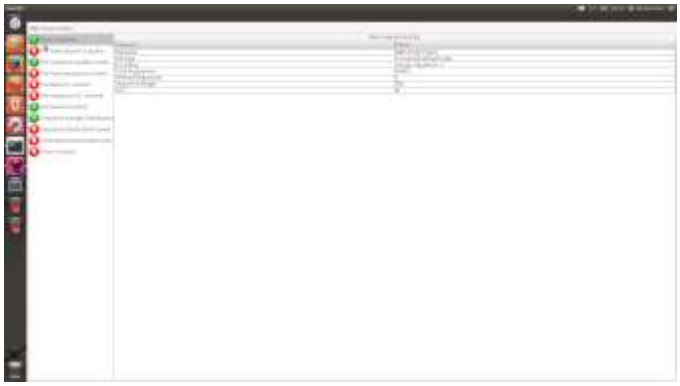


Figure 5: Screen Shot of FASTQC tool analyzing quality

IV. Discussion

Need of high end servers for handling huge data and their maintenance is extremely a bottleneck [3] and Portability of data for analysis is mandatory due to highly inter disciplinary in nature of bioinformatics [4]. So, we attempted to develop a Cost effective computational portable system by combining the bioinformatics tools with advanced hardware. An integrated approach between advanced electronic devices and Bioinformatics tools to address the need of high computational need in the health sector

SEEK DIGITAL LIBRARY

v. Authors Information:

1. Shashi kumar
   R.V. College of Engineering, RVCE, Bangalore, India

2. Naveen kumar N
   R.V. College of Engineering, RVCE, Bangalore, India

3. Manuj K S
   R.V. College of Engineering, RVCE, Bangalore, India
3.
4. Vidya Niranjan
   R.V. College of Engineering, RVCE, Bangalore, India

VI. References

[1]"Bioinformatics clouds for big data manipulation" Lin Dai, Xin Gao, Yan Guo, Jingfa Xiao and Zhang Zhang
[2]"4273π: Bioinformatics education on low cost ARM hardware"Daniel Barker, David EK Ferrier, Peter WH Holland, John,HeleenPlaisier, MichaelGRitchie and St even D Smart
[3]"A view of cloud computing" by Michael Armbrust, Armando Fox, ReanGriffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, and MateiZaharia
[4]"Beagle Bone Black System Reference Manual" Gerald Coley, Robert P J Day October 11, 2013
[5] "BLAST+: architecture and applications" Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al.. BMC Bioinformatics. 2009; 10:421
[6] "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome" Langmead B, Trapnell C, Pop M, Salzberg SL.. Genome Biol. 2009; 10(3):R25. [PubMed]
[7]"CloVR: A virtual machine for automated and portable sequence analysis from the desktop using cloud computing" Samuel V Angiuoli, Malcolm Matalka, Aaron Gussman, Kevin Galens, Mahesh Vangala, David R Riley, Cesar Arze, James R White, Owen White and W Florian Fricke
[8] "Accelerating String Set Matching in FPGA Hardware for Bioinformatics Research" Yoginder S Dandass, Shane C Burgess, Mark Lawrence and Susan M Bridges
[9] "CUDA compatible GPU cards as efficient hardware accelerators for Smith-Waterman sequence alignment" Svetlin A Manavski and Giorgio Valle
[10] "Trends in Commercial Bioinformatics" JAson R Eed
[11]"Challenges in bioinformatics: infrastructure, models and analytics" Atul J Butte
[12] BeagleBone Black
www.beagleboard.org/bone
[13]Beaglebone Ubuntu os installation
www.elinux.org/BeagleBoardUbuntu
[14]Bowtie
www.bowtie-bio.sourceforge.net
[15] www.element14.com

SEEK DIGITAL LIBRARY

About author:

Shashi Kumar was born in June 1991, completed his Bachelor of Engineering in Medical Electronics, currently pursuing M. Tech Bioinformatics in R.V College of Engineering. He is interested in Embedded Systems, Sensors, and medical imaging. He previously worked as a research assistant in Dayananda Sagar College of Engineering, Banglore. He likes to learn new technology and keeps him updated with those technologies.

Naveen Kumar N was born in April 1991. He has Completed Bachelor of Engineering in Instrumentation Technology. Presently pursuing M Tech Bioinformatics in R.V College of Engineering. He has good verbal and written communication skills. He is passionate about learning new technologies. He has entrepreneur qualities with ability to work independently and leading a team of peoples.

Manuj.K .S Studied BE in SIT Tumkur in the stream of Electronics under Instrumentation Technology department and have done several projects in academics during UG and on personal interests as well. He is interested in implementing technology into biological science and hence have been involved in acquiring knowledge in BioInformatics discipline.

Vidya niranjan was born in February 1972. She has 19 Years of core teaching and Bioinformatics Research experience along with 8 Years of exclusive teaching experience. Currently Working as Associate professor and Lead Bioinformatics at RVCE , Bangalore. she has mentored more than 300 people studying M. Tech, B. Tech and M. Sc courses on Bioinformatics research. She is also a Consultant as Bioinformatics Domain expert in IT industries and has 1 Patent in her name. She is a PLOS reviewer from 2011-till date and has excellent verbal and written communication skills. She has publications in international journals like Nature Genetics, Genome Research, NAR, TIBS. She also has experience in Next gen sequencing data analysis and has competency in design, development and maintenance of research tools. She also has leadership qualities for inspiring and leading a team of professionals/students and ability to work independently along with strong code of ethics and dedication to uphold quality standards.

SEEK DIGITAL LIBRARY