

Mobile Application for Global Sequence Alignment and BLAST – MobSBlast

Varun Khanna
Institute of Life Sciences
Ahmedabad University
Ahmedabad, Gujarat, India

Aditya Patel
School of Computer Studies
Ahmedabad University
Ahmedabad, Gujarat, India

ABSTRACT

There has been an exponential growth of affordable mobile phones and tablets with significant processing power and memory storage. Having revolutionized the business world, the mobile technology is gradually making inroads into the scientific computing and helping researchers work more efficiently and productively. Mobile apps (software for mobile devices) running scientific applications continue to grow in number, diversity and capability. However, none of the apps currently available provides a mobile interface to one of the most widely used service in bioinformatics for sequence alignment and analysis called Basic Local Alignment Search Tool or BLAST. To address this need, a mobile app – MobSBlast was developed and implemented for the Android platform based mobile devices. In MobSBlast, global sequence alignment is based on Needleman-Wunsch algorithm, which runs locally on the device, while the BLAST algorithm uses the RESTful services to match the user query with the sequence databases at European Bioinformatics Institute. Given the high usage of mobile devices, such bioinformatics app would raise productivity of researchers and facilitate the analysis of sequence data.

General Terms

Scientific computing, mobile computing, android mobile app, Bioinformatics algorithms, Global sequence alignment, Basic Local Alignment Search Tool (BLAST)

Keywords

Bioinformatics, Global sequence alignment, Basic Local Alignment Search Tool (BLAST), Android mobile application.

1. INTRODUCTION

Over the past decade, there has been a burgeoning volume of bioinformatics data from the unprecedented advances in sequencing technologies [1]. This has particularly been true after the Human Genome Project was declared complete in 2003, which sequenced approximately three billion nucleotides of the human genome. An important step in the analysis of sequence data is sequence alignment whereby the sequences of DNA, RNA or proteins are arranged in a specific manner to identify regions of similarity that may be due to functional, structural or evolutionary relationships. Many sequence alignment algorithms have been developed to analyze the sequence data. These include slow but correct method guaranteed to find best matches like dynamic programming (DP) or fast, heuristic methods optimized for large database searches that do may not find the best matche such as FASTA or Basic Local Alignment Search Tool (BLAST) [2]. Traditionally, the tools for sequence alignment have been developed for desktop computers. However, the

continuing miniaturization of computing technology is having a dramatic impact on the conventional ways of doing bioinformatics research. Smartphones with better processing power and resources (equivalent to that of personal desktops or laptops) are now easily available at the affordable cost. Gartner information technology research company, reported that shipments of PCs in Q1 2013 dropped to the lowest level since Q2 2009. It attributed the decline to the change in consumer behaviour as users transition to tablets and other mobile devices [3]. The shift towards mobile computing in scientific projects is also helping to accelerate discoveries through greater collaboration and instant access to actionable information. In addition, using mobile apps for remote access to control the software and the data generated allow scientists to have more flexibility and freedom. Therefore, the demand of mobile apps for scientific research, particularly in bioinformatics is also increasing [4].

1.1 Currently available related mobile apps

Following is a brief summary of the mobile apps available for pairwise sequence alignment (see Figure 1).



Figure 1: Related mobile apps for pairwise sequence alignment.

- Oh BLAST It! [5] – It is a mobile application developed by Hemal N. Varambhia. It helps users to determine the taxonomy of a DNA sequence using EBI-EMBL and NCBI's BLAST sequence searching algorithms. It validates the sequence and other search parameters, regularly polls the BLAST sequence search it sent and informs the user when it is completed.
- DNA Alignment [6] – This mobile application is offered by Blink Droid. It aligns DNA sequences using the mobile phone. Both global and local alignment options are available for users.
- DNA Sequence [7] – This sequence alignment tool is developed by Samuel Antonio Klein. It provides the capabilities of local and global alignment in an easy to use interface where users can input two sequences and choose various parameters of alignment for nucleotides or protein sequences.

Keeping in view that the purpose of Smartphone apps is for the convenience, we proposed stripped down, user friendly, touch screen interface to the BLAST service at European Bioinformatics Institute. In addition to the BLAST, users can also perform global sequence alignment (GSA) through dynamic programming. We compared the performance of our mobile application with other apps currently available in the market.

2. TECHNICAL ARCHITECTURE AND STUDY DESIGN

In this section, we briefly describe the two modules of MobSBlast (android app is available on Google playstore URL: <https://play.google.com/store/apps/details?id=org.scs.mobsblast.main>). The first module is based on GSA, where the alignment spans the entire length of the queries to find the best possible match. GSA method is most appropriate when the query sequence is similar and of approximately equal length. The detailed working of the GSA algorithm is mentioned in [8]. The second module of app is based on Basic

Local Alignment Search Tool algorithm. BLAST was developed by Altschul et al. in 1990 [2]. It is based on local similarity search strategy. It compares nucleotide or protein sequences to the sequence database and then calculates the statistical significance of matches. BLAST helps researchers find functional and evolutionary relationships between sequences as well as assist in identifying members of gene families. We have summarized the technical architecture of the app in Figure 2. Users can choose to run either global sequence alignment or BLAST on the main screen.

2.1 User runs Global Sequence Alignment algorithm

When a user submits two query sequences to MobSBlast app it will run the GSA algorithm locally on the mobile device in order to find out the best global alignment (Figure 3). The alignment results are displayed on the screen along with the input parameters such as match, mismatch score and gap penalty. The app also allows the users to store the results in the mobile device.

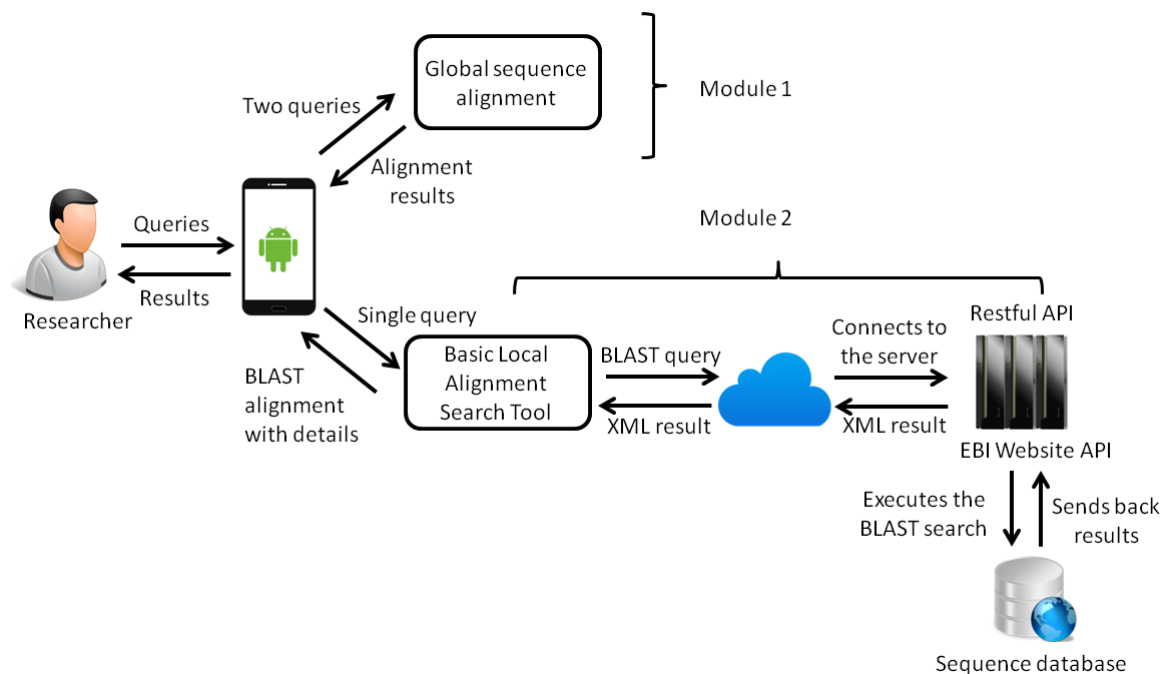


Figure 2: Technical architecture of MobSBlast.

2.2 User runs BLAST algorithm

When the user inputs a query sequence and other mandatory parameters such as the database name and BLAST type, the data is fed to the algorithm on the EBI BLAST server (Figure 4). In the background, RESTful API services are invoked with pre-specified parameters. This generates a jobid unique for a job submitted. The status for the submitted job, i.e

RUNNING, FINISHED, ERROR, FAILURE, NOT_FOUND, are retrieved in the form of response after the query is executed. The output is retrieved in multiple formats such as (XML, JSON, image, HTML, plain text) as per the requirements of the user. The output mainly consists of the alignment score, expectation value (e-value) and alignment between the query and the target sequence.



Figure 3: Steps involved in running global sequence alignment on MobSBlast.

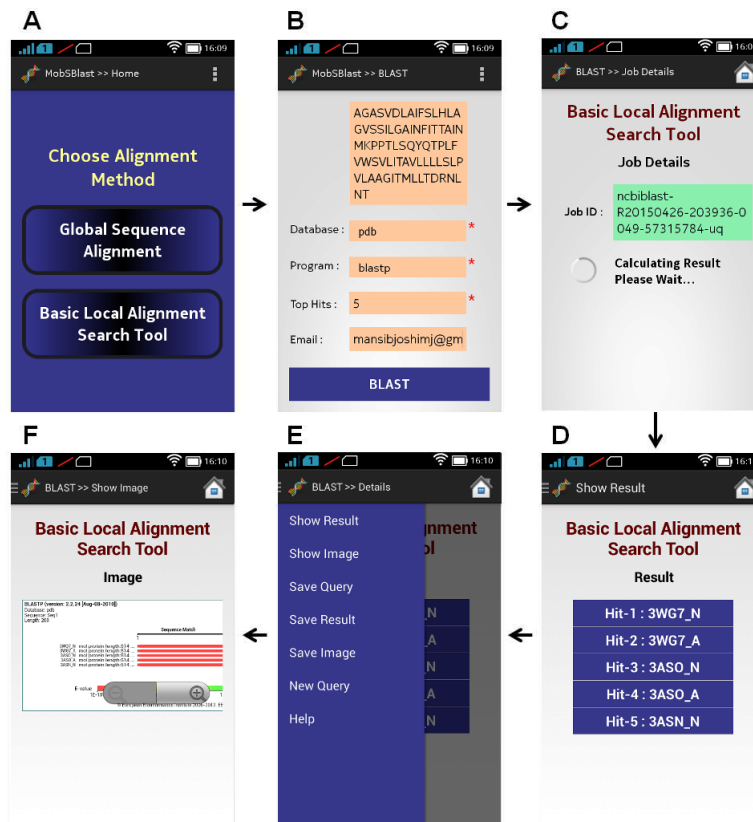


Figure 4: Steps involved in running BLAST on MobSBlast.

3. RESULTS AND DISCUSSION

The summary of the compared features of MobSBlast and other apps is present in Table 1.

Table 1. Comparative summary of the features and functionalities of MobSBlast with other apps.

Functionality	MobSBlast	DNA Alignment	Oh BLAST It!
Dual functionality of global sequence alignment and BLAST	✓	✗	✗
Facility to specify match, mismatch and gap penalty by the researcher	✓	✗	–
Load the sequences from the phone memory	✓	✗	✗
Allows to save results in the phone memory	✓	✗	✗
Image representation of BLAST result	✓	–	✗
Limit the number of hits in BLAST	✓	–	✗
Allow sequences of length above 400 characters	✓	✗	–

We note that MobSBlast is the first application which provides mobile interface to the BLAST service hosted at EBI. It also has some other interesting features such as limiting the number of hits in the BLAST output and representing the BLAST result in form of an image. Additionally, it allows users to perform pairwise global sequence alignment. In order to compare MobSBlast app with other apps, GSA and BLAST algorithms implemented in MobSBlast were executed on a mobile device with a dual core processor clocking in at 1.2GHz, carrying 1.75 GB RAM and powered by Android version 4.4.2 (KitKat) OS. The execution time of different apps running global sequence alignment algorithm are tabulated in Table 2.

Table 2. Comparison of the average execution time for running global sequence alignment in MobSBlast and other apps.

Query length	Average execution time in seconds	
	MobSBlast	DNA Alignment
20	1.30	0.79
200	1.65	0.91
400	1.97	Does not allow
600	2.36	Does not allow
800	2.91	Does not allow

We note that the execution time is proportional to the size of the query in MobSBlast and in general it requires a little bit more amount of time to complete alignments as compared to other apps. However, with the increase in the length of sequences MobSBlast fares better than other apps. We also note that except for MobSBlast none of the other apps allow sequence queries greater than 400 characters in length as an input to the GSA algorithm. Similarly, we noted the execution time of the BLAST implementation in MobSBlast. The execution time of BLAST experiments are shown in Figure 5. The execution time of BLAST algorithm in MobSBlast increases proportional to the query length however even for longer queries above 400 characters in length the app performs reasonably well. Also, it should be noted that BLAST query execution time on EBI server at any point of time can vary depending on the current status, processing load and number of concurrent request being processing by the EBI server.

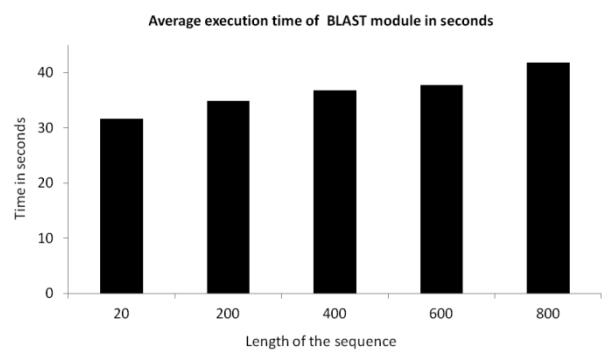


Figure 5: Average execution time for different length sequences in BLAST module of MobSBlast.

The experimental results indicate the technical feasibility and scope of MobSBlast app for running bioinformatics algorithms such as pairwise sequence alignment on android mobile devices. However, there is future scope for improving the execution time by parallelizing the algorithms using threads or by using dedicated high performance cloud servers.

4. CONCLUSIONS AND FUTURE WORK

From our survey of pair wise sequence alignment apps we found that there were various limitations in the existing apps and none of the available apps provided the pairwise sequence alignment and BLAST together. To address this issue, we proposed a new mobile app called “MobSBlast” to overcome the limitations of the existing apps. We have successfully designed, developed and implemented the mobile app on android platform mobile devices. Prototype implementation indicates that MobSBlast is efficient and would be useful in life science research. In near future, it is possible to have further enhancements in this application that would be useful to design specific and complex workflows in the domain of life sciences research. In order to provide the full capability of sequence alignment, we also plan to include the local sequence alignment module in the next version of the app. We see a healthy demand for mobile bioinformatics apps and associated cloud-based services in future. The combination of these developments will result in transition from conventional ways of desktops computing to mobile computing.

5. ACKNOWLEDGMENTS

The authors acknowledge Professor Alok Dhawan, Director, Institute of Life Sciences, Ahmedabad University for his encouragement and guidance during this project. Communication number assigned to this manuscript was ils52 for internal records.

6. REFERENCES

- [1] Green ED, Guyer MS. 2011. Charting a course for genomic medicine from base pairs to bedside. *Nature* 470 (7333):204-213. doi:10.1038/nature09764
- [2] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215 (3):403-410. doi:10.1016/S0022-2836(05)80360-2
- [3] Rivera J. 2013. Gartner Says Worldwide PC Shipments in the First Quarter of 2013 Drop to Lowest Levels Since Second Quarter of 2009 (<http://www.gartner.com/newsroom/id/2420816>). Accessed 10 April 2015
- [4] Shen H. 2015. Smartphones set to boost large-scale health studies, <http://www.nature.com/news/smartphones-set-to-boost-large-scale-health-studies-1.17083>). Accessed 12 May 2015
- [5] Varambhia HN. 2013. Oh BLAST It!, <https://play.google.com/store/apps/details?id=com.bioinformaticsapp>. Accessed 12 May 2015
- [6] Droid B. 2010. DNA Alignment, <https://play.google.com/store/apps/details?id=blink.dna.align>). Accessed 12 May 2015
- [7] Klein SA. 2014. DNA Sequence, https://play.google.com/store/apps/details?id=br.com.samuelfklein.dna.phonegap.DNA_Sequence). Accessed 12 May 2015
- [8] Needleman SB, Wunsch CD. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 48 (3):443-453