

Customizing Bioinformatics Graduate Programs for Diversified Student Backgrounds

Huanmei Wu, Oindrila Raha,
Bioinformatics Program
School of Informatics and Computing
Indianapolis, IN, USA
{hw9, oraha}@iupui.edu SOIC,

Jian Zhang
School of Computer Engineering,
Shenzhen Polytechnic,
Guangdong, Shenzhen, P.R.China,
jzha930@szpt.edu.cn

Abstract— For graduate programs, usually, the curricula are developed based on the generic learning outcomes of the matching undergraduate programs. However, most graduate programs have diversified students from various backgrounds with undergraduate degrees from an assortment of majors. Even for graduate students coming from the same undergraduate majors, they might come from different countries with differing undergraduate learning outcomes. It is even more challenging for graduate programs where there are no corresponding undergraduate programs. The graduate program curriculum should be customized based on the backgrounds of special student groups. This paper will describe the redesign of our professional Bioinformatics MS program for incoming domestic and international students with diverse backgrounds. The MS students in Bioinformatics program may have a previous degree in biotechnology, computer science, biology, computer engineering, electrical engineering, biomedical engineering, or other science and engineering fields. Thus, it is of great significance to categorize the program specific competencies and student learning outcomes from their previous study. It will help to customize the student specific plan of study in the Bioinformatics MS program so that they can complete the graduate study with the job-ready skills. For example, detailed studies have been carried out for the Bioinformatics related programs in India, where a substantial population of our Bioinformatics MS students comes from. In India, Bioinformatics is a growing subject and has emerged as an independent program from biomedical engineering and biotechnology. The Bioinformatics in India is taught at in all different degree levels: Bachelors, Masters, PhD, and certificates. Graduate students from India who are familiar with the Indian education system have helped to compare and contrast the various Bioinformatics related degree programs, including both Bioinformatics and Biotechnology degree programs of BS, MS, B. Tech. (Bachelor of Technology), M.Tech. (Master of Technology), M.Phil (Master of Philosophy), PhD, and integrated (such as BS+MS or MS+PhD) programs. The corresponding program competencies and the student learning outcomes are tabulated and compared. The job market in India and the USA are also analyzed. The information has been used to redesign our MS Bioinformatics program, including the alternative prerequisites, the different sequences of courses, and the

diverse plans of study for students with various backgrounds. For example, some existing 3-credit introductory courses (such as Programming in Bioinformatics) are re-designed into various 1-credit common course modules (such as Programming in R, or Python, or Perl). Students are provided with more flexibility to select specific course modules based on their backgrounds and future career interests. New course modules have also been developed based on the job market needs, such as big data analysis for Bioinformatics. These student-oriented and career-ready customization of graduate programs will better serve our expanded student groups and provide a better workforce for the job market.

Keywords—*Student-oriented; Bioinformatics Curricula; International Students*

I. INTRODUCTION

Bioinformatics is a combination of subjects that improves methods and software tools for understanding genetic, protein or biological data. As an interdisciplinary subject of science, Bioinformatics combines genetics, molecular biology, computer science, biostatistics, engineering, and mathematics to investigate and understand biological data [1].

Bioinformatics is a new field comparatively to other field of science as genetics, biotechnology or engineering. In developing countries like India, bioinformatics is still in its infancy. Bioinformatics in India is still a developing subject.

Bioinformatics was initially introduced in India as a part of biotechnology or molecular biology. As a subject, bioinformatics was introduced in India as a diploma or certificate program in Pune University in the year 2002. The duration of the diploma/certificate program is six months to one year. It took some time for Bioinformatics to be accepted as subject in the universities. Still, many universities offer it as a certificate program.

At present, bioinformatics is taught at different program level such as bachelors, masters and in doctorate level. Bioinformatics students are getting opportunities in bioinformatics job field. Many students pursue their studies abroad specially in the USA. There are good percentage of students who leverage their education in bioinformatics at different universities.

Generally, students after completing their bachelors or masters studies in India continue their master's education from

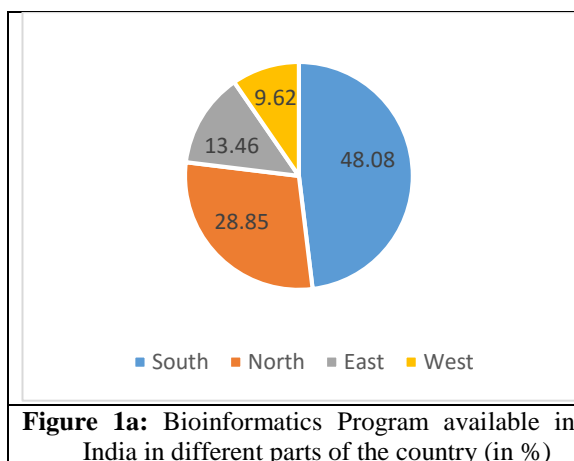


Figure 1a: Bioinformatics Program available in India in different parts of the country (in %)

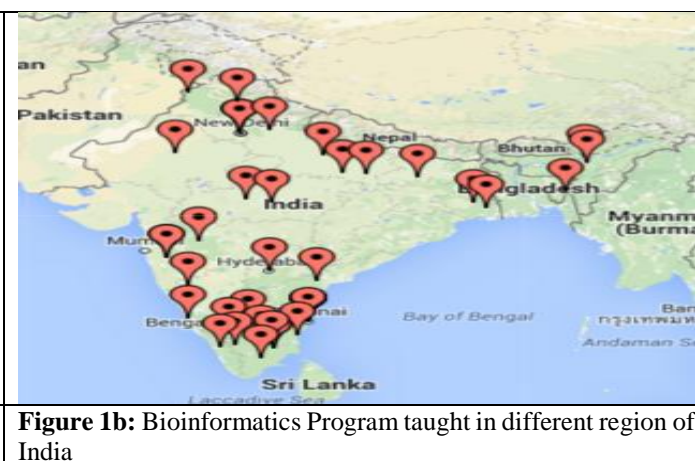


Figure 1b: Bioinformatics Program taught in different region of India

the USA. It is estimated that around thirteen percent (13%) of international student enrollment is from India [2; 3; 4]. Students experience a different environment in the USA. Students from different background face adjustment problem such as learning about new concepts of bioinformatics. They also need to adjust to multi-cultural environment in the USA. They need to live as an independent, and isolated from the family. The students earn their money, cook their own food, learn to drive their own car and take care of themselves is a part of the journey of USA education system. The journey is interesting and sometimes it is bit exhilarating. There are many blogs on the journey to the USA from Indian students and all agree to one point that the first semester is the hardest part [3; 5; 6].

This paper aims to analyze Bioinformatics as a subject taught in India. The different programs and departments under which bioinformatics is taught. This paper examines the job skill set required in the market. This paper also evaluates the prerequisites for joining Masters in bioinformatics in the USA universities.

II. BIOINFOMATICS: THE SUBJECT

A. Bioinformatics Subdisciplines

In the last few years, the term "Bioinformatics" has been used by genetists/molecular biologists to specify the subject constituent of genomics and is considered to include a computer-assisted analysis of all data on sequences of DNA. However, many scientists such as biologists, physicists, mathematicians, have accepted "Bioinformatics" as a substitute for science cybernetics. Thus, "Bioinformatics" embraced genomics and all sections of the biological science [7]. Bioinformatics include study of information processes (storage, transfer, and processing of information, etc.) of macromolecules to the higher animals and human. As mentioned earlier, bioinformatics utilizes and applies computational methods for analysis of large collections of biological data[8; 9].

Bioinformatics is a field which similar and again different from biological computational biology or computation.

Biological computation practices bioengineering and biology to make biological computers, whereas, Bioinformatics uses computation to understand biology. Bioinformatics' basic function is to organize and analyze basic biological data such as protein, DNA and RNA [10; 11; 12]. Computational biology concentrates on theoretical models of biological systems, similar to mathematical biology does with mathematical models.

Analysis predominantly focuses on three types of datasets available in molecular biology. There are important sub-disciplines in bioinformatics. These include:

- Development of algorithms and biostatistics, which assess relationships among data sets. The analysis and interpretation of various types of data (nucleotide and amino acid sequences, protein domains, and protein structures), and
- The search for text of scientific papers or journals.

The other sub discipline includes "relationship data" from metabolic pathways, taxonomy trees, and protein function. Bioinformatics involves a number of computational techniques such as sequence and structural alignment, database design and data mining, geometry of protein or macromolecular, phylogenetic tree creation, gene finding, and expression data clustering. Bioinformatics is a practical discipline.

B. Bioinformatics in different Programs in India

In India, if we divide the country in four regions based on geographical zones, we can observe that in South part of the country has more than fifty percent (50%) of the credit for offering Bioinformatics as a subject (Fig. 1). Bioinformatics is taught as a science subject and as well as an engineering subject in different universities in India. Majority of universities offers it as a science subject (34% vs 23%) (Fig.2). Program duration is different from the US. Bachelor is taught as four year program in the USA. In India, Bioinformatics is offered as a variety of program offerings. It has been offered as Bachelor of Science (BS/BSc.), Master of Science (MS/MSc.), Bachelor of technology (B.Tech.) and Master of Technology (M.Tech.). B.Tech. is an undergraduate academic degree. It is conferred after completion of a four-year program of studies at an accredited university and is also recognized in the USA.

In India, a Master of Engineering/Master of Science in Engineering (M.E.) or Master of Technology/Master of Science in Technology (M.Tech.) degree is a postgraduate program in

engineering or technology field. This is generally a two or two-and-half year specialization program in a specific branch of engineering or a technical field. M.Phil is advanced postgraduate research degree with thesis option. Engineering Programs in India are generally structured as research degree and considered to be parallel or equal to similar with M.Sc. and M.Phil programs in Science [13]. Bioinformatics is offered as both three year (BS/BSc.) and four year (B.Tech.) programs and as well as science and engineering subject. This is generally followed by Masters in either science or engineering. Universities which offers either MS/M.Tech./M.Phil., also, generally offers PhD (Table 1). When we compare both Bioinformatics and biotechnology program distribution in India, we observe that biotechnology as a subject is offered in more number of universities than bioinformatics.

C. Bioinformatics in India

Bioinformatics program in India is taught under different schools or departments. In the USA, bioinformatics is generally taught under school of informatics, system biology, computational biology, computer science, biology or molecular biology, bioinformatics [1; 10; 14]. Whereas in India, Bioinformatics is taught under mainly three categories: biotechnology, computer science and Bioinformatics.

Multidisciplinary sciences include biophysics, mathematics, applied sciences, agricultural science and or combination of all of them. The *wordle* explains the distribution of department names. It is evident that bioinformatics is mainly taught under the bioinformatics followed by biotechnology department (Fig. 3a and 3b).

An introduction to main areas of bioinformatics include: introduction to bioinformatics, databases, genomics, DNA and protein sequences, protein structures. This also include a practice of the most common computational tools used in Bioinformatics. In the Fig. 4, we have created the *wordle* to understand which courses are emphasized in different degree programs in bioinformatics. We have selected few premier universities/institutes to understand the curriculum of the subject [14; 15; 16; 17; 18]. We have separated the course in B.Tech., M.Sc. and M.Tech.. We can observe from *wordle* (Fig. 4) that database study and statistics are taught in every level and are emphasized.

Both master’s program has gene expression or microarray study. On reviewing the course material it is repeated in all eight courses in M.Tech. program. Till now next-generation sequencing (NGS) has not found its place in the curriculum in India. Programming language are taught in all the engineering programs. The courses for programming includes Python, Perl, Java or C language. Perl has found its foothold in most of the programs. Perl is repeated in all M.Tech. , B.Tech. and MS programs and closely followed by Python.

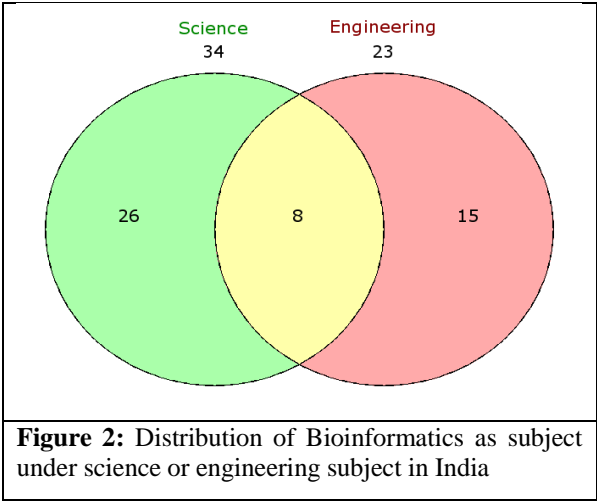


Figure 2: Distribution of Bioinformatics as subject under science or engineering subject in India

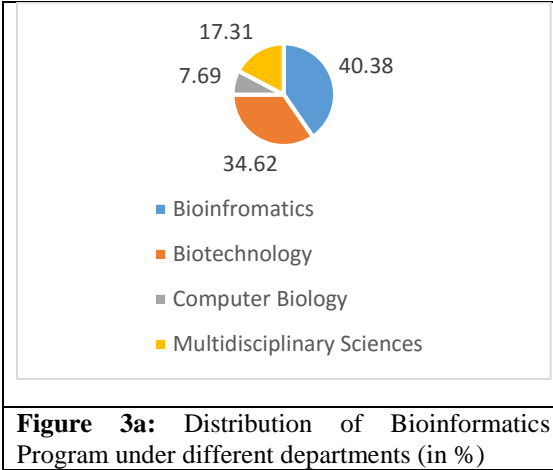


Figure 3a: Distribution of Bioinformatics Program under different departments (in %)

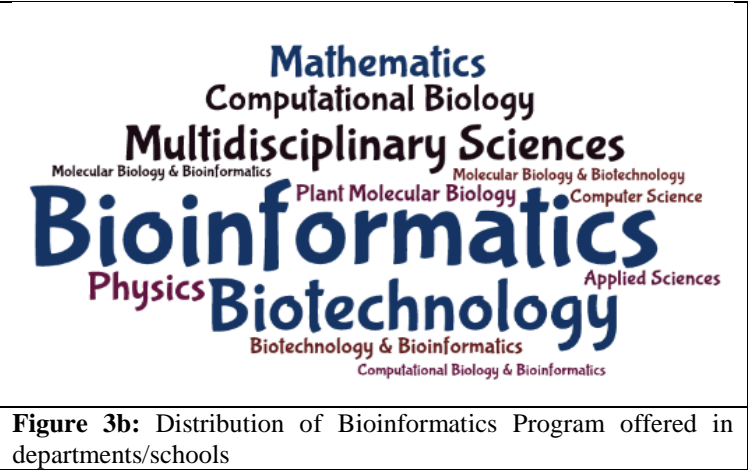
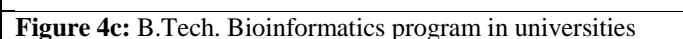
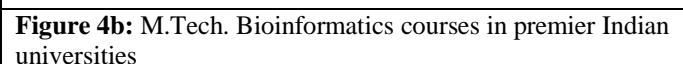
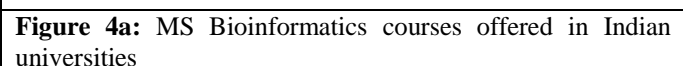


Figure 3b: Distribution of Bioinformatics Program offered in departments/schools

Universities Teaching in different degrees	Bioinformatics (%) N=52	Biotechnology (%) N=117
B.Sc + B.Tech	-	1.71
B.Sc +M.Sc+ B.Tech	-	0.85
B.Sc only	1.92	3.42
B.Sc+B.Tech+M.Tech	1.92	0.85
B.Sc+M.Sc	1.92	5.13
B.Sc+M.Sc+B.Tech	-	2.56
B.Sc+M.Sc+M.Phil	-	2.56
B.Sc+M.Sc+Phd+M.Phil	-	3.42
B.Tech only	5.77	11.97
B.Tech+M.Tech	5.77	10.26
B.Tech+M.Tech+PhD	7.69	6.84
Diploma/Certificate only	3.85	1.71
M.Phil only	0.00	1.71
M.Sc only	26.92	18.80
M.Sc+B.Tech	1.92	1.71
M.Sc+B.Tech+M.Tech	-	1.71
M.Sc+B.Tech+M.Tech+Ph D	-	3.42
M.Sc+B.Tech+PhD	1.92	0.85
M.Sc+M.Tech	1.92	1.71
M.Sc+M.Tech+MPhil	1.92	0.85
M.Sc+M.Tech+PhD	7.69	0.85
M.Sc+M.Tech+PhD+M.Ph il	-	2.56
M.Sc+PhD	17.31	9.40
M.Tech only	9.62	1.71
PhD only	1.92	0.85



A. Job opportunities and Requirements for Bioinformaticians

Understanding genomics and proteomics data is the goal of bioinformaticians. Progresses in information technology, especially in storage and speed of internet, has allowed easy access to the field. With genetics as subject is now come from bench to bedside and is becoming more as a personalized medicine. This has led to pharmaceutical companies also using expertise of bioinformaticians [20]. Thus, bioinformatics research and development (R&D) units are being recognized in many pharmaceutical and drug companies. Bioinformatics is thus considered as number one career in the field of biology [12]. In India, the job market is growing. We can see the distribution of Bioinformatics in different field (Fig. 5).

We can split the activities in bioinformatics requirements in two areas (1) the health care and (2) the analysis of biological data. Analysis activity in Bioinformatics in healthcare field is mainly based on coding. The biological data analysis involves development of techniques to predict the function of RNA sequences, clustering protein sequences into families of related sequences and the development of protein models, aligning similar proteins and generating phylogenetic trees to examine evolutionary relationships.

example, job market has requirements for fresh graduates with fewer skill sets. Academics requires more in-depth knowledge and higher skill sets (Fig. 6).

- Training of students in the fundamental disciplines relevant to bioinformatics.
- Training of students in It to acquire the essential IT and Bioinformatics skills.
- To develop the competency in problem solving.

In Indian job market, programming skills are considered as an assets. Jobs in the industry mainly concentrate on programming skills. The programming language currently in demand are Perl, Python, Java, C++.

It is important to verify the fact that how many percentage of students are going abroad to pursue higher studies in Bioinformatics. There has been no study or reports yet. Students who come from diverse background such as Biotechnology or Genetics, face struggles when thrown into the world of computation.. Informatics students are more skilled in the computational areas. It is very essential for students to have basic knowledge of bioinformatics and programming.

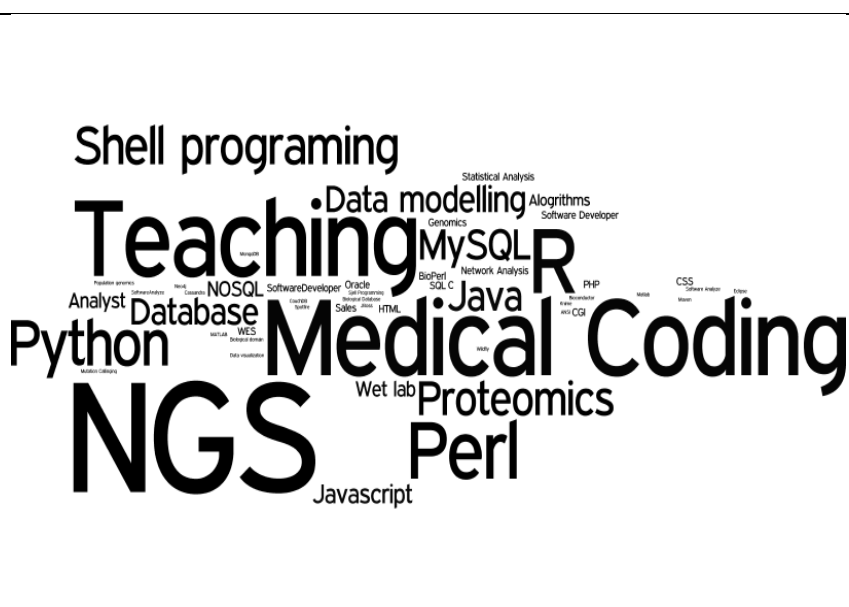
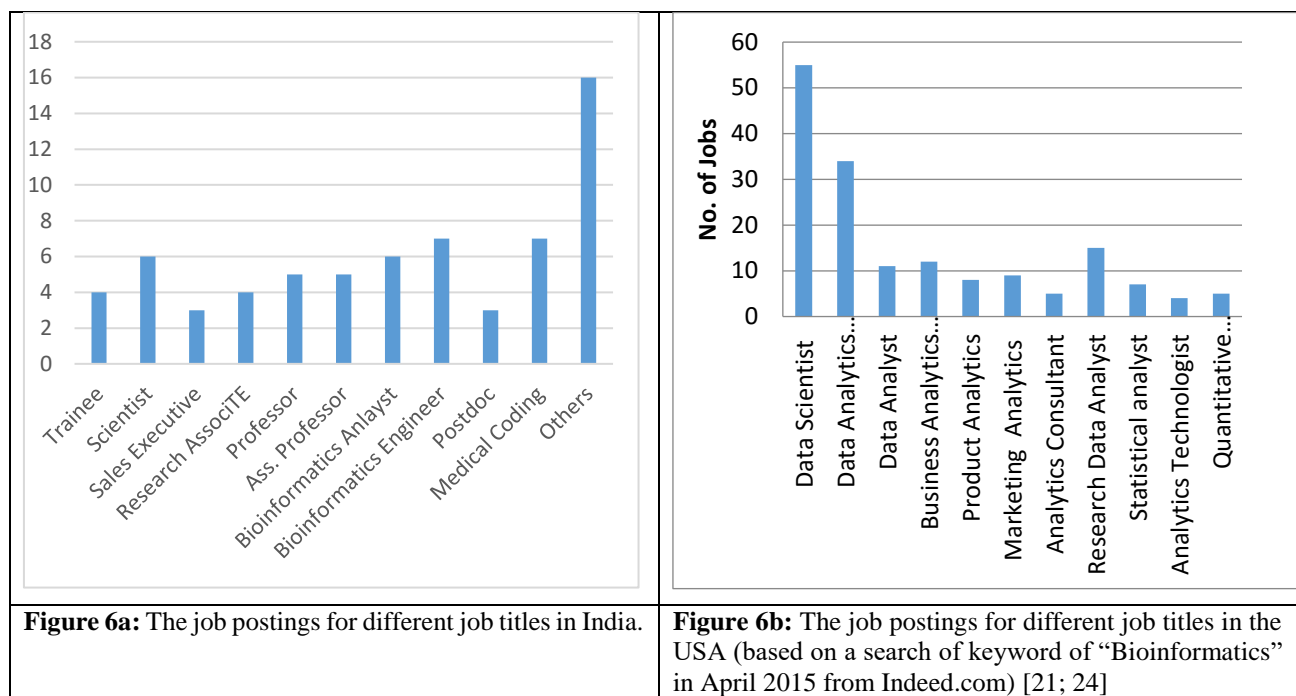


Figure 5b: The required skill sets for Bioinformatics job market.



Knowledge of using NCBI, Uniprot, KEGG tools should be compulsory. Students with biology background need to understand the concept of programming. Hard core programming language such as C++, C or JAVA need not be mandatory for students joining in bioinformatics course in India or abroad. But, they should be aware of the concept of programming and terms such as loops, iterations, object-oriented. The current job market requires sound knowledge of R, Python. These programming languages are taught in different online courses such as Coursera, Datacamp, Khan Academy. It's advisable to students to go through these sites and acquire knowledge.

IV. COMPARISION

On comparing IUPUI course structure with premier Indian universities, we observed that students from biology program need to have following knowledge: Programming basics: Familiarize with Definition, symbols, commands, in short have basics of programming [24]. It will be preferable if a prerequisite course in programming are made compulsory. These subject enable students to understand of basics of bioinformatics. *Introduction to Bioinformatics* as course is a transition between biology and computation. Students should have basic statistical programming language like R, scripting language such as Perl or Python, and a compiled language such as C, C++ or Java (Fig. 7).

Students with different degree program, subject background and skill sets join Master's program in the USA. Students with B.Sc. program from India has only fifteen years of education. The requirement for enrollment for Master's program in the USA universities is sixteen years of education. Hence, students need to complete Pre-Master's education. The education includes Bachelor level courses.

It is important to mention that students from computer science also join bioinformatics program. These may not be common in India, but in many US universities, students with four-five yrs. experience in IT sector job take Bioinformatics Master's program. For them, the programming part might be easy but they lack knowledge of biology. Hence, introduction to Bioinformatics is essential to these category of students. An introduction to Molecular Biology or Genetics and Biochemistry is essential.

There has been many studies from different countries where they have described the development of Bioinformatics course over a period. Studies from Africa has shown that the progress of Bioinformatics in combating deadly diseases such as Malaria and HIV. The studies also emphasizes on the improvement of the subject, so that they can compete with developed countries [25; 26; 27]. Ranganathan *et al.*, has given number of updates on bioinformatics subject in Asia-Pacific region [16; 23; 28].

V. CONCLUSION

Bioinformatics in India is still developing and needs more universities to come up with proper course structure. Government of India have taken many steps for the growth of bioinformatics as a subject. Several research institutes have now dedicated departments for projects. Some of them are funded by Government of India. This aided in the development of bioinformatics resources and publications in international journals of high impact factor. In India, with its strengths in information technology, increasing investments in bioinformatics infrastructure and human resource development, is poised to play a greater role in the global landscape in future. Inclusion of basic programming knowledge will also help students who are willing to join the USA universities for higher studies.

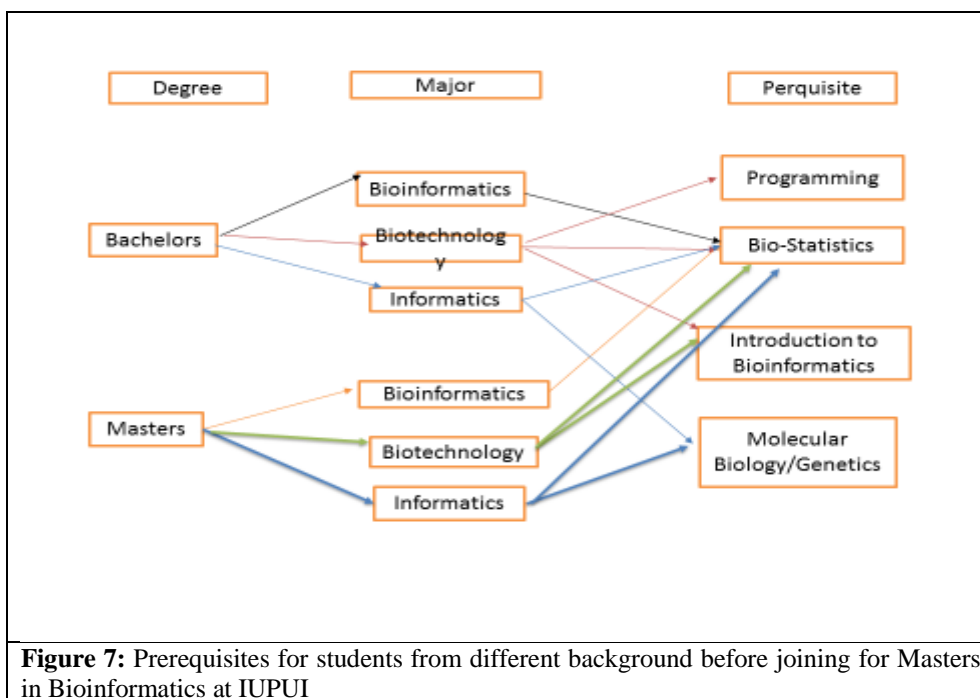


Figure 7: Prerequisites for students from different background before joining for Masters in Bioinformatics at IUPUI

REFERENCES

- [1] Kulkarni-Kale U, Sawant S, and Chavan V, "Bioinformatics education in India". *Brief Bioinform.* 11 2010 Nov 616-25.
- [2] <http://www.iie.org/Services/Project-Atlas/United-States/International-Students-In-US#.V3VKIgrLJU>.
- [3] <http://www.usnews.com/education/best-colleges/articles/2014/11/17/number-of-international-college-students-continues-to-climb>.
- [4] <http://wenr.wes.org/2013/12/indian-study-abroad-trends-past-present-and-future/>.
- [5] <http://www.usnews.com/education/best-colleges/articles/2012/08/28/6-challenges-for-international-students-in-college>.
- [6] <http://www.msmbainusa.com/articles/entrance-preparations/challenges-faced-by-indian-students-in-foreign-campuses/>.
- [7] Arora JR, "Growth of Biotechnology in India- A tribute to Dr. S. Ramachandran ", Narosa Publishing House, Biotechnology 1992.
- [8] Magana AJ, Taleyarkhan M, Alvarado DR, Kane M, Springer J, and Clase K, "A Survey of Scholarly Literature Describing the Field of Bioinformatics Education and Bioinformatics Educational Research", 2014.
- [9] <http://www.bioinformation.net/>.
- [10] Gupta S, Chavan S, Deobagkar DN, and Deobagkar DD, "Bio/chemoinformatics in India: an outlook". *Brief Bioinform.* 16 2015 Jul 710-31.
- [11] Ramachandran S, and Kolaskar AS, Bioinformation systems in India, computer handling and discrimination of data. in: CODATA, (Ed.), In: Glaeser PS, Elsevier Science B.V, 1987, pp. 394-9.
- [12] Glassdoor.com.
- [13] https://en.wikipedia.org/wiki/Master_of_Engineering.
- [14] www.bii.in/programs/distancelearning/overview.html.
- [15] bioinformaticsweb.net/courses.html.
- [16] Ranganathan S, "Bioinformatics Education—Perspectives and Challenges". *PLoS Comput Biol* 1 2005 e52.
- [17] www.jagranjosh.com/bioinformatics-colleges.
- [18] www.rxpgonline.com/article1566.html.
- [19] Yang JY, Yang MQ, Zhu M, Arabnia HR, and Deng Y, "Promoting synergistic research and education in genomics and bioinformatics". *BMC Genomics* 9 2008 I1.
- [20] R. Altman, "A curriculum for bioinformatics: the time is ripe". *Bioinformatics* 14 1998 549-550.
- [21] Indeed.com.
- [22] Monster.com.
- [23] Ranganathan S, "Towards a career in bioinformatics". *BMC Bioinformatics* 10 2009 S1.
- [24] Wu H, and Palani A, "Bioinformatics Curriculum Development and Skill Sets for Bioinformaticians", FIE, 2015.
- [25] Fatumo SA, Adoga MP, and Ojo OO, "Computational Biology and Bioinformatics in Nigeria". *PLoS Comput Biol.* 10 2014 e1003516.
- [26] Karikari TK, "Bioinformatics in Africa: The Rise of Ghana?". *PLoS Computational Biology* 11 2015 e1004308.
- [27] Nashiru O, Huynh C, Doumbia S, Kissinger JC, Isokpehi RD, and Olorunsogo OO, "Building bioinformatics capacity in West Africa". *Afr J Med Med Sci.* 36 2007.
- [28] Ranganathan S, Hsu WL, Yang UC, and Tan TW, "Emerging strengths in Asia Pacific bioinformatics". *BMC Bioinformatics* 9 2008 S1.