

Research Survey

RISE OF BIOINFORMATICS AND COMPUTATIONAL BIOLOGY IN INDIA: A LOOK THROUGH PUBLICATIONS

Anjali Srivastava, Ankita Srivastava, Arshiya and Suman K Mallik*

BTIS-SDIC, CSIR-Central Drug Research Institute, Lucknow – 226021, India

Abstract: Computational biology and bioinformatics have been part and parcel of biomedical research for few decades now. However, the institutionalization of bioinformatics research took place with the establishment of Distributed Information Centres (DISCs) in the research institutions of repute in various disciplines by the Department of Biotechnology, Government of India. Though, at initial stages, this endeavor was mainly focused on providing infrastructure for using information technology and internet based communication and tools for carrying out computational biology and *in-silico* assisted research in varied arena of research starting from disease biology to agricultural crops, spices, veterinary science and many more, the natural outcome of establishment of such facilities resulted into new experiments with bioinformatics tools. Thus, Biotechnology Information Systems (BTIS) grew into a solid movement and a large number of publications started coming out of these centres. In the end of last century, bioinformatics started developing like a full-fledged research subject. In the last decade, a need was felt to actually make a factual estimation of the result of this endeavor of DBT which had, by then, established about two hundred centres in almost all disciplines of biomedical research. In a bid to evaluate the efforts and outcome of these centres, BTIS Centre at CSIR-CDRI, Lucknow was entrusted with collecting and collating the publications of these centres. However, when the full data was compiled, the DBT task force felt that the study must include Non-BTIS centres also so as to expand the report to have a glimpse of bioinformatics publications from the country.

Keywords: Bioinformatics; Computational Biology; India; BTIS Net

Note - Coloured Figures available on Journal Website in "Archives" Section

Introduction

Bioinformatics and computational biology like all walks of life, has had a drastically rapid growth after the information technology boom. The speed of new knowledge generated specially in biological sphere after the revolution in genetics could be managed only because the information technology could cater to the huge data analysis at significantly

high speed with the introduction of huge storage capacity and high speed hardware and useful bioinformatics tools, software and servers. This brought the area of bioinformatics into the centre stage of all research in the arena of biological world.

Though, knowledge generations in all other fields are easily recognizable and evaluation of growth of the subject is comparatively easy to make, bioinformatics has a marked limitation. In spite of the fact that the area has had a tremendous impact, its presence is not as evident as is seen with respect to other areas. There are issues of scholars and researchers still considering it to be just a tool and not a full-fledged area of science. Though everyone

Corresponding Author: **Suman K. Mallik**

E-mail: sumanmallik@cdri.res.in

Received: September 3, 2017

Accepted: September 22, 2017

Published: September 23, 2017

accepts that in today's world no new knowledge in biological world can be imagined without using one or other bioinformatics tools and bioinformatics is indispensable, full recognition of bioinformatics as a subject is yet to take place (Barlette *et al*, 2017).

Bioinformatics and computational biology, as term quite common and one gets million of result giving these terms as keywords in any search engine. However bioinformatics as a distinct subject has been classified by only few databases (Benton, 1996).

However, Bioinformatics grew in the last three decades all over the world with tremendous speed. International realization that bioinformatics and computational biology had to be nurtured led all major universities and institutions around the world to have the subjects as the major subject area in their curriculum (Curioso *et al*, 2008). Governments funding agencies also put lots of emphasis on encouraging development of sufficient infrastructure and manpower in bioinformatics. In India, the realization was simultaneous with the world and government established a robust bioinformatics program within the Department of Biotechnology, Government of India. The establishment of Sectoral Centres of BTIS NET helped push the momentum of bioinformatics facilities across the length and breadth of country. The Centres, almost two hundred in numbers, have contributed a great deal in bringing bioinformatics at par with the international level of understanding (Krishnaswamy *et al*, 2016).

Since India is one of the few countries where the government took a serious initiative more than three decades ago to accelerate the growth of the then emerging area of research, it was appropriate to evaluate whether the efforts of Department of Biotechnology of Govt. of India (DBT) yielded the desired result (Gopal *et al*, 2015).

The BTIS Centre of DBT at Central Drug Research was entrusted upon this task of making a detailed survey of the outcome of these Centers. Evaluation of any such endeavor is somewhat difficult tasks as the outputs are many times not measurable especially in shorter span of time. All the efforts do not result into a tangible outcome and thus, evaluation result many times become subjective.

However, in this particular study tangible outcome such as research publications generated

out of the efforts of BTIS centers as well as other organization have been taken into consideration.

It was observed that there was no effort to compile Indian publications till 2010. Lack of any compilation led to speculation and criticism of the efforts of Dept of biotechnology efforts in supporting bioinformatics. DBT had to substantiate the outcome of the financial support provided to BTIS Centers.

The compiled result in this direction was quite encouraging and a compendium was published by DBT. This compendium was not only appreciated but the process of collecting information continued hereafter. DBT encouraged us to have annual compilation of the Centers and even started annual incentive awards for best papers in bioinformatics and computational biology.

The present presentation is all about the result of all these efforts of collection and compilation. The data on publications will be presented such as number of publications, subject-wise papers and results based on various parameters. The actual problem in collecting such data will also be elaborated giving same glimpses of how they could be dealt with. Overall, this presentation is an effort to look at the growth of bioinformatics in India through the publications in reputed journals by a large number of research organizations.

The aim of this project was to compile the Bioinformatics literature for the period 2002 to 2016 from DBT funded institutions as well as other research groups involved in bioinformatics and computational biology. The purpose also was to assess the literature growth year wise to see if the research publications have experienced any upward or downward trends over the years both number-wise and quality-wise. We also wanted to see if the publications in bioinformatics were generated only from few niche and bigger institutions or they were being published from smaller institutions as well. To it was also aimed at mapping the geographic distribution of the literature to see if the publications are generated from few geographical pocket of the country or otherwise. Since funding patterns of DBT to the Centers as well as the levels of infrastructure and manpower available at different groups of Centers were drastically uneven, the endeavor was also to see how different groups of Centers performed. Thus, it was undertaken to study the distribution of Papers among various groups of Bioinformatics Centers (Centers of Excellence (Coe),

Distributed Information Centers (DICs), Distributed Sub-Information Centers (SDICs) and Bioinformatics Facilities (BIF).

Materials and Methods

Data were collected from Annual Reports of Institutions, Report from Coordinators; Reports sent to DBT from each Centre and publication lists of the Centres were downloaded from Web of Knowledge, Pubmed / science direct / open access sites etc.

So the metadata of all publications of BTIS centres of all categories were compiled. The source of data was mostly the reports of the co-coordinators of the centres provided to the DBT for the renewal of the project or for the compilation of the proceedings to the co-ordinations Annual meetings.

Since many of the centres did not necessarily provide such information and many papers were not available in the online databases, individual contacts were made to compile the data. Apart from that website and personal profile of authors of the member institutions, annual reports of the organization available in their website were also used for data collection.

We listed each publication with author wise i.e. single, double or multi author separating in individual year. Since there were many institutions which are not part of the BTIS Network, especially in recent years, from where a large number of publications in bioinformatics have been published, any result on Indian publications status would be erroneous if only BTIs centres publication were to be included in the study. So, it was decided to include other papers as well which were not published under the BTIS NET banner. Data were collected from contacting such authors as there are varieties of research work. Apart from web of science and Pubbmed as a source, non-BTIS Institutions authors names were also suggested by senior members of the DBT task force members. Annual Reports of IITs, NISERS, ISERS, etc, were downloaded. Mails were sent to each individual/scientists to get their publication list so as to make the compilation as comprehensive as possible.

Criteria for Inclusion of Data

Since bioinformatics is hidden in all aspects of biomedical research arena and is not taken as a

specific field by many, the authors tended to send lots of papers to be taken as bioinformatics and computational biology publications. Another major reason was that the authors tried to send as maximum number as possible as the data was being compiled by the funding agency (DBT). After having compiled all the metadata or publications, it was realized that the list was too big to even imagine. There were papers which just touched small aspect of bioinformatics and some even had no element of any computational work involved or had mostly experimental data.

So the task force decided to have strict criteria for inclusion of a paper in this list. The criteria set was: At least 50% of the work mentioned in a publication must have bioinformatics or computational biology component. Application of these criteria led to discard of a very large portion of publications from the list.

Results and Discussion

After the data was finally compiled and the papers selected on the basis of the criteria set up a hard copy of the compilation was published in a hard copy format in 2010. The same data was bifurcated quantitatively under various categories such as institutions, centre group, authors, subject etc and result were compiled.

Growth of number of Publications over the Years

Figure 1 gives an illustration of the number of publications during the period of fifteen years from 2002 to 2016. It is observed that highest no of papers were published 2016 (1051) followed by 2014 and 2015 and there has been steady growth of publications since 2002. The research productivity in Bioinformatics is on the increase though not uniform. At the start of the new Millennium, the growth is less while in 2016 the growth rate is more. The findings vividly indicate that the growth of literature in Bioinformatics is on increase.

State wise DBT Institutions

The data was looked from the geographical angle of the location of the institution. It is seen that just like the existence pattern of the location of the BTIS Centres all over the country, the publications have also been made from the entire length and breadth of the country (Fig 2).

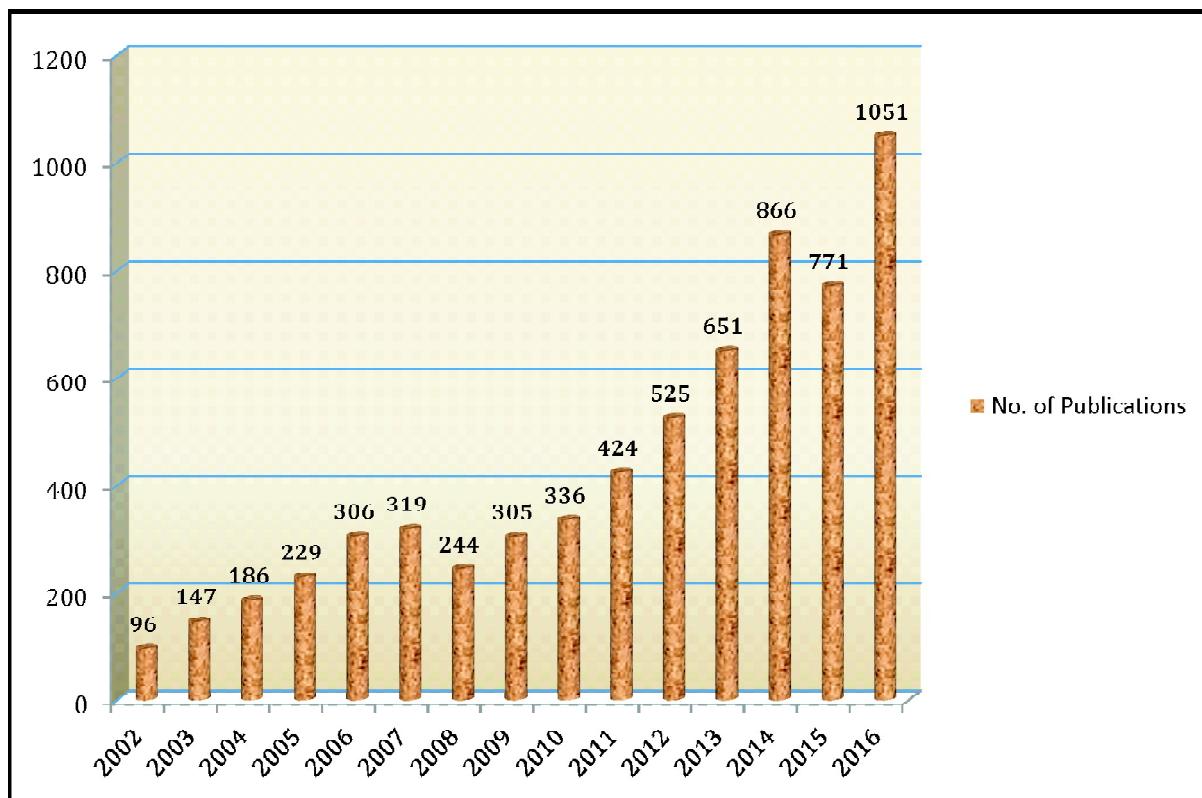


Figure 1 : Year-wise publications in Bioinformatics and Computational Biology (2002-2016)



Figure 2: Publications in Bioinformatics from all over the country.

Ranking of States as per the number of Publications

Table 1 shows the state-wise DBT Centres in India. As shown in list in Delhi there are 17 Centres, Tamil Nadu 17, Uttar Pradesh 14 followed by Assam 11, Karnataka 9 and so on. Bottom seven states Bihar, Goa, Jharkhand, Nagaland and Pondicherry, Telangana and Andaman and Nicobar Islands (UT), all have only one Centre.

Due to more no of DBT centers, Delhi as a state is at number 1 position (Fig. 3) in terms of publishing research papers, followed by Uttar Pradesh and Tamil Nadu, and so on.

Growth Pattern - State-wise

Using data on the largest Indian states, we summarize growth patterns across the states during the period 2013-2016. The chart shows that all the states, with the exception of Assam, Meghalaya, Tamilnadu and West Bengal publications growth is substantially noteworthy (Fig 3).

Publications from Various Categories of Centres (COEs, DICs, SDICs and BIFs)

Figure 4 depicts the chronological distribution of papers published each by the groups named as BIF, SDIC, DIC, CoE Institutions during 2013 to 2016. From this distribution of publications it is reflected that during year 2013 published papers were quite less in comparison to those in further productive years. It is observed that highest no of papers were published by BIF in 2014 (374), SDIC in 2016 (431), DIC in 2016 (105), CoE (164) in 2016.

Table 1
State-wise distribution of publications

State	DBT Institutions	State 2	DBT Institutions 2
Delhi	17	Uttarakhand	4
Tamil Nadu	17	Chhattisgarh	3
Uttar Pradesh	14	Gujarat	3
Assam	11	Meghalaya	3
Karnataka	9	Rajasthan	3
Kerala	9	Arunachal Pradesh	2
West Bengal	9	Mizoram	2
Andhra Pradesh	8	Sikkim	2
Odisha	7	Tripura	2
Haryana	6	Bihar	1
Jammu and Kashmir	6	Goa	1
Maharashtra	6	Jharkhand	1
Madhya Pradesh	5	Nagaland	1
Himachal Pradesh	4	Pondicherry	1
Manipur	4	Telangana	1
Punjab	4	Andaman and Nicobar Islands	1

SDIC, DIC, CoE Institutions during 2013 to 2016. From this distribution of publications it is reflected that during year 2013 published papers were quite less in comparison to those in further productive years. It is observed that highest no of papers were published by BIF in 2014 (374), SDIC in 2016 (431), DIC in 2016 (105), CoE (164) in 2016.

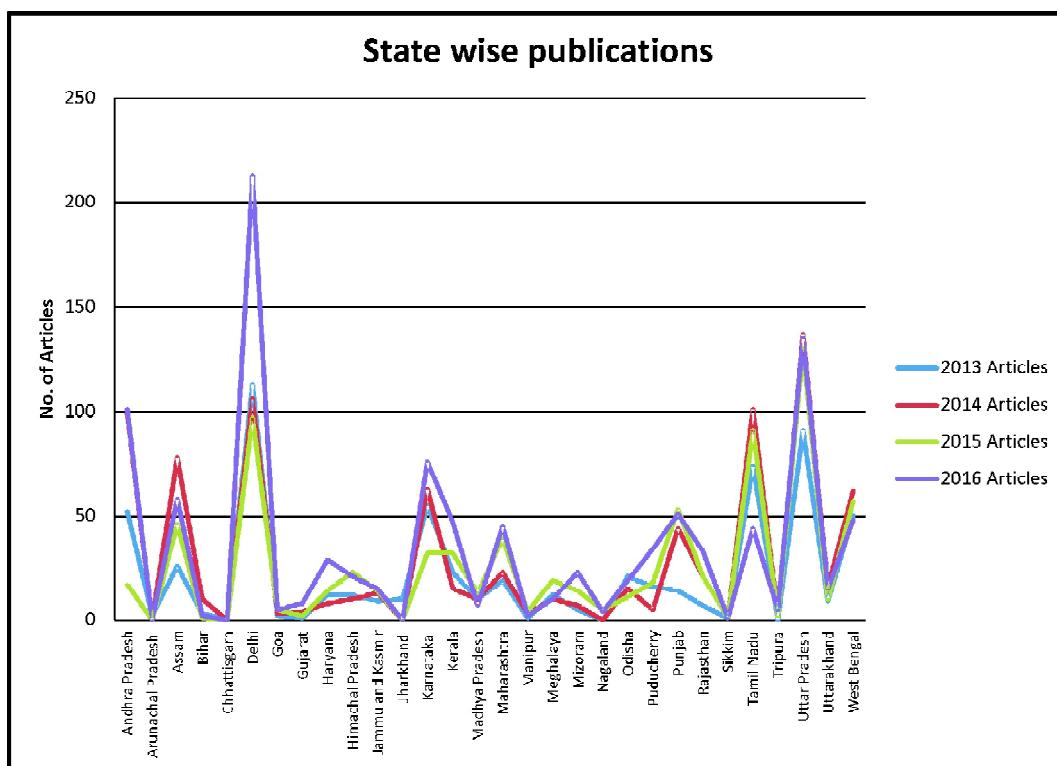


Figure 3: Growth pattern - State-wise

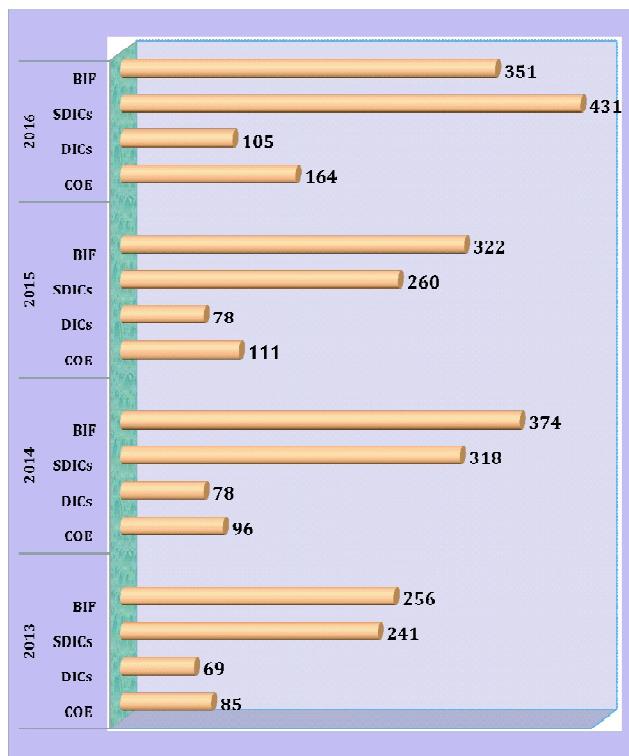


Figure 4 : Contribution of research papers during 2013 to 2016 by each group of Centres.

Conclusion

One very evident conclusion that can be drawn from the compilation of the output of Indian endeavor in the area of bioinformatics and Computational Biology have had a tremendous impact not only in terms of number of publications and but also in terms of quality of the research output. The growth has not only been steady but has even seen a very steep trajectory. Though we have not presented here the graphs in terms of quality of journals in which they were published mainly due to non-availability of an acceptable yardstick, the conventional yardsticks such as JCR impact factor and Scimago indices demonstrate that quality of publications also are quite impressive. The Bioinformatics publications from India can be found in any internationally acclaimed journals including those published by international societies and commercial publishers. We will be presenting those data in our next publications.

However, there are few elements which need attention of the scientists as well as the funding organizations. There is significantly lower number of papers from multiple organizations. Collaborations need to be strengthened among institutions, both within the BTISNET family as well

as among other organizations. Continuity of work also has to be maintained. There are subject areas where some good publications are found but the works continuation is not exhibited in further publications of the research group.

Overall, the Bioinformatics and Computational Biology in India is at par with any part of the world and the funding by government agencies have yielded desired results.

Acknowledgement

Support in compilation and analysis by Ms Akanksha Shukla and Neha Verma, Apprentice Trainees at KRC, CSIR-CDRI is highly appreciated. We acknowledge the BTIS project of the Department of Biotechnology for providing infrastructure and grants including Project Fellowship to the first three authors, AS, AS and A. CDRI Communication Number 9566.

Abbreviations

DISC, Distributed Information Sub-Centres; COE, Centre of Excellence; DIC, Distributed Information Centres; SDIC, Sub-Distributed Information Centres; BIF, Bioinformatics Facilities; DBT, Department of Biotechnology; BTIS, Biotechnology Information Systems; BTISNET, Biotechnology Information Systems Network; JCR, Journal Citation Reports

References

- Bansard, J.Y., Rebholz-Schuhmann, D., Cameron, G., Clark, D., Van Mulligen, E., Beltrame, E., Barbolla, E., Martin-Sanchez Fdel, H., Milanesi, L., Tollis, I., Van der Lei, J. and Coatrieux, J. L. (2007). Medical informatics and bioinformatics: A bibliometric study. *IEEE Trans Inf Technol Biomed.* 11, 237-243.
- Bartlett, A., Penders, B. and Lewis, J. (2017). Bioinformatics: indispensable, yet hidden in plain sight? *BMC Bioinformatics.* 18, 311.
- Benton, D. (1996). Bioinformatics - principles and potential for a new multidisciplinary tool. *Trends Biotechnol.* 14, 261-276.
- Manoharan, A., Kanagavel, B., Muthuchidamaram, A. and Kumaravel, J.P.S. (2011). Bioinformatics Research - An Informetric View International Conference on Information Communication and Management IPCSIT vol.16, IACSIT Press, Singapore.
- Curioso, W. H., Hansen, J. R., Centurion-Lara, A., Garcia, P. J., Wolf, F. M., Fuller, S., Holmes, K. K. and Kimball, A. M. (2008). Evaluation of a joint Bioinformatics and Medical Informatics international course in Peru. *BMC Med Educ* 8, 1.
- Gopal Manya, R. and Sudhier, K.G. P. (2015). A quantitative analysis of bioinformatics literature in India. *Inf. Stud.* 21, 93-110.

Hurrell, S (2004). Data Collection Challenges. Case Study Projects 1-12.

Krishnaswamy, S. and Madhan Mohan, T. (2016). The largest distributed network of bioinformatics centres in the world: Biotechnology Information

System Network (DBT-BTISNET). Curr. Sci. 110, 556-561.

Madhani, P.M. (2011). Indian Bioinformatics Growth Opportunities and Challenges. 'PRERANA: J Manag Thought Pract. 3, 7-17.