A stepping-stone to developing bioinformatics in Pakistan

Shahid Manzoor[1], Adnan Niazi[2], Erik Bongcam-Rudloff[3]

[1]Department of Information Technology, University of the Punjab, Lahore, Pakistan; [2]Department of Immunology, Genetics and Pathology, Science for Life Laboratory Uppsala, Uppsala University, Uppsala, Sweden; [3]SLU Global Bioinformatics Centre, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden.

Introduction

Bioinformatics employs a wide range of 'informatics' techniques to analyse and extract information associated with large-scale biological data. *In silico* tools and methods have been pivotal for DNA and protein analysis, including those for sequence translation and alignment, gene finding, gene annotation, protein structure prediction and phylogenetic reconstruction. Advances in computational and molecular biology research, and application of high-throughput next-generation sequencing technologies in areas such as genomics, transcriptomics and proteomics, has taken bioinformatics to an even higher level (Cole *et al.*, 2014); (Wang and Zhang, 2013). Academic groups, research consortia and industries worldwide are intensively using bioinformatics as a tool to address life science related questions. This field of study is relatively new in Pakistan: it was introduced in 2002, when Muhammad Ali Jinnah University, now Capital University of Science and Technology (CUST), took the initiative and started an undergraduate degree programme in bioinformatics for the first time. Later, a similar programme was introduced by the COMSATS Institute of Information Technology (CIIT), another renowned university in Pakistan and member of EMBnet since 2006. HEC and the aforementioned universities played a vital role in developing and promoting awareness of bioinformatics, and the importance of education and training in this field, among scientific communities in Pakistan. Consequently, several universities launched similar programmes at both graduate and postgraduate levels.

Importance of bioinformatics in Pakistan

Bioinformatics cannot be overlooked in a country like Pakistan because of its unique genetic resources, in terms of human population and biodiversity of crops and animal species. Pakistan produces a large variety of agricultural products, such as cotton, wheat, rice, sugarcane, fruit and vegetables, in addition to cattle and poultry. Furthermore, its geographical features, and the presence of various ethnicities with familial and social characteristics in a population of over 200 million, is a valuable resource for the study of genetic disorders, such as Down syndrome, Fragile X syndrome, intellectual disability, psoriasis, schizophrenia, deafness, Alzheimer's disease, albinism and epilepsy.

Bioinformatics is widely practiced within the pharmaceutical industry in the development of health-care products, and also in agriculture and environmental protection (Lyall, 1996); (Xue and Zhao, 2008). The growth of the pharmaceutical industry demands advanced tools and methods for the discovery of drug targets (Fagan and Swindells, 2000), for drug design (Kelly and Clark, 2003) and for the identification of new disease markers to improve early diagnosis and develop new therapeutic strategies. Pakistan is aiming to raise its research standards in agricultural, biotechnological and biomedical sectors by exploiting bioinformatics approaches in many life-science domains.

A step forward to improve bioinformatics in Pakistan

To meet the demand for bioinformatics-oriented research and education in Pakistan, a key step was taken by the HEC through the approval of an "*Overseas scholarship for MS/MPhil leading to PhD in bioinformatics*" programme in 2006. The HEC offered 50 scholarships in bioinformatics (Ilyas and Sadique, 2011), aiming to create a critical mass of highly qualified researchers. The inclusion of this specialised human resource in bioinformatics in major research organisations and institutes aimed to boost research activities of Pakistan, and to allow the development of new projects with great economic returns. To this end, a collaboration was established with our Swedish colleagues during the EMBnet Annual General Meeting in Torremolinos (ES) (Rahman and Chohan, 2010). Scholarships were assigned to talented graduate students in selected fields of science. Applicants were shortlisted and called for interview after a graduate-assessment test. In 2008, an expert panel (comprising Erik Bongcam-Rudloff, Shahid Nadeem Chohan, Raheel Qamar and a representative from the HEC) nominated

Reports

10 candidates for higher studies at SLU. This was the first time (2009) that SLU had set up a bioinformatics MSc course with students selected in this way. All students started PhD programmes after completing their MSc studies. These scholars were awarded PhD degrees after successful public defence of their doctoral theses; the research topics of these theses are listed in Table 1.

Table 1. List of PhD theses under the programme.

Title of the project	Main Supervisor	Co Supervisor(s)
Bioinformatics screening for candidate mutations underlying phenotypic traits in domestic animals. ISBN 978-91-576-8132-4	Prof. Göran Andersson	Prof. Leif Andersson, Prof. Kerstin Lindblad-Toh, Prof. Erik Bongcam-Rudloff
Bioinformatics analysis of bacterial pathogens from East African camels. ISBN 978-91-576-8242-0	Prof. Erik Bongcam-Rudloff	Dr. Etienne de Villiers & Dr. Richard Bishop (Kenya)
Computational and comparative investigations of syntrophic acetate-oxidising bacteria (SAOB). ISBN 978-91-576-8060-0	Prof. Erik Bongcam-Rudloff	Prof. Anna Schnürer, Dr. Bettina Müller
Bioinformatics studies on the mechanisms of gene regulation in vertebrates. ISBN 978-91-576-8112-6	Prof. Göran Andersson	Prof. Leif Andersson
In silico analysis of Treponema and Brachyspira genomes. ISBN 978-91-576-8240-6	Prof. Erik Bongcam-Rudloff	Dr. Anna Rosander, Dr. Desiree Jansson
Mapping and functional characterisation of candidate genes and mutations for chicken growth. ISBN 978-91-576-8046-4	Prof. Örjan Carlborg	Dr. Stefan Marklund, Dr. Anna Johansson
Genome-wide analyses of *Bacillus amyloliquefaciens* strains provide insights into their beneficial role on plants. ISBN 978-91-576-8080-8	Prof. Erik Bongcam-Rudloff	Prof. Johan Meijer, Dr. Sarosh Bejai
Bioinformatics analysis of whole genome sequencing data. ISBN 978-91-576-8064-8	Prof. Leif Andersson	Prof. Lars Rönnegård, Prof. Erik Bongcam-Rudloff
Towards High-Throughput Phenotypic and Systemic Profiling of *in vitro* Growing Cell Populations using Label-Free Microscopy and Spectroscopy: Applications in Cancer Pharmacology. ISBN: 978-91-554-9082-9	Prof. Mats Gustafsson	Dr. Mårten Fryknäs, Dr. Ulf Hammerling
Integrated Computational and Experimental Approaches for Accelerated Drug Combination Discovery and Development: Applications in Cancer Pharmacology. ISBN: 978-91-554-9177-2	Prof. Mats Gustafsson	Prof. Rolf Larsson, Dr. Claes Andersson

The scholars also published more than 40 research articles in peer-reviewed journals during their PhD programmes (Figure 1).

The success of this education scheme had a wide resonance in the country, and several institutes subsequently launched degree programmes in bioinformatics. Currently, around 20 universities and research institutes are offering bioinformatics programmes at undergraduate and postgraduate levels (see Table 2).

These educational programmes will provide Pakistan with specialised human resources able to improve its research capability and its competitiveness at an international level.
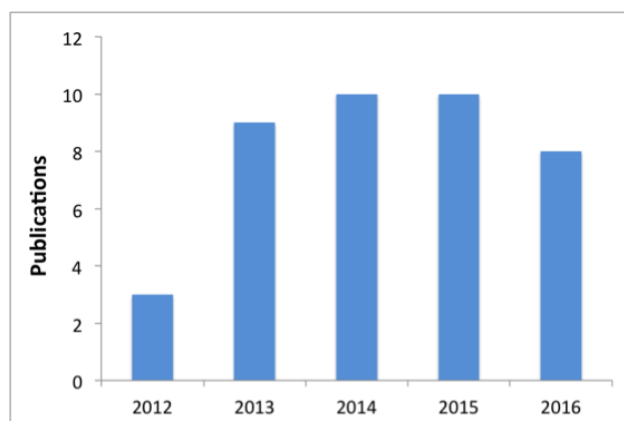

Figure 1. Number of articles published by scholars of the programme..

EMBnet.journal
Bioinformatics in Action

Reports

Table 2. Institutes offering education programmes in bioinformatics in Pakistan.

Sr.#	Institute	City	Programme
1	Baqai Medical University	Karachi	BS
2	Capital University of Science and Technology	Islamabad	BS/MS
3	Comsats Institute of Information Technology	Islamabad	BS/MS
4	Comsats Institute of Information Technology	Sahiwal	BS
5	Federal Institute of Health Sciences	Lahore	BS
6	Forman Christian College	Lahore	BS
7	Government College University	Faisalabad	BS/MS
8	Government Postgraduate College Mandian	Abbottabad	BS
9	Hazara University	Mansehra	BS/MS
10	International Islamic University	Islamabad	BS/MS
11	Khushal Khan Khattak University	Karak	BS
12	National University of Sciences and Technology	Islamabad	MS
13	Qarshi University	Lahore	BS
14	Quaid-e-azam University	Islamabad	BS/MPhil/PhD
15	Shaheed Benazir Bhutto Women University	Peshawar	BS/MPhil
16	Sir Syed University of Engineering & Technology	Karachi	BS
17	The Superior University	Lahore	BS
18	University of Agriculture	Faisalabad	BS
19	University of Sindh	Jamshoro	MPhil
20	Virtual University of Pakistan	Lahore	BS/MS

Conclusions

This joint venture between Pakistan's HEC and SLU (SE) was a stepping-stone to expand the scope of bioinformatics in educational and research centres in Pakistan. In addition, it provided important international networking opportunities within the field. Similar efforts in the future will allow Pakistan to augment the pool of skilled bioinformaticians to meet its challenging needs in different research fields. The Masters and PhD programmes developed via this initiative are also being used as a model for bioinformatics education in several countries in Asia and Africa.

Acknowledgments

References

1. Cole C, Krampis K, Karagiannis K, Almeida JS, Faison WJ *et al.* (2014) Non-synonymous variations in cancer and their effects on the human proteome: workflow for NGS data biocuration and proteome-wide analysis of TCGA data. *BMC Bioinformatics* **15**, 28. http://dx.doi.org/10.1186/1471-2105-15-28

2. Fagan R, Swindells M (2000) Bioinformatics, target discovery and the pharmaceutical/biotechnology industry. *Curr Opin Mol Ther* **2**, 655–661

3. Ilyas M, Sadique S, Masood K, Qamar R, Chohan SN (2011) The development of computational biology in Pakistan: still a long way to go. *PLoS Comput Biol* **7**, e1001135. http://dx.doi.org/10.1371/journal.pcbi.1001135

4. Kelly DE, Clark A (2003) Modern approaches to drug discovery and design: setting the scene. *Biochem Soc Trans* **31**, 428. http://dx.doi.org/10.1042/bst0310428

5. Lyall A (1996) Bioinformatics in the pharmaceutical industry. *Trends in Biotechnology* **14**, 308–312

6. Rahman N, Chohan SN (2010) Pakistan EMBnet node: progress report. *EMBnet.news* **15**(4), 35–36. http://journal.embnet.org/index.php/embnetnews/article/view/48

7. Wang X, Zhang B (2013) customProDB: an R package to generate customized protein databases from RNA-Seq data for proteomics search. *Bioinformatics* **29**, 3235–3237. https://dx.doi.org/10.1093/bioinformatics/btt543

8. Xue J, Zhao S, Liang Y, Hou C, Wang J (2008) Bioinformatics and its Applications in Agriculture. In Li D (ed.), *Computer And Computing Technologies In Agriculture* **2**, 977–982. http://dx.doi.org/10.1007/978-0-387-77253-0_29