

Infrastructural Needs and Capacity Building in Bioinformatics for Sustainable Development in Developing Countries

Abad Shah

Research & Development
Center of Computer Science
University of Engineering and
Technology
Lahore, Pakistan.

Khadim H. Asif

Research & Development
Center of Computer Science
University of Engineering and
Technology
Lahore, Pakistan.

Syed Ahsan

Research & Development
Center of Computer Science
University of Engineering and
Technology
Lahore, Pakistan.

Abstract. *Bioinformatics is the field dedicated mainly to the development and use of computational methods and tools for the understanding of biological data and information, and holds great promise for poverty alleviation and disease control. Developed countries have invested generously in the research and development of bioinformatics. Developing countries must follow suite. These countries should setup the infrastructure for the research and development of bioinformatics and related areas. To achieve this, some of the steps, actions and the required resources are: i) availability of vast DNA information in the public databases, ii) a number of computational tools freely available either on-line or as an open source, iii) to train human resource and capacity building, iv) setting up of bioinformatics resource facilities, and v) collaboration with developed countries. In this paper we describe the need and the required actions to motivate bioinformatics research and development in the developing countries.*

Keywords: Biotechnology, Disease control, Genomic Gap, Biodiversity, Capacity Building

1 Introduction

In the last 15 years, there has been a tremendous growth in the amount of biological data available to biological scientists and researchers. This has been possible due to the huge funding this emerging field has attracted from various government and private agencies in various developed countries all over the world and also because of advancement in hardware Technology such as Micro Arrays [3] Bioinformatics cannot be disregarded by any country intending to remain up-to-date in the biomedical, biotechnological and agricultural sectors. The rapid growth of genetics and its application provide an unparalleled opportunity to advance the living standards of all. We however feel that developing countries will lose out on the benefits of genomics research unless action is taken to strengthen their participation in bioinformatics. Developing countries should take

initiatives to manage their own specific data on indigenous biological species, on local epidemiology and biodiversity programs. The relevant advantages of these initiatives for developing countries would be technology transfer and capacity building. Consequently, these initiatives could be critical to the development of the respective countries. In the first section of the paper, we identify the stakes which the developing countries have in bioinformatics. The remaining paper suggests various measures which must be adapted to facilitate bioinformatics research and development in these countries.

2 Improving Animal Production and Health in Developing Countries

Biodiversity, which includes all plants, animals, and microorganisms, can be measured and expressed in different units such as genes, individuals, populations, species, ecosystems, communities, and landscapes. The world's greatest biodiversity is mostly retained in developing countries located in the tropics and includes large numbers of etiological agents of infectious and parasitic diseases whose impact on both human and animal health has been significant [11]. Because of the number of species involved and the relatively low level of exploration of biodiversity, several organisms are still waiting to be discovered and explored in terms of genetics and genomics, mainly for a better understanding of our own natural genetic resources. Several genome projects have been started and completed in recent years and those that have targeted microbes are the most abundant, mainly because the relatively small genomes of these organisms facilitate sequencing of the whole genome in a short time at large sequencing centers [11]. Consequently, eukaryotic genomes that are larger require significantly more time, manpower, and budgets. Specifically in the area of livestock, cattle, pigs, chicken, sheep, horses, rabbits, turkeys, and aquaculture, species have differing levels of ongoing genome initiative. While these represent some of the most important species with respect to impact

on the livestock industry, many species remain to be studied in a genomic context. Concerning human and animal pathogens, several efforts have been directed at the sequencing of the genomes of *Trypanosoma* spp., *Leishmania major*, *Plasmodium* spp., *Schistosoma* spp., *Theileria* spp., *Toxoplasma gondii*, *Eimeria tenella*, *Giardia lamblia*, *Entamoeba* spp., *Brugia malayi*, *Mycobacterium* spp., and *Candida albicans*. A few other examples of species with genome sequencing initiatives are *Theileria annulata*, *Cowdria (Ehrlichia) ruminantium*, and *Babesia bovis*. Regarding vectors of parasitic diseases, one of the most relevant species has been sequenced now - *Anopheles gambiae*. Meanwhile, *Aedes aegypti*, *Glossina morsitans morsitans*, *Culex pipiens*, and *Amblyomma americanum* genomes are in different sequencing stages. Initiatives for Triatominae species, vectors of human trypanosomiasis, are just starting, and a project on the transcriptome of some tissues of *Lutzomyia longipalpis* being funded by the Brazilian Council for Research and Development [3]. Again, these represent very important species, but many others, affecting especially the developing countries, remain to be studied. They appear unlikely to enter the genomics area, because they affect typically poor countries where there is neither a market nor a profit for drug and vaccine development by private companies. These diseases have been called “neglected diseases,” and an initiative “Drugs for Neglected Diseases initiative” to target those specifically affecting human health was recently begun. Considering this, one important question to ask is, “who is going to explore the genetics and genomics of relevant livestock species and pathogens that are restricted to developing countries, with no apparent relevance to developed countries?” The most probable answer is “nobody,” unless scientists in developing countries receive training, financial support, and capacity building to do this for themselves. Otherwise, the problem will increasingly affect their livestock or humans.

3 Agricultural Applications of Biotechnology and the Potential for Biodiversity

Agriculture in the twenty-first century is facing unprecedented challenges. An additional 2 billion people will have to be fed over the next 30 years from an increasingly fragile natural resource base. More than 842 million people are chronically hungry, most of them in rural areas of poor countries, and billions suffer from micronutrient deficiencies, an insidious form of malnutrition caused by the poor quality of, and lack of diversity in, their habitual diet (Genetics in developing countries, www.who.int/genomics). The Green Revolution taught us that technological innovation - higher-yielding seeds and the inputs required to make them grow - can bring enormous benefits to poor people

through enhanced efficiency, higher incomes and lower food prices. This virtuous cycle of rising productivity, improving living standards and sustainable economic growth has lifted millions of people out of poverty.

There is clear promise that biotechnology can contribute to meeting these challenges. Biotechnology can overcome production constraints that are more difficult or intractable with conventional breeding. It can speed up conventional breeding programs and provide farmers with disease-free planting materials. It can create crops that resist pests and diseases, replacing toxic chemicals that harm the environment and human health, and it can provide diagnostic tools and vaccines that help control devastating animal diseases. It can improve the nutritional quality of staple foods such as rice and cassava and create new products for health and industrial uses [7].

Agriculture and biodiversity are closely linked to biotechnology; the sustainability of protected biodiversity areas depends to a large extent on what transpires in intervened adjacent areas. Sound agricultural intensification can prevent further deterioration of protected landscapes. In traditional agricultural areas, modern biotechnology offers an opportunity not only to increase productivity and significantly reduce the use of off-farm chemical inputs, but also to enhance natural agro-biological systems. Biotechnology has now become a preferred tool for adding value to biodiversity by allowing more effective identification and utilization of genes and derived products [10]. Recent advances in genomic, proteomic and metabolomic research offer unprecedented opportunities for the search, identification, and commercial utilization of biological products and molecules in the pharmaceutical, nutraceutical, agricultural, and environmental sectors. This is crucial in a region where the agro-industry sector contributes up to 60% to the internal gross product (IGP), where more than half of the world's population resides, and where there is also a growing demand for food quantity and quality and mitigation of environmental degradation [8].

The developing world concentrates major biodiversity hotspots of the world. The region is also a center of origin and diversity of a number of species that sustain current world food supply (e.g., rice, wheat, potato, sweet-potato, corn, tomato, beans, cassava, peanuts, pineapple, cacao, chili pepper, and papaya). Furthermore, the greatest number of flowering plants with unusual sources of compounds for food and agriculture, as well as for the biopharmaceutical, nutraceutical, cosmetic, and environmental industries, exists in this region. This comparative advantage needs to be explored and sustainability utilized by integrating the emerging biotechnologies with the region's rich traditional knowledge on the properties and attributes of its biological resources. The world market for biological

resource-derived ingredients and molecules in 1999 was US\$925 billion for different sectors and industries [12].

4 Improving Health and Disease Control through Drug and Vaccine Discovery

Millions of people in developing countries die each year from diseases that could be prevented or treated cheaply and easily through the development and use of a handful of biotechnologies. Knowledge from man and pathogen genomes can revolutionize healthcare and result in the development of new vaccines and drugs that could benefit millions [4].

Infectious and parasitic diseases represent a great threat to the economic survival of the poorest developing countries. For example malaria affects the health and wealth of nations and individuals alike. In Africa today, malaria is understood to be both a disease of poverty and a cause of poverty. Malaria has significant measurable direct and indirect costs, and has recently been shown to be a major constraint to economic development. For developing economies this has meant that the gap in prosperity between countries with malaria and countries without malaria has become wider every single year [1].

Annual economic growth in countries with high malaria transmission has historically been lower than in countries without malaria. Economists believe that malaria is responsible for a 'growth penalty' of up to 1.3% per year in some African countries [8]. When compounded over the years, this penalty leads to substantial differences in GDP between countries with and without malaria and severely restrains the economic growth of the entire region [12]. The direct costs of malaria include a combination of personal and public expenditures on both prevention and treatment of the disease. Personal expenditures include individual or family spending on insecticide treated mosquito nets (ITNs), doctors' fees, anti-malarial drugs, transport to health facilities, support to the patient and sometimes an accompanying family member during hospital stays. Public expenditures include spending by government on maintaining health facilities and health care infrastructure, publicly managed vector control, education and research. In some countries with a heavy malaria burden, the disease may account for as much as 40% of public health expenditure, 30-50% of inpatient admissions, and up to 50% of outpatient visits [12]. The indirect costs of malaria include lost productivity or income associated with illness or death [8]. This might be expressed as the cost of lost workdays or absenteeism from formal employment and the value of unpaid work done in the home by both men and women. In the case of death, the indirect cost includes the discounted future lifetime earnings of those who die.

To implement effective public health policies to deal with malaria, it is essential for the developing countries that are the most affected to understand the biology of the disease. Although *Plasmodium* research is being done in developed and developing countries alike, but without a critical mass of qualified people, proper regulation, and ethical guidelines in developing countries, the research could actually widen the gap between rich and poor. In spite of the huge investments in drug discovery and disease control, the bulk of genomics research is not directed at the diseases effecting developing countries (Genetics in developing countries, www.who.int/genomics). The reason behind this is that research is largely market driven by market forces and its products are probably unaffordable to the poor living in the developing countries. This discrepancy risks the creation of a "genomics divide" between rich and poor countries, such that 90% of those with health problems will not benefit from what is arguably the most golden era of biology and medicine [12].

An international panel of experts, including many from developing countries, recently ranked the top 10 biotechnologies most likely to impact in the next 5 to 10. These promising cutting-edge technologies for health in developing countries are [13].

- i) Easy-to-use molecular diagnostic tests for TB, hepatitis C, HIV-AIDS, malaria and other diseases, which detect the presence or absence of pathogen-associated molecules, such as DNA or protein, in a patient's blood or tissues;
- ii) Recombinant vaccines against infectious diseases, produced through genetic engineering, which promise to be safer, cheaper and possibly easier to store and transport than traditional vaccines;
- iii) Reducing pollution and making water safe to drink through bioremediation – the potential exploitation of micro-organisms with remarkable biochemical properties;
- iv) Creating microbicides for female-controlled protection against sexually transmitted diseases like HIV, both with and without contraceptive effect;
- v) Better drug and vaccine delivery methods that avoid the use of needles and reduce cross contamination;
- vi) Bioinformatics to identify drug targets and to examine pathogen-host interactions;
- vii) Nutrition-enriched crops to counter specific deficiencies, such as vitamin A-rich "Golden Rice" to improve health for millions without a balanced diet;

viii) Sequencing pathogen genomes to understand their biology and identify new antimicrobials;

ix) Recombinant technology to make therapeutic products (e.g. insulin, interferons) more affordable to help fight such diseases as diabetes, now emerging as a major public health problem throughout the world;

x) Combinatorial chemistry for drug discovery.

According to the same authors, some experts have suggested the formation of an international commission to govern the application of biotechnology. However, an initiative originating in the developed world may take the form of just another international commission, with token participation from only a few scientists from developing countries, and may not be the best way forward. If not properly harnessed, the experts warn, the breakthroughs being produced through genomics will dramatically widen rather than narrow the gap between developed and developing countries. A more acceptable approach may be to truly involve and empower scientists and institutions in developing countries, working together in equal partnership [7].

An international initiative should also involve scientists and institutions from the more genomically advanced developing countries such as Brazil, China, Malaysia, and Thailand, which have made significant inroads and are shining examples of national commitment to the potential of genomics. Modern biomathematics research does not require more resources than any other field of computer science; almost all processes can be efficiently designed and modeled on a personal computer or workstation. Public bioinformatics resources, such as databanks and software tools that are crucial for biotechnology projects, are today available via the internet. Scientists need only a computer and an internet connection of a certain quality to use them. If these conditions exist, the situation of a developing country biologist is no different than that of an academic biologist in an industrialized country. However, the step from theoretical bioinformatics to applied bioinformatics, intending to produce software from an algorithm, is not an easy one. It requires a supportive R&D climate that generates a local need for such research, and an appropriate computer science infrastructure.

5 Building Bioinformatics Capacity

Access to the results of bioinformatics, that is, teaching the users to understand bioinformatics, is a general problem, and in this respect expertise is the bottleneck. Of the computational solutions that are available, many are underutilized because few biologists are trained in computer science, and therefore, are unable to utilize

complex computational methodologies. Similarly, computer scientists typically are unaware of the biology problems that need to be solved and/or do not understand the biology problems well enough to know what software tools and technologies could be applied. As a consequence, much of the vital new biological information goes unused [5]. The first challenge is to teach the fundamentals of bioinformatics to university students, which is even more complicated since senior professors and science managers are often not yet familiar enough with the methodology, not only in developing countries. Unfortunately, a major obstacle in training the next generation of bioinformaticians is that few biology students are interested in computer science, and few computer science students are attracted to biology [9]. This problem is exasperated by an inherent difficulty in opening channels of communication across these traditionally divergent academic cultures. We suggest a collaborative approach in teaching bioinformatics to university students. In this approach, we refer to the development of a prototypical process that includes: i) biologists and computer scientists working side-by-side on site to solve problems, ii) biologists leading the formation and testing of a hypothesis, with computer scientists observing and assisting, iii) computer scientists developing software tools with the input of the biologists, and, iv) computer scientists and biologists learning about the access, design, and development of databases and Web sites targeted at users with little to no computer training [9].

The second source of difficulty is a problem of building and maintaining capacity. There are few university-level bioinformatics curricula worldwide, and very few in developing countries. Moreover, graduates are often immediately absorbed by the prestigious universities, or pharmaceutical or agri-biotech multinationals. The same problem, however, plagues most European and American universities.

The same problem, however, plagues most European and American universities. Furthermore, instead of maintaining bioinformatics research groups, many universities choose to support bioinformatics teaching in the general framework of biological curricula, using knowledgeable user-level teachers involved in other areas of biological research at the same university.

In the USA, this will change with new initiatives from the NIH and the Howard Hughes Medical Institute to establish 20 national bioinformatics programs of excellence. In Europe pharmaceutical companies have founded an industrial collaboration centered on the European Bioinformatics Institute (EBI) in Hinxton, UK, that helps them train their own bioinformaticians. The UK Research Councils coordinate their efforts in creating new training and research facilities. The first programs

specifically addressing the bioinformatics needs of developing countries was initiated in 1990 by the International Centre for Genetic Engineering and Biotechnology (ICGEB, see box). An estimated US\$ 200,000 per year is spent in these programs to establish ICGEBnet, a high level computing environment freely available for developing country scientists, as well as for theoretical and practical courses in several disciplines, including bioinformatics [13]. To date ICGEBnet has served over 1300 registered users and a comparable number of course participants from Asia, Africa, South America and (mostly Central and Eastern) Europe. A European initiative, the European Molecular Biology Network (EMBnet), was organized in 1989 to help the spread of bioinformatics data and techniques in the European Union. Although EMBnet is mainly active in Europe, it recently adopted some non-European countries including Argentina, China, India and South Africa as members, and has sponsored several courses, partly in collaboration with the ICGEB, outside Europe. EMBnet can therefore be considered a worldwide network of centers of bioinformatics expertise that spreads bioinformatics knowledge through a local infrastructure and local training courses. Such centers can be established at university campuses, a practice that could be recommended to developing countries. Other programs include a recent initiative of the International Council of Scientific at Israel's Weizmann Institute aimed at organizing courses in Middle Eastern countries. There is a growing amount of teaching material on the internet itself, and one can expect that the availability of the internet will help to compensate for the missing training course

6 Tools Needed and the Role of Web

While all branches of modern biology make some use of computers, molecular biology and especially genetic engineering could not even exist without them. Given a certain research and computer infrastructure, developing countries may have relatively easy access to the products of bioinformatics. Bioinformatics produces databanks such as collections of protein sequences for biology and biotechnology, as well as computer software for analysis. Users such as biotechnologists and academic biologists may choose to install these components on a local computer system, or access them over the internet, using the publicly available databanks.

There are number of useful resources / tools available freely on the web. Sequence and other non-bibliographic databases are the central, most important type of information resource in this field.

There are hundreds of databases that might be considered relevant to bioinformatics. There are specialized databases for each species, and separate databases for different types of information (nucleic acid sequences, protein sequences, protein structures, biochemical and biophysical information, etc.). There is also a great redundancy of databases, with multiple databases covering nearly the same information for the same organisms. This situation arose in part from many researchers developing their own databases in their own formats over the years, and from databases developing in parallel in Europe, Japan and the United States. The situation is further complicated by the existence of several versions or mirrors of the same database on different servers (each with varying degrees of currency or completeness), and by the sharing of records between databases. For example, the Entrez search system draws data from SWISS-PROT but only includes SWISS-PROT records for proteins that are based upon nucleotide sequence data that meet the criteria for inclusion in GenBank. A search of SWISS-PROT through another interface may retrieve more records as well as more

Users such as biotechnologists and academic biologists may choose to install these components on a local computer system, or access them over the internet, using the publicly available databanks. However, their future use of this technology hinges on the availability of bioinformatics knowledge in the public domain. For this reason, developing countries may also want to manage their own specific data on indigenous biological species, on local epidemiology and biodiversity programs. These tasks clearly require that statisticians and informatics experts become advanced users of bioinformatics software and develop a capability to solve problems locally. This process does not require large resources itself but will allow developing countries to further investigate their own biological resources.

7 North- South and South- South Collaboration

It is widely believed that scientists in developing countries will be able to significantly contribute to the progress of this field since the capital investment needed for bioinformatics research may be much smaller than for experimental biological sciences. However, the advancement of bioinformatics science in developing countries requires not only competent human resources, but also good ideas and problems. Without good communication with state-of-the-art experimental laboratories, there is a clear danger that small bioinformatics teams in developing countries may try to address irrelevant questions. Without good communication with state-of-the-art bioinformatics laboratories, there is a danger of using suboptimal technology to address the problems [8]. To avoid these

dangers, there must be forums where bioinformatics scientists in developing countries interact with their counterparts in developed countries, as well as with biologists from various disciplines. This can be done through scientific publications as well as meetings and seminars. There are several good international bioinformatics conferences. The fact that they are mostly organized in the developed world limits the access of scientists from developing countries. Moreover, there are problems that are of urgent importance for developing countries, such as those relating to agriculture and biodiversity.

A more acceptable approach may be to truly involve and empower scientists and institutions in developing countries, working together in equal partnership. An international initiative should also involve scientists and institutions from the more genomically advanced developing countries such as Brazil, China, Malaysia, and Thailand, which have made significant inroads in bioinformatics.

Moreover, intra- and inter-hemispheric scientific collaboration should be encouraged and supported to increase the chances for success. In Brazil, the Ministry of Science and Technology has stepped forward to further such initiatives, co-supporting collaborative genomics and bioinformatics projects [2]. The need for the establishment of working groups on genomics and bioinformatics in developing countries as well as the improvement and strengthening of collaborative research projects between developed and developing countries.

8 Conclusion

Bioinformatics and Biotechnology holds the key to progress and prosperity in developing countries and is indeed the enabling technology for several fields of biomedical and agricultural research. Unfortunately, the research infrastructure for bioinformatics is almost non-existent in these countries which most need it. In this paper we outlined the benefits developing countries can reap many times over by investing in bioinformatics and biotechnology research. To achieve this, these countries need to establish an infrastructure and an R&D environment. At present, bioinformatics has successfully been applied only in those developing countries where the infrastructural and resource requirements are met, such as Brazil, China, India, Thailand and Mexico.

Education is the key to developing biotechnology in developing countries and active research and commercialization of research are needed to promote sustainable development in this field. To facilitate this process bioinformatics / biocomputing should be introduced to universities, and the establishment of small software groups and companies should be encouraged.

These countries need to: i) Re-energize academic institutions and public sector research to strengthen their science base; ii) Train people to use, adapt and innovate biotechnologies; Encourage regional and international cooperation to exchange knowledge ; iii) Improve the policy environment (including intellectual property laws and regulation); iv) Foster private sector growth, encouraging it to address local health needs, and strengthening linkages between public and private sectors to create new biotechnology goods and services.

We strongly feel that the developed countries should do more to help developing countries attain self-reliance in bioinformatics. Most of the genome sequencing projects to date did not include the participation of developing country scientists. In spite of the fact that bioinformatics methods are freely accessible, and the use of bioinformatics spreads freely through the internet and it helps developing countries to catch up with industrialized countries, there is clearly a gap, a “genomic divide” between the developing and the industrialized world, which must be consciously narrowed. All this is based, however, on the principle that information resources worldwide remain freely accessible. If this should change in the future, it might widen the North-South gap in biotechnology. The most likely scenario to avoid or help bridge the genomics divide is the improved collaboration between scientists in developing and developed countries and the promotion of South-South collaboration, thus contributing towards the development of science in poor countries more in need of it.

Acknowledgement

We like to thank Higher Education Commission (HEC) Government of Pakistan for financial support of this research work.

References

- [1] African summit on roll back malaria, Abuja, Nigeria, 25th April 2000
- [2] Alberto M.R. Dávila^{a,C}, Mário Steindel^{b,C} and Edmundo C. Grisard^{b,C}, *Ann. N.Y. Acad. Sci.* 1026: 41–46 (2004). doi: 10.1196/annals.1307.005
- [3] Andrade, M.A. and Sander, C. (1997), "Bioinformatics: from genome data to biological knowledge." *Current Opinion on Biotechnology* No. 8(6), pp. 675-83.
- [4] Bradbury, E.M. and Pongor, S. (1999), *Structural Biology and Functional Genomics*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

- [5] Burley, S.K., Almo, S.C., Bonanno, J.B., Capel, M., Chance, M.R., Gaasterland, T., Lin, D., Sali, A., Studier, F.W. and Swaminathan, S. (1999) "Structural genomics: beyond the human genome project." *Nature Genetics*, No. 23(2), pp. 151-157.
- [6] Counsell, Damian. 2001. *Bioinformatics FAQ*. : <http://bioinformatics.org/faq/>, January 16, 2002
- [7] Doernberg, D. 1993. Computer Literacy Interview With Donald Knuth. http://www1.fatbrain.com/interviews/knuth_interview.html November 19, 2001. Genetics in developing countries, www.who.int/genomics
- [8] Harvard International Review, 'Enabling Genetics Promise in the Developing World'. *International Health*, Vol. 27 (1) - Spring 2005
- [9] Maglia and J. Leopold, "A Collaborative Approach to Teaching Bioinformatics Using Web-Accessible Resources and Biological Fieldwork", *Proceedings of World Conference on Educational Multimedia, Hypermedia, and Telecommunications (ED-MEDIA 2003)*, Honolulu, Hawaii, June 2003, pp. 2972-2976, (winner of conference Best Work Award).
- [10] Prasit Palittapongarnpim I U P A C *Pure Appl. Chem.* Vol. 74, No. 6, pp. iv (2002) <http://www.utoronto.ca/jcb/>
- [11] Pongor, S. and Landsman, D. (1999), "Bioinformatics and the Developing World." *Biotechnology and Development Monitor*, No. 40, p. 10-13.
- [12] Press Release WHO/28 Executive summary for Economics of Malaria for Int'l Development, Harvard University and the London School of Hygiene and Tropical Medicine, April 25 2000.
- [13] Top 10 Biotechnologies for Improving Health in Developing Countries', University of Toronto Joint Centre for Bioethics. October 3, 2002.
- [14] Tara Acharya 1 , Nandini K Kumar 2 , Vasantha Muthuswamy 2 , Abdallah S Daar 1 and Peter A Singer Published:19 May 2004 *Health Research Policy and Systems* 2004,2 :1(available from:<http://www.health-policy-systems.com/content/2/1/1>).