African Journal of Biotechnology Vol. 3 (2), pp. 117-122, February 2004
Available online at http://www.academicjournals.org/AJB
ISSN 1684–5315 © 2004 Academic Journals

Perspective

Opportunities in Africa for training in genome science

Daniel K. Masiga[1]* and Raphael D. Isokpehi[2]

[1]International Centre of Insect Physiology and Ecology (ICIPE), P.O. Box 30772, 00100-Nairobi, Kenya.
[2]South African National Bioinformatics Institute (SANBI), University of the Western Cape, Bellville 7535, South Africa.

Genome science is a new type of biology that unites genetics, molecular biology, computational biology and bioinformatics. The availability of the human genome sequence, as well as the genome sequences of several other organisms relevant to health, agriculture and the environment in Africa necessitates the development and delivery of several types and levels of training that will enhance the use of genome data and the associated computational resources. A survey of initiatives that provide opportunities for training in genome science is presented. Current efforts to increase the ability of African scientists to computationally process and analyse genomic and post-genomic data have the potential to produce excellent scientists who perform cutting-edge, hypothesis-based research, and who will accelerate the continent's scientific and technological development.

Key words: Bioinformatics, computational biology, genome science, networking.

INTRODUCTION

The tremendous capacity for high-throughput determination and computational analysis of the sequence of nucleotide bases of the genetic material (DNA) or genome of organisms has led to the description of genomics or genome science as "a new way of looking at biology that unites genetics, molecular biology, computational biology and bioinformatics" (Gibson and Muse, 2002). Genome sequencing projects have resulted in the growth of computational resources for storage and analysis of diverse and complex data derived from these projects, as well as post-genomic studies such as parallel gene and protein expression profiling to understand genomic activity (Gibson, 2003; Hanash, 2003). Some of the organisms whose genomes have been sequenced are relevant to human health, agriculture and the environment in Africa (Gardner et al., 2002; Holt et al., 2002). This necessitates the development and delivery of several types and levels of training that will enhance the use of these computational resources as genome research tools by African scientists. One of the effects of the integration of biology and information technology is that it has provided access to genome data to anyone, anywhere with some training in bioinformatics.

Bioinformatics has become a central discipline in enabling researchers in a number of developing countries to make significant advances in various biological disciplines. However, in Africa, extremely wide disparities exist in human resources and infrastructure for access and utilization of genome data. Nonetheless, research and training in computational biology and bioinformatics has attracted much attention in developing countries because it does not require the huge capital and technological infrastructure required for genetics, biochemistry and molecular biology research (Oduola et al., 2002). Though the computational aspects of genome science considerably advance the process of generating hypotheses, there is no doubt that its growth in Africa, as elsewhere, needs to be accompanied by the infrastructure to test these hypotheses to discover new methods and tools that will address biomedical, agricultural and environmental problems.

There is an increased awareness that the benefits of genome science should reach the global community of peoples, rather than merely the most affluent, avoiding the 'genome divide' (Calva et al., 2002). According to Claire Fraser, the President and Director of The Institute

*Corresponding author. E-mail: dmasiga@icipe.org.

of Genome Research (TIGR), USA, “As scientists in developing countries increase their ability to acquire and use the fruits of the genomics revolution, they will become more important players in the scientific world” (Fraser, 2002). It is feared that if the challenge of developing tools and products to combat infectious diseases is left exclusively in the hands of market forces, or addressed by *laissez-faire* scientific and technological policies, genomics will increase the divide between the rich and the poor, instead of bridging it (Singer and Daar, 2001; WHO, 2002; Morel, 2003). The UNDP/World Bank/World Health Organization Special Programme for Research and Training in Tropical Diseases (TDR) has played a central role in efforts to bridge this gap, by initiating the generation of knowledge about several parasite genomes, and establishing regional training centres in bioinformatics (Oduola et al., 2002). The regional courses are designed to "train the trainer". The potential impact therefore goes way beyond those trained at the centres, hence expanding the network of scientists with training in bioinformatics. In addition, some examples of research outputs from such training courses have started to emerge (Isokpehi 2003; Isokpehi and Hide, 2003).

In writing this article, we hold the view that unless the entire global community obtains access to cutting-edge genome science, the incredible growth in available resources will have little impact on scientific and technological development, particularly in developing countries. Weatherall (2003) argues for the transfer of genomic technologies into developing countries to provide a base for future developments in genomics as well as building local capacities for monitoring genomic research. Some countries in the south have good infrastructure and have made considerable advances in genome science. An example is Brazil, where a government-funded consortium called Organisation for Nucleotide Sequencing and Analysis (ONSA) was formed to sequence and analyse the genome of the plant pathogen *Xylella fastidiosa* (Simpson et al., 2000), and has led to additional genome projects (Da Silva et al., 2002). Furthermore, the South-South Initiative for Tropical Disease Research (SSI-TDR; http://www.ssi-tdr.net) established by the TDR is an effort to encourage inter- and intra-continental collaboration among research groups in disease endemic countries. Genome science and its applications are a major thrust of the SSI-TDR. The above-mentioned efforts are significant, however more partnerships including North-South will need to be established to prevent applications of genome science becoming a means to widen the gap between the developed and developing world (Alwan and Modell, 2003).

The existence in Africa of national and continental societies of genetics, biochemistry and molecular biology indicates that several types and levels of training are available in the non-computational aspects of genome science. Thus, emphasis in this article is on opportunities for training in computational aspects of genome science in Africa. We also offer suggestions on how competence to perform cutting-edge hypothesis-based research facilitated by bioinformatics could be achieved.

OPPORTUNITIES FOR DEGREE TRAINING IN BIOINFORMATICS

There is not much information on the availability of undergraduate and postgraduate bioinformatics training in Africa. However, the South African National Bioinformatics Institute (SANBI), University of the Western Cape offers honours, masters and doctoral degree programmes in bioinformatics. The SANBI degree programmes accept students from outside South Africa. An increasing number of universities in Africa that offer courses in life sciences are now teaching bioinformatics modules or courses. It is important for substantial investment in defining the bioinformatics education curriculum for such courses (Pearson, 2001) in order to produce world-class scientists that can contribute positively to the development of genome science in Africa. Educators interested in starting or strengthening bioinformatics courses can benefit from workshops on education in bioinformatics organized by a special interest group of the International Society for Computational Biology (ISCB). The website for the fourth meeting is http://surya.bic.nus.edu.sg/web04/main.html. The recent establishment in South Africa of a government-supported national network of institutions, the National Bioinformatics Network (NBN) presents new opportunities for training African students. The NBN is discussed elsewhere in this article.

Collaborative research projects in genome science should be encouraged to include postgraduate student training in bioinformatics. Such collaborations could incorporate research mostly undertaken at home institutions, with registration and partial supervision with an institution (located anywhere in the world) able to provide the bioinformatics training. A list of bioinformatics and computational biology certificate and degree programmes worldwide is available from the International Society for Computational Biology website (http://www.iscb.org).

SHORT-TERM TRAINING WORKSHOPS

The establishment of a TDR regional training centre for bioinformatics and applied genomics, at SANBI in October 2001, has been a key catalyst in the growth of the number of African scientists with some basic training in bioinformatics. During the two courses held in 2002 and 2003, nearly 50 participants from countries representing all regions of Africa were trained. In 2003, additional funding from The Wellcome Trust facilitated a

training workshop on genome annotation using Artemis and the Artemis Comparison Tool (ACT) (http://www.sanbi.ac.za/mrc/artemis.html). The third TDR training course will take place in early February 2004.

A number of initiatives to hold regional, national and pre-conference courses are also in progress in other parts of Africa, some with support from staff of the regional training centre at SANBI. In 2002, the Howard Hughes Medical Institute, USA sponsored a workshop held in Kampala, Uganda and organised by the Med Biotech Laboratories, University of Pennsylvania and the U.S. National Academies. The West African Bioinformatics Training Workshop organised within the context of the West African Biotechnology Workshop Series (WABWS) took place at the University of Ibadan, Nigeria from 26 May – 7 June 2003. The WABWS event also included a Symposium on Bioinformatics Education, Training, Research and Development in West Africa. In May 2003, a one-week course was held at the University of Sfax, Tunisia. Furthermore, the 2003 meetings of the Biochemical Society of Kenya (BSK) in Nairobi and the Federation of African Societies of Biochemistry and Molecular Biology (FASBMB) in Yaounde had pre-workshops on bioinformatics. Table 1 is a list of some initiatives that run annual short-term courses in Africa with a substantial bioinformatics component as well as online resources that could be useful to African scientists interested in genome science.

ON-LINE TRAINING

Globally accessible Internet courses can constitute a key pathway for capacity building in genome science. Many institutions that provide courses make their materials available on-line. Perhaps the foremost organization providing on-line training in bioinformatics is the S-star alliance (http://www.s-star.org). Established as an alliance of six universities on five continents (Lim et al., 2003), this alliance has as its mission the provision of an open introductory course in bioinformatics anywhere, over the Internet. In 2003, the alliance grew to seven universities and increased the number of lectures from 12 to 14. Africa is represented by the University of the Western Cape, with Professor Winston Hide, the Director of SANBI, contributing a lecture on gene transcript analysis. In the first trial course held in 2001, 156 students participated. Out of 96 who followed through the entire course, 9 were from Africa (Lim et al., 2003). In 2002, there were 14 participants from Africa out of 97, while in 2003 Africa was represented by 7 out of 140 (Annette Badenhorst, personal communication). S-star aims to develop courses, with modules and evaluations that are approved by the educators and the institutions they represent. One of the aims is to develop an integrated modular learning environment that allows a student to select from both pre-requisite modules and advanced modules in order to build a comprehensive programme in genomics and bioinformatics. SANBI has incorporated the S-star course as a part of its masters degree programme. Improved internet connectivity and computer access as well as increased awareness of S-star courses at institutional levels could increase the number of students from Africa participating in this online resource.

NETWORKING AND PARTNERSHIPS

African Regional Postgraduate Programme in Insect Science

The African Regional Postgraduate Programme in Insect Science (ARPPIS) has been running at ICIPE for about 20 years, during which more than 200 postgraduates from nearly 30 participating universities in Africa have been trained. Albeit with substantial funding from outside the continent, such long-term functional units, with participants in disparate geographical areas demonstrate their utility and longevity. From these experiences, we can build capacity in genome sciences by networking. Although only a small proportion of students within ARPPIS currently benefit from training in genome science, the experience in long-term networking can facilitate considerable growth in this area among universities in Africa.

National Bioinformatics Network (South Africa)

The recent establishment in South Africa of the National Bioinformatics Network (NBN) (http://www.nbn.ac.za) is an excellent example of solid support from government to contribute to genome science capacity in Africa. The mission of the NBN is “to develop capacity in bioinformatics in South Africa, especially among disadvantaged groups, and to perform world-class bioinformatics research”. We propose that this is a model that can be replicated or remodelled in other African countries, and in some cases may be more suitable for regional networks. A feature of the NBN is that, although it is a bioinformatics network, the location of nodes at universities offers opportunities for collaborative research with university departments with capacity in different fields, such as genetics, biochemistry and molecular biology. Such collaborations offer the opportunity to test the results of computational experiments in the laboratory. The success of NBN will influence the development of bioinformatics capacity beyond South Africa. One way this is likely to happen is through the inclusion of postgraduate students and postdoctoral scientists from other African countries at NBN nodes. Additionally, the model presented by the NBN could be used to establish new networking initiatives that

Table 1. Some initiatives relevant to genome science in Africa.

Initiative
Training courses
Regional Training Course on Bioinformatics Applied to Tropical Diseases in Africa South African National Bioinformatics Institute, University of the Western Cape, South Africa http://www.sanbi.ac.za/tdrcourse tdrcourse@sanbi.ac.za
The Biochemical Society of Kenya Computational Biology (Bioinformatics) Training Workshop http://www.icipe.org/bionet/Workshop/index.htm
West African Biotechnology Workshops Series http://www.wabw.org info@wabw.org
Bioinformatics Training Course University of Sfax, Tunis
South African National Bioinformatics Institute Training Information http://www.sanbi.ac.za/mrc/Training.html
National Bioinformatics Network (South Africa) Courses http://www.nbn.ac.za/Education/courses.html
S-star Alliance Bioinformatics Education http://www.s-star.org
Societies/Networks
African Bioinformatics Network http://www.abionet.org
African Society of Human Genetics http://www.afshg.org
Federation of African Societies of Biochemistry and Molecular Biology http://www.fasbmb.org.za
National Bioinformatics Network (South Africa) http://www.nbn.ac.za
South-South Initiative for Tropical Disease Research http://www.ssi-tdr.net
Mailing Lists
SANBI Bioinformatics mailing List http://fling.sanbi.ac.za/mailman/listinfo/bioinformatics
ABioNET mailing list http://fling.sanbi.ac.za/mailman/listinfo/abionet
Access to On-line Journals
BioMed Central http://www.biomedcentral.com
PubMed Central http://www.pubmedcentral.nih.gov/
Health Internetwork Access to Research Initiatives (HINARI) http://www.healthinternetwork.org
International Network for the Availability of Scientific Publications (INASP) http://www.inasp.info/ajol/

transcend national borders, considering that some countries will have very few potential members.

TIGR and ILRI Collaboration

A collaborative effort between TIGR and the International Livestock Research Institute (ILRI) in Nairobi to sequence and annotate the genome of *Theileria parva* is considered a good model of how scientists can help bridge the North-South gap in genomic research by collaborating on projects of mutual interest (Fraser, 2002). This project has led to the development of good human and infrastructural resources at ILRI. Establishment of similar collaborations with national institutions have the potential to increase the capacities of developing countries to develop diagnostics, vaccines, and other tools to address unique problems that are considered of little economic value outside these countries. It is encouraging that programmes such as the TIGR International Travel Fellowships, which foster collaborations between TIGR and universities, governmental agencies or not-for-profit research institutes in Group 77 countries have been established.

African Bioinformatics Network

The African Bioinformatics Network (ABioNET) was formed during the first Regional Training Course on Bioinformatics Applied to Tropical Diseases in Africa held at SANBI in January 2002. ABioNET, which is now an affiliate of the International Society for Computational Biology, has as its key objective to "support and expand areas of bioinformatics expertise in Africa through human resource development, technical training, scientific exchanges, outreach and awareness at all levels of scientific manpower". To date ABioNET meetings have taken place during the TDR supported bioinformatics training workshop. The 2004 meeting will include lectures by selected African scientists on the application of bioinformatics in their research in tropical diseases. In addition, through the ABioNET mailing list, information on bioinformatics training and conferences has been distributed. The network needs to accelerate the establishment of its formal administrative structures and membership in order to function as the key bioinformatics network in Africa.

SUMMARY

There is a substantial need to provide opportunities in Africa for training in the ability to computationally process and analyse data generated from genome sequencing projects especially those that are relevant to human health, agriculture and the environment in the continent. Such training opportunities have the potential to produce excellent scientists who perform cutting-edge hypothesis-based research. A number of initiatives in Africa involving international funding agencies, government and not-for-profit institutions have made significant progress in providing several types and levels of training opportunities. However, increased commitment from African governments is required especially in the provision of infrastructure to establish and sustain continental, regional and national networks. Finally, African scientists should be proactive in seeking international collaborations in genome science.

ACKNOWLEDGEMENTS

We acknowledge the financial support of UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR) and thank colleagues who provide useful comments on the manuscript. The authors are recipients of a Collaborative Research Grant within the context of the TDR's South-South Initiative for Tropical Diseases Research. RDI is a Claude Harris Leon Foundation Fellow.

REFERENCES

Alwan A, Modell B (2003). Recommendations for introducing genetics services in developing countries. Nat. Rev. Genet. 4:61-68.

Calva E, Cardoso MJ, Gavilondo JV (2002). Avoiding the genomics divide. Trends Biotechnol. 20:368-370.

Da Silva AC, Ferro JA, Reinach FC, Farah CS, Furlan LR, Quaggio RB, Monteiro-Vitorello CB, et al. (2002). Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. Nature 417:459-463.

Fraser C (2002). Genomics for North-South collaboration. http://www.scidev.net/Opinions/index.cfm?fuseaction=readOpinions&itemid=75&language=1 Accessed January 19, 2004

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, et al. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419:498-511.

Gibson G (2003). Microarray analysis: Genome-scale hypothesis scanning. PLoS Biol. 1: 028-029.

Gibson G, Muse SV (2002). A primer of genome science. Sinauer Associates Inc. Publishers, U.S.A.

Hanash S (2003). Disease proteomics. Nature 422:226-232

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, et al. (2002). The genome sequence of the malaria mosquito *Anopheles gambiae*. Science 298:129-149.

Isokpehi RD (2003). Bioinformatics and drug target selection for malaria control. Trop. J. Pharm. Res. 2:123-124.

Isokpehi RD, Hide WA (2003). Integrative analysis of intraerythrocytic differentially expressed transcripts yields novel insights into the biology of *Plasmodium falciparum*. Malaria J. 2:38.

Lim YP, Hoog J, Gardner P, Ranganathan S, Andersson S., Subbiah S, Tan TW, et al. (2003). The S-Star trial bioinformatics course: An on-line learning success. Biochem. Mol. Biol. Edu. 31:20-23.

Morel C (2003). Neglected diseases: Under-funded research and inadequate health interventions. Can we change this reality? EMBO Rep. 4:S35-S38.

Oduola A, Wayling S, Morel C (2002). Genomics and bioinformatics for tropical diseases. TDRnews 68:6-7.

Pearson WR (2001). Training for bioinformatics and computational biology. Bioinformatics 17:761-762.

Simpson AJ, Reinach FC, Arruda P, Abreu FA, Acencio M, Alvarenga R, Alves LM, et al. (2000). The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis. Nature 406:151-157.

Singer PA, Daar AS (2001). Harnessing genomics and biotechnology to improve global health equity. Science 294:87-89.

Weatherall D (2003). Genomics and Global Health: Time for a Reappraisal. Science 302, 597-599.

WHO (2002). Genomics and World Health. World Health Organization, Switzerland.