Status of Bioinformatics in Southern Africa: Challenges and Opportunities



Yasmina Jaufeerally-Fakim[1], Daneshwar Puchooa[1] and Luke Mumba[2]

[1] University of Mauritius, SANBio Bioinformatics Node, Mauritius
[2] Director Southern African Network for Biosciences (SANBio), New Partnership for Africa's Development (NEPAD), South Africa

Background

The Southern African Network for Biosciences (SANBio) includes twelve countries in Southern Africa namely Angola, Botswana, Malawi, Mauritius, Mozambique, Namibia, Lesotho, Swaziland, Seychelles, South Africa, Zambia and Zimbabwe. In 2005 SANBio started its function to promote Science and Technology for the benefit of the people of the region. SANBio draws its programme of work from the Africa's Science & Technology Consolidated Plan of Action (CPA) [1]. The CPA was adopted in 2005 by the African Ministerial Conference on Science and Technology (AMCOST) at its second conference in Dakar, Senegal. The same document was endorsed a year later by the AU Summit in Khartoum, Sudan. The CPA articulates Africa's common objective of socio-economic transformation and full integration into the world economy. It reaffirms the continent's collective action for using S&T for meeting the developmental goals of Africa with key pillars being capacity building, knowledge production and technological innovation. The CPA recognizes that S&T in Africa is plagued by such factors as weak or no links between industry and S&T institutions, a mismatch between R&D activities and national industrial development strategies and goals. The consequence of these weaknesses is that research findings in public institutions, including universities, do not get accessed and used by local industries especially small and medium enterprises. The CPA comprises of three key areas: research and development programmes; improvement in policy conditions and building innovation mechanisms; and implementation, funding and governance strategies.

CPA R&D Programmes and Implementation

The programmes contained in the CPA are implemented through regional networks of centres of excellence, consisting of hubs and nodes. The objectives of these networks are: to improve quality of and access to infrastructure and facilities; develop further institutional and political regulations; improve the human skill base; obtain political and civil society support; strengthen the capacity of regional institutions; integrate R&D into sectoral programmes; improve the applicability of S&T towards the Millennium Development Goals and Sustainable Development; and to develop innovative funding instruments and build international partnerships. Research and Development Programmes of the CPA consists of five clusters. Under each cluster there are several programmes. The clusters are: *Cluster 1, Biodiversity, Biotechnology and Indigenous Knowledge*. This cluster focuses on the conservation and sustainable use of biodiversity; safe development and application of biotechnology; securing and using Africa's indigenous knowledge base. *Cluster 2: Energy, Water and Desertification*. This includes building a sustainable energy base by increasing rural and urban access to environmentally-sound energy sources and technologies; securing and sustaining water to ensure sustainable access to safe and adequate clean water supply and sanitation; combating drought and desertification by improving scientific understanding and sharing of information on the causes of and extent of drought and desertification in Africa. *Cluster 3: Material Sciences, Manufacturing, Laser Technology and Post- Harvest Technology*. This includes the development of new and improvement of existing infrastructure by building new skills or expertise in material sciences, promoting the sharing of physical infrastructure and exchange of scientific information and the promotion of public sector partnerships in material sciences research and innovation. *Cluster 4: Information and Communication Technologies; and Space Science and Technologies*. This in-

cludes the creation of experts engaged in computer science, information systems as well as informatics; building skills in software research and development. It also includes the establishment of the African Institute of Space Science. *Cluster 5: Mathematical Sciences*. This includes the establishment of an African Mathematical Institutes aimed at strengthening the African Mathematical Institutes network that was constituted in 2005 with the sole purpose of building a new generation of African scientists and technologists with excellent quantitative problem-solving skills.

SANBio operates from its secretariat based at the Council for Scientific and Industrial Research (CSIR) campus in Pretoria, South Africa. It has been active in the setting up of regional projects whereby several countries participate. Among those are: The scientific validation of traditional medicines for HIV treatment, conservation and utilization of plant genetic resources in the SADC (Southern African Development Community) region for food security, Fish biodiversity of inland rivers of Southern Africa, Mushroom production, Livestock Development and Capacity building in bioinformatics. Each project is managed by a country node under the supervision of the SANBio Director.

Capacity Building in Bioinformatics

The University of Mauritius is the regional node for the project on capacity building in Bioinformatics. The main objective of the action is to provide training to scientists in the countries of Southern Africa for the development and utilization of bioinformatics in research institutions and universities. It will create a platform whereby research in biosciences will be strengthened through the incorporation of bioinformatics within the current activities. Such an initiative will have a regional impact and improve the perspectives of the research output.

Conventional approaches for research projects that are under way in the region in biological sciences are based on methods of molecular biology, biochemistry and biotechnology. Most of the universities have well-developed departments for these subjects. They also have their computer science departments and this action will bridge them with the aim of setting up facilities for bioinformatics. The justification for such a project is the fact that Southern Africa has yet to tap genomics and expression data that have become available in the recent past. Most other regions of the world are actively utilising and applying such information for bringing solutions to their priority problems. Southern African countries face massive difficulties in tackling infectious diseases, food production, animal health, environmental degradation and population growth. Malawi has several scientific set ups where research in being undertaken in molecular epidemiology of HIV, malaria and other human diseases [2,3]. The School of Veterinary Medicine of the University of Zambia and the Livestock and Pest Research Centre of Zambia are conducting a major work on the tick-borne diseases such as Trypanosomiasis. More than 75 % of the livestock are found in areas where the tsetse fly, which is the vector for this disease, can be found. Current work there include epidemiology of human trypanosomiasis, comparative studies of cysteine proteases of *Trypanosoma brucei* and *T. congolese*, characterisation of Theileria parasites and immunogenicity of Theileria sporozoites.

The Theileria genome information is available [4] and the scientists in Zambia will benefit from training in using genome browsers and other databases for extracting useful information on this organism. Works on sequencing other Theileria species are also under way at ILRI (International Livestock Research Institute) in Kenya. Tapping genomic and expression data, which have become available for many parasites, can lead to more efficient control of diseases through the design of better vaccines and drugs.

An important fish biodiversity project is under way in Malawi. The lakes of Southern Africa are heavily exploited for food production. Lake Malawi is known to house five hundred endemic fish species and follow up of decrease in stocks is necessary to assess the probable loss in biodiversity. DNA sequences for phylogenetic comparison are being used for better diversity management of the fish resources of the seven main rivers of Southern Africa which are the Zaire, Zambezi, Okavango, Limpopo, Orange, Ruvuma and Cunene. They run through eleven countries and make up 6.7 km^2 of catchment area.

The region also hosts a major project on the Conservation of Plant Genetic Resources. The SADC Plant Genetic Resources Centre (SPGRC) is located in Zambia and the project aims at strengthening capacities in other countries for: Assessment of genetic diversity, identification of

important genetic traits, conservation measures, utilisation and food security and sustainable livelihoods.

At the University of Mauritius, a bioinformatics group has been set up to evaluate the current needs and to initiate research. The group is presently working on microbial genomes comparison [5].

South Africa stands as an opportunity for neighbouring countries, as bioinformatics is well implemented among scientific institutions and universities. High quality research output is evidence for this [6-9]. A strong network has been set up and is functioning with many nodes which provide short courses.

The specific objectives of this project are:

- setting up of a network of scientists from the Southern African region,
- organise core facilities,
- Initiate the development of courses in bioinformatics,
- provide specific training to academics and students,
- assist research students in obtaining short-attachment in bioinformatics centres.

Scientists in the region have high hopes that training in the utilisation of bioinformatics resources and tools will enable them to enhance the quality of teaching and research being undertaken. Above all, such tools will accelerate development of diagnostics, vaccines and other medical and veterinary products that are directly relevant to national priorities.

An advisory working group has been set up and comprises of the following members:

- Prof. Luke Mumba, Director SANBio
- Dr. Y. Jaufeerally-Fakim, SANBio Bioinformatics Node Coordinator
- Mrs. Chimwemwe Chamdimba, M & E Manager, NEPAD Office Science and Technology
- Prof. Fourie Joubert, University of Pretoria
- Prof. Oleg Reva, University of Pretoria
- Dr. Etienne de Villiers, ILRI (International Livestock Research Institute) Kenya
- Prof. Erik Bongcam-Rudloff, The Linnaeus Centre for bioinformatics-SLU, Uppsala, Sweden
- Dr. Eija Korpelainen, CSC-IT Center for Science, Finland

A first meeting was held at the University of Pretoria, Computational Biology department in January 2009. Participants came from Malawi, Zambia, Namibia, Botswana, Mozambique, Tanzania, Mauritius and South Africa. A preliminary needs assessment was conducted to work out the types of training that will be required. The project planning meeting was held in Mauritius last July and the series of activities to follow was finalised. The areas for training were identified as follows:

1. *Genome Browsers:* an introduction to a browser (e.g., Human or Rice genome browser (TIGR), ENSEMBL, UCSC), general features, how to understand the display, and work with the features available.
2. *Biodiversity and Evolutionary Genetics:* tools for sequence alignment, motifs searching, finding recombinants (HyPhy/ PAML). phylogenetic programs; how to interpret output from such analysis.
3. *Pathogen genomics:* comparative genomics, how to compare genomes, find homologs or conserved sequences. Using for example the *Plasmodium falciparum* or *Mycobacterium tuberculosis* genomes.
4. *Assembly and Annotation:* tools for genome assembly; how annotation of different genomes is conducted.

It is expected that this initiative will enable more projects to follow and therefore securing extra funding for a flagship project will be an essential part.

Internet Access

The digital divide between Europe and Africa has been addressed by the e-Africa commission

Figure 1. Participants at a SANBio Regional Bioinformatics Training Course, University of Pretoria, South Africa, 2009.

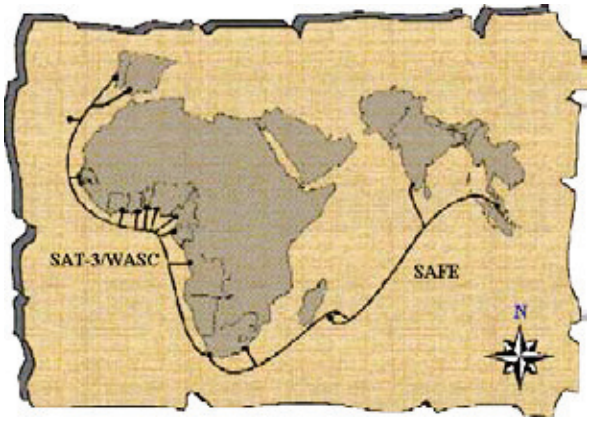

Figure 2. The South Africa Far East cable links Europe with countries of Western and Southern Africa and the Far East.

of NEPAD. The prices of internet access in Africa were among the highest in the world. This is about to change. High speed digital links have been made possible by the deep sea cable SAFE (South Africa Far East) which is 13,104 km and has points in Melkbosstrand and Mtunzini South Africa, St Paul in Reunion Island, Baie du Jacotet in Mauritius, Kochi in India and Penang Malaysia. It has capacity upgradable to 130 Gbits per second. This is important as it will allow easy and fast Internet traffic which is essential for sequence data requiring transfer of very large files.

In July this year the Seacom cable owned by African companies, went live thus bringing high speed connections to South Africa, Mozambique, Tanzania, Kenya, Uganda and Asia. It is 17,000 km and already many universities are benefitting from it (http://www.seacom.mu/intro.html).

The Regional Communications Infrastructure Program (RCIP 3) has brought high speed connection to Eastern and Southern Africa through the links from Malawi, Mozambique and Tanzania to the sea cable running up the Eastern coast of Africa.

Figure 3. Connection through SEACOM.

References

1. Africa's Science & Technology Consolidated Plan of Action (2006). NEPAD Office of Science & Technology. D.S Print Media, Johannesburg. Pp72. ISBN : 978-0620-37633-4.
2. Gaoqian F, Aitken E, Yosaatmadja F, Kalilani L, Meshnick SR, Jaworowski A, Simpson JA, Rogerson SJ (2009) Antibodies to variant surface antigens of Plasmodium falciparum-infected erythrocytes are associated with protection from treatment failure and the development of anemia in pregnancy. The Journal of infectious diseases 200(2):299-306.
3. Kalilani L, Atashili J (2006) Measuring additive Interactions using odds ratio. Epidemiologic Perspectives and Innovations 3:5.
4. Bioinformatics in Africa [http://www.lirmm.fr/france_afrique/20060708_NEPAD.pdf]
5. Khoyratty Sher-ullah SS, Souza MT Jr., Jaufeerally-Fakim Y (2008) Structural analysis of catalase from two Musa accessions FHIA 18 and Williams and from Ravenala Madagascariensis. In Silico Biology 8: 413-425.
6. Collins NE, Liebenberg J, de Villiers EP, Brayton KA, Louw E, Pretorius A, Faber FE, van Heerden H, Josemans AI, van Kleef M, Steyn HC, van Strijp, Birkholtz LM, Wrenger C, Joubert F, Wells GA, Walter RD, Louw AI (2003) Parasitespecific inserts in the bifunctional Sadenosylmethionine decarboxylase/ornithine decarboxylase of Plasmodium falciparum modulate catalytic activities and domain interactions. Biochem J. 377:439-448.
7. Klockgether J, Wurdemann D, Reva O, Wiehlmann L, Tummler B. (2007) Diversity of the abundant pKLC102/PAGI-2 family of genomic islands in Pseudomonas aeruginosa. J. Bacteriol. 189:2443-2459.
8. Reva, O, Weinel C, Weinel M, Böhm K, Stjepandic D, Hoheisel J, Tümmler B (2006). Functional genomics of stress response in Pseudomonas putida KT2440 J. Bacteriol. 188: 4079-4092.
9. Wells GA, Birkholtz LM, Joubert F, Walter RD, Louw AI (2006) Novel properties of malarial Sadenosylmethionine decarboxylase as revealed by structural modelling. J. Mol. Graph. Model. 24:309-318.