

# Towards a bioinformatics network for Latin America and the Caribbean (LACBioNet)

Wim M Degraeve,<sup>1</sup> Reynaldo Vargas,<sup>2</sup> Fernando Alvarez,<sup>3</sup> Julio Collado-Vides,<sup>4</sup> Luis Nuñez,<sup>5</sup> José Luis Ramirez,<sup>6</sup> Oscar Grau<sup>7</sup>

<sup>1</sup>Departamento de Bioquímica e Biologia Molecular, Instituto Oswaldo Cruz-Fiocruz, Rio de Janeiro, Brazil; <sup>2</sup>Universidad Católica Boliviana, Cochabamba, Bolivia; <sup>3</sup>Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay; <sup>4</sup>CIFN-UNAM, Cuernavaca, Morelos, Mexico; <sup>5</sup>Centro de Cálculo Científico, Universidad de Los Andes (CeCalCULA), Mérida, Venezuela; <sup>6</sup>Instituto de Estudios Avanzados – IDEA, Caracas, Venezuela; <sup>7</sup>Instituto de Bioquímica y Biología Molecular, Facultad de Ciencias Exactas, UNLP, La Plata, Argentina

**Abstract:** Bioinformatics is increasingly recognised as a crucial field for research and development in the biological sciences, and forms an integral part of genomics, proteomics and modern biotechnology. Worldwide participation is important, and scientists in developing countries can contribute to this field. Regional networks for bioinformatics are highly beneficial for capacity strengthening and cooperation, and for establishing productive interactions between scientists in the fields of biological and informatics sciences. Such a network (LACBioNet) is being organised for Latin America and the Caribbean. Its immediate goals include the organisation and extension of nodes and services, information and communication, research and development in different specialty fields of bioinformatics, and training and human resource development.

**Keywords:** bioinformatics, computational biology, networking, LACBioNet

## Introduction

Bioinformatics is widely recognised as pivotal to research and development in agricultural, veterinary and human health sciences, and biotechnology. The amount of biological data is growing at exponential speed ever since the onset of modern sequencing techniques. The organisation of such data, flexible retrieval, further processing, and innovative analysis and representation of data, all contribute to our understanding and capacity to construct more refined models of the biological world. Computer-assisted planning and research is now essential for managing complex areas such as vaccine and drug development, and is vital for understanding the functioning and regulation of biological processes and the study of host-parasite interactions, epidemiology, molecular evolution and phylogeny, population genetics and so on.

Many genome projects of (micro)organisms (at this moment more than 60) (KEGG 2002) have been successfully completed, and more than 150 of a wide variety of model organisms and organisms with medical, veterinarian or agricultural importance are under way (KEGG 2001). The need for new developments in the field of diagnostics, drugs and biotechnology as a whole make worldwide participation a necessity, as globalisation through distributed technology, opportunities and knowledge must be developed. Indeed,

such new developments and insights will come about through a creative and large-scale cooperative analysis, which is currently hampered by insufficient computing power, software, optimised algorithms and manpower. Several research centres in Latin America and the Caribbean are active in genomics, structural biology and drug design (see Table 1). Their interests also include microarray analysis, proteomics, phylogeny, database design etc. Trained bioinformaticists are crucial to these fields.

Moreover, it has also been observed that bioinformatics and computational biology are disciplines where scientists in developing countries can more effectively contribute to the world scene, as computer hardware is less demanding across many applications. But critical mass and interaction between groups is certainly a bonus, alongside creativity and insight. Thus, the organisation of regional networks for bioinformatics is highly beneficial. However, improvements to research capacity through training, sustained funding and collaborative projects (ie capacity strengthening) are urgently needed in the different fields and specialties, and interfaces should be established between disciplines in informatics,

Correspondence: Wim M Degraeve, DBBM-Fiocruz, Av Brasil 4365, 21045-900 Rio de Janeiro, RJ, Brazil; tel +55 21 2598 4205; fax +55 21 2590 3495; email wdegrave@fiocruz.br

**Table 1** Web addresses of major centres in Latin America involved in bioinformatics

<i>Country and centre</i>	<i>Web access</i>	<i>Main activities</i>
<b>Argentina</b>		
Universidad de la Plata	www.biol.unlp.edu.ar/	EMBnet node, access, training
Fundacion de Informática Médica	www.fim.org.ar	Medical informatics
Universidade Nacional de G San Martin	www.iib.unsam.edu.ar/genomelab/	Bioinformatics in genome analysis
<b>Bolivia</b>		
Universidad Catolica Boliviana	http://iiaa.ucbca.edu.bo	Software development, training, Expasy mirror
<b>Brazil</b>		
Fundação Oswaldo Cruz	www.dbbm.fiocruz.br www.procc.fiocruz.br	Access, training, bioinformatics in genome analysis
Laboratorio Nacional de Computação Científica	www.lncc.br	Access, training, bioinformatics in genome analysis
Federal University of Rio de Janeiro	http://biomat.org	Training, bioinformatics
Pontificia Universidade Católica do Rio de Janeiro	www.inf.puc-rio.br	Training, bioinformatics
Universidade de São Paulo	www.bioinfo.usp.br www.lemb.icb.usp.br/Virus/ http://verjo19.iq.usp.br	Training, bioinformatics in genome analysis
Fundação de Amparo à Pesquisa do Estado de São Paulo	http://watson.fapesp.br/onsa/Genoma3.htm http://watson.fapesp.br/structur/menusgp.htm	Bioinformatics in genome analysis Bioinformatics in structural biology
Instituto Ludwig de Pesquisa sobre o Câncer	www.ludwig.org.br/ORESTES/	Bioinformatics in genome analysis
Universidade de Campinas	www.nib.unicamp.br/proj.htm www.ic.unicamp.br www.lbi.ic.unicamp.br	Medical informatics, bioinformatics in genome analysis, training
Federal University Rio Grande do Sul	www-gppd.inf.ufrgs.br	Cluster and parallel computing
Universidade Federal de Pernambuco	www.cin.ufpe.br	Training, bioinformatics
<b>Chile</b>		
Universidad de Chile	www.dcc.uchile.cl	EMBnet node, access, training, bioinformatics
<b>Colombia</b>		
Universidad Nacional de Colombia	http://dis.unal.edu.co/	Training, bioinformatics
<b>Cuba</b>		
Centro de Ingeniería Genética y Biotecnología	www.cu.embnet.org	EMBnet node, access, training, bioinformatics
<b>Mexico</b>		
Universidad Central de Mexico	http://embnet.cifn.unam.mx www.cifn.unam.mx/Computational_Genomics/ http://clusters.unam.mx	EMBnet node, access, training, bioinformatics Bioinformatics in genome analysis, clusters
<b>Venezuela</b>		
Centro Nacional de Calculo Cientifico	www.cccalc.ula.ve	EMBnet node, access, training, bioinformatics
Instituto Venezolano de Investigaciones Cientificas	http://cbe.ivic.ve	Training, bioinformatics in structural biology

biochemistry, biology, pharmacology and so on. Such a network was initiated for Latin America and the Caribbean at a recent planning meeting<sup>1</sup> during an international training workshop in bioinformatics applied to genome studies, sponsored by UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases – TDR, Burroughs-Wellcome Trust and others (2001 May 21–Jun 15; Fiocruz, Brazil). Seventeen participants representing

16 institutions from 8 countries in Latin America and the Caribbean discussed network goals and directives.

## Network formation and planning

A number of researchers and centres in the region (see Table 1) are already active in the field of bioinformatics.<sup>2</sup> A survey of such groups and centres is underway with the aim to develop a widespread network, initiate cooperation and

stimulate capillarization. Integrated activities, meetings, training workshops and exchanges are being established. The scope of the regional bioinformatics/computational biology network for Latin America and the Caribbean (LACBioNet), is broadly understood as the use of information technology for the management and modelling of biological data. It also includes the use in epidemiology, aspects of public health planning and intervention, genetics etc.

Key players in different countries in the region (see Table 1) have been invited to list and contact known centres and people active in the field. Online registration is available and includes subscription to a discussion list, and also available is an associated database of bioinformaticists in the region.<sup>3,4</sup> Network activities will be conducted in English, Spanish and Portuguese.

LACBioNet will focus on the following activities:

- organisation and extension of nodes and services
- information and communication
- research and development
- training and human resources development.

### *Organisation and extension of nodes and services*

The internet infrastructure, network extension and bandwidth are improving rapidly in many countries in the region, with Brazil<sup>5</sup> and Mexico<sup>6</sup> leading the way. Additional information on national network structures and their plans for local network expansion can be found for Panama,<sup>7</sup> Peru,<sup>8</sup> Chile,<sup>9</sup> Venezuela<sup>10</sup> and Argentina.<sup>11</sup> Furthermore, Internet2 has started in the region, providing connection with Mexico, and, since 2001 as part of the Americas Path Network project (AMPATH),<sup>12</sup> also with Chile, Puerto Rico, Brazil and Argentina. Connection with the national research networks of other Latin American and Caribbean countries is expected to occur gradually in the near future. These links will be used for interconnection and cooperation amongst academic networks for the development of new applications and protocols.

However, many research institutes in Latin America do not currently have adequate facilities for bioinformatics, lacking either hardware, software or skilled staff for system management and user support, and/or trained bioinformaticists. Therefore centres that can provide individual users access to high performance facilities, software and databases are widely used (Table I). Usually such centres also play a nucleating role for new groups with growing research capability, and often act, at least to some extent, as training centres. Several centres in the region are

EMBNET nodes for their country, or receive other sources of support and funding from national or international funding bodies. They offer a variety of services to users, such as processing power with suitable software for different applications, good connectivity and bandwidth, and 24 hour/7 day operation. Such centres should exchange experiences, interfaces and scripts, and improve user support as much as possible. However, security issues for remote access are an ever-increasing problem, and very often it is hard to find and keep system managers, as non-profit or governmental centres usually cannot offer competitive salaries.

### *Information and communication*

To promote integration between centres, programmes and initiatives, it is vital to establish a directory of groups and individuals active in bioinformatics within the region, and create a virtual centre for bioinformatics and computational biology.<sup>13,14</sup> Initiatives are in progress to improve communication through discussion lists, web pages, distance learning and the use of videoconferencing. Also, a virtual library for bioinformatics is being organised, providing links to electronic journals in the field, and encouraging the submission of manuscripts from the region.

### *Research and development*

Research and development in collaboration with others needs to include the following fields (see Table 1):

- development and optimisation of algorithms, parallel processing and grid applications
- database structuring, methods for data mining, interfaces, scripting and visualisation
- general protein and nucleic acid sequence analysis, genomics, sequence assembly and annotation, clustering and comparative genomics
- functional genomics, analysis of proteomics and microarrays
- molecular evolution and phylogeny
- public health planning and intervention modelling (including geo-referencing)
- population genetics and biostatistics
- molecular modelling, docking, drug development etc.

Through a directory of groups and individuals, one can find research partners, set up joint projects, and exchange students and researchers. At a later stage, specific programme coordinators will be assigned for the different fields of activities to maximise integration. Specific funding (WHO/TDR, United Nations University, Burroughs-Wellcome trust, CYTED, Fogarty and others) will have to be sought to

promote the research and development in each speciality field. Software development and commercialisation are not excluded. Steps in such a direction are still incipient in Latin America, and bioinformatics is mostly restricted to governmental institutions. However, the growing interest in biotechnology (supported by bioinformatics) and medical informatics might bring a quick change to this picture.

### *Training and human resources development*

As bioinformatics is still incipient in many research institutes, there is an urgent need to train users at entry, medium and advanced levels, using hands-on training courses, online and distance learning, undergraduate and graduate training, and student exchange between network centres. In many centres (see Table 1) formal undergraduate and graduate programmes are being organised using a similar core curriculum, which can facilitate the exchange of students, staff and teaching modules. However, it is necessary to optimise integration, in multidisciplinary approaches, between informatics, genomics and biology/biomedical researchers. This is not always an easy task as there is a wide gap between the vocabulary and concepts in the respective areas. For example, the word 'database' often calls up very different images amongst researchers and informaticians, and the translation of a biological problem, explained in molecular biology jargon, into an algorithm or new application often struggles with linguistics and the translation of concepts. Special advanced training also has to be organised for tool developers, system managers and system security, trainers, and individuals who can generate research in bioinformatics.

An international training initiative in bioinformatics/computational biology applied to genome research was initiated by WHO/TDR<sup>15</sup> with a 'train-the-trainer' workshop<sup>16</sup> organised at Fiocruz, Rio de Janeiro, May 21–June 15, 2001. This was followed by the establishment of four regional centres for training in bioinformatics and applied genomics, in Africa,<sup>17</sup> Asia<sup>18,19</sup> and Latin America.<sup>20</sup> Other training efforts are also ongoing in Mexico<sup>21</sup> and Venezuela.<sup>22</sup>

Regional collaborative networking in bioinformatics has been functioning successfully in the Asia-Pacific region since 1998, with the Asian-Pacific Bioinformatics Network (APBioNET).<sup>23</sup> Similar to the Latin American and Caribbean effort, an African bioinformatics network is being organised with high hopes of further contributions and interactions in the field of bioinformatics/computational biology. Support and partnerships from centres in North America and other

advanced countries are highly beneficial to the success of collaborative network organisations, and such partnerships are extremely welcome.

## **Acknowledgements**

We thank the United Nations University,<sup>24</sup> Programa de Biotecnología para América Latina y el Caribe (BIOLAC)<sup>25</sup> and the UNDP/World Bank/WHO –TDR<sup>26</sup> for their support. We also thank Leonardo Silva Leite and Jonathan Scott for technical support.

## **Notes**

- <sup>1</sup> Planning meeting, 2001 Jun 13–15; Oswaldo Cruz Foundation, Fiocruz, Rio de Janeiro, Brazil.
- <sup>2</sup> Refer to <http://www.dbbm.fiocruz.br/latinbioinfo/institutes/>
- <sup>3</sup> <http://www.dbbm.fiocruz.br/helpdesk/>
- <sup>4</sup> [latinbioinfo@procc.fiocruz.br](mailto:latinbioinfo@procc.fiocruz.br)
- <sup>5</sup> Brazil, <http://www.rmp.br>
- <sup>6</sup> Mexico, <http://www.cudi.edu.mx>
- <sup>7</sup> Panama, <http://www.senacyt.gob.pa>
- <sup>8</sup> Peru, <http://www.rcp.net.pe>
- <sup>9</sup> Chile, <http://www.reuna.cl>
- <sup>10</sup> Venezuela, <http://www.reacciun.ve>
- <sup>11</sup> Argentina, <http://www.retina.ar>
- <sup>12</sup> AMPATH, <http://www.ampath.fiu.edu>
- <sup>13</sup> See <http://www.dbbm.fiocruz.br/helpdesk/> (in English).
- <sup>14</sup> <http://www.cecalc.ula.ve/bioinformatica/> (in Spanish).
- <sup>15</sup> WHO/TDR, <http://www.who.int/tdr/grants/awards/bioinformatics-10-01.htm>
- <sup>16</sup> Information and acknowledgment of sponsors is available at <http://www.dbbm.fiocruz.br/helpdesk/courses/>
- <sup>17</sup> Africa, <http://www.sanbi.ac.za/>
- <sup>18</sup> Asia, <http://www.icgeb.res.in/>
- <sup>19</sup> Asia, <http://www.mahidol.ac.th/>
- <sup>20</sup> Latin America, <http://icb.ime.usp.br/>
- <sup>21</sup> Mexico, <http://embnet.cifn.unam.mx/talleres/>
- <sup>22</sup> Venezuela, <http://www.cecalc.ula.ve/BIOINFORMATICA/thas/>
- <sup>23</sup> APBioNET, <http://www.apbionet.org>
- <sup>24</sup> United Nations University, <http://www.unu.edu/>
- <sup>25</sup> BIOLAC, <http://www.biolac.unu.edu>
- <sup>26</sup> UNDP/World Bank/WHO –TDR, <http://www.who.int/tdr/>

All websites accessed 14 March 2002

## **References**

- KEGG 2002. Compete genomes in KEGG [online]. Accessed 13 March 2002. URL: [http://www.genome.ad.jp/kegg/catalog/org\\_list.html](http://www.genome.ad.jp/kegg/catalog/org_list.html)
- KEGG 2001. Genome sequencing projects [online]. Accessed 13 March 2002. URL: [http://www.genome.ad.jp/dbget-bin/get\\_htext?Genome\\_Projects+-n](http://www.genome.ad.jp/dbget-bin/get_htext?Genome_Projects+-n)