A review of Bioinformatics training applied to research in Molecular Medicine, Agriculture and Biodiversity in Costa Rica and Central America

Allan Orozco, Jessica Morera, Sergio Jiménez and Ricardo Boza

Submitted: 4th February 2013; Received (in revised form): 13th April 2013

Abstract

Today, Bioinformatics has become a scientific discipline with great relevance for the Molecular Biosciences and for the Omics sciences in general. Although developed countries have progressed with large strides in Bioinformatics education and research, in other regions, such as Central America, the advances have occurred in a gradual way and with little support from the Academia, either at the undergraduate or graduate level. To address this problem, the University of Costa Rica's Medical School, a regional leader in Bioinformatics in Central America, has been conducting a series of Bioinformatics workshops, seminars and courses, leading to the creation of the region's first Bioinformatics Master's Degree. The recent creation of the Central American Bioinformatics Network (BioCANET), associated to the deployment of a supporting computational infrastructure (HPC Cluster) devoted to provide computing support for Molecular Biology in the region, is providing a foundational stone for the development of Bioinformatics in the area. Central American bioinformaticians have participated in the creation of as well as co-founded the Iberoamerican Bioinformatics Society (SOIBIO). In this article, we review the most recent activities in education and research in Bioinformatics from several regional institutions. These activities have resulted in further advances for Molecular Medicine, Agriculture and Biodiversity research in Costa Rica and the rest of the Central American countries. Finally, we provide summary information on the first Central America Bioinformatics International Congress, as well as the creation of the first Bioinformatics company (Indromics Bioinformatics), spin-off the Academy in Central America and the Caribbean.

Keywords: *Bioinformatics; molecular medicine; biodiversity; training; research; Central America*

INTRODUCTION

Bioinformatics is a science that acquires, stores, organizes, processes, manages and distributes massive amounts of data and information of biological character using sophisticated computing infrastructures and software. As a scientific discipline, Bioinformatics actually manages more data and information than most other theoretical research areas. Information used in Bioinformatics comes from knowledge gathered in many scientific areas and sub-areas, such as Biology, Informatics, Physics, Chemistry, Pharmacy, Electronics, Biostatistics, Agronomy and Medicine [1, 2], requiring complex integration procedures.

Corresponding author. Allan Orozco, School of Medicine (Bioinformatics Program), University of Costa Rica, San Pedro de Montes de Oca, 2060 San José, Costa Rica. Tel.: +506-215448; Fax: +506-25114570; E-mail: bioinformatica.em@ucr.ac.cr

Allan Orozco Solano is the creator and coordinator of the Bioinformatics Project at the School of Medicine (University of Costa Rica) and Director of the Bioinformatics Central American Network.

Jessica Morales is a researcher at the School of Medicine, University of Costa Rica, San Pedro de Montes de Oca, 2060 San José, Costa Rica.

Sergio Jimenez is a professor at the School of Medicine, University of Costa Rica, San Pedro de Montes de Oca, 2060 San José, Costa Rica.

Ricardo Boza is the Director at the School of Medicine, University of Costa Rica, San Pedro de Montes de Oca, 2060 San José, Costa Rica.

Among the areas of interest targeted by Bioinformatics, we would like to highlight the molecular biosciences (Biotechnology, Biomedicine, Biochemistry and Molecular Biology) and the 'Omics' (derived from the Latin *oma*, which means 'group or mass') sciences such as Genomics, Proteomics, Metabolomics, Transcriptomics, Epigenomics, Metagenomics, Nutrigenomics, etc. These sciences are fundamentally in charge of data generation, integration and the annotation of genes, proteins, biomolecules and metabolic networks, with the purpose of better understanding the dynamical function and properties of different organisms.

To better address the problems implied by each of these different areas, Bioinformatics resorts to complex combinations of mathematical, information technological, computational, networking and hardware tools, resulting in novel techniques, methods, algorithms and specialized databases.

These new tools produced by bioinformaticians enable experimental scientists and applied practitioners to carry out an integral analysis of the large variety of results that can nowadays be obtained from experimentation and routine field tests. It is worth noting here a relevant distinction often being made among two closely related and frequently confused disciplines: Bioinformatics (where the stress is put on data and information management) and Biocomputing (where number crunching is stressed instead). Both of them are so closely related that the distinction between them is often blurred and that the two of them are usually combined in common practice by most professionals.

We can take as an example the application to the study of cancer by using Structural Biology, to better characterize relevant proteins and their interactions, and Bioinformatics for the analysis of cancer gene expression in different clinical setups using microarrays. Each of these disciplines will yield large amounts of completely different data as a result of their analyses and, given their size, the results must necessarily be analyzed with the help of computers, using bioinformatic and biocomputing tools. Integration of interdisciplinary data from heterogeneous specialties usually transcends the field of expertise of specialists in each of the fields involved. Therefore, a professional is required to master and integrate knowledge obtained from inherently disparate disciplines, and to combine disjoint techniques to find the solutions sought.

Generally speaking, a Bioinformatics professional must therefore be able to integrate knowledge and tools from different sources to provide solutions to problems in the fields of Life Sciences, often and more so every day with the uprising of high-throughput experimental technologies, using high-performance computing techniques. This last remark helps us highlight another important element in this discipline: its highly dynamic nature that quickly and continuously evolves to match and adapt to experimental developments. To accomplish these professional qualifications, it is imperative to have stable Bioinformatics academic programs available, and that these programs can quickly adapt to new needs to deliver the education and training required on demand [2–4].

THE CENTRAL AMERICAN REGION AND COSTA RICA

Central America, being at the continent's central region, is also considered a sub-continent that connects North America with South America. It is constituted by seven countries, Belize, Guatemala, El Salvador, Honduras, Nicaragua, Costa Rica and Panama. In absolute regional terms, it is populated by approximately 45 million people and has one of the planet's most important biodiversities (flora and fauna). Costa Rica is a small Central American country with an area of 51.100 km^2 and a population close to 4.5 million people. Some of its main characteristics are to have no army and to have one of the oldest modern democracies in the world. It also has one of the best high-technology indicators in Latin America. Only Costa Rica harbors 5% of the world's biodiversity [5, 6]. This scenario gives the region an extraordinary value from a Bioinformatics perspective, with a great potential for the study of drug research and the detection of new active principles with potential application in Molecular Medicine, Agrigenomics and Nutrigenomics.

UNDERGRADUATE EDUCATION IN BIOINFORMATICS

Biology and Informatics

Costa Rica has a variety of education programs in Biological Sciences and Informatics. Since 2009 to 2012, 4573 students graduated from private universities in careers related to Computing Engineering, Systems Informatics, Software Development, Biological Sciences and Education. On the other hand, 579 students graduated in other related careers such as Biological Sciences, Science Education with emphasis in Biology, etc. At the undergraduate level, there are several Bioinformatics courses in Costa Rica on careers such as Biotechnology, Biology, Medicine, Odontology, Pharmacy and Agronomy. All these careers belong to programs in the public university. Other countries of the region, such as Honduras (Autonomous University of Honduras) and Guatemala (Del Valle University), have given Bioinformatics courses and workshops at this level as well.

GRADUATE-LEVEL EDUCATION IN BIOINFORMATICS

Master's Degree Program

The Costa Rican Institute of Technology (TEC) started delivering the first optional graduate-level courses in Computational Biology as part of the program in their Master's Degree in Computing. In 2012, the University of Costa Rica's Medical School initiated a more specific Master in Bioinformatics and Systems Biology, prominently positioned within its postgraduate program in Biomedical Sciences. This Master was started with 12 students, increasing in 2013 to 29 inscribed students, with 65% of the accepted students coming from Informatics areas. This Master's Degree is pioneer at the Central American region. Its professors are graduates from Harvard, Houston-TX, Karolinska, Heidelberg, Glasgow, Freiburg, University College of London and the Autonomous University of Madrid. In designing this training program, the initial proposal (originally written in 2006) used Oxford University's Bioinformatics Master's Degree as its reference (Table 1).

Table 1: Syllabus summary of the Master in Bioinformatics and Systems Biology organized by the University of Costa Rica

Course	Credits
I	
Bioinformatics	2
Programming	2
Biological sequence databases and analysis	2
Biostatistics and experimental design	2
II	
Computational genomics and proteomics	3
Computational paradigms	2
Thesis research I	8
III	
Data mining	2
Modeling and simulation of biomolecular structures	3
Thesis research II	8
IV	
Systems biology	3
Pharmacogenomics	3
Thesis research III	8
V	
Biological system modeling and mathematical optimization	3
Immunoinformatics	3
Thesis presentation and defense	6
Total	60

One credit equals 3 working hours per week during 15 weeks.

A major goal of this initiative has been to open new development routes for the country. In Costa Rica, there is not yet a strong market need for Bioinformatics, except for some indirect opportunities like outsourcing requests from foreign companies (mainly pharmaceutical companies from the West of USA) to local IT companies.

On the other hand, with >1300 companies, IT represents 14.8% of gross domestic product (5.8% of gross domestic product corresponding to digital services) and represents 33% of total exports in Costa Rica, making this country the first exporter of high technology in Latin America. In this context, creating new professional specializations that join the pharmaindustrial demands to the IT knowledge base and to the Life Sciences leads naturally to Bioinformatics and arises spontaneously as a major strategic decision to avoid losing competitiveness in the global market and to open new working opportunities for young students. This is obviously a great challenge for a relative newcomer in the field that needs continued efforts to educate professionals and attract companies with this local profile.

PhD Program

In 2013, the University of Costa Rica will start, for the first time in the region, a Computing Sciences and Informatics PhD program with one of its main research lines for the final thesis being Bioinformatics and Computational Biology. Eight students will begin the program with scholarships from the Costa Rican Ministry of Science and Technology.

BIOINFORMATICS EDUCATION AND APPLICATIONS ON MEDICINE, AGRONOMY AND REGIONAL BIODIVERSITY

Molecular Medicine

Bioinformatics tools are commonly used on research but, sadly, their presence is still wanting in the clinical sector, where they can be used for the assessment of risks on pathologies such as obesity, cancer and diabetes. It is already well established that the clinical patient's genome is a major morbidity determinant and hence should become a primary source of information that may allow medical doctors to take better decisions on the prevention, diagnosis and prognosis of genetically influenced diseases using the large bulk of knowledge amassed by Molecular Medicine. However, currently there is no clear consensus about the use of these bioinformatics tools within standard clinical practice in hospitals [7].

One of the main causes for this absence has been the lack of continuous and stable bioinformatics education that would allow various professionals to learn how to interpret the various laboratory analyses available using the full panel of computer-based tools offered by Bioinformatics. Even though many gene characterization studies are now available, as well as tests for molecular markers and mutations associated with predisposition to develop different diseases, the know-how and expertise accrued by researchers rarely finds its way to medical practitioners and, ultimately, patients in a practical way [6, 7].

Since 2004, a translational education corpus has been in development with the purpose of helping the Central American clinical specialists to get an up-to-date education of Bioinformatics technologies, methods and strategies that may enable them to carry out genomic studies in areas such as monogenic and complex diseases and in Pharmacogenomics. This educational block consists basically in mastering Bioinformatics and Computational Biology tools, and has had as its main goals to achieve a better preparation of professionals; to bring the students, basic professionals and clinical experts closer to practical applications and, ultimately, to reach the patient, enabling delivery of personalized medicine. This program has proven helpful in reducing the breach among the previously remotely related fields of Molecular Medicine and Bioinformatics [8].

The training topics covered in these courses included analysis of microarrays, genomic variations, Pharmacogenomics, biocomputing analysis of monogenic and complex diseases, large-scale Bioinformatics (Web Services and Workflows) and molecular interaction networks. Each of the program's topics has usually been associated with technological solutions offered and developed by the University of Costa Rica to address real-world problems faced by the isthmus' hospital sector. This training was delivered in-person and virtually. On the virtual section, students had to develop real clinical profile cases applying selected Bioinformatics technology catalogs. The tutorials associated to these courses helped demonstrate the practical solution of actual real-world questions and were tailored to specific model requirements and to the participants' professional profiles (Table 1).

The orientation of these training events towards Bioinformatics has resulted in a translation of common trends in research in the Life Sciences to actual practice in the clinical fields. The integration of the themes involved has channeled well-chosen sets of analytical options to medical researchers, giving them the opportunity of applying Bioinformatics technologies to their various specialization fields. One of the most recent examples is the Central American Bioinformatics and Health Sciences Education Workshop, given through INISA (National Institute of Health Research) in Costa Rica in 2010. The workshop was titled 'Bioinformatics and Molecular Biocomputing applied to the Health Sciences' and attracted the participation of experts from the oncology research sector.

Microbiology

The Biotechnology Research Center (CIB) is a research unit from the Biology School at the Costa Rica's Technological Institute. Among their research lines they work on Biomedicine, vegetal tissue culture, renewable energies, research and production of biocontrollers, genetic characterization and Microbiology, where they set special emphasis on the development and application of Bioinformatics tools. Building on their expertise, they have organized together with UNU-BIOLAC (United Nations University–Biotechnology for Latin America and the Caribbean) Bioinformatics education workshops applied to Microbiology and Metagenomics. As an example, in 2010, they organized a workshop titled 'Applied Microbial Bioinformatics' with support of instructors from the Western Illinois University.

Pharmacology

The University of Costa Rica's Pharmacy Faculty organized its first Bioinformatics workshop in 2010. The course addressed in depth topics such as molecular simulations and modeling, docking and drug design. Ten students from Pharmacy and Biotechnology participated in the course.

In 2011, the course 'Introduction to Bioinformatics applied to drug development' was given by the Federal University of Rio de Janeiro, Brazil. It was organized by the University of Costa Rica's Medical School and had an attendance of 15 people among students and professionals.

In June 2012, another workshop was organized on 'Bioinformatics and Computational Biology with emphasis in Pharmacogenomics' by an alliance of the University of Costa Rica's Pharmacy and Medical School, Iberoamerican Science and Technology Program (CYTED), Iberoamerican Bioinformatics Society (SOIBIO) and Global Bioinformatics Network (European Molecular Biology Network, EMBnet); 12 students participated, coming mainly from the Pharmacy Faculty. Topics covered included Web services in molecular systems, XML programming, analysis of microarrays and gene expression on CYP2D6, CYP2D9 and molecular docking.

Agriculture

The first GENECHIP National Workshop took place in 2008, with the running line 'Genomic, Gene Identification and Bioinformatics'. The workshop was based on a proposal presented by the Horticulture Department of the Wisconsin-Madison University to the Technical Institute of Costa Rica's Agronomy Department and with support from the Costa Rica's Science and Technology Ministry. The main purpose of the course was the analysis of a simplified microarray (created from a microscope slide using as genomic model *Arabidopsis thaliana*) to study gene expression during the photosynthesis process. The workshop had a participation of 20 people, mainly teachers of local Schools of Biology.

EARTH (Agriculture School of the Tropical Humid Region) University has organized three Bioinformatics and Agriculture workshops between 2008 and 2011. The main goal pursued was to tailor training delivered to the needs of the agricultural field. As an example of their structure, the 2010 workshop was partially supported with a grant from CYTED (510RT0391), had an attendance of 25 people, included students and professionals on Biotechnology, Biology, Pharmacy, Microbiology, Biomedical Sciences and Agronomy and coming from a variety of countries, both in the tropical and Central American regions like Guatemala, Panama, Costa Rica or Colombia. The workshops included academic training topics such as sequence analysis, biological databases, molecular structures, microarrays, workflow design and modeling and quantum and molecular dynamics. These topics were usually developed in conjunction with case studies applied to the agricultural sector using free and open-source software (FLOSS).

The workshop organized by EARTH in 2011 was named 'Bioinformatics and Computational Biology Fundamentals', had a prevailing focus on Bioinformatics applications on agriculture and biodiversity and was supported by EMBnet, SOIBIO and CYTED. It had an attendance of 24 participants with educational backgrounds ranging from students to practicing professionals, and coming not only from Central America (Costa Rica, Guatemala, Belize and Panama) but also from Haiti, Ecuador, Colombia and Kenya. Lectures were coupled to practical sessions specially adapted for the course goals and making use of FLOSS. The use of FLOSS was deemed especially relevant, as it ensured students could take home the tools mastered during the workshop and start using them straight away afterwards.

Biological prospection and biodiversity

Costa Rica has 169 protected areas (26.3% of the national territory) [8], which allows it to become a natural laboratory of great importance for biological prospection research. INBIO (National Biodiversity Institute) was established in 1989. Its main purpose is to register and obtain knowledge from the biological diversity of Costa Rica. To maintain this knowledge, they developed a biological information system called American Genus Ant (ATTA). This modular system with 22 taxonomic levels is used to maintain a taxonomic registry of 2 million species (animals, plants and fungi) through barcodes complemented with geo-referencing and enriched with meta-information on potential uses and applications of the different organisms. Many of the registered specimens are maintained on the Institute's own biological collections to facilitate further research and development.

Given the wealth of information that is being stored and maintained, it is no surprise that this information system has subsequently attracted the attention and become an important bioinformatics tool of the pharmaceutical industry.

Besides collecting information on biodiversity and maintaining reference specimens, INBIO functions also include training and education on sustainable biodiversity. In this role, INBIO organizes Informatics courses applied to Biodiversity and Bioinformatics. In 2011, they organized their second international Bioinformatics workshop 'Spreading Bioinformatics knowledge: a tool for knowledge development with added value', 50 people attended the workshop with training extending from students to diverse professionals. This workshop was celebrated with support from the Joint Genome Institute, Aberdeen University and Wisconsin University.

Finally, INBIO is using Bioinformatics to support their research in the metagenomic analysis of microbial communities related to sub-social beetles of the *Passalidae* family and in the analysis of ant mega-colonies and their environment using genomic sequencing to explore differential bacterial advantages (*Klebsiella* vs. *Pantoea)* in nitrogen fixation for *Atta cephalotes* ants that culture fungi for their subsistence [8].

Snake venom

The University of Costa Rica's Clodomiro Picado Institute (ICP) is a research center focused on the study of snake venom and antidotes. Currently, this Institute is deeply involved in the development of a research program in Proteomics to study toxic proteins, with technical support from Computational Biology. Their research program is complemented by an educational program teaching Bioinformatics applications for the analysis of translation, concentrating mainly on sequence analysis and equipment use. They have an agreement with the Príncipe Felipe Biomedical Institute in Valencia, Spain [9].

By coupling Proteomics and Bioinformatics they have been able to achieve significant success in classifying different phylogenetic and toxicity mechanisms and designing specific serums for *Bothrops asper* snakes from different geographical regions.

This work provided a clear example on how joint interdisciplinary efforts that bring together snake protein toxicology, Pharmacogenomics and Bioinformatics can lead to new perspectives and help design new drugs.

In 2012, the Interamerican Agricultural Institute (IICA), an organization that stimulates and promotes agricultural development and rural well-being of the Iberoamerican population, organized the first national educational technical forum related to Bioinformatics and Nanobiotechnology to seek possible technology convergences in agriculture innovation. This forum highlighted the relevance of Bioinformatics for the identification of bacterial and fungal genes that attack different crops and to help define strategies to fight them. It also called attention to the growing role of integrative nanobiotechnology, which is allowing the creation of new fertilizers and herbicides, *in situ* monitoring of environmental factors and plant growth as well as prolongation of food's useful life, among other applications [10].

BIOINFORMATICS INFRASTRUCTURE

The first Bioinformatics infrastructure of the Central American region was installed at the Informatics Center of the University of Costa Rica, promoted by University of Costa Rica Medical School. It is composed by 60 SGI rackable system servers and a network storage system with capacity for 60 TB distributed in four independent racks. The cluster runs on the Rocks Mamba Server 6.2 operating system (stable version of Red Hat Linux for servers). This infrastructure provides the technical support for the Central American Bioinformatics Network [11] as well as a Galaxy-based platform for aggregated services that delivers support for applied biomedical data analysis (www.bioinformatica.ucr.ac.cr:8080).

The Central American Bioinformatics Network (BIOCANET) is composed of a group of high-profile researchers and professionals from different Central American countries, working on the area's Universities, Institutes, Foundations and Research Centers. Currently, this organization is based on seven interconnected nodes (one per country), with a central node in Costa Rica, dedicated to Bioinformatics research and led by the Bioinformatics Research Laboratory (BREL) from University of Costa Rica's Medical School. Among its associated research projects in terms of bioinformatics data processing are gene molecular characterization, CFTR, SPINK 1, PRSS1, ATP7b, development of Web services catalogs on molecular medicine and biodiversity and increasing the agriculture competitiveness of the region [11].

The main objective of BIOCANET is to develop a Central American cooperation network of biocomputing services for research, technology development and Bioinformatics education associated to project developments on Biodiversity, Biotechnology, Molecular Medicine, Genomics, Proteomics and other Omics in general. This network has three main pillars: Biodiversity, Environmental and

Molecular Biosciences. These three large areas have Bioinformatics and Computational Biology as an integrator layer providing the technological support they all need.

The first meeting of the Bioinformatics Central American Network defined that a major goal to address was the need to shorten the existing knowledge gap in the discipline through specialized education. Experts from Guatemala (Del Valle University), Honduras (Autonomous University of Honduras), Costa Rica (University of Costa Rica) and Panama (Smithsonian Tropical Research Institute (STRI)) attended the meeting. Experts agreed on the development of a Bioinformatics virtual platform for training courses and education, for Central America and the Caribbean. This platform will be deployed over the actual HPC Bioinformatics cluster located at the University of Costa Rica's Rodrigo Facio head office.

RedBioNaGual is Guatemala's Bioinformatics National Network and has been spun off as a subnetwork of BIOCANET. RedBioNaGual has been made possible thanks to SENACYT (Science and Technology National Secretary), which approved its funding in June 2012. The stated goal of the network is to strengthen Bioinformatics application to Genetics, contributing towards increasing Guatemala's agricultural competitiveness, and its main weapons will be the development of a solid training and education schedule in Guatemala.

The network has scheduled 14 Bioinformatics education workshops between 2013 and 2014 as part of the activities to support the Central American Bioinformatics Network. These 14 workshops will tackle a wide variety of topics, from generic IT to specialized Bioinformatics: node creation and maintenance, Wiki design, network and Web site design, international Bioinformatics networks, virtual servers (for IT professionals and researchers), general concepts in Genetics and Biostatistics applied to Bioinformatics, techniques for the functional validation of sequences of interest and analysis of gene expression data, Comparative Genomics, expressed sequence tag (EST) analysis, integrative compilation of microbial promoters, generation of molecular markers, genetic identification (marker selection criteria, analysis of results, marker development methods, such as microsatellites, single nucleotide polymorphism (SNPs), etc.), genetic mapping techniques, generation of genetic maps, analysis of quantitative trait loci (QTLs), transcriptomics and biodiversity and taxonomic databases. All of these workshops will be organized with the collaboration of many institutions, such as Del Valle University; EARTH University; University of Costa Rica; Wisconsin University; National Institute of Health, Bethesda, MD, USA; National Autonomous University of Mexico; Santander's Industrial University of Colombia, USDA (United States Department of Agriculture) and UW Madison.

IBEROAMERICAN SOCIETY OF BIOINFORMATICS AND COSTA RICA

SOIBIO (Iberoamerican Society of Bioinformatics) is an international scientific society founded in 2009 to promote research and development of Bioinformatics and Computational Biology in the large international region of Iberoamerica. SOIBIO seeks its goals helping individuals, novel groups and national societies with an interest in these fields both at the academic and the professional level. This help can take various forms, such as facilitating international exchanges, collaboration, education and development in all topics related to its scope. Additionally, SOIBIO provides a gateway to connect them with other international societies and organizations with an interest in the area, like the EMBnet, the Asian Pacific Bioinformatics Network or the International Society for Computational Biology (ISCB) [12].

Costa Rica has been involved in SOIBIO since its origins, being a co-founder of the Society, and actively participates with representatives on the education committee and other international activities. Some of the organization's outstanding achievements include the B3BC education networks (*Bioinformatics, Biotechnology, Biocuration and Computational Biology)* and Global Organization for Bioinformatics Learning, Education and Training (GLOBLET, http://www.mygoblet.org/), whose main objective is to develop a strategy and general vision of global bioinformatics education and activities. The activities of this meeting will be supported by Bioinformatics Training Network), Netherlands Bioinformatics Centre, the Swiss Institute of Bioinformatics EMBL-European Bioinformatics Institute and EMBnet [13–15].

CENTRAL AMERICAN CONGRESS OF BIOINFORMATICS

The first Central America International Bioinformatics Congress was organized in 2012. This activity was officially supported by the Costa Rica Ministry of Science and Technology (MICIT), the Ministry of Exterior Commerce (COMEX) and the National Council for Scientific Research (CONICIT). The Congress was accompanied by satellite education activities and a general course on Bioinformatics delivered by members of the Genetics Department of Cambridge University. The event had a participation of >100 people from 26 different countries and reflected the variety of interests in Bioinformatics in the region.

As the interest in Bioinformatics in the region has been spreading to more institutions and over more different areas, it became increasingly more appealing to bring together all practitioners in the field to share their various interests and points of view, to become aware of each others work and to realize the full potential and breadth of Bioinformatics in the region. For these reasons, the Congress was planned as a multidisciplinary convergence approach to Bioinformatics, spanning application fields like Oncology, Molecular Biology, Disease Control (e.g. 'Chagas' disease or Paludism) or Pharmacogenomics (e.g. beta-lactamases), and analytical approaches like data mining, Computational Biology, Systems Biology, Quantum Molecular Dynamics, neural networks, next-generation sequencing data analysis, etc.

As we have already hinted across this article, a large part of the initial development of Bioinformatics in the region has taken advantage of collaborations with institutions abroad. These cooperations have been conversely reflected in the notorious participation of many international institutions in the event, both from Central America like Guatemala, Panama or Costa Rica, etc.; the rest of America like Argentina, Colombia, Chile, USA (Florida, Virginia) and Europe, most notably University of Cambridge in UK and various universities and research institutions in Spain.

PARTICIPATIONS IN INTERNATIONAL EVENTS FROM THE ISTHMUS

The initial steps in the development of a Bioinformatics community profited considerably from the establishment of cooperations with mature groups overseas, and as local groups have been reaching maturity, it is natural that they start reaching overseas as well. Besides the mentioned collaborations, this outreach materializes as active participation in international events to present local developments to the wider international Bioinformatics community.

From 2009 to 2012, the University of Costa Rica's Medical School has participated in events of international relevance either by itself or in representation of the Central American region. Some of these events stood out for promoting new education alternatives on Computational Biology. For example, the ISCB's presentation in Costa Rica, Africa Joint Conference on Bioinformatics of Infectious Disease 2009, on Bioinformatics, Education and Curricular Development sections: Changes and Opportunities in Bamako, Mali.

Participations have also taken place as presentation of summaries on poster categories: ISCB 2012 California, EEUU: 'BiocaNet: Central American Bioinformatics Network', ISCB–Latin America, 2012 in Chile: 'Bioinformatics software development in Central America: A growth alternative in education and computer technology', on the field of molecular modeling of protein and DNA virtual structures and finally on European Conference of Computational Biology in Basel, Suiza: 'BiocaNET: A platform applied in Central America's clinical genomic medicine'. These were the first Central American Bioinformatics participations on relevant global events (School of Medicine, University of Costa Rica).

INITIAL ACADEMIC ENTREPRENEURIAL UNDERTAKINGS

Another indicator of academic maturity is the ability to generate practical knowledge of immediate application in society and to spin off new enterprises that can transfer and exploit this knowledge, increasing the competitiveness of the region and employing newly formed specialists.

In 2012, the University of Costa Rica through its Research Vice-Rectory approved the creation of the AUGE initiative (University Management Accelerator). Its main objective is the creation of companies from initiatives started by students, graduates and professors of the University. From this proposal, several professors from Schools of Medicine, Economy and Computing, along with several entrepreneurs from the area founded Indromics Bioinformatics (www.indromics.com), a spin off company of Bioinformatics services and education [16]. One of the main focus areas of the company is regional

Bioinformatics training and education. To strengthen this, the company has built a specialized Bioinformatics laboratory with virtual equipment for training in diverse Bioinformatics topics (Galaxy, Bioconductor, Taverna and EMBOSS). This space has capacity for a maximum of 15 people.

This is a good example of the coordination between academy, company and government to create companies from the academic sector and increase job creation opportunities for human resources trained and graduated at the University [16]. Indromics is the first Bioinformatics company in Central America from an academic background.

CONCLUSIONS AND FUTURE PROSPECTS

Development of Bioinformatics in the region started out of specific local needs in various institutions that were initially addressed largely in cooperation with other international partners, mainly from the USA and European Union (where Bioinformatics was better developed) but also from other regions with similar problems and interests, most notably in Latin America. Commonality of language has also been a great help, meaning that a large number of collaborations took place within Iberoamerica, most notably Spain.

The importance of deploying local training programs has become more evident as the local institutions have started to obtain results dependent on Bioinformatics results and have realized its full potential. Progressively, more institutions are starting to deliver courses, either of general scope or specialized to their own interest. These activities have taken place mostly with little coordination, as they were driven mainly by local needs: addition of new initiatives has been reflecting the pace at which each separate institution has been developing its own capabilities [17].

The role of pioneers in the field can be determinant to direct the growth of the discipline: although it may be unintended, in the initial stages, other institutions lagging behind look at them for directions and leadership, a role that to date has been fulfilled by University of Costa Rica. In this sense, the organization of the first Masters' and PhD degrees in collaboration with international partners is having a major role in promoting development of the discipline; the organization of the first international congress has enabled for the first time to bring together interested professionals from a variety of fields; and the nucleation of a Central American network in Bioinformatics is seeding the creation of similar national networks in each country. Although adopting such an active role in coordinating and bringing together all the, to date, independent initiatives carries a significant payload, we find that it is most rewarding in enabling fast development of partnerships and cooperative initiatives, and in contributing to the development of the region.

In this sense, we feel that the early creation of contacts and partnerships between groups with an interest in Bioinformatics, even if their respective application areas are dissimilar, has a major impact in speeding up the development of the discipline. Isolated development has commonly been identified as a major hindrance by most groups, with coordination and expertise exchange activities being the most welcome improvement for their advance. We therefore consider now that waiting for independent groups to mature to start cooperations will only introduce significant delays, and strongly advice that cooperations be established and new groups be integrated in existing initiatives as early as possible to speed up Bioinformatics development.

Although some groups are quickly reaching maturity, as evidenced by their participation in external international events, there is still a large way to walk if we want to develop a competitive community in the region. The creation of coordination networks is a great leap forward to exchange expertise and promote collaborative development, and is already resulting in a roadmap and work plan towards the consecution of this objective. A relevant part will be creation of an innovation framework where emergent bioinformaticians will be able to develop their own applications. We intend to reduce external dependency and build our own support systems and education programs, emphasizing education in IT technologies applied to biological and medical problems of interest in a region with such a rich biodiversity and the introduction of advanced experimental equipment (like NGS sequencers). It is worth noting that lack of this equipment did not preclude scientists from carrying out their research (by off-shoring the analysis to the European Union, USA or Asia), but we feel that first-hand experience will strongly contribute to the development of local skills and to reduce external dependency.

To summarize, there is still a lot to be done, and success will depend on maintenance and strengthening of international collaborations, but most of all, on the attitude and effort of all the players involved, who will be required to invest hard work, responsibility and a high level of compromise if they want to succeed in bringing the region up to speed to match the highest international standards.

Acknowledgements

Special thanks to Dr. Ricardo Boza, Director of the Medical School and to the University of Costa Rica's Rectory. To FreeBIT-CYTED (510RT0391), Ibero-American Network of Converging Technologies-CYTED, SOIBIO, EMBnet (Global Bioinformatics Network), MICIT (Costa Rica's Science and Technology Ministry), COMEX, CINDE (Costa Rica Investment Promotion Agency), CONICIT and to the Molecular Biosciences PhD Program of the Autonomous University of Madrid, Spain. Finally to Dr. María Victoria Schneider (EMBL-European Bioinformatics Institute, UK), José R. Valverde (CNB-CSIC Spanish National Center for Biotechnology-Spanish National Research Council, Spain), Diana Soto, Alexander Mora, Anabel González, Alejandro Cruz, Keylor Rojas, Man Sai, Jorge Torres, Mauricio Valverde, Pedro Rocha, Dennis Céspedes, David Holmes (FCV-Life Science Foundation, Chile), Emiliano Barreto (UNAL-National University of Colombia, Bogotá), Pamela Pennington, Georgina Gomez, Jorge Pérez, Gustavo Fontecha, Sergio Vargas, Cecilia Díaz, Alfonso Valencia (Spanish National Bioinformatics Institute), Ana Tamayo (EARTH), Adrian Pinto, Luis Jimenez, Ronald Gonzalez, Nien Weng, Luis Rojas, Luis Ruiz, Edwin Baez, Laura Vázquez, David Judge (Cambridge University, UK), Marcelo Jenkins (PROSIC, Information Society and Knowledge), Marta Guzmán and Adriana Gonzalez for integral support to this work.

References

1. Brass A. Bioinformatics education—a UK perspective. *Bioinformatics* 2000;**16**:77–8.
2. Orozco A. La Bioinformática y su importancia en Costa Rica. http://www.nacion.com/2012–09-25/Opinion/La-bioinformatica-y-su-importancia-en-Costa-Rica.aspx (29 December 2012, date last accessed).
3. Altman RB. A curriculum for Bioinformatics: the time is ripe. *Bioinformatics* 1988;**14**:549–50.
4. Counsell D. A review of Bioinformatics educations in the UK. *Brief Bioinformatics* 2003;**14**:7–21.
5. Central America Bioinformatics Network (BioCANET). http://www.soibio.org/RedCentroamericanaDeBioinformatica/ (28 December 2012, date last accessed).
6. Orozco A. Strategic aspect by technological for the academic and business development of the bioinformatics in Costa Rica and Central America regions. *BMC Bioinformatics* 2007. I (Suppl I), 1–3.
7. Orozco A. XXXIV Congreso de la Sociedad Española de Bioquímica y Biología Molecular. *Bioinformática y Medicina Molecular: Caso De Formación Mediante Tecnologías Aplicadas en Biología Molecular Computacional Dirigido Hacia el Contexto Clínico en España*, 5–8 September 2011, Barcelona), p. 102.
8. National Biodiversity Institute (INBIO). http://www.inbio.ac.cr (27 December 2012, date last accessed).
9. The Clodomiro Picado Institute. http://www.icp.ucr.ac.cr (28 December 2012, date last accessed).
10. Inter-American Institute for Cooperation on Agriculture (IICA). http://www.webiica.iica.ac.cr (10 January 2013, date last accessed).
11. Bioinformatics Platform. http://www.bioinformatica.ucr.ac.cr:8080 (10 February 2013, date last accessed).
12. Training at SOIBIO. Iberoamerican Society for Bioinformatics. http://www.soibio.org/es/ (20 December 2012, date last accessed).
13. Global Organization for Bioinformatics Learning, Education and Training. http://www.mygoblet.org/ (25 December 2012, date last accessed).
14. Schneider MV, Watson J, Attwood TK, *et al*. Bioinformatics training: a review of challenges, actions and support requirements. *Brief Bioinformatics* 2010;**11**:544–51.
15. Schneider MV, Walter P, Blatter MC, *et al*. Bioinformatics Training Network (BTN): A community resource for bioinformatics trainers. *Brief Bioinformatics* 2012;**13**:383–9.
16. AUGE. http://www.augeucr.com/auge.html (30 December 2012, date last accessed).
17. Fernandez P. The GTBP training programme in Portugal. *Brief Bioinformatics* 2010;**2**:626–34.