Proceedings of the "Bioinformatics of African Diseases and Pathogen Vectors" Conference (Nairobi 2007) - Editorial

Olivier Gascuel, Alia Benkahla, Winston Hide, Dan Masiga, Trushar Shah

Infection, Genetics and Evolution xxx (2008) xxx–xxx

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Editorial

This special issue of *Infection, Genetics and Evolution* is devoted to the *Bioinformatics of African Pathogens and Vectors* conference, which was held in Nairobi (Kenya), 28th May–2nd June 2007. During the first 4 days, the conference was hosted by the International Centre of Insect Physiology and Ecology (ICIPE) and followed a standard format with keynote speakers, oral presentations and posters. During the last 2 days a selection of young African scientists were trained on bioinformatics software (Artemis for genome annotation) and databases (SwissProt, and VectorBase for genomic data of disease vectors) by renowned teachers at the International Livestock Research Institute (ILRI). The Centre National de la Recherche Scientifique (CNRS, France) and Laboratoire d'Informatique de Robotique et de Microélectronique de Montpellier (LIRMM, France) provided funding to cover the travel and local expenses of African researchers and invited speakers. This special issue provides all abstracts of the talks and posters presented at the conference, along with a selection of eight articles derived from the conference, which have been rigorously reviewed and improved thanks to the journal's editorial board and anonymous referees.

This conference was the first on this theme in Africa. The aim was to provide a forum for discussions and the development of future collaborations amongst bioinformaticians, mathematicians, computer scientists and biologists, whose research focuses on the bioinformatics of African pathogens and their vectors. The meeting also offered young and emerging scientists from Africa the opportunity to meet scientists from the international community. More than 80 researchers attended the conference, including 65 from Africa representing 17 countries (Burkina Faso, Cameroon, Côte d'Ivoire, Ethiopia, Kenya, Malawi, Mali, Mauritius, Morocco, Nigeria, Rwanda, Senegal, South Africa, Sudan, Tanzania, Tunisia and Uganda), and the others from France, Mexico, Switzerland, UK and USA. We had 10 excellent keynote speakers, who discussed a wide range of research from genomics to drug design. This conference was the first opportunity for many African scientists to meet in person. The students decided to set up a Regional Student Group (RSG-Africa) of the International Society for Computational Biology (ISCB); they have already organised several workshops and training courses, and RSG-Africa currently includes about 400 members. Moreover, during the conference, a joint initiative was set up to build an African bioinformatics network, with the aim of developing pathogen and disease vector genome and protein databases. A first meeting of this network was held at Abuja (Nigeria, April 2008) with support from WHO-TDR.

Bioinformatics emerged about 15 years ago, with the acceleration of large sequencing programs. Since the 1980s, the size of GenBank is doubling every 16 months. We currently have complete genomes of more than 1000 species, and the new sequencing technologies (454, Illumina, etc.) will further accelerate data acquisition, with very large projects such as the sequencing of 1000 human genomes started in early 2008. Moreover, several other large-scale data are available to support and complete results extracted from genomes, namely transcriptomes, proteomes and interactomes. While the genomic data enable the discovery of the genes, i.e. the biological actors, these post-genomic data help in understanding the play. Bioinformatics has a key role in the analysis of all of these data. This science is used to search genes within genomes, elucidate the protein structures, predict the biological functions, build and simulate models of interactions and regulations, and suggest new targets for diagnostics and drugs.

Africa has a special interest in bioinformatics in this respect, with the focus on endemic diseases, their epidemiology and evolution, their causative agents whose biological mechanisms are still largely unknown, their vectors and their complex coevolution with parasites and hosts, and the design of new drugs and treatments. Sequencing projects were recently initiated in Africa on specific pathogens, e.g. *Theileria parva* which affects cattle in sub-Saharan countries and was sequenced by ILRI and TIGR in 2005. However, a huge amount of genomic and post-genomic data related to African diseases are available and produced worldwide at an ever increasing speed. An emblematic example is malaria, with *Plasmodium falciparum* and *Anopheles gambiae* genomes both sequenced in 2002 by international consortiums. However, these data are still poorly understood and require continuous bioinformatics efforts to extract all of the biological information they contain. For example, about two thirds of *P. falciparum* genes are still classified as hypothetical, so a lot of work remains to functionally annotate all of these genes and discover new drug and vaccine targets. Bioinformatics plays a key role in this quest, by integrating and comparing different sources of information (sequences, transcriptomes, proteomes, etc.), and thus suggesting new clues and hypotheses to be tested in wet labs. African researchers are in a good position to act in this direction. For example, a project discussed during the conference is to promote the maintenance of major pathogen and vector subsections of parent databases (SwissProt, GeneDB) by Africans, e.g. on malaria (Mali) or tsetse (South Africa). African researchers could take advantage of the availability of clinical data on a broad range of human diseases and integrate them together with functional genomics data to enhance the overall understanding of the diseases, thereby promoting diagnostics/drug/vaccine design processes. Moreover, bioinformatics is an essential component of molecular biology today. It is expected that bioinformatics will become increasingly important and may soon represent 1/4 of all researchers and other personnel in biology laboratories. It is thus essential for Africa to develop research in this domain on the continent and train young researchers, and this was one of the main goals of the conference.

The main themes of the conference were:

- *Genomics*, with keynotes on: "Translation of genomics in endemic countries" (W. Hidde, ZA), "Genomics and emerging

10.1016/j.meegid.2008.09.002

viral infectious diseases" (A. Djikeng, CA, USA), and "Taenia genomics" (E. Morett, MX).
- *Functional genomics and databases*, with keynotes on: "Data mining parasite genomes" (M. Berriman, UK), "Resources and databases for HIV" (T. de Oliveira, ZA), and "The annotation of proteins from pathogens in UniProtKB/Swiss-Prot" (A. Bairoch, SW).
- *Evolution*, with a keynote on "Trypanosomia, host/parasite/vector coevolution" (M. Tibayrenc, FR).
- *Structure, target discovery and drug design*, with keynotes on "Information superstructure for protozoan aquaporins" (R. Isokpehi, US), "In silico strategies for target discovery" (E. Maréchal, FR), and "In silico drug design" (M. Afshar, FR).

The eight papers selected in this issue deal with various aspects of the bioinformatics of African pathogens and vectors (genomics, gene expression, evolution, epidemiological markers, and new vaccine and drug target discovery):

- Megy et al. introduce VectorBase, a genomics resource for invertebrate vectors of human pathogens. High-throughput genome sequencing techniques have now reached vectors, e.g. *Anopheles gambiae* (malaria), *Culex pipiens* (lymphatic filariasis and West Nile fever) and *Rhodnius prolixus* (Chagas disease). Vector genomes are being sequenced to gain further insight into their biology; to develop new control strategies or understand the limitations of current strategies (e.g. insecticide resistance); to analyse the mechanisms driving their evolution; and to perform exhaustive analyses of the gene repertoire. In this context, VectorBase is involved in the annotation of vector genomes and provides a portal for access to the genomic information.
- Mulder et al. present a comparative analysis to study expanded gene families in *Mycobacterium tuberculosis*, the main agent of tuberculosis, causing millions of deaths annually. The ability of organisms to evolve with enhanced pathogenicity appears to be at least partially provided by gene duplication. This evolutionary mechanism results in expansion of gene families, thereby providing the organism with extra gene copies and thus the opportunity to evolve new functions. Comparative genomics tools are used to compare the proteomes of over 80 pathogenic and non-pathogenic micro-organisms, including several mycobacteria, to identify unique proteins and determine the extent of family expansion in *M. tuberculosis*.
- Albufera et al. present a study on the molecular characterisation of *Salmonella* isolates using REP-PCR and RAPD analyses. The objective is to assess these molecular methods to trace the clonality of *Salmonella* strains and produce epidemiological markers to detect contaminated food. REP-PCR shows greater discriminatory power in differentiating closely related strains, and a dendrogram approach indicates that most human isolates are clustered separately from food-source isolates.
- Ghouila et al. describe a new method to cluster and interpret gene expression data, and its application to macrophage genes. Clustering is one of the main routes to analyse such data. Genes that show similar expression patterns are clustered together and assumed to be co-regulated and involved in similar functions. The authors propose a new algorithm, named Multi-SOM, to achieve this task and deal with the notoriously difficult problem of estimating the number of clusters. They use Multi-SOM for the analysis of macrophage gene expression data generated in vitro from the same individual blood infected with 5 different pathogens. Gene ontology tools show that the obtained clusters are biologically meaningful.
- Githui et al. discuss the conservation of dynein light chains within *Plasmodium falciparum* species. This parasite belongs to the phylum Apicomplexa, and several proteins associated with apical organelles are believed to be crucial in the host (human) invasion process. Most notably, dynein is a multisubunit motor protein, which is presumed to be involved in vesicular transport in infected erythrocytes. Dynein light chains (Dlcs) consist of three different families: LC8, TcTex1/2, and LC7/roadblock. The data presented demonstrate that *P. falciparum* Dlcs sequences and functional domains show high sequence similarity within the species, but that only the group1 (LC8) has a high similarity to human orthologs. TcTex1/2 and LC7/roadblock have low similarity to human orthologs, and this sequence variation could be targeted for vaccine or drug development.
- Guerfali et al. present an informatics approach to predict epitopes of *Leishmania major* proteins in susceptible and resistant murine models of infection. They address the problem of selecting appropriate proteins as vaccine candidates, to minimize the number of peptides that would otherwise be classically assayed using an expensive and laborious experimental mouse model. They propose an immuno-informatics approach to analyze *L. major* secreted proteins in order to identify putative MHC class I binding peptides, which could be retained by murine BALB/c (susceptible model) and C57BL/6 (resistant model) MHC class I molecules and presented to CD8+ T cells. Surprisingly, a higher number of *L. major* peptides are predicted to bind BALB/c molecules and very few or none to bind C57BL/6 molecules.
- Fatumo et al. describe a new computational method to find novel potential drug targets of *Plasmodium falciparum* by analysing the metabolic network of knock-out strains in silico. This method analyses the topology of the metabolic network of *P. falciparum* to identify essential enzymes as possible drug targets. The essentiality of a reaction in the metabolic network is investigated by deleting (knocking-out) such a reaction in silico. With this, 70 essential reactions are identified, a refined list of 22 new potential candidate targets for *P. falciparum* is proposed, and there is reasonable evidence that half of them are valid targets.
- Saidani et al. review the potential and limits of in silico target discovery in the malaria framework. They describe how in silico organization of genomic and post-genomic information of all partners involved in malaria (human patient, *Plasmodium* parasite and *Anopheles* vector), complying with knowledge of the disease in etiologic terms, appears to be an efficient source of information to help selecting, but also discarding target candidates. Some limitations in our capacity to explore the stored biological information, due to the current genomic annotation quality, level of database integration, and the performance of existing analytic and mining tools, are discussed. In silico strategies to assess the feasibility of bringing a target to a therapeutic development pipeline, in terms of target "drugability", are introduced.

We trust that everyone who attended the *Bioinformatics of African Pathogens and Vectors* conference has very pleasant memories of the special relaxed atmosphere, with plenty of time for discussion and strong scientific content. Organizers from ICIPE, ILRI and LIRMM played a major role in this success, especially Vincent Lefort and Rosekellen Njiru. We would like to conclude this editorial by thanking them again for all of their work, patience and efficiency.

Bioinformatic of African Pathogens and Vectors

(May 28th–June 2nd 2007 ILRI and ICIPE Campus, Nairobi, Kenya http://www.lirmm.fr/france_afrique/Nairobi2007/)

Sponsors: International Centre of Insect Physiology and Ecology (ICIPE, Kenya), International Livestock Research Institute

Please cite this article in press as: Gascuel, O., Infect. Genet. Evol. (2008), doi:10.1016/j.meegid.2008.09.002

(ILRI, Kenya), Centre National de la Recherche Scientifique (CNRS, France), African Society of Bioinformatics and Computational Biology (ASBCB).

Organizing Committee, co-chairs: Daniel Masiga (ICIPE, Kenya), Trushar Shah (BecA, Kenya); members: Vincent Lefort (CNRS, France), Chantal Pacteau (CNRS, France), Anne Corval (CNRS, France), Etienne de Villiers (ILRI, Kenya).

Program Committee, co-chairs: Alia BenKahla (Tunisia), Winston Hide (South Africa); members: Ezekiel Adebiyi (Nigeria), Jaco de Ridder (South Africa), Scott Hazelhurst (South Africa), Dan Jacobson (South Africa), Beatrice Kilel (Kenya), Daniel Masiga (Kenya), Nicky Mulder (South Africa), Trushar Shah (Kenya), Etienne de Villiers (Kenya).

International Scientific Committee, chair: Olivier Gascuel (CNRS, France); members: Raphael Isokpehi (Jackson State University, USA), Daniel Masiga (ICIPE, Kenya).

Guest Editors
Olivier Gascuel*
LIRMM, CNRS - U. Montpellier 2, France

Alia BenKahla
Institut Pasteur de Tunis, Tunisia

Winston Hidde
SANBI, U. Western Cape, South Africa

Dan Masiga
ICIPE, Nairobi, Kenya

Trushar Shah
BecA, Nairobi, Kenya

*Corresponding author at:
Projet Méthodes et algorithmes pour la Bioinformatique, LIRMM, CNRS - U. Montpellier 2, 161 rue Ada, 34392 Montpellier, France.
Tel.: +33 4 67 41 85 47; fax: +33 4 67 41 85 00
E-mail address: gascuel@lirmm.fr
(O. Gascuel)

Please cite this article in press as: Gascuel, O., Infect. Genet. Evol. (2008), doi:10.1016/j.meegid.2008.09.002