BIOINFORMATICS, THE CLEARING-HOUSE MECHANISM, AND THE CONVENTION ON BIOLOGICAL DIVERSITY

MARCOS SILVA

Clearing-House Mechanism, Secretariat of the Convention on Biological Diversity, United Nations Environment Programme, 393 Saint-Jacques Street, Suite 300, Montreal, Quebec, H2Y 1N9, Canada

Abstract.—This paper discusses the relevance of bioinformatics to the Convention on Biological Diversity. It also discusses the role of the Convention's Clearing-House Mechanism and the Biosafety Clearing-House, and their role in assisting Parties and other Governments with issues pertaining to bioinformatics in general. Finally, it reviews many work programs established under the Convention and the potential role of bioinformatics in assisting with their development and implementation.

Key words.—biodiversity informatics, biodiversity convention, international cooperation

The word bioinformatics does not have a uniform usage: it may refer primarily to genomics (Sugden and Pennisi 2000, Ouzounis and Valencia 2003), or may also include more general issues related to biology and computation (Scott 2004). This paper defines bioinformatics as the application of tools of computation and analysis to capture and interpretation of biological data (Bayat 2002).

For many reasons, the Conference of the Parties (COP) to the Convention on Biological Diversity is paying increased attention to issues related to computational biology and to bioinformatics in general. Indeed, data collection and dissemination, data mining modeling. development. and interoperability and further enhancement of databases, and visualization of data were explicitly identified by the COP during its seventh meeting, and are found throughout the work programmes of many thematic areas and cross-cutting issues of the Convention.

This paper discusses the development and role of the Convention's Clearing-House Mechanism (CHM), synergies with the Biosafety Clearing-House (BCH), and CHM's support for initiatives related to bioinformatics in general. It also examines the significance of bioinformatics in assisting Parties in implementation of obligations under the Convention, and CHM's role in facilitating activities by Parties and Governments to exploit benefits arising from the evolving bioinformatics global infrastructure.

The view is also presented that development of bioinformatics tools and resources is having positive impacts on the ability of Parties to meet the three objectives of the Convention more effectively and to manage their biodiversity resources better. This trend is reflected by the overall impact of bioinformatics in the programme areas and cross-cutting issues of the Convention and its related activities, especially work related to the 2010 target.

THE CLEARING-HOUSE MECHANISM

CHM was created pursuant to Article 18, paragraph 3, of the Convention, to promote and facilitate technical and scientific cooperation between State Parties. As identified in its strategic plan (CBD 1999), CHM has three primary objectives: cooperation, promotion, and facilitation of technical and scientific cooperation; information exchange; and network development. CHM's original mandate has been broadened to include matters pertaining to information exchange (Article 17 of the Convention) and the Biosafety Clearing-House (BCH), pursuant to Article 20, paragraph 1, of the Cartagena Protocol on Biosafety. In general, then, issues related to bioinformatics fall within the purview of the CHM.

THE BIOSAFETY CLEARING-HOUSE

The BCH was established as part of the CHM pursuant to Article 20.1 (CBD 2001) of the Cartagena Protocol on Biosafety to:

- (a) Facilitate the exchange of scientific, technical, environmental and legal information on, and experience with living modified organisms;
- (b) Assist Parties in the implementation of the Protocol, taking into account the special needs of developing State Parties, in particular the least developed and small island developing States among them, and countries with economies in transition as well as countries that are centres of origin and centres of genetic diversity.

In this broad context, CHM recommended the technical and architectural framework that was eventually implemented in the pilot phase of the BCH, and it continues to oversee the framework's technical functioning and enhancements. In addition, CHM was responsible for the technical architecture of the BCH pilot phase; protocols and standards supporting interoperability among disparate and distributed databases; and design, modules, and text of the draft BCH Toolkit.

The BCH continues to evolve, particularly in the context of decisions made at the first meeting of the COP, serving as the meeting of the Parties to the Cartagena Protocol on Biosafety. Indeed, at the meeting, it was decided to approve the transition of the BCH pilot phase to its fully operational phase. Of interest are decisions pertaining to modalities of the BCH which include use of a decentralized Internet-based system, adherence to common formats, use of controlled vocabularies and metadata, and providing means to facilitate interoperability of data among disparate and remote databases (CBD 2004b).

SYNERGIES BETWEEN CHM AND BCH

It should be noted that the BCH may be the first instance in which Parties have the opportunity to register information electronically to fulfill obligations under a legally binding international treaty. Indeed, it would be difficult, if not impossible, to imagine an effective and functioning Cartagena Protocol without a Web-based BCH. In other words, Parties have a legally binding obligation to register data under the Biosafety Protocol, and this fact alone differentiates the narrower, more concise BCH mandate from the very broad CHM mandate.

In contrast, the very broad CHM mandate was developed to promote and facilitate technical and cooperation among Parties scientific and Governments. Although COP has repeatedly called for Parties to establish CHM National Focal Points, Parties are not legally obliged to do so. In addition, CHM operates in a much broader environment, in which information exchange is but one of its activities, albeit a singularly essential one. Even more, although CHM has assumed responsibility for information exchange under the Convention, this activity falls under the prerogative of Article 17 of the Convention, and not Article 18.3.

However, the uniqueness of BCH can be understood as but one of the functions of CHM: providing a secure, stable, and authenticated data registration, searching, and retrieval mechanism in support of activities and obligations undertaken by Parties. If BCH is further understood as being part of CHM (per Article 20.1 of the Cartagena Protocol), then, in this narrow context, it becomes possible to delineate synergies between the two clearing-houses, especially relating to issues pertaining to bioinformatics in general.

Capacity Building

First, both clearing-houses implement activities to assist Parties in raising national and regional capacities. With regard to biosafety, capacities required are clearly associated with obligations under the Protocol, particularly the use and navigation of the BCH. CHM, again, operates within the entire framework of the Convention, including technical issues associated with BCH. The synergies, therefore, between these two clearing-houses deal primarily with technical issues, namely assisting countries and regions to gain adequate capacity in the acquisition, on-going support and use of new information technologies.

Formats, Protocols, and Standards

Work on recommendations pertaining to adoption and use of data formats, protocols, and standards is a concern held by CHM in relation to its general work on information exchange and in its development of a BCH that can function in and accommodate a distributed, interoperable data and information environment. CHM's work with formats, protocols, and standards was formalized by a decision at the fifth meeting of the COP, in which the Executive Secretary was requested to identify possible formats, protocols, and standards for improved exchange of biodiversity-related data, information, and knowledge, including national reports, biodiversity assessments, and Global Biodiversity Outlook reports, and convene an informal meeting on this issue (CBD 2000a).

With regard to the BCH, the meeting of the technical experts on BCH, held in Montreal, Canada (September 2000), recommended that the pilot phase of BCH be developed "...using a combination of centralized/decentralized information systems to offer the Biosafety Clearing-House the necessary flexibility for better coordination of the submission of data while ensuring timeliness and links to complementary distributed information" (CBD 2000b). This recommendation entailed formulation

of data formats, protocols, and standards, for use with BCH and for adoption by Parties desiring to make their national biosafety clearing-houses interoperable with BCH, ensuring that BCH can function in a distributed information environment.

Metadata

Recognition of the significance of metadata to the development of CHM was voiced by the CHM Informal Advisory Committee (November 2001), with the recommendation to use the Dublin Core as the metadata standard for the CBD web site, and to continue further development of controlled vocabularies and metadata standards. A core group was constituted to examine this issue and to report at an Informal Meeting on Formats, Protocols, and Standards for Improved Exchange of Biodiversityrelated Information (Montreal, February 2002). This meeting gave further support for use of the Dublin Core as the metadata standard for the CBD web site. This work on metadata was ported to the BCH.

Controlled Vocabularies

Use of controlled vocabularies by CHM or BCH national focal points becomes an essential tool in assisting Parties and Governments in making information interoperable. The Secretariat has made controlled vocabularies developed by CHM and BCH available through its website and through the CHM and BCH toolkits.

In many ways, these activities complement those of other organizations, such as the Global Biodiversity Information Facility (GBIF) and the Inter-American Biodiversity Information Network (IABIN), and reflect elements needed for effective use of bioinformatics as a support element throughout activities related to the program areas and cross-cutting issues of the Convention.

BIOINFORMATICS AND THE CONVENTION

It is within this context of promotion of technical and scientific cooperation, capacity-building, adherence to common formats, protocols and standards, and controlled vocabularies, that the CHM strives to assist Parties and Governments to make use of resources from bioinformatics initiatives and projects, GBIF's Data Portal being a case in point. Indeed, the need to search, retrieve, and analyze information hosted in very large distributed databases of biological data is increasingly important to the work of the COP. Note, for example, the increasing use of Web services, particularly XML and SOAP (Simple Object Access Protocol), to disseminate software applications that cannot be distributed as source code (Jamison 2003), to search and aggregate data from different databases (Stein 2002), and to assist in issues related to invasive alien species (Morris 2004). Discussion of a few activities, program areas, and cross-cutting issues of the Convention will illustrate these points.

In the Strategic Plan of the Convention, Parties emphasized the need for more effective and coherent implementation of the three objectives of the Convention, and agreed to reduce current rates of biodiversity loss significantly at global, regional, and national levels by 2010. Parties also adopted a framework to facilitate assessment of progress towards 2010 and communication of this assessment, to promote coherence among Convention work programs, and to provide a flexible framework within which national and regional targets may be set and indicators identified. The framework includes seven focal areas. The COP identified indicators for assessing progress towards the 2010 target at the global level, and goals and sub-targets for each focal area, as well as a general approach for integration of goals and sub-targets into Convention work programs.

Arguably, the challenge before the Parties regarding the 2010 target is how to quantify and measure existing biodiversity, and how to quantify its loss or conservation. Indeed, regarding measurement and quantification, one result of the meeting on "2010 – The Global Biodiversity Challenge" (London, May 2003) was:

"there is a need *to make the biodiversity data that exists more readily accessible* and available in a timely manner. Actions to achieve this would include:

- a) disseminating information in *appropriate formats* for potential users;
- b) using *best-practice* in information management and dissemination;
- c) supporting the development and implementation of tools, standards and protocols for *data exchange* that allow more effective sharing of information;
- d) establishing *inter-operable* electronic databases that allow for more effective integration of information from multiple sources in real time;
- e) improving use of the *Internet* as a tool for access and dissemination of biodiversity data, including increasing access to the Internet;

f) reviewing the *adequacy* of the existing data, assessing gaps and the action that needs to be taken to fill them."

(CBD 2004a)

The meeting also emphasized that "assessment is necessary of the *datasets currently available*, either through compilation of national-level data or through remote sensing, and of the processes for maintaining these data, in order to determine their potential value in addressing monitoring and assessment of achievement of the 2010 target, and their ability to contribute indicators" (CBD 2004a).

Such an exercise would necessitate assessment of relevant large-scale databases and review of issues related to their access. In addition, it would also demand adherence to common formats, protocols, and standards, as well as access to taxonomic databases. specimen and observational data databases, and remotely sensed data, all integrated via geographic information systems, analytical tools, and data mining programs. In all, these activities point to Schmidt et al.'s (2004) comment that the exponential growth of biological databases is establishing the need for high-performance computing (HPC) in bioinformatics to analyze distributed data more optimally. It may also illustrate the need for systems with search and analysis functions analagous to BLAST1 (basic local alignment search tool capable of searching databases for genes with similar nucleotide structure; Bayat 2004).

Regarding issues related to taxonomy and the Global Taxonomy Initiative (GTI), the COP pointed to the role of organizations and initiatives concerned with bioinformatics and taxonomy. For example, in its work program, it calls for facilitation of improved effective infrastructure for access to taxonomic information, with priority on ensuring access to information concerning elements of native biodiversity to countries of origin.

Principal actors identified include ECOPORT, GBIF, Species2000, Integrated Taxonomic Information System (ITIS), Tree of Life, ISIS, and BioNET International, as well as large-scale biosystematics research institutions and other stakeholders of taxonomic information, in collaboration with CHM. Such partnerships take on added urgency in light of new bioinformatics projects attempting to use DNA barcodes to identify specimens (Herbert et al. 2004). Arguably, establishment of an Internet-based "...public library of DNA barcodes linked to named specimens could provide a new master key for identifying species, one whose power will rise with increased taxon coverage and with faster, cheaper sequencing (Herbert et al. 2004)."

Similar emphasis on need for taxonomic data resources for identification and monitoring is inherent in decisions regarding invasive alien species. For example, COP requested the Executive

Table I: Decision VI/23, Paragraph 28: Alien species that threaten ecosystems, habitats or species.

- 1. Compilation and dissemination of case-studies submitted by Parties, other Governments and organizations, best practices and lessons learned, drawing upon, as appropriate, tools listed in information document UNEP/CBD/SBSTTA/6/INF/3 and the "Toolkit" compiled by the Global Invasive Species Programme;
- 2. Further compilation and preparation of anthologies of existing terminology used in international instruments relevant to invasive alien species, and to develop, and update as necessary, a non-legally binding list of terms most commonly used;
- 3. Compilation and making available lists of procedures for risk assessment/analysis and pathway analysis which may be relevant in assessing the risks of invasive alien species to biodiversity, habitats and ecosystems;
- 4. Identification and inventory of existing expertise relevant to the prevention, early detection and warning, eradication and/or control of invasive alien species, and restoration of invaded ecosystems and habitats, which may be made available to other countries, including the roster of experts for the Convention on Biological Diversity;
- 5. Development of databases and facilitated access to such information for all countries including repatriation of information to source countries, through, *inter alia*, the clearing-house mechanism;
- 6. Development of systems for reporting new invasions of alien species and the spread of alien species into new areas.

Secretary to support development and dissemination of technical tools and related information on prevention, early detection, monitoring, eradication, and/or control of invasive alien species. Moreover, the same decision requests several activities that would profit from an efficient global bioinformatics infrastructure (Table 1), particularly once it becomes

¹ <u>http://www.ncbi.nlm.nih.gov/BLAST/</u>.

possible to resolve synonyms of scientific names of species that refer to the same taxonomic concepts, an issue currently under discussion by GBIF.

The need for a well-functioning global bioinformatics infrastructure is explicit in other Convention-related program areas and cross-cutting issues. For example, Phase III of the proposed process for periodic assessment of status and trends of biological diversity in dry and sub-humid lands calls for data collection, processing, and communication according to agreed guidelines and mechanisms (CBD 2004c).

The work program on mountain biodiversity likewise points to the need "...[t]o improve the infrastructure for data and information management for accurate assessment and monitoring of mountain biological diversity and develop associated databases" and "...[e]nhance and improve the technical capacity at a national level to monitor mountain biological diversity, benefiting from the opportunities offered by the clearing-house mechanism of the Convention on Biological Diversity, including the development of associated databases as required at the global scale to facilitate exchange" (CBD 2004c).

An implicit recognition regarding the link between monitoring and need for better data access is the invitation to Parties by the COP to "[i]mprove and update national and regional databases on protected areas and consolidate the World Database on Protected Areas (WDPA) as key support mechanisms in the assessment and monitoring of protected area status and trends" (CBD 2004c).

The COP also recognized the important links between the mandate of the CHM and issues related to the transfer of technology. The work program on technology transfer is unequivocal on the role of the CHM as the "...central mechanism for the exchange of information on and facilitation of technology transfer and technical and scientific cooperation relevant for the Convention on Biological Diversity" (CBD 2004c). Furthermore, COP requested the Executive Secretary, in collaboration with Parties, the informal advisory committee of the CHM, and relevant organizations and initiatives, to work on "development of advice and guidance on the use of new information exchange formats, protocols and standards to enable interoperability among relevant existing systems of national and international information exchange, including technology and patent databases" (CBD 2004c).

WORKING RELATIONSHIPS WITH BIOINFORMATICS INITIATIVES AND ORGANIZATIONS

This concern for access to data in support of programme goals and activities of the Convention is reflected throughout other decisions from the seventh meeting of the COP. In response to these decisions and to ones from previous COPs, the Secretariat established memoranda of understanding with a number of bioinformatic initiatives and organizations with a view to facilitate access to data resources by Parties and Governments, repatriate information and enhance national and regional capacities with regard to the adherence to and use of common formats, protocols, and standards.

For example, a memorandum of understanding was completed with GBIF to establish a framework of collaboration between the CBD and GBIF secretariats to further common goals. These goals include facilitating development and implementation of approaches, technologies, and best practices necessary to access, share, and disseminate biodiversity data at the species, ecosystem, and genetic levels via the Internet. Furthermore, the CBD Secretariat is an *ex-officio* member in GBIF Governing Council, and participates in many of its working committees.

The potential impact of this working relationship should not be underestimated, given GBIF's development of its data portal; the Electronic Catalogue of Names of Known Organisms; common formats, protocols, and standards to facilitate interoperability; and its efforts to enhance national and regional capacities regarding bioinformatics. These projects are, in part, efforts to solve problems related to accessing and using data, or

"The scientific and data management communities have expressed a number of concerns in recent years regarding the collection and dissemination of data. Considerable amounts of data are still held by scientists or institutes that have not released their findings (Reynolds et al. 1997). Many data are generated to be analyzed once, published, and often not visited again (Gosz 1994). Data suppliers either are disconnected from wide-area networks, lack a standard mechanism for informing potential users (Walker et al. 1992), or prefer to advertise and distribute their products through more traditional means. The bottleneck in sharing information may be not knowing where to find useful information. In other cases, though information may exist on the network, no systematic directories exist to guide a user through the thousands of data sources."

(Xu et el. 2003).

A similar working relationship has been established with IABIN to facilitate development and implementation of technologies and best practices necessary to share knowledge and information relevant to biodiversity conservation and sustainable management. In response to the above, the Secretariat and IABIN are working to promote adoption of interoperability standards, share expertise. collaborate in development and implementation of programs pertaining to use of biodiversity information and management tools, and collaborate in enhancing national and regional capacities.

Another venue for collaboration regarding data exchange has been development of international thematic focal points under the CHM. To date, four international thematic focal points have been established: Birdlife International, the Global Invasive Species Programme (GISP), GTI, and NatureServe.

Several actions have been initiated as a result of these international thematic focal points. For example, the Secretariat is supportive of Birdlife International's efforts to develop a bi-national ecoregion-based clearing-house mechanism for the dry forests of Peru and Ecuador (the Tumbesian Endemic Bird Area). As well, the Secretariat is working with GISP in development of a global network of information on invasive species. NatureServe has made its databases of information available through the CHM, and is cooperating in enhancing national and regional capacities.

These few examples serve to illustrate the impact of bioinformatics in the work program of the Secretariat, and the role of CHM in facilitating and promoting bioinformatics at national and regional levels. Inarguably, access to and exchange of data and information are fundamental to the Convention's work program, and are key elements in ensuring development and enhancement of national and regional capacities.

CONCLUSION

The potential and impact of bioinformatics and the role of the CHM to the Convention's work programs and activities should not be underestimated. Bioinformatics ensures success of activities related to the three objectives of the Convention, and offers Parties and Governments means by which to implement obligations under the Convention more effectively, the need for accurate data to measure biodiversity loss in relation to the 2010 target being a case in point.

As technologies evolve, and more data and information become available for mining and analysis, bioinformatics will most likely continue to gain prominence and receive increased investment. This increase may indicate an increased role for CHM, including more active and collaborative work programs with bioinformatics organizations and initiatives, particularly initiatives to develop local access to regional and global information networks. Put another way, as stated by Laihonen et al. (2003), in their discussion of geospatially structured biodiversity information as a component of a regional CHM: "The regional viewpoint combined with the exploitation of geo-referenced information through modern information technology can offer opportunities lacking in coarse-grained national level CHMs. In further development of mechanisms for biodiversity information sharing, this should be seen as a resource enhancing our knowledge and wisdom on biodiversity at the national and global levels."

After all, an important premise of biodiversityrelated research is the need to examine variables under a holistic approach. This goal can be achieved only through a common and agreed upon strategy of information sharing; adherence to common formats, protocols, and standards; and equitable access to technologies and expertise. Such agreements are possible only when Parties, Governments, and regions have equitable access to technologies and knowledge. For these reasons, Parties should continue to strengthen and support the CHM as the global tool for biodiversity related technical and scientific cooperation, particularly in light of work requiring access to bioinformatics resources and expertise.

References

- Bayat, A. 2002. Science, medicine, and the future: Bioinformatics. British Medical Journal 324:1018-1022.
- Hebert, P. D. N., M. Y. Stoeckle, T. S. Zemlak, C. M. Francis. 2004. Identification of birds through DNA barcodes. PLoS Biology 2:e312.
- Jamison, D. Curtis. (2003). Open bioinformatics. Bioinformatics 19:679–680.
- Laihonen, P., M. Rönka, H. Tolvanen and R. Kalliola. 2003. Geospatially structured biodiversity information as a component of a

regional biodiversity clearing house. Biodiversity and Conservation 12:103-120.

- Morris, R. A. 2004. Web services: Who, what, why, where, and when? Pp. 37-41 *in* Meeting On Implementation of a Global Invasive Species Information Network: Proceedings of a Workshop (E. Sellers, A. Simpson, J. P. Fisher and S. Curd-Hetrick, eds.). National Biological Information Infrastructure, Reston, Virginia.
- Ouzounis, C. A. and A. Valencia. 2003. Early bioinformatics: The birth of a discipline—a personal view. Bioinformatics 19:2176-2190.
- Schmidt, B., L. Feng, A. Laud and Y. Santoso. 2004. Development of distributed bioinformatics applications with GMP. Concurrency and Computation: Practice and Experience 16:945-959.
- Scott, R. L. 2004. Bioinformatics: Essay review. Perspectives in Biology and Medicine 47:135-139.
- CBD. 1999. Strategic Plan of the Clearing-House Mechanism. (UNEP/CBD/COP/5/INF/3²)
- CBD. 2000a. Report of the Fifth Meeting of the Conference of the Parties to the Convention on Biological Diversity (UNEP/CBD/COP/5/23³).
- CBD. 2000b. Report of the Meeting of Technical Experts on the Biosafety Clearing-House (UNEP/CBD/BS/TE-BCH/1/5⁴).
- CBD. 2001. Report of the Intergovernmental Committee for the Cartagena Protocol on Biosafety on the Work of its First Meeting $(UNEP/CBD/ICCP/1/9^5)$.
- CBD. 2004a. Consideration of the results of the meeting on "2010–The Global Biodiversity Challenge" (UNEP/CBD/COP/7/INF/22⁶).
- CBD. 2004b. Report of the first meeting of the Conference of the Parties serving as the meeting of the Parties to the Protocol on Biosafety (UNEP/CBD/BS/COP-MOP/1/15⁷).

- CBD. 2004c. Report of the Seventh Meeting of the Conference of the Parties (UNEP/CBD/COP/7/21⁸)
- Stein, L. 2002. Creating a bioinformatics nation. Nature 417:119-120.
- Sugden, A. and E. Pennisi. 2000. Diversity digitized. Science 289:2305.
- Xu, H., D. Wang X. Sun. 1999. Biodiversity Clearing-House Mechanism in China: present status and future needs. Biodiversity and Conservation 9:361–378.

The views expressed in this paper are those of the author and do not reflect the official position of the Convention on Biological Diversity.

² <u>http://www.biodiv.org/doc/meetings/cop/cop-05/information/cop-05-inf-03-en.pdf</u>.

³ <u>http://www.biodiv.org/doc/meetings/cop/cop-05/official/cop-05-23-en.doc.</u>

⁴ http://www.biodiv.org/doc/meetings/bch/tebch-01/official/tebch-01-05en.pdf.

⁵ <u>http://www.biodiv.org/doc/meetings/bs/iccp-01/official/iccp-01-09-en.doc.</u>

⁶ http://www.biodiv.org/doc/meetings/cop/cop-07/information/cop-07inf-22-en.doc.

⁷ http://www.biodiv.org/doc/meetings/bs/mop-01/official/mop-01-15en.doc.

^{8 &}lt;u>http://www.biodiv.org/doc/meetings/cop/cop-07/official/cop-07-21-part1-en.doc.</u>