

Geographical Bioinformatics Systems

Tariq Rahim Soomro, Fujairah College, Fujairah, UAE, soomro@fc.ac.ae

Ghassan Al-Qaimari, Fujairah College, Fujairah, UAE, ghassan@fc.ac.ae

Hassan Wahba, Al-Ain University of Science & Technology, Al-Ain, UAE, hasanwahba@yahoo.co.uk

Abstract

Geographical Bioinformatics Systems (GBS) represent the marriage between two distinct and well established fields, GIS & Bioinformatics. Bioinformatics refer to the use of computers to handle biological information—that is, the use computers to store, compare, retrieve, analyze and predict the composition or structure of bio-molecules. On the other hand, the term Geographic Information Systems (GIS) refers to the computerized database management system for the capture, storage, retrieval, analysis, and display of spatial data. GIS is no longer restricted to the fields of computer science, geography, cartography or environmental science, but nowadays GIS tools and techniques are successfully used and implemented in many other fields of life, such as advertising, agribusiness, banking, entertainment, estate management, finance, food services, health services, insurance, manufacturing, pharmaceutical services, retail and utilities, social services, travel services and many more. This paper explores the merger of these two distinct fields GIS & Bioinformatics, and also explores the prospect for research and development in these two areas.

1. Introduction

In this paper we use the term Geographical Bioinformatics Systems (GBS) to describe the merger between GIS and Bioinformatics. Geographical Information Systems (GIS) refer to a complex technology, starting with the digital representation of landscapes captured by cameras, digitizers or scanners, in some cases transmitted by satellite, and then with the help of computer systems, storing, checking, manipulating and enhancing images to be analyzed and displayed as data referenced to the earth. GIS today is a well established field with many practical applications. The power of GIS lies not only in the ability to visualize spatial relationship, but beyond the space to a holistic view of the world with its many interconnected subsystems and complex relationship [1]. Bioinformatics on the other hand is an exciting and rapidly growing field that uses computer technology and information science to study biological information and structures. More specifically, it is the science of developing computer databases and algorithms to facilitate and expedite biological research, particularly in genomics [2]. In simple terms, it is an application of computer technology to the management of biological information. Computers here are used to gather,

store, analyze and integrate biological and genetic information.

The collaboration of these two distinct fields, GIS and Bioinformatics was initiated by Virginia Tech Office of GIS and Remote Sensing (OGIS) and the Virginia Bioinformatics Institute (VBI), in 2001, when a conference focused on the interface between GIS and Bioinformatics was organized to study the prospects of research in these two fields. The participants of conference strongly indorsed the merger between the two fields, as GIS and Bioinformatics have much in common, most notably digital maps, large databases, and research involving visualization, pattern recognition, and most significantly analysis of data [3].

This paper explores the future research in this new multidisciplinary field. The paper is organized as follow: section 2 reviews the applications of GBS; section 3 highlights future research trends in GBS; section 4 discusses future directions and concluding remarks.

2. Applications of GBS

The authors propose the term Geographical Bioinformatics Systems (GBS) to describe a new field of research and development that utilizes expertise from two well-established fields, GIS and Bioinformatics. The term GBS refers to the application of computer science in the management of biological information using the spatial and temporal maps and databases to collect and study the complex patterns and analyze its effects. Researchers in [3] agreed to use GIS applications to find and track large pattern, for example, geographic distribution of cancer and other diseases in human, animal and plant populations etc. They are also exploring the potential of using GIS to visualize and enhance the presentation of Bioinformatics data, and to identify the opportunities Bioinformatics presents to the GIS research community. They are also exploring from another perspective the use of Bioinformatics methodologies in GIS aiming at enhancing current GIS techniques, and identifying new approaches to pattern recognition and data analysis that could be used specifically for GBS purposes.

The following is a short review of GIS applications that analyze and visualize Bioinformatics data to help Bioinformatics experts study biological

phenomena. These applications represent the solution for various Bioinformatics databases from GIS perspectives.

READ-IT (Research, Education, and Dissemination via Information Technology) is a Bioinformatics/GIS program aimed at protecting chimpanzees. It is a system for the integration of technology; research and education, through a project designed to study and protect chimpanzees in Tanzania. This research is designed to develop “interesting and compelling content for integrated research and educational opportunities for undergraduate, graduate and professional students at Virginia Tech.” Some of those are involved with the “Bush to Base Bio-Informatics” and the GIS components of the program. In this program, field researchers gather physiological data on the chimps, and code and process the data using handheld modules with Global Positioning Systems (GPS) capability. The modules are designed to interface with a Virginia Tech-based Web-enabled server designed to share information with authorized users [4].

Micro-Mar is a database designed especially for dynamic representation of marine microbial biodiversity. It is created to collect DNA diversity information from marine prokaryotes for biogeographical and ecological analysis. The database aims to integrate molecular data and taxonomic affiliation with biogeographical and ecological features that will allow experts to have dynamic representation of the marine microbial diversity embedded in a user friendly web interface. In this program, GIS option provides an interface for selecting a particular sampling location on the world map and getting all the sequences from the location and their details [5].

GenoSIS (Genome Data Interpretation Using GIS) is a genome spatial information system based on the idea of “spatial genomics”, applying concepts and tools of spatial analysis and GIS to the interpretation of genome data. It uses “off the shelf” GIS software, for example, ESRI’s ArcGIS and Oracle Spatial, to reuse existing spatial analysis, classification, querying, and visualization tools for genome data analysis [6].

GEOBASE is a simple geographical information system designed for personal computers. This system allows representation and elementary analysis of geographically coded information, and handles two kinds of data: maps and facts, where map data describe the basis on which the fact data are located [7].

Gene Information System is a biomedical text-mining system for gene information discovery. It is

a biomedical text-mining system focused on four types of gene-related information: biological functions, associated diseases, related genes and gene-gene relations. The aim of this system is to provide researchers with a user-friendly bio-information service [8].

Fungi is designed at University of Michigan to explore spatial relationships between specimen data and other data using a GIS as one of the bioinformatics method. This project makes it possible to create maps showing the relationship between the location of specimens and political units (states and counties of USA), major highways, major lakes and rivers, and federal land ownership [9].

Map Viewer is a tool designed at the National Center for Biotechnology Information (NCBI) that allows a user to view an organism’s complete genome, integrated maps for each chromosome (when available), and/or sequence data for a genomic region of interest [10].

MultiStore is a research infrastructure designed at University of Buffalo for supporting integrated research in specific targeted areas of computer science in the fields including Multimedia, Visualization, GIS and Bioinformatics. The research objectives is to develop computational theories and algorithms for storing, managing, analyzing, querying, and visualizing multi-dimensional data sets that are generated from the related fields [11].

ETI Projects at ETI Bioinformatics, in Amsterdam, is an NGO in operational relations with UNESCO, participating in two GIS projects funded by NWO, the Dutch national science foundation. The first project records and identifies the impact of global change on the biological diversity of the North Sea and second project monitors climate changes in the Indonesian coral reef biotas. For these projects, ETI is building databases, programming queries and preparing maps and map layers, using MapIt software package [12].

The majority of the above applications in Bioinformatics are in the form of sequence analysis (e.g. DNA), genome annotation, computational evolutionary biology, measuring biodiversity, analysis of gene expression, analysis of regulations, analysis of protein expression, analysis of mutations in cancer, prediction of protein structure, comparative genomics, modeling biological systems, high-throughput image analysis and protein-protein docking; all these research areas have analysis, prediction, comparative studies and modeling. On the other hand, GIS is basically a spatial database software tool, which has also analysis, prediction, comparison and modeling

capabilities. Given that there are many cases where the spatial organization of genome features has been shown to have biological significance; this consequently leads researchers to consider extending GIS to handle Bioinformatics. Recent studies related to gene finding and genome annotation, provide evidence of complex spatial interrelationships of genome features, such as genes being alternatively spliced, or nested, or overlapping. In addition, conservation of gene order in microbes has been a subject of a great deal of analysis [6]. An example of the this integration trend between the two fields is the recent work to use “predictive methods” for the analysis of complex medical data in relation to the environment [13].

3. Future of GBS

Bioinformatics is a rising field that seeks to apply the tools and techniques of computer science to the management and analysis of biological data. Bioinformatics involves the incorporation of computers, software tools, and databases in an effort to address biological questions. Bioinformatics approaches are often used for major initiatives that generate large data sets. Bioinformatics is a new science and a new way of thinking that could potentially lead to many relevant biological discoveries. The future of GBS is in the integration of multiple Bioinformatics databases and in the utilization of the visualization techniques available in GIS for the purpose of identifying biological trends and relationships. For example, the integration of a wide variety of data sources such as clinical and genomic data will allow specialists to use disease symptoms to predict genetic mutations and vice versa. Another future area of research in GBS is large-scale comparative genomics. For example, the development of tools that can do multiple ways of comparisons between genomes will push forward the discovery rate in bioinformatics. Along these lines, the modeling and visualization of full networks of complex systems could be used in the future to predict how the system reacts to a drug, for example [14].

There are many aspects where intelligent GIS techniques played an important role when utilized in existing Bioinformatics system to compute accurate, best, error-free, and fast results [15]. The trend will continue toward building intelligent GBS systems that incorporate data mining, collaborative agents, and sophisticated visualization techniques to analyze, predict and present information more effectively.

Geographical Bioinformatics Systems (GBS) will allow researchers to do a far better job of monitoring, quantifying, and predicting the impact

environment phenomena (such as climate changes, emerging and resurgent infectious diseases) has on human-health. In public health, GIS can play a vital role to resolve issues that required spatial analysis and spatial attention [16].

GIS, which emerged several decades ago, has matured to the point that the incorporation of its intelligent techniques in GBS will potentially help scientists to effectively track diseases in populations. The integration of Bioinformatics data, genomics and proteomics data will be telling the story of what takes place at the cell and sub-cell level in individual diseased organisms and will soon enable epidemiologists using GIS techniques to capture the how of disease outbreaks. Both fields are heavily relying on mining, managing, accessing, and analyzing large amounts of data [3].

4. Conclusion

GIS and Bioinformatics share some common characteristics, most importantly is the fact that their applications relate to environment, that they are characterized by large dedicated databases; and that most of their data manipulation takes place using widely adopted and well-supported commercial or open source applications [17]. Since disease mapping is a major focus of spatial epidemiology, we believe the future role of GIS will become more crucial in handling disease outbreak surveillance systems and in helping decision makers interpret and analyze both, routine- and outbreak-related health data. That is, it will help to provide insight into possible causes of diseases, clusters of disease outbreaks across particular geographic areas, and the evolution of disease outbreaks. GIS will effectively help scientists interpret disease patterns with the help of built-in query capabilities, and they will be able to select different data layers to query for disease-related information by selecting geographical areas of interest [18].

The Bioinformatics arena is very broad and encompasses many problems, such as finding gene sequences, construction of molecular pathway, prediction of protein structure etc. Extending Bioinformatics applications with a GIS interfaces, with their spatial tools and techniques, enable these applications to overcome their limitations and enhances their ability to visualize data. GBS is a hybrid and emerging field of research and development, capable of handling spatial biological queries, and analyzing data for the purpose of modeling and predicting disease behavior and other biological phenomena.

5. References

- [1] Designing Issues of Components-Based ZGISObjects, Tariq Rahim Soomro, Kougen Zheng and Yunhe Pan, 4th International Conference on Computational Intelligence & Multimedia Applications, 2001, Proceeding by IEEE Computer Society.
<http://csdl.computer.org/comp/proceedings/iccima/2001/1312/00/13120338abs.htm>
- [2] UNO Bioinformatics Home page, Retrieved April 2008,
<http://www.unomaha.edu/bioinformatics/>
- [3] VBI Virginia Tech Conference Puts New Research Area on the Map, GIS Expert Michael Goodchild Echos Its Value, 2001, Retrieved April 2008,
www.vbi.vt.edu/pr/press_releases/press_20010628_news_gisexpert.htm
- [4] Bioinformatics/GIS program aim is to protect chimpanzees, Science Blog, 2003, Retrieved April 2008,
www.scienceblog.com/community/older/2003/A/20037439.html
- [5] Micro-Mar: a database for dynamic representation of marine microbial biodiversity, Ravindra Pushker, Giuseppe D'Auria, Jose Carlos Alba-Casado and Francisco Rodrigues-Valera, Journal of BioMed Central (BMC) Bioinformatics 2005, Retrieved April 2008,
<http://www.biomedcentral.com/1471-2105/6/222>
- [6] GenoSIS: Genome Data Interpretation Using GIS, Mary E. Dolan, Constance Holden, M Kate Beard, and Carol J, Bult, 2002, Retrieved April 2008,
<http://gis.esri.com/library/userconf/proc02/pap0719/p0719.htm>
- [7] GEOBASE: a simple geographical information system on a personal computer, Philippe Blaise and Cesare Gessler, Oxford University Press, Oxford Journals of Bioinformatics 1990, Retrieved April 2008,
<http://bioinformatics.oxfordjournals.org/cgi/content/abstract/7/2/155>
- [8] GIS: a biomedical Text-mining system for gene information discovery, Jung-Hsien Chiang, Hsu-Chun Yu and Huai-Jen Hsu, Oxford University Press, Oxford Journals of Bioinformatics 2004, Retrieved April 2008,
<http://bioinformatics.oxfordjournals.org/cgi/content/abstract/20/1/120>
- [9] Fun facts about Fungi, Robert Fogel, Ivins, 2000, Retrieved April 2008,
<http://herbarium.usu.edu/fungi/FunFacts/bioinform.htm>
- [10] A Science Primer, Just the facts: a basic introduction to the science underlying NCBI resources, 2004, Retrieved April 2008,
www.ncbi.nlm.nih.gov/About/primer/bioinformatics.html
- [11] Overview, Multimedia Database Research Group, University of Buffalo, Retrieved April 2008,
www.cse.buffalo.edu/DBGROUP/Infrastructure/overview.html
- [12] ETI Bioinformatics – Geographical information systems (GIS), Retrieved April 2008,
www.eti.uva.nl/Products/gis.php
- [13] MPBA, Machine Learning, GIS and Bioinformatics, Retrieved April 2008,
<http://mpa.itc.it/>
- [14] What is Bioinformatics, Joanne Fox (UBC Bioinformatics Centre), Retrieved April 2008,
http://bioinformatics.ubc.ca/about/what_is_bioinformatics/
- [15] GIS: A weapon to combat the crime, Tariq Rahim Soomro, Syed Mehmood Raza Naqvi and Zheng Kougen, 5th World Multiconference on Systemics, Cybernetics and Informatics (SCI-2001) and 7th International Conference on Information Systems Analysis and Synthesis (ISAS-2001), July 22-25, 2001, Orlando, USA.
<http://200.75.139.138:1081/ProceedingSCI/volume1/2001.htm>
- [16] Public Health GIS News and Information, 2007, Retrieved April 2008,
http://www.cdc.gov/nchs/about/otheract/gis/gis_publichealthinfo.htm
- [17] Cyberinfrastructure, Data, and Libraries, Part 2, Libraries and the Data Challenge: Roles and Actions for Libraries, Anna Gold, Massachusetts Institute of Technology, D-Lib Magazine, September/October 2007, Volume 13 Number 9/10, ISSN 1082-9873,
<http://www.dlib.org/dlib/september07/gold/09gold-pt2.html>
- [18] Real-Time Syndromic Surveillance, 2006, Retrieved April 2008,
<http://www.esri.com/news/arcuser/0206/geostat2of2.html>

Copyright © 2008 by the International Business Information Management Association (IBIMA). All rights reserved. Authors retain copyright for their manuscripts and provide this journal with a publication permission agreement as a part of IBIMA copyright agreement. IBIMA may not necessarily agree with the content of the manuscript. The content and proofreading of this manuscript as well as and any errors are the sole responsibility of its author(s). No part or all of this work should be copied or reproduced in digital, hard, or any other format for commercial use without written permission. To purchase reprints of this article please e-mail: admin@ibima.org.