Advances in Life Science and Technology
ISSN 2224-7181 (Paper) ISSN 2225-062X (Online)
Vol.28, 2015

The Role and Application of Bioinformatics in Plant Disease Management

Kumlachew Alemu
Department of Plant Science, Assosa University, P.O.Box 18, Assosa, Ethiopia
E-mail of corresponding Author: kum.alemu@gmail.com

Abstract
In recent years, rapid developments in genomics and proteomics have generated a large amount of biological data. Drawing conclusions from these data requires sophisticated computational analyses. Bioinformatics is the interdisciplinary science of interpreting biological data using information technology and computer science. The importance of this new field of inquiry will grow as we continue to generate and integrate large quantities of genomic, proteomic, and other data. As the amount of data grows exponentially, there is a parallel growth in the demand for tools and methods in data management, visualization, integration, analysis, modeling and prediction. Bioinformatics plays an essential role in today's plant pathology with regards to the development of new plant diagnostic tools. Pathogen is among the traits considered in the primary interest of plant bioinformatics. The contribution of bioinformatics advances made possible the mapping of the entire genomes of many organisms in just over a decade. These discoveries, along with current efforts to determine gene and protein functions, have improved the ability to understand the root causes of plant diseases and find new cures. Furthermore, many future bioinformatics innovations will likely be spurred by the data and analysis demands of the life sciences. Bioinformatics have many practical applications in current plant disease management with respect to the study of host-pathogen interactions, understanding the disease genetics and pathogencity factor of a pathogen which ultimately help in designing best management options. This review paper describes some of the key concepts and databases used in bioinformatics, with an emphasis on those relevant to plant science. It also covers some aspects with regards to the role application of this endeavor science in today's plant disease management strategies.
Keywords: Bioinformatics, Database, Diagnostic tool, Pathogen, Plant disease,

1. Introduction

Over the past few decades, major advances in the field of molecular biology, coupled with advances in genomic technologies, have led to an explosive growth in the biological information generated by the scientific community (Vassilev *et al.,* 2006).Bioinformatics is an emerging interdisciplinary area of science and technology encompassing a systematic development and application of information technology solutions to handle biological information by addressing biological data collection and warehousing, data mining, database searches, analyses and interpretation, modeling and product design (Degrave *et al*., 2002; Xiong, 2009; Jayaram and Dhingra, 2010).

Being an interface between modern biology and informatics it involves discovery, development and implementation of computational algorithms and software tools that facilitate an understanding of the biological processes with the goal to serve primarily agriculture and healthcare sectors with several spinoffs (Jayaram and Dhingra, 2010). Bioinformatics is more of a tool than a discipline, the tools for analysis of biological data (Kumor, 2009).

Recent advances in genomic technologies have led to an explosion of data and a huge growth in bioinformatics within both plant biotechnology and the broader biomedical sciences (Edwards and Batley, 2004). In a developing country, bioinformatics has a key role to play in areas like agriculture where it can be used for increasing the nutritional content, increasing the volume of the agricultural produce and implanting disease resistance (Jayaram and Dhingra, 2010).

The ultimate goal of the field is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be discerned. That is to create better understanding of a living cell and how it functions at the molecular level (Autenrieth *et al*., 2005). By analyzing raw molecular sequence and structural data, bioinformatics research can generate new insights and provide a global perspective of the cell (Xiong, 2009).

Bioinformatics is a rapidly developing branch of biology and is highly interdisciplinary, using techniques and concepts from informatics, statistics, mathematics, chemistry, biochemistry, physics, and linguistics. It has many practical applications in different areas of biology and medicine. The common activities in bioinformatics includes, mapping and analyzing DNA and protein sequences, aligning different DNA and protein sequences to compare them and creating and viewing 3-D models of protein structures (Luscombe *et al.,* 2001; Autenrieth *et al*., 2005;Nilges and Linge, 2009).

The purpose of bioinformatics is to understand the genetic and molecular basis of all biological processes in plants that are relevant to the specie. This understanding is fundamental to allow efficient exploitation of plants as biological resources in the development of new cultivars with improved quality and reduced economic and environmental costs. This knowledge is also vital for the development of new plant diagnostic tools. Traits considered of primary interest are, pathogen and abiotic stress resistance, quality traits for plant, and reproductive traits determining yield (Koltai and Volpin, 2003). Bioinformatics also play a significant role in diagnostic pathology which paves the way towards better understanding of plant-pathogen biological network, which ultimately help in management of the disease (Koltai and Volpin, 2003). This paper reports recent advancement in molecular diagnosis of plant disease in particular to see the application of bioinformatics in detection and diagnosis of plant pathogenic microorganisms.

2. Scope of Bioinformatics

Bioinformatics has emerged as an important discipline within the biological sciences that allows scientists to decipher and manage the vast quantities of data that are now available (Ojo and Maxwell, 2010). It consists of two subfields: the development of computational tools and databases, and the application of these tools and databases in generating biological knowledge to better understand living systems (Xiong, 2009). These two subfields are complementary to each other. The tool development includes writing software for sequence, structural, and functional analysis, as well as the construction and crating of biological databases. These tools are used in three areas of genomic and molecular biological research, including molecular sequence analysis, molecular structural analysis and molecular functional analysis (Tramontano, 2009; Xiong, 2009).

The areas of sequence analysis includes sequence alignment, sequence database searching, motif and pattern discovery, gene and promoter finding, reconstruction of evolutionary relationships and genome assembly and comparison. Structural analyses include protein and nucleic acid structure analysis, comparison, classification and prediction. The functional analysis includes gene expression profiling, protein–protein interaction prediction, protein sub cellular localization prediction, metabolic pathway reconstruction and simulation (Rao *et al.,* 2008; Xiong, 2009).The three aspects of bioinformatics analysis are not isolated, but often interact to produce integrated results. For example, protein structure prediction depends on sequence alignment data; clustering of gene expression profiles requires the use of phylogenetic tree construction methods derived in sequence analysis. Sequence based promoter prediction is related to functional analysis of co-expressed genes. Gene annotation involves a number of activities, which include distinction between coding and non-coding sequences, identification of translated protein sequences and determination of the gene's evolutionary relationship with other known genes (Xiong, 2009).

3. Importance of Bioinformatics

Bioinformatics apply the principles of information sciences and technologies to complex life science data (Ojo and Maxwell, 2010). In order to study how normal cellular activities are altered in different disease states, the biological data must be combined to form a comprehensive picture of these activities. Therefore, the field of bioinformatics has evolved such that the most pressing task now involves the analysis and interpretation of various types of data, including nucleotide and amino acid sequences, protein domains, and protein structures (Wales, 2009).

The analysis of nucleic acid sequences, protein structure/function relationships, genome organization, regulation of gene expression, interaction of proteins and mechanisms of physiological functions, can all benefit from a bioinformatics approach and cannot be achieved without it. Nucleic acid and protein sequence data from many different species and from population samplings provides a foundation for studies leading to new understandings of evolution and the natural history of life (Tramontano, 2009).

The tasks of Bioinformatics involve the analysis of sequence information. This involves the following activities; identifying the genes in the DNA sequences from various organisms, identifying families of related sequences and the development of models, aligning similar sequences and generating phylogenetic trees to examine evolutionary relationships, Finding all the genes and proteins of a genome from a given sequence of amino acids and predicting active sites in the protein structures to attach drug molecules (Rao *et al.,* 2008). According to Kumor (2009) and Asthana (2009), there are three important sub-disciplines within bioinformatics: the development of new algorithms and statistics with which to assess relationships among members of large data sets; the analysis and interpretation of various types of data including nucleotide and amino acid sequences, protein domains, and protein structures; and the development and implementation of tools that enable efficient access and management of different types of information (Rao *et al.,* 2008).

4. Bioinformatics Tool

4.1. Biological databases

Biological databases are archives of consistent data that are stored in a uniform and efficient manner. These databases contain data from a broad spectrum of molecular biology areas. A simple database might be a single file containing many records, each of which includes the same set of information (Xiong, 2009). Databases are composed of computer hardware and software for data management. The chief objective of the development of a database is to organize data in a set of structured records to enable easy retrieval of information. Each record, also called an *entry*, should contain a number of fields that hold the actual data items. To retrieve a particular record from the database, a user can specify a particular piece of information, called *value*, to be found in a particular field and expect the computer to retrieve the whole data record. This process is called making a query. Although data retrieval is the main purpose of all databases, biological have a higher level of requirement, known as knowledge discovery, which refers to the identification of connections between pieces of information that were not known when the information was first entered (Xiong, 2009).

There are two kinds of database; these are primary and secondary database. Primary or archived databases contain information and annotation of DNA and protein sequences, DNA and protein structures and DNA and protein expression profiles. Secondary or derived databases are so called because they contain the results of analysis on the primary resources including information on sequence patterns or motifs, variants and mutations and evolutionary relationships (Rao *et al.,* 2008).

4.2. Software and tools

Software tools for bioinformatics range from simple command-line tools, to more complex graphical programs and standalone web-services available from various bioinformatics companies or public institutions. BLAST (Basic Local Alignment Sequence Tool) is one of a number of generally available programs for doing sequence alignment. It remains the fastest means by which to identify specific sequences in large datasets and enables the rapid annotation of novel sequences. Although BLAST is the standard tool for identifying sequence similarities in large datasets, there are several options for assembling sequence datasets, the choice of which depends on hardware availability, dataset size, data format, structure and the genetic structure of the organism (Edwards and Batley, 2004).

5. Bioinformatics in Plant Disease Management

Genomic studies focused on whole genome analysis, have opened up a new era for biology in general and for agriculture in particular. Along with the use of genetic plant models and the progress in sequencing agriculturally important organisms, the combination of bioinformatics and functional genomics globally enhance agricultural genomics. These studies are likely to pave the way towards better understanding of plant-pathogen biological network, and eventually to lead to break thoughts in promotion of plant resistance to pests (Koltai and Volpin, 2003). Bioinformatics plays a great role in plant disease management. Some of the areas in which bioinformatics is involved in plant diseases management are indicated below.

5.1. Study of host- pathogen interaction

The interaction between plants and their pathogens is complex. The study of these interactions has a long and rich history in science, with plant pathologists tackling these complex systems first with classical tools, such as physiology, histology, microbiology, plant breeding and genetics, and more recently with advanced biochemistry and molecular biology approaches (Ojo and Maxwell, 2010). Plant pathogens have evolved a broad set of proteins that enable a stealthy entry into the plant cell and facilitate the evasion of host defenses (Vencato *et al*., 2006). Among other defenses, plants have evolved a series of proteins that monitor their cells for signs of infection. Downstream of these monitors is a signaling and response system triggered upon infection (Rao *et al.,* 2008).

The molecular basis of the host-pathogen interaction is now much better understood, as a result of the development of genomic data and tools (Koltai and Volpin, 2003). For example, the complete genomic sequence is available for a model plant, Arabidopsis, and for one of its bacterial pathogens, *Pseudomonas syringae*pv. tomato DC3000. Detailed molecular analyses of these two organisms have revealed much about plant defenses. Modern genomics tools, including applications of bioinformatics and functional genomics, allow scientists to interpret DNA sequence data and test hypotheses on a broader scale than previously possible (Anonymous, 2005). In the last 5 to 10 years, many of the critical host proteins that detect the presence of pathogens have been characterized. Numerous components of the plant signaling system have also been identified that function downstream of the detection molecules. In parallel, the pathogen proteins that are used to suppress host defenses and drive the infection process (so-called effector proteins) have also been identified, using molecular biological technologies and genetics (Anonymous, 2005).

5.2. Study of Disease genetics

Advances in molecular biology, plant pathology, and biotechnology have made the development of different techniques are designed to detect plant diseases early, either by identifying the presence of the pathogen in the plant (by testing for the presence of pathogen DNA) or the molecules (proteins) produced by either the pathogen or the plant during infection. These techniques require minimal processing time and are more accurate in identifying pathogens (Koltai and Volpin, 2003). In the past, disease genetics has focused on monogenic diseases in which the expression of a particular variant of a single gene will, in the vast majority of cases, lead to disease. There are innumerable monogenic diseases, each of which affects only a small number of individuals. In contrast, disease genetics research is now focused on identification of genes associated with common diseases. These common diseases are multi-factorial (i.e. dependent on complex interactions between numerous environmental factors and a number of alternative forms of genes called disease susceptibility genes) and polygenic (involving more than one gene in their multi-factorial pathogenesis (Middleton, 2000). The overall goal of disease genetics is to identify how genetic variation can influence disease susceptibility and to improve understanding of the molecular processes resulting in clinically overt disease. New treatments can then be designed to target these molecular processes to prevent and treat the disease (Rao *et al.,* 2008).

Typically, new disease susceptibility genes have been identified using a combination of linkage and association studies. The linkage studies involve collection of DNA samples and extensive clinical phenotypic data from multiple members of affected families. Markers are typed throughout the genome, and using linkage analysis algorithms, chromosomal regions harboring disease genes are identified (Rao *et al.,* 2008). The regions are identified using highly informative markers on the basis of their chromosomal location by taking advantage of the meiotic process of recombination as apparent in families segregating for the disease. Markers closest to the disease gene show the strongest correlation with disease patterns in families (Kruglyak, 1999). These linkage studies allow identification of a region on a chromosome and large portions of the DNA that may be linked to a specific disease (Rao *et al.,* 2008).

5.3. Identification of pathogencity factors of a pathogen

Pathogensity is the capacity of a pathogen to overcome genetically determined host resistance. The identification of genes involved in host–pathogen interactions is important for the elucidation of mechanisms of disease resistance and host susceptibility. A traditional way to classify the origin of genes sampled from a pool of mixed cDNA is through sequence similarity to known genes from either the pathogen or host organism or other closely related species. This approach does not work when the identified sequence has no close homologues in the sequence databases (Emmersen *et al.,* 2007).

It was reported that bioinformatics play a role in identification of the HrpL region and type III Secretion System effector proteins of *Pseudomonas syringae*pv. *phaseolicola*1448A. The HrpL alternative sigma factor activates the expression of multiple genes that are essential to the plant pathogencity of *Pseudomonas syringae* (Willis and Kinscherf, 2004). The most important of these are genes encoding the type III secretion system (T3SS) and effector proteins that are injected by the T3SS into host cells. The T3SS is encoded by hypersensitive response and pathogencity (*hrp*) and *hrp*conserved (*hrc*) genes. Effector proteins are encoded by avirulence (*avr*) and Hrp outer protein (*hop*) genes, whose different names reflect the phenotype used to discover them (Lindeberg *et al*., 2005). Mutations were constructed in the DC3000 homologs and found to reduce bacterial growth in host These results establish the utility of the bioinformatic or candidate gene approach to identifying effectors and other genes relevant to pathogenesis in pathogen genomes (Vencato *et al*., 2006).Similarly, it was reported that identification and characterization of potato protease inhibitors able to inhibit pathogencity and growth of *Botrytis cinerea* (Hermosa *et al.,* 2006)

5.4. Developing disease resistance cultivars

Durable disease resistance is a major but elusive goal of many crop improvement programs. Genomic approaches will have a significant impact on efforts to ameliorate plant diseases by increasing the definition of and access to gene pools available for crop improvement. This approach will involve the detailed characterization of the many genes that confer resistance, as well as technologies for the precise manipulation and deployment of resistance genes. Genomic studies on pathogens are providing an understanding of the molecular basis of specificity and the opportunity to select targets for more durable resistance (Michelmore, 2003). As goal of plant genomics is to understand the genetic and molecular basis of all biological processes in plants that are relevant to the species. This understanding is fundamental to allow efficient exploitation of plants as biological resources in the development of new cultivars with improved quality and reduced economic and environmental costs. This knowledge is also vital for the development of new plant diagnostic tools Traits considered of primary interest are, pathogen and abiotic stress resistance, quality traits for plant, and reproductive traits determining yield (Vassilev *et al.,* 2006). A genome program can now be envisioned as a highly important tool for plant improvement (Vassilev *et al.,* 2005). Such an approach to identify key genes and understand their function will result in a quantum leap in plant improvement. Additionally, the ability to examine gene expression will allow us to understand how plants respond to and interact with the physical environment and management practices (Vassilev *et al.,* 2006). Bioinformatics plays several roles in breeding for disease resistance. It will be important for acquiring and organizing large amounts of information. It will also allow the visualization of information from heterogeneous datasets to facilitate the selection of superior individuals.

Disease resistance is only one of several traits under selection in a breeding program. Bioinformatics will therefore play an increasing role in integrating phenotypic and pedigree information for agronomic as well as resistance traits (Vassilev *et al.,* 2006). Improved algorithms and increased computing power will make it possible to simulate and optimize selection strategies as well as to model the epidemiology of pathogens (Michelmore, 2003). According to Vassilev *et al.* (2005) the key role of bioinformatics for plant improvements may include: to encourage the submission of all sequence data into the public domain, through repositories, to provide rational annotation of genes, proteins and phenotypes, and to elaborate relationships both within the plants' data and between plants and other organisms to provide data including sequence information, information on mutations, markers, maps, functional discoveries, and other.

5.5. Production of disease free planting materials

Most agricultural biotechnologies involve tissue culture and DNA-based markers for germplasm conservation, production of disease-free planting material, and assistance to genetic improvement. More recently, Latin America countries such as, Argentina, Brazil, Colombia, Honduras, Mexico and Uruguay have commercially grown transgenic crops. Advanced biotechnologies, such as genetic sequencing and microarray genomics, are differentially utilized in some Latin America countries, with Brazil being at the forefront, for characterization, mapping, and trait screening for important crops and pathogens. With this regard the role of bioinformatics is indispensable (Roca *et al*., 2004).

Conclusion

The contribution of bioinformatics to the development of plant functional genomics is already very apparent, as is its potential in other areas of fundamental plant cell biology. It's role in development of related areas such as crop protection, detection of allergenicity of genetically modified crops and systematic indispensable. With the increasingly large amounts of biological data, integration with information technology has become essential. Originally started as a specialty for storage of data and as a tool kit for analyzing data, bioinformatics now encompasses many emerging areas like, evolutionary studies, protein structure-function prediction, gene expression studies etc. It may not be long before bioinformatics becomes a hypothesis driven molecular science bridging the gap between the genome and the organism, with data providing a platform for validation and new product development. The major challenge for the plant science community over the next few years is the need to extend genomics from models to crops. In particular, a priority is the integration of genomic and agronomic data, including the disease management aspects, are important to in increasing the productivity of the crop per unit area. Few concerted efforts have been made to harness the large historic repositories of information relating to crop phenotypic traits available from variety trailing and the literature, to provide a functional link with the underlying genomics. Despite the pitfalls, there is no doubt that bioinformatics is a field that holds great potential for revolutionizing biological research in the coming decades.

10. References

Anonymous, (2005). Pseudomonas versus Arabidopsis: models for genomic research into plant disease resistance.www.actionbioscience.org.

Asthana, V. (2009). Bioinformatics at Present: Description and disciplines in bioinformatics.http://bionet.awardspace.info/index.htm

Degrave .W, Huynh, C., Roos, D., Oduola, A. and More, C. M. (2002.Bioinformatics for Disease Endemic Countries: Opportunities and Challenges in Science and Technology Development for Health.IUPA.

Edwards, D. and Batley, J. (2004). Plant bioinformatics: from genome to Phenome. *Biotechnology* 22 (5): 232-237.

Emmersen, J., Rudd, S. Mewes, H.W. and Tetko, I. V. 2007.Separation of sequences from host–pathogen interface using triplet nucleotide frequencies*. Fungal Genetics and Biology* 44:231-241.

Hermosa, M.R., Turra, D., Fogliano, V., Monte, E. and Lorito, M. (2006).Identification and characterization of potato protease inhibitors able to inhibit pathogenicity and growth of *Botrytis cinerea*. *Physiological and Molecular Plant Pathology* 68: 138–148.

Jayaram, B. and Dhingra, P. (2010).Bioinformatics for a better tomorrow. Indian Institute of Technology, HauzKhas, New Delhi - 110016, India.

Koltai,H. and Volpin, H. (2003). Agricultural genomics: an approach to plant protection. *European journal of plant pathology* 109: 101-108.

Kumor, S. (2009). Introduction to bioinformatics, scope and potential of bioinformatics. http://www.bioinformaticsweb.org/catagory/bioinformatics.html.

Kruglyak, L. (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics* 22: 139-144.

Lindeberg, M., Stavrinides, J., Chang, J.H., Alfano, J. R., Collmer, A., Dangl, J. L., Greenberg, J. T., Mansfield, J.W., and Guttman, D.S. (2005). Proposed guidelines for a unified nomenclature and phylogenetic analysis of type III Hop effector proteins in the plant pathogen *Pseudomonas syringae*. Mol. Plant-Microbe Interact. 18:275-282.

Luscombe, N. M., Greenbaum, D. and Gerstein, M. (2001). What is Bioinformatics? A Proposed Definition and Overview of the Field. New Haven, USA.

Michelmore, R. W. (2003). The impact zone: genomics and breeding for durable disease resistance. *Current Opinion in Plant Biology* 6:397–404.

Middleton, H, (2000). From gene-specific tests to pharmacogenetics.*Community Genetetics*3: 198-203.

Nilges, M. and Linge, J.P. (2009). Bio–Informatique Structurale. Institut Pasteur, F–75015 Paris, France.

Ojo, O.O. and Maxwell, O. (2010).Incorporating bioinformatics into biological science education in Nigeria: prospects and challenges, *Infection, Genetics and Evolution*: 10.1016/j.meegid2010.11.015.

Rao, V. S., Das, S.K.., Rao, V.J and Srinubabu, G. (2008).Recent developments in life sciences research: Role of Bioinformatics. *African Journal of Biotechnology* 7 (5): 495-503.

Roca, W., Espinoza, C. and Panta, A. (2004).Agricultural Applications of Biotechnology and the Potential for Biodiversity Valorization in Latin America and the Caribbean. *Journal of Agro Biotechnology Management and Economics*7(1&2): 13-22.

Tramontano, A.L. (2009). Bioinformatics. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.DOI: 10.1002/9780470015902.a0001900.pub2.

Vassilev. D., Leunissen, J.A., Atanassov, A., Nenov, A. and Dimov, G. (2005). Application of bioinformatics in plant breeding. Wageningen University, Netherland.

Vassilev.D, Nenov, A. Atanassov, A. Dimov, G. and Getov, L. (2006).Application of bioinformatics infruit plant breeding. *Journal of fruit and ornamental plant research*14:145-162.

Vencato, M., Tian, F. Alfano, J. R. Buell, C. R. Cartinhour, S. DeClerck, G. A. Guttman, D. S. Stavrinides, J. Joardar, V. Lindeberg, M. Bronstein, P. A. Mansfield, J. W. Myers, C. R. Collmer, A. and Schneider, D. J. (2006). Bioinformatics-Enabled Identification of the HrpLRegulon and Type III Secretion System Effector Proteins of *Pseudomonas syringae*pv. *phaseolicola*1448A. MPMI 19(11): 1193–1206.

Wales, J. (2009). Bioinformatics http://en.wikipedia.org/wiki/Bioinformatics# Web_services_in_bioinformatics.

Willis, D.K., and Kinscherf, T.G. (2004). Global regulation in *Pseudomonas syringae*. Pages 223-238 in: Pseudomonas, Vol. 2. J. L. Ramos, ed. Kluwer Academic/Plenum Publishers, New York, U.S.A.

Xiong, J. (2009). Essential bioinformatics: introduction and biological databases. Cambridge University press, USA. http://www.cambidge.org/catalogue.