

BIOINFORMATICS A RELIABLE TOOL IN AGRICULTURE

Abstract

Bioinformatics is a multidisciplinary field of study that creates tools and software for analysing biological data. It aids in the evaluation and understanding of complex data in the numerous branches of biology and closely related sciences. It is possible to think of bioinformatics as a computer-based scientific discipline that combines computer science, biology, and mathematics to analyse and understand data from the fields of genomics and proteomics. The main components of bioinformatics are the collection and analysis of databases and the development of software tools and algorithms as a tool for the interpretation of biological data. It aids in the examination of vast amounts of disorganised data in agricultural research and the organisation of the data into the appropriate research route. As a result of the massive amounts of data created in the biological sciences, bioinformatics has evolved and grown. Understanding genetics and the molecular systems that drive various plant activities necessitates the use of omics, bioinformatics, and computational techniques. Using various bioinformatics methodologies and databases, results may be reviewed, saved, annotated, presented, and retrieved to aid in a better understanding of biological system investigation. Bioinformatics is important in the interpretation and analysis of data relevant to proteins, nucleotides, amino acid sequences, and a variety of other cellular biology investigations. Bioinformatics is the deployment and construction of software and solutions for the integration of many forms of plant phenomics data. It can examine the relationship between massive datasets using statistics and algorithms to figure out the structure, function, and protein, as well as locate the genes in a sequence. Agriculture bioinformatics, often known as agri-

Auothers

Goskula kiran

Ph.D Scholar
Department of Plant Pathology in Bidhan
Chandra Krishi Viswavidyalaya
West Bengal, India

Ankam shashank

Ph.D Scholar
Department of Plant Physiology in Bidhan
Chandra Krishi Viswavidyalaya
West Bengal, India

G.Razia sultana begum

Ph.D Scholar
Department of PlantPathology in Bidhan
Chandra Krishi Viswavidyalaya
West Bengal, india

Ponaganti Shiva kishore

Ph.D Scholar
Department of Genetics and Plant Breeding
Bidhan Chandra Krishi Viswavidyalaya
West Bengal, India

Amitava mondal

Ph.D Scholar
Department of PlantPathology in Bidhan
Chandra Krishi Viswavidyalaya
West Bengal, India

informatics, is a growing field of study. Novel genes were identified using computer software to improve seed quality, add micronutrients to plants for human health (nutritional genomics), and build plants to withstand or deal with metals (phytoremediation).

Keywords: Bioinformatics, Biological sciences, Genomics.

I. INTRODUCTION

Bioinformatics is a multidisciplinary field of study that creates tools and software for analysing biological data. It aids in the evaluation and understanding of complex data in the numerous branches of biology and closely related sciences. It is possible to think of bioinformatics as a computer-based scientific discipline that combines computer science, biology, and mathematics to analyse and understand data from the fields of genomics and proteomics. The main components of bioinformatics are the collection and analysis of databases and the development of software tools and algorithms as a tool for the interpretation of biological data. It aids in the examination of vast amounts of disorganised data in agricultural research and the organisation of the data into the appropriate research route. As a result of the massive amounts of data created in the biological sciences, bioinformatics has evolved and grown. Understanding genetics and the molecular systems that drive various plant activities necessitates the use of omics, bioinformatics, and computational techniques. Using various bioinformatics methodologies and databases, results may be reviewed, saved, annotated, presented, and retrieved to aid in a better understanding of biological system investigation. Bioinformatics is important in the interpretation and analysis of data relevant to proteins, nucleotides, amino acid sequences, and a variety of other cellular biology investigations. Bioinformatics is the deployment and construction of software and solutions for the integration of many forms of plant phenomics data. It can examine the relationship between massive datasets using statistics and algorithms to figure out the structure, function, and protein, as well as locate the genes in a sequence. Agriculture bioinformatics, often known as agri-informatics, is a growing field of study. Novel genes were identified using computer software to improve seed quality, add micronutrients to plants for human health (nutritional genomics), and build plants to withstand or deal with metals (phytoremediation).

II. TOOLS USED IN BIOINFORMATICS

Research in the biological sciences may be completed with the help of a variety of instruments and databases. The foundation of recent advances in genomic and proteomic research lies in these instruments and databases. Tools and databases can be divided into the following categories based on study of a specific field of life science:

Primary sequence analysis tools
Phylogenetic sequence analysis tools
Tools for protein structure-function analysis
Databases

- 1. Primary sequence analysis tools:** Understanding different biomolecule components is referred to as sequence analysis. As an example, proteins, DNA, and RNA all have specific functions. Muhammad, A.M., et al. 2014. A variety of tools have been created depending on the functionality to be used. Table I includes some of these.
- 2. Phylogenetic sequence analysis tools:** Phylogenesis is the field of study that deals with evolution. The techniques used in phylogenetic analysis are focused on recreating the evolutionary relationships between related animals or molecules; they also aim to forecast

some as-yet-unknown properties of molecules; monitor gene flow; and establish genetic relatedness. Khan, F. A., et al., 2014 Table I is a list of some of these.

3. **Tools for protein structure-function analysis:** Various techniques, which are listed in Table I, have been developed to forecast proteins based on their distinctive structure and function.
4. **Databases:** It has an enormous amount of data about biological elements including proteins, polymers, and nucleic acids. A specific key inside the databases identifies each component. For those who plan to pursue careers in biological research, including students, researchers, and scientists, it is highly helpful. Table II lists the equipment.

Table 1: Tools used in Bioinformatics

Classification of tools	Tool	Description	Reference
Primary sequence analysis tools	HMMER	With the use of this programme, databases of homologous protein sequences may be searched.	R. D. Finn <i>et al.</i> , 2011
	Clustal Omega	This application may be used to achieve multiple sequence alignment.	F. Sievers <i>et al.</i> , 2011
Phylogenetic sequence analysis tools	JStree	a readily accessible library for modifying phylogenetic trees.	A. Boc <i>et al.</i> , 2012
	Jalview	a fine-tuning the alignment-based alignment editor.	A.M. Waterhouse., 2009
Tools for protein structure-function analysis	CATH	a technology that organises proteins into categories semi-automatically.	I. Sillitoe <i>et al.</i> , 2013
	RaptorX	protein structure prediction using either a single template or many templates.	M. Källberg <i>et al.</i> , 2012
	HADDOCK	explains how biomolecular complexes like protein-protein and protein-DNA are modelled and interact.	S. J. Vries <i>et al.</i> , 2010
	SMART	A straightforward tool that retrieves modular architecture information provides many descriptions of proteinquery.	I. Letunic <i>et al.</i> , 2012
	STRING	a database used to forecast protein interactions.	A. Franceschini <i>et al.</i> , 2013
	MIMO	a dynamical graphmatching tool for effectively comparing biological processes.	P. Di Lena <i>et al.</i> , 2013

Table 2: Databases used in Bioinformatics

Databases	Description	Reference
CMAP	Complement Map Database to aid the complement community and researchers in relevant fields in finding new connections.	K. Yang <i>et al.</i> , 2013
Dicty Base	Dictyostelium discoideum database.	P. Gaudet <i>et al.</i> , 2011
Ensembl	Include eukaryotic genomes with annotations.	P. Flicek <i>et al.</i> , 2013
European	Nucleotide Archive Records and displays information on experimental procedures that centre on nucleotide sequencing.	R. Elanchezhian., 2012
GenBank	A repository for nucleotide sequences and a member of the international nucleotide sequence database (INSD).	Benson <i>et al.</i> , 2011
Medherb	Database of plants with therapeutic value.	M.I. Rajoka <i>et al.</i> , 2014
Pfam	Protein family groupings.	R. D. Finn <i>et al.</i> , 2011
Prosite	Gives details on protein families, conserved domains, and active protein locations.	C.J. Sigrist <i>et al.</i> , 2013
Rfam	Sequence alignments representing a variety of RNA families.	S. W. Burge <i>et al.</i> , 2013
SGMP	Proteins involved in signal transduction pathways are included in the Signaling Gateway Molecule Pages (SGMP) database, which offers structured data about them.	A.R. Dinasarapu <i>et al.</i> , 2011

III. AGRICULTURE APPLICATIONS OF BIOINFORMATICS

1. Crops: Using bioinformatics, it is possible to learn more about the genome sequences of different crops, such as Arabidopsis, rice, pea, papaya, etc. The following are some methods for identifying them:

- **Arabidopsis:** TAIR, UK Crop Net, Web Ace
- **Rice:** www.riceweb.org/research/Res_ntbio.html
- **Pea, papaya:** European sequencing trait (EST) and Quantitative Trait Loci (QTL) mapping.

- 2. Insect resistance:** Insect resistance has been developed in plants including corn, cotton, and potatoes. The insect-resistant genes from microorganisms like *Bacillus thuringiensis* (bacteria) have been placed in the plants to make them insect resistant, and this technique has been effective as a result. The term "Bt corn" refers to the insect-resistant corn. This approach has raised plant production and improved the nutritional content of plants while reducing the quantity of pesticides used on plants. Golden rice, which has had the rice genome stimulated with vitamin A, is one example of a crop with higher nutritional value.

IV. ROLES OF BIOINFORMATICS

Today, bioinformatics is used in all of biology's key fields. In genomics and bioinformatics have supported genome sequencing and have demonstrated their effectiveness in finding the genes, in phylogenetic comparison, and in the discovery of transcription factor binding sites of the genes (Liu et al., 1995; Thijs G. et al., 2002), to mention a few. Microarray technology has given scientists access to the world of transcripts (Spellman et al., 1998; Eisen et al., 1998). Microarray data analysis tools are provided by bioinformatics. These tools include image processing methods that read out the data, visualisation tools that give biologists a quick hint, preprocessing methods (Durbin et al., 2002) that eliminate systematic noise in the data, and clustering methods (Eisen et al., 1998; Sheng et al., 2003) that identify genes that behave similarly under various experimental conditions. Bioinformatics is useful in proteomics for the analysis of protein structures and the identification of the sequence locations involved in protein-protein interactions. In order to better comprehend biology at the system level, bioinformatics is beginning to show promise in resolving genetic networks (Segal et al., 2003). Finally, to model the dynamics of cellular interactions, bioinformatics is utilised in the research of the metabolome to examine cellular dynamics.

Agri-Informatics in Genome Sequencing: Lower prices, better-quality food, and higher-quality raw materials are the goals of developments in plant genomics. Food quality and quantity may be enhanced and improved via the application of biotechnology, bioinformatics, and genomics. The sequencing of the rice genome has influenced how bioinformatics and biotechnological technologies are applied (Edwards and Batley 2004). The genomic sequencing of plants and animals will be extremely beneficial to the agriculture industry. Finding genes in certain genomes may be useful for agricultural genetic research thanks to bioinformatics approaches and procedures. The unusual genetic information may be utilised in the future to create crops that are resistant to disease, pests, and drought, as well as to improve the quality of cattle and make them more productive, healthy, and disease-resistant. Crop plants are used extensively in many different sectors as well as play a significant part in human nutrition. It informs us of plant genetic research and its commercial significance. All of the major biological science fields were connected to bioinformatics. It facilitates the comparison of phylogenetic studies and the discovery of transcriptional factors that bind the locations of the genes (Liu et al., 1995; Thijs G. et al., 2002). Microarray data may be analysed using techniques provided by bioinformatics. These bioinformatics tools evaluate data, show it, offer tips for processing it, assist prevent data mistakes, and aid in comprehending how genes behave under various experimental circumstances (Durbin et al., 2002; Eisen et al., 1998; Sheng et al., 2003). The aim of bioinformatics in proteomics is to support the investigation of protein function and the identification of the sequence locations where protein-protein interactions take place. Bioinformatics assists in elucidating the genetic

networks in biology to research the systemic level (Segal *et al.*, 2003). Bioinformatics is ultimately employed to investigate the dynamics of a cell while studying metabolomes. The data sequence enables the explicit genetic study to be carried out with the help of database access. The co-linearity of the genetic and molecular genomes allows for the simple exchange of data, indicating significant conservation of genome structure between species. The aims of genome research include the finding of sequenced genes and the clarification of their activity through metabolic studies and reverse genetic examinations of gene knockouts. Over 20% of projected genes, which manifest as groups of related genes, are responsible for the development of a sizable portion of gene families. Multiple alignments may be used to determine how many genes are present within gene families, allowing previously undiscovered genes to be identified. This understanding opens up new approaches for analysing the patterns of gene expression in plants.

The information offered by modern technology, such as the database recorded DNA microarray expression data, will be useful for understanding functional genomics in plant biology. Expressed sequence tags (ESTs) also enable "digital northern" comparisons of gene expression levels, which can offer early hints about regulatory mechanisms that are yet unknown. The extensive effort being done on genomic mapping is tapped into via bioinformatics resources. This site facilitates the identification of ergonomically important genes, allowing the genetic engineering of agricultural plants chosen based on the quality of the end products. It does this by comparing crop plants and model species. Beyond expectations, nutritional genomics biotechnology techniques have been developed with the use of bioinformatics resources to genetically modify and increase the world's food supply for an expanding population. As a result, bioinformatics may now be employed to speed up the application of basic findings to agriculture. When it comes to agriculture, predictive plant growth modification will have an impact at a time when issues like food security, the disappearance of arable land, environmental protection, and The effects of climate change are well established.

V. APPLICATIONS OF BIOINFORMATICS IN PLANT BIOTECHNOLOGY

Bioinformatics applications give a variety of data kinds, such as protein domains and architectures, nucleotide and amino acid sequences, and expression patterns from numerous species. Bioinformatics is essential in a variety of biological domains. The use of bioinformatics, which gives extensive genetic data on a variety of plant species, has benefited the study of plants as a biological resource for humans. The purpose of this article is to introduce some of the major principles, techniques, and applications of bioinformatics as they relate to plant biotechnologies. Scientific discovery in the life sciences is developing significantly as a result of the use of bioinformatics and computational biology in the field of plant biology. Plant scientists have used sequencing technology to uncover the genetic architecture of many plant and microbial species, as well as their proteome, transcriptome, metabolome, and even metabolic pathways. A whole genome sequence, comprising DNA, RNA, and protein sequences, may be obtained from an organism's genome using the most basic technique currently being used in research: sequence analysis. Whole genome sequencing provides an initial idea of their use by estimating the number of species that are grouped into these groups. A full sequence of data made up of both coding and non-coding parts can serve as a necessary precursor for every functioning gene that defines the unique characteristics that organisms possess. Exons, introns, regulators, and promoters are all

present in the resulting sequence, which typically produces a significant amount of genomic data. As next-generation sequencing (NGS) and other omics technologies used to explore plant genomics progress, a greater number of sequenced plant genomes will become available to the public. Because of the advancement and use of bioinformatics, scientists can now assemble, store, and organise these massive amounts of data in a well-organized database.

Plant biotechnology may be studied utilising a range of databases and software programmes from the bioinformatics discipline. The study of plant genomes using next-generation sequencing (NGS) and bioinformatics has created a large amount of data over time. As of 2021, almost 21,000 plant genomes were accessible via the National Center for Biotechnology Information (NCBI) database, which may be viewed at <https://www.ncbi.nlm.nih.gov/>. The sciences of molecular biology, biochemistry, and genetics are the emphasis of NCBI's information collecting and analysis operations. In contrast to the NCBI database, which is not solely dedicated to plant genomes, EnsemblPlants was built specifically to access plant genomes. EnsemblPlant is part of the Ensembl project, which began in 1999 intending to automatically annotate the genome, merge the results with other publicly available biological data, and create an open-access archive or database online for biological data. EnsemblPlant contains data on polymorphic loci, population structure, genotype, linkage, and phenotype, in addition to the genome sequence, gene models, and functional annotation of the appropriate plant species.

- 1. Bioinformatics for plant breeding:** As was already said, a sizable amount of biological data is produced as a result of the advancement of next-generation sequencing (NGS) and other sequencing technologies, which must be kept in databases. Since whole genome sequences are easily accessible in databases, it is possible to freely link genomes based on gene sequence, probable function, or genetic map location. The bioinformatics software may be used to develop prediction hypotheses and incorporate the required phenotypes into plants from a complicated combination by concentrating on the genetic markers that perform well and offer a higher level of breeding dependability.
- 2. Biotic and abiotic stress management:** The investigation and analysis of the plant transcriptome in response to biotic and abiotic stress necessitates the use of bioinformatics tools. Furthermore, by using bioinformatics expertise on plant and crop genomes, the agricultural community may find the needed gene across genomes from different species and understand how it affects crops. Several genome databases offer gene expression profiling in addition to data storage to determine how a gene could be expressed in cells or tissues at the transcript level. The disease resistance geneenzyme and its related transcription factor can be discovered using in silico genomic technology. These genes and transcription factors aid the body's defence mechanism in the case of stress.
- 3. Bioinformatics for resistance in plant pathological studies:** Bioinformatics is becoming more crucial in plant pathology as novel methods for plant diagnostics are created. Because of developments in bioinformatics, the whole genomes of several species have been mapped in a little more than 10 years. These successes, together with ongoing efforts to elucidate the activities of genes and proteins, have improved our ability to understand the fundamental causes of plant diseases and identify innovative preventive strategies. Modern plant disease management makes extensive use of bioinformatics,

especially in the research of host-pathogen interactions, understanding of the genetics of the illness, and pathogenicity of a pathogen, all of which are inextricably linked to the creation of efficient management strategies. The relationships between plant pathogens and their hosts are dynamic. In the past, plant pathologists have studied these links using traditional approaches including physiology, histology, microbiology, plant breeding, and genetics. More recently, however, they have applied advanced biochemistry and molecular biology techniques. These interactions have been extensively studied by science for a very long time (Ojo and Maxwell, 2010). A wide range of proteins that allow a stealthy entry into the plant cell and make it much easier to get past the host defence mechanisms have emerged in plant infections (Vencato et al., 2006). In addition to various defences, plants have developed a variety of proteins that monitor their cells for symptoms of infection. When these monitors begin to slip, that's when you can tell whether there are any new infections (Rao et al., 2008). We now have a far better grasp of the molecular underpinnings of the host-pathogen relationship because of the advancement of genomic technology and knowledge (Koltai and Volpin, 2003). For example, it is possible to get the whole genomic sequence of the model plant *Arabidopsis* and *Pseudomonas syringae* pv. *Tomato DC3000*, which causes one of the diseased plants' bacterial infections. Additionally, the plant signalling system has other components that continue to function even after the detecting molecules have been identified. In addition, by using genetics and molecular biology methods, the pathogen proteins that would be employed to weaken host defences and accelerate the infection process have also been found (Anonymous, 2005). It is well acknowledged that bioinformatics is recognised for identifying the harmful components of a pathogen. Genomic technology will greatly influence the steps taken to reduce wanton plant diseases by extending the range and accessibility of gene pools that may be used for crop improvement. This plan will make use of technology that allows for precise modification and usage of resistance genes as well as an in-depth examination of the various resistance genes. Understanding the molecular basis of specificity and choosing targets for more robust resistance are both possibilities provided by pathogen genomic investigations (Michelmore, 2003). Additionally, this comprehension is necessary for the creation of fresh plant diagnostic methods. The qualities thought to be of main importance include those that impact yield, plant quality characteristics, and resistance to pathogens and abiotic stress (Vassilev *et al.*, 2006). Now, one might consider a genome programme as a vital tool for enhancing plants (Vassilev *et al.*, 2005).

One of the challenges facing contemporary agriculture in meeting the growing need for food as the world's population grows is crop loss brought on by the disease. The study of plant pathogens is essential for identifying pathogens, understanding disease aetiology, understanding disease resistance, and understanding how diseases affect the economy, among other things. The Plant Disease Resistance Gene Database (PRGdb), a highly developed bioinformatics database with hundreds of plant species, was created to assist plant genome research in the identification and prediction of plant disease resistance genes. This widely accessible platform not only preserves resistance genes but also provides a wide range of tools for investigation and the identification of additional R genes. Understanding plant infections and stress tolerance are crucial for enhanced crop breeding, and these topics may also be researched using bioinformatics. NGS and other sequencing technologies will increase the amount of plant genome data that is accessible in all public databases, allowing for the identification of genomic variants and the

prediction of the structure and function of proteins. Bioinformatics is used extensively in the process of breeding for disease resistance. Massive volumes of data will need to be collected and compiled, therefore it will be essential. It will also enable the presentation of data from many sources, which will facilitate the identification of remarkable individuals. Tissue culture and DNA-based markers are used by the majority of agricultural biotechnological methods to preserve germplasm, help in genetic improvement, and generate disease-free planting material. Recently, commercial transgenic crop production has begun in several Latin American countries, including Argentina, Brazil, Mexico, and others. Brazil is leading the way in the application of cutting-edge biotechnologies for the characterisation, mapping, and trait screening of important crops and diseases in certain Latin American countries. These technologies include genetic sequencing and microarray genomics. The function of bioinformatics in this situation is significant (Roca *et al.*, 2004). The vital demand for bioinformatics tools would be extremely helpful for the development of the agricultural industry for the benefit of mankind.

As a result of using molecular inheritable analysis to probe phytopathogenic fungi, a wide range of genes intertwined in fungal pathogenicity have been discovered and characterised (Idnurm and Howlett 2001 Knogge 1998; Sweigard *et al.* 1998). These include genes involved in the detoxification of antifungal composites produced by shops (Bowyer *et al.* 1995; Straney and VanEtten 1994), biosynthesis of phytotoxic composites (Panaccione *et al.* 1992), breakdown of the host factory cuticle (Tonukari *et al.* 2000; Walton 1994), conidia birth (Hamer and Givan 1990), appressorium conformation and function (Balhadère and Talbot 2001 Clergeot *et al.* 2001 Silué *et al.* 1998; Talbot *et al.* 1993), and amino acid metabolism (Balhadère *et al.* 1999), as well as those involved in conserved signalling pathways (Kronstad 1997; Xu and Hamer 1996). still, the capability to beget factory complaint is a complicated particularity and phytopathogenic fungi show significant variability in both their embryonic biology and the feathers of complaint symptoms they induce (Agrios 1988; Bowyer 1999). The disquisition of the inheritable regulation of pathogenic development in a small number of experimentally compliant pathogens, most specially *Ustilago maydis* and *Magnaporthe grisea*, has been made possible by molecular biology; still, it's clear that our understanding of the biology of fungal pathogens is still fairly limited.

A new and potent system for analysing colourful fungal infections is anticipated to be made available with the arrival of genome-wide analysis. For the first time, it'll be doable, for case, to characterise all the inheritable rudiments expressed during spore germination, infection-related development, and factory towel irruption, as well as those expressed under the direction of the major signal-transduction pathways necessary for pathogenesis (Idnurm and Howlett 2001 Talbot and Foster 2001). analogous to this, relating the gene databases and genomic structures of colourful fungus provides the chance to test propositions about the mechanisms that distinguish different phytopathogenic species and their evolutionary connections. nonetheless, several brand-new difficulties come on with these fantastic chances to consolidate our understanding of fungal infections. The first of them will presumably include the creation of specialised bioinformatic tools that are intended to store and study the enormous volume of genomic data that's most likely to be produced by the phytopathogenic fungus. Then, we give a brief overview of the pathogenic fungal genome-wide exploration that has been

conducted so far as well as the informatic coffers that have been created. We also go through the creation of a brand-new relational database that contains a sizeable chance of the EST data from phytopathogenic fungi that's presently accessible, as well as the procedures used to make it and the guiding generalities that guided its design.

4. The current status of fungal genomics: No harmful fungus has, to far, had its genome completely sequenced in a public-sector laboratory, hence the application of genomic methods to examine fungal conditions has lagged behind that of many other species. While two other fungal genome sequences, those of the filamentous fungus *Neurospora crassa* and the fission-inducing *Schizosaccharomyces pombe*, have recently been finished (<http://www-genome.wi.mit.edu/annotation/fungi/neurospora>), only one fungal genome sequence, that of *Saccharomyces cerevisiae*, is entirely accessible to the public. The Stanford genome database (<http://genome-www.stanford.edu/Saccharomyces/>) and the incentive protein database (<http://www.incyte.com/sequence/proteome/index.shtml>), which give a system of interrogating and reacquiring specific gene information from among them, 200 open reading frames that have so far been linked in the *cerevisiae* genome sequence has significantly increased the mileage of the incentive genome. These databases give data on mutant phenotypes and transcriptional profile analysis trials that track how the genome expresses itself in colourful surroundings. Incentive microarray analysis has been used, for illustration, to assay changes in situations of reiterations caused by heat shock, cold shock, the switch from galactose to glucose as the top carbon source (Lashkari *et al.* 1997), the switch from turmoil to respiration (DeRisi *et al.* 1997), diauxic growth shift from rich nutrient conditions to starvation stress (Wodicka *et al.* 1997), cell-cycle progression (Cho *et al.* 1998), sporulation (Chu *et al.* 1998), and gene expression in several nonsupervisory mutant strains (DeRisi *et al.* 1997). Analogous to this, periodical analysis of gene expression (savant) is a veritably dependable fashion that has been considerably used to look at the patterns of gene expression across the board in incentive (Yamamoto *et al.* 2001). (Kal *et al.* 1999; Velculescu *et al.* 1997). The link between protein accumulation and cornucopia has been investigated using savant frequency tables (Gygi *et al.* 1999). This is critical for further comparison of transcriptional profiling and proteomic data. At a specific stage of the development of incentive cells, savant frequency tables can provide an estimate of the significance of a given gene (Jansen and Gerstein 2000). Cluster analysis of genome-wide expression data from microarray and savant studies has made it possible to group together genes with comparable patterns of expression. For case, clustering of gene expression data has been used to categorise genes with analogous functions in incentives and humans (Eisen *et al.* 1998). To the topmost extent possible, the incentive community has served from the collection of all of this data into centralised data libraries.

The majority of yeast genes are now being deliberately disrupted, and the yeast proteome and metabolome are being studied in order to create and deploy the next generation of object-relational databases. The Genome Information Management System (GIMS), created by the University of Manchester (Manchester, U.K.) (<http://www.cs.man.ac.uk/norm/gims>), is an obvious example of a next-generation bioinformatics resource. It provides a setting where researchers may carry out more than 50 distinct analytic activities using the genetic data that is accessible. Long, text-based queries and other complicated querying tools are also permitted by the environment (Paton *et al.* 2000). Transcriptional profiling and proteomic analysis are significantly more

successful at exposing the coordinated action of large gene sets than a simple, manual study of the data. Better pattern-identification algorithms should make it easier to ask more complicated queries of stored genomic data.

Even though there is a stark difference between bioinformatic tools created for academics studying yeast and those now accessible to those studying phytopathogenic fungus, these databases show what may be done in a relatively short amount of time when access to genomic formation is available. As a result, the public sector's relative underfunding of fungal genomics has been a major topic of concern. Fungal genome information is critically needed (Pennisi 2001). In a recent assessment, for instance, just 16 (4%) of the 379 genome project websites looked at dealing with fungus, and of those, only seven dealt with harmful species (Yoder and Turgeon 2001). The sequences of several phytopathogenic fungi are expected to become available shortly, according to recent developments, including the Whitehead Institute's (Cambridge, MA, U.S.A.) announcement of the Fungal Genomes Initiative at the 21st Fungal Genetics Conference (Pacific Grove, CA, U.S.A.; March 2001). With the use of the same whole-genome shotgun method used to sequence *N. crassa*, up to 15 fungi will have their genomes sequenced as part of this new endeavour. The way phytopathogenic fungi are investigated will undoubtedly change as a result of the quantity of new genetic data, thus it is to be hoped that the programme will soon acquire the necessary funding. The genomes of the end-rot pathogen *Botrytis cinerea* and the powdery mildew fungus *Blumeria graminis* will both be sequenced as part of similar, albeit smaller scale, projects in Europe. These projects will also add to the wealth of new information that will likely be made public in the next three to five years. The recent discovery of syntenic regions between the genomes of the saprotroph *N. crassa* and the fungal pathogen *M. grisea*, in contrast to the limited similarity between some yeasts like *S. cerevisiae* and *Candida albicans*, shows the potential for surprises from genome sequence analysis (Seoighe *et al.* 2000). (Hamer *et al.* 2001).

5. Development of phytopathogenic fungal EST database: Most fungal research teams lack the resources necessary to comprehensively sequence a fungus genome. Although it is restricted by the developmental stage at which mRNA is extracted, single-pass, partial sequencing of either the 3' or 5' ends of complementary DNA (cDNA) clones to produce a set of ESTs offers a low-cost method to identify significant gene inventories in the absence of gene information. EST synthesis has been used to find genes in a wide range of species, although a limited number of phytopathogenic fungi have just lately been included (Keon *et al.* 2000; Rauyaree *et al.* 2001; Thomas *et al.* 2001). Five plant diseases are among the few fungal species having publically accessible EST data inventories, which are often only provided in flat-file format with minimal annotation (Skinner *et al.* 2001). Other EST sets that have been produced from various diseases are multiplying and are awaiting online publication. We developed and put into operation a database comprising EST information from three pathogenic fungi: *Magnaporthe grisea*, the cause of the rice-blast disease; *Blumeria graminis*, the cause of the barley powdery mildew; and *Mycosphaerella graminicola*, the cause of the septoria blotch of wheat. The database was designed to be user-friendly and to give the community of researchers studying phytopathogenic fungus information that they could use right away. By sequencing a library created by Dr. Sarah Gurr of the University of Oxford in Oxford, United Kingdom, using infected plant material and an EST dataset created by the Carlsberg

Institute in Copenhagen, Denmark, it was possible to get ESTs from *B. graminis* (Thomas *et al.* 2001). 2,701 unisequences were included in the sample, with a 499 base pair average for each sequence. The majority of the *M. grisea* ESTs were produced by Professor Daniel Ebbole at Texas A&M University and Dr. Ralph Dean at the North Carolina State Biotechnology Center (Research Triangle Park, NC, U.S.A.; formerly Clemson University, Clemson, SC, U.S.A.). They were downloaded from the EMBL database (<http://www.ebi.ac.uk/embl/index.html>). The collection included 1,839 unique sequences with an average length of 851 base pairs. Three cDNA libraries made at IACR-Long Ashton (Bristol, UK) by Dr. DDr. Hargreaves and Dr John Keon were sequenced to extract ESTs from *Myco sphaerella graminicola* (Keon *et al.* 2000). Two libraries were built using plant material that had been infected by fungi and another using mycelium that had been cultivated in liquid culture. The dataset included 2,926 uni sequences with an average length of 667 base pairs. EST sequences are frequently one-pass sequences (sequence readings from a single DNA strand only) and inevitably contain mistakes. Because there are so many sequences encoding frequently expressed mRNAs, EST data frequently show substantial inherent redundancy. Using these sequences to create a nonredundant set of unique sequences (uni sequences) through cluster assembling can improve the quality of EST data (by using Sequencher; Gene Codes Corporation, Ann Arbor, MI, U.S.A.). In addition to removing duplication from the dataset, this increases sequence accuracy and results in longer sequences. Consensus sequences that have been assembled and unassembled singletons can both be found in Unisequences that are produced from EST datasets.

Using the sequences of known genes as a comparison point, putative functions were ascribed to each unsequenced gene. The National Center for Biotechnology Information's database (<http://www.ncbi.nlm.nih.gov>) was queried using the blast method (Altschul *et al.* 1990; Bethesda, MD, United States). To compare the unsequenced sequences with those in the database of known protein sequences, the blast software converts the raw data into protein sequences in all six reading frames. For each unsequenced, the top five most similar sequences were obtained (with expectation values smaller than 1×10^{-5}). The expectation value may be thought of as a measure of the quality or strength of the match between the sequences; the lower the expectation value, the stronger the match. The cutoff number chosen here is more stringent than the suggested cutoff value of 10^{-2} (Anderson and Brass 1998), below which the matches were deemed significant in 98% of the cases. Each unsequenced was given a potential product or function based on these similarity scores. It is exceedingly speculative in certain circumstances to identify a putative function to an unsequenced based just on sequence similarity. We also found ESTs in each EST set that represented sequences from host plant contamination or from the vectors used to build the cDNA libraries. The Munich Information Center for Protein Sequences (MIPS; Neuberger, Germany) employed a hierarchical classification system to categorise the unit sequence by function based on the assignments (Mewes *et al.* 1997). Based on the specific metabolic route or cellular activity that each gene product is engaged in, this system categorises the gene products. The remaining sequences represented a broad spectrum of genes that are involved in several fundamental metabolic processes, as well as DNA, RNA, protein synthesis, protein sorting, and cellular transport. The genes for melanin production, fungal toxin manufacturing, detoxification of plant defence chemicals, breakdown of

plant cell walls, fungal growth, and signal transduction were also present, suggesting that they may play a role in fungal pathogenicities.

The three phytopathogenic fungi that are included in our database have a limited number of published gene sequences. *B. graminis*, *Mycosphaerella graminicola*, and *M. grisea* all had 196 entries in the NCBI protein database, respectively, after searches. Our database is intended to serve as a tool for gene discovery as its primary use. A researcher can utilise the EST sequence to create primers to make it possible to clone a gene they are interested in reasonably quickly by searching the database for that gene. The EST dataset includes a substantially bigger number of sequences than those currently accessible in the public databases if researchers are interested in finding homologues existing in the genome of species included in our database and have already successfully cloned a certain gene. EST datasets for phytopathogenic fungus will be contributed to this ongoing effort. Three additional phytopathogenic fungi, *Botrytis cinerea*, *Fusarium graminearum* (*Gibberella zeae*), and *F. sporotrichioides*, will shortly have their EST sequences uploaded. The database will eventually serve as a repository for all publicly accessible EST information from phytopathogenic fungus. Additionally, additional capabilities will be added to the web-based interface that will enable more complex text queries and allow users to search the database according to functions, enabling a deeper examination of the metabolic pathways found in each fungus. The cDNA clones from which the EST sequences were derived are being used by the research teams that provided the EST data from *B. graminis* and *Mycosphaerella graminicola* to create microarrays that will be used to track global changes in gene expression in these organisms under circumstances relevant to pathogenesis. The database will include the information from these experiments. The database was created as a part of the Consortium for the Genomics of Microbial Eukaryotes (COGEME) project, which may be found online at <http://www.cogeme.man.ac.uk/>. As soon as possible, information from our database will be added to the GIMS (Paton *et al.* 2000) for thorough comparison with the *S. cerevisiae* genome. Finding the elements required for pathogenesis is one aim of research into the genomes of pathogenic fungi. To find pathogen-specific orthologues, sequencing data from fungal pathogens may be compared to nonpathogenic fungi like *N. crassa* (genes that are present in pathogenic fungi but not in nonpathogenic fungi). However, it's probable that many of the pathogenicity components present in phytopathogenic fungi have analogues in nonpathogenic fungi, and that these analogues have undergone changes to their expression patterns or activities to become essential in pathogenesis. For instance, the mitogen-activated protein (MAP) kinase-encoding genes *PMK1* and *MPS1* in *M. grisea* have functional homologs in *S. cerevisiae* and are crucial for signal-transduction pathways linked to illness (Xu and Hamer 1996; Xu *et al.* 1998). The identification of genes with high expression levels during key pathogenic stages such as appressorium formation, conidiation, and penetration of plant tissues will also be possible with the use of microarray analysis. It is believed that the COGEME database will prove to be a useful tool for the scientific community studying phytopathogenic fungi in order to find new genes and ascertain their roles as there are currently no publicly accessible, finished genome sequences for these organisms.

Crop loss due to disease is one of the difficulties in modern agriculture to satisfy the increasing demand for nutrition along with the increase in the global population. Regarding pathogen identification, illness aetiology, disease resistance, and economic

effect, among other things, the study of plant pathogens is crucial. To aid plant genome research in the discovery and prediction of plant disease resistance genes, the Plant Disease Resistance Gene Database (PRGdb), a sophisticated bioinformatics resource containing hundreds of plant species, was established. In addition to maintaining resistance genes, this easily accessible platform makes a variety of tools available for research and the discovery of new R genes. In advanced crop breeding, it is necessary to comprehend plant pathogens and stress resistance, both of which can be studied using bioinformatics. More plant genome data will be available in all public databases thanks to NGS and other sequencing technologies, which will also make it possible to recognize genomic variations and predict the structure and function of proteins. Breeding for disease resistance involves the use of bioinformatics in numerous ways. It will be crucial for gathering and compiling massive amounts of data. To make the selection of exceptional people easier, it will also enable the presentation of data from disparate sources. The majority of agricultural biotechnological approaches use tissue culture and DNA-based markers to aid in genetic improvement, produce disease-free planting material, and save germplasm. Transgenic crops have lately been produced commercially in Latin American nations like Argentina, Brazil, Mexico, and many other countries. Brazil is at the forefront of the use of advanced biotechnologies, such as genetic sequencing and microarray genomics, for the characterization, mapping, and trait screening of significant crops and pathogens in some Latin American nations. In this context, bioinformatics plays a crucial role (Roca *et al.*, 2004). The crucial need for bioinformatics tools would be most benevolent for the advancement of the agriculture sector for the betterment of mankind.

REFERENCES

- [1] Agrios, G. N. 1988. Plant Pathology. Academic Press, San Diego, CA, U.S.A. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
- [2] A. Boc, A. B. Diallo and V. Makarenkov, "T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks", *Nucleic Acids Res.*, vol. 40, pp. 573-579, 2012.
- [3] A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn and M. Simonovic et al, "STRING v 9.1: protein-protein interaction networks, with increased coverage and integration", *Nucleic Acids Res.*, vol. 41, pp. 808-815, 2013.
- [4] Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.
- [5] A.M. Muhammad, S. Ujala and A. Niaz, "Use of Bioinformatics Tools in Different Spheres of Life Sciences", *J. Data Mining, Genomics and Proteomics*, No.5, pp. 2, 2014.
- [6] A.M. Waterhouse, J.B. Procter, D. M. Martin, M. Clamp and G. J. Barton, "Jalview Version 2--a multiple sequence alignment editor and analysis workbench", *Bioinfo.*, vol. 25, pp. 1189-1191, 2009.
- [7] Anderson, I., and Brass, A. 1998. Searching DNA databases for similarities to DNA sequences: When is a match significant? *Bioinformatics* 14:349-356.
- [8] A.R. Dinasarapu, B. Saunders, I. Ozerlat, K. Azam and S. Subramaniam, "Signaling gateway molecule pages--a data model perspective", *Bioinfo.*, vol. 27, pp. 1736-1738, 2011.
- [9] Balhadère, P. V., and Talbot, N. J. 2001. PDE1 encodes a P-type ATPase involved in appressorium-mediated plant infection by *Magnaporthe grisea*. *Plant Cell* 13:1987-2004.

- [10] Balhadère, P. V., Foster, A. J., and Talbot, N. J. 1999. Identification of pathogenicity mutants of the rice blast fungus *Magnaporthe grisea* by insertional mutagenesis. *Mol. Plant-Microbe Interact.* 12:129-142.
- [11] Bowyer, P. 1999. Plant disease caused by fungi: Phytopathology. Pages 294-321 in: *Molecular Fungal Biology*. R. P. Oliver and M. Schweizer, eds. Cambridge University Press, Cambridge
- [12] Blätke, M. A., Szymanski, J. J., Gladilin, E., Scholz, U., & Beier, S. (2021). Advances in Applied Bioinformatics in Crops. *Frontiers in Plant Science*, 12, 640394. <https://doi.org/10.3389/fpls.2021.640394>.
- [13] Bolser, D., Staines, D. M., Pritchard, E., & Kersey, P. (2016). Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomics data. In *Plant bioinformatics* (pp. 115-140). Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-3167-5_6.
- [14] Bowyer, P., Clarke, B. R., Lunness, P., Daniels, M. J., and Osbourn, A. E. 1995. Host range of a plant pathogenic fungus determined by a saponin detoxifying enzyme. *Science* 267:371-374.
- [15] Caligari, P. D. S., & Brown, J. (2017). Plant breeding, practice. <https://doi.org/10.1016/B978-0-12-394807-6.00195-7>
- [16] Cho, K. T., Portwood, J. L., Gardiner, J. M., Harper, L. C., Lawrence-Dill, C. J., Friedberg, I., & Andorf, C. M. (2019). MaizeDIG: maize database of images and genomes. *Frontiers in Plant Science*, 10, 1050. <https://doi.org/10.3389/fpls.2019.01050>.
- [17] Cho, R. J., Campbell, M. J., Winzeler, E. A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T. G., Gabrielian, A. E., Landsman, D., Lockhart, D. J., and Davis, R. W. 1998. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* 2:65-73.
- [18] Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P. O., and Herskowitz, I. 1998. The transcriptional program of sporulation in budding yeast. *Science* 282:699-705.
- [19] C.J. Sigrist, E. Castro, L. Cerutti, B. A. Cuche and N. Hulo et al, “New and continuing developments at PROSITE”, *Nucleic Acids Res.*, vol. 41, pp. 344-347. 2013.
- [20] Clergeot, P. H., Gourgues, M., Cots, J., Laurans, F., Latorse, M. P., Pepin, R., Tharreau, D., Notteghem, J. L., and Lebrun, M. H. 2001. PLS1, a gene encoding a tetraspanin-like protein, is required for penetration of rice leaf by the fungal pathogen *Magnaporthe grisea*. *Proc. Natl. Acad. Sci. U.S.A.* 98:6963-6968.
- [21] D.A. Benson, I.K. Mizrachi, K. Clark, D.J. Lipman, J. Ostell et al, “GenBank”, *Nucleic Acids Res.*, vol. 40, pp. 48-53, 2012.
- [22] DeRisi, J. L., Iyer, V. R., and Brown, P. O. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278:680-686.
- [23] Durbin, B.P., Hardin, J.S., Hawkins, D.M., Roche, D.M. 2002. A variance-stabilizing transformation for gene-expression microarray data, *Bioinformatics*, 18(Suppl. 1): s105.
- [24] Edwards, D., & Batley, J. (2004). Plant bioinformatics: from genome to phenome. *TRENDS in Biotechnology*, 22(5), 232-237.
- [25] Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, 95(25), 14863-14868.
- [26] F. A. Khan, C. D. Phillips and R. J. Baker, “Timeframes of speciation, reticulation, and hybridization in the bulldog bat explained through phylogenetic analyses of all genetic transmission elements”, *Systematic Biol.*, vol. 63, pp. 96-110, 2014.
- [27] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson and K. Karplus, et al, “Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega”, *Molecular and Systematic Biol*, vol. 7, pp. 539, 2011.
- [28] Gomez-Casati, D. F., Busi, M. V., Barchiesi, J., Peralta, D. A., Hedin, N., & Bhadauria, V. (2018). Applications of bioinformatics to plant biotechnology. *Current issues in molecular biology*, 27(1), 89-104. <https://doi.org/10.21775/cimb.027.089>.
- [29] Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S. G. 1996. Life with 6000 genes. *Science* 274:546, 563- 567.

- [30] Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. 1999. Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* 19:1720-1730.
- [31] Howe, K. L., Contreras-Moreira, B., De Silva, N., Maslen, G., Akanni, W., Allen, J., ... & Flicek, P. (2020). Ensembl Genomes 2020—enabling non-vertebrate genomic research. *Nucleic acids research*, 48(D1), D689-D695. <https://doi.org/10.1093/nar/gkz890>.
- [32] I. Letunic, T. Doerks and P. Bork, “SMART 7: recent updates to the protein domain annotation resource”, *Nucleic Acids Res.*, vol. 40, pp. 302-305, 2012.
- [33] I. Sillitoe, A. L. Cuff, B. H. Dessailly, N. L. Dawson and N. Furnham et al., “New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures”, *Nucleic Acids Res.*, vol. 41, pp. 490-498, 2013.
- [34] Jansen, R., and Gerstein, M. 2000. Analysis of the yeast transcriptome with structural and functional categories: Characterizing highly expressed protein. *Nucleic Acids Res.* 28:1481-1488.
- [35] Kal, A. J., van Zonneveld, A. J., Benes, V., van den Berg, M., Koerkamp, M., Albermann, K., Strack, N., Ruijter, J. M., Richter, A., Dujon, B., Ansong, W., and Tabak, H. F. 1999. Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two carbon sources. *Mol. Biol. Cell* 10: 1859-1872.
- [36] Keon, J., Bailey, A., and Hargreaves, J. 2000. A group of expressed cDNA sequences from the wheat fungal leaf blotch pathogen, *Mycosphaerella graminicola* (*Septoria tritici*). *Fungal Genet. Biol.* 29: 118-133.
- [37] Koltai, H. and Volpin, H. (2003). Agricultural genomics: an approach to plant protection. *European journal of plant pathology* 109: 101-108.
- [38] Knogge, W. 1998. Fungal pathogenicity. *Curr. Opin. Plant Biol.* 1:324- 328.
- [39] Kronstad, J. W. 1997. Virulence and cAMP in smuts, blast, and blight. *Trends Plant Sci.* 2:193-199.
- [40] Idnurm, A., and Howlett, B. J. 2001. Pathogenicity genes of phytopathogenic fungi. *Mol. Plant Pathol.* 2:241-255.
- [41] Knogge, W. 1998. Fungal pathogenicity. *Curr. Opin. Plant Biol.* 1:324- 328.
- [42] Kushwaha, U. K. S., Deo, I., Jaiswal, J. P., & Prasad, B. (2017). Role of bioinformatics in crop improvement. *Glob J Sci Front Res D Agric Vet*, 17, 13-24.
- [43] Kyriakidou, M., Tai, H. H., Anglin, N. L., Ellis, D., & Strömvik, M. V. (2018). Current strategies of polyploid plant genome sequence assembly. *Frontiers in plant science*, 9, 1660. <https://doi.org/10.3389/fpls.2018.01660>.
- [44] K. Yang, A. R. Dinasarapu, E. S. Reis, R. A. Deangelis and D. Ricklin et al, “CMAP: Complement Map Database”, *Bioinfo.*, vol. 29, pp. 1832- 1833, 2013.
- [45] Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., Brown, P. O., and Davis, R. W. 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl. Acad. Sci. U.S.A.* 94:13057-13062.
- [46] Liu, J.S., Neuwald, A.F., Lawrence, C.E. 1995. Bayesian models for multiple local sequence alignment and Gibbs sampling strategies, *J. Amer. Stat.*, 90: 1156.
- [47] Liu, J. S., Neuwald, A. F., & Lawrence, C. E. (1995). Bayesian models for multiple local sequence alignment and Gibbs sampling strategies. *Journal of the American statistical Association*, 90(432), 1156-1170.
- [48] Maloy, S., & Hughes, K. (Eds.). (2013). *Brenner's encyclopedia of genetics*. Academic Press. <https://doi.org/10.1016/B978-0-12-374984-0.00155-8>
- [49] Mathur, M. (2018). Bioinformatics challenges: a review. *Bioinformatics*, 3(6).
- [50] Mewes, H. W., Albermann, K., Heumann, K., Liebl, S., and Pfeiffer, F. 1997. MIPS: A database for protein sequences, homology data and yeast genome information. *Nucleic Acids Res.* 25:28-30.
- [51] Michelmore, R. W. (2003). The impact zone: genomics and breeding for durable disease resistance. *Current opinion in plant biology*, 6(4), 397-404.

- [52] M.I. Rajoka, S. Idrees, S. Khalid and B. Ehsan, "Medherb: An Interactive Bioinformatics Database and Analysis Resource for Medicinally Important Herbs", *Curr. Bioinfo.*, vol. 9, pp. 23-27, 2014.
- [53] M. Källberg, H. Wang, S. Wang, J. Peng and Z. Wang et al., "Templatebased protein structure modeling using the RaptorX web server", *Nat. Protocols*, vol. 7, pp. 1511- 1522, 2012.
- [54] Nishad, R., Ahmed, T., Rahman, V. J., & Kareem, A. (2020). Modulation of plant defense system in response to microbial interactions. *Frontiers in Microbiology*, *11*, 1298.<https://doi.org/10.3389/fmicb.2020.01298>.
- [55] Nobuta, K., & Meyers, B. C. (2005). Pseudomonas versus Arabidopsis: models for genomic research into plant disease resistance. *BioScience*, *55*(8), 679-686.
- [56] Normand, E. A., & Van den Veyver, I. B. (2019). Next-Generation Sequencing for Gene Panels and Clinical Exomes. *Human Reproductive and Prenatal Genetics*, 553-575.<https://doi.org/10.1016/B978-0-12-813570-9.00025-5>
- [57] Ojo, O. O., & Omabe, M. (2011). Incorporating bioinformatics into biological science education in Nigeria: prospects and challenges. *Infection, Genetics and Evolution*, *11*(4), 784-787.
- [58] Panaccione, D. G., Scott-Craig, J. S., Pocard, J. A., and Walton, J. D. 1992. A cyclic peptide synthetase gene required for pathogenicity of the fungus Cochliobolus carbonum on maize. *Proc. Natl. Acad. Sci. U.S.A.* 89:6590-6594.
- [59] P. Gaudet, P. Fey, S. Basu, Y. A. Bushmanova and R. Dodson et al., "dictyBase update 2011: web 2.0 functionality and the initial steps towards a genome portal for the Amoebozoa", *Nucleic Acids Res.*, vol. 39, pp. 620-624, 2011.
- [60] Platten, J. D., Cobb, J. N., & Zantua, R. E. (2019). Criteria for evaluating molecular markers: Comprehensive quality metrics to improve marker-assisted selection. *PLoS one*, *14*(1), e0210529.<https://doi.org/10.1371/journal.pone.0210529>.
- [61] Paton, N. W., Khan, S. A., Hayes, A., Moussouni, F., Brass, A., Eilbeck, K., Goble, C. A., Hubbard, S., and Oliver, S. G. 2000 Conceptual modelling of genomic information. *Bioinformatics* 16:548-558.
- [62] P. Di Lena, G. Wu, P. L. Martelli, R. Casadio and C. Nardini, "MIMO: an efficient tool for molecular interaction maps overlap", *BMC Bioinfo.*, vol. 14, pp. 159, 2013.
- [63] Pennisi, E. 2001. The push to pit genomics against fungal pathogens. *Science* 292:2273-2274.
- [64] Rao, V. S., Das, S. K., Rao, V. J., & Srinubabu, G. (2008). Recent developments in life sciences research: Role of bioinformatics. *African Journal of Biotechnology*, *7*(5).
- [65] Paton, N. W., Khan, S. A., Hayes, A., Moussouni, F., Brass, A., Eilbeck, K., Goble, C. A., Hubbard, S., and Oliver, S. G. 2000 Conceptual modelling of genomic information. *Bioinformatics* 16:548-558. Pennisi, E. 2001. The push to pit genomics against fungal pathogens. *Science* 292:2273-2274.
- [66] P. Flicek, M. R. Amode, D. Barrell, K. Beal, S. Brent et al, "Ensembl 2012", *Nucleic Acids Res.*, vol. 40, pp. 84-90, 2012.
- [67] Rauyaree, P., Choi, W., Fang, E., Blackmon, B., and Dean, R. A. 2001. Genes expressed during early stages of rice infection with the rice blast fungus Magnaporthe grisea. *Mol. Plant Pathol.* 2:347-354.
- [68] R. D. Finn, J. Clements and S. R. Eddy, "HMMER web server: interactive sequence similarity searching", *Nucleic Acids Res.*, vol. 39, pp. 29-37, 2011.
- [69] R. D. Finn, A. Bateman, J. Clements, P. Marco and C. Penny et al, "Pfam: the protein families database", *Nucleic Acids Res.*, volume 42, pp. 222-230, 2014.
- [70] R. Elanchezhian, "Application of Bioinformatics in Agriculture", 2012, link-<https://www.researchgate.net/publication/281291354> [last accessed 22, Oct, 2017].
- [71] Rhee, S. Y., Dickerson, J., & Xu, D. (2006). Bioinformatics and its applications in plant biology. *Annual review of plant biology*, *57*(1), 335-360.<https://doi.org/10.1146/annurev.arplant.56.032604.144103>
- [72] 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.

- [73] Sanseverino, W., Hermoso, A., D'Alessandro, R., Vlasova, A., Andolfo, G., Frusciante, L., ... & Ercolano, M. R. (2012). PRGdb 2.0: towards a community-based database model for the analysis of R-genes in plants. *Nucleic acids research*, 41(D1), D1167-D1171. <https://doi.org/10.1093/nar/gks1183>
- [74] Sanseverino, W., Roma, G., De Simone, M., Faino, L., Melito, S., Stupka, E., ... & Ercolano, M. R. (2010). PRGdb: a bioinformatics platform for plant resistance gene analysis. *Nucleic acids research*, 38(suppl_1), D814-D821. <https://doi.org/10.1093/nar/gkp978>.
- [75] Seoighe, C., Federspiel, N., Jones, T., Hansen, N., Bivolarovic, V., Surzycki, R., Tamse, R., Komp, C., Hulzar, L., Davis, R. W., Scherer, S., Tait, E., Shaw, D. J., Harris, D., Murphy, L., Oliver, K., Taylor, K., Rajandream, M. A., Barrell, B. G., and Wolfe, K. H. 2000. Prevalence of small inversions in yeast gene order evolution *Proc. Natl. Acad. Sci. U.S.A.* 97:14433-14437.
- [76] Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., Friedman, N. 2003. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Gen.*, 34(2): 166.
- [77] Silué, D., Tharreau, D., Talbot, N. J., Clergeot, P.-H., Notteghem, J. L., and Lebrun, M.-H. 1998. Identification and characterization of *apf1*– in a non-pathogenic mutant of the rice blast fungus *Magnaporthe grisea* which is unable to differentiate appressoria. *Physiol. Mol. Plant Pathol.* 53:239-251.
- [78] Sheng, Q., Moreau, Y., De Moor, B. 2003. Biclustering microarray data by Gibbs sampling, *Bioinformatics*, 19(Suppl. 2): ii 196.
- [79] Straney, D. C., and VanEtten, H. D. 1994. Characterization of the PDA1 promoter of *Nectria haematococca* and identification of a region that binds a pisatin-responsive DNA binding factor. *Mol. Plant-Microbe Interact.* 7:256-266.
- [80] Sweigard, J. A., Carroll, A. M., Farrall, L., Chumley, F. G., and Valent, B. 1998. *Magnaporthe grisea* pathogenicity genes obtained through insertional mutagenesis. *Mol. Plant-Microbe Interact.* 11:404-412
- [81] Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC et al (2022) Database resources of the national center for biotechnology information. *Nucleic Acids Res* 50(D1):D20–D26. <https://doi.org/10.1093/nar/gk>.
- [82] Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., & Friedman, N. (2003). Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature genetics*, 34(2), 166-176.
- [83] Sheng, Q., Moreau, Y., & De Moor, B. (2003). Biclustering microarray data by Gibbs sampling. *Bioinformatics*, 19(suppl_2), ii196-ii205.
- [84] S. J. Vries, M. Dijk and A. M. Bonvin, “The HADDOCK web server for datadriven biomolecular docking”, *Nat. Protocol.*, vol. 5, pp. 883-897, 2010.
- [85] Skinner, W., Keon, J., and Hargreaves, J. 2001. Gene information for fungal plant pathogens from expressed sequences. *Curr. Opin. Microbiol.* 4:381-386.
- [86] Straney, D. C., and VanEtten, H. D. 1994. Characterization of the PDA1 promoter of *Nectria haematococca* and identification of a region that binds a pisatin-responsive DNA binding factor. *Mol. Plant-Microbe Interact.* 7:256-266.
- [87] S. W. Burge, J. Daub, R. Eberhardt, J. Tate and L. Barquist et al, “Rfam 11.0: 10 years of RNA families”, *Nucleic Acids Res.*, vol. 41, pp. 226- 232, 2013.
- [88] Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., Futcher, B. 1998. Comprehensive identification of cell cycle-regulated genes of the yeast *saccharomyces cerevisiae* by microarray hybridization, *Molecular Biology of the Cell*, 9: 3 273.
- [89] Sweigard, J. A., Carroll, A. M., Farrall, L., Chumley, F. G., and Valent, B. 1998. *Magnaporthe grisea* pathogenicity genes obtained through insertional mutagenesis. *Mol. Plant-Microbe Interact.* 11:404-412
- [90] Tan, Y. C., Kumar, A. U., Wong, Y. P., & Ling, A. P. K. (2022). Bioinformatics approaches and applications in plant biotechnology. *Journal of Genetic Engineering and Biotechnology*, 20(1), 1-13.

- [91] Talbot, N. J., and Foster, A. J. 2001. Genetics and genomics of the rice blast fungus *Magnaporthe grisea*: Developing an experimental model for understanding fungal diseases of cereals. *Adv. Bot. Res.* 34:287-311.
- [92] Talbot, N. J., Ebbole, D. J., and Hamer, J. E. 1993. Identification and characterisation of MPG1 a gene involved in pathogenicity from the rice blast fungus *Magnaporthe grisea*. *Plant Cell* 5:1575-1590.
- [93] Thijs, G., Marchal, K., Lescot, M., Rombauts, S., De Moor, B., Rouze, P., Moreau, Y. 2002A Gibbs Sampling method to detect over-represented motifs in upstream regions of coexpressed genes, *Journal of Computational Biology*, 9(2): 447.