

Computational Biology as an aid in Deciphering vital insights on Plant Pathogens and their Interaction with hosts

Anurag Kashyap*, Borsha Rani Baruah and PD Nath

Department of Plant pathology, Assam Agricultural University, Jorhat, 785001, Assam, India

*For correspondence - anuragkashyap11@gmail.com

Running title – Deciphering Vital Insights on Plant Pathogens and their Interaction

Abstract

Computational biology or bioinformatics has its roots in the dawn of genomic revolution when research activities were started to sequence the genome of various organisms. Refinement in sequencing technologies and potential of genomic research resulted in meteoric growth of biological information such as sequences of DNA, RNA and protein requiring databases for efficient storage, management and retrieval of the biological information. Also, computational algorithms for analysis of these colossal data became a vital aspect of biological sciences. As in other biological sciences, computational biology and genome sequencing gave a new lease of life to plant pathology research, enabling an improved knowledge of the molecular basis of host-pathogen interactions. At present, these genome information of various plant pathogenic fungi, bacteria, virus, phytoplasma etc are stored and managed in various databases. Besides databases for information on nucleotide and proteins sequences of plant pathogens, databases are available for resistance genes for facilitating resistance breeding as well as bioprospecting and allele mining. Analysis of stored information, through bioinformatics tools, can reveal vital information of plant pathogens. Computational algorithms are now an intrinsic part of plant pathology providing valuable solutions in various aspects of plant disease research such as designing diagnostic molecular tools, ascertaining genetic variability among pathogens, taxonomic and phylogenetic studies of plant pathogens,

prediction of pathogenicity genes to name a few. Bioinformatics tools give valuable insights into the mechanism of pathogenesis and also help in identifying the underlying strategy involved in the initiation of plant immune response.

Keywords: Bioinformatics, Databases, Genome sequencing, Plant pathology research

Introduction

Food security is paramount issue to humanity as food is the most basic necessity for the survival of each individual. Rapid population growth, 6.07 billion in 2000 to projected 9 billion in 2050 (1), will further pose immense pressure on the food availability among the individuals. According to FAO report, due to deficit in food supply, about 12.5 per cent of the total world population suffers from chronic undernourishment (2), with situations worst in the developing countries.

Various factors stand as hurdles in achieving food security, includes depletion of biodiversity, global water crisis, climate change, land degradation, crop and livestock diseases etc. With no exigency plans in hand, diseases affecting food crops can have devastating effects on food availability. Estimates differ, but about 14.1 per cent of the crop loss occurs due to diseases (3) and jeopardize the current situation of food supply. The causal agents of diseases in crops accounts a vast number of plant pathogens ranging from few nucleotides in viroids to eukaryotic fungi. The effect of the phytopathogens on food crops range from development of mild insignificant

symptoms to catastrophes, in which vast areas of the crops are ravaged. Epidemics due to plant pathogens can be accounted to Irish famine in 1845, Great Bengal famine in 1942, Southern corn leaf blight epidemic in 1972 etc. In recent times, a lineage of wheat stem rust, Ug99 widely prevalent in the wheat fields across several countries in Africa and the Middle East and anticipated to spread further through these regions and possibly afield, has the potential of causing wheat production desolation with 100% crop loss (4).

Due to diversification in population of plant pathogens in time, latitude and genotype, it is very difficult to formulate control measures. Therefore, to strive the destruction in yield they create, it is necessary to delineate the real cause and explore for the remedies. The classical approaches applied in plant pathology like disease diagnosis by macroscopic symptom observation, microscopic observation of pathogen, transmission test to indicator plants, etc, have been invaluable over the years and will continue to remain so, but the recent advances in molecular and genomic research have revolutionized the approaches. The various processes underlying the interactions of pathogen and host at the molecular level are revealed through advent of molecular biology in plant pathology utilizing various techniques like molecular cloning, polymerase chain reaction, gel electrophoresis, macromolecule blotting and probing, etc. Genesis of genome sequencing further complimented molecular biology enabling a better knowledge at the gene level, of the strategies and underlying mechanisms of pathogenesis of plant pathogens.

The recent explosion in genome sequencing and high throughput sequencing technology has brought about unprecedented quantity of biological information to practically any biological field. With bacteriophage MS2 having RNA genome (5) being the first organism to mark the beginning of genome sequencing which further led to multinational genomic collaborative projects like Human Genome Project (6) and International Rice Genome Sequencing Project (7). In respect of plant pathogens, there are currently complete

genome sequences from various plant pathogenic fungi, bacteria and representatives of all plant virus genera are publicly available. This rapid rise in enormous amount of biological data amounting to billions of base pairs has led to an absolute requirement of computerized system for storage, organization and management and efficient retrieval of databases for further analysis. Thus this resulted in emergence of the discipline of bioinformatics or computational biology.

The computational analysis of biological data, *i.e.*, Bioinformatics which includes investigation of the information stored in the form of nucleotide and protein sequences in diverse biological databases has huge application in various fields of sciences including plant pathology. It has worked wonders in the study of plant pathogens, solving problems which otherwise would have been impossible without its use. It not only deals in creation of databases for efficient storage and retrieval but also in development of tools and algorithms for further analysis of stored data.

DNA sequence analysis using Bioinformatics

Ranging from prokaryotic bacteria to multi-cellular higher mammals large amount of sequence data are being published and analysis of this data involves various tasks which includes sequence assembly, annotation, motif finding, sequence alignment and phylogenetic trees (8). Sequencing through shotgun approach in which 1000 base pair long DNA fragments are sequenced with significant over coverage followed by assembling. Annotation follows assembling where genes are identified, marked out and their functions are turned out (9). Non-coding DNA contains regulatory regions where proteins called transcription factors (TFs) bind to turn on genes. TFs typically bind to small motifs, so the task is to find overrepresented short motifs in larger quantities of sequence. In the last two tasks, it is very useful to compare genomes of previously sequenced species. "Comparative genomics" is becoming a very important subfield (10). Detection and alignment of homologous sequence is an important task here. Given sequence data from

different species, it is useful to reconstruct their phylogenetic relationship. Algorithms exist for all these tasks, but all are evolving with increasing understanding of the function of non-coding DNA, increasing mathematical and algorithmic sophistication in the methods, and increasing raw computational power available to tackle these tasks.

Bioinformatics databases for plant pathology research : Databases, assemblage of structured, systematic and up-to-date information regarding the sequence of the three most important biomolecule *i.e.* DNA, RNA and protein are of the types, the nucleotide databases and the protein databases. The three primary nucleotide databases, which together comprise the International Nucleotide Sequence Database Collaboration are GenBank (National Centre for Biotechnology Information), EMBL (European Molecular Biology Laboratory) and DDJB (DNA Databank of Japan) (11). On the other hand, the leading protein sequence databases are Protein Data Bank (PDB), Universal Protein Resource (UniProt), Protein Information Resource (PIR) and SwissProt (12). These databases contain nucleotide sequence of thousands of organisms which include various plant pathogens *viz.*, fungi, virus, bacteria, phytoplasma etc, which are regularly used by researchers to reveal various facts regarding their pathogenicity and virulence. Besides these, there are databases solely devoted to plant pathology research.

Comprehensive Phytopathogen Genomics Resource (CPGR): This database was created to cater the requirement of 806 genome and transcriptome sequences of plant pathogens *viz.*, bacteria, fungi, oomycete, nematode, virus and viroid, to plant pathologist and other researchers in the field of plant pathology (13). Evolution of diagnostic molecular markers and hefty diagnostic protocols for plant pathogens through employment of genomics is the major objective of this database.

Plant Resistance Genes Database (PRGdb): It is a web accessible open-source database which

holds information about not less than 16 000 R-genes (resistance genes) in 192 plant species defined by 115 pathogens belonging to diverse groups (14). PRGdb is the first database of its kind dedicated solely for R-genes in plants, which are one of the most important genes in resistance breeding that can significantly reduce losses caused by pathogens.

PathoPlant: A new database, PathoPlant, combining information of specific plant-pathogen interactions on organism level and data about signal transduction on molecular level related with plant pathogenesis has been developed. It is a web-based service and the entries in the database are chained to associated in-house records as well as to entries in extrinsic databases such as SWISS-PROT, GenBank, PubMed, and TRANSFAC. At the present time, PathoPlant stores information regarding 104 plant-pathogen interactions in relation to 47 plant species and 29 pathogens, with the corresponding references and hyperlinks annotated (15).

Application of Bioinformatics in plant pathology research : With the help of genomic information of plant pathogens, stored in the databases, various tasks can be carried out through bioinformatics tools which are nothing but computer programs or algorithms. A vast array of computer programs are available, many of which are freely accessible in the web.

Comparative genomics to ascertain variability in the isolates of phytopathogens : In plant pathology, determination of genetic variability among isolates of plant pathogen is a routine work but the morphological variation gives only an idea of the genetic variability. Sequencing of plant pathogens followed by comparison with sequences of other isolates, utilizing bioinformatics tools, gives reliable and clear cut evidence of genetic variability. An algorithm, Basic Local Alignment Search Tool (BLAST) is one such program commonly used by Plant Pathologist for rapidly aligning and correlating a DNA or protein sequence in question with a database of sequences (16). The contrast of

nucleotide or protein sequences to the sequence databases and the significance of matches are evaluated statistically, which thereby determines the underlying functional and evolutionary relationships existing in the sequences and also helps to identification of the members of gene families (17). Genome sequencing and comparative genomics have been utilized in study of various plant pathogens like *Rhizoctonia solani* (18), *Puccinia graminis* f.sp. *tritici* (19), etc.

Prediction of genes in the genomic sequence of phytopathogens : Analysis of the genes constituting the biological sequence of the organism *i.e.* gene prediction can be executed through bioinformatics using algorithms like BLAST, GeneScan, GeneMark, Glimmer, GeneBuilder, Grail etc (20) based on previously available gene sequence database of various phytopathogens which have been experimentally verified. Many genes are discovered with additional help by comparison with known genes from other organisms, or searching for a match to a known protein sequence and the most popular tool to align genetic or protein sequences is BLAST. Zhou *et al.* (21) identified 306 new CDSs (Coding DNA sequence) in phytopathogenic bacteria *Xanthomonas campestris* involving a connected policy of bioinformatic, postgenomic and genetic approaches.

Bioinformatics approach in designing primers for PCR based diagnosis of plant pathogens : Nucleic acid based/ molecular method *i.e.* Polymerase Chain Reaction (PCR) method is one of the most reliable, accurate methods of plant pathogen detection and diagnosis utilizing primers, specifically fashioned for detection of plant pathogens. These specific primers for plant pathogens are designed based on genomic information of the pathogen in databases. Some of the online primer design sites are Gene Fisher, Do Primer, Primer 3, Web Primer etc (22). For accelerated and meticulous detection of three *Xanthomonas* species *viz.*, *X. axonopodis* pv. *poinsettiicola* (XAP), *X. hyacinthi* (XH) and *X. campestris* pv. *zantedeschiae* (XCZ), Back *et al.* (23) developed

a species-specific PCR assay and the species-specific primers were created from variable regions of the draft genome sequence data.

Phylogenetic analysis of phytopathogens through Bioinformatics approach : The evolutionary history based taxonomical classification of organisms, *i.e.*, Phylogenetics, an elemental part of the science of systematics that utilizes characteristics of organisms to create their phylogeny. Using bioinformatics approach, information on evolutionary relationships could be obtained through DNA sequence analysis, using those genes which are universally present and play essential role as these genes are likely to be highly conserved allowing little change over large span of evolutionary time, which are 18S rDNA in eukaryotes and 16 S rDNA in prokaryotes. Comparative analyses of 582 genomes across 75 genus groups, obtained from the NCBI genomes database in a high throughput approach was carried out utilizing the software EDGAR (Efficient Database framework for comparative Genome Analyses using BLAST score Ratios) and the results were integrated into the database. A quick survey of evolutionary relationships is executed through the software, simplifying the process of attaining new biological insights into the differential gene content of corresponding genomes (24).

Bioinformatics in fungal taxonomy in species identification : In fungal taxonomy, various techniques are being implemented for fungal identification and resolving the taxonomic positions. Of the various methods, Ribosomal DNA (rDNA) sequences is one of the most useful techniques aiding in fungal taxonomy. The three regions coding for the 5.8S, 18S, and 28 S rRNA form the ribosomal gene cluster. 5.8 rRNA gene, parts of the Large Subunit rRNA (LSU) 1 and 2 regions and Small Subunit rRNA (SSU) genes are useful for delineating of higher taxonomic ranks, such as classes and phyla and others for separation of genera and species (e.g. Internally Transcribed Spacer (ITS) 1 and 2 regions). In fungi, identification of species and molecular phylogeny, sequence analogy of the ITS region is extensively utilized (25) due to high copy number of rDNA

genes, which makes its amplification easier even from minute quantities of DNA and also, high degree of variation in present between closely related species. "FHiTINGS", an open source software program, for rapid identification, classification and analyzing of ITS DNA sequences produced in fungal ecology studies utilizing the output of a BLASTn (blastall) search through next-generation DNA sequencing (26). Public databases like GenBank, EMBL, DDBJ and Molecular Mycology Research Laboratory provide a huge amount of ITS sequences of fungi for species and race identification.

Bioinformatics in group and sub group identification of Phytoplasma : Phytoplasma, cell wall less bacteria, inhabiting in phloem of infected plants, are responsible for causing numerous plant diseases. Being obligate in nature and unculturable *in vitro*, the launch of web based research tool, *iPhyClassifier* provided greater scopes for rapid recognition and classification of varied population of phytoplasma into groups and sub groups on the basis of restriction fragment length polymorphism (RFLP) pattern similarity coefficient scores (21).

Bioinformatics in the study of plant pathogen interaction : In cellular functions like initiation of pathogenesis and maintaining infection, a major role is being played by protein-protein interactions (PPIs). An understanding of the molecular basis of pathogenesis is primarily based to a certain extent on the PPI network between a host and pathogen. Such informations can be best utilized for management of diseases through designing of pesticides which could target specific pathogen protein so as to avert disease inducing harmful interaction. For the purpose of experimental characterization of the molecular dynamics of plant-pathogen interactions, a model system was adopted where a major bacterial leaf pathogen, *Pseudomonas syringae* pv. *tomato* strain DC3000 (*PstDC3000*) was asserted to infect the plant host *Arabidopsis thaliana* (27). An internet-based database, PHI-base, classifies experimentally verified pathogenicity, virulence and effector genes from fungal, Oomycete and bacterial pathogens,

which infect animal, plant, fungal and insect hosts. It has the potential in the exploration of genes in agronomically important pathogens, which might be the targets for chemical intervention (28).

Protein-protein docking : In an endeavor for identification and cataloging of physical interactions between pairs or groups of proteins, a combination of bioinformatics and structural biology is utilized through prediction of interaction between protein-protein. An understanding of protein-protein interactions (PPIs) is of prime importance in the attempt for scrutiny of intracellular signaling pathways and modelling of protein complex structures. There is a close relation between the fields of protein-protein interaction prediction and protein-protein docking, which attempts to employ geometric and steric considerations to fit two proteins of recognized structure into a bound complex. Prognosis of the three dimensional structure of the macromolecular complex of interest which would occur in a living organism is the eventual objective of docking. The most widely used Protein-Ligand docking softwares are AutoDock, AADS, Surflex, VoteDock, SODOCK, SwissDock etc (29).

Docking as a tool against plant viruses : Docking, a method utilized for determination of the desired orientation of one molecule to a second when both are attached to each other forming a stable complex (30), which could be utilized for studying the mechanism of an enzymatic reaction, recognition of all possible attachment modes for ligands, screening a database and prognosis of ligation position of small molecule or anti viral agents to their object proteins (31). Docking acts a powerful resource for structure based drug discovery (32). A study on homology modeling and docking in *Mimosa yellow vein virus* resulted in screening of inhibitors against the virus protein and opened up scopes for designing antiviral agents against *Geminiviruses* in the future (33).

Secretome analysis in fungal pathogens : Secretome, the secreted proteins, is helpful in studying the underlying mechanisms involved in host pathogen interaction in case of fungal

Table 1. List of some important plant pathogens whose complete genome have been sequenced

Scientific Name	Disease caused	Genome Size	Reference
Plant pathogenic Fungi			
<i>Magnaporthe grisea</i>	Rice blast	40 Mb	Dean <i>et al.</i> (37)
<i>Ustilago maydis</i>	Smut	20.5 Mb	Kamper <i>et al.</i> (38)
<i>Fusarium oxysporum</i> f. sp. <i>lycopersici</i>	Wilt	60 Mb	Ma <i>et al.</i> (39)
<i>Botrytis cinerea</i>	Grey mold	38.8 Mb	Amselem <i>et al.</i> (40)
<i>Sclerotinia sclerotiorum</i>	White mold	39.6 Mb	Amselem <i>et al.</i> (40)
<i>Puccinia graminis</i> f. sp. <i>tritici</i>	Black rust	89 Mb	Duplessis <i>et al.</i> (41)
<i>Mycosphaerella graminicola</i>	Blotch	33.6 Mb	Goodwin <i>et al.</i> (42)
<i>Colletotrichum graminicola</i>	Anthraco nose	57.4 Mb	O'Connell <i>et al.</i> (43)
<i>Macrophomina phaseolina</i>	Charcoal rot	49 Mb	Islam <i>et al.</i> (44)
<i>Fusarium graminearum</i>	Head scab	36 Mb	King <i>et al.</i> (45)
Plant Pathogenic Bacteria			
<i>Xylella fastidiosa</i>	Pierce's disease, Citrus variegated chlorosis disease	2,679,306 bp	Simpson <i>et al.</i> (46)
<i>Agrobacterium tumefaciens</i>	Crown gall	2,841,581 bp	Goodner <i>et al.</i> (47)
<i>Ralstonia solanacearum</i>	Bacterial wilt	3,716,413 bp	Salanoubat <i>et al.</i> (48)
<i>Xanthomonas axonopodis</i> pv. <i>citri</i>	Citrus canker	5,175,554 bp	da Silva <i>et al.</i> (49)
<i>Xanthomonas campestris</i> pv. <i>campestris</i> Black rot of crucifers		5,148,708 bp	Qian <i>et al.</i> (50)
<i>Pseudomonas syringae</i>	Blight, wilt	6,093,698 bp	Feil <i>et al.</i> (51)
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i>	Bacterial blight of rice	4,941,439 bp	Lee <i>et al.</i> (52)

pathogens. Genome sequencing and various bioinformatic analysis are carried out to predict secretomes produced by the fungi in establishing infection in their host. But all-inclusive secretomic profiles of fungal pathogens are lacking as the process is in the budding stage. Predicted secretomes have been used in interspecies comparative analysis of important fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea* causing necrosis in plants (34). Analysis of the secretomes in plant pathogenic fungi *Fusarium graminearum* (35), *Phytophthora plurivora* (36) etc have been carried out for

understanding the association of the pathogen with their host plants.

Conclusion

Bioinformatics has a tremendous potential to assist Plant Pathologists in their efforts to reduce crop losses due to diseases. Plant Pathologist around the world has already acknowledged this fact and bioinformatics has evolved to be an elemental part of plant pathology research. Bioinformatics tools give valuable insights into the mechanism of pathogenesis and also help in figuring out an explanation regarding the methods and mechanisms associated in

Table 2. List of some important bioinformatics software used in the study of plant pathogens

Tool name	Description	Web Address	Reference
BLAST (Basic local alignment search tool)	Program utilized for identifying regions of local similarity between nucleotide or protein sequences and comparing them to sequence databases and thereby calculating the statistical significance of matches	https://blast.ncbi.nlm.nih.gov/	Altschul <i>et al.</i> (16)
GeneFisher	Web-based program for designing degenerate primers in an user interaction mode	bibiserv.techfak.uni-bielefeld.de/genefisher2/	Giegerich <i>et al.</i> (53)
GENSCAN	A software to acquire a DNA sequence and find the open reading frames that accord to genes	genes.mit.edu/GENSCAN.html	Burge and Karlin (54)
GeneMarkS	A program for prokaryotic gene annotation	http://exon.gatech.edu/Genemark/genemarks.cgi	Besemer <i>et al.</i> (55)
ProbCons	A protein multiple sequence alignment program	https://toolkit.tuebingen.mpg.de/probcons	Do <i>et al.</i> (56)
MEGA4 (Molecular Evolutionary Genetics Analysis)	A tool for operating automatic and manual sequence alignment in an integrated manner, inferring phylogenetic trees, mining web-based databases, evaluating rates of molecular evolution and testing evolutionary hypotheses	www.mega-software.net/	Kumar <i>et al.</i> (57)
AutoDock	Docking tool, working in an automated manner devised for prediction of how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure	autodock.scripps.edu/	Trott and Olson (58)
RDP3	Program suitable for characterization of recombination events in DNA-sequence alignments	web.cbio.uct.ac.za/~darren/rdp.html	Martin <i>et al.</i> (59)
SwissDock	Web based service for prediction of molecular interactions that might occur between a protein in question and a small molecule	www.swissdock.ch/	Grosdidier <i>et al.</i> (60)
Primer3	Tool for designing and analyzing primers for PCR and real time PCR experiments	simgene.com/Primer3	Untergasser <i>et al.</i> (61)
iPhyClassifier	Internet-based search tool for rapid recognition and classification of varied phytoplasmas	plantpathology.ba.ars.usda.gov/cgi-bin/resource/iphyclassifier.cgi	Zhao <i>et al.</i> (62)

triggering of plant immunity which can assist in improving plant disease resistance. Field of computational biology particularly homology modeling and docking have the potential to discover various unconventional chemicals against many plant pathogens including plant viruses where no effective direct methods of control is available at present in field conditions. The achievements made in Plant Pathology through bioinformatics or computational biology are plenty, but this is just a tip of the iceberg and many more breakthroughs are yet to be made in near future.

References

1. United Nations Population Division. (2010). World Population Prospects, the 2010 Revision. <http://esa.un.org/unpd/wpp/index.htm>. Retrieved on 2nd February, 2016.
2. FAO, WFP, IFAD. (2012). The State of Food Insecurity in the World 2012. Economic growth is necessary but not sufficient to accelerate reduction of hunger and malnutrition, Rome, FAO.
3. Agrios, G.N. (2006). Plant Pathology. Fifth Edition. Elsevier Academic Press. pp. 4.
4. McKie, R. and Rice, X. (2007). Millions face famine as crop disease rages. Science The Observer.
5. Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., et al. (1976). Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature, 260: 500 – 7.
6. Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., et al. (2001). Initial sequencing and analysis of the human genome. Nature, 409 : 860-921.
7. Tyagi, A.K., Khurana, J.P., Khurana, P., Raghuvanshi, S., Gaur, A., Kapur, A., et al. (2004). Structural and functional analysis of rice genome. J Genet, 83(1): 79-99.
8. Rhee, S.Y., Dickerson, J., Xu, D. (2006). Bioinformatics and its Applications in Plant Biology. Ann Rev Plant Biol, 57: 335–59.
9. Yandell, M. and Ence, D. (2012). A beginner's guide to eukaryotic genome annotation. Nature Rev Genet, 13: 329-42.
10. Touchman, J. (2010). Comparative Genomics. Nature Education Knowledge, 3(10): 13.
11. Batley, J. and Edwards, D. (2009). Genome sequence data: management, storage, and visualization. BioTechniques, 46 : 333-6.
12. Baxervanis, A. and Quellet, F. (2001). Bioinformatics: a practical guide to the analysis of genes and proteins. A John Wiley & Sons, INC., Publication. pp. 47.
13. Hamilton, J.P., Neeno-Eckwall, E.C., Adhikari, B.N., Perna, N.T., Tisserat, N., Leach, J.E., et al. (2011). The Comprehensive Phytopathogen Genomics Resource: a web-based resource for data-mining plant pathogen genomes. Database.
14. Sanseverino, W., Roma, G., Simone, M.D., Faino, L., Melito, S., Stupka, E., et al. (2010). PRGdb: a bioinformatics platform for plant resistance gene analysis. Nucleic Acids Res, 38: 814-21.
15. Bülow, L., Schindler, M., Choi, C., Hehl, R. (2004). PathoPlant®: A Database on Plant-Pathogen Interactions. In Silico Biology 4, 0044.
16. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic Local Alignment Search Tool. J Mol Biol, 215: 403–10.

17. Lobo, I. (2008). Basic Local Alignment Search Tool (BLAST). *Nature Education*, 1(1): 215.
18. Hane, J.K., Anderson, J.P., Williams, A.H., Sperschneider, J., Singh, K.B. (2014). Genome Sequencing and Comparative Genomics of the Broad Host-Range Pathogen *Rhizoctonia solani* AG8. *PLoS Genetics*, 10(5): e1004281.
19. Upadhyaya, N.M., Garnica, D.P., Karaoglu, H., Sperschneider, J., Nemri, A., Xu, B., et al. (2014). Comparative genomics of Australian isolates of the wheat stem rust pathogen *Puccinia graminis* f. sp. *tritici* reveals extensive polymorphism in candidate effector genes. *Front Plant Sci*, 5: 759.
20. Troy, C.S., MacHugh, D.E., Bailey, J.F., Magee, D.A., Loftus, R.T., Cunningham, P., et al. (2001). Genetic evidence for Near-Eastern origins of European cattle. *Nature*. 410: 1091.
21. Zhou, L., Vorhölter, F.J., He, Y.Q., Jiang, B.L., Tang, J.L., Xu, Y., et al. (2011). Gene discovery by genome-wide CDS re-prediction and microarray-based transcriptional analysis in phytopathogen *Xanthomonas campestris*. *BMC Genomics*. 12: 359.
22. Abd-Elsalam, K.A. (2003). Bioinformatic tools and guideline for PCR primer design. *Afr J Biotechnol*, 2(5): 91-95.
23. Back, C.G., Lee, S.Y., Lee, B.J., Yea, M.C., Kim, S.M., Kang, I.K., et al. (2015). Development of a Species-specific PCR Assay for Three *Xanthomonas* Species, Causing Bulb and Flower Diseases, Based on Their Genome Sequences. *Plant Pathol J*, 31(3): 212–18.
24. Blom, J., Albaum, S.P., Doppmeier, D., Pühler, A., Vorhölter, F.J., Zakrzewski, M., et al. (2009). EDGAR: A software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinformatics*, 10: 154.
25. Nilsson, R.H., Kristiansson, E., Ryberg, M., Hallenber, N., Karl-Henrik Larsson, K.H. (2008). Intraspecific *ITS* Variability in the Kingdom *Fungi* as Expressed in the International Sequence Databases and Its Implications for Molecular Species Identification. *Evol Bioinforma*, 4: 193–201.
26. Dannemiller, K.C., Reeves, D., Bibby, K., Yamamoto, N., Peccia, J. (2014). Fungal High-throughput Taxonomic Identification tool for use with Next-Generation Sequencing (FHiTINGS). *J Basic Microbiol*, 54: 315–21.
27. Sahu, S.S., Weirick, T. and Kaundal, R. 2014. Predicting genome-scale *Arabidopsis-Pseudomonas syringae* interactome using domain and interolog-based approaches. *BMC Bioinformatics*, 15 (Suppl 11) : S13.
28. Winnenburg, R., Urban, M., Beacham, A., Baldwin, T.K., Holland, S., Lindeberg, M., et al. (2008). PHI-base update: additions to the pathogen host interaction database. *Nucleic Acids Res*, 36: D572-76.
29. Sousa, S.F., Ribeiro, A.J., Coimbra, J.T.S., Neves, R.P.P., Martins, S.A., Moorthy, H.N.S., et al. (2013). Protein-ligand docking in the new millennium - A Retrospective of 10 years in the field. *Curr Med Chem*, 20(5): 2296–314.
30. Lengauer, T. and Rarey, M. (1996). Computational methods for biomolecular docking. *Curr Opin Struct Biol*, 6 (3): 402–6. doi :10.1016/S0959-440X(96)80061-3.

31. Gaur, R.K., Hohn, T., Sharma, P. (2014). Plant Virus-Host Interaction: Molecular Approaches and Viral Evolution. Elsevier Academic Press. pp. 153-5.
32. Meng, X.Y., Zhang, H.X., Mezei, M., Cui, M. (2011). Molecular Docking: A powerful approach for structure-based drug discovery. *Curr Comput Aided Drug Des*, 7(2): 146–57.
33. Prajapat, R., Marwal, A., Sahu, A., Gaur, R.K. (2011). Phylogenetics and in silico docking studies between coat protein of Mimosa yellow vein virus and whey alpha-lactalbumin. *Am J Biochem Mol Biol*, 1(3): 265-74.
34. Heard, S., Brown, N.A., Hammond-Kosack, K. (2015). An Interspecies Comparative Analysis of the Predicted Secretomes of the Necrotrophic Plant Pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS One*, 10(6): e0130534.
35. Brown, N.A., Antoniw, J., Hammond-Kosack, K. (2012). The Predicted Secretome of the Plant Pathogenic Fungus *Fusarium graminearum*: A Refined Comparative Analysis. *PLoS One*, <http://dx.doi.org/10.1371/journal.pone.0033731>.
36. Severino, V., Farina, A., Fleischmann, F., Dalio, R.J.D., Maro, A.D., Scognamiglio, M., et al. (2014). Molecular Profiling of the *Phytophthora plurivora* Secretome: A Step towards Understanding the Cross-Talk between Plant Pathogenic Oomycetes and Their Hosts. *PLoS One*, <http://dx.doi.org/10.1371/journal.pone.0112317>.
37. Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., et al. (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature*, 434 (7036): 980–6.
38. Kämper, J., Kahmann, R., Bölker, M., Ma, L.J., Brefort, T., Saville, B.J., et al. (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature*. 444 (7115): 97–101.
39. Ma, L.J., van der Does, H.C., Borkovich, K.A., Coleman, J.J., Daboussi, M.J., Di Pietro, A., et al. (2010). Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*, 464(7287): 367–73.
40. Amselem, J., Cuomo, C.A., van Kan, J.A., Viaud, M., Benito, E.P., Couloux, A., et al. (2011). Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genetics*, 7 (8): e1002230.
41. Duplessis, S., Cuomo, C.A., Lin, Y.C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., et al. (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc Natl Acad Sci U.S.A.*, 108 (22): 9166–71.
42. Goodwin, S.B., M'Barek, S.B., Dhillon, B., Wittenberg, A.H.J., Crane, C.F., Hane, J.K., et al. (2011). Finished Genome of the Fungal Wheat Pathogen *Mycosphaerella graminicola* Reveals Dispensome Structure, Chromosome Plasticity, and Stealth Pathogenesis. *PLoS One*, <http://dx.doi.org/10.1371/journal.pgen.1002070>.
43. O'Connell, R.J., Thon, M.R., Hacquard, S., Amyotte, S.G., Kleemann, J., Torres, M.F., et al. (2012). Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nature Genet*, 44 (9): 1060–5.
44. Islam, M.S., Haque, M.S., Islam, M.M., Emdad, E.M., Halim, A., Hossen, Q.M., et al. (2012). Tools to kill: genome of one of

- the most destructive plant pathogenic fungi *Macrophomina phaseolina*. BMC Genomics, 13 (1): 493. doi:10.1186/1471-2164-13-493.
45. King, R., Urban, M., Hammond-Kosack, M.C.U., Hassani-Pak, K., Hammond-Kosack, K.E. (2015). The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. BMC Genomics, 16: 544.
46. Simpson, A.J.C., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M., Alves, L. M. C. et al., (2000). The genome sequence of the plant pathogen *Xylella fastidiosa*. Nature. 406 (6792): 151–7.
47. Goodner, B., Hinkle, G., Gattung, S., Miller, N., Blanchard, M., Qurollo, B., et al. (2001). Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens* C58. Science. 294 (5550): 2323–8.
48. Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., et al. (2002). Genome sequence of the plant pathogen *Ralstonia solanacearum*. Nature, 415(6871): 497–502.
49. da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., et al. (2002). Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. Nature, 417 (6887): 459–63.
50. Qian, W., Jia, Y., Ren, S.X., He, Y.Q., Feng, J.X., Lu, L.F., et al. (2005). Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*. Genome Res, 15(6): 757–67.
51. Feil, H., Feil, W.S., Chain, F., Larimer, P., DiBartolo, G., Copeland, A., et al. (2005). Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. tomato DC3000. Proc Natl Acad Sci U.S.A. 102 (31): 11064–9.
52. Lee, B.M., Park, Y.J., Park, D.S., Kang, H.W., Kim, J.G., Song, E.S., et al. (2005). The genome sequence of *Xanthomonas oryzae* pathovar *oryzae* KACC10331, the bacterial blight pathogen of rice. Nucleic Acids Res, 33 (2): 577–86.
53. Giegerich, R., Meyer, F., Schleiermacher, C. (1996). GeneFisher Software support for the detection of postulated genes. In Proceedings of the Fourth Conference on Intelligent Systems for Molecular Biology. ISMB 96. AAAI Press, Menlo Park, CA. pp. 68–78.
54. Burge, C. and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. J Mol Biol, 268: 78-94.
55. Besemer, J., Lomsadze, A., Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res, 29: 2607-18.
56. Do, C.B., Mahabhashyam, M.S., Brudno, M., Batzoglou, S. (2005). ProbCons: Probabilistic consistency-based multiple sequence alignment. Genome Res, 15(2): 330-40.
57. Kumar, S., Dudley, J., Nei, M., Tamura, K. (2008). MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Briefings in Bioinformatics, 9: 299-06.

58. Trott, O. and Olson, A.J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem*, 31(2): 455–61.
59. Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefevre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*, 26(19): 2462–3.
60. Grosdidier, A., Zoete, V., Michielin, O. (2011). SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.* doi: 10.1093/nar/gkr366.
61. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Res.* doi: 10.1093/nar/gks596.
62. Zhao, Y., Wei, W., Lee, I.M., Shao, J., Suo, X., Davis, R.E. (2013). The iPhyClassifier, an interactive online tool for Phytoplasma classification and taxonomic assignment. *Methods Mol Biol*, 938: 329-38.