



## PRESIDENTIAL ADDRESS

## The science of “omics” for plant pathologists\*

M. ANANDARAJ

ICAR-Indian Institute of Spices Research, Kozhikode 673 012, Kerala, India



It was a moment of exhilaration and joy when I heard from the Secretary of the Indian Phytopathological Society that I have been elected unanimously as the president of the society. I was overwhelmed by the warmth and confidence of the members and would ever remain grateful for this kind gesture. The society has grown to this level by the vision, foresight and untiring efforts of pioneer leaders in plant pathology and I recollect their contributions with reverence. During my tenure as president the executive committee and general body were gracious enough to agree for opening a new zone in the Northeast for the benefit of the students from this region. I am sure the society will be growing and contributing not only for the growth of the science of plant pathology but to help and advice policy makers in matters relating to plant disease management and reduction of crop losses to enhance productivity of crop plants.

The growing population of India is poised to overtake China in about a decade and half and India will be the most populous country in the world. This is putting the onus on the Indian agricultural scientists to grow more food, feed, fibre and fodder under shrinking natural resources like land, water and non availability of labour. The global warming and climate change could affect the plant diseases and there is increased pressure on the plant pathologists to minimize crop losses.

### Major plant pathogens

The molecular plant pathology based on surveys has listed the following as top ten pathogens of virus, bacteria, fungi, oomycetes and nematodes internationally.

#### Virus

1) *Tobacco mosaic virus*, 2) *Tomato spotted wilt virus*, 3) *Tomato yellow leaf curl virus*, 4) *Cucumber mosaic virus*, 5) *Potato virus Y*, 6) *Cauliflower mosaic virus*, 7) *African cassava mosaic virus*, 8) *Plum pox virus*, 9) *Bromemosaic virus* and 10) *Potato virus X*, (Scholthof *et al.*, 2011).

#### Bacterial pathogens

1) Pathovars of *Pseudomonas syringae*, 2) *Ralstonia solanacearum*; 3) *Agrobacterium tumefaciens*; 4)

*Xanthomonas oryzae* pv. *oryzae*; 5) *Xanthomonas campestris* pathovars; 6) Pathovars of *Xanthomonas axonopodis*; 7) *Erwinia amylovora*; 8) *Xylella fastidiosa*; 9) *Dickeya (dadantii and solani)* and 10) *Pectobacterium carotovorum* and *P. atrosepticum* (John Mansfield *et al.*, 2012).

#### Fungal pathogens

1) *Magnaporthe oryzae*, 2) *Botrytis cinerea*, 3) *Puccinia* spp., 4) *Fusarium graminearum*, 5) *Fusarium oxysporum*, 6) *Blumeria graminis*, 7) *Mycosphaerella graminicola*, 8) *Colletotrichum* spp., 9) *Ustilago maydis* and 10) *Melampsora lini* (Dean *et al.*, 2012).

#### Oomycete pathogens

1) *Phytophthora infestans*, 2) *Hyaloperonospora arabidopsidis* and *Phytophthora ramorum*, 4) *Phytophthora sojae*, 5) *Phytophthora capsici*, 6) *Plasmopara viticola*, 7) *Phytophthora cinnamomi*, 8) *Phytophthora parasitica* and *Pythium ultimum* and 10) *Albugo candida* (Kamoun *et al.*, 2014).

#### Plant parasitic nematodes

1) *Meloidogyne* spp., 2) *Heterodera* and *Globodera* spp., 3) *Pratylenchus* spp. 4) *Radopholus similis*, 5) *Ditylenchus dipsaci*, 6) *Bursaphelenchus xylophilus*, 7) *Rotylenchulus reniformis*, 8) *Xiphinema index*, (9) *Nacobbus aberrans* and 10) *Aphelenchoides besseyi* (John *et al.*, 2013).

In India no such study or survey has been done. However the Indian Council of Agricultural Research New Delhi has been funding network projects on major diseases of National Importance that included, wilt diseases, virus diseases and on *Phytophthora*. During XI plan (2007/12) a major project involving *Phytophthora*, *Fusarium* and *Ralstonia* has been launched and running in 20 Research Institutes ([www.phytofura.net.in](http://www.phytofura.net.in)).

There are ever increasing populations of aggressive pests and pathogens evolving to add to woes of global warming. There are reports of new strains of viruses on crop plants, new diseases such as late blight of tomato with newer strains of *Phytophthora infestans* that has created epidemics, new races of *Fusarium* on banana etc. Indian Phytopathological society being one of the

\*Presidential address delivered on 16 March, 2015 during 67th meeting of IPS held at the ICAR-Indian Institute of Spices Research, Kozhikode (Calicut), Kerala  
Email: arajiisr@gmail.com

largest societies, this august body of plant pathologists have always taken up the challenges and won several laurels. I am confident that the younger generations of plant pathologists are competent enough to meet the challenges and make use of the technologies available today.

The annual review of plant pathology is completing 52 years since it was first published and has been a source of information for researchers. The focus has been changing over the years and the reviews bring about latest information. A look at the latest issues reveals the focus is on climate change at the unraveling of host pathogen interaction and molecular level. The advances in bioinformatics along with genome sequencing technology have added to the growth of the new frontiers of "omics" sciences. It is providing newer tools to plant pathologists to understand the host-pathogen interaction and deploy newer and cheaper methods of pathogen control.

### Genomics

The recent development in methods for analyzing the structure and function of genes, which may be collectively termed "genomics", represents a new paradigm with broad implications. Technological advances in genomics continue to transform modern biology. Instead of characterizing genes one or a few at a time, it is now possible to determine the complete nucleotide sequence of all of the genes in an organism and to measure the amount of mRNA corresponding to all of the genes. Although comprehensive information of this kind is currently available for only a few organisms, it seems likely that comparable levels of information will rapidly become available for most widely studied organisms, including several higher plants.

Germplasm resources of crop plants, comprising of cultivars and their crop wild relatives are important reservoirs of natural genetic variations, originated from a number of genetic events as a response to environmental stresses and selection. Exploitation of these genetic diversities is vital to overcome problems associated with narrow genetic base of modern cultivars. Many agriculturally important variations such as yield, quality, tolerance to abiotic stresses, and some forms of disease resistance are controlled by poly genes that are highly influenced by gene in to environment (G × E) interactions as demonstrated for the spice crop turmeric across the country (Anandaraj *et al.*, 2014). These complex traits are referred to as quantitative trait loci (QTLs) and it is challenging to identify QTLs based on only phenotypic evaluation. Identification of QTLs of agronomic importance and its utilization in crop improvement further requires mapping of these QTLs in a genome of crop species using molecular markers. Genome projects of representative food crops yielded not only the deeper knowledge on evolution and also the functional/candidate genes responsible for the quality and yield traits of these systems. The identification of QTLs led to the identification of marker assisted selection

(MAS) that are helping plant breeders towards development of elite genotypes for sustainable crop production (Goff *et al.*, 2002; Deshmukh *et al.*, 2010; Amarawathi *et al.*, 2008; Soltis and Soltis, 1999). The genomics aided understanding of disease resistance, the allelic variation of cultivated and wild genotypes with associated and the development of new approaches to this problem are expected to reduce the time for new resistant varieties to be developed.

The combination of an increased understanding of the pathogen's genome, as well as the responses that occur in both the pathogen and the host on infection, will open up new methods for controlling diseases in crops. The input from the characterization of the genomes of beneficial microorganisms that forms main microenvironment for the growth promotion, stress tolerance and yield increase will greatly facilitate the perfect marriage of these two different fields of science towards increasing yield and for reducing environmental hazards that may be associated with the current agronomic use of available fungicides and insecticides.

The 1001 Genomes Project was launched at the beginning of 2008 to discover the whole-genome sequence variation in 1001 strains (accessions) of the reference plant *Arabidopsis thaliana* (Almira *et al.*, 2003; Bevan and Walsh 2005). The complete genome sequences of over 80 accessions were released in early 2010 by the Max Planck Institute, and many more have been added since by the Salk Institute, the Gregor Mendel Institute and Monsanto. As of September 2014, over 1100 lines have been sequenced ([www.1001genomes.org](http://www.1001genomes.org)). The 150 tomato genome re sequencing project was recently initiated with an objective to reveal and explore extant genetic variation in tomato, and will provide a major boost to identification of valuable alleles. The project aims to sequence 83 genotypes, including 30 wild accessions, 43 land races and 10 old varieties.

The advent of next generation sequencing platforms are the technological basis for the genome projects with their high throughput application for whole genome sequencing of all organisms. The "generation" refers to the chemistry and technology used by the sequencing process. First generation generally refers to Sanger sequencing. "Next-generation", is generally used to refer to any of the high-throughput methods which were developed after Sanger. The commercially available NGS technologies such as Roche/454 (<http://www.454.com/>), Solexa/ Illumina (<http://www.illumina.com/>) and ABISOLiD ([http://www3.appliedbiosystems.com/AB\\_Home/applicationstechnologies/SOLiDSystemSequencing/index.htm](http://www3.appliedbiosystems.com/AB_Home/applicationstechnologies/SOLiDSystemSequencing/index.htm)) have already demonstrated the potential to circumvent the limiting factors of Sanger sequencing. De novo sequencing can also be undertaken using these sequencing technologies. Generation of a whole genome sequence assembly by alignment of small sequence fragments without the availability of a reference genome is tedious, if not impossible, at present.

Genetic variation can be assayed using a variety of molecular markers. Once molecular markers have been linked to a trait of interest, these markers can be used to select desired lines from a large-scale population through marker-assisted selection (MAS), which saves both costs and time. Furthermore, the availability of gene and transcript sequence data in the public domain has made it possible to develop molecular markers from genes, which have been designated genic molecular markers (GMMs) or functional markers. The development and application of such molecular markers is gaining momentum because their discovery is inexpensive and putative functions can often be deduced by homology searches. Because these markers represent functional units, they are useful for assaying functional diversity in natural populations or germplasm collections and are valuable anchor markers for comparative mapping, evolutionary studies and for MAS. The next generation sequencing tools emerged in 2005 as replacement to the low throughput and high cost of first-generation methods (David *et al.*, 2003). With the advancement in Next gen sequencing technologies and less cost investment re-sequencing of any organism became easy. Whole genome re-sequencing aims to sequence the individual whose reference genome is already known. As reference genome sequences become increasingly available for many species, cataloguing sequence variations and understanding their biological consequences have become major research goals. The DNA re-sequencing is of sequencing a DNA region for an individual given that a reference sequence for this region is already available for the specific organism. Targeted re-sequencing is a variation of re-sequencing where only a small subset of the genome is sequenced, such as the exome, a particular chromosome, a set of genes or a region of interest (van Dijk *et al.*, 2014; Glen, 2011; Chung, 2014).

Genomics-assisted breeding approaches have greatly advanced with the increasing availability of genome and transcriptome sequence data for several model plant and crop species. Complete and/or draft genome sequences have become available for several organisms. Previously, most genome and transcriptome sequencing projects used Sanger sequencing methodology. The length of Sanger sequence reads has now increased from 450 bases to more than 1 kb. Owing to growing interest in human genome resequencing, a new generation of sequencing technologies have emerged. These next-generation sequencing technologies are able to generate DNA sequence data inexpensively and at a rate that is several orders of magnitude faster than that of traditional technologies.

### Plant associated microbial genome projects

The two different regions of the plant the stems, leaves, flowers and the roots, have received very different attention in terms of scientific investigations. The above-ground plant portions have been more amenable to morphological/biochemical characterization. The below ground (root) exploration is comparatively less not only

they are non sterile environment with both pathogenic/beneficial microbial population but also difficult to separate from the soil medium. With the continuing development of genomic tools it is becoming easier to pinpoint the contribution and characteristics of root associated microbes also as of the other microbes associated in the aerial parts of the plants. An international public effort on plant associated microbial genome analysis with the prioritized list of plant associated microbes to be sequenced by the American Phytopathological Society (APS) for discussions (Workshop on Genomic Analysis of Plant-Associated Microorganisms) was held in Washington, D.C., on 9 to 11 April 2002 as five year project, which formed the foundation on plant associated microbial genome sequencing (<http://genomenewsnetwork.org>). The Indian Institute of Spices Research Kozhikode has undertaken sequencing of an isolate of *Phytophthora capsici* the foot rot pathogen of black pepper with the help of The Centre for Genomic Analysis (TCGA), New Delhi. This is the first draft genome sequence and annotation of a *P. capsici* (05-06) isolate, infecting black pepper in India. Sequencing was done by combination of two next-generation sequencing (NGS) technologies (Illumina and Roche/454) in 2010 for the isolate 05-06 followed by isolate (98-93) in 2012 to bring significant insights in to the genetic variations that exist in Indian subcontinent having diverse agro-ecological climates. Hybrid assembly has been done to analyze the reads from both the NGS technology (combining different NGS technology results more accurate assembly). Using genome-scale analysis, we estimated the multi-locus phylogenetic relationship and distances among the Clade 2 species with twelve marker sequences. Finally, from the annotated data, we have predicted 1339 effector protein domains that pathogen secrete for enabling parasitic infection and reproduction. The complete role of different effectors proteins in pathogenicity for *P. capsici* has yet to be elucidated. Comparative genomics between the isolate revealed the common gene counts as 2039, 6095 genes unique to 98-93 and 4034 genes unique to 05-06 isolate. The unique gene variation towards its pathogen city needs to be investigated using gene expression analysis and proteomics approaches.

### Implications

Several representative of microbial sequence information (<http://microbesonline.org>, <http://www.e-fungi.org.uk>, <http://pfgd.org>) aided basic understanding of genome structure, organization of these organisms along with the evolutionary path. Many genes with functional significance and other genomic elements involved in plant-microbe interaction (Winson *et al.*, 2011) enhanced the information to the protection of food resources and environment.

The JGI Genome Portal (<http://genome.jgi.doe.gov>) provides unified access to all JGI genomic databases and analytical tools. The JGI maintains extensive data management systems and specialized analytical capabilities to manage and interpret complex genomic data.



Arabidopsis.org complete genome sequence along with gene structure, gene product information, gene expression, DNA and seed stocks, genome maps, genetic and physical markers, publications, and information about the Arabidopsis research community. The genomics aided understanding of disease resistance, the allelic variation of cultivated and wild genotypes with associated (QTL) elite function for example, and the development of new approaches to this problem are expected to reduce the time for new resistant varieties to be developed.

The combination of an increased understanding of the pathogen's genome, as well as the responses that occur in both the pathogen and the host on infection, will open up new methods for controlling diseases in crops. The input from the characterization of the genomes of beneficial microorganisms that forms main microenvironment for the growth promotion, stress tolerance and yield increase will greatly facilitate the perfect marriage of these two different fields of science towards increasing yield and for reducing environmental hazards that may be associated with the current agronomic use of available fungicides and insecticides.

### Bioinformatics for plant pathologists

Bioinformatics involves the technology that uses computers for storage, retrieval, manipulation, and distribution of information related to biological macromolecules such as DNA, RNA, and proteins and often considered computational molecular biology. The emphasis here is on the use of computers because most of the tasks in genomic data analysis are highly repetitive or mathematically complex. The use of computers is absolutely indispensable in mining genomes for information gathering and knowledge building. Bioinformatics consists of two complementary fields: the computational tools include writing software for sequence and databases. The application of these tools and databases in generating biological knowledge for better understand living systems.

In host pathogen interaction plants exerts multilayer defense response that includes physical barriers induction of hypersensitive reaction, programmed cell death, induction of defense proteins and synthesis of antimicrobial proteins. Crop breeding for enhanced disease resistance is mainly depends on locating resilient genes along with the understanding the plant- pathogen interaction (Collinge *et al.*, 2010). The proteomic studies reveal the functional players for each cellular process and are more informative than the genomic studies which tells only what is possible theoretically in the system. Large scale proteomics approaches enable studies that are impossible to perform by other molecular techniques and allowing for global look on the effects of plant-pathogen interaction, disease development, defense and the protein network. Such studies would lead to identify the candidate proteins that would be useful in future breeding as well as future generation fungicides.

### Open-source bioinformatics software

Many free and open-source software tools have existed and continued to grow since the 1980s. The combination of a continued need for new algorithms for the analysis of emerging types of biological readouts, the potential for innovative *in silico* experiments, and freely available open code bases have helped to create opportunities for all research groups to contribute to both bioinformatics and the range of open-source software available, regardless of their funding arrangements. The open source tools often act as incubators of ideas, or community-supported plug-ins in commercial applications. They may also provide *de facto* standards and shared object models for assisting with the challenge of bioinformation integration. The range of open source software packages include titles such as Bioconductor, Bioperl, Biopython, Biojava, BioJS, BioRuby, Bioclipse, EMBOSS, NET Bio, Taverna workbench and UGENE. To maintain this tradition and create further opportunities the non-Profit open Bioinformatics foundation has supported the annual Bioinformatics Open Source Conference since 2000.

### Web services in bioinformatics

Basic bioinformatics services are classified by the EBI(<http://www.ebi.ac.uk/services>) into three categories: Sequence Search Services, Multiple Sequence Alignment and Biological Sequence Analysis (<http://www.ebi.ac.uk/services>; <http://molbiol-tools.ca><http://sites.univ-provence.fr/~wabim/english/logligne.html>; <http://www.expasy.org/links.html>; <http://www.expasy.org/tools/>; <http://www.ncbi.nlm.nih.gov/pubmed/>; [http://www.uniprot.org/help/update\\_notification](http://www.uniprot.org/help/update_notification); <http://ugene.unipro.ru/>; <http://sequenceeasy.aeyilmaz.com.html>; <http://emboss.bioinformatics.nl/>; <http://galaxyproject.org/>)

The availability of these service-oriented bioinformatics resources demonstrate the applicability of web-based bioinformatics solutions, and range from a collection of standalone tools with a common data format under a single, standalone or web-based interface, to integrative, distributed and extensible bioinformatics workflow management systems.

### Taxonomy of pathogens

For plant pathologists taxonomy of the organism is very relevant. From phenotypic characterization to gene based sequences and the present multigene approaches gives a clear demarcation of the differences among pathogen populations (Almeida *et al.*, 2010; Kroon *et al.*, 2004; Maiden *et al.*, 1998; Martin and Tooley, 2003) The fact that the number of *Phytophthora* species described after 2000 is far more in number than that has been described till then is because of the molecular tools available presently (O'Sullivan, 2013). At Indian Institute of Spices Research, Kozhikode, whole genome sequencing of two isolates of *Phytophthora* infecting black pepper revealed the existence of enormous diversity. The genome size of

98-93 is 46.1Mb while that of isolate 05-06 is 66.8Mb. There were only 2039 genes common among both isolates whereas 6095 unique genes noticed in 98-93 and 4039 were unique to 05-06. The genome of 05-06 also revealed the presence of several effectors genes some are common as reported in JGI and about 52 different effector type unique to this isolate. This clearly shows the diversity and non-conformity with the characters reported for *P. capsici* on other crops worldwide warranting re-description of the species complex found in black pepper.

Classical bacteriological techniques used for strain typing include Gram staining, sugar utilization analysis, and serological testing. The main problem with such tests is that the phenotypic expression of a trait may not necessarily reflect the underlying genetics involved. Multi locus sequence typing is first described by Maiden *et al.* (1998). It is a molecular typing of many bacterial and fungal pathogens which deals with sequencing of 400-600bp from at least 6 housekeeping genes. The unambiguous characterization of strains of a pathogen is crucial for addressing questions relating to its epidemiology, population and evolutionary biology. The house keeping genes that perform basic metabolic function, present in most members of the study organism and are under selection. Organism with identical alleles in the sequence locus is denoted as sequence type (ST). Sequence data on ST are ideal for strain characterization, can be readily compared between laboratories.

### Multi Locus Sequence Typing (MLST) advantages

Housekeeping genes are rich in variability than 16s rDNA and the long length of the analysed sequence allows better differentiation of the organism. ITS sequences in some circumstances fail to discriminate among closely related taxa. Similarly, phylogenetic analyses and were based on coding sequence (Martin and Tooley, 2003; Kroon *et al.*, 2004), with a relatively low mutation rate and therefore real limited target sites for diagnostic development. Introns and intergenic portions of the nuclear and mitochondrial genome used in MLST are more variable and therefore more promising targets for diagnostic development. Since MLST is based on nucleotide sequence and not DNA fragment pattern matching (like PFGE or PCR fingerprinting), the method is highly portable and STs can be compared using web-based databases that enable the global spread of bacterial pathogens to be tracked. Sequence types may also be generated without the need to culture bacteria since PCR is used to amplify the MLST genes from a sample, hence it can be a viable surrogate to metagenomics from various sample types. Sequence with at least 6 genes reduces obscure of the resulting phylogeny.

### MLST in genomic Era

High throughput whole genome sequencing of organism can locate numerous gene targets and comparison with the bioinformatic softwares for the characterization of

microorganisms. These new gene targets would assist the fine tuned identification/ diagnosis of organism combing with the available genes for MLST. A novel software- Automated Selection of Typing Target Subsets (AuSeTTS) allows intelligent selection of optimal targets for pathogen strain typing. The objective of this software is to maximize both discriminatory power, using Simpson's index of diversity (D), and concordance with existing typing methods, using the adjusted Wallace coefficient (AW). The program interrogates molecular typing results for panels of isolates, based on large target sets, and iteratively examines each target, one-by-one, to determine the most informative subset (O'Sullivan *et al.*, 2013). MLST efforts to better understand the epidemiology, ecology, clinical microbiology, and virulence of pathogenic organism. Advances in MLST research would bring clearer appreciation of the taxonomy and in population structure of the microorganisms. Recent whole genome sequencing and comparative genomics provides opportunity to develop more new genes for MLST analysis and development of diagnostic gene array gene chips for high throughput genotyping of any organism.

### Transcriptomics

The transcriptome encompasses the set of transcripts from a cell or a population of cells, which include protein-coding mRNAs and non-coding small RNAs (e.g. ribosomal, tRNA, miRNA). Traditionally, transcriptome profiling, or transcriptomics, has focused on quantifying gene expression. EST sequencing has traditionally been the core technology used for the discovery of reference transcripts. However, it has some inherent limitations, such as low throughput, high cost and a long experimental cycle. With the advent of Ultra high-throughput sequencing (UHTS) technologies, it is now possible to obtain highly resolved structural information of RNA populations on a high-throughput platform. This includes mapping transcript initiation and termination sites, splice junctions and post-transcriptional modifications. Such information will lead to a better understanding of the functional elements within the genome and the discovery of novel developmental or environmental regulatory networks. Whole-genome or whole transcriptome analyses have become a realistic option for genetic non-model organisms, even for individual laboratories, and will soon be standard practice in molecular studies.

The transcriptome has been sequenced from rhizome tissue samples after challenge inoculation with *R. solanacearum* using next generation technologies for short read (Illumina platform). It involves analysis of GC content, repeat content, putative functions, gene families, transcription factor encoding genes and its comparative analysis with and *C. amada*. The identification of many defense related genes differentially expressed provides many insights to resistance mechanism to *R. solanacearum* and for studying pathways involved in responses to pathogen. We also identified several candidate genes that may underline the difference in

resistance to *R. solanacearum* between ginger and mango ginger (Prasath *et al.*, 2014).

### Proteomics

Proteins are vital parts of any living organisms and the main components of the physiological metabolic pathways of cells. The term proteomics was coined in 1997 to make an analogy with the genomics, the study of genes. The word proteomics is the blend of protein and genome and was coined by Wilkins *et al.* (2007). The proteomics is the entire complement of proteins. The study of proteins particularly their structure and function is called as proteomics. The term proteomics can be defined as the systematic analysis of proteome, the protein complement of genome (Pandey and Mann, 2000). This technique allows the global analysis of gene products in various tissues and physiological states of cells. With the completion of genome sequencing projects and the development of analytical methods for protein characterization, proteome becomes a major field of functional genomics. The mechanism of gene action in plant pathogen interaction is not a strait forward transcription, translation and protein synthesis. There are ever complicated processes of alternative splicing and various regulations and post translational modifications. Proteomics is the study of protein properties (expression level, post-translational modification, interactions etc.) on a large scale to obtain a global, integrated view of disease processes, cellular processes and networks at the protein level. The emergence of the concept proteomics is recent with the focus to deal with the issues concerning the function and regulation of the sequenced genes from whole genome and cDNA sequencing projects (Wasinger *et al.*, 1995). A key requirement in understanding the functional elements of the genome is accurate annotation of protein-coding genes. Most gene structures in newly sequenced organisms are based on computational predictions, often unsupported by experimental evidence; when available, experimental validation is usually based on cDNA analysis (Jones *et al.*, 2013). The cDNAs derived from RNA genes and aberrant transcripts as well as protein-coding genes, ultimate confirmation of the latter is needed at the protein level. The use of proteomics data to experimentally validate gene annotations has recently become an increasingly valuable complement to cDNA efforts. Several inconsistencies in genome annotation were verified experimentally using proteomics. The work by Merrihew *et al.* (2008) confirms that mass spectrometry is a powerful experimental tool for annotating sequenced genomes. In any sequenced whole genome, at least 30% to 50% of predicted gene products would have no known homologues or show too little sequence homology (less than 30% identity) to known proteins to make a reliable functional annotation.

Recent advancement in proteomics (MS, protein purification, post translational modifications and 3D structure determination) offers promising windows to characterize unknown genes by applying these technologies using proteins. Structural proteomics have

also showed its application in correcting sequence based functional annotation.

The overall aim of the proteomic study is characterizing the complex network of cell regulation. This analysis is required to determine which proteins are conditionally expressed, how strongly and whether any post translational modifications are affected. The greatest challenge of proteomics is the reproducible fractionation of the complex protein mixture while retaining the quantitative and qualitative relationships. Since 2-D PAGE is capable of resolving over 1,800 proteins in a single gel (Choe and Lee, 2003) it is important as the primary tool of proteomics where multiple proteins must be separated for parallel analysis.

The general work flow in a 2-D gel follows first dimension separation, second dimension separation, staining, imaging and image analysis. The computer assisted image analysis software is used that allows spot detection, spot quantification, gel comparison and statistical analysis on the protein image. Image master TM, PD Quest TM and Z3 are the popular image analyzers.

At Indian Institute of Spices Research Kozhikode involving studies on black pepper-*Phytophthora capsici* interaction an efficient protein extraction protocol was standardized (Umadevi and Anandaraj, 2015). The study indicated the mechanisms of resistance in a moderately resistant variety (IISR-Shakthi) in comparison with a susceptible variety (Subhakara) leading to an insight on defense mechanism in these genotypes (Umadevi *et al.*, 2015).

### Mass spectroscopy in protein analysis

Currently, there are 2 complementary approaches in mass spectroscopy (MS) based proteomics bottom-up and top-down. The conventional peptide-based bottom-up shotgun proteomics involves in-gel or in-solution proteolytic digestion of proteins with enzymes, usually trypsin, into many pieces of small peptides (1-3 kDa) before MS analysis. This approach is well suited for protein identification, which only requires a small portion of sequence coverage (10-20 amino acid residues) to identify the protein from the database (Steen *et al.*, 2004). However, it is technically challenging to extract the intact proteins from gel matrices with a high recovery rate (Fridriksson *et al.*, 1999) Thus, gel-based separation is not applicable in top-down MS. Recently, solution-based isoelectric focusing, coupled with a multiplex tube gel electrophoresis separation device, referred to as gel-eluted liquid fraction entrapment electrophoresis, has been developed for intact protein separation based on their MWs and applied to proteins (10-250 kDa) with a high resolution and a high recovery rate (Zabrouskov *et al.*, 2006). Nevertheless, the surfactant SDS is still present in the sample so the proteins need to be precipitated in organic solvent and resolubilized in MS-compatible buffers. These analyses are known as peptide mass fingerprinting or multidimensional protein identification technology (Mud PIT).



## Host pathogen interactions

Plant disease resistance mechanisms were grouped into two categories *viz.*, innate and R-gene mediated resistance. The innate immunity is the first line of resistance mechanism and is mediated by the recognition of conserved pathogen- or microbe-associated molecular patterns (P/MAMPs) by plant pattern recognition receptors (PRRs) initiates the P/MAMP-triggered immune (PTI) response, which may occasionally result in Hypersensitive response (Mandadi Kranthi *et al.*, 2013) and Herbivore associated molecular patterns (HAMPs; Mithofer and Wilhelm Boland, 2008) released during attempted pathogen invasion. Innate immunity is sufficient to effectively combat majority of microbes, but few microbes with the ability to escape/suppress innate immunity (by evolving effectors) become the pathogens of the respective plant species. To combat pathogenic microbes, plant have further evolved the R-gene mediated defence mechanism (second line of defence), which operates through the specific recognition of pathogen-derived effectors or their reaction products. It was generalized that, the defense responses involving HR/ cell death were mediated by R-genes (mostly mediated through SA); and defense responses involving the accumulation PR proteins, antimicrobial compounds and cell wall reinforcements etc., were mediated by innate immunity (mostly mediated through ET/JA). It was also observed that R-gene mediated resistance is suitable only for biotrophic pathogens, while necrotrophs often take advantage of R-gene reactions for their pathogenicity. The quantitative defense exerted by innate immunity seems to be the only option plants have for combating necrotrophs (McDowell and Dangl, 2000). However, even after extensive number of R-genes cloned, the downstream process leading to the ultimate defense response is still elusive. Although innate immunity and R-gene mediated defense mechanisms operate by employing different kinds of receptors/defense molecules, recent progress in plant disease resistance research revealed the significant overlap in the downstream components of both the resistance mechanisms (Tsuda *et al.*, 2008). These findings, make the classification of resistance mechanisms into two distinct categories (*viz.*, innate and R-gene mediated) less dependable and inadequate. Further, display of heterogeneous lifestyles of the pathogens (as biotrophs, necrotrophs and hemibiotrophs) at different stages, adds to the weakness of the conventional classification. These inadequacies led to the proposal of the 'zig-zag' model, which includes components related to both innate and R-gene mediated responses and gave importance to the signal threshold generated through the combined action of both mechanisms, in the development of resistance response (Jones and Dangl, 2006). Hence defects in the components of either (innate or R-gene mediated) of the resistance mechanisms will contribute to the increased susceptibility, similarly presence of additional components in either of the mechanisms will lead to the increased resistance. This led to the idea that, resistance and susceptibility are not binary alternatives, but a continuum of possible interactions ranging from complete resistance to extreme susceptibility.

## Proteogenomics in black pepper

Black pepper is an important Spice crop in tropical countries. With the absence of whole genome sequence/ EST data on this crop, our lab undertook 2D separation of leaf proteins coupled with LTQ-Orbitrap mass spectrometry platform to bring out information on black pepper genes (Data Unpublished). This proteogenomics approach provided insight in to the entire protein spectra and identified many different important proteins with biochemical, physiological and defense significance. Hence, it is proving its applicability in functional genomics even in orphan crops in terms of genome sequencing attempts.

## CONCLUSIONS

Fungi and oomycetes are the crucial agents of many of the world's most serious plant diseases. They are able to get into the intact surfaces of host plants, rapidly establishes the infections leading to significant yield loss in large scale. Proteomics is a practical tool for dissecting molecular mechanisms underlying the plant pathogen interaction. Proteome changes in the host plant as it becomes infected/ resistant the pathogen can be traced back to the interaction between a single host and pathogen. The resulting biochemical change are of great interest because they may give insight into critical switch points in the defense related pathways that could be manipulated to engineer host plants with increased resistance and immunity to the pathogen. In horticulture crops, tomato, pepper and pear are extensively studied for the defense protein there by identification of defense genes. Apart from these studies stages of pathogen development and secretome analysis are also emerging areas of proteomes in this aspect. Some proteomic studies have been focused on *in vitro* fungal responses upon exogenous addition of substances that mimic host target compounds.

The knowledge on pathogenesis related proteins (PR) proteins not only would explore the mechanism of plant-pathogen interaction, but also would guide to develop the plant either through selection or to develop newer molecules to avoid the pathogen. The science of effectronomics, phosphoproteomics etc are paving way for better understanding of the complex processes involved. The science of proteomics can be better utilized for management of plant diseases. The scientists of Indian Institute of Spices Research are in the process of developing strategies to manage viral diseases in black pepper in the field by utilizing the knowledge of the interaction between the viruses and the host plant by using the proteomics information. During the virus infection several normal physiological processes are affected that include assimilation of several nutrients. By supplementing such nutrients and circumventing the pathogenic processes by minimizing the damage the diseased vines were nursed back to normal health (Data unpublished). This gives an alternate strategy to manage viral diseases in the field than the usual recommendation of reducing the inoculums by eradication of affected vines.

Similarly the transcriptome data generated during the pathogenic process between *Phytophthora capsici* and *Piper nigrum* and *P. colubrinum* is paving the way for better understanding of the infection process.

The responsibilities of plant pathologists increase with the new challenges facing crop production in the changing weather and emerging new diseases. The new tools available will certainly be useful to face all challenges swiftly and uphold the traditions of proving the critic wrong. The oldest science of “omics” relevant to plant pathologists is not the science as it is understood in genomic era but the art of managing diseases and the ‘economics’ of crop production.

## REFERENCES

- Almeida, N.F., Yan, S. and Cai, R. (2010). PAMPD, a multilocus sequence typing and analysis database and website for plant - associated microbes. *Phytopathol.* **100**: 208-205.
- Almira, E.C., Panayotova, N. and Farmerie, W.G. (2003). Capillary DNA Sequencing: Maximizing the Sequence Output. *J. Biomol. Tech.* **14**: 270-277.
- Amarawathi, Y., Singh, R. and Singh, A.K. (2008). Mapping of quantitative trait loci for basmati quality rice (*Oryza sativa* L.). *Mol. Breeding* **21**: 49-65.
- Anandaraj, M., Prasath, D., Kandiannan, K., John Zachariah, T., Srinivasan, V., Jha, A.K., Singh, B.K., Singh, S.K., Pandey, V.P., Singh, S.P., Shobha, N., Jana, J.C., Ravindrakumar, K. and Maheshwari, Uma (2014). Genotype by environment interaction effects on yield and curcumin in turmeric (*Curcuma longa* L.). *Ind. Crop Prod.* **53**: 358-364.
- Bevan, M. and Walsh, S. (2005). The *Arabidopsis* genome: A foundation for plant research. *Genome Res.* **15**: 1632-1642.
- Choe, L.H. and Lee, K.H. (2003). Quantitative and qualitative measure of intra laboratory two-dimensional protein gel reproducibility and the effects of sample preparation, sample load, and image analysis. *Electrophoresis* **24**: 3500-3507.
- Chung, W.C., Chen, C.C. and Ho, J.M. (2014). Cloud DOE: A User-Friendly Tool for Deploying Hadoop Clouds and Analyzing High-Throughput Sequencing Data with Map Reduce. *PLoS ONE* **9**(6): e98146.
- Collinge, D.B., Jorensen, H.J.L., Lund, O.S. and Lyngkjaer, M.F. (2010) Engineering pathogen resistance in crop plants – Current trends and future prospects. *Ann. Rev. Phytopathol.* **48**: 269-291.
- David, J. Galas, Stephen, J. and McCormack (2003). An Historical Perspective on Genomic Technologies. *Curr. Issues Mol. Biol.* **5**: 12-128.
- Dean, R., Van Kan, J.A., Pretorius, Z.A., Hammond-Kosack, K.E., Di Pietro, A., Spanu, P.D., Rudd, J.J., Dickman, M., Kahmann, R., Ellis, J. and Foster, G.D. (2012) The Top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* **13**(4): 414-30.
- Deshmukh, R., Singh, A. and Jain, N. (2010). Identification of candidate genes for grain number in rice (*Oryza sativa* L.). *Funct. Integr. Genomics* **10**: 339-347.
- Fridriksson, E.K., Baird, B. and McLafferty, F.W. (1999). Electrospray mass spectra from protein electro eluted from sodium dodecylsulfate polyacrylamide gel electrophoresis gels. *J. Am. Soc. Mass. Spectrom* **10**: 453-455.
- Glenn, T.C. (2011). Field guide to next-generation DNA sequencers. *Mol. Ecol. Resour.* **11**: 759-769.
- Goff, S.A., Ricke, D. and Lan, T.H. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**: 92-100.
- Hashemi, S.M. and Bardsiri, A.K. (2012). Cloud Computing Vs. Grid Computing. *ARPN J. Syst. Software* **2**:188-194.
- Johnston, J.M., Arcus, V.L. and Morton, C.J. (2003). Crystal structure of a putative methyltransferase from *Mycobacterium tuberculosis*: misannotation of a genome clarified by protein structural analysis. *J. Bacteriol.* **185**: 4057-4065.
- Jones, J.T, Haegeman, A., Danchin, E.G., Gaur, H.S., Helder, J., Jones, M.G., Kikuchi, T., Manzanilla-López, R., Palomares-Rius, J.E., Wesemael, W.M. and Perry, R.N. (2013). Top 10 plant-parasitic nematodes in molecular plant pathology. *Mol. Plant Pathol.* **14**(9): 946-961.
- Jones, J.D.G. and Dangle, J.L. (2006). The plant immune system. *Nature* **444**: 323-329.
- Kamoun, S., Furzer, O., Jones, J.D.G., Judelson, H.S., Ali, G.S., Dalio, R.J.D., Roy, S.G., Schena, L., Zambounis, A., Panabieres, F., Cahill, D., Ruocco, M., Figueiredo, A., Chen, X.R., Hulvey, J., Stam, R., Lamour, K., Gijzen, M., Tyler, B.M., Grunwald, N.J., Mukhtar, M.S., Tome, D.F.A., Tor, M., Van Den Ackerveken, G., McDowell, J., Daayf, F., Fry, W.E., Lindqvist-Kreuzer, H., Meijer, H.J.G., Petre, B., Ristaino, J., Yoshida, K., Birch, P.R.J. and Govers, F. (2014). The Top 10 oomycete pathogens in molecular plant pathology. *Mol. Plant Pathol.* **16**(4): 413-434.
- Kroon, L.P., Bakker, F.T., van den Bosch, G.B., Bonants, P.J. and Flier, W.G. (2004). Phylogenetic analysis of *Phytophthora* species based on mitochondrial and nuclear DNA sequences. *Fungal Genet. Biol.* **41**: 766-782.
- Maiden, M.C.J., Bygraves, J.A. and Fei, E. (1998). Multilocus sequence typing: a portable approach to the identification of clones within population of pathogenic microorganisms. *Proc. Nat. Acad. Sci. USA*, **95**: 3140-3145.
- Mandadi Kranthi, K. and Scholthof, Karen-Beth G. (2013). Plant Immune Responses Against Viruses: How Does a Virus Cause Disease? *Plant Cell* **25**: 1489-1505.
- Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., Dow, M., Verdier, V., Beer, S.V., Machado, M.A., Toth, I., Salmond, G. and Foster, G.D. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology *Mol. Plant Pathol.* **13**(6): 614-629.
- Martin, F.N. and Tooley, P.W., (2003). Phylogenetic relationships among *Phytophthora* species inferred from sequence analysis of the mitochondrially encoded cytochrome oxidase I and II genes. *Mycologia* **95**, 269–284.
- McDowell, J.M. and Dangle, J.L. (2000). Signal transduction in the plant immune response. *Trends Biochem. Sci.* **25**: 79-82.
- Merrihew, G.E., Davis, C. and Ewing, B. (2008). Use of shotgun proteomics for the identification, confirmation and correction of *C. elegans* gene annotations. *Genome Res.* **18**, 1660–1669.
- Mithofer, Axel and Boland, W. (2008). Recognition of Herbivory-Associated Molecular Patterns. *Plant Physiol.* **146**: 825-831.
- O’Sullivan, M.V.N., Vitali, S. and Gwendolyn, L.G. (2013). Software for selecting the most informative sets of genomic loci for multi-target microbial typing. *BMC Bioinformatics* **14**: 1-8
- Pandey, A. and Mann, M. (2000). Proteomics to study genes and genomes. *Nature* **405**: 837-846.



- Prasath, D., Karthika, R., Habeeba, N.T., Suraby, E.J., Rosanna, O.B., Shaji, A., Eapen, S.J. Deshpande, U. and Anandaraj, M.** (2014). Comparison of the transcriptome of ginger (*Zingiber officinale* Rosc.) and mango ginger (*Curcuma amada* Roxb. ) in response to bacterial wilt infection. *PLoS ONE* **9**(6): e 99731.
- Scholthof, K.B., Adkins, S., Czosnek, H., Palukaitis, P., Jacquot, E., Hohn, T., Hohn, B., Saunders, K., Candresse, T., Ahlquist, P., Hemenway, C. and Foster, G.D.** (2011). Top 10 plant viruses in molecular plant pathology. *Mol. Plant Pathol.* **12**(9): 938-954.
- Soltis, D.E. and Soltis, P.S.** (1999). Polyploidy: recurrent formation and genome evolution. *Trends Ecol. Evol.* **14**: 348-352.
- Steen, H. and Mann, M.** (2004). The abc's (and xyz's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* **5**: 699-711.
- Tsuda, K., Sato, M. and Glazebrook, J.** (2008). Interplay between MAMP-triggered and SA-mediated defense responses. *Plant J.* **53**: 763-775.
- Umadevi, P. and Anandaraj, M.** (2015). An efficient protein extraction method for proteomic analysis of black pepper (*Piper nigrum* L.) and generation of protein map using nano LC-LTQ Orbitrap mass spectrometry. *Plant Omics.* **8**(6): 500-507.
- Umadevi, P., Johnson, K. and Anandaraj, M.** (2015). Towards understanding the black pepper-*Phytophthora* pathosystem using integrated transcriptome and proteome datasets. 3<sup>rd</sup> International Symposium on Phytophthora 9-12 September, Bengaluru.
- van Dijk, E.L., Auger, H., Jaszczyszyn, Y. and Thermes, C.** (2014). Ten years of next-generation sequencing technology. *Trends Genet.* **30**: 418-426.
- Wasinger, V.C., Cordwell, S.J. and Cerpa-Poljak, A.** (1995). Progress with gene-product mapping of the mollicutes: mycoplasma genitalium. *Electrophoresis* **16**: 1090-1094.
- Wilkins, M.R., Appel, R.D., Williams, K.L. and Hotrasser, D.F.** (2007) (eds). Proteome Research Concepts, Technology and Application. <http://www.Springer.com/978-3-540-71240>.
- Winson, G.L., David, K.W.L. and Fleming, L.** (2001). Pseudomonas genome Database: Improved comparative analysis and population genomics capability for Pseudomonas genomes. *Nucleic. Acids Res.* **39**: 596-600.
- Zabrouskov, V., Han, X.M. and Welker, E.** (2006). Stepwise deamidation of ribonuclease at five sites determined by top down mass spectrometry. *Biochem.* **45**: 987-992.
- van Dijk, E.L., Auger, H., Jaszczyszyn, Yan and Thermes, C.** (2014). Ten years of next-generation sequencing technology. *Trends Genet.* **30**: 418-426.