CHAPTER-5

CONCLUSIONS

Plant diseases are caused by environmental stress, deficiencies and biological infectious agents including viroids, viruses, bacteria and fungi. Both agriculture scientists and microbiologists have been attracted to this field of research because of the need to identify the molecular agents and respective processes involved in causing infectious diseases in economically important plants.

To handle and control the biotic damage of plants by phytopathogens, it is very important to get knowledge about the responsible biomolecules, their roles and respective mechanism responsible for their pathogenicity in plants. Such approaches may be helpful in defining sophisticated biocontrol strategies in near future in the field of agriculture science to control the damage.

Detailed analyses of microbial pathogens have been documented by various scientists and research groups and it was found that their virulence factors display significant similarities at genomic and proteomic levels with each other. It was also observed that these phytopathogens are distinguished from their nonpathogenic neighbors by the presence of some specific biomolecules which can be considered as pathogenic markers i.e., virulence factors, genes and metabolic pathways. Mostly, the reason behind these molecular level differences in pathogenic and nonpathogenic bacteria is the gene acquisition process. It has been well reported that in some pathogens such virulence genes are acquired by horizontal gene transfer during evolution. Thus, distantly related pathogens have turned out to harbor closely related virulence genes and same metabolic pathways. Availability of DNA and protein sequence data for various organisms and knowledge of some known metabolic genes and their respective pathways provide the base to analyze the role and type of interaction between genes involved in various pathways and to develop a global gene regulatory network.
Since the last two decades, various attempts have been made by scientists to control the damage of plants and their products by phytopathogens. Out of all the attempts, the most potential and effective approach has been to develop drugs or therapeutic agents against the phytopathogen. Rational drug development process, which is faster than conventional approaches, requires the knowledge about the potential drug targets as an initial step for drug development process. Prediction of such targets through experimental analysis is very expensive and time consuming process, thus it requires the use of computational approaches for such purposes. All presently available computational approaches for drug target predictions are designed for human pathogens and none of them is suitable for phytopathogens, because of the unavailability of the data for phytopathogens and their hosts. Thus development of new computational approaches that suit drug target prediction in phytopathogens bypassing the needs of complex data is the need of the hour.

This report describes development of computational drug design strategies for phytopathogens *Pseudomonas* and *Xanthomonas* as models, with following objectives: to predict virulence associated genes in plant pathogens, and assigning their role in different metabolic pathways; Phylogenetic footprinting of virulence-related genes; Prediction of horizontal gene transfer of virulence genes; *In silico* drug (therapeutic) target prediction.

Data mining for all six phytopathogens was done to predict those genes which are well documented as virulence or pathogenic genes, belonging to ‘virulence/pathogenic gene families’ i.e., type III secretion system, effectors, adhesins and hrp. All mined virulence genes were considered as reference genes for further study to achieve the aims. Sequence of all these mined virulence genes, available at National Center for Biotechnology Information (NCBI), were downloaded in FASTA format for further use.

With the help of downloaded sequence, by using them as ‘seed information’, through Blastx, attempt was made to predict probable virulence genes from complete microbial world. A manual search of Blast results, using
stringent filtering, was carried out to select probable virulence genes. In total hundreds of genes, through each model pathogen, were found as probable virulence genes. The number of predicted probable virulence genes was directly dependent on the amount of seed information. Few of these predicted genes are already known to act as virulence, but most of them are, till now, not reported as virulence genes, thus considered as probable virulence genes. These probable virulence genes now can be used by scientific community to get more details.

Microarray gene expression analysis provides knowledge about the expression of genes in given condition, which further can be used to group the functionally associated genes. Still scientific communities have very less information about the expression patterns of virulence and virulence-associated genes, which can act as potential information to understand the growth and interaction of phytopathogens with their hosts. We downloaded cDNA-microarray data from GEO database, after normalization we clustered them. As a result we found genes with similar expression patterns at high association level with the known virulence genes and consider them as virulence associated genes; we also observed their expression level with the help of heat map.

Annotation can provide better understanding about genes; therefore attempt was made to get related information about predicted probable virulence genes. For annotation purpose we downloaded gene sequence of all probable virulence genes and treated them with Blast2GO. Results of Blast2GO provided biological process and functional annotations. Annotations for biological process suggested that in all the six phytopathogens, majority of probable genes were involved in cellular process, biological regulation and localization. Molecular functional annotation suggested that the major activities in which most of the probable virulence genes were involved are catalytic, binding and transport activity.

Regulatory elements play a major role in the expression and regulation of each and every gene. One such element is the transcription factor binding site (Regulatory motifs); in microbes such sites are located in upstream region
of the genes, where transcription factors are bound to regulate genes for their expression. Attempts were made to predict these regulatory sites in virulence genes, using phylogenetic footprinting approach. For few of virulence genes, mostly in *Pseudomonas* pathovars, upstream region, where TFBS are generally located, was not found with sufficient length, which suggests that these genes belong to some operon. For rest of genes extracted upstream sequence was treated with MEME and its results was manually observed for regulatory motifs. Comparisons of predicted TFBS with known microbial TFBS suggested that, the predicted TFBS found as sequence homologs to TFBS of *B. japonicum, B. subtilis, E. coli, H. pylori*, and *P. aeruginosa*. Most of these known TFBS are known to belong to sigma factor binding sites. Width of predicted TFBS was not found as similar and it varied from 8-50 nucleotides.

Virulence genes show some specific distribution throughout different groups of organisms. Using comparative sequence analysis, attempt was made to check distribution profile of virulence genes. A series of sequence comparison was done with the help of the Blast, using various available sequence databases at NCBI. Manual analysis was done to check the various distribution patterns of virulence genes and it was concluded that these virulence genes shared one of following five distribution patterns: family specific, class specific, common to microbes, horizontally transferred and universal. It was observed that more than 50 % of the virulence genes were considered as family specific, which suggests that virulence genes are generally not distributed widely, specifically in case of the *Pseudomonas* pathogens and they may be responsible behind the specific pathogenicity of *Pseudomonas* pathogens against their respective host. Observation for horizontal transferred genes suggest that, in case of *Pseudomonas* it is very less, out of all only two-three genes were found to be transferred horizontally but in case of *Xanthomonas*, proportion of horizontally transferred genes is relatively high, which was also supported by the genome dissimilarity value (δ-value). GC percentage of gene also supported the horizontal transfer of some of the proposed genes but not to all.
Phytopathogens cause a major damage to a wide range of plants, so it is very important to control them. The most important and successful method is to use chemical drug against them. To develop a drug through rational or information based drug development process, as a first step we required targets so that against them drug can be designed to control pathogens. An important aspect of the present work was to predict potential drug targets for all phytopathogens. To achieve this aim, first we designed genome based ‘F_{4/5}-Phase subtractive approach’, which was based on gene sequence pattern and relative synonymous codon usage biasing in genes (RSCU_{gene}). As first step, essential genes for all phytopathogens were predicted through Blastx facility of Database of Essential Genes, using member of γ-Proteobacteria as target data. Further we compared these essential genes with their respective host to subtract those essential genes which share homology with their host plants, to avoid threat to plants. It is very important that targets always remain available in the cells, so that the drug can come in contact and show its therapeutic effects, thus codon usage based RSCU_{gene} value was predicted for all passed gene. Higher the RSCU_{gene} value (>1) means higher the expression of genes. Around 15-20% of passed genes indicate RSCU_{gene} values > 1, thus their chance of expression are higher. Analysis of available cDNA microarray for *P. syringae* pv. phaseolicola also suggest similar level of gene expression for those genes which are predicted as high expressed genes by RSCU_{gene}. On the basis of the RSCU_{gene} values we again subtract those genes whose values were observed below 1 by considering them unsuitable as drug target and select genes with RSCU_{gene} value >1 and considered them as probable drug targets. In number they are 22 for *P. syringae* pv. phaseolicola, 24 for *P. syringae* pv. syringae, 21 for *P. syringae* pv. tomato, 08 for *X. compestris* pv. vesicatoria, 06 for *X. axonopodis* pv. citri and 07 for *X. oryzae* pv. oryzae.

Effective number of codon (Nc) was also calculated for all passed genes and their relation with RSCU_{gene} was observed and it was found that they are inversely correlated with each other with ‘Pearson correlation’ value (r) ~ -0.8, thus lower the Nc value means higher the gene expression.
To get the importance of probable drug targets gene’s products in pathogens, attempt was made to predict their probable biological pathways, through Blast2GO. Blast2GO analysis suggest pathways for only very few target genes, the reason behind this low number, might be the unavailability of sufficient pathway related information in KEGG database. Predicted pathways suggested that, predicted probable virulence genes are involved in biologically important pathways, thus, very important for survival of pathogenic bacteria.

It has been well documented that *Phaseolus vulgaris* and other beans, which act as host for *P. syringae* phaseolicola and *P. syringae* pv. syringae, show symbiotic relations with the members of nitrogen fixating bacteria i.e *Rhizobium etli* and other members of *Rhizobium*. To avoid the accidental killing of such beneficial bacteria, we checked the homology between the probable drug targets of *P. syringae* pv. phaseolicola and *P. syringae* pv. syringae and proteins of *Rhizobium etli* and *Rhizobium leguminosarum* bv. phaseoli. We observed that only two gene products of *P. syringae* pv. phaseolicola and three gene products of *P. syringae* pv. syringae are found non-homologs to Rhizobia. Thus, to control *P. syringae* pv. phaseolicola and *P. syringae* pv. syringae in field condition only these non-homolog targets can be considered as potential drug targets against drugs which will be designed to apply on seed as well as field condition and rest of the drug targets only considerable when drug will be designed to apply on off field conditions or only on seeds.

Till date none of the report is documented regarding commonism with the host of *Xanthomonas* and other bacteria; thus all the above reported probable drug targets can be used as potential drug targets against all types of drugs.