IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Volume-4, Special Issue-3, May 2016 E-ISSN: 2347-2693

Possible Drug Targets in Human Pathogens

Vishwanatha[1] K S Jagadeesh[2] Muralidhara B K[3] Siddaramappa[4]

[1]Asst.Professor, Department of CSE, Sri Krishna Institute of Technology, Bengaluru.
[2]Head of the Department-CSE, Sri Krishna Institute of Technology, Bengaluru.
[3]Asst.Professor, Department of CSE, Moodalkatte Institute of Technology, Kundapura
[4]Asst.Professor, Department of ISE, City Engineering College, Bengaluru.

Available online at: www.ijcseonline.org

Abstract— When antibiotics were first introduced in the 1940's, they were hailed as miracle drugs, and quickly provided effective therapy for many of the more dangerous pathogens then prevalent. However, resistance to these antimicrobials developed quickly. The World Health Organisation report into antimicrobial resistance published online, notes that formerly curable bacterial diseases are on the increase. For example, 98% of all South-East Asian gonorrhoea cases are presently multi-drug resistant, while up to 60% of nosocomial infections in the developed world are caused by drug-resistant and often opportunistic pathogens. Infections with rare virulent micro-organisms like Acinetobacter are also on the increase and opportunistic bacterial infections such as Pseudomonas aeruginosa and Salmonella spp. are becoming more common. Several factors contributing to this phenomenon during five decades of antibiotic mishandling have included: health workers misdiagnosing illness or providing the wrong prescription, patients failing to adhere to treatment, and the misuse of antimicrobials in animals with secondary effects observed in humans. To help counteract these problems, advances in technology can be used to hasten the hunt for new drug and vaccine targets. Bioinformatics itself can be defined as utilising large databases of biological information with specific in silico tools to complement traditional wet laboratory-based biology.

Keywords : Genomes, Putative Target Database, endonuclease fragments,Perl and Bioperl.

INTRODUCTION

Infectious diseases caused by various pathogenic bacteria are considered to be a major public health problem globally. Several pathogens have been reported to challenge the existing treatment regime by developing drug resistance and in several cases effective vaccines are yet to be developed. Although, researches are going on to develop effective drugs and vaccines, the efforts are not yet successful due to the dynamic adaptability, frequent phase and antigenic modifications, variations in major virulence factors, and adoptive mutations. The advent of several microbial complete genome sequences along with development of various bioinformatics tools, made it faceable for in silico analysis of the genomes and subsequent drug discovery against deadly human pathogen. To date, NCBI genome database has listed approximately 4704 fully sequenced microbial genomes including pathogenic bacteria and computational approaches based on subtractive genomics have successfully been used to identify drug targets in many pathogenic bacteria.

The first whole organism to be sequenced was that of bacteriophage ΦX174, by Sanger. His group also went on to complete the genome of bacteriophage λ, making use of cloned restriction endonuclease fragments. Whole genome sequencing is now commonplace; the first genome from a free-living organism, Haemophilus influenzae was published in 1995, and others followed rapidly. Such a flood of raw sequence information obviously requires refinement and further analysis. Both in silico and in vitro research can assist in this area.

In the field of raw genomic sequence annotation, in silico prediction of genes and open reading frames (ORFs) enables the rapid identification of coding regions in prokaryotic sequences. One of the difficulties in discovering microbial genes in a stretch of sequence is that the gene density is greatly increased in prokaryotes. At its simplest, one assumes that any ORF above a reasonable threshold (~300 bps) contains a coding sequence or gene, but in doing so, smaller moieties may be missed. Bacterial genomes may also contain overlapping coding regions due to frameshift mutations. Therefore, the preferred prediction software used in prokaryotic genome interpretation is GLIMMER, which caters for overlapping ORFs and high gene density using an interpolated Markov Model (IMM). GeneScan may also be used, although it is more suited to the analysis of eukaryotic genomes. With post-genomic research comes the rapid growth of proteomics, the study of a particular species' complete protein repertoire encoded for by its genome. Combining 2-D gel electrophoretic separation of complex protein mixtures with protein analytical methods like peptide massspectrometry allows identification of expressed proteins at a set point in time under certain conditions. Bioinformatics can play an important support role for

proteomics, permitting recognition of known proteins through the use of peptide mass databases like MOWSE (Molecular Weight SEarch), and thus highlighting uncharacterized proteins for further research.

Gene expression profiling using DNA microarrays allows researchers to monitor changes to an organism's mRNA expression either in response to a set of specified conditions, or compared to another genome. Combining genomics, biotechnology and computational biology, this method is ideal for analysing bacterial species, as virtually all of a microbe's ORFs can be fitted onto one microarray chip. An important role for bioinformatics evolved when the large microarray data output required investigation. By using algorithms that can group together similar elements i.e. genes displaying a homologous function, and reject redundant or dissimilar entries in the data, the output can be reduced to a manageable size.

BIOINFORMATICS TOOLS

The subtractive genomics approach is used to identify the putative drug targets for human pathogens. Tool is being developed using the language perl/bioperl and uses MySQL database to store the information.

Genomics

Genomics is the study of the genomes of organisms. The field includes intensive efforts to determine the entire DNA sequence of organisms and fine-scale genetic mapping efforts. The field also includes studies of intragenomic phenomena such as heterosis, epistasis, pleiotropy and other interactions between loci and alleles within the genome. In contrast, the investigation of the roles and functions of single genes is a primary focus of molecular biology or genetics and is a common topic of modern medical and biological research.

The first free-living organism to be sequenced was that of *Haemophilus influenzae* (1.8 Mb) in 1995, and since then genomes are being sequenced at a rapid pace. As of July 2010, NCBI contains the complete sequence of about 2469 viruses, 1101 bacterial species and roughly 38 eukaryote organisms (425 in progress) out of which 18 is belongs to fungi, 6 belongs to plants, 5 belongs to animals.

Subtractive Genomics

Subtractive genomic approach is one of the most useful strategies helpful in identification of potential targets. The approach works by subtracting the genes or proteins homologous to both host and the pathogen and identify those set of gene or proteins which are essential for the pathogen and are exclusively present in the pathogen. Subtractive genomic approach is employed to identify novel target in *salmonella typhi*.

Protein subcellular localization

Protein subcellular localization prediction involves the computational prediction of where a protein resides in a cell. Prediction of protein subcellular localization is an important component of bioinformatics-based prediction of protein function and genome annotation, and it can aid the identification of drug targets.

BIOPERL

BioPerl is a collection of Perl modules that facilitate the development of Perl scripts for bioinformatics applications. It has played an integral role in the Human Genome Project. It is an active open source software project supported by the Open Bioinformatics Foundation.

BioPerl provides software modules for many of the typical tasks of bioinformatics programming. These include:

- Accessing nucleotide and peptide sequence data from local and remote databases
- Transforming formats of database/ file records
- Manipulating individual sequences
- Searching for similar sequences
- Creating and manipulating sequence alignments
- Searching for genes and other structures on genomic DNA
- Developing machine readable sequence annotations

Bioperl is a large, object-oriented toolkit of interacting perl modules useful for building bioinformatics solutions in Perl. The collection of modules in the bioperl-live repository contains the core functionality. Additional packages for creating graphical interfaces (bioperl-gui), setting up persistent ORM storage in RDMBS (bioperl-db), running and parsing the results from hundreds of bioinformatics applications (bioperl-run), and software to automate bioinformatic analyses (bioperl-pipeline) are all available.

4 REQUIREMENT SPECIFICATION

The tool that will be developed will be a standalone tool useful to identify the putative drug targets in human pathogens. Later, the tool can be made an online application. The tool is being developed using the language perl/bioperl, as it supports many of the operations necessary for the implementation of the tool through its modules. It also uses MySQL database to store the information.

Operating system: *Windows*
Language: *PERL/ BIOPERL*
Tool: *ActivePerl with core libraries*

Database: MySQL
Additional Libraries: Bioperl libraries

The initial version of the tool uses NCBI online server to perform BLAST of query sequence against the sequences present in the database. Since NCBI provides a standalone BLAST tool to perform the same, it can be used in later versions of the tool, as it reduces the load on the NCBI online server.

RESULTS

The increasing number of complete bacterial genomes available in the public databases offers new opportunities for understanding the relationship between genotype and phenotype using in-silico genome comparisons. Subtractive genome analysis is an attempt to link genome content and phenotypic features according to the presence or absence of genes. The method is based on the assumption that the genes responsible for a specific function are conserved during evolution but lost in those genomes not showing that phenotype.

Therefore, this method is used to search for those genes which are present in a group of genomes having a common phenotype, but which are absent in another group not showing this phenotype, as for instance the capacity to grow in the presence of an antibiotic or the ability to synthesize an outer membrane. The objective of subtractive genome analysis was to find out the essential proteins, which play a key role in survival of bacteria within human and identify them as drug target to block the bacterial pathogenesis. The tool being developed uses the above subtractive genomics approach to find the non-homologous genes present in the pathogens which are absent in humans.

CONCLUSIONS

Large genomic sequencing projects of pathogens as well as human genome leads to immense genomic and proteomic data which would be very beneficial for the novel target identification in pathogens. Subtractive genomic approach is one of the most useful strategies helpful in identification of potential targets. A number of approaches for new vaccine development exist, such as sub-unit protein and DNA vaccines, recombinant vaccines, auxotrophic organisms to deliver genes and so on. Testing such candidates is tedious and expensive. *In-silico* approaches enable us to reduce substantially the number of such candidates to test and speed up drug discovery with least toxicity.

BIBLIOGRAPHY

[1] Kishore R. Sakharkar, Meena K. Sakharkar and Vincent T. K. Chow, "A novel genomics approach for the identification of drug targets in pathogens, with special reference to *Pseudomonas aeruginosa"*
[2] Chan-Eng Chong, Boon-San Lim, Sheila Nathan and Rahmah Mohamed, "*In silico* analysis of *Burkholderia pseudomallei* genome sequence for potential drug targets"
[3] Anirban Dutta, Shashi Kr. Singh, Payel Ghosh, Runni Mukherjee, Sayak Mitter and Debashis Bandyopadhyay, "*In silico* identification of potential therapeutic targets in the human pathogen *Helicobacter pylori"*
[4] Bhawna Rathi, Aditya N. Sarangi, Nidhi Trivedi, "Genome subtraction for novel target definition in *Salmonella typhi"*
[5] Gupta Sunil Kumar, Singh Sarita, Gupta Manish Kumar, Pant KK and Seth PK, "Definition of Potential Targets in *Mycoplasma pneumoniae* Through Subtractive Genome Analysis"
[6] Ren Zhang, Hong-Yu Ou and Chun-Ting Zhang, "DEG: a database of essential genes".
[7]Debmalya Barh, Anil Kumar, Amarendra Narayana Misra, "Genomic Target Database (GTD): A database of potential targets in human pathogenic bacteria"