Pathway Reconstruction of *Pseudomonas aeruginosa* by the integrated analysis of Genomics and Metabolomics: An initiation towards the identification of therapeutic markers

Aarthi Rashmi B [1], Harishchander A[2],Swetha T N[3],Shada Fathima M[4],Pravitha K S[4],Uma K[5] , Sowmya P[6], Noreen A Khan[7], Swathi T V[8], Nimitha P V[9] ,Deepak G[10], Thangaraj T L [11]

[1,2,3,4,5,6,7,7,9,10,11]*Department of Bioinformatics, Sri Krishna Arts and Science College, Coimbatore, Tamilnadu, India*
E-mail : aarthirashmib@skasc.ac.in

Abstract

Analysis of the Genome and Metabolites from the reads generated by high throughput technologies and reconstructing the forward and reverse reactions of metabolism in microbes is a challenging task to execute. The Challenge was addressed by the annotation of Genome and Metabolites from high throughput annotation servers like Webgstalt, DAVID and KAAS and performing a secondary analysis with respect to the principles of thermodynamics. Initially, analysis of the genome and metabolite mapping resulted in an association between 10,000 compounds (metabolites) and 30,000 thousand enzymes to carry out the forward and reverse reactions to activate and deactivate the metabolism but on the basis of thermodynamics only 638 compounds and 381 enzymes were stable on the basis of the free energy and the rate of the forward reaction is greater than the rate of reverse reaction and the maintenance of equilibrium to identify the circulation of an active metabolism between the enzymes and the secondary metabolites to infer the identification of active cascades in the process of signal transduction. In conclusion, the the forward reaction between glucosyltransferase and oligosaccharides were identified as an active cascade to drive the metabolism of Pseudomonas aeruginosa and inhibiting those targets will be useful in identifying the potential inhibitors to treat the infection caused by Pseudomonas aeruginosa.

Keywords: *Genome, Metabolites, Pseudomonas aeruginosa, Webgstalt, DAVID, KAAS, Enzymes, Compounds*.

1. Introduction

Pseudomonas aeruginosa is an encapsulated, Gram-negative, rod-shaped, asporogenous, and monoflagellated bacterium that can cause disease in plants and animals, including humans. It is a rod about 1-5 µm long and 0.5-1.0 µm wide The genome of P. aeruginosa

consists of a relatively large circular chromosome (5.5–6.8 Mb) that carries between 5,500 and 6,000 open reading frames *P. aeruginosa* is a multidrug resistant pathogen recognized for its ubiquity, its intrinsically advanced antibiotic resistance mechanisms and it has natural resistance to antibiotics. It thrives not only in normal atmospheres, but also in low-oxygen atmospheres. It uses a wide range of organic material for food; in animals, its versatility enables the organism to infect damaged tissues or those with reduced immunity.
It thrives on moist surfaces, this bacterium is also found on and in medical equipment, causing cross-infections in hospitals and clinics. *P.aeruginosa* relies on iron as a nutrient source to grow. Furthermore, excessively high levels of iron can be toxic to P. aeruginosa. To overcome this and regulate proper intake of iron, P. aeruginosa uses siderophores, which are secreted molecules that bind and transport iron. *P. aeruginosa*. can degrade aromatic hydrocarbons such as methylbenzenes, which are the by-products of petroleum industries. Methylbenzenes are considered as environmental contaminants that are present in the atmosphere, underground and soils, and in surface water. *P. aeruginosa* can break down toluene and can be used in pollution control.

Metabolic network reconstruction and simulation allows for an in-depth insight into the molecular mechanisms of a particular organism. In particular, these models correlate the genome with molecular physiology.A reconstruction breaks down metabolic pathways into their respective reactions and enzymes, and analyzes them within the perspective of the entire network. A reconstruction collects all of the relevant metabolic information of an organism and compiles it in a mathematical model. In general, the process to build a reconstruction is as follows:
a. Draft a reconstruction
b. Refine the model
c. Convert model into a mathematical/computational representation
d. Evaluate and debug model through experimentation

The integration of biochemical metabolic pathways with rapidly available, annotated genome sequences has developed what are called genome-scale metabolic models. Mechanically speaking, the process of reconstructing prokaryotic and eukaryotic metabolic networks is essentially the same. Eukaryote reconstructions are typically more challenging because of the size of genomes, coverage of knowledge, and the multitude of cellular compartments. A reconstruction is a systematic verification and compilation of data from various sources that takes into account all of the discrepancies. Metabolic comparisons can be performed between various organisms of the same species as well as between different organisms.

Metabolic network reconstructions and models are used to understand how an organism or parasite functions inside of the host cell. A reconstruction model serves as a first step to deciphering the complicated mechanisms surrounding disease. These models can also look at the minimal genes necessary for a cell to maintain virulence. The next step would be to use the predictions and postulates generated from a reconstruction model and apply it to discover novel biological functions such as drug-engineering and drug delivery techniques.

2. Materials and methods

So as to annotate the genome and metabolites of the organism as well as to find out the the rate of forward reaction and backward reaction, identify the circulation of an active metabolism between the enzymes and the secondary metabolites to infer the identification of active cascades in the process of signal transduction.

2.1 DAVID

DAVID - Database for Annotation, Visualization and Integrated Discovery .It is a free online bioinformatics resource originally developed by Dr. Glynn Dennis. All tools in the DAVID Bioinformatics Resources aim to provide functional interpretation of large lists of genes derived from genomic studies, e.g microarray and proteomics studies.

The DAVID Bioinformatics Resources consists :
a. DAVID Gene Functional Classification Tool
b. DAVID Functional Annotation Tool
c. DAVID Gene ID Conversion Tool
d. DAVID Gene Name Viewer
e. DAVID NIAID Pathogen Genome Browser

The expanded DAVID Knowledgebase now integrates almost all major and well-known public bioinformatics resources. It uses a single-linkage method to agglomerate tens of millions of diverse gene/protein identifiers and annotation terms from a variety of public bioinformatics databases. For any uploaded gene list, the DAVID Resources now provides not only the typical gene-term enrichment analysis, but also new tools and functions that allow users to condense large gene lists into gene functional groups, It allows convertion between gene/protein identifiers, cluster redundant and heterogeneous terms into groups, search for interesting and related genes or terms, dynamically view genes from their lists on bio-pathways.

2.2 KAAS

KAAS - KEGG Automatic Annotation Server It provides functional annotation of genes by BLAST or GHOST comparisons against the manually curated KEGG GENES database. The result contains KO (KEGG Orthology) assignments and automatically generated KEGG pathways. KAAS works best when a complete set of genes in a genome is known. The BBH (bi-directional best hit) method is used to assign orthologs. KAAS can also be used for a limited number of genes. The SBH (single-directional best hit) method is used to assign orthologs. When the query consists of large numbers of sequences and / or sequences from mixture of species such as those from metagenome sequencing project, the GHOSTX search and SBH method is used.

2.3 WEBGSTALT

WebGestalt (WEB-based Gene SeT Analysis Toolkit) is a functional enrichment analysis webtool. It provides emphasise on providing user friendly interfaces which could directly translate into publication- ready figures. It supports three-well established and complementary methods for enrichment analysis, including Over-Representation Analysis(ORA),Gene Set Enrichment Analysis(GSEA), and Network Topology – based Analysis (NTA).

3. Result

With the help of various tools like DAVID, KAAS and WEBGSTALT annotation of the genome as well as the metabolites were performed, also secondary analysis of the metabolites was performed , and resulted in obtaining an association between 10000 compounds (metabolites) and 30000 enzymes to carry out the forward and reverse reactions to activate and deactivate the metabolism, on the basis of thermodynamics principle(free energy) only 638 compounds and 381 compounds were stable whose rate of forward reaction is greater than reverse reaction.
Thus it is found that forward reaction between glucosyltransferase and oligosaccharides were identified as an active cascade to drive the metabolism of Pseudomonas aeruginosa.

Below table gives the details regarding Delta G ,charge and kegg pathways of different metabolites and enzymes whose association results in forward reaction.

Table.1

Compound name	Delta G	Charge
phospho-heptosyl-phospho-heptosyl-heptosyl-kdo2-lipidA_c0	-2460.38	-10
glucosyl-inner core oligosaccharide lipid A_c0	-2976.94	-11
glucosyl-glucosyl-galactosyl-glucosyl-inner core oligosaccharide lipid A_c0	-3452.38	-11
glucosyl-galactosyl-glucosyl-inner core oligosaccharide lipid A_c0	-3293.9	-11
galactosyl-glucosyl-inner core oligosaccharide lipid A_c0	-3135.42	-11
inner core oligosaccharide lipid A_c0	-2818.46	-7
kdo-phospho-heptosyl-phospho-heptosyl-heptosyl-kdo2-lipidA_c0	-2696.21	-11
core oligosaccharide lipid A_c0	-3647.28	-11
phospho-heptosyl-heptosyl-kdo2-lipidA_c0	-2052.98	-8
heptosyl-phospho-heptosyl-heptosyl-kdo2-lipidA_c0	-2247.88	-8

Enzyme	Delta G	KEGG pathways
pyridoxine 5'-phosphate synthase_c0	-22.13	Vitamin B6 metabolism
dimethyallyl diphosphate:NADP+ oxidoreductase_c0	-23.02	Alanine, aspartate and glutamate metabolism

UDP-L-rhamnose:flavonol-3-O-D-glucoside L-rhamnosyltransferase_c0	-19.42	Ubiquinone and other terpenoid-quinone biosynthesis
Uroporphyrinogen-III carboxy-lyase_c0	-19.6	Porphyrin and chlorophyll metabolism
5,10-methylenetetrahydrofolate,NADPH:dUMP C-methyltransferase_c0	-16.47	Pyrimidine metabolism
R05221_c0	-14.03	Porphyrin and chlorophyll metabolism
(S)-2,3-Epoxysqualene mutase (cyclizing, lanosterol-forming)_c0	-86.57	Steroid biosynthesis
5,10-Methylenetetrahydrofolate:dUMP C-methyltransferase_c0	-15.88	Pyrimidine metabolism
Pyridoxine 5-phosphate:oxygen oxidoreductase_c0	-28.04	Vitamin B6 metabolism
L-tryptophan:oxygen 2,3-oxidoreductase (decyclizing)_c0	-92.1	Tryptophan metabolism

Table.2

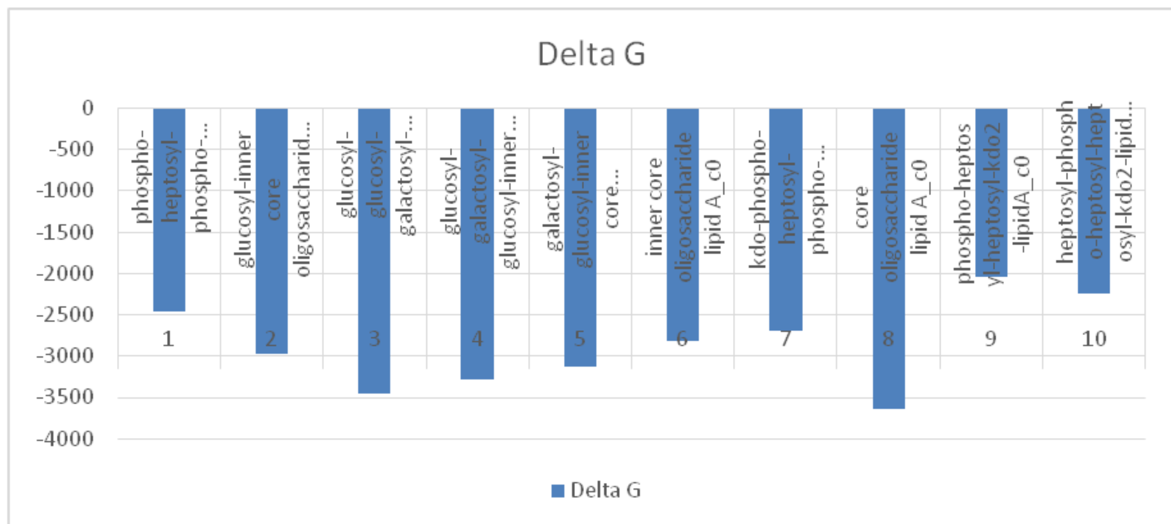

Figure.1

4. Conclusion

Thus by performing genome annotation and metabolite mapping of pseudomonas aeroginosa,we obtained a specific metabolite and enzyme association namely glucosyltransferase and oligosaccharides which are an active cascade to drive the metabolism of Pseudomonas aeruginosa. All these association between the metabolite and enzyme was obtained using the tools like KAAS, DAVID and WEBGSTALT. And considering the organism it thrives even in low-oxygen level areas and can eaily affect

damaged tissues, also as found out metabolism of the organism is activated due to the association of glucosyltransferase and oligosaccharides, therefore by finding out a potential inhibitor we can inhibit the metabolism of the organism thereby preventing these from infecting the tissues.