Infection, Genetics and Evolution 27 (2014) 300–308

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Unraveling novel broad-spectrum antibacterial targets in food and waterborne pathogens using comparative genomics and protein interaction network analysis

Ankush Jadhav, Buvaneswari Shanmugham, Anjana Rajendiran, Archana Pan *

Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Pondicherry 605014, India

ARTICLE INFO

Article history:
Received 20 March 2014
Received in revised form 31 July 2014
Accepted 7 August 2014
Available online 14 August 2014

Keywords:
Foodborne
Waterborne
Broad-spectrum
Target
Comparative genomics
Protein network

ABSTRACT

Food and waterborne diseases are a growing concern in terms of human morbidity and mortality worldwide, even in the 21st century, emphasizing the need for new therapeutic interventions for these diseases. The current study aims at prioritizing broad-spectrum antibacterial targets, present in multiple food and waterborne bacterial pathogens, through a comparative genomics strategy coupled with a protein interaction network analysis. The pathways unique and common to all the pathogens under study (*viz*., methane metabolism, D-alanine metabolism, peptidoglycan biosynthesis, bacterial secretion system, two-component system, C5-branched dibasic acid metabolism), identified by comparative metabolic pathway analysis, were considered for the analysis. The proteins/enzymes involved in these pathways were prioritized following host non-homology analysis, essentiality analysis, gut flora non-homology analysis and protein interaction network analysis. The analyses revealed a set of promising broad-spectrum antibacterial targets, present in multiple food and waterborne pathogens, which are essential for bacterial survival, non-homologous to host and gut flora, and functionally important in the metabolic network. The identified broad-spectrum candidates, namely, integral membrane protein/virulence factor (MviN), preprotein translocase subunits SecB and SecG, carbon storage regulator (CsrA), and nitrogen regulatory protein P-II 1 (GlnB), contributed by the peptidoglycan pathway, bacterial secretion systems and two-component systems, were also found to be present in a wide range of other disease-causing bacteria. Cytoplasmic proteins SecG, CsrA and GlnB were considered as drug targets, while membrane proteins MviN and SecB were classified as vaccine targets. The identified broad-spectrum targets can aid in the design and development of antibacterial agents not only against food and waterborne pathogens but also against other pathogens.

1. Introduction

Food and waterborne bacterial pathogens are one of the leading causes of mortality and morbidity in developing as well as developed nations. The illnesses attributed to these pathogens are devastating, having a negative impact on economic growth. According to the CDC (Centers for Disease Control and Prevention), foodborne pathogens alone cause 76 million illnesses, about 325,000 hospitalizations and a death toll of 5000, while waterborne pathogens cause 4–32 million acute gastrointestinal illness each year in the United States (Colford et al., 2006; Messner et al., 2006; Newell et al., 2010). A significant proportion of food and waterborne outbreaks are caused solely by bacteria. The important bacterial pathogens that are transmitted through intake of contaminated food and water are *Escherichia coli*, *Campylobacter jejuni*, *Listeria monocytogenes*, *Staphylococcus aureus*, *Salmonella enterica*, *Shigella*, *Vibrio* and *Yersinia* species. The foremost infections that are caused by these pathogens include gastrointestinal infections, listeriosis, meningitis, septicemia, salmonellosis, shigellosis, cholera, gastroenteritis and bacteremia (Rayan and Ray, 2004). Infants, young children, elderly people, pregnant women and people with immunodeficiency are usually susceptible to food and waterborne diseases (Lund and O'Brien, 2011). The incidence of these diseases has increased over time with various etiological agents propelling in and with the escalation of resistance to existing antibiotic therapy. A majority of food and waterborne pathogens have become resistant to currently used antibiotics. This scenario accentuates the need to develop new drugs to replace existing antibiotics. Target identification is a key step in the drug discovery process.

* Corresponding author. Tel.: +91 413 2654584; fax: +91 413 2655211.
E-mail addresses: archana@bicpu.edu.in, archanpan@gmail.com (A. Pan).

http://dx.doi.org/10.1016/j.meegid.2014.08.007
1567-1348/

Toward this end, the present study aims to identify broad-spectrum targets in different food and waterborne bacterial pathogens. The discovery of broad-spectrum targets facilitates designing broad-spectrum antibacterial agents that are effective against a wider spectrum of microorganisms. The benefit of using broad-spectrum agents is that it is not necessary to identify the infecting pathogen with certitude on commencing the treatment.

Computational approaches have been widely used to search for novel drug targets as experimental approaches are tedious, time consuming, expensive and often result in very few target candidates. Comparative genomics, subtractive genomics and protein network analysis are extensively used in the prediction and identification of potential drug targets in numerous life threatening pathogens (Barh et al., 2011; Butt et al., 2012; Raman and Chandra, 2008; Shanmugham and Pan, 2013; Sharma and Pan, 2012). In principle, the former two approaches rely on two criteria, namely, essentiality and selectivity/specificity, for prioritizing therapeutic candidates. A potential target must be essential for the growth and survival of the pathogen and exclusive to the pathogen in order to avoid undesirable cross-reactivity with host proteins. Protein network analysis is based on a protein's functional importance in a metabolic network (Kushwaha and Shakya, 2010; Raman et al., 2008). Understanding the interaction mechanism between pathogen and host requires further investigation via host-pathogen protein–protein interaction studies (Zhou et al., 2013a,b, 2014). The availability of complete genome sequences for several pathogenic microorganisms in combination with bioinformatics methods and databases is of great assistance in reducing the problem of searching for potential drug targets/broad-spectrum targets from a large list of gene/protein polls. In the present study, we have uncovered five broad-spectrum antibacterial targets by analyzing 14 food and waterborne bacterial pathogens through a comparative genomics strategy coupled with protein interaction network analysis.

2. Materials and methods

Various clinically important bacterial food and waterborne pathogens, as summarized in Table 1, were subjected to a broad-spectrum antibacterial target identification protocol using comparative genomics and protein interaction network analysis. The complete workflow of the analyses is shown in Fig. 1.

2.1. Comparative genomics approach

2.1.1. Comparative metabolic pathway analysis

The metabolic pathways of the host and 14 food and waterborne pathogens (Table 1) under study were retrieved from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al., 2012). The pathways of the host and pathogens were manually compared to identify pathways unique to the pathogen. Those pathways that were absent in the host and present in the pathogen were selected as unique pathways. The unique pathways identified by KEGG were also cross-checked using IntPath, a comprehensive integrative database for analysis of model organisms and important pathogen pathways (Zhou et al., 2012). The pathways that were common to all pathogens were then identified from the list of unique pathways. Unique pathways that are conserved among several pathogens are potential candidates to harbor broad-spectrum antimicrobial targets. As the objective of the present study is to identify broad-spectrum drug targets, the proteins involved in common pathogen-specific pathways were used exclusively as the input dataset. The amino acid sequences of these common pathway proteins with a valid EC number were retrieved from the KEGG database and passed through the further analyses.

2.1.2. Host non-homology analysis

All protein sequences resulting from the comparative metabolic pathway analysis were subjected to a protein BLAST (BLASTp) search (Altschul et al., 1990) against the non-redundant database of the human proteome, with the *e*-value threshold set at 0.0001 and bit score cutoff at 100 (Jadhav et al., 2013). Sequences which showed *e*-value > 0.0001 with bit score < 100 were marked as proteins having no significant homologs in human and were selected for a domain similarity test. The domain architecture of the non-homologous proteins was compared with that of human proteins using Pfam (Punta et al., 2012) and the SMART database (Schultz et al., 1998). Those pathogen proteins having no similar domain in the human proteome were sorted out and considered for further analysis in the broad-spectrum target prioritization protocol.

2.1.3. Essentiality analysis

The Database of Essential Genes (DEG) is a repository of genes indispensable to support cellular life that comprises essential genes from 31 prokaryotic and 10 eukaryotic species (Zhang and Lin, 2009). The functions encoded by essential genes are considered a foundation of life and thus are expected to be common to all cells. Query genes homologous to essential genes are likely to be essential for bacterial survival. Thus, the non-homologous protein sequences resulting from host non-homology analysis were subject to a BLASTp search against DEG 10.0. Any protein which showed hits with *e*-value < 0.0001, bit score > 100 and similar function was considered as orthologous to gene products in the DEG (Jadhav et al., 2013) and likely to be essential in pathogens. In addition, these essential gene products were subjected to a Cluster of Orthologous Groups (COG) search (Tatusov et al., 2000) in order to assess their conservation in other pathogens, as it is believed that the essential genes are conserved among pathogens (Kaufimann, 2006).

2.1.4. Gut flora non-homology analysis

Approximately 10^{14} harmless microorganisms, known as gut flora, inhabit the intestinal tract of human and are crucial for maintaining human health (Jacobs et al., 2009). The unintentional inhibition of proteins in gut microbes would eliminate the beneficial gut flora population, eventually increasing the colonization by pathogenic bacteria and causing adverse effects in the host (Savage, 1977). Hence, pathogen proteins homologous to the gut flora proteome should not be considered as drug targets. In the present study, the shortlisted unique, essential and conserved pathogenic proteins were compared with proteomes of 79 gut flora organisms (Supplementary Text 1) using BLASTp (Altschul et al., 1990; Shanmugham and Pan, 2013). Those proteins showing *e*-value > 0.0001 with bit score < 100 and non-homologous to $\geqslant$40 gut flora organisms were considered as non-homologous to gut flora proteomes (Jadhav et al., 2013).

2.2. Protein interaction network analysis

A protein–protein interaction network of the selected proteins was constructed using STRING 9.05, which includes more than 1100 completely sequenced organisms (Szklarczyk et al., 2011). In this database, the functional associations among proteins were derived by both experimental and computational techniques. It has been reported by Zhou and Wong (2011) that protein interaction networks from the STRING database with high confidence interactors have good agreement with correlated gene expression profiles, coherent informative GO term annotations and conservation in other organisms. However, the STRING database also has certain limitations, including the absence of a specific source on each protein–protein interaction (PPI) (Zhou and Wong, 2011). In