Genomics 104 (2014) 24–35

Contents lists available at ScienceDirect

Genomics

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

Prioritizing drug targets in *Clostridium botulinum* with a computational systems biology approach

Syed Aun Muhammad [a,b], Safia Ahmed [a,*], Amjad Ali [f,*], Hui Huang [b], Xiaogang Wu [b], X. Frank Yang [e], Anam Naz [f], Jake Chen [b,c,d]

[a] Department of Microbiology, Quaid-I-Azam University Islamabad, Pakistan
[b] Indiana Center for Systems Biology and Personalized Medicine, Indiana University–Purdue University, Indianapolis, IN, USA
[c] School of Informatics, Indiana University, Indianapolis, IN, USA
[d] Department of Computer & Information Science, School of Science, Purdue University, Indianapolis, IN, USA
[e] Department of Microbiology and Immunology, Indiana University School of Medicine, Indianapolis, IN, USA
[f] Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad, Pakistan

ARTICLE INFO

Article history:
Received 24 March 2014
Accepted 5 May 2014
Available online 13 May 2014

Keywords:
Drug targets prioritization
Clostridium botulinum
In silico analysis
Interactome

ABSTRACT

A computational and *in silico* system level framework was developed to identify and prioritize the antibacterial drug targets in *Clostridium botulinum* (*Clb*), the causative agent of flaccid paralysis in humans that can be fatal in 5 to 10% of cases. This disease is difficult to control due to the emergence of drug-resistant pathogenic strains and the only available treatment antitoxin which can target the neurotoxin at the extracellular level and cannot reverse the paralysis. This study framework is based on comprehensive systems-scale analysis of genomic sequence homology and phylogenetic relationships among *Clostridium*, other infectious bacteria, host and human gut flora. First, the entire 2628-annotated genes of this bacterial genome were categorized into essential, non-essential and virulence genes. The results obtained showed that 39% of essential proteins that functionally interact with virulence proteins were identified, which could be a key to new interventions that may kill the bacteria and minimize the host damage caused by the virulence factors. Second, a comprehensive comparative COGs and blast sequence analysis of these proteins and host proteins to minimize the risks of side effects was carried out. This revealed that 47% of a set of *C. botulinum* proteins were evolutionary related with *Homo sapiens* proteins to sort out the non-human homologs. Third, orthology analysis with other infectious bacteria to assess broad-spectrum effects was executed and COGs were mostly found in Clostridia, Bacilli (Firmicutes), and in alpha and beta Proteobacteria. Fourth, a comparative phylogenetic analysis was performed with human microbiota to filter out drug targets that may also affect human gut flora. This reduced the list of candidate proteins down to 131. Finally, the role of these putative drug targets in clostridial biological pathways was studied while subcellular localization of these candidate proteins in bacterial cellular system exhibited that 68% of the proteins were located in the cytoplasm, out of which 6% was virulent. Finally, this framework may serve as a general computational strategy for future drug target identification in infectious diseases.

1. Introduction

Drug target identification is the first step in the drug discovery process [1] and the completion of pathogenic bacterial genome sequences has increased the momentum within the field of drug discovery and vaccine target identification [2,3]. Because of the availability of both pathogen and host genome sequences, it has become easier to identify drug targets at the genomic level for any given pathogen [4]. Traditionally, targets have been identified through functional knowledge of individual protein molecules, where their function has been well characterized. The cost of research progress in the pharmaceutical industry has been growing steeply and steadily in the last decade, but the amount of time required bringing a new product to market remains around ten to fifteen years [5]. *In silico* methods have the advantages of speed, low cost and, even more importantly, providing a systems view of the whole microbe. This enables researchers to ask questions that are otherwise difficult to address experimentally. There is a progressive development in drug discovery procedures from conventional ligand based drug discovery to structural and targeted based drug designing approaches by targeting the key molecular pathways of the diseases. Using systems biology concepts and understanding the

* Corresponding authors.
E-mail addresses: aunmuhammad78@yahoo.com (S.A. Muhammad), safiamrl@yahoo.com (S. Ahmed), amjad_uni@yahoo.com (A. Ali), h.huang@medeolinx.com (H. Huang), wu33@iupui.edu (X. Wu), xfyang@iupui.edu (X.F. Yang), anam.naz88@live.com (A. Naz), jakechen@iupui.edu, jyc298@yahoo.com (J. Chen).

http://dx.doi.org/10.1016/j.ygeno.2014.05.002
0888-7543/

microbe as a whole opens up new prospects for computational drug target identification [6].

In recent years, the drug discovery strategies are shifting progressively to genomic, proteomic and metabolomic approaches [7] to identify novel drug targets for the design of new defenses against antibiotic-resistant pathogens [8]. Currently, genomics and, more specifically *in silico* comparative and functional genomics, are being widely used to identify novel drug and vaccine targets in order to develop effective antibacterial agents and vaccines against bacterial pathogens that are either resistant to existing antibacterial regimens or for which a suitable vaccine is not available [9,10]. It has been proposed that a systems-level analysis of the genes, proteins, and interactions involved is the key to gaining insights into routes required for drug targets and drug resistance. One of the prerequisites of such an analysis is the existence of a comprehensive protein interactome of the relevant pathogen [11].

Botulism, a rare infection caused by *Clostridium botulinum* (*Clb*) that produces one of several toxins collectively known as botulinum neurotoxins, the most potent toxins known to man and induce a potentially fatal paralytic condition in humans [12]. The toxin enters the body in one of three ways: colonization of the digestive tract by the bacterium in children (infant botulism) or adults (adult intestinal toxemia), ingestion of toxin from foods (food-borne botulism), or contamination of a wound (wound botulism) by the bacterium. The ratio of this deadly infection is increased if cure is not appropriate and instant [13] and can be fatal in 5 to 10% of cases. Approximately an average of 145 cases of botulism occurs in the U.S. per year and of these around 15% is foodborne botulism, 65% are infant botulism, and 20% are wound botulism (CDC Reports, 2011). In the west of the U.S. Mississippi River, it is most common and approximately half of the cases of food-borne botulism are caused by home-canned foods while wound botulism has been reported in California and caused by black-tar heroin injection [14].

Though a number of antibiotics are recently available, the rise of drug-resistant pathogenic strains has made *C. botulinum* difficult to control [15]. Moderate resistance to chloramphenicol, tetracycline, cephalosporins and nalidixic acid [16] and a high level of resistance to cycloserine, nitroimidazole, gentamycin, sulfamethoxazole, and trimethoprim by *Clb* have been observed [17]. The only available treatment for botulism is administration of botulinum antitoxin, which can arrest the progression of paralysis and decrease its duration. However, antitoxin against the botulism neurotoxins can only target the toxins at the extracellular level and cannot reverse the paralysis caused by botulism. In addition, an antibody against the botulism neurotoxins can cause severe hypersensitivity reactions and is limited to prophylaxis use [18].

Here, we report an exclusive computational and database system-level framework to confidently find and prioritize clostridial drug targets that is based on protein- and ortholog-network analyses, genome sequence analyses and evolutionary studies (Fig. 1). This systems-level paradigm (i) integrates the use of biological network tools and databases with systems-level information about drug targets, (ii) retrospectively and prospectively assesses network-based relationships between essential and virulent proteins, (iii) analyzes the evolutionary diversity and similarity of orthologs between species, (iv) minimizes the risks of disturbing the normal ecosystem of endogenous human gastrointestinal microbial flora by comparative genomic analysis, (v) develops and reports the integrated unique biochemical and functional metabolic pathways, and (vi) predicts subcellular localization of

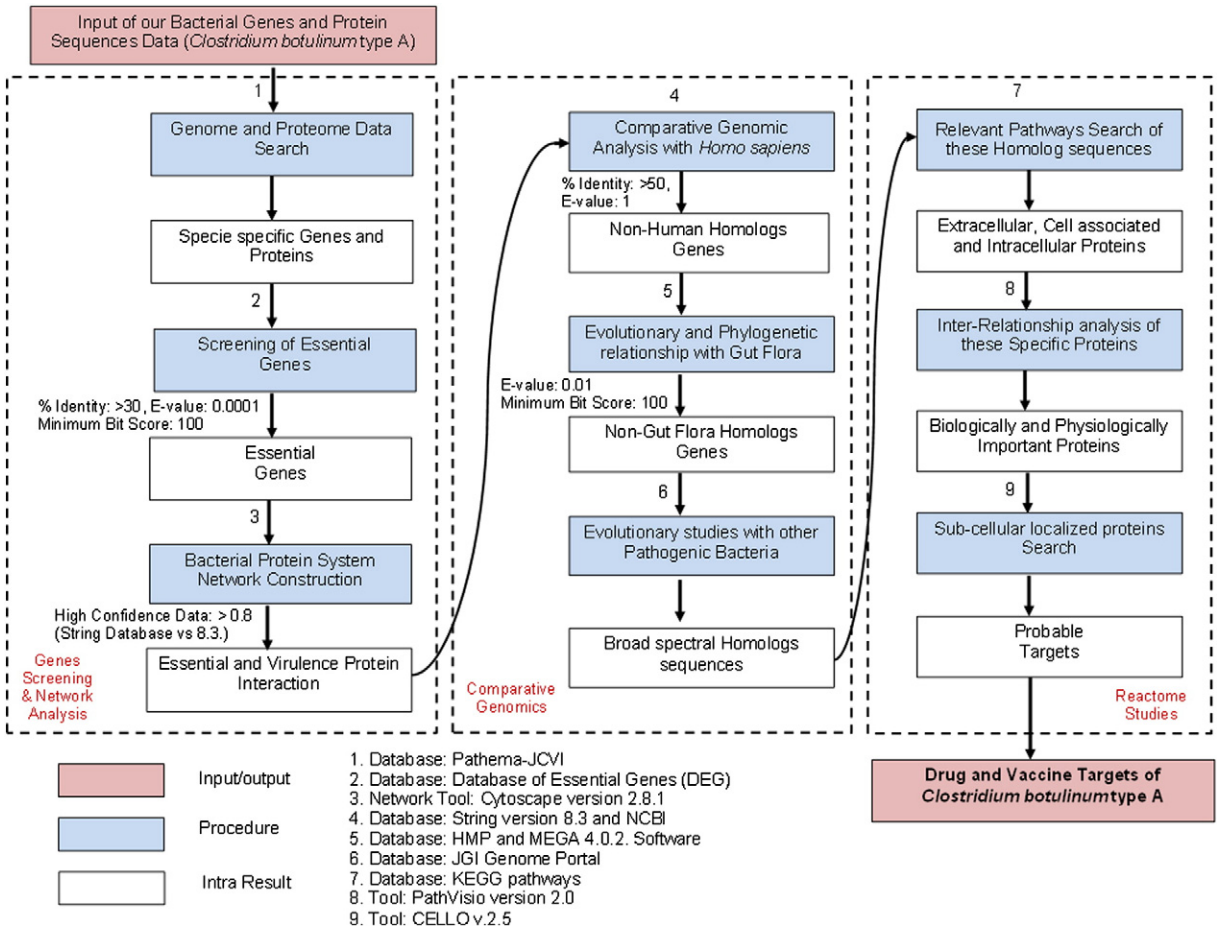

Fig. 1. Steps involved in target prioritization in *C. botulinum* type A by computational approach.