Subtractive Genomic Analysis for Identification of Novel Drug Targets and Vaccine Candidates against *Bartonella bacilliformis* subsp. Ver097

Md. Tahsin Khan¹, Araf Mahmud¹, Md. Asif Iqbal¹, Mahmudul Hasan^{2*}

¹Department of Genetic Engineering and Biotechnology, Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh.

²Department of Pharmaceuticals and Industrial Biotechnology, Sylhet Agricultural University, Sylhet-3100, Bangladesh.

***Corresponding Author:**

Mahmudul Hasan,

Assistant professor

Department of Pharmaceuticals and Industrial Biotechnology

Sylhet Agricultural University, Sylhet-3100, Bangladesh.

008801723698461, mhasan.pib@sau.ac.bd

Abstract

Bartonella bacilliformis the causative agent of Carrión's disease, one of the truly neglected tropical diseases found in Peru, Colombia and Ecuador. Recent evidence predicts that *Bartonella bacilliformis* subsp. ver097 can emerge as an antibacterial resistant strain and hence identification of novel drug targets is a crying need. Subtractive genome analysis of *B. bacilliformis* subsp. ver097 was successfully done in order to address the challenges. Various computational tools and online based servers were used to screen out human homologous proteins of pathogen and proteins involved in common metabolic pathways of host and pathogen. Only 7 proteins involved in pathogen specific pathways were further analyzed to identify membrane proteins. 'Flagellar biosynthesis protein FlhA' and 'ABC transporter permease' were found to be novel as targets according to DrugBank database. To avoid side effects in human while administering drugs, human 'anti-targets' analysis was performed to confirm non-homology of selected novel drug targets. Both predicted proteins also showed probability of antigenicity prediction through VaxiJen, however, 'Flagellar biosynthesis protein FlhA' showed broad spectrum conservancy with *Bartonella* strains. Therefore, Flagellar biosynthesis protein FlhA could facilitate the development of novel drugs and therapeutic compounds along with vaccines for efficient treatment of infections caused by *Bartonella bacilliformis* subsp. ver097.

Keywords: *Bartonella bacilliformis*, Carrión's disease, Subtractive genomics, Drug target, Vaccine

Introduction

Bartonella bacilliformis, a gram-negative, intracellular, facultative, aerobic[1] and a member of the alpha-2 subgroup Proteobacterium is the causative agent of Carrión's disease in humans. This pathogen is highly motile by 3 to 10 micrometers long polar lophotrichous flagella and appendages consisting of approximately 42kDa flagellin protein subunits[2]. Carrión's disease has been reported as an endemic in the river valleys of Peru, Ecuador as well as Colombia[3, 4]. However, Peru is considered as the most important endemic area for Bartonellosis[5]. The only known natural reservoir for this pathogen is human. *Bartonella bacilliformis* is believed to be transmitted by the bites of a sand fly, *Lutzomyia verrucarum* as well as other possible closely related sand flies[4].

Carrión's disease can occur in two clinical forms; verruga peruana-generally is a non-fatal form characterized by nodular cutaneous eruptions, and the other is Oroya fever, a frequently fatal disease in absence or delay of antibiotic treatments[2]. Verruga-peruana generally persists 3 to 6 months and wart-like hemangiomatous lesions are produced in the skin along with mucous membranes due to the entrance of the bacteria in the endothelial cells. The lesions are classified as miliary (small reddish papules less than 3 mm in diameter), mular (nodular tumors greater than 5 mm in diameter) and diffuse subdermal nodules[6]. However, Oroya fever, commonly found in children [8] is a life-threatening acute phase in which erythrocytes are destroyed by the spleen due to the entrance of bacteria in erythrocytes[7], and it has a high fatality rate (up to 88%) while untreated [8]. General malaise, myalgia, headache, hepato/splenomegaly, palor, jaundice besides fever are the main symptoms of this phase[3]. Some patients may develop a transient immunosuppression, as a consequence of the septic state which results in the eve of

other opportunistic infections, for example, toxoplasmosis, tuberculosis, salmonellosis, malaria, shigellosis, histoplasmosis and, pneumocystis phase [3, 9, 10].

The majority of pathogens causing infections and diseases are resistant to more than one drug[11]. Though studies suggested that most of *Bartonella* species including *B. bacilliformis* are highly susceptible to most of antibiotics, such as beta-lactams, aminoglycosides, chloramphenicol, rifampin, macrolides, tetracyclines, cotrimoxazole, fluoroquinolones[12] and gentamicin[1], but there is a potential risk of the spontaneous selection of successful antibiotic resistant *B. bacilliformis* clone in a clinical setting. For examples, few studies developed in vitro has showed the ability of *B. bacilliformis* to acquire chromosomal mutations leading to high levels of antibiotic resistance [13,14]. Lack of comprehensive analysis of *B. bacilliformis* genomic structure is an obstacle in developing species-specific new therapeutic compounds and vaccine candidates [5].

In recent years, identification of species-specific drug targets and vaccine candidates has been performed on several pathogenic strains through subtractive genomic approaches [15-18]. Analysis of mutational divergence of core genes revealed the emergence of multiple sub-species or outlier for *B. bacilliformis* [19, 20]. In the study, proteome of *Bartonella bacilliformis* supsp. ver097 was retrieved for applying subtractive genomic approaches. Various computational tools were used to identify proteins that are essential for the survival of the pathogen. Host non-homology analysis, metabolic pathways analysis were performed to avoid both cross-reactivity of drugs with human proteins and involvement of pathogenic proteins in host metabolism respectively. This study was further augmented to identify membrane proteins showing novelty as drug targets. Antigenicity testing was performed to check whether the identified novel drug targets. The study would be helpful to develop effective broad spectrum drugs target and vaccine candidates against *Bartonella bacilliformis*.

Materials and methods

The whole proteome of *Bartonella bacilliformis* subsp. ver097 was analyzed according to subtractive genomic approach to recognize novel drug targets as well as vaccine candidates. The overall workflow is illustrated in Figure 1.

Retrieval of Complete Proteome (Set0)

The whole proteome of *B. bacilliformis* subsp. ver097 (Assembly GCF_000709795.1) was retrieved from NCBI Genome database (<https://www.ncbi.nlm.nih.gov/genome>).

Identification of Paralogous Sequences (Set1)

To remove the paralogous sequences from the *B. bacilliformis* subsp. ver097 proteome, the protein sequences of set0 were subjected to CD-Hit suite[21](http://weizhongli-lab.org/cdhit_suite/cgi-bin/index.cgi?cmd=cd-hit). The server is programmed to take input as a fasta format sequence database which produces output as a set of non-redundant as well as representative sequences. To eliminate redundant sequences through this program, cutoff score 0.6 of sequence identity was set. Paralogous proteins with more than 60% identity were excluded from set0. Moreover, proteins <100 amino acids were excluded from set0. The remaining proteins were listed as set1 as non-paralogous protein sequences.

Identification of Protein Sequences Non-homologous to the Proteome of Human (Set2)

The aim of this step was to avoid functional similarity with human proteome in order to prevent bindings of therapeutic compounds to the active site of the host homologous proteins. Non-paralogous proteins of the pathogen were analyzed using BLASTp against human Refseq proteome in Ensemble Genome Database (<https://uswest.ensembl.org/Multi/Tools/Blast?db=corE>)[22]. Proteins of the pathogen were considered host homologous proteins if any significant hit above the threshold value 10^{-4} was

found. These homologous proteins were filtered while non-homologous proteins were listed as set2.

Identification of Essential Non-homologous Proteins (Set3)

Host non-homologous proteins of the pathogen listed in set2 were subjected to the Database of Essential Genes (DEG)[23]. DEG15.2 server contains 53,885 essential genes along with 786 essential non-coding sequences. By selecting all the strains of organisms present DEG 15.2 server, BLASTp was carried out using threshold value 10^{-10} and minimum bit score 100 as parameters. Proteins hit with expectation value $\leq 10^{-100}$, identity $\geq 25\%$ were listed in set3 considering as essential proteins of the pathogen. Products of essential genes constitute a minimal genome for a bacterium, that play key roles in the field of synthetic biology[24].

Analysis of Metabolic Pathway (Set4-a and Set-b)

Metabolic pathways of *B. bacilliformis* subsp. ver097 were analyzed against the human metabolic pathways through KEGG server[25]. KEGG is Kyoto Encyclopedia of Genes and Genomes which contains complete metabolic pathways present in living organisms. All metabolic pathways of *B. bacilliformis* and host were collected from KEGG PATHWAY Database[26] using their respective three letter KEGG organism code; 'bbk' for *B. bacilliformis* and 'has' for *H. sapiens*. A comparison was made in order to recognize the unique metabolic pathways only present in the pathogen, while the remaining pathways of the pathogen were grouped as common.

Identified host non-homologous, essential proteins of *B. bacilliformis* subsp. ver097 were subjected to BLASTp through KAAS server[27] at KEGG database. Proteins present only in the unique metabolic pathways of the pathogen were listed as set4-a and proteins involved in

common pathways or only in host pathways were excluded. KAAS provides functional annotation of genes through a blast comparison against the manually curated database of KEGG GENES. KO (KEGG Orthology) assignments specify metabolic proteins. KAAS also generates KEGG pathways which indicate these metabolic proteins. Proteins having no KO assignment were listed as set4-b.

Prediction of Subcellular Localization of set4-a

Since *B. bacilliformis* subsp. ver097 is a gram-negative bacterium, proteins of this bacterium can be recognized from five feasible subcellular locations. These are cytoplasm, inner membrane, periplasm, outer membrane and extracellular space. Proteins functioning in cytoplasm can be used as putative drug targets while surface membrane proteins can be considered both as drug targets and vaccine candidates. To predict subcellular localization of shortlisted proteins of set4-a, PSORTb v3.0.2 (<http://www.psort.org/psortb/index.html>) [28], CELLO v.2.5[29] (<http://cello.life.nctu.edu.tw/>), ngLOC[30] (<http://genome.unmc.edu/ngLOC/index.html>) servers were used. The best score generated by these servers for a location was counted. Final location of the proteins was fixed by considering the reported location of more than two servers. Only membrane proteins were listed in set5 as potential drug targets and vaccine candidates.

Identification Novel Drug Targets by Screening and Druggability Analysis (Set6-a and Set6-b)

Membrane proteins of set5 were further analyzed to check the novelty as drug target. Proteins were screened out through database of DrugBank 5.1.0 [31] while using default parameters. DrugBank 5.1.0 is a database containing entries including 2556 approved small molecule drugs, 5125 experimental drugs, 962 approved biotech drugs and 112 nutraceuticals. Moreover 511

non-redundant protein sequences such as drug target, enzyme, carrier or transporter are linked to these drug entries. Presence of drug targets in the database indicating the same function was listed as druggable targets as set6-a, while the absence of drug entries for a protein indicated its novelty as a drug target. Since our aim is to identify novel drug targets for *B. bacilliformis* subsp. ver097, proteins showing no similarity after screened through DrugBank database, were listed in set6-b.

‘Anti-target’ Analysis of Essential, Nonhomologous and Novel Drug Targets

Drugs or therapeutic compounds designed to bind and inhibit the action of a protein present in the pathogen may dock with some crucial proteins of host serving pharmacological effects in the host. These proteins of the host are termed as ‘anti-targets’. In human, ‘anti-targets’ include ether-a-go-go related gene (hGER), P-glycoproteins (P-gp), the pregnane X receptor (PXR) and constitutive androstane receptor (CAR). Additionally some membrane receptors, e.g. dopaminergic D2, adrenergic α_{1a} , serotonergic 5-HT_{2C}, and the muscarinic M1 are included as ‘anti-targets’. Total 210 ‘anti-targets’ were reported in the literature with their accession numbers[32]. The corresponding protein sequences were fetched from NCBI Protein database and these accession numbers are provided in Supplementary file SF1. Novel drug targets were subjected to BLASTp analysis in NCBI blast program against these human ‘anti-targets’ setting an E-value <0.005, query length >30%, identity <25% as parameters. Proteins showing <25% identity were listed as set7.

Prediction of Antigenicity (Set8)

To predict the protective antigens as well as subunit vaccines, novel drug target proteins were screened through VaxiJen v2.0 server[33](<http://www.ddg->

pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) using threshold value 0.4. Protein sequences showing antigenicity >0.4 were listed in set8. This server was developed in order to allow classification of antigen based on the physiochemical properties. VaxiJen is the first server to predict of protective antigens as alignment independent.

Conservancy Analysis of Predicted Sequences with Other Strains

Conservancy pattern of the predicted sequences with other classically used strains is essential for determining the range of drug spectrum among the entire homologous bacterial community. Conservancy analysis of the predicted drug target sequences of was done by BLASTp suite of NCBI server (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>). Here, in case of performing protein-protein BLAST, all of the parameters were kept default except *Bartonella bacilliformis* in organism option.

Results and Discussion

Identification of novel drug targets, as well as vaccine candidates against *B. bacilliformis* subsp. ver097 was the prime concerned target of the study, and here, subtractive genomic analysis of the entire *B. bacilliformis* subsp. ver097 proteome was done by using different computational tools and database. A brief eyesight of the subtractive genomic approach was presented in Table 1.

Identification of Paralogous Sequences

Entire proteome of *B. bacilliformis* subsp. ver097 contains 1097 proteins. Paralogous protein sequences of the pathogen were identified through CD-hit server. Total 8 paralogous clusters were found among set0 proteins of the pathogen. The paralogous sequence above >60% similarity were removed leaving 1087 non-paralogous protein sequences. Furthermore, total

proteins containing <100 amino acids were filtered out from 1087 non-paralogous protein sequences assuming these smaller proteins unlikely represent the essentiality[34,35]. Total 997 non-duplicate proteins which contain ≥ 100 amino acids in the sequences were listed in set1.

Identification of Human non-homologous Proteins

Proteins found to be involved in common cellular systems emerged as homologous in course of time between bacteria and human[35-37]. As a result, drugs and therapeutic compounds developed to bind proteins of any pathogen should avoid unintentional blocking of host proteins. To avoid cross-reactivity with host proteins, all non-paralogous proteins of the pathogen (set1) were subjected to BLASTp against Refseq proteome of *Homo sapiens* in Ensemble genome browser 92. The result revealed that total 394 proteins were found to be homologous with the human proteins. These proteins were removed while 603 non-homologous proteins were listed in set2.

Selection of Essential Proteins

The most promising drug targets are essential proteins because most antibacterial compounds are designed to dock in essential gene products[38]. Screening through DEG 15.2 server, 114 proteins were found to be similar above the parameters. These proteins were selected and listed in set3 as essential proteins for the survival of *B. bacilliformis* subsp. ver097. Proteins encoded by essential genes unique to an organism may be considered as species-specific therapeutic targets[39].

Analysis of Metabolic pathway

KEGG server contains 95 metabolic pathways for *B. bacilliformis* subsp. ver097 and 325 pathways for human. Through manual comparison, 24 metabolic pathways were found to be pathogen-specific and are provided in Supplementary Table 1. Proteins involved in these unique pathways can be selected to as drug targets.

Non-homologous essential proteins subjected to BLASTp in KAAS server at KEGG revealed that 82 proteins among 114 were assigned both KO (KEGG Orthology) and metabolic pathways. To avoid undesirable cross-reactions with human pathways, 75 proteins among 82 were excluded, leaving only 7 proteins involved in pathogen-specific pathways. These unique pathway proteins were listed in set4-a and are provided in Table 2. Only Set4-a proteins were used to identify novel drug targets and vaccine candidates in our study.

Among 114 proteins, 27 were found not to be involved in metabolic pathways but were assigned KO (KEGG Orthology). Since these proteins had successfully passed through host non-homology test and essentiality test, in future they can be used as drug targets as well as vaccine candidates. These 27 proteins are provided in Table 3. Only 5 proteins among 114 listed as set4-b were found not to be assigned KO meaning that they are not involved in any metabolic pathway of the pathogen and the host. They are provided in Table 4.

Subcellular Location Analysis of Proteins Involved in Pathogen-Specific Pathways

It was observed for many proteins which can span multiple localization sites and hence localization is one of the most important aspects of any therapeutic target [40]. Proteins listed in set4-a were passed through PSORTb v3.0.2, CELLO v.2.5 and ngLOC servers. Counting the best score for any predicted location for a protein, the final result were selected considering two

conditions, 1) the indifferent location of protein predicted by all three servers, and we selected the locations of 6 proteins by this consideration; 2) the indifferent location predicted by any two servers and we selected the location of 1 protein by this consideration. 4 proteins were to be present in inner membrane while the rest 3 were found to be present in the cytoplasm. The results are provided in Supplementary Table 2. Excluding the cytoplasmic proteins, only inner membrane proteins were listed in set5 for further analysis and the accession numbers are provided in Table 5. More than 60% of therapeutic and drug targets are represented membrane proteins[41]. There are several advantages of targeting membrane proteins for treatment, for example, their function can easily be studied by computer-based tools and approaches rather than wet-lab processes. Computational tools can be used to predict the secondary structure since membrane proteins generally have a unique structure and show a propensity to form secondary structure[42]. In future, structure-based drug designing will be possible[43] by using ab initio modeling method, while no experimental information or 3D structure of high resolution is available[42,44].

Identification of Novel Drug Targets by Screening and druggability Analysis

Inner membrane proteins were screened through DrugBank Database to identify novelty of proteins which be used as drug targets. Two proteins showed similarity with the drug targets available at the database. They are provided in Table 6. Since our aim was to find novel drug targets we filtered these two proteins and listed the remaining two proteins in set5 which were found to be absent in the database. The broad spectrum therapeutic compounds medicated for a pathogenic strain or group of strains may transfer genes to other pathogenic strains as well as cause mutational changes. Due to overuse and misuse of antibiotics as well as the lack of designing new drugs by the pharmaceutical industry, the antibiotic resistance crisis has

emerged[41,42]. To meet the challenge of this crisis, the selected novel drug targets in this can play a vital role in developing new therapeutic compounds.

‘Anti-targets’ Analysis of Novel Drug Targets

Many drug candidates were withdrawn from markets due to showing carcinogenic effects, and that is why cross reactivity and carcinogenic checking is crucial for developing an effective drug molecule. [43-48]. Though host non-homologous proteins were removed from the the non-paralogous protein sequences of the pathogen, this analysis was performed to avoid any kind of cross-reactivity and toxic effects due to docking between the drugs administered for the pathogen and the host ‘anti-targets’. Set5 proteins were successfully screened through BLASTp with 210 reported ‘anti-targets’ present in human. No evidence of similarity was seen and hence all of set5 proteins were listed in set6 considering non-homologous to host ‘anti-targets’.

Prediction of Antigenicity (Set8)

New genomics-based approaches or alternatively known as reverse vaccinology have been demonstrated to be an effective as well as powerful in discovering vaccine candidates [49-52]. Development of safe recombinant vaccines on the basis of the small number of antigenic protein sequences is emerging as the most attractive even cost-effective solution to combat with infectious disease [53-55]. Each novel drug target of *B. bacilliformis* subsp. ver097 in set7 was screened through VaxiJen v2.0 server. The server predicted ‘Flagellar biosynthesis protein FlhA (WP_041849879.1)’ and ‘ABC transporter permease (WP_052472869.1)’ as probable antigens of *B. bacilliformis* subsp. Ver097 and the result is provided in Table 7. Development of vaccine candidates against these membrane proteins can contribute to the control and eradication of infectious diseases of human which type of studies have already been reported [56,57]. .

Conservancy Analysis of Predicted Sequences with Other Strains

Protein-protein BLASTp revealed the conservancy pattern of ‘Flagellar biosynthesis protein FlhA (WP_041849879.1)’ and ‘ABC transporter permease (WP_052472869.1)’ from *B. bacilliformis* subsp. Ver097 with other strains of *B. bacilliformis*. ‘ABC transporter permease’ of *B. bacilliformis* subsp. Ver097 showed minimal conservancy with other *Bartonella bacilliformis* strains while ‘Flagellar biosynthesis protein FlhA’ of *B. bacilliformis* subsp. Ver097 showed above 99% conservancy with homologous protein of other classically used *Bartonella bacilliformis* strains (Table 8). Predicted protein always should be conserved in a wide range of bacterial strains for ensuring an effective drug target against a specific bacterial pathogen [58]. Here, ‘Flagellar biosynthesis protein FlhA’ displayed a higher conservancy pattern with homologous proteins of other *Bartonella bacilliformis* strains from different geographical locations. So, ‘Flagellar biosynthesis protein FlhA’ could be a potent broad spectrum drug target for *Bartonella bacilliformis*.

Conclusion

Subtractive genomic approach was successfully used throughout the study, and revealed that one protein of *Bartonella bacilliformis subsp. ver097* can be used as novel drug target and vaccine candidate. Inhibition of this inner membrane protein (Flagellar biosynthesis protein FlhA) will be helpful to combat against the infectious diseases caused by *Bartonella bacilliformis* pathogen since this protein is involved in pathogen-specific ‘flagellar assembly’ metabolic pathway. Furthermore chances of cross-reactivity between drugs and human proteins have been reduced since they showed no similarity against the human proteome and ‘anti-targets’. Hence the

development of new vaccines and novel drugs or therapeutic compounds could be a promising
footsteps to the eradication of diseases progression by *Bartonella bacilliformis*.

Data Availability

Not Applicable

Conflict of Interest

There is no conflict of interest regarding the publication of this paper

Acknowledgment

We are highly thankful to Md. Ismail Hosen (Department of Biochemistry and Molecular
Biology, University of Dhaka, Dhaka 1000, Bangladesh) for his cordial advice and suggestions.

Funding source

There was no funding source for this project.

Supplementary Materials

Supplementary file S1: 'Anti-targets' protein sequences (210) fetched from NCBI Protein database that can cause undesirable side effects during drug interactions.

Supplementary Tables

Supplementary Table 1: List of unique metabolic pathways of the pathogen

Supplementary Table 2: Results of subcellular location of 7 non-homologous, essential proteins only involved in the unique metabolic pathways of the pathogen

References

1. Silva-Caso W, Pons MJ, Ruiz J, del Valle-Mendoza J. Antibiotic resistance in *Bartonellabacilliformis* clinical isolates from an endemic area of Peru. Journal of global antimicrobial resistance. 2015 Sep 1;3(3):222-3. (<https://doi.org/10.1016/j.jgar.2015.05.004>)
2. Ihler GM. *Bartonellabacilliformis*: dangerous pathogen slowly emerging from deep background. FEMS Microbiology Letters. 1996 Oct 1;144(1):1-1. (<https://doi.org/10.1111/j.1574-6968.1996.tb08501.x>)
3. Maguiña C, Gotuzzo E. Bartonellosis: new and old. Infectious disease clinics of North America. 2000 Mar 1;14(1):1-22. ([https://doi.org/10.1016/s0891-5520\(05\)70215-4](https://doi.org/10.1016/s0891-5520(05)70215-4))
4. Alexander B. A review of bartonellosis in Ecuador and Colombia. The American journal of tropical medicine and hygiene. 1995 Apr 1;52(4):354-9. (<https://doi.org/10.4269/ajtmh.1995.52.354>)

5. Henriquez-Camacho C, Ventosilla P, Minnick MF, Ruiz J, Maguina C. Proteins of Bartonellabacilliformis: Candidates for Vaccine Development. International Journal of Peptides. 2015;2015:1. (<https://doi.org/10.1155/2015/702784>)
6. Maguiña C, Guerra H, Ventosilla P. Bartonellosis. Clinics in dermatology. 2009 May 1;27(3):271-80. (<https://doi.org/10.1016/j.clindermatol.2008.10.006>)
7. Henriquez C, Hinojosa JC, Ventosilla P, Infante B, Merello J, Mallqui V, Verastegui M, Maguina C. Report of an unusual case of persistent bacteremia by Bartonellabacilliformis in a splenectomized patient. The American journal of tropical medicine and hygiene. 2004 Jul 1;71(1):53-5. (<https://doi.org/10.4269/ajtmh.2004.71.53>)
8. Huarcaya E, Maguiña C, Torres R, Rupay J, Fuentes L. Bartonellosis (Carrion's Disease) in the pediatric population of Peru: an overview and update. Brazilian Journal of Infectious Diseases. 2004 Oct;8(5):331-9. (<https://doi.org/10.1590/s1413-86702004000500001>)
9. Gray GC, Johnson AA, Thornton SA, Smith WA, Knobloch J, Kelley PW, Escudero LO, Huayda MA, Wignall FS. An epidemic of Oroya fever in the Peruvian Andes. The American journal of tropical medicine and hygiene. 1990 Mar 1;42(3):215-21. (<https://doi.org/10.4269/ajtmh.1990.42.215>)
10. Maguiña Vargas C, Ugarte-Gil C, Breña Chávez P, Ordaya Espinoza E, VentosillaLópez P, HuarcayaCastilla E, Henriquez Camacho C. Actualización de la enfermedad de Carrión. Revistamédicaherediana. 2008 Jan;19(1):36-41. (<https://doi.org/10.20453/rmh.v19i1.975>)
11. Uddin R, Saeed K. Identification and characterization of potential drug targets by subtractive genome analyses of methicillin resistant Staphylococcus aureus.

- Computational biology and chemistry. 2014 Feb 1;48:55-63.
(<https://doi.org/10.1016/j.compbiolchem.2013.11.005>)
12. Sobraquès M, Maurin M, Birtles RJ, Raoult D. In vitro susceptibilities of four Bartonellabacilliformis strains to 30 antibiotic compounds. Antimicrobial agents and chemotherapy. 1999 Aug 1;43(8):2090-2. (<https://doi.org/10.1128/aac.43.8.2090>)
13. Gomes C, Martínez-Puchol S, Ruiz-Roldán L, Pons MJ, del Valle Mendoza J, Ruiz J. Development and characterisation of highly antibiotic resistant Bartonellabacilliformis mutants. Scientific reports. 2016 Sep 26;6:33584. (<https://doi.org/10.1038/srep33584>)
14. Biswas S, Raoult D, Rolain JM. Molecular mechanisms of resistance to antibiotics in Bartonellabacilliformis. Journal of antimicrobial chemotherapy. 2007 Apr 21;59(6):1065-70. (<https://doi.org/10.1093/jac/dkm105>)
15. Uddin R, Jamil F. Prioritization of potential drug targets against P. aeruginosa by core proteomic analysis using computational subtractive genomics and Protein-Protein interaction network. Computational biology and chemistry. 2018 Jun 1;74:115-22. (<https://doi.org/10.1016/j.compbiolchem.2018.02.017>)
16. Sarangi AN, Aggarwal R, Rahman Q, Trivedi N. Subtractive genomics approach for in silico identification and characterization of novel drug targets in Neisseria Meningitides Serogroup B. J ComputSciSyst Biol. 2009;2(5):255-8. (<https://doi.org/10.4172/jcsb.1000038>)
17. Hossain T, Kamruzzaman M, Choudhury TZ, Mahmood HN, Nabi AH, Hosen M. Application of the subtractive genomics and molecular docking analysis for the identification of novel putative drug targets against Salmonella enterica subsp.

- entericaserovar Poona. BioMed Research International. 2017;2017.
(<https://doi.org/10.1155/2017/3783714>)
18. Mondal SI, Ferdous S, Jewel NA, Akter A, Mahmud Z, Islam MM, Afrin T, Karim N. Identification of potential drug targets by subtractive genome analysis of *Escherichia coli* O157: H7: an in silico approach. *Advances and applications in bioinformatics and chemistry: AABC*. 2015;8:49. (<https://doi.org/10.2147/aabc.s88522>)
19. Paul S, Minnick MF, Chattopadhyay S. Mutation-driven divergence and convergence indicate adaptive evolution of the intracellular human-restricted pathogen, *Bartonellabacilliformis*. *PLoS neglected tropical diseases*. 2016 May 11;10(5):e0004712. (<https://doi.org/10.1371/journal.pntd.0004712>)
20. Ruiz J, Pons MJ. Revisiting *Bartonellabacilliformis* MLST. *Infection, Genetics and Evolution*. 2018 Jun 2. (<https://doi.org/10.1016/j.meegid.2018.05.032>)
21. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*. 2010 Jan 6;26(5):680-2. (<https://doi.org/10.1093/bioinformatics/btq003>)
22. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*. 2010 Jan 6;26(5):680-2. (<https://doi.org/10.1093/bioinformatics/btq003>)
23. Luo H, Lin Y, Gao F, Zhang CT, Zhang R. DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic acids research*. 2013 Nov 15;42(D1):D574-80. (<https://doi.org/10.1093/nar/gkt1131>)

24. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*. 2000 Jan 1;28(1):27-30. (<https://doi.org/10.1093/nar/28.1.27>)
25. Paduano F, Forbes AG. Extended LineSets: a visualization technique for the interactive inspection of biological pathways. In *BMC proceedings* 2015 Dec (Vol. 9, No. 6, p. S4). BioMed Central. (<https://doi.org/10.1186/1753-6561-9-s6-s4>)
26. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic acids research*. 2007 Jul 1;35(suppl_2):W182-5. (<https://doi.org/10.1093/nar/gkm321>)
27. Damte D, Suh JW, Lee SJ, Yohannes SB, Hossain MA, Park SC. Putative drug and vaccine target protein identification using comparative genomic analysis of KEGG annotated metabolic pathways of *Mycoplasma hyopneumoniae*. *Genomics*. 2013 Jul 31;102(1):47-56. (<https://doi.org/10.1016/j.ygeno.2013.04.011>)
28. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*. 2010 May 13;26(13):1608-15. (<https://doi.org/10.1093/bioinformatics/btq249>)
29. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. *Proteins: Structure, Function, and Bioinformatics*. 2006 Aug 15;64(3):643-51. (<https://doi.org/10.1002/prot.21018>)
30. King BR, Guda C. ngLOC: an n-gram-based Bayesian method for estimating the subcellular proteomes of eukaryotes. *Genome biology*. 2007 May;8(5):R68. (<https://doi.org/10.1186/gb-2007-8-5-r68>)

31. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research*. 2017 Nov 8;46(D1):D1074-82. (<https://doi.org/10.1093/nar/gkx1037>)
32. Shanmugham B, Pan A. Identification and characterization of potential therapeutic candidates in emerging human pathogen *Mycobacterium abscessus*: a novel hierarchical in silico approach. *PloS one*. 2013 Mar 19;8(3):e59126. (<https://doi.org/10.1371/journal.pone.0059126>)
33. Doytchinova IA, Flower DR. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. *Vaccine*. 2007 Jan 15;25(5):856-66. (<https://doi.org/10.1016/j.vaccine.2006.09.032>)
34. Kumar GS, Sarita S, Kumar GM, Pant KK, Seth PK. Definition of potential targets in *Mycoplasma Pneumoniae* through subtractive genome analysis. *J. Antivir. Antiretrovir*. 2010;2:038-41. (<https://doi.org/10.4172/jaa.1000020>)
35. Haag NL, Velk KK, Wu C. In silico identification of drug targets in methicillin/multidrug-resistant *Staphylococcus aureus*. *Int. J. Adv. Life Sci*. 2012;4(1-2):21-32. (<https://doi.org/10.1101/034306>)
36. Hediger MA, Turk E, Wright EM. Homology of the human intestinal Na⁺/glucose and *Escherichia coli* Na⁺/proline cotransporters. *Proceedings of the National Academy of Sciences*. 1989 Aug 1;86(15):5748-52. (<https://doi.org/10.1073/pnas.86.15.5748>)
37. Swango KL, Hymes J, Brown P, Wolf B. Amino acid homologies between human biotinidase and bacterial aliphatic amidases: putative identification of the active site of

- biotinidase. *Molecular genetics and metabolism*. 2000 Feb 29;69(2):111-5.
(<https://doi.org/10.1006/mgme.2000.2959>)
38. Zhang R, Ou HY, Zhang CT. DEG: a database of essential genes. *Nucleic acids research*. 2004 Jan 1;32(suppl_1):D271-2. (<https://doi.org/10.1093/nar/gkh024>)
39. Judson N, Mekalanos JJ. TnAraOut, a transposon-based approach to identify and characterize essential bacterial genes. *Nature biotechnology*. 2000 Jul;18(7):740.
(<https://doi.org/10.1038/77305>)
40. Golkar Z, Bagasra O, Pace DG. Bacteriophage therapy: a potential solution for the antibiotic resistance crisis. *The Journal of Infection in Developing Countries*. 2014 Feb 13;8(02):129-36. (<https://doi.org/10.3855/jidc.3573>)
41. Michael CA, Dominey-Howes D, Labbate M. The antimicrobial resistance crisis: causes, consequences, and management. *Frontiers in public health*. 2014 Sep 16;2:145.
(<https://doi.org/10.3389/fpubh.2014.00145>)
42. Bartlett JG, Gilbert DN, Spellberg B. Seven ways to preserve the miracle of antibiotics. *Clinical Infectious Diseases*. 2013 Feb 12;56(10):1445-50.
(<https://doi.org/10.1093/cid/cit070>)
43. Hossain M, Chowdhury DU, Farhana J, Akbar MT, Chakraborty A, Islam S, Mannan A. Identification of potential targets in *Staphylococcus aureus* N315 using computer aided protein data analysis. *Bioinformation*. 2013;9(4):187.
(<https://doi.org/10.6026/97320630009187>)
44. Fung M, Thornton A, Mybeck K, Wu JH, Hornbuckle K, Muniz E. Evaluation of the characteristics of safety withdrawal of prescription drugs from worldwide pharmaceutical

markets-1960 to 1999. *Drug Information Journal*. 2001 Jan;35(1):293-317.

(<https://doi.org/10.1177/009286150103500134>)

45. Fenner H. Evaluation of the efficacy and safety of NSAIDs. A new methodological

approach. *Scandinavian Journal of Rheumatology*. 1989 Jan 1;18(sup80):32-9.

(<https://doi.org/10.3109/03009748909103710>)

46. Bakke OM, Manocchia M, de Abajo F, Kaitin KI, Lasagna L. Drug safety

discontinuations in the United Kingdom, the United States, and Spain from 1974 through

1993: a regulatory perspective. *Clinical Pharmacology & Therapeutics*. 1995 Jul

1;58(1):108-17. ([https://doi.org/10.1016/0009-9236\(95\)90078-0](https://doi.org/10.1016/0009-9236(95)90078-0))

47. Moses PL, Schroeder B, Alkhatib O, Ferrentino N, Suppan T, Lidofsky SD. Severe

hepatotoxicity associated with bromfenac sodium. *The American journal of*

gastroenterology. 1999 May;94(5):1393. ([https://doi.org/10.1111/j.1572-](https://doi.org/10.1111/j.1572-0241.1999.01093.x)

[0241.1999.01093.x](https://doi.org/10.1111/j.1572-0241.1999.01093.x))

48. Galan MV, Potts JA, Silverman AL, Gordon SC. The burden of acute nonfulminant drug-

induced hepatitis in a United States tertiary referral center. *Journal of clinical*

gastroenterology. 2005 Jan 1;39(1):64-7. ([https://doi.org/10.1016/s0016-5085\(03\)83786-](https://doi.org/10.1016/s0016-5085(03)83786-1)

1)

49. Rappuoli R. Reverse vaccinology. *Current opinion in microbiology*. 2000 Oct 1;3(5):445-

50. ([https://doi.org/10.1016/s1369-5274\(00\)00119-3](https://doi.org/10.1016/s1369-5274(00)00119-3))

50. Maignani V, Rappuoli R, Pizza M. Reverse vaccinology: a genome-based approach for

vaccine development. *Expert opinion on biological therapy*. 2002 Dec 1;2(8):895-905.

(<https://doi.org/10.1517/14712598.2.8.895>)

- 517 51. Doro F, Liberatori S, Rodríguez-Ortega MJ, Rinaudo CD, Rosini R, Mora M, Scarselli
518 M, Altindis E, D'Aurizio R, Stella M, Margarit I. Surfome analysis as a fast track to
519 vaccine discovery identification of a novel protective antigen for group B streptococcus
520 hypervirulent strain COH1. *Molecular & cellular proteomics*. 2009 Jul 1;8(7):1728-37.
521 (<https://doi.org/10.1074/mcp.m800486-mcp200>)
- 522 52. Hasan M, Azim KF, Begum A, Khan NA, Shammi TS, Imran AS, Chowdhury IM, Urme
523 SR (2019a) Vaccinomics strategy for developing a unique multi-epitope monovalent
524 vaccine against Marburg marburgvirus. *Infection, Genetics and Evolution* 70:140-157
525 (<https://doi.org/10.1016/j.meegid.2019.03.003>)
- 526 53. Bogush ML, Velikodvorskaya TV, Lebedev YB, Nikolaev LG, Lukyanov SA, Fradkov
527 AF, Pliyev BK, Boichenko MN, Usatova GN, Vorobiev AA, Andersen GL. Identification
528 and localization of differences between *Escherichia coli* and *Salmonella typhimurium*
529 genomes by suppressive subtractive hybridization. *Molecular and General Genetics*
530 MGG. 1999 Dec 1;262(4-5):721-9. (<https://doi.org/10.1007/s004380051134>)
- 531 54. Hasan M, Hakim A, Iqbal A, Bhuiyan FR, Begum MK, Sharmin S and Abir RA (2015b)
532 Computational study and homology modeling of phenol hydroxylase: key enzyme for
533 phenol degradation. *Int J ComputBioinfoIn Silico Model* 4(4):691-698
- 534 55. Joy ZF, Purkaystha A, Das NK, Al-Hakim, Chakrabarty S and Hasan M (2019)
535 Screening for Alternative Sources of L-Asparaginase Used in Acute Lymphoblastic
536 Leukaemia (ALL) Treatment: An In Silico Approach. *Bioinformatics and Proteomics Open*
537 *Access Journal* 3(1): 1-9
- 538 56. Azim KF, Hasan M, Hossain MN, Somana SR, Hoque SF, Bappy MN, Chowdhury AT,
539 Lasker T. Immunoinformatics approaches for designing a novel multi epitope peptide
540 vaccine against human norovirus (Norwalk virus). *Infection, Genetics and Evolution*.
541 2019 Jun 21:103936. (<https://doi.org/10.1016/j.meegid.2019.103936>)

57. Hasan M, Ghosh P, Azim K, Mukta S, Abir RA, Nahar J, Khan MM. Reverse vaccinology approach to design a novel multi-epitope subunit vaccine against avian influenza A (H7N9) virus. *Microbial pathogenesis*. 2019b;130:19-37. (<https://doi.org/10.1016/j.micpath.2019.02.023>)
58. Gupta A, Pal SK, Pandey D, Fakir NA, Rathod S, Sinha D, SivaKumar S, Sinha P, Periera M, Balgam S, Sekar G. PknB remains an essential and a conserved target for drug development in susceptible and MDR strains of M. Tuberculosis. *Annals of clinical microbiology and antimicrobials*. 2017 Dec;16(1):56. (<https://doi.org/10.1186/s12941-017-0234-9>).

Tables

Table 1: Subtractive genomic analysis scheme

<i>S. No.</i>	<i>Steps</i>	<i>B. bacilliformissubsp. ver097</i>
1	<i>Total number of proteins</i>	1097
2	Removed non-paralogous (>60% identical) and smaller proteins (>100 amino acids) in CD-Hit and	100
3	Number of proteins nonhomologous to H. sapiens using BLASTp (E value 10^{-3})	603
4	Essential proteins in DEG 15.2 server (E value $\leq 10^{-100}$, Bit score >100)	114
5	Essential Proteins involved only in unique metabolic pathways (KAAS at KEGG)	7

5	Proteins assigned KO (KEGG Orthology) but not in any pathway	27
6	Essential membrane proteins using PSORTb, CELLO, ngLOC,	4
7	Essential proteins found to be novel in DrugBank 5.1.0 (using default parameters)	2
8	Novel drug target proteins non-homologous to ‘anti-targets’ using BLASTp (E value <0.005, Identity <25%, Query length >30%)	2
9	Proteins showing antigenicity using VaxiJen v2.0 (Threshold value >0.4)	2
10	Highly conserved protein	1

Table 2: Metabolic proteins involved only in pathogen-specific pathways

<i>S. No.</i>	<i>KO Assignment</i>	<i>Protein ID</i>	<i>Protein name</i>	<i>Pathways</i>
1	K14981	WP_041848966.1	DNA-binding response regulator	Two-component system
2	K14980	WP_041849190.1	histidine kinase	Two-component system
3	K13599	WP_041849407.1	sigma-54-dependent Fis family transcriptional regulator	Two-component system

4	K01928	WP_041849726.1	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase	Peptidoglycan biosynthesis Lysine biosynthesis
5	K03587	WP_041849727.1	penicillin-binding protein 2	Peptidoglycan biosynthesis beta-lactam resistance
6	K02400	WP_041849879.1	flagellar biosynthesis protein FlhA	Flagellar assembly
7	K02034	WP_052472869.1	ABC transporter permease	Quorum sensing

Table 3: Proteins assigned KO but not assigned any pathway

<i>KO assignment</i>	<i>Accession No.</i>	<i>Protein name</i>
K03296	WP_041848830.1	AcrB/AcrD/AcrF family protein
K03466	WP_041848841.1	DNA translocase FtsK
K03089	WP_041848844.1	RNA polymerase sigma factor RpoH
K19689	WP_041848873.1	aminopeptidase
K02600	WP_041848898.1	transcription termination/

		antitermination protein NusA
K07391	WP_041848977.1	ATP-binding protein
K03496	WP_041849016.1	ParA family protein
K03497	WP_041849017.1	chromosome partitioning protein ParB
K03641	WP_041849127.1	translocation protein TolB
K03797	WP_041849219.1	S41 family peptidase
K03568	WP_041849289.1	metalloprotease TldD
K02483	WP_041849309.1	DNA-binding response regulator
K02198	WP_041849313.1	hemelyaseNrfEFG subunit NrfE
K04485	WP_041849362.1	DNA repair protein Rada
K07082	WP_041849372.1	endolytictransglycosylaseMltG
K03499	WP_041849408.1	Trk system potassium transporter TrkA
K07018	WP_041849624.1	alpha/beta hydrolase
K09014	WP_041849626.1	Fe-S cluster assembly protein SufB
K09456	WP_041849647.1	acyl-CoA dehydrogenase
K03761	WP_041849702.1	MFS transporter
K03086	WP_041849735.1	RNA polymerase sigma factor RpoD
K03498	WP_041849746.1	metal ABC transporter ATPase
K03924	WP_041849766.1	MoxR family ATPase
K09015	WP_052472887.1	Fe-S cluster assembly protein SufD
K03820	WP_080772740.1	apolipoprotein N-acyltransferase
K04744	WP_080772760.1	LPS-assembly protein LptD
K00528	WP_041849825.1	ferredoxin--NADP reductase

567 **Table 4:** Proteins having no KO assignment

<i>S. No.</i>	<i>Accession No.</i>	<i>Protein Name</i>
1	WP_041849180.1	M23 family peptidase
2	WP_041849317.1	ssensor histidine kinase
3	WP_041849617.1	ribonuclease E/G
4	WP_041849690.1	TerC family protein
5	WP_041849779.1	DegT/DnrJ/EryC1/StrS

aminotransferase family protein

Table 5: List of membrane proteins

<i>Accession No.</i>	<i>Protein</i>	<i>Location</i>
WP_041849190.1	histidine kinase	Inner Membrane
WP_041849727.1	penicillin-binding protein 2	Inner Membrane
WP_041849879.1	flagellar biosynthesis protein FlhA	Inner Membrane
WP_052472869.1	ABC transporter permease	Inner Membrane

Table 6: List of druggable proteins

<i>Accession No.</i>	<i>Protein Name</i>	<i>DrugBank ID</i>	<i>Drug Name</i>	<i>Category</i>	<i>Organism Name</i>
WP_041849190.1	histidine kinase	DB02731	Ethylmercurit hiosalicylic acid	Experimental	Thermotogamaritima (strain ATCC 43589 / MSB8 / DSM 3109 / JCM 10099)
WP_041849727.1	penicillin-binding protein 2	DB00303	Ertapenem	Approved, Investigational	Haemophilusinfluenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd)
		DB00671	Cefixime	Approved, Investigational	Haemophilusinfluenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd)

Table 7: Novel drug target proteins showing antigenicity

<i>Accession No.</i>	<i>Protein Name</i>	<i>VaxiJenScore</i>
WP_041849879.1	flagellar biosynthesis protein FlhA	0.6098
WP_052472869.1	ABC transporter permease	0.4516

576 **Table 8:** Conservancy analysis Flagellar biosynthesis protein FlhA; (BLASTp results of
577 WP_041849897.1 against other *Bartonellabacilliformis* strains)

<i>S. No.</i>	<i>Sequence id</i>	<i>Organism</i>	<i>Identity</i>
1	WP_005767817.1	<i>Bartonellabacilliformis</i> KC583	99%
2	EKS43057.1	<i>Bartonellabacilliformis</i> INS	99%
3	EYS88603.1	<i>Bartonellabacilliformis</i> San Pedro600-02	99%
4	EYS94419.1	<i>Bartonellabacilliformis</i> Peru-18	99%
5	KEG15906.1	<i>Bartonellabacilliformis</i> CUSCO5	99%
6	KEG19809.1	<i>Bartonellabacilliformis</i> Peru38	99%
7	AMG86163.1	<i>Bartonellabacilliformis</i>	99%
8	KZM37516.1	<i>Bartonellabacilliformis</i>	99%
9	KZN21555.1	<i>Bartonellabacilliformis</i>	99%
10	KEG22213.1	<i>Bartonellabacilliformis</i> Ver075	99%
11	KEG16163.1	<i>Bartonellabacilliformis</i> Cond044	99%
12	EYS91027.1	<i>Bartonellabacilliformis</i> str. Heidi Mejia	99%
13	KEG18152.1	<i>Bartonellabacilliformis</i> Hosp800-02	99%
14	KEG21780.1	<i>Bartonellabacilliformis</i> VAB9028	99%
15	KEG23155.1	<i>Bartonellabacilliformis</i> CAR600-02	99%

578 ***Figures Caption:***

579 **Figure 1:** An outline of Subtractive Genomics approach to *Bartonellabacilliformis* subsp.Ver097
580 proteomes for identification of novel drug candidates.

581

Proteome Retrieval

- *Legionella pneumophila subsp. pneumophila* (strain Philadelphia 1 / ATCC 33152 / DSM 7513)
- Proteome retrieved from UniProt (UniProt ID UP000000609) containing 2930 proteins

Resource Used

- NCBI
- CD-Hit suite
- BLASTp
- DEG Server
- KEGG
- KAAS
- PSORTb, CELLO, ngLOC
- DrugBank