Genomics 102 (2013) 47–56

Contents lists available at SciVerse ScienceDirect

Genomics

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

Putative drug and vaccine target protein identification using comparative genomic analysis of KEGG annotated metabolic pathways of *Mycoplasma hyopneumoniae*

Dereje Damte [a], Joo-Won Suh [b,c], Seung-Jin Lee [a], Sileshi Belew Yohannes [a], Md. Akil Hossain [a], Seung-Chun Park [a,*]

[a] Laboratory of Veterinary Pharmacokinetics & Pharmacodynamics, College of Veterinary Medicine, Kyungpook National University, Daegu 702-701, Republic of Korea
[b] Center for Nutraceutical and Pharmaceutical Materials, Science Campus, Myongji University, San 38-2, Namdong, Cheoin-Gu, Yongin, Gyeonggi 449-728, Republic of Korea
[c] Division of Bioscience and Bioinformatics, Science Campus, Myongji University, San 38-2, Namdong, Cheoin-Gu, Yongin, Gyeonggi 449-728, Republic of Korea

ARTICLE INFO

Article history:
Received 12 February 2013
Accepted 18 April 2013
Available online 26 April 2013

Keywords:
Mycoplasma hyopneumoniae
KEGG metabolic pathways
Therapeutic targets
Vaccine candidate

ABSTRACT

In the present study, a computational comparative and subtractive genomic/proteomic analysis aimed at the identification of putative therapeutic target and vaccine candidate proteins from Kyoto Encyclopedia of Genes and Genomes (KEGG) annotated metabolic pathways of *Mycoplasma hyopneumoniae* was performed for drug design and vaccine production pipelines against *M.hyopneumoniae*. The employed comparative genomic and metabolic pathway analysis with a predefined computational systemic workflow extracted a total of 41 annotated metabolic pathways from KEGG among which five were unique to *M. hyopneumoniae*. A total of 234 proteins were identified to be involved in these metabolic pathways. Although 125 non homologous and predicted essential proteins were found from the total that could serve as potential drug targets and vaccine candidates, additional prioritizing parameters characterize 21 proteins as vaccine candidate while druggability of each of the identified proteins evaluated by the DrugBank database prioritized 42 proteins suitable for drug targets.

1. Introduction

Mycoplasma hyopneumoniae, the causative agent of enzootic pneumonia, plays an important role in the swine industry worldwide causing enormous economic loss affecting up to 80% of pigs around the world [14]. The control of enzootic pneumonia by antimicrobial medication and/or vaccination under field condition has been reportedly inadequate, requiring additional optimized through herd management practices and housing conditions [32,42]. This strongly indicates the need to search new therapeutic targets and vaccine candidates that would offer better protection, particularly optimizing both medication and vaccination in preventing adhesion and colonization of *M. hyopneumoniae* to the respiratory tract and minimizing the high economic loss incurred.

The arrival of the post-genomic era has brought with it the possibility of genome-wide application of a rational new drug target and vaccine candidate selection methodology such as comparative and reductive genomics using computational approaches with integrated data from genomics, proteomics, and metabolomics [7,28].

In recent years the number of whole genome sequenced microbial species has increased rapidly with already more than 2000 prokaryotic whole genome sequences reported. Among those reported, several *Mycoplasma* species have had their whole genome sequenced including four strains of the swine pathogen *M. hyopneumoniae* [29,34,41]. Taking advantage of the whole genome sequence and the need for new therapeutic target and vaccine candidate of *M. hyopneumoniae*, we report here the first computational comparative genomic/proteomic analysis aimed at the identification of putative therapeutic targets and vaccine candidate proteins of *M. hyopneumoniae*.

2. Results

2.1. Identification of metabolic pathways in M. hyopneumoniae

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a database resource that integrates genomic, chemical, and systemic functional information. In particular, gene catalogs in the completely sequenced genomes are linked to higher-level systemic functions of the cell, the organism, and the ecosystem. As a result, KEGG is widely used as a reference knowledge base for integration and interpretation of large-scale datasets generated by genome sequencing and other high-throughput experimental technologies. Currently, the KEGG database incorporates 41 different pathways for the complete genome sequenced *M. hyopneumoniae* strain 232, 7448 and J [29,34,41] and 40 for strain 168. In addition, a total of 257 for *M. hyopneumoniae* swine host (*Sus scrofa*; TAX:9823) and 261 for human (*Homo sapiens*; TAX:9606) pathways is contained in the database. The polycyclic aromatic hydrocarbon degradation (00624) pathway was not present in strain 168. Strain 232, 7448 and J of *M. hyopneumoniae* had been reported to share about 663 (95%) of their genes which was

* Corresponding author at: Department of Pharmacology, College of Veterinary Medicine, Kyungpook National University, Daegu 702-701, Republic of Korea.
E-mail address: parksch@knu.ac.kr (S.-C. Park).

0888-7543/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2013.04.011

extended as representative of the core genome of *M. hyopneumoniae* [30]. However, this genome size was not inclusive of strain168. Hence, the reference *M. hyopneumoniae* 232 strain with a total of 691 gene products was selected as representative for further analysis. Accordingly, identification of the unique and common pathways of the pathogen with the reference strain *M. hyopneumoniae* 232 was conducted manually with the host pathways according to our work flow chart (Fig. 1). Five pathways were identified as unique for the pathogen with the remaining 36 pathways shared by both pathogen and host (Table 1). The comparison with the human metabolic pathways revealed no difference and all the shared pathways with the pathogen host were also shared by the human. Furthermore, the number of proteins involved in each pathway was identified (Table 1).

2.2. Non homologous and non homologous essential proteins

A total of 234 proteins were identified involved in both unique and common pathways of the pathogen (Table S2). As a protein was found to be involved in more than one pathway, among the 234 proteins only two were found to be involved only in unique pathways of the pathogens while the rest were shared either by both or involved only in the common pathways. The comparisons of these protein sequences against the swine host proteome using the special feature of NCBI BLASTP revealed a total of 191 proteins with no hit against the host proteome and these proteins were identified as non homologous proteins of the pathogen (Table S3). Both of the two proteins from the unique pathways were found non homologous with no hits to the host proteome. All the non homologous proteins identified were further filtered by homology searching against 9129 essential genes from 22 bacteria in DEG (Table S1). This comparison performed by using prokaryotic BLASTP from DEG detected 125 essential proteins of *M. hyopneumoniae* by similarities, which were also non-homologous to swine host proteins (Table S4). The identified essential proteins had a hit ranging from 1–20 of the 22 bacteria essential proteins in the DEG. The highest hit of 107 was observed in *M. pulmonis* UAE CTIP essential proteins. The total hits by each of the 22 bacteria in the DEG against all 125 identified essential proteins above the threshold score are presented in Fig. 2. One of the two proteins was involved only in the unique pathways mapped to essential proteins of the DEG database. However, all the non-homologous essential proteins represent an attractive dataset that could be exploited for future drug design and vaccine production against *M. hyopneumoniae*.

2.3. Prioritization of target proteins

Although all the 125 identified non homologous proteins of the pathogen represent potential targets for therapeutic and vaccine candidates, prioritization and more filtration of the identified proteins could help to minimize the time, labor and resources for developing the therapeutic agent and optimize the success of getting the best drug and/or vaccine against the pathogen. Hence, additional parameters that determine the suitability of a drug and vaccine target were used to characterize the identified target proteins. The characterization of subcellular localization by CELLO server identified 21 membrane proteins. From determined molecular weight in Swiss-Prot database we identified only 4 proteins with molecular mass >110 while the TMHMM identified 18 transmemebrane proteins. We found no single entry for experimental confirmed 3D structure in

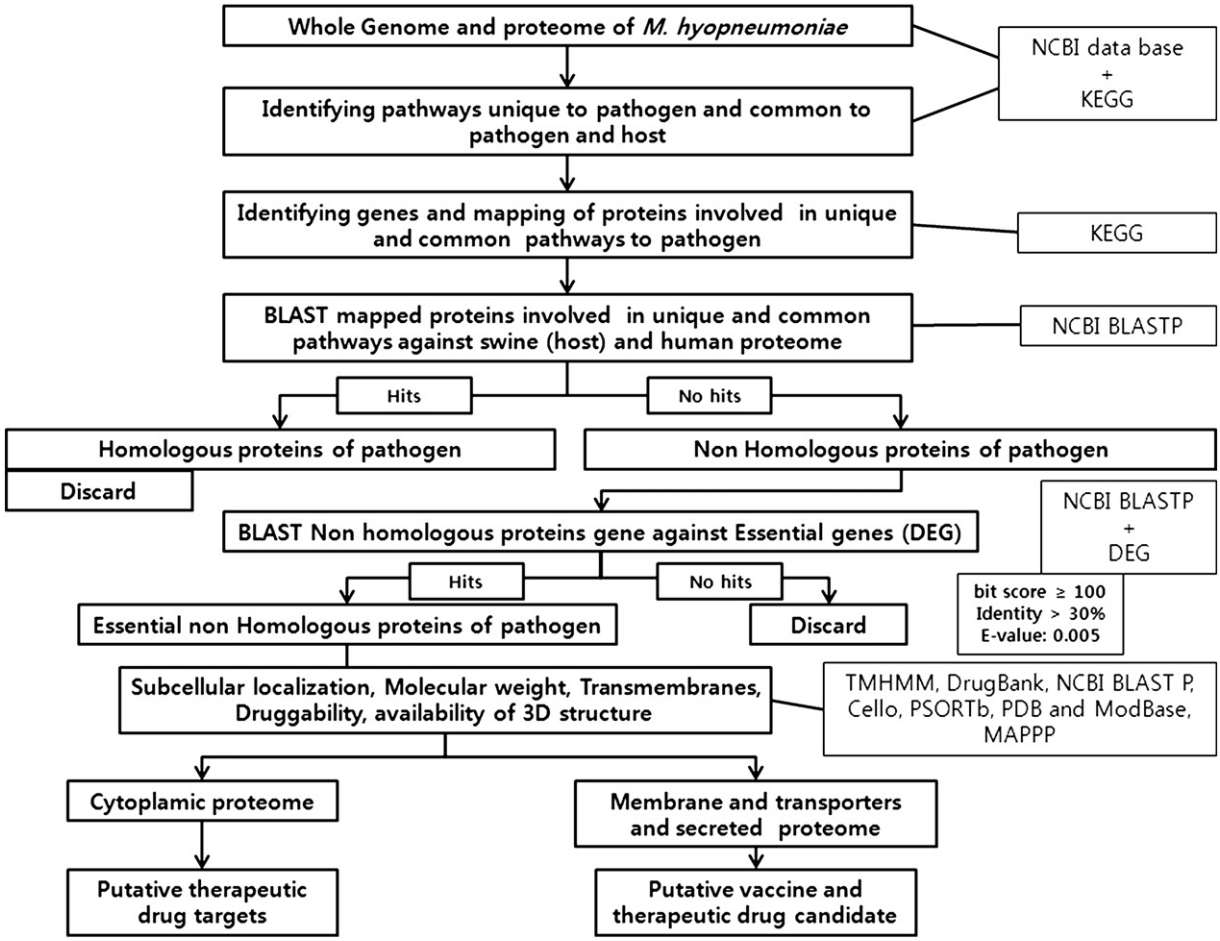

Fig. 1. Illustration of predefined comparative and subtractive genomics systemic workflow. Steps involved in computational comparative genomics based target identification in *M. hyopneumoniae.*

Table 1
Unique and common metabolic pathways of *M. hyopneumoniae* with reference to its swine host and human.

No.	Unique pathways (KEGG)	KEGG pathway ID	Total proteins
1	Streptomycin biosynthesis	00521	3
2	Polycyclic aromatic hydrocarbon degradation	00624	1
3	Methane metabolism	00680	8
4	Phosphotransferase system (PTS)	02060	12
5	Bacterial secretion system	03070	8
No.	Common pathways (KEGG)	KEGG pathway ID	Total proteins
1	Glycolysis/gluconeogenesis	00010	19
2	Citrate cycle (TCA cycle)	00020	5
3	Pentose phosphate pathway	00030	11
4	Fructose and mannose metabolism	00051	9
5	Galactose metabolism	00052	2
6	Ascorbate and aldarate metabolism	00053	9
7	Oxidative phosphorylation	00190	11
8	Purine metabolism	00230	21
9	Pyrimidine metabolism	00240	23
10	Glycine, serine and threonine metabolism	00260	4
11	Cysteine and methionine metabolism	00270	5
12	Valine, leucine and isoleucine degradation	00280	3
13	Taurine and hypotaurine metabolism	00430	2
14	Selenocompound metabolism	00450	2
15	Glutathione metabolism	00480	1
16	Starch and sucrose metabolism	00500	3
17	Amino sugar and nucleotide sugar metabolism	00520	7
18	Glycerolipid metabolism	00561	3
19	Inositol phosphate metabolism	00562	8
20	Glycerophospholipid metabolism	00564	6
21	Pyruvate metabolism	00620	10
22	Propanoate metabolism	00640	5
23	Riboflavin metabolism	00740	2
24	Nicotinate and nicotinamide metabolism	00760	5
25	Lipoic acid metabolism	00785	2
26	Aminoacyl-tRNA biosynthesis	00970	23
27	ABC transporters	02010	21
28	Ribosome	03010	47
29	RNA degradation	03018	4
30	RNA polymerase	03020	3
31	DNA replication	03030	12
32	Protein export	03060	10
33	Base excision repair	03410	5
34	Nucleotide excision repair	03420	6
35	Mismatch repair	03430	10
36	Homologous recombination	03440	12

PDB database but only 7 have no 3D model structure in ModBase among the 125 non homologous essential proteins. All these results are summarized and presented in Table S4. Although the identified 21 membrane proteins could directly be taken as potential vaccine candidates, we have further characterized them by the MMPPP for identification of any predicted antigenic peptide epitopes. All the 21 proteins were found to have antigenic peptide epitope with a cleavage probability of 1 and a binding score ranging from 0.6125 to 0.8946. The results of all the identified predicted antigenic peptide epitopes are summarized in Table S5.

2.4. Druggability of therapeutic targets

The BLASTP search aligning the non homologous essential proteins to the list of drug targeted proteins downloaded from DrugBank was used to examine the druggability of each of the non homologous essential proteins. The BLASTP alignment search identified 42 *M. hyopneumoniae* proteins highly similar to the binding partners of all the drug target proteins (Table 2). Among these 22 were FDA-approved drugs or nutraceuticals or approved under investigation or biotechs and the remaining 20 are small molecule compounds or biotechs under experimentation. Additionally, the identified drug targeted proteins produced a list of experimental, approved drugs, nutraceuticals and biotechs that binds to these. In Table 3 we summarized the identified target protein binding partners of FDA-approved drug list of the 22 *M. hyopneumoniae* target proteins.

2.5. Distributions of essential proteins/genes to metabolic pathways

The distribution and association of the essential gene/proteins were compared in two stages; first, next to the identification of the non homologous essential genes of *M. hyopneumoniae* and second, after similarity search of the non homologous with the DrugBank database, i.e. before and after the non homologous essential proteins/genes were analyzed by homology search to DrugBank database of all drug targets (Fig. 3). The distribution of the number of non homologous essential proteins/genes after DrugBank search reduced by ~65% and the pathways by ~32% which also led to the shift of the priority targets, ribosome, pyrimidine metabolism, purine metabolism, ABC transporters and aminoacyle-tRNA biosynthesis pathways according to concentrations in descending order (Fig. 3A) to ribosome, aminoacyle-tRNA biosynthesis and purine (Fig. 3B). Although the non homologous essential genes/proteins in ABC transporters had a high number of hits their similarity score remains under threshold.

3. Discussion

Recently, the search for drug targets using computational methods and integrated 'omics' data, such as genomics, proteomics, and metabolomics, has received much attention and has been increasing continuously, for instance, comparative and subtractive genomics have been widely used for the prediction and identification of potential therapeutic targets and vaccine candidate proteins in numerous pathogenic bacteria and also fungi [2,9,11,23,43]. These computational methods coupled with the availability of four strains of whole genome sequenced *M. hyopneumoniae* provide us the prospect to perform the first computational comparative and subtractive genomic analysis aimed at the identification of putative therapeutic targets and vaccine candidate proteins of *M. hyopneumoniae*. In the current study the comparisons of initial information of metabolic pathways obtained from the KEGG database revealed the same type and number of metabolic pathways among the four strains of *M. hyopneumoniae* except for the polycyclic aromatic hydrocarbon degradation pathway which was not present in *M. hyopneumoniae* strain 168. The conservation of all the metabolic pathways more specifically genes involved in the metabolic pathways among the strains is a major determining factor in the attempt to identify species specific therapeutic drug target and vaccine candidate proteins. Moreover, previous studies have also indicated conservation among genomes is itself a predictive feature for essentiality of genes in predicting drug targets [11]. Hence, our further insight in genes involved in the pathway of polycyclic aromatic hydrocarbon degradation revealed a single gene characterized as DNA methylase. The protein sequence of this gene product was found less by 24 amino acids in strain 168 suggesting the reason for the KEGG in not predict the pathway in this strain [29,34,41]. Our further insight into each of the metabolic pathways revealed that all the proteins/genes involved in the unique metabolic pathways of bacterial secretion system and methane metabolism were shared with common pathways. This shows that essentially the two pathways are not unique to the pathogen, although they were unique in the direct comparisons of the pathogen pathway list against the swine host.

The possibilities of selecting targets using computational approaches with integrated 'omics' data rely on searching for those genes/proteins that are absent in the host and/or non homologous to the host proteome but present in the pathogen as designing a drug specific to such targets will have effect only on the pathogen and not any other aspect of the host biology [8]. In this aspect our results of the non homologous proteins of *M. hyopneumoniae* through the NCBI BLASTP search against

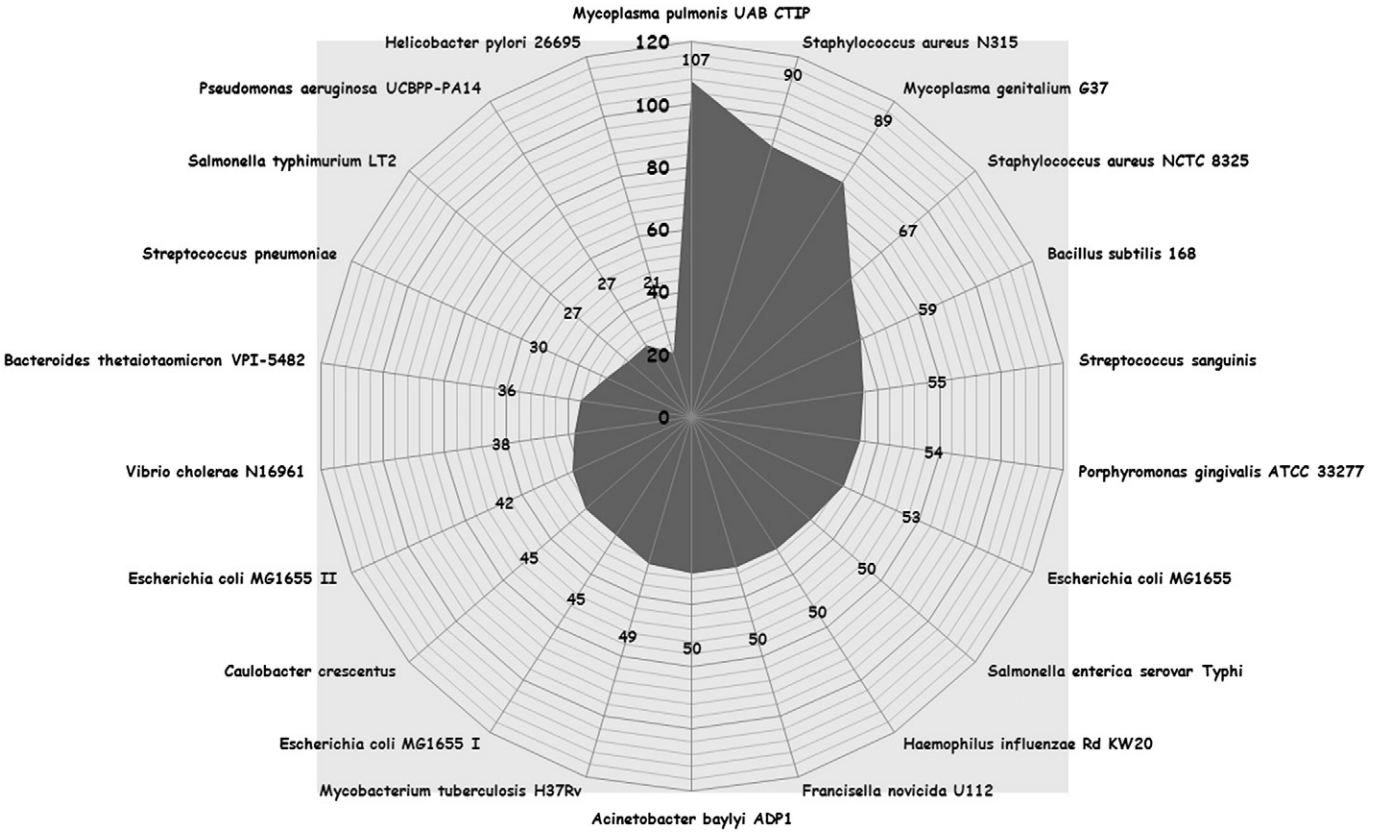

Fig. 2. Frequency of hits of *M. hyopneumoniae* proteins by the 22 bacteria in DEG. Distance from the center are the number of hits of the 125 proteins by essential proteins of the 22 bacteria in DEG.

the pathogen host was much higher compared with a similar study of human pathogen *Mycoplasma genitalium (M. genitalium)* in which only 79 non homologous proteins were identified [7]. This prompted us to repeat the homology search against the human proteome, yet it had resulted in a similar outcome with only a single additional homologous protein to the human proteome but non homologous to the pathogen host. However, the number and/or proportion of non homologous proteins identified from unique pathways were lower in *M. hyopneumoniae* than *M. genitalium*.

The potential of the identified non homologous proteins to be a therapeutic target of a given pathogen gene product is dependent on two broad types of information. First, the role of the gene in the growth and survival of the pathogen defining the essentiality of genes/proteins and second, the likelihood of being able to develop a compound that modulates that target into a drug defining the druggability of the target proteins [1]. These parameters have been used in drug target prioritization in *Mycobacterium ulcerans* [8] *Burkholderia pseudomallei* [9] and *M. genitalium* [8]. Although the search for drug targets among essential genes is a useful approach and considered an important parameter for the prioritization of targets, the inability of this methodology to predict some of the successful drug targets (false negatives) and inclusions of non drug targets (false positives) is a limitation of this practice [1,11]. There have been several attempts to define or predict essential genes in infectious agents through systematic disruption of candidate genes and transposon mutagenesis, for example, in *Streptococcus pneumoniae (S. pneumoniae)*, *Mycobacterium tuberculosis* (*M. tuberculosis*) and *M. genitalium* [17,38,40]. However, for the vast majority of infectious organisms, high throughput experimental interrogation of essentiality is currently not possible, and efforts to define essentiality have been based on bioinformatic predictions [1,11,43]. Consistent with this, 7 of 8 computationally predicted essential genes via DEG-based homology search were also experimentally validated as essential for *Yersinia pseudotuberculosis (Y. pseudotuberculosis)* [12]. Here, we have attempted to predict the essentiality of non homologous proteins of *M. hyopneumoniae* following a similar fashion based on bioinformatic predictions through homology search with the already identified essential gene/protein database from 22 bacteria in DEG [45,46]. This approach filtered and reduced 34.03% of the non homologous proteins of *M. hyopneumoniae* protein targets. Although the reduction was twice higher proportionally compared with a similar report on *M. genitalium* which had reduced by 15.18%, the number of essential non homologous proteins was still higher in this study. However, this is not unexpected as a previous experiment in global transposon mutagenesis of *M. genitalium* and *M. pulmonis* reported 382 and 310 protein-coding essential genes, respectively [15,19]. This was also evidenced by the number of total hits observed for the essential genes of *M. hyopneumoniae* by *M. genitalium* next to *M. pulmonis* UAB CTIP and *Staphylococcus aureus*(*S. aureus*) N315 (Fig. 2). Hence, the difference could be attributed for the higher similarity or homologous proteins identified in the human proteome by *M. genitalium* than *M. hyopneumoniae*.

Advances in genome sequencing, bioinformatics and cheminformatics coupled with experimental data have shown several additional factors such as molecular mass (<110), subcellular localization, whether the protein is transmembrane and availability of 3D structural information that can aid in determining the suitability of therapeutic targets besides identifying the proteins of the pathogen as non homologous and essential protein [1,7,8]. Proteins with smaller molecular mass were found very likely to be soluble and easier to purify [12]. The characterization and prioritization of the proteins using their molecular mass resulted in much of the identified non homologous essential proteins to be less than 110 (96.73%, n = 126) suggesting higher possibility of experimental verification of many of the identified proteins for drug targets (Table S4). Furthermore, among the critical reasons mentioned for knowledge of the localization of proteins within cellular compartments, the factor that aids in determining the suitability of therapeutic targets, is understanding their function which yields important insight that can lead to

Table 2
Non-homologous essential proteins of *M. hyopneumoniae* similar to binding partners of FDA approved drugs, experimental small molecule compounds, or nutraceutical compounds as inferred from DrugBank database using BLASTP.

No.	Gene name	Protein ID	Pathways involved name (ID)	Protein existence	Subcellular location	TMHMM[a] no.	3D Exp[b]	3D Model	MW (KDa)
1	fba	Q602D6	Glycolysis/gluconeogenesis (00010) Pentose phosphate pathway (00030) Fructose and mannose metabolism (00051) Methane metabolism (00680)	Predicted	Cytoplasmic	No	No	Yes	31
2	ictD	P0C0J2	Glycolysis/gluconeogenesis (00010) Cysteine and methionine metabolism (00270) Pyruvate metabolism (00620) Propanoate metabolism (00640)	[c]Inferred H	Cytoplasmic	No	No	Yes	34
3	pdhA	Q601D7	Glycolysis/gluconeogenesis (00010) Citrate cycle (TCA cycle) (00020) Pyruvate metabolism (00620)	Predicted	Cytoplasmic	No	No	Yes	42
4	pgi	Q600A8	Glycolysis/gluconeogenesis (00010) Pentose phosphate pathway (000300 Starch and sucrose metabolism (00500) Amino sugar and nucleotide sugar metabolism (00520)	Inferred H	Cytoplasmic	No	No	Yes	46
5	pdhD	Q5ZZX0	Glycolysis/gluconeogenesis (00010) Citrate cycle (TCA cycle) (00020) Glycine, serine and threonine metabolism (00260) Valine, leucine and isoleucine degradation (00280) Pyruvate metabolism (00620)	Inferred H	Cytoplasmic	No	No	Yes	51
6	hpt	Q601T4	Purine metabolism (00230)	Predicted	Cytoplasmic	No	No	Yes	20
7	nrdE	Q601P3	Purine metabolism (00230) Pyrimidine metabolism (00240)	Inferred H	Membrane	No	No	Yes	82
8	adk	Q601J4	Purine metabolism (00230)	Inferred H	Cytoplasmic	No	No	Yes	24
9	rpoA	Q601I9	Purine metabolism (00230) Pyrimidine metabolism (00240) RNA polymerase (03020)	Inferred H	Cytoplasmic	No	No	Yes	37
10	rpoB	Q5ZZS1	Purine metabolism (00230) Pyrimidine metabolism (00240) RNA polymerase (03020)	Inferred H	Membrane	No	No	Yes	138
11	cmk	Q601X7	Pyrimidine metabolism (00240)	Inferred H	Cytoplasmic	No	No	Yes	26
12	trxB	Q601C1	Pyrimidine metabolism (00240) Selenocompound metabolism (00450)	Inferred H	Cytoplasmic	No	No	Yes	34
13	upp	Q600B3	Pyrimidine metabolism (00240)	Predicted	Cytoplasmic	No	No	Yes	23
14	mmsA	Q601P8	Valine, leucine and isoleucine degradation (00280) Inositol phosphate metabolism (00562) Propanoate metabolism (00640)	Inferred H	Cytoplasmic	No	No	Yes	54
15	pta	Q600E9	Taurine and hypotaurine metabolism (00430) Pyruvate metabolism (00620) Propanoate metabolism (00640) Methane metabolism (00680)	Predicted	Cytoplasmic	No	No	Yes	35
16	metG	Q600P5	Selenocompound metabolism (00450) Aminoacyl-tRNA biosynthesis (00970)	Predicted	Cytoplasmic	No	No	Yes	62
17	nagB	Q5ZZW6	Amino sugar and nucleotide sugar metabolism (00520)	Predicted	Cytoplasmic	No	No	Yes	29
18	gatB	Q602C0	Aminoacyl-tRNA biosynthesis (00970)	Inferred H	Cytoplasmic	No	No	Yes	54
19	tyrS	Q601Y5	Aminoacyl-tRNA biosynthesis (00970)	Inferred H	Cytoplasmic	No	No	Yes	48
20	pheS	Q601U5	Aminoacyl-tRNA biosynthesis (00970)	Predicted	Cytoplasmic	No	No	Yes	26
21	serS	Q601S3	Aminoacyl-tRNA biosynthesis (00970)	Predicted	Cytoplasmic	No	No	Yes	47
22	gltX	Q601G1	Aminoacyl-tRNA biosynthesis (00970)	Inferred H	Cytoplasmic	No	No	Yes	55
23	valS	Q5ZZL4	Aminoacyl-tRNA biosynthesis (00970)	Inferred H	Cytoplasmic	No	No	Yes	97
24	ptsI	Q600I5	Phosphotransferase system (PTS) (02060)	Inferred H	Cytoplasmic	No	No	Yes	64
25	rpsB	Q601Z3	Ribosome (03010)	Inferred H	Cytoplasmic	No	No	Yes	36
26	rpsG	Q601W7	Ribosome (03010)	Inferred H	Cytoplasmic	No	No	Yes	18
27	rpsL	Q601W6	Ribosome (03010)	Inferred H	Cytoplasmic	No	No	Yes	15
28	rplC	Q601L5	Ribosome (03010)	Inferred H	Cytoplasmic	No	No	Yes	25
29	rplD	Q601L4	Ribosome (03010)	Inferred H	Cytoplasmic	No	No	Yes	24
30	rpsS	Q601L1	Ribosome (03010)	Inferred H	Cytoplasmic	No	No	Yes	10
31	rpsC	Q601K8	Ribosome (03010)	Inferred H	Cytoplasmic	No	No	Yes	25
32	rpsH	Q601K0	Ribosome (03010)	Inferred H	Cytoplasmic	No	No	Yes	14
33	rpsE	Q601J7	Ribosome (03010)	Inferred H	Cytoplasmic	No	No	Yes	24
34	rpsM	Q601J1	Ribosome (03010)	Inferred H	Cytoplasmic	No	No	Yes	16
35	rplK	Q600J6	Ribosome (03010)	Inferred H	Cytoplasmic	No	No	Yes	16
36	rpsD	Q5ZZW4	Ribosome (03010)	Inferred H	Cytoplasmic	No	No	Yes	24
37	rpsI	Q5ZZN6	Ribosome (03010)	Inferred H	Cytoplasmic	No	No	Yes	15
38	lig	Q601T8	DNA replication (03030) Base excision repair (03410) Nucleotide excision repair (03420) Mismatch repair (03430)	Inferred H	Cytoplasmic	No	No	Yes	79
39	ftsY	Q602E2	Protein export (03060) Bacterial secretion system (03070)	Predicted	Cytoplasmic	No	No	Yes	37
40	secA	Q601A7	Protein export (03060) Bacterial secretion system (03070)	Inferred H	Cytoplasmic	0	No	Yes	112

(continued on next page)

Table 2 (*continued*)

No.	Gene name	Protein ID	Pathways involved name (ID)	Protein existence	Subcellular location	TMHMM[a] no.	3D Exp[b]	3D Model	MW (KDa)
41	recA	Q602A9	Homologous recombination (03440)	Inferred H	Cytoplasmic	No	No	Yes	36
42	ruvB	Q600N3	Homologous recombination (03440)	Inferred H	Cytoplasmic	No	No	Yes	36

[a] Number of transmembrane protein predicted by THMMH.
[b] Experimentally confirmed 3D structure.
[c] Inferred homology.

improving therapeutics and diagnosis [18]. The membrane localization of 21 proteins from the non homologous essential proteins identified using CELLO server were in line with the prediction of 18 transmembrane by TMHMM which were among the 21 predicted earlier. These targets represent potential vaccine candidates for development of an effective vaccine against *M. hyopneumoniae* as many of the vaccines that protect by eliciting antibody mediated immunity appear to be secreted toxins and/or highly expressed, surface exposed molecules [10]. The new genomic based approaches shown to be very powerful for the discovery of vaccine candidates are labor intensive and time consuming in that the identification of the few protective antigens requires the screening of a large number of recombinant proteins in biological assays usually involving animal models [16,23]. Hence, the current approach could be a response to the highly desirable strategy that substantially reduces the number of proteins to be tested.

Although *M. hyopneumoniae* is a highly studied swine respiratory pathogen with the whole genome of 4 of its strains already sequenced, there is no single experimentally verified protein within the PDB database, indicating the gap in the available study of protein structure and the skewedness of research areas in *M. hyopneumoniae*. However, most of the essential non homologs have a 3D model in Modbase. Structure libraries of drug target proteins play a key role in drug design, enabling computational methods to score and rank the predicted affinity between drugs and targets. This process is known as virtual screening and has reduced the costs of experimental high-throughput assays. Virtual screening is knowledge-based and requires structural information of the target and ligand (ligand-based screening) or of the target alone (target-based screening) [19,35].

The second determinant for the potential of a non homologous protein to be a therapeutic target, the druggability of the protein, was determined with the homology search against the DrugBank target proteins. This search besides its advantage of identification of the pathogen proteins that are highly similar to the binding partners of all drug targets of DrugBank and reducing the testable proteins to 42, produced a list of different types of drugs in the DrugBank that bind to proteins of the pathogen. However, the protein sequence similarity search had limitations in that it does not assure the structure or biding pockets identicalness of the identified similar proteins. Our attempt to further validate and characterize the proteins through 3D homology search had also failed for the absence of experimentally confirmed 3D structure in the databases of PDB as mentioned above and also in the literature. Among the identified FDA drug lists none matched to the known virulence factors such as P97 adhesion protein. This could be due to the non essential nature of the virulence factors that would exclude them from the list as essential proteins of the pathogen.

Finally when we come to specific targets identified, from the unique metabolic pathways, a single protein i.e. phosphoenolpyruvate-protein phosphotransferase (ptsI) enzyme from phosphotransferase system (PTS) pathway was identified as drug target against *M. hyopneumoniae.* This pathway is a major mechanism of carbohydrate metabolism and critical for growth and survival of the bacteria. This was the only target protein identified from the unique pathways of *M. hyopneumoniae* not shared with common pathways and also previously reported in *M. genitalium* [7]. Although there were several essential non homologous identified protein targets from unique pathways of the pathogen observed to be involved in multiple pathways of the common pathways, there were no drug targets identified shared by the unique pathway (Fig. 3). However, there was good number of vaccine candidates identified from essential non homologous proteins shared by both unique and common pathways as several of them were predicted to be membrane proteins (Table S4). Several of the targets were found to be involved in multiple pathways and targeting these proteins may lead to development of more potent drugs.

Both fructose-bisphosphate aldolase (fba) an enzyme that catalyzes a reversible reaction required for both glycolysis and gluconeogenesis and glucose-6-phosphate isomerase (pgi) an enzyme that catalyzes the reversible isomerization of D-glucose 6-phosphate to D-fructose 6-phosphate in the glycolysis pathway which have been previously reported as targets of antifungal and anti-protozoa drugs, respectively [31,37] were found as drug targets for experimental small molecule compounds in DrugBank in the current study. DNA ligase (lig), a key enzyme that catalyzes the formation of phosphodiester bonds at single stranded or double stranded breaks between adjacent 5′ phosphoryl and 3′ hydroxyl groups of DNA, is an important enzyme for survival because of its involvement in major cellular processes like DNA replication/repair and recombination [13]. However, this target albeit previous reports as drug target of antibiotics specially in overcoming current drug resistance issues did not achieve a threshold similarity with any of the FDA approved drug targets from DrugBank. Besides the above three enzymes involved in multiple pathways of *M. hyopneumoniae* with hits for only experimental drug targets of the DrugBank with no previous reports was phosphate acetyltransferase (pta).

L-lactate dehydrogenase (L-LDH) protein also called immunogenic protein p36 was among the drug targets found to be involved in multiple common pathways and highly similar to the binding partner of one FDA-approved and four experimental drug targets from DrugBank (Table 3). This protein has been previously confirmed to produce strong immunogenic reaction, however antibodies raised against *M. hyopneumoniae* do not inhibit its growth or metabolic activities as confirmed experimentally [21]. Hence, the potential as drug target remains open for experimental validation. The other two drug target proteins identified involved in multiple common pathways with hits of FDA approved drug targets, are methylmalonate-semialdehyde dehydrogenase (mmsA) and dihydrolipoyl dehydrogenase (pdhD) the functions of which have been studied in *Bacillus subtilis* for the earlier [39] and *M. hyopneumoniae* itself for the later [33] which have indicated a critical role in the energy metabolism and transcriptions of an organism symptomatic of the possibility of their potential role in the development of a drug against *M. hyopneumoniae*. Furthermore, the higher distribution of FDA approved drug targets in ribosome and aminoacyl-tRNA biosynthesis common pathways of the pathogen (Fig. 3) prompted us for more insight in these pathways. We observed an increasing interest in the assembly of bacterial ribosomes as a potential target for new antibiotics especially with the recently detailed review. In this review the assembly of ribosomes which begins with transcription of rRNA from multiple copies of rRNA operons had been reported to disfavor resistance by the need to mutate multiple copies unlike mutations in the singly encoded ribosomal proteins, and for this reason antibiotics that target the rRNA had become attractive [6].

Aminoacyl-tRNA biosynthesis, the second common pathway with higher distribution of identified targets, had been also previously reported in *M. genitalium* for its multiple target enzymes whose inhibition could

Table 3
Non-homologous essential proteins of *M. hyopneumoniae* similar to binding partners of FDA approved drugs as inferred from DrugBank database using BLASTP and the list of FDA approved drugs for the targets.

NO	Protein ID	Pathways involved name (ID)	DrugBank ID	Drug name list	Drug group
1	P0C0J2	Glycolysis/gluconeogenesis (00010), cysteine and methionine metabolism (00270), pyruvate metabolism (00620), propanoate metabolism (00640)	DB00157	NADH	Approved, nutraceutical
2	Q601D7	Glycolysis/gluconeogenesis (00010), citrate cycle (TCA cycle) (00020), pyruvate metabolism (00620)	DB00157	NADH	Approved, nutraceutical
3	Q5ZZX0	Glycolysis/gluconeogenesis (00010), citrate cycle (TCA cycle) (00020), glycine, serine and threonine metabolism (00260), pyruvate metabolism (00620)	DB01169, DB00262, DB00143, DB00157	Arsenic trioxide, Carmustine, NADH	Approved, investigational; approved; approved, nutraceutical (2)
4	Q601P3	Purine metabolism (00230), pyrimidine metabolism (00240)	DB00242, DB00631, DB01073, DB00441, DB01005	Cladribine, Clofarabine, Fludarabine, Gemcitabine, Hydroxyurea	Approved, investigational (2); approved (3)
5	Q601I9	Purine metabolism (00230), pyrimidine metabolism (00240), RNA polymerase (03020)	DB00615	Rifabutin	Approved
6	Q5ZZS1	Purine metabolism (00230), pyrimidine metabolism (00240), RNA polymerase (03020)	DB00615, DB01045, DB01220	Rifabutin, Rifampin, Rifaximin	Approved (2); approved, investigational
7	Q601C1	Pyrimidine metabolism (00240), selenocompound metabolism (00450)	DB00548	Azelaic acid	Approved
8	Q600B3	Pyrimidine metabolism (00240)	DB00548	Azelaic acid	Approved
9	Q601P8	Valine, leucine and isoleucine degradation (00280), inositol phosphate metabolism (00562), propanoate metabolism (00640)	DB00157, DB00165, DB00534, DB00139, DB00116, DB00755, DB00162, DB01612, DB00822, DB00536, DB00435, DB00727	NADH, Pyridoxine, Chlormerodrin, Succinic acid, Tetrahydrofolic acid, Vitamin A, Amyl nitrite, Disulfiram, Guanidine, Nitric oxide, Nitroglycerin	Approved, nutraceutical (5); approved (5); approved, investigational; approved, nutraceutical, investigational
10	Q602C0	Aminoacyl-tRNA biosynthesis (00970)	DB00142, DB00130	L-Glutamic Acid, L-Glutamine	approved, nutraceutical; approved, nutraceutical, investigational
11	Q601Y5	Aminoacyl-tRNA biosynthesis (00970)	DB00135	L-Tyrosine	approved, nutraceutical
12	Q601S3	Aminoacyl-tRNA biosynthesis (00970)	DB00133	L-Serine	approved, nutraceutical
13	Q601G1	Aminoacyl-tRNA biosynthesis (00970)	DB00142	L-Glutamic Acid	approved, nutraceutical
14	Q5ZZL4	Aminoacyl-tRNA biosynthesis (00970)	DB00161, DB00167, DB00410	L-Valine, L-Isoleucine, Mupirocin	approved, nutraceutical (2); approved, investigational
15	Q601W6	Ribosome (03010)	DB00479, DB06696, DB00452, DB00798, DB01172, DB00994, DB00955, DB00919, DB01082, DB00560, DB00684	Amikacin, Arbekacin, Framycetin, Gentamicin, Kanamycin, Neomycin, Netilmicin, Spectinomycin, Streptomycin, Tigecycline, Tobramycin	Approved (10); approved, investigational
16	Q601L5	Ribosome (03010)	DB01256	Retapamulin	approved
17	Q601L4	Ribosome (03010)	DB01321	Josamycin	approved
18	Q601L1	Ribosome (03010)	DB00759, DB00560	Tetracycline, Tigecycline	approved
19	Q601J1	Ribosome (03010)	DB00560	Tigecycline	approved
20	Q5ZZW4	Ribosome (03010)	DB00453, DB00618, DB00254, DB00256, DB01017, DB00595	Clomocycline, Demeclocycline, Doxycycline, Lymecycline, Minocycline, Oxytetracycline	Approved (4); approved, investigational (2)
21	Q5ZZN6	Ribosome (03010)	DB00453, DB00618, DB00254, DB00256, DB01017, DB00596, DB01301, DB00560	Clomocycline, Demeclocycline, Doxycycline, Lymecycline, Minocycline, Oxytetracycline, Rolitetracycline, Tigecycline	Approved (6); approved, investigational (2)
22	Q600N3	Homologous recombination (03440)	DB00173	Adenine	approved, nutraceutical

lead to disruption of protein biosynthesis that would in turn attenuate bacterial growth in both experimental and infectious conditions [7].

In summary the computational comparative and subtractive genomic/proteomic analysis of different metabolic pathways resulted in a step wise reduction (Fig. 4) and identification of several proteins of *M hyopneumoniae* that can be targeted for effective drug design and vaccine development. As many of the identified drug targets have been reported to play a role in the critical metabolic pathways that regulate bacterial growth, energy metabolism, protein biosynthesis and essential nutrient uptake, a systematic approach to develop drugs against these targets would help to combat the chronic swine pneumonia caused by this pathogen. As the drugs developed this way are also expected to be specific to the pathogen the development of antibacterial drug resistance and toxicity for the host could be minimized. Experimental investigation on dual functions of a single identified target protein and in combination as drug and vaccine candidate at the same time is underway.

4. Materials and methods

4.1. Genomic proteome databases and alignment tools

The Refseq genomic nucleotide sequences and genomic Refseq protein sequence of four strains of *M. hyopneumoniae* (Refseq: NC_006360.1, NC_007332.1, NC_007295.1 and NC_017509.1) and standalone release of NCBI BLASTP+ version 2.2.27(ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/) were downloaded on a local machine and BLASTP was installed. Additionally, sequences of experimentally verified essential genes and proteins from 22 Gram-positive and Gram-negative bacteria (Table S1) from the latest update (version 7; November 11, 2012) of the Database of Essential Genes (DEG) [45,46] containing 6711 drug entries, including 1441 FDA-approved small molecule drugs, 134 FDA-approved biotech (protein/peptide) drugs, 84 nutraceuticals, and

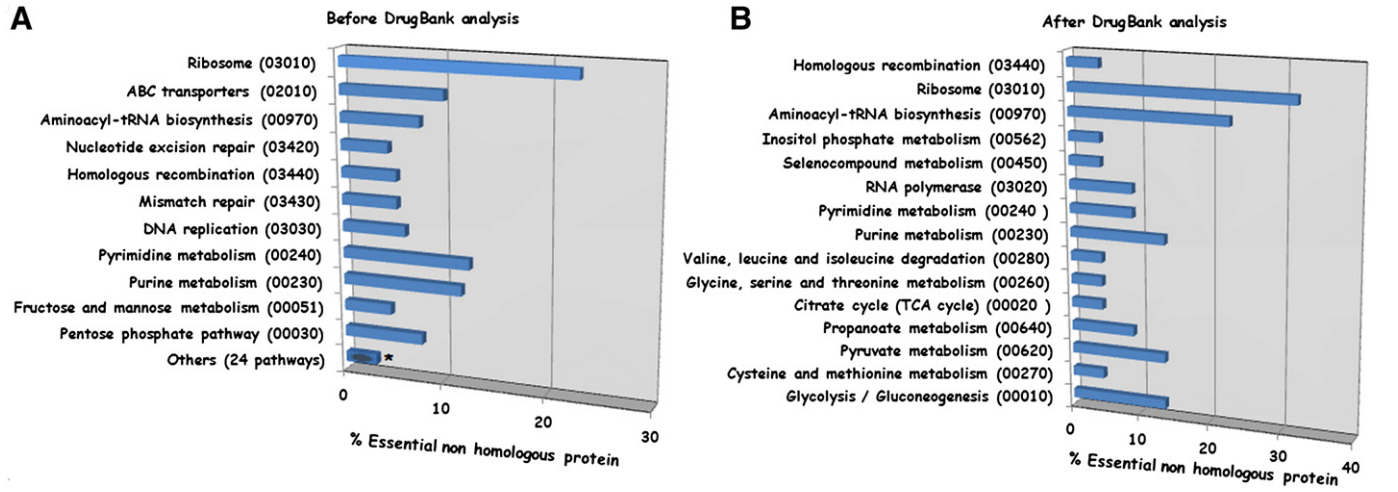

Fig. 3. Distribution of essential proteins in metabolic pathways. Percentage distribution of essential non homologous proteins of *M. hyopneumoniae* into their associated pathways before (A) and after (B) the essential non homologous proteins/genes were analyzed by homology search of DrugBank database. (* 24 metabolic pathways whose essential non homologous percentage distribution ranges from 0.8 to 4%).

5084 experimental drugs with additional 4231 non-redundant protein (i.e. drug target/enzyme/transporter/carrier) sequences linked to these drug entries from the DrugBank (http://www.drugbank.ca/) were downloaded and saved locally [26]. For genome wide analysis of putative target proteins of the pathogen, the locally saved databases and the search engine were used as presented in the systemic work flow (Fig. 1) and as described below deliberated for the reductive analysis of comparative genomics for the identification of putative therapeutic and vaccine targets against *M. hyopneumoniae*. Besides the locally saved databases, other online databases and NCBI BLAST search were used as required.

4.2. Host and pathogen metabolic pathway analysis using KEGG

The Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg) through its link to NCBI databases was used for genome wide pathways analysis of the four strains of whole genome sequenced *M. hyopneumoniae* and retrieval of their information [24,25]. A list of metabolic pathways and the respective information of the host (swine), human and the pathogen *M. hyopneumoniae* were retrieved and saved locally. A manual comparison was conducted and pathways that did not appear in the host and human but were present in the pathogen, according to the KEGG database annotations were selected as unique pathways to *M. hyopneumoniae*, whereas the remaining pathways were listed as common pathways. Proteins from common and unique pathways were identified and the respective amino acid sequences were selected from the NCBI downloaded pathogen genome database.

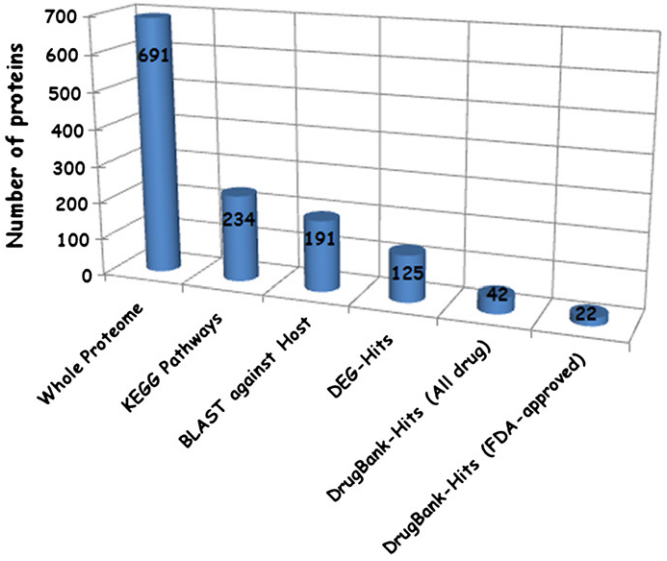

Fig. 4. Summary of target protein identification. Number of proteins in stepwise subtractive analysis of whole proteome of *M. hyopneumoniae* in drug target and vaccine candidate protein identification.

4.3. Selection of non homologous and non homologous essential pathogen proteins

A two step comparison of the pathogen proteome from both unique and common pathways, first with host proteome for identification of non homologous proteome of the pathogen and then a second step similarity search of the identified non homologous proteins of the pathogen against the protein sequences of the essential genes/proteins from the database of essential gens (DEG) for the identification of the essential non homologous proteins of the pathogen were performed with BLASTP analysis. The comparative searches against the host were restricted to the host proteins through an option available under BLASTP parameters. Hits were filtered on the basis of expectation value (e value) inclusion threshold being set to 0.005, and a minimum bit score of 100. Proteins that didn't have hits below the e value inclusion threshold were picked out as non homologous proteins of the pathogen. These proteins were further screened on the basis of essentiality using DEG at cut off e value of 10^{-10} and a minimum bit score of 100 using DEG microbial BLASTP (Zhang et al., 2004).

4.4. Prioritization of essential non homologous proteins of the pathogen for drug target

The molecular and structural criteria proposed to aid in prioritizing suitable therapeutic targets in pathogenic microorganisms [1] such as calculation of molecular weight (MW) using computational tools and drug target-associated literature available in the Swiss-Prot database [5], prediction of biological significance and subcellular localization of therapeutic targets using CELLO v.2.5 (multi-class support vector machine classification system) [44] and TMHMM v2.0 (Transmembrane predictions; http://www.cbs.dtu.dk/services/TMHMM/) [27] were

performed on the identified target proteins. Additionally, experimentally and computationally solved 3D structures were detected by searching the Protein Data Bank (PDB) (http://www.rcsb.org/pdb) [3,4] and ModBase (http://salilab.org/modbase) [36] databases, respectively. The MAPPP (MHC-I Antigenic Peptide Processing Prediction/ http://www.mpiib-berlin.mpg.de/MAPPP/expertquery.html) with the SYFPEITHY matrix which combines existing prediction tools for proteasomal processing and MHC class I anchoring were used for peptide prediction [20].

4.5. Druggability of essential non homologous proteins of the pathogen

The druggability potential, the likelihood of being able to develop a drug-like compound to modulate the target, of each identified drug target was evaluated by aligning the pathogen target with the DrugBank contents of all and FDA-approved drugs, small molecule drugs, biotech (protein/peptide) drugs and nutraceuticals. A BLASTP with default parameters was performed to align the potential drug targets from the pathogen against the list of protein targets of compounds of the mentioned data bases within DrugBank with a selection criterion for filtering as described previously for identification of drug targets in bacterial genomes [22].

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2013.04.011.

Acknowledgments

This work was supported in part by a grant from the Next-Generation BioGreen 21 Program (No. PJ009007), the Rural Development Administration, Republic of Korea and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2011-0021670).

References

[1] F. Aguero, B. Al-Lazikani, M. Aslett, M. Berriman, F.S. Buckner, R.K. Campbell, S. Carmona, I.M. Carruthers, A.W. Chan, F. Chen, G.J. Crowther, M.A. Doyle, C. Hertz-Fowler, A.L. Hopkins, G. McAllister, S. Nwaka, J.P. Overington, A. Pain, G.V. Paolini, U. Pieper, S.A. Ralph, A. Riechers, D.S. Roos, A. Sali, D. Shanmugam, T. Suzuki, W.C. Van Voorhis, C.L. Verlinde, Genomic-scale prioritization of drug targets: the TDR targets database, Nat. Rev. Drug. Discov. 7 (2008) 900–907.
[2] U. Amineni, D. Pradhan, H. Marisetty, In silico identification of common putative drug targets in *Leptospira interrogans*, J. Chem. Biol. 3 (2010) 165–173.
[3] H.M. Berman, K. Henrick, H. Nakamura, Announcing the worldwide Protein Data Bank, Nat. Struct. Biol. 10 (2003) 980.
[4] H.M. Berman, K. Henrick, H. Nakamura, J.L. Markley, The Worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data, Nucleic Acids Res. 35 (2007) D301–D303.
[5] B. Boeckmann, A. Bairoch, R. Apweiler, M.C. Blatter, A. Estreicher, E. Gasteiger, M.J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout, M. Schneider, The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003, Nucleic Acids Res. 31 (2003) 365–370.
[6] A.M. Bruce, Inhibition of bacterial ribosome assembly: a suitable drug target? Microbiol. Mol. Biol. Rev. 73 (1) (2009) 22–35.
[7] A.M. Butt, S. Tahir, I. Nasrullah, M. Idrees, J. Lu, Y. Tong, *Mycoplasma genitalium*: a comparative genomics study of metabolic pathways for the identification of drug and vaccine targets, Infect. Genet. Evol. 12 (1) (2012) 53–62.
[8] A.M. Butt, I. Nasrullah, S. Tahir, Y. Tong, Comparative genomics analysis of *Mycobacterium ulcerans* for the identification of putative essential genes and therapeutic candidates, PLoS One 7 (8) (2012) e43080.
[9] C.E. Chong, B.S. Lim, S. Nathan, R. Mohamed, In silico analysis of *Burkholderia pseudomallei* genome sequence for potential drug targets, In Silico Biol. 6 (2006) 341–346.
[10] F. Doro, S. Liberatori, M.J. Rodríguez-Ortega, C.D. Rinaudo, R. Rosini, M. Mora, M. Scarselli, E. Altindis, R. D'Aurizio, M. Stella, I. Margarit, D. Maione, J.L. Telford, N. Norais, G. Grandi, Surfome analysis as a fast track to vaccine discovery: identification of a novel protective antigen for Group B *Streptococcus* hypervirulent strain COH1, Mol. Cell. Proteomics 8 (7) (2009) 1728–1737.
[11] M.A. Doyle, R.B. Gasser, B.J. Woodcroft, R.S. Hall, S.A. Ralph, Drug target prediction and prioritization: using orthology to predict essentiality in parasite genomes, BMC Genomics 11 (2010) 222.
[12] M. Duffield, I. Cooper, E. McAlister, M. Bayliss, D. Ford, P. Oyston, Predicting conserved essential genes in bacteria: in silico identification of putative drug targets, Mol. Biosyst. 6 (2010) 2482–2489.
[13] N. Dwivedi, D. Dube, J. Pandey, B. Singh, V. Kukshal, R. Ramachandran, R.P. Tripathi, NAD(+)-dependent DNA ligase: a novel target waiting for the right inhibitor, Med. Res. Rev. 28 (4) (2008) 545–568.
[14] E. Fano, C. Pijoan, S. Dee, Dynamics and persistence of *Mycoplasma hyopneumoniae* infection in pigs, Can. J. Vet. Res. 69 (3) (2005) 223–228.
[15] C.T. French, P. Lao, A.E. Loraine, B.T. Matthews, H. Yu, K. Dybvig, Large-scale transposon mutagenesis of *Mycoplasma pulmonis*, Mol. Microbiol. 69 (1) (2008) 67–76.
[16] G. Grandi, Genomics and proteomics in reverse vaccines, Methods Biochem. Anal. 49 (2006) 379–393.
[17] J.I. Glass, N. Assad—Garcia, N. Alperovich, S. Yooseph, M.R. Lewis, M. Maruf, C.A. Hutchison, H.O. Smith, J.C. Venter, Essential genes of a minimal bacterium, Proc Natl Acad Sci U S A. 103 (2006) 425–430.
[18] E. Glory, R.F. Murphy, Automated subcellular location determination and high—throughput microscopy, Dev Cell. 12 (2007) 7–16.
[19] M.A. Grant, Protein structure prediction in structure-based ligand design and virtual screening, Comb. Chem. High Throughput Screen. 12 (10) (2009) 940–960.
[20] J. Hakenberg, A.K. Nussbaum, H. Schild, H.G. Rammensee, C. Kuttler, H.G. Holzhütter, P.M. Kloetzel, S.H. Kaufmann, H.J. Mollenkopf, MAPPP: MHC class I antigenic peptide processing prediction, Appl. Bioinforma. 2 (3) (2003) 155–158.
[21] A. Haldimann, J. Nicolet, J. Frey, DNA sequence determination and biochemical analysis of the immunogenic protein P36, the lactate dehydrogenase (LDH) of *Mycoplasma hyopneumoniae*, J. Gen. Microbiol. 139 (2) (1993) 317–323.
[22] A.G. Holman, P.J. Davis, J.M. Foster, C.K. Carlow, S. Kumar, Computational prediction of essential genes in an unculturable endosymbiotic bacterium, Wolbachia of *Brugia malayi*, BMC Microbiol. 9 (2009) 243.
[23] A.K. Johri, L.C. Paoletti, P. Glaser, M. Dua, P.K. Sharma, G. Grandi, R. Rappuoli, Group B *Streptococcus*: global incidence and vaccine development, Nat. Rev. Microbiol. 4 (2006) 932–942.
[24] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, M. Hirakawa, KEGG for representation and analysis of molecular networks involving diseases and drugs, Nucleic Acids Res. 38 (2010) D355–D360.
[25] M. Kanehisa, S. Goto, M. Hattori, K.F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, M. Hirakawa, From genomics to chemical genomics: new developments in KEGG, Nucleic Acids Res. 34 (2006) D354–D357.
[26] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, Y. Djoumbou, R. Eisner, A.C. Guo, D.S. Wishart, DrugBank 3.0: a comprehensive resource for 'omics' research on drugs, Nucleic Acids Res. 39 (2011) D1035–D1041.
[27] A. Krogh, B. Larsson, G. von Heijne, E.L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, J. Mol. Biol. 305 (2001) 567–580.
[28] S. Kumar, K. Chaudhary, J.M. Foster, J.F. Novelli, Y. Zhang, S. Wang, D. Spiro, E. Ghedin, C.K. Carlow, Mining predicted essential genes of *Brugia malayi* for nematode drug targets, PLoS One 2 (11) (2007) e1189.
[29] W. Liu, Z. Feng, L. Fang, Z. Zhou, Q. Li, S. Li, R. Luo, L. Wang, H. Chen, G. Shao, S. Xiao, Complete genome sequence of *Mycoplasma hyopneumoniae* strain 168, J. Bacteriol. 193 (2011) 1016–1017.
[30] W. Liu, L. Fang, M. Li, S. Li, S. Guo, R. Luo, Z. Feng, B. Li, Z. Zhou, G. Shao, H. Chen, S. Xiao, Comparative genomics of *Mycoplasma*: analysis of conserved essential genes and diversity of the pan-genome, PLoS One 7 (4) (2012) e35698, http://dx.doi.org/10.1371/journal.pone.0035698.
[31] S.S. Loo, D.P. Blake, A. Mohd-Adnan, R. Mohamed, K.L. Wan, *Eimeria tenella* glucose-6-phosphate isomerase: molecular characterization and assessment as a target for anti-coccidial control, Parasitology 137 (8) (2010) 1169–1177.
[32] D. Maes, J. Segales, T. Meyns, M. Sibila, M. Pieters, F. Haesebrouck, Control of *Mycoplasma hyopneumoniae* infections in pigs, Vet. Microbiol. 126 (4) (2008) 297–309.
[33] J.N. Matic, J.L. Wilton, R.J. Towers, A.L. Scarman, F.C. Minion, M.J. Walker, S.P. Djordjevic, The pyruvate dehydrogenase complex of *Mycoplasma hyopneumoniae* contains a novel lipoyl domain arrangement, Gene 319 (2003) 99–106.
[34] F.C. Minion, E.J. Lefkowitz, M.L. Madsen, B.J. Cleary, S.M. Swartzell, G.G. Mahairas, The genome sequence of *Mycoplasma hyopneumoniae* strain 232, the agent of swine mycoplasmosis, J. Bacteriol. 186 (2004) 7123–7133.
[35] T.I. Oprea, H. Matter, Integrating virtual screening in lead discovery, Curr. Opin. Chem. Biol. 8 (4) (2004) 349–358.
[36] U. Pieper, B.M. Webb, D.T. Barkan, D. Schneidman-Duhovny, A. Schlessinger, H. Braberg, Z. Yang, E.C. Meng, E.F. Pettersen, C.C. Huang, R.S. Datta, P. Sampathkumar, M.S. Madhusudhan, K. Sjölander, T.E. Ferrin, S.K. Burley, A. Sali, ModBase, a database of annotated comparative protein structure models, and associated resources, Nucleic Acids Res. 39 (2011) D465–D474.
[37] A. Rodaki, T. Young, A.J. Brown, Effects of depleting the essential central metabolic enzyme fructose-1,6-bisphosphate aldolase on the growth and viability of *Candida albicans*: implications for antifungal drug target discovery, Eukaryot. Cell 5 (8) (2006) 1371–1377.
[38] C.M. Sassetti, D.H. Boyd, E.J. Rubin, Comprehensive identification of conditionally essential genes in mycobacteria, Proc. Natl. Acad. Sci. 98 (22) (2001) 12712–12717.
[39] F. Talfournier, C. Stines-Chaumeil, G. Branlant, Methylmalonate-semialdehyde dehydrogenase from *Bacillus subtilis*: substrate specificity and coenzyme A binding, J. Biol. Chem. 286 (25) (2011) 21971–21981.
[40] J.A. Thanassi, S.L. Hartman-Neumann, T.J. Dougherty, B.A. Dougherty, M.J. Pucci, Identification of 113 conserved essential genes using a high-throughput gene disruption system in *Streptococcus pneumoniae*, Nucleic Acids Res. 30 (14) (2002) 3152–3162.
[41] A.T. Vasconcelos, H.B. Ferreira, C.V. Bizarro, S.L. Bonatto, M.O. Carvalho, P.M. Pinto, D.F. Almeida, L.G. Almeida, R. Almeida, L. Alves-Filho, E.N. Assuncao, V.A. Azevedo, M.R. Bogo, M.M. Brigido, M. Brocchi, H.A. Burity, A.A. Camargo,

S.S. Camargo, M.S. Carepo, D.M. Carraro, J.C. de Mattos Cascardo, L.A. Castro, G. Cavalcanti, G. Chemale, R.G. Collevatti, C.W. Cunha, B. Dallagiovanna, B.P. Dambros, O.A. Dellagostin, C. Falcao, F. Fantinatti-Garboggini, M.S. Felipe, L. Fiorentin, G.R. Franco, N.S. Freitas, D. Frias, T.B. Grangeiro, E.C. Grisard, C.T. Guimaraes, M. Hungria, S.N. Jardim, M.A. Krieger, J.P. Laurino, L.F. Lima, M.I. Lopes, E.L. Loreto, H.M. Madeira, G.P. Manfio, A.Q. Maranhao, C.T. Martinkovics, S.R. Medeiros, M.A. Moreira, M. Neiva, C.E. Ramalho-Neto, M.F. Nicolas, S.C. Oliveira, R.F. Paixao, F.O. Pedrosa, S.D. Pena, M. Pereira, L. Pereira-Ferrari, I. Piffer, L.S. Pinto, D.P. Potrich, A.C. Salim, F.R. Santos, R. Schmitt, M.P. Schneider, A. Schrank, I.S. Schrank, A.F. Schuck, H.N. Seuanez, D.W. Silva, R. Silva, S.C. Silva, C.M. Soares, K.R. Souza, R.C. Souza, C.C. Staats, M.B. Steffens, S.M. Teixeira, T.P. Urmenyi, M.H. Vainstein, L.W. Zuccherato, A.J. Simpson, A. Zaha, Swine and poultry pathogens: the complete genome sequences of two strains of *Mycoplasma hyopneumoniae* and a strain of *Mycoplasma synoviae*, J. Bacteriol. 187 (2005) 5568–5577.

[42] I. Villarreal, T. Meyns, J. Dewulf, K. Vranckx, D. Calus, F. Pasmans, F. Haesebrouck, D. Maes, The effect of vaccination on the transmission of *Mycoplasma hyopneumoniae* in pigs under field conditions, Vet. J. 188 (1) (2011) 48–52.

[43] C. Volker, J.R. Brown, Bioinformatics and the discovery of novel anti-microbial targets, Curr. Drug Targets Infect. Disord. 2 (4) (2002) 279–290.

[44] C.S. Yu, Y.C. Chen, C.H. Lu, J.K. Hwang, Prediction of protein subcellular localization, Proteins 64 (2006) 643–651.

[45] R. Zhang, Y. Lin, DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes, Nucleic Acids Res. 37 (2009) D455–D458.

[46] R. Zhang, H.Y. Ou, C.T. Zhang, DEG, a database of essential genes, Nucleic Acids Res. 32 (2004) D271–D272.