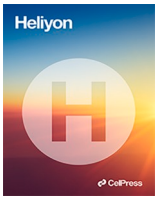
Heliyon 9 (2023) e18623

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Identification of vaccine candidates against *rhodococcus equi* by combining pangenome analysis with a reverse vaccinology approach

Lu Liu [a], Wanli Yu [a], Kuojun Cai [a], Siyuan Ma [a], Yanfeng Wang [a], Yuhui Ma [b], Hongqiong Zhao [a,*]

[a] College of Veterinary Medicine, Xinjiang Agricultural University, Urumqi 830052, Xinjiang, China
[b] Zhaosu Xiyu Horse Industry Co., Ltd. Zhaosu County 835699, Yili Prefecture, Xinjiang, China

ARTICLE INFO

Keywords:
Rhodococcus equi
Pangenome
Reverse vaccinology
Potential vaccine candidates
Epitopes

ABSTRACT

Rhodococcus equi (*R. equi*) is a zoonotic opportunistic pathogen that can cause life-threatening infections. The rapid evolution of multidrug-resistant *R. equi* and the fact that there is no currently licensed effective vaccine against *R. equi* warrant the need for vaccine development. Reverse vaccinology (RV), which involves screening a pathogen's entire genome and proteome using various web-based prediction tools, is considered one of the most effective approaches for identifying vaccine candidates. Here, we performed a pangenome analysis to determine the core proteins of *R. equi*. We then used the RV approach to examine the subcellular localization, host and gut flora homology, antigenicity, transmembrane helices, physicochemical properties, and immunogenicity of the core proteins to select potential vaccine candidates. The vaccine candidates were then subjected to epitope mapping to predict the exposed antigenic epitopes that possess the ability to bind with major histocompatibility complex I/II (MHC I/II) molecules. These vaccine candidates and epitopes will form a library of elements for the development of a polyvalent or universal vaccine against *R. equi*. Sixteen *R. equi* complete proteomes were found to contain 6,238 protein families, and the core proteins consisted of 3,969 protein families (~63.63% of the pangenome), reflecting a low degree of intraspecies genomic variability. From the pool of core proteins, 483 nonhost homologous membrane and extracellular proteins were screened, and 12 vaccine candidates were finally identified according to their antigenicity, physicochemical properties and other factors. These included four cell wall/membrane/envelope biogenesis proteins; four amino acid transport and metabolism proteins; one cell cycle control, cell division and chromosome partitioning protein; one carbohydrate transport and metabolism protein; one secondary metabolite biosynthesis, transport and catabolism protein; and one defense mechanism protein. All 12 vaccine candidates have an experimentally validated 3D structure available in the protein data bank (PDB). Epitope mapping of the candidates showed that 16 MHC I epitopes and 13 MHC II epitopes with the strongest immunogenicity were exposed on the protein surface, indicating that they could be used to develop a polypeptide vaccine. Thus, we utilized an analytical strategy that combines pangenome analysis and RV to generate a peptide antigen library that simplifies the development of multivalent or universal vaccines against *R. equi* and can be applied to the development of other vaccines.

* Corresponding author.
E-mail address: zhaohongqiong@sina.com (H. Zhao).

https://doi.org/10.1016/j.heliyon.2023.e18623
Received 27 April 2023; Received in revised form 12 July 2023; Accepted 24 July 2023
Available online 25 July 2023

1. Introduction

Rhodococcus equi (*R. equi*) is a gram-positive, aerobic, facultative intracellular actinobacterium that causes severe pulmonary and extrapulmonary pyogranulomatous infections in different animal species and humans [1]. *R. equi* shows a worldwide distribution and a high incidence, and long-term treatment and nursing costs and high mortality rates have led to considerable economic losses at farms where *R. equi* is endemic. *R. equi* infection has particularly severe outcomes in young foals (1–6 months of age) and immunocompromised persons. In foals, it typically manifests as a life-threatening purulent broncho pneumonic disease, and the mortality rate for *R. equi*-infected untreated foals ranges from 70% to 80% [2,3]. Even with treatment, ~30% mortality has been reported in *R. equi*-infected foals [4]. In addition, this disease affects the development of the equine racing industry, as foals that have recovered from pneumonia are less likely to race as adults. Notably, this infection also often occurs in immunosuppressed people. The mortality rate has been reported to range from 50% to 55% in HIV-infected individuals [5,6], threatening global public health security.

Prevention and control of *R. equi* infection is problematic for several reasons. First, *R. equi* is ubiquitous in the environment, and virulent isolates can be found at horse farms in the air, soil and feces, leading to environmental management challenges [7]. Second, early screening for disease has limited accuracy. The insidious progression of *R. equi* pneumonia results in marked pathology by the time clinical signs manifest. Consequently, treatment is generally prolonged, expensive and not always successful [3,8,9]. The standard treatment for *R. equi* pneumonia in foals is dual antimicrobial therapy with a combination of macrolide (erythromycin, azithromycin, or clarithromycin) and rifampin [10]. Over the past 20 years, the overuse of macrolides and rifampicin has led to the rapid evolution of multidrug-resistant (MDR) strains of *R. equi* [11,12]. In 2022, the European Food Safety Authority identified *R. equi* as one of the most relevant antimicrobial-resistant bacteria in the EU for horses [13]. These reports indicate that the prevention of R. equi infection rather than treatment is becoming critical.

Vaccines are the most effective way to control infection with pathogenic microorganisms, but there is currently no safe and effective vaccine against *R. equi*. Traditional types of vaccines, including live, killed and attenuated (physical and chemical) vaccines, have proven to be ineffective [14,15]. In addition, second-generation molecular-based vaccines, including DNA, genetically attenuated and subunit vaccines, have provided inadequate protection of foals [16,17]. The development of an effective universal vaccine depends on identifying and selecting common protective antigens and epitopes [18]. However, the pathogenic mechanism of *R. equi* is unclear; only virulence-associated proteins (Vaps) as immunodominant surface-expressed proteins have been widely investigated in vaccine development to prevent *R. equi* infection. Therefore, more potential antigens must be identified to develop a safe and effective protective vaccine against *R. equi*. With the ever-increasing volume of information on microbial genomes, the pangenome and reverse vaccinology (RV) are now attractive options for tackling vaccine development. RV is a predictive bioinformatics analytical approach that can be used to identify all potential protective antigens and epitopes in a pathogenic microbial genome or proteome [19]. This approach significantly reduces the time needed to develop a vaccine and provides reasonable targets that are identified via selection criteria. Since the development of a universal vaccine against serogroup *B Neisseria meningitidis* (*MenB*) in 2000 [20], RV has been extensively used to develop therapies against infections with many other organisms, such as *Acinetobacter baumannii* [21], *Shigella dysenteriae* [22], *Brucella spp.* [23], and *Mycobacteroides abscessus* [24]. Moreover, since *R. equi* strains demonstrate low intraspecies genomic variability, pangenome analysis may be a suitable strategy for analyzing the available complete genome/proteome of *R. equi* to explore potential protective antigens.

In this study, we performed pangenome analysis to identify conserved core proteins in *R. equi* strains of different host origins. We

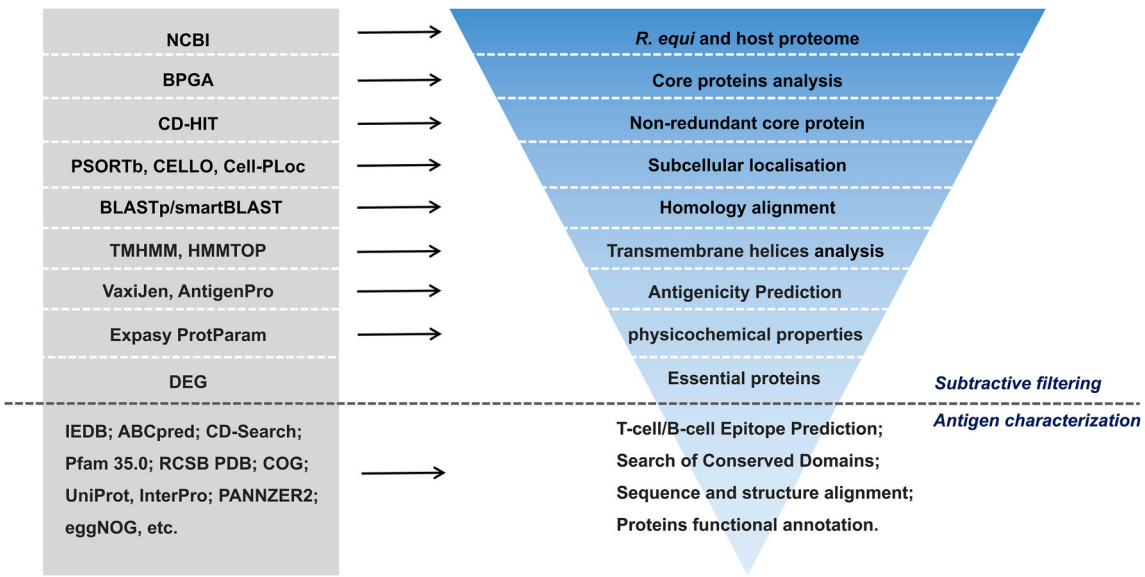


Fig. 1. Workflow of screening vaccine candidates for *R. equi* by pangenome analysis combined with reverse vaccinology.

used the RV approach to examine the subcellular localization, host homology, antigenicity, transmembrane helices, physicochemical properties and immunogenicity of the core proteins to select all potential vaccine candidates. The vaccine candidates were then subjected to epitope mapping to predict the exposed antigenic epitopes that possess the ability to bind with major histocompatibility complex I/II (MHC I/II) molecules. Based on these antigens and epitopes, the study will form a library of elements that can be utilized in the development of a polyvalent or universal vaccine against *R. equi*, reducing the time for researchers to develop a protective vaccine against *R. equi*.

2. Material and methods

The workflow diagram shown in Fig. 1 summarizes the protocol used for the identification of potential vaccine candidates via pangenome analysis and an RV approach.

2.1. R. equi genome/proteome data source

FASTA format complete genome/proteome sequences of 16 *R. equi* strains sequenced by PacBio were downloaded from the NCBI assembly database (https://www.ncbi.nlm.nih.gov/assembly/) (Accessed March 29, 2022). The accession numbers, strain names, host and other parameters are listed in Table A.1.

2.2. Pangenome analysis

The 16 *R. equi* strains selected for analysis were isolated from human, horse and environmental samples from four countries. To identify the core proteins, the 16 proteomes were analyzed using the Bacterial Pan-Genome Analysis tool (BPGA, v.1.3, Indian Institute of Chemical Biology, Kolkata) with the default parameters. In the input preparation for the clustering step, option number 4 (use any protein FASTA files) was chosen. To ensure fast and accurate clustering, USEARCH was used as the default protein clustering tool (identity cutoff = 0.5) [25]. The CD-HIT web server (http://www.bioinformatics.org/cd-hit/index.php) was used for core proteins sequence clustering, removing redundant (or highly similar) sequences (set threshold = 0.5), to obtain the nonredundant (nr) core proteins database.

2.3. Subcellular localization

Proteins localized on the membrane and in the extracellular space are considered good vaccine targets, as they can have direct contact with host cells and contain multiple antigenic determinants [26]. Core proteins with subcellular localization were analyzed using the PSORTb v3.0.2 (https://www.psort.org/psortb/index.html), CELLO v.2.5 (http://cello.life.nctu.edu.tw/) and Cell-PLoc 2.0 (http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/) servers. Proteins with the following predicted scores were excluded: PSORTb $\leq$9 and CELLO $\leq$2. Proteins that were predicted to have membrane and extracellular localization using at least two tools were selected for further analysis.

2.4. Host homology analysis

Nonhost homologous proteins were selected to avoid host autoimmunity [27]. Membrane and extracellular proteins were subjected to a host homology search using BLASTp (https://BLASTp.ncbi.nlm.nih.gov/BLASTp.cgi) with the following parameters: database = reference proteins (refseq_protein); organisms = *Homo sapiens* (GCF_000001405.40), *Equus caballus* (GCF_002863925.1), *Bos taurus* (GCF_002263795.2), *Ovis aries* (GCF_016772045.1) and *Sus scrofa* (GCF_000003025.6); and E-value cut off = 0.01. Antigens that showed >30% identity to any host protein and a bit score >100 were excluded.

2.5. Transmembrane helices

Proteins with multiple transmembrane helices are difficult to purify and hence are not considered efficient vaccine candidates [28]. The number of transmembrane helices in the nonhost homologous proteins was predicted using the TMHMM-2.0 (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0) and HMMTOP v.2.0 (http://www.enzim.hu/hmmtop/html/submit.html) tools, and proteins with <2 transmembrane helices were selected [29].

2.6. Protein antigenicity

Antigenicity was predicted using two tools: AntigenPro (http://scratch.proteomics.ics.uci.edu/index.html), which computes antigenicity based on amino acid sequence features, and VaxiJen v.2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html), which computes antigenicity based on the physicochemical properties of amino acid sequences [30,31]. Antigens with a prediction score of more than 0.5 for both tools are considered candidate antigens.

2.7. Protein physicochemical properties

The physicochemical properties of the vaccine candidates were analyzed using the Expasy ProtParam server (https://web.expasy.org/protparam/). The physicochemical properties analyzed were molecular weight (MW), theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY). Proteins with an MW of less than 110 kDa are considered good vaccine targets; however, proteins with an MW of less than 10 kDa are weakly antigenic [32]. Therefore, proteins in the MW range of 10 kDa to approximately 110 kDa were selected. Proteins with an aliphatic index >50 have good thermostability and were selected. Proteins with an instability index >40 are deemed unstable and were not selected.

2.8. Essential protein analysis

Essential proteins are indispensable for the survival of an organism, which makes them an attractive target for the development of effective vaccines and drugs [33]. The Database of Essential Genes (DEG, v.15, http://origin.tubic.org/deg/public/index.php/index) is a database of essential genes and proteins. To determine the criticality of the proteins, the candidate antigens were subjected to BLASTp with the DEGs (bit score >100, E-value cut off = 1e-10).

2.9. Similarity to gut flora proteins

Proteins significantly different from the host's gut flora proteins were selected to prevent the induced immune response from having any side effects on the host. Essential proteins were subjected to a gut flora homology search using SmartBLAST. Candidate antigens that showed >70% identity to any gut flora protein were excluded.

2.10. Functional annotation

To further characterize the vaccine candidates, in addition to noting the one-line annotations provided by the NCBI, we performed thorough manual annotations to determine the most likely biological functions of the selected vaccine candidates. For this, we used the following annotation servers: the Cluster of Orthologous Groups (COG) (https://www.ncbi.nlm.nih.gov/research/cog/), the Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://weizhongli-lab.org/metagenomic-analysis/server), InterPro (https://www.ebi.ac.uk/interpro/), Pannzer2 (http://ekhidna2.biocenter.helsinki.fi/sanspanz/) and eggNOG-mapper (http://eggnog-mapper.embl.de/). Furthermore, conserved domains were analyzed using the NCBI CD-search (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) and Pfam 35.0 (http://pfam.xfam.org/) servers [34].

2.11. Virulence factor database BLASTp

Virulence factors support bacterial adhesion, colonization and invasion within the host to promote disease progression. The Virulence Factor Database (VFDB) (http://www.mgc.ac.cn/cgi-bin/VFs/v5/main.cgi) is an extensive collection of known pathogenic bacterial virulence factors [35]. Consequently, VFDB (accessed September 3, 2022) was used for the identification of virulence proteins. Candidate antigens were subjected to a BLASTp (E-value cut off = 1e-10) with the VFDB to identify potential virulence proteins.

2.12. AntigenDB BLASTp

AntigenDB (https://webs.iiitd.edu.in/raghava/antigendb/antBLASTp.html) is a database of antigens that provides comprehensive information about a wide range of experimentally validated antigens [36]. To further characterize the candidate antigens, they were subjected to BLASTp with AntigenDB (E-value cut off = 1e-10).

2.13. Epitope mapping

B-cell and T-cell epitopes are highly immunodominant elements and can stimulate significant specific immune responses. The ABCPred server (https://webs.iiitd.edu.in/raghava/abcpred/index.html) was used (threshold value > 0.8) to predict linear B-cell epitopes in the potential vaccine candidates using artificial neural networks (65.93% accuracy) [37]. The resultant 20-mer B-cell epitopes were subsequently analyzed for T-cell epitopes for binding with MHC I and MHC II alleles using the IEDB analysis resource NetMHCpan EL 4.1 (http://tools.iedb.org/mhci/) and recommended 2.22 (http://tools.iedb.org/mhcii/) tools, respectively [38,39]. Epitopes with an MHC I pre-rank ≤0.2 and an MHC II pre-rank ≤10 were selected. The epitope density (ED) of each potential vaccine candidate was calculated by dividing the number of predicted epitopes by the protein length, and the potential vaccine candidates were further characterized by ED.

T-cell epitopes must have certain immunogenic features that enable them to stimulate either CD4 or CD8 T cells. Hence, MHC I immunogenicity (http://tools.iedb.org/immunogenicity/) and CD4 T-cell immunogenicity (http://tools.iedb.org/CD4episcore/) analysis tools from the IEDB analytical resources were further applied to predict the immunogenicity of the identified epitopes. Epitopes with an MHC I immunogenicity score >0 and a CD4 T-cell comprehensive score <60 were selected. Toxicity and antigenicity were also analyzed to shortlist MHC I and MHC II immunogenic epitopes. Vaxijen v2.0 was used to assess the antigenicity of the

epitopes (threshold value > 0.5). ToxinPred (http://crdd.osdd.net/raghava/toxinpred/) was used to screen for nontoxic epitopes.

2.14. Protein structure and comparative modeling

Comparative homology modeling using Swiss-Model (https://swissmodel.expasy.org/) was applied to determine the availability of the 3D structures of the potential vaccine candidates, and those with the highest sequence identity/similarity were selected. Exposure of an epitope on the protein surface facilitates the generation of a strong immune response. PyMOL (v.2.5, DeLano Scientific LLC), a user-sponsored molecular visualization system, was employed to reveal the topology of the epitopes in the protein structures.

The tools used in the computer analysis of the potential vaccine candidates are listed in Table A.2.

3. Results

3.1. Pangenome analysis of R. equi

The 16 *R. equi* proteomes were found to contain 6,238 protein families, and 3,992 core proteins were identified by BPGA (identity cutoff = 0.5). Among these core proteins, 3,969 nonredundant protein sequences were retrieved using CD-HIT (threshold = 50%), which is equivalent to 63.63% of the pangenome protein families and reflects a low degree of intraspecies genomic variability. According to the core- and pangenome fitting curve, the parameter 'b' = 0.123339, indicating that the pangenome is still open but may be closed soon. This means that the acquisition of new genes by the species will no longer affect the pangenome after some time (Fig. 2A). COG distribution analysis revealed that most of the core and accessory proteins are involved in metabolic regulation and that most of the unique proteins are involved in information storage and processing (Fig. 2B). Furthermore, KEGG functional classification revealed that most of the core, accessory and unique genes were involved in metabolic regulation (Figure A.1). The *R. equi* pangenome phylogenetic tree shows that the analyzed *R. equi* isolates are located at short genetic distances from each other, indicating strong intraspecies genetic relatedness (Fig. 2C).

3.2. Prioritization of potential vaccine candidates

3.2.1. Subcellular localization

Of the 3,969 core proteins identified, 2,410 proteins localize to the cytoplasm, 687 proteins localize to the membrane, 135 proteins are extracellular and 737 proteins have unknown localization. Membrane and extracellular proteins are higher-priority vaccine candidates due to their closer contact with host cells and hence their ability to initiate immune responses. Thus, 822 membrane and extracellular proteins were selected for further analysis (Table A.3).

3.2.2. Nonhost homology analysis

Host-homologous proteins were discarded because they can generate host autoimmunity. From the group of 822 proteins, 339

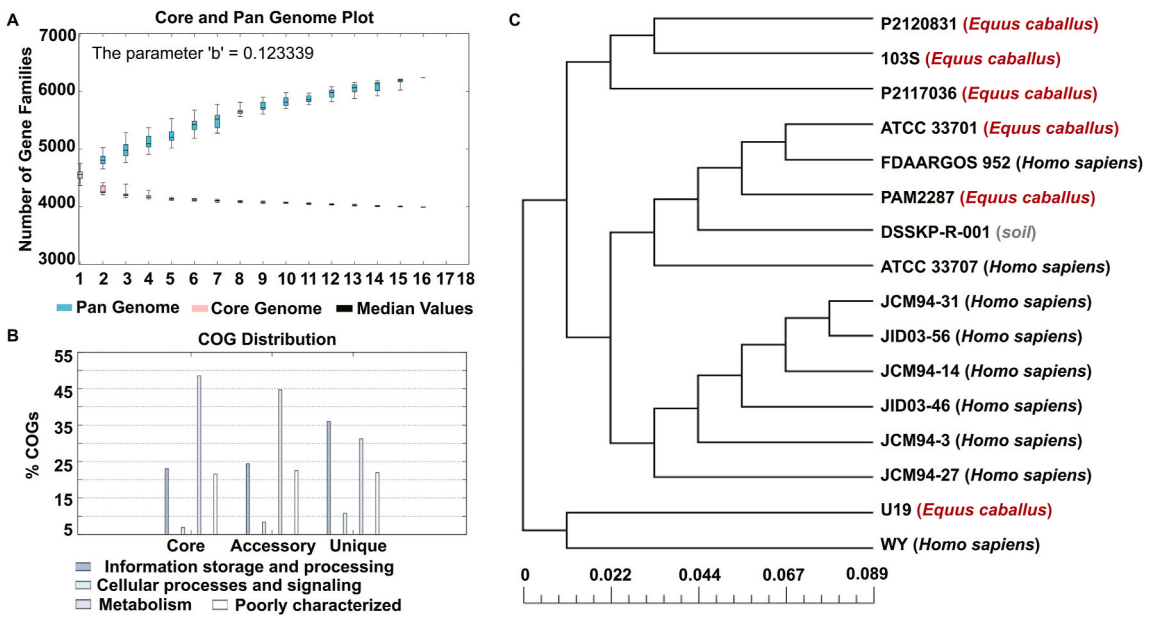


Fig. 2. Analysis of the *R. equi* core- and pan-genome. (A) Core- and pan-genome growth trends. (B) COG distribution of the core, accessory and unique proteins. (C) Phylogenetic tree of the pan-genome of 16 *R. equi* strains.

proteins with homology to *H. sapiens*, *E. caballus*, *B. taurus*, *O. aries* and *S. scrofa* were identified and removed. This left 483 nonhost homologous proteins (Table A.4).

3.2.3. Prediction of transmembrane helices

Among the 483 nonhost homologous proteins selected, 407 proteins were predicted by TMHMM v.2.0 and HMMTOP 2.0 to contain ≥ 2 transmembrane helices. These were then removed from the candidate protein list, leaving 76 proteins for further analysis.

3.2.4. Antigenicity analysis

These 76 proteins were then subjected to antigenicity analysis using the VaxiJen v.2.0 and AntigenPro tools (cutoff = 0.5). The results showed that 46 of the 76 proteins were more antigenic, and these were selected for further evaluation.

3.2.5. Physicochemical properties

The physicochemical properties of the 46 proteins were predicted using the Expasy ProtParam server. Selecting proteins with an MW of 10 kDa–~110 kDa, an aliphatic index >50 and an instability index <40 resulted in 29 proteins being identified for the subsequent analytical steps (Table A.5).

3.2.6. Essential protein analysis

From the 29 proteins obtained in the previous step, 12 essential proteins that aligned with DEG entries were retained (Fig. 3) (Table A.6).

3.2.7. Similarity to gut flora proteins

These 12 essential proteins were then subjected to a sequence similarity analysis using SmartBLAST, and all 12 proteins were determined to be potential vaccine candidates, as detailed in Table 1 (Table A.7).

3.3. Functional annotation of potential vaccine candidates

Functional analysis of proteins helps to understand their biological significance, including their physiological and biochemical activities and their roles in biological processes. According to the definition of COG protein function, the 12 potential vaccine candidates identified here fall into six different functional groups: Four are in the cell wall/membrane/envelope biogenesis group; four are in the amino acid transport and metabolism group; one is in the cell cycle control, cell division, chromosome partitioning group; one is in the carbohydrate transport and metabolism group; one is in the secondary metabolite biosynthesis, transport and catabolism group; and one is in the defense mechanisms group (Fig. 4). According to the definition of KEGG_ko protein function, these 12 proteins fall into four different functional categories: Five are in the cellular processes group, five are in the metabolism group, one is in the glycan biosynthesis and metabolism group and one is in the unclassified group (Fig. 4). Furthermore, when the 12 proteins were manually annotated using a conservative domain search and various protein annotation tools, it was found that most of the 12 proteins were associated with bacterial adhesion, invasion and evasion (Table 2).

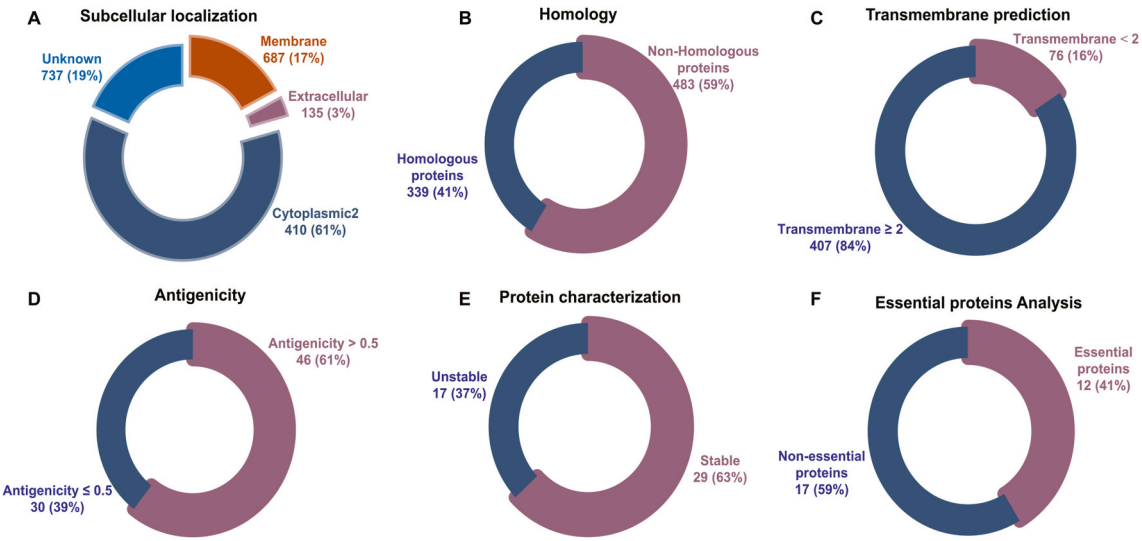


Fig. 3. Prioritization of the potential vaccine candidates for *R. equi*. (A) Subcellular localization analysis of core proteins. (B) Host homology analysis of membrane proteins and extracellular proteins. (C) Prediction of transmembrane helix of nonhost homologous proteins. (D) Antigenicity analysis of proteins with transmembrane helix<2. (E) Physicochemical properties of proteins with Antigenicity>0.5. (F) Essential proteins screened by the DEG database.

Table 1
Information about the potential vaccine candidates for *R. equi*.

Accession Number	Length (aa)	MW (kDa)	Localization	PredHel (TMH)	PredHel (HMM)	Antigenicity (VaxiJen)	Antigenicity (AntigenPro)	Instability index	In vivo half-life (h)	In vitro half-life (h)
WP_013414512.1	629	67.517	Extra	1	1	0.5518	0.886341	29.49	> 20 (yeast) > 10 (*E. coli*)	30 (mammalian reticulocytes)
WP_013414996.1	537	57.327	Extra	0	1	0.5134	0.860102	29.16	> 20 (yeast) > 10 (*E. coli*)	30 (mammalian reticulocytes)
WP_013417558.1	534	57.163	Extra	0	1	0.5714	0.601849	31.91	> 20 (yeast) > 10 (*E. coli*)	30 (mammalian reticulocytes)
WP_013417385.1	525	55.187	Extra	1	1	0.6264	0.894264	28.54	> 20 (yeast) > 10 (*E. coli*)	30 (mammalian reticulocytes)
WP_013415902.1	524	55.596	Extra	0	1	0.5516	0.610999	25.21	> 20 (yeast) > 10 (*E. coli*)	30 (mammalian reticulocytes)
WP_013414386.1	495	51.090	Extra/ Memb	1	1	0.6319	0.952852	31.42	> 20 (yeast) > 10 (*E. coli*)	30 (mammalian reticulocytes)
WP_013416802.1	384	38.111	Extra	0	0	0.5418	0.637049	33.54	> 20 (yeast) > 10 (*E. coli*)	30 (mammalian reticulocytes)
WP_013415053.1	361	38.024	Extra	1	1	0.5315	0.906752	16.67	> 20 (yeast) > 10 (*E. coli*)	30 (mammalian reticulocytes)
WP_013415127.1	277	28.188	Extra	0	0	0.6976	0.838859	37.50	> 20 (yeast) > 10 (*E. coli*)	30 (mammalian reticulocytes)
WP_044990738.1	538	56.807	Extra	0	1	0.6378	0.740046	27.45	> 20 (yeast) > 10 (*E. coli*)	30 (mammalian reticulocytes)
WP_080668398.1	421	44.154	Extra	0	0	0.5508	0.846806	29.42	> 20 (yeast) > 10 (*E. coli*)	30 (mammalian reticulocytes)
WP_022596925.1	380	38.189	Extra/ Memb	0	0	0.6907	0.819534	36.83	> 20 (yeast) > 10 (*E. coli*)	30 (mammalian reticulocytes)

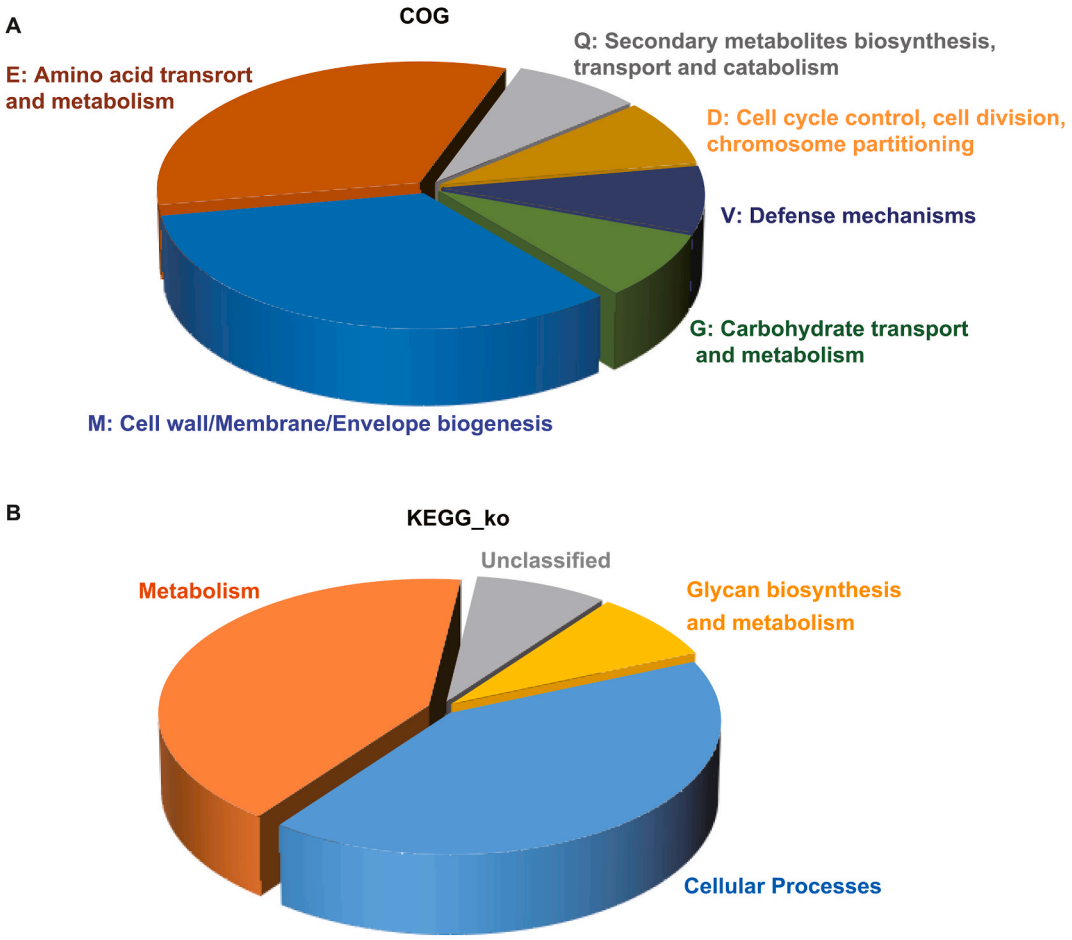


Fig. 4. Functional annotation of potential vaccine candidates. The categorisation of the 12 potential vaccine candidates according to the COG (A) and KEGG_ko (B) protein function definitions.

AntigenDB provides comprehensive information about a wide range of experimentally validated antigens. The AntigenDB BLASTp results showed that of the 12 candidates, six are homologous to experimentally validated antigens, including alpha/beta hydrolase-fold protein (WP_013414512.1), penicillin-binding protein 2 (WP_013414386.1), NlpC/P60 family protein (WP_013416802.1), esterase family protein (WP_013415053.1), M23 family metallopeptidase (WP_013415127.1) and serine hydrolase (WP_080668398.1).

The VFDB contains comprehensive information about the virulence factors of bacterial pathogens. The VFDB BLASTp results showed that of the 12 proteins, four are homologous to reported virulence factors: Alpha/beta hydrolase-fold protein (WP_013414512.1) and esterase family protein (WP_013415053.1) are homologous to Antigen 85 (FbpA/B/C) of *Mycobacterium tuberculosis* (*M. tb*), NlpC/P60 family protein (WP_013416802.1) is homologous to PPE family proteins of *M. tb*, and M23 family metallopeptidase (WP_013415127.1) is homologous to LasA of *Pseudomonas aeruginosa*. These four virulence-related proteins are also experimentally validated antigens and thus may have greater potential as vaccine candidates than other proteins (Table 2).

3.4. Epitope mapping and structural analysis of potential vaccine candidates

Comparative homology modeling using Swiss-Model was applied to determine the availability of the 3D structure of each potential vaccine candidate. All 12 proteins have an experimentally validated 3D structure available in the protein data bank (PDB) (sequence identity >35%). A set of 20-mer B-cell epitopes was predicted for each protein, and the B-cell epitopes were screened for the presence of T-cell epitopes that can bind to both MHC I and MHC II alleles. In total, 176 B-cell epitopes, 1,464 MHC I epitopes (9-mer) and 1,490 MHC II epitopes (15-mer) were found within the 12 potential vaccine candidates (Table A.8). The overall ED of each of the 12 candidates was calculated. The four candidates that aligned with entries in the VFDB and AntigenDB were also found to have the highest ED values (Table 3).

Further filtering resulted in the selection of epitopes with an antigenicity score >0.5, an MHC I epitope immunogenicity score >0 and an MHC II epitope immunogenicity score <60. Subsequently, the T-cell epitopes in each candidate that have the potential to

Table 2
Functional annotations of the potential vaccine candidates.

Accession number	One-line annotation (NCBI)	Conserved domain (CD-search)	Description (Pfam)	Description (Pannzer2)	COG category	KEGG_ko category	Function annotation (eggNOG, Pannzer, InterPro)	VFDB blast	AntigenDB blast
WP_013414512.1	Alpha/beta hydrolase-fold protein	FrmB	Esterase, LGFP	Antigen 85-A	M: Cell wall/ Membrane/ Envelope biogenesis	Metabolism	GO:0050348 GO:0016787 GO:0004144	Antigen 85 (FbpA/B/C) [*M. tb*]	Antigen 85-A/B/C [*M. tb*]
WP_013414996.1	ABC transporter substrate-binding protein	DdpA	SBP_bac_5	Putative high affinity substrate-binding lipoprotein	E: Amino acid transport and metabolism	Cellular processes	GO:0055085 GO:0043190	–	–
WP_013417558.1	ABC transporter substrate-binding protein	DdpA	SBP_bac_5	Peptide/nickel transport system substrate-binding protein	E: Amino acid transport and metabolism	Cellular processes	GO:0055085 GO:0043190	–	–
WP_013417385.1	ABC transporter substrate-binding protein	DdpA	SBP_bac_5	Putative oligopeptide/ dipeptide ABC transporter substrate-binding lipoprotein	E: Amino acid transport and metabolism	Cellular processes	GO:0055085 GO:0043190	–	–
WP_013415902.1	ABC transporter substrate-binding protein	DdpA	SBP_bac_5	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	E: Amino acid transport and metabolism	Cellular processes	GO:0055085 GO:0043190	–	–
WP_013414386.1	Penicillin-binding protein 2	FtsI	Transpeptidase	Penicillin-binding protein 2	M: Cell wall/ Membrane/ Envelope biogenesis	Glycan biosynthesis and metabolism	GO:0051301 GO:0008658 GO:0008955	–	Penicillin-binding protein 2 [*Staphylococcus aureus*]
WP_013416802.1	NlpC/P60 family protein	Spr	DUF4226, NLPC_P60	Glycoside hydrolase	M: Cell wall/ Membrane/ Envelope biogenesis	Metabolism	GO:0016787	p60 (iap/cwhA) [*Listeria monocytogenes*]	Protein p60 [L. monocytogenes] PPE family protein [*M. tb*]
WP_013415053.1	Esterase family protein	FrmB	Esterase	Antigen 85-A	Q: Secondary metabolites biosynthesis, transport and catabolism	Metabolism	GO:0016787 GO:0016746	Antigen 85 (FbpA/B/C) [*M. tb*]	Antigen 85A/B/C [*M. tb*]
WP_013415127.1	M23 family metallopeptidase	NlpD	Peptidase_M23	Putative metallopeptidase	D: Cell cycle control, cell division, chromosome partitioning	Metabolism	GO:0016787 GO:0016021	LasA (lasA) [*Pseudomonas aeruginosa*]	Proline-rich 28 kDa antigen [*M. tb*]
WP_044990738.1	ABC transporter substrate-binding protein	DdpA	SBP_bac_5	Putative oligopeptide/ dipeptide ABC transporter substrate-binding lipoprotein	M: Cell wall/ Membrane/ Envelope biogenesis	Cellular processes	GO:0055085 GO:0043190	–	–
WP_080668398.1	Serine hydrolase	AmpC	Beta-lactamase	D-alanyl-D-alanine carboxypeptidase	V: Defense mechanisms	Metabolism	GO:0032259 GO:0043412 GO:0044260 GO:0044238 GO:0004180 GO:0070008 GO:0008168	–	Lipoprotein lpqH/ lprG [*M. tb*]
WP_022596925.1	PQQ-dependent sugar dehydrogenase	YliI superfamily	GSDH	Glucose/sorbosone dehydrogenase, lipoprotein LppZ	G: Carbohydrate transport and metabolism	Unclassified	GO:0008876	–	–

bind to the maximum number of MHC I and II class alleles were selected. Finally, 16 MHC I epitopes and 13 MHC II epitopes were selected from 1,464 MHC I epitopes and 1,490 MHC II epitopes, respectively (Table A.7). PyMOL was employed to reveal the topology of the epitopes within the protein structures. These epitopes satisfactorily display surface exposure and thus can be exploited as targets for vaccines against *R. equi* (Fig. 5).

4. Discussion

In the absence of efficacious and safe antibiotics, vaccination represents the best strategy to combat MDR pathogens; however, no licensed commercial vaccine has been made available against *R. equi* to date. In addition, all candidate *R. equi* vaccines under development have been found to have limitations due to instability, a limited protection period or reactogenicity [21–26]. This evidence strongly suggests that there is a need to search for additional vaccine targets in *R. equi* genomes that would achieve effective long-lasting protection. In this study, a pangenome analysis and RV approach were adapted to identify potential vaccine candidates for *R. equi*.

4.1. Methodology

The identification of antigens that are common to several strains can contribute to the development of effective universal or polyvalent vaccines [17,18]. Additionally, analysis of a single genome does not consider the influence of genetic variability on a bacterial species. For these reasons, pangenome analysis was used to identify core proteins shared by different *R. equi* strains that have potential as vaccine candidate antigens. *R. equi* has a low degree of intraspecies genomic variability [25]. Therefore, pangenome analysis of the *R. equi* genomes or proteomes available in the NCBI database can be used to accurately screen for universal candidate antigens. The core proteins were filtered to screen for potential vaccine candidates. Since all vaccine candidates are antigens, but not all antigens are effective vaccine candidates for various reasons, we applied an RV approach and used multiple physicochemical properties and immunogenicity tools to identify high-priority vaccine candidates. For example, membrane and extracellular proteins are appropriate vaccine targets due to their close proximity and contact with host cells and hence are more likely to stimulate an immune response [26]. Proteins with homology to host proteins induce an autoimmune response, and thus, such proteins were eliminated from the list of preselected proteins [27]. Antigenicity is another important feature of a potential vaccine candidate; therefore, nonhost homology proteins with antigenic probability values > 0.5 were considered for the development of potential vaccine candidates in this study [30,31]. Proteins with multiple transmembrane helices, with an MW > 110 kDa and that are unstable were discarded because they are difficult to clone, express and purify [29]. Furthermore, essential genes are indispensable for the survival of an organism, and are therefore considered effective therapeutic and vaccine targets [32–39]. Based on all of these criteria, the selected proteins had <2 helices and an MW in the required range and were essential for the bacteria to survive. Our analytical strategy integrated pangenome analysis of proteome sequences and an RV approach that utilized web-based prediction tools, which resulted in rational computational predictions that allowed us to generate a manageable list of vaccine targets. This strategy significantly reduces the time needed to develop a vaccine and provides reasonable vaccine targets that are identified based on selection criteria. Similar strategies have been implemented by Hassan et al., Shahid et al. and Jalal et al. [21,22]. Finally, we identified 12 vaccine candidates from 3,969 core proteins.

As we aimed to identify universal antigens among isolates obtained from various sources, it is possible that we could have missed some strain-specific antigens. In fact, one important vaccine candidate is missing from our list that is recognized as the most promising vaccine candidate for *R. equi*: Vaps. All virulent strains of *R. equi* contain a plasmid that encodes Vaps. Interestingly, plasmids derived from equine isolates are of the pVAPA type, whereas those from swine strains are typically of the pVAPB type, and plasmids obtained from bovine isolates are primarily of the pVAPN type. Human *R. equi* isolates carry either a pVAPA- or pVAPB-type plasmid or no plasmid at all [40]. The Vaps sequence is not highly conserved among the different plasmid types [25], which this may explain why Vaps was filtered out and not included in our pangenome or RV analysis. Bacterial protective antigens contain polysaccharides, which are often not included in predictions. In addition, only targets that are expressed in the organism are of interest. However, the RV target

Table 3
The epitope density (ED) of the 12 potential vaccine candidates.

Accession number	One-line Annotation (NCBI)	PDB >35%	Length (aa)	B-cell ED	MHC I ED	MHC IIED	Overall ED
WP_013414512.1	Alpha/beta hydrolase-fold protein	6SX4	629	0.035	0.184	0.245	0.464
WP_013414996.1	ABC transporter substrate-binding protein	5U4O	537	0.034	0.317	0.304	0.654
WP_013417558.1	ABC transporter substrate-binding protein	4OET	534	0.030	0.352	0.243	0.625
WP_013417385.1	ABC transporter substrate-binding protein	5U4O	525	0.027	0.248	0.310	0.585
WP_013415902.1	ABC transporter substrate-binding protein	3M8U	524	0.034	0.269	0.296	0.599
WP_013414386.1	Penicillin-binding protein 2	3UPP	495	0.034	0.364	0.313	0.711
WP_013416802.1	NlpC/P60 family protein	4HPE	384	0.023	0.151	0.141	0.315
WP_013415053.1	Esterase family protein	4H18	361	0.039	0.343	0.288	0.670
WP_013415127.1	M23 family metallopeptidase	5Ja1L	277	0.022	0.375	0.152	0.549
WP_044990738.1	ABC transporter substrate-binding protein	6HLX	538	0.028	0.214	0.290	0.532
WP_080668398.1	Serine hydrolase	4Y7P	421	0.033	0.166	0.295	0.494
WP_022596925.1	PQQ-dependent sugar dehydrogenase	3DAS	380	0.034	0.179	0.237	0.450

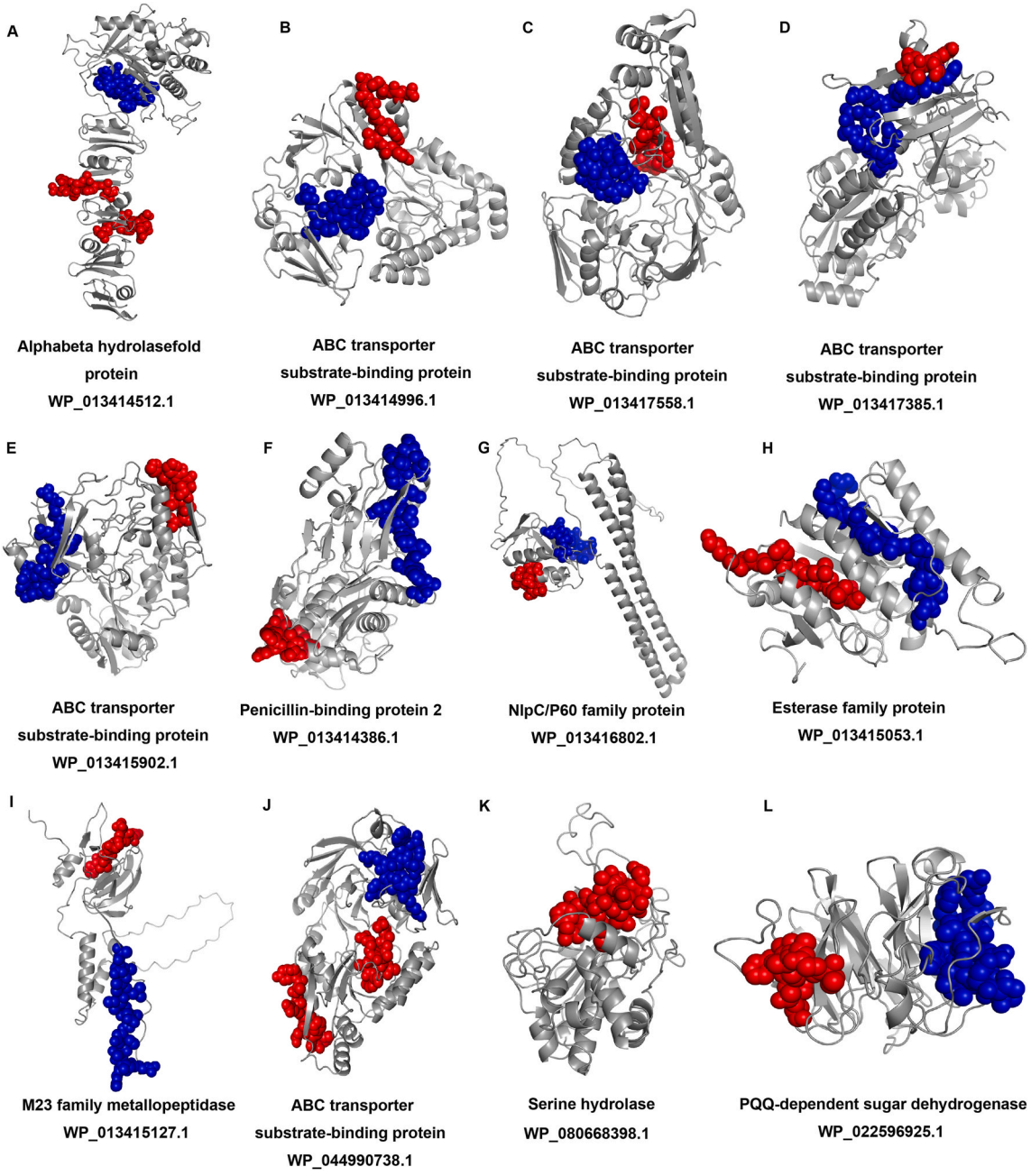


Fig. 5. The 3D structures and epitope mapping of the potential vaccine candidates. The Swiss-Model was used to perform comparative homology modeling of the 12 potential vaccine candidates (A–L). The 9-mer MHC I epitopes are shown as red spheres on the grey-coloured protein structure, and the 15-mer MHC II epitopes are shown as blue spheres on the grey-coloured protein structure.

screening methodology is based on the identification of pathogen genomes and amino acid sequences translated from genomes and cannot be used to determine whether potential targets are actually expressed. Therefore, further experimental validation is needed to test the function and immunogenicity of these proteins identified using this approach.

4.2. Potential vaccine candidates

All the proteins in the selected group of 12 are membrane and extracellular proteins, are nonhost homologous, have stable physicochemical properties, are essential for the survival of the microorganism and are immunogenic in nature. In this section, these proteins are further described.

The first group of proteins consists of five ATP-binding cassette (ABC) transporter substrate-binding family proteins (WP_013414996.1, WP_013417558.1, WP_013417385.1, WP_013415902.1 and WP_044990738.1). ABC transporters are essential for the uptake of nutrients and other molecules that cannot be produced by *R. equi* [41]. In addition, ABC transporters play key roles in the pathogenesis of infection, facilitating pathogenic mechanisms such as biofilm formation, adherence and invasion of host cells, intracellular survival and nasopharyngeal colonization [42]. It has been reported that ABC transporters can induce immunity; hence, they are potential vaccine and drug targets [43].

The second group of proteins is *M. tb* virulence factor homologs. They include the Fbp (A/B/C)/antigen 85 (Ag85) homologs alpha/beta hydrolase-fold protein (WP_013414512.1) and esterase family protein (WP_013415053.1) and the PE/PPE protein homolog NlpC/P60 family protein (WP_013416802.1). In *M. tb* infections, high levels of Ag85 are secreted, eliciting strong humoral and cellular immune responses, and PE/PPE proteins mediate the immune responses via cell-surface adhesion or receptor binding. These proteins have been intensely investigated as vaccines against *M. tb* in animal models [44,45]. *R. equi* and *M. tb* are both actinomycetes, and the Ag85 and PE/PPE homologs are also considered potential virulence-associated determinants of *R. equi* [46]. Targeting these proteins is a promising strategy for developing effective *R. equi* vaccines.

Proteins in the third group are associated with antibiotic resistance and include penicillin-binding protein 2 (PBP2) (WP_013414386.1) and serine hydrolase (WP_080668398.1). PBP2 plays a key role in building the cell wall in several bacteria by catalyzing the biosynthesis of peptidoglycan, and it is also involved in bacterial resistance to β-lactams [47]. In addition, Morgan et al. have shown that PBP2 displays immunogenic properties [48]. The AmpC conserved domain of serine hydrolase is also involved in bacterial resistance to β-lactams, and AmpC expression levels affect cellular fitness and virulence [49]. Targeting this protein that confers β-lactam resistance would have a severe effect on bacterial colonization and survival.

M23 family metallopeptidase (WP_013415127.1) is a peptidoglycan hydrolase that plays an important role in cell growth and division and bacterial virulence [50]. The NlpD lipoprotein conserved domain has been shown to be required for cell division and is also an essential virulence factor [51]. Furthermore, mutant NlpD has proven to be a superior vaccine candidate by providing effective immunity, indicating that bacterial M23 peptidase family members may be used as vaccine targets to prevent infection [52].

PQQ-dependent sugar dehydrogenase (WP_022596925.1) is an oxidoreductase that belongs to the YliI superfamily. Its main functions are carbohydrate transport and metabolism [53]. It is an *M. tb* lipoprotein Z (LppZ) homolog. Chen et al. demonstrated that LppZ exhibits strong immunogenicity during *M. tb* infection in both humans and mice and has the ability to trigger effective innate and cellular immunity [54]. Therefore, it is worthwhile to explore LppZ as a potential vaccine candidate for *R. equi*.

Ideal vaccines induce both innate immunity and humoral and cell-mediated adaptive immunity. It has been proven that proteins with high ED have stronger immunogenicity than those with low ED [55]. However, most of the currently available web-based prediction tools used in RV are not adaptable to equine vaccinology. For example, ProPred-I/II, NetMHC-4.0 and NetMHCII-2.3 are used to predict epitopes, but they are not specifically trained to predict epitopes that bind to equine MHC alleles; rather, they are trained to predict epitopes that bind to human and other animal MHC alleles, such as those of mice. In the absence of equine MHC alleles to predict epitopes, epitope predictions were conducted using human alleles, which have a high identity with equine MHC alleles [56,57]. In this study, 16 MHC I (9-mer) and 13 MHC II (15-mer) high-affinity epitopes were identified in the 12 potential vaccine candidates, and these could be included in a polyvalent vaccine. Epitope mapping and protein 3D structure analysis showed that all the predicted epitopes were exposed on the protein surface. This suggests that these epitopes have a high potential of being recognized by equine MHC alleles and eliciting a humoral and cellular immunity response in equines. Among the 12 potential vaccine candidates, four proteins (WP_013414386.1, WP_013415053.1, WP_013414996.1 and WP_013415127.1) are the most interesting vaccine candidates because they are predicted to have high ED and strong immunogenicity. Further investigation is needed to understand the functions of these proteins and their potential as protective antigens.

5. Conclusions

Here, we provide an analytical strategy that combines pangenome analysis with an RV approach. This strategy significantly reduces the time needed to develop a vaccine and provides reasonable vaccine targets that are identified based on selection criteria. Using this strategy, we generated a manageable list of 12 potential vaccine candidates and 29 antigenic epitopes of *R. equi*. These proteins are strongly associated with bacterial survival and virulence. Further experimental validation is needed to test the immunogenicity and protection level of these proteins.

Author contribution statement

Lu Liu: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.
Wanli Yu: Analyzed and interpreted the data; Wrote the paper.
Kuojun Cai: Analyzed and interpreted the data.
Siyuan Ma: Yanfeng Wang: Contributed reagents, materials, analysis tools or data.
Yuhui Ma: Performed the experiments; Contributed reagents, materials, analysis tools or data.
Hongqiong Zhao: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Data availability statement

Data will be made available on request.

Funding

This work was supported by the Xinjiang Uygur Autonomous Region Regional Collaborative Innovation Project-Shanghai Cooperation Organization Science and Technology Partnership Program and International Science and Technology Cooperation Program [2022E01027], and Xinjiang Uygur Autonomous Region Graduate Student Innovation Program [XJ2023G125].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e18623.

References

[1] J.A. Vázquez-Boland, W.G. Meijer, The pathogenic actinobacterium *Rhodococcus equi*: what's in a name? Mol. Microbiol. 112 (2019) https://doi.org/10.1111/mmi.14267.
[2] S. Takai, S. Ohbushi, K. Koike, S. Tsubaki, H. Oishi, M. Kamada, Prevalence of virulent *Rhodococcus equi* in isolates from soil and feces of horses from horse-breeding farms with and without endemic infections, J. Clin. Microbiol. 29 (1991), https://doi.org/10.1128/jcm.29.12.2887-2889.1991.
[3] G. Muscatello, *Rhodococcus equi* pneumonia in the foal - Part 1: pathogenesis and epidemiology, Vet. J. 192 (2012), https://doi.org/10.1016/j.tvjl.2011.08.014.
[4] S. Giguère, S. Jacks, G.D. Roberts, J. Hernandez, M.T. Long, C. Ellis, Retrospective comparison of azithromycin, clarithromycin, and erythromycin for the treatment of foals with *Rhodococcus equi* pneumonia, J. Vet. Intern. Med. 18 (4) (2004) 568–573, https://doi.org/10.1892/0891-6640(2004)18<568:rcoaca>2.0.co;2.
[5] D. Speck, I. Koneth, M. Diethelm, I. Binet, A pulmonary mass caused by *Rhodococcus equi* infection in a renal transplant recipient, Nat. Clin. Pract. Nephrol. 4 (2008), https://doi.org/10.1038/ncpneph0833.
[6] J.C. Sunstrum, J.C. Sunstrum, *Rhodococcus equi* infection in patients with and without human immunodificiency virus infection, Rev. Infect. Dis. 13 (1990), https://doi.org/10.1093/clinids/13.1.139.
[7] S. Petry, C. Sévin, M.A. Fleury, F. Duquesne, N. Foucher, C. Laugier, M. Henry-Amar, J. Tapprest, Differential distribution of vapA-positive *Rhodococcus equi* in affected and unaffected horse-breeding farms, Vet. Rec. (2017) 181, https://doi.org/10.1136/vr.104088.
[8] M.K. Chaffin, N.D. Cohen, G.P. Blodgett, M. Syndergaard, Evaluation of hematologic screening methods for early detection of *Rhodococcus equi* pneumonia in foals, J. Equine Vet. Sci. 32 (2012) S20, https://doi.org/10.1016/j.jevs.2012.08.049.
[9] H.T. Vechi, E.T.G. de Oliveira, M.R. de Freitas, F. Rossi, M.H.M.F. de Britto, M.D.M. Alves, Chronic cavitary pneumonia by *Rhodococcus equi* in a highly prevalent tuberculosis country: a diagnosis challenge, Rev. Inst. Med. Trop. Sao Paulo 60 (2018), https://doi.org/10.1590/S1678-9946201860074.
[10] S. Giguère, Treatment of Infections Caused by *Rhodococcus equi*, Veterinary Clinics of North America - Equine Practice 33 (2017). https://doi.org/10.1016/j.cveq.2016.11.002.
[11] H. Liu, Y. Wang, J. Yan, C. Wang, H. He, Appearance of multidrug-resistant virulent *Rhodococcus equi* clinical isolates obtained in China, J. Clin. Microbiol. 52 (2014), https://doi.org/10.1128/JCM.02925-13.
[12] S. Álvarez-Narváez, S. Giguère, N. Cohen, N. Slovis, J.A. Vázquez-Boland, Spread of multidrug-resistant *Rhodococcus equi*, United States, Emerg. Infect. Dis. 27 (2021), https://doi.org/10.3201/eid2702.203030.
[13] S.S. Nielsen, D.J. Bicout, P. Calistri, E. Canali, J.A. Drewe, B. Garin-Bastuji, J.L. Gonzales Rojas, C. Gortázar, M. Herskin, V. Michel, M.Á. Miranda Chueca, B. Padalino, P. Pasquali, H.C. Roberts, H. Spoolder, K. Ståhl, A. Velarde, A. Viltrop, C. Winckler, F. Baldinelli, A. Broglia, L. Kohnle, J. Alvarez, Assessment of listing and categorisation of animal diseases within the framework of the Animal Health Law (Regulation (EU) No 2016/429): antimicrobial-resistant *Rhodococcus equi* in horses, EFSA J. 20 (2022), https://doi.org/10.2903/j.efsa.2022.7081.
[14] J.N. Rocha, N.D. Cohen, A.I. Bordin, C.N. Brake, S. Giguère, M.C. Coleman, R.C. Alaniz, S.D. Lawhon, W. Mwangi, S.D. Pillai, Oral administration of electron-beam inactivated *rhodococcus equi* failed to protect foals against intrabronchial infection with live, virulent R. Equi, PLoS One 11 (2016), https://doi.org/10.1371/journal.pone.0148111.
[15] A.E. Whitehead, V.R. Parreira, J. Hewson, J.L. Watson, J.F. Prescott, Development of a live, attenuated, potential vaccine strain of *R. equi* expressing vapA and the virR operon, and virulence assessment in the mouse, Vet. Immunol. Immunopathol. 145 (2012) 479–484, https://doi.org/10.1016/j.vetimm.2011.10.011.
[16] M.M. Trevisani, E.S. Hanna, A.F. Oliveira, S.A. Cardoso, M.C. Roque-Barreira, S.G. Soares, Vaccination of mice with virulence-associated protein G (VapG) antigen confers partial protection against *Rhodococcus equi* infection through induced humoral immunity, Front. Microbiol. 8 (2017), https://doi.org/10.3389/fmicb.2017.00857.
[17] K.L. Lohmann, A. Marianela Lopez, S.T. Manning, F.J. Marques, R. Brownlie, A.L. Allen, A.E. Sangster, G. Mutwiri, V. Gerdts, A. Potter, H.G.G. Townsend, Failure of a VapA/CpG oligodeoxynucleotide vaccine to protect foals against experimental *Rhocococcus equi* pneumonia despite induction of VapA-specific antibody and interferon-γ response, Can. J. Vet. Res. 77 (2013).
[18] R. Servín-Blanco, R. Zamora-Alvarado, G. Gevorkian, K. Manoutcharian, Antigenic variability: obstacles on the road to vaccines against traditionally difficult targets, Hum. Vaccines Immunother. 12 (2016), https://doi.org/10.1080/21645515.2016.1191718.
[19] S. Vivona, J.L. Gardy, S. Ramachandran, F.S.L. Brinkman, G.P.S. Raghava, D.R. Flower, F. Filippini, Computer-aided biotechnology: from immuno-informatics to reverse vaccinology, Trends Biotechnol. 26 (2008), https://doi.org/10.1016/j.tibtech.2007.12.006.
[20] M. Pizza, V. Scarlato, V. Masignani, M.M. Giuliani, B. Aricò, M. Comanducci, G.T. Jennings, L. Baldi, E. Bartolini, B. Capecchi, C.L. Galeotti, E. Luzzi, R. Manetti, E. Marchetti, M. Mora, S. Nuti, G. Ratti, L. Santini, S. Savino, M. Scarselli, E. Storni, P. Zuo, M. Broeker, E. Hundt, B. Knapp, E. Blair, T. Mason, H. Tettelin, D. W. Hood, A.C. Jeffries, N.J. Saunders, D.M. Granoff, J.C. Venter, E.R. Moxon, G. Grandi, R. Rappuoli, Identification of vaccine candidates against serogroup *B meningococcus* by whole-genome sequencing, Science (1979) (2000) 287, https://doi.org/10.1126/science.287.5459.1816.
[21] F. Shahid, T. Zaheer, S.T. Ashraf, M. Shehroz, F. Anwer, A. Naz, A. Ali, Chimeric vaccine designs against *Acinetobacter baumannii* using pan genome and reverse vaccinology approaches, Sci. Rep. 11 (2021), https://doi.org/10.1038/s41598-021-92501-8.
[22] K. Jalal, T. Abu-Izneid, K. Khan, M. Abbas, A. Hayat, S. Bawazeer, R. Uddin, Identification of vaccine and drug targets in *Shigella dysenteriae* sd197 using reverse vaccinology approach, Sci. Rep. 12 (2022), https://doi.org/10.1038/s41598-021-03988-0.
[23] X. Zai, Y. Yin, F. Guo, Q. Yang, R. Li, Y. Li, J. Zhang, J. Xu, W. Chen, Screening of potential vaccine candidates against pathogenic *Brucella spp.* using compositive reverse vaccinology, Vet. Res. 52 (2021), https://doi.org/10.1186/s13567-021-00939-5.

[24] H.A. Dar, S. Ismail, Y. Waheed, S. Ahmad, Z. Jamil, H. Aziz, H.F. Hetta, K. Muhammad, Designing a multi-epitope vaccine against *Mycobacteroides* abscessus by pangenome-reverse vaccinology, Sci. Rep. 11 (2021), https://doi.org/10.1038/s41598-021-90868-2.
[25] N.M. Chaudhari, V.K. Gupta, C. Dutta, BPGA-an ultra-fast pan-genome analysis pipeline, Sci. Rep. 6 (2016), https://doi.org/10.1038/srep24373.
[26] G. Grandi, Bacterial surface proteins and vaccines, F1000, Biol. Rep. (Wash. D C) 2 (2010), https://doi.org/10.3410/B2-36.
[27] B. Shanmugham, A. Pan, Identification and characterization of potential therapeutic candidates in emerging human pathogen Mycobacterium abscessus: a novel hierarchical in silico approach, PLoS One 8 (2013), https://doi.org/10.1371/journal.pone.0059126.
[28] M. Dalsass, A. Brozzi, D. Medini, R. Rappuoli, Comparison of open-source reverse vaccinology programs for bacterial vaccine antigen discovery, Front. Immunol. 10 (2019), https://doi.org/10.3389/fimmu.2019.00113.
[29] G.E. Tusnády, I. Simon, The HMMTOP transmembrane topology prediction server, Bioinformatics 17 (2001), https://doi.org/10.1093/bioinformatics/17.9.849.
[30] I.A. Doytchinova, D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, BMC Bioinf. 8 (2007), https://doi.org/10.1186/1471-2105-8-4.
[31] C.N. Magnan, M. Zeller, M.A. Kayala, A. Vigil, A. Randall, P.L. Felgner, P. Baldi, High-throughput prediction of protein antigenicity using protein microarray data, Bioinformatics 26 (2010), https://doi.org/10.1093/bioinformatics/btq551.
[32] A. Jadhav, B. Shanmugham, A. Rajendiran, A. Pan, Unraveling novel broad-spectrum antibacterial targets in food and waterborne pathogens using comparative genomics and protein interaction network analysis, Infect. Genet. Evol. 27 (2014), https://doi.org/10.1016/j.meegid.2014.08.007.
[33] H. Luo, Y. Lin, T. Liu, F.L. Lai, C.T. Zhang, F. Gao, R. Zhang, DEG 15, an update of the Database of Essential Genes that includes built-in analysis tools, Nucleic Acids Res. 49 (2021), https://doi.org/10.1093/nar/gkaa917.
[34] G.F. Ejigu, J. Jung, Review on the computational genome annotation of sequences obtained by next-generation sequencing, Biology 9 (2020), https://doi.org/10.3390/biology9090295.
[35] B. Liu, D. Zheng, S. Zhou, L. Chen, J. Yang, VFDB 2022: a general classification scheme for bacterial virulence factors, Nucleic Acids Res. 50 (2022), https://doi.org/10.1093/nar/gkab1107.
[36] H.R. Ansari, D.R. Flower, G.P.S. Raghava, AntigenDB: an immunoinformatics database of pathogen antigens, Nucleic Acids Res. 38 (2009), https://doi.org/10.1093/nar/gkp830.
[37] A.A. Malik, S.C. Ojha, N. Schaduangrat, C. Nantasenamat, ABCpred: a webserver for the discovery of acetyl- and butyryl-cholinesterase inhibitors, Mol. Divers. 26 (2022), https://doi.org/10.1007/s11030-021-10292-6.
[38] E. Karosiene, C. Lundegaard, O. Lund, M. Nielsen, NetMHCcons: a consensus method for the major histocompatibility complex class i predictions, Immunogenetics 64 (2012), https://doi.org/10.1007/s00251-011-0579-8.
[39] H.H. Bui, J. Sidney, B. Peters, M. Sathiamurthy, A. Sinichi, K.A. Purton, B.R. Mothé, F.V. Chisari, D.I. Watkins, A. Sette, Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications, Immunogenetics 57 (2005), https://doi.org/10.1007/s00251-005-0798-y.
[40] A. Valero-Rello, A. Hapeshi, E. Anastasi, S. Alvarez, M. Scortti, W.G. Meijer, I. MacArthur, J.A. Vázquez-Bolanda, An invertron-like linear plasmid mediates intracellular survival and virulence in bovine isolates of *Rhodococcus equi*, Infect. Immun. 83 (2015), https://doi.org/10.1128/IAI.00376-15.
[41] D.C. Rees, E. Johnson, O. Lewinson, ABC transporters: the power to change, Nat. Rev. Mol. Cell Biol. 10 (2009), https://doi.org/10.1038/nrm2646.
[42] A.R. Kerr, P.V. Adrian, S. Estevão, R. De Groot, G. Alloing, J.P. Claverys, T.J. Mitchell, P.W.M. Hermans, The Ami-AliA/AliB permease of *Streptococcus pneumoniae* is involved in nasopharyngeal colonization but not in invasive disease, Infect. Immun. 72 (2004), https://doi.org/10.1128/IAI.72.7.3902-3906.2004.
[43] M.C. Hung, M.V. Humbert, J.R. Laver, R. Phillips, J.E. Heckels, M. Christodoulides, A putative amino acid ABC transporter substrate-binding protein, NMB1612, from *Neisseria meningitidis*, induces murine bactericidal antibodies against meningococci expressing heterologous NMB1612 proteins, Vaccine 33 (2015), https://doi.org/10.1016/j.vaccine.2015.07.032.
[44] T. Sharma, A. Alam, A. Ehtram, A. Rani, S. Grover, N.Z. Ehtesham, S.E. Hasnain, The *Mycobacterium tuberculosis* PE_PGRS protein family acts as an immunological decoy to subvert host immune response, Int. J. Mol. Sci. 23 (2022), https://doi.org/10.3390/ijms23010525.
[45] P.K. Giri, I. Verma, G.K. Khuller, Enhanced immunoprotective potential of *Mycobacterium tuberculosis* Ag85 complex protein based vaccine against airway Mycobacterium tuberculosis challenge following intranasal administration, FEMS Immunol. Med. Microbiol. 47 (2006), https://doi.org/10.1111/j.1574-695X.2006.00087.x.
[46] M. Letek, P. González, I. MacArthur, H. Rodríguez, T.C. Freeman, A. Valero-Rello, M. Blanco, T. Buckley, I. Cherevach, R. Fahey, A. Hapeshi, J. Holdstock, D. Leadon, J. Navas, A. Ocampo, M.A. Quail, M. Sanders, M.M. Scortti, J.F. Prescott, U. Fogarty, W.G. Meijer, J. Parkhill, S.D. Bentley, J.A. Vázquez-Boland, The genome of a pathogenic *Rhodococcus*: cooptive virulence underpinned by key gene acquisitions, PLoS Genet. 6 (2010), https://doi.org/10.1371/journal.pgen.1001145.
[47] E. Sauvage, F. Kerff, M. Terrak, J.A. Ayala, P. Charlier, The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis, FEMS Microbiol. Rev. 32 (2008), https://doi.org/10.1111/j.1574-6976.2008.00105.x.
[48] M. Hill, A.E. Deghmane, M. Segovia, M.L. Zarantonelli, G. Tilly, P. Blancou, G. Bériou, R. Josien, I. Anegon, E. Hong, C. Ruckly, A. Antignac, M. El Ghachi, I. G. Boneca, M.K. Taha, M.C. Cuturi, Penicillin binding proteins as danger signals: meningococcal penicillin binding protein 2 activates dendritic cells through toll-like receptor 4, PLoS One 6 (2011), https://doi.org/10.1371/journal.pone.0023995.
[49] M. Pérez-Gallego, G. Torrens, J. Castillo-Vera, B. Moya, L. Zamorano, G. Cabot, K. Hultenby, S. Albertí, P. Mellroth, B. Henriques-Normark, S. Normark, A. Oliver, C. Juan, Impact of AmpC derepression on fitness and virulence: the mechanism or the pathway? mBio 7 (2016) https://doi.org/10.1128/mBio.01783-16.
[50] A. Wysocka, E. Jagielska, Ł. Łężniak, I. Sabała, Two new M23 peptidoglycan hydrolases with distinct net charge, Front. Microbiol. 12 (2021), https://doi.org/10.3389/fmicb.2021.719689.
[51] A.J. Darwin, V.L. Miller, Identification of *Yersinia enterocolitica* genes affecting survival in an animal host using signature-tagged transposon mutagenesis, Mol. Microbiol. 32 (1999), https://doi.org/10.1046/j.1365-2958.1999.01324.x.
[52] A. Tidhar, Y. Flashner, S. Cohen, Y. Levi, A. Zauberman, D. Gur, M. Aftalion, E. Elhanany, A. Zvi, A. Shafferman, E. Mamroud, The NIpD lipoprotein is a novel Yersinia pestis virulence factor essential for the development of plague, PLoS One 4 (2009), https://doi.org/10.1371/journal.pone.0007023.
[53] A. Oubrie, H.J. Rozeboom, B.W. Dijkstra, Active-site structure of the soluble quinoprotein glucose dehydrogenase complexed with methylhydrazine: a covalent cofactor-inhibitor complex, Proc. Natl. Acad. Sci. U. S. A. 96 (1999), https://doi.org/10.1073/pnas.96.21.11787.
[54] Y. Chen, J.N. Xiao, Y. Li, Y.J. Xiao, Y.Q. Xiong, Y. Liu, S.J. Wang, P. Ji, G.P. Zhao, H. Shen, S.H. Lu, X.Y. Fan, Y. Wang, *Mycobacterial* lipoprotein Z triggers efficient innate and adaptive immunity for protection against Mycobacterium tuberculosis infection, Front. Immunol. 10 (2019), https://doi.org/10.3389/fimmu.2018.03190.
[55] W. Liu, Y.H. Chen, High epitope density in a single protein molecule significantly enhances antigenicity as well as immunogenecity: a novel strategy for modern vaccine development and a preliminary investigation about B cell discrimination of monomeric proteins, Eur. J. Immunol. 35 (2005), https://doi.org/10.1002/eji.200425749.
[56] S. Mashima, Comparative sequence analysis of equine and human MHC class II DQB genes, Cytogenet. Genome Res. 102 (2003), https://doi.org/10.1159/000075748.
[57] A.L. Gustafson, R.L. Tallmadge, N. Ramlachan, D. Miller, H. Bird, D.F. Antczak, T. Raudsepp, B.P. Chowdharya, L.C. Skow, An ordered BAC contig map of the equine major histocompatibility complex, Cytogenet. Genome Res. 102 (2003), https://doi.org/10.1159/000075747.