

Pangenomics-Driven Reverse Vaccinology for the Discovery of New Vaccine Candidates Against *Bordetella pertussis*

Bioinformatics and Biology Insights

Volume 20: 1–10

© The Author(s) 2026

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/11779322261442522

journals.sagepub.com/home/bbi

Majid Abdelouahab^{1,2}  and Diawara Idrissa^{1,2,3,4}

Abstract

Despite widespread vaccination, pertussis resurgence persists due to waning immunity and emerging resistant strains in *Bordetella pertussis*. This study employs a pangenomics-driven reverse vaccinology approach to identify novel vaccine candidates. Analyzing 160 genomes revealed a closed pan-genome, with approximately 60% conserved genes, including 3389 core genes of which 1312 participate in pathogen-specific pathways. Non-homologous proteins were identified by comparison against human and microbiome proteomes, yielding 205 candidates. Essentiality assessment via the Database of Essential Genes (DEG) refined this to 63 non-homologous essential proteins. A multi-criteria selection process evaluated purifiability based on physicochemical properties and transmembrane helices, accessibility (extracellular or secreted localization), and immunogenicity through antigenicity prediction and B-cell epitope mapping. This pipeline culminated in 11 high-potential vaccine targets. The in silico methodology offers rapidity, cost-effectiveness, and reduced side effects compared with conventional vaccinology, though experimental validation is essential for confirmation.

Keywords

Bordetella pertussis, pertussis resurgence, reverse vaccinology, pan-genomics, immuno-bioinformatics

Received: 18 December 2025; accepted: 30 March 2026

Introduction

Pertussis, commonly known as whooping cough, is a highly contagious respiratory infection caused by the Gram-negative coccobacillus *Bordetella pertussis*. This pathogen primarily affects the upper and lower respiratory tracts in humans, leading to paroxysmal cough with or without whoops, leukocytosis, alveolitis, and bronchopneumonia in vulnerable populations like infants, which can result in potentially fatal outcomes.¹

Despite the availability of vaccination programs, pertussis remains a major global public health concern. According to the World Health Organization, pertussis exhibits cyclical epidemic peaks every 2 to 5 years, with multiple large-scale outbreaks reported over the past decade.² Although reported cases declined markedly during the COVID-19 pandemic, a substantial resurgence has been observed in the post-pandemic period, reaching alarming levels worldwide.^{3,4}

Vaccination has long been the cornerstone of pertussis prevention and has substantially reduced disease incidence worldwide. The shift from DTwP to DTaP vaccines was driven by the latter's improved safety profile and tolerability due to reduced reactogenicity, making DTaP preferable for pediatric use.⁵ Waning immunity, antigenic variations in circulating strains, and diminished natural boosting from

subclinical infections have been associated with the resurgence of pertussis in highly vaccinated populations.^{5,6} Moreover, DTaP vaccines, which target a limited set of antigens, exhibit limitations that may restrict the breadth and durability of the immune response.^{7,8} Together, these limitations underscore a critical gap, highlighting the urgent need for next-generation vaccines with enhanced and broader protective efficacy.

¹Research Technology Platforms, Mohammed VI University of Sciences and Health (UM6SS), Casablanca, Morocco

²Mohammed VI Higher Institute of Biosciences and Biotechnologies, Mohammed VI University of Sciences and Health (UM6SS), Casablanca, Morocco

³Research Laboratory of Microbiology, Infectious Diseases, Allergology and Pathogen Surveillance (LARMIAS), Mohammed VI Faculty of Medicine, Mohammed VI University of Sciences and Health (UM6SS), Casablanca, Morocco

⁴Mohammed VI Center for Research and Innovation (CM6RI), Rabat, Morocco

Corresponding author:

Majid Abdelouahab, Research Technology Platforms, Mohammed VI University of Sciences and Health (UM6SS), Mohamed Taieb Naciri Boulevard, Hay Hassani District, Casablanca 82403, Morocco.
Email: amajid@um6ss.ma



The persistence and resurgence of pertussis are further illuminated by the genetic evolution of *B. pertussis*, providing critical context for understanding vaccine inefficacy. Comparative genomic studies have reported that genomic variations (polymorphisms and mutations in key virulence genes) affecting virulence-associated factors and surface antigens have increased, underscoring how evolutionary pressures—such as vaccine selection—may drive adaptation.^{9–11} This genetic evolution limits the effectiveness of vaccines targeting a restricted number of antigens and highlights the limitations of traditional vaccine target discovery and development approaches, which are often protracted and constrained by the intricate biology of pathogens.¹² In contrast, reverse vaccinology harnesses computational tools to systematically screen for antigenic proteins, accelerating discovery and facilitating comprehensive genomic analysis that allows the capture of strain diversity while identifying conserved and functionally relevant targets.^{13,14}

Pangenomics, a subset of this paradigm, examines the complete genomic repertoire across strains to pinpoint conserved elements, such as those involved in essential metabolic processes or signaling pathways, ensuring broad applicability irrespective of strain-specific variations.^{15,16} By integrating multiple genomes, this approach enables the identification of the core genome and conserved genes shared across all strains, as well as the non-core genome; classified into accessory and singleton genomes; that contributes to phenotypic variability and adaptation.¹⁷

This study addresses these challenges by employing pangenomics reverse vaccinology—a sophisticated, data-driven framework—to identify novel vaccine candidates from *B. pertussis* genomes. By integrating pangenomics with subtractive genomics, we identified and targeted core genome; encoded proteins that are essential for bacterial survival or persistence, while excluding homologous proteins to human and microbiota proteomes to enhance specificity, these proteins are surface-exposed or secreted and immunity-induced for certainty.

Materials and Methods

This *in silico* study was carried out at the research technology platforms of Mohammed VI University of Sciences and Health (UM6SS) in Casablanca, Morocco. To systematically identify potential vaccine targets, a pangenomics-driven *in silico* approach was employed. This strategy is based on the comparative analysis of available *B. pertussis* genomes to define conserved, pathogen-specific gene sets, followed by a rational prioritization framework. The overall workflow, illustrated in Figure 1, integrates pangenomics analysis and successive bioinformatic filtering steps to select proteins with high vaccine potential. This approach aims to enhance target specificity while minimizing cross-reactivity with host and microbiota components.

Pangenome Analysis

Prior to initiating the pangenome analysis, a series of preparatory steps are required, starting with the acquisition of

genome data. Ten sequencing reads of *B. pertussis* isolates in FASTQ format were retrieved from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) (Home–SRA–NCBI) (Supplemental Table S1).¹⁸ This retrieval was followed by quality control using FastQC (v0.12.1) (<https://github.com/s-andrews/FastQC.git>),¹⁹ after which adapters and low-quality reads were trimmed with Fastp (v0.24.1) (<https://github.com/OpenGene/fastp.git>).²⁰ The trimmed raw reads, originally downloaded from NCBI-SRA, underwent *de novo* assembly employing SPAdes (v4.2.0) (<https://github.com/ablab/spades.git>), with default parameters, to generate a novel complete genome of *B. pertussis*,²¹ with the assembly outcomes subsequently assessed via Quast (v5.3.0) (<https://github.com/ablab/quast.git>) to evaluate standard assembly metrics such as number of contigs, total assembly length, largest contig, N50, and L50.²² Finally, genome annotation was performed using Prokka (v1.14.5) (<https://github.com/tseemann/prokka.git>) with the parameters `–outdir`, `–prefix`, `bordpt`, and `–locustag BP`, specifying the output directory, file prefix and locus tag prefix used for the analysis. Prokka used to delineate additional attributes and essential elements (such as coding sequences, rRNA, tRNA, and others) within the *de novo* assembled genome.²³ Therefore, the pangenome analysis was conducted using Roary (v3.13.0) (<https://github.com/sanger-pathogens/Roary.git>), analyzing a set of 150 Prokka-generated GFF3 files of complete *B. pertussis* genome (Supplemental Table S2), retrieved from NCBI RefSeq database and the newly annotated genome, all of which were quality-checked prior to analysis using Quast. The purpose of this analysis was to construct the pan-genome and delineate core versus non-core gene sets for comparative genomic interpretation.²⁴ The pangenome analysis was run with the parameters `–e –n –v –p 8 –f` with default settings including the default 95 % BLASTP identity threshold and default paralog splitting behavior, which separates paralogous gene copies into distinct clusters and excludes unresolved paralogs from the core alignment.

Identification of Host and Pathogen Metabolic Pathways

Through a systematic workflow (Figure 1), this study represents the first computational and subtractive genomics analysis of various metabolic pathways to identify potential vaccine targets in *B. pertussis*. To derive metabolic pathway-based information, 2 steps were performed; the first step is functional annotation, involved uploading a list of 3389 core genes to an automatic annotation server called BlastKOALA (v3.1) (<https://www.kegg.jp/blastkoala/>) developed by KEGG (Kyoto Encyclopedia of Genes and Genomes) to correlate each gene with its corresponding KEGG Ortholog (KO) so the metabolic pathway.²⁵ Data were extracted with reference to the metabolic pathways and identification numbers for the human host. Subsequently, a comparison was conducted between the *B. pertussis* and human pathways, and pathways present in the pathogen but absent in the host were selected as unique to *B. pertussis*.

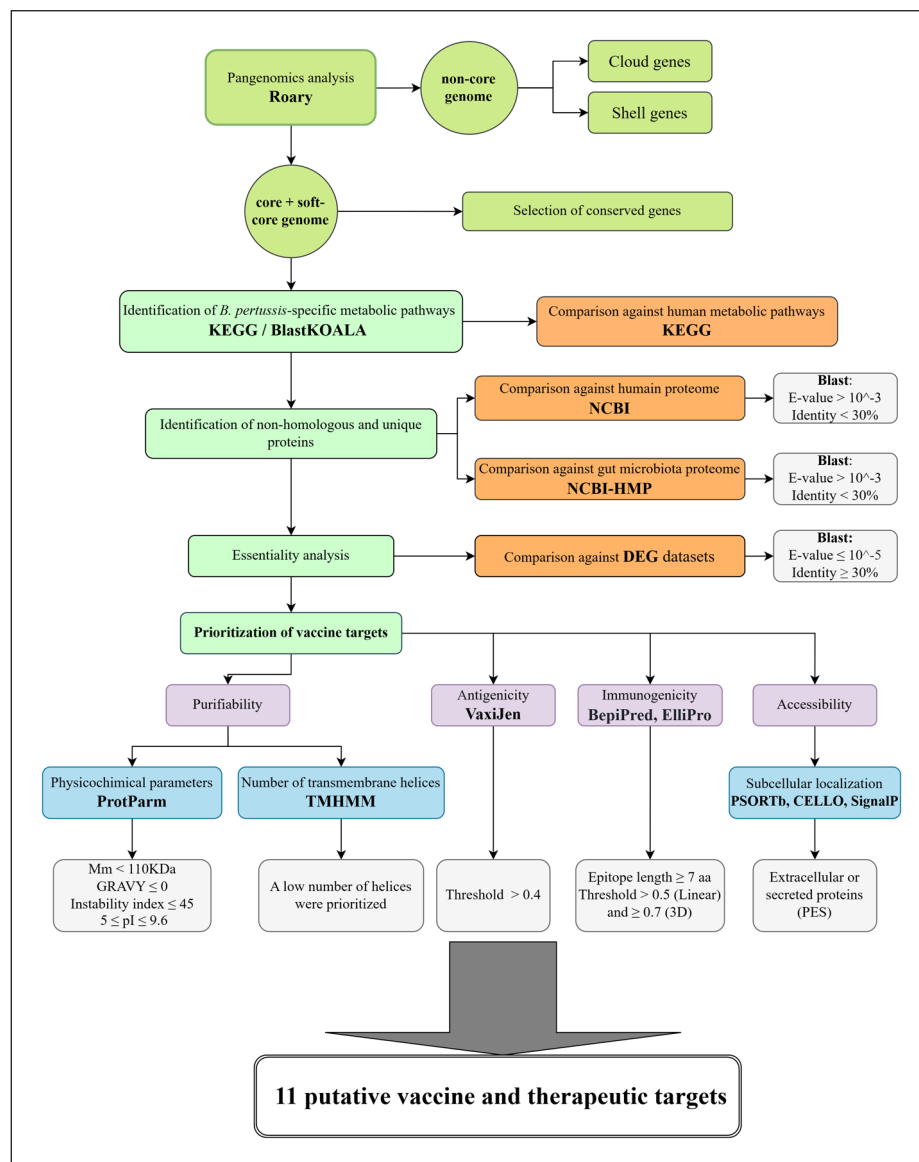


Figure 1. Schematic representation of the pangenomics-driven reverse vaccinology pipeline applied to prioritize targets in *B. pertussis* leading to the identification of 11 putative vaccine and therapeutic targets.

Identifying the Non-Homologues and Unique Proteins

This step was performed to eliminate proteins from common pathways which serve to minimize potential side effects, the perturbation of the protein function and avoid inducing an immune response against the microbiota. To execute this step, 2-step comparisons were conducted between human and pathogen proteomes to identify non-homologous proteins of *B. pertussis* and between the microbiota and pathogen proteomes to identify unique proteins of *B. pertussis*. Initially, all human protein were retrieved from Refseq database ftp site (v May 23, 2018) (Index of /refseq/H_sapiens/mRNA_Prot) and only proteins from pathogen-specific pathways were subjected to BLASTP (v2.16.0) analysis, with E-value equal to 10^{-3} , against the human proteome.²⁶ Subsequently, non-homologous proteins were compared using BLASTP, with

E-value equal to 10^{-3} , analysis against the microbiota proteome, which yield from NIH Human microbiome project (https://www.ncbi.nlm.nih.gov/datasets/genome/?bioproject=PRJNA43021&assembly_level=3:3). In each case, hits were filtered based on E-value $> 10^{-3}$, identity $< 30\%$, query coverage $< 50\%$, and others as default. Proteins that did not produce hits meeting these criteria were selected as non-homologous and unique proteins.

Identification of Essential Proteins for *Bordetella pertussis*

After identifying all non-homologous and unique proteins, we further filtered them based on their essentiality, specifically, targeting proteins critical for *B. pertussis* survival. Essential proteins in cellular organisms are indispensable for viability and replication, making them promising targets for antimicrobial treatments. We compared the list of

these proteins to a database of experimentally verified essential genes from approximately 30 bacterial species in the Database of Essential Genes (DEG: v Sep. 1, 2020) (<http://origin.tubic.org/deg/public/index.php/download>).²⁷ An E-value threshold of 10^{-5} was used to screen essential proteins of *B. pertussis* using DEG microbial BLASTP. Significant hits were obtained by filtering based on E-value $\leq 10^{-5}$, identity $\geq 30\%$ and a minimum query coverage of 50%.

Prioritization of Vaccine Targets

B. pertussis essential genes were further evaluated on the basis of their subcellular localization, physicochemical properties, and immunogenicity analysis. This analysis is mandatory to characterize the antigenicity of these essential genes and validate them as vaccine targets.

Subcellular Localization

The subcellular localization prediction and secretion screening was performed on the non-homologous and essential proteins identified in *B. pertussis*. PSORTb (v3.0.3) (<https://psort.org/psortb/>), was applied first, generating confidence scores from 0 to 10 for each compartment and retaining predictions with the highest score (≥ 7.5); proteins scoring below this threshold were further analyzed with CELLO (v2.5) (<https://cello.life.nctu.edu.tw/>) using normalized probability scores and a reliability index (0-5).^{28,29} SignalP (v5.0) (<https://services.healthtech.dtu.dk/services/SignalP-5.0/>) was additionally employed to detect signal peptides and secretion pathways based on probability scores ≥ 0.7 . In cases of disagreement or low confidence, consensus evaluation was applied, prioritizing PSORTb predictions and using CELLO and SignalP for complementary validation.³⁰ The retained proteins predicted as surface-exposed or secreted (outer membrane, periplasmic, or extracellular), which were advanced to the downstream physicochemical and antigenicity analyses.

Physicochemical Parameters

Initially, information regarding non-homologous, unique, and essential proteins was evaluated to prioritize suitable vaccine targets; however, to minimize the time required for vaccine testing and development, it is preferable to evaluate these vaccine targets by several molecular and structural criteria.³¹ This involved calculating the molecular weight, which must be less than 110 kDa, and the GRAVY score, which should be less than or equal to zero. These 2 parameters facilitate the purification of targeted proteins.³² The stability index, which must be less than or equal to 45, indicates protein stability. For pI, there is no critical value, but an extreme pI value could cause issues; thus, the pI should range between 5 and 9 and differ from human pH. These calculations were performed using the ProtParam (<https://web.expasy.org/protparam/>) computational web tool.³³ The next criteria involved employing TMHMM

(v2.0) (<https://services.healthtech.dtu.dk/services/TMHMM-2.0/>), which used to predict the number of the transmembrane helices of the proteins. The selection was made if a protein has 1 or 0 transmembrane helices.^{34,35}

Prediction of Antigenicity and Immunogenicity

Proteins classified as PSE or secreted are those that first interact with the immune system³⁶; thus, we predicted their ability to elicit a strong immune response by evaluating their antigenicity using VaxiJen (v2.0) (<https://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>). VaxiJen employs to its alignment-independent approach, which evaluates protein antigenicity based on the physicochemical properties of amino acids rather than sequence similarity. Proteins with a score > 0.4 were considered probable antigens with significant immunogenic potential.^{37,38} Subsequently, we proceeded to predict of potentially immunogenic epitopes in a given protein sequence, that may significantly reduce wet lab effort needed to discover the epitopes required for the design of vaccines and for immunodiagnosics. The goal of B cell epitope prediction was to identify immunogenic epitopes on potential antigens, encompassing both linear (continuous) and discontinuous (conformational) categories, capable of binding to B cell receptors, thereby activating B lymphocytes and eliciting a humoral immune response characterized by the secretion of neutralizing antibodies.^{39,40} Linear B-cell epitopes were predicted employing BepiPred (v3.0) (<https://services.healthtech.dtu.dk/services/BepiPred-3.0/>) with a threshold of 0.5 and a minimum residue length of 7 amino acids.⁴¹ For conformational epitopes, which necessitate a 3D protein structure model, the Protein Data Bank (PDB) (<https://www.rcsb.org/>) was initially queried to ascertain the availability of structures for the target proteins; in cases where no structure existed, a homology model was generated using SWISS-MODEL (<https://swissmodel.expasy.org/>).⁴² The corresponding .pdb files were subsequently imported into the ElliPro tool, applying a protrusion index threshold ≥ 0.7 and a minimum residue length of 7 amino acids, which uses a combination of geometric features and the propensity of single amino acids in an antigen for conformational epitope prediction.^{39,43}

Results

Analysis Pangenomics

Using Roary_plots for visualization of pangenomics analysis results give an overview of *B. pertussis* pangenomics type (Figure 2). A total of 5747 genes clusters was found and classified into 3 categories, the core genome (core and soft core) with 3442 genes which present in more than 95% of strains (143 strains) occupied 59.9% of genome which mean a strong conservation of genes and the non-core genome which shell genome with 553 genes present between the 15% and 95% of strains and cloud genome correspond to 1752 of genes which present in less than 15%.

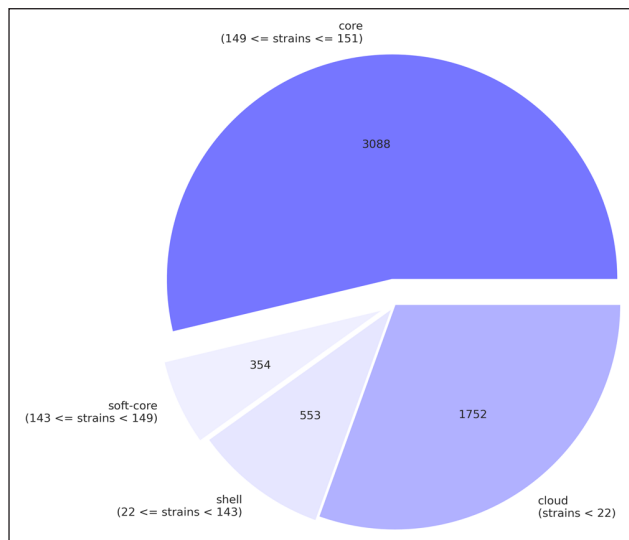


Figure 2. Distribution of 5747 genes into the core (core + soft-core) and the non-core (shell + cloud) genome categories.

The Shell and Cloud genome are occupying, respectively, 9.6% and 30.5% of genome with a total of 60.1%. This distribution, core genome is more than 50% and the ratio non-core genome on core genome strict (without soft genome) equal around 0.75 which is inferior to 1, confirms that the *B pertussis* pangenomics is closed, which is also confirmed by Parkhill et al.⁹

Identifying the Non-Homologues and Unique Proteins

Proteins identified as human homologs were excluded in the initial step, as such targets could adversely affect host metabolism. Unique metabolic pathways specific to *B pertussis*, but absent in the human host, encompass proteins that represent promising candidates for drugs and vaccines. In addition, certain proteins within shared pathways between the pathogen and host may be exclusive to *B pertussis*; those involved in multiple pathways, particularly if non-homologous, emerge as strong vaccine candidates owing to their essential functions. Submission of the *B pertussis* core gene list to KEGG yielded annotations for 1993 genes across 48 pathways, including 1312 genes associated with pathogen-specific pathways and 275 genes shared with humans. Proteins within unique pathways, such as lipopolysaccharide biosynthesis, oxidative phosphorylation, and bacterial secretion systems, were deemed robust vaccine targets. Protein sequences from both common and unique pathways were subsequently aligned against the human proteome via NCBI BLASTP to detect non-homologous proteins, resulting in 1010 sequences exhibiting no matches to the human proteome and thus likely to avoid eliciting autoimmune responses. These non-homologous proteins were further compared against the microbiota proteome, with 205 proteins identified as unique to *B pertussis* and selected for subsequent evaluation.

Identification of Essential Proteins for *Bordetella pertussis*

Being non-human and involved in unique metabolic pathways are not the sole criteria for selecting favorable vaccine targets. Identifying proteins that regulate key factors, such as nutrient uptake, growth, secretion systems, virulence, and pathogenicity, is of great importance for disrupting pathogen functions and survival.^{27,44} Such proteins are considered essential for the pathogen. However, not all essential proteins are non-homologous. Therefore, pathogen proteins that are both non-homologous and essential represent more attractive vaccine targets.⁴⁵ The non-homologous proteins identified through resistance gene analysis were further screened based on their essentiality. A total of 63 *B pertussis* proteins were found to be essential, as they are involved in activities significant to the pathogen's biological pathways (Table 1).

Prioritization of Vaccine Targets

Sub-Cellular Localization. Subcellular localization is a key functional attribute of a protein, as many cellular processes are compartmentalized; thus, predicting the localization of unknown proteins can offer valuable insights into their roles and assist in selecting candidates for further study.³¹ In bacteria, all proteins are synthesized in the cytoplasm, where many remain to perform specific functions, while others contain export signals that direct them to different cellular compartments. In Gram-negative bacteria, compartments include the cytoplasmic membrane, periplasm, inner and outer membranes, and extracellular space.²⁸ Although most proteins are confined to a single compartment, some span multiple locations.

The subcellular localization of 63 non-homologous essential proteins of *B pertussis* was evaluated using a combination of prediction tools: PSORTb, followed by cross-validation with CELLO and SignalP. Of these, 27 proteins were predicted to be cytoplasmic, 15 membrane-associated, 18 periplasmic, 9 located in the inner membrane, and 3 extracellular (Figure 3).

Subcellular localization is a critical factor in identifying effective drug and vaccine targets; proteins localized in the cytoplasm or membrane are often considered for drug development, while those secreted or exposed on the bacterial surface are ideal candidates for vaccines.⁴⁶ The next step involves refining these candidates further through multilevel analysis of physicochemical parameters.

Physicochemical Parameters. We initially screened the secreted or extracellularly exposed proteins of *B pertussis* for potential vaccine targets, followed by the application of additional prioritization parameters, including physicochemical properties, to refine the selection. These properties were critical in the early stages of vaccine discovery. Most proteins had a molecular weight of less than 110 kDa and a GRAVY score of ≤ 0 , indicating solubility (hydrophilic proteins), which supports their potential for experimental study in vaccine development. In addition, an

Table 1. Non-Homologous and Essential *B. pertussis* Proteins Retained After Comparison Against DEG Database.

Protein	
BP_00041: Cytochrome c biogenesis protein	BP_01223: HTH-type transcriptional regulator/antitoxin HigA
BP_00165: Phospholipid transport system substrate-binding protein	BP_01332: HTH-type transcriptional regulator/antitoxin MqsA
BP_00233: 3-deoxy-D-manno-octulosonic acid kinase	BP_01841: Nickel/cobalt transporter (NicO) family protein
BP_00260: Lipopolysaccharide heptosyltransferase II	BP_01457: Lipopolysaccharide export system permease protein
BP_00302: Uroporphyrin-III C-methyltransferase	BP_01576: Outer membrane lipoprotein carrier protein
BP_00311: Pilus assembly protein CpaC	BP_01641: Rod shape-determining protein MreD
BP_00325: Outer membrane protein	BP_01804: Transmembrane sensor
BP_00358: DNA-binding protein H-NS	BP_01367: Uncharacterized protein
BP_00399: Hemolysin activation/secretion protein	BP_01857: ptxB; pertussis toxin subunit 2
BP_00400: fimD, fimC, Outer membrane usher protein	BP_01858: ptxD; pertussis toxin subunit 4
BP_00465: Type III secretion protein K	BP_01859: ptxE; pertussis toxin subunit 5
BP_00466: Type III secretion protein L	BP_01860: ptxC; pertussis toxin subunit 3
BP_00563: Cytolysin-activating lysine-acyltransferase	BP_01890: Primosomal replication protein N
BP_00614: Cell division protein FtsL	BP_02451: 3-demethoxyubiquinol 3-hydroxylase
BP_00630: Sulfonate transport system permease protein	BP_02414: Poly[(R)-3-hydroxyalkanoate] polymerase subunit PhaC
BP_00640: D-threonine aldolase	BP_02637: LPS-assembly protein
BP_00805: Outer membrane protein assembly factor BamE	BP_02638: N-acetylmuramate I-kinase
BP_00815: Methylglutamate dehydrogenase subunit B	BP_02690: Uncharacterized protein
BP_00836: Protoporphyrinogen IX oxidase	BP_02889: Malonyl-CoA O-methyltransferase
BP_00850: Outer membrane protein, multidrug efflux system	BP_02793: Lipopolysaccharide export system protein LptA
BP_00966: Succinate dehydrogenase cytochrome b subunit	BP_03002: Uncharacterized protein involved in response to NO
BP_01025: Putative oxidoreductase	BP_02906: LPS-assembly lipoprotein
BP_01027: Uncharacterized protein	BP_03011: DNA polymerase III subunit chi
BP_01055: IclR family transcriptional regulator, pca regulon regulatory protein	BP_03013: Lipopolysaccharide export system permease protein
BP_01063: Putative spermidine/putrescine transport system substrate-binding protein	BP_03066: Sulfite reductase (NADPH) hemoprotein beta-component
BP_01143: Aerobic carbon-monoxide dehydrogenase large subunit	BP_03079: Cytochrome c oxidase cbb3-type subunit
BP_01156: Xanthine dehydrogenase accessory factor	BP_03088: Thiol-disulfide interchange protein DsbC
BP_01189: Tungstate transport system substrate-binding protein	BP_03126: General bacterial porin, GBP family
BP_01207: Outer membrane protein assembly factor BamB	BP_03239: L-histidine Nalpha-methyltransferase
BP_01214: ATP phosphoribosyltransferase regulatory subunit	BP_0350: Multicomponent K ⁺ : H ⁺ antiporter subunit E
BP_03749: Homoserine kinase type II	BP_03892: Cell division protein ZapA
BP_03797: Uncharacterized protein	

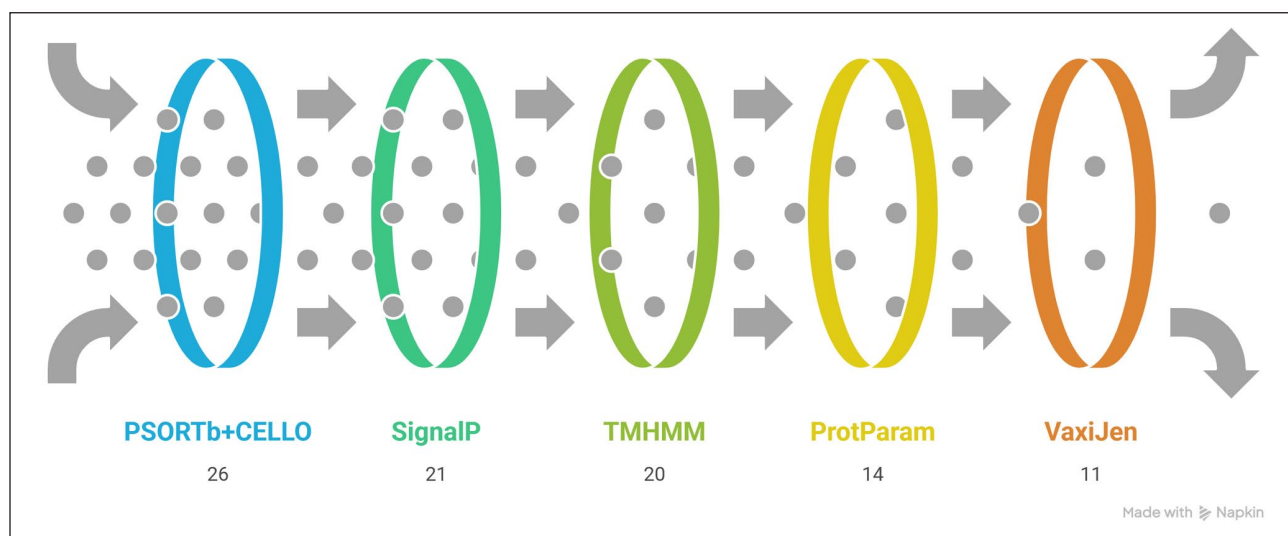
**Figure 3.** A multi-selection screening pipeline for the characterization and prioritization of 63 essential proteins using subcellular localization, signal peptide, transmembrane topology, physicochemical, and antigenicity prediction tools (Made using Napkin).

Table 2. Analysis of Physicochemical Properties (Molecular Mass, pI, Stability, GRAVY), Subcellular Localization, Transmembrane Helix Number, Signal Peptide Prediction, and Antigenicity.

Protein	Mm (Da)	pI	Instability index	GRAVY	Localization	TM	SignalP	Vaxijen
BP_00041: Cytochrome c biogenesis protein	76995.9	6.5	38.57	−0.03	Cytoplasmic-Membrane	4	Absent	0.514
BP_00399: Hemolysin activation/secretion protein (fhaC)	64445	9.6	29.66	−0.37	Extracellular	1	Present	0.599
BP_00400: fimD, fimC, Outer membrane usher protein	40467.8	7.6	30.39	−0.06	Extracellular	1	Present	0.642
BP_00805: outer membrane protein assembly factor BamE	20488.8	6	43.97	−0.64	Periplasmic	0	Present	0.561
BP_00850: outer membrane protein, multidrug efflux system	53097.8	6.3	40.33	−0.2	Extracellular	0	Present	0.685
BP_01207: Outer membrane protein assembly factor BamB	31498.4	8.7	19.04	−0.04	Extracellular	0	Present	0.601
BP_01857: ptxB; pertussis toxin subunit 2	24829	8.6	46.48	−0.12	Extracellular	0	Present	0.394
BP_01859: ptxE; pertussis toxin subunit 5	13094	5.5	42.18	0.2	Secreted	0	Present	0.486
BP_02637: LPS-assembly protein	88299.7	8	36.21	−0.45	Extracellular	0	Present	0.564
BP_02906: LPS-assembly lipoprotein	24231.2	5.6	32.7	−0.36	Periplasmic	1	Present	0.459
BP_03126: General bacterial porin, GBP family	30129.1	5	20.56	−0.18	Extracellular	0	Present	0.643

instability index of ≤ 45 signifies protein stability. Although the isoelectric point (pI) lacks a critical threshold, proteins with extreme pI values may pose challenges; thus, a pI range of 5 to 9.6 is preferred, though not strictly codified. Furthermore, estimation of transmembrane helices in the proteins revealed all 11 proteins to have either 0 or 1 transmembrane helices, (except Cytochrome c biogenesis protein) which to its easy ability to be easily purified in experimental analysis (Figure 3 and Table 2).

Prediction of Antigenicity and Immunogenicity

Antigenicity. The proteins classified as secreted or PSE were analyzed using VaxiJen software. This analysis revealed that all 11 proteins have a probability above the threshold of 0.4, indicating that they are immunogenic and capable of generating antibodies, thus making them strong candidates for vaccine targets (Figure 3 and Table 2).

Immunogenicity. Peptide vaccines are more convenient and safer than conventional vaccines, as they include only immunogenic epitopes rather than the full 3-dimensional protein structure. B-cell epitopes are antigenic determinants that are recognized and bound by receptors on the surface of B lymphocytes. A total of 103 linear B-cell epitopes were identified across the 11 examined proteins of *B. pertussis*, employing a threshold score exceeding 0.5 and a minimum epitope length of 7 amino acids (Table 3). For conformational epitope detection, 3-dimensional structures of all proteins were modeled using the SWISS-MODEL homology-modeling server, with the structure exhibiting the highest sequence identity and coverage selected for each protein. The resulting PDB files were subsequently imported into the ElliPro

tool, applying a threshold of 0.7 and a minimum epitope length of 7 amino acids. This analysis identified 24 conformational B-cell epitopes from the 11 proteins under investigation (Table 3).

Discussion

Face to decline of vaccine efficacy and increase of pertussis cases reported around the world, it was suggested the necessity of finding novel therapeutic strategy to improve the vaccines and to combat this resurgence. The purpose of this study was to identifying new vaccine targets against *B. pertussis*, based on pangenomics approach, via a list of 3389 core genes. The efficacy of a vaccine depends on the careful selection of appropriate antigens. Critical criteria for antigen evaluation include (1) non-homology to human proteins, to minimize cross-reactivity and potential adverse effects; (2) accessibility, determined by the antigen's subcellular location, to ensure it can be targeted by the immune system; and (3) immunogenicity, defined as the ability to stimulate B-cell and/or T-cell responses. These factors collectively enhance the vaccine's potential to activate protective immunity while avoiding interference with host protein functions.

The pangenomics approach, which is applied in this study, successfully reduced the number of candidate proteins from 1010 to 11. These proteins were selected as vaccine targets for treating *B. pertussis* infections due to their high potential among identified non-homologous proteins and their significant role in the pathogenicity of *B. pertussis*, suggesting that targeting them could disrupt the pathogen's survival and colonization. Of these 11 proteins, 5 have experimentally confirmed antigenicity and immunogenicity: 2 pertussis toxin (Ptx) subunits S2 (BP_01857) and S5 (BP_01859),^{6,47} 2 LPS assembling proteins (BP_02637 and

Table 3. Number of Predicted Linear and Conformational Epitopes for the 11 Candidate Targets.

Protein	Linear	Conformational
BP_00041: Cytochrome c biogenesis protein	19	3
BP_00399: Hemolysin activation/secretion protein (FhaC)	14	4
BP_00400: fimD, fimC, Outer membrane usher protein	9	2
BP_00805: outer membrane protein assembly factor BamE	5	3
BP_00850: Outer membrane protein, multidrug efflux system	12	3
BP_01207: Outer membrane protein assembly factor BamB	11	1
BP_01857: ptxB; pertussis toxin subunit 2	3	2
BP_01859: ptxE; pertussis toxin subunit 5	3	1
BP_02637: LPS-assembly protein	18	2
BP_02906: LPS-assembly lipoprotein	2	2
BP_03126: General bacterial porin, GBP family	7	1

BP_02906),⁴⁸ and the outer membrane protein assembly factors BamB (BP_01207),⁴⁸ regarding the outer membrane protein (BP_03126) belonging to the general bacterial porin (GBP) family, needs further characterization to confirm its identity.⁴⁹

Two pertussis toxin (Ptx) subunits, S2 (BP_01857) and S5 (BP_01859), are widely recognized as major virulence factors expressed by *B. pertussis*, which contribute significantly to bacterial adhesion and immune modulation, making them key components of *B. pertussis* vaccines (DTaP and DTwP).^{6,47} Our identification of S2 and S5 through a pangenomics approach confirms their conservation and relevance as vaccine targets. However, vaccine-induced selective pressure has led to observed variability in pertussis toxin genes, which may affect the effectiveness of subunits such as S2 and S5.^{50,51} Consequently, a vaccination strategy targeting stable genomic regions and continuous monitoring of antigenic variability are required.

Two LPS assembling proteins (BP_02637 and BP_02906) are highly appeared to be the novel vaccine targets, replacing LPS, which are the major component of DTwP and the main responsible for the reactogenicity.⁴⁷ Assembling proteins, such as LptD, whose immunogenicity was demonstrated in a study done by Dorji, showed that mice vaccinated with LptD alone exhibited significantly lower bacterial loads compared with those vaccinated solely with DTaP.⁴⁸

The outer membrane protein assembly factors BamE (BP_00805) and BamB (BP_01207) are components of the β -barrel assembly machinery (BAM) complex. This complex facilitates the insertion and folding of β -barrel proteins into the outer membrane (OM) of Gram-negative bacteria.⁵² Targeting the BAM complex holds potential not only for vaccine development but also for developing potentiators that enhance OM permeability. The

immunogenicity of BamB was demonstrated in vivo by Dorji, who showed that mice vaccinated with BamB alone exhibited significantly lower bacterial loads compared with those vaccinated solely with DTaP.⁴⁸ However, the immunogenicity of BamE remains to be fully evaluated and warrants further investigation.

Outer membrane protein (BP_03126) belonging to the general bacterial porin (GBP) family are found in the outer membrane of Gram-negative bacteria, which is responsible for antibiotics permeability.⁵³ Among the proteins of this family in *B. pertussis*, OmpP is the only porin identified as highly immunogenic, making it a promising and essential vaccine candidate with confirmed humoral immune responses in both animals and humans.⁴⁹ Further characterization of this outer membrane protein is necessary to confirm its identity and optimize its use in vaccine development.

Hemolysin activation/secretion protein, FhaC (BP_00399), is an outer membrane protein that acts as a transporter in the 2-partner secretion pathway (TPS). Its essentiality has been confirmed in the secretion of filamentous hemagglutinin (FHA).⁵⁴ Furthermore, Xu and Hu have noted that secretion systems have changed under vaccine pressure, and they suggested that targeting FhaC is an innovative approach.⁵⁵ Therefore, it is suggested that the immunogenicity of FhaC needs to be demonstrated.

The outer membrane usher protein, FimC (BP_00400), is predicted to be an integral outer membrane (OM) β -barrel usher protein involved in fimbriae (Fim2/3) assembly.⁵⁶ Despite the critical role of FimC in fimbriae assembly and its potential as a novel vaccine target to disrupt fimbriae assembly and reduce *B. pertussis* colonization, no studies to date have specifically investigated its immunogenicity.

The outer membrane protein (BP_00850), part of the multidrug efflux system, is a relevant element within the microbial repertoire that contributes to multidrug resistance (MDR) and the failure of anti-infectious therapies⁵⁷; thus, it is considered an attractive target for the development of new vaccines. It has been reported that the multidrug efflux system of *B. pertussis* exhibits low activity, which poses challenges for vaccine manufacturers due to the pathogen's sensitivity, potentially reducing the yield of biomass for vaccine production.⁵⁸

Protein involved in the biogenesis of cytochrome c (BP_00041), an essential component of bacterial energy metabolism, is critical for cellular function. The entire system appears to operate on the outer surface of the cytoplasmic membrane,⁵⁹ which limits its potential as a direct vaccine target, but it could be exploited for antimicrobial strategies and used as a potential drug target.

Conclusions

Unlike classical vaccinology methods, which rely on in vitro bacterial cultivation for antigen identification, reverse vaccinology leverages genomic analysis to predict vaccine targets more rapidly and cost-effectively, while enhancing specificity and minimizing risks of autoimmune reactions. These predicted targets subsequently undergo

immunogenicity assessments to precisely characterize the elicited immune responses, thereby facilitating advanced testing of potential candidates. In this study, 160 *B. pertussis* genomes, the etiological agent of whooping cough, were analyzed using bioinformatics approaches to identify novel vaccine targets. This methodology yielded 11 proteins as putative vaccine and drug targets, several of which require rigorous in vivo validation to confirm their immunogenicity, protective efficacy, and safety profile before integration into novel vaccine formulations. Collectively, the findings underscore the substantial relevance of this investigation to pertussis vaccine development and highlight the necessity of in vivo testing to optimize the application of these results.

ORCID iD

Majid Abdelouahab  <https://orcid.org/0009-0000-3670-9036>

Consent and Ethics

Not applicable (this is an in silico study only).

Author Contributions

Majid Abdelouahab: Conceptualization; Methodology; Data curation; Visualization; Writing – original draft; Investigation.

Diawara Idrissa: Validation; Supervision; Writing – review & editing.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplemental Material

Supplemental material for this article is available online.

References

- Guiso N, Bassinet L, Coqueluche. *EMC Mal Infect*. 2005;2:84-96. doi:10.1016/j.emcmi.2004.11.001
- WHO immunization data portal—Detail page. Immunization data. Accessed September 21, 2025. <https://immunization-data.who.int/global/wiise-detail-page>
- CDC. Pertussis in other countries. Whooping cough (Pertussis). June 26, 2024. Accessed September 21, 2025. <https://www.cdc.gov/pertussis/php/global/index.html>
- Bhattacharya S. Response to pertussis in the post-COVID-19 era: resurgence, diagnosis, and management. *Infect Chemother*. 2025;57:327. doi:10.3947/ic.2025.0010
- Burdin N, Handy LK, Plotkin SA. What is wrong with pertussis vaccine immunity? The problem of waning effectiveness of pertussis vaccines. *Cold Spring Harb Perspect Biol*. 2017;9:a029454. doi:10.1101/cshperspect.a029454
- Cherry JD. Why do pertussis vaccines fail. *Pediatrics*. 2012;129:968-970. doi:10.1542/peds.2011-2594
- Kim AR, Sette A, da Silva Antunes R. Adaptive immune response to *Bordetella pertussis* during vaccination and infection: emerging perspectives and unanswered questions. *Expert Rev Vaccines*. 2024;23:705-714. doi:10.1080/14760584.2024.2383745
- Warfel JM, Zimmerman LI, Merkel TJ. Acellular pertussis vaccines protect against disease but fail to prevent infection and transmission in a nonhuman primate model. *Proc Natl Acad Sci*. 2014;111:787-792. doi:10.1073/pnas.1314688110
- Parkhill J, Sebaihia M, Preston A, et al. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet*. 2003;35:32-40. doi:10.1038/ng1227
- Preston A, Parkhill J, Maskell DJ. The *Bordetellae*: lessons from genomics. *Nat Rev Microbiol*. 2004;2:379-390. doi:10.1038/nrmicro886
- Li Z, Xiao F, Hou Y, et al. Genomic epidemiology and evolution of *Bordetella pertussis* under the vaccination pressure of acellular vaccines in Beijing, China, 2020–2023. *Emerg Microbes Infect*. 2025;14:2447611. doi:10.1080/22221751.2024.2447611
- Vishweshwaraiah YL, Dokholyan NV. Toward rational vaccine engineering. *Adv Drug Deliv Rev*. 2022;183:114142. doi:10.1016/j.addr.2022.114142
- Rappuoli R. Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine*. 2001;19:2688-2691. doi:10.1016/S0264-410X(00)00554-5
- Monterrubio-López GP, Delgadillo-Gutiérrez K. [Reverse vaccinology: strategy against emerging pathogens]. *Rev Medica Inst Mex Seguro Soc*. 2021;59:233-241.
- Tettelin H, Riley D, Cattuto C, Medini D. Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol*. 2008;11:472-477. doi:10.1016/j.mib.2008.09.006
- Mira A, Martín-Cuadrado AB, D'Auria G, Rodríguez-Valera F. The bacterial pan-genome: a new paradigm in microbiology. *Int Microbiol*. 2010;13:45-57. doi:10.2436/20.1501.01.110
- Tettelin H, Medini D, eds. *The Pangenome: Diversity, Dynamics and Evolution of Genomes*. Springer International Publishing; 2020.
- Kodama Y, Shumway M, Leinonen R; International Nucleotide Sequence Database Collaboration. The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Res*. 2012;40:D54-D56. doi:10.1093/nar/gkr854
- Leggett RM, Ramirez-Gonzalez RH, Clavijo BJ, Waite D, Davey RP. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Front Genet*. 2013;4:288. doi:10.3389/fgene.2013.00288
- Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34:i884-i890. doi:10.1093/bioinformatics/bty560
- Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo assembler. *Curr Protoc Bioinformatics*. 2020;70:e102. doi:10.1002/cpbi.102
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29:1072-1075. doi:10.1093/bioinformatics/btt086
- Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:2068-2069. doi:10.1093/bioinformatics/btu153
- Sitto F, Battistuzzi FU. Estimating pangenomes with roary. *Mol Biol Evol*. 2020;37:933-939. doi:10.1093/molbev/msz284
- Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res*. 2010;38:D355-360. doi:10.1093/nar/gkp896
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403-410. doi:10.1016/S0022-2836(05)80360-2

27. Zhang R. DEG: a database of essential genes. *Nucleic Acids Res.* 2004;32:271D-272D. doi:10.1093/nar/gkh024
28. Gardy JL, Laird MR, Chen F, et al. PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics.* 2005;21:617-623. doi:10.1093/bioinformatics/bti057
29. Yu C, Chen Y, Lu C, Hwang J. Prediction of protein subcellular localization. *Proteins Struct Funct Bioinforma.* 2006;64:643-651. doi:10.1002/prot.21018
30. Petersen TN, Brunak S, Von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 2011;8:785-786. doi:10.1038/nmeth.1701
31. Agüero F, Al-Lazikani B, Aslett M, et al. Genomic-scale prioritization of drug targets: the TDR Targets database. *Nat Rev Drug Discov.* 2008;7:900-907. doi:10.1038/nrd2684
32. Duffield M, Cooper I, McAlister E, Bayliss M, Ford D, Oyston P. Predicting conserved essential genes in bacteria: in silico identification of putative drug targets. *Mol Biosyst.* 2010;6:2482-2489. doi:10.1039/c0mb00001a
33. Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server. In: Walker JM, ed. *The Proteomics Protocols Handbook*. Humana Press; 2005:571-607.
34. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001;305:567-580. doi:10.1006/jmbi.2000.4315
35. Rizwan M, Naz A, Ahmad J, et al. VacSol: a high throughput in silico pipeline to predict potential therapeutic targets in prokaryotic pathogens using subtractive reverse vaccinology. *BMC Bioinformatics.* 2017;18:106. doi:10.1186/s12859-017-1540-0
36. Barh D, Tiwari S, Jain N, et al. In silico subtractive genomics for target identification in human bacterial pathogens. *Drug Dev Res.* 2011;72:162-177. doi:10.1002/ddr.20413
37. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics.* 2007;8:4. doi:10.1186/1471-2105-8-4
38. Xiang Z, He Y. Genome-wide prediction of vaccine targets for human herpes simplex viruses using Vaxign reverse vaccinology. *BMC Bioinformatics.* 2013;14 suppl 4:S2. doi:10.1186/1471-2105-14-S4-S2
39. Sun P, Ju H, Liu Z, et al. Bioinformatics resources and tools for conformational B-cell epitope prediction. *Comput Math Methods Med.* 2013;2013:943636. doi:10.1155/2013/943636
40. EL-Manzalawy Y, Honavar V. Recent advances in B-cell epitope prediction methods. *Immunome Res.* 2010;6:S2. doi:10.1186/1745-7580-6-S2-S2
41. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.* 2017;45:W24-W29. doi:10.1093/nar/gkx346
42. Waterhouse AM, Studer G, Robin X, Bienert S, Tauriello G, Schwede T. The structure assessment web server: for proteins, complexes and more. *Nucleic Acids Res.* 2024;52:W318-W323. doi:10.1093/nar/gkae270
43. Kant R, Khan MS, Chopra M, Saluja D. Artificial intelligence-driven reverse vaccinology for *Neisseria gonorrhoeae* vaccine: prioritizing epitope-based candidates. *Front Mol Biosci.* 2024;11:1442158. doi:10.3389/fmolb.2024.1442158
44. Yuan B, Cheng A, Wang M. Polysaccharide export outer membrane proteins in gram-negative bacteria. *Future Microbiol.* 2013;8:525-535. doi:10.2217/fmb.13.13
45. Sakharkar KR, Sakharkar MK, Chow VT. A novel genomics approach for the identification of drug targets in pathogens, with special reference to *Pseudomonas aeruginosa*. *In Silico Biol.* 2004;4:355-360.
46. Karp PD, Krummenacker M, Paley S, Wagg J. Integrated pathway-genome databases and their role in drug discovery. *Trends Biotechnol.* 1999;17:275-281. doi:10.1016/S0167-7799(99)01316-5
47. Mattoo S, Cherry JD. Molecular pathogenesis, epidemiology, and clinical manifestations of respiratory infections due to *Bordetella pertussis* and other *Bordetella* subspecies. *Clin Microbiol Rev.* 2005;18:326-382. doi:10.1128/CMR.18.2.326-382.2005
48. Dorji D, Graham RM, Singh AK, Ramsay JP, Price P, Lee S. Immunogenicity and protective potential of *Bordetella pertussis* biofilm and its associated antigens in a murine model. *Cell Immunol.* 2019;337:42-47. doi:10.1016/j.cellimm.2019.01.006
49. Gregg KA, Wang Y, Warfel J, et al. Antigen discovery for next-generation pertussis vaccines using immunoproteomics and transposon-directed insertion sequencing. *J Infect Dis.* 2023;227:583-591. doi:10.1093/infdis/jiac502
50. Hausman SZ, Burns DL. Use of pertussis toxin encoded by *ptx* genes from *Bordetella bronchiseptica* to model the effects of antigenic drift of pertussis toxin on antibody neutralization. *Infect Immun.* 2000;68:3763-3767. doi:10.1128/IAI.68.6.3763-3767.2000
51. Mooi FR, He Q, van Oirschot H, Mertsola J. Variation in the *Bordetella pertussis* virulence factors pertussis toxin and pertactin in vaccine strains and clinical isolates in Finland. *Infect Immun.* 1999;67:3133-3134. doi:10.1128/IAI.67.6.3133-3134.1999
52. Lundquist K, Billings E, Bi M, Wellnitz J, Noinaj N. The assembly of β -barrel membrane proteins by BAM and SAM. *Mol Microbiol.* 2021;115:425-435. doi:10.1111/mmi.14666
53. Kerek Á, Román I, Szabó Á, et al. Antibiotic resistance genes in *Escherichia coli*—literature review. *Crit Rev Microbiol.* 2025;52:1-35. doi:10.1080/1040841X.2025.2492156
54. Melvin JA, Scheller EV, Noël CR, Cotter PA. New insight into filamentous hemagglutinin secretion reveals a role for full-length FhaB in *Bordetella* virulence. *mBio.* 2015;6:e01189-01115. doi:10.1128/mBio.01189-15
55. Xu Z, Hu D, Lu LDW, et al. Genomic dissection of the microevolution of Australian epidemic *Bordetella pertussis*. *Emerg Microbes Infect.* 2022;11:1460-1473. doi:10.1080/22221751.2022.2077129
56. Melvin JA, Scheller EV, Miller JF, Cotter PA. *Bordetella pertussis* pathogenesis: current and future challenges. *Nat Rev Microbiol.* 2014;12:274-288. doi:10.1038/nrmicro3235
57. Alcalde-Rico M, Hernando-Amado S, Blanco P, Martínez JL. Multidrug Efflux pumps at the crossroad between antibiotic resistance and bacterial virulence. *Front Microbiol.* 2016;7:1483. doi:10.3389/fmicb.2016.01483
58. MacArthur I, Belcher T, King JD, Ramasamy V, Alhammadi M, Preston A. The evolution of *Bordetella pertussis* has selected for mutations of *acr* that lead to sensitivity to hydrophobic molecules and fatty acids. *Emerg Microbes Infect.* 2019;8:603-612. doi:10.1080/22221751.2019.1601502
59. Beckett CS, Loughman JA, Karberg KA, Donato GM, Goldman WE, Kranz RG. Four genes are required for the system II cytochrome *c* biogenesis pathway in *Bordetella pertussis*, a unique bacterial model. *Mol Microbiol.* 2000;38:465-481. doi:10.1046/j.1365-2958.2000.02174.x