Process Biochemistry 122 (2022) 128–148

Contents lists available at ScienceDirect

Process Biochemistry

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

An *in silico* reverse vaccinology study of *Brachyspira pilosicoli*, the causative organism of intestinal spirochaetosis, to identify putative vaccine candidates

Myron Christodoulides [a,*], Daysiane de Oliveira [b], David W. Cleary [a,c], Maria Victoria Humbert [a], Ricardo A. Machado-de-Ávila [b], Roberto M. La Ragione [d]

[a] *Molecular Microbiology, School of Clinical and Experimental Sciences, University of Southampton Faculty of Medicine, Southampton General Hospital, Southampton, England SO16 6YD, UK,*
[b] *Programa de Pós-Graduação em Ciências da Saúde, Universidade do Extremo Sul Catarinense, Criciúma, 88806-000 Santa Catarina, Brazil*
[c] *NIHR Southampton Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust, Southampton, UK*
[d] *Department of Pathology and Infectious Diseases, School of Veterinary Medicine, Faculty of Health and Medical Sciences, University of Surrey, Guildford, England GU2 7AL, UK*

ARTICLE INFO

Keywords:
Brachyspira pilosicoli
Vaccine
Reverse vaccinology
in silico
B-cell chimera

ABSTRACT

Brachyspira pilosicoli is a zoonotic bacterium that can cause intestinal spirochaetosis (IS) in avian species (AIS), pigs (PIS) and humans (HIS). In the absence of vaccines to prevent infections, we used genome-based reverse vaccinology (RV) to identify putative *B. pilosicoli* vaccine candidates. Genome sequence of *B. pilosicoli* strain B2904, an AIS isolate, was analysed with PSORTb3, CELLO, $SOSUI_{GramN}$, LipoP, SignalP-5.0, TMHMM, BLAST 2.12.0 +, PDB database, SEED Viewer, eggNOG-mapper, UniProt, VaxiJen and Vaxign2, and Tblastn to generate a RV list of putative vaccine candidates. We also generated a linear B-cell chimera antigen using Blast-p, Emini Surface Accessibility Prediction, ABCpred, Expasy ProtParam and PepCalc programs. RV defined a list of 162 proteins containing 48 Outer Membrane (OM), 27 OM/Extracellular, 27 Extracellular, 4 Periplasm, 2 Surface, 2 Cytoplasm and 52 Unknown proteins. The list was characterised by an abundance of SPII lipoproteins. We found that genes encoding amino acid sequences of 146/162 (90%) proteins were present in 19 other *B. pilosicoli* genomes. A linear B-cell chimera antigen was generated from the amino acid sequences of 18 OM and Extracellular proteins. Our contemporary RV study represents a starting point for a comprehensive vaccine development strategy for preventing intestinal spirochaetosis.

1. Introduction

Brachyspira (previously *Treponema*, *Serpula* and *Serpulina*) is the sole genus of the family *Brachyspiraceae* within the order Spirochaetales, phylum Spirochaetes [1] and contains nine officially recognised species [2]. *Brachyspira (B.) spp.* are flagellated, anaerobic, aero-tolerant, slow-growing Gram-negative spirochaetes that inhabit the large intestines of animals and birds, where they are intimately associated with the colonic or caecal mucosa. Several of these species are pathogenic to primarily pigs and poultry but can also infect other animals and humans. Infection with *B. hyodysenteriae* (classical agent), *B. hampsonii* or *B. suanatina* causes swine dysentery (SD), a severe colitis in pigs [3–5]. *B. murdochii* and *B. pilosicoli* also infect pigs but cause milder colitis symptoms [6]. Infection of chickens with either *B. intermedia* or *B. alvinipulli* causes avian intestinal spirochaetosis (AIS) [7]. *B. innocens* is an enteric commensal of pigs, chickens and rats, and no disease has been associated with this species [2]. However, *B. innocens* infection was associated with poor performance and below target egg production in free-range flocks [8]. *B. aalborgi* has been reported to cause histologically confirmed intestinal spirochaetosis (IS) in humans (HIS) [9].

By contrast to the other *Brachyspira spp., B. pilosicoli* has a very broad host range [2,10], and it is capable of infecting chickens to cause AIS [11,12], as well as wild ducks [13,14], domesticated turkeys [15], pheasants [16,17], rodents (mice, rats and guinea pigs) [18,19], dogs [20–22], horses [23], zoo birds, marsupials, opossums and non-human primates [2]. *B. pilosicoli* is the sole causative agent of porcine

* Corresponding author.
E-mail address: mc4@soton.ac.uk (M. Christodoulides).

https://doi.org/10.1016/j.procbio.2022.08.014
Received 16 March 2022; Received in revised form 26 July 2022; Accepted 11 August 2022
Available online 18 August 2022

intestinal spirochaetosis (PIS), which is distinct from the more severe swine dysentery [24]. PIS is characterised by diarrhoea and poor growth [10,24], and AIS in chickens is associated with production of wet faeces, bloody faeces, diarrhoea, lower growth rate, a late-onset of egg laying, faecally contaminated eggs and a reduction in the numbers and quality of eggs laid, as well as lethargy and depression [10–12]. Severe clinical symptoms can result in increased flock mortality. *B. pilosicoli* would appear to have global prevalence and spirochaetal infections have been reported in the UK, continental Europe, Scandinavia, North America, Oceania, Iran, Malaysia and South America [12,25–29]. AIS and PIS are under-reported diseases and there is likely to be a significant economic impact in food production, globally. No economic analysis of the overall costs of PIS to the pig industry worldwide is available, to our knowledge, whereas the economic loss to the poultry industry associated with AIS has been estimated to be approximately £ 18 million per year in the UK [12]. Extrapolation of these losses suggest that combined economic losses to both industries globally may be in the region of 1–2 billion USD annually.

The zoonotic potential of *B. pilosicoli* has been recognised [10,23] and human gut colonisation has been reported [30,31]. Individuals diagnosed with HIS may have one or more non-specific clinical symptoms, e.g. abdominal pain, change in bowel habits, pseudo-appendicitis, irritable bowel, diverticulitis, chronic diarrhoea and rectal bleeding [32–36]. Risk factors associated with zoonotic infection of humans by *B. pilosicoli* include exposure to, and ingestion of, faecal contaminated water as potentially the most important route of transmission [10,13,23,37], living rurally and/or among animals, crowding, socioeconomic depression, travel to-and-from less economically developed countries, and positive HIV status [2]. *B. pilosicoli* is regarded as a potential human enteric pathogen and infection may be under reported. In part, this is due to its specific growth requirements, and, because it grows slowly, it is often overgrown by other more rapidly growing bacteria of the intestinal microbiota, which hampers its detection using routine culture based diagnostic methods [2].

Treatment of AIS, PIS and HIS involves the use of antibiotics, but resistance has been observed [38]. Current antibiotic therapy for HIS involves co-amoxicilline and metronidazole, and pleuromutilins (notably tiamulin), and macrolides and lincosamides are currently used for AIS and PIS [12]. Good animal husbandry is also useful for reducing zoonotic transmission [2] and other potential treatments include the use of probiotics, e.g. oral treatment of chickens with *Lactobacillus reuteri* has been shown to reduce *B. pilosicoli*-induced pathology and colonisation [39]. Ideally, prevention would be preferred for livestock animals, but there are no commercially available vaccines to prevent AIS or PIS. Previous experimental vaccine strategies have included testing recombinant proteins with Freund's Incomplete Adjuvant and testing bacterin, i.e. formalised whole bacteria cells, in chicken, mouse and pig models. In addition, in the early to mid-2000's, Novartis Animal Vaccines (NAV) funded the first sequencing of a *B. pilosicoli* strain (95/100) at Murdoch University in Perth, Western Australia, aiming to subject it to reverse vaccinology (RV) to identify putative vaccine candidates [40]. A patent covers seven high priority candidates identified form this RV (https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2007113001), but further animal studies were not done due to lack of funding. Thus, experimental vaccines and identified antigens have not been translated to commercial products and this serves to highlight the paucity of knowledge of candidate *B. pilosicoli* vaccine antigens.

In the current study we have used broader and more contemporary genome-and proteome-based Reverse Vaccinology (RV) to identify the surface-exposed proteome (i.e. the proteins that are potentially able to be expressed) of a reference *B. pilosicoli* strain B2904 that causes AIS, and used this information to predict putative vaccine candidates. RV uses the complete genomic information of a microorganism to inform the complete antigen repertoire, from which vaccine candidates can be selected by using bioinformatic algorithms [41]. RV was first reported in 2000 by Rappuoli and colleagues at Novartis for the discovery of potential antigens from the genetic information of a single strain of serogroup B *Neisseria meningitidis* (MenB), which led to the development of the Bexsero/4CMenB vaccine to prevent meningococcal infections [41]. Based on the RV data generated, we produced *in silico* a B-cell chimera antigen from predicted linear B-cell epitopes of a selection of candidate proteins.

2. Materials and methods

2.1. In silico analysis and computational tools for RV

The whole genome sequence for a representative reference *B. pilosicoli* (strain B2904) isolate causing AIS was downloaded from the NCBI GenBank webpage [42] and uploaded onto RAST annotation server [43]. The *B. pilosicoli* NCBI accession number is NC_018607.1 The latest genome assembly and annotated genome and protein amino acid sequences have all remained unchanged since 2016. PSORTb3 [44] was used to identify subcellular localization prediction based on amino acid sequence identity, and then CELLO [45] and $SOSUI_{GramN}$ [46] were used to provide additional information that was not provided by PSORTb3. A consensus localization was then predicted for the proteins from these three programs, where two or more agreed. LipoP server [47] was used to predict lipoproteins and SignalP-5.0 [48,49] to predict signal peptides and their probable cleavage site in secreted proteins. Proteins identified as SPII peptidase specific were renamed as lipoproteins. In general, proteins with SPI peptidase sequences were excluded. Presence or absence of aromatic amino acid residues in the C-terminal region of proteins was checked manually, as this may suggest probable β-barrel structure, characteristic of membrane proteins [50]. Predicted β-barrel integral membrane proteins will likely be difficult to express/purify and were considered for exclusion from the candidate list. We used the TMHMM Server 2.0 program to predict transmembrane helices [51].

For those proteins with unknown localization and/or function, amino acid sequences were also compared against the whole non-redundant GenBank database excluding Brachyspirales using BLAST [52,53], and against the PDB database [54] in order to gain some insight from known PDB structures. Hypothetical/unknown protein sequences were also searched for descriptors such as 'cell surface', 'virulence', 'outer', 'secreted', 'extracellular', 'export', 'binding' and 'toxin', and retained and analysed further. Proteins containing 1, 2 or 4 cysteine residues were retained, whereas proteins with 3 or 5 or more cysteine residues were removed; although the latter help to stabilise protein structure, their presence may originate problems in recombinant protein expression [55]. The *B. pilosicoli* annotated genome was also browsed and compared against a non-pathogenic *Escherichia coli* K12 genome and a genome of the commensal *Lactobacillus reuteri* (F275) that is used as a probiotic supplement for poultry [39], using the SEED Viewer software [56]. In general, though not exclusively, the few proteins like other homologues in *E. coli* and *L. reuteri* were excluded. All proteins predicted as cytoplasmic, cytoplasmic/inner membrane, smaller than 150 amino acids in length were also excluded. The protein sequences were also run through eggNOG-mapper [57], which combines Cluster of Orthologous Genes and Gene Ontology terms to provide potential information on putative preferred names and functions, especially of hypothetical proteins. Further searches in the literature (PubMed), and in UniProt [58], and with Blast-p for homology searching, provided additional information for many of the identified proteins.

Allergenic proteins and allergenic regions in a protein were predicted *in silico* using the AlgPred 2.0 web server, with a cut-off value of 0.5 [59]. Hypothetical/unknown protein amino acid sequences were also run through the STRING database [60] and the NCBI's Conserved Domain Database (CDD) [61].

There has been very little genomic data available for *B. pilosicoli*. Recently, we introduced 12 newly sequenced draft genome assemblies to the literature, a 4-fold increase in the number of examined isolates [62]. These genomes [62] and 7 other genomes available in Genbank

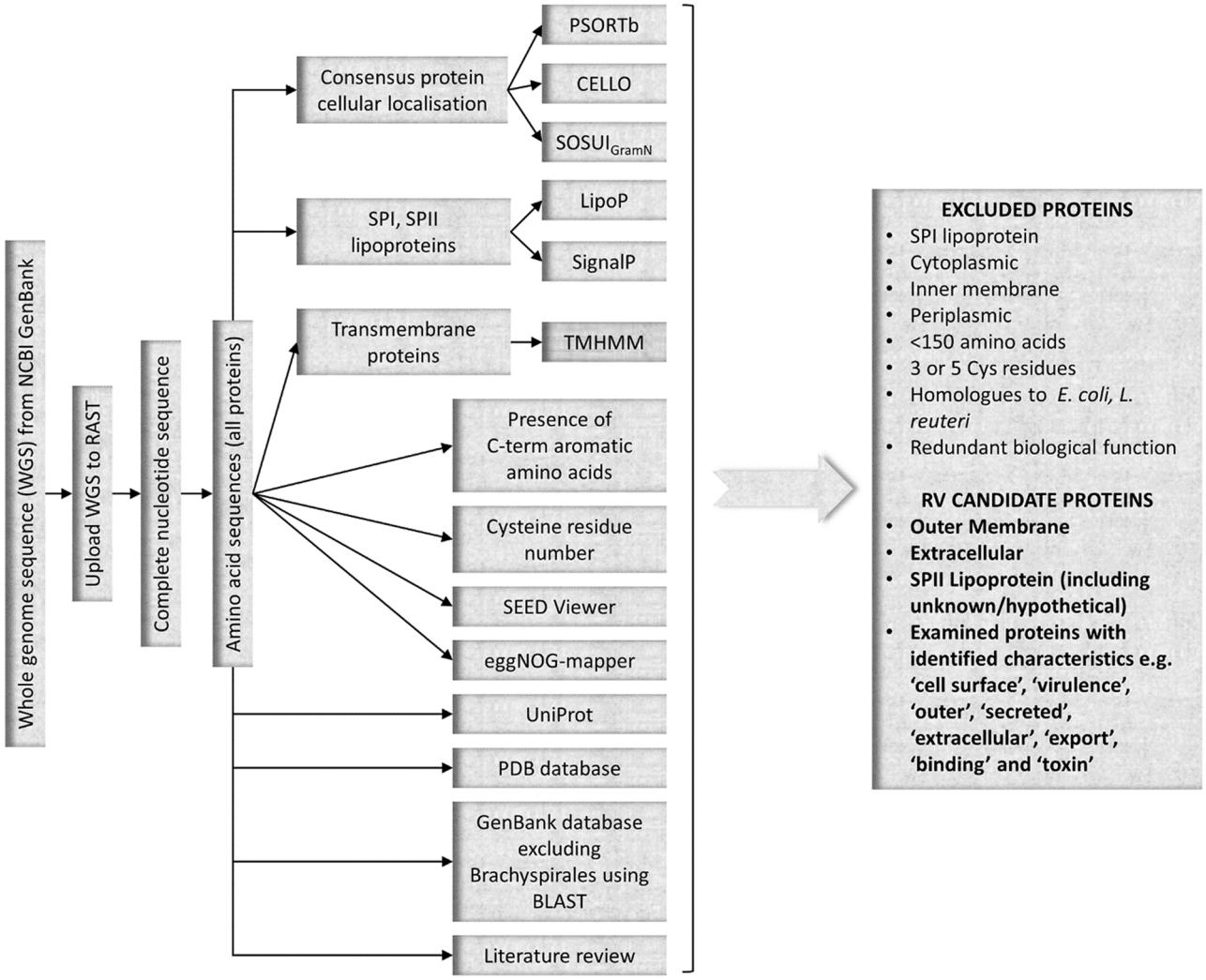

Fig. 1. The scheme outlining the bioinformatics tools used in this study. The scheme starts from a whole genome sequence that is translated to output the organism proteome, which is then analysed with a variety of bioinformatics programs to produce a final list of candidate proteins that can be explored as potential vaccine antigens for AIS, PIS and potentially HIS.

were searched for the presence of genes encoding for the candidate protein amino acid sequences from *B. pilosocoli* strain B2904 identified by RV. Each protein amino acid sequence was searched against these 19 genomes that are present in the *Brachyspira pilosicoli* (taxid:52584) RefSeq Genome Database (refseq_genomes) Genbank, using Tblastn program.

2.2. VaxiJen and Vaxign2 analyses

Protein amino acid sequences identified by the RV workflow were also analysed further with VaxiJen: Prediction of Protective Antigens and Subunit Vaccines program for alignment-independent prediction of protective antigens [63]. Antigen classification with VaxiJen is based solely on the physicochemical properties of proteins without depending on sequence alignments. A threshold value of > 0.4 was computed as this showed the highest accuracy for predicting immunogens with the models [63].

The same protein amino acid sequences identified by RV were analysed with Vaxign2 (Vaccine Design) Dynamic Vaxign Analysis tool, which is a vaccine target prediction and analysis system based on RV principles including localisation (PSORTb), transmembrane helices (TMHMM), adhesion probability (SPAAN), signal peptide (SignalP), epitope prediction with IEDB (The Immune Epitope Database and Analysis Resource) [64] and physicochemical properties (Propy). Vaxign2 presents Vaxign-ML, a pipeline for machine learning-based Vaccine Target Prediction, which give 'Protegenicity' scores, i.e. percentage probabilities that the antigens are suitable vaccine candidates. Vaxign2 also determined any similarities to human, mouse and pig host proteins, using BLAST [65]. Any similarities to chicken host proteins were also done using BLAST against chickens (taxid:9031).

2.3. Linear B-cell epitope prediction and construction of a B. pilosicoli B-cell chimera antigen

Selected proteins from the RV were analysed to identify linear B-cell epitopes to develop a chimera antigen. Each selected protein was analysed with the following workflow: 1) with Blast-p to check for similarity with other proteins; 2) with the Emini Surface Accessibility Prediction program, highlighting amino acids with a score > 1.0 [66]; 3) ABCpred (Artificial neural network based B-cell epitope prediction server, underlining epitopes with a score >0.85 [67]; 4) analyses of the results from both programs and choosing a common epitope; 5) constructing a chimera protein using all of the epitopes selected. The order of the peptide amino acid sequences were chosen based on their position in their original protein of origin, i.e. epitopes closer to the C-terminus region in the original protein were put closer to the C-terminus of the

Table 1
Final list of *B. pilosicoli* proteins identified by reverse vaccinology.

SeqID	Feature ID *B. pilosicoli* B2904	Putative Name and Function//eggNOG-mapper preferred name	Consensus Localization	Protein characteristic (if known)	Amino Acids	Molecular Weight (kDa)
247	fig|6666666.171271.peg.247	*C. botulinum* toxin (265 homology over short distance)	Cytoplasm	Toxin	428	50.35
58	fig|6666666.171271.peg.58	Flagellar basal-body rod protein FlgF//FlhO, flagellar basal-body rod protein	Cytoplasm	Extracellular	282	32.18
2325	fig|6666666.171271.peg.2325	Flagellar motor switch protein FliN	Extracellular	Unknown	348	36.52
1751	fig|6666666.171271.peg.1751	*Serpulina (Brachyspira) hyodysenteriae* variable surface /protein_	Extracellular	Surface	197	21.29
2319	fig|6666666.171271.peg.2319	FIG137478: Hypothetical protein/weak sequence similarity to Ring-infected erythrocyte surface antigen *P. falciparum*	Extracellular	Surface	1057	122.57
2125	fig|6666666.171271.peg.2125	Flagellar hook-length control protein FliK	Extracellular	Secreted	505	57.27
2126	fig|6666666.171271.peg.2126	Flagellar basal-body rod modification protein FlgD	Extracellular	Secreted	233	25.65
2573	fig|6666666.171271.peg.2573	Flagellin protein FlaA//FlaB, component of the core of the flagella	Extracellular	Periplasm	286	31.33
1843	fig|6666666.171271.peg.1843	Hypothetical protein	Extracellular	Lipoprotein (SPII)	168	19.25
382	fig|6666666.171271.peg.382	Cell surface protein//Leucine rich repeats (6 copies)	Extracellular	Lipoprotein (SPII)	182	20.64
400	fig|6666666.171271.peg.400	Hypothetical protein//GlgB, 1,4-alpha-glucan branching enzyme activity	Extracellular	Lipoprotein (SPII)	246	28.97
2275	fig|6666666.171271.peg.2275	Hypothetical protein	Extracellular	Lipoprotein (SPII)	299	34.15
1066	fig|6666666.171271.peg.1066	FIG00439184: hypothetical protein//amino acid activation for non-ribosomal peptide biosynthetic process	Extracellular	Lipoprotein (SPII)	312	32.68
1309	fig|6666666.171271.peg.1309	FIG00438996: hypothetical protein	Extracellular	Lipoprotein (SPII)	341	39.17
1802	fig|6666666.171271.peg.1802	Sialidase (EC 3.2.1.18)	Extracellular	Lipoprotein (SPII)	455	47.4
1804	fig|6666666.171271.peg.1804	Sialidase (EC 3.2.1.18)	Extracellular	Lipoprotein (SPII)	472	49.82
2490	fig|6666666.171271.peg.2490	Hypothetical protein	Extracellular	Lipoprotein (SPII)	136	14.23
2016	fig|6666666.171271.peg.2016	Hypothetical protein	Extracellular	Lipoprotein (SPII)	158	17.05
1886	fig|6666666.171271.peg.1886	Hypothetical protein	Extracellular	Lipoprotein (SPII)	261	27.83
2574	fig|6666666.171271.peg.2574	Hypothetical protein	Extracellular	Lipoprotein (SPII)	296	33.4
1398	fig|6666666.171271.peg.1398	Hypothetical protein/contains *a* SCP-like extracellular protein domain, found in virulence-associated extracellular proteins/ *B. subtilis* extracellular protein 30/112 27%//Cysteine-rich secretory protein family	Extracellular	Extracellular	257	28.58
1638	fig|6666666.171271.peg.1638	Hypothetical protein	Extracellular	Extracellular	205	23.09
1838	fig|6666666.171271.peg.1838	Flagellar basal-body rod protein FlgG	Extracellular	Extracellular	264	28.67
2557	fig|6666666.171271.peg.2557	FIG00438166: hypothetical protein	Extracellular	Extracellular	296	34.17
1949	fig|6666666.171271.peg.1949	Hypothetical protein	Extracellular	Extracellular	361	38.22
2646	fig|6666666.171271.peg.2646	Peptidase M30, hyicolysin//cell wall binding repeat	Extracellular	Extracellular	386	44.85
1332	fig|6666666.171271.peg.1332	Flagellar hook-associated protein FlgL	Extracellular	Extracellular	417	46.32
2145	fig|6666666.171271.peg.2145	Flagellar hook protein FlgE	Extracellular	Extracellular	441	47.11
1312	fig|6666666.171271.peg.1312	Flagellar hook-associated protein FliD	Extracellular	Extracellular	735	82.47
2167	fig|6666666.171271.peg.2167	Flagellar filament outer layer protein FlaA, putative	Outer Membrane	Unknown	275	31.36
1917	fig|6666666.171271.peg.1917	Hemolysin	Outer Membrane	Secreted	346	40.14
629	fig|6666666.171271.peg.629	Outer membrane protein assembly factor YaeT precursor//TolC, efflux transmembrane transporter activity	Outer Membrane	OMP	484	54.71
2449	fig|6666666.171271.peg.2449	Hypothetical protein	Outer Membrane	OMP	201	21.94
228	fig|6666666.171271.peg.228	FIG00438828: hypothetical protein	Outer Membrane	OMP	222	26.17
1002	fig|6666666.171271.peg.1002	FIG00437507: hypothetical protein	Outer Membrane	OMP	225	26.58

(*continued on next page*)

Table 1 (*continued*)

SeqID	Feature ID *B. pilosicoli* B2904	Putative Name and Function//eggNOG-mapper preferred name	Consensus Localization	Protein characteristic (if known)	Amino Acids	Molecular Weight (kDa)
1702	fig|6666666.171271.peg.1702	FIG00437470: hypothetical protein	Outer Membrane	OMP	230	25.66
410	fig|6666666.171271.peg.410	FIG00437886: hypothetical protein	Outer Membrane	OMP	234	25.67
559	fig|6666666.171271.peg.559	Hypothetical protein	Outer Membrane	OMP	240	27.76
1390	fig|6666666.171271.peg.1390	Hypothetical protein	Outer Membrane	OMP	242	28.35
977	fig|6666666.171271.peg.977	Hypothetical protein	Outer Membrane	OMP	254	28.63
558	fig|6666666.171271.peg.558	Hypothetical protein	Outer Membrane	OMP	255	29.24
976	fig|6666666.171271.peg.976	Hypothetical protein	Outer Membrane	OMP	256	29.24
1235	fig|6666666.171271.peg.1235	FIG00438510: hypothetical protein	Outer Membrane	OMP	261	29.94
1317	fig|6666666.171271.peg.1317	FIG00438006: hypothetical protein	Outer Membrane	OMP	300	34.37
448	fig|6666666.171271.peg.448	*B. burgdorferi* predicted coding region BBJ25	Outer Membrane	OMP	334	38.54
455	fig|6666666.171271.peg.455	*B. burgdorferi* predicted coding region BBJ29	Outer Membrane	OMP	351	41.9
1517	fig|6666666.171271.peg.1517	FIG00438343: hypothetical protein	Outer Membrane	OMP	361	40.6
840	fig|6666666.171271.peg.840	Peptidoglycan-binding protein LysM//TmpB, Treponemal (spirochaetal) membrane protein	Outer Membrane	OMP	377	42.47
387	fig|6666666.171271.peg.387	*Serpulina (Brachypsira) hyodysenteriae* variable surface /protein	Outer Membrane	OMP	386	42.05
227	fig|6666666.171271.peg.227	FIG00438231: hypothetical protein	Outer Membrane	OMP	438	47.23
44	fig|6666666.171271.peg.44	Putative outer membrane efflux protein	Outer Membrane	OMP	452	51.76
1529	fig|6666666.171271.peg.1529	FIG00439356: hypothetical protein//CsgG curli production assembly transport component	Outer Membrane	OMP	489	55.21
2554	fig|6666666.171271.peg.2554	TolA protein//TmpB, Treponemal (spirochaetal) membrane protein	Outer Membrane	OMP	566	62.43
239	fig|6666666.171271.peg.239	FIG00437458: hypothetical protein//Tetratricopeptide repeats	Outer Membrane	OMP	645	74.89
2019	fig|6666666.171271.peg.2019	Outer membrane vitamin B12 receptor BtuB//cobalamin-transporting ATPase activity	Outer Membrane	OMP	647	74.67
785	fig|6666666.171271.peg.785	FIG00439155: hypothetical protein	Outer Membrane	OMP	683	76.65
938	fig|6666666.171271.peg.938	Hypothetical protein	Outer Membrane	OMP	763	86.62
2620	fig|6666666.171271.peg.2620	Outer membrane protein assembly factor YaeT precursor	Outer Membrane	OMP	872	100.23
2235	fig|6666666.171271.peg.2235	TreP; involved in the TonB-independent uptake of proteins	Outer Membrane	OMP	911	106.12
1303	fig|6666666.171271.peg.1303	TPR Domain containing protein	Outer Membrane	OMP	916	107.17
1904	fig|6666666.171271.peg.1904	FIG00438929: hypothetical protein//lipopolysaccharide transport	Outer Membrane	OMP	967	111.64
980	fig|6666666.171271.peg.980	Putative exported protein	Outer Membrane	OMP	1010	113.22
693	fig|6666666.171271.peg.693	Cell division protein FtsH (EC 3.4.24.-)//CiaB, TPR repeat	Outer Membrane	OMP	1118	128.74
2621	fig|6666666.171271.peg.2621	FIG00437896: hypothetical protein//protein secretion	Outer Membrane	OMP	1177	132.26
1221	fig|6666666.171271.peg.1221	UPF0192 protein all5100 precursor//Spirochaete alpha-2-macroglobulin family	Outer Membrane	OMP	1893	215.17
2232	fig|6666666.171271.peg.2232	Tia invasion determinant virulence//Has lipid A 3-O-deacylase activity.	Outer Membrane	OMP	205	23.57
2472	fig|6666666.171271.peg.2472	OmpA superfamily similarities//Flagellar Motor Protein	Outer Membrane	OMP	139	16.2
1506	fig|6666666.171271.peg.1506	SusC/RagA family TonB-linked outer membrane protein	Outer Membrane	OMP	218	23.98
2119	fig|6666666.171271.peg.2119	SusC/RagA family TonB-linked outer membrane protein (88%)	Outer Membrane	OMP	375	43.24
1434	fig|6666666.171271.peg.1434	OstA family protein//LptA, lipopolysaccharide binding	Outer Membrane	OMP	427	49.27
624	fig|6666666.171271.peg.624	TonB dependent receptor//cobalamin-transporting ATPase activity	Outer Membrane	OMP	445	51.41
2067	fig|6666666.171271.peg.2067	Phosphate-selective porin O and P superfamily protein (71%)	Outer Membrane	OMP	327	39.34

(*continued on next page*)

Table 1 (*continued*)

SeqID	Feature ID *B. pilosicoli* B2904	Putative Name and Function//eggNOG-mapper preferred name	Consensus Localization	Protein characteristic (if known)	Amino Acids	Molecular Weight (kDa)
1723	fig|6666666.171271.peg.1723	Hypothetical protein	Outer Membrane	Lipoprotein (SPII)	356	41.62
1087	fig|6666666.171271.peg.1087	Hypothetical protein	Outer Membrane	Lipoprotein (SPII)	485	56.9
1636	fig|6666666.171271.peg.1636	FIG00437393: hypothetical protein	Outer Membrane	Lipoprotein (SPII)	587	66.63
1387	fig|6666666.171271.peg.1387	FIG00437743: hypothetical protein	Outer Membrane	Lipoprotein (SPII)	912	106.49
1187	fig|6666666.171271.peg.1187	FIG00439107: hypothetical protein	Outer Membrane	Extracellular	397	46.05
597	fig|6666666.171271.peg.597	FlgA flagellar protein *Thermatoga* 22/76 (29%)	Outer Membrane/ Extracellular	Unknown	220	25.62
1829	fig|6666666.171271.peg.1829	Flagellar hook-associated protein FlgK (64%)	Outer Membrane/ Extracellular	Unknown	222	24.62
1294	fig|6666666.171271.peg.1294	BatC//PFAM Sporulation and spore germination	Outer Membrane/ Extracellular	Unknown	321	37.46
1728	fig|6666666.171271.peg.1728	Membrane protein containing Diverse 7TM receptor, extracellular region 2 (74%)	Outer Membrane/ Extracellular	Membrane	443	52.17
81	fig|6666666.171271.peg.81	Hypothetical protein	Outer Membrane/ Extracellular	Lipoprotein (SPII)	163	18.32
71	fig|6666666.171271.peg.71	Hypothetical protein	Outer Membrane/ Extracellular	Lipoprotein (SPII)	164	18.41
2229	fig|6666666.171271.peg.2229	FIG00437836: hypothetical protein	Outer Membrane/ Extracellular	Lipoprotein (SPII)	175	18.98
2017	fig|6666666.171271.peg.2017	Hypothetical protein	Outer Membrane/ Extracellular	Lipoprotein (SPII)	176	19.94
2312	fig|6666666.171271.peg.2312	Hypothetical protein	Outer Membrane/ Extracellular	Lipoprotein (SPII)	176	20.5
1435	fig|6666666.171271.peg.1435	FIG00437342: hypothetical protein	Outer Membrane/ Extracellular	Lipoprotein (SPII)	188	21.59
1901	fig|6666666.171271.peg.1901	Hypothetical protein	Outer Membrane/ Extracellular	Lipoprotein (SPII)	195	22.43
2018	fig|6666666.171271.peg.2018	Hypothetical protein	Outer Membrane/ Extracellular	Lipoprotein (SPII)	199	22.39
981	fig|6666666.171271.peg.981	Hypothetical protein	Outer Membrane/ Extracellular	Lipoprotein (SPII)	200	23.16
2068	fig|6666666.171271.peg.2068	Hypothetical protein	Outer Membrane/ Extracellular	Lipoprotein (SPII)	207	23.88
1869	fig|6666666.171271.peg.1869	FIG00438647: hypothetical protein	Outer Membrane/ Extracellular	Lipoprotein (SPII)	242	26.3
1217	fig|6666666.171271.peg.1217	FIG00438757: hypothetical protein	Outer Membrane/ Extracellular	Lipoprotein (SPII)	246	28.55
1677	fig|6666666.171271.peg.1677	Hypothetical protein	Outer Membrane/ Extracellular	Lipoprotein (SPII)	247	27.02
1599	fig|6666666.171271.peg.1599	FIG00438621: hypothetical protein	Outer Membrane/ Extracellular	Lipoprotein (SPII)	377	42.32
429	fig|6666666.171271.peg.429	Sialic acid-induced transmembrane protein YjhT(NanM), possible mutarotase	Outer Membrane/ Extracellular	Lipoprotein (SPII)	384	41.82
828	fig|6666666.171271.peg.828	Hypothetical protein	Outer Membrane/ Extracellular	Lipoprotein (SPII)	393	45.71
1473	fig|6666666.171271.peg.1473	Ankyrin repeat-containing protein//response to abiotic stimulus	Outer Membrane/ Extracellular	Lipoprotein (SPII)	528	62.24
724	fig|6666666.171271.peg.724	FIG00439020: hypothetical protein//Involved in the TonB-independent uptake of proteins	Outer Membrane/ Extracellular	Lipoprotein (SPII)	2330	269.18
2613	fig|6666666.171271.peg.2613	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA (TC 3.A.1.5.1)// extracellular solute-binding protein, family 9	Outer Membrane/ Extracellular	Lipoprotein (SPII)	534	60.56
1762	fig|6666666.171271.peg.1762	FIG00438985: hypothetical protein	Outer Membrane/ Extracellular	Lipoprotein (SPII)	494	57.28
819	fig|6666666.171271.peg.819	Tetratricopeptide repeat protein/Tfp pilus assembly protein	Outer Membrane/ Extracellular	Extracellular	631	72.96
129	fig|6666666.171271.peg.129	Flagellar hook-associated protein FlgK	Outer Membrane/ Extracellular	Extracellular	637	70.68
240	fig|6666666.171271.peg.240	FIG00438249: hypothetical protein/weak similarity to cell adhesion molecule-like protein 1 [*Littorina littorea*]	Outer Membrane/ Extracellular	Adhesion	240	27.43
2026	fig|6666666.171271.peg.2026	Flagellar filament outer layer protein//FlaA	Periplasm	Periplasm	316	35.71
235	fig|6666666.171271.peg.235	Ngo MtrE OM channel (26% similarity)	Periplasm	OMP	197	21.47
2198	fig|6666666.171271.peg.2198	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA (TC 3.A.1.5.1)//extracellular solute-binding protein, family 8	Periplasm	Lipoprotein (SPII)	533	60.09
894	fig|6666666.171271.peg.894		Periplasm	Lipoprotein (SPII)	537	60.82

(*continued on next page*)

Table 1 (*continued*)

SeqID	Feature ID *B. pilosicoli* B2904	Putative Name and Function//eggNOG-mapper preferred name	Consensus Localization	Protein characteristic (if known)	Amino Acids	Molecular Weight (kDa)
		Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA (TC 3.A.1.5.1)//extracellular solute-binding protein, family 10				
1775	fig|6666666.171271.peg.1775	Cell surface protein//regulation of response to stimulus	Surface	Surface	239	26.12
434	fig|6666666.171271.peg.434	Variable surface protein VspH	Surface	Surface	387	42.46
1028	fig|6666666.171271.peg.1028	Hypothetical protein/canonical leader peptide +canonical cleavage site AXA/Predicted - Integrin alpha-11//ATP-independent chaperone mediated protein folding	Unknown	Unknown	159	18.62
2323	fig|6666666.171271.peg.2323	Flagellar biosynthesis protein FliL; controls the rotational direction of flagella during chemotaxis	Unknown	Unknown	181	20.3
2166	fig|6666666.171271.peg.2166	flagellar filament outer layer protein (FlaA-2)	Unknown	Unknown	227	26.06
878	fig|6666666.171271.peg.878	Flagellar motor rotation protein MotB//tpn50, ompA family	Unknown	Unknown	271	31.63
1888	fig|6666666.171271.peg.1888	Toxin A	Unknown	Toxin	263	30.24
1767	fig|6666666.171271.peg.1767	Cell surface protein//regulation of response to stimulus	Unknown	Surface	65	7.19
1771	fig|6666666.171271.peg.1771	Cell surface protein//regulation of response to stimulus	Unknown	Surface	161	17.88
1750	fig|6666666.171271.peg.1750	*Serpulina (Brachyspira) hyodysenteriae* variable surface /protein	Unknown	Surface	122	13.73
2077	fig|6666666.171271.peg.2077	ABC transporter substrate-binding protein (88%)	Unknown	Periplasm	240	26.47
2655	fig|6666666.171271.peg.2655	Flagellin protein FlaA//FliC, component of the core of the flagella	Unknown	Periplasm	278	30.16
1510	fig|6666666.171271.peg.1510	Flagellin protein FlaA//FliC, component of the core of the flagella	Unknown	Periplasm	290	32.03
1083	fig|6666666.171271.peg.1083	Gliding motility protein GldG ABC transporter//PFAM ABC-type uncharacterised transport system	Unknown	Periplasm	245	27.68
2619	fig|6666666.171271.peg.2619	Outer membrane protein (Omp)H precursor	Unknown	OMP	142	16.53
1010	fig|6666666.171271.peg.1010	FIG00437452: hypothetical protein	Unknown	OMP	178	20.84
1772	fig|6666666.171271.peg.1772	Chitin binding protein	Unknown	Lipoprotein (SPII)	150	16.33
1661	fig|6666666.171271.peg.1661	FIG00437880: hypothetical protein//response to heat	Unknown	Lipoprotein (SPII)	153	16.33
125	fig|6666666.171271.peg.125	FIG00437359: hypothetical protein	Unknown	Lipoprotein (SPII)	162	18.94
2255	fig|6666666.171271.peg.2255	Hypothetical protein	Unknown	Lipoprotein (SPII)	163	18.69
2028	fig|6666666.171271.peg.2028	FIG00437611: hypothetical protein	Unknown	Lipoprotein (SPII)	164	18.79
532	fig|6666666.171271.peg.532	FIG00439291: hypothetical protein	Unknown	Lipoprotein (SPII)	168	18.91
631	fig|6666666.171271.peg.631	Hypothetical protein	Unknown	Lipoprotein (SPII)	169	19.97
1081	fig|6666666.171271.peg.1081	Hypothetical protein	Unknown	Lipoprotein (SPII)	171	19.02
833	fig|6666666.171271.peg.833	FIG00437338: hypothetical protein	Unknown	Lipoprotein (SPII)	174	20.17
1076	fig|6666666.171271.peg.1076	FIG00437875: hypothetical protein	Unknown	Lipoprotein (SPII)	177	21.15
2314	fig|6666666.171271.peg.2314	Hypothetical protein	Unknown	Lipoprotein (SPII)	180	20.95
2011	fig|6666666.171271.peg.2011	Hypothetical protein	Unknown	Lipoprotein (SPII)	186	20.69
691	fig|6666666.171271.peg.691	Hypothetical protein	Unknown	Lipoprotein (SPII)	193	22.3
2130	fig|6666666.171271.peg.2130	Hypothetical protein//chlorophyll binding	Unknown	Lipoprotein (SPII)	194	21.98
841	fig|6666666.171271.peg.841	FIG00437768: hypothetical protein	Unknown	Lipoprotein (SPII)	195	22.14
145	fig|6666666.171271.peg.145	Hypothetical protein	Unknown	Lipoprotein (SPII)	196	23.1
2129	fig|6666666.171271.peg.2129	Hypothetical protein//chlorophyll binding	Unknown	Lipoprotein (SPII)	198	22.06
234	fig|6666666.171271.peg.234	Hypothetical protein	Unknown	Lipoprotein (SPII)	204	22.79
1216	fig|6666666.171271.peg.1216	Hypothetical protein	Unknown	Lipoprotein (SPII)	207	23.99

(*continued on next page*)

Table 1 (*continued*)

SeqID	Feature ID *B. pilosicoli* B2904	Putative Name and Function//eggNOG-mapper preferred name	Consensus Localization	Protein characteristic (if known)	Amino Acids	Molecular Weight (kDa)
1711	fig|6666666.171271.peg.1711	Hypothetical protein	Unknown	Lipoprotein (SPII)	212	24.19
1801	fig|6666666.171271.peg.1801	TPR domain protein	Unknown	Lipoprotein (SPII)	236	27.05
2151	fig|6666666.171271.peg.2151	FIG00437918: hypothetical protein	Unknown	Lipoprotein (SPII)	241	27.36
957	fig|6666666.171271.peg.957	Ankyrin repeat-containing protein//response to abiotic stimulus	Unknown	Lipoprotein (SPII)	251	28.49
1637	fig|6666666.171271.peg.1637	FIG00437393: hypothetical protein	Unknown	Lipoprotein (SPII)	288	33.1
2006	fig|6666666.171271.peg.2006	FIG00438121: hypothetical protein//Pla, Protein of unknown function (DUF3089)	Unknown	Lipoprotein (SPII)	321	36.98
1433	fig|6666666.171271.peg.1433	FIG00438846: hypothetical protein	Unknown	Lipoprotein (SPII)	339	39.37
2182	fig|6666666.171271.peg.2182	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA (TC 3.A.1.5.1)// extracellular solute-binding protein, family 5	Unknown	Lipoprotein (SPII)	531	61.07
892	fig|6666666.171271.peg.892	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA (TC 3.A.1.5.1)// extracellular solute-binding protein, family 6	Unknown	Lipoprotein (SPII)	536	60.58
1127	fig|6666666.171271.peg.1127	COG1649 predicted glycoside hydrolase//YddW, PFAM Uncharacterised BCR, COG1649	Unknown	Lipoprotein (SPII)	604	70.78
2276	fig|6666666.171271.peg.2276	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA (TC 3.A.1.5.1)// Extracellular solute-binding protein, family 5	Unknown	Lipoprotein Periplasm (SPII)	522	58.9
189	fig|6666666.171271.peg.189	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA (TC 3.A.1.5.1)// extracellular solute-binding protein, family 6	Unknown	Lipoprotein Periplasm (SPII)	527	60.46
487	fig|6666666.171271.peg.487	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA (TC 3.A.1.5.1)// extracellular solute-binding protein, family 7	Unknown	Lipoprotein Periplasm (SPII)	529	59.92
763	fig|6666666.171271.peg.763	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA (TC 3.A.1.5.1)// extracellular solute-binding protein, family 11	Unknown	Lipoprotein Periplasm (SPII)	546	62.78
30	fig|6666666.171271.peg.30	FIG00438389: hypothetical protein/Canonical lipobox at the N-term, aromatic residue at the C-term	Unknown	Lipoprotein (SPI)	264	31.32
233	fig|6666666.171271.peg.233	Hypothetical protein//LPP20 lipoprotein	Unknown	Lipoprotein (SPI)	332	38.31
2282	fig|6666666.171271.peg.2282	FIG00438972: hypothetical protein	Unknown	Lipoprotein (SPI)	380	43.68
1490	fig|6666666.171271.peg.1490	BatC/aero-tolerance-like exported protein	Unknown	integral component of membrane	290	33.28
2440	fig|6666666.171271.peg.2440	Tetratricopeptide repeat protein	Unknown	Extracellular	388	45.33

chimera, and vice versa with N-terminus located epitopes. The sequence GPGPG was used as a spacer between the epitopes to provide flexibility and K and C residues were added at the C- and N-termini to improve stability of the chimera. 6) Properties of the chimera protein were analysed by Blast-p and the Expasy ProtParam tool [68] to compute molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY). PepCalc was used to estimate water solubility [69]. ABCpred (Windows: 16 Threshold = 0.8) was finally used on the chimera to check if the epitopes were still presented as epitopes in the protein. *In silico* structure prediction was attempted using the ProteinPredict program [70,71] and SWISS-MODEL [72].

3. Results and discussion

3.1. *B. pilosicoli* reverse vaccinology

The scheme outlining the bioinformatics tools used in this study is summarised in Fig. 1. The whole genome sequence for *Brachyspira pilosicoli* strain B2904, an isolate causing AIS, was downloaded from the NCBI GenBank and uploaded onto RAST. The latest assembly of the complete genome nucleotide sequence is shown in Supplementary Dataset 1 (GCA_000296575.1_ASM29657v1_genomic.fna). The complete nucleotide sequence encodes for 2679 protein coding sequences (Supplementary Dataset 2) and the amino acid sequences for these proteins are shown in Supplementary Dataset 3 and were used for all subsequent proteome analyses.

The *B. piloscoli* proteome was examined for potential protein localisation with PSORTb3 (Supplementary Dataset 4), CELLO (Supplementary Dataset 5) and $SOSUI_{GramN}$ (Supplementary Dataset 6). However, the three programs could not find consensus for the localisation of some of the proteins, and these were recorded as 'Unknown'. The consensus localisation in Table 1 was derived from agreement of at least two of the programs. LipoP server (Supplementary Dataset 7) and SignalP (Supplementary Dataset 8) programs were used to identify the SPI and SPII peptidase lipoproteins and the presence of signal peptides, respectively. The proteins were also examined for the presence or absence of C-terminal aromatic amino acid residues to aid predicting β-barrel integral transmembrane proteins, and for the presence of transmembrane helices using the TMHMM server (Supplementary Dataset 9). The TMHMM data showed that most of the selected proteins did not have transmembrane regions; where one was recorded, this was generally in the N-terminal signal peptide sequence. For those proteins with 'unknown localization', amino acid sequences were also compared against the whole non-redundant GenBank database excluding Brachyspirales using BLAST and manually against the PDB database. This enabled the localities for some of the 'unknown' proteins to be identified. The eggNOG-mapper

program was also useful for identifying the preferred names and functions for some of the proteins (Supplementary Dataset 10).

In general, the exclusion criteria were all proteins that were finally predicted as cytoplasmic, cytoplasmic/inner membrane, smaller than 150 amino acids in length, and/or like other homologues in a non-pathogen *E. coli* and a commensal *L. reuteri* (Supplementary Dataset 11), and if they contained 3 or 5 or more cysteine residues (Supplementary Dataset 12). There were 1175 out of 2679 (43%) proteins in *E. coli* K12 and 818 out of 2679 (31%) proteins in *L. reuteri* that shared similarity with *B. pilosicoli* proteins, ranging from ~19–72% similarity (Supplementary Dataset 11). Of the 1175 *E.coli* proteins, 880 (75%) and 183 (16%) were localised in the cytoplasm and cytoplasmic membrane respectively, with 23 (2%) that were periplasmic and 84 (7%) that were of unknown localisation. Only 3 extracellular and 2 OM proteins had some similarities with *B. pilosicoli* proteins. A similar distribution was observed with the 818 *L. reuteri* proteins, of which 634 (78%) and 130 (16%) were found in the cytoplasm and cytoplasmic membrane respectively, and 6 (1%) in the periplasm, 47 (6%) unknown localisation and with only one OM protein with some similarity in sequence.

As a result of these *in silico* analyses, there were 197 proteins identified (see Final_list > final_removal datasheet in Supplementary Dataset 13). Proteins localised within the outer membrane (OM) or identified as extracellular or on the surface are important candidates, since they suggest potential immune recognition. We identified 57 proteins localised within the OM, 30 as extracellular and 2 as surface exposed. In addition, there were 36 proteins for which no precise distinction between localisation in the OM or extracellular compartments could be made. Thus, a total of 125 (63%) proteins are potentially exposed for immune recognition. We also identified 62 (31%) proteins for whom a consensus localisation was unknown. We also retained a smaller number of proteins (6%) that were localised within the periplasm (n = 6) and cytoplasm (n = 2) and either within the OM or inner membrane (n = 1). These were retained and examined further as described below.

For the proteins within each of the consensus localisation compartments, *in silico* analyses could annotate many of them according to fundamental characteristics, e.g. whether they were lipoproteins, or OM proteins (OMP), extracellular or periplasmic proteins, or whether they had key descriptors such as 'toxin', 'surface' or 'membrane' (Table 1). Ascribed putative names and functions, coupled with preferred names provided by eggNOG-mapper was also useful for describing individual proteins and the rationale for their retention for future study. Further searches in the literature (PubMed) and in UniProt provided additional information for some of the identified proteins. However, they also showed that the consensus localisations for several of the proteins was wrong and that their biological function(s) was redundant with respect to vaccine potential, and therefore these candidates could also be excluded.

3.2. Cytoplasm consensus localisation proteins

We excluded all cytoplasmic proteins; however, we did identify SeqID247 in this compartment, which shared homology with secreted *Clostridium botulinum* toxin. SeqID58, identified putatively as a flagellar basal-body rod protein FlgF or FlhO, was also retained, as other flagellar-associated proteins appeared to be extracellular (Supplementary Dataset 13).

3.3. Extracellular consensus localisation proteins

The extracellular consensus localisation compartment contained 12 SPII lipoproteins, 10 extracellular proteins, and 2 of each that were either surface, secreted, periplasmic or unknown proteins (Supplementary Dataset 13). Indeed, many of the lipoproteins in the whole *B. pilosicoli* list were designated with SPII sequences (n = 85/197, 43%). Lipoproteins with SPII peptidase sequences are generally surface-exposed in Gram-negative organisms: for example, in the spirochaete *Borrelia burgdorferi*, approximately two-thirds of the over 120 lipoproteins expressed by the pathogen localise to the surface [73]. Zuckert et al. suggested a mechanism for surface localisation whereby the various surface-targeted lipoproteins interact with a "holding" chaperone protein in the periplasm, and this chaperone delivers these surface lipoproteins to an OM lipoprotein "flippase" complex. This ''flippase'' enables translocation of lipoproteins through the OM and to their ultimate anchoring within the surface leaflet of the OM. Examination of the proteome shows that *B. pilosicoli* produces two proposed peptidoglycan lipid II flippase MurJ proteins (Feature ID fig|6666666.171271.peg.462 and fig|6666666.171271.peg.463; Supplementary_Dataset_2), suggesting that the identified SPII lipoproteins in this organism can potentially be surface-localised.

Of the extracellular SPII lipoproteins, 9 were hypothetical in name and function, one was a putative cell surface protein (SeqID382) and 2 were sialidases (SeqID1802, SeqID1804). Secreted sialidase enzymes that can cleave sialic acid monosaccharides found on the end of glycan chains of various mammalian secreted proteins, cell surface proteins and lipids, often of mucosal surfaces, are used by several bacterial pathogens to facilitate colonization of the mammalian host and/or increase pathogenesis [74]. There were 5 hypothetical proteins amongst the 10 extracellular proteins. One extracellular protein, SeqID1398, contains a cysteine-rich secretory SCP-like extracellular protein domain, found in virulence-associated extracellular proteins and shows weak homology (27%) to a *Bacillus subtilis* extracellular protein 30/112. SeqID2646 is identified as a peptidase M30/hyicolysin protein: interestingly, *Staphylococcus hyicus*, which infects pigs, cattle and chickens also secretes a mature hyicolysin extracellularly [75]. In pigs, *S. hyicus* infection can cause exudative dermatitis, necrosis of the tips of the ears, and tail-biting. One surface protein (SeqID1751) shared identity with a variable surface protein of *Serpulina (Brachyspira) hyodysenteriae*, and the other surface protein (SeqID2319) had weak sequence similarity to the Ring-infected erythrocyte surface antigen (RESA) of *Plasmodium falciparum*. A recombinant RESA protein has been used as a component of a malarial vaccine that reduced *P. falciparum* density and exerted selective pressure on parasite populations in a human vaccine trial in Papua New Guinea [76]. Flagellar-associated proteins were also evident as extracellular proteins (e.g. FlgG (SeqID1838), FlgL (SeqID1332), FlgE (SeqID2145), FliD (SeqID1312)) and secreted/periplasmic/surface/unknown proteins, i.e. FliK (SeqID2125), FlgD (SeqID2126), FlaA/FlaB (SeqID2573) and FliN (SeqID2325). The potential of *Brachyspira* flagellar proteins as vaccine targets has been attempted with a recombinant FlaB1 38 kDa recombinant flagellar protein of *B. hyodysenteriae*. Vaccination of pigs with this antigen in a mineral-oil paraffin adjuvant did produce an antibody response to the protein, but vaccination was not protective against experimental infection with *B. hyodysenteriae* [77].

The proteins that could be excluded due to incorrect identification of localisation and redundant biological function included SeqID454 of unknown function, although it does share 28% homology with LolA protein, which is generally a periplasm located carrier protein involved in lipoprotein transport from the inner membrane to the OM [78]. In addition, although SeqID45 is localised extracellularly, its suggested homology to β-lactamase inhibitor suggests that it may be a periplasmic protein. The extracellular protein SeqID1161 is the largest protein in the list, with 5397 amino acid residues and a molecular weight of 629 kDa. Its excessive size might suggest that it is subject to extracellular processing, but a BLAST analysis identified this protein as a possible DNA Directed RNA Polymerase or apolipoprotein A1/A4/E domain-containing protein, and thus more probably located in the cytoplasm (Supplementary Dataset 13).

3.4. OM consensus localisation proteins

This compartment contained 57 proteins, which were identified as 46 OM proteins (OMPs), 5 SPII lipoproteins, 1 each of a periplasmic SPII

lipoprotein, a periplasmic protein, an extracellular protein and a secreted protein, and 2 unknown proteins (Supplementary Dataset 13). Several interesting candidates were identified. SeqID629 and SeqID2620 are identified as a homologue of YaeT OMP assembly factor protein from *Escherichia coli* [79]. There were several OMPs identified with association to efflux pumps, e.g. SeqID44 and TolC (also identified as SeqID629) [80]. SeqID840 defined an OMP with homology to the treponemal membrane protein B (TmpB): a previous study has shown that immunization of guinea pigs with recombinant TmpB antigen induced protection against challenge infection with *Treponema pallidum* [81], and *B. pilosicoli* TmpB has been recognised by hyper-immune sera from challenged pigs [82]. TmpB was also named by eggNOG-mapper in SeqID2554 (and also named as a TolA protein, reportedly located in the *E.coli* periplasm predominantly [83]).

The CsgG curli production assembly transport component OMP (SeqID1529) is required for the secretion of curli, which are functional amyloid fibres that are the major protein component of the biofilm extracellular matrices formed by Bacteroidetes and Proteobacteria [84] and have significant roles in pathogenesis. Curli fimbriae are believed to be involved in the survival of avian pathogenic *E. coli* (APEC) outside of the host and to play roles in early colonisation of the host and in bacterial persistence and invasion [85–87]. Furthermore, a RV study of *Haemophilus parasuis*, which causes contagious porcine Glässer's disease, identified CsgG as a potent immunogenic protein [88]. SeqID2019 is the OMP BtuB, which in *E. coli* functions to transport cyanocobalamin (vitamin B12) [89]. The Tia invasion determinant virulence OMP (SeqID2232) has homology with the 25-kDa Tia OMP of enterotoxigenic *E. coli*, which has been reported to act as an adhesin and invasin [90]. SeqID1434 is a protein related to OstA, which is a base pH-induced OMP in *E.coli* [91] and a homologue of the β-barrel protein LptD, and whose expression was up-regulated in kanamycin-resistant *E. coli* [92]. SeqID2472 is an OMP with similarities to the proteins of the OmpA family, which play major roles in bacterial adhesion, invasion, intracellular survival, inflammatory processes and evasion of host innate immune responses. Notably, OmpA family proteins have been evaluated as experimental vaccine antigens for many Gram-negative pathogens causing human and veterinary infections [93]. We also identified SeqID2067 as an OMP with 71% homology to a phosphate-selective porin of the O and P families, but it is possible that these anionic porins are expressed only during phosphate-starvation, as observed for porin O from *P. aeruginosa* [94]. There were several OMP linked to TonB, a cytoplasmic membrane protein that is part of a protein complex that traverses the periplasm to interact with OMP and provide energy for the active transport of substrates across the OM and periplasm [95]. These included SeqID1506 and SeqID2119 (the predicted TonB receptor proteins SusC/RagA) and another TonB receptor (SeqID624) that might function as a transporter. Conversely, we identified an OMP putatively called TreP (SeqID2235) that was suggested to play a role in TonB-independent protein uptake. SeqID693 is an OMP identified either as a homologue of Cia (*Campylobacter* invasion antigen) B, a secreted virulence protein used by *Campylobacter jejuni* to invade cultured mammalian cells [96], or as FtsH, which is an essential inner membrane protease in *E. coli* [97]. Clearly, a virulence function would make this OMP attractive as a vaccine candidate and is worth exploring.

Other OMP of interest included a *Serpulina (Brachyspira) hyodysenteriae* variable surface protein (SeqID387), a putative exported protein (SeqID980) and a LPS transport protein (SeqID1904). Interestingly, SeqID1221, was identified as an OMP of the spirochaetal α2-macroglobulin family. It is known that metazoan α_2-macroglobulins inhibit proteases produced by pathogens, and conversely, bacteria may have acquired these proteins from their metazoan hosts and use them to block host antimicrobial defences. It has also been suggested that these proteins might be useful targets for enhancing vaccine efficacy [98]. *B. pilosicoli* also produces a haemolysin in the OM (SeqID1917) and this is a virulence factor that could be a potential vaccine target. Several studies have shown the potential of *Brachyspira* haemolysin proteins as potential vaccine targets: for example, vaccination of pigs with a *B. hyodysenteriae tlyA* haemolysin mutant offered partial protection against challenge with a virulent strain [99]. In addition, vaccination of pigs with weakly haemolytic *B. hyodysenteriae* bacteria provided significant protection against subsequent challenge with a virulent strain [100]. The flagellar filament outer layer protein FlaA (SeqID2167) is a putative vaccine candidate for the reasons described above.

Within the OM compartment, we noted the presence of 26 hypothetical proteins, including 2 that were homologues of unknown proteins from the spirochete *Borrelia burgdorferi* (SeqID448, SeqID455). These cannot be excluded from future study as they may contain vaccine antigens.

However, several of the proteins characterised as OMP could be excluded due to incorrect localisation or redundant biological function (s), following individual examination with UniProt analysis and literature review of BLAST homologues (Supplementary Dataset 13). ArcD (SeqID506), identified as a transmembrane protein with arginine:ornithine exchange activity and localised instead to the inner cytoplasmic membrane of other Gram-negative bacteria such as *Pseudomonas aeruginosa* [101]; Hvp32 (SeqID2682) is a protein associated with the tail of the VSH-1 prophage released by *Brachyspira spp.* [102]; SeqID362, an OMP with 53% homology to the molybdenum cofactor guanylyltransferase enzyme, which is located in the cytoplasm of *E.coli*; ModA (SeqID1454), a molybdenum ABC transporter, localised to the periplasm; Seq ID1953, a phage-associated protein; SeqID1727, a protein with homology (74%) to a hemin ABC transporter substrate-binding protein, which are usually localised in the periplasm; SeqID495, a arylsulfate Ig-like domain sulfotransferase, or assT, which in other bacteria, e.g. *Klebsiella,* has been suggested to be located in the periplasm [103]; and two peptidylprolyl cis-trans isomerase (PPIase) proteins, one possibly ppiD (SeqID1034) and the other SurA (SeqID2374). Bacterial PPIases associated with OM have significant potential as vaccine antigens [104], but these two can probably be excluded as they are recognised as periplasmic chaperone proteins.

3.5. OM/Extracellular consensus localisation proteins

There were 36 proteins for which no precise distinction between localisation in the OM or extracellular compartments could be made (Supplementary Dataset 13). The majority of these, 26/36 (72%) were SPII lipoproteins. We identified several possible candidates, e.g. SeqID2613, OppA family protein, which is a potential vaccine candidate by analogy to the OppA proteins of *H. parasuis*, which were identified as potent immunogenic proteins [88]. SeqID429 is identified as YjhT, a possible mutarotase. A homologue has been identified in the periplasm of *E. coli* [105], and the protein converts α-N-acetylneuraminic acid to the β-anomer and accelerates the equilibrium between the α and β-anomers. YjhT may play a role in virulence by allowing sialidase-negative bacteria to compete successfully for low amounts of extracellular α-N-acetylneuraminic acid, and the removal of sialic acid from the environment might be advantageous to the bacterium by dampening host responses.

There were 17 hypothetical SPII lipoproteins, of which two had putative descriptors, i.e. SeqID724, which may be involved in TonB-independent uptake of proteins and SeqID1473, which may be involved in bacterial response to abiotic stimuli. Of the remaining proteins, we identified a membrane protein, SeqID1728, which had homology to extracellular regions of 7TM receptors, which can be cell-surface located in a manner somewhat analogous to microbial rhodopsin proteins [106]. We also identified a protein called BatC (SeqID1294): recently, a pan-genome RV approach identified a homologue BatC (LB_056) as a promising vaccine candidate for *Leptospira interrogans* [107]. There were also flagellar-associated proteins FlgA (SeqID597), FlgK (SeqID129, SeqID1829) and a putative Type 4 Pilus (Tfp) assembly protein (SeqID819). There was also a protein hypothetically identified as an adhesion protein, SeqID240, with weak similarity

to a cell adhesion molecule-like protein 1 of the common periwinkle *Littorina littorea*.

However, several proteins identified as OM/extracellular could be excluded from the list (Supplementary Dataset 13) due to incorrect identification of localisation and redundant biological function(s). These include SeqID43, identified as MtrC, an efflux pump component, which is located under the inner leaflet of the OM and interacting with MtrE in the OM as part of the MtrCDE multi-drug pump, e.g. of *Neisseria gonorrhoeae* [108]; SeqID246, arylsulfate Ig-like domain sulfotransferase; SeqID614, xylF, a D-xylose-binding lipoprotein attached to the inner membrane; SeqID2358, a hypothetical intracellular S-adenosyl-L-methionine hydroxide adenosyltransferase enzyme [109]; SeqID2386, a periplasmic nucleoside ABC transporter; SeqID603, LolA; SeqID1499, a Type IV pilus biogenesis protein PilO, which is located in the inner membrane [110]; SeqID152, Hvp 101 protein, which is associated with the prophage VSH-1 of *B. hyodysenteriae* [111]; SeqID1820, a SPII lipoprotein probably located in the cytoplasm; SeqID479 was annotated as a penicillin-binding protein (PBP) outer and inner membrane//serine threonine protein kinase. Since PBPs are involved in peptidoglycan synthesis, SeqID479 is probably located in the inner membrane or periplasm and can be excluded.

3.6. Periplasm consensus localisation proteins

We identified 6 proteins in the periplasm, 3 of which were SPII lipoproteins, 1 was an OMP and 2 were periplasmic (Supplementary Dataset 13). The SPII lipoproteins SeqID894 and SeqID2198 are defined as OppA ABC-type transporter family proteins. SeqID235 protein has partial homology to the *N. gonorrhoeae* MtrE efflux pump OMP and was retained. The rationale for this retention was the fact that surface loops of the gonococcal MtrE are potential vaccine targets and antibodies to MtrE are bactericidal [112]. The flagellar unit FlaA SeqID2026 is analogous to SeqID2167 described above. However, MglB (SeqID2288, Supplementary Dataset 13) can probably be excluded as it is defined as a periplasmic D-galactose-binding protein, as can the PPIase putatively identified as PpiB (SeqID839), which is probably a cytoplasmic cyclophilin.

3.7. Surface consensus localisation proteins

We identified two surface proteins, SeqID1775 annotated as a cell surface protein possibly involved in the regulated response to stimuli, and SeqID434, the variable surface protein VspH (Supplementary Dataset 13). The Vsp proteins (also known as Bhmp) have been reported as major OMPs in *B. hyodysenteriae* and antibodies from pigs infected with *B. hyodysenteriae* have been shown to react with Vsp conformational epitopes [113]. Rabbit antibodies to recombinant Vsp proteins have also been reported to inhibit, to varying degrees, the adherence of *B. hyodysenteriae* bacteria to pig intestinal epithelial cells [114].

3.8. Unknown consensus localisation proteins

There were 63 proteins with an unknown consensus localisation (Supplementary Dataset 13). We identified SeqID2619 as the OMP OmpH and one hypothetical OMP (SeqID1010), 3 surface proteins (SeqID1750, SeqID 1771, SeqID 1767) and a putative toxin A (SeqID1888). There were 4 periplasmic proteins, of which 2 were flagellar components (SeqID1510, SeqID 2655) and 2 were ABC transporter-associated proteins (SeqID2077, SeqID 1083). Interestingly, SeqID1083 is a putative gliding motility protein GldG and a similarly named protein has been identified as a virulence factor in the fish pathogen *Flavobacterium psychrophilum* [115]. The observation that disruption of GldG in this pathogen resulted in impaired secretion of extracellular enzymes and adhesins and reduced host colonization and infectivity, suggested that this protein may be a therapeutic target. There were 31 SPII lipoproteins, of which 24 were hypothetical function proteins. The remaining 7 included two OppA family proteins (SeqID892, SeqID 2182), a lipoprotein responding to abiotic stimuli (SeqID957), a chitin-binding protein (CBP, SeqID1772), which is of interest given the increase in reports of bacterial chitinases and CBPs acting as virulence factors [116] and YddW (SeqID1127), a glycoside hydrolase that has been localised to the OM in *E. coli* [117].

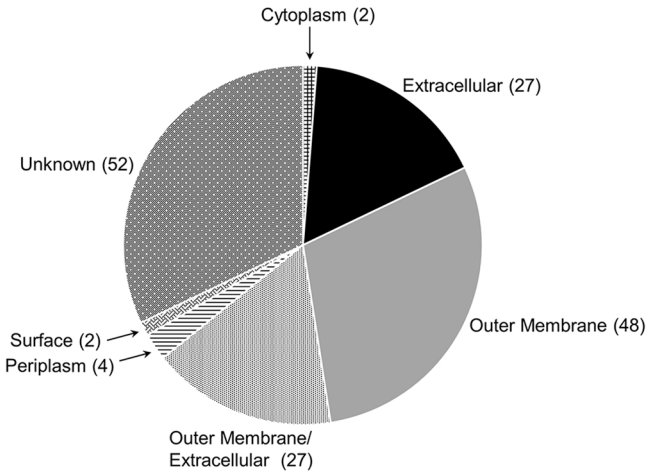

Fig. 2. Venn diagram showing the distribution of proteins identified by reverse vaccinology amongst cellular compartments of *B. pilosicoli* B2904. Brackets include the number of proteins per compartment.

There were 6 periplasmic SPII lipoproteins, which included 4 OppA family member proteins (SeqID189, SeqID 487, SeqID 763, SeqID 2276), which are retained. RV also identified 5 SPI lipoproteins for further study, of which 3 were of hypothetical function (SeqID30, SeqID 233, SeqID 2282). There was also an extracellular protein with TPR repeats (SeqID2440). We also identified BatC in SeqID1490. Finally, there were 8 proteins of unknown localisation and unknown descriptor, some of which had putative names. There were two flagellar-associated proteins, FliL (SeqID2323) and FlaA-2 (SeqID2166) and a third one (SeqID878), annotated as either MotB or Tpn50, which is a member of the OmpA family and related to the putative Tpn50 OM porin of *Treponema pallidum* [118]. A hypothetical protein SeqID1028 could have chaperone function.

Several proteins could be excluded from the list (Supplementary Dataset 13) after re-examination of cellular localisation and biological function: SeqID2380, a glycerate kinase enzyme, probably cytoplasmic in location; XylF (SeqID615), a D-xylose-binding lipoprotein attached to the inner membrane; the periplasm-located proteins CpdB (SeqID1797) that belongs to the 5'-nucleotidase family, TRAP transporter (SeqID207) and SPI Lipoprotein ABC-type nitrate/sulfonate/bicarbonate transport system (SeqID1405); CapA (SeqID913), an inner membrane located capsule biosynthesis protein; SPI Lipoprotein LytB (SeqID1126), which is probably associated with the cellular membrane; SeqID457, which has homology to *P. aeruginosa* penicillin-binding protein 3, found in the periplasm; SeqID1262, a methyl-accepting chemotaxis protein I (mcp), which are cellular transmembrane proteins; SeqID437, related to the cytoplasmic SagB Streptolysin S biosynthesis protein [119]; and SeqID1477, the soluble periplasmic protein YfiH [120].

3.9. Application of VaxiJen and Vaxign software to predict antigenicity and vaccine target prediction

The *B. pilosicoli* RV list produced with the final round of exclusions from all of the cellular compartments described above, contained a total of 162 proteins localised into 48 OM proteins, 27 OM/Extracellular proteins, 27 Extracellular proteins, 4 Periplasm proteins, 2 Surface proteins, 2 Cytoplasm proteins and 52 Unknown proteins (Fig. 2, Supplementary Dataset 13). We then analysed this subset of 162 proteins in

Table 2
Experimental *B. pilosicoli* vaccines.

Antigen	Vaccine	Efficacy	Reference	SeqID
34 kDa carboxy-terminus of Bpmp72 OMP	Recombinant protein emulsified with Freund's incomplete adjuvant (FIA)	Vaccination of laying chickens lead to a highly significant reduction in the duration of colonization by *B. pilosicoli*. Fewer vaccinated chickens had abnormal caecal contents after infection, compared to control animals, but not statistically significant.	[121]	2554
Putative oligopeptide-binding proteins	Recombinant proteins P-1 and P-3 emulsified with FIA	Subcutaneous vaccination of mice with proteins induced systemic and local colon IgG antibody responses. After experimental infection, the cumulative number of colonization days was significantly less in vaccinated mice than in control mice. Fewer mice in the P-1 group were colonized compared to control mice.	[122]	189, 487, 763, 892, 894, 2182, 2198, 2276, 2613
67 kDa membrane-associated ATPase subunit of Clp protease (ClpX)	Recombinant protein with FIA	Subcutaneous vaccination of mice induces antibodies that recognize the 67 kDa protein in whole-cell preparations of *B. pilosicoli.*	[123]	606
Bacterin (formalised whole bacteria cells)	Bacterin with FIA	Vaccination of pigs induced a primary and secondary serological response to *B. pilosicoli*, as measured by ELISA on sonicated whole bacterial cells. Vaccination did not delay or significantly reduce the onset of faecal excretion of *B. pilosicoli* after experimental challenge. Conversely, simple consumption of a rice-based diet delayed and reduced the onset of faecal excretion of *B. pilosicoli* after experimental challenge.	[141]	–

order to try and refine the list. We analysed each protein amino acid sequence using both the VaxiJen: Prediction of Protective Antigens and Subunit Vaccines program and Vaxign2 (Vaccine Design) Dynamic Vaxign Analysis tool (Supplementary Dataset 13). VaxiJen predicted that 20 of the proteins in the RV list were 'Non-antigens', and the majority of 142 were 'Antigens' (Supplementary Dataset 13). The non-antigens included cytoplasmic proteins (SeqID147 and SeqID58), several extracellular proteins (SeqID2275, SeqID2646, SeqID1332), OM proteins (SeqID1917 hemolysin, SeqID1303, SeqID2472 OmpA superfamily protein, SeqID1087), OM/Extracellular proteins (SeqID1728, SeqID1217, SeqID724) surface protein SeqID1775, and several unknown proteins (SeqID1767, SeqID1771, SeqID1010, SeqID125, SeqID841, SeqID145, SeqID1801).

Using Vaxign2/Vaxign-ML, 130/140 proteins (93%) had a 'Protegenicity' score > 90, with 6/140 having a score of $> 80\%$ and the remaining 6 ranging from 58% to 77%. Interestingly, within the 20 proteins identified as 'Non-antigens' by Vaxijen, 16 had a 'Protegenicity' score of > 90, 3 had scores $> 80\%$ and 1 had a lower score of 62%. For example, the non-antigen SeqID1917, identified as a secreted hemolysin had a 'Protegenicity' score of 93. Even the 'cytoplasmic' proteins, i.e. the toxin and flagellar basal body rod protein had 'Protegenicity' scores of 91. The discrepancy between the two programs probably reflects the differences between the two tools, e.g. VaxiJen allows antigen classification solely based on the physicochemical properties of amino acids within the protein sequences without sequence alignment, and Vaxign2/Vaxign-ML analyses both biological and physicochemical features. Vaxign2/Vaxign-ML also showed that only one protein, SeqID957, a SPII lipoprotein antigen with a 'Protegenicity' score of 91, showed cross-reactivity with human, pig and mouse host protein, but not with chicken. A *B. pilosicoli* vaccine containing this antigen, if immunogenic in vivo, would only be useful for chickens and not pigs. None of the proteins showed any similarities to chicken host proteins as determined by BLAST against chickens taxid:9031. Taking into account all of the findings, including the VaxiJen and Vaxign2/Vaxign-ML analyses, no further refinement to the list was warranted.

3.10. Final *B. pilosicoli* RV protein list

The final list of 162 proteins is shown in Table 1. Given the economic losses due to AIS and PIS, it is surprising that vaccine development and testing has been limited to the few antigens described in Table 2. Although some of these antigens have demonstrated an ability to reduce bacterial colonization, none of them have progressed further into commercial development. Moreover, the choice of Freund's emulsion as the adjuvant is likely to be superceded today with safer and more efficacious adjuvants. The superiority of our study is that it has expanded significantly the repertoire of potential vaccine candidates that can be tested experimentally. As described above, many of these candidates have corollaries with antigens in other pathogens that have shown promise as vaccine candidates.

Some of the antigens within the experimental vaccines described in Table 2 were identified by our RV. For example, the nucelotide sequence encoding the 34 kDa carboxy-terminus of Bpmp72 OMP from *B. pilosicoli* strain P43/6/78 [121] maps onto the *B. pilosicoli* B2904 complete genome (99%) at 2632578 – 2634331 (Seq ID CP003490.1) and encodes for SeqID2554, the putative treponemal OMP, TmpB, in our study. In

Table 3
The high priority antigens from *B.pilosicoli* identified by the Novartis Animal Vaccines (NAV) Group and their presence or absence from our new RV list.

NAV antigen accession number	Identity	Similarity to *B. pilosicoli* proteins in new RV study
AAK14801.1	Variable surface protein (VspF) (BLAST 100% for *B. hyodysenteriae* protein)	Partial similarity to variable surface family protein SeqID387
AAC10219.1	VspD (BLAST 100% for *B. hyosysenteriae* protein)	Partial similarity to variable surface family protein SeqID387
AAB47846.1	OMP *Treponema pallidum* flagellar filament outer layer protein FlaA	Several proteins identified in new RV as FlaA, i.e. SeqID2573 (13% amino acid sequence similarity), SeqID2167 (26% similarity), SeqID2026 (15% similarity) and SeqID2166 (22% similarity).
ZP_00300183.1 (ADK30893.1)	Integral membrane protein MviN putative virulence factor	First 94 amino acids are identical to SeqID462, one of the *B. pilosicoli* MurJ 'flippase' proteins
ZP_00330300.1 (ADK30237.1)	Preprotein translocase, subunit SecA	SeqID1528 is similar but absent in the new RV list.
AAC82625.1	Surface antigen BspA (BLAST 100% match to BspA of *Tannerella forsythia*)	No similarity to any protein in new RV list. Has minimal homology to *B. pilosicoli* BspA-like protein (BLAST; 46% identity for 7% query cover) and a leucine-rich repeat domain-containing protein (40% of 53% query cover).
NP_713795	100% identical to WP_000443308.1 OMPA family protein of *Leptospira interrogans*	No similarity to the OmpA proteins SeqID2472 or SeqID878 in new RV list.

addition, the recombinant putative oligopeptide-binding proteins P-1 and P-3 from *B. pilosicoli* strain 95/1000 used in another experimental vaccine [122] were identified from a panel of 6 Open Reading Frames (ORFs) that encoded for oligopeptide-binding proteins in this strain. Specifically, the amino acid sequences of expressed ORFs 1, 2, 3 and 4 showed varying homology with the family of OppA proteins in our *B. pilosicoli* B2904 RV list (SeqID189, SeqID 487, SeqID 763, SeqID 892, SeqID 894, SeqID 2182, SeqID 2198, SeqID 2276, SeqID 2613) (Table 2). For example, SeqID189 shared 99.8% amino acid sequence similarity with ORF1 protein, SeqID2182 shared 99.9% similarity with ORF2 protein and SeqID892 shared 99.7% similarity with ORF4 protein (Supplementary Figure 1). By contrast, the most disparate proteins, SeqID2163 and ORF5, shared only 20% amino acid sequence similarity. Another putative vaccine antigen, the 67 kDa membrane-associated ATPase subunit of Clp protease (ClpX), was initially identified in strain 95/1000, and the product of the gene (Genbank AY466377.1) [123] is identical to SeqID606 (fig|6666666.171271.peg.606, Supplementary Dataset 2) in the genome of *B. piloscoli* B2904. However, we excluded this protein from our RV list, as PSORTB3, CELLO and $SOSUI_{GramN}$ all identified this as a cytoplasmic protein. Moreover, the previous study only identified the ability of recombinant ClpX to generate an antibody response in mice that recognised native protein in Western blots and no data on protection were provided [123]. An earlier study demonstrated that colonic infection with *B. pilosicoli* induced serum IgG in convalescent pigs that reacted with specific OMPs of 47, 54 and 64 kDa, but their identify within our RV list is unknown [124].

It is worth also discussing whether the antigens identified in the first RV study by NAV were present in our RV list. From the patent search data, there were seven high priority antigens identified (Table 3). The amino acids sequences for these seven protiens were run through BLAST against *B. pilosicoli* B2904 to look for homology, and then the RV list in Table 1 was searched for their presence. Of these, AAK14801.1 and AAC10219.1 showed partial similarity to variable surface family protein SeqID387, and AAB47846.1 identifed as a protein homologous to FlaA OMP of *T. pallidum* showed similarity, to various degrees, with the FlaA proteins of *B. pilosicoli* (SeqID2573, SeqID2167, SeqID2026 and SeqID2166). The first 94 N-terminal amino acids of ADK30893.1, the putative virulence factor MviN identified by NAV, and of SeqID462, one of the *B. pilosicoli* MurJ 'flippase' proteins identified by our RV study, were identical. ADK30237.1, the Preprotein translocase, subunit SecA, was similar to SeqID1528, but this was absent in our RV list. Both AAC82625.1 and NP_713795 were absent also from our RV list. Thus, the NAV study and our more contemporary RV study showed some concordance and some disparity, and higlighted the fact that the *B. pilosicoli* proteome contains significantly more potential vaccine candidates.

A potentially important factor for further refinement of the final list could be the possibility of the proteins being allergens or containing possible allergenic regions. We ran the 162 candidate proteins through the AlgPred 2.0 program, which identified 112 of the proteins as possible 'Non-Allergens' and 50 proteins as possible 'Allergens' (Supplementary Dataset 14). These possible 'Allergens' included one cytoplasmic toxin, two extracellular/surface, one extracellular/secreted, four extracellular, one outer membrane/secreted, 17 outer membrane, eight outer membrane/extracellular, one surface and 15 unknown location proteins. Some of these proteins are considered major vaccine candidates, e.g. VSP proteins, RESA, the sialidases, hemolysin (evidently a potential allergen), CsgG curli, TmpB,CiaB, tpn50/OmpA, Cbp, BatC and several flagellar-associated proteins. However, the balance between immunogenicity and allergenicity for candidate proteins should be examined before any decision is made for final exclusion.

There are many proteins within the list of vaccine candidates that are annotated as hypothetical/unkown (Table 1). In order to estimate their potential role, we ran the protein sequences of all 162 candidates throught the STRING database and the NCBI Conserved Domain Database (CDD). Interestingly, both database searches did not provide additional information on the identities, functions or superfamilies of the hypothetical proteins, but did serve to confirm the names and functions of the proteins already annotated in Table 1 using other *in silico* tools (data not shown).

3.11. Presence of RV proteins in other B. pilosicoli strains

In our study, RV was done with just the one genome of *B. pilosicoli* strain B2904, which has been used in several experimental infection studies [39,125–128]. We next examined if the proteins identified in the *B. pilosicoli* B2904 strain RV list were present in other *B. pilosicoli* strains. The 19 strains in Genbank that we examined included 12 that we sequenced recently as part of a study to catalogue the intraspecific genomic diversity of the organism more comprehensively [62], and strains 95/1000, WesB, P43/6/78, SP16, PC538III-hc, and two NCTC strains (13046 and 12874). All of the strains were isolated from chickens, except for 95/1000, P43/6/78 and SP16, which were isolated from pigs, and WesB and PC538III-hc, which were isolated from humans. The provenance of the NCTC strains has not yet been released. We found that 146 out of 162 (90%) of the protein amino acid sequences were present in all 19 strains (Supplementary Dataset 15). There were significant differences in the presence of these 16 proteins amongst the strains. The extracellular proteins SeqID1751 and SeqID382 were absent in 1 strain each, SeqID1638 was absent in 2 strains, and SeqID1949 and SeqID2016 were absent in 8 and 13 strains respectively. OMP SeqID2119 was absent in 12 strains, and the OMP/Extracellular proteins SeqID1901, SeqID2018 and SeqID2017 were absent in 1, 2 and 13 strains, respectively. Of the unknown proteins, SeqID2006, SeqID1081, SeqID2011 and SeqID631 were absent in 1, 13, 13 and 15 strains respectively. Notably, we found 4 proteins that were present only in

strain B2904: SeqID235 (Ngo MtrE OM channel (26% similarity), SeqID145 (hypothetical SPII lipoprotein), SeqID234 (hypothetical SPII lipoprotein) and SeqID233 (hypothetical SPI lipoprotein).

The variable abscence of these proteins suggests that any vaccine containing any of these antigens possibly may not provide wide coverage of circulating strains, and they could probably be excluded as viable vaccine candidates. By contrast, the presence of a large number of proteins amongst strains isolated from chickens, pigs and humans suggests that a cross-species vaccine could be considered for development. Another factor that might complicate vaccine development is that we observed variability within the amino acid sequences for many of the proteins amongst different strains isolated from the different hosts. This suggests both intra- and inter-host immune pressure and could focus vaccine development efforts on the most conserved proteins. However, a limitation of these analyses is that we only examined 19 genomes, and their presence in a larger number of genomes would need to studied before conclusively excluding these potential antigens, or for proposing a vaccine for use across species.

Thus, the use of a single strain in our preliminary study could be viewed as a limitation, but bears similarity with the single isolate RV approach used by Novartis to successfully generate the licensed vaccine Bexsero [41]. A pangenome approach could be useful to potentially identify those genes that are present in the core genome, as well as the potential prediction of pathogenicity islands that are stable across the genomes and contain these genes. However, a major obstacle to these approaches is the paucity of complete genome sequences available for *B. pilosicoli* [62]. A useful tool for doing comparative genomics to identify core and pangenomes is the EDGAR software platform [129], but this only contains the sequences for the *B. pilosicoli* isolates B2904, 95/1000, WesB and P43/6/78. An expanded analysis of variation could be done with the *Brachyspira* isolate database in PubMLST [130], which contains 182 *B. pilosicoli* isolates, but their genome assemblies and complete allele sequences are currently absent - unlike the 26,700 meningococcal genomes in the PubMLST database. Thus, true representative antigen coverage and diversity amongst the *B. pilosicoli* requires signficantly more genomes to interogate.

It is also worth commenting on the use of the RAST program in our study to annotate the genome sequence of strain B2904 for RV. The NCBI entry for this strain has been annotated using the Prokaryotic Genome Annotation Pipeline (PGAP). That two annotation pipelines gave slightly different output, even on an identical assembly, is not without precedent, particularly for RAST versus PGAP [131]. Thus, to a certain extent, work that uses outputs from these pipelines are always at the mercy of their respective idiosyncrasies and determining which is the "best" annotation, is still an unsolved computational problem. We downloaded the NCBI entry (Accession: NC_018607) and compared the annotation to that generated with RAST using a freely available python tool 'compare-annotations'. The output is summarised as follows: Features in RAST assembly = 2679; Features in PGAP assembly = 2507; Features that were an exact match = 2076. Of the features that were different, 238 were a result of longer ORFs identified by PGAP and 35 which were shorter. Of the features that were missing between the two annotations, 158 were in PGAP but not in RAST and 330 were in RAST but not PGAP. In terms of hypothetical annotations, 441 were consistent between annotations. A total of 347 were no longer hypothetical in the PGAP annotation, and 47 hypotheticals in the RAST annotation. When examining the similarities and differences (Supplementary Dataset 16), it is interesting to note that the RAST list contains some proteins identified as candidates that are absent in PGAP list, e.g. BatC, several cell surface proteins, chitin binding protein, MotB, GldG (gliding motility), *Serpulina hyodysenteriae* variable surface protein, ZnuB. No potential candidates that are named in the PGAP list are absent in the RAST list. The majority of proteins were cytoplasmic in localization.

Our defined RV list (Table 1) shares some commonality with the *B. pilosicoli* surfaceome (proteins bound to the cell surface) and exoproteome (proteins present in the extracellular milieu) characterised previously in *B. pilosicoli* using a proteomic shotgun approach [132]. Several virulence factors and potential vaccine candidates, e.g. VspD and FlaA, amongst others, were detected within the surfaceome/exoproteome. Interestingly, it did appear that many of the most abundant proteins in both of these compartments would be excluded as vaccine candidates, due to their localisation within the periplasm, cytoplasm, or inner membrane, and by their functional roles. There is one other report in the literature of an *in silico* analysis of a *Brachyspira spp.*, which was done with only partial genomic sequence data from a single *B. hyodysenteriae* isolate [133], using a combination of SignalP, PSORTb (Version 2.0), Lipop and SpLip, and TMpred (membrane-spanning prediction) programs. Nevertheless, it still served to identify potential *B. hyodysenteriae* vaccine candidate molecules. For future studies, it would be interesting to examine the proteomes of the *B. hyodysenteriae* genome sequences that are publically available, for the presence of antigens in the *B. pilosicoli* RV list for the development of potential pan--*Brachyspira* species vaccines.

3.12. Conclusions and future work

In this study, we have used a genome-based RV approach to generate a candidate list of proteins from *B. pilosicoli* that can be examined as vaccine candidates. The limitations of our study are well-established limitations of RV, e.g. the inability to identify potential non-protein candidates and the possible elimination of vaccine candidates as a result of false negative predictions by any of the bioinformatic tools used. Furthermore, RV tends to generate large lists of candidates that would be time-consuming to process. It is possible that refinements of protein candidate lists could be made using additional filtering algorithm tools [134] or computational pipelines, for example ReVac [135], EpitoCore [136] and PanRV [137]. Regardless, the number of protein candidates identified by any *in silico* approach/pipeline is likely still to be considerable. Indeed, although the number of potential candidates numbers 162, it is worth mentioning that in the first RV study that lead to the development of Bexsero, Novartis identified 570 ORFs from the meningococcal strain MC58, from which they were able to express 350 in *E.coli* and use the recombinant forms of each protein to vaccine mice and examine biological responses. Murine sera were then screened by western blot analysis of total cell lysates or OM vesicles to evaluate whether the protein was actually expressed by the bacteria and to determine its localization. Surface expression/secretion of the protein was then tested by ELISA and flow cytometry on whole-cell bacteria. Functional bactericidal assays were done to evaluate the complement-mediated killing activity of the antibodies, since bactericidal activity is the known correlate of protection for humans. The team demonstrated the surface expression for 91 of the proteins and 29 were able to induce a bactericidal response. Further examination of alleic diversity and expression levels reduced this number eventually to 5 antigens for inclusion in the vaccine [41].

Thus, testing vaccine potential of *B. pilosicoli* candidates would involve, initally, the production of individual recombinant proteins, which would likely lead to the exclusion of some proteins due to technical issues associated with their expression and/or purification. For example, within the *B. pilosicoli* RV list, there are 11 proteins of molecular weight > 100 kDa (i.e. Extracellular protein SeqID2139; OMPs SeqID2620, SeqID2235, SeqID1303, SeqID1904, SeqID980, SeqID693, SeqID2621, SeqID1221, SeqID1387; and Extracellular/OMP protein SeqID724), which might be challenging to produce. An informed choice on which proteins to prioritise for testing may be necessary for practical reasons: for example, the first collection of candidates could be known OMPs and extracellular proteins that have homologues in other organisms that are involved in pathogenesis and may have been examined for vaccine potential. This collection could include proteins such as the variable surface RESA-like protein, the two sialidases, hyicolysin, TmpB, CsgG curli, Tia invasion determinant, OmpA, CiaB, haemolysin, OppA family proteins, YjhT, BatC, MtrE OMP, VspH, GldG, CBP and Tpn50

Table 4
Linear B-cell epitopes chosen from a selection of *B. pilosicoli* vaccine candidates to construct a B-cell chimera protein antigen.

	Protein	Peptide sequence
1	fig|6666666.171271.peg.2319 - Hypothetical protein/weak sequence similarity to Ring-infected erythrocyte surface antigen P. falciparum	YIQNNNKEYMDMNKIK
2	fig|6666666.171271.peg.1802 - Sialidase (EC 3.2.1.18)	KKASTAPGGEGPTKEDEVKPP
3	fig|6666666.171271.peg.1804 - Sialidase (EC 3.2.1.18)	GGLEPPKEEDNTAGK
4	fig|6666666.171271.peg.2646 - Peptidase M30, hyicolysin//cell wall binding repeat	EIYKDIANAPEEYKHTY
5	fig|6666666.171271.peg.840 - Peptidoglycan-binding protein LysM// TmpB, Treponemal (spirochaetal) membrane protein	LFPKYYKVQYRKVG
6	fig|6666666.171271.peg.1529 - FIG00439356: hypothetical protein// CsgG curli production assembly transport component	GLTKNTRFKVYS
7	fig|6666666.171271.peg.2232 - Tia invasion determinant virulence	PIRVEFEYLYKNGLEVNNYPNNID
8	fig|6666666.171271.peg.2472 - OmpA superfamily similarities// Flagellar Motor Protein	EGHIDSSEVRYMNKNTVYN
9	fig|6666666.171271.peg.693 - Cell division protein FtsH (EC 3.4.24.-)//CiaB, TPR repeat	DDTSEKEEPKQEDTDNLDDLDSILD
10	fig|6666666.171271.peg.2613 - Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA (TC 3.A.1.5.1)// extracellular solute-binding protein, family 9	MTERKTDEKIVMEVNTNYYDKESI
11	fig|6666666.171271.peg.429 - Sialic acid-induced transmembrane protein YjhT(NanM), possible mutarotase	GKQSGKGSDKVYEYDLATKETKELAPVPNQASR
12	fig|6666666.171271.peg.1294 - BatC//PFAM Sporulation and spore germination	TPPSYMRNNNEP
13	fig|6666666.171271.peg.235 - Ngo MtrE OM channel (26% similarity)	DRQRLDKLMKEVLLQQTSGMVDE
14	fig|6666666.171271.peg.434 - Variable surface protein VspH	AALATTYKNIDEANRG
15	fig|6666666.171271.peg.1083 - Gliding motility protein GldG ABC transporter//PFAM ABC-type uncharacterised transport system	EYKSAAGKNISFEIIDA
16	fig|6666666.171271.peg.1772 - Chitin binding protein	QEDSVIKEQIRTKIQQYNKDK
17	fig|6666666.171271.peg.878 - Flagellar motor rotation protein MotB// tpn50, ompA family	MALSADYLFANIETNENNDPL
18	fig|6666666.171271.peg.1917 - Hemolysin	LFKIGESYYNEKNYNSA

All epitopes were selected using IEDB (Emini Surface Acessibility) and ABCpred (Artificial neural network-based B-cell epitope prediction server). Gray highlighted amino acids are results from IEDB and underlined amino acids are results from ABCpred.

OMP (Table 4).

However, an alternative strategy to generating individual recombinant proteins is to potentially develop a chimera antigen *in silico* containing predicted linear B-cell epitopes from this first collection of candidate proteins. This would be a strategy for generating immune protection principally dependent on antibody production. Although the mechanisms of host immunity and protection against *B. pilosicoli* infection are poorly understood [2], both mice [138] and pigs [82] experimentally infected with *B. pilosicoli* strains have been reported to develop serum antibody responses, suggesting that B-cell antibody mediated immune responses may be important. In our study, we identified the linear B-cell epitope with the best antigenic and physiochemical properties for each of the 18 proteins using IEDB Emini Surface Acessibility and ABCpred and these are shown in Table 4. The complete data for each individual protein are shown in Supplementary Dataset 17. We then constructed the *B. pilosicoli* linear B-cell chimera antigen with the peptide order as shown in Fig. 3, with GPGPG used as a spacer to provide flexibility and K and C residues at both the C- and N-terminus to improve protein stability. The chimera antigen had 437 residues, a molecular weight of ~46.8 kDa, good water solubility, and it was stable with appropriate half-lives (Fig. 3). Final Blast-p analysis of the chimera antigen showed that of the 29 sequences identified with similarity, 22 principally belonged to *B. pilosicoli*, three to *B. hampsonii* and one to *B. intermedia*, and there was very low and irrelevant similarity with a canarypox virus protein (27%), a hypothetical protein from the brownbanded bambooshark *Chiloscyllium punctatum* (24%) and a trichohyalin-like protein from the American lobster *Homarus americanus* (Supplementary Dataset 18). Final ABCpred analysis showed that the epitopes within the B-cell chimera antigen were still predicted to be linear B-cell epitopes (data not shown).

Close examination of the proteins selected for producing a B-cell chimera antigen showed that three of them had homologues by name within the RV list; thus, OppA family contained 9 named proteins, and there were two proteins that were named TmpB and two named BatC (Table 1). Full sequence alignments are shown in Supplementary Dataset 19. Using ABCpred to identify the putative linear B-cell epitopes, there is good homology between the different OppA proteins within the linear B-cell epitope sequence that is in the chimera antigen (Fig. 4; Supplementary Dataset 20), whereas there was only a small epitope region of homology in the BatC protein SeqID1490 compared with SeqID1294, and no homology in the TmpB protein SeqID2554 compared with SeqID840 (Fig. 4). Thus, using Emini and ABCpred, we identified the sequence RDRLISEGYLTKDSEEEQKLSQTI as a predicted linear B-cell epitope in TmpB SeqID2554 and sequence KLPTSPEIYYNMGT as a predicted linear B-cell epitope in BatC SeqID1490, although this protein did not appear to have good predicted immunogenicity (Supplementary Dataset 21). There is an option, therefore, to insert these new epitopes into the B-cell chimera and/or use them to replace the epitope belonging to the other similarly named protein.

However, a limitation of the B-cell chimera approach is that the selected B-cell epitopes are likely to be continuous/linear rather than discontinuous/conformational. This is a potential issue for immunogenicity as recognition of conformational B-cell epitopes by the immune system is normally essential for inducing protection. In addition to the

A) *B. pilosicoli* B-cell chimera protein amino acid sequence

CKKASTAPGGEGPTKEDEVKPPGPGPGEGHIDSSEVRYMNKNTVYNGPGPGGGLEPPKEEDNTAGKGPGPGPIRVEFEYLYKNGLEVNNYPNNIDGPGPGD
RQRLDKLMKEVLLQQTSGMVDEGPGPGEYKSAAGKNISFEIIDAGPGPGEYKSAAGKNISFEIIDAGPGPGMTERKTDEKIVMEVNTNYYDKESIGPGPGGL
TKNTRFKVYGPGPGGKQSGKGSDKVYEYDLATKETKELAPVPNQASGPGPGDDTSEKEEPKQEDTDNLDDLDSILDGPGPGMALSADYLFANIETNENNDPL
GPGPGYIQNNNKEYMDMNKIKGPGPGEIYKDIANAPEEYKHTYGPGPGLFPKYYKVQYRKVGGPGPGLFKIGESYYNEKNYNSAGPGPGQEDSVIKEQIRTK
IQQYNKDKGPGPGTPPSYMRNNNEPKKKC

ORDER OF EPITOPES: 2-8-3-7-13-15-14-10-6-11-9-17-1-4-5-18-16-12

B) PepCalc and ProtParam physiochemical properties of *B. pilosicoli* B-cell chimera protein

Physiochemical Properties	
Number of residues:	437
Molecular weight:	46815.2 g/mol
Extinction coefficient:	33280 M^{-1} cm^{-1}
Iso-electric point:	pH 4.7
Net charge at pH 7:	-12.9
Estimated solubility:	Good water solubility

Estimated half-life
The N-terminal of the sequence considered is C (Cys)

The estimated half-life is: 1.2 hours (mammalian reticulocytes, in vitro)
>20 hours (yeast, in vivo)
>10 hours (Escherichia coli, in vivo)

Instability index:

The instability index (II) is computed to be 33.38
This classifies the protein as stable

Aliphatic index: 48.88

Grand average of hydropathicity (GRAVY): -1.084

Fig. 3. Amino acid sequence and physiochemical properties of the *B. pilosicoli* B-cell chimera antigen. A) The peptide order was chosen based on their original position in their parent protein, i.e. epitopes closer to C-terminus in the original protein were placed closer to C-terminus of the chimera protein, and similarly for N-terminus located proteins. B) PepCalc and ProtParam were used to examine the physiochemical properties of the chimera antigen.

OppA Family

```
B-cell Chimera                IIADGPGPGEYKSAAGKNISFEIIDAGPGPGMTERKTDEKIVMEVNTNYYDKESIGPGPG   202
fig|6666666.171271.peg.2613   AYFPVRKDIVEKYGDDWSRNPETYIVNGAYVMTERKTDEKIVMEVNTNYYDKESIVAKKI   236
fig|6666666.171271.peg.763    CYVPVREDIINKYGDDWTWNSESYIGNGAYKMTERKPDELIAFELNTNYWDYKNQVAKKI   249
fig|6666666.171271.peg.2276   IFMPLREDIINTYGDEWTLKPETYIVNGAYTMTERLADEKIVFSANKNYYNPNEQVAKKI   236
fig|6666666.171271.peg.2182   PFYPVREDIINEYGDKWTLKPATYIGNGAFKMTERNFDKSIILERNTNYWNNENTKPNKL   235
fig|6666666.171271.peg.487    IFSPLRADYIEDN-EKWTFYPNTYIGNGPYKMIERKVDESISLELNTNYWNKDNMVANRI   231
fig|6666666.171271.peg.189    IYSPLRKDIIEKYGDAWTQ--EEYIGNGPFFVKEHINNDKIVMEKNTNYWNKETIIPQTI   228
fig|6666666.171271.peg.892    TFYPVRKDIIEKYGDSWSLNVESYIGNGPFVTTEINQDESIIMVKNTNYWAFNDVVPEKL   238
fig|6666666.171271.peg.894    TFYPIREDIIEKYGDTWTINPESYIGNGSYVMIERNIDNNIIMVKNTNYWDYENLVPEKI   239
fig|6666666.171271.peg.2198   TFYPVRKDIIEENKDNWTLSPDTYIGNGPFTLVERRTDDRLVMIKNTNYWNSENIIPEKL   235
                                                          *   :. : :  *.**:  .
```

TmpB Family

```
B-cell Chimera                APEGPGPGLFPKYYKVQYRKVGGPGPGLFKIGESYYNEKNYNSAGPGPGQEDSVIKEQIR   406
fig|6666666.171271.peg.840    TIVETTGVLFPKYYKVQYRKVGTDSLWKIASYDFIYGDGNLWKKIYEANKDKIKD---PN   342
fig|6666666.171271.peg.2554   VNTEGKVTVLPQYYVVVRRVPLTDALWRIAGYSYIYNNPIEWYRIYEANRNILRDPDNPD   517
```

BatC Family

```
B-cell Chimera                RTKIQQYNKDKQEDSVIKEQIRTKIQQYNKDKGPGPGTPPSYMRNNNEPKKKC          437
fig|6666666.171271.peg.1294   ANNSNISSSIFYDILKPKENTTRDIKTMRPATTTYSSTPPSYMRNNNEPKEKYNTNNYNY   304
fig|6666666.171271.peg.1490   NNNQN----------DEEQNAQTDIDRLLDSLRQ--------YRKDKENDQQYYGGGRIDK   288
                                                          * :::::
```

Fig. 4. Comparison of B-cell epitopes in OppA family proteins and TmpB and BactC proteins. The linear B-cell epitope identified in the chimera antigen and the source proteins are highlighted in yellow. The epitope regions in the homologue/family proteins for OppA and BatC are shown in green. OppA SeqID487 has the lowest similarity, which is probably due to the most variability of the amino acids within the sequence. BatC SeqID1490 has a small epitope region similarity of four amino acids and no epitope similarity was identified in TmpB SeqID2554.

Fig. 5. *In silico* structure prediction using ProteinPredict and SWISS-MODEL. The B-cell chimera amino acid sequence was run through the ProteinPredict and SWISS-MODEL programs; the former generated a set of protein parameters including predicted secondary structure, whereas the latter could only predict structure for a short sequence of amino acids (proline 100 to glycine 149).

physiochemical properties of the polypeptide shown in Fig. 3, we attempted *in silico* structure prediction using the ProteinPredict program. This showed that the secondary structure of the chimera contained a small amount of helix and strand, but was dominated by other probably disordered regions (Fig. 5). Solvent accessibility was predominately exposed. The Relative B-Value (PROFbval) predicts flexible and rigid residues in proteins, and the chimera has intermediate values in the majority. High PROFbval values (flexible) have been reported to correlate with biological activities such as antigenic recognition and catalytic activity, whereas low values (rigid) correlate with active sites in enzymes [70,71]. However, the B-cell chimera sequence could not be aligned to any other sequence, and SWISS-MODEL could only generate secondary structure from proline 100 to glycine 149, which consisted of α-helix and disordered region (Fig. 5).

Although we have developed a candidate *B. pilosicoli* B-cell chimera antigen, the possibility of T-cell mediated immune responses being important for protection should not be ignored. However, for predictive T-cell epitope mapping and MHC-1 binding, neither IEDB nor the NetCTLpan 1.1 Server [139] contain chicken as a MHC source species, but do have pig and human and selections of MHC alleles. In addition, neither pig nor chicken is available as a species for MHC-II binding predictions in either IEDB or the NetMHCII 2.3 server [140]. Thus, this makes it difficult to identify T-cell epitopes for *B. pilosicoli* protein vaccines for pigs and chickens; by contrast, T-cell epitopes can be identified for proteins that could be used in humans, but this is of a lesser priority for vaccine development than vaccines for veterinary use.

Studies subsequent to examining the first collection of candidate proteins could involve interrogating the remainder of the list (Table 1), which includes the many hypothetical SPII OM/Extracellular lipoproteins that could provide effective vaccine antigens. Furthermore, identifying the function(s) of these hypothetical proteins would increase our knowledge of the biology of *B. pilosicoli.* However, it should be noted that there are no reports, to our knowldege, on the use of *B. pilosicoli* OM vesicles (OMV) as a potential vaccine, which is perhaps surprising, given that OMV could contain many of the potential candidates in our RV list and provide cheap manufacture. The use of formalinised whole bacterial cells has been shown to be ineffective [141].

Vaccination of chickens and/or pigs would be used to assess if recombinant protein and chimera antigen candidates are immunogenic and able to confer protection against spirochaetosis caused by subsequent *Brachyspira* infection. A consideration of adjuvant and/or delivery system may be needed at this stage, and the use of aluminium hydroxide and commercial oil emulsions found in many licensed animal vaccines may be inappropriate, if protein conformation, especially for OMPs, is important for functional immune responses [142,143]. Ranking the efficacy of protective antigens in these models would identify the most promising proteins for the production of either single antigen vaccine or combination vaccines. Vaccination studies would demonstrate also whether vaccines containing multiple recombinant proteins could be replaced by a B-cell chimera antigen to generate an antibody-dependent immune response. The obvious advantage of using chimera antigens is targeted epitope specificity and the manufacture of a single recombinant protein antigen. Furthermore, choosing antigens with minimal protein amino acid sequence variability and stable expression amongst *B. piosicoli* isolates, or peptide epitopes that are conserved across the proteins expressed by *B. pilosicoli*, would be factors for final selection.

Thus, our broad and contemporary RV study is the starting point for a comprehensive vaccine development strategy for preventing AIS principally and could lead to new vaccines within this decade.

CRediT authorship contribution statement

MC, DdO, RMdA and MVH carried out the *in silico* analyses; MC wrote the first draft; MC, DdO, RMdA, MVH and RML reviewed and edited and revised the manuscript.

Conflict of interest statement

The authors have no competing interests.

Data Availability

All data generated in this study and presented in Supplementary Datasets 1 – 21 and Supplementary Figure 1 at url https://eprints.soton.ac.uk/449222/.

Acknowledgements

MC and RML were in receipt of a Houghton Trust UK (Avian Pathology) research award (unnumbered) for research into diseases of avian species and domestic poultry. We are grateful to Dr. Maria Scarselli, Glaxo Smith Kline Vaccines Srl, Siena, Italy, for instruction in *in silico* reverse vaccinology.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.procbio.2022.08.014.

References

[1] W. Ludwig, J. Euzeby, W.B. Whitman, *Draft taxonomic outline of the Bacteroidetes, Planctomycetes, chlamydiae, spirochaetes, Fibrobac-teres, Fusobacteria, Acidobacteria, Verrucomicrobia, Dictyoglomi and Gemmatimonadetes*. Bergey Vol 4 ⟨www.bergeysorg/outlines/2008⟩, p. Outlinepdf01/05/12 2008.
[2] D.J. Hampson, The spirochete *Brachyspira pilosicoli*, enteric pathogen of animals and humans, Clin. Microbiol. Rev. 31 (1) (2018).
[3] M. Mushtaq, S. Zubair, T. Rasback, E. Bongcam-Rudloff, D.S. Jansson, *Brachyspira suanatina* sp. nov., an enteropathogenic intestinal spirochaete isolated from pigs and mallards: genomic and phenotypic characteristics, BMC Microbiol 15 (2015) 208.
[4] T. La, N.D. Phillips, D.J. Hampson, An investigation into the etiological agents of swine dysentery in australian pig herds, PLoS One 11 (12) (2016), e0167424.
[5] N.S. Mirajkar, N.D. Phillips, T. La, D.J. Hampson, C.J. Gebhart, Characterization and recognition of *Brachyspira hampsonii* sp. nov., a novel intestinal spirochete that is pathogenic to pigs, J. Clin. Microbiol. 54 (12) (2016) 2942–2949.
[6] J. Osorio, A. Carvajal, G. Naharro, P. Rubio, T. La, N.D. Phillips, D.J. Hampson, Identification of weakly haemolytic *Brachyspira* isolates recovered from pigs with diarrhoea in Spain and Portugal and comparison with results from other countries, Res. Vet. Sci. 95 (3) (2013) 861–869.
[7] L.J. Mappley, R.M. La Ragione, M.J. Woodward, *Brachyspira* and its role in avian intestinal spirochaetosis, Vet. Microbiol. 168 (2–4) (2014) 245–260.
[8] D.G.S. Burch, B.W. Strugnell, A. Steventon, E.N. Watson, C. Harding, Survey of 222 flocks in Great Britain for the presence of *Brachyspira* species and their effect on production. Proceedings of the 5th International Conference on Colonic Spirochaetal Infections in Animals and Humans Leon, Spain 2009.
[9] K. Hovind-Hougen, A. Birch-Andersen, R. Henrik-Nielsen, M. Orholm, J. O. Pedersen, P.S. Teglbjaerg, E.H. Thaysen, Intestinal spirochetosis: morphological characterization and cultivation of the spirochete *Brachyspira aalborgi* gen. nov., sp. nov, J. Clin. Microbiol. 16 (6) (1982) 1127–1136.
[10] J.L. Smith, Colonic spirochetosis in animals and humans, J. Food Prot. 68 (7) (2005) 1525–1534.
[11] C.P. Stephens, D.J. Hampson, Intestinal spirochete infections of chickens: a review of disease associations, epidemiology and control, Anim. Health Res Rev. 2 (1) (2001) 83–91.
[12] C.I. Le Roy, L.J. Mappley, R.M. La Ragione, M.J. Woodward, S.P. Claus, *Brachyspira pilosicoli*-induced avian intestinal spirochaetosis, Micro Ecol. Health Dis. 26 (2015) 28853.
[13] S.L. Oxberry, D.J. Trott, D.J. Hampson, *Serpulina pilosicoli*, waterbirds and water: potential sources of infection for humans and other animals, Epidemiol. Infect. 121 (1) (1998) 219–225.
[14] D.S. Jansson, M. Persson, U. Zimmerman, K.E. Johansson, Phenotypic and genetic diversity among intestinal spirochaetes (genus *Brachyspira*) in free-living wild mallards (*Anas platyrhynchos*) sampled in southern Sweden, Syst. Appl. Microbiol. 34 (8) (2011) 566–575.
[15] H.L. Shivaprasad, G.E. Duhamel, Cecal spirochetosis caused by *Brachyspira pilosicoli* in commercial turkeys, Avian Dis. 49 (4) (2005) 609–613.
[16] D.M. Webb, G.E. Duhamel, M.R. Mathiesen, N. Muniappa, A.K. White, Cecal spirochetosis associated with *Serpulina pilosicoli* in captive juvenile ring-necked pheasants, Avian Dis. 41 (4) (1997) 997–1002.
[17] D.S. Jansson, C. Brojer, D. Gavier-Widen, A. Gunnarsson, C. Fellstrom, *Brachyspira spp. (Serpulina spp.)* in birds: a review and results from a study of Swedish game birds, Anim. Health Res. Rev. 2 (1) (2001) 93–100.

[18] A. Backhans, K.E. Johansson, C. Fellstrom, Phenotypic and molecular characterization of *Brachyspira spp*. isolated from wild rodents, Environ. Microbiol. Rep. 2 (6) (2010) 720–727.
[19] A. Backhans, D.S. Jansson, A. Aspan, C. Fellstrom, Typing of *Brachyspira spp*. from rodents, pigs and chickens on Swedish farms, Vet. Microbiol. 153 (1–2) (2011) 156–162.
[20] S.L. Oxberry, D.J. Hampson, Colonisation of pet shop puppies with *Brachyspira pilosicoli*, Vet. Microbiol. 93 (2) (2003) 167–174.
[21] G.E. Duhamel, D.J. Trott, N. Muniappa, M.R. Mathiesen, K. Tarasiuk, J.I. Lee, D. J. Hampson, Canine intestinal spirochetes consist of *Serpulina pilosicoli* and a newly identified group provisionally designated "*Serpulina canis*" sp. nov, J. Clin. Microbiol 36 (8) (1998) 2264–2270.
[22] A. Hidalgo, P. Rubio, J. Osorio, A. Carvajal, Prevalence of *Brachyspira pilosicoli* and "*Brachyspira canis*" in dogs and their association with diarrhoea, Vet. Microbiol. 146 (3–4) (2010) 356–360.
[23] D.J. Hampson, S.L. Oxberry, T. La, Potential for zoonotic transmission of *Brachyspira pilosicoli*, Emerg. Infect. Dis. 12 (5) (2006) 869–870.
[24] D.J. Trott, T.B. Stanton, N.S. Jensen, G.E. Duhamel, J.L. Johnson, D.J. Hampson, *Serpulina pilosicoli* sp. nov., the agent of porcine intestinal spirochetosis, Int. J. Syst. Bacteriol. 46 (1) (1996) 206–215.
[25] D.J. Hampson, C.P. Stephens, S.L. Oxberry, Antimicrobial susceptibility testing of *Brachyspira intermedia* and *Brachyspira pilosicoli* isolates from Australian chickens, Avian Pathol. 35 (1) (2006) 12–16.
[26] L. Bano, G. Merialdi, P. Bonilauri, G. Dall'Anese, K. Capello, D. Comin, G. Cattoli, V. Sanguinetti, D.J. Hampson, F. Agnoletti, Prevalence, disease associations and risk factors for colonization with intestinal spirochaetes (*Brachyspira* spp.) in flocks of laying hens in north-eastern Italy, Avian Pathol. 37 (3) (2008) 281–286.
[27] S.E. Myers, P.A. Dunn, N.D. Phillips, T. La, D.J. Hampson, *Brachyspira intermedia* and *Brachyspira pilosicoli* are commonly found in older laying flocks in Pennsylvania, Avian Dis. 53 (4) (2009) 533–537.
[28] M.M. Amin, N.D. Phillips, T. La, I.D. Robertson, D.J. Hampson, Intestinal spirochaetes (*Brachyspira* spp.) colonizing flocks of layer and breeder chickens in Malaysia, Avian Pathol. 43 (6) (2014) 501–505.
[29] M.J. Zarabi, A. Khanzadi, S. Razmyar, J. Intestinal, colonization of different *Brachyspira* spp. in laying hens, Iran. J. Vet. Med. 8 (3) (2014) 213–218.
[30] N.L. Trivett-Moore, G.L. Gilbert, C.L. Law, D.J. Trott, D.J. Hampson, Isolation of *Serpulina pilosicoli* from rectal biopsy specimens showing evidence of intestinal spirochetosis, J. Clin. Microbiol 36 (1) (1998) 261–265.
[31] A.S. Mikosza, T. La, W.B. de Boer, D.J. Hampson, Comparative prevalences of *Brachyspira aalborgi* and *Brachyspira (Serpulina) pilosicoli* as etiologic agents of histologically identified intestinal spirochetosis in Australia, J. Clin. Microbiol. 39 (1) (2001) 347–350.
[32] D.F. Carpentieri, S. Souza-Morones, J.S. Gardetto, H.M. Ross, K. Downey, K. Ingebo, E. Siaw, Intestinal spirochetosis in children: five new cases and a 20-year review of the literature, Pedia Dev. Pathol. 13 (6) (2010) 471–475.
[33] A. Gad, R. Willen, K. Furugard, B. Fors, M. Hradsky, Intestinal spirochaetosis as a cause of longstanding diarrhoea, Ups. J. Med. Sci. 82 (1) (1977) 49–54.
[34] J. Gan, C. Bryant, D. Arul, C. Parmar, Intestinal spirochaetosis mimicking acute appendicitis with review of the literature, BMJ Case Rep. 2017 (2017).
[35] L. Marthinsen, R. Willen, B. Carlen, E. Lindberg, G. Varendh, Intestinal spirochetosis in eight pediatric patients from Southern Sweden, APMIS 110 (7–8) (2002) 571–579.
[36] S. Nishii, M. Higashiyama, S. Ogata, S. Komoto, S. Ito, A. Mizoguchi, H. Terada, H. Furuhashi, T. Takajo, K. Shirakabe, C. Watanabe, K. Tomita, S. Nagao, S. Miura, R. Hokari, Human intestinal spirochetosis mimicking ulcerative colitis, Clin. J. Gastroenterol. 11 (2) (2018) 145–149.
[37] D.J. Trott, A.S. Mikosza, B.G. Combs, S.L. Oxberry, D.J. Hampson, Population genetic analysis of *Serpulina pilosicoli* and its molecular epidemiology in villages in the eastern Highlands of Papua New Guinea, Int. J. Syst. Bacteriol. 48 (Pt 3) (1998) 659–668.
[38] D.J. Hampson, K. Lugsomya, T. La, N.D. Phillips, D.J. Trott, S. Abraham, Antimicrobial resistance in *Brachyspira* - an increasing problem for disease control, Vet. Microbiol. 229 (2019) 59–71.
[39] L.J. Mappley, M.A. Tchórzewska, A. Nunez, M.J. Woodward, P.M. Bramley, R. M. La Ragione, Oral treatment of chickens with *Lactobacillus reuteri* LM1 reduces *Brachyspira pilosicoli*-induced pathology, J. Med. Microbiol. 62 (Pt 2) (2013) 287–296.
[40] P. Wanchanthuek, M.I. Bellgard, T. La, K. Ryan, P. Moolhuijzen, B. Chapman, M. Black, D. Schibeci, A. Hunter, R. Barrero, N.D. Phillips, D.J. Hampson, The complete genome sequence of the pathogenic intestinal spirochete *Brachyspira pilosicoli* and comparison with other *Brachyspira* genomes, PLoS One 5 (7) (2010), e11455.
[41] M. Pizza, V. Scarlato, V. Masignani, M.M. Giuliani, B. Arico, 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 287 (5459) (2000) 1816–1820.
[42] E.W. Sayers, M. Cavanaugh, K. Clark, J. Ostell, K.D. Pruitt, I. Karsch-Mizrachi, GenBank, Nucleic Acids Res. 47 (D1) (2019) D94–D99.
[43] R.K. Aziz, D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, G.J. Olsen, R. Olson, A.L. Osterman, R. A. Overbeek, L.K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G.D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST server: rapid annotations using subsystems technology, BMC Genom. 9 (2008) 75.
[44] N.Y. Yu, J.R. Wagner, M.R. Laird, G. Melli, S. Rey, R. Lo, P. Dao, S.C. Sahinalp, M. Ester, L.J. Foster, F.S. Brinkman, PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes, Bioinformatics 26 (13) (2010) 1608–1615.
[45] C.S. Yu, C.J. Lin, J.K. Hwang, Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions, Protein Sci. 13 (5) (2004) 1402–1406.
[46] K. Imai, N. Asakawa, T. Tsuji, F. Akazawa, A. Ino, M. Sonoyama, S. Mitaku, SOSUI-GramN: high performance prediction for sub-cellular localization of proteins in gram-negative bacteria, Bioinformation 2 (9) (2008) 417–421.
[47] A.S. Juncker, H. Willenbrock, G. Von Heijne, S. Brunak, H. Nielsen, A. Krogh, Prediction of lipoprotein signal peptides in Gram-negative bacteria, Protein Sci. 12 (8) (2003) 1652–1662.
[48] J.J. Almagro Armenteros, K.D. Tsirigos, C.K. Sonderby, T.N. Petersen, O. Winther, S. Brunak, G. von Heijne, H. Nielsen, SignalP 5.0 improves signal peptide predictions using deep neural networks, Nat. Biotechnol. 37 (4) (2019) 420–423.
[49] H. Nielsen, Predicting secretory proteins with SignalP, Methods Mol. Biol. 2017 (1611) 59–73.
[50] M.B. Ulmschneider, M.S.P. Sansom, Amino acid distributions in integral membrane protein structures, Biochim. Et. Biophys. Acta BBA Biomembr. 1512 (1) (2001) 1–14.
[51] S. Möller, M.D. Croning, R. Apweiler, Evaluation of methods for the prediction of membrane spanning regions, Bioinformatics 17 (7) (2001) 646–653.
[52] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (3) (1990) 403–410.
[53] G.M. Boratyn, A.A. Schaffer, R. Agarwala, S.F. Altschul, D.J. Lipman, T. L. Madden, Domain enhanced lookup time accelerated BLAST, Biol. Direct 7 (2012) 12.
[54] S.K. Burley, H.M. Berman, G.J. Kleywegt, J.L. Markley, H. Nakamura, S. Velankar, Protein data bank (PDB): the single global macromolecular structure archive, Methods Mol. Biol. 1607 (2017) 627–641.
[55] A. de Marco, Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli, Micro Cell Fact. 8 (2009) 26.
[56] R. Overbeek, T. Begley, R.M. Butler, J.V. Choudhuri, H.Y. Chuang, M. Cohoon, V. de Crécy-Lagard, N. Diaz, T. Disz, R. Edwards, M. Fonstein, E.D. Frank, S. Gerdes, E.M. Glass, A. Goesmann, A. Hanson, D. Iwata-Reuyl, R. Jensen, N. Jamshidi, L. Krause, M. Kubal, N. Larsen, B. Linke, A.C. McHardy, F. Meyer, H. Neuweger, G. Olsen, R. Olson, A. Osterman, V. Portnoy, G.D. Pusch, D. A. Rodionov, C. Rückert, J. Steiner, R. Stevens, I. Thiele, O. Vassieva, Y. Ye, O. Zagnitko, V. Vonstein, The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes, Nucleic Acids Res. 33 (17) (2005) 5691–5702.
[57] J. Huerta-Cepas, D. Szklarczyk, D. Heller, A. Hernandez-Plaza, S.K. Forslund, H. Cook, D.R. Mende, I. Letunic, T. Rattei, L.J. Jensen, C. von Mering, P. Bork, eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses, Nucleic Acids Res. 47 (D1) (2019) D309–D314.
[58] C. UniProt, UniProt: the universal protein knowledgebase in 2021, Nucleic Acids Res. 49 (D1) (2021) D480–D489.
[59] N. Sharma, S. Patiyal, A. Dhall, A. Pande, C. Arora, G.P.S. Raghava, AlgPred 2.0: an improved method for predicting allergenic proteins and mapping of IgE epitopes, Brief. Bioinf. 22 (4) (2021).
[60] D. Szklarczyk, A.L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N.T. Doncheva, J.H. Morris, P. Bork, L.J. Jensen, C.V. Mering, STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, Nucleic Acids Res. 47 (D1) (2019) D607–D613.
[61] A. Marchler-Bauer, Y. Bo, L. Han, J. He, C.J. Lanczycki, S. Lu, F. Chitsaz, M. K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, F. Lu, G. H. Marchler, J.S. Song, N. Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng, L.Y. Geer, S.H. Bryant, CDD/SPARCLE: functional classification of proteins via subfamily domain architectures, Nucleic Acids Res. 45 (D1) (2017) D200–D203.
[62] A. Pandey, M.V. Humbert, A. Jackson, J.L. Passey, D.J. Hampson, D.W. Cleary, R. M. La Ragione, M. Christodoulides, Evidence of homologous recombination as a driver of diversity in *Brachyspira pilosicoli*, Micro Genom. 6 (12) (2020).
[63] I.A. Doytchinova, D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, BMC Bioinforma. 8 (2007) 4.
[64] R. Vita, S. Mahajan, J.A. Overton, S.K. Dhanda, S. Martini, J.R. Cantrell, D. K. Wheeler, A. Sette, B. Peters, The Immune Epitope Database (IEDB): 2018 update, Nucleic Acids Res. 47 (D1) (2019) D339–D343.
[65] E. Ong, M.F. Cooke, A. Huffman, Z. Xiang, M.U. Wong, H. Wang, M. Seetharaman, N. Valdez, Y. He, Vaxign2: the second generation of the first Web-based vaccine design program using reverse vaccinology and machine learning, Nucleic Acids Res. 49 (W1) (2021) W671–W678.
[66] E.A. Emini, J.V. Hughes, D.S. Perlow, J. Boger, Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide, J. Virol. 55 (3) (1985) 836–839.
[67] S. Saha, G.P. Raghava, Prediction of continuous B-cell epitopes in an antigen using recurrent neural network, Proteins 65 (1) (2006) 40–48.
[68] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. Wilkins, R. Appel, A. Bairoch, Protein Identification and Analysis Tools on the ExPASy Server, in: J. M. Walker (Ed.), The Proteomics Protocols Handbook, Humana Press, New York, 2005, pp. 571–607.

[69] S. Lear, S.L. Cobb, Pep-Calc.com: a set of web utilities for the calculation of peptide and peptoid properties and automatic mass spectral peak assignment, J. Comput. Aided Mol. Des. 30 (3) (2016) 271–277.

[70] A. Schlessinger, G. Yachdav, B. Rost, PROFbval: predict flexible and rigid residues in proteins, Bioinformatics 22 (7) (2006) 891–893.

[71] M. Bernhofer, C. Dallago, T. Karl, V. Satagopam, M. Heinzinger, M. Littmann, T. Olenyi, J. Qiu, K. Schütze, G. Yachdav, H. Ashkenazy, N. Ben-Tal, Y. Bromberg, T. Goldberg, L. Kajan, S. O'Donoghue, C. Sander, A. Schafferhans, A. Schlessinger, G. Vriend, M. Mirdita, P. Gawron, W. Gu, Y. Jarosz, C. Trefois, M. Steinegger, R. Schneider, B. Rost, PredictProtein - predicting protein structure and function for 29 years, Nucleic Acids Res. 49 (W1) (2021) W535–W540.

[72] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T.A.P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: homology modelling of protein structures and complexes, Nucleic Acids Res. 46 (W1) (2018) W296–W303.

[73] W.R. Zückert, Secretion of bacterial lipoproteins: through the cytoplasmic membrane, the periplasm and beyond, Biochim. Biophys. Acta 1843 (8) (2014) 1509–1516.

[74] A.L. Lewis, W.G. Lewis, Host sialoglycans and bacterial sialidases: a mucosal perspective, Cell Microbiol. 14 (8) (2012) 1174–1182.

[75] S. Ayora, P.E. Lindgren, F. Götz, Biochemical properties of a novel metalloprotease from *Staphylococcus hyicus subsp. hyicus* involved in extracellular lipase processing, J. Bacteriol. 176 (11) (1994) 3218–3223.

[76] B. Genton, I. Betuela, I. Felger, F. Al-Yaman, R.F. Anders, A. Saul, L. Rare, M. Baisor, K. Lorry, G.V. Brown, D. Pye, D.O. Irving, T.A. Smith, H.P. Beck, M. P. Alpers, A recombinant blood-stage malaria vaccine reduces *Plasmodium falciparum* density and exerts selective pressure on parasite populations in a phase 1-2b trial in Papua New Guinea, J. Infect. Dis. 185 (6) (2002) 820–827.

[77] J.D. Gabe, R.J. Chang, R. Slomiany, W.H. Andrews, M.T. McCaman, Isolation of extracytoplasmic proteins from *Serpulina hyodysenteriae* B204 and molecular cloning of the flaB1 gene encoding a 38-kilodalton flagellar protein, Infect. Immun. 63 (1) (1995) 142–148.

[78] S. Rao, G.T. Bates, C.R. Matthews, T.D. Newport, O.N. Vickery, P.J. Stansfeld, Characterizing membrane association and periplasmic transfer of bacterial lipoproteins through molecular dynamics simulations, Structure 28 (4) (2020) 475–487.e3.

[79] S. Kim, J.C. Malinverni, P. Sliz, T.J. Silhavy, S.C. Harrison, D. Kahne, Structure and function of an essential component of the outer membrane protein assembly machine, Science 317 (5840) (2007) 961–964.

[80] V. Koronakis, J. Eswaran, C. Hughes, Structure and function of TolC: the bacterial exit duct for proteins and drugs, Annu Rev. Biochem. 73 (2004) 467–489.

[81] K. Wicher, L.M. Schouls, V. Wicher, J.D. Van Embden, S.S. Nakeeb, Immunization of guinea pigs with recombinant TmpB antigen induces protection against challenge infection with *Treponema pallidum* Nichols, Infect. Immun. 59 (12) (1991) 4343–4348.

[82] V. Casas, A. Rodríguez-Asiain, R. Pinto-Llorente, S. Vadillo, M. Carrascal, J. Abian, *Brachyspira hyodysenteriae* and *B. pilosicoli* Proteins Recognized by Sera of Challenged Pigs, Front. Microbiol. 8 (2017) 723.

[83] J.C. Lazzaroni, P. Germon, M.C. Ray, A. Vianney, The Tol proteins of *Escherichia coli* and their involvement in the uptake of biomolecules and outer membrane stability, FEMS Microbiol Lett. 177 (2) (1999) 191–197.

[84] P. Goyal, P.V. Krasteva, N. Van Gerven, F. Gubellini, I. Van den Broeck, A. Troupiotis-Tsaïlaki, W. Jonckheere, G. Péhau-Arnaudet, J.S. Pinkner, M. R. Chapman, S.J. Hultgren, S. Howorka, R. Fronzes, H. Remaut, Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG, Nature 516 (7530) (2014) 250–253.

[85] R.M. La Ragione, W.A. Cooley, M.J. Woodward, The role of fimbriae and flagella in the adherence of avian strains of *Escherichia coli* O78:K80 to tissue culture cells and tracheal and gut explants, J. Med. Microbiol. 49 (4) (2000) 327–338.

[86] R.M. La Ragione, A.R. Sayers, M.J. Woodward, The role of fimbriae and flagella in the colonization, invasion and persistence of *Escherichia coli* O78:K80 in the day-old-chick model, Epidemiol. Infect. 124 (3) (2000) 351–363.

[87] R.M. La Ragione, M.J. Woodward, Virulence factors of *Escherichia coli* serotypes associated with avian colisepticaemia, Res Vet. Sci. 73 (1) (2002) 27–35.

[88] M. Hong, J. Ahn, S. Yoo, J. Hong, E. Lee, I. Yoon, J.K. Jung, H. Lee, Identification of novel immunogenic proteins in pathogenic *Haemophilus parasuis* based on genome sequence analysis, Vet. Microbiol 148 (1) (2011) 89–92.

[89] P.J. Bassford Jr., R.J. Kadner, Genetic analysis of components involved in vitamin B12 uptake in *Escherichia coli*, J. Bacteriol. 132 (3) (1977) 796–805.

[90] J.G. Mammarappallil, E.A. Elsinghorst, Epithelial cell adherence mediated by the enterotoxigenic *Escherichia coli* tia protein, Infect. Immun. 68 (12) (2000) 6595–6601.

[91] L. Wu, X.M. Lin, X.X. Peng, From proteome to genome for functional characterization of pH-dependent outer membrane proteins in *Escherichia coli*, J. Proteome Res. 8 (2) (2009) 1059–1070.

[92] H. Li, D.F. Zhang, X.M. Lin, X.X. Peng, Outer membrane proteomics of kanamycin-resistant *Escherichia coli* identified MipA as a novel antibiotic resistance-related protein, FEMS Microbiol. Lett. 362 (11) (2015).

[93] A.W. Confer, S. Ayalew, The OmpA family of proteins: roles in bacterial pathogenesis and immunity, Vet. Microbiol. 163 (3–4) (2013) 207–222.

[94] R.J. Siehnel, C. Egli, R.E. Hancock, Polyphosphate-selective porin OprO of *Pseudomonas aeruginosa:* expression, purification and sequence, Mol. Microbiol. 6 (16) (1992) 2319–2326.

[95] K.D. Krewulak, H.J. Vogel, TonB or not TonB: is that the question? Biochem. Cell Biol. 89 (2) (2011) 87–97.

[96] M.E. Konkel, B.J. Kim, V. Rivera-Amill, S.G. Garvis, Bacterial secreted proteins are required for the internaliztion of *Campylobacter jejuni* into cultured mammalian cells, Mol. Microbiol. 32 (4) (1999) 691–701.

[97] S. Langklotz, U. Baumann, F. Narberhaus, Structure and function of the bacterial AAA protease FtsH, Biochim. Biophys. Acta 1823 (1) (2012) 40–48.

[98] A. Budd, S. Blandin, E.A. Levashina, T.J. Gibson, Bacterial alpha2-macroglobulins: colonization factors acquired by horizontal gene transfer from the metazoan genome? Genome Biol. 5 (6) (2004) R38.

[99] D.R. Hyatt, A.A. ter Huurne, B.A. van der Zeijst, L.A. Joens, Reduced virulence of *Serpulina hyodysenteriae* hemolysin-negative mutants in pigs and their potential to protect pigs against challenge with a virulent strain, Infect. Immun. 62 (6) (1994) 2244–2248.

[100] T. La, N.D. Phillips, F. Coiacetto, D.J. Hampson, An atypical weakly haemolytic strain of *Brachyspira hyodysenteriae* is avirulent and can be used to protect pigs from developing swine dysentery, Vet. Res. 50 (1) (2019) 47.

[101] J.P. Bourdineaud, D. Heierli, M. Gamper, H.J. Verhoogt, A.J. Driessen, W. N. Konings, C. Lazdunski, D. Haas, Characterization of the arcD arginine:ornithine exchanger of *Pseudomonas aeruginosa*. Localization in the cytoplasmic membrane and a topological model, J. Biol. Chem. 268 (8) (1993) 5417–5424.

[102] S.B. Humphrey, T.B. Stanton, N.S. Jensen, R.L. Zuerner, Purification and characterization of VSH-1, a generalized transducing bacteriophage of *Serpulina hyodysenteriae*, J. Bacteriol. 179 (2) (1997) 323–329.

[103] M.C. Baek, S.K. Kim, D.H. Kim, B.K. Kim, E.C. Choi, Cloning and sequencing of the *Klebsiella* K-36 *astA* gene, encoding an arylsulfate sulfotransferase, Microbiol Immunol. 40 (8) (1996) 531–537.

[104] M.V. Humbert, H.L. Almonacid Mendoza, A.C. Jackson, M.C. Hung, M. K. Bielecka, J.E. Heckels, M. Christodoulides, Vaccine potential of bacterial macrophage infectivity potentiator (MIP)-like peptidyl prolyl cis/trans isomerase (PPIase) proteins, Expert Rev. Vaccin. 14 (12) (2015) 1633–1649.

[105] E. Severi, A. Müller, J.R. Potts, A. Leech, D. Williamson, K.S. Wilson, G. H. Thomas, Sialic acid mutarotation is catalyzed by the *Escherichia coli* beta-propeller protein YjhT, J. Biol. Chem. 283 (8) (2008) 4841–4849.

[106] O.P. Ernst, D.T. Lodowski, M. Elstner, P. Hegemann, L.S. Brown, H. Kandori, Microbial and animal rhodopsins: structures, functions, and molecular mechanisms, Chem. Rev. 114 (1) (2014) 126–163.

[107] L. Zeng, D. Wang, N. Hu, Q. Zhu, K. Chen, K. Dong, Y. Zhang, Y. Yao, X. Guo, Y. F. Chang, Y. Zhu, A novel pan-genome reverse vaccinology approach employing a negative-selection strategy for screening surface-exposed antigens against leptospirosis, Front. Microbiol 8 (2017) 396.

[108] T.K. Janganan, V.N. Bavro, L. Zhang, M.I. Borges-Walmsley, A.R. Walmsley, Tripartite efflux pumps: energy is required for dissociation, but not assembly or opening of the outer membrane channel of the pump, Mol. Microbiol. 88 (3) (2013) 590–602.

[109] H. Deng, C.H. Botting, J.T. Hamilton, R.J. Russell, D. O'Hagan, S-adenosyl-L-methionine:hydroxide adenosyltransferase: a SAM enzyme, Angew. Chem. Int Ed. Engl. 47 (29) (2008) 5357–5361.

[110] V. Karuppiah, R.F. Collins, A. Thistlethwaite, Y. Gao, J.P. Derrick, Structure and assembly of an inner membrane platform for initiation of type IV pilus biogenesis, Proc. Natl. Acad. Sci. USA 110 (48) (2013) E4638–E4647.

[111] E.G. Matson, M.G. Thompson, S.B. Humphrey, R.L. Zuerner, T.B. Stanton, Identification of genes of VSH-1, a prophage-like gene transfer agent of *Brachyspira hyodysenteriae*, J. Bacteriol. 187 (17) (2005) 5885–5892.

[112] A. DeRocco, H. Staats, G. Sempowski, M. Ventevogel, A. Jerse, Abstract O51. Development of MtrE, the outer membrane channel of the MtrCDE and FarAB-MtrE active efflux pump systems, as a gonorrhea vaccine. 19th International Pathogenic Neisseria Conference, Asheville, North Carolina, USA, 2014.

[113] T.D. Witchell, D.E. Hoke, D.M. Bulach, S.A. Coutts, S.J. Driesen, S.J. Cordwell, B. Adler, The major surface Vsp proteins of *Brachyspira hyodysenteriae* form antigenic protein complexes, Vet. Microbiol. 149 (1–2) (2011) 157–162.

[114] M. Gömmel, S. Barth, C. Heydel, G. Baljer, W. Herbst, Adherence of *Brachyspira hyodysenteriae* to porcine intestinal epithelial cells is inhibited by antibodies against outer membrane proteins, Curr. Microbiol. 66 (3) (2013) 286–292.

[115] D. Pérez-Pascual, T. Rochat, B. Kerouault, E. Gómez, F. Neulat-Ripoll, C. Henry, E. Quillet, J.A. Guijarro, J.F. Bernardet, E. Duchaud, More than gliding: involvement of GldD and GldG in the virulence of *Flavobacterium psychrophilum*, Front. Microbiol. 8 (2017) 2168.

[116] H.T. Tran, N. Barnich, E. Mizoguchi, Potential role of chitinases and chitin-binding proteins in host-microbial interactions during the development of intestinal inflammation, Histol. Histopathol. 26 (11) (2011) 1453–1464.

[117] A.A. Yakhnina, T.G. Bernhardt, The Tol-Pal system is required for peptidoglycan-cleaving enzymes to complete bacterial cell division, Proc. Natl. Acad. Sci. USA 117 (12) (2020) 6777–6783.

[118] J.M. Hardham, L.V. Stamm, Identification and characterization of the *Treponema pallidum* tpn50 gene, an ompA homolog, Infect. Immun. 62 (3) (1994) 1015–1025.

[119] V. Nizet, B. Beall, D.J. Bast, V. Datta, L. Kilburn, D.E. Low, J.C. De Azavedo, Genetic locus for streptolysin S production by group A streptococcus, Infect. Immun. 68 (7) (2000) 4245–4254.

[120] S. Parveen, M. Reddy, Identification of YfiH (PgeF) as a factor contributing to the maintenance of bacterial peptidoglycan composition, Mol. Microbiol. 105 (5) (2017) 705–720.

[121] T. La, N.D. Phillips, D.J. Hampson, Vaccination of chickens with the 34kDa carboxy-terminus of Bpmp72 reduces colonization with *Brachyspira pilosicoli* following experimental infection, Avian Pathol. 48 (1) (2019) 80–85.

[122] A. Movahedi, D.J. Hampson, Evaluation of recombinant *Brachyspira pilosicoli* oligopeptide-binding proteins as vaccine candidates in a mouse model of intestinal spirochaetosis, J. Med. Microbiol. 59 (Pt 3) (2010) 353–359.
[123] A. Movahedi, D.J. Hampson, Distribution of the clpX gene in *Brachyspira* species and reactivity of recombinant *Brachyspira pilosicoli* ClpX with sera from mice and humans, J. Med. Microbiol. 56 (Pt 7) (2007) 930–936.
[124] P. Zhang, N.A. Witters, G.E. Duhamel, Recovery from colonic infection elicits serum IgG antibodies to specific *Serpulina pilosicoli* outer membrane antigens (SPOMA), Adv. Exp. Med. Biol. 473 (1999) 191–197.
[125] L.J. Mappley, M.A. Tchórzewska, W.A. Cooley, M.J. Woodward, R.M. La Ragione, Lactobacilli antagonize the growth, motility, and adherence of Brachyspira pilosicoli: a potential intervention against avian intestinal spirochetosis, Appl. Environ. Microbiol. 77 (15) (2011) 5402–5411.
[126] L.J. Mappley, M.L. Black, M. AbuOun, A.C. Darby, M.J. Woodward, J. Parkhill, A. K. Turner, M.I. Bellgard, T. La, N.D. Phillips, R.M. La Ragione, D.J. Hampson, Comparative genomics of Brachyspira pilosicoli strains: genome rearrangements, reductions and correlation of genetic compliment with phenotypic diversity, BMC Genom. 13 (2012) 454.
[127] L.J. Mappley, M.A. Tchórzewska, A. Nunez, M.J. Woodward, R.M. La Ragione, Evidence for systemic spread of the potentially zoonotic intestinal spirochaete Brachyspira pilosicoli in experimentally challenged laying chickens, J. Med Microbiol 62 (Pt 2) (2013) 297–302.
[128] M.J. Woodward, L. Mappley, C. Le Roy, S.P. Claus, P. Davies, G. Thompson, R. M. La Ragione, Drinking water application of Denagard® Tiamulin for control of Brachyspira pilosicoli infection of laying poultry, Res Vet. Sci. 103 (2015) 87–95.
[129] J. Blom, J. Kreis, S. Spänig, T. Juhre, C. Bertelli, C. Ernst, A. Goesmann, EDGAR 2.0: an enhanced software platform for comparative gene content analyses, Nucleic Acids Res. 44 (W1) (2016) W22–W28.
[130] K.A. Jolley, M.C. Maiden, BIGSdb: scalable analysis of bacterial genome variation at the population level, BMC Bioinf. 11 (2010) 595.
[131] L. Berrios, B. Ely, Achieving accurate sequence and annotation data for caulobacter vibrioides CB13, Curr. Microbiol. 75 (12) (2018) 1642–1648.
[132] V. Casas, S. Vadillo, C. San Juan, M. Carrascal, J. Abian, The exposed proteomes of *Brachyspira hyodysenteriae* and *B. pilosicoli*, Front. Microbiol. 7 (2016) 1103.
[133] Y. Song, T. La, N.D. Phillips, M.I. Bellgard, D.J. Hampson, A reverse vaccinology approach to swine dysentery vaccine development, Vet. Microbiol. 137 (1–2) (2009) 111–119.
[134] M. Dalsass, A. Brozzi, D. Medini, R. Rappuoli, Comparison of open-source reverse vaccinology programs for bacterial vaccine antigen discovery, Front. Immunol. 10 (2019) 113.
[135] A. D'Mello, C.P. Ahearn, T.F. Murphy, H. Tettelin, ReVac: a reverse vaccinology computational pipeline for prioritization of prokaryotic protein vaccine candidates, BMC Genom. 20 (1) (2019) 981.
[136] T.S. Fiuza, J. Lima, G.A. de Souza, EpitoCore: mining conserved epitope vaccine candidates in the core proteome of multiple bacteria strains, Front Immunol. 11 (2020) 816.
[137] K. Naz, A. Naz, S.T. Ashraf, M. Rizwan, J. Ahmad, J. Baumbach, A. Ali, PanRV: pangenome-reverse vaccinology approach for identifications of potential vaccine candidates in microbial pangenome, BMC Bioinforma. 20 (1) (2019) 123.
[138] R.E. Sacco, D.W. Trampel, M.J. Wannemuehler, Experimental infection of C3H mice with avian, porcine, or human isolates of *Serpulina pilosicoli*, Infect. Immun. 65 (12) (1997) 5349–5353.
[139] T. Stranzl, M.V. Larsen, C. Lundegaard, M. Nielsen, NetCTLpan: pan-specific MHC class I pathway epitope predictions, Immunogenetics 62 (6) (2010) 357–368.
[140] K.K. Jensen, M. Andreatta, P. Marcatili, S. Buus, J.A. Greenbaum, Z. Yan, A. Sette, B. Peters, M. Nielsen, Improved methods for predicting peptide binding affinity to MHC class II molecules, Immunology 154 (3) (2018) 394–406.
[141] D.J. Hampson, I.D. Robertson, T. La, S.L. Oxberry, D.W. Pethick, Influences of diet and vaccination on colonisation of pigs by the intestinal spirochaete *Brachyspira (Serpulina) pilosicoli*, Vet. Microbiol. 73 (1) (2000) 75–84.
[142] A. Facciolà, G. Visalli, A. Laganà, A. Di Pietro, An overview of vaccine adjuvants: current evidence and future perspectives, Vaccinines 10 (5) (2022).
[143] J.L. Turley, E.C. Lavelle, Resolving adjuvant mode of action to enhance vaccine efficacy, Curr. Opin. Immunol. 77 (2022), 102229.