Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Iron Uptake in Pasteurellaceae

Zachary Phillips

mrzacharyphillips@outlook.com

Griffith University https://orcid.org/0000-0002-9507-1295

Research Article

Keywords: Pasteurellaceae, iron uptake, phase-variation, metabolism, nomenclature, Actinobacillus pleuropneumoniae, Aggregatibacter actinomycetemcomitans, Avibacterium paragallinarum, Gallibacterium anatis, Glaesserella parasuis, Haemophilus ducreyi, Haemophilus haemolyticus, Haemophilus influenzae, Haemophilus parainfluenzae, Histophilus somni, Mannheimia haemolytica, Pasteurella multocida

Posted Date: October 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10023473/v3

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Iron Uptake in Pasteurellaceae

Author: Zachary N Phillips
Author Affiliation: none
Contact: mrzacharyphillips@outlook.com

Abstract

Iron is essential to *Pasteurellaceae*, a bacterial family that includes opportunistic human commensals and numerous human, zoonotic and veterinary pathogens of medical and economic significance. This analysis examines 12 *Pasteurellaceae*, including *Actinobacillus pleuropneumoniae*, *Aggregatibacter actinomycetemcomitans*, *Avibacterium paragallinarum*, *Gallibacterium anatis*, *Glaesserella parasuis*, *Haemophilus ducreyi*, *Haemophilus haemolyticus*, *Haemophilus influenzae*, *Haemophilus parainfluenzae*, *Histophilus somni*, *Mannheimia haemolytica* and *Pasteurella multocida*. Fully annotated genomes were surveyed for outer membrane iron uptake proteins. Novel proteins and previously overlooked alleles within protein families were catalogued. A total of 44 outer membrane iron uptake protein families were identified and organised under a unified nomenclature system. The findings reveal the dynamic and diverse iron acquisition repertoire of *Pasteurellaceae*. This analysis substantially expands current knowledge of iron uptake factors in *Pasteurellaceae* and provides a standardised nomenclature system for these factors.

Keywords: Pasteurellaceae, iron uptake, phase-variation, metabolism, nomenclature

Introduction

To our knowledge, no recent study has comprehensively examined iron uptake across a bacterial family. Prior studies and reviews are rarely exhaustive. Factors such as genetic conservation, multiple copies of a gene, antigenic/phase-variation and trends across genus are rarely evaluated. This study focuses on outer membrane (OM) iron uptake proteins from clinically and economically important members of the *Pasteurellaceae* family. The members of the *Pasteurellaceae* family under examination, with accompanying material reviewing their importance, are: *Actinobacillus pleuropneumoniae* [1], *Aggregatibacter actinomycetemcomitans* [2], *Avibacterium paragallinarum* [3, 4], *Gallibacterium anatis* [5], *Glaesserella parasuis* [6], *Haemophilus ducreyi* [7], *Haemophilus haemolyticus* [8], *Haemophilus influenzae* [9, 10], *Haemophilus parainfluenzae [11]*, *Histophilus somni* [12], *Mannheimia haemolytica* [13, 14] and *Pasteurella multocida* [15, 16]. Among these are significant human pathogens, opportunistic human commensals, zoonotic pathogens and economically important veterinary pathogens. Despite the members of the *Pasteurellaceae* family inhabiting a broad range of hosts, nearly all are associated with respiratory tract colonisation or infection.

The effectiveness of antibiotics is waning faster than new therapies are being developed. Few vaccines are long lasting and provide protection against all alleles of a target factor. Bacteria may adapt to vaccine targets through means such as antigenic variation. Biological iron and its uptake are ubiquitously necessary for growth and survival of *Pasteurellaceae*. OM iron uptake factors have long been considered in vaccine/treatment development but faced issues such as the diversity and redundancy of bacterial iron uptake systems, antigenic- and phase-variation, and lack of data availability. More recent efforts to broadly target iron uptake have provided encouraging results [17]. Modern genome collections such as the NCBI database contain an unprecedented and expanding trove of data, and analysis tools have rapidly improved over the past decade. With these improvements, targeting iron uptake therapeutically may finally be achievable, should a ‘complete picture’ of the factors involved be revealed.

Iron in the Host

Biological iron is reactive and potentially damaging to the host. To protect the host, iron is stored in a lower reactive state in a carrier molecule. These include heme [18], ferritin [19], haemoglobin [20], haptoglobin [21], hemopexin [22], myoglobin [23], lactoferrin [24] and transferrin [25]. Competition for iron between bacteria and host is an ongoing promoter of evolution [26]. Iron acquisition is a prominent target for the humoral immune response [27, 28]. Heme is typically the most abundant iron source available to bacteria in an animal host [18]. Pathogenic members of the *Pasteurellaceae* family commonly have multiple redundant iron uptake systems, which can undergo variable expression [29].

Bacterial Iron Uptake Pathways

Iron is not passively acquired by *Pasteurellaceae* and is actively taken up via OM proteins. Infectious or commensal bacteria target host iron carrier molecules to source biological iron. Iron uptake in *Pasteurellaceae* typically involves an OM bound 22-stranded β-barrel [30]. This is true for many of the iron uptake OM proteins of *Pasteurellaceae* [31]. Multiple lineages of these proteins in the *Pasteurellaceae* family have been studied to varying degrees. Previous literature covering relevant β-barrels accompanies their sub-section as part of the **Protein Analysis**. Siderophores are secreted iron-chelating molecules that bind to iron and then associate with a receptor, such as an OM 22-stranded β-barrel [32, 33]. Similarly, hemophores are secreted or membrane associated proteins which bind heme, then associate with an OM 22-stranded β-barrel [34]. These receptors can function independently while also specifically

interacting with either siderophore or hemophore-bound iron. Secreting siderophores and hemophores provides bacteria a method of gathering iron out of the range of the OM iron uptake proteins.

Variable Gene Expression

There are multiple mechanisms that can impact expression of OM iron uptake proteins. This study focuses on those that most impact vaccine development and host immune efforts – phase-variation, gene copy number and allelic diversification. Fine-tuning mechanisms such as Fur regulation [35], methylation [36] and graded gene expression have not been evaluated in this study.

ON/OFF Phase-variable Gene Expression

Genes can phase-vary through frameshift changes in their open reading frame. This happens during replication, facilitated by a genomic region with repeated, or highly similar sequences. These sequences can be a single nucleotide **(see Figure 1 a)**, or more complicated, such as a tetranucleotide repeat sequence **(see Figure 1 b).** Phase-variation provides contingencies in a changing environment. Expressing a gene isn't always beneficial due to the metabolic cost of doing so. Phase-variable expression of OM iron uptake proteins offers bacteria extra contingencies on top of other factors such as Fur regulation and antigenic variation [18]. Gram negative surface factors, such as OM iron uptake proteins, are particularly prone to **ON**/**OFF** phase-variable expression [37].

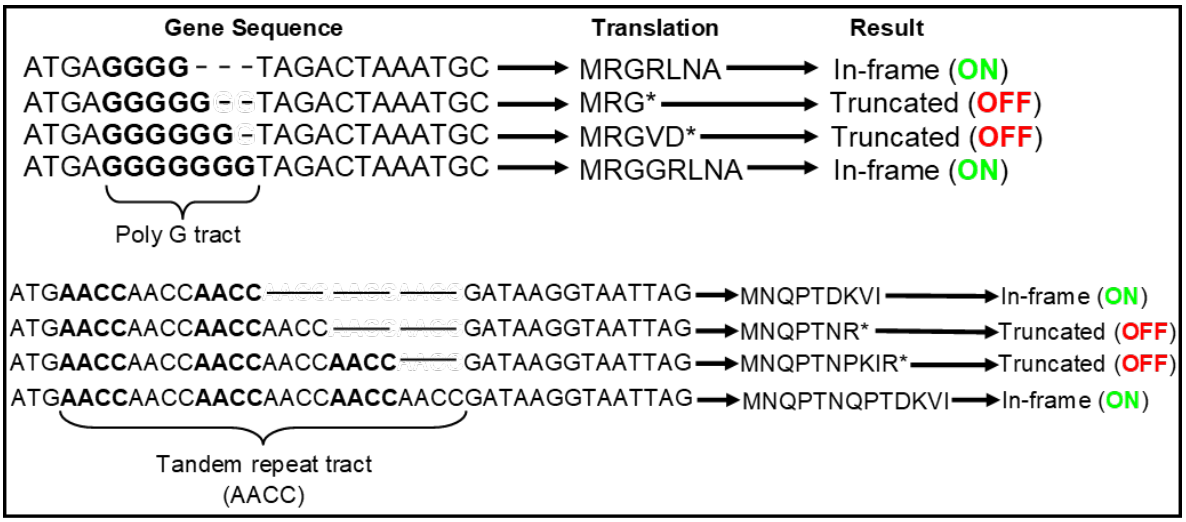


Figure 1: Phase-variable gene expression through a polynucleotide (in this case, G) tract and a tetra-nucleotide (in this case, AACC) repeat tract. Slipped strand mispairing during bacterial replication causes repeat regions to be gained or lost. The gene turns **ON**/**OFF** as the number of repeats expands or contracts. This results in a frameshift to the open reading frame downstream of the change, turning the gene **ON** or **OFF**. * Indicates a premature stop codon in the protein sequence.

Multiple Copies of a Gene

Gene copy number is a frequently overlooked aspect of bacterial genetics. Multiple alleles of a gene may be misinterpreted as a simple gene duplication event. The presence of multiple similar (but not identical) gene copies is likely underreported, and, despite the availability of modern genomic data, cases are rarely investigated. Gene redundancy can provide functional resilience, where one copy may acquire deleterious mutations while an alternative copy retains

functionality. It may also increase overall transcriptional output, leading to greater potential protein production.

Within this study, three broad forms of a gene "copy" are recognised. The first is produced via a duplication event, where an **identical copy** of a gene can be found more than once in the genome. The second is when a **non-identical copy** of a gene is present i.e. when two genes share high (>80%) sequence homology but aren't identical. This may reflect a beneficial gene duplication event generations prior. The final instance is where **multiple genes of the same family** are observed within a single genome. These proteins likely share a function, though are distinct.

Historically, gene copies have been un- or under- reported. This may be due to the reliance of methods such as PCR, which can overlook multiple copies of similar genes. Genes of Hgps in *H. influenzae* have all three types of gene copies, described previously [29]. This can also be observed in *P. multocida* with its multiple available Hgp alleles. It's likely that the latter two types of gene copies (multiple nonidentical copies and multiple alleles) provide a contingency against host immune efforts. Further, some gene copies are both ON/OFF phase-variably expressed and immunologically distinct. As all three instances of gene copies impact vaccine design, they have been assessed within this study.

A Family Approach

Previous approaches have overlooked many iron uptake factors and failed to address nomenclature issues. This study provides a broader evaluation of OM iron uptake proteins available to *Pasteurellaceae*. A wider view is crucial, as members of *Pasteurellaceae* often co-occupy the same host species at the same anatomical site. For example, *A. pleuropneumoniae* and *G. parasuis* (pigs), *H. somni* and *M. haemolytica* (cows), *A. actinomycetemcomitans* and *H. influenzae* (humans). Furthermore, some members of the *Pasteurellaceae* family can colonise multiple different host species, such as *P. multocida* (cows, pigs, birds, humans, cats, dogs, *etc*), and can move virulence factors between species. Sharing one site in the same host provides members of the *Pasteurellaceae* family an opportunity for genetic exchange of virulence factors, such as iron uptake genes. This has been observed in the exchange of an antibiotic resistance plasmid between *A. pleuropneumoniae* and *G. parasuis* [38]. *Pasteurellaceae* are naturally competent [39], and may share anatomical sites enabling genetic exchange across genera and species. A broader 'family approach' allows these observations to be interpreted in a more complete evolutionary and functional context. This is also important in the design of vaccine/treatment targets. For example, *H. somni* and *M. haemolytica* both infect cattle. An effective vaccine target may target a single essential pathway across multiple organisms.

Methods

Iron Uptake Genes

OM Iron uptake factors were identified through a combination of literature curation, homology-based searches and manual genome inspection. The analysis included outer-membrane proteins and secreted factors associated with iron acquisition. Proteins were included where their role in iron acquisition had been described in published studies or where their involvement was supported by sequence similarity to characterised proteins, conserved domain structure or genomic context. Inclusion did not necessarily imply that the function of a protein had been experimentally confirmed in every species examined. This distinction is discussed for individual proteins in the Protein Analysis section.

Published studies were surveyed to compile previously reported iron-acquisition systems, with particular emphasis on outer-membrane-associated receptors. A summary of the genes and proteins identified in the literature for each organism is provided in the corresponding Previous Literature sections.

Genomes were manually inspected for genes commonly associated with iron-acquisition systems, including ABC transporters, TonB-dependent energy transduction systems and genes involved in heme metabolism (e.g. *hemA*, *recA*, *tonB*, *exbB* and *exbD*). These components were used as indicators to identify potential iron-acquisition systems in the surrounding genomic regions. Outer-membrane factors located near these factors were then investigated.

Homology searches were conducted using the NCBI BLAST web interface with default search parameters for BLASTn and tBLASTn. Searches were conducted one species at a time, with the search restricted to genomes from species for which fully annotated genomes were available in NCBI. Previously characterised iron uptake proteins identified in the literature were used as query sequences to identify homologous sequences in the selected genomes. Retrieved sequences were manually collated and compared by multiple sequence alignment using Clustal Omega. Sequence relatedness, conserved domain structure and genomic context were considered when assigning proteins to families and sub-alleles.

Although published evidence supports a role in iron uptake for nearly all proteins included in this study, equivalent functional evidence was not available for every species examined. For example, the *hxuABC* locus has been experimentally characterised as an iron uptake system in *H. influenzae*, but orthologues in *A. paragallinarum* and *M. haemolytica* lack equivalent functional validation. Proteins without direct functional evidence in the species examined were therefore treated as putative iron uptake factors rather than experimentally confirmed factors.

Truncated sequences were identified during translation of nucleotide sequences retrieved from BLAST searches. Potential explanations for apparent truncations, including phase-variable regions containing tandem or simple sequence repeats and alternative start codons, were also investigated. Predicted protein structures were examined as an additional check of protein integrity. In this study, descriptions such as “potentially functional” and “likely functional” refer to the potential to encode an intact protein and do not necessarily indicate an experimentally demonstrable biological activity.

Proteins investigated in this study were also compared by sequence alignment with representative β-barrels involved in the uptake of non-iron nutrients, including SusC and MalA, and with YncD, a TonB-dependent receptor associated with pyrroloquinoline quinone uptake.

No significant homology was detected between these proteins and the iron uptake factors identified in this study.

Functional predictions of outer-membrane proteins were further assessed using the online SIFTER server [40] using UniProt entries as an input (**Supplementary File 15**). SIFTER scores in the range of 0.4–0.6 were observed for known heme-binding proteins. A score threshold of 0.4 was therefore used as an indicator of potential involvement in iron or heme acquisition. SIFTER predictions were considered alongside sequence similarity, conserved domain structure and genomic context, rather than as independent experimental confirmation of protein function.

Organism Selection

This study utilised fully annotated genomes obtained from the NCBI database. Species were selected for inclusion if at least five fully annotated genomes were available. Fully annotated genomes were selected to facilitate the identification of phase-variable regions and improve the resolution of genomic variability. The number of genomes analysed for each species is provided in **Table 1**. Highly similar clonal or duplicate genomes were removed before comparative analysis to reduce redundancy in the dataset.

The number of available genomes varied substantially between species at the time of analysis, reflecting differences in sequencing efforts. Notably, well-studied human pathogens such as *H. influenzae* had more suitable genomes (approximately 150), whereas less extensively studied species such as *A. actinomycetemcomitans* had fewer (n = 26).

Nomenclature of Iron Uptake Genes and Proteins across *Pasteurellaceae*

Due to inconsistencies in historical gene and protein naming across *Pasteurellaceae*, a standardised nomenclature system was applied based on previously proposed guidelines for bacterial gene naming [41]. Inconsistent nomenclature obscures relationships between proteins and complicates comparative analysis. For example, Haemoglobin-haptoglobin binding proteins in *H. influenzae, H. somni, A. pleuropneumoniae and A. actinomycetemcomitans* are named Hgps [42-45], whereas in *P. multocida* they are named Hg**b**s [46] and the corresponding orthologue in *H. ducreyi* is named HupA [47]. This is not to be confused with HupA, a DNA-binding protein in *P. multocida*, or Hup, the well documented iron uptake outer membrane protein found in many members of *Pasteurellaceae* (**Table 2**). These proteins share ~50% Amino Acid Identity (AAID), and are different alleles, except in the case of *H. influenzae* and *A. actinomycetemcomitans*, for which HgpA ***is*** essentially the same protein. We propose unifying nomenclature across all organisms.

To address the naming issues, existing gene and protein names were reviewed to establish a consensus of nomenclature. Proteins were sorted into protein families that kept their consensus abbreviation/prefix (e.g. Hgp). Arbitrary suffixes were identified and reassigned with a number e.g. changing HgpA to Hgp1. The suffix was assigned based on order of publication date (first published = Hgp1). The earliest published suffix designation was prioritised, when appropriate. Where no *Pasteurellaceae*-specific name existed, orthologues were assigned based on established homologues from other organisms, such as *E. coli*. In cases where a gene was part of a larger locus, such as the *hxuABC* locus, the gene retained the suffix as it provides positional data as to the locus order. This differs from HgpA, in which the suffix ‘A’ merely indicated that it was the first instance of this protein family identified in that organism.

Protein families were defined according to the evolutionary relationships between proteins, informed by sequence similarity and structural conservation. Genomic context and conserved domain structure were also considered when characterising proteins and assessing their likely functions. For example, β-barrels with a near-identical backbone grouped into a single 'protein family.' Most 22-stranded TonB-dependent β-barrels within a family share >40% AAID. Sub-alleles tend to share >60% AAID. Another example, the Omp family (except OmpW), all share >30% AAID and group phylogenetically. Hup alleles share >30% AAID and group phylogenetically. These thresholds were not consistently applicable to all protein families. For example, a consistent allele threshold could not be established for hTbps. However, in all cases sequences assigned as the same allele have >80% AAID. For example, under the proposed system, all Hgp1 proteins share >60% AAID, whereas Hgp1 and Hgp2 share approximately 50% AAID. Hgp1 was further divided into distinct sub-alleles, which shared ~70% (>60% but <80%) AAID. All Hgps with the same allele suffix had >80% AAID. Following the system proposed by *Demerec et al*. [41], decimal suffixes were added to distinguish these sub-alleles: Hgp1.1, Hgp1.2 and Hgp1.3. **Supplementary File 16** contains representative amino acid sequences of all catalogued proteins, which can be used to generate a phylogenetic tree and a percent identity matrix for further analysis.

This unified nomenclature was applied across all iron uptake proteins identified in this study to enable consistent comparison and functional analysis. Many factors in this study lacked prior literature, had only been assigned an accession number, or had been named something misleading.

Structural Prediction

The AlphaFold 3 online server was used for protein structural prediction [48]. PDB structures were visualised with images exported via the PyMOL Molecular Graphics System, Version 3.0 (Schrödinger, LLC).

Results and Discussion

Genomes across the 12 *Pasteurellaceae* species shown in **Table 1** (*below*) were examined. Iron uptake genes were widely distributed in most genomes. *H. ducreyi* has the smallest average genome of organisms examined and has the fewest OM iron uptake proteins observed. However, this was not an observable trend across the other members of *Pasteurellaceae* examined. For example, the organism with the largest average genome size, *M. haemolytica* (2.7Mb), had 26 iron uptake proteins, and the organism with the second-lowest average genome size, *H. influenzae* (1.9Mb), had 34 iron uptake proteins. The ability to colonise multiple hosts or association with pathogenicity also did not correlate with an increased number of iron uptake proteins. For example, *P. multocida*, which can infect a range of hosts, had 28 iron uptake proteins. The human commensal bacterium *H. haemolyticus* had 37. Further, anatomical niche and microbial competition appear more important than host identity. We suggest host niche (including anatomical site, local microbiome diversity and heme sources) exerts more selective pressures to expand or contract the iron uptake protein repertoires vs factors such as genome size, genetic lineage or host range. These observations do not establish a direct relationship between anatomical niche and iron uptake protein composition. The close phylogenetic relationships among the species examined, and their host-associated niches, make it difficult to distinguish niche-associated selection from shared evolutionary history. Niche complexity is therefore proposed as a possible contributor to variation in iron uptake protein repertoires, rather than a demonstrated explanation for the distribution of individual iron uptake protein families. Comparative analyses across more bacterial families occupying similar anatomical niches may help clarify the contributions of niche-associated selection and shared lineage to OM iron uptake protein repertoire diversity.

Name	Genome Size	Strains Examined	Primary Host	IROMP Genes
A. pleuropneumoniae	~2.35Mb	32	Pigs	30
A. actinomycetemcomitans	~2.10Mb	26	Humans	13
A. paragallinarum	~2.55Mb	11	Birds	16
G. anatis	~2.60Mb	9	Birds	24
G. parasuis	~2.25Mb	28	Pigs	22
H. ducreyi	~1.65Mb	8	Humans	3
H. haemolyticus	~1.90Mb	5	Humans	37
H. influenzae	~1.90Mb	109	Humans	34
H. parainfluenzae	~2.00Mb	18	Humans	38
H. somni	~2.10Mb	24	Cows	14
M. haemolytica	~2.70Mb	99	Cows	26
P. multocida	~2.30Mb	108	Multiple	28

Table 1: Organism Information. Iron uptake proteins were detected in listed *Pasteurellaceae*. The sum of these genes includes alleles and sub-alleles detected in this study. The number of strains examined varied from 5 to 109. Clonal strains/genomes were excluded from this study. Average genome size is in mega base-pairs (Mb).

Gene	*A. pleu*	*A. act*	*A. para*	*G. ana*	*G. para*	*H. duc*	*H. hae*	*H. inf*	*H. para*	*H. som*	*M. hae*	*P. mul*
curA							4		1		2	
fcmA	3	1			1		*T		1	1	1	
fecA		1					1		1			1
fepA		1							3			
fhuE					*T						1	
fhuH							1	1				
fhuI	3		2	1	1				1			
foxA											1	
fupA	1		1	1	*T				1			
gasA				1								
hasR		1										1
hex			1				1	1				1
fre	1			1	1		1		1		2	
hemR						1		1		1		1
hgp	1	1				1	4	11		1		5
hhr			1				2	1	1			2
hmbR	*T		1						4		2	1
hup		1			1		1	1	1	2	1	1
hxuA					1			1			1	
hxuB			1	1	1			1			1	
hxuC			1	1	1			1			1	
hxuD			1	1								
ireA							1					
irpA	1				2					1	2	3
iup			1									
miu											1	
nbrA		1	1					1		1		2
omhr			1									1
ompM	1			4	1		2	1	1		1	
ompO									2			
ompU	1	1		2	1		2	2	2	1	1	1
ompW	1	1	1	1	1					1	1	1
pE		1					1	1	1			1
pfhR												1
pmuA		1			1		2	2	1			
rup							1					
squA			1				1					
squB			1				1					
tbpA	2	1			1		1	1		1	1	
tbpB/htbpB	13	1	1	10	6		8	7	15	2	6	4
tbpC										1		1
tbpD										1		
tdX	1					1						
viuA									1			

Table 2: Gene Presence. This is the distribution of the 44 OM or secreted iron uptake protein families observed in examined *Pasteurellaceae*. The names of the organisms have been shortened: *A. pleuropneumoniae* = *A. pleu*, *A. actinomycetemcomitans* = *A. act*, *A. paragallinarum* = *A. para*, *G. anatis* = *G. ana*, *G. parasuis* = *G. para*, *H. ducreyi* = *H. duc*, *H. haemolyticus* = *H. hae*, *H. influenzae* = *H. inf*, *H. parainfluenzae* = *H. para*, *H. somni* = *H. som*, *M. haemolytica* = M. hae and *P. multocida* = *P. mul.* This table shows if *any* of the genomes surveyed contained the corresponding gene. For a statistical breakdown of results see **Supplementary File 13** and **14**. The number of distinct alleles are represented by the numbers in the table. e.*g*. *H. haemolyticus* genomes contained four (4) distinct *curA* alleles across isolates. Sub-alleles have not been included in the tally. *T indicates a gene was present but not functional (truncated). hTbpB and TbpB proteins have been collapsed into one group which includes all structurally similar proteins like Hemophilin, OmlA, PlpE and Fishing

Rods. Genes that have been highlighted blue encode OM β-barrels, orange encode secreted factors, and grey export machinery.

Factors Impacting Immunity

We examined key factors that may impact vaccine and therapy design. These include Phase-variable expression, the migration of iron uptake factors and presence of multiple genes/alleles in organisms.

Phase-variable (ON/OFF) Expression

Organism	*fepA*	*fcmA*	*foxA*	*hhr*	*hemR*	*hgp*	*hmbR*	*hup*	*hxuD*	*irpA*	*omhr*	*ompM*	*nbrA*	*tbpA*	*tbpC*	*tbpD*	*h/tbpB*
A. pleuropneumoniae												Poly A					Poly C
A. actinomycetemcomitans	Poly G													Poly A			Poly A
A. paragallinarum							Poly T		Poly A		Poly C		Poly C				
G. anatis																	Poly G
G. parasuis										Tandem							Poly A
H. ducreyi																	
H. haemolyticus																	
H. influenzae						Tandem							Tandem				
H. parainfluenzae				Poly A													
H. somni		Poly G			Poly C	Poly C		Poly G		Poly G			Poly G		Poly G	Poly G	Poly C
M.haemolytica			Poly A				Poly T			Poly A							Poly T
P. multocida				Poly A		Poly A							Poly G				Poly C

Table 3: ON/OFF phase-variable iron uptake factors across *Pasteurellaceae*. Green cells indicate at least one gene is present in genomes surveyed. Red cells indicate a gene is present but truncated. Orange cells show a gene is present and ON/OFF phase-variable. Alternate forms of variable expression (e.g. periscope proteins) have not been included. The mode genes phase-vary has been indicated - via polynucleotide tract ('poly _') or via tandem repeats ('tandem'). As this table is highly reductive, more information can be found in the respective gene/protein section (below) and in the **Supplementary Material**.

Phase-variable expression was observed in iron uptake factors across the examined *Pasteurellaceae*, except for *H. ducreyi* and *H. haemolyticus*. In the case of *H. haemolyticus*, the absence of observed phase-variable genes may reflect the small number of genomes surveyed (n = 5). However, periscope proteins in the form of Fishing Rods were identified in *H. haemolyticus*. *H. somni* appeared to have a particularly high frequency of ON/OFF phase-variable genes, with 9 of 14 unique iron uptake outer membrane genes showing evidence of phase-variation. It's possible that the proliferation of phase-variable iron uptake factors in *H. somni* accommodates a reduced iron uptake repertoire. Likewise, *H. haemolyticus*, which had a large repertoire (37) has no phase-variable iron uptake genes.

The genetic mechanisms allowing ON/OFF phase-variable expression are not conserved among similar genes in different organisms. For example, *H. influenzae hgps* phase-vary via tandem repeats, whereas *H. somni hgps* phase-vary via poly-C tracts. Similarly, *irpA* phase-varies via tandem repeats in *G. parasuis* but via poly-G or poly-A tracts in *H. somni* and *M. haemolytica*. The capacity for phase-variation therefore does not appear to be linked to a specific gene, but rather to genetic features already present within the host genome. For example, the *hgp8* allele is present in both *H. influenzae* and *P. multocida* isolates. Despite the genes sharing near-identical nucleotide sequences, only the *H. influenzae hgp8* has a tandem repeat tract. The earliest *P. multocida* isolate carrying *hgp8* was collected in 1947, whereas the *H. influenzae* isolates carrying *hgp8* were collected in 2016. The high sequence similarity is consistent with a shared evolutionary origin and may reflect horizontal gene transfer, while the difference in repeat structure suggests that the capacity for phase-variation may have emerged or been lost independently. However, the direction and timing of a transfer, and the evolutionary origin of the repeat tract, cannot be established from these observations alone.

Further support for this observation can be seen when comparing *hgp* genes in *H. haemolyticus* and *H. influenzae*. Although these organisms share four near-identical *hgp* alleles, only the *H. influenzae* genes appear to be phase-variable. Importantly, this is not limited to *hgp* genes and was observed for multiple phase-variable genes examined in this study as seen in **Table 3** (e.g. *tbpC*). These observations indicate genetic elements allowing ON/OFF phase-variable expression may be independent mobile genetic elements.

Gene Transfer

The distribution of *hgp1.3* and the hemophilin locus raises the possibility of genetic exchange among *Pasteurellaceae*. The *hgp1.3* allele was present in all examined *A. actinomycetemcomitans* genomes and one *H. influenzae* genome (HiB). Both organisms inhabit the human respiratory tract, providing a possible opportunity for genetic exchange, although the observed distribution does not establish direct transfer between these species. Similarly, the hemophilin locus was identified in two confirmed *H. influenzae* isolates, despite being traditionally associated with *H. haemolyticus*. These distributions warrant further investigation into the evolutionary relationships and possible transfer of iron-acquisition genes. Additional examples are provided in the Supplementary Material.

Evidence of transfer was also observed between organisms that do not typically occupy the same host/niche. *P. multocida* is associated with human infection resulting from pet bites/scratches [16] and rarely occupies the human respiratory tract where *H. influenzae* is found [16, 49]. The *hgp6* gene was common in *P. multocida* but uncommon in *H. influenzae*, where it was detected across typeable, *aegyptius* and Non-typeable *Haemophilus influenzae* (NTHi) isolates. Its distribution across these groups is consistent with a gene that is not restricted to a single *H. influenzae* lineage. Together with its distribution in *P. multocida*, this raises the possibility of horizontal transfer or circulation through a broader gene pool, although additional evolutionary analysis would be required for confirmation.

Gene transfer may also occur across bacterial families. The *Vibrio cholerae* iron uptake gene *viuA* was identified in 22% of *H. parainfluenzae* genomes and is also present in other *Pasteurellaceae* not examined in this study (e.g. *A. segnis*, *L. koalarum*). Other examples are *ireA* and *rup* genes found in *H. haemolyticus*. These two genes share a locus in *H. haemolyticus* but may have originated from different organisms. The *ireA* gene may have come from *E. coli* where it is commonly observed [158, 159]. Notably, the *ireA* gene was observed in *E. coli* plasmids (data not shown). The *rup* gene is not present in *E. coli*, and instead likely came from *N. meningitidis* where the gene is more abundant (though not previously identified).

Collectively, these observations are consistent with the movement of genes across bacterial species, genera and families, although the methods of transfer remain unresolved. The observed variation in gene distribution supports the concept of a dynamic accessory antigen repertoire rather than a uniformly conserved set of iron uptake factors across *Pasteurellaceae*. If antigen distribution is fluid, a broad antigen repertoire may be required for effective long-term vaccine and therapeutic strategies.

Gene Copies and Alleles

Gene presence alone does not equate to gene expression, and gene expression does not necessarily correspond to protein production. An additional and underappreciated factor in bacterial genetics is the presence of multiple related genes within individual genomes. Gene duplication and allelic diversification are frequently underreported or excluded from

comparative analyses, potentially due to reductive approaches to detection (such as PCR). However, tracking both duplicate genes and allelic diversity is essential for rational vaccine design and for understanding the functional repertoire of these organisms.

We sought to standardise nomenclature across iron acquisition factors while explicitly retaining duplicate genes, including both identical and non-identical copies. This is a fundamental step in accurately defining virulence-associated iron uptake repertoires within *Pasteurellaceae*.

Multiple copies or alleles of FepA, FcmA, Hgp, Hup, IrpA, OmpM/O/U, TbpA and numerous h/TbpBs were identified in some genomes. These findings demonstrate that individual genomes can encode multiple related iron acquisition factors. Such redundancy may contribute to differences in iron acquisition or antigenic diversity, although the impacts on immune evasion, host adaptation and niche specialisation remain unknown. Three examples of apparent gene duplication were identified: *hgp2* in NTHi strain NCTC12699, *tbpC* in *P. multocida* strain HB01 and *pfhR2* in *P. multocida* strain KVNON-213. The duplicated copies were identical, with no sequence variation between copies. The consequences of these duplications for protein function or bacterial fitness remain unknown.

Allelic diversity had only been partially characterised for a small subset of proteins within this study, and usually only within individual species. We attempted to define distinct alleles using a consistent and sensible nomenclature system. This was necessary as existing nomenclature had become inconsistent and difficult to interpret. An example highlighted in the **Methods** section is that the first Hgp identified in an organism has sometimes been published as “HgpA.” Some proteins designated 'HgpA' share >80% Amino Acid Identity (AAID), whereas others share only ~50% AAID and represent distinct alleles. ‘Hg**b**A’ of *P. multocida* and HupA of *H. ducreyi* are allelic variants of Hgps. The new system makes it clear these proteins are in one protein family and clarifies if they’re the ‘same’ protein or a separate allele. This nomenclature system can accommodate new alleles as they’re discovered. **Table 4** (below) shows proteins with multiple alleles.

This analysis and nomenclature clarification showed extensive allelic diversity across *Pasteurellaceae*. It also revealed notable similarities between organisms as discussed in the **Gene Transfer** section (above). These findings inform the genetic diversity of iron-acquisition proteins, which may inform future studies of virulence, host adaptation and vaccine antigen selection.

Protein	CurA	FcmA	FecA	FepA	FhuE	FhuI	Fup	HasR	Hex	Fre	HemR	Hgp	Hhr	HmbR	Hup	HxuA	HxuB	HxuC	IrpA	Nbr	OMHR	OmpM	OmpO	OmpU	OmpW	PE	PmuA	TbpA	TdX
Alleles	4	7	2	3	2	6	4	2	3	6	3	14	4	8	6	3	4	4	8	4	2	11	2	8	4	4	3	6	2

Table 4: Proteins with multiple alleles within *Pasteurellaceae* examined. This tally does not include sub-alleles. Many of the proteins were observed in more than one organism. Many (but not all) organisms had unique alleles of these shared proteins. hTbpB/TbpBs (including hemophilin) have not been included due to the complexity of this broad group of proteins.

Genomic Rearrangement

Loci were observed to have undergone genomic rearrangement. This type of event produced *tbpC* of *P. multocida* [50] and *H. somni* [51], which is a merged form of *tbpA/B* genes. Importantly, TbpC can function as a standalone protein in place of the separate TbpA/B proteins.

One *fecA1* gene in *P. multocida* strain HN01 may have undergone a genomic rearrangement event that merged the existing *fecA1* with the adjacent 3′ end of an upstream SLAM of unknown function. The SLAM usually appears to be intertwined with a beta-propeller. A transposase gene is located immediately adjacent to *P. multocida fecA1* genes. Transposases are associated with genetic rearrangement events [52]. Whether this gene produces a functional protein remains unclear. Further, *P. multocida* strain Pm1 had a unique *pfhR2* gene. This *pfhR2* gene had a 5′ end which appears to have been merged with the adjacent upstream ABC transporter ATP-binding cassette. It is unknown if this new merged protein preserves the function of both individual genes.

Limitations

Several limitations should be considered in this study. The number of genomes available varied substantially between species, affecting the detection of rare genes and alleles. Clonal, duplicate genomes were removed before comparative analysis to reduce redundancy in the dataset and limit overrepresentation of closely related isolates. However, this does not eliminate sampling bias or account for phylogenetic relationships among the remaining genomes. Differences in geographic, temporal, clinical and host sampling may also influence the observed distributions of iron uptake genes. Consequently, the repertoires described here represent those observed in the genomes examined and may not capture the full diversity within each species. Apparent associations between gene distributions and host or niche should therefore be interpreted cautiously, as they may reflect sampling patterns or shared evolutionary history rather than host adaptation.

Iron Uptake Repertoires

We suggest *Pasteurellaceae* iron acquisition systems utilise a dynamic accessory antigen repertoire. Horizontal gene transfer, gene duplication, phase-variation and allelic diversification continuously reshape the available iron uptake pathways. Rather than evolving entirely new systems, *Pasteurellaceae* tend to repurpose and redistribute existing iron acquisition genes across species, genera and even bacterial families. This dynamic has important implications for host adaptation, competition during co-colonisation and in the design of vaccines and therapeutics. The data herein suggests broadly targeting iron uptake pathways may prove more effective than targeting individual antigens.

This is the first examination of these factors at the bacterial family level. Although this study was limited to 12 members of *Pasteurellaceae*, it is the largest and most thorough study of these factors to date. A visual guide to the outer membrane/secreted proteins of each organism has been included below, along with primary findings/discussion. Included in this information is the rate at which genes were detected in organisms, and the rate at which genes were predicted to be functional.

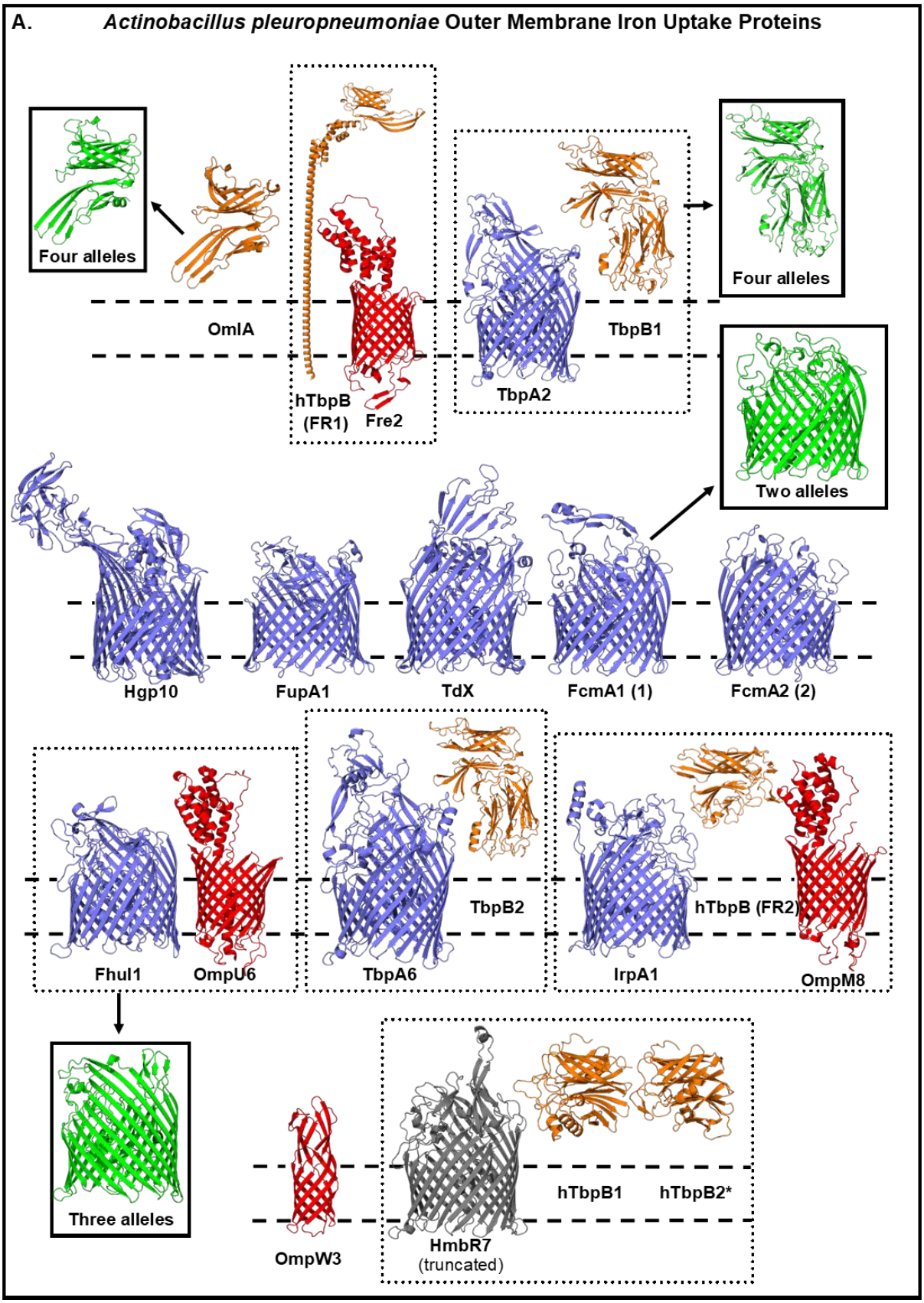


Figure 2: *A. pleuropneumoniae* iron uptake proteins. The rate (%) at which genes were observed in organisms as well as the rate these genes were predicted to be functional: ***omlA*** 100%; 100%, ***fre2*** 97%; 47%, ***hTbpB*** **(Fr1)** 100%; 25%, ***tbpA2*** **(1)** 100%; 100%, ***tbpB*** **(1)** 100%; 100%, ***hgp10*** 100%; 97%, ***fupA1*** 94%; 40%, ***tdX*****:** 100%; 97%, ***fcmA1*** 100%; 97%,

fcmA2* (2)** 100%:100%, ***fhuI1 100%;94%, ***ompU6*** 100%: 97%, ***tbpA6* (2)** 100%; 47%, ***tbpB* (2)** 100%; 78%, ***irpA1*** 100%; 100%, ***hTbpB* (FR2):** 100%; 100%, ***ompM8*** 100%: 100%, ***ompW3*** 100%; 100%, ***hmbR7*** 100%; 0%, ***hTbpB* (1)** 100%; 97%, **hTbpB (2)*** 100%; 97%. *Indicates a gene was prone to ON/OFF phase-variable expression. Dotted boxes around proteins indicate proteins are likely co-expressed as part of a wider locus.

A. pleuropneumoniae isolates were sourced from a diverse geography and serotypes from various time points from 1964 onward. Nearly all isolates were from pigs, with a single isolate from a lamb host.

The isolate from the lamb host had a truncated *hgp10* gene, which was the only instance of this gene being truncated in isolates. **Supplementary File 1** contains the genomic information for *A. pleuropneumoniae*. Further trends observed in **Supplementary File 1** included the association of *omlA* group 3 with a functional Fishing Rod 1 but not a functional *fupA1* gene. The *fupA1* gene was also truncated in all isolates with *omlA* group 1, whereas group 2 and 4 *omlA* genes tended to have a functional *fupA1* gene. *tbpA6* was functional in isolates with the *omlA* group 1 and 2 genes. These observations align with the prior use of *omlA* alleles to detect and group serotypes [53].

Previous Literature Identified 17/30 *A. pleuropneumoniae* iron uptake proteins

TbpA and TbpB have been previously described in *A. pleuropneumoniae* [54, 55], including antigenically distinct serovars of TbpBs. Additionally, an Hgp (published as HgpA) [54] which binds porcine haemoglobin [56], OmpW [57], FhuA (Ferric hydroxamate receptor) [54], CirA (APJL_0554 from strain JL03 – called Fup1 herein) [58], FrpB (referenced as APJL_1312, named FcmA herein) [58], TdX (APJL_1922) [58] and a separate FrpB (APJL_0286) [58] were also found in prior literature. Another publication noted *A. pleuropneumoniae* FrpB is an orthologue to FrpB/FetA of *Neisseria spp* [59]. OmlA has been described in *A. pleuropneumoniae* [60] and is likely involved in iron uptake [61]. There are multiple alleles (four) of *omlA* which are associated with different serovars of *A. pleuropneumoniae* [53, 62].

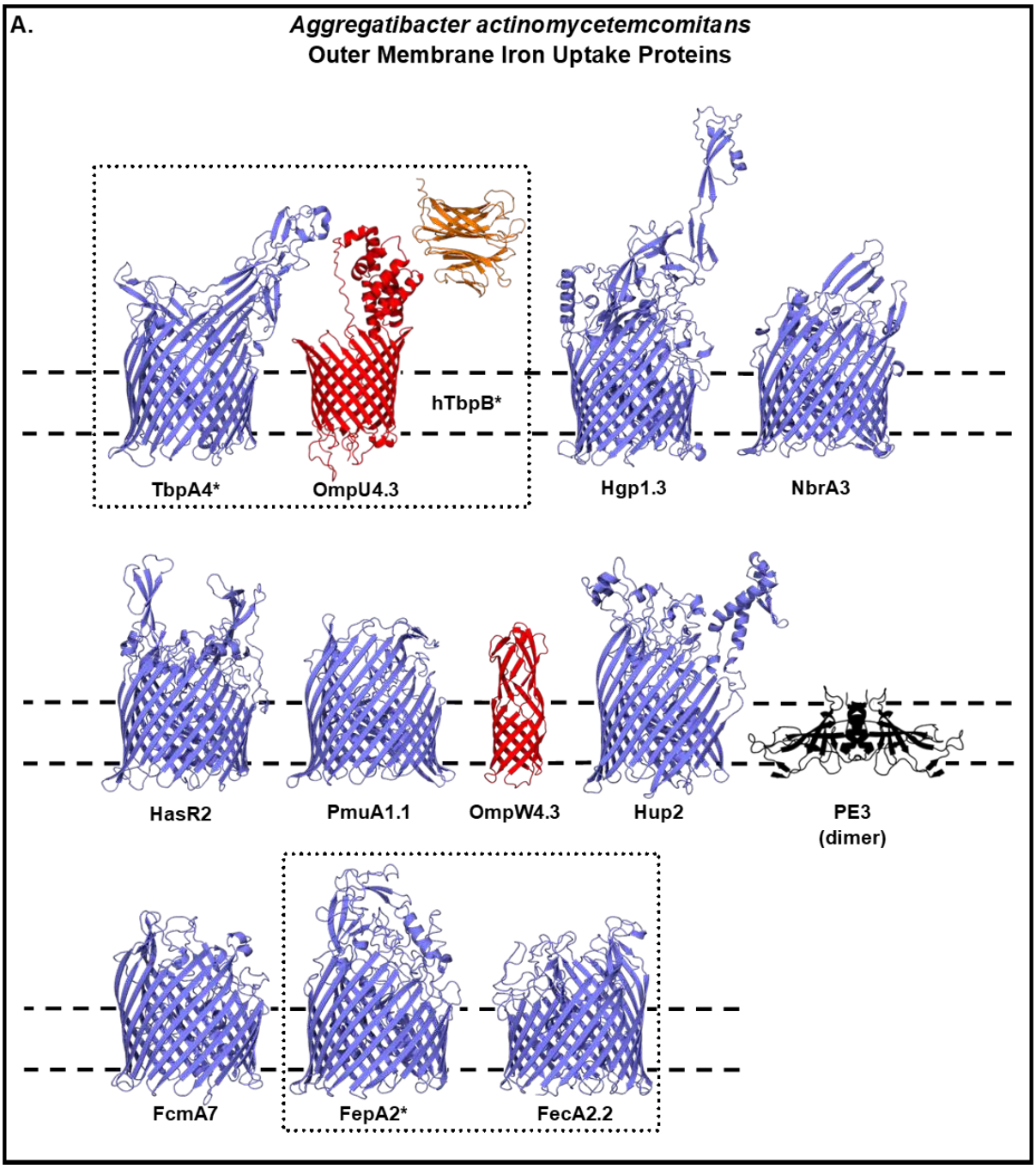


Figure 3: Iron uptake proteins of *A. actinomycetemcomitans*. The rate (%) at which genes were observed in organisms as well as the rate (%) these genes were predicted to be functional: ***tbpA4*** 100%; 35%, ***ompU4.3*** 100%; 96%, ***hTbpB*** 100%; 69%, ***hgp1.3*** 100%; 50%, ***nbrA3.2*** 100%; 85%, ***hasR2*** 100%; 100%, ***pmuA1.1*** 62%; 50%, ***ompW4.3*** 100%; 69%, ***hup2*** 100%; 69%, ***pE3*** 100%; 100%, ***fcmA6*** 100%; 100%, ***fepA2*** 88%; 26%, ***fecA2.2*** 100%; 54%. Proteins marked with '*' are prone to ON/OFF phase-variable expression. Dotted boxes around proteins indicate proteins are likely co-expressed as part of a wider locus.

Previous Literature Identified 3/13 *A. actinomycetemcomitans* iron uptake proteins

A previous summary from 2016 covered the known literature of the time well [63]. Three outer-membrane-associated iron uptake proteins have been previously identified, named TbpA, HgpA and HasR [45] (TbpA4, Hgp1.3 and HasR2 herein).

Isolates were primarily of European or USA origin from 1991 onward. Notably, there is an overrepresentation of functional *fecA2.2* genes in aggressive periodontitis isolates (from a range of years and geographically distant locations). Further insights may be garnered from **Supplementary File 2**.

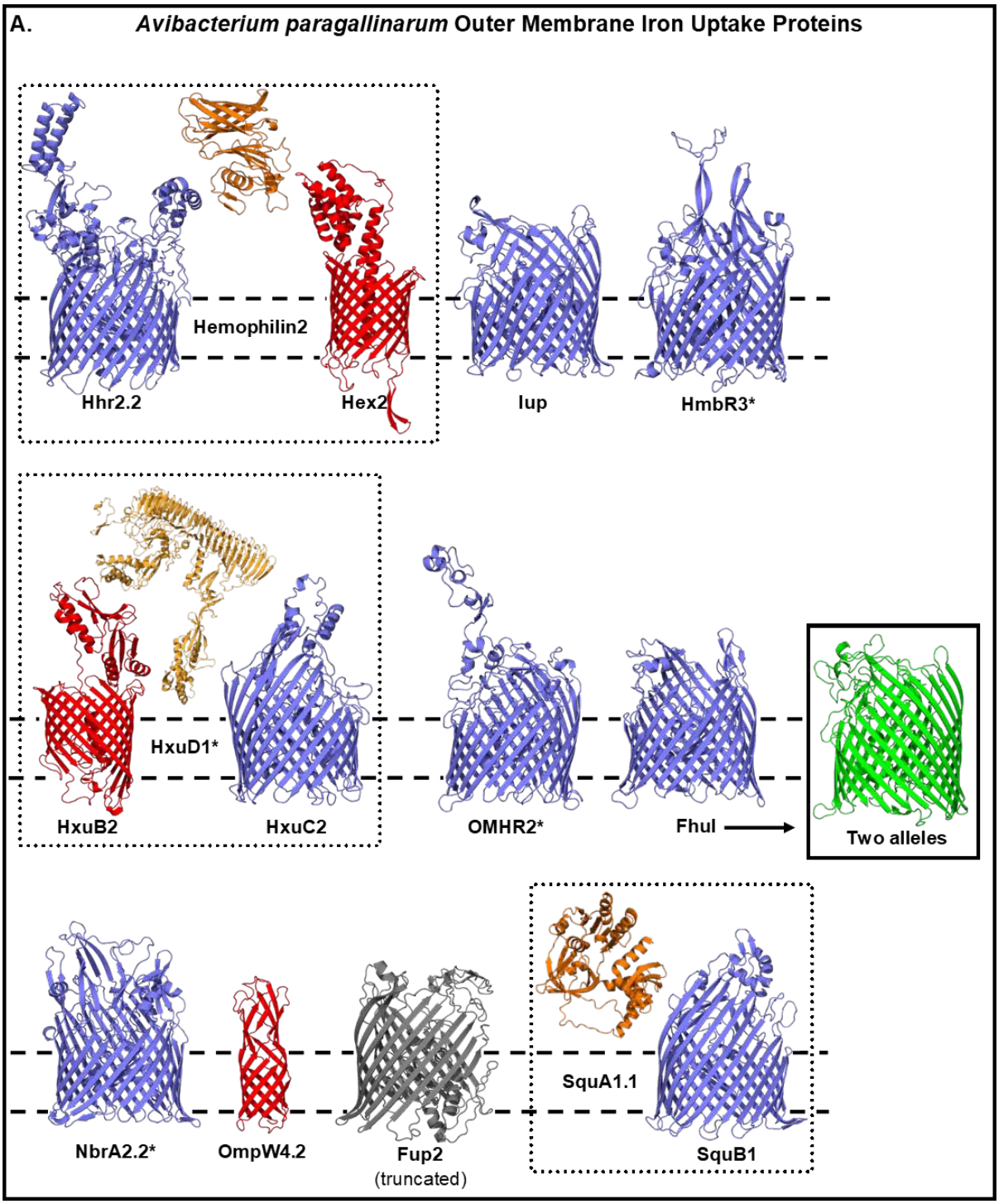


Figure 4: *A. paragallinarum* iron uptake proteins. The rate at which genes were found in genomes and the rate at which they were predicted to be functional: ***hhr2.2*** 9%; 100%, ***hemophilin2*** 9%; 100%, ***hex2***: 9%; 100%, ***iup*** 45%; 80%, ***hmbR3*** 100%; 91%, ***hxuB2*** 100%; 100%, ***hxuD1*** 100%; 100%, ***hxuC2*** 100%; 100%, ***OMHR2*** 100%; 91%, ***fhuI4*** 100%; 91%, ***nbrA2.2*** 100%; 100%, ***ompW4.2*** 100%; 100%, ***fup2*** 64%; 0%, ***squA1.1*** 55%; 100%, ***squB1***

55%; 100%. Proteins marked with '*' are prone to ON/OFF phase-variable expression. Dotted boxes around proteins indicate proteins are likely co-expressed as part of a wider locus.

A. paragallinarum isolates were collected from 2008 onwards from diverse locations. The genomes available were limited to nontypical isolates and serotypes A and C. Serotype B genomes are underrepresented in this dataset. Only five of the 16 iron uptake proteins had been previously identified. Notable findings include the presence of the hemophilin locus in one genome and the identification of five phase-variable genes. Further, though not clonal, isolates from China all contained the *squ* locus and the truncated *fup2* gene. **Supplementary File 3** contains further detailed information. Genomes appear to be from two lineages, which were not associated with serotype. Lineage one contained the *squ* locus, *fhuI5* and *fup2* gene, whereas lineage two lacked these and had the *fhuI4* gene.

Previous Literature Identified 5/16 *A. paragallinarum* iron uptake proteins

Screening *A. paragallinarum* publications resulted in surprisingly few specific mentions of iron uptake proteins. A 2021 publication identified four TonB-dependent receptors upregulated in iron-limited conditions, although the authors did not attempt to name protein products or identify orthologues in other species [64]. That publication didn't establish whether genes were phase-variably expressed or provide sufficient sequence information for closer comparison to findings herein. Contact with these authors was attempted but they did not respond. Previous outer membrane proteins thought to be involved in iron uptake are described by mass as 53, 62, 66, and 94 kDa proteins. Two of these (the 53 and 94 kDa) proteins were found to bind chicken ovotransferrin [65]. More have been suggested but not named [66], which were again listed by mass as "60, 68 and 93 kDa outer membrane proteins." Although we could suggest which of the proteins identified herein this applies to, as no sequence information is available for these mentioned proteins it would only be conjecture.

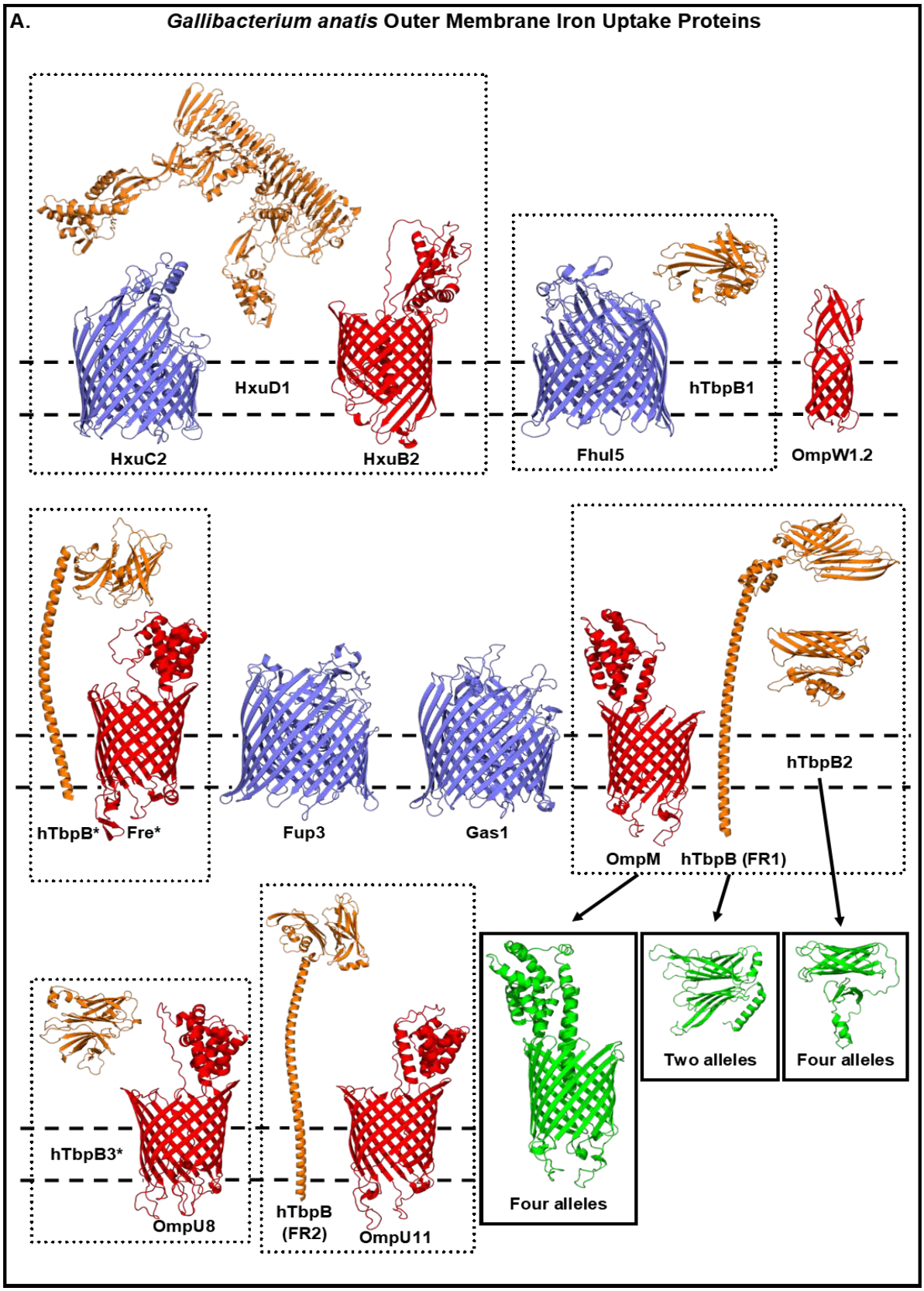

A.
Gallibacterium anatis Outer Membrane Iron Uptake Proteins
HxuD1
HxuC2
HxuB2
hTbpB1
Fhul5
OmpW1.2
hTbpB*
Fre*
Fup3
Gas1
hTbpB2
OmpM
hTbpB (FR1)
hTbpB3*
OmpU8
hTbpB
(FR2)
OmpU11
Four alleles
Two alleles
Four alleles

Figure 5: *G. anatis* iron uptake proteins. The rate at which genes were found in genomes and the rate at which they were predicted to be functional: ***hxuC2*** 89%; 75%, ***hxuD1*** 78%; 100%, ***hxuB2*** 78%; 100%, ***fhuI5*** 11%; 100%, ***hTbpB1*** 11%; 100%, ***ompW1.2*** 100%; 100%, ***hTbpB*** 78%; 71%, ***fre6*** 78%; 71%, ***fup3*** 78%; 100%, ***gasA*** 89%; 100%, ***ompM*** 100%; 89%, ***hTbpB*** (FR1) 33%; 100%, ***hTbpB2*** 78%; 100%, ***hTbpB*** 44%; 100%, ***ompU8*** 44%; 75%, ***hTbpB*** (FR2) 44%;100%, ***ompU11*** 44%; 100%. Proteins marked with '*' are prone to ON/OFF phase-variable expression. Dotted boxes around proteins indicate proteins are likely co-expressed as part of a wider locus.

Isolates were collected from 2011 onwards from diverse geographic locations. Although *G. anatis* is associated with poultry, two isolates were recovered from cow and human hosts. There was no clear correlation between geography and gene distribution. Further details are provided in **Supplementary File 4**.

G. anatis genomes contain a relatively small number of β-barrel receptors, ranging from one to three per genome. No single β-barrel receptor was conserved across all genomes, indicating the absence of a core β-barrel repertoire. In contrast, *G. anatis* encoded a comparatively large number of SLAMs and associated hTbpBs, representing the highest SLAM diversity observed among the *Pasteurellaceae* examined. Two *G. anatis* hTbpBs were phase-variable (**Table 3**), and two were classified as Fishing Rods. The highly reduced repertoire of β-barrel receptors, but increased number of hTbpBs (and secretion machinery) supports the prior suggestion that β-barrel receptors accommodate nonspecific uptake of iron from variable sources [67].

Previous Literature Identified 5/24 *G. anatis* iron uptake proteins

One unnamed heme/siderophore receptor [68] (GasA herein), the *hxuABC* locus [69] and *ompW* [70] have been identified in *G. anatis*.

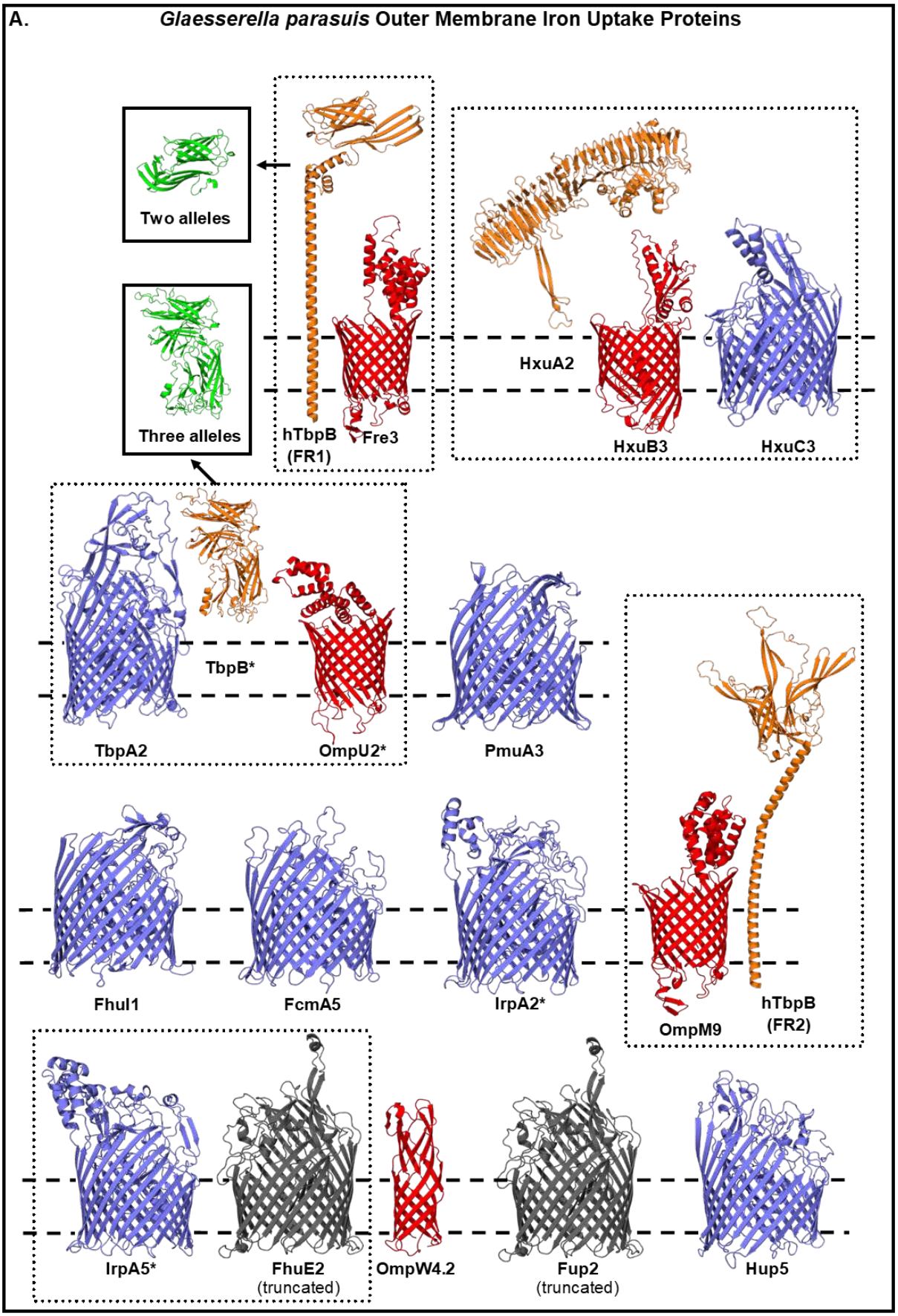


Figure 6: *G. parasuis* iron uptake proteins. The rate at which genes were found in genomes and the rate at which they were predicted to be functional: ***hTbpB*** 64%; 50%, ***fre3*** 64%; 83%,

hxuA2 100%; 93%, ***hxuB3*** 100%; 100%, ***hxuC3*** 100%; 100%, ***tbpA2*** 100%; 100%, ***tbpB*** 100%; 100%, ***ompU2*** 100%; 100%, ***pmuA3*** 43%; 17%, ***fhul1*** 64%; 28%, ***fcmA5*** 61%; 82%, ***irpA2*** 100%; 100%, ***ompM9*** 79%; 82%, ***htbpB*** (FR2) 50%; 100%, ***irpA5*** 100%; 100%, ***fhuE2*** 39%; 0%, ***ompW4.2*** 21%; 33%, ***fup2*** 43%; 0%, ***hup5*** 21%; 100%. Proteins marked with '*' are prone to ON/OFF phase-variable expression. Dotted boxes around proteins indicate proteins are likely co-expressed as part of a wider locus.

Isolates were primarily collected in China from 1986 onward. Examining **Supplementary File 5** supports the division of *G. parasuis* into two main clades [71]. One clade tends to have *fcmA5*, *irpA5*, *fhuE2* and *pmuA3* genes, whereas the other tends to lack these genes and contains the truncated *fup2* gene. The *hxu* locus, the *tbp* locus (including *ompU2*) and *irpA2* appear to comprise a stable core of iron uptake factors for *G. parasuis*.

Previous Literature Identified 8/22 *G. parasuis* iron uptake proteins

Iron uptake in *G. parasuis* is under-characterised with only 8 of the 22 iron uptake proteins identified. There are also three lineages of *G. parasuis* which may have complicated previous investigation. Previously identified iron uptake proteins included HxuA/B/C, TbpA/B, FhuA (*HAPS_RS00485)* and B8F8J9 (HI0997 - the SLAM beside TbpA/B) [72]. One publication identified a gene involved in iron uptake - *cirA* [73], named *fhuA* subsequently [74] and herein.

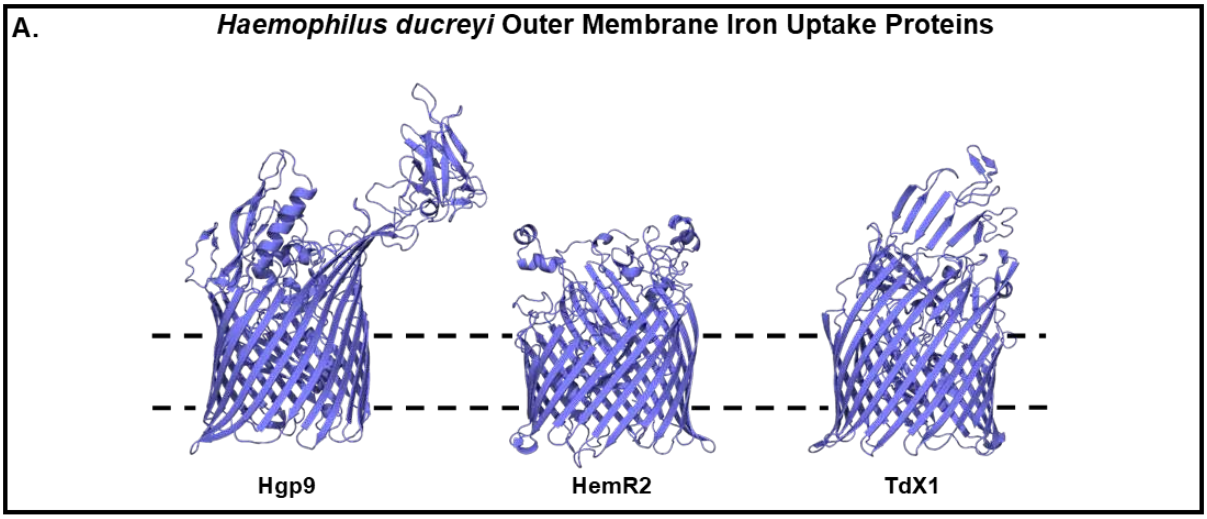


Figure 7: *H. ducreyi* iron uptake proteins. The rate at which genes were found in genomes and the rate at which they were predicted to be functional: ***hgp9*** 100%; 100%, ***hemR2*** 100%; 88%, ***tdX1*** 100%; 100%.

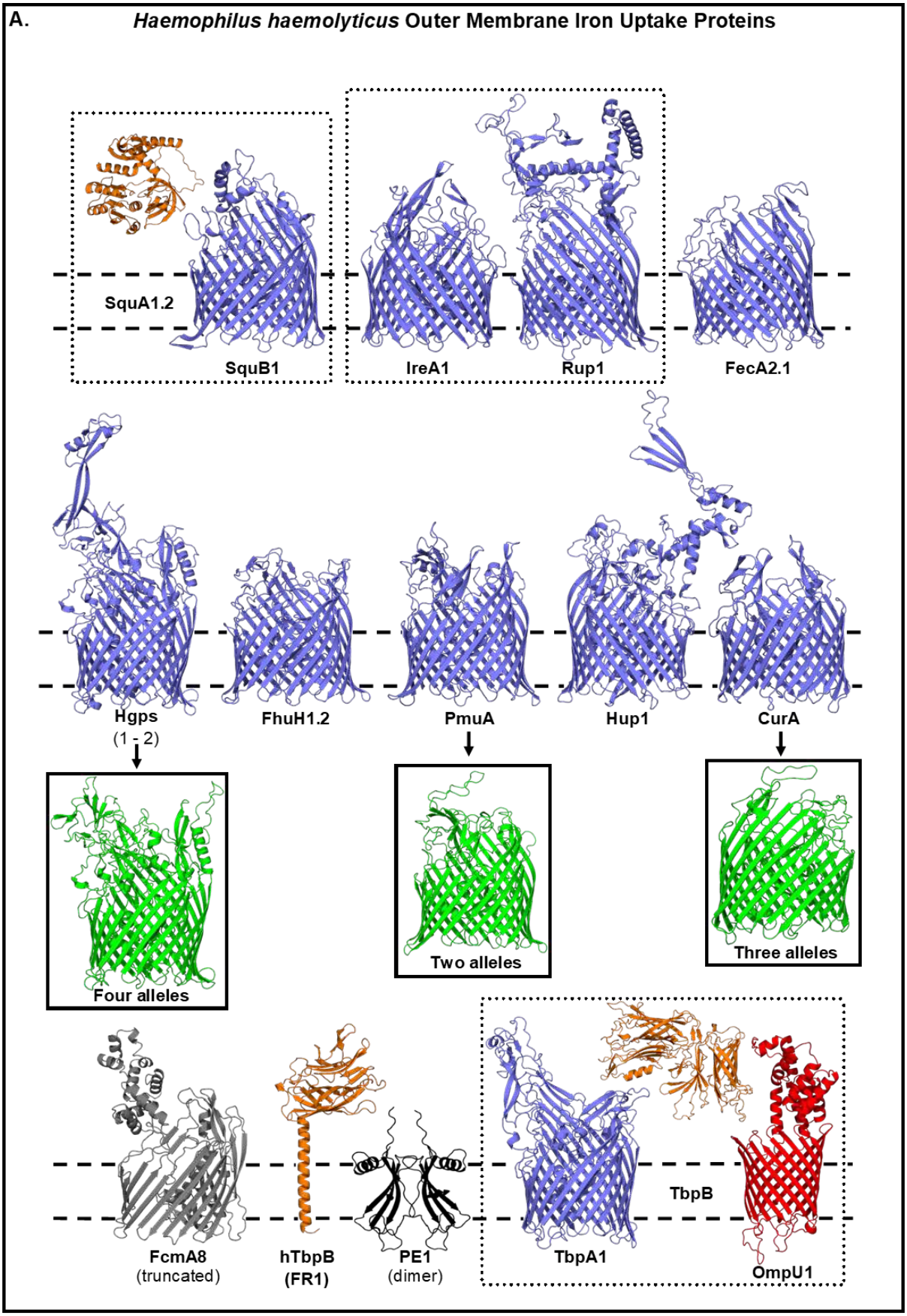
In agreement with prior literature *H. ducreyi* had three iron uptake outer membrane proteins [47, 75, 76]. This was the only organism investigated in which additional unpublished factors were not found. The nomenclature of the previously named *tdhA* and *hupA* was adjusted to *hemR2* and *hgp9*. As seen in **Table 1** *H. ducreyi* had the fewest iron uptake factors compared to other *Pasteurellaceae*. The second-lowest was *A. actinomycetemcomitans* which had 13. This highly reduced diversity of factors may relate to the smaller genome size of *H. ducreyi* at ~1.65Mb (**Table 1**). However, other members of *Pasteurellaceae* with small genomes included *H. haemolyticus* at ~1.9Mb, which had the second highest number of factors at 37. As such, there appears to be no obvious link between genome size and number of iron uptake factors. The fact that *H. ducreyi* only colonises humans also isn't likely the driving force behind the reduced iron uptake repertoire, as both *H. ducreyi* and *H. haemolyticus* only occupy human hosts. More likely to influence iron uptake repertoire is the anatomical site that *H. ducreyi* infects. *H. ducreyi* infects the skin of humans and does not cause bacteraemia [77]. *H. ducreyi*

does not tolerate host systemic conditions, and perishes in temperatures above 35ºC [78]. Most *Pasteurellaceae* appear to utilise iron uptake repertoire expansion and often phase-variation as a means of niche adaptation. Human skin has a complex immune structure that is highly capable of mounting a lasting adaptive response to infection [79]. Despite having just three identified iron uptake surface factors which lacked phase-variable expression, immunity against *H. ducreyi* does not confer lasting protection [80]. This suggests neither extensive receptor diversity nor phase-variable expression are prerequisites for persistence within human skin. How *H. ducreyi* evades or withstands host immune responses remains unclear.

Despite being highly host and niche adapted, *H. ducreyi* does not appear to have exclusive iron acquisition systems; Hgp proteins were observed broadly throughout *Pasteurellaceae*, HemR proteins were identified in *H. influenzae*, *H. somni* and *P. multocida*, and TdX was observed in *A. pleuropneumoniae*. Notably, none of the three factors identified in *H. ducreyi* were restricted to human-associated *Pasteurellaceae* alone. The absence of unique iron uptake systems in *H. ducreyi* suggests adaptation to host niches occurs through modification of existing iron acquisition pathways rather than via evolution of new protein families.

Previous Literature Identified 3/3 *H. ducreyi* iron uptake proteins

All three iron uptake proteins have been previously identified in *H. ducreyi*: TdhA (a HemR orthologue) [75], HupA (Hgp orthologue) [47] and TdX [76]. Unlike HupA and TdhA, TdX expression is not upregulated in heme replete conditions, but rather downregulated [76].


A.
Haemophilus haemolyticus Outer Membrane Iron Uptake Proteins
SquA1.2
SquB1
IreA1
Rup1
FecA2.1
Hgps
(1 - 2)
FhuH1.2
PmuA
Hup1
CurA
Four alleles
Two alleles
Three alleles
FcmA8
(truncated)
hTbpB
(FR1)
PE1
(dimer)
TbpA1
TbpB
OmpU1

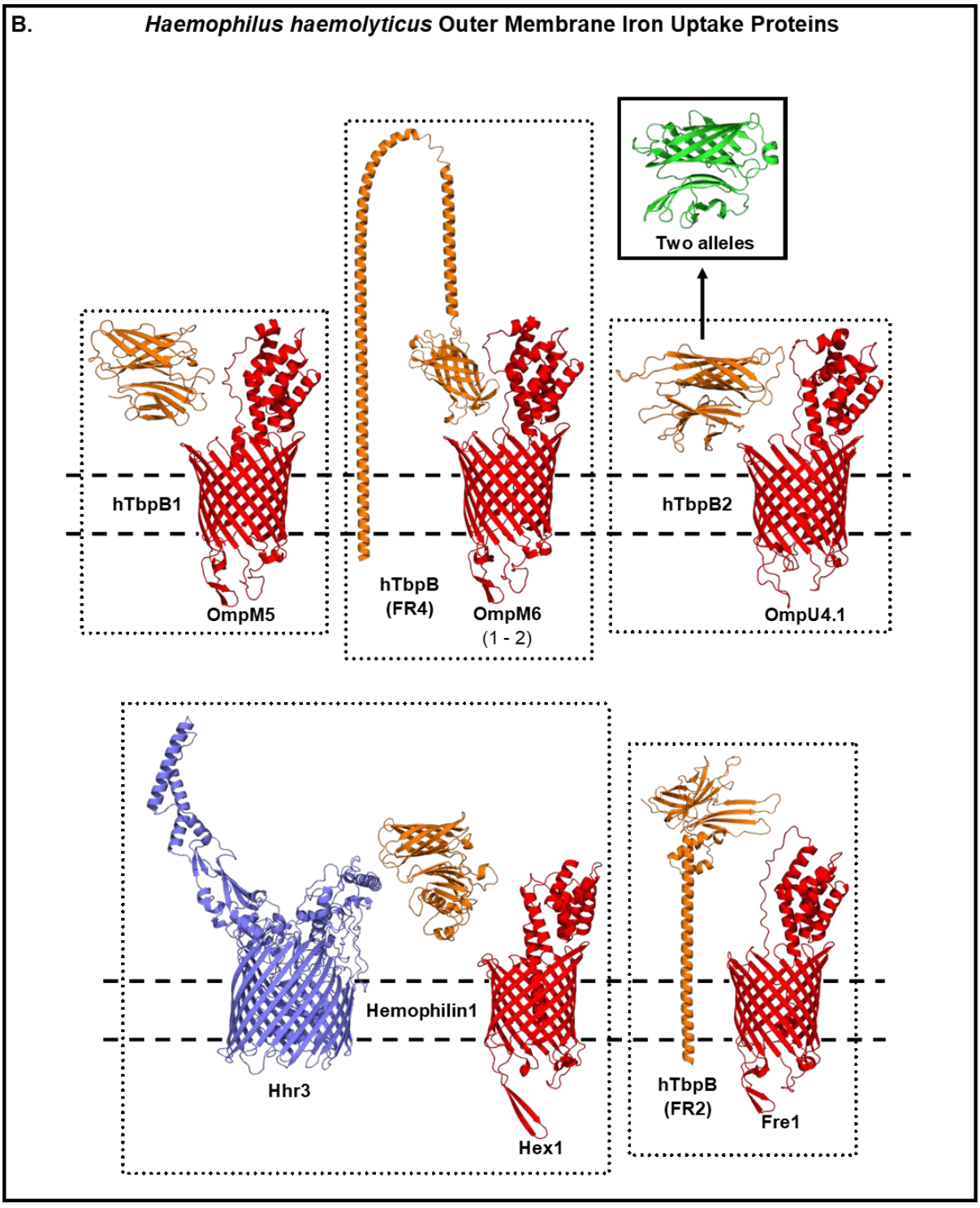


Figure 8: *H. haemolyticus* iron uptake proteins. The rate at which genes were found in genomes and the rate at which they were predicted to be functional: ***squA1.2*** 20%; 100%, ***squB1*** 20%; 100%, ***ireA1*** 40%; 100%, ***rup1*** 40%; 50%, ***fecA2.1*** 40%; 100%, ***hgps*** 100%; 100%, ***fhuH1.2*** 20%; 100%, ***pmuA*** 60%; 100%, ***hup1*** 100%; 100%, ***curA*** 100%; 100%, ***fcmA8*** 20%; 0%, ***hTbpB*** **(FR1)** 20%; 100%, ***pE1*** 100%; 100%, ***tbpA1*** 40%; 100%, ***hTbpB*** 40%; 100%, ***ompU1*** 40%; 100%, ***hTbpB1*** 20%; 100%, ***ompM5*** 20%; 100%, ***hTbpB*** **(FR4)** 100%; 80%, ***ompM6*** 100%; 80%, ***hTbpB2*** (Note: this can be a FR) 60%; 100%, ***ompU4.1*** 80%; 100%, ***hhr3*** 80%; 100%, ***hemophilin1*** 80%; 100%, ***hex1*** 80%; 100%, ***hTbpB*** **(FR2)**

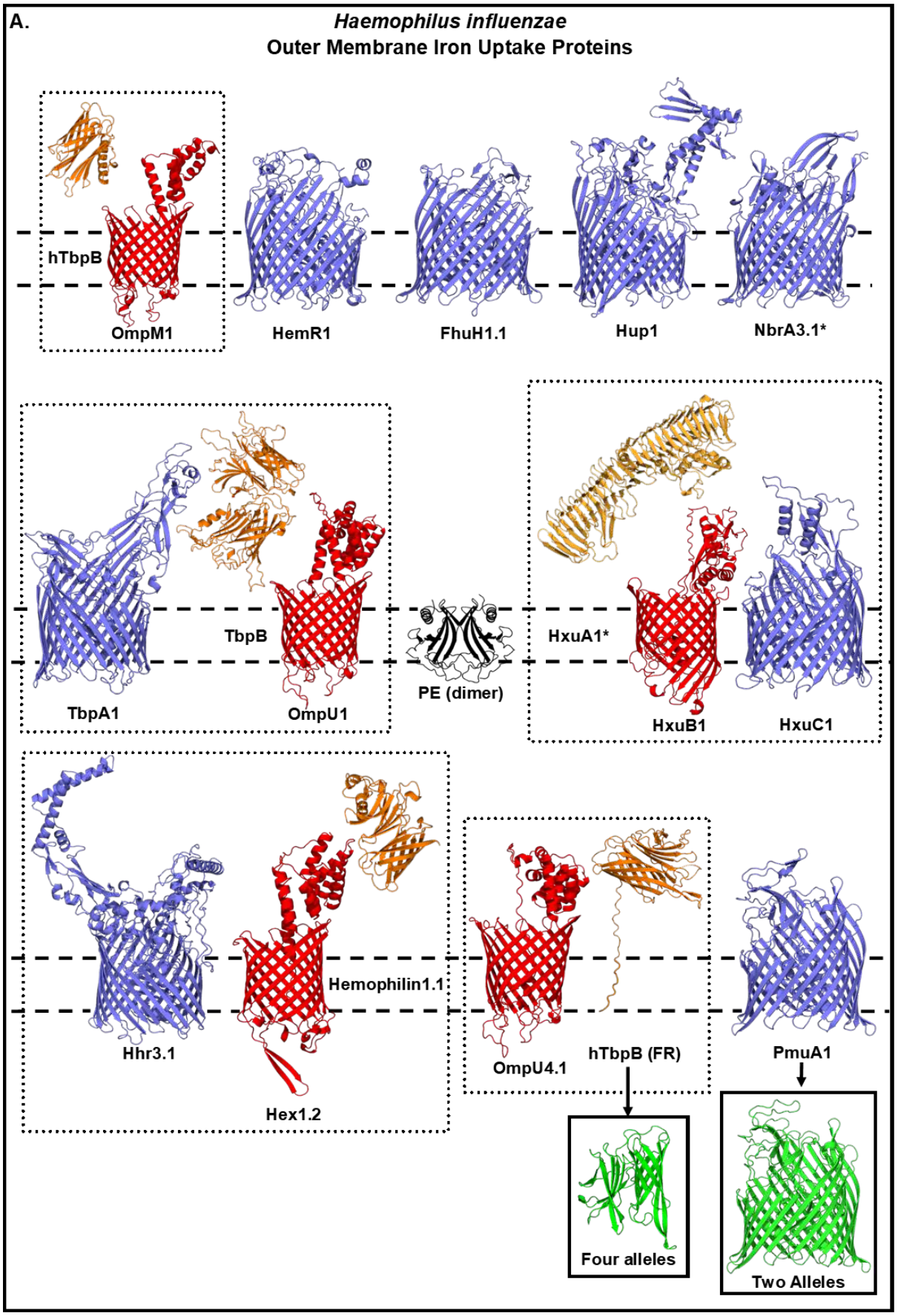
20%; 20%, ***fre1*** 20%; 20%. Dotted boxes around proteins indicate proteins are likely co-expressed as part of a wider locus.

H. haemolyticus had substantial iron uptake outer membrane and secreted proteins. Although only seven had been described previously, 37 unique factors were identified across the five genomes examined. Given the limited number of genomes available, the true diversity of iron uptake genes and alleles in this species is likely considerably larger. *H. haemolyticus* exhibited extensive allelic diversity despite the small sample size. In contrast to other members of the *Haemophilus* genus (except *H. ducreyi*), no ON/OFF phase-variable iron uptake factors were identified. This suggests that *H. haemolyticus* generates diversity in iron acquisition through expansion and diversification of its receptor repertoire rather than through phase-variable expression.

All *Pasteurellaceae* examined shared some common factor with *H. haemolyticus*. Further, all *Pasteurellaceae* examined shared at least one outer membrane β-barrel receptor with *H. haemolyticus*, except *G. anatis*. *G. anatis* had the lowest number of β-barrels in genomes (1-3) and appears to favour SLAMs and hTbpBs over β-barrels. *H. haemolyticus* sharing factors with other *Pasteurellaceae*, even those exclusively found in avian hosts, is particularly noteworthy. This further supports the notion that iron uptake factors are repurposed rather than spontaneously developed, regardless of host niche.

Previous Literature Identified 7/37 *H. haemolyticus* iron uptake proteins

There are few reports of iron uptake proteins in *H. haemolyticus*. This is likely because it's primarily considered a commensal of the human airway, and rarely a pathogen. *H. haemolyticus* is closely related to *H. influenzae*. Literature includes a hemophore named hemophilin, which appears to inhibit growth of *H. influenzae* isolates [81]. Further, the presence of Hup, HemR and an Hgp in *H. haemolyticus* isolates has been described [82]. This same study reported the presence of the *hxuABC* locus rarely in *H. haemolyticus* isolates [82]. Perhaps, due to the small number of genomes available (n = 5), this could not be verified here.


A.
Haemophilus influenzae
Outer Membrane Iron Uptake Proteins
hTbpB
OmpM1
HemR1
FhuH1.1
Hup1
NbrA3.1*
TbpA1
TbpB
OmpU1
PE (dimer)
HxuA1*
HxuB1
HxuC1
Hhr3.1
Hemophilin1.1
Hex1.2
OmpU4.1
hTbpB (FR)
PmuA1
Four alleles
Two Alleles

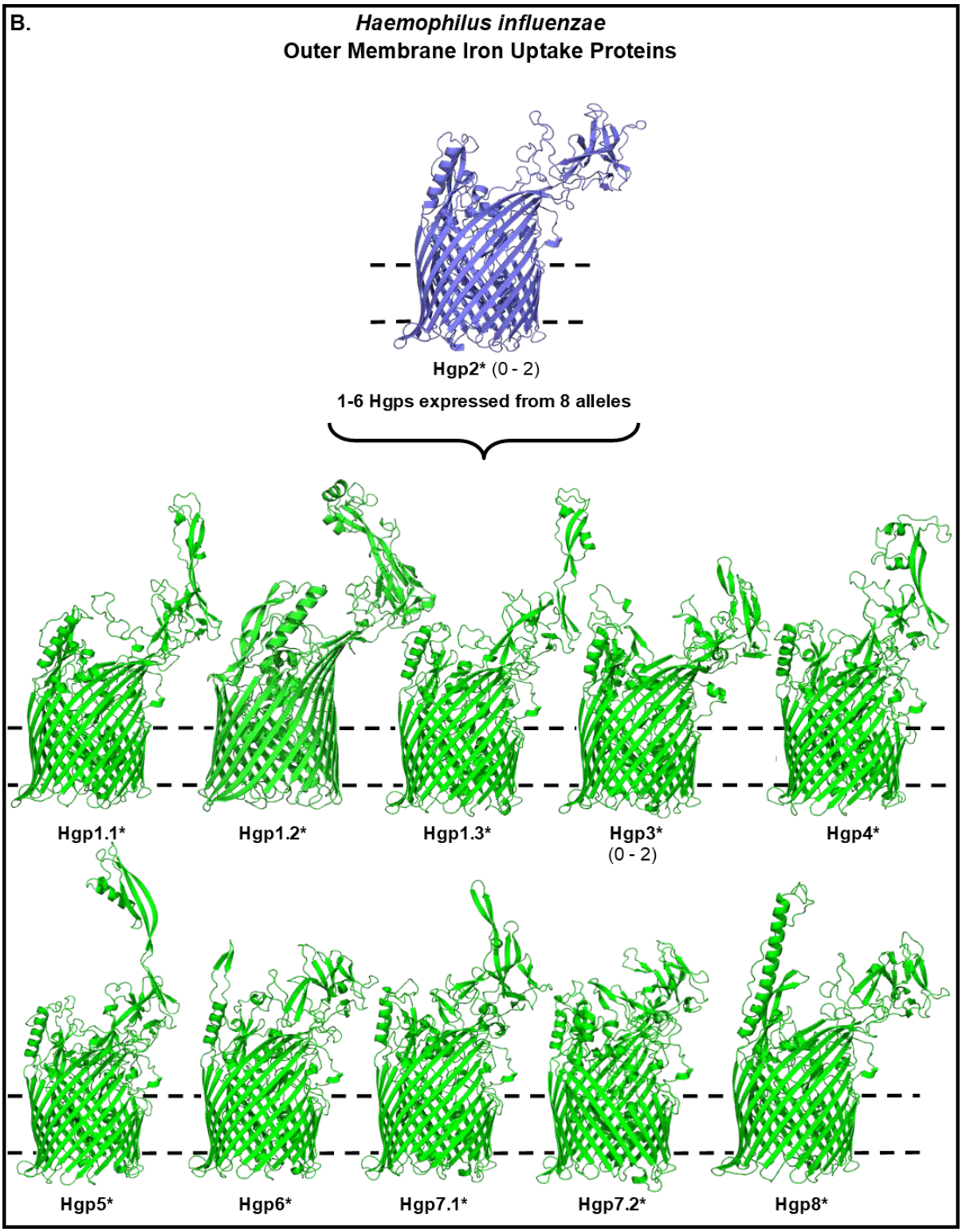


Figure 9: Iron Uptake Outer Membrane repertoire across *H. influenzae*. Note: this includes typeable and non-typeable isolates. **Supplementary File 8** and **Supplementary File 13** contain the statistical data for these genes/proteins. Statistics were separated into NTHi, typeable *H. influenzae* and *aegyptius* isolates. Proteins marked with '*' are prone to ON/OFF phase-variable expression. Dotted boxes around proteins indicate proteins are likely co-expressed as part of a wider locus.

As a prominent human pathogen, *H. influenzae* has been extensively studied and had the largest body of prior literature. 20 of the 34 iron uptake factors (including alleles and sub-alleles) identified in this study had been previously described. A substantial number of genomes were available for analysis (n = 109) and spanned a broad geographic range and collection dates from 1940 onward. Such a broad collection of samples has allowed a detailed overview of iron uptake diversity in *H. influenzae*.

The dataset (**Supplementary File 8**) indicates that *hasR* is largely absent from modern *H. influenzae* isolates. Collected in 1940, *H. influenzae* type B strain NCTC8468 contains *hasR*, whereas more recent genomes lack this gene, except NTHi strain 10P129H1. All genomes were confirmed as *H. influenzae* and not misclassified *H. haemolyticus*. *hasR* is immunodominant in *P. multocida* [83], suggesting that immunogenicity may have contributed to the loss of this gene. The *rapA* gene has replaced the *hasR* gene in modern isolates and has been associated with antibiotic resistance [84]; however, it is also present in the isolate carrying *hasR*.

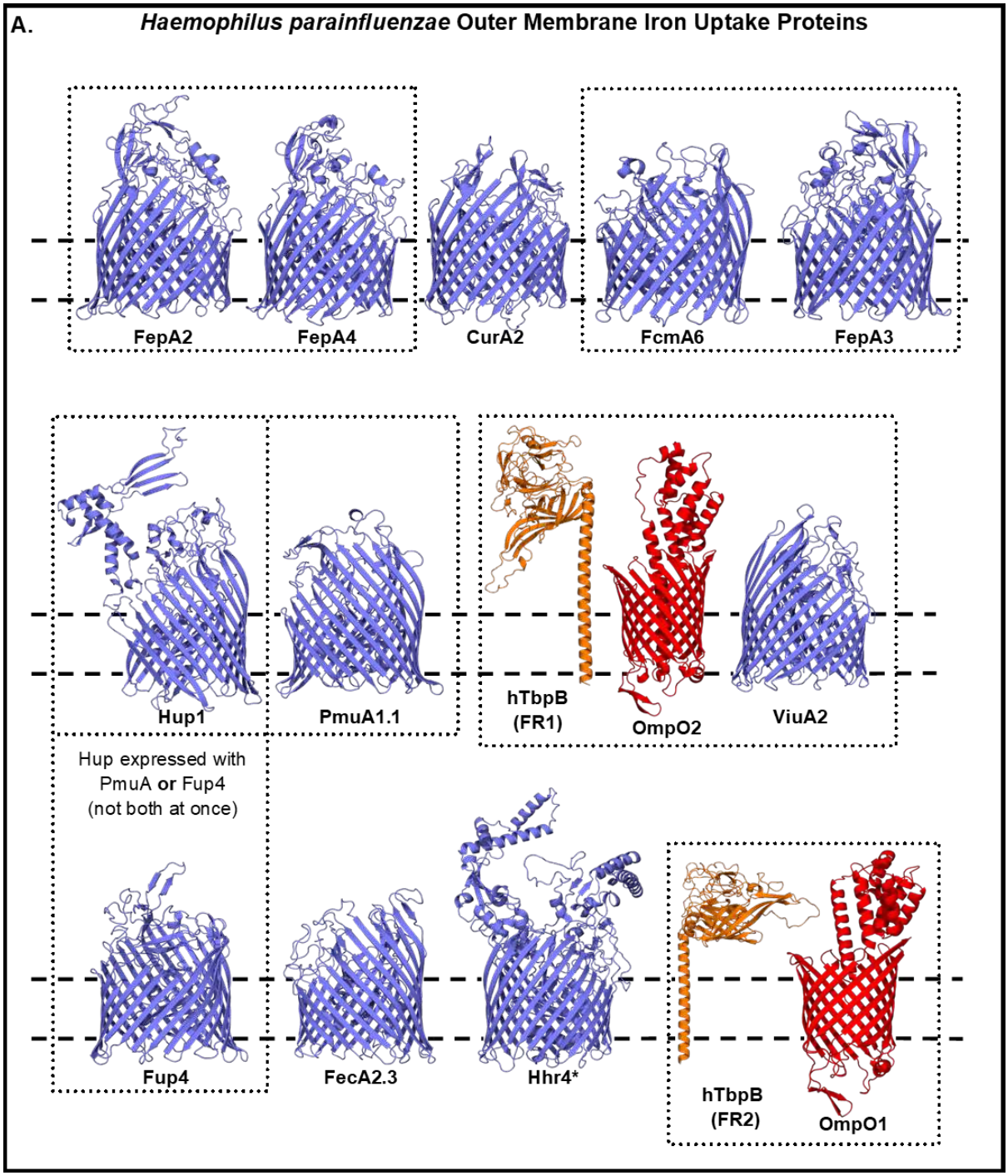
The *hemophilin* locus, previously associated with *H. haemolyticus* and proposed to mediate competition with *H. influenzae* [81], was identified in two *H. influenzae* isolates. The two isolates with the *hemophilin* locus (including receptor and secretion machinery) were confirmed to be *H. influenzae* isolates (and did not mistakenly include *H. haemolyticus* isolates).

Although multiple iron uptake outer membrane proteins are present and highly conserved across *H. influenzae* strains, none show diversity comparable to Hgps. This study expands the known diversity of Hgp alleles in *H. influenzae*, building on previous work [29]. The redundancy of heme acquisition systems is consistent with adaptation to a human host. Notably, hgp8 was detected only in isolates from Taiwan collected in 2016. These isolates were not clonal, indicating that its distribution was not restricted to a single clonal lineage among the genomes examined. Hgp8 from *P. multocida* shares approximately 90% AAID with *H. influenzae* Hgp8. This high sequence identity, together with the gene's distribution across the two species, is consistent with possible horizontal gene transfer, although the direction and timing of any transfer cannot be established from these data alone.

Dividing genomes into typeable, NTHi and *aegyptius* isolates revealed additional patterns in gene distribution (**Supplementary File 8**). For example, *nbrA* was present in 37% of *aegyptius* isolates but was nearly ubiquitously present in NTHi and typeable strains. In contrast, typeable genomes lacked *fhuH* but frequently (75%) contained an additional *ompU* gene (and accompanying *hTbpBs*). These patterns may have implications for vaccine antigen selection.

Previous Literature Identified 20/34 *H. influenzae* iron uptake proteins

H. influenzae has the most previous literature of the organisms examined within this study. Despite evidence that an increased number of iron uptake proteins correlates with increased virulence [82], the field has largely stagnated in the past decade. Including an extra five allelic variants of Hgps previously published [29] 20 iron uptake proteins had been identified versus the 34 observed in this study. The previously identified OM iron uptake proteins include Hup [85], HgpA-G [29], TbpA/B [86], HxuABC, PE [87], HemR (HI0113), HI1369 (NbrA herein) [88], FhuH [89] and one hTbpB [90].


A.
Haemophilus parainfluenzae Outer Membrane Iron Uptake Proteins
FepA2
FepA4
CurA2
FcmA6
FepA3
Hup1
PmuA1.1
hTbpB
(FR1)
OmpO2
ViuA2
Hup expressed with
PmuA or Fup4
(not both at once)
Fup4
FecA2.3
Hhr4*
hTbpB
(FR2)
OmpO1

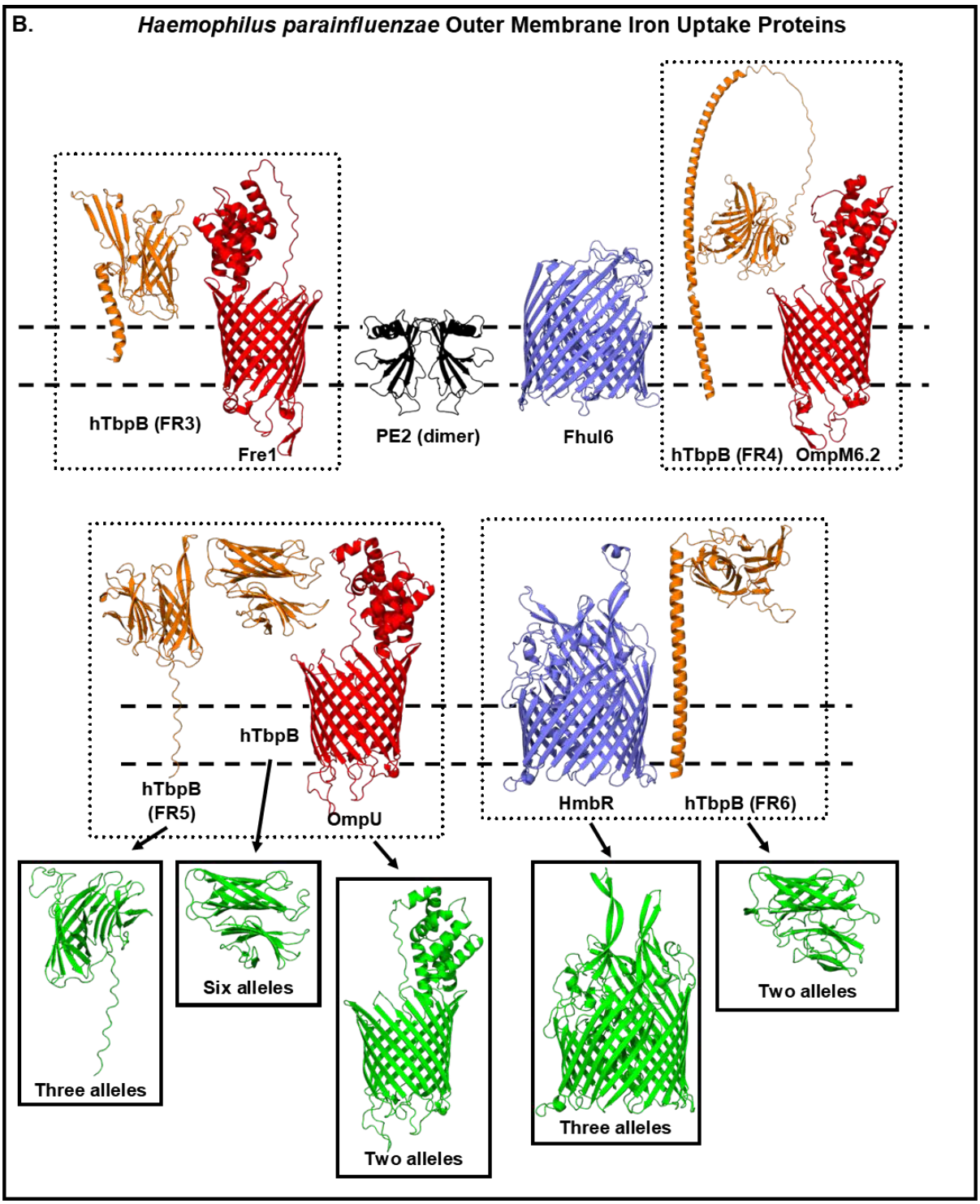


Figure 10: *H. parainfluenzae* iron uptake proteins. The rate at which genes were found in genomes and the rate at which they were predicted to be functional: ***fepA2*** 11%; 50%, ***fepA4*** 11%; 50%, ***curA2*** 100%; 100%, ***fcmA6*** 39%; 100%, ***fepA3*** 100%; 100%, ***hup1*** 28%; 100%, ***pmuA1.1*** 78%; 100%, ***htbpB* (FR1)** 6%; 100%, ***ompO2*** 6%; 100%, ***viuA2*** 22%; 75%, ***fupA4*** 11%; 100%, ***fecA2.3*** 28%; 100%, ***hhr4*** 11%; 100%, ***hTbpB* (FR2)** 28%; 100%, ***ompO1*** 28%; 80%, ***hTbpB* (FR3)** 28%; 80%, ***fre1*** 22%; 100%, ***pE2*** 100%; 100%, ***fhul6*** 6%; 100%, ***hTbpB* (FR4)** 17%; 100%, ***ompM6.2*** 17%; 100%, ***hTbpB* (FR5)** 61%; 91%, ***hTbpB*** 89%; 100%, ***ompU*** 94%; 100%, ***hmbR*** 78%; 93%, ***hTbpB* (FR6)** 39%; 100%. Proteins marked with '*' are prone to ON/OFF phase-variable expression. Dotted boxes around proteins indicate proteins are likely co-expressed as part of a wider locus.

18 *H. parainfluenzae* genomes were examined revealing 38 iron uptake factors, two of which had prior literature. This was the largest pool of iron uptake factors observed for an organism. Genes ranged from a conserved 'core' few (*fepA3*, *pE2*, *curA2*) to genes not observed in other *Pasteurellaceae* examined (e.g. *viuA*, *ompO*).

Previous Literature Identified 2/38 *H. parainfluenzae* iron uptake proteins

There are likely many more allelic variants and iron uptake proteins that remain undetected in *H. parainfluenzae* as only 18 fully annotated genomes were available for study vs ~150 in *H. influenzae*. Unlike *H. influenzae*, *H. parainfluenzae* doesn't require heme for sustained culture. Literature covering iron uptake in *H. parainfluenzae* is nearly non-existent. One older publication (from 1990) identified the presence of a separate *E. coli* ferric enterobactin receptor gene *fepA* in *H. parainfluenzae* isolates [91], though this could not be confirmed herein and is distinct from the *fepA* alleles identified herein. There is also evidence that a hemophilin homologue can be found in *H. parainfluenzae* [81], however we could not verify this finding, and the authors of the publication may have inadvertently used a *H. haemolyticus* isolate. A more recent publication described a hemophilin-like protein in *H. parainfluenzae* [92]. This protein is structurally similar to hemophilin but is not a strict orthologue (a hTbpB herein).

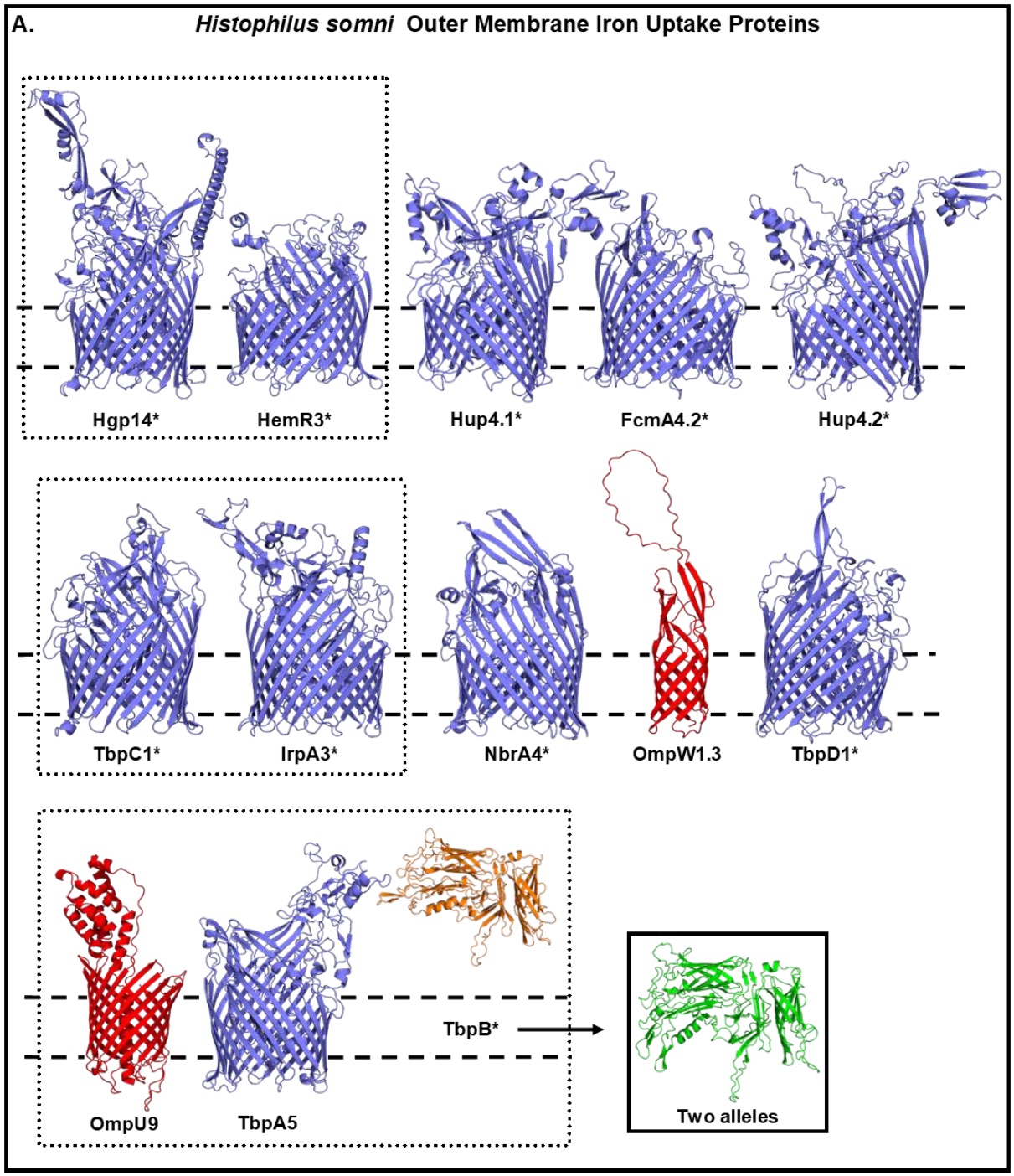


Figure 11: *H. somni* iron uptake proteins. The rate at which genes were found in genomes and the rate at which they were predicted to be functional: ***hgp14*** 100%; 79%, ***hemR3*** 92%; 86%, ***hup4.1*** 100%; 21%, ***fcmA4.2*** 88%; 100%, ***hup4.2*** 100%; 71%, ***tbpC1*** 100%; 88%, ***irpA3*** 100%; 96%, ***nbrA4*** 100%; 88%, ***ompW1.3*** 100%; 83%, ***tbpD1*** 100%; 83%, ***ompU9*** 96%; 83%, ***tbpA5*** 100%; 67%, ***tbpB*** 100%; 83%. Proteins marked with '*' are prone to ON/OFF phase-variable expression. Dotted boxes around proteins indicate proteins are likely co-expressed as part of a wider locus.

H. somni isolates were collected from 1983 onward and limited in geographical origin to either Canada or the USA. The reduced spread of geographical origin likely biased this dataset. Seven of the 14 *H. somni* iron uptake proteins observed herein had prior literature. The previously identified *hgp* of *H. somni* (hgp14) and one *tbpC/B* were previously reported to be

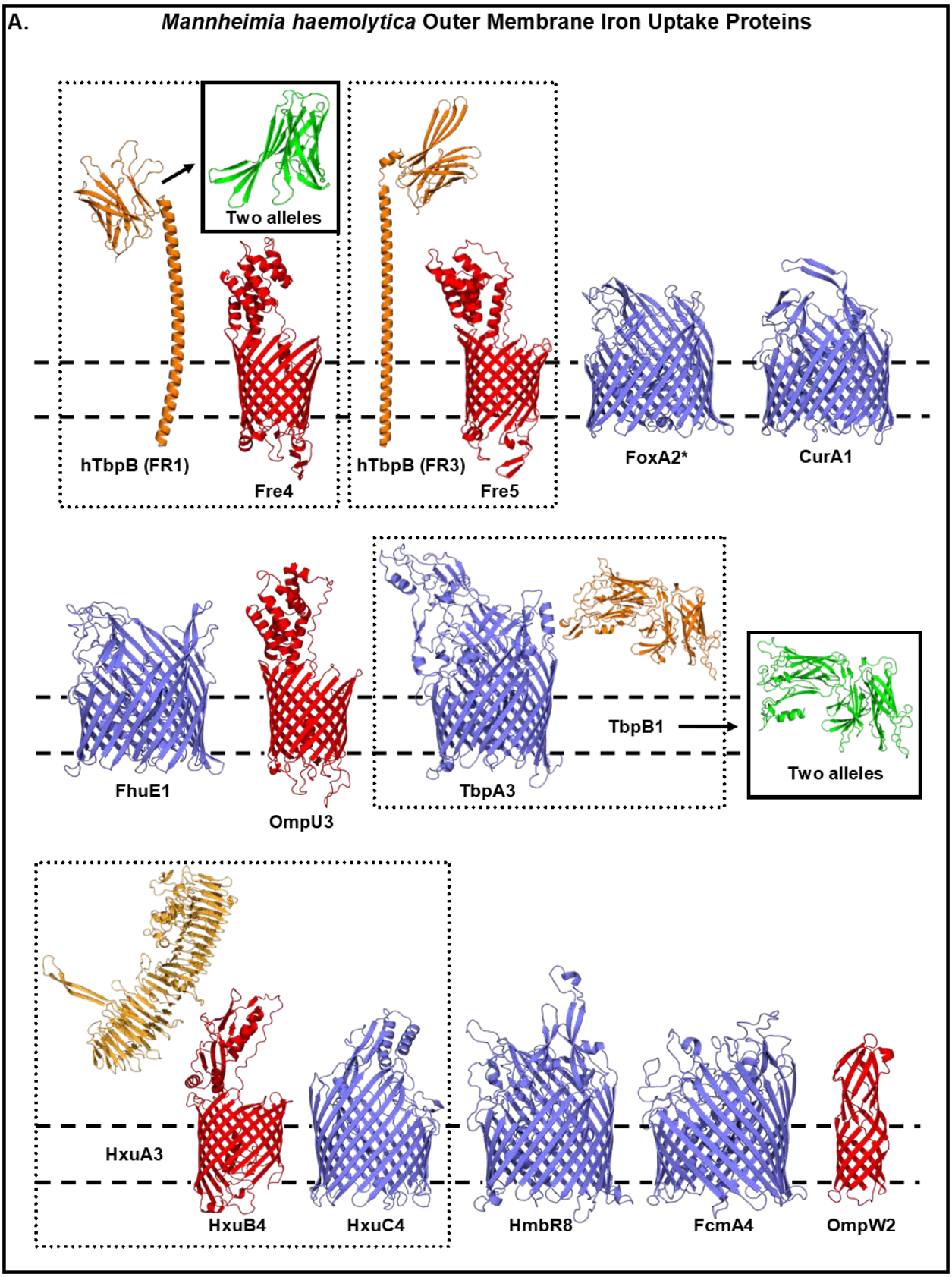
prone to phase-variable expression [44, 93]. *H. somni* appears to have the most phase-variable iron uptake genes of the *Pasteurellaceae* examined in this study (**Table 3**). Nine of the 14 factors appeared to possess either a poly C or G tract near the 5′ (start) of the gene. Phase-variable genes enable immune evasion and allow efficient expression of situationally beneficial factors**.** This is particularly relevant to *H. somni*, as it appears to only have access to relatively few iron uptake factors versus other *Pasteurellaceae* examined (i.e. *P. multocida* with 28).

H. somni was the only *Pasteurellaceae* examined to contain three *tbp* associated receptors – *tbpA*, *tbpC* and *tbpD*. The latter genes are likely phase-variably expressed. A broad BLAST search indicated the *tbpD1* gene appears exclusively in *H. somni*, with unnamed alleles observed in *Campylobacter mucosalis and Campylobacter concisus* (data not shown). *C. mucosalis* is primarily found in pigs *and C. concisus* in humans. It is unclear if the three transferrin binding proteins have been retained in genomes due to host adaptation, immune evasion or for functional redundancy. Notably, all three loci have a factor that is phase-variably expressed.

H. somni was again unique among the *Pasteurellaceae* examined as genomes may express two non-identical, phase-variable *hup* genes. All genomes contain two *hup* genes, though only ~21% contained a functional *hup4.1* gene whereas *hup4.2* was functional in 71% of genomes. It is unclear if nutrient availability or host immune pressures are promoting the redundancy of *hup* genes and transferrin binding protein receptors (A, C, D) in *H. somni*.

Previous Literature Identified 7/14 *H. somni* iron uptake proteins

There are seven previously identified OM iron uptake proteins of *H. somni*; two TbpA homologues [93], TbpB/C [51], HgpA (which was identified as ON/OFF phase-variable) [44], OmpW and FcmA (named FetA in the publication) [31]. However, the previously cited publication also reported unnamed porins, hypothetical proteins and other OM proteins that may be included herein. We identify a total of 14 distinct OM iron uptake proteins in *H. somni* genomes.


A.
Mannheimia haemolytica Outer Membrane Iron Uptake Proteins
Two alleles
hTbpB (FR1)
Fre4
hTbpB (FR3)
Fre5
FoxA2*
CurA1
FhuE1
OmpU3
TbpA3
TbpB1
Two alleles
HxuA3
HxuB4
HxuC4
HmbR8
FcmA4
OmpW2

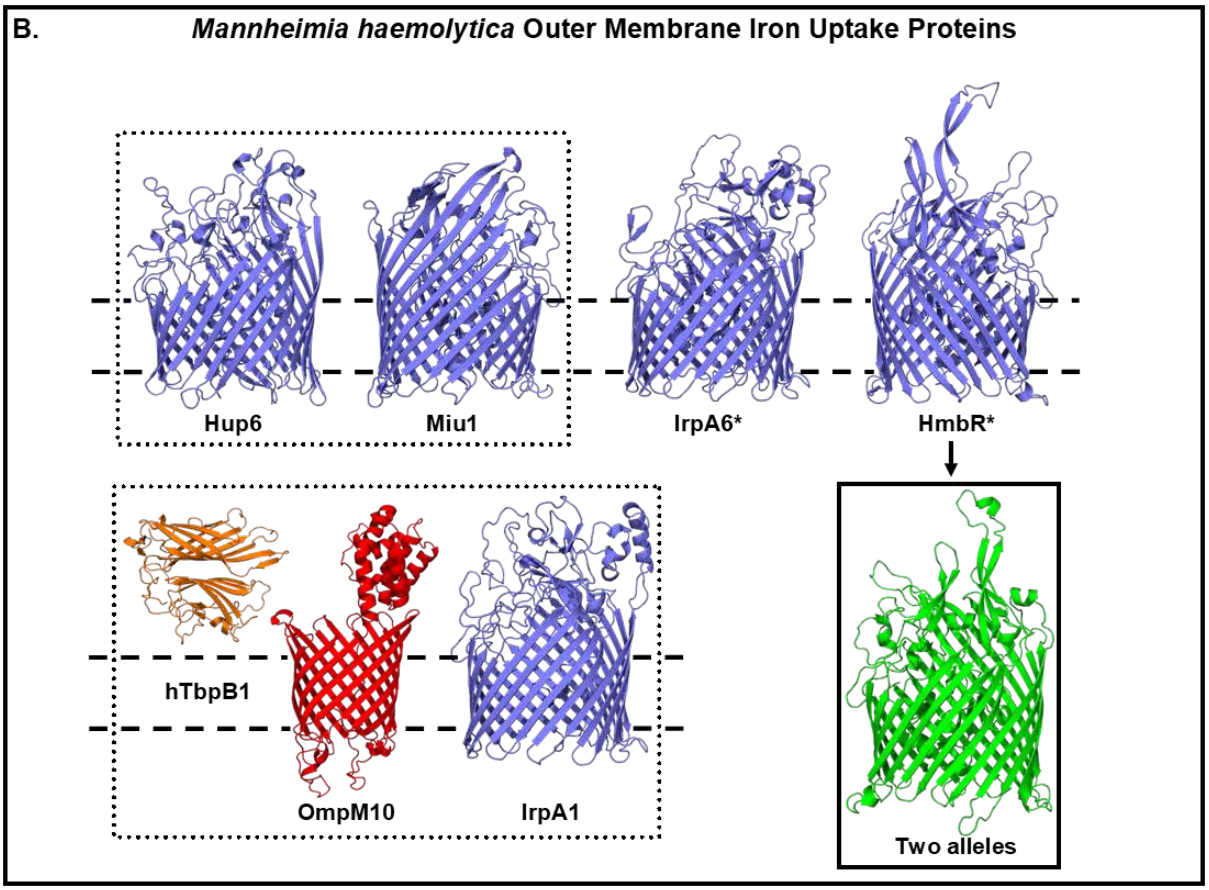


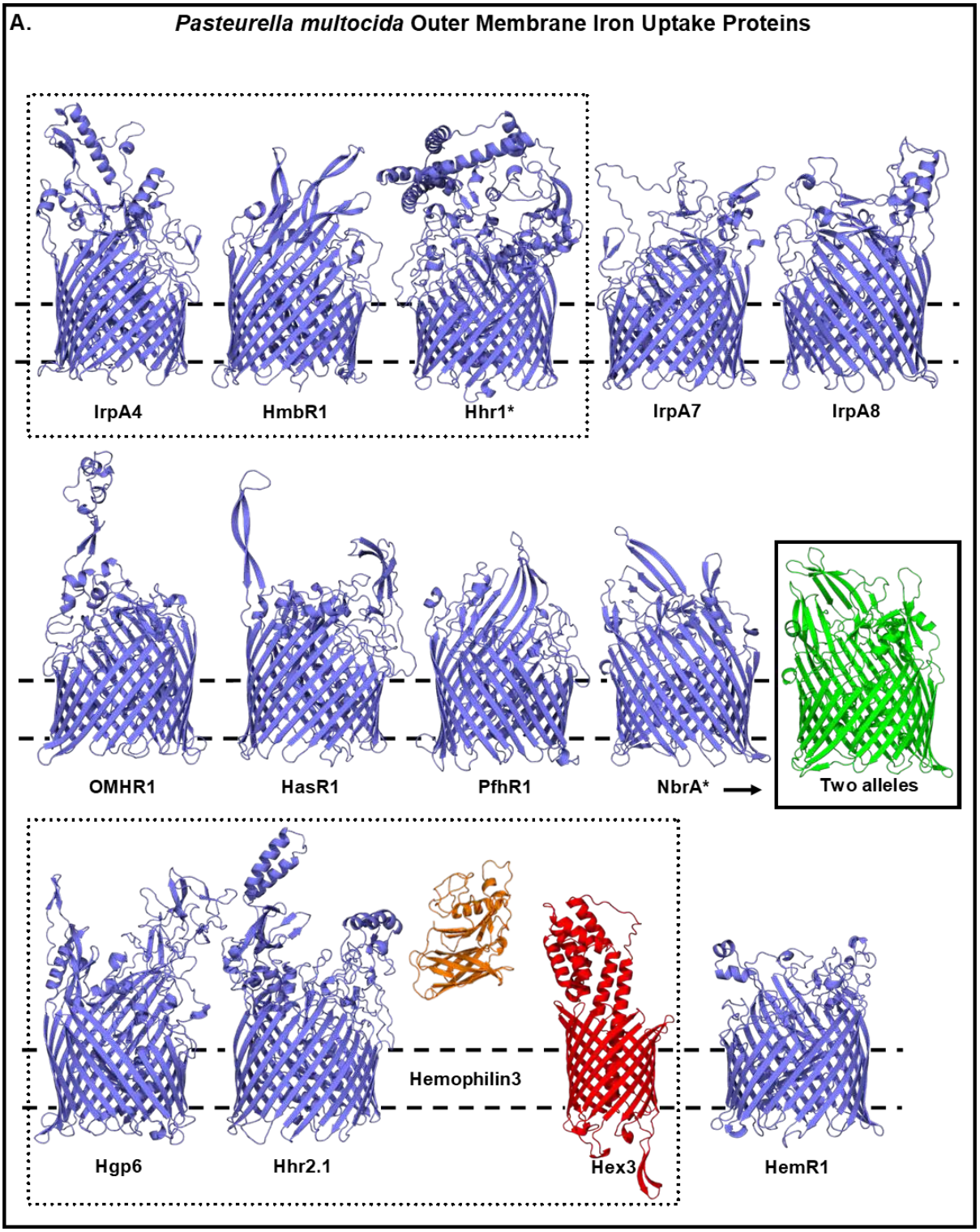
Figure 12: *M. haemolytica* iron uptake proteins. The rate at which genes were found in genomes and the rate at which they were predicted to be functional can be found in **Supplementary File 13**. Further, the division between genome types (1 and 2) and serotypes may be examined in **Supplementary File 11**. Proteins marked with '*' are prone to ON/OFF phase-variable expression. Dotted boxes around proteins indicate proteins are likely co-expressed as part of a wider locus.

M. haemolytica is a significant livestock (cattle/sheep) pathogen with Genotype 2 being a primary cause of bovine respiratory disease [94]. Strains included in this study primarily came from Canada and various States of the USA but also included European/UK isolates. Further, isolates were largely from cows, with 7% of isolates from lamb hosts. Despite prior study, only 11 of the predicted 26 iron uptake proteins had been previously identified.

Figure 12 (above) shows all available iron uptake outer membrane factors identified across *M. haemolytica* genomes. However, different genotypes of *M. haemolytica* are associated with commensal (Genotype 1) versus invasive (Genotype 2) disease. Striking differences are apparent when comparing trends from Genotype 1 versus Genotype 2 isolates (**Supplementary File 11**). Prior investigation comparing the two genotypes noted individual iron uptake factors were present in either Genotype 1 or 2 isolates [95]. Neither Genotype 1 nor 2 isolates had a *hmbR8* gene or *curA1.2* and *fcmA4.1.3* sub-alleles. Genotype 1 isolates were associated with one *tbpB* allele, a truncated *hxuC* gene and a functional *fhuE* gene. Genotype 2 genomes strictly lacked a functional *fhuE* and *hup6* gene, had a low rate of functional *foxA2* genes, had one *hTbpB* (Fishing Rod) allele associated with the *fre4* gene and contained one particular *tbpB* allele (associated with the *tbp* locus).

Previous Literature Identified 11/26 *M. haemolytica* iron uptake proteins

M. haemolytica has multiple previously identified OM iron uptake proteins, including two HmbR proteins, TbpA/B, [96], Irp (IrpA herein) [97], *HxuBC*, FcmA (named FetA in the publication), OmpW, FhuA (FoxA2 herein) and FhuE [31]. The study of Irp in *M. haemolytica* identified IrpA as an orthologue of the *Neisseria* Irp and included that it was ON/OFF phase-variably expressed [97]. An *irpA* gene was found in type 1 Genotypes more frequently than type 2 (pathogenic) Genotypes [98]. HxuA is mentioned briefly by *Gioia et al.* [99].



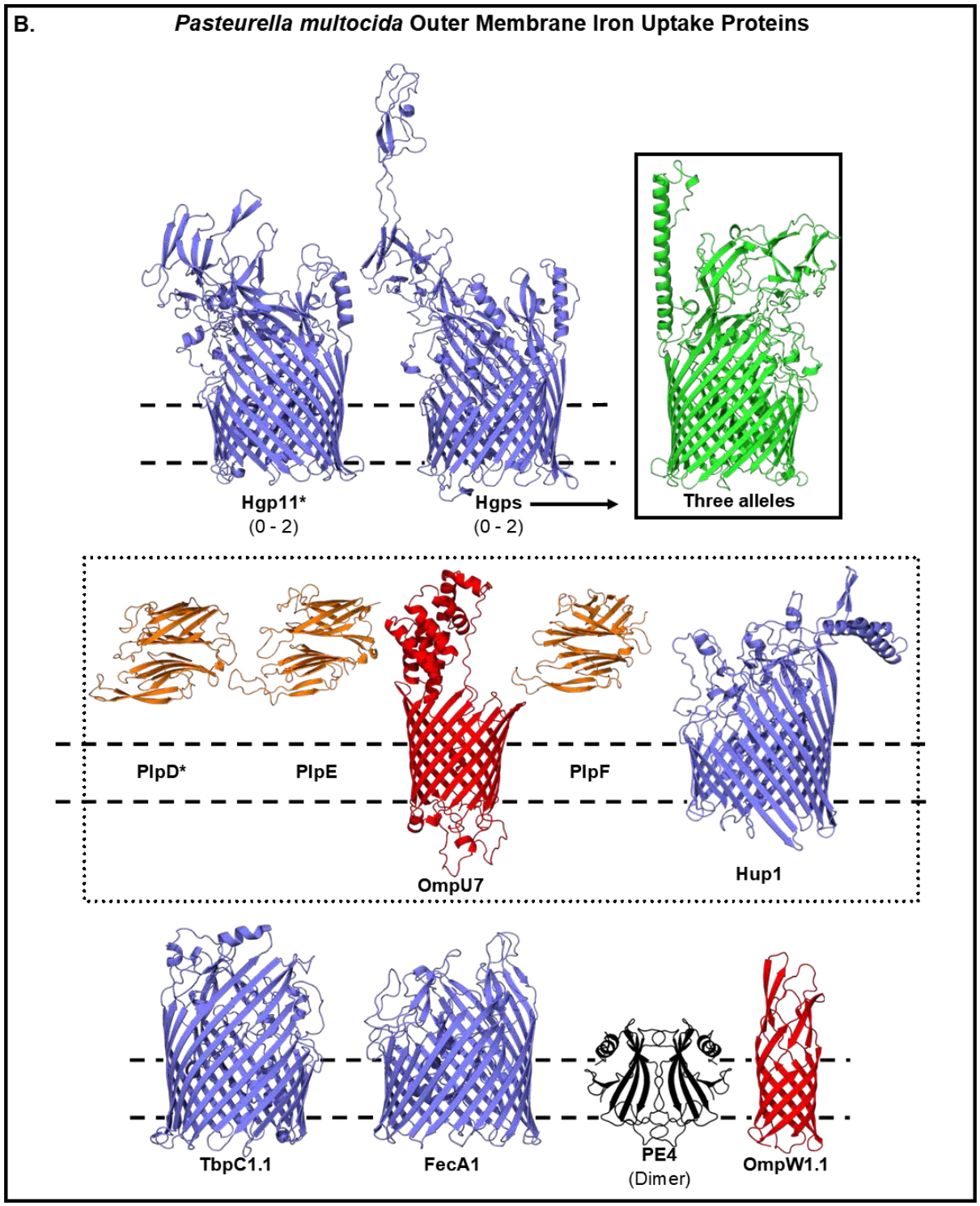


Figure 13: *P. multocida* iron uptake proteins. Proteins marked with '*' are prone to ON/OFF phase-variable expression. Dotted boxes around proteins indicate proteins are likely co-expressed as part of a wider locus. **Supplementary File 11** contains the statistical data for these genes in *P. multocida*. Host origin heavily impacted the distribution of genes.

P. multocida uses a flexible, host-associated iron acquisition receptor repertoire. Isolates were primarily from bird, cow, pig, human and rabbit hosts. Isolates originated from a diverse range of locations, with representation appearing proportional to the number of livestock in these areas. 41 genomes were excluded from statistical calculations because they appeared identical to other strains.

Trends can be observed when comparing isolates from different hosts. Host appears to have a greater influence on the iron uptake repertoire than factors such as geography or year of isolate collection. Examples of stark difference in gene presence are seen in genes such as *fecA*, *hasR*, *hhr* (standalone), *hgps*, *hup*, *nbrA*, *omhr*, *plpE/F and tbpC*. For example, The *hup1* gene was present in 8% of avian isolates (including ducks, chickens, turkeys) but could be seen in the majority of cow (65%) and rabbit (86%) isolates, *hemR1* was associated with cows and pig hosts, but not bird hosts. Further, the standalone (orphaned) *hhr* gene was observed in most (73%) human isolates but no rabbit hosts. **Supplementary File 11** contains statistical data to compare gene presence across hosts.

Hgp in *P. multocida*

Five distinct Hgp alleles were observed in *P. multocida* encoded by *hgp6, hgp8, hgp11, hgp12 and hgp13*. Two of these have been previously identified under the names *hgbA* [46] (our *hgp6*) and *hgbB* [100] (our *hgp11*), including the two nonidentical copies of *hgp11* [50]. Individual *P. multocida* genomes appear to have between two and three functional *hgp* genes. Strains frequently contain two nonidentical copies of *hgp11*. The *hgp11* gene also appears to be prone to ON/OFF phase-variable switching, possessing a poly-A tract approximately 1,200bp into the gene that alternated between seven and eight adenines in different genomes. This is an unusual site for a phase-variable element.

All	*hgp6*	*hgp11*	*hgp12*	*hgp13*	*hgp8*
% Presence	91%	100%	37%	5%	3%
% Functional	96%	97%	78%	80%	100%
Avian					
% Presence	93%	100%	13%	3%	3%
% Functional	97%	96%	100%	100%	100%
Bovine					
% Presence	87%	48%	77%	6%	3%
% Functional	96%	100%	96%	100%	100%
Swine					
% Presence	100%	83%	33%	0%	0%
% Functional	100%	80%	100%	0%	0%
Human					
% Presence	73%	100%	36%	18%	9%
% Functional	88%	85%	50%	50%	100%

Table 5: Distribution of *hgps* across *P. multocida* isolates from different hosts.

As shown in **Table 5** (*above*), there is a high association between *hgp11* and avian hosts (**Supplementary File 11**). Likewise, *hgp12* was associated with bovine hosts. The *hgp6* gene appears nearly ubiquitously present in genomes.

Previous Literature Identified 18/28 *P. multocida* iron uptake proteins

P. multocida has abundant, redundant iron uptake pathways. As *P. multocida* is of particular interest to poultry and livestock industries it is perhaps unsurprising that just over half of the OM iron uptake proteins identified herein have been mentioned in prior literature. These include HgbA (Hgp6), HgbB1 (Hgp11), HgbB2 (Hgp11 second copy) and TbpA [50], *pfhR, hasR [101], hemR*, *fecB/C/D* (but oddly not *fecA*) [102], PM0803 (NbrA) [103], OmpW, FecA, IrpA (two) and PM1282 (OM Hemin Receptor) [104]. Further, PM0576 (HemR), PM0040 (*pfhR*), PM0741 (*hmbR*), PM1081 (*irpA7*), PM1428 (*irpA8*) and HgbA (*hgp6*) have been found

to bind heme in strain Pm70 [105]. A SLAM and associated lipoprotein (PlpE) have been identified but not previously linked to iron metabolism [106, 107].

Summary

The findings support a dynamic accessory repertoire of iron uptake genes across *Pasteurellaceae*, shaped by differences in gene presence, copy number, phase-variation and allelic diversity. The distribution and sequence similarity of some genes are consistent with horizontal gene transfer, suggesting that these factors can move across species and genera.

We describe a total of 44 iron uptake protein families across medically and economically important *Pasteurellaceae*. Nearly all genes and protein families have literature supporting their role in iron uptake. At a species level, less than half (106/285; ~37%) of specific genes/proteins had previous literature identifying them in their respective organisms. This tally included proteins unreported in *Pasteurellaceae* and allelic variants within protein families that were previously unrecognised.

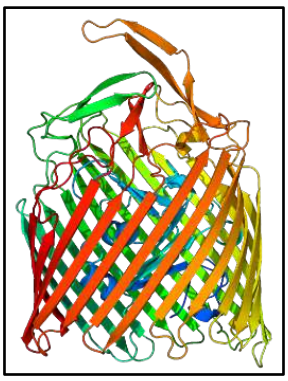
We observed multiple interconnected strategies for iron uptake across the *Pasteurellaceae* examined. Some organisms favour receptor expansion. This was observed in *A. actinomycetemcomitans*, which primarily relies on outer membrane β-barrel receptors. Others appear to favour secreted SLAM associated factors. This was observed with *G. anatis*, which possesses comparatively few β-barrel receptors but an abundance of hTbpBs. A third strategy involves expansion (or contraction) of phase-variable iron uptake factors. For example, *H. somni* appears to favour generating population diversity via ON/OFF phase-variation over increasing the number of iron uptake genes in strains. These strategies are not mutually exclusive and likely represent different solutions to the common challenge of acquiring iron while adapting to the host.

These observations suggest that *Pasteurellaceae* iron uptake systems constitute a dynamic accessory repertoire rather than a fixed species-specific ‘core’. This has important implications for understanding host adaptation, virulence evolution, and the development of effective vaccines and therapies. Standardised nomenclature and family-wide analyses provide a unified perspective on iron acquisition that informs future investigations of these systems and their role in pathogenesis.

Protein Analysis

This section contains an introduction and summary of each protein examined. Prior literature has been included, where available. We have attempted to provide gene/protein names adhering to traditional nomenclature [41], pre-modification molecular weight (kDa), the predicted 3D structure (produced in AlphaFold 3 [48]), example amino acid sequence and a list of the *Pasteurellaceae* in which genes were observed. Factors leading to variable expression that would impact vaccine development efforts are included. This includes ON/OFF phase-variable gene expression. Further, we have included a simplified representation of the locus of the gene. Arrows (-> or <-) indicate the direction of the gene in the locus. **Supplementary File 16** contains representative protein sequences.

Catecholate uptake receptor A (CurA)
Structure: β-barrel
Molecular weight: ~70kDa
Example Protein Sequence: ManH_CurA1.2
Found in: *H. haemolyticus* (4), *H. parainfluenzae*, *M. haemolytica* (2)
No. of Alleles: Four
Variable expression: No
Locus: *cur* (*H. parainfluenzae*)
-> *curA* [~1980bp]
-> *curB* (helical backbone metal receptor) [~780bp]
-> *curC* (iron transporter permease) [~850bp]
-> *curD (*ABC transporter) [~730bp]

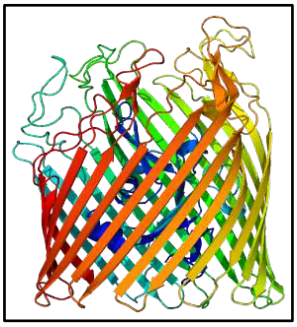


CurA

No prior literature describing the *cur* locus (named herein) was found. *curA* was present and functional in all *H. haemolyticus*, *H. parainfluenzae* and *M. haemolytica* genomes. In *H. parainfluenzae curA* sits on the *cur* locus (this study). The SIFTER results for CurA in *H.* parainfluenzae (UniProt entry A0ABN0EVH8) suggest the *cur* locus may be involved in heme binding (= 0.41) (**Supplementary File 15).** The SIFTER results also showed a high probability for colicin transmembrane transporter activity (= 0.8) and starch binding (= 0.69), though c*urA* lacks notable similarity to known colicin and starch uptake genes (data not shown). Further, the *cur* operon has the characteristics of an iron uptake locus, producing CurA (OM β-barrel receptor) and ABC import machinery in the form of CurB (helical backbone metal receptor), CurC (iron transporter permease) and CurD (ABC transporter ATP-binding protein). However, the *M. haemolytica* and *H. haemolyticus cur* locus appears more complicated with additional genes, including predicted heme ABC transporters. In *M. haemolytica*, the *curA1.1* gene is ~50kbp upstream of a Fishing Rod; likely involved in iron uptake, and 10kb downstream of the *fhu* locus. Should future study reveal that CurA is a cobalamin/B12 receptor, it's noteworthy that there's competition for iron for both cobalamin receptors and siderophore receptors [108].

In *M. haemolytica,* CurA appears to have two forms - dubbed CurA1.1 and CurA1.2. CurA1.1 was common, present in 97% of *M. haemolytica* genomes and ~10kb downstream of the *fhu* locus. Even in the minimal amount of *H. haemolyticus* examined (n = 5), four alleles were detected. Only one form of *curA*/CurA was present in *H. parainfluenzae*, which was also one of the alleles seen in *H. haemolyticus* genomes (*curA2*). The exact role of the *cur* locus remains unknown, though its structure, associated locus, high presence and tendency to remain functional in genomes suggest it plays an important role in iron metabolism.

Ferric citrate receptor A (FecA)
Structure: β-barrel
Molecular weight: ~76kDa
Example Protein Sequence: AggA_FecA2.2
Found in: *A. actinomycetemcomitans*, *H. haemolyticus*, *H. parainfluenzae*, *P. multocida*
No. of Alleles: Two
Variable expression: No
Locus: Associated with *fep* locus
<- *fepA* [~2400bp]
<- ABC transporter ATP-binding protein [~750bp]
<- iron ABC transporter permease [~1000bp]
<- iron ABC transporter substrate-binding protein [~1000bp]
-> iron ABC transporter substrate-binding protein [~680bp]
-> *fecA* [~2000bp]



FecA

FecA is a 22-stranded β-barrel previously associated with iron uptake in *P. multocida* [104], but not described in *A. actinomycetemcomitans*, *H. haemolyticus* or *H. parainfluenzae*, where it was found in this study. *fecA* appears more abundant in *A. actinomycetemcomitans* genomes, which all had a *fecA* gene, though only ~54% were predicted to be functional **(Supplementary File 1)**. Sorting *P. multocida* genomes by host source reveals *fecA* may also be important during infection of cattle (65% gene presence, 100% functional) and pigs (92% gene presence, 91% functional). *P. multocida* isolated from rabbit hosts has a notably lower rate of *fecA* presence (14%) and no functional *fecA* genes. This may be an artifact of the small sample size of *P. multocida* isolated from rabbit hosts in this study (n = 7) or because of the geographical distribution of the isolates (being from China or France).

Ferric enterobactin receptor A (FepA)
Molecular weight: ~70kDa
Example Protein Sequence: HaeP_FepA2
Found in: *A. actinomycetemcomitans*, *H. parainfluenzae* (3)
No. of Alleles: Three (plus one in *E. coli*)
Variable expression: Possibly ON/OFF Phase-variable in *A. actinomycetemcomitans* via short poly-G (TATGGGGCCG).
Locus: *fep*.
<- *fepA* [~2400bp]
<- ABC transporter ATP-binding protein [~750bp]
<- iron ABC transporter permease [~1000bp]
<- iron ABC transporter substrate-binding protein [~1000bp]
-> iron ABC transporter substrate-binding protein [~680bp]
-> *fecA* [~2000bp]



FepA

The enterobactin siderophore uptake protein FepA was first described in *E. coli* [109] and binds ferric catecholates [110]. FepA is a 22-stranded β-barrel. An orthologue of this protein, named PfeA, has also been found in *P. aeruginosa* [111]. We have preserved the legacy name "FepA," however, *fepA* appears to be the last of multiple genes in the *fep* locus and shouldn't necessarily have the "A" identifier. As such, we have held off naming the rest of what looks to be the *fep* locus, which contains typical ABC machinery associated with iron uptake. FepA orthologues are found in two of the *Pasteurellaceae* examined – *A. actinomycetemcomitans* and *H. parainfluenzae*. We have reserved the name "FepA1" for the original *E. coli* FepA protein. FepA has also been observed in *A. baumannii* (52% AAID) as a siderophore uptake

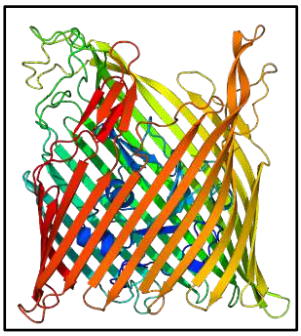
protein (named PirA) [112]. As FepA has long been associated with iron uptake via siderophores the orthologues found in *Pasteurellaceae* were included in this study.

All *H. parainfluenzae* isolates contained at least one functional *fepA* gene, with 2/18 isolates having a second *fepA* allele identical to that found in *A. actinomycetemcomitans*. Since these two organisms occupy the same space anatomically (human respiratory tract), it's good evidence that these organisms can share important genes, such as FepA. A short poly-G tract observed in many strains of *A. actinomycetemcomitans* have a 1bp change (C-to-T) to truncate the gene, which would otherwise be functional. *A. actinomycetemcomitans* isolates had a relatively high *fepA* gene presence (88%) but had a low rate (26%) of *fepA* functionality.

The false *fhuA* gene

There are many unrelated genes named *fhuA*. FhuA of *E. coli* [113] is the original, well-studied iron uptake protein. However, there's also *fhuA* of *G. parasuis* [74], the *fhu* locus of *H. influenzae* and *A. pleuropneumoniae* (including *fhuA*) [114] and *fhuA* of *Neisseria meningitidis* [115]. These are all involved in iron uptake and include an outer-membrane β-barrel, yet only *fhuA* of *A. pleuropneumoniae* and *G. parasuis* are homologues. The *fhuA* gene of *N. meningitidis* is notable because it has an orthologue in *A. paragallinarum* (this study), though shares just 24% AAID to FhuA of *E. coli* (data not shown). Homologous genes can occasionally be observed across the otherwise unrelated *fhu* locus of the above organisms; *fhuC* of *A. pleuropneumoniae* is homologous (~50% AAID) to the *E. coli fhuC* [114]. However, the unique genes and locus need distinct names. As our focus is on the outer membrane receptors, rather than internal machinery, we have renamed the unique (non *E. coli*) *fhuA* genes to the following: *fhuI* in *G. parasuis* [74] (including homologues across *Pasteurellaceae* including *A. pleuropneumoniae* [114]), *fhuH* in *H. influenzae* [114] (including *H. haemolyticus*) and *iup* in *A. paragallinarum* (this study). Importantly, FhuE of *E. coli* [116] can be observed in *M. haemolytica* [31]. Refreshingly, this appears to be named correctly - although the *M. haemolytica* FhuE has only ~34% AAID to FhuE of *E. coli* we consider it a true orthologue. Iup, though previously named FhuA in Neisseria *meningitidis* [115], neither sits on the *fhu* locus nor has significant homology to any proteins transcribed on a *fhu* locus, which is why it was given an alternate prefix.

Ferric coprogen receptor E (FhuE)
Structure: β-barrel
Molecular weight: 79kDa
Example Protein Sequence: ManH_FhuE1
Found in: *G. parasuis* (truncated), *M. haemolytica.*
No. of Alleles: Two
Variable expression: No
Locus: *fhu*
-> *fhuC* (51% ID to *E. coli* FhuC) [~620bp]
-> *fhuD* (29% ID to *E. coli* FhuD) [~880bp]
-> *fhuB* (38% ID to *E. coli* FhuB) [~1970bp]
-> *fhuE2* [~2170bp]

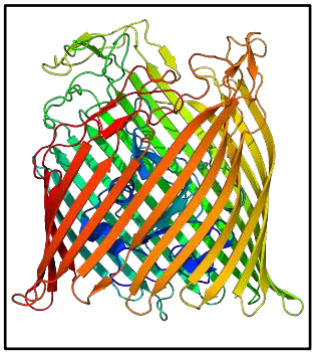


FhuE

This gene has been identified and named previously in *E. coli* as a 22-stranded β-barrel ferric coprogen receptor (FhuE) [116]. In *Pasteurellaceae* the *fhuE* gene had previously only been identified in *M. haemolytica* [31]. Although *M. haemolytica* FhuE has ~34% AAID to FhuE of *E. coli* it does appear to be a true homologue. We have accordingly designated the *M. haemolytica* gene/protein FhuE2, reserving FhuE1 for *E. coli*. FhuE2 appears to be produced exclusively in Genotype 1 *M. haemolytica* isolates, with all isolates containing the gene, which is nearly always (93%) functional. Genotype 2 *M. haemolytica* isolates did not have the *fhuE2* gene. 39% of *G. parasuis* genomes also contained a *fhuE2* gene, though none were predicted to be functional. The presence of the truncated gene in *G. parasuis* genomes was strongly

associated with the *fcmA5* and *irpA5* genes, and had an inverse relationship with the *fup2* gene, which is also truncated.

Ferric coprogen receptor H (FhuH)
Structure: β-barrel
Molecular weight: ~80kDa
Example Protein Sequence: HaeH_FhuH1.2
Found in: *H. influenzae*, *H. haemolyticus*
No. of Alleles: One
Variable expression: No.
Locus: *fhu* locus
-> *fhuC* (51% ID to *E. coli* FhuC) [~620bp]
-> *fhuD* (29% ID to *E. coli* FhuD) [~880bp]
-> *fhuB* (38% ID to *E. coli* FhuB) [~1970bp]
-> *fhuH* (20/22/20% ID to *E. coli fhuA/E/I*) [~2170bp]



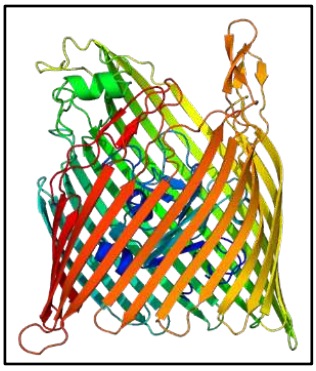
FhuH

Prior to this publication, this gene/protein hadn't been formally named in *Pasteurellaceae*. FhuH is a 22-stranded β-barrel. An orthologue of this gene/protein has been identified as being involved in iron uptake in *Bordetella spp.* - named *Bordetella* ferric complex receptor B (BfrB) by *Parkhill et al.* [117]. The second B in BfrB indicates the order in which it was found, and not a locus position, which doesn't adhere to historic naming conventions [41]. To better adhere to nomenclature guidelines, we propose calling this gene/protein *fhuH*/FhuH, as it is the eighth (thus, H) gene associated with the *fhu* locus.

Bordetella spp., *H. influenzae* and *H. haemolyticus* are respiratory pathogens, which may explain the shared presence of this gene. *fhuH* has been identified in NTHi as an iron uptake receptor associated with the *fhu* locus but remained unnamed (referred to as r2846.1777) in that publication [89]. FhuH has low Amino Acid Identity (AAID) to other *fhu* locus receptors of *E. coli*; FhuA (~20% AAID) and FhuE (~22% AAID). However, FhuA and FhuE have a low ~20% AAID. The previously established role of the *fhu* locus in hydroxamate siderophore uptake and metabolism [113, 118] heavily supports this protein's role as a siderophore receptor directly involved in iron uptake.

fhuH was only seen in *H. influenzae* and *H. haemolyticus* genomes. FhuH from *H. influenzae* and *H. haemolyticus* shared 75% AAID (**Supplementary File 11**). *fhuH* was notably absent from typeable *Haemophilus influenzae* isolates but had a 30% gene presence in NTHi and 26% in *aegyptius* strains (this study). Further, where *fhuH* was present, it was nearly always predicted to be functional. Aside from NTHi, *fhuH* was observed in one *H. haemolyticus* genome – strain 2019-19. This isolate was previously confirmed to be *H. haemolyticus* and not a case of mistaken identity of *H. influenzae* [119].

Ferric coprogen receptor I (FhuI)
Structure: β-barrel
Molecular weight: 78kDa
Example Protein Sequence: ActP_FhuI2
Found in: *A. pleuropneumoniae* (3), *A. paragallinarum* (2), *G. anatis*, *G. parasuis*, *H. parainfluenzae*
No. of Alleles: Six
Variable expression: No
Locus: *fhu (*A. pleuropneumoniae) *ompU* 11kDa from fhu *locus*
-> *fhuC* (53% ID to *E. coli* FhuC) [~470bp]
-> *fhuD* (27% ID to *E. coli* FhuD) [~950bp]
-> *fhuB* (38% ID to *E. coli* FhuB) [~1950bp]
-> *fhuI* (30/20% ID to E. coli FhuA/E) [~2020bp]

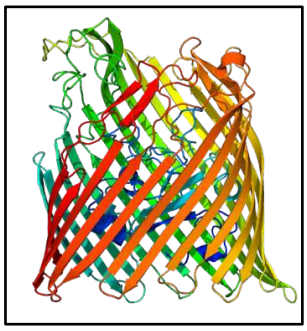


FhuI

FhuI has been reported as FhuA in *G. parasuis* [74], though (as explained in the *above* '*fhuA*' section) it has relatively low sequence homology (~30%) to FhuA of either *N. meningitidis* or *E. coli*. FhuI has also been previously named CirA in *G. anatis* [68]. As both alternate names are already in use by other, heterologous β-barrels involved in iron uptake [120-122] we will continue to use *fhuI*/FhuI to describe this gene/protein. This gene is also a part of the well-established *fhu* locus, so the 'fhu' prefix is appropriate, and the next alphabetical suffix was used (I). Much like FhuH, FhuI expands the known members of the *fhu* locus.

All *A. paragallinarum* and *A. pleuropneumoniae* genomes had a *fhuI* gene, with ~90% having a functional *fhuI* gene. The high gene presence suggests *fhuI* is important to these organisms. The other organisms with *fhuI* had low gene presence and/or functionality rates (*G. parasuis*; 64% presence, 28% functional, *G. anatis;* 11% presence, 100% functional, *H. parainfluenzae;* 6% presence, 100% functional). Within *A. pleuropneumoniae* genomes three *fhuI* alleles were found, one of which was shared with *G. parasuis* (*fhuI1*). Similarly, *A. paragallinarum* and *G. anatis* share a *fhuI* allele (*fhuI5*).

Ferrioxamine receptor (FoxA)
Structure: β-barrel
Molecular weight: 78kDa
Example Protein Sequence: ManH_FoxA
Found in: *M. haemolytica*
No. of Alleles: One
Variable expression: ON/OFF phase-variable in *M. haemolytica* (Poly A – AAAAAAAAGATG)
Locus: Undetermined



FoxA

This gene/protein was previously identified in *M. haemolytica* as FhuA in [31]. Since FhuA is already in use as a placeholder for many genes/proteins (see **The false *fhuA* gene** section), a new name was needed. This protein shares ~35% AAID with the previously identified FoxA of *Yersinia enterocolitica* and does appear to be homologous [123]. FoxA also shares 33% AAID with FhuA of *E. coli* [123], appearing to be related. Other Fhu β-barrels also such as FhuE and FhuI share homology to this protein (data not shown). Without more information 'FoxA' was adopted, though this gene/protein may need improved nomenclature in future.

Of the organisms examined in this study, the *foxA* gene was only found in *M. haemolytica*, with all genomes containing the gene and ~70% of genes predicted to be functional. However, when *M. haemolytica* isolates are separated into Genotypes all Genotype 1 isolates have a functional gene whereas only ~50% of Genotype 2 genomes have a functional gene.

foxA can further be observed in *Mannheimia granulomatis* (cattle, deer, rabbits), *Pasteurella bettyae* (humans, cats, dogs), *Neisseria dentiae* (cows), *Pelistega ratti* (rats) and *Lonepinella koalarum* (koalas). Although *P. bettyae* is a human opportunistic/zoonotic pathogen in the *Pasteurellaceae* family, no fully annotated genomes were available for examination and therefore could not be confidently included in this study. Since the bacteria that have a *foxA* gene inhabit a variety of hosts, it may be that FoxA plays an important role in broad iron acquisition in mammals.

FcmA (Ferric Metabolism Protein A)
Structure: β-barrel
Molecular weight: 72kDa
Example Protein Sequence: ActP_FcmA1
Found in: *A. actinomycetemcomitans, A. pleuropneumoniae.* (3), *G. parasuis*, *H. haemolyticus* (truncated), *H. parainfluenzae*, *H. somni* and *M. haemolytica*
No. of alleles: Six
Variable expression: ON/OFF Phase-variable in *H. somni* (Poly G – TAGGGGGGGGGGT)
Locus: Undetermined



FcmA

This protein has been under-reported in *Pasteurellaceae* but has been identified in *A. pleuropneumoniae* as FrpB [59]. This protein was first identified in *Moraxella catarrhalis* and named 'outer membrane protein B' (CopB) [124], then described in *N. meningitidis* as Fe-regulated protein B (FrpB) [125, 126]. Though referred to as FrpB in *N. meningitidis*, an orthologue in *N. gonorrhoeae* was shown to function as a receptor for the siderophore enterobactin and was hence named FetA (ferric enterobactin transport) [127]. As noted in the prior study, the binding affinity of FetA to enterobactin was relatively low, indicating that an enterobactin-related siderophore (not enterobactin itself) may be the primary ligand of this protein. It can also be found in *Helicobacter pylori [128].* Neither CopA nor FrpB use appropriate nomenclature [41] as in each case the 'B' represents the order in which each protein was discovered, rather than its position in the locus. FetA is also already in use and involved in iron metabolism [129]. As such, the new name 'Ferric Metabolism Protein A' (FcmA) was assigned.

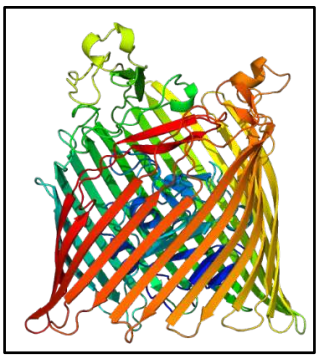
These are a family of 22-stranded β-barrel OM proteins. The many FcmA of *Pasteurellaceae* are ~40-50% identical to *Neisseria* FrpB/FetA proteins. Examining the immunogenicity of each outer loop of the β-barrel of *N. meningitidis* FrpB showed that loop 11 was immunogenic but highly variable, producing strain specific antibodies that shielded other, more stably expressed loops 5 and 3 [130]. *Pasteurellaceae* FcmA don't appear to have the variable (amino acid) loop sequences as seen in *Neisseria* [130], though some *Pasteurellaceae* FcmA appear to phase-vary ON/OFF (an alternative means of immune evasion) as observed in *H. somni fcmA*.

FcmA appears to be important to *A. pleuropneumoniae*, *A. actinomycetemcomitans* and *M. haemolytica*, which have a gene presence and functional rate approaching 100%. *H. somni* also has *fcmA* in most (88%) genomes, all of which were predicted to be functional. The other organisms that have the *fcmA* gene (*G. parasuis*, *H. haemolyticus* and *H. parainfluenzae*)

have a functional gene in less than 50% of isolates examined. The distribution of *fcmA* across *Pasteurellaceae* (found in 7/12 of the organisms examined) suggest it is an important iron uptake factor. The abundance of FcmA is comparable to better known factors TbpA and Hgps and may be suitable as a broad target for vaccines and therapies.

In *A. pleuropneumoniae fcmA* is likely transcribed with the *uvrC* gene. UvrC is required for iron and heme utilization [131]. Additional organisms with *fcmA* include *Nicoletella semolina* (horse host) and *Neisseria shayeganii* (human host).

Ferric uptake protein (FupA)
Structure: β-barrel
Molecular weight: 109kDa
Example Protein Sequence: GalA_Fup3
Found in: *A. pleuropneumoniae*, *A. paragallinarum* (truncated), *G. anatis*, *G. parasuis* (truncated), *H. parainfluenzae*
No. of Alleles: Four
Variable expression: No
Locus: *fup* – variable by organism
H. parainfluenzae
-> *fupA* (β-barrel) [~2200bp]
-> *fupB* (ABC transporter) [~1760bp]



FupA

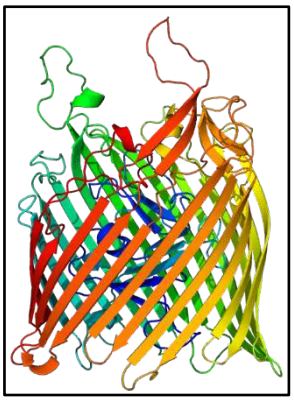
G. anatis
-> *fupA* (β-barrel) [~2200bp]
-> *fupB* (ABC transporter) [~1750bp]

A. pleuropneumoniae
-> *fupC* (ABC transporter substrate-binding protein) [~1000bp]
-> *fupB* (ABC transporter permease) [~1000bp]
-> *fupA* (β-barrel) [~2200bp]

This 22-stranded β-barrel has been previously associated with iron uptake, identified as CirA (APJL_0554 from strain JL03) in *A. pleuropneumoniae* [58]. However, an unrelated *cir* locus, including a β-barrel named *cirA*, has already been identified in *E. coli* [132]. Although these two β-barrels are both involved in iron uptake, they share low AAID (~21%). As such, new names were needed for this receptor and locus. The locus was given the name Ferric uptake protein locus (*fup*), and contains a variety of co-expressed genes, such as *fupB* and *fupC*. In *A. pleuropneumoniae* this locus has a typical iron uptake locus structure – a receptor (*fupA*) accompanied by ABC transport machinery in the form of a permease (*fupB*) and an iron binding protein (*fupC*).

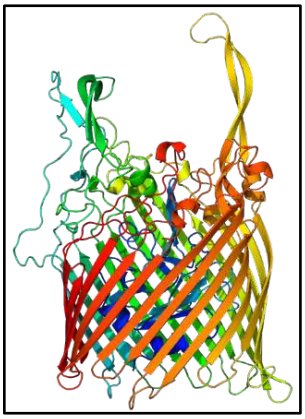
The *fup* gene is relatively common across the *Pasteurellaceae* investigated, though has a low functional rate. 78% of *G. anatis* genomes have a functional *fup* gene. The *fup* gene was present in 94% of *A. pleuropneumoniae* isolates examined, though only predicted to be functional in 40%. Approximately half of *G. parasuis* and *A. paragallinarum* isolates also had a *fup* gene, though none were functional.

Gallibacterium anatis siderophore receptor (GasA)
Structure: β-barrel
Molecular weight: ~74kDa
Example Protein Sequence: GalA_GasA
Found in: *G. anatis*
No. of Alleles: 1
Variable expression: No
Locus: *gas*
-> *gasA* [~2000bp]
-> *gasB* (exbB gene) [~450bp]
-> *gasC* (exbD gene) [~390bp]
-> *gasD* (tonB) [~600bp]


Gas

Identified and investigated previously, though not named, this 22-stranded β-barrel (herein named GasA) was found to be under Fur regulation [68]. This locus was given the name *Gallibacterium anatis* siderophore receptor (*gas*) as it was first described in *G. anatis* [68]. The locus was assigned the '*gas*' prefix and each gene on the locus assigned the suffix (A-D). The genes *gasB-D* encode energy transduction proteins (exbB, exbD and tonB) and are important for internal iron/siderophore processing [133]. As the focus of this study was OM proteins, further classification of the extra genes in the *gas* operon was not attempted. Of the *Pasteurellaceae* examined in this study, *gasA* was only found in *G. anatis* (89% presence, 100% functional). There are orthologues of *gasB-D* in *H. somni*, though it lacks *gasA* - the gene for the receptor. The *gas* locus can also be found in other *Pasteurellaceae* not investigated in this study, such as *Haemophilus pittmaniae, Basfia succiniciproducens*, *Exercitatus varius* and *Mesocricetibacter intestinalis* (data not shown).

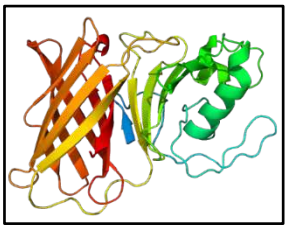
Heme Acquisition System Receptor (HasR)
Structure: β-barrel
Molecular weight: 96kDa
Example Protein Sequence: PasM_HasR
Found in: *A. actinomycetemcomitans*, *P. multocida.*
No. of Alleles: 2
Variable expression: No
Locus: No


HasR

HasR is an immunogenic OM β-barrel protein that functions as a hemophore receptor but can also bind free heme [134-136]. The hemophore associated with HasR is called HasA and has only been reported in genera outside the *Pasteurellaceae*, such as *Serratia* and *Yersinia* [136, 137]. An association has been shown between the hemophore Hemophilin and a HasR orthologue in *H. haemolyticus [81, 138]*. *P. multocida* HasR has been previously identified [134]. Prior to this study, *hasR* had also been identified in *A. actinomycetemcomitans* [45]. HasR is likely important to *A. actinomycetemcomitans* as all strains examined had a functional *hasR* gene (**Supplementary File 1**). Though unreported, the *hasR* gene can also be found in some clinically significant *Campylobacter* species, such as *C. sputorum* (livestock) and *C. ureolyticus* (data not shown).

Only 63% of *P. multocida* isolates contained the *hasR* gene (97% predicted to be functional). However, important trends can be observed when separating *P. multocida* isolates by host origin: *hasR* was present and predicted to be functional in all isolates from pigs and rabbits. Further, *hasR* was present in 87% of isolates from cows, with 93% of these predicted to be functional. Conversely, bird isolates have just 25% gene presence, while human isolates had 64%. This interesting spread of gene distribution suggests *hasR* plays an important role in determining host susceptibility to a strain of *P. multocida.* This may be valuable information for vaccine development and highlights the importance of including factors such as host origin of the strain under investigation.

Hemophilin
Structure: one-lobed β-strand protein
Molecular weight: 29kDa
Example Protein Sequence: HaeH_Hemophilin
Found in: *A. paragallinarum*, *H. haemolyticus*, *H. influenzae* and *P. multocida*
No. of Alleles: Three
Variable expression: No
Locus: *hemophilin*
Avibacterium paragallinarum
-> *hhr* [~3000bp]
-> *hemophilin* [~780bp]
-> *hex* [~1450bp]


Hemophilin

Hemophilin is a hemophore first described in *H. haemolyticus* [81]. Hemophilin can inhibit the growth of *H. influenzae,* which was initially thought to lack the appropriate Hemophilin receptor [83]. More recently (2024), hemophilin homologues, as well as the hemophilin export machinery and receptor, were described in many other bacteria, including *H. influenzae* [138]. A putative iron uptake β-barrel surface protein was also described and modelled in the initial publication [81]. Discussed later, we have given the hemophilin receptor the name ‘Hhr’ and the associated exporter the name ‘Hex.’

Apart from *H. haemolyticus*, in which hemophilin was expected to be present, Hemophilin was rarely found in *H. influenzae* (two isolates), *A. paragallinarum* (one isolate) and *P. multocida* (three isolates). It has been suggested that Hemophilin may prove valuable as a therapy option given its ability to sequester iron away from *H. influenzae* [83]. However, this relies on *H. influenzae* not expressing the hemophilin receptor Hhr, which is not always the case with isolates.

Hex (Hemophilin Exporter)
Structure: Surface Lipoprotein Assembly Modulator (SLAM)
Molecular weight: 57kDa
Example Protein Sequence: AviP_Hex2
Found in: *A. paragallinarum, H. haemolyticus, H. influenzae, P. multocida*
No. of Alleles: Three
Variable expression: No
Locus: *hemophilin*
Avibacterium paragallinarum
-> *hhr* [~3000bp]
-> *hemophilin* [~780bp]
-> *hex* [~1450bp]


Hex

The hemophilin exporter (Hex) is a SLAM expressed with Hemophilin and the hemophilin heme receptor. Hex is a relatively unique SLAM that doesn't group phylogenetically with the other SLAMs herein. Any organism that produces hemophilin contains at least one *hex* gene.

Heme-hemopexin Utilisation Proteins A, B, C and D (HxuA, B, C, D)
Structure: (See Figure 14) **HxuA** - right-handed β-helix, **HxuB** – exporter (two-partner secretion system), **HxuC** – 22-stranded β-barrel, **HxuD** - right-handed β-helix
Molecular weight: HxuA: 101kDa, **HxuB**: 64kDa, **HxuC**: 81kDa, **HxuD**: 141kDa
Example Protein Sequence: HaeI_HxuA, HaeI_HxuB, HaeI_HxuC, AviP_HxuD
Locus found in: *A. paragallinarum, G. anatis, G. parasuis, H. influenzae, M. haemolytica*
No. of Alleles: HxuA: 3, HxuB: 4**, HxuC:** 4**, HxuD:** 1

Organism	HxuA	HxuB	HxuC	HxuD
A. paragallinarum		HxuB2	HxuC2	HxuD1
G. anatis		HxuB2	HxuC2	HxuD1
G. parasuis	HxuA2	HxuB3	HxuC3	
H. influenzae	HxuA1	HxuB1	HxuC1	
M. haemolytica	HxuA3	HxuB4	HxuC4	

Table 6: Allelic distribution of HxuA/B/C in examined *Pasteurellaceae*.
Variable expression: HxuA: Tandem repeat domain (AATGGC) in immunogenic loops (*H. influenzae* only), **HxuB:** No, **HxuC:** No, **HxuD:** Not enough information
Locus: *hxu*
A. paragallinarum
<- *hxuD* [~3820bp]
<- *hxuB* [~1700bp]
-> *hxuC* [~2160bp]
-> *hutX* [~510bp]
-> *hutZ* [~510bp]

H. influenzae
-> *hxuC* [~2160bp]
-> *hxuB* [~1700bp]
-> *hxuA* [~2780bp]

Haemopexin is a glycoprotein found in human plasma and bodily fluids. Hemopexin binds heme with a high affinity, forming a heme-haemopexin complex [139]. The *Pasteurellaceae* heme-hemopexin utilisation (*hxu*) locus traditionally encodes three proteins: HxuA, HxuB and

HxuC. This locus has been reported in *G. anatis* [69], *H. influenzae* [87], *G. parasuis* [72], *M. haemolytica* [31, 99] and *H. haemolyticus* [82]. HxuC is an outer membrane 22-stranded β-barrel protein responsible for internalising free heme (from any source) [67]. HxuB is part of a two-partner secretion (TPS) system and is allegedly responsible for HxuA export/assembly. HxuA is exported and tethered to the outer membrane via lipid anchor, where it frees heme from the heme-hemopexin complex. For example, HxuB exports HxuA and HxuA frees heme that can be internalised by HxuC [67, 140, 141]. However, HxuB is a uncommonly unique porin [142], and few investigations have confirmed its role as solely part of a TPS system. There's also evidence that HxuB is not essential to the haemopexin acquisition process [140]. More information about the TPS system can be found elsewhere [143]. Intriguingly, HxuA can be released, or bound to the cell [144].

We did not find the *hxu* locus in *H. haemolyticus* as reported elsewhere [82]. We did find the *hxu* locus in *A. paragallinarum.* To our knowledge it has not been reported previously. The *hxu* locus found in *A. paragallinarum* is virtually identical to that of *G. anatis*, which may be due to their shared host (birds). We have also expanded the *hxu* locus to include *hxuD* – a gene encoding a protein that likely functions as an alternative to *hxuA*. HxuA and HxuD share ~35% AAID and are both a right-handed β-helix protein. It appears organisms either produce HxuA or HxuD, but they always retain the *hxuB* and *hxuC* genes (**Table 6**).

Where the *hxu* locus could be observed in a genus it was present in virtually all isolates, except *G. anatis*, which still ranged above an ~80% gene presence. Genes were generally all functional in these organisms, with two exceptions. *G. anatis* has a reduced number of functional (75%) *hxuC* genes. Comparing Genotype 1 and 2 *M. haemolytica* isolates showed a stark difference in *hxuC* gene functional rates. Despite all Genotype 1 and 2 *M. haemolytica* isolates having a full *hxu* locus, Genotype 2 isolates had a low (11%) rate of functional *hxuC* genes. This contrasts the 100% functional rate of *hxuC* genes in Genotype 1 *M. haemolytica* isolates. Further, the lack of a functional *hxuC* gene appears to correlate with a functional *hup* and *fhuE* gene in Genotype 2 *M. haemolytica.* It's possible that Hup and/or FhuE act as a secondary receptor for HxuA in *M. haemolytica*.

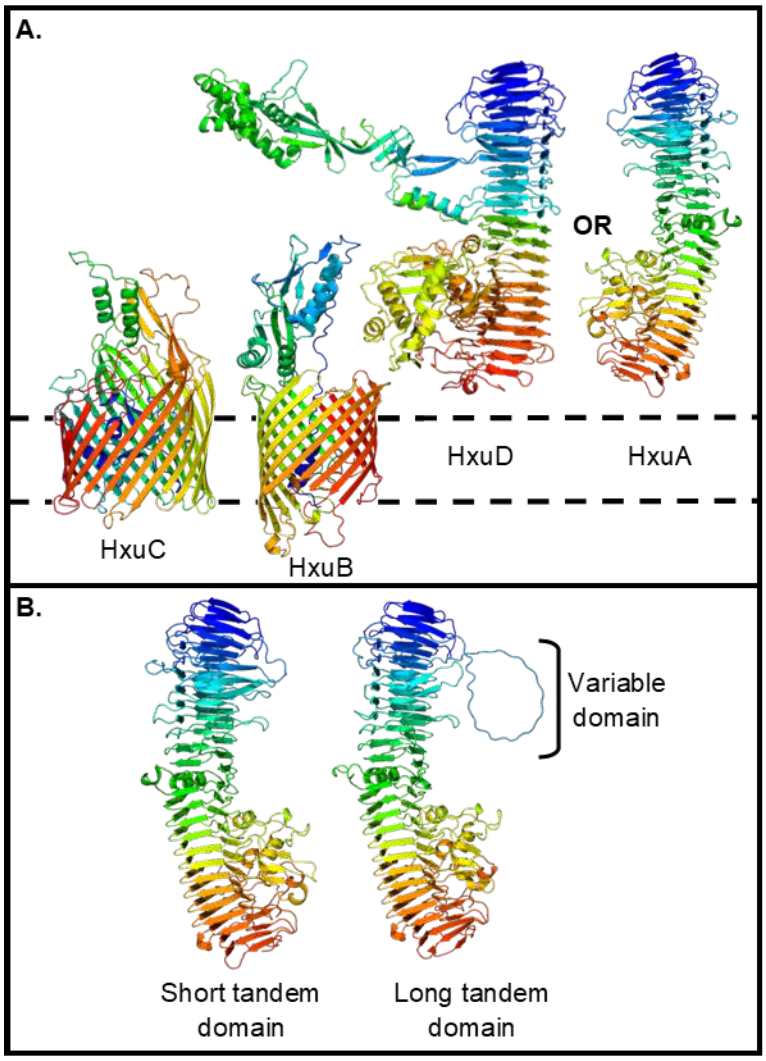


Figure 14: A. HxuA, HxuB, HxuC and HxuD proteins. **B.** Three-dimensional depiction (AlphaFold) of the immunogenic tandem repeat domain of HxuA of *H. influenzae*.

Tandem Repeats in a HxuA Loop Domain

40 out of the 109 *hxuA* loci found in *H. influenzae* strains contained a tandemly repeating domain. This domain expands and contracts, changing the length of a region of HxuA with no known function. The increase or decrease in tandem units doesn't turn the gene ON/OFF, so is not ON/OFF phase-variation. It's also unlikely to be a periscope protein, such as the Fishing Rods, as HxuA is secreted and not bound to the cell surface. However, similar to periscope proteins, this appears to be a case of length–dependent structural variation. **Figure 14 B** shows the structure of *H. influenzae* HxuA and highlights the tandem repeat domain. The exact role of the *H. influenzae* HxuA tandem repeated domain is unknown, however, may be involved in immune evasion. Repeat regions are highly immunogenic [145], so a tandem repeat domain could draw antibodies away from the functional domains of HxuA. Further, immune efforts against functional domains can disrupt protein interactions with target molecules [146]. As HxuA is not associated with the cell, retaining function is more important than immune evasion. A similar strategy has been observed in *Plasmodium falciparum*, which utilises short tandem repeat protein sequences to generate 'decoy' antibodies [147].

Hemin Receptor (HemR)
Structure: β-barrel
Molecular weight: 85kDa
Example Protein Sequence: PasM_HemR
Found in: *H. ducreyi*, *H. influenzae*, *H. somni*, *P. multocida*
No. of Alleles: Three
Variable expression: ON/OFF phase-variable in *H. somni* (Poly-C; TTCCCCCCCCCTT)
Locus: No.

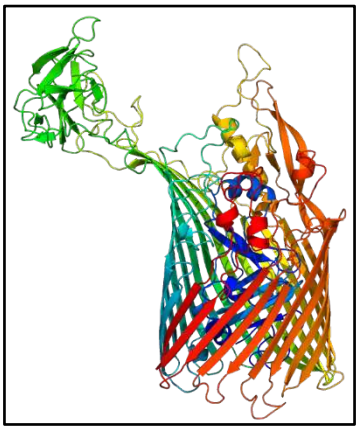


HemR

The *Pasteurellaceae* hemin receptor HemR is an OM-associated TonB-dependent 22-stranded β-barrel protein that binds and extracts host heme [148]. HemR appears to be an orphaned β-barrel, with no associated iron uptake machinery or locus. A prior investigation reported *hemR* in 87% of commensal NTHi isolates, and the higher rate of 99% in invasive isolates [149]. 47 amino acid polymorphisms have also been described in Non-typeable *H. influenzae* (NTHi) HemR, with some more associated with a virulence outcome [148]. These polymorphisms are not the same as the different alleles identified herein. HemR has also been previously identified in *P. multocida* [102, 105], *H. haemolyticus* [82] and in *H. ducreyi* (named therein TdhA) [75]. We have additionally identified *hemR* in *H. somni* (this study).

The *hemR* gene is relatively abundant in the genus where it can be observed; *H. ducreyi, H. influenzae* and *H. somni* have a high rate (100%, 97% and 92% respectively) of *hemR* gene presence. Superficially, *P. multocida* has a *hemR* gene presence of ~70%, however isolates from swine and rabbit hosts have 100% gene presence, with all *hemR* predicted to be functional. Further, *P. multocida* isolates from human and avian hosts have gene presence rates of ~55% and 45% respectively. This suggests that *hemR* may be more important during infection of swine and rabbits than during infection of human and avian host. This information improves our understanding of transmission between different hosts.

Haemoglobin-Haptoglobin Binding Proteins (Hgps)
Structure: β-barrel
Molecular weight: 118kDa
Example Protein Sequence: HaeI_Hgp1.1
Found in: *A. pleuropneumoniae,* *A. actinomycetemcomitans, H. ducreyi, H. haemolyticus* (4), *H. influenzae* (11), *H. somni, P. multocida* (5)
No. of Alleles: 14
Variable expression: ON/OFF phase-variable in *H. somni* via poly C tract (TTCCCCCCCCCTC), *H. influenzae* via tandem repeat tract (CCAA) and *P. multocida* via poly A tract (AAGAAAAAAAA)
Locus: Variable



Hgp

Organism	*hgp1*	*hgp2*	*hgp3*	*hgp4*	*hgp5*	*hgp6*	*hgp7*	*hgp8*	*hgp9*	*hgp10*	*hgp11*	*hgp12*	*hgp13*	*hgp14*
A. pleuropneumoniae										Hgp10				
A. actinomycetemcomitans	Hgp1.3													
H. ducreyi									Hgp9					
H. haemolyticus	Hgp1.1	Hgp2	Hgp3				Hgp7.2							
H. influenzae	Hgp1.1 (up to 2), Hgp1.2, Hgp1.3	Hgp2 (up to 3)	Hgp3 (up to 3)	Hgp4	Hgp5 (up to 2)	Hgp6	Hgp7.1, Hgp7.2	Hgp8						
H. somni														Hgp14
P. multocida						Hgp6		Hgp8			Hgp11 (up to 2)	Hgp12	Hgp13	

Table 7: Distribution of Hgp alleles in examined *Pasteurellaceae*.

Haemoglobin-haptoglobin binding proteins (Hgps) are TonB-dependent 22-stranded β-barrel proteins that bind and extract heme from haemoproteins such as the haemoglobin-haptoglobin (Hb-Hp) and myoglobin-haptoglobin (Mb-Hp) complex [150, 151]. Hgps use ABC transporter machinery to chaperone internalised heme. Hgps are common to *Pasteurellaceae* but can rarely be found in non-*Pasteurellaceae* species. Hgps have been identified in *A. actinomycetemcomitans* [152], *A. pleuropneumoniae* [42, 56, 153], *H. ducreyi* [47], *H. haemolyticus* [82], *H. influenzae [43, 154, 155], H. somni* [156] and *P. multocida* [46, 100].

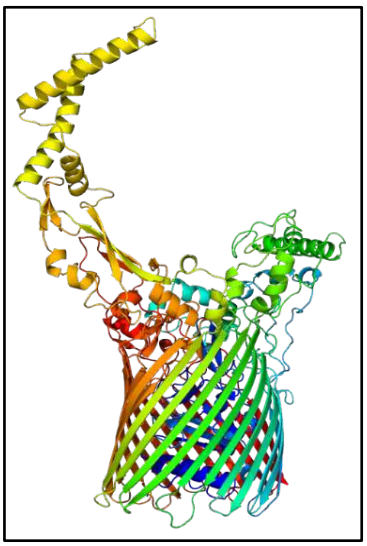
Hgps are generally annotated as Hgp in *H. influenzae [43, 154, 155],* Hg**b** in the Genus *Pasteurella* (including *P. multocida*) [46, 50] and have further alternate nomenclature, such as *hupA* in *H. ducreyi* [47]. For brevity, in this publication haemoglobin-haptoglobin binding proteins (and proteins with significant homology of >40%) will be referred to as Hgps. Further, there was a need to rectify issues with prior nomenclature. Prior to this publication, Hgps have acquired a suffix based on the order in which they were discovered. For example, the first Hgp found was named HgpA [43]. This goes against historic naming conventions [41], and is confusing when comparing Hgps from different strains/species. For example, HgpA in *H. influenzae* is not HgpA of *A. pleuropneumoniae* but *is* homologous to HgpA of *A. actinomycetemcomitans*. We have provided a new, sensible naming system for Hgps, which also accommodates variable sub-alleles within alleles. With this new system, we have identified 14 allelic variants of Hgps across *Pasteurellaceae* examined. The identification of these is important to vaccine/therapy design as they are likely all immunogenically unique.

Our study shows Hgps are particularly abundant in *H. influenzae, H. haemolyticus* and *P. multocida* (**Table 7**). *H. influenzae* isolates contained 2-6 unique *hgp* genes from eight unique alleles (and three sub-alleles*). H. haemolyticus* contained 1-2 *hgp* genes from four alleles. *P. multocida* contained 1-3 unique *hgp* genes from a mix of five alleles. The increased number and levels of redundancy of *hgp* genes suggests these genes have a key importance to these organisms. *hgp* genes appear important to all organisms with the gene, except *A. actinomycetemcomitans*; despite all isolates of *A. actinomycetemcomitans* containing an *hgp* gene, it was only predicted to be functional in 50% of isolates. This appears to be the exception, as the remaining organisms with an *hgp* gene (*H. ducreyi*, *A. pleuropneumoniae* and *H. haemolyticus*) all contained an *hgp* gene, of which virtually all were functional. Hgps are clearly an understudied and underreported iron uptake factor across *Pasteurellaceae*.

Hgps are sometimes expressed in an ON/OFF phase-variable manner, containing tandem repeats shortly after their start codon [155, 157]. All *hgps* in *H. influenzae* appear to be phase-variable via (CCAA) tandem repeat tract at the 5′ end of the gene. This means that although all *H. influenzae* genomes contain at least two *hgp* genes, it may not express these genes. More likely is that it alternates these generationally to inhibit host immune efforts. Further phase-variable expression can be observed in *H. somni hgps*, which contains a 5′ poly C tract**,** and *hgp11* of *P. multocida* which contains a poly A tract around the middle of the gene.

Hgps phase-vary during *H. influenzae* COPD infection, which suggests the ability to phase-vary is a virulence determinant [158]. Prior work revealed a single *H. influenzae* strain contains up to six phase-variable *hgp* genes, with strains regularly having non-identical copies of *hgpB* and *hgpC [29]*. Separate to the allelic variation and phase-variation, the immunogenic outer loops of Hgps have high sequence variability [29], further complicating investigation. Of note, an Hgp orthologue in *Aeromonas hydrophila* was found to offer the highest protection efficacy vs other OM proteins [159]. With little recent study of Hgps in *Pasteurellaceae*, we build on previous findings of Hgps to better characterise this important family of proteins.

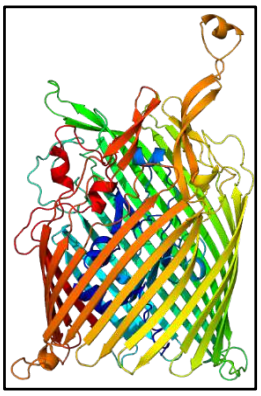
Hemophilin heme receptor (Hhr)
Structure: β-barrel
Molecular weight: 125kDa
Example Protein Sequence: PasM_Hhr
Found in: *A. paragallinarum*, *H. haemolyticus* (2), *H. influenzae, H. parainfluenzae*, *P. multocida* (2)
No. of Alleles: Four
Variable expression: poly A tract at 3′ end of gene in *H. parainfluenzae* (TAGAAAAAAAAAGT) and *P. multocida*.
Locus: *hemophilin*. Also found orphaned as a standalone gene.
Avibacterium paragallinarum
-> *hhr* [~3000bp]
-> *hemophilin* [~780bp]
-> *hex* [~1450bp]


Hhr

Prior literature on Hhr is scarce. Though not named, Hhr was modelled and mentioned briefly in the initial publication describing hemophilin [81]. Hhr is an OM-associated 22-stranded β-barrel that is usually expressed with hemophilin and the hemophilin exporter. However, Hhr can be observed as a standalone gene in *P. multocida* and *H. parainfluenzae* and lacks the typically associated *hemophilin* and *hex* genes. β-barrels involved in iron uptake can broadly bind and uptake multiple iron sources. It's likely Hhr is beneficial as a standalone protein and functions as a miscellaneous iron receptor for factors like heme. The *hhr* gene may be phase-variably expressed in *P. multocida* and *H. parainfluenzae* via poly-A tract. The poly-A tract is at the 3′ end of the gene, which results in a truncated protein, rather than the typical ON/OFF switching where a poly-nucleotide tract or tandem repeat tract is at the 5′ (start) of the gene. *In vitro* analysis is needed to confirm whether this poly-A tract affects gene expression across generations. It does suggest Hhr is under immune pressure, which is not an uncommon observation regarding iron uptake factors in these organisms.

Hhr was rarely observed in the *Pasteurellaceae* examined, with two exceptions: in *H. haemolyticus* and *P. multocida* isolates from humans. Most (80%; 4/5) *H. haemolyticus* isolates had the full hemophilin locus (with *hhr*). The single isolate without a *hhr* lacked the entire hemophilin locus. *P. multocida* appears to have a more complicated relationship with *hhr*. Unlike with *H. haemolyticus*, which has one allele or another in the same genetic location, *P. multocida* isolates contained two *hhr* genes at separate genetic locations. One is far more abundant and associated with *irpA4* and *hmbR1* genes. The other location contains *hemophilin*, hemophilin exporter (*hex*), a separate *hhr* and an *hgp* gene. The 'full' *hhr* (hemophilin) locus was only observed in two pig and one rabbit isolate of *P. multocida*. No *P. multocida* isolates with the hemophilin locus also had a standalone *hhr* gene. However, there were only three isolates to examine with the full hemophilin locus, and the standalone gene was only observed in ~44% of isolates. It's possible isolates with both *hhr* genes would be seen in a larger sample pool. The standalone *hhr* gene was overrepresented in *P. multocida* isolates from humans – with 73% of human isolates containing the gene (*hhr1*). This is in stark contrast to avian and rabbit isolates, which contained 28% and no (0%) *hhr* genes, respectively. The standalone *hhr* gene may provide a virulence advantage to *P. multocida* in human infection.

Haemoglobin Receptor (HmbR)
Structure: β-barrel
Molecular weight: 86kDa
Example Protein Sequence: PasM_HmbR1
Found in: *A. paragallinarum*, *A. pleuropneumoniae* (truncated), *H. parainfluenzae* (4), *M. haemolytica*, *P. multocida*
No. of Alleles: Eight
Variable expression: ON/OFF phase-variable via tandem T tract at 5′ end of gene in *A. paragallinarum* (GCGTTTTTTTGAA) and M. haemolytica (TCTAAAAAAAAATTG).
Locus: *hmbR*



HmbR

HmbR was first described as a β-barrel OM haemoglobin receptor in *Neisseria meningitidis* [160]. It was later found to be ON/OFF phase-variably expressed in *N. meningitidis* [161]. HmbR phase-varied ON is virulence associated in *N. meningitidis* outbreaks [162]. Little research has been conducted on the HmbR proteins of the *Pasteurellaceae* family, which share >40% amino acid identity to meningococcal HmbR. Some members of the *Pasteurellaceae* family have multiple *hmbR* genes, such as *M. haemolytica*, genomes of which have been reported to have two *hmbR* genes – *hmbR1* and *hmbR2* [99]. HmbR has been identified in *A. pleuropneumoniae* [163]. Another study observed a *hmbR* gene in *P. multocida*, but only referred to it as PM0741 [105].

All *hmbR* genes in *A. pleuropneumoniae* were truncated. This is one of a few instances where a gene is present in all isolates of an organism, but none are predicted to be functional. Although some instances of *hmbR* are phase-variable, this doesn't appear to be the case with the truncated *A. pleuropneumoniae hmbR* genes.

All *A. paragallinarum* isolates contained *hmbR* which was functional in 91% of cases. All *M. haemolytica* isolates had a *hmbR* gene, which was functional. *M. haemolytica* strain NCTC10208, which was an isolate from a lamb host, contained a unique *hmbR* gene (*hmbR8*). Nearly all *P. multocida* isolates had a *hmbR* gene, which was functional in 89% of isolates.

A previously unidentified *tbpB* gene is often found beside *hmbR* in *H. parainfluenzae* and the truncated *hmbR* in *A. pleuropneumoniae* (two *tbpB* genes). If these two proteins have a similar relationship to TbpB and TbpC, which are structurally similar, the Fishing Rod feeds a target (such as heme) to HmbR. In *P. multocida* the *hmbR1* gene sits on a locus between *irpA4* and *hhr1*. This means there are three situations where the *hmbR* gene can be observed; as a standalone gene, as part of a small locus with coexpressed *tbpBs* and as part of a larger locus with two other β-barrels. The large number of HmbR alleles, its phase-variable expression in multiple organisms and its distribution across hosts suggest HmbR is immunogenic and important to *Pasteurellaceae* survival.

Heme Utilisation Protein (Hup)
Structure: β-barrel
Molecular weight: 106kDa
Example Protein Sequence: HaeH_Hup
Found in: *A. actinomycetemcomitans*, *G. parasuis*, *H. haemolyticus*, *H. influenzae*, *H. parainfluenzae*, *H. somni* (2), *M. haemolytica*, *P. multocida*
No. of Alleles: Six
Variable expression: ON/OFF phase-variably expressed in *H. somni* (both alleles; poly-G; GTGGGGGGGG, ATGCGGGGGGGGT)
Locus: Variable (e.g. expressed with *miu* in *M. haemolytica, ompM* in *P. multocida*)



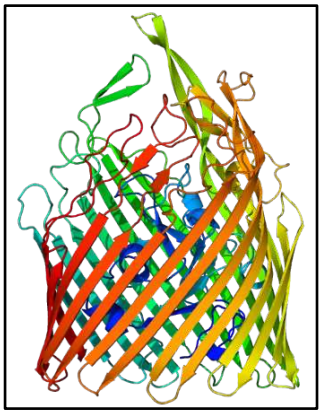
Hup

Hup was first described in *H. influenzae* as a ~100kDa 22-stranded β-barrel protein that sequesters iron from the haemoglobin complex or free heme [85]. Hup has been shown to be a virulence determinant and core protein to *H. influenzae* [82, 88]. Hup has orthologues in other *Haemophilus* species, such as *H. parainfluenzae* and *H. haemolyticus* [82]. A *hup* orthologue referred to as 'CirA' was identified in Genotype 2 *M. haemolytica* [31]. However, it appears *hup* has a wide distribution in *Pasteurellaceae*, with poorly or unreported orthologues in *G. parasuis*, *M. haemolytica*, *H. somni* and *A. actinomycetemcomitans* (this study).

There is significant variability in *hup* distribution among genera. For example, no *M. haemolytica* Genotype 2 isolates had a functional *hup* gene, whereas Genotype 1 isolates do. Further, when sorting *P. multocida* by host origin, *hup* is present in just 8% of avian isolates vs 86% of rabbit isolates. The high gene prevalence across multiple genera, the presence of multiple (six) alleles, the ability of *hup* to undergo phase-variable ON/OFF switching, and that more than one allele can be observed in *H. somni* indicate Hup is immunogenic and important to *Pasteurellaceae*.

The previously unreported *hup* gene in *P. multocida* is associated with three hTbpB genes (named *plpE*, *plpD* and *plpF* herein) and the SLAM *ompU7*.

Iron Responsive Element A (IreA)
Structure: β-barrel
Molecular weight: 76kDa
Example Protein Sequence: HaeH_IreA2
Found in: *H. haemolyticus*
No. of Alleles: One (Plus one in *E. coli*)
Variable expression: No
Locus: *ire/rup.*
-> *ireX* (ABC transporter) [~1000bp]
-> *ireX* (ABC transporter) [~1000bp]
-> *ireX* (Siderophore ABC transporter ATP-binding) [~800bp]
-> *ireX* (Iron-siderophore binding protein) [~900bp]
-> *rup* [~2080bp]
-> *ireA* [~3080bp]

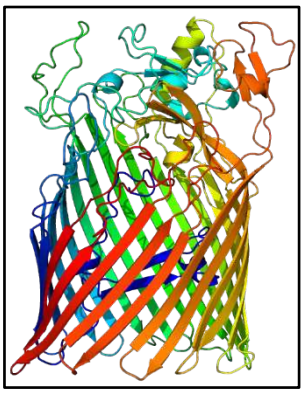


IreA

H. haemolyticus has an orthologue of the previously described *E. coli ireA* gene [164, 165]. In *E. coli*, *ireA* is flanked by a transposase and *E. coli* genomes can have multiple copies of this gene. This suggests *ireA* may be a transposable element. Previous study has highlighted this

protein as a likely siderophore receptor involved in iron uptake and as a virulence factor [165, 166]. The *E. coli* IreA protein has high (48%) AAID to *H. haemolyticus* IreA (data not shown), so likely serves a similar function, though it has not been investigated in *Pasteurellaceae*. We have designated ireA of *E. coli ireA1,* as it was the first instance where this was named, and the orthologue found in *H. haemolyticus ireA2*. The *ireA* gene also appears to sit on a larger locus in *H. haemolyticus*. There's little prior information regarding the *ireA* locus. It was not within the scope of this study to investigate the original IreA of *E. coli* to tease out a locus, so we have temporarily, informally named the genes proximal to IreA and Rup "IreX."

Iron Regulated Protein (IrpA; PM0745/PM1081/PM1428)
Structure: β-barrel
Molecular weight: 78kDa
Example Protein Sequence: ActP_IrpA1
Found in: *A. pleuropneumoniae*, *G. parasuis* (2), *H. somni*, *M. haemolytica* (2), *P. multocida* (3)
No. of Alleles: Eight
Variable expression: ON/OFF phase-variable expression via 5′ poly A tract in *M. haemolytica* (ATGAAAAAAAATTA), poly G in *H. somni* (ACAGGGGGGGATG) and tandem repeat tracts (AACAACAG$_n$ and TC$_n$) in *G. parasuis*.
Locus: Associated with *hTbpB* and *ompM* in *A. pleuropneumoniae* and *M. haemolytica*.

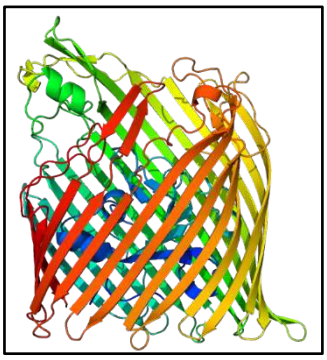


IrpA

IrpAs are an underreported family of proteins found in multiple members of the *Pasteurellaceae* family. They have also been identified in *N. meningitidis* [167]. This is a family of homologous 22-stranded β-barrel proteins, named as phase-variable iron regulated protein (Irp) in *M. haemolytica* [97]. IrpAs are fur regulated and upregulated in iron restricted growth conditions in *P. multocida* [104, 168]. Herein, we have adopted the name *irpA* because it is the first gene in a larger locus in *A. pleuropneumoniae* (*irpA1 -> hTbpB_FR6 -> ompM8*). IrpA7 and IrpA8 (described by their gene identifier PM1081 and PM1428) were found to bind haemin and haemoglobin in *P. multocida* [105]. In *N. meningitidis* the Irp orthologue is called ZnuD, and shares >40% amino acid identity to IrpAs of the *Pasteurellaceae* family [169] so it falls within the same protein family. ZnuD of *N. meningitidis* is primarily associated with zinc uptake, though the *irp* locus of *P. multocida* is Fur regulated, not Zur regulated like in *N. meningitidis*. Iron may be a secondary target of ZnuD in *N. meningitidis*, as it binds zinc and heme when expressed in *E. coli* [167]. In other bacteria, heme uptake systems can moonlight as zinc uptake systems, such as HmuR from *Y. pestis* [170], supporting this possibility.

When *irpA* was present in organisms, at least one *irpA* tended to be functional in all isolates. In cases where organisms expressed more than one *irpA* gene (up to three in *P. multocida*), they often expressed all *irpA* genes. However, *irpA4* in *Pasteurella multocida* isolates from cow and pig hosts tended to be present but truncated. IrpA appears to be highly important to *Pasteurellaceae* that infect pigs and cows. IrpA was absent from *Pasteurellaceae* that infected humans and birds (i.e. The *Haemophilus* genus, *A. actinomycetemcomitans*, *G. anatis* and *A. paragallinarum*). IrpA may be a target of interest to the pig and beef industries.

Iron Uptake Protein (Iup)
Structure: β-barrel
Molecular weight: 81kDa
Example Protein Sequence: AviP_Iup
Found in: *A. paragallinarum*
No. of Alleles: One
Variable expression: No
Locus: Unnamed
-> *hemA* (Glutamyl-tRNA reductase) [~1300bp]
<- Siderophore reductase [~520bp]
-> *iup* [~2110bp]

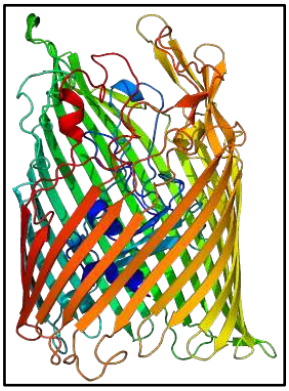


Iup

Iup was undocumented in *Pasteurellaceae*, likely because it was only observed in *A. paragallinarum*. It remains unclear whether *iup* is part of a larger operon. An orthologue has been documented in *N. meningitidis* but was named FhuA [115]. This is discussed in detail later in the 'false *fhuA*' section. Briefly, there are too many heterologous β-barrels named FhuA: FhuA of *N. meningitidis* [115], FhuA of *A. pleuropneumoniae* [171], FhuA of *G. parasuis* [74] and FhuA of *E. coli* [113]. To start to address this issue, which is in the scope of this study, a new name was given to this gene/protein – **I**ron **U**ptake **P**rotein (Iup). This can be adjusted with an appropriate suffix should it be shown to be part of a larger operon.

The *iup* gene has two proximal genes: a siderophore reductase and *hemA*. HemA is involved in iron homeostasis [172]. Siderophore reductases are involved in iron metabolism [173]. Since these genes sit proximal to *iup*, it's likely Iup plays a role in iron metabolism. As a 22-stranded β-barrel, Iup also has homology to other iron uptake factors.

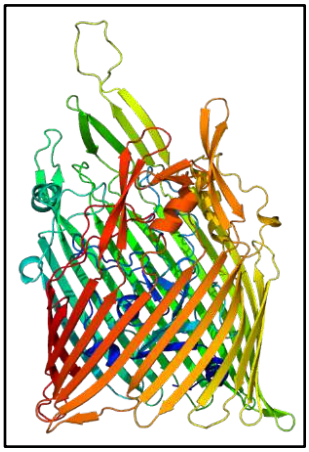
Mannheimia iron uptake protein (Miu)
Structure: β-barrel
Molecular weight: 80kDa
Example Protein Sequence: ManH_Miu
Found in: *M. haemolytica*
No. of Alleles: One
Variable expression: No
Locus: Expressed with *hup6*



Miu

Miu, named herein, was only detected in *M. haemolytica*. Miu is an OM 22-stranded β-barrel. The *miu* gene is present in all *M. haemolytica* isolates, functional in all Genotype 1 isolates and 85% of Genotype 2 isolates. The *miu* gene sits adjacent to the *hup6* gene. As *hup* has a confirmed association with iron uptake [85] it is highly likely *miu* is also involved in iron acquisition, albeit limited to *M. haemolytica*.

Nutrient Binding Receptor A (NbrA)
Structure: β-barrel
Molecular weight: 92kDa
Example Protein Sequence: PasM_Nbr1
Found in: *A. actinomycetemcomitans, A. paragallinarum, H. influenzae, H. somni, P. multocida* (2)
No. of Alleles: Four
Variable expression: ON/OFF phase-variable via 5′ poly C tract in *A. paragallinarum* (ACGTCCCCATA), poly G tract in *H. somni* (CTGGGGGGGGGTA), poly G in one *P. multocida* gene (Nbr1; TGCGGGGGGGGGCCG) and a tandem repeat in *H. influenzae* (AACAACC$_n$).
Locus: *nbr*
-> *nbrA* [~2370bp]
-> *nbrB* (metal protease) [~2770bp]



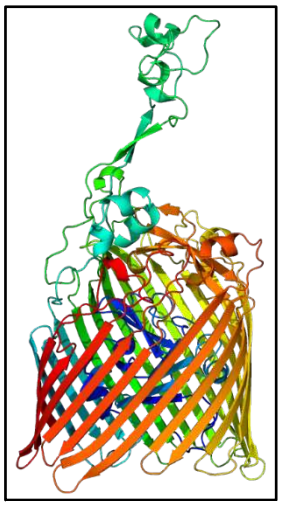
NbrA

These are an understudied and poorly characterised family of homologous (~40-50% AAID) 22-stranded β-barrel proteins that have been previously referred to by their gene identifiers - PM0803 (*P. multocida*) and HI1369 (*H. influenzae*). *PM0803* is upregulated in iron restricted growth conditions of *P. multocida* [103]. HI1369 was identified by *Whitby et al.* [88] as a β-barrel likely involved in iron uptake. PM0803 was similarly described by *Boyce et al.* [103]. *HI1369* was later identified as ON/OFF phase-variable by *Power et al.* [174]. The *nbrA* gene is proximal to a metal degradation protein (metal protease) in the M16 family metallopeptidases. This class of metalloenzyme is associated with iron and zinc uptake [175]. This doesn't directly confirm involvement in iron uptake, but between the structure, Fur regulation [103] and proximity to genes involved in iron metabolism, it strongly suggests NbrA has a role in iron uptake. As NbrA have characteristics typical of other OM iron uptake proteins, we have included it in this study. Outside of the *Pasteurellaceae* family, this protein has orthologues in human commensal and pathogenic bacteria *Suttonella indologenes*, *Shewanella algae* (emerging pathogen) and *Morganella morganii* (data not shown).

The *nbrA* gene was ON/OFF phase-variably expressed in four of the six instances of the gene. The method of phase-variation also changed depending on organism – with poly G tracts in *H. somni* and *P. multocida*, an unusually short poly-C tract in *A. paragallinarum* and a seven-nucleotide tandem repeat tract in *H. influenzae*. As ON/OFF phase-variation provides contingencies against host immune efforts, and most of the *nbrA* genes undergo a form of phase-variable expression, it's highly likely *nbrA* is immunogenic and subject to immune pressure. This is unsurprising, as many *Pasteurellaceae* OM factors involved in iron uptake are similarly prone to phase-variable expression. NbrA was also abundant in organisms where it was found – approaching 100% functional gene presence in most cases. The exception was in *H. influenzae* biogroup *aegyptius* isolates, of which ~30% had a functional *nbrA* gene. The isolates with the *nbrA* gene were all from the same geographical location – Japan. This may be an artifact of the overrepresentation of Japanese *aegyptius* isolates. *P. multocida* isolates from cows also had a somewhat reduced instance of the *nbrA* gene (65%).

The *nbrA* gene was also a rare instance in which a gene was common in *H. influenzae* but no other members of the *Haemophilus* genus. The other example is with the *hxu* locus. NbrA may therefore be a good *H. influenzae*-specific target for vaccines and therapies, though is complicated by the ON/OFF phase-variable expression of the *nbrA* gene. Regardless, NbrA is a highly abundant and important factor for *Pasteurellaceae* isolated from a broad range of hosts.

Outer Membrane Hemin Receptor (OMHR)
Structure: β-barrel
Molecular weight: 88kDa
Example Protein Sequence: PasM_OMHR1
Found in: *A. paragallinarum, P. multocida.*
No. of Alleles: Two
Variable expression: ON/OFF Phase-variable expression via poly-C tract at the 5′ end of the gene in A. paragallinarum (AATCCCCCCAAT).
Locus: No



OMHR

Data relating to OMHR is scarce. OMHR has been described in *A. paragallinarum* and *P. multocida* [104]. OMHR is a 22-stranded β-barrel [168] that binds hemin [105] and was therefore included in this investigation.

The *omhr* gene was present and functional in nearly all isolates of *A. paragallinarum*. However, it is likely susceptible to ON/OFF phase-variable expression in the bird pathogen, which means expression is not stable despite all isolates having the gene. OMHR was also observed in *P. multocida*. Trends could be observed depending on from which host a *P. multocida* isolate was collected. Nearly all genes that were present were predicted to be functional in *P. multocida* isolates. Cow and pig isolates approached 100% gene presence, while human isolates had 55% gene presence and avian isolates had 65% gene presence. This presence in avian hosts did not correlate with a similar presence in *A. paragallinarum*. This discrepancy suggests that the *omhr* gene may have a broader role in iron uptake, which is not necessarily host adapted (e.g. it may bind a variety of different host factors). Alternatively, it may be that there is a benefit for *P. multocida* isolates to not express *omhr* in an avian host – which may be driven by the pre-existing host exposure to the *A. paragallinarum* OMHR allele. Despite only two alleles of OMHR being found in the isolates examined, other alleles exist in *Pasteurellaceae* not examined here, such as *Chelonobacter oris* isolated from tortoise (data not shown).

Surface Lipoprotein Assembly Modulators

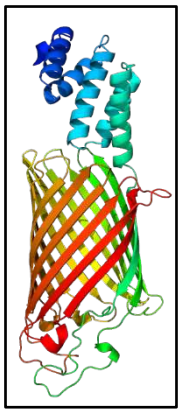
Many iron uptake factors, such as siderophores and 22-stranded TonB-dependent β-barrels, have accompanying outer membrane export machinery called **S**urface **L**ipoprotein **A**ssembly **M**odulator (SLAMs). SLAMs may function as low-affinity iron uptake proteins, however this appears to be a secondary function. SLAMs are typically proximal to their exported target gene in the genome. If the SLAM is removed the function of the associated siderophore (and receptor) is diminished [106]. For example, HphA is a siderophore of *Acinetobacter baumannii* and is secreted by an accompanying SLAM (HsmB) [176]. When the SLAM is removed, the siderophore is not secreted. Although the associated TonB-dependent receptor HphA does remain functional, it loses access to distal heme with no siderophore secretion [176]. OmpU is a previously identified (and named) SLAM of *H. influenzae* that is Fur regulated and involved in iron uptake [88]. There are multiple unexamined siderophore-like proteins associated with OmpU orthologues within the *Pasteurellaceae* family. There is also the similarly named porin OmpW, which has been described in *P. multocida* [177] and has an orthologue named OmpW

in *E. coli* [177]. Importantly, OmpW is not in the same protein family as OmpU and is not a SLAM – OmpU is a 14-stranded β-barrel, where OmpW is an eight-stranded β-barrel.

Nomenclature of SLAMs/*omps* is complicated. It was in the scope of this study to change the "omp" prefix of OmpU established in *H. influenzae* to match that of *Neisseria* – where the protein family is simply called SLAM/*slam*. *Neisseria* has two SLAM variants – dubbed 'SLAM1' and 'SLAM2' [178]. However, this would cause further nomenclature issues, as the *Neisseria* system also doesn't adhere to traditional nomenclature propositions [41]. Further, *ompU* in *H. influenzae* is part of the *tbp* locus [88, 106]. *ompU* is analogous to *hxuB* of the *hxuABC* locus. If a new nomenclature system was introduced for OmpU of *H. influenzae*, HxuB would also likely need renaming. As the renaming of the 'omp' prefix appears to become more complicated the more literature we involve, the prefix will remain unchanged from the original *Pasteurellaceae* publications in which it was named [88, 177].

Proteins that appear to be in the same family as OmpU were identified – OmpM, OmpO and Fishing Rod Exporter (Fre). These are all associated with iron uptake, and likely function as SLAMs for proximal genes. These appear to be similarly locus associated with iron uptake, though have no prior literature. The 'M' and 'O' suffix was added as these were the next available in the alphabetical sequence for *omps*. Fre (discussed later) was associated with Fishing Rods and was named accordingly.

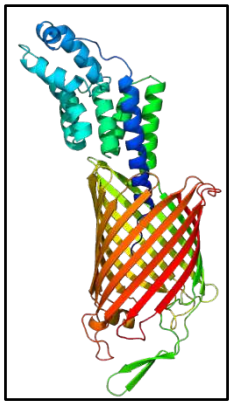
Outer Membrane Protein M (OmpM)
Structure: SLAM
Molecular weight: 50kDa
Example Protein Sequence: HaeI_OmpM1
Found in: *A. pleuropneumoniae, G. anatis* (4)*, G. parasuis, H. haemolyticus* (2), *H. influenzae*, *H. parainfluenzae, M. haemolytica*
No. of Alleles: Nine
Variable expression: No
Locus: Associated with hTbpB. Associated with *irpA* in *A. pleuropneumoniae* and *M. haemolytica.*



OmpM

OmpM is a SLAM with a 14-stranded β-barrel structure and appears to be widely distributed across *Pasteurellaceae*. This gene has some association with the *irp* locus, as was observed in *A. pleuropneumoniae* and *M. haemolytica,* but is generally associated with an hTbpB which is assumed to be exported via OmpM.

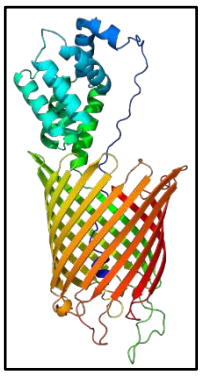
Outer Membrane Protein O (OmpO)
Structure: SLAM
Molecular weight: 57kDa
Example Protein Sequence: HaeP_OmpO1
Found in: *H. parainfluenzae* (2)
No. of Alleles: Two
Variable expression: No
Locus: Yes – with *hTbpB* Fishing Rods



OmpO

OmpO has the typical SLAM 14-stranded β-barrel structure. OmpO was not widely distributed in *Pasteurellaceae* investigated, appearing in five (28%) *H. parainfluenzae* isolates, with one isolate containing a second OmpO allele (OmpO2). OmpO is expressed with an upstream hTbpB that is also a Fishing Rod. The SLAM OmpO likely functions as an exporter for this hTbpB.

Outer Membrane Protein U (OmpU)
Structure: SLAM
Molecular weight: 56kDa
Example Protein Sequence: HaeI_OmpU1
Found in: *A. actinomycetemcomitans*, *A. pleuropneumoniae*, *G. anatis* (2), *G. parasuis*, *H. haemolyticus* (2), *H. influenzae* (2), *H. parainfluenzae* (2), *H. somni*, *P. multocida*
No. of Alleles: Six
Variable expression: No
Locus: Variable - associated with *fhuI* in *A. pleuropneumoniae*.

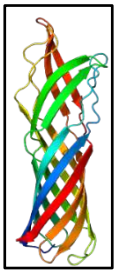


OmpU

OmpU is a previously characterised 14-stranded β-barrel SLAM involved in iron uptake [88]. It is from this previous literature that we based the continued naming scheme of Omps. OmpU has been specifically named in *P. multocida* [177] and *H. influenzae* [88]. Of note, an unrelated 16-stranded β-barrel has also been named OmpU in the *Vibrio* genus [179-181].

OmpU was observed to be proximal to the *tbp* locus in *A. actinomycetemcomitans*, *H. influenzae*, *H. haemolyticus*, *H. somni* and *G. parasuis*. This has been observed previously in *H. influenzae* [88], but further highlights an association with iron uptake. OmpU was also observed beside the *hup3* gene and multiple *plp* genes in *P. multocida*, which had not been previously shown.

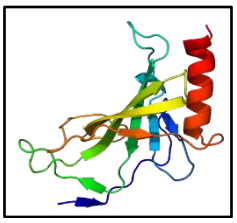
Outer Membrane Protein W (OmpW)
Structure: Eight-stranded β-barrel
Molecular weight: 23kDa
Example Protein Sequence: PasM_OmpW1.1
Found in: *A. actinomycetemcomitans, A. paragallinarum, A. pleuropneumoniae, G. anatis, G. parasuis, H. somni, M. haemolytica, P. multocida*
No. of Alleles: Four
Variable expression: No
Locus: Associated with *irpA5* and *fhuE2* in *G. parasuis*.


OmpW

OmpW is an eight-stranded β-barrel porin previously identified as being involved in iron uptake in *E. coli* [182], *P. multocida* [177] and *A. pleuropneumoniae* [57]. These proteins share >40% AAID, are orthologous and sit in a single protein family. Unfortunately, unrelated proteins (of similar structure) have also been named OmpW in other organisms, such as *A. baumannii* [183]. This nomenclature has not been amended herein as the initial naming of OmpW in *E. coli* was correctly retained in the *Pasteurellaceae* examined. However, herein, the various alleles of OmpW have been assigned an appropriate suffix. Unlike the other Omps in this study, OmpW is not a SLAM. Instead, OmpW is much smaller (23kDa) and forms an eight-stranded β-barrel [182, 184]. This porin likely forms a channel for iron influx or efflux.

The *ompW* gene was observed in all *Pasteurellaceae* genera examined except *Haemophilus* (*H. ducreyi, H. haemolyticus, H. influenzae, H. parainfluenzae*). Organisms with *ompW* tended to have them in virtually all isolates, except for *G. parasuis*, which only had *ompW* in 21% (n = 6) of isolates. Of these *G. parasuis* isolates with the *ompW* gene, even fewer had a functional gene (n = 2). *A. actinomycetemcomitans* also had a high gene presence (100%) but only 69% of genes were predicted to be functional. The most striking difference regarding *ompW* genes was seen comparing *M. haemolytica* Genotype 1 and 2 isolates. Although all *M. haemolytica* isolates contained the *ompW* gene, just 7% of Genotype 1 isolates had a functional *ompW* gene where Genotype 2 isolates all had a functional *ompW* gene. A similar finding was reported in a previous comparison of *M. haemolytica* Genotype 1 and 2 isolates [31]. In *M. haemolytica* genomes, *ompW* was observed immediately upstream of the thioredoxin gene. Thioredoxin is associated with iron regulation and metabolism [185, 186].

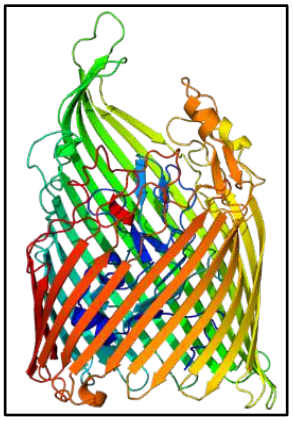
Protein E (PE)
Structure: Porin
Molecular weight: 18kDa
Example Protein Sequence: HaeH_PE1
Found in: *A. actinomycetemcomitans*, *H. haemolyticus*, *H. influenzae*, *H. parainfluenzae*, *P. multocida*
No. of Alleles: Four
Variable expression: No
Locus: No


PE

Protein E (PE), named in *H. influenzae*, is involved in iron uptake [87]. PE has been reported and investigated as a vaccine target in *H. influenzae* [187]. PE has not been previously identified in *A. actinomycetemcomitans, H. haemolyticus, H. parainfluenzae* or *Pasteurella multocida*. PE dimerises, forming a transmembrane porin and is one of few TonB-independent iron uptake systems available to the *Pasteurellaceae* family. Vaccines targeting PE have been shown to provide protective efficacy against NTHi [187, 188].

Protein E was found in 5/12 of the *Pasteurellaceae* examined. In organisms with PE, it was present and functional in nearly all genomes (approaching 100% in both cases; **Supplementary File 12**). PE has been established to be involved in hemin binding in *H. influenzae* [87, 189], which is why it was included in this investigation.

***Pseudomonas fluorescens* heme receptor (PfhR)**
Structure: β-barrel
Molecular weight: 82kDa
Example Protein Sequence: PasM_PfhR2
Found in: *P. multocida*
No. of Alleles: One (plus one in *P. fluorescens*)
Variable expression: No
Locus: Likely
-> ABC transporter [~1700bp]
-> *pfhR* [~2180bp]
-> Metallo-hydrolase [~710bp]

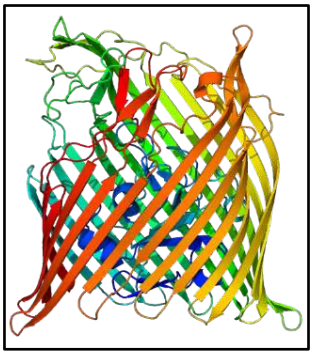


PfhR

PfhR is a 22-stranded OM β-barrel. The *pfhR* gene was observed in *P. multocida*, where it was previously named and identified as a homologue to the *Pseudomonas fluorescens* protein of the same name [101]. PfhR (PM0040) was found to bind heme in *P. multocida* [105]. A *pfhR* gene was in all *P. multocida* genomes examined and was functional in most instances. When considering host origin of *P. multocida* isolates, cow isolates had a notably lower proportion of functional *pfhR* genes than isolates from other hosts (65% functional vs ~100%). This does not appear to be attributable to collection date, geography or MLST.

PfhR appears to sit on a larger operon in *P. multocida*, with proximal upstream and downstream genes that are likely involved in iron uptake. However, as we are unsure of the size of the operon, we have not attempted to name the surrounding proteins.

Pasteurellaceae metal uptake A (PmuA)
Structure: β-barrel
Molecular weight: 82kDa
Example Protein Sequence: AggA_PmuA1.1
Found in: A. actinomycetemcomitans, G. parasuis,
H. haemolyticus (2), H. influenzae (2), H. parainfluenzae
No. of Alleles: Three
Variable expression: No.
Locus: *pmu*
-> *pmuA* [~2140bp]
-> *pmuB* (ABC transporter ATP-binding protein) [~1740bp]



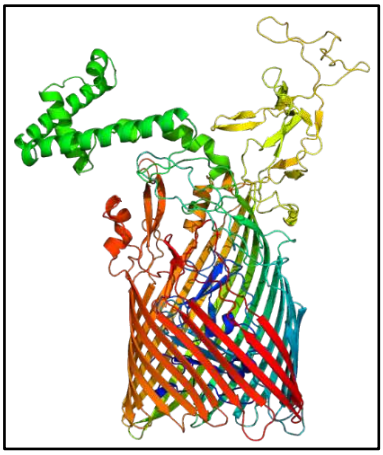
PmuA

This locus was previously unnamed, so was designated the *Pasteurellaceae* metal uptake (*pmu*) locus. The *pmu* locus contains at least two genes – *pmuA* and *pmuB*. *pmuB* encodes ABC machinery associated with metal metabolism/processing and is downstream of the *pmuA* gene, which encodes the OM β-barrel receptor. Running PmuA through the SIFTER server [40] indicated a high likelihood of involvement in iron uptake via ferrichrome transporter activity (= 0.45) or siderophore uptake transmembrane transporter activity (= 0.42). This is similar to known siderophore receptors such as ViuA (SIFTER value of 0.43). Further, PmuA was included in this study due to the structural similarity to other iron uptake proteins, as well as the *pmuA* gene proximity to the ABC transport machinery (*pmuB*). The *pmuA* gene was found in *A. actinomycetemcomitans, G. parasuis, H. parainfluenzae, H. haemolyticus* (two alleles)

and *H. influenzae* (two alleles). Interestingly, the *pmuA* gene was also observed in a *Caudoviricetes* bacteriophage. This group of bacteriophage can infect a range of bacterial hosts [190]. Recent evidence indicates there is interplay between virus and bacterial iron uptake [191]. *pmuA* can also be found in *L. koalarum, H. pittmaniae, P. ratti, C. pharyngocola* and *V. psittacicida* (data not shown).

Amino acid identity among PmuA varies more than typical β-barrels in the same protein family, e.g. using *H. parainfluenzae* PmuA as a reference, *H. haemolyticus* PmuA shares 38% AAID. Importantly, when analysed alongside other β-barrels, PmuA proteins cluster together, and only have a 15-20% AAID to other known β-barrels.

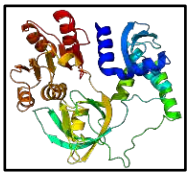
Respiratory-tract metal uptake protein (Rup)
Structure: β-barrel
Molecular weight: 116kDa
Example Protein Sequence: HaeH_Rup1
Found in: *H. haemolyticus*
No. of Alleles: One
Variable expression: No
Locus: *ire*
-> *ireX* (ABC transporter) [~1000bp]
-> *ireX* (ABC transporter) [~1000bp]
-> *ireX* (Siderophore ABC transporter ATP-binding) [~800bp]
-> *ireX* (Iron-siderophore binding protein) [~900bp]
-> *rup* [~2080bp]
-> *ireA* [~3080bp]

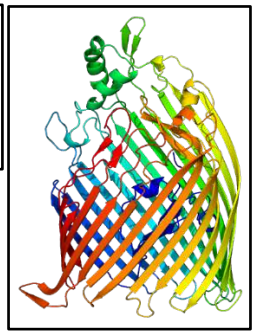


Rup

This study only observed *rup* in *H. haemolyticus* as part of the larger *ire* locus. Cursory examination found the gene/protein was present, though unnamed in *Neisseria* species (*meningitidis*, *subflava*, *lactamica, cinerea*). Rup can also be seen in *Haemophilus parahaemolyticus* genomes, one *H. parainfluenzae* genome that wasn't included in this study (strain C2008001710), *Haemophilus pittmaniae* and *Aggregatibacter segnis*. As these are primarily respiratory tract colonisers this gene/protein was named **R**espiratory-tract metal **u**ptake **p**rotein (*rup*/Rup). The unique nomenclature surrounding the *ire* locus in *Pasteurellaceae* is why this gene was given a unique prefix, rather than simply adding it into the *ire* operon. Rup is a 22-stranded β-barrel likely involved in iron uptake. As a 22-stranded β-barrel, it has structural homology to iron uptake proteins. The *rup* gene is often (but not always) found beside another established 22-stranded β-barrel involved in iron uptake, named iron-responsive element A – *ireA*. The proximity to a known iron uptake protein and ABC machinery further supports an involvement in iron uptake. No UniProt entry was eligible for processing with the SIFTER server for further analysis.

Siderophore acquisition locus A-C (*squA*, squB, squ*C*)
Structure: SquA: Siderophore-interacting protein,
SquB: β-barrel
Molecular weight: SquA: 42kDa, SquB: 70kDa
Example Protein Sequence: AviP_Squ1.1, AviP_SquB1
Found in: *A. paragallinarum*, *H. haemolyticus*
No. of Alleles: One
Variable expression: No
Locus: *squABC*
-> *squA* (siderophore-interacting protein) [750bp]
-> *squB* (β-barrel) [~1900bp]
-> *squC* (major facilitator superfamily transporter) [~1200bp]


SquA


SquB

The *squ* locus consists of *squA*, encoding a siderophore-interacting protein (more information: [192, 193]), *squB*, encoding a 22-stranded β-barrel and *squC* which encodes a major facilitator superfamily transporter (more information: [194]) gene likely providing internal iron-siderophore transport. This locus is found in *A. paragallinarum*, *H. haemolyticus* (this study) but also in *Neisseria subflava* (data not shown). The *H. haemolyticus* and *N. subflava squ* locus share high sequence identity. A more recent study examining iron uptake in *A. paragallinarum* used strain AVPG2015 as their reference strain [64], which unfortunately lacks the s*quABC* locus. The *squ* locus may play a lesser or more circumstantial, role in iron acquisition as it was only observed in 55% of *A. paragallinarum* genomes and 20% *H. haemolyticus* genomes.

Transferrin Binding Proteins

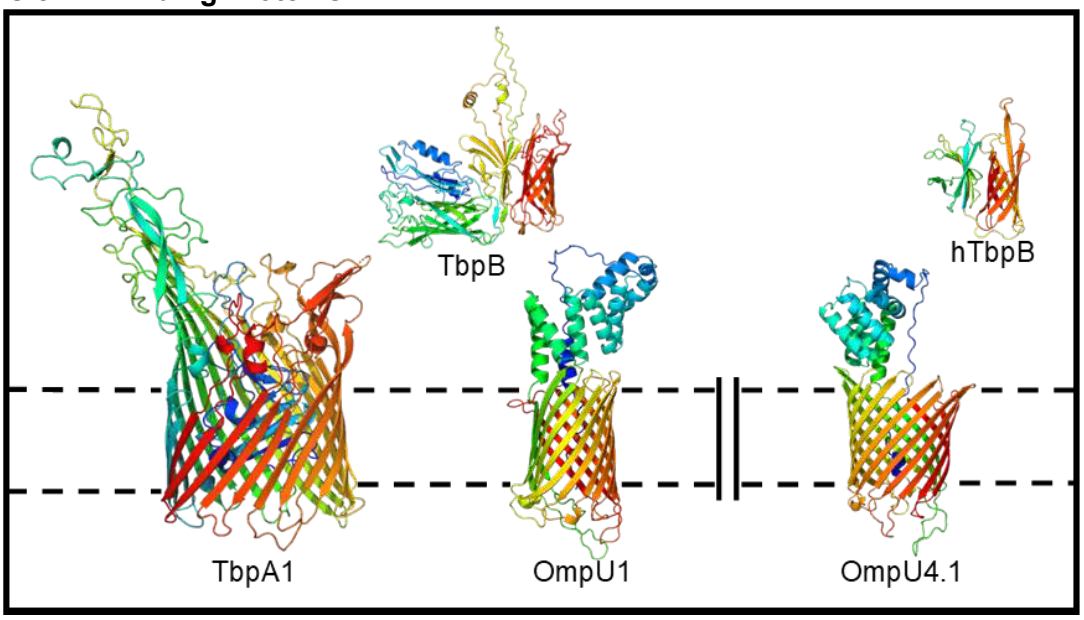


Figure 15: ***tbp* locus products in *H. influenzae*.** Multiple Tbp genes can be observed at different locus positions in *H. influenzae* isolates. In the first example, the locus encodes TbpA1, a TbpB and OmpU1. All of which are membrane-associated. A separate, unnamed locus expresses a hemi-TbpB and OmpU (allele 4.1). It's likely the Omp associated with the *tbp* loci behaves as an exporter for the secreted TbpB/hTbpB proteins. Aside from the two included examples, there are other instances where a TbpB can be observed in H. influenzae genomes.

Organism	TbpA	TbpC	TbpD
A. actinomycetemcomitans	TbpA4		
A. pleuropneumoniae	TbpA2, TbpA6		
G. parasuis	TbpA2		
H. haemolyticus	TbpA1		
H. influenzae	TbpA1		
H. somni	TbpA5	TbpC1.2	TbpD1
M. haemolytica	TbpA3		
P. multocida		TbpC1.1	

Table 8: Distribution of *tbp* locus β-barrels across Pasteurellaceae examined.

The *tbp* locus is responsible for transferrin uptake in the *Pasteurellaceae* family. It produces the β-barrel receptors TbpA and/or TbpD, a secreted TbpB or hTbpB protein (which associates with extracellular transferrin), and another β-barrel receptor that is a hybrid of *tbpA/B* genes referred to herein as TbpC [195]. There is also a gene encoding a SLAM named OmpU that is associated with the *tbp* locus, though this hasn't been widely reported. *tbpC* has been previously found in two *Pasteurellaceae*: *P. multocida* [50] and *H. somni* [51], and is a merged form of the *tbpA/B* genes which now functions as a standalone protein. *Pasteurellaceae* usually have the *tbpA/B* or *tbpC* locus alone, but can have both, and multiple copies of TbpA/B/C (this study). Some *Pasteurellaceae* also lack these genes and may not utilise transferrin. Although all Tbp proteins likely bind host transferrin, TbpA, TbpC and TbpD occupy the bacterial outer membrane as β-barrel proteins, where TbpB is a lipid-tail-anchored outer membrane protein that branches into the extracellular environment (**Figure 16**). Since TbpB extends outward from the cell and binds transferrin, it is thought that it feeds transferrin to the TbpA counterpart for uptake. Traditional understanding of TbpA/B are that they bind host specific transferrin, though exceptions where binding atypical host transferrins have been observed e.g. *H. somni*, a bovine pathogen, produces TbpC which specifically binds bovine transferrin, however *H. somni* also produces TbpB and a separate TbpA, which can bind sheep and goat transferrin [51]. *H. somni* also produces TbpD, which is closely related to TbpA and appears to be expressed as part of the same locus.

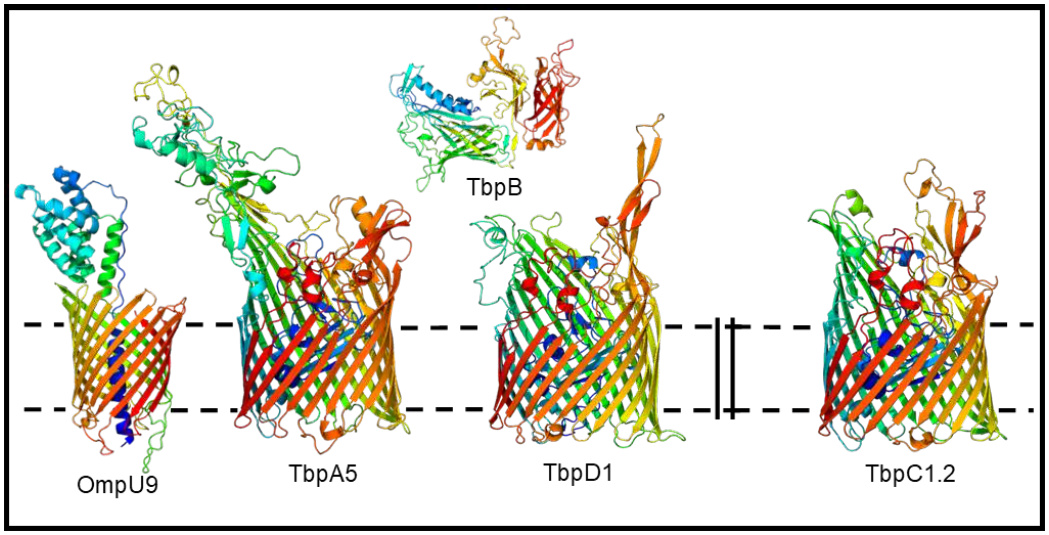


Figure 16: *H. somni* Tbps. *H. somni* genomes contain a locus with *ompU*, *tbpA, tbpB, tbpC* genes, and a separate *tbpD* gene. TbpC is not associated with the main *tbp* locus and is expressed with another iron uptake β-barrel IrpA (not shown).

Interest in vaccines targeting transferrin binding proteins is increasing, reviewed elsewhere [196]. A major hurdle has been addressing the highly variable loop domains of the β-barrel proteins, which is also observed in Tbps of the *Pasteurellaceae* family. Older studies of the homologous Tbps in *Neisseria gonorrhoeae* highlight the difficulty in generating protective antibodies against variable loop domains, however this particular study did not use modern modelling software and did not account for capsular interference [197]. More recent studies have shown that it is possible to successfully target select conserved surface epitopes, and that TbpB may prove useful as a scaffold for displaying vaccine targets [198]. We have separated Tbps into their subcategories of β-barrel receptors (TbpA/D), the secreted tbpB/hTbpB, the hybrid protein TbpC and the associated SLAM OmpU/M.

Transferrin Binding Protein A (TbpA)
Structure: (see **Figure 16**) β-barrel
Molecular weight: ~102kDa
Example Protein Sequence: HaeI_TbpA1
Found in: *A. pleuropneumoniae* (2),
A. actinomycetemcomitans, G. parasuis,
H. haemolyticus, H. influenzae, H. somni,
M. haemolytica
No. of Alleles: Six (**Table 8**)
Variable expression: ON/OFF Phase-variable via
poly tract in *A. actinomycetemcomitans* (CAGAAAAAAACGT).
Locus: *tbp*

TbpA is a well-established *Pasteurellaceae* β-barrel protein involved in iron acquisition via transferrin uptake. TbpA was observed in 7/12 of the *Pasteurellaceae* examined. We propose renaming the previously identified TbpA2 [51] in *H. somni* to TbpD, which better adheres to established nomenclature guidelines [41]. The amino acid sequence for the previously named TbpA2 [51] is also relatively low compared to typical TbpA homologues and should be designated a standalone gene/protein that's still part of the *tbp* locus – thus TbpD. This is necessary, as six TbpA alleles (TbpA1-6) were identified across the examined *Pasteurellaceae*, including TbpA2. In some cases, two of the organisms examined had highly similar TbpAs (the same allele): *A. pleuropneumoniae* and *G. parasuis* (TbpA2), *H. influenzae* and *H. haemolyticus* (TbpA1). This is perhaps unsurprising, as they both share the same host (pigs and humans, respectively).

The Neisserial TbpA has a loop domain essential for function, where even a single amino acid change impedes transferrin binding [199]. As TbpA of *Neisseria* is homologous to TbpA found across *Pasteurellaceae*, which share the same loop domain, it's likely this region needs to be conserved to protect function in both families. If this essential loop domain were to be targeted by host immune efforts it would have the dual action of impeding function of TbpA via 'clogging' the binding domain, while also helping to target this metabolically important protein.

In cases where TbpA was present in an organism it was found in all genomes examined, except for *H. haemolyticus*, which only had a 40% *tbpA* gene presence. Also, virtually all tbpA genes were functional, except for those of *A. actinomycetemcomitans*, which were only functional in ~35% of cases. TbpA alleles (TbpA2 and TbpA6) present in all *A. pleuropneumoniae* genomes have not been previously reported. In the second, undocumented locus the *tbpB* gene appears to be transcribed, but the *tbpA* gene lacks a traditional start codon. Despite the lack of a clear start codon the rest of the gene is in frame.

Transferrin Binding Protein C (TbpC) and D (TbpD)
Structure: (see **Figure 16**) β-barrel
Molecular weight: ~90kDa
Example Protein Sequence: PasM_TbpC1.1, HisS_TbpD1
Found in: TbpC: *H. somni*, *P. multocida.*
TbpD: *H. somni*
No. of Alleles: One of each TbpC and TbpD (**Table 8**)
Variable expression: TbpC: Yes – ON/OFF phase-variable
via 5′ poly G tract in *H. somni* (ATAGGGGGGGGTTC).
TbpD: Yes - ON/OFF phase-variable
via 5′ poly G tract in *H. somni* (ATAGGGGGGGGTTC)
Locus: *tbp*

Explained in the (*above*) transferrin binding protein introduction, TbpA, TbpC and TbpD are all 22-stranded β-barrels associated with the *tbp* locus. We have needed to modify prior nomenclature to accommodate the findings herein, which shows that TbpC and TbpD are distinct from TbpA, sharing ~30 and ~28% AAID to TbpA, respectively. TbpC and TbpD share ~35% AAID to each other. TbpC has previously been referred to as TbpA2 [195], however that doesn't adhere to suggested nomenclature guidelines [41]. This indicates that rather than being allelic variants of TbpA, TbpC and TbpD are unique (albeit related) proteins (**Table 8**). It's highly likely they behave as receptors for a specific target (e.g. goat, sheep and cow transferrin) and can take up various accessory sources of iron. Prior literature has shown the presence of two TbpA-like proteins in *H. somni* [51]. However, genomes tended to contain one of each TbpA, TbpC, TbpD. The *tbpC* gene was previously established to be phase-variable [93]. TbpD is also phase-variable (this study) and shares the same 5′ nucleotide region that allows ON/OFF switching (ATA**GGGGGGGGG**TTC).

Transferrin Binding Protein B (TbpB)
Structure: (see **Figure 16**) one or two lobed β-strand protein
Molecular weight: Variable 45kDa (hTbpB) – 72kDa (TbpB)
Example Protein Sequence: HisS_TbpB2, GalA_hTbpB1
Found in: *A. pleuropneumoniae* (7), *A. actinomycetemcomitans*, *G. anatis* (5), *G. parasuis* (2), *H. haemolyticus* (2), *H. influenzae* (6), *H. parainfluenzae* (6), *H. somni* (2), *M. haemolytica* (2)
No. of Alleles: >60 found
Variable expression: May be ON/OFF via
poly nucleotide sequence, variable length
via tandem repeat (see **Fishing Rod** section below), or both.
Locus: *tbp,* variable

Excluding hemophilin (and true orthologues), more than 60 TbpB-like proteins were found in the *Pasteurellaceae* examined. These proteins sit in a group with a wide distribution across Gram negative bacteria [92, 106]. Common to immune exposed proteins, the surface (N lobe) of TbpBs tends to be hypervariable, where the inner components (including the β-strands) are highly conserved. The C lobe is conserved despite appearing equally immune exposed as the N lobe. Structural comparison of the C and N lobes show that these regions are structurally similar – both an eight stranded β-barrel with a unique handle domain [200]. Two forms of TbpBs have been identified in *Neisseria spp.* – isotypes I and II [201]. However, the 'isotype I and II' nomenclature [201] lacks a strict definition and relies on molecular weight of the protein. This doesn't work universally and is complicated by phase-variable domains in TbpBs. Further, an 'isotype 1' TbpB has been previously named factor H binding lipoprotein in *H. influenzae* [90] and may associate with heme [202]. For simplicity and clarity, we propose calling the single-lobed TbpBs hemi-transferrin binding protein B (hTbpB, for short). "Hemi" due to their

structure being half the regular TbpB. In *Neisseria*, TbpBs outnumber hTbpBs, though in *Pasteurellaceae* there are more hTbpBs than TbpBs. Given the large number of hTbpBs and TbpBs found, we have not named each individual h/TbpB identified. Instead, we have provided a temporary identifier to temporarily differentiate them. There were cases where existing nomenclature identified and named a hTbpB/tbpB, such as OmlA of *A. pleuropneumoniae* (including alleles) [53] and PlpE of *P. multocida [60]*.

Prior study of Neisseria hTbpBs suggested the presence of a tetratricopeptide repeat domain present in all hTbpBs [178]. However, the 5′ repeat regions do not appear to fit the criteria to be classified as a tetratricopeptide repeat domain as previously defined [203]. The sequences are also not recognised by the tetratricopeptide repeat recognition tool TPRpred [204] (data not shown). Instead, we propose the highly variable repeat domains of these genes indicate that they are periscope proteins [205], and that the repeat domains are not functional motifs like those observed in other ligand binding proteins [206].

hTbpBs are broadly accepted to be involved in iron uptake. There are cases where a hTbpB has a well-investigated role in iron acquisition, for example Hemophilin (described in *H. haemolyticus*) binds heme [81, 83]. However, the structurally similar protein *Pasteurella multocida* lipoprotein E (PlpE) has no known association with iron uptake. A PlpE homologue named OmlA has also been described in *A. pleuropneumoniae* [60].

There's debate about the function of the N-terminal domain vs the C-terminal domain [207, 208]. *Moraes et al.* suggested that this distinction between C and N lobes implies the C lobe is more important to function and frequently 'hidden' from the immune system via bound protein, where the N lobe is perhaps more frequently immune exposed and free to vary [200]. When comparing hTbpBs to the larger TbpB proteins, some of the hTbpBs align with the N-terminus, however some also align with the C-terminus. This is curious, since it has been suggested that the C-terminal lobe of Neisserial TbpBs are not involved in binding [207], though this was later refuted by *DeRocco et al.* [208]. *DeRocco et al.* also provided a sequence guide to separating the C- and N-terminal lobes, which was used herein to confirm standalone C-terminal hTbpBs (data not shown). Many *tbpB* and *hTbpB* genes examined within this study were Fishing Rods – a type of periscope protein described below. Function aside, TbpBs have been found to be effective vaccine targets in pigs [209], and show promise as antigenic peptide carriers in human vaccine efforts [198].

Separate from Fishing Rods, some hTbpBs (and TbpBs) were prone to phase-variable ON/OFF expression via polynucleotide tract. Such is the case with hTbpB3 from *A. pleuropneumoniae* (ACT**CCCCC**TAA) and TbpB1 (gene 6) of *H. somni* (TA**CCCCCCCCC**T) and two hTbpBs from *G. anatis* (GT**GGGGGGGGGGGG**GT, TGA**TTTTTTTT**GAA). This was observed in both hTbpBs and TbpBs.

Fishing Rods (FR)
Structure: variable length hTbpB (see **Figure 17**)
Molecular weight: Variable depending on length of tandem region.
Example Protein Sequence: GlaP_hTbpB_FR1
Found in: *A. pleuropneumoniae* (2), *G. anatis* (3), *G. parasuis* (3) , *H. haemolyticus* (4), *H. influenzae*, *H. parainfluenzae* (9), *M. haemolytica* (3)
Variable expression: Tandem repeat length–dependent structural variation.
Locus: Variable – associated with SLAM (*fre*, *ompO, ompM, ompU*)

Fishing rods fall into the recently described group of 'periscope proteins' [205] and SLAM dependent surface lipoproteins [106]. The small amount of literature involving these proteins [106] did not comprehensively include their structure (hTbpBs vs TbpBs), their propensity to undergo tandem repeat length–dependent structural variation or predict a role in iron uptake. Fishing rods are abundant across *Pasteurellaceae* and appear heavily associated with iron uptake.

The term 'Fishing Rod' herein refers to an outer membrane protein that has three defining structural elements: a lipid anchor sequence, a tandemly repeating domain encoding an alpha-helix (which varies in length generationally) and a (functional) hemi-Transferrin Binding Protein B (hTbpB) domain. Importantly, this is a broad category of proteins, which are not all necessarily closely phylogenetically related. Describing a protein as a "Fishing Rod" doesn't mean it shares significant sequence homology with another protein with this same description, only that it has the three key features listed above.

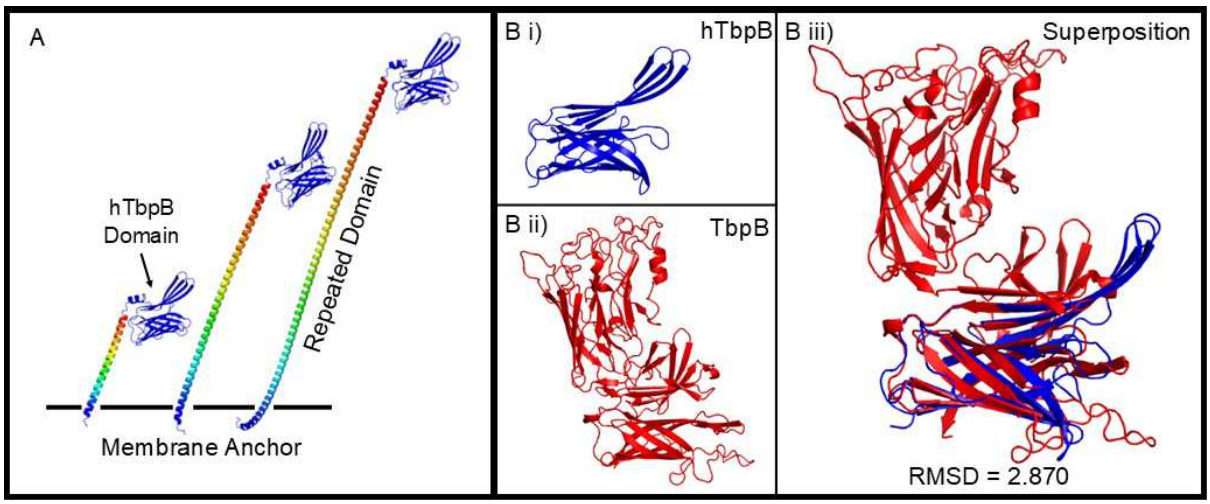


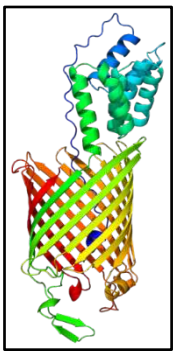
Figure 17: From *M. haemolytica* isolates examined in this study: **A)** A Fishing Rod from Strain IMT47952 has two tandem repeated units (short). Strain 184 (medium length) has eight repeated units (middle). Strain 1451 has 13 repeated units (long). Rather than ON/OFF phase-variable expression, the change in repeated sequence number causes variable length of protrusion from the outer membrane.

Fishing Rods are membrane-bound and contain an 8-stranded β-barrel domain (hTbpB) at the 3′ end of the protein. This extends from the bacterial cell surface via repeated alpha-helix domains, encoded by a tandem repeat sequences at the 5′ end of the gene. Fishing Rods contain a lipid-anchor sequence (before the alpha-helix domain) and generally have adjacent export machinery in the form of a SLAM. The number of tandem repeats can increase or decrease with each cell division/generation. This does not result in ON/OFF phase-variation as is the case with previously discussed tandem repeats. As has been established with periscope proteins [205], a change to the number of tandem repeats changes the distance the functional (hTbpB) domain extends from the outer membrane. The change in reach of the functional domain impacts interaction with host immune factors, target molecule (e.g. heme), and other membrane anchored proteins. A longer Fishing Rod offers the advantage of reaching target molecules at greater distances. It also allows for a thicker bacterial capsule and can hand-off a target molecule to a receptor from a longer distance. However, extending further from the outer membrane comes with the risk of increased immune interaction.

TbpB and hTbpB proteins typically use membrane anchored β-barrels as a receptor for their target, once captured. Further study is needed as the targets of the many Fishing Rods found

in this study are unknown, as are their corresponding receptor. However, evidence suggests Fishing Rods play a role in iron acquisition; the Fishing Rod functional domains have low amino acid sequence homology to the C lobe of TbpBs but have high structural similarity. This is also the case for different families of 22-stranded TonB-dependent β-barrels which are typically involved in iron (or similar metabolite) uptake. Despite having similar targets and overall function, the amino acid identity homology between different 22-stranded TonB-dependent β-barrel families is generally <30%.

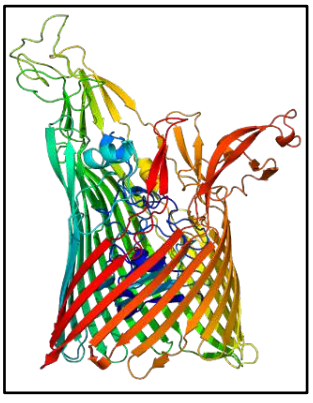
Fishing Rod Exporter (Fre)
Structure: SLAM
Molecular weight: 24kDa
Example Protein Sequence: GalA_Fre
Found in: *A. pleuropneumoniae*, *G. parasuis*, G. anatis, *H. influenzae* *H. haemolyticus*, *H. parainfluenzae*, *M. haemolytica* (2)
Variable expression: No.
Locus: Associated with Fishing Rods



Fre

Named herein, the Fishing Rod Exporter (Fre) is associated with Fishing Rods (described above). Fre genes (fre) are closely associated with Fishing Rod genes as part of the same locus. However, not every Fishing Rod gene has an accompanying Fre; all organisms that produce a Fishing Rod do encode at least one Fre. Fishing Rod Exporters are structurally similar to the exporters of two-partner secretion systems [210]. It's likely these exporters are important to the secretion of Fishing Rods. In *G. anatis* there is an ON/OFF phase-variably expressed *hTbpB* gene beside a *fre* gene. Though it had most of the characteristics of a Fishing Rod, including an alpha-helix that did vary between isolates, the *hTbpB* gene lacked a tandem repeat tract and therefore did not meet the prior set definition of a Fishing Rod. This instance shows that the Fishing Rod Exporter is not always associated with a Fishing Rod or hTbpB and may act as a general-purpose SLAM for hTbpBs from other organisms.

TonB-dependent protein X (TdX)
Structure: β-barrel
Molecular weight: 96kDa
Example Protein Sequence: Hduc_TdX1
Found in: *A. pleuropneumoniae, H. ducreyi.*
No. of Alleles: 2
Variable expression: No
Locus: *tdX*
-> *tdX* [~2490bp]
-> Mechanosensitive ion channel family protein [~910bp]

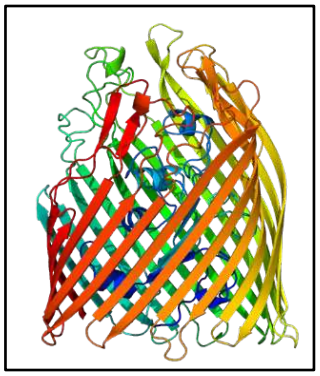


TdX

TdX is a 22-strand β-barrel found functional in virtually all *A. pleuropneumoniae* and *H. ducreyi* isolates (this study). TdX was first named and identified in *H. ducreyi* [76], but also observed as APJL_1922 in *A. pleuropneumoniae* [58]. In both cases, these proteins were associated with iron uptake [58, 76]. These two proteins are different alleles within the TdX protein family – herein named TdX1 (*H. ducreyi)* and TdX2 (*A. pleuropneumoniae)*. The downstream possible second gene of the TdX locus encodes a mechanosensitive ion channel family protein, which forms a quaternary iron channel in the cell membrane. This gene was observed beside *tdX* in both *A. pleuropneumoniae* and *H. ducreyi* and may be part of a larger operon.

The initial designation of TonB-dependent protein X (TdX) to this gene/protein was meant as a placeholder until further information on this protein was found. It was within the scope of this study to provide a more appropriate name. However, it is premature to rename this gene/protein without further information about its specific target and whether the tdX gene belongs to a larger operon. Thus, TdX was retained, but remains a placeholder.

***Vibrio cholerae* ferric vibriobactin receptor (ViuA)**
Structure: β-barrel
Molecular weight: ~78kDa
Example Protein Sequence: Hpara_ViuA2
Found in: *H. parainfluenzae*
No. of Alleles: One (plus one in Vibrio cholerae).
Variable expression: No
Locus: *viu*



ViuA

<-*viuA* (β-barrel receptor) [~2060bp]
<-*viuB* (ABC transport machinery) [~1720bp]
<-*viuC* (ABC transport machinery) [~1730bp]
<-*viuD* (Major facilitator superfamily transporter) [~1230bp]

ViuA is another 22-stranded β-barrel. This gene/protein has no prior literature in *Pasteurellaceae*. However, it has ~40% AAID to the previously described *Vibrio cholerae* ferric vibriobactin receptor ViuA [211] and the orthologue ferric vulnibactin receptor VuuA of *Vibrio vulnificus* [212]. These homologous genes/proteins are within the same protein family. ViuA was named first, so this nomenclature was adopted for orthologues in *Pasteurellaceae*. The initial *Vibrio cholerae* ViuA lacks a suffix that distinguishes the allele, so it was designated 'ViuA1' and the newly identified allele in *H. parainfluenzae* was assigned 'ViuA2.' The *viuA2* gene was only observed in *H. parainfluenzae*. However, it can also be found in other *Aggregatibacter spp.* and in human soil, water and sewage pathogen *Providencia stuartii* (data not shown). Because prior literature clashed with the structure of the *viu* operon in *Pasteurellaceae* the *viu* locus was named in reverse: *viuD* is the first gene in the operon; *viuA* is the fourth gene in the operon.

Of the 18 *H. parainfluenzae* isolates examined within this study the *viu* locus was only present in four (22%), and functional in three. The distribution of the isolates with the *viu* locus suggests the gene is not restricted to a clonal population, as it was observed in isolates from the USA, Europe and Africa obtained 18 years apart.

Conclusion

Inconsistent nomenclature obscures biological relationships. We have provided a modern nomenclature system based on historical precedent. This nomenclature system was applied to bacterial genes involved in iron uptake across *Pasteurellaceae* examined within this study. Simultaneously, new factors were found in these organisms. Targeting bacterial surface factors is a prominent strategy for vaccine design. We show that *Pasteurellaceae* iron acquisition systems are highly dynamic and diverse, undergoing continual change rather than forming a static 'core' repertoire of factors. Understanding the behaviour of these systems is important for rational vaccine design. To produce vaccines and therapies that provide long-term and broad protection we must cast a deeper and wider net. No single vaccine target is likely to remain effective indefinitely. Factors allowing bacteria to bamboozle the immune system include phase-variation, allelic diversification, redundancy, gene duplication and antigenic drift. Identifying and overcoming these issues is a necessary step in designing robust vaccines against important systems, including iron uptake. A pan-species approach is also likely necessary, since pathogenic bacteria may share factors via horizontal gene transfer. Vaccines and therapies targeting an array of iron uptake factors may prove effective in achieving broad protection with longevity. This study informs the design of future vaccines and treatments while furthering our understanding of bacterial pathobiology.

Funding Statement

This work was conducted without internal or external funding or remuneration.

Conflict of Interest Statement

No conflict of interest.

Acknowledgements

We thank Prof. Stanley Spinola for their correspondence on the nomenclature surrounding TdX.

References

1. Sassu, E.L., et al., *Update on Actinobacillus pleuropneumoniae-knowledge, gaps and challenges.* Transbound Emerg Dis, 2018. **65 Suppl 1**: p. 72-90.
2. Raja, M., F. Ummer, and C.P. Dhivakar, *Aggregatibacter actinomycetemcomitans - a tooth killer?* J Clin Diagn Res, 2014. **8**(8): p. Ze13-6.
3. Xu, Y., et al., *Characterization of emergent Avibacterium paragallinarum strains and the protection conferred by infectious coryza vaccines against them in China.* Poultry Science, 2019. **98**: p. 6463-6471.
4. Blackall, P.J., *Infectious coryza: overview of the disease and new diagnostic options.* Clin Microbiol Rev, 1999. **12**(4): p. 627-32.
5. Abd El-Ghany, W.A., et al., *Gallibacterium anatis infection in poultry: a comprehensive review.* Trop Anim Health Prod, 2023. **55**(6): p. 383.
6. Zhang, B., et al., *Update on the pathogenesis of Haemophilus parasuis infection and virulence factors.* Vet Microbiol, 2014. **168**(1): p. 1-7.
7. Ruiz Del Pino, M., et al., *Clinical significance of isolation of Haemophilus no ducreyi in genital samples. Systematic review.* Enferm Infecc Microbiol Clin (Engl Ed), 2023. **41**(8): p. 468-484.
8. Gilsdorf, J.R., *What the pediatrician should know about non-typeable Haemophilus influenzae.* J Infect, 2015. **71 Suppl 1**: p. S10-4.
9. Wen, S., et al., *Molecular epidemiology and evolution of Haemophilus influenzae.* Infect Genet Evol, 2020. **80**: p. 104205.
10. Sriram, K.B., et al., *Nontypeable Haemophilus influenzae and chronic obstructive pulmonary disease: a review for clinicians.* Crit Rev Microbiol, 2018. **44**(2): p. 125-142.
11. Onafowokan, O.O., R. Mateo, and H.J.R. Bonatti, *A Series of Haemophilus parainfluenzae Surgical Infections and Review of the Literature.* Surg Infect (Larchmt), 2021. **22**(9): p. 940-947.
12. Timsit, E., et al., *Prevalence and antimicrobial susceptibility of Mannheimia haemolytica, Pasteurella multocida, and Histophilus somni isolated from the lower respiratory tract of healthy feedlot cattle and those diagnosed with bovine respiratory disease.* Vet Microbiol, 2017. **208**: p. 118-125.
13. Confer, A.W. and S. Ayalew, *Mannheimia haemolytica in bovine respiratory disease: immunogens, potential immunogens, and vaccines.* Anim Health Res Rev, 2018. **19**(2): p. 79-99.
14. Rice, J.A., et al., *Mannheimia haemolytica and bovine respiratory disease.* Anim Health Res Rev, 2007. **8**(2): p. 117-28.
15. Mostaan, S., et al., *Pasteurella multocida Vaccine Candidates: A Systematic Review.* Avicenna J Med Biotechnol, 2020. **12**(3): p. 140-147.
16. Piorunek, M., B. Brajer-Luftmann, and J. Walkowiak, *Pasteurella Multocida Infection in Humans.* Pathogens, 2023. **12**(10).
17. Alteri, C.J., et al., *Mucosal immunization with iron receptor antigens protects against urinary tract infection.* PLoS Pathog, 2009. **5**(9): p. e1000586.
18. Runyen-Janecky, L.J., *Role and regulation of heme iron acquisition in gram-negative pathogens.* Front Cell Infect Microbiol, 2013. **3**: p. 55.
19. Gehrer, C.M., et al., *Advances in Ferritin Physiology and Possible Implications in Bacterial Infection.* Int J Mol Sci, 2023. **24**(5).
20. Ahmed, M.H., M.S. Ghatge, and M.K. Safo, *Hemoglobin: Structure, Function and Allostery.* Subcell Biochem, 2020. **94**: p. 345-382.
21. di Masi, A., et al., *Haptoglobin: From hemoglobin scavenging to human health.* Mol Aspects Med, 2020. **73**: p. 100851.
22. Detzel, M.S., et al., *Revisiting the interaction of heme with hemopexin.* Biol Chem, 2021. **402**(6): p. 675-691.

23. Saita, Y., et al., *Plasma myoglobin indicates muscle damage associated with acceleration/deceleration during football.* J Sports Med Phys Fitness, 2023. **63**(12): p. 1337-1342.
24. Baker, E.N., H.M. Baker, and R.D. Kidd, *Lactoferrin and transferrin: functional variations on a common structural framework.* Biochem Cell Biol, 2002. **80**(1): p. 27-34.
25. Gomme, P.T., K.B. McCann, and J. Bertolini, *Transferrin: structure, function and potential therapeutic actions.* Drug Discov Today, 2005. **10**(4): p. 267-73.
26. Wu, Z., et al., *A zero-sum game or an interactive frame? Iron competition between bacteria and humans in infection war.* Chin Med J (Engl), 2022. **135**(16): p. 1917-1926.
27. Kuang, H., et al., *Humoral regulation of iron metabolism by extracellular vesicles drives antibacterial response.* Nat Metab, 2023. **5**(1): p. 111-128.
28. Weiss, G., *Macrophage vesicles starve bacteria of iron.* Nature Metabolism, 2023. **5**(1): p. 10-12.
29. Phillips, Z.N., et al., *Examination of phase-variable haemoglobin-haptoglobin binding proteins in non-typeable Haemophilus influenzae reveals a diverse distribution of multiple variants.* FEMS Microbiol Lett, 2022. **369**(1).
30. Krewulak, K.D. and H.J. Vogel, *Structural biology of bacterial iron uptake.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2008. **1778**(9): p. 1781-1804.
31. Wynn, E.L. and M.L. Clawson, *Differences between predicted outer membrane proteins of Pasteurella multocida, Histophilus somni, and genotype 1 and 2 Mannheimia haemolytica strains isolated from cattle.* Genome, 2022. **65**(2): p. 115-121.
32. Mosbahi, K., et al., *Bacterial iron acquisition mediated by outer membrane translocation and cleavage of a host protein.* Proceedings of the National Academy of Sciences, 2018. **115**(26): p. 6840-6845.
33. Kramer, J., Ö. Özkaya, and R. Kümmerli, *Bacterial siderophores in community and host interactions.* Nat Rev Microbiol, 2020. **18**(3): p. 152-163.
34. Cescau, S., et al., *Heme acquisition by hemophores.* Biometals, 2007. **20**(3-4): p. 603-13.
35. Troxell, B. and H.M. Hassan, *Transcriptional regulation by Ferric Uptake Regulator (Fur) in pathogenic bacteria.* Front Cell Infect Microbiol, 2013. **3**: p. 59.
36. Phillips, Z.N., et al., *Phasevarions of bacterial pathogens - phase-variable epigenetic regulators evolving from restriction-modification systems.* Microbiology, 2019.
37. Phillips, Z.N., et al., *Phase-variable bacterial loci: how bacteria gamble to maximise fitness in changing environments.* Biochem Soc Trans, 2019. **47**(4): p. 1131-1141.
38. Zhu, C., et al., *A Florfenicol-Resistant Plasmid Shuttling Between Actinobacillus pleuropneumoniae and Glaesserella parasuis.* Microb Drug Resist, 2024.
39. Mell, J.C. and R.J. Redfield, *Natural competence and the evolution of DNA uptake specificity.* J Bacteriol, 2014. **196**(8): p. 1471-83.
40. Sahraeian, S.M., K.R. Luo, and S.E. Brenner, *SIFTER search: a web server for accurate phylogeny-based protein function prediction.* Nucleic Acids Res, 2015. **43**(W1): p. W141-7.
41. Demerec, M., et al., *A proposal for a uniform nomenclature in bacterial genetics.* Genetics, 1966. **54**(1): p. 61-76.
42. Pawelek, P.D. and J.W. Coulton, *Hemoglobin-binding protein HgbA in the outer membrane of Actinobacillus pleuropneumoniae: homology modelling reveals regions of potential interactions with hemoglobin and heme.* J Mol Graph Model, 2004. **23**(3): p. 211-21.
43. Jin, H., et al., *Characterization of hgpA, a gene encoding a haemoglobin/haemoglobin-haptoglobin-binding protein of Haemophilus influenzae.* Microbiology (Reading), 1999. **145 (Pt 4)**: p. 905-914.
44. Tremblay, Y.D., F. Bahrami, and D.F. Niven, *Acquisition of haemoglobin-bound iron by Histophilus somni.* Vet Microbiol, 2006. **114**(1-2): p. 104-14.

45. Rhodes, E.R., et al., *Iron acquisition in the dental pathogen Actinobacillus actinomycetemcomitans: what does it use as a source and how does it get this essential metal?* Biometals, 2007. **20**(3-4): p. 365-77.
46. Bosch, M., et al., *Characterization of the Pasteurella multocida hgbA gene encoding a hemoglobin-binding protein.* Infect Immun, 2002. **70**(11): p. 5955-64.
47. Stevens, M.K., et al., *A hemoglobin-binding outer membrane protein is involved in virulence expression by Haemophilus ducreyi in an animal model.* Infect Immun, 1996. **64**(5): p. 1724-35.
48. Abramson, J., et al., *Addendum: Accurate structure prediction of biomolecular interactions with AlphaFold 3.* Nature, 2024. **636**(8042): p. E4.
49. Beyt, B.E., Jr., et al., *Human pulmonary pasteurellosis.* Jama, 1979. **242**(15): p. 1647-8.
50. Farahani, M.F., M. Esmaelizad, and A.R. Jabbari, *Investigation of iron uptake and virulence gene factors (fur, tonB, exbD, exbB, hgbA, hgbB1, hgbB2 and tbpA) among isolates of Pasteurella multocida from Iran.* Iran J Microbiol, 2019. **11**(3): p. 191-197.
51. Pogoutse, A.K. and T.F. Moraes, *Transferrin Binding Protein B and Transferrin Binding Protein A2 Expand the Transferrin Recognition Range of Histophilus somni.* J Bacteriol, 2020. **202**(14).
52. Martinez-Gutierrez, C.A., Y. Huang, and L.M. Bobay, *Prevalence and Evolutionary Implications of Genome Rearrangements in Bacteria.* Genome Biol Evol, 2026. **18**(2).
53. Rossi, C.C., et al., *Characterization of the omlA gene from different serotypes of Actinobacillus pleuropneumoniae: A new insight into an old approach.* Genet Mol Biol, 2013. **36**(2): p. 243-51.
54. Jacques, M., *Surface polysaccharides and iron-uptake systems of Actinobacillus pleuropneumoniae.* Can J Vet Res, 2004. **68**(2): p. 81-5.
55. Wilke, M., B. Franz, and G.F. Gerlach, *Characterization of a large transferrin-binding protein from Actinobacillus pleuropneumoniae serotype 7.* Zentralbl Veterinarmed B, 1997. **44**(2): p. 73-86.
56. Shakarji, L., et al., *Fhua and HgbA, outer membrane proteins of Actinobacillus pleuropneumoniae: their role as virulence determinants.* Can J Microbiol, 2006. **52**(4): p. 391-6.
57. Chen, X., et al., *Involvement of the Actinobacillus pleuropneumoniae ompW Gene in Confrontation of Environmental Pressure.* Front Vet Sci, 2022. **9**: p. 846322.
58. Li, L., et al., *Catecholamines promote Actinobacillus pleuropneumoniae growth by regulating iron metabolism.* PLoS One, 2015. **10**(4): p. e0121887.
59. Buettner, F.F., et al., *Analysis of the Actinobacillus pleuropneumoniae HlyX (FNR) regulon and identification of iron-regulated protein B as an essential virulence factor.* Proteomics, 2009. **9**(9): p. 2383-98.
60. Pandher, K., A.W. Confer, and G.L. Murphy, *Genetic and immunologic analyses of PlpE, a lipoprotein important in complement-mediated killing of Pasteurella haemolytica serotype 1.* Infect Immun, 1998. **66**(12): p. 5613-9.
61. Chung, J.W., et al., *Immunoproteomic analyses of outer membrane antigens of Actinobacillus pleuropneumoniae grown under iron-restricted conditions.* Vet Microbiol, 2012. **159**(1-2): p. 187-94.
62. Gram, T., et al., *An Actinobacillus pleuropneumoniae PCR typing system based on the apx and omlA genes--evaluation of isolates from lungs and tonsils of pigs.* Vet Microbiol, 2000. **75**(1): p. 43-57.
63. Stacy, A., et al., *Microbial Community Composition Impacts Pathogen Iron Availability during Polymicrobial Infection.* PLoS Pathog, 2016. **12**(12): p. e1006084.
64. Huo, C., et al., *The Transcriptomic and Bioinformatic Characterizations of Iron Acquisition and Heme Utilization in Avibacterium paragallinarum in Response to Iron-Starvation.* Front Microbiol, 2021. **12**: p. 610196.

65. Ogunnariwo, J.A. and A.B. Schryvers, *Correlation between the ability of Haemophilus paragallinarum to acquire ovotransferrin-bound iron and the expression of ovotransferrin-specific receptors.* Avian Dis, 1992. **36**(3): p. 655-63.
66. Abascal, E.N., et al., *Identification of iron-acquisition proteins of Avibacterium paragallinarum.* Avian Pathol, 2009. **38**(3): p. 209-13.
67. Cope, L.D., et al., *A gene cluster involved in the utilization of both free heme and heme:hemopexin by Haemophilus influenzae type b.* J Bacteriol, 1995. **177**(10): p. 2644-53.
68. Chantes-Guerra, A., et al., *A putative siderophore receptor of Gallibacterium anatis 12656-12 under Fur control also binds hemoglobin.* Front Microbiol, 2022. **13**: p. 951173.
69. Johnson, T.J., et al., *Genome analysis and phylogenetic relatedness of Gallibacterium anatis strains from poultry.* PLoS One, 2013. **8**(1): p. e54844.
70. Bager, R.J., et al., *In silico prediction of Gallibacterium anatis pan-immunogens.* Vet Res, 2014. **45**(1): p. 80.
71. Gong, X., et al., *Genomic insight into the diversity of Glaesserella parasuis isolates from 19 countries.* mSphere, 2024. **9**(9): p. e0023124.
72. Álvarez-Estrada, Á., et al., *Transcriptomics of Haemophilus (Glässerella) parasuis serovar 5 subjected to culture conditions partially mimetic to natural infection for the search of new vaccine antigens.* BMC Vet Res, 2018. **14**(1): p. 326.
73. Sack, M. and N. Baltes, *Identification of novel potential virulence-associated factors in Haemophilus parasuis.* Vet Microbiol, 2009. **136**(3-4): p. 382-6.
74. del Río, M.L., et al., *Molecular characterization of Haemophilus parasuis ferric hydroxamate uptake (fhu) genes and constitutive expression of the FhuA receptor.* Vet Res, 2006. **37**(1): p. 49-59.
75. Iida, T. and K. Yamamoto, *Cloning and expression of two genes encoding highly homologous hemolysins from a Kanagawa phenomenon-positive Vibrio parahaemolyticus T4750 strain.* Gene, 1990. **93**(1): p. 9-15.
76. Leduc, I., et al., *Evaluation of the repertoire of the TonB-dependent receptors of Haemophilus ducreyi for their role in virulence in humans.* J Infect Dis, 2008. **197**(8): p. 1103-9.
77. Bong, C.T., M.E. Bauer, and S.M. Spinola, *Haemophilus ducreyi: clinical features, epidemiology, and prospects for disease control.* Microbes Infect, 2002. **4**(11): p. 1141-8.
78. Alfa, M., *The laboratory diagnosis of Haemophilus ducreyi.* Can J Infect Dis Med Microbiol, 2005. **16**(1): p. 31-4.
79. Nguyen, A.V. and A.M. Soulika, *The Dynamics of the Skin's Immune System.* Int J Mol Sci, 2019. **20**(8).
80. Brothwell, J.A., et al., *Interactions of the Skin Pathogen Haemophilus ducreyi With the Human Host.* Front Immunol, 2020. **11**: p. 615402.
81. Latham, R.D., et al., *A heme-binding protein produced by Haemophilus haemolyticus inhibits non-typeable Haemophilus influenzae.* Mol Microbiol, 2020. **113**(2): p. 381-398.
82. Hariadi, N.I., et al., *Comparative Profile of Heme Acquisition Genes in Disease-Causing and Colonizing Nontypeable Haemophilus influenzae and Haemophilus haemolyticus.* J Clin Microbiol, 2015. **53**(7): p. 2132-7.
83. Fulte, S., et al., *Heme sequestration by hemophilin from Haemophilus haemolyticus reduces respiratory tract colonization and infection with non-typeable Haemophilus influenzae.* mSphere, 2024. **9**(3): p. e0000624.
84. Lynch, S.V., et al., *Role of the rapA gene in controlling antibiotic resistance of Escherichia coli biofilms.* Antimicrob Agents Chemother, 2007. **51**(10): p. 3650-8.
85. Morton, D.J., et al., *Identification of a haem-utilization protein (Hup) in Haemophilus influenzae.* Microbiology (Reading), 2004. **150**(Pt 12): p. 3923-33.
86. Loosmore, S.M., et al., *Cloning and expression of the Haemophilus influenzae transferrin receptor genes.* Mol Microbiol, 1996. **19**(3): p. 575-86.

87. Rodríguez-Arce, I., et al., *Moonlighting of Haemophilus influenzae heme acquisition systems contributes to the host airway-pathogen interplay in a coordinated manner.* Virulence, 2019. **10**(1): p. 315-333.
88. Whitby, P.W., et al., *The iron/heme regulated genes of Haemophilus influenzae: comparative transcriptional profiling as a tool to define the species core modulon.* BMC Genomics, 2009. **10**: p. 6.
89. Morton, D.J., et al., *Identification of a siderophore utilization locus in nontypeable Haemophilus influenzae.* BMC Microbiol, 2010. **10**: p. 113.
90. Fleury, C., et al., *Identification of a Haemophilus influenzae factor H-Binding lipoprotein involved in serum resistance.* J Immunol, 2014. **192**(12): p. 5913-23.
91. Williams, P., et al., *Utilization of enterobactin and other exogenous iron sources by Haemophilus influenzae, H. parainfluenzae and H. paraphrophilus.* J Gen Microbiol, 1990. **136**(12): p. 2343-50.
92. Shin, H.E., et al., *Prevalence of Slam-dependent hemophilins in Gram-negative bacteria.* J Bacteriol, 2024. **206**(6): p. e0002724.
93. Ekins, A., et al., *Haemophilus somnus possesses two systems for acquisition of transferrin-bound iron.* J Bacteriol, 2004. **186**(13): p. 4407-11.
94. Cozens, D., et al., *Pathogenic Mannheimia haemolytica Invades Differentiated Bovine Airway Epithelial Cells.* Infect Immun, 2019. **87**(6).
95. Deschner, D. and J.E. Hill, *Identification of genes that differentiate Mannheimia haemolytica genotypes 1 and 2 using a pangenome approach.* PLoS One, 2025. **20**(10): p. e0325338.
96. Samaniego-Barrón, L., et al., *Two outer membrane proteins are bovine lactoferrin-binding proteins in Mannheimia haemolytica A1.* Vet Res, 2016. **47**(1): p. 93.
97. Graham, M.R. and R.Y. Lo, *A putative iron-regulated TonB-dependent receptor of Mannheimia (Pasteurella) haemolytica A1: possible mechanism for phase variation.* Vet Microbiol, 2002. **84**(1-2): p. 53-67.
98. Clawson, M.L., et al., *Differences between predicted outer membrane proteins of genotype 1 and 2 Mannheimia haemolytica.* BMC Microbiol, 2020. **20**(1): p. 250.
99. Gioia, J., et al., *The genome sequence of Mannheimia haemolytica A1: insights into virulence, natural competence, and Pasteurellaceae phylogeny.* J Bacteriol, 2006. **188**(20): p. 7257-66.
100. Cox, A.J., et al., *Functional characterization of HgbB, a new hemoglobin binding protein of Pasteurella multocida.* Microb Pathog, 2003. **34**(6): p. 287-96.
101. Li, N., et al., *High and low-virulent bovine Pasteurella multocida capsular type A isolates exhibit different virulence gene expression patterns in vitro and in vivo.* Vet Microbiol, 2016. **196**: p. 44-49.
102. Paustian, M.L., et al., *Transcriptional response of Pasteurella multocida to defined iron sources.* J Bacteriol, 2002. **184**(23): p. 6714-20.
103. Boyce, J.D., et al., *Analysis of the Pasteurella multocida outer membrane sub-proteome and its response to the in vivo environment of the natural host.* Proteomics, 2006. **6**(3): p. 870-80.
104. Jatuponwiphat, T., et al., *Iron-associated protein interaction networks reveal the key functional modules related to survival and virulence of Pasteurella multocida.* Microb Pathog, 2019. **127**: p. 257-266.
105. Bosch, M., et al., *Pasteurella multocida contains multiple immunogenic haemin- and haemoglobin-binding proteins.* Vet Microbiol, 2004. **99**(2): p. 103-12.
106. Hooda, Y., C.C.L. Lai, and T.F. Moraes, *Identification of a Large Family of Slam-Dependent Surface Lipoproteins in Gram-Negative Bacteria.* Front Cell Infect Microbiol, 2017. **7**: p. 207.
107. Wu, J.R., et al., *Protective immunity conferred by recombinant Pasteurella multocida lipoprotein E (PlpE).* Vaccine, 2007. **25**(21): p. 4140-8.
108. Kadner, R.J. and K.J. Heller, *Mutual inhibition of cobalamin and siderophore uptake systems suggests their competition for TonB function.* J Bacteriol, 1995. **177**(17): p. 4829-35.

109. Wookey, P. and H. Rosenberg, *Involvement of inner and outer membrane components in the transport of iron and in colicin B action in Escherichia coli.* J Bacteriol, 1978. **133**(2): p. 661-6.
110. Buchanan, S.K., et al., *Crystal structure of the outer membrane active transporter FepA from Escherichia coli.* Nat Struct Biol, 1999. **6**(1): p. 56-63.
111. Dean, C.R. and K. Poole, *Cloning and characterization of the ferric enterobactin receptor gene (pfeA) of Pseudomonas aeruginosa.* J Bacteriol, 1993. **175**(2): p. 317-24.
112. Moynié, L., et al., *Structure and Function of the PiuA and PirA Siderophore-Drug Receptors from Pseudomonas aeruginosa and Acinetobacter baumannii.* Antimicrob Agents Chemother, 2017. **61**(4).
113. Bonhivers, M., et al., *FhuA, an Escherichia coli outer membrane protein with a dual function of transporter and channel which mediates the transport of phage DNA.* Biochimie, 1998. **80**(5-6): p. 363-9.
114. Mikael, L.G., et al., *Molecular cloning and characterization of the ferric hydroxamate uptake (fhu) operon in Actinobacillus pleuropneumoniae.* Microbiology (Reading), 2002. **148**(Pt 9): p. 2869-2882.
115. Klee, S.R., et al., *Molecular and biological analysis of eight genetic islands that distinguish Neisseria meningitidis from the closely related pathogen Neisseria gonorrhoeae.* Infect Immun, 2000. **68**(4): p. 2082-95.
116. Sauer, M., K. Hantke, and V. Braun, *Ferric-coprogen receptor FhuE of Escherichia coli: processing and sequence common to all TonB-dependent outer membrane receptor proteins.* J Bacteriol, 1987. **169**(5): p. 2044-9.
117. Parkhill, J., et al., *Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica.* Nat Genet, 2003. **35**(1): p. 32-40.
118. Mademidis, A. and W. Köster, *Transport activity of FhuA, FhuC, FhuD, and FhuB derivatives in a system free of polar effects, and stoichiometry of components involved in ferrichrome uptake.* Mol Gen Genet, 1998. **258**(1-2): p. 156-65.
119. Tanaka, E., et al., *High-Level Quinolone-Resistant Haemophilus haemolyticus in Pediatric Patient with No History of Quinolone Exposure.* Emerg Infect Dis, 2022. **28**(1): p. 104-110.
120. Zárate-Bonilla, L.J., et al., *Computational modeling and preliminary iroN, fepA, and cirA gene expression in Salmonella Enteritidis under iron-deficiency-induced conditions.* Poult Sci, 2014. **93**(1): p. 221-30.
121. Long, H., Y. Feng, and Z. Zong, *Global molecular epidemiology of the incomplete CirA protein related to cefiderocol resistance in Klebsiella pneumoniae: a genome-based study.* Microbiol Spectr, 2025. **13**(5): p. e0141024.
122. Song, Q., et al., *Proteomics analysis reveals that CirA in Aeromonas hydrophila is involved in nutrient uptake.* Environ Microbiol, 2023. **25**(5): p. 977-989.
123. Bäumler, A., et al., *Survey on newly characterized iron uptake systems of Yersinia enterocolitica.* Zentralbl Bakteriol, 1993. **278**(2-3): p. 416-24.
124. Helminen, M.E., et al., *A major outer membrane protein of Moraxella catarrhalis is a target for antibodies that enhance pulmonary clearance of the pathogen in an animal model.* Infect Immun, 1993. **61**(5): p. 2003-10.
125. Pettersson, A., et al., *Molecular characterization of FrpB, the 70-kilodalton iron-regulated outer membrane protein of Neisseria meningitidis.* Infect Immun, 1995. **63**(10): p. 4181-4.
126. Thompson, S.A., et al., *Neisseria meningitidis produces iron-regulated proteins related to the RTX family of exoproteins.* J Bacteriol, 1993. **175**(3): p. 811-8.
127. Carson, S.D., et al., *Ferric enterobactin binding and utilization by Neisseria gonorrhoeae.* J Bacteriol, 1999. **181**(9): p. 2895-901.
128. Delany, I., et al., *Iron-dependent transcription of the frpB gene of Helicobacter pylori is controlled by the Fur repressor protein.* J Bacteriol, 2001. **183**(16): p. 4932-7.
129. Mulé, M.P., et al., *Iron-dependent gene expression in Actinomyces oris.* J Oral Microbiol, 2015. **7**: p. 29800.

130. Kortekaas, J., et al., *Shielding of immunogenic domains in Neisseria meningitidis FrpB (FetA) by the major variable region.* Vaccine, 2007. **25**(1): p. 72-84.
131. Wang, J., et al., *Riemerella anatipestifer UvrC is required for iron utilization, biofilm formation and virulence.* Avian Pathol, 2024. **53**(4): p. 247-256.
132. Nikaido, H. and E.Y. Rosenberg, *Cir and Fiu proteins in the outer membrane of Escherichia coli catalyze transport of monomeric catechols: study with beta-lactam antibiotics containing catechol and analogous groups.* J Bacteriol, 1990. **172**(3): p. 1361-7.
133. Higgs, P.I., R.A. Larsen, and K. Postle, *Quantification of known components of the Escherichia coli TonB energy transduction system: TonB, ExbB, ExbD and FepA.* Mol Microbiol, 2002. **44**(1): p. 271-81.
134. Prado, M.E., S.M. Dabo, and A.W. Confer, *Immunogenicity of iron-regulated outer membrane proteins of Pasteurella multocida A:3 in cattle: molecular characterization of the immunodominant heme acquisition system receptor (HasR) protein.* Vet Microbiol, 2005. **105**(3-4): p. 269-80.
135. Létoffé, S., et al., *Activities of the Serratia marcescens heme receptor HasR and isolated plug and beta-barrel domains: the beta-barrel forms a heme-specific channel.* J Bacteriol, 2005. **187**(13): p. 4637-45.
136. Létoffé, S., J.M. Ghigo, and C. Wandersman, *Secretion of the Serratia marcescens HasA protein by an ABC transporter.* J Bacteriol, 1994. **176**(17): p. 5372-7.
137. Kumar, R., et al., *The hemophore HasA from Yersinia pestis (HasAyp) coordinates hemin with a single residue, Tyr75, and with minimal conformational change.* Biochemistry, 2013. **52**(16): p. 2705-7.
138. Grossman, A.S., et al., *Bacterial hemophilin homologs and their specific type eleven secretor proteins have conserved roles in heme capture and are diversifying as a family.* J Bacteriol, 2024. **206**(6): p. e0044423.
139. Morgan, W.T. and U. Muller-Eberhard, *Interactions of porphyrins with rabbit hemopexin.* J Biol Chem, 1972. **247**(22): p. 7181-7.
140. Fournier, C., A. Smith, and P. Delepelaire, *Haem release from haemopexin by HxuA allows Haemophilus influenzae to escape host nutritional immunity.* Mol Microbiol, 2011. **80**(1): p. 133-48.
141. Cope, L.D., et al., *Involvement of HxuC outer membrane protein in utilization of hemoglobin by Haemophilus influenzae.* Infect Immun, 2001. **69**(4): p. 2353-63.
142. Moslavac, S., et al., *Conserved pore-forming regions in polypeptide-transporting proteins.* Febs j, 2005. **272**(6): p. 1367-78.
143. Baelen, S., et al., *Structure of the secretion domain of HxuA from Haemophilus influenzae.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2013. **69**(Pt 12): p. 1322-7.
144. Wong, J.C., et al., *Affinity, conservation, and surface exposure of hemopexin-binding proteins in Haemophilus influenzae.* Infect Immun, 1995. **63**(6): p. 2327-33.
145. Bachmann, M.F. and R.M. Zinkernagel, *The influence of virus structure on antibody responses and virus serotype formation.* Immunol Today, 1996. **17**(12): p. 553-8.
146. Forthal, D.N., *Functions of Antibodies.* Microbiol Spectr, 2014. **2**(4): p. 1-17.
147. Hou, N., et al., *Low-Complexity Repetitive Epitopes of Plasmodium falciparum Are Decoys for Humoural Immune Responses.* Front Immunol, 2020. **11**: p. 610.
148. LaCross, N.C., C.F. Marrs, and J.R. Gilsdorf, *Otitis media associated polymorphisms in the hemin receptor HemR of nontypeable Haemophilus influenzae.* Infect Genet Evol, 2014. **26**: p. 47-57.
149. Xie, J., et al., *Identification of new genetic regions more prevalent in nontypeable Haemophilus influenzae otitis media strains than in throat strains.* Journal of clinical microbiology, 2006. **44**(12): p. 4316-4325.

150. Morton, D.J., et al., *Utilization of myoglobin as a heme source by Haemophilus influenzae requires binding of myoglobin to haptoglobin.* FEMS Microbiology Letters, 2006. **258**(2): p. 235-240.
151. Morton, D.J., et al., *Differential utilization by Haemophilus influenzae of haemoglobin complexed to the three human haptoglobin phenotypes.* FEMS Immunol Med Microbiol, 2006. **46**(3): p. 426-32.
152. Hayashida, H., K. Poulsen, and M. Kilian, *Differences in iron acquisition from human haemoglobin among strains of Actinobacillus actinomycetemcomitans.* Microbiology (Reading), 2002. **148**(Pt 12): p. 3993-4001.
153. Srikumar, R., et al., *Molecular cloning of haemoglobin-binding protein HgbA in the outer membrane of Actinobacillus pleuropneumoniae.* Microbiology (Reading), 2004. **150**(Pt 6): p. 1723-1734.
154. Morton, D.J., et al., *Effect of multiple mutations in the hemoglobin- and hemoglobin-haptoglobin-binding proteins, HgpA, HgpB, and HgpC, of Haemophilus influenzae type b.* Infect Immun, 1999. **67**(6): p. 2729-39.
155. Ren, Z., et al., *hgpB, a gene encoding a second Haemophilus influenzae hemoglobin- and hemoglobin-haptoglobin-binding protein.* Infect Immun, 1998. **66**(10): p. 4733-41.
156. Siddaramappa, S., et al., *Horizontal gene transfer in Histophilus somni and its role in the evolution of pathogenic strain 2336, as determined by comparative genomic analyses.* BMC Genomics, 2011. **12**: p. 570.
157. Ren, Z., et al., *Role of CCAA nucleotide repeats in regulation of hemoglobin and hemoglobin-haptoglobin binding protein genes of Haemophilus influenzae.* J Bacteriol, 1999. **181**(18): p. 5865-70.
158. Pettigrew, M.M., et al., *Haemophilus influenzae genome evolution during persistence in the human airways in chronic obstructive pulmonary disease.* Proc Natl Acad Sci U S A, 2018. **115**(14): p. E3256-e3265.
159. Liu, Y., et al., *The protective efficacy of forty outer membrane proteins based DNA vaccines against Aeromonas hydrophila in zebrafish.* Aquaculture Reports, 2022. **27**: p. 101381.
160. Stojiljkovic, I., et al., *The Neisseria meningitidis haemoglobin receptor: its role in iron utilization and virulence.* Mol Microbiol, 1995. **15**(3): p. 531-41.
161. Richardson, A.R. and I. Stojiljkovic, *HmbR, a hemoglobin-binding outer membrane protein of Neisseria meningitidis, undergoes phase variation.* Journal of bacteriology, 1999. **181**(7): p. 2067-2074.
162. Sevestre, J., et al., *Differential expression of hemoglobin receptor, HmbR, between carriage and invasive isolates of Neisseria meningitidis contributes to virulence: lessons from a clonal outbreak.* Virulence, 2018. **9**(1): p. 923-929.
163. Deslandes, V., et al., *Transcriptional profiling of Actinobacillus pleuropneumoniae under iron-restricted conditions.* BMC Genomics, 2007. **8**: p. 72.
164. Russo, T.A., et al., *IroN functions as a siderophore receptor and is a urovirulence factor in an extraintestinal pathogenic isolate of Escherichia coli.* Infect Immun, 2002. **70**(12): p. 7156-60.
165. Russo, T.A., U.B. Carlino, and J.R. Johnson, *Identification of a new iron-regulated virulence gene, ireA, in an extraintestinal pathogenic isolate of Escherichia coli.* Infect Immun, 2001. **69**(10): p. 6209-16.
166. Li, Y., et al., *Iron-regulated gene ireA in avian pathogenic Escherichia coli participates in adhesion and stress-resistance.* BMC Vet Res, 2016. **12**(1): p. 167.
167. Kumar, P., S. Sannigrahi, and Y.L. Tzeng, *The Neisseria meningitidis ZnuD zinc receptor contributes to interactions with epithelial cells and supports heme utilization when expressed in Escherichia coli.* Infect Immun, 2012. **80**(2): p. 657-67.
168. Wheeler, R., *Outer membrane proteomics of Pasteurella multocida isolates to identify putative host-specificity determinants.* Bioscience Horizons: The International Journal of Student Research, 2009. **2**(1): p. 1-12.

169. Stork, M., et al., *An outer membrane receptor of Neisseria meningitidis involved in zinc acquisition with vaccine potential.* PLoS Pathog, 2010. **6**(7): p. e1000969.
170. Si, M., et al., *The Type VI Secretion System Engages a Redox-Regulated Dual-Functional Heme Transporter for Zinc Acquisition.* Cell Rep, 2017. **20**(4): p. 949-959.
171. Mikael, L.G., et al., *fhuA of Actinobacillus pleuropneumoniae encodes a ferrichrome receptor but is not regulated by iron.* Infect Immun, 2003. **71**(5): p. 2911-5.
172. Leasure, C.S., et al., *Maintenance of heme homeostasis in Staphylococcus aureus through post-translational regulation of glutamyl-tRNA reductase.* J Bacteriol, 2023. **205**(9): p. e0017123.
173. Josts, I., et al., *Structural insights into a novel family of integral membrane siderophore reductases.* Proc Natl Acad Sci U S A, 2021. **118**(34).
174. Power, P.M., et al., *Simple sequence repeats in Haemophilus influenzae.* Infect Genet Evol, 2009. **9**(2): p. 216-28.
175. Grinter, R., et al., *FusC, a member of the M16 protease family acquired by bacteria for iron piracy against plants.* PLoS Biol, 2018. **16**(8): p. e2006026.
176. Bateman, T.J., et al., *A Slam-dependent hemophore contributes to heme acquisition in the bacterial pathogen Acinetobacter baumannii.* Nature Communications, 2021. **12**(1): p. 6270.
177. Revanaiah, Y., et al., *Sequence and structural analysis of OmpW protein of Pasteurella multocida strains reveal evolutionary conservation among members of Pasteurellaceae along with its homologues.* Gene Reports, 2018. **14**.
178. Hooda, Y., et al., *Slam is an outer membrane protein that is required for the surface display of lipidated virulence factors in Neisseria.* Nat Microbiol, 2016. **1**: p. 16009.
179. Lv, T., et al., *Outer membrane protein OmpU is related to iron balance in Vibrio alginolyticus.* Microbiol Res, 2020. **230**: p. 126350.
180. Nakasone, N. and M. Iwanaga, *Characterization of outer membrane protein OmpU of Vibrio cholerae O1.* Infect Immun, 1998. **66**(10): p. 4726-8.
181. Li, H., W. Zhang, and C. Dong, *Crystal structure of the outer membrane protein OmpU from Vibrio cholerae at 2.2 Å resolution.* Acta Crystallogr D Struct Biol, 2018. **74**(Pt 1): p. 21-29.
182. Zhang, P., et al., *OmpW is positively regulated by iron via Fur, and negatively regulated by SoxS contribution to oxidative stress resistance in Escherichia coli.* Microb Pathog, 2020. **138**: p. 103808.
183. Shahryari, S., P. Mohammadnejad, and K.A. Noghabi, *Screening of anti-Acinetobacter baumannii phytochemicals, based on the potential inhibitory effect on OmpA and OmpW functions.* R Soc Open Sci, 2021. **8**(8): p. 201652.
184. Hong, H., et al., *The outer membrane protein OmpW forms an eight-stranded beta-barrel with a hydrophobic channel.* J Biol Chem, 2006. **281**(11): p. 7568-77.
185. Bisio, H., et al., *A New Class of Thioredoxin-Related Protein Able to Bind Iron-Sulfur Clusters.* Antioxid Redox Signal, 2016. **24**(4): p. 205-216.
186. Anjou, C., et al., *From ubiquity to specificity: The diverse functions of bacterial thioredoxin systems.* Environ Microbiol, 2024. **26**(6): p. e16668.
187. Ysebaert, C., et al., *A Protein E-PilA Fusion Protein Shows Vaccine Potential against Nontypeable Haemophilus influenzae in Mice and Chinchillas.* Infect Immun, 2019. **87**(8).
188. Behrouzi, A., et al., *Recombinant truncated E protein as a new vaccine candidate against nontypeable H. influenzae: Its expression and immunogenic evaluation.* Microb Pathog, 2017. **110**: p. 431-438.
189. Al Jubair, T., et al., *Haemophilus influenzae stores and distributes hemin by using protein E.* Int J Med Microbiol, 2014. **304**(5-6): p. 662-8.
190. Stante, M., et al., *Four Novel Caudoviricetes Bacteriophages Isolated from Baltic Sea Water Infect Colonizers of Aurelia aurita.* Viruses, 2023. **15**(7).
191. Khosravi, A., et al., *Phage-Mediated Iron Acquisition by Pseudomonas aeruginosa.* bioRxiv, 2026.

192. Trindade, I.B., et al., *Structure and reactivity of a siderophore-interacting protein from the marine bacterium Shewanella reveals unanticipated functional versatility.* J Biol Chem, 2019. **294**(1): p. 157-167.
193. Liu, C., Y. Han, and Q. Ma, *Structural analysis of the siderophore-interacting protein from Vibrio anguillarum and its implications in classification of Vibrio homologs.* Biochem Biophys Res Commun, 2024. **739**: p. 150979.
194. Drew, D., et al., *Structures and General Transport Mechanisms by the Major Facilitator Superfamily (MFS).* Chem Rev, 2021. **121**(9): p. 5289-5335.
195. Schryvers, A.B., *Targeting bacterial transferrin and lactoferrin receptors for vaccines.* Trends Microbiol, 2022. **30**(9): p. 820-830.
196. Ewasechko, N.F., S. Chaudhuri, and A.B. Schryvers, *Insights from targeting transferrin receptors to develop vaccines for pathogens of humans and food production animals.* Front Cell Infect Microbiol, 2022. **12**: p. 1083090.
197. Cornelissen, C.N., et al., *Antigenic and sequence diversity in gonococcal transferrin-binding protein A.* Infect Immun, 2000. **68**(8): p. 4725-35.
198. Fegan, J.E., et al., *Utility of Hybrid Transferrin Binding Protein Antigens for Protection Against Pathogenic Neisseria Species.* Frontiers in Immunology, 2019. **10**.
199. Greenawalt, A.N., et al., *Point Mutations in TbpA Abrogate Human Transferrin Binding in Neisseria gonorrhoeae.* Infect Immun, 2022. **90**(11): p. e0041422.
200. Moraes, T.F., et al., *Insights into the bacterial transferrin receptor: the structure of transferrin-binding protein B from Actinobacillus pleuropneumoniae.* Mol Cell, 2009. **35**(4): p. 523-33.
201. Renauld-Mongénie, G., et al., *Role of transferrin receptor from a Neisseria meningitidis tbpB isotype II strain in human transferrin binding and virulence.* Infect Immun, 2004. **72**(6): p. 3461-70.
202. Tsiftsoglou, S.A. and E. Gavriilaki, *A potential bimodal interplay between heme and complement factor H 402H in the deregulation of the complement alternative pathway by SARS-CoV-2.* Infect Genet Evol, 2024. **126**: p. 105698.
203. Cerveny, L., et al., *Tetratricopeptide repeat motifs in the world of bacterial pathogens: role in virulence mechanisms.* Infect Immun, 2013. **81**(3): p. 629-35.
204. Karpenahalli, M.R., A.N. Lupas, and J. Söding, *TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences.* BMC Bioinformatics, 2007. **8**: p. 2.
205. Whelan, F., et al., *Periscope Proteins are variable-length regulators of bacterial cell surface interactions.* Proc Natl Acad Sci U S A, 2021. **118**(23).
206. Grove, T.Z., A.L. Cortajarena, and L. Regan, *Ligand binding by repeat proteins: natural and designed.* Curr Opin Struct Biol, 2008. **18**(4): p. 507-15.
207. Krell, T., et al., *Insight into the structure and function of the transferrin receptor from Neisseria meningitidis using microcalorimetric techniques.* J Biol Chem, 2003. **278**(17): p. 14712-22.
208. DeRocco, A.J. and C.N. Cornelissen, *Identification of transferrin-binding domains in TbpB expressed by Neisseria gonorrhoeae.* Infect Immun, 2007. **75**(7): p. 3220-32.
209. Frandoloso, R., et al., *Nonbinding site-directed mutants of transferrin binding protein B exhibit enhanced immunogenicity and protective capabilities.* Infect Immun, 2015. **83**(3): p. 1030-8.
210. Jacob-Dubuisson, F., et al., *Two-partner secretion: as simple as it sounds?* Res Microbiol, 2013. **164**(6): p. 583-95.
211. Butterton, J.R., et al., *Cloning, sequencing, and transcriptional regulation of viuA, the gene encoding the ferric vibriobactin receptor of Vibrio cholerae.* J Bacteriol, 1992. **174**(11): p. 3729-38.
212. Webster, A.C. and C.M. Litwin, *Cloning and characterization of vuuA, a gene encoding the Vibrio vulnificus ferric vulnibactin receptor.* Infect Immun, 2000. **68**(2): p. 526-34.

Protein Sequences

>HaeH_FhuH1.2
MKKSFLLTTVASSILLAMNTAYAEENVEQLDTIEVSADASQTGLSKEFAGGQVASGSRIGILG
NKSNLESSFATTSYTNKFIQDKQARSVADVLKNDPNVRVARGFGNFQESYFMRGFITNSDD
ILYNGLYGLLPRQYIASEMIERVEVQRGSSAFLNGMAPGGGSIGGTIALMPKRAQAEDLNRL
NLGYSSNERANFSADISRRLGENREFGVRLNAAHTEGESAIDNDNARNTVFHLGLDWEVKK
VRLSADLGYQKNYLGSPRPSVTLSGVTMVPSTPNPRSNWGQGWTYSEERDIFGTFRAEYD
LHPNLTAYAAYGYRYGKEANVLANLRVSNINGNGTEYRFDNTRRNIIHTSEIGLKGHFTTNE
VEHDVVLSANRYQEKRKNAYVMDWKNTFATNLYHPTFISSIAYSANALKGNDLSSPALQHR
ATLNSVAIGDTLSVLDKKLQVTLGARWQQIVSRDFAYNTGVEGGYYNDSRISPSVGVVYRIL
PELSIYGNYIESLAQGDTAPSTAVNKGETQKPYISKQKEIGIKYESQSGFGTSIALFSTEKPR
GYLDSNNYYTTKGKDRHNGAEINVYGQITSNIRLLGGVTFLQAKQRKTGSNTTEGKYTIGVP
KFQGNLGLEYDVRALEGLTLESRITYTGSSYADSQNTLKVKGWTKVDVGARYIAMIGNTPIT
LRARIDNLTNKKYWESVGGYPNYGYLVSGAPRTFSLNASIDF

>ManH_CurA1.2
MLIFQINIKNFHIFAYYIFILLLYNSSKKIIPKEIIMKKNLITTAISVALASTAFAQNEMTQLDSVTV
YSAYATPINQDKTASSVTILTEKDFSERNATYVSDVLKTVPGVAIGQQQGGRGTLTSLFLRGA
ESRHTAVVIDGVKVNPINIGNFDFGGLPISNIERIEVLRGEQSALWGSSAMGGVVYITTKSGL
YKEKPFNAEVDLGLGSNNTRDASATLSGFHNGFYYALHGDSHRTKGISALSKNHFSYTTET
GSEVKTGGASERDGFHRDNGSLRLGYDLGNKGVEVLAAQSSQTVHIDGYNSDVSGEYSR
TRNQTFKLGGYWGNEQELLKHQANISQFNSKATHFGSNARYSNEKQLNANYQLDVNFDRE
GEVTQAVSLLTDYAKTRYTSDKYLREKTLSEKSAALEYRLFTEQDHSFSISGRYTDNSQFKN
SITGRISGAYRLSPNFRLHASLGKAIKNPSITDLYGWDASYVGNPNLKAEKSRGGDLGLLIES
TDKKHSLDVTYFARVVNDLFSSKGNNCQMVSSPWGDYLQCANYQSINLEDKSRIKGVELA
YQGKWSEKLTSFANYTFTRTKDSTQKELLRRPKHVANAGVNYQISEEFGSSTSLSYVGKRI
DSSFPSRVKMPSYTLVNLGVNYQLSKQLNVYVNLNNLFNKKYENVLGYGQDGRNIYVGLK
GSF

>ManH_FhuE2
MNTLSKAIKLALYSSGIGIISTAYAETVTHLDTVEVSAETQNASSEETGRYAQKKAKGSNGM
ALSAKDTPQSVTAVTHQQIRDQNLNTIAKALEATHGVSVSLLDRGRYSFSARGFGIDKVKVD
GMDLKVNNQWTTGENTGNSVLYDRVEVVRGATGLTTGAGDPSASVNLVRKRANSISRYTI
LEAGIGRYADFNTMIDHSQALNNSGSVRGRFIAEHKNGNTFIKNEKERYTTLYGNIEADLTDT
TKLGLGVSYQKEKRDAALWGGLPAFYSDGGETNWSRSKNASPEWAYWDNESVNYFADIT
QKLGEKWELALKGNYRVGKHNSELFYFSGSSLNRQDGLGWSPWPGKFKNDNTQSNVQL
DLTGSFSAWGLEHDAAIGIQYNRAHRNSYAADNSTFDGAFDFNNWNGIYARPNWGSMSEK
YYQNLKEFGVYVSTRLRLTDRLSMLLGSRISSYRTEGTFYGSTQNFHANDVWTPYAGLTFD
INPQNTVYVSYTDIFKPQSQVDINNKLLSPVTGETYEAGWKGSYFNDNLQTQISVFETRQDN
LAQSAGINPNTLLAYYTASKGAKVRGFEVEASGNVTEKLKLSAGYTQWKGDDASGNPINTT
QPRKQFKFFATYDMNRLVQGLTVGAGINWQSRIYTKHATYGEYGQKAYTLVNLMARYQFN
DNFSAQLNIDNLFDKKYRNALSFSQYSYGEPFYTRLNLRYEF

>ActP_FhuI2
MKKTFFYSTVASAVALAISPAIAQETAVLDEVSVVGSVSKAGKVEYMTPKSVSVLTEEQIKEK
GYKQADEIGRYEAGILSQPYGGDIDTHDWIRIRGFSPTLRIDGTAMYSFGFYHWQPNMYGL
ESIEVVKGADSLTYGSAQSGGLINLVSKRPTKEPKGELKFTLGNRNERGFSADYSNVISENV
RFRLVGDYNRRDGEPKGSYMDAYYFAPSLSWDISKNTSLTLLASIQHDDGRPSSGFFPAGY
ERLSSLGLSNRTNFAENVPEYLRRNQYSVGYEFTHKFSDDLVYQSNYRYAYSNKDQFSASI
GSDGSSLSTMTARGKTFTHSLDNRLSKTWAIDRIENILTVGFDYTTGKNYAANGYGSEILSN
LYGNYPTAAQPFGAVSPYVNKERQLGVYLQNQFRFDNRLLVNLGARRDFAKGDYGDEGY
DINHNSYSAGVMYITDSGLSPFYNYSESFRPISGSNINGQIVRYKPYEGTQHEVGFKYLPSF
ADGTLSIAYFDLKEKNGLVQGVVNGQVTQLQTEKAKSKGVELQADMKLNDNISTLLAYTYT

NAKVNNARKVIIPRHTYSAFVNYSFTNEALAGLKIGTGIRYIGTSTNGVTEVKIPAYTLVDLMA
KYDLNKNWTAQVNVSNLSNKKYISACDTYWCYMGEGRKVTANLTYKF

>AggA_FecA2.2
MKKTLLYTAIAMACLPAYAEEVFTLGQIEVIADRSTDLSTTRIEQADLQKNNQTNVAEVAKTT
PGVFLDRSGARNEHNLLVRGFKANRVPVFIDGIPVYVPYDGNMDIGRFTTFDLSRIDISKGA
SSVLYGANTLGGAVNLITQKPTKPFEGTIGYGFAHGRSGSTGNNQTYFNLGSKQDYFYAQL
SGSFLEKQGLQLSRHYHRVLPSGDDGARAENSVQRDKKLSLKVAFTPNATDEYALVLSTQ
KSAKQQLLYSGKTACSERYWRWPQWDKDSIYFLSHTQFDSHNIYLNTKAFFDTFKNELDSF
DDAGFSRQTGSNRNSYYRDYAYGTGFELGGDLTSNDTLKFSAILKQDVHREHNSNEPTAV
DKDRTYSFGLENTYRFSERTKLITGMSFDRREAKRAENYQARCPTTNVRSQSICPFNVADK
NAFNYQIKLVHSFDEQDEFSLGFAKKTRFPTMKDRYSYRLNRAEPNPFLDPEIAYHYEAGYT
RTFGNWLRLDGALFYSEVRDAIELVAHPTISNKEQNQNYGKEVFKGVELAAAIFATDNLTLG
ANYTYTRAKNKTYTNFIVRDIPQHKFFAYVDWKIVPNLSLYVSQEAEHGRYSRDGFGSRAP
TVRLSGFGNTNAKLTYNVTEQFIVEAGVFNLFDKNYYYEAGYPEAGRVYFSNLKYRF

>HaeP_FepA2
MPGEHSMYKGNHLKTTSFFALSTLTIALFSSYSYGEEKLDEIKVTAENQVKQSLGSSLVTAK
DLEKRPATNDVSEVLRTMPGVNLTGNSSTGQRGNKRQIDIRGMGPENTLILVDGKPVTSRN
AERYGVRGERNSRGDSNWVPVEAIEKIEVLRGPSAARYGSGAMGGVVNIITKPVTNDLHGS
LSYYTNQPEDSDEGATNRVGFNLSGALIKDVLQFRLYGNWNKTQADEVGINGDPAIAGREG
VRNKDINGRLVWNINEKNKLTLDSGFSRQGNIYNGDTQNSSAGLYNNKNYPETKTLAGSKA
ETARLYRQNYALTYEGKFDHFDNKTYITFDKTKNSRLPEYLAGGPEGSYSSTSAFSDSILKN
TRFSSEFYIPFTLGVEHMLTVGFEGVHSSLDDASSMTQLLGGRRVGNKVVYDLKERLPSGI
SNVRGGHSSQTEWAVFVEDNISATQSTILTPSLRYDYNTYSGSNVSGGLNFLQSLNDDWKI
KGGIARAYKAPNLYQTNPNYLLFTLGQGCPTNVPNNKTCYFQGNSDIKPETSWNKEIGIEYD
KDGLLASVAYFRNDYRNKIVSDGEFIGLTNLGNYLYKWGNANRAVIEGFEGNLTLPLIQDKL
NWVNNFTYMHKTKNKDTGNPLSIVPKYTLNSSLSYQITDSLDAMLTYTQYGKQKSRKNPEN
RMENGNNKYVNQLAGSEDIGSYAVWGLSAGYNWKDTISIRVGVSNLFDKRIYRSSSSTNYN
AHTYNEPGRAYYVTVKYTF

>ManH_FoxA2
MKIKKYSLSLITLAIVAPQVLAETTSQAELNAIEVKEKLAATTGYIPVYTTVGTKTLTSLARTPF
SINVVSDDLMHDQAVRSVTDSVAYTSGMMGGYRGNNSLIEISIRGIGNKSDGSTIPTYLNGS
RYQTAFEIDPFFLESINVIKGPNSILYGQANPGGVMDIITKKAGSNPHRHLQFVVGNHQRYQI
GLDIEENLSNTLSARLIGQLERTNWNTQFVKEKGGAFAPSILWKPTTNTELNVYAYYAKKPK
AGDRNFLPKEGTINSVNGKKLPYDMFLSDLNYHGLKAEQMQVGYSFSHKFTPNLTFRQNA
SYHNGEENFKNLVYWGDVGSVLERRARHWDVKTQEINLDNQLEWVFSTANINHTLLAGIDY
KRVRKDLDNYMGIAPDFDWQNPVYGVTVTEPSLFSSEVKHLKQLGVYLQDQIEIGNFDFLL
GGRYDRAKNSNQDRLNSKTENNQEGKFTWRTGVVYNFENGIAPYISYSTSFVPSAEKDAQ
GNFLKPTTAGQWEVGMKYQPTQDTLLTLAGFKITQKDLANYQWQTRSYEQIGEVQTKGIEV
SLNQQLSDKFSIAASYAYLKKEITADKDGTTIGKTQWGVPRHQASLWAKYKFFDKLTAGVG
VRYMGTTFGDNQNTFTVPAYTLYDMMLAYDFMPNLNLQLNAQNLTNKKYVASCANNFSCF
YGKDRNLSATLNYHF

>ActP_FcmA1
MRFSPSKLSMAVFLATTSLAYAQENTVNLDVVSVVSENAGAKSRTNVVTLKDLNKGTKEDL
RGVLSAEPAINFGGGTGGTSQWTTIRGMGQDQIDIKVDNTYSDAQIFHHQSRFMLDPSLIKR
IDVRKGAGSASAGIGATSGSIEATTVDAKDLLREGQDFGFKLNAGVSSNKGHSQGASVYGK
SGAFDALLSGNWETAENYKGGKGYSNKEGSDTVKNSALGQRSLLAKFGVDINEDHRVVLS
HRQERYHGERALREEFDFSQDWTTTASLTSEQRAAGYQLGAAAGVDRNGRPTRYLLDKN
GALVANDANNDPRYRITTQNTTNLEYIGKNLGLINEVKANAYLLENSREETDTNSKTKVVIKG
ANINFDSEIGSDHFVKYGVNYRHQEGKPNSLEAFAENRRTRRLEATGVTNNQEKEDIGAYV

EGIWGFGPVTMTTGLRYDHFDFKANDGKKVSDGAFSPSVGLIWEVVQGLSLNTNLNYATR
SPRFYEVALSGSSARSVADDLKAEKARNTEIGFNYDFNDQISLNGSYFWQKVKDAHAVLNN
TLVNSALLRNKGYELGAAYKYSGFTFRLGVADSKPETWVFNDASLDSTVYAVATGRTWTT
GLSYKFENPSLEIGWKGRFVEGETGTPSRGSNATTAPVKQSGYGVNDFYASWDANKNLTL
NFAVNNAFNKNYKSHSQRAGATSLPSAGRDFRVNVNYTF

>GalA_Fup3
MKAKKIFFYISISLSFSLYVNANSDFSDDDTLEQVEVITELERYKATNKLKSDVNLSLLGQQTA
FTAPISVVNYNQQAIEQTEARNIVDAIAKMDASVMNFGGETNTLSGIYVRGLQLDARQFSVN
GLAGLHSSYNSPLSAVASAQLIKGASTATVGMDPEGSSGASVNIQTKRATDTPINILGMRWY
SQDRTQPYFDFGRRFGNNNQWGIRISGLYRDGDTTREQFSERNKEIAIGADYQGDKLRAAI
DYMYTKRATKGGRARIQDIQDLDFALATAPDGAINLIPHWSGQTTEDQTVMATFEYDLPHQ
MILSGGIGHMASKYYGAFGQIRMTDPQTYNIQSMRAMDYKINTTSANLKLHGDFNTHSISHN
WNIGWDSVIRTRDFKQSPVIDYKKVTTFRNINIYHPILPSAPTYKALIQSTDEKVQAHSVAFAD
TVGLFEDVLRLTLGGRYQWIKQTDYQSNIHANSHRFSPMITLAYQPTSDLIWYANYLEDLEP
GAVDEDGVMSKPLVSKQIELGVRKNWGEQYTTTFSLYQLSRPGTITSKAANNYPRKAGEQ
QGKERNRGIEFNLYANLLNGHLRPNLGISYNKGELIDYANYAGQIINGVQVASPRIIAKAGIE
WDTPFIQGLTLNSAIQYYGSSYQDYAKKYKFPSYTTVDIGAKYTFKLQQQQFTVKLGVENLF
NKNYWQVQRGRYDRSFAVLGMPRTYWANVEFSF

>GalA_GasA1
MKKTSFVLLPLAAFISGVVVADEADNLDAIEVVSDNLSPQKASLSAGSLAKVRQATTTADILR
SVPGVNVNGARSVVQRYSIRGVSEEYLTVTVDGARQNGYAFHHAGNYGIDPDILKRVNVDV
GANSVVTGAGSLGGAIRFETVEAADLLADGENFGAKFKWGYGSNADSNQAATTLYGRAGG
LDVLGYFNYRHQENGKDGNGVENANKGHLSNYLFKAKYNIDDAQWVKFSAERYNNTAVSC
YRANFNFCLGDVPKPGEQGYNETNHGKALTELLRKTYTLSYGIDPKDNPYLNFKANLYNTE
TDMSSMGKPQTNIRTVGGTVSNISEIDLANTHHSITFGGEYYNTKSRALNAPDKSYQPRIDS
TSVYLEDKIAVGDFLITPGVRYDYYQADLDKDFDKSYKRFSKALGLKYLLTDDLAVFANYTEI
FKGPDVGEVYLRGTRNYLPSIDAIRGNNKEAGISFAKADIFGNDDFSLTAKYFQTKYNHFNT
NTITEQGSGYQDIGEVKVKGVEVSTAYRINNFSINAGYARARSEQLDPLFGFYNLTALPDTG
DKYTLGLSYVMPDYGVELGWNTIWVRNITTYTAGIERCKGKNCDRTPKKMSTYKDSYSVSN
VYATWSPKQLPSLELTAGIDNIFDKAYKDQATKYYAASDNELGRNYKFTVSYKF

>PasM_HasR1
MQKQQPYPIHLGIFLMLGLPTWAFSQANLEKSTINKLETILVNESEEKNKFDENLIKTYLSSG
SYSYLSQSDISTFRGSSVGDFLSGVPGVIVGNKRNSGALSVNIRGIANENRVPVWIDKGLQS
VPSYQGYAGSSTRTYLDPDLISQVEIEKGPSLQMDATGATGGVVRVDTLRWQDIIPQGKNW
GVRLKLGTMTNTVSPPPYYTRGGYQTKYISKCLSNDTGLCQTQTYAPNARYSSHGFDLNA
YNYSLAFANKWQNADLVLAYAKRKQGNYFVGRHGQTPVIESIKFEEDSVEVKEPRVHEDVE
IGSLTFKENRSTLYRPGEEALNTSQDNTSYLAKINVYNDVHRLGLAYRHYHSRFGEIMSSILN
FRAYGALQGEGTEVKVDSYHANYSYNPTTPYVNLNVNAYFTDSDSSNFTPFIEEYGYSLSS
RHAHFLVSKQKGLSIENTSIFQLNDKPFTLKYGLAHSYERIYQPRNAQARVRAKGYPEDAIG
PLYIRDGKRKEWSAFVAANYPITSWLKADIGLRYLQSTIYDYIVRTERVNIGGALVPNPNGPG
NIWVEKYKDVVHKQAPVKNKGMSPIVMLTFEPINGVQIYTKYAEALRSPSLFQATKGWSMS
ATADNLEQLRPERAKNWEAGINLFYENLGGKDNILGFKLAYFNNRIKDYLTRSYSPKDKVTQ
TINIQSAQFKGIELSAYYDMGKFYTKLAGTYYTKTKFCLTAEQAGKGEQCNSGYVYRSNLNN
AVPPRLNLHATLGTRLFEQKLDIGARYSYYSKRLVSVLSAERFVNTSSIEWAPYSLVDLYAN
YNVSNNLKLTMTMDNVFNRYYLDINNMGLNTAPGRTLHLGLEYRF
>HaeH_Hemophilin1.1
MKKLSKIAVAVMACGVVGFTQAQVVGNVSTDTNQTRYIKIKAGEKDGKAGVEIYDSSIPNDP
AVLSKTANNKGQSMALMAERADKWISHITGVAKKDSNGVVVAKLYKMPKLTLIMPDHTGLG
RLSFKQVGNLDTYYGEWENVSGQTTQEKNVSVYYAGSNPTKRLPKGKAVYQVKGINQYN

DFNRALMTGKFNVDFDAGTIEGSLSKPDLKLSIQSQINKSKATFEGEATAEGITGKSEGRFY
GAKAEGLAGMATFASKPEYNTAFGGTKN

>AviP_Hex2
MLKPMKNYFLILLGVALLRALPSYANENTEQGLQLWQDLQQKINQQQKEAALFTVPNNQKE
NEITLDGQSFSVANTEADVGQALYIAVNQQLWHYAEKFLQQYRQFPQHRKELVWFAQGAL
ARQRGELIKAEAEYQKLLESEPTFLRGQLDLARILFENKKNKEASKLFSQLVQMPLPNEVNA
LLKEYLQVLKEREQWHGELRVGYRYQKNINQSPEHYRCLLFSGSTCIVQRATPKAINAKGW
GYELNLNRHFNLIGNHGINIYFNSEGQYYPMAQEYNDTTFKTHIGYTFQDAKRELIISPLFEY
ELFANKRYYHAFGGHLEWMQSLLKKHFLNVQLEYKHLSYNSNYSAVENAGISSLYTTWYYT
PHYNTTLFIGGDGKYRNVTDKTNAYKMLGTRFGLYHQFNEKLNLTLLASIKRTNYLRYNAILE
TKRKDNELIYQAILAFPNIFYEKLTPNLIIKRTINRSNVDWLYSYKQTELQFNLHFIF

>HaeI_HxuA1
MYKLNVISLIILTTCSGAAYASTPDFPQHHKTVFGTVTIEKTTADKMTIKQGSDKAQIDWKSF
DIGQKKEVKFEQPNEHAVAYNRVIGGNASQIQGKLTANGKVYLANPNGVIITQGAEINVAGL
LATTKDLERISENSNSYQFTRRTKDRQVLKEGLVLKDGQVVKEGQVINEGNITAQDFVVLNG
DEVINKGNINVEKNSTINGKVYLSSGYNFTFTLPDSGISVALEDNTVQGIVKNEGSIKAGEITL
SAKGRKQALDSLVMNNGVLEATKVSNKNGKVVLSADNVELNNESNIKGEIVTFGADVTSNK
ELKDNIKITSKTGSKVTSPKINFTGKSVNINGNFGREDSTTHYKDEFKKLNTEVNIDVPDNENI
RIADIEDNTGTGTTGTGTSSFIQTGALSSLLANNGKVNLKGNNVNISGRIHIDSFRGSDSLLKL
TNKGHIDINNADIHSKGRLFFITSLQNEEDFKSNITITDSKINLGNGAMGLGRSVDEKDYDNR
WQKTEGSQRKKFDVKMSNVEFNQVDDVILAGGFEKVNLDKIVATGQTNFYIDGGVSRNGR
KYEYGVLDLDKRTQLSELNQGRRRWGYYYDLELDMNRAYLYRFDLFATKNTGRSTIKDTEI
NISNSNINLKNGFVHLLAEKIKLDNSKIDITFDKDNSQDTLAQTNRLGMNGKVSMINSHIKIVG
DEKEGISPTGTYATMFLIGELIGEKSSIFVKSHQGYTFKTDGNTKIAGKYSKEDLKITAINTGG
RAAEEVLINGALGSADNDANIANMAFTIGDSANTKTTIENADITALAPNGGTAYLSSKDVEIEV
KPNSNFTFFELPREKNLNQTKINGASTKLSERGFARLYDKINGVRASNLSAEQLNVTDASEK
IINTKLVSSLDVEKLVSVAVCDAGNGCEEQQFGDKGNNTKVSVGELEAEQ

>HaeI_HxuB1
MKMRPRYSVIASAVSLGFVLSKSVMALGQPDTGSLNRELEQRRIQSEAKPSGELFNQTANS
PYTAQYKQGLKFPLTQVQILDRNNQEVVTDELAHILKNYVGKEVSLSDLSNLANEISEFYRH
NNYLVAKAILPPQEIEQGTVKILLLKGNVGEIRLQNHSALSNKFVSRLSNTTVNTSEFILKDEL
EKFALTINDVPGVNAGLQLSAGKKVGEANLLIKINDAKRFSSYVSVDNQGNKYTGRYRLAAG
TKVSNLNGWGDELKLDLMSSNQANLKNARIDYSSLIDGYSTRFGVTANYLDYKLGGNFKSL
QSQGHSHTLGAYLLHPTIRTPNFRLSTKVSFNHQNLTDKQQAVYAKQKRKINSLTAGIDGS
WNLIKDGTTYFSLSTLFGNLANQTSEKKHNAVENFQPKSHFTVYNYRLSHEQILPKSFAFNI
GINGQFADKTLESSQKMLLGGLSGVRGHQAGAASVDEGHLIQTEFKHYLPVFSQSVLVSSL
FYDYGLGKYYKNSQFLEQGVKNSVKLQSVGAGLSLSDAGSYAINVSVAKPLDNNINNADKH
QFWLSMIKTF

>HaeI_HxuC1
MRFSKLSLAIATTLVTANALAQSVELDSINVIATRDPSRFAYTPEKQSKDSLLSKQATSVAAA
LEDIPNVDIRGGSRSIAQKPNIRGLSDNRVVQVIDGVRQNFDLAHRGSYFLPMSLIQEIEVIK
GPSSSLWGSGALGGVVAMRTPNALDLLKNNDKFGVKIRQGYQTANNLSERDVSVFAANDK
FDVLISGFYNNADNLRTGKGNKLNNTAYKQFGGLAKFGWQINDANRVELSHRETRFKQTAP
SNNEVENELTNEQITDQIREFHKPNNGSPPKAKPSQEEFYSGVKTRFGSVSYLTDQQIPDQ
STVFNYYLTPDNPYLNTHIALYNNKTIEKEQRKVSGVKDQTKLTTRGINLRNSSELSHISFVY
GVDYMRDKIRTERGTNNTDAKFRAEPYNANSNTTGVYLIAHIPLFGEKLLISPSVRYDHYDT
SSKTVKYKDNHLSPATKLTWIVTNWLDFTAKYNEAFRAPSMQERFVSGAHFGANLIGGLDTI
NRFVANPNLRPETAKNKEITANLHFDSLFKQGDKFKIEATYFRNDVKDFINLKIFNDANTNKN

ASASANGALLPLPTKSQYQNITNARLSGIELQAQYQTERLTLFTNYGSTKGKDKDSGEALSN
IAASKIGVGVNYALVKDKFTVGATVTHYAAQRRVPKDHGVTYPSYILTDLRATYAPLKGEWK
NLRLDFALENLFDRKYQPAFSLMEGTGRNAKISAVYSF

>AviP_HxuD1
MKTSLSLISLAILSSSAIANNIDVNALPKNHKLTQGEASISQNNNAMTIDQKSDRVSIDWESFN
IGKQASVTFNQPTFSSIAYNRVTGADVSTIQGKLNANGRVFLANPNGVIFTKDAQVNVGAILA
TTKEVEKLAEKPTRSYTRDEANFKRKTDKEGEIRNEGKIQAEEAIYLVGDKVVNNGELRLTG
NPEEVIKLNYSGNLDGIAVKNKTVLLAAGDSFTFDLANFSISLTGKQVAGLIENKGAIISNDRIL
LTAKGEDQLKHSAVNNSGLLEANDIAIEGGKIVLTEQSDIRYGKDSRNGATLKIIQTGKDKDA
SIKIGGKITALAQTDNVKQATYHHNDRFQSGKMARDEQGKILYTVDKKTIAKAEQSENGLIG
DDGKRYTNQSDVVFVISAKNGGYQTDSRGYRPEFNKKFDALRLPYRDERILEIGYNEKNGE
LSDIKIIKRNDKLDSINSKIADILEEKIKAKDFSFLKKNELDSLTTKVIEEQKQELINKLGINEKQL
SLLSGDQTVTVEERVYTAPAVYLLKSNVTQLLKEVQQNKPDAKFEDLTLLASILNNRRPVSQ
GIDNSKTKTGFLNDEKKQIYPLMVSQSLSEKLKQNIALPLTTISTQTQLEIQNAEFNHIEGGKH
NVMILANNAKDNEATIRLQRSHINLGGGAFGIGSTTGFVPLQENILQVKQENGQLISLYNVTG
TSLKQKTLNELLKNIKFYDFDSKFQTKDNQQISLPELWSQFVENPTIQNKNVENLNFYTFVKE
NFKAFVQDKLEKLPSQVDTVLSRSTFSDGSTINSSWKKIEVAKQATNIPGLTAHQYAPFNVE
LNEVTFQKSKDFIIADGFKTVQLNKVTSKDEIGIYINAGNQRSYFSKETNLSENVKKQGYEYLI
ANLEERITRGNFSVLNERYDFNQYNWDVNNKPFTRKLAHSERATRAANRELNRAFAEWTG
DATSQRSHIDGTKITIRDSQLNSKDGFMHLLAKDVDIQNSHLTVNYARPLYDDFMPKASKIGI
NAQNIRLDHSHLFVSGEDKLNTSAMPGAAGIFLIGDVVGKNSSIYAKTYQGYSIRTSGKTTLS
GESSAKDLRITVIHTGPYPKDSGSVFGANSADYFIEGIGFDVGDIIQRDEKGNVISGTTLSNLE
FNIFAPNGARAFKSYDNSSKDYVPNIKVMKSAIFNYFERPRIVSLASNVVTGKSQALSLRDM
LRNIDQIQGVKASNLGLQQATNTNLPTSLNQIDNIIQNVSLPEERSEVEICDASSQECSTLPS
KMNGKVSVGEIL

>HaeH_Hup1
MKKVIKLNLITLCLINTLSISIVDAKAEETLDQIDVVEKNVANDKKPFIEAKAKSTREHIFKDTQT
IDQVIRSIPGAFTQQDKGSGVVSVNIRGENGLGRVNTMVDGITQTFYSTSMDSGQSGGSSQ
FGAAIEPNFIAGVDVNKSNFSGSNGINTLSGSANFRTLSVNDVITDDKPFGLIVKGMTGSNAT
KSNFMTMAAGRKWLDNGGYVGVAYGYSQREVSQDYRIGGGERLSSLGQDILAKEKEAYF
RNAGYVLNSAGQWTPDLNKNHWSCNAPSPTFHGSTEKTTSSNDLTGETETIWTDSNCITHI
EKSNIYDPEETETSVKNIYKNQARKNILQQLGAGKKPEDIEELQKGDDGIEKTDKSFERNKD
QYSVAPIEPGSLQSRSRSHLLKFEYGDDHHTLGAQLRTLDNKIGSRKIENQNYQVNYNFNN
NSYLDLNLMVAHNLGKTIYPKGGFFSGWRVTDKLITKNVANIIDINNSYTFLLPKEIDLKATLG
FNYITNEYSKNRFPDELSLFSKEASHDPGRYNYLGKFEGTRTLLPQRSVILQPSGKQKFKTV
YFDTALSKGIYHLDYSVNFTHYAFNGEYVGYENTPTQINEPILHKSGHKTAINHSATLSADIN
DYFMPFFTYSRTHRMPNIQEMFFSQVSDVGVNTALKPERAETYQLGFNTYKKGVFTKEDVL
GLKLVGYRSFIKNYIHNVYGVWWRDGVVPDWASENRFRFTIAHQNYKPIVKKSGVELEINY
DMGRFFANLSYAYQRTNQPTNYADASPRPNNASKEDILKQGYGLSRISMLPKDYGRLELGS
RWFDQKLTLGIAARYYGKSKRATIQEDYINGSQFELNTSHERTYYAIKKTEEINKQPIILDLHV
SYEPVKDLIIKAEIQNLLDKRYVDPLDSGNDAASQRYYSSLNDSICSKNNTCEDGGKDKSVL
YNFARGRTYILSLNYKF

>PasM_HemR1
MVMDKNLMKGCVFLLIVGCGIQIGLASNPNPPDVDELLPIIVNADEDNKLPGRSVLKQKNIDQ
QQADNAADLINILPGVNMAGGFRPGGQTLNINGMGDAEDVRVQLDGATKSFEKYQQGSIFI
EPELLRKVTVDKGNYSPQYGNGSFAGTVKFETKDATDFLKENQKIGGLFKYGNNSNNNQK
TYSTALVLQNEQKNIDLLLFGSVRNASNYTRPDKSKILFSKNNQKSGLIKVNWQITPEHLLTL
SSVYGIHKGWEPWAAKRDVMSRPTEKEIKRYGIDVAWKRKLVYRDQKDESYSLKYRYLPE
NNKWINLSVQLSYSKTEQNDTRHEKVTSSFLGTLGNKSWITYSDLTFDISNTSTLNIGRAEH
ELLFGLQWLKNKRNTLMYHKEGVKKADYNYGYFQPYYMPSGRQYTQAFYLQDQIKWQNF

LFTGGIRYDHINNIGQKNLAPRYNDISAGHDYSQKNYNGWSYYLGLKYDVNHYLSLFTNFSK
TWRAPVIDEQYETQYSQASVSATSLNLEKEMINQTRVGGIITLNHLFQENDAFQFRTTYFYN
RGKNEIFKTRGVNCVENALDVDNSVCPKIISNYRNLPGYVIQGAELEAYYQSSYLFGGLTYS
YVKGKRDTSPRNPWSKTSTWIAETPPRKATATLGFNIPEYYFTAGWRAEFVRKQDRSPLSN
DSKASYWALPSSKGYSLHSVFFSWSPTKIKGMNFKVTVDNLFNRPYYPYLGELASGTGRN
VKFSLTQQF

>HaeI_Hgp1.1
MLPNQPTNQPTNQPTNQPTNQPTNQDSNVSEQLEQINVSGSTENSDTKTPPKIAETVKTAK
KLEKEQAQDVKDLVRYETGITVVEAGRFGNSGFAVRGVDENRVAVQIDGLHQAETISSQGF
KELFEGYGNFNNTRNTAEIETLKQVTIRKGADSLKSGSGALGGSVSFDTKDARDYLLNKNY
YASYKRGYNTADNQNLQTLSLAGRYKYFDAIAVITSRKGHELENYGYKNYNDRIQGREREK
ADPYRRTQESKLLKFAFQPTENHRLSVVVDLYKQTSKGHDFSYTLKPNTEYRTYGEEELRH
TNDKVNRKNLAFTYENFTETPFWDTLKISYSHQKITTTARTDDYCDGNDKCALAGNPLGMK
YNQDNQLVGKDGNSVQYKDIDKETKEQKVAGSKELTKNGTWSWQKVDWDAVQAKNPGK
TIGKYCIGDDDDKNNQCTYEVKTRQKENTFEINGTKYDLLSEAHKNVISDEQRLPTNSSYLF
SCDGLNCNKKMIQGFNKDGNTVDIPFKVIEKNGKKYAKTEAVAKDQLSGPYLFMPNKKGYQ
TNLWSQRDLTSETKQINLDLTKHLELGKTQHDLSYGGLWSEMEKSMTNISGDSPMNVKWW
AQYPHNCDIFLPSSTPNGAPTLNPERTNTLCNNSNVYSFLIPVKTKTGALYFINDFRVNSHIA
FNFGYRYDRVKYDPEYIPGKTPKIPDDMVVNLYVKQPTFDDTKVNLPPEELRKKEANAAANI
KAIVQPKKFSASSYSVGTTLDPLNWLRLQAKYGKAFRAPTSDEIYFTFLHPDFSIRPNRDLQ
AETAKTKELALTLHNEIGYVTTSVFDTRYRNFIDLAYKGLYNVQRHSKLRPYHTYQNVNRPN
AKVTGLEIASQISLGKMTQFLDGLSLSYKFTYQKGKIDGNVPMNAIQPKTSVYGVSYAHPAD
KFGLDLYFTHVSAKNAEDTYNMFYKEEKKTDSTIKWRSESYTTIDLLGYIKPIKNLTLRAGVY
NLTNRKYITWDSARSIRPFGTSNMINKDTGLGINRFYAPGRNYRMSVQFEF

>PasM_Hhr1
MFVMKKSNQLKHHSLAMGLLLMLLNLSVAQATTLETIVVGGKSGDKGQAGKDEIFTKDQVT
EYKSKKEIETFHGQSVSDLFSGITGVYSGDARNGGAVDPIIRSAWGQGRIPVLVDDTEQSITI
WRGYAGVSNRNYLDPFLISSISVEKGPNLNRSLRTGSAGTVRMNTLSVDDIVKPGEKWGIE
LKVETANNSIKKRPGPYIFGEDYRKVSQLGGAELGEWAAFFKGDDRQEPRSGGRNKFFND
NAIRLAVATKQDKFEGLAAYAYRNKGNYYAGSKGGHRYGEGKPMERNVNLYSDDPYVPYI
AKIYYPKYETPNTSYESKSLLLKGKYYINDYSNVNMNFRQSDIQYGDIMPSRLGQVWGGLG
TVLEWPLANVKQKSISFNYELSPPDNKWIDLKIGAWALWNKTLTNTAGGSPMDVLFSDQQY
TNMRMEAINEAYSIVGPQYMQMHPDWNFDFDDPNFRKEVEKELVKLMKEKTEAYLNSGTA
TPNTDGIFNTQKAHAQFARDNHWGLTISNRFELHPRLNLDVMANYRRETLETETVFGLWEK
YQLPGAMSGTCSPTNPNACKVGISSDAKAAPRKGTRNEFNAGFRFEYMPTDWLMLTAGM
KYTHYKSKDKGLQEKIKDMNKEAIQNPSVIKLSLLKVRELEEVDQAIAERYKKTYNRQIELTK
QFAEKDPKYDTLLSQGMTAEEICQQAGFADSTCYPYFRAERDARFEEFKLANNFPELSAEE
KEFEEKYFNGQGSGMKYDLVKEFFWYKNEYGNYSGEAFPTFNGEVDPVSLKELVENPITG
KNEEVIPRNFEGAGNLLAGDKIYRYRLEGSAKGEHIPLTDAQWEAMKKAERKDHAWAPSFS
ATVFLSPNVRIYARYDETKRMPSLFEDTIGYSVAVLTPLYKRKPEHSKNIEIGYVHDLRGFFP
RLRRADIRLNWYKNTTKNIFDRDANYEMRQFEKRILEGLELQGRYDQGNFFADLGISYNIRN
KFCDKESAIINNDFSIRGNNIHPVCVNGGNANGYLKNTILPKYSITSNLGMRLLDEKLELGTRL
VYHTNVKDTRNKSLFEAGFPGYSNVPRWTPVFLVDAYINYQVNEHLLLTLSGTNLTDRYYL
DPMTRSSMPAPGRTVKFGLTARF

>PasM_HmbR1
MYYLSIEESAMKYPLSYKNIARSIPFLSFIAFPLYAQETTELEQITVQESATAEVNKTSPTVISK
SATTIQNEMIRDTRDLVRYTTDVGISDNGRFLKGFAMRGVEDNRVGISIDGVSLPDSEENSL
YARYGNFNNSRLSIDPELIQTIDIVRGSDSFNAGSGSLGGGVNYNTLDPQHIVKAGNSVGAL
LRGSYASKNREWVRTLGIGYVGEKFDALLMYSQRTGHEFKSRGSGPEFRYSSSQYPDPVT
QRFHNYLAKMNYQINDNQRIGLTLNGQTGGRYIDERSYTLMGSQWREADDQNERLNANLY
YIYAPSTGWLAYSKFDLDYQKTDLAAVNYKGGRHFTTDAKELNEIYDRRMKTVFTRGSVEL
NAQPVHFYGEHTLTIKGYVSQRDFKNINQDRIGIGTNYDTQYHYTIQYPIRTKQYGLSLKDHV

RWNDTFSSHLGLRYDHTKLKPKELNAPCSKACLEEGKPKPTRFSTVSTFAGLEAQLSPSW
MLGYNISTGYRVPTASEMFFSFTNAYGTWKSNPSLKPEKSINHTLSLKGNSEKGLLDLTLYQ
TNYRHFLFEQESLIQRTEMRYGRPYTYQSQEQQMVNLDKAKIYGVELKTHVNLDQMIAVIP
QGFKFYAALGYSKGKLSNNASLLSIQPLKIILGLDYEATNGKWAIFNRLTYLGEKRASDAKVY
EIKRRCTEFVTETDPWSGEQITRCKKRELYPDLSTYKHLNKSAFVFDTFGYYKITDDITFRAG
IYNLFNKKYHTWDALRGINANSTLNSVDREGKGLQRFYAPGRNYAASLEIRF

>HaeH_IreA2
MNPFKIKLSLLSLSILAALATNAAATEQNTENLDAIVVTASGTTQELRDAPASISVVNQAKLRS
HPSNRLENALQDIPGVNVSGSNANRTDIAIRGLPADYTLIMVDGRRQNTRESRPNGNGGFE
GGFIPPVSAIERIEVVRGPMSSLYGSDAMGGVVNIITKNAAKEWTGSVSTGFTVHDKSDFGD
GYHGDFFVSGPLVKDILGLQVYGGGNYRQEDNLIGASNKNKDQNLNAKLTFTPVDKQTFIF
EAGRHRLEKDETPGKTIDLTTTRGTSVTPNNPSSTRASRNHWSITHQADWDLMTSELSFSQ
EIAKREVSVNGVMNTRKPEIKNNVFDAKFMLPLSNHFFVFGGQVQQAKLQDDSVTRVLTKT
LPNGHKASRAVFENSKNSVKQSALFLEDRIDIGKDILLTLGSRLDHHEKYGSHWSPRVYAVY
HINDNFSLKGGVSKAFKAPSIREISESYVTSTEAGAGVIYGNKDLKPETSVNEEIALVYTNDN
GYNASITLFNTDFKNKLTSHYTGLTDLVMKAKLYEYANVGRANIKGTEIATHIPFNEHWNVD
MSYTYQHSKRQSDEDISASKMSLRGLPLDNTPKHAASAKLNWQMNDALSAYTRIAYKGKQI
WANPRNGDSKNHYRTRAAYTTVDLGTTYKFNPNVSLSLAVLNIGNEIGERVNTNGGSWTV
EEGRRFWTNLNVTF

>ActP_IrpA1
MSNLSPDHAVVADSLLAKQVEILRGSSTLLYASSSPAGIVNVVDKRIPTEIPEKGYEVELNSR
FDTAAKEKVGALGATFGIGKHIAVRAEGLTRHSDNYRVPGINLGERLNYVPDTYNKSKVGTL
GLSFVGEQGYIGASYSKRRDNYGLPGHNHKFDFCIGHIYGNKQGKYAYTYLYPHLIGEENIG
SNPHFHCGTDHAEDGTHSHDNPFGHDHDHTHPGPWVDLESKRFDVKAELRQPFKGIDKIK
VSYADADYYHDEKDAGVLATRYHKQLKKDQDYGKPVNIFKNRGKNARLEIYHAPLGGLTGV
WGVQYQTQKSSMHAPKDREVKFPLVENTNKQMSLFGIEQYMWDNFALEFAGRVEKQKIEI
EYDRNEIKRLQDHYRISGGKQVEPDLSPYNQNAYAYSSTLNWFFHPDYQLSFTASHNERFP
TPMELYYHGQHIATNSFEYGNKDLKKEQSNNVELGLGYQTERVGYKVNVYYNHFKNYIYNE
NLFRENQLFMRRYNQAKARFYGIEAEASYRFNDKYQATIFGDMVRGWLTNLPPLTVNSDYS
VFKDYLPKDAKPGEDYLIYRADQNTPRTPPVRLGFRFNAEFTPNWSGDLELIRTFTQRRTS
QLEYITEGNTMLNIGVAYSNKWKDLDYKISLNGTNLLNQPVYIHTSYHQFVPQTGRNFILVVN
VKF

>AviP_Iup
MKLACPLNFPLKTTALLVISVCSSKALYAEEISSSTEYMAVLPTIDVVTTQETANTKGYVGYE
EAQATRNLLTIKEMPQTIDVINIQKNKNYGTNDLSSILEGNAGVDATYDMRGENIYLRGFQAD
ANDIYRDGIRESGQVRRSTANIERVEILKGPSSILYGRSNGGGVINMVSKFANFTTSRNIGVT
YGSWNSRSVNLDVNQKINENVAVRLTSELSAAEAYRYGVRSKGRMFSPSISLQSDDGRLQ
WVGQYTYDYQWRIPDRNPAKSVYDEMGIGYRNSFFRDGDYVDDKLQVWRSDLKYFINDQ
WLVNWQLAYRQADQDFDHYFAGTYSATDKTLKQSYAWQKTRNKTFTNNITFNGEFDTASL
KHKVTIGLDYSQEERHPILAVLRNQKIDPFLSRYQWPARQHPNATVNNRHKAYSTGIFVQDL
ISLTDNVKVLLGGRYDFYRFNSTNIKQERRDTKGHSFSPNVGVVWEVTPEHTLYASFNRSF
SPYGGRSYLGISTDQKDVFNASPEYNQQYEVGIKSDWFNRTLTTTLSAYQLERRNIRYRPN
KDILDVWAVRGKDQSKGVELSLLGQLAPKWYIRSSVGWMVAKIKEDKQNPQNNNRTLNNT
GNFTSNLFVRYVPVEKFYLETGLTHLGKRYYFNGNQQTILPSFTRVDAMVGYNLNPVNITFA
VSNLFNKTYWRSDSMPGNPRSFNLRLTYMF

>ManH_Miu
MQIKQTVLAIAITPFISQLSLAQETKISSIELDTIEVNEQQENLLKQQAVMSNEDLVRYSTDMG
VPQTGRFGPQGFNIRGVEGNQVAVTVDGISQAASNNYATFSRYSYYNKSRPEIDTELMSDV
TIEKGSSTKVGNGALGGSVNYKTKTVSDILMKGRQFGGMLKYTYNGQNKENIETAGIAADG
QYADALVMYSHRQGHETQTTSKGPDIMGSGRGLPNPYKNNQYSYLAKLGLKWNDENRLE
LKTSKQRSKNAGQELSYSTVTLRDFLDVQEIETYGVTYLYTPKNSILENIKVAVTKQDTEISGI

NDQITIARGTKTDNTKRAFNTKNVEGSLGVTFKPVDLGNTEHKLSTELGYGVSDFSFDMQT
YTAGITPTDQNMQSPAKTKSFHILMGDSILLTQNLSAHLNARYEQSKLHSDDNSLSDKTFNN
WAGDIGLNYQLSPVWQIGYKLSRGFRTPTASEMFFNREVQGRGASQTFLANTNLKPETSL
GNEISLVGNGSLGSLSVAAYHTHYRNLIDLATIGNNIYQSQNVNRAKIKGVDINGTLDLNTAW
NAIPQGFELNGGFGYARGKRNDGTDLLTVQPLKALFGIGYRAVNNDWAIHLNGTYTQGKKA
KYAQQYAILANTGYLSYDKSLTTYPYLSNSATVFDLVAHKKFGERAIVKVGVYNLFNKKYWT
WDSLRTIGITSPGVANSITGEGLNRYLAPERYVMASVEFKF

>PasM_Nbr1
MEKIDMESAKNPLKKTTLALLCCSTAFSLSAKTDTDSDKNHFLTEIVVYADQNKSMSSTQSV
TQDDMKKSPVTNGNITDYLRSNPHVRYENSDQNGLQRGEIKPENISINGADYQQTTFFVDN
VNINNDMGFGSDLFDGTMATVPFANHSQGYFFDANLLSSIVVHDSNVSASLGGFAGGAVV
AKTKQYDGKDRLKFSYRTTDASWAKFKVEDKDLERFKNAIPEGSVAEFQPKYSKHFFNITA
EKGLSENLGMVIGLSRRTSSIQQSRQINPQGDRDKQTHTRRSDNALLNFNLTPNDKHRFEL
GFRYSNYRERKFFNTNIDSNVFDYHRAYGVTFSWINALQSGILTTTLAYDNFDDTRKSASTY
MKTTLTDDGEEYTEGGMGNSQLNQKNLHTSLEYAMNPFNLGSIEHSVSLGGIYQATKYRFT
RHSDAVGELYTPNWLDNNSNDIYDELTLAQRNIAKKGTVKTRYQNIALYVEDLMTWKNLEF
RAGLRLERDDYLKNTNLAPRTVFRYKPFEDTAFSVGWNRYYGRSFASMKLSEGIFKLDGH
DTFRYKDLSQFKTPYSDELSFGVEQYVANLAFHLKYILRDNKQRIVLQEEDVMLNGEKKKLR
YYQRGKDYKTNVLTFQISTQAPWEFGPTRWTSALAFDWLDSKAIDHGRGYNGSTPVILDGK
LMTYEQMLKKVNAYKETWGLRLNLDMFVPKFDLSWANTIYVKPPTTLTERVSSNTPEVYRS
YDYGTYTQWDTSLRWQPTFAEKHRPYIKLDVLNVLNKTRKGAGPNGQDLGIYTPGREFWL
EVGYEF

>PasM_OMHR1
MMLCKPNHLGLLLLISEVVFADAQLDEIHVVESVVNGKASLTALASQVRISKEKLDNVKSGTL
GEIVTNVSGVQNNYFGPNAGRPMIRGLSGQRVQIMQNDMPIHDMAAVSSNLAISANLASVK
EIQIIKSGVASVLYGGKSLGGAVNLTDDAIPQHIAEKAFRGEVEWRAGLNAPNTYQVQLQGN
NQVNFAWYLDASLSEISSRKIPGRSKDGICYDYEYLKTNTKLRSQCQVNIEKLSSINPKYFKY
ISQFYLDHYKNPDYGLSEGDQYTNMSAGGKNPLNPEYVEGSPYYAEKFGPLEEYVPYPNG
RIPNAHSSTKQLTVGASYLHDKGYIGASFFHFNTRYGVPGFSYLASRAASQRKSYLPVSVT
NETNRLNLKSQLNALVMPISKVEIQLGFQKSRDKEWLGTHLSSAFHTYNDYYRSSFEHQPL
FNHHVLGTIGVEWNRQKIFSSGKDSYLPNVNMDTKSLFILEKIDVSPFVVEFGQRWDKVKYV
VDTTNHISSRGKGGAYSKNRTFFVQNTHLGLLFQPVDTWYLKLQRSWQERAPEVNELYAD
NNHYALLIEENGDGRLKKEQSDNWELSTGVNLEQFNAELTWFQRNFDNFHYLGYTGISRA
GLAVKEWRQAPLRMNGVELDLTYQHQTPHFGDLTWHLFADYVHHKLTLAPDHIDAKRIGYL
PHLPTSRIGGDLTVQWQRWRGFTSVTHYVKQKRVTKEVDDELPAKKFTLVDLGLSYVYPLT
RATLEISLNIQNLTNREARLNNSPLRFLAPLPGRNARLGVKLMF

>HaeI_OmpM1
MLFIQIMKNIIVFLSLCLISSNIYSETRAINYNSSDIKQNPAIIEMLLLDAIDEDNINGIESLLIDYQ
RLHNARLDLIAIAKFKIGQYQFNNYQNNKSLKNFQLAKLIFNFPPLHFYIIEQYLSVLKARESW
QFSLGGNYLQTKNVNNTSNERHIENTGFVKNEDMLPQSAHGFAYYFDAQKNWNIAGSHYL
HFANNLYGKNYWNNHNYDEITNRTYLGYQYQNAKYKLVLKPFYERQWLGGHRYNWANGA
RAEYSLNLSKNWQISTALELSQLRYFTQADRNGTIKLASVTFIWQPSDKGYYYLGSDFIRET
TRIKQYSNDMKALRLGWRQNWGYAIASQINGSIALKQYKDFASLGGILPLNKIRRDKIYSLNL
TLWKQDWQYLGFTPKVQFRWKKQESNLPSMFSYSEKYVQMLVEKDF

>HaeP_OmpO1
MKKSTLLLTLASIALPTYAQTSPAPVQERDPLQNEASMPNINQKNLKEDLLRKQRLSEQSAM
AQENQGNSVARFTGKQLAENPALLERLFLEALVNPNKAVLPVYIQIYESVPNADRSLIDWGK
AILEREEDLNKSVASYRKLISDFPENNFIRYQLAETLFYNQEYEAAKTQFERLRASTQNAQD
VAVFDKFIEVIDSKENWNFSFSTTFLNDKNLANSAKQGTKMAVGSGEITYNTPRQTGKGLGL
SIGADKRWSLSNGKYIAFDSSMSSRYYWDNKKYNDLIGHIGLGYGYSDARFNMQVTPYISK
RWYGGGLNGGDSLKQYTDTYGASLSMGYWLSQNFKYSALYNFGYERYRKEIDKRNYNGA

VHSLTNSLMFIPSSTQFWSISLDLTKKYAADRSNAYNRIGTRLTWGQEWPLGITTSTTLGYA
KRHYLEQSFIGMKQRNNEYSASVSLWHKQFHYAGFTPKVTWSYSKTKSNMAIYSYDKNQV
FIEVGKSF

>HaeI_OmpU1
MKNGVKQISFLSLIGLSLIGLSLTNVAWAKVARPKNDTLTNTIQSAELKTSSFSSMPKKEIPN
RHIISLSKSQLAQHPRLVLRGLILALYQNNTQAVQLLLPLYKQFPQQDNFLLTWAKAIEAREQ
GDLTQSIAYYRELFARDVSLLPLRYQLAQALFFNYENESAKIQFEKLRTEVDDEKFLGVIDQY
LLTLNQRNQWIWQVGLNFLNDDNLNNAPKSGTKIGNWTAWEKESGKGVGYSLSVEKKWP
WADHFFSKTMFNGNGKYYWDNKKYNEATLRIGGGLGYQTALVEVSLFPFQEKRWYAGGS
SGTNTMKQYADKLGIRLENVDWLSKTWQISTALEYGKSRYKIRKHLDGNYYFISSTLFYLPK
STQFWFVGMDFHRENTQALDNAYQQKTLRFGWGQDWSHGISSRFTFSYANRVYREKDLI
GIQQKNREYTTTITLWHRNIHFMGLTPKLSWDYQKSTSNHAFYRYDKNRIYLEIGKIF

>HisS_TbpB2
MTSFKLLGFSVLSVALLSACSSGKGGFDLDDVEHTPPSSSGSSRPTYQDVPTGQRQQEIVE
EINSPALGYATEIPRRNISPMPTTGTKESNARVAITAQQVAPLSMPFNSIKEDFIKRLIAENTK
KNARGRDVKYFDDTDDVLFAHDGSNLAHKRDLQYVRVGYVLGTRKIELVFSHDKKTRDQF
PAGWVGYVFYQGTSPAVTLPTQTVTYKGYWDFVSDAFNERTLAEDFTQENSSATSNIPGN
QIGATSMDALVNRKVSGEKINIAHSAEFTADFGSKKLSGELKSNGYVSRIENEQQDVKTRYK
IDADIKGNRFVGSATAQEKSHTIFGKDADKRLEGGFFGPKAEELAGKFLTDDNSLFVVFGAK
RESKGDEKLETRFDAVKISTDSNKLEKETMDTFGNAAYLVLDGRQFPLVPESNAGTTGAGN
TGKNEFISTIDGSHLNKTNHENHKKYKVTVCCSNLVYVKFGSYGEQTTANDASNSTAGAAT
THNSTLTNGHLFLTGERTSLTDMAKQSGAAKYIGTWQANFLSSKGQVGSVDAGDPRNDSG
KSRAEFDVNFGGKTVTGKFFDADGIQPALTMDSTKIEGNGFSGTAKTTGSLQLDKGSTGAG
ITVTFTDAKVDGAFYGPNAEEIGGTITSNGTGDKVGGVFGAKRQELSQQK

>GalA_hTbpB1
MRKSIKFSLLALCISNVIAVNTVFAAEQNQEKSPSPTLSKQQLIAEINQLNKEISAKEKTAVEL
NNQIKDLKSKITKAKQDLQSKIKERDALTNSLQASSKEKAQLEEQIKSLTAQLADLQKKIDPK
VVELVKQKNTAETQLKKVNADVSSKTNQLSQLTDQIKQSTTKVEQLKAQIKPLETKISGLNS
QVKQQQDKLTKIAAERKAVTTKIAEYTKNQEWVKQYELDSDGIRWRDIYFDRNNRSLSKSN
FAKTSNNAMSTHIEVITDELGNPKISDFYTPTLPLSQYKANPSKINEVLISSAEPENKEKAVGK
AFFVNQNYSSYVAWRPVMLKAYNPEIIQALSYGYQGDVDYVARTDYKKGVDIKLQAQQRYE
SVKAKTPQITYRGYMLNVGGKSGDITLTADLDKNVVNGRVTNRIVNPLQDGRDLLLKNGQIS
VDYRGITFRGTPGRAIIPVENNPNNLPFDRANFEGIFAGKNMEEVVGKISYLPNEASSVFGG
TQATK

>GlaP_hTbpB_FR1
MSLKKVTLASLILLGLTACGSSGGGSSDSTVNQNTSQTAEQQAQADKLAAEKAEKERLAKE
AERLAAEQAEKERLAKEEAERLAAEQAEKERLAKEEAERLAAEQAEKERLAKEEAERLAAE
QAEKERLAKEEAERLAAEQAGKERLAKEAERLAAEQAEKERLAKEAERLAAEQAEKERLAK
EAERLAAEQAEKDRLAKEEAERLAAEEKARQEKLASLVAIAKEAGLTDSQAMQVAKNNMTS
SEDQFIQAIETFKIDSAKGFSSNTYRQGFNHLNYSASSSASNRCVNSVCTSWSGTTTTGKTL
YNQKYSVVSGDYSARQGLLNGSWIDESSFSNIQIKGLKTQENAVPNEGSAIYLGKAFKGQT
AGELTYNVNFATRIGSGSISGLPVGQVTLGEANINGTALSGEATAGYGGGYSLEFFGPKAEE
VGGTMYLGADDMYGLSGTRGEIQK

>GalA_Fre
MWRDAQFVGKIIPFVEQNWFGVHRYNHQFGVKLFSQIQLSEHWLWQNRLANMRKHYADT
SLADRYNGNENTVSTTIIWQDNTHWQSWLNLELVSDRLKEKSSSSLKWASGFGIAYYQKQ
WGTQINVRYIDRHFADHHYLMHFKRHDREYQLNGAIWYQKLSWRGFLPKLHIRYQRINSNI
PALYSRKNQELFFTIEKQF

>PasM_OmpW1.1

MKKTVLALGVMAALVAGSAVAHQAGSVIVRGGPILVVPNASTNHDVFKFDVNSNAQLGLTA
TYMATDNLGVELLAATPFSHEITLGNTLVGKTKHLPPSLYAQYYFLDKDAKARPYVGAGVNY
TTFFSEKAVLNGVTDLKLKDSWGPAFNAGVDIQVADNLFLNTAIWYAKIKSKATFKLGDVEH
KVNVKLDPTVFFVGLGYRF

>HaeH_PE1
MKKIILTLSLGLLTACSAQTQKAEQNNVKLTPPTDVRSGYVRLVKNVNYYIDSESIWVDNQE
PQIVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAA
PKKQKKHTLSLTPDTTLYNAAQIICANYGRAQSSDKK

>PasM_PfhR2
MSFKHKTLALFVAHACCTSALAENAATTLEPIVVSDLSHTTLNLDQNKLEKESPKDLKAIFAT
TPNINVIHTGHAQLGDIEIRGMGSSREIFATGANRVTMELDGMDISPSFYFGHSSRHGRQYF
DPSDLKRVEIHKGPNSQGVAGHVRFQTKDPRDYLLPNQRTGAQLRAGYLGDSDAYYVGIT
GATLLDEHSSALVSYTRRWFNEFNNKGGLDVTGSQRTKSNPSSGYSNAVNSKLRYSPNDR
HKFTLNLQHYDLKHTAYLEDSLGTTTTRRGTKTVHHNTNIQKNQRHAIAFSHDMQQTTAFF
DHLHWQIALQQTKSTSRNTGAVTSTSGSPPPSTPKFSQERSLDGFKTKTISLKTEFNKSLG
QHVVHELHYGLKLQYSQMQALRQTQSLNEQGSNTRTSAFFPTQQQWQSKLHLSDRISFGK
SGLSLTPSIHLTQIRIKPKTENVSKKNREQLFTYKDTAIGYGLRVDYALNEANLLSLNYQHAT
RLPGYGENNAQSYGHWPAKPNPHLQPETSDGIELSWRSAGAIGQQTTTLFYNRYNDLIYLD
TTACYADRTGQVPCDLANEKGRSYSYGIEFDGKLNLDTIGFAQGTYLNAGFAYSKGKTANK
QPQGRLDPLTGFVGLGYQQPMDVWGIEGKLKFAAKKKTKDLPANQGFEGLPGYAVVDLTA
YYNVTKQLYLGIGIYNVLDKKYARWAMARGDIKHGNYDKHTEVGRHFGANIRYHF

>AggA_PmuA1.1
MKFKLSLISTALLTAVSMTAYAEEKTENLEQINVEDTGIKQNGYQTTGTSVVSKAEVPVLDTP
NTVNILSTQLLQDRKPDSLIDALYNVSGVSQANTLGGMFDSIQKRGFGYNRDNSIMRNGLQ
AGPAKNFSATTEIVEVLKGPASVLYGIQDPGGVVNVITKKPQQAPKYVVGGTLGNNSLWGT
ELDFTGGLGNGFAYRFIYDKQEKNYWRNFGKIKNTTYAPSLSWENDTTKVLLAYEHKDILEP
FDRGANLLTGTNSLPDIPVSRRLDEPNNETTAKTDNIDFKVEHKLNDHWKLNAGYSYARYK
YFYNQARITAINVNTRTATRAIEKQEGDQRVHSGTVNLVGEFGIGDIANRFVVGVDAMRNIR
DLGPIYNGGYTNSRINIDNPVYTNPVSPVNNSNGNAYQYNYLKTLGVYVQDTAYFTDSFIVT
GGLRYEYFDQFAGRHGLNGTKTGNTDQYDGKLLYQLGAVYKFTPNFAIFANYAESFRPQS
SIAATVDGGLKPEEGKSIEIGTKYENDRFNATISLFNINKRNVGENVTVNGVTQLNIVGKQRS
RGVEFDLNGQLTDNLSVAANYTYTNVKSLENARYPNAIGKQLSGVPKHQASLFLAYNVGEF
DFGNIRFGGGARYLGSWYAYNASYANSYKLPHAVVYDSFIAYDTKISGKKVSFQLNGKNLT
DKTYFPSTSGNATNTLIPVALGYGREFIFNTKIEF

>HaeH_Rup1
MPNNVVFDKQTNSTSLLGFSLSLISLAIFAFHPLTQAESLEQINVVDESPANSTAGSKSYTIKS
MNTATGLRISGKDTPQSVSVVTRKQLDDRAIHTLEDAMKNTTGVNVVHDSGLQTRFLSRGF
YVDQIGEDGITTNVGGRSGYTAKIDVVPSTDLAVYDHIEVVRGATGLTQSNSEPGGTINLIRK
RPTDKFQHLGEITVDHRGSVRTMLDVSGPMNKEKNVRGRLVGVDEKSHSFKQDVKGSKQ
MVYGVFESNIRDKGLFTLGGMYQKINEVPDFAGVMLPCENPKVAAFSSRPACQDPLKFPRS
TYLGESWSRLKGDKYNLFSETKLFLDNDWELSLELSYTQNNSDAKVGQYFIKDEHAAGVSG
ADATSFLTEKGEVIPFEPKDKALEKLASYRKDAKTEFESKKADYLKNGFNQQDFEQYKNKRI
NDRALLRQQCIDEGTLDFFCDGYYPALTDEEYKQDFINKELADQGISDNANERFPNRLYDS
AYNNRDVINRRFNYMPMRYQKKDKQYGIKLNLKGNYELFDRSHDFYLGYAYNNENIESTYL
EIFQKKYRVRPSTGTIHAQFDGLCEVHPLGSNKLPTEKGNREPDWDKYNEYGNVSIYEPDC
KNAIKTKYKPMLDVSGKQIYKIDDFGNKVPALEPVYELDANGNKIQEVDEKGNPIFVGFAGTT
PVYKHVKIPDDHILAQYNYAKYFNKNQTHTITASNRFNATDKLHILTGLSFVRFTTSQQKEMP

VRYGEPITNFQSASNVKADNEHYKASMTGHKFTPYLGITYDLTDNHSIYVSYTKIFKQQDNV
DVTTKTTLPPLVGTNYEVGWKGAFLDNRLNASLALFTLEQKNRTVVDFGFVPDSTGRQGSF
QTIAKPIGRVESKGFEIEAAGNLTENWQLFLGYTYNKSKYKNAQEVNAERLEKNTKATPYNF
SNFTPVQMFRLATSYHIPNTKLTIGGGASAQSPTSSLYGIHQAGYALWDANIQYEFNPHAKL
SLIGTNLTDKIYFENNYNRTRGMNNFYGRPRTVMMKLDWTF

>AviP_SquA1.1
MEYQMTRTQIANIDEKLDIIDHLNQDHSDELLVIANYYSQNSNYTKARIADIFEEGVIIDLFHID
QGQSQLYIPFKIKGSLEEKILYLAYYALTKLGENLKGNKRQFFTVKGKSRPSKNIMRIIINGTM
PFPEYYATYTYGFVLKQINRPKEIHSDKKKKGFFMRGMDHLFIWLLRILNSKNRMKLIKHINR
DIRLYTLRSSTQNDEGQYTQGAIDIYLHGDTPGRRWVEQLKEGDIIFSRSKLPDKHPHLRSG
KNVLIGDETAYPALAGILELWNNPTPPVVLILMNHPEEKAYFKDVPMPEGTELHYLIYAEKKQ
SELVIEELKKLSDLTGTWGGLERREAQKVRHYIRHIFSIDGTKNHIKGYWEIK

>AviP_SquB1
MDDNSVDIVVDGISNGKSGLARPLLDVERIEVAKGPQGTLFGTKAEAGSVMIKTTDPKQEFE
ANIGVNAGNLNLRGINGLLNVPFSEQFSFRLATQLERFDDYIKDADTGKPLNRKTNEAIQAKL
RWNDGEYNDAILSVYHDQRKNFLPIILSEPFSFSTQTNGLPHSAYRKNSGISLKYMHGFDFA
VLESTTAYHYHRANVNRPLRPLDMLGVFYNAANIPQSLHPILDQYYYQNKNNRQNVNERVK
QFSQELKFTSETNSGLIWVAGAYFEKRKRNFTYDAIRSTQQLSPAVLLGNDPFNAIIDRDFD
YETKALFGELTFPLTESLKAVAGARYSHEQLNYRAIYMPNADLAKPSYTEKHKISDNFLSGR
LGVNYALTSEWRLYAIQSLGNKFGGFADYGTNIAYGKDNQPYKSAKVISSEIGSKFLTNNGK
FGFDVSLFNTKVKNDHITITLYPSYLTGTANADTRSRGAELGIYWQLAEQLKLKQEITYLDTK
VTKVPATARNITAKGNRLPQAAKWSGALSIAYESKPFNFGFMGESQLLSDLSIRYVGSRYA
QPDNVQKLGHYMLLNLSIGLQNKHHNILLWSKNLTNRKYHAFGIMPGYAGLPSSGRTFGVN
YSYAF

>HaeI_TbpA1
MTKKPYFRLSIISCLLISCYVKAETQSIKDTKEAISSEVDTQSTEDSELETISVTAEKIRDRKDN
EVTGLGKIIKTSESISREQVLNIRDLTRYDPGISVVEQGRGASSGYSIRGMDRNRVALLVDGL
PQTQSYVVQSPLVARSGYSGTGAINEIEYENVKAVEISKGGSSSEYGNGALAGSVTFQSKS
AADILEGDKSWGIQTKNAYSSKNKGFTHSLAVAGKQGGFEGLAIYTQRNSIETQVHKDALK
GVQSYDRLIATTDKSSGYFVIQGECPNGDDKCAAKPPATLSTQSETVSVSDYTGANRIKPN
PMKYESQSWFLRGGYHFSEQHYIGGIFEFTQQKFDIRDMTFPAYLSPTERRDDSSRSFYPM
QDHGAYQHIEDGRGVKYASGLYFDEHHRKQRVGIEYIYENKNKAGIIDKAVLSANQQNIILDS
YMRHTHCSLYPNPSKNCRPTLDKPYSYYRSDRNVYKEKHNMLQLNLEKKIQQNWLTHQIV
FNLGFDDFTSALQHKDYLTRRVIATADSIPRKPGETGKPRNGLQSQPYLYPKPEPYFAGQD
HCNYQGSSSNYRDCKVRLIKGKNYYFAARNNMALGKYVDLGLGIRYDVSRTKANESTISVG
KFKNFSWNTGIVIKPTEWLDLSYRLSTGFRNPSFSEMYGWRYGGKNDEVYVGKFKPETSR
NQEFGLALKGDFGNIEISHFSNAYRNLIAFAEELSKNGTGKGNYGYHNAQNAKLVGVNITAQ
LDFNGLWKRIPYGWYATFAYNQVKVKDQKINAGLASVSSYLFDAIQPSRYIIGLGYDHPSNT
WGINTMFTQSKAKSQNELLGKRALGNNSRDVKSTRKLTRAWHILDVSGYYMANKNIMLRL
GIYNLFNYRYVTWEAVRQTAQGAVNQHQNVGSYTRYAASGRNYTLTLEMKF

>PasM_TbpC1.1
MLTLVFSVIRRSFMKFKYCVIHQLVLSGSLIFPTLSFADSDSTKTTPNKASYLEEISVVATTVQ
HIFSSTQSTEQLRKNMVSNERDLFRYEVGIGVPESGRSGSNGFAIRGVDKDRVAVIVDGIPQ
AESTISTSARYSTERHNGNINNIEYENVSSLKVQKGAASVMYGSGALGGTVEFTTKDIEDFV
EPGRHLGFLSKTGYTSKNREYRQVIGVGGKGEHFFGFVQLTKRWGHETINNGKGTDILGE
HRGKPNPLNYYTTSWLTKVGYDINNTHRFTLFLEDRREKKLTEEKTLGLSDAVRFANDQTP
YLRYGIEYRYNGLSWLETVKLFLAKQKIEQRSALQEFDINNRNKLDSTMSFVYLQRQNIARG

EFSTSPLYWGPSRHRLSAKFEFRDKFLENMNKHFTFRPWQINTFRQQGRNNYTEVFPVKS
REFSFSLMDDIKIGELLHLGLGGRWDHYNYKPLLNSQHNINRTQRLPYPKTSSKFSYQLSLE
YQLHPSHQIAYRLSTGFRVPRVEDLYFEDRGKSSSQFLPNPDLQPETALNHEISYRFQNQY
AHFSVGLFRTRYHNFIQEREMTCDKIPYEYNRTYGYCTHNTYVMFVNEPEAVIKGVEVSGA
LNGSAFGLSDGLTFRLKGSYSKGQNHDGDPLKSIQPWTVVTGIDYETEGWSVSLSGRYSA
AKKAKDAIETEYTHDKKVVKQWPHLSPSYFVVDFTGQVNLSKNVILNMGVFNLFNRDYMTW
DSAYNLFTRGYTSRSVRANSPGINRFTAPKRNFAASVEIRF

>HisS_TbpD1
MFFDQNRQQTLLRFSAVTLPLFSFNAIAQFNDSATLDQIIVETEQVSAKTSSTNSQQISRGLV
WNEQDLVRWQTGVTVTEGGRGGANGYTMRGVESDRVAMSVDGISAVESYMPRFYYIKGF
YNGNRNSTEIENLSSVEFNKGANSLTQGSGALGGSVSMRTKNVHDFVDSNKTVGFYAKTA
YSSRNTEYRQVIGGFKQGNAEGLLQYTYRHAHENKNYYSGKLKDVAYCGILPNGDQRDIFP
QLCGRSRVLPDNVDYHSRSWLAKLGYHFTPQHFVQGFFENLEQSRKILEKSFYASNRQEA
GDITPYQRFGLSYNYTPTESWLNKFSVQLAQQKVSQQANSNQYGVSPQDYNIITDKRLYKT
EQKHYQLDVSFIANEIAGKFAHHQLSGGLGYHWGKLENRNQEYGAYSRIKQREFTIQQPVA
RQAWYAYVQDLISINQKWGANAGIRFDHYRYKPTTSALKYDLASQKVEVLPKKQFSAVNFS
LGLNYSLTDSTLLSYEFSTGFKAPKIEEMYFDLKGASDVQYLPNINLKPEKAQNHELSLTKEG
DNYQLSLAVFYTKYRDFIDVETKPKIGSRQRTNWATKPITKYTEYYLDGVNYQQINIDKADVT
GLDFHARANGKIVGLSKSWYAILKASYAKGRKNNGTAMLPIQPFSTTLGLGYQADNDKWNV
LFSGRYVSAKKAKDAIDIPNVNSLQLDVDKKTGVVSDFRQPKPYRFLSSSYVVFDLTAQYRL
NSHFTLNAGIFNLFNRKYTTWDTLRQIKENGALGDVYNSGKGLERFTAAGRNFALSLEMKF

>HaeD_TdX1
MKLNKISLFVSLSLFSTQMYANTELGTIYVKGNVNSTSSGKYSFTEKLFNTKLRANSHISDAL
ASLPAVQVSNTHNGNTLGEITPQNISFHGERYYNNNFMLNGMGINDNINPIGLGANGVINVN
AVKTINPEHIPIGHPQAFWLSPDLIGKVDVLESNVSSRYGSFTGGAVDAQLKEPNAERASGA
ISYRTTHSAWTRFHLDGAYKDEFQRATSPLLQPKFDKHETTIQLNQPLSAHSALLFAYSRQQ
SAVPQHQQFLHKWVGRHRRNESLLLSYKNEVNDNNRLLANAIYAPHYADEYLDNALDGRF
RISGGGLLTNIKWENYNRLGLLTSELSYRHNRTQTKYDADTLYRYTKTDSINWISDPNTEEA
TKGGIGKRYTEQKSVQLKQHLQFKKLELAHTTHEASVGWEWQYNQSALKQPKLARQYTGA
RKYYAEGKPFDIENCTDCIPKEQYFTLKVDYLPLNGKVNHQRLAFYLEDKIKWRDVTFVPGM
RFDYAQFTHKWNIAPRTYLDYDLFGQDKTHFSAGFNRYYAGDFVEYKLRSAFNASNEYDR
DTYDSEWKLTKSQLNIAYKGNKIKTPYSDEWNVGVNQKIANSVWQLKWVKRQSKDQFITQ
VDFDAKPQQRWLSNLGKGTHHTISLEVENIAPIQTQYAELGWQAGLAWTKARTNQTRDYT
QMDWNNFGVKKMLHNGKLDTVDHLPAQDFNLPWKGYAQLDSYFPSLNLRWAQRINYESA
SNQYVLNGFRCKKEELACQGYEGSVVKVNKRRNPHSITLDWFFDWKKPLTNNRILSVNMSI
LNVLNRVQKANQVTEDHENYYQTYRPGRQFWLGIKYEW

>HaeP_ViuA2
MNKLSPRLSLISLCLMSAYSFSTHAEESQQTTVLDEITVTGEKFERSQSSTGSSTTVVTGDQ
FKREANLLSATQLLKRDVNILDTGLGNDLPTVRGVDGSGPAVGAVAFFAGSRPRLNMQIDG
RTSSYNELAFGTKSLWDMKQVEIYRGAQSYAQGRNAIAGAVVMTSNDPTQEWEGAAKLN
MGNHRLAQTAALISGPVVKDELAFRLSVDHQQRETAVDLPHYDPVGNPRWFKATNTRAKL
LWTPSALPDLYSRLTFNHLNARAPQSETELQSHSPRYTPERPVFQTRSASTIWDIGYQLSE
RWKWENKLVYTHFIHDRKTTSPFNTALPPNRRGVPARVDGNEFQIEPIIKYESEKYRGLFGL
FYFNAKQDESVTMLNGRIARTPITTNFNDKTKTKAAFGEITFTPDIPFELTISARYEQEHHQR
KGKSAMFSINRDKKYNVFLPKADIAWKINDDQRLGFKVGKGYNPGGAGVTFGVPYTSYEYD
AEYVWNYELYHRWTSADKRLRINSNLFYNDYKDMQLPFTLGPNSIVIRNADKVVTYGAEINT
EWQATEKLALNAGIGMLKTDIKRYPNSSIEGNKLARAPSFSGKVGANYRLLDHLEIGTNYSY
NSSYYSTADNLANGKVGHYDQLDVYLAYDFKHARITLYADNVLNSRKDILLVPRSGDITRQP
ERQIGLSTELRF

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupFile7H.haemolyticus.xlsx
- SupFile8H.influenzae.xlsx
- SupFile9Haemophilusparainfluenzae.xlsx
- SupFile10H.somni.xlsx
- SupFile12P.multocida.xlsx
- SupFile14SIFTER.xlsx
- SupFile15SimpleTally.xlsx
- SupFile16RepProteinSequences.docx
- SupFile1A.pleuropneumoniae.xlsx
- SupFile2A.actinomycetemcomitans.xlsx
- SupFile4G.anatis.xlsx
- SupFile5G.parasuis.xlsx
- SupFile6Haemophilusducreyi.xlsx
- SupFile3A.paragallinarum.xlsx
- SupFile11M.haemolytica.xlsx
- SupFile13Stats.xlsx