1

2 **Genomic differentiation of *Mannheimia haemolytica* genotypes**

3

4 Darien Deschner, Janet E. Hill*

5 Department of Veterinary Microbiology, University of Saskatchewan, Saskatoon, Canada

6

7 *Corresponding author

8 E-mail: Janet.Hill@usask.ca

9

Abstract

Mannheimia haemolytica is an opportunistic bacterial pathogen associated with the economically costly bovine respiratory disease. Two genotypes have been described, of which genotype 2 is more strongly associated with disease. Several previous studies have investigated the genomic differences between the genotypes and/or the major serotypes (1, 2 and 6) of *M. haemolytica*, however we still lack a clear basis for the greater disease association of genotype 2 (serotypes 1 and 6) and demonstrations of phenotypic differences are scarce. This work builds upon previous investigations to identify genes that differentiate the two genotypes with a particular focus on genes that may play a role in virulence and fitness in the respiratory tract microbiome. We identified 422 genotype differentiating genes in a collection of 208 *M. haemolytica* genomes (62 genotype 1, 146 genotype 2). Genotype differentiating genes including genotype-associated variants of a TonB-dependent siderophore receptor homolog, transferrin binding protein B, leukotoxin A, and IgA1 proteases. We also identified a genotype 1 associated lytic transglycosylase, and a genotype 2 specific highly immunogenic outer membrane lipoprotein. These results expand our knowledge of what differentiates the genotypes of *M. haemolytica* and provides information that can be utilized to inform diagnostic strategies. The results also highlight the relative lack of information regarding *M. haemolytica* from sheep and goats and geographic regions other than Canada and the United States.

Introduction

Mannheimia haemolytica is a gram-negative opportunist bacterial pathogen associated with bovine respiratory disease (BRD) [1,2]. The beef industry is estimated to lose more than $3

31 billion per year globally due to BRD [3,4]. Two genotypes of *M. haemolytica* have been described
32 based on whole genome sequencing and single-nucleotide polymorphism (SNP) analysis and
33 multiple serotypes have also been defined [5,6]. The three most common serotypes (1, 2 &, 6)
34 have been found to correspond to genotypes with serotypes 1 & 6 corresponding to genotype 2
35 (G2) and serotype 2 corresponding to genotype 1 (G1) [1,5,7,8]. Genotype 2 has been more
36 frequently associated with clinical cases of BRD, however the genetic and phenotypic basis for
37 this apparent association is unclear [1,9–14].

38 Differences in disease association between strains of the same species of bacteria is
39 relatively common. For example, commensal and pathogenic strains of *E. coli* differ in their
40 possession of virulence factors such as adhesins, toxins such as the Shiga toxin, and other effectors
41 such as the cytotoxic necrotizing factors [15]. Other species known to exhibit differences in disease
42 outcome based on strain diversity include *Staphylococcus aureus*, *Mycobacterium tuberculosis*,
43 *Campylobacter jejuni*, *Listeria monocytogenes*, and *Salmonella enterica* [16]. In addition to
44 virulence factors, adaptations that confer increased fitness, such as greater ability to acquire iron
45 and other nutrients, can contribute to disease association by allowing a pathogenic strain to
46 outcompete non-pathogenic strain(s) present in the same environment [17].

47 Previous research into the virulence factors of *M. haemolytica* has focused on leukotoxin,
48 lipopolysaccharide, adhesins and various proteases [11,18–20]. The response of *M. haemolytica*
49 to iron limitation has also been studied, however most previous studies have only examined the
50 virulence factors and metabolism of serotype 1 isolates [21]. Serotype associated differences in
51 leukotoxin have been demonstrated, including differing degrees of cytoxic activity [1,22,23].
52 Serotype 1 has also been demonstrated to invade differentiated bovine bronchial epithelial cells *in*
53 *vitro* while serotype 2 was unable to, suggesting that serotype 1 (included in G2) likely possesses

54 virulence factors facilitating this invasion that serotype 2 (included in G1) lacks [24]. Expanding
55 on these previous observations and examining them through a genotype rather than serotype lens
56 will help elucidate the clinical relevance of genotype in *M. haemolytica*. Also, as diagnostic
57 laboratories continue to move toward molecular methods and away from techniques like serology
58 for many pathogens, understanding the association of genotype with pathogenicity is increasingly
59 important.

60 Several previous studies have investigated the phenotypic and genetic differences between
61 the two genotypes of *M. haemolytica.* Phenotypic differences described to date are distinct matrix-
62 assisted laser desorption/ionization - time of flight mass spectrometry (MALDI-TOF) profiles and
63 colony morphology [25,26]. A pangenome analysis of 69 *M. haemolytica* genomes identified 112
64 G1 specific, and 179 G2 specific genes including seven G2 specific predicted outer membrane
65 protein encoding genes that were hypothesized to play a role in the greater disease association of
66 genotype 2 [27]. Another study examined a collection of 11 *M. haemolytica* genomes for potential
67 differences in virulence factors between the serotypes 1, 2 and 6 [28]. The authors reported more
68 prophages in serotypes 1 and 6 than in serotype 2, which could influence virulence due to the
69 presence of virulence factors in these prophages. Self-targeting CRISPR spacer sequences were
70 detected in serotypes 1 and 6, which may enhance immune evasion or adhesion during infection
71 by modulating sialic acid expression on the cells surface. ARG-carrying ICEs were more prevalent
72 in serotypes 1 and 6 than serotype 2, which could enhance survival of antibiotic exposure [28]. In
73 our own previous investigation of macrolide resistance in *M. haemolytica,* we found that genotype
74 was not associated with differences in phenotypic resistance after accounting for the prevalence of
75 ICEs, however we did identify a resistance-associated single nucleotide polymorphism in the L22
76 ribosomal protein that was exclusive to G2 [29].

77 In addition to the demonstration of genotype differentiation by MALDI-TOF [25], a loop-
78 mediated amplification (LAMP) assay has also been developed to differentiate the *M. haemolytica*
79 genotypes [30]. The rationale for application of such genotype specific diagnostics beyond
80 research studies and the potential value of their adoption in routine diagnostic investigations
81 depends on a better understanding of the roles of *M. haemolytica* genotypes in bovine respiratory
82 disease. In the current study, we take advantage of the greater number of *M. haemolytica* whole
83 genome sequences now available to identify genetic differences between the two genotypes. In our
84 examination of >200 *M. haemolytica* genomes, particular attention was given to genotype
85 differentiating genes that are predicted to encode virulence factors and those implicated in
86 processes that affect fitness in the respiratory tract such as iron acquisition and carbohydrate
87 utilization.

88 Materials & Methods

89 Genome sequences

90 *M. haemolytica* whole genome sequences defined as complete by NCBI were downloaded
91 from the NCBI Genome database on March 25, 2024, using the NCBI Datasets command line
92 tools [31]. To avoid duplication all .gca files that had a matching .gcf file were removed. The
93 remaining genome sequences were genotyped using an *in silico* PCR assay in Unipro UGENe
94 v45.1 and primers from a genotype specific LAMP assay [30,32]. Primers from the LAMP assay
95 included a pair targeting G1 specific adhesin pseudogene B1, G2 adhesin G (CP017538.1 locus
96 tag: BG586_06285; Forward: 5'-GCA CAT TAA AAT ATA GCA GCT TTG-3'; Reverse: 5'-
97 CTA AGC CAG AGT GAT CCG-3'), and leukotoxin D (*lktD*) (CP005972.1 locus tag:

98 F382_07400; Forward: 5'-ACT GAA AA TTC ACA CTA TAG GTG-3'; Reverse: 5'-GCT AAT
99 ATG TTT AAT TCG ACC AGT T-3'). The *lktD* primers were used to confirm identity of genomes
100 as *M. haemolytica*. In initial testing G1 specific primers failed to produce any predicted products
101 in our G1 reference strain and so were discarded. New G1 specific primers targeting the G1
102 specific adhesin pseudogene B1 were developed using Primer 3 v 4.1.0 (CP017495.1 locus tag
103 BG548_01640; Forward: 5'-ACT AAT CTG AAG AGC GGC GT-3'; Reverse: 5'-TTG GTG
104 GTT GCT GCT TGT AC-3') [33–35]. Successfully genotyped NCBI genomes were combined
105 with a set of 99 genome sequences previously generated by our lab (BioProject:
106 PRJNA1088094)(30) and reference genomes (G1 = CP017518; G2 = CP017519) for both
107 genotypes to generate the final collection of 209 genomes (62 G1, 147 G2) (S1 Table).

108 Pangenome construction, annotation, and classification

109 The pangenome of 209 genomes was constructed using PanTools v 4.3.2 with FASTA
110 files, gff3 annotation files, and genotype designation as inputs [36–38]. Genomes were assigned a
111 new unique identifier based on their input order during pangenome construction (S1 Table).

112 Pantools groups protein sequences from the input genomes into homology groups based on
113 their similarity. Homology groups also contain the nucleotide sequences encoding the constituent
114 protein sequence(s). Homology groups were generated using the group command with the
115 relaxation parameter set to 2. The relaxation parameter is a combination of four sub-parameters
116 (intersection rate, similarity threshold, mcl inflation, and contrast) where a lower number
117 represents a stricter set of thresholds. Given the large number of genomes and anticipated sequence
118 variation, we set the relaxation parameter to 2 to reduce the chances of minor differences
119 preventing similar proteins from being grouped together.

120 Genome metrics such as genome size, GC content, number of genes, etc. were determined
121 using the metrics command in PanTools [36]. One genome (Genome 130/38241 S-19) was
122 approximately twice as large as all others and appeared to represent a chimeric assembly. This
123 genome was excluded from all subsequent analyses reducing the genome collection to a total of
124 208 (62 G1, 146 G2).

125 Homology groups were classified as core (present in ≥ 90%, 187/208 genomes), unique
126 (present in only one single genome), or accessory (present in 2-186 genomes). Homology groups
127 were additionally classified as either shared (present in ≥ 90% genomes of one genotype),
128 exclusive (solely present in genomes of one genotype), or specific (present in ≥ 90% genomes of
129 one genotype, and no genomes of the other genotype) [36]. To allow for instances in which a
130 homology group was strongly associated with one genotype yet still present in small numbers of
131 the other genotype, all homology groups not initially classified as core or unique were further
132 manually examined. Homology groups that were present in ≥ 95% (60/62 G1, 138/146 G2) of one
133 genotype while also being present in ≤ 5% (3/62 G1, 7/146 G2) of the other genotype were
134 classified as being genotype associated.

135 Genomes were functionally annotated using InterProScan v 5.64-96.0 with SignalP v 4.1
136 database integrated [39–41]. Biosynthetic gene cluster annotations were generated using
137 antiSMASH 7.1.0 with default parameters [42]. InterPro and antiSMASH annotations were added
138 to the pangenome, with InterProScan annotations including InterPro, Pfam, GO, and SignalP
139 annotations [36].

140 Multiple sequence alignments of constituent nucleotide sequences were generated for all
141 homology groups in the pangenome with MAFFT and consensus sequences from the alignments
142 were generated to represent the genotype specific and associated homology groups [36,43].

143 Consensus sequences were used as queries for BLASTx searches using the bacterial/archaeal
144 translation code against the NCBI refseq_select database [44–48]. All other parameters were kept
145 as default. Searches were performed March 20-25, 2025. The top hit for each search was manually
146 recorded to provide further evidence of the predicted function, facilitate human interpretation, and
147 identify functions that may not have been annotated by any of the other approaches. The entire
148 pangenome database is available in S1 Dataset [43].

149 The nucleotide sequences of several previously characterized genes upregulated in
150 response to iron limitation, and previously identified proteases were used as BLAST queries
151 against the pangenome using the blast command in PanTools to determine if these previously
152 characterized genes were present in our dataset [18,21,36]. Both alignment threshold and minimum
153 identity were set to 90. Query nucleotide sequences are available S2 Table.

154 Tet(H) gene abundance

155 Tet(H) genes were used as a proxy for the presence of integrative and conjugative elements
156 (ICE) due to the strong correlation between presence of tet(H) and presence of ICE [12,49]. The
157 reference tet(H) nucleotide sequence (Accession Number: Y15510.1 (nucleotides 2099-3301))
158 was downloaded from the comprehensive antibiotic resistance database (CARD) on July 26, 2024,
159 and used as nucleotide BLAST query against the pangenome database using the inbuilt PanTools
160 blast command. Minimum identity and alignment threshold were both set to 90%.

161 Leukotoxin operons

162 Reference nucleotide sequences for each of the component genes of the *M. haemolytica*
163 leukotoxin operon (*lktCABD*) were downloaded from GenBank (AF314508.2) and used as queries
164 against the pangenome BLAST database with percent identity and alignment threshold both set to

165 90%. Consensus nucleotide sequences of G1 and G2 versions of all leukotoxin operon genes were
166 extracted and translated in Geneious 2024.0.5. Predicted protein sequences for the G1 and G2
167 versions of all leukotoxin operon genes were aligned against one another in Geneious 2024.0.5.

168 Annotation of carbohydrate active enzymes

169 Carbohydrate active enzymes (CAZymes) were annotated using a local installation of
170 dbCAN v 4.1.1 software with integrated SignalP v 4.1 database and DIAMOND, HMMER, and
171 HMMER dbcan_sub prediction algorithms [50]. Annotation was performed on each genome
172 sequence individually using a custom BASH script [43]. dbCAN annotations that had supporting
173 annotations from at least three tools, and contained a predicted signal peptide were selected for
174 further analysis.

175 Protein structure prediction and visualization

176 Protein sequences were submitted to the AlphaFold Server for structural prediction [51].
177 Models with the highest-ranking score were downloaded and visualized in PyMol version 2.5.2
178 (Schrödinger, LLC).

179 Statistics

180 A Fisher's exact test and odds ratio comparison was performed to determine if the
181 prevalence of tet(H) was different between genotypes. Both tests were performed in Microsoft
182 Excel for Microsoft 365 using Real Statistics Resource Pack. A p-value ≤ 0.05 was considered
183 significant.

184 A one-way ANOVA (Kruskal-Wallis), and post-hoc pairwise Mann-Whitney t-tests were
185 performed in GraphPad PRISM v 10.4.1 to determine if there was a difference in genome size
186 between G1 and G2 with or without ICE. A p-value ≤ 0.05 was considered significant.

187 Results & Discussion

188 Description of genome collection

189 *M. haemolytica* genomes designated as complete were retrieved from NCBI and after
190 removal of duplicates, 121 genomes were available for further analysis. *In silico* PCR with
191 modified primers from a previously published genotyping assay resulted in the expected amplicons
192 for 108/121 genomes. No *in silico* genotyping PCR products were obtained for 13 genomes and
193 so they were removed from further analyses leaving 108 genomes.

194 Three of the 13 isolates that could not be genotyped also failed to produce any predicted
195 product for the *lktD in silico* PCR. Two of these genomes have since been removed by the RefSeq
196 curation team for either containing too many frameshift mutations (GCF_900474405) or failing to
197 pass a completeness check (GCF_023516295). All three genome sequences that failed to produce
198 an *lktD* product were from sheep isolates (NCTC 10643, NCTC 10609) or from culture collections
199 from the 1960s (ATCC 29701). Of the other ten genomes that were not genotyped by *in silico*
200 PCR, four were from a German study examining macrolide resistance in *M. haemolytica* [49], and
201 three were from sheep isolates collected in 1951 or 1968 and sequenced as part of the United States
202 Department of Agriculture (USDA) *M. haemolytica* sequencing project (BioProject:
203 PRJNA172854). The last three isolates were all generated from non-nasopharyngeal cattle samples

204 as part of either the USDA sequencing project (GCF_000819525, GCF_023516255) or a German
205 study (GCF_020971705) and had a CheckM completeness <92.5%.

206 Given the variety of the sources of the isolates we were unable to genotype it is difficult to
207 determine a general cause, however the assay developed may struggle with genomes derived from
208 ovine isolates, non-North American isolates, and non-nasopharyngeal isolates. Whether the failure
209 of the *in silico* assay to genotype isolates from these origins represents true genotypic differences
210 is unclear and requires further investigation. Difficulties in genotyping these isolates likely reflects
211 the lack of diversity in available *M. haemolytica* genomes. For example, of the 158 annotated
212 complete *M. haemolytica* genome sequences available in the NCBI genome database as of April
213 4, 2025, only 22 were from sources other than collections from the USDA or the University of
214 Saskatchewan.

215 The successfully genotyped genomes were combined with 99 genomes from our previous
216 work (BioProject Accession PRJNA1088094) and reference genomes for G1 and G2 as described
217 in the Methods. One additional genome was removed following discovery that it was likely a
218 chimera, leaving a final dataset 208 *M. haemolytica* whole genomes with 62 G1 and 146 G2
219 representatives (S1 Table). Most genomes were from isolates originating from feedlot cattle in
220 either Canada or the United States between 2003 and 2021. The lack of geographic diversity and
221 absence of genomes from ovine isolates means that the results presented in this work may not
222 represent the characteristics of *M. haemolytica* genomes outside of the Canadian and American
223 feedlot industries.

Genome properties & Tet(H) abundance

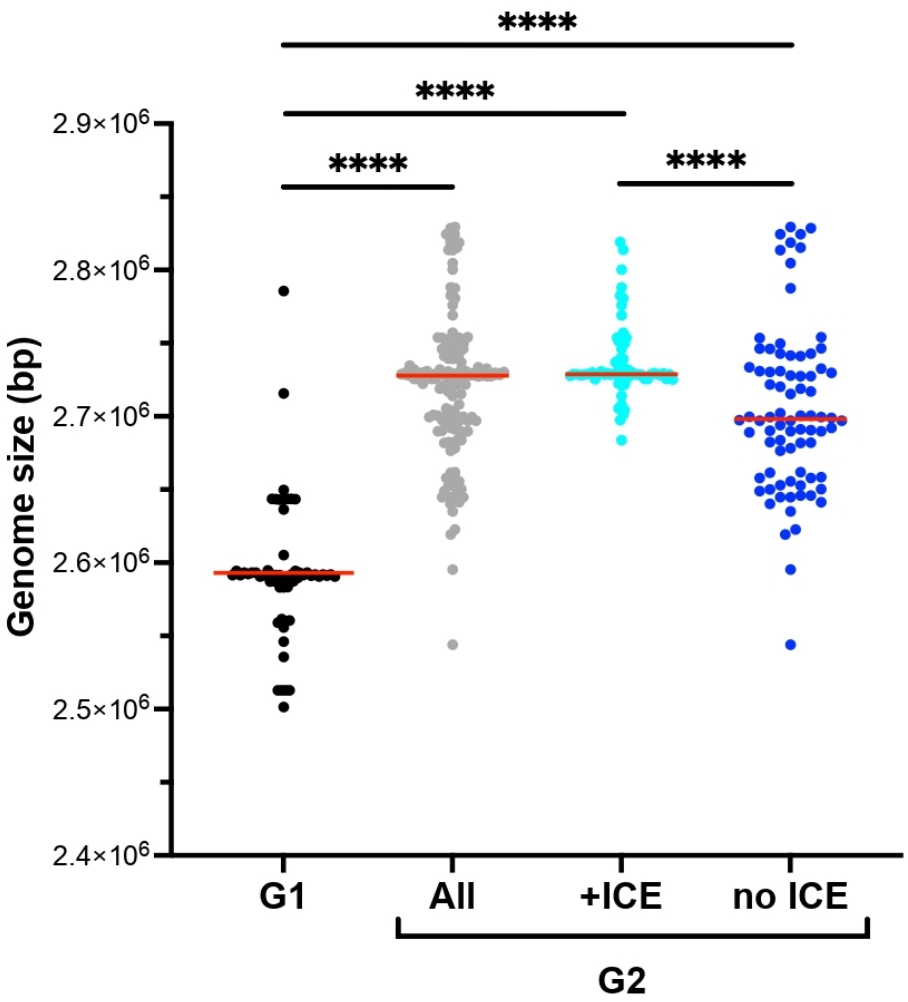
The two genotypes were compared to one another on a broad set of genome metrics, such as genome size, GC content, etc., to determine if there were any high-level differences in the genomic structure and content of the two genotypes. G2 genomes were larger, had more predicted protein-coding sequences, and had slightly lower GC content than G1 genomes. No differences were identified in the number of tRNA or rRNA genes or the proportion of the genome covered by genes (Table 1).

Table 1. Comparison of genome features of G1 and G2. Values are the average ± standard deviation. P-value ≤ 0.05 was considered significant.

Metric	**Genotype 1 (n = 62)**	**Genotype 2 (n = 146)**	**t-test *P*-value**
Genome Size (Mb)	2.592 ± 0.045	2.719 ± 0.048	<2e-16
GC Content (%)	41.14 ± 0.026	41.04 ± 0.031	<2e-16
Gene Count	2,561.7 ± 100.9	2,798.9 ± 116.4	<2e-16
Genome Covered by Genes (%)	88.17 ± 1.5	88.73 ± 1.2	0.0097
tRNA Count	64.31 ± 1.70	64.17 ± 2.04	0.61
rRNA Count	18.98 ± 4.3	19.68 ± 2.4	0.24

The larger size of G2 genomes may be attributable to a greater prevalence of integrative and conjugative elements (ICEs), large (~20 kb to >500 kb) mobile genetic elements that are known to be associated with antimicrobial resistance and are more frequently identified in G2 genomes [5,29]. To investigate this possibility, we used the proxy measure of the presence of tetracycline resistance gene tet(H) due to its strong association with ICEs [12,49]. Tet(H) was identified significantly more frequently in G2 (67/146, 45.9%) than G1 (1/62, 1.6%) (Fisher's *p*-value = 4.39×10^{-12} Odds Ratio: 0.02 ± 1.02), however, G2 genomes were larger than G1 regardless

241 of the presence of ICE (Figure 1). The larger size of G2 genomes regardless of presence of ICE
242 suggests that G2 genomes carry a wider repertoire of genes that may allow them to be more
243 adaptable to a wider variety of environments than the smaller G1 genomes. Considering average
244 bacterial gene sizes, the difference in the median genome size of G2 without ICE and G1 (~100
245 kb) could represent dozens of G2-specific genes. Additionally, the greater prevalence of ICEs in
246 G2 suggests that this more disease associated genotype is also more likely to possess antimicrobial
247 resistance genes.



248

249 **Figure 1.** Genome sizes of Genotype 1 (G1), Genotype 2 (G2), G2 possessing integrative and
250 conjugative elements (ICEs), and G2 without ICEs (no ICE). Red bars indicate mean. Significant

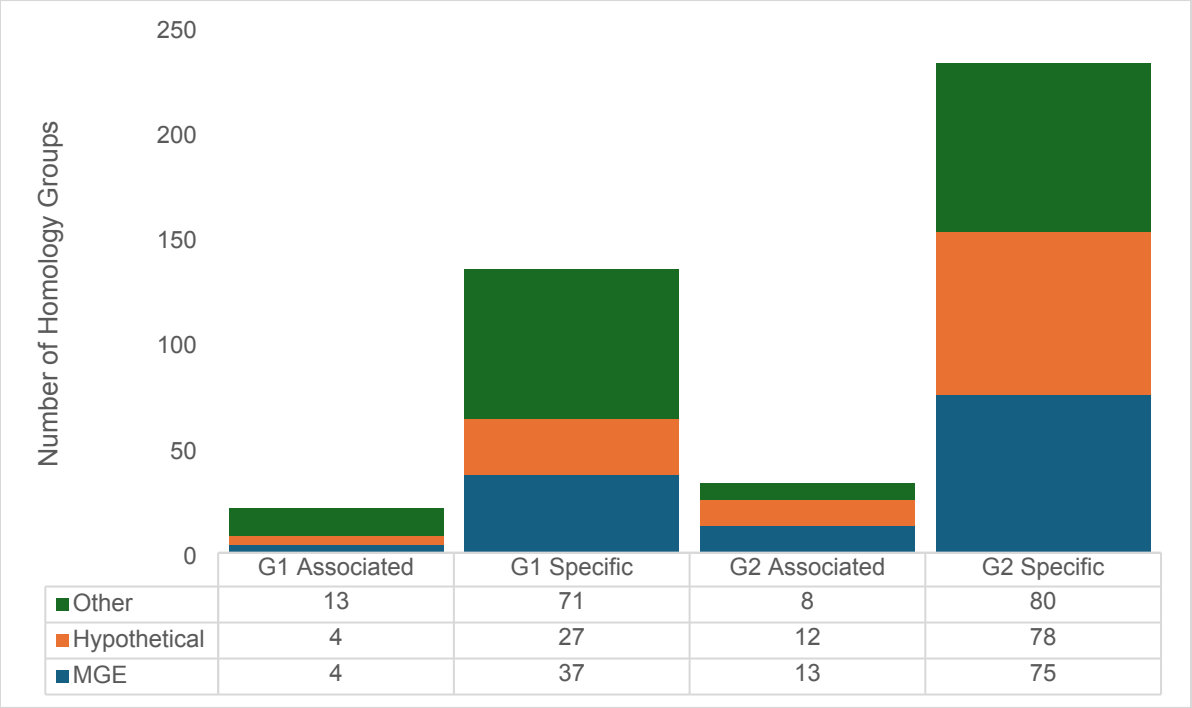
251 differences (P < 0.0001) are indicated (one way ANOVA (Kruskal-Wallis), post-hoc pairwise
252 Mann-Whitney t-tests).

253 Pangenome analysis

254 The pangenome calculation resulted in the identification of "homology groups": clusters
255 of sequences from input genomes that exceed a given similarity threshold. We then examined the
256 distribution of homology groups across all *M. haemolytica* genomes to identify those that are core,
257 unique, accessory, shared, genotype specific or genotype associated as defined in the methods. A
258 total of 4,038 homology groups were identified of which 2,007 (49.7%) were core, 1,473 (36.48%)
259 were accessory, and 558 (13.82%) were unique. Subsequent analyses focused on accessory
260 homology groups as these groups were most likely to differ in prevalence between the two
261 genotypes. We identified 280 G1 exclusive, 2,169 G1 shared, and 135 G1 specific homology
262 groups and 790 G2 exclusive, 2,342 G2 shared, and 233 G2 specific homology groups. To focus
263 our attention on the homology groups most likely to define and differentiate genotypes, the shared
264 or exclusive categories were further examined to identify those homology groups that were present
265 in ≥ 95% of one genotype and ≤ 5% of the other genotype, resulting in 21 G1 associated, and 33
266 G2 associated homology groups (S3-S4 Tables). Identification of genotype specific and associated
267 genes within this relatively large genome collection further reinforces previous findings that the
268 two genotypes carry distinct sets of genes [27]. A substantial proportion of the proteins encoded
269 by the specific and associated homology groups were mobile genetic element (MGE) associated
270 (transposases, toxin-antitoxin systems, prophage associated proteins, etc.) or hypothetical or
271 unknown function proteins (Figure 2).

272 The larger number of G2 specific or associated homology groups compared to G1 specific
273 or associated groups is consistent with the larger size of G2 genomes (Figure 1) and the larger

274 number of specific and associated homology groups encoding MGE-associated proteins in G2
275 suggests that these elements are more common in G2 than G1, consistent with previous
276 observations [9,12,27,28,49]. The number of hypothetical/unknown proteins is also unsurprising
277 as ~22 % of all domain families in Pfam v32.0 are defined as "domains of unknown function" and
278 a recent examination of >30,000 bacterial genomes and metagenomes showed that 40-60% of
279 genes are of unknown function [52,53]. Homology groups predicted to encode
280 unknown/hypothetical proteins, and MGE-associated proteins were excluded from further analysis
281 in this study but certainly represent an opportunity for future investigation.



282

283 **Figure 2.** Genotype specific/associated homology groups. Mobile Genetic Element (MGE)
284 associated proteins included transposases, plasmid toxin-antitoxin systems, prophage associated
285 proteins, etc. Hypothetical proteins included those annotated as hypothetical proteins as well as
286 proteins annotated with a domain of unknown function (DUF) which did not appear to be MGE-
287 associated and did not have any putative functions. Proteins that did not fall into either of the two
288 preceding categories were placed into the Other category. G1 = Genotype 1 G2 = Genotype 2.

289

290 Based on inspection of the complete inventory of pangenome homology groups (S3-4
291 Tables), we identified a subset of specific/associated homology groups that may be associated with
292 reported differences between disease association of G1 and G2 (Table 2).

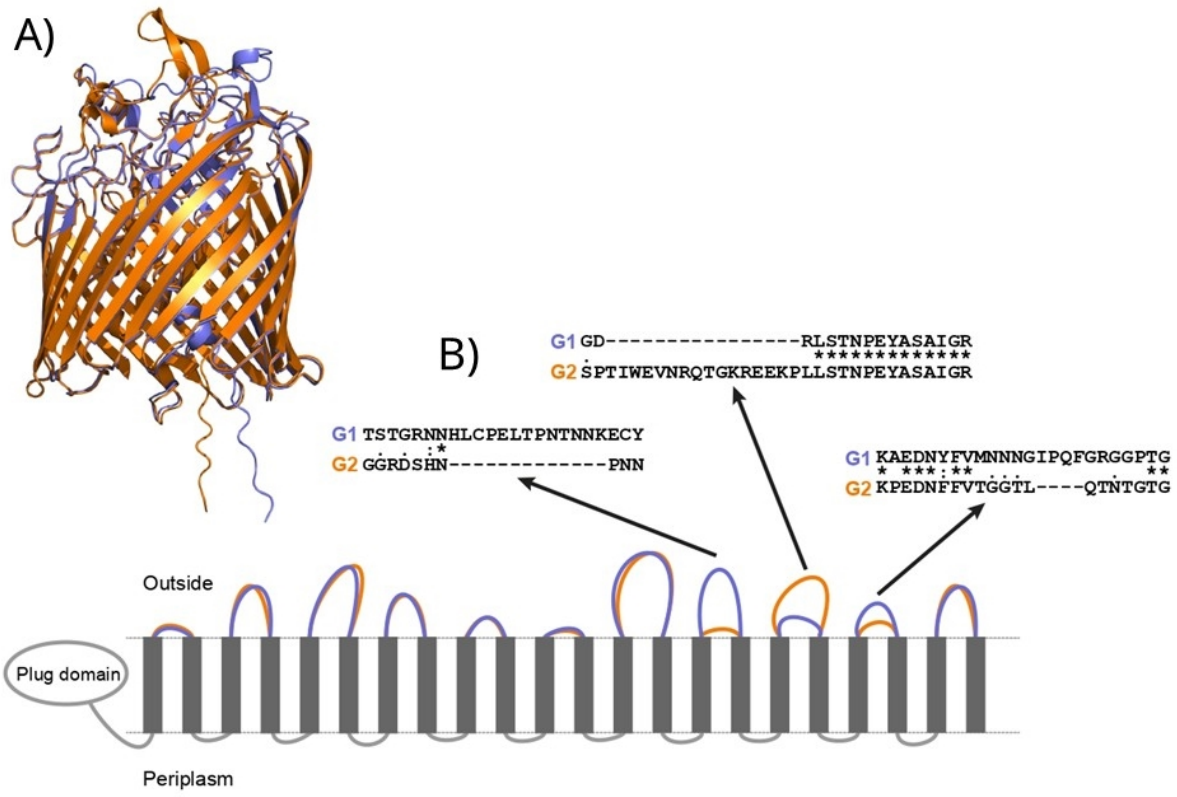
293 **Table 2.** Genotype association and predicted functions of selected homology groups. Annotation
294 notes summarize the functional annotations from InterProScan for each homology group.
295 BLASTX search used consensus nucleotide sequence of the homology group as query against
296 the refseq_select database. Reported top BLASTx hits were from *M. haemolytica* unless
297 otherwise stated.

Homolog group	Genotype association	InterProScan Annotation	BLASTx hit (top)	Predicted Function
2042480	G2 specific	TonB-dependent receptor.	TonB-dependent siderophore receptor	Iron Acquisition
2042481	G1 associated	TonB dependent receptor protein.	TonB-dependent siderophore receptor	Iron Acqusition
2041941	G2 specific	TonB Dependent Receptor	TonB-dependent receptor plug-domain containing protein	Iron Acquisition
2944545	G2 specific	Transferrin Binding Protein (TpbB)	Transferrin-binding protein-like solute binding protein	Iron Acquisition
2040773	G1 specific	Transferrin Binding Protein	Transferrin-binding protein-like solute binding protein. *Mannheimia pernigra*	Iron acquisition
2042189	G2 specific	Peptidase S6/Autotransporter/IgA1 protease	S6 Family Peptidase	Host immune evasion
2040817	G1 associated	Peptidase/Pectate Lyase/IgA1 Protease	S6 family peptidase	Host immune evasion
2044609	G2 specific	N/A	Outer membrane lipoprotein PlpE.	Highly immunogenic outer membrane protein
2042729	G1 associated	MltB-like Transglycosylase	Lytic murein transglycosylase. *Mannheimia glucosida*	Cell wall remodeling/recycling

298

299 Homology groups 2042480 (G2 specific) and 2042481 (G1 associated) were both
300 annotated as encoding a TonB-dependent siderophore receptor. The consensus amino acid
301 sequences of these homology groups were 84.4% identical, and both contained consensus TonB-
302 dependent transporter (TBDT) motifs. In addition, the predicted 3D structures demonstrated the
303 canonical 22 β-strand barrel structure with an N-terminal plug domain characteristic of TBDTs
304 (Figure 3A). The amino acid sequence differences between the G1 and G2 sequences were
305 concentrated in the extracellular loops responsible for substrate binding, suggesting that these two
306 proteins may differ in their ability to bind their substrate [54,55] or they may bind different

307 substrates (Figure 3B). TonB dependent receptors (TBDTs) are transporters often involved in iron
308 acquisition, specifically in the uptake of small, secreted iron scavenging molecules known as
309 siderophores [54,55]. Homology groups 2042480 and 2042481 also produced BLAST hits to a
310 previously characterized siderophore (ferric enterobactin) receptor (*frpB*) homolog that was
311 upregulated in *M. haemolytica* in response to iron limitation [21]. Identification of a putative
312 siderophore receptor in *M. haemolytica* is unexpected as *M. haemolytica* is not known to produce
313 or utilize siderophores and has been demonstrated to be unable to utilize ferric enterobactin so it
314 is likely that the identified TBDTs have some other function besides siderophore uptake [11,21].



315

316 **Figure 3.** A) Predicted 3D structures of Genotype 1 (G1, Purple) & Genotype 2 (G2, Orange)
317 ferric enterobactin receptor (frpB) homolog proteins. 3D structure conforms to canonical TonB-
318 dependent receptor structures. B) 2D representation of tertiary structure of G1 and G2 frpB

319 homologs. Each grey rectangle represents one of the 22 β-strand sheets, and the size of the
320 extracellular/periplasmic loops are shown. Differences in the amino acid sequence for select
321 extracellular loops are shown in further detail. Difference in extracellular loops may contribute to
322 differences in substrate binding efficiency and/or result in interactions with completely different
323 substrates.
324

325 G2 specific homology group 2041941 was also predicted to encode a TBDT, and no similar
326 G1 specific/associated protein was identified and the consensus sequence of 2041941 did not
327 correspond to previously identified genes upregulated in response to iron limitation [21]. The
328 identification of another G2 specific TBDT suggests that G2 may have a greater ability to uptake
329 iron and/or other substrates. TBDT have been implicated in transport of a variety of substrates in
330 Gram-negative bacteria including siderophores as well as vitamin B_{12} and carbohydrates [55].
331 Further examination is required to determine the function of this TBDT and whether it provides a
332 competitive advantage to G2.

333 Other iron acquisition proteins include transferrin binding proteins A and B (*tpbAB*), both
334 of which are found in *M. haemolytica* [11]. Transferrin is an iron-carrying glycoprotein found in
335 mammalian circulatory systems that transports iron throughout the body and serves as a sink for
336 any excess free iron released into the circulatory system [56]. TpbB is a surface lipoprotein that
337 extends outwards from the outer membrane to the extracellular space where it binds to transferrin
338 and brings the transferrin closer to TbpA, a TBDT in the outer membrane that transports the
339 transferrin into the periplasm [56]. To investigate any genotype associations of transferrin binding
340 proteins, previously characterized *tbpA* and *tbpB* nucleotide sequences from *M. haemolytica* were
341 used as BLAST queries against the pangenome database. Core homology group 2044546
342 contained hits for *tbpA* while *tbpB* hits were solely identified in G2 specific homology group
343 2044545 with no hits in any G1 genomes. These results suggest that *tbpA* is highly conserved

344 within *M. haemolytica* and likely plays a key role in iron metabolism. In contrast while no BLAST
345 hits were produced to the query *tbpB* nucleotide sequence in G1 genomes, G1 specific homology
346 group 2040773 was predicted to encode for a transferrin binding protein. The amino acid
347 sequences of G1 and G2 variants of the putative tbpB (homology groups 2040773 and 2044545)
348 were compared and found to be 56.5% identical and 81.2% similar. Despite the relatively low
349 similarity of these two proteins, the predicted 3D structures were very similar to one another as
350 well as the crystal structure for the TbpB from *Actinobacillus pneumoniae* (S1 Figure). A
351 difference in the TbpB structure between the genotypes was observed in the N lobe with G2
352 appearing to have a longer alpha-helix structure in the N-lobe, which is the region of the protein
353 involved in transferrin binding [56]. Genotype variation of a transferrin binding protein further
354 suggests they may differ in their ability to acquire iron, which may partially explain observed
355 differences in disease association between the two genotypes. The genotype variants of TbpB
356 identified here may also reflect host preferences of the genotypes as TbpB is selective for host
357 transferrin and differences between TbpB proteins has been partially attributed to this host
358 selectivity [56].

359 IgA proteases are metalloproteases that cleave immunoglobulin A (IgA), a critical
360 component of the mucosal immune system. Two IgA proteases (IgA1 and IgA2) have been
361 previously identified and characterized in a G2 (serotype 1) *M. haemolytica* isolate [18]. We
362 identified a G2 specific IgA protease (homology group 2042189) and a G1 associated IgA protease
363 (2040817) (Table 2). To determine the relationship of these sequences to the previously reported
364 IgA proteases, a BLAST search was performed using previously reported nucleotide sequences
365 for IgA1 and IgA2 as queries against the pangenome database (IgA1:NZ_DS264709 Locus Tag:
366 MHA_RS00165IgA2: NZ_DS264612.1 Locus Tag: MHA_RS09765). At least one BLAST hit

367 was returned for IgA2 in all but five of the included genomes with hits grouped into five homology
368 groups (2041035, 2041036, 2041399, 2041400, and 2043723). Homology groups 2041305 and
369 2041306 were accessory, while 2041399 and 2041400 were unique suggesting some minor
370 variations are likely present within the pangenome for this IgA2 protease. Homology group
371 2043723, however, was a core group suggesting that this IgA2 protease is a core gene of *M.*
372 *haemolytica* and is not genotype differentiating. In contrast, a BLAST hit was returned for IgA1
373 in 146/209 of the genomes with hits grouped into one of three homology groups (2040817,
374 2041298, and 2042189). Homology group 2041298 was found in only one genome so was not
375 examined any further. Group 2040817 is a G1 associated group while 2042189 is a G2 specific
376 group and thus these groups appear to represent genotype variants of an IgA protease.

377 Comparison of the consensus sequences of the two IgA1-like groups revealed that the
378 nucleotide sequence for 2040817 is 2,281 bp while the sequence for 2042189 is 4,512 bp (94%
379 identical across 2,281 shared positions). The G1 associated consensus sequence is much smaller
380 than the previously reported 1,503 amino acids (4,509 bp) for IgA1 and 1,398 amino acids (4,194
381 bp) for IgA2 while the G2 specific consensus sequence is slightly larger than both previously
382 reported IgA protease sequences. The smaller size of the G1 sequence suggests that this sequence
383 may represent a truncated protease or difficulties in the gene finding process due to the large size
384 of these genes [18]. The slightly larger size of the G2 specific sequence may similarly be due to
385 gene finding issues or in generating the consensus sequence. While the identified G1 specific
386 sequence may not represent a complete protease sequence, the observation of a G2 specific
387 sequence and possibility of genotype variation remains of interest and warrants further
388 investigation. Should both genotypes have differing IgA protease sequences the potential

389 difference(s) in their ability to evade the host immune system may contribute to differences in
390 disease association.

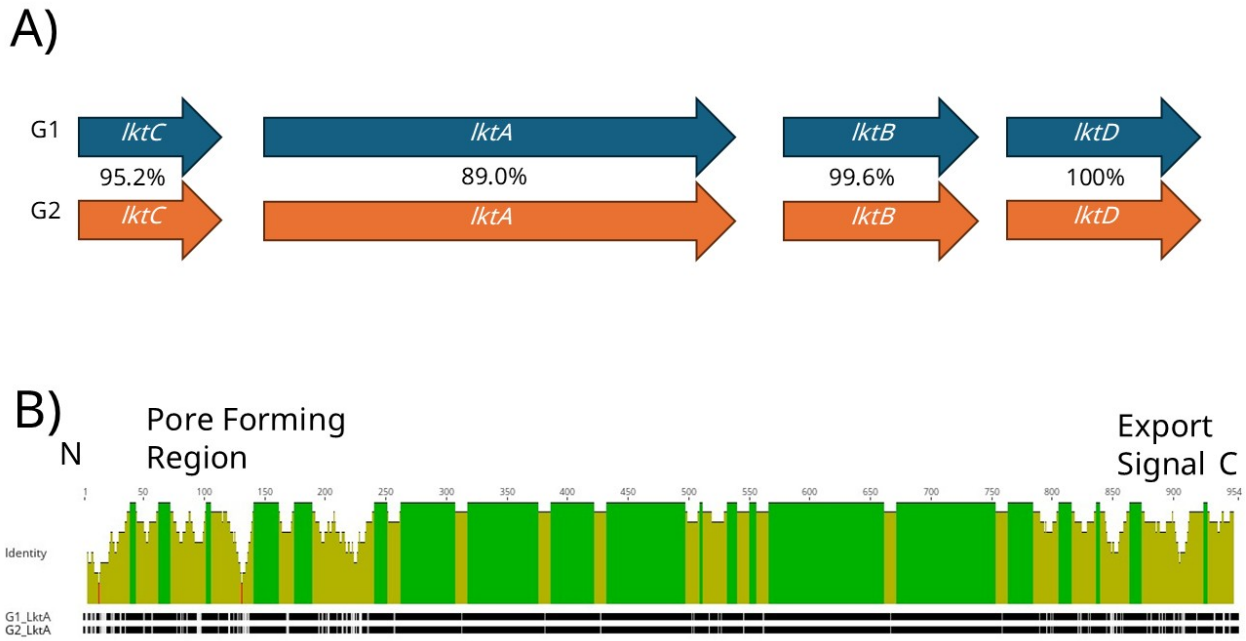
391 *Pasteurella* lipoproteins (Plp) are a group of several immunogenic outer membrane
392 lipoproteins identified in *M. haemolytica* [57]. G2 specific homology group 2044609 was
393 predicted to encode PlpE, which has been investigated as a potential vaccine target for *M.*
394 *haemolytica* because of its immunogenicity [58–60]. Aside from the immunogenic role of PlpE
395 relatively little is known regarding its function although it is important in complement-mediated
396 killing of *M. haemolytica* cells in bovine hosts. Previous investigations of PlpE sequence diversity
397 across serotypes showed that the serotype 1 and 6 (G2) sequences were highly conserved whereas
398 serotype 2 (G1) PlpE sequences were extremely heterogenous, however, antibodies raised against
399 serotype 1 PlpE cross reacted with serotype 2 bacteria [61]. No G1 associated or specific *plpE*
400 gene was identified in this study which may be due to the previously observed heterogeneity of
401 this sequence in G1 genomes or potentially due to the relative lack of available G1 genomes
402 compared to G2. The identification of G2 specific PlpE suggests that G2 isolates may induce
403 stronger immune reactions than G1 isolates, however further research is needed to investigate the
404 prevalence of PlpE in G1and the immunogenicity of G1 PlpE compared to G2-derived PlpE.

405 Homology group 2040817 (G1 specific) was predicted to encode a lytic transglycosylase
406 and is explored in more depth below in the discussion of carbohydrate active enzymes (CAZymes).

407 Leukotoxin operon

408 Leukotoxin (Lkt) is a critical virulence factor in *M. haemolytica* and is produced by both
409 genotypes (all serotypes) [62]. Lkt is encoded in an operon consisting of four genes *(lktCABD*).
410 *lktA* encodes the protoxin, *lktC* encodes the acylating enzyme essential for post-translational

411 modification and activation of the protoxin, and the products of *lktB* and *lktD* are involved in
412 secretion of the toxin [23,63]. *lktCABD* genes were identified and the predicted protein sequences
413 for the Lkt operon genes from G1 and G2 were aligned (Figure 4A). Within the Lkt operon the
414 predicted proteins of the *lktC, lktB* and *lktD* genes were 95.2%, 99.6% and 100% identical,
415 respectively, suggesting these genes are under conservation pressure due to their roles in regulating
416 the secretion and function of Lkt [63,64]. In contrast, LktA protein sequences were only 89%
417 identical. Differences in the predicted LktA proteins of G1 and G2 were concentrated in the N-
418 terminal and C-terminal regions of the protein (Figure 4B). In leukotoxins the N-terminus is
419 involved in pore-forming activities while the C-terminus is involved in export and host immune
420 system recognition [63]. As such any differences in the N-terminal and/or C-terminal regions may
421 potentially influence the activities of the Lkt and thus may result in virulence differences between
422 the two genotypes.



423
424 **Figure 4.** A) Percent identity between the predicted protein sequence of each of the four genes in
425 the leukotoxin operon in G1 and G2. B) Sequence alignment of G1 and G2 LktA amino acid

426 sequence with a sliding window size of 10. A larger peak indicates a greater number of matching
427 amino acids within the window for that position.
428

429 In a previous study, Davies *et al*. examined the sequence diversity of leukotoxin in 31 *M.*
430 *haemolytica,* 6 *Mannheimia glucosida,* and 4 *Pasteurella* [*Bibersteinia*] *trehalosi* isolates
431 identified eight *M. haemolytica* allelic variants of *lktA* [23]. Five of the alleles (A6-A10) were
432 identified solely in ovine isolates, allele A3 was only identified in bovine isolates, while A1 and
433 A2 alleles were identified in both ovine and bovine isolates. The A1 allele was subdivided into 5
434 groups (A1.1-A1.5) with A1.1 solely occurring in bovine isolates from serotype 1 (n = 3) and 6 (n
435 = 1) isolates (corresponding to G2). In bovine isolates from serotype 2 (corresponding to G1), the
436 authors reported detection of alleles A2, A2.1, A2.2, and A3. LktA sequences of study genomes
437 were compared with representative amino acid sequences for each allele and subgroup downloaded
438 from NCBI Genbank. Within G2, 141/146 LktA sequences were 100% identical to the A1.1 allele
439 and another 4/146 were >99.5% identical to A1.1, consistent with the observations of Davies *et al*.
440 (S5-7 Tables). The only exception was the LktA from Genome 2, that shared 100% amino acid
441 identity with A1.3 and A1.5 (S7 Table). Interestingly, all but one of the G1 LktA sequences were
442 99.5-100% identical to the A8.1 allele (Genome 157 was identical to A2) (S5 and S8-9 Tables).
443 The A8.1 allele was found in serotype 2 (G1) isolates by Davies *et al*., however, all the isolates in
444 that study were ovine (n = 3), and A8.1 was also detected in ovine serotype 7 isolates (n = 3). The
445 identification of A8.1 across many bovine origin genomes and nearly complete absence of A2
446 alleles suggests that G1 from Canada and the U.S.A has diverged from the G1 found in the United
447 Kingdom where most of the isolates used by Davies *et al*. were collected and further demonstrates
448 that A8.1 is not restricted to ovine isolates [23]. Whether there is a relationship between specific
449 LktA alleles and respiratory disease severity in cattle warrants further investigation.

450 ## CAZymes

451 Carbohydrate active enzymes (CAZymes) are enzymes whose function has been
452 biochemically verified to be involved in the metabolism of carbohydrates. We chose to investigate
453 any differences in CAZymes between the two genotypes since these may indicate differences in
454 ability to metabolize certain carbohydrates, and thus in the nutritional niches of the genotypes [65].
455 We were particularly interested in enzymes predicted to be secreted into either the periplasmic or
456 extracellular space (i.e. contain a signal peptide) and are thus most likely to be involved in
457 interactions with the external environment. A total of six CAZymes containing signal peptides
458 were identified in the pangenome (S10 Table) (50). At least one CAZyme was predicted in all
459 genomes except for Genome 86/GCF_022404935, which was removed from this analysis.
460 CAZymes identified were members of families GH23 (lytic transglycosylase), GH33 (sialidase),
461 GH13_19 (amylose glucosidase), GT9 (glycosyltransferase), GH102 (lytic transglycosylase) and
462 GH103 (lytic transglycosylase). Only GH103 (homology group 2042729) was genotype
463 associated, present in 60/61 (98.4%) G1 genomes and 2/146 (1.37%) G2 genomes. The GH103
464 family consists of a group of lytic transglycosylases that cleave the β-1,4 linkage between *N*-
465 acetylmuramoyl and *N*-acetylglucosaminyl residues in peptidoglycan and belong to family 3 of
466 the transglycosylases with structures similar to lysozyme [66,67]. Numerous biological functions
467 have been attributed lytic transglycosylases, however the only clear role is that they are involved
468 in cell wall remodelling [66–69]. Whether this genotype difference in lytic transglycosylases is
469 associated with phenotypic differences in colony morphology of the two genotypes [26] or other
470 differences in cell wall maintenance will require further investigation.

471 # Conclusions

472 We identified a total of 422 *Mannheimia haemolytica* genotype differentiating genes (135
473 G1 specific, 21 G1 associated, 233 G2 specific, and 33 G2 associated), including 125 with
474 unknown functions and 132 that are associated with mobile genetic elements. Genotype associated
475 genes and variants that may contribute to differences in disease association include TonB-
476 dependent transporters, IgA proteases, a highly immunogenic outer membrane protein (PlpE), a
477 lytic transglycosylase, and genotype variants of the leukotoxin gene. The list of genotype
478 differentiating genes provides a shortlist of targets that warrant further investigation to provide a
479 mechanistic understanding of phenotypic differences between the *M. haemolytica* genotypes and
480 inform rational strategies for *M. haemolytica* and bovine respiratory disease diagnostics.

481 # Acknowledgements

482 The authors are grateful to Dr. David Palmer (Department of Chemistry, University of
483 Saskatchewan) for assistance with the protein structure alignments, and to Dr. Maarten Voordouw
484 (Department of Veterinary Microbiology, University of Saskatchewan) for assistance with
485 statistical analysis and approaches.

486 # Supporting Information

487 **S1 Table**. Summary of Genomes Included in Pangenome Analysis

488

489 **S2 Table**. Accession numbers of sequences used as BLAST queries against pangenome database.

490

491 **S3 Table**. Overview of pangenome classification of all homology groups. Genes matching is the
492 number of genes that were found to correspond to the given homology group, as this referred to
493 genes the number may be greater than or less than the number of genomes examined. For
494 pangenome classification: core = group was present in ≥90% of all genomes (187/209); unique =
495 Group was present in exactly one genome; Accessory = Group was present in somewhere between
496 2-186 genomes. For phenotype (Genotype) classification: Shared = Group was present in ≥90%
497 of the genomes of the given genotype; exclusive = Group was present solely in genomes of the
498 given genotype; Specific = Group was exclusive to one genotype and present in ≥ 90% of that
499 genotype; Associated = Group was present in ≥90% of the genomes of one genotype and ≤5% of
500 the other genotype.
501
502 **S4 Table.** Genotype Specific and Associated Homology Groups. Annotation notes column
503 provides summary of the interpro, pfam and GO functional annotations. Genes matching is the
504 number of genes that were found to correspond to the given homology group, as this referred to
505 genes the number may be greater than or less than the number of genomes examined. For
506 pangenome classification: core = group was present in ≥90% of all genomes (187/209); unique =
507 Group was present in exactly one genome; Accessory = Group was present in somewhere between
508 2-186 genomes. For phenotype (Genotype) classification: Shared = Group was present in ≥90%
509 of the genomes of the given genotype; exclusive = Group was present solely in genomes of the
510 given genotype; Specific = Group was exclusive to one genotype and present in ≥ 90% of that
511 genotype; Associated = Group was present in ≥90% of the genomes of one genotype and ≤5% of
512 the other genotype.
513

514 **S5 Table**. Pairwise comparisons of amino acid sequences of genotype 1 (G1) and genotype 2 (G2)
515 Leukotoxin A consensus sequences against representatives of LktA alleles from Davies *et al*. 2001
516
517 **S6 Table**. Pairwise comparison of all genotype 1 LktA nucleotide sequences.
518
519 **S7 Table**. Pairwise comparison of genotype 1 LktA amino acid sequences that had a divergent
520 LktA nucleotide sequence to representative sequences of LktA alleles identified in Davies *et al*.
521 2001.
522
523 **S8 Table**. Pairwise comparison of all genotype 2 LktA nucleotide sequences.
524
525 **S9 Table**. Pairwise comparison of genotype 1 LktA amino acid sequences that had a divergent
526 LktA nucleotide sequence to representative sequences of LktA alleles identified in Davies *et al*.
527 2001.
528
529 **S10 Table**. Summary of CAZyme Analysis Results. Only CAZymes that were predicted to contain
530 a signal peptide and were annotated by 3 of the utilized annotation tools (dbCAN_sub,
531 DIAMOND, & HMMER) were included. No CAZymes were predicted in genome
532 86/GCF_022404935 and so was excluded from this analysis.
533
534 **S1 Figure**. Predicted structures of G1 and G2 variants of a predicted transferrin binding protein B
535 (TbpB). (A) The crystal structure of TbpB from *Actinobacillus pleuropneumoniae* (PDB accession
536 3HOL) was used as a reference. Major domains include the N-lobe and C-lobe, each containing a

537 beta-barrel and a “handle”. The reference structure is “rainbow” coloured from N terminus (blue)
538 to C terminus (red). (B) Alignment of G1 and G2 variants. (C) G2 aligned to reference. (D) G1
539 aligned to reference. The G1 variant has an elongated alpha-helix in the N-lobe handle domain
540 relative to G2 and the reference. The N-lobe has been demonstrated to interact with transferrin.
541 Structures and alignments were visualized in PyMol.
542
543

544 # References

545 1. Rice JA, Carrasco-Medina L, Hodgins DC, Shewen PE. *Mannheimia haemolytica* and
546 bovine respiratory disease. Animal Health Research Reviews. 2007;8: 117–128.
547 doi:10.1017/S1466252307001375

548 2. Highlander SK. Molecular genetic analysis of virulence in *Mannheimia (Pasteurella)*
549 *haemolytica*. FBL. 2001;6: 1128–1150. doi:10.2741/highland

550 3. DeDonder KD, Apley MD. A literature review of antimicrobial resistance in pathogens
551 associated with bovine respiratory disease. Animal Health Research Reviews. 2015;16:
552 125–134. doi:10.1017/S146625231500016X

553 4. Blakebrough-Hall C, McMeniman JP, González LA. An evaluation of the economic effects
554 of bovine respiratory disease on animal performance, carcass traits, and economic outcomes
555 in feedlot cattle defined using four BRD diagnosis methods. Journal of Animal Science.
556 2020;98: skaa005. doi:10.1093/jas/skaa005

557 5. Clawson ML, Murray RW, Sweeney MT, Apley MD, DeDonder KD, Capik SF, et al.
558 Genomic signatures of *Mannheimia haemolytica* that associate with the lungs of cattle with
559 respiratory disease, an integrative conjugative element, and antibiotic resistance genes.
560 BMC Genomics. 2016;17: 982. doi:10.1186/s12864-016-3316-8

561 6. Biberstein EL, Gills M, Knight H. Serological types of Pasteurella hemolytica. Cornell Vet.
562 1960;50: 283–300.

563 7. Al-Ghamdi GM, Ames TR, Baker JC, Walker R, Chase CCL, Frank GH, et al. Serotyping
564 of *Mannheimia (Pasteurella) haemolytica* Isolates from the Upper Midwest United States. J
565 VET Diagn Invest. 2000;12: 576–578. doi:10.1177/104063870001200617

566 8. Mason C, Errington J, Foster G, Thacker J, Grace O, Baxter-Smith K. *Mannheimia*
567 *haemolytica* serovars associated with respiratory disease in cattle in Great Britain. BMC
568 Veterinary Research. 2022;18: 5. doi:10.1186/s12917-021-03121-3

569 9. Klima CL, Alexander TW, Hendrick S, McAllister TA. Characterization of *Mannheimia*
570 *haemolytica* isolated from feedlot cattle that were healthy or treated for bovine respiratory
571 disease. Can J Vet Res. 2014;78: 38–45.

572 10. Arnal JL, Fernández A, Vela AI, Sanz C, Fernández-Garyzábal JF, Cid D. Capsular type
573 diversity of *Mannheimia haemolytica* determined by multiplex real-time PCR and indirect
574 hemagglutination in clinical isolates from cattle, sheep, and goats in Spain. Veterinary
575 Microbiology. 2021;258: 109121. doi:10.1016/j.vetmic.2021.109121

576 11. Confer AW, Martino P. *Mannheimia haemolytica*. CABI Compendium. 2017;CABI
577 Compendium: 70912. doi:10.1079/cabicompendium.70912

578 12. Klima CL, Zaheer R, Cook SR, Booker CW, Hendrick S, Alexander TW, et al. Pathogens
579 of Bovine Respiratory Disease in North American Feedlots Conferring Multidrug
580 Resistance via Integrative Conjugative Elements. Journal of Clinical Microbiology.
581 2014;52: 438–448. doi:10.1128/jcm.02485-13

582 13. Holman DB, McAllister TA, Topp E, Wright A-DG, Alexander TW. The nasopharyngeal
583 microbiota of feedlot cattle that develop bovine respiratory disease. Vet Microbiol.
584 2015;180: 90–95. doi:10.1016/j.vetmic.2015.07.031

585 14. Noyes NR, Benedict KM, Gow SP, Booker CW, Hannon SJ, McAllister TA, et al.
586 *Mannheimia haemolytica* in feedlot cattle: prevalence of recovery and associations with
587 antimicrobial use, resistance, and health outcomes. J Vet Intern Med. 2015;29: 705–713.
588 doi:10.1111/jvim.12547

589 15. Kaper JB, Nataro JP, Mobley HL. Pathogenic *Escherichia coli*. Nat Rev Microbiol. 2004;2:
590 123–140. doi:10.1038/nrmicro818

591 16. Nysten J, Sofras D, Dijck PV. One species, many faces: The underappreciated importance
592 of strain diversity. PLOS Pathogens. 2024;20: e1011931. doi:10.1371/journal.ppat.1011931

593 17. Hacker J, Carniel E. Ecological fitness, genomic islands and bacterial pathogenicity. A
594 Darwinian view of the evolution of microbes. EMBO Rep. 2001;2: 376–381.
595 doi:10.1093/embo-reports/kve097

596 18. Ayalew S, Murdock BK, Snider TA, Confer AW. *Mannheimia haemolytica* IgA-specific
597 proteases. Vet Microbiol. 2019;239: 108487. doi:10.1016/j.vetmic.2019.108487

598 19. Singh K, Ritchey JW, Confer AW. *Mannheimia haemolytica*: Bacterial–Host Interactions in
599 Bovine Pneumonia. Vet Pathol. 2011;48: 338–348. doi:10.1177/0300985810377182

600 20. Kisiela DI, Czuprynski CJ. Identification of *Mannheimia haemolytica* Adhesins Involved in
601 Binding to Bovine Bronchial Epithelial Cells. Infection and Immunity. 2009;77: 446–455.
602 doi:10.1128/iai.00312-08

603 21. Roehrig SC, Tran HQ, Spehr V, Gunkel N, Selzer PM, Ullrich HJ. The response of
604 *Mannheimia haemolytica* to iron limitation: Implications for the acquisition of iron in the
605 bovine lung. Veterinary Microbiology. 2007;121: 316–329.
606 doi:10.1016/j.vetmic.2006.12.013

607 22. Davies RL, Baillie S. Cytotoxic activity of *Mannheimia haemolytica* strains in relation to
608 diversity of the leukotoxin structural gene lktA. Vet Microbiol. 2003;92: 263–279.
609 doi:10.1016/s0378-1135(02)00408-x

610 23. Davies RL, Whittam TS, Selander RK. Sequence Diversity and Molecular Evolution of the
611 Leukotoxin (lktA) Gene in Bovine and Ovine Strains of *Mannheimia (Pasteurella)*
612 *haemolytica*. J Bacteriol. 2001;183: 1394–1404. doi:10.1128/jb.183.4.1394-1404.2001

613 24. Cozens D, Sutherland E, Lauder M, Taylor G, Berry CC, Davies RL. Pathogenic
614 *Mannheimia haemolytica* Invades Differentiated Bovine Airway Epithelial Cells. Infect
615 Immun. 2019;87: 10.1128/iai.00078-19. doi:10.1128/iai.00078-19

616 25. Loy JD, Clawson ML. Rapid typing of *Mannheimia haemolytica* major genotypes 1 and 2
617 using MALDI-TOF mass spectrometry. J Microbiol Methods. 2017;136: 30–33.
618 doi:10.1016/j.mimet.2017.03.002

619 26. Wynn EL, Schuller G, Loy JD, Workman AM, McDaneld TG, Clawson ML.
620 Differentiation of *Mannheimia haemolytica* genotype 1 and 2 strains by visible phenotypic

621 characteristics on solid media. Journal of Microbiological Methods. 2020;171: 105877.
622 doi:10.1016/j.mimet.2020.105877

623 27. Clawson ML, Schuller G, Dickey AM, Bono JL, Murray RW, Sweeney MT, et al.
624 Differences between predicted outer membrane proteins of genotype 1 and 2 *Mannheimia*
625 *haemolytica*. BMC Microbiology. 2020;20: 250. doi:10.1186/s12866-020-01932-2

626 28. Klima CL, Cook SR, Zaheer R, Laing C, Gannon VP, Xu Y, et al. Comparative Genomic
627 Analysis of *Mannheimia haemolytica* from Bovine Sources. PLOS ONE. 2016;11:
628 e0149520. doi:10.1371/journal.pone.0149520

629 29. Deschner D, Voordouw MJ, Fernando C, Campbell J, Waldner CL, Hill JE. Identification
630 of genetic markers of resistance to macrolide class antibiotics in *Mannheimia haemolytica*
631 isolates from a Saskatchewan feedlot. Appl Environ Microbiol. 2024;90: e00502-24.
632 doi:10.1128/aem.00502-24

633 30. Dassanayake RP, Clawson ML, Tatum FM, Briggs RE, Kaplan BS, Casas E. Differential
634 identification of *Mannheimia haemolytica* genotypes 1 and 2 using colorimetric loop-
635 mediated isothermal amplification. BMC Research Notes. 2023;16: 4. doi:10.1186/s13104-
636 023-06272-8

637 31. O'Leary NA, Cox E, Holmes JB, Anderson WR, Falk R, Hem V, et al. Exploring and
638 retrieving sequence and metadata for species across the tree of life with NCBI Datasets. Sci
639 Data. 2024;11: 732. doi:10.1038/s41597-024-03571-y

640 32. Okonechnikov K, Golosova O, Fursov M, UGENE team. Unipro UGENE: a unified
641 bioinformatics toolkit. Bioinformatics. 2012;28: 1166–1167.
642 doi:10.1093/bioinformatics/bts091

643 33. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3--
644 new capabilities and interfaces. Nucleic Acids Res. 2012;40: e115. doi:10.1093/nar/gks596

645 34. Kõressaar T, Lepamets M, Kaplinski L, Raime K, Andreson R, Remm M. Primer3_masker:
646 integrating masking of template sequence with primer design software. Bioinformatics.
647 2018;34: 1937–1938. doi:10.1093/bioinformatics/bty036

648 35. Koressaar T, Remm M. Enhancements and modifications of primer design program
649 Primer3. Bioinformatics. 2007;23: 1289–1291. doi:10.1093/bioinformatics/btm091

650 36. Jonkheer EM, van Workum D-JM, Sheikhizadeh Anari S, Brankovics B, de Haan JR, Berke
651 L, et al. PanTools v3: functional annotation, classification and phylogenomics.
652 Bioinformatics. 2022;38: 4403–4405. doi:10.1093/bioinformatics/btac506

653 37. Sheikhizadeh S, Schranz ME, Akdel M, de Ridder D, Smit S. PanTools: representation,
654 storage and exploration of pan-genomic data. Bioinformatics. 2016;32: i487–i493.
655 doi:10.1093/bioinformatics/btw455

656 38. Anari SS, Ridder D de, Schranz ME, Smit S. Pangenomic read mapping. bioRxiv; 2019. p.
657 813634. doi:10.1101/813634

658 39. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-
659 scale protein function classification. Bioinformatics. 2014;30: 1236–1240.
660 doi:10.1093/bioinformatics/btu031

661 40. Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, et al. The
662 InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 2021;49:
663 D344–D354. doi:10.1093/nar/gkaa977

664 41. Nielsen H. Predicting Secretory Proteins with SignalP. Methods Mol Biol. 2017;1611: 59–
665 73. doi:10.1007/978-1-4939-7015-5_6

666 42. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, et al. antiSMASH 7.0:
667 new and improved predictions for detection, regulation, chemical structures and
668 visualisation. Nucleic Acids Res. 2023;51: W46–W50. doi:10.1093/nar/gkad344

669 43. Deschner D, Hill JE. Genomic Differentiation of *Mannheimia haemolytica* genotypes.
670 FigShare; 2025. doi:10.6084/m9.figshare.28872536

671 44. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+:
672 architecture and applications. BMC Bioinformatics. 2009;10: 421. doi:10.1186/1471-2105-
673 10-421

674 45. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST
675 and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res.
676 1997;25: 3389–402.

677 46. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool.
678 J Mol Biol. 1990;215: 403–10.

679 47. Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA
680 sequences. J Comput Biol. 2000;7: 203–214. doi:10.1089/10665270050081478

681 48. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA. Database
682 indexing for production MegaBLAST searches. Bioinformatics. 2008;24: 1757–1764.
683 doi:10.1093/bioinformatics/btn322

684 49. Kostova V, Hanke D, Kaspar H, Fiedler S, Schwarz S, Krüger-Haker H. Macrolide
685 resistance in *Mannheimia haemolytica* isolates associated with bovine respiratory disease
686 from the German national resistance monitoring program GERM-Vet 2009 to 2020. Front
687 Microbiol. 2024;15. doi:10.3389/fmicb.2024.1356208

688 50. Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y. dbCAN3: automated carbohydrate-
689 active enzyme and substrate annotation. Nucleic Acids Research. 2023;51: W115–W121.
690 doi:10.1093/nar/gkad328

691 51. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure
692 prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630: 493–500.
693 doi:10.1038/s41586-024-07487-w

694 52. Vanni C, Schechter MS, Acinas SG, Barberán A, Buttigieg PL, Casamayor EO, et al.
695 Unifying the known and unknown microbial coding sequence space. Brown CT, Storz G,
696 Brown CT, Smith B, editors. eLife. 2022;11: e67667. doi:10.7554/eLife.67667

697 53. Lobb B, Tremblay BJ-M, Moreno-Hagelsieb G, Doxey AC. PathFams: statistical detection
698 of pathogen-associated protein domains. BMC Genomics. 2021;22: 663.
699 doi:10.1186/s12864-021-07982-8

700 54. Ferguson AD, Deisenhofer J. TonB-dependent receptors-structural perspectives. Biochim
701 Biophys Acta. 2002;1565: 318–332. doi:10.1016/s0005-2736(02)00578-3

702 55. Silale A, Van Den Berg B. TonB-Dependent Transport Across the Bacterial Outer
703 Membrane. Annu Rev Microbiol. 2023;77: 67–88. doi:10.1146/annurev-micro-032421-
704 111116

705 56. Pogoutse AK, Moraes TF. Iron acquisition through the bacterial transferrin receptor. Crit
706 Rev Biochem Mol Biol. 2017;52: 314–326. doi:10.1080/10409238.2017.1293606

707 57. Ayalew S, Shrestha B, Montelongo M, Wilson AE, Confer AW. Identification and
708 immunogenicity of *Mannheimia haemolytica* S1 outer membrane lipoprotein PlpF.
709 Vaccine. 2011;29: 8712–8718. doi:10.1016/j.vaccine.2011.08.074

710 58. Ayalew S, Confer AW, Blackwood ER. Characterization of immunodominant and
711 potentially protective epitopes of *Mannheimia haemolytica* serotype 1 outer membrane
712 lipoprotein PlpE. Infect Immun. 2004;72: 7265–7274. doi:10.1128/IAI.72.12.7265-
713 7274.2004

714 59. Confer AW, Ayalew S, Panciera RJ, Montelongo M, Wray JH. Recombinant *Mannheimia*
715 *haemolytica* serotype 1 outer membrane protein PlpE enhances commercial *M. haemolytica*
716 vaccine-induced resistance against serotype 6 challenge. Vaccine. 2006;24: 2248–2255.
717 doi:10.1016/j.vaccine.2005.11.036

718 60. Pandher K, Confer AW, Murphy GL. Genetic and immunologic analyses of PlpE, a
719 lipoprotein important in complement-mediated killing of Pasteurella haemolytica serotype
720 1. Infect Immun. 1998;66: 5613–5619. doi:10.1128/IAI.66.12.5613-5619.1998

721 61. Ayalew S, Blackwood ER, Confer AW. Sequence diversity of the immunogenic outer
722 membrane lipoprotein PlpE from *Mannheimia haemolytica* serotypes 1, 2, and 6. Vet
723 Microbiol. 2006;114: 260–268. doi:10.1016/j.vetmic.2005.11.067

724 62. Oppermann T, Busse N, Czermak P. *Mannheimia haemolytica* growth and leukotoxin
725 production for vaccine manufacturing — A bioprocess review. Electronic Journal of
726 Biotechnology. 2017;28: 95–100. doi:10.1016/j.ejbt.2017.06.001

727 63. Narayanan SK, Nagaraja TG, Chengappa MM, Stewart GC. Leukotoxins of gram-negative
728 bacteria. Vet Microbiol. 2002;84: 337–356. doi:10.1016/s0378-1135(01)00467-9

729 64. Larsen J, Pedersen AG, Davies RL, Kuhnert P, Frey J, Christensen H, et al. Evolution of the
730 leukotoxin promoter in genus *Mannheimia*. BMC Evolutionary Biology. 2009;9: 121.
731 doi:10.1186/1471-2148-9-121

732 65. Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-
733 active enzyme database: functions and literature. Nucleic Acids Res. 2022;50: D571–D577.
734 doi:10.1093/nar/gkab1045

735 66. Dik DA, Marous DR, Fisher JF, Mobashery S. Lytic transglycosylases: concinnity in
736 concision of the bacterial cell wall. Crit Rev Biochem Mol Biol. 2017;52: 503–542.
737 doi:10.1080/10409238.2017.1337705

738 67. Martinez-Bond EA, Soriano BM, Williams AH. The mechanistic landscape of Lytic
739 transglycosylase as targets for antibacterial therapy. Curr Opin Struct Biol. 2022;77:
740 102480. doi:10.1016/j.sbi.2022.102480

741 68. Suvorov M, Lee M, Hesek D, Boggess B, Mobashery S. Lytic Transglycosylase MltB of
742 *Escherichia coli* and Its Role in Recycling of Peptidoglycan Strands of Bacterial Cell Wall.
743 J Am Chem Soc. 2008;130: 11878–11879. doi:10.1021/ja805482b

744 69. Scheurwater E, Reid CW, Clarke AJ. Lytic transglycosylases: bacterial space-making
745 autolysins. Int J Biochem Cell Biol. 2008;40: 586–591. doi:10.1016/j.biocel.2007.03.018

746