

Galactocerebroside biosynthesis pathways of *Mycoplasma* species: an antigen triggering Guillain–Barré–Stohl syndrome

Erika Gaspari¹  Jasper J. Koehorst,¹  Joachim Frey,²  Vitor A.P. Martins dos Santos^{1,3}  and Maria Suarez-Diez¹ 

¹Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Wageningen, the Netherlands.

²Vetsuisse, University of Bern, Bern, Switzerland.

³LifeGlimmer GmbH, Berlin, Germany.

Summary

Infection by *Mycoplasma pneumoniae* has been identified as a preceding factor of Guillain–Barré–Stohl syndrome. The Guillain–Barré–Stohl syndrome is triggered by an immune reaction against the major glycolipids and it has been postulated that *M. pneumoniae* infection triggers this syndrome due to bacterial production of galactocerebroside. Here, we present an extensive comparison of 224 genome sequences from 104 *Mycoplasma* species to characterize the genetic determinants of galactocerebroside biosynthesis. Hidden Markov models were used to analyse glycosyl transferases, leading to identification of a functional protein domain, termed M2000535 that appears in about a third of the studied genomes. This domain appears to be associated with a potential UDP-glucose epimerase, which converts UDP-glucose into UDP-galactose, a main substrate for the biosynthesis of galactocerebroside. These findings clarify the pathogenic mechanisms underlining the triggering of Guillain–Barré–Stohl syndrome by *M. pneumoniae* infections.

Introduction

Mycoplasma species are bacteria representing the smallest free-living organisms on earth. They include several pathogens infecting human, animals and plants. *Mycoplasma* is characterized by the lack of a cell wall, reason for which glycolipids of the membrane are exposed to the host's immune system upon an infection. *Mycoplasma pneumoniae* is a human respiratory pathogen causing atypical (or walking) pneumonia, accounting for approximately 20–30% of all types of pneumonia worldwide (Liu *et al.*, 2009; Varma-Basil *et al.*, 2009; Waites & Talkington, 2004; Zhang *et al.*, 2016; Waites *et al.*, 2017). Diagnosis of *M. pneumoniae* infections is currently performed mostly by PCR tests but remains complicated at an early stage of infection (Miyachi *et al.*, 2009). Since about a decade, research to improve early diagnosis has focused on glycolipid antigens present on the membrane of *M. pneumoniae* (Matsuda, 2015). The percentage of glycolipids in *M. pneumoniae*'s membrane varies between 6% and 10% of total lipids (Gaspari *et al.*, 2019).

For long, *M. pneumoniae* has been suspected as a potential preceding factor of Guillain–Barré–Stohl syndrome (GBS) (Ang *et al.*, 2002; Yuki, 2007), which occurs at a frequency of about 5% of the total cases of past *M. pneumoniae* infections (van den Berg *et al.*, 2014; Meyer Sauter *et al.*, 2016). The GBS is an autoimmune neurological disorder that is potentially life threatening. *Campylobacter jejuni* is the first microorganism that was associated with post-infectious outbreak of GBS (Rees *et al.*, 1995) and has been found to perform galactocerebroside biosynthesis (Hao *et al.*, 1998). Galactocerebroside has been shown to be immunogenic to a low degree in *M. pneumoniae* infections (Kusunoki *et al.*, 2001; Susuki *et al.*, 2004), and it is postulated that *M. pneumoniae* triggers GBS by inducing anti-galactocerebroside IgG (Meyer Sauter *et al.*, 2018; Smolders *et al.*, 2019). In a clinical study about, a third of patients with central nervous system infections by *M. pneumoniae* revealed anti-GalC antibodies indicating that *Mycoplasma pneumoniae* also might induce other CNS symptoms by other mechanisms (Christie *et al.*, 2007).

Galactocerebroside, also called galactosylceramide, is a sphingolipid, more specifically a cerebroside,

Received 20 June, 2020; accepted 22 February, 2021.

For correspondence: *E-mail erika.gaspariwur@gmail.com; Tel. +31647889939. **E-mail maria.suarezdiez@wur.nl.

Microbial Biotechnology (2021) 0(0), 1–11

doi:10.1111/1751-7915.13794

Funding information

This work has received funding from European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement n. 634942 (MycSynVac), as well as from TTW-NWO project SafeChassis, grant agreement n.15814.

© 2021 The Authors. *Microbial Biotechnology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

characterized by a galactosyl head group. A similar compound is glucocerebroside, alias glucosylceramide, which instead consists in a cerebroside where the monosaccharide head group is glucose. Glycosphingolipids such as galactocerebroside and glucocerebroside are typically synthesized in bacteria by the enzymes, ceramide galactosyltransferase (or galactosylceramide synthase – reaction EC 2.4.1.47) and ceramide glucosyltransferase (or glucosylceramide synthase – reaction EC 2.4.1.80) of the glycosyltransferase family. The enzyme ceramide galactosyltransferase appears in viruses and cellular organisms; it synthesizes the biosynthesis of galactocerebroside by binding UDP-galactose to a ceramide molecule, releasing UDP.

The glycosyltransferase of *M. pneumoniae*, encoded by gene *mpn483*, has been shown to synthesize galactosylceramide (most likely the beta-variant) using ceramide and UDP-glucose as substrates, both with phosphatidylglycerol or cardiolipin as activators. *M. pneumoniae* has access to ceramide and galactose. It imports ceramide both from *in vitro* growth in axenic medium and from the host *in vivo* during infection (Klement *et al.*, 2007). In addition, it can use the fatty acid chains from incorporated ceramide in other lipids to build up ceramide-based glycolipids. Moreover, *M. pneumoniae* favours the import of glucose *in vivo*, albeit not *in vitro*, in which galactose is preferred (Plackett *et al.*, 1969). Finally, it is postulated but not confirmed that *M. pneumoniae* contains a potential epimerase converting UDP-glucose into UDP-galactose (Dandekar *et al.*, 2000).

Characterization of the galactocerebroside biosynthesis pathway in *M. pneumoniae* will further clarify pathogenic mechanisms and can greatly impact the development of methods early detection and diagnosis of GBS. Moreover, these data are essential for the design of alternative metabolic pathways for *M. pneumoniae* avoiding galactocerebroside formation and to identify alternative *Mycoplasma* species devoid of galactocerebroside for biomedical applications.

We present here a comparative analysis investigation of 9 strains of *M. pneumoniae* and additional 103 currently genome-sequenced *Mycoplasma* species. The goal is to identify the genetic determinants of galactocerebroside biosynthesis and to further characterize the proteins in this pathway.

Experimental procedures

All available and complete *Mycoplasma* genome sequences were retrieved from the NCBI Genome repository. Overall, 224 genome sequences were obtained, belonging to 104 species that are listed in File S1. All genome sequences were re-annotated using the SAPP pipeline (Koeberst *et al.*, 2017). Gene prediction was

performed using Prodigal 2.6.3 (Hyatt *et al.*, 2010), and protein sequences were annotated using InterProScan 5.36.75.0 (Jones *et al.*, 2014) to assign functional domains. Annotation data were stored in a triple-store (GraphDB) (Güting, 1994) in a linked data format using the GBOL ontology as schema (van Dam *et al.*, 2019) and systematically queried using SPARQL.

Hidden Markov Models (HMMs) were built and searched for with HMMER v3.3 (Eddy, 1998) – through commands *hmmbuild* and *hmmsearch* – on multiple sequence alignments (MSAs) performed with Clustal Omega 1.2.4 (Sievers *et al.*, 2011; Madeira *et al.*, 2019).

Sequence logos have been generated with WebLogo version 3 (Crooks *et al.*, 2004).

Results

We studied the functionalities associated with the *M. pneumoniae* genome, through an analysis of protein domains, to describe the biosynthesis pathway that leads to the formation of galactocerebroside and compare it to the pathways found in other mycoplasmas. Therefore, we analysed 224 genomes of 104 *Mycoplasma* species (given in File S1), comprising 213 strains. To ensure uniform annotation and a consistent comparison, all genomes were re-annotated.

The synthesis of glycosphingolipid such as galactocerebroside, in bacteria, needs a glycosyltransferase enzyme, linking a sugar (galactose) to a phospholipid (ceramide) and building the glycosyl bond. Our functional analysis identified 4 genes containing a glycosyltransferase domain in *M. pneumoniae*. These have been found in the 9 analysed strains and correspond to locus tags *mpn028*, *mpn483*, *mpn075* and *mpn064* in *M. pneumoniae* M129. While *mpn064* (*deoA*) codes for a thymidine phosphorylase (EC 2.4.2.2) that contains a 'glycosyl transferase family 3' domain with InterPro identifier IPR000312, *mpn028*, *mpn483* and *mpn075* contain a 'glycosyltransferase 2-like' domain, (IPR001173), associated with proteins that have been linked to glycosphingolipids biosynthesis pathways and, specifically, to proteins that have been proven to own glycosyltransferase activity (Sobhanifar *et al.*, 2016). Their lengths and E-values of associated glycosyltransferase are

Table 1. The three enzymes MPN_028, MPN_075 and MPN_483 of *Mycoplasma pneumoniae* M129 matching the InterPro domain IPR001173 ("Glycosyltransferase 2-like"), with E-values and gene length.

<i>Mycoplasma pneumoniae</i> M129 Enzyme	E-value: glycosyltransferase 2-like domain	Gene length (bp)
MPN_028	2.7×10^{-26}	899
MPN_483	1.2×10^{-25}	1025
MPN_075	1.2×10^{-21}	899

shown in Table 1. The glycosyltransferase with more significant E-value is MPN_028, while the one known to synthesize galactocerebroside is MPN_483 (Klement *et al.*, 2007).

The InterPro IPR001173 domain comprises two Pfam domains: 'Glyco_trans_2_3' (PF13632) and 'Glyco_transf_2' (PF00535); this last is found in two glycosyltransferases of *C. jejuni*, with E-values of 10^{-29} (Putative galactosyltransferase – UniProtKB Q8KWR2) and 10^{-27} (Beta-1,3-galactosyltransferase coded by gene *cgtB* – UniProtKB Q5DT13). Focusing on the *Mycoplasma* species, we therefore assumed the protein sequences containing the Pfam domain PF00535 to be the ones associated with galactocerebroside synthesis. Thus, we continued our analysis focusing on PF00535 and discarded further analysis of PF13632. The distribution of the E-values of this signature, presented in Fig. 1, shows a bimodal distribution. The presence of two peaks suggests two similar but distinct domains. Thus, we re-built HMMs on the two separate groups obtaining two new domains, M100535 and M200535. Then, E-values for these domains on the sequences were re-calculated, as indicated in Fig. 1. This approach results in two groups of protein sequences each matching its corresponding motif with much higher significance, that is much lower E-values. All the strains of *M. pneumoniae* contain only the second motif M200535. The HMM for M200535 motif is provided in File S2. The two glycosyltransferases of *C. jejuni* match M200535 with much higher significance than PF00535: the beta-1,3-galactosyltransferase with E-value 10^{-43} and putative galactosyltransferase with E-value 10^{-42} . An additional glycosyltransferase family 2 protein matching M200535 with E-value 10^{-42} was identified in *C. jejuni*, and full results are given in File S3.

The two new M100535 and M200535 motifs show some substantial differences: the sequence alignments of the proteins carrying the domains reveal M200535 to be almost a twice as long as M100535. Most differences between the two motifs are present in the first part of the domain, where M200535-containing sequences show a predominance of aspartic acid in positions 344, 348, 428, 430, 436 and 455 (Fig. 2). Consensus sequences for both M100535 and M200535 are provided in File S4.

Occurrences of the domains M100535, M200535, here defined, and of domains PF00534, PF13439 or PF13641 associated with alternative glycosyltransferases are summarized in Table 2. In total, 73 out of the 104 *Mycoplasma* species analysed match at least one glycosyltransferase domain. It should be noted that any genome containing either M100535 or M200535 also contains PF00535.

In silico functional analysis on *M. pneumoniae* reveals that *mpn257* contains Pfam domain 'GDP-mannose 4,6 dehydratase' (PF16363) comprised in the InterPro domain IPR016040 (E-value: 2.4×10^{-46}), found in

sequences annotated as UDP-6-glucose 4-epimerases that interconvert UDP-glucose into UDP-galactose. Therefore, we assume *M. pneumoniae* has the ability of converting UDP-glucose into UDP-galactose.

We can conclude the pathway for galactocerebroside synthesis in *Mycoplasma pneumoniae in vivo* is most likely as represented in Fig. 3.

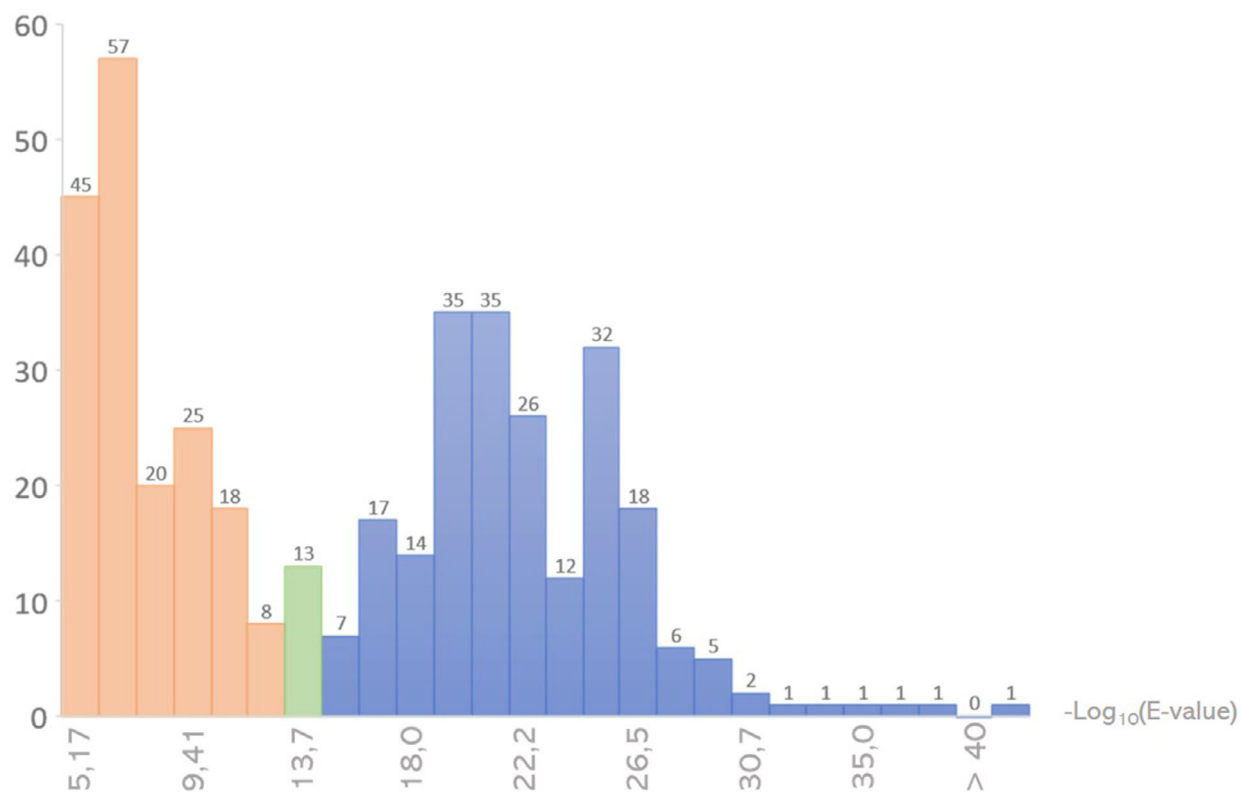
Domain Pfam PF01370 is associated with a functionally equivalent epimerase. Interestingly, analysis of all *Mycoplasma* species shows that almost only the species containing at least one domain, M200535 matched the UDP-glucose-epimerase domains Pfam PF16363 and/or PF01370 (Fig. 4). Exceptions are *M. sp. Bg1* (not containing any glycosyltransferase domain but matching an epimerase domain) and *Mycoplasma iowae* (containing a glycosyltransferase domain different from M200535).

Discussion

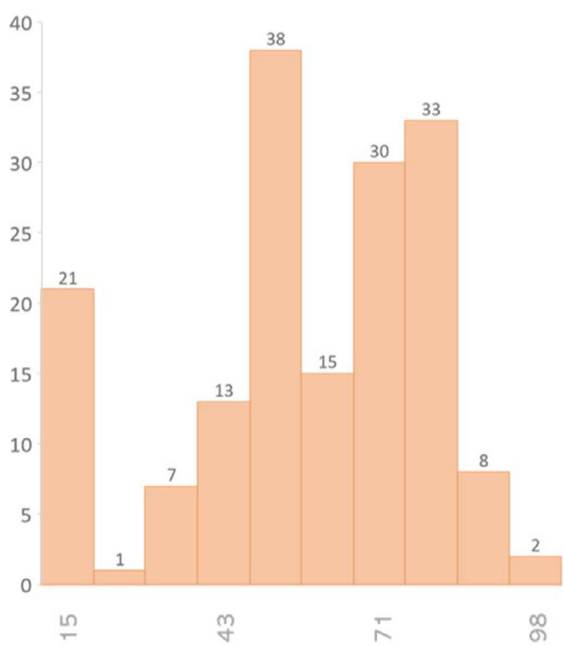
Our results show that all strains of *M. pneumoniae* analysed have at least three glycosyltransferases encoded in the genome (in *M. pneumoniae* strain M129 these are MPN_075, MPN_028 and MPN_483) that can potentially perform the synthesis of galactocerebroside. The highly significant match of M200535 with *C. jejuni* galactosyltransferases indicates that the sequences responsible of galactocerebroside synthesis contain the functional domain M200535, which, in this study, was found to be a distinctive motif with specific feature, while previously included in the Pfam domain PF00535. The *Mycoplasma* species and strains containing at least a protein with the functional domain M200535 are the only *Mycoplasma* (with the two exceptions *M. iowae* and the taxonomically not yet defined *M. sp. Bg1*) showing concomitant presence of a UDP-glucose epimerase domain, which converts UDP-glucose into UDP-galactose. UDP-galactose is used in the galactocerebroside synthesis and favoured, over UDP-glucose, as a substrate by the glycosyltransferase operating the linkage. In fact, M200535 contains conserved aspartic acid residues, which in other microorganisms such as *E. coli* have been found to be catalytically essential for glycosyltransferase to exploit their β -transferase activity on UDP-sugars (Griffiths *et al.*, 1998).

Our analysis pinpoints that almost all the *Mycoplasma* species infecting humans contain the domain M200535 (Fig. 4). Other *Mycoplasma* species without the functional domain M200535 show an absence of UDP-glucose epimerase. About half of them lack any domain associated with glycosyltransferase function; therefore, this set of species would be most likely unable to synthesize galactocerebroside that triggers GBS. Among the human-related *Mycoplasma* species, those with no glycosyltransferase domain are *M. hominis* and *M. orale*, while among respiratory tract-related species, present in

PF00535

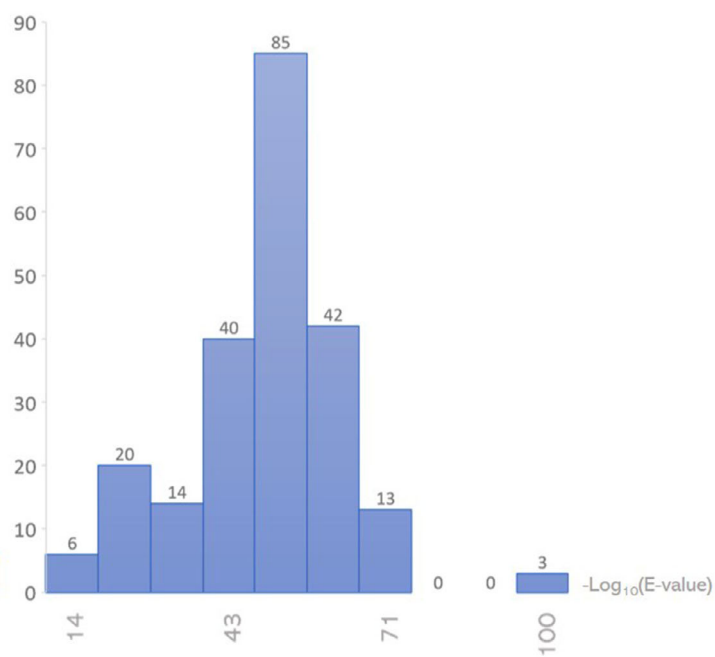


M100535



N = 178
M = 341

M200535



N = 227
M = 349

Fig. 1. Top centre) Distribution of number of expected hits of all the *Mycoplasma* protein sequences to the Pfam domain PF00535 ('Glyco_transf_2'). Expected hits are reported with correspondent $-\log_{10}(\text{E-value})$. The histogram shows two different peaks: new HMMs were built on the two groups of sequences. The first group, $-\log_{10}(\text{E-value})$ lower or equal to 13.4, is marked in orange and contains 178 sequences and the second group, $-\log_{10}(\text{E-value})$ higher or equal to 13.6, is marked in blue and contains 227 sequences. The 4 sequences with $-\log_{10}(\text{E-value})$ equal to 13.5 (comprised in the green bar) are used in both groups. Lower) distribution of matches to the two new motifs. The two new motifs obtained are called M100535 (left) and M200535 (right) and have amino acids lengths $M = 341$ and $M = 349$ respectively.

organisms other than humans, are the calves' *M. dispar*, the swine's *M. flocculare* and *M. hyopneumoniae*, the caprine's *M. ovipneumoniae*, dog's *M. spumans* and turtle's *M. testudineum*. The rest of the species without functional domain M200535, and therefore lacking UDP-glucose epimerase, comprise *Mycoplasma* species with one or more glycosyltransferase enzymes containing domains. Glycosyltransferases with domains such as M100535 do not show conserved presence of aspartic acid residues, indicating they might be unable to use UDP-sugars as substrate and, in consequence, to perform the synthesis of galactocerebroside. Human *Mycoplasma* species with glycosyltransferases lacking the domain M200535 have not been detected, while the following non-human respiratory tract-related species have been identified with glycosyltransferase and absence of M200535: tortoise's *M. agassizii*, domestic animal's *M. arginine*, bovine *M. bovirhinis* and *M. canadense*, wild bird's *M. buteonis*, goat's *M. capricolum*, pigeon's *M. columborale*, canine *M. cynos*, pig's *M. hyorhinis*, chicken's *M. iners* and *M. synoviae*, turkey's *M. melagris*, dog's *M. molare*, mink's *M. mustelae*, seal's *M. phocidae* and *M. phocirhinis*, avian *M. pullorum* and lion's *M. simbae*.

Our work leads to the suggestion of genetic modifications that would validate the hypotheses formulated by computational analysis. The characterization of the pathway indeed suggests which genes should be a primary target for genetic modifications to avoid biosynthesis of galactocerebroside in *M. pneumoniae*. The most intuitive strategy would be the knock-out of the genes encoding for the glycosyltransferases blocking the transfer of galactose to ceramide. However, this seems to be non-trivial: in a global transposon mutagenesis inactivation experiment of *M. genitalium*, the gene encoding for the glycosyltransferase MG_517, homologous of MPN_483 in *M. pneumoniae* (Klement *et al.*, 2007), remained untouched suggesting the gene encoding for this enzyme is essential (Glass *et al.*, 2006). Although it is proven that *mg517* and *mpn483* share 77% of gene sequence similarity, the enzymatic activities slightly differ in terms of specificity (Andrés, 2011). The essentiality of *mpn483*, not excluded by gene transposon analysis (Lluch-senar *et al.*, 2015), is expected as *M. pneumoniae* uses this enzyme to perform synthesis of many other lipids in the membrane (Klement, 2007), which we

know are crucial for its survival (Gaspari *et al.*, 2020). In the same way, the gene transposon analysis conducted by Lluch-Senar *et al.* suggests *mpn075* and *mpn257*, respectively, coding for the glycosyltransferase

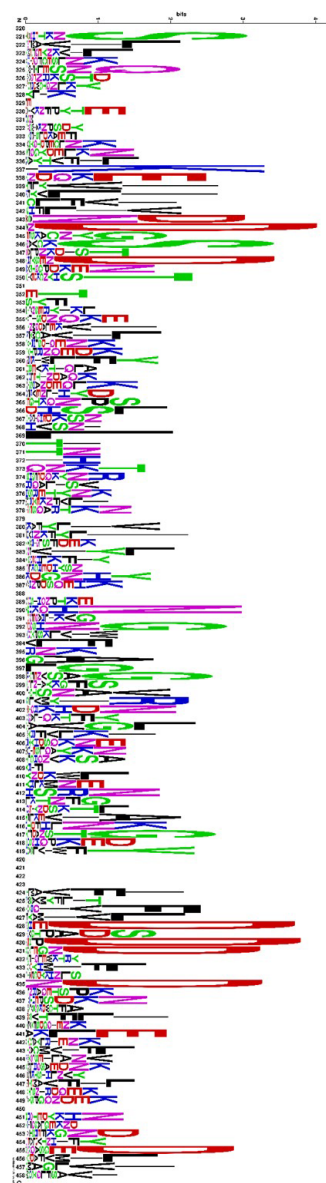


Fig. 2. Sequence logo of protein domains M200535 from sequence position 320 to 458. M200535 is abundant in aspartic acid (D) in positions 344, 348, 428, 430, 436 and 455.

Table 2. Number of glycosyltransferases found in each *Mycoplasma* species analysed for domains M100535, M200535, PF00534, PF13439 and PF13641.

Species	N. glycosyltransferases found with domain				
	M100535	M200535	PF00534	PF13439	PF13641
<i>M. agalactiae</i> ^a	1	1/2	0	0	1
<i>M. agassizii</i>	1	0	1	0	0
<i>M. alligatoris</i>	1	2	0	0	1
<i>M. alvi</i>	5	3	5	0	2
<i>M. amphoriforme</i>	0	5	1	0	0
<i>M. anatis</i>	1	1	0	0	0
<i>M. anseris</i>	1	0	0	0	0
<i>M. arginini</i>	1	0	0	0	0
<i>M. arthritidis</i>	1	0	0	0	0
<i>M. bovirhinis</i>	1	0	0	0	0
<i>M. bovirhinis</i>	1	0	0	0	0
<i>M. bovis</i>	1	1	0	0	1
<i>M. bovovulvi</i>	0	2	0	0	0
<i>M. buteonis</i>	1	0	0	0	0
<i>M. californicum</i> ^a	1/2	1	1	1	2
<i>M. canadense</i>	1	0	0	0	0
<i>M. canis</i>	1	0	0	0	0
<i>M. capricolum</i>	1	0	0	0	1
<i>M. cloacale</i>	1	0	0	0	0
<i>M. collis</i>	1	1	1	0	0
<i>M. colombinum</i>	1	0	0	0	0
<i>M. columborale</i>	1	0	0	0	0
<i>M. conjunctivae</i>	0	1	1	1	0
<i>M. cricetuli</i>	1	0	0	0	0
<i>M. crocodyli</i>	1	2	0	0	1
<i>M. cynos</i>	1	0	0	0	0
<i>M. elephantis</i> ^b	2/3	2/3	0	0	0
<i>M. falis</i>	1	0	1	1	0
<i>M. felis</i>	1	0	0	0	0
<i>M. fermentans</i>	1	1	1	1	1
<i>M. gallinarum</i>	1	1	0	0	1
<i>M. gallisepticum</i>	1	1	0	0	0
<i>M. gallopavonis</i>	1	0	0	0	0
<i>M. genitalium</i> ^a	1	2/3/4	0	0	0
<i>M. gireddii</i>	2	0	0	0	0
<i>M. glycophyllum</i>	1	0	0	0	0
<i>M. hyorhinis</i>	1	0	0	0	0
<i>M. imitans</i>	0	1	0	0	0
<i>M. iners</i>	1	0	0	0	0
<i>M. iowae</i>	2	0	0	0	1
<i>M. leachii</i>	1	0	0	0	0
<i>M. leonicaptivi</i>	1	0	0	0	0
<i>M. lipofaciens</i>	1	1	0	0	0
<i>M. melagris</i>	1	0	0	0	0
<i>M. moatsii</i>	2	6	2	0	1
<i>M. mobile</i>	1	2	1	0	0
<i>M. molare</i>	1	0	0	0	0
<i>M. mustelae</i>	1	0	0	0	0
<i>M. mycoides</i>	1	2	0	0	3
<i>M. penetrans</i>	0	3	0	0	1
<i>M. phocidae</i>	1	0	0	0	0
<i>M. phocirhinis</i>	1	0	1	1	0
<i>M. pirum</i>	2	9	4	1	1
<i>M. pneumoniae</i>	0	3	0	0	0
<i>M. primate</i>	0	3	0	0	0
<i>M. pullorum</i>	1	0	0	0	0
<i>M. pulmonis</i>	1	0	0	0	0
<i>M. putrefaciens</i>	1	0	0	0	0
<i>M. simbae</i>	1	0	1	1	0
<i>M. sp. 5H</i>	0	3	1	0	0
<i>M. sp. Bg1</i>	0	0	0	0	0
<i>M. sp. CAG:472</i>	1	1	2	3	0

Table 2. (Continued)

Species	N. glycosyltransferases found with domain				
	M100535	M200535	PF00534	PF13439	PF13641
<i>M. sp. CAG:611^a</i>	0	2/3	3	2	1
<i>M. sp. CAG:776</i>	1	9	3	2	2
<i>M. sp. CAG:877</i>	1	6	5	4	0
<i>M. sp. CAG:956</i>	3	7	2	1	1
<i>M. sp. G5847</i>	3	1	0	0	1
<i>M. sp. PE</i>	1	2	1	0	0
<i>M. sp. UBA710</i>	2	4	0	0	0
<i>M. sturni</i>	1	0	0	0	0
<i>M. subdolum</i>	1	0	0	0	0
<i>M. synoviae</i>	1	0	0	0	0
<i>M. testudinis</i>	2	7	4	0	0
<i>M. yeatsii</i>	1	0	0	0	0

a. The number of glycosyltransferases among the different strains of the same species, else the same number must be considered for all the strains of the indicated species.

b. One of the glycosyltransferase domains matches both M100535 and M200535.

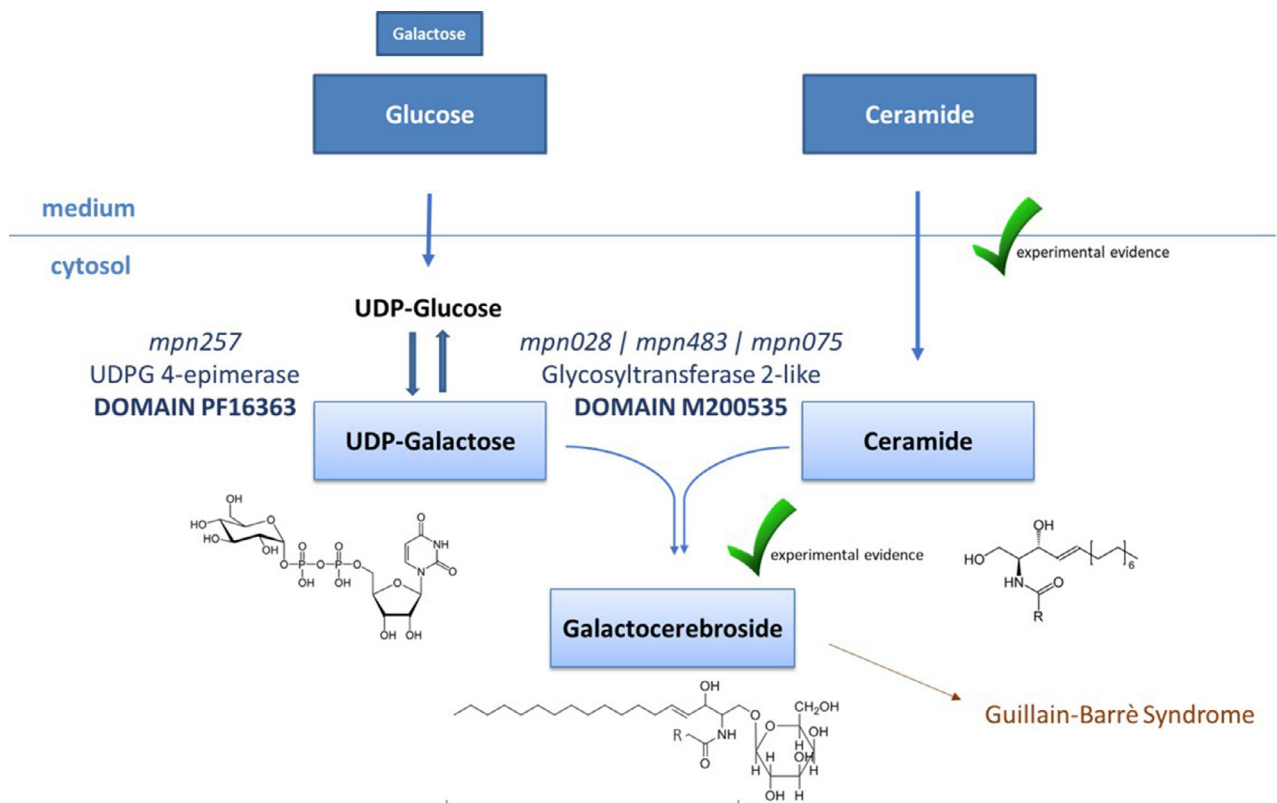


Fig. 3. In vivo galactocerebroside biosynthesis pathway in *Mycoplasma pneumoniae* according to functional analysis. *M. pneumoniae* favours the import on glucose over galactose in vivo. Once in the cytosol, UDP-glucose is converted into UDP-galactose by the epimerase containing the functional domain PF16363, encoded by gene *mpn257*. UDP-galactose is then used by one of the glycosyltransferases encoded by *mpn028*, *mpn483* or *mpn075*, all containing the functional domain M200535. The import of ceramide and the synthesis of galactocerebroside are supported by experimental evidences.

MPN_075 and the epimerase MPN_257, might be essential, at the contrary of *mpn028*, coding for the third glycosyltransferase, which is reported to be non-essential. A validation of our computational analysis would

consist in knocking out *mpn028* and replacing *mpn483* and *mpn075* with genes coding for glycosyltransferases that do not possess the motif M200535. To facilitate the genetic modification, genes coding for

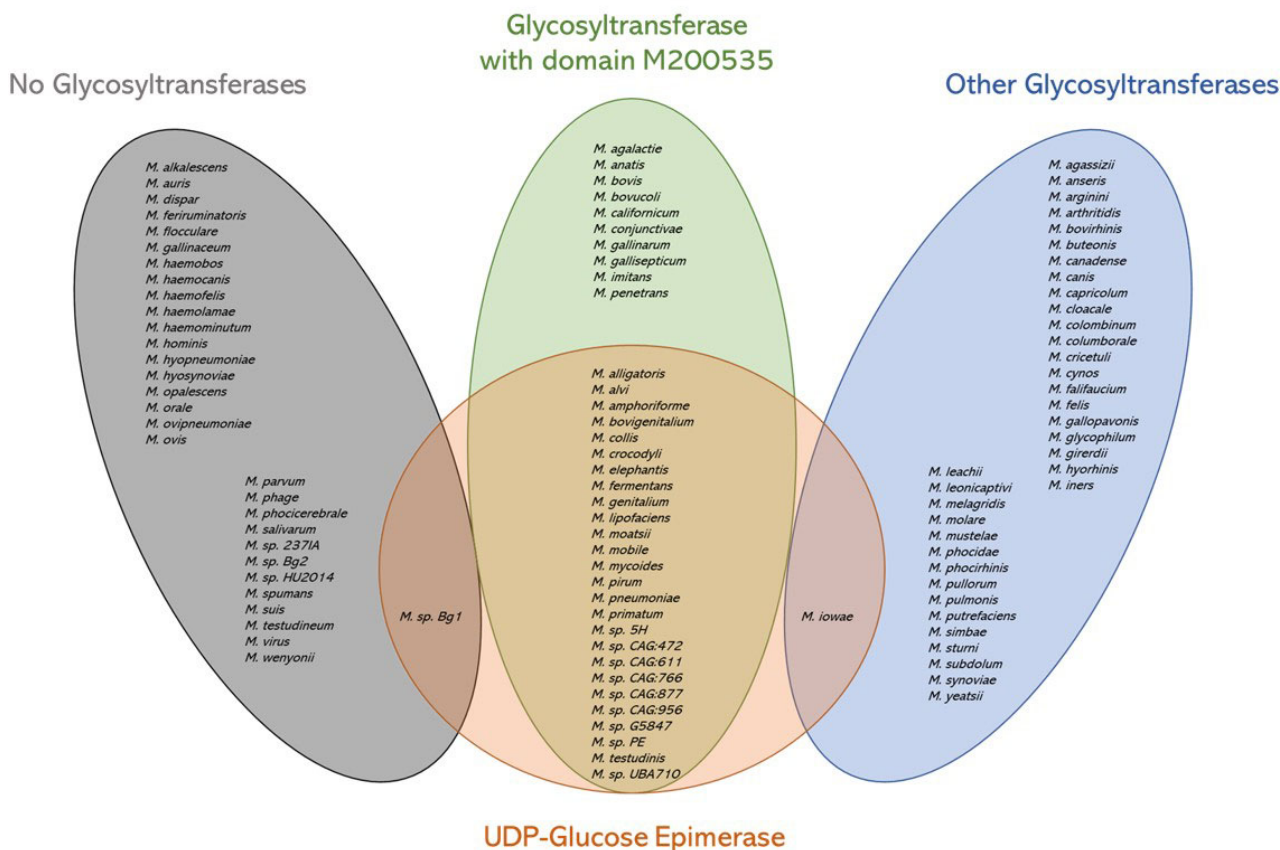


Fig. 4. Venn diagram of the three glycosyltransferase-representing groups of *mycoplasmas* (no glycosyltransferase – 31 species – in grey, at least one glycosyltransferase with domain M200535 – 36 species – in green, other glycosyltransferases with domains different from M200535 – 37 species – in blue) intersecting the group of *Mycoplasmas* containing at least one UDP-glucose epimerase (28 species – in red). Except for *M. sp. Bg1* and *M. iowae*, only species with M200535 show presence of UDP-glucose epimerase (26 species).

glycosyltransferases of other *mycoplasmas* should be chosen, among the ones not containing the motif M200535 (Fig. 4). However, the essentiality of genes *mpn075* and *mpn257* remains to be clarified, due to the uncertainty of the gene transposon analysis method in establishing gene essentiality. Moreover, the essentiality of the gene arises only after several strain passages (Lluch-Senar *et al.*, 2015). The knock-out of *mpn257*, coding for the UDP-glucose epimerase, would indeed constitute an additional layer of safety: the limited import of galactose *in vivo* is not a sufficient condition for assuming limiting quantities for biosynthesis of galactocerebroside, as the bacterium could take the needed amount of galactose from glucose conversion.

M. pneumoniae is the *Mycoplasma* species which metabolome, transcriptome and proteome are among the best-studied, hence, giving access to profound basic research opportunities and development of novel approaches in biotechnology and biomedicine (Yus *et al.*, 2009; Maier *et al.*, 2013; Chen *et al.*, 2016; Trusart *et al.*, 2017; Yus *et al.*, 2019), among which live

attenuated vaccines (www.mycosynvac.org) and respiratory tract-related live biotherapeutics (www.pulmobio.com). Therefore, an experimental application of our computational results would represent a suitable approach to render this bacterium safe in its numerous applications. In fact, other attempts to avoid biosynthesis of galactocerebroside seem to be not trivial: it would not be possible to knock-out genes involved in sphingolipid transporter since this transporter would not only import ceramide but also very important lipids such as sphingomyelin that were shown to be essential to the survival of *M. pneumoniae* (Gaspari *et al.*, 2020). Instead, the introduction of a ceramidase, disassembling ceramide into its sphingosine backbone and fatty acids chains, is another potential strategy to integrate. In bacteria, this enzyme has been found in *Pseudomonas aeruginosa* (Kita *et al.*, 2000; Okino *et al.*, 1998) and *Mycobacterium tuberculosis* (Okino *et al.*, 2010) to be neutral (Tani *et al.*, 2004) and reversible (Ito *et al.*, 2014), so not always favouring degradation of ceramide but also its synthesis, according to the environmental

and cytosolic conditions. However, *M. pneumoniae* affinity for sphingomyelin is unique, therefore, the degradation of ceramide into sphingosine and fatty acids could be of advantage for *M. pneumoniae* to build up sphingomyelin, which it typically imports unchanged from the medium (Worliczek *et al.*, 2007). The *in silico* characterization of the protein domain that might be responsible of the galactocerebroside biosynthesis has potential impact as a target for drugs related to post-infectious GBS: the elucidation of the pathway and its analysis on mycoplasmas could be used to assess the risk of post-infectious GBS development, helping in the development of therapeutical strategies for early diagnose and/or control of GBS.

Moreover, in this manuscript we report a group of mycoplasmas that lacks both UDP-glucose epimerase, converting UDP-glucose into UDP-galactose, and any glycosyltransferase domain, therefore unable to complex lipids with sugars. This group consists of *M. alkalescens*, *M. auris*, *M. dispar*, *M. feriruminatoris*, *M. flocculare*, *M. gallinaceum*, *M. haemobos*, *M. haemocanis*, *M. haemofelis*, *M. haemolamae*, *M. haemominutum*, *M. hominis*, *M. hyopneumoniae*, *M. hyosynoviae*, *M. opalescens*, *M. orale*, *M. ovipneumoniae*, *M. ovis*, *M. parvum*, *M. phage*, *M. phocicerebrale*, *M. salivarum*, *M. spumans*, *M. suis*, *M. testudineum*, *M. virus*, *M. wenyonii* and the uncharacterized *Mycoplasma* species named 2371A, Bg2 and HU2014. Lacking the whole machinery for galactocerebroside synthesis, this group of *Mycoplasma* species may be considered suitable for biomedical applications in terms of reducing the risk of post-infectious GBS.

Conclusion

All *Mycoplasma pneumoniae* strains have genes encoding for glycosyltransferases, of which at least two are essential and at least one has been proved to encode for an enzyme (MPN_483 in M129) that can synthesize glycosphingolipids such as galactocerebroside. Most likely, MPN_028 and MPN_075 perform the same synthesis, as they show high significant match for motif M200535. This motif was found as well in *C. jejuni*, the first microorganism linked to GBS through galactocerebroside biosynthesis. While the access of *M. pneumoniae* to galactose *in vivo*, when the import of glucose is at higher rate, remains unclear, the presence of a UDP-glucose epimerase MPN_257, converting UDP-glucose into UDP-galactose, could make the synthesis of galactocerebroside possible. We can conclude that all wild-type strains of *M. pneumoniae* are potentially capable of synthesizing galactocerebroside and will most likely be able to do so even with substitution or knock-out of the galactosyltransferase MPN483 or the epimerase MPN_257, which will be problematic for certain medical applications, that is human

vaccines and live biotherapeutics. Similarly, this will be the case for *Mycoplasma* species with the functional domain M200535, presenting also a UDP-glucose epimerase domain. Our data show a set of *Mycoplasma* species that could serve as alternatives for such biomedical applications or that could provide pathway genes to modify *M. pneumoniae* accordingly.

Conflict of interest

Patent application n. EP20174842.3 by the Wageningen University & Research. Author Vitor A.P. Martins dos Santos has interests in the company LifeGlimmer GmbH. All authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions

JF suggested the study. EG, JJK, JF, MS-D and VAPMdS conceived the project. JJK performed the functional domain analysis. EG developed the motif model. EG wrote the manuscript with input of MS-D, JF, VAPMdS and JJK. JF, MS-D and VAPMdS supervised the project. All authors read and approved the manuscript.

Data Availability Statement

The authors declare that all the data supporting the modelling are available within the paper and its supplementary information.

References

- Andrés, E. (2011) Expression and characterization of a mycoplasma genitalium glycosyltransferase in membrane glycolipid biosynthesis. *J Biol Chem* **286**: 35367–35379. <https://doi.org/10.1074/jbc.M110.214148>
- Ang, C.W., Tio-Gillen, A.P., Groen, J., Herbrink, P., Jacobs, B.C., Van Koningsveld, R., *et al.* (2002) Cross-reactive anti-galactocerebroside antibodies and *Mycoplasma pneumoniae* infections in Guillain-Barré syndrome. *J Neuroimmunol* **130**: 179–183. [https://doi.org/10.1016/s0165-5728\(02\)00209-6](https://doi.org/10.1016/s0165-5728(02)00209-6)
- Chen, Z., Shao, X., Dou, X., Zhang, X., Wang, Y., Zhu, C., *et al.* (2016) Role of the mycoplasma pneumoniae/interleukin-8/neutrophil axis in the pathogenesis of pneumonia. *PLoS One* **11**(1): e0146377. <https://doi.org/10.1371/journal.pone.0146377>
- Christie, L. J., Honarmand, S., Yagi, S., Ruiz, S., and Glaser, C.A.. Anti-galactocerebroside testing in *Mycoplasma pneumoniae*-associated encephalitis. *J Neuroimmunol* **189**: 129–131. <https://doi.org/10.1016/j.jneuroim.2007.06.017>
- Crooks, G. E., Hon, G., Chandonia, J., and Brenner, S. E. (2004). *WebLogo: A Sequence Logo Generator*, 1188–1190. <https://doi.org/10.1101/gr.849004.1>

- Dandekar, T., Huynen, M., Regula, J.T., Ueberle, B., Zimmermann, C.U., Andrade, M.A., *et al.* (2000) Re-annotating the *Mycoplasma pneumoniae* genome sequence: adding value, function and reading frames. *Nucleic Acids Res* **28**: 3278–3288. <https://doi.org/10.1093/nar/28.17.3278>
- Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics* **14**: 755–763. <https://doi.org/10.1093/bioinformatics/14.9.755>
- Gaspari, E., Malachowski, A., Garcia-morales, L., Burgos, R., Serrano, L., and Vitor, A.P. (2020) *Model-driven design allows growth of Mycoplasma pneumoniae on serum-free media*, **6**, pp. 1–19. <https://doi.org/10.1038/s41540-020-00153-7>
- Glass, J.I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M.R., Maruf, M., *et al.* (2006) Essential genes of a minimal bacterium. *Proc Natl Acad Sci USA* **103**: 425–430. <https://doi.org/10.1073/pnas.0510013103>
- Griffiths, G., Cook, N. J., Gottfridson, E., Lind, T., Lidholt, K., and Roberts, I. S. (1998). Characterization of the glycosyltransferase enzyme from the *Escherichia coli* K5 capsule gene cluster and identification and characterization of the glucuronyl active site. *J Biol Chem* **273**: 11752–11757. <https://doi.org/10.1074/jbc.273.19.11752>
- Güting, R. H. (1994). *GraphDB: A Data Model and Query Language for Graphs in Databases*.
- Hao, Q., Saida, T., Kuroki, S., Nishimura, M., Nukina, M., Obayashi, H., and Saida, K. (1998) Antibodies to gangliosides and galactocerebroside in patients with Guillain-Barré syndrome with preceding *Campylobacter jejuni* and other identical infections. *J Neuroimmunol* **81**: 116–126. [https://doi.org/10.1016/S0165-5728\(97\)00166-5](https://doi.org/10.1016/S0165-5728(97)00166-5)
- Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119. <https://doi.org/10.1186/1471-2105-11-119>
- Ito, M., Okino, N., and Tani, M. (2014) New insight into the structure, reaction mechanism, and biological functions of neutral ceramidase. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **1841**: 682–691. <https://doi.org/10.1016/j.bbalip.2013.09.008>
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>
- Kita, K., Okino, N., and Ito, M. (2000) Reverse hydrolysis reaction of a recombinant alkaline ceramidase of *Pseudomonas aeruginosa*. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **1485**: 111–120. [https://doi.org/10.1016/S1388-1981\(00\)00029-9](https://doi.org/10.1016/S1388-1981(00)00029-9)
- Klement, M. (2007). *The advantages of being small: Glycosyltransferases in many dimensions and glycolipid synthesis in Mycoplasma Pneumoniae*. Institutionen för biokemi och biofysik.
- Klement, M.L.R., Öjemyl, L., Tagscherer, K.E., Widmalm, G., and Wieslander, Å. (2007) A processive lipid glycosyltransferase in the small human pathogen *Mycoplasma pneumoniae*: involvement in host immune response. *Mol Microbiol* **65**: 1444–1457. <https://doi.org/10.1111/j.1365-2958.2007.05865.x>
- Koehorst, J.J., van Dam, J.C.J., Saccenti, E., Martins dos Santos, V.A.P., Suarez-Diez, M., and Schaap, P.J. (2017) SAPP: functional genome annotation and analysis through a semantic framework using FAIR principles. *Bioinformatics* **34**: 1401–1403. <https://doi.org/10.1093/bioinformatics/btx767>
- Kusunoki, S., Shiina, M., and Kanazawa, I. (2001) Anti-GalC antibodies in GBS subsequent to mycoplasma infection: evidence of molecular mimicry. *Neurology* **57**: 736–738. <https://doi.org/10.1212/WNL.57.4.736>
- Liu, Y., Ye, X., Zhang, H., Xu, X., Li, W., Zhu, D., and Wang, M. (2009) Antimicrobial susceptibility of mycoplasma pneumoniae isolates and molecular analysis of macrolide-resistant strains from Shanghai, China. *Antimicrob Agents Chemother* **53**: 2160–2162. <https://doi.org/10.1128/AAC.01684-08>
- Lluch-senar, M., Delgado, J., Chen, W., Lloréns-rico, V., Reilly, F. J. O., Wodke, J. A. H., *et al.* (2015). *Defining a minimal cell: essentiality of small ORFs and ncRNAs in a genome-reduced bacterium*, 1–7.
- Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., *et al.* (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47**: W636–W641. <https://doi.org/10.1093/nar/gkz268>
- Maier, T., Marcos, J., Wodke, J.A.H., Paetzold, B., Liebeke, M., Gutiérrez-Gallego, R., and Serrano, L. (2013) Large-scale metabolome analysis and quantitative integration with genomics and proteomics data in *Mycoplasma pneumoniae*. *Mol Biosyst* **9**: 1743–1755. <https://doi.org/10.1039/C3MB70113A>
- Matsuda, K. (2015) A novel therapeutic strategy for mycoplasma infectious diseases. *Personalized Med Universe* **4**: 32–39. <https://doi.org/10.1016/j.pmu.2015.04.005>
- Meyer Sauter, P.M., Huizinga, R., Tio-Gillen, A.P., Roodbol, J., Hoogenboezem, T., Jacobs, E., *et al.* (2016) *Mycoplasma pneumoniae* triggering the Guillain-Barré syndrome: a case-control study. *Ann Neurol* **80**: 566–580. <https://doi.org/10.1002/ana.24755>
- Meyer Sauter, P. M., Huizinga, R., Tio-Gillen, A. P., Drenthen, J., Unger, W. W. J., Jacobs, E., van Rossum, A., and Jacobs, B. C. (2018). Intrathecal antibody responses to GalC in Guillain-Barré syndrome triggered by *Mycoplasma pneumoniae*. *J Neuroimmunol* **314**: 13–16. <https://doi.org/10.1016/j.jneuroim.2017.11.011>
- Miyachi, A., Miyazaki, A., Shingu, Y., Matsuda, K., Dohi, H., and Nishida, Y. (2009) Synthesis and absolute structures of *Mycoplasma pneumoniae* β -glyceroglycolipid antigens. *Carbohydr Res* **344**: 36–43. <https://doi.org/10.1016/j.carres.2008.09.028>
- Okino, N., Ikeda, R., and Ito, M. (2010). Expression, purification, and characterization of a recombinant neutral ceramidase from mycobacterium tuberculosis. *Bioscience, Biotechnology, and Biochemistry* **74(2)**: 316–321. <https://doi.org/10.1271/bbb.90645>
- Okino, N., Tani, M., Imayama, S., and Ito, M. (1998) Purification and characterization of a novel ceramidase from *pseudomonas aeruginosa*. *J Biol Chem* **273**: 14368–14373. <https://doi.org/10.1074/jbc.273.23.14368>
- Plackett, P., Marmion, B.P., Shaw, E.J., and Lemcke, R.M. (1969) Immunochemical analysis of mycoplasma

- pneumoniae. *Aust J Exp Biol Med Sci* **47**: 171–195. <https://doi.org/10.1038/icb.1969.19>
- Rees, J.H., Soudain, S.E., Gregson, N.A., and Hughes, R.A.C. (1995) *Campylobacter jejuni* Infection and Guillain-Barré Syndrome. *N Engl J Med* **333**: 1374–1379. <https://doi.org/10.1056/NEJM199511233332102>
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignment using Clustal Omega. *Mol Syst Biol* **7**: 539. <https://doi.org/10.1038/msb.2011.75>
- Smolders, J., Jacobs, B.C., Tio-Gillen, A.P., Nijhuis, F., and Verrips, A. (2019) *Mycoplasma pneumoniae* and antibodies against galactocerebroside in a 9-year-old boy with encephalitis. *Neuropediatrics* **50**: 54–56. <https://doi.org/10.1055/s-0038-1676054>
- Sobhanifar, S., Worrall, L.J., King, D.T., Wasney, G.A., Baumann, L., Gale, R.T., *et al.* (2016) Structure and mechanism of *Staphylococcus aureus* β -glucosyltransferase involved in methicillin resistance. *PLoS Pathog* **12**: e1006067.
- Susuki, K., Odaka, M., Mori, M., Hirata, K., and Yuki, N. (2004) Acute motor axonal neuropathy after *Mycoplasma* infection. *Neurology* **62**: 949–956. <https://doi.org/10.1212/01.WNL.0000115123.42929.FD>
- Tani, M., Okino, N., Sueyoshi, N., and Ito, M. (2004) Conserved amino acid residues in the COOH-terminal tail are indispensable for the correct folding and localization and enzyme activity of neutral ceramidase. *J Biol Chem* **279**: 29351–29358. <https://doi.org/10.1074/jbc.M404012200>
- Trussart, M., Yus, E., Martinez, S., Baù, D., Tahara, Y.O., Pengo, T., *et al.* (2017) Defined chromosome structure in the genome-reduced bacterium *Mycoplasma pneumoniae*. *Nat Commun* **8**: 14665. <https://doi.org/10.1038/ncomms14665>
- van Dam, J.C.J., Koehorst, J.J., Vik, J.O., Martins dos Santos, V.A.P., Schaap, P.J., and Suarez-Diez, M. (2019) The Empusa code generator and its application to GBOL, an extendable ontology for genome annotation. *Scientific Data* **6**: 254. <https://doi.org/10.1038/s41597-019-0263-7>
- van den Berg, B., van der Eijk, A. A., Pas, S. D., Hunter, J. G., Madden, R. G., Tio-Gillen, A. P., *et al.* (2014). Guillain-Barré syndrome associated with preceding hepatitis E virus infection. *Neurology*, **82**, 491 LP – 497. <https://doi.org/10.1212/WNL.0000000000000111>
- Varma-Basil, M., Dwivedi, S.K.D., Kumar, K., Pathak, R., Rastogi, R., Thukral, S.S., *et al.* (2009) Role of *Mycoplasma pneumoniae* infection in acute exacerbations of chronic obstructive pulmonary disease. *J Med Microbiol* **58**: 322–326. <https://doi.org/10.1099/jmm.0.003335-0>
- Waites, K.B., and Talkington, D.F. (2004) *Mycoplasma pneumoniae* and its role as a human pathogen. *Clin Microbiol Rev* **17**: 697–728. <https://doi.org/10.1128/CMR.17.4.697-728.2004>
- Waites, K.B., Xiao, L., Liu, Y., Balish, M.F., and Atkinson, T.P. (2017) *Mycoplasma pneumoniae* from the Respiratory Tract and Beyond. *Clin Microbiol Rev* **30**: 747–809. <https://doi.org/10.1128/CMR.00114-16>
- Worliczek, H.L., Kämpfer, P., Rosengarten, R., Tindall, B.J., and Busse, H.-J. (2007) Polar lipid and fatty acid profiles – Re-vitalizing old approaches as a modern tool for the classification of mycoplasmas? *Syst Appl Microbiol* **30**: 355–370. <https://doi.org/10.1016/j.syapm.2007.03.004>
- Yuki, N. (2007) Ganglioside mimicry and peripheral nerve disease. *Muscle Nerve* **35**: 691–711. <https://doi.org/10.1002/mus.20762>
- Yus, E., Lloréns-Rico, V., Martínez, S., Gallo, C., Eilers, H., Blötz, C., Stülke, J., Lluch-Senar, M., and Serrano, L. (2019) Determination of the gene regulatory network of a genome-reduced bacterium highlights alternative regulation independent of transcription factors. *Cell Systems* **9**: 143–158.e13. <https://doi.org/10.1016/j.cels.2019.07.001>
- Yus, E., Maier, T., Michalodimitrakis, K., Van Noort, V., Yamada, T., Chen, W.H., *et al.* (2009) Impact of genome reduction on bacterial metabolism and its regulation. *Science* **326**: 1263–1268. <https://doi.org/10.1126/science.1177263>
- Zhang, Y., Zhou, Y., Li, S., Yang, D., Wu, X., and Chen, Z. (2016) The clinical characteristics and predictors of refractory *Mycoplasma pneumoniae* pneumonia in children. *PLoS One* **11**: e0156465. <https://doi.org/10.1371/journal.pone.0156465>

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

File S1. List of *Mycoplasma* genome sequences used in the study, obtained from NCBI repository.

File S2. Motif M200535, suggested to be responsible for galactocerebroside biosynthesis, as Hidden Markov Model.

File S3. Result of HMM search of motif M200535 in *Campylobacter jejuni*.

File S4. Consensus sequences for motifs M100535 and M200535.