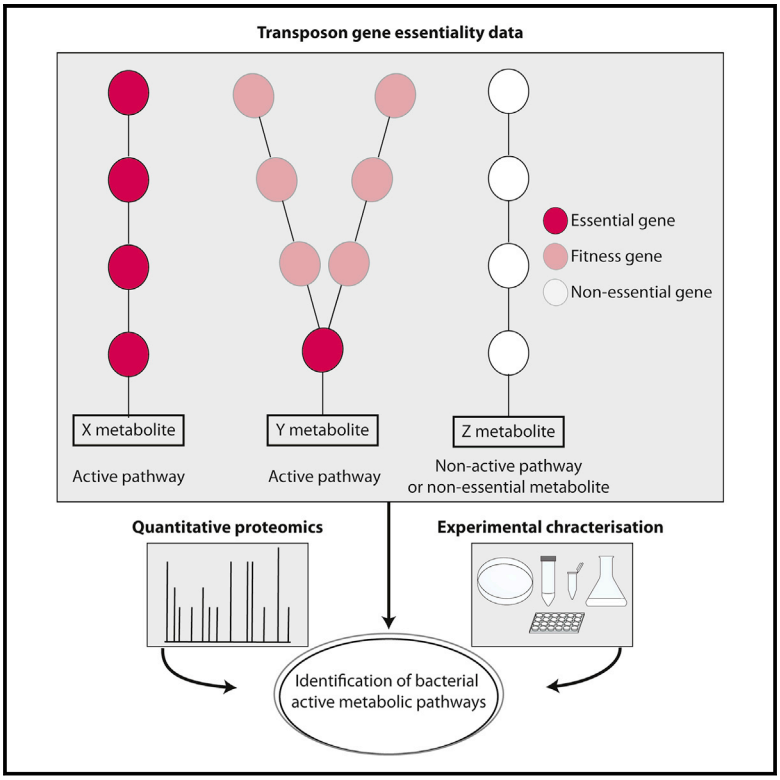
Cell Reports

Resource

Inferring Active Metabolic Pathways from Proteomics and Essentiality Data

Graphical Abstract



Highlights

- Active metabolic bacterial pathways are identified
- Integration of gene essentiality and proteomics allow prediction of active pathways
- Glucose-dependent growth is restored in *Mycoplasma agalactiae*
- Two Mycoplasma species show different usage of metabolic pathways

Authors

Ariadna Montero-Blay,
Carlos Piñero-Lambea,
Samuel Miravet-Verde,
Maria Lluch-Senar, Luis Serrano

Correspondence

luis.serrano@crg.eu

In Brief

Montero-Blay et al. identify active metabolic pathways in bacteria by integrating gene essentiality data and quantitative proteomics. Predictions agree with experimental information and show substantial differences in usage and directionality of metabolic pathways in bacteria with high degree of gene similarity.

Montero-Blay et al., 2020, Cell Reports *31*, 107722
June 2, 2020
https://doi.org/10.1016/j.celrep.2020.107722

CellPress

Cell Reports

CellPress
OPEN ACCESS

Resource

Inferring Active Metabolic Pathways from Proteomics and Essentiality Data

Ariadna Montero-Blay,[1] Carlos Piñero-Lambea,[1] Samuel Miravet-Verde,[1] Maria Lluch-Senar,[1] and Luis Serrano[1,2,3,4,*]
[1]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain
[2]Universitat Pompeu Fabra (UPF), Barcelona, Spain
[3]ICREA, Pg. Lluis Companys 23, 08010 Barcelona, Spain
[4]Lead Contact
*Correspondence: luis.serrano@crg.eu
https://doi.org/10.1016/j.celrep.2020.107722

SUMMARY

Here, we propose an approach to identify active metabolic pathways by integrating gene essentiality analysis and protein abundance. We use two bacterial species (*Mycoplasma pneumoniae* and *Mycoplasma agalactiae*) that share a high gene content similarity yet show significant metabolic differences. First, we build detailed metabolic maps of their carbon metabolism, the most striking difference being the absence of two key enzymes for glucose metabolism in *M. agalactiae*. We then determine carbon sources that allow growth in *M. agalactiae*, and we introduce glucose-dependent growth to show the functionality of its remaining glycolytic enzymes. By analyzing gene essentiality and performing quantitative proteomics, we can predict the active metabolic pathways connected to carbon metabolism and show significant differences in use and direction of key pathways despite sharing the large majority of genes. Gene essentiality combined with quantitative proteomics and metabolic maps can be used to determine activity and directionality of metabolic pathways.

INTRODUCTION

Evolution has shaped bacterial metabolism during millions of years, allowing these microorganisms to successfully colonize an enormous variety of environments, hosts, or tissues within a particular host (Rottem, 2003). Aside from its interest from an evolutionary point of view, understanding bacterial metabolism is key for a wide variety of applications. For instance, bacteria are attractive workhorses for the green production of valuable compounds, such as those related with the denim industry, that have been classically produced using environmentally hazardous processes (Hsu et al., 2018). In addition, studying the metabolism of a bacterium is essential to understand its interaction with the bacterium's host. Specifically, it has recently been shown that the microbiome metabolism can affect dramatically drug pharmacokinetics (Zimmermann et al., 2019), which therefore should be taken into account for personalized medicine purposes. Lastly, a detailed knowledge of all chemical reactions that take place in a particular strain could help to develop attenuated vaccines or bactericidal therapies based on selectively toxic metabolic intermediates.

Understanding metabolism requires being able to build an accurate metabolic map with no dead-end reactions or futile loops. This could be hindered by the fact that some metabolic enzymes can use different substrates and produce different products. For instance, fructose bisphosphate aldolase (Fba) is a key enzyme of the glycolytic pathway that catalyzes the conversion of fructose 1,6-bisphosphate (fructose-1,6BP) to dihydroxyacetone phosphate (DHAP) and glyceraldehyde phosphate (GAP). However, the same enzyme can also perform other less well-known reactions, such as conversion of fructose 1-phosphate (fructose-1P) to DHAP and glyceraldehyde (GA) or conversion of sedoheptulose 1,7-bisphosphate (sedoheptulose-1,7BP) to DHAP and erythrose 4-phosphate (erythrose-4P), among others. This catalytic versatility of metabolic enzymes could explain why, in the experimentally validated metabolic reconstruction performed in *M. pneumoniae* (Yus et al., 2009), some of the predicted reactions were not supported by any known enzymatic activity encoded by *M. pneumoniae* genome.

Here, we addressed whether it is possible to discern active and important metabolic pathways using available omics data, such as proteomics and transposon-determined gene essentiality. The basis for this hypothesis is that, in principle, pathways that produce essential metabolites for the cell must be composed of enzymes that are either essential (e.g., cannot be deleted or disrupted) or necessary for fitness (e.g., disruption triggers a growth impairment). In other words, if the pathway leading to an essential metabolite is unique, all enzymes involved should be essential; in contrast, if there is an alternative pathway with less capacity or whose use could detract from making another important metabolite, the enzymes involved should be for fitness. Enzymes involved in pathways that are not actively used under certain conditions or that generate non-essential metabolites will be non-essential, and their deletion should not compromise bacterial growth. Furthermore, we could imagine that, in general, proteins in a pathway with no branches should

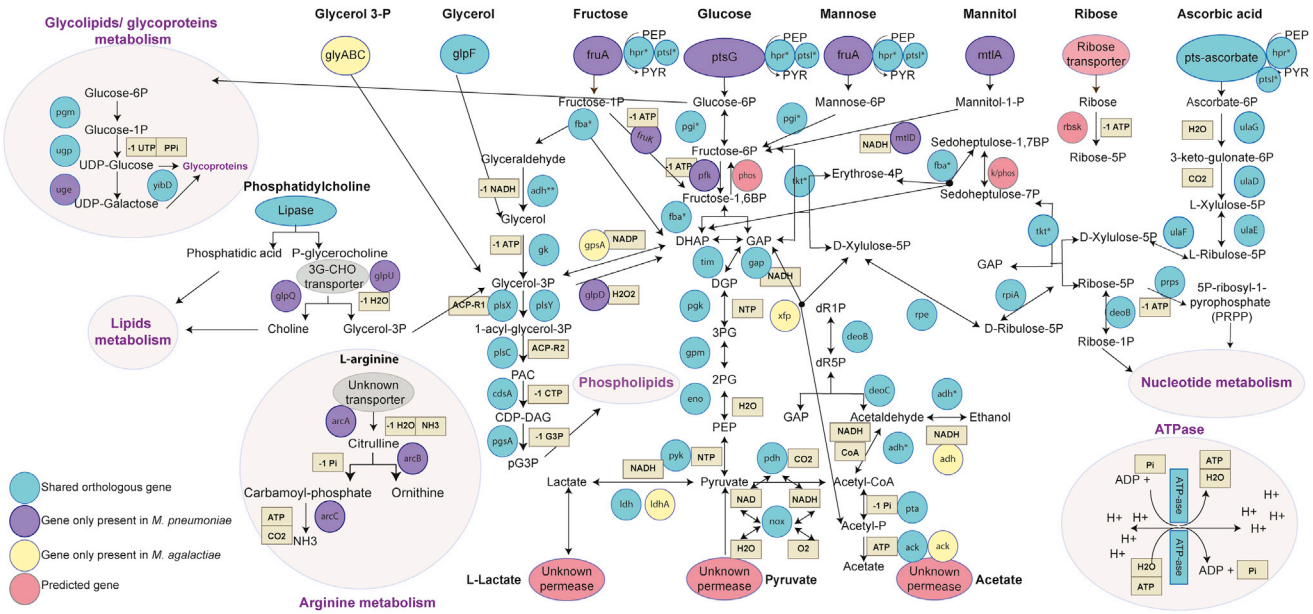


Figure 1. Genetic Context for Selected Metabolic Genes in *M. agalactiae* and *M. pneumoniae*

Panel showing genes involved in glycolipids, glycoproteins, arginine, and sugar metabolism and their link to generate DNA or RNA precursors. Blue circles indicate orthologous genes present in both species, purple circles genes exclusively present in *M. pneumoniae*, and yellow circles genes present only in *M. agalactiae*. Pink indicates assigned genes for which the gene assignation is not known. Single asterisks (*) indicate protein-coding genes participating in multiple reactions in the cell and double asterisks (**) genes whose encoded proteins are described to have multiple functionalities but for which the absence of the antecedent metabolite in one of the species analyzed here suggests that it does not perform that function in that species. Arrows connecting different metabolites indicate directionality of chemical reactions; molecules or cofactors involved in the different reactions are highlighted in squared boxes.

exhibit less variation in terms of protein abundance, although proteins upstream or downstream of a branching point will have different concentration than those in the linear pathway and therefore more variability in terms of protein abundance. Depending on the activity of the different branches, protein concentrations will be different for each branch. By combining both concepts, we potentially would be able to predict the activity and importance of metabolic active pathways. This information could be used in metabolic flux analysis or bacterial engineering. In fact, the use of transposon insertional profile analysis has been used in other bacteria for identifying specific metabolic genes in a determined condition (Ochsner et al., 2017) or as a tool to validate existing genomic-based metabolic reconstruction and flux balance analysis (Yang et al., 2014).

To see whether active metabolic pathways could be inferred from proteomics and essentiality data, we selected two bacterial species in the Mycoplasma genus: *Mycoplasma pneumoniae* and *Mycoplasma agalactiae*, which comprises species with streamlined genomes and simplified metabolic complexity. For *M. pneumoniae*, a species that infects the human lung and causes atypical pneumonia (Waites and Talkington, 2004), there is a high-resolution gene essentiality study (Lluch-Senar et al., 2015), metabolic reconstruction (Yus et al., 2009), measurement of metabolites and fluxes (Maier et al., 2013), metabolic modeling, and flux balance analysis (FBA) (Wodke et al., 2013). In the case of *M. agalactiae*, a metabolic analysis with isotope tracking using different carbon sources is available for the closely related species *Mycoplasma bovis* (Masukagami et al., 2017). *M. bovis* and *M. agalactiae* share 89.6% of their genes (Qi et al., 2012) and have a 99% sequence similarity in their 16S rRNA (Pettersson et al., 1996). Furthermore, for each of the genes involved in carbohydrates, lipids, amino acids, nucleotides, and vitamins metabolism in *M. agalactiae*, it has been found an orthologous gene in the *M. bovis* genome evidencing their metabolic analogy (Table S1). For both bacteria, there are high-resolution essentiality studies available (Lluch-Senar et al., 2015; Montero-Blay et al., 2019).

Here, we reconstructed the metabolic pathways using genome and metabolite information available. Then, we experimentally determined carbon sources that allowed robust growth and determined protein relative concentration by mass spectroscopy. Finally, we mapped gene essentiality on the metabolic maps. Comparisons of the metabolic active pathways predicted by our omics analysis with those found in the literature for both bacterial species showed an excellent agreement.

RESULTS

Generation of the Metabolic Maps

We constructed metabolic maps with all possible enzymes involved in carbon metabolism that are capable of producing ATP, as well as with the direct branches coming out of them, into nucleotides, lipids, glycolipids, and glycoproteins for *M. pneumoniae* and *M. agalactiae* (Figure 1).

For the generation of the metabolic map, we used genomics data from *M. pneumoniae* (Himmelreich et al., 1996) and *M. agalactiae* (Montero-Blay et al., 2019). This information was consistent with KEGG database (Kanehisa and Goto, 2000).

Though there are computational methods to fulfil missing reactions or existing metabolic gaps (Karp et al., 2018; Ponce-de-León et al., 2020), the simplicity of Mycoplasmas with around 700 genes allowed us to manually curate the metabolic pathways using the maps for *M. pneumoniae* of Yus et al. (2009) and Wodke et al. (2013) as references. A table for orthologous genes from *M. pneumoniae*, *M. agalactiae*, and *M. bovis* has been generated in this work (Table S1).

Comparative Genomics Reveals Fundamental Differences in *M. agalactiae* and *M. pneumoniae* Carbon Metabolism

At first glance, *M. agalactiae* and *M. pneumoniae* appear to share a common core of metabolic reactions (Figure 1, blue circles). However, a more detailed look reveals that *M. agalactiae* misses some of the pathways involved in carbon metabolism that are present in *M. pneumoniae* (Figure 1, purple circles). Likewise, there are certain enzymatic activities missing in *M. pneumoniae*, which are only encoded by *M. agalactiae* genome (Figure 1, yellow circles).

M. agalactiae lacks genes encoding transporters involved in the metabolism of glucose phosphotransferase system (PTS) (*ptsG*), fructose and mannose (*fruA*), mannitol (*mtlA*), and glycerophosphocholine (*glpU*). Accordingly with this lack of transporters, all the cytoplasmic processing enzymes whose activity is exclusively involved in the metabolism of these carbon sources are also absent in the genome of *M. agalactiae* (i.e., *fruK* for fructose, *mtlD* for mannitol, and *glpQ* for glycerophosphocholine). The only exception to this is the maintenance of cytoplasmic enzymes involved in glucose and mannose metabolism that is discussed below.

Regarding carbon metabolism, *M. agalactiae* has some unique features. Initially, as indicated above, it lacks a PTS-glucose transporter as well as phosphofructokinase (Pfk) and can therefore only enter into glycolysis from the phosphate pentose pathway (PPP) or using glycerol-3P. Second, whereas the genes encoding lactate dehydrogenase, alcohol dehydrogenase, and acetate kinase are present in a single copy in *M. pneumoniae* genome (i.e., *MPN674*, *MPN564*, and *MPN533*, respectively), these same genes are duplicated in the *M. agalactiae* genome (*MAGA7784*_RS00770/*MAGA7784_RS02675*, *MAGA7784_RS02235*A/*MAGA7784_RS02235*B, and *MAGA7784_RS00725*/*MAGA7784_RS02690*, respectively). Third, *M. agalactiae* contains unique genes not present in *M. pneumoniae* (*xfp* and *gpsA*). The enzyme D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase (Xfp) reversibly catalyzes the conversion of D-xylulose to GAP and acetyl-P. This reaction, which is not present in *M. pneumoniae*, connects the PPP to glycolysis and has been detected in a vast range of organisms, from bacteria to yeast (BRENDA: EC4.1.2.9). The GpsA enzyme converts reversibly glycerol-3P into DHAP, using NADPH as a reducing cofactor. In *M. pneumoniae*, the reaction connecting glycerol metabolism and glycolysis is irreversible and is catalyzed by GlpD, producing H_2O_2.

For arginine metabolism, there are genomic but no functional differences between the species: *M. agalactiae* does not carry any gene involved in arginine metabolism, although *M. pneumoniae* contains three genes (i.e., *arcA*, *arcB*, and *arcC*) whose protein products could metabolize arginine (to obtain one molecule of ATP). However, whereas this pathway is fully active in other *Mycoplasma* species (such as *Mycoplasma fermentans*), *M. pneumoniae* has lost its ability to metabolize arginine, as indicated by the presence of severe truncations in the genes encoding enzymes of this pathway (Rechnitzer et al., 2013).

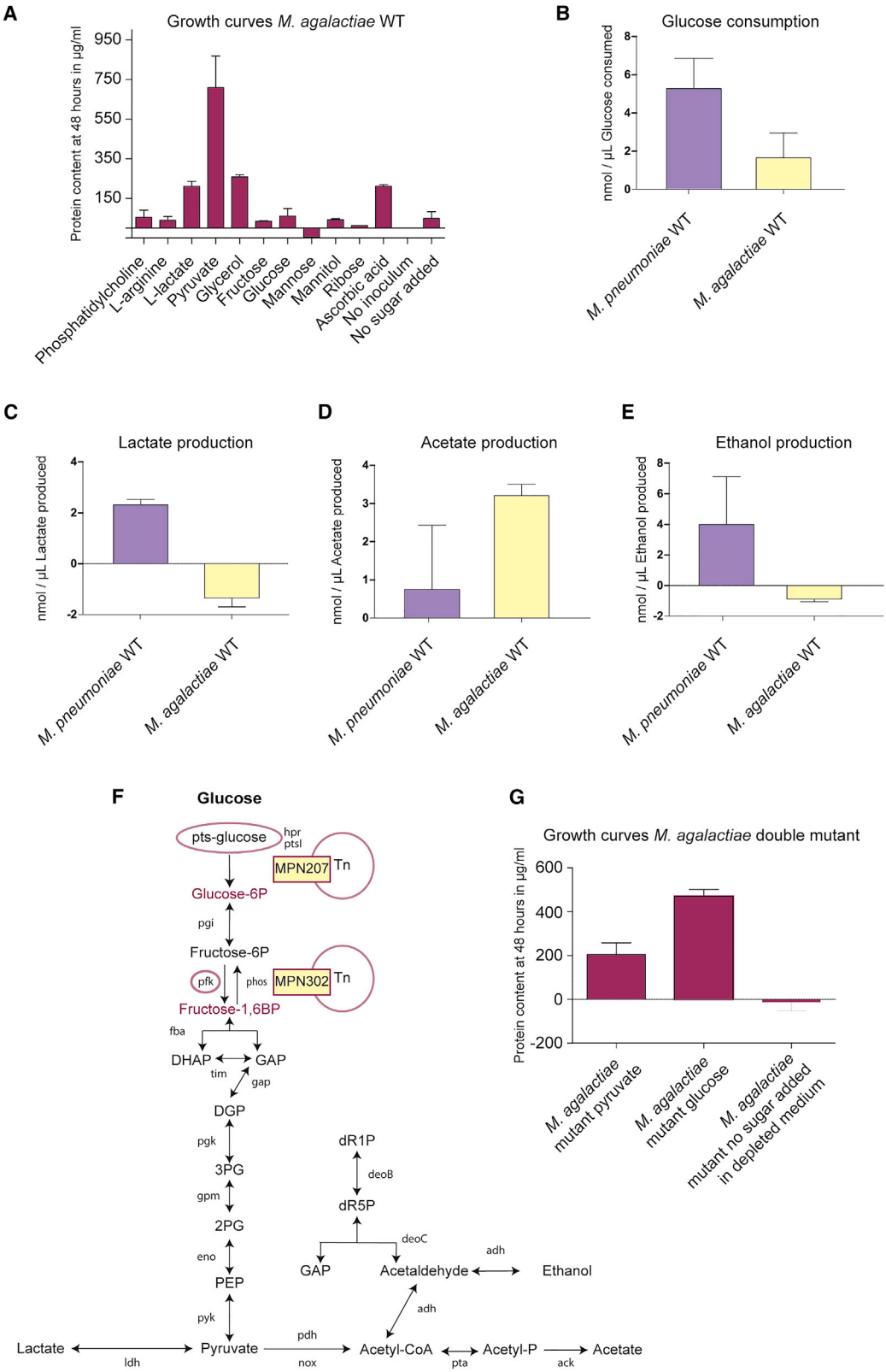
For glycolipid metabolism, both species contain a functional pathway to produce uridin diphosphate (UDP)-glucose from glucose-6P. However, the conversion of UDP-glucose to UDP-galactose is only enzymatically supported in *M. pneumoniae*, but not in *M. agalactiae*. Both metabolites can be used to generate glycoproteins as well. In line with this, it has been experimentally demonstrated that the levels of UDP-glucose are much higher in the *M. agalactiae* counterpart, *M. bovis*, than in the *M. pneumoniae*–related species *M. gallisepticum* (Masukagami et al., 2017).

Regarding phospholipid production, both organisms share the same set of genes. However, *M. pneumoniae* can only import glycerol through a permease (i.e., GlpF), as the predicted ABC transporter for glycerol-3P encoded by its genome has been shown to carry out other functions (Großhennig et al., 2013). In contrast, aside from the GlpF permease, *M. agalactiae* genome also encodes a dedicated ABC transporter for glycerol-3P that, in contrast with the one predicted for *M. pneumoniae*, shares a high homology with the validated glycerol-3P uptake system from *Escherichia coli*.

Filling Gaps in the Metabolic Map

There are genes present in both species (*fba* and *pgi*) whose maintenance in *M. agalactiae* is difficult to explain, given the absence of key enzymes in the import and metabolism of glucose. Fba catalyzes the reversible split of fructose-1,6BP into DHAP and GAP as part of the glycolytic pathway present in *M. pneumoniae*, and its substrate (i.e., fructose-1,6BP) is produced by Pfk, which is absent in *M. agalactiae*. However, as mentioned in the introduction, Fba can also reversibly catalyze the synthesis of sedoheptulose-1,7BP from DHAP and erythrose-4P (Vanyushkina et al., 2014) and therefore participate in the PPP. This less-known enzymatic activity of Fba would explain the absence of transaldolase coding gene in both species as well as in other Mycoplasma species (Kamminga et al., 2017; Vanyushkina et al., 2014). Transaldolase is the enzymatic activity that usually connects PPP and glycolysis in most living forms. Of note, the Fba-mediated connection of PPP and glycolysis requires the presence of a sedoheptulose-7P kinase to convert this metabolite into sedoheptulose-1,7BP and/or a phosphatase that performs the opposite reaction (see below). These two metabolites have been identified in *M. pneumoniae* (sedoheptulose-7P; Maier et al., 2013) and *M. bovis* (sedoheptulose-7P and sedoheptulose-1,7BP; Masukagami et al., 2017; Table S2). The existence of both metabolites supports the presence of these two activities.

The reversible enzyme Pgi converts glucose-6P into fructose-6P. This step is isolated from the rest of glycolysis in *M. agalactiae*, as it does not have the Pfk or Pts proteins (Figure 1). On the other hand, there is experimental evidence in *M. bovis* that ^{13}C-glycerol can be converted into fructose-6P



(legend on next page)

(Masukagami et al., 2017). This suggests the existence of a phosphatase that converts fructose-1,6BP into fructose-6P. The metabolic product of this phosphatase activity would then be converted into glucose-6P, which enters into the glycolipids and glycoproteins pathway that leads to essential cell metabolites. In *E. coli*, different phosphatases (i.e., YieH, YbiV, YidA, and YaeD) could dephosphorylate fructose-1,6BP into fructose-6P. Of these, YidA is the enzyme with lowest K_m (Kuznetsova et al., 2006). *M. agalactiae* has different sugar phosphatases (e.g., *MAGA7784_RS01220*, *MAGA7784_RS00145*, and *MAGA7784_RS03930*), and one of these is related to YidA in *E. coli* (*MAGA7784_RS03930*; Figure S1). Thus, we propose that *MAGA7784_RS03930* could catalyze the conversion of fructose-1,6BP into fructose-6P.

Experimental Validation of Carbon Sources Capable of Supporting Growth in *M. agalactiae*

For *M. pneumoniae*, the carbon sources capable of promoting growth have been previously analyzed (Yus et al., 2009). This bacterium grows best when using glucose as carbon source but is also able to grow with mannose, fructose, ribose, glycerol, ascorbate (Yus et al., 2009), and phosphoglycerol choline (Großhennig et al., 2013), although to a much lesser extent. However, it cannot use pyruvate, lactate, mannitol, or arginine.

One initial obstacle for determining the carbon sources required for *M. agalactiae* was our lack of a chemically defined medium in which the compounds present could be controlled. Notably, Mycoplasmas are characterized by their reduced metabolic and biosynthetic capacities. For this reason, the broth used for their growth, termed Hayflick medium, contains animal-derived serum with unspecified composition, complemented by the addition of the preferred carbon source for each *Mycoplasma* species. The presence of serum in the broth likely explains the modest growth observed for *M. agalactiae* in plain Hayflick medium without any added carbon source (Table S3). This limited growth could mask the effect of exogenous added carbon sources. To prevent this, we generated a "depleted" version of the rich medium in which all carbohydrate likely provided by the serum have been exhausted (see STAR Methods section; Table S3). We then supplemented this depleted medium with 11 different carbon sources, including phosphatidylcholine, L-arginine, L-lactate, pyruvate, glycerol, fructose, glucose, mannose, mannitol, ribose, and ascorbic acid, and then assessed the growth of *M. agalactiae.*

No significant growth of *M. agalactiae* was observed in the depleted medium (Figure 2A). As expected based on the metabolic map (Figure 1), we observed significant growth when the depleted medium was supplemented with ascorbic acid or glycerol (Figure 2A). Supplementing with those sugars requiring metabolic enzymes or transporters absent from the *M. agalactiae* genome, such as phosphatidylcholine arginine, ribose, glucose, fructose, or mannose, did not lead to significant growth (Figure 2A). We found that lactate and especially pyruvate supported robust growth of the bacterium (Figure 2A; Table S3). This is surprising because these metabolites do not support growth in *M. pneumoniae* and suggests the presence of specific transporters for these metabolites in *M. agalactiae* and the existence of a reverse gluconeogenic flow. Only a gluconeogenic flow starting from pyruvate could provide the required precursors for phospholipids and nucleotide metabolism in a depleted medium supplemented with pyruvate.

In agreement with the metabolic maps, glucose is consumed by *M. pneumoniae*, but not by *M. agalactiae* (Figure 2B). This is in line with the observation that *M. bovis*, a species closely related to *M. agalactiae*, cannot metabolize glucose (Masukagami et al., 2017). *M. pneumoniae* produces lactate as sub-product of its metabolism (Yus et al., 2009), although *M. agalactiae* grown in medium with pyruvate seems to consume the traces of lactate present in the non-depleted version of the medium (Figure 2C). Notably, although both species produced acetate, the amount of acetate produced by *M. agalactiae* was four times that of the concentration measured for *M. pneumoniae* (3.2 nmol/μL for *M. agalactiae* versus 0.8 nmol/μL for *M. pneumoniae*; Figure 2D). Finally, it seems that some ethanol could be produced in *M. pneumoniae* (Figure 2E).

Overall, when metabolizing glucose, *M. pneumoniae* generated the metabolic sub-products of lactate, acetate, and, possibly, some ethanol (Table S4). In contrast, *M. agalactiae* grown in presence of pyruvate mainly generated acetate as sub-product, and lactate was consumed from the medium. These results indicated that, although *M. pneumoniae* used glycolysis for ATP production (2 ATP per molecule or 4 ATP per molecule of glucose, depending of type of sub-product; lactate or acetate, respectively), *M. agalactiae* converted lactate, pyruvate, and/or glycerol into acetate to generate only one ATP per molecule. It would also produce two ATPs per molecule when using ascorbate to produce acetate.

Generation of a Synthetic *M. agalactiae* Strain Capable of Metabolizing Glucose

Glycolysis is an energy conversion pathway used for many organisms to convert glucose to pyruvate. Glucose is the main carbon source used for *Mycoplasmas*, but the *M. agalactiae* and *M. bovis* species are unable to use glucose for ATP production. Strikingly, despite being unable to metabolize glucose, *M. agalactiae* and *M. bovis* still conserve eight out of the ten

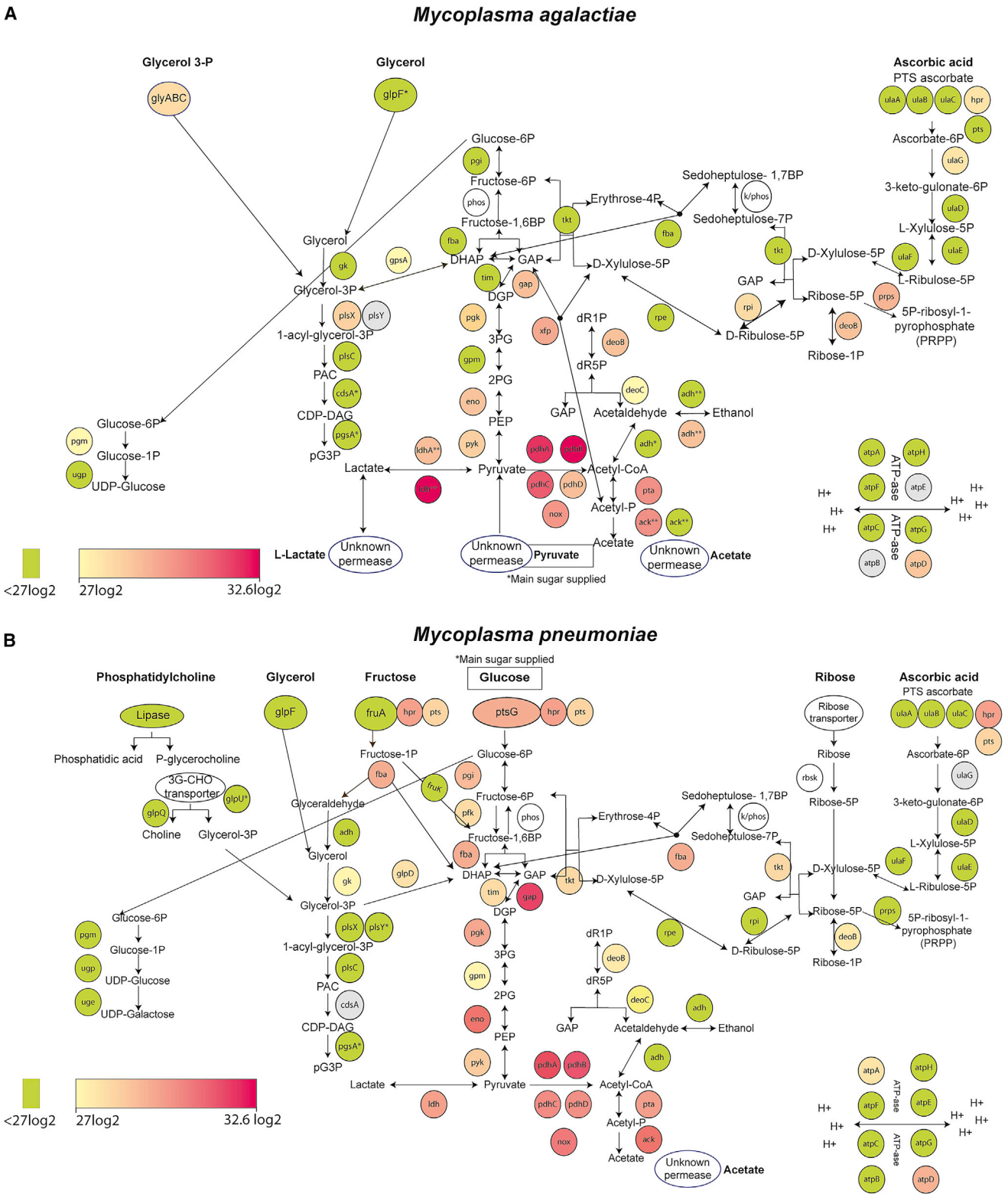
Figure 2. Growth Measurement of *M. agalactiae* under Different Culture Conditions and Secondary Metabolites Measurement for *M. agalactiae* and *M. pneumoniae*

(A) Measurement of protein content after 48 h of growth for *M. agalactiae* cultures, under various culture conditions using different carbon sources in depleted medium.

(B–E) Measurement of glucose consumption or lactate, acetate, or ethanol production in *M. pneumoniae* (left) or *M. agalactiae* (right).

(F) Reconstruction of the glycolysis pathway and pyruvate metabolism for *M. agalactiae* is indicated by black letters. Genes introduced via transposon from *M. pneumoniae*, to restore complete glycolytic functionality in *M. agalactiae*, are indicated by pink letters.

(G) Measurement of protein content after 48 h of growth for the synthetic strain of *M. agalactiae* that has restored glycolysis when using pyruvate or glucose as a carbon source. The number of replicates for the measurements represented in this figure is two. Error bars are representing the range of the distribution.



(legend on next page)

genes involved in the glycolytic pathway. We wondered whether these remaining genes were still functional in glycolysis or could be involved in other functionalities. To address this point, we decided to complement the *M. agalactiae* genome with the two genes involved in the glycolytic pathway missing in these species (coding for PTS-glucose transporter and Pfk protein). These two proteins are encoded by the genes *MPN207* and *MPN302*, respectively, in *M. pneumoniae* (Figure 2F). We cloned the genes either together or individually in a transposon under the control of a strong native regulatory region of *M. agalactiae* (see construct information in Table S5) and transformed these constructs into *M. agalactiae.* The expression of both heterologous proteins was verified by mass spectrometry (Table S6).

We assessed the ability of the engineered *M. agalactiae* strains to grow in depleted medium supplemented with glucose as the carbon source. Whereas the single mutants were unable to grow, as inferred from their inability to increase the culture biomass (Table S3), the synthetic strain carrying copies of both *MPN207* and *MPN302* grew even better than in depleted medium supplemented with pyruvate (Figure 2G). Moreover, this synthetic strain consumed glucose and produced acetate/lactate (Table S4; glucose consumption). Altogether, these results demonstrated that the glycolytic enzymes present in *M. agalactiae* have retained their functionality and could work in both glycolysis and gluconeogenesis senses.

Proteomics Abundance Quantification Is Associated with Metabolic Pathway Activity

As stated in the introduction, we wanted to find out whether integration of protein copy numbers into a metabolic map could give a clue about the magnitude of metabolic fluxes. To this end, we performed free-label mass spectrometry analyses of *M. agalactiae* and *M. pneumoniae* to obtain a quantitative proteome overview of both species (Tables S1 and S7). For visualization, we show the relative protein abundance following a color code based on the log2 normalized average area of the three top peptides of each metabolic enzyme for both species (see STAR Methods section, Figure 3, and Table S1 for orthologous gene protein comparison).

Alcohol dehydrogenase (Adh) and lactate dehydrogenase (Ldh) were among the proteins that exhibit larger differences in protein abundance between *M. agalactiae* and *M. pneumoniae*. *M. agalactiae* carries a gene duplication for Adh, and one of the isoforms is more highly expressed in this bacterium than in *M. pneumoniae* (i.e., 29.0 log2 versus 22.3 log2, respectively). Ldh also has two gene copies in *M. agalactiae* coding for two different isoforms, named Ldh and LdhA. LdhA showed a moderate expression level (i.e., 28.8 log2) and has limited amino acid identity with Ldh from *M. pneumoniae*. The Ldh proteins from both species share a high amino acid identity and therefore are likely performing the same role, but their expression levels are quite different (i.e., hereinafter referred to as those protein levels whose difference is >1 log2) in the two species (i.e., 32.6 log2 for *M. agalactiae* versus 28.7 log2 for *M. pneumoniae*; 3.89 log2 fold change with $p = 0.01$).

In addition, there are certain proteins whose relative expression levels are higher in *M. pneumoniae* than in *M. agalactiae.* For instance (and as expected), proteins involved in the glycolytic pathway present in both species (i.e., Pgi, Fba, Tim, Gap, Pgk, Gpm, Eno, and Pyk) are predominantly expressed at higher levels in *M. pneumoniae* than in *M. agalactiae*. This is especially evident for those enzymes placed in the upper part of the glycolytic pathway (i.e., Pgi, Fba, Tim, and Gap), in which the differences in their relative abundances between the two species are remarkable (i.e., 3.2 log2 fold change $p = 0.04$, 3.7 log2 fold change $p = 0.005$, 1.9 log2 fold change $p = 0.1$, and 2.1 log2 fold change $p = 0.6$, respectively). This indicates that, for *M. pneumoniae*, there is a continuous flow from glucose to pyruvate but that, for *M. agalactiae*, there is a break at the level of Tim, in agreement with its incapability to use glucose (see Discussion).

There are two other proteins related with glycolysis whose relative expression levels are higher in *M. pneumoniae* than in *M. agalactiae*: PtsI and Hpr. These proteins are involved in the internalization and subsequent phosphorylation of different carbon sources, such as ascorbate, glucose, mannose, and fructose. Whereas both Hpr and PtsI were highly expressed in *M. pneumoniae* (29.7 log2 and 28.2 log2, respectively), they were expressed at lower levels in *M. agalactiae* (27.7 log2 and 24.1 log2, respectively, and 2.0 log2 fold change $p = 0.001$; 4.1 log2 fold change $p = 0.09$ for Hpr and PtsI). This observation can be explained by the fact that *M. pneumoniae* has more carbon-source transporters that require the use of Hpr and PtsI than *M. agalactiae*. For the PPP, the expression of the transketolase (Tkt) enzyme was also higher in *M. pneumoniae* than in *M. agalactiae* (28.2 log2 versus 24.8 log2, respectively). This enzyme not only catalyzes an intermediate reaction in the PPP but is also the unique entry point to the PPP from glycolysis in *M. pneumoniae*; in contrast, *M. agalactiae* can also connect to the PPP through the more abundant Xfp enzyme. For ascorbate, both species showed lower but close to similar (hereinafter referred to as those differences ≤ 1 log2) abundances (27.8 log2 for *M. pneumoniae* and 26.7 log2 for *M. agalactiae*), suggesting that this metabolite has a subsidiary role as a carbon source.

Finally, for the metabolism of glycerol/phospholipids as well as glycolipids/glycoproteins, we see similar low levels (e.g., for GlpK enzyme 27.3 log2 and 26.5 log2 for *M. pneumoniae* or *M. agalactiae*, respectively) for the enzymes involved in both species. In summary, for *M. agalactiae*, higher expression levels are found in those enzymes involved in pyruvate and lactate metabolism, although for *M. pneumoniae*, higher expression

Figure 3. Proteomics Abundance for Selected Metabolic Pathways in *M. agalactiae* and *M. pneumoniae*

Genes whose encoded protein abundance is represented for *M. agalactiae* growing in pyruvate (A) or *M. pneumoniae* growing in glucose (B) in urea samples. The color scale represents a gradient according to protein abundance in the set of metabolic enzymes selected, showing a range of minimum (yellow) to maximum (magenta). Proteins with expression levels lower than 27 log2 are shown as min and represented in green. Double asterisks (**) indicates duplicated genes in *M. agalactiae.* White circles indicate protein-coding genes for proteins with unassigned genes and gray circles undetected proteins by mass spectrometry. A single asterisk (*) indicates proteins not detected in proteomics samples using urea as lysis buffer but rescued from SDS samples.

levels are found in general for all glycolytic enzymes and proteins involved in carbohydrate import.

As a general observation derived from the proteomics analysis, proteins belonging to a section of metabolic pathways without bifurcation points (hereinafter referred to as sub-pathways) tend to have similar expression levels. To further demonstrate this, we defined a set of linear sub-pathways for both *M. pneumoniae* and *M. agalactiae* (see STAR Methods). Subsequently, the coefficients of variation (CV) for protein abundance were calculated for each individual sub-pathway (observed) as well as for artificial sub-pathways whose members were randomly sampled from the set of metabolic genes of each strain (expected; Table S8; Figure S2). In general, most of the observed sub-pathways showed less CV than expected. However, some exceptions to this trend are observed. For instance, sub-pathway two in *M. pneumoniae* showed a CV similar to the expected if genes were randomly sampled. This might be related with the fact that genes belonging to this sub-pathway are expressed at low levels (i.e., close to the detection limit), which could impair accurate protein abundance determination. Of note, despite these exceptions, when the CV of all the sub-pathways analyzed for *M. pneumoniae* were pooled together, the mean of them was significantly lower than the one observed for artificial sub-pathways for natural quantification or $\log_2$ transformed values, respectively (Figures 4A and 4B). Moreover, a similar result was obtained in the *M. agalactiae* analysis (Figures 4C and 4D). These results support the idea that linear sub-pathways have less variation in protein abundance than expected by chance. Furthermore, this correlation in protein abundance cannot be explained by operon composition, as numerous members of the analyzed sub-pathways belong to different transcriptional units (Table S9). Therefore, similar protein expression levels in consecutive proteins might be taken as an indication of a functional pathway lacking bifurcation points. Strikingly, this correlation in protein abundance for enzymes belonging to the same sub-pathway seems to be a dynamic phenomenon. In fact, *M. pneumoniae* naturally expresses low levels of enzymes related with fructose metabolism and consequently grows poorly when this sugar is added as major carbon source. However, adaptation to this sugar (i.e., several passages with fructose as main carbon source present in the medium) results in a significant increase of the expression levels of those genes involved in fructose metabolism (Yus et al., 2009; Figure S3).

High-Resolution Transposon Density and Its Application for Flux Metabolic Directionality Ascertainment

We propose that gene essentiality could provide insights about the activity of different metabolic pathways of a cell, which together with information about protein abundance would complement the information provided by metabolomics studies. The assumption behind this is that the level of essentiality of a gene will be related to the importance of the pathway and the flux that goes through it. However, this exclusively applies for essential and fitness genes, although for non-essential genes, the flux going through it does not necessarily have to be zero.

To address this, we plotted the metabolic maps of *M. agalactiae* and *M. pneumoniae* together with a color-coded scale of transposon insertion density (see STAR Methods, Figures 5A and 5B, and Table S10). Insertion density is related to the impact that disruption of the gene has on the fitness of the bacterium (Lluch-Senar et al., 2015; Montero-Blay et al., 2019).

Our essentiality/metabolic map in *M. agalactiae* reflected an essential footprint along all the genes involved in converting pyruvate into acetate (i.e., the four subunits of the pyruvate dehydrogenase complex, NADH oxidase, phosphate acetyltransferase, and acetate kinase). Out of the two copies, the more abundantly expressed *ack* gene showed an essential profile (shown in magenta), whereas the other less-abundant isoform was clearly non-essential (shown in white). This divergence in the essentiality profile suggests that these two isoforms have different functionalities, as if they were performing the same role it would impossible to find an essential character for any of the isoforms. Conversely, the lactate-related genes *ldh* and *ldhA* were found as fitness under these conditions. This falls in line with the fact that *M. agalactiae*, in order to generate ATP, produces acetate rather than lactate as main sub-product of pyruvate metabolism. Unfortunately, the dispensability of the lactate pathway in *M. agalactiae* prevents to ascertain whether the two isoforms Ldh and LdhA have the same function. In contrast, in *M. pneumoniae*, the entire pathway from pyruvate to acetate showed a fitness footprint, as did the Ldh enzyme. This is expected, as *M. pneumoniae* glycolysis produces two ATPs before bifurcation into lactate or acetate and therefore does not need the ATP produced when going from pyruvate to acetate.

Even though *M. agalactiae* is unable to use glucose as a carbon source, the genes *pyk*, *eno*, *gpm*, *pgk*, and *gap* involved in the lower part of the glycolytic pathway showed an essential profile. This was not the case for the Fba and Pgi enzymes (discussed above). As the essentiality map was performed with pyruvate as the main carbon source, this suggests the existence of a gluconeogenic flux (i.e., reverse glycolysis) from pyruvate to GAP in this bacterium. The essential footprint extended across most of the genes (i.e., *xfp*, *rpi*, *rpe*, *deoB*, and *prps*) involved in the conversion of GAP to 5-phospho-D-ribosyl α-1-pyrophosphate (PRPP), ribose-5P, and ribose-1P, all of which are critical precursors for DNA and RNA metabolism. The *xfp* gene shows a fitness profile that can be explained, as the role of Xfp converting GAP into xylulose-5P can also be mediated by Tkt, and further agrees with the fact that we did not observe growth of *M. agalactiae* with ribose. In *M. pneumoniae*, the entry into ribose-5P and ribose-1P could be direct, as it can grow on ribose (as shown its presence in human serum in a range of 7–100 μM; see Table S11) and as indicated by the non-essential character of the *rpe* and *rpi* genes and the very low fitness for Tkl.

The whole metabolic pathway of ascorbic acid is clearly non-essential in both species. This pathway would represent a route for generation of nucleotide precursors that would not require a connection of GAP to the PPP. This suggests that the amount of ascorbic acid present in the medium in which both species are grown is rather low (see Table S11) and, consequently, that the generation of DNA and RNA precursors is performed mainly through the link of glycolysis/gluconeogenesis to the PPP in *M. agalactiae* and/or by importing ribose in *M. pneumoniae.*

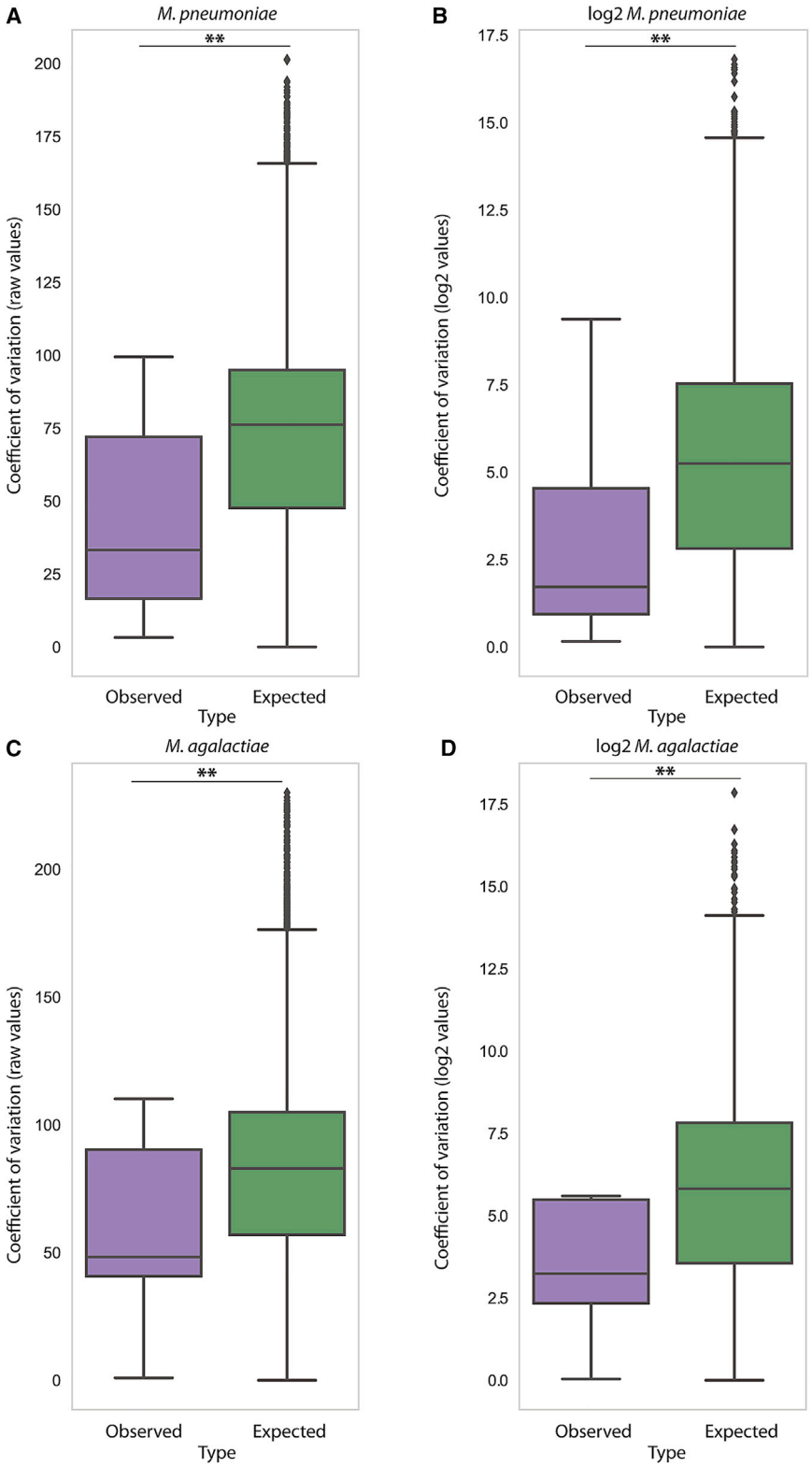
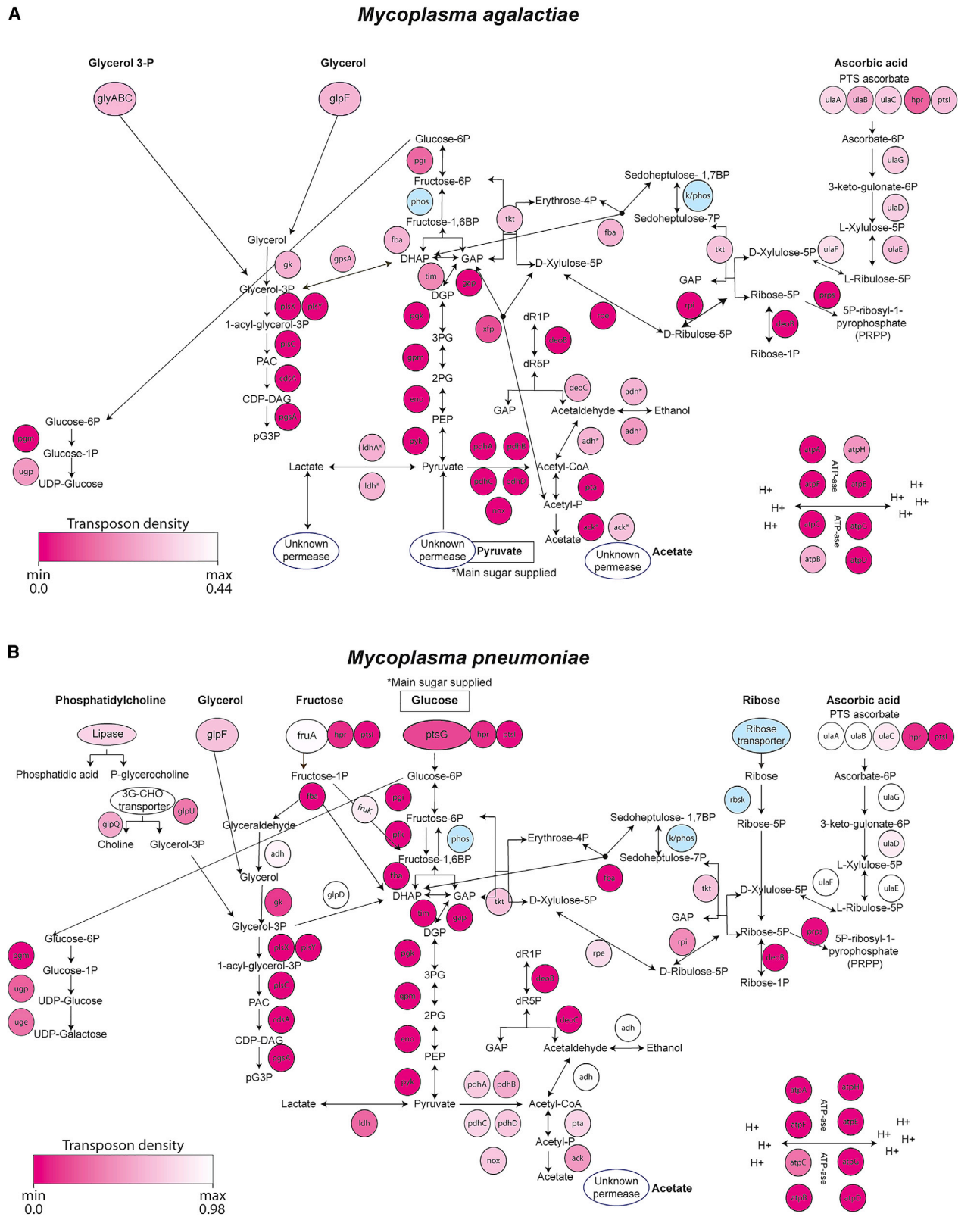


Figure 4. Study of the Protein Variability in Linear Pathways

For each of the strains, the coefficient of variation (CV) (standard deviation corrected by the mean) for linear sub-pathways was calculated and pooled together. In box plot (A and B) for *M. pneumoniae* and (C and D) for *M. agalactiae* using raw quantification values and a $\log_2$ transformation, respectively, is shown. In purple is represented the pooled CV from the set of sub-pathways and in green the pooled expected CV calculated based on artificial sub-pathways whose members were randomly sampled from the set of metabolic genes of each strain. The p-values obtained for *M. pneumoniae* are p = 0.005 and p = 0.012 for raw and log2 values, respectively. For *M. agalactiae*, the p-values obtained are for raw and log2 values, respectively, p = 0.045 and p = 0.008.



(legend on next page)

Lastly, looking into the glycerol pathway, we found that the glycerol permease (GlpF) is fitness in both *M. agalactiae* and *M. pneumoniae*. We also saw that GlpD, which links glycerol to glycolysis in *M. pneumoniae*, was not essential. GpsA, which does the same in a reversible manner, is fitness in *M. agalactiae.* In *M. pneumoniae*, we found that glucose is not converted to glycerol-3P (Maier et al., 2013). Glycerol-3P is essential for generating phospholipids and therefore it should be provided by transport and phosphorylation of glycerol and/or by hydrolysis of phosphatidylcholine present in the serum that complements the medium. These two possible ways of obtaining glycerol-3P explain the fitness character of both pathways observed in *M. pneumoniae.* In contrast, *M. agalactiae* does not have the enzymes required for production of glycerol-3P from phosphatidylcholine, and it therefore needs to import glycerol by (1) the GlpF permease, (2) glycerol-3P through the glycerol ABC transporter, or (3) obtain it from DHAP via GpsA. Again, this redundancy of pathways to obtain glycerol-3P in *M. agalactiae* explains the fitness character of the genes involved in all these pathways. In this respect, it is worth noting that a small amount of glycerol is required for growth in *M. pneumoniae*, suggesting it could have another role aside from generating glycerol-3P (Yus et al., 2009).

In summary, our essentiality maps agree with our metabolic maps and experimental analyses of carbon sources. We observed that *M. agalactiae* could be performing a reverse glycolysis to enter into the PPP to generate nucleotide precursors. *M. pneumoniae* uses glucose to produce ATP through the glycolytic pathway but also links this pathway to the PPP to generate nucleotide precursors.

DISCUSSION

The ascertainment of essential and active or non-active metabolic pathways is the starting point for performing metabolic engineering for different purposes, such as the development of vaccination strains, deletion of virulence factors, or the enhancement of biosynthetic capacities. However, even for *Mycoplasmas*, which are bacterial cells characterized by their limited biosynthetic pathways, metabolism constitutes an intricate and interconnected network of reactions.

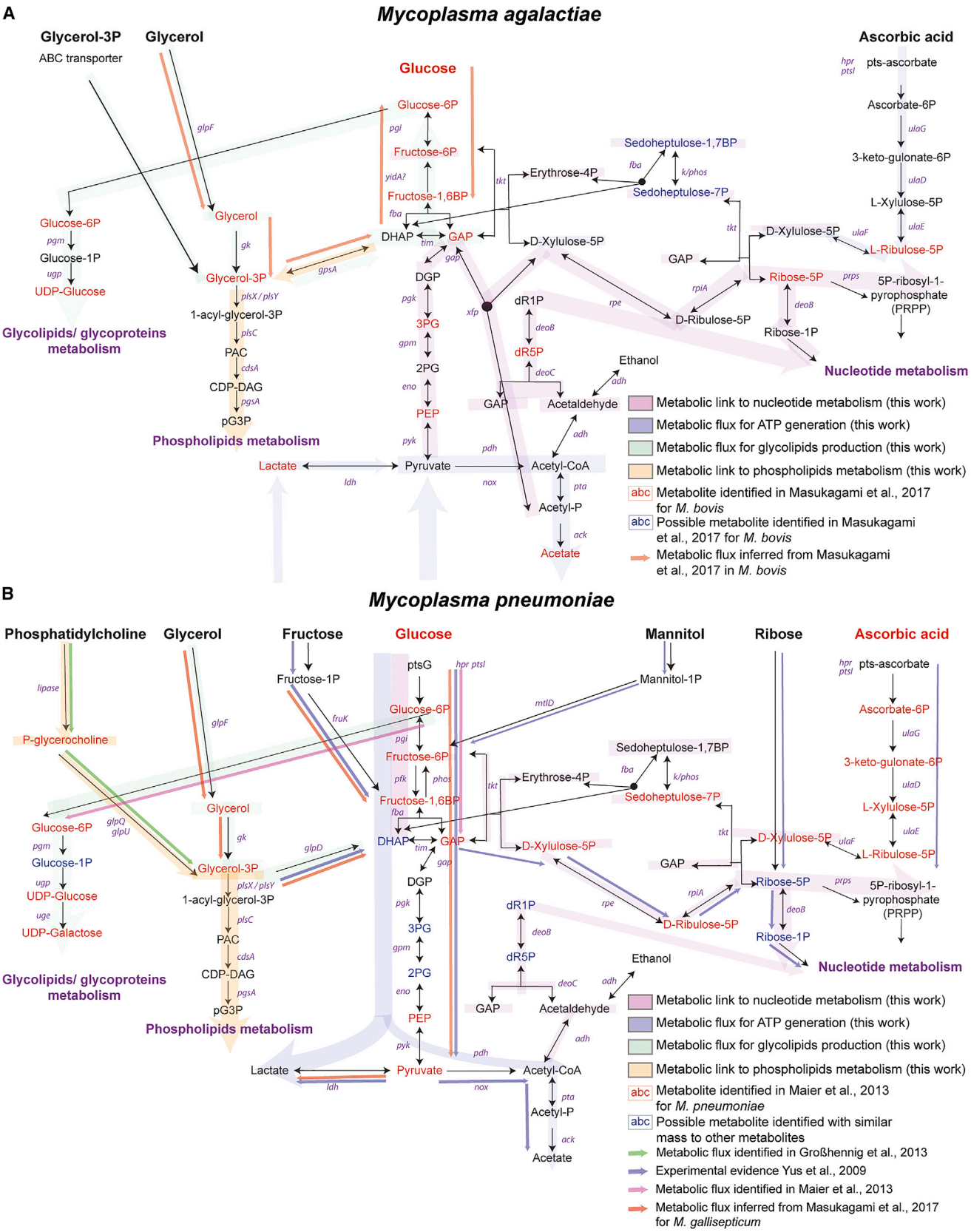
In this work, we integrated genomics, proteomics, and gene essentiality data at high resolution for *M. agalactiae* and *M. pneumoniae* as an example of how to elucidate metabolic active pathways without experimental metabolomics determination. These predictions are validated by FBA analysis in *M. pneumoniae* and experimental information collected previously for *M. pneumoniae* (and its closely related species, *M. gallisepticum*, and in the case of *M. agalactiae*, for the closely related species *M. bovis*; Masukagami et al., 2017). These two species share 100% of all enzymes involved in metabolic pathways linked to the use of carbon sources (Table S1).

First, we manually reconstituted the metabolic map regarding the use of carbon sources to produce energy and its connections to nucleotides, lipids, glycolipids, and glycoproteins. Besides postulating the existence of additional players to fulfil the logics of the pathway architecture, we also propose alternative functionalities for proteins. The proposal of specific glycolytic enzymes with multiple functionalities has been contemplated previously for *Mycoplasmas* (Pollack et al., 2002). In *M. pneumoniae*, the pyruvate kinase produces nucleosides triphosphate (NTP), thereby replacing a nucleotide diphosphate kinase. This cannot happen in *M. agalactiae* when using pyruvate as carbon source, which gives gluconeogenesis, but not glycolysis. Thus, we propose that acetate kinase genes could be involved in the generation of NTPs. It has been described that this enzyme can also use cytidine diphosphate (CDP), guanosine diphosphate (GDP), inosine diphosphate (IDP), thymidine diphosphate (TDP), or UDP (BRENDA:EC2.7.2.1). We then combined all experimental evidence for active metabolic fluxes in *M. agalactiae* (mainly coming from *M. bovis*; Masukagami et al., 2017) and in *M. pneumoniae* (Yus et al., 2009; Maier et al., 2013; Wodke et al., 2013; Figures 6A and 6B, respectively). In both cases, we have experimental data regarding (1) carbon sources that allow growth, (2) metabolites identified, and (3) ^{13}C-glucose and/or ^{13}C-glycerol fluxes. For *M. pneumoniae*, the majority of the metabolites shown in Figure 6 were detected, indicating that the different pathways shown are active (to various extents). We have experimental evidence showing a limited flux of glucose going into the PPP or the phospholipid pathway (Maier et al., 2013) and a major flux going to acetate and lactate production. However, both pathways can be used, as *M. pneumoniae* can grow not only on ribose and ascorbate but also on glycerol or phosphatidylcholine. For *M. agalactiae*, experiments with ^{13}C-glucose and ^{13}C-glycerol (from *M. bovis*; Masukagami et al., 2017) demonstrate that there is no flux from glucose to glycolysis and that glycerol is used not only for making phospholipids but also to enter into gluconeogenesis, thereby linking glucose-6P to the generation of glycolipids and glycoproteins. Furthermore, the growth of this bacterium with glycerol or ascorbate shows that, in the absence of pyruvate and lactate, it can carry out at least a set of the reactions involved in the canonical glycolysis. This is further supported by our finding that the introduction of the missing Pfk and PTS-glucose transporter enzymes enables *M. agalactiae* growth using glucose as carbon source. Thus, this bacterium could perform glycolysis or gluconeogenesis, depending on the availability of carbon sources.

We have also integrated the experimental data analyzed in this manuscript from genomics, gene transposon density, and proteomics data for *M. pneumoniae* and *M. agalactiae* to postulate active metabolic pathways for their metabolism of carbohydrates, nucleotide precursors, glycolipids, and phospholipids when grown with glucose or pyruvate, respectively, and

Figure 5. Insertional Transposon Density Landscape for *M. agalactiae* and *M. pneumoniae* for Selected Metabolic Enzymes

Sets of metabolic genes in *M. agalactiae* in presence of pyruvate (A) or in *M. pneumoniae* in presence of glucose (B) are represented in a color gradient, with gene-insertional transposon density adjusted to the specific coverage obtained for each of the strains. Minimal transposition density is indicated by magenta (showing a more essential profile), and maximum transposon density is represented in white (showing a non-essential profile). Blue circles indicate predicted genes but still unassigned. Duplicated genes for *M. agalactiae* are displayed with an asterisk (*).



(legend on next page)

compared that to the ones determined experimentally (Figures 6A and 6B for *M. agalactiae* and *M. pneumoniae*, respectively).

In *M. agalactiae*, when grown on medium with pyruvate, ATP production is mediated by the conversion of pyruvate to acetate, as indicated by the essential footprint and high abundance of all enzymes involved in this metabolic conversion. The enzyme catalyzing the conversion of lactate to pyruvate (i.e., Ldh) is not essential. However, the high expression levels of Ldh in *M. agalactiae* suggest that the production of ATP from lactate might be a common phenomenon for this species. We showed that *M. agalactiae* can grow using either ascorbate or glycerol as a carbon source. However, the enzymes involved in glycerol import and its phosphorylation or in ascorbate metabolism have a fitness profile or are non-essential, respectively. This is logical, as pyruvate was the main carbon source in our essentiality study. Glycerol, although essential for phospholipids, can be obtained by two different transport systems as well as from gluconeogenesis.

The generation of nucleotide precursors using pyruvate as carbon source in *M. agalactiae* follows a strategy based on a gluconeogenic flux (i.e., reverse glycolysis) from pyruvate to GAP. This hypothesis is supported by the essentiality footprint found for all the genes involved this metabolite conversion, despite the fact that the glycolytic pathway is truncated in this bacterium. To finally enter into the PPP and produce in this way nucleotide precursors, GAP has to be converted to xylulose-5P as a first step. In *M. agalactiae*, this conversion of GAP to xylulose-5P can be mediated by either Tkt or Xfp. Tkt needs fructose-6P and GAP to do this. However, we observed that, although this metabolite (fructose-6P) can be produced by phosphatase hydrolysis of fructose-1,6BP, the abundance of Fba was significantly lower than that of the downstream glycolytic enzymes (Figure 3A). This, together with the low essentiality profile of Fba and the quasi-essential character of the *xfp* gene for *M. agalactiae*, seems to indicate that this conversion is mainly performed by Xfp. Indeed, this assumption is further supported by the expression levels of these enzymes in *M. agalactiae*, as that for Xfp (i.e., 29.3 log2) is much higher than that for Tkt (i.e., 24.8 log2). Thus, in M. *agalactiae*, there is a flux starting at pyruvate and ending in ribose-5P that enables the synthesis of nucleotide precursors, such as PRPP.

We postulate that synthesis of glycolipids and glycoproteins might be mediated by two different fluxes in *M. agalactiae*. We propose that, in *M. agalactiae*, glucose-6P, the major building block for glycolipids synthesis, can be generated thanks to the existence of a complete gluconeogenic flux from pyruvate to glucose-6P. In this way, the consecutive activity of the glycolytic enzymes Pyk, Eno, Gpm, Pgk, Gap, and Fba, as well as a putative phosphatase, most likely the YidA-related phosphatase encoded by MAGA7784_RS03930 (Figure 6) would enable this conversion. All the genes coding for these enzymes show an essential profile (e.g., *pyk*, *eno*, *gpm*, *pgk*, and *gap*) or a quasi-essential profile (e.g., *pgi* and *fba*). Alternatively, the generation of glucose-6P could be mediated by a partial gluconeogenic flux starting at glycerol. However, as discussed in the previous section, *M. agalactiae* can only produce glycerol-3P from DHAP or import glycerol or glycerol-3P, but it cannot produce it from phosphatidylcholine. The fitness profile of the respective enzymes suggests that these pathways are all active. Finally, we observed that all the enzymes going from glycerol-3P to phospholipids have an essential profile in both *M. agalactiae* and *M. pneumonia*e. This indicates the importance of generating glycerol-3P.

Whereas most of our predictions are based on the essentiality study done for *M. agalactiae*, the proteomics data strongly appear to corroborate our hypothesis. In fact, it is curious to observe that expression levels of all the enzymes involved in the lower part of the glycolytic pathway (i.e., Gap, Pgk, Gpm, Eno, and Pyk) are higher than the ones observed for the enzymes found in the upper part (i.e., Fba and Pgi). This observation falls in line with the fact that the lower part of the glycolytic pathway feeds both nucleotide metabolism and glycolipids metabolism, whereas the enzymes in the upper part are intended to exclusively feed glycolipid metabolism. In contrast, *M. pneumoniae* shows a different metabolic strategy. In this species, ATP generation is mediated through glycolysis, as indicated by the protein abundance and having an essential footprint present of all the enzymes in the whole pathway from glucose to pyruvate. Through this set of reactions, two molecules of ATP are obtained, which can be increased to four if pyruvate is further metabolized to acetate rather than to lactate. Alternatively, under glucose depletion conditions, *M. pneumoniae* can also use alternative carbon sources, such as fructose or mannose, which enter glycolysis at different point-load levels.

Regarding the PPP, *M. agalactiae* can link glycolysis and the PPP using the activity of either Tkt or Xfp, but the genome of *M. pneumoniae* only encodes for Tkt. Intriguingly, however, the *tkt* gene shows a low fitness character in both species, which makes sense for *M. agalactiae*, given the alternative route based on Xfp, but not for *M. pneumoniae*, which lacks the *xfp* gene. The

Figure 6. Integration of Metabolic Information Collected for *M. agalactiae* and *M. pneumoniae* to Date with the Metabolic Information Observed in This Work

(A) For *M. agalactiae*, pink indicated integration of metabolic active pathways predicted in this work for nucleotide precursors, blue metabolic pathways for obtaining energy, green metabolic pathways for generating glycolipids or glycoproteins, and yellow metabolic pathways linked to phospholipids metabolism. Red letters indicate metabolites detected for *M. bovis*, and blue letters indicate possible metabolites identified in the same work (Masukagami et al., 2017). Orange arrows indicate active metabolic fluxes inferred from metabolomics studies for *M. bovis* (Masukagami et al., 2017).

(B) Integration of metabolic information inferred from this work for *M. pneumoniae* with experimental information collected in previous works. Metabolic active pathways proposed here for nucleotide metabolism are highlighted in pink, those involved in pathways for energy generation in blue, metabolic pathways for glycolipids and glycoproteins production in green, and phospholipids metabolism in yellow. Red letters indicate metabolites identified previously for *M. pneumoniae* (Maier et al., 2013) and blue letters metabolites whose masses make it difficult for precise ascertainment. Green arrows indicate metabolic pathways described for *M. pneumoniae* in Großhennig et al. (2013), pink arrows metabolic pathways predicted in Maier et al. (2013), orange arrows metabolic fluxes predicted for *M. gallisepticum* in Masukagami et al. (2017), and blue arrows active metabolic pathways ascertained from experimental validation for *M. pneumoniae* in Yus et al. (2009).

reason could be that *M. pneumoniae* can use ribose as carbon source as well as import nucleosides (Yus et al., 2009), although *M. agalactiae* does not grow with ribose. Finally, the link to glycolipids metabolism is quite obvious in *M. pneumoniae*, given its ability to directly import and phosphorylate glucose into glucose-6P, which can be later employed for glycolipids synthesis. Regarding phospholipids in *M. pneumoniae*, the enzymes downstream of glycerol-3P are essential and have as *M. agalactiae* similar expression levels, although those involved in glycerol-3P generation have a fitness profile, as the bacterium has three possible routes: from fructose-1P; from phosphatidylcholine; and by importing it directly. Essentiality analysis suggests that the fructose flux must be very low. In this case, glycolysis cannot produce glycerol-3P, as the GlpD enzyme produces DHAP from glycerol-3P in an irreversible manner.

The study of bacterial metabolism may give a hint about their lifestyle within a host. In the case of *M. pneumoniae*, a lung-colonizing bacterium, the essentiality profile for its active metabolic pathways might be different in *in vivo* conditions, where phosphatidylcholine and phosphatidylglycerol are the main carbon sources present in the lung alveolar surfactant. The fact that *M. agalactiae* or *M. bovis* have the ability to metabolize acetate or pyruvate, products of glycolysis, could suggest that they could internalize into the cytoplasm of eukaryotic cells (Bürki et al., 2015) and use pyruvate as substrate.

Summing up, integration of the metabolic pathway genomics, the experimental analysis on carbon sources supporting growth, the quantitative proteome, and the essentiality analyses allows us to predict the metabolic pathways active under the conditions studied. Notably, our predictions agree not only with the metabolic studies done in *M. pneumoniae* and *M. bovis* but also with the experimental validation of carbon sources capable to promote bacterial growth done in this study. Also, in a metabolic reconstruction performed in minimal cell JCVI-syn3A, the usefulness of gene essentiality and proteomics to help reconstructing the metabolic activity of the bacterium was also shown (Breuer et al., 2019). Thus, protein quantification together with gene essentiality under different experimental conditions could be a way to complement metabolic flux analyses that use ^{13}C-tagged compounds or metabolic modeling. To conclude, it should be noted that, although in some aspects the information provided by proteomics and gene essentiality might seem redundant, both approaches support each other and offer complementary information. Gene essentiality shows essential pathways and, when more than one pathway is connected to an essential metabolite, indicates whether they can replace each other. Protein abundance in pathways connected to a metabolite gives an indication of the preferred flow. Combination of both could offer additional insights. In *M. pneumoniae*, the expression levels of proteins converting pyruvate into lactate and acetate are comparable. Notwithstanding, the essentiality profile of those genes differs, showing *ldh* a lower density of transposon insertions and therefore pointing to lactate as main metabolism sub-product for *M. pneumoniae.* On the other hand, there are other situations in which proteomics data could clarify the information provided by gene essentiality. For instance, when there is redundancy in the pathways to produce a given metabolite, such as xylulose-5P in *M. agalactiae*, the essentiality profile could suggest which pathway is absorbing most of the flux (i.e., *xfp* is much more essential than the alternative pathway based on *fba* and *tkt* encoded activities), but this can be further corroborated with protein abundance (i.e., Xfp is detected at levels almost ten times higher than Tkt).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability Statement
  - Data and Code Availability Statement
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Experimental metabolic analysis of M. agalactiae
  - Generation of a M. agalactiae strain able to metabolize glucose
  - Secondary metabolite measurement
  - Preparing samples for mass spectrometry analysis
  - Chromatographic and mass spectrometric analyses
  - Data analysis
  - Protein abundance variability analysis
  - Gene essentiality
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.107722.

ACKNOWLEDGMENTS

The proteomics analyses were performed in the CRG/UPF Proteomics Unit, which is part of the Proteored, PRB3, and is supported by grant PT 17/0019 of the PE I+D+i 2013-2016, funded by ISCIII and ERDF. We are especially grateful to Guadalupe Espadas-Garcia and Eduard Sabidó for technical and scientific guidance for proteomics. We thank V.A. Raker for manuscript editing. This project was financed by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program, under grant agreement nos. 634942 (MycoSynVac) and 670216 (MYCOCHASSIS). We acknowledge support of the Spanish Ministry of Science and Innovation to the EMBL partnership, the Centro de Excelencia Severo Ochoa, and the CERCA Programme / Generalitat de Catalunya.

AUTHOR CONTRIBUTIONS

A.M.-B. performed the experiments. A.M.-B. and S.M.-V. performed data analysis. L.S., M.L.-S., and C.P.-L. contributed with ideas and direct supervision of the project. All authors participated in evaluating results and discussions about the project. A.M.-B., C.P.-L., and L.S. wrote the paper. A.M.-B. and S.M.-V. prepared tables. A.M.-B. prepared figures. L.S., C.P.-L., S.M.-V., and M.L.-S. reviewed the manuscript. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 7, 2019
Revised: March 4, 2020
Accepted: May 12, 2020
Published: June 2, 2020

REFERENCES

Beer, L.A., Liu, P., Ky, B., Barnhart, K.T., and Speicher, D.W. (2017). Efficient quantitative comparisons of plasma proteomes using label-free analysis with MaxQuant. Methods Mol. Biol. *1619*, 339–352.

Breuer, M., Earnest, T.M., Merryman, C., Wise, K.S., Sun, L., Lynott, M.R., Hutchison, C.A., Smith, H.O., Lapek, J.D., Gonzalez, D.J., et al. (2019). Essential metabolism for a minimal cell. eLife *8*, e36842.

Bürki, S., Frey, J., and Pilo, P. (2015). Virulence, persistence and dissemination of Mycoplasma bovis. Vet. Microbiol. *179*, 15–22.

Chiva, C., Olivella, R., Borràs, E., Espadas, G., Pastor, O., Solé, A., and Sabidó, E. (2018). QCloud: a cloud-based quality control system for mass spectrometry-based proteomics laboratories. PLoS ONE *13*, e0189209.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Großhennig, S., Schmidl, S.R., Schmeisky, G., Busse, J., and Stülke, J. (2013). Implication of glycerol and phospholipid transporters in Mycoplasma pneumoniae growth and virulence. Infect. Immun. *81*, 896–904.

Güell, M., van Noort, V., Yus, E., Chen, W.-H., Leigh-Bell, J., Michalodimitrakis, K., Yamada, T., Arumugam, M., Doerks, T., Kühner, S., et al. (2009). Transcriptome complexity in a genome-reduced bacterium. Science *326*, 1268–1271.

Himmelreich, R., Hilbert, H., Plagens, H., Pirkl, E., Li, B.C., and Herrmann, R. (1996). Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae. Nucleic Acids Res. *24*, 4420–4449.

Hsu, T.M., Welner, D.H., Russ, Z.N., Cervantes, B., Prathuri, R.L., Adams, P.D., and Dueber, J.E. (2018). Employing a biochemical protecting group for a sustainable indigo dyeing strategy. Nat. Chem. Biol. *14*, 256–261.

Hunter, J.D. (2007). Matplotlib: a 2D graphics environment. Comput. Sci. Eng. *9*, 90–95.

Junier, I., Unal, E.B., Yus, E., Lloréns-Rico, V., and Serrano, L. (2016). Insights into the mechanisms of basal coordination of transcription using a genome-reduced bacterium. Cell Syst. *2*, 391–401.

Kamminga, T., Slagman, S.-J., Bijlsma, J.J.E., Martins Dos Santos, V.A.P., Suarez-Diez, M., and Schaap, P.J. (2017). Metabolic modeling of energy balances in Mycoplasma hyopneumoniae shows that pyruvate addition increases growth rate. Biotechnol. Bioeng. *114*, 2339–2347.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30.

Karp, P.D., Weaver, D., and Latendresse, M. (2018). How accurate is automated gap filling of metabolic models? BMC Syst. Biol. *12*, 73.

Kuznetsova, E., Proudfoot, M., Gonzalez, C.F., Brown, G., Omelchenko, M.V., Borozan, I., Carmel, L., Wolf, Y.I., Mori, H., Savchenko, A.V., et al. (2006). Genome-wide analysis of substrate specificities of the Escherichia coli haloacid dehalogenase-like phosphatase family. J. Biol. Chem. *281*, 36149–36161.

Lluch-Senar, M., Delgado, J., Chen, W.-H., Lloréns-Rico, V., O'Reilly, F.J., Wodke, J.A., Unal, E.B., Yus, E., Martínez, S., Nichols, R.J., et al. (2015). Defining a minimal cell: essentiality of small ORFs and ncRNAs in a genome-reduced bacterium. Mol. Syst. Biol. *11*, 780.

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.

Mao, F., Dam, P., Chou, J., Olman, V., and Xu, Y. (2009). DOOR: a database for prokaryotic operons. Nucleic Acids Res. *37*, D459–D463.

Masukagami, Y., De Souza, D.P., Dayalan, S., Bowen, C., O'Callaghan, S., Kouremenos, K., Nijagal, B., Tull, D., Tivendale, K.A., Markham, P.F., et al. (2017). Comparative metabolomics of *Mycoplasma bovis* and *Mycoplasma gallisepticum* reveals fundamental differences in active metabolic pathways and suggests novel gene annotations. mSystems *2*, e00055-17.

Miravet-Verde, S., Ferrar, T., Espadas-García, G., Mazzolini, R., Gharrab, A., Sabido, E., Serrano, L., and Lluch-Senar, M. (2019). Unraveling the hidden universe of small proteins in bacterial genomes. Mol. Syst. Biol. *15*, e8290.

Montero-Blay, A., Miravet-Verde, S., Lluch-Senar, M., Piñero-Lambea, C., and Serrano, L. (2019). SynMyco transposon: engineering transposon vectors for efficient transformation of minimal genomes. DNA Res. *26*, 327–339.

Ochsner, A.M., Christen, M., Hemmerle, L., Peyraud, R., Christen, B., and Vorholt, J.A. (2017). Transposon sequencing uncovers an essential regulatory function of phosphoribulokinase for methylotrophy. Curr. Biol. *27*, 2579–2588.e6.

Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis *20*, 3551–3567.

Pettersson, B., Uhlén, M., and Johansson, K.E. (1996). Phylogeny of some mycoplasmas from ruminants based on 16S rRNA sequences and definition of a new cluster within the hominis group. Int. J. Syst. Bacteriol. *46*, 1093–1098.

Pollack, J.D., Myers, M.A., Dandekar, T., and Herrmann, R. (2002). Suspected utility of enzymes with multiple activities in the small genome Mycoplasma species: the replacement of the missing "household" nucleoside diphosphate kinase gene and activity by glycolytic kinases. OMICS *6*, 247–258.

Ponce-de-León, M., Apaolaza, I., Valencia, A., and Planes, F.J. (2020). On the inconsistent treatment of gene-protein-reaction rules in context-specific metabolic models. Bioinformatics *36*, 1986–1988.

Qi, J., Guo, A., Cui, P., Chen, Y., Mustafa, R., Ba, X., Hu, C., Bai, Z., Chen, X., Shi, L., and Chen, H. (2012). Comparative geno-plasticity analysis of Mycoplasma bovis HB0801 (Chinese isolate). PLoS ONE *7*, e38239.

Rechnitzer, H., Rottem, S., and Herrmann, R. (2013). Reconstitution of an active arginine deiminase pathway in Mycoplasma pneumoniae M129. Infect. Immun. *81*, 3742–3749.

Rottem, S. (2003). Interaction of mycoplasmas with host cells. Physiol. Rev. *83*, 417–432.

Seabold, S., and Perktold, J. (2010). Statsmodels: econometric and statistical modeling with Python. In Proceedings of the 9th Python in Science Conference, S. van der Walt and J. Millman, eds. (Skipper Seabold), pp. 92–96.

Vanyushkina, A.A., Fisunov, G.Y., Gorbachev, A.Y., Kamashev, D.E., and Govorun, V.M. (2014). Metabolomic analysis of three Mollicute species. PLoS ONE *9*, e89312.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al.; SciPy 1.0 Contributors (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272.

Vizcaíno, J.A., Csordas, A., Del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., et al. (2016). 2016 update of the PRIDE database and its related tools. Nucleic Acids Res. *44*, 11033.

Waites, K.B., and Talkington, D.F. (2004). *Mycoplasma pneumoniae* and its role as a human pathogen. Clin. Microbiol. Rev. *17*, 697–728.

Wiśniewski. (2016). Quantitative Evaluation of Filter Aided Sample Preparation (FASP) and Multienzyme Digestion FASP Protocols. Anal Chem. In press. https://doi.org/10.1021/acs.analchem.6b00859.

Wodke, J.A.H., Puchałka, J., Lluch-Senar, M., Marcos, J., Yus, E., Godinho, M., Gutiérrez-Gallego, R., dos Santos, V.A.P.M., Serrano, L., Klipp, E., and

Maier, T. (2013). Dissecting the energy metabolism in Mycoplasma pneumoniae through genome-scale metabolic modeling. Mol. Syst. Biol. *9*, 653.

Yang, H., Krumholz, E.W., Brutinel, E.D., Palani, N.P., Sadowsky, M.J., Odlyzko, A.M., Gralnick, J.A., and Libourel, I.G.L. (2014). Genome-scale metabolic network validation of Shewanella oneidensis using transposon insertion frequency analysis. PLoS Comput. Biol. *10*, e1003848.

Yus, E., Maier, T., Michalodimitrakis, K., van Noort, V., Yamada, T., Chen, W.H., Wodke, J.A., Güell, M., Martínez, S., Bourgeois, R., et al. (2009). Impact of genome reduction on bacterial metabolism and its regulation. Science *326*, 1263–1268.

Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R., and Goodman, A.L. (2019). Separating host and microbiome contributions to drug pharmacokinetics and toxicity. Science *363*, eaat9931.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
Mycoplasma pneumoniae M129	Colleague	Richard Herrmann
Mycoplasma agalactiae 7784	Colleague	Christine Citti
Mycoplasma agalactiae MPN302+MPN207 synthetic strain	This paper	NA
Escherichia coli 5-alpha High Efficiency	NEB	C2987P
Chemicals, Peptides, and Recombinant Proteins		
L-phosphatidylcholine	Sigma-Aldrich	Cat#61755-25G
Ethanol	Merck Millipore	Cat#108543
Glycerol	Sigma-Aldrich	Cat#G5516-1L
D-Fructose	Sigma-Aldrich	Cat#F0127-10MG
D-Glucose	Sigma-Aldrich	Cat#D8270-1KG
D-Ribose	Sigma-Aldrich	Cat#R1757-10G
D-Mannose	Sigma-Aldrich	Cat#M6020-25G
D-Mannitol	Sigma-Aldrich	Cat#M4125-100G
Sodium Pyruvate	Sigma-Aldrich	Cat#P2256-25G
L-Ascorbic Acid	Sigma-Aldrich	Cat#A0278-25G
L-Lactic Acid	Sigma-Aldrich	Cat#199257-5G
Critical Commercial Assays		
Ethanol Assay	Megazyme	K-ETOH
Glucose Colorimetric/Fluorometric Assay Kit	Biovision	Cat#K606
Lactate Colorimetric/Fluorometric Assay Kit	Biovision	Cat#K607
Acetate Colorimetric/Fluorometric Assay Kit	Biovision	Cat#K658
Pierce BCA Protein Assay Kit	Thermo Fisher Scientific	Cat#23225
Deposited Data		
Proteomics raw data for *M. pneumoniae* M129	This paper	Pride repository: PXD015800
Proteomics raw data for *M. agalactiae* 7784	This paper	Pride repository: PXD015800
Proteomics raw data for synthetic strain *M. agalactiae*	This paper	Pride repository: PXD015800
Recombinant DNA		
pMTnGm-SynMyco + EF-tu-MPN207 + EF-tu-MPN302 See Table S5	This work	Addgene ID: 145779
pMTnGm-SynMyco + EF-tu-MPN207 See Table S5	This work	N/A
pMTnGm-SynMyco + EF-tu-MPN302 See Table S5	This work	N/A
Software and Algorithms		
Proteome Discoverer Software 2.0	Thermo Fisher Scientific	https://www.thermofisher.com/order/catalog/product/OPTON-30795
Mascot Search Engine 2.6	Matrix Science	http://www.matrixscience.com/mascot_support_v2_6.html
Microbial Genomes Database Default values	MGDB	http://mbgd.genome.ad.jp/
BlastP	BlastP suite	https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome
Python 3.6	Python Software Foundation	https://www.python.org
Statsmodel Python 3.6	Seabold and Perktold, 2010	Python package

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Matplotlib Python 3.6	Hunter, 2007	Python package
RandomPackage Python 3.6	Lib/random.py	Python package
Scipy Python 3.6	Virtanen et al., 2020	Python package
DOOR database	Mao et al., 2009	http://161.117.81.224/DOOR2/

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Luis Serrano (luis.serrano@crg.eu).

Materials Availability Statement

Plasmid generated in this work has been deposited in Addgene: ID 145779.

Data and Code Availability Statement

This study generated unique proteomics datasets. The datasets generated have been deposited in Pride Repository ID PXD015800. The code supporting this work is available in the following link https://github.com/CRG-CNAG/MetabolicPathVariationCoeff

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The strain 7784 was used for wild-type *M. agalactiae* (kindly provided by Dr. Christine Citti), and the strain M129, for wild-type *M. pneumoniae (*ATTC 29342, subtype 1). Both species were grown at 37°C in standard Hayflick medium (for each adjusted 500 mL of Hayflick medium at pH 7.6 it contains 7.3 g PPLO broth, 10 μg phenol red, 11.9 g HEPES, 100 mL inactivated fetal horse serum, and 500,000 units sterile penicillin G) supplemented with 0.5% sodium pyruvate, pH 7.6 (Sigma-Aldrich) for *M. agalactiae* or 0.5% glucose, pH 7.6 (Sigma-Aldrich) for *M. pneumoniae*. When indicated, *M. agalactiae* was grown in 'depleted Hayflick medium' supplemented with the desired carbon source. *M. agalactiae* was grown in suspension (180 rpm, 37°C), while *M. pneumoniae* was grown attached to a flask without shaking at 37°C.

METHOD DETAILS

Experimental metabolic analysis of M. agalactiae

We started all experiments with a protein concentration of 2 μg/ml of *M. agalactiae* cells in the depleted medium. "Depleted medium" was generated by inoculating 2 L of complete Hayflick medium (lacking pyruvate) with *M. agalactiae*, growing at 37°C with 180 rpm orbital shaking for 48 h, centrifuging twice at 9408 g for 10 min at 4°C, discarding the pellet, and filtering the medium using Stericup®-GP Filter Units polyethersulfone membrane (0.22 μm pore size) (Millipore Express® PLUS, Z660507-12EA). Depleted medium was stored at 4°C until use, and a single batch was used for all experiments in this work when indicated. Different carbon sources were prepared in concentrated stocks and then diluted to a fixed final concentration. All components analyzed were dissolved in sterile Milli-Q water except L-α-phosphatidylcholine, which was dissolved in ethanol (Merck, 108543). The final concentration used was 0.174% for L-arginine (Sigma-Aldrich, A5006-100), $4x10^{-3}$% for L-α-phosphatidylcholine (Sigma-Aldrich, 61755-25G), 0.05% for glycerol (Sigma-Aldrich, G5516-1L), and 0.1% for D-fructose (Sigma-Aldrich, F0127-10MG), D-glucose (Sigma-Aldrich, G8270-1KG), D-ribose (Sigma-Aldrich, R1757-10G-A), D-mannose (Sigma-Aldrich, M6020-25G), D-mannitol (Sigma-Aldrich, M4125-100G), sodium pyruvate (pH adjusted to 7.8; Sigma-Aldrich, P2256-25G), L-ascorbic acid (pH adjusted to 7.8; Sigma-Aldrich, A0278-25G), and L-lactic acid (provided in solution, pH adjusted to 7.3; Sigma-Aldrich, 199257-5G,). For each culture, *M. agalactiae* stock was inoculated in 50 mL Falcon tubes (Fisher Scientific, 14-432-22) filled with 10 mL depleted medium supplemented with the different carbon sources. After 48 h of growth, cultures were centrifuged twice and washed with one ml PBS before final resuspension in 250 μL of lysis solution (4% SDS, 100 mM HEPES). Growth was measured by protein content using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, 23225) following manufacturer's protocol. For growth curves based on protein content for *M. agalactiae* WT and growth comparison of *M. agalactiae* WT in Hayflick medium versus the depleted version see Table S3.

Generation of a M. agalactiae strain able to metabolize glucose

Wild-type *M. agalactiae* was transformed via transposon with two *M. pneumoniae* genes using plasmids described in Table S5. The transposon plasmid was generated following the Gibson assembly method (Gibson et al., 2009). DNA was isolated from NEB® 5-alpha High Efficiency (C2987P). Clones were isolated using LB agar plus ampicillin (100 μg/ml) plates and confirmed by sequencing (Eurofins Genomics). Cultures of *M. agalactiae* (25 ml) were grown 48 h in Hayflick medium. After, medium was removed, and 10 mL

fresh Hayflick medium was added. After 3 h, cultures were centrifuged at 9408 g at 4°C, washed three times with chilled electroporation buffer (EB; 272 mM sucrose, 8 mM HEPES, pH 7.4), and resuspended in 300 μl chilled EB. Cells (50 μl) were then mixed with 1.5 μg DNA, incubated 20 min on ice, and electroporated in 0.1-cm electro cuvettes using a BIO-RAD Gene Pulser Xcell apparatus, using the settings of 1250 V / 25 μF / 100 Ω. Immediately after the pulse, 420 μl fresh Hayflick was added, and cells were incubated for 90 min (*M. agalactiae*) at 37°C before seeding on agar plates. Culture (100 μl) was then inoculated into 10 mL Hayflick medium with 100 μg/ml gentamicin, 0.25% glucose, and 0.25% pyruvate for a progressive metabolic adaptation. After 4 days, cultures were centrifuged at 9408 g, medium was discarded, and the cell pellet was resuspended in 10 mL Hayflick medium with 1% glucose and 100 μg/ml gentamicin and grow further. After 2 days, the culture of *M. agalactiae* strain was collected in 1 mL of Hayflick medium to generate the primary stock of cells. The expression of both heterologous proteins from *M. pneumoniae* in *M. agalactiae* have been detected by mass spectrometry (Table S6). The growth of *M. agalactiae* synthetic strain double mutant (expressing both heterologous proteins) or single mutants (expressing only single protein; or MPN207 or MPN302) has been assessed by protein content after 24 hours of growth in the depleted version of the medium supplemented with glucose (Table S3).

Secondary metabolite measurement

For each metabolite the concentration was measured at the initial point and then after 48 h of growth. *M. agalactiae* and *M. pneumoniae* cultures were grown for 48 h in standard Hayflick medium (supplemented with pyruvate or glucose, respectively). For *M. agalactiae,* the culture was centrifuged at 4°C for 10 min at 9408 g, and the supernatant was collected and placed on ice. For *M. pneumoniae* (which grows attached), supernatant was collected directly and placed on ice. Concentrations were measured for ethanol with the enzymatic kit for ethanol assay (Megazyme, K-ETOH), for glucose with the Glucose Colorimetric/Fluorometric Assay Kit (Biovision, #K606), for L-lactate with the Lactate Colorimetric/Fluorometric Assay Kit (Biovision, #K607), and for acetate with the Acetate Colorimetric Assay Kit (Biovision, #K658); all methods followed the manufacturers' instructions. Analyses of the different metabolites were performed on different days and collected in Table S4.

Preparing samples for mass spectrometry analysis

For *M. pneumoniae,* 50 μg of cells were inoculated in two T25 flask with 5 mL Hayflick medium supplemented with glucose 0.5% and grown for 48 h. Cells were then washed twice with PBS and scraped for final collection in a 1.5 mL collection tube. For *M. agalactiae*, 5 mL cultures inoculated in Falcon tubes were grown for 48 h in standard Hayflick medium supplemented with sodium pyruvate 0.5%. Cultures were then centrifuged and washed twice in PBS. For both species, pelleted cells were then resuspended ten times with a 25-gauge needle to separate individual cells and resuspended in 150 μL of urea lysis buffer (6 M urea, 0.2 M NH_4CO_3) or SDS lysis buffer (4% SDS, 100 mM HEPES). For samples resuspended in urea buffer, two biological replicates were performed for each species. The protein concentration of each lysate was measured using the Pierce BCA Protein kit following manufacturer's protocol. For urea samples, 10 μg of lysate samples were reduced with 30 nmol dithiothreitol (37°C, 60 min) and alkylated in the dark with 60 nmol iodoacetamide (25°C, 30 min). The resulting protein extract was first diluted to 2 M urea with 200 mM ammonium bicarbonate for digestion with endoproteinase LysC (1:10 LysC weight: Cell lysate weight, overnight at 37°C; Wako, cat # 129-02541) and then diluted 2-fold with 200 mM ammonium bicarbonate for trypsin digestion (1:10 w:w, 8 h at 37°C; Promega, cat # V5113). Samples with SDS lysis buffer were reduced with 90 nmol dithiothreitol (30 min at 56°C) and alkylated in the dark with 180 nmol iodoacetamide (30 min at 25°C). Following the filter-aided sample preparation (FASP) method described in Wiśniewski, 2016, samples were then digested with 3 μg LysC (Wako, cat # 129-02541) overnight at 37°C, and then with 3 μg of trypsin (Promega, cat # V5113) for eight hours at 37°C. After digestion, peptide mixes were acidified with formic acid and desalted with a MicroSpin C18 column (The Nest Group, Inc) prior to LC-MS/MS analysis.

Chromatographic and mass spectrometric analyses

Samples were analyzed using a LTQ-Orbitrap Velos Pro-mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) coupled to an EASY-nLC 1000 (Thermo Fisher Scientific (Proxeon), Odense, Denmark). Peptides were loaded onto a 2-cm Nano Trap column with an inner diameter of 100 μm packed with ^{18}C particles of 5 μm particle size (Thermo Fisher Scientific) and separated by reversed-phase chromatography using a 25-cm column with an inner diameter of 75 μm, packed with 1.9 μm ^{18}C particles (Nikkyo Technos Co., Ltd. Japan). Chromatographic gradients started at 93% buffer A, 7% buffer B with a flow rate of 250 nl/min for 5 min, and gradually increased to 65% buffer A, 35% buffer B over 120 min. After each analysis, the column was washed for 15 min with 10% buffer A, 90% buffer B. Buffer A was 0.1% formic acid in water, and buffer B, 0.1% formic acid in acetonitrile. The mass spectrometer was operated in positive ionization mode with nano spray voltage set at 2.1 kV and source temperature at 300°C. Ultramark 1621 was used for external calibration of the FT mass analyzer prior to analyses, and an internal calibration was performed using the background polysiloxane ion signal at m/z 445.1200. Acquisition was performed in data-dependent acquisition (DDA) mode, and full MS scans with 1-micro scans at a 60,000 resolution were used over a mass range of m/z 350–2000 with detection in the Orbitrap. Auto gain control (AGC) was set to $1 \cdot 10^6$, dynamic exclusion (60 s) and charge state filtering disqualifying singly charged peptides was activated. In each cycle of DDA analysis, following each survey scan, the top twenty most intense ions with multiple charged ions above a threshold ion count of 5000 were selected for fragmentation. Fragment ion spectra were produced via collision-induced dissociation (CID) at normalized collision energy of 35% and they were acquired in the ion trap mass analyzer. AGC was set to $1 \cdot 10^4$, and an isolation window of 2.0 m/z, an activation time of 10 ms, and a maximum injection time of 100 ms were used. Data

was acquired with Xcalibur software v2.2. Digested bovine serum albumin (New England Biolabs cat # P8108S) and was analyzed between each sample run to avoid sample carryover and to assure stability of the instrument. Finally QCloud (Chiva et al., 2018) was used to control instrument longitudinal performance during the project.

Data analysis

Acquired spectra were analyzed using the Proteome Discoverer software suite (v2.0, Thermo Fisher Scientific) and the Mascot search engine (v2.6, Matrix Science (Perkins et al., 1999)). Data were searched against a *M. agalactiae* strain 7784 database (76,678 entries) or *M. pneumoniae* M129 database (87,070 entries) plus a list of common contaminants and all the corresponding decoy entries (Perkins et al., 1999; Beer et al., 2017). For peptide identification, a precursor ion mass tolerance of 7 ppm was used for the MS1 level, trypsin was used as an enzyme, and up to three missed cleavage sites were allowed. The fragment ion mass tolerance was set to 0.5 Da for MS2 spectra. Oxidation of methionine and N-terminal protein acetylation were used as variable modifications, with carbamidomethylation on cysteine set as a fixed modification. False discovery rate (FDR) in peptide identification was set to a maximum of 5%. Peptide quantification data were retrieved from the ''Precursor ion area detector'' node from Proteome Discoverer (v2.0) using 2 ppm mass tolerance for the peptide extracted ion current (XIC). We then determined the area under the curve for each unique peptide and select the top three to determine the average area under the curve for each protein. The raw proteomics data were deposited in the PRIDE repository (Vizcaíno et al., 2016) with the dataset identifier PXD015800. The obtained dataset was normalized assuming equal total protein content for both species (using two biological replicates per species). Total protein content of sample 170718_TFLS_01_01 from *M. pneumoniae* was used as a reference for normalization (Tables S1 and S7). Once the area under the curve for each protein were normalized we determined the log2 of the area for each orthologous gene in the two species and we determined the difference in log2. A t test statistical analysis has been performed for each of the selected orthologous genes, two-tailed distribution and assuming unequal variances; later we corrected the obtained p values for multiple tests using Benjamini-Hochberg correction. For the specific comparison of individual orthologous genes p values were calculated assuming one-tailed distribution and assuming unequal variances. Orthology between *M. pneumoniae* M129, *M. agalactiae* 7784 and *M. bovis* PG45 was generated using the Microbial Genomes Database (MBGD) server considering default parameters and curated for duplicated using BlastP. We added up mass spectrometry areas of duplicated forms in *M. agalactiae* when the two copies were conserved respect the same protein in *M. pneumoniae* (Table S1). The average area under the curve of the three best-flying peptides was used to estimate the relative concentration of each protein. Data for both species were normalized assuming equal total protein content. As proteins highlighted with asterisk (*) in Figure 3 were not detected in the urea samples, the values obtained in SDS are shown after normalization, with an assumption of equal total protein content. The protein encoded by gene MPN637 in *M. pneumoniae* was not detected in these experiments but data was inferred from other *M. pneumoniae* proteomics datasets published in Miravet-Verde et al. (2019) with an average value of 25.75 log2. To represent protein abundance in Figure 3, all enzymes with an expression value of lower than 27 log2 were arbitrarily represented with same color code (green) to indicate ''minimum.'' Note that below this value, peptide detection becomes random, and protein concentrations cannot be accurately estimated. Thus, in Figure 3, protein expression levels range from 27 to 32.6 log2. The upper value corresponds to the maximum expression value obtained for the set of genes involved in sugar metabolism for the two species and is represented in magenta color (i.e., *ldh*, MAGA7784_RS02675).

Protein abundance variability analysis

In order to study if the protein abundance variability is reduced when considering proteins pertaining to a unique linear metabolic pathway, we extracted eight linear sub-pathways from the studied set of metabolic pathways (Table S1). Please note that for *M. agalactiae* the number of sub-pathways analyzed is seven because enzyme Pgi was included in linear sub-pathway one (genes involved in glycolipids) while in *M. pneumoniae* Pgi was grouped with Pfk, enzyme absent in *M. agalactiae*.

The selection was based on single entry and single exit genes (nodes with one entry edge and one exit edge) (Table S8). For both species separately, the coefficient of variation was calculated (CV, standard deviation corrected by the mean) and a background model of expected CVs by random sampling (1000 iterations with each of them a size equal to the number of proteins in the sub-pathway analyzed). The same process was repeated using raw quantification (natural) and using a $\log_2$ transformation (log) (Figures S2A and S2B for *M. pneumoniae* and Figures S2C and S2D for *M. agalactiae*). In parallel, we merged all sub-pathways together (independent for each species and alternatively with raw values and logarithmic values) and compared using Wilcoxon signed-rank tests the distribution of observed CVs to a distribution of average expected CVs by each linear sub-pathway (Table S8; Figure 4). P values obtained for *M. pneumoniae* for raw and log2 values are p value = 0.005 and p value = 0.012, respectively. P values obtained for *M. agalactiae* for raw and log2 values are p value = 0.045 and p value = 0.008, respectively

To unravel whether the similarity in protein abundances observed for enzymes involved in the same sub-pathway arises as a consequence of belonging the same transcriptional unit we showed operon distributions in *M. pneumoniae* from previously available literature (Güell et al., 2009; Junier et al., 2016). In the case of *M. agalactiae*, operon distribution information was collected from DOOR database, which predicts operon organization based on genomic features (Mao et al., 2009) using as reference organism the strain NC_013948.

The information of operon distribution for both strains is available in Table S9.

Gene essentiality

Transposon density values were different for both species, with a higher coverage obtained in the *M. pneumoniae* (Table S10 for the reanalyzed *M. pneumoniae* dataset) compared to *M. agalactiae* dataset (Montero-Blay et al., 2019). In both datasets, the culture medium used to perform the experiments was Hayflick supplemented with 0.5% glucose in the case of *M. pneumoniae* or 0.5% sodium pyruvate for *M. agalactiae*. The maximum values of transposon density observed for a gene in these species were 0.98 and 0.40 for *M. pneumoniae* and *M. agalactiae*, respectively. To represent these values in Figure 4, the range of transposon density values was normalized for each species independently. In Figure 4, colors are shown along a min–max scale after normalization for each of the species (from low transposon density in magenta, to high transposon density in white).

QUANTIFICATION AND STATISTICAL ANALYSIS

For the detection and analysis of proteomics samples we have used the Proteome Discoverer Software 2.0 and Mascot Search Engine 2.6 (default parameters). The exploration of orthologous genes for *M. bovis*, *M. agalactiae* and *M. pneumoniae* have been performed using software BlastP and the Microbial Genomes Database (MGDB, default parameters) as reference. The data analysis of the proteomics samples have been performed using general Python programming language and the 'statsmodel' package. For both species, the average (in log2) and standard deviation were calculated when signal was detected in both replicates. Then, a differential expression analysis was performed and fold-change was extracted. From this, a p-value extracted by two-tailed t-test for individual samples assuming unequal variances. Those p-values were corrected by Benjamini-Hochberg (FDR 5%) and -log10 of p-value was calculated, where highest values associate to the most different values (Table S1). For protein variability analysis, we used Python 3.6 and the packages 'statsmodel' for statistical analysis, 'random' for randomization of the sets, and 'matplotlib' for visualization. Script to reproduce the analysis of the variability between and within metabolic pathways using shotgun proteomics from *M. pneumoniae* and *M. agalactiae* can be accessed through this link: https://github.com/CRG-CNAG/MetabolicPathVariationCoeff . Data is presented in Table S8.