Essential Metabolism for a Minimal Cell

Marian Breuer[1], Tyler M. Earnest[1], Chuck Merryman[2], Kim S. Wise[2], Lijie Sun[2], Michaela R. Lynott[2], Clyde A. Hutchison III[2], Hamilton O. Smith[2], John D. Lapek Jr.[3], David J. Gonzalez[3], Valérie de Crécy-Lagard[4], Drago Haas[4], Andrew D. Hanson[5], Piyush Labhsetwar[1], John I. Glass[2], and Zan Luthey-Schulten[1]

[1]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801
[2]J. Craig Venter Institute, La Jolla, CA 92037
[3]Department of Pharmacology and School of Pharmacy, University of California at San Diego, La Jolla, CA 92093
[4]Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611
[5]Horticultural Sciences Department, University of Florida, Gainesville, FL 32611

December 12, 2018

Abstract

JCVI-syn3A, a robust minimal cell with a 543 kbp genome and 493 genes, provides a versatile platform to study the basics of life. Using the vast amount of experimental information available on its precursor, *Mycoplasma mycoides capri*, we assembled a near-complete metabolic network with 98 % of enzymatic reactions supported by annotation or experiment. The model agrees well with genome-scale *in vivo* transposon mutagenesis experiments, showing a Matthews correlation coefficient of 0.59. The genes in the reconstruction have a high *in vivo* essentiality or quasi-essentiality of 92 % (68 % essential), compared to 79 % *in silico* essentiality. This coherent model of the minimal metabolism in JCVI-syn3A at the same time also points toward specific open questions regarding the minimal genome of JCVI-syn3A, which still contains many genes of generic or completely unclear function. In particular, the model, its comparison to *in vivo* essentiality and proteomics data yield specific hypotheses on gene functions and metabolic capabilities; and provide suggestions for several further gene removals. In this way, the model and its accompanying data guide future investigations of the minimal cell. Finally, the identification of 30 essential genes with unclear function will motivate the search for new biological mechanisms beyond metabolism.

1 Introduction

Establishing the core requirements of cellular life is a fundamental challenge of biology. The question of the minimal set of biochemical functions necessary for a cell to grow and replicate has been studied from a number of angles for more than 20 years. It has long been suggested [1] that a model to study the basics of cellular life would be the mycoplasmas—a group of bacteria with small genomes (580–1350 kbp [2, 3]) lacking a cell wall, which evolved via extreme genome reduction from low GC content Gram-positive ancestors [4]. Mycoplasmas exist as parasites or saprotrophs and are adapted to scavenging nutrients and cellular building blocks from their niche environments, which enabled them to lose many metabolic capabilities.

The genome of the human urogenital pathogen *Mycoplasma genitalium* (580 kbp, 525 genes overall, 482 for proteins, 43 for RNAs), sequenced in 1995 [3], is the smallest genome of any autonomously replicating cell found in nature and has thus been deemed a close approximation to a minimal genome [5]. In particular, different efforts have been undertaken to establish a minimal set of genes based on the near-minimal *M. genitalium* genome. A comparison of the first two sequenced bacterial genomes (the Gram-positive *M. genitalium* [3] and the Gram-negative *Haemophilus influenzae* [6]) yielded 256 orthologous genes that were suggested to approximate a minimal set of bacterial genes [7]; a subsequent comparative study, including genomes from several free-living and endosymbiotic bacteria, proposed a minimal set of 206 genes [8]. A limitation of this approach lies in the possibility of the same function being fulfilled by non-orthologous proteins in different organisms, in which case it would not be captured by searching for orthologous genes. Transposon mutagenesis studies [9] avoid this drawback by directly probing the dispensability of individual genes in a single organism via random gene disruption, and testing for viability. In *M. genitalium*, such studies have suggested 382 out of the 482 protein-coding genes to be essential [5].

An important limitation of deriving a minimal gene set from essentiality information on individual genes lies in the fact that more than one gene can fulfill the same function, and while neither gene is essential individually, at least one of them has to be present in a functional minimal genome. Thus, while removal of either gene would be nonlethal, removing both would create a synthetic lethality. This problem can, in principle, be circumvented by sequential gene deletion starting from a given wild-type organism (as partially done for *Escherichia coli* and *Bacillus subtilis* [10, 11]), with testing for viability and growth rate after each deletion. In principle, this would not only yield the information on a minimal genome, but also would produce a living organism controlled by such a genome. However, the time and resource costs of minimizing a genome by serial deletion of dispensable genes are prohibitive.

In 2016, we developed a successful bottom-up approach to design a minimal genome and create a living cell controlled by it [12]. Starting with the gene sequence from the 1,079 kbp genome of the ruminant pathogen *Mycoplasma mycoides capri* serovar LC GM12, a minimal genome of 531 kbp was designed and constructed containing 473 genes (438 protein-coding genes and 35 genes for RNAs) [12]. The resulting strain, JCVI-syn3.0 (NCBI GenBank: CP014940.1 [13]), has a genome smaller than that of any independently-replicating cell found in nature and is considered to be our "working approximation to a minimal cell". This achievement was the culmination of a series of breakthroughs in synthetic biology. In 2007, the successful transplantation of an *M. mycoides capri* LC GM12 genome into a *Mycoplasma capricolum* recipient cell was reported [14], transforming the recipient cell to the species of the implanted DNA. In 2008, the complete genome of *M. genitalium* was synthesized from scratch, starting with chemically synthesized oligonucleotides and stepwise recombination *in vitro* and subsequently in *Saccharomyces cerevisiae* (yeast), using the available genetic manipulation tools to assemble the genome as a plasmid inside the yeast cell [15]. These methods enabled the construction of JCVI-syn1.0, the first cell controlled by a synthetic genome (NCBI GenBank: CP002027.1) [16, 17]. This was accomplished by synthesizing of a copy of the *M. mycoides capri* LC GM12 genome along with vector sequences that allowed cloning in *E. coli* and yeast, and its subsequent transplantation into *M.*

capricolum recipient cells to yield JCVI-syn1.0. These milestones enabled the synthesis of reduced versions of the JCVI-syn1.0 genome with subsequent transplantation into *M. capricolum* to test for viability. The genome reduction process was guided by transposon mutagenesis studies on the original JCVI-syn1.0 genome, as well as on intermediate reduced genome versions. Successful genome minimization depended on identifying both essential genes, whose disruption is immediately lethal, and quasi-essential genes, whose disruption causes an observable growth disadvantage. Quasi-essential genes were identified by observing if cells with potentially growth-reducing gene disruptions were outgrown during sufficiently long competition experiments, so that cells sampled from much later generations no longer contained the disrupted gene. Three cycles of genome design, assembly and growth testing yielded JCVI-syn3.0 [12].

JCVI-syn3.0 contains all the genes of JCVI-syn1.0 that are required for growth. This includes both essential and quasi-essential genes. Individually non-essential genes were removed in the design for JCVI-syn3.0 to the greatest extent possible without causing synthetic lethality or a major growth disadvantage from simultaneous knockouts. However, in a few cases, genes that appear to be non-essential were retained for ease of genome design and construction. Intriguingly, the role of a considerable fraction of the minimal genome of JCVI-syn3.0 could not be elucidated in spite of extensive bioinformatic analyses. At the time of publication of the minimal cell, 149 of the genes ($\sim$31 % of the genome) could not be assigned a completely specific biological function. Assignment to a broad functional category could not even be made for a subset of 79 genes. These genes of unknown or poorly defined function potentially point towards required features of cellular life yet to be discovered.

The original minimal cell JCVI-syn3.0 genome was assembled by combining individually-minimized 1/8$^{\text{th}}$ chromosome segments [12]. Phenotypic traits of JCVI-syn3.0 included extensive filamentation and vesicle formation during growth, and a doubling time of 2–3 hours (compared to the spheroidal morphology and 1-hour doubling time conferred by the JCVI-syn1.0 genome). To address these phenotypic disadvantages, an alternative design of segment 6 was found to restore consistent morphologic features and increase the growth rate, while retaining a near-minimal genome. This new design incorporated 19 additional genes from JCVI-syn1.0 segment 6 that were not present in JCVI-syn3.0, including those encoding the cell partitioning proteins FtsZ and SepF along with others of unknown function; in addition, two other genes retained in JCVI-syn3.0 segment 6 were removed. The complete genome sequence of this strain, termed JCVI-syn3A, is available through NCBI under the accession number CP016816.2. [18] This entry contains 498 genomic features, however 3 of those are pseudo-genes and two are genes required for cloning in yeast.

JCVI-syn3A has a doubling time of $\sim$2 h and consistently forms spherical cells of approximately 400 nm in diameter. With a 543 kbp genome containing 493 genes of which 452 code for proteins and 38 for RNAs, JCVI-syn3A still has a smaller genome than any natural autonomously replicating cell while providing a robust and versatile platform to study the basics of life. In particular, this minimal cell opens up the possibility to pursue the construction of a complete *in silico* model including the function of all genes. The map of protein coding genes (Figure 1) clearly shows the fundamental importance of Syn3A as a platform to study the principles of life. Among the model bacteria *E. coli* and the related and well-studied [21–25] *Mycoplasma pneumoniae*, JCVI-syn3A has the smallest ratio of genes involved in metabolism to those in genetic information processing. With 91 it also has the smallest number of genes that are considered to have no known (unclear) function compared to 311 and 1,790 for *M. pneumoniae* and *E. coli* respectively (see Table 1 and Supplementary File 1C–1D for an itemized account of the functional categories for the three genomes).

A model for ribosome biogenesis that includes DNA replication, transcription, translation, and ribosome assembly in slow growing *E. coli* has already been developed [26, 27]. As its components have on average 50 % sequence identity to those genetic information processing genes in JCVI-syn3A, this model is assumed to be applicable to JCVI-syn3A as well. Hence, the next important step in modeling JCVI-syn3A is the reconstruction of its metabolic network.

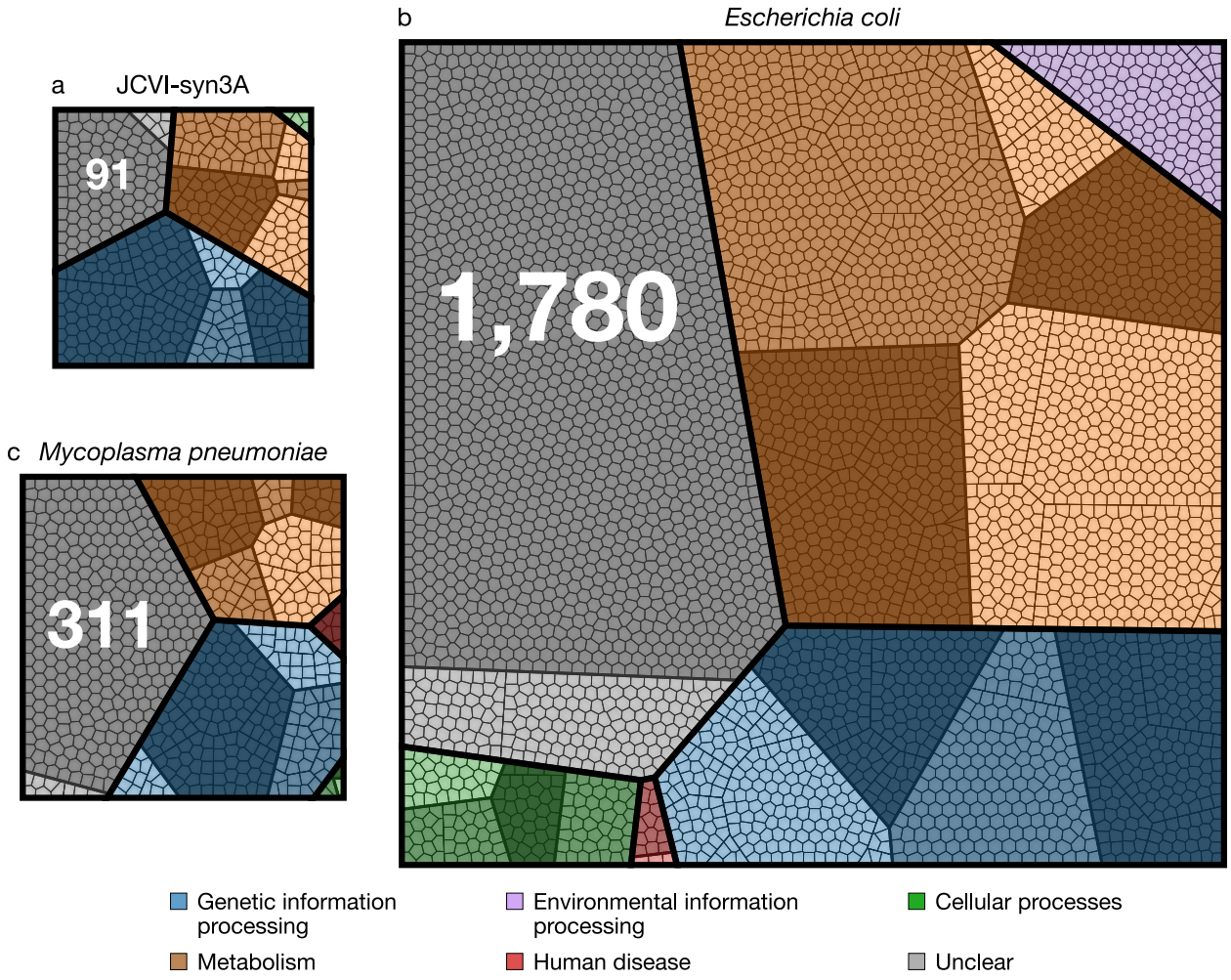


FIGURE 1 Comparison of protein coding genes in the genomes of JCVI-syn3A (NCBI GenBank: CP016816.2 [18]), *M. pneumoniae*, (NCBI GenBank: U00089.2 [19]), and *E. coli* (NCBI GenBank: NC_012967.1 [20]) with 452, 688, and 4,637 coding genes, respectively. Each color represents a primary functional class, each contiguous shaded region corresponds to a secondary functional class, within each of the shaded regions the bold lines separate tertiary functional classes, finally each polygonal cell represents a single gene. The functional class hierarchy is presented in Supplementary File 1A. The ratio of metabolic to genetic information processing genes—0.67, 0.79, and 2.23 respectively—is smallest for JCVI-syn3A. The JCVI-syn3A genome contains both the smallest absolute number of genes of unclear function and the smallest percentage, 91 (20 %), compared to *M. pneumoniae* with 311 (45 %) and *E. coli* with 1,790 (38 %).

The metabolic reconstruction presented here is based on the curated genome annotation, extensive experimental information from the literature on *M. mycoides capri* and other mycoplasma species, and accompanying transposon insertion and proteomics data. Our model features 338 reactions organized in nine subsystems (see Supplementary File 1B), involving 304 metabolites, catalyzed by gene products of 155 genes, thus covering a third of the genes of JCVI-syn3A. The reconstruction process enabled us to suggest several annotation refinements and updates, and yielded a metabolic network that is fairly complete.

Together with the reconstructed biomass composition of JCVI-syn3A and estimates of its reaction flux constraints and energy expenses, the reconstructed metabolic network was cast into a flux-balance analysis (FBA) model [28]. FBA yields the set of steady-state reaction fluxes through a metabolic network that maximize a pre-defined objective function, e.g. production of cellular biomass. The solution space of possible fluxes is constrained not only by the steady-state assumption, but also by specific flux limits accounting for maximal uptake/secretion rates or cellular energy expenses. If these flux limits are not known, the network stochiometry predicts the biomass yield achieved by the cell, i.e. gram biomass produced/gram carbon source taken up (or equivalently biomass production rate/carbon substrate uptake rate). If flux constraints, in particular substrate uptake rates are known or can be assumed, the yield as growth rate per uptake rate can be converted to an absolute growth rate. While measurements to derive such flux constraints are not available yet for JCVI-syn3A, some measurements are available from other mycoplasmas and bacteria that have the same high-affinity glucose transporter (PtsG) found in JCVI-syn3A. Using the glucose uptake rate measured in *M. pneumoniae* [25] (which is similar to the one measured in slow-growing *E. coli* [29]) and other constraint estimates allows us to provisionally predict a growth rate for JCVI-syn3A; this model growth rate is however sensitive to the assumed uptake rate (see Sensitivity analysis in Appendix 1). In this article, the growth rate predicted by the model is presented with the understanding that for the aforementioned reasons, the prediction is provisional and and comes with a degree of uncertainty. This uncertainty has no bearing on the prediction of *in silico* gene essentialities (see below), which can be obtained by removing certain genes in the model and their associated reactions, and testing whether FBA still predicts a nonzero growth rate for the resulting *in silico* knockout.

This FBA model for JCVI-syn3A allows for the analysis of the properties of minimized metabolism in JCVI-syn3A. In particular, gene essentiality can be compared between the model and experimental transposon mutagenesis data. Random gene disruption by bombardment with transposon insertions [5, 9] was used to identify non-essential genes in JCVI-syn1.0 that to the most part were removed during the construction of JCVI-syn3.0 [12]; here, genome-scale transposon mutagenesis studies were carried out on JCVI-syn3A to survey the individual essentiality of all its remaining genes. We find that transposon- and model-derived gene essentiality agree well, with every *in silico* essential gene being at least quasi-essential *in vivo* (i.e. removal might not be immediately detrimental, but give a growth disadvantage). The metabolic reconstruction allows us to rationalize the non-essentiality of some genes, and to propose possible further gene removals in JCVI-syn3A. These suggestions from the model are of particular interest as transposon mutagenesis experiments only probe the individual essentiality of genes and do not yield information on which genes could be removed simultaneously. The metabolic construction, on the other hand, allows us to suggest which genes might be simultaneously removed. At the same time, *in silico* and *in vivo* essentiality also show some discrepancies, which lead us to postulate new hypotheses about specific gene functions or metabolic capabilities. Protein expression profiles of essential and non-essential genes, classified by either transposon mutagenesis studies or FBA *in silico* gene knockouts, were not found to differ significantly–possibly indicating by and large the absence of expression regulation that would discriminate gene products based on their essentiality. Finally, the reconstruction process as well as the gene essentiality comparison have yielded a number of informed hypotheses and suggestions for specific experiments that will guide the ongoing experimental investigation of gene functions in the minimal cell.

2 Results

While the minimal cell JCVI-syn3A is a new organism with little experimental data yet available, its natural precursor *M. mycoides capri* has been studied in depth, which informed all aspects of the metabolic model. To refer to genes in the JCVI-syn3A genome, we use the locus names of the form MMSYN1_xxxx as used in the annotation of JCVI-syn1.0 [12] to allow us to discuss genes deleted in JCVI-syn3A more clearly. The MMSYN1_ prefix is omitted for brevity. Understanding the *in vivo* essentiality of genes in JCVI-syn3A is an important first step to the development of the metabolic reconstruction: this is presented first in Section 2.1. Using the protein expression profiles measured for JCVI-syn3A, the biomass composition of JCVI-syn3A is then derived in Section 2.2, as well as the biomass reaction used in the model. The construction and justification of the metabolic model is presented in Section 2.3. The steady-state fluxes obtained from the model are then compared in Section 2.4 to experimental fluxes, as well as to protein abundance-based flux limits. The metabolic energy usage of JCVI-syn3A is analyzed in Section 2.5. Gene essentiality obtained from *in silico* gene knockouts is presented in Section 2.6. Finally, we compare protein expression profiles between essential and non-essential proteins as identified in the model or *in vivo* in Section 2.7.

2.1 Transposon mutagenesis experiments probe *in vivo* gene essentiality

Transposon insertion mutagenesis studies were performed in order to probe the dispensability of individual genes in JCVI-syn3A (see Methods). In this experiment, transposons are randomly inserted into the chromosomes of a population of cells that is then plated under selection for a drug resistance gene carried by the transposon [5, 9]. After transferring to a liquid culture ("passage zero", P_0), four serial passages are performed. DNA from the pooled colonies is isolated and sequenced to determine the location of transposon insertions within the genome at the beginning or at the end of the experiment. When determining transposon locations at the beginning of the experiment, P_1 is used over P_0 to limit any contamination from the DNA of non-viable cells. The number of insertions observed in a coding region can then be used to infer the importance of that gene. We note that not every insertion will necessarily obliterate a gene's function. A graphical presentation of the essentiality classification along with the distribution of transposon insertions over a portion of the genome is presented in Figure 2. It shows that *secA*/0095 is heavily hit with insertions in the $3'$ 25 % of the gene (but practically nowhere else), however SecA is certainly essential because it is a necessary component of the protein translocase, which inserts proteins such as transporters into the membrane. While the absence of gene products for genes carrying transposon inserts has not been confirmed experimentally, genes with relatively high insertion counts are more likely to be functionally disrupted and thus non-essential. Genes that are not required by the organism to grow in rich media will contain many transposon insertions ("non-essential" genes), whereas genes required for cell viability will be sparsely hit by transposon insertions. To identify genes whose disruption might not be immediately detrimental but might cause a growth defect apparent later on, sequencing of the transposon mutagenesis library was performed on P_4 cells as well. Cells with a gene disruption that is not immediately lethal but causes a growth disadvantage will be outgrown after P_4, and the number of insertions for that gene will then significantly decrease from P_1 to P_4. These genes are denoted "quasi-essential".

A Poisson mixture model was used to partition the coding regions into two sets of genes based on the transposon insertion density. By comparing the assignment of genes into classes of sparse and dense transposon insertions between P_1 and P_4, essentiality can be inferred. This classification method considers a gene to be essential if it has been classified to have sparse transposon insertions in both P_1 and P_4, quasi-essential if it was classified to have dense transposon insertions for P_1 and sparse insertions for P_4, and non-essential if the gene was classified as densely hit for both P_1 and P_4. See Methods for a complete description of the classification method. Figure 2–Figure supplement 3 shows the fit of

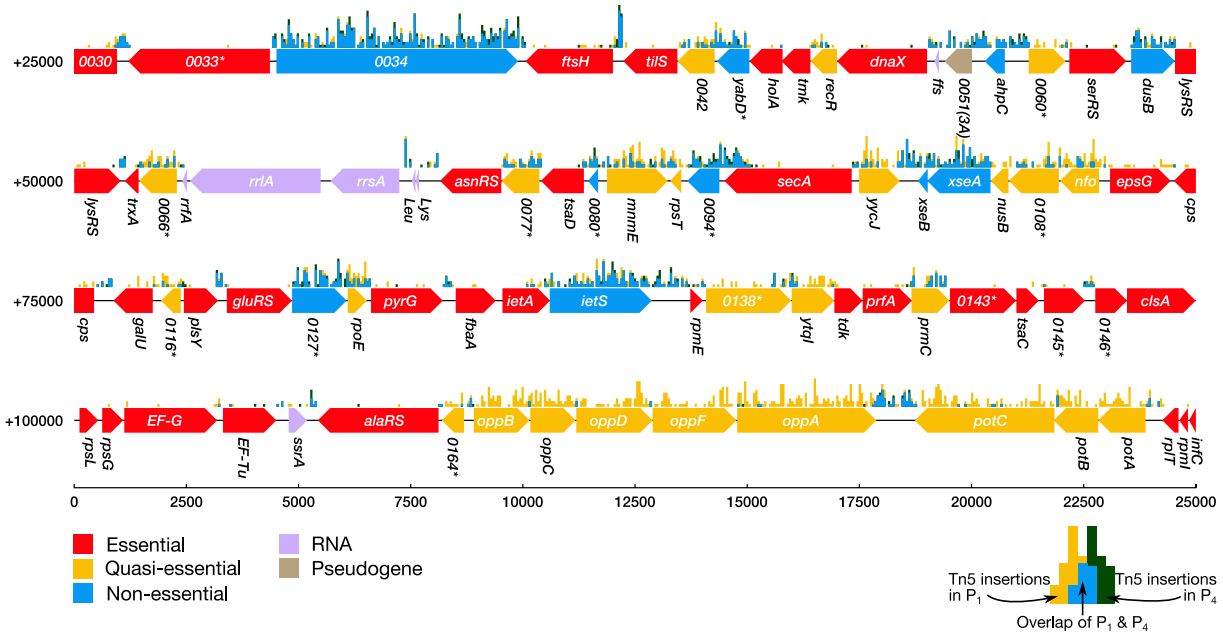


FIGURE 2 Classification of gene essentiality from transposon insertion data using a Poisson mixture model for a representative region of the JCVI-syn3A genome. Coding regions are colored by their predicted class: red (essential), yellow (quasi-essential), blue (non-essential). Lavender regions denote RNA and light brown regions are pseudogenes. The distributions of transposon insertions in passage 1 and passage 4 are represented by yellow and dark green histograms respectively (bin size of 50 bp). The overlap of the two histograms is highlighted in blue. When a common gene name is not available, the four-digit locus tag for JCVI-syn1.0 is used instead. Locus number identifiers with the (3A) suffix represent newly identified open reading frames in JCVI-syn3A which are missing from the JCVI-syn1.0 annotation. Asterisks mark genes with unknown functionality.

Figure supplement 1 Classification of gene essentiality from transposon insertion data using a Poisson mixture model for 0–275000 bp. Coding regions are colored by their predicted class: red (essential), yellow (quasi-essential), blue (non-essential). Lavender regions denote RNA, light brown regions are pseudogenes, and green regions are markers used to construct and implant the genome. The distributions of transposon insertions in passage 1 and passage 4 are represented by yellow and dark green histograms respectively (bin size of 50 bp). The overlap of the two histograms is highlighted in blue. When a common gene name is not available, the four-digit locus tag for JCVI-syn1.0 is used instead. Locus number identifiers with the (3A) suffix represent represent newly identified open reading frames in JCVI-syn3A which are missing from the JCVI-syn1.0 annotation. Asterisks mark genes with unknown functionality.

Figure supplement 2 Classification of gene essentiality from transposon insertion data using a Poisson mixture model for 275000–543379 bp.

Figure supplement 3 Distribution of transposon insertion counts for P_1 (panel a) and P_4 (panel b) compared to the distribution inferred through the Poisson mixture model. To separate genes labeled "non-essential" by the mixture model, but that showed a significant decrease in insertion counts from P_1 to P_4, *k*-means clustering was used on the ratios of transposon insertion rates in P_1 and P_4 for the genes labeled "non-essential". Panel c shows how the genes were divided into two clusters such that the first cluster (blue) contains quasi-essential genes and the second contains truly non-essential genes.

the model to the distribution of transposon insertion counts per gene.

In six out of 452 instances, the mixture model failed to classify the gene, either due to low assignment confidence or due to increased insertions from P_1 to P_4. The short ribosomal proteins S9 (*rpsI*/0637), L27 (*rpmA*/0499), and L31 (*rpmE*/0137) were manually assigned as essential since they are necessary to construct a functional ribosome. The gene *secA*/0095 could not be automatically classified since the mixture model predicted the gene to be more heavily hit with insertions in P_4 than in P_1; it was assigned as essential as it is a major component of the translocase assembly. The insertions occurred in the C-terminus linker domain considered to be important for binding to phospholipids and preprotein translocation. A gene of unclear function (0235) was predicted by the model to be essential at a slightly higher probability than quasi-essential (0.471 vs. 0.416 respectively), however it was manually assigned to be quasi-essential following its previous assignment in JCVI-syn2.0 [12]. Thioredoxin (*trxA*/0065) was assumed to be essential since its associated reductase (*trx*/0819) was predicted to be essential by the mixture model. Only one gene was misclassified: the ATP synthase subunit ε (*atpC*/0789), initially classified as non-essential, was manually reassigned to essential since all other ATPase subunits (*atpD*/0790 through *atpB*/0796) were essential according to the mixture model. The majority of transposon insertions in *atpC*/0789 are found in the 3′ region, similar to the pattern seen in *secA*. However, it is possible that the ε subunit may not actually be essential since in *M. pneumoniae*, transposon insertions into the *atpC* (MPN597) lead to viable cells with decreased cytadherence activity [30].

The set of genes classified quasi-essential could potentially include essential genes which cannot be discriminated using these transposon insertion data. For these misclassified genes, it is possible that although expression of the gene product essential for cell growth has been halted, previously translated essential proteins remain in the cell in sufficient quantities to maintain cell growth and division up to P_1. A further discussion of this argument is presented in Section 3.2 and Appendix 1.

The genes identified as non-essential by the Poisson mixture model may contain "weakly" quasi-essential genes, i.e. disrupted genes which confer a minor growth disadvantage. This behavior would manifest as a decrease in transposon insertions between P_1 and P_4, but not such a steep decline that the gene is observed with little to no insertions. To identify these "weakly" quasi-essential genes, the genes classified as non-essential were subjected to further classification using k-means clustering of the ratio of the number of transposon insertions in P_4 to P_1 assuming two clusters (see Figure 2–Figure supplement 3). Of the 118 genes initially classified non-essential, 42 were reclassified as quasi-essential.

The assignment of essentiality classes and distribution of transposon insertions over the entire genome are presented in Figure 2–Figure supplement 1-2, and Supplementary File 3. Genomic positions of transposon insertions are listed in Supplementary File 2. Figure 3 summarizes the breakdown of the essential, quasi-essential, and non-essential genes according to the functional classes. Of the 452 coding genes in JCVI-syn3A, 60 % are essential, 25 % are quasi-essential, and 15 % are non-essential by this analysis. The detailed breakdown of the JCVI-syn3A genome into these classes (Table 1) shows that of the 91 genes of unclear function, 30 are essential, 32 are quasi-essential, and 29 are non-essential. Those 30 essential genes could represent new biological mechanisms not yet defined and should motivate the search to discover their function [31].

Since on average only one transposon insertion occurs per cell and the identification of insertion locations within the genome is performed over an ensemble of cells, these transposon mutagenesis studies can only reveal *individual* gene essentialities. To probe the essentiality of groups of genes, one would need to perform targeted multiple knockout studies. However, for metabolic genes, flux balance analysis of the metabolic reconstruction can predict the essentiality of groups of genes. In Section 3.5, the individual gene essentiality results are expanded to include the assignment of essentiality to combinations of genes *in silico*, leading to potential combinations of genes to remove to further minimize

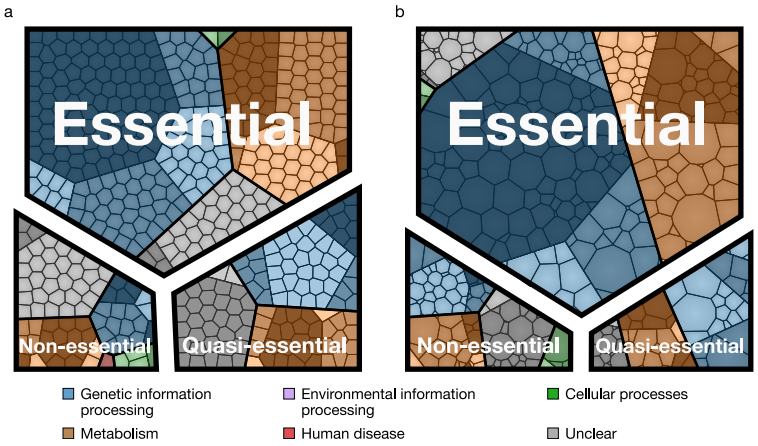


FIGURE 3 Essential, quasi-essential, and non-essential protein coding genes in JCVI-syn3A across four functional classes. (a) Distribution across genome (cell areas all equal). (b) Distribution across proteome (cell areas proportional to protein copy number in an average cell). Among non-essential proteins, the three most abundant ones are *ftsZ*/0522, the peptidase 0305 and 0538 (unclear function). A detailed breakdown of the JCVI-syn3A genome into these classes is available in Table 1.

TABLE 1 Breakdown of protein coding genes in JCVI-syn3A into functional classes.

Functional hierarchy		Protein		Genes		Essentiality			
		%	# unique	%	# unique	# E	# Q	# N	# model
Cellular Processes	Cell Growth	1.02	4	0.88	4	1	0	3	0
	Defense	0.23	2	0.44	2	1	0	1	1
	Subtotal	1.25	6	1.33	6	2	0	4	1
Genetic Information Processing	DNA Maintenance	5.07	38	8.41	38	25	9	4	3
	Folding, Sorting and Degradation	9.58	25	5.53	25	18	7	0	7
	Transcription	3.92	14	3.32	15	8	5	2	0
	Translation	39.5	129	29.7	134	95	28	11	25
	Subtotal	58.1	206	46.9	212	146	49	17	35
Metabolism	Biosynthesis	4.27	29	6.86	31	26	4	1	27
	Central Carbon Metabolism	16.4	46	10.4	47	26	10	11	44
	Energy Metabolism	0.47	4	0.88	4	2	1	1	1
	Membrane Transport	9.37	54	12.6	57	37	16	4	46
	Other Enzymes	1.12	4	0.88	4	2	1	1	1
	Subtotal	31.6	137	31.6	143	93	32	18	119
Unclear	Kegg ortholog defined	1.04	8	1.77	8	3	2	3	0
	No Kegg ortholog	7.98	71	18.4	83	27	30	26	0
	Subtotal	9.02	79	20.1	91	30	32	29	0
Total		100.	428	100.	452	271	113	68	155

the genome. The classifications of the genes used in the metabolic reconstruction are shown in Table 2.

Preliminary triple knockout experiments involving various sets of non-essential genes lead to cells with greatly impaired growth rates (data not shown). The fact that ∼15 % of the genes in JCVI-syn3A are individually non-essential is not inconsistent with the near-minimality of the genome as a whole: it is not possible to remove all non-essential genes without vastly decreasing the growth rate or outright killing the cell. Furthermore, a genome comprised only of essential and quasi-essential genes is non-viable as well, since the removal of a non-essential gene can cause a previously quasi-essential gene to become non-essential in the new construct. As JCVI-syn3A grows more slowly than JCVI-syn1.0 (2 h doubling time vs. 1 h), a gene disruption that in JCVI-syn1.0 led to outgrowth by unaffected competitor cells might still survive through passage four in JCVI-syn3A. As a result, genes that were classified as quasi-essential in JCVI-syn1.0 can appear non-essential in JCVI-syn3A, and could in principle be removed as well—for the price of some gradual further decrease in growth rate. This lack of a clear cutoff again underscores the "trade-off between genome size and growth rate" taking place during genome minimization [12].

2.2 Biomass composition and reaction

The cellular components of JCVI-syn3A fall into three categories: macromolecules, lipids and capsule, and small molecules and ions. Appendix 1-Table 1 lists the mass fractions for all components included in the JCVI-syn3A biomass composition. These mass fractions are used to derive the coefficients in the biomass reaction depicted in Figure 4 for each component based on its molecular weight. The different biomass components are summarized below with the full discussion and derivation in Appendix 1. The growth-associated maintenance (GAM) ATP cost shown in Figure 4 is described in Section 2.3.8.

2.2.1 Macromolecules

The macromolecular mass fractions are based on the experimental composition of *M. mycoides capri* [32], which is assumed to provide a very good approximation for JCVI-syn3A (which was derived from an *M. mycoides capri* substrain). The DNA fraction is slightly increased from the reported 5 % to 5.5 %, which corresponds to exactly one chromosome in JCVI-syn3A. Assuming almost all RNA to be present as ribosomal RNA (rRNA) and around 4600 bases per ribosome yields an upper limit of ∼670 ribosomes per average cell. This number would correspond to ∼20,000 ribosomes in *E. coli* when scaled by cell volume, which is within the observed growth dependent range of 8,000–73,000 for *E. coli* [33]. The RNA base composition is based upon the close relative *M. mycoides capri* serovar LC (*M. mycoides capri* LC) strain Y [34]. (*M. mycoides capri* LC Y is referred to as "*M. mycoides mycoides* goat strain Y" in the older literature, but has recently been included in subspecies *capri* [35].) The DNA composition is determined by the GC content of the genome (24 %). The absolute number of proteins can be estimated from the average protein molecular weight in JCVI-syn3A, which is obtained from the proteomics studies reported in Section 2.7. (See also "Mass Spectrometry Based Proteomics" in Methods for experimental details.) For an average cell, this amounts to ∼77,000 proteins; the resulting protein volume density in a 400 nm spherical cell is 2.3×10^6 proteins/μm^3, which compares well to the estimated density of 3.5-4.4 $\times 10^6$ proteins/μm^3 in *E. coli* [36]. The protein amino acid composition is computed directly from the proteomics data.

In addition to a generic protein species (describing the average JCVI-syn3A protein), two specific proteins are included: acyl carrier protein (ACP, *acpA*/0621) and dUTPase (*dut*/0447). ACP carries a 4′-phosphopantetheine moiety in its holo form, and including the holoenzyme in the biomass equation enforces flux through the corresponding prosthetic group attachment reaction. dUTPase is included for technical reasons discussed in Section 2.3.2. We use

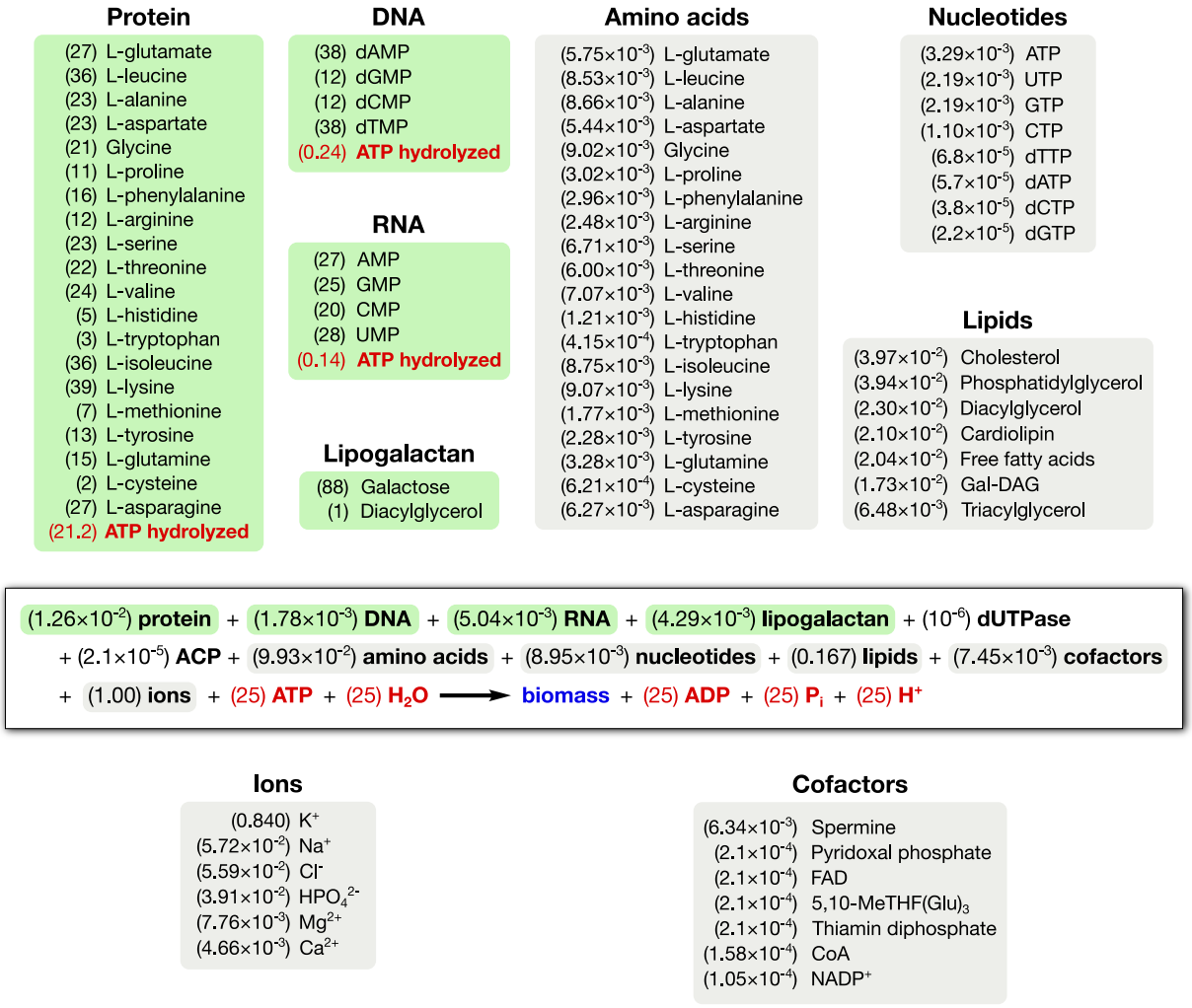


FIGURE 4 Biomass reaction equation for JCVI-syn3A. This reaction consumes biomass precursors (macromolecules, lipids, capsule, small molecules) (black) and consumes energy in the form of ATP (red) to produce biomass (blue). Values in parentheses are the stoichiometric coefficients in mmol compound per gram cellular dry weight (mmol gDW^{-1}). The macromolecular compositions are highlighted in green (stoichiometric coefficients within the macromolecule, unitless) and the compositions of lipids and small molecule pools are highlighted in gray (mmol gDW^{-1}). ATP expenses within green boxes denote total macromolecular synthesis costs (based on macromolecular fractions in the biomass) and the ATP expense in the main equation denotes the nonquantifiable part of the growth-associated maintenance cost (GAM; see Section 2.3.8).

cellular abundances of 138 (ACP) and 10 (dUTPase), derived from the proteomics experiments. The resulting mass fractions are then subtracted from the total protein mass fraction.

2.2.2 Lipids and capsule

Based on the experimental lipid composition of *M. mycoides capri* serovar capri PG3 [37] and *M. mycoides capri* LC Y [38], the model includes the phospholipids phosphatidylglycerol and cardiolipin, the glycolipid monogalactosyl-diacylglycerol (Gal-DAG), cholesterol, diacylglycerol and free fatty acids. For fatty acids, palmitic acid (C16:0) and oleic acid (C18:1 cis-9) are considered to be the two most important representatives, and an average "fatty acid" with a molecular weight averaged between palmitate and oleate is used in all lipid species. In addition, a galactan polysaccharide capsule is included in the biomass. *M. mycoides capri* LC GM12 (the strain from which JCVI-syn3A is derived) is known to produce a galactan polysaccharide (specifically, poly-β-1$\rightarrow$6-galactofuranose) [39]; while it is not yet experimentally known whether the minimal cell still produces this galactan, genetic features suggest it does. As other *M. mycoides capri* LC strains form a polysaccharide capsule but secrete negligible amounts of polysaccharide [40], the JCVI-syn3A galactan is assumed to form a capsule as well and is included as polygalactosyl-diacylglycerol (lipogalactan).

2.2.3 Small molecules & ions

In addition to the macromolecules and lipids, we also include pools of free amino acids, nucleotides and deoxynucleotides in our biomass, as well as cofactors and ions expected to be needed in JCVI-syn3A. A minimal medium for JCVI-syn3A has yet to be obtained, so we use the minimal media reported for *M. mycoides capri* LC Y [41] and *M. pneumoniae* [23] as a guideline for required ions and cofactors: Any compound present in a minimal medium is required by the cell, and the compound or its downstream product(s) need to be included in the biomass composition. From the two media mentioned, all inorganic ions are included in the JCVI-syn3A biomass composition except for sulfate, for which there is no known need in JCVI-syn3A. The vitamin choline from the *M. pneumoniae* minimal medium is also excluded as *M. mycoides capri* does not synthesize its own phosphatidylcholine [38]. Transition metal ions are also not included. While these media are used to determine which compounds to include in the biomass composition, their intracellular concentrations/mass fractions are obtained from measurements in *M. mycoides capri* [42–44] or other mycoplasmas [45] or taken from the iJW145 *M. pneumoniae* [25] and iJO1366 *E. coli* models [46].

2.3 Metabolic reconstruction

The metabolic reconstruction of JCVI-syn3A features 338 reactions involving 304 metabolites, catalyzed by gene products of 155 genes, thus covering a third of the genome. The scope of the reconstruction includes all reactions associated with providing the components of the reconstructed biomass (see Figure 4). Not covered are metabolic functions outside the "core" functions, in particular RNA modifications and damage repair reactions. While many RNA modification enzymes are known already, the prevalence of specific RNA modifications in the RNA pool is not yet known. A few RNA modification enzymes are however discussed with regard to folate metabolism in Section 3.3. The majority of damage reactions and possible repair thereof are mostly not yet known, and are hence omitted save for two genes in cofactor and nucleotide metabolism. Approximately 30 genes pertaining to RNA modification are listed in our KEGG ortholog search as "Genetic Information Processing" and will be included in a future model for ribosome biogenesis and tRNA biogenesis.

The model reactions are organized in nine subsystems, which are listed in Supplementary File 1B together with their respective number of reactions and genes included. Among these subsystems, "Biomass production" contains the biomass reaction discussed in Section 2.2. "Exchange" contains the model reactions that describe metabolite exchange with the media. All other subsystems are discussed in detail in the following subsections. Figure 5 shows a global view of the metabolic network (excluding biomass and exchange reactions). Individual maps for the subsystems for central, nucleotide, cofactor, lipid, macromolecule and amino acid metabolism are depicted in Figure 6–7 and Figure 9–12. Table 2 lists all genes included in the reconstruction, together with their *in vivo* and *in silico* essentiality.

Experimentally, JCVI-syn3A is grown in the rich and not fully defined SP4 media [47, 48], since a defined media supporting its normal growth has yet to be obtained. Consequently, a rich *in silico* medium that provides for all biomass precursors the cell can take up is assumed, with glucose as the only energy source.

2.3.1 Central metabolism

A schematic diagram of central metabolism in JCVI-syn3A is provided in Figure 6. The only annotated sugar importer in the JCVI-syn3A genome is the glucose PTS system comprising PtsI (*ptsI*/0233), PtsH (*ptsH*/0694), Crr (glucose-specific IIA component, *crr*/0234) and PtsG (*ptsG*/0779). The phosphate-transfer reaction chain from phosphoenolpyruvate (PEP) to glucose [50] is lumped together into a single reaction that imports glucose and phosphorylates it with PEP. The presence of ManA (mannosamine-6-phosphate isomerase, *manA*/0435) and NagB (glucosamine 6-phosphate deaminase, *nagB*/0726) suggests possible utilization of mannose and glucosamine as well, which is supported by mutation studies suggesting that in *M. mycoides mycoides* strain T1, PtsG is involved in the uptake of all three sugars [51]. Thus PTS-mediated combined uptake and phosphorylation is included for glucose, glucosamine and mannose. Furthermore, the presence of a putative *nagA* operon (0493 through 0495) suggests possible uptake of N-acetylmannosamine or N-acetylneuraminate (sialic acid). Therefore, an uptake reaction for N-acetylmannosamine is included, noting that this is probably not imported through the glucose PTS system, as concomitant phosphorylation would render the putative N-acetylmannosamine kinase NagC (0495) redundant. Lacking more information, we assume that this import reaction is mediated by an ATP-binding cassette (ABC) transporter instead.

As uptake and secretion measurements for JCVI-syn3A are not yet available, similar uptake rates to those measured in *M. pneumoniae* [25] are used, as has been done in previous mycoplasma models [52]. Since glucose is considered to be the only energy source, upper limits of 0.0 $\mathrm{mmol\,gDW^{-1}\,h^{-1}}$ on mannose, glucosamine and N-acetylmannosamine uptake are applied. A maximum glucose uptake rate of 7.4 $\mathrm{mmol\,gDW^{-1}\,h^{-1}}$ measured in mid-exponential phase *M. pneumoniae* is used. We note that since mannose and glucosamine would compete for the same PTS importer with glucose, their uptake would not increase the overall sugar uptake.

Four possible sugar sources are assumed to feed into glycolysis via fructose-6-phosphate (F6P): glucose, mannose, glucosamine and N-acetylmannosamine. Only the putative N-acetylmannosamine-6-phosphate epimerase NanE (*nanE*/0494) has been annotated, however, it seems likely that this gene forms part of an operon together with its two adjacent genes. The RAST annotation pipeline [53] suggests the putative ROK family gene 0495 codes for N-acetylmannosamine kinase (NagC). Its genomic context suggests 0493, annotated as a putative dipeptidase, codes for N-acetylglucosamine-6-phosphate deacetylase (NagA). This would complete the N-acetylmannosamine utilization pathway and would be consistent with the putative amide cleavage functionality.

Starting from fructose-6-phosphate, the annotation of JCVI-syn3A contains a complete glycolytic pathway as well as the non-oxidative branch of the pentose phosphate pathway up to the nucleotide precursor phosphoribosylpyrophosphate (PRPP), in line with experimental studies on *M. mycoides capri* LC Y [54, 55]. As in other mycoplasmas [4], the gene for transaldolase (TALA) has not been identified, however transaldolase activity has been

TABLE 2 Genes modeled in the metabolic reconstruction. The "MMSYN1_" prefix on the locus tags has been omitted for brevity. The reaction column provides the specific reaction name or general description of the gene (if involved in multiple reactions). Reaction names may appear multiple times if there are multiple gene products that can catalyze that reaction. Column Ess_{Tn5} contains a · if the gene is non-essential, a □ if it is quasi-essential, or a ■ if it has been determined to be essential through the transposon mutagenesis experiments. A dagger in this column indicates that the automatic essentiality assignment required manual intervention. Column Ess_{FBA} contains a · if the gene is non-essential or a ■ if it has been determined to be essential through FBA. Loci marked with an asterisk are genes that are non-essential only "technically" with respect to FBA (see Section 2.6). The doubling times predicted by FBA for non-essential genes were all 2.02 h, with the exception of single knockouts of loci *pdhC*/0227 through *ackA*/0230, which all had doubling times of 3.22 h,; locus *punA*/0747 with a doubling time of 2.04 h; and locus *gltP*/0886 with a doubling time of 2.03 h.

Locus	Reaction	Ess_{Tn5}	Ess_{FBA}
Amino acid metabolism			
0381*	AHCi	·	·
0163	ALATRS	■	■
0535	ARGTRS	■	■
0076	ASNTRS	■	■
0287	ASPTRS	■	■
0837	CYSTRS	■	■
0687	GLNTRAT	■	■
0688	GLNTRAT	■	■
0689	GLNTRAT	■	■
0126	GLUTRS_Gln	■	■
0405	GLYTRS	■	■
0288	HISTRS	■	■
0519	ILETRS	■	■
0634	LEUTRS	■	■
0064	LYSTRS	■	■
0432*	MAT	■	·
0012	METTRS	■	■
0528	PHETRS	■	■
0529	PHETRS	■	■
0282	PROTRS	■	■
0133	Peptidase 1	·	·
0305	Peptidase 2	□	·
0444	Peptidase 3	□	·
0479	Peptidase 4	□	·
0061	SERTRS	■	■
0222	THRTRS	■	■
0308	TRPTRS	■	■
0613	TYRTRS	■	■
0260	VALTRS	■	■
Central metabolism			
0230	ACKr	■	·
0493	AGDC	□	·
0495	AMANK	·	·
0494	AMANPEr	·	·
0732	DRPA	·	·
0213	ENO	■	■
0131	FBA	■	■
0726	G6PDA	·	·
0607	GAPD	■	■
0451	GAPDP	■	·
0475	LDH_L	■	·
0435	MAN6PI	·	·
0227	PDH_E2/_acald	·	·
0228	PDH_E3	·	·
0220	PFK	■	■
0445	PGI	■	■
0606	PGK	■	■
0729	PGM	■	■
0831	PRPPS	■	■
0229	PTAr	■	·
0221	PYK	■	■
0262	RPE	□	■
0800	RPI	□	■
0316	TKT	□	■
0727	TPI	■	■
Cofactor metabolism			
0823	5FTHFPGS	□	■
0390	FMETTRS	□	■
0291	FMNAT	■	■
0443	FTHFCL	■	■
0799	GHMT	□	■
0684	MTHFC	□	■
0259	NADK	■	■
0378	NADS	■	■
0614	NCTPPRT	■	■
0380	NNATr	■	■
Lipid metabolism			
0621	ACP	■	■
0419	ACPPAT	■	■
0513	ACPS	■	■
0512	AGPAT	■	■
0117	APG3PAT	■	■
0139	BPNT	□	■
0147	CLPNS	■	■
0697	DAGGALT	□	·
0114	DAGPST/DAGGALT	■	■
0304	DASYN	■	■
0420	FAKr	■	■
0616	FAKr	■	■
0617	FAKr	□	■
0115	GALU	■	■
0218	GLYK	■	■
0733	PGMT/PPM	■	■
0214	PGPP	□	■
0875	PGSA	■	■
0113	PSSYN	■	■
0813	UDPG4E	■	■
0814	UDPGALM	■	■
Macromolecules			
0394	Lon	□	■
0650	Met peptidase	■	■
0201	Pept. deformylase	■	■
Nucleotide metabolism			
0651	(D)ADK	■	■
0413	ADPT	■	■
0129	CTPS	■	■
0347	CYTK	□	■
0515	DCMPDA	□	·
0447	DUTPDP	■	■
0203	GK	■	■
0216	GUAPRT	□	■
0747	PNP	□	·
0344	PPA	■	■
0771	RNDR	□	■
0772	RNDR	■	■
0773	RNDR	□	■
0140	TMDK1/DURIK1	■	■
0045	TMPK	■	■
0065	TRDR	■†	■
0819	TRDR	■	■
0537	UMPK	■	■
0798	UPPRT	■	■
0330	dAdn kinase 1	·	·
0382	dAdn kinase 2	·	·
Transport			
0822	5FTHFabc	□	■
0876	AA permease 1	·	·
0878	AA permease 2	□	·
0789	ATPase	■†	■
0790	ATPase	■	■
0791	ATPase	■	■
0792	ATPase	■	■
0793	ATPase	■	■
0794	ATPase	■	■
0795	ATPase	■	■
0796	ATPase	■	■
0879	CA2abc	■	■
0836	COAabc	■	■
0642	EcfA	■	■
0643	EcfA	■	■
0641	EcfT	■	■
0233	GLCpts	■	■
0234	GLCpts	■	■
0694	GLCpts	■	■
0779	GLCpts	■	■
0886	GltP	□	·
0685	Kt6	■	■
0686	Kt6	■	■
0401	LIPTA	·	·
0787	MG2abc	■	■
0314	NACabc	■	■
0165	Opp	□	·
0166	Opp	□	·
0167	Opp	□	·
0168	Opp	□	·
0169	Opp	□	·
0345	P5Pabc	■	■
0425	PIabc	■	■
0426	PIabc	■	■
0427	PIabc	■	■
0877	RIBFLVabc	■	■
0008	RNS	■	■
0009	RNS	■	■
0010	RNS	■	■
0011	RNS	■	■
0195	SPRMabc	□	■
0196	SPRMabc	□	■
0197	SPRMabc	□	■
0706	THMPPabc	□	■
0707	THMPPabc	□	■
0708	THMPPabc	□	■

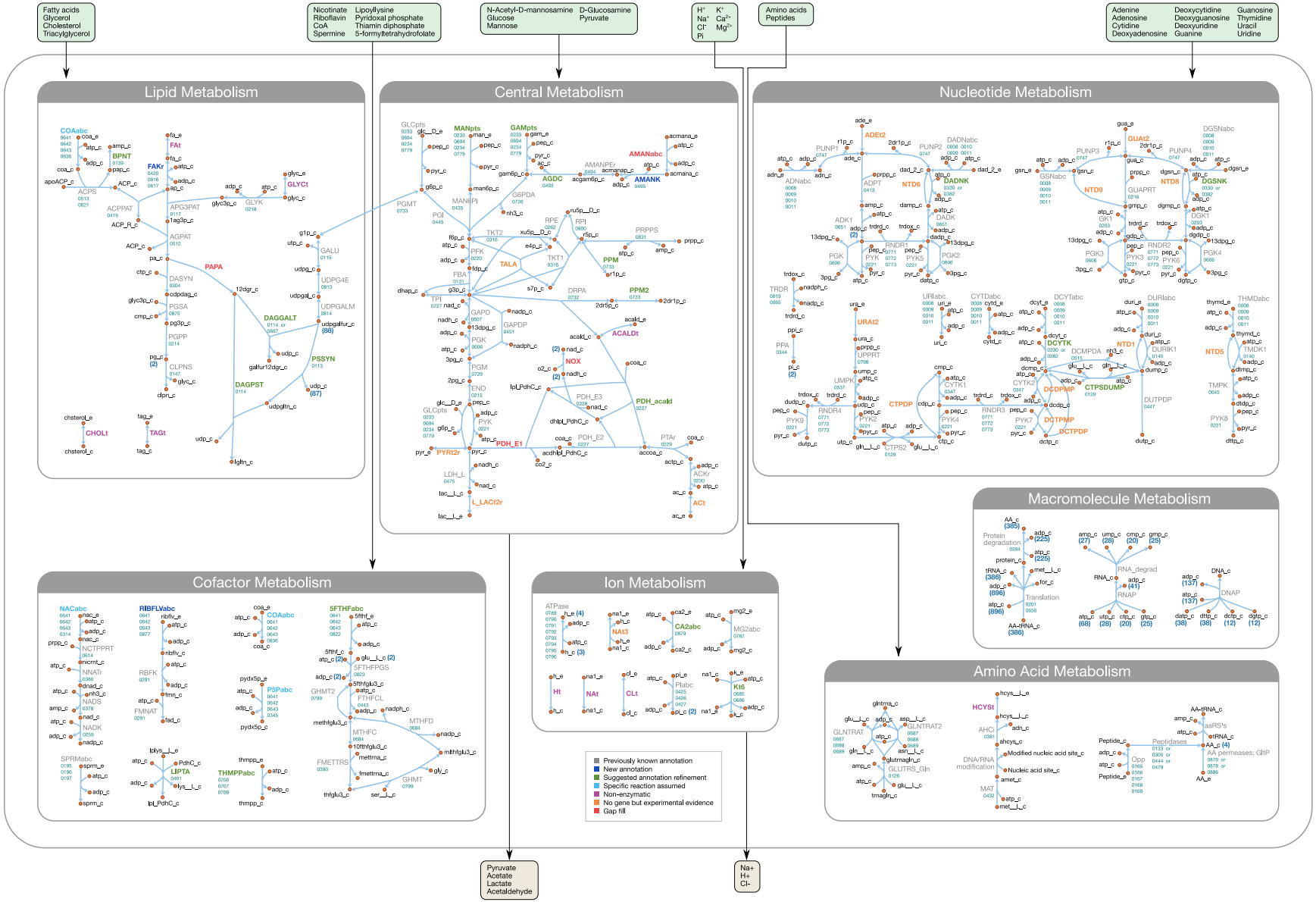


FIGURE 5 Overview of the metabolic reconstruction of JCVI-syn3A, drawn with Escher [49]. Orange nodes represent metabolites, labeled by their short names in the model (black); the suffixes "_c" and "_e" denote cytoplasmic and extracellular compartments, respectively. For clarity, H_2O, H^+, PP_i and P_i are generally omitted as reactants. Blue edges represent (enzymatic or spontaneous) reactions, labeled by reaction name (gray labels) and associated gene loci (gene-protein-reaction (GPR) rules, turquoise; omitting "MMSYN1_" prefix). Blue parenthesized numbers denote reactants (products) which are consumed (produced) in stoichiometric quantities greater than one. In this map and subsequent maps, the following color scheme for highlighted reactions is used—blue: reaction based on new annotation, light green: reaction based on suggested annotation refinement, cyan: specific reaction assumed for generic annotation, light violet: non-enzymatic reaction, orange: reaction not accounted for by gene yet but supported by experimental evidence, and red: reaction included based on gap filling. Small boxes list metabolites that can be taken up (green boxes) or secreted (brown boxes) under physiological conditions.

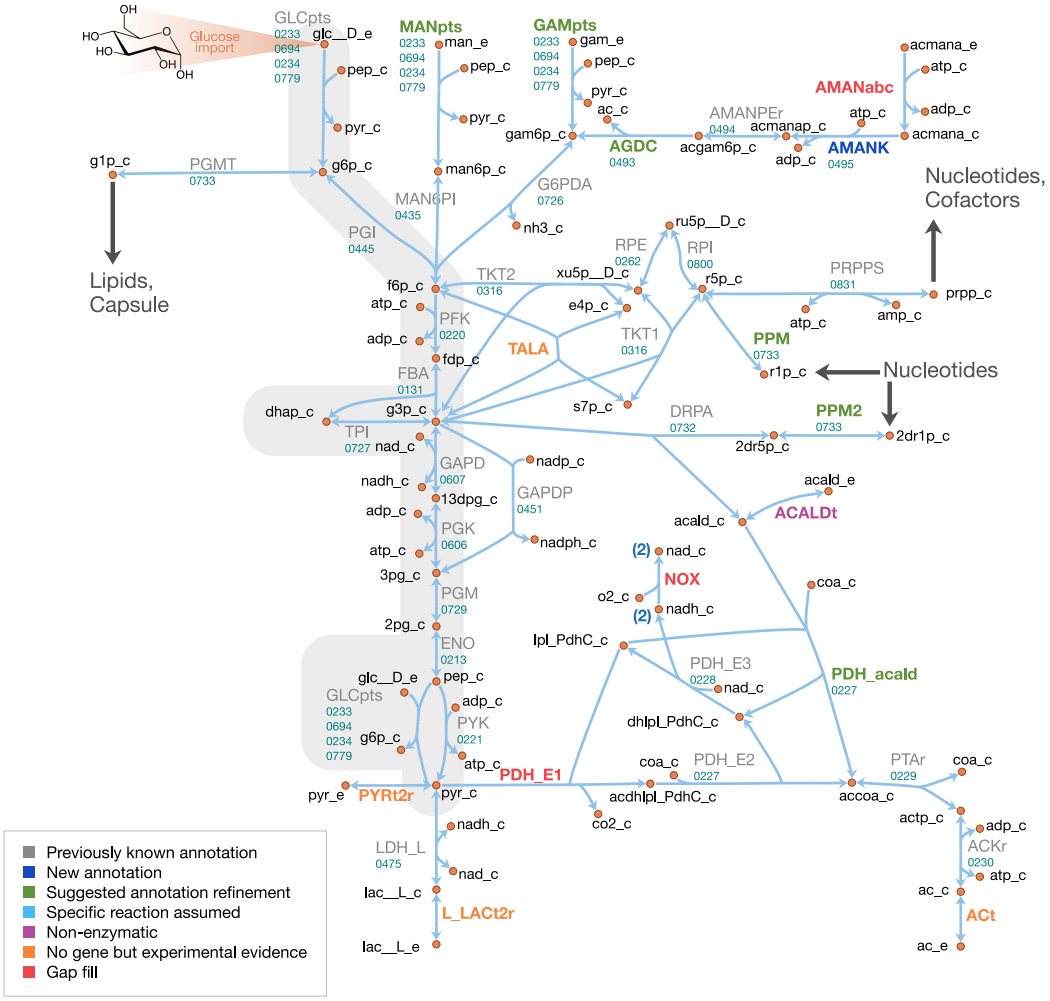


FIGURE 6 Central metabolism in JCVI-syn3A. Map components and labels as in Figure 5. Big arrows denote incoming or outgoing connections to other parts of the metabolic network. For context, the node representing glucose transport has been labeled explicitly and glycolysis has been highlighted in gray.

Figure supplement 1 Steady-state fluxes through central metabolism in JCVI-syn3A. Map components and labels as in Figure 5, with gene loci/gene-protein-reaction rules omitted. Numbers after reaction labels denote steady-state reaction fluxes in mmol $gDW^{-1}\,h^{-1}$; edge color corresponds to the absolute value of the carried flux—gray to blue to purple to red, from low to high flux. For reversible reactions, the reaction progresses from the white to the filled arrowhead.

detected experimentally [56]. In particular, it has been detected in *M. mycoides capri* LC Y cell extracts [54], albeit the detected specific activity was rather weak (13 μmol min^{-1}g cell $protein^{-1}$) and the level of background noise in that study is not known. Thus this reaction is included in the model but we note that the evidence seems ambiguous.

While the completion of the glycolytic pathway via lactate dehydrogenase (LDH, *ldh*/0475) is possible, several genes from the acetate fermentation branch have been deleted, namely the E1 subunit of pyruvate dehydrogenase (PDH_E1, MMSYN1_0225 and MMSYN1_0226) and NADH oxidase (NOX, MMSYN1_0223). However, the remaining subunits of the PDH complex (PDH_E2 and PDH_E3, *pdhC*/0227 and *pdhD*/0228), as well as the path from acetyl-CoA to acetate, are still present in the genome. NAD regeneration could possibly be carried out by one of the remaining oxidoreductases of unclear functionality. Until further information becomes available, the remaining PDH complex is assumed to act on pyruvate to yield acetyl-CoA and finally acetate.

Another possible function for the remaining PDH complex would be oxidation of acetaldehyde to acetyl-CoA, which would not require a decarboxylation in the absence of the PDH_E1 subunit (see Appendix 1). Phosphopentomutase (PPM) and deoxyribose phosphate aldolase (DRPA) activity have been experimentally observed in *M. mycoides capri* LC Y [54], enabling the breakdown of deoxyribose 1-phosphate (dR1P) into glyceraldehyde 3-phosphate (G3P) and acetaldehyde (acald). A gene for deoxyribose phosphate aldolase has been annotated in JCVI-syn3A (*deoC*/0732). A strong candidate for phosphopentomutase activity is the putative phosphoglucomutase (PGMT) gene *deoB*/0733. It is preceded by the deoxyribose phosphate aldolase gene (*deoC*/0732) and (in the original JCVI-syn1.0 genome) succeeded by the pyrimidine nucleoside phosphorylase MMSYN1_0734, i.e. it is neighboring two genes responsible for dR1P production and dR5P breakdown, respectively. At the same time, it shows some similarity (21 % sequence identity) to the phosphopentomutase TK1777 from the archaeon *Thermococcus kodakaraensis*. TK1777 showed activity mainly against d1RP, but also weaker activity against glucose 1-phosphate (G1P) [57]. Thus *deoB*/0733 is assumed to be the gene responsible for both activities.

Secretion of acetate [58], lactate [59] and pyruvate [41] has been observed in *M. mycoides*; with mutational data on *M. mycoides mycoides* T1 [51] indicating a common transporter for pyruvate and lactate. While it is not clear how mycoplasmas secrete acetate, lactate and pyruvate, proton symporters have been suggested for lactate and acetate [25] and such reactions are assumed in other mycoplasma models [52, 60, 61]. The genome of JCVI-syn3A contains several annotated efflux proteins, but all of these show features of ATP-coupled transporters, suggesting they are not involved in lactate or acetate export. Thus, lactate, pyruvate and acetate secretion reactions are included as proton symports, noting that for the purposes of this model, this is equivalent to assuming secretion of neutral acid species. The acetate secretion rate is constrained to a maximum of 6.9 mmol gDW^{-1} h^{-1} following Wodke et al. [25], and the lactate secretion rate is kept unconstrained as the optimal FBA solution will always route as much flux as possible through the acetate pathway, yielding one more ATP per pyruvate. The pyruvate secretion rate is also left unconstrained; this reaction only carries flux under certain *in silico* gene knockout conditions.

2.3.2 Nucleotide metabolism

A schematic diagram of nucleotide metabolism in JCVI-syn3A is presented in Figure 7. The JCVI-syn3A annotation contains a putative ribonucleoside (RNS) ABC import system (*rnsD*/0008 through *rnsB*/0011), which is assumed to import all nucleosides (ribo- and deoxyribo-), but no individual bases or free ribose, in accordance with the experimental characterization of the ribonucleoside ABC importer in *Streptococcus mutans* [62]. While intact nucleotides are rarely taken up as a whole, there have been reports of *M. mycoides capri* LC Y being capable of taking up deoxymononucleotides [63, 64]. However, no gene has been identified for this functionality in that strain, nor is there any hint that the minimized JCVI-syn3A still possesses this ability. We note that competition experiments suggest distinct

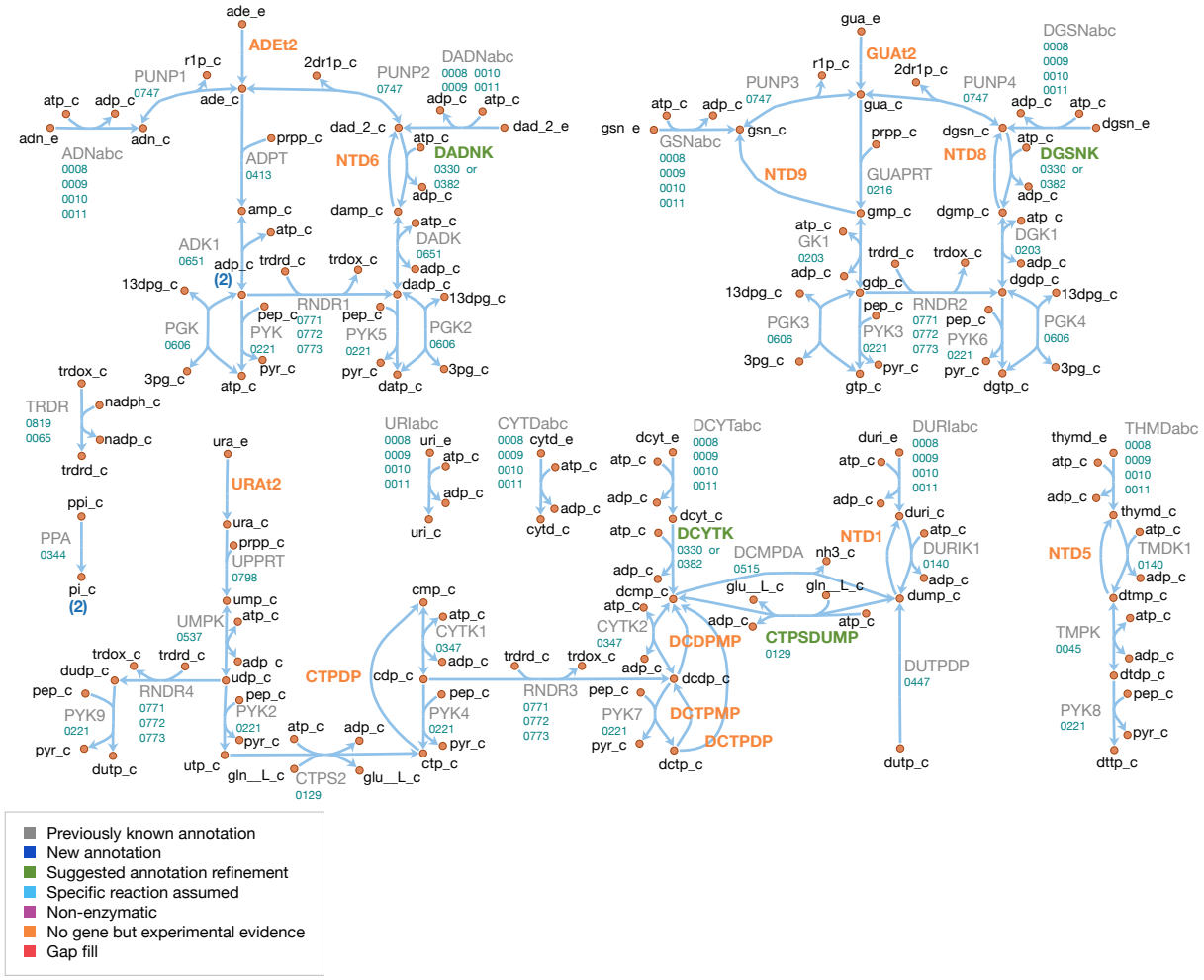


FIGURE 7 Nucleotide metabolism in JCVI-syn3A. Map components and labels as in Figure 5.

uptake systems for nucleosides and nucleotides [65], i.e. the RNS importer should not allow for nucleotide uptake. The presence of several nucleoside kinases as well as phosphoribosyltransferases in JCVI-syn3A suggests that this nucleotide uptake ability, if present at all, cannot provide nucleotides in sufficient amounts and thus no nucleotide uptake reactions are included. The mechanism of nucleobase uptake in mycoplasmas has not been established, but a proton symport mechanism has been suggested [65]. This mechanism is used in the model as well, as done in other mycoplasma models [25, 52, 60, 61].

The further reconstruction of nucleotide metabolism is aided by the extensive experimental studies on nucleotide salvage pathways in *M. mycoides capri* LC Y [34, 43, 44, 66–68] and *M. mycoides mycoides* SC [69, 70]. The reactions detected or inferred are in agreement with the existing annotations, and help to refine possible specificities and suggest additional functionalities.

The genome of JCVI-syn3A contains three deoxynucleoside kinases: *tdk*/0140, *dak1*/0330, and *dak2*/0382. The thymidine kinase, *tdk*/0140, is assumed to phosphorylate both thymidine and deoxyuridine. As further discussed in Appendix 1, it is furthermore assumed that *dak1*/0330 and *dak2*/0382 both act on deoxyadenosine, deoxyguanosine and deoxycytidine, but not significantly on any ribonucleosides. Therefore, AMP, GMP and UMP are only formed directly from their respective bases by the corresponding phosphoribosyltransferases (*hptA*/0216, *apt*/0413, and *upp*/0798).

The genome of JCVI-syn3A contains several mononucleotide kinases (*tmk*/0045, *gmk*/0203, *cmk*/0347, *pyrH*/0537, and *adk*/0651) that can phosphorylate all (deoxy-)mononucleotides except for dUMP [67], but in line with other mycoplasmas [71], the genome of JCVI-syn3A contains no gene for nucleoside diphosphate kinase (*ndk*). Instead, the glycolytic enzymes phosphoglycerate kinase (PGK) and pyruvate kinase (PYK) have been found to phosphorylate other dinucleotides besides ADP in several mycoplasmas [71]. Specifically, PYK was found to act on all (deoxy-)dinucleotides and PGK was found to act on all purine (deoxy-)dinucleotides, but not on pyrimidines. These reactions complete the pathways from the mononucleotides to the final (deoxy-)trinucleotides. We note that the apparent absence of cytidine kinase activity implies that the only route to cytidine nucleotides goes through CTP synthase (CTPS, *pyrG*/0129; aminating UTP to CTP). All deoxytrinucleotides except dTTP can be obtained either from their deoxynucleosides or from the corresponding ribodinucleotide through the action of ribonucleotide diphosphate reductase (RNDR, *nrdE*/0771 through *nrdF*/0773).

In addition to these synthetic pathways, JCVI-syn3A also contains several catabolic reactions. The phosphorolysis of purine nucleosides observed in *M. mycoides capri* LC Y [34, 66] can be carried out by purine nucleoside phosphorylase (PNP, *punA*/0747). PNP is assumed to also act on purine deoxynucleosides, as this activity has been demonstrated in *M. capricolum* and *M. gallisepticum* [72]. However, no pyrimidine nucleoside phosphorylase activity is assumed to be left in JCVI-syn3A (see Appendix 1).

Hydrolase activity against several mononucleotides (GMP, dAMP, dGMP, dUMP and dTMP), a dinucleotide (dCDP) and several trinucleotides (CTP, dCTP and dUTP) has been experimentally observed [44, 66, 68]. A putative dUTPase is annotated in JCVI-syn3A (*dut*/0447). We note the presence of several hydrolases of unclear function that may possibly carry out the other reactions. Thus, all of these hydrolysis reactions are included in the network, without the assignment of a specific gene. Experimental studies suggest a common enzyme for all deoxymononucleotidase reactions, but separate enzymes for dUTP and dCTP hydrolysis [68].

While the observation of dUTPase activity in the natural *M. mycoides capri* does not itself imply that this activity has to be present in JCVI-syn3A, possible DNA incorporation of dUTP (formed from UDP through RDNR and subsequent phosphorylation) is a problem all organisms must contend with, as exemplified by the essentiality of dUTPase in *E. coli* [73] and *S. cerevisiae* [74]. The situation is exacerbated in JCVI-syn3A after the deletion of the repair enzyme uracil-DNA glycosylase (MMSYN1_0436), such that hydrolysis of dUTP is the only apparent defense

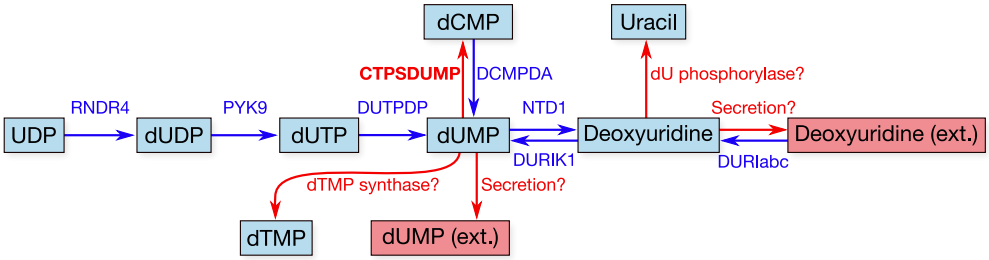


FIGURE 8 Apparent dead-end of dUMP/deoxyuridine and possible solutions. Internal metabolites are highlighted with cyan boxes, external ones with red boxes. Blue arrows denote reactions incorporated during model reconstruction—no reaction leads away from the dUMP/deoxyuridine pair. Red arrows denote hypothetical reactions that could possibly solve this dead-end. In the model, we have adopted the hypothetical CTP synthase reaction converting dUMP to dCMP (see also Figure 7; CTPSDUMP).

mechanism against its incorporation into DNA. Inclusion of this reaction (Figure 7; DUTPDP) is therefore warranted even if the annotation of the candidate gene *dut*/0447 is only tentative. To avoid RNDR being *in silico* essential solely by virtue of being the only source of dUTP in the network (through reactions RNDR4 and PYK9; see Figure 7), dUTPase is included directly in the biomass, rather than enforcing a minimal flux through dUTP formation.

The breakdown of dUTP to dUMP however raises the question of the downstream degradation. The uptake requirement of *M. mycoides capri* for some form of thymine [41, 44], in spite of availability of dUMP, indicates the absence of thymidylate synthase activity (an otherwise common usage for dUMP as a precursor for dTMP), in line with the lack of an annotation for a thymidylate synthase gene in JCVI-syn3A. While the aforementioned hydrolysis reaction (Figure 7; NTD1) would enable degradation of dUMP to deoxyuridine, the deletion of the pyrimidine nucleoside phosphorylase MMSYN1_0734 renders deoxyuridine a dead-end. Thus, the issue arises of how JCVI-syn3A disposes of the dUMP/deoxyuridine formed. The first possibility is through pyrimidine nucleoside phosphorylase activity either by some unidentified paralog of MMSYN1_0734, or by some side activity of the purine nucleoside phosphorylase (*punA*/0747). The second possibility would be the export of either dUMP or deoxyuridine. We note that the possibility to recycle deoxyuridine through pyrimidine nucleoside phosphorylase in the natural *M. mycoides capri* (and thereby also dUMP after its dephosporylation) renders an additional dedicated export system for either metabolite unlikely, but side activity of some other system would be possible.

While the aforementioned dependence of *M. mycoides capri* on external thymine/thymidine rules out any thymidylate synthase activity high enough to meet cellular dTTP needs, such activity has been reported for *M. mycoides mycoides* SC [75]; however, the reported activity was extremely low (∼10 pmol/min/mg cell protein) and no responsible gene could be identified. If such activity was present in JCVI-syn3A as well at a higher level, it might provide for a way to dispose of dUMP. Furthermore, the presence of deoxycytidylate deaminase (DctD, *dctD*/0515) enables the conversion of dCMP to dUMP (which is used in wild-type *M. mycoides capri* to ultimately convert thymine to thymidine [44]). No experimental evidence is available of this enzyme running in the reverse direction to aminate dUMP with free ammonia to form dCMP. However, CTP synthase (CTPS, *pyrG*/0129), which catalyzes the conversion of UTP to CTP, spends ATP and uses glutamine as an amino donor, which suggests that DctD catalyzing the amination without ATP and from free ammonia is unlikely. Instead, the question arises whether CTPS in JCVI-syn3A may have a relaxed substrate specificity. The *E. coli* CTPS was found to not act on UMP [76, 77] or dUTP [78], but may have activity against UDP [76, 77]. The *Lactococcus lactis* enzyme has been found to act on dUTP [79] as does one of the isozymes in *S. cerevisiae* [80]. As relaxed enzymatic substrate specificity in mycoplasmas is a common phenomenon, a further broadening of the substrate range of CTPS in JCVI-syn3A seems possible. The preceding hypothetical mechanisms are summarized in Figure 8. Currently, the available data does not allow for the determination of which of these potential dUMP disposal

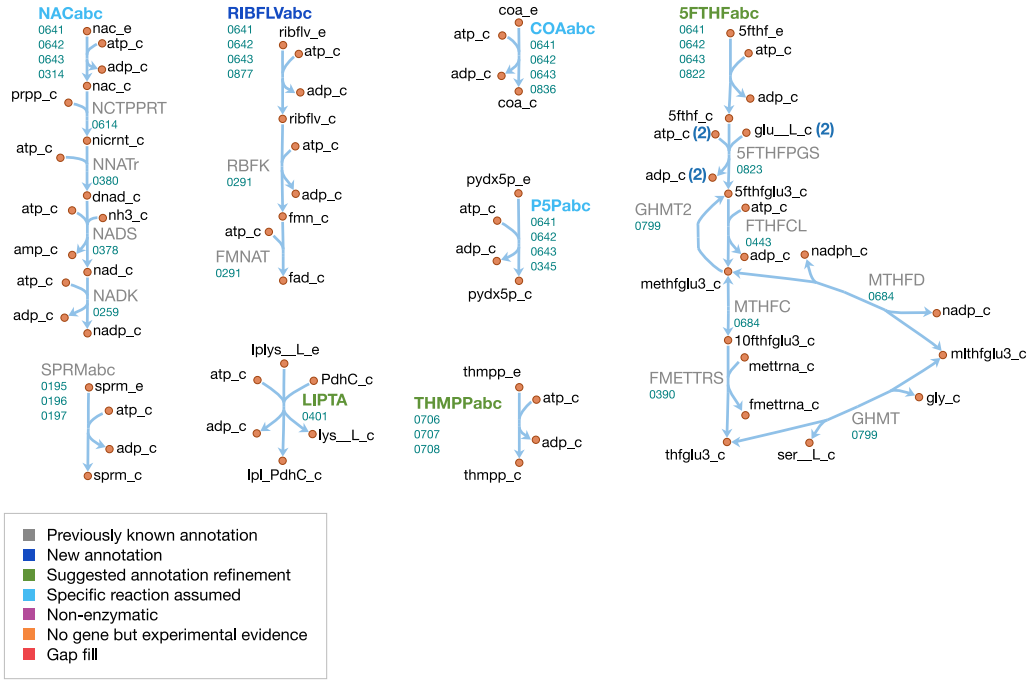


FIGURE 9 Cofactor metabolism in JCVI-syn3A. Map components and labels as in Figure 5.

mechanisms occurs in JCVI-syn3A; however, the increased substrate spectrum of CTP synthase currently seems the most plausible, thus it is included in the model to deal with the produced dUMP.

2.3.3 Cofactor metabolism

The cofactor metabolism of JCVI-syn3A is shown in Figure 9. Many vitamins are known to be taken up through the energy coupling factor (ECF) system [81], which consists of a membrane permease EcfT, a dimer of the ATPase EcfA, and a substrate-binding subunit EcfS [82]. The gene for the latter can either occur in a cluster with the genes for EcfA and EcfT, or several *ecfS* genes can be spread across the genome, their protein products displaying distinct substrate specificities but interacting with a common EcfT$(EcfA)_2$ module in the membrane. The JCVI-syn3A genome annotation lists three consecutive genes *ecfT*/0641, *ecfA*/0642 and *ecfA*/0643 and four *ecfS* genes spread throughout the genome (*ecfS*/0314, *ecfS*/0345, *folT*/0822, and *ecfS*/0836). Folate, riboflavin, coenzyme A, nicotinate, and pyridoxal are all imported through the ECF system. We note that the absence of related salvage enzymes in JCVI-syn3A necessitates the uptake of complete coenzyme A; this is in accordance with the experimental requirement of *M. mycoides capri* LC Y for coenzyme A rather than coenzyme A precursors [41, 59], as well as the apparently incomplete coenzyme A salvage pathway already in JCVI-syn1.0. As no kinase has been identified for pyridoxal, pyridoxal phosphate is assumed to be imported directly.

For the case of folate, *folT*/0822 is assumed to be the necessary substrate-binding unit due to sequence conservation [83] and its adjacency to *folC* [84]. While the exact form of folate in the SP4 medium is not known, we note that generally, 5-formyl-tetrahydrofolate (5-formyl-THF, folinic acid) is the most stable folate derivative, and is known to be imported by FolT [84]. Furthermore, *M. mycoides capri* LC Y was found to be unable to utilize folate itself [85], in line with JCVI-syn3A lacking a gene coding for dihydrofolate reductase. It is thus assumed that folate is taken up in the form of 5-formyl-THF. Genes for the proteins driving the folate cycle consist of GlyA (*glyA*/0799), FolD (*folD*/0684) and Fmt (*fmt*/0390), together with the repair enzyme YgfA (*ygfA*/0443/FTHFCL). YgfA is not only

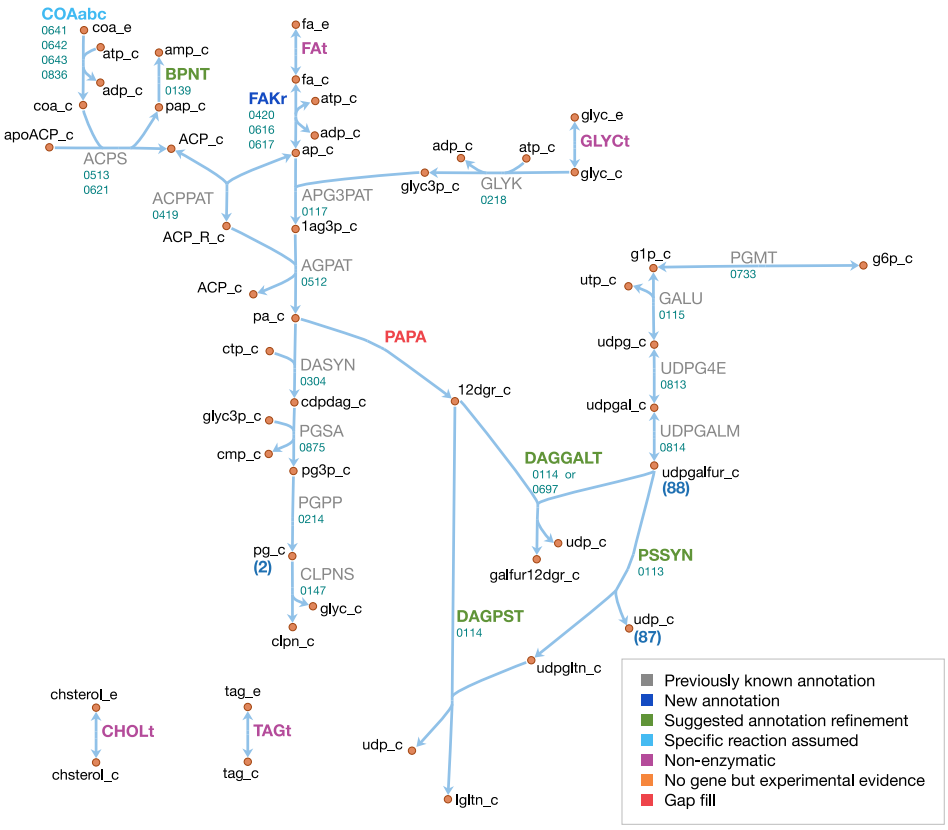


FIGURE 10 Lipid and capsule metabolism in JCVI-syn3A. Map components and labels as in Figure 5.

required to utilize imported 5-formyl-THF, but also to recycle the 5-formyl-THF produced from the hydrolysis of 5,10-methenyl-THF (a side reaction of GlyA [86]). The RAST annotation pipeline [53] suggests that the putative membrane protein gene 0877 is *ribU*, coding for the riboflavin-specific ECF component. As the substrate specificity of the remaining EcfS components (*ecfS*/0314, *ecfS*/0345 and *ecfS*/0836) is unclear, they are tentatively assigned to nicotinate, pyridoxal phosphate, and coenzyme A respectively. For downstream conversion, the genome contains pathways for NAD(P) and FMN/FAD formation from nicotinate and riboflavin.

Spermine and thiamine have their own uptake systems (*potC*/0195 through *potA*/0197 and *thiB*/0706 through 0708, respectively). For thiamine, the deletion of the corresponding diphosphokinase (MMSYN1_0261) suggests that thiamine diphosphate (ThDP) must be taken up directly. Sequence and structural information suggest this to be possible (see Appendix 1 and Appendix 1-Figure 1). While free lipoate is a component of the minimal media for *M. mycoides capri* LC Y [41] and *M. pneumoniae* [23] and is a possible ECF system substrate [81], two putative lipoyl transferases have been deleted in JCVI-syn3A (MMSYN1_0224 and MMSYN1_0464), such that free lipoate cannot be used to provide the lipoyl moiety of pyruvate dehydrogenase subunit E2 (*pdhC*/0227). Instead, simultaneous import and transamidation of lipoate from lipoyllysine onto PdhC, catalyzed by the peptidase 0401, is tentatively assumed in the model (see Appendix 1).

2.3.4 Lipids and capsule

The lipid and capsule metabolism of JCVI-syn3A is presented in Figure 10. In the metabolic reconstruction, the import of four lipid components is necessary: free fatty acids, glycerol, cholesterol and triacylglycerol. These are all assumed to be imported through passive processes. Free fatty acids have been found to be incorporated into the membrane of the mollicute *Spiroplasma floricola* both actively and via passive diffusion [87]. Glycerol is usually imported through dedicated transport systems, yet the glycerol permease GlpF (MMSYN1_0217) has been deleted in JCVI-syn3A. However, it is known that cells can take up glycerol by passive membrane permeation [88, 89] and physicochemical data suggests this could provide sufficient glycerol uptake to fuel the lipid synthesis needs of JCVI-syn3A (see Appendix 1). Cholesterol is known to be incorporated into membranes spontaneously [90, 91] and has been suggested to be incorporated by simple physical absorption in *M. mycoides capri* cells as well [92]. Triacylglycerol was identified as a membrane component in *M. mycoides capri*, but it is not known whether it is still included in the membrane of JCVI-syn3A, so a passive uptake reaction for it is included, noting that the presence of triacylglycerol in the biomass expression then only affects the model by lowering the amounts needed of other lipid species.

The existing annotation with the refinements for the two glycosyltransferases *epsG*/0113 and *cps*/0114 contains nearly complete pathways to produce all membrane components identified in the biomass, with the only gaps occurring in the fatty acid phosphorylation and diacylglycerol (DAG) production pathways. JCVI-syn3A shares the fatty acid utilization pathway from *Staphylococcus aureus* [93], which starts with phosphorylation of free fatty acids in the membrane and subsequent binding to glycerol phosphate (by *plsY*/0117) and acyl carrier protein (ACP, by *plsX*/0419). Holo-ACP is formed from apo-ACP and coenzyme A by ACP synthase (*acpS*/0513), releasing adenosine 3′,5′-bisphosphate (pAp). The DHH phosphoesterase family protein *ytqI*/0139 is 30 % identical to the experimentally confirmed bifunctional oligoribonuclease/pAp phosphatase NrnA from *M. pneumoniae* (MPN140) [94]. Thus, it is assumed that *ytqI*/0139 catalyzes the degradation of pAp to AMP.

Fatty acid kinase consists of an kinase FakA and a fatty acid-binding protein FakB. Both JCVI-syn3A and *S. aureus* contain two copies of FakB, and these have been demonstrated in *Staphylococcus aureus* to display distinct substrate specificities, one acting preferably on unsaturated fatty acids and the other on saturated fatty acids. Assuming similar specificities in JCVI-syn3A would be consistent with the assumed prevailing fatty acids (palmitate and oleate). No annotation for FakA exists in JCVI-syn3A, however, Parsons et al. [93] reports the location of the *fakA* gene in *M. pneumoniae* (MPN547), which shows 33 % sequence identity to *fakA*/0420 and shares the same genomic context (located upstream of *plsX*). Thus it is assumed that *fakA*/0420 is the missing FakA subunit, completing the lipid assembly pathway from free fatty acids to cardiolipin.

As discussed above, JCVI-syn3A has the pathway from glucose-6-phosphate to UDP-galactofuranose, the sugar building block for galactosyl-diacylglycerol (Gal-DAG) and the galactan lipopolysaccharide. In *M. genitalium*, DAG is produced from phosphatidate via phosphatidate phosphatase (PAPA). While no such enzyme is annotated in JCVI-syn3A, and a BLAST search with the *M. genitalium* gene (MPN455) against JCVI-syn3A scores no hits, we note that the presence of a number of unassigned phosphatases in the genome of JCVI-syn3A makes it plausible that one of them could act on phosphatidate. An alternative possibility might be phosphatidate phosphatase side activity by phosphatidylglycerophosphatase (*pgpA*/0214), which has been reported for phosphatidylglycerophosphatase B (PgpB) in *E. coli* [95]. With no gene yet for phosphatidate phosphatase but plausible candidates, this reaction is included as a gap fill.

The product DAG serves as the lipid moiety for the synthesis of lipogalactan (catalyzed by *epsG*/0113 and *cps*/0114, see Section 2.2) and Gal-DAG. Lacking further evidence, Gal-DAG is assumed to be formed by either of the two glycosyltransferases, *cps*/0114 and 0697.

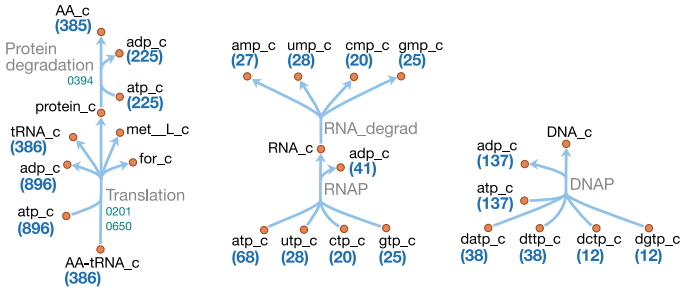


FIGURE 11 Macromolecule metabolism in JCVI-syn3A. Map components and labels as in Figure 5. The detailed (amino acid-specific) stoichiometry of the protein synthesis and degradation reactions can be found in Supplementary File 4. Protein synthesis reactions for the proteins explicitly included in the model (apo-ACP, dUTPase and PdhC) are analogous to the translation reaction shown and are therefore not included in the map.

Possibly related to lipid metabolism, the existing annotation lists a putative choline/ethanolamine kinase (0906); however, as *M. mycoides capri* LC is experimentally known to not produce phosphatidylcholine [38], we assume that this kinase has some yet to be determined substrate and do not include it in the model.

2.3.5 Macromolecules and amino acids

Schematics of the macromolecular and amino acid reaction networks are provided in Figure 11 and Figure 12. The genome of JCVI-syn3A contains a putative oligopeptide ABC importer (Opp/Ami, *oppC*/0166 through *oppA*/0169). In *Lactococcus lactis*, the Opp system has been found to import peptides of four to at least 35 amino acids with little dependence of uptake rates on peptide length or amino acid composition [96, 97]. For the sake of simplicity, peptides imported are assumed to be representative homotetrapeptides of all amino acids except cysteine, since cysteine contained in peptides cannot be easily utilized by mycoplasmas [23]. The model assumes that any of four peptidases (*ietS*/0133, 0305, 0444, and 0479) can split these peptides into individual amino acids.

In addition, the glutamate/aspartate permease *gltP*/0886, as well as two amino acid permeases of unknown specificity (0876 and 0878) have been identified in JCVI-syn3A. The substrate specificities of these two amino acid permeases are not known. However, *M. mycoides capri* LC Y has been found capable of taking up all amino acids in their free form [41] (glutamic and aspartic acid not investigated); thus, the least constraining assumption is made that both permeases can take up all amino acids, except for glutamic and aspartic acid, whose uptake is already enabled by GltP. Proton symport reactions are assumed for each amino acid except glutamate and aspartate, which are symported by GltP while translocating two protons per substrate, as observed in *E. coli* [98].

In order to distinguish between free nucleotides and amino acids and those incorporated in nucleic acids and proteins, explicit macromolecular synthesis reactions (DNA replication, RNA transcription, protein translation) are included that consume amino acids and nucleotides according to the assumed macromolecular compositions. These reactions produce representative DNA, RNA and protein species that enter the biomass reaction according to the mass fractions in Appendix 1-Table 1. Similar macromolecular synthesis costs as in *E. coli* [99, 100] are assumed, i.e. 1.37 ATP/nucleotide in DNA synthesis, 0.41 ATP/nucleotide in RNA synthesis, 2 ATP/amino acid in tRNA charging, and 2.32 ATP/peptide bond in protein synthesis. Protein products are represented as species containing 385 amino acids (based on the average protein length obtained from the JCVI-syn3A proteomics, and accounting for the N-terminal methionine cleaved off the nascent peptide–see below). DNA and RNA are represented as species of 100 bases each to keep the nucleotide reaction coefficients small (and also since no average RNA length is known).

In addition, separate translation reactions are included for three additional proteins: ACP and PdhC, whose prosthetic group attachment is included in the model; and dUTPase, which is included explicitly in the biomass. Protein

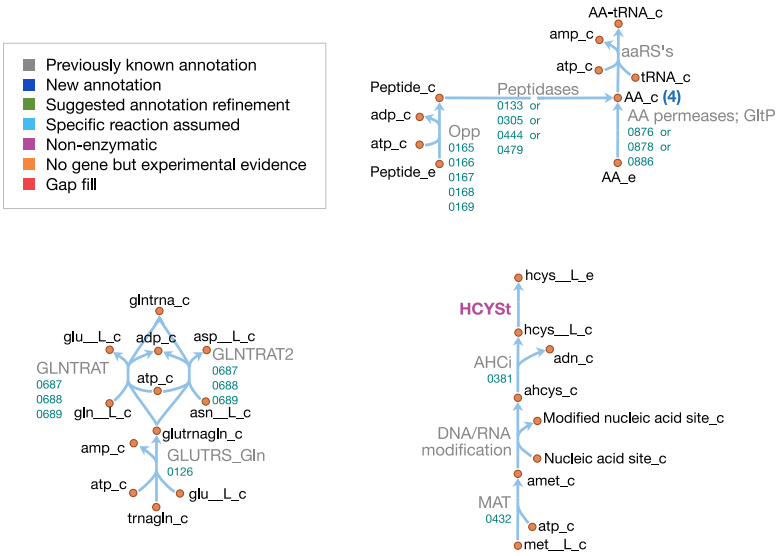


FIGURE 12 Amino acid metabolism in JCVI-syn3A. Map components and labels as in Figure 5. As amino acid metabolism in JCVI-syn3A constitutes sets of analogous reactions (for each amino acid or peptide), we use generic reactions in the upper right part of the map. The ABC importer Opp catalyzes tetrapeptide uptake reactions in the model ([amino acid]4abc in Supplementary File 4); the AA permeases (incl. GltP) catalyze amino acid proton symport reactions ([amino acid]t2[p]r in Supplementary File 4). The peptidases catalyze peptide hydrolysis reactions ([amino acid]4P in Supplementary File 4). The aminoacyl tRNA synthetases ("aaRS's" in the map) catalyze charging of tRNAs ([amino acid]TRS in Supplementary File 4). Synthesis of Gln-tRNA$^{\mathrm{Gln}}$ requires transamidation of initially mischarged Glu-tRNA$^{\mathrm{Gln}}$ and the corresponding reactions are shown on the lower left. In the *S*-adenosylmethionine pathway on the lower right, we note that nucleic acid modification reactions (indicated by the edge labeled "DNA/RNA modification") were not included in the model due to lack of sufficient information on kind and abundance of nucleic acid modifications in JCVI-syn3A.

translation in the model uses charged tRNA that are produced from one synthetase for each amino acid, except for glutamine. Instead, glutamyl-tRNA(Gln) is transamidated to glutaminyl-tRNA(Gln) via an amidotransferase (*gatB*/0687 through *gatC*/0689) using glutamine or asparagine as amino donor. Translation is assumed to be initiated with formylated methionyl initiator tRNA (fMet-tRNA$^{\mathrm{fMet}}$). While the natural *M. mycoides capri* has been found to be able to initiate protein synthesis with the unformylated species [85], the folate cycle (including the Met-tRNA$^{\mathrm{fMet}}$ formyltransferase *fmt*/0390) is quasi-essential in JCVI-syn3A and the assumption is hence made that formylation cannot be omitted in JCVI-syn3A. Translation reactions for both apo-ACP and the generic protein species include deformylation and methionine cleavage, in line with ca. 80 % of proteins in a proteome assumed to have the initial methionine cleaved [101], and based on the alanine in second position of the apo-ACP sequence *acpA*/0621, favorable for methionine cleavage [101]. For PdhC (*pdhC*/0227) and dUTPase (*dut*/0447), only deformylation is considered, as the phenylalanine and isoleucine in second position of their sequences are not favorable for methionine cleavage [101]. The excess formate is assumed to be secreted by passive means.

Explicit protein and RNA degradation reactions are included in the model to account for protein and mRNA turnover in the cell and to regenerate the amino acid and nucleotide pools respectively. In line with most other mycoplasma genomes [102], the JCVI-syn3A genome only contains two AAA+ proteases, Lon (*lon*/0394) and (putatively) FtsH (*ftsH*/0039). Lon has been found to degrade ssrA-tagged peptides in place of ClpXP in *Mesoplasma florum* [102] and *M. pneumoniae* [103]. It is tentatively assumed to be the main protease for protein turnover in general in the metabolic reconstruction. An ATP expense of 225 ATP per protein of 385 residues is assumed (see Appendix 1).

Lower bounds on the protein and RNA degradation reactions of 3.5×10^{-4} mmol gDW^{-1} h^{-1} and 7.7×10^{-3} mmol gDW^{-1} h^{-1}, respectively, are imposed assuming similar degradation rates as in *M. pneumoniae* [25], which shares the protease Lon. Assuming that the entire proteome is subject to turnover (as observed in *M. pneumoniae* [24]),

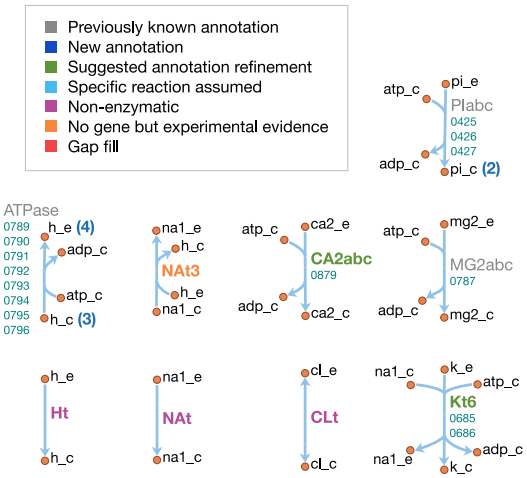


FIGURE 13 Ion transport reactions in JCVI-syn3A. Map components and labels as in Figure 5.

the protein degradation constraint corresponds to an average protein half life time of 25 h, which describes the time an average protein persists throughout cell divisions before being actively degraded; it is based on ^{13}C mass-spectrometry measurements of protein degradation in *M. pneumoniae* that corrected for protein dilution by cell growth [24]. The protein degradation constraint thus accounts for the ATP expense arising from the slow (compared to the cell cycle time) but non-zero active protein degradation. With an ATP consumption of 225 ATP per degraded protein, protein breakdown thus requires an ATP expense of 0.079 mmol gDW^{-1} h^{-1}. In addition, the imposed protein degradation flux implies an additional ATP cost due to the required additional protein synthesis (see Section 2.3.8). The RNA degradation constraint corresponds to an mRNA half-life time of around 1 min, assuming mRNA to account for ∼5 % of all cellular RNA [104, 105] (and assuming other RNA degradation to be negligible in the exponential growth phase). This is comparable to the mRNA half-life time of 2 min determined in *Mycoplasma gallisepticum* [106]. While RNA degradation itself is assumed to not consume ATP, it incurs an indirect ATP cost due to the required additional RNA synthesis, analogous to the situation for protein degradation.

Finally, we note the presence of methionine adenosyltransferase (*metK*/0432), which produces the methyl donor *S*-adenosylmethionine used by several nucleic acid methylation enzymes; these nucleic acid modifications are however not included in the model due to lack of sufficient information on kind and abundance of such modifications in JCVI-syn3A. The demethylation product *S*-adenosylhomocysteine is broken down by *S*-adenosylhomocysteine hydrolase (*mtnN*/0381) to homocysteine. As there is no obvious way for further breakdown of homocysteine in JCVI-syn3A, it is assumed to be secreted for the time being. As a precedent, secretion—rather than further conversion—of homocysteine has been suspected in *Plasmodium falciparum* [107]. As the magnitude of the required efflux rate is expected to be small, secretion is furthermore assumed to occur via passive permeation.

2.3.6 Ion uptake

The JCVI-syn3A annotation lists importers for magnesium (*mgtA*/0787) and phosphate (Pst system, *pstS*/0425 through *pstB*/0427). Danchin and Fang [83] have suggested the putative magnesium import gene *corA*/0879 could import calcium as well based on its similarity to the *B. subtilis* calcium importer YloB [108]. While the sequence identity of *corA*/0879 to YloB (CAB31439) is lower than between YloB and the deleted calcium importer MMSYN1_0246 (22 % vs. 30 %), we note that the putative calcium binding sites in YloB [108] are equally well conserved in both JCVI-syn3A genes, and include *corA*/0879 as the calcium importer. Active transport processes for Na^+ and K^+ are

inferred from biochemical evidence (see Appendix 1). Based on this, an ATP-consuming K^+/Na^+ antiport reaction is catalyzed by KtrAB (*natA*/0685 and *trkA*/0686) in the model, and an Na^+/H^+ antiport reaction is included as well without a specific gene assignment. Import systems have not been identified for sodium or chloride, but chloride has been proposed to freely permeate the membrane of *M. mycoides capri* [109] and sodium has been suggested to "leak" through membranes of *M. gallisepticum* [45]. Alternatively, leakage through some other transporter cannot be ruled out. Thus passive uptake reactions are assumed for sodium and chloride. Finally, a passive H^+ influx reaction is used, which is counteracted by the proton-extruding ATPase. Lacking more information, all ion uptake rates in the model are left unconstrained.

2.3.7 Damage reactions

On top of the enzymes involved in required metabolic reactions, we note the presence of possible damage reactions and repair enzymes to repair this damage. As discussed, 5,10-methenyl-THF is hydrolyzed to 5-formyl-THF as a side reaction of GlyA [86] and YgfA catalyzes the opposite direction. 5-formyl-THF is not only the only form of THF taken up in the model, but is also a potent inhibitor of folate-related enzymes [110], and as such YgfA is an important repair enzyme. To account for the known damage reaction of 5,10-methenyl-THF hydrolysis, a small lower bound ($0.01\ \mathrm{mmol\,gDW^{-1}\,h^{-1}}$) is imposed to ensure non-zero flux through this reaction. Reconversion of 5-formyl-THF to 5,10-methenyl-THF via YgfA consumes 1 ATP per 5-formyl-THF and thus requires $0.01\ \mathrm{mmol\,gDW^{-1}\,h^{-1}}$ ATP. Another damage preemption/repair enzyme included in the model is dUTPase (*dut*/0447), which hydrolyses dUTP and prevents its erroneous incorporation into DNA. Rather than enforcing a minimal flux through this reaction, dUTPase is included directly in the biomass for technical reasons (see Section 2.3.2).

2.3.8 GAM/NGAM

In order to account for cellular energy expenses, the growth- and non-growth-associated maintenance costs (GAM and NGAM) need to be included in the model. The GAM describes the energy cost associated with cellular growth, and therefore enters the biomass reaction as a certain amount of ATP spent per unit biomass production. The NGAM describes the basic, growth rate-independent cellular energy requirements and is therefore frequently implemented as a lower constraint on a separate ATP hydrolysis reaction [25, 46, 111]. Both parameters can be measured experimentally (e.g. from chemostat measurements [112]), however, to our knowledge no such measurements are available for any mycoplasma, and thus the cellular energy expenses must be estimated differently.

The GAM consists of a component that can be related to macromolecular synthesis energy costs and a non-quantifiable part. The macromolecular synthesis costs are outlined in Section 2.3.5; for the macromolecular composition of JCVI-syn3A, they yield a total cost of $21.54\ \mathrm{mmol\,gDW^{-1}}$ ATP. For the non-quantifiable part, $25\ \mathrm{mmol\,gDW^{-1}}$ ATP is used following the previously published *M. pneumoniae* model [25]. Together with the ATP costs for macromolecular synthesis, this yields a total GAM of $46.54\ \mathrm{mmol\,gDW^{-1}}$, which is comparable to the $53.95\ \mathrm{mmol\,gDW^{-1}}$ estimated for *E. coli* [46]. The use of these values is supported by the overall conservation of the gene expression apparatus.

The NGAM captures non-growth related energy expenses, such as macromolecular turnover and maintenance of the transmembrane pH gradient, which has been suggested to be a considerable energy expense in mycoplasmas [25]. These two energy expenses are included in the JCVI-syn3A model as an approximate NGAM. Rather than accounting for all NGAM expenses through a single ATP hydrolysis reaction, the total NGAM is distributed across several reactions (akin to the quantifiable GAM fraction). The total macromolecular turnover costs can be obtained from the assumed protein and RNA degradation fluxes ($3.5 \times 10^{-4}\ \mathrm{mmol\,gDW^{-1}\,h^{-1}}$ and $7.7 \times 10^{-3}\ \mathrm{mmol\,gDW^{-1}\,h^{-1}}$, respectively; see

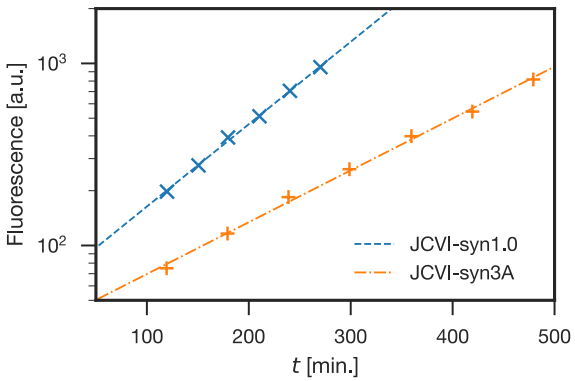


FIGURE 14 Comparison of growth curves of JCVI-syn1.0 and JCVI-syn3A. JCVI-syn1.0 has a doubling time of 66 min (blue; "×" markers), whereas JCVI-syn3A has a doubling time of 105 min (orange; "+" markers). Doubling times (t_d) were calculated as described in Methods, plotting fluorescence staining of cellular DNA vs. time, fitted by exponential regression curves. The regression curves for JCVI-syn1.0 and JCVI-syn3A have R^2 values of 0.9986 and 0.9976, respectively.

Section 2.3.5); the ATP costs in protein and RNA synthesis; and the associated costs for tRNA charging and nucleotide phosphorylation, respectively. Specifically, a constant part of the protein/RNA synthesis fluxes (and, preceding these, tRNA charging and nucleotide phosphorylation) is routed through protein/RNA degradation, and only the surplus beyond this constant flux actually contributes to model growth. These turnover-associated ATP costs amount to 2.73 mmol gDW^{-1} h^{-1} in the model for JCVI-syn3A. The maintenance of a transmembrane pH gradient in *M. mycoides capri* PG3 has been experimentally demonstrated [113, 114] and an approximate value for the H^+ extrusion rate by the transmembrane ATPase has been obtained [114]. The observed rate of ∼70 nmol/min/mg cellular protein corresponds to a proton flux of ∼2.30 mmol gDW^{-1} h^{-1} in JCVI-syn3A, which translates to an ATP consumption of 0.57 mmol gDW^{-1} h^{-1} (4 H^+/ATP). This value is hence implemented as a lower bound on the model ATPase reaction. The resulting total NGAM in the model for JCVI-syn3A is then 3.30 mmol gDW^{-1} h^{-1}, which comes out similar to the iJO1366 model for *E. coli* (3.15 mmol gDW^{-1} h^{-1}) [46].

2.4 Steady state fluxes

With the chosen parameters, the model yields an *in silico* growth rate of $\mu = 0.34\ h^{-1}$, corresponding to a doubling time of $t_d = 2.02$ h ($t_d = \ln(2)/\mu$); this comes close to the experimental doubling time of ca. 105 min (see Figure 14). This exact agreement is sensitive to the choices for uptake/secretion and GAM parameters however (see Sensitivity analysis in Appendix 1 and Appendix 1-Figure 2), and the *in silico* growth rate should thus be more understood as a provisional prediction. This does not constitute a problem for the subsequent analyses; when the impact of *in silico* gene knockouts on growth rate is studied in Section 2.6, nearly all knockouts either abolish the growth rate entirely (lethal knockouts) or do not affect it at all. Thus, this analysis is not affected by the rather qualitative nature of the growth rate prediction by the model.

In the rest of this section, the steady-state fluxes produced by the model are compared to literature data, and to theoretical flux limits obtained from protein abundances and enzyme turnover numbers (k_{cat}). While no fluxomics data is yet available for JCVI-syn3A, some experimental fluxes from the literature allow for a few comparisons. The purine incorporation flux into RNA has been determined for *M. mycoides capri* LC Y [66] as 0.5–1.0 nmol/min/mg cellular protein, corresponding to 0.016–0.033 mmol gDW^{-1} h^{-1}. This is close to the *in silico* net flux of ATP and

GTP into RNA of 0.047 and 0.043 mmol gDW^{-1} h^{-1}, respectively. Here, the net flux is defined as the difference between NTP consumed by the RNA polymerase reaction and NMP released by the RNA degradation reaction. The *in silico* K^+ uptake of 0.29 mmol gDW^{-1} h^{-1} also falls within a factor of two of the experimental uptake rate of 0.49 mmol gDW^{-1} h^{-1} (15 nmol/min/mg cellular protein) in *M. mycoides capri* PG3 [115]. These comparisons serve as an internal consistency check on the model, demonstrating that the *in silico* uptake/incorporation rates as resulting from biomass composition and growth rate indeed reproduce the experimental values.

Furthermore, it has been reported that *M. mycoides capri* LC Y spends 10 % of its glucose uptake on polysaccharide capsule production [38]. The *in silico* fluxes through central metabolism are depicted in Figure 6–Figure supplement 1. As can be seen, the *in silico* flux through phosphoglucomutase (PGMT, *deoB*/0733, leading from glucose-6-phosphate to galactan and Gal-DAG in Figures 6 and 10) is lower than the experimental value, amounting to only 1.8 % (0.135 mmol gDW^{-1} h^{-1}, from 7.4 mmol gDW^{-1} h^{-1} glucose taken up). However, the model does qualitatively reproduce the further splitting between galactan and Gal-DAG production, which gives a ratio of ∼22:1 (UDP-galactofuranose consumption of 0.129 mmol gDW^{-1} h^{-1} vs. 0.006 mmol gDW^{-1} h^{-1}), compared to a ratio of ∼64:1 from ^{3}H labeling [38]. Thus, with the chosen parameters, the model reproduces several experimental fluxes; with the only significant difference occurring in capsule production.

FBA flux predictions were also compared to reaction flux bounds (V_{max}) calculated from protein abundances and enzyme turnover numbers (k_{cat}) [116, 117]. The protein abundances were derived from proteomics experiments (see Section 2.7) and the turnover numbers were extracted from the BRENDA database [118] (see Figure 15, Figure 15–Figure supplement 1 and Appendix 1). V_{max} values could be obtained for 105 "non-pseudo" reactions (i.e. excluding exchange, biomass and macromolecular reactions). Of these, 86 had non-zero fluxes. The zero-flux reactions include for example reactions pertaining to alternative sugars, which are unused in the assumed medium. Of the reactions with non-zero fluxes, only 19 reactions required fluxes in the FBA optimal solution higher than their proteomics-derived V_{max} (see Figure 15–Figure supplement 1A). The reaction with the lowest V_{max}/flux ratio is adenylate kinase, which is predicted to carry a flux of 2.23 mmol gDW^{-1} h^{-1}, compared to a proteomics-derived V_{max} of only 0.01 mmol gDW^{-1} h^{-1}. However, the k_{cat} for this enzyme as found for *B. subtilis* in BRENDA is 0.053 s^{-1}, which falls in the lower 10th percentile of the k_{cat} data for all reactions in the model. The second-smallest V_{max} to flux ratio is found for aspartate-tRNA synthetase (0.11 mmol gDW^{-1} h^{-1} in model vs. 0.005 mmol gDW^{-1} h^{-1} from proteomics/k_{cat}); other amino acyl-tRNA synthetases with significantly low V_{max}/flux ratio (< 0.25) are the ones for threonine and serine. These three amino acyl-tRNA synthetases have the smallest k_{cat} numbers among all amino acyl-tRNA synthetases. The third-lowest V_{max} to flux ratio is found for fructose bisphosphate aldolase (7.21 mmol gDW^{-1} h^{-1} in model, 0.39 mmol gDW^{-1} h^{-1} from proteomics/k_{cat}). This protein has 227 copies in the cell on average, which places it among the least abundant proteins in central metabolism. Furthermore, the k_{cat} value for this enzyme retrieved from BRENDA for *Bacillus cereus* is 2.95 s^{-1}, which is also one of the lowest values found among proteins in central metabolism. The only other reactions with a V_{max}/flux ratio less than 0.25 are: DASYN (*cdsA*/0304), which produces the lipid intermediate CDP-diacylglycerol; ACPS (*acpS*/0513), which attaches the 4′-phosphopantetheine to apo-ACP; and GUAPRT (*hptA*/0216), which produces GMP from guanine. DASYN has a V_{max} to flux ratio of 0.19, i.e. only slightly decreased. ACPS has both the lowest k_{cat} among the model enzymes (0.001 s^{-1}) and one of the lowest protein abundances (just one copy per cell per our proteomics data). If either value turned out to be not accurate, this could easily raise the V_{max}/flux ratio above the current level of 0.07. However, the discrepancy observed for GUAPRT is interesting in the light of the aforementioned mononucleotide uptake capabilities in *M. mycoides capri* [63, 64] (see Section 2.3.2). While there is no other evidence for the possible conservation of this capability in JCVI-syn3A, this flux bound might suggest that this uptake capability still at least partially exists in JCVI-syn3A, and might be worth investigating. All other reactions with V_{max} lower than the FBA flux differ by less than a factor of four; even though the FBA flux exceeds the estimated V_{max}, the disagreement

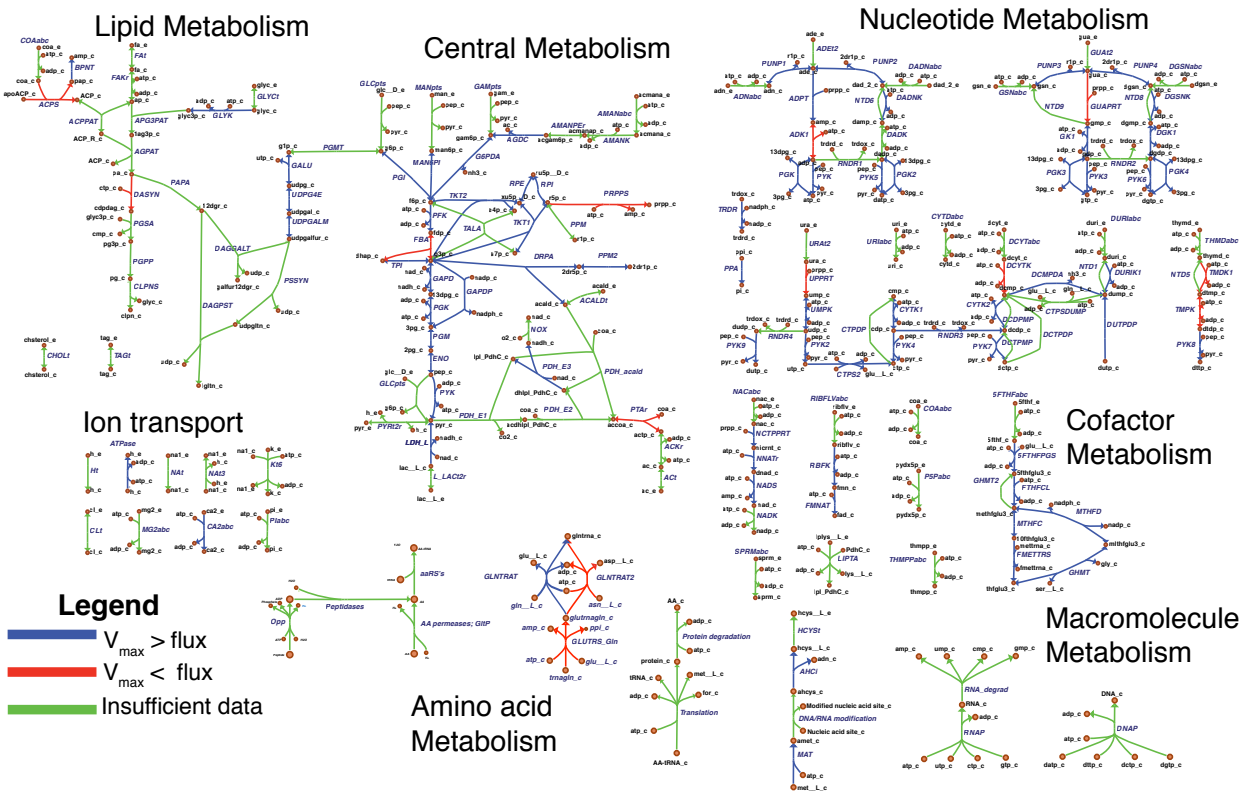


FIGURE 15 Comparison of FBA steady-state fluxes v to maximal fluxes $V_{\max}$ obtained from protein abundances and turnover numbers from BRENDA and the literature. Map components and labels as in Figure 5, with reaction highlighting and gene loci/gene-protein-reaction rules omitted. Each edge is colored according to the ratio between $V_{\max}$ and v: Blue indicates $V_{\max} > v$, red indicates $V_{\max} < v$ and green indicates that no $V_{\max}$ could be obtained (because of either missing turnover number or missing protein abundance; or because reaction is not enzymatic to begin with).
Figure supplement 1 Statistics of FBA steady-state fluxes v vs. maximal fluxes $V_{\max}$ comparison (see Figure 15). A: Summary of $V_{\max}$ vs. v comparison over all 253 non-exchange reactions in the model. Red, blue, green: Meaning as in Figure 15. Green-striped: Subset of green set—reactions without $V_{\max}$ that pertain to transport, which usually do not have an EC number associated with them. Gray: Reactions with $v = 0$ in the FBA solution (thus $V_{\max} > v$ always fulfilled). B: Histogram of $V_{\max}/v$ over the blue and red subset in panel A.

is rather modest. For the reactions that do show higher disagreement, we note that the k_{cat} values obtained tend to be on the lower end either within the respective group of reactions, or across the model. This suggests that these k_{cat} values might merit closer investigation. Overall, the proteomics-derived bounds are consistent with the FBA fluxes, with only a handful of reactions showing significant discrepancies.

It should be noted that the rates of enzyme-catalyzed reactions *in vivo* are typically less than $V_{\max}$ to allow the cell to respond to increases in substrate concentration. Accordingly, $V_{\max}$ is expected to be greater than the metabolic flux necessary to sustain the cell, such that the flux required under typical growth conditions can be achieved without enzyme saturation. In line with this argument, a histogram of $V_{\max}$ values for reactions in the model shows the bulk of reactions to have a $V_{\max}$ 1–3 orders of magnitudes higher than the flux required in the FBA solution (see Figure 15–Figure supplement 1B).

2.5 Energy usage

Table 3 breaks down the energy consumption in JCVI-syn3A (as percentage of total ATP consumption, see Methods). The upper five categories correspond to individual subsystems of the metabolic model. The lower five categories provide a breakdown of GAM and NGAM expenses into individual components. As discussed in Section 2.3.8, a part

TABLE 3 Cellular ATP expenses by category (in percent of total ATP consumption).

Category	ATP expense [%]
Nucleotide metabolism	3.6
Pentose phosphate pathway	1.7
Lipid metabolism	0.7
Cofactor metabolism	0.1
Transport	3.4
$GAM_{Macromolecules}$	18
$GAM_{tRNA\ charging}$	16
$GAM_{Nonquant}$	41
$NGAM_{Turnover}$	13
$NGAM_{ATPase}$	2.7

of the protein and RNA synthesis (and, by extension, the tRNA charging and nucleotide phosphorylation fluxes) is routed through protein and RNA degradation, constituting the turnover-associated part of the NGAM; the resulting fraction of total ATP consumption is listed as "$NGAM_{Turnover}$" in Table 3. Analogously, "$NGAM_{ATPase}$" denotes the ATP expense for the ATPase-associated part of the NGAM. "$GAM_{Macromolecules}$" and "$GAM_{tRNA\ charging}$" denote the ATP expenses for growth-associated protein/RNA synthesis (subsystem "Macromolecules" in the model) and tRNA charging (subsystem "Amino acid metabolism"), respectively. Analogously, "Nucleotide metabolism" only includes ATP expenses beyond RNA turnover (i.e. NTP production for growth-associated nucleic acid synthesis and nucleotide usage in other subsystems). "$GAM_{Nonquant}$" denotes the non-quantifiable fraction of the GAM.

In line with JCVI-syn3A relying heavily on uptake of pre-formed precursors and further conversion through salvage pathways only, it spends only ∼6 % of ATP on (small molecule) metabolic processes (i.e. lipids, cofactors and nucleotides, plus PRPP synthesis in the pentose phosphate pathway). The vast majority of energy (75 %) is spent directly on growth, i.e. macromolecular synthesis and tRNA charging and the non-quantifiable contribution to GAM. A modest fraction of ∼16 % of cellular energy expenses falls to the NGAM (macromolecular turnover and ATPase). These numbers stand in striking difference to *M. pneumoniae*, for which non-growth associated maintenance accounts for 71–88 % of total cellular ATP consumption (in the accounting for *M. pneumoniae*, the NGAM does not include protein/RNA turnover) [25]. This correlates with *M. pneumoniae* having a doubling time between 8 and 20 h [23, 25], i.e. four to ten times slower than JCVI-syn3A.

The ATP breakdown also reveals that in spite of the minimal cell's heavy reliance on uptake of pre-formed precursors, transport processes only account for ∼3 % of ATP consumption. While the optimal FBA solution only takes up amino acids through the permeases (0876, 0878, and *gltP*/0886) using proton symport reactions, the ATP expense on transport does not increase significantly (only to ∼5 %) when forcing amino acid uptake through the ATP-consuming Opp peptide importer. This illustrates how JCVI-syn3A can maintain a relatively fast growth rate in spite of its extreme genome minimization and reliance on fermentative ATP production: By importing pre-formed precursors or recovering them through salvage reactions, the cell expends a minimal amount of energy to obtain the final macromolecular precursors and passes this savings in energy along to the production of biomass.

The other important currency in the cell are reduction equivalents in the form of NADPH, which in JCVI-syn3A is produced by GapN (GAPDP, *gapdh*/0451) and, in tiny amounts, by FolD (MTHFD, *folD*/0684). The only consumer of NADPH in the model is ribodinucleotide reductase (RNDR, *nrdE*/0771 through *nrdF*/0773). *In vivo* however, NADPH is expected to also be needed for expenses not captured by the model, including RNA modification (dihydrouridine synthesis) and response to oxidative stress: The reduction of protein disulfide bonds formed by oxidative stress is mediated by thioredoxin [119], and coenzyme A disulfide reductase (*cdr*/0887) serves to reduce coenzyme A disulfide

dimers to the free thiol-carrying monomers. NADPH production through GAPDP diverts flux from the ATP-producing GAPD/PGK branch in glycolysis, effectively incurring an ATP cost for NADPH production. In order to probe the cellular capacity for NADPH production, Appendix 1-Figure 2G shows a plot of *in silico* doubling time as a function of imposed NADPH consumption (imposed via an artificial NAPDH oxidation reaction with O_2, introduced for testing purposes). Within a considerable margin, the doubling time rises shallowly with NADPH consumption: E.g., at 3.5 mmol $\mathrm{gDW}^{-1}\,\mathrm{h}^{-1}$ (a quarter of the maximally possible flux through GAPDP), the model doubling time only rises by 25 % to ∼2.5 h. This suggests that even though NADPH usage is not fully captured by the model, the cell should be able to accommodate a considerable amount of NADPH demand without strong impact on the growth rate (see also Sensitivity analysis in the Appendix 1).

Finally, there is also some experimental information that allows for a comparison of cellular energetics, specifically of basal energy expenses. In Benyoucef et al. [114], the residual acid secretion in *M. mycoides capri* PG3 in a saline buffer after inhibition of ATPase has been measured to be around 110 nmol/min/mg cellular protein (corresponding to ∼3.6 mmol $\mathrm{gDW}^{-1}\,\mathrm{h}^{-1}$), which can be compared to the corresponding *in silico* acid secretion (which in turn is connected to ATP production). The measurements were performed in a saline buffer containing glucose but no other nutrients for growth. Under these conditions the cell is not able to grow [42], but should be able to meet its basic energetic needs. Furthermore, since ATPase was inhibited with N,N′-dicyclohexylcarbodiimide (which abolishes both proton transduction and ATPase activity [120]), it should not consume ATP anymore under the experimental conditions. These conditions are simulated by setting the lower bound on ATPase proton extrusion to 0.0 mmol $\mathrm{gDW}^{-1}\,\mathrm{h}^{-1}$ and changing the objective function in FBA from maximal growth rate to minimal glucose uptake. A residual acid secretion of 1.3–2.6 mmol $\mathrm{gDW}^{-1}\,\mathrm{h}^{-1}$ results, which depends on the assumed lactate to acetate ratio, and falls within a factor of ∼2 of the experimental value. This suggests that the basic cellular energy expenses—protein degradation, RNA and protein synthesis under non-growth conditions—are described reasonably well by this model. At the same time, hypotheses can be made as to what energy expenses could account for the observed remaining discrepancy. One expected factor is the unknown actual NADPH demand (and resulting effective ATP cost). In addition, a possibly significant energy sink not covered yet by the model are metabolite repair functions, of which thus far only two are included in the model, namely 5-formyl-THF cyclo-ligase (*ygfA*/0443) and dUTPase (*dut*/0447). Metabolite repair usually consumes energy [121], and it would be interesting to see to what extent this could account for the current underestimation of basal energy expenses.

2.6 *In silico* gene knockouts and mapping to *in vivo* essentiality

In addition to studying fluxes of the unperturbed model, the FBA framework also allows to study the impact of *in silico* gene disruptions by simulating knockouts in COBRApy [122], i.e. by removing all reactions associated with a gene of interest from the model and calculating the growth rate from the resulting model. A knockout is defined to be lethal if the resulting growth rate is zero or the FBA problem becomes infeasible. By this definition, 123 of the 155 genes included in the model are essential (79%). In this analysis, two genes are currently non-essential *in silico* for "technical" reasons: *metK*/0432 (methionine adenosyltransferase, MAT) and *mtnN*/0381 (*S*-adenosylhomocysteine (SAH) hydrolase, reaction ID: AHCi). These genes are part of the *S*-adenosylmethionine (SAM) pathway and would be connected through nucleic acid methylation reactions (consuming SAM and producing SAH), which were not included in the model due to missing experimental information. As these two reactions currently cannot carry flux, it does not make sense to consider their *in silico* essentiality in the comparison to experiment.

An individual breakdown of *in silico* gene essentialities is provided in Table 2, which lists all genes modeled *in silico*, together with their catalyzed reaction (or general description for genes with several reactions, like the

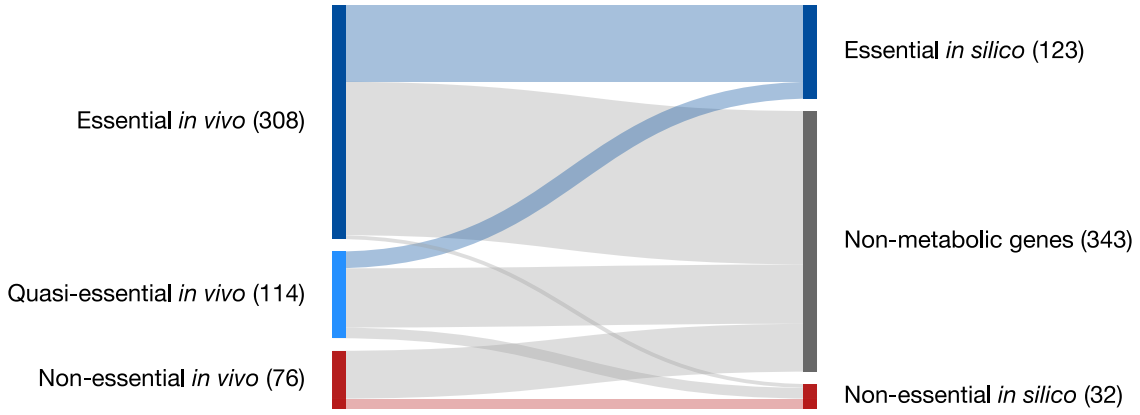


FIGURE 16 Partitioning of genes classified as essential, quasi-essential, and non-essential by transposon mutagenesis experiments into those which are *in silico* essential, *in silico* non-essential, and not modeled ("Non-metabolic"). All genes are included (i.e., also RNA genes and pseudogenes).
Figure supplement 1 *In silico* double-gene knockouts between genes that are non-essential in single-gene knockouts. Among the individually non-essential genes, a double knockout of the gene pair (0876, 0878) is the only lethal combination (red). This knockout corresponds to simultaneously removing both amino acid permeases, thus preventing cysteine uptake. Simultaneous knockout of the glutamate/aspartate permease *gltP*/0886 and any Opp gene (*oppB*/0165 through *oppA*/0169) is non-lethal *in silico*, as the model will under these circumstances produce glutamate through the hypothesized dUMP breakdown reaction CTPSDUMP and, to a lesser extent, through the reaction CTPS2 (both catalyzed by *pyrG*/0129). Glutamate production through *pyrG*/0129 is not expected to be able to meet cellular demands *in vivo*. If flux through CTPSDUMP is set to zero in the model, a double knockout of *gltP*/0886 and Opp becomes lethal *in silico*.

TABLE 4 Confusion matrices for gene essentiality prediction. "All genes" denotes agreement/disagreement between the model prediction and the transposon mutagenesis experiment considering all genes in the metabolic reconstruction (excluding the two "technical non-essential" genes). "Excluding AA" repeats the same comparison as "All genes", with genes for amino acid utilization (uptake and peptidase genes) excluded. See Table 2 for individual gene essentialities *in silico* and *in vivo*.

	All genes		Excluding AA	
Exp. / Model	Essential	Non-essential	Essential	Non-essential
Essential	101	4	101	4
Quasi-essential	22	14	22	4
Non-essential	0	12	0	10

peptidases), and their essentiality *in silico* and *in vivo*. Genes non-essential *in silico* are found in amino acid, central and nucleotide metabolism as well as transport reactions, and only one gene in lipid metabolism. Some non-essentialities are functionally connected. For example, as the peptide importer Opp (*oppB*/0165 through *oppA*/0169) is non-essential *in silico*, the downstream peptidases have to be non-essential as well. A further analysis of *in silico* essentialities is featured in Sections 3.2–3.4.

These *in silico* essentialities can be compared to the experimental transposon data (see Section 2.1). Figure 16 shows an overall mapping between *in silico* and *in vivo* essentiality (including all genes in JCVI-syn3A, including RNA genes, pseudogenes and the two "technical non-essentials"). A more detailed analysis for the genes included in the model is presented in Table 4, which displays the confusion matrix for the *in silico* to *in vivo* comparison, i.e. the distribution of model genes among the *in silico* and *in vivo* classifications. The left table represents the breakdown for all model genes except the two technical non-essentials, while the right table shows the breakdown if genes related to amino acid utilization are also excluded (see below). Whilst the model only distinguishes essential and non-essential genes, the experimental classification includes quasi-essentiality, which falls somewhere in between essentiality and non-essentiality. Thus, for any evaluation of predictive performance of the model, some assumption has to be made with regard to the *in vivo* quasi-essential genes. Table 5 summarizes several statistics obtained for specific cases, discussed in the following.

TABLE 5 Accuracy, sensitivity, specificity and Matthews correlation coefficient calculated for several scenarios. QE as E: Treating *in vivo* quasi-essentials as essentials; QE as NE: Treating quasi-essentials as non-essentials; No QE genes: Excluding all genes quasi-essential *in vivo*; QE as E, no AA genes: Excluding genes related to amino acid utilization, and treating quasi-essentials as essentials; QE as NE, no AA genes: Excluding genes related to amino acid utilization, and treating quasi-essentials as non-essentials.

	Accuracy	Sensitivity	Specificity	MCC
QE as E	88 %	87 %	100 %	0.59
QE as NE	83 %	96 %	54 %	0.59
No QE genes	97 %	96 %	100 %	0.85
QE as E, no AA genes	94 %	94 %	100 %	0.72
QE as NE, no AA genes	82 %	96 %	39 %	0.46

Two limiting cases of interest are treating all quasi-essential genes as either *in vivo* essential (1) or *in vivo* non-essential (2). Given that the identification of quasi-essential genes was crucial for the successful genome minimization in JCVI-syn3.0 [12], treating these genes as essential might be the biologically more relevant assumption. If all quasi-essential genes are considered essential (i.e. adding the numbers in the second row in Table 4, left matrix to the first row), the model displays an accuracy of 88 %. (Accuracy = (TP + TN)/total; we opt to define essential genes as "positive" and non-essential genes as "negative", so that a true positive gene (TP) is essential in model and experiment; a false positive (FP) is essential in the model but non-essential in experiment; a true negative (TN) is non-essential in model and experiment; and a false negative (FN) is non-essential in the model but essential in experiment. "Total" is the sum of all genes included in the analysis.) The resulting sensitivity (TP/(TP + FN)) is 87 %, while the specificity (TN/(TN + FP)) is 100 %: All *in silico* essential genes are at least quasi-essential *in vivo*, so there are no "strong" false positive predictions (of genes to be essential that are actually non-essential *in vivo*). If, alternatively, all quasi-essential genes are considered non-essential *in vivo* (adding the numbers in the second row to the third row in the left confusion matrix in Table 4), the accuracy comes out a bit lower at 83 %; the sensitivity increases to 96 % while the specificity drops to 54 %. This low specificity can be explained by considering the comparatively low number of *in vivo* non-essentials among the genes included in the model (12): Considering all quasi-essentials (two thirds of which are essential in the model) to be non-essential as well then leads to a large relative fraction of "non-essentials" not detected by the model, even though the overall accuracy does not change much compared to case (1). As a more balanced measure of model prediction performance, Table 5 also features the Matthews correlation coefficient (MCC) in the last column, which can range from -1.0 (perfect disagreement) via 0.0 (same agreement as a random model) to 1.0 (perfect agreement). For both cases described above (treating quasi-essentials as either all essential or all non-essential), the MCC comes out to $\sim$0.59.

While this does not amount to perfect agreement, we note that the quasi-essentials in the middle row in Table 4 (upper confusion matrix) actually encompass the vast majority of false model predictions. Thus, in addition to the two limiting cases presented above, it is also instructive to consider the prediction performance when including only those genes that can be classified as essential or non-essential *in vivo*, i.e. those genes that can be compared to the model classification without further assumptions. In this case, the specificity reaches 100 % as in case (1) above, as there are again no false positives; the sensitivity reaches the same value as in case (2) above (96 %) as there are only 4 false negatives; and the accuracy increases to 97 % in this case. The MCC also comes out higher at 0.85. This demonstrates that the lower MCC and other metrics obtained before really arise from the large number of quasi-essential genes included in the model, that are inherently difficult to describe in an FBA model: For example, nucleic acid stabilization by polyamines is a known essential process, and the minimal media for both *M. mycoides capri* LC Y [41] and *M. pneumoniae* [23] hence include spermine, which is thus a biomass component in the model for JCVI-syn3A. While this renders the corresponding uptake genes (*potC*/0195 through *potA*/0197) essential in the model, they are only

quasi-essential *in vivo* (see Table 2).

Similarly, it is of interest to consider one set of genes whose functionality is difficult to capture precisely based on the currently available information, namely the genes pertaining to uptake and utilization of amino acids (in free or peptide form): As can be seen in Table 2, from the overall 14 "weak" false negative predictions (*in silico* non-essential genes that are quasi-essential *in vivo*), 10 comprise the peptide importer Opp (*oppB*/0165 through *oppA*/0169), two amino acid permeases (0878 and *gltP*/0886) and three of the four peptidases (0305, 0444 and 0479). As further discussed in Section 3.2, the *in vivo* essentiality of these genes is likely affected by their exact substrate profiles and maximal uptake rates. If all 12 genes related to amino acid utilization (i.e. the genes above plus *ietS*/0133 and 0876) are excluded from the prediction comparison, the right confusion matrix in Table 4 is obtained. The resulting metrics are listed in the last two rows of Table 5, where the remaining quasi-essentials are again included in the *in vivo* essentials (upper row) or in the non-essentials (lower row). As can be seen, in the first case, the accuracy and sensitivity both rise to 94 % compared to the full set of genes (88 % and 87 %, row 1 in Table 5); the MCC rises to 0.72. In the second case, the specificity drops from 54 % to 39 % compared to the full set of genes (row 2 in Table 5), and the MCC decreases to 0.46. However, this must be seen in the light of the excluded genes comprising mainly weak false negatives, i.e. quasi-essential genes that are non-essential *in silico*, and no weak false positives (quasi-essentials that are essential *in silico*). Thus, even though genes are excluded that show disagreement between model and experiment, the agreement worsens because these genes happened to be classified as "true negatives" in case (2). The improved model metrics in case (1) for excluding amino acid genes thus seem more relevant.

In summary, this analysis demonstrates an overall good agreement between model and experiment, which is mainly impacted by the *in vivo* quasi-essential genes, whose essentiality is inherently difficult to capture in an FBA model. The disagreements observed (quasi-essential genes, and a few strong false negatives) are discussed in detail in Sections 3.2–3.4. Some of them can be rationalized, while others lead to new hypotheses.

Finally, performing *in silico* double knockouts (Figure 16–Figure supplement 1) yields just one synthetic lethality (i.e. lethality of a two-gene knockout where the individual knockouts are non-lethal)–namely, a double knockout of the two amino acid permeases 0876 and 0878, which prevents the cell from acquiring cysteine.

2.7 Abundances of essential and non-essential proteins

Absolute cellular abundances (molecules per average cell) of JCVI-syn3A proteins were obtained from mass spectrometry based proteomics and the assumed protein dry mass fraction. Relative and absolute protein abundances were used in the reconstruction of the JCVI-syn3A biomass composition (see Section 2.2) and estimates of the $V_{\max}$ for reactions in the metabolic model. They also served for the further study of the JCVI-syn3A proteome, both with respect to the fraction of proteins with known functions, and in regard to expression of essential vs. non-essential proteins.

Comparing the overall JCVI-syn3A proteomics breakdown in Figure 17a to the genome breakdown in Figure 1 shows that the "Unclear" fraction is even smaller in the proteome than in the genome, suggesting that at least a generic function can be immediately assigned to >90 % of the proteome. Furthermore, proteins classified as "Metabolism" alone account for ∼25 % of the proteome. Considering all proteins included in the FBA model (i.e. also the synthetases classified as "Genetic Information Processing") covers a subset of 40 % of the proteome. Thus, studying expression features for genes in the model should yield relevant insights into the proteome as a whole.

Figure 17b compares distributions of absolute protein abundances between *in silico* essential, *in silico* non-essential and all proteins. Figure 17c shows the same comparison based on the transposon mutagenesis classification of essentiality (also including quasi-essential genes). As can be seen, the expression profiles for essential and non-essential

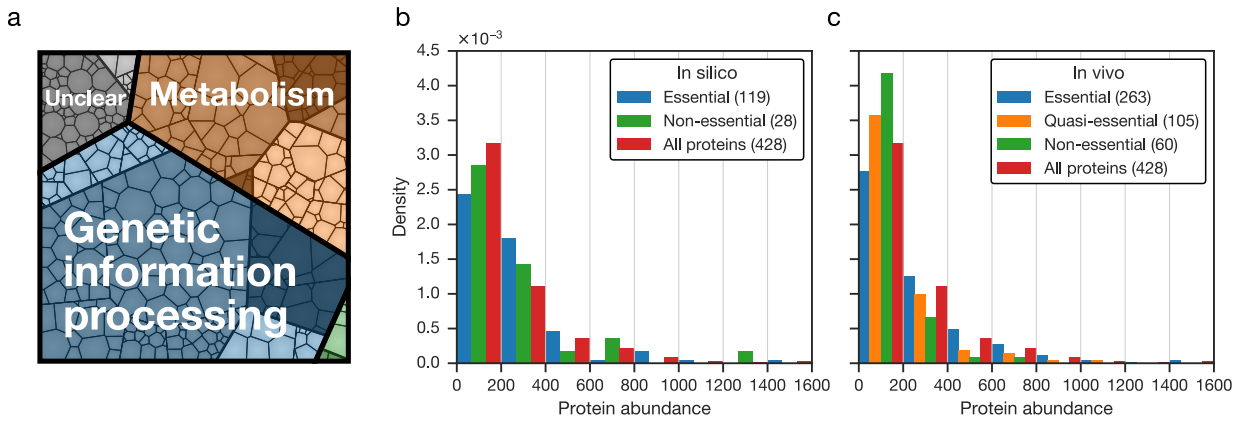


FIGURE 17 Distributions of absolute protein abundances (number of molecules per average cell) in JCVI-syn3A. (a) Breakdown of the JCVI-syn3A proteome into functional classes. The area of each cell is proportional to its relative abundance. (b) Absolute abundances of model-included metabolic proteins essential or non-essential *in silico* compared to all protein abundances. "Technical non-essential" proteins are not included (see Section 2.6). (c) Absolute abundances for proteins classified by *in vivo* essentiality from transposon mutagenesis experiments.
Figure supplement 1 Exceedence plots of abundances for proteins classified by *in silico* or *in vivo* essentiality. The exceedence at a given protein abundance value x is the fraction of the protein set displaying an abundance higher than x. Left panel: Model-included proteins (classified by *in silico* essentiality) compared to all proteins. Right panel: Proteins classified by *in vivo* (transposon-based) essentiality.

proteins are qualitatively similar both to each other and to the expression profile of all proteins in JCVI-syn3A. This holds for both the genome-wide transposon data-based comparison (Figure 17c) and the comparison for the subset of (mostly) metabolic FBA genes (Figure 17b). While this does not yet allow for strong conclusions, it does suggest the presence of little regulation, if at all, that would discriminate gene products based on their essentiality. This conclusion would be in line with the small number of identified regulatory proteins left in the genome of JCVI-syn3A.

3 Discussion

3.1 Completeness of the model

The creation of the first minimal bacterial cell JCVI-syn3.0 in 2016 provided a powerful platform for understanding the basics of life. As a first step along this road, we have combined the genetic information of JCVI-syn3A with the extensive amount of experimental information available for the natural *M. mycoides capri* and assembled a metabolic reconstruction and FBA model for the minimal cell. The majority of reactions in this model are supported by experimental evidence on the parent organism and related mycoplasmas. The model is near-complete with regard to accounting for the biomass components, describes cellular energy expenses well, shows good agreement with experimental transposon insertion data, and importantly has relatively few non-essential metabolic genes. It thus provides a foundation to study the features of the minimal metabolic network.

The metabolic networks of lipids and cofactors are both functionally nearly minimal and in their reconstruction nearly complete. The reconstructed lipid network is consistent with all membrane components known from the biomass composition (save for the small fraction of triacylglycerol, which might or might not still be produced in JCVI-syn3A) and contains no redundant features (except for one more glycosyltransferase than required by the current reconstruction). The only remaining gap in lipid metabolism is the missing gene for phosphatidate phosphatase. In cofactor metabolism, the remaining questions are the substrate specificities of the EcfS transporter subunits, and the proposed lipoate uptake mechanism. Amongst the ion transport reactions, a gene for the Na^+/H^+ antiporter remains to be identified. Central and

TABLE 6 Summary of features of the metabolic model for JCVI-syn3A. "Non-pseudo" reactions exclude exchange, biomass and macromolecule reactions. "Annotation-supported" includes all non-pseudo reactions that have a gene assigned. "Passive" reactions are transport processes assumed to take place without protein mediation. The meaning of "technical" non-essential genes is explained in Section 2.6.

Model overview	Genes	155	
	Genome coverage	31 %	
	Metabolites	304	
	Reactions (total)	338	
	Reactions (non-pseudo)	244	
Reaction breakdown (% of non-pseudo)	Annotation-supported	209	86 %
	Passive	14	6 %
	Gap fills with exp. evidence	17	7 %
	Gap fills without exp. evidence	4	2 %
	Supported (annotation/exp./passive)	240	98 %
Essentiality *in silico*	Essential genes	123	79 %
	Non-essential genes	30	19 %
	"Technical" non-essential genes	2	1 %
Essentiality *in vivo*	Essential genes	308	62 %
	Quasi-essential genes	114	23 %
	Non-essential genes	76	15 %

nucleotide metabolism display a number of potential redundancies (see Table 2), and several important reactions not accounted for by a gene yet. In central metabolism, these include the *in silico* essential transaldolase (TALA) reaction; and the reactions PDH_E1, NOX and export of lactate and acetate, all of which are required in the model to maintain the experimental doubling time of ∼2 h. In nucleotide metabolism, nucleobase uptake is an essential function still unaccounted for.

Even so, we obtain a number of gap-filled reactions of only 21—a fraction of 6 % of all model reactions, or 9 % of the 244 "non-pseudo" reactions. (Non-pseudo reactions are the subset of individual chemical or transport reactions in the model. This includes all model subsystems except for the artificial exchange and biomass reactions; and the macromolecule reactions, which describe non-metabolic processes taking place in all cells that are therefore not relevant for the number of gap fills.) These 21 gap fills are obtained from the total number of 35 non-pseudo reactions without assigned gene after subtraction of 14 passive transport reactions assumed to take place without protein mediation (see e.g. discussion on passive permeative glycerol uptake in Appendix 1). This number of gap fills is considerably lower than in comparable models (see Section 3.7). Furthermore, from these 21 gap fills, only 4 are not supported by experimental evidence. Hence, 98 % of all non-pseudo reactions are justified through gene assignments and/or experimental evidence, or are assumed to be passive. Table 6 summarizes the overall features of the model.

We also note that there are good candidates for many of the missing functions: The NOX reaction could conceivably be carried out by an oxidoreductase of unspecified function. Both 0029 and *fre*/0302 code for putative oxidoreductases. The gene *fre*/0302 in particular has been suggested to be the missing NADH oxidase [83] and might be a candidate for investigation. While no gene for transaldolase has thus far been identified in any mycoplasma, the alternative route proposed in Vanyushkina et al. [123] (see Section 2.3.1) would just require a phosphatase reaction, which could plausibly be carried out by one of a number of hydrolases in JCVI-syn3A of thus far unknown function. The same holds for further phosphatase reactions, including the phosphatidate phosphatase (PAPA) reaction in lipid metabolism and a number of hydrolase reactions in nucleotide metabolism. Substrate screening, informed by the metabolic reconstruction, might therefore be of interest for the hydrolases of unknown function. Finally, some of the reactions without assigned gene are transport processes (e.g. lactate/acetate export, nucleobase uptake). It stands to reason that these processes might be carried out by some of the many membrane proteins in JCVI-syn3A whose

function could not be identified yet.

Our metabolic model of JCVI-syn3A thus features an overall quite complete metabolic network, and even though a small percentage of reactions could currently not yet be assigned to a gene, the presence of genes catalyzing these reactions is plausible and the majority of these reactions are supported by experimental evidence. We therefore believe that comparing our model to the experimental transposon mutagenesis data is informative. In the following we discuss the comparison of *in silico* and *in vivo* (transposon mutagenesis-based) essentiality. While a number of genes can be discussed individually (Section 3.2), two pathways need to be discussed as a whole: the folate cycle (Section 3.3) and the pentose phosphate pathway (Section 3.4). Overall, the analysis suggests a few new hypotheses and even yields suggestions for some genes or groups of genes that could still be removed from the genome of JCVI-syn3A to minimize the genome even further. In this way, it complements the transposon mutagenesis data that can only probe individual essentialities—and simultaneous knockouts prove challenging in experiment (see Section 2.1).

3.2 Interpretation of individual gene essentialities

The most remarkable observation in comparing individual gene essentialities are three *in silico* non-essential genes whose *in vivo* (quasi-)essentiality is challenging to rationalize, and is therefore mysterious. These are the last two enzymes of the acetate branch (phosphate acetyltransferase and acetate kinase, *pta*/0229 and *ackA*/0230) in central metabolism and dCMP deaminase (*dctD*/0515) in nucleotide metabolism. The former two are essential *in vivo* while the remaining pyruvate dehydrogenase subunits (*pdhC*/0227 and *pdhD*/0228) are not. The non-essentiality of *pdhC/D* implies that JCVI-syn3A can grow without acetate fermentation, as also predicted by the model. This however raises the question what other essential function phosphate acetyltransferase and acetate kinase perform. A conceivable possibility could be that knockout of *pta*/0229 would lead to sequestration of coenzyme A as acetyl-coenzyme A if pyruvate dehydrogenase itself was still active. This could impede any function that coenzyme A could have as a cellular redox buffer. Similarly, knocking out *ackA*/0230 with the remainder of the pathway still active could lead to accumulation of acetyl phosphate, whose capacity as a nonenzymatic protein acetylation agent [124] might become deleterious for excess concentrations. dCMP deaminase converts dCMP to dUMP, which in the apparent absence of pyrimidine nucleoside phosphorylase is a dead end for which we hypothesize back-conversion to dCMP (see Section 2.3.2). Thus, dCMP deaminase neither provides a relevant nucleotide nor does it seem to contribute to balancing cellular nucleotide pools; its quasi-essentiality therefore completely eludes rationalization. For all of these three enzymes, it might be interesting to reduce their expression and study the impact on the metabolome.

A number of other discrepancies suggest specific new biological hypotheses. “Weak” false negatives (i.e. genes non-essential *in silico* but quasi-essential *in vivo*) occur in central, nucleotide, lipid and amino acid metabolism. In central metabolism, the assumed N-acetylglucosamine-6-phosphate deacetylase, 0493, is the only *in vivo* quasi-essential gene along the entire N-acetylmannosamine utilization pathway (with all other genes being non-essential *in vivo*), which strongly suggests some other (or additional) function for this gene. Lactate dehydrogenase (*ldh*/0475) is essential *in vivo* but non-essential *in silico*—suggesting that the residual capacity of pyruvate dehydrogenase is not sufficient to sustain cell growth on its own, or that the truncated complex is non-functional entirely. Furthermore, it suggests that pyruvate cannot be secreted at the same rate as lactate and/or that the assumed residual NADH oxidase activity would not be high enough to regenerate NAD+ for the GAPD/PGK branch in lower glycolysis (assuming a limit to how much NADPH produced through the alternative GAPDP branch the cell can utilize). The idea of a lower pyruvate transport rate is supported by reports of *M. mycoides* oxidizing external pyruvate 2-3 times slower than lactate [125, 126]. In nucleotide metabolism, the purine nucleoside phosphorylase (PNP, *punA*/0747) is quasi-essential *in vivo*, which suggests that either the nucleobase uptake capacities in JCVI-syn3A do not suffice to cover purine base demand; or that the

flux through the pentose phosphate pathway on its own does not suffice to provide all required PRPP, and additional synthesis starting from R1P released by PNP is necessary (see also Section 3.4). In lipid metabolism, the second glycosyltransferase, 0697, assumed to be redundant with *cps*/0114 for synthesis of Gal-DAG (and hence non-essential *in silico*), is quasi-essential *in vivo*—suggesting that either the activity of *cps*/0114 is not high enough to cover both Gal-DAG and capsule production, or that *cps*/0114 only catalyzes capsule production and is not involved in Gal-DAG synthesis at all. In amino acid metabolism, ten genes related to amino acid uptake and utilization are non-essential *in silico* but quasi-essential *in vivo*: namely, all genes of the Opp peptide importer (*oppB*/0165 through *oppA*/0169), three of the peptidases (0305, 0444 and 0479), the glutamate/aspartate permease GltP (*gltP*/0886) and one of the amino acid permeases (0878). The quasi-essentiality of Opp (and, possibly functionally related, three peptidases) is consistent with *M. mycoides capri* LC Y requiring alanine in peptide form for optimal growth [41]; the growth rate decreases a bit if instead only free alanine is provided. The quasi-essentiality of Opp suggests that this effect could be more pronounced in the SP4 media, which might not have free alanine at the 1 mM concentration used in Rodwell [41]. The *in vivo* quasi-essentiality of at least one amino acid permease (0878) is plausible in light of peptide-incorporated cysteine being not easily utilized by mycoplasmas [23]. In line with this, an *in silico* knockout of both amino acid permeases (0876 and 0878) is lethal. However, the *in vivo* quasi-essentiality of the glutamate/aspartate permease GltP is surprising in the light of *M. mycoides capri* LC Y reportedly not depending on these two amino acids being provided in the media [41]. If GltP is knocked out *in silico*, the model instead acquires glutamic and aspartic acid in their peptide form through the peptide importer Opp. Additionally imposing and gradually decreasing an upper limit on Glu/Asp peptide uptake however also gradually decreases the *in silico* growth rate, demonstrating the model dependence on Glu/Asp uptake. (The proposed dUMP disposal reaction CTPSDUMP is switched off in this test, since it would otherwise allow for unrealistically high glutamate production.) The latter observation would be consistent with the *in vivo* quasi-essentiality of GltP if the Glu/Asp peptide uptake capability through Opp was truly limited. The ability of *M. mycoides capri* to grow without Glu/Asp must then depend on some functionality removed during minimization of the genome.

Three "weak" false positives (essential *in silico* and quasi-essential *in vivo*) suggest relaxed substrate specificities or additional functionalities. In lipid metabolism, the quasi-essentiality of *fakB*/0617 suggests that the specificities of the two *fakB* genes might be relaxed and less complementary than in *S. aureus* [93] (albeit the overlap of activity could not be sufficient to maintain a stable membrane composition in the long run). In nucleotide metabolism, cytidylate kinase (CYTK, *cmk*/0347) is only quasi-essential *in vivo*, but it is not apparent how the cell would produce dCTP in its absence, or deal with CMP (produced from RNA breakdown). The most obvious hypothesis would be relaxed substrate specificity of uridylate kinase (UMPK, *pyrH*/0537) to also act on CMP and dCMP (possibly with weaker activity), as is the case for the eukaryotic enzyme. Guanine phosphoribosyltransferase (GUAPRT, *hptA*/0216) is the only phosphoribosyltransferase that is only quasi-essential. This suggests either some small but not negligible guanosine kinase activity or some capability to import intact GMP. Unfortunately, Neale et al. [63] only studied uptake of deoxynucleotides and CMP in *M. mycoides capri*, so that the possibility of GMP uptake seems conceivable but is not directly supported by experiment. However, it is noteworthy that GUAPRT is one of the few reactions whose $V_{\max}$ obtained from proteomics and turnover numbers is significantly smaller than the steady-state flux demanded by the FBA model (with a ratio of 0.07), and the only such reaction within nucleotide metabolism (see Section 2.4); this lends further support to the hypothesis of additional routes to GMP.

The remaining discrepancies of individual genes can be rationalized by genetic context or biological interpretation. In central metabolism, the NADPH-producing GapN (*gapdh*/0451) is non-essential in the model (with RNDR as the only consumer of NADPH) but essential *in vivo*. As discussed in Section 2.5, there are most likely other cellular NADPH demands not captured yet by the model, which would explain the *in vivo* essentiality. In nucleotide and lipid metabolism, the genes *nrdE*/0771 and *nrdF*/0773 (subunits of RNDR) and *pgpA*/0214 (PGPP) are essential in the

FBA model but only quasi-essential in the transposon data. However, the RNDR maintenance gene *nrdI* (*nrdI*/0772) is essential *in vivo*, in line with the *in silico* essentiality of RNDR. Similarly, the gene downstream from *pgpA* in the lipid synthesis pathway, cardiolipin synthase (*clsA*/0147), is essential, just as the entire pathway is *in silico*. This suggests that *nrdE/F* and *pgpA* might also be essential but not identified as such by the transposon mutagenesis analysis, possibly because their knockouts do not become lethal immediately (see Section 2.1). In Appendix 1, we provide a possible rationalization for these genes in terms of enzyme dilution and required fluxes. Finally, the genes for spermine uptake (*potC*/0195 through *potA*/0197), $3',5'$-adenosine bisphosphate breakdown (*ytqI*/0139) and the peptidase Lon (*lon*/0394) are all quasi-essential *in vivo*, suggesting that lack of nucleic acid stabilization by spermine, lack of protein turnover or buildup of $3',5'$-adenosine bisphosphate are not detrimental immediately—a circumstance not captured by the steady-state FBA model. This could be another possible example of time-delayed lethality upon knocking out an essential gene—where the detrimental effect might take time to manifest even if all protein flux were to cease quickly.

In addition to these discrepancies, we note a total of 12 genes non-essential both *in silico* and *in vivo*. These are strong candidates for removal attempts, observing that synthetic lethalities must be avoided (see also Section 3.5).

3.3 The role of folate metabolism

While all folate-related genes (the putative uptake gene *folT*/0822, the folylpolyglutamate synthase *folC*/0823 and all genes of the folate cycle) in cofactor metabolism are essential *in silico*, they are only quasi-essential *in vivo*, with the exception of the gene for 5-formyl-THF cyclo-ligase (*ygfA*/0443, which is essential *in vivo*). The *in vivo* essentiality of YgfA could mean inhibitory actions of YgfA's substrate 5-formyl-THF outside folate metabolism (see Appendix 1). The *in silico* essentiality of all folate genes arises *per construction* in the model since formylation of Met-tRNA$^{\text{fMet}}$ for translation is assumed, and a THF derivative is included in the biomass.

Intriguingly however, Neale et al. [85] demonstrated the ability of *M. mycoides capri* to omit Met-tRNA$^{\text{fMet}}$ formylation, and to initiate protein synthesis with the unformylated species, in the absence of folate in the media–without significant change in doubling time. This raises the question if there is a way to remove the folate-related genes in JCVI-syn3A. The discrepancy between *M. mycoides capri*, which can grow without folate, and JCVI-syn3A, where folate genes are quasi-essential, might arise out of the different experimental setups: The classification of genes in our transposon mutagenesis experiments is based on outgrowth of slow-growers in a competition experiment, whereas Neale et al. [85] necessarily could only study the impact of folate free media on the entire culture. Probing the fitness of a genotype by competition experiments is known to be more sensitive than mere growth rate measurements, as has e.g. been observed in studies of rRNA modification enzyme knockouts in *E. coli* [127, 128]: While the knockouts affected the log-phase growth rate little to none, they did lead to the mutants being outgrown by the wild type in competition experiments to different degrees. If a similar situation were the case for folate usage in *M. mycoides capri* as well, it could mean that folate non-utilization yields a fitness disadvantage that was not amenable to detection in Neale et al. [85].

Alternatively, the removal of some translation-related genes, like RNA modification enzymes, during genome minimization might have made the cell more susceptible to further interference with translation by disrupting folate-related genes, preventing both Met-tRNA$^{\text{fMet}}$ formylation and tRNA wobble position uridine modification (through MnmEG: *mnmE*/0081, *mnmG*/0885). (JCVI-syn3A also contains the folate-dependent 23S rRNA modification gene *RlmFO*/0434, which is non-essential in the transposon data however.) It is generally known that translational mistakes, such as those introduced by enzyme knockouts, can sum up and become detrimental once a certain threshold is crossed. In the genome of JCVI-syn3A, there already seems to be a precedent for this threshold scenario: Thiouridine tRNA modification through IscS and MnmA (*iscS*/0441 and *mnmA*/0387) appears to be non-essential in *M. mycoides capri*

LC Y, which can grow without a precursor for the IscS cofactor pyridoxal phosphate [41]. However, the transposon mutagenesis data for JCVI-syn3A indicate both genes to be essential, which must hence be due to the removal of other genes. Thus, if the genes could be identified whose removal similarly rendered folate usage quasi-essential, reintroducing them might enable one to delete folate metabolism in JCVI-syn3A entirely. A possible candidate for such a gene might be the deleted *rmsB*/MMSYN1_0204. RmsB catalyzes the methylation of 16S rRNA C967 using *S*-adenosylmethionine and plays a role in translation initiation in *E. coli*—both in binding of $tRNA^{fMet}$ [128] and in fidelity of initiation [129]. Removal of *rmsB* in JCVI-syn3A might thus have exacerbated the effects of further translation perturbations by non-formylation of Met-$tRNA^{fMet}$ resulting from folate gene knockouts. This might thus be an example of different routes of minimization yielding different minimal genomes.

3.4 A partial bypass to the pentose phosphate pathway

While all genes from the pentose phosphate pathway are essential *in silico*, they are only quasi-essential *in vivo*—except for the final gene, *prs*/0831 (PRPPS), which converts ribose 5-phosphate (R5P) to the pathway end product PRPP. This is noteworthy because with *deoB*/0733 (PGMT/PPM), there is an alternative route to R5P, bypassing the rest of the pentose phosphate pathway by providing pentose sugars from nucleoside breakdown. We note that PRPP itself is still needed, as is also evidenced by the essentiality of its utilizing enzymes encoded by the genes *apt*/0413 (ADPT), *upp*/0798 (UPPRT), and 0164 (NCTPPRT) in nucleotide and cofactor metabolism. However, the fact that the remainder of the pathway is only quasi-essential *in vivo* suggests that the demand on R5P could partially be covered by R1P from purine nucleoside phosphorylase (PNP, *punA*/0747; catalyzing reactions PUNP1–4).

For every equivalent of R1P released, PNP will also yield as the other product an equivalent of free purine base, which, in order to be utilized, would consume again the equivalent of PRPP generated from the R1P released. In this way, PNP would allow to exactly meet the PRPP demand for AMP and GMP synthesis; in the absence of pyrimidine nucleoside phosphorylase in JCVI-syn3A however (see Section 2.3.2), the PRPP demand of uracil would still not be met (and neither would the much smaller demand of nicotinate ribonucleotide synthesis). For PNP alone to cover the R1P/PRPP demand, it would need to degrade more purine nucleosides and release purine bases than can be used in the cell, and the cell would have to secrete purine bases at the same rate as uracil is taken up and phosphoribosylated.

This conclusion is supported by an *in silico* test: Setting the base symport reactions in the model to reversible for testing purposes indeed renders Tkt, Rpi and Rpe non-essential *in silico*—but not Prs. For example, upon knocking out the first reaction of the pathway (TKT2), we observe that the flux through PRPPS equals that through PPM (providing R5P), which in turn equals the sum of the fluxes through PUNP1 and PUNP3 (providing R1P from nucleoside breakdown). Accordingly, PNP (*punA*/0747) becomes essential *in silico* under these circumstances. Furthermore, the model secretes about as much purine bases as it takes up uracil (nicotinate uptake accounting for the very small difference).

It is unclear how biologically realistic this scenario is, but proton-symport coupled secretion of bases released from nucleosides has been suggested for mycoplasmas, and has even been proposed to contribute to the transmembrane proton gradient [65]. If such a nucleobase export capability existed but did not enable purine base secretion at the same rate as uracil is taken up at optimal growth rate, this would explain both why the pentose phosphate pathway is not fully essential *in vivo*, and why it is nonetheless still quasi-essential, as the R1P bypass would still not suffice to cover PRPP demands.

Except for the above test, we keep base uptake irreversible in the model as the biological feasibility of secretion is not known and cannot be easily implemented in the model. Specifically, it allows for feeding of excess (d)R1P into

glycolysis, which renders PGI and the PTS system non-essential *in silico* even if fluxes through the pentose phosphate pathway and DRPA are constrained with the V_{max} values derived from proteomics and k_{cat} values (see Section 2.4). For testing purposes, the above test with knocking out TKT2 therefore also had the DRPA flux set to zero.

3.5 Targeted gene removal experiments

The 12 true negatives from the essentiality comparison (genes non-essential both *in silico* and *in vivo*) are strong candidates for attempts to remove more genes, and hence minimize the genome of JCVI-syn3A even further. Suggested gene removal experiments are listed in Table 7 and are discussed in the following. Except for *folT*/0822 in the last two suggested experiments, all listed genes can be removed simultaneously *in silico*, as any interdependencies between them do not pertain to essential functions. The resulting *in silico* doubling time is 3.2 h, i.e. the same doubling time observed by knocking out any single gene along the acetate branch.

All of the true negatives are either individual genes or belong to short pathways. The genes *manA*/0435 (mannose 6-phosphate isomerase) and *deoC*/0732 (deoxyribose phosphate aldolase) could be removed individually. The whole acetylmannosamine branch (*nanE*/0494, 0495 and *nagB*/0726) could be removed, with the exception of the proposed N-acetylglucosamine 6-phosphate deacetylase 0493, which per the transposon essentiality data seems to have some other/additional function. The remaining subunits of pyruvate dehydrogenase (*pdhC*/0227 and *pdhD*/0228) and the proposed lipoylpeptide importer 0401 could also be removed together per our reconstruction and transposon data. The remaining individual functions include the two deoxyadenosine kinases (*dak1*/0330 and *dak2*/0382), a peptidase (*ietS*/0133) and an amino acid permeases (0876). The experimental non-essentiality of the peptidase and the amino acid permease supports the model assumption of broad substrate profiles and hence some redundancy among peptidases and amino acid permeases. The individual non-essentiality of the two deoxynucleoside kinases is also consistent with the assumed overlapping substrate profiles (see Appendix 1). However, it is not known whether the ribodinucleotide reductase (RNDR, *nrdE*/0771 through *nrdF*/0773) on its own could supply all dNTPs if both deoxyadenosine kinases were removed. Thus, simultaneous removal of *dak1*/0330 and *dak2*/0382 might incur a synthetic lethality or growth defect.

Furthermore, the observed quasi-essentiality of folate metabolism, in conjunction with the experimental observations of folate-free growth [41, 85], raises the question if even this subsystem could be removed somehow. If this were the case, it would allow deleting a number of genes at once: On top of the five quasi-essential folate genes, these would also include the 5-formyl cycloligase, *ygfA*/0443, as the metabolic source of 5-formyl-THF would be gone; the remaining THF-dependent RNA modification enzymes (MnmEG: *mnmE*/0081, *mnmG*/0885; and *RlmFO*/0434); and the peptide deformylase *def*/0201, as nascent peptides would not be formylated anymore to begin with. Thus, a removal of ten genes might be possible if JCVI-syn3A could be shown to grow without folate usage, or be enabled to do so (e.g. by reintroducing translation-related gene(s) whose removal might have raised the importance of Met-tRNA^{fMet} formylation; see Section 3.3).

To probe the fitness cost of folate non-usage in JCVI-syn3A more specifically and study whether this fitness cost is affected by the genome minimization, it would be of interest to carry out a specific competition study of a *folT* knockout vs. the wild type for both JCVI-syn3A and JCVI-syn1.0. If folate non-usage only occurs a significant fitness cost in JCVI-syn3A but not in JCVI-syn1.0, this would imply that JCVI-syn3A could be re-enabled to grow without folate as well.

TABLE 7 List of suggested gene removal experiments.

Gene(s)	Description	Remark
manA/0435	Mannose 6-phosphate isomerase	
deoC/0732	Deoxyribose phosphate aldolase	
nanE/0494, 0495, *nagB*/0726	N-acetylmannosamine 6-phosphate branch	
pdhC/0227, *pdhD*/0228, 0401	pdhCD and proposed lipoate importer	
dak1/0330, *dak2*/0382	Deoxynucleoside kinases	Synthetic lethality possible
ietS/0133	Peptidase	
0876	Amino acid permease	
folT in JCVI-syn3A	Folate uptake and usage	Competition experiment with wild type to probe fitness cost in JCVI-syn3A
folT in JCVI-syn1.0	Folate uptake and usage	Competition experiment with wild type to probe fitness cost in JCVI-syn1.0

3.6 Suggestions for further experimental study

As presented in the preceding sections, the comparison between *in silico* and *in vivo* essentiality yielded a number of hypotheses and suggested several possible gene removal experiments (Table 7). Similarly, the metabolic reconstruction itself yielded a number of informed hypotheses, as well as raised specific questions. While the minimal genome has been experimentally obtained, understanding all genetic functions both individually and as a system remains an ongoing challenge. Thus, the hypotheses and questions raised in this work provide invaluable help in the ongoing effort to completely understand the minimal genome. In Table 8, we provide a list of suggested experiments other than gene removal/knockout studies, sorted by category and providing a rationale for each experiment.

TABLE 8 List of suggested experiments, with rationale behind each suggestion.

Experiment	Rationale
Nutrient utilization	
Detect lipoylpeptide uptake	Lipoate is cofactor for PDH, whose functionality is unclear after E1 subunit deletion.
Detect nucleotide uptake	Activity reported for *M. mycoides capri* without gene assignment; alternative routes present in JCVI-syn3A, but activity not ruled out.
Detect nucleobase uptake	Activity reported for *M. mycoides capri*, and uracil uptake essential in model.
Demonstrate growth on thiamine diphosphate	Structural data and deletion of thiamine diphosphokinase suggest thiamine diphosphate (ThDP) uptake. Inability to grow on ThDP would imply unidentified kinase activity.
Demonstrate growth on pyridoxal phosphate	With no pyridoxal kinase identified, growth on pyridoxal phosphate (P5P) assumed; inability to grow on P5P would imply unidentified kinase activity.
Demonstrate growth on acetylmannosamine	Reconstruction suggests operon 0493 through 0495 to be *nagA*/*nanE*/*nagC*; growth on acetylmannosamine would support assignment and imply uptake capability.

TABLE 8 (cont.)

Experiment	Rationale
Demonstrate growth on mannose or glucosamine	Literature suggests mannose and glucosamine import through PtsG; and downstream enzymes are present.
Metabolite production/secretion	
Detect production of acetate	Acetate pathway has been partially removed, but several of the remaining enzymes remain essential in transposon mutagenesis experiments.
Investigate production of lipogalactan capsule	Genetic evidence suggests capsule is still being produced.
Detect secretion of deoxyuridine or dUMP	Deoxyuridine/dUMP is currently dead-end; secretion would be one possible solution.
Enzymatic activity	
Detect pyruvate oxidation	Conversion of pyruvate to acetyl-CoA by truncated PDH complex has been assumed for the time being but is not supported by experiment.
Detect oxidation of acetaldehyde	Conversion of acetaldehyde to acetyl-CoA would provide alternative explanation for presence of truncated PDH complex in JCVI-syn3A in spite of deletion of first subunit.
Detect NOX activity	NADH oxidase (NOX) has been deleted but activity would be necessary for PDH activity against both pyruvate and acetaldehyde.
Detect transaldolase activity	Activity is essential in model and has been detected in *M. mycoides capri*; no known gene in any mycoplasma though.
Detect sedoheptulose-1,7-bisphosphate phosphatase activity	Reaction would provide bypass to transaldolase reaction.
Detect phosphatidate phosphatase activity	Gene present in *M. pneumoniae* and reaction is missing link to diacylglycerol, but no gene identified in JCVI-syn3A.
Assess phosphatase activity against dCTP, dCDP, GMP, dAMP, dGMP, dUMP, dTMP; pyrophosphatase activity against CTP, dCTP	Activities observed in *M. mycoides capri* but no gene identified in JCVI-syn3A.
Detect deoxyuridine phosphorylase activity	Gene has been removed in JCVI-syn3A (MMSYN1_0734), but deoxyuridine/dUMP is currently dead-end, raising the question whether function is carried out by some other gene.
Detect thymidylate synthase activity	Extremely low activity detected in *M. mycoides mycoides* SC, but no gene identified. Reaction would be alternative solution to deoxyuridine/dUMP dead-end.
Specific gene function	
Determine substrates for deoxynucleoside kinase *dak2*/0382	Presence of *dak2*/0382 in addition to the characterized *dak1*/0330 suggests different substrate profile.
Verify CTPS activity against dUMP	CTPS (*pyrG*/0129) converting dUMP to dCMP seems most plausible solution to deoxyuridine/dUMP dead-end.
Check PGPP activity against phosphatidate	Activity observed for PgpB in *E. coli*; activity for PGPP (*pgpA*/0214) would provide missing link to diacylglycerol in apparent absence of phosphatidate phosphatase gene.
Check UMPK activity against CMP and dCMP	Substrate profile for UMP kinase similar to eukaryotic enzyme could explain quasi-essentiality of CMP kinase.

TABLE 8 (cont.)

Experiment	Rationale
Change of expression levels	
Knock out deoxynucleoside kinases & over express RNDR	RNDR and deoxynucleoside kinases provide redundant routes to deoxydinucleotides, suggesting one pathway might be sufficient if expression increased.
Reduce expression of RNDR and knock out putative dUTPase simultaneously	RNDR is currently only known source of dUTP. If RNDR knockdown would make the putative dUTPase gene *dut*/0447 nonessential as well, this would corroborate the putative assignment.

3.7 Comparison to *M. pneumoniae*

M. pneumoniae is an important systems biology model organism that has been extensively studied [21–25] so a comparison to its metabolic map should be of interest. With a published metabolic reconstruction (iJW145 [25]) that includes 304 reactions involving the products of 145 genes it is similar in size to the metabolic reconstruction of the minimal cell JCVI-syn3A with its 338 reactions and 155 genes. Utilizing the vast experimental information on *M. mycoides capri*, the natural precursor of JCVI-syn3A, as well as information on JCVI-syn3A homologs in other organisms, enabled us to obtain a smaller percentage of gap fills (i.e. model reactions assumed to be enzymatic yet having no gene assigned) of 6 % out of all model reactions, compared to 25 % in the *M. pneumoniae* model iJW145; or 9 % for JCVI-syn3A vs. 32 %, if exchange, macromolecular and biomass reactions are excluded from the total number of reactions in each model. The JCVI-syn3A model yields a higher degree of *in silico* essentiality (79 % vs. 56 % for the 131 "metabolic proteins" in the *M. pneumoniae* model [25])—reflecting the minimization of the JCVI-syn3A genome. This higher degree of essentiality is also reflected in the differences in individual reactions presented in Supplementary File 6 (see also Methods for details on the model comparison).

Excluding exchange, macromolecular and biomass reactions, a core of 126 reactions is shared between the models, including glycolysis, the pentose phosphate pathway, reactions from nucleotide, cofactor and lipid salvage pathways, and tRNA charging. However, *M. pneumoniae* has 116 reactions not present in JCVI-syn3A, which mainly includes uptake and utilization of additional sugar sources, further nucleotide conversions, more extensive cofactor salvage reactions, and additional lipid-related reactions. Some of these reactions were present in JCVI-syn1.0 but were removed during minimization of the genome to JCVI-syn3A. Furthermore, some of the differences are technical in nature, e.g. the choice to model amino acid uptake as ABC import reactions in *M. pneumoniae*, or the decision to include amino acid secretion reactions there. Interestingly, in spite of the much smaller genome of JCVI-syn3A, its reconstruction still contains 120 reactions not present in the *M. pneumoniae* model. While a number of these arise from a more detailed description of various transport processes (nucleosides, peptides and ions), we note the presence of some functionalities not present or known in *M. pneumoniae*. These include the production of a polysaccharide capsule (in addition to the monogalactosyl-lipid), some alternative sugar sources specific to JCVI-syn3A and also specific nucleotide conversion and breakdown reactions, perhaps most notably the presence of the essential damage preemption enzyme dUTPase (*dut*/0447, see Figure 8).

4 Conclusion

We have presented a comprehensive metabolic reconstruction and FBA model of the minimal cell JCVI-syn3A, informed by the extensive experimental information available for the natural precursor, *M. mycoides capri*, *in vivo* transposon mutagenesis and proteomics data. The metabolic model is near complete with regards to accounting for all biomass components, with known metabolic functions not included mainly pertaining to damage repair/pre-emption and RNA modification. The high quality of the model is exemplified by the strong support for the network, with 98 % of enzymatic reactions in the model justified through gene assignments and/or experimental evidence; and by its good agreement with experimental transposon mutagenesis data showing 92 % of the genes included in the model to be essential or quasi-essential. The essential metabolism of this minimal cell consists of only a few subsystems that are only minimally connected with each other. The subsystems for lipids, amino acids, nucleotides and cofactors contain only salvage pathways. An energy analysis shows how this reliance on salvage pathways enables the cell to only spend 9 % of its produced ATP on precursor transport and processing while maintaining a doubling time of 2 h. The experimental transposon mutagenesis data probe individual gene essentialities, which together with the metabolic model point to a few possible remaining redundancies. Comparison with *M. mycoides capri* further suggests that folate metabolism only became quasi-essential by removal of other genes, underlining how different routes of genome minimization could yield different minimal genomes. Model and accompanying experimental data thus not only reveal properties of the minimal metabolic network, but also yield an extensive list of suggested experiments to test the resulting hypotheses. The model, together with the accompanying transposon mutagenesis and proteomics data, provides an excellent foundation for further studies of the minimal cell.

Acknowledgements

We thank James Daubenspeck and Kevin Dybvig for information on the minimal cell capsule; Emile van Schaftingen for helpful suggestions on possible reactions in central metabolism; and David Bianchi for help with Figure 15–Figure supplement 1. M.B. gratefully acknowledges an NSF Postdoctoral Fellowship through the Center for the Physics of Living Cells (CPLC) at the University of Illinois at Urbana-Champaign (grant NSF PHY 1430124). T.E. and Z.L.S. were supported through grant NSF PHY 1430124. M.B., T.E. and Z.L.S. were supported through grant NSF MCB-1244570. P.L. was supported by the DOE/BER (ORNL 4000134575) as part of the Adaptive Biosystems Imaging Focus at ORNL (Oak Ridge National Laboratory). J.L. was supported through grant NIH K12 GM06852. V.C.L. and A.H. were supported through grant NSF MCB-1611711. D.G. is supported by the Ray Thomas Edwards Foundation and the University of California Office of the President. C.M., K.W., C.H., H.O.S. and J.G. were supported by the J. Craig Venter Institute. The bacterium JCVI-syn3A was created using funding from the J. Craig Venter Institute, Synthetic Genomics, Inc., and the Defense Advanced Research Projects Agency's Living Foundries program (contract HR0011-12-C-0063).

Conflict of interest

H.O.S. is on the Board of Directors and cochief scientific officer of Synthetic Genomics, Inc. (SGI). C.A.H. is a consultant for SGI. H.O.S., C.A.H., and the J. Craig Venter Institute (JCVI) hold SGI stock and/or stock options. SGI and JCVI have filed patent applications related to the design and construction of JCVI-syn3.0.

Supplementary files

Supplementary File 1 Supplementary File 1A: Hierarchy used in gene function classification. Adapted from Liebermeister et al. [130] for mycoplasmas. Supplementary File 1B: Breakdown of the JCVI-syn3A metabolic reconstruction. Reactions in subsystem "Exchange" act as sources (sinks) for metabolites taken up from (secreted into) the medium. The other subsystems are discussed in detail in the main text, Section 2.2 and Section 2.3. Supplementary File 1C: Breakdown of protein coding genes in *M. pneumoniae* [19] into functional classes. Supplementary File 1D: Breakdown of protein coding genes in *E. coli* [20] into functional classes.

Supplementary File 2 Transposon insertion nucleotide positions.

Supplementary File 3 Transposon insertion counts and assignment of gene essentiality from both transposon mutagenesis and FBA.

Supplementary File 4 Reactions and metabolites included in the metabolic reconstruction.

Supplementary File 5 Data from proteomics experiments.

Supplementary File 6 Comparison of the JCVI-syn3A metabolic reconstruction to that of *M. pneumoniae* published by Wodke et al. [25].

Supplementary File 7 Flux constraints derived from proteomics and turnover numbers and comparison to FBA fluxes.

Supplementary File 8 Known metabolic reactions removed during genome minimization from JCVI-syn1.0 to JCVI-syn3A.

Supplementary File 9 FBA model in sbml format.

Supplementary File 10 FBA model in json format.

Supplementary File 11 ESCHER network map in json format.

Transparent reporting form

Methods

Model construction

A genome-scale FBA model requires the reconstruction of the network of metabolic reactions, the assembly of the cellular biomass composition and necessary reaction constraints (e.g. substrate uptake and ATP consumption). The biomass composition of JCVI-syn3A was assembled based on experimental information available for *Mycoplasma mycoides capri* (in a few instances using information from other organisms). The reconstruction of the metabolic network began with the curated annotation published for JCVI-syn3.0 [12] (which also contained annotations for all genes removed from JCVI-syn1.0). As done in other models [60], an existing curated model was used as a reference to construct a first draft reconstruction. Initially, an FBA model for *M. pneumoniae* [25] was used, keeping all reactions whose enzymes had an equivalent in JCVI-syn3A. Information from MetaCyc [131], KEGG [132], and an extensive evaluation of primary literature was then used to add reactions for the remaining metabolism-related genes in JCVI-syn3A, as well as reactions without a gene but supported by experimental evidence (including the assembled biomass composition). Experimental evidence was also used to exclude certain candidate reactions. Finally, a few reactions were added as gap-fills to complete the respective pathways. Metabolite and reaction IDs were matched to BiGG

TABLE 9 Key resources table.

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Strain, strain background	JCVI-syn3A	JCVI; this article	GenBank accession number:CP016816.2	[1]
Strain, strain background	JCVI-syn1.0	doi:10.1126/science.1190719	GenBank accession number:CP002027.1	[1]
Genetic reagent	terTufPuro transposome	doi:10.1126/science.aad6253		Constructed by the JCVI
Genetic reagent	Yeast tRNA	Life Technologies, Carlsbad, CA, USA	15-401-029	
Genetic reagent	EZ-Tn5-Transposase	Lucigen, Madison, WI, USA	TNP92110	
Sequence-based reagent	Custom forward primer	Integrated DNA technologies, San Diego, CA, USA		[2]
Commercial assay or kit	Nextera XT DNA library preparation kit	Illumina, San Diego, CA, USA	FC-131-1024	
Chemical compound, drug	Quant-iT^{TM}PicoGreen	Molecular Probes, Eugene, OR, USA	P7589	
Chemical compound, drug	Puromycin	Molecular Probes, Eugene, OR, USA	A1113802	
Software, algorithm	COBRApy	doi:10.1186/1752-0509-7-74		
Software, algorithm	CLC Genomics Workbench	QIAGEN Bioinformatics, Redwood City, CA, USA		
Software, algorithm	ProteomeDiscoverer 2.1.0.81	Thermo Fisher Scientific		

[1] Bacterial strains JCVI-syn3A and JCVI-syn1.0 will be made available to qualified researchers by the JCVI and Synthetic Genomics, Inc. under a material transfer agreement. Note that United States scientists must obtain a United States Veterinary Permit for Importation and Transportation of Controlled Materials and Organisms and Vectors from the U. S. Department of Agriculture Animal and Plant Health Inspection Service. The organisms require Biosafety Level 2 containment.
[2] Used for marker-specific sequencing with PCR; sequence under "Tn5 mutagenesis–Experimental method".

IDs [133, 134] when possible, otherwise IDs akin to BiGG IDs were assigned. Additionally, KEGG compound IDs were assigned to metabolites using the KEGG API; and InChI keys were assigned using the API for the Chemical Translation Service [135].

Flux constraints for certain reactions were based on *in vivo* measurements, other models or physicochemical parameters. Reaction reversibilities were based on information from MetaCyc [131] and eQuilibrator [136], inferred by analogy (e.g., fatty acid kinase was set as reversible like acetate kinase) or determined from biochemical context (e.g., H^+ diffusive influx is set to irreversible, in accordance with *in vivo* flow direction).

Flux-balance analysis

Model assembly and flux-balance analysis [28] were carried out in COBRApy [122], a Python module for constraint-based modeling. In flux-balance analysis, a system of n reaction equations featuring in total m reactants is represented as a stoichiometric matrix $\boldsymbol{S}$ of dimensions $m \times n$, where the element S_{ij} denotes the stoichiometric coefficient of reactant i in reaction j (negative for reactants, positive for products). A given set of fluxes through each reaction in the system is represented as a flux vector $\vec{v}$ of length n. Any steady-state flux vector then belongs to the solution space of the equation $\boldsymbol{S} \cdot \vec{v} = \vec{0}$. This solution space is further constrained by any other constraints imposed on individual fluxes of the form $V_{\min,j} < v_j < V_{\max,j}$. A default upper bound $V_{\max}$ of 1000 mmol $\mathrm{gDW}^{-1}\,\mathrm{h}^{-1}$ was used for all reactions and a default lower bounds of -1000 mmol $\mathrm{gDW}^{-1}\,\mathrm{h}^{-1}$ and 0 mmol $\mathrm{gDW}^{-1}\,\mathrm{h}^{-1}$ were used for reversible and irreversible reactions, respectively. Specific constraints were chosen to account for uptake, secretion and ATP consumption restrictions. An optimal flux vector or set of flux vectors within the constrained solution space is then found by maximizing a particular objective function by means of linear programming. We picked biomass production as our objective function, so that the optimal flux vector describes the optimal growth under the chosen constraints.

As the solution to the flux optimization may not be unique, parsimonious FBA (pFBA) [137] is employed to obtain a unique solution. In pFBA, the optimal growth rate obtained by using the original objective function (biomass production in our case) is subsequently set as a constraint and a new objective function is defined with a coefficient of -1 for all reactions not part of the original objective function. Optimizing the flux vector under this objective function then yields the solution with the smallest sum of individual fluxes, corresponding to minimal enzyme usage in a biological context. Reversible reactions are split into two irreversible reactions for this purpose so as to avoid negative

fluxes being maximized rather than minimized.

Calculation of energy usage by subsystem

In order to analyze the energy consumption in the metabolic model for JCVI-syn3A, the consumption of ATP equivalents per subsystem was calculated. The term "ATP equivalent" is used to account for the fact that phosphorylation of all dinucleotides in JCVI-syn3A is assumed to be carried out by the glycolytic enzymes phosphoglycerate kinase and pyruvate kinase, so that the phosphate donors are 1,3-diphosphoglycerate (1,3-DPG) and phosphoenolpyruvate (PEP) instead of ATP (whose role in dinucleotide phosphorylation is effectively bypassed). For all model reactions not involving 1,3-DPG or PEP, the production or consumption of ATP equivalents was calculated from the number of phosphate bonds formed or broken in each reaction producing or consuming ATP multiplied by the flux through that reaction in the FBA solution. Interconversion of ATP and ADP produces/consumes one phosphate bond. Hydrolysis of ATP to AMP (e.g. in tRNA charging) was counted as consuming two phosphate bonds, since the free pyrophosphate can only be hydrolyzed further to two individual phosphates. Consumption of the ATP moiety as a whole (e.g. in NAD+ synthesis) was also counted as consuming two phosphate bonds, accounting for the phosphorylation steps from AMP to ATP; the energy spent in AMP is already accounted for in other reactions (nucleoside uptake and PRPP synthesis for adenine phosphoribosylation). The flux through adenylate kinase (ADK1) phosphorylating AMP to ADP is already accounted for by counting ATP–>AMP hydrolysis as two phosphate bonds; it is thus ignored to avoid double-counting. To properly account for 1,3-DPG and PEP as phosphate donors for trinucleotide production, the fluxes for the ATP-producing PGK and PYK reactions were set equal to the sum of all PGK or PYK fluxes, respectively, in order to obtain the total number of ATP equivalents produced. In turn, the PGK and PYK model reactions phosphorylating dinucleotides other than ADP were counted as consuming one ATP equivalent each. Accuracy and correct accounting of the calculated ATP equivalent creation and consumption fluxes were verified by confirming that all individual fluxes thus calculated added up to zero.

To obtain the total ATP equivalent consumption percentage per category in Table 3, the consumption fluxes for all reactions in a given category were added up and normalized by the total ATP equivalent consumption flux in the model. (Central metabolism as the only source of ATP is not included in Table 3.) In doing so, an own category "$\text{NGAM}_{\text{Turnover}}$" was introduced to include all energy expenses attributable to protein and RNA turnover. This includes the ATP spent on protein degradation itself, as well as the fractions of protein synthesis, RNA synthesis, tRNA charging and phosphorylation of mononucleotides to trinucleotides that produce protein and RNA for turnover only (as determined from the protein and RNA degradation reaction constraints). The remainder of the protein and RNA synthesis fluxes then produces protein and RNA to be consumed in the biomass equation; hence, the associated energy consumption is part of the quantifiable fraction of the growth-associated maintenance (GAM) cost, and is hence included in "$\text{GAM}_{\text{Macromolecules}}$" in Table 3 (DNA synthesis being the other cost included). Similarly, "$\text{GAM}_{\text{tRNA charging}}$" is the fraction of energy expense in tRNA charging attributable to growth-associated protein synthesis. The consumption of the ATP moiety in RNA and biomass production was included in nucleotide metabolism, in order to stay consistent with the definition of the GAM to only include the ATP *hydrolyzed* for growth (including macromolecular synthesis), but not the consumption of the ATP moiety as a precursor (see also Figure 4). Accordingly, the ATP hydrolyzed in RNA synthesis was included under "$\text{GAM}_{\text{Macromolecules}}$". Finally, PRPPS (PRPP synthase) is part of central metabolism in the model but as a reaction is independent from energy production in glycolysis. It was hence assigned its own subsystem ("Pentose phosphate pathway") for the purposes of energy usage breakdown.

Model comparison

Reactions between the models for JCVI-syn3A and for *M. pneumoniae* (iJW145) [25] were compared programmatically by associating with each reaction in either model a set of involved metabolites, excluding water, P_i, and H^+. By comparing only the involved metabolites, differences in stoichiometry, reversibility, and mass balance between the two models are not considered. To develop a common language of metabolites between the two models, a mapping from the chemical name in the model SBML file to a KEGG compound identifier (C number) was constructed. The KEGG API was used to search for a C number based on the substrate description in the SBML files. When a C number was not found for a particular substrate, the mapping was created by hand. The name to compound map was verified manually by comparing the name given in the model to the name given in the KEGG database to that C number.

A reaction in the JCVI-syn3A model was determined to be equivalent to a reaction in the iJW145 model if the metabolite sets associated with each reaction were equal. The resulting grouping of reactions into common or model-specific reactions was then manually curated to distinguish reactions where different directionalities/reversibilities between models arose from different roles of these reactions in the model (irreversible amino acid influx in JCVI-syn3A vs. irreversible amino acid efflux in iJW145, which has a separate set of ATP-driven amino acid uptake reactions).

Growth curve measurements

Growth and rate measurements of minimized synthetic cells have been described in detail elsewhere [12]. Briefly, cells were grown in SP4 medium to mid-late log phase in static cultures, then diluted in fresh pre-warmed (37 °C) medium. Subsequent samples obtained over time were centrifuged to remove medium, cells were lysed with dilute detergent, and released dsDNA was measured using the fluorescent stain Quant-iT^{TM} PicoGreen® (Molecular Probes®, Eugene, OR, USA). Fluorescence was measured in a 96-well format using a FlexStation 3 fluorimeter (Molecular Devices, San Jose, CA, USA). The net relative fluorescence units (RFU) of samples (after subtracting RFU from a medium control lacking cells), were plotted as $\ln(RFU)$ vs. time from which the doubling times, τ_d were calculated from the slopes of exponential regression curves (k) as

$$\tau_d = \frac{\ln 2}{k}.$$

Rates were measured from log–linear portions of the growth curve. To avoid minor variables such as batch differences among medium preparations and temperature fluctuations, strains with different genomes were compared under identical conditions and within a single experiment. The accuracy and reproducibility of the measurements (reflected in the observed R^2 values, see Figure 14) allowed the use of single samples, as also observed previously [12].

Tn5 mutagenesis

Experimental methods

We used the procedure described in Hutchison et al. [12] with minor modifications. A single experiment was performed, which however yielded ∼92,000 transformed colonies (see below), and hence ∼92,000 insertion events across a genome with 493 genes. This was deemed to yield sufficient statistics. Preparation of terTufPuro transposomes was as described. JCVI-syn3A cells were grown in SP4 media until reaching pH 6.3–6. For one transformation reaction, we used 8.8 ml of culture. The cells were centrifuged for 15 min at 4700 rpm at 10 °C in a 50-ml tube. The pellet was resuspended in 3 ml of S/T buffer (Tris 10 mM, sucrose 0.5 M, pH 6.5). The resuspended cells were centrifuged for 15 min at 4700 rpm at 10 °C and the supernatant was removed. The pellet was resuspended in 250 µl of 0.1 M $CaCl_2$ and incubated for

30 min on ice. Transposomes (2 µl) and yeast tRNA (10 µg) (Life Technologies, Carlsbad, CA, USA) were mixed gently with the cells. Two ml of 70 % poly(ethylene glycol) (PEG) 6000 (Sigma) dissolved in S/T buffer was added. We allowed a maximum of 2 min in contact with PEG at room temperature. The components were mixed well during the 2 min of incubation. S/T buffer (20 ml) was added immediately after 2 min and mixed well. The tube was centrifuged at 8 °C for 15 min at 10,000×g. The supernatant was discarded and thoroughly drained from the tube by inversion onto a Kimwipe. The cells were resuspended well in 1 ml of warm SP4 media and incubated for 3 h at 37 °C and then plated on SP4 agar with 2 µg/ml of puromycin (Sigma). The plates were incubated for 3–4 days at 37 °C.

An estimated 92,000 colonies were harvested from the plates in 20 ml of SP4 media (passage zero, P_0) and a 45 µl sample was added to 45 ml SP4 media containing puromycin 2 µg/ml and grown for 24 h (passage one, P_1). A 45-µl sample of P_1 culture was added to 45 ml of fresh SP4 media (P_2) and grown for 48 h. Two more passages (P_3 and P_4) were done.

DNA preparations from each passage were done as described in Hutchison et al. [12] and DNA preparations were additionally purified by gel electrophoresis. A Nextera XT DNA library preparation kit was used for paired-end library construction (Illumina, San Diego, CA, USA) by the manufacturer's protocol with the following change. A forward primer 5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNN-NNNGCCAACGACTACGCACTAG designed by us and the reverse primer from the kit were used for the PCR amplification to achieve marker-specific sequencing [138]. To locate points of Tn5 insertion in the JCVI-syn3A genome, sequence reads were searched for the Tn5 19-bp terminus followed by an exact 30-bp match to genome sequence. The Tn5 to genome junction point identified the insertion location. "Duplicate insertions" (i.e., insertions found in sequences repeated in the genome) were ignored in all further analyses since they could not be unequivocally assigned to a single gene. Tn5 insertions were displayed using CLC Genomics Workbench (QIAGEN Bioinformatics, Redwood City, CA, USA).

Classification of genes

To place genes into "essential", "quasi-essential", and "non-essential" classifications, a simple statistical model of transposon insertion was used. It assumes that the positions of insertions within a gene are unbiased (Poisson distributed) and that the number of insertions in a particular gene is described by one of two possible distributions. The two distributions separate the genes into groups with few and many insertions. The probability to observe n_i insertions for gene i, which has a sequence length of ℓ_i is

$$P(n_i|\ell_i) = p_{\mathrm{lo}} \mathrm{e}^{k_{\mathrm{lo}}\ell_i} \frac{(k_{\mathrm{lo}}\ell_i)^{n_i}}{n_i!} + p_{\mathrm{hi}} \mathrm{e}^{k_{\mathrm{hi}}\ell_i} \frac{(k_{\mathrm{hi}}\ell_i)^{n_i}}{n_i!}, \tag{1}$$

where k_{lo} and k_{hi} are the transposon insertion rates for the few and many insertion distributions, and p_{lo} and p_{hi} are the probabilities that the insertions follow the few or many insertion distribution, respectively, such that $p_{\mathrm{lo}} + p_{\mathrm{hi}} = 1$. The model is fit to the experimental data using expectation–maximization [139].

The probability that a particular observation of n_i transposon insertions for a gene of length ℓ_i follows the fewer insertion distribution for passage j, is

$$P_{\mathrm{lo}}(n_{ij}|\ell_i, j) = p_{\mathrm{lo},j} \mathrm{e}^{k_{\mathrm{lo},j}\ell_i} \frac{(k_{\mathrm{lo},j}\ell_i)^{n_{ij}}}{n_{ij}!}. \tag{2}$$

Comparing passage 1 and passage 4, the classification probabilities are

$$P(\text{Essential}|n_{i1},n_{i4},\ell_i) = P_{\text{lo}}(n_{i1}|\ell_i,1)\cdot P_{\text{lo}}(n_{i4}|\ell_i,4) \tag{3a}$$

$$P(\text{Quasi-essential}|n_{i1},n_{i4},\ell_i) = [1-P_{\text{lo}}(n_{i1}|\ell_i,1)]\cdot P_{\text{lo}}(n_{i4}|\ell_i,4) \tag{3b}$$

$$P(\text{Non-essential}|n_{i1},n_{i4},\ell_i) = [1-P_{\text{lo}}(n_{i1}|\ell_i,1)]\cdot [1-P_{\text{lo}}(n_{i4}|\ell_i,4)]. \tag{3c}$$

A gene is assigned to a category if the classification probability is greater than 0.5. Genes where the classification probabilities are all less than 0.5 (labeled "unclassifiable" in Figure 2 and Figure 2–Figure supplement 1-2, or where $P_{\text{lo}}(n_{i1}|\ell_i,1)\cdot[1-P_{\text{lo}}(n_{i4}|\ell_i,4)] > 0.5$ are not classifiable by this method and were manually assigned an essentiality class.

To differentiate weakly quasi-essential genes from non-essential genes, the genes identified as non-essential were further classified using k-means clustering (provided by the SciPy [140] library) of the ratio of transposon insertion counts in P_4 to P_1.

Mass Spectrometry Based Proteomics

Cell preparation With the objective of studying protein expression changes along the growth curve (unrelated to the current study), mass spectrometry with tandem mass tag labeling was carried out on JCVI-syn3A samples from different time points along a growth curve. In the current study, we use the data from the first time point (logarithmic phase). JCVI-syn3A cells used for proteomic analysis were grown as described previously [12] and in "Growth curve measurements" above, using SP4 medium that contained heat inactivated horse serum (Invitrogen$^{\text{TM}}$) in lieu of FBS. Static cultures were sampled at different times to determine the culture stage (measured as described in "Growth curve measurements"). Six centrifuge bottles each containing approximately 130 mL of a logarithmic phase culture were centrifuged (10,000×g, 15 min, 20 °C) and the cell pellets were drained and resuspended in a small volume of medium. The suspensions were pooled, redistributed in 1-mL volumes and again centrifuged (16,000×g, 5 min, 20 °C). The resulting pellets were drained and used immediately for lysis. In total, three pellets were obtained from independently grown cultures as biological replicates. Samples of two further time points (early plateau and plateau phase) were prepared in an analogous fashion, with three biological replicates each.

Cell lysis and protein digestion Cells were lysed in a buffer comprised of 3 % SDS, 75 mM NaCl, 1 mM NaF, 1 mM β-glycerophosphate, 1 mM sodium orthovanadate, 10 mM sodium pyrophosphate, 1 mM PMSF and 1X Roche Complete mini EDTA free protease inhibitors in 50 mM HEPES, pH 8.5 [141]. Lysates were passed through a 21-gauge needle 20 times and sonicated for 5 min to ensure full lysis. Debris was pelleted by centrifugation at 14,000 rpm for 5 min, with resultant supernatants used for downstream processing. Briefly, proteins were reduced with DTT [142] and precipitated with methanol-chloroform [143] before re-suspension in 1 M urea in 50 mM HEPES, pH 8.5 for digestion. Digestion was performed in a two-step process, 1) with LysC overnight at room temperature, 2) with trypsin for 6 h at 37 °C. Digestion was quenched with TFA, and peptides desalted with C18 solid-phase extraction [144]. Dried peptides were re-suspended in 50 % acetonitrile/5 % formic acid and quantified via BCA assay.

Tandem mass tag labeling and fractionation Lyophilized peptides were re-suspended in 30 % dry acetonitrile in 200 mM HEPES, pH 8.5 and 8 μL of the appropriate tandem mass tag (TMT) reagent was added to each sample, incubated for 1 h at room temperature, and quenched with 5 % hydroxylamine. Labeled samples were then acidified with 1 % trifluoroacetic acid. Differentially labeled samples were pooled into multiplex experiments and then desalted via solid-phase extraction and lyophilized. Samples were fractionated by basic pH reverse-phase liquid chromatography, using a 4.6 mm $\times$ 250 mm C18 column on an Ultimate 3000 HPLC (Thermo Fisher Scientific, Waltham, MA, USA). In total, 96 fractions were collected and combined in a concatenated manner [145], lyophilized and re-suspended in 5 % formic acid/5 % acetonitrile for identification and quantification by LC-MS2/MS3.

LC-MS2/MS3 analysis All LC-MS2/MS3 experiments were performed on an Orbitrap Fusion mass spectrometer with an in-line Easy-nLC 1000 with chilled autosampler (Thermo Fisher Scientific). Peptides were separated on columns that were packed with C4 resin (5 μm, 100 Å), followed by C18 resin (3 μm, 200 Å) and then to a final length of 30 cm with C18 (1.8 μm, 12 Å). Peptides were eluted with a linear gradient from 11 to 30 % acetonitrile in 0.125 % formic acid over 165 min at a flow rate of 300 nL/min and heating the column to 60 °C. Electrospray ionization was achieved by applying 2000 V through a stainless-steel T-junction. Mass spectrometer settings were as previously described [146].

Data processing and analysis Data were processed using the ProteomeDiscoverer 2.1.0.81 software package (Thermo Fisher Scientific). The built-in version of SequestHT [147] was utilized to assign identities to MS2 spectra searching against the JCVI-syn3A database downloaded from NCBI. The database was appended to include a decoy database comprised of all protein sequences in reversed order for downstream false discovery estimation [148–150]. Search parameters included a 50 ppm MS1 mass tolerance [151], 0.6 Da fragment ion tolerance, fully-enzymatic trypsin with a maximum of two missed cleavages per peptide, static modifications of 10-plex TMT tags on lysines and peptide n-termini and carbamidomethylation of cysteines. Variable modifications included oxidation of methionines and phosphorylation of serine, threonine and tyrosine residues. Data were filtered to a peptide and protein false discovery rate of less than 1 % using the target-decoy search strategy [150]. Peptides matching to multiple proteins were assigned to the protein containing the largest number of matched redundant peptides following the law of parsimony [151]. TMT reporter ion intensities were extracted from MS3 spectra for quantitative analysis. Spectra used for quantitation had to meet the requirements of greater than 10 average signal-to-noise per label and isolation interference of less than 25 % [152]. Data were normalized as previously described [146, 153]. In order to convert relative protein abundances obtained from the mass spectrometry data to absolute cellular abundances (i.e. number of molecules per cell for each protein species), the average protein length and amino acid composition, and hence average molecular weight, were calculated from the relative abundances and known protein sequences. The molecular weight of the average JCVI-syn3A protein was then used to estimate the total number of all proteins in JCVI-syn3A based on the protein dry mass fraction and cellular dry weight (see Section 2.2). This estimated total number of proteins was used to scale relative abundances of proteins in the proteome to absolute abundances in the average cell of JCVI-syn3A. The mass spectrometry data has been deposited on MassIVE with accession number 000081687 (ftp://MSV000081687@massive.ucsd.edu, password: JCVISYN3A). The ProteomeXchange accession number is PXD008159. [154]

Omics scale visualization

Voronoi treemaps (Figures 1, 3, and 17a) were constructed following Liebermeister et al. [130]. Briefly, the genetic loci were associated with a KEGG orthology (KO) identifier [132]. Mappings between KO identifiers and locus tags were

acquired from KEGG Genomes for *M. pneumoniae* (T00006) and *E. coli* (T00944). A mapping between genes and KO identifiers for JCVI-syn3A was derived from the locus tag/KO map for *M. mycoides capri* LC str. 95010 (T01478) by matching *M. mycoides capri* genes to JCVI-syn3A genes using a reciprocal best hit BLASTp search [155]. Since the KO identifier, in general, can associate multiple functionality to a single ortholog, it was necessary to choose a single function for each ortholog. Initially, the KO/function assignment was taken from Liebermeister et al. [130]. Mycoplasma specific genes were then added to this hierarchy manually. Genes for which no ortholog could be assigned, but were well annotated in the genome were also added to the hierarchy manually. Voronoi treemaps were constructed by first using the freely available software described by Nocaj and Brandes [156] to generate the vertices of the polygons comprising the Voronoi tessellation, then rendering the resulting treemap using Cairo [157].

References

[1] Morowitz H (1984). The completeness of molecular biology. *Israel Journal of Medical Sciences*, **20**(9), pp. 750–753. doi:10.1002/ar.23031.

[2] Herrmann R (1992). 9. genome structure and organization. *American Society for Microbiology, Washington, DC (USA).*, pp. 157–168. doi:10.1007/s003359900222.

[3] Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchman JL, Weidman JF, Small KV, Sandusky M, Fuhrmann J, Nguyen D, Utterback TR, Saudek DM, Phillips CA, Merrick JM, Tomb JF, Dougherty BA, Bott KF, Hu PC, and Lucier TS (1995). The minimal gene complement of Mycoplasma genitalium. *Science*, **270**(5235), p. 197. doi:10.1126/science.270.5235.397.

[4] Pollack JD, Williams MV, and McElhaney RN (1997). The comparative metabolism of the mollicutes (mycoplasmas): The utility for taxonomic classification and the relationship of putative gene annotation and phylogeny to enzymatic function in the smallest free-living cells. *Critical Reviews in Microbiology*, **23**(4), pp. 269–354. doi:10.3109/10408419709115140.

[5] Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA, Smith HO, and Venter JC (2006). Essential genes of a minimal bacterium. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(2), pp. 425–430. doi:10.1073/pnas.0510013103.

[6] Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, McKenney K, Sutton G, FitzHugh W, Fields C, Gocayne JD, Scott J, Shirley R, lng Liu L, Glodek A, Kelley JM, Weidman JF, Phillips CA, Spriggs T, Hedblom E, Cotton MD, Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Brandon RC, Fine LD, Fritchman JL, Fuhrmann JL, Geoghagen NSM, Gnehm CL, McDonald LA, Small KV, Fraser CM, Smith HO, and Venter JC (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science*, **269**(5223), p. 496. doi:10.1126/science.7542800.

[7] Mushegian AR and Koonin EV (1996). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proceedings of the National Academy of Sciences*, **93**(19), pp. 10268–10273. doi:10.1073/pnas.93.19.10268.

[8] Gil R, Silva F, Pereto J, and Moya A (2004). Determination of the core of a minimal bacterial gene set. *Microbiology and Molecular Biology Reviews*, **68**(3), pp. 518–. doi:10.1128/MMBR.68.3.518-537.2004.

[9] Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, and Venter JC (1999). Global transposon mutagenesis and a minimal mycoplasma genome. *Science*, **286**(5447), pp. 2165–2169. doi:10.1126/science.286.5447.2165.

[10] Juhas M, Reu DR, Zhu B, and Commichau FM (2014). Bacillus subtilis and Escherichia coli essential genes and minimal cell factories after one decade of genome engineering. *Microbiology*, **160**(11), pp. 2341–2351. doi:10.1099/mic.0.079376-0.

[11] Pósfai G, Plunkett G, Fehér T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, Burland V, Harcum SW, and Blattner FR (2006). Emergent properties of reduced-genome Escherichia coli. *Science*, **312**(5776), pp. 1044–1046. doi:10.1126/science.1126439.

[12] Hutchison CA, Chuang RY, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L, Pelletier JF, Qi ZQ, Richter RA, Strychalski EA, Sun L, Suzuki Y, Tsvetanova B, Wise KS, Smith HO, Glass JI, Merryman C, Gibson DG, and Venter JC (2016). Design and synthesis of a minimal bacterial genome. *Science*, **351**(6280), p. aad6253. doi:10.1126/science.aad6253.

[13] Hutchison CA, Chuang RY, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L, Pelletier JF, Qi ZQ, Richter RA, Strychalski EA, Sun L, Suzuki Y, Tsvetanova B, Wise KS, Smith HO, Glass JI, Merryman C, Gibson DG, and Venter JC (2016). Synthetic bacterium JCVI-Syn3.0, complete genome. URL www.ncbi.nlm.nih.gov/nuccore/CP014940.1. Accession no. CP014940.1.

[14] Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, Hutchison CA, Smith HO, and Venter JC (2007). Genome transplantation in bacteria: Changing one species to another. *Science*, **317**(5838), pp. 632–638. doi:10.1126/science.1144622.

[15] Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA, Merryman C, Young L, Noskov VN, Glass JI, Venter JC, Hutchison CA, and Smith HO (2008). Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science*, **319**(5867), pp. 1215–1220. doi:10.1126/science.1151721.

[16] Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA, Smith HO, and Venter JC (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, **329**(5987), pp. 52–56. doi:10.1126/science.1190719.

[17] Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA, Smith HO, and Venter JC (2010). Synthetic Mycoplasma mycoides JCVI-syn1.0 clone sMmYCp235-1, complete sequence. URL www.ncbi.nlm.nih.gov/nuccore/CP002027.1. Accession no. CP002027.1.

[18] Glass JI (2017). Synthetic bacterium JCVI-Syn3.0 strain 6d, complete genome. URL www.ncbi.nlm.nih.gov/nuccore/CP016816.2. Accession no. CP016816.2.

[19] Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li B, Herrmann R, Dandekar T, Huynen M, Regula J, Ueberle B, Zimmermann C, Andrade M, Doerks T, Sanchez-Pulido L, Snel B, Suyama M, Yuan Y, Herrmann R, and Bork P (2014). Mycoplasma pneumoniae M129, complete genome. URL www.ncbi.nlm.nih.gov/nuccore/U00089.2. Accession no. U00089.2.

[20] Jeong H, Barbe V, Vallenet D, Choi SH, Lee C, Lee SW, Vacherie B, Yoon S, Yu DS, Cattolico L, Hur CG, Park HS, Segurens B, Blot M, Schneider D, Studier F, Oh T, Lenski R, Daegelen P, and Kim J (2017). Escherichia coli B str. REL606, complete genome. URL www.ncbi.nlm.nih.gov/nuccore/NC_012967.1. Accession no. NC_012967.1.

[21] Guell M, van Noort V, Yus E, Chen WH, Leigh-Bell J, Michalodimitrakis K, Yamada T, Arumugam M, Doerks T, Kuhner S, Rode M, Suyama M, Schmidt S, Gavin AC, Bork P, and Serrano L (2009). Transcriptome complexity in a genome-reduced bacterium. *Science*, **326**(5957), pp. 1268–1271. doi:10.1126/science.1176951.

[22] Kuhner S, van Noort V, Betts MJ, Leo-Macias A, Batisse C, Rode M, Yamada T, Maier T, Bader S, Beltran-Alvarez P, Castano-Diez D, Chen WH, Devos D, Guell M, Norambuena T, Racke I, Rybin V, Schmidt A, Yus E, Aebersold R, Herrmann R, Bottcher B, Frangakis AS, Russell RB, Serrano L, Bork P, and Gavin AC (2009). Proteome organization in a genome-reduced bacterium. *Science*, **326**(5957), pp. 1235–1240. doi:10.1126/science.1176343.

[23] Yus E, Maier T, Michalodimitrakis K, van Noort V, Yamada T, Chen WH, Wodke JAH, Guell M, Martinez S, Bourgeois R, Kuhner S, Raineri E, Letunic I, Kalinina OV, Rode M, Herrmann R, Gutierrez-Gallego R, Russell RB, Gavin AC, Bork P, and Serrano L (2009). Impact of genome reduction on bacterial metabolism and its regulation. *Science*, **326**(5957), pp. 1263–1268. doi:10.1126/science.1177263.

[24] Maier T, Schmidt A, Güell M, Kühner S, Gavin AC, Aebersold R, and Serrano L (2011). Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Molecular Systems Biology*, **7**(1), p. 511. doi:10.1038/msb.2011.38.

[25] Wodke JAH, Puchałka J, Lluch-Senar M, Marcos J, Yus E, Godinho M, Gutierrez-Gallego R, dos Santos VAPM, Serrano L, Klipp E, and Maier T (2013). Dissecting the energy metabolism in Mycoplasma pneumoniae through genome-scale metabolic modeling. *Molecular Systems Biology*, **9**(1), p. 653. doi:10.1038/msb.2013.6.

[26] Earnest TM, Lai J, Chen K, Hallock MJ, Williamson JR, and Luthey-Schulten Z (2015). Toward a whole-cell model of ribosome biogenesis: Kinetic modeling of ssu assembly. *Biophysical Journal*, **109**(6), pp. 1117–1135. doi:10.1016/j.bpj.2015.07.030.

[27] Earnest TM, Cole JA, Peterson JR, Hallock MJ, Kuhlman TE, and Luthey-Schulten Z (2016). Ribosome biogenesis in replicating cells: Integration of experiment and theory. *Biopolymers*, **105**(10), pp. 735–751. doi:10.1002/bip.22892.

[28] Orth JD, Thiele I, and Palsson BØ (2010). What is flux balance analysis? *Nature Biotechnology*, **28**(3), pp. 245–248. doi:10.1038/nbt.1614.

[29] Fuhrer T, Fischer E, and Sauer U (2005). Experimental identification and quantification of glucose metabolism in seven bacterial species. *Journal of bacteriology*, **187**(5), pp. 1581–1590.

[30] Shimizu T, Kimura Y, Kida Y, Kuwano K, Tachibana M, Hashino M, and Watarai M (2014). Cytadherence of Mycoplasma pneumoniae induces inflammatory responses through autophagy and toll-like receptor 4. *Infection and Immunity*, **82**(7), pp. 3076–3086. doi:10.1128/iai.01961-14.

[31] Alberts B (2011). A grand challenge in biology. *Science*, **333**(6047), pp. 1200–1200. doi:10.1126/science.1213238.

[32] Razin S, Argaman M, and Avigan J (1963). Chemical composition of mycoplasma cells and membranes. *Microbiology*, **33**(3), pp. 477–487. doi:10.1099/00221287-33-3-477.

[33] Bremer H and Dennis PP (2008). Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus*, **3**(1). doi:10.1128/ecosal.5.2.3.

[34] Mitchell A and Finch LR (1977). Pathways of nucleotide biosynthesis in Mycoplasma mycoides subsp. mycoides. *Journal of Bacteriology*, **130**(3), pp. 1047–1054. doi:10.1201/b16752-73.

[35] Manso-Silvan L, Vilei EM, Sachse K, Djordjevic S, Thiaucourt F, and Frey J (2009). Mycoplasma leachii sp. nov. as a new species designation for Mycoplasma sp. bovine group 7 of leach, and reclassification of Mycoplasma mycoides subsp. mycoides LC as a serovar of Mycoplasma mycoides subsp. capri. *International Journal of Systematic and Evolutionary Microbiology*, **59**(6), pp. 1353–1358. doi:10.1099/ijs.0.005546-0.

[36] Milo R (2013). What is the total number of protein molecules per cell volume? a call to rethink some published values. *BioEssays*, **35**(12), pp. 1050–1055.

[37] Archer D (1975). Modification of the membrane composition of Mycoplasma mycoides subsp. capri by the growth medium. *Microbiology*, **88**(2), pp. 329–338. doi:10.1099/00221287-88-2-329.

[38] Plackett P (1967). The glycerolipids of Mycoplasma mycoides. *Biochemistry*, **6**(9), pp. 2746–2754. doi:10.1021/bi00861a015.

[39] Schieck E, Lartigue C, Frey J, Vozza N, Hegermann J, Miller RA, Valguarnera E, Muriuki C, Meens J, Nene V, Naessens J, Weber J, Lowary TL, Vashee S, Feldman MF, and Jores J (2016). Galactofuranose in Mycoplasma mycoides is important for membrane integrity and conceals adhesins but does not contribute to serum resistance. *Molecular Microbiology*, **99**(1), pp. 55–70. doi:10.1111/mmi.13213.

[40] Bertin C, Pau-Roblot C, Courtois J, Manso-Silván L, Tardy F, Poumarat F, Citti C, Sirand-Pugnet P, Gaurivaud P, and Thiaucourt F (2015). Highly dynamic genomic loci drive the synthesis of two types of capsular or secreted polysaccharides within the Mycoplasma mycoides cluster. *Applied and environmental microbiology*, **81**(2), pp. 676–687. doi:10.1128/aem.02892-14.

[41] Rodwell A (1969). A defined medium for Mycoplasma strain Y. *Microbiology*, **58**(1), pp. 39–47. doi:10.1099/00221287-58-1-39.

[42] Leblanc G and Le Grimellec C (1979). Active K+ transport in Mycoplasma mycoides var. capri. net and unidirectional K+ movements. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, **554**(1), pp. 156–167. doi:10.1016/0005-2736(79)90015-4.

[43] Mitchell A and Finch LR (1979). Enzymes of pyrimidine metabolism in Mycoplasma mycoides subsp. mycoides. *Journal of Bacteriology*, **137**(3), pp. 1073–1080. doi:10.1201/b16752-73.

[44] Neale G, Mitchell A, and Finch LR (1983). Pathways of pyrimidine deoxyribonucleotide biosynthesis in Mycoplasma mycoides subsp. mycoides. *Journal of Bacteriology*, **154**(1), pp. 17–22. doi:10.1201/b16752-73.

[45] Linker C and Wilson TH (1985). Sodium and proton transport in Mycoplasma gallisepticum. *Journal of Bacteriology*, **163**(3), pp. 1250–1257. doi:10.20506/rst.28.3.1940.

[46] Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, and Palsson BØ (2011). A comprehensive genome-scale reconstruction of Escherichia coli metabolism. *Molecular Systems Biology*, **7**(1), p. 535. doi:10.1038/msb.2011.65.

[47] Glass JI, Smith HO, Hutchison CA, Alperovich NY, and Assad-Garcia N (2015). Minimal bacterial genome. U.S. Patent Application: 14/733,743.

[48] Tully JG, Rose DL, Whitcomb RF, and Wenzel RP (1979). Enhanced isolation of Mycoplasma pneumoniae from throat washings with a newly modified culture medium. *Journal of Infectious Diseases*, **139**(4), pp. 478–482. doi:10.1093/infdis/139.4.478.

[49] King ZA, Dräger A, Ebrahim A, Sonnenschein N, Lewis NE, and Palsson BØ (2015). Escher: A web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS Computational Biology*, **11**(8), p. e1004321. doi:10.1371/journal.pcbi.1004321.

[50] Postma P, Lengeler J, and Jacobson G (1993). Phosphoenolpyruvate: Carbohydrate phosphotransferase systems of bacteria. *Microbiological Reviews*, **57**(3), pp. 543–594. doi:10.1016/s0014-5793(01)02705-3.

[51] Lee D, Miles R, and Beezer A (1986). Isolation and microcalorimetric characterisation of glucose-negative and pyruvate-negative mutants of Mycoplasma mycoides subsp. mycoides. *FEMS Microbiology Letters*, **34**(3), pp. 283–286. doi:10.1016/0378-1097(86)90370-8.

[52] Ferrarini MG, Siqueira FM, Mucha SG, Palama TL, Jobard É, Elena-Herrmann B, Vasconcelos ATR, Tardy F, Schrank IS, Zaha A, and Sagot MF (2016). Insights on the virulence of swine respiratory tract mycoplasmas through genome-scale metabolic modeling. *BMC genomics*, **17**(1), p. 1. doi:10.1186/s12864-016-2644-z.

[53] Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, and Stevens R (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Research*, **42**(D1), pp. D206–D214. doi:10.1093/nar/gkt1226.

[54] Cocks BG, Brake FA, Mitchell A, and Finch LR (1985). Enzymes of intermediary carbohydrate metabolism in Ureaplasma urealyticum and Mycoplasma mycoides subsp. mycoides. *Microbiology*, **131**(9), pp. 2129–2135. doi:10.1099/00221287-131-9-2129.

[55] Mitchell A, *The ribonucleotides of Mycoplasma mycoides* (University of Melbourne, School of Biochemistry, 1976). doi:10.1201/b16752-73.

[56] Desantis D, Tryon VV, and Pollack JD (1989). Metabolism of mollicutes: The embden-meyerhof-parnas pathway and the hexose monophosphate shunt. *Microbiology*, **135**(3), pp. 683–691. doi:10.1099/00221287-135-3-683.

[57] Rashid N, Imanaka H, Fukui T, Atomi H, and Imanaka T (2004). Presence of a novel phosphopentomutase and a 2-deoxyribose 5-phosphate aldolase reveals a metabolic link between pentoses and central carbon metabolism in the hyperthermophilic archaeon thermococcus kodakaraensis. *Journal of Bacteriology*, **186**(13), pp. 4185–4191. doi:10.1128/jb.186.13.4185-4191.2004.

[58] Rodwell A and Rodwell ES (1954). The breakdown of carbohydrates by asterococcus mycoides, the organism of bovine pleuropneumonia. *Australian journal of biological sciences*, **7**(1), pp. 18–30.

[59] Rodwell A (1967). The nutrition and metabolism of mycoplasma: Progress and problems. *Annals of the New York Academy of Sciences*, **143**(1), pp. 88–109. doi:10.1111/j.1749-6632.1967.tb27649.x.

[60] Suthers PF, Dasika MS, Kumar VS, Denisov G, Glass JI, and Maranas CD (2009). A genome-scale metabolic reconstruction of Mycoplasma genitalium, iPS189. *PLoS Computational Biology*, **5**(2), p. e1000285. doi:10.1371/journal.pcbi.1000285.

[61] Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival Jr B, Assad-Garcia N, Glass JI, and Covert MW (2012). A whole-cell computational model predicts phenotype from genotype. *Cell*, **150**(2), pp. 389–401. doi:10.1016/j.cell.2012.05.044.

[62] Webb AJ and Hosie AH (2006). A member of the second carbohydrate uptake subfamily of ATP-binding cassette transporters is responsible for ribonucleoside uptake in Streptococcus mutans. *Journal of Bacteriology*, **188**(23), pp. 8005–8012. doi:10.1128/jb.01101-06.

[63] Neale G, Mitchell A, and Finch LR (1984). Uptake and utilization of deoxynucleoside 5′-monophosphates by Mycoplasma mycoides subsp. mycoides. *Journal of Bacteriology*, **158**(3), pp. 943–947. doi:10.1201/b16752-73.

[64] Youil R and Finch LR (1988). Isolation and characterization of Mycoplasma mycoides subsp. mycoides mutants deficient in nucleoside monophosphate transport. *Journal of Bacteriology*, **170**(12), pp. 5922–5924. doi:10.1128/jb.170.12.5922-5924.1988.

[65] Maniloff J, *Mycoplasmas: Molecular biology and pathogenesis* (Zondervan, 1992).

[66] Mitchell A, Sin IL, and Finch LR (1978). Enzymes of purine metabolism in Mycoplasma mycoides subsp. mycoides. *Journal of Bacteriology*, **134**(3), pp. 706–712. doi:10.1201/b16752-73.

[67] Neale G, Mitchell A, and Finch LR (1983). Enzymes of pyrimidine deoxyribonucleotide metabolism in Mycoplasma mycoides subsp. mycoides. *Journal of Bacteriology*, **156**(3), pp. 1001–1005. doi:10.1201/b16752-73.

[68] Cocks BG, Youil R, and Finch LR (1988). Comparison of enzymes of nucleotide metabolism in two members of the Mycoplasmataceae family. *International Journal of Systematic and Evolutionary Microbiology*, **38**(3), pp. 273–278. doi:10.1099/00207713-38-3-273.

[69] Wang L, Westberg J, Bölske G, and Eriksson S (2001). Novel deoxynucleoside-phosphorylating enzymes in mycoplasmas: Evidence for efficient utilization of deoxynucleosides. *Molecular Microbiology*, **42**(4), pp. 1065–1073. doi:10.1046/j.1365-2958.2001.02700.x.

[70] Welin M, Wang L, Eriksson S, and Eklund H (2007). Structure-function analysis of a bacterial deoxyadenosine kinase reveals the basis for substrate specificity. *Journal of Molecular Biology*, **366**(5), pp. 1615–1623. doi:10.1016/j.jmb.2006.12.010.

[71] Pollack JD, Myers MA, 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: A Journal of Integrative Biology*, **6**(3), pp. 247–258. doi:10.1089/15362310260256909.

[72] McElwain MC and Pollack JD (1987). Synthesis of deoxyribomononucleotides in mollicutes: Dependence on deoxyribose-1-phosphate and ppi. *Journal of Bacteriology*, **169**(8), pp. 3647–3653. doi:10.1128/jb.169.8.3647-3653.1987.

[73] el Hajj HH, Zhang H, and Weiss B (1988). Lethality of a dut (deoxyuridine triphosphatase) mutation in Escherichia coli. *Journal of Bacteriology*, **170**(3), pp. 1069–1075. doi:10.1128/jb.170.3.1069-1075.1988.

[74] Gadsden MH, McIntosh E, Game JC, Wilson PJ, and Haynes R (1993). dUTP pyrophosphatase is an essential enzyme in Saccharomyces cerevisiae. *The EMBO journal*, **12**(11), p. 4425. doi:10.1002/j.1460-2075.1993.tb06127.x.

[75] Wehelie R, Eriksson S, Bölske G, and Wang L (2010). Thymidylate synthases of Mycoplasma mycoides subsp. mycoides SC and Ureaplasma parvum are flavin-dependent. *Veterinary Microbiology*, **145**(3), pp. 265–272. doi:10.1016/j.vetmic.2010.03.029.

[76] Lieberman I (1956). Enzymatic amination of uridine triphosphate to cytidine triphosphate. *Journal of Biological Chemistry*, **222**(2), pp. 765–775. doi:10.1021/ja01614a107.

[77] Long CW and Pardee AB (1967). Cytidine triphosphate synthetase of Escherichia coli B I. purification and kinetics. *Journal of Biological Chemistry*, **242**(20), pp. 4715–4721. doi:10.3724/sp.j.1005.2012.00371.

[78] Scheit KH and Linke HJ (1982). Substrate specificity of CTP synthetase from Escherichia coli. *The FEBS Journal*, **126**(1), pp. 57–60. doi:10.1111/j.1432-1033.1982.tb06745.x.

[79] Willemoës M and Sigurskjold BW (2002). Steady-state kinetics of the glutaminase reaction of CTP synthase from Lactococcus lactis. *The FEBS Journal*, **269**(19), pp. 4772–4779. doi:10.1046/j.1432-1033.2002.03175.x.

[80] Pappas A, Park TS, and Carman GM (1999). Characterization of a novel dUTP-dependent activity of CTP synthetase from Saccharomyces cerevisiae. *Biochemistry*, **38**(50), pp. 16671–16677. doi:10.1021/bi9920127.

[81] Rodionov DA, Hebbeln P, Eudes A, ter Beek J, Rodionova IA, Erkens GB, Slotboom DJ, Gelfand MS, Osterman AL, Hanson AD, and Eitinger T (2009). A novel class of modular transporters for vitamins in prokaryotes. *Journal of Bacteriology*, **191**(1), pp. 42–51. doi:10.1128/jb.01208-08.

[82] Erkens GB, Majsnerowska M, ter Beek J, and Slotboom DJ (2012). Energy coupling factor-type abc transporters for vitamin uptake in prokaryotes. *Biochemistry*, **51**(22), pp. 4390–4396. doi:10.1021/bi300504v.

[83] Danchin A and Fang G (2016). Unknown unknowns: Essential genes in quest for function. *Microbial Biotechnology*, **9**(5), pp. 530–540. doi:10.1111/1751-7915.12384.

[84] Eudes A, Erkens GB, Slotboom DJ, Rodionov DA, Naponelli V, and Hanson AD (2008). Identification of genes encoding the folate- and thiamine-binding membrane proteins in firmicutes. *Journal of Bacteriology*, **190**(22), pp. 7591–7594. doi:10.1128/jb.01070-08.

[85] Neale G, Mitchell A, and Finch LR (1981). Formylation of methionyl-transfer ribonucleic acid in Mycoplasma mycoides subsp. mycoides. *Journal of Bacteriology*, **146**(2), pp. 816–818. doi:10.1201/b16752-73.

[86] Stover P and Schirch V (1990). Serine hydroxymethyltransferase catalyzes the hydrolysis of 5,10-methenyltetrahydrofolate to 5-formyltetrahydrofolate. *Journal of Biological Chemistry*, **265**(24), pp. 14227–14233. PMID: 2201683.

[87] Tarshis M and Salman M (1992). Uptake of a fluorescent-labeled fatty acid by spiroplasma floricola cells. *Archives of Microbiology*, **157**(3), pp. 258–263. doi:10.1007/bf00245159.

[88] McElhaney RN, De Gier J, and Van der Neut-Kok E (1973). The effect of alterations in fatty acid composition and cholesterol content on the nonelectrolyte permeability of acholeplasma laidlawii b cells and derived liposomes. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, **298**(2), pp. 500–512. doi:10.1016/0005-2736(73)90376-3.

[89] Eze MO and McElhaney RN (1981). The effect of alterations in the fluidity and phase state of the membrane lipids on the passive permeation and facilitated diffusion of glycerol in Escherichia coli. *Microbiology*, **124**(2), pp. 299–307. doi:10.1099/00221287-124-2-299.

[90] McLean L and Phillips M (1981). Mechanism of cholesterol and phosphatidylcholine exchange or transfer between unilamellar vesicles. *Biochemistry*, **20**(10), pp. 2893–2900. doi:10.1021/bi00513a028.

[91] Bittman R, Clejan S, and Hui SW (1990). Increased rates of lipid exchange between Mycoplasma capricolum membranes and vesicles in relation to the propensity of forming nonbilayer lipid structures. *Journal of Biological Chemistry*, **265**(25), pp. 15110–15117. doi:10.1002/bit.25864.

[92] Rigaud JL and Leblanc G (1980). Effect of membrane cholesterol on action of phospholipase A2 in Mycoplasma mycoides var. capri. *European Journal of Biochemistry*, **110**(1), pp. 77–84. doi:10.1111/j.1432-1033.1980.tb04842.x.

[93] Parsons JB, Broussard TC, Bose JL, Rosch JW, Jackson P, Subramanian C, and Rock CO (2014). Identification of a two-component fatty acid kinase responsible for host fatty acid incorporation by Staphylococcus aureus. *Proceedings of the National Academy of Sciences*, **111**(29), pp. 10532–10537. doi:10.1073/pnas.1408797111.

[94] Postic G, Danchin A, and Mechold U (2012). Characterization of NrnA homologs from mycobacterium tuberculosis and Mycoplasma pneumoniae. *RNA*, **18**(1), pp. 155–165. doi:10.1261/rna.029132.111.

[95] Dillon DA, Wu WI, Riedel B, Wissing JB, Dowhan W, and Carman GM (1996). The Escherichia coli pgpB gene encodes for a diacylglycerol pyrophosphate phosphatase activity. *Journal of Biological Chemistry*, **271**(48), pp. 30548–30553. doi:10.1074/jbc.271.48.30548.

[96] Detmers FJ, Kunji ER, Lanfermeijer FC, Poolman B, and Konings WN (1998). Kinetics and specificity of peptide uptake by the oligopeptide transport system of Lactococcus lactis. *Biochemistry*, **37**(47), pp. 16671–16679. doi:10.1021/bi981712t.

[97] Doeven MK, Abele R, Tampé R, and Poolman B (2004). The binding specificity of OppA determines the selectivity of the oligopeptide ATP-binding cassette transporter. *Journal of Biological Chemistry*, **279**(31), pp. 32301–32307. doi:10.1074/jbc.m404343200.

[98] Tolner B, Ubbink-Kok T, Poolmann B, and Konings WN (1995). Cation-selectivity of the l-glutamate transporters of Escherichia coli, Bacillus stearothermophilus and Bacillus caldotenax: Dependence on the environment in which the proteins are expressed. *Molecular Microbiology*, **18**(1), pp. 123–133. doi:10.1111/j.1365-2958.1995.mmi_18010123.x.

[99] Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, and Palsson BØ (2007). A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, **3**(1), p. 121. doi:10.1038/msb4100155.

[100] Neidhardt FC, Ingraham JL, and Schaechter M, *Physiology of the bacterial cell: A molecular approach*, vol. 20 (Sinauer Sunderland, 1990). doi:10.1038/348401a0.

[101] Frottin F, Martinez A, Peynot P, Mitra S, Holz RC, Giglione C, and Meinnel T (2006). The proteomics of N-terminal methionine cleavage. *Molecular & Cellular Proteomics*, **5**(12), pp. 2336–2349. doi:10.1074/mcp.M600225-MCP200.

[102] Gur E and Sauer RT (2008). Evolution of the ssrA degradation tag in mycoplasma: specificity switch to a different protease. *Proceedings of the National Academy of Sciences*, **105**(42), pp. 16113–16118. doi:10.1073/pnas.0808802105.

[103] Ge Z and Karzai AW (2009). Co-evolution of multipartite interactions between an extended tmRNA tag and a robust Lon protease in mycoplasma. *Molecular Microbiology*, **74**(5), pp. 1083–1099. doi:10.1111/j.1365-2958.2009.06923.x.

[104] Westermann AJ, Gorski SA, and Vogel J (2012). Dual rna-seq of pathogen and host. *Nature Reviews Microbiology*, **10**(9), p. 618.

[105] Rosenow C, Saxena RM, Durst M, and Gingeras TR (2001). Prokaryotic rna preparation methods useful for high density array analysis: comparison of two approaches. *Nucleic acids research*, **29**(22), pp. e112–e112.

[106] Kirk R and Morowitz H (1969). Ribonucleic acids of mycoplasma gallisepticum strain a5969. *American journal of veterinary research*, **30**(2), pp. 287–293.

[107] Beri D, Balan B, Chaubey S, Subramaniam S, Surendra B, and Tatu U (2017). A disrupted transsulphuration pathway results in accumulation of redox metabolites and induction of gametocytogenesis in malaria. *Scientific Reports*, **7**, p. 40213. doi:10.1038/srep40213.

[108] Raeymaekers L, Wuytack E, Willems I, Michiels C, and Wuytack F (2002). Expression of a P-type Ca 2+-transport ATPase in Bacillus subtilis during sporulation. *Cell Calcium*, **32**(2), pp. 93–103. doi:10.1016/s0143-4160(02)00125-2.

[109] Schummer U and Schiefer H (1983). Electrophysiology of mycoplasma membranes. *The Yale Journal of Biology and Medicine*, **56**(5-6), p. 413. doi:10.1002/0471142727.mb2805s106.

[110] Stover P and Schirch V (1992). Enzymic mechanism for the hydrolysis of 5,10-methenyltetrahydropteroylglutamate to 5-formyltetrahydropteroylglutamate by serine hydroxymethyltransferase. *Biochemistry*, **31**(7), pp. 2155–2164. doi:10.1021/bi00122a037.

[111] Feist AM, Scholten JC, Palsson BØ, Brockman FJ, and Ideker T (2006). Modeling methanogenesis with a genome-scale metabolic reconstruction of methanosarcina barkeri. *Molecular systems biology*, **2**(1).

[112] Thiele I and Palsson BØ (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, **5**(1), pp. 93–121. doi:10.1038/nprot.2009.203.

[113] Benyoucef M, Rigaud JL, and Leblanc G (1981). The electrochemical proton gradient in mycoplasma cells. *The FEBS Journal*, **113**(3), pp. 491–498. doi:10.1111/j.1432-1033.1981.tb05090.x.

[114] Benyoucef M, Rigaud JL, and Leblanc G (1981). Gradation of the magnitude of the electrochemical proton gradient in mycoplasma cells. *European Journal of Biochemistry*, **113**(3), pp. 499–506. doi:10.1111/j.1432-1033.1981.tb05091.x.

[115] Benyoucef M, Rigaud JL, and Leblanc G (1982). Cation transport mechanisms in Mycoplasma mycoides var. capri cells. Na+-dependent K+ accumulation. *Biochemical Journal*, **208**(3), pp. 529–538. doi:10.1042/bj2080529.

[116] Labhsetwar P, Cole JA, Roberts E, Price ND, and Luthey-Schulten ZA (2013). Heterogeneity in protein expression induces metabolic variability in a modeled Escherichia coli population. *Proceedings of the National Academy of Sciences*, **110**(34), pp. 14006–14011. doi:10.1073/pnas.1222569110.

[117] Labhsetwar P, Melo MC, Cole JA, and Luthey-Schulten Z (2017). Population FBA predicts metabolic phenotypes in yeast. *PLOS Computational Biology*, **13**(9), p. e1005728. doi:10.1371/journal.pcbi.1005728.

[118] Scheer M, Grote A, Chang A, Schomburg I, Munaretto C, Rother M, Söhngen C, Stelzer M, Thiele J, and Schomburg D (2010). BRENDA, the enzyme information system in 2011. *Nucleic Acids Research*, **39**(suppl_1), pp. D670–D676. doi:10.1093/nar/gkq1089.

[119] Ben-Menachem G, Himmelreich R, Herrmann R, Aharonowitz Y, and Rottem S (1997). The thioredoxin reductase system of mycoplasmas. *Microbiology*, **143**(6), pp. 1933–1940. doi:10.1099/00221287-143-6-1933.

[120] Hermolin J and Fillingame R (1989). H+-ATPase activity of Escherichia coli F1F0 is blocked after reaction of dicyclohexylcarbodiimide with a single proteolipid (subunit c) of the F0 complex. *Journal of Biological Chemistry*, **264**(7), pp. 3896–3903. doi:10.1111/j.1432-1033.2004.04143.x.

[121] Linster CL, Van Schaftingen E, and Hanson AD (2013). Metabolite damage and its repair or pre-emption. *Nature Chemical Biology*, **9**(2), p. 72. doi:10.1038/nchembio.1141.

[122] Ebrahim A, Lerman JA, Palsson BØ, and Hyduke DR (2013). COBRApy: Constraints-based reconstruction and analysis for Python. *BMC Systems Biology*, **7**(1), p. 74. doi:10.1186/1752-0509-7-74.

[123] Vanyushkina AA, Fisunov GY, Gorbachev AY, Kamashev DE, and Govorun VM (2014). Metabolomic analysis of three mollicute species. *PLoS One*, **9**(3), p. e89312. doi:10.1371/journal.pone.0089312.

[124] Weinert BT, Iesmantavicius V, Wagner SA, Schölz C, Gummesson B, Beli P, Nyström T, and Choudhary C (2013). Acetyl-phosphate is a critical determinant of lysine acetylation in E. coli. *Molecular Cell*, **51**(2), pp. 265–272.

[125] Miles R, Beezer A, and Lee D (1985). Kinetics of utilization of organic substrates by mycoplasma mycoides subsp. mycoides in a salts solution: a flow-microcalorimetric study. *Microbiology*, **131**(8), pp. 1845–1852.

[126] Shahram M, Nicholas RA, Miles RJ, Wood AP, and Kelly DP (2009). Kinetics of substrate oxidation and hydrogen peroxide production by mycoplasma mycoides subsp. mycoides large colony (lc) type and mycoplasma mycoides subsp. capri. *Research in veterinary science*, **87**(3), pp. 364–366.

[127] Persaud C, Lu Y, Vila-Sanjurjo A, Campbell JL, Finley J, and OConnor M (2010). Mutagenesis of the modified bases, m^5U1939 and $\psi 2504$, in Escherichia coli 23S rRNA. *Biochemical and Biophysical Research Communications*, **392**(2), pp. 223–227. doi:10.1016/j.bbrc.2010.01.021.

[128] Burakovsky DE, Prokhorova IV, Sergiev PV, Milon P, Sergeeva OV, Bogdanov AA, Rodnina MV, and Dontsova OA (2012). Impact of methylations of m^2G966/m^5C967 in 16S rRNA on bacterial fitness and translation initiation. *Nucleic Acids Research*, **40**(16), pp. 7885–7895. doi:10.1093/nar/gks508.

[129] Arora S, Bhamidimarri SP, Bhattacharyya M, Govindan A, Weber MH, Vishveshwara S, and Varshney U (2013). Distinctive contributions of the ribosomal P-site elements m^2G966, m^5C967 and the C-terminal tail of the S9 protein in the fidelity of initiation of translation in Escherichia coli. *Nucleic Acids Research*, **41**(9), pp. 4963–4975. doi:10.1093/nar/gkt175.

[130] Liebermeister W, Noor E, Flamholz A, Davidi D, Bernhardt J, and Milo R (2014). Visual account of protein investment in cellular functions. *Proceedings of the National Academy of Sciences*, **111**(23), pp. 8488–8493. doi:10.1073/pnas.1314810111.

[131] Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M, Paley S, Rhee SY, Shearer AG, Tissier C, Walk TC, Zhang P, and Karp PD (2008). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research*, **36**(Database), pp. D623–D631. doi:10.1093/nar/gkm900.

[132] Kanehisa M, Sato Y, Kawashima M, Furumichi M, and Tanabe M (2015). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, **44**(D1), pp. D457–D462. doi:10.1093/nar/gkv1070.

[133] Schellenberger J, Park JO, Conrad TM, and Palsson BØ (2010). BiGG: A biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, **11**(1), p. 213. doi:10.1186/1471-2105-11-213.

[134] King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, Ebrahim A, Palsson BØ, and Lewis NE (2016). BiGG models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, **44**(D1), pp. D515–D522. doi:10.1093/nar/gkv1049.

[135] Wohlgemuth G, Haldiya PK, Willighagen E, Kind T, and Fiehn O (2010). The chemical translation servicea web-based tool to improve standardization of metabolomic reports. *Bioinformatics*, **26**(20), pp. 2647–2648.

[136] Noor E, Haraldsdóttir HS, Milo R, and Fleming RM (2013). Consistent estimation of Gibbs energy using component contributions. *PLoS Computational Biology*, **9**(7), p. e1003098. doi:10.1371/journal.pcbi.1003098.

[137] Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G, Purvine SO, Lopez-Ferrer D, Weitz KK, Eils R, König R, Smith RD, and Palsson BØ (2010). Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology*, **6**(1), p. 390. doi:10.1038/msb.2010.47.

[138] Wright MS, Mountain S, Beeri K, and Adams MD (2017). Assessment of insertion sequence mobilization as an adaptive response to oxidative stress in Acinetobacter baumannii using IS-seq. *Journal of Bacteriology*, **199**(9), pp. e00833–16. doi:10.1128/jb.00833-16.

[139] Titterington DM, The em algorithm, variational approximations and expectation propagation for mixtures. In KL Mengersen, CP Robert, and DM Titterington, eds., *Mixtures: Estimation and Applications*, chap. 1, pp. 1–29 (John Wiley & Sons, Ltd, 2011). doi:10.1002/9781119995678.

[140] Jones E, Oliphant T, and Peterson P (2001–). SciPy: Open source scientific tools for Python. [Online; accessed 2016-01-19].

[141] Villén J and Gygi SP (2008). The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nature Protocols*, **3**(10), pp. 1630–1638. doi:10.1038/nprot.2008.150.

[142] Haas W, Faherty BK, Gerber SA, Elias JE, Beausoleil SA, Bakalarski CE, Li X, Villen J, and Gygi SP (2006). Optimization and use of peptide mass measurement accuracy in shotgun proteomics. *Molecular & Cellular Proteomics*, **5**(7), pp. 1326–1337. doi:10.1074/mcp.m500339-mcp200.

[143] Wessel D and Flügge U (1984). A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical Biochemistry*, **138**(1), pp. 141–143. doi:10.1016/0003-2697(84)90782-6.

[144] Tolonen AC and Haas W (2014). Quantitative proteomics using reductive dimethylation for stable isotope labeling. *Journal of Visualized Experiments: JoVE*, (89). doi:10.3791/51416.

[145] Wang Y, Yang F, Gritsenko MA, Wang Y, Clauss T, Liu T, Shen Y, Monroe ME, Lopez-Ferrer D, Reno T, Moore RJ, Klemke RL, Camp DG, and Smith RD (2011). Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human mcf10a cells. *Proteomics*, **11**(10), pp. 2019–2026. doi:10.1002/pmic.201000722.

[146] Lapek JD, Lewinski MK, Wozniak JM, Guatelli J, and Gonzalez DJ (2017). Quantitative temporal viromics of an inducible hiv-1 model yields insight to global host targets and phospho-dynamics associated with protein vpr. *Molecular & Cellular Proteomics*, **16**(8), pp. 1447–1461. doi:10.1074/mcp.m116.066019.

[147] Eng JK, McCormack AL, and Yates JR (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*, **5**(11), pp. 976–989. doi:10.1016/1044-0305(94)80016-2.

[148] Peng J, Elias JE, Thoreen CC, Licklider LJ, and Gygi SP (2003). Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC- MS/MS) for large-scale protein analysis: The yeast proteome. *Journal of Proteome Research*, **2**(1), pp. 43–50. doi:10.1021/pr025556v.

[149] Elias JE, Haas W, Faherty BK, and Gygi SP (2005). Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nature Methods*, **2**(9), pp. 667–675. doi:10.1038/nmeth785.

[150] Elias JE and Gygi SP (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods*, **4**(3), pp. 207–214. doi:10.1038/nmeth1019.

[151] Huttlin EL, Jedrychowski MP, Elias JE, Goswami T, Rad R, Beausoleil SA, Villén J, Haas W, Sowa ME, and Gygi SP (2010). A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell*, **143**(7), pp. 1174–1189. doi:10.1016/j.cell.2010.12.001.

[152] Ting L, Rad R, Gygi SP, and Haas W (2011). Ms3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nature Methods*, **8**(11), pp. 937–940. doi:10.1038/nmeth.1714.

[153] Lapek Jr JD, Greninger P, Morris R, Amzallag A, Pruteanu-Malinici I, Benes CH, and Haas W (2017). Detection of dysregulated protein-association networks by high-throughput proteomics predicts cancer vulnerabilities. *Nature Biotechnology*, **35**(10), p. 983. doi:10.1038/nbt.3955.

[154] Lapek JD and Gonzalez DJ (2017). Proteomic analysis of JCVI-Syn3A. URL `ftp://MSV000081687@massive.ucsd.edu`. Accession no. 000081687.

[155] Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ (1990). Basic local alignment search tool. *Journal of Molecular Biology*, **215**(3), pp. 403–410. doi:10.1006/jmbi.1990.9999.

[156] Nocaj A and Brandes U (2012). Computing Voronoi treemaps: Faster, simpler, and resolution-independent. *Compututer Graphics Forum*, **31**(3pt1), pp. 855–864. doi:10.1111/j.1467-8659.2012.03078.x.

[157] Packard K, Worth C, Esfahbod B, et al. (2013–2018). Cairo. https://cairographics.org/.

[158] Lazarev V, Levitskii S, Basovskii YI, Chukin M, Akopian T, Vereshchagin V, Kostrjukova E, Kovaleva G, Kazanov M, Malko D, et al. (2011). The complete genome and proteome of acholeplasma laidlawii. *Journal of bacteriology*, pp. JB–05059.

[159] Folmsbee M, Howard G, and McAlister M (2010). Nutritional effects of culture media on mycoplasma cell size and removal by filtration. *Biologicals*, **38**(2), pp. 214–217.

[160] Bratbak G and Dundas I (1984). Bacterial dry matter content and biomass estimations. *Applied and environmental microbiology*, **48**(4), pp. 755–757. doi:10.1007/bf01347859.

[161] Cooper S and Helmstetter CE (1968). Chromosome replication and the division cycle of escherichia colibr. *Journal of Molecular Biology*, **31**(3), pp. 519 – 540. ISSN 0022-2836. doi:https://doi.org/10.1016/0022-2836(68)90425-7. URL http://www.sciencedirect.com/science/article/pii/0022283668904257.

[162] Andrés E, Martínez N, and Planas A (2011). Expression and characterization of a Mycoplasma genitalium glycosyltransferase in membrane glycolipid biosynthesis potential target against Mycoplasma infections. *Journal of Biological Chemistry*, **286**(41), pp. 35367–35379. doi:10.1074/jbc.m110.214148.

[163] Klement MLR, Öjemyr L, Tagscherer KE, Widmalm G, and Wieslander Å (2007). A processive lipid glycosyltransferase in the small human pathogen Mycoplasma pneumoniae: Involvement in host immune response. *Molecular Microbiology*, **65**(6), pp. 1444–1457. doi:10.1111/j.1365-2958.2007.05865.x.

[164] Razin S (1969). Structure and function in mycoplasma. *Annual Reviews in Microbiology*, **23**(1), pp. 317–356. doi:10.1146/annurev.mi.23.100169.001533.

[165] McElhaney RN and Tourtellotte ME (1969). Mycoplasma membrane lipids: Variations in fatty acid composition. *Science*, **164**(3878), pp. 433–434. doi:10.1126/science.164.3878.433.

[166] Razin S, Yogev D, and Naot Y (1998). Molecular biology and pathogenicity of mycoplasmas. *Microbiology and Molecular Biology Reviews*, **62**(4), pp. 1094–1156. doi:10.1007/b113360.

[167] Bertin C, Pau-Roblot C, Courtois J, Manso-Silván L, Thiaucourt F, Tardy F, Le Grand D, Poumarat F, and Gaurivaud P (2013). Characterization of free exopolysaccharides secreted by Mycoplasma mycoides subsp. mycoides. *PLoS One*, **8**(7), p. e68373. doi:10.1371/journal.pone.0068373.

[168] Morgan JL, Strumillo J, and Zimmer J (2013). Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature*, **493**(7431), pp. 181–186. doi:10.1038/nature11744.

[169] Wheatley RW, Zheng RB, Richards MR, Lowary TL, and Ng KK (2012). Tetrameric structure of the GlfT2 galactofuranosyltransferase reveals a scaffold for the assembly of mycobacterial arabinogalactan. *Journal of Biological Chemistry*, **287**(33), pp. 28132–28143. doi:10.1074/jbc.m112.347484.

[170] Daubenspeck JM, Bolland JR, Luo W, Simmons WL, and Dybvig K (2009). Identification of exopolysaccharide-deficient mutants of Mycoplasma pulmonis. *Molecular Microbiology*, **72**(5), pp. 1235–1245. doi:10.1111/j.1365-2958.2009.06720.x.

[171] Plackett P (1967). The synthesis of polar lipids by mycoplasma. *Annals of the New York Academy of Sciences*, **143**(1), pp. 158–164. doi:10.1046/j.1432-1327.2000.01709.x.

[172] Saxena IM, Brown Jr RM, Fevre M, Geremia RA, and Henrissat B (1995). Multidomain architecture of beta-glycosyl transferases: implications for mechanism of action. *Journal of Bacteriology*, **177**(6), p. 1419. doi:10.1128/jb.177.6.1419-1424.1995.

[173] Buttery S and Plackett P (1960). A specific polysaccharide from Mycoplasma mycoides. *Microbiology*, **23**(2), pp. 357–368. doi:10.1099/00221287-23-2-357.

[174] Jordan DS, Daubenspeck JM, and Dybvig K (2013). Rhamnose biosynthesis in mycoplasmas requires precursor glycans larger than monosaccharide. *Molecular Microbiology*, **89**(5), pp. 918–928. doi:10.1111/mmi.12320.

[175] Daubenspeck JM, Liu R, and Dybvig K (2016). Rhamnose links moonlighting proteins to membrane phospholipid in mycoplasmas. *PLoS One*, **11**(9), p. e0162505. doi:10.1371/journal.pone.0162505.

[176] Luecke H and Quiocho FA (1990). High specificity of a phosphate transport protein determined by hydrogen bonds. *Nature*, **347**(6291), p. 402. doi:10.1002/pro.2449.

[177] Elias M, Wellner A, Goldin-Azulay K, Chabriere E, Vorholt JA, Erb TJ, and Tawfik DS (2012). The molecular basis of phosphate discrimination in arsenate-rich environments. *Nature*, **491**(7422), pp. 134–137. doi:10.1107/s0108767313099583.

[178] Raven JA, *Energetics and transport in aquatic plants* (AR Liss, 1984). doi:10.4319/lo.1974.19.6.0912.

[179] Voorde JV, Gago F, Vrancken K, Liekens S, and Balzarini J (2012). Characterization of pyrimidine nucleoside phosphorylase of Mycoplasma hyorhinis: Implications for the clinical efficacy of nucleoside analogues. *Biochemical Journal*, **445**(1), pp. 113–123. doi:10.1042/bj20112225.

[180] Sippel KH, Robbins AH, Reutzel R, Boehlein SK, Namiki K, Goodison S, Agbandje-McKenna M, Rosser CJ, and McKenna R (2009). Structural insights into the extracytoplasmic thiamine-binding lipoprotein p37 of Mycoplasma hyorhinis. *Journal of Bacteriology*, **191**(8), pp. 2585–2592. doi:10.1128/jb.01680-08.

[181] Sippel KH, Venkatakrishnan B, Boehlein SK, Sankaran B, Quirit JG, Govindasamy L, Agbandje-McKenna M, Goodison S, Rosser CJ, and McKenna R (2011). Insights into Mycoplasma genitalium metabolism revealed by the structure of mg289, an extracytoplasmic thiamine binding lipoprotein. *Proteins: Structure, Function, and Bioinformatics*, **79**(2), pp. 528–536. doi:10.1002/prot.22900.

[182] Russell RB and Barton GJ (1992). Multiple protein sequence alignment from tertiary structure comparison: Assignment of global and residue confidence levels. *Proteins: Structure, Function, and Bioinformatics*, **14**(2), pp. 309–323. doi:10.1002/prot.340140216.

[183] Humphrey W, Dalke A, and Schulten K (1996). VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, **14**, pp. 33–38. doi:10.1016/0263-7855(96)00018-5.

[184] Eargle J, Wright D, and Luthey-Schulten Z (2006). Multiple alignment of protein structures and sequences for VMD. *Bioinformatics*, **22**(4), pp. 504–506. doi:10.1093/bioinformatics/bti825.

[185] Paik SH, Chakicherla A, and Hansen JN (1998). Identification and characterization of the structural and transporter genes for, and the chemical and biological properties of, sublancin 168, a novel lantibiotic produced by Bacillus subtilis 168. *Journal of Biological Chemistry*, **273**(36), pp. 23134–23142. doi:10.1074/jbc.273.36.23134.

[186] Dirix G, Monsieurs P, Marchal K, Vanderleyden J, and Michiels J (2004). Screening genomes of gram-positive bacteria for double-glycine-motif-containing peptides. *Microbiology*, **150**(5), pp. 1121–1126. doi:10.1099/mic.0.27040-0.

[187] Menon AS, Waxman L, and Goldberg AL (1987). The energy utilized in protein breakdown by the ATP-dependent protease (La) from Escherichia coli. *Journal of Biological Chemistry*, **262**(2), pp. 722–726. doi:10.1046/j.0014-2956.2001.02668.x.

[188] Benyoucef M, Rigaud JL, and Leblanc G (1982). Cation transport mechanisms in Mycoplasma mycoides var. capri cells. the nature of the link between K+ and Na+ transport. *Biochemical Journal*, **208**(3), pp. 539–547. doi:10.1042/bj2080539.

[189] Tholema N, der Brüggen MV, Mäser P, Nakamura T, Schroeder JI, Kobayashi H, Uozumi N, and Bakker EP (2005). All four putative selectivity filter glycine residues in KtrB are essential for high affinity and selective K+ uptake by the KtrAB system from vibrio alginolyticus. *Journal of Biological Chemistry*, **280**(50), pp. 41146–41154. doi:10.1074/jbc.m507647200.

[190] Letunic I and Bork P (2016). Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, **44**(W1), pp. W242–W245. doi:10.1093/nar/gkw290.

[191] Ciccarelli FD, Doerks T, Von Mering C, Creevey CJ, Snel B, and Bork P (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science*, **311**(5765), pp. 1283–1287. doi:10.1126/science.1123061.

[192] Sirand-Pugnet P, Lartigue C, Marenda M, Jacob D, Barré A, Barbe V, Schenowitz C, Mangenot S, Couloux A, Segurens B, et al. (2007). Being pathogenic, plastic, and sexual while living with a nearly minimal bacterial genome. *PLoS genetics*, **3**(5), p. e75.

APPENDIX 1-TABLE 1 Reconstructed biomass composition of JCVI-syn3A, listing the fraction of each component as percent of cellular dry mass.

Category	Component	Fraction [%]
Macromolecules	Protein	54.727
Total: 76.521 %	RNA	16.274
	DNA	5.5
	Acyl carrier protein	0.018
	dUTPase	0.003
Lipids & capsule	Lipogalactan capsule	6.368
Total: 17.563 %	Phosphatidylglycerol	2.944
	Cardiolipin	2.944
	Cholesterol	1.534
	Diacylglycerol	1.366
	Gal-DAG	1.31
	Triacylglycerol	0.549
	Fatty acid	0.549
Small molecules & ions	Potassium	3.285
Total: 5.916 %	Phosphate	0.375
	Chloride	0.198
	ATP	0.167
	L-lysine	0.133
	Sodium	0.131
	Spermine	0.128
	L-isoleucine	0.115
	GTP	0.115
	L-leucine	0.112
	UTP	0.106
	L-glutamate	0.085
	L-valine	0.083
	L-asparagine	0.083

Category	Component	Fraction [%]
Small molecules	L-alanine	0.077
& ions (cont'd)	L-aspartate	0.072
	L-threonine	0.071
	L-serine	0.071
	Glycine	0.068
	CTP	0.053
	L-phenylalanine	0.049
	L-glutamine	0.048
	L-arginine	0.043
	L-tyrosine	0.041
	L-proline	0.035
	L-methionine	0.026
	Magnesium	0.019
	L-histidine	0.019
	Calcium	0.019
	FAD	0.016
	5,10-meTHF$(Glu)_3$	0.015
	CoA	0.012
	Thiamin diphosphate	0.009
	NADP+	0.008
	L-tryptophan	0.008
	L-cysteine	0.008
	Pyridoxal phosphate	0.005
	dTTP	0.003
	dATP	0.003
	dCTP	0.002
	dGTP	0.001

Appendix 1

Further model details and justification

In the following, we discuss specific aspects of the derivation of the biomass composition and of the reconstruction of the metabolic network that were not covered in the main text.

Biomass composition

Appendix 1-Table 1 lists the mass fractions for all components included in the JCVI-syn3A biomass composition (Figure 4). Mass fractions for the different biomass components are obtained from various sources; since the mass fractions thus determined initially add up to ∼106 %, we finally rescale all mass fractions to a total of 100 %. (In doing so, we keep the DNA fraction fixed to ensure a total DNA mass corresponding to one chromosome.) These rescaled numbers are the ones shown in Appendix 1-Table 1.

Applicability of macromolecular composition The macromolecular composition is based on *M. mycoides capri* [32] whose genome is approximately twice as large as the one of JCVI-syn3A. However, the protein dry mass fraction in *M. mycoides capri* appears to be a good initial approximation for JCVI-syn3A. A protein dry mass fraction of ∼40–60 % is generally observed in bacteria, e.g. in different mycoplasmas [32] or *E. coli* [33]. In particular, we note that *Acholeplasma laidlawii* PG8, in spite of having a 1.5-times larger genome [158] than *M. mycoides capri* but a

comparable cell size [159], shows a similar protein dry mass fraction (55 %) to *M. mycoides capri* (58 %) [32].

While no such argument can be made *per se* for the conservation of the RNA content, we note that the assumed RNA dry mass fraction agrees reasonably well with ribosomal protein abundance from our proteomics data: The average copy number across all ribosomal proteins in JCVI-syn3A comes out at 340 copies. If this number was interpreted as an estimate of the total number of ribosomes, it would come reasonably close to the upper limit of ∼670 ribosomes per average cell if all RNA was present as ribosomal RNA (see Section 2.2.1). The presence of the same number of rRNA operons (2) in both JCVI-syn1.0 and JCVI-syn3A is also consistent with the assumption of comparable rRNA contents in the two organisms. Thus, the RNA content from *M. mycoides capri* should provide a reasonable approximation for the RNA content in JCVI-syn3A.

The only required adaptation was a slight increase in the DNA mass fraction from the 5 % reported for *M. mycoides capri* [32] to 5.5 %, since this corresponds to exactly one chromosome of a 543,379 bp genome in a 400 nm spherical cell with a dry weight of ∼10.2 fg. The dry weight is obtained assuming a density of 1.1 g/ml, which has been found in different bacterial cells [160]; and around 4.8 μl water/mg cellular protein as measured in *M. mycoides capri* serovar capri PG3 [42], corresponding to a cellular water content of 72 % in JCVI-syn3A. Bacterial cells in exponential growth phase can on average contain more than one chromosome [161]; until such data becomes available for JCVI-syn3A though, we stick to the assumption of one chromosome per cell, since this stays close to the DNA dry mass fraction of 5 % reported for *M. mycoides capri*. (E.g., 1.5 chromosomes in an average cell would already imply a dry mass fraction of 8.25 %.) As an aside, the above means that the reported DNA fraction for *M. mycoides capri* [32] is lower than expected for a whole *M. mycoides capri* chromosome. A similar discrepancy has been observed in the reconstruction of the *M. genitalium* biomass [61].

Details on lipid composition The overall lipid composition of *M. mycoides capri* serovar capri PG3 has been studied previously [37] and found to comprise phospholipids, glycolipids, cholesterol, free fatty acids and (mono-,di-,tri-)glycerides. Mono- and diglycerides could not be distinguished and are included as diglycerides in the model. In addition, it has been shown that the only phospholipids produced by *M. mycoides capri* LC Y are phosphatidylglycerol and cardiolipin; no phosphatidylcholine or -ethanolamine are synthesized, although lecithin can be incorporated as a whole if present in the media [38]. The phospholipid fraction is assumed to be equal parts phosphatidylglycerol and cardiolipin. The same study also identified the glycolipid as monogalactosyl-diacylglycerol (Gal-DAG) [38]. JCVI-syn3A contains the pathway from glucose-6-phosphate to UDP-galactofuranose, and the glycosyltransferases *cps*/0114 and 0697 are both 20 % identical to the experimentally confirmed glucosyl-/galactosyltransferases MG517 from *M. genitalium* [162] and MPN483 from *M. pneumoniae* [163]. We therefore assume that the glycolipid in JCVI-syn3A is a Gal-DAG as well.

For the fatty acids, palmitic acid (C16:0) and oleic acid (C18:1 cis-9) are considered to be the two most important representatives. While it is known that the abundance of different fatty acids in mycoplasma cell membranes can be affected by the media composition [37, 164, 165], we note that these two fatty acids are the only ones in the minimal media for *M. mycoides capri* LC Y [41] and *M. pneumoniae* [23] and also yielded among the highest growth rates and cell yields in a screen of fatty acid combinations [59]. We thus define a metabolite species "fatty acid" with an average molecular weight between palmitate and oleate, which is used in all lipid species.

Genetic evidence for capsule production in JCVI-syn3A While mycoplasmas lack a cell wall [166], several *Mycoplasma* spp. do produce capsular polysaccharides (CPS) or secrete polysaccharides into the medium (exopolysaccharides, EPS) [40]. In particular, *M. mycoides capri* LC GM12, the strain from which JCVI-syn3A is derived, has

been shown to produce a galactan (specifically, poly-β-1→6-galactofuranose) [39]; and other *M. mycoides capri* LC strains have been demonstrated to produce a galactan CPS but secrete negligible amounts of EPS [40]. This galactan has been suggested to play a role in membrane integrity in *M. mycoides capri* LC GM12, and deletion of the *glf* gene decreases the growth rate, possibly due to increased energy expenses to maintain cell homeostasis [39]. While it is not yet experimentally known whether the minimal cell still produces this galactan polysaccharide, genetic features suggest it does. A galactan CPS is therefore included in the JCVI-syn3A biomass composition, with a dry weight fraction of 6.77 % before rescaling (see next subsection).

The assumption of capsule production in JCVI-syn3A rests on two glycosyltransferases, whose homologs in another mycoplasmas are candidates for capsule synthesis and attachment. Specifically, the two putative glycosyltransferases *epsG*/0113 and *cps*/0114 have homologs in *M. mycoides mycoides* PG1$^{\mathrm{T}}$, EpsG/MSC_0108 (84 % sequence identity to *epsG*/0113) and cps/MSC_0109 (92 % sequence identity to *cps*/0114). Based on *in silico* and preliminary experimental studies, a tentative mechanism has been suggested for galactan capsule synthesis in *M. mycoides mycoides* PG1$^{\mathrm{T}}$ [167]. Based on its predicted high structural similarity to both the cellulose synthase BcsA of *Rhodobacter sphaeroides* [168] as well as the galactosyltransferase GlfT of *Mycobacterium tuberculosis* [169], the glycosyltransferase EpsG of *M. mycoides mycoides* has been hypothesized to cytoplasmically polymerize UDP-galactofuranose and export it to the cell exterior [167], a mechanism generally assumed for bacteria with a single membrane and no periplasmic space [170]. Polysaccharide synthesis independent from the mono-Gal-DAG would also be consistent with experimental observations indicating that the latter does not serve as precursor for the galactan polymer in *M. mycoides mycoides* strain V5 [171]. cps would then be a candidate to attach the galactan chain to the cell membrane, a hypothesis supported by its differential expression in capsulated vs. non-capsulated colony variants of *M. mycoides mycoides* [167]. As the sequences of *epsG*/0113 and *cps*/0114 of JCVI-syn3A are very similar to their homologs, and in particular the conserved motifs DXD and QXXRW (common for processive enzymes [167, 172]) are also present in *epsG*/0113, we assume the same tentative mechanism to apply in JCVI-syn3A as well. The likely lipid acceptor for the galactan chain would then be DAG which is the substrate for the glycosyltransferases MG517 in *M. genitalium* [162] and MPN483 in *M. pneumoniae* [163] (and which we already assume to be the substrate for mono-Gal-DAG production by *cps*/0114 and 0697 as well).

Composition and mass fraction of capsule *M. mycoides mycoides* was found to contain ∼10 % dry weight of galactose in form of galactan (the only other carbohydrate species being ribose from nucleic acids) [173], suggesting this polysaccharide to completely account for the 8.1 % carbohydrate measured in *M. mycoides mycoides* cell residues defatted prior to measurement [32] (monogalactosyl-lipids removed during defatting possibly contributing to the remaining difference). We therefore assume that the 6.5 % carbohydrate in defatted *M. mycoides capri* [32] consists mainly of galactan.

The intact lipopolysaccharide moiety studied in Buttery and Plackett [173] contained around 4 % lipid. If we assume the latter to be palmityl oleyl glycerol (594.95 g/mol), then the average galactan chain length must be 88 galactosyl residues ($C_6H_{10}O_5$: 162.006 g/mol). We note that with ∼0.5 nm per monomer, this yields a polysaccharide chain of ca. 40–50 nm, which seems consistent with the capsular thickness visible in TEM imaging of *M. pulmonis* [170]. We therefore include Gal_{88}-DAG as the lipogalactan moiety in our biomass (before rescaling) with a dry weight fraction of 6.77 % (6.5 % carbohydrate + 0.27 % lipid).

As an aside on capsule composition, we note that the monosaccharide rhamnose was unexpectedly not detected using GC/MS (gas chromatography/mass spectrometry) in the minimal cell (personal communication with James Daubenspeck and Kevin Dybvig). It is present in wild-type *M. mycoides capri*. Mycoplasmas can convert oligomeric

but not monomeric glucose to rhamnose [174]. The enzyme that catalyzes rhamnose synthesis has not been identified. Rhamnose is thought to link proteins to phospholipids as a mechanism of trafficking proteins to the membrane. Proteins that are cytoplasmic when not associated with rhamnose, such as the glycolytic enzyme enolase, moonlight on the cell surface when modified by the addition of rhamnose and phospholipid [175]. Because mutants lacking rhamnose have not been described in global transposon mutagenesis studies of mycoplasmas, it was thought that this system was essential for mycoplasmas and possibly other bacteria.

Details on small molecule pool composition In addition to macromolecules, lipids and capsule, pools of free amino acids, nucleotides and deoxynucleotides are also included in the biomass, as well as cofactors and ions expected to be needed in JCVI-syn3A. A minimal medium for JCVI-syn3A has yet to be obtained, so the minimal media reported for *M. mycoides capri* LC Y [41] and *M. pneumoniae* [23] are used as a guideline for required ions and cofactors: Any compound present in a minimal medium is required by the cell, and the compound or its downstream product(s) need to be included in the biomass composition. From the two media mentioned, all inorganic ions are included in the JCVI-syn3A biomass composition except for sulfate. There is no known biological need for sulfate in JCVI-syn3A, since sulfur needed for certain tRNA modifications can be derived from cysteine via cysteine desulfurase (*iscS*/0441). Also, there is no obvious transporter candidate: A putative sulfate import system has been identified in *Mycoplasma hyopneumoniae* [52] (MHP168_157/158) but their homologs have been deleted in JCVI-syn3A (MMSYN1_0192 and MMSYN1_0193). The only other anion import system, the phosphate import system Pst (*pstS*/0425 through *pstB*/0427) is known to be highly selective for phosphate over sulfate [176, 177] and hence is no plausible candidate for sulfate uptake. Finally, we note that while the minimal media for *M. mycoides capri* LC Y [41] includes $MgSO_4$, a study on this organism's inorganic requirements [59] only reported a need for Mg^{2+}.

Transition metal ions (possibly present as trace contaminants in the media) are also not included in the biomass. While the need for divalent cations other than Mg^{2+} and Ca^{2+} is not clear, it was observed that addition of 1 μM of Mn^{2+}, Zn^{2+}, Co^{2+} or Fe^{2+} actually inhibited growth of *M. mycoides capri* LC Y [59].

From the nine vitamins needed by *M. pneumoniae*, we exclude choline based on experimental evidence that *M. mycoides capri* does not synthesize its own phosphatidylcholine [38]. For the remaining vitamins, there is an apparent need in JCVI-syn3A as cofactors or coenzymes (or in the case of spermine, to stabilize nucleic acids) and they (or their final forms) are thus included in the biomass. For lipoate, an uptake reaction is included in the model (see below), but lipoate is not actually included in the biomass in any form, as the holo-protein, lipoyl-PdhC (*pdhC*/0227), itself turns out to be nonessential in our model. We note that Rodwell [41] reports pyridoxal or folate derivatives are not required by *M. mycoides capri* LC Y, which is consistent with reports of *M. mycoides capri* LC Y being able to omit methionyl-tRNA formylation in the absence of folate derivatives without impact on growth [85]. However, the transposon mutagenesis data for JCVI-syn3A gives folate-related enzymes as quasi-essential, and the pyridoxal phosphate-dependent IscS (*iscS*/0441) as essential. Both pyridoxal phosphate and tetrahydrofolate are therefore included in the biomass.

Intracellular concentrations/mass fractions for vitamins and ions are taken from the iJO1366 *E. coli* model [46], except for potassium and sodium (which have been determined in *M. mycoides capri* PG3 [42]), chloride (which has been determined in *Mycoplasma gallisepticum* [45]), and phosphate (which is taken from the *M. pneumoniae* model [25]). Mass fractions for free nucleotides and deoxynucleotides have been determined in *M. mycoides capri* LC Y [43, 44], and relative mole fractions of free amino acids are approximated to match the average protein composition of JCVI-syn3A. The total amino acid mass fraction is assumed to resemble that of *M. pneumoniae* [25].

Metabolic reconstruction

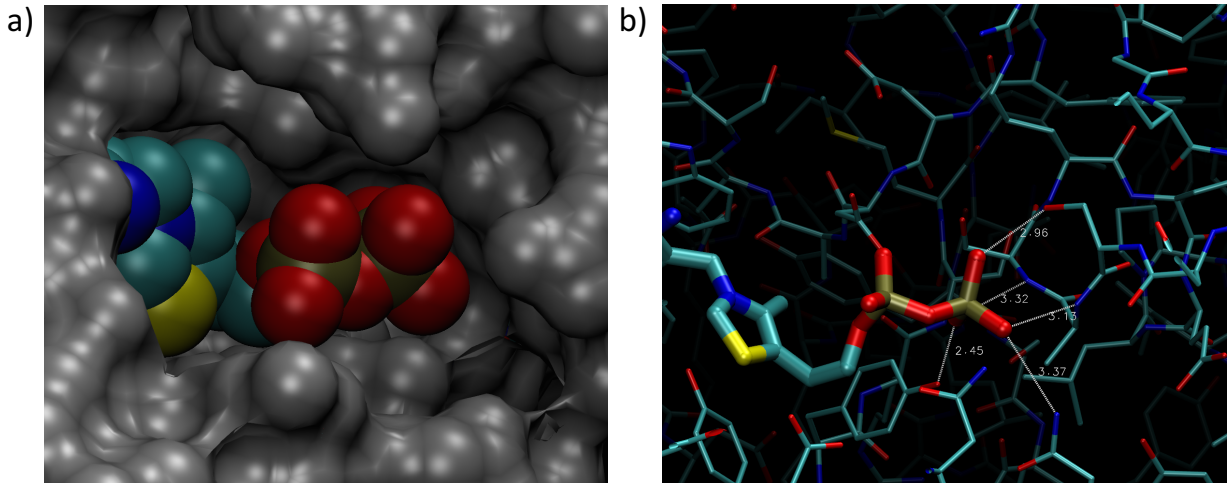
Possible oxidation of acetaldehyde Oxidation of acetaldehyde to acetyl-CoA could provide a rationale for the only partial deletion of pyruvate dehydrogenase (PDH) in JCVI-syn3A, as it would not require a decarboxylation step in the absence of the PDH_E1 subunit (MMSYN1_0225 and MMSYN1_0226). Cocks et al. [54] ruled out alcohol dehydrogenase or acetaldehyde dehydrogenase activity in *M. mycoides capri* LC Y, consistent with the absence of a corresponding gene in JCVI-syn3A; however, the assay mixture compositions did not mention coenzyme A, so that acetaldehyde oxidation to acetyl-CoA instead of acetate might not have been detectable. Therefore, the model tentatively includes an acetaldehyde oxidation reaction catalyzed by PdhC (*pdhC*/0227). Furthermore, assuming the membrane permeability of acetaldehyde to be comparable to that of acetamide, which is five times higher than for glycerol [178], suggests that acetaldehyde could also passively leave the cell. Thus a direct secretion reaction is included for acetaldehyde.

Presence and absence of specific nucleoside kinase and phosphorylase reactions The substrate profile of the two nucleoside kinases *dak1*/0330 and *dak2*/0382 is inferred from *in vitro* [70] and *in vivo* [43, 66] studies. Both proteins show significant sequence identity to the *M. mycoides mycoides* SC kinase MSC_0388. *dak1*/0330 (previously annotated as deoxyguanosine kinase) is identical to MSC_0388 up to a single C-terminal residue, and the putative deoxynucleoside kinase *dak2*/0382 [12] shares 37 % identity with MSC_0388. MSC_0388 was shown to act on deoxyadenosine, deoxyguanosine and deoxycytidine, with weaker activity against adenosine and guanosine and very weak activity against cytidine [70]. Together with the thymidine kinase *tdk*/0140, this should cover the phosphorylation of all deoxyribonucleosides in JCVI-syn3A, leaving the question of the role of the putative third deoxynucleoside kinase *dak2*/0382. The sequence of this protein shows some changes in the active site residues compared to the crystal structure for MSC_0388 [70]; however, Tyr43 close to the ribosyl-C2 in the *M. mycoides mycoides* SC crystal structure (possibly responsible for the preference for deoxyribonucleosides by blocking the space of a C2-hydroxyl group) is preserved in *dak2*/0382 (Tyr45), suggesting the same preference for deoxyribonucleosides. Lacking further information, a similar substrate profile for this third kinase is assumed as for the deoxyadenosine kinase, and deoxyadenosine/-guanosine/-cytidine phosphorylation is considered to be carried out by either enzyme. One possible explanation for the presence of both kinases could be complementary activities *in vivo*: Activity of MSC_0388 for deoxyguanosine/-cytidine was found to be strongly inhibited by the presence of deoxyadenosine [69], so that in *in vivo*, not all reactions might be carried out by both kinases. The fact that almost no activity was found against cytidine suggests that the cytidine kinase activity observed in cell extracts of *M. mycoides capri* LC Y [43] arises from uridine kinase instead (uridine kinase activity was partially inhibited by cytidine [43]), which has been deleted in JCVI-syn3A (MMSYN1_0491). Weak activity against adenosine and guanosine has been observed [70], but this activity was not found in cell-free extracts [66] and the presence of the corresponding phosphoribosyltransferases in the minimized genome suggests that it could not provide AMP and GMP in sufficient quantities. Thus adenosine (or guanosine) kinase reactions are not included, and consequently there is no direct conversion of adenosine to AMP or guanosine to GMP. Instead, it is assumed that AMP, GMP and UMP are formed from their respective bases by the corresponding phosphoribosyltransferases alone (*hptA*/0216, *apt*/0413, and *upp*/0798).

While purine nucleoside phosphorylase activity is present in JCVI-syn3A (see main text), it appears as if no such activity exists for pyrimidines in JCVI-syn3A. Cytidine phosphorylase activity has been ruled out in *M. mycoides capri* LC Y cell extracts [43]. Phosphorolysis of uridine [34, 43, 54], deoxyuridine and thymidine [44] has been observed; furthermore, the latter two activities could be attributed to the same enzyme [44]. This is in agreement with the experimentally observed activity of MHR_0565 [179], the homolog in *Mycoplasma hyorhinis* HUB-1 of

the putative pyrimidine nucleoside phosphorylase MMSYN1_0734 (45 % sequence identity): MHR_0565 was found to phosphorolyse thymidine, deoxyuridine, and uridine, but not cytidine or deoxycytidine. As MMSYN1_0734 has been deleted from JCVI-syn3A, we assume that JCVI-syn3A no longer exhibits pyrimidine nucleoside phosphorolysis activity. We note that the *M. mycoides capri* thymine uptake requirement can also be met by thymidine in the growth media [44], which means that there is no need for a thymidine phosphorylase activity by some unidentified paralog of MMSYN1_0734.

Uptake forms of thiamine and lipoate Deletion of thiamine diphosphokinase (MMSYN1_0261) suggests that thiamine diphosphate (ThDP) must be taken up directly. The substrate-binding protein 0708 shows similarity to both the ThDP-binding Cypl from *Mycoplasma hyorhinis* [180] (24 % identity) and the thiamine-binding MG289 from *M. genitalium* [181] (23 %). However, a sequence alignment reveals that the diphosphate-stabilizing interactions in Cypl are largely missing in 0708, as is the case in MG289. Sippel et al. [181] note that the mere absence of these interactions would not yet exclude ThDP binding, an idea supported by a structural alignment of ThDP-bound Cypl and MG289 suggesting that ThDP could bind to the MG289 binding pocket as well (see Appendix 1-Figure 1). In light of this, the conservative assumption is made that thiamine is directly taken up as ThDP.



APPENDIX 1-FIGURE 1 Thiamin diphosphate (ThDP) from the *Mycoplasma hyorhinis* Cypl crystal structure (pdb: 3EKI) overlaid onto the crystal structure of MG289 (pdb: 3MYU). (Structures aligned using STAMP [182] in VMD [183, 184].) a): Space-filling view, with MG289 in gray and ThDP in color. The pyrophosphate tail of ThDP from the Cypl structure would have an appropriate cavity in MG289 as well. b): Visualization of hydrogen bonds for the same alignment. All possible hydrogen bonds are shown between potential donor and acceptor heavy atoms within 3.5 Å or less of each other. Even in the absence of the residues involved in pyrophosphate binding in Cypl [180], the alignment suggests other side group and backbone interactions could still allow for pyrophosphate binding.

Lipoate is a required cofactor for PdhC (*pdhC*/0227), but the deletion of two putative lipoyl transferases (MMSYN1_0224 and MMSYN1_0464) suggests an alternative lipoylation mechanism for PdhC. Such an alternative mechanism would be transamidation using a lysine- or peptide-bound lipoate, for which putative peptidases with a covalent mechanism would be candidates. Among the remaining genes in JCVI-syn3A, 0401 has been annotated by RAST as homologous to the sublancin 168 lantibiotic transporter [185], which exports bactericidal peptides and simultaneously cleaves off a leader peptide containing a double-glycine motif in the process at its peptidase C39 domain. However, no candidate for such a double-glycine motif peptide has been identified in JCVI-syn3A; this is consistent with the occurence of peptidase C39 domain proteins in other mycoplasmas without apparent candidate double-glycine motif peptides [186]. This suggests another function for 0401; it could potentially import lipoyllysine or a lipoylpeptide

and catalyze the transamidation of the lipoyl moiety onto the lipoyl-binding domain of PdhC. To account for the required lipoylation of PdhC in the absence of lipoyl transferases, this mechanism of uptake and transamidation of lipoyllysine is tentatively included in the model for JCVI-syn3A.

Feasibility of permeative glycerol uptake In the absence of a dedicated glycerol importer, passive permeative glycerol uptake is assumed. Using a permeation coefficient of 50 nm/s for glycerol through a phospholipid bilayer [178] and assuming an external glycerol concentration of 6 mg/l in SP4 medium [61], an upper limit on glycerol uptake 0.193 mmol $\mathrm{gDW}^{-1}\,\mathrm{h}^{-1}$ is obtained for a 0.4 micron diameter spherical cell, assuming a density of 1.1 g/ml [160] and a cellular water content of 72 % [42], as in the biomass calculations. We note that the optimal FBA solution of our model demands 0.064 mmol $\mathrm{gDW}^{-1}\,\mathrm{h}^{-1}$ glycerol uptake, so that this passive uptake is not growth-limiting, and the ability of the model cell to grow is not dependent on the exact permeation coefficient *in vivo*, which might not be the same as for the pure phospholipid bilayer.

ATP costs of protein turnover The AAA+ protease Lon (*lon*/0394) is assumed to be the main protease for turnover in JCVI-syn3A. Lon from *E. coli* has been found to decompose proteins of different length and composition into oligopeptides of 10–20 residues, consuming an amount of ATP per hydrolyzed peptide bond in the process that depends on pH and ADP concentration but not on the substrate [187]. Lon from *Salmonella typhimurium* reportedly has a similar ATP/peptide bond stoichiometry to *E. coli* Lon [187], suggesting that the stoichiometry for *E. coli* can be used to approximate the stoichiometry in other species as well. At pH 7.5 (approximate cytosolic pH of *M. mycoides capri* in neutral medium [113]) and 0.5 mM of the inhibitor ADP (comparable to the concentration in *M. mycoides capri* LC Y [43]), Lon consumes 9 ATP per peptide bond [187]. Assuming breakdown to oligopeptides of $\sim$15 residues yields an ATP expense of 225 ATP per protein of 385 residues in the model. The resulting oligopeptides would then be further broken down to individual amino acids by the other peptidases without expense of energy.

Derivation of Na^+ and K^+ active transport reactions Potassium and sodium transport in *M. mycoides capri* is known to involve several functionalities. Specifically, from studies on *M. mycoides capri* PG3, three functionalities have been proposed [115, 188]:

- A K^+ uniport functionality that can only import K^+ until its chemical potential is in equilibrium.
- A K^+/Na^+ antiport functionality that consumes ATP and is able to concentrate K^+ inside the cell (beyond equilibrium).
- An Na^+/H^+ antiport functionality that extrudes Na^+, powered by the proton gradient established by the proton-extruding ATPase.

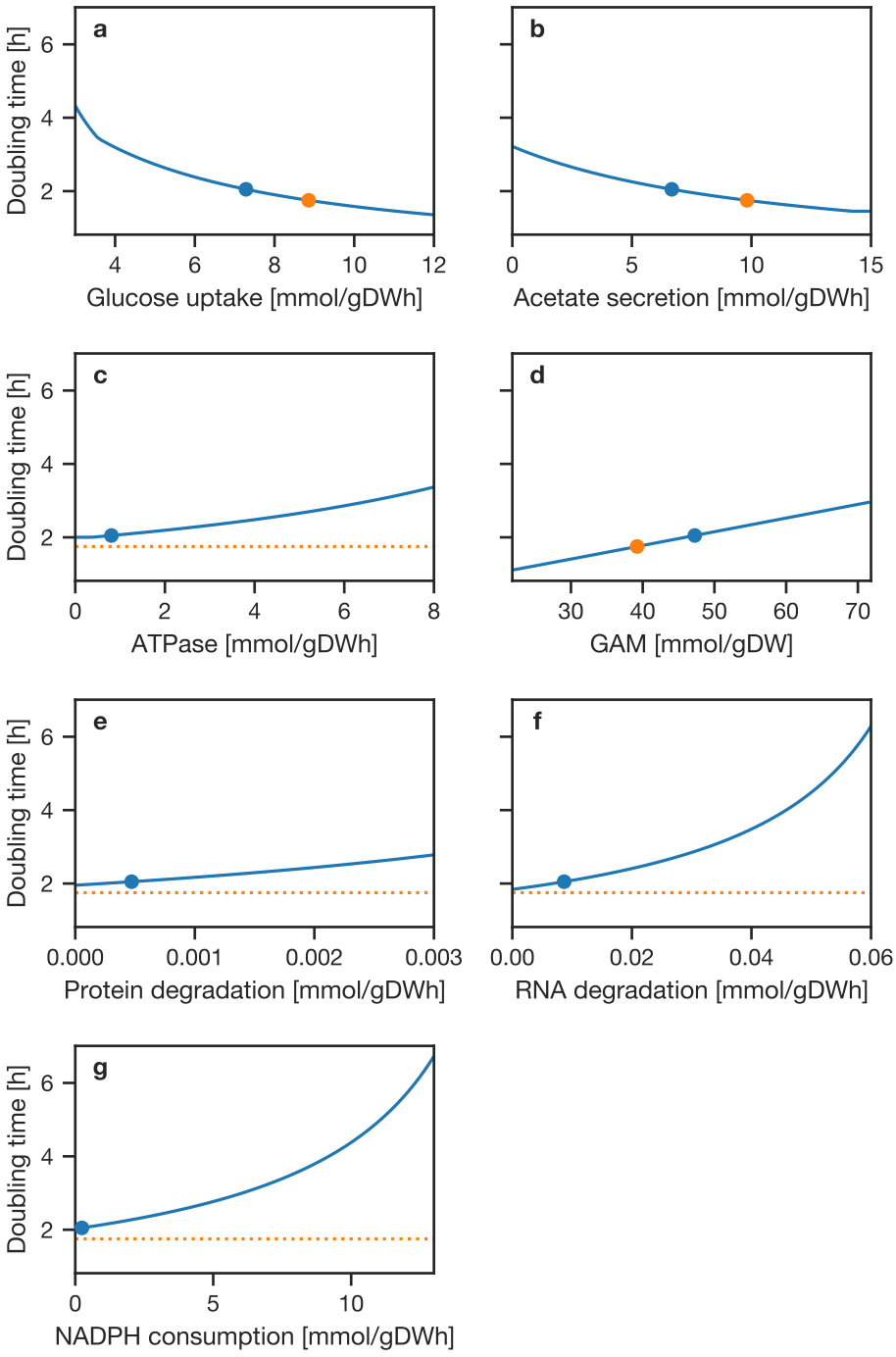
The two K^+ transport functionalities compare well to the properties of the KtrAB K^+ import system (*natA*/0685 and *trkA*/0686) as characterized in *Vibrio alginolyticus* [189]. KtrB alone was found to slowly import K^+ (in an Na^+-independent manner), and also import Na^+. The complete KtrAB system, in contrast, was found to import K^+ two orders of magnitude faster and in an Na^+-dependent manner, but no longer imported Na^+ (extrusion not studied). These findings would be consistent with the KtrAB system providing both the passive K^+ uniport functionality as well as the active K^+/Na^+ antiport functionality. We thus include an ATP-consuming K^+/Na^+ antiport reaction catalyzed by KtrAB (*natA*/0685 and *trkA*/0686) in the model (but not the uniport reaction, since it does not participate in yielding the concentrated K^+ level in the cytoplasm). We also include an Na^+/H^+ antiport reaction (without any gene assignment for the time being).

Sensitivity analysis

As the constraint parameters used in our model stem from other mycoplasmas, the sensitivity of the model to variations in the parameters is of particular interest. Appendix 1-Figure 2 shows the doubling time obtained for varying different constraints around the respective value chosen in the model (indicated by a blue filled circle in each plot). In addition, we calculated elasticities of the doubling time with respect to each constraint per the formula:

$$E_c\, t_\mathrm{d}|_{c=c_0} = \frac{c_0}{t_\mathrm{d}|_{c=c_0}} \left.\frac{\partial t_\mathrm{d}}{\partial c}\right|_{c=c_0} \tag{4}$$

for a given constraint c at its reference value c_0 (i.e. the value used in the model) while keeping all other constraints fixed at their reference values. These elasticities describe the ratio in relative changes of growth rate and constraint for varying the given constraint around its reference value. (I.e., a (locally constant) elasticity of 0.1 would imply a growth rate change by 1 % for a change in the constraint by 10 %.)



APPENDIX 1-FIGURE 2 Sensitivity analysis of model doubling time with respect to model constraints. In each panel, the stated parameter was varied over the indicated range and the model doubling time calculated while keeping all other constraints constant. A: Maximal glucose uptake. B: Maximal acetate secretion. C: ATPase ATP cost. D: GAM ATP cost. E: Protein degradation rate. F: RNA degradation rate. G: Imposed NADPH consumption. The blue circle marks the value used in the FBA model and resulting doubling time; the orange circle indicates the parameter that would yield the experimental doubling time. If there is no value of the parameter which would yield the experimental doubling time, a horizontal line is plotted.

Low elasticities (≤ 0.1) are obtained for the ATPase efflux (0.03), protein degradation (0.04) and RNA degra-

dation (0.1), i.e. the three parameters that together determine the model NGAM–demonstrating the model's relative insensitivity to variations in these parameters. A moderate elasticity of 0.37 is observed for acetate secretion, while higher elasticities of 0.80 and 0.86 are found for glucose uptake and the total GAM, respectively (ignoring signs). While this elasticity suggests a considerable sensitivity of model growth rate on the estimated GAM, we note that nearly half of the GAM (21.54 mmol gDW^{-1}) can be accounted for by macromolecular synthesis cost, specifically protein synthesis cost ($\sim$21.2 mmol gDW^{-1} in the model). The protein synthesis cost depends on the ATP cost per amino acid in the proteome ($\sim$4 ATP/amino acid) which is deemed conserved, and on the protein fraction in the cellular dry mass. The protein dry mass fraction is based on that of the natural *M. mycoides capri*, which is assumed to provide a reasonable first approximation for JCVI-syn3A (see also Further model details and justification in Appendix 1). The quantifiable fraction of the GAM thus carries considerably less uncertainty than the non-quantifiable fraction.

It is thus instructive to consider the two contributions to the GAM in our model separately, by decomposing the total GAM (G) into its quantifiable ($\mathrm{G_q}$) and nonquantifiable ($\mathrm{G_{nq}}$) parts, $\mathrm{G} = \mathrm{G_q} + \mathrm{G_{nq}}$; and evaluating the elasticity with respect to $\mathrm{G_q}$ and $\mathrm{G_{nq}}$ separately. We have:

$$\frac{\partial t_\mathrm{d}}{\partial \mathrm{G_q}} = \frac{\partial t_\mathrm{d}}{\partial \mathrm{G}} \times \frac{\partial \mathrm{G}}{\partial \mathrm{G_q}} = \frac{\partial t_\mathrm{d}}{\partial \mathrm{G}} \tag{5}$$

which analogously holds for $\mathrm{G_{nq}}$. Furthermore, since $\mathrm{G_0} = \mathrm{G_{q,0}} + \mathrm{G_{nq,0}}$:

$$t_\mathrm{d}|_{\mathrm{G}=\mathrm{G_0}} = t_\mathrm{d}\big|_{\substack{\mathrm{G_q}=\mathrm{G_{q,0}} \\ \mathrm{G_{nq}}=\mathrm{G_{nq,0}}}} \tag{6}$$

where as in Eq. 11, the subscript 0 denotes the value adopted in the model for each quantity. Summing the elasticities with respect to $\mathrm{G_q}$ and $\mathrm{G_{nq}}$ then yields:

$$\begin{aligned}
&E_{\mathrm{G_q}}\, t_\mathrm{d}\big|_{\substack{\mathrm{G_q}=\mathrm{G_{q,0}} \\ \mathrm{G_{nq}}=\mathrm{G_{nq,0}}}} + E_{\mathrm{G_{nq}}}\, t_\mathrm{d}\big|_{\substack{\mathrm{G_q}=\mathrm{G_{q,0}} \\ \mathrm{G_{nq}}=\mathrm{G_{nq,0}}}} \\
&= \frac{\mathrm{G_{q,0}}}{t_\mathrm{d}\big|_{\substack{\mathrm{G_q}=\mathrm{G_{q,0}} \\ \mathrm{G_{nq}}=\mathrm{G_{nq,0}}}}} \left.\frac{\partial t_\mathrm{d}}{\partial \mathrm{G_q}}\right|_{\substack{\mathrm{G_q}=\mathrm{G_{q,0}} \\ \mathrm{G_{nq}}=\mathrm{G_{nq,0}}}} + \frac{\mathrm{G_{nq,0}}}{t_\mathrm{d}\big|_{\substack{\mathrm{G_q}=\mathrm{G_{q,0}} \\ \mathrm{G_{nq}}=\mathrm{G_{nq,0}}}}} \left.\frac{\partial t_\mathrm{d}}{\partial \mathrm{G_{nq}}}\right|_{\substack{\mathrm{G_q}=\mathrm{G_{q,0}} \\ \mathrm{G_{nq}}=\mathrm{G_{nq,0}}}} \\
&= \frac{\mathrm{G_{q,0}}}{t_\mathrm{d}\big|_{\substack{\mathrm{G_q}=\mathrm{G_{q,0}} \\ \mathrm{G_{nq}}=\mathrm{G_{nq,0}}}}} \left.\frac{\partial t_\mathrm{d}}{\partial \mathrm{G}}\right|_{\substack{\mathrm{G_q}=\mathrm{G_{q,0}} \\ \mathrm{G_{nq}}=\mathrm{G_{nq,0}}}} + \frac{\mathrm{G_{nq,0}}}{t_\mathrm{d}\big|_{\substack{\mathrm{G_q}=\mathrm{G_{q,0}} \\ \mathrm{G_{nq}}=\mathrm{G_{nq,0}}}}} \left.\frac{\partial t_\mathrm{d}}{\partial \mathrm{G}}\right|_{\substack{\mathrm{G_q}=\mathrm{G_{q,0}} \\ \mathrm{G_{nq}}=\mathrm{G_{nq,0}}}} \\
&= \frac{\mathrm{G_{q,0}} + \mathrm{G_{nq,0}}}{t_\mathrm{d}\big|_{\substack{\mathrm{G_q}=\mathrm{G_{q,0}} \\ \mathrm{G_{nq}}=\mathrm{G_{nq,0}}}}} \left.\frac{\partial t_\mathrm{d}}{\partial \mathrm{G}}\right|_{\substack{\mathrm{G_q}=\mathrm{G_{q,0}} \\ \mathrm{G_{nq}}=\mathrm{G_{nq,0}}}} \\
&= \frac{\mathrm{G_0}}{t_\mathrm{d}|_{\mathrm{G}=\mathrm{G_0}}} \left.\frac{\partial t_\mathrm{d}}{\partial \mathrm{G}}\right|_{\mathrm{G}=\mathrm{G_0}} \\
&= E_\mathrm{G}\, t_\mathrm{d}|_{\mathrm{G}=\mathrm{G_0}}
\end{aligned}$$

I.e., the elasticity with regard to the total GAM is equal to the sum of elasticies with regard to its quantifiable and non-quantifiable fraction. This yields elasticities of the model doubling time with respect to $\mathrm{G_q}$ and $\mathrm{G_{nq}}$ of 0.40

and 0.46, respectively. The former, then, describes the elasticity with respect to a parameter where comparatively less uncertainty is in fact expected; from the overall high elasticity of 0.86 with respect to the GAM, only 0.46 fall to a quantity of significant uncertainty.

Regarding the high elasticity of 0.8 with regard to glucose uptake rate, we note that since the carbon source uptake directly determines the overall ATP production in the model, a significant sensitivity of model growth rate with regard to that uptake rate is in fact to be expected. For example, we observe a growth rate elasticity of 1.01 with respect to glucose uptake rate (10 mmol $\mathrm{gDW}^{-1}\,\mathrm{h}^{-1}$) for the *E. coli* model iJO1366 [46]. For our model, where glucose uptake rate and other parameters are adopted from other organisms and models, this means that the growth rate predicted should hence be more considered a qualitative prediction, and provisional until corresponding measurements become available for JCVI-syn3A.

Finally, we note that the only consumer of reduction equivalents in the model is currently the ribodinucleotide reductase system (RNDR). There are however likely other significant demands for reduction equivalents that currently can not be quantified, like repair of oxidative damage of proteins (cysteine oxidation to disulfides). To probe the possible impact of this demand, we introduced an artificial NADPH oxidation reaction for testing purposes and calculated the doubling time as a function of imposed NADPH consumption (panel G), as NADPH production through GAPDP diverts flux from the ATP-producing GAPD/PGK branch. As the NADPH consumption through this artificial oxidation reaction is absent/zero in our model outside of this test, the associated elasticity at this point is necessarily also zero and thus not informative. Similarly, the elasticity has very small positive values for small nonzero values of NADPH consumption (e.g. 0.0005 for an imposed NADPH consumption of 0.01 mmol $\mathrm{gDW}^{-1}\,\mathrm{h}^{-1}$). It is thus of interest to consider the elasticity at higher NADPH consumption values. At 1.0 mmol $\mathrm{gDW}^{-1}\,\mathrm{h}^{-1}$, the observed elasticity is still low with 0.06. The *in silico* growth rate at this value is also still 2.14 h, demonstrating that while we cannot accurately capture NADPH demand in our model, the model can sustain a certain level of NADPH demand without significant impact on growth rate.

Tolerable protein dilution as possible cause of time-delayed gene disruption lethality

As discussed in Section 2.1 in the main text, a gene can in principle appear quasi-essential in the transposon mutagenesis analysis if its disruption is lethal in principle (rather than just causing a growth defect) but will take time to take effect. One reason for such time-delayed lethality of a gene knockout could be an initial gene product abundance that is high enough to sustain cellular demands over several generations (during which the protein concentration is diluted by half at each cell division). Specific candidates for such a scenario would be genes identified as quasi-essential where not only the FBA model, but also biological context would otherwise strongly suggest the gene to be essential. As mentioned in Section 3.2 in the main text, this is the case for the genes *nrdE*/0771 and *nrdF*/0773 (subunits of RNDR) and *pgpA*/0214 (PGPP).

The capacity of the cell to tolerate disruption of a gene and maintain the required metabolic fluxes with an ever-decreasing protein abundance (twofold dilution at cell division, no new synthesis of functional protein due to gene disruption) could be estimated by the ratio of the protein abundance based flux constraint $V_{\max}$ to the reaction flux required by the model. The estimated $V_{\max}$ for PGPP is $\sim$2.4 mmol $\mathrm{gDW}^{-1}\,\mathrm{h}^{-1}$, assuming an initial cellular abundance of $\sim$170 copies. (I.e., assuming an abundance equal to the average copy number across the proteome; PgpA was not quantified in the proteomics experiment.) The PGPP flux required by the optimal FBA solution is 0.028 mmol $\mathrm{gDW}^{-1}\,\mathrm{h}^{-1}$, i.e. $\sim$100 times lower. Assuming twofold dilution of the initial PgpA copy number at each cell division would then allow the cell to divide 6-7 times before the flux limit would not suffice anymore to sustain the flux required for optimal growth. I.e., for these first 6-7 generations, the cell would not experience any loss of fitness yet.

A caveat lies in the turnover number k_{cat} found for PGPP, which stems from *E. coli* and a slightly different substrate (phosphatidate instead of phosphatidylglycerophosphate).

A similar argument could be made for RNDR. The total flux through all four RNDR reactions required by the model is 0.005 mmol gDW^{-1} h^{-1}. Assuming a similar k_{cat} for all four dinucleotide substrates would yield a total flux limit V_{max} of 0.378 mmol gDW^{-1} h^{-1}, $\sim$70 times higher than the required flux; assuming simple dilution, this would then allow for six cell divisions before the flux limit does not support the optimal flux anymore, similar to the situation for *pgpA*. It should be noted however that RNDR is essential in the model not as a source of deoxynucleotides (which can be obtained from deoxynucleosides as well), but rather as the only consumer of NADPH (produced by *folD*/0684/MTHFD). Accordingly, the *in silico* flux through the RNDR reactions equals the NADPH production via MTHFD, and amounts to 12 % of the total dA/G/CDP production flux of 0.038 mmol gDW^{-1} h^{-1}, with the bulk of the flux in the model instead carried by the deoxynucleoside/-mononucleotide kinases. (dTDP cannot be made through RNDR and thus is not relevant for the RNDR flux limit.) *In vivo*, other NADPH sinks should remove the need to use RNDR for NADP regeneration, so that the *in vivo* flux should only be determined by the deoxynucleotide synthesis partioning between RNDR and the deoxynucleoside kinases for deoxyadenosine, -guanosine and -cytidine. The deoxynucleoside kinases *dak1*/0330 and *dak2*/0382 are running not too far from their V_{max} limit in the model: The reaction DADNK (accounting for 70 % of the flux through *dak1*/0330 and *dak2*/0382) has a V_{max}/FBA flux ratio of 2.4; as elaborated in the main text, running enzymes close to V_{max} creates unstable conditions in case of changing substrate concentrations. RNDR might then be required *in vivo* to take flux load off the deoxynucleoside kinases, which would explain why it was still essential even if not needed as an NADPH sink. At the same time, since the deoxynucleoside kinases are capable of carrying most of the flux, the *in vivo* RNDR flux can be expected to not significantly exceed the one observed in the model. Thus, the scenario of flux capacity buffering against enzyme dilution as a cause for the apparent quasi-essentiality of the RNDR genes *nrdE* and *nrdF* seems conceivable even if more definitive statements cannot be made.

Interpretation of YgfA essentiality amidst a quasi-essential folate metabolism

5-formyl-THF cyclo-ligase (*ygfA*/0443) is the only essential folate-related enzyme whereas all other genes in folate cycle and uptake are merely quasi-essential *in vivo*. This seems plausible at first given that 5-formyl-THF is a known inhibitor of folate-related enzymes [110], rendering prevention of its buildup by YgfA an important metabolic damage repair function. However, for JCVI-syn3A a paradox arises: The potential inhibition targets for 5-formyl-THF, namely the other folate enzymes, are themselves only quasi-essential—raising the question why preventing buildup of 5-formyl-THF should then be essential. Unlike a knockout of any other individual folate-related enzyme, 5-formyl-THF buildup as a result of YgfA knockout would inhibit several folate-related enzymes at once. However, if this were the reason for the essentiality of YgfA, the uptake protein FolT should be essential too. In addition, an individual knockout of GlyA (GMHT, *glyA*/0799) or FolD (MTHFD/MTHFC, *folD*/0684) would also in effect disrupt the complete folate cycle. This suggests that 5-formyl-THF buildup might have detrimental effects outside its known range of interactions with folate-related enzymes—or that there are essential folate-related genes in JCVI-syn3A yet to be identified. It should in this context be noted however that *ygfA* is classified as essential in the transposon data with a probability of only 0.58, compared to a probability of being quasi-essential of 0.42—rendering the classification much less certain than for nearly all other genes classified as essential. As an aside, the proposed folate uptake gene *folT*/0822 is the only *ecfS* gene that is quasi-essential rather than fully essential—lending further support to its assignment as *folT*.

Proteomics derived constraints

Genome-scale metabolic reconstructions can be used to predict flux distributions through the metabolic network. Metabolic reconstructions represent the network topology of all metabolic reactions that can occur in a given organism. The network topology imposes a linear dependence between fluxes through the various reactions based on the stoichiometry and connectivity of the metabolites involved, assuming steady state conditions. Typically the number of reactions (variables) is higher than the number of metabolites (and hence mass balance equations) which results in an under-determined system of linear equations for reaction fluxes. Additional constraints based on the environment (growth medium), experimentally determined housekeeping requirements (NGAM *etc.*) and thermodynamics (reversibilities) further reduce the possible flux values through individual reactions (see Section 2.3 in the the main text). This solution space of feasible fluxes can be further reduced based on proteomics constraints [116, 117]. Proteomics constraints are applied using the Gene-Protein-Reaction (GPR) rules which list all genes whose products catalyze a particular reaction. The copy number of proteins available in the cell combined with their catalytic capacity gives an upper limit on the flux possible through a particular reaction by

$$V_{\max} = \frac{1000\, N_{\text{prot}}\, k_{\text{cat}}}{N_{\text{A}}\, m_{\text{cell}}}. \tag{7}$$

Here, $V_{\max}$ is the upper bound on the flux through a particular reaction with units of mmol $\text{gDW}^{-1}\,\text{h}^{-1}$, k_{cat} is the turnover number of a given enzyme (in 1/hr), N_{A} is Avogadro's constant, m_{cell} is the dry weight of a single JCVI-syn3A cell, and N_{prot} is the copy number of a given protein per cell. For reactions catalyzed by a protein complex, the lowest copy number among the measured subunits is used for N_{prot}, whereas in the case of reactions which can be catalyzed by multiple enzymes independently, the sum of the copy numbers of all isozymes is used for N_{prot}. In such cases it is required that all isozymes are measured to estimate a $V_{\max}$.

Protein copy numbers are obtained from quantitative proteomics (see "Mass Spectrometry Based Proteomics" in Methods). The median of relative abundances obtained from the three replicates of time point 1 (exponential growth phase) is used. Relative abundances are converted to absolute abundances (cellular copy numbers) per the total protein biomass fraction (see Section 2.2 in main text) and protein average molecular weight as reconstructed from the proteomics data. This yields an estimated total number of ~77,000 protein molecules per cell. k_{cat} values are obtained from BRENDA [118] with careful consideration for reaction substrate, physiological conditions of measurement and source species for the enzyme of interest, as presented in the following section.

For each EC number, the BRENDA database was queried for all available turnover data across all organisms. Since kinetic data is available for many different substrates, the results must be filtered to only include the natural substrates. The chemical names in the BRENDA database are not regular and can be specified under many synonyms. Thus, the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [132] was used to specify the preferred names of the substrates and products. To derive a mapping between all possible synonyms and the preferred chemical name, the set of all natural product and substrate names from BRENDA were compiled for each EC number. A Python script was used to efficiently construct the synonym to preferred name mapping by displaying the KEGG chemical names prefixed with a number and allowing the user to indicate which of the compounds listed in the BRENDA entries are equivalent. After regularizing the chemical names using the synonym map, and filtering out entries which do not correspond to the natural substrates/products, a unique entry must be chosen for each EC number.

To choose the most appropriate database entry, we first define an ordering function

$$\omega(E) = [e, d, m, t, p, c], \tag{8}$$

over the database entries E, where the quantities in the list are the selection criteria in decreasing order of importance. First, e is 1 if the entry's commentary indicates that the temperature of the measurement was performed above 50 °C, allowing for extremophiles to be filtered out. Second, the phylogenetic distance, d, between *Mycoplasma mycoides* and the entry organism is considered. The distance utilizes the dataset available at the Interactive Tree of Life (iTOL) [190] which was first described by Ciccarelli et al. [191]. For organisms missing from the Ciccarelli et al. data set, if there is a species available with the same genus, that distance is used. Otherwise, the species is classified into Gram-positive or gram-negative bacterium, eukaryote, or archaeon and assigned the largest distance to a member of those groups. Third, the commentary field is searched for the strings "wild-type" and "mutant". Entries containing "wild-type" but not "mutant" ($m = 0$) are preferred over entries with neither string present ($m = 1$). Then entries containing "mutant" are considered, with entries containing both "wild-type" and "mutant" ($m = 2$) preferred over those with "mutant" alone ($m = 3$). Fourth, the commentary field is searched for temperature and the entry closest to 30 °C preferred ($t = |T - 30^\circ\,\mathrm{C}|$). If no temperature is listed, it is treated as a 5 °C difference. Fifth, the commentary field is searched for pH ($p = |\mathrm{pH} - 7.0|$), preferring entries closest to pH 7.0. If no pH is listed, the entry is treated as having a difference of 0.1 units from pH 7.0. Finally, the entry with the shortest commentary string is preferred ($c = length(\texttt{commentary})$). Since longer comments are more likely to be for measurements taken at non-standard conditions, shorter comments are more likely to be appropriate for our purposes. The group of entries with the minimum value of the ordering function, $\omega(E)$, is then selected (lexicographically) for further processing.

At this point, if there is no unique best candidate in the selected group, the median turnover number among the remaining candidates is taken as k_{cat}. When there are an even number of values, the geometric mean is taken of the two middle values since it is possible that the two values could span multiple orders of magnitude. This prevents the larger value from dominating the median. The V_{max} values used as well as the k_{cat} entries used to compute them are available in Supplementary File 7, along with the reason that particular value was chosen. The column V_{min} in Supplementary File 7 refers to V_{max} for the reverse reaction assuming the forward direction as shown in the reaction equation.

There were instances where substrates identified using KEGG did not match the substrate in the model reaction. In these cases, k_{cat} values were assigned manually following the above selection criteria. In addition, k_{cat} measurements found during the literature survey for the metabolic reconstruction which were not available in BRENDA were used over the automated selections if they were better matches with respect to Equation 8. These manual changes are also listed in Supplementary File 7.

Comparison of JCVI-syn1.0 and JCVI-syn3A

It is instructive to consider the differences between JCVI-syn1.0 and JCVI-syn3A. As outlined above, in regards to the biomass composition, the overall protein dry mass fraction is not expected to differ much, and the RNA content of *M. mycoides capri* is expected to provide a reasonable approximation for JCVI-syn3A as well. An important change in the biomass composition is of course the smaller size of the JCVI-syn3A genome (by a factor of two); the different genome is also expected to affect the proteome composition, which in the JCVI-syn3A reconstruction is given directly by the JCVI-syn3A proteomics data.

When considering the functional content of the genome in JCVI-syn3A versus the one in JCVI-syn1.0, it is informative to consider their respective origins. While the genes encoded by the near minimal bacterial cell JCVI-syn3A are a subset of the gene content of both the naturally occurring bacterium *M. mycoides capri* and JCVI-syn1.0 (the wild type organism encodes a few genes deleted from JCVI-syn1.0 to reduce possible pathogenicity and to facilitate genome synthesis), the minimized and full-size genomes were evolved for life under very different conditions. The naturally occurring bacterium evolved for life in both the upper and lower respiratory systems of goats. Initial infection

would take place in the upper respiratory system, which has different epithelial cells and a lower temperature than would be encountered in the lower respiratory system, i.e. the lungs, which are about 37 °C. As an example of those evolutionary forces, DNA methylation analysis of the closely related goat pathogen *Mycoplasma capricolum* showed 315 GATGA sites with adenine methylated at the first A in cells grown at 30 °C and a different 15 sites when the cells were grown at 37 °C (data not shown). We assume the methylation machinery that causes this is an adaptation to enable the bacterium to grow in the two different goat milieus. Wild type *M. mycoides capri* likely has similar methylation machinery; however when we designed the JCVI-syn3A genome, the only environment we wanted the organism to grow in was SP4 media. Those DNA methylases are no longer present. Because of their high mutation rate and apparent evolutionary pressure to minimize their gene content, mycoplasmas tend to lose genes that are not essential for life in their preferred host [65, 192]. The entire gene content of wild type *M. mycoides capri* is optimal for life in a goat. Removal of any gene would probably result in that mutant failing to survive long in an infected animal or herd of animals. In JCVI-syn3A, the evolutionary pressure, which of course is channeled through the choices it JCVI designers made in choosing its gene content, is quite different. It was not based on ideal growth, but rather on coming up with a genome that had as few genes as possible, but could still divide in two hours or less in SP4 growth media. Gene content is based on an artificial criterion of minimization rather than evolutionary advantage.

To elucidate the impact of genome reduction on metabolic capabilities, Supplementary File 8 lists known reactions expected to be present in JCVI-syn1.0 but absent in JCVI-syn3A after genome minimization. The format is the same as in Supplementary File 4. The list contains 53 reactions connected to 48 removed genes of metabolic function: 20 genes (17 reactions) pertain to uptake of alternative sugars and conversion to glycolytic intermediates, demonstrating the wider scope of carbon sources JCVI-syn1.0 can use to generate energy. The removal of five genes in nucleotide metabolism further tightened the nucleotide precursor requirements in JCVI-syn3A, which cannot utilize uridine and thymine anymore and also requires adenine (in free base or nucleoside form) since guanine cannot serve as precursor for all purines anymore. Seven genes in JCVI-syn1.0 enable the degradation of certain amino acids as further ATP sources.

Depending on the information available, some of the listed reactions are described in a generic fashion. This list is not meant to be exhaustive in the sense of providing a complete reconstruction for JCVI-syn1.0, but contains reactions encountered during the reconstruction process for JCVI-syn3A and/or that could easily be inferred from the JCVI-syn1.0 genome annotation.