Science
AAAS

Impact of Genome Reduction on Bacterial Metabolism and Its Regulation
Eva Yus, *et al.*
Science **326**, 1263 (2009);
DOI: 10.1126/science.1177263

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

Impact of Genome Reduction on Bacterial Metabolism and Its Regulation

Eva Yus,[1] Tobias Maier,[1] Konstantinos Michalodimitrakis,[1] Vera van Noort,[2] Takuji Yamada,[2] Wei-Hua Chen,[2] Judith A. H. Wodke,[1] Marc Güell,[1] Sira Martínez,[1] Ronan Bourgeois,[1] Sebastian Kühner,[2] Emanuele Raineri,[1] Ivica Letunic,[2] Olga V. Kalinina,[2,3] Michaela Rode,[2] Richard Herrmann,[3] Ricardo Gutiérrez-Gallego,[4] Robert B. Russell,[2] Anne-Claude Gavin,[2] Peer Bork,[2]* Luis Serrano[1,6]

To understand basic principles of bacterial metabolism organization and regulation, but also the impact of genome size, we systematically studied one of the smallest bacteria, *Mycoplasma pneumoniae.* A manually curated metabolic network of 189 reactions catalyzed by 129 enzymes allowed the design of a defined, minimal medium with 19 essential nutrients. More than 1300 growth curves were recorded in the presence of various nutrient concentrations. Measurements of biomass indicators, metabolites, and ^{13}C-glucose experiments provided information on directionality, fluxes, and energetics; integration with transcription profiling enabled the global analysis of metabolic regulation. Compared with more complex bacteria, the *M. pneumoniae* metabolic network has a more linear topology and contains a higher fraction of multifunctional enzymes; general features such as metabolite concentrations, cellular energetics, adaptability, and global gene expression responses are similar, however.

Accurate representation of cellular networks through mathematical models is a central goal of integrative systems biology. For this purpose, all components and reactions of a target system should be listed and validated, and their quantitative relations should be determined and analyzed in the context of the physiology of the organism (*1*). We have selected *Mycoplasma pneumoniae*, a human pathogen that causes atypical pneumonia (*2*), as a model organism for bacterial and archaeal systems biology. Similar to other Mollicutes, *M. pneumoniae* has undergone a massive genome reduction to include only 689 protein coding genes, 231 of which have unknown function (table S1) (*3*), yet it can be cultivated in vitro without helper cells (*4*). The genome reduction of *M. pneumoniae* favors its suitability as a systems biology model because it largely follows genome size–scaling principles (fig. S1) (*5*). We manually reconstructed and validated the metabolic network of *M. pneumoniae* and studied its regulation, complementing analyses of the transcriptome (*6*) and the proteome organization (*7*).

The metabolism of *M. pneumoniae* has been studied biochemically (*8*) and computationally (*9*). We integrated these approaches in a framework that maximized coverage and accuracy (*10*). To build a comprehensive metabolic network, we complemented the reactions from the Kyoto Encyclopedia of Genes and Genomes (KEGG; www.genome.jp/kegg) with activities obtained manually from the literature and new annotations (fig. S2 and tables S1 to S5) (*11*). We also considered other genomic (co-occurrence in one operon), sequence (homology to known enzymes), and structural information (identification of catalytic residues to ensure enzyme functionality) (Fig. 1A and figs. S2 and S3). For example, we identified an incomplete ascorbate pathway through sequence analyses and filled the gap by assigning a critical enzyme [L-ascorbate-6-phosphate lactonase (*mpn497*)] on the basis of sequence homology, predicted activity (metal-dependent hydrolase), and its position in the ascorbate operon (*mpn492* to *mpn497*). For pathways in which only one enzyme was missing, we closed the gap by adding an unassigned reaction (for example, transketolase activity in the pentose phosphate pathway). Putative enzymes missing conserved catalytic residues were discarded (for example, Mpn255 and Mpn673 enzymes of the terpenoid pathway). Lastly, for enzymes that could carry out more than one reaction, we removed the reactions that were decoupled from pathways and those for which the substrate was unavailable. The final result was a map without gaps, isolated reactions, or open metabolic loops (Fig. 2).

A number of alternative pathways, interactions between pathways, as well as missing enzymes still needed to be validated, and reaction directionalities had to be inferred. For this, we used two different experimental strategies. We first used the rich medium (fig. S4) to validate the pathway functionality in various carbon sources. As expected from the map (Fig. 2), all known carbon sources except mannitol supported growth to various extents (figs. S5 and S6) (*12*). Using ^{13}C-glucose labeling, we validated (for example) the predicted connection between glycolysis, the pentose phosphate pathway, and lipid synthesis (fig. S3 and table S6), and ruled out the proposed production of aspartate from pyruvate (*13*). For our second strategy, we developed on the basis of the metabolic map a defined medium (Fig. 1A and table S7) with which we could validate other pathways (such as vitamin metabolism) (fig. S10) and reaction directionalities that could not be studied in rich medium (such as the synthesis of uracyl and thymine nucleotides from cytosine) (figs. S7 and S8). The low number of amino acid permeases and transporters and the existence of a peptide importer (*oppB-F* cluster) (table S1) suggested a requirement for peptides in the medium, which we confirmed experimentally (fig. S9).

We systematically tested the defined medium in more than 1300 experiments in order to properly assess all the components necessary for survival. We replaced these components with simpler building blocks in order to obtain a defined, minimal medium that contains only 26 components (19 of which are essential) (Fig. 1A). This medium, as predicted from our metabolic map and comparative analysis, also supports growth of *M. genitalium* (figs. S11 and S12). On the basis of these experiments, we estimated the upper flux limits for the use of the various nutrients (fig. S13). The medium implicitly validates the reconstructed metabolic map (Fig. 2), which consists of 189 reactions (table S2): 169 are catalyzed by the products of 140 known genes, and 20 are not yet assigned to any gene (table S4). The map includes 74 essential metabolic genes and 34 conditionally essential ones (depending on medium composition), which is in agreement with essentiality as determined by means of transposon mutagenesis analyses (with a 96% overlap) (fig. S14 and table S8) (*14*). A total of 32 enzymes (25%) are multifunctional; they have more than one activity and together catalyze 91 reactions (48% of the total) (table S3). With respect to previous genome annotations (*3*, *15*), we assigned new or refined functions to 57 metabolic genes (plus 30 non-metabolic genes; see the new annotations in table S1). The above strategy could more generally be used to design media to grow axenically hard-to-culture bacteria, as was done for the recalcitrant *Tropheryma whipplei* (*16*) and might be applicable in the context of increasing metagenomics efforts.

Analysis of the metabolism of *M. pneumoniae* reveals that it is more linear than that of larger bacteria, such as *Bacillus subtilis* (Fig. 1B). Furthermore, *M. pneumoniae* has a wider metabolic network diameter (shortest biochemical pathway averaged over all pairs of substrates), although the diameter has been reported to increase with the logarithm of the network size (*17*). The

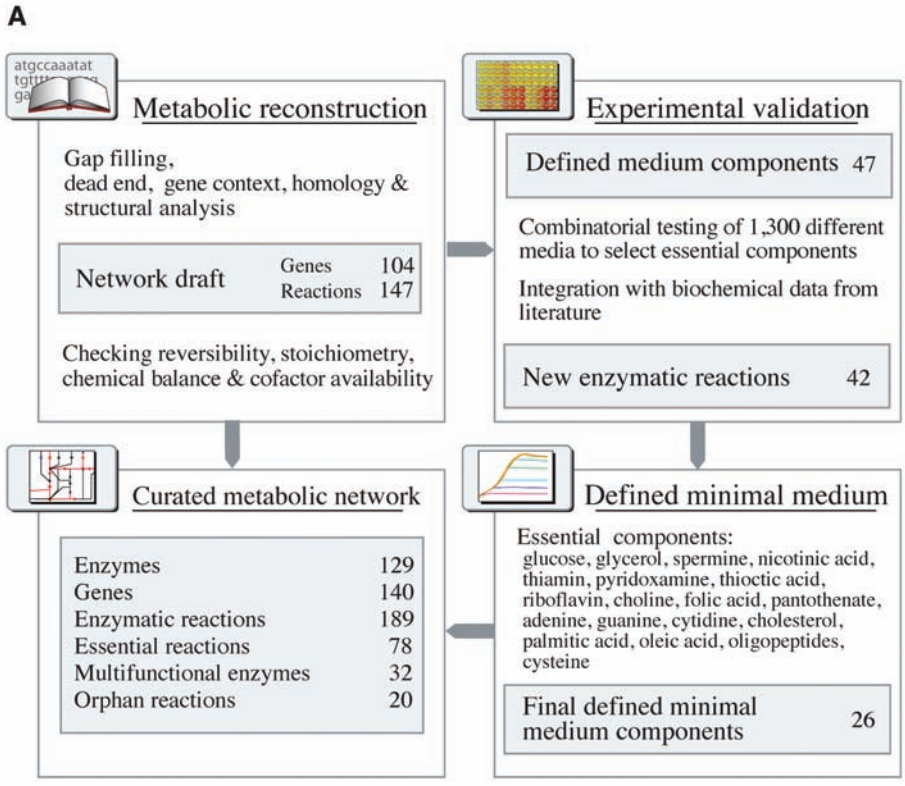
[1]Centre for Genomic Regulation (CRG) and Universitat Pompeu Fabra, Avenida Dr. Aiguader 88, 08003 Barcelona, Spain. [2]European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, D-69117 Heidelberg, Germany. [3]Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 127994, Russia. [4]Zentrum für Molekulare Biologie Heidelberg (ZMBH), Im Neuenheimer Feld 282, 69120 Heidelberg, Germany. [5]Institut Municipal d'Investigació Médica–Hospital del Mar, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Avenida Dr. Aiguader 88, 08003 Barcelona, Spain. [6]Institució Catalana de Recerca i Estudis Avançats, Lluis Companys 23, Barcelona 08010, Spain.

*To whom correspondence should be addressed. E-mail: luis.serrano@crg.es, bork@embl.de

greater linearity and the wider diameter of the network suggest that it is less interconnected and contains fewer parallel paths. Thus, the *M. pneumoniae* network is less redundant both in terms of enzyme paralogy and in network topology. Yet, the distribution of the number of metabolites per reaction is similar to other organisms (fig. S14). This is partly achieved by an increased fraction of multifunctional enzymes as compared with that in larger bacteria, as happens in endosymbionts (*17*). We did not find any evidence of *M. pneumoniae* multifunctional enzymes being more conserved than others. This suggests the larger number could be due to function acquisition that is not present (or detected) in their homologs. This might represent a more general mechanism expected to facilitate further genome reduction (Fig. 1, B and C). The increased linearity and limited redundancy in the metabolic network suggest limited robustness and adaptability to external factors (*18*): Of the metabolic enzymes, 60% are essential (*19*), in contrast to only 15% in *Escherichia coli* (www.shigen.nig.ac.jp/ecoli/pec/index.jsp).

M. pneumoniae has a relatively long duplication time (at least 8 hours) in comparison with *E. coli* or *L. lactis* (20 min), both in culture (*20*) and in the presence of host cells (*21*). Slow growth in genome-reduced, pathogenic bacteria has been proposed to be the result of (i) less efficient enzymatic activity that is explained by the accumulation of mutations resulting from genetic drift (*22*), (ii) a reduced number of ribosomal RNA (rRNA) operons, and/or (iii) other mechanisms related to the adaptation to a pathogenic lifestyle. To understand the causes of slow growth, it is necessary to measure the overall energetics of the metabolic network (Fig. 2) as well as the changes in macromolecules (Fig. 3A) and metabolites along the growth curve (Fig. 3B).

We used the metabolic map, the measured protein concentration (10 fg of protein per cell), and the estimated turnover rates of macromolecules (~20 hours for proteins and ~7 min for mRNA) (table S9 and fig. S15) to estimate the rate of glucose uptake required to duplicate a cell every 8 hours at 18,000 to 24,000 glucose molecules per second [assuming that the majority of adenosine triphosphate (ATP) is used for biomass production] (*10*). This figure closely matched the experimentally determined value under exponential growth: ~19,000 glucose molecules per cell per second (Fig. 3C) (*10*). When cultures approached stationary phase (Fig. 3A), the rate increased to ~45,000 glucose molecules per cell per second (Fig. 3C), concomitantly with the increased transcription of many glycolytic and fermentation genes (Fig. 3D and tables S10 and S11). In both cases, at least 95% of the glucose carbon was found in lactate and acetate (Fig. 3B and fig. S16), implying that the glucose is used primarily for energy production. At the fastest glucose consumption rate, assuming all ATP were devoted to biomass production, *M. pneumoniae* could divide about every



B

	M. pneumoniae	*L. lactis*	*B. subtilis*	*E. coli*
Mean path length (# of reactions)	8.17	5.37	5.67	6.12
Degree average	3.26	8.13	7.54	8.67
Branching nodes (Degree>2)	95 (62%)	381 (80%)	438 (78%)	677 (76%)

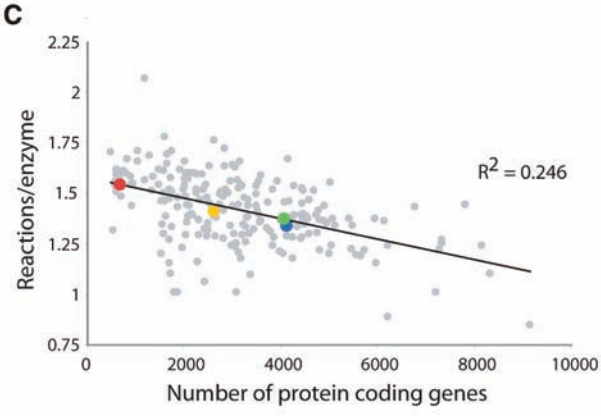


Fig. 1. Metabolic network development and properties and minimal medium design. (**A**) Schematic diagram of the process leading to *M. pneumoniae* metabolic network reconstruction and the design of a minimal medium. (**B**) Comparison of *M. pneumoniae* metabolic network properties with those of other model bacteria. (**C**) Quantification of enzyme multifunctionality among prokaryotic genomes. *M. pneumoniae*, red; *L. lactis*, yellow; *B. subtilis*, green; *E. coli*, blue; and other bacterial species, gray.

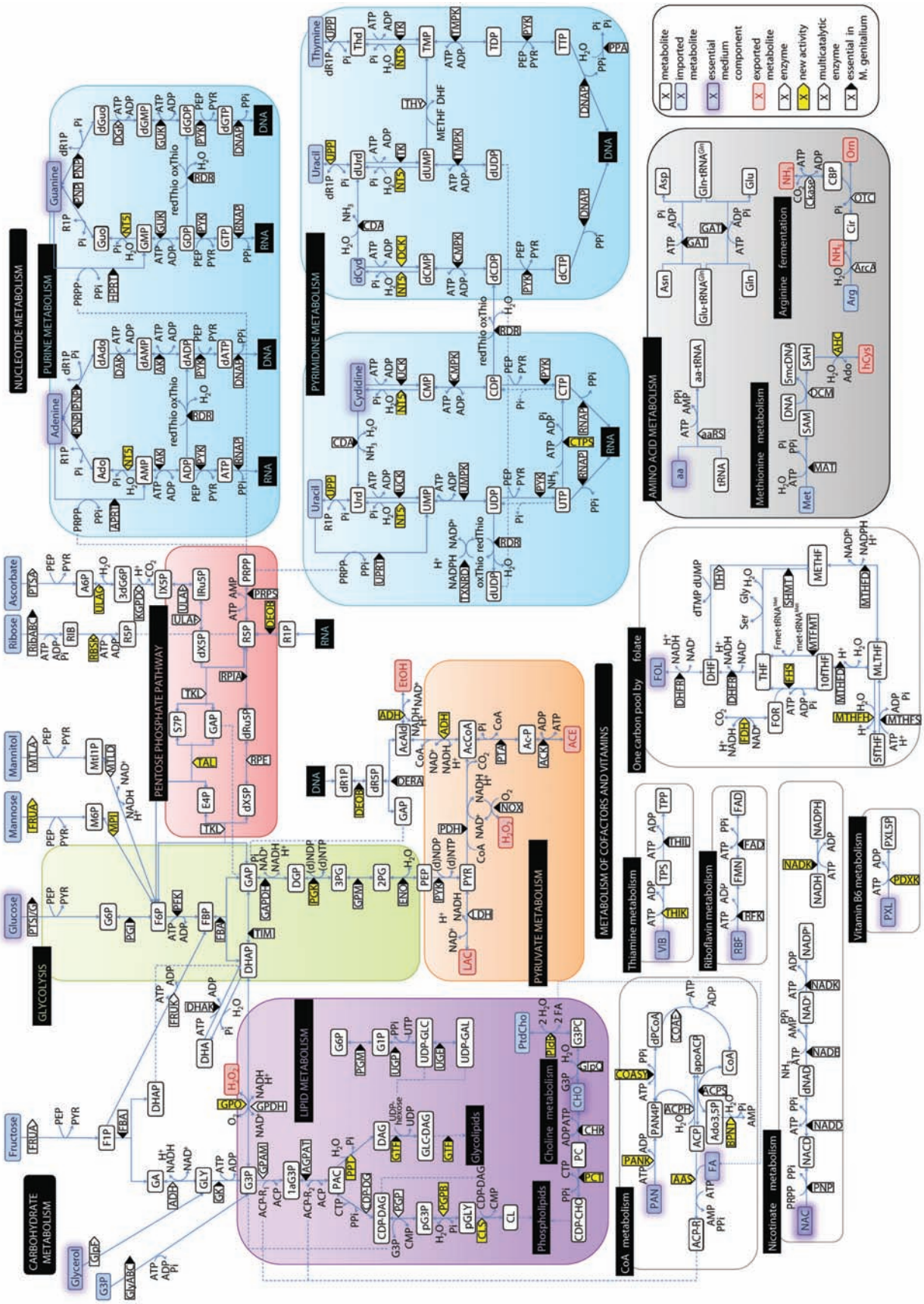


Fig. 2. Metabolic map of *M. pneumoniae*. Main metabolites are shown as boxes, and enzymes and transporters are shown as pentagons. Input metabolites are indicated in blue, and output products are indicated in red. New enzymatic activities determined in this study are displayed in yellow, and enzymes catalyzing multiple reactions are bold. Essential enzymes (according to the mutagenesis study in *M. genitalium*) are indicated with a black triangle. Minimal medium components have been shadowed in blue. See the bottom-right legend for details, fig. S12 and table S2 for description of the enzymatic reactions and enzymes, and table S25 for metabolite abbreviations. aaRS, aminoacyl-tRNA synthase.

3 hours. However, most of the energetic parameters (the concentration of glycolytic intermediates fructose-1,6-biphosphate, glycerol-3-phosphate, phosphoenolpyruvate, glucose-6-phosphate, fructose-6-phosphate, ribose-5-phosphate, and glycerone phosphate, as well as glucose uptake) that we measured were similar to those of larger bacteria (table S9) (*10*), which suggests comparable enzyme efficiencies. This similarity extended to regulatory processes seen in *Lactococcus lactis* (*23*). For example, as in *L. lactis*, we observed both a shift from mostly mixed-acid to homolactic fermentation and an acceleration of glycolysis when the medium acidifies (Fig. 3, A and B); the drop in O_2 concentration relieves inhibition of lactate dehydrogenase (*10*, *22*, *24*). Also, the ATP yield per fermented glucose (two to four ATP, depending on lactate or acetate fermentation) is the same as in *L. lactis* (table S9).

Given all of the above, we cannot explain the slow growth of *M. pneumoniae* on the basis of glycolytic efficiency or ATP yield. One of the main differences compared with fast dividing bacteria is the number of rRNA operons per genome [just one in *M. pneumoniae* and six in *L. lactis* (fig. S17) and five to 10 times proportionally fewer ribosomes as compared with those of *E. coli* (table S10)] (*7*). In many bacteria, the number of ribosomes correlates with the division rate (*25*). For *M. pneumoniae*, we see a correlation of changes in biomass duplication speed with the number of ribosomes but not with the glycolytic rate (Fig. 3, C and D, and fig. S17). We thus suggest that the slow division rate of *M. pneumoniae* is not due to less efficient energy production but to the limit in protein biosynthesis capacity. This small pathogenic bacterium does not appear to be optimized for biomass production. Instead, more complex strategies for fitness, such as suppression of growth by other microorganisms (*26*) or optimization of interactions with host cells, might determine growth rate in small organisms.

It has been suggested that genome-reduced organisms have limited adaptability to external factors (*24*). To determine the capacity of *M. pneumoniae* to respond to environmental changes, we performed three types of experiments. First, we followed the changes in gene expression from the exponential growth phase to the stationary phase (Fig. 4A). Analysis of changes in gene expression (validated by means of tiling arrays and quantitative polymerase chain reaction) (fig. S18) at different points along the growth curve showed that a large part of the transcriptome can be grouped into four time-dependent expression clusters (Fig. 4A, figs. S19 and S20, and tables S11 and S12). These clusters can be regarded as two pairs of anticorrelating patterns, indicating a complex regulation. Subsequent analysis by means of mass spectrometry for a subset of enzymes showed correlation between changes in mRNA and protein abundance (fig. S18 and table S10). For example, the production of lactate by lactate dehydrogenase (Mpn674|Ldh) revealed the close temporal coordination of gene and protein expression and metabolite turnover (Fig. 3, B and D, and table S10).

Second, we analyzed the response of *M. pneumoniae* to specific individual metabolic perturbations encountered as the population grows, such as low pH, accumulation of fermentation end-products, and sugar and amino acid starvation, as well as to more complex stimuli, such as entry into the stationary phase (Fig. 4, B to D, and tables S13 to S16). We found coordinated changes in gene expression specific to each condition (Fig. 4B and fig. S21). For example, there was a general inhibition of transcription and translation upon glucose deprivation and an increase of ATP proton pump genes at pH 6.5 (Fig. 4, B and C). Induction of the stringent response (a global response to the absence of amino acids) results in up-regulation of peptide and amino acid transporters (Fig. 4D). Also, a specific repression of the Thr-tRNA synthetase gene (*mpn553*) (table S17), which is a core component of a tRNA synthetase complex (*7*), suggests its possible regulatory role in complex assembly and therefore in regulation of translation. We found some common responses to multiple stresses. Some were known, such as the

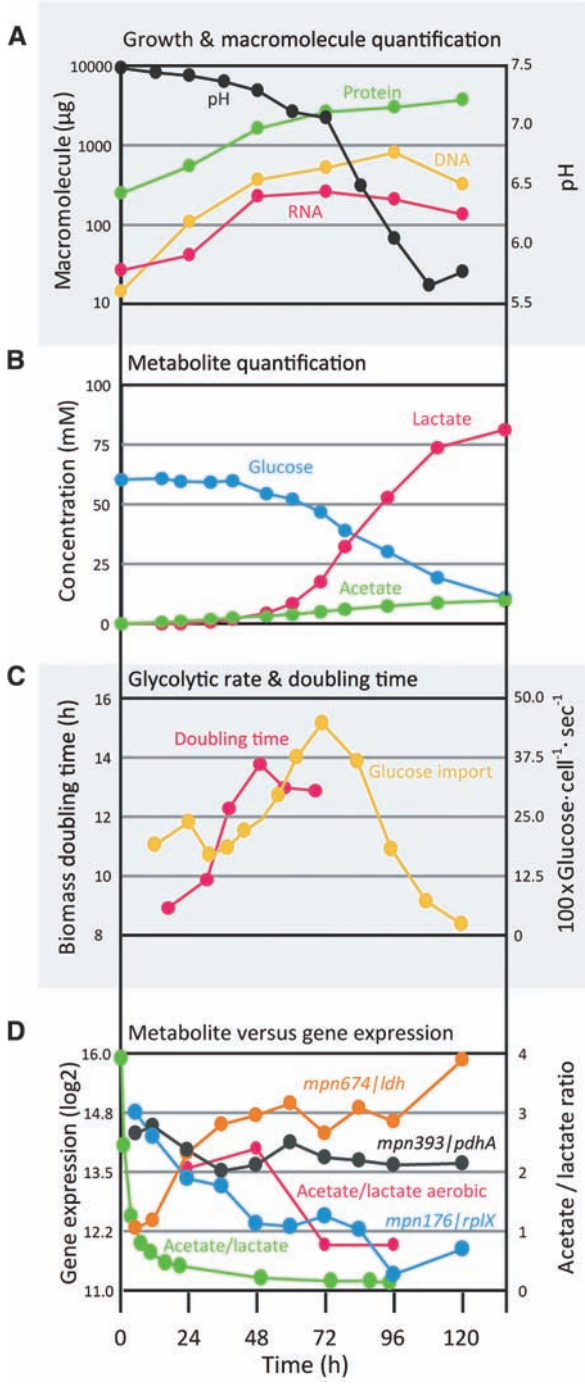


Fig. 3. Determination of various metabolic parameters in growing cultures. Consistently generated heterogeneous data, all derived by using a rich medium, are compared through time in hours (*x* axis). (**A**) *M. pneumoniae* growth determined by monitoring the decrease in extracellular pH and the concomitant changes in the amount of protein, DNA, and total RNA. (**B**) Determination of glucose consumption and its fermentation to lactate and acetate. (**C**) Changes in the number of glucose molecules imported by a cell per second and comparison with the biomass doubling time. (**D**) Changes in gene expression of a representative ribosomal protein (*rplX*) and two enzymes [*ldh* and *pdhA* (a component of the pyruvate dehydrogenase complex), which are enzymes from the two fermentation branches] and the relation with the shift from acetate to lactate production (the ratio between acetate and lactate is shown in green, which can be compared with that of cells grown in the presence of oxygen, shown in red).

down-regulation of ribosomal proteins or peptide importers, which is common to all stresses. Others, like the up-regulation of *ldh* and glycerol-3-P dehydrogenase (*mpn051*), were unexpected and suggest additional functions for these proteins during stress (Fig. 4B).

Third, we adapted the cells by means of serial passage (15 passages) to efficient growth in other carbon sources (fructose, mannose, and glycerol) (tables S18 to S20). Fructose adaptation resulted in overexpression of fruA and fruK (>3 $\log_2$), and mannose-adapted cells overexpressed the mannitol importer (>5 $\log_2$) (tables S19 and S20). Thus, *M. pneumoniae* shows surprising adaptation capability similar to that reported for *E. coli* (*27*).

The coordinated changes in gene expression along the growth curve, the specific responses to many various metabolic perturbations, and the adaptability of the cells to various carbon sources indicate that *M. pneumoniae* retains some robustness and adaptability despite its extreme genome reduction.

Compared with more complex bacteria, *M. pneumoniae* lacks the majority of transcription factors (TFs) regulating metabolic gene factors [such as the catabolite regulation protein (CRP)], major sigma factors, and other regulators (*28*). Gene assignment on the basis of sequence analysis (table S1), in some cases validated through copurification with the RNA polymerase complex (such as *mpn266|spxA*) (*7*), revealed four TFs (*mpn239|gntR*, *mpn329|fur*, and *mpn124|hrcA*), the general sigma 70 factor (*mpn352|sigA*), two putative sigma-like factors (*mpn626|sigD* and *mpn424|ylxM*), and a putative DNA-binding protein (*mpn241|whiA*) (fig. S2 and table S1). Despite this apparently reduced gene regulatory toolbox, both environmental stresses (*6*) and metabolic insults induced complex, specific transcriptional responses; comparison with more complex bacteria showed sim-

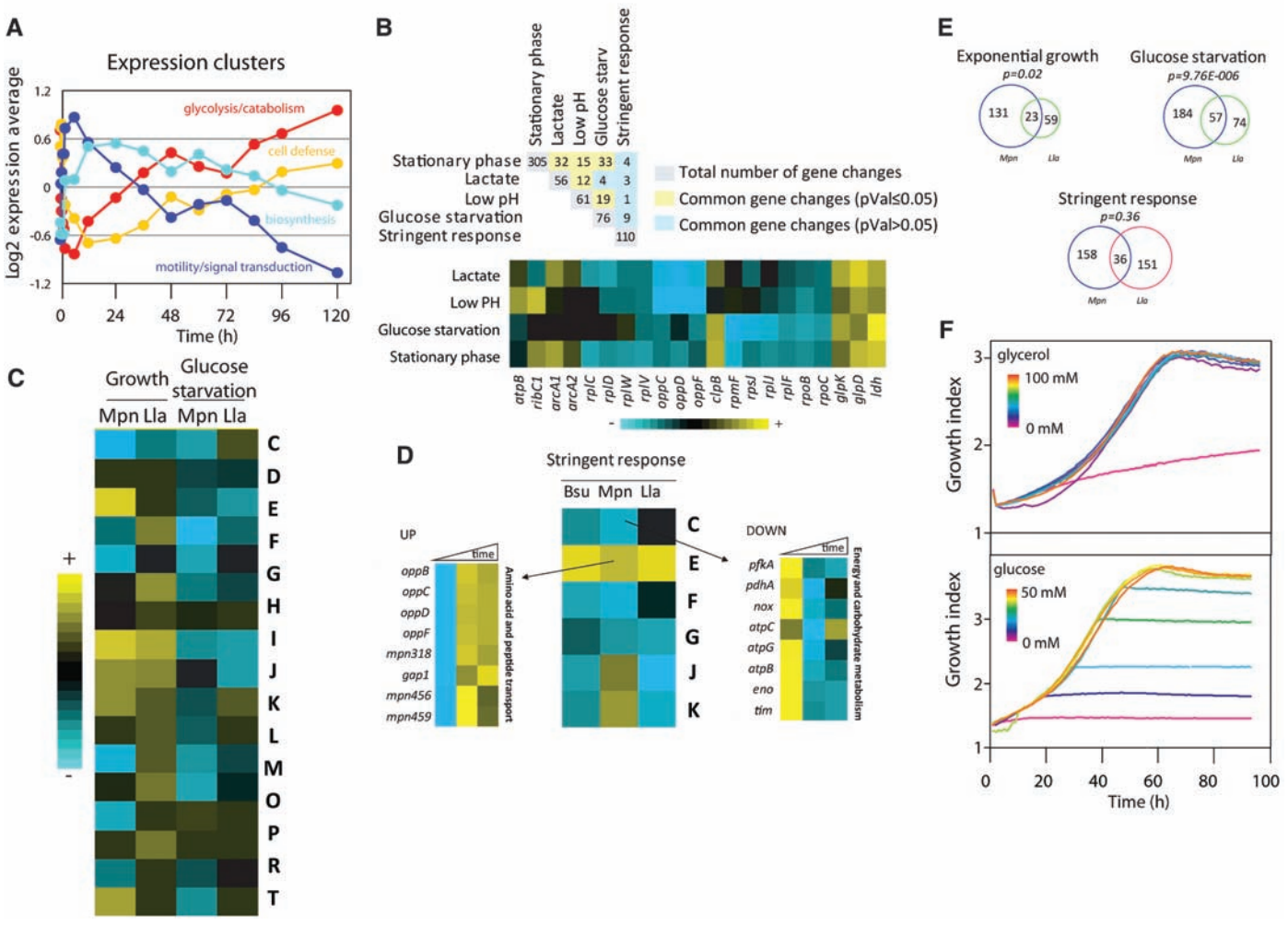


Fig. 4. Regulation of metabolism. (**A**) Representative plot of the four main gene co-expression clusters identified along the growth curve, named after the main functional classes of the genes involved. (**B**)(Top) Overlap between changes in gene expression under various stresses: lactate (80 mM buffered lactate), low pH (pH 6.5), glucose or amino acid starvation (stringent response), and the entry into stationary phase. (Bottom) Heat map of the genes found to be commonly up- or down-regulated under stress and growth inhibition. (**C**) Gene Ontology functional classification of genes significantly regulated during exponential growth and glucose deprivation in *M. pneumoniae* (*Mpn*) or *L. lactis* (*Lla*). The average of the significant changes within each category is depicted. C, energy production and conversion; D, cell division and chromosome partitioning; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme metabolism; I, lipid metabolism; J, translation, ribosomal structure, and biogenesis; K, transcription; L, DNA replication, recombination, and repair; M, cell envelope biogenesis, outer membrane; O, posttranslational modification, protein turnover, and chaperones; P, inorganic ion transport and metabolism; R, general function; T, signal transduction mechanisms. (**D**) Stringent response expression pattern was compared with that of *L. lactis* and *B. subtilis* (Bsu). The average of the significant changes upon stringent response induction (with norvaline) is shown. (**E**) Venn diagrams showing the overlap in *M. pneumoniae* and *L. lactis* of ortholog genes under various metabolic conditions. High *P* value in the case of the stringent response indicates that it is not statistically significant. (**F**) (Top) Growth curve of cells growing in minimal medium plus increasing amounts of glycerol. (Bottom) Glucose titration in Hayflick is shown for comparison.

ilarities but also some specific differences in regulation of gene expression (tables S21 to S23). For example, we observed an increase in mRNA and protein expression of glycolytic enzymes concomitant with the increase of glycolytic rate upon medium acidification (Fig. 3, C and D, fig. S20, and table S15), very similar to what has been described in *L. lactis* cultures (*29*). Response to glucose starvation was also similar to that of *L. lactis* (Fig. 4, C and E, and table S21) (*30*). Part of the stringent response, such as the induction of peptide and amino acid transporters and down-regulation of carbohydrate catabolism (*31*), was conserved in *M. pneumoniae* (table S22); other mechanisms, such as the repression of ribosomal protein operons or rRNA synthesis, were not observed (Fig. 4D and fig. S22). This is in agreement with the proposed involvement of the RNA polymerase omega subunit (missing in *M. pneumoniae*) in sensing guanosine pentaphosphate/tetraphosphate [(p)ppGpp] and thus arresting rRNA biosynthesis (*32*).

We believe it unlikely that the conserved responses, and the specific differences in regulation, can be caused only by combinations of the few TFs that regulate operons and suboperons, even if one includes regulation through antisense RNA (*6*). The presence of genes for synthesis or degradation of a number of chemical messengers, such as (p)ppGpp (*mpn397|spoT*), AppppA (*mpn273|hit1*), or c-di-AMP (*mpn244|disA*) (fig. S2 and table S1) (*33*), implies that signaling mechanisms have been preserved despite genome reduction. For example, overexpression of the spoT gene that regulates (p)ppGpp levels (*31*) results in substantial changes in gene expression, mainly related to the stringent response (table S24). The presence of genes coding for a Ser/Thr phosphatase (*mpn247|ptc1*), two protein kinases (Ser/Thr/Tyr kinase *mpn248|prkC* and *mpn223|hrpK*, an HPr kinase), and the differential phosphorylation of key metabolic enzymes under various growth conditions (*33*) suggest posttranslational control. Also, metabolites such as glycerol regulate gene expression at the base of the fermentation branches in *M. pneumoniae* (*34*) as well as glucose import (*35*). This explains why glycerol is essential in the minimal medium in a concentration-independent manner (Fig. 4F).

Our results suggest that complex metabolic regulation can be achieved in a streamlined genome despite the absence of the respective TFs probably because of a combination of transcriptional regulators, posttranslational modifications, and small molecules, including chemical messengers and metabolites.

Taken together, our newly established *M. pneumoniae* resource, containing a manually annotated metabolic map, full annotations, reactome, consistently measured growth curves, and gene expression profiles corresponding to an extensive list of metabolites, should facilitate integrative systems biology studies at a high resolution. Comparison with more complex bacteria revealed systemic features associated with genome streamlining, which should be examined in other small bacteria. Despite its apparent simplicity, we have shown that *M. pneumoniae* shows metabolic responses and adaptation similar to more complex bacteria, providing hints that other, unknown regulatory mechanisms might exist.

References and Notes

1. A. M. Feist, B. O. Palsson, *Nat. Biotechnol.* **26**, 659 (2008).
2. K. B. Waites, D. F. Talkington, *Clin. Microbiol. Rev.* **17**, 697 (2004).
3. T. Dandekar *et al.*, *Nucleic Acids Res.* **28**, 3278 (2000).
4. T. Proft, R. Herrmann, *Mol. Microbiol.* **13**, 337 (1994).
5. E. van Nimwegen, *Trends Genet.* **19**, 479 (2003).
6. M. Güell *et al.*, *Science* **326**, 1268 (2009).
7. S. Kühner *et al.*, *Science* **326**, 1235 (2009).
8. J. D. Pollack, *Trends Microbiol.* **5**, 413 (1997).
9. M. Pachkov, T. Dandekar, J. Korbel, P. Bork, S. Schuster, *Gene* **396**, 215 (2007).
10. Materials and methods are available as supporting material on *Science* Online.
11. J. D. Pollack, M. V. Williams, R. N. McElhaney, *Crit. Rev. Microbiol.* **23**, 269 (1997).
12. S. Halbedel, C. Hames, J. Stulke, *J. Bacteriol.* **186**, 7936 (2004).
13. J. T. Manolukas, M. F. Barile, D. K. Chandler, J. D. Pollack, *J. Gen. Microbiol.* **134**, 791 (1988).
14. C. A. Hutchison *et al.*, *Science* **286**, 2165 (1999).
15. R. Himmelreich *et al.*, *Nucleic Acids Res.* **24**, 4420 (1996).
16. P. Renesto *et al.*, *Lancet* **362**, 447 (2003).
17. E. Zientz, T. Dandekar, R. Gross, *Microbiol. Mol. Biol. Rev.* **68**, 745 (2004).
18. L. Z. Pipe, M. J. Grimson, *Mol. Biosyst.* **4**, 192 (2008).
19. J. I. Glass *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 425 (2006).
20. B. M. Hasselbring, J. L. Jordan, R. W. Krause, D. C. Krause, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 16478 (2006).
21. S. F. Dallo, J. B. Baseman, *Microb. Pathog.* **29**, 301 (2000).
22. J. B. Smart, T. D. Thomas, *Appl. Environ. Microbiol.* **53**, 533 (1987).
23. N. A. Moran, *Cell* **108**, 583 (2002).
24. M. Cocaign-Bousquet, S. Even, N. D. Lindley, P. Loubiere, *Appl. Microbiol. Biotechnol.* **60**, 24 (2002).
25. M. Nomura, *J. Bacteriol.* **181**, 6857 (1999).
26. B. Teusink *et al.*, *J. Biol. Chem.* **281**, 40041 (2006).
27. E. Oxman, U. Alon, E. Dekel, *Evolution* **62**, 1547 (2008).
28. W. Goebel, S. Lory, *Curr. Opin. Microbiol.* **9**, 123 (2006).
29. S. Even, N. D. Lindley, M. Cocaign-Bousquet, *Microbiology* **149**, 1935 (2003).
30. S. Even, N. D. Lindley, P. Loubiere, M. Cocaign-Bousquet, *Mol. Microbiol.* **45**, 1143 (2002).
31. K. Potrykus, M. Cashel, *Annu. Rev. Microbiol.* **62**, 35 (2008).
32. C. E. Vrentas, T. Gaal, W. Ross, R. H. Ebright, R. L. Gourse, *Genes Dev.* **19**, 2378 (2005).
33. H. C. Su, C. A. Hutchison 3rd, M. C. Giddings, *BMC Microbiol.* **7**, 63 (2007).
34. S. Halbedel *et al.*, *J. Mol. Biol.* **371**, 596 (2007).
35. S. Halbedel, C. Hames, J. Stulke, *J. Mol. Microbiol. Biotechnol.* **12**, 147 (2007).

36. The authors thank V. Benes and S. Schmidt of the Gene Core facility at EMBL (Heidelberg, Germany) for hybridization of samples in custom made arrays; A. Wieslander (Stockholm University, Sweden) for advice on the lipid requirements for the minimal medium; J. Marcos del Aguila (Pompeu Fabra University, Barcelona, Spain) for conceptual and experimental input on the gas chromatography–mass spectrometry measurements; T. Doerks (EMBL, Heidelberg, Germany) for operon analysis and help in annotation; and J. Leigh-Bell for editorial help. L.S. is an Institució Catalana de Recerca i Estudis Avançats professor.

Supporting online Material

www.sciencemag.org/cgi/content/full/326/5957/1263/DC1
Materials and Methods
SOM Text
Figures S1 to S22
Tables S1 to S25
References

3 June 2009; accepted 2 October 2009
10.1126/science.1177263

Transcriptome Complexity in a Genome-Reduced Bacterium

Marc Güell,[1] Vera van Noort,[2] Eva Yus,[1] Wei-Hua Chen,[2] Justine Leigh-Bell,[1] Konstantinos Michalodimitrakis,[1] Takuji Yamada,[2] Manimozhiyan Arumugam,[2] Tobias Doerks,[2] Sebastian Kühner,[2] Michaela Rode,[2] Mikita Suyama,[2]* Sabine Schmidt,[2] Anne-Claude Gavin,[2] Peer Bork,[2]† Luis Serrano[1,3]†

To study basic principles of transcriptome organization in bacteria, we analyzed one of the smallest self-replicating organisms, *Mycoplasma pneumoniae*. We combined strand-specific tiling arrays, complemented by transcriptome sequencing, with more than 252 spotted arrays. We detected 117 previously undescribed, mostly noncoding transcripts, 89 of them in antisense configuration to known genes. We identified 341 operons, of which 139 are polycistronic; almost half of the latter show decaying expression in a staircase-like manner. Under various conditions, operons could be divided into 447 smaller transcriptional units, resulting in many alternative transcripts. Frequent antisense transcripts, alternative transcripts, and multiple regulators per gene imply a highly dynamic transcriptome, more similar to that of eukaryotes than previously thought.

Although large-scale gene expression studies have been reported for various bacteria (*1*–*7*), comprehensive strand-specific data sets are still missing, limiting our understanding of operon structure and regulation. Similarly, the number of classified noncoding RNAs in bacteria has recently been expanded (*8*), but a complete and unbiased repertoire is still not available. To obtain a blueprint of bacterial transcription, we combined the robustness and versatility of spotted arrays [62 independent conditions and 252 array experiments (*9*)], the superior res-