The following resources related to this article are available online at www.sciencemag.org (this information is current as of January 13, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/cgi/content/full/326/5957/1268

Supporting Online Material can be found at:
http://www.sciencemag.org/cgi/content/full/326/5957/1268/DC1

A list of selected additional articles on the Science Web sites **related to this article** can be found at:
http://www.sciencemag.org/cgi/content/full/326/5957/1268#related-content

This article **cites 27 articles**, 11 of which can be accessed for free:
http://www.sciencemag.org/cgi/content/full/326/5957/1268#otherarticles

This article has been **cited by** 3 articles hosted by HighWire Press; see:
http://www.sciencemag.org/cgi/content/full/326/5957/1268#otherarticles

This article appears in the following **subject collections**:
Molecular Biology
http://www.sciencemag.org/cgi/collection/molec_biol

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 stream-

lining, 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.

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-

olution of strand-specific tiling arrays (Fig. 1A) (designed after genome resequencing, table S1), and the mapping capacity of RNA deep sequencing [direct strand-specific sequencing (DSSS)] (Fig. 1A and fig. S1) to analyze one of the smallest bacteria that can live outside a host cell, *Mycoplasma pneumoniae*, with 689 annotated protein-coding genes and 44 noncoding RNAs (ncRNAs).

Considering DSSS under reference conditions (*9*) and 43 tiling arrays from four time series (growth curve, heat shock, DNA damage, and cell cycle arrest) (table S8), we observed the expression of all genes. Using a segmentation algorithm for the tiling arrays (*10*), we identified an additional 117 regions with no previous annotation (table S2) (*9*). These regions were further confirmed by DSSS (Fig. 1B and fig. S1) and in four cases by quantitative polymerase chain reaction (table S3). Sequence similarity with known proteins revealed the presence of two previously undescribed protein-coding genes, one pseudogene, one N-terminal truncation, and five 5′ extensions of known genes (table S2). The remaining 108 transcripts are probably regulatory rather than

¹Centre for Genomic Regulation (CRG), Universitat Pompeu Fabra, Barcelona, Spain. ²European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. ³Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain.

*Present address: Center for Genomic Medicine, Graduate School of Medicine, University of Kyoto, Kyoto, Japan.
†To whom correspondence should be addressed. E-mail: bork@embl.de (P.B.); luis.serrano@crg.es (L.S.)

structural RNAs, because comparison of their predicted secondary structures with the ones of coding genes does not show any substantial difference (*9*). Eighty-nine of them are antisense with respect to previously annotated genes. Out of the nonoverlapping ones, two of them (NEW87 and NEW8) are conserved in *M. genitalium* and could be involved in DNA replication and repair, and in peptide transport, respectively (*9*) (figs. S3 to S5). In total, 13% of the coding genes are covered by antisense; this is twice more than in yeast (7%) (*11*) and about half of what was reported for plants (22.2%) (*12*, *13*) or humans (22.6%) (*14*). Antisense transcripts may affect expression of the overlapping functional sense transcripts through several mechanisms (*15*): Double-stranded RNA-dependent mechanisms require coexpression with their targets (*16*), whereas transcriptional interference implies the mutual exclusion of sense and antisense transcripts (*17*, *18*). In *M. pneumoniae*, we observed a predominance of double-stranded RNA mechanisms as in mammals (*19*) (47% positive correlation versus 2% negative correlation). In addition, we detected a reduced expression level of genes targeted by antisense transcripts, as reported in some prokaryotes (*9*, *18*) (fig. S6).

We identified operon boundaries through sharp transcription changes in the tiling reference condition by using local convolution methods (Fig. 1A) (*9*, *20*). More than 90% of the operons (139 polycistronic and 202 monocistronic oper-

ons, table S4) were well supported by DSSS reads [DSSS alone was not sufficient to unambiguously characterize operons (fig. S2) (*9*)]. Most polycistronic operons contain two or three genes (Fig. 1C, fig. S7, and table S4); the largest one is the ribosomal operon containing 20 genes. For the majority of operons, we observed a canonical or slightly altered version of a standard sigma 70 promoter region (fig. S8), with transcription starts located within 60 base pairs (fig. S9) upstream of the translation start (*6*). In contrast to previous suggestions (*21*), we observed, as proposed by others (*22*), a preferential use of termination hairpins for the tight regulation of gene expression (Fig. 1, A and D, and table S5). Moreover, we found that almost half of the consecutive genes within polycistronic operons show a decay behavior (Fig. 1A and fig. S1), indicating that such staircase-like expression is a widespread phenomenon in bacteria (*9*).

Analysis of the 43 tiling arrays and integration with 252 spotted arrays representing 173 independent conditions, some of them from time series, revealed context-dependent modulation of operon structures involving repression or activation of operon internal genes, as well as of genes located at the beginning or end (Fig. 2, A and B, fig. S10, and table S5). In some cases this modulation can be assigned to specific environmental changes. Down-regulation of the first four genes of the *ftsZ* operon involved in the initiation of cell division corresponds to entry into stationary phase (Fig.
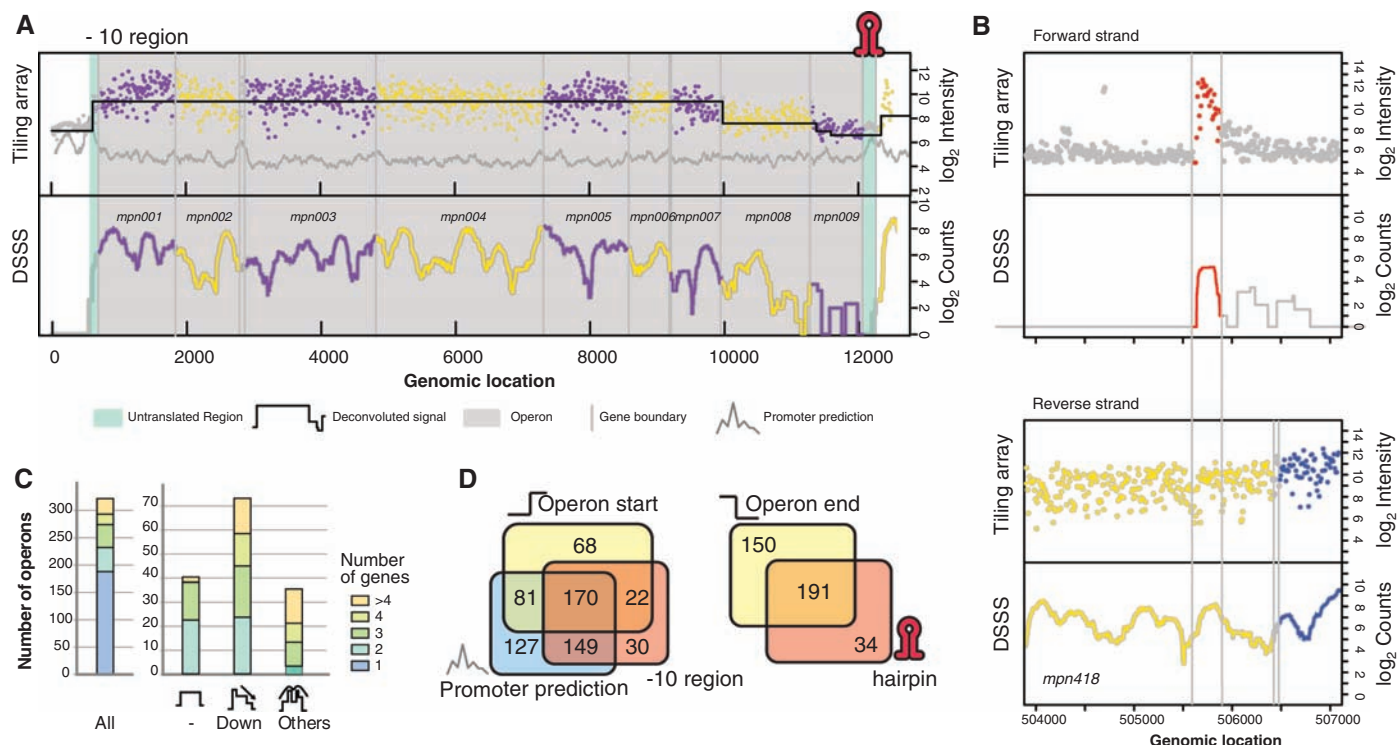

Fig. 1. Transcriptome feature in the reference condition. (**A**) The first operon in the genome on the forward strand has a staircase behavior, meaning that the consecutive genes have lower and steady expression levels. (**B**) Example of an antisense RNA transcript. (**C**) Analysis of staircase operons. (Left) All reference operons subdivided by the number of protein-coding genes they contain. (Right) All reference operons subdivided by their staircase behavior (see bottom graphs). (**D**) (Left) Overlap of operon starts and single-gene starts with previously identified −10 promoter sequence motifs in *M. pneumoniae* (*29*) and predicted promoters based on hexamers. (Right) Overlap of operon ends and single-gene ends with predicted transcription termination hairpins.

Fig. 2. Alternative operon structure. The continuous lines in (A) and (B) indicate expression level measured with tiling arrays. (**A**) Alternative transcripts discovery pipeline (*9*). Reference operon 001 is split into three suboperons. (Top) Tiling and DSSS under reference conditions. (Middle) Specific expression changes for genes **dNAa** and *xdj1* involved in DNA repair and replication. (Bottom) The coexpression matrices correspond to the final conditional operon splitting by 252 arrays. (**B**) Two examples of conditional operons are presented. (Top) Specific induction of the middle genes in operon 126 when the cells reach stationary phase. (Bottom) Repression of the first four genes of the operon 129 involved in cell division, when the cells reach stationary phase. (**C**), Example of heat shock–induced genes sharing the known CIRCE element. The calculated consensus sequence is represented below.
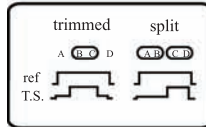

A Operon splitting

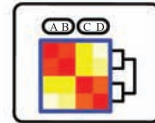STEP I: Reference operon definition

| monocistronic | 202 |
| polycistronic | 139 |

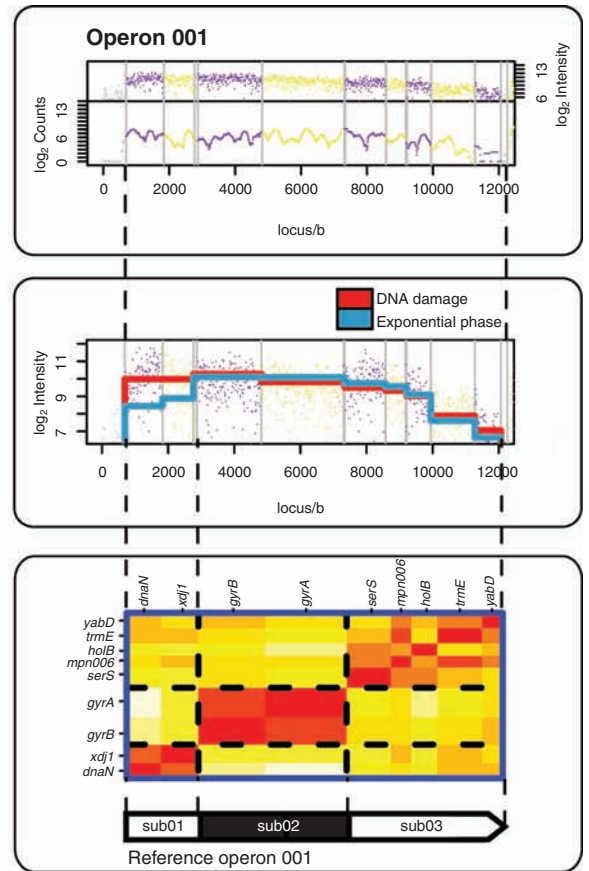STEP II: Reference operon splitting based on multiple condition tiling arrays

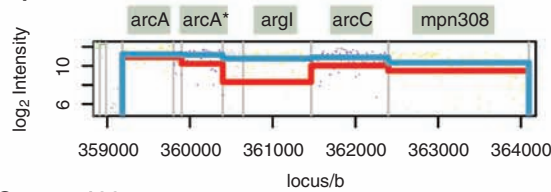| monocistronic | 236 |
| polycistronic | 135 |

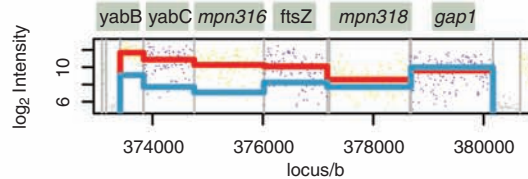STEP III: Further splitting based on 252 spotted microarray

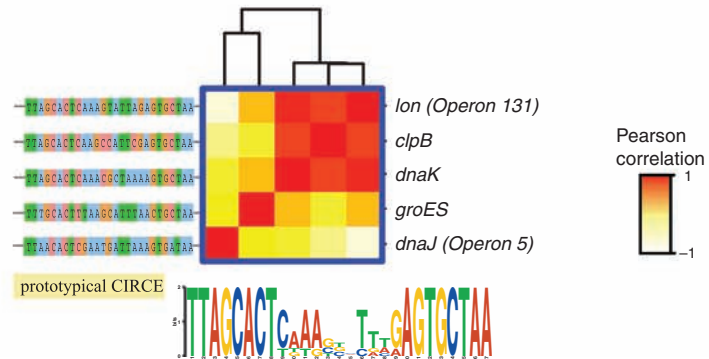| monocistronic | 336 |
| polycistronic | 111 |

B Examples of suboperon dynamics

C Differential expression of dnaJ and groES despite CIRCE element

2B, lower panel). An increase in the expression of arginine fermentation genes (*arcA*, *arcI*, and *arcC*) (Fig. 2B) in stationary phase could be a mechanism to cope with acidification (*23*). We found formal evidence for a total of 447 transcriptional units (336 monocistronic and 111 polycistronic), implying a high rate of alternative transcripts (42%) in this bacterium under the conditions studied, similar to that in eukaryotes (40%, although still under debate) (*24*) and archea (40% in *H. salinarum*) (*25*). We found that genes that are split into different suboperons tend to belong to different functional categories (*9*). Thus, although genome reduction leads to longer operons accommodating genes with different functions (*26*), the latter can still retain internal transcription and termination sites under certain conditions.

The high frequency of alternative transcripts of *M. pneumoniae* genes hints at a situation similar to that in eukaryotes, where many factors contribute to the regulation of gene expression. To further support this hypothesis, we used gene expression clustering under the 62 distinct conditions (table S7) to identify groups of coexpressed genes and their possible common regulatory motifs. Using a correlation cutoff of 0.65, we identified 94 coexpression groups (table S6 and fig. S11), encompassing 416 genes. Thirty of the clusters contained genes from more than two operons. Of these, 14 share a specific sequence motif in their upstream region and another 8 have a specific combination of motifs (fig. S12), which might drive the coexpression (for example, 4 of the 14 motifs are found at splitting sites inside operons). The rest of the genes did not group together, implying complex and multiple levels of regulation orchestrated by the various environmental conditions. This is exemplified by the five heat shock–induced genes containing a regulatory CIRCE (controlling inverted repeat of chaperone expression) element (*27*) (Fig. 2C). Not all of them clustered together, indicating at least one other regulatory element. Similarly, overexpression of a transcription factor (Fur, ferric uptake regulator) reveals a common motif in all genes significantly changing expression, although they belong to different coexpression clusters (fig. S13 and table S6).

Our work revealed an unanticipated complexity in the transcriptome of a genome-reduced bacterium. This complexity cannot be explained by the presence of eight predicted transcription factors (*26*). Furthermore, the fact that the proteome organization is not explainable by the genome organization (*28*) indicates the existence of other regulatory processes. The surprisingly frequent expression heterogeneity within operons, the change of operon structures leading to alternative transcripts in response to environmental perturbations, and the frequency of antisense RNA, which might explain some of these expression changes, suggest that transcriptional regulation in bacteria resemble that of eukaryotes more than previously thought.

References and Notes

1. B. Tjaden *et al.*, *Nucleic Acids Res.* **30**, 3732 (2002).
2. D. W. Selinger *et al.*, *Nat. Biotechnol.* **18**, 1262 (2000).
3. N. B. Reppas, J. T. Wade, G. M. Church, K. Struhl, *Mol. Cell* **24**, 747 (2006).
4. C. M. Nelson *et al.*, *BMC Genomics* **9**, 364 (2008).
5. T. Akama *et al.*, *J. Bacteriol.* **191**, 3321 (2009).
6. P. T. McGrath *et al.*, *Nat. Biotechnol.* **25**, 584 (2007).
7. A. Toledo-Arana *et al.*, *Nature* **459**, 950 (2009).
8. J. Vogel, E. G. Wagner, *Curr. Opin. Microbiol.* **10**, 262 (2007).
9. Materials and methods are available as supporting material on *Science* Online.
10. W. Huber, J. Toedling, L. M. Steinmetz, *Bioinformatics* **22**, 1963 (2006).
11. Z. Xu *et al.*, *Nature* **457**, 1033 (2009).
12. X. J. Wang, T. Gaasterland, N. H. Chua, *Genome Biol.* **6**, R30 (2005).
13. S. R. Henz *et al.*, *Plant Physiol.* **144**, 1247 (2007).
14. X. Ge, Q. Wu, Y. C. Jung, J. Chen, S. M. Wang, *Bioinformatics* **22**, 2475 (2006).
15. M. Lapidot, Y. Pilpel, *EMBO Rep.* **7**, 1216 (2006).
16. S. Brantl, E. G. Wagner, *EMBO J.* **13**, 3599 (1994).
17. G. Andre *et al.*, *Nucleic Acids Res.* **36**, 5955 (2008).
18. S. Brantl, *Curr. Opin. Microbiol.* **10**, 102 (2007).
19. S. Katayama *et al.*, *Science* **309**, 1564 (2005).
20. S. D. Hooper *et al.*, *Mol. Syst. Biol.* **3**, 72 (2007).
21. T. Washio, J. Sasayama, M. Tomita, *Nucleic Acids Res.* **26**, 5456 (1998).
22. M. J. de Hoon, Y. Makita, K. Nakai, S. Miyano, *PLOS Comput. Biol.* **1**, e25 (2005).
23. A. Budin-Verneuila, E. Maguin, Y. Auffraya, S. Dusko Ehrlich, V. Pichereaua, *Lait* **84**, 8 (2004).
24. S. Boue, I. Letunic, P. Bork, *Bioessays* **25**, 1031 (2003).
25. T. Koide *et al.*, *Mol. Syst. Biol.* **5**, 285 (2009).
26. E. Yus *et al.*, *Science* **326**, 1263 (2009).
27. L. J. Chang, W. H. Chen, F. C. Minion, D. Shiuan, *Biochem. Biophys. Res. Commun.* **367**, 213 (2008).
28. S. Kühner *et al.*, *Science* **326**, 1235 (2009).
29. J. Weiner III, R. Herrmann, G. F. Browning, *Nucleic Acids Res.* **28**, 4488 (2000).
30. We thank the Genomics core facility at EMBL (Heidelberg, Germany), J. Lozano for help with statistical analysis, and the Ultrasequencing Unit at CRG. This work was funded by the Foundation Marcelino Botín, the Ministry of Education of Spain (MEC)–Consolider, and the European Research Council. M.G. is funded by the Spanish MEC–Formación Profesorado Universitario. V.N. is funded by the Netherlands Organization for Scientific Research.

Crystal Structure of the Catalytic Core of an RNA-Polymerase Ribozyme

David M. Shechner,[1,2] Robert A. Grant,[2] Sarah C. Bagby,[1,2] Yelena Koldobskaya,[3] Joseph A. Piccirilli,[3] David P. Bartel[1,2]*

Primordial organisms of the putative RNA world would have required polymerase ribozymes able to replicate RNA. Known ribozymes with polymerase activity best approximating that needed for RNA replication contain at their catalytic core the class I RNA ligase, an artificial ribozyme with a catalytic rate among the fastest of known ribozymes. Here we present the 3.0 angstrom crystal structure of this ligase. The architecture resembles a tripod, its three legs converging near the ligation junction. Interacting with this tripod scaffold through a series of 10 minor-groove interactions (including two A-minor triads) is the unpaired segment that contributes to and organizes the active site. A cytosine nucleobase and two backbone phosphates abut the ligation junction; their location suggests a model for catalysis resembling that of proteinaceous polymerases.

The RNA world hypothesis proposes that early life forms lacked DNA and coded proteins, depending instead on RNA for both chemical catalysis and information storage (*1*). Central to this RNA world would have been polymerase ribozymes able to replicate RNA. Among the efforts to generate ribozymes with such ability, the most productive have started with the class I RNA ligase ribozyme (*2–4*). This ribozyme was originally isolated from a large pool of random sequences (*5*, *6*). It has since been improved by mutation and selection and has served as a platform for modeling ribozyme evolution in vitro (*6–8*). Because it rapidly promotes a reaction with chemistry identical to that catalyzed by proteinaceous enzymes that replicate RNA (Fig. 1A) (*6*), the ligase has provided the catalytic engine for more sophisticated RNA enzymes that use nucleoside triphosphates and the information from an external RNA template to synthesize short strands of RNA (*2*, *3*, *9*). Although more efficient with some templates than with others, this primer-extension reaction is gen-

[1]Whitehead Institute for Biomedical Research and Howard Hughes Medical Institute, 9 Cambridge Center, Cambridge, MA 02142, USA. [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Department of Chemistry, Department of Biochemistry and Molecular Biology, and Howard Hughes Medical Institute, University of Chicago, Chicago, IL 60637, USA.

*To whom correspondence should be addressed. E-mail: dbartel@wi.mit.edu