Here, we report the genome-wide identification of small RNAs associated with transcription start sites (TSSs), termed tssRNAs, in Mycoplasma pneumoniae. tssRNAs were also found to be present in a different bacterial phyla, Escherichia coli. Similar to the recently identified promoter-associated tiny RNAs (tiRNAs) in eukaryotes, tssRNAs are associated with active promoters. Evidence suggests that these tssRNAs are distinct from previously described abortive transcription RNAs. ssRNAs have an average size of 45 bases and map exactly to the beginning of cognate full-length transcripts and to cryptic TSSs. Expression of bacterial tssRNAs requires factors other than the standard RNA polymerase holoenzyme. We have found that the RNA polymerase is halted at tssRNA positions in vivo, which may indicate that a pausing mechanism exists to prevent transcription in the absence of genes. These results suggest that small RNAs associated with TSSs could be a universal feature of bacterial transcription.

Molecular Systems Biology 8; Article number 585; doi:10.1038/msb.2012.16
Subject Categories: RNA; microbiology & pathogens
Keywords: non-coding RNAs; small RNAs; transcription; transcriptomics
an abrupt raise and a flat plateau (Supplementary Figure S2A). We also observed more complex patterns, such as two or more overlapping distributions, indicative of multiple promoters (Supplementary Figure S2B).

In order to quantify these small RNAs automatically and consistently, we developed a computational method that identifies narrow peaks with a flat plateau significantly above the background (Figure 1A; Table I; Supplementary information). Automatic analysis of *M. pneumoniae* small transcripts obtained from stationary phase cells (Yus et al., 2009) allowed 1339 ± 116 small RNAs to be identified (using data from three independent biological replicates; Supplementary Tables S3 and S4), of which 457 ± 28 (~34%) were located <10 bases away from a manually annotated TSS (annotated in this study based on the published data in Guell et al., 2009) (Table I; Supplementary Table S1). In total, ~73% of the *M. pneumoniae* TSSs have an automatically assigned small RNA (Table I). Visual inspection of the missing TSSs reveals that all of them have an associated RNA that was not identified by the algorithm due to a complicated shape or low height (Supplementary Figure S4B). In this way, we identified 1371 small RNAs on the plus strand from the *E. coli* data set. Using the Regulon database (Gama-Castro et al., 2008) to extract a high-confidence group of sigma 70-dependent TSSs on the plus strand (Table I; Supplementary Table S2), we reproducibly found that a somewhat smaller (but still large) proportion of active TSSs have associated small RNAs in the stationary phase (44 ± 5%; Table I). In both species, the number of small RNAs decreased in the exponential phase (from 1239 to 818 for *M. pneumoniae*, and from 1371 to 684 for *E. coli*; Table I). The distance from the small RNA start position to the experimentally determined TSS of the cognate full-length transcript overlapped significantly (P = 0.00015; Supplementary Figure S4B), with differences between the small RNA starting positions and the annotated TSSs of ~0.5 ± 10 bases in *M. pneumoniae*, and of ~3.5 ± 12 bases in *E. coli*. At each TSS, we observed a dominant species of small RNAs, with some minor ones that started at the same point but had slightly different lengths (Supplementary Figure S5). These results suggest that these newly identified small RNAs are transcribed from the promoters of the corresponding cognate full-length transcripts. Hence, we named these ‘transcription start site-associated’ RNAs (tssRNAs). We propose that tssRNAs could be used as markers for promoters in uncharacterized genomes. They could also help to identify ambiguous TSSs, for example in regions where transcripts overlap or when no clear boundaries (e.g., start and end) can be observed at the RNA level (see scheme in Figure 1A, and example in Figure 2A).

We next applied a number of independent approaches to confirm the existence of tssRNAs and to rule out possible technical artifacts. We found that (i) tssRNAs were also observed when using RNA-seq protocols that do not require fractionation of RNAs by size (using TrueSeq, from Illumina) (Figures 1B and 2B); (ii) tssRNAs were unequivocally detected on tiling arrays hybridized with total cDNA (see Supplementary information; Figures 1B and 2D) and, similarly, deep sequencing of non-size-fractionated RNAs showed a clear peak at the TSSs from low expression genes (allowing co-detection of the cognate tssRNAs) (Supplementary Figure...
S6B); (iii) direct hybridization of fluorescently labeled small RNAs (<65 bases) onto tiling arrays further substantiated the presence of tssRNAs (Figure 1B; Supplementary Figure S6A); and (iv) the existence of tssRNAs was directly confirmed by Northern blot analyses (Figure 3B; Supplementary Figure S3). Since the 5’ ends of bacteria transcripts bear a triphosphate, while RNA degradation products have a single phosphate, this treatment should remove all RNA degradation products (Sharma et al., 2010). tssRNAs remained largely unaffected by this treatment, while nascent RNA (Taft et al., 2009). To see if this is also the case in bacteria, we exposed the purified small RNAs to terminator 5’-phosphate-dependent exonuclease (Tex) and analyzed as in (A). Second panel: untreated control. Treated samples contained a higher ratio of tssRNAs with respect to degradation products. This demonstrates that tssRNAs are capped with 5’-triphosphate and are thus nascent transcripts. Lower panel: mRNA levels of the genomic region as measured by DSSS. Untreated samples were subtracted from TAP-treated samples to remove the background value (see Supplementary Methods). Average normalized counts (shown in red) were estimated for positions where tssRNAs start. To determine the null distribution for random selected points in the genome (shown in gray), 500 groups with equal sizes as the tssRNAs were selected. To determine the number of TSSs for which an associated tssRNA was automatically detected. T otal tssRNAs indicates the total number of tssRNAs that were automatically detected.

**Figure 2** Properties of tssRNAs. (A) Example of tssRNA identification by DSSS of RNAs with fewer than 65 bases in a *M. pneumoniae* cytoadherence operon (mpn309-mpn312). ORFs are shown in alternative colors. A second, internal TSS (TSS2) was easily detected due to the presence of a tssRNA, and revealed the presence of a suboperon or alternative transcript. (B) Size distribution of tssRNAs by a RNA-seq method that does not involve RNA fragmentation (TrueSeq) showed an average size of 45 bases. Pile-ups were built by selecting reads containing the indicated sizes, and normalized by the mean number of counts. Untreated samples were used as a control. Treated samples contained a higher ratio of tssRNAs with respect to degradation products. This demonstrates that tssRNAs are capped with 5’-triphosphate and are thus nascent transcripts. Lower panel: mRNA levels of the genomic region as measured by DSSS. (D) Plot showing the distribution of log2 values for tiling array data around the TSS start site of 49 genes that showed a clear tssRNA over the mRNA level. There was a peak at the TSS proximity that spans about three bases in a 44-base length, similar to that found by RNA-seq (see Materials and methods).

**Table 1** Presence of tssRNAs at TSSs

| Organism                  | TSS (strands) | TSS tssRNAs (bp) | Total tssRNAs (bp) |
|---------------------------|---------------|------------------|---------------------|
| *M. pneumoniae* (+ strand)| 309           | 210–236”         | 586–652”            |
| *M. pneumoniae* (− strand)| 317           | 219–249”         | 535–705”            |
| *E. coli* (+ strand; active) | 220           | 97               | 1371               |

The number of TSSs identified in *M. pneumoniae*, and a subset of high confidence sites in the plus strand of *E. coli* as defined in the Regulon database* (Supplementary Table S2), are listed. *The number of *E. coli* promoters found to be active in the ultrasequencing analysis under the conditions tested is indicated. TSS tssRNAs indicates the number of TSSs for which an associated tssRNA was automatically detected. Total tssRNAs indicates the total number of tssRNAs that were automatically detected. (*) Values represent the range of two independent experiments.

© 2012 EMBO and Macmillan Publishers Limited Molecular Systems Biology 2012 3
the level of other (background) small RNAs was reduced (Figure 2C; Supplementary Figure S3), consistent with the view that tssRNAs are primary transcripts and not the result of endonucleolytic cleavage. However, this experiment cannot distinguish from other possible mechanisms that produce such an 5' end, like specific endonucleolytic cleavage near the 5' end, or 3'-to-5' RNase activity with some degree of protection of the first 40–50 bases. The fact that tssRNAs are observed to be isolated entities, without a corresponding long transcript (Supplementary Figure S1), would exclude that they are generated by degradation, since the RNAs we observed could not arise from a longer RNA. Even if secondary structure of the 5' untranslated region (UTR) could explain their resistance to degradation in some cases, it is very unlikely that this could apply to all the detected tssRNAs. In fact, we did not observe any particular enrichment in three-dimensional structures at the 5' of M. pneumoniae genes (Supplementary Figure S14). In sum, this supports the idea that they are primary transcripts resulting from transcription and not the result of degradation.

To further determine the length of tssRNAs, we analyzed the following various experimental results from M. pneumoniae: (i) standard electrophoresis of total RNA gave an apparent size of 35–40 bases (Supplementary Figure S3); (ii) Northern blot of tssRNA promoters driving YFP (see below) were within similar range (Supplementary Figure S12); (iii) tiling array showed a size of around 44 bases (Figure 2D); (iv) deep sequencing of RNA fractionated in various size ranges (<65, 65–100, 100–150, 150–200, and >200 bases) showed a consistent enrichment of tssRNAs in the size range 15–65 bases (Supplementary Figure S6E); and (v) using the TrueSeq kit from Illumina, an improved DSSS method that does not involve any size selection step (see Supplementary information), we verified that the tssRNAs were in the size range of 35–55 bases (Figure 2B). All methods offer a congruent view of the tssRNAs ranging in length from 35 to 55 bases.

In E. coli, we observed a similar pattern, but with slightly smaller sizes for the tssRNAs (33–40 bases; Supplementary Figure S8). Thus, the bacterial tssRNAs are clearly larger in size than the described abortive transcripts found in vitro and in vivo (which range from 6 to 17 bases; see Goldman et al., 2009). Moreover, in-vitro transcription (IVT) performed with M. pneumoniae genomic DNA resulted in a pattern similar to that observed in vivo (Supplementary Figure S7B) but did not reveal the presence of any tssRNA.
recognition sites that appear randomly in the genome. be that these are synthesized from ‘cryptic’ promoters, that is, with the start of a full-length transcript. One explanation could be that tssRNAs at other genomic positions that are not associated with tssRNA (Table I). However, we also found a large number of tssRNAs within 50 bases from the ATG of a gene inside an operon. Intergenic indicates tssRNAs located between open reading frames, and intragenic peaks are inside open reading frames but >50 bases away from the initial ATG. AverageGenome reflects the average values over the whole chromosome.

(Supplementary Figure S7A). In sum, these results indicate that tssRNA are distinct from abortive transcripts, and that tssRNA synthesis requires the endogenous RNA polymerase machinery.

The majority of TSSs in M. pneumoniae have an associated tssRNA (Table I). However, we also found a large number of tssRNAs at other genomic positions that are not associated with the start of a full-length transcript. One explanation could be that these are synthesized from ‘cryptic’ promoters, that is, promoter-like sequences that appear randomly in the genome. We thus scored the quality of putative RNA polymerase recognition sites -- 10 regions (Pribnow boxes), which are the most conserved regions in M. pneumoniae promoters (Guell et al., 2009) along the chromosome (‘Pribnow_score’; Supplementary Methods; Table II). The score was based on an analysis of the sequences upstream of the manually annotated TSSs (as determined by transcriptome analysis; Guell et al., 2009) (Supplementary Table S1), or by 5’ sequencing (Weiner et al., 2000) (see Supplementary Methods). Our analysis showed that all tssRNA upstream regions have a better Pribnow score than the average value of a random sequence (taking the whole M. pneumoniae genome into account), meaning that they have promoter-like features. Analyzing the 25 bases upstream of the tssRNA start sites for the best-scored Pribnow boxes showed that they are located at the right distance of around 10 bases upstream (according to the previously determined distance of 9 ± 3 bases; see Shultzberger et al., 2007) (Table II). Moreover, these regions have classic Pribnow sequences (of TANAAT, where N can be any base; Supplementary Figure S9), indicating they are true cryptic promoters. Consistent with this, we found RNA polymerase to be bound to them (see below).

Of the tssRNAs that did not map to the TSS of a long transcript, 34% are close to a translation start codon, about 21% are intragenic, and 48% are intergenic (in stationary phase; Table II). We analyzed the upstream promoter-like sequences to determine whether intragenic tssRNAs can be considered to be background and can thus be used to distinguish the true positives. However, all three sets had a good Pribnow score (although worse than that of the TSS-associated tssRNAs) at the right distance to the TSS (Table II, Supplementary Figure S9), and thus all could represent true TSSs. We additionally observed a positive correlation between the Pribnow score and the expression level of the tssRNA (Supplementary Figure S10). These results suggest that tssRNAs found at intergenic and intragenic regions reflect true TSSs from promoters that are likely to originate from random sequences, or from promoters that are activated under specific conditions. Considering the base composition of M. pneumoniae, we estimated a probability of having 1562 Pribnow boxes (TANAAT) in the genome, a figure that is around 30% larger than the actual one (of 1131 TANAAT sequences in the genome).

To demonstrate that tssRNAs not associated with long transcripts are the result of spurious transcription, we made three M. pneumoniae tssRNA promoter constructs that had a good Pribnow score and supported a high level of expression (Supplementary Figure S12A–C) but that did not produce a corresponding full-length transcript (Supplementary Table S5). These promoters were fused to the yellow fluorescent protein (YFP-Venus) gene. We did not detect any YFP expression from any of the constructs (as shown for two cases; Figure 3A), even when they were trimmed to a minimum length (i.e., they were ‘leaderless’, which improves the signal for a synthetic promoter) or when a ribosomal recognition sequence (Shine-Dalgarno) was added (Figure 3A). Adding the first 20 bases of the tssRNA did not influence the expression levels from either of the two tssRNAs promoters. We confirmed by Northern blot that these promoters did not yield full-length transcripts but rather only tssRNAs (Figure 3B; Supplementary Figure S12D). On the other hand, promoters that produce mRNAs and associated tssRNAs (Figure 3A), or even rRNAs, produced detectable Venus expression and long transcripts, as well as tssRNAs (Figure 3B; Supplementary Figure S12). These results suggest that although a good promoter will support RNA polymerase recruitment and tssRNA production, DNA features other than the Pribnow box are needed to produce full-length RNAs from productive transcription.

So far, we have confirmed the existence of native tssRNAs that are associated with full-length transcripts in both E. coli and M. pneumoniae. However, it is still unclear whether these are co-regulated by the same promoter sequences and thus expressed to the same extent, or whether the tssRNAs could be independently expressed. In M. pneumoniae, tssRNA expression levels correlate weakly with that of the corresponding mRNA (R = 0.54; see Supplementary Figure S13). However, when comparing the expression levels of tssRNAs with those of the cognate full-length transcripts in M. pneumoniae in both exponential and stationary phases, we observed an important relative increase of tssRNAs expression only in the stationary phase (P = 3.52 × 10⁻³², two-sample t-test), when transcription is known to be repressed (Table I; Supplementary Figure S15 and Table S6). This could indicate that tssRNA production is driven independently from its associated full-length RNA, and/or that it depends on other protein factors that determine transcription. To test for these possibilities, we first performed chromatin immunoprecipitation analyses of the two RNA polymerase subunits, α and β, in M. pneumoniae (MPN191 [RpoA and MPN515 [RpoB, respectively], followed by DNA ultrasequencing (ChIP-seq) and DNase protection
assays. These results revealed that the RNA polymerase indeed binds to both the productive (i.e., associated with long associated transcripts) and unproductive (isolated) tssRNAs (Supplementary Figure S16). The RNA polymerase was found to be located both at the promoter region (~10), a position at which it is known to stall and produce abortive transcripts prior to initiation of transcription elongation (Goldman et al., 2009), and at some nucleotides downstream of the TSS, where it could produce tssRNAs (around the +30 position; Figure 3C).

Positioning at downstream regions is more prominent in unproductive, isolated tssRNAs (Supplementary Figure S17), despite the fact that the overall affinity of the RNA polymerase is generally lower at these promoters, which on average have slightly worse Pribnow scores (Table II). This would indicate that RNA polymerase pausing is more likely to occur in non-productive promoters. Altogether, these results suggest that, once elongation has started, RNA polymerase pausing could be a mechanism to avoid spurious transcription at any place where a Pribnow box sequence is present. tssRNAs could thus represent a transcriptional checkpoint to ensure that the RNA polymerase machinery is correctly assembled (e.g., that the sigma factor is lost and the correct elongation factors are recruited) (Roberts et al., 2008; Yang and Lewis, 2010; Burmann and Rosch, 2011). This would further guarantee that there is no unnecessary transcription, avoiding the energy expense and preventing unwanted products, such as truncated proteins or transcripts that are antisense to essential genes (Supplementary Figure S18).

Conclusions

We identified and validated a putative new and distinct class of bacterial RNAs that are associated with TSSs, which we have termed tssRNAs. These have an average size of ~45 bases and exhibit dynamic behavior not necessarily concomitant with that of the cognate gene. The absence of tssRNA synthesis in vitro indicates that their expression requires additional native factors that would ensure the accurate elongation/termination of transcription (Nudler and Gottesman, 2002). While the results of our experiments indicate that tssRNAs could be primary transcripts, based on their 5′ triphosphate, it is still possible that other mechanisms could produce the 5′ ends, such as specific endonucleolytic cleavage near the 5′ end or 3′-to-5′ RNase activity with some degree of protection of the first 40–50 bases. However, we consider this to be unlikely, since: (i) almost all mRNAs are associated with tssRNAs; (ii) isolated tssRNAs are not associated with longer RNAs; (iii) promoters of isolated tssRNAs fused to the Venus protein did not produce full-length RNAs, while mRNA promoters did; (iv) tssRNAs overlap with pausing sites, as shown by RNA polymerase ChIP-seq. We hypothesize that the incorrect assembly of processive RNA polymerase complexes could lead to premature termination of RNA transcripts, which in turn could result in deleteriously truncated proteins. tssRNAs could be part of a regulatory mechanism that prevents transcription from starting before the correct RNA polymerase complex is assembled. It is still unclear which additional factors are involved in this process, or which sequences determine that a promoter will ensure productive transcription. In addition, it cannot be ruled out that tssRNAs have a role on their own. We expect that the data presented here will inspire future studies to address these questions. One practical and immediate application of tssRNAs is in high-throughput studies, where tssRNAs (identified by sequencing the small RNA fraction) could be analyzed in combination with transcriptomic data to identify promoters in bacterial species.

Materials and methods

*M. pneumoniae* RNA (in the size groups of ~15–65, 65–100, 100–150, 150–200, and over 200 bases) and *E. coli* small RNA (~15–65 bases) from cells in exponential and stationary phases were subjected to direct strand-specific sequencing as previously described (Vivancos et al., 2010). tssRNAs were identified with an algorithm that takes into consideration their particular shape and context. High-resolution tiling arrays and Pribnow score calculations were performed as previously described (Gueli et al., 2009). IVT; chromatin immunoprecipitation, and northern and western blotting are described in the Supplementary Methods. Error intervals represent the standard deviation.

Sequencing and tiling array data have deposited in the NCBI Short Read Archive (http://www.ncbi.nlm.nih.gov/sra) and Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) as data sets SRA051821 and GSEI4019, respectively.

Supplementary information

Supplementary information is available at the Molecular Systems Biology website (www.nature.com/msb).

Acknowledgements

This work was supported by an European Research council (ERC) advanced grant, the Fundación Marcelino Botín, the Spanish Ministry of Research and ICREA (Institució Catalana de Recerca i Estudis Avançats). We thank the CRG ultrasequencing facility (Rebecca Curley and Heinz Himmelbauer) and the EMBL Genomics Core (Tomi Bähr-Ivacevic and Vladimir Benes) for technical advice. We kindly thank Veronica Raker for editorial help.

Author contributions: MG and EY designed the study, carried out the experiments, analyzed the data, prepared the figures, and wrote the manuscript. WH helped in the bioinformatics analysis of the tssRNAs and contributed to the software developed in this work. APV helped in the ultrasequencing data of the tssRNAs. MLS helped with the sample preparation. JD helped with the data processing. LS designed the study, performed the simulations, analyzed the data, discussed the results, and commented on the manuscript. ACG and PB contributed to the study design, discussed the results, and commented on the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Burmann BM, Rosch P (2001) The role of E. coli Nus-factors in transcription regulation and transcription/translation coupling: from structure to mechanism. *Transcription* **2**: 130–134

Chekanova JA, Gregory BD, Reverdatto SV, Chen H, Kumar R, Hooker T, Yazaki J, Li P, Skiba N, Peng Q, Alonso J, Brukhin V, Grossniklaus U, Ecker JR, Belostotsky DA (2007) Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome. *Cell* **131**: 1340–1353
Cserzo M, Turu G, Varnai P, Hunyady L (2010) Relating underrepresented genomic DNA patterns and tRNAs: the rule behind the observation and beyond. Biol Direct 5: 56
Gama-Castro S, Jimenez-Jacinto V, Peralta-Gil M, Santos-Zavaleta A, Penalola-Spinola MI, Contreras-Moreira B, Segura-Salazar J, Muniz-Rascado L, Martinez-Flores I, Salgado H, Bonavides-Martinez C, Abreu-Goodger C, Rodriguez-Penagos C, Miranda-Rios J, Morett E, Merino E, Huerta AM, Trevino-Quintanilla L, Collado-Vides J (2008) RegulonDB (version 6.0): gene regulation model of Escherichia coli K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. Nucleic Acids Res 36(Database issue): D120–D124
Goldman SR, Ebright RH, Nickels BE (2009) Direct detection of abortive RNA transcripts in vivo. Science 324: 927–928
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, Serrano L (2009) Transcriptome complexity in a genome-reduced bacterium. Science 326: 1268–1271
Hsu LM (2002) Promoter clearance and escape in prokaryotes. Biochim Biophys Acta 1577: 191–207
Kapanidis AN, Margeat E, Ho SO, Kortkhonjia E, Weiss S, Ebright RH (2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. Science 314: 1144–1147
Kuhner S, van Noort V, Betts MJ, Leo-Macias A, Batisse C, Rode M, Yamada T, Maier T, Rader 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 et al (2009) Proteome organization in a genome-reduced bacterium. Science 326: 1235–1240
Miura F, Kawaguchi N, Sese J, Toyoda A, Hattori M, Morishita S, Ito T (2006) A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. Proc Natl Acad Sci USA 103: 17846–17851
Nudler E, Gottesman ME (2002) Transcription termination and anti-termination in E. coli. Genes Cells 7: 755–768
Roberts JW, Shankar S, Filter JJ (2008) RNA polymerase elongation factors. Annu Rev Microbiol 62: 211–233
Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermuller J, Reinhardt R, Stadler PF, Vogel J (2010) The primary transcriptome of the major human pathogen Helicobacter pylori. Nature 464: 250–255
Shi Y, Tyson GW, DeLong EF (2009) Metatranscriptomics reveals unique microbial small RNAs in the ocean’s water column. Nature 459: 266–269
Shultzaberger RK, Chen Z, Lewis KA, Schneider TD (2007) Anatomy of Escherichia coli sigma70 promoters. Nucleic Acids Res 35: 771–788
Taft RJ, Glazov EA, Cloonan N, Simons C, Stephen S, Faulkner GJ, Lassmann T, Forrest AR, Grimmond SM, Schroder K, Irvine K, Arakawa T, Nakamura M, Kubosaki A, Hayashida K, Kawazu C, Murata M, Nishiyori H, Fukuda S, Kawai J et al (2009) Tiny RNAs associated with transcription start sites in animals. Nat Genet 41: 572–578
Taft RJ, Hawkins PG, Mattick JS, Morris KV (2011) The relationship between transcription initiation RNAs and CTCF-binding factor (CTCF) localization. Epigenetics Chromatin 4: 13
Vivancos AP, Guell M, Dohm JC, Serrano L, Himmelbauer H (2010) Strand-specific deep sequencing of the transcriptome. Genome Res 20: 989–999
Weiner 3rd J, Herrmann R, Browning GF (2000) Transcription in Mycoplasma pneumoniae. Nucleic Acids Res 28: 4488–4496
Yang X, Lewis PJ (2010) The interaction between bacterial transcription factors and RNA polymerase during the transition from initiation to elongation. Transcription 1: 66–69
Yus E, Maier T, Michalodimitrakis K, van Noort V, Yamada T, Chen WH, Wodke JA, 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 et al (2009) Impact of genome reduction on bacterial metabolism and its regulation. Science 326: 1263–1268

Molecular Systems Biology is an open-access journal published by European Molecular Biology Organization and Nature Publishing Group. This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License.