868–877 *Nucleic Acids Research, 2010, Vol. 38, No. 3* *Published online 18 November 2009*
doi:10.1093/nar/gkp1032

Deep sequencing-based discovery of the *Chlamydia trachomatis* transcriptome

Marco Albrecht[1], Cynthia M. Sharma[2], Richard Reinhardt[3], Jörg Vogel[2] and Thomas Rudel[1,*]

[1]University of Würzburg, Biocentre, Department of Microbiology, Am Hubland, D-97074 Würzburg, [2]Max Planck Institute for Infection Biology, RNA Biology, Charitéplatz 1, D-10117 Berlin and [3]Max Planck Institute for Molecular Genetics, Ihnestrasse 63–73, D-14195 Berlin, Germany

Received October 10, 2009; Accepted October 20, 2009

ABSTRACT

***Chlamydia trachomatis* is an obligate intracellular pathogenic bacterium that has been refractory to genetic manipulations. Although the genomes of several strains have been sequenced, very little information is available on the gene structure of these bacteria. We used deep sequencing to define the transcriptome of purified elementary bodies (EB) and reticulate bodies (RB) of *C. trachomatis* L2b, respectively. Using an RNA-seq approach, we have mapped 363 transcriptional start sites (TSS) of annotated genes. Semi-quantitative analysis of mapped cDNA reads revealed differences in the RNA levels of 84 genes isolated from EB and RB, respectively. We have identified and in part confirmed 42 genome- and 1 plasmid-derived novel non-coding RNAs. The genome encoded non-coding RNA, ctrR0332 was one of the most abundantly and differentially expressed RNA in EB and RB, implying an important role in the developmental cycle of *C. trachomatis*. The detailed map of TSS in a thus far unprecedented resolution as a complement to the genome sequence will help to understand the organization, control and function of genes of this important pathogen.**

INTRODUCTION

Chlamydia trachomatis is the major cause of bacterial sexually transmitted diseases and ocular infections leading to blindness with hundreds of millions of new cases per year (1). The outcome of infection correlates with certain serovars: Serovars A–C invade mucosal epithelia in the ocular tissue which can lead to trachoma. Serovars D–K infect the urogenital tract causing sexually transmitted diseases. The LGV serovars L1, L2 and L3 invade lymph nodes causing the sexually transmitted systemic syndrome LGV (lymphogranuloma venereum). Chlamydiae are obligate intracellular gram-negative bacteria with an innate biphasic developmental cycle (2). The infection starts with the phagocytosis of the metabolically inactive elementary bodies (EB) by the eukaryotic cell (3). EB differentiate to metabolically active reticulate bodies (RB) which replicate in a vacuole inside the host cell. RB re-differentiate to EB, which are then released from the cells to initiate a new cycle of infection.

Since genetic tools to manipulate the genome and methods to culture the bacteria outside the host cell are lacking, genome sequence analysis has been the main approach to get insight into the biology of all *Chlamydiales*. The genomes of representatives of all biovariants of *C. trachomatis* have been sequenced (4–6) and subsequent genome comparison has unveiled important information on their evolution. The genomes of the representatives of all four serovars show an extremely high degree of conservation. They are similar in size (~1.04 Mbp) with variations of only 5000 bp. Moreover, 846 coding sequences (CDS) out of the 889–920 CDS are common to all serovars (5). Chlamydial genomes exhibit a very high coding density of 90% indicating a highly optimized usage of the coding capacity of their genomes. However, CDSs were identified by computational analysis and some if not most of the differences in the CDSs of the genomes are due to the use of different gene prediction algorithms rather than the real gene content (5). Only recently, transcriptome studies using microarrays (7) and reverse transcription PCR (8) have permitted a first evaluation of gene expression patterns in *Chlamydia*. As these studies were based on the predicted CDSs, the real chlamydial transcriptome has not been determined yet.

As a consequence of the high coding density, intergenic regions (IGRs) are rare and small in *C. trachomatis*. This is noteworthy since the IGRs of other bacterial genomes often harbor the genes of small regulatory RNAs (sRNAs)

*To whom correspondence should be addressed. Tel: +49 931 31 84401; Fax: +49 931 888 4402;
Email: thomas.rudel@biozentrum.uni-wuerzburg.de

(9–13). Searches for bacterial sRNAs have been performed mainly by computational predictions combined with experimental verification (10–12). Other approaches successfully identified new sRNAs by way of cDNA cloning of small-sized RNA species (14,15), and detection on tiling arrays (16–19). Yet another method has been the co-precipitation of sRNAs with Hfq, a conserved sRNA-binding protein in bacteria (20), and the subsequent identification of Hfq-associated transcripts on whole genome microarrays (21) or by deep sequencing of cDNA (22) [a.k.a. RNA-seq (23)]. Although an Hfq homolog has not been known in *Chlamydia*, several sRNA have been predicted; however, one sRNA has been experimentally validated and studied (24,25). The sRNA IhtA is involved in the translational regulation of histone-like protein Hc-1.

Recently, the study of *Vibrio cholera, Bacillus anthracis* and *Burkholderia cenocepacia* has demonstrated the feasibility of the unbiased identification of sRNAs by deep sequencing approaches (26–28). Here we have analysed the primary transcriptome of *C. trachomatis* L2b/UCH-1/proctitis by selectively sequencing cDNA libraries enriched for primary transcripts. Besides new open reading frames (ORFs) that were missed in previous genome annotation, we identified and validated several sRNAs. The most abundant RNA discovered here seems specific for *C. trachomatis* spp. Another abundant sRNA is encoded on the pathogenicity-associated cryptic plasmid. Our data demonstrate the power of RNA-seq for direct RNA identification and the discovery of the structure and control of genes in genetically inaccessible organisms such as obligate intracellular bacteria.

MATERIALS AND METHODS

Infection and isolation of bacterial RNA

HeLa229 (ATCC CL-2.1) cells were cultured in DMEM (Invitrogen) containing 10% FBS (Biochrom) and infected with *C. trachomatis* strain L2b with a MOI of 5 for 24 h. *C. trachomatis* containing cells were collected and disrupted with glass beads. *Chlamydia* were isolated by differential centrifugation followed by density gradient centrifugation as described before (29) with several modifications to achieve better separation of EB and RB. Cell debris was pelleted at 1500 rcf for 10 min at 4°C followed by centrifugation of the *Chlamydia* containing supernatant at 30 000 rcf for 30 min. The bacterial pellet was washed twice in sucrose-phosphate-glutamate (SPG) buffer (250 mM sucrose, 3.8 mM KH_2PO_4, 7.2 mM K_2HPO_4, 0.5 mM L-glutamate, pH 7.4) and resuspended in 2 ml SPG using a 26G syringe needle. The *C. trachomatis* suspension was layered on top of a Percoll Plus (GE Healthcare) solution containing 0.25 M sucrose. EB and RB were separated based on their different buoyant densities on an *in situ* forming continuous density gradient by centrifugation at 30 000 rcf for 30 min in a fixed angle rotor. Two distinct bands were visible containing EB (lower) and RB (upper). These were collected and washed twice in SPG buffer. Purity of EB and RB preparations was verified by electron microscopy (Supplementary Figure S5). Pelleted bacteria were resuspended in Trizol (Invitrogen), mechanically disrupted in a homogenizer (FastPrep, MP Biomedicals) and RNA was isolated according to the manufacturers protocol. Contaminating DNA was digested by DNAseI (Fermentas) and absence of DNA was controlled by PCR. RNA quality was determined on a Bioanalyzer (Agilent). Absence of 18S and 28S ribosomal RNA peaks supported the purity of the bacteria preparation.

Preparation of cDNA and sequencing

cDNA cloning and pyrosequencing was performed as described before (30) but omitting size fractionation of RNA prior to cDNA synthesis. Equal amounts of total RNA were used for the generation of all cDNA libraries. Primary transcripts of total RNA were enriched by selective degradation of RNAs containing 5′ mono-phosphate (5′P) by treatment with a 5′ P-dependent exonuclease as it will be described in full detail in the context of a comprehensive primary transcriptome study of *Helicobacter pylori* (Sharma *et al.*, submitted). Primary bacterial transcripts (most mRNAs and sRNAs) are protected from exonucleolytic degradation by their tri-phosphate (5′PPP) RNA ends. For linker ligation RNA was treated with tobacco acid pyrophosphatase to generate 5′-monophosphates. After addition of specific 5′-linkers with unique tags for each library and poly-A-tailing, the RNA was converted into a cDNA library. Four cDNA libraries were generated in total: total RNA and total RNA enriched for primary transcripts from EB and RB, respectively. The libraries were then sequenced on a Roche/454 GS-FLX system using the 'Amplicon A' and the 'SR70 sequencing kits'.

Analysis of sequences and statistics

After clipping of 5′-linker and poly-A-tails all sequences longer than 17 nucleotides (nt) were considered for BLAST search. The sequences were aligned to the *C. trachomatis* L2b/UCH-1/proctitis genome and plasmid (NC_010280 and NC_010285) using WU-BLAST 2.0. For visualization of BLAST hit locations graph files were calculated and loaded into the Integrated Genome Browser (Affymetrix) as previously described (31). For comparative quantification of gene expression, sequence numbers were counted for every ORF including the 5′-UTR if a distinct TSS was present. Transcripts with at least 20 sequences detected in total were considered for expression analysis. Individual sequence numbers were normalized to the total number of mRNA sequences per library. The threshold for a classification as differentially expressed was 2-fold.

Northern blot analysis of sRNAs

To confirm the expression and size of putative transcripts total RNA from *Chlamydia* was separated on a denaturing 10–15% polyacrylamide gel containing 8M urea and then transferred to a nylon membrane followed by covalent cross-linking by UV irradiation. 24-mer DNA oligonucleotides antisense to the putative bacterial RNAs were end labeled with (γ^{32}P)-ATP, hybridized at

45°C in hybridization buffer (Rapid-Hyb, GE Healthcare) and washed with washing buffer (2× SSC, 0.1% SDS). Blots were exposed to phosphor storage screens (Fujifilm) which were then scanned by Typhoon 9200 imager (GE Healthcare).

RESULTS

Sequencing of *C. trachomatis* total RNA

In order to obtain a comprehensive image of the total transcriptome of *C. trachomatis*, we analysed cDNA libraries of total RNA by deep sequencing analysis. To allow detection and quantification of transcripts which differ between EB and RB, these developmentally distinct forms of *C. trachomatis* were purified, RNA was isolated and EB and RB cDNA libraries were generated. To identify primary transcriptional start sites, we sequenced two libraries for each growth form: one generated from the original, untreated total RNA, and the other following enrichment of primary transcripts by selective enzymatic degradation of processed RNA species (see 'Materials and methods' section for details). The resulting four different libraries, derived from total RNA and RNA enriched for primary transcripts of EB and RB, were then subjected to deep sequencing to achieve a semi-quantitative comparison of gene expression. In total, we analyzed 338 678 sequences of the four cDNA libraries. After removal of 5′-linkers and poly-A tail sequences, only reads longer than 17 nt (309 695 reads; 91.44%) were considered for further analysis. Clipped read lengths up to 120 nt with 67.9% longer than 60 nt were obtained with similar read length distribution in all libraries (Figure 1A). We then used the WU-Blast (http://blast.wustl.edu/) algorithm to map 263 949 sequences to either the *C. trachomatis* genome (NC_010280, 94.5% of sequence reads) or the *C. trachomatis* cryptic plasmid (NC_010285, 5.5% of sequence reads). Sequencing read numbers per library ranged from 73 784 to 98 753 and were sorted by classes of RNA as shown in Figure 1B. Plasmid reads represented 8.4% of the RB and 0.9% of the EB libraries, respectively. The fraction of mRNA reads was reduced by the enrichment for primary transcripts since processing and degradation products were removed by the enzymatic enrichment procedure. The RB libraries contained a large fraction of sequences which were neither found on the chlamydial genome nor on the plasmid. Further analyses unveiled, that these sequences were derived from contaminating human host cell mitochondria which apparently co-purified with RB during gradient purification.

Annotation of transcription start sites

The above mentioned treatment of RNA enriches for sequence reads whose 5′-end is a transcriptional start site. To globally map the TSS of the *C. trachomatis* transcriptome, mapped reads along the whole genome were visualized by calculation of graph files displayed by the Integrated Genome Browser (IGB). TSS were manually annotated by inspection of sequenced regions upstream of Open reading frames (ORFs). Sequencing reads starting at exactly the same nucleotide position defined the TSS. This precise mapping of the TSS unveiled a broad variation in the length of the 5′-untranslated regions (UTRs) with a peak at 21–40 nt and a number of 5′-UTRs ranging over 100 nt in length (shown in Supplementary Figure S1). Three transcripts belong to the group of leaderless mRNAs since the TSS equals the translation start (CTLon_464, 600, 684) (32). Further investigation and probably re-annotation is needed for five transcripts where the TSS is located downstream of the annotated translation start (CTLon_0311, 0537, 0742, 0755, 0757). These genes contain an alternative ORF downstream of the TSS which is supported by a putative ribosome binding site. For several genes two distinct TSS could be identified which indicates a differential regulation of gene expression by variation in 5′ UTR length. In total, sequence reads could be detected for 548 out of 934 annotated genes (58.7%). Thereof TSS could be identified for 356 (65.0%) annotated genes located on the bacterial chromosome comprising 317 putative protein coding genes and 39 tRNA and rRNA genes (for details see Supplementary Table S1). For the remaining expressed genes, TSS could not be specified since RNAs were not abundant enough to be identified or expressed as an operon-like transcript. Furthermore, we could identify 16 distinct TSS in intergenic regions which lack an ORF and therefore indicate non-coding transcripts. In addition, we identified 25 TSS representing small non-coding transcripts that are either partially or completely antisense to annotated ORFs (Supplementary Table S2). Furthermore, we detected nine long overlapping antisense transcripts, most of them representing extended 3′ UTRs (Figure S4).

Differences in the EB and RB transcriptome

The *C. trachomatis* L2b/UCH-1/proctitis annotation contains 874 putative ORFs. A comparison of relative read numbers of a given RNA between EB and RB indicated a set of differentially expressed genes in the two chlamydial developmental cycle stages (shown in Tables 1 and Supplementary Table S2). The separate sequencing of EB and RB transcripts revealed 38 genes which are more abundant in EB and could be crucial for host cell invasion and infectivity. This subset includes 18 hypothetical proteins of unknown function. The 46 genes which were found to be upregulated in RB include outer membrane proteins and genes involved in bacterial division. Twelve genes are of unknown function.

Besides this set of overlapping genes, we also identified deviations from previous predictions (Table 1). One interesting example was the most abundant transcript accounting for 20% and 78% of all putative mRNA sequence reads in the RB and EB libraries, respectively. Further analysis of this transcript revealed its localization inside a previously annotated ORF *CTLon_0332* (homolog to *CT081*). The RNA was 97 nt shorter than predicted for the annotated ORF and started at the −19 position from the annotated translation start (Supplementary Figure S2). The sequencing profile indicated processing from a larger transcript rather than a defined TSS. Northern blot analysis

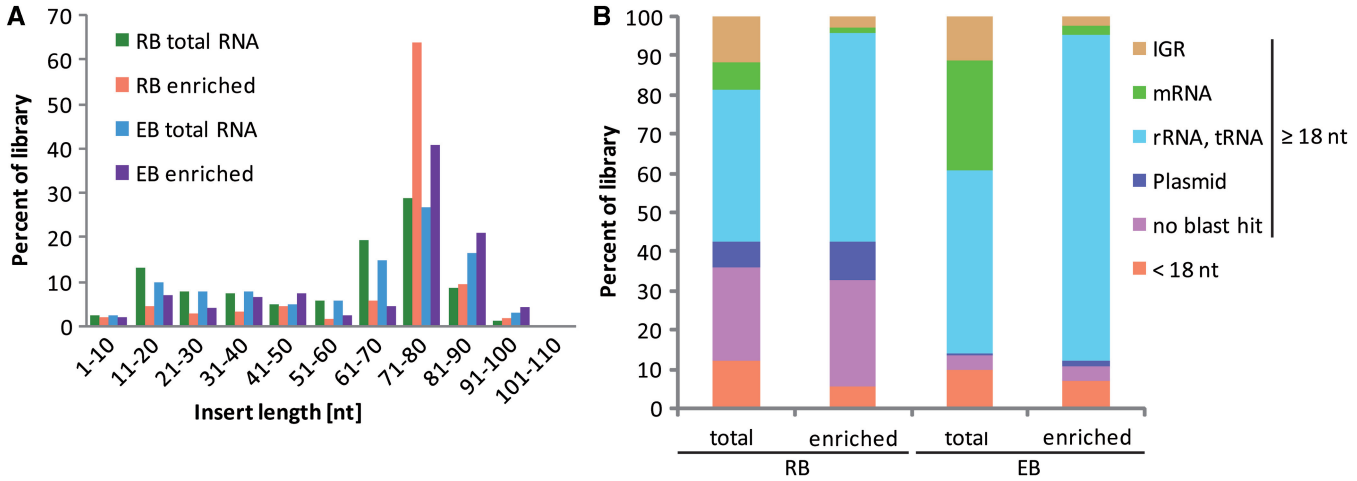

Figure 1. Characterization of the cDNA libraries. (**A**) Length distribution of reads after 5′ end-linker and polyA-tail clipping of four sequenced *C. trachomatis* cDNA libraries generated from either total RNA or total RNA enriched for primary transcripts of reticulate bodies (RB) or elementary bodies (EB), respectively. Shown are relative numbers of groups of sequence length in relation to the total number of reads per library. (**B**) Sequence read distribution of the four cDNA libraries grouped into different classes of RNAs. Transcripts antisense to ribosomal RNAs, transfer RNAs and messenger RNAs are not shown since they represent < 1% of total sequences. Transcripts located in intergenic regions (IGR) can either be sRNA candidates, part of 5′- or 3′-UTRs of mRNAs or unannotated coding genes. The majority of the fraction of reads that could not be mapped to the chlamydial genome corresponds to contaminating host cell RNA, mainly mitochondrial RNAs.

revealed two highly abundant RNA species of ~80 and 240 nt in size (Figure 2, ctrR0332) which fits very well to the 89 and 242 nt calculated from the sequencing data. The major RNA molecule does not contain an ORF. This data suggest that these transcripts encode regulatory RNAs rather than peptides.

Although our RNA sequencing data argued against the existence of ORFs in these RNAs, we screened proteome databases performed with *C. trachomatis* for the presence of protein sequences possibly translated from the region covering these RNAs (33,34). No such proteins could be identified supporting the prediction of a non-coding RNA encoded in region 416 302–416 543.

We next performed BLAST searches for similar sequences in other bacteria. Sequences with significant homology were only present in all sequenced *C. trachomatis* strains (99–100%) and in *C. muridarium* Nigg (89%), a *Chlamydia* species isolated from mice with an average similarity of 90% to *C. trachomatis* orthologous genes (35). These data suggest in summary, that CTLon_0332 represents a highly abundant non-coding RNA specific for human and mouse *C. trachomatis* isolates. We have therefore renamed ORF CTLon_0332 to ctrR0332 for *C. trachomatis* ncRNA0332.

Identification of small RNAs in intergenic regions or antisense to ORFs

Besides reads mapping to the annotated region of the genome we also found a number of transcripts in intergenic regions and transcripts partially or completely antisense to annotated ORFs. Based on manual inspection for such transcripts with a defined primary transcription start site, 16 putative small RNAs located in intergenic regions as well as 25 small antisense RNAs could be detected from the sequencing data. In addition, one abundant processed RNA derived from a tRNA containing precursor could be identified (ctrR1, Figure 2). The sRNA candidates included ihtA, the only previously identified chlamydial sRNA (25).

To test whether these transcripts reflected small RNAs, total RNA from an EB-RB-mixture was analysed by northern blot hybridisation. Expression of 9 out of 12 tested sRNAs was successfully validated by northern blot analysis (Figure 2A), arguing for the presence of multiple sRNAs in *C. trachomatis*. Figure 2B shows the genomic location of these novel sRNAs. Six sRNAs (ctrR1–3, 5–7) are located in intergenic regions. ctrR4 encodes two sRNAs and the larger transcript overlaps the antisense located gene 403 and ctrR8 is located antisense to gene 807. As shown in Supplementary Figure S3 most of these sRNAs show a stable secondary structure when analyzed by the RNAfold algorithm (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi).

An *in silico* search for putative targets regulated by the novel sRNAs was performed using the TargetRNA algorithm (36). The results were filtered for sRNA:target interactions incorporating the translation start codon or a putative ribosome binding site. Candidate targets are listed in Table 2 in the order of their binding score, starting with the highest. Most targets have at least a 7 to 9 nt seed sequence. None of the newly identified sRNA is conserved within different species other than Chlamydiae. However, some motifs of ctrR2, -3, -6, -8 and the plasmid encoded sRNA pL2-sRNA1 are conserved among other species. Targets which are putatively bound by these are underlined in Table 2.

The chlamydial cryptic plasmid encodes a highly abundant sRNA

Virtually all isolates of *C. trachomatis* harbour a cryptic plasmid of 7.5 Kb (37). A total of 14 459 cDNAs mapped to the chlamydial cryptic plasmid which has a size of

Table 1. Transcripts enriched in EB and RB

Gene	Description	Gene no.	Homolog	Microarray (8)
Elementary bodies				
	Pseudogen	CTLon_0548	CT300	–
	Hypothetical protein	CTLon_0186	CT814.1	Late
	Hypothetical protein	CTLon_0332	CT081	Late
scc2	Type III secretion chaperone (low calcium response protein H)	CTLon_0833	CT576	Late
copD	Putative type III secretion system protein	CTLon_0836	CT579	Late
	Hypothetical protein	CTLon_0250	CT875	Late
	Hypothetical protein	CTLon_0333	CT082	Late
omcA	Cysteine-rich outer membrane protein	CTLon_0699	CT444	Late
	Hypothetical protein	CTLon_0880	CT622	Late
	Hypothetical protein	CTLon_0834	CT577	Late
	Hypothetical protein	CTLon_0242	CT867	–
	Hypothetical protein	CTLon_0185	CT814	Late
crpA	Cysteine-rich membrane protein	CTLon_0697	CT442	Late
	Putative integral membrane protein	CTLon_0398	CT147	Very late
	Putative lipoprotein	CTLon_0700	CT444.1	–
	Hypothetical protein	CTLon_0428	CT181	Late
ltuB	Late transcription unit B protein	CTLon_0331	CT080	Late
copB	Putative type III secretion system membrane protein	CTLon_0835	CT578	Late
	Hypothetical protein	CTLon_0609	CT357	–
	Hypothetical protein	CTLon_0255	CT005	Late
	Putative protein ligase	CTLon_0285	CT035	Very late
	Hypothetical protein	CTLon_0608	CT356	Late
omcB	60 kDa cysteine-rich outer membrane protein	CTLon_0698	CT443	Late
	1-Acyl-sn-glycerol-3-phosphate acyltransferase	CTLon_0144	CT775	Late
	Putative integral membrane protein	CTLon_0617	CT365	Late
ltuA	Late transcription unit A protein	CTLon_0629	CT377	Midlate I
	Hypothetical protein	CTLon_0461	CT214	Late
	Putative oxidoreductase	CTLon_0627	CT375	Late
	Hypothetical protein	CTLon_0240	CT865	–
	ABC transporter, ATP-binding protein	CTLon_0427	CT180	–
	Hypothetical protein	CTLon_0334	CT083	Late
mdhC	Malate dehydrogenase	CTLon_0628	CT376	Midlate I
	Pseudogen	CTLon_0610	CT358	–
	Hypothetical protein	CTLon_0536	CT288	Late
aas	Long chain fatty acid-[acyl-carrier-protein] ligase	CTLon_0145	CT776	Late
	Hypothetical protein	CTLon_0477	CT229	–
	Hypothetical protein	CTLon_0028	CT659	Late
rpiA	Ribose-5-phosphate isomerase A	CTLon_0460	CT213	Late
Reticulate bodies				
pmpC	Polymorphic outer membrane protein	CTLon_0667	CT414	Midlate I
nrdA	Ribonucleotide-diphosphate reductase subunit alpha	CTLon_0199	CT827	Midlate I
	Hypothetical protein	CTLon_0267	CT017	Late
	Hypothetical protein	CTLon_0859	CT602	–
cydA	Cytochrome *d* ubiquinol oxidase subunit I	CTLon_0263	CT013	Midlate I
	Putative type III secretion system chaperone	CTLon_0294	CT043	Midlate I
	Putative cation efflux protein	CTLon_0678	CT423	–
	Type III secretion structural protein (outer membrane ring)	CTLon_0043	CT674	Midlate I
sctJ	Type III secretion system protein, membrane component	CTLon_0816	CT559	Midlate I
	Hypothetical protein	CTLon_0251	CT001	Late
tal	Transaldolase B	CTLon_0561	CT313	Midlate I
ftsH	Cell division protein	CTLon_0213	CT841	Late
ompA	Major outer membrane protein	CTLon_0050	CT681	Midlate II
pkn5	Putative serine/threonine-protein kinase (TTSS effector protein)	CTLon_0042	CT673	–
ruvB	Holliday junction DNA helicase B	CTLon_0291	CT040	–
	Putative lipoprotein	CTLon_0501	CT253	–
ihfA	Integration host factor alpha-subunit	CTLon_0515	CT267	Midlate I
	Putative helicase	CTLon_0077	CT708	Midlate I
pmpF	Polymorphic outer membrane protein	CTLon_0245	CT870	Midlate I
	Hypothetical protein	CTLon_0624	CT372	Midlate I
	Hypothetical protein	CTLon_0152	CT783	Midlate II
	Tyrosine-specific transport protein	CTLon_0190	CT818	–
clpC	ATP-dependent Clp protease	CTLon_0534	CT286	Midlate II
	Hypothetical protein	CTLon_0537	CT289	Midlate I
copN	Low calcium response protein E (TTSS effector protein)	CTLon_0340	CT089	Midlate I
pmpG	Polymorphic outer membrane protein	CTLon_0246	CT871	Midlate I
uhpC	Putative sugar phosphate permease	CTLon_0801	CT544	Midlate I
pmpH	Polymorphic outer membrane protein	CTLon_0247	CT872	Midlate I
rplM	50S ribosomal protein L13	CTLon_0376	CT125	–
	Hypothetical protein	CTLon_0876	CT618	Midlate I
pmpE	Polymorphic outer membrane protein	CTLon_0244	CT869	Late
	Hypothetical protein	CTLon_0572	CT324	–
dppD	ABC transport protein, ATPase component	CTLon_0059	CT690	–
pmpB	Polymorphic outer membrane protein	CTLon_0666	CT413	Midlate I
	Hypothetical protein	CTLon_0521	CT273	Midlate I
euo	Hypothetical protein	CTLon_0702	CT446	–
sfhB	Ribosomal large subunit pseudogenuridine synthase D	CTLon_0027	CT658	–
	Hypothetical protein	CTLon_0607	CT355	–
	Putative DNA methyltransferase	CTLon_0733	CT477	–
hct2	Histone-like protein 2	CTLon_0297	CT046	Late
	Hypothetical protein	CTLon_0879	CT621	Midlate I
ndk	Nucleoside diphosphate kinase	CTLon_0757	CT500	Midlate I
pmpI	Polymorphic outer membrane protein	CTLon_0249	CT874	Midlate I
pgsA	CDP-diacylglycerol–glycerol-3-phosphate 3-phosphatidyltransferase	CTLon_0752	CT496	Midlate I
nqrA	Na(+)-translocating NADH-quinone reductase subunit A	CTLon_0002	CT634	Midlate I
	Hypothetical protein	CTLon_0635	CT382.1	–

A semi-quantitaive analysis of differentially expressed genes returned 84 protein coding genes, 38 overrepresented in EB and 46 more abundant in RB. Sequence read numbers were counted as for each ORF including the 5′-UTR if present and normalized by the total number of mRNA transcripts for each library. Genes with a total number of at least 20 sequence reads and a 2-fold regulation were considered as differentially expressed. A comparison to a microarray based gene expression study shows a high correlation of genes abundant in the late infection cycle phase to EB and genes abundant in the midlate phases to RB.

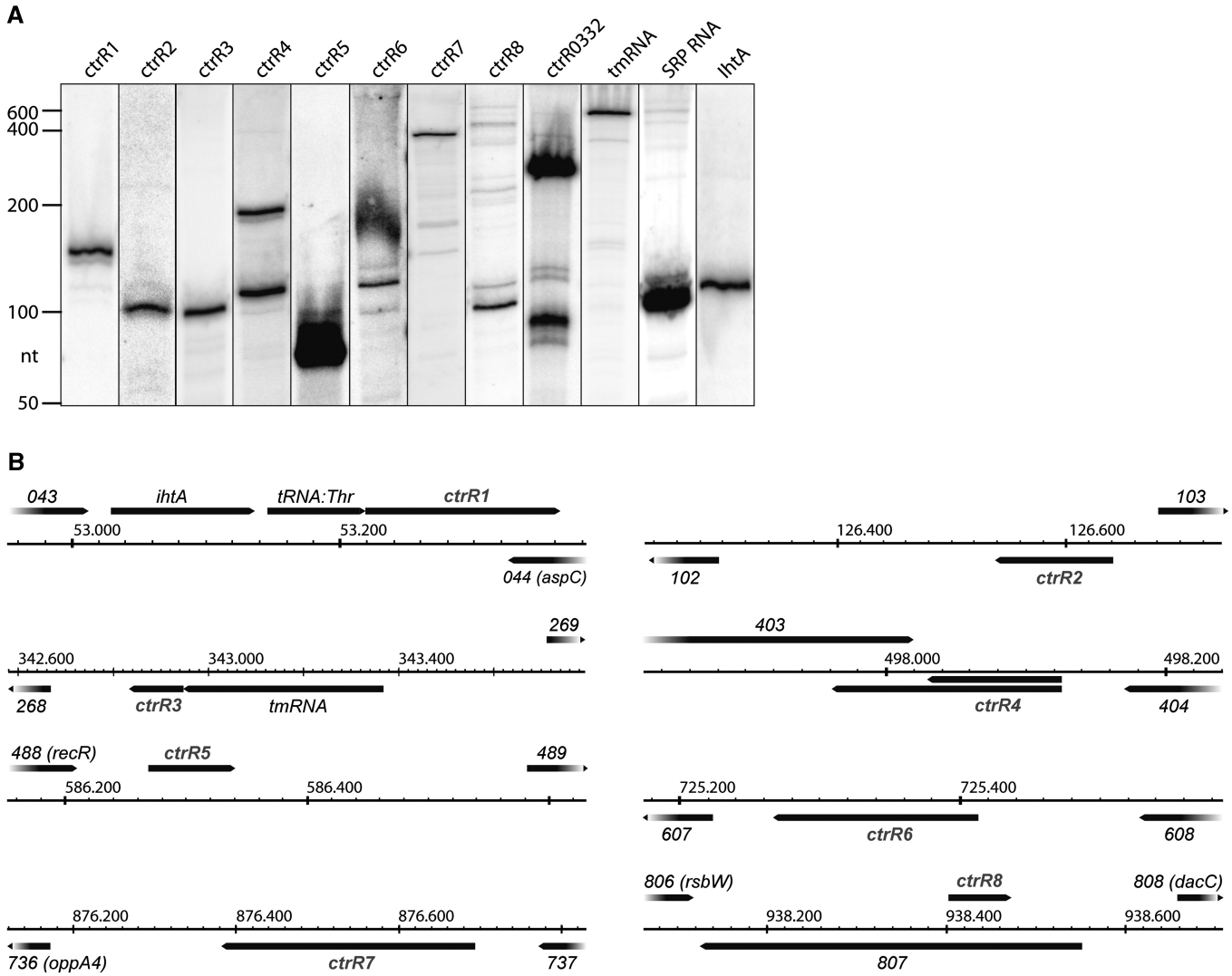

Figure 2. Validation of putative small RNAs by northern analysis. (**A**) Expression of eight out of twelve new candidate sRNAs could be confirmed by northern blotting whereas the length of the probed RNA corresponded to the calculated length from the sequencing data. The genomic location of the sRNAs is shown in (**B**). ctrR1 is not a primary transcript but presumably processed from a larger transcript containing tRNA:Thr. ctrR2, 3, 5–7 are located intergenic and primary transcripts. Two sRNAs are transcribed from ctrR4 and the longer transcript overlaps the gene 403 on the opposite strand. ctrR8 is located antisense to gene 807. *ctrR0332* has previously been identified as ORF CTLon_0332, but represents two RNA species processed from a larger transcript lacking an ORF. Details on the genomic location and the sequence of *ctrR0332* are given in Supplementary Figure S2. Two housekeeping RNAs (tmRNA and SRP RNA) and ihtA is the only so far reported chlamydial sRNA are shown in the right lanes. Genomic locations, sequence read numbers and hybridization probes are given in Supplementary Table S2.

7.5 Kb and contains nine ORFs, eight of which on the minus strand. Figure 3A shows a screenshot of the IGB showing cDNA reads mapped on the plasmid, including the annotated ORFs and regions of high transcriptional activity. We found a highly abundant (>10 000 reads) small transcript pL2-sRNA1 of ~80 nt in length. This sRNA is antisense to a predicted ORF coding for the putative partitioning protein PCTB_7 (pL2-07a, Figure 3A). Moreover, high transcriptional activity can be detected antisense to the only predicted ORF on the plus strand, *pL2-02* (38), encoding virulence plasmid integrase, pGP8-*D*. Northern blot experiments confirmed the expression of several small transcripts in 90–400 nt size range (Figure 3B, probes p4800, p4963, p5120) from this plasmid region which lacks an ORF on the minus strand. The largest RNA species has a size of ~420 nt and is recognized by all probes. The second smallest RNA is ~220 nt long, and was detected by the probes p4963 and p5120. It has the same TSS as the longer RNA and could be derived by 3′ processing of the latter end. The smaller RNA fragments, e.g. the 90 nt fragment detected by p4963 are most likely processing products of larger fragments, since no further TSS could be identified. TSS of the plasmid encoded transcripts are indicated by arrows in Figure 3A including the start coordinates.

We identified TSS for seven of the nine ORFs currently annotated on the plasmid. The remaining two annotated genes are expressed but lack a detectable TSS. The sequencing data suggests that pL2-04 and -05 as well as pL2-07a and -07 are transcribed from a common transcript, respectively. Interestingly, two distinct TSS could be identified for the pL2-02 transcript whereas the

Table 2. Putative targets of chlamydial non-coding RNAs

sRNA	Start	End	Length	RB read number	EB read number	Putative target mRNAs
ihtA	53 029	53 137	109	77	17	hctA, 317, 097, mgtE
ctrR1	53 220	53 363	144	136	227	819, 370, 505
ctrR2	1 26 638	1 26 536	103	1	2	492, tufA, nlpD, 753, 443, dnaA, 035, 730, secG
ctrR3	3 42 947	3 42 838	110	981	1761	248, 794, 190, pfrA, 218, 394, mip, rodA, 294, rbfA, def, ptsN_2, sctU
ctrR4	4 98 126	4 97 963	164	19	3	xerC, pknD, 626
ctrR5	5 86 269	5 86 341	73	8	3	dacC, gnd
ctrR6	7 25 411	7 25 268	144	75	19	824, rplL, 854, kdsA, 751, secG, ppa, 219
ctrR7	8 76 694	8 76 384	311	71	15	copB, 847, glnQ, 531, ssc2, 641, 623, 652, lpdA, ihfA, 597, tig, nusG, aroA
ctrR8	9 38 403	9 38 509	107	59	1	497, murE, eno, 080, 511, 013, 316, dnaQ
ctrR0332	4 16 302	4 16 641	340	1039	16 960	333, atpK, 292, 586, gnd, 822, glgC
pL2-sRNA1	213	292	90	8332	1075	305, sucA, dnaG

Based on binding probability of the newly identified small non-coding RNAs putative binding partners were identified using the TargetRNA algorithm. Hits were filtered for binding of the sRNA to the translation start codon or the ribosome binding site. None of the sRNAs is conserved in other bacteria than Chlamydiae. However some seed sequences are conserved among at least two other bacterial phyla. Targets putatively bound by these conserved sequence elements are underlined. Putative target mRNAs are listed in order of their binding score starting with the highest score.

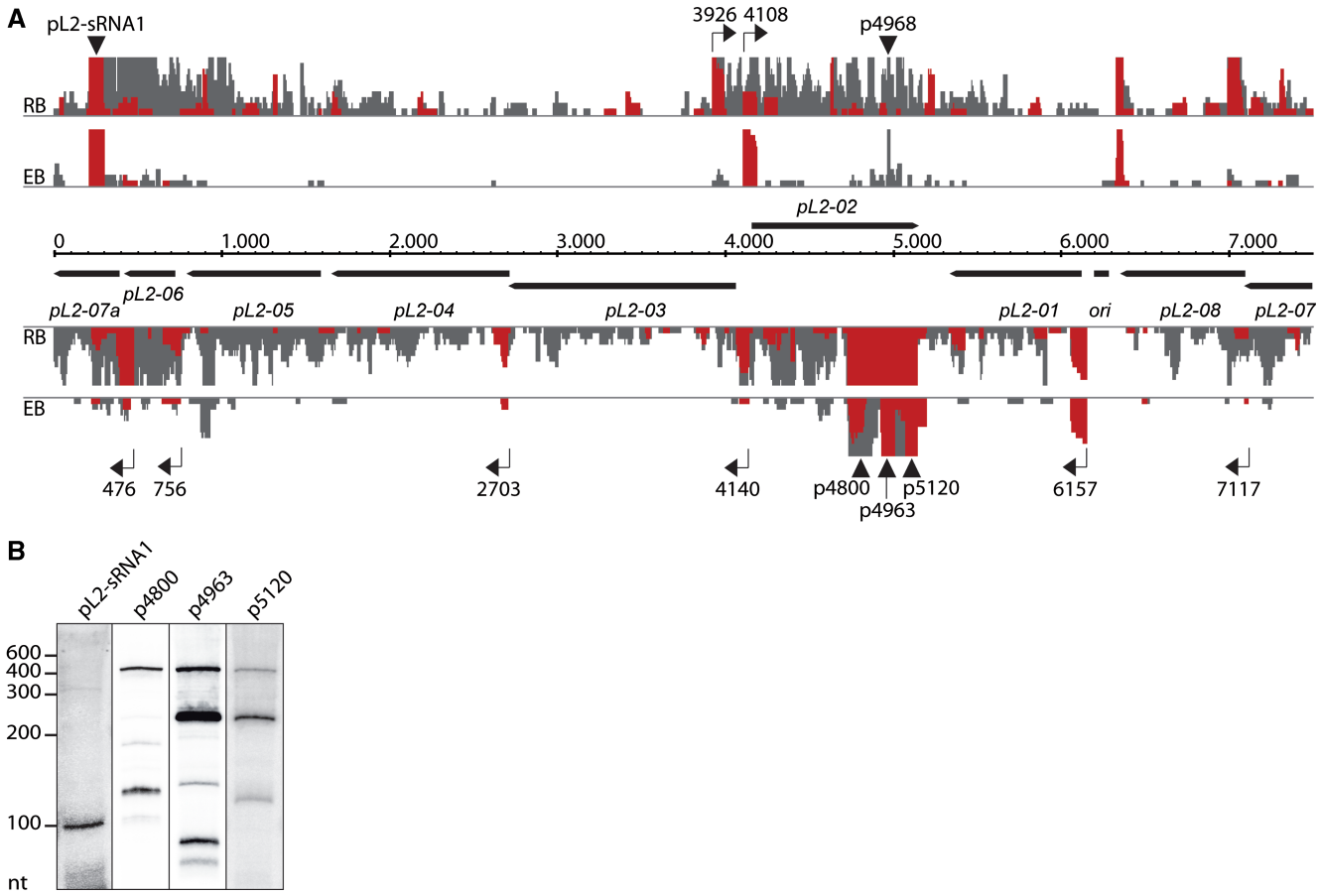

Figure 3. Transcriptome of the cryptic plasmid. (**A**) Calculated sequence graphs for *C. trachomatis* cryptic plasmid. Graphs show the number of sequence reads for every nucleotide of the plasmid up to the cut-off value of 10. Total RNA library reads are shown in grey and are overlaid by red graphs for the TSS enriched libraries. Annotated ORFs are shown as black bars for the plus strand (upper) and minus strand (lower), respectively. Note that total numbers for single transcripts are much higher than 10 but omitted for better visualization. TSS are indicated by horizontal arrows marked with the start positions. Hybridization probe binding sites used for northern detection are marked by vertical arrowheads and the probe name. (**B**) Northern analysis of *C. trachomatis* plasmid transcripts reveals a highly abundant small RNA pL2-sRNA1 of ~100 nt in length antisense to pL2-07a. The region located antisense to the ORF pL2-02 encodes several transcripts in the range of ~80–450 nt. Probe binding sites are marked by an arrow in (A). The names of the transcripts correspond to the first nucleotide of the probe binding site.

transcript starting at −240 nt is enriched in RB and the transcript starting at −58 nt is enriched in EB. This suggests a transcriptional regulation by alternative promoters depending on the developmental cycle phase.

DISCUSSION

The availability of genome sequences for many different bacterial species and their successive annotation has initiated transcriptome analysis as a means to measure gene activity. Using microarray technology, annotated RNAs have been quantified very successfully in various biological conditions. A shortcoming of these probe-dependent approaches is, however, the indirect nature of the RNA quantification which allows only the detection and quantification of previously predicted RNAs with only limited alterations. Moreover, the precise starts of the 5′ ends of most bacterial RNAs cannot be determined by conventional transcriptome analyses but require investigation of individual genes by 5′-RACE or RNA run-off experiments. These limitations are especially relevant for bacteria that cannot be genetically manipulated and grown under laboratory conditions like obligate intracellular bacteria. The gene structure and function of this large group of bacteria including many important human pathogens remains therefore poorly understood.

We have applied RNA-seq to define the TSS of chlamydial genes at the single nucleotide level. Due to the high resolution of the technology, we identified a bona fide non-coding RNA ctrR0332 as a species-specific and most abundant RNA located in an annotated ORF (CTLon_0332) in *C. trachomatis*. In addition, numerous plasmid- and chromosomally encoded new sRNA species could be identified by sequencing and subsequently verified by northern blot analysis.

The definition of 356 TSS in a single sequencing run, most of them to the single nucleotide level, demonstrates the power of the method applied. The enzymatic depletion of RNAs with 5′ monophosphates was an essential step for the resolution of the TSS. Noise caused by RNA break-down products in many cases presumably would have complicated the unequivocal identification of TSS in the sequences obtained from untreated RNA. Such a precise mapping of TSS can now be used to characterize the respective promoters and other gene control elements by computational approaches.

Most of the discovered TSS clearly correlated with annotated ORFs. Others are located antisense and overlap unexpressed annotated ORFs like e.g. CTLon_0175, CTLon_0184 and CTLon_0239. It has been shown previously that these transcripts are expressed and contain ORFs (39). However, numerous annotated transcripts could not be detected, some of them probably due to the limited depth of sequencing or the low expression. On the other hand, we do not expect all annotated 934 genes to have a defined TSS, since *C. trachomatis* genes are frequently organized in operons such as the ten operons coding for a type three secretion system (40). The operon structure can be analysed by comparing total RNA and enriched cDNA libraries. In enriched libraries a distinct TSS is present only for the first gene of an operon, whereas sequence reads are obtained for genes located downstream in the non-enriched libraries. Because of the polycistronic organisation of transcripts, the number of TSS is expected to be lower than the total ORF number. Collectively, the number of 356 identified TSS is likely to approach the real number of native transcript ends in *C. trachomatis*.

Differential transcriptomics of EB and RB, the developmentally different stages of the chlamydial developmental cycle, has been performed before from chromosomal genes but not from plasmid-encoded genes (7,8). These studies used infected cells at different stages of the cycle as a source to prepare RNA. Since the infection course is asynchronous, later time points result in a mixture of EB and RB. To overcome this challenge and to avoid massive background of host RNAs, we generated libraries from purified EB and RB providing proof of principle for successful transcriptome analysis from purified chlamydial particles.

We compared the relative abundances of RNAs from the different chlamydial developmental stages with the published microarray expression data. Nicholson *et al.* (8) classified clusters of temporal gene expression with 'early' (12 h p.i.), 'mid-late' (18 h p.i.), 'late' (24 h p.i.) and 'very late' (36 h p.i.). EB are metabolically inactive, and transcripts present in EB assure immediate translation in early infection events. Such transcripts are produced in the RB stage and stored in the EB, and therefore accumulate during the end of the infection cycle when the RB to EB transition is completed. These genes correlate with 'late' and 'very late' gene expression from microarray studies using RNA from time course experiments (Table 1). Most of the transcripts overrepresented in EB are of unknown function with the interesting exception of the highly overrepresented RNA of an operon consisting of the four genes *CTLon_0833-0836*. This transcript encodes components of the T3SS such as the chaperone scc2, a protein of unknown function, and the ATPase copB and copD, all of which is consistent with the crucial role of the T3SS early in infection (40). In the 'mid-late' stage, the RB is the predominant form of *Chlamydia*. Of the 48 genes we found to be up-regulated in RB, 27 were classified as 'mid-late' and 5 as 'late' genes. Arguably, the use of different platforms and the temporal spectrum of RB-EB-conversion limit a direct comparison of expression data, i.e. obtained by microarrays versus direct transcriptome sequencing of purified bacteria. Nevertheless, our semi-quantitative analysis of relative RNA abundances encoded by chromosomal genes from purified bacterial stages by deep-sequencing correlated very well with the previous microarray analyses of infected cells. This indicates that the method is also suitable for relative quantitative measurements of RNA abundance.

Of all annotated ORFs, the most abundant transcript was that of ORF CTLon_0332. Its transcript quantity is comparable to the class of rRNAs or tRNAs, although it was previously annotated as mRNA. Precise mapping of the TSS and RNA length by northern blot revealed two highly abundant non-coding RNAs. They are shorter than

the annotated transcript, do not contain an ORF and are present only in *C. trachomatis* and the closely related mouse strain *C. muridarium*. It is noteworthy that CTLon_0332 is about tenfold overrepresented in EB which may indicate a role in the EB-RB transition.

The experimental prediction of 16 potential non-coding RNAs in chromosomal IGRs and 25 located antisense within annotated ORFs was another remarkable finding. A total of 9 sRNAs was validated on northern blots, which firmly demonstrates that *C. trachomatis* expresses multiple sRNAs. Since chlamydial genomes are small and highly optimized, intergenic regions are rare. The presence of these sRNAs may indicate a strong evolutionary selection for their expression in *C. trachomatis* due to their putative regulatory functions. The only previously known chlamydial sRNA also identified here was IhtA involved in translational regulation of histone-like protein Hc-1 (25), supporting an important role of sRNAs in regulating gene expression *C. trachomatis*. Most bacteria express the conserved RNA-binding protein Hfq, a pleiotropic regulator involved in the sRNA-mediated control of mRNA stability or translation (20). Chlamydia lack a Hfq homologue but the translational control of Hc-1 expression by the *trans*-encoded sRNA IhtA constitutes a riboregulatory event that is typically facilited by Hfq in *E. coli*, the model bacterium of sRNA research. Whether a protein functionally similar to Hfq controls sRNA function in *C. trachomatis* remains to be shown. Clearly, the sRNAs identified here provide a departure point to discover a general sRNA-binding protein in *Chlamydia* by biochemical methods, e.g. aptamer-facilitated affinity purification (41).

Our results demonstrate that the chlamydial cryptic plasmid also expresses several sRNAs. Some were found in antisense orientation to ORFs implying a role in the regulation of plasmid-coded genes. One of these, pL2-sRNA1 was among the most abundant RNAs identified in *C. trachomatis* with more than 10 000 reads. pL2-sRNA1 codes for an RNA antisense to the gene for the putative partitioning protein PCTB_7, a ParA homolog. ParA proteins are membrane bound ATPases involved in the partitioning of chromosomes and plasmids during replication. It is therefore likely that pL2-sRNA1 is involved in the control of plasmid partitioning. The role of the other plasmid-encoded sRNAs is less clear. They may also be involved in the control of plasmid replication as has been shown for many different replicons (42). Yet another possibility is the regulation of chromosomal gene expression. Plasmid-free strains of *C. trachomatis* down-regulate the activity of chromosomal *glgA* encoding glycogen synthase leading to accumulation of glycogen in the chlamydial inclusion and an attenuated phenotype in mouse genital infection model (43).

Whereas the general and precise identification of TSS will help to better understand the genome organization, the previously unknown highly abundant sRNAs described here raise several questions on their role in the physiology of *C. trachomatis*. Transcriptome analyses by deep-sequencing of other chlamydial species are on the way and will ultimately help to clarify whether species-specific abundant sRNAs are common among other members of the *Chlamydiales*.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Georg Krohne for the preparation of electron micrographs.

FUNDING

Federal Ministry of Education, Science, Research and Technology (BMBF NGFN: 01GS08200 to J.V., R.R. and T.R.); the European Community FP6 IP SIROCCO (Silencing RNAs: Organisers and Coordinators of Complexity in eukaryotic Organisms: LSHG-CT- 2006-037900 to J.V. and T.R.). Funding for open access charge: Federal Ministry of Education, Science, Research and Technology (BMBF NGFN: 01GS08200).

Conflict of interest statement. None declared.

REFERENCES

1. Grayston,J.T. and Wang,S. (1975) New knowledge of chlamydiae and the diseases they cause. *J. Infect. Dis.*, **132**, 87–105.
2. Moulder,J.W. (1982) The relation of basic biology to pathogenic potential in the genus Chlamydia. *Infection*, **10(Suppl. 1)**, S10–S18.
3. Byrne,G.I. and Moulder,J.W. (1978) Parasite-specified phagocytosis of Chlamydia psittaci and Chlamydia trachomatis by L and HeLa cells. *Infect. Immun.*, **19**, 598–606.
4. Stephens,R.S., Kalman,S., Lammel,C., Fan,J., Marathe,R., Aravind,L., Mitchell,W., Olinger,L., Tatusov,R.L., Zhao,Q. *et al.* (1998) Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. *Science*, **282**, 754–759.
5. Thomson,N.R., Holden,M.T., Carder,C., Lennard,N., Lockey,S.J., Marsh,P., Skipp,P., O'Connor,C.D., Goodhead,I., Norbertzcak,H. *et al.* (2008) Chlamydia trachomatis: genome sequence analysis of lymphogranuloma venereum isolates. *Genome Res.*, **18**, 161–171.
6. Carlson,J.H., Porcella,S.F., McClarty,G. and Caldwell,H.D. (2005) Comparative genomic analysis of Chlamydia trachomatis oculotropic and genitotropic strains. *Infect. Immun.*, **73**, 6407–6418.
7. Belland,R.J., Zhong,G., Crane,D.D., Hogan,D., Sturdevant,D., Sharma,J., Beatty,W.L. and Caldwell,H.D. (2003) Genomic transcriptional profiling of the developmental cycle of Chlamydia trachomatis. *Proc. Natl Acad. Sci. USA*, **100**, 8478–8483.
8. Nicholson,T.L., Olinger,L., Chong,K., Schoolnik,G. and Stephens,R.S. (2003) Global stage-specific gene regulation during the developmental cycle of Chlamydia trachomatis. *J. Bacteriol.*, **185**, 3179–3189.
9. Gottesman,S. (2002) Stealth regulation: biological circuits with small RNA switches. *Genes Dev.*, **16**, 2829–2842.
10. Vogel,J. and Sharma,C.M. (2005) How to find small non-coding RNAs in bacteria. *Biol. Chem.*, **386**, 1219–1238.
11. Altuvia,S. (2007) Identification of bacterial small non-coding RNAs: experimental approaches. *Curr. Opin. Microbiol.*, **10**, 257–261.
12. Vogel,J. and Wagner,E.G. (2007) Target identification of small noncoding RNAs in bacteria. *Curr. Opin. Microbiol.*, **10**, 262–270.
13. Waters,L.S. and Storz,G. (2009) Regulatory RNAs in bacteria. *Cell*, **136**, 615–628.
14. Vogel,J., Bartels,V., Tang,T.H., Churakov,G., Slagter-Jager,J.G., Huttenhofer,A. and Wagner,E.G. (2003) RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.*, **31**, 6435–6443.

15. Kawano,M., Reynolds,A.A., Miranda-Rios,J. and Storz,G. (2005) Detection of 5′- and 3′-UTR-derived small RNAs and cis-encoded antisense RNAs in Escherichia coli. *Nucleic Acids Res.*, **33**, 1040–1050.
16. Landt,S.G., Abeliuk,E., McGrath,P.T., Lesley,J.A., McAdams,H.H. and Shapiro,L. (2008) Small non-coding RNAs in Caulobacter crescentus. *Mol. Microbiol.*, **68**, 600–614.
17. McGrath,P.T., Lee,H., Zhang,L., Iniesta,A.A., Hottes,A.K., Tan,M.H., Hillson,N.J., Hu,P., Shapiro,L. and McAdams,H.H. (2007) High-throughput identification of transcription start sites, conserved promoter motifs and predicted regulons. *Nat. Biotechnol.*, **25**, 584–592.
18. Selinger,D.W., Cheung,K.J., Mei,R., Johansson,E.M., Richmond,C.S., Blattner,F.R., Lockhart,D.J. and Church,G.M. (2000) RNA expression analysis using a 30 base pair resolution Escherichia coli genome array. *Nat. Biotechnol.*, **18**, 1262–1268.
19. Toledo-Arana,A., Dussurget,O., Nikitas,G., Sesto,N., Guet-Revillet,H., Balestrino,D., Loh,E., Gripenland,J., Tiensuu,T., Vaitkevicius,K. *et al.* (2009) The Listeria transcriptional landscape from saprophytism to virulence. *Nature*, **459**, 950–956.
20. Valentin-Hansen,P., Eriksen,M. and Udesen,C. (2004) The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Mol. Microbiol.*, **51**, 1525–1533.
21. Zhang,A., Wassarman,K.M., Rosenow,C., Tjaden,B.C., Storz,G. and Gottesman,S. (2003) Global analysis of small RNA and mRNA targets of Hfq. *Mol. Microbiol.*, **50**, 1111–1124.
22. Sittka,A., Sharma,C.M., Rolle,K. and Vogel,J. (2009) Deep sequencing of Salmonella RNA associated with heterologous Hfq proteins in vivo reveals small RNAs as a major target class and identifies RNA processing phenotypes. *RNA Biol.*, **6**, 266–275.
23. Wang,Z., Gerstein,M. and Snyder,M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, **10**, 57–63.
24. Livny,J., Brencic,A., Lory,S. and Waldor,M.K. (2006) Identification of 17 Pseudomonas aeruginosa sRNAs and prediction of sRNA-encoding genes in 10 diverse pathogens using the bioinformatic tool sRNAPredict2. *Nucleic Acids Res.*, **34**, 3484–3493.
25. Grieshaber,N.A., Grieshaber,S.S., Fischer,E.R. and Hackstadt,T. (2006) A small RNA inhibits translation of the histone-like protein Hc1 in Chlamydia trachomatis. *Mol. Microbiol.*, **59**, 541–550.
26. Yoder-Himes,D.R., Chain,P.S., Zhu,Y., Wurtzel,O., Rubin,E.M., Tiedje,J.M. and Sorek,R. (2009) Mapping the Burkholderia cenocepacia niche response via high-throughput sequencing. *Proc. Natl Acad. Sci. USA*, **106**, 3976–3981.
27. Passalacqua,K.D., Varadarajan,A., Ondov,B.D., Okou,D.T., Zwick,M.E. and Bergman,N.H. (2009) Structure and complexity of a bacterial transcriptome. *J. Bacteriol.*, **191**, 3203–3211.
28. Liu,J.M., Livny,J., Lawrence,M.S., Kimball,M.D., Waldor,M.K. and Camilli,A. (2009) Experimental discovery of sRNAs in Vibrio cholerae by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res.*, **37**, e46.
29. Newhall,W.J., Batteiger,B. and Jones,R.B. (1982) Analysis of the human serological response to proteins of Chlamydia trachomatis. *Infect. Immun.*, **38**, 1181–1189.
30. Berezikov,E., Thuemmler,F., van Laake,L.W., Kondova,I., Bontrop,R., Cuppen,E. and Plasterk,R.H. (2006) Diversity of microRNAs in human and chimpanzee brain. *Nat. Genet.*, **38**, 1375–1377.
31. Sittka,A., Lucchini,S., Papenfort,K., Sharma,C.M., Rolle,K., Binnewies,T.T., Hinton,J.C. and Vogel,J. (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet.*, **4**, e1000163.
32. Moll,I., Grill,S., Gualerzi,C.O. and Blasi,U. (2002) Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. *Mol. Microbiol.*, **43**, 239–246.
33. Skipp,P., Robinson,J., O'Connor,C.D. and Clarke,I.N. (2005) Shotgun proteomic analysis of Chlamydia trachomatis. *Proteomics*, **5**, 1558–1573.
34. Shaw,A.C., Gevaert,K., Demol,H., Hoorelbeke,B., Vandekerckhove,J., Larsen,M.R., Roepstorff,P., Holm,A., Christiansen,G. and Birkelund,S. (2002) Comparative proteome analysis of Chlamydia trachomatis serovar A, D and L2. *Proteomics*, **2**, 164–186.
35. Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Utterback,T., Berry,K. *et al.* (2000) Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39. *Nucleic Acids Res.*, **28**, 1397–1406.
36. Tjaden,B., Goodwin,S.S., Opdyke,J.A., Guillier,M., Fu,D.X., Gottesman,S. and Storz,G. (2006) Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Res.*, **34**, 2791–2802.
37. Palmer,L. and Falkow,S. (1986) A common plasmid of Chlamydia trachomatis. *Plasmid*, **16**, 52–62.
38. Ricci,S., Cevenini,R., Cosco,E., Comanducci,M., Ratti,G. and Scarlato,V. (1993) Transcriptional analysis of the Chlamydia trachomatis plasmid pCT identifies temporally regulated transcripts, anti-sense RNA and sigma 70-selected promoters. *Mol. Gen. Genet.*, **237**, 318–326.
39. Jensen,K.T., Petersen,L., Falk,S., Iversen,P., Andersen,P., Theisen,M. and Krogh,A. (2006) Novel overlapping coding sequences in Chlamydia trachomatis. *FEMS Microbiol. Lett.*, **265**, 106–117.
40. Hefty,P.S. and Stephens,R.S. (2007) Chlamydial type III secretion system is encoded on ten operons preceded by sigma 70-like promoter elements. *J. Bacteriol.*, **189**, 198–206.
41. Said,N., Rieder,R., Hurwitz,R., Deckert,J., Urlaub,H. and Vogel,J. (2009) In vivo expression and purification of aptamer-tagged small RNA regulators. *Nucleic Acids Res.*, doi:101093/nar/gkp719 [Epub ahead of print]..
42. Weaver,K.E. (2007) Emerging plasmid-encoded antisense RNA regulated systems. *Curr. Opin. Microbiol.*, **10**, 110–116.
43. Carlson,J.H., Whitmire,W.M., Crane,D.D., Wicke,L., Virtaneva,K., Sturdevant,D.E., Kupko,J.J. III, Porcella,S.F., Martinez-Orengo,N., Heinzen,R.A. *et al.* (2008) The Chlamydia trachomatis Plasmid Is a Transcriptional Regulator of Chromosomal Genes and a Virulence Factor. *Infect. Immun.*, **76**, 2273–2283.