2020–2031 Nucleic Acids Research, 2012, Vol. 40, No. 5 Published online 12 November 2011
doi:10.1093/nar/gkr904

Genome-wide transcriptome analysis of the plant pathogen *Xanthomonas* identifies sRNAs with putative virulence functions

Cornelius Schmidtke[1,*], Sven Findeiß[2,3], Cynthia M. Sharma[4], Juliane Kuhfuß[1], Steve Hoffmann[3,5], Jörg Vogel[4], Peter F. Stadler[2,3,5,6,7,8,9] and Ulla Bonas[1,*]

[1]Department of Genetics, Martin-Luther-Universität Halle-Wittenberg, Institute for Biology, D-06099 Halle, Germany, [2]Institute for Theoretical Chemistry, University of Vienna, A-1090 Vienna, Austria, [3]Department of Computer Science and Interdisciplinary Centre for Bioinformatics, University of Leipzig, D-04107 Leipzig, [4]Institute for Molecular Infection Biology, University of Würzburg, D-97080 Würzburg, [5]LIFE – Leipzig Research Center for Civilization Diseases, University of Leipzig, D-04107 Leipzig, [6]Fraunhofer Institute for Cell Therapy and Immunology, RNomics Group, [7]Max Planck Institute for the Mathematics in Science, D-04103 Leipzig, Germany, [8]Center for non-coding RNA in Technology and Health, University of Copenhagen, DK-1870 Frederiksberg, Denmark and [9]The Santa Fe Institute, Santa Fe, 87501 New Mexico, USA

Received March 31, 2011; Revised and Accepted October 5, 2011

ABSTRACT

The Gram-negative plant-pathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* (*Xcv*) is an important model to elucidate the mechanisms involved in the interaction with the host. To gain insight into the transcriptome of the *Xcv* strain 85–10, we took a differential RNA sequencing (dRNA-seq) approach. Using a novel method to automatically generate comprehensive transcription start site (TSS) maps we report 1421 putative TSSs in the *Xcv* genome. Genes in *Xcv* exhibit a poorly conserved −10 promoter element and no consensus Shine-Dalgarno sequence. Moreover, 14% of all mRNAs are leaderless and 13% of them have unusually long 5′-UTRs. Northern blot analyses confirmed 16 intergenic small RNAs and seven *cis*-encoded antisense RNAs in *Xcv*. Expression of eight intergenic transcripts was controlled by HrpG and HrpX, key regulators of the *Xcv* type III secretion system. More detailed characterization identified sX12 as a small RNA that controls virulence of *Xcv* by affecting the interaction of the pathogen and its host plants. The transcriptional landscape of *Xcv* is unexpectedly complex, featuring abundant antisense transcripts, alternative TSSs and clade-specific small RNAs.

INTRODUCTION

At a staggering pace new high-throughput sequencing technologies have helped to unveil the transcriptional complexity of many organisms in all kingdoms of life (1–3). The recently developed differential RNA sequencing approach (dRNA-seq) has yet added a new perspective. dRNA-seq, based on a selective enrichment of native 5′-ends, has been shown to accurately and cost-effectively identify transcription start sites (TSSs) and RNA processing sites for whole genomes (4). In addition to the obvious advantages for the analysis of 5′-UTR or promoter elements, dRNA-seq allows distinguishing independently transcribed short non-coding and coding RNAs from post-transcriptional processes such as maturation (4). However, a fully-automated method to annotate and statistically evaluate TSSs in large dRNA-seq data sets has been missing so far. Here, we sketch a procedure to automatically identify TSSs.

Transcriptome analyses in plant pathogenic bacteria so far mainly focused on coding regions and the regulon controlling type III secretion [e.g. (5,6)]. A

*To whom correspondence should be addressed. Tel: +345 5526291; Fax: +345 5527277; Email: ulla.bonas@genetik.uni-halle.de
Correspondence may also be addressed to Cornelius Schmidtke. Tel: +345 5526345; Fax: +345 5527277; Email: cornelius.schubert@genetik.uni-halle.de

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

recent deep sequencing analysis of *Pseudomonas syringae* identified many small RNA (sRNA) candidates, most of which, however, await validation by independent methods (7).

The Gram-negative plant pathogenic γ-proteobacterium *Xanthomonas campestris* pv. *vesicatoria* (*Xcv*) is the causal agent of bacterial spot disease on pepper and tomato and is of great economic importance in regions with a warm and humid climate (8). *Xcv* serves as a model system to elucidate the molecular communication between plant pathogens and their hosts and to characterize bacterial virulence strategies. Genome analysis predicted 4726 open reading frames (ORFs) in the *Xcv* strain 85–10 (9), yet the overall gene structure and non-coding RNA output of this model pathogen are still poorly understood.

Essential for pathogenicity of *Xcv* on susceptible host plants is the type III secretion (T3S) system, encoded by the *hrp* [hypersensitive response (HR) and pathogenicity] gene cluster (10). In *Xcv*, as in most Gram-negative bacterial pathogens, the T3S nanomachine translocates a suite of effector proteins into the plant cell where they manipulate host cellular processes to the benefit of the pathogen, e.g. by suppression of basal plant defense responses (9,11–13). *hrp* mutants do not grow in plant tissue, and they no longer cause disease in susceptible plants and the HR in resistant plants (10). The HR is a local, rapid programmed cell death at the site of infection, which coincides with arrest of bacterial multiplication in the plant (14,15).

The T3S system is transcriptionally induced in certain minimal media and in the plant (16,17). Key regulatory proteins are the OmpR-type response regulator HrpG, which is activated by unknown plant signals and controls the expression of a genome-wide regulon including *hrp*, type III effector and putative virulence genes (16–19). HrpG-mediated activation of gene expression depends in most cases on the AraC-type transcriptional activator HrpX (18), which binds to a conserved motif (plant-inducible promoter; PIP box) in the promoters of target genes (20). The identification of a point mutation in HrpG (termed HrpG*), which renders the protein constitutively active, was key for the analysis of T3S and the identification of putative virulence factors that are cotranscribed with the T3S system (19,21). An open question was whether virulence gene expression in *Xcv* is post-transcriptionally regulated, for instance by sRNAs. Here, we provide for the first time an insight into the transcriptional landscape of a plant pathogenic bacterium and the involvement of sRNAs in its virulence.

MATERIALS AND METHODS

RNA isolation for 454 pyrosequencing, RACE analysis and northern blot

RNA was isolated from NYG-grown *Xcv* strains 85–10 and 85* (exponential growth phase) by phenol extraction and treated with DNase I (Roche). For RACE and northern blot analyses, RNA was isolated from NYG-grown *Xcv* strains in exponential and stationary growth phases, as described (22). RACE analyses were carried out as described (23) with modifications [for detailed information see Supporting Information (SI)]. Northern blots were performed as described (24) using 10 µg RNA, 5–10 pmol [γ-^{32}P]-ATP end-labeled oligodeoxynucleotides (Supplementary Table S1). Hybridization signals were visualized with a phosphoimager (FLA-3000 Series, Fuji). Northern blot hybridizations were performed at least twice with independently isolated RNA.

Construction of cDNA libraries for dRNA-seq and 454 pyrosequencing

Prior to RNA treatment and cDNA synthesis, equal amounts of RNA from the two *Xcv* strains 85–10 and 85* were mixed. dRNA-seq libraries were prepared according to Sharma *et al*. (2010) and sequenced with a Roche 454 sequencer using FLX and Titanium chemistry (see SI).

Annotation of transcription start sites

We aimed at the automated identification of TSSs based on the discrimination between narrow clusters of dRNA-seq reads that might represent a TSS and the distribution of individual read starts. The density of read starts varies across the genome and can be modeled locally by a Poisson distribution with a parameter λ. We used fixed-length intervals of size l to determine $\lambda_r = s_r/l$ from the number of read starts s_r in the region r. The parameter λ_{ave} models the average genome wide arrival rate of read starts. λ is defined as λ_r/λ_{ave}. The corresponding Poisson distribution $F(k,\lambda)$ describes the probability that at most k read starts are observed at a given genomic position. We used library 1 to determine λ_m for the background distribution of read starts. Similarly, library 2 was used to obtain λ_p to model the distribution biased towards the TSS.

A TSS is defined as the genomic position at which the observed number of read starts in library 2 significantly exceeds the background distribution of read starts in library 1. The significance of a putative TSS was determined as follows: for each genomic position, the difference of the number of read starts P in library 2 and M in library 1, $D = P–M$, was calculated. The difference of two Poisson distributed variables, D, follows a Skellam distribution (25) whose cumulative distribution function is given by

$$F(D,\lambda_p,\lambda_m) = \sum_{d=-\infty}^{D} e^{-(\lambda_p+\lambda_m)}\left(\frac{\lambda_p}{\lambda_m}\right)^{\frac{d}{2}} J_{|d|}(2\sqrt{\lambda_p\lambda_m});\ d \in Z$$

where $J_{|d|}$ is the modified Bessel function of the first kind and integer order $|d|$. Furthermore, $1 - F(D,\lambda_p,\lambda_m)$ represents the probability that a difference of at least D read starts is observed given the normalized rates of read starts λ_p and λ_m. To reduce the influence of window sizes and local variation of transcriptional activity a sliding window of size x was shifted by y nucleotides along the genome

and each site was tested $t = x/y$ times for being a TSS. The p-value was obtained using the geometric mean

$$p = \sqrt[t]{\prod_{i=1}^{t} p_i}$$

where p_i denotes the P-value obtained in the *i-th* test. Note that only sites with a minimum expression of three read starts within a distance of ≤ 5 nt were tested. Furthermore, we excluded sites in the vicinity of perfectly aligned hit blocks, i.e. stacks of hits that all share a common 5′- and 3′-end. To determine λ_r, we selected a region size of 500 nt. For the sliding window approach an offset of 50 nt was used. All potential TSSs significant to the $p = 0.05$ level are listed in Supplementary Table S2. In order to achieve a high positive predictive value for data sets of similar size, these parameters have been fixed globally in our study and may have to be adjusted for the application of the method to other data sets.

Evaluation of the automated TSS annotation method

To evaluate the predictive power of the automated TSS annotation method we used *Helicobacter pylori* and its manually curated TSS map (4) as reference. A data set of comparable size to the *Xcv* data set was generated. Reads overlapping with annotated tRNA or rRNA genes were excluded. From the *H. pylori* data set 40 385 mapped reads of the treated library and 49 845 reads of the untreated library were randomly selected and contained 392 manually annotated TSSs which were used as reference class. TSSs were predicted using the same parameter settings (500 nt window size, 50 nt offset; 0.05 p-value cutoff) as for the *Xcv* data set. 566 genomic positions met the criteria for being TSS candidates, i.e. the clustering of at least three read starts. These positions represent putative TSSs and were statistically evaluated with the automatic TSS annotation approach, according to (26). The results are summarized in an extended confusion matrix (Supplementary Table S9).

Estimation of expression level

To estimate the expression level of CDSs in *Xcv* likely to exhibit a proximal promoter, we selected 1276 annotated CDSs in a head-to-head arrangement. The set comprised 549 CDSs with and 727 without annotated TSS. Due to the limited sequencing depth of our data set we combined reads of both libraries and evaluated the coverage of the first 100 nt of CDSs (Supplementary Figure S2).

Detailed information about additional methods is provided in SI.

Further supporting information and the raw sequencing data are available at the official institutional website of the University of Leipzig (http://www.bioinf.uni-leipzig.de/publications/supplements/10-035).

RESULTS

Mapping of sequencing reads

To analyze the primary transcriptome of *Xcv*, total RNA of strain 85–10 and its derivative 85* were mixed (SI and Supplementary Table S1). *Xcv* strain 85* carries a chromosomal point mutation in *hrpG* (*hrpG**) leading to expression of the Hrp-regulon. cDNAs were synthesized from total RNA (untreated library; hereafter library 1) and RNA enriched for primary transcripts (treated library; hereafter library 2), respectively (4). dRNA-seq analysis resulted in 160 349 reads for library 1 and 149 596 reads for library 2. A total of 84% of the reads were mapped to the *Xcv* genome using the program segemehl (27). As previously described, *Xcv* contains two identical copies of the 5S, 23S and 16S rRNA clusters, respectively, and 56 tRNA loci (9). A total of 63% of the reads of library 1 and 68% of library 2 reads mapped to these genes although the processed rRNAs and tRNAs were expected to be depleted in library 2. Closer examination revealed that the majority of tRNA-read starts in library 2 correspond to the presumed RNase P processing sites rather than TSSs (Supplementary Figure S1). To verify our observations we analyzed all reads overlapping tRNAs in the *Helicobacter pylori* dRNA-seq data set (4), which supports our findings (Supplementary Figure S1). The abundance of library 2 tRNA reads mapping to putative RNase P processing sites might be due to stable secondary structures formed after RNase P cleavage thus protecting mature tRNAs from exonuclease degradation. We, therefore, discarded the reads mapping to rRNA and tRNA loci and analyzed the remaining 49 845 and 40 385 reads in more detail. While reads of library 1 cover entire genes, the read starts of library 2 are shifted towards the 5′-end of primary transcripts, which permits precise mapping of the TSS of a given gene (Figure 1A, e.g. *XCV0520*), as described (4).

A statistical model to annotate TSSs

Most of the TSS maps published to date are derived from tedious manual inspection of sequencing data (4,24,28) or using *ad hoc* heuristics complemented by manual inspection (29–31). Here, we aimed at the automated identification of TSSs based on well-defined criteria, i.e. to discriminate between potential TSSs and the background distribution of read starts. This background, however, is not uniform across the genome but varies depending on gene expression levels. We therefore modeled read starts by Poisson distributions depending on the expression level in a well-defined genomic neighbourhood. Comparing the two libraries, a TSS is defined as a position where the observed difference of read starts in both libraries significantly exceeds the expected differences of read starts modeled by a Skellam distribution from which p-values are readily derived (see 'Materials and Methods' section).

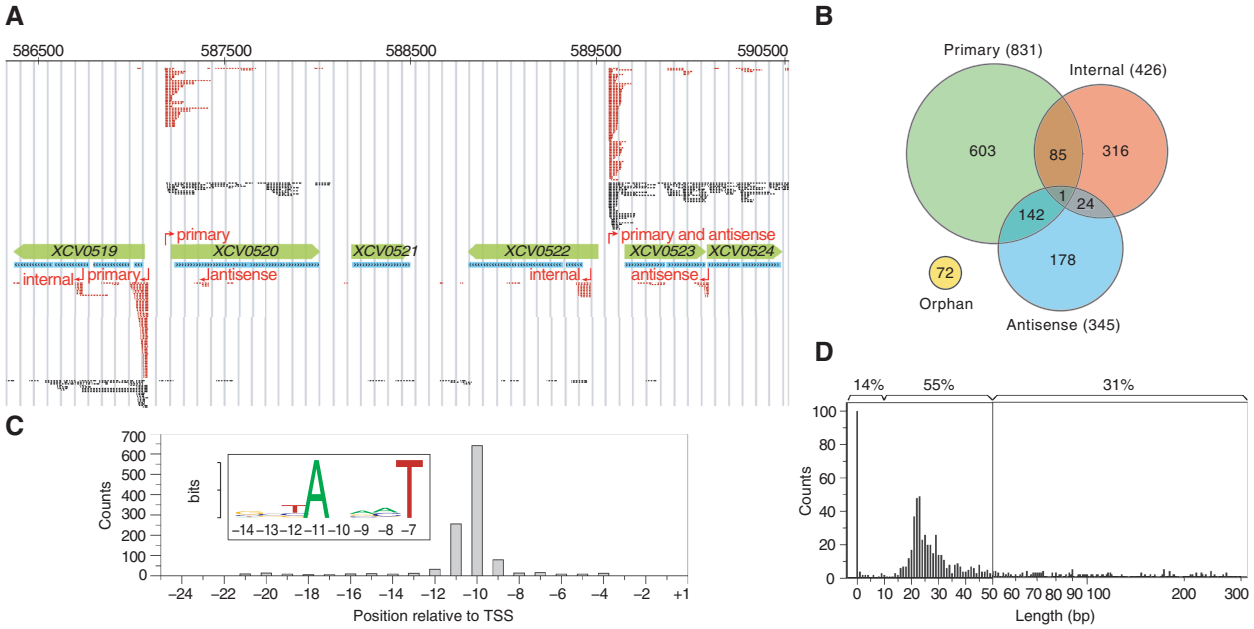

Figure 1. Identification of TSSs, promoter elements and analysis of 5′-UTRs. (**A**) Distribution of dRNA-seq reads in the chromosomal locus of *Xcv* 85–10 spanning genes *XCV0519* to *XCV0524*. Annotated CDSs and RNAcode high-scoring segments are highlighted in green and blue, respectively. Sequencing reads of library 1 (black) and library 2 (red) are shown on top for the (+)-strand and below for the (−)-strand. Predicted TSSs and corresponding classes are indicated in red. (**B**) Venn diagram illustrating the TSS classes. TSSs found maximal 300-bp upstream of coding sequences are classified as primary. Internal TSSs are found within and antisense TSSs on the opposite strand of genes (±100 bp). Orphan TSSs do not belong to other classes. (**C**) Sequence analysis identified a T/A-rich promoter element for 1205 of 1421 putative TSSs. The histogram depicts the position of the conserved sequence pattern relative to the annotated TSSs at position +1. (**D**) 5′-UTR length distribution. The *x*-axis is split into linear (0–50) and logarithmic (51–300) scales. The top of the histogram gives the percentage of leaderless (≤10 bp), short (≤50 bp) and longer UTRs (>50 bp).

Annotation of TSSs

In total, 1372 chromosomal and 49 TSSs on the large plasmid pXCV183 of *Xcv* (Figure 1B and Supplementary Table S2) were identified. The data confirm TSSs determined previously for selected pathogenicity genes, e.g. *hrcU* and *hrpB1* (20,32). Nevertheless, the majority of TSSs annotated in our study should be considered as putative. TSSs were classified into four categories, i.e. (i) primary TSSs located up to 300 bp 5′ of an annotated translation start, (ii) internal TSSs within an annotated coding sequence (CDS), (iii) antisense TSSs that map to the opposite strand of CDSs ± 100 bp and (iv) orphan TSSs that do not belong to the other three categories. Most of the annotated TSSs are primary TSSs (831) and probably correspond to the 5′-end of mRNAs. Overall, CDSs that lack an assigned TSS exhibit much lower expression levels than CDSs with an annotated TSS (see 'Materials and Methods' section and Supplementary Figure S2).

As illustrated in Figure 1B, TSSs can belong to more than one category, e.g. the assumed primary TSS of *XCV0523* is also antisense to *XCV0522* (Figure 1A). Interestingly, 10% (86/831) of primary TSSs are also classified as internal. Thus, some neighboring CDSs previously supposed to be cotranscribed as part of a polycistronic mRNA can also be transcribed from alternative promoters. As illustrated for *XCV0522* (Figure 1A), we identified 71 putative TSSs which are located within the first 50 bp of annotated CDSs suggesting that previously annotated translation starts have to be revisited (Supplementary Table S3). Furthermore, 345 TSSs are located antisense to annotated genes. Interestingly, 41% of these TSSs are also classified as primary TSSs, including 16 TSSs that correspond to overlapping mRNAs in an antisense orientation (Supplementary Table S2). 49 antisense TSSs are positioned in the 3′-region (±100 bp) of annotated sense genes (Supplementary Table S4). In total, antisense reads map to 22% of all nucleotides that belong to annotated CDSs irrespective of read numbers, the presence of a TSS and the expression of the corresponding CDSs. The majority of these antisense reads lack automatically assigned TSSs and do not accumulate in clusters and thus, might not be originated from defined antisense genes. We also compared the sense- and antisense-read coverage of all annotated CDSs in *Xcv* and did not observe a correlation (data not shown).

Most bacterial δ^{70}-dependent promoters contain conserved sequence elements, i.e. −35 (TTGACA) and −10 (TATAAT) elements present in *Escherichia coli* (33). In *Xcv*, there is a weakly conserved T/A-rich motif in the proximity of −10 regions, however, other conserved promoter elements and a Shine-Dalgarno (SD) motif are missing (Figure 1C). This might be due to the high G+C content (65%) of the *Xcv* genome (9) and is discussed below.

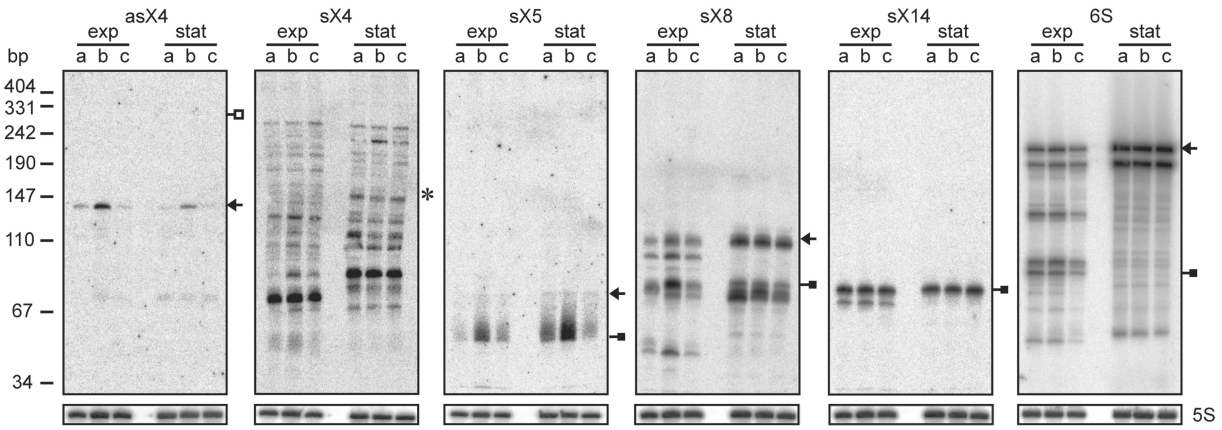

Figure 2. Expression of selected *Xcv* sRNAs and antisense RNAs depends on HrpG and HrpX. Total RNA isolated from exponential (exp) and stationary phase cultures (stat) of (a) *Xcv* strain 85–10, (b) 85–10 expressing *hrpG** from pFG72-1 and (c) 85–10Δ*hrpX* carrying pFG72-1 was analyzed by northern blot. Arrows and filled squares indicate signals corresponding to the expected full-length RNA and processing products obtained by transcriptome sequencing, respectively. The open square indicates the expected size of full-length asX4 determined by RACE analysis. The expected size of sX4 according to the sequencing data is marked by an asterisk. 5S rRNA (lower panel) was probed as loading control.

Analysis of 5′-UTRs revealed unexpected size diversity

The lengths of 5′-UTRs deduced from 831 putative primary TSSs range from 0 to >300 bp, with the majority being between 10 and 50 bp (Figure 1D). Surprisingly, 14% of the mRNAs (118 of 831) are leaderless, i.e. their 5′-UTR consists of <10 bp with respect to the annotated genome sequence of *Xcv* (9). Many of the corresponding genes presumably have housekeeping functions (Supplementary Table S5). In addition, the 5′-UTRs of type III effectors were manually inspected. TSSs of 11 described type III effectors from *Xcv* strain 85–10 (9,13) were mapped in this study (Supplementary Table S5). The promoter regions of nine effector genes contain a PIP box (consensus TTCG-N_{16}-TTCG) (20). The assumed lengths of the 5′-UTRs of *avrBs2*, *xopE2*, *xopJ1* and *xopO* are average. Curiously, the *avrRxv* mRNA is leaderless, and six mRNAs (*avrBs1*, *xopAA*, *xopB*, *xopC*, *xopD* and *xopN*) (9,13) contain unusually long 5′-UTRs, ranging from 173 to 678 bp. Consequently, the CDSs of some effector genes might be considerably larger than previously described (9). Overall, 13% of the *Xcv* 5′-UTRs are unusually long (150–300 bp; Supplementary Table S5).

Northern blot analysis confirmed 23 sRNAs in *Xcv*

A computational scan for known RNA elements in *Xcv* identified already annotated tRNAs, rRNAs and the recently described ptaRNA1 (34). In addition, we identified eight putative riboswitches and widely conserved RNAs, i.e. RNase P-, RtT-, SRP-, tmRNA and 6S-RNA (Figure 2 and Supplementary Table S6). Based on our dRNA-seq data, most of these transcripts were strongly expressed and TSSs were annotated for four of the housekeeping RNAs and five of the predicted riboswitches (Supplementary Table S6). The genes located downstream of the riboswitch candidates are either known to be involved in the respective riboswitch-controlled pathways in other bacteria or, as in case of yybP/ykoY candidates, presumably encode membrane proteins (35–37) (Supplementary Table S6).

Prior to the automated TSS prediction, we selected 89 sRNA candidates by manual inspection of the sequencing data with a focus on intergenic regions. We used northern blot analysis to experimentally validate sRNA candidates and analyzed their potential coregulation with the T3S system. To this end, RNA was isolated from exponential and stationary phase cultures of NYG-grown *Xcv* strains 85–10, 85–10 expressing HrpG* and a derivative lacking *hrpX* (85–10Δ*hrpX*p*hrpG**), respectively. Northern hybridizations confirmed 23 new sRNAs, whereas remaining candidates either appeared to correspond to longer transcripts, i.e. UTRs of mRNAs, or were poorly detectable. The latter can be explained by their low abundance in the dRNA-seq data (data not shown).

After completion of bioinformatic analyses, seven verified sRNAs turned out to correspond to *cis*-encoded antisense RNAs, termed asX1-7 (Table 1, Figure 2 and Supplementary Figure S3). We detected dRNA-seq reads mapping to both antisense RNA and mRNA for six of these transcripts and a few reads mapping to the CDS complementary to *asX4*, respectively (data not shown). The remaining 16 sRNAs mapped to intergenic regions and were termed sX1-15 and 6S (Table 1, Figures 2, 3A, 4A and Supplementary Figure S3). Intriguingly, three sRNAs (sX15, asX6, asX7) are encoded on the large plasmid, two of which (asX7 and sX15) are in antisense orientation to each other (Table 1 and Supplementary Figure S3). Most sRNA genes were constitutively expressed under the conditions tested, and appeared to accumulate in stationary growth phase either due to higher transcription rates or increased stability, e.g. sX14 and 6S (Figure 2). Interestingly, expression/accumulation of five intergenic sRNAs and three antisense RNAs was affected by the key regulators of *hrp* gene expression, HrpG and HrpX, suggesting a role of these sRNAs or their targets in the interaction of *Xcv* with the plant. HrpX-dependent induction of sRNA expression was observed for asX4, sX5, sX8 (Figure 2) and sX12 (see below), whereas sX11 appeared to be HrpG/HrpX-dependently repressed

Table 1. Verified sRNAs (sX) and antisense RNAs (asX) in *Xcv*

RNA (Strand)	TSS category[a]	Start-Stop[b]	Library 2[c]	Library 1[c]	Expected length (nt)[d]	Detected length (nt)[e]	HrpG/HrpX dependency[f]	Conservation[g]
sX1 (−)	primary: *XCV0067*	**78978**–78799	8	5	180	190	–	A (5); B (4); C1 (4); C2 (5); C3 (6); D1–3 (4); E (3)
		78978–78816	8	5	163	170		
		78978–78937	8	5	42	50		
sX2 (+)	orphan	**85087**–85196	11	0	110	110	–	A; B; C1–3; D1–3
		85109–85196	0	5	88	85		
sX3 (+)	orphan	**1233578**–1233669	28	5	92	85	–	A
		1233611–1233669	0	14	59	60		
sX4 (+)	orphan	**1235373**–1235528	5	1	156	280–300	stability	A; B; C1–3
sX5 (−)	–	1899107–1899037	2	0	71	70	HrpX–induced	A; B; C1–3; D1–3; G1–2
		1899085–1899037	35	41	49	50		
sX6 (+)	antisense: *XCV1748*	**1971505**–1971845	15	1	341	350	–	A; B; C1–3; D1–3
sX7 (+)	–	1995660–1995754	1	57	95	85	–	A; B; C1–3; D1–3; E
sX8 (+)	orphan	**2740875**–2740992	7	3	118	110	HrpX–induced	A; F1; H; I; J
		2740875–2740956	7	3	82	85		
sX9 (+)	–	2929816–2929890	2	1	75	75	–	A; B
sX10 (+)	–	3850318–3850496	25	17	179	180	–	A (5); B (4); C1 (4); C2 (5); C3 (6); D1–3 (4); E (3)
		3850415–3850496	14	6	82	80		
sX11 (+)	orphan	**4069950**–4070087	35	10	138	130	HrpG/HrpX–repressed	A; B; C1–3; D1–3
		4070000–4070087	0	5	88	100		
sX12 (+)	–	4358796–4358873	4	2	78	67	HrpX–induced	A; B; C1–3; D1–3
sX13 (−)	–	4810196–4810082	2	73	115	105	–	A; B; C1–3; D1–3; E; G1–2
sX14 (−)	–	5040690–5040599	2	16	92	85	–	A; B; C–3; D1–3
sX15 (+)	antisense: *XCVd0106* and primary: *XCVd0107*	#**116378**–116536	80	8	159	150	–	Ap; G1 (10); G2
		#116438–116536	2	4	99	100		
6S (+)	orphan	**4037865**–4038084	2627	442	220	220	–	A; B; C1–3; D1–3; E; F1–4; G1–2
		4037865–4037952	2627	442	88	90		
asX1 (+)	–	447108–447223	2	0	116	110	stability	A; B; C1–3; D1–3;E; F1–4; G1–2
asX2 (−)	orphan	**3290997**–3290913	12	2	85	75	–	A; B
asX3 (+)	orphan	**4498825**–4499001	16	0	177	70–600	–	A; D1–3
asX4 (−)	antisense: *XCV4106*	**4701994**–4701856	23	7	139	140	HrpX–induced	A; B; C1–3; D1–3
		4701994–4701686	23	7	309	–		
asX5 (+)	orphan	**4757260**–4757445	21	2	186	190	stability	A
		4757360–4757445	1	7	86	85		
asX6 (+)	antisense: *XCVd0099*	#**109851**–109943	35	6	93	90	–	Ap; Bp; Kp; Lp
		#**109851**–109875	35	6	25	30		
asX7 (−)	–	#116528–116459	4	4	70	60	–	Ap; G1 (10); G2

[a]Classification of the automatically annotated TSS (Figure 1B). [b]The 5′- and 3′-positions of the respective dRNA-seq-read clusters on the *Xcv* chromosome and plasmid pXCV183 (indicated by #). Positions highlighted in bold indicate an automatically annotated TSS (see SI; Supplementary Table S2). Underlined numbers correspond to transcript ends which were verified by 5′- and 3′-RACE, respectively. The underlined 3′-end of asX4 was identified only by RACE. [c]Number of read starts at the respective start position given in column 'Start-Stop'. [d]Transcript length deduced from dRNA-seq. [e]transcript size and [f]HrpG/HrpX dependency of sRNA/antisense RNA accumulation determined by northern blot (Figures 2, 3, 4 and Supplementary Figure S3); 'stability' indicates altered amounts of sRNA processing products in dependency of HrpG and/or HrpX. Constitutive expression is indicated by '-'. [g]sequence conservation among other bacteria (see SI). Strains containing homologous sequences and the respective accession numbers are given below. Numbers in brackets indicate the number of homologous sequences in the respective strains if more than one homolog was identified.

A: *X. campestris* pv. *vesicatoria* 85–10 (NC_007508)
Ap: *X. campestris* pv. *vesicatoria* 85–10 plasmid pXCV183 (NC_007507)
B: *X. axonopodis* pv. *citri* 306 (NC_003919)
Bp: *X. axonopodis* pv. *citri* 306 plasmid pXAC64 (NC_003922)
C1: *X. campestris* pv. *campestris* ATCC 33913 (NC_003902)
C2: *X. campestris* pv. *campestris* 8004 (NC_007086)
C3: *X. campestris* pv. *campestris* B100 (NC_010688)
D1: *X. oryzae* pv. *oryzae* MAFF 311018 (NC_007705)
D2: *X. oryzae* pv. *oryzae* KACC 10331 (NC_006834)
D3: *X. oryzae* pv. *oryzae* PXO99A (NC_010717)
E: *X. albilineans* GPE PC73 (NC_013722)
F1: *Xylella fastidiosa* 9a5c (NC_002488)
F2: *Xylella fastidiosa* Temecula1 (NC_004556)
F3: *Xylella fastidiosa* M12 (NC_010513)
F4: *Xylella fastidiosa* M23 (NC_010577)
G1: *Stenotrophomonas maltophilia* K279a (NC_010943)
G2: *S. maltophilia* R551-3 (NC_011071)
H: *Burkholderia xenovorans* LB400 (NC_007951)
I: *Acidovorax* sp. JS42 (NC_008782)
J: *Bordetella petrii* DSM 12804 (NC_010170)
Kp: *Ralstonia solanacearum* CMR15 plasmid pRSC35 (FP885893)
Lp: *X. citri* plasmid pXcB (AY228335)

(Supplementary Figure S3). In case of sX4 (Figure 2) and the antisense RNAs asX1 and asX5 (Supplementary Figure S3) the sRNA stability appeared to depend on HrpG and HrpX as well as on the growth phase.

Processing of sRNAs

In general, the dRNA-seq data and northern blots suggest that *Xcv* sRNAs do not accumulate as primary transcripts but undergo growth-phase dependent processing. However, in most cases the apparent sizes of full-length and processed sRNAs in northern blots were in agreement with the dRNA-seq data, e.g. sX8 and 6S RNA (Figure 2 and Table 1). In addition to full-length and processing products, northern blots detected unexpectedly long signals, up to 900 nt, for the antisense RNAs asX1, asX2, asX3, asX6 and asX7 (Supplementary Figure S3). These signals may be caused by alternative termination of transcription. The sequencing data also suggest that sX7, sX13 and sX14 represent processing products of longer transcripts since reads mapping to these loci are predominantly found in library 1, and no TSS was identified in library 2 (Table 1). For selected RNAs the 5′- and 3′-ends were determined by RACE (Table 1). While the 5′-end of the antisense RNA asX4 is identical to the TSS identified by dRNA-seq, the 3′-region is 170 nt longer suggesting the presence of a processing site.

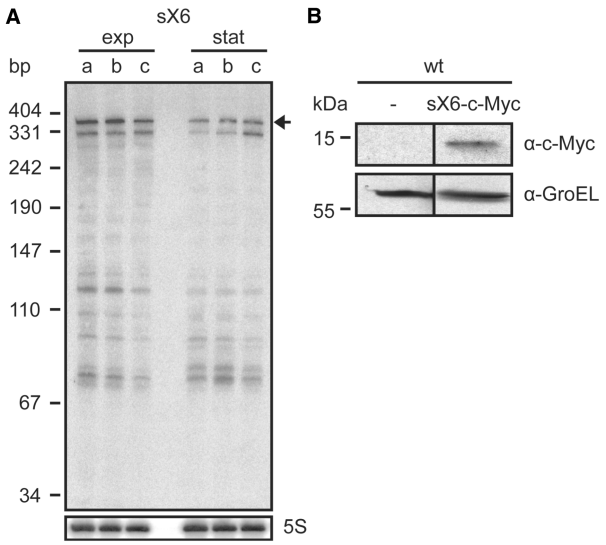

Figure 3. sX6 encodes a small protein. (**A**) Expression analysis of the sX6 transcript. Total RNA isolated from exponential (exp) and stationary phase cultures (stat) of (a) *Xcv* strain 85–10, (b) 85–10 expressing *hrpG** from pFG72-1 and (c) 85–10Δ*hrpX* carrying pFG72-1 was analyzed by northern blot. The expected signal according to sequencing data is indicated by an arrow. 5S rRNA (lower panel) was probed as loading control. (**B**) Expression of the sX6 protein. Derivatives of *Xcv* strain 85–10 (wt) carrying promoterless empty vector pBRM-P (−) and *sX6*-c-Myc expression construct, respectively, were grown to $OD_{600} = 0.7$. Protein extracts were analyzed by immunoblotting using c-Myc epitope-specific and GroEL-specific antibodies.

Phylogenetic distribution of sRNAs from *Xcv*

While *sX3* and *asX5* are unique for *Xcv*, homology searches revealed that 10 sRNA genes are exclusively found in sequenced *Xanthomonas* species that encode a *hrp*-T3S system (Table 1). Four of the latter sRNAs, including sX12 described in more detail below, and asX5 were coregulated with the T3S system.

Two intergenic sRNAs, sX1 and sX10 (Table 1; Supplementary Figure S3) are highly similar in sequence and structure. Three additional homologous genes are predicted and expressed in *Xcv* and might therefore be considered as an sRNA family. As three to six copies of members of this gene family are found in other *Xanthomonas* species (Table 1), we propose a functional redundancy of the respective sRNAs.

Interestingly, 10 homologs of the plasmid-encoded and complementary *Xcv sX15* and *asX7* genes are present in the chromosome of *Stenotrophomonas maltophilia* strain K279a (38) (Table 1). Moreover, *asX6*, which is also

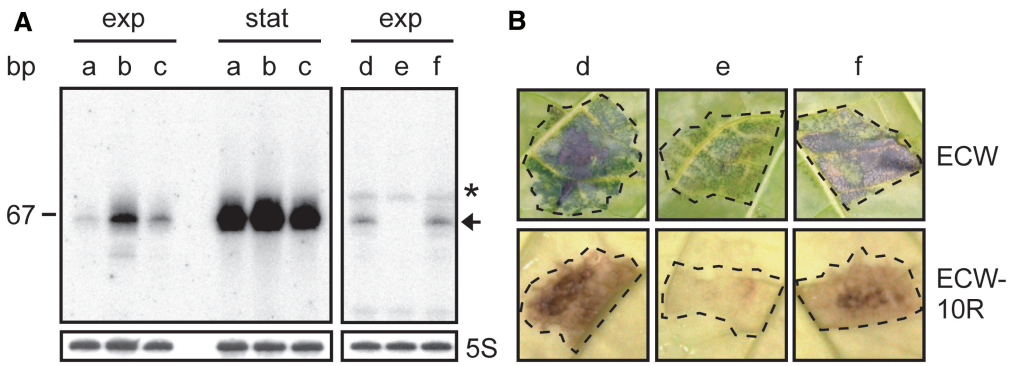

Figure 4. *sX12* is involved in virulence of *Xcv*. (**A**) *sX12* is HrpX-dependently expressed. Total RNA isolated from exponential (exp) and stationary phase cultures (stat) of (a) *Xcv* strain 85–10, (b) *Xcv* expressing *hrpG** from pFG72-1 and (c) a derivative deleted in *hrpX* and carrying pFG72-1 was analyzed by northern blot. The right panel shows a northern blot with RNA from (d) *Xcv* strain 85–10 and (e) an *sX12* deletion mutant carrying empty vector pLAFR6, respectively, and (f) an *sX12* deletion mutant ectopically expressing *sX12* from p*sX12*. The expected RNA size is indicated by an arrow. The asterisk denotes an unspecific signal. 5S rRNA (lower panel) was probed as loading control. (**B**) *sX12* contributes to virulence and the HR. Strains used in (A) (right panel) were inoculated at a density of 1.25×10^8 CFU ml^{-1} into leaves of susceptible ECW and resistant ECW-10R pepper plants. Disease symptoms were photographed at 7 days post-inoculation (dpi). The HR was visualized by ethanol bleaching of the leaves at 2 days post-inoculation. Dashed lines indicate the inoculation site.

located on pXCV183 of *Xcv*, is conserved in plasmids of *X. axonopodis* pv. *citri* strain 306 (39), *Ralstonia solanacearum* strain CMR15 (40) and *X. citri* (41) (Table 1). A rather erratic phylogenetic distribution was observed for *sX8* since homologs are predicted in a small subset of the known genomes of both beta- and gamma-proteobacteria (Table 1). Interestingly, this holds true also for the gene cluster upstream of *sX8* which suggests a common evolutionary origin of this region. This type of phylogenetic pattern has in particular been observed for toxin/anti-toxin systems and suggests frequent horizontal transmissions (42).

sX6 encodes a small protein

Using RNAcode, a program, which was applied for the detection of novel protein coding genes in *E. coli* (43), 24 potential short ORFs were predicted in the *Xcv* genome (see SI and Supplementary Table S7). dRNA-seq reads mapped to 12 of these loci. One example is sX6 (341 nt), which is constitutively expressed (Figure 3A) and has a predicted coding capacity of 80 amino acids including a signal peptide in the N-terminal region. We generated a translational fusion of sX6 with a C-terminal c-Myc epitope tag, under control of the native *sX6* promoter (44), and introduced the expression construct into *Xcv* strain 85–10. As shown in Figure 3B, a fusion protein of the predicted molecular mass (~12 kDa) was detectable in protein extracts of *Xcv*.

Besides *sX6*, TSSs for two of the predicted ORFs with a coding capacity of 36 and 67 amino acids, respectively, were predicted (Supplementary Table S7). Interestingly, homologs of genes for the three small proteins are exclusively found in xanthomonads encoding a *hrp*-T3S system.

sX12 contributes to virulence

The fact that several sRNAs are expressed under control of the T3S system regulators suggested a possible role in virulence. Here, we focused on sX12 whose size of 78 nt was confirmed by 5′- and 3′-RACE (Table 1). As mentioned above, expression of *sX12* is HrpX-dependently induced and accumulates in stationary growth phase (Figure 4A). To assess the contribution of *sX12* to virulence we generated a deletion mutant derivative of strain 85–10 (Δ*sX12*). While growth of strain Δ*sX12 in planta* was as wild-type (Supplementary Figure S4), plant reactions were altered. Disease symptoms in leaves of infected susceptible (ECW) and the HR in resistant (ECW-10R) pepper plants were delayed with strain Δ*sX12* when compared to the wild-type (Figure 4B). The Δ*sX12* mutant phenotype was complemented by ectopic expression of *sX12* under control of its own promoter (Figure 4). We also performed T3S assays to analyze whether the delay in plant reactions by strain Δ*sX12* might be due to reduced protein levels of T3S system components, e.g. the conserved apparatus component HrcJ, or the secretion of T3S substrates, i.e. the translocon protein HrpF. However, the detected protein amounts and the secretion of HrpF were comparable for the wild-type and the Δ*sX12* mutant (Supplementary Figure S4).

DISCUSSION

The dRNA-seq-based analysis of the *Xcv* transcriptome led to remarkable insights into the transcriptional landscape of this important model plant pathogen and identified an sRNA with a role in virulence. In this study, we have devised a new method to automatically generate maps of TSSs for dRNA-seq data sets alleviating the need for manual inspection and allowing application of dRNA-seq also for larger genomes than *Xcv*. In contrast to earlier dRNA-seq approaches, mostly based on laborious manual inspection of sequencing data (4,24,28), the presented computational approach provides a measure of statistical confidence and ensures that predictions are comparable between different studies as demonstrated by our comparative analysis between manual and automated annotation of the previously published *H. pylori* transcriptome (4). While the sensitivity of 82% demonstrates the method's capability of recovering manually annotated TSS at exactly the same position, a positive predictive value of 72% indicates its reliability (Supplementary Table S9). However, to dynamically adjust parameters such as significance levels the method remains subject to further research. We used only exact matches of the manual and automated TSS map for this analysis. The number of false positives and negatives might therefore be overestimated and suffers from biases introduced by manual inspection. Several parameters including window sizes to determine local expression levels, minimum coverage and significance thresholds to control for sensitivity and specificity have been fixed globally for this study.

We annotated 1421 putative TSSs in *Xcv* (Supplementary Table S2) including riboswitches and genes for conserved housekeeping and novel sRNAs. Interestingly, 178 TSSs correspond to antisense transcripts including six that map to type III effector genes and to *hrcC*, which is transcriptionally induced by HrpX and essential for T3S and pathogenicity (Supplementary Tables S2 and S4) (10,45). The potential role of post-transcriptional regulation in *Xcv* is further supported by the finding that 22% of all nucleotides that belong to annotated CDSs are covered by antisense reads. Nevertheless, the majority of these reads might be derived from promiscuous transcription initiation as it was also suggested for *E. coli* (46). It remains to be clarified whether the identified antisense transcripts in *Xcv* represent functional gene products or the transcription itself has a regulatory function.

We identified 831 putative primary TSSs, which were assigned to 17.35% of the 4726 annotated CDSs (Figure 1B and Supplementary Table S2) (9). Similarly, in the archeon *Methanosarcina mazei* TSSs for ~20% of the CDSs were assigned (24). A considerably larger number of TSSs corresponding to 60% of the CDSs were recently mapped in the plant symbiont *Sinorhizobium meliloti* (31) and the human pathogen *H. pylori* (>50%) (4). This might be explained by the plethora of conditions analyzed and/or the higher number of sequencing reads and is supported by our finding that TSSs in *Xcv* are predominantly assigned to

CDSs with high expression level whereas CDSs without assigned TSS are generally weakly or not expressed (Supplementary Figure S2).

In *Xcv*, the majority of 5′-UTRs appears to be <50 bp (Figure 1D), which is characteristic for bacteria (3). Surprisingly, there is no clear consensus sequence for ribosome binding. A recent study (47) analyzed the evolutionary process of translation initiation in prokaryotes and found that a SD-initiated translation in xanthomonads is unlikely. In good agreement with this, we identified an unexpected high number of leaderless mRNAs in *Xcv* (Supplementary Table S5) suggesting an alternative mechanism of ribosome guidance. In *Xcv*, transcription of 82% of the leaderless mRNAs starts with AUG which was shown to be essential for stable ribosome binding to these transcripts in *E. coli* (48).

Unusually long 5′-UTRs as identified here for *Xcv* might be indicative of extensive post-transcriptional regulation, e.g. by sRNA-mediated modulation of mRNA translation or transcript stability. Since also 5′-UTRs of genes that encode type III effector proteins are unusually long (Supplementary Table S5), this might indicate a role of these 5′-UTRs in virulence. For instance, the genes for the type III effector proteins XopN and XopAA, shown to be important virulence factors of *Xcv* (49,50), comprise 5′-UTRs of 173 and 477 bp, respectively (Supplementary Table S5). In *H. pylori*, mRNAs of genes involved in pathogenesis also carry long 5′-UTRs (4).

Another potential implication of the high number of unusually long 5′-UTRs in *Xcv* is that the respective CDSs might be longer than predicted by the genome annotation as shown recently for the type III effector protein XopD (51). On the other hand, a number of CDSs are presumably shorter than annotated, because 71 internal TSSs are located within the first 50 bp of annotated CDSs (Supplementary Table S3). We also identified 12 expressed new loci with potential coding capacity (Supplementary Table S7) exemplified by *sX6* that encodes an 80 amino acid protein (Figure 3). Hence, this study contributes to a first refinement of CDS annotation in *Xcv*.

sRNAs represent important post-transcriptional regulators involved in a variety of processes such as quorum sensing (52) and virulence (53). In this study, the combination of manual and automatic inspection of the cDNA sequencing data and northern blots verified 23 sRNAs in *Xcv*, seven of which represent antisense RNAs (Table 1). For six of the antisense RNAs we also detected expression of the complementary mRNAs. It should be noted, however, that our data do not allow distinguishing between cells that express both transcripts at the same time and cells that either express the mRNA or the antisense RNA.

Notably, expression of five intergenic sRNAs and three antisense RNAs verified in this study was affected by the master regulators of *Xcv* virulence, HrpG and/or HrpX (Table 1) (16,18,19). Coregulation of sRNA expression with the T3S system clearly suggests a role of these transcripts in the interaction of *Xcv* with its host plant. As a proof-of-principle, we have demonstrated that *sX12* contributes to virulence of *Xcv* (Figure 4B). Lack of *sX12* does not affect bacterial growth inside the host and T3S, i.e. bacterial fitness is not impaired (Supplementary Figure S4). What might be the targets of sX12? Preliminary experiments did not reveal an effect of the absence of sX12 on selected *hrp* (T3S) genes, i.e. transcript and protein accumulation was unaltered. Instead of regulating mRNA targets, sX12 might control gene expression in a different manner, e.g. by binding to proteins, DNA or metabolites. Furthermore, *sX12* might impinge on the efficiency of the T3S system, similar to the *Salmonella typhimurium* sRNA IsrJ which accumulates under infection conditions. IsrJ positively contributes to invasion and effector translocation (54).

After our analysis was complete, the identification of eight sRNAs in *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) strain PXO99A was reported (55). In agreement with our data, the *Xoo* sRNAs, *Xoo*3, *Xoo*4 and *Xoo*6, represent orthologs of the *Xcv* RNAs sX14, asX4 and sX1, respectively (Table 1). Contrary to *Xoo*4 (55), which is 145 nt, our analyses revealed that asX4 in *Xcv* is 309-nt long and encoded antisense to an annotated CDS. We also identified potential TSSs for the *Xcv* homologs of *Xoo*1 and *Xoo*5, whereas *Xcv* lacks homologs of described bacterial sRNA genes except for housekeeping RNAs. Vice versa, the majority of sRNAs identified in *Xcv* is restricted to the genera *Xanthomonas*, *Xylella* and *Stenotrophomonas* (Table 1) and thus, reflects the current taxonomy (56). An estimation of the total number of sRNAs in *Xcv* is hampered by the relatively small number of sequence reads and the fact that, for example, TSSs of sRNA genes in the proximity (≤300 bp) of downstream CDSs are classified as primary TSSs (see *sX1*; Table 1).

A remarkable finding of this study is the indication of frequent processing of *Xcv* sRNAs, which appears to be growth-phase dependent. In several studies, sRNA processing was shown to affect sRNA activity, e.g. GlmZ from *E. coli* which is cleaved and thus inactivated (57,58). The *E. coli* sRNA IstR-1 is rendered inactive by RNase III-dependent cleavage upon sRNA-mRNA interaction (59). In contrast, MicX from *Vibrio cholerae* is stabilized by RNaseE-mediated cleavage which does not impair its interaction with target-mRNAs (60). Which ribonucleases are involved in processing of the *Xcv* sRNAs is not known.

The analysis of additional knock-out mutants is needed to assess sRNA functions in *Xcv*. In case of virulence phenotypes, a challenge will be the identification of the targets. Besides possible effects of sRNAs on mRNAs the target can also be an RNA-binding protein. To the best of our knowledge, the only reported sRNAs involved in the regulation of virulence gene expression in plant pathogenic bacteria are members of the RsmB family which was studied in *Erwinia carotovora* ssp. *carotovora*. RsmB antagonizes the RNA-binding protein RsmA that acts as translational repressor (61–63). Although a major virulence function was reported for RsmA from *X. campestris* pv. *campestris* the interacting sRNAs are not known yet (5). The latter is complicated by the lack of CsrB/RsmB sequence homologs in xanthomonads.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supporting Information (SI), Supplementary Figures S1–4, Supplementary Tables S1–9 and Supplementary References [64–76].

ACKNOWLEDGEMENTS

The authors are grateful to B. Rosinsky and C. Kretschmer for technical assistance. The authors thank Richard Reinhardt (MPI for Molecular Genetics, Berlin, Germany) for 454 sequencing and Daniela Büttner for helpful comments on the manuscript.

FUNDING

Deutsche Forschungsgemeinschaft as part of the priority program 'Sensory and Regulatory RNAs in Prokaryotes' (SPP 1258, to U.B., P.F.S. and J.V.); 'Graduiertenkolleg' (GRK 1591, to U.B.); Bundesministerium für Bildung und Forschung ('GenoMik-Plus' network, to U.B.); LIFE Leipzig Research Center for Civilization Diseases, Universität Leipzig; European Social Fund and the Free State of Saxony. Funding for open access charge: Deutsche Forschungsgemeinschaft as part of the priority program 'Sensory and Regulatory RNAs in Prokaryotes' (SPP 1258, to U.B., P.F.S. and J.V.).

Conflict of interest statement. None declared.

REFERENCES

1. van Vliet,A.H. (2010) Next generation sequencing of microbial transcriptomes: challenges and opportunities. *FEMS Microbiol. Lett*., **302**, 1–7.
2. Croucher,N.J. and Thomson,N.R. (2010) Studying bacterial transcriptomes using RNA-seq. *Curr. Opin. Microbiol*., **13**, 619–624.
3. Sorek,R. and Cossart,P. (2010) Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat. Rev. Genet*., **11**, 9–16.
4. Sharma,C.M., Hoffmann,S., Darfeuille,F., Reignier,J., Findeiss,S., Sittka,A., Chabas,S., Reiche,K., Hackermüller,J., Reinhardt,R. *et al*. (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature*, **464**, 250–255.
5. Chao,N.X., Wei,K., Chen,Q., Meng,Q.L., Tang,D.J., He,Y.Q., Lu,G.T., Jiang,B.L., Liang,X.X., Feng,J.X. *et al*. (2008) The *rsmA*-like gene $rsmA_{Xcc}$ of *Xanthomonas campestris* pv. *campestris* is involved in the control of various cellular processes, including pathogenesis. *Mol. Plant-Microbe Interact*., **21**, 411–423.
6. Plener,L., Manfredi,P., Valls,M. and Genin,S. (2010) PrhG, a transcriptional regulator responding to growth conditions, is involved in the control of the type III secretion system regulon in *Ralstonia solanacearum*. *J. Bacteriol*., **192**, 1011–1019.
7. Filiatrault,M.J., Stodghill,P.V., Bronstein,P.A., Moll,S., Lindeberg,M., Grills,G., Schweitzer,P., Wang,W., Schroth,G.P., Luo,S. *et al*. (2010) Transcriptome analysis of *Pseudomonas syringae* identifies new genes, noncoding RNAs, and antisense activity. *J. Bacteriol*., **192**, 2359–2372.
8. Jones,J.B., Stall,R.E. and Bouzar,H. (1998) Diversity among xanthomonads pathogenic on pepper and tomato. *Annu. Rev. Phytopathol*., **36**, 41–58.
9. Thieme,F., Koebnik,R., Bekel,T., Berger,C., Boch,J., Büttner,D., Caldana,C., Gaigalat,L., Goesmann,A., Kay,S. *et al*. (2005) Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. vesicatoria revealed by the complete genome sequence. *J. Bacteriol*., **187**, 7254–7266.
10. Bonas,U., Schulte,R., Fenselau,S., Minsavage,G.V., Staskawicz,B.J. and Stall,R.E. (1991) Isolation of a gene-cluster from *Xanthomonas-campestris* pv. *vesicatoria* that determines pathogenicity and the hypersensitive response on pepper and tomato. *Mol. Plant-Microbe Interact*., **4**, 81–88.
11. Szczesny,R., Büttner,D., Escolar,L., Schulze,S., Seiferth,A. and Bonas,U. (2010) Suppression of the AvrBs1-specific hypersensitive response by the YopJ effector homolog AvrBsT from *Xanthomonas* depends on a SNF1-related kinase. *New Phytol*., **187**, 1058–1074.
12. Büttner,D. and Bonas,U. (2010) Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiol. Rev*., **34**, 107–133.
13. White,F.F., Potnis,N., Jones,J.B. and Koebnik,R. (2009) The type III effectors of *Xanthomonas*. *Mol. Plant Pathol*., **10**, 749–766.
14. White,F.F., Yang,B. and Johnson,L.B. (2000) Prospects for understanding avirulence gene function. *Curr. Opin. Plant Biol*., **3**, 291–298.
15. Klement,Z. (1982) In: Mount,M.S. and Lacy,G.H. (eds), *Phytopathogenic Prokaryotes*, Vol. 2. Academic Press, New York, pp. 149–177.
16. Wengelnik,K. and Bonas,U. (1996) HrpXv, an AraC-type regulator, activates expression of five of the six loci in the *hrp* cluster of *Xanthomonas campestris* pv. vesicatoria. *J. Bacteriol*., **178**, 3462–3469.
17. Schulte,R. and Bonas,U. (1992) Expression of the *Xanthomonas campestris* pv. vesicatoria *hrp* gene cluster, which determines pathogenicity and hypersensitivity on pepper and tomato, is plant inducible. *J. Bacteriol*., **174**, 815–823.
18. Wengelnik,K., Van den Ackerveken,G. and Bonas,U. (1996) HrpG, a key *hrp* regulatory protein of *Xanthomonas campestris* pv. *vesicatoria* is homologous to two-component response regulators. *Mol. Plant-Microbe Interact*., **9**, 704–712.
19. Noel,L., Thieme,F., Nennstiel,D. and Bonas,U. (2001) cDNA-AFLP analysis unravels a genome-wide *hrpG*-regulon in the plant pathogen *Xanthomonas campestris* pv. *vesicatoria*. *Mol. Microbiol*., **41**, 1271–1281.
20. Koebnik,R., Krüger,A., Thieme,F., Urban,A. and Bonas,U. (2006) Specific binding of the *Xanthomonas campestris* pv. vesicatoria AraC-type transcriptional activator HrpX to plant-inducible promoter boxes. *J. Bacteriol*., **188**, 7652–7660.
21. Wengelnik,K., Rossier,O. and Bonas,U. (1999) Mutations in the regulatory gene *hrpG* of *Xanthomonas campestris* pv. vesicatoria result in constitutive expression of all *hrp* genes. *J. Bacteriol*., **181**, 6828–6831.
22. Hartmann,R.K., Bindereif,A., Schön,A. and Westhof,E. (2005) *Handbook of RNA biochemistry*, Vol. 2. Wiley-VCH, Weinheim, Germany, pp. 636–637.
23. Argaman,L., Hershberg,R., Vogel,J., Bejerano,G., Wagner,E.G., Margalit,H. and Altuvia,S. (2001) Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr. Biol*., **11**, 941–950.
24. Jäger,D., Sharma,C.M., Thomsen,J., Ehlers,C., Vogel,J. and Schmitz,R.A. (2009) Deep sequencing analysis of the *Methanosarcina mazei* Gö1 transcriptome in response to nitrogen availability. *Proc. Natl Acad. Sci. USA*, **106**, 21878–21882.
25. Skellam,J.G. (1946) The frequency distribution of the difference between two Poisson variates belonging to different populations. *J. R. Stat. Soc. Ser. A*, **109**, 296.
26. Fawcett,T. (2006) An introduction to ROC analysis. *Pattern Recogn. Lett*., **27**, 861–874.
27. Hoffmann,S., Otto,C., Kurtz,S., Sharma,C.M., Khaitovich,P., Vogel,J., Stadler,P.F. and Hackermüller,J. (2009) Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput. Biol*., **5**, 10.1371/journal.pcbi.1000502.
28. Albrecht,M., Sharma,C.M., Reinhardt,R., Vogel,J. and Rudel,T. (2010) Deep sequencing-based discovery of the *Chlamydia trachomatis* transcriptome. *Nucleic Acids Res*., **38**, 868–877.

29. Wurtzel,O., Sapra,R., Chen,F., Zhu,Y., Simmons,B.A. and Sorek,R. (2010) A single-base resolution map of an archaeal transcriptome. *Genome Res.*, **20**, 133–141.
30. Mitschke,J., Georg,J., Scholz,I., Sharma,C.M., Dienst,D., Bantscheff,J., Voss,B., Steglich,C., Wilde,A., Vogel,J. *et al.* (2011) An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proc. Natl Acad. Sci. USA*, **108**, 2124–2129.
31. Schlüter,J.P., Reinkensmeier,J., Daschkey,S., Evguenieva-Hackenberg,E., Janssen,S., Jänicke,S., Becker,J.D., Giegerich,R. and Becker,A. (2010) A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium *Sinorhizobium meliloti*. *BMC Genomics*, **11**, 1111–1245.
32. Fenselau,S. and Bonas,U. (1995) Sequence and expression analysis of the *hrpB* pathogenicity operon of *Xanthomonas campestris* pv. *vesicatoria* which encodes eight proteins with similarity to components of the Hrp, Ysc, Spa, and Fli secretion systems. *Mol. Plant-Microbe Interact.*, **8**, 845–854.
33. Burgess,R.R. and Anthony,L. (2001) How sigma docks to RNA polymerase and what sigma does. *Curr. Opin. Microbiol.*, **4**, 126–131.
34. Findeiß,S., Schmidtke,C., Stadler,P.F. and Bonas,U. (2010) A novel family of plasmid-transferred anti-sense ncRNAs. *RNA Biol.*, **7**, 120–124.
35. Barrick,J.E., Corbino,K.A., Winkler,W.C., Nahvi,A., Mandal,M., Collins,J., Lee,M., Roth,A., Sudarsan,N., Jona,I. *et al.* (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc. Natl Acad. Sci. USA*, **101**, 6421–6426.
36. Wang,J.X. and Breaker,R.R. (2008) Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine. *Biochem. Cell Biol.*, **86**, 157–168.
37. Winkler,W.C. and Breaker,R.R. (2005) Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol.*, **59**, 487–517.
38. Crossman,L.C., Gould,V.C., Dow,J.M., Vernikos,G.S., Okazaki,A., Sebaihia,M., Saunders,D., Arrowsmith,C., Carver,T., Peters,N. *et al.* (2008) The complete genome, comparative and functional analysis of *Stenotrophomonas maltophilia* reveals an organism heavily shielded by drug resistance determinants. *Genome Biol.*, **9**, R74.
39. da Silva,A.C., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida,N.F., Alves,L.M. *et al.* (2002) Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature*, **417**, 459–463.
40. Remenant,B., Coupat-Goutaland,B., Guidot,A., Cellier,G., Wicker,E., Allen,C., Fegan,M., Pruvost,O., Elbaz,M., Calteau,A. *et al.* (2010) Genomes of three tomato pathogens within the *Ralstonia solanacearum* species complex reveal significant evolutionary divergence. *BMC Genomics*, **11**, 379.
41. El Yacoubi,B., Brunings,A.M., Yuan,Q., Shankar,S. and Gabriel,D.W. (2007) In planta horizontal transfer of a major pathogenicity effector gene. *Appl. Environ. Microbiol.*, **73**, 1612–1621.
42. Makarova,K.S., Wolf,Y.I. and Koonin,E.V. (2009) Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol. Direct*, **4**, 10.1186/1745-6150-1184-1119.
43. Washietl,S., Findeiß,S., Müller,S.A., Kalkhof,S., von Bergen,M., Hofacker,I.L., Stadler,P.F. and Goldman,N. (2011) RNAcode: robust discrimination of coding and noncoding regions in comparative sequence data. *RNA*, **17**, 578–594.
44. Szczesny,R., Jordan,M., Schramm,C., Schulz,S., Cogez,V., Bonas,U. and Büttner,D. (2010) Functional characterization of the Xcs and Xps type II secretion systems from the plant pathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria*. *New Phytol.*, **187**, 983–1002.
45. Wengelnik,K., Marie,C., Russel,M. and Bonas,U. (1996) Expression and localization of HrpA1, a protein of *Xanthomonas campestris* pv. vesicatoria essential for pathogenicity and induction ofthe hypersensitive reaction. *J. Bacteriol.*, **178**, 1061–1069.
46. Dornenburg,J.E., Devita,A.M., Palumbo,M.J. and Wade,J.T. (2010) Widespread antisense transcription in *Escherichia coli*. *MBio*, **46**, e00024–10.
47. Nakagawa,S., Niimura,Y., Miura,K. and Gojobori,T. (2010) Dynamic evolution of translation initiation mechanisms in prokaryotes. *Proc. Natl Acad. Sci. USA*, **107**, 6382–6387.
48. Brock,J.E., Pourshahian,S., Giliberti,J., Limbach,P.A. and Janssen,G.R. (2008) Ribosomes bind leaderless mRNA in *Escherichia coli* through recognition of their 5′-terminal AUG. *RNA*, **14**, 2159–2169.
49. Kim,J.G., Li,X., Roden,J.A., Taylor,K.W., Aakre,C.D., Su,B., Lalonde,S., Kirik,A., Chen,Y., Baranage,G. *et al.* (2009) *Xanthomonas* T3S Effector XopN Suppresses PAMP-Triggered Immunity and Interacts with a Tomato Atypical Receptor-Like Kinase and TFT1. *Plant Cell*, **21**, 1305–1323.
50. Morales,C.Q., Posada,J., Macneale,E., Franklin,D., Rivas,I., Bravo,M., Minsavage,J., Stall,R.E. and Whalen,M.C. (2005) Functional analysis of the early chlorosis factor gene. *Mol. Plant-Microbe Interact.*, **18**, 477–486.
51. Canonne,J., Marino,D., Noel,L.D., Arechaga,I., Pichereaux,C., Rossignol,M., Roby,D. and Rivas,S. (2010) Detection and functional characterization of a 215 amino acid N-terminal extension in the *xanthomonas* type III effector XopD. *PLoS One*, **5**, e15773.
52. Bejerano-Sagie,M. and Xavier,K.B. (2007) The role of small RNAs in quorum sensing. *Curr. Opin. Microbiol.*, **10**, 189–198.
53. Papenfort,K. and Vogel,J. (2010) Regulatory RNA in bacterial pathogens. *Cell Host Microbe*, **8**, 116–127.
54. Padalon-Brauch,G., Hershberg,R., Elgrably-Weiss,M., Baruch,K., Rosenshine,I., Margalit,H. and Altuvia,S. (2008) Small RNAs encoded within genetic islands of *Salmonella typhimurium* show host-induced expression and role in virulence. *Nucleic Acids Res.*, **36**, 1913–1927.
55. Liang,H., Zhao,Y.T., Zhang,J.Q., Wang,X.J., Fang,R.X. and Jia,Y.T. (2011) Identification and functional characterization of small non-coding RNAs in *Xanthomonas oryzae* pathovar *oryzae*. *BMC Genomics*, **12**, 101186/1471-2164-1112-1187.
56. Cutino-Jimenez,A.M., Martins-Pinheiro,M., Lima,W.C., Martin-Tornet,A., Morales,O.G. and Menck,C.F. (2010) Evolutionary placement of Xanthomonadales based on conserved protein signature sequences. *Mol. Phylogenet. Evol.*, **54**, 524–534.
57. Kalamorz,F., Reichenbach,B., Marz,W., Rak,B. and Görke,B. (2007) Feedback control of glucosamine-6-phosphate synthase GlmS expression depends on the small RNA GlmZ and involves the novel protein YhbJ in *Escherichia coli*. *Mol. Microbiol.*, **65**, 1518–1533.
58. Urban,J.H. and Vogel,J. (2008) Two seemingly homologous noncoding RNAs act hierarchically to activate *glmS* mRNA translation. *PLoS Biol.*, **6**, e64.
59. Vogel,J., Argaman,L., Wagner,E.G. and Altuvia,S. (2004) The small RNA IstR inhibits synthesis of an SOS-induced toxic peptide. *Curr. Biol.*, **14**, 2271–2276.
60. Davis,B.M. and Waldor,M.K. (2007) RNase E-dependent processing stabilizes MicX, a *Vibrio cholerae* sRNA. *Mol. Microbiol.*, **65**, 373–385.
61. Cui,Y., Chatterjee,A., Liu,Y., Dumenyo,C.K. and Chatterjee,A.K. (1995) Identification of a global repressor gene, *rsmA*, of *Erwinia carotovora* subsp. carotovora that controls extracellular enzymes, N-(3-oxohexanoyl)-L-homoserine lactone, and pathogenicity in soft-rotting Erwinia spp. *J. Bacteriol.*, **177**, 5108–5115.
62. Liu,Y., Cui,Y., Mukherjee,A. and Chatterjee,A.K. (1998) Characterization of a novel RNA regulator of *Erwinia carotovora* ssp. *carotovora* that controls production of extracellular enzymes and secondary metabolites. *Mol. Microbiol.*, **29**, 219–234.
63. Cui,Y., Chatterjee,A., Yang,H. and Chatterjee,A.K. (2008) Regulatory network controlling extracellular proteins in *Erwinia carotovora* subsp. carotovora: FlhDC, the master regulator of flagellar genes, activates *rsmB* regulatory RNA production by affecting *gacA* and *hexA* (*lrhA*) expression. *J. Bacteriol.*, **190**, 4610–4623.
64. 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.

65. Bailey,T.L. and Elkan,C. (1995) The value of prior knowledge in discovering motifs with MEME. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **3**, 21–29.
66. Blanchette,M., Kent,W.J., Riemer,C., Elnitski,L., Smit,A.F., Roskin,K.M., Baertsch,R., Rosenbloom,K., Clawson,H., Green,E.D. *et al.* (2004) Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.*, **14**, 708–715.
67. Hertel,J., de Jong,D., Marz,M., Rose,D., Tafer,H., Tanzer,A., Schierwater,B. and Stadler,P.F. (2009) Non-coding RNA annotation of the genome of *Trichoplax adhaerens*. *Nucleic Acids Res.*, **37**, 1602–1615.
68. Will,S., Reiche,K., Hofacker,I.L., Stadler,P.F. and Backofen,R. (2007) Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comput. Biol.*, **3**, e65.
69. Büttner,D., Nennstiel,D., Klüsener,B. and Bonas,U. (2002) Functional analysis of HrpF, a putative type III translocon protein from *Xanthomonas campestris* pv. vesicatoria. *J. Bacteriol.*, **184**, 2389–2398.
70. Rossier,O., Van den Ackerveken,G. and Bonas,U. (2000) HrpB2 and HrpF from *Xanthomonas* are type III-secreted proteins and essential for pathogenicity and recognition by the host plant. *Mol. Microbiol.*, **38**, 828–838.
71. Canteros,B.I. (1990), Ph.D. thesis. *University of Florida, Gainesville, FL.*
72. Menard,R., Sansonetti,P.J. and Parsot,C. (1993) Nonpolar mutagenesis of the *ipa* genes defines IpaB, IpaC, and IpaD as effectors of *Shigella flexneri* entry into epithelial cells. *J. Bacteriol.*, **175**, 5899–5906.
73. Bonas,U., Stall,R.E. and Staskawicz,B. (1989) Genetic and structural characterization of the avirulence gene *avrBs3* from *Xanthomonas campestris* pv. *vesicatoria*. *Mol. Gen. Genet.*, **218**, 127–136.
74. Huguet,E., Hahn,K., Wengelnik,K. and Bonas,U. (1998) *hpaA* mutants of *Xanthomonas campestris* pv. *vesicatoria* are affected in pathogenicity but retain the ability to induce host-specific hypersensitive reaction. *Mol. Microbiol.*, **29**, 1379–1390.
75. Figurski,D.H. and Helinski,D.R. (1979) Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided *in trans*. *Proc. Natl Acad. Sci. USA*, **76**, 1648–1652.
76. Gardner,P.P., Daub,J., Tate,J., Moore,B.L., Osuch,I.H., Griffiths-Jones,S., Finn,R.D., Nawrocki,E.P., Kolbe,D.L., Eddy,S.R. *et al.* (2010) Rfam: Wikipedia, clans and the 'decimal' release. *Nucleic Acids Res.*, **39**, D141–145.