Transcriptome changes in STSV2-infected Sulfolobus islandicus REY15A undergoing continuous CRISPR spacer acquisition

León Sobrino, Carlos; Kot, W.P.; Garrett, Roger Antony

Published in:
Molecular Microbiology

DOI:
10.1111/mmi.13263

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Molecular Microbiology (2016) 99(4), 719–728 ■
doi: 10.1111/mmi.13263
First published online 27 November 2015

Transcriptome changes in STSV2-infected *Sulfolobus islandicus* REY15A undergoing continuous CRISPR spacer acquisition

Carlos León-Sobrino,[1] Witold P. Kot[2] and Roger A. Garrett[1]*
[1]*Archaea Centre, Department of Biology, University of Copenhagen, Copenhagen DK-2200, Denmark.*
[2]*Department of Environmental Science, Aarhus University, Roskilde, Denmark.*

Summary

A transcriptome study was performed on *Sulfolobus islandicus* REY15A actively undergoing CRISPR spacer acquisition from the crenarchaeal monocaudavirus STSV2 in rich and basal media over a 6 day period. Spacer acquisition preceded strong host growth retardation, altered transcriptional activity of four different CRISPR-Cas modules and changes in viral copy numbers, and with significant differences in the two media. Transcript levels of proteins involved in the cell cycle were reduced, whereas those of DNA replication, DNA repair, transcriptional regulation and some antitoxin–toxin pairs and transposases were unchanged or enhanced. Antisense RNAs were implicated in the transcriptional regulation of adaptation and interference modules of the type I-A CRISPR-Cas system, and evidence was found for the occurrence of functional co-ordination between the single CRISPR-Cas adaptation module and the functionally diverse interference modules.

Introduction

CRISPR-Cas adaptive immune systems occur in most Archaea and in many Bacteria. Immunity is generated by the uptake of short DNA fragments (protospacers) from invading genetic elements, primarily viruses and plasmids and their insertion into host CRISPR arrays as spacers (reviewed in Barrangou and van der Oost, 2013). CRISPR loci are then transcribed and processed into crRNAs, each carrying part of a repeat sequence and a spacer sequence. After assembly of crRNAs into protein interference complexes, they can target and cleave DNA of the invading genetic element, or transcripts, within the matching protospacer sequences.

Members of the crenarchaeal order Sulfolobales are infected by a range of viruses exhibiting very diverse morphologies and genomes (Prangishvili *et al.*, 2006; Pina *et al.*, 2011), and they generally exhibit multiple CRISPR-Cas systems, primarily of types I-A and III-B (Vestergaard *et al.*, 2014). A few species, including *Sulfolobus solfataricus* P2 and *Sulfolobus islandicus* REY15A, have been established for laboratory studies because they can host a range of diverse viruses and are readily cultured and manipulated genetically (reviewed in Manica and Schleper, 2013; Garrett *et al.*, 2015).

Sulfolobus islandicus REY15A carries CRISPR loci 1 and 2 with 115 and 93 spacer-repeat units, respectively, which share identical repeat and leader sequences. In addition, it encodes a type I-A CRISPR-Cas system consisting of adaptation and interference modules, and two functionally different type III-B interference modules (denoted Cmr-α and Cmr-β), as well as a single Cas6 CRISPR RNA processing enzyme (Guo *et al.*, 2011). The type I-A interference module cleaves DNA, whereas the type III-B interference modules target transcripts (Cmr-β) or transcripts and DNA (Cmr-α) (Hale *et al.*, 2012; Deng *et al.*, 2013; Zebec *et al.*, 2014; Peng *et al.*, 2015).

The monocaudavirus STSV2 was selected for this transcriptome study because it can propagate stably, and in high yield, in different laboratory *Sulfolobus* hosts and, as it does not cause cell lysis, the infected cells are amenable to long-term studies (Erdmann *et al.*, 2014a). STSV2 has also been shown to undergo CRISPR spacer acquisition, at about 12 days post infection (dpi) when coinfected with the monocaudavirus SMV1 (Erdmann *et al.*, 2014b). Transcriptome studies were previously performed on different *Sulfolobus* species infected with the turreted icosahedral virus STIV (Ortmann *et al.*, 2008; Maaty *et al.*, 2012a,b), rudivirus SIRV2 (Kessler *et al.*, 2004; Okutan *et al.*, 2013; Quax *et al.*, 2013) and fuselloviruses SSV1 and SSV2 (Ren *et al.*, 2013) but in none of these studies was CRISPR-Cas spacer acquisition shown to be active.

The present study was designed primarily to examine interactions between STSV2 and *S. islandicus* REY15A

Accepted 28 October, 2015. *For correspondence. E-mail garrett@bio.ku.dk; Tel. +45 35322010; Fax +45 35322128

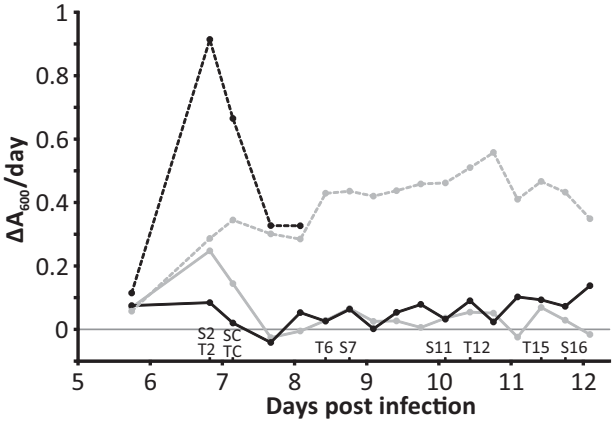

Fig. 1. Growth rates of control and STSV2-infected *S. islandicus* cultures grown in TYS (black) or SCV (gray) media. A_{260} measurements were made over 6–12 days and show cell density variations with time. Time points of the TYS (T) and SCV (S) samples selected for RNA sequencing are indicated. Control samples grown in the two media are represented by dashed lines; the control culture in TYS medium was discontinued due to cell saturation.

during CRISPR-Cas adaptation and interference. CRISPR spacer acquisition was induced after STSV2 infection and was ongoing over at least 6 days. Transcript level changes for protein and RNA genes, of host and virus, were examined at successive time intervals over the 6 day period. In addition, the experiment was performed in rich and basal media in order to examine the influence of nutrient availability on the viral DNA replication levels and on CRISPR-Cas adaptation (Erdmann *et al.*, 2014b).

Results

STSV2 infection of S. islandicus

STSV2 virions were isolated from an infected culture of *Sulfolobus tengchongensis* HB52 and infected in the laboratory *Sulfolobus* strain *S. islandicus* REY15A (Erdmann *et al.*, 2014a,b), and cultured in (i) rich TYS-medium (T) and (ii) basal SCV-medium (S). Samples were taken at 8 h intervals over 6–12 days post infection (dpi), and total RNA and DNA were extracted from each sample.

Strong growth retardation occurred at 6 dpi in TYS medium and at 7 dpi in SCV-medium. Thereafter, infected samples exhibited similar very low growth rates in both media, with a minor recovery occurring after 12 days in TYS medium (Fig. 1). After 12 days, samples were diluted in fresh medium and cell concentrations were measured for a day. Near-control growth levels were observed in TYS medium, but they remained low in SCV medium (data not shown).

STSV2 does not form visible plaques on Gelrite plates (Erdmann *et al.*, 2014a) and, therefore, intracellular STSV2:host molar genome ratios were measured at regular intervals by qPCR. In both media, STSV2 infection produced about 2:10 virus : host genome ratios at 6 dpi. The viral copy number increased at 8 dpi and reached a maximum ratio of 7:10 in TYS samples, and of 13:10 in SCV samples at 11 dpi (Fig. 2; Table S1). The peaks were followed by a final steep decrease in both media, likely coincident with virus release and CRISPR-Cas-induced viral DNA cleavage.

S. islandicus *and STSV2 transcriptomes*

Ten libraries were constructed from the purified RNA samples. They were representative of: (i) early low viral copy number phase (~ 7 dpi, T2/S2); (ii) viral DNA increase (~ 8.5 dpi, T6/S7); (iii) viral peak (~ 10 dpi, T12/S11); and (iv) reversion to low copy number (~ 11.5 dpi, T15/S16) (Fig. 1), as estimated from the virus : host genomic ratios measured by qPCR (Fig. 2). All virus-infected samples, except S2, were taken after the onset of growth retardation (Fig. 1), and uninfected samples served as controls. Libraries were pooled and loaded onto two Illumina sequencing flow cells yielding lanes of single-end HiSeq 50 bp reads and of paired-end MiSeq 250 bp reads, and the sequences were aligned on the host and viral genomes. Total numbers of transcript reads were obtained for each sample at the different time points (Table S2). Transcript abundances were calculated as Reads per Kilobase and Million mapped reads (RPKM), normalized and compared to those of uninfected control transcript libraries for both

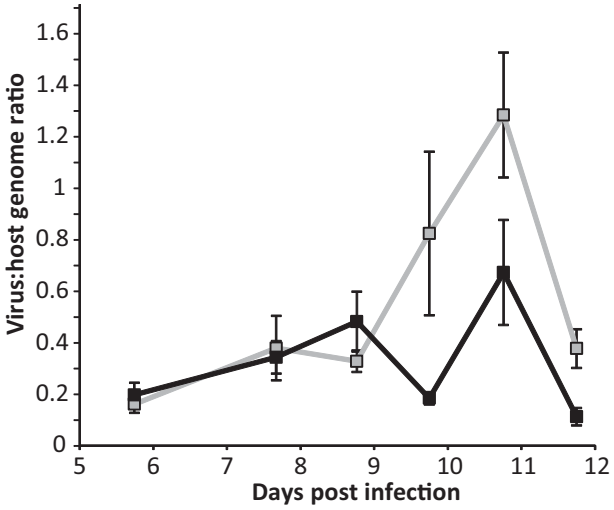

Fig. 2. Intracellular STSV2–host genome ratios determined for infected cultures grown in TYS (black) and SCV (gray) media. Estimated ratios assumed that 75% of the cellular population contained two chromosomes over the 6 day period. Genome ratios were measured in triplicate samples by qPCR. Error bars represent the technical standard deviation calculated as the standard deviation of all ratio pairs for each data point.

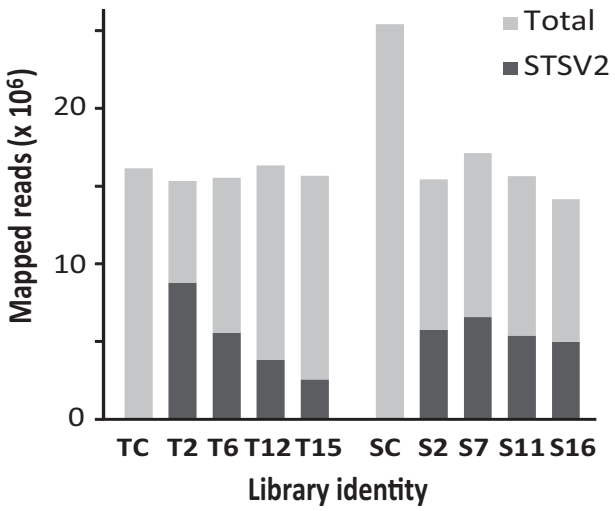

Fig. 3. Total number of transcript reads for each sample mapped to STSV2 and *S. islandicus* genomes.

media. Transcript levels of the host and viral genes in each sample are presented in Table S3. Transcript level changes of at least twofold in three or more consecutive libraries, or in two consecutive libraries in both media, were considered [qualitatively] significant. Non-aligned transcript sequences were assembled into contigs and subjected to BLAST searches against public sequence databases; most yielded imperfect alignments to host or STSV2 DNA, and no matches were detected to possible contaminating crenarchaeal genetic elements (www.ebi.ac.uk/genomes/archaealvirus.html). The fraction of viral transcripts decreased from about 60% to 16% of the total transcripts over the 6 days in TYS medium but remained constant at about 36% in SCV medium (Fig. 3).

CRISPR-Cas systems

The CRISPR loci and gene cassettes of the adaptation, and functionally diverse type I-A and type III-B interference modules, are illustrated in Fig. 4. Transcript levels from CRISPR loci and associated *cas* genes were quantified for uninfected and virus-infected samples. The values were used to infer virus-induced changes in the activities of the single CRISPR-Cas adaptation gene cassette (*csa1*, *cas1*, *cas2* and *cas4*) and of the type I-A and two type III-B interference cassettes, and they were compared for the two media.

Transcripts did not map uniformly on the CRISPR loci. In uninfected samples, CRISPR locus 2 yielded, on average, 72% of all detected CRISPR transcripts with a highest abundance for spacer regions 1–13, 26–44, 55–63 and 66–72 and with the leader-adjacent region being most prominent, whereas for CRISPR locus 1, there was a strong transcript bias to spacers 21 and 22. On infection, although overall transcript levels from both CRISPR loci increased, the biases did not change significantly. Transcription remained at a constant level in TYS medium but gradually increased in SCV medium over the 6 day period (Fig. 5), with corresponding two- to eightfold increases in transcripts of the Cas6 CRISPR RNA processing enzyme (SiRe0772). Quantification of CRISPR transcript levels was approximate because of a preferential loss of short transcripts, including crRNAs, during library preparation.

In uninfected cultures, few transcripts mapped to the CRISPR adaptation gene cassette but moderate transcript levels were observed for the interference *cas* and *cmr* gene cassettes. On infection, transcript levels from the former increased strongly, although in TYS medium there was subsequently a moderate decrease with time (Figs 5 and 6A). Transcript yields from the interference gene cassette increased strongly, especially in TYS medium (Figs 5 and 6B). In contrast, transcript levels from the type III-B interference gene cassettes did not increase; the Cmr-β module was initially downregulated in both media, with a slow recovery over the 6 days in TYS medium, whereas the Cmr-α module was weakly upregulated in TYS medium but remained reduced in SCV medium (Fig. 5). Transcript levels for the non-core Cas proteins Cmr7 (SiRe0603) and Csx1 (SiRe0884), associated with type III-B interference (Deng *et al.*, 2013), also decreased (Table S3).

Proteins Csa3a (SiRe0764) and Csa3b (SiRe0765) have been implicated in regulation of the adaptation and type I-A interference gene cassettes, respectively, and the former is a transcriptional activator (Liu *et al.*, 2015), and its transcript levels were enhanced significantly on infection in TYS medium, coincident with the onset of adaptation, and then gradually decreased with time. In contrast, transcripts of Cas3b remained at a low level after infection

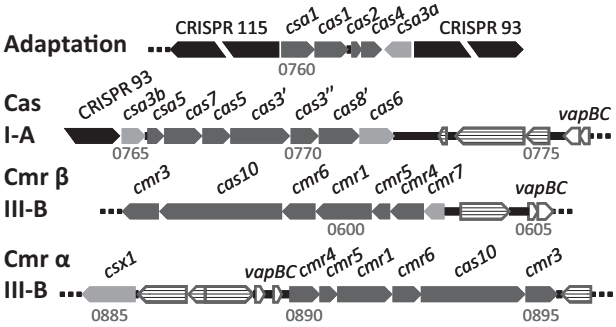

Fig. 4. Overview of the adaptation and interference CRISPR-Cas modules of *S. islandicus* REY15A. CRISPR loci 1 (115 repeat-spacer units) and 2 (95 units) are indicated. Core and non-core *cas* genes are shaded dark and light gray respectively. *vapBC* antitoxin–toxin gene pairs are labeled and mobile element-related genes are striped.

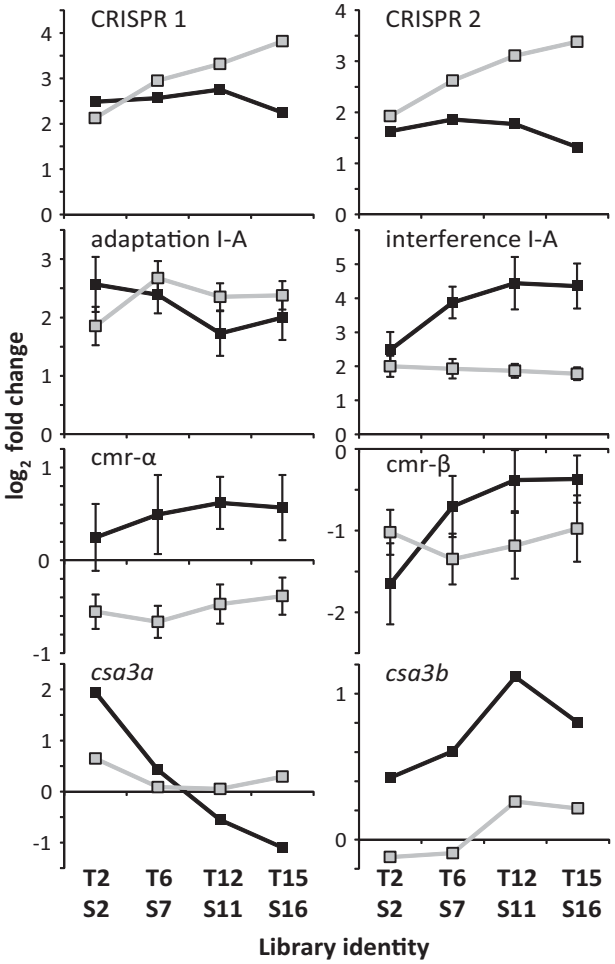

Fig. 5. Changes in transcript levels of the CRISPR-Cas gene cassettes with time. Results are presented for STSV2-infected samples in TYS medium (black lines) and SCV medium (gray lines). They are shown for the adaptation cassette, the type I-A interference cassette, the two type III-B Cmr cassettes and the regulator genes. Values are given as averages of gene log2 fold-change for each group of genes. Data for individual genes are given in Table S3.

(Fig. 5). Transcripts from the adaptation gene cassette provided evidence for alternative regulatory modes. While many spanned *cas1-cas2*, the paired-end read data indicated an additional transcriptional start site upstream of *cas2* (Fig. 6A). Moreover, a high level of antisense transcripts encompassed *cas2* and the upstream region of *cas4* in TYS samples, with lower yields in SCV samples. The antisense RNAs increased to about 34% of all transcripts from *cas2-cas4* at about 7 dpi in TYS and then gradually decreased (Fig. 6A). Evidence of internal regulation was also observed for the type I-A interference gene cassette. Transcriptional termination occurred immediately downstream from *cas7*, thereby dividing the cassette into high expression *csa5* and *cas7* and low expression *cas5*, *cas3*, *cas3′* and *cas8′* regions (Fig. 6B). Antisense transcripts also mapped to the terminal region of *csa5*, with highest yields in SCV medium at 7 dpi that decreased with time.

Spacer acquisition from STSV2

Leader end regions of both CRISPR loci were amplified by PCR to detect *de novo* CRISPR spacers that yielded larger PCR products in agarose gels (Erdmann and Garrett, 2012). Additional bands were detected from 6 dpi in TYS medium, and they increased in intensity and size with time, consistent with continuous spacer acquisition. In contrast, SCV samples yielded no well resolved larger PCR products indicating either the absence, or a low level, of spacer acquisition (data not shown).

In order to quantify spacer acquisition and determine the origin of the *de novo* spacers, and to test further for spacer acquisition in SCV samples, newly acquired spacers were identified from leader-end CRISPR regions in paired-end transcript reads. A total of 2842 reads were analyzed of which 431 carried unique *de novo* spacers in TYS and SCV samples. They mapped fairly evenly around the circular STSV2 genome and between DNA strands. Moreover, the gradual increase with time in both the numbers of unique new spacers, and the fraction of

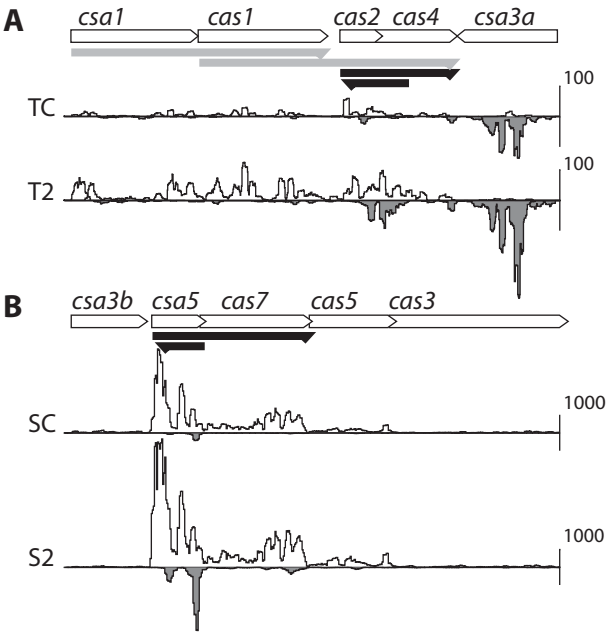

Fig. 6. Antisense and other newly detected transcripts from the type I-A adaptation and interference cassettes.
A. Adaptation gene cassette from TYS sample T2. B. Type I-A interference gene cassette (initial region only) from the SCV sample S2. Scale bars at the right indicate the level of transcript reads from the forward (above line) and reverse strands (below line). Novel transcripts are indicated by black arrows; gray arrows denote transcripts detected earlier from the *Sulfolobus* type I-A adaptation gene cassette (Liu *et al.*, 2015).

Table 1. Summary of the numbers of *de novo* STSV2 spacers sequenced in the different cultures derived from selected paired-end reads at the CRISPR leader ends. The total number of leader-end reads examined for each library are given together with the total numbers of *de novo* spacers identified. The results show the fraction of reads carrying one to four *de novo* spacers.

Library	Leader-end CRISPR reads	*De novo* spacer reads	Number of *de novo* spacers			
			1	2	3	4
TYS control	120	–	–	–	–	–
T2	149	24	20	–	4	–
T6	207	30	27	3	–	–
T12	369	89	72	12	5	–
T15	244	72	53	14	3	2
SCV control	94	1	–	1	–	–
S2	160	8	6	2	–	–
S7	302	22	18	4	–	–
S11	434	31	26	3	2	–
S16	763	70	60	9	1	–

transcript reads containing them, in both media, reinforced the PCR results for TYS samples that spacer acquisition was continuous (Table 1).

The cellular fraction that had undergone spacer acquisition was estimated from the ratio of transcripts carrying new spacer sequences to those with wild-type CRISPR sequences at the leader end (Table 1). The results indicated that at least 30% of the TYS cells, and 9% of the SCV cells, had undergone spacer acquisition over 6–12 dpi.

In summary, the transcript data underlined that the CRISPR-Cas immune response was more active in TYS medium, yielding more unique *de novo* spacers, lower viral copy numbers and decreasing levels of viral transcripts over the 6 day period. In contrast, SCV samples displayed a low level of new spacers, higher viral copy numbers and fairly constant levels of viral transcripts (Figs 2 and 3; Table 1).

Cell cycle, DNA replication, transcriptional regulation and toxins

During retarded growth and spacer acquisition, transcripts from cell cycle-related genes and especially the ESCRT/*cdv* operon (SiRe1173–SiRe1175) were reduced (Table 2). Nevertheless, many core genes were still expressed. Thus, transcript levels of the main DNA polymerase B homologue (SiRe1451) were not significantly repressed in infected samples while those of different DNA repair enzymes were enhanced, including DNA polymerase 2 (Dpo2 – SiRe0615) in TYS medium, DNA polymerase 4 (Dpo4 – SiRe0236) in SCV medium and the HerA-NurA operon (SiRe0061–0064) in both media (Table 2).

Altered transcript levels of basal transcriptional regulators were also consistent with major changes in cellular activity. Those of the main TFB factor (SiRe1144) were strongly reduced in both media, with a late recovery in TYS medium, which paralleled that of the *cdv* operon. In contrast, transcript levels of other TFB homologues (SiRe1555 and SiRe1717) changed little (Table S3). TFE can enhance transcription initiation at suboptimal promoters, and TFEα (SiRe1730) transcripts were upregulated in both media. Moreover, the transcript abundance for TFEβ (SiRe1145) decreased in TYS samples with time while that of termination factor Tfs (SiRe1707) increased only in SCV samples (Table 2). Furthermore, from a total of 66 genes encoding predicted transcriptional regulators, transcripts from 12 were upregulated and a further 11 were downregulated, with a few showing media-specific responses (Table 2 and S3).

Moreover, transcripts of several host antitoxin–toxin gene pairs were activated and may have contributed to retarded growth. They included nine of the 21 annotated *vapBC* pairs (Shah and Garrett, 2013), and another pair

Table 2. Significantly altered transcript levels from host genes involved in the cell cycle, DNA replication and repair, and basal transcriptional regulation.

Gene ID	Annotation	TYS	SCV
Cell cycle			
SiRe1173	CdvA	– (early)	–
SiRe1174	CdvB	– (early)	–
SiRe1175	CdvC	– (early)	–
DNA replication and repair			
SiRe0061	NurA 5′-3′ nuclease	+	+
SiRe0062	Rad50	+	+
SiRe0063	Mre11	+	+
SiRe0064	HerA	+	+
SiRe0236	DNA polymerase 4		+
SiRe0614	DNA polymerase 2 amino-end	+	
SiRe0615	DNA polymerase 2 elongation subunit	+	
SiRe0668	DNA-binding 7 kDa protein		+
SiRe0930	Holliday junction-type resolvase		–
SiRe1124	Reverse gyrase	– (early)	–
SiRe1176	DNA topoisomerase I	+	
SiRe1211	Replication initiator protein WhiP		–
SiRe1431	Holliday junction-type resolvase	+	+
SiRe1625	XerC/D integrase-recombinase	– (early)	–
SiRe1747	RadA		+
SiRe0240	RecA-like, KaiC	+	+
SiRe2648	DNA-binding 7 kDa protein	–	–
Basal transcription factors			
SiRe1144	TFB	– (early)	–
SiRe1145	TFE-β	– (late)	
SiRe1707	Tfs		+
SiRe1730	TFE-α	+	+

On infection: +, upregulation; –, downregulation. Blank cells indicate that there was no significant (> 2-fold) change in transcript levels in 2 or more samples from each series. 'Early' samples were taken at 7 and 8 dpi, 'late' samples at 10 and 11 dpi. Precise transcript fold-change data are given in Table S3.

Table 3. Transcript levels of host antitoxin–toxin gene pairs.

Genes	Antitoxin–toxin	TYS	SCV
SiRe0402-SiRe0403	VapBC	+ (early)	+ (early)
SiRe0698-SiRe0699	VapBC	+ (early)	+ (early)
SiRe0743-SiRe0744	VapBC	+ (early)	+
SiRe1476-SiRe1477	VapBC	+	+
SiRe2067-SiRe2068	VapBC	–	+
sSiRe2070-SiRe2071	VapBC	+	+
SiRe2072-SiRe2073	VapBC	+ (early)	+
SiRe2115-SiRe2116	VapBC	+	+
SiRe2294-SiRe2295	VapBC	+	+
SiRe0947	HEPN	–	

On infection: +, upregulation; –, downregulation. 'Early' samples were taken at 7 and 8 dpi, 'late' samples at 10 and 11 dpi. Precise transcript fold-change data are given in Table S3.

(SiRe-2067–SiRe2068) was repressed in TYS and enhanced in SCV medium (Table 3).

Changes in transcript levels of STSV2

Viral transcript levels were high in all samples despite the early low moi values (~ 0.2). A gradual decrease in transcripts was observed in TYS, but not in SCV samples, consistent with enhanced CRISPR interference and reduced viral copy numbers in TYS samples. Most transcripts mapped to a few viral genes, especially those encoding the main coat protein gp36 and adjacent gp37; other highly expressed genes, in all samples, included gp03, gp08 and gp43, all of unknown function.

Although the virus cycle could not be synchronized, phases were distinguishable from transcript level patterns, especially for early (7 and 8.5 dpi) and late (10 and 11.5 dpi) samples. Strong transcription occurred early for gene cluster gp70 to 75, and late for gp01, gp02 and gp04 near the putative replication origin, the multi-domain virion protein gp31, the putative multi-ribbon–helix–helix domain and for the putative regulators gp21, gp22 and gp23, and glycosyl transferase gp61. Transcripts of two putative OrfB transposases gp50 and gp66 were also upregulated late. The transcriptional regulator cluster gp17–gp25 also showed media and time-dependent variations in transcript levels, and antisense transcription was observed from gp24 and gp25 in the SCV sample S2 preceding growth retardation, suggesting an involvement in virus cycle progression together with gp01–gp04 (Tables S3 and S4).

Host and virus gene annotations

The short HiSeq transcript data provided evidence for many non-annotated genes (Guo *et al.*, 2011; Erdmann *et al.*, 2014a). Two hundred and six additional putative genes were mapped in the host genome, 26 of which exhibited known protein motifs (Table S3). In addition, 1288 potential antisense RNAs were located and 604 operon structures could be assigned with confidence based on paired-end read data (Table S3). In the STSV2 genome, two additional small highly expressed genes were located between gp54–gp55 (58535–58687) and gp57–gp58 (60144–60281). In contrast to the host, transcriptional read-through was common between adjacent inverted viral genes, and 3′-UTR transcripts from gp12–gp13, gp40–gp41 and gp44–gp45 invaded one another, possibly indicative of a regulatory role. Sixteen viral operons were manually curated using coverage-based statistics (Table S3).

Discussion

CRISPR-Cas module activation

Cell populations underwent spacer acquisition over 6–12 dpi, coincident with retarded growth, as judged by PCR results and transcript sequence analyses. Transcript levels from CRISPR loci, and the type I-A adaptation and interference modules, and Cas6, were strongly upregulated. In contrast, transcripts of type III-B interference modules underwent minor changes (Cmr-α) or were downregulated (Cmr-β) in both media during spacer acquisition, indicative of a possible functional co-ordination between different DNA targeting modules. Biases observed in the yields of CRISPR RNA processing products were consistent with earlier *Sulfolobus* studies (Wurtzel *et al.*, 2010; Deng *et al.*, 2013).

Production of antisense RNAs from type I-A adaptation and interference gene cassettes in TYS and SCV media, respectively, pointed to another level of regulation than that provided by Csa3a and Csa3b, respectively (Vestergaard *et al.*, 2014; Liu *et al.*, 2015). The antisense RNAs that peaked at 7 dpi and gradually decreased with time could inactivate transcripts of the putative toxins Cas2 and Csa5 (Koonin and Makarova, 2013; He *et al.*, 2014). Antisense *cas2-cas4* RNA was primarily observed in the early TYS sample, whereas *csa5* antisense RNA levels peaked prior to cell retardation in SCV samples.

CRISPR adaptation and DNA replication

High overall transcription was observed for DNA polymerase genes during retarded growth consistent with continuing replication of the viral genomes. Moreover, transcript levels were strongly increased for some DNA repair enzymes including the type B DNA polymerase 2 implicated in damage repair, and the type Y DNA polymerase 4 involved in lesion bypass (Choi *et al.*, 2011), in TYS and SCV medium, respectively. Their upregulation was coincident with increasing *de novo* CRISPR spacers and the inferred enhanced expression of the type I-A CRISPR-Cas

components. The results correlate with spacer acquisition being dependent on DNA replication in *S. islandicus* (Erdmann *et al*., 2014b) and receives further support from studies on bacterial CRISPR-Cas type I-E system which showed that formation of single strand breaks at DNA replication forks can stimulate adaptation and ensure that spacer acquisition occurs primarily from viral DNA in non-dividing cells (Levy *et al*., 2015). This is also consistent with adaptation genes being more actively transcribed in TYS medium where more nutrients are available for replication.

Host growth retardation

In an earlier study, attempts to induce CRISPR adaptation by infecting *Sulfolobus* hosts with STSV2 alone were unsuccessful and only minor growth retardation effects were observed, likely reflecting that the infection level was too low (Erdmann *et al*., 2013; 2014b). In the present study, when CRISPR adaptation was activated both infected and uninfected cells stopped dividing a few days after infection, maintaining low virus levels but with transcript levels of DNA replication and repair proteins remaining similar to, or higher than, those of uninfected control cultures. The nutrient dependency of growth retardation suggested that there is a threshold in intracellular virus activity. We infer that a host-dependent response ensures that uninfected cells also undergo retarded growth, consistent with SSV9 inducing population-wide growth arrest of *S. islandicus* at low moi values (Bautista *et al*., 2015).

Strongly enhanced upregulation of antitoxin-toxin gene pairs was coincident with growth retardation. Seven of 21 of *vapBC* host gene pairs, and one MNT-HEPN pair, were upregulated in infected samples and, moreover, four of the VapC toxins (SiRe0403, SiRe0699, SiRe0744 and SiRe2171) were upregulated in rudivirus SIRV2-infected *S. islandicus* LAL14/1 that also underwent growth retardation (Quax *et al*., 2013).

Antitoxin–toxin gene pairs are often clustered with CRISPR loci and *cas* genes for the Sulfolobales (Garrett *et al*., 2011), and it has been suggested that CRISPR spacer acquisition can activate toxins and induce cell arrest (Erdmann *et al*., 2014b). In the present study, it was difficult to distinguish growth inhibition effects resulting from viral infection from those induced by CRISPR adaptation. The strongest evidence for the latter is provided by the Csa5 toxin. In the SCV S2 sample, taken just prior to the onset of growth inhibition, there is a strong downregulation of antisense RNA against the Csa5 mRNA. The potential upregulation of the toxin, just prior to growth retardation, would also be consistent with earlier evidence showing that inhibition of Csa5 activity prevents the onset of cell death in SIRV2-infected *S. islandicus* REY15A (He *et al*., 2014).

Media-specific effects

For bacteria, richer media tend to produce shorter reproductive cycles and increased burst sizes of bacteriophages, thereby increasing viral dispersal (Abedon *et al*., 2001). The present results suggest that for STSV2, the CRISPR-Cas immune response counteracted such effects. STSV2 was initially more actively transcribed in the richer TYS medium but spacer acquisition was also more active, and the onset of growth retardation occurred earlier, consistent with STSV2 being more virulent in this medium. However, the increased efficiency in the uptake of new spacers and the strong upregulation of transcripts from the type I-A interference gene cassette in this medium probably produced cleavage of viral DNA, fewer viral transcripts and led to the recovery of cell growth after 12 days in TYS medium.

Comparison with other studies of virus-infected Sulfolobus

Infection with STIV and SIRV2 also produced upregulation of type I-A CRISPR interference gene cassettes, whereas for the type III interference modules transcripts of Cmr-β were upregulated while those of Cmr-α were unchanged or repressed (Maaty *et al*., 2012a; Quax *et al*., 2013). The data suggest that activation of type III-B interference responses is dependent on different factors including the nature of the virus, the medium and the presence of matching CRISPR spacers.

A direct comparison of our results with those of published transcriptome studies of virus-infected *Sulfolobus* species is limited because no evidence was presented for the occurrence of CRISPR adaptation earlier. Nevertheless, some general properties of the viral infection were shared including host growth retardation, observed for the lytic *Sulfolobus* viruses STIV, SIRV2, SMV1 and SSV9 (Ortmann *et al*., 2008; Erdmann and Garrett, 2012; Quax *et al*., 2013; Erdmann *et al*., 2014b; Bautista *et al*., 2015). In transcriptional studies of STIV and SIRV2, the *cdv* operon was downregulated only for SIRV2, but for both viruses transcripts of DNA polymerases remained stable, consistent with ongoing viral propagation, and transcript abundance for DNA repair enzymes was enhanced. Moreover, as for STSV2 infection, complex changes occurred in the transcript levels of the main transcription factors (Ortmann *et al*., 2008; Quax *et al*., 2013).

In conclusion, we have shown that major transcript abundance changes occur as a result of *Sulfolobus* viral infection and activation of CRISPR adaptation. Some cellular changes are common to those observed for different *Sulfolobus* viral infections in the absence of CRISPR adaptation. At present we know relatively little about detailed gene regulation mechanisms of the *Sulfolobus*

viruses and their hosts (Reeve, 2003; Werner, 2013), but the evidence for antisense RNAs in modulating expression of the I-A CRISPR adaptation and interference modules underlines that regulatory pathways in *Sulfolobus* species can be very complex. This inference was reinforced by the identification of up to 1288 putative host-encoded antisense RNAs, some of which have been characterized for other *Sulfolobus* species, including some complementary to CRISPR RNAs (Lillestøl *et al*., 2006; 2009; Deng *et al*., 2012) and many matching to mRNAs (Tang *et al*., 2005; Wurtzel *et al*., 2010). Clearly, the gene regulation possibilities of virus-infected *Sulfolobus* hosts are enormous.

Experimental procedures

STSV2 virus infection

STSV2 from a CsCl gradient-purified stock (Erdmann *et al*., 2014a) was used to infect *Sulfolobus tengchongensis* to obtain a fresh virus preparation. The infected cells were cultured in *Sulfolobus* medium supplemented with 0.2% trypton, 0.1% yeast extract and 0.2% sucrose (TYS medium) at 78°C and pH 3.5. Two liters of culture were centrifuged, and cell-free medium was filtered through 0.2 μm filters (Thermo Fisher Scientific, Waltham, MA) and 1,000,000 MWCO PES filters (Sartorius, Stedim, Goettingen, Germany). Virions were isolated by filtration. Density gradient purification was avoided to minimize virion damage and loss of infectivity. The virus particles were washed and resuspended in *Sulfolobus* medium. Filtrate samples were stained in 2% uranyl acetate and examined for homogeneity in a JEM-1010 transmission electron microscope (Jeol, Tokyo, Japan).

An *S. islandicus* REY15A culture, grown to late log phase (A_{600} = 0.8), was infected with STSV2. Cells were harvested from 14 ml culture, washed in *Sulfolobus* medium, resuspended in 400 μl of virus filtrate suspension and incubated at 75°C for 2.75 h. An moi value was estimated from qPCR of a separate culture at 1 dpi to be approximately 0.5. Free virions were removed by washing in *Sulfolobus* medium, and infected cells were resuspended in two 10 ml tubes containing this medium supplemented with (i) TYS-medium (0.2% trypton, 0.1% yeast extract and 0.2% sucrose) and (ii) SCV-medium (0.2% sucrose, 0.2% (w/v) vitamin-free casamino acids (Difco Vitamin Assay, BD, Franklin Lakes, NJ, USA) and a mixed vitamin solution (Zillig *et al*., 1993), both carrying 0.02 mg/ml uracil (Deng *et al*., 2009).

After 32 h, the culture was used to inoculate six 50 ml bottles, three with TYS medium and three with SCV medium, each with added uracil, to A_{600} = 0.02, and grown at 75°C with mixing. Five milliliter samples were taken at approximately 8 h intervals over 6 days, and A_{600} values were measured. One milliliter was taken from each sample for DNA extraction, washed in *Sulfolobus* medium to remove free virions, and cell pellets were stored at −20°. The remaining 4 ml of each sample was centrifuged, and the pellet was quickly frozen in liquid N_2 and stored at −80°C for total RNA extraction. The sample volume was restored by adding fresh pre-warmed medium for long-term cultivation.

STSV2 concentration

STSV2 does not form visible plaques in Gelrite plates, and therefore viral infection levels were estimated by qPCR of total DNA samples to determine virus : host genome ratios using primer pairs 5′-CAGGCAAGCGTGTAGTTTCA-3′ and 5′-TCAGGTCACTCTTGTTAGACCTT-3′ for STSV2_gp17, and 5′-GGGTCATCGAAAGTTAGCCTCA-3′ and 5′-CGGC GAATTTATTTGTGGCCA-3′ for *S. islandicus* SiRe1288, encoding ribosomal protein L3. Genome ratios were averaged from triplicate measurements of each target, using serial dilutions of purified DNA from virus and host as reference. It was inferred that about 75% of cells carry two chromosome copies during most of the *Sulfolobus* cell cycle (Lindås *et al*., 2008). The technical standard deviation was calculated as the standard deviation of all nine possible combinations of virus : host ratios for each data point.

DNA and RNA extraction

DNA was extracted using 1:1 phenol-chloroform and 15 μg glycogen (GlycoBlue, Thermo Fisher Scientific), followed by isopropanol precipitation and resuspension in 40 μl 10 mM Tris-HCl, pH 6.5 and 1 mM EDTA. $A_{260/230}$ and $A_{260/280}$ ratios were measured using NanoDrop (Thermo Scientific). Samples were diluted to 1 ng/ml prior to qPCR. Total RNA was extracted using TRIzol reagent (Sigma Aldrich, St Louis, MO). RNA was precipitated by adding 15 μg glycogen (GlycoBlue) in isopropanol. Possible DNA contaminations were removed by treatment with DNAse I (Life Technologies, Carlsbad, CA), and the RNA was resuspended in 40 μl Tris-HCl, pH 6.5, 1 mM EDTA and quality tested with NanoDrop (Thermo Scientific) and polyacrylamide-urea gel electrophoresis. RNA concentrations were estimated using Qubit (Life Technologies) prior to library preparation. Samples with sufficient RNA (> 50 ng/μl; $A_{260/280}$ and $A_{260/230}$ > 2) that were representative of each of the virus phases defined by qPCR were used to construct Illumina libraries. The 0.5 μg of total RNA was processed with the bacterial low input ScriptSeq Complete Kit (Epicentre). This involved RNA fragmentation such that true transcript ends could be underrepresented in the libraries. Library concentrations were measured with a Qubit fluorimeter (Life Technologies) and pooled to equal amounts. Quality testing and insert sizes were determined with a BioAnalyzer (Agilent Technologies, Santa Clara, CA) prior to loading into a flow cell channel in HiSeq2000 (Illumina, San Diego, CA) for 50 bp single-end reads, and MiSeq using a MiSeq Reagent kit v2 (Illumina) for 250 bp paired-end reads. Adaptors were trimmed prior to analysis using the Biopieces bioinformatics toolset (www.biopieces.org).

Transcriptome analysis

The 166.7 million single-end and 34,5 million paired-end reads were analyzed for the 10 libraries (Table S2). Reads were mapped independently on the host and viral genomes to avoid copy number-associated effects using ERANGE (Mortazavi *et al*., 2008) included in the CLC Genomics Workbench (CLC Bio, version 7.0, Qiagen, Hilden, Germany). Default settings of the algorithm for sequence matching were:

mismatch cost 2, gap opening and extension costs 3, minimal alignment/similarity region 80%. Calculated RPKM values were normalized by quantiles (Bolstad *et al*., 2003), and proportional fold-changes were assigned a statistical confidence threshold of $P = 0.05$ as calculated from Kal's Z-test (Kal *et al*., 1999) and applied FDR (False Discovery Rate) *P*-value corrections (Benjamini and Hochberg, 1995).

Comparisons with transcriptome data for SIRV2-infected *S. islandicus* LAL14/1 (Quax *et al*., 2013) involved accessing data available at ArrayExpress database (http://ww.ebi.ac.uk/arrayexpress), accession number E-MTAB-1660. These data were downloaded and processed using the procedure described above for STSV2-*S. islandicus* REY15A.

Operon structures were determined from single-base resolution coverage maps of aligned paired-end read data. Regions between adjacent genes, in the same orientation, were compared with a two-sample *t*-test to analyze whether coverage values correspond to a coherent group comprising the intergenic region and the flanking ORFs, with a *P*-value cutoff of 0.05 using the R statistics suite (http://www.r-project.org). Intergenic regions shorter than 5 bases were assumed positive. The accuracy of this test was examined visually for a subset of 156 intergenic regions in the host for all libraries and for the sum of coverage values of all, and the SCV-medium control library was chosen as the most accurate. Visual examination of 165 host intergenic regions revealed a low proportion of false negative predictions (10 out of 27) and false positives were rare (9 out of 138). STSV2 operon structures were determined using the same procedure and curated manually.

Novel transcripts were detected based on coverage (length ≥ 42 bp, coverage of ≥ 6 reads in ≥ 4 libraries) where those mapping to the reverse strand of annotated ORFs were classified as antisense RNAs. All transcript boundary estimates were approximate. Putative domains of newly annotated proteins were predicted from Blastx searches against the public archaeal RefSeq database. All the transcriptome data are available at E-MTAB-4027 in the ArrayExpress database (www.ebi.ac.uk/arrayexpress).

Monitoring adaptation

Spacer acquisition was monitored by amplifying leader-proximal regions of both CRISPR loci using GTCCATAGGAGGACCAGCTTTC and CCAACCCCTTAGTTC CTCCT CTATAG for locus 1, and GTTCCTTCCACTATGGGACTAG GAAC and CGTCACTGACACCATATTTATAC for locus 2. The fraction of the cellular population that underwent adaptation was estimated from image quantification of CRISPR locus bands run in 1% agarose gels. Leader proximal CRISPR reads were collected by selecting paired-end records containing a 36 nt sequence corresponding to the full repeat sequence CTTTCAATTCTATAGTAGATTAGC and 12 nt from the first spacer unit of CRISPR loci 1 and 2, on either strand using the Biopieces bioinformatics toolset (http://www.biopieces.org). Reads were aligned using BLAST to a local database of *S. islandicus* REY15A (ENA CP002425) and STSV2 (ENA NC_020077) using word size 28 nt, mismatch cost 3 and gap cost 2. Only forward reads of each library insert were used for quantification.

Acknowledgements

The research was supported by grants from the Danish Natural Science Research Council and Copenhagen University. We thank Shiraz A Shah, Susanne Erdmann, Xu Peng, Qunxin She and other members of the Danish Archaea Centre for helpful discussions. Lars Hestbjerg Hansen and Søren Sørensen generously provided access to the Copenhagen University Nucleic Acid Sequencing Facility. The authors have no conflicts of interest.

References

Abedon, S.T., Herschler, T.D., and Stopar, D. (2001) Bacteriophage latent-period evolution as a response to resource availability. *Appl Environ Microbiol* **67:** 4233–4241.

Barrangou, R., and van der Oost, J. (eds) (2013) *CRISPR-Cas Systems*. Berlin, Heidelberg: Springer.

Bautista, M.A., Zhang, C., and Whitaker, R.J. (2015) Virus-induced dormancy in the archaeon *Sulfolobus islandicus*. *mBio* **6:** e02565–14.

Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* **57:** 289–300.

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19:** 185–193.

Choi, J.-Y., Eoff, R.L., Pence, M.G., Wang, J., Martin, M.V., Kim, E.-J., *et al*. (2011) Roles of the four DNA polymerases of the crenarchaeon *Sulfolobus solfataricus* and accessory proteins in DNA replication. *J Biol Chem* **286:** 31180–31193.

Deng, L., Zhu, H., Chen, Z., Liang, Y.X., and She, Q. (2009) Unmarked gene deletion and host–vector system for the hyperthermophilic crenarchaeon *Sulfolobus islandicus*. *Extremophiles* **13:** 735–746.

Deng, L., Kenchappa, C.S., Peng, X., She, Q., and Garrett, R.A. (2012) Modulation of CRISPR locus transcription by the repeat binding protein Cbp1 in *Sulfolobus*. *Nucleic Acids Res* **40:** 2470–2480.

Deng, L., Garrett, R.A., Shah, S.A., Peng, X., and She, Q. (2013) A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. *Mol Microbiol* **87:** 1088–1099.

Erdmann, S., and Garrett, R.A. (2012) Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Mol Microbiol* **85:** 1044–1056, Corrigendum Mol Microbiol 86, 757.

Erdmann, S., Shah, S.A., and Garrett, R.A. (2013) SMV1 virus-induced CRISPR spacer acquisition from the conjugative plasmid pMGB1 in *Sulfolobus solfataricus*. *Biochem Soc Trans* **41**.

Erdmann, S., Chen, B., Huang, X., Deng, L., Liu, C., Shah, S.A., *et al*. (2014a) A novel single-tailed fusiform *Sulfolobus* virus STSV2 infecting model *Sulfolobus* species. *Extremophiles* **18:** 51–60.

Erdmann, S., Le Moine Bauer, S., and Garrett, R.A. (2014b) Inter-viral conflicts that exploit host CRISPR immune systems of *Sulfolobus*. *Mol Microbiol* **91:** 900–917.

Garrett, R.A., Vestergaard, G., and Shah, S.A. (2011)

Archaeal CRISPR-based immune systems: exchangeable functional modules. *Trends Microbiol* **19:** 549–556.
Garrett, R.A., Shah, S.A., Erdmann, S., Liu, G., Mousaei, M., León-Sobrino, C., *et al*. (2015) CRISPR-Cas adaptive immune systems of the Sulfolobales: unravelling their complexity and diversity. *Life* **5:** 783–817.
Guo, L., Brügger, K., Liu, C., Shah, S.A., Zheng, H., Zhu, Y., *et al*. (2011) Genome analyses of Icelandic strains of *Sulfolobus islandicus*: model organisms for genetic and virus-host interaction studies. *J Bacteriol* **193:** 1672–1680.
Hale, C.R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., *et al*. (2012) Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol Cell* **45:** 292–302.
He, F., Chen, L., and Peng, X. (2014) First experimental evidence for the presence of a CRISPR toxin in *Sulfolobus*. *J Mol Biol* **426:** 3683–3688.
Kal, A.J., van Zonneveld, A.J., Benes, V., van den Berg, M., Koerkamp, M.G., Albermann, K., *et al*. (1999) Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two different carbon sources. *Mol Biol Cell* **10:** 1859–1872.
Kessler, A., Brinkman, A.B., van der Oost, J., and Prangishvili, D. (2004) Transcription of the rod-shaped viruses SIRV1 and SIRV2 of the hyperthermophilic archaeon *Sulfolobus*. *J Bacteriol* **186:** 7745–7753.
Koonin, E.V., and Makarova, K.S. (2013) CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes. *RNA Biol* **10:** 679–686.
Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M., Amitai, G., *et al*. (2015) CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520:** 505–510.
Lillestøl, R.K., Redder, P., Garrett, R.A., and Brügger, K. (2006) A putative viral defence mechanism in archaeal cells. *Archaea* **2:** 59–72.
Lillestøl, R.K., Shah, S.A., Brügger, K., Redder, P., Phan, H., Christiansen, J., and Garrett, R.A. (2009) CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. *Mol Microbiol* **72:** 259–272.
Lindås, A.C., Karlsson, E.A., Lindgren, M.T., Ettema, T.J., and Bernander, R. (2008) A unique cell division machinery in the Archaea. *Proc Natl Acad Sci USA* **105:** 18942–18946.
Liu, T., Li, Y., Wang, X., Ye, Q., Li, H., Liang, Y., *et al*. (2015) Transcriptional regulator-mediated activation of adaptation genes triggers CRISPR *de novo* spacer acquisition. *Nucleic Acids Res* **43:** 1044–1055.
Maaty, W.S., Steffens, J.D., Heinemann, J., Ortmann, A.C., Reeves, B.D., Biswas, S.K., *et al*. (2012a) Global analysis of viral infection in an archaeal model system. *Front Microbiol* **3:** 411.
Maaty, W.S., Selvig, K., Ryder, S., Tarlykov, P., Hilmer, J.K., Heinemann, J., *et al*. (2012b) Proteomic analysis of *Sulfolobus solfataricus* during *Sulfolobus* turreted icosahedral virus infection. *J Proteome Res* **11:** 1420–1432.
Manica, A., and Schleper, C. (2013) CRISPR-mediated defense mechanisms in the hyperthermophilic archaeal genus *Sulfolobus*. *RNA Biol* **10:** 671–678.
Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5:** 621–628.
Okutan, E., Deng, L., Mirlashari, S., Uldahl, K., Halim, M., Liu, C., *et al*. (2013) Novel insights into gene regulation of the rudivirus SIRV2 infecting *Sulfolobus* cells. *RNA Biol* **10:** 875–885.
Ortmann, A.C., Brumfield, S.K., Walther, J., McInnerney, K., Brouns, S.J.J., van de Werken, H.J.G., *et al*. (2008) Transcriptome analysis of infection of the archaeon *Sulfolobus solfataricus* with *Sulfolobus* turreted icosahedral virus. *J Virol* **82:** 4874–4883.
Peng, W., Feng, M., Feng, X., Liang, Y.X., and She, Q. (2015) An archaeal CRISPR type III-B system exhibiting distinctive RNA targeting features and mediating dual RNA and DNA interference. *Nucleic Acids Res* **43:** 406–417.
Pina, M., Bize, A., Forterre, P., and Prangishvili, D. (2011) The archeoviruses. *FEMS Microbiol Rev* **35:** 1035–1054.
Prangishvili, D., Forterre, P., and Garrett, R.A. (2006) Viruses of the Archaea: a unifying view. *Nat Rev Microbiol* **4:** 837–848.
Quax, T.E.F., Voet, M., Sismeiro, O., Dillies, M.-A., Jagla, B., Coppée, J.-Y., *et al*. (2013) Massive activation of archaeal defense genes during viral infection. *J Virol* **87:** 8419–8428.
Reeve, J.N. (2003) Archaeal chromatin and transcription. *Mol Microbiol* **48:** 587–598.
Ren, Y., She, Q., and Huang, L. (2013) Transcriptomic analysis of the SSV2 infection of *Sulfolobus solfataricus* with and without the integrative plasmid pSSVi. *Virology* **441:** 126–134.
Shah, S.A., and Garrett, R.A. (2013) Archaeal type II toxin-antitoxins. In *Prokaryotic Toxin-Antitoxins*. Gerdes, K. (ed.). Berlin Heidelberg: Springer, pp. 225–238.
Tang, T.-H., Polacek, N., Zywicki, M., Huber, H., Brügger, K., Garrett, R.A., *et al*. (2005) Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. *Mol Microbiol* **55:** 469–481.
Vestergaard, G., Garrett, R.A., and Shah, S.A. (2014) CRISPR adaptive immune systems of Archaea. *RNA Biol* **11:** 156–167.
Werner, F. (2013) Molecular mechanisms of transcription elongation in Archaea. *Chem Rev* **113:** 8331–8349.
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.
Zebec, Z., Manica, A., Zhang, J., White, M.F., and Schleper, C. (2014) CRISPR-mediated targeted mRNA degradation in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res* **42:** 5280–5288.
Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janekovic, D., Hain, J., *et al*. (1993) Screening for Sulfolobales, their plasmids and their viruses in Icelandic solfataras. *Syst Appl Microbiol* **16:** 609–628.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site: