

## Genome Analyses of Icelandic Strains of *Sulfolobus islandicus*, Model Organisms for Genetic and Virus-Host Interaction Studies<sup>∇</sup>

Li Guo,<sup>1†</sup> Kim Brügger,<sup>2†</sup> Chao Liu,<sup>2†</sup> Shiraz A. Shah,<sup>2†</sup> Huajun Zheng,<sup>3</sup> Yongqiang Zhu,<sup>3</sup> Shengyue Wang,<sup>3</sup> Reidun K. Lillestøl,<sup>2</sup> Lanming Chen,<sup>2</sup> Jeremy Frank,<sup>2</sup> David Prangishvili,<sup>4</sup> Lars Paulin,<sup>5</sup> Qunxin She,<sup>2‡</sup> Li Huang,<sup>1‡\*</sup> and Roger A. Garrett<sup>2‡\*</sup>

State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, No. 1 West Beichen Road, Chaoyang District, Beijing 100101, China<sup>1</sup>; Archaea Centre, Department of Biology, Copenhagen University, Ole Maaløes Vej 5, DK-2200N Copenhagen, Denmark<sup>2</sup>; Shanghai MOST Key Laboratory of Disease and Health Genomics, Chinese National Human Genome Center at Shanghai, Shanghai 201203, China<sup>3</sup>; Molecular Biology of the Gene in Extremophiles Unit, Institut Pasteur, rue Dr Roux 25, 75724 Paris Cedex, France<sup>4</sup>; and DNA Sequencing and Genomics Laboratory, Institute of Biotechnology, University of Helsinki, 00790 Helsinki, Finland<sup>5</sup>

Received 10 December 2010/Accepted 16 January 2011

The genomes of two *Sulfolobus islandicus* strains obtained from Icelandic solfataras were sequenced and analyzed. Strain REY15A is a host for a versatile genetic toolbox. It exhibits a genome of minimal size, is stable genetically, and is easy to grow and manipulate. Strain HVE10/4 shows a broad host range for exceptional crenarchaeal viruses and conjugative plasmids and was selected for studying their life cycles and host interactions. The genomes of strains REY15A and HVE10/4 are 2.5 and 2.7 Mb, respectively, and each genome carries a variable region of 0.5 to 0.7 Mb where major differences in gene content and gene order occur. These include gene clusters involved in specific metabolic pathways, multiple copies of VapBC antitoxin-toxin gene pairs, and in strain HVE10/4, a 50-kb region rich in glycosyl transferase genes. The variable region also contains most of the insertion sequence (IS) elements and high proportions of the orphan *orfB* elements and SMN1 miniature inverted-repeat transposable elements (MITEs), as well as the clustered regular interspaced short palindromic repeat (CRISPR)-based immune systems, which are complex and diverse in both strains, consistent with them having been mobilized both intra- and intercellularly. In contrast, the remainder of the genomes are highly conserved in their protein and RNA gene synteny, closely resembling those of other *S. islandicus* and *Sulfolobus solfataricus* strains, and they exhibit only minor remnants of a few genetic elements, mainly conjugative plasmids, which have integrated at a few tRNA genes lacking introns. This provides a possible rationale for the presence of the introns.

Iceland has been a rich source of hyperthermophilic crenarchaea over the past 3 decades and especially of acidothermophilic members of the order *Sulfolobales*. Many *Sulfolobus islandicus* strains (“Island” is German for “Iceland”) have also yielded many novel viruses showing varied and sometimes unique morphologies and exceptional genome contents. These properties are consistent with these viruses constituting an archaeal lineage distinct from those of bacteria and eukarya, and they have now been classified into several new viral families (38, 63). In addition, a family of conjugative plasmids has been characterized, with most members deriving from Iceland, which appear to conjugate by a mechanism unique to the archaeal domain (18, 37).

Although the availability of genome sequences of *Sulfolobus*

strains and their genetic elements has yielded important insights into the biology of these model crenarchaea, a major impediment to more detailed insights has been the paucity of robust and versatile vector-host systems for genetic studies. A few *Sulfolobus* species have been successfully employed as hosts for such systems, including *Sulfolobus solfataricus* strains P1 and 98/2 (22, 58), *Sulfolobus acidocaldarius* (57), and *S. islandicus* strain REY15A (54). To date, the genetic tools developed for the latter host are the most versatile and include the following: (i) *Sulfolobus-Escherichia coli* shuttle vectors carrying either viral or plasmid replication origins (50); (ii) conventional and novel gene knockout methodologies (14, 62), and (iii) a D-arabinose-inducible expression system with a *lacS* reporter gene system (35). The *S. islandicus* system has also been employed successfully to demonstrate the dynamic character of the clustered regular interspaced short palindromic repeat (CRISPR)-based immune systems of *Sulfolobus* when challenged with genetic elements carrying matching viral gene and protospacers maintained under selection (20). These developments necessitated the determination of the genome sequence of *S. islandicus* strain REY15A as a prerequisite for successful exploitation of the genetic systems.

A second Icelandic strain, *S. islandicus* strain HVE10/4, has been employed as a broad laboratory host for propagating diverse *Sulfolobus* viruses and conjugative plasmids (63) and

\* Corresponding author. Mailing address for R. A. Garrett: Archaea Centre, Department of Biology, Copenhagen University, Ole Maaløes Vej 5, DK-2200N Copenhagen, Denmark. Phone: 045-353-22010. Fax: 045-353-22128. E-mail: garrett@bio.ku.dk. Mailing address for L. Huang: State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, No. 1 West Beichen Road, Chaoyang District, Beijing 100101, China. Phone: 086-10-64807430. Fax: 086-10-64807429. E-mail: huangl@sun.im.ac.cn.

† These authors contributed equally.

‡ The last three authors are joint senior authors.

∇ Published ahead of print on 28 January 2011.

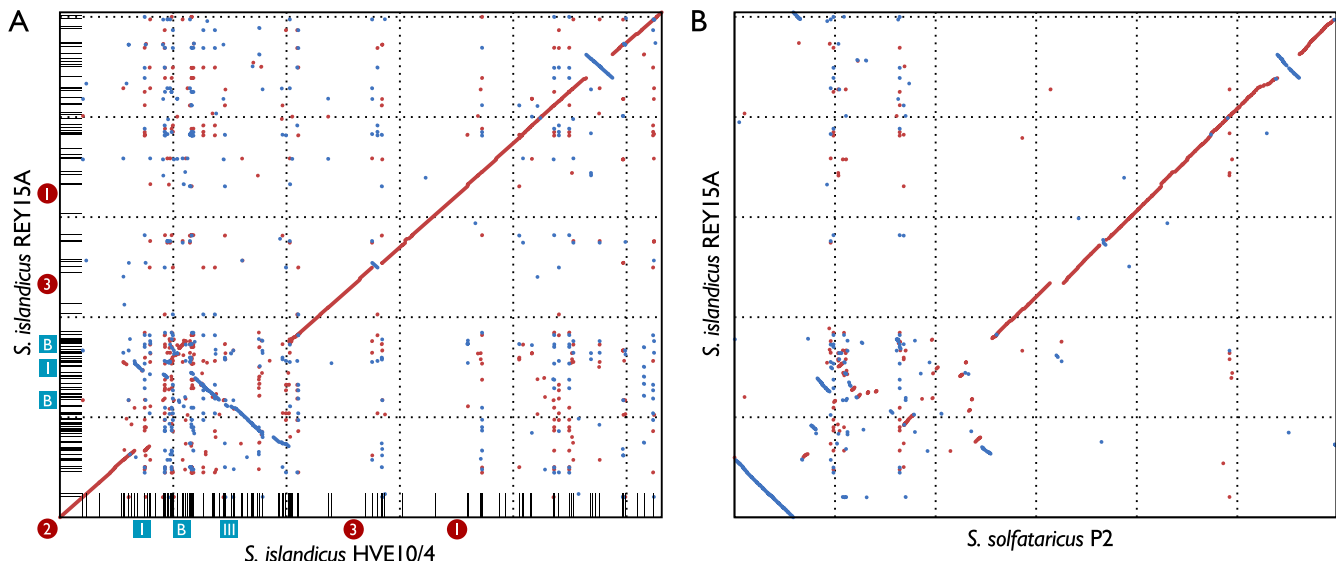


FIG. 1. (A) Dot plot of the two Icelandic genomes showing the approximate levels of sequence synteny. The large variable regions extend from about 0.35 to 1.0 Mb. Transposase genes are denoted by black lines along the axes. Putative origins of replication adjacent to the *cdc6* and *whiP* genes are indicated with red circles, while the families of the CRISPR/Cas (I and III) and Cmr (B) modules are indicated by blue squares. (B) Dot plot of the *S. islandicus* REY15A and *S. solfataricus* P2 genomes.

was selected for in-depth studies of their life cycles and host interactions. This effort received added impetus with the demonstration that some genetic elements show exceptional and sometimes unique properties of their viral life cycles or conjugative mechanisms (3, 8, 18, 40). Therefore, the genome sequence of *S. islandicus* strain HVE10/4 was also determined.

The genome sequences of two Icelandic strains, REY15A and HVE10/4, were analyzed and compared and contrasted with one another and with genomes of other *S. solfataricus* and *S. islandicus* strains isolated from different geographical locations, including Naples, Italy; Kamchatka, Russia; Lassen Volcanic National Park; and Yellowstone National Park (44, 53).

**MATERIALS AND METHODS**

**Genome sequencing.** *S. islandicus* strains REY15A and HVE10/4 were colony purified three times and cultured essentially as described earlier (11). Total DNA was extracted from the cells using phenol-chloroform and further purified by CsCl density-gradient centrifugation. For strain REY15A, sequencing of shotgun libraries with a 454 GS FLX sequencer yielded 324,123 reads with 31-fold genome coverage. For strain HVE10/4, DNA was sonicated to yield fragments in the size range of 1.5 to 4.0 kb, and clone libraries were generated in pUC18 using the SmaI site. Sequencing was performed on MegaBace 1000 sequencers to yield approximately 3-fold sequence coverage, and the sequencing data were combined with a sequencing run using a 454 FLX sequencer to yield approximately 10- to 15-fold coverage. The genome sequences were assembled using the phred/phrap/consed package, contigs were linked by combinatorial PCR using primers matching to each contig end, and the PCR products were sequenced to close the gaps. Remaining ambiguous sequence regions in the genome were identified and resolved by generating and sequencing PCR products. Both genomes were annotated automatically and refined manually.

**Sequence analyses.** Open reading frames (ORFs) were predicted with Glimmer (13). Frameshifts were detected and checked by sequencing after manual annotation, and the remaining frameshifts were considered to be authentic. Functional assignments of ORFs are based on searches against GenBank (<http://www.ncbi.nlm.nih.gov/>) and the Conserved Domain Database (CDD) ([www.ncbi.nlm.nih.gov/cdd/](http://www.ncbi.nlm.nih.gov/cdd/)). tRNA genes were located with tRNAscan-SE (26). Potential noncoding RNAs were predicted by comparison with the untranslated RNAs characterized for *S. solfataricus* and *S. acidocaldarius*, in terms of sequence similarity and gene context (see Results). Putative insertion sequence

(IS) elements were identified by BLASTN search against the IS Finder database (<http://www-is.biotoul.fr/>). All annotations were manually curated using Artemis software (47).

**RESULTS**

**Genome general properties.** Genomes of the two Icelandic strains were sequenced using a combination of sequencing strategies. *S. islandicus* REY15A was determined primarily by 454 sequencing, while strain HVE10/4 was obtained by a combination of Sanger and 454 sequencing at approximately 30-fold and 10-fold coverage, respectively. Protein-coding genes were annotated in Artemis (47), where start codons for single genes and first genes of *Sulfolobus* operons were generally located 25 to 30 bp downstream from the archaeal hexameric TATA-like box and only genes within operons were preceded by Shine-Dalgarno motifs, of which GGUG predominates (56). Where alternative start codons were juxtapositioned, we selected the most probable on the basis of its position relative to the putative promoter and/or Shine-Dalgarno motifs or experimental data from closely related organisms.

Dot plots of the two genomes demonstrate long sections of gene synteny. One region of about 0.5 to 0.7 Mb exhibits extensive gene shuffling, and there is a smaller region with a 200-kb inversion bordered by shuffled genes (Fig. 1A). Some of the minor irregularities in the dot plot were attributable to insertion or integration events. The synteny is maintained, to a large degree, when each genome is compared to that of *S. solfataricus* P2, despite the occurrence of a large inversion in the latter, and this is illustrated in a dot plot for the genomes of strain REY15A and *S. solfataricus* P2 (Fig. 1B). This extensive gene synteny is surprising, given the high level of transpositional activity occurring in *S. solfataricus* (Table 1) (7, 30, 41). A similar pattern was also observed when other pairs of *S. islandicus* genomes from different geographical locations were

TABLE 1. Summary of genetic properties obtained from genomes of two Icelandic *S. islandicus* strains and other available *S. solfataricus* and *S. islandicus* strains

Characteristic	Genetic properties obtained from genomes of <sup>a</sup> :													
	SsoP2	Sso98/2	REY15A	HVE10/4	LD8.5	LS2.15	M16.4	M16.27	M14.25	YG57.14	YN15.51			
Origin	Naples, Italy	Unknown	Reykjanes, Iceland	Hvergaard, Iceland	Lassen, USA	Lassen, USA	Kamchatka, Russia	Kamchatka, Russia	Kamchatka, Russia	Yellowstone, USA	Yellowstone, USA	Yellowstone, USA	Yellowstone, USA	Yellowstone, USA
GenBank accession no.	AE006641	CP001402	CP002425	CP002426	CP001731	CP001399	CP001402	CP001401	CP001400	CP001403	CP001404			
Genome size (Mb)	3.0	2.7	2.5	2.7	2.7	2.7	2.6	2.7	2.6	2.7	2.8			
No. of:														
Conserved genes (total, 1,679)	675	656	765	847	837	842	848	823	797	869	840			
Unique single genes (total, 1,346)	190	138	118	114	209	100	100	75	49	113	140			
Transporters (total, 15)	11	11	13	14	11	11	12	11	14	12	10			
VapBC antitoxin-toxins	20	18	16	18	21	24	21	21	21	20	19			
Glycosyl transferases (50-kb region)	Absent	Absent	Absent	15	Absent	15	15	15	15	Absent	15			
Conserved noncoding RNAs	123	81	44	42	42	44	39	42	43	42	42			
Transposases/IS elements	168	158	75	65	68	60	34	47	45	103	130			
MITEs (families)	155 (6)	133 (6)	9 (2)	11 (2)	4 (1)	4 (1)	10 (2)	5 (2)	7 (2)	5 (1)	5 (1)			
Cmr family(ies) <sup>b</sup>	D, 2× B (B)	B	2× B	B	B	2× B, D	B	B, D	B, D	3× B	E			
CRISPR/Cas family(ies) <sup>a</sup>	I, II (2× I)	II (2× I)	I	I, III	I, III (I)	I (II)	I	I, II	I, II	I	I			

<sup>a</sup> Letters and numbers in parentheses for the Cmr and CRISPR/Cas modules families (25, 48) denote the numbers and families of putatively defective modules generally lacking essential genes.

<sup>b</sup> Lassen, Lassen Volcanic National Park; Yellowstone, Yellowstone National Park.

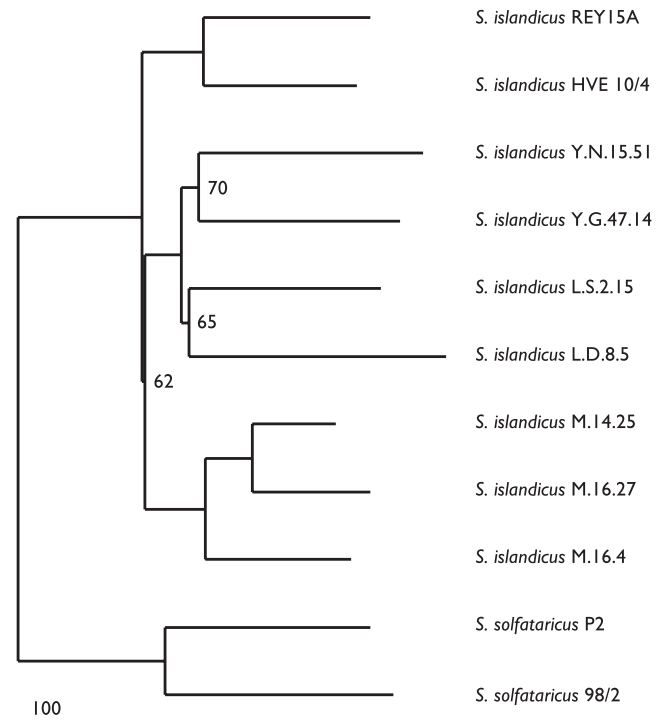


FIG. 2. Neighbor-joining tree based on a gene content matrix, including the conserved, core, and unique genes for each available *S. islandicus* and *S. solfataricus* genome (Table 1). The branch lengths represent the number of differences between the strains in terms of the presence or absence of individual genes. The data for the tree were prepared using methods described earlier (44, 48). Only bootstrap values below 100% for the individual branches are given.

compared (48), consistent with a high level of conservation of gene synteny for all the *S. solfataricus* and *S. islandicus* genomes.

A phylogenetic tree derived from the available genomes clusters together *S. islandicus* strains from different geographical locations (44), with *S. solfataricus* strains P2 and 98/2 being more distantly related (Fig. 2 and Table 1). The nucleotide sequence identity for the concatenated core genes of the two *S. islandicus* genomes (Fig. 1A) is 99.6%, and between all the *S. islandicus* genomes, it is about 99%. The relatively long branches for individual strains (Fig. 2) arise mainly from differences in gene content of the large variable regions (Fig. 1A). The degree of sequence identity between the concatenated core genes of the *S. islandicus* and *S. solfataricus* genomes is about 90% (Fig. 2).

Three origins of chromosome replication, demonstrated experimentally for *S. solfataricus* and *S. acidocaldarius* (27, 46), are well conserved with respect to both the DNA sequence and flanking gene organization in both of the genomes, albeit with the origin *oriC2* being inverted relative to the genomes of *S. solfataricus* P2 and *S. islandicus* strain YN1551 (Fig. 1B). Origin *oriC1* lies immediately upstream of *cdc6-1*, *oriC2* is close to *cdc6-3*, while *oriC3* is positioned downstream of the *whiP* gene (Fig. 1A). The two *cdc6* genes and the *whiP* gene encode putative replication initiators (45).

**Large variable region.** The genomes carry two types of variable regions. The large region, constituting 20 to 25% of each

TABLE 2. Integration events at tRNA genes showing the sizes and origins of the residual integrated genes

tRNA	Intron present	Integration event <sup>a</sup>		Conserved?
		REY15A	HVE10/4	
Val—TAC	No	SiRe1242-1247 conj plasmid	No insert	No
Phe—GAA	No	SiRe1321-1323 conj plasmid	SiH1399-1402 conj plasmid	Yes
Met—CAT	Yes	SiRe1465-1479, 12 kb + IS, pNOB8 integrase, unknown	No insert	No
Glu—TTC	No	SiRe1484-1490 conj plasmid	SiH1561-1574 conj plasmid	Yes
Ala—GGC	No	<i>intN</i> fragment	<i>intN</i> fragment	Yes
Thr—GGT	No	SiRe2413-2417 IS/MITEs SSV	SiH2464-24672 SSV	Partly
Pro—GGG	Yes	<i>intN</i> fragment	<i>intN</i> fragment	Yes
His—GTG	No	SiRe1787-1792, 7 kb + IS	No insert	No

<sup>a</sup> SSV, spindle-shaped fusellovirus; conj, conjugative; int, integrase.

genome, extends approximately from positions encompassing 0.3 to 0.8 Mb and 0.3 to 1.0 Mb for strains REY15A and HVE10/4, respectively (Fig. 1A). The other class is represented mainly by regions downstream from tRNA genes, where integration events have occurred (Table 1; also, see below). The large variable region contains about 60% of the potentially transposable IS elements and most of the nonautonomous mobile elements, as well as many degenerate copies of the former (Fig. 1A). It carries some gene clusters, which are present in one or more of the *Sulfolobus* genomes, including operons and gene cassettes associated with metabolic pathways, and it contains the diverse CRISPR/Cas and Cmr modules (Table 1; also, see below). It generally lacks essential genes; for example, no tRNA genes or replication origins are present, and thus, it appears to constitute a region where nonessential genes are collected, interchanged, and exchanged intercellularly and where genetic innovation occurs.

**Integration sites.** tRNA gene integration events in *Sulfolobus* genomes predominantly involve conjugative plasmids and fuselloviruses, and these were also the genetic elements most commonly isolated from acidic hot springs in Iceland (63). Most integration events occur via an archaea-specific mechanism, whereby a viral/plasmid integrase gene recombines into a host tRNA gene and partitions (32). The capture of a genetic element in a chromosome leaves a trace because the *intN* fragment overlapping the tRNA gene is generally maintained, even if the remainder of the genetic element degenerates or is deleted (51, 52) (Table 2).

For strains REY15A and HVE10/4, remnants of integrated elements adjoin eight and five tRNA genes, respectively (Table 2). Most of the integrated genes derive from conjugative plasmids, and fuselloviral genes were detected only at tRNA<sup>Thr</sup>[GGT] in each strain, with an integrated region of unknown origin at tRNA<sup>Met</sup>[CAT] in strain REY15A. All of the integrated elements are highly degenerate, with IS elements or miniature inverted-repeat transposable elements (MITEs) inserted downstream from the tRNA genes (Table 2). Given the possibility of multiple integrations of genetic elements occurring at a given tRNA gene, it is difficult to analyze unambiguously the origins of residual integrated genes (42).

In contrast to the two Icelandic strains, the other *S. solfataricus* and *S. islandicus* genomes carry intact genetic elements bordered by *intN* and *intC* fragments that are all potentially excisable (44, 52). They each show evidence of 2 to 7 tRNA gene integration events, in which the most conserved sites are tRNA<sup>Pro</sup>[GGG] and tRNA<sup>Ala</sup>[GGC], with less common events

at tRNA<sup>Leu</sup>[GAG] and different alleles of tRNA<sup>Arg</sup> (Table 2). For the integrated tRNA genes of the Icelandic strains, there was no significant correlation between the identity of the tRNA anticodon and the frequency of codon usage or between the encoded amino acid and the average number of amino acids in the genome-encoded proteins.

**Anti-integration role for tRNA introns.** Each genome carries 45 tRNA genes and 2 to 3 pseudo-tRNA genes all located in conserved regions. Sixteen of the tRNA genes contain introns immediately 3' to the anticodon, varying in size from 12 to 65 bp, and in contrast to many archaeal tRNA genes, none were detected at other sites (29), although putatively degenerate introns, lacking the capacity to form splicing sites, occur in D-loop regions of tRNA<sup>Glu</sup>[CTC] and tRNA<sup>Glu</sup>[TTC]. Moreover, the tRNA genes and introns are highly conserved in sequence between the two genomes, and also with the other six *S. islandicus* genomes, with very few base changes occurring between the introns of a given tRNA. This high level of tRNA and intron sequence conservation extends to *S. solfataricus* P2, with only very minor differences observed for about one-third of the genes, and it reinforces the concept that the RNA introns are functionally important (5).

A possible function for the tRNA introns, suggested by the above-described analyses, is that they provide protection against integration of genetic elements into tRNA genes. Integration can be disadvantageous in that pre-tRNA transcription can be impaired. Only two intron-carrying tRNA genes showed evidence of integration events (Table 2). For the tRNA<sup>Met</sup>[CAT] gene copies, an intact integrase gene is located downstream from the tRNA gene, while for the tRNA<sup>Pro</sup>[GGG], an overlapping *intN* fragment is present, but the overlapping sequence does not extend to the intron, suggesting that the intron entered after the integration event. This is consistent with the latter integration event being the most conserved, and probably the most ancient, among *Sulfolobus* species.

**IS elements and the versatile *orfB* element.** Each genome carries a limited range of IS element types, with some in multiple copies (Table 3). The IS elements are clustered in the variable genomic region and also downstream from tRNA genes that have undergone integration events (Fig. 1A). Many of these elements appear to be intact, carrying the inverted terminal repeats (ITRs) required for transposition, but exhibit fragmented transposase genes, which are unlikely to be restored by programmed translational frameshifting, as was observed for some bacterial transposases of the IS1 and IS3



TABLE 3. Properties of the IS elements, transposases, and MITEs in the Icelandic genomes<sup>a</sup>

Element	Family	No. of:					Conserved genome positions
		ORFs	REY15A copies	Intact TPases in REY15A	HVE10/4 copies	Intact TPases in HVE10/4	
ISC796	IS1	1	5	0	4	1	1
ISC1043	ISL3	1	1	0	1	0	1
ISC1048	IS630	1	10	0	12	0	3
ISC1058	IS5	1	2	0	1	0	1
ISC1078	IS630	1	1	0	1	0	0
ISC1190	IS110	1	3	1	1	0	0
ISC1200	ISH3	1	22	11	7	3	0
ISC1205	ISCNY	2	3	0	4	2	1
ISC1229	IS110	1	4	2	10	9	1
ISC1234	IS5	1	8	6	7	5	1
ISC1332	IS256	2	1	1	1	1	0
ISC1395	IS630	1/2	2	1	0	0	0
ISC1733	IS200/IS605	2	8	8	2	2	1
ISC1921	IS607	2	1	1	0	0	0
ISSis1 (pARN4)	IS6	1	4	2	5	4	0
ISSto2	IS6	1	2	1	2	2	0
OrfB	IS605	1	19 (8)	18	18 (2)	17	0
SM3A		0	2	0	2	0	2
SMN1		0	7	7	9	9	0

<sup>a</sup> The nomenclature used for IS elements and MITEs follows that which was used previously (2, 6, 16). For the OrfB elements, the numbers in parentheses indicate the numbers of copies that are physically linked to ISC1200 elements. TPase, transposase.

families (28). Although some of these elements may be mobilizable by transposases acting in *trans*, for over one-third of the IS families present, there is no encoded transposase (Table 3). Potentially, the most active elements are ISC1200 and ISC1234 in both genomes and ISC1229 in strain HVE10/4 (Table 3). The two Icelandic *S. islandicus* strains, together with those from Kamchatka, Russia, carry the lowest number of IS elements (Table 1), many of which are inactive.

*orfB* elements of family IS605, together with elements of the IS6 family (Table 3), are considered to represent the few classes of transposable elements that are ancestral to the archaeal domain (16). *orfB* occurs alone, or together with a transposase gene, *orfA*, in the IS200/605 family of transposable elements. They lack ITRs, and both element types occur commonly in viruses and conjugative plasmids of the *Sulfolobales* (18, 40) (Table 3). Exceptionally, strain REY15A and HVE10/4 genomes carry 11 and 16 nearly identical copies of the single *orfB* elements in unconserved genomic positions, respectively. This is consistent with these being the most active transposable elements in each genome (Table 3), although it remains uncertain whether they are autonomous or require an OrfA in *trans* for mobility (16). In addition, the *orfB* elements are exceptionally adaptable, because a further 8 and 2 copies are physically coupled to copies of ISC1200 for strains REY15A and HVE10/4, respectively (Table 3), and are potentially cotransposable.

***Sulfolobus* MITEs.** Only two MITE types were detected in multiple copies in each genome, SMN1 (320 bp) and SM3A (164 bp) (Table 3), and both of which are capable of non-autonomous transposition in different *S. islandicus* strains, facilitated by transposases of ISC1733 and ISC1058, respectively (2, 4, 43). All SMN1 copies are located immediately downstream from the sequence TTTAA, but none occur at conserved positions within the two genomes. Clearly, the SMN1

MITEs are active in both of the genomes, as is ISC1733, which encodes the mobilizing transposase (Table 3), and they appear to be cleanly excised when mobilized, in agreement with the results of an earlier induced excision in the *S. islandicus* strain REN1H1 (2). Although most SMN1 copies lie in intergenic regions, and may or may not affect regulatory signals, some appear to inactivate or alter genes. Thus, in strain REY15A, an AAA<sup>+</sup> ATPase (SiRe0883) and a hypothetical gene (SiRe0925) have incurred insertions in their promoters, and in strain HVE10/4, SMN1 copies partially overlap with two genes (SiH0773/2472), generating altered ORF sequences.

In contrast, the two SM3A copies are conserved in position in each genome, consistent with the mobilizing transposase encoded in ISC1058 being degenerate in both genomes. Nevertheless, each SM3A copy retains the conserved 8-bp inverted terminal repeat of the ISC1058 element (and unconserved 9-bp direct repeats resulting from the transposition event) and can potentially be mobilized if a transposase-encoding ISC1058 element enters the cell. Their maintenance as intact elements may result from one SM3A copy overlapping with the start of a conserved C/D box RNA gene (3), which may alter its transcriptional properties, while the other lies between promoters of two conserved protein genes and may influence their relative transcriptional levels. SM3A occurs in a few copies in each of the sequenced *S. islandicus* genomes, whereas SMN1 is limited to the Icelandic and three Kamchatka strains, where it occurs in 1 to 5 copies (Table 1).

**Strain-specific metabolic pathways.** Each Icelandic strain shows a few specific metabolic properties. Thus, the REY15/A strain carries an operon (SiRe0441-0445) encoding enzymes implicated in nitrate reduction and nitrite extrusion, suggesting that it can use nitrate as a terminal electron acceptor for anaerobic respiration. The operon is located in the variable region and has been observed previously only for two other

archaea, *S. islandicus* strains M.14.25 and M.16.27. The larger genome of strain HVE10/4 exclusively carries a urease operon (SiH0978-0983) predicted to encode enzymes involved in the hydrolysis of urea to  $\text{NH}_4$  and  $\text{CO}_2$  and previously found only in the archaea *Sulfolobus tokodaii*, *Metallosphaera sedula*, and *Cenarchaeum symbiosum*. Moreover, uniquely for a *Sulfolobus* species, strain HVE10/4 also carries several genes predicted to encode hydrogenases and hydrogenase maturation enzymes (SiH0883-0892) in the variable region, which suggests that the strain may be able to grow anaerobically.

A 50-kb region of strain HVE10/4 in the variable region (SiH0447-0489) is bordered by IS elements and carries 15 predicted glycosyl transferase genes (group 1 and family 2), constituting about half of the genome copies, interspersed almost exclusively with genes of unknown function and a gene encoding a predicted polysaccharide biosynthesis enzyme. It is well established that *Sulfolobus* S-layer proteins SlaA and SlaB (SiRe1612/1 and SiH1691/0, respectively) are heavily glycosylated (36), but the relatively low G+C content of the region suggests that it has been inserted and has an alternative unknown function. The genome region is absent from strain REY15A and from some of the other *S. islandicus* strains (Table 1).

**Transporters.** *Sulfolobus* strains utilize different sugars and carbohydrates as carbon and energy sources (19), consistent with their coding capacity for solute ABC transporters. A total of 15 different ABC transporters were identified, of which strain REY15A carries 12 and strain HVE10/4 contains 14. Of these, 11 ABC transporters are present in *S. solfataricus* P2 (53), 6 in *S. tokodaii* (23), but only 3 in *S. acidocaldarius* (9). The other *S. islandicus* genomes each carry 10 to 14 ABC transporters (44) (Table 1). In both of the Icelandic genomes, many ABC transporter genes are located in the variable region (Fig. 1A) and are often flanked by transposons, consistent with their being subjected to loss or gain events.

The ABC transporters are diverse, and some of their solute specificities have been identified for other *Sulfolobus* strains (15, 24). Cellobiose, maltose, and arabinose transporters are present in both of the Icelandic genomes and most other sequenced *S. solfataricus* and *S. islandicus* genomes, although a few *S. islandicus* strains lack one of the systems, as follows: the arabinose system is absent from strain YG5714, while the maltose system is not present in strains YN1551 and LD215. Strikingly, the transporter of glucose, the preferred carbon source for many microbes, is present only in the Icelandic strains, *S. islandicus* strains M1415 and YG5714, and in *S. solfataricus* P2. The lack of specific ABC transporters suggests either that glucose is an uncommon nutrient in hot environments or that another ABC transporter can facilitate glucose transport. One ABC transporter encoded in the variable region of strain HVE10/4 (SiH0899-0903), flanked by IS elements, appears to be unique in public sequence databases.

**Toxin-antitoxin systems.** Four of the eight families of anti-toxin-toxin complexes characterized for free-living bacteria also occur in archaea, of which the VapBC family is by far the most abundant (34) and is the main antitoxin-toxin family that we detected in the *Sulfolobus* strains. The Icelandic strains REY15A and HVE10/4 carry 17 and 18 *vapBC* gene pairs, respectively (Table 1), as well as 2 *vapC*-like gene copies coupled to other genes. They are distributed throughout the genomes, with several

TABLE 4. Summary of *Sulfolobus* conserved noncoding RNA genes located in the two Icelandic genomes

RNA	Function/modification	No. of indicated RNAs in genome of strain:	
		REY15A	HVE10/4
C/D box	rRNA	18	16
C/D box	tRNA	4	4
C/D box	Unknown	3	3
H/ACA box	tRNA	2	2
Noncoding	Unknown	31	27
Total		58	53

located in the variable region, and only five gene pairs are conserved in sequence and gene contexts in both strains (SiRe0698/SiH0636, SiRe2073/SiH2137, SiRe2171/SiH2227, SiRe2294/SiH2344, and SiRe2626/SiH2689). Sequence alignments and tree-building exercises demonstrated that the sequences of both antitoxins and toxins within each genome are very diverse and can be classified into subtypes (data not shown), consistent with their functional diversity and targeting of different cellular sites. These data also indicate, for given gene pairs, that the subtypes of VapB and VapC do not always correspond, implying that some gene pairs may have exchanged partners.

**Reading frame shifts and mRNA intron splicing.** Examples of translational reading frame shifts yielding single polypeptides have been demonstrated experimentally for *S. solfataricus* P2 (10). For two of these, a predicted transketolase (SiRe1696/8 and SiH1776/8) and a putative *O*-sialoglycoprotein endopeptidase (SiRe1569/70 and SiH1648/9), the *S. islandicus* genes overlap in a similar way and are likely to undergo reading frame shifts. In contrast to *S. solfataricus* P2,  $\alpha$ -fucosidase (SiRe2185 and SiH2241) is a single gene, as is the predicted dihydrolipoamide acyltransferase gene (SiH0582), located only in strain HVE10/4. Very few transposase genes present in IS elements (Table 3) carry a single reading frame shift that could be expressed as a single protein via translational reading frame shifts (28).

Transcripts of the intron-carrying *cbf5* genes (SiRe1607/8 and SiH1686/7) have been demonstrated to be spliced by the archaeal splicing enzyme at the mRNA level in some crenarchaea (60). Other mRNAs, including those encoding the XPD helicase (SiRe1685/SiH1765), have been predicted to undergo splicing, but experimental support is lacking (5).

**Noncoding RNAs.** Many untranslated RNAs have been characterized for *S. solfataricus* and *S. acidocaldarius* using a variety of techniques, including probing cell extracts for RNA with K-turn binding motifs and generating cDNA libraries of total cellular RNA extracts, as well as numerous antisense RNAs (33, 55, 59, 61). Most of these RNAs were characterized for nucleotide length and partial sequence, and several were detected by more than one experimental approach. We have reanalyzed all these different RNA entities and have annotated the *S. islandicus* RNA homologs which are conserved in both sequence and gene contexts. The total number of RNA genes and their putative functions are given (Table 4).

As for other archaeal hyperthermophiles, each genome carries many C/D box RNAs that methylate primarily rRNAs and tRNAs (Table 4). In strains REY15A and HVE10/4, 18 and 16

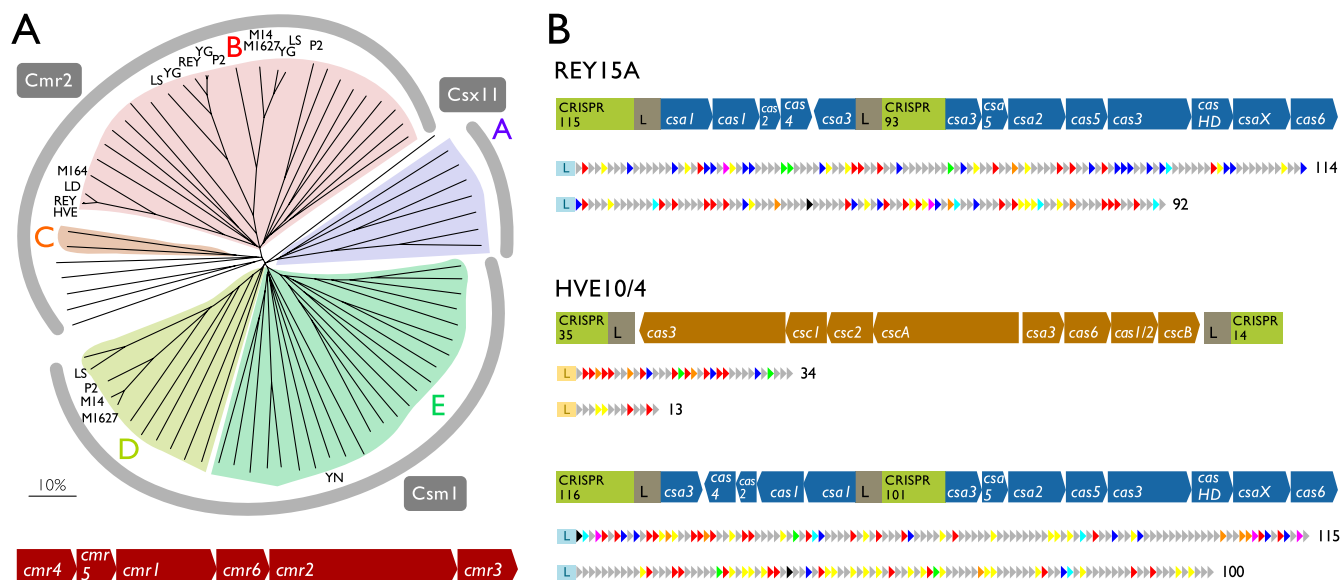


FIG. 3. (A) Phylogenetic tree of Cmr2 and its homologues in all sequenced archaeal genomes generates 5 families, A to E. The two Icelandic strains carry family B Cmr modules, for which the gene order is shown. Other sequenced *S. islandicus* and *S. solfataricus* strains also carry Cmr modules of family B and, less frequently, families D or E, as indicated on the tree. (B) Schematic representations of the CRISPR/Cas cassettes in the two Icelandic strains, together with the contents of their CRISPR loci. Strain REY15A carries a single family I CRISPR/Cas cassette (blue), whereas HVE10/4 carries cassettes from families III and I (orange and blue, respectively). Compositions of the individual CRISPR loci are shown, where each triangle represents a spacer-repeat unit. Significant spacer matches to sequenced viruses and plasmids are color coded (red, rudiviruses; orange, lipothrixviruses; yellow, fuselloviruses; green, bicaudaviruses; turquoise, turreted icosahedral viruses; blue, conjugative plasmids; and violet, cryptic plasmids).

C/D box RNAs target rRNAs, respectively, while 4 modify tRNAs and a further 3 have unknown targets. Two copies of H/ACA RNA genes are present in each genome which, together with the aPus7 protein (SiRe1836 and SiH1908), generate pseudouridine-35 in pre-tRNA<sup>Tyr</sup> transcripts (31). Each of these C/D box and H/ACA box RNA genes can be detected in the other available *S. islandicus* genomes, which underlines their functional importance. Of these, only three RNA genes characterized for other *Sulfolobus* strains, Sso-sR4, Sso-sR8, and Sso-92, were not located in any *S. islandicus* genomes (33, 55). For the numerous noncoding RNAs of unknown function, similar contents were found for the two Icelandic strains (Table 4) and for the other *S. islandicus* strains, with only a few variations (Table 1), thereby underlining their functional importance.

**Diversity of the CRISPR-based immune systems.** The CRISPR/Cas and Cmr modules all lie within the large variable regions. They show marked heterogeneity in the number and family (25, 48) and are unconserved in position between the genomes (Fig. 1A). Whereas REY15A carries one paired CRISPR/Cas module of the family I type and two family B Cmr modules, HVE10/4 contains two paired CRISPR/Cas modules of family I and III types and a single family B Cmr module (48) (Fig. 3A and B). This diversity of CRISPR-based systems also extends to the other *S. solfataricus* and *S. islandicus* genomes (Table 1). Although the gene content and organization of the paired family I CRISPR/Cas modules are quite conserved among crenarchaea (48), exceptionally, for strain HVE10/4, the internal group of *cas* genes located between the two leader regions is inverted (Fig. 3B), indicative of a rearrangement having occurred within the module, possibly via the identical inverted repeat sequences of the bordering leader regions (Fig. 3B).

The CRISPR loci of strain REY15A carry 115 and 93 spacer-repeat units centered at position 733,000, while those of HVE10/4 contain 116 and 101 repeat-spacer units and 35 and 14 repeat-spacer units centered at positions 364000 and 745000, respectively (Fig. 1A). No spacer sequence identity was detected within, or between, the two Icelandic strains or with the other *S. solfataricus* and *S. islandicus* genomes. None of the available fully sequenced *S. islandicus* genomes (Table 1) have any spacers in common, in contrast to the *S. solfataricus* strains P1, P2, and 98/2, which all share many identical spacers (17, 25) despite their being as distant from one another, phylogenetically, as the *S. islandicus* strains (Fig. 2). Thus, it seems that diversification of genomic CRISPR loci can occur either by simple spacer turnover or by horizontal transfer of whole or partial CRISPR/*cas* cassettes. There is increasing evidence for the latter mechanism being the most common one in *S. islandicus* strains (17, 21).

Since many of the characterized viruses and plasmids of *Sulfolobus* derive from Iceland, we analyzed the degree to which CRISPR spacer sequences of the Icelandic strains yielded significant matches to genetic element sequences using an earlier approach examining nucleotide and translated sequences of the spacers (25, 49). Several significant sequence matches were detected for both of the genomes, primarily to rudiviruses, fuselloviruses, and conjugative plasmids, all of which are abundant in Icelandic hot springs (63), but also were detected in smaller numbers to other viruses and cryptic plasmids (Fig. 3B).

## DISCUSSION

The genome analyses underline the potential importance of *S. islandicus* strain REY15A as a model organism for molec-



ular genetic studies of the *Sulfolobales*, and crenarchaea in general, for a variety of reasons. The genome size of 2.5 Mb is minimal for a *Sulfolobus* species; moreover, the incidence of mobile elements is relatively low (Table 1), and stable deletion mutants can be readily isolated (14, 20). Furthermore, the high incidence of diverse ABC transporter systems (Table 1) may explain why *S. islandicus* (and *S. solfataricus*) is most commonly isolated from enrichment cultures obtained from terrestrial acidic hot springs, which is in contrast to, for example, *S. acidocaldarius*, which carries only three ABC transporters (9, 44, 63).

The relatively high incidence of deletion mutants obtained from strain REY15A occurs despite the presence of several transposable elements. However, in both of the Icelandic strains, many of the IS elements are degenerate or carry disrupted transposase genes (Table 3), consistent with the “copy-and-paste” transpositional mechanism of most classes of *Sulfolobus* IS elements and their undetectably low reversibility rate (4, 41). The inability to remove the elements by spontaneous deletion, which does occur in many bacteria (16), may also explain the presence of antisense RNAs in *Sulfolobus* species to regulate transposase activity (55). The Icelandic strains do, however, carry many copies of orphan *orfB* elements and SMN1 MITEs, which are mobilized by a “cut-and-paste” mechanism presumably through *OrfA* encoded in IS element *ISC1733* (2, 16). The SMN1 MITEs appear to be specific to the Icelandic and Kamchatka strains (Table 1), and they can generate genetic novelty, reversibly, by extending open reading frames, in contrast to the other *Sulfolobus* MITEs, which carry many potential stop codons in all reading frames (43). The absence of most of the known *Sulfolobus* MITEs, except SM3A, probably reflects the much lower diversity of the mobilizing transposases present (Table 3). Many of these elements are located in the large variable region where genetic diversification occurs, including the uptake and loss of operons and gene cassettes and rearrangements of mainly nonessential genes. A similar variable genetic region in many genetic elements of *Sulfolobus* has also been observed (e.g., see reference 18).

Many questions concerning the exceptional molecular and cellular properties of crenarchaeal organisms remain to be resolved. They include the functions of the multiple and highly diverse gene pairs encoding VapBC antitoxin-toxins. For hyperthermophilic *Sulfolobus* species, in particular, their presence and variety could be a prerequisite for adaptation to life under extreme, and sometimes rapidly varying, temperature and pH conditions, as well as to survival in nutrient-poor environments possibly by optimizing the quality control of gene expression (12, 34). They may also be related to the sulfobiotics implicated in killing competitor *Sulfolobus* cells (39). The crystal structure of a VapC toxin from the crenarchaeal hyperthermophile *Pyrobaculum aerophilum* implicated the protein in exonuclease activity (1), but the multiplicity and wide sequence diversity of the *vapBC* genes suggest that the toxins target different cellular or molecular sites.

Strain HVE10/4 has been used as a host for a variety of genetic elements, mainly from Iceland, which were likely to be genetically close to the Icelandic host (63). The genome analyses provide few insights into why it is a good host, especially since it appears to carry a type 1 restriction-modification sys-

tem (SiH1435 to SiH1437). Moreover, the CRISPR/Cas and CRISPR/Cmr modules of strain HVE10/4 are relatively complex, as they also are for strain REY15A and other *Sulfolobus* strains. Their activities have also been demonstrated, at least for strain REY15A, by challenging the CRISPR/Cas systems with vector-borne matching protospacers maintained under selection, which produced deletions of the matching spacers (20). The puzzle remains as to why the *Sulfolobus* CRISPR-based systems are so complex, given that many of the viruses and plasmids coexist at low copy numbers and are nonlytic. One possibility is that the CRISPR/Cmr system primarily has a regulatory role, with antisense crRNAs (CRISPR RNAs) targeting viral mRNAs. Whatever the reason, the genetic closeness of strains REY15A and HVE10/4 suggests that the former may also be a broad host for viruses and plasmids, with the added advantage that genetic manipulation systems are now available, and our preliminary studies with fuselloviruses and conjugative plasmids support this supposition.

#### ACKNOWLEDGMENTS

This research was supported by grants from the National Natural Science Foundation of China (grants 306210165, 30730003, and 30870058) to L.H., a grant from the Danish Research Council for Technology and Production (grant 09-062932) to Q.S., and grants from the Danish Natural Science Research Council (grant 272-08-0391) and Danish National Research Foundation to R.A.G.

#### REFERENCES

- Arcus, V. L., K. Bäckbro, A. Roost, E. L. Daniel, and E. N. Baker. 2004. Distant structural homology leads to the functional characterisation of an archaeal PIN domain as an exonuclease. *J. Biol. Chem.* **279**:16471–16478.
- Berkner, S., and G. Lipps. 2007. An active nonautonomous mobile element in *Sulfolobus islandicus* REN1H1. *J. Bacteriol.* **189**:2145–2149.
- Bize, A., et al. 2009. A unique virus release mechanism in archaea. *Proc. Natl. Acad. Sci. U. S. A.* **106**:11306–11311.
- Blount, Z. D., and D. W. Grogan. 2005. New insertion sequences of *Sulfolobus*: functional properties and implications for genome evolution in hyperthermophilic archaea. *Mol. Microbiol.* **55**:312–325.
- Brügger, K., X. Peng, and R. A. Garrett. 2007. *Sulfolobus* genomes: mechanisms of rearrangement and charge, p. 95–104. In R. A. Garrett and H.-P. Klenk (ed.), *Archaea: evolution, physiology, and molecular biology*. Blackwell Publishing, Oxford, United Kingdom.
- Brügger, K., et al. 2002. Mobile elements in archaeal genomes. *FEMS Microbiol. Lett.* **206**:131–141.
- Brügger, K., E. Torarinsson, P. Redder, L. Chen, and R. A. Garrett. 2004. Shuffling of *Sulfolobus* genomes by autonomous and non-autonomous mobile elements. *Biochem. Soc. Trans.* **32**:179–183.
- Brumfield, S. K., et al. 2009. Particle assembly and ultrastructural features associated with the replication of the lytic archaeal virus *Sulfolobus* turreted icosahedral virus. *J. Virol.* **83**:5964–5970.
- Chen, L., et al. 2005. The genome of *Sulfolobus acidocaldarius*, a model organism of the Crenarchaeota. *J. Bacteriol.* **187**:4992–4999.
- Cobucci-Ponzano, B., et al. 2010. Functional characterisation and high-throughput proteomic analysis of interrupted genes in the archaeon *Sulfolobus solfataricus*. *J. Proteome Res.* **9**:2496–2507.
- Contursi, P., et al. 2006. Characterisation of the *Sulfolobus* host-SSV2 virus interaction. *Extremophiles* **10**:615–627.
- Cooper, C. R., A. J. Daugherty, S. Tachdjian, P. H. Blum, and R. M. Kelly. 2009. Role of *vapBC* toxin-antitoxin loci in the thermal stress response of *Sulfolobus solfataricus*. *Biochem. Soc. Trans.* **37**:123–126.
- Delcher, A. L., K. A. Bratke, E. C. Powers, and S. L. Salzberg. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**:673–679.
- Deng, L., H. Zhu, Z. Chen, Y. X. Liang, and Q. She. 2009. Unmarked gene deletion and host-vector system for the hyperthermophilic crenarchaeon *Sulfolobus islandicus*. *Extremophiles* **13**:735–746.
- Elferink, M. G., S. V. Albers, W. N. Konings, and A. J. Driessen. 2001. Sugar transport in *Sulfolobus solfataricus* is mediated by two families of binding protein-dependent ABC transporters. *Mol. Microbiol.* **39**:1494–1503.
- Filé, J., P. Signier, and M. Chandler. 2007. Insertion sequence diversity in archaea. *Microbiol. Mol. Biol. Rev.* **71**:121–157.
- Garrett, R. A., et al. 2011. CRISPR-based immune systems of the *Sulfolobales*: complexity and diversity. *Biochem. Soc. Trans.* **39**:51–57.



18. Greve, B., S. Jensen, K. Brügger, W. Zillig, and R. A. Garrett. 2004. Genomic comparison of archaeal conjugative plasmids from *Sulfolobus*. *Archaea* **1**:231–239.
19. Grogan, D. W. 1989. Phenotypic characterization of the archaeobacterial genus *Sulfolobus*: comparison of five wild-type strains. *J. Bacteriol.* **171**:6710–6719.
20. Gudbergstottir, S., et al. 2011. Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* **79**:35–49.
21. Held, N. L., A. Herrera, H. Cadillo-Quiroz, and R. J. Whitaker. 2010. CRISPR associated diversity within a population of *Sulfolobus islandicus*. *PLoS One* **5**:e12988.
22. Jonuscheit, M., E. Martusewitsch, K. M. Stedman, and C. Schleper. 2003. A reporter gene system for the hyperthermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector. *Mol. Microbiol.* **48**:1241–1252.
23. Kawarabayasi, Y., et al. 2001. Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain 7. *DNA Res.* **8**:123–140.
24. Koning, S. M., S. V. Albers, W. N. Konings, and A. J. Driessen. 2002. Sugar transport in (hyper)thermophilic archaea. *Res. Microbiol.* **153**:61–67.
25. Liljestøl, R. K., et al. 2009. CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. *Mol. Microbiol.* **72**:259–272.
26. Lowe, T. M., and S. R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**:955–964.
27. Lundgren, M., A. Andersson, L. Chen, P. Nilsson, and R. Bernander. 2004. Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. *Proc. Natl. Acad. Sci. U. S. A.* **101**:7046–7051.
28. Mahillon, J., and M. Chandler. 1998. Insertion sequences. *Microbiol. Mol. Biol. Rev.* **62**:725–774.
29. Marck, C., and H. Grosjean. 2003. Identification of BHB splicing motifs in intron-containing tRNAs from 18 archaea: evolutionary implications. *RNA* **9**:1516–1531.
30. Martusewitsch, E., C. W. Sensen, and C. Schleper. 2000. High spontaneous mutation rate in the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by transposable elements. *J. Bacteriol.* **182**:2574–2581.
31. Muller, S., et al. 2009. Deficiency of the tRNA<sup>Tyr</sup>:Ψ35-synthase aPus7 in Archaea of the Sulfolobales order might be rescued by the H/ACA sRNA-guided machinery. *Nucleic Acids Res.* **37**:1308–1322.
32. Muskhelishvili, G., P. Palm, and W. Zillig. 1993. SSV1-encoded site-specific recombination system in *Sulfolobus shibatae*. *Mol. Gen. Genet.* **273**:334–342.
33. Omer, A. D., M. Zago, A. Chang, and P. P. Dennis. 2006. Probing the structure and function of an archaeal C/D-box methylation guide sRNA. *RNA* **12**:1708–1720.
34. Pandey, D. P., and K. Gerdes. 2005. Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res.* **33**:966–976.
35. Peng, N., Q. Xia, Z. Chen, Y. X. Liang, and Q. She. 2009. An upstream activation element exerting differential transcription activation on an archaeal promoter. *Mol. Microbiol.* **74**:928–939.
36. Peyfoon, E., et al. 2010. The S-layer glycoprotein of the crenarchaeote *Sulfolobus acidocaldarius* is glycosylated at multiple sites with chitobiose-linked N-glycans. *Archaea* pii:754101.
37. Prangishvili, D., et al. 1998. Conjugation in archaea: frequent occurrence of conjugative plasmids in *Sulfolobus*. *Plasmid* **40**:190–202.
38. Prangishvili, D., P. P. Forterre, and R. A. Garrett. 2006. Viruses of the Archaea: a unifying view. *Nat. Rev. Microbiol.* **4**:837–848.
39. Prangishvili, D., et al. 2000. Sulfolobocins, specific proteinaceous toxins produced by strains of the extremely thermophilic archaeal genus *Sulfolobus*. *J. Bacteriol.* **182**:2985–2988.
40. Prangishvili, D., et al. 2006. Structural and genomic properties of the hyperthermophilic archaeal virus ATV with an extracellular stage of the reproductive cycle. *J. Mol. Biol.* **359**:1203–1216.
41. Redder, P., and R. A. Garrett. 2006. Mutations and rearrangements in the genome of *Sulfolobus solfataricus* P2. *J. Bacteriol.* **188**:4198–4206.
42. Redder, P., et al. 2009. Four newly isolated fuselloviruses from extreme geothermal environments reveal unusual morphologies and a possible inter-viral recombination mechanism. *Environ. Microbiol.* **11**:2849–2862.
43. Redder, P., Q. She, and R. A. Garrett. 2001. Non-autonomous elements in the crenarchaeon *Sulfolobus solfataricus*. *J. Mol. Biol.* **306**:1–6.
44. Reno, M. L., N. L. Held, C. J. Fields, P. V. Burke, and R. J. Whitaker. 2009. *Sulfolobus islandicus* pan-genome. *Proc. Natl. Acad. Sci. U. S. A.* **106**:8605–8610. (Erratum, **106**:18873.)
45. Robinson, N. P., and S. D. Bell. 2007. Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **104**:5806–5811.
46. Robinson, N. P., et al. 2004. Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell* **116**:25–38.
47. Rutherford, K., et al. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* **16**:944–945.
48. Shah, S. A., and R. A. Garrett. 2011. CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. *Res. Microbiol.* **162**:27–38.
49. Shah, S. A., N. R. Hansen, and R. A. Garrett. 2009. Distributions of CRISPR spacer matches in viruses and plasmids of crenarchaeal acidothermophiles and implications for their inhibitory mechanism. *Trans. Biochem. Soc.* **37**:23–28.
50. She, Q., et al. 2008. Host-vector systems for hyperthermophilic archaeon *Sulfolobus*, p. 151–156. In S.-J. Liu and H. L. Drake (ed.), *Microbes and the environment: perspective and challenges*. Science Press, Beijing, China.
51. She, Q., X. Peng, W. Zillig, and R. A. Garrett. 2001. Gene capture in archaeal chromosomes. *Nature* **409**:478.
52. She, Q., B. Shen, and L. Chen. 2004. Archaeal integrases and mechanisms of gene capture. *Biochem. Soc. Trans.* **22**:222–226.
53. She, Q., et al. 2001. The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. U. S. A.* **98**:7835–7840.
54. She, Q., et al. 2009. Genetic analyses in the hyperthermophilic archaeon *Sulfolobus islandicus*. *Biochem. Soc. Trans.* **37**:92–96.
55. Tang, T.-H., et al. 2005. Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.* **55**:469–481.
56. Torarinsson, E., H.-P. Klenk, and R. A. Garrett. 2005. Divergent transcriptional and translational signals in Archaea. *Environ. Microbiol.* **7**:47–54.
57. Wagner, M., et al. 2009. Expanding and understanding the genetic toolbox of the hyperthermophilic genus *Sulfolobus*. *Biochem. Soc. Trans.* **37**:97–101.
58. Worthington, P., V. Hoang, F. Perez-Pomares, and P. Blum. 2003. Targeted disruption of the alpha-amylase gene in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J. Bacteriol.* **185**:482–488.
59. Wurtzel, O., et al. 2010. A single-base resolution map of an archaeal transcriptome. *Genome Res.* **20**:133–141.
60. Yokobori, S., et al. 2009. Gain and loss of an intron in a protein-coding gene in Archaea: the case of an archaeal RNA pseudouridine synthase gene. *BMC Evol. Biol.* **9**:198.
61. Zago, M. A., P. P. Dennis, and A. D. Omer. 2005. The expanding world of small RNAs in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.* **55**:1812–1828.
62. Zhang, C., et al. 2010. Revealing the essentiality of multiple archaeal *pca* genes using a mutant propagation assay based on an improved knockout method. *Microbiology* **156**:3386–3397.
63. Zillig, W., et al. 1998. Genetic elements in the extremely thermophilic archaeon *Sulfolobus*. *Extremophiles* **2**:131–140.