Molecular Microbiology (2005) **55**(1), 312–325 doi:10.1111/j.1365-2958.2004.04391.x

New insertion sequences of *Sulfolobus*: functional properties and implications for genome evolution in hyperthermophilic archaea

Zachary D. Blount[†] and Dennis W. Grogan*
Department of Biological Sciences, University of Cincinnati, Cincinnati, OH 45221, USA.

Summary

Analyses of complete genomes indicate that insertion sequences (ISs) are abundant and widespread in hyperthermophilic archaea, but few experimental studies have measured their activities in these hosts. As a way to investigate the impact of ISs on *Sulfolobus* genomes, we identified seven transpositionally active ISs in a widely distributed *Sulfolobus* species, and measured their functional properties. Six of the seven were found to be distinct from previously described ISs of *Sulfolobus*, and one of the six could not be assigned to any known IS family. A type II 'Miniature Inverted-repeat Transposable Element' (MITE) related to one of the ISs was also recovered. Rates of transposition of the different ISs into the *pyrEF* region of their host strains varied over a 250-fold range. The *Sulfolobus* ISs also differed with respect to target-site selectivity, although several shared an apparent preference for the *pyrEF* promoter region. Despite the number of distinct ISs assayed and their molecular diversity, only one demonstrated precise excision from the chromosomal target region. The fact that this IS is the only one lacking inverted repeats and target-site duplication suggests that the observed precise excision may be promoted by the IS itself. Sequence searches revealed previously unidentified partial copies of the newly identified ISs in the *Sulfolobus tokodaii* and *Sulfolobus solfataricus* genomes. The structures of these fragmentary copies suggest several distinct molecular mechanisms which, in the absence of precise excision, inactivate ISs and gradually eliminate the defective copies from *Sulfolobus* genomes.

Accepted 13 September, 2004. *For correspondence. E-mail grogandw@email.uc.edu; Tel. (+1) 513 556 9748; Fax (+1) 513 556 5299. [†]Present address: Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI, USA.

Introduction

Transposable elements (TEs) cause dramatic changes in genes and genomes via molecular events that do not depend on the proximity or similarity of the DNA sequences affected. These events include inactivation of functional genes by insertion, activation of cryptic genes by positioning of a promoter 5′ to the coding region, deletion or inversion of DNA adjacent to the TE, and stable incorporation of DNA transferred from outside the lineage (Kleckner *et al*., 1975; Nevers and Sadler, 1977; Ciampi *et al*., 1982; Prentki *et al*., 1986). In addition to these TE-promoted changes, homologous recombination between dispersed copies of a TE can rearrange large genomic segments (Haack and Roth, 1995). These properties make TEs, which occur in nearly all organisms, a major source of genomic plasticity.

Insertion sequences (ISs) are the smallest TEs capable of independent transposition, and represent the most abundant TEs in microbial genomes. A typical IS consists of a transposase-encoding gene flanked by inverted repeats (IRs), which provide the recognition and cleavage sites for the transposase (Mahillon and Chandler, 1998). *In situ*, an IS is usually bounded by short direct repeats (DRs) that represent target-site duplications (TSDs) resulting from transposition. In prokaryotes, the transpositional mode of IS propagation often inactivates genes and *cis*-acting sequences, because of the high informational density of prokaryotic DNA. Furthermore, unlike antibiotic resistance transposons, ISs encode no extra genes that can benefit the host directly. Thus, while transposition has been associated with adaptation of bacterial lineages in certain situations (Chao *et al*., 1983), ISs are more commonly considered to be classic examples of 'selfish DNA' (Doolittle *et al*., 1984).

The sequencing of complete genomes has revealed many new and diverse ISs in a widening circle of microbial hosts; accordingly, the effect of ISs on prokaryotic genomes has become an increasingly significant question in microbiology. Analyses of the ISs found in genomes by sequencing provide valuable information on the past activity of these TEs, but quantitative assessment of IS functions remains largely confined to those that occur in bacterial hosts with well-developed genetic methods. This

reflects the technical difficulty of detecting IS transposition without the aid of specialized genetic selections. Because of this difficulty, functional characterization of new ISs has not kept pace with their discovery by genome sequencing.

Availability of quantitative genetic assays has particular importance for investigating ISs in hyperthermophilic archaea (HA). Although extensive sequencing indicates that ISs occur widely in anaerobic HA (DiRuggiero *et al*., 2000; Brügger *et al*., 2002), the genetic techniques needed to analyse transposition experimentally are most feasible in the aerobic, acidophilic HA of the genus *Sulfolobus*. Fortunately, complete genome sequences reveal an abundance and diversity of ISs in *Sulfolobus solfataricus* strain P2 and *Sulfolobus tokodaii* (She *et al*., 2001; Kawarabayasi *et al*., 2001; Brügger *et al*., 2002). Indeed, *S. solfataricus* strain P2 contains far more copies of ISs than any other prokaryote yet sequenced (Chandler and Mahillon, 2002; Brügger *et al*., 2004). Analyses of these genomic sequences have revealed clustering and complex interdigitation of *Sulfolobus* ISs, as well as small elements related to full-length ISs (Redder *et al*., 2001; Brügger *et al*., 2002). These results, combined with evidence of extensive IS-catalysed rearrangements in *Sulfolobus* genomes (Brügger *et al*., 2004), argue that ISs should play roles in HA similar to the roles played in bacteria. On the other hand, ISs are also subject to the genetic processes of their hosts, and a number of these processes may differ between HA and mesophilic bacteria. HA grow optimally at temperatures that destabilize the structure of DNA, for example (Stetter, 1996); they are also phylogenetically distant from the bacteria and eucarya in which transposition mechanisms have been analysed (Woese *et al*., 1990), and certain of their DNA-metabolizing enzymes exhibit novel biochemical properties (Lasken *et al*., 1996; Belova *et al*., 2001; Lipps *et al*., 2003). As a result, basic questions regarding the functional diversity of ISs and the evolutionary significance of different IS-catalysed events cannot be answered by analogy to mesophilic bacteria, but need to be addressed directly by experimentation based on quantitative genetic assays.

For example, of the dozens of ISs identified by sequencing in the *S. solfataricus* P2 genome (She *et al*., 2001), only four have been observed experimentally to transpose (Martusewitsch *et al*., 2000), and IS transposition in *S. tokodaii* has yet to be reported. Does the complement of active ISs vary among *Sulfolobus* species, or among isolates of the same species? Second, the four ISs observed to transpose in *S. solfataricus* inserted into different sites within a particular chromosomal region (Martusewitsch *et al*., 2000). Does this reflect different target-site specificities among different *Sulfolobus* ISs, or, alternatively, a general lack of specificity? Third, numerous examples of 'Miniature Inverted-repeat Transposable Elements' (MITEs), which have IRs corresponding to known ISs but no transposase gene, have been identified in the *S. solfataricus* and *S. tokodaii* genomes (Redder *et al*., 2001; Brügger *et al*., 2002; Brügger *et al*., 2004). The large numbers of these 'defective' ISs, and their sequence contexts, suggest repeated transposition over evolutionary time scales. Is transposition of MITEs in *Sulfolobus* genomes sufficiently frequent to be documented experimentally? Fourth, the *S. solfataricus* and *S. tokodaii* genomes also contain many other partial copies of ISs distinct from these MITEs (Brügger *et al*., 2002). Do these genomes harbour additional families of IS fragments? What do the structures of these fragments reveal about the processes that remove ISs from *Sulfolobus* genomes?

In order to expand the range of ISs available for addressing such questions, we isolated *Sulfolobus* strains from geographically diverse populations and screened them for active TEs. Here we report the recovery, structural features and functional properties of seven functional ISs, including the smallest and largest yet found in *Sulfolobus* spp., and one type II MITE. The results reveal considerable evolutionary and functional diversity of TEs in this genus of HA, and provide the first experimental measurements of the frequencies and molecular characteristics of certain IS-mediated events in archaea growing at extremely high temperature.

Results

Recovery of active ISs from natural Sulfolobus *populations*

Heterotrophic thermoacidophiles were isolated from Yellowstone National Park and Lassen Volcanic National Park in the western USA, and the Kamchatka peninsula of eastern Siberia (Whitaker *et al*., 2003). Although the clonally purified strains examined in this study exhibited a range of growth phenotypes, they appeared, by molecular criteria, to be conspecific. For example, nucleotide sequences representing the 5′ end of the *pyrB* gene, the *pyrEF* promoter, the *pyrE* gene or the *pyrF* gene from several representative isolates all yielded pair-wise nucleotide identities of about 99%. This nucleotide divergence lies well below that observed within most prokaryotic species as typically defined (Palys *et al*., 1997). In addition, a 1400 nt interval encompassing this region was found to be highly divergent from the corresponding chromosomal regions of *S. tokodaii* (55.2% nt identity) and *Sulfolobus acidocaldarius* (55.5% nt identity), less divergent from *S. solfataricus* strain P2 (90.9% nt identity), and nearly identical to Icelandic isolate REN1H1 (98.9% nt identity over the *pyrE* gene) (Zillig *et al*., 1994). These relationships agree with the extensive multilocus sequence typing of 75 similar isolates drawn from this same collection (Whitaker

Table 1. Insertion sequences recovered by genetic selection.[a]

Designation	Length (bp)	G+C[b]	IR (bp)	TSD (bp)	ORFs	Family	Closest relative, host	Transposase, % identity/similarity[c]
ISC735	735	41%	18	8	1	IS6	ISt847, *S. tokodaii*	36/55
ISC796	796	43%	21	8	1	IS1	ISt796, *S. tokodaii*	88/94
ISC1057	1057	41%	8+6[d]	8	1	IS5	ISC1058, *S. solfataricus* P2	83/90
ISC1058b	1058	39%	8+6[d]	8	1	IS5	ISC1058, *S. solfataricus* P2	72/82
ISC1205	1204–1211	45–46%	17–20	4–7	1	None	ISC1217, *S. solfataricus* P2	32/48
ISC1290[e]	1279–1288	43%	34	5	1	IS5	ISC1290, *S. solfataricus* P2	92/93
ISC1926	1926	41%	0	0	2	IS200/IS605	ISC1913, *S. solfataricus* P2	90/95

a. Table incorporates data from multiple isoforms of the indicated ISs.
b. For comparison, the target region of the host chromosome is 34.3 mol% G+C.
c. Compared with closest known, full-length relative (listed in previous column).
d. The compound IRs had the following hyphenated structure: (TSD)-octanucleotide IR-hexanucleotide spacer-hexanucleotide IR-(IS interior). Also, ISC1057 isoforms lacked the final nt at their 3′ ends complementary to the first nt of the first IR.
e. The various isoforms observed in the present study were considered equivalent to ISC1290 of *S. solfataricus.*
IR, inverted repeat; TSD, target-site duplication.

et al., 2003). The multilocus analyses indicated that the corresponding *Sulfolobus* isolates from these regions are conspecific with those isolated by Zillig and co-workers from Iceland and informally designated '*Sulfolobus islandicus*' by them (Zillig *et al*., 1994).

In order to screen for transpositionally active ISs, liquid cultures were plated on growth medium containing 5-fluoro-orotic acid (FOA) plus uracil to select spontaneous loss-of-function mutants of the *pyrE* and *pyrF* genes or their common promoter. The corresponding region of the chromosome of the mutants was then amplified by polymerase chain reaction (PCR) to detect enlarged loci. This is the approach used by Martusewitsch *et al*. (2000) for *S. solfataricus* strain P1, and represents an example of the more general 'gene trap' strategy for recovering ISs (Gay *et al*., 1985; Solenberg and Bergett, 1989; Szeverenyi *et al*., 1996; Beck *et al*., 2002). Enlarged loci were common among the mutants; most were 0.8 to about 2 kb larger than the native gene, although a few cases of enlargement by less than 0.2 kb were observed.

Molecular features of insertion sequences

Sequencing of enlarged PCR products revealed seven distinct ISs of differing lengths and structural features, which are described in Table 1. Sequence analysis permitted identification of putative open reading frames (ORFs), and BLAST searches using predicted transposase primary sequences enabled most of the new ISs to be assigned to families that have been identified previously (Mahillon and Chandler, 1998). The molecular features of the seven ISs may be described as follows.

ISC735. ISC735 represents the smallest IS yet identified in *Sulfolobus*. BLAST searches revealed no significant similarity to any catalogued nucleotide sequences other than two small fragments of 77 and 25 nt in the *S. solfataricus* genome. A single putative ORF occupies 88% of the element's length and is predicted to encode a transposase of 214 amino acids. A region between aa41 and aa208 was found to encode a conserved COG3316 protein domain (Fig. 1). This domain is characteristic of transposases related to IS240-A, a member of the IS6 family. Phylogenetic analysis of transposases accordingly placed ISC735 into this family (Fig. 2).

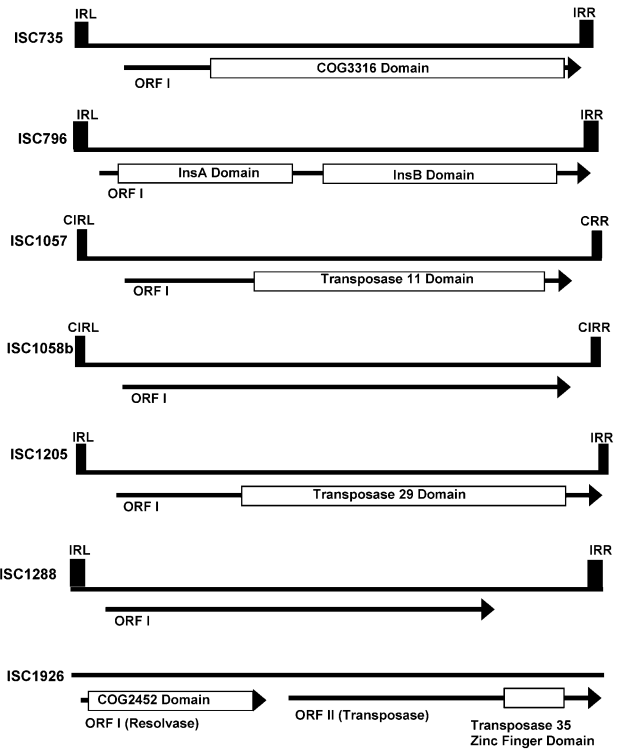

Fig. 1. Molecular organization of new insertion sequences. Diagrams (not drawn to scale) illustrate the major sequence features of the ISs analysed in this study. IRR, IRL, simple inverted repeats; CIRR, CIRL, compound inverted repeats (see Table 1); ORF, open reading frame, with direction of transcription shown by arrow.

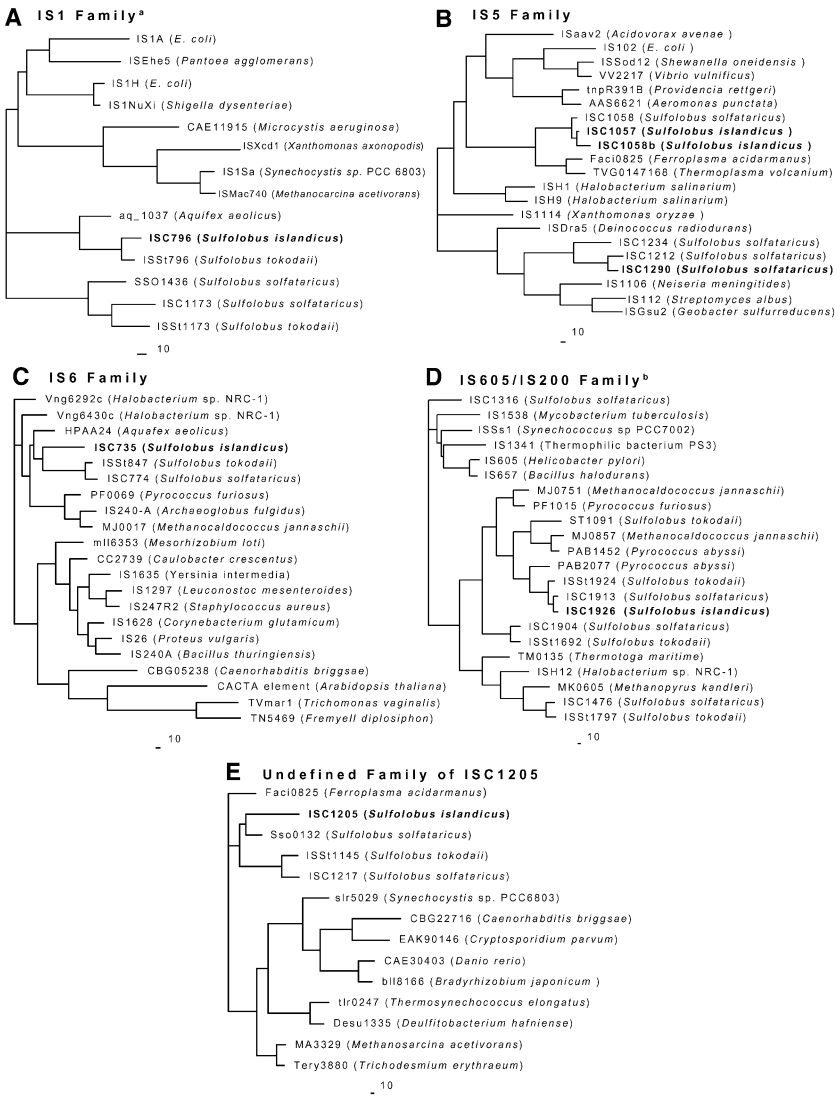

Fig. 2. Phylogenetic analysis of new insertion sequences. Maximum parsimony phylograms of the IS families represented by the IS discovered in the course of this study (shown in bold). The first species in which an IS was found is shown in parentheses. Taxa lacking IS or Tn designations represent elements known only from putative transposase sequences. Trees were constructed from T-Coffee alignments of transposase primary sequences. Non-HA IS1 elements, designated by subscript (a), were represented in data sets by InsB primary sequences; IS605/IS200 data sets, designated by subscript (b), were constructed using only transposase sequences.

ISC796. ISC796 shows 84% identity at the nucleotide level to the ISSt796 element present in five complete copies in the *S. tokodaii* genome. A similarly close relationship was also detected to an IS present in several fragmented copies in the *S. solfataricus* genome. ISC796 possesses a single putative ORF of 734 bp that occupies 92% of the element. Archaeal and bacterial promoters were detected upstream of this ORF, with the latter predicted to be very strong. The putative transpose is 244 amino acids long and exhibits high similarity to conserved protein domains characteristic of the IS1 family. Most known IS1 elements possess two ORFs: one encodes a InsA domain with a DNA binding function, while the other encodes a InsB catalytic domain, and programmed translational frameshifting permits the two ORFs to be expressed as a single, functional transposase (Mahillon and Chandler, 1998). In contrast, IS1 elements identified previously in HA (such as ISSt796) possess a single ORF encoding both domains. The predicted transposase of ISC796 is consistent with this latter structure (Fig. 1). The sequence between aa9 and aa97 showed alignment scores of 92% to the IS1 pfam03811/InsA domain. Similarly, the sequence between aa113 and aa231 displayed alignment scores of 95.9% with the COG1662 conserved domain corresponding to the IS1 InsB domain. This latter region of ISC796 also showed the presence of an IS1-type DDE domain (Asp-Asp-Glu) identical to that of ISSt796 (Ohta *et al*., 2002).

ISC1057 and ISC1058b. Depending on the particular isoform, ISC1057 shows 84–86% nucleotide identity to ISC1058, an element of the IS5 family present in 14 complete copies in the *S. solfataricus* P2 genome. ISC1057 encodes a single predicted ORF constituting 85% of the element's length, and preceded by a moderate to strong archaeal promoter and a presumptive ribosomal binding

site. The putative transposase encoded is 299 amino acids long, with a conserved Transposase 11 domain over the last two-thirds of the primary sequence (Fig. 1). This domain corresponds to the catalytic domain of the DDE transposase/integrase superfamily. In spite of their similarities, ISC1057 and ISC1058 differ significantly in their IR structure. ISC1057 has compound IRs with asymmetric terminal IRs of eight and seven bases, followed by conserved six-base spacer sequences and a second set of six-base IRs. In contrast, ISC1058 displays simple IRs of 19 bases.

ISC1058b is a relative of ISC1057 that displays essentially the same basic characteristics in genetic organization and IR structure. While the two are closely related, we treated them separately in the present analysis. This decision was based on the 88–93% nucleotide identity among the known examples and significant differences in the primary structures of their putative transposases (Fig. 1).

ISC1205. The nucleotide sequence of ISC1205 displayed no significant relationship to any known IS, although several fragmented copies of a close relative were identified in both *S. solfataricus* and *S. tokodaii* (discussed below). ISC1205 encodes a single ORF that occupies 90% of its length and is preceded by a presumptive ribosome binding site and a putatively strong promoter. The predicted 366-amino-acid transposase exhibits a Transposase 29 archaeal transposase conserved domain stretching from aa95 to aa340 (Fig. 1), shared by the transposase of ISC1217 from *S. solfataricus*. The highest protein similarity based on the ISC1205 transposase (75% identity and 86% similarity over 131 aa by BLASTP) came from *S. solfataricus* P2, ORF SSO0132. This ORF is encoded by one of the partial copies of a close relative of ISC1205 detected by BLASTN (Fig. 2). Much lower-scoring matches were found to the transposases of ISSt1145 (32% identity and 53% similarity over 279 aa) of *S. tokodaii*, and ISC1217 (32% identity and 48% similarity over 370 aa) (Fig. 2). These distantly related ISs do not belong to any of the IS families defined previously (Mahillon and Chandler, 1998). Thus, although we did identify several unclassified ISs related to ISC1205 (Fig. 2), we did not assign it to an IS family.

ISC1290. A 1288 nt IS recovered by screening showed 96% nucleotide identity to ISC1290, an element present in four partial or complete copies in *S. solfataricus*. This high degree of similarity led us to treat this IS as an isoform of ISC1290 (see *Experimental procedures*). The transposase of this isoform displayed lower levels of relationship to those of ISC1234 and ISSt1319. Although all of these ISs belong to an archaeal subtree of the IS5 family (Fig. 2), no IS5-related conserved domains were detected in the transposase sequence (Fig. 1).

ISC1926. ISC1926 is the longest IS yet found in *Sulfolobus*; it is also the only one we recovered that lacks terminal IRs and does not generate TSDs. Its nucleotide sequence shows close relationship to ISC1913 of *S. solfataricus* (90% identity) and ISSt1924 from *S. tokodaii* (78% identity). Analysis predicted two similarly oriented ORFs at nt42 to nt707 and nt785 to nt1918; both are preceded by putative ribosome binding sites, although no promoters were detected. ORF I encodes a putative resolvase of 221 amino acids with a COG2452 DNA Integrase/Resovase conserved domain yielding an alignment score of 100% from aa12 to aa210. This domain is characteristic of the highly conserved resolvases of the IS605/200 family. ORF II encodes a predicted transposase of 378 amino acids, displaying a Transposase 35 domain (a zinc finger domain with a DNA binding function) that showed an alignment score of 100% from aa260 to aa332 (Fig. 1). The presence of a putative resolvase gene indicates that ISC1926 may transpose by a replicative mechanism (Mahillon and Chandler, 1998).

Non-IS insertions

In the course of screening spontaneous mutations by PCR, mutant loci were sometimes observed to be enlarged only slightly. In most of these cases, sequencing revealed tandem duplications of the native sequence; similar duplications are relatively common among spontaneous *pyrE* mutations in *S. acidocaldarius* (Grogan *et al*., 2001). In one case, however, sequencing revealed an insertion of 138 bp near the 3′ end of the *pyrE* gene, which included a 10 bp duplication of the site of insertion. The 128 bp segment within this TSD had ends very similar to those of ISC1057 and ISC1058 (Fig. 3A), but the rest of the insertion exhibited only low similarity to other portions of these ISs (data not shown). This organization, in which the ends correspond to those of a known IS, but the central region does not, is a defining characteristic of type II MITEs (Oosumi *et al*., 1996).

Using the 128 bp sequence, BLASTN identified two short sequences of the *S. solfataricus* P2 genome that are nearly identical to each other and marginally similar to the 128 nt MITE (*P*-values of about 10^{-6}); these were later determined to be two copies of SM3A (Redder *et al*., 2001). No other prokaryotic genomes, including those of *S. tokodaii* and *Aeropyrum pernix*, yielded BLASTN matches. The first 11 bp of the two *S. solfataricus* sequences matched the 128 nt MITE exactly and represent the region of most extensive identity to it (data not shown). The two *S. solfataricus* sequences exhibited only low similarity to ISC1057 and ISC1058b.

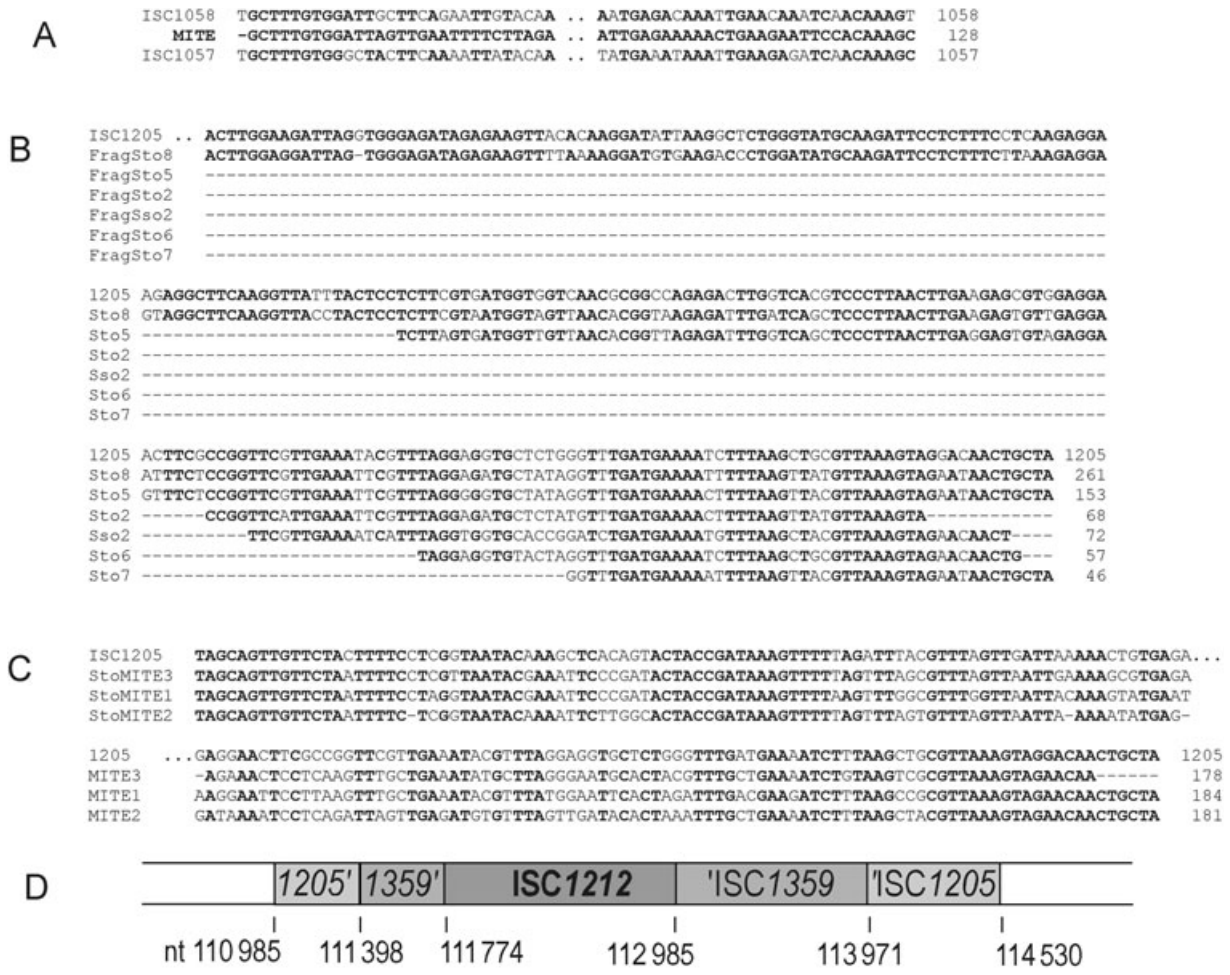

Fig. 3. IS-related sequences.
A. MITE recovered as a *pyrE* insertion. The ends of the element are shown aligned with the ends of two full-length ISs: ISC1058, identified by Martusewitsch *et al.* (2000) in *S. solfataricus* P1, and ISC1057, found in the same host strain as the indicated MITE. Nucleotides in boldface type are those that match the MITE; ellipses indicate the central regions omitted from the figure. Numerals to the right indicate the position of the last nt shown.
B. ISC1205-related fragments in sequenced genomes. Five sequences found in *S. tokodaii* (designated Sto2, 5, 6, 7 and 8) and one in *S. solfataricus* P2 (designated Sso2) are shown aligned with the 3′ end of ISC1205. Nucleotides common to all the sequences are shown in boldface. Other notations are as for (A).
C. Putative MITEs of *S. tokodaii* related to ISC1205. Sequences of the three MITEs are shown aligned to the corresponding sequences (ends) of ISC1205. Other notations are as for (B).
D. Site of multiple IS transpositions in the genome of *S. solfataricus* P2. Numerals in boxes identify corresponding intervals of ISC1205-, ISC1359- and ISC1212-like ISs; tick marks indicate sites of interruption. Boundaries between the corresponding insertion sequences are located by nucleotide positions within the *S. solfataricus* P2 genome, shown below the map.

Rates of transposition

Arguably the most important functional property of a TE is its ability to transpose to a new region of DNA. To our knowledge, however, rates of transposition have not been measured for any TE of HA. In order to quantify the rates of transposition of the new *Sulfolobus* ISs, we first measured rates of mutation to FOA resistance in the corresponding host strains, using small-scale fluctuation tests. These rates represent the sum of all spontaneous mutations that inactivate either the *pyrE* gene, the *pyrF* gene or their common promoter (Martusewitsch *et al.*, 2000; Thia-Toong *et al.*, 2002). Using PCR, we then determined the proportion of these loss-of-function mutations that resulted from IS transposition into these loci. The rates that we measured (Table 2) ranged from about 1×10^{-8} to about 2.6×10^{-6} insertions per cell division and were often characteristic of the IS. ISC796, for example, transposed at relatively low frequencies in all tests, whereas ISC1058b exhibited uniformly high frequencies. The data also suggest variation among host strains, however, as seen by the fact that ISC1205 transposed at low rates in two of the strains evaluated but at high rates in the other three (Table 2). A genetic consequence of the variation in

Table 2. Rates of transposition.

Isolate[a]	ISC	Total rate[b]	Mutants analysed	Insertions found	Proportion	Transposition rate[b]
K00 8-41	735	8.1	25	11	0.440	3.6
Y01 88′-13	796	4.2	18	1	0.056	0.2
Y01 90′-18	796	2.6	18	1	0.056	0.1
K00 3-19	1057	26	27	24	0.889	23
K00 3-8	1057	4.5	24	21	0.875	3.9
Y00 51′-90	1057	1.3	12	8	0.667	0.9
Y00 58-73	1057	21	15	7	0.467	9.7
Y99 9-16	1058b	37	23	16	0.696	26
K00 12-8	1205	25	4	2	0.500	13
K00 16-1	1205	25	6	2	0.333	8.5
L00 14	1205	8.1	11	1	0.091	0.7
L00 24	1205	8.4	19	8	0.421	3.5
L00 4	1205	5.5	19	1	0.053	0.3
Y00 16-37	1290	9.4	19	12	0.632	6
L00 11	1926	6.5	15	3	0.200	1

a. Strain in which the fluctuation tests and mutant analyses were performed (see *Experimental procedures*).
b. Rates are expressed as the number of mutational events per 10^7 cell divisions.

rates is that gene inactivation (spontaneous mutation) was dominated by IS transposition in some isolates but not in others.

Target-site selectivity

Another functional property affecting the impact of TEs on a host genome is the specificity that determines the site of an insertion. To evaluate this property, we selected a number of independent transpositions into the mutational target region of about 1250 bp (defined by the *pyrEF* promoter and the two structural genes) and identified the sites of insertion. The locations of the independent insertions (Table 3) showed that the *Sulfolobus* ISs differ both quantitatively and qualitatively with respect to the site specificity of their transposition. ISC735 was the most specific; only one target site was used in a total of six independent transposition events analysed by sequencing. ISC1205 and ISC1057 exhibited lower selectivity, using certain sites repeatedly but not exclusively.

In addition to exhibiting differing levels of sequence specificity, these ISs also exhibited differences in the nucleotide sequence of the preferred target. Except for the closely related ISC1057 and ISC1058b, target sites used by the different ISs were all distinct, so that no insertion site was used by two unrelated ISs (Table 3). In addition, alignment of the different sites of insertion did not reveal any obvious sequence motifs common to the various ISs (Table 3). However, the data did suggest that certain regions of DNA attracted transposition that was not spe-

Table 3. Target-site specificity of transposition.

Isolate	Mutant(s)[a]	ISC	Position[b]	Flanking host sequences[c]
Y99 9-16	1.4	1057	736	TGATTTAAACGTTGA GGAAATTATCGTTGA
	2.2	1057	964	TATATTAAGAAACGT TATAAGAGAGATAAG
	5.3	MITE	469	GAAAAATTAGGAGTC AAATTACACTCTTTA
	5.4	1058b	−17	TATTTAAATTCTTTT TCACAGACTCTCTAC
Y00 51′-90	2.5	1058b	−20	GTATATTTAAATTCT TTTTCACAGACTCTC
	4.2	1057	−20	GTATATTTAAATTCT TTTTCACAGACTCTC
	1.3	1057	−16	ATTTAAATTCTTTTT CACAGACTCTCTACG
	2.2, 3.4	1057	964	TATATTAAGAAACGT TATAAGAGAGATAAG
Y00 16-37	6.5	1290	23	TCGCAGAAGTATTAC TCGAAAGGAAATTAT
	4.5	1290	583	AGGCGTTAAAGGATC TCTAGATGAATTAAA
	3.3	1290	732	TAAATGATTTAAACG TTGAGGAAATTATCG
K00 8-41	3.1, 4.3	735	−14	TTTAAATTCTTTTCA CAGACTCTCTACGTA
	5.1, 6.1			
K00 12-8	6.2	1205	143	ATATAGTTAATCAAG CTATAAAGAAGGTAA
K00 16-1	5.2, 5	1205	149	TAATCAAGCTATAAA GAAGGTAAAAGATAT
L00 24	5.2	1205	113	GACCTTTACCAAATT ATCCAGAATTTTACG

a. First numeral indicates the culture plated; second numeral indicates the FOA-resistant colony analysed. Multiple mutants per entry indicate repeated insertion into the same site.
b. Location of the insertion relative to the first nt of the *pyrE* coding sequence. Thus, the TATA box for *pyrEF* is nt −32 to −25, and positions beyond nt587 lie in the *pyrF* gene.
c. The 15 nt immediately 5′ and 3′ of the insertion point respectively.

Table 4. Frequency of reversion by precise excision.

ISC	Mutant background	Location	Cultures plated[a]	Aggregate cfu[b]	Aggregate frequency[c]	P_0[d]	Reversion rate[e]
735	K00 8-41	Promoter	15	5.66	<2	>0.933	<1.5
735	K00 8-41	Promoter	15	5.68	<2	>0.928	<1.5
800	Y00 16-37	3′ end *pyrE*	10	2.11	<5	>0.900	<3.5
800	Y00 16-37	5′ end *pyrE*	8	1.52	<7	>0.875	<5
1057	K00 3-8	Promoter	13	6.53	<2	>0.923	<1.5
1057	K00 3-8	5′ end *pyrE*	15	5.47	<2	>0.929	<1.5
1205	K00 12-8	Promoter	8	3.02	<3	>0.875	<3
1205	K00 12-8	5′ end *pyrE*	9	1.92	<5	>0.875	<4
1205	L00 38	Promoter	12	3.17	<3	>0.917	<3
1205	L00 38	5′ end *pyrE*	13	5.88	<2	>0.923	<1.5
1290	Y00 15.1-14	5′ end *pyrE*	9	3.46	<3	>0.889	<2.5
1926	L00 11	3′ end *pyrE*	8	1.34	300	0.75	11.9

a. Number of independent cultures plated for phenotypic reversion.
b. Aggregate number of cells from all cultures in the previous column, in units of 10^9 cfu.
c. Aggregate revertant frequency (ratio of revertants to total number of cells), in units of 10^{-10} per cfu.
d. Proportion of independent cultures yielding no revertants.
e. Mutation frequency calculated from P_0 and the average number of cells in the *Sulfolobus* cultures, in units of 10^{-10} per cfu.

cific to a particular nucleotide sequence. For example, various target sites used by ISC1057 and ISC1058b tended to cluster in an interval 15–20 nt ahead of the *pyrE* coding sequence, into which ISC735 also inserted (Table 3). By comparison to the homologous region of the *S. acidocaldarius* chromosome (Thia-Toong *et al*., 2002), this interval is about halfway between the *pyrEF* promoter (boxA) and the mRNA start site.

Precise excision

In bacteria, the various processes that remove TEs are broadly classified as either 'precise' or 'imprecise' excision. Precise excision, which includes all events that delete the IS and one equivalent of any TSD, is not necessarily more frequent than imprecise excision, but it has important consequences for the host because it restores the sequence, and therefore function, of the interrupted gene. Precise excision accordingly represents the only mechanism that completely reverses an insertion mutation in one step. As non-essential genes inactivated by TEs can be nevertheless beneficial to the host, correspondingly strong selective pressure can exist for precise excision of these insertions which does not apply to imprecise excision. Precise excision also differs mechanistically from imprecise excision in being mediated in most cases by host functions rather than TE-encoded functions (Eigner and Berg, 1981; Lundblad *et al*., 1984).

Despite this special genetic significance, precise excision had not been investigated, to our knowledge, in any of the HA. We therefore assayed for the precise excision of several ISs by selecting phenotypic revertants of corresponding *pyrE* insertion mutants in fluctuation tests. As summarized in Table 4, the observed rates of precise excision were below detection for all but ISC1926. This result was unexpected, given the molecular diversity of the ISs being evaluated (Figs 1 and 2; Table 1) and the potential of these assays to detect reversion rates below 5×10^{-10} per cell division. The result could not be attributed to unfavourable plating conditions: (i) the same cultures yielded colonies on uracil-supplemented medium with high efficiency of plating under these conditions, (ii) selective plates were incubated twice as long as necessary to yield visible colonies on the same medium supplemented with uracil, (iii) the revertant colonies grew to a readily visible size (>1 mm) under these conditions and (iv) cross-feeding caused faint haloes to develop around the revertant colonies in the later stages of incubation, demonstrating that the pyrimidine-starvation conditions were not lethal for the cell population over this period of time. The lack of phenotypic revertants could also not be attributed to a special property of one IS or one insertion site, because, with the sole exception of ISC735, at least two distinct insertions of each of these ISs were evaluated for reversion.

Related sequences in other genomes

In the course of analysing the new ISs, we found that several of them had extensive similarity to sequences in *Sulfolobus* genomes. In *S. tokodaii*, for example, BLASTN with the ISC796 sequence identified five complete copies of a closely related IS, plus at least five partial copies; several partial copies were similarly found in *S. solfataricus*. BLASTN with the ISC1205 sequence revealed only partial copies in both genomes. BLASTN with ISC735 revealed two small fragments in *S. solfataricus* but no related sequences in *S. tokodaii*. These various copies had not been annotated as IS or IS fragments in either genome (Kawarabayasi *et al*., 2001; She *et al*., 2001).

We analysed the ISC1205-related sequences in greater detail in order to gain insight into the molecular processes that form partial IS copies. Eight such sequences were found in *S. tokodaii*, and two in *S. solfataricus*; together, these 10 partial copies represent three classes of mutated IS. The first and largest class (five copies in *S. tokodaii* and one in *S. solfataricus*) consists of varying lengths of the 3′ end of the element (as defined by the orientation of its transposase; Fig. 3B). The sequences seem consistent with a single large deletion (or multiple overlapping smaller deletions) extending past the 5′ end of each of six ancestral IS copies of the ISC1205 relative. The postulated deletion events appear to have been independent, as each partial copy exhibits a different end-point. In addition, the various copies have numerous substitutions with respect to each other, and three have small deletions of the extreme 3′ end.

The second class of partial IS includes the remaining three copies in *S. tokodaii*, which can be explained by three similar but distinct deletions of the central portion of the ISC1205 relative, leaving about 90 nucleotides intact at each end (Fig. 3C). This is the structure of type I MITEs, which can be represented as internal deletion mutants of functional ISs (Oosumi *et al*., 1996). Examples of other type I MITEs have been found in HA (Redder *et al*., 2001). The type I MITEs found in *S. tokodaii* are slightly divergent from each other (79–86% sequence identity) and from ISC1205, but the alignment (Fig. 3C) suggests a common deletion event relating all three to an ISC1205-like ancestor. It should be noted that, while the *S. tokodaii* genome contains no complete IS whose sequences match those of the MITEs shown in Fig. 3C, it does have a complete copy of ISSt1145, a distant relative of ISC1205 with low overall sequence similarity but very similar IRs. This raises the possibility that these MITEs may remain active in *S. tokodaii* via 'heterologous' complementation, i.e. complementation by a rather distantly related IS. It is also interesting to note that Redder *et al*. (2001) independently identified a number of type II MITEs (designated 'SM2 elements') in *S. solfataricus* that seem to be derived from ISC1217, the closest known functional relative of ISC1205.

The third class of ISC1205 fragment is represented by a single segment in the *S. solfataricus* genome. This sequence has about 90% nt identity to ISC1205, and displays a deletion of 309 bp near the centre of the element. In addition, it is interrupted by about 2600 nt of DNA, segments of which display homology to the putative ISs ISC1212 and ISC1359 found elsewhere in the P2 genome (Brügger *et al*., 2002). Examination of the boundaries of these various segments indicates an order of events leading to this complex structure (Fig. 3D). The ISC1205 relative, for example, is flanked by DRs of 6 bp, consistent with the TSDs observed for ISC1205. This indicates that the ISC1205 relative transposed into this region of the chromosome before the centre of its transposase gene was deleted. Similarly, the ISC1359 segment is flanked by DRs of 4 bp of the ISC1205 ORF, implying that ISC1359 transposed into the ISC1205 relative in its current location. Finally, the copy of ISC1212 is intact and flanked by 6 bp DRs of ISC1359, indicating that ISC1212 was the last to transpose, thereby inactivating the ISC1359 copy.

Discussion

IS diversity in Sulfolobus *spp.*

Despite the extreme conditions of their habitats and relative isolation of individual populations (Whitaker *et al*., 2003), HA have proven to be a surprisingly rich source of transposable genetic elements (Brügger *et al*., 2002). For example, despite identification of at least 25 ISs in the genome of *S. solfataricus* strain P2 (She *et al*., 2001), a completely distinct IS was discovered fortuitously in another *S. solfataricus* strain cultivated from the same geothermal area (Ammendola *et al*., 1998). Results like these suggest that *Sulfolobus* populations harbour an extensive diversity of ISs and that many more of these TEs remain to be discovered by sequencing and genetic screening of additional *Sulfolobus* isolates. Our experimental results reinforce this idea, and more than double the number of ISs demonstrated experimentally to transpose in HA. Of seven ISs recovered from one widely distributed *Sulfolobus* species, we considered only one to be an isoform of a previously identified IS. Furthermore, one of the remaining six, ISC1205, could not be assigned to any recognized IS family, although it exhibited distant relationships to several ISs that also have not been assigned to any family (Fig. 2). These homologues, all identified by BLAST, include ISC1217 of *S. solfataricus* (32% transposase identity/48% similarity over 369 aa), ISSt1145 of *S. tokodaii* (33%/54% over 282 aa), and hypothetical ISs in *Ferroplasma acidarmanus* (28%/45% over 169 aa), *Bradyrhizobium* (30%/52% over 100 aa), a *Desulfitobacterium* sp. (27%/46% over 129 aa) and *Caenorhabditis briggsae* (25%/49% over 136 aa). These data suggest that ISC1205 belongs to a yet unnamed IS family with diverse representatives in all three domains of life.

Do different Sulfolobus *species, or conspecific isolates, have different complements of active ISs?*

Nucleotide sequences of the *pyrBEF* region suggest that the host strains examined in this study belong to a *Sulfolobus* species related to *S. solfataricus* but distinct from it. The pattern of transposition that we observed among these isolates thus provides an informative comparison to

the pattern reported for *S. solfataricus* strain P1 (Martusewitsch *et al*., 2000). For example, only one of the ISs observed to transpose in *S. solfataricus* P1 (ISC1058) is closely related to any of the ISs we observed to transpose. This demonstrates that closely related *Sulfolobus* species can differ greatly with respect to their complement of active ISs. Furthermore, we observed differences with respect to the active IS complement among different strains of the same species. For example, aside from ISC1057 and ISC1058b, which are close relatives of each other, we observed very few cases in which more than one IS was observed to transpose in a given strain, despite examining numerous isolates and as many as 10 independent transposition events per isolate. This contrasts dramatically with *S. solfataricus*, in which a single isolate (strain P1) yielded four unrelated ISs in only seven transposition events (Martusewitsch *et al*., 2000). Although our results could, in principle, be explained by a lower activity of ISs than in *S. solfataricus*, we note that several of the ISs were capable of frequent transposition. We similarly doubt that many of the isolates in our study contain only one IS in an intact form. Although this cannot ruled out without exhaustive analysis of each isolate, we note that Southern blotting detected multiple copies of ISC796 that were conserved among strains from the Lassen population, and that IS-specific PCR detected full-length copies of at least two different ISs in about 90% of more than 100 isolates screened from all three geographical regions (Z.D. Blount, unpubl.).

Do Sulfolobus *ISs differ with respect to target-site selectivity and other functional properties?*

In addition to exhibiting different transposition frequencies, the ISs were also diverse with respect to qualitative properties such as the degree of target-site specificity. Within the *pyrEF* region, ISC735 used only one site in six independent transpositions, whereas the closely related ISC1057 and ISC1058b showed a moderate preference with regard to sequence. The latter two ISs also exhibited a preference for inserting into particular intervals of the chromosomal target, despite variation in the specific position. One of the more prominent of these hot-spots for insertion lies halfway between the TATA box and the transcription start site predicted by transcript analysis of the homologous region in *S. acidocaldarius* (Thia-Toong *et al*., 2002). Interestingly, the sequence used exclusively by ISC735 also occurs in this interval, and the homologous interval in *S. solfataricus* strain P1 serves as the preferred insertion site for ISC1058 (Martusewitsch *et al*., 2000).

This pattern resembles the tendency of many bacterial ISs to insert in or near promoters (Mahillon and Chandler, 1998) and cannot be explained readily by some property of the FOA selection in *Sulfolobus* spp. that enables insertions in or near the promoter region to support faster growth than insertions in other parts of the *pyrE–pyrF* interval. In particular, as the basis of the selection is loss of function, insertions into the promoter region would not be favoured by virtue of a partial phenotype. This is confirmed experimentally by the fact that in *S. acidocaldarius*, leaky *pyrE* alleles are at a growth disadvantage in the selection, and are not recovered under conditions of higher stringency (Grogan *et al*., 2001). Alternatively, there is no obvious mechanism to explain how insertion mutations before the coding sequences, whether they disrupt the promoter itself or exert polarity on the *pyrF* gene, would result in less enzymatic activity than insertions into the *pyrE* or *pyrF* coding sequences. The insertional preference observed in this region of *Sulfolobus* genomes may thus represent a functional similarity of IS transposition in HA and mesophilic bacteria involving the increased accessibility or reactivity of certain DNA regions. The high level of target-site specificity we observed for ISC735 would seem to make it unique within the IS6 family, however, as no other members of this family have displayed such specificity (Chandler and Mahillon, 2002).

Do MITEs retain the ability to transpose in Sulfolobus *genomes?*

MITEs are extremely short repetitive sequences with structures corresponding to 'ends-only' TEs. They are found in all three domains, and can be extremely abundant in eukaryotic genomes (Brügger *et al*., 2002; Yang and Hall, 2003). MITEs can also be abundant in HA genomes, as demonstrated by *S. solfataricus* strain P2 (Redder *et al*., 2001; Brügger *et al*., 2004). To our knowledge, all identifications of MITEs have been based on genome sequences, and none have been confirmed experimentally to transpose. Our observation of a MITE insertion event in *Sulfolobus* thus provides the first experimental demonstration of MITE transposition, and suggests that such transposition can be sufficiently frequent to affect genome evolution in *Sulfolobus*. As the host strain has an active copy of a cognate IS (ISC1057), the observed insertion is consistent with the hypothesis that MITEs are mobilized by transposase produced from complete ISs having the corresponding IRs. In addition, the MITE we recovered by transposition has a type II structure in which IRs are separated by intervening sequence lacking homology to any known IS. Simple internal deletion of a functional IS therefore cannot readily explain this class of MITEs (Oosumi *et al*., 1996), and it is not yet clear how these structures originate. It should also be noted that similar 'gene trap' screening of new *S. solfataricus* isolates has revealed transposition of another, unrelated type II MITE (Z.D. Blount, unpubl. results), suggesting consider-

able diversity and activity of these TEs in *Sulfolobus* genomes. The present study also detected new, putative type I MITEs in the genome of *S. tokodaii* based solely on their relatedness to ISC796 and ISC1205 (see below).

Precise excision of ISC1926

The available data seem generally consistent with the hypothesis that much of the precise excision of TEs in bacteria reflects spontaneous deletion events promoted by the short IRs and DRs at the TE boundaries (Eigner and Berg, 1981; Glickman and Ripley, 1984; Schaaper *et al*., 1986; Perkins-Balding *et al*., 1999). It is thus significant that the only insertion mutation observed to excise precisely in our study has no IRs or DRs to facilitate such deletion, whereas all the ISs with such repeats, including an extensive (34 bp) IR in the case of the ISC1290 isoform, failed to excise at detectable frequencies. These results imply that short, non-tandem DRs, including those associated with IRs, do not promote frequent deletion in this *Sulfolobus* species. A similar conclusion was supported by a fundamentally different analysis in *S. acidocaldarius* (Grogan and Hansen, 2003), thus reinforcing the idea that *Sulfolobus* spp., and perhaps other HA, are deficient in one or more 'pathways' of spontaneous deletion that predominate in bacteria (Glickman and Ripley, 1984; Schaaper *et al*., 1986). It remains difficult to define distinct bacterial pathways of spontaneous deletion in terms of specific genes required, and genetic manipulation of the process in HA remains daunting in technical terms. However, in bacteria and yeast it has been possible to identify host mutations that accelerate the precise excision of TEs, which implicates these genes as suppressors of DR-dependent, RecA-independent pathways of deletion. The host genes include *ssb*, *polA*, *topA*, *MutSLH*, *dam*, *uvrD*, *uup*, and special alleles of *recBC* in *Escherichia coli*, which mediate a range of DNA transactions, and *POL1* and *POL3* in yeast, which encode the DNA polymerases that synthesize the lagging strand (Lundblad *et al*., 1984; Gordenin *et al*., 1992; Reddy and Gowrishankar, 2000). It must be noted, however, that tandem DRs are deleted efficiently in *S. acidocaldarius*, implying that the separation of DRs attenuates deletion between them (Grogan and Hansen, 2003), and that such an effect has been documented in other microbial systems, as well (Scearce *et al*., 1991; Chédin *et al*., 1994).

Comparing phenotypic reversion of *pyrE*::ISC1926 with spontaneous deletion formation in the *S. acidocaldarius pyrE* gene reveals a quantitative discrepancy, however, in that the former was about 10% as frequent as the latter (1×10^{-9} versus 1×10^{-8} per cell respectively). If deletion end-points are determined approximately randomly (as appears to be the case in *S. acidocaldarius*), only a very small fraction of all spontaneous deletion events within *pyrE* (on the order of one in 3.6×10^5 deletions) would effect precise excision of any given insertion. Thus, the phenotypic reversion of a *pyrE*::ISC1926 mutation is not predicted to occur at an observable frequency by random, spontaneous deletion of the type documented in *S. acidocaldarius*. One possible explanation for our results therefore is that ISC1926 carries out or assists its own excision, a situation with precedent among bacterial TEs (Shen *et al*., 1987). We note, for example, that self-catalysed precise excision at about 1% of the rate of transposition into the target gene has been observed with transposon Tn*10* in *E. coli* (Shen *et al*., 1987), and a similar ratio would be sufficient to explain our results. In any case, the molecular basis of ISC1926 transposition and precise excision warrants further study, particularly as TEs that have no IRs and generate no TSDs have not been studied extensively.

What do the structures of fragments reveal concerning IS removal?

In contrast to ISC1926, all other *Sulfolobus* ISs that we tested exhibited no phenotypic reversion of a *pyrE* insertion, despite strong selection. This provides experimental evidence that gene inactivation by these ISs, which are more typical of *Sulfolobus* ISs than ISC1926 in terms of structural features (Martusewitsch *et al*., 2000; Brügger *et al*., 2002), is largely irreversible. This has important implications for genome evolution. For example, although it does not imply permanence of the corresponding ISs, it does restrict their elimination to mechanisms such as (i) imprecise excision promoted by the IS transposase (i.e. abortive transposition) (Shen *et al*., 1987) and (ii) spontaneous mutation of the types occurring generally in the host genome. To the extent that mechanisms (i) leave significant portions of the IS in the genome, or are not excessively more frequent than (ii), *Sulfolobus* genomes should contain segments of inactivated ISs that are in the process of accumulating spontaneous mutations.

Sequence analyses provide several lines of molecular evidence for incremental modes of IS removal from *Sulfolobus* genomes. For example, we found four cases in which an IS related to those we have discovered seems to have been successfully eliminated from a *Sulfolobus* genome: (i) an ISC796 relative and (ii) an ISC735 relative eliminated from *S. solfataricus*, and ISC1205 relatives eliminated from (iii) *S. solfataricus* and (iv) *S. tokodaii*. These cases are defined by the genomes having multiple fragments of the IS but no full-length copy; in all cases, the differences among the fragments suggest independent deletion events. Other cases, in which the full-length IS nevertheless remains in the genome, have been documented in genome annotations (Kawarabayasi *et al*., 2001; She *et al*., 2001). Brügger *et al*. (2002) noted the

abundance of these partial IS copies and postulated the existence of a mechanism that specifically inactivates ISs when they become too numerous in a genome. Our experimental analysis of precise excision suggests an alternative explanation for the observed abundance of these partial copies. Without precise excision, the complete eradication of an IS copy is forced to be a slow, multistep process. In this situation, proliferation of an IS in a genome and the inevitable inactivation of individual copies by small mutations produces partially degraded intermediates in the genome, and the abundance of these relics reflects the relative rates of their production versus removal. Although comprehensive testing of either hypothesis is beyond the scope of the present study, our results supplement the available genomic sequence data by providing new examples of partial IS copies for such analyses.

Finally, in qualitative terms, our examination of the ISC1205-related fragments in sequenced genomes reveals at least two molecular processes by which IS may be inactivated and partially removed. One is a mechanism of removal, typified by most of the ISC1205 fragments in *S. tokodaii*, that leaves only the element's 3′ end. It will be of interest to determine whether this mode of deletion occurs with other IS and other *Sulfolobus* spp. The other is a mechanism of inactivation, illustrated by the nearly full-length relic in *S. solfataricus* P2, representing insertion of other genetic elements into the IS. Similar relics apparently generated by this inactivation mechanism occur elsewhere in the P2 genome (Brügger *et al*., 2002), and remain fully consistent with the non-essential nature of ISs and the fact that spontaneous mutation in *S. solfataricus* is dominated by IS transposition (Martusewitsch *et al*., 2000).

Experimental procedures

Strains and growth conditions

Samples of water and sediment were collected from a number of acidic geothermal springs in the Norris Geyser Basin, Crater Hills, and Geyser Creek areas of Yellowstone National Park in northwestern Wyoming (USA), Devil's Kitchen and Bumpass Hell areas of Lassen Volcanic National Park in northern California (USA), and Uzon Caldera, Geyser Valley, and Mutnovsky Volcano areas of the Kamchatka peninsula (Russia), as previously described (Whitaker *et al*., 2003). Samples were plated directly (i.e. without enrichment) on dextrin-tryptone medium and incubated for 8–15 days at 78°C. Individual colonies were cultured in liquid medium, clonally purified on plates, catalogued, and preserved at −70°C using techniques similar to those previously described for *S. acidocaldarius* (Grogan and Gunsalus, 1993). Isolate designations (e.g. 'Y00 51′-90') incorporate the following sampling and isolation data: region (Yellowstone, Lassen or Kamchatka) and year, sample number (hyphen) clone number.

Unless otherwise noted, media and growth conditions were identical to those used for *S. acidocaldarius*, except that 0.2% D-xylose was replaced by 0.2% Dextrin-10 (Fluka). Spontaneous *pyr* mutants were selected by spreading aliquots of liquid cultures on plates containing 150 g of FOA and 20 g of uracil per ml of medium. Fluctuation tests were conducted by inoculating sets of at least five tubes, each containing 3 ml of liquid medium, with one isolated colony per tube of the isolate or mutant under investigation. When the cultures reached a density of about 10^8 cfu ml^{-1}, the cells were plated on selective media. For assays of transposition, the selective plates contained uracil and FOA, and rates of total spontaneous mutation were calculated from the distribution of the number mutant colonies, as previously described (Jacobs and Grogan, 1997). The fraction of mutants represented by insertions was determined by PCR screening of mutants from the fluctuation tests. For assays of phenotypic reversion, the cells harvested from liquid cultures were washed in sterile dilution buffer and plated on medium containing glutamine and acid-hydrolysed casein. Because most cultures in the reversion assays yielded no mutants, rates were estimated by the P_0 method of Lea and Coulson (1949).

DNA analyses

DNA was extracted from mutant cultures using the guanidinium thiocyanate procedure of Pitcher *et al*. (1989). PCR was then used to amplify the *pyrE* and *pyrF* loci, their common promoter region, and the 3′ end of the *pyrB* gene. Primers for these amplifications were based on the *S. solfataricus pyr* operon sequence (Martusewitsch *et al*., 2000). To amplify *pyrE* and the promoter region-*pyrB* end only, the primers SsoINTER1for (5′-CGAATATTCTAAAGTAGTCATCTCTGG-3′) and SsopyrE1rev (5′-CGGGATCCATTGCTAATATTACT CTAG-3′) were used. Amplification of the entire *pyrBEF* region required SsoINTER1for and Sso*pyrF*1rev (5′-TTCC TCGTGTAGATTTTCCCC). Reaction mixtures contained dNTPs, Taq DNA polymerase and buffer (Promega or New England Biolabs), ≈50 ng of genomic DNA, as well as the appropriate primers. Temperature cycling (2 min at 94°C, followed by 25 cycles of 1 min at 94°C, 1 min at 57°C and 1 min at 70°C) was carried out in an MJ Research. PTC-100 programmable thermocycler, and products were run on 1% agarose gels.

A second PCR series with nested primers was then used with those mutants displaying enlarged loci to roughly localize the insertion site within the target region. The primers used were as follows: promoter region: SsoINTER1for and SsoINTER2rev (5′-ACTAACCTTACCTGATGTTAAAACG-3′); first third of *pyrE*: SsopyrE2for (5′-GAAGATCTCTACGTAT GAATTTCGC-3′) and PyrEmid-2-rev (5′-CCATAGGCTCTT TAAGGTTACAAGC-3′); middle third of *pyrE*: PyrEmid-1-for (5′-GCTTGTAACCTTAAAGAGCCTATGG-3′) and PyrEmid-4-rev (5′-TGCGTCTGAAACTTTACCTCC-5′); last third of *pyrE*: PyrEmid-3-for (5′-TCCATATGAGAAAGCAACATTGG-3′) and SsopyrE1rev (5′-CGGGATCCATTGCTAATATTACTCTAG-3′). For nucleotide sequencing, the portion of the target containing the insertion was amplified, and the amplification product purified from the primers and salts and transferred to water using Millipore Microcon® YM-100 filters. DNA sequencing was performed by the Cincinnati Children's Hospital

Sequencing Facility, using ABI PRISM dye-terminator reagents. Text files were edited for miscalls by visual inspection of electropherograms.

Based on the common identity of their sequences, an interval of 1400 nt was compiled from several of the isolates. This sequence included a portion of the *pyrB* gene, the bidirectional promoter, the *pyrE* gene and the *pyrF* gene. Because of its much closer relationship to the individual sequences analysed, the 1400 nt sequence compiled from experimental isolates was used in the present study in place of the *S. solfataricus* P2 sequence as the reference for analysis of insertions and other mutations.

Assembled IS DNA sequences were used in BLAST searches (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) to identify similar IS in the NCBI database, as well as fragments in sequenced genomes. Putative ORFs were identified using the ORF finder in the NEBCUTTER program (http://tools.neb.com/NEBcutter). Putative protein sequences were then used in BLAST searches to identify related transposase sequences, conserved domains and family assignment. Putative protein characteristics were determined using the resources of the ExPASy site (http://us.expasy.org/). Putative transposase amino acid sequences were aligned with those of related ISs identified from BLAST searches or contained in the IS Finder database (http://www-is.biotoul.fr) using T-Coffee (Notredame *et al*., 2000). Phylogenetic reconstruction from aligned amino acid sequences was performed with PAUP* 4.0 for Macintosh using maximum parsimony analysis (Swofford, 2003).

Sequence designations and accession numbers

In the present study, an 'insertion sequence' was considered to encompass all observed sequence variants exhibiting more than 90% nt identity in the transposase-encoding region; conversely, variants with 90% or less identity were given distinct designations. Naming followed a convention similar to that of Martusewitsch *et al*. (2000): 'ISC' is followed by the length, in nucleotides, of the first isoform documented at the sequence level. Nucleotide sequences for each of the seven ISs described in the study have been deposited in GenBank under Accession No. AY671942 to AY671948.

Acknowledgements

We thank R. Whitaker for collecting hot spring samples from Lassen Volcanic National Park and the Mutnovsky Volcano region of Kamchatka. We thank G. Bell and J. Hansen for help with the isolation and preservation of strains, L. Hoffman and J. Holmes for PCR analysis of mutants in fluctuation tests, and C. Borland for comments on the manuscript. This work was supported by Grant MCB 9733303 from the National Science Foundation

References

Ammendola, S., Politi, L., and Scandurra, R. (1998) Cloning and sequencing of ISC1041 from the archaeon *Sulfolobus solfataricus* MT-4, a new member of the IS30 family of insertion elements. *FEBS Lett* **428:** 217–223.

Beck, P., Dingermann, T., and Winckler, T. (2002) Transfer RNA gene-targeted retrotransposition of *Dictyostelium* Tre5-a into a chromosomal UMP synthase gene trap. *J Mol Biol* **318:** 273–285.

Belova, G.I., Prasad, R., Kozyavkin, S.A., Lake, J.A., Wilson, S.H., and Slesarev, A.I. (2001) A type IB topoisomerase with DNA repair activities. *Proc Natl Acad Sci USA* **98:** 6015–6020.

Brügger, K., Redder, P., She, Q., Confalonieri, F., Zivanovic, Y., and Garrett, R.A. (2002) Mobile genetic elements in archaeal genomes. *FEMS Microbiol Lett* **206:** 131–141.

Brügger, K., Torarinsson, E., Redder, P., Chen, L., and Garrett, R.A. (2004) Shuffling of *Sulfolobus* genomes by autonomous and non-autonomous mobile elements. *Biochem Soc Trans* **32:** 179–183.

Chandler, M., and Mahillon, J. (2002) Insertion sequences revisited. In *Mobile DNA II*. Craig, N., Craigie, R., Gellert, M., and Lamhowitz, A. (eds). Washington, DC: American Society for Microbiology Press, pp. 305–366.

Chao, L., Vargas, C., Spear, B.B., and Cox, E.C. (1983) Transposable elements as mutator genes in evolution. *Nature* **303:** 633–635.

Chédin, F., Dervyn, E., Dervyn, R., Ehrlich, S.D., and Noirot, P. (1994) Frequency of deletion formation decreases exponentially with distance between short direct repeats. *Mol Microbiol* **12:** 561–569.

Ciampi, M., Schmid, M., and Roth, J. (1982) Transposon Tn*10* provides a promoter for transcription of adjacent sequences. *Proc Natl Acad Sci USA* **79:** 5016–5020.

DiRuggiero, J., Dunn, D., Maeder, D.L., Holley-Shanks, R., Chatard, J., Horlacher, R., *et al.* (2000) Evidence of recent lateral gene transfer among hyperthermophilic archaea. *Mol Microbiol* **38:** 684–693.

Doolittle, W.F., Kirkwood, T.B.L., and Dempster, M.A.H. (1984) Selfish DNAs with self-restraint. *Nature* **307:** 501–502.

Eigner, C., and Berg, D.E. (1981) Excision of transpson Tn*5* is dependent on the inverted repeats but not on the transposase function of Tn*5*. *Proc Natl Acad Sci USA* **78:** 459–463.

Gay, P., Coq, W., Steinmetz, M., Berkelman, T., and Kado, C. (1985) Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J Bacteriol* **164:** 918–921.

Glickman, B.W., and Ripley, L.S. (1984) Structural intermediates of deletion mutagensis: a role for palindromic DNA. *Proc Natl Acad Sci USA* **81:** 512–516.

Gordenin, D.A., Malkova, A.L., Peterzen, A., Kulikov, V.N., Pavlov, Y.I., Perkins, E., and Resnick, M.A. (1992) Transposon Tn*5* excision in yeast: influence of DNA polymerases alpha, delta, and epsilon and repair genes. *Proc Natl Acad Sci USA* **89:** 3785–3789.

Grogan, D.W., and Gunsalus, R.P. (1993) *Sulfolobus acidocaldarius* synthesizes UMP via a standard *de novo* pathway: results of a biochemical-genetic study. *J Bacteriol* **175:** 1500–1507.

Grogan, D.W., and Hansen, J.E. (2003) Molecular characteristics of spontaneous deletions in the hyperthermophilic archaeon *Sulfolobus acidocaldarius*. *J Bacteriol* **185:** 1266–1272.

Grogan, D.W., Carver, G., and Drake, J.W. (2001) Genetic

fidelity under harsh conditions: analysis of spontaneous mutation in the thermophilic archaeon *Sulfolobus acidocaldarius*. *Proc Natl Acad Sci USA* **98:** 7928–7933.

Haack, K., and Roth, J. (1995) Recombination between chromosomal IS200 elements supports frequent duplication formation in *Salmonella typhimurium*. *Genetics* **141:** 1245–1252.

Jacobs, K.L., and Grogan, D.W. (1997) Rates of spontaneous mutation in an archaeon from geothermal environments. *J Bacteriol* **179:** 3298–3303.

Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., *et al.* (2001) Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain 7. *DNA Res* **8:** 123–140.

Kleckner, N., Chan, R.K., Tye, B.-K., and Botstein, D. (1975) Mutagenesis by insertion of a drug-resistance element carrying an inverted repetition. *J Mol Biol* **97:** 561–575.

Lasken, R.S., Schuster, D.M., and Rashtchian, A. (1996) Archaebacterial DNA polymerases tightly bind uracil-containing DNA. *J Biol Chem* **271:** 17692–17696.

Lea, D.E., and Coulson, C.A. (1949) The distribution of the numbers of mutants in bacterial populations. *J Genet* **49:** 264–285.

Lipps, G., Rother, S., Hart, C., and Strauss, G. (2003) A novel type of replicative enzyme harbouring ATPase, primase, and DNA polymerase activity. *EMBO J* **22:** 2516–2525.

Lundblad, V., Taylor, A.F., Smith, G.R., and Kleckner, N. (1984) Unusual alleles of *recB* and *recC* stimulate excision of inverted-repeat transposons Tn*10* and Tn*5*. *Proc Natl Acad Sci USA* **81:** 824–828.

Mahillon, J., and Chandler, M. (1998) Insertion sequences. *Microbiol Mol Biol Rev* **62:** 725–774.

Martusewitsch, E., Sensen, C., and Schleper, C. (2000) High spontaneous mutation rate in the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by transposable elements. *J Bacteriol* **182:** 2574–2581.

Nevers, P., and Sadler, H. (1977) Transposable genetic elements as agents of gene instability and chromosomal rearrangements. *Nature* **268:** 109–115.

Notredame, C., Higgins, D., and Heringa, J. (2000) T-Coffee: a novel method for multiple sequence alignments. *J Mol Biol* **203:** 205–217.

Ohta, S., Tsuchida, K., Choi, S., Sekine, Y., Shiga, Y., and Ohtsubo, E. (2002) Presence of a characteristic D–D–E motif in IS1 transposase. *J Bacteriol* **184:** 6146–6154.

Oosumi, T., Garlick, B., and Belknap, W.R. (1996) Identification of putative nonautonomous transposable elements associated with several transposon families in *Caenorhabditis elegans*. *J Mol Evol* **43:** 11–18.

Palys, T., Nakamura, L.K., and Cohan, F.M. (1997) Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *Internat J Syst Bacteriol* **47:** 1145–1156.

Perkins-Balding, D., Duval-Valentin, G., and Glasgow, A.C. (1999) Excision of IS492 requires flanking target sequences and results in circle formation in *Pseudomonas atlantica*. *J Bacteriol* **181:** 4937–4948.

Pitcher, D., Saunders, N., and Owen, R. (1989) Rapid extraction of bacterial genomic DNA with guanidinium thiocyanate. *Lett Appl Microbiol* **8:** 151–156.

Prentki, P., Teter, B., Chandler, M., and Galas, D. (1986) Functional promoters created by the insertion of transposable element IS1. *J Mol Biol* **191:** 383–393.

Redder, P., She, Q., and Garrett, R.A. (2001) Non-autonomous mobile elements in the crenarchaeon *Sulfolobus solfataricus*. *J Mol Biol* **306:** 1–6.

Reddy, M., and Gowrishankar, J. (2000) Characterization of the *uup* locus and its role in transposon excisions and tandem repeat deletions in *Escherichia coli*. *J Bacteriol* **182:** 1978–1986.

Scearce, L.M., Pierce, J.C., McInroy, B., and Masker, W. (1991) Deletion mutagenesis independent of recombination in bacteriophage T7. *J Bacteriol* **173:** 869–878.

Schaaper, R.M., Danforth, B.N., and Glickman, B.W. (1986) Mechanisms of spontaneous mutagenesis: an analysis of the spectrum of spontaneous mutation in the *Escherichia coli lacI* gene. *J Mol Biol* **189:** 273–284.

She, Q., Singh, R., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M., *et al.* (2001) Complete genome of the crenarchaeote *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci USA* **98:** 7835–7840.

Shen, M.M., Raleigh, E.A., and Kleckner, N. (1987) Physical analysis of Tn*10*- and IS10-promoted transpositions and rearrangements. *Genetics* **116:** 359–369.

Solenberg, P., and Bergett, S. (1989) Method for selection of transposable DNA and characterization of a new insertion sequence, IS493, from *Streptomyces lividans*. *J Bacteriol* **171:** 4807–4813.

Stetter, K.O. (1996) Hyperthermophilic procaryotes. *FEMS Microbiol Rev* **18:** 149–158.

Swofford, D.L. (2003) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.* Sunderland, MA: Sinauer Associates.

Szeverenyi, I., Hodel, A., Arber, W., and Olasz, F. (1996) Vector for IS element entrapment and functional characterization based on turning on expression of distal promoterless genes. *Gene* **174:** 103–110.

Thia-Toong, T.-L., Roovers, M., Durbecq, V., Gigot, D., Glansdorff, N., and Charlier, D. (2002) Genes of *de novo* pyrimidine biosynthesis from the hyperthermophilic crenarchaeote *Sulfolobus acidocaldarius*: novel organization in a bipolar operon. *J Bacteriol* **184:** 4430–4441.

Whitaker, R.J., Grogan, D.W., and Taylor, J.W. (2003) Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301:** 976–978.

Woese, C.R., Kandler, O., and Wheelis, M. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* **87:** 4576–4579.

Yang, G.J., and Hall, T.C. (2003) MAK, a computational tool kit for automated MITE analysis. *Nucleic Acids Res* **31:** 3659–3665.

Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janekovic, D., Hain, J., *et al.* (1994) Screening for Sulfolobales, their plasmids and their viruses in Icelandic solfataras. *Syst Appl Microbiol* **16:** 609–628.