J Appl Genet 50(4), 2009, pp. 421–430

Original article

CRISPR elements in the Thermococcales: evidence for associated horizontal gene transfer in *Pyrococcus furiosus*

M.C. Portillo, J.M. Gonzalez

Institute for Natural Resources and Agrobiology, IRNAS-CSIC, Sevilla, Spain

Abstract. The presence and distribution of CRISPR (clustered regularly interspaced short palindrome repeat) elements in the archaeal order Thermococcales were analyzed. Four complete genome sequences from the species *Pyrococcus abyssi*, *P. furiosus*, *P. horikoshii*, and *Thermococcus kodakaraensis* were studied. A fragment of the genome of *P. furiosus* was flanked by CRISPR elements upstream and by a single element downstream. The composition of the gene sequences contained in this genome fragment (positions 699013 to 855319) showed significant differences from the other genes in the *P. furiosus* genome. Differences were observed in the GC content at the third codon positions and the frequency of codon usage between the genes located in the analyzed fragment and the other genes in the *P. furiosus* genome. These results represent the first evidence suggesting that repeated CRISPR elements can be involved in horizontal gene transfer and genomic differentiation of hyperthermophilic Archaea.

Keywords: CRISPR, gene transfer, genome, hyperthermophilic Archaea, *Pyrococcus furiosus.*

Introduction

Genomic regions consisting of tandem repeat DNA elements, typically of 21–47 base pairs (bp) in length, separated by nonrepetitive spacer sequences of approximately similar length, have been identified in about half of bacterial and most archaeal genomes (Mojica et al. 1995; Jansen et al. 2002; Godde and Bickerton 2006; Lillestol et al. 2006). Recent reports suggest that these short, regularly spaced repeat and spacer regions (CRISPRs = clustered regularly interspaced short palindrome repeats) constitute a microbial immune system acting similarly to eukaryotic RNA interference, to target and neutralize foreign DNA from phage and plasmid sources (Makarova et al. 2006; Sorek et al. 2008). Compelling evidence in support of this hypothesis has been recently presented (Barrangou et al. 2007), demonstrating that *Streptococcus thermophilus* responded to viral predation by integrating new spacers into a CRISPR system. Furthermore, addition and removal of spacers and repetitive elements modified the phage-resistance phenotype of bacteria (Barrangou et al. 2007). The incorporation of short fragments of foreign DNA into the genome of prokaryotes at CRISPR loci has been reported (Bolotin et al. 2005; Pourcel et al. 2005; Godde and Bickerton 2006). These studies only analyze the origin of spacers, and the addition or deletion of repeated elements.

Work by Godde and Bickerton (2006) has suggested that horizontal gene transfer acts upon CRISPR-associated (*cas*) genes likely involving the use of conjugation. This transfer of genetic material is carried out through phages (Pourcel et al. 2005) and megaplasmids (Godde and Bickerton 2006), suggesting the potential for transferring large fragments of DNA (i.e. >40 Kb), although the spread of antibiotic resistance genes has been suggested to be limited by CRISPR elements (Marraffini and Sontheimer 2008). A recent study (Tyson and Banfield 2008), based on comparative genomics of populations of *Leptospirillum*, pro-

Received: May 4, 2009. Accepted: July 3, 2009.
Correspondence: J.M. Gonzalez, Instituto de Recursos Naturales y Agrobiologia, IRNAS-CSIC, Avda. Reina Mercedes 10, 41012 – Sevilla, Spain; e-mail: jmgrau@irnase.csic.es

poses recent lateral transfer of CRISPR loci, followed by significant loss and gain of spacer and repeated element sequences. Diversification of CRISPR systems and their role in shaping species evolution require further investigation (DeBoy et al. 2006).

The presence of CRISPR systems in hyperthermophilic Archaea has been reported for several species (Lillestol et al. 2006; Sorek et al. 2008). Among other Archaea, CRISPRs have been found in species of *Methanocaldococcus, Sulfolobus, Pyrobaculum,* and *Pyrococcus*. At present, there are 4 complete genomes of species described within the order Thermococcales, including the genera *Pyrococcus* and *Thermococcus*: *P. abyssi*, *P. furiosus*, *P. horikoshii*, and *T. kodakaraensis*. This study analyzed the presence and distribution of CRISPR sequences in these genomes and focused on a potential horizontal gene transfer event suggested from newly found individual elements identical to CRISPR elements within the *P. furiosus* genome.

Materials and methods

The genome sequences and gene annotations from the members of the order Thermococcales were obtained from the microbial genomes database (National Center for Biotechnology Information; http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi). Four genomes were studied within this order: *Pyrococcus abyssi*, *P. furiosus*, *P. horikoshii*, and *Thermococcus kodakaraensis*. Searches for repetitive sequences throughout these genomes were performed by the software FIRES (short for Finding Repetitive Sequences) written in C by us and run in a computer under the Linux operating system. This program finds every sequence repeated over a determined number of times and over a length of a specified number of nucleotides. The repeated sequences found in these genomes, their number of repetitions, and location, were reported in the program output. (FIRES is available online at http://www.microextreme.net/software). The detected repeated sequences were compared with those reported at the CRISPR database (http://crispr.u-psud.fr/crispr/CRISPRdatabase.php) for the 4 analyzed genomes.

The possibility of transfer of genetic material associated with repeated sequences was analyzed following the procedure proposed by Garcia-Vallvé et al. (2000), based on the frequency of codon usage (FCU) and the GC content at the third codon position for the genes in the target DNA fragment and the whole genome. GC content at the third codon position and cumulative differences in FCU were compared using portions of the genomic sequence of equivalent length (in number of genes) to the targeted fragment being analyzed. FCU for a codon *j* encoding an amino acid *i* was calculated as:

$$\mathrm{FCU_j} = x_{ij} \,/\, \sum_{j=1}^{n_{ij}} x_{ij}$$

where n_{ij} represents the number of codons encoding for a single amino acid *i*.

The average FCU_{jk} for a genomic fragment containing n_k gene sequences was estimated as:

$$\mathrm{A_{FCUj}} = \sum_{k=1}^{n_k} FCU_{\mathrm{jk}} \,/\, \mathrm{n_k}$$

The cumulative difference in A_{FCUj} for all the 64 codons between the genes in genome fragment A with respect to the genes in another portion of genome B was computed as:

$$\mathrm{CD_{AB}} = \sum_{j=1}^{64} \left| A_{FCUj\cdot A} - A_{FCUj\cdot B} \right|$$

Differences between the values of A_{FCUj} for the genes in a selected genomic fragment and the values for the rest of the genome were statistically compared by a *G*-test for goodness-of-fit according to Sokal and Rohlf (1981). Comparisons between the GC content at the third codon position for the genes contained in 2 genomic fragments were compared by the Student *t*-test (Sokal and Rohlf, 1981).

Dendrograms to show the distance between the genes in the different gene classes within the *P. furiosus* genome and genes in the analyzed fragment, as well as the relationship of the genes in the analyzed fragment with the genes in the genomes of 4 species of Thermococcales, were constructed using the UPGMA algorithm (Garcia-Vallvé et al. 2000) based on the A_{FCUj} values for the genes contained in each genome or genome fragment.

Results and discussion

CRISPR-type motifs have been reported in most Bacteria and Archaea. They have been assigned a major role in controlling viral re-infections (Barrangou et al. 2007) and could be indicative of previous viral insertions (Andersson and Banfield 2008). To our knowledge, there has been no reported evidence of the potential implication of these repeats in the insertion of novel DNA fragments in microbial genomes, with the exception of

transfer events for *cas* genes (Godde and Bickerton 2006; Tyson and Banfield 2008). With this perspective in mind, the presence and distribution of CRISPR elements and identical motifs were analyzed in the Thermococcales.

The repeated elements detected in the 4 studied genomes from the Thermococcales are reported in Table 1, which represents an update from previous estimates, mainly on the detection of a couple of individual elements identical to those repeated sequences forming CRISPR. The repeated elements in the circular genomes are mainly distributed from near the beginning of the genome sequence up to 240 degrees. No CRISPR elements were detected between 240 and 358 degrees in the studied genomes. The length of these repeated sequences is up to 30 bases. The spacer regions between repeats were generally between 46 and 60 bp in length. The number of repeated elements per CRISPR structure was variable, with groups of up to 67 in *P. horikoshii* and isolated conserved sequences in *P. furiosus*. Every analyzed genome presented CRISPR elements, although their number per genome varied from 60 in *P. abyssi* to 185 in *P. furiosus*. Within a single CRISPR group of elements, these repeated elements maintained its orientation, although different CRISPR groups could be found in both strands. The repeated sequences were well conserved within a single CRISPR group. Different repeated sequences could be detected within a genome, although they showed an elevated proportion of conserved nucleotides (Table 1). Even if distinct genomes within the Thermococcales are compared, the repeated motifs are generally well conserved.

The detected repeated elements have been compared to those previously reported (CRISPR database), and some novel elements have been discovered. The presence at a couple of sites of nonrepeated elements identical to the repeated elements in CRISPR sequences was only found in the genome of *P. furiosus* (Table 1). These sequences, identical to the repeated elements found as individual elements in *P. furiosus* genome, were further investigated. Specifically, a fragment of about 156 kbp, located between positions 699013 and 855319 in the *P. furiosus* genome, presented differential characteristics when compared to the average values for the genes in this genome. This genomic fragment was surrounded by a standard series of 47 repeated CRISPR elements upstream, and downstream by a single, nonrepeated sequence identical to the previous ones in the opposite strand. The annotated genes proposed in that genome fragment are listed in Table 2.

The distinctive characteristics detected for the genes located in the analyzed fragment are shown in Figure 1. The GC content at the third codon position for the ORFs included in this fragment was 48%, while the value for the total number of ORFs

Table 1. Updated summary of CRISPR elements detected in the Thermococcales. Nucleotides different from the corresponding base of the first sequence for each species are indicated in grey.

Species	Position		Repeats	Strand	Most representative element
	start	end			
Pyrococcus abyssi	147936	149453	23	+	CTTTCAATTCTATTTTAGTCTTATTGGAAC
	482332	482850	7	+	CTTTCCACACTACTA AGTTCTACGGAAAC
	1090078	1090376	3	+	CTTTCAATTCTATTTTAGTCTTATTGGAAC
	1760079	1761876	27	–	CTTTCAATTCTATTTTAGTCTTATTGGAAC
Pyrococcus furiosus	27110	30570	52	–	CTTTCAATTCTATTTTAGTCTTATTGGAAC
	260726	262123	21	–	CTTTCAATTCTTTTGTAGTCTTATTGGAAC
	275806	276006	3	–	CTTTCCACACGATTACGTTC TACTGAAAC
	312417	313951	23	–	CTTTCAATTCTTTTGTAGTCTTATTGGAAC
	623137	625196	3	+	CTTTCAATTCTATTTTGGTCTTATTGTAAC
	695555	699012	47	–	CTTTCAATTCTATTTTAGTCTTATTGTAAC
	855320	855349	1	+	CTTTCAATTCTATTTTAGTCTTATTGGAAC
	1046763	1046791	1	+	TTTTCAATTCTATTTTAGTCTTATTCA
	1064094	1065563	22	+	CTTTCAATTCTTTTGTAGTCTTATTGGAAC
	1091109	1091879	12	+	CCTTCAATTCTTTTGTAGTCTTATTGTAAC
Pyrococcus horikoshii	32925	34102	18	+	CTTTCCACACTATT TAGTTCTACGGAAAC
	150437	152186	26	–	CTTTCCACACTACT AAGTTCTACGGAAAC
	637754	638938	18	–	CTTTCAATTCTCTTATAGTCTTATTGGAAC
	932787	938330	67	+	CTTTC ACACTATT TAGTTCTACGGAAAC
	974376	994804	7	–	GTTTCAATTCTATTTTAGTCTTATTGGAAC
	1117827	1118993	18	+	CTTTCAATTCTATTTTAGTCTTATTGGAAC
Thermococcus kodakaraensis	373052	374093	16	+	ATTTCAATTCTCTTAGAGTCTTATTGCAAC
	469066	470587	23	+	GTTTCAATTCTCTTAGAGTCTTATTGCAAC
	833513	835999	37	–	CTTTCAATTCTCTTAGAGTCTTATTGCAAC

in this genome was 39%, and the values from classes of ORFs (of similar length to the analyzed fragment) ranged from 35% to 41% (except the fragments containing the analyzed sequence). Thus, differences about 9% and lower than 4% were observed for the analyzed fragment and the other ORF classes in the *P. furiosus* genome, respectively (Figure 1). Considering these classes of ORFs, the average GC content at the third codon position in the genes of the studied fragment was 3-fold higher (Student *t*-test; $P < 0.001$) than the average value for the other ORF classes in this ge-

Table 2. List of genes described in the analyzed fragment of the *P. furiosus* genome (positions 699013 to 855319), showing gene ID, gene annotation, and the species with the closest gene homologue.

Gene ID	Gene annotation	Species of closest gene homologue
PF0688	subtilisin-like protease	*Thermococcus onnurineus*
PF0689	hypothetical protein	*Thermococcus onnurineus*
PF0690	hypothetical protein; glutaredoxin/thioredoxin-like protein	*Thermococcus onnurineus*
PF0691	hypothetical protein	*Thermococcus onnurineus*
PF0692	hydroxylamine reductase [EC:1.7.-.-]	*Thermococcus barophilus*
PF0693	hypothetical protein	*Thermococcus barophilus*
PF0694	flavoprotein	*Thermococcus barophilus*
PF0695	hypothetical protein; hypothetical protein	*Thermococcus onnurineus*
PF0696	putative carbohydrate-binding protein	*Thermococcus onnurineus*
PF0697	multidrug resistance protein	*Aeropyrum pernix*
PF0698	hypothetical protein	*Physcomitrella patens*
PF0699	hypothetical protein	none
PF0700	x-pro dipeptidase, N-terminal fragment	*Thermococcus barophilus*
PF0701	x-pro dipeptidase, N-terminal fragment	*Thermococcus barophilus*
PF0702	x-pro dipeptidase	*Thermococcus onnurineus*
PF0703	hypothetical protein	*Thermococcus onnurineus*
PF0704	hypothetical protein	*Thermococcus barophilus*
PF0705	cytochrome c-type biogenesis protein	*Thermococcus onnurineus*
PF0706	hypothetical protein	*Thermococcus onnurineus*
PF0706.1n	hypothetical protein	*Thermococcus onnurineus*
PF0707	hypothetical protein	*Methanocaldococcus fervens*
PF0708	hypothetical protein	*Thermotoga maritima*
PF0709	hypothetical protein	*Pyrobaculum arsenaticum*
PF0710	hypothetical protein	none
PF0711	hypothetical protein	*Thermococcus barophilus*
PF0712	hypothetical protein	*Thermococcus* sp.
PF0713	hypothetical protein	*Thermococcus barophilus*
PF0714	hypothetical protein	*Burkholderia phytofirmans*
PF0715	NAD(P)H oxidase	*Thermococcus onnurineus*
PF0716	3-hydroxyisobutyrate dehydrogenase [EC:1.1.1.31]	*Methanoregula boonei*
PF0717	putative methyltransferase	*Archaeoglobus fulgidus*
PF0718	hypothetical protein	Candidate Division TM7
PF0719	hypothetical protein	*Thermosipho africanus*
PF0720	hypothetical protein	*Thermotoga neapolitana*
PF0721	hypothetical protein; NADPH-dependent FMN reductase, putative	*Thermococcus onnurineus*
PF0722	alkyl hydroperoxide reductase subu C [EC:1.11.1.15]	*Thermococcus barophilus*
PF0723	hypothetical protein; high-affinity iron transporter	*Thermococcus barophilus*
PF0724	hypothetical protein	*Thermococcus onnurineus*
PF0725	hypothetical protein; CoA-binding domain-containing protein	*Thermococcus onnurineus*
PF0726	hypothetical protein	*Thermococcus onnurineus*
PF0727	hypothetical protein	*Thermococcus onnurineus*
PF0728	hypothetical protein; ubiquinone/menaquinone biosynthesis methyltransferase [EC:2.1.1.-]	*Thermococcus onnurineus*
PF0729	mulitcopper oxidase domain-containing protein	*Methanospirillum hungatei*
PF0730	hypothetical protein	*Thermococcus kodakaraensis*
PF0731	copper-transporting ATPase, p-type	*Thermococcus barophilus*
PF0731.1n	hypothetical protein	none

Table 2 cont.

Gene ID	Gene annotation	Species of closest gene homologue
PF0732	ABC-2 type transport system ATP-binding protein	*Bacillus clausii*
PF0732.2n	hypothetical protein	none
PF0733	putative ABC transporter	*Synechococcus* sp.
PF0734	hypothetical protein	none
PF0735	hypothetical protein	*Caldicellulosiruptor saccharolyticus*
PF0736	hypothetical protein	none
PF0736.1n	hypothetical protein	none
PF0737	hypothetical protein; putative membrane protein	*Thermococcus barophilus*
PF0738	ubiquinone/menaquinone biosynthesis methyltransferase	*Thermococcus onnurineus*
PF0738.1n	hypothetical protein	*Thermococcus barophilus*
PF0739	AsnC family transcriptional regulator	*Thermococcus kodakaraensis*
PF0740	heavy-metal transporting cpx-type ATPase [EC:3.6.1.-]	*Thermococcus kodakaraensis*
PF0741	hypothetical protein	*Thermococcus kodakaraensis*
PF0742	putative ferritin	*Thermococcus onnurineus*
PF0743	hypothetical protein	*Thermococcus onnurineus*
PF0744	ABC transporter; ATP-binding cassette, subfamily C	*Thermococcus kodakaraensis*
PF0745	hypothetical protein; multiple antibiotic resistance protein	*Thermococcus* sp.
PF0746	hypothetical protein; zinc transporter, ZIP family	*Thermococcus kodakaraensis*
PF0747	putative proline dipeptidase	*Thermococcus onnurineus*
PF0748	hypothetical protein	*Thermococcus onnurineus*
PF0749	hypothetical protein	*Thermococcus* sp.
PF0750	putative NADH oxidase	*Pyrococcus abyssi*
PF0751	flavoprotein	*Thermococcus kodakaraensis*
PF0752	thioredoxin peroxidase	*Thermococcus kodakaraensis*
PF0753	2-oxoglutarate ferredoxin oxidoreductase subunit beta [EC:1.2.7.3]	*Thermococcus kodakaraensis*
PF0754	2-oxoglutarate ferredoxin oxidoreductase subunit alpha [EC:1.2.7.3]	*Thermococcus kodakaraensis*
PF0755	glyceraldehyde-3-phosphate dehydrogenase (NADP) [EC:1.2.1.9]	*Thermococcus kodakaraensis*
PF0756	transposase	*Pyrococcus* sp.
PF0757	hypothetical protein	*Thermococcus* sp.
PF0758	hypothetical protein	*Pyrococcus horikoshii*
PF0759	hypothetical protein	*Thermococcus kodakaraensis*
PF0760	hypothetical protein; transposase	*Thermococcus barophilus*
PF0761	hypothetical protein	*Pyrococcus abyssi*
PF0762	hypothetical protein	*Pyrococcus abyssi*
PF0763	hypothetical protein	none
PF0764	DEXX-box ATPase	*Pyrococcus abyssi*
PF0765	NDP-sugar dehydrogenase	*Pyrococcus abyssi*
PF0766	putative dehydrogenase	*Pyrococcus abyssi*
PF0767	pleiotropic regulatory protein DegT	*Thermococcus gammatolerans*
PF0768	acetyl/acyl transferase related protein	*Pyrococcus abyssi*
PF0769	glycosyl transferase	*Thermococcus gammatolerans*
PF0770	UTP-glucose-1-phosphate uridylyltransferase [EC:2.7.7.9]	*Thermococcus gammatolerans*
PF0771	UDP-glucose 6-dehydrogenase [EC:1.1.1.22]	*Pyrococcus abyssi*
PF0772	hypothetical protein	*Pyrococcus horikoshii*
PF0773	hypothetical protein	none
PF0774	hypothetical protein	*Methanocaldococcus jannaschii*
PF0775	hypothetical protein	*Thermococcus barophilus*
PF0776	hypothetical protein	*Thermococcus barophilus*
PF0777	hypothetical protein	*Pyrococcus abyssi*
PF0776.1n	hypothetical protein	*Thermococcus* sp.
PF0778	hypothetical protein	*Archaeoglobus fulgidus*
PF0779	hypothetical protein	*Archaeoglobus fulgidus*
PF0780	transposase	*Pyrococcus* sp.
PF0781	hypothetical protein	*Thermococcus kodakaraensis*
PF0782	hypothetical protein	*Thermococcus barophilus*

Table 2 cont.

Gene ID	Gene annotation	Species of closest gene homologue
PF0783	hypothetical protein	Pyrococcus abyssi
PF0784	hypothetical protein; glycosyl transferase, group 1	*Methanosaeta thermophila*
PF0785	transposase	*Pyrococcus woesei*
PF0785.2n	hypothetical protein; glycosyl transferase, group 1	*Methanococcoides burtonii*
PF0785.3n	hypothetical protein; transposase	*Sulfurimonas denitrificans*
PF0786	transposase	*Riemerella* sp.
PF0787	hypothetical protein	*Methanococcoides burtonii*
PF0788	glycosyl transferase	*Thermococcus barophilus*
PF0789	hypothetical protein; glycosyl transferase	*Pelotomaculum thermopropionicum*
PF0790	hypothetical protein	*Methanoculeus marisnigri*
PF0791	hypothetical protein	*Thermococcus gammatolerans*
PF0792	hypothetical protein; asparagine synthase [EC:6.3.5.4]	*Thermococcus gammatolerans*
PF0793	hypothetical protein	*Methanococcus maripaludis*
PF0794	UDP-N-acetylglucosamine 2-epimerase [EC:5.1.3.14]	*Pelotomaculum thermopropionicum*
PF0795	capsular polysaccharide biosynthesis protein	*Pyrococcus abyssi*
PF0796	galactoside O-acetyltransferase	*Methanosarcina mazei*
PF0797	hypothetical protein	*Geobacter* sp.
PF0798	glycosyl transferase	*Methanococcus vannielii*
PF0799	hypothetical protein	*Pelotomaculum thermopropionicum*
PF0800	transposase	*Riemerella anatipestifer*
PF0801	hypothetical protein	none
PF0802	hypothetical protein	*Pyrococcus horikoshii*
PF0803	hypothetical protein	*Pyrococcus horikoshii*
PF0804	MoxR-like ATPase [EC:3.6.3.-]	*Pyrococcus horikoshii*
PF0805	hypothetical protein	*Pyrococcus horikoshii*
PF0806	hypothetical protein	*Pyrococcus horikoshii*
PF0807	hypothetical protein	*Pyrococcus horikoshii*
PF0808	hypothetical protein	*Pyrococcus horikoshii*
PF0809	hypothetical protein	*Pyrococcus horikoshii*
PF0810	hypothetical protein	*Pyrococcus horikoshii*
PF0811	hypothetical protein; oligosaccharyl transferase	*Thermococcus onnuriensis*
PF0812	hypothetical protein	*Pyrococcus abyssi*
PF0813	hypothetical protein	*Pyrococcus abyssi*
PF0814	hypothetical protein	*Thermococcus kodakaraensis*
PF0815	hypothetical protein	*Thermococcus onnuriensis*
PF0816	hypothetical protein	*Thermococcus kodakaraensis*
PF0817	transposase	*Pyrococcus* sp.
PF_t015	tRNA-Glu	
PF_t016	tRNA-Ile	
PF0818	aminohydrolase	*Pyrococcus abyssi*
PF0819	50S ribosomal protein L14e	*Pyrococcus abyssi*
PF0820	cytidylate kinase [EC:2.7.4.14]	*Pyrococcus horikoshii*
PF0821	rpl34e; 50S ribosomal protein L34e	*Pyrococcus horikoshii*
PF0822	hypothetical protein	*Pyrococcus horikoshii*
PF0823	hypothetical protein, sodium-driven multidrug efflux pump protein	*Thermococcus kodakaraensis*
PF0824	hypothetical protein, sodium-driven multidrug efflux pump protein	*Thermococcus kodakaraensis*
PF0825	prolyl endopeptidase; prolyl oligopeptidase [EC:3.4.21.26]	*Pyrococcus horikoshii*
PF0826	putative transport membrane protein	*Pyrococcus abyssi*
PF0827	putative RNA methyltransferase	*Pyrococcus abyssi*
PF0828	N-type ATP pyrophosphatase superfamily protein	*Pyrococcus abyssi*
PF_s016	small nucleolar RNA, posttranscriptional modification of rRNA	*Thermococcus barophilus*
PF0829	hypothetical protein	*Pyrococcus furiosus*
PF0830	hypothetical protein	*Thermococcus onnurineus*
PF0831	hypothetical protein	*Thermococcus barophilus*
PF0832	hypothetical protein	*Thermococcus barophilus*
PF0833	low temperature requirement C-like protein	*Pyrococcus horikoshii*

Table 2 cont.

Gene ID	Gene annotation	Species of closest gene homologue
PF0834	hypothetical protein; ABC-2 type transport system permease protein	*Thermococcus* sp.
PF0835	ABC-2 type transport system ATP-binding protein	*Thermococcus* sp.
PF0836	adenosylcobinamide-GDP ribazoletransferase [EC:2.7.8.26]	*Thermococcus kodakaraensis*
PF0837	adenosylcobinamide-GDP ribazoletransferase, C-terminal fragment [EC:2.7.8.26]	*Thermococcus barophilus*
PF0838	hypothetical protein	*Petrotoga mobilis*
PF0839	hypothetical protein	*Thermococcus* sp.
PF0840	nucleotidyltransferase	*Thermococcus* sp.
PF0841	hypothetical protein	*Pyrococcus abyssi*
PF0841.1n	hypothetical protein	*Thermococcus kodakaraensis*
PF0842	hypothetical protein	*Thermococcus kodakaraensis*
PF0842.1n	hypothetical protein	*Thermococcus kodakaraensis*
PF0843	hypothetical protein	*Thermococcus barophilus*
PF0844	hypothetical protein	*Pyrococcus abyssi*
PF0845	indolepyruvate ferredoxin oxidoreductase, alpha subunit [EC:1.2.7.8]	*Pyrococcus horikoshii*
PF0846	hypothetical protein	*Pyrococcus horikoshii*
PF0847	hypothetical protein	*Thermococcus kodakaraensis*
PF0848	hypothetical protein	*Pyrococcus horikoshii*
PF0849	hypothetical protein	*Pyrococcus horikoshii*
PF_s017	small nucleolar RNA, posttranscriptional modification of rRNA	*Pyrococcus horikoshii*
PF0850	hypothetical protein	*Pyrococcus horikoshii*
PF0851	DtxR family transcriptional regulator, Mn-dependent transcriptional regulator	Pyrococcus horikoshii
PF0852	hypothetical protein; putative MFS transporter, AGZA family, xanthine/uracil permease	*Pyrococcus horikoshii*
PF0853	5'-methylthioadenosine phosphorylase [EC:2.4.2.28]	*Pyrococcus abyssi*
PF_t017	tRNA-Arg	
PF0854	endo-1,4-beta-glucanase B	*Thermococcus* sp.
PF0855	hypothetical protein	*Pyrococcus horikoshii*
PF_t018	tRNA-Ala	
PF0856	isopentenyl-diphosphate delta-isomerase [EC:5.3.3.2]	Pyrococcus abyssi
PF0855.1n	hypothetical protein	*Pyrococcus horikoshii*
PF0857	ferrous iron transport protein B; ferrous iron transport protein B	*Pyrococcus abyssi*
PF0858	hypothetical protein; ferrous iron transport protein A	*Pyrococcus abyssi*
PF0859	hypothetical protein; TatD DNase family protein [EC:3.1.21.-]	*Pyrococcus abyssi*
PF0860	BtpA family protein	*Pyrococcus horikoshii*
PF0861	phosphoglucosamine mutase [EC:5.4.2.10]	*Thermococcus gammatolerans*
PF0861.1n	hypothetical protein	none
PF0862	tyrosyl-tRNA synthetase [EC:6.1.1.1]	*Pyrococcus horikoshii*
PF0863	hypothetical protein, adenylate cyclase, class 2 [EC:4.6.1.1]	*Pyrococcus horikoshii*
PF0864	AsnC family transcriptional regulator	*Pyrococcus horikoshii*
PF0865	3-octaprenyl-4-hydroxybenzoate carboxy-lyase UbiX [EC:4.1.1.-]	*Pyrococcus abyssi*
PF0866	hypothetical protein	*Pyrococcus horikoshii*
PF0867	hypothetical protein	*Thermococcus barophilus*
PF0868	mannose-1-phosphate guanylyltransferase [EC:2.7.7.13]	*Pyrococcus abyssi*
PF0869	aspartyl-tRNA synthetase [EC:6.1.1.12]	*Pyrococcus horikoshii*
PF0870	hypothetical protein	*Pyrococcus horikoshii*
PF0871	hypothetical protein	*Pyrococcus horikoshii*
PF0872	hypothetical protein	*Pyrococcus horikoshii*
PF0873	hypothetical protein	*Pyrococcus horikoshii*
PF0874	membrane dipeptidase	*Pyrococcus horikoshii*
PF0875	hypothetical protein	*Pyrococcus horikoshii*
PF0876	50S ribosomal protein L15e	*Pyrococcus horikoshii*
PF0876.1n	hypothetical protein	*Pyrococcus horikoshii*
PF0877	hypothetical protein	none
PF0878	hypothetical protein	*Pyrococcus horikoshii*
PF0879	hypothetical protein	none

Table 2 cont.

Gene ID	Gene annotation	Species of closest gene homologue
PF0880	DEXX-box ATPase	*Pyrococcus horikoshii*
PF0881	ABC-2 type transport system ATP-binding protein	*Thermococcus kodakaraensis*
PF0882	hypothetical protein	*Thermococcus kodakaraensis*
PF0883	hypothetical protein; ABC-2 type transport system permease protein	*Pyrococcus horikoshii*
PF0884	hypothetical protein	*Thermococcus kodakaraensis*

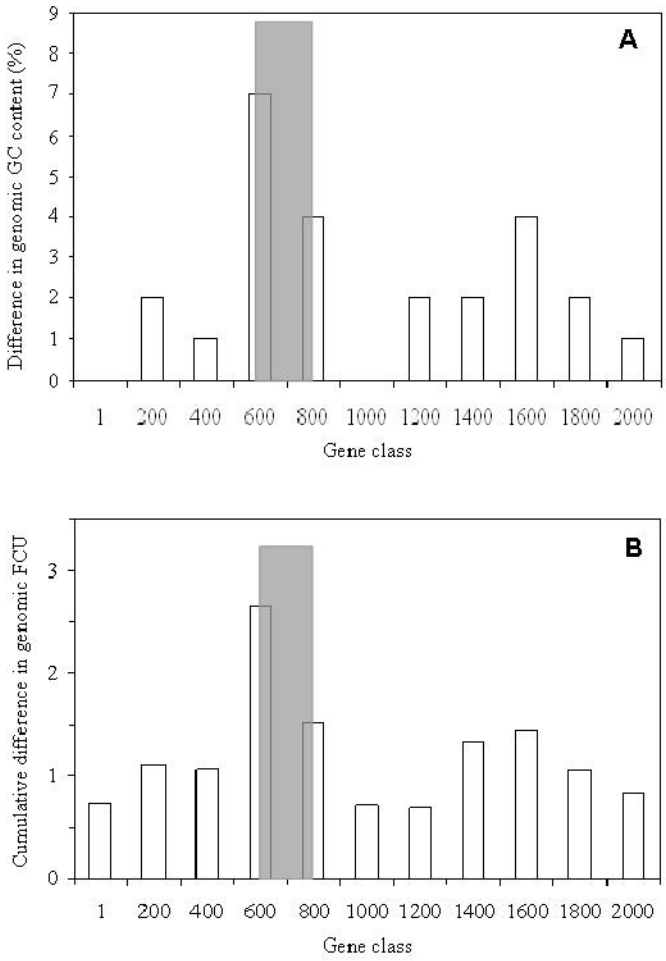

Figure 1. Comparison of the detected differences in GC content at the third codon position (A) and cumulative differences in frequency of codon usage (FCU), i.e. CD values (B) for ordered classes of genes along the *Pyrococcus furiosus* genome (white bars) and the genes in the genomic fragment located between positions 699013 and 855319 bp (grey bar). These differences are estimated with respect to the values obtained for the genes in the whole genome.

nome. No differences were detected at the first and second codon positions. The cumulative differences in FCU for the genes within the analyzed fragment were clearly different (2-3-fold) from the corresponding values estimated for the whole *P. furiosus* genome (Figure 1). The average FCU values (A_{FCUj}) for the genes within the analyzed fragment, with respect to those calculated for the genes in the whole genome, were significantly different (*G*-test for goodness-of-fit; $P < 0.001$). Besides, the percentage of putative genes encoding hypothetical proteins in the analyzed fragment was 61.1%, while the percentage for these genes located in the rest of the *P. furiosus* genome was 48.7%. Another genomic fragment surrounded by CRISPR elements and an individual element in the genome of *P. furiosus* (Table 1) could not be distinguished as significantly different from the average genome values of GC content at the third codon position and FCU.

A dendrogram based on the average FCU values for the genes contained in each of the genome fragments analyzed in Figure 1 showed that the analyzed fragment (positions 699013 to 855319 in the *P. furiosus* genome) diverged from the other genomic fragments of similar length (Figure 2A).

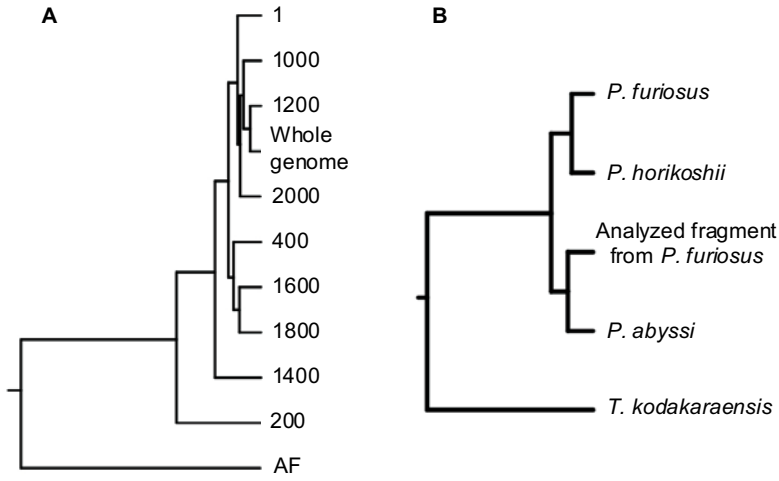

Figure 2. Dendrograms showing the proximity of different portions of *P. furiosus* genome and the divergence of the analyzed fragment (AF; positions 699013 to 855319 in the *P. furiosus* genome) (A) and the relationship of this genomic fragment to the genomes of various species of Thermococcales (B). These dendrograms were obtained on the basis of the average frequency of codon usage (FCU) for the genes present in each of the genomic fragments (as in Figure 1) or genomes considered.

The analyzed genome fragment is located in the dendrogram (Figure 2B) as closely related to *P. abyssi,* and separated phylogenetically from the characteristic FCU found in the genes of *P. furiosus* genome and the other genome sequences within the Thermococcales. These results, together with 78.3% of genes in the analyzed fragment showing the closest homology to Thermococcales genes (Table 2), suggest that this fragment probably originated from within the Thermococcales rather than from a phylogenetically distant microorganism.

CRISPRs contain spacers, i.e. nonrepeated sequences originating from previous viral infections (Barrangou et al. 2007, Brouns et al. 2008). While this role of CRISPR sequences has been demonstrated, other potential functions of these repeated sequences remain to be confirmed. For instance, gene transfer of leader genes (*cas*) (Jansen et al. 2002; Godde and Bickerton 2006) and the unidirectional addition of new viral sequences as spacers to CRISPR groups (Tyson and Banfield 2008) have been reported. Besides, the CRISPR machinery might be counteracting the spread of antibiotic resistance genes from phages and plasmids in Bacteria (Marraffini and Sontheimer 2008). Although the presence of repeated elements might facilitate recombination phenomena, there are no previous reports on the involvement of CRISPR elements in genomic insertions or deletions of CRISPR- and *cas*-independent DNA fragments between prokaryotic genomes.

Results from this study confirm horizontal gene transfer associated with CRISPR elements in *P. furiosus*. The analyzed fragment shows distinctive characteristics, which suggest its foreign origin. Thus, this study supports the involvement of CRISPR elements in the insertion of DNA fragments from diverse origins into archaeal genomes. The presence of repeated elements might facilitate the insertion of DNA fragments by recombination and significantly contribute to the evolution of species and their genomic differentiation. Further analyses of different bacterial and archaeal groups will be required to determine the magnitude and importance of the proposed lateral gene transfer-associated with CRISPR elements.

Acknowledgements. We acknowledge funding from the Government of Andalusia (BIO288), as well as the use of computer resources and assistance provided by the Barcelona Supercomputing Center (Spanish Supercomputing Network).

REFERENCES

Andersson AF, Banfield JF, 2008. Virus population dynamics and acquired virus resistance in natural microbial communities. Science 320: 1047–1050.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315: 1709–1712.

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD, 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151: 2551–2561.

Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, et al. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321: 960–964.

DeBoy RT, Mongodin EF, Emerson JB, Nelson KE, 2006. Chromosome evolution in the Thermotogales: large-scale inversions and strain diversification of CRISPR sequences. J Bacteriol 188: 2551–2561.

Garcia-Vallvé S, Romeu A, Palau J, 2000. Horizontal gene transfer in bacterial and archaeal complete genomes. Gen Res 10: 1719–1725.

Godde JS, Bickerton A, 2006. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal gene transfer among prokaryotes. J Mol Evol 62: 718–729.

Jansen R, van Embden JDA, Gaastra W, Schouls LM, 2002. Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol 43: 1565–1575.

Lillestol RK, Redder P, Garrett RA, Brugger K, 2006. A putative viral defence mechanism in archaeal cells. Archaea 2: 59–72.

Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV, 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct 1.7: 1–26.

Marraffini LA, Sontheimer EJ, 2008. CRISPR interference limits horizontal gene transfer in Staphylococci by targeting DNA. Science 322: 1843–1845.

Mojica FJM, Ferrer C, Juez G, Rodrigues-Valera F, 1995 Long stretches of short tandem repeats are present in largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. Mol Microbiol 17: 85–93.

Pourcel C, Salvignol G, Vergnaud G, 2005. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 151: 653–663.

Sokal RR, Rohlf FJ, 1981. Biometry, 2nd edn. New York: Freeman WH & Co.

Sorek R, Kunin V, Hugenholtz P, 2008. CRISPR – a widespread system that provides acquired resistance against phages in Bacteria and Archaea. Nat Rev Microbiol 6: 181–186.

Tyson GW, Banfield JF, 2008. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ Microbiol 10: 200–207.