

MicroCorrespondence

Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria

Sir,

A peculiar type of repeated element has been detected in different prokaryotes and the occurrence of similar elements in very distant phylogenetic groups is being reported subsequent to genomic sequencing. A comparative study of these peculiar elements, aimed at determining the common structural and sequence features, as well as their phylogenetic distribution, will contribute to elucidate their biological relevance.

These sequences share multiple features which are unique as a whole, being easily distinguishable from any other recurrent motif, and arising as a new family of prokaryotic repeats. They are repeated short elements generally occurring in clusters, but their main peculiarity is the layout: they are always regularly spaced by unique intervening sequences of constant length. For the sake of clarity, and ensuing from the mentioned characteristics, we will refer to the members of this family of repeats as Short Regularly Spaced Repeats (SRSRs).

Using a specific computer program, we have performed a SRSRs search in the completed microbial genomes and the available partial genome sequences of those close to completion. The organisms in which SRSRs have currently been found are listed in Table 1. In summary, the SRSRs are widespread among the various physiological and phylogenetic groups, probably being present in all the Archaea and hyperthermophilic Bacteria, in at least some members of the cyanobacteria and proteobacteria lineages, as well as in the two subgroups of Gram-positive bacteria (the low and high GC content groups). They thus represent the most widely distributed family of repeats among prokaryotic genomes.

The main features of the SRSRs are summarized in Table 1. They are typically short partially palindromic sequences of 24–40 bp, containing inner and terminal inverted repeats of up to 11 bp (see Fig. 1). Although isolated elements have been detected, the SRSR elements are generally arranged in clusters (up to 14 per genome) of repeated units spaced by unique intervening 20–58 bp sequences. The extent of the clusters is particularly noteworthy in the Archaea.

Received 20 December, 1999; revised 5 January, 2000; accepted 10 January, 2000. *For correspondence. E-mail fmojica@ua.es; Tel. (+34) 965 903 870; Fax (+34) 965 909 494.

The SRSRs are very homogeneous within a genome, most of them being identical. However, there are examples of heterogeneity, specially in Archaea. Various SRSR sequences with less than 85% similarity can be distinguished in *Pyrococcus abyssi*, *Archaeoglobus fulgidus*, *Aeropyrum pernix* and *Methanococcus jannaschii*. In the latter, two clusters with 25 and five units of the same element were initially reported (Bult *et al.*, 1996, *Science* **273**: 1058–1073). We have found 12 additional loci and three different SRSR elements, with more than 5 bp changes.

The sequence is conserved in members of the same phylogenetic group, and there is a high percentage of similarity even among domains (see Fig. 1), indicative of a common origin. Phylogenetic distance and the degree of sequence conservation closely concur. *Haloferax volcanii* differs from *Haloferax mediterranei* in 3 out of 30 bp, and *Pyrococcus horikoshii* differs from *Pyrococcus abyssi* in 2 out of 29 bp. The high degree of homology between *Escherichia coli* and *Salmonella typhi* is remarkable, with one difference out of 29 bp.

The terminal and inner-inverted repeats of each element are the most conserved regions of the SRSRs (Fig. 1), suggesting that they must be playing an essential role.

In *M. jannaschii*, *Methanobacterium thermoautotrophicum*, *A. fulgidus*, *Thermotoga maritima*, *A. pernix* and *Mycobacterium tuberculosis*, some SRSR clusters are followed by larger (> 300 bp) repeated elements. This association is not detectable in other microorganisms, nor is its possible relevance known.

A general location pattern of the SRSRs loci is not recognizable. There is, however, a remarkable coincidence. Possible chromosomal origins of replication have recently been proposed for the Archaea *M. thermoautotrophicum* and *P. horikoshii* (Lopez *et al.* 1999, *Mol Microbiol* **32**: 883–886). In both cases, two clusters of SRSRs are located one to each side of the proposed origin of replication. The distance to the origin is similar, and relatively short, for both clusters (200 and 270 kb in *M. thermoautotrophicum*, 40 and 78 kb in *P. horikoshii*). The early and simultaneous appearance of the SRSR clusters in the nascent molecules can be interpreted as being indicative of their relevance.

Besides the sequence conservation, other remarkable features of this family of tandem repeats are the palindromic nature and regular spacing of the SRSR elements. The size of the repeated unit and the presence

Table 1. Main features of the SRSRs.

Organism	SRSR size (bp)	Spacing (bp)	Number of clusters	SRSR units per cluster	Reference
Archaea					
<i>H. volcanii</i>	30	ND	≥2	ND	Mojica <i>et al.</i> (1995) <i>Mol Microbiol</i> 9 : 13–21
<i>H. mediterranei</i>	30	33–39	3	21/ ND / ND	Mojica <i>et al.</i> (1995) <i>Mol Microbiol</i> 9 : 13–21
<i>M. jannaschii</i>	28–30	31–51	7 ^A + 6 ^B + 1 ^C	4–25	Bult <i>et al.</i> (1996) <i>Science</i> 273 : 1058–1073 and this work
<i>M. thermoautotrophicum</i>	30	34–38	2	124/47	This work
<i>A. fulgidus</i>	37 ^A /30 ^B	≈ 37	1 ^A + 2 ^B	42 ^A /48 ^B /60 ^B	This work
<i>S. solfataricus</i>	25	≈ 40	≥2	94/102	Sensen <i>et al.</i> (1998) <i>Extremophiles</i> 2 : 305–312
<i>P. abyssi</i>	29 ^A /30 ^B	26–43	1 ^A + 2 ^B	7 ^A /22 ^B /27 ^B	This work
<i>P. horikoshii</i>	29	34–58	3	18/26/66	Kawarabayasi <i>et al.</i> (1998) <i>DNA Res</i> 5 : 55–76
<i>A. pernix</i>	24 ^A /23 ^B	37–52	2 ^A + 1 ^B	19 ^A /27 ^A /42 ^B	Kawarabayasi <i>et al.</i> (1999) <i>DNA Res</i> 6 : 83–101
Bacteria					
<i>T. maritima</i>	30	39–40	8	2–40	Nelson <i>et al.</i> (1999) <i>Nature</i> 399 : 323–329
<i>A. aeolicus</i>	29	36–38	1	6	This work
<i>E. coli</i>	29	32–33	3	2/7/13	Nakata <i>et al.</i> (1989) <i>J Bacteriol</i> 171 : 3553–3556 and this work
<i>S. typhi</i>	29	32	≥1	6	This work
<i>C. jejuni</i>	36	30	1	5	This work
<i>Y. pestis</i>	28	32–33	2	6/9	This work
<i>C. difficile</i>	29	36–38	4 ^A + 2 ^B	5–17	This work
<i>M. tuberculosis</i>	36	38–40	1	Variable	Hermans <i>et al.</i> (1991) <i>Infect Immun</i> 59 : 2695–705
<i>Calothrix sp.</i>	37	35–41	>1	5	Masepohl <i>et al.</i> (1996) <i>Biochim Biophys Acta</i> 1307 : 20–36
<i>Anabaena sp.</i>	37	32–43	>1	17	Masepohl <i>et al.</i> (1996) <i>Biochim Biophys Acta</i> 1307 : 20–36
Mitochondria					
<i>V. faba</i>	40	20–35	1	6	Flamand <i>et al.</i> (1992) <i>Plant Mol Biol</i> 19 : 913–923

A,B, Types of SRSRs distinct (more than 3 bp differences) within the same microorganism. ND, Not determined.

of inner short inverted repeats are characteristics that concur with those of recognizing sites for certain DNA-binding proteins. The regular spacing of the SRSR elements locate the inverted repeats to the same side of the DNA chain. Although cooperative binding to free proteins cannot be excluded, this peculiar arrangement, with such a length of regularly positioned sites, would rather suggest the need for a solid attachment to a cellular structure that is consequently organized. This would be in agreement with the previously proposed role in replicon partitioning for the SRSRs of haloarchaea (Mojica *et al.* 1995, *Mol Microbiol* **9**: 13–21)

The question emerges here as to whether the SRSRs have a common function in prokaryotes, or whether their presence is reminiscent of ancient sequences and their role diverged with evolution. The universality, phylogeny and biological significance of this peculiar family of repeats arises as an item to be elucidated.

Acknowledgements

This work was financed by a research grant from the Conselleria de Cultura Educació i Ciència, Generalitat Valenciana (GV97-VS-25–82). E.S. holds a graduate fellowship from the

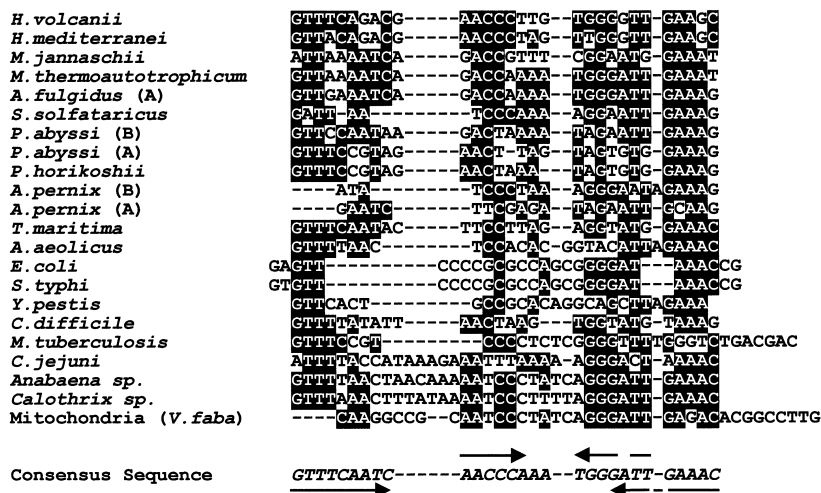


Fig. 1. Alignment of the SRSRs. Highlighted blocks indicate positions occupied by the most frequent base in the aligned sequence. Only the most abundant type of SRSR element has been considered for *M. jannaschii* and *Clostridium difficile*. Two types of SRSR (A and B) present in *P. abyssi*, *A. pernix* and *A. fulgidus* have been aligned. A consensus sequence with the most frequent base at each position in the alignment is included. Arrows indicate the palindromic character of the SRSRs.

Conselleria de Cultura Educació i Ciència, Generalitat Valenciana.

The sequence data of unfinished genomes were produced by the S.typhi (*Salmonella typhi*), the C.jejuni (*Campylobacter jejuni*), the Y.pestis (*Yersinia pestis*), and the C.difficile (*Clostridium difficile*) Sequencing Groups at the Sanger Centre and can be obtained from <ftp://ftp.sanger.ac.uk/pub/pathogens/st>, <ftp://ftp.sanger.ac.uk/pub/pathogens/cj>, <ftp://ftp.sanger.ac.uk/pub/pathogens/yp> and <ftp://ftp.sanger.ac.uk/pub/pathogens/cd> respectively.

Francisco J. M. Mojica,^{1*} Cesar Díez-Villaseñor,¹ Elena Soria¹ and Guadalupe Juez²

¹*División de Microbiología, Departamento de Fisiología, Genética y Microbiología, Universidad de Alicante, Apartado 99, 03080 Alicante, Spain.* ²*División de Microbiología, Campus de San Juan, Universidad Miguel Hernández, 03550 Alicante, Spain.*