# Prevalence of clustered regulatory interspaced short palindromic repeat (CRISPR)-like sequences in mitis-group streptococci

Y. MAEDA[*†], C. E. GOLDSMITH[*], W. A. COULTER[‡], C. MASON[‡], J. S. G. DOOLEY[†], C. J. LOWERY[†], W. J. SNELLING[†] and J. E. MOORE[*†]

[*]*Northern Ireland Public Health Laboratory, Department of Bacteriology, Belfast City Hospital;* [†]*School of Biomedical Sciences, University of Ulster, Coleraine; and* [‡]*School of Dentistry, Queen's University of Belfast, Royal Group of Hospitals, Belfast, Northern Ireland, UK*

*Accepted: 24 December 2010*

**ABSTRACT**

Clustered regulatory interspaced short palindromic repeats (CRISPRs) have been discovered in many bacteria and archaea. Many CRISPR-like sequences have been identified in an increasing number of studies on the function of CRISPRs. One CRISPR-like sequence of approximately 240 base pairs has been found to be highly conserved within 11 genome sequences of *Streptococcus pneumoniae*. A specific CRISPR-like polymerase chain reaction (PCR) assay was designed with the novel primers CRISPR 5F (forward primer) 5′-CTA ATY TCA TAA CCA TAR GAA TC-3′ and CRISPR 3R (reverse primer) 5′-GAT AAR ATC CTY TAA WCT TCT AG-3′ to detect the presence of this CRISPR-like sequence in pneumococci, as well as in viridans-group streptococci (VGS). This study investigates the prevalence of this CRISPR-like sequence in *S. pneumoniae* and 12 viridans-group streptococcal species and shows its existence to be shared by the majority of *S. pneumoniae* and, to a lesser extent, *S. mitis*. This CRISPR-like sequence was also found in *S. australis* and it is highly conserved among these strains, suggesting possible biological functional differences from true CRISPR because this CRISPR-like sequence has relatively few repeat numbers, and adjacent homology of CRISPR-associated (*cas*) genes was absent. The sharing of this CRISPR-like sequence between pneumococci, the mitis group and other VGS, as well as its high sequence homology, may suggest close evolutionary emergence of this sequence between these species.

KEY WORDS: DNA primers.
Molecular biology.
Polymerase chain reaction.
Streptococcus.

## Introduction

Clustered regulatory interspaced short palindromic repeats (CRISPRs) are repeat sequences that exist in many bacteria and archea. When using up-to-date genome sequence databases, CRISPRs are present in approximately 40% of bacteria and in many archeae.[1] They have a specific structure which consists of a few hundred base pairs (bp) of leader sequence, 24–47 bp of direct repeats and a unique spacer region between the direct repeat regions. Protein coding regions adjacent to CRISPR structure have been identified and are designated CRISPR-associated (*cas*) genes. The repeat number varies from species to species and is usually two or more.[2] There has been speculation that CRISPRs provide bacteriophage resistance to bacteria because some spacer sequences have high similarity to bacteriophages.[3] This has been studied in *Streptococcus thermophilus* and showed a direct correlation between numbers of newly acquired spacer sequences and the level of bacteriophage resistance.[4]

With the increasing number of CRISPRs discovered, some questionable CRISPR structures have been identified. Questionable CRISPRs are defined as having two to three direct repeats and/or not having 100% identical direct repeats. In addition, when deciding if a sequence is CRISPR or CRISPR-like, consideration has to be given to whether or not the sequence is part of a coding region, as CRISPRs are usually found in non-coding regions.[2] These CRISPR-like sequences have been observed in many bacterial species and these repeat sequences appear to be unique to each species.[5]

Currently, 11 genome sequences of *S. pneumoniae* are available in GenBank and all possess common CRISPR-like sequences, which comprise two direct repeats with a spacer region between them. This study aims to investigate whether this CRISPR-like sequence is present in only *S. pneumoniae* or is shared by closely related species (mitis group of viridans-group streptococci [VGS]) and to examine phylogenetic relationships between pneumococci and the other VGS using this sequence marker.

*Correspondence to: Professor John E. Moore, Northern Ireland Public Health Laboratory, Department of Bacteriology, Belfast City Hospital, Belfast BT9 7AD, Northern Ireland, UK*
*Email: jemoore@niphl.dnet.co.uk*

## Materials and methods

### *Bacterial isolates*

Streptococcal organisms ($n$=190) belonging to the VGS were used in this study. All organisms had been previously isolated from clinical specimens at Belfast City Hospital and

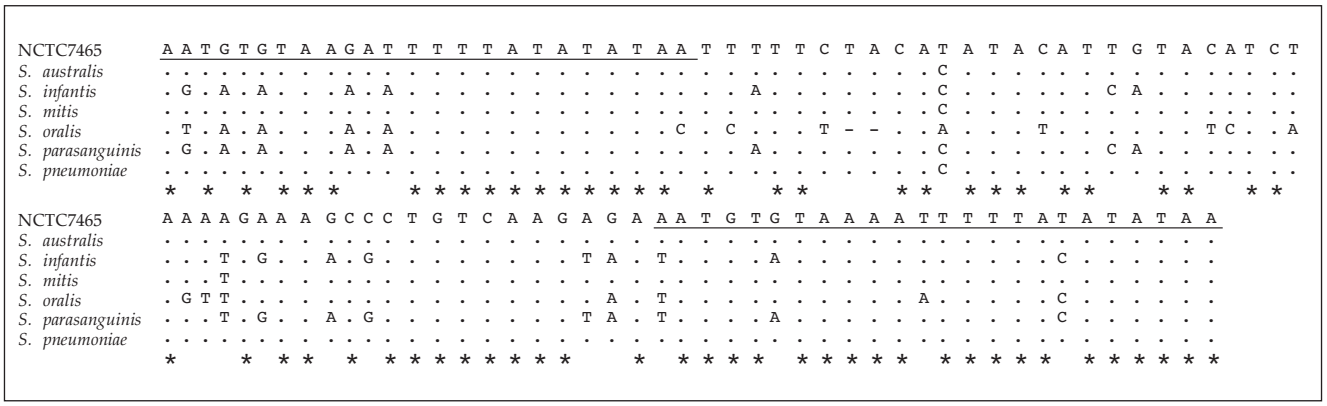


**Table 1.** Sequence alignment of CRISPR-like sequences of representative strains. Underlines denote direct repeat regions.

identified by molecular means to species level by *rnp*B gene sequence analysis. These included *S. salivarius* (*n*=47), *S. mitis* (*n*=36), *S. sanguinis* (*n*=25), *S. oralis* (*n*=20), *S. pneumoniae* (*n*=19), *S. parasanguinis* (*n*=16), *S. infantis* (*n*=11), *S. gordonii* (*n*=7), *S. anginosus* (*n*=4), *S. cristatus* (*n*=2), *S. australis* (*n*=1), *S. mutans* (*n*=1) and *S. agalactiae* (*n*=1). The reference strains of *S. pneumoniae*, NCTC 7465 and ATCC 49619 were included.

### *DNA extraction*

Bacterial strains were subcultured on Columbia blood agar supplemented with 5% (v/v) horse blood (Oxoid, Basingstoke, UK) for 24 h at 37˚C under microaerophilic conditions. All DNA isolation procedures were carried out in a Class II biological safety cabinet (MicroFlow, UK) in a room separated from that used to set up nucleic acid amplification reaction mixes and also from the 'post-PCR' room in accordance with the Good Molecular Diagnostic Procedure (GMDP) guidelines of Millar *et al.*,[6] in order to minimise contamination. Bacterial genomic DNA was extracted from a few colonies of each isolate using the Roche High Purity PCR Template Preparation Kit (Roche Diagnostics, Sussex, UK), in accordance with the manufacturer's instructions. Extracted DNA was stored at –20˚C prior to polymerase chain reaction (PCR) amplification.

### *PCR amplification*

A primer set to amplify CRISPR-like sequences was designed *in silico* based on genome sequences of *S. pneumoniae* in GenBank which were shown to have CRISPR-like sequences by employing CRISPRFinder.[5] The sequence of each primer was CRISPR 5F (forward primer) 5'-CTA ATY TCA TAA CCA TAR GAA TC-3' and CRISPR 3R (reverse primer) 5'-GAT AAR ATC CTY TAA WCT TCT AG-3'. Polymerase chain reaction mixes (25 µL) contained 1 µL DNA template, 10 mmol/L Tris-HCl (pH 8.3) 50 mmol/L KCl, 2.0 mmol/L $MgCl_2$, 200 µmol each dNTP, 0.6 unit *Thermophilus aquaticus* (*Taq*) DNA polymerase (New England Biolabs, Hertfordshire, UK) and 10 µmol each primer. Following a 'hot start', the reaction mixtures were subjected to the following thermal cycling parameters in a GeneAmp PCR system 9700 thermocycler (Applied Biosystems, Warrington, UK): 95˚C for 5 min, followed by 35 cycles of 95˚C for 30 sec, 54˚C for 30 sec, and 72˚C for 45 sec, and a final extension at 72˚C for 7 min. After amplification, the PCR products were visualised on 1.5% (w/v) agarose gel in 0.5 x TBE buffer, followed by staining with ethidium bromide (0.5 µg/mL) and ultraviolet (UV) illumination with a gel image analysis system (UVP Products, Cambridge, UK).

### *DNA sequencing and analysis*

Amplicons for sequencing were purified with a QIAquick PCR purification kit (Qiagen, West Sussex, UK), according to the manufacturer's instructions. The PCR amplicons containing CRISPR-like sequences were determined using the BigDye Terminator Cycle Sequencing Kit and an ABI 3100 genetic analyser (Applied Biosystems). Sequencing analysis was carried out (Genetyx, Tokyo, Japan) and a phylogenetic tree was constructed by the neighbour-joining method[7] with 1000 bootstrap replicates using MEGA 4.0.2.[8]

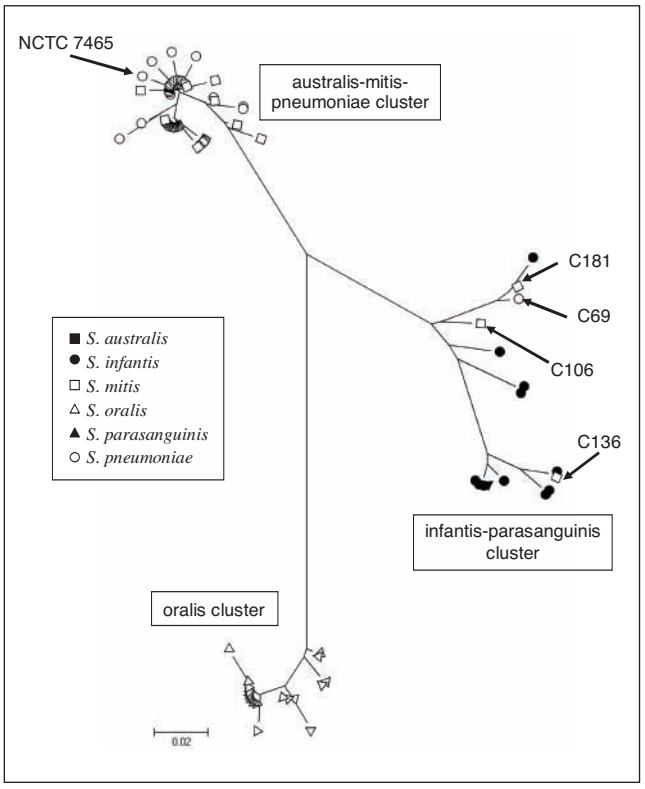


**Fig. 1.** Phylogenetic tree constructed from 96 bp of CRISPR-like sequences of 84 streptococcal strains using the neighbour-joining method. The evolutionary distances were computed using the Maximum Composite Likelihood method.[11]

## Results

About 240 bp of PCR amplicons containing CRISPR-like sequences were obtained from *S. australis* (1/1, 100%), *S. infantis* (11/11, 100%), *S. mitis* (35/36, 97.2%), *S. oralis* (19/20, 95%), *S. parasanguinis* (1/16, 6.3%) and *S. pneumoniae* (21/21, 100%). The CRISPR-like sequence regions of representative strains are shown in Figure 1. Average sequence similarities of each species to *S. pneumoniae* NCTC 7465 were as follows: *S. australis* (98.9%), *S. infantis* (78.1%), *S. mitis* (95.9%), *S. oralis* (73.4%), *S. parasanguinis* (77.5%) and *S. pneumoniae* (97.2%). Sequence analysis revealed no direct repeat sequences in *S. infantis, S. oralis* and *S. parasanguinis*.

Complete identity of the direct repeat regions was found in 21 *S. mitis* and 17 *S. pneumoniae*, with only one nucleotide substitution seen in seven and three isolates, respectively. Three *S. mitis* and one *S. pneumoniae* isolates showed CRISPR-like sequence homology of 91.5–94.9% to *S. infantis*, which did not possess conserved direct repeat regions and were shown to incorporate in the *S. infantis* cluster when a phylogenetic tree was constructed.

Distinctive clusters were built between *S. infantis* and *S. oralis* strains, although neither showed direct repeat regions in this CRISPR-like sequence. Nucleotide CRISPR-like sequences of *S. australis* C89, *S. mitis* C31 and *S. pneumoniae* NCTC 7465 and C10 have been deposited in GenBank under accession numbers HM061637–HM061640.

## Discussion

Oral streptococci are largely composed of members of the VGS, which currently encompasses 20 species that are regarded as commensal inhabitants of the oropharyngeal cavity and the gastrointestinal and genital tracts. On the basis of 16S rRNA sequence homology, these bacteria are categorised into four groups: the salivarius rRNA homology group, the mitis group, the anginosus group, and the mutans group.

Recently, several CRISPRs have been described in the genomes of various bacteria and archaea, where these repeats range in size from 23 bp to 47 bp. They usually show some dyad symmetry but may not be truly palindromic, and such repeats are separated by spacers of similar length. Horizontal gene transfer (HGT) may be an important mechanism for the exchange of essential genetic material between VGS in the mouth and upper respiratory tract and *S. pneumoniae*.

The recent description of CRISPRs in the archaea and bacteria has caused much excitement in terms of microbial evolution. To date, 183 CRISPRs have been identified in 46 archaeal genomes sequenced and 594 CRISPRs have been located in the 535 bacterial genomes fully sequenced. To date, the function of CRISPRs is not entirely clear. An *in silico* analysis of pneumococci and VGS demonstrated the presence of several confirmed and putative new CRISPR regions. The sharing of CRISPR sequence homology between the pneumococci suggests the presence of a unique signature sequence diagnostic marker that could aid identification and genotyping of the pneumococci. CRISPRs have been identified in several species of the genus *Streptococcus*, including *S. agalactiae*, *S. dysgalactiae* subsp. *equisimilis*, *S. dysgalactiae* subsp. *zooepidemicus*, *S. gordonii*, *S. mutans, S. pyogenes, S. sanguinis* and *S. thermophilus*. In particular, CRISPRs in *S. thermophilus* have been studied in depth as it involves industrial interest of quality control of various fermented dairy products.[9] However, no reports on the prevalence of CRISPR-like sequences in the genus *Streptococcus* have appeared to date.

The present study describes for the first time the existence of CRISPR-like sequences in *S. australis, S. mitis* and *S. pneumoniae*, which were not found in the other 10 streptococcal species examined in this study. Although only one *S. australis* isolate was investigated in this study, direct repeat regions within CRISPR-like sequences of these three species (*S. australis, S. mitis* and *S. pneumoniae*) were highly conserved, suggesting those CRISPR-like sequences may have a potential role that is different from true CRISPRs. True CRISPRs are usually found with CRISPR-associated (*cas*) genes adjacent to upstream leader sequences or downstream of the last direct repeat region, which possess endonuclease activity and are thought to contribute to the acquisition of new spacer regions that confer bacteriophage resistance.[10] No *cas* gene homology was found in the 11 genome sequences of *S. pneumoniae* by BLAST analysis with *cas* gene sequences of *S. gordonii* and *S. thermophilus*, which are closely-related species to *S. pneumoniae* (data not shown).

This study showed that a few outliers of *S. mitis* and *S. pneumoniae* made a cluster with *S. infantis*, suggesting there were interspecies homologous recombinations, as their sequence identities showed >91% similarity and the nucleotide sequence substitution patterns were similar to *S. infantis*. However, why such recombinational events were not observed in *S. oralis*, which is more closely related to *S. mitis* and *S. pneumoniae*, is unclear. Unfortunately, due to the lack of specificity, it has not been possible to exploit the presence of this CRISPR sequence within *S. pneumoniae* to design a pneumococcal detection assay based on this CRISPR-like region. Equally, with its absence from certain *S. mitis* isolates, it is impossible to attribute this CRISPR-like region as specific to the *S. mitis* group.

In conclusion, the 240 bp CRISPR-like sequence found in this study was widely distributed in *S. pneumoniae* and, to a lesser extent, in *S. mitis*. This CRISPR-like sequence was also shown to be present in *S. australis*, which suggests that other species of mitis-group streptococci may have common CRISPR-like sequences. Sharing of this CRISPR-like sequence between pneumococci, the mitis group and other VGS, as well as its high sequence homology, suggests close evolutionary emergence of this sequence between these species, and work is now required to define the functionality of this highly conserved CRISPR-like region. □

*This work was supported by a project grant from the HPSS (NI) Research & Development Office Commissioned Research in Antimicrobial Resistance Action Plan (AMRAP) (COM/2730/04).*

## References

1 Kunin V, Sorek R, Hugenholtz P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol* 2007; **8** (4): R61.

2 Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 2007; **8**: 172.

3 Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 2005; **151** (8): 2551–61.
4 Barrangou R, Fremaux C, Deveau H *et al*. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 2007; **315** (5819): 1709–12.
5 Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 2007; **35** (Web Server issue): W52–7.
6 Millar BC, Xu J, Moore JE. Risk assessment models and contamination management: implications for broad-range ribosomal DNA PCR as a diagnostic tool in medical bacteriology. *J Clin Microbiol* 2002; **40** (5): 1575–80.
7 Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987; **4** (4): 406–25.
8 Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007; **24** (8): 1596–9.
9 Mills S, Coffey A, McAuliffe OE, Meijer WC, Hafkamp B, Ross RP. Efficient method for generation of bacteriophage-insensitive mutants of *Streptococcus thermophilus* yoghurt and mozzarella strains. *J Microbiol Methods* 2007; **70** (1): 159–64.
10 Karginov FV, Hannon GJ. The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol Cell* 2010; **37** (1): 7–19.
11 Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci USA* 2004; **101** (30): 11030–5.