Epidemiol. Infect. (1997), **118**, 125–135

Characterization of the Lancefield group C streptococcus 16S-23S RNA gene intergenic spacer and its potential for identification and sub-specific typing

N. CHANTER, N. COLLIN, N. HOLMES, M. BINNS AND J. MUMFORD

Animal Health Trust, P.O. Box 5, Newmarket, Suffolk, UK

(*Accepted* 27 *October* 1996)

SUMMARY

The 16S-23S RNA gene intergenic spacers of isolates of *Streptococcus equi* ($n = 5$), *S. zooepidemicus* ($n = 5$), *S. equisimilis* ($n = 3$) and *S. dysgalactiae* ($n = 2$) were sequenced and compared. There were distinct regions within the spacer, arranged in the order 1–9 for all *S. equi* and one *S. zooepidemicus* isolate and 1,2 and 4–9 for the remaining isolates. Region 4 was identical to the $tRNA^{ala}$ gene found in the 16S-23S intergenic spacers of other streptococci. Regions 1, 5, 6 and 7 had distinct variations, each conserved in different isolates. However, amongst the intergenic spacers there were different combinations of variant regions, suggesting a role for DNA recombination in their evolution. The intergenic spacer of all isolates of *S. equi* and one *S. zooepidemicus* isolate were almost identical. Primers derived from the variant sequences of regions 1 and 5 to 6 were used to group all *S. zooepidemicus* ($n = 17$) and *S. equi* ($n = 5$) into 1 of 8 types by polymerase chain reaction; three *S. zooepidemicus* isolates typed the same as *S. equi*. *S. equi* and *S. zooepidemicus* were clearly distinguishable from *S. equisimilis* and *S. dysgalactiae* which had shorter regions 5 and 6 and no region 7. Most homology for the group C sequences was found in previously published sequences for the 16S-23S intergenic spacers of *S. anginosis*, *S. constellatus*, *S. intermedius*, *S. salivarius* and *S. agalactiae*. A 75-90 nucleotide length shared with *S. anginosus* and *S. intermedius* in opposite orientations in the two main variants of region 6 supported the role for DNA recombination in the evolution of the spacer. The 16S-23S intergenic spacers indicate that *S. zooepidemicus* was the archetypal species for *S. equi* and that both are genetically more distant from *S. equisimilis* and *S. dysgalactiae*. The intergenic spacer can be used to identify specifically the group C streptococci and as an epidemiological marker for *S. zooepidemicus*.

INTRODUCTION

Streptococcus equi and *S. zooepidemicus*, the causes (or aggravating factor) of an important proportion of infectious disease problems of the horse, belong to Lancefields antigenic group C. *S. equi* causes a distinct disease syndrome, strangles, in which the lymph nodes of the head are typically invaded, become abscessated and rupture through the skin or mucous membranes usually leading to recovery; in 1–5% of cases death follows metastases to almost any organ of the body. *S. equi* can be carried for long periods in the guttural pouch and yet remain undetected by culture from nasopharyngeal swabs [1].

S. zooepidemicus is associated with lower airway disease during the training of thoroughbred horses [2], foal pneumonia [3], endometritis [4] and abortion [5]. Equids of all types commonly suffer a series of infections by *S. zooepidemicus*, particularly when young. This might be related to antigenically distinct surface proteins [6] which may be protective immunogens [7], analogous to the M-proteins of *S. pyogenes*

[8], which determine susceptibility according to the immune experience of the host; *S. equi* appears to have just one type of M-like protein. Currently, there is no clear definition of these antigens or tests for their detection. In contrast to *S. equi*, *S. zooepidemicus* is not only found as a pathogen in cattle, pigs and man but may also act in a commensal role; it is commonly found colonizing the mucous membranes of healthy equids.

S. equisimilis, another group C streptococcus, is generally regarded as a common commensal in the horse although it is often found in lesions together with *S. equi* or *S. zooepidemicus*. It is, however, regarded as a potential pathogen in the pig and man.

The objectives of this study were to investigate the use of the 16S-23S RNA gene intergenic spacer as the basis for a PCR test for the detection of *S. equi*, potentially for identification of culture negative carrier animals, and as a means of additionally subtyping *S. zooepidemicus* to provide simpler techniques for epidemiological studies than culture and the typing of surface proteins. *S. equisimilis* was studied because of the need to differentiate its presence in equine lesions and *S. dysgalactiae* was included as the remaining Lancefield group C species.

The 16S-23S RNA gene intergenic spacer of bacteria has been used for identification and epidemiological sub-typing [9–11]. Sequences in the 16S and 23S RNA genes are highly conserved across most bacteria so the intergenic spacer is readily accessible for analysis through polymerase chain reaction tests (PCR). The presence of several RNA gene operons in each bacterial cell has increased the sensitivity of PCR detection. The presence of conserved differences between most species of bacteria coupled with other differences within species and the presence of different operons in the same cell, has provided conserved variations needed both for highly specific identification and for sub-specific typing.

Taxonomic uncertainty within the streptococci is being unravelled using the 16S RNA gene and the 16S-23S intergenic spacer. For example, the oral '*S. milleri* group' of *S. anginosus*, *S. intermedius* and *S. constellatus*, a cause of purulent infections, was shown to comprise a distinct taxon by 16S RNA sequence analysis [12] even though there is species specific variation in this region of DNA [13]; by using 16S-23S RNA sequences, *S. anginosus* could be distinguished from *S. constellatus* but not from *S. intermedius* [14].

Taxonomic uncertainty also exists within the Lancefield group C streptococci. *S. equi* and *S. zooepidemicus* have been classified as *S. equi* subspecies *equi* and subspecies *zooepidemicus* on the basis of DNA-DNA hybridization [15]. However, multilocus enzyme electrophoresis demonstrated that *S. zooepidemicus* is likely to have been the archetypal species from which *S. equi* was derived [16]. The 16S-23S intergenic spacer data described here support this view and, consequently, we have chosen not to use the subspecies nomenclature.

Table 1. *Location and year of isolation of the Lancefield group C streptococci*

Species	Isolate	Location (premises, horse)	Year
S. equi	3938	Cambridgeshire	1990
S. equi	4047	Northamptonshire	1990
S. equi	758	Norfolk	1993
S. equi	1026	Suffolk	1995
S. equi	1224	Surrey	1995
S. zooepidemicus	6393	Suffolk (R)	1991
S. zooepidemicus	3682	Sussex (X)	1992
S. zooepidemicus	3685	Sussex (Y)	1992
S. zooepidemicus	4020	Sussex (Z)	1992
S. zooepidemicus	3711	Norfolk	1992
S. zooepidemicus	6016	Berkshire (A)	1993
S. zooepidemicus	K3	Suffolk (U)	1993
S. zooepidemicus	2809	Berkshire (B)	1993
S. zooepidemicus	6416	North Yorkshire	1993
S. zooepidemicus	145	Suffolk (S)	1994
S. zooepidemicus	148	Berkshire (A)	1994
S. zooepidemicus	460	Suffolk (S)	1994
S. zooepidemicus	461	Suffolk (S)	1994
S. zooepidemicus	1741	Buckinghamshire (R)	1994
S. zooepidemicus	1732	Buckinghamshire (R)	1994
S. zooepidemicus	3454	Suffolk (T)	1994
S. zooepidemicus	46b	Suffolk (U)	1995
S. equisimilis	27	Suffolk	1994
S. equisimilis	585	Surrey	1994
S. equisimilis	1402	Berkshire	1995

MATERIALS AND METHODS

Bacterial isolates and culture conditions

Five isolates of *S. equi* from nasopharyngeal swabs of 5 different ponies on 5 separate premises and 17 isolates of *S. zooepidemicus*, all from nasopharyngeal swabs or tracheal washes, taken trans-endoscopically, from different thoroughbred horses and ponies on 12 separate premises throughout Great Britain and three separate isolates of *S. equisimilis* (Table 1), were identified by their reactivity with antibody to group C antigen and by their sugar fermentation reactions.

Table 2. *Primers used to amplify the 16S-23S RNA gene intergenic spacer or subsections of it*

Primer	Sequence (5′-3′)	Reference
a	AGTCGTAACAAGGTAAGCCG	17
b	GGTACCTTAGATGTTTCAGTTC	10
c	TTGTACACACCGCCCGTCA	10
d	AAAAAGGAAGCACGTTTAGCG	This paper
e	AAAAAGGAACACGTTTAGCG	This paper
f	T AAAAAAGGAAGCATGTTTGGAAG	This paper
g	CCGTCTGTTAGTATCCTGTTT	This paper
h	A C AAACTTTTCTTTGTATCCTAG	This paper

Two strains of *S. dysgalactiae* (NCTC 4335 and 4669) were obtained from the National Collection of Type Cultures. All were grown overnight at 37 °C in Todd Hewitt broth (Oxoid).

Extraction of DNA for PCR

The growth from 1·5 ml of culture was centrifuged at 10000 g for 10 min, resuspended in 0·6 ml of 8 M guanidine thiocyanate and heated at 95 °C for 20 min. After centrifugation at 10000 g for 15 min the supernatant fraction was extracted using the Wizard DNA clean up system (Promega) according to the manufacturer's instructions with elution of DNA in deionized water.

PCR conditions

A 25 μl reaction volume for PCR contained 25 pM of primers b and c (Table 2), 0·25 μl of extracted target DNA sample, 0·5 μl of 10 mM dNTP's and 2·5 μl of 10 × buffer V (Advanced Biotechnologies) and 0·5 units of Taq DNA polymerase (Advanced Biotechnologies). The thermocycling conditions after heating to 94 °C for 5 min (during which the DNA sample was added) were 40 cycles of 94 °C for 1 min, an annealing temperature of 58 °C for 2 min and 72 °C for 3 min before a final chain extension step for 7 min at 72 °C. The same conditions were used for PCR with primers d to h except that the annealing temperature was 55 °C.

Sequencing

PCR products were sequenced using primers a and b (Table 2), the ABI PRISM™ dye terminator cycle sequencing kit with AmpliTaq DNA polymerase FS and an ABI PRISM automated sequencing machine according to the manufacturer's instructions (Perkin Elmer).

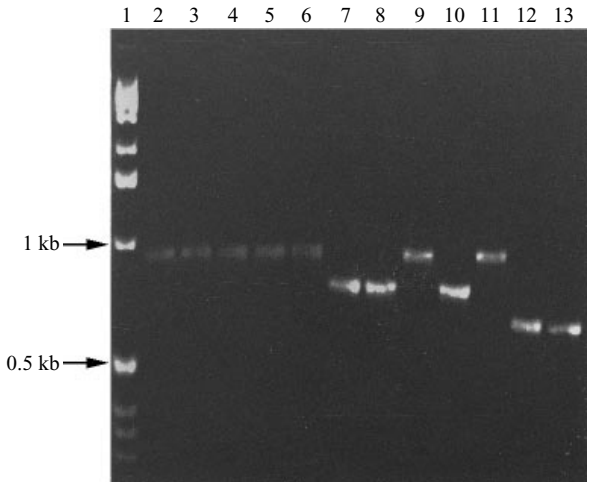

Fig. 1. Agarose gel electrophoresis of PCR products from primers b and c which amplify the 16S-23S RNA gene intergenic spacer. Lane 1: 1 kb standards. Lanes 2–6: separate isolates of *S. equi*. Lane 7: *S. zooepidemicus* 2809. Lane 8: *S. zooepidemicus* 3682. Lane 9: *S. zooepidemicus* 461. Lane 10: *S. zooepidemicus* 3685. Lane 11: *S. zooepidemicus* K3. Lane 12: *S. dysgalactiae*. Lane 13: *S. equisimilis*.

Agarose gel electrophoresis

PCR products were analysed by agarose gel electrophoresis in 0·04 M Tris acetate pH 8·3 containing 0·001 M EDTA. PCR products from primers b and c and d–h were separated in 1 and 3% agarose, respectively, and detected with a UV transilluminator after ethidium bromide staining. The sizes of PCR products were estimated by their electrophoretic mobilities relative to those of the 1 kb DNA ladder (GibcoBRL).

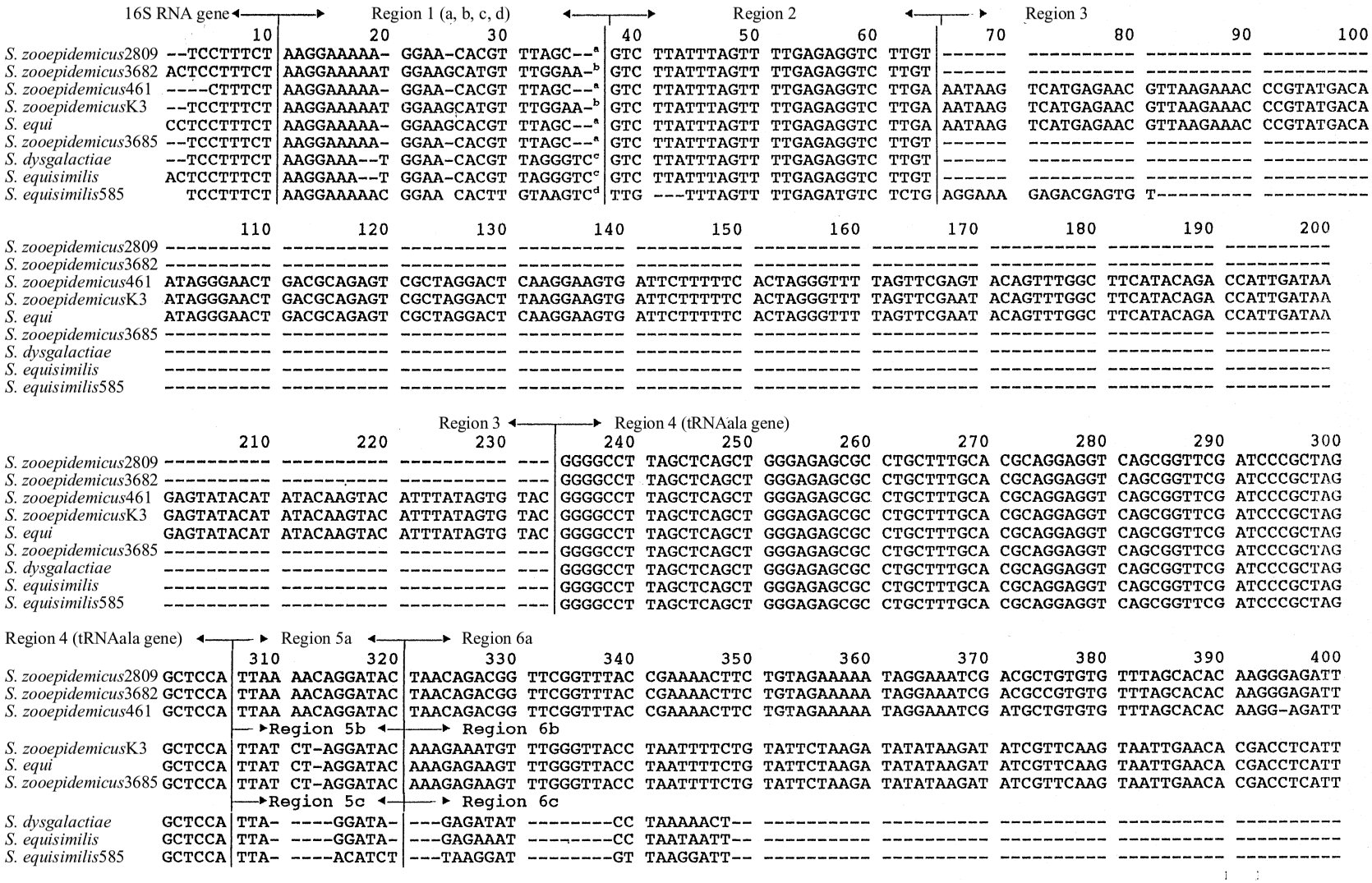

Fig. 2. For legend see facing page.

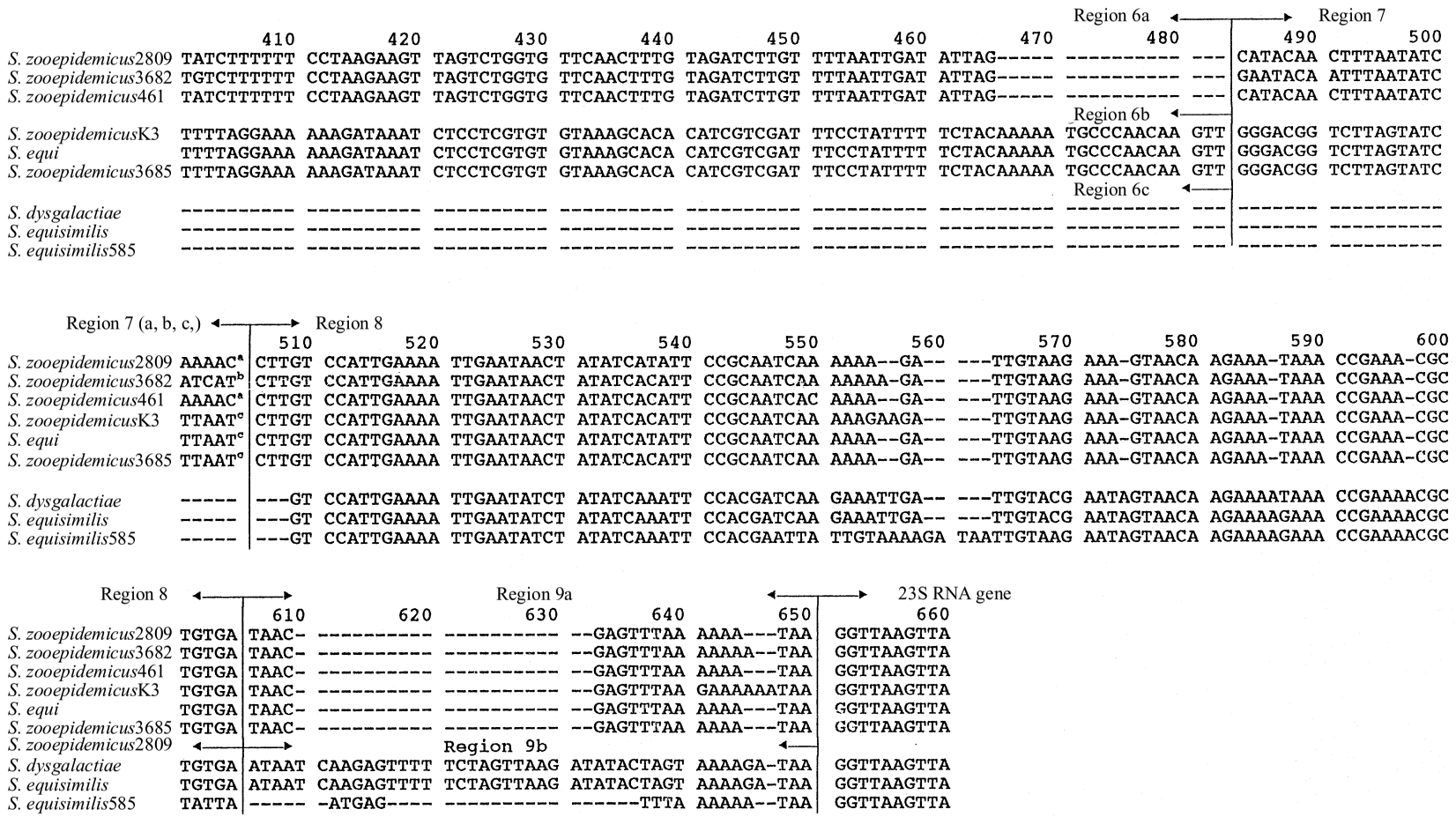

Fig. 2. Multiple alignment of the 16S-23S ribosomal RNA intergenic spacer sequences for Lancefield Group C streptococci; stuffer regions to achieve alignment are indicated by ---.

DNA sequence analyses

The complete sequences of the intergenic spacers were assembled from overlapping sequence derived from both strands, and multiple alignments of these were produced with the aid of DNASIS (Hitachi). The Genbank data base was searched for homologies using BLAST.

RESULTS

PCR amplification of the 16S-23S intergenic spacer using primers b and c (derived from conserved regions within the 16S and 23S ribosomal RNA genes; see Table 2), respectively, produced a single amplicon of approximately 850 bp from all five *S. equi* isolates (Fig. 1). In contrast, three isolates of *S. zooepidemicus* produced a single amplicon of approximately 750 bp whilst one of the *S. zooepidemicus* isolates produced an amplicon similar in size to that produced by the *S. equi* isolates. Another *S. zooepidemicus* isolate produced an intermediate sized amplicon of approximately 800 bp. PCR amplification of *S. dysgalactiae* and *S. equisimilis* produced similar but smaller amplicons of approximately 650 bp (Fig. 1).

The 16S-23S intergenic amplicons of the isolates of *S. equi* and *S. zooepidemicus*, two isolates of *S. dysgalactiae* and three isolates of *S. equisimilis* were sequenced using primers a and b (the sequence of primer a is closer to the end of the 16S RNA gene than that of primer c). The entire sequence for the intergenic spacers was obtained from the complementary overlap between the sequences derived from each of the primers. The sequences for these isolates were multiply aligned (see Fig. 2). Since the sequences for the *S. equi* isolates were identical, with the exception of two bases which will be described later, and the sequences for the two *S. dysgalactiae* isolates were identical, a representative sequence for each of these species is in Figure 2. Two of three sequences from three distinct isolates of *S. equisimilis* were identical and virtually identical with the sequence for *S. dysgalactiae*; the third sequence (*S. equisimilis* 585) was distinct and a representative of each of the two *S. equisimilis* sequences is in the multiple alignment.

Analysis of the multiple alignment revealed that the intergenic spacers had distinct sub-regions with an ordered relationship to each other. Some of these regions were identified by characteristic variations in sequence between some isolates but which were common to others. The regions with variants were punctuated by other regions which were homologous in most of the isolates and species. On the basis of which combination of variant regions made up a particular intergenic spacer, all five of the *S. zooepidemicus* isolates could be separately identified. One of these combinations was essentially identical to the sequence for the 16S-23S intergenic spacer of *S. equi*. The intergenic spacers of the other *S. zooepidemicus* isolates shared some of these regions with *S. equi* to a greater or lesser extent.

Overall each isolate had at least eight regions. All *S. equi* and two of the *S. zooepidemicus* isolates had an additional region (region 3). One of the regions (region 6) was significantly shorter for the *S. equisimilis* and *S. dysgalactiae* isolates.

Region 1 (positions 11–37 of the multiple alignment) followed the end of the 16S ribosomal RNA gene. Although the region showed strong homology between all of the Lancefield group C streptococci there were three clearly identifiable variants on the basis of nucleotides at positions 28 and 33–37. Variants 1a, 1b and 1c were represented at least twice and 1d was present in *S. equisimilis* 585. Two of the five *S. equi* isolates had an additional nucleotide in this region which was a guanidine at position 25 of the alignment.

Region 2 (positions 38–64 of the multiple alignment) was virtually identical between all of the Lancefield group C streptococci with the exception of the last nucleotide and nucleotides 41–43 of the alignment in *S. equisimilis* 585.

Region 3 (positions 64–233 of the alignment) was present in all of the *S. equi* isolates and two of the *S. zooepidemicus* isolates. This region was more than 97% homologous between all isolates sequenced. Two of the five *S. equi* isolates had an additional nucleotide in this region which was a guanidine between positions 135 and 136 of the alignment; there was no correlation between the presence or absence of this nucleotide and the presence or absence of the additional guanidine in region 1.

Region 4 (positions 234–306 of the alignment) was present and identical in all of the isolates. This region was identical with the sequence for the transfer RNA (tRNA) gene for alanine found in several other streptococcal 16S-23S intergenic spacers (see below).

Region 5 (positions 307–320 of the alignment) was a short sequence following the tRNA gene which was largely homologous between all isolates, although somewhat truncated in *S. equisimilis* and *S. dysgalactiae*. For *S. equi* and *S. zooepidemicus*, there were two variants, 5a and 5b; 5b was one base shorter

Table 3. *A 16S-23S intergenic typing scheme for Lancefield Group C streptococci based on combinations of regions of the 16-23S intergenic spacer illustrated in Figure 1*

16S-23S intergenic type*	Regions	Representative example
A1	1a, 5a, 6a, 7a	*S. zooepidemicus* 2809
A2	1b, 5a, 6a, 7b	*S. zooepidemicus* 3682
B1†	1a, 5b, 6b, 7c	*S. zooepidemicus* 3685
C1†	1a, 3, 5a, 6a, 7b	*S. zooepidemicus* 461
D1	1a, 3, 5b, 6b, 7c	*S. equi* (5/5 isolates tested)
D2	1b, 3, 5b, 6b, 7c	*S. zooepidemicus* K3
E1	1c, 5c, 6c	*S. dysgalactiae* and *S. equisimilis*

* The capital letter is determined by the variant of regions 6 with or without region 3. The numerical suffix is determined by the variant of region 1.

† Types B2 (1b, 5b, 6b, 7c) and C2 (1b, 3, 5a, 6a, 7b) might be expected to occur because region 1b was found together with regions 5a, 6a and 7c (although in different isolates) and region 1b was found together with regions 5b and 6b in other types.

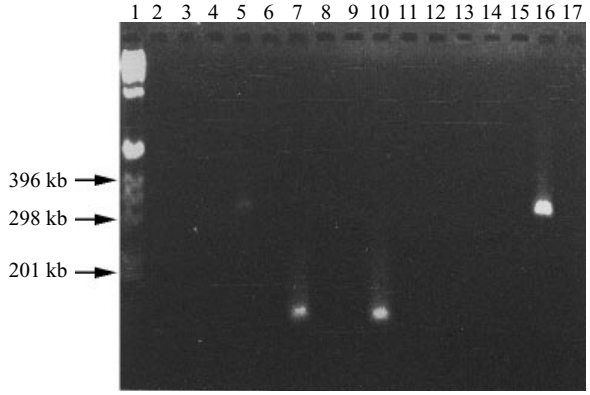

Fig. 3. Example of the PCR typing scheme for *S. zooepidemicus* based on the 16S-23S RNA gene intergenic spacer; agarose gel electrophoresis of PCR products from primers d+e, f, g and e (see Table 1) which amplify subsections of the Lancefield group C streptococcus intergenic spacer. Lane 1: 1 kb standards. Lanes 2–5: reactions with primers d+e and g (giving products from types A1 and C1 – see Table 2). Lanes 6–9: reactions with primers f and g (giving products from types A2 and C2). Lanes 10–13: reactions with primers d+e and h (giving products from types B1 and D1). Lanes 14–17: reactions with primers f and h (giving products from types B2 and D2). Lanes 2, 6, 10 and 14: *S. zooepidemicus* 3682 (type A2). Lanes 3, 7, 11 and 15: *S. zooepidemicus* 3685 (type B1). Lanes 4, 8, 12 and 16: *S. zooepidemicus* K3 (type D2). Lanes 5, 9, 13 and 17: *S. zooepidemicus* 461 (type C1).

and also differed from 5a at three positions. Possession of 5a and 5b coincided with region 6a or 6b, respectively (see below).

Region 6 (positions 321–491 of the alignment) was present in three clearly identifiable and distinct variants (6a, 6b and 6c) and each was represented in at least two isolates. Variant 6a was found in four of the *S. zooepidemicus* isolates, one of which also had region 3; variant 6a had an A/G variation in position 402 which coincided with the possession of variants 1a or 1b. Variant 6b (17 nucleotides longer than 6a) was found in all the *S. equi* isolates and two of the *S. zooepidemicus* isolates. One of the *S. zooepidemicus* isolates with 6b lacked region 3. Variant 6c, which was found in *S. dysgalactiae* and *S. equisimilis*, was just 17 bases long and, as such, was much smaller than 6a or 6b.

Region 7 (positions 492–505 of the alignment) was another short region showing some homology between the *S. equi* and *S. zooepidemicus* isolates, which was absent from *S. equisimilis* and *S. dysgalactiae* but which was present in 3 variants (7a, 7b and 7c) each represented in at least two isolates. Variant 7a was found only in the *S. zooepidemicus* isolates without region 3 but with region 6a. Variant 7b was found in *S. zooepidemicus* isolates with or without region 3 but only those with region 6a. Region 7c was found only in isolates of *S. equi* or *S. zooepidemicus* with region 6b. Those isolates of *S. zooepidemicus* with variants 7a and 7b possessed variants 1a and 1b, respectively.

Region 8 (positions 506–600 of the alignment) was 86% homologous between all the *S. equi*, *S. zooepidemicus*, *S. dysgalactiae* isolates and 2 of 3 of the *S. equisimilis* isolates; the third *S. equisimilis* isolate was 75% homologous in this region.

Region 9 (positions 601–650 of the alignment) was

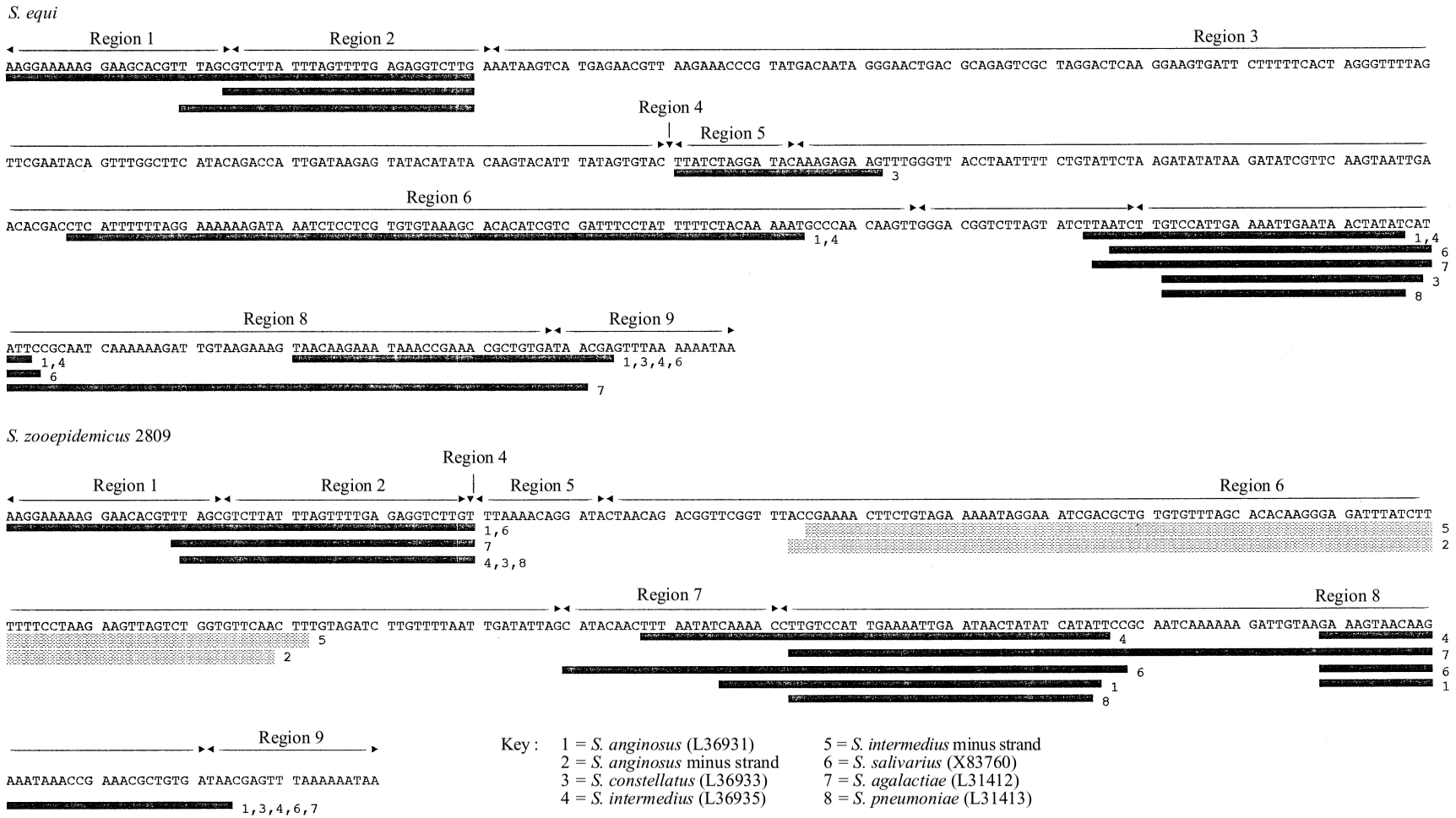

Fig. 4. Homologies (> 75%; homologies of less than 90% were due to short lengths of additional nucleotides between homologous regions) of the 16S-23S RNA gene intergenic spacers of *S. equi* and *S. zooepidemicus* 2809 (which together contain all variations of regions 1–3, 5–6 and 8–9) with the sequences on the DNA databases for ribosomal genes from other streptococci (accession number given in the key). The $tRNA^{ala}$ gene was excluded from the analysis since it is present in all of the streptococcal intergenic species.

a short region which preceded the 23S RNA gene. It was largely homologous between the isolates of *S. equi* and *S. zooepidemicus* and varied only in the number of adenosine nucleotides before the final TAA triplet. At 20–23 bases long it was 21–24 bases shorter than this region in *S. dysgalactiae* and 2 of 3 *S. equisimilis* isolates which showed little similarity to this region in *S. equi* and *S. zooepidemicus* apart from the string of adenosines which preceded a final TAA triplet; the third *S. equisimilis* isolate (585) more closely resembled *S. equi* and *S. zooepidemicus* in this region.

Each of the regions with variants within the 16S-23S intergenic spacer was present in at least 2 of the 15 isolates sequenced, irrespective of which other variant regions were possessed. Since the variant regions occurred so frequently within so few isolates and were present in different combinations with variants of other regions, the pattern of their possession by different isolates would lend itself to a typing system for Lancefield Group C streptococci, particularly as a subtyping scheme for *S. zooepidemicus*.

A proposed typing scheme based on the variant regions of the intergenic spacer is shown in Table 3 which contains representatives detected in this study. All possible combinations of the different variant regions might not be expected since, amongst the isolates tested, some variants were always found with some others. Variant region 5a was always present with 6a, 5b was always present with 6b and 7c but 5a, for example, was not found with 6b or 7c; variant regions 1a and 1b, although both found with 7c, were never found with 7b or 7a, respectively. The isolates of *S. equi* seemed universally to possess 1b together with 6b and 7c.

Four separate PCR tests were designed which would identify the 16S-23S intergenic type on the basis of the presence of regions 1a or 1b separately combined with 5a/6a or 5b/6b. It was predicted that only one of the four reactions would give a product and that the intergenic type could be identified on the basis of which reaction worked and the size of its product. An equal mixture of primers d and e was used for region 1a, primer f was used for region 1b, primer g for region 5a–6a and primer h for region 5b–6b. A PCR product of 144 or 148 base pairs was predicted for intergenic types A1, A2, B1 and B2 and of 317–323 base pairs for types C1, C2, D1 and D2 (which contain region 3). *S. equi* and *S. zooepidemicus* strains 2809, 3682, 3685, K3 and 461 were tested with these primers and they each gave a product of the correct size and from the combination of primers predicted from the sequences of their 16S-23S intergenic spacers; in each case, products did not result from PCRs with the other three combinations of primers (examples of four typing results are in Fig. 3). Twelve other *S. zooepidemicus* isolates from different horses on different premises were tested. Eight (6393, 4020, 3711, 6016, 6414, 148, 460, 46b, Table 1) belonged to type A1, two (1741, 1732) belonged to type D1 (the *S. equi* type) and one (1732) belonged to type B1. The twelfth isolate (145) typed as A2 and C1. There were two possible explanations for this. One was that the original culture was mixed and the other was that, unlike the other isolates, this one possessed different versions of the ribosomal RNA gene operons, one which contained the A2 intergenic type and the other the C1 type. The isolate was subcloned twice, taking care to subculture single colony picks, and was re-tested. Only growth giving the C1 intergenic type was recovered, suggesting that a contaminant of the A2 type was present in the original culture.

The sequence for the intergenic spacer of *S. equi* and *S. zooepidemicus* 2809 (which together contain all the variant regions found in these two bacteria) were screened for homology on the DNA sequence database. The sequences showing the strongest homologies were from the 16S-23S intergenic spacers of other streptococci. All these sequences contained the $tRNA^{ala}$ gene and none of them contained region 3 detected in all *S. equi* and some *S. zooepidemicus*. Homologous sequences were otherwise found to DNA just downstream of the 16S RNA gene (regions 1 and 2; none was entirely homologous to region 1a or 1b) and to region 8 (Fig. 4). A 75–90 nucleotide sequence shared with *S. anginosus* and *S. intermedius* was found in regions 6a and 6b in opposite orientations. Region 5b was shared with *S. constellatus*. The 30–35 nucleotides at the start of region 8 were 70% homologous to a sub-region of the minus strand of the 23S-5S RNA gene intergenic spacer of *S. salivarius* and *S. thermophilus*.

DISCUSSION

S. equi and *S. zooepidemicus* together and *S. equisimilis* and *S. dysgalactiae* formed two distinct groups in terms of both the sizes and sequences of their 16S-23S intergenic spacers. The structure of the spacers for *S. equi* and *S. zooepidemicus* appeared to be dominated largely by two versions of regions 1 and

5–6 coupled with the presence or absence of region 3. The occurrence of *S. zooepidemicus* isolates with a spacer identical to that of *S. equi* and *S. zooepidemicus* isolates with very distinct spacers or spacers with an intermediate relatedness to the '*S. equi* type' suggested that each type of intergenic spacer was derived by DNA recombination. The 75–90 nucleotides shared with *S. anginosus* and *S. intermedius* in regions 6a and 6b in opposite orientations support this interpretation.

Since regions 1b, 5a and 6a were not found in *S. equi* and regions 1a or 1b and 5a with 6a and 5b with 6b were each found in *S. zooepidemicus*, it seems likely that *S. zooepidemicus* was the archetypal species for *S. equi*. This was previously proposed on the basis of multilocus enzyme electrophoresis comparisons [16]. However, taken together the 16S-23S intergenic spacer and a small number of enzymes represent comparatively few loci and, as such, may not be representative of overall chromosomal relationships. Nonetheless, for most bacteria tested, the ribosomal operons have proved to be good indicators of the relatedness established by other taxonomic means. Additionally, the intergenic spacers in this case have the advantage that well-conserved sequences which might have been derived by recombination can help to determine relatedness and perhaps the order of evolution. *S. zooepidemicus* with a spacer containing regions 1b, 5a and 6a are unlikely to have been derived from *S. equi* whereas *S. equi* may have evolved from *S. zooepidemicus* with regions 1a, 5b and 6b.

The intergenic spacers of the Lancefield group C streptococci all contained the gene for $tRNA^{ala}$ as is the case in the other streptococci. Apart from this tRNA gene most homology was found in the DNA database with the intergenic spacers from *S. anginosus*, *S. constellatus*, *S. intermedius*, *S. agalactiae* and *S. salivarius*. The homologies found in region 6 taken together with those in regions 1, 2 and 8 for sequences in the intergenic spacers of *S. anginosus* and *S. intermedius* suggest that these bacteria have a common ancestry. Whether region 3 and other parts of region 6 were lost or acquired with evolution is uncertain. The homology of the sequences at the start of region 8 with the minus strand of the 23S-5S intergenic spacer of other streptococci is probably related to a role in secondary structure of the RNA gene operon transcript [18] and it can be predicted that these sequences will be found in the 23S-5S intergenic spacer of the Lancefield group C streptococci.

The ribosomal gene operon is usually present in bacteria in 7–10 copies per chromosome. In some bacteria there are different versions of the operon in the same cell giving more than one product when PCR amplified from primers in the 16S and 23S genes. Gel electrophoresis of the intergenic spacer amplified from the 16S and 23S RNA genes here revealed only one band for each isolate of Lancefield group C streptococcus and the sequence data were of a quality compatible with the presence of a single template. This suggests that the copies of this operon in the chromosome of these bacteria are homogeneous. The only isolate which appeared to have two types of intergenic spacer when amplified from regions 1 and 5–6 was found to be a mixed culture.

The different combinations of variant regions within the intergenic spacer of different *S. zooepidemicus* isolates and the conserved nature of the different variants of each region in isolates from widely different sources suggested the basis of a typing system. Such a system would aid epidemiological studies of the dynamics of *S. zooepidemicus* infection and the incidence of disease and commensal carriage in relation to infection history with different types. However, only five isolates were sequenced here and it is possible that other variants of the nine regions of the intergenic spacer remain to be characterized. Nonetheless, PCR tests of a further 12 distinct isolates on the basis of the variants in regions 1, 5 and 6 produced a clear result with all falling into a type already sequenced or one that was predicted to occur. This system would represent the first practical way of typing these bacteria since it is straightforward and depends on well-characterized reagents. In contrast, serotyping of *S. zooepidemicus* would depend on specific antisera to well-characterized antigens and, currently, these are not available.

ACKNOWLEDGEMENTS

The authors thank the HGMP Resource Centre for access to their computing facilities at Hinxton, Cambridge.

REFERENCES

1. Wood JLN, Burrell MH, Roberts CA, Chanter N, Shaw Y. Streptococci and *Pasteurella* spp. associated with disease of the equine lower respiratory tract. Eq Vet J 1993; **25**: 314–8.
2. Wood JLN, Dunn K, Chanter N, De Brauwere N. Persistent infection of *Streptococcus equi* and the epidemiology of strangles. Vet Rec 1993; **133**: 375.
3. Hoffman A, Laerant V, Prescott JF, Rosendal S,

Thorsen J. Association of microbiologic flora with clinical, endoscopic, and pulmonary cytologic findings in foals with distal respiratory tract infection. Am J Vet Res 1993; **54**: 1615–22.
4. Widders PR, Warner S, Huntington PJ. Immunisation of mares to control endometritis caused by *Streptococcus zooepidemicus*. Res Vet Science 1995; **58**: 75–81.
5. Giles RC, Donahue JM, Hong CB, *et al*. Causes of abortion, stillbirth, and perinatal death in horses: 3,527 cases (1986–1991). J Am Vet Med Assoc 1993; **203**: 1170–5.
6. Moore BO, Bryans JT. Type specific antigenicity of group C streptococci from diseases of the horse. Eq Infect Dis 1970; **II**: 231–8.
7. Timoney JF, Mukhtar MW. The protective M proteins of equine group C streptococci. Vet Microbiol 1993; **37**: 389–95.
8. Robinson JH, Kehoe MA. Group A streptococcal M proteins: virulence factors and protective antigens. Immunol Today 1992; **13**: 362–7.
9. Jensen MA, Webster J, Straus N. Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal DNA spacer polymorphisms. Appl Environ Microbiol 1993; **59**: 945–52.
10. Kostman JR, Alden MB, Mair M, Edlind TD, LiPuma JJ, Stull TL. A universal approach to bacterial molecular epidemiology by polymerase chain reaction ribotyping. J Infect Dis 1995; **171**: 204–8.
11. Kostman JR, Edlind TD, LiPuma JL, Stull TL. Molecular epidemiology of *Pseudomonas cepacia* determined by polymerase chain reaction ribotyping. J Clin Microbiol 1992; **30**: 2084–7.
12. Bentley RW, Leigh JA, Collins MD. Intrageneric structure of *Streptococcus* based on comparative analysis of small-subunit rRNA sequences. Int J Syst Bacteriol 1991; **41**: 487–94.
13. Jacobs JA, Schot CS, Bunschoten AE, Schouls LM. Rapid species identification of '*Streptococcus milleri*' strains by line blot hybridization: identification of a distinct 16S rRNA population closely related to *Streptococcus constellatus*. J Clin Microbiol 1996; **34**: 1717–21.
14. Whiley RA, Dule B, Hardie JM, Hall LMC. Heterogeneity among 16S-23S rRNA intergenic spacers of species within the '*Streptococcus milleri* group'. Microbiol 1995; **141**: 1461–7.
15. Farrow JAE, Collins MD. Taxonomic studies on streptococci of serologic groups C, G and L and possibly related taxa. System Appl Microbiol 1984; **5**: 483–93.
16. Jorm LR, Love DN, Bailley GD, McKay GM, Briscoe DA. Genetic structure of populations of beta haemolytic Lancefield group C streptococci from horses and their association with disease. Res Vet Science 1994; **57**: 292–9.
17. Barry T, Colleran G, Glennon M, Dunican LK, Gannon F. The 16S/23S ribosomal spacer region as a target for DNA probes to identify eubacteria. PCR Meth Appl 1991; **1**: 51–6.
18. Apirion D, Miczak A. RNA processing in prokaryotic cells. BioEssays 1993; **15**: 113–20.