JOURNAL OF BACTERIOLOGY, May 1999, p. 3144–3154 Vol. 181, No. 10
0021-9193/99/$04.00+0
Copyright © 1999, American Society for Microbiology. All Rights Reserved.

Genetic Diversity of the Streptococcal Competence (*com*) Gene Locus

ADRIAN M. WHATMORE,* VICTORIA A. BARCUS, AND CHRISTOPHER G. DOWSON

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom

Received 15 October 1998/Accepted 28 February 1999

The *com* operon of naturally transformable streptococcal species contains three genes, *comC*, *comD*, and *comE*, involved in the regulation of competence. The *comC* gene encodes a competence-stimulating peptide (CSP) thought to induce competence in the bacterial population at a critical extracellular concentration. The *comD* and *comE* genes are believed to encode the transmembrane histidine kinase and response regulator proteins, respectively, of a two-component regulator, with the *comD*-encoded protein being a receptor for CSP. Here we report on the genetic variability of *comC* and *comD* within *Streptococcus pneumoniae* isolates. Comparative analysis of sequence variations of *comC* and *comD* shows that, despite evidence for horizontal gene transfer at this locus and the lack of transformability of many *S. pneumoniae* strains in the laboratory, there is a clear correlation between the presence of a particular *comC* allele and the cognate *comD* allele. These findings effectively rule out the possibility that the presence of noncognate *comC* and *comD* alleles may be responsible for the inability to induce competence in many isolates and indicate the importance of a functional *com* pathway in these isolates. In addition, we describe a number of novel CSPs from disease-associated strains of *S. mitis* and *S. oralis*. The CSPs from these isolates are much more closely related to those from *S. pneumoniae* than to most CSPs previously reported from *S. mitis* and *S. oralis*, suggesting that these particular organisms may be a potential source of DNA in recombination events generating the mosaic structures commonly reported in genes of *S. pneumoniae* that are under strong selective pressure.

Many members of the oral streptococci are naturally transformable, being able to take up naked DNA from the extracellular environment (16). Homologous recombination of foreign DNA into the host chromosome following transformation is believed to play a major role in the evolution of these bacteria. This notion is illustrated by both the rapid emergence of penicillin resistance following the acquisition of low-affinity penicillin binding proteins (4) and evidence for the occurrence of frequent recombination events in the evolution of virulence factors in *Streptococcus pneumoniae* (5, 20). Competence for transformation in streptococci is not constitutive, as it is in *Neisseria* species (24), but is regulated by genes of the recently characterized *com* locus (18). The *com* operon contains three genes, *comC*, *comD*, and *comE*, encoding a competence-stimulating peptide (CSP), histidine kinase, and a response regulator, respectively (1, 7, 9, 19). Two genes located elsewhere on the chromosome, *comA* and *comB*, encode proteins responsible for the export of CSP from the cell (11, 27). CSP is thought to induce competence when a critical extracellular concentration is reached. The *comD*-encoded transmembrane histidine kinase is believed to be a receptor for the CSP and to phosphorylate a *comE*-encoded transcription regulator, producing an active form that up regulates both the *comCDE* operon and, presumably, a number of other genes involved in competence development.

Despite the fact that many *S. pneumoniae* strains appear untransformable under laboratory conditions, the CSP-encoding gene, *comC*, is thought to be ubiquitous (22). Sequencing studies recently identified two distinct alleles encoding *S. pneumoniae* CSP, *comC1* and *comC2* (21). It appears that the vast majority of strains carrying the *comC1* allele cannot be induced to competence with the *comC2*-encoded peptide CSP-2 and vice versa (21). Recently, it was suggested that all members of the mitis and anginosus phylogenetic groups of streptococci possess homologues of the *comCDE* operon (10). The CSP-encoding genes of several such isolates have been sequenced, revealing a range of structurally related but biologically distinct peptides both between species and among organisms recognized as a single species, such as *S. mitis* (9, 10, 21). A complete understanding of the specificity of competence induction is crucial for understanding the biology of *S. pneumoniae* and other naturally transformable streptococci, especially with regard to the horizontal transfer of antibiotic resistance and virulence markers within these organisms.

Here we describe a study designed to confirm and extend current knowledge of the genetic diversity of *comC* and the corresponding regulatory genes, *comD* and *comE*. The aims of the study were as follows. First, we aimed to characterize the *comC* genes of a diverse range of *S. pneumoniae* isolates in order to extend understanding of the genetic variation of the CSP, to link these findings to competence phenotype, and to provide a well-characterized set of isolates for further studies. Second, we aimed to characterize the 5′ region of *comD* (encoding the CSP receptor motif) from the same strains; while *comC* variation has been examined previously with a limited number of strains of *S. pneumoniae*, there are currently no data describing genetic variation of the corresponding *comD* genes. One feasible explanation for the reported lack of ability to transform at least 50% of strains tested with synthetic CSP (21), which has yet to be addressed, is the possibility of noncognate *comC* alleles and *comD* alleles in particular strains. A recent report that horizontal gene transfer may occur between the *com* operons of distinct streptococcal species, generating mosaic genes (10), highlights the need to examine this possibility. We therefore performed a comparative analysis of *comC* and *comD* alleles within a set of diverse *S. pneumoniae* isolates. Third, as an understanding of interspecies signalling may have important implications for understanding the population biology of the naturally transformable streptococci, we also identified a number of novel *comC* alleles from isolates of oral

* Corresponding author. Mailing address: Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom. Phone: 44-1203-528359. Fax: 44-1203-523701. E-mail: a.m.whatmore@warwick.ac.uk.

streptococci that are apparently closely related to pneumococci and that may act as donors in horizontal gene transfer events with pneumococci. Finally, we analyzed the complete *comCDE* region from a limited number of strains with variant *comC* alleles in order to add to the limited knowledge of diversity across the whole *com* operon.

MATERIALS AND METHODS

Strains. The designations and sources of the *S. pneumoniae* strains used in this study are shown in Table 1. Strains were routinely cultured on brain heart infusion (BHI) agar containing 5% (vol/vol) defibrinated sheep blood at 37°C in 5% CO_2. Strains were stored frozen at −80°C in BHI broth with 15% (vol/vol) glycerol until needed. The non-*S. pneumoniae* streptococcal strains, used in screening for a *comC* PCR product, were as follows (sources of strains are available on request): mitis group—*S. mitis* NCTC10712, NS51T, NCTC3166, NCTC1080, NCTC3165, HV51, PP53, K208, NCTC7864, NCTC11189, Col15, Col16, Col17, and Col18, *S. oralis* NCTC11427$^{\mathrm{T}}$, 20070NS, 20003NS, Col19, Col21, Col22, and Col24, *S. sanguis* 12088NS, 2397, NCTC7863$^{\mathrm{T}}$, and 13, *S. parasanguis* 55895, and *S. cristae* CR311 and CC5A; anginosus group—*S. milleri* NCTC10708, *S. constellatus* F436 and NMH4, *S. anginosus* NCTC11062, and *S. intermedius* 415-87, AM6425, and HW7, salivarius group—*S. vestibularis* NCTC12166$^{\mathrm{T}}$ and *S. salivarius* NCTC8618$^{\mathrm{T}}$; and other groups—group G streptococcus strains 91.2153, 91.2388, and 11555, group K streptococcus strain NCTC11389, *S. iniae* NCFB5389, *S. adjacens* X193, *S. cricetus* NCFB2720, *S. ferus* NCFB2721, *S. rattus* NCFB2723, *S. defectivus* DA4, *S. ceocorum* NCFB2674, and *S. sobrinus* NCFB2724.

Purification of streptococcal chromosomal DNA. Chromosomal DNA from each strain was obtained by harvesting the confluent overnight growth from two or three heavily inoculated agar plates into 1 ml of 50 mM Tris-HCl–10 mM EDTA (pH 8.0). Cell lysates were obtained by sequential addition (at 37°C) of 5 μl of lysozyme (10 mg ml^{-1}), (10 min), 5 μl of proteinase K (10 mg ml^{-1}) (30 min), and 40 μl of 20% (wt/vol) Sarkosyl. The clear viscous lysates were extracted once with phenol and once with chloroform and precipitated in 2.5 volumes of ethanol with 10% (vol/vol) 3 M sodium acetate (pH 5.2). The resulting DNA was washed with 70% (vol/vol) ethanol, resuspended in 10 mM Tris-HCl–1 mM EDTA (pH 7.5), and stored at −20°C.

Induction of competence by a CSP. Stimulation of competence by a synthetic CSP was tested essentially as described by Pozzi et al. (21). Bacteria were grown overnight on BHI agar supplemented with 4% (vol/vol) sterile defibrinated sheep blood at 37°C in an atmosphere of 5% CO_2 before being resuspended in BHI broth to an optical density at 620 nm of approximately 0.01. The cultures were incubated at 37°C until the optical density at 620 nm reached 0.4 to 0.5 (mid-exponential to late exponential phase) and were frozen in 15% glycerol at −80°C. For transformation, the frozen cultures were diluted 1:20 in C+y medium (25) containing 0.16% (wt/vol) bovine serum albumin, 0.01% (wt/vol) $CaCl_2$, and 100 ng of synthetic CSP-1 or CSP-2 ml^{-1} (both generously provided by D. Morrison, University of Illinois). Chromosomal DNA from a pneumococcus carrying a spectinomycin resistance cassette (14) was added to a concentration of 1 μg ml^{-1}. The transformation reaction mixture was kept at 37°C for 150 min before samples were plated on BHI blood agar supplemented with spectinomycin at 200 ng μl^{-1}. Approximately 5×10^7 cells were plated from each transformation reaction, and the induction of competence was judged to have taken place by the appearance of significant numbers (>500) of spectinomycin-resistant transformants on the selective plates after overnight incubation at 37°C.

Analysis of *comC* and *comD* allelic variations. The sequences of the oligonucleotides used as PCR primers in this study are shown in Table 2. The sequence flanking *comC* in strain R6 was obtained by performing inverse PCR to amplify a PCR product from a *Dra*I-generated chromosomal digest by use of primers 1 and 2 designed against the strain Rx CSP-encoding sequence (7). PCR products were cloned in pTAg (R & D Systems) according to the manufacturer's instructions. Plasmids were purified from the resulting clones with Wizard Plus Minipreps (Promega), and inserts were sequenced by use of primers corresponding to the vector sequence and novel primers to "walk" along the sequence. The *comC*-flanking sequence obtained was then used to design primers 3 and 4, which were used to amplify a *comC*-containing PCR product from chromosomal DNA. PCR was performed with standard parameters at an annealing temperature of 48°C for 32 cycles. PCR products were purified by passage through Microcon 100 columns (Amicon) and were sequenced directly by the cycle sequencing method with an ABI 373A automated sequencing system. Later, when the *comD* sequence had been determined, primer 3 was used in conjunction with primer 5 to amplify a region encompassing both *comC* and the 5′ variable region of *comD*, and the 5′ sequence of *comD* was determined directly in the same manner.

Cloning and sequencing of the *comCDE* operon. PCR products corresponding to the entire *comC* gene and the entire *comD* gene and most or all of *comE* were obtained with forward primer 6, corresponding to the upstream Arg-tRNA gene (9), or primer 3, located just upstream of *comC*, in conjunction with primer 7, corresponding to the downstream Glu-tRNA gene (9), or primer 8, located in the 3′ region of the *comE* gene. PCR products obtained with these primers were cloned in pTAg as described above, and the sequence of the entire insert was

TABLE 1. Designations, sources, and serotypes of bacterial isolates used in this study

Strain[a]	Source (date) of isolation[b]	Serotype[c]	Allele[d]
VA1	United States (1983)	19	*comC1*
Pn60 (143G)	Spain (1993)	19A	*comC1*
Pn16 (110K/70)	Papua New Guinea (NK)	42	*comC1*
Pn12 (53139/72)	Papua New Guinea (1972)	6	*comC1*
960	Oxford, UK (1994)	6B	*comC1*
670	Spain (1988)	6B	*comC1*
954	Oxford, UK (1994)	6B	*comC1*
964	Oxford, UK (1995)	6B	*comC1*
950	Oxford, UK (1995)	18C	*comC1*
Pn17 (46/68)	Papua New Guinea (1968)	46	*comC1*
878	Kenya (1990)	10F	*comC1*
Pn15 (N943/69)	Papua New Guinea (1969)	12	*comC1*
871	Kenya (1990)	18C	*comC1*
952	Oxford, UK (1995)	22F	*comC1*
29044	Czechoslovakia (1987)	14	*comC1*
R6	United States (1930)	NT	*comC1*
Rst7	NK (NK)	NK	*comC1*
Pn6 (9/122)	Brighton, UK (1988)	NK	*comC1*
Col14	UK (1993)	NT	*comC1*
Col6	UK (1993)	NT	*comC1*
881	Kenya (1990)	NT	*comC1*
Col8	UK (1993)	NT	*comC1*
Pn109	UK (1995)	1	*comC1*
873	Kenya (1990)	8	*comC1*
Pn112	Cambridge, UK (1995)	3	*comC1*
Pn111	Ashford, UK (1995)	3	*comC1*
7731	Equine isolate (1982)	3	*comC1*
Pn148	Equine isolate (1987)	3	*comC1*
Pn146	Equine isolate (1987)	3	*comC1*
41G	NK (NK)	NK	*comC2.1*
9858	Brighton, UK (1988)	NK	*comC2.1*
Col12	UK (1993)	NT	*comC2.1*
Pn5 (9/121)	Brighton, UK (1988)	NK	*comC2.1*
Col11	UK (1993)	NT	*comC2.1*
Col7	UK (1993)	NT	*comC2.1*
Col11	UK (1993)	NT	*comC2.1*
Col13	UK (1993)	NT	*comC2.1*
872	Kenya (1990)	15A	*comC2.1*
967	Oxford, UK (1994)	35F	*comC2.1*
Pn107	Oxford, UK (1995)	1	*comC2.1*
Pn108	Oxford, UK (1995)	1	*comC2.1*
Pn110	Norwich, UK (1995)	3	*comC2.1*
875	Kenya (1990)	16F	*comC2.1*
880	Kenya (1990)	13	*comC2.1*
953	Oxford, UK (1994)	6A	*comC2.1*
951	Oxford, UK (1994)	6A	*comC2.1*
85G	Mexico (1992)	23	*comC2.1*
859	Liverpool, UK (1996)	23F	*comC2.1*
81G	Mexico (1992)	23F	*comC2.1*
3G	Mexico (1993)	23F	*comC2.2*
Pn8 (DN87/629)	Oldham, UK (1987)	23	*comC2.2*
Pn24 (264)	Spain (NK)	23	*comC2.2*
577	UK (1987)	23	*comC2.2*
SP1	Spain (1989)	23	*comC2.2*
Pn25 (267)	Spain (1994)	23	*comC2.2*
F5	France (1992)	6B	*comC2.2*
Pn59 (138G)	Spain (1993)	NK	*comC3*
Pn13 (Kagnane/73)	Papua New Guinea (1973)	14	*comC4*
101/87	Spain (1987)	NT	*comC5*
874	Kenya (1990)	NT	*comC6.1*
Col19	UK (1993)		*comC6.2*
Col15	UK (1992)		*comC7*
Col18	UK (1993)		*comC7*
Col16	UK (1993)		*comC8*
NCTC10712			*comC9*

[a] In the vast majority of cases, both *comC* and 5′ *comD* sequences were examined. However, in the case of strains shown in boldface, only the sequence of *comC* was determined. All strains were *S. pneumoniae* except for Col19 (*S. oralis*) and Col15, Col18, Col16, and NCTC10712 (*S. mitis*).

[b] NK, not known. UK, United Kingdom.

[c] NT, nontypeable.

[d] The *comC* allele of each strain, as determined by sequencing, is indicated. Both alleles *comC2* and *comC6* are subdivided by a coding change within the leader peptide.

TABLE 2. Sequences of primers used in this study

Primer	Primer sequence (5′→3′)	Location or use
1	CAAAGCTACAAATCGTTCCAAT	Inverse PCR
2	GATGAGGTTGTCAAAATTCTTC	Inverse PCR
3	TGACAGTTGAGAGAATCTT	5′ to *comC*
4	CTTTTCTATTTATTTGACCT	3′ to *comC*
5	TAGTTCCAAATGGAAATA	5′ of *comD*
6	CATAGCTCAGCTGGATAGAGCA TTCGCCTTC	Arg-tRNA gene
7	GGCGGTGTCTTAACCCCTTGAC CAACGGACC	Glu-tRNA gene
8	WGAAATIGGWSAACGAT	3′ of *comE*

obtained by use of flanking primers corresponding to the vector sequence and a series of internal primers.

Phylogenetic analysis. Preliminary analysis and alignment of sequences were performed with the DNAStar package. Phylogenetic analysis and tree construction were performed with the program MEGA (15). Trees were constructed by the UPGMA method with the Jukes-Cantor correction (12), and the bootstrap confidence level of internal branches was estimated from 500 resamplings of the data.

Nucleotide sequence accession numbers. The EMBL accession numbers for the sequences reported in this paper are AJ240738 to AJ240795.

RESULTS

Sequences of *S. pneumoniae comC* alleles. When this work was being performed, the coding sequence of only a single *comC* gene was available (7). In order to obtain the flanking sequences of *comC*, thus facilitating the amplification and sequencing of the entire *comC* gene from multiple isolates of pneumococci, inverse PCR was performed with strain R6. By use of primers 1 and 2, a PCR product of approximately 2.6 kb was obtained from self-ligated, *Dra*I-digested chromosomal DNA and cloned into pTAg, and regions flanking *comC* were sequenced by use of both primers 1 and 2 and primers for the flanking vector sequence. On the basis of this sequence, primers 3 and 4 were designed in order to attempt to amplify *comC* from a diverse range of *S. pneumoniae* strains and from closely related streptococcal species.

PCR amplifications with primers 3 and 4 were performed with DNA isolated from a range of geographically, serotypically, and temporally diverse *S. pneumoniae* isolates (Table 1). A PCR product of about the expected size (337 bp) was obtained from all 60 *S. pneumoniae* strains examined. PCR products were sequenced in full, and six distinct mature CSPs, designated CSP-1 to CSP-6 (encoded by *comC1* to *comC6*, respectively), were identified. The predicted amino acid sequences of the CSPs are shown in Fig. 1A. The vast majority of strains of *S. pneumoniae* contained one of the two major alleles, *comC1* (29 of 60) or *comC2* (27 of 60), corresponding to CSP-1 or CSP-2, respectively, as previously described (21). However, four other alleles were detected, although only one example of each of these alleles was seen in the strains examined. The *comC3* allele of strain Pn59, isolated from Spain in 1993, is identical to a third pneumococcal allele recently reported (22). Pn13, an isolate from Papua New Guinea, possesses a unique but closely related allele (*comC4*) encoding a CSP with a deletion of a 3-amino-acid repeat relative to Pn59. The *comC5* allele was found only in strain 101/87, a Spanish atypical bile-insoluble *S. pneumoniae* strain (2), the *comC6* allele was found only in strain 874, a Kenyan strain isolated from an human immunodeficiency virus-seropositive, asymptomatic pneumococcal carrier.

Induction of competence by a CSP. The ability of synthetic CSP-1 and CSP-2 to induce competence in a subset of strains containing *comC1* or *comC2* was examined by monitoring the ability of each CSP to induce transformation to spectinomycin resistance (Table 3). As reported previously (21), the vast majority of strains could be induced to competence only by their congruent CSP, although two strains (Pn16 and VA1)

```
A
                                           ↓
CSP-1   (P)     MKNTVKLEQFVALKEKDLQKIKGG EMRLSKF----FRDFI----LQRKK
CSP-2.1 (P)     MKNTVKLEQFVALKEKDLQKIKGG EMRISRI----ILDFL----FLRKK
CSP-2.2 (P)     MKNTVKLEQFVALKEKDLQNIKGG EMRISRI----ILDFL----FLRKK
CSP-3   (P)     MKNTVKLEQFVALKEKDLQNIQGG EMRKMNEKSFNIFNFFNFFNFFRRR
CSP-4   (P)     MKNTVKLEQFVALKEKDLQNIQGG EMRKMNEKSFNIFNFFNF---FRRR
CSP-5   (P)     MKNTVKLEQFVDLKEKDLQEIQGG ESRLPKI----LLDFL----FLRKK
CSP-6.1 (P)     MKNTVKLEQFVALKEKDLQKIKGG EMRLPKI----LRDFI----FPRKK

B
                                           ↓
CSP-6.2 (O)     MKNTVKLEQFVALKEKDLQEIKGG EMRLPKI----LRDFI----FPRKK
CSP-7   (M)     MKNTVKLEQFVALKEKDLQEIRGG ESRVSRI----ILDFL----FQRKK
CSP-8   (M)     MKNTVKLEQFVALKEKDLQNIKGG ESRISDI----LLDFL----FQRKK
CSP-9   (M)     MKNTVKLEQFVALKEKDLQKIQGG EMRKPDG---------ALFNLFRRR
S. mitis B5     MKNTVKLEQFVALKEKDLQKIKGG ESRLPKI----RFDFI----FPRKK
S. mitis Hu8    MKNTVKLEQFVALKEKDLQKIQGG EMRKSNN---------NFFHFLRRI
S. mitis 12261  MKNTVNLDKFVELKEKDLQNIQGG EIRQTHN---------IFFNFFKRR
S. oralis 11427 MKNTEKLEQFKEVTEAELQEIRGG DKRLP--------YFFKHLFSNRTK
S. oralis 20066 MKNTVKLEQFKEVTEAELQEIRGG DWRIS--------ETIRNLIFPRRK
                **** .*..* ...* .**.*.** . *                .
```

FIG. 1. Alignments of the predicted amino acid sequences of the nine distinct CSPs characterized in this study. Asterisks represent amino acids conserved across all sequences, while dots represent conservative amino acid substitutions. Gaps in the alignment are represented by dashes. The predicted cleavage point of the mature peptide from the leader sequence, based on homology with other double-glycine-type leader peptides, is indicated by an arrow. The CSP-2 and CSP-6 groups are subdivided by coding changes in the leader peptide. The species of origin was *S. oralis* (O), *S. pneumoniae* (P), or *S. mitis* (M). (A) Sequences of the six distinct CSPs obtained from *S. pneumoniae* isolates. (B) Sequences obtained from isolates characterized as *S. mitis* or *S. oralis* alongside the sequences of three *S.mitis* CSPs and two *S. oralis* CSPs characterized by Håvarstein et al. (10).

TABLE 3. Induction of competence in strains containing *comC1* and *comC2* by artificial CSPs[a]

Strain	Allele	CSP1	CSP2
Col8	*comC1*	+	−
29044	*comC1*	+	−
R6	*comC1*	+	−
Pn15	*comC1*	+	−
670	*comC1*	−	−
7731	*comC1*	−	−
Pn109	*comC1*	−	−
Pn111	*comC1*	−	−
Pn112	*comC1*	−	−
Pn146	*comC1*	−	−
Pn148	*comC1*	−	−
Pn60	*comC1*	−	−
Pn6	*comC1*	−	−
Col14	*comC1*	−	−
VA1	*comC1*	+	+
Pn16	*comC1*	+	+
81G	*comC2*	−	+
Sp1	*comC2*	−	+
3G	*comC2*	−	+
Pn5	*comC2*	−	+
Pn25	*comC2*	−	−
Pn8	*comC2*	−	−
41G	*comC2*	−	−
Col11	*comC2*	−	−
Pn107	*comC2*	−	−
Pn108	*comC2*	−	−
Pn110	*comC2*	−	−
Col12	*comC2*	−	−
577	*comC2*	−	−

[a] The ability of a CSP to induce competence was monitored by the ability to induce the uptake of a spectinomycin resistance marker (+, competence induced, −, competence not induced).

appeared to develop competence in the presence of either CSP. However, and again as reported previously (21), less than 50% of strains were found to be rendered transformable even in the presence of their specific CSP.

Analysis of *comC* and *comD* allelic variations. As described above and as demonstrated previously (21), at least 50% of the strains appeared untransformable in the laboratory even in the presence of their specific CSP. There are many potential explanations for this phenomenon, notably, encapsulation (23, 26). However, one possible explanation, in view of the occurrence of multiple CSPs in *S. pneumoniae* and which has not been formally examined, is noncongruence (and therefore lack of binding) between the *comC*-encoded CSP and the *comD*-encoded CSP receptor protein within an individual strain. The data of Håvarstein et al. (10), indicating that horizontal gene transfer occurs between the *com* loci of distinct streptococcal species, provides supportive evidence that such a situation could arise. We investigated this possibility of noncognate *comC* and *comD* alleles within individual strains by examining the 5′ 384 bp of the *comD* sequence (encoding the putative CSP binding domain) of most of the strains from which the *comC* sequence had already been determined. This examination facilitated a comparative analysis of *comC* and *comD* sequences within a large number of diverse pneumococcal isolates.

The results of this analysis are presented in the form of two phylogenetic trees representing the sequences of the entire *comC* gene (Fig. 2A) and the 5′ end (first 384 bp) of the *comD* gene (Fig. 2B). The *comC* tree illustrates the relationship between the two major alleles, *comC1* (upper half of tree) and *comC2* (lower half of tree), which are subdivided by one silent change and one coding change (in the leader peptide), respectively; the amino acid sequences of the mature peptides are unaltered. The remaining sequences, representing strains containing *comC3* to *comC6*, are rather divergent from those of strains containing the major alleles. For the two major groupings, representing alleles *comC1* and *comC2*, allele groups for *comC* and *comD* remain entirely congruent: there is no mixing of the 27 *comC1* strains with the 22 *comC2* strains within the *comD* tree. Two distinct groups of *comD* sequences, *comD1* and *comD2*, are apparent; these groups are entirely congruent with the *comC1* and *comC2* groups. Like the *comC* sequences, the *comD* sequences of strains with *comC3* to *comC6* are all divergent from the two major groupings, *comD1* and *comD2*. Therefore, in spite of previous reports of evidence of horizontal gene transfer within the *com* locus (10), *comC1* and *comD1* group sequences and *comC2* and *comD2* group sequences are always paired within the diverse collection of *S. pneumoniae* strains examined in this study. Thus, the lack of transformability of many strains is due to factors other than horizontal gene transfer resulting in mismatched alleles of *comC* and *comD*.

Screening of other streptococci for *comC* homologues. It is thought that the horizontal gene transfer of DNA from oral streptococcal species to *S. pneumoniae* has played an important role in the evolution of this organism. We therefore used our primer set to screen isolates of a range of oral and other streptococci to investigate organisms with CSPs closely related to those of pneumococci, which may have an increased likelihood of acting as DNA donors. The results of this screening are summarized in Table 4. Only a small proportion of the *S. mitis* and *S. oralis* isolates screened yielded *comC* PCR products with this primer set, suggesting considerable intraspecies diversity within the *com* operon. PCR products were sequenced from four *S. mitis* strains (NCTC10712, Col15, Col16, and Col18) and a single *S. oralis* strain (Col19). The strains with the prefix Col were obtained from A. Efstratiou of the Central Public Health Laboratory, Colindale, England, and were unusual oral streptococci in that they were associated with chest infections and/or pneumonia. The strains were typed on the basis of their reactions with optochin, bile insolubility, quellung reaction, and biochemical profile. The predicted amino acid sequences of the CSPs are shown in Fig. 1B. Three distinct alleles were seen in the *S. mitis* strains. Strains Col15 and Col18 possess a novel allele, *comC7*. Strain Col16 possesses a unique and novel allele, *comC8*. However, *S. mitis* NCTC10712 was found to possess a CSP, encoded by *comC9*, unique in this study but identical to that previously reported from *S. mitis* B6 (10). The single CSP identified in an *S. oralis* strain (Col19) was found to be identical, apart from one amino acid change in the leader sequence, to that seen in pneumococcal isolate 874 described above (CSP-6).

CSP-7 and CSP-8, from organisms classified as *S. mitis*, are distinct from any of four CSPs recently reported from *S. mitis* strains (10), as demonstrated in the alignments shown in Fig. 1B. Likewise, Col19 is typed as *S. oralis* but has a CSP very different from the two previously reported *S. oralis* CSPs (10). The relationships between these sequences are demonstrated in the dendrogram shown in Fig. 3, constructed from CSP amino acid sequences. This dendrogram is not intended to be a phylogenetic interpretation but is intended to be used merely as a simple visual representation of sequence relationships. Thus, it can be seen that CSP-6, CSP-7, and CSP-8 are all much more closely related to the common pneumococcal CSPs, CSP-1 and CSP-2, than to most of the previously described *S. mitis* or *S. oralis* CSPs. One other previously reported *S. mitis* CSP, from strain B5 (10), also falls within this group. Likewise, pneumococcal CSP-3 and CSP-4 which, at least in our sample,

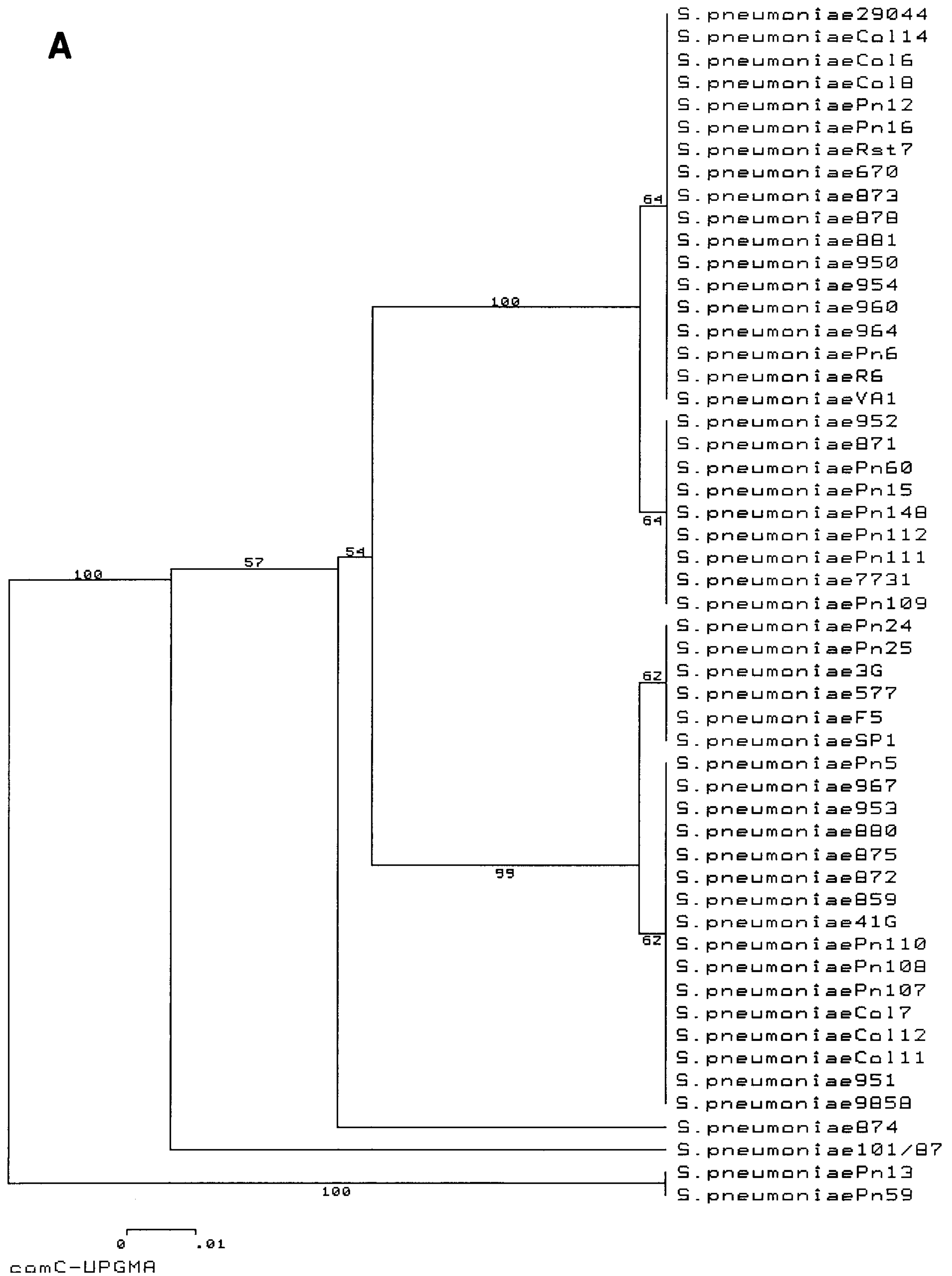

FIG. 2. UPGMA trees constructed with DNA sequence data from *comC* (A) and the 5′ 384 bp of *comD* (B) illustrating the complete congruence of CSP (*comC*) and receptor (*comD*) sequences. The numbers at internal branches represent the bootstrap confidence levels of particular branches estimated from 500 resamplings of the data set.

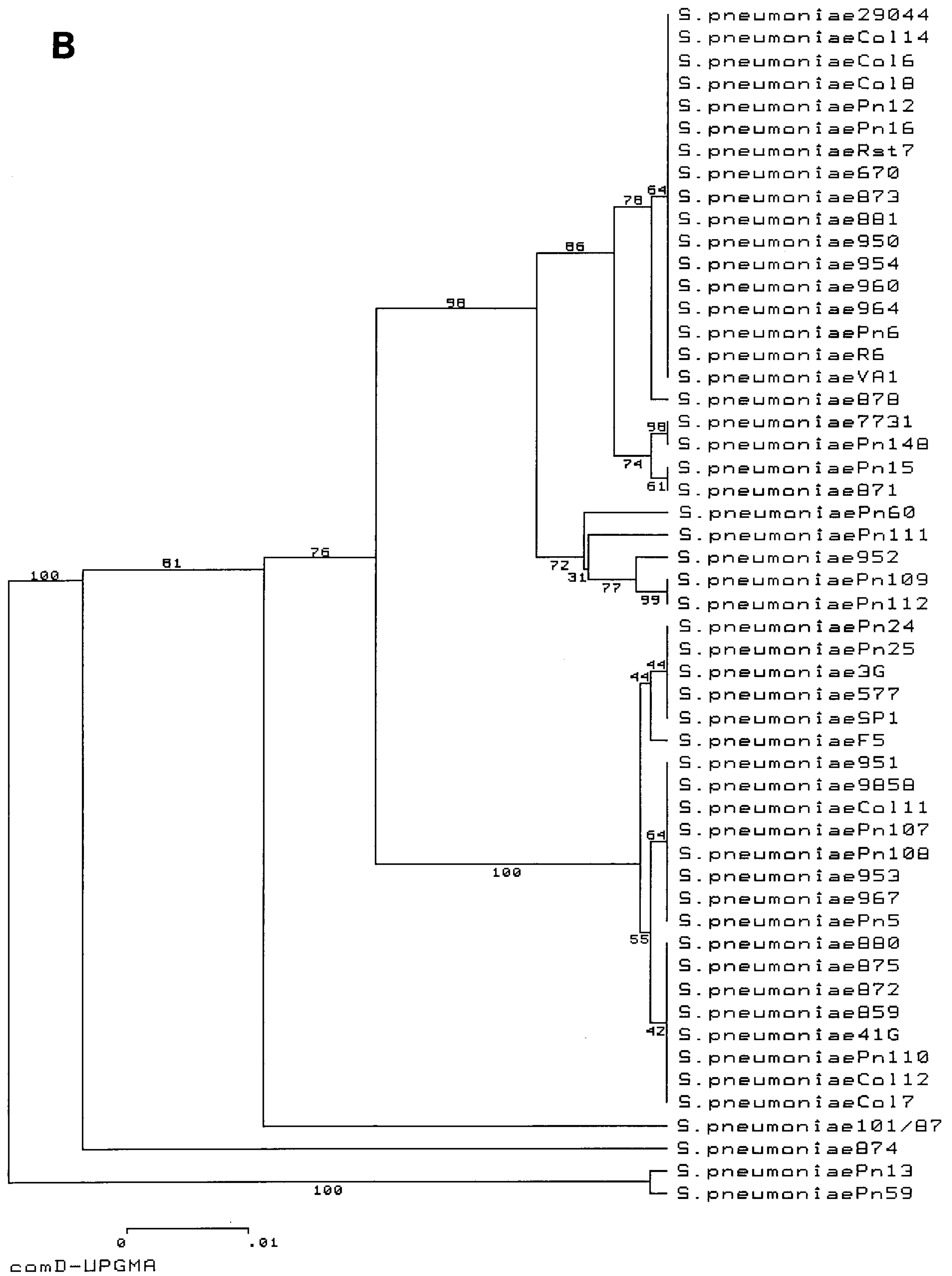

FIG. 2—*Continued.*

TABLE 4. Amplification of *comC* by PCR from streptococcal strains[a]

Group[b]	No. of PCR-positive strains/ total no. of strains tested
Mitis	
S. pneumoniae	60/60
S. mitis	4/14
S. oralis	1/7
S. sanguis	0/4
S. parasanguis	0/1
S. cristae	0/2
Anginosus	0/6
Salivarius	0/2
Other	0/15

[a] Purified streptococcal DNA was screened with primers 3 and 4.
[b] Full details of the strains examined are given in Materials and Methods.

appear much less common, appear most closely related to some *S. mitis* CSPs.

Variations across the whole *comCDE* locus. Primers corresponding to conserved flanking regions were used to amplify a PCR product representing the complete *comC* and *comD* genes and a large portion of the *comE* gene from six strains with divergent CSPs. The samples were cloned and sequenced in full in order to examine the nature and extent of genetic variations in the *comD* and *comE* genes of these strains. The strains examined were *S. pneumoniae* F5, containing the second major *comC* allele (*comC2*) and from which the sequences of *comD* and *comE* have not yet been reported; *S. pneumoniae* Pn59, *S. pneumoniae* Pn13, and atypical bile-insoluble *S. pneumoniae* 101/87 (2), each of which contains atypical pneumococcal CSPs; and *S. oralis* Col19 and *S. mitis* NCTC10712.

The levels of sequence divergence vary significantly across the *comD* and *comE* genes. As might be expected, the region encoding the N-terminal membrane-spanning domain of ComD, presumably containing the CSP receptor, is the most variable part of the *comD* gene, with up to 10.24% nucleotide divergence seen in a comparison of the 5′ 420 bp of the *S. pneumoniae* sequences. The region encoding the C-terminal kinase domain is much more conserved, displaying a maximum of 1.77% variation in the remaining 3′ 906 bp of *comD* in a comparison of the same sequences. The distribution of coding changes is illustrated by an alignment (Fig. 4) of the six sequences characterized here with the published *S. pneumoniae* Rx ComD sequence. Transmembrane segments and the possible topology of the ComD protein were determined with the program TmPred (25a). This analysis suggested that the ComD protein most likely contains seven transmembrane helices, in agreement with the structure predicted for the ComD protein of *Streptococcus gordonii* (9). The approximate locations (amino acids) of the transmembrane helices were predicted to be as follows: 3 to 20 (outside-inside), 31 to 51 (in-out), 47 to 74 (out-in), 85 to 104 (in-out), 121 to 140 (out-in), 161 to 180 (in-out), and 188 to 204 (out-in). The vast majority of the coding variation in *comD* is confined to regions corresponding to amino acids 1 to 63 and 114 to 151, suggesting that the ComD receptor motif is located within a surface-exposed region contained within these sequences. When the ComD sequences of strains containing the two major *S. pneumoniae comC* alleles, *comC1* and *comC2*, were compared virtually all variation was located in the region from amino acids 4 to 59; at least in strains containing these two major alleles, this result appears to narrow the location of the receptor segment to within this region. As might be expected, the *comE*-encoded response regulator protein is highly conserved even in strains with divergent *comC* and *comD* alleles (data not shown). Nucleotide divergence within *comE* ranged from 0.16 to 1.27% in the four *S. pneumoniae* strains examined in this study.

Horizontal gene transfer between *com* loci. A clear example of a recombination event generating mosaic structure can be seen by comparison of the sequences of *comCDE* of *S. pneumoniae* Pn13 and Pn59 with the previously published sequence of strain Rx. Figure 5 shows an alignment of polymorphic residues seen in a comparison of these sequences. It can readily be seen that the sequences of the Pn13 and Pn59 *comC* and 5′ *comD* regions are very similar to each other and divergent from the Rx sequence. However, beyond nucleotide 621 (located within the *comD* gene), the sequences of Pn13 and Pn59 diverge such that the Pn13 sequence is characteristic of the Rx sequence. This observed mosaic structure, indicative of horizontal gene transfer, is significant at a P level of <0.01 (chi-square test) (17). Similarly, a comparison of the divergent *comCD* regions of Pn13 and Pn59 with the *comCD* region of *S.*

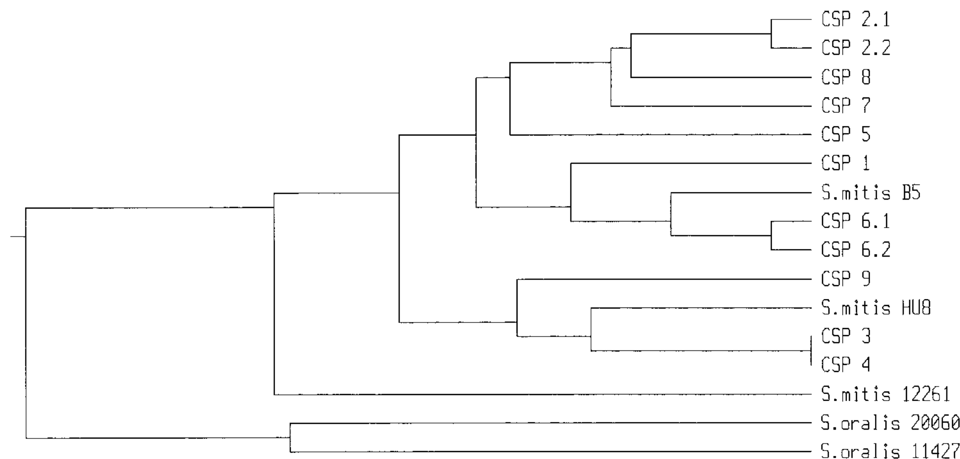

FIG. 3. UPGMA tree demonstrating the relationships between ComC (CSP) proteins characterized in this study (CSP-1 and CSP-9) and previously identified CSPs.

```
                    1                                                                     70
S.pneumoniaeRx      MDLFGFGTVI VHFLIISHSY HFICKGQINR KELFVFGAYT LLTEIVFDFP LYILYLDGLG IERFLFPLGL
S.pneumoniaeF5      ...L...... .......... RL...R.... ...YI..... ......LE.S F.L...KI.. ..........
S.pneumoniaePN13    ..F.LLVDL. LY........ RL..D..... ....F..... ......L... F.L....... .AT.......
S.pneumoniaePN59    ..F.LLVDL. LY........ RL..D..... ....F..... ......L... F.L....... .AT.......
S.oralisCOL19       .N........ G......... RL......KS ...YF..I.. ..V.A.LELS F.L...EIV. ..........
S.mitisNCTC10712    ..F.LVVDF. LY...L.... .L......KA ...Y...... ..V.A.LEL. F.F.S..... .VT.......
S.pneumoniae101/87  ......I... .......... RL........ ...Y...... ......LELF F.L.N...A. ..........

                    71                                                                   140
S.pneumoniaeRx      YSYFRWMKQY ERDRGLFLSL LLSLLYESTH NFLSVTFSSI TGDNFVLQYH FPFFFVVTVL TYFVTLKIIY
S.pneumoniaeF5      .......... .......... .......... .......... ...S...... .......... ..........
S.pneumoniaePN13    .......... .....R.... .......... .......... .........Y GL........ ..........
S.pneumoniaePN59    .......... .......... .......... .......... .........Y GL........ ..........
S.oralisCOL19       ......I... .......... .......... .......... ......S... D......... ....V....H
S.mitisNCTC10712    ......I... .......... .......... .......... .........Y GL..LAIST. ....V....H
S.pneumoniae101/87  ....L.I... .......... .......... .......... ......S... A....A.... ..I.IVT..R

                    141                                                                  210
S.pneumoniaeRx      YFHLELAYFD EDYLYPFLKK VFFALLLLHI VSFVSDMVST IKHLNSFGSI LSSIVFISLL LTFFAMNSHK
S.pneumoniaeF5      .......... K......... .......... .......... .......... .......... ..........
S.pneumoniaePN13    .......... K......... .......... .......... .......... .......... ..........
S.pneumoniaePN59    .......... .......... .......... .......... .......... .......... ..........
S.oralisCOL19       ......S... K......... .......... .......... .......... .......... ..........
S.mitisNCTC10712    .......... K......... .........V A......... .......... .......... .I......Y.
S.pneumoniae101/87  ......N... K......... .......... .......... .......... .......... ..........

                    211                                        #                         280
S.pneumoniaeRx      VQMEKEIALK QKKFEQKHLQ NYTDEIVGLY NEIRGFRHDY AGMLVSMQMA IDSGNLQEID RIYNEVLVKA
S.pneumoniaeF5      .......... R......... .......... .......... .......... .......... ..........
S.pneumoniaePN13    .......... .......... .......... .......... .......... .......... ..........
S.pneumoniaePN59    .......... .........R .......... .......... .......... .......... .V........
S.oralisCOL19       E......... .......... .......... .......... .......... ....D..... ..........
S.mitisNCTC10712    E......... .......... .......... .......... .......... .......... .V........
S.pneumoniae101/87  G......... .......... .......... .......... .......... ....D..... .V........

                    281                                                                  350
S.pneumoniaeRx      NHKLRSDKYT YFDLNNIEDS ALRSLVAQSI VYARNNGVEF TLEVKDTITK LPIELLDLVR IMSVLLNNAV
S.pneumoniaeF5      .......... .......... .......... .......... .......... .......... ..........
S.pneumoniaePN13    .......... .......... .......... .......... .......... .......... ..........
S.pneumoniaePN59    .......... .......... .......... .......... .......... .......... ..........
S.oralisCOL19       .......... .......... .......... .......... .......... .......... ..........
S.mitisNCTC10712    .......... .......... .......... .......... .........E .S........ ..........
S.pneumoniae101/87  ...?...... .......... .......... .......... ......I..R ...D...... ..........

                    351                                                                  420
S.pneumoniaeRx      EGSADSYKKQ MEVAVIKMET ETVIVIQNSC KMTMTPSGDL FALGFSTKGR NRGVGLNNVK ELLDKYNNII
S.pneumoniaeF5      .......... .......... ...M...... .......... .......... .......... ..........
S.pneumoniaePN13    .......... .......... .......... .......... .......... .......... ..........
S.pneumoniaePN59    .......... .......... .......... .?........ .......... .......... ..........
S.oralisCOL19       .......... .......... .......... .......... .......... .......... ..........
S.mitisNCTC10712    .......... ........DV .......... ....N..E.. .I........ ...L...... .I....D...
S.pneumoniae101/87  .......... .......... .......... .......... .......... .......... ..........

                    421                  442
S.pneumoniaeRx      LETEMEGSTF RQIIRFKREF E*
S.pneumoniaeF5      .......... .......... ..
S.pneumoniaePN13    .......... .......... ..
S.pneumoniaePN59    .......... ........K. ..
S.oralisCOL19       .......... .......... ..
S.mitisNCTC10712    ......D... .......... ..
S.pneumoniae101/87  .......... .........V ..
```

FIG. 4. Alignment of the predicted amino acid sequences of the ComD proteins from streptococcal strains in comparison to the published *S. pneumoniae* Rx ComD sequence (19). Only residues which differ from the Rx sequence are shown. Identical residues are shown by dots. The active-site histidine residue of ComD is indicated by a number sign, while the predicted stop codon is indicated by an asterisk.

```
                                                 1111111111111111111111111111111
                         5668888888889999999999900000000111222222233333333444555
                         4042345678901234567890123456801401245601234567836712 3

S.pneumoniae Rx          AGA--------------------------TGCAACCGGATATTTTACAAAAGTGA
S.pneumoniae Pn13        GTCAAAATGAATGAAAAGTCCTTTAATACTTTTAATTC----------GGA---
S.pneumoniae Pn59        GTCAAAATGAATGAAAAGTCCTTTAATACTTTTAATTC.A...CTTTGGA---

                                                                               1
                         1111111111111111111112222222222333333344455555555568890
                         5555666666788889999999900002334702255772390223336925880
                         4567135699234801234570134426739375878714948125001914 1

S.pneumoniae Rx          AATAGAAGTATGGTGGACGGTTGTCCATGGTAAAGATTCCTACTTCACGATAA
S.pneumoniae Pn13        ----ATGC-TCTTATAGATTGATATTGAATCCCCCCCGTTCGTGGTGTA....
S.pneumoniae Pn59        ----ATGC-TCTTATAGATTGATATTGAATCCCCCCC.TTCGTGGTGT.GCGG


                         1111111111111111111111112
                         0000112445677777778899990
                         2267881287501358893538897
                         0273471458151960361051705

S.pneumoniae Rx          TGCTGTTTGTGGGATAAGTAGAGAA          comC  1-150
S.pneumoniae Pn13        ...G..................A....          comD  171-1496
S.pneumoniae Pn59        CAT.AGCCACATAGCGCACC.GACG          comE  1493-
```

FIG. 5. Distribution of polymorphic sites among the sequences of the *com* operons of *S. pneumoniae* Rx, Pn59, and Pn13. Numbering, shown above the sequence, begins at the *comC* start codon, with residues identical to those in strain Rx shown by dots and gaps in the sequence alignment shown by dashes.

mitis (NCTC10712) also suggests that the polymorphisms seen in Pn13 and Pn59 may be the result of horizontal transfer of DNA encoding both the CSP and the congruent receptor region of ComD from an *S. mitis* strain. These data support the previous observation that horizontal gene transfer, involving acquisition of the CSP and congruent receptor protein, has occurred during the evolution of the *com* locus (10).

DISCUSSION

The data presented here characterize nine distinct CSP variants; five of these represent novel CSP variants not previously reported. However, all the CSP variants conform to previously described characteristics, being small cationic peptides with a negatively charged N terminus, with a positively charged C terminus, and possessing double-glycine-type leader peptides (6, 8). The leader peptides are highly conserved among all the CSPs, with almost every residue being identical or showing a conservative substitution (Fig. 1). Similarly, the negatively charged N-terminal amino acid and the arginine residue of the mature peptide are completely conserved as, to a large extent, are the three C-terminal arginine and/or lysine residues. In contrast, the central regions of the mature peptide are highly variable, suggesting that these are the regions which confer peptide specificity and which interact with the ComD receptor, while the flanking regions are functionally constrained.

Six distinct CSPs were found in isolates which have been classified as *S. pneumoniae*. The vast majority of *S. pneumoniae* isolates possess one of two major CSP variants reported previously (CSP-1 or CSP-2). However, the occurrence of closely related CSP-3 and CSP-4 in strains which were isolated about 20 years apart in different parts of the world but which are, as far as we are aware, representative of typical pneumococci suggests that a significant subpopulation of pneumococci which possess these very distinct CSPs may exist. CSP-3 was reported in a single South African isolate of a distinct serotype (22), indicative of probable long-term and worldwide prevalence of such isolates. The biological significance of these isolates is unclear, although one might expect that some genetic isolation could develop because of the possibility of reduced genetic exchange with other pneumococci.

Two other strains, 101/87 and 874, classified as *S. pneumoniae* and possessing distinct CSPs, CSP-5 and CSP-6, may represent either atypical or incorrectly classified organisms. In this respect, one should remember that distinguishing pneumococci and other members of the oral streptococci in the clinical microbiology laboratory can be problematic. Strain 874 was obtained by us as part of a study of the population genetic structure of carried pneumococcal isolates and was classified as *S. pneumoniae* by conventional criteria, including bile solubility, optochin sensitivity, and the results of an Accuprobe *S. pneumoniae* culture identification test (Gen-Probe). However, almost all housekeeping genes examined in this organism are distinct from those of other pneumococci that we have examined (unpublished data), raising doubts about the classification of this strain as *S. pneumoniae*. Demonstration that the CSP carried by this organism is virtually identical to that carried by an *S. oralis* isolate, Col19, supports such doubts. Strain 101/87 is a well-studied organism isolated from the blood of a patient with pneumonia; although initially classified as *S. mitis* on the basis of biochemical tests indicating a nontypeable alpha-hemolytic strain resistant to bile and optochin, it was eventually classified as an atypical *S. pneumoniae* strain on the basis of the use of specific DNA probes (2). Given the background of these strains, it was therefore not surprising to isolate novel CSPs from them.

There have been no previous studies of the genetic diversity of the *comD* gene of *S. pneumoniae*. Sequencing of the 5′ region of *comD* from 53 strains for which *comC* had already been characterized allowed a comparative analysis of the relationship between the alleles of these two genes. The region sequenced encompasses the most variable segment of *comD* and indeed contains virtually all of the diversity seen when the complete *comD* sequences of strains containing the two major

comC alleles, *comC1* and *comC2*, are examined (Fig. 4). Thus, the approach of sequencing only the 5′ region of *comD* is a valid one. Despite the inability in both this study and previous studies to induce competence even with the cognate synthetic CSP in up to 50% of strains, a complete correlation between *comC* alleles and apparently matching *comD* alleles was found in these strains. All CSP-1-containing strains had closely related *comD* alleles (*comD1* group) which were substantially different from the *comD* alleles found in all CSP-2-containing strains (*comD2* group). Thus, the inability to induce competence in many strains cannot be due to the lack of a ComD receptor protein which is cognate for the CSP. Other factors must therefore be responsible for the observed nontransformability of many strains; these could include capsule production (23, 26) or nonoptimal in vitro culture conditions.

In spite of the complete congruence of the *comC* and *comD* alleles seen here (i.e., *comC1* with the *comD1* group and *comC2* with the *comD2* group), both this study and previous studies have revealed data illustrating that horizontal gene transfer was involved in the evolution of the *com* locus. Håvarstein et al. (10) recently examined the *comCDE* operon from several representatives of the mitis and anginosus phylogenetic groups of streptococci (13) and reported three instances of apparent mosaic structure in comparisons of distinct *Streptococcus gordonii* or distinct milleri group isolates, indicative of horizontal gene transfer. In agreement with the findings reported above of cognate *comC* and *comD* alleles in all strains examined, all of the putative recombination events appeared to involve the transfer of *comC* and the 5′ region of *comD* (i.e., the receptor-encoding segment) in conjunction. Presumably, horizontal gene transfer events generating noncongruent *comC* and *comD* alleles must occur in nature but are either rare or rapidly selected against.

In the second part of this study, we used our primer set flanking *comC* to screen isolates of streptococcal species for organisms which might produce closely related CSPs. The rationale behind this strategy was that the horizontal gene transfer of DNA from other streptococcal species is believed to play an important role in generating the genetic diversity of *S. pneumoniae*. Mosaic structures which result from horizontal gene transfer have been demonstrated in a number of genes. For example, *pbp* genes, encoding penicillin binding proteins, display mosaic structures which impart β-lactam resistance and which are thought to result from the horizontal transfer of DNA from oral streptococci to pneumococci (4). Similarily, a number of genes encoding putative virulence factors of *S. pneumoniae* possess mosaic structures believed to result from the horizontal transfer of DNA originating in oral streptococci (5, 20).

In some of the instances of horizontal gene transfer reported above, the exact DNA donor has not been identified (3). Organisms containing closely related CSPs might be expected to act as donors in such events simply because their own CSP might be able to induce some degree of competence in *S. pneumoniae*, thus increasing the probability of uptake of their DNA. Despite the fact that Håvarstein et al. (10) have recently reported the presence of the *com* operon in all members of the mitis and anginosus phylogenetic groups, the primer set used here appeared to successfully amplify PCR products only from strains with very closely related *com* operons. Thus, for example, Håvarstein et al. (10) amplified the *com* operon by using primers to flanking tRNA genes from strains such as *Streptococcus oralis* NCTC11427 and *Streptococcus sanguis* NCTC7863, which were PCR negative in our study. A comparison of the recently published sequences of these *com* operons demonstrates that these sequences have only 58% homology and 74% homology, respectively, with the primer 3 sequence; thus, no PCR product would be expected under the rather stringent amplification conditions used in this study. However, our PCR primers did successfully yield PCR products from a small subset of *S. mitis* and *S. oralis* isolates and, as predicted, the *comC* sequences obtained from these organisms were generally more closely related to *S. pneumoniae* sequences than to other, previously determined *S. mitis* and *S. oralis* sequences. This apparent subset of *S. mitis* and *S. oralis* isolates may represent an important group of organisms, and it is possible that they are a source of DNA generating mosaic genes following horizontal transfer of DNA. We are now actively examining this possibility. These findings also support the growing belief that *S. mitis* and *S. oralis* may actually be rather poorly defined species containing a wide range of rather disparate organisms.

ACKNOWLEDGMENTS

This work was supported by grants from the Wellcome Trust, the Medical Research Council (United Kingdom), and the National Institutes of Health (United States).

We are grateful to Paul Pickerill and Maggie Yeo for expert technical assistance and to A. Efstratiou of the Central Public Health Laboratory, Colindale, United Kingdom, for providing isolates of *S. oralis* and *S. mitis*.

REFERENCES

1. **Cheng, Q., E. A. Campbell, A. M. Naughton, S. Johnson, and H. R. Masure.** 1997. The *com* locus controls genetic transformation in *Streptococcus pneumoniae*. Mol. Microbiol. **23:**683–692.
2. **Diaz, E., R. López, and J. L. García.** 1992. Role of the major pneumococcal autolysin in the atypical response of a clinical isolate of *Streptococcus pneumoniae*. J. Bacteriol. **174:**5508–5515.
3. **Dowson, C. G., T. J. Coffey, C. Kell, and R. A. Whiley.** 1993. Evolution of penicillin resistance in *Streptococcus pneumoniae*; the role of *Streptococcus mitis* in the formation of a low affinity PBP2B in *S. pneumoniae*. Mol. Microbiol. **9:**635–643.
4. **Dowson, C. G., T. J. Coffey, and B. G. Spratt.** 1994. Origin and molecular epidemiology of penicillin-binding-protein-mediated resistance to β-lactam antibiotics. Trends Microbiol. **2:**361–366.
5. **Dowson, C. G., V. Barcus, S. King, P. Pickerill, A. Whatmore, and M. Yeo.** 1997. Horizontal gene transfer and the evolution of resistance and virulence determinants in *Streptococcus*. J. Appl. Microbiol. **83:**42S–51S.
6. **Håvarstein, L. S., H. Holo, and I. F. Nes.** 1994. The leader peptide of colicin V shares consensus sequences with leader peptides that are common among peptide bacteriocins produced by Gram positive bacteria. Microbiology **140:**2383–2389.
7. **Håvarstein, L. S., G. Coomaraswamy, and D. A. Morrison.** 1995. An unmodified heptadecapeptide induces competence for genetic transformation in *Streptococcus pneumoniae*. Proc. Natl. Acad. Sci. USA **92:**11140–11144.
8. **Håvarstein, L. S., D. B. Diep, and I. F. Nes.** 1995. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. Mol. Microbiol. **16:**229–240.
9. **Håvarstein, L. S., P. Gaustad, I. F. Nes, and D. A. Morrison.** 1996. Identification of the streptococcal competence-pheromone receptor. Mol. Microbiol. **21:**863–869.
10. **Håvarstein, L. S., R. Hakenbeck, and P. Gaustad.** 1997. Natural competence in the genus *Streptococcus*: evidence that streptococci can change pherotype by interspecies recombinational exchanges. J. Bacteriol. **179:**6589–6594.
11. **Hui, F., and D. A. Morrison.** 1991. Competence for transformation in *Streptococcus pneumoniae*: nucleotide sequence analysis shows *comA*, a gene required for competence induction, to be a member of the bacterial ATP-dependent transport protein family. J. Bacteriol. **173:**372–381.
12. **Jukes, T. H., and C. R. Cantor.** 1969. Evolution of protein molecules, p. 21–132. *In* H. N. Munro (ed.), Mammalian protein metabolism. Academic Press, Inc., New York, N.Y.
13. **Kawamura, Y., X.-G. Hou, F. Sultana, H. Miura, and T. Ezaki.** 1995. Determination of 16S rRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptococcus*. Int. J. Syst. Bacteriol. **45:**406–408.
14. **Kell, C. M., U. K. Sharma, C. G. Dowson, C. Town, T. S. Balganesh, and B. G. Spratt.** 1993. Deletion analysis of the essentiality of penicillin-binding proteins 1A, 2B and 2X of Streptococcus pneumoniae. FEMS Microbiol. Lett. **106:**171–176.
15. **Kumar, S., K. Tamura, and M. Nei.** 1993. MEGA: Molecular Evolutionary Genetic Analysis software for microcomputers. CABIOS **10:**189–191.

16. **Lorenz, M. G., and W. Wackernagel.** 1994. Bacterial gene transfer by natural genetic transformation in the environment. Microbiol. Rev. **58:**563–602.
17. **Maynard Smith, J.** 1992. Analysing the mosaic structure of genes. J. Mol. Evol. **34:**126–129.
18. **Morrison, D. A.** 1997. Streptococcal competence for genetic transformation: regulation by peptide pheromones. Microb. Drug Resist. **3:**27–37.
19. **Pestova, E. V., L. S. Håvarstein, and D. A. Morrison.** 1996. Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an auto-induced peptide pheromone and a two-component regulatory system. Mol. Microbiol. **21:**853–862.
20. **Poulson, K., J. Reinholdt, C. Jespersgaard, K. Boye, T. A. Brown, M. Hauge, and M. Kilian.** 1998. A comprehensive genetic study of streptococcal immunoglobulin A1 proteases: evidence for recombination within and between species. Infect. Immun. **66:**181–190.
21. **Pozzi, G., L. Masala, F. Iannelli, R. Manganelli, L. S. Håvarstein, L. Piccoli, D. Simon, and D. A. Morrison.** 1996. Competence for genetic transformation in encapsulated strains of *Streptococcus pneumoniae*: two allelic variants of the peptide pheromone. J. Bacteriol. **178:**6087–6090.
22. **Ramirez, M., D. A. Morrison, and A. Tomasz.** 1997. Ubiquitous distribution of the competence related genes *comA* and *comC* among isolates of *Streptococcus pneumoniae*. Microb. Drug Resist. **3:**39–52.
23. **Ravin, A. W.** 1959. Reciprocal capsular transformation of pneumococci. J. Bacteriol. **77:**296–309.
24. **Solomon, J. M., and A. D. Grossman.** 1996. Who's competent and when: regulation of natural genetic competence in bacteria. Trends Genet. **12:**150–155.
25. **Tomasz, A., and R. D. Hotchkiss.** 1964. Regulation of the transformability of pneumococcal cultures by macromolecular cell products. Proc. Natl. Acad. Sci. USA **51:**480–487.
25a. **Worley, K. C.** BCM Search Launcher. 18 September 1998, revision date. [Online.] TmPred program. http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html. [9 April 1999, last date accessed.]
26. **Yother, J., L. S. McDaniel, and D. E. Briles.** 1986. Transformation of encapsulated *Streptococcus pneumoniae*. J. Bacteriol. **168:**1463–1465.
27. **Zhou, L., F. M. Hui, and D. A. Morrison.** 1995. Competence for genetic transformation in *Streptococcus pneumoniae*: organisation of a regulatory locus with homology to two lactococcin A secretion genes. Gene **153:**25–31.