International Journal of Systematic Bacteriology (1999), **49**, 1181–1192 Printed in Great Britain

Species identification and phylogenetic relationships based on partial HSP60 gene sequences within the genus *Staphylococcus*

Anita Y. C. Kwok,[1,3,5] Shey-Chiang Su,[1,2] Robert P. Reynolds,[1,2,3] Sue J. Bay,[4,5] Yossef Av-Gay,[1,3] Norman J. Dovichi[4,5] and Anthony W. Chow[1,2,3,5]

Author for correspondence: Anthony W. Chow. Tel: +1 604 875 4148. Fax: +1 604 875 4013. e-mail: tonychow@interchange.ubc.ca

[1,2] Departments of Medicine[1] and Microbiology and Immunology[2], University of British Columbia, Vancouver, BC, Canada

[3] Vancouver Hospital Health Sciences Centre, 2733 Heather Street, Vancouver, BC, Canada

[4] Department of Chemistry, University of Alberta, Edmonton, AB, Canada

[5] Division of Infectious Diseases, G. F. Strong Research Laboratories, and the Canadian Bacterial Diseases Network, Vancouver, BC, Canada

The phylogenetic relationships among 36 validly described species or subspecies within the genus *Staphylococcus* were investigated by cloning and sequencing their 60 kDa heat-shock protein (HSP60) genes using a set of universal degenerate HSP60 PCR primers. The cloned partial HSP60 DNA sequences from nine *Staphylococcus aureus* strains were highly conserved (97–100% DNA sequence similarity; mean 98%), indicating that the HSP60 gene of multiple isolates within the same species have little microheterogeneity. At the subspecies level, DNA sequence similarity among members of *S. aureus*, *Staphylococcus schleiferi*, *Staphylococcus cohnii* and *Staphylococcus capitis* ranged from 91 to 98%. At the interspecies level, sequence similarity among 23 distinct species of staphylococci ranged from 74 to 93% (mean 82%). By comparison, the highest sequence similarity of *Bacillus subtilis* and *Escherichia coli* with members within the genus *Staphylococcus* was only 70 and 59%, respectively. Importantly, phylogenetic analysis based on the neighbour-joining distance method revealed remarkable concordance between the tree derived from partial HSP60 gene sequences and that based on genomic DNA–DNA hybridization, while 16S rRNA gene sequences correlated less well. The results demonstrate that DNA sequences from the highly conserved and ubiquitous HSP60 gene offer a convenient and accurate tool for species-specific identification and phylogenetic analysis of staphylococci.

Keywords: *Staphylococcus*, HSP60 gene sequences, 16S rRNA gene sequences, DNA–DNA hybridization, phylogenetic analysis

INTRODUCTION

According to the current List of Bacterial Names with Standing in Nomenclature (Euzéby, 1997; updated 18 January 1999), the genus *Staphylococcus* comprises 36 species, nine of which also contain subdivisions with subspecies designation. The majority of these staphylococci are coagulase-negative, with the exception of *Staphylococcus aureus*, *Staphylococcus intermedius*, *Staphylococcus delphini*, *Staphylococcus schleiferi* subsp. *coagulans* and some strains of *Staphylococcus hyicus* (Kloos & Bannerman, 1995). Although previously considered to be harmless commensals of the skin and mucous membranes, coagulase-negative staphylococci, along with *S. aureus*, have emerged as the leading cause of bacteremia associated with nosocomial infections (Kloos & Bannerman, 1995). Thus, there is a pressing need to accurately and rapidly identify and speciate clinically important coagulase-negative staphylococci isolates. Unfortunately, the taxonomic classification and species identification of these micro-organisms remain problematic.

Conventional taxonomic and identification schemes

Abbreviation: HSP60, 60 kDa heat-shock protein.

The GenBank accession numbers for the sequences reported in this paper are AF033622, AF036322–AF036324, AF053568–AF053587, AF060184–AF060191 and U92809.

Table 1. List of strains and GenBank accession numbers of partial HSP60 and 16S rRNA gene sequences of staphylococci

Species	HSP60 DNA		16S rRNA			
	ATCC or isolate no.	GenBank no.	ATCC or isolate no.*	GenBank no.	Size of 16S rRNA gene	*E. coli* 16S rRNA nucleotide 33 homologue†
S. aureus	ATCC 25923‡	AF053568	–	–	–	–
S. aureus	ATCC 27217	AF060184	–	–	–	–
S. aureus	ATCC 25178	AF060185	–	–	–	–
S. aureus	ATCC 12598	AF060186	–	–	–	–
S. aureus	ATCC 10832	AF060187	–	–	–	–
S. aureus	ATCC 13565	AF060188	–	–	–	–
S. aureus	ATCC 14458	AF060189	–	–	–	–
S. aureus	ATCC 19095	AF060190	–	–	–	–
S. aureus	ATCC 27664	AF060191	–	–	–	–
S. aureus	Strain 912	D14711§	ATCC 29740	AF015929§	1442	33
S. aureus subsp. *anaerobius*	ATCC 35844^T‡	AF036323	–	–	–	–
S. aureus subsp. *aureus*	ATCC 12600^T‡	AF036324	–	–	–	–
S. arlettae	ATCC 43957‡	AF053580	DSM 20676	Z26888§	771	38
S. auricularis	–	–	ATCC 33753^T	D83358§	1475	34
S. capitis subsp. *capitis*	ATCC 27840‡	AF036322	Same	L37599§	1469	33
S. capitis subsp. *ureolyticus*	ATCC 49324‡	AF053587	–	–	–	–
S. caprae	ATCC 35538‡	AF053574	CCM3573^T	Z26890§	771	38
S. carnosus	–	–	DSM 20501	Z26891§	771	38
S. caseolyticus	ATCC 13548^T‡	AF053577	Same	D83359§	1478	35
S. chromogenes	–	–	ATCC 43764^T	D83360§	1475	35
S. cohnii subsp. *cohnii*	ATCC 29974^T‡	AF053582	Same	D83361§	1477	35
S. cohnii subsp. *ureolyticus*	ATCC 49330^T‡	AF053581	–	–	–	–
S. delphini	ATCC 49171‡	AF053571	–	–	–	–
S. delphini	Strain Heidy	AF019774§	–	–	–	–
S. epidermidis	9759‡	AF029245	ATCC 14990^T	D83363§	1475	35
S. epidermidis	9759	U13618§	–	–	–	–
S. equorum	–	–	DSM 20674	Z26895§	771	38
S. felis	–	–	ATCC 49168^T	D83364§	1475	35
S. gallinarum	ATCC 35539‡	AF053579	Same	D83366§	1477	35
S. haemolyticus	ATCC 29970^T‡	U92809	Same	D83367§	1475	35
S. hominis	ATCC 27844^T‡	AF053572	Same	L37061§	1468	33
S. hyicus	ATCC 11249^T‡	AF019778§	Same	D83368§	1476	35
S. intermedius	CFDD	AF019773§	ATCC 29663^T	D83369§	1476	35
S. kloosii	ATCC 43959‡	AF053575	DSM 20676^T	Z26898§	771	38
S. lentus	ATCC 29070^T‡	AF053586	Same	D83370§	1480	35
S. lugdunensis	ATCC 43809^T‡	AF053570	Same	Z26899§	771	38
S. pulvereri	–	–	ATCC 51698^T	AB009942§	1497	53
S. saccharolyticus	–	–	ATCC 14953^T	L37602§	1527	42
S. saprophyticus	ATCC 15305^T‡	AF053578	Same	D83371§	1477	35
S. schleiferi subsp. *schleiferi*	ATCC 43808^T‡	AF033622	CD22-1	D83372§	1476	33
S. schleiferi subsp. *coagulans*	ATCC 49545‡	AF053585	–	–	–	–
S. sciuri	ATCC 29060‡	AF053583	NCTC 12103	Z26901§	772	38
S. simulans	ATCC 27848^T‡	AF053584	Same	D83373§	1476	35
S. succinus	–	–	AMG-D1^T	AF004220§	1548	56

Table 1. (*cont.*)

Species	HSP60 DNA		16S rRNA			
	ATCC or isolate no.	GenBank no.	ATCC or isolate no.*	GenBank no.	Size of 16S rRNA gene	*E. coli* 16S rRNA nucleotide 33 homologue†
S. vitulus	ATCC 51145^T‡	AF053576	–	–	–	–
S. warneri	ATCC 27836^T‡	AF053569	Same	L37603§	1470	33
S. xylosus	ATCC 29971^T‡	AF053573	Same	D83374§	1477	35
B. subtilis	W168	M81132§	TB11	AF058766§	1512	53
E. coli	WK-12	AE000487§	Ocular isolate	AF076037§	1144	33

* These isolates were used for 16S rRNA sequence similarity analysis shown in Table 3.

† Position of nucleotide in 16S rRNA gene which aligned with nucleotide 33 in the 16S rRNA homologue of *E. coli*.

‡ Isolates used for HSP60 sequence similarity analysis shown in Table 2.

§ Previously reported GenBank sequence.

have relied heavily on phenotypic characteristics such as colonial morphology and biochemical profiles, including cell wall peptidoglycan (Schumacher-Perdreau *et al*., 1983) and teichoic acid (Endl *et al*., 1983) composition, cellular fatty acid analysis (Birnbaum *et al*., 1994; O'Donnell *et al*., 1985), acid fermentation products (Kloos & Bannerman, 1995), phage typing (Martin de Nicolas *et al*., 1995), serotyping (Pillet & Orta, 1981) and antibiotic susceptibility patterns (Kloos, 1997). These phenotypic characterizations have severe limitations, in part because of variable expression of certain traits, and ambiguity in the interpretation of their end point reactions (Birnbaum *et al*., 1991).

Among molecular taxonomic methods, DNA–DNA hybridization and 16S rRNA sequences have been the most proficient for phylogenetic analysis of the genus *Staphylococcus* (Kloos, 1997). DNA–DNA hybridization determines DNA relatedness by the relative binding of single-stranded DNA in reassociation reactions under stringent or non-stringent conditions, and is the current standard for defining the species or subspecies designation of staphylococci (Kloos & Bannerman, 1995; Wayne *et al*., 1987). Members of the same species generally demonstrate relative DNA binding values of 70% or greater under non-stringent conditions (50% or greater under stringent conditions), whereas organisms representing different species within the same genus have values less than 70% (Kloos, 1997). However, DNA–DNA hybridization is not suited for defining phylogenetic relationships of micro-organisms above the genus level because of the high degree of sequence dissimilarity (Stackebrandt *et al*., 1992). In contrast, 16S rRNA sequence analysis is well adapted for higher order classifications, but may not discriminate effectively among related members within a given genus or species because of the high degree of sequence similarity (Stackebrandt & Goebel, 1994).

Thus, both approaches have limitations for taxonomic classification and phylogenetic relationship of micro-organisms. Recently, the use of DNA sequences from genes encoding highly conserved proteins for performing phylogenetic and taxonomic analysis, such as the gene sequences of the 60 kDa heat-shock protein (HSP60) (Goh *et al*., 1996) or the B subunit DNA gyrase protein (Yamamoto & Harayama, 1995), has been evaluated. HSP60 genes, which encode highly conserved housekeeping proteins that assist in proper protein folding (also known as molecular chaperonins), are ubiquitous in both prokaryotes and eukaryotes. Viale *et al*. (1994) and Gupta (1995) previously observed that evolutionary trees drawn from the protein sequences of these molecules in eubacteria demonstrate remarkable similarity to those derived from 16S rRNA genes. Furthermore, the HSP60 homologues in mycobacteria (HSP65) were found to be useful for species identification and taxonomic classification of the genus *Mycobacterium* (Steingrube *et al*., 1995; Pai *et al*., 1997). We previously reported that PCR-amplified DNA probes prepared from partial HSP60 genes of *S. aureus*, *S. epidermidis*, *S. haemolyticus*, *S. saprophyticus*, *S. lugdunensis* and *S. schleiferi* were species-specific in dot-blot hybridization against a panel of 55 *Staphylococcus* species (Goh *et al*., 1996).

In the current study, we report the use of partial HSP60 gene sequences for the phylogenetic analysis of 36 validly described (Euzéby, 1997) staphylococcal strains representing 28 different staphylococcal species and subspecies. Results were compared to phylogenetic relationships based on DNA–DNA hybridization and published 16S rRNA gene sequences.

METHODS

Staphylococcal isolates. A collection of 36 staphylococcal strains, including nine *S. aureus* and 27 other validly

described *Staphylococcus* species or subspecies, were used for this study (Table 1). They were obtained either from the American Type Culture Collection (ATCC), from W. Kloos of North Carolina State University, Raleigh, NC, USA, or from our own collection obtained from the Clinical Microbiology Laboratory of the Vancouver Hospital and Health Sciences Centre, Vancouver, British Columbia, Canada. All isolates were grown in brain-heart infusion (BHI) broth, and subcultured overnight on BHI agar plates for examination of purity and colony characteristics. A single colony was picked for DNA extraction, PCR amplification with HSP60 degenerate primers, cloning and sequencing. Genomic DNA was extracted using the InstaGene matrix (Bio-Rad) according to manufacturer's instructions, with the exception that lysostaphin (Sigma) was added (3 U ml^{-1} final concentration) to facilitate cell lysis.

PCR amplification. The PCR mixture consisted of 15 μl (~ 50 ng) of InstaGene DNA extract, 50 mM KCl, 10 mM Tris (pH 8·3), 1·5 mM $MgCl_2$ (GIBCO), 200 μM of each dNTP (Pharmacia Biotech), 2·5 U *Taq* DNA polymerase (GIBCO), and 0·5 μM each of the HSP60 degenerate primers, designated H279 and H280, as previously described (Goh *et al.*, 1996). These primers, with the nucleotide sequence of 5′-GAATTCGAIIIIGCIGGIGA(TC)GGIACIACIAC-3′ and 5′-CGCGGGATCC(TC)(TG)I(TC)(TG)ITCICC-(AG)AAICCIGGIGC(TC)TT-3′, respectively, amplify an anticipated 600 bp HSP60 DNA. The H279 primer had an *Eco*RI restriction enzyme digest site while H280 had a *Bam*HI digest site (both underscored in the above sequences). The PCR thermal cycling conditions were 3 min at 95 °C for 1 cycle, followed by 30 cycles of 37 °C for 30 s, 72 °C for 1 min and 94 °C for 30 s. The last cycle was at 72 °C for 10 min. After PCR amplification, a sample of each reaction mixture was analysed by electrophoresis on a 2% TAE (Tris-acetate-EDTA) agarose gel at 100 V for 1 h, and DNA was visualized with ethidium bromide under UV light.

Purification of 600 bp HSP60 PCR products, cloning and transformation. The amplified 600 bp PCR products were purified using the QIAquick PCR purification kit (Qiagen). In instances where multiple bands were visualized on the gel, the 600 bp band was cut out and DNA was purified using the QIAquick gel extraction kit. Cloning was performed using the TA Cloning vector pCR2.1 (Invitrogen) as described in the manufacturer's protocol. During the early phase of this study, cloning of some PCR amplified DNA was performed using the plasmid vector pUC19. Kanamycin (50 μg ml^{-1}) was used to select for transformants. The presence of a correct insert was verified by PCR amplification using two primers (designated A1 and B2), which flank the multiple cloning site of the vector and amplify a 900 bp product containing the 600 bp HSP60 DNA. The sequences of these primers were: 5′-GCTTCCGGCTCGTATGTTGTGTG-3′ and 5′-AAAGGGGGATGTGCTGCAAGGCG-3′, respectively. The PCR thermal cycling conditions for screening were: 96 °C for 30 s for 1 cycle, followed by 25 cycles of 94 °C for 10 s, 55 °C for 30 s and 72 °C for 1 min, with the final cycle at 72 °C for 2 min.

DNA sequencing. Transformed *Escherichia coli*, containing plasmids with the correct PCR insert, was cultured overnight in BHI broth containing 100 μg ampicillin ml^{-1}. Plasmids from 3 ml overnight broth were purified using the Wizard Miniprep (Promega) according to the manufacturer's instructions. DNA sequencing was performed by the fluorescence-based dideoxy chain-termination method using the universal M13(−20) forward and M13 reverse sequencing primers in an automated DNA sequencer (Applied Biosystems model 373A). Emission data from the fluorescence-tagged reaction mixtures were analysed with the proprietary Macintosh-based software SeqEd (version 1.2.0).

Data analysis. Sequence analysis was performed with the entire 600 bp cloned fragment omitting the primer sequences used to amplify the HSP60 genes (sizes of different clones varied between 552 and 555 bp). Edited sequences were used for similarity searches of the NCBI BLAST databases (Altschul *et al.*, 1997). Multiple sequence alignments were performed using the CLUSTAL W program, version 1.7 (Thompson *et al.*, 1994). The published HSP60 gene sequences of the Gram-positive bacterium *Bacillus subtilis* (GenBank accession no. M81132), and the Gram-negative bacterium *E. coli* (AE000487) were included for comparison. Aligned DNA sequences were also translated into protein sequences using the Genetic Data Environment (GDE) program, version 2.3 (Smith, 1997). Phylogenetic analysis was performed using the PHYLIP program package, version 3.57 (Felsenstein, 1995). The unrooted phylogenetic tree was constructed by both the maximum-parsimony method (Fitch, 1971) and the neighbour-joining method using the Jukes–Cantor correction for multiple substitutions according to the one-parameter model (Jukes & Cantor, 1969). Bootstrapping was performed using 500 iterations. Phylogenetic trees were also generated based on the translated partial HSP60 protein sequences. For comparison, a similar phylogenetic tree was generated using the first 800 nucleotides from the 5′-end of previously published 16S rRNA gene sequences of 29 *Staphylococcus* species currently available in the GenBank databases (isolates, accession numbers and length of available gene sequences are shown in Table 1).

RESULTS

Cloned 600 bp partial HSP60 gene sequences of staphylococci

A library of 36 cloned 600 bp HSP60 PCR products representing 28 validly described *Staphylococcus* species and subspecies was generated. These PCR products were cloned and sequenced, and their GenBank accession numbers are listed in Table 1. Errors attributed to sequencing when compared to an internal standard of 650 bases were less than 0·2% and therefore negligible. To further evaluate the reproducibility of DNA extraction, PCR amplification, cloning and sequencing procedures, the partial HSP60 genes of *S. cohnii* subsp. *cohnii* and *S. cohnii* subsp. *ureolyticus* were each cloned and sequenced in two separate and independent experiments. There was 100% agreement between the two independent studies in the consensus sequences derived from both strands for both strains.

Similarity searches of staphylococcal HSP60 DNA sequences

Searches were performed in the NCBI BLAST databases for DNA sequence similarity with our 36 PCR-generated staphylococcal partial HSP60 gene sequences. The highest similarity scores (and lowest

Table 2. Sequence similarity (%) of partial HSP60 gene and protein sequences of different staphylococci based on CLUSTALW (1.74) alignment

1, *S. arlettae*; 2, *S. aureus*; 3, *S. aureus/anaerobius*; 4, *S. aureus/aureus*; 5, *S. capitis/capitis*; 6, *S. capitis/ureolyticus*; 7, *S. caprae*; 8, *S. caselyticus*; 9, *S. cohnii/cohnii*; 10, *S. cohnii/ureolyticus*; 11, *S. delphini*; 12, *S. epidermidis*; 13, *S. gallinarum*; 14, *S. haemolyticus*; 15, *S. hominis*; 16, *S. hyicus*; 17, *S. intermedius*; 18, *S.kloosii*; 19, *S. lentus*; 20, *S. lugdunensis*; 21, *S. saprophyticus*; 22, *S. schleiferi/coagulans*; 23, *S. schleiferi/schleiferi*; 24, *S. sciuri*; 25, *S, simulans*; 26, *S. vitulus*; 27, *S. warneri*; 28, *S. xylosus*; 29, *B. subtilis*; 30, *E. coli*. Percentage sequence similarity of HSP60 proteins and HSP60 genes are shown in the upper and lower triangles, respectively. Darkly shaded rectangles denote sequence similarity between pairs within the same species. Shaded enclosed rectangles denote sequence similarity between the most similar pairs of different species. Open rectangles denote sequence similarity of the least similar pairs of different species.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
1	–	92	93	93	92	92	92	85	93	94	91	91	95	92	92	92	88	96	85	94	95	93	92	85	88	85	95	96	70	52
2	85	–	99	99	94	95	95	83	91	92	90	93	91	94	94	89	88	92	83	94	92	90	90	83	89	83	96	91	69	54
3	85	98	–	100	95	95	95	83	91	92	91	94	91	94	95	89	88	93	84	94	93	91	90	83	90	83	96	92	70	54
4	85	98	98	–	95	95	95	83	91	92	91	94	91	94	95	89	88	93	83	94	93	91	90	83	90	83	96	92	70	54
5	83	86	86	86	–	99	99	84	89	90	90	95	92	95	96	88	88	91	82	96	91	89	89	82	91	82	95	91	69	53
6	83	86	86	86	91	–	100	85	90	91	91	96	93	96	96	88	88	92	83	97	91	90	89	83	91	83	96	91	69	53
7	83	85	85	86	86	86	–	85	90	91	91	96	93	96	96	88	88	92	83	97	91	90	89	83	91	83	96	91	69	53
8	77	77	77	77	76	77	75	–	84	85	85	83	87	83	84	85	79	85	87	84	86	85	84	89	79	90	85	85	73	53
9	87	84	84	84	81	80	84	74	–	97	88	90	94	89	89	88	85	92	84	91	97	89	88	84	84	84	93	96	70	53
10	88	85	85	85	81	81	83	75	93	–	89	91	95	90	90	89	86	94	85	92	98	90	89	85	85	85	94	97	71	54
11	82	80	81	81	79	81	80	77	78	80	–	90	90	89	90	95	90	91	84	90	89	95	95	84	88	84	91	89	71	52
12	80	83	83	83	84	83	83	75	81	80	78	–	92	93	94	88	86	91	82	94	91	89	89	81	88	81	95	90	70	52
13	86	82	82	83	83	82	82	76	84	84	81	80	–	92	92	89	88	93	84	94	96	90	89	84	87	84	94	95	71	53
14	85	86	87	87	85	85	84	77	84	86	81	82	85	–	99	86	89	91	81	94	90	88	88	83	90	83	95	91	69	54
15	84	84	85	85	85	85	86	76	84	85	81	81	85	89	–	87	89	91	82	95	91	89	88	83	90	83	95	91	69	55
16	81	80	81	81	79	78	79	74	81	81	82	79	82	80	81	–	88	91	85	90	90	98	97	86	86	85	90	90	71	53
17	82	81	81	82	80	81	82	77	80	81	93	79	80	82	82	84	–	87	85	88	86	90	89	80	86	80	90	87	67	51
18	85	83	83	83	81	82	81	77	83	83	80	81	82	84	86	80	81	–	85	92	94	91	91	85	86	85	94	94	70	52
19	86	79	79	79	77	78	77	77	77	78	78	75	79	80	78	77	77	79	–	83	85	85	85	95	78	95	85	85	68	53
20	83	86	85	85	86	86	83	75	82	83	79	82	84	85	86	81	80	82	80	–	92	91	91	83	90	83	95	92	70	53
21	82	85	85	85	83	83	83	76	87	87	80	83	85	83	84	80	81	85	77	84	–	90	90	85	86	85	95	97	71	53
22	82	82	82	82	81	82	82	78	83	83	83	80	82	83	81	85	84	81	79	82	82	–	99	85	88	85	91	90	71	52
23	83	82	82	82	81	82	82	77	83	83	84	80	82	83	81	86	84	82	79	81	82	98	–	86	88	85	92	91	72	53
24	82	79	79	79	79	79	80	79	80	81	79	76	80	81	82	80	80	81	88	81	80	80	80	–	79	97	84	85	71	53
25	83	81	82	81	81	82	81	75	79	81	80	81	80	84	81	78	81	80	76	83	81	81	81	78	–	79	90	86	68	53
26	79	78	78	78	76	77	77	77	78	79	77	74	80	79	79	76	78	78	89	79	77	79	79	90	77	–	84	85	72	53
27	85	87	87	87	86	84	84	75	86	86	80	83	84	86	87	81	81	86	78	87	86	82	82	80	82	78	–	94	71	54
28	85	82	82	82	81	83	83	75	86	86	79	79	84	84	85	81	81	83	78	83	87	83	82	80	79	78	83	–	70	53
29	65	66	65	65	65	67	66	70	66	65	65	66	69	64	66	66	66	66	66	65	66	67	68	66	65	67	64	66	–	53
30	55	59	59	59	59	58	59	56	55	58	59	55	57	55	56	56	58	56	55	55	56	57	57	55	56	53	58	58	62	–

Table 3. Sequence similarity (%) of 16S rRNA gene sequences of different *Staphylococci* based on CLUSTAL W (1.74) alignment

Species: 1, *S. arlettae*; 2, *S. aureus*; 3, *S. auricularis*; 4, *S. capitis*; 5, *S. caprae*; 6, *S. carnosus*; 7, *S. caseolyticus*; 8, *S. chromogenes*; 9, *S. cohnii*; 10, *S. epidermidis*; 11, *S. equorum*; 12, *S. felis*; 13, *S. gallinarum*; 14, *S. haemolyticus*; 15, *S. hominis*; 16, *S. hyicus*; 17, *S. intermedius*; 18, *S. kloosii*; 19, *S. lentus*; 20, *S. lugdunensis*; 21, *S. pulvereri*; 22, *S. saccharolyticus*; 23, *S. saprophyticus*; 24, *S. schleiferi*; 25, *S. sciuri*; 26, *S. simulans*; 27, *S. succinus*; 28, *S. warneri*; 29, *S. xylosus*; 30, *B. subtilis*; 31, *E. coli*.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	–																													
2	93	–																												
3	96	96	–																											
4	99	97	97	–																										
5	96	90	96	95	–																									
6	96	90	96	95	96	–																								
7	92	94	93	94	91	92	–																							
8	96	96	97	97	96	96	94	–																						
9	97	97	97	97	98	97	94	97	–																					
10	99	98	97	99	95	96	94	97	97	–																				
11	96	91	96	95	98	96	92	96	98	96	–																			
12	96	96	97	97	95	96	94	98	96	97	95	–																		
13	95	97	97	97	98	96	94	97	98	97	97	97	–																	
14	97	98	97	98	97	96	94	97	98	98	96	97	98	–																
15	97	97	97	98	96	95	94	97	98	98	96	97	97	99																
16	97	96	97	97	96	96	94	98	97	97	96	98	97	97	97	–														
17	96	96	97	97	96	96	94	98	97	97	96	98	97	97	97	99	–													
18	96	90	96	95	97	97	92	96	97	95	97	95	97	96	96	96	96	–												
19	95	96	95	96	94	94	95	96	96	97	94	96	96	96	96	96	96	94	–											
20	97	91	97	96	97	96	92	96	97	97	97	96	96	98	97	96	96	97	95	–										
21	95	96	95	96	95	94	94	96	96	96	94	96	96	96	96	96	96	94	99	95	–									
22	99	98	96	99	95	96	93	97	97	99	95	96	96	97	97	97	96	95	96	97	95	–								
23	95	97	96	97	97	94	93	96	98	97	97	96	98	98	97	96	96	96	96	96	94	94	–							
24	96	96	97	97	96	96	94	98	97	97	96	98	97	97	97	98	99	95	96	96	96	97	96	–						
25	96	90	94	95	95	95	93	95	95	95	95	95	95	95	95	96	95	94	98	96	98	95	95	95	–					
26	96	96	96	97	95	96	94	96	97	97	95	96	96	97	97	96	96	95	96	95	95	96	96	96	94					
27	96	97	97	97	98	96	94	97	98	97	98	96	99	98	97	97	97	98	96	97	96	95	98	97	95	96	–			
28	97	97	97	98	95	95	94	97	98	98	96	97	98	98	98	97	97	96	96	97	96	97	97	97	95	96	98	–		
29	96	97	96	97	98	95	94	97	98	97	98	96	99	98	98	97	97	97	96	97	96	96	99	97	95	96	99	98	–	
30	90	91	91	91	90	90	91	91	91	91	90	91	90	91	91	91	91	90	91	90	91	89	91	91	91	91	91	90	91	–
31	74	73	76	76	73	73	76	76	76	76	73	76	76	76	76	76	76	74	76	74	76	76	76	76	74	76	76	77	76	77

probability of a random match) were obtained with the sequences from *S. aureus*, *S. epidermidis*, *S. intermedius*, *S. hyicus* and *S. delphini*, each matching with the HSP60 DNA sequences of the corresponding *Staphylococcus* species currently available in the databases (Table 1).

Sequence similarity in partial HSP60 genes and proteins among different *Staphylococcus* species and subspecies

Alignments of HSP60 DNA and protein sequences from nine different *S. aureus* isolates and 28 different *Staphylococcus* species or subspecies were obtained using the CLUSTAL W program. DNA sequence similarity among the nine different *S. aureus* isolates ranged from 97 to 100% (mean similarity 98% among 36 pairwise comparisons among these nine isolates). Protein sequence similarity among these nine isolates ranged from 98 to 100% (mean similarity 99%) (data not shown).

Sequence similarity in partial HSP60 DNA and proteins among 28 different *Staphylococcus* species or subspecies as well *E. coli* and *B. subtilis* is shown in Table 2. At the subspecies level, the HSP60 DNA sequence similarity among paired members of the same species ranged from 91 to 98%, with the most similar pairs between *S. aureus* subsp. *aureus* and *S. aureus* subsp. *anaerobius* and between *S. schleiferi* subsp. *schleiferi* and *S. schleiferi* subsp. *coagulans* (98% sequence similarity each), and the least similar pair between *S. capitis* subsp. *capitis* and *S. capitis* subsp. *ureolyticus* (91%). At the interspecies level, HSP60 DNA sequence similarity among the entire set of 23 different *Staphylococcus* species ranged from 74 to 93% (mean 82%) among 253 pairwise comparisons, with the most similar pair between *S. intermedius* and *S. delphini* (93%), and the least similar pairs between *S. caseolyticus* and *S. hyicus*, *S. caseolyticus* and *S. cohnii* subsp. *cohnii*, and *S. epidermidis* and *S. vitulus* (74% for each pair, respectively). Among the different *Staphylococcus* species studied, *S. caseolyticus* was the least similar in HSP60 DNA sequences, with similarity values ranging from 74 to 79% (mean 76%) when aligned with other members of the genus *Staphylococcus*. By comparison, the maximum similarity in the HSP60 DNA sequences between the Gram-positive *B. subtilis* and different members within the genus

Table 4. CLUSTAL W alignment of partial HSP60 DNA (upper) and translated protein (lower) sequences showing hot spot region

Symbols: '*', identical or conserved residues in all sequences in the alignment; ':', conserved substitutions; '.', semiconserved substitutions. Fonts in upper section: regular, C; bold, A; italic regular, T. Fonts in lower section: regular, small hydrophobic residues (AVFPMILW); bold, acidic residues (DE); italic regular, basic residues (RHK); italic bold, hydroxyl, amine or basic residues (STYCNGQ).

```
                    91        100       110       120       130       140       153
S. arlettae         CAAAGCTGT---AGATGTTGCGATTACAGCA-TTACA-TGATATTTCACAAAAAGTT--GAAAATAAAAA
S. aureus           CAAAGCAGT---TAAAGTTGCTGTTGAAGCG-TTACA-TGAAAATTCTCAAAAAGTT--GAAAATAAAAA
S. aureus subsp. anaerobius
                    CAAAGCAGT---TAAAGTTGCTGTTGAAGCG-TTACA-TGAAAATTCTCAAAAAGTT--GAAAATAAAAA
S. aureus subsp. aureus
                    CAAAGCAGT---TAAAGTTGCTGTTGAAGCG-TTACA-TGAAAATTCTCAAAAAGTT--GAAAATAAAAA
S. capitis subsp. capitis
                    TAAAGCCGT---TAAAATAGCTGTAGAAGCT-TTACA-TGATATCTCTCAAAAAGTT--GAAAATAAAAA
S. capitis subsp. ureolyticus
                    TAAAGCCGT---CAAAGTAGCTGTAGAAGCA-TTACA-TGACATTTCTCAAAAAGTT--GAAAACAAAAA
S. caprae           CAAAGCAGT---AAAAGTTGCTGTTGAAGCA-CTTCA-TGATATTTCTCAAAAAGTT--GAAAATAAAAA
S. caseolyticus     CAAAGCAGT---AGCTGTAGCACTAGAAGAACTTGCA--GCAATCTCTAAAACAGTA--TCAAGCAAAGA
S. cohnii subsp. cohnii
                    CAAAGCTAT---AGAAGTAGCGATTGAAGCG-TTACA-TGAAATTTCACAAAATGTA--GATAATAAAAA
S. cohnii subsp. ureolyticus
                    CAAAGCTGT---AGAAGTAGCGATTGAAGCT-TTACA-TGAAATTTCACAAAATGTA--GATAATAAAAA
S. delphini         TAAAGCAGT---GGCGGTCGCAATTGAATCA-TTACA-CGACATTTCGCAAAAAGTT--GAAAACCAAGA
S. epidermidis      CAAAGCAGT--GC-AAGTGGCTATAGAAGCG-CTCCA-TGAGATTTCTCAAAAGGTT--GAAAATAAGAA
S. gallinarum       TAAAGCTGT---TGAAGTTGCAATAACTGCT-TTACA-TGATATTTCTCAAAAAGTA--GAAAATAAAAA
S. haemolyticus     TAAAGCAGT---AAGAGTAGCTGTACAAGCT-CTACA-CGATATTTCTCAAAAAGTT--GAAAATAAAAA
S. hominis          TAAAGCAGT---CAGAGTAGCAGTTGAAGCA-TTACA-TGACATTTCTCAAAAAGTT--GAAAATAAAAA
S. hyicus           TAAAGCGGT---AGCTGTTGCAATAGAATCG-TTACA-CAATATTTCTCAAAAAGTA--GAAAATAAAGA
S. intermedius      TAAAGCTGT---GGCAGTCGCAATTGAATCA-TTACA-CAATATTTCTCAAAAAGTT--GAAAACAAAGA
S. kloosii          TAAAGCAGT---TGAAGTAGCAATTGAAGCT-TTACA-CGGTATCTCTCAAAAAGTT--GAAAATAAAAA
S. lentus           TAAGGCAGT---TAAAGTTGCTTTAGAAGAA-TTACA-CGAGATTTCACAACCTGTA--GAGAAGAAAGA
S. lugdunensis      TAAAGCAGT---TAAAGTAGCTATCGAAGCA-TTACA-TGATATATCACAAAAAGTT--GAAAATAAAAA
S. saprophyticus    TAAAGCTGT---AGAAGTAGCAATTGAGGCA-TTACA-TGAAATTTCACAAAATGTA--GATAATAAAAA
S. schleiferi subsp. coagulans
                    TAAAGCGGT---AGCTGTAGCGATTGAAGCA-CTTCA-TAATATTTCTCAAAAAGTA--GAAAATAAAGA
S. schleiferi subsp. schleiferi
                    TAAAGCGGT---AGCTGTAGCGATTGAAGCA-CTTCA-TAATATTTCTCAAAAAGTA--GAAAATAAAGA
S. sciuri           TAAAGCAGT---TAAAGTAGCTTTAGAAGAA-TTACA-CAATATCTCTCAACCAGTT--GAGAAAAAAGA
S. simulans         TTTAGCAGTAGGCGAAGCGG---TACAAGCT-TTGCA-TGATCAATCTCAAAAAGTT--GAAAATAAAAA
S. vitulus          TAAAGCGGT---TAAAGTAGCTTTAGAAGAA-TTACAACGA-ATTTCACAACCTGTC--GAGAAAAAAGA
S. warneri          TAAAGCAGT---TCAAGTAGCAGTTGAAGCT-TTACA-TGAAATTTCTCAAAAAGTA--GAAAATAAAAA
S. xylosus          TAAAGCTGT---AGAAGTAGCAATTAATGCA-TTACA-TGATATTTCTCAAAATGTA--GACAATAAAAA
                       **  *          *  *          * **          **  **  **          *   *  *
B. subtilis         ACAAGCTGT---AGCGGTTGCGATCGAA-AACTTAAA-AGAAATTTCTAAGCCAATC--GAAGGCAAAGA
E. coli             CAAAGCGGT---TACCGCTGCAGTTGAAGAACTGAAAGCGCTGTC-CGTACCATGCTCTGACTCTAAA--
                       **  *          *  *          *          *  *          *

                    31                  40                                  51
S. arlettae         --K--A--V--D--V--A--I--T--A--L--H--D--I--S--Q--K--V--E--N--K--N
S. aureus           --K--A--V--K--V--A--V--E--A--L--H--E--N--S--Q--K--V--E--N--K--N
S. aureus subsp. anaerobius
                    --K--A--V--K--V--A--V--E--A--L--H--E--N--S--Q--K--V--E--N--K--N
S. aureus subsp. aureus
                    --K--A--V--K--V--A--V--E--A--L--H--E--N--S--Q--K--V--E--N--K--N
S. capitis subsp. capitis
                    --K--A--V--K--I--A--V--E--A--L--H--D--I--S--Q--K--V--E--N--K--N
S. capitis subsp. ureolyticus
                    --K--A--V--K--V--A--V--E--A--L--H--D--I--S--Q--K--V--E--N--K--N
S. caprae           --K--A--V--K--V--A--V--E--A--L--H--D--I--S--Q--K--V--E--N--K--N
S. caseolyticus     --K--A--V--A--V--A--L--E--E--L--A--A--I--S--K--T--V--S--S--K--E
S. cohnii subsp. cohnii
                    --K--A--I--E--V--A--I--E--A--L--H--E--I--S--Q--N--V--D--N--K--N
S. cohnii subsp. ureolyticus
                    --K--A--V--E--V--A--I--E--A--L--H--E--I--S--Q--N--V--D--N--K--N
S. delphini         --K--A--V--A--V--A--I--E--S--L--H--D--I--S--Q--K--V--E--N--Q--E
S. epidermidis      --K--A--V--Q--V--A--I--E--A--L--H--E--I--S--Q--K--V--E--N--K--N
S. gallinarum       --K--A--V--E--V--A--I--T--A--L--H--D--I--S--Q--K--V--E--N--K--N
S. haemolyticus     --K--A--V--R--V--A--V--Q--A--L--H--D--I--S--Q--K--V--E--N--K--N
S. hominis          --K--A--V--R--V--A--V--E--A--L--H--D--I--S--Q--K--V--E--N--K--N
S. hyicus           --K--A--V--A--V--A--I--E--S--L--H--N--I--S--Q--K--V--E--N--K--E
S. intermedius      --K--A--V--A--V--A--I--E--S--L--H--N--I--S--Q--K--V--E--N--K--E
S. kloosii          --K--A--V--E--V--A--I--E--A--L--H--G--I--S--Q--K--V--E--N--K--N
S. lentus           --K--A--V--K--V--A--L--E--E--L--H--E--I--S--Q--P--V--E--K--K--E
S. lugdunensis      --K--A--V--K--V--A--I--E--A--L--H--D--I--S--Q--K--V--E--N--K--N
S. saprophyticus    --K--A--V--E--V--A--I--E--A--L--H--E--I--S--Q--N--V--D--N--K--N
S. schleiferi subsp. coagulans
                    --K--A--V--A--V--A--I--E--A--L--H--N--I--S--Q--K--V--E--N--K--E
S. schleiferi subsp. schleiferi
                    --K--A--V--A--V--A--I--E--A--L--H--N--I--S--Q--K--V--E--N--K--E
S. sciuri           --K--A--V--K--V--A--L--E--E--L--H--N--I--S--Q--P--V--E--K--K--E
S. simulans         --L--A--V--G--E--A--V--Q--A--L--H--D--Q--S--Q--K--V--E--N--K--N
S. vitulus          --K--A--V--K--V--A--L--E--E--L--Q--R--I--S--Q--P--V--E--K--K--E
S. warneri          --K--A--V--Q--V--A--V--E--A--L--H--E--I--S--Q--K--V--E--N--K--N
S. xylosus          --K--A--V--E--V--A--I--N--A--L--H--D--I--S--Q--N--V--D--N--K--N
                       *  :        *  :        *          *  :     .  .  :  :
B. subtilis         --Q--A--V--A--V--A--I--E--N--L--K--E--I--S--K--P--I--E--G--K--E
E. coli             --K--A--V--T--A--A--V--E--E--L--K--A--L--S--V--P--C--S--D--S--K
                       *  :        *  :        *          *             .     .  :
```

Staphylococcus was 70% (mean 66%), and that between the Gram-negative *E. coli* and different members within the genus *Staphylococcus* was 59% (mean 57%). This degree of HSP60 DNA relatedness is comparable to the HSP60 DNA sequence similarity of 62% between *B. subtilis* and *E. coli*.

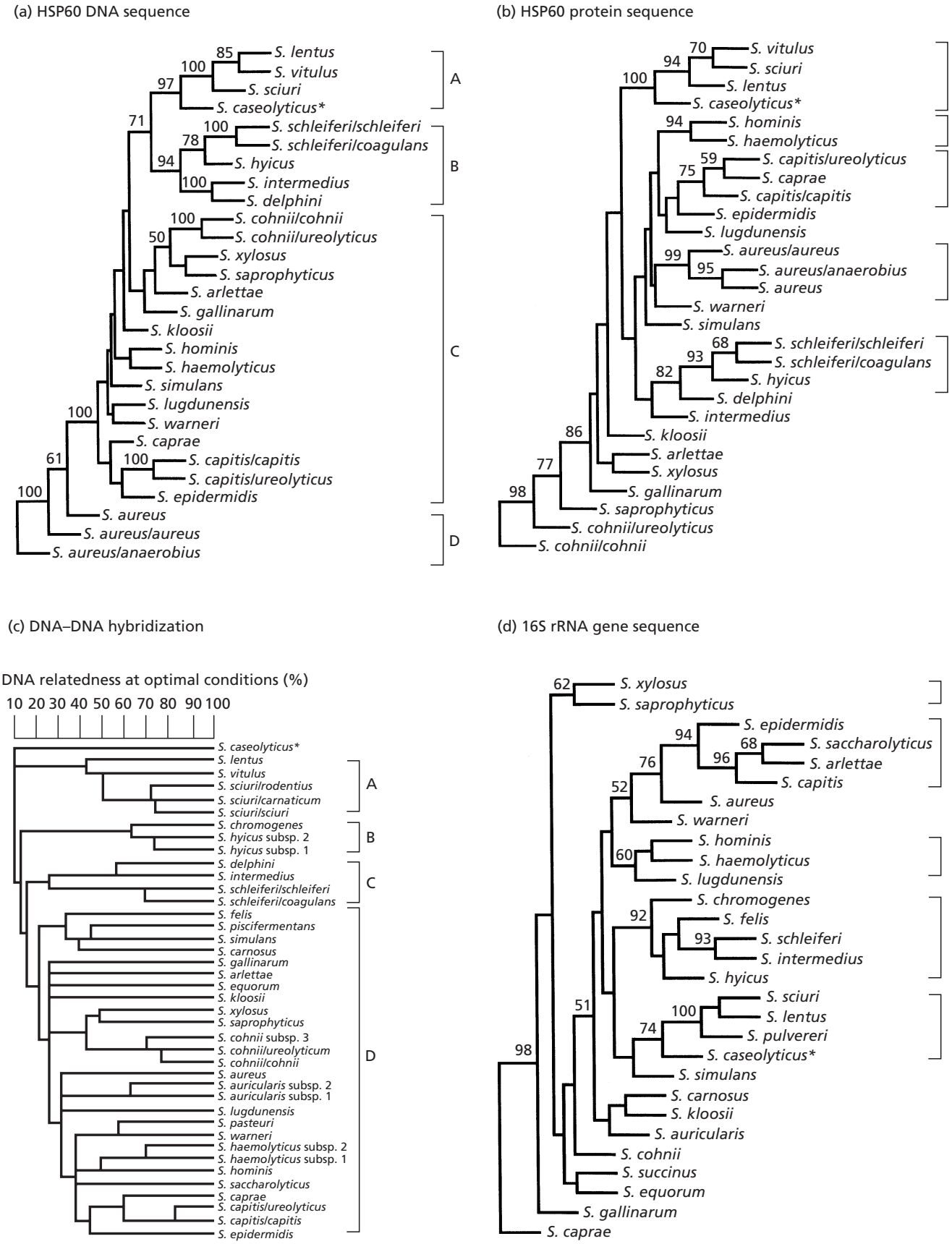

Fig. 1. For legend see facing page.

HSP60 protein sequence similarity among these 28 different *Staphylococcus* species and subspecies was also examined (Table 2). As expected from the degeneracy of the genetic code, the amino acid sequence similarity between any given pair was always higher than that of the corresponding DNA sequences between the same pair. The amino acid sequence similarity values ranged from 78% (between *S. lentus* and *S. intermedius*) to 100% (between *S. aureus* subsp. *aureus* and *S. aureus* subsp. *anaerobius*, and between *S. caprae* and *S. capitis* subsp. *ureolyticus*, respectively) among 378 pairwise comparisons. The HSP60 proteins of *S. vitulus*, *S. lentus* and *S. sciuri* appear to be highly related to each other as demonstrated by their high similarity scores (97% between *S. vitulus* and *S. sciuri*, 95% between *S. vitulus* and *S. lentus*, and 95% between *S. lentus* and *S. sciuri*). This contrasts with the mean of 90% similarity for the entire group of 28 *Staphylococcus* species and subspecies. Conversely, these three species along with *S. caseolyticus* appeared to share the least similarity in HSP60 protein sequences with other members of the genus *Staphylococcus* since their mean similarity scores when aligned with other members of *Staphylococcus* species (84·6, 84·7, 84·8 and 84·8% for *S. lentus*, *S. caseolyticus*, *S. vitulus* and *S. sciuri*, respectively) were lower than the mean similarity score for the entire group (90%). The percentage similarity in HSP60 protein sequences of *B. subtilis* and *E. coli* with different members within the genus *Staphylococcus* were comparable to that of their corresponding HSP60 DNA sequences (mean 70 and 53%, respectively).

Comparison of sequence similarity in 16S rRNA genes with HSP60 DNA among different *Staphylococcus* species

Similarity values for 16S rRNA gene sequences among 29 different *Staphylococcus* species ranged from 92 to 99% (mean 96%) (Table 3). These values were consistently higher than those for HSP60 gene sequences (range 74–93%, mean 82%) (Table 2). This indicates that 16S rRNA sequences are less discriminatory among the different *Staphylococcus* species as compared to HSP60 DNA sequences. The data in Table 3 also reveal that 16S rRNA gene sequences of these *Staphylococcus* species were more related to *B. subtilis* (mean similarity 91%; range 90–91%) than to *E. coli* (mean similarity 75%; range 73–77%).

Analysis of HSP60 DNA and protein sequences by multiple sequence alignment

The DNA multiple sequence alignments of 28 *Staphylococcus* species and subspecies demonstrated the presence of both conserved and variable regions distributed throughout the partial HSP60 genes (data not shown). However, a region spanning from nucleotide positions 90–153 was identified which appeared more variable than other regions in the DNA sequences of these genes (Table 4). This appeared to be a 'hot spot' region encoding highly variable amino acid residues corresponding to amino acid residues 30–51. In addition, the partial HSP60 proteins of four *Staphylococcus* species (*S. caseolyticus*, *S. sciuri*. *S. lentus* and *S. vitulus*) shared a similar pattern of amino acid substitutions in five positions [residues 39 (A/S → E), 69 (Y → F), 89 (N/S → K), 151 (A/S → T) and 175 (M → L)], which differed from all the other staphylococci.

Phylogenetic analysis

Phylogenetic analysis of the partial HSP60 gene sequences was performed using both the neighbour-joining and maximum-parsimony methods in the PHYLIP software package, version 3.57c. After 500 bootstrapping replications, the consensus trees derived from the two methods were virtually identical and only the consensus tree by the neighbour-joining method is shown (Fig. 1a). This is contrasted to the consensus tree derived by the neighbour-joining method from HSP60 protein sequences (Fig. 1b). Based on HSP60 DNA sequences, the 28 staphylococcal species and subspecies can be grouped into four major divisions, A, B, C and D (Fig. 1a). *S. lentus*, *S. vitulus* and *S. sciuri* are highly related to each other within Division A, and together with *S. caseolyticus*, they are readily discriminated from the other staphylococci species. Similarly, *S. intermedius* and *S. delphini* are highly related to each other, and with *S. schleiferi* and *S. hyicus* within Division B. *S. aureus* appears to stand out from all other staphylococci in Division D. The remaining staphylococcal species are loosely grouped into Division C.

The consensus tree based on HSP60 protein sequences yielded essentially similar results, although the divisions were not as clear-cut (Fig. 1b). Thus, *S. vitulus*, *S. sciuri*, *S. lentus* and *S. caseolyticus* were

Fig. 1. Phylogenetic relationships among different *Staphylococcus* species revealed by (a) consensus tree (unrooted) derived by the neighbour-joining method based on partial HSP60 gene sequences after 500 bootstrapping iterations (bootstrap values above 50% are shown at branch points); (b) consensus tree (unrooted) derived by the neighbour-joining method based on partial HSP60 protein sequences after 500 bootstrapping iterations; (c) genomic DNA relationships of *Staphylococcus* species and subspecies based on DNA–DNA hybridization [adapted with permission from Kloos *et al.* (1997)]; and (d) consensus tree (unrooted) derived by the neighbour-joining method based on 16S rRNA gene sequences from GenBank databases after 500 bootstrapping iterations. **S. caseolyticus* has recently been reclassified as *Macrococcus caseolyticus* (Kloos *et al.*, 1998).

again in a unique cluster similar to Division A in Fig. 1(a). *S. hominis* and *S. haemolyticus* were highly related. *S. aureus* and the two subspecies were again in a single cluster with high bootstrap values (99%), while *S. schleiferi*, *S. hyicus* and *S. delphini* were again grouped together, although the bootstrap values in the protein consensus tree (82%) were not as high that in the DNA consensus tree (94%).

The consensus tree based on HSP60 partial gene sequences was remarkably similar to that based on DNA–DNA hybridization reported by Kloos (1997) (Fig. 1c). Thus, *S. lentus*, *S. vitulus* and *S. sciuri* again formed a unique cluster (similar to Division A in the HSP60 DNA tree). *S. delphini*, *S. intermedius* and *S. schleiferi* were again grouped together (similar to Division B in the HSP60 DNA tree), while *S. hyicus* was in a separate cluster with *S. chromogenes* (the latter was not examined by HSP60 DNA sequencing). The remaining staphylococcal species again formed a large and loosely related grouping (similar to Division C in the HSP60 DNA tree). Within this large grouping, similar clusters were seen as in the consensus HSP60 DNA tree (e.g. *S. xylosus*, *S. saprophyticus* and *S. cohnii*; *S. haemolyticus* and *S. hominis*; *S. caprae*, *S. capitis* and *S. epidermidis*). The highly related species of *S. felis*, *S. piscifermentans* and *S. carnosus* were unfortunately not studied by HSP60 DNA sequencing.

In contrast to HSP60 gene sequences, the consensus tree based on published 16S rRNA gene sequences of 29 *Staphylococcus* species correlated less well with the DNA–DNA hybridization tree (Fig. 1d). Although similarities in clustering patterns were seen, the bootstrap values were generally not high. Similarly to DNA–DNA hybridization and to the consensus HSP60 DNA tree, *S. epidermidis* and *S. capitis* were grouped together, as were *S. xylosus* and *S. saprophyticus*; *S. hominis* and *S. haemolyticus*; *S. schleiferi* and *S. intermedius*; and *S. sciuri* and *S. lentus*. However, a number of discordances were observed. These include the phylogenetic position of *S. aureus*, *S. felis*, *S. carnosus* and *S. simulans*, among others.

DISCUSSION

The accuracy of conventional methods for species identification and taxonomic classification of staphylococci based on phenotypic characteristics is limited (reported to range from 50 to ~ 70%) (Birnbaum *et al.*, 1991; Ieven *et al.*, 1995; Kloos & Bannerman, 1995). Rosypal *et al.* (1966) were among the first to use G+C ratios and DNA base composition as a molecular tool for the classification of staphylococci. Since then, DNA–DNA reassociation and 16S rRNA sequence analysis have emerged as two powerful and more accurate molecular tools for the taxonomic classification and phylogenetic analysis of micro-organisms (Stackebrandt & Goebel, 1994; Olsen *et al.*, 1994), including staphylococci (Kloos 1997; Gribaldo *et al.*, 1997; Zakrzewska-Czerwinska *et al.*, 1995). However, although useful for higher hierarchical classifications above the genus level, sequence analysis of 16S rRNA may not provide sufficient resolving power as DNA–DNA reassociation in discriminating between highly related micro-organisms of different species within the same genus (Fox *et al.*, 1992). For example, designation of a species within a given genus would generally include strains with approximately 70% or greater DNA–DNA relatedness and 5 °C or less ΔT_m (Wayne *et al.*, 1987). This degree of DNA–DNA relatedness generally corresponds to 97·5% or greater 16S rRNA sequence similarity (Stackebrandt & Goebel, 1994). Thus, it is clear that the resolving power of DNA–DNA hybridization is substantially higher than that of 16S rRNA sequence analysis for highly related organisms. For this reason, DNA–DNA hybridization has remained the reference standard for species designation within the genus *Staphylococcus* (Kloos, 1997). The value of 16S rRNA sequence analysis lies in the finding that two organisms having less than 97·5% 16S rRNA sequence similarity would not be expected to have more than 60–70% DNA similarity, and are therefore unlikely to be related at the genus level. This obviates the need for more laborious DNA reassociation studies if their 16S rRNA sequence similarity is low. Indeed, it would be valuable to identify an alternative target to 16S rRNA genes that offers similar discriminatory power at the species level that DNA–DNA hybridization provides currently.

Our results of the phylogenetic analysis of staphylococci indicate that the highly conserved HSP60 gene may be such an alternative. Firstly, HSP60 genes are ubiquitous in both prokaryotes and eukaryotes, and encode highly conserved housekeeping proteins that are essential for the survival of these cells. With rare exceptions [*Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Streptomyces albus*, *Sinorhizobium* (*Rhizobium*) *meliloti*] (Gupta, 1995), only single copies of the gene are present in each genome. These characteristics may render them less subject to random mutations or intraspecies variation. Viale *et al.* (1994) and Gupta (1995) independently observed that evolutionary trees drawn from the protein sequences of these molecules in eubacteria demonstrate remarkable similarity to those derived from 16S rRNA genes, and in some instances, provide additional insights not available from 16S rRNA sequence analysis. We previously reported that PCR-amplified DNA probes prepared from partial HSP60 genes of *S. aureus*, *S. epidermidis*, *S. haemolyticus*, *S. saprophyticus*, *S. lugdunensis* and *S. schleiferi* were species-specific in dot-blot hybridization against a panel of 55 *Staphylococcus* species (Goh *et al.*, 1996). A more recent study using reverse checkerboard hybridization has extended these observations to the identification of a further 25 *Staphylococcus* species (Goh *et al.*, 1997). No false-positive result was observed with 24 negative-control isolates of Gram-positive and Gram-negative bacteria. A false-negative result was observed only with *S. hyicus*. In this instance, the PCR-amplified HSP60

DNA from four bovine *S. hyicus* isolates failed to hybridize with the *S. hyicus* type strain, a porcine isolate. HSP60 DNA sequence data revealed that the bovine *S. hyicus* isolates differed from the type strain at 14% of 552 bases (i.e. only 86% similarity), suggesting that these bovine isolates could be a subspecies of *S. hyicus*. Of interest, species-specific allelic variation in the GroEL homologue of *Mycobacterium* species (HSP65) has also been utilized for rapid identification and taxonomic classification of mycobacteria in the clinical setting (Pai *et al.*, 1997; Steingrube *et al.*, 1995).

In the current study, we have cloned and sequenced the partial HSP60 genes from 36 staphylococci, representing 28 validly described *Staphylococcus* species or subspecies. Our results provide direct evidence for the presence of species-specific DNA sequences within the HSP60 genes of staphylococci, and further support the utility of the HSP60 gene as a universal target for the species identification of staphylococci. Furthermore, we demonstrated remarkable agreement in the consensus trees derived from partial HSP60 gene sequences and DNA–DNA reassociation for staphylococci, and that the phylogenetic tree derived from HSP60 DNA sequences correlated better than that from 16S rRNA sequences. HSP60 DNA sequences were also more discriminatory than 16S rRNA sequences, since sequence similarity values for 16S rRNA genes among different *Staphylococcus* species (Table 3) were consistently higher (range 92–99%; mean 96%) compared to partial HSP60 gene sequences (range 74–93%; mean 82%) (Table 2). The higher sequence similarity values for 16S rRNA genes compared to HSP60 genes may be due to the likelihood that genes encoding RNA are more evolutionarily ancient and conserved compared to genes encoding proteins. Our findings confirm the earlier observation that 16S rRNA sequence analysis lacks resolving power in discriminating between related species within the same genus (Fox *et al.*, 1992). Similar to genomic DNA–DNA reassociation results, the interspecies HSP60 DNA sequence similarity among the staphylococcal isolates varied from 74 to 93%, while subspecies sequence similarity ranged from 91 to 98%. By comparison, the highest sequence similarity of *B. subtilis* and *E. coli* with members within the genus *Staphylococcus* was only 70 and 59%, respectively (mean 66 and 57%). This suggests that HSP60 DNA sequence analysis could be an alternative to DNA–DNA hybridization or 16S rRNA sequence analysis for species identification and taxonomic classification within the genus *Staphylococcus*. Additionally, we identified a 'hot spot' corresponding to nucleotide positions 90–153 in the HSP60 gene sequences of staphylococci, a region which may be critical for the species-specific nature of these DNA sequences among staphylococci. Further research will be required to determine whether HSP60 gene sequences can be similarly utilized for taxonomic classification and phylogenetic analysis of other bacteria besides staphylococci, and whether this 'hot spot' region in the HSP60 gene is also present among bacteria other than staphylococci.

It should not be surprising that HSP60 DNA sequences were more discriminatory than protein sequences in our phylogenetic studies in light of the degeneracy of the genetic code, and the triplet nucleotide coding of specific amino acid residues. Thus, in the pairwise alignments, sequence similarities based on HSP60 proteins were consistently higher than those based on HSP60 DNA for all of the *Staphylococcus* species studied (Table 2). Analysis of the translated partial HSP60 protein sequences confirmed that all the proteins within the genus *Staphylococcus* as well as *B. subtilis* and *E. coli* are highly conserved. Another interesting observation is that the translated partial HSP60 proteins of four *Staphylococcus* species, *S. caseolyticus*, *S. sciuri*, *S. lentus* and *S. vitulus*, were highly related, and all share a characteristic pattern of amino acid substitutions at positions 39, 69, 89, 151 and 175 that were different from all the other *Staphylococcus* species (data not shown). It is also known that these four *Staphylococcus* species are uniquely different from all other *Staphylococcus* species in that cytochrome *c* is present in their electron transport system but not in other *Staphylococcus* species (Kloos & Bannerman, 1995). It is unknown whether the common amino acid substitutions observed in the HSP60 proteins of these four species are in any way related to their unique cytochrome oxidase activity.

Finally, it is of some interest to note that based on HSP60 DNA sequences, *S. caseolyticus* was the least similar to the other *Staphylococcus* species studied, thus lending additional support to the recent decision to remove this organism taxonomically from the genus *Staphylococcus*, and to describe it as *Macrococcus caseolyticus* gen. nov., comb. nov. (Kloos *et al.*, 1998).

In summary, our results indicate that DNA sequence analysis of the highly conserved and ubiquitous HSP60 genes may offer advantages over both DNA–DNA hybridization and 16S rRNA sequencing in defining the taxonomy and phylogenetic relationships within the genus *Staphylococcus*. In addition, our sequence data provide direct evidence for the presence of species-specific HSP60 gene sequences, which can be employed as a genotypic method for the species, identification of staphylococci, in particular coagulase-negative staphylococci.

ACKNOWLEDGEMENTS

We are grateful to Dr W. Kloos of North Carolina State University, Raleigh, NC, USA, for providing some of the strains studied.

REFERENCES

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Birnbaum, D., Kelly, M. & Chow, A. W. (1991). Epidemiologic typing systems for coagulase-negative staphylococci. *Infect Control Hosp Epidemiol* **12**, 319–326.

Birnbaum, D., Herwaldt, L., Low, D. E., Noble, M., Pfaller, M., Sherertz, R. & Chow, A. W. (1994). Efficacy of microbial identification system for epidemiologic typing of coagulase-negative staphylococci. *J Clin Microbiol* **32**, 2113–2119.

Endl, J., Seidl, H. P., Fiedler, F. & Schleifer, K. H. (1983). Chemical composition and structure of cell wall teichoic acids of staphylococci. *Arch Microbiol* **135**, 215–223.

Euzéby, J. P. (1997). List of bacterial names with standing in nomenclature: a folder available on the Internet. *Int J Syst Bacteriol* **47**, 590–592.

Felsenstein, J. (1995). PHYLIP – Phylogeny Inference Package (version 3.57c). Seattle, WA: University of Washington.

Fitch, W. M. (1971). Toward defining the course of evolution: minimum change for a specified tee topology. *Syst Zool* **20**, 406–416.

Fox, G. E., Wisotzkey, J. D. & Jurtshuk, P., Jr (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee soecies identity. *Int J Syst Bacteriol* **42**, 166–170.

Goh, S. H., Potter, S., Wood, J. O., Hemmingsen, S. M., Reynolds, R. P. & Chow, A. W. (1996). HSP60 gene sequences as universal targets for microbial species identification: studies with coagulase-negative staphylococci. *J Clin Microbiol* **34**, 818–823.

Goh, S. H., Santucci, Z., Kloos, W. E., Faltyn, M., George, C. G., Driedger, D. & Hemmingsen, S. M. (1997). Identification of *Staphylococcus* species and subspecies by the chaperonin 60 gene identification method and reverse checkerboard hybridization. *J Clin Microbiol* **35**, 3116–3121.

Gribaldo, S., Cookson, B., Saunders, N., Marples, R. & Stanley, J. (1997). Rapid identification by specific PCR of coagulase-negative staphylococcal species important in hospital infection. *J Med Microbiol* **46**, 45–53.

Gupta, R. S. (1995). Evolution of the chaperonin families (Hsp60, Hsp10 and Tcp-1) of proteins and the origin of eukaryotic cells. *Mol Microbiol* **15**, 1–11.

Ieven, M., Verhoeven, J., Pattyn, S. R. & Goosens, H. (1995). Rapid and economical method for species identification of clinically significant coagulase-negative staphylococci. *J Clin Microbiol* **33**, 1060–1063.

Jukes, T. H. & Cantor, C. R. (1969). Evolution of protein molecules. In *Mammalian Protein Metabolism*, vol. 3, pp. 21–132. Edited by H. N. Munro. New York: Academic Press.

Kloos, W. E. (1997). Taxonomy and systematics of staphylococci indigenous to humans. In *The Staphylococci in Human Disease*, pp. 113–137. Edited by K. B. Crossley & G. L. Archer. New York: Churchill Livingstone.

Kloos, W. E. & Bannerman, T. L. (1995). *Staphylococcus* and *Micrococcus*. In *Manual of Clinical Microbiology*, 6th edn, pp. 282–298. Edited by P. R. Murray, E. J. Baron, M. A. Pfaller, F. C. Tenover & R. H. Yolken. Washington, DC: American Society for Microbiology.

Kloos, W. E., Ballard, D. N., George, C. G., Webster, J. A., Hubner, R. J., Ludwig, W., Schleifer, K.-H., Fiedler, F. & Schubert, K. (1998). Delimiting the genus *Staphylococcus* through description of *Macrococcus caseolyticus* gen. nov., comb. nov. and *Macrococcus equipercicus* sp. nov., *Macrococcus bovicus* sp. nov. and *Macrococcus carouselicus* sp. nov. *Int J Syst Bacteriol* **48**, 859–877.

Martin de Nicolas, M. M., Vindel, A. & Saez-Nieto, J. A. (1995). Epidemiological typing of clinically significant strains of coagulase-negative staphylococci. *J Hosp Infect* **29**, 35–43.

O'Donnell, A. G., Nahaie, M. R., Goodfellow, M., Minnikin, D. E. & Hajek, V. (1985). Numerical analysis of fatty acid profiles in the identification of staphylococci. *J Gen Microbiol* **131**, 2023–2033.

Olsen, G. J., Woese, C. R. & Overbreek, R. (1994). The winds of (evolutionary) change: breathing new life into microbiology. *J Bacteriol* **176**, 1–6.

Pai, S., Esen, N., Pan, X. & Musser, J. M. (1997). Routine rapid mycobacterium species assignment based on species-specific allelic variation in the 65-kilodalton heat shock protein gene (hsp65). *Arch Pathol Lab Med* **121**, 859–864.

Pillet, J. & Orta, B. (1981). Species and serotypes in coagulase-negative Staphylococci. In *Staphylococci and Staphylocoocal Infections*, pp. 147–152. *Zentralblatt fur Bakteriologie, Mikrobiologie und Hygiene. I. Abteilung Supplemente*, vol. 10. Edited by J. Jeljaszewicz, W. Hryniewicz, L. M. Switalski & S. Szmigielski. Stuttgart & New York. Gustav Fischer Verlag.

Rosypal, S., Rosypalova, A. & Horejs, J. (1966). The classification of micrococci and staphylococci based on their DNA base composition and Adansonian analysis. *J Gen Microbiol* **44**, 281–292.

Schumacher-Perdreau, F., Rotering, H. & Pulverer, G. (1983). Cell wall analysis and taxonomy of staphylococci. *Zentbl Bakteriol Mikrobiol Hyg 1 Abt Orig A* **256**, 25–36.

Smith, S. (1997). *Genetic Data Environment*. Harvard Genome Laboratory. Chicago: University of Illinois.

Stackebrandt, E. & Goebel, B. M. (1994). A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* **44**, 846–849.

Stackebrandt, E., Liesack, W. & Witt, D. (1992). Ribosomal RNA and rDNA sequence analysis. *Gene* **115**, 255–260.

Steingrube, V. A., Gibson, J. L., Brown, B. A., Zhang, Y., Wilson, R. W., Rajagopalan, M. & Wallace, R. J., Jr (1995). PCR amplification and restriction endonuclease analysis of a 65-kilodalton heat shock protein gene sequence for taxonomic separation of rapidly growing mycobacteria. *J Clin Microbiol* **33**, 149–153.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighing, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Viale, A. M., Arakaki, A. K., Soncini, F. C. & Ferreyra, R. G. (1994). Evolutionary relationships among eubacterial groups as inferred from GroEL (chaperonin) sequence comparisons. *Int J Syst Bacteriol* **44**, 527–533.

Wayne, L. G., Brenner, D. J., Colwell, R. R. & 9 other authors (1987). International Committee on Systematic Bacteriology. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* **37**, 463–464.

Yamamoto, S. & Harayama, S. (1995). PCR amplification and direct sequencing of *gyrB* genes with universal primers and their application to the detection and taxonomic analysis of *Pseudomonas putida* strains. *Appl Environ Microbiol* **61**, 1104–1109.

Zakrzewska-Czerwinska, J., Gaszewska-Mastalarz, A., Lis, B., Gamian, A. & Mordarski, M. (1995). *Staphylococcus pulvereri* sp. nov., isolated from human and animal specimens. *Int J Syst Bacteriol* **45**, 169–172.