The Journal of Immunology

RESEARCH ARTICLE | JANUARY 01 2001

egc, A Highly Prevalent Operon of Enterotoxin Gene, Forms a Putative Nursery of Superantigens in *Staphylococcus aureus* FREE

Sophie Jarraud; ... et. al

J Immunol (2001) 166 (1): 669–677.
https://doi.org/10.4049/jimmunol.166.1.669

egc, A Highly Prevalent Operon of Enterotoxin Gene, Forms a Putative Nursery of Superantigens in *Staphylococcus aureus*

Sophie Jarraud,* Marie Alix Peyrat,† Annick Lim,‡ Anne Tristan,* Michèle Bes,* Christophe Mougel,§ Jerome Etienne,* François Vandenesch,* Marc Bonneville,† and Gerard Lina[1]*

The recently described staphylococcal enterotoxins (SE) G and I were originally identified in two separate strains of *Staphylococcus aureus*. We have previously shown that the corresponding genes *seg* and *sei* are present in *S. aureus* in tandem orientation, on a 3.2-kb DNA fragment (Jarraud, J. et al. 1999. *J. Clin. Microbiol.* 37:2446–2449). Sequence analysis of *seg-sei* intergenic DNA and flanking regions revealed three enterotoxin-like open reading frames related to *seg* and *sei*, designated *sek*, *sel*, and *sem*, and two pseudogenes, ψ *ent1* and ψ *ent2*. RT-PCR analysis showed that all these genes, including *seg* and *sei*, belong to an operon, designated the enterotoxin gene cluster (*egc*). Recombinant SEG, SEI, SEK, SEL, and SEM showed superantigen activity, each with a specific Vβ pattern. Distribution studies of genes encoding superantigens in clinical *S. aureus* isolates showed that most strains harbored such genes and in particular the enterotoxin gene cluster, whatever the disease they caused. Phylogenetic analysis of enterotoxin genes indicated that they all potentially derived from this cluster, identifying *egc* as a putative nursery of enterotoxin genes. ***The Journal of Immunology*, 2001, 166: 669–677.**

Staphylococcus aureus produces a wide variety of toxic proteins, including the staphylococcal enterotoxins (SEs)[2] A-E and G-J, the toxic shock syndrome toxin-1 (TSST-1), and the exfoliative toxins (ETs) A and B. These toxins were initially described as being responsible for specific acute staphylococcal toxemia syndromes, such as toxic shock syndrome (TSS) and staphylococcal scarlet fever (SSF) (both due to TSST-1, SEB, and SEC), scalded skin syndrome (SSSS, due to the ETs), and staphylococcal food poisoning (due to the SEs) (1, 2).

SEs and TSST-1 share common structural and biological properties, suggesting that they derived from a common ancestor (3). They display significant homology in their primary sequence and secondary and tertiary structures (3). Based on amino acid sequence comparisons, SEs have been divided into several groups; one includes SEA, SEE, SEJ, SED, and SEH, and another SEB and SEC, whereas SEG and SEI could not be clearly attributed to a specific group (4, 5). Biologically, SEs and TSST-1 exhibit superantigen activity, stimulating polyclonal T cell proliferation through coligation between MHC class II molecules on APCs and the variable portion of the T cell Ag receptor β-chain (TCR Vβ) (3). The pattern of Vβ activation is specific for each of these superantigens (3). T cell/APC activation by these toxins leads to the release of various cytokines/lymphokines and IFN, enhances endotoxic shock, and causes T and B cell immunosuppression, all of which may hinder the immune response against bacterial infection (5–8).

SEG and SEI are recently described SEs (4). We have previously reported the involvement of these toxins in TSS and SSF (9). The SEG and SEI genes (*seg* and *sei)* were originally identified in two separate strains (4), but we have shown that, when present, *seg* and *sei* coexist in all clinical isolates of *S. aureus* examined to date (9). Moreover, we found that the two genes were in tandem orientation on the same 3.2-kb DNA fragment. As Munson et al. (4) have found that the *seg* transcript is unusually large (~6.7 kb), we postulated that the *seg* transcript might encode additional genes, including *sei*. The aim of this study was to identify and characterize the genes that are cotranscribed with *seg*.

Materials and Methods

Strains

S. aureus A900322, isolated from a patient with TSS, was shown to have the genotype *sea*$^-$, *seb*$^-$, *sec*$^-$, *sed*$^-$, *see*$^-$, *seg*$^+$, *seh*$^-$, *sei*$^+$ (9), and was used as a *seg* and *sei* reference strain. *S. aureus* RN450 (*sea*$^-$, *seb*$^-$, *sec*$^-$, *sed*$^-$, *see*$^-$, *seg*$^-$, *seh*$^-$, *sei*$^-$) was used as a negative control for SE genes. *S. aureus* MJB1316 (a gift from Sibyl Munson, University of Wisconsin, Madison, WI), an RN450 derivative that contains the cloned *seg* gene on the staphylococcal expression vector pRN5548 (4), was used as *seg* positive control. The following *S. aureus* strains were used to check the specificity of PCR amplification: FDA-S6 (ATCC 13566 (*sea*$^+$ *seb*$^+$)), FRI-137 (ATCC 19095 (*sec*$^+$ *seg*$^+$ *seh*$^+$ *sei*$^+$)), FRI-1151 m (*sed*$^+$), FRI-326 (ATCC 27664 (*see*$^+$)), FRI-569 (ATCC 51811 (*seh*$^+$)), FRI-1169 (*tst*$^+$), TC-7 (*eta*$^+$ *seg*$^+$ *sei*$^+$), and TC-146 (*etb*$^+$ *seg*$^+$ *sei*$^+$) (9). Two hundred thirty *S. aureus* clinical isolates were collected by the Center National de Référence des Toxémies à Staphylocoques (Lyon, France) between January 1998 and December 1999. They were isolated from 58 patients with *S. aureus* infection (arthritis, skin infection, pneumonia, or infective endocarditis), 102 patients with acute toxemia (TSS, SSF, or SSSS), and 70 asymptomatic nasal carriers. All strains were collected from hospitals located throughout France and were identified as *S. aureus* by their ability to coagulate citrated rabbit plasma (bioMérieux, Marcy-l'Etoile, France) and to produce a clumping factor (Staphyslide Test; bioMérieux). *Escherichia coli* TG1 was used for plasmid amplification and genetic manipulations.

*Centre Nationale des Toxémies à Staphylococques, Faculté de Médecine Laennec, Lyon, France; †Institute National de la Santé et de la Recherche Médicale Unité 463, Institut de Biologie, Nantes, France; ‡Institute National de la Santé et de la Recherche Médicale Unité 277, Institut Pasteur, Paris, France; and §Unité Mixte de Recherche Centre National de la Recherche Scientifique 5557, Laboratoire d'Ecologie Microbienne, Universite Claude Bernard, Villeurbanne, France

Received for publication July 26, 2000. Accepted for publication October 4, 2000.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] Address correspondence and reprint requests to Dr. Gerard Lina, Centre Nationale des Toxémies à Staphylococques, EA1655, Faculté de Médecine Laennec, Rue G Parradin, 69372 Lyon Cedex 08, France. E-mail address: geralina@univ-lyon1.fr

[2] Abbreviations used in this paper: SE, staphylococcal enterotoxin; ET, exfoliative toxin; ORF, open reading frame; SD, Shine-Dalgarno; SSF, staphylococcal scarlet fever; SSSS, scalded skin syndrome; TSS, toxic shock syndrome; TSST, TSS toxin.

0022-1767/01/$02.00

DNA amplification and sequencing

DNA was extracted from A900322 cultures and used as a template for amplification with primers sei-1 and seg-2 (Table I) in conditions described in detail elsewhere (9). Primers wsei and wseg (Table I) were designed following identification of suitable hybridization sites in the *sei* and *seg* genes and were compatible with the Clontech Genome Walker kit (Ozyme; Montigny-Le Bretonneux, France), which is suitable for cloning unknown DNA sequences adjacent to a known sequence. This kit was used, according to the supplier's instructions, to identify *sei* and *seg* flanking regions using primers hindIII and wsei (Table I) on a *Hin*dIII chromosomal digest for the amplification of the *sei*-upstream region; and primers hpaI and wseg (Table I) on an *Hpa*I chromosomal digest for the amplification of the *seg*-downstream region. PCR products were analyzed by electrophoresis through 0.8% agarose gels (Sigma, St. Louis, MO), purified using the High Pure PCR Product Purification kit (Boehringer Mannheim, Meylan, France), and sequenced (Genome Express, Grenoble, France). Sequences were compiled, analyzed, and compared using Blast (http://www.ncbi.nlm.nih.gov/BLAST), GeneJokey, and ClustalX software (European Bioinformatics Institute, Cambridge, U.K., http://www.ebi.ac.uk) (10).

Phylogenic reconstruction

The sequences of SE-related genes were obtained from GenBank: *sea* accession number M18970; *seb* accession number M11118; *sec1* accession number X05815; *sed* accession number M28521; *see* accession number M21319; *seh* accession number U11702; and *ent* accession number U93688. Nucleotide sequences of these genes and open reading frames (ORFs) encoded by *egc* were aligned using the multiple alignment ClustalX software (10). The evolutionary distances were determined by the method of Kimura, and these values were used to construct a dendrogram by means of the neighbor-joining method using the Phylip package (European Bioinformatics Institute). At least 1000 bootstrap trees were generated for each data set to investigate the stability of phylogenic relationships using the Seqboot module of Phylip package. Similar phylogenic analysis was conducted using the corresponding amino acids sequences.

Toxin-gene detection

Sequences specific for *sea-e*, *seg-i*, *tst*, *eta*, and *etb*, encoding SEA-E, SEG-I, TSST-1, ETA, and ETB, respectively, were detected by PCR, as previously described (9). DNA from clinical isolates was extracted from cultures and used as a template for amplification with the primers described in Table I (Eurogentec, Seraing, Belgium). Amplification of *gyrA* was used as a control to confirm the quality of each DNA extract and the absence of PCR inhibitors (9). All PCR products were analyzed by electrophoresis through 1% agarose gels (Sigma).

Detection of bacterial RNA by RT-PCR

Total RNA was extracted from staphylococcal cultures by using RNeasy spin columns (Qiagen, Courtaboeuf, France). cDNA was synthesized using Ready-To-Go RT-PCR beads (Pharmacia Biotech, Orsay, France) by incubating 0.1 μg of total RNA with the following pairs of primers (primer 5′, sel3), (sel-4, sel-5), (sel1, sel2), (invsel2, invsem1), (sem1, invsei1), (sei1, sei2), (invsei2, ψent2), (ψent1, invsek1), (sek1, sek2), (invsek2, invseg1), (seg1, seg2), (invseg2, primer 3′) (Table I). The reaction mixtures were incubated with each primer pair described above, at 42°C for 30 min for reverse transcription, followed by 30 cycles of amplification (1-min denaturation at 94°C, 1-min annealing at 55°C, and 1-min extension at

Table I. *Sequences of primers used in this study*[a]

Primer	Location within *egc*[b]	Oligonucleotide Sequence (5′-3′)
wseg	5014–5040	CGTCTCCACCTGTTGAAGGAAGAGGAG
wsei	2494–2525	GTATTGTCCTGATAAAGTGGCCCCTCCATAC
primer5′	1–19	GTCCCGTTAGGAGTCATAC
sel-3	427–449	GCATTGTTTACACTACATATTGC
sel-4	244–270	CTGTTTGTTCAATAGTAAGTAGGATTG
sel-5	557–581	GTTGATACAATTGATTTTACTGTCG
sel-1	481–506	AGTTTGTGTAAGAAGTCAAGTGTAGA
sel-2	630–660	ATCTTTAAATTCAGCAGATATTCCATCTAAC
invsel-2	630–660	GTTAGATGGAATATCTGCTGAATTTAAAGAT
rsel-1	443–461	CAGAATTCTGTAGTGTAAACAATGCATATGCAAATG
rsel-2	1123–1156	GCCTGCAGTTATGTAAATAAATAAACATCAATATGATAGTC
sem-1	1785–1811	CTATTAATCTTTGGGTTAATGGAGAAC
invsem-1	1785–1811	GTTCTCCATTAACCCAAAGATTAATAG
sem-2	2085–2110	TTCAGTTTCGACAGTTTTGTTGTCAT
rsem-1	1469–1495	CAGAATTCTTTTGCTATTCGCAAAATCATATCGCA
rsem-2	2126–2155	GCCTGCAGTCAACTTTCGTCCTTATAAGATATTTCTAC
sei-1	2260–2281	CTCAAGGTGATATTGGTGTAGG
invsei-1	2260–2281	CCTACACCAATATCACCTTGAG
sei-2	2886–2915	GTTACTATCTACATATGATATTTCGACATC
invsei-2	2886–2915	GATGTCGAAATATCATATGTAGATAGTAAC
rsei-1	2262–2289	CAGAATTCCAAGGTGATATTGGTGTAGGTAACTTAA
rsei-2	2883–2918	GCCTGCAGTTAGTTACTATCTACATATGATATTTCGACATCAAG
ent-1	3352–3375	ACGTAGATTTGTTTGGGACAAACT
ent-2	3502–3529	GTGCTGTTATGTTTTTCTTATTAGTAGG
sek-1	3969–3988	ACGTGGCAATTAGACGAGTC
invsek-1	3969–3988	GACTCGTCTAATTGCCACGT
sek-2	4415–4444	GATTGATCTTGATGATTATGAGAATGAAAG
invsek-2	4415–4444	CTTTCATTCTCATAATCATCAAGATCAATC
rsek-1	3885–3918	CAGAATTCGAAGTAGACAAAAAAGATTTAAAGAAAAAATCTG
rsek-2	4531–4568	GCCTGCAGTTAATCTTTATATAAAAATACATCAATATGATAATTAG
seg-1	4979–5003	TAAGGGAACTATGGGTAATGTAATG
invseg-1	4979–5003	CATTACATTACCCATAGTTCCCTTA
seg-2	5514–5541	GAACAAAAGGTACTAGTTCTTTTTTAGG
invseg-2	5514–5541	CCTAAAAAAGAACTAGTACCTTTTGTTC
rseg-1	4926–4950	CAGAATTCCAACCCGATCTTAAATTAGACGAAC
rseg-2	5593–5627	GCCTGCAGTCAGTGAGTATTAAGAAATACTTCCATTTTAATAC
primer3′	6163–6189	CTTTAACCTCATAAATTAGCAGTAGTC

[a] Restriction sites for *Eco*RI and *Pst*I are underlined.
[b] *egc*, Enterotoxin gene cluster.

72°C). The RT-PCR products were then analyzed by electrophoresis through 1% agarose gel. RNA extracts were tested for DNA contamination by preincubating the reaction mixtures at 95°C for 10 min to inactivate reverse transcriptase before the RT-PCR.

Production and purification of recombinant enterotoxins

Primers were designed following identification of suitable hybridization sites in *sel, sem, sei, sek,* and *seg* (Table I). The 5′ primers were chosen within the coding sequence of each gene, omitting the region predicted to encode the signal peptide, as determined by hydrophobicity analysis according to Kyte and Doolitttle (11) with GeneJockey software and SignalP V1.1 World Wide Web Prediction Server (http://www.cbs.dtu.dk/services/SignalP/) (12); the 3′ primers were chosen to overlap the stop codon of each gene. A restriction site was included in each primer (Table I). DNA was extracted from A900322 or MJB1316 and used as a template for PCR amplification. PCR products and plasmid DNA were prepared using the Qiagen plasmid kit. PCR fragments were digested with *Eco*RI and *Pst*I (Boehringer Mannheim) and ligated (T4 DNA ligase; Boehringer Mannheim) with the pMAL-c2 expression vector from New England Biolabs (Ozyme) digested with the same restriction enzymes. The resulting plasmids were transformed into *E. coli* TG1. The integrity of the ORF of each construct was verified by DNA sequencing of the junction between pMAL-c2 and the different inserts. The fusion proteins were purified from cell lysates of transfected *E. coli* by affinity chromatography on an amylose column according to the supplier's instructions (New England Biolabs).

T cell proliferation assays

PBL from healthy donors were cultured in 24-well plates (10^6 cells/well) in RPMI 1640 medium supplemented with 8% pooled human serum and 10 μg/ml recombinant staphylococcal toxin. rIL-2 (50 IU/ml) was added on day 5. When necessary, T cell cultures were diluted in IL-2-supplemented medium until TCR analysis. We used as controls T cells from the same donors that were stimulated with 0.5 μg/ml *Phaseolus vulgaris* leucoagglutinin (PHAL) (Sigma).

Flow cytometry

The following mAb (mAb; specificity indicated in brackets) were used for flow cytometry: E2.2E7.2 (Vβ2), LE89 (Vβ3), IMMU157 (Vβ5.1), 3D11 (Vβ5.3), CRI304.3 (Vβ6.2), 3G5D15 (Vβ7), 56C5.2 (Vβ8.1/8.2), FIN9 (Vβ9), C21 (Vβ11), S511 (Vβ12), IMMU1222 (Vβ13.1), JU74 (Vβ13.6), CAS1.1.13 (Vβ14), Tamaya1.2 (Vβ16), E17.5F3 (Vβ17), BA62.6 (Vβ18), ELL1.4 (Vβ20), IG125 (Vβ21.3), IMMU546 (Vβ22), and HUT78.1 (Vβ23). These mAb, and CD4- and CD8-specific mAb, were purchased from Beckman/Coulter/Immunotech (Marseille, France). Cells were phenotyped by indirect immunofluorescence, as described previously (13). Briefly, cells were incubated with unconjugated mAb for 30 min at room temperature, then washed and incubated with FITC-conjugated rabbit anti-mouse Ig antiserum (BioAtlantic, Nantes, France) for 30 min on ice. After washing, cells were analyzed by flow cytometry on a FACScan apparatus (Becton Dickinson, Mountain View, CA) using the LYSYS II software package on a FACstation.

Immunoscope analysis

Total RNA was extracted using the Trizol reagent (Life Technologies, Gaithersburg, MD). TCR β-chain-specific primers were as described previously (14), and reverse transcription, PCR amplification, and run-off steps were performed as reported previously (15). Fluorescent DNA products were loaded on a sequencing gel and analyzed with the Immunoscope software (16).

Statistical analysis

χ^2 test was used to determine whether the distribution of *egc*, *sea*, *seb*, *sec*, *sed*, *see*, *seh*, *tst*, *eta*, and *etb* was significantly different in isolates from asymptomatic nasal carriers and patients with *S. aureus* infection or acute toxemia; $p < 0.05$ was considered statistically significant.

Results and Discussion

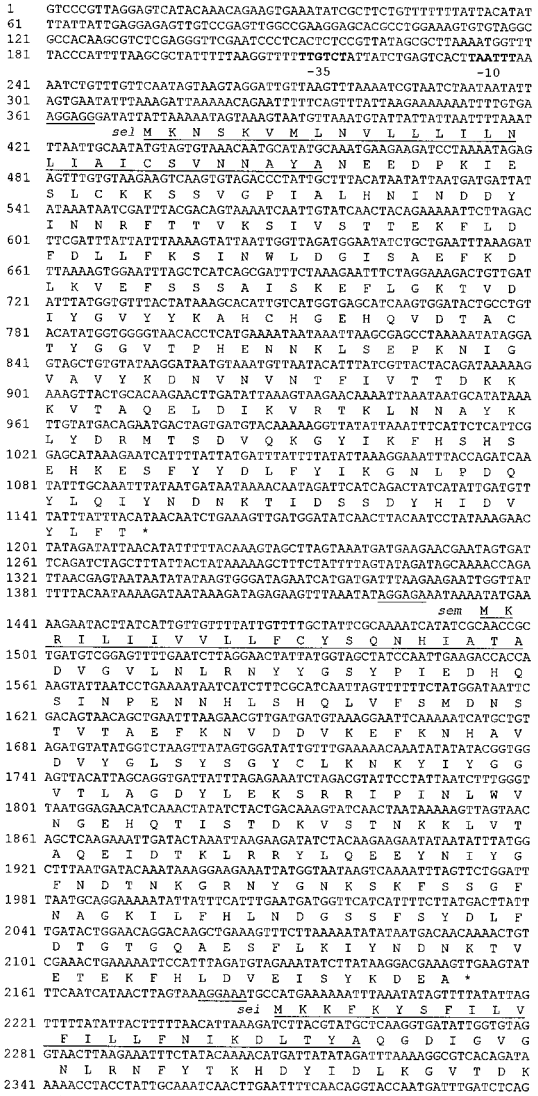
Identification of the seg *and* sei *flanking regions*

When this work was initiated, the coding regions of only *seg* and *sei* were available, and the two genes were known to be in tandem orientation, separated by a 1.9-kb DNA fragment in *S. aureus* strain A900322 (4, 9). A 3.2-kb fragment was thus amplified by PCR with primers sei1 and seg2 and was then sequenced. The intergenic 1.9-kb DNA sequence contained three open reading frames (ORF1, 2, and 3) of 399, 327, and 777 bp, respectively. Comparison of the deduced amino acid sequences of these ORFs with translated sequences from GenBank showed that the putative proteins corresponding to these ORFs had substantial sequence similarities to known SEs: ORF1 exhibited homology to the N-terminal region of SEB; ORF2 to the C-terminal region of SEC; and ORF3 to SEA (Table II). The PCR "walking" strategy was

Table II. *Percentage of amino acid sequence identity among the staphylococcal superantigenic toxins*[a]

	% Sequence Identity															
Toxin	SEA	SEB	SEC1	SED	SEE	TSST-1	SEG	SEH	SEI	SEJ	ENT	ORF1[b] ENT1	ORF2[c] ENT2	ORF3[d] SEK	ORF4[e] SEL	ORF5[f] SEM
SEA	100	33	30	50	83	NS	27	37	39	64	31	35	32	39	37	35
SEB		100	68	35	32	26	43	33	31	33	30	47	63	32	36	29
SEC1			100	31	29	30	41	27	26	30	26	46	66	29	33	26
SED				100	52	NS	27	35	33	51	33	35	35	38	39	41
SEE					100	26	27	35	35	63	32	34	31	39	37	37
TSST-1						100	33	NS	NS	NS	NS	NS	31	NS	NS	NS
SEG							100	34	28	29	28	33	49	31	30	28
SEH								100	33	35	28	31	36	38	34	31
SEI									100	34	67	30	34	31	31	57
SEJ										100	33	33	37	38	42	33
ENT											100	30	33	28	29	57
ENT1												100	NS	26	29	28
ENT2													100	33	43	38
SEK														100	42	28
SEL															100	31
SEM																100

[a] Amino acid sequences were compared by using the Blast 2 sequences method with open gap of 11 and extension gap penalties of 1 (10).
[b] Blast search result in GenBank using ORF1: homology with SEB [identities = 56/119 (47%), positives = 84/119 (70%), gaps = 6/119 (5%)].
[c] Blast search result in GenBank using ORF2: homology with SEC [identities = 69/103 (66%), positives = 90/103 (86%)].
[d] Blast search result in GenBank using ORF3: homology with SEA [identities = 101/258 (39%), positives = 151/258 (58%), gaps = 11/258 (4%)].
[e] Blast search result in GenBank using ORF4: homology with SEJ [identities = 103/238 (43%), positives = 141/238 (58%), gaps = 12/238 (5%)].
[f] Blast search result in GenBank using ORF5: homology with SEI [identities = 135/235 (57%), positives = 175/235 (74%), gaps = 1/235 (0%)].



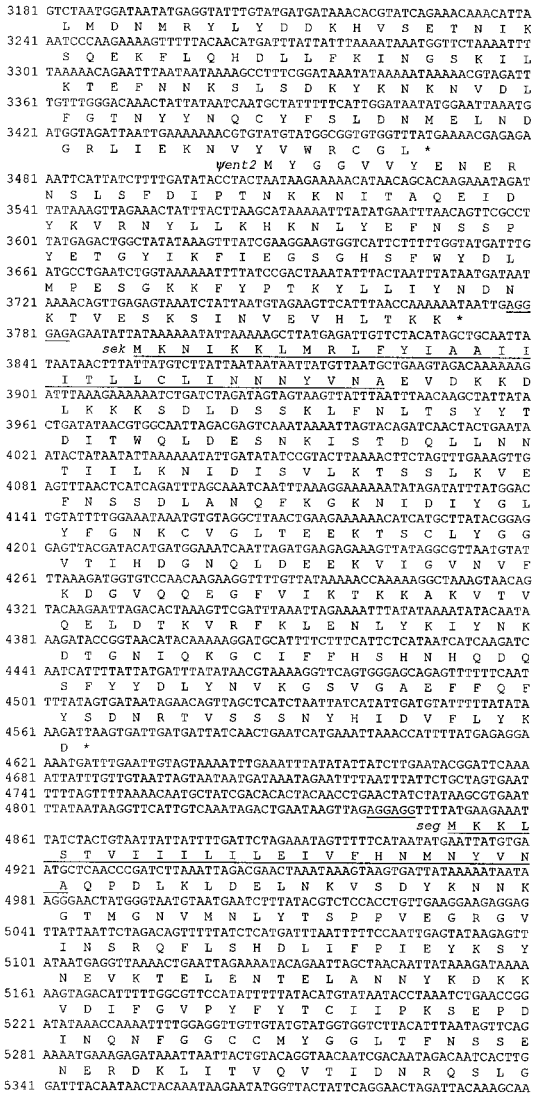


FIGURE 1. Complete *egc* nucleotide sequence. The deduced amino acid sequence of each ORF, with its putative signal peptide (underlined) and its stop codon (asterix), is indicated below. The putative SD sequence (underlined), the proposed −10 and −35 consensus sequences (bold characters), and the inverted repeats representing the putative transcription terminator (horizontal arrows) are also indicated. These sequence data are available from GenBank under accession number AF285760.

chosen to identify the *seg* and *sei* flanking regions. The use of primers wsei and hindIII on *Hin*dIII digests allowed us to amplify and sequence the 3.2 kb of DNA upstream of *sei*. Analysis of this sequence showed two significant ORFs (ORF4 and ORF5) of 783 and 720 bp, respectively. ORF4 exhibited homology with SEJ, and ORF5 with SEI (Table II). The use of primers wseg and hpaI on *Hpa*I digests amplified a 0.8-kb fragment downstream of *seg*. Sequence analysis of this fragment revealed no other significant ORFs. The concatenated sequence of *seg-sei*-intergenic, -upstream and -downstream regions was validated by sequencing a 6.189-kb PCR fragment encompassing the whole region (Fig. 1). Although *sei* in strain A900322 was 100% homologous with the sequence

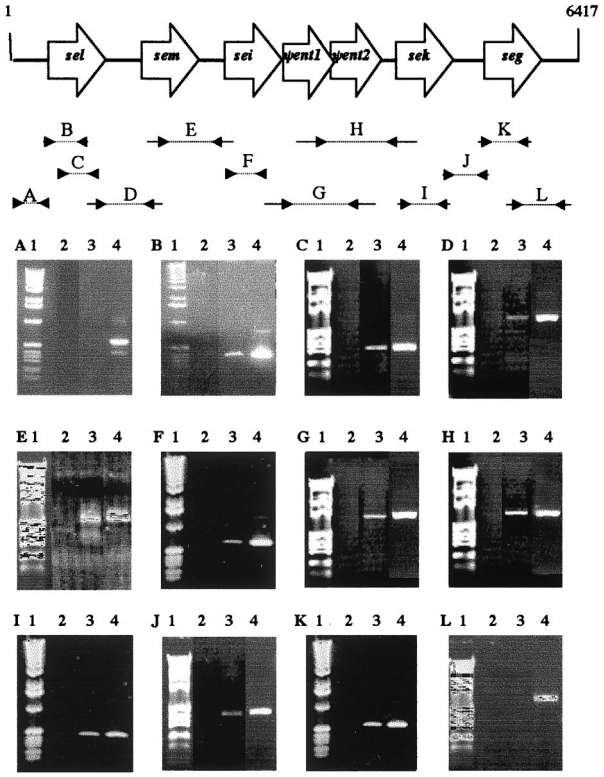


FIGURE 2. Analysis of *egc* transcripts by RT-PCR. cDNA was prepared from *S. aureus* A900322 total RNA and subjected to PCR using the primer pairs A to K (schematically represented in the *upper part* of the figure, and described in *Materials and Methods*). *A–K*, Correspond to the results obtained using the corresponding primer pairs. *Lane 1*, Molecular size marker; *lane 2*, RT-PCR negative control (RT-PCR with heat-inactivated reverse transcriptase); *lane 3*, RT-PCR from extract of A900322; *lane 4*, PCR positive control (A900322 DNA as template).

deposited in GenBank (accession number AF064774), *seg* in strain A900322 showed one mutation, corresponding to a Leu→Pro substitution at position 29. This new variant was designated SEG_{L29P}.

As ORFs 1–5 were homologous but not identical with any known enterotoxins, we speculated that they corresponded to new enterotoxins. However, ORF1 and 2 were at least 50% shorter than any of the known enterotoxins. ORF-1 possesses a satisfactory Shine-Dalgarno (SD) sequence (TGGAGT-N7-AUG, consensus AGGAGG-N6/10-AUG) but, in comparison with SEB, to which it is highly related, shows a large deletion of its 3′ end, which corresponds to a region that is essential for biological (superantigenic) activity (17). ORF2 has neither an SD sequence nor a signal peptide, and resembles an N-terminal-truncated SEC. Accordingly, ORF1 and 2 were designated ψ ent1 and 2, respectively, meaning they represent pseudogenes with no likely biological function. In contrast, ORFs 3, 4, and 5 had sizes consistent with active enterotoxin-like molecules. ORF5 possesses a satisfactory SD sequence and translation start site, whereas ORF3 and ORF4 have an adequate SD sequence in front of a noncanonical, although suitable (18), translation start site (ATT) coding thethionine (Fig. 1). Thus, ORF3, ORF4, and ORF5 were designated *sek*, *sel*, and *sem*, respectively. Thus, the 6301-bp DNA region identified in this study contains *seg* and *sei* plus three potential enterotoxin genes (*sek*, *sel*, and *sem*) and two pseudogenes (Ψ*ent1*, Ψ*ent2*), all in the same orientation (Fig. 1). We designated this region *egc* for "enterotoxin gene cluster." With the exception of plasmid pIB485, which contains *sed* and *sej* in opposite orientations separated by 895 nucleotides, and the staphylococcal pathogenicity island, which contains *tst* and *ent* separated by 10.234 kb (19), no such gene cluster organization has ever been described for enterotoxin genes. It is likely that this organization was generated through gene duplication and variation from an ancestral gene. The proposed molecular mechanism involved in *egc* formation is unequal crossing-over,

Table III. *Flow cytometry analysis of PBL from donor A and B, cultured for 12 days with 10 μg/ml of SEI or SEK*[a]

	SEI[c]		SEK[c]	
TCR Vβ[b]	Donor A	Donor B	Donor A	Donor B
1	6.8	10.1		
5.1	36.0	34.0		
5.2	10.4	11.0		
5.3	5.8	4.0		
9			43.5	45.1
23	2.0	2.0		

[a] Results are expressed as percentage of total cells in the culture.
[b] Results for TCR Vβ2, Vβ3, Vβ4, Vβ6.7, Vβ7, Vβ8, Vβ9, Vβ11, Vβ12, Vβ13.2, Vβ13.6, Vβ14, Vβ16, Vβ17, Vβ18, Vβ20, Vβ21.3, and Vβ22 were negatives.
[c] CD4/CD8 ratios were of 13.4 (donor A) and 21.7 (donor B) in SEI cultures (preferred expansion of the CD4 subset) and of 2.9 and 3.3 in SEK cultures (no gross alteration of CD4/CD8 in this case).

Table IV. *Flow cytometry analysis of PBL from donor C and D, cultured for 12 days with 10 μg/ml of SEG, SEI, SEL, or SEM*[a]

	PHAL[c] (0.5 μg/ml)	SEG[c]		SEI[c]		SEL[c]		SEM[c]	
TCR Vβ[b]	Donor C	Donor C	Donor D	Donor C	Donor D	Donor C	Donor D	Donor C	Donor D
1	2.8			7.8	10.0				
5.1	5.7			35.4	32.5	9.1	22.3		
5.3	0.7			3.7	4.4				
7	0.5					34.0	42.1		
13.6	1.3	0.4	1.5						
14	3.5	55.5	65.0						
18	0.2							1.1	
21.3	1.5							30.7	27.1
22	2.5					1.8	2.9		
23	0.5			2.0	2.4				

[a] Results are expressed as percentage of total cells in the culture.
[b] Results for TCR Vβ2, Vβ3, Vβ4, Vβ5.2, Vβ4, Vβ6.7, Vβ8, Vβ9, Vβ11, Vβ12, Vβ13.2, Vβ16, Vβ17, and Vβ20 were negatives.
[c] CD4/CD8 ratios were of 3.7 (donor C) for PHAL culture, of 4.4 (donor C) and 3.3 (donor D) for SEG cultures, of 6.8 and 6.2 for SEI cultures, of 3.3 and 3.2 for SEL cultures, and of 6.0 and 7.2 for SEM cultures.

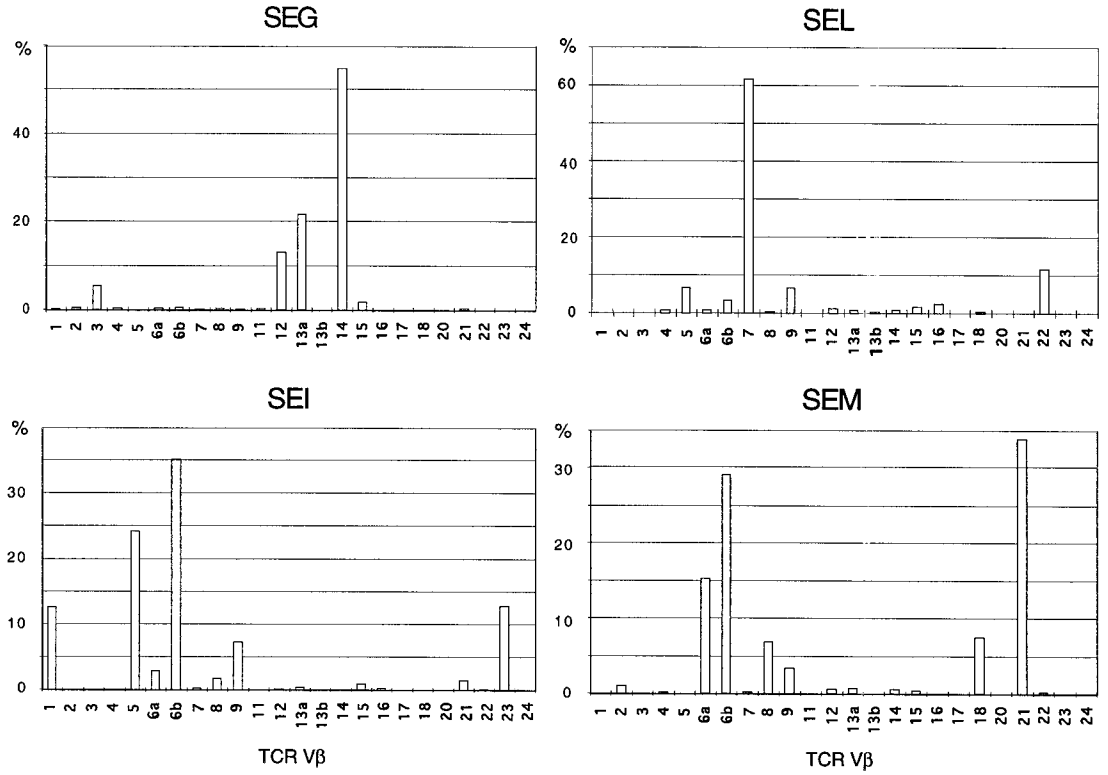


FIGURE 3. Analysis of the TCR Vβ repertoire of superantigen-stimulated cultured PBL by semiquantitative immunoscopy. The results shown are those obtained with T cell lines derived from donor D (Table III). In all cases, junctional transcripts derived from amplified Vβ segments showed diverse lengths with a pseudogaussian distribution (14, 16) (data not shown). Similar results were obtained with superantigen-stimulated cells from donor C.

which can be generated when recombination occurs between nonallelic regions by misalignment (20). This hypothesis is supported by the presence of homologous genes and pseudogenes in the same DNA region.

Transcriptional analysis

As mentioned above, Munson et al. (4) reported an unusually large (6.7-kb) *seg* transcript. To investigate whether this transcript was polycistronic, i.e., encoded one or more of the ORFs identified in *egc*, c-DNA was generated from strain A900322 total RNA by reverse transcription and amplified by PCR using primer pairs located within each gene and bracketing adjacent genes. Abundant RT-PCR products (B to K) of the expected size were obtained using the corresponding primer pairs (Fig. 2). In contrast, no RT-PCR product A (primer 5′, sel3) nor L (primer invseg2 and primer 3′) was obtained (Fig. 2). These results suggest that the seven genes and pseudogenes composing *egc* are cotranscribed, and that the 5′ and 3′ ends of the transcript must be close to the beginning of *sel* and to the end of *seg*, respectively. Sequence analysis revealed putative −10 and −35 promotor sequences (TTGTCT-N15-TAATTT-N134-ATT) upstream of the *sel* start codon. The 3′ end may lie at an inverted repeat at position 6018–6067, which is a potential transcription termination signal, 5830 nucleotides downstream of the putative transcription start site. These results suggest that *egc* is an operon. However, we could not rule out the coexistence of alternative transcription start sites and/or termination sites resulting in partial *egc* transcription. The size of the *egc* transcript was slightly shorter than that previously estimated by means of Northern blot analysis (6.7 kb) by Munson et al. This discrepancy is most probably due to technical reasons, as Northern blot analysis permits only a rough estimate of RNA size.

Superantigen activity

The association of related genes that are cotranscribed suggested that the resulting peptides might have complementary effects on the host's immune response. One hypothesis was that gene recombination had created new variants of toxins differing by their superantigen profiles. Purified recombinant SEL, SEM, SEI, SEK, and SEG_{L29P} expressed in *E. coli* were studied for their ability to induce selective expansion of T cells bearing particular TCR Vβ regions in short-term PBL culture. As shown in Tables III and IV, recombinant SEL SEM, SEI, and SEK consistently induced selective expansion of distinct sets of Vβ subpopulations. By contrast, SEG_{L29P} failed to trigger expansion of any of the 23 Vβ subsets. The sum of results obtained with each of these recombinant toxins globally corresponded to the selective expansion of Vβ subpopulations induced by crude supernatant of staphylococcal culture of strains that harbored *egc* (data not shown). This suggested that the maltose-binding protein portion of the fusion toxins did not significantly influence the Vβ specificity of these superantigens. To investigate whether the L29P mutation could explain the lack of superantigen activity, a rSEG with an L29 codon was constructed from *S. aureus* strain MJB1316 (which contains the cloned *seg* on a plasmid) and then expressed in *E. coli*, and the superantigen activity of this toxin was tested. SEG_{L29}.induced selective expansion of Vβ14 and, to a lesser extent, Vβ13.6,0T cells (Table III). The L29P mutation thus accounts for the complete loss of superantigen activity. Computer modeling of the two-dimensional structure (21) of the wild-type and mutated proteins revealed no major conformational differences between the two proteins (not shown). It is likely that L29 is located at a position crucial for proper

Table V. *Toxin production by 160* S. aureus *strains associated with various clinical syndromes*

Type of Infection	No. of Strains Tested	No. (%) of Toxin-Positive Strains[a] and p Value[b]									
		seg sei	*sea*	*seb*	*sec*	*sed*	*see*	*seh*	*tst*	*eta*	*etb*
Nasal carriage	70	40 (57)-[c]	13 (19)-	5 (7)-	13 (19)-	10 (14)-	0 (0)-	3 (4)-	8 (11)-	1 (1)-	0 (0)-
Suppurative disease	58	39 (67) NS	7 (12) NS	11 (20) NS	5 (9) NS	3 (5) NS	0 (0) NS	6 (10) NS	7 (20) NS	0 (0) NS	0 (0) NS
Arthritis	10	6 (60)	2 (20)	2 (20)	0 (0)	1 (10)	0 (0)	0 (0)	1 (10)	0 (0)	0 (0)
Furunculosis	13	11 (85)	2 (15)	4 (31)	1	1 (8)	0 (0)	2 (15)	2 (15)	0 (0)	0 (0)
Cutaneous abscess	10	7 (70)	2 (20)	1 (10)	3 (8)	1 (10)	0 (0)	2 (20)	3 (30)	0 (0)	0 (0)
Pneumonia	18	11 (61)	1 (6)	5 (28)	0 (0)	0 (0)	0 (0)	1 (6)	0 (0)	0 (0)	0 (0)
Endocarditis	7	4 (57)	0 (0)	0 (0)	1 (14)	0 (0)	0 (0)	1 (14)	1 (14)	0 (0)	0 (0)
TSS	63	34 (54) NS	21 (33) NS	6 (10) NS	5 (8) NS	8 (13) NS	0 (0) NS	4 (6) NS	24 (38) $p = 0.04$	5 (8) NS	3 (5) NS
Acute presentation of TSS	32	19 (59) NS	11 (34) NS	3 (9) NS	4 (12) NS	4 (12) NS	0 (0) NS	0 (0) NS	14 (44) .07	2 (6) NS	1 (3) NS
Mild presentation of TSS (SSF)	31	15 (48) NS	10 (32) NS	3 (10) NS	1 (3) NS	4 (13) NS	0 (0) NS	4 (13) NS	10 (32) NS	3 (10) NS	2 (6) NS
SSSS	39	36 (92) $p = 0.03$	0 (5) NS	3 (8) NS	1 (3) NS	0 (0) NS	0 (0) NS	1 (3) NS	1 (3) NS	29 (74) $p < 0.001$	26 (77) $p < 0.001$
Generalized exfoliative syndrome	27	25 (96)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (4)	0 (0)	18 (67)	18 (67)
Bullous impetigo	12	11 (92)	2 (17)	3 (25)	1 (8)	0 (0)	0 (0)	0 (0)	1 (8)	11 (92)	8 (77)

[a] All genes were detected by PCR assay as described in *Materials and Methods*.
[b] χ^2 test was used to determine whether the distribution of *egc*, *sea*, *seb*, *sec*, *sed*, *see*, *seh*, *tst*, *eta*, and *etb*, were significantly different in *S. aureus* isolates from asymptomatic nasal carriers (reference group) and patients with either *S. aureus* suppurative diseases or acute; $p < 0.05$ was considered statistically significant.
[c] Reference group for statistical analysis.

superantigen/MHC II interaction. In addition to the selective expansion of TCR Vβ subsets observed with the different toxins, flow cytometry revealed preferential expansion of CD4 T cells in SEI and SEM cultures (Table III). By contrast, the CD4/CD8 ratios in SEK-, SEL-, and SEG-stimulated T cell lines were close to those in fresh PBL. This phenomenon, which was observed with cells from several donors, may reflect a variable contribution of the CD4 coreceptor to the T cell activation process, depending on the affinity of the TCR for the superantigen/MHC complex (22, 23).

To document the TCR Vβ composition of superantigen-stimulated T cell lines and the clonal diversity of the expanded TCR Vβ subsets, the size distribution of PCR-amplified TCR β-chain junctional products was studied using the Immunoscope technique (14–16). Results of this molecular analysis were in good overall agreement with those obtained by flow cytometry, as similar dominant TCR Vβ subsets were identified with the two approaches (Fig. 3, Tables III and IV, and data not shown for rSEK). Slight discrepancies observed in some instances may have been due to selective expansion of particular members of a given Vβ subfamily not recognized by available mAb (e.g., Vβ6 or Vβ13), and to the fact that TCR Vβ frequencies were estimated with a semiquantitative PCR technique, which might have led to slight over- or underestimation of particular Vβ subsets. Additionally, Immunoscope analysis showed that the complementarity-determining region 3 size distribution of TCR β-chain junctional transcripts within expanded Vβ subsets was pseudogaussian in all superantigen-stimulated cultures, reflecting a high level of polyclonality (data not shown). This was further confirmed by sequence analysis of TCR β junctional transcripts derived from some expanded TCR Vβ subsets (e.g., V$\beta5^+$ cells in SEL and SEI cultures) (not shown). Taken together, these TCR repertoire studies confirmed the superantigenic nature of the new toxins identified in this study.

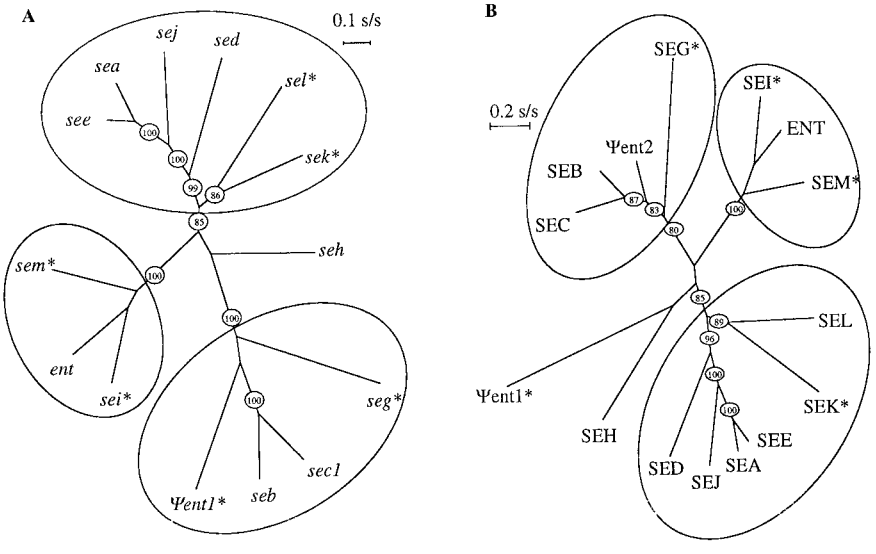


FIGURE 4. Reconstruction of phylogenic tree of SE genes (*A*) and toxins (*B*). The nucleic, and peptidic, sequences of all SE-related genes and toxins were obtained from GenBank. They were aligned using the multisequence alignment program ClustalX. Phylogenic relationships were inferred using the Phylip software package. The evolutionary distances were determined by the method of Kimura, and these values were used to construct a dendrogram by means of the neighbor-joining method. As the Phylip package was not able to confidently branch *ψent2*, this gene is not presented in the tree. The numbers at the nodes are the proportion of 1000 bootstrap resamplings that support the topology shown. Only bootstrap values >70% are indicated. Genes belonging to *egc* and toxins encoded by *egc* are indicated by an asterisk. Monophyletic groups of genes and toxins are circled.

Distribution of egc *in human* S. aureus *isolates*

We then analyzed the distribution of *egc* (by PCR amplification encompassing *sei* to *seg*) and that of all known enterotoxins (by selective PCR (9)) in 230 *S. aureus* strains isolated in various clinical settings (nasal carriage, suppurative infection, and toxemia). As shown in Table V, the majority of the isolates harbored gene(s) encoding superantigenic toxins, whatever the clinical setting. *seg-sei* (and thus *egc*) were present in most toxemia strains (59% in TSS, 48% in SSF, and 92% in SSSS), and also in most strains associated with suppurative infections (67%) and nasal carriage (57%). Moreover, *egc* appeared to be the most frequent superantigens in *S. aureus*, whatever the clinical setting. The prevalence of *egc* in strains associated with SSSS, a disease caused by ETs, was significantly higher than that in nasal carriage strains (χ^2 test, $p = 0.03$) (Table V). This could reflect the clonal origin of the strains associated with SSSS, as previously suggested by phage typing, pulsed-field gel electrophoresis, and amplified fragment length polymorphism (24, G. Lina, manuscript in preparation). The strains associated with TSS were significantly more frequently TSST-1 producers than were nasal carriage strains (χ^2 test, $p = 0.04$) (Table V), whereas no significant difference was observed between the two groups of strains regarding the presence of *egc*. Thus, the superantigens produced by *egc* must have a role other than the induction of toxemia. As each toxin encoded by *egc* was associated with a complementary pattern of Vβ subset usage, a bacterium that produces such a panel of superantigens theoretically has a marked capacity for stimulating polyclonal T cell proliferation and thus for inducing several deleterious effects, including immune anergy by T cell suppressor activity, B cell depletion, and inhibition of Ab responses (6–8). We speculate that the apparent redundancy of these superantigens confers a selective advantage toward colonization and/or invasion of human and not only for toxemia.

Phylogenic analysis of enterotoxin genes

The high prevalence of *egc* among staphylococcal isolates raises the possibility that this locus acts as a reservoir of enterotoxin genes. Phylogenic analysis was conducted, including all known enterotoxins and enterotoxin-like toxins in *S. aureus*. Phylogenetic trees, constructed from the nucleic acid sequences of these genes and from the amino acid sequences of the corresponding toxins by using the neighbor-joining method, were superimposable. The position of Ψ*ent1* and its products was unstable, as reflected by the low bootstrap value at the node from which they branched (52.5% and 62%, respectively). As the Phylip package was not able to confidently branch ψ *ent2*, this gene is not presented in the tree. All other nodes were well supported (>70% bootstrap values) (Fig. 4). We identified three monophyletic groups within the tree: one composed of *sea*, *see*, *sej*, *sed*, *sek*, *sel*, and *seh*; another including *seb*, *sec*, ψ *ent2*, *seg*, and probably ψ *ent1*; and a third including *sei*, *sem*, and *ent* (a putative enterotoxin identified in the staphylococcal pathogenicity island (19)). Each of these clusters contained one or more genes encoded in *egc*. Remarkably, each of the predicted *egc* products showed the strongest homology with one of the known enterotoxins encoded outside *egc* on monocistronic loci. This phylogenic organization could be interpreted as suggesting that gene ancestors of enterotoxin genes encoded outside *egc* derive from *egc*. Thus, *egc* would appear to be an enterotoxin gene nursery. The mechanism by which gene diversity has been generated in *egc* and then exported on the mode of a single gene to other regions of the chromosome remains to be elucidated.

In conclusion, we have identified an enterotoxin gene cluster in *S. aureus*, which was probably generated from an ancestral gene through gene duplication and variation. This cluster is an operon, encoding SEG, SEI, and three new enterotoxins designated SEL, SEM, and SEK. All these toxins exhibit superantigenic properties associated with specific Vβ subsets. The wide distribution of *egc* in clinical isolates suggests that it is beneficial for *S. aureus*. Finally, phylogenetic analysis of all known enterotoxins indicates that they all potentially derived from this cluster, inferring that *egc* is in an enterotoxin gene nursery.

Note added in proof. While the present article was under review, Williams et al. (25) reported the discovery of a novel genetic locus within *S. aureus* that encodes a cluster of at least five exotoxin-like proteins designated the staphylococcal exotoxin-like genes 1 to 5 (set1 to set5). Comparison of the nucleotide sequences of set1-5 with that of *egc* revealed that the two clusters are distinct.

Acknowledgments

We thank Gregoire Cozon (Unité d'Immunologie, Hôpital de la Croix-Rousse, Lyon, France), Patrick Blanco (Baylor Institute, Dallas, Texas), and Mohamed Hamidou (Department of Internal Medicine, University Hospital, Nantes, France) for preliminary characterization of staphylococcal superantigens; Philippe Kourilsky (Institut National de la Santé et de la Recherche Médicale Unité 277, Institut Pasteur, Paris, France) for scientific advises; and Nicole Viollant, Christine Courtier, and Christine Gardon for technical assistance.

References

1. Arbuthnott, J. P., D. C. Coleman, and J. S. de Azavedo. 1990. Staphylococcal toxins in human disease. *J. Appl. Bacteriol. Symp. Suppl. 19:101S.*
2. Lina, G., Y. Gillet, F. Vandenesch, M. E. Jones, D. Floret, and J. Etienne. 1997. Toxin involvement in staphylococcal scalded skin syndrome. *Clin. Infect. Dis. 25:1369.*
3. Marrack, P., and J. Kappler. 1990. The staphylococcal enterotoxins and their relatives. *Science 248:705.*
4. Munson, S. H., M. T. Tremaine, M. J. Betley, and R. A. Welch. 1998. Identification and characterization of staphylococcal enterotoxin types G and I from *Staphylococcus aureus. Infect. Immun. 66:3337.*
5. Dinges, M. M., P. M. Orwin, and P. M. Schlievert. 2000. Exotoxins of *Staphylococcus aureus. Clin. Microbiol. Rev. 13:16.*
6. Hu, H. L., W. D. Cornwell, T. J. Rogers, and Y. S. Lin. 1996. In vivo analysis of a superantigen-induced T cell suppressor factor. *Cell. Immunol. 167:285.*
7. Lussow, A. R., and H. R. MacDonald. 1994. Differential effects of superantigen-induced "anergy" on priming and effector stages of a T cell-dependent antibody response. *Eur. J. Immunol. 24:445.*
8. Sundstedt, A., S. Grundstrom, and M. Dohlsten. 1998. T cell- and perfor-independent depletion of B cells in vivo by staphylococcal enterotoxin A. *Immunology 95:76.*
9. Jarraud, S., G. Cozon, F. Vandenesch, M. Bes, J. Etienne, and G. Lina. 1999. Involvement of enterotoxins G and I in staphylococcal toxic shock syndrome and staphylococcal scarlet fever. *J. Clin. Microbiol. 37:2446.*
10. Tatusova, T. A., and T. L. Madden. 1999. Blast 2 sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett. 174:247.*
11. Kyte, J., and F. R. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol. 157:105.*
12. Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng. 10:1.*
13. Vié, H., S. Chevalier, R. Garand, J. P. Moisan, V. Praloran, M. C. Devilder, J. F. Moreau, and J. P. Soulillou. 1989. Clonal expansion of lymphocytes bearing the γδ T-cell receptor in a patient with large granular lymphocyte disorder. *Blood 74:285.*
14. Pannetier, C., J. P. Levraud, A. Lim, J. Even, and P. Kourilsky. 1996. The Immunoscope technique for analysis of T-cell repertoires. In *The Human Antigen T-Cell Receptor: Selected Protocols and Applications.* J. Oksenberg, ed. Landes, R.G. Company, Austin, p. 162.
15. Even, J., A. Lim, I. Puisieux, L. Ferradini, P. Y. Dietrich, A. Toubert, T. Hercend, F. Triebel, C. Pannetier, and P. Kourilsky. 1995. T-cell repertoires in healthy and diseased human tissues analyzed by T-cell receptor β-chain CDR3 size determination: evidence for oligoclonal expansions in tumors and inflammatory diseases. *Res. Immunol. 146:65.*
16. Pannetier, C., M. Cochet, S. Darche, A. Casrouge, M. Zoller, and P. Kourilsky. 1993. The sizes of the CDR3 hypervariable regions of the murine T-cell receptor β chains vary as a function of the recombined germ-line segments. *Proc. Natl. Acad. Sci. USA 90:4319.*

17. Metzroth, B., T. Marx, M. Linnig, and B. Fleischer. 1993. Concomitant loss of conformation and superantigenic activity of staphylococcal enterotoxin B deletion mutant proteins. *Infect. Immun. 61:2445.*
18. Falconi, M., M. Brombach, C. O. Gualerzi, and C. L. Pon. 1991. In vivo transcriptional pattern in the *infC* operon of *Bacillus stearothermophilus. Mol. Gen. Genet. 227:60.*
19. Lindsay, J. A., A. Ruzin, H. F. Ross, N. Kurepina, and R. P. Novick. 1998. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus. Mol. Microbiol. 29:527.*
20. Lewin, B. 1990. Structural genes evolve in families. In *Gene IV.* B. Lewin, ed. Oxford University Press, New York, p. 497.
21. Geourjon, C., and G. Deleage. 1995. SOPMA: significant improvements in protein secondary structure prediction by prediction from multiple alignments. *Comput. Appl. Biosci. 11:681.*
22. Hong, S. C., G. Waterbury, and C. A. Janeway, Jr. 1996. Different superantigens interact with distinct sites in the Vβ domain of a single T cell receptor. *J. Exp. Med. 183:1437.*
23. Huang, B., A. Yachou, S. Fleury, W. A. Hendrickson, and R. P. Sekaly. 1997. Analysis of the contact sites on the CD4 molecule with class II MHC molecule: co-ligand versus co-receptor function. *J. Immunol. 158:216.*
24. Murono, K., K. Fujita, and H. Yoshioka. 1988. Microbiologic characteristics of exfoliative toxin-producing *Staphylococcus aureus. Pediatr. Infect. Dis. J. 7:313.*
25. Williams, R. J., J. M. Ward, B. Henderson, S. Poole, B. P. O'Hara, M. Wilson, and S.P. Nair. 2000. Identification of a novel gene cluster encoding staphylococcal endotoxin-like proteins: characterization of the prototypic gene and its protein product, SET1. *Infect. Immun. 68:4407.*