JOURNAL OF CLINICAL MICROBIOLOGY, Aug. 2005, p. 3985–3994 Vol. 43, No. 8
0095-1137/05/$08.00+0 doi:10.1128/JCM.43.8.3985–3994.2005

Comparative Sequencing of the Serine-Aspartate Repeat-Encoding Region of the Clumping Factor B Gene (*clfB*) for Resolution within Clonal Groups of *Staphylococcus aureus*

Larry Koreen,[1,2] Srinivas V. Ramaswamy,[3]† Steven Naidich,[4] Irina V. Koreen,[1] Gavin R. Graff,[5] Edward A. Graviss,[3] and Barry N. Kreiswirth[2]*

New Jersey Medical School and Graduate School of Biomedical Sciences, University of Medicine and Dentistry of New Jersey, Newark, New Jersey 07103[1]; Public Health Research Institute, International Center for Public Health, Newark, New Jersey 07103[2]; Department of Pathology, Baylor College of Medicine, Houston, Texas 77030[3]; eGenomics, New York, New York 10013[4]; and Department of Pediatrics, Penn State College of Medicine, Hershey, Pennsylvania 17033[5]

Received 23 December 2004/Returned for modification 12 February 2005/Accepted 16 March 2005

Molecular techniques such as *spa* typing and multilocus sequence typing use DNA sequence data for differentiating *Staphylococcus aureus* isolates. Although *spa* typing is capable of detecting both genetic micro- and macrovariation, it has less discriminatory power than the more labor-intensive pulsed-field gel electrophoresis (PFGE) and costly genomic DNA microarray analyses. This limitation hinders strain interrogation for newly emerging clones and outbreak investigations in hospital or community settings where robust clones are endemic. To overcome this constraint, we developed a typing system using DNA sequence analysis of the serine-aspartate (SD) repeat-encoding region within the gene encoding the keratin- and fibrinogen-binding clumping factor B (*clfB* typing) and tested whether it is capable of discriminating within clonal groups. We analyzed 116 *S. aureus* strains, and the repeat region was present in all isolates, varying in sequence and in length from 420 to 804 bp. In a sample of 36 well-characterized genetically diverse isolates, *clfB* typing subdivided identical *spa* and PFGE clusters which had been discriminated by whole-genome DNA microarray mapping. The combination of *spa* typing and *clfB* typing resulted in a discriminatory power (99.5%) substantially higher than that of *spa* typing alone and closely approached that of the whole-genome microarray (100.0%). *clfB* typing also successfully resolved genetic differences among isolates differentiated by PFGE that had been collected over short periods of time from single hospitals and that belonged to the most prevalent *S. aureus* clone in the United States. *clfB* typing demonstrated in vivo, in vitro, and interpatient transmission stability yet revealed that this locus may be recombinogenic in a primarily clonal population structure. Taken together, these data show that the SD repeat-encoding region of *clfB* is a highly stable marker of microvariation, that in conjunction with *spa* typing it may serve as a DNA sequence-based alternative to PFGE for investigating genetically similar strains, and that it is useful for analyzing collections of isolates in both long-term population-based and local epidemiologic studies.

Strain typing techniques for *Staphylococcus aureus*, the leading cause of nosocomial infections (28), have become widely used. These techniques aid in both local/short-term epidemiologic outbreak investigations and global/long-term population-based studies of methicillin-resistant and -susceptible *S. aureus* (MRSA and MSSA, respectively). Macrorestriction digests using pulsed-field gel electrophoresis (PFGE) have been shown to be highly effective in outbreak settings (48). Multilocus enzyme electrophoresis (MLEE), multilocus sequence typing (MLST), and sequence analysis of the repeat region within the coagulase gene, i.e., *coa* typing, are effective techniques for analyzing *S. aureus* strains in long-term study settings (10, 42), and the sequence analysis of the repeat region within the protein A gene, *spa* typing, can effectively be used in both settings (22, 41). However, *spa* typing, like MLST and other techniques, is capable of only a certain degree of resolution beyond which clonal groups of isolates cannot be subdivided. This limitation hinders strain interrogation for newly emerging clones and outbreak investigations in hospital or community settings where robust clones, such as the *spa* type 2 clone, which is the most prevalent *S. aureus* strain associated with health care-related infections in the United States (25) and also the strain that recently has acquired resistance to vancomycin (55), are endemic. In order to overcome this constraint, a highly informative locus capable of discriminating within clonal groups is necessary.

For species such as *Bacillus anthracis*, markers have been found that subdivide closely related strains (36). However, for *S. aureus*, no such marker exists aside from PFGE, a technique that is difficult to standardize, analyze, and database (41, 52). DNA sequence-based techniques overcome these limitations of PFGE and are considerably faster to perform (9, 24, 41). However, *spa* typing alone, which has the single highest discriminatory power of the DNA sequence-based *S. aureus* typing techniques (22, 37), sometimes fails to discriminate between two closely related strains that can be differentiated by PFGE and is often considered inferior to PFGE with regard to discriminating rapidly accumulating genetic microvariation

* Corresponding author. Mailing address: PHRI TB Center, International Center for Public Health, 225 Warren St., Newark, NJ 07103. Phone: (973) 854-3240. Fax: (973) 854-3241. E-mail: barry@phri.org.
† Present address: Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115.

(22, 41, 46). It is therefore important to combine the use of *spa* typing with another locus that can be analyzed for genetic variation when microresolution of strains is sought. This approach has been successfully applied to *Neisseria meningitidis* (12) and *Enterococcus faecalis* (27). Furthermore, because *spa* typing has been shown to be in agreement with MLST (1, 6, 31–34), a typing technique that helps to resolve differences among similar *spa* types would also be useful in subdividing the clonal complexes identified by MLST.

Potential gene candidates for providing high-level genetic resolution are those that encode MSCRAMMs (microbial surface components recognizing adhesive matrix molecules) (15) such as *spa*, because they contain variable repeat region sequences and because the surface proteins they encode interact with the environment and may change accordingly. The clumping factor B gene (*clfB*) of *S. aureus* encodes an MSCRAMM protein that binds to fibrinogen (30) and keratin and facilitates *S. aureus* colonization in human nares (32, 53). Clumping factor B has also been implicated in the pathogenesis of *S. aureus*-induced endocarditis (11). Furthermore, *clfB* has an unusual repeat region encoding a directly repeating serine-aspartate (SD) dipeptide that, presumably due to nucleotide mutations and ease of repeat duplication/deletion via slipped-strand mispairing (51) during replication, had appeared to be a highly evolving region that could be used to complement *spa* typing (L. Koreen, S. Ramaswamy, S. Naidich, E. A. Graviss, and B. Kreiswirth, Abstr. 103rd Gen. Meet. Am. Soc. Microbiol. 2003, abstr. C-429, 2003). Here we report on the development and evaluation of this new DNA sequence-based typing scheme for *S. aureus* using the *clfB* SD repeat region for indexing genetic microvariation within groups of closely related strains and for differentiating among identical *spa* types. In order to explore the unique discriminatory capability of *clfB* typing, evolutionary analyses were performed to investigate the selection pressures on this repeat region.

MATERIALS AND METHODS

Bacterial strains. The following 116 *S. aureus* strains were analyzed to determine whether use of the SD repeat-encoding region within the *clfB* gene (*clfB* typing) provided suitable discriminatory power for differentiating strains deemed closely related by other molecular markers discussed below. Thirty-six strains, which had been selected from over 2,000 isolates, were recovered from 10 countries on four continents over a period of 4 decades and included the fully sequenced COL strain. These 36 strains formed a highly diverse collection representing the most abundant lineages and the breadth of genetic variation of *S. aureus* (14). These strains (11 MRSA and 25 MSSA) had previously been analyzed with a whole-genome DNA microarray, MLEE, PFGE, and *spa* and *coa* typing (14, 22); included were 14 clonal strains with electrophoretic type (ET) 234 in lineage H1 by MLEE associated with toxic shock syndrome (TSS) and well differentiated by the DNA microarray (14). Fourteen other strains were recovered during an *S. aureus* carriage and infection surveillance study in 2002 from patients at a New York City hospital over a 2-week period. These strains had *spa* type 2 and type 2-related genotypes (the most prevalent *S. aureus* clonal lineage found in U.S. hospitals [25]). Another 21 strains were collected from cystic fibrosis patients at a Midwestern hospital, and these strains also had *spa* type 2 and type 2-related genotypes. The in vitro stability of the *clfB* repeat region was tested using a strain that was passed extensively in the laboratory for 6 weeks (44); three isolates picked from single colonies from the first week and three from the last week were *clfB* typed (analysis of three isolates picked from single colonies is sufficient for finding existing genotypic discordance among isolates from the same source ([5]). The in vivo stability of *clfB* typing was tested using three carriage isolates obtained over a 21-month period from each of four hemodialysis patients consistently carrying strains of the same PFGE-determined genotype as part of our laboratory's longitudinal hemodialysis patient *S. aureus* carriage study (L. Koreen, C. Kutler, B. Mathema, R. Abder, B. Shopsin, W. Eisner, B. Saïd-Salim, B. Raucher, N. Levin, A. Kaufman, B. Koll, and B. Kreiswirth, Abstr. 40th Annu. Meet. Infect. Dis. Soc. Am. 2002, abstr. 125, 2002). Two other strains obtained from the same patient over a 3-month period in this hemodialysis study were identical according to PFGE analysis except for a two-band difference, which was interpreted as being directly due to the loss of the methicillin resistance element, staphylococcal cassette chromosome *mec*, based on the difference in size between the two PFGE bands and on methicillin sensitivity and *mecA* Southern hybridization testing of the strains. These two strains were used to test whether *clfB* typing was excessively variable, a disadvantage occasionally attributed to PFGE (12, 17). Also, 18 strains from a well-characterized Centers for Disease Control and Prevention collection of strains obtained from different outbreaks (43, 47, 50) were used to study interpatient transmission stability of the *clfB* repeat region. These strains had been previously *spa* typed; seven identified as outbreak I MRSA strains (strains SB-3, -5, -10, -12, -15, -19, and -20), were obtained from the Iowa Veterans Affairs Medical Center, and four others, identified as outbreak II MSSA strains (strains SB-2, -4, -6, and -11), were isolated from a contaminated anesthetic (41). Finally, seven strains (Newman, 476, 252, MW2, NCTC 8325, Mu50, and N315) whose *clfB* sequences were already available in public databases were also *clfB* typed.

Molecular analysis. DNA was isolated (41) and the *clfB* SD repeat region of each isolate was PCR amplified with a Geneamp System 9700 thermocycler (Applied Biosystems, Inc., Foster City, CA) using the following primers: *clfB*F, 5′-CAG CAG TAA ATC CGA AAG ACC C-3′; *clfB*R, 5′-CAC CTT TAG GAT TTG ATG GTG C-3′. Unincorporated nucleotides and primers were removed with a QIAGEN Qiaquick PCR purification kit (Valencia, CA). DNA sequencing reactions were performed with a BigDye Terminator cycle sequencing kit (Applied Biosystems, Inc., Foster City, CA). Sequencing reaction products were purified using Centrisep columns (Princeton Separations, Adelphia, NJ). Sequence data generated with an ABI 377 (Applied Biosystems, Inc., Foster City, CA) automated instrument were assembled and edited electronically with the ALIGN, EDITSEQ, and MEGALIGN programs (DNASTAR, Madison, WI). Contigs were built with maximal stringency using SEQUENCHER v. 4.1.2 (Gene Codes Corporation, Ann Arbor, MI). Final sequence size was verified by correlation with PCR amplicon size. Each unique *clfB* repeat region nucleotide sequence was given a *clfB* type numeric identifier.

spa and *coa* typing, macrorestriction analysis using PFGE, and analysis using the DNA microarray have been previously described (7, 22, 41, 42). Briefly, a *spa* type is defined by the makeup of the variable number tandem repeat region in the 3′ end of the staphylococcal protein A gene (*spa*). The different repeats, designated randomly with letters (A to Z, A2, and B2, etc.), that comprise a *spa* type vary from one another by at least one point mutation and are generally each composed of 24 bp. The different types of organization of a repeat region are termed repeat profiles and range from 1 to 16 repeats in length. *coa* typing is similar to *spa* typing, except the variable number tandem repeat region in the coagulase gene consists of 81-bp repeats. Within each typing scheme, isolates with similar repeat profiles have in common sequential point mutations and were grouped together as part of the same numeric *spa* or *coa* lineage (sublineages indicated with letters). For the previously performed PFGE analysis of the diversity collection (22), isolates with patterns with differences of up to six bands were considered as possibly related (48) and had been grouped together into the same alphabetic lineage, with every unique pattern within a lineage given a secondary numeric code. Patterns not falling into any lineages were identified with numeral 1. DNA microarray experiments performed by Fitzgerald et al. (14) on each of the 36 diversity collection strains used in this study assayed for the presence of over 90% of the open reading frames within *S. aureus*. Based on the presence or absence of these 2,817 open reading frames, hierarchical cluster analysis had been previously used to construct a dendrogram showing relatedness among these 36 strains (14).

Comparing genetic markers and evolutionary analysis. The percent concordance between any two typing techniques for a particular set of isolates was calculated as previously described with cross-classification analysis of all possible pairs of those isolates (22, 38, 39). Simpson's index of diversity, which indicates the probability that among a group of isolates any two randomly selected isolates will have different genotypes (19), was used to measure discriminatory power. Molecular evolutionary analyses of the *clfB* repeats were performed as previously described for *spa* and *coa* repeats (22). Using MEGA version 2.1 (23), the overall mean Nei-Gojobori (Jukes-Cantor-corrected) method (29) with pairwise deletion handling of gaps and standard error determined with 1,000 bootstrap replications was used to calculate the average number of synonymous substitutions per synonymous site (dS) and nonsynonymous substitutions per nonsynonymous site (dN). Subsequently, to determine selection pressure on the repeats, *Z* tests were done with the null hypothesis, dN = dS, and the following three alternative

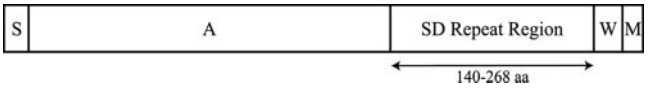

FIG. 1. Structural organization of the *clfB* protein. Repeat region ranged in size from 140 to 268 amino acids (aa), making the total protein approximately 800 to 928 amino acids long (21). S, signal sequence; A, binding region for keratin and fibrinogen; W, wall-spanning region; M, membrane-spanning region (21, 32).

hypotheses: dN ≠ dS (test of neutrality), dN > dS (positive selection), and dN < dS (purifying selection). Other testing with chi-square/Fisher's exact tests was performed using EpiInfo 2002 (Centers for Disease Control and Prevention, Atlanta, GA). Statistical significance was determined with a P value of <0.05.

RESULTS

Choice of *clfB* for high-resolution strain typing. *clfB* encodes the clumping factor B protein, whose structural organization is shown in Fig. 1 (21). This gene was chosen as a typing target because of its unique repeat region containing direct repeats that each encode three SD dipeptides. The overall repeat region of *clfB* (an example from strain NCTC 8325, whose genome has been fully sequenced, is shown in Fig. 2) is typically larger than that of *spa*, and the individual repeats are 18 bp (TCN-GAY-TCN-GAY-AGY-GAY, with N equaling A, C, G, or T and Y equaling C or T), 6 bp shorter than the *spa* repeats. For these reasons we believed there would be increased slipped-strand mispairing (51) during replication and recombination, resulting in more genetic variation in the repeat region of *clfB* than in *spa* among related strains. Furthermore, of different genes encoding SD repeat-containing proteins in *S. aureus*, only *clfB* was found in all strains assayed previously (35, 49), and *clfB* had a repeat region that was somewhat smaller, and thus easier to sequence and analyze, than other genes, such as *clfA* (21).

***clfB* SD repeats, types, and lineages.** One hundred sixteen *S. aureus* strains, whose origins are described in Materials and Methods, were characterized by *clfB* genotyping. All strains had the *clfB* gene present. Initially, PCR amplicons were analyzed for restriction fragment length polymorphisms using BamHI and Tsp45I enzymes to test whether strains could be well differentiated by restriction fragment length polymorphism patterns alone. However, it was determined that the restriction digests did not produce adequate strain resolution (data not shown). DNA sequencing of all *clfB* SD repeat regions was then performed. The sizes of the *clfB* SD repeat region PCR amplicons ranged from 627 to 1,011 bp. The average number of repeats in each strain's repeat region was 38 (range, 24 to 46 repeats), and the average repeat region size was 677 bp (range, 420 to 804 bp). Due to point mutations, a total of 81 different repeats (each given a numeric identifier) were identified among the strains tested (Table 1). Each repeat varies from another by at least one point mutation; occasionally, this variation created a nonsynonymous change.

Seventy-two (72) of the repeats had the standard 18-bp length, and nine of the repeats had 12-bp lengths. The 12-bp repeats (each given a numeric and an asterisk identifier) were the result of slippage events where, presumably through slipped-strand mispairing during replication, either 12-bp segments were deleted from two contiguous repeats (six from each repeat) or 6 bp was deleted from only one repeat. Immediately downstream of any slippage event, the repeating DNA sequence continued in standard fashion.

```
5'-cagcagtaaatccgaaagacccaactccagggccgccggttgacccagaaccaagtccagac
     A  V  N  P  K  D  P  T  P  G  P  P  V  D  P  E  P  S  P  D
   ccagaaccagaaccaacgccagatccagaaccaagtccagacccagaaccggaaccaagc
    P  E  P  E  P  T  P  D  P  E  P  S  P  D  P  E  P  E  P  S
   ccagacccggatccggattcggattcagacagtgactcaggctcagacagcgactcaggt
    P  D  P  D  P  D  S  D  S  D  S  D  S  G  S  D  S  D  S  G
   tcagatagcgactcagaatcagatagcgattcggattccagacagtgattcagattcagac
    S  D  S  D  S  E  S  D  S  D  S  D  S  D  S  D  S  D  S  D
   agcgactcagaatcagatagcgactcagaatcagatagtgagtcagattcagacagtgac
    S  D  S  E  S  D  S  D  S  E  S  D  S  E  S  D  S  D  S  D
   tcggactcagacagtgattcagactcagatagcgattcagactcagatagcgattcagac
    S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D
   tcagacagcgattcagattcagacagcgactcagattcagacagcgactcagactcagat
    S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D
   agcgactcagactcagacagcgactcagattcagatagcgattcagactcagacagcgac
    S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D
   tcagactcagacagcgactcagactcagatagcgactcagattcagatagcgattcagac
    S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D
   tcagacagcgactcagattcagatagcgattcggactcagacagcgattcagattcagac
    S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D
   agcgactcagactcggatagcgattcagattcagatagcgattcggattcagacagtgat
    S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D
   tcagattcagacagcgactcagactcggatagcgactcagactcagacagcgattcagac
    S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D
   tcagatagcgactcagactcggatagcgactcggattcagatagcgactcagactcagat
    S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D  S  D
   agtgactccgattcaagagttacaccaccaaaataatgaacagaaagcaccatcaaatcct
    S  D  S  D  S  R  V  T  P  P  N  N  E  Q  K  A  P  S  N  P
   aaaggtg-3'
    K  G
```

FIG. 2. SD repeat-encoding region of *clfB* in *S. aureus* strain NCTC 8325. Both DNA and amino acid sequences of the SD repeat-encoding region and immediate flanking regions are shown. The underlined nucleotides are the primers used to amplify the repeat region. Nucleotides/amino acids in bold type show the first and last complete repeats, which frame the repeat region portion (composed, in this example, of 36 direct 18-bp repeats) used for *clfB* typing.

A computerized search algorithm was designed (eGenomics, New York, N.Y.) to take the full amplicon sequence input and automatically find the SD repeat-encoding region and identify all individual repeats. The algorithm was the following. The start site for repeat typing in the full amplicon is the sequence TCN-GAY that is found as part of the first instance of GAT-TCN-GAY in the amplicon. Every 18-mer thereafter defines a repeat, unless the fifth codon (i.e., the 13th to 15th nucleotides of an individual repeat) equals TCN. If the 13th to 15th nucleotides are TCN, then slippage has occurred and the 12 nucleotides preceding the TCN are made into one repeat (indicated with an asterisk). The signal to end the repeat typing in the full-amplicon sequence is with the nucleotide right before the first TCN-GAT-TCA-AGA.

In a manner analogous to that for the *spa* types (22, 41), each *clfB* type (i.e., *clfB* allele) was given a numeric identifier and was defined by the composition, number, and order of repeats (termed a repeat profile) within the repeat region. The *clfB* types that had slippage events were given numeric identifiers and asterisks (if two slippage events occurred, then that *clfB* type was given two asterisks, and so forth). There was a total of 37 different *clfB* types found in this study (Table 2). Also, analogous to the procedure described previously for *spa* and *coa* typing (6, 22, 41, 42), *clfB* lineages were formed by grouping strains with similar *clfB* repeat profiles together (Table 2), as genetic relatedness is suggested by the presence of identical point mutations. Using a global sequence alignment program, similar *clfB* profile groupings were obtained (data not shown).

TABLE 1. Nucleotide sequences of 81 numbered *clfB* repeats[a]

Repeat no.	Sequence	Repeat no.	Sequence
1	TCGGATTCAGACAGTGAC	42	TCAGGTTCAGACAGTGAC
2	TCAGGCTCAGACAGCGAC	43	TCGGACTCAGATAGCAAC
3	TCAGGTTCAGATAGCGAC	44	TCGGACTCAGACAGTGAG
4	TCAGAATCAGATAGCGAT	45	TCGGATTCAGACAGCGAT
5	TCGGATTCAGACAGTGAT	46	TCGGACTCAGACAGTGAC
6	TCAGATTCAGACAGCGAC	47*	TCTGATTCAGAC
7	TCAGAATCAGATAGCGAC	48*	TCAGACAGTGAT
8	TCAGATTCAGATAGCGAT	49	TCGAATTCAGACAGCGAT
9	TCAGAATCAGATAGTGAG	50	TCAGACTCAGACAGTGAC
10	TCAGATTCAGACAGTGAC	51	TCTGATTCAGACAGCGAC
11	TCGGACTCAGACAGTGAT	52	TCAGATTCAGATAGTGAT
12	TCAGACTCAGATAGCGAT	53	TCGGATTCAGACAGCGAC
13	TCAGACTCAGATAGCGAC	54	TCGGATTCAGACAGTGAG
14	TCAGACTCAGACAGCGAC	55	TCAGATTCAGACAGTGAG
15	TCGGACTCAGACAGCGAT	56	TCTGATTCAGACAGTGAG
16	TCAGACTCGGATAGCGAT	57	TCTGATTCAGATAGTGAC
17	TCAGACTCGGATAGCGAC	58	TTAGACTCAGACAGTGAG
18	TCAGACTCAGACAGCGAT	59	TCAGATTCAGATAGTGAC
19	TCGGATTCAGATAGCGAC	60	TCAGACTCAGACATCGAT
20	TCAGACTCAGATAGTGAC	61	TCTGATTCAGACAGCGAT
21	TCAGGTTCAGATAGCGAT	62*	TCAGACTCAGAT
22	TCGGATTCAGATAGCGAT	63*	TCATACAGCGAT
23	TCAGATTCGGACAGCGAT	64	TCCGATTCAGATAGTGAT
24	TCAGATTCAGATAGCAAC	65	TCCGATTCAGACAGCGAT
25	TCAGAATCAGACAGTGAT	66*	TCGGATTCAGAC
26	TCAGAATCAGACAGCGAT	67*	TCAGACAGTGAC
27	TCAGAATCAGACAGTGAC	68	TCCGATTCAGATAGCGAT
28	TCAGATTCAGATAGTGAG	69	TCAGATTCAGACAGTGAT
29	TCAGACTCAGACAGTGAG	70	TCAGACTCGGACAGCGAC
30	TCAGATTCAGACAGCGAT	71	TCAGATTCATACAGCGAT
31	TCAGAATCAGACAGTGAG	72	TCAGGCTCAGATAGTGAC
32	TCGGACTCAGATAGCGAC	73*	TCAGACTCAGAC
33	TCAGAATCAGACAGCGAC	74*	TCAGATAGCGAT
34	TCGGATTCAGATAGTGAC	75	TTAGACTCAGATAGCGAT
35	TCAGGCTCAGACAGCGAT	76	TCAGACTCGGATAGCAAC
36	TCGGACTCAGATAGCGAT	77	TCAGACTCGGATAGCTAT
37	TCGGATTCAGACAACGAT	78	TCAAACTCAGACAGTGAG
38	TCAGATTCAGATAGCGAC	79	TCAAACTCAGATAGTGAC
39	TCAGACTCAGACAGTGAT	80	TCAGACCCAGACAGTGAG
40	TCGGACTCAGATAGTGAC	81*	TCAGATAGCGAC
41	TCGGATTCGGACAGTGAC		

[a] Asterisks indicate repeats 12 bp in length.

clfB types, similar to *coa* types described previously (22), could be organized into either nine or seven (for deeper phylogenetic classification) lineages entitled *clfB* lineages I and II, respectively (i.e., counting *clfB* lineages 3A, 3B, and 3C separately or as a single lineage). Of note is that the *clfB* types where slippage occurred did not all have a common *clfB* repeat profile; that is, they did not all fall into the same lineage (Table 2).

Diversity collection of 36 strains. (i) Discriminatory power. The 36 strains representing the breadth of genetic diversity in *S. aureus* that were studied previously using a whole-genome microarray (14) and *spa* and *coa* typing (22) were *clfB* typed (Fig. 3). There was a total of 17 *clfB* types and eight *clfB* lineages I. Simpson's index of diversity for *clfB* typing alone was 91.0%, compared with 97.3% for *spa* typing (22). However, when strains were genotyped with *spa* typing in combination with *clfB* typing, the index of diversity increased to 99.5%, with 34 of the 36 strains being assigned different genotypes. This closely approximated the DNA microarray, whose index of diversity was 100.0%. Thus, *clfB* on its own appears not to have exceedingly strong resolving power, but when combined with *spa* typing, it increases *spa* typing's resolving power greatly, as opposed to markers such as *coa* and PFGE that do not (22).

There were three PFGE types, three *spa* types, and one combined *spa*-PFGE type that were found among at least two isolates in this collection of 36 strains. In all cases *clfB* typing was able to differentiate these strains with genotypes that initially appeared identical in at least two groups. Of the 14 ET 234 TSS strains, which previously had been difficult to discriminate using only *spa* typing (22) and which resulted in an 83.5% index of diversity (resolving 9 genotypes), *spa* typing combined with *clfB* typing had a 96.7% index of diversity (resolving 12 genotypes). Of the 14 ET 234 strains, 6 were of *spa* type 33, all of which had the same A2 PFGE pattern. *clfB* typing resolved four different genotypes from three different *clfB* lineages among these six strains (Fig. 4A). These data, as well as the microarray findings, indicate that there are considerable genetic differences between these strains.

(ii) Evidence for recombination. In cross-classification analysis of all possible pairs of the 36 isolates, individual *clfB* types were 85% concordant with PFGE types and 89% concordant

TABLE 2. *clfB* types, lineages, and repeat profiles organized by *clfB* lineage

clfB type	*clfB* lineage	Repeat profile[a]
2	1	10-2-21-22-13-23-23-13-24-1-25-26-27-8-7-28-14-19-8-5-7-29-8-7-28-8-18-30-18-14-4-4-6-31-8-7-28-10-32-33-12-32
12	1	10-2-21-22-13-23-23-13-24-54-28-10-36-12-18-30-18-14-4-4-69-33-12-38-22-36-18-27-8-32
4	2	10-8-7-28-6-21-1-35-5-27-8-7-28-10-36-37-33-8-5-6-7-9-10-12-38-18-33-12-39-33-12-38-22-40-13-33-8-18-6-22-32
19	2	10-8-7-28-6-21-45-1-35-5-27-8-7-28-10-36-37-33-8-5-6-33-9-10-12-38-18-33-12-6-22-32
17*	3A	1-2-3-4-54-6-8-7-26-1-8-7-28-10-36-37-14-38-33-13-14-8-18-73*-74*-18-13-33-13-33-75-14-8-18-14-12-39-30-27-8-32
23	3B	1-2-3-4-5-6-4-7-8-8-4-5-6-7-9-10-11-12-12-14-8-15-6-16-6-17-34
25	3B	1-2-3-4-5-6-4-7-8-8-4-5-6-7-9-10-11-12-12-18-6-13-14-8-14-14-12-33-8-15-6-16-6-17-34
30	3B	1-2-3-4-5-6-4-7-8-8-4-5-6-7-9-10-11-12-12-18-6-14-13-14-8-14-14-12-14-8-15-6-16-6-17-34
9	3B	1-2-3-4-5-6-4-7-8-8-4-5-6-7-9-10-11-12-12-18-6-33-13-14-14-12-14-8-15-6-16-6-17-34
22	3B	1-2-3-4-5-6-4-7-8-8-4-5-6-7-9-10-11-12-12-18-6-33-13-14-4-14-14-12-14-8-15-6-16-6-17-34
24	3B	1-2-3-4-5-6-4-7-8-8-4-5-6-7-9-10-11-12-12-18-6-33-13-14-8-14-12-14-8-15-6-16-6-17-34
3	3B	1-2-3-4-5-6-4-7-8-8-4-5-6-7-9-10-11-12-12-18-6-33-13-14-8-14-14-12-14-8-15-6-16-6-17-34
21	3B	1-2-3-4-5-6-4-7-8-8-4-5-6-7-9-10-11-12-12-18-6-33-13-14-8-14-14-12-14-8-15-6-17-34
28	3B	1-2-3-4-5-6-4-7-8-8-4-5-6-7-9-10-11-12-12-18-6-6-13-14-8-14-14-12-14-8-15-6-16-6-17-34
26	3B	1-2-3-4-5-6-7-4-5-6-4-7-8-8-4-5-6-7-9-10-11-12-12-18-6-33-13-14-8-14-14-12-14-8-15-6-16-6-17-34
27	3B	1-2-3-4-5-6-7-4-5-6-4-7-8-8-4-5-6-7-9-10-11-12-12-18-6-33-13-33-8-14-14-12-14-8-15-6-16-6-17-34
18	3C	1-2-3-4-5-6-4-7-8-8-8-5-6-7-9-10-11-12-12-18-6-6-13-14-8-14-14-13-8-14-8-15-6-16-8-5-6-17-18-13-17-19-20
34	3C	1-2-3-4-5-6-4-7-8-8-8-5-6-7-9-10-11-12-12-18-6-6-14-8-14-14-13-8-14-8-11-6-17-18-13-17-19-20
7	3C	1-2-3-4-5-6-4-7-8-8-8-5-6-7-9-10-11-12-12-18-6-6-14-8-14-14-13-8-14-8-15-6-16-8-5-6-17-18-13-17-19-20
13	3C	1-2-3-4-5-6-4-7-8-8-8-5-6-7-9-10-11-12-12-18-6-6-14-8-14-14-13-8-14-8-15-6-16-8-5-6-17-18-13-70-19-20
29	3C	1-2-3-4-5-6-4-7-8-8-8-5-6-7-9-10-11-12-12-18-6-6-14-8-15-6-77-8-5-6-17-18-13-17-19-20
8	3C	1-2-3-4-5-6-4-7-8-8-8-5-6-7-9-10-11-12-12-6-6-13-14-8-14-14-13-8-14-8-15-6-16-8-5-6-17-17-19-20
1	3C	1-2-3-4-5-6-4-7-8-8-8-5-6-7-9-10-11-12-12-6-6-13-14-8-14-14-13-8-14-8-15-6-16-8-5-6-17-18-13-17-19-20
35	3C	1-2-3-4-5-6-7-9-10-11-12-12-18-6-6-13-14-8-14-14-13-8-14-8-15-6-16-8-5-6-17-18-13-17-19-20
36	4	34-2-21-34-20-78-10-40-8-30-44-8-46-40-40-79-30-22-30-26-1-19-45-46-68-31-8-27-80-30-18-29-55-30-32
5*	5	41-2-42-43-44-43-44-30-1-32-45-46-36-45-47*-48*-14-49-19-50-1-36-51-52-20-30-53-54-55-30-56-6-36-19-53-52-20-30-53-54-51-36-19-32
6*	5	41-2-42-43-44-43-44-30-1-32-45-46-36-45-47*-48*-14-49-19-50-1-36-51-52-20-30-53-54-55-30-56-6-36-19-53-52-20-30-53-54-6-36-19-32
10	5	41-2-42-43-46-36-14-49-45-57-30-58-30-36-53-6-30-15-19-59-30-6-52-20-30-54-6-36-22-29-52-60-19
14	6	41-2-42-44-30-29-20-30-54-30-29-20-30-61-29-20-71-61-51-54-52-53-8-53-20-50-8-44-52-53-8-53-20-50-8-19-64-1-68-32
11**	6	41-2-42-44-30-29-20-30-54-30-29-20-30-61-29-62*-63*-61-51-54-52-53-8-53-20-50-8-44-52-53-8-53-20-50-8-19-64-53-65-61-50-64-66*-67*-68-32
16*	6	41-72-50-68-44-52-53-8-53-20-50-50-8-19-64-53-65-45-50-64-66*-67*-68-32
33	7	53-21-45-14-21-53-31-8-7-28-10-36-18-14-38-18-38-28-59-36-18-6-13-28-10-13-8-18-30-14-14-30-18-38-30-38-6-8-18-14-17-34
15	7	53-21-45-14-21-53-31-8-7-28-10-36-18-14-38-18-7-28-59-36-18-14-13-28-10-12-18-30-14-14-4-18-38-18-38-14-4-18-6-17-34
37*	7	53-21-45-14-21-53-31-8-7-28-10-36-18-14-38-18-7-28-59-36-18-14-13-28-10-12-18-30-14-14-4-18-38-18-81*-14-4-18-6-17-34
32	7	53-21-45-14-21-53-31-8-7-28-10-36-18-14-38-18-7-28-59-36-18-14-13-28-10-13-18-30-14-14-8-18-38-18-38-14-8-18-6-17-34
20	7	53-21-45-14-21-53-31-8-7-28-10-36-18-14-38-18-7-28-59-36-18-14-13-28-10-13-8-18-30-14-14-4-18-38-18-38-18-6-76-34
31	7	53-21-45-14-21-53-31-8-7-28-10-36-18-14-38-18-7-28-59-36-18-14-13-28-10-13-8-18-30-14-14-8-18-38-18-38-14-8-18-6-17-34

[a] Repeat profiles are grouped by lineage. Double spaces group repeat profiles by the seven *clfB* lineages II. A combination of double and single spaces group repeat profiles by the nine *clfB* lineages I. *clfB* types that had a single slippage event were given a single asterisk, and if two slippage events occurred, then that *clfB* type was given two asterisks.

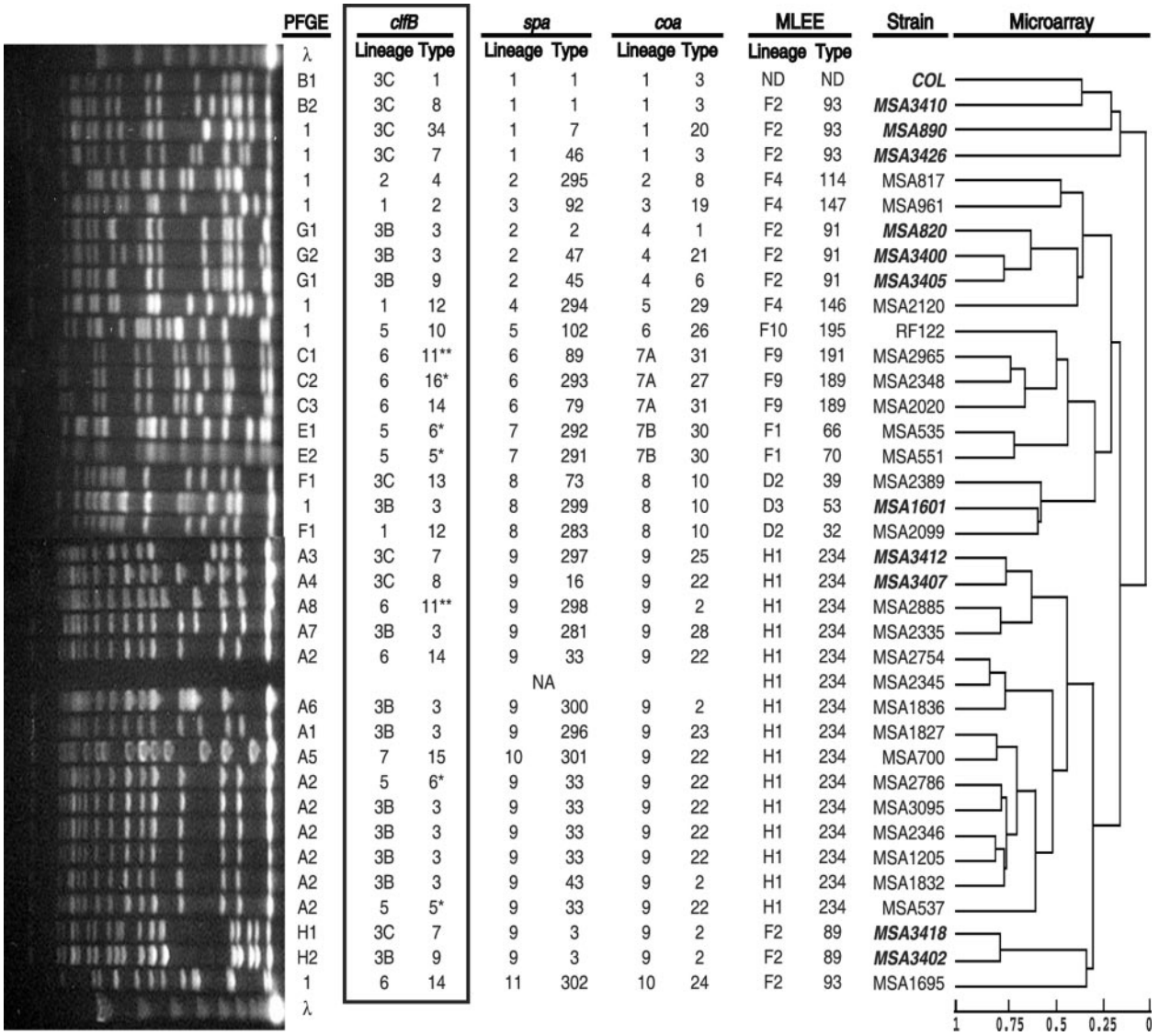

FIG. 3. Molecular characterization of diversity collection strains. From right to left: (i) a dendrogram (adapted from reference 14 with permission of the publisher) providing estimated relationships of 36 strains based on a whole-genome DNA microarray with a scale showing the Pearson correlation coefficient for each node (0, totally unrelated; 1, identical); (ii) the list of strains (MSA, Musser *S. aureus*) (bold italicized text indicates MRSA strains); (iii) MLEE lineage and ET for each strain; (iv) *coa*, *spa*, and *clfB* lineages and types (note that *clfB* data are outlined in a box); (v) PFGE patterns and types are shown (patterns not falling into any lineages were identified with numeral 1; λ, molecular weight standard of lambda DNA concatemers from New England BioLabs). NA, strain was not available. (Adapted from reference 22.)

with *spa* types. Further, *clfB* lineages I were 76% concordant with *spa* lineages. These results indicate that there was general agreement between *clfB* typing and typing with PFGE and *spa*. However, there is evidence of recombination at the *clfB* locus, as identical *clfB* types were found among isolates from completely different lineages as determined by microarray, *spa*, *coa*, PFGE, and MLEE typing (Fig. 3). For example, *clfB* type 14 is found in three strains from three different lineages (Fig. 3). In fact, no *clfB* type encountered more than once in this collection was confined to a single *spa* or *coa* lineage, whereas each *coa* type found in more than one strain, barring one exception, was associated with only a single *spa* lineage ($P = 0.002$), and each *spa* type found in more than one strain was associated with only a single *coa* lineage ($P = 0.005$). It also appeared that identical PFGE or *spa* genotypes tended to be split up by *clfB* types with different lineages in approximately 50% of cases (Fig. 3). This implies that recombination occurs and is important, as are slipped-strand mispairing and single nucleotide polymorphisms, etc., in the ability of *clfB* typing to discriminate among closely related strains.

***clfB* typing for discrimination among *spa* type 2 strains.** *spa* type 2 strains have become the most prevalent hospital-associated *S. aureus* clone in the United States (25). These strains are often endemic in hospital environments and well suited for testing the ability of *clfB* typing to detect genetic variation and distinguish among seemingly related strains. PFGE typing of 12 *spa* type 2 isolates obtained from nine patients over a period of 2 weeks from a New York City hospital resulted in an index

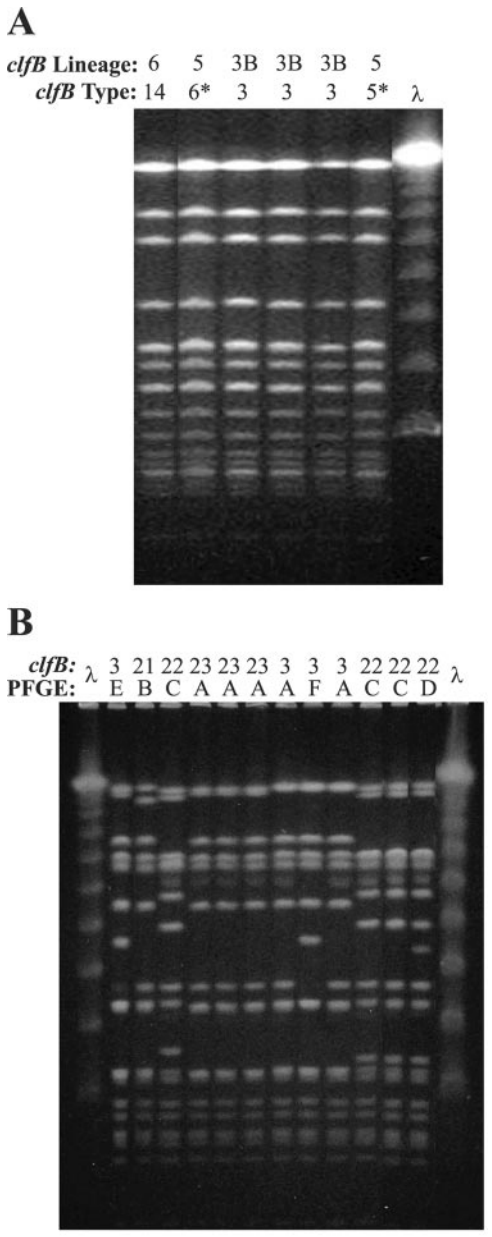

FIG. 4. (A) Six *spa* type 33 strains with the same A2 PFGE pattern from the diversity collection, which were differentiated by *clfB* typing. Individual *clfB* types and lineages are shown. Note that all strains were also differentiated by the whole-genome microarray. Asterisks indicate *clfB* types that had slippage events. (B) Twelve *spa* type 2 strains differentiated by PFGE and *clfB* typing. Individual *clfB* types are shown by their numerical identifiers, and individual PFGE types are shown by letters (note that each unique PFGE pattern was labeled here with a separate letter). The results of PFGE and *clfB* typing were in 79% agreement. λ, molecular weight standard.

of diversity of 80.3% (six genotypes), and *clfB* typing resulted in an index of diversity of 77.3% (four genotypes). The results of *clfB* and PFGE typing (Fig. 4B) were in 79% direct concordance in cross-classification analysis; that is, the majority of isolate pairs when considered either identical or different by one genotyping technique were given the same designation by the other technique.

Not including isolates with identical *clfB* or PFGE types from the same patient, there was a total of 29 *spa* type 2 strains among the 116 strains used in this study. Remarkably, Simpson's index of diversity for *clfB* typing these 29 strains was 92.9% (10 *clfB* types).

Reproducibility and stability of *clfB* typing. The reproducibility of the *clfB* sequencing results was verified, as many PCR amplicons were sequenced multiple times and identical sequences were always obtained. However, to formally test the reproducibility of the method, two strains had their DNA isolated twice each, and two additional strains had their DNA isolated three times each; all strains subsequently underwent full processing, and the resulting *clfB* type of a strain always matched that of the other member of the pair/triplicate.

This repeat region has in vitro stability, as an isolate passed extensively in the laboratory for 6 weeks (44) retained the same *clfB* type. Interestingly, it was noted that this isolate did undergo some nucleotide polymorphisms elsewhere in its genome during its passage (44), yet this went undetected when *clfB* typing was used, which apparently is thus not prone to being excessively variable. Another example of *clfB* typing not being excessively variable and maintaining its ability to recognize that two strains are indeed from the same source comes from work with two strains obtained during our laboratory's hemodialysis study from a patient over a 3-month period. The PFGE patterns were identical except for a two-band difference that was directly due to the loss of the methicillin-resistance determining element, staphylococcal cassette chromosome *mec* (data not shown). The *clfB* types for these two strains were identical, whereas reliance only on the PFGE data could have led to the false conclusion that the strains, which had different PFGE patterns, could possibly be derived from two different sources, a problem that has been noted for PFGE in previous work with *S. aureus* and other species (12, 17).

Three carriage isolates, obtained over a 21-month period in the same aforementioned hemodialysis study, from each of four persistent carrier patients (i.e., patients who consistently harbored *S. aureus* strains that maintained the same PFGE pattern over time) had identical *clfB* types, indicating that this region also has high in vivo stability. Furthermore, *clfB* types remained the same for the group of seven outbreak I MRSA strains obtained from different patients during a well-characterized outbreak (41, 47). Of the four outbreak II MSSA strains from another well-characterized outbreak (41, 47), three had the same *clfB* types and one was different. This one strain with a different *clfB* type was also given an unrelated genotype by 8 of 13 (62%) additional genotyping techniques, implying that this strain was probably erroneously included as part of the outbreak (41, 43). These data also indicate that *clfB* has interpatient transmission in vivo stability, allowing it to be useful for outbreak investigations in areas where clones are endemic, which require high-resolution techniques for determining whether an outbreak has occurred.

Evolutionary pressure on *clfB* repeats. The 18-bp *clfB* repeat has six codons, TCN-GAY-TCN-GAY-AGY-GAY, where N is A, C, G, or T and Y is C or T, encoding six amino acids, S-D-S-D-S-D, respectively. Among all of the 18-bp repeats found in this study, there was never a serine in the first or third amino acid positions that was encoded by AGY, and there was never a serine in the fifth amino acid position encoded by TCN

($P < 0.0000001$). This finding demonstrates a strong codon usage bias (i.e., preferential use of certain codons). However, because the instances of codon usage bias are within the same gene and physically adjacent to one another within the same repeat, this is strong evidence for selection occurring at synonymous, or silent, sites (i.e., selection pressure for the use of certain codons at certain positions, even though the codons not being used would not alter the amino acid sequence).

While moving from selection analysis at the codon usage level to selection analysis at the nucleotide level for evolutionary pressure to alter amino acids, it was found that the 81 *clfB* repeats had a dS/dN value of 8.0 (a ratio of <1 indicates positive selection, a ratio of 1 indicates no selection pressure [i.e., neutral evolution], and a ratio of >1 indicates purifying selection). The dS value, i.e., the number of synonymous substitutions per potential synonymous site, was 0.80 (standard error, 0.16) and the dN value, i.e., the number of nonsynonymous substitutions per potential nonsynonymous site, was 0.10 (standard error, 0.05). A *Z* test for detecting purifying selection on the *clfB* repeats was highly significant ($P = 0.00004$). Therefore, the *clfB* repeat region is under strong purifying selection, indicating that the SD amino acids are under selection pressure not to change.

Because of this purifying selection against amino acid alterations and because *S. aureus* is considered a highly clonal species with very little independent assortment of genes (13, 22), it was surprising to find recombination at the *clfB* locus, as described above. This recombination suggests *clfB* may be under positive selection pressure at the macrolevel of the full repeat region and not at the nucleotide level to change amino acids individually (as evidenced by the high dS/dN value). It is also interesting that *clfB* type 3/lineage 3B was the most common *clfB* type found in this study—it was found throughout many different *spa* and *coa*, etc., lineages (Fig. 3). Additionally, all 50 strains used in this study that were of *spa* type 2 or related to *spa* type 2 had *clfB* types from lineage 3B. Only strains from this group seemed to have had no outside *clfB* type recombine into their genetic backgrounds, and this may suggest that *clfB* types across the *S. aureus* species are being driven towards the *clfB* types of lineage 3B—the sole lineage that characterizes the prevalent *S. aureus* strain in the United States (25).

DISCUSSION

As *S. aureus* strains such as *spa* type 2 become more widespread (25), it will become increasingly difficult to distinguish among them. Our laboratory's genotyping experience has shown that hospitals where *S. aureus* strains are endemic have difficulty discriminating among these strains when an outbreak is suspected, as all isolates are usually assigned the same genotype. However, *S. aureus* is constantly developing genetic variation that can be harnessed for investigational purposes, and *clfB* appears to be a marker that can detect recent genetic variation that other markers cannot. This study demonstrated that DNA sequencing of the SD repeat-encoding region of *clfB* subdivides the highly prevalent *spa* type 2 group and also other identical *spa* and PFGE clusters, which were discriminated by a DNA microarray. If two strains with the same *spa* or MLST genotype are suspected of being part of an outbreak and have different *clfB* types, they probably are not likely to be from the same outbreak, in light of the stability of the *clfB* locus. However, if they have the same *clfB* type and if such a conclusion is warranted by the supporting infection control and epidemiological data, they probably are likely to be from the same outbreak. Furthermore, because *clfB* typing was useful in discriminating among the collection of strains representing the breadth of diversity in *S. aureus*, *clfB* typing may also be used by bacterial population geneticists for interrogating large clonal groups of strains for previously undetected or newly emerging strain subclusters. Markers with similar utility have been successfully employed for other bacterial species, such as group A *Streptococcus* species (45).

The combined mutation rate (discriminatory power) of *spa* and *clfB* typing is virtually the same as the overall chromosome mutation rate detected by PFGE methodology and is occasionally even greater. However, *clfB* typing may technically be more difficult than *spa* typing, even though sequencing technology is improving, because of the large average *clfB* repeat region size (677 bp) compared to that of *spa* (41). Yet, it is this increased repeat region size that allows for more genetic variation to accumulate, along with increased amounts of slipped-strand mispairing due to the shorter 18- and 12-bp individual repeats, which contributes to the discriminatory ability of *clfB* typing. Nevertheless, *spa* typing should be used as the initial genetic marker for typing strains, and if further discrimination is necessary (as here with the ET234 TSS-associated strains from the diversity collection that were differentiated by the microarray but not by *spa* or PFGE typing), *clfB* typing can be used. The combination of *spa* and *clfB* typing may then serve as a DNA sequence-based alternative to image-based genotyping techniques, such as PFGE, which are known to be difficult to standardize, analyze, and database (26). If verification of strain lineage were sought, MLST or DNA microarray analysis would then be appropriate.

Close examination of the *clfB* repeat region revealed that it is under three different types of evolutionary pressure. First, there is a high level of statistically significant selection at silent sites, as evidenced by the serines in the first and third amino acid positions of the six-amino-acid-long repeat encoded only by the TCN codon, whereas the serine found in the fifth amino acid position is encoded only by the AGY codon. Serine is the only amino acid with fourfold (TCN) and twofold (AGY) degenerate codons that cannot interconvert with a single mutation (3, 8). The chance of two mutations happening in the same codon in one generation is $\sim 10^{-18}$ and thus highly unlikely (Hiroshi Akashi [Institute of Molecular Evolutionary Genetics, Pennsylvania State University], personal communication). This selection at silent sites is not due to codon usage bias resulting from relative tRNA abundance (20), gene expression rates and control (16), or translational speed or accuracy (2, 4), as both TCN and AGY codons are repeatedly selected for at every 18 bp throughout the *clfB* repeat region. More likely, the selection is due to preferred mRNA or protein structure, as seen in other cases (40, 54, 56). This can be studied further by substituting codons for one another and assessing the resulting mRNA and/or protein structure stability. Another possible explanation for this selection at silent sites is that by increasing the repeat length, the rate of slipped-strand mispairing during replication is decreased (22, 51), ensuring less variation in

overall repeat region lengths. Repeat region length requirements have been shown to be important in the functioning of another SD repeat-containing protein in *S. aureus* (18). The use of both TCN and AGY codons creates an 18-bp repeat (which encodes SDSDSD) as opposed to the 6-bp repeat (which encodes SD) that would be formed by the use of only one of the codons. Therefore, if serine is highly conserved, as it is in the *clfB* repeat region, the placement of the TCN and AGY codons together in the same repeat increases the repeat's length and the overall repeat region stability.

The second type of evolutionary pressure on the *clfB* repeat region is at the amino acid level, which undergoes purifying selection to maintain its SD amino acid composition. The third type of evolutionary pressure is at the macrolevel of the entire *clfB* repeat region, which appears to be recombinogenic, while other parts of the *S. aureus* genome are not (13, 22). This property, along with the polymorphic nature of the repeats' nucleotide compositions and organizations, contributes to the ability of *clfB* typing to resolve differences within clonal groups of *S. aureus*. The possibility that *clfB* may be recombinogenic has been discussed elsewhere (13) as part of a hitchhiking effect explanation for why a nearby (in terms of chromosomal location to *clfB*) neutral housekeeping gene used in MLST (*arcC*) appears to have undergone recombinational replacements. For this reason, *clfB* typing must be used in combination with *spa* typing. Although *clfB* types from closely related *spa* type clusters were usually similar, identical *clfB* types were found in distant lineages of *S. aureus*, indicating that this region may be recombining and under selection pressure, possibly due to the role of *clfB* in binding to host keratin and fibrinogen. There may also be positive selection towards an optimal *clfB* type, possibly the *clfB* type within the highly successful *spa* type 2-related strains, because this *clfB* type is found in many other *spa* lineages. Additionally, we are presently investigating potentially different keratin-binding affinities among *clfB* types that may help provide insight into the process of colonization with *S. aureus*. In summary, the *clfB* repeat region has in vitro stability, long-term in vivo stability, and interpatient transmission stability. The data indicate that when combined with *spa* typing, *clfB* typing is a highly stable marker of microvariation within related strains, has discriminatory power comparable to those of PFGE and whole-genome microarray analysis, and is useful for analyzing collections of *S. aureus* isolates in both long-term population-based and local epidemiologic studies.

ACKNOWLEDGMENT

We are grateful to William Eisner for assistance with performance of PFGE.

REFERENCES

1. **Aires de Sousa, M., C. Bartzavali, I. Spiliopoulou, I. S. Sanches, M. I. Crisostomo, and H. de Lencastre.** 2003. Two international methicillin-resistant *Staphylococcus aureus* clones endemic in a university hospital in Patras, Greece. J. Clin. Microbiol. **41:**2027–2032.
2. **Akashi, H.** 1994. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. Genetics **136:**927–935.
3. **Brenner, S.** 1988. The molecular evolution of genes and proteins: a tale of two serines. Nature **334:**528–530.
4. **Bulmer, M.** 1991. The selection-mutation-drift theory of synonymous codon usage. Genetics **129:**897–907.
5. **Cespedes, C., B. Saïd-Salim, M. Miller, S.-H. Lo, B. Kreiswirth, R. J. Gordon, P. Vavagiakis, R. S. Klein, and F. D. Lowy.** 2005. The clonality of *Staphylococcus aureus* nasal carriage. J. Infect. Dis. **191:**444–452.
6. **Crisostomo, M. I., H. Westh, A. Tomasz, M. Chung, D. C. Oliveira, and H. de Lencastre.** 2001. The evolution of methicillin resistance in *Staphylococcus aureus*: similarity of genetic backgrounds in historically early methicillin-susceptible and -resistant isolates and contemporary epidemic clones. Proc. Natl. Acad. Sci. USA **98:**9865–9870.
7. **de Lencastre, H., M. Chung, and H. Westh.** 2000. Archaic strains of methicillin-resistant *Staphylococcus aureus*: molecular and microbiological properties of isolates from the 1960s in Denmark. Microb. Drug Resist. **6:**1–10.
8. **Diaz-Lazcoz, Y., A. Henaut, P. Vigier, and J. L. Risler.** 1995. Differential codon usage for conserved amino acids: evidence that the serine codons TCN were primordial. J. Mol. Biol. **250:**123–127.
9. **Enright, M. C., N. P. Day, C. E. Davies, S. J. Peacock, and B. G. Spratt.** 2000. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. J. Clin. Microbiol. **38:**1008–1015.
10. **Enright, M. C., and B. G. Spratt.** 1999. Multilocus sequence typing. Trends Microbiol. **7:**482–487.
11. **Entenza, J. M., T. J. Foster, D. Ni Eidhin, P. Vaudaux, P. Francioli, and P. Moreillon.** 2000. Contribution of clumping factor B to pathogenesis of experimental endocarditis due to *Staphylococcus aureus*. Infect. Immun. **68:**5443–5446.
12. **Feavers, I. M., S. J. Gray, R. Urwin, J. E. Russell, J. A. Bygraves, E. B. Kaczmarski, and M. C. Maiden.** 1999. Multilocus sequence typing and antigen gene sequencing in the investigation of a meningococcal disease outbreak. J. Clin. Microbiol. **37:**3883–3887.
13. **Feil, E. J., J. E. Cooper, H. Grundmann, D. A. Robinson, M. C. Enright, T. Berendt, S. J. Peacock, J. M. Smith, M. Murphy, B. G. Spratt, C. E. Moore, and N. P. Day.** 2003. How clonal is *Staphylococcus aureus*? J. Bacteriol. **185:**3307–3316.
14. **Fitzgerald, J. R., D. E. Sturdevant, S. M. Mackie, S. R. Gill, and J. M. Musser.** 2001. Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. Proc. Natl. Acad. Sci. USA **98:**8821–8826.
15. **Foster, T. J., and M. Hook.** 1998. Surface protein adhesins of *Staphylococcus aureus*. Trends Microbiol. **6:**484–488.
16. **Grantham, R., C. Gautier, M. Gouy, M. Jacobzone, and R. Mercier.** 1981. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. **9:**r43–r74.
17. **Harmsen, D., H. Claus, W. Witte, J. Rothganger, D. Turnwald, and U. Vogel.** 2003. Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for *spa* repeat determination and database management. J. Clin. Microbiol. **41:**5442–5448.
18. **Hartford, O., P. Francois, P. Vaudaux, and T. J. Foster.** 1997. The dipeptide repeat region of the fibrinogen-binding protein (clumping factor) is required for functional expression of the fibrinogen-binding domain on the *Staphylococcus aureus* cell surface. Mol. Microbiol. **25:**1065–1076.
19. **Hunter, P. R., and M. A. Gaston.** 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol. **26:**2465–2466.
20. **Ikemura, T.** 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J. Mol. Biol. **151:**389–409.
21. **Josefsson, E., K. W. McCrea, D. Ni Eidhin, D. O'Connell, J. Cox, M. Hook, and T. J. Foster.** 1998. Three new members of the serine-aspartate repeat protein multigene family of *Staphylococcus aureus*. Microbiology **144:**3387–3395.
22. **Koreen, L., S. V. Ramaswamy, E. A. Graviss, S. Naidich, J. M. Musser, and B. N. Kreiswirth.** 2004. *spa* typing method for discriminating among *Staphylococcus aureus* isolates: implications for use of a single marker to detect genetic micro- and macrovariation. J. Clin. Microbiol. **42:**792–799.
23. **Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei.** 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17:**1244–1245.
24. **Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. USA **95:**3140–3145.
25. **McDougal, L. K., C. D. Steward, G. E. Killgore, J. M. Chaitram, S. K. McAllister, and F. C. Tenover.** 2003. Pulsed-field gel electrophoresis typing of oxacillin-resistant *Staphylococcus aureus* isolates from the United States: establishing a national database. J. Clin. Microbiol. **41:**5113–5120.
26. **Murchan, S., M. E. Kaufmann, A. Deplano, R. de Ryck, M. Struelens, C. E. Zinn, V. Fussing, S. Salmenlinna, J. Vuopio-Varkila, N. El Solh, C. Cuny, W. Witte, P. T. Tassios, N. Legakis, W. van Leeuwen, A. van Belkum, A. Vindel, I. Laconcha, J. Garaizar, S. Haeggman, B. Olsson-Liljequist, U. Ransjo, G. Coombes, and B. Cookson.** 2003. Harmonization of pulsed-field gel electrophoresis protocols for epidemiological typing of strains of methicillin-resistant *Staphylococcus aureus*: a single approach developed by consensus in 10 European laboratories and its application for tracing the spread of related strains. J. Clin. Microbiol. **41:**1574–1585.
27. **Nallapareddy, S. R., R. W. Duh, K. V. Singh, and B. E. Murray.** 2002. Molecular typing of selected *Enterococcus faecalis* isolates: pilot study using

multilocus sequence typing and pulsed-field gel electrophoresis. J. Clin. Microbiol. **40:**868–876.
28. **National Nosocomial Infections Surveillance System.** 1999. National Nosocomial Infections Surveillance (NNIS) System report, data summary from January 1990-May 1999, issued June 1999. Am. J. Infect. Control **27:**520–532.
29. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:**418–426.
30. **Ni Eidhin, D., S. Perkins, P. Francois, P. Vaudaux, M. Hook, and T. J. Foster.** 1998. Clumping factor B (ClfB), a new surface-located fibrinogen-binding adhesin of *Staphylococcus aureus*. Mol. Microbiol. **30:**245–257.
31. **O'Brien, F. G., T. T. Lim, F. N. Chong, G. W. Coombs, M. C. Enright, D. A. Robinson, A. Monk, B. Said-Salim, B. N. Kreiswirth, and W. B. Grubb.** 2004. Diversity among community isolates of methicillin-resistant *Staphylococcus aureus* in Australia. J. Clin. Microbiol. **42:**3185–3190.
32. **O'Brien, L. M., E. J. Walsh, R. C. Massey, S. J. Peacock, and T. J. Foster.** 2002. Staphylococcus aureus clumping factor B (ClfB) promotes adherence to human type I cytokeratin 10: implications for nasal colonization. Cell. Microbiol. **4:**759–770.
33. **Oliveira, D. C., A. Tomasz, and H. de Lencastre.** 2001. The evolution of pandemic clones of methicillin-resistant *Staphylococcus aureus*: identification of two ancestral genetic backgrounds and the associated *mec* elements. Microb. Drug Resist. **7:**349–361.
34. **Oliveira, D. C., A. Tomasz, and H. de Lencastre.** 2002. Secrets of success of a human pathogen: molecular evolution of pandemic clones of meticillin-resistant *Staphylococcus aureus*. Lancet Infect. Dis. **2:**180–189.
35. **Peacock, S. J., C. E. Moore, A. Justice, M. Kantzanou, L. Story, K. Mackie, G. O'Neill, and N. P. Day.** 2002. Virulent combinations of adhesin and toxin genes in natural populations of *Staphylococcus aureus*. Infect. Immun. **70**: 4987–4996.
36. **Read, T. D., S. L. Salzberg, M. Pop, M. Shumway, L. Umayam, L. Jiang, E. Holtzapple, J. D. Busch, K. L. Smith, J. M. Schupp, D. Solomon, P. Keim, and C. M. Fraser.** 2002. Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. Science **296:**2028–2033.
37. **Robinson, D. A., and M. C. Enright.** 2003. Evolutionary models of the emergence of methicillin-resistant *Staphylococcus aureus*. Antimicrob. Agents Chemother. **47:**3926–3934.
38. **Robinson, D. A., S. K. Hollingshead, J. M. Musser, A. J. Parkinson, D. E. Briles, and M. J. Crain.** 1998. The IS1167 insertion sequence is a phylogenetically informative marker among isolates of serotype 6B *Streptococcus pneumoniae*. J. Mol. Evol. **47:**222–229.
39. **Robles, J. C., L. Koreen, S. Park, and D. S. Perlin.** 2004. Multilocus sequence typing is a reliable alternative method to DNA fingerprinting for discriminating among strains of *Candida albicans*. J. Clin. Microbiol. **42:**2480–2488.
40. **Shen, L. X., J. P. Basilion, and V. P. Stanton, Jr.** 1999. Single-nucleotide polymorphisms can cause different structural folds of mRNA. Proc. Natl. Acad. Sci. USA **96**:7871–7876.
41. **Shopsin, B., M. Gomez, S. O. Montgomery, D. H. Smith, M. Waddington, D. E. Dodge, D. A. Bost, M. Riehman, S. Naidich, and B. N. Kreiswirth.** 1999. Evaluation of protein A gene polymorphic region DNA sequencing for typing of *Staphylococcus aureus* strains. J. Clin. Microbiol. **37:**3556–3563.
42. **Shopsin, B., M. Gomez, M. Waddington, M. Riehman, and B. N. Kreiswirth.** 2000. Use of coagulase gene (*coa*) repeat region nucleotide sequences for typing of methicillin-resistant *Staphylococcus aureus* strains. J. Clin. Microbiol. **38:**3453–3456.
43. **Smeltzer, M. S., A. F. Gillaspy, F. L. Pratt, and M. D. Thames.** 1997. Comparative evaluation of use of *cna*, *fnbA*, *fnbB*, and *hlb* for genomic fingerprinting in the epidemiological typing of *Staphylococcus aureus*. J. Clin. Microbiol. **35:**2444–2449.
44. **Somerville, G. A., S. B. Beres, J. R. Fitzgerald, F. R. DeLeo, R. L. Cole, J. S. Hoff, and J. M. Musser.** 2002. In vitro serial passage of *Staphylococcus aureus*: changes in physiology, virulence factor production, and *agr* nucleotide sequence. J. Bacteriol. **184:**1430–1437.
45. **Stockbauer, K. E., D. Grigsby, X. Pan, Y. X. Fu, L. M. Mejia, A. Cravioto, and J. M. Musser.** 1998. Hypervariability generated by natural selection in an extracellular complement-inhibiting protein of serotype M1 strains of group A *Streptococcus*. Proc. Natl. Acad. Sci. USA **95:**3128–3133.
46. **Tang, Y. W., M. G. Waddington, D. H. Smith, J. M. Manahan, P. C. Kohner, L. M. Highsmith, H. Li, F. R. Cockerill III, R. L. Thompson, S. O. Montgomery, and D. H. Persing.** 2000. Comparison of protein A gene sequencing with pulsed-field gel electrophoresis and epidemiologic data for molecular typing of methicillin-resistant *Staphylococcus aureus*. J. Clin. Microbiol. **38:** 1347–1351.
47. **Tenover, F. C., R. Arbeit, G. Archer, J. Biddle, S. Byrne, R. Goering, G. Hancock, G. A. Hebert, B. Hill, R. Hollis, et al.** 1994. Comparison of traditional and molecular methods of typing isolates of *Staphylococcus aureus*. J. Clin. Microbiol. **32**:407–415.
48. **Tenover, F. C., R. D. Arbeit, R. V. Goering, P. A. Mickelsen, B. E. Murray, D. H. Persing, and B. Swaminathan.** 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J. Clin. Microbiol. **33:**2233–2239.
49. **Tristan, A., L. Ying, M. Bes, J. Etienne, F. Vandenesch, and G. Lina.** 2003. Use of multiplex PCR to identify *Staphylococcus aureus* adhesins involved in human hematogenous infections. J. Clin. Microbiol. **41:**4465–4467.
50. **van Belkum, A., J. Kluytmans, W. van Leeuwen, R. Bax, W. Quint, E. Peters, A. Fluit, C. Vandenbroucke-Grauls, A. van den Brule, H. Koeleman, W. Melchers, J. Meis, A. Elaichouni, M. Vaneechoutte, F. Moouens, N. Maes, M. Struelens, F. Tenover, and H. Verbrugh.** 1995. Multicenter evaluation of arbitrarily primed PCR for typing of *Staphylococcus aureus* strains. J. Clin. Microbiol. **33:**1537–1547.
51. **van Belkum, A., S. Scherer, L. van Alphen, and H. Verbrugh.** 1998. Short-sequence DNA repeats in prokaryotic genomes. Microbiol. Mol. Biol. Rev. **62:**275–293.
52. **van Belkum, A., W. van Leeuwen, M. E. Kaufmann, B. Cookson, F. Forey, J. Etienne, R. Goering, F. Tenover, C. Steward, F. O'Brien, W. Grubb, P. Tassios, N. Legakis, A. Morvan, N. El Solh, R. de Ryck, M. Struelens, S. Salmenlinna, J. Vuopio-Varkila, M. Kooistra, A. Talens, W. Witte, and H. Verbrugh.** 1998. Assessment of resolution and intercenter reproducibility of results of genotyping *Staphylococcus aureus* by pulsed-field gel electrophoresis of *Sma*I macrorestriction fragments: a multicenter study. J. Clin. Microbiol. **36:**1653–1659.
53. **Walsh, E. J., L. M. O'Brien, X. Liang, M. Hook, and T. J. Foster.** 2004. Clumping factor B, a fibrinogen-binding MSCRAMM of *Staphylococcus aureus*, also binds to the tail region of type I cytokeratin 10. J. Biol. Chem. **279:**50691–50699.
54. **Wang, L., and S. R. Wessler.** 2001. Role of mRNA secondary structure in translational repression of the maize transcriptional activator Lc(1,2). Plant Physiol. **125:**1380–1387.
55. **Weigel, L. M., D. B. Clewell, S. R. Gill, N. C. Clark, L. K. McDougal, S. E. Flannagan, J. F. Kolonay, J. Shetty, G. E. Killgore, and F. C. Tenover.** 2003. Genetic analysis of a high-level vancomycin-resistant isolate of *Staphylococcus aureus*. Science **302:**1569–1571.
56. **Xie, T., D. Ding, X. Tao, and D. Dafu.** 1998. The relationship between synonymous codon usage and protein structure. FEBS Lett. **434:**93–96.