Research article

Natural population dynamics and expansion of pathogenic clones of *Staphylococcus aureus*

Damian C. Melles, Raymond F.J. Gorkink, Hélène A.M. Boelens, Susan V. Snijders, Justine K. Peeters, Michael J. Moorhouse, Peter J. van der Spek, Willem B. van Leeuwen, Guus Simons, Henri A. Verbrugh, and Alex van Belkum

The population structure of *Staphylococcus aureus* carried by healthy humans was determined using a large strain collection of nonclinical origin (*n* = 829). High-throughput amplified fragment length polymorphism (AFLP) analysis revealed 3 major and 2 minor genetic clusters of *S. aureus*, which were corroborated by multilocus sequence typing. Major AFLP cluster I comprised 44.4% of the carriage isolates and showed additional heterogeneity whereas major AFLP groups II and III presented 2 homogeneous clusters, including 47.3% of all carriage isolates. Coanalysis of invasive *S. aureus* strains and epidemic methicillin-resistant *S. aureus* (MRSA) revealed that all major clusters contained invasive and multiresistant isolates. However, clusters and subclusters with overrepresentation of invasive isolates were also identified. Bacteremia in elderly adults, for instance, was caused by a IVa cluster–derived strain significantly more often than by strains from other AFLP clusters. Furthermore, expansion of multiresistant clones or clones associated with skin disease (impetigo) was detected, which suggests that epidemic potential is present in pathogenic strains of *S. aureus*. In addition, the virulence gene encoding Panton-Valentine leukocidin was significantly enriched in *S. aureus* strains causing abscesses and arthritis in comparison with the carriage group. We provide evidence that essentially any *S. aureus* genotype carried by humans can transform into a life-threatening human pathogen but that certain clones are more virulent than others.

Introduction

*Staphylococcus aureus* is a major human pathogen capable of causing a wide range of infections. Over the last 25 years, the incidence of both community-acquired and hospital-acquired *S. aureus* infections has increased (1, 2). It is assumed that most of the infections derive from nasal carriage (3–7) with the nose acting as the primary ecological reservoir of *S. aureus* in humans (8).

In order to perform detailed comparative genomics and population genetics for a bacterial species, the limited availability of adequate strain collections needs to be resolved. Although the general population structure of *S. aureus* has been studied previously (9–11), these studies were biased because of the use of mostly clinical isolates and collections of nosocomial-acquired methicillin-resistant *S. aureus* (MRSA). The population structure of naturally occurring methicillin-susceptible *S. aureus* isolated from the nose of persons living in the community has never been described in detail.

There is controversy over whether all strains of *S. aureus* have equal disease-invoking potential or whether invasive disease is associated with particularly virulent genotypes. Day et al. had to rescind their conclusion that clonal *S. aureus* strains that are most successful in colonizing humans also show increased virulence potential (9, 12). They finally concluded that there is no significant difference in population structure between *S. aureus* carriage and disease-associated strains. When Peacock et al. focused on the presence of putative virulence determinants rather than overall genome polymorphism, it was concluded that 7 of these determinants were significantly more present in invasive isolates of *S. aureus* (13). Whether this increased virulence gene density may be specific for certain phylogenetic branches or lineages of *S. aureus*, however, remains unclear.

Most recent studies have assessed the population structure of *S. aureus* using multilocus sequence typing (MLST) (9, 14, 15). This molecular typing method characterizes bacterial isolates on the basis of the sequence of internal fragments of 7 housekeeping genes, representing the stable “core” of the bacterial genome. For each gene fragment, the different sequences are translated into distinct alleles, and each isolate is defined by the combination of alleles of the 7 housekeeping loci (the allelic profile or sequence type [ST]) (14). In contrast, whole genome typing methods, including amplified fragment length polymorphism (AFLP) (16, 17), document the contribution of accessory genetic elements as well as genome-core polymorphisms. AFLP is a method that scans for polymorphism in actual restriction sites but also among the nucleotides bordering these sites. As such, it documents nucleotide sequence variation, insertions, and deletions across genomes (16). This may be a more comprehensive approach for coming to a full understanding of staphylococcal genome diversity and evolution.

Nonstandard abbreviations used: AFLP, amplified fragment length polymorphism; CC, clonal complex; ETA, exfoliative toxin A; hAFLP, high-throughput AFLP; MLST, multilocus sequence typing; MRSA, methicillin-resistant *Staphylococcus aureus*; PC, principal component; PCA, principal component analysis; PVL, Panton-Valentine leukocidin; ST, sequence type.

Conflict of interest: The authors have declared that no conflict of interest exists.

Citation for this article: J Clin Invest. 114:1732–1740 (2004). doi:10.1172/JCI200423083.
We collected over 2,000 nonclinical \textit{S. aureus} isolates from various groups of healthy individuals in the Dutch community over an extended period of time, creating a unique population-based strain collection. To assess differences in the virulence potential of various strains of \textit{S. aureus}, insight into the natural (nonclinical) population structure is essential. We here present whole-genome scanning by high-throughput AFLP (ht-AFLP) of a random selection of these \textit{S. aureus} strains ($n = 829$). Strains were derived from the anterior nares of healthy children (1–18 years) and healthy elderly adults (>55 years) from the Rotterdam area (The Netherlands) (18, 19). To investigate genomic differences between these carriage strains and invasive isolates, contemporary isolates from blood, deep-seated, and soft-tissue infections from the same geographical area were included ($n = 164$). Furthermore, a collection of international epidemic MRSA strains ($n = 21$) (20) and \textit{S. aureus} isolates from Rotterdam children with nonbul- lous impetigo ($n = 40$) were included (21). The prevalence of the clinically relevant \textit{mecA} and Panton-Valentine leukocidin (PVL) genes was determined for all strains as well.

\textbf{Results}

\textit{Genetic diversity of \textit{S. aureus}.} Using the set of 1,056 \textit{S. aureus} strains, a total of 155,232 AFLP fragments were generated, covering 147 different marker fragments per strain. These outcomes are visualized in Figure 1. The dendrogram on the left (y axis) reveals bacterial clustering. The bar on the right of this figure delineates the presence of 3 major (I, II, III) and 2 minor (IVa, IVb) branches, as identified by principal component analysis (PCA) (Figure 2, A and B). Unsupervised cluster analysis of the 1,056 strains (Figure 3) clearly demonstrates that the AFLP data represent 2 distinct, homogenous major clusters (II and III) and, conversely, several other smaller subclusters that could be divided into 1 major and 2 minor AFLP clusters by PCA (I, IVa, IVb). MLST analysis indeed revealed additional heterogeneity in AFLP group I, identifying different clonal complexes (CCs) [CC5, CC8, CC15 (Figure 4)]. CCs are defined as clusters of closely related STs where single differences in the allelic profile are tolerated (9). In contrast, AFLP clusters II and III harbor single CCs, CC30 and CC45, respectively. These 2 major CCs embrace almost half (47.3\%) of all carriage isolates. Clusters IVa and
IVb are associated with CC22 and CC121, respectively (Figure 4). The unsupervised correlation-analysis showed that minor AFLP group IVa also consists of different subclusters (Figure 3).

The horizontal dendrogram in Figure 1 shows the clustering of the 147 AFLP markers, many of which segregate in separate groups. These groups are cluster specific; some of these are shown as boxes A–G (Figure 1). The colored bars on the right represent the distribution of the invasive strains (in children and elderly adults), the impetigo isolates, the MRSA strains, and the reference strains across the dendrogram. The central area of this figure best defines the 5 phylogenetic lineages. The clustered green and red boxes show components of genomic diversity within the 5 S. aureus subpopulations. The distributions of the carriage isolates as well as the invasive isolates are also visualized in Figure 3.

**Carriage versus disease-causing strains of S. aureus.** Among the S. aureus isolates from healthy individuals with nasal carriage, 3 major genetic clusters (I, II, III) could be identified, comprising 760 of the 829 (91.7%) carriage strains. Two minor clusters (IVa, IVb) embrace the remaining 69 (8.3%) carriage strains. All 5 clusters coherently contain carriage strains isolated from children as well as elderly adults (Figure 5). The distribution of the pediatric and geriatric strains across the 5 AFLP clusters was overlapping (Figure 3). However, healthy children more often carried strains from cluster IVa+b than did healthy elderly adults (Fisher’s exact test, \( P < 0.0001 \)), whereas cluster I strains were slightly overrepresented in elderly adults (Fisher’s exact test, \( P = 0.03 \)). Although carriage isolates from children and elderly adults were equally divided in AFLP cluster II (Figure 5), a clone strongly associated with carriage isolates in children could be identified (Fisher’s exact test, \( P < 0.0001 \)) (area between the dotted lines in Figure 3).

Contemporary invasive S. aureus strains (\( n = 164 \)) from children and elderly adults were distributed across all phylogenetic branches and widely scattered across the AFLP dendrogram (Figures 1 and 3). The population structure of contemporary carriage isolates and invasive isolates in the same geographical area appeared to be strongly overlapping (Figure 5). However, bacteremia in elderly adults was significantly more often caused by a IVa cluster–derived strain (Fisher’s exact test, \( P = 0.0095 \)). Additional analysis of AFLP cluster IVa showed that this group consists of distinct subclusters (Figure 3). Therefore it would be incorrect to define this minor cluster as a single invasive clone. Apparently, several subclusters, plotted next to each other by 3D (AFLP cluster IVa), contain proportionately more bacteremia isolates from elderly adults than carriage isolates from the same group. Statistical analysis of major cluster I revealed 2 subclusters potentially associated with invasive disease. Subcluster II (Figure 3) contains proportionately more bacteremia isolates from elderly adults (\( n = 5; 5.6\% \)) in comparison with carriage isolates from elderly adults (\( n = 6; 1.4\% \)) (Fisher’s exact test, \( P = 0.027 \)). Subcluster Ia shows overrepresentation of invasive isolates from both children and elderly adults (\( n = 11; 6.7\% \)) in comparison with carriage isolates from both groups (\( n = 20; 2.4\% \)) (Fisher’s exact test; \( P = 0.01 \) (Figure 3).

The invasive strains from children in major cluster I are associated with hospital-acquired disease (Fisher’s exact test, \( P = 0.01 \)) (data not shown). No significant difference was found in the distribution of isolates from individuals with invasive community-acquired disease versus invasive hospital-acquired disease in the other clusters.

All 5 clusters contain S. aureus strains isolated from children with community-acquired nonbullous impetigo. The distribution is shown in Figures 1 and 5. Compared to pediatric carriage strains,
impetigo isolates were more frequently found in cluster IVb (Fisher’s exact test, \( P < 0.0001 \)) and less frequently in cluster II (Fisher’s exact test, \( P = 0.01 \)) (Figure 5), suggesting clonal expansion of a certain genotype associated with impetigo (22).

**MecA- and PVL-positive strains.** The 21 international epidemic MRSA strains are present in several clusters and subclusters of *S. aureus* (Ic, Id, Ij, II, III, IVa). This strain collection comprises epidemic MRSA from Belgium, Finland, France, Greece, Spain, Germany, and the United Kingdom (20). AFLP cluster I contained disproportionately more MRSA isolates as compared to the other clusters (Fisher’s exact test, \( P = 0.006 \)) (Figure 5), suggesting that many of these epidemic MRSA strains are derived from a common cluster I ancestor. Notably, most of the 21 MRSA strains are located in subcluster Ic (\( n = 9 \); 42.9%; Fisher’s exact test, \( P < 0.0001 \)).

None of the Dutch carriage and clinical *S. aureus* isolates (\( n = 1033 \)) included in this study harbor the *mecA* gene, which is consistent with the reported low MRSA prevalence in the Dutch population (23). Four PVL-positive *S. aureus* strains (1.0%) were found in the pediatric carriage group (\( n = 400 \)) and 1 (0.2%) in the elderly adults carriage group (\( n = 429 \)). Three of 146 (2.1%) blood-culture isolates carried the PVL gene, 2 of which derived from pediatric patients. Seven of the 18 (38.9%) invasive strains isolated from deep-seated or soft-tissue infections in children were PVL positive. There was no significant difference in the presence of PVL when comparing the carriage isolates and invasive blood-culture isolates. In contrast, *S. aureus* strains causing abscesses and arthritis were significantly enriched in the presence of PVL (38.9%) in comparison with the pediatric carriage group (1.0%) (Fisher’s exact test, \( P < 0.0001 \)) and in comparison with the pediatric bacteremia isolates (3.6%) (Fisher’s exact test, \( P = 0.0005 \)). All impetigo strains (\( n = 40 \)) were PVL negative.

**Sequence assessment of AFLP markers.** To determine the origin of genetic polymorphism, the nucleotide sequence of a set of 81 AFLP markers was established, 60 of which were located in cluster-specific marker boxes (Figure 1, boxes A–G). Nineteen of the remaining 21 markers were not cluster specific and were present in almost all 1,056 *S. aureus* strains (red area on the right side of box G, Figure 1). Interestingly, 66 (81.5%) of the 81 markers showed homology to all 3 completed (MW2, Mu50, N315) and 4 unfinished (252, 476, COL, NCTC 8325) genome sequences of *S. aureus*. Conversely, 7 (8.6%) of the 81 markers appear to be absent in all 3 completed genomes (MW2, Mu50, N315) of *S. aureus* (see Supplemental Table 1; supplemental material available at http://www.jci.org/cgi/content/full/114/12/1732/DC1). Of these 7 marker fragments, 3 showed homology to the *S. aureus* strain 252 epidemic MRSA-16 (EMRSA-16), 1 to bacteriophage Φ exfoliative toxin B (ETB) plasmid DNA. Only 2 did not match with any GenBank entry and may represent novel (or hypermutable) *S. aureus* genome segments. The 15 markers that
appear to be absent in at least 1 of the known genome sequences of
*S. aureus* may be associated with (larger) deletions, the existence
of which has been described before (24).

The cluster-specific marker boxes are associated with different
AFLP clusters. Box A is associated with a part of cluster IVa,
B with cluster IVb, C (present) and G (absent) with cluster III,
D (present) and F (absent) with cluster II, and box E with cluster I
(Figure 1). The majority (78.0%) of the markers present in boxes A–D
(not present in cluster I) showed homology to the reference strains
Mu50 and N315, which are located in AFLP cluster I. This sug-
gests that these markers become cluster specific by point muta-
tions rather than genomic rearrangement (deletions or insertions).
This is also supported by additional data (Supplemental Figure 1),
which show that cluster-defining markers in each box are widely
spread across the *S. aureus* chromosome.

Cluster-specific box B, associated with cluster IVb (with overrepresentation of impetigo-isolates), comprises both viru-
ulence genes bacteriophage Φ ETA DNA and ETB plasmid DNA
(Supplemental Table 1).

**Discussion**

The incidence of *S. aureus* bacteremia is rising and has more than
doubled over the past 25 years in some Western European coun-
tries. This increase coincides with a growing rate of community-
acquired disease (in proportion to hospital-acquired disease) and
the epidemic emergence of nosocomial MRSA strains (2). Elderly
adults are most frequently affected, particularly those with addi-
tional predisposing risk factors. Although a part of this effect
may be explained by host susceptibility and population aging,
this does not fully explain the current and drastic rise in the
number of infections. However, little is known about possible
bacterial determinants and whether or not these are associated
with changes in the virulence of *S. aureus*.

We previously showed that AFLP analysis using optimal enzyme
and primer combinations is an excellent tool for assessing genetic
polymorphism in the clonal microorganism *M. tuberculosis* (17).
For *S. aureus* we used the enzyme combination MboI-Csp61, which resulted
in a fingerprint of about 70 polymorphic AFLP fragments well distrib-
uted within the size range of 100 bp to 600 bp in a single AFLP reac-
tion. However, a potential limitation of this AFLP approach is in the
randomness of the restriction sites for *MboI* and *Csp61*. For instance,
genomic islands with underrepresentation of these restriction sites
will not be fully scanned for polymorphism. However, if currently
known genomic sequences of *S. aureus* (Mu50 and N315) are analyzed
by computer for the occurrence of these sites, the average number of
fragments (useful for AFLP) generated per genome is 4.373, and the
average length of the fragments is 200 bp. This suggests that coverage
is indeed random and, in this respect, AFLP provides more of a whole
gene- and genome-scanning approach than MLST, for instance.

In the present study, the population structure of *S. aureus*, iso-
lated from the nose of healthy individuals in the Rotterdam area
(The Netherlands), has been determined. Using ht-AFLP, we ana-
lyzed 147 polymorphic markers for 1,056 *S. aureus* strains. Two large
unbiased strain collections of a nonclinical origin were used. These
collections were obtained from children (<19 years) and elderly
adults (>55 years) with nasal carriage of *S. aureus*. ht-AFLP analysis
revealed the existence of 3 major (I, II, III) and 2 minor (IVa, IVb)
phylogenetic branches. Major AFLP group I could be subdivided
into 10 different subclusters, indicating its heterogeneity. In con-
trast, major clusters II and III showed a very homogenous nature.

Carriage isolates were not completely randomly distributed over
these clusters. AFLP cluster IVa+b represents more carriage strains
from children (isolated in 2002) than carriage strains from elderly
adults (isolated from 1997–1999). A clonal expansion associated
with carriage isolates in children was also observed in AFLP group
II. Conversely, AFLP group I embraces more carriage strains from
elderly adults. Apparently, a certain degree of bacterial popula-
tion heterogeneity exists between the 2 groups included. Whether
the different sampling moments or the different age categories
are fundamental to the observed differences is currently unclear.
Danish studies from the late 1970s demonstrated that waves of
phage types of *S. aureus* go through human populations (25, 26).
Such a phenomenon may explain our current observations. How-
ever, the proportionate distribution of the major phylogenetic
branches (I, II, III) within the overall population of *S. aureus*, iso-
lated from humans, appears to be fairly stable over time and com-
parable for children and elderly adults.
The 5 AFLP clusters identified in this study match with the major CCs as defined by MLST (http://www.mlst.net/). These MLST-based CCs have been defined by studying carriage, invasive, and MRSA isolates mainly from the United Kingdom. The top 5 CCs in the MLST database are CC8, CC30, CC5, CC22, and CC45 (9, 15, 27, 28). We studied more than 1,000 strains isolated in the Rotterdam region (The Netherlands), and we identified essentially the same CCs. Apparently, these clonal clusters have spread successfully in the United Kingdom and The Netherlands and probably worldwide. All large-scale molecular typing studies of nonclinical isolates of S. aureus have been performed using geographically biased strain collections, including our present analysis. However, considering the overlap in MLST types and the similarity in prevalence of certain major clonal clusters, it is supposed that geographical bias is not a confounding factor. An ongoing analysis of strains derived from Indonesian carriers corroborated this hypothesis. The Indonesian strains clustered in the same groups (I to IV), although there was a difference in the relative numbers of isolates per cluster. No new AFLP clusters were identified (Melles et al., unpublished data).

ht-AFLP clusters II and III, identical to MLST CCs 30 and 45, respectively, account for almost half (47%) of all carriage isolates in our population, which suggests that these 2 CCs have evolved to be very successful in colonizing humans.

The above-mentioned observations suggest that virulent and nonvirulent strains are probably not fundamentally different from each other: clinical isolates and MRSA from international sources fall into the same main clusters as carriage isolates. In addition, invasive S. aureus strains, mainly blood-culture isolates (89%), were found to be widely distributed across all 5 AFLP groups in this study. This suggests that strains from each of the genetic clusters are essentially able to cause invasive disease. On the other hand, 1 minor AFLP cluster (IVA) and 2 AFLP subclusters (Ia and If) contained proportionately more invasive isolates. Direct cross-infection is not a likely explanation for this finding because these epidemiologically nonrelated invasive strains were isolated from patients in different medical departments over a period of several years. There have been controversial reports on the existence of hypervirulent lineages of S. aureus (9, 12), but our data suggest that not all S. aureus strains share the same invasive potential. This is not in agreement with recent findings from the Oxford (United Kingdom) region where MLST data suggest that virulence may not be associated with clonal lineages (15). Rather, a relationship between genetic background and disease type is thought to be primarily dependent on the presence of certain toxin genes only (29).

### Table 1

<table>
<thead>
<tr>
<th>Number of Staphylococcus aureus strains included in this study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Carriage</td>
</tr>
<tr>
<td>Invasive — blood culture</td>
</tr>
<tr>
<td>Hospital acquired</td>
</tr>
<tr>
<td>Community acquired</td>
</tr>
<tr>
<td>Invasive — deep-seated or soft-tissue infection</td>
</tr>
<tr>
<td>Hospital acquired</td>
</tr>
<tr>
<td>Community acquired</td>
</tr>
<tr>
<td>Unknown</td>
</tr>
<tr>
<td>Impetigo</td>
</tr>
<tr>
<td>MRSA</td>
</tr>
<tr>
<td>Reference strains</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>
S. aureus strains isolated in the Rotterdam area from patients with nonbullous impetigo showed less clonal diversity than bacteremia-associated strains. Although each of the 5 clusters was found to contain impetigo-derived isolates, AFLP cluster IVb clearly represented significantly more isolates causing impetigo. An explanation for this clonal expansion in impetigo could be the facile spread of this disease. A recent study by Koning et al. concerning nonbullous S. aureus impetigo concluded that a combination of staphylococcal virulence and resistance genes determines the development and course of nonbullous impetigo (22).

The 21 international epidemic MRSA strains included in this study were spread across several lineages, indicating that methicillin resistance has developed in distinct phylogenetic subpopulations of S. aureus, which has been described before (10, 30). MLST studies have placed most of the MRSA in 5 major clusters: CC5, CC8, CC22, CC30, and CC45 (10, 27, 31, 32). Figure 4 shows similar dissemination for strains from a nonclinical origin as well. Notably, computer algorithms used to solve the S. aureus population structure based on MLST data (eBURST; ref. 28) generate clustering identical to that found in the AFLP Spotfire and OmniViz analyses. This provides a solid experimental and mathematical framework for all conclusions drawn from our AFLP data.

We have also assessed the prevalence of mecA and the PVL genes in the carriage and disease-causing populations of S. aureus. All carriage isolates (n = 829) were mecA negative, corroborating data showing the insignificant spread of MRSA in the Dutch community (23). Also, all clinical isolates (n = 204) were mecA negative, as opposed to proportions of above 10% in many of the other European countries, including those sharing borders with The Netherlands (33). PVL is a toxin associated with skin infections (furuncles), community-acquired MRSA infections, and necrotizing pneumonia (31, 34, 35). PVL prevalence in a S. aureus population of nonclinical isolates has never been studied accurately. We found a very low prevalence of 0.6% in a large (n = 829) carriage collection. In this study PVL was carried in 2.1% of blood-culture isolates. However, a significantly higher prevalence of PVL (38.9%) was found in S. aureus strains causing abscesses and arthritis. This is in agreement with the proposed involvement of PVL in severe and invasive (soft-tissue) staphylococcal infections (31, 35).

Our sequence analysis for the clustered AFLP markers suggested that genetic diversity among clusters is primarily caused by point mutation rather than by large-scale deletions or insertions (15). Ultimate proof for this hypothesis should be provided by detailed physical mapping and large-scale sequencing studies, however. Furthermore, we provide indirect proof that 7 genome sequences quite accurately represent the genetic potential of S. aureus as a species; only 4 of 81 marker sequences did not match with the 7 known S. aureus whole genome sequences.

In conclusion, we have solved the population structure of S. aureus of nonclinical origin. Three major and 2 minor phylogenetic branches were identified in our geographically restricted group. Inclusion of invasive S. aureus strains and international-epidemic MRSA revealed that within all major clusters, invasive and multiresistant isolates could be identified. However, clusters and subclusters with overrepresentation of bacteremia-associated isolates were identified. Expansion of multiresistant clones or clones associated with skin disease (impetigo) was observed as well. We suggest that essentially any S. aureus genotype that is carried by humans can transform into a life-threatening human pathogen, but strains from some clonal lineages are more virulent than others.

### Methods

#### Bacterial strains

Two strain collections provided nonclinical S. aureus carriage isolates from healthy individuals. These collections were obtained from 2 study cohorts involving children and elderly adults. In addition, various clinical isolates were included. Contemporary invasive S. aureus strains, isolated from children and elderly adults from the same geographic region, were cultured from normally sterile sites in hospitalized patients with clinical signs of S. aureus infection. Community-acquired invasive disease was defined as isolation of S. aureus from patients within 48 hours of admission; hospital-acquired was defined as isolation of S. aureus 48 hours or longer after admission. The different subcollections are described in more detail below.

In total, 3,198 children from Rotterdam (The Netherlands), aged between 1 and 19 years and participating in the national 2002 Meningococcal Vaccination Campaign, were enrolled (19). A team of 10 research nurses and medical doctors obtained a single nasopharyngeal swab per child at the time of vaccination. S. aureus was isolated from 1,116 children. All isolates were stored at –80°C in broth containing glycerol. A random sample of 400 S. aureus carriage isolates was drawn.

The second collection originated from a community-based prospective study of elderly adults in Rotterdam (The Netherlands) (18). From 3,851 persons aged over 55 years, nasal swab cultures were obtained between April 1, 1997, and December 31, 1999. S. aureus strains were isolated from 1,043 elderly adults. All isolates were stored at –80°C in glycerol-containing broth. A random sample of 429 carriage isolates was drawn.

Seventy-four clinical S. aureus isolates were retrospectively collected from children with invasive S. aureus disease identified in Sophia Children’s University Hospital (Rotterdam, The Netherlands) (2000–2002). Fifty-six isolates derived from blood cultures and 18 isolates were obtained from deep-seated (arthritis; n = 4) or soft-tissue (abscess; n = 14) infections. Ninety clinical isolates from elderly adults (+55 years) were obtained from persons with S. aureus bacteremia identified in Erasmus MC (Rotterdam, The Netherlands) (1997–1999). Forty S. aureus strains obtained from lesions of children suffering from impetigo were randomly drawn from a collection described by Koning et al. (21).

Twenty-one international epidemic MRSA strains were obtained from the HARMONY collection (http://www.harmony-microbe.net), described by Murchan et al. (20). Finally, we included 2 reference strains. N315 is an MRSA strain isolated in 1982, and Mu50 is an MRSA strain with reduced susceptibility to vancomycin isolated in 1997. For both strains genome sequences have been determined (36). In total, 1,056 S. aureus strains were included (Table 1).

#### Cultures, DNA isolation, and detection of mecA and PVL genes

Bacteria were grown overnight at 37°C on Columbia III agar (BD) supplemented with 5% horse blood. Three to 5 colonies were suspended in TEG buffer (25 mM Tris, 10 mM EDTA, 50 mM glucose) containing lysostaphin (50 μg/ml) and incubated at 37°C for 1 hour. DNA was extracted with the MagNA Pure LC DNA Isolation Kit III (Bacteria, Fungi) using the MagNA Pure LC Instrument (Roche Diagnostics) and stored at –20°C. We determined the presence of the mecA and PVL genes by PCR (34, 37).

#### AFLP

AFLP analysis has been performed as described by Van den Braak et al. (17). Using the predictive software package Recomb (Keygene NV) (38), the optimal enzyme and primer combinations were selected. Bacterial DNA was digested with the enzymes MboI and Cip6I, and the linker oligonucleotide pair for MboI (5′-TCGTCGATCTGCCTACC-3′ and 5′-GATCGGTACGAGTC- TAC-3′) and for Cip6I (5′-GAGCTAGTACGCTG-3′ and 5′-TACGTA- GACTCAT-3′) were ligated. Subsequently, a nonspecific preamplification was performed using the MboI primer (5′-GTCGATCGTACGAGTC-3′) and Cip6I primer (5′-GACGATGATCTGACTC-3′). In the final amplification, a 31P-labeled MboI primer containing 1 selective nucleotide (either +C or +G) and a Cip6I primer containing 2 selective nucleotides (+TA) were used.

Amplified material was analyzed using standard polyacrylamide slabgels and

**Downloaded from [http://www.jci.org](http://www.jci.org) on December 28, 2017. [https://doi.org/10.1172/JCI23083)**
subsequent autoradiography. Marker fragments were scored and a binary table scoring marker fragment absence (0) or presence (1) was constructed.

After excision of some selected AFLP fragments from dried gels, re-amplification followed by double-strand sequence analysis was performed (17, 38). The sequence of several additional fragments was determined by computer analysis. The size of the fragments in combination with the selective nucleotides of the AFLP primers facilitated adequate mapping of the fragments on the staphylococcal genome sequence. These fragments were further analyzed by BLAST searching (http://www.ncbi.nlm.nih.gov/BLAST/) (39) against the 3 completed (MW2, NC_003923.1; Mu50, NC_002758.1; N315, NC_002745.2) and 4 unfinished (252, NC_002952; 476, NC_002953; COL, NC_002951; NCTC 8325, NC_002954) genomic sequences of S. aureus. The BLAST results enabled computer-mediated genomic localization and gene annotations of the AFLP fragments.

MLST. MLST was carried out for 56 S. aureus strains using DNA arrays (40). The selected strains were equally distributed across the AFLP dendrogram by selecting 1 out of 10 carriage or invasive strains isolated from children, going from top to bottom through the AFLP dendrogram (Figure 1). MLST data for the 21 epidemic MRSA strains are available at the MLST home page (http://www.mlst.net/) (20).

Data analysis. The method used for 2D clustering of the AFLP data was agglomerative (successive) hierarchical. This was performed using the unweighted pair group method with arithmetic mean (UPGMA). The similarity metric used was Tanimoto (Spotfire DecisionSite 7.2, Spotfire), which defines similarity for binary data (0 and 1) based on the number of positive attributes that 2 records have in common. The resulting dendrogram was ordered by value.

PCA is a standard multivariate method used to reduce the dimensional space of the data to its principal components (PCs) (41, 42). PCA aims to reduce a large number of variables that explain most of the variation in the data (43). It is basically a rotation of axes after centering data to the means of the variables, the rotated axes being the PCs, which are linear combinations of the original variables. The PC computation is displayed as a 3D scatter plot in which the position along the axes shows the PCA score of the strain. PCA was used to identify subgroups of AFLP clusters as hidden by 2D representation of hierarchical clustering. The distribution of the strains in the 5 phylogenetic branches was defined on the basis of PCA. Hierarchical cluster analysis and PCA were performed using Spotfire DecisionSite 7.2 software.

The OmniViz package (OmniViz Inc.) was used to perform and visualize the results of unsupervised cluster analysis in a correlation visualization. This correlation visualization tool displays pairwise correlations among the samples calculated by Pearson’s correlation coefficient (44). In order to reveal correlation patterns, a matrix-ordering method is applied to rearrange the samples. The ordering algorithm starts with the most correlated sample pair and, through an iterative process, sorts all the samples into correlated blocks. Each sample is joined to a block in an ordered manner so that a correlation trend is formed within a block with the most correlated samples at the center. The blocks are then positioned along the diagonal of the plot in a similar ordered manner. As the resultant visualization is symmetrical about the diagonal, half the matrix display is replaced by appropriate clinical data.

To compare the distribution of strain categories in different phylogenetic lineages, Fisher’s exact test was used. A 2-sided P value of less than 0.05 was considered significant.

Acknowledgments

We gratefully acknowledge all participants, personnel, and researchers of the nasopharyngeal carriage study in healthy Dutch children and the community-based prospective cohort study in elderly adults for participating in and facilitating this study. Furthermore, we thank Alewijn Ott for statistical advice. The research described in this communication has in part been facilitated by a grant provided by the Dutch Ministry of Economic Affairs (BTS 00145). AFLP is a registered trademark of Keygene NV, and the AFLP technology is covered by patents (US006045994A, EP0534888B1) and patent applications owned by Keygene NV.

Received for publication August 19, 2004, and accepted in revised form October 19, 2004.

Address correspondence to: Damian C. Melles, Erasmus MC, University Medical Center Rotterdam, Department of Medical Microbiology & Infectious Diseases, Room L-313, Dr. Molewaterplein 40, 3015 GD Rotterdam, The Netherlands. Phone: 31-10-463-3510; Fax: 31-10-463-3875; E-mail: d.melles@erasmusmc.nl.
exfoliative toxin B, Panton-Valentine leukocidin, and the multidrug resistance plasmid pSK41.


