A Geographic Variant of the *Staphylococcus aureus* Panton-Valentine Leukocidin Toxin and the Origin of Community-Associated Methicillin-Resistant *S. aureus* USA300

F. Patrick O’Hara, Nicolas Guex, J. Michael Word, Linda A. Miller, Julie A. Becker, Stacey L. Walsh, Nicole E. Scangarella, Joshua M. West, Ribhi M. Shawar, and Heather Amrine-Madsen

1Evolutionary and Structural Bioinformatics, 2Infectious Diseases Medicine Development Center, and 3Department of Microbiology, GlaxoSmithKline, Collegeville, Pennsylvania; 4Evolutionary and Structural Bioinformatics, GlaxoSmithKline, Research Triangle Park, North Carolina

(See the editorial commentary by John and Lindsay, on pages 175–8.)

**Background.** The majority of recent community-associated methicillin-resistant *Staphylococcus aureus* (MRSA) infections in the United States have been caused by a single clone, USA300. USA300 secretes Panton-Valentine leukocidin (PVL) toxin, which is associated with highly virulent infections.

**Methods.** We sequenced the PVL genes of 174 *S. aureus* isolates from a global clinical sample. We combined phylogenetic reconstruction and protein modeling methods to analyze genetic variation in PVL.

**Results.** Nucleotide variation was detected at 12 of 1726 sites. Two PVL sequence variants, the R variant and the H variant, were identified on the basis of a substitution at nt 527. Of sequences obtained in the United States, 96.7% harbor the R variant, whereas 95.6% of sequences obtained outside the United States harbor the H variant; 91.3% of MRSA isolates harbor the R variant, and 91.3% of methicillin-susceptible strains harbor the H variant. A molecular model of PVL shows 3 mechanisms by which the amino acid substitution may affect PVL function.

**Conclusions.** All sampled PVL genes appear to share a recent common ancestor and spread via a combination of clonal expansion and horizontal transfer. US isolates harbor a variant of PVL that is strongly associated with MRSA infections. Protein modeling reveals that this variant may have functional significance. We propose a hypothesis for the origin of USA300.

Methicillin-resistant *Staphylococcus aureus* (MRSA) has been a major cause of nosocomial infections since it was first described in 1961 [1], and hospital-associated strains (HA-MRSA) are frequently resistant to multiple classes of antibacterials [2]. Beginning in the 1990s, outbreaks of community-associated MRSA (CA-MRSA) have become increasingly frequent, and some CA-MRSA strains have acquired multidrug resistance [3]. CA-MRSA infections in the United States are predominantly caused by a single epidemic clone, USA300 [4–6], which has multilocus sequence typing (MLST) sequence type (ST) 8, pulsed-field gel electrophoresis (PFGE) type USA300, and staphylococcal chromosomal cassette mec (SCCmec) type IV. First described in 2000, USA300 rapidly supplanted the original US CA-MRSA clone known as MW2 (PFGE type 400 and ST1) and has begun to displace traditional HA-MRSA strains [5, 7]. Both USA300 and MW2 express the Panton-Valentine leukocidin (PVL) genes [8, 9], which are associated with highly virulent infections of the skin and soft tissue as well as necrotizing pneumonia and necrotizing fasciitis [10, 11]. The PVL genes (*lukF-PV* and *lukS-PV*) encode a cytotoxin that has multiple functions: it causes concentration-dependent necrosis and apoptosis of human polymorphonuclear neutrophils (PMNs) [12], lysis of human monocytes and macrophages [13], activation of calcium channels [14], and global changes in gene transcription [15]. PVL has been shown to be sufficient to cause necrotizing pneumonia in a murine model [15].
CA-MRSA from other countries are genetically distinct but can also harbor PVL and SCCmec type IV [16]. To date, none of the sequenced S. aureus genomes publicly available represent PVL-positive methicillin-susceptible S. aureus (MSSA) strains.

The lukSF-PV genes are carried on a prophage, and there is in vitro evidence that the genes may be transferred from one strain to another via phage transduction [17]. Thus, PVL genes may be transmitted vertically within the same clones or horizontally to another via phage transduction [17]. Thus, PVL genes may be transferred from one strain to another via phage transduction [17].

Within the studies to date have focused on evaluating the genetic variation among different clones. Although variation in the genetic back-transmitted vertically within the same clones or horizontally to another via phage transduction [17]. Thus, PVL genes may be transferred from one strain to another via phage transduction [17].

PVL genes in vitro evidence that the genes may be transferred from one strain to another. We could amplify the full fragment from 174 of these isolates. We demonstrate that all US MRSA isolates in the sample possess a unique geographic variant of PVL. Furthermore, we show that the geographic variant is distinguished by an amino acid substitution. Using molecular modeling, we provide evidence that this substitution may have functional significance and may contribute to the success of USA300. Our data allow us to suggest a hypothesis for the origin of the USA300 clone.

METHODS

Bacterial isolates. Single S. aureus isolates were obtained during the baseline visit from the skin specimens of adult and pediatric patients enrolled in 3 global clinical trials of uncomplicated skin infections. All isolates were collected in 2004–2005 from investigator sites in North America (including 13 US states), Germany, Russia, India, South Africa, Peru, and Costa Rica. Isolates were stored frozen at −70°C and subcultured on Luria-Bertani medium for testing. Multiplex polymerase chain reaction (PCR) technology was used to characterize all S. aureus isolates on the basis of the presence or absence of the PVL and meca genetic determinants [18]. Using a PCR-based test for the lukSF-PV genes, we determined that 177 isolates were PVL positive. We could amplify the full fragment from 174 of these isolates, so these 174 isolates (57 MRSA and 117 MSSA) were included in the present study. MLST was performed on all isolates, as described by Enright et al. [19]. The MLST sequences were compared with those in the public database (http://www.mlst.org) to generate STs, and the STs were grouped into clonal complexes by use of the eBURST analysis tools [20, 21]. A clonal complex is defined as a group of STs in which each ST shares at least 6 of 7 identical loci with at least 1 other ST in the group. Clonal complexes are named after the most common ST in the complex.

Phylogenetic analyses. Contigs were assembled using Sequencer (version 4.5; Gene Codes), homologous DNA sequence alignments were generated using ClustalX (version 1.83) [22], and sequences were analyzed and manually edited in GeneDoc (version 2.6) [23]. Sequences of lukSF-PV from the published genomes of the CA-MRSA strains USA300 (CP000255 region: 1546170–1548087) and MW2 (BA000033 region: 1529381–1531298) were collected from GenBank and aligned for use as comparators. Statistical parsimony (haplotype) networks were generated using TCS (version 1.21) [24], with gaps treated as missing data. The statistical parsimony connection limit was set at 99%.

Molecular modeling. A crude model of the LukSF-PV pore complex as an octamer was built from the crystal structure of the related heptameric α-hemolysin pore (protein data bank identifier 7ahl [25]) using Swiss-PdbViewer [26] and MAGE [27] by superimposing the crystal structures of the LukF-PV monomer (1pvl [28]) and the LukS-PV monomer (1t5r [29]) onto chain B of the α-hemolysin structure.

RESULTS

PVL sequence variation among S. aureus isolates. A 1726-bp fragment comprising 90% of the full sequence of the lukSF-PV genes from 174 isolates was amplified by PCR, and the products were sequenced. The sequence is highly conserved, but nucleotide variation is seen at 12 sites (figure 1). The lukS-PV sequence is more variable than the lukF-PV sequence; 8 of the 12 variable sites are found in lukS-PV. Eight of the 12 variable sites contain minor variants; the minor allele was present in only 1–5 sequences. Four of these minor variable sites result in amino acid change. At the remaining 4 variable sites, a major variant was observed in >5 sequences. One of these major variable sites results in an amino acid change.

We identified 13 unique haplotypes among the sampled sequences. The sequence of the most common haplotype is identical to that of USA300. The sequences can be assigned to 4 major haplotype groups on the basis of variation at positions 527, 663, and 1396 of the nucleotide sequence. Sixty-three sequences have

Sequencing of the lukSF-PV genes. An internal fragment of the lukSF-PV coding region starting 124 bp downstream of the start of lukS-PV and ending 69 bp upstream of the end of lukF-PV was amplified by PCR, and both strands were sequenced. The following primers were used for amplification and sequencing: PVL1F (5’-gggatat-tggtcgctaggtatgcaa-3’), PVL2F (5’-gacgtggaagtagcaca-3’), PVL1R (5’-ctgatcttcccccaactctc-3’), PVLInt3R (5’-tattggagagccacctgtgc-3’), PVLInt2R (5’-caaatcagtgtaacctgagc-3’), PVLInt2F (5’-caactgtcaactagctgccgtaa-3’), PVLInt1R (5’-ttcctcctaaagtaaacctgccg-3’), and PVLInt3F (5’-gagggactatgcccagagatgtac-3’). Sequences have been submitted to GenBank under the accession numbers EF571668–EF571841.
a G at position 527, which results in an arginine (R) at aa 176. The remaining 111 have an A at 527, resulting in a histidine (H) at aa 176. These will be referred to hereafter as the “R variant” and the “H variant,” respectively. The H variant can be further broken into groups H1, H2, and H3.

Variation at aa 176 correlates strongly with the geography of the isolate harboring the PVL sequence and is also correlated with the presence of meca (table 1). Of the 63 R variant sequences, 58 are from the United States. Of the remaining 111 sequences, 109 are found outside the United States. The H1 and H2 variants are commonly found in South Africa and India, whereas the H3 variant is common in Europe (Germany and Russia). The R variant is strongly associated with presence of meca; 53 of 63 isolates are meca positive. The opposite is true for

<table>
<thead>
<tr>
<th>PVL sequence variant</th>
<th>meca status</th>
<th>Geographic distribution</th>
<th>Distribution among common clonal complexes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>Negative</td>
<td>United States</td>
</tr>
<tr>
<td>R</td>
<td>53</td>
<td>10</td>
<td>58</td>
</tr>
<tr>
<td>H</td>
<td>5</td>
<td>106</td>
<td>2</td>
</tr>
<tr>
<td>H1</td>
<td>2</td>
<td>51</td>
<td>1</td>
</tr>
<tr>
<td>H2</td>
<td>3</td>
<td>38</td>
<td>1</td>
</tr>
<tr>
<td>H3</td>
<td>0</td>
<td>17</td>
<td>0</td>
</tr>
</tbody>
</table>
isolates carrying the H variant. Only 5 of 111 isolates carrying the H variant of PVL also carried mecA.

After performing MLST and eBURST analyses, we were able to assign each isolate to a clonal complex (table 1). These analyses reveal that 57 of 63 isolates harboring the R variant of PVL are found in the same clonal complex, CC8. The H variant is frequently found in CC121 and CC22 but is dispersed among several clonal complexes. H1 is found predominantly in South Africa and India and is normally associated with CC22. H2 is also common in South Africa and India but is associated with CC121. H3 is common in Europe and is associated with CC121.

Recent evolutionary history of PVL and the origins of PVL-positive MRSA. To reconstruct the evolutionary history of the PVL genes, we used TCS to estimate a phylogenetic network of 174 PVL sequences (figure 2). Branches indicating a single-nucleotide substitution connect ellipses representing the PVL alleles sampled in this study. All PVL sequences are connected on the network at a 99% confidence level without the need to include hypothetical missing alleles or loops, pointing to the absence of any ambiguities in connecting the sampled haplotypes. This indicates that the relationships among the individual PVL alleles are fully resolved. On the basis of statistical parsimony, the H variant is predicted to be the ancestor of the R variant. This is consistent with the fact that the H variant has a broader geographic distribution, is spread among several clonal complexes, and contains more genetic variation (11 variant genotypes vs. 2 for the R variants). Statistical parsimony predicts that the H2 variant is the likely ancestral allele that gave rise to the other H variants. This sequence is identical to other published sequences, including USA1100 from Japan and A980470 from France. CC121, which is predominantly MSSA, serves as the reservoir for the H2 variant. Through a combination of clonal expansion and horizontal exchange, the H2 variant spread and gave rise to the H3 and H1 variants, which are found in most PVL-positive isolates around the world; these isolates are predominantly mecA negative. A single nucleotide substitution differentiates the H3 variant from the rare R variant, which is represented by a single sequence among the sampled isolates. This sequence is identical to the lukSF-PV sequence found in the MW2 strain, and the isolate is ST1 and mecA positive, just like MW2. The phylogenetic history of PVL represented in the network allows us to reconstruct the origins of the PVL-positive CA-MRSA strains MW2 and USA300. The MW2 strain originated when a PVL-negative strain from CC1 obtained PVL from a PVL-positive CA-MRSA strain from CC121, either before or after the acquisition of mecA by CC1. The PVL in MW2 was then transferred to a CC8 strain harboring SCCmec type IV, leading to the formation of the USA300 strain. This strain rapidly became the predominant CA-MRSA strain in the United States between its discovery in 2000 and the time of this study (2004–2005).

Functional implications of the histidine to arginine substitution at position 176. Substitution of arginine for histidine changes the physiochemical properties of the site. The side chain of histidine contains an aromatic imidazole ring that can make 2 hydrogen bonds and that, because it has a pK_a that ranges from ~6 to 7, is either neutral or positively charged depending on the local environment [30]. The side chain of arginine is considerably longer, ending in a guanidinium group that can make 5 hydrogen bonds and is always positively charged under physiological conditions (pK_a ~12.5 [31]).

Because Jayasinghe and Bayley [32] have proposed that the LukFS-PV pore complex is an octamer consisting of alternating LukF-PV and LukS-PV monomers, we constructed a crude model of such a pore (see Methods for details) to assess the functional relevance of the mutation H176R (H148R in the crystal structure of LukS-PV and in our model). On the basis of this model, it appears that the mutation is located at the surface of LukS-PV, in its interaction surface with LukF-PV (figure 3A and 3B), suggesting that changes at this site may have functional consequences for the PVL toxin and lending itself to several intriguing hypotheses.

First, H176(H148) is predicted to be located near a region of the LukF-PV binding partner with the sequence GNIXSG, a motif conserved among members of the LukF-PV family except for residue x, which is Y (tyrosine) in LukF-PV and N (asparagine) in HlgB and LukD, and both residues are capable of accepting hydrogen bonds. In our model, position x is the LukF-PV residue closest to H176(H148). Perhaps the tyrosine side chain is large enough to form hydrogen bonds with both histidine and
arginine, whereas the shorter asparagine side chain can interact only with arginine. It is therefore possible that the arginine allele enables LukS-PV to interact with additional partners besides LukF-PV (e.g., LukD and HlgB) and thus increases the diversity and numbers of pores formed, essentially acting as an amplifier.

Assuming that the mechanism of pore assembly for LukSF-PV is similar to that of α-hemolysin [33], the octamer is assembled at the surface of the membrane after receptor recognition and before the conformational change that results in the formation of the pore itself. Interestingly, position 176(148) is on the same side and near the F163-P175 loop, which corresponds to loop M174-N193 of α-hemolysin, shown to interact specifically with caveolin-1 at the cell surface [34]. Assuming this loop would play a similar role in LukS-PV, position 176(148) may influence recognition of LukS-PV cognate membrane receptor and help initiate oligomer recruitment. Indeed, LukS-PV has been shown to possess a high specific affinity for a unique receptor on PMNs and monocytes [35].

A final hypothesis for functional change at position 176(148) involves an interaction between Arg176(Arg148) and the 3’ end of the LukF-PV subunit. Because of its location in a β sheet, the arginine could adopt a nearly fully extended side chain conformation (the most frequently observed conformation for an arginine located in a β sheet, according to Lovell et al. [36]). This conformation is compatible with the LukS-PV structure, and, after moving M300 to an allowed region of the Ramachandran plot in the model, it could allow the positively charged Arg176(Arg148) to form a strong ionic interaction with the negatively charged carboxy-terminus of LukF-PV (figure 4A). The side chain of His176(His148) does not reach as far and may not be positively charged (figure 4B). It is worth noting that the last 2 residues of monomeric LukF-PV (M300 and S301) have high B factors (indicating increased thermal motion) relative to any other residues of LukF-PV. Therefore, it is possible that the mutation affects the ability of LukS-PV to interact with LukF-PV, thus influencing pore stability or the rate of pore formation.

**DISCUSSION**

Despite being gathered from isolates with diverse histories, including clonal complex membership, geography, and patient demographics, lukSF-PV sequences are highly conserved at the nucleotide level, indicating that they all share a recent common ancestor. The observed patterns of divergence of lukSF-PV sequences, combined with MLST analysis, suggest that genes encoding PVL can rapidly spread via a combination of 2 mechanisms: clonal expansion and horizontal gene transfer. The prevalence of identical R variant sequences in the United States, 90% of which are associated with CC8, reveals that these PVL genes have primarily spread rapidly via a single, highly successful clone. A similar pattern is observed in the H3 variant in Europe, where 83% of sampled isolates are associated with CC121. In contrast, the H1 variant is most often associated with CC22, but CC22 isolates account for only 47% of the H1 variant sequences observed in this study. H1 variant sequences are also found in 8 other clonal complexes. This pattern suggests that CC22 serves as the primary reservoir for H1 variant genes that spread via horizontal gene transfer to other clonal complexes at a relatively high rate. The spread of highly ecologically successful virulent clones is an established property of the population biology of S. aureus [37, 38]. The exchange of mobile genetic elements, particularly when associated with virulence and resistance determinants, is also predicted to be a significant force driving the evolution of the accessory genome [39]. The lukSF-PV gene has been shown to be located within the sequence of a prophage, and the capacity for phage-mediated transfer has been demonstrated...
in vitro [17]. Our survey of lukSF-PV sequence variation provides evidence supporting both of these mechanisms.

The geographic distribution of the major variants of lukSF-PV is striking. The presence of arginine at position 176 marks 58 of 60 isolates sampled from the United States, including all MRSA collected from the United States. Interestingly, this identical variant is found to be broadly distributed in the United States, including states across the continental United States as well as Hawaii. This variant is associated with a particularly virulent strain of CA-MRSA known as the USA300 clone, which has been shown to have a wide geographic distribution in the United States [4]. Outside the United States, 106 of 111 sampled isolates have histidine at residue 176. All but 5 of these samples are MSSA, and all non-US MRSA strains collected in this study harbor the H variant. Together with the observation that most CC8 isolates are MRSA, this suggests that the acquisition of PVL in a US clone of MRSA led to the development of the highly successful and virulent USA300 strain.

Previous investigations into the evolutionary history of MRSA have concluded that genetically distinct lineages have acquired SCCmec multiple times in independent horizontal transfer events [40, 41]. Data presented here indicates that multiple lineages possess lukSF-PV genes that all derive from a recent common ancestor, indicating that independent horizontal transfer events of PVL also contribute to the evolution of S. aureus. Furthermore, we show evidence that lukSF-PV may be acquired by strains that are predominantly MRSA (e.g., CC8), whereas previous models have predicted that CA-MRSA originated with the acquisition of SCCmec type IV by PVL-positive MSSA [42, 43]. Countries in which the frequency of MRSA is still low may benefit from analysis of MSSA strains that are PVL positive in order to anticipate the spread of lukSF-PV to MRSA clones.

In addition to its utility as an epidemiologic marker, the substitution at aa 176 may have functional implications for the PVL toxin. Molecular modeling suggests that the substitution could enable LukS-PV to form pores with subunits other than LukF-PV, amplifying the rate of pore formation. It could also alter interaction with the yet-unknown cognate cell-surface receptor, required for competent pore formation. Of particular interest is the hypothesis that an arginine at site 176 can extend further than a histidine and thus can interact with the 3’ end of the LukF-PV protein, stabilizing the LukFS-PV interaction. A more stable interaction between the LukS and LukF subunits could allow faster and more efficient pore formation, which could be critical to toxicity, because there is evidence that PVL function is concentration dependent [12]. This proposal is well supported by the molecular model and implies that the R variant is more fit than the H variant.

The relationship between PVL and the fitness of pathogenic clones is unclear at this time, but the connection between PVL and virulence has been the subject of considerable study. PVL is strongly associated with highly virulent strains [44], particularly those that cause necrotizing pneumonia [45]. Indeed, PVL-positive strains and purified PVL toxin have been shown to be sufficient to cause pneumonia in mice [15]. The addition of PVL to a relatively avirulent strain significantly increases PMN apoptosis [12], but other factors affect PMN lysis in more virulent strains [46]. Although the association between PVL and serious skin abscesses in humans [11] is at odds with the observation that PVL does not significantly affect abscesses in mice [46], it should be noted that PVL does not target mouse neutrophils [47] and that staphylococcal virulence in mice generally requires

Figure 4. Proposed structural significance of arginine at position 176(148). A, Close-up of the interaction between the histidine (H variant) at position 176(148) of LukS-PV, which may or may not carry a positive charge, and the negatively charged carboxy-terminus of LukF-PV. B, Proposed strong ionic interaction between the positively charged arginine (R variant) at position 176(148) and the carboxy-terminus. For clarity, only 1 LukF (orange) and 1 LukS (blue) monomer “head” are represented. The “stalk” of the octamer is represented in dark gray.

192 • JID 2008:197 (15 January) • O’Hara et al.
higher concentrations of enterotoxins than those needed in humans [48]. The observation that PVL significantly alters gene expression in S. aureus [15] suggests that PVL may affect the virulence and fitness of the organism in diverse ways. The clinical relevance of PVL in human infections and the functional implications of sequence variants of PVL warrant further investigation.

Analysis of the phylogenetic network of lukSF-PV sequences in light of what we know about their geography, clonality, and mecA status allows us to reconstruct the evolutionary history of the R variant of PVL present in US MRSA, including the USA300 strain. The H variant is older than the R variant. This is supported by its broader geographic distribution, its distribution among several clonal complexes, and its increased sequence diversity relative to that of the R variant. The older variant is more frequently associated with MSSA, whereas the newer variant is associated with CA-MRSA, suggesting that PVL-positive MSSA has acted as a reservoir for PVL genes that are subsequently incorporated into CA-MRSA lineages. Statistical parsimony predicts that the R variants are derived from the H3 variant, which is found in isolates that are all mecA negative, and predominantly CC121. A single nucleotide substitution led to the R variant found in the MW2 lineage, and this PVL was transferred into CC8, leading to the development of USA300, which spread rapidly in the United States.

Using a diversity of approaches, including protein modeling and phylogenetics, we have revealed a geographic variant of the PVL toxin containing an amino acid substitution that may have functional importance. Hopefully, the insights that have been gained into the evolutionary history of current CA-MRSA strains such as USA300 will help to predict and even mitigate the emergence of future strains.

Acknowledgment

We thank the GlaxoSmithKline United States–based sequencing facility, under the management of Ganesh Sathe, for its efforts on this project.

References


Origin of CA-MRSA USA300 • JID 2008:197 (15 January) • 193

Downloaded from https://academic.oup.com/jid/article-abstract/197/2/187/810075 by guest on 09 November 2018


