INFECTION AND IMMUNITY, Mar. 2005, p. 1856–1860 Vol. 73, No. 3
0019-9567/05/$08.00+0 doi:10.1128/IAI.73.3.1856–1860.2005

Factors Characterizing *Staphylococcus epidermidis* Invasiveness Determined by Comparative Genomics

Yufeng Yao,[1] Daniel E. Sturdevant,[1] Amer Villaruz,[1] Lin Xu,[2] Qian Gao,[2] and Michael Otto[1]*

Laboratory of Human Bacterial Pathogenesis, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, The National Institutes of Health, Hamilton, Montana,[1] and Key Laboratory of Medical Molecular Virology, Shanghai Medical College, Fudan University, Shanghai, China[2]

Received 27 July 2004/Returned for modification 10 September 2004/Accepted 17 October 2004

Virulence mechanisms of the leading nosocomial pathogen *Staphylococcus epidermidis* are poorly understood. We used microarray-based genome-wide comparison of clinical and commensal *S. epidermidis* strains to identify putative virulence determinants. Our study revealed high genetic variability of the *S. epidermidis* genome, new markers for invasiveness of *S. epidermidis*, and potential targets for drug development against *S. epidermidis* infections.

Staphylococcus epidermidis is the most prevalent cause of nosocomial infections, costing the public health system ~$1 billion/year in the United States alone (18). Usually an innocuous commensal microorganism on human skin, *S. epidermidis* can cause severe infection after penetration of epidermal and mucosal barriers, which frequently occurs in the hospital during the insertion of indwelling medical devices (18). *S. epidermidis* mostly lacks components that are easily recognized as virulence factors, such as toxins or aggressive degradative exoenzymes (18). Furthermore, genetic manipulation of *S. epidermidis* is very difficult. For these reasons we have a serious lack of knowledge about the basis of *S. epidermidis* virulence. However, discovering the genes that determine success of *S. epidermidis* as an opportunistic pathogen is a crucial prerequisite step for designing therapeutic interventions directed to control *S. epidermidis* infections.

In contrast to studies performed with some other pathogenic bacteria, approaches using in vivo expression technology to identify virulence genes in *Staphylococcus* have proven problematical (3, 9). Low infectivity and the resulting difficulty to establish reproducible animal infection models further complicate the use of this technology with *S. epidermidis*. Therefore, we used comparative genomics of clinical and benign strains as an alternative approach to identify *S. epidermidis* virulence determinants. This approach has been used to characterize virulence factors in the opportunistic pathogen *Pseudomonas aeruginosa* (20). In our study, 22 strains isolated from prostheses infections and 20 strains isolated from the skin of healthy individuals were analyzed by DNA/DNA hybridization of genomic DNA on a whole-genome *S. epidermidis* microarray. The microarray contained a 70mer oligonucleotide of every gene found in the genome of *S. epidermidis* RP62A (sequence available at www.tigr.org). We have described synthesis and characterization of the microarray previously (21). Specifically, we have verified that all oligonucleotides hybridize with control DNA isolated from strain RP62A. The strains used in the present study were a subset of an essentially nonclonal collection (5) and were further screened to exclude related strains that were sometimes found in the same patient. We determined the degree of relatedness of the strains by microarray analysis (Fig. 1). As anticipated from the preselection, microarray data confirmed that the strains were not clonal. Results of the distribution of individual genes were analyzed by Fisher's exact test. A P value of <0.05 was considered significant.

A total of 939 (36%) genes in the control strains and 425 (16%) genes among clinical strains lacked a hybridization signal in at least one strain, indicating absence or significant mutation. These data reveal considerable, previously unknown genetic variability of the *S. epidermidis* genome. A total of 59 genes showed a significantly disproportionate distribution between the two groups (Table 1 and Table 2). Also, 39 genes were found to be more frequent among clinical strains than among commensal strains (Table 1). Importantly, these genes included the *ica* locus, which encodes the biosynthetic machinery for the exopolysaccharide PIA (6), and genes related to the insertion sequence IS*256*. *ica* and IS*256* are among the very few factors that have been described as determinants of virulence and markers for invasiveness of *S. epidermidis* (6, 7, 15, 23). Notably, these findings validated our approach and further confirmed the use of *ica* and IS*256* as markers for invasiveness of *S. epidermidis*. However, many other genes revealed an even more significant difference. Particularly, a gene encoding a 190-kDa cell surface protein with similarity to a streptococcal hemagglutinin binding protein showed the most pronounced difference. It was present in 16 of 22 (73%) clinical strains but in only 5 of 20 (25%) control strains ($P = 0.0026$). We validated the distribution of the gene coding for this protein by analytical PCR, using a different strain collection containing strains from different infections and skin strains from Shanghai, China (Table 3). The results confirmed that the gene occurred significantly more frequently among clinical isolates than among isolates from healthy individuals ($P = 0.03$).

So-called MSCRAMMs (for "microbial surface components recognizing adhesive matrix molecules") are believed to play

* Corresponding author. Mailing address: 903 S. 4th St., Hamilton, MT 59840. Phone: (406) 363-9283. Fax: (406) 375-9677. E-mail: motto@niaid.nih.gov.

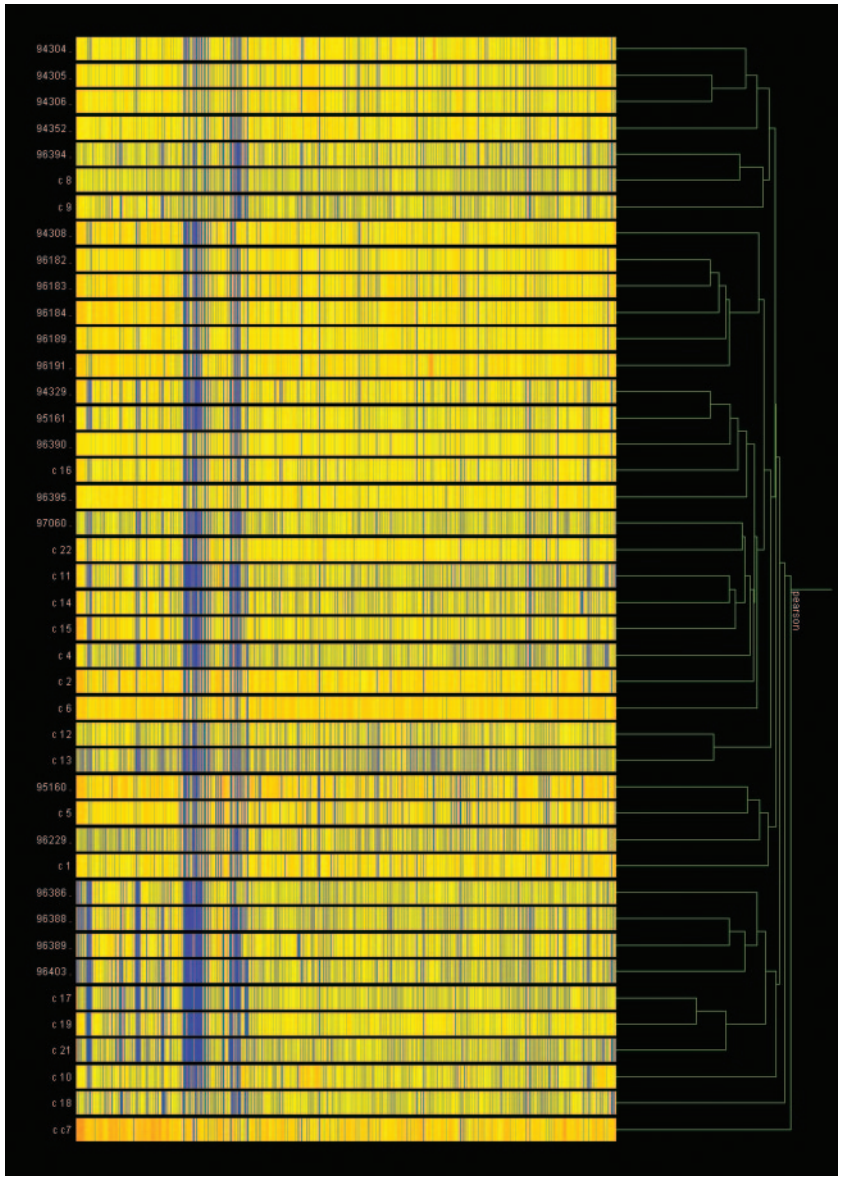

FIG. 1. Relatedness of strains used for microarray-based investigation of gene distribution in this study. Strain names for strains from infections are according to Galdbart et al. (5), by whom the strains were first described. Control strains are numbered from c1 to c22 and are also from that study. Absence of genes is shown in blue; presence of genes is shown in yellow and red. A dendrogram is shown on the right. The analysis was performed with GeneSpring software.

an eminent role in bacterial pathogenesis during the establishment of infection (13). *S. epidermidis* MSCRAMMs, e.g., the fibrinogen-binding protein Fbe, are presently under intense investigation for use as drug targets or antigens for vaccine development (12, 14). Many but not all MSCRAMMs have an LPXTG motif for linkage to the bacterial cell surface (10). The 190-kDa protein gene lacked a clearly distinguishable LPXTG motif but revealed repeat regions and a putative cell wall binding domain that are typical for MSCRAMMs (13). We detected 10 putative MSCRAMM genes with an LPXTG motif in the *S. epidermidis* genome. Remarkably, 8 of these 10 putative MSCRAMM genes were absent from at least one strain in our study, as previously shown for the *sdrF* gene (11), indicating high genetic variability for this class of surface proteins (data not shown). The two MSCRAMMs found to be present in all strains were SE0828 and SE1682, two yet uncharacterized proteins. However, the 190-kDa protein gene was the only putative MSCRAMM that appeared significantly more frequently among invasive strains, suggesting a crucial role for this protein in *S. epidermidis* pathogenesis.

Several genes with a previously proposed role in virulence were more frequently present in invasive strains. For example, the lipoprotein signal peptidase LspA is required for the secretion of lipoproteins, which represent surface-attached extracellular proteins that may be involved in various virulence mechanisms (2, 19). Furthermore, SsaA is an abundant extracellular antigenic protein in *S. epidermidis* for which a role in pathogenesis has been proposed (8). Moreover, we detected

TABLE 1. Genes more frequent among invasive strains of *S. epidermidis*

Protein designation(s) or description[a]	Gene[a]	No. of control strains with gene present (total = 20)	No. of clinical strains with gene present (total = 22)	*P* (Fisher's exact test)
Cell surface protein (similar to streptococcal Hemagglutinin binding protein)	SE2251	5	16	0.00265532
Hypothetical protein	SE0240	13	22	0.00287342
UDP-*N*-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopimelate-D-alanyl- D-alanyl ligase (*murF*)	SE1680	13	22	0.00287342
Unknown	SE1646	14	22	0.00738879
Conserved hypothetical protein	SE0864	12	21	0.00770752
Proline-betaine transporter homologue	NA	3	12	0.0106516
Similar to long chain fatty acid CoA ligase	SE0344	7	17	0.0116155
Intercellular adhesion protein C (*icaC*)	NA	7	17	0.0116155
Antibiotic transport-associated protein	NA	6	15	0.0170854
Lipoprotein signal peptidase LspA	SE0871	6	15	0.0170854
Conserved hypothetical protein	SE1952	15	22	0.0182257
Acetate-CoA ligase	SE2161	15	22	0.0182257
Conserved hypothetical protein	SE0692	13	21	0.0182257
Phospho-*N*-muramic acid-pentapeptide translocase	SE0857	13	21	0.0182257
Conserved hypothetical protein	SE1071	15	22	0.0182257
Unknown	NA	15	22	0.0182257
Unknown	NA	10	19	0.0185603
Transposase for IS*256*	NA	1	8	0.0220505
Conserved hypothetical protein	SE0329	8	17	0.02653
Unknown	NA	8	17	0.02653
Hypothetical protein (*Lactobacillus gassen*)	NA	7	16	0.0286499
Putative 4-diphosphocytidyl-2C-methyl-D-Erythritol synthase	SE0319	4	12	0.0289139
Arsenic efflux pump protein (ArsB)	SE0135	5	13	0.0334195
Oligoendopeptidase	SE1065	2	9	0.0353177
Ribulose-phosphate 3-epimerase	SE0897	6	14	0.0365301
Transposase for insertion sequence element IS*257* in transposon Tn4003	SE0079	6	14	0.0365301
Conserved hypothetical protein	SE2232	11	19	0.0400408
Conserved hypothetical protein	SE0142	11	19	0.0400408
Hypothetical protein	NA	14	21	0.0408025
Conserved hypothetical protein	SE0420	16	22	0.043286
Secretory antigen SsaA	NA	16	22	0.043286
Phosphatidylglycerophosphate synthase	SE0960	16	22	0.043286
Urease accessory protein (*Bacillus halodurans*)	SE1866	16	22	0.043286
Conserved hypothetical protein	SE0693	16	22	0.043286
Hypothetical protein (*Bacillus anthracis* A2012)	NA	16	22	0.043286
Unknown	NA	16	22	0.043286
Transposase (IS256, TN4001)	NA	1	7	0.047124
IS*256* transposase	NA	1	7	0.047124
IS*256* transposase	NA	1	7	0.047124

[a] Gene annotation according to reference 22 (modified).

two genes encoding murein synthesis enzymes in this group, a phosphomuraminic acid pentapeptide translocase-encoding gene and the *murF* gene. *murF* and other murein synthesis genes are critical for bacterial survival, and their gene products are under current investigation as potential novel drug targets (4). The absence of a DNA hybridization signal for an essential gene like *murF* is likely due to significant gene mutation. A mutated *murF* has been shown to influence methicillin resistance in *Staphylococcus aureus* (17). A similar role of *murF* in *S. epidermidis* might cause the observed lower gene frequency in control strains compared to clinical strain results. Taken together, our data help to emphasize the importance of specific genes among a variety of proposed virulence factors for further investigation of *S. epidermidis* pathogenesis.

The genes that were more frequent in invasive strains comprised several resistance genes such as genes coding for an antibiotic transport-associated protein and the arsenic efflux pump ArsB (1). In accordance with the latter finding, the gene coding for ArsD, the *trans*-acting repressor of the arsenic resistance operon, was less common among invasive strains. On the other hand, some putative antibiotic resistance genes showed a significantly higher frequency in the control group. Further, a proline-betain transporter homolog gene was more frequent among invasive strains than in commensal strains,

TABLE 2. Genes more frequent among commensal strains of *S. epidermidis*

Protein designation(s) or description[a]	Gene[a]	No. of control strains (total = 20)	No. of clinical strains (total = 22)	*P* (Fisher's exact test)
Conserved hypothetical protein	SE1875	20	9	2.25E-05
Sugar transporter	SE0123	14	3	0.00038208
Beta-lactamase	SE1608	19	10	0.00063793
Arsenical resistance operon trans-acting ArsD	SE0138	15	6	0.00265532
Accumulation-associated protein (AAP)	SE0175	20	14	0.00377649
Copper-transporting ATPase (*copA*)	SE0126	20	14	0.00377649
ATP-dependent DNA helicase	SE1590	19	12	0.00417491
Multidrug resistance protein	SE0239	16	8	0.00584389
Potassium-transporting ATPase B-chain homologue	NA	6	0	0.00738879
FdhD protein homolog	NA	14	6	0.0124069
Transposase	NA	20	16	0.0216122
Transcription regulator MarR family	SE1837	19	14	0.0220505
Potassium transporting ATPase A-chain Homologue, truncated	NA	11	4	0.0231032
MoaA molybdenum cofactor biosynthesis protein A	SE1841	18	13	0.0353177
Unknown	NA	9	3	0.0400408
Two-component response regulator	SE1969	20	17	0.0491825
Hypothetical protein	SE2152	20	17	0.0491825
Fosfomycin resistance protein (*fosB*)	SE0231	20	17	0.0491825
Hypothetical protein	NA	20	17	0.0491825
Signal recognition particle	SE0910	20	17	0.0491825

[a] Gene annotation according to reference 22 (modified).

which is in contrast to the assumption that an osmoprotective factor is required during life on the skin rather than during infection. Moreover, recent findings indicate that the arsenic resistance operon might also be involved in osmoprotection (16). These data suggest interesting, previously unexpected roles for antibiotic resistance genes and osmoprotection factors in *S. epidermidis*.

In conclusion, our study revealed high genetic variability of *S. epidermidis* as a species. We identified several markers for *S. epidermidis* invasiveness, which included proposed virulence factors, confirming the validity of our approach and the role in pathogenesis of these factors. Most importantly, our study also identified genes with unknown function for use as potential novel drug targets. We are presently investigating the contribution to virulence of several of the detected putative novel virulence factors.

TABLE 3. Distribution of gene SE2251 in strains from Huashan hospital, Shanghai, determined by analytical PCR

Source	No. of strains with gene SE2251[a]	Total no. of strains analyzed[b]
Catheter	1	6
Blood	1	13
Urine	10	12
Peritoneal dialysis fluid	2	2
Cerebrospinal fluid	2	2
EPS (encapsulating peritoneal sclerosis)	1	2
Bile	1	1

[a] Number of skin samples, 26; number of infectious samples, 28.
[b] Number of skin samples, 52; number of infectious samples, 38.

We thank Donald J. Meyer, Partek Inc., for statistical analysis, Nevine El Solh, Institut Pasteur, for *S. epidermidis* strains, and Frank R. DeLeo for critically reading the manuscript.

REFERENCES

1. **Broer, S., G. Ji, A. Broer, and S. Silver.** 1993. Arsenic efflux governed by the arsenic resistance determinant of *Staphylococcus aureus* plasmid pI258. J. Bacteriol. **175:**3480–3485.
2. **Cabanes, D., P. Dehoux, O. Dussurget, L. Frangeul, and P. Cossart.** 2002. Surface proteins and the pathogenic potential of *Listeria monocytogenes*. Trends Microbiol. **10:**238–245.
3. **Coulter, S. N., W. R. Schwan, E. Y. Ng, M. H. Langhorne, H. D. Ritchie, S. Westbrock-Wadman, W. O. Hufnagle, K. R. Folger, A. S. Bayer, and C. K. Stover.** 1998. *Staphylococcus aureus* genetic loci impacting growth and survival in multiple infection environments. Mol. Microbiol. **30:**393–404.
4. **El Zoeiby, A., F. Sanschagrin, and R. C. Levesque.** 2003. Structure and function of the Mur enzymes: development of novel inhibitors. Mol. Microbiol. **47:**1–12.
5. **Galdbart, J. O., A. Morvan, N. Desplaces, and N. el Solh.** 1999. Phenotypic and genomic variation among *Staphylococcus epidermidis* strains infecting joint prostheses. J. Clin. Microbiol. **37:**1306–1312.
6. **Heilmann, C., O. Schweitzer, C. Gerke, N. Vanittanakom, D. Mack, and F. Götz.** 1996. Molecular basis of intercellular adhesion in the biofilm-forming *Staphylococcus epidermidis*. Mol. Microbiol. **20:**1083–1091.
7. **Kozitskaya, S., S. H. Cho, K. Dietrich, R. Marre, K. Naber, and W. Ziebuhr.** 2004. The bacterial insertion sequence element IS*256* occurs preferentially in nosocomial *Staphylococcus epidermidis* isolates: association with biofilm formation and resistance to aminoglycosides. Infect. Immun. **72:**1210–1215.
8. **Lang, S., M. A. Livesley, P. A. Lambert, W. A. Littler, and T. S. Elliott.** 2000. Identification of a novel antigen from *Staphylococcus epidermidis*. FEMS Immunol. Med. Microbiol. **29:**213–220.
9. **Lowe, A. M., D. T. Beattie, and R. L. Deresiewicz.** 1998. Identification of novel staphylococcal virulence genes by in vivo expression technology. Mol. Microbiol. **27:**967–976.
10. **Mazmanian, S. K., G. Liu, H. Ton-That, and O. Schneewind.** 1999. *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. Science **285:**760–763.
11. **McCrea, K. W., O. Hartford, S. Davis, D. N. Eidhin, G. Lina, P. Speziale, T. J. Foster, and M. Hook.** 2000. The serine-aspartate repeat (Sdr) protein family in *Staphylococcus epidermidis*. Microbiology **146**(Pt. 7)**:**1535–1546.
12. **Nilsson, M., L. Frykberg, J. I. Flock, L. Pei, M. Lindberg, and B. Guss.** 1998. A fibrinogen-binding protein of *Staphylococcus epidermidis*. Infect. Immun. **66:**2666–2673.

13. **Patti, J. M., B. L. Allen, M. J. McGavin, and M. Hook.** 1994. MSCRAMM-mediated adherence of microorganisms to host tissues. Annu. Rev. Microbiol. **48:**585–617.
14. **Pei, L., and J. I. Flock.** 2001. Lack of *fbe*, the gene for a fibrinogen-binding protein from *Staphylococcus epidermidis*, reduces its adherence to fibrinogen coated surfaces. Microb. Pathog. **31:**185–193.
15. **Rupp, M. E., P. D. Fey, C. Heilmann, and F. Götz.** 2001. Characterization of the importance of *Staphylococcus epidermidis* autolysin and polysaccharide intercellular adhesin in the pathogenesis of intravascular catheter-associated infection in a rat model. J. Infect. Dis. **183:**1038–1042.
16. **Scybert, S., R. Pechous, S. Sitthisak, M. J. Nadakavukaren, B. J. Wilkinson, and R. K. Jayaswal.** 2003. NaCl-sensitive mutant of *Staphylococcus aureus* has a Tn917-lacZ insertion in its *ars* operon. FEMS Microbiol. Lett. **222:** 171–176.
17. **Sobral, R. G., A. M. Ludovice, S. Gardete, K. Tabei, H. De Lencastre, and A. Tomasz.** 2003. Normally functioning *murF* is essential for the optimal expression of methicillin resistance in *Staphylococcus aureus*. Microb. Drug Resist. **9:**231–241.
18. **Vuong, C., and M. Otto.** 2002. *Staphylococcus epidermidis* infections. Microbes Infect. **4:**481–489.
19. **Witke, C., and F. Götz.** 1995. Cloning and nucleotide sequence of the signal peptidase II (*lsp*)-gene from *Staphylococcus carnosus*. FEMS Microbiol. Lett. **126:**233–239.
20. **Wolfgang, M. C., B. R. Kulasekara, X. Liang, D. Boyd, K. Wu, Q. Yang, C. G. Miyada, and S. Lory.** 2003. Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*. Proc. Natl. Acad. Sci. USA **100:**8484–8489.
21. **Yao, Y., D. E. Sturdevant, and M. Otto.** 2005. Genome-wide analysis of gene expression in *Staphylococcus epidermidis* biofilms: insights into *S. epidermidis* biofilm pathophysiology and the role of phenol-soluble modulins in biofilm formation. J. Infect. Dis.**191:**289–298.
22. **Zhang, Y. Q., S. X. Ren, H. L. Li, Y. X. Wang, G. Fu, J. Yang, Z. Q. Qin, Y. G. Miao, W. Y. Wang, R. S. Chen, Y. Shen, Z. Chen, Z. H. Yuan, G. P. Zhao, D. Qu, A. Danchin, and Y. M. Wen.** 2003. Genome-based analysis of virulence genes in a non-biofilm-forming *Staphylococcus epidermidis* strain (ATCC 12228). Mol. Microbiol. **49:**1577–1593.
23. **Ziebuhr, W., V. Krimmer, S. Rachid, I. Lossner, F. Götz, and J. Hacker.** 1999. A novel mechanism of phase variation of virulence in *Staphylococcus epidermidis*: evidence for control of the polysaccharide intercellular adhesin synthesis by alternating insertion and excision of the insertion sequence element IS*256*. Mol. Microbiol. **32:**345–356.

Editor: F. C. Fang