International Journal of Infectious Diseases 91 (2020) 22–31

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

Collagen adhesion gene is associated with bloodstream infections caused by methicillin-resistant *Staphylococcus aureus*

Yasunori Iwata[a,b,*,1], Kenji Satou[g,1], Kengo Furuichi[m], Ikuko Yoneda[b], Takuhiro Matsumura[d], Masahiro Yutani[d], Yukako Fujinaga[d], Atsushi Hase[g], Hidetoshi Morita[h], Toshiko Ohta[i], Yasuko Senda[a], Yukiko Sakai-Takemori[a], Taizo Wada[a], Shinichi Fujita[a], Taito Miyake[b,f], Haruka Yasuda[e], Norihiko Sakai[b,c], Shinji Kitajima[b,f], Tadashi Toyama[b,f], Yasuyuki Shinozaki[b,f], Akihiro Sagara[b,f], Taro Miyagawa[b,f], Akinori Hara[b,f], Miho Shimizu[b,f], Yasutaka Kamikawa[b,f], Kazuho Ikeo[l], Shigeyuki Shichino[j,k], Satoshi Ueha[j,k], Takuya Nakajima[j,k], Kouji Matsushima[j,k], Shuichi Kaneko[f], Takashi Wada[b,e]

[a] *Division of Infection Control, Kanazawa University, Kanazawa, Japan*
[b] *Division of Nephrology, Kanazawa University, Kanazawa, Japan*
[c] *Division of Blood Purification, Kanazawa University, Kanazawa, Japan*
[d] *Department of Bacteriology, Kanazawa University, Kanazawa, Japan*
[e] *Department of Nephrology and Laboratory Medicine, Kanazawa University, Kanazawa, Japan*
[f] *Department of Disease Control and Homeostasis, Kanazawa University, Kanazawa, Japan*
[g] *Faculty of Electrical and Computer Engineering, Kanazawa University, Kanazawa, Japan*
[h] *Graduate School of Environmental and Life Science, Okayama University, Okayama, Japan*
[i] *University of Tsukuba, Tsukuba, Japan*
[j] *Department of Molecular Preventive Medicine, University of Tokyo, Tokyo, Japan*
[k] *Division of Molecular Regulation of Inflammatory and Immune Diseases, Research Institute of Biomedical Sciences, Tokyo University of Science, Noda, Japan*
[l] *Laboratory of DNA Data Analysis, National Institute of Genetics, Shizuoka, Japan*
[m] *Division of Nephrology, Kanazawa Medical University School of Medicine, Ishikawa, Japan*

ARTICLE INFO

Article history:
Received 20 August 2019
Received in revised form 1 November 2019
Accepted 1 November 2019

Keywords:
MRSA
Bloodstream infection
Cna
Whole genome sequencing

ABSTRACT

Objectives: Methicillin-resistant *Staphylococcus aureus* (MRSA) causes hospital- and community-acquired infections. It is not clear whether genetic characteristics of the bacteria contribute to disease pathogenesis in MRSA infection. We hypothesized that whole genome analysis of MRSA strains could reveal the key gene loci and/or the gene mutations that affect clinical manifestations of MRSA infection.
Methods: Whole genome sequences (WGS) of MRSA of 154 strains were analyzed with respect to clinical manifestations and data. Further, we evaluated the association between clinical manifestations in MRSA infection and genomic information.
Results: WGS revealed gene mutations that correlated with clinical manifestations of MRSA infection. Moreover, 12 mutations were selected as important mutations by Random Forest analysis. Cluster analysis revealed strains associated with a high frequency of bloodstream infection (BSI). Twenty seven out of 34 strains in this cluster caused BSI. These strains were all positive for collagen adhesion gene (*cna*) and have mutations in the locus, those were selected by Random Forest analysis. Univariate and multivariate analysis revealed that these gene mutations were the predictor for the incidence of BSI. Interestingly, mutant CNA protein showed lower attachment ability to collagen, suggesting that the mutant protein might contribute to the dissemination of bacteria.
Conclusions: These findings suggest that the bacterial genotype affects the clinical characteristics of MRSA infection.
© 2019 The Author(s). Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

* Corresponding author at: Division of Infection Control, Division of Nephrology, Kanazawa University, 13-1 Takara-machi, Kanazawa 920-8641, Japan.
E-mail address: iwatay@staff.kanazawa-u.ac.jp (Y. Iwata).
[1] These authors contributed equally to this manuscript.

https://doi.org/10.1016/j.ijid.2019.11.003
1201-9712/© 2019 The Author(s). Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Methicillin-resistant *Staphylococcus aureus* (MRSA) is a common pathogenic bacterium that causes nosocomial and community-acquired infections. An annual report of a nationwide survey in Japan revealed MRSA was detected at 100% of the hospitals in 2017 (JANIS, 2019). Moreover, MRSA infection can cause severe illness not only in immunocompromised patients but also in healthy people. Although anti-MRSA therapeutic agents are clinically available, MRSA infection can be life threatening.

Whole genome sequencing (WGS) provides robust genetic identification of bacteria (Köser et al., 2012a). WGS reveals single-nucleotide polymorphisms (SNPs) and insertions/deletions (indels), allowing phylogeny tracing of MRSA outbreaks (Harris et al., 2013; Harris et al., 2010; Köser et al., 2012b). Moreover, a recent study raised the possibility that genome sequence could predict the virulence in MRSA (Laabei et al., 2014).

We hypothesized that whole genome analysis could reveal key gene loci and/or mutations that have an impact on the clinical manifestations of MRSA infection. To explore this possibility, we performed whole genome analysis of MRSA with a focus on genes related to the toxins, drug resistance, and adhesion/biofilm formation of the bacteria. We compared the clinical features of infection in clusters based on the similarity of SNPs and indels.

Subjects and methods

Patient characteristics

The baseline clinical data were collected before MRSA infection or at the same time as isolation of the bacteria. The laboratory data during MRSA infection were collected from the patients who were treated with anti-MRSA antibiotics. Performance status (PS) was graded with the score developed by the Eastern Cooperative Oncology Group. Vancomycin, teicoplanin, arbekacin, daptomycin, and linezolid were considered to be anti-MRSA antibiotics.

Bacterial isolation and identification

The MRSA strains were isolated from patients who were admitted to Kanazawa University Hospital from 1998 to 2015. MRSA was identified according to colony morphology on mannitol salt sugar with egg yolk agar, Gram staining, and detection of coagulase production by the Staphyslide test (Eiken Chemical Co., Ltd., Tokyo, Japan).

Biofilm formation assay

Biofilm formation was investigated using the microtiter plate biofilm assay as described previously (Cassat et al., 2014). Briefly, diluted bacteria were incubated overnight on 96-well plates (Corning, NY) or collagen-I coated 96-well plates (Corning, NY). The attached bacteria were fixed in 100% ethanol and stained with crystal violet after aspiration of bacterial cultures. The absorbance of the cultures was read at a wavelength of 595 nm.

DNA sequencing and analysis

DNA was isolated from each MRSA strain using the Nucleo Spin Tissue kit (MACHREY-NAGEL, Düren) according to the manufacturer's protocol. WGS data from Hiseq sequencing system (Illumina, San Diego, CA, USA) was generated from 1 μg of DNA isolated from the bacteria. Extracted DNA was fragmented using Covaris S220 (Covaris, Woburn, MA, USA). Fragmented DNA was purified using Agencourt AMPure XP (Beckman Coulter, Inc. Brea, CA, USA). DNA libraries were constructed using the TruSeq DNA HT Sample Prep Kit (Illumina) according to the manufacturer's protocol. Constructed DNA libraries were analyzed using an Agilent Bioanalyzer with the High Sensitivity Kit (Agilent, Santa Clara, CA, USA). The sequenced data were aligned to a reference isolate (HO 5096 0412) using Bowtie 2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) and SAM tools (http://samtools.sourceforge.net/). Then, SNPs and indels were detected with VarScan (http://varscan.sourceforge.net/). Clustering was performed using Cluster 3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm) based on the similarity of whole genome sequences. The domains of amino acid replacements were detected using the database InterPro (http://www.ebi.ac.uk/interpro/). All the sequence data are available from the DDBJ database (accession number DRA006521.)

Preparation of CNA Protein

Preparation of CNA proteins with/without mutation were prepared as described before (Matsumura et al., 2015; Zong et al., 2005). Briefly, purified DNA from *MRSA* strains with/without can mutation were used as a template for the amplification of DNA encoding CNA by PCR. The amplified DNAs were inserted into pET-28b(+) (Merck Millipore). Recombinant CNAs were expressed in *E. coli* strain Rosetta2 (Merck Millipore) and purified using HisTrap HP (GE Healthcare). The purity of CNA proteins were confirmed by SDS–PAGE (Supplementary Figure S2). Primer sets are listed in Supplementary Table S1.

Binding assay (enzyme-linked immunosorbent assay, ELISA)

The 96-well plates (IWAKI) were coated with type I collagen (1.0 μg/well) from calf skin (Sigma-Aldrich) or rat tail for overnight at 4 °C. The wells were washed with phosphate-buffered saline (PBS) containing 0.05% Tween20 (Sigma-Aldrich) (PBS-T) and blocked with 1.0% bovine serum albumin (BSA, Sigma-Aldrich)/PBS-T for overnight at 4 °C. CNA proteins were added to a well and incubated for 1 h at 37 °C. After washing, anti-His tag antibody (MBL) was added and incubated for 1 h at 37 °C. After washing, anti–mouse IgG conjugated with HRP (Jackson ImmunoResearch) was added and incubated for 1 h at 37 °C. Plates were washed again and then incubated with substrate solution (ABTS, Roche) for 20 min at 37 °C, and absorbance values at Abs_{405} were measured. To analyze the binding of CNA with type I collagen from pig tendon, we used collagen coated plate (IWAKI). CNA proteins were added to plates, and the binding CNA proteins were detected with anti-His tag antibody and anti–mouse IgG conjugated with HRP.

RNA isolation and quantitative real-time PCR

Total RNA was isolated from fresh overnight cultures of MRSA strains using the High Pure RNA Isolation Kit (Roche Diagnostics, Tokyo, Japan). Quantitative real-time polymerase chain reaction (PCR) based on SYBR Green fluorescence (Bio-Rad, Tokyo, Japan) was performed on the Villa 7 Real-Time PCR System (Thermo Fisher Scientific, Tokyo, Japan). The sequences of primers are listed in Supplementary Table S1. Data were analysed by the delta-delta Ct method. Primer sets are listed in Supplementary Table S1.

Statistics

Data were presented as the mean ± SEM as determined with StatView software. The data of patients and bacterial strains were compared with chi-squared test, Fisher exact test, unpaired t-test, Kruskal–Wallis test as appropriate. Random Forest analysis was performed with Random Forest package in *R* (https://www.r-project.org/). Univariate and multivariate logistic regression

Table 1
The correlation between gene mutation and disease onset in MRSA infection.

SNPs	Gene symbpl/protein name	Correlation
● 2591486:T->C	*fnbA*	0.446870375
● 2072101:A->T	transposase	0.44327757
● 557941:C->T	*rpoC*	0.416631156
722116:C->G	*fruA*	0.416631156
1590871:A->G	*gnd*	0.416631156
2100500:A->G	*groEL*	0.416631156
● 2810892:T->C	*cna*	0.416631156
● 1943824:A->T	transposase	0.409960309
26304:C->T	*yycG*	0.408723827
423385:T->G	type I restriction modification system modification protein	0.408723827
423720:C->T	type I restriction modification system modification protein	0.408723827
424708:A->G	type I restriction modification system modification protein	0.408723827
457961:C->T	LysR-family regulatory protein	0.408723827
457964:C->T	LysR-family regulatory protein	0.408723827
457967:T->C	LysR-family regulatory protein	0.408723827
583105:C->T	*sdrD*	0.408723827
● 586186:A->G	*sdrD*	0.408723827
719268:C->A	DeoR family regulatory protein	0.408723827
722155:G->A	*fruA*	0.408723827
735366:A->G	para-aminobenzoate synthase component	0.408723827
823298:GCTA->G	*clfA*	0.408723827
823834:A->T	*clfA*	0.408723827
963063:C->T	*prfC*	0.408723827
1744911:C->T	*lysP*	0.408723827
1808046:C->T	*harA*	0.408723827
1808146:T->G	*harA*	0.408723827
1874261:G->C	Crc-like protein	0.408723827
1996064:T->C	*putP*	0.408723827
1996067:T->C	*putP*	0.408723827
2100284:C->T	*groEL*	0.408723827
2100287:C->T	*groEL*	0.408723827
2100293:G->A	*groEL*	0.408723827
2100506:T->A	*groEL*	0.408723827
2161825:A->G	*kdpD*	0.408723827
2268562:T->G	FecCD transport family protein	0.408723827
2278721:A->T	*opuD2*	0.408723827
2278730:A->C	*opuD2*	0.408723827
2291107:T->C	*lacR*	0.408723827
2297985:T->C	*hysA2*	0.408723827
2297988:T->C	*hysA2*	0.408723827
2365778:T->C	*ureD*	0.408723827
2460923:C->T	*ureD*	0.408723827
2480960:T->C	*nasD*	0.408723827
2481005:A->G	*nasD*	0.408723827
2512275:A->G	ABC transporter ATP-binding protein	0.408723827
2592104:A->T	*fnbA*	0.408723827
2652300:T->C	TetR family regulatory protein	0.408723827
2738381:T->A	*clfB*	0.408723827
2738384:A->G	*clfB*	0.408723827
2738397:G->T	*clfB*	0.408723827
2738399:A->T	*clfB*	0.408723827
2739308:C->T	*clfB*	0.408723827
2591489:T->A	*fnbA*	0.408723827
2072210:T->C	transposase	0.401250282

Blue columns indicate adhesion/biofilm related genes and green column indicates resistant gene. Closed circle indicates the mutations, those were raised as important mutations by Random Forest analysis.

analysis were performed to identify the hazard ratios for bloodstream infection (BSI).

Ethics statement

Informed consent was obtained from all subjects. All data analyzed were anonymized. All experiments were performed in accordance with approved guidelines of Kanazawa University. This study was conducted with the approval of the ethics committee of Kanazawa University (IRB approval number:1522).

Results

Adhesion and biofilm-related gene mutations are associated with disease manifestation in MRSA infection

WGS identified 136686 indels/SNPs from 154 isolated MRSA. We excluded the mutations in the loci, those encoded putative, hypothetical or no protein. Totally, 39712 mutations were analyzed in this study. We explored the relationship between indels/SNPs and disease manifestations in MRSA infection. Each gene mutation was evaluated for correlations with infectious disease onset/bacterial colonization. The correlation coefficients for these manifestations were then added. The top 54 mutations, which showed fifth highest correlation coefficient, are listed in Table 1. The gene mutation 2591486:T->C in the *fnbA* locus showed the highest correlation with infectious disease onset. There were 13 indels/SNPs detected in loci related to adhesion and biofilm formation, while one SNP was detected in the drug-resistance locus. None were in a toxin-related locus. These data indicate that mutations in the genes related to adhesion/biofilm formation may contribute to disease manifestation in MRSA infection.

Twelve mutations were raised as important mutations for infectious disease onset of MRSA with Random Forest model

The top 1000 mutations, which showed high correlation with clinical manifestation, were analyzed with Random Forest regression model. The importance of 30 out of 1000 mutations were predicted in the model (Figure 1a). To better characterize this approach, the appearance in the top 30 important mutations was analyzed as a frequency by repeating the performance 20 times. Two hundred mutations were raised as candidates of important mutations in this method (Figure 1b). Among them, 12 mutations showed over 50 % of appearance in the top 30 important mutations with Random Forest. Moreover, 6 of 12 mutations were also listed in the top 54 mutations; those showed high correlation (Table 1). Thus, these 12 were thought to be important for infectious disease onset of MRSA.

Cluster analysis of 154 MRSA strains

WGS was used to arrange 154 strains into clusters according to SNPs and indels. Three strains had patterns of SNPs and indels that differed from the other 151 strains (Figure 2). We then examined 12 mutations; those were selected by Random Forest analysis in the cluster. Almost all the strains in cluster C possessed these 12 mutations. Of these mutations, 2810892:T → C at collagen adhesion gene (*cna*) locus, showed different expression patterns in each cluster. Interestingly, the presence/absence patterns of the *cna* locus matched the WGS cluster pattern. Therefore, we designated clusters positive for the *cna* locus as clusters A (n = 18) and C (n = 34) and those negative for the *cna* locus as clusters B (n = 76) and D (n = 23) (Figure 2). ITS genotype (Fujita et al., 2005), the culture specimen and the wards of the patients in each cluster are described in Supplementary Figure S1.

Baseline characteristics of patients in each cluster

To compare host factors between clusters, the basal characteristics of patients were evaluated in each cluster as shown in Table 2. The patients in cluster A were younger than those in the other clusters (age [years], cluster A, 44.7 ± 7.3; B, 63.5 ± 2.3; C, 62.6 ± 3.3; D, 66.3 ± 3.0). Performance status was higher in cluster A than the other clusters (cluster A, 0.6 ± 0.3; B, 2.0 ± 0.3; C, 1.7 ± 0.3; D, 2.4 ± 0.4). Serum albumin levels in cluster B patients were lower than those in other clusters (cluster A, 3.5 ± 0.2; B, 2.9 ± 0.1; C, 3.2 ± 0.2; D, 3.1 ± 0.2). The other parameters did not differ between the clusters.

Clinical features of patients in each cluster

To evaluate the clinical impact of the genetic differences between strains, patient clinical manifestations were compared

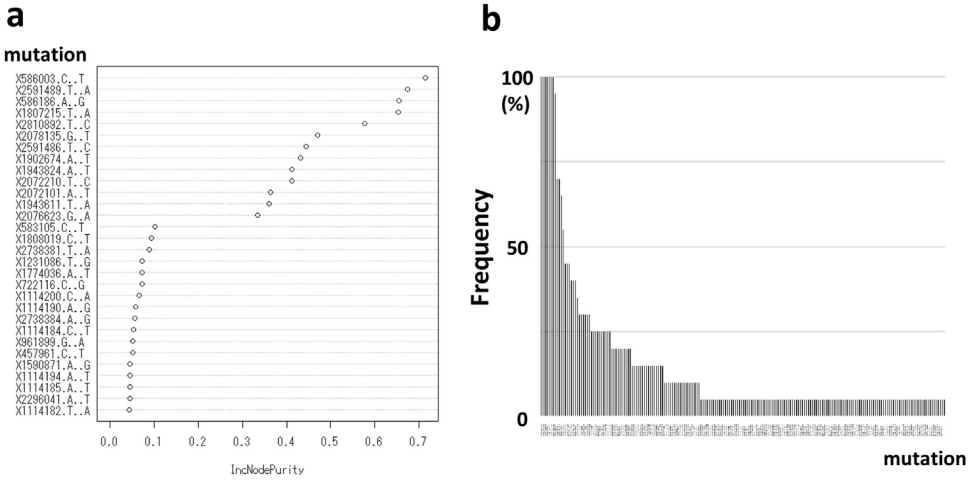

Figure 1. Random Forest analysis revealed 12 important mutations for infectious disease onset.
Top 1000 mutations, which showed high correlation with clinical manifestation, were analyzed with Random Forest regression model. The importance of 30 mutations was predicted in the model (a). To better characterize this approach, the appearance in the top 30 important mutations was analyzed as a frequency by repeating the performance 20 times. Two hundred mutations were raised as candidates of important mutations in this method (b). Among them, 12 mutations showed over 50 % of appearance in the top 30 important mutations with Random Forest. Thus, these 12 of 1000 mutations were thought to be important for infectious disease onset of MRSA.

Figure 2. Phylogenetic clustering of MRSA.
Whole genome sequencing divided 154 strains into some clusters according to SNPs and indels. Three strains had different patterns of SNPs and indels from 151 strains. We then examined 12 mutations, selected by Random Forest analysis, in the cluster. Almost all the strains in cluster C possessed these 12 mutations. Of these mutations, 2810892: T → C at *cna* locus showed a different expression patterns in each cluster. The present/absent patterns of *cna* locus matched the cluster pattern of the whole genome sequence. Thus, we designated the clusters that were positive *cna* locus as clusters A and C, and the clusters that were negative *cna* locus as clusters B and D. Gray box shows the strains that caused BSI.

between clusters. Notably, the rate of BSI was higher in cluster C patients than that in the other cluster patients (Figures 2 and 3). In particular, catheter infection was the leading cause of BSI in cluster C patients. The percentages of patients needing treatment with anti-MRSA antibiotics and of deceased patients were similar in each cluster (Figure 3).

Laboratory data during infection in each cluster

To compare the severity of infection, we analyzed the laboratory data acquired during therapy with anti-MRSA antibiotics from all surviving patients. The platelet counts were lower in cluster C patients during the therapy (platelet count $\times 10^4/\mu L$:

Table 2
The character of patients before MRSA infection.

	A	B	C	D
Male/female	10/8	50/26	21/13	17/6
Age	44.7 ± 7.3^{abc}	63.5 ± 2.3	62.6 ± 3.3	66.3 ± 3.0
Ptients on dialysis	1	2	2	0
Performance status	0.6 ± 0.3^{abc}	2.0 ± 0.3	1.7 ± 0.3	2.4 ± 0.4
Platelets ($\times 10^4/\mu L$)	22.8 ± 2.9	25.8 ± 1.4	20.4 ± 1.4	19.7 ± 2.1
Creatinine (mg/dL)	0.8 ± 0.2	1.1 ± 0.2	1.1 ± 0.2	0.7 ± 0.1
T. bilirubin (mg/dL)	0.5 ± 0.1	0.8 ± 0.1	2.4 ± 0.4	1.9 ± 0.7
T. protein (mg/dL)	6.5 ± 0.2	6.2 ± 0.1	6.4 ± 0.3	6.4 ± 0.2
Albumin (g/dL)	3.5 ± 0.2^{a}	2.9 ± 0.1	3.2 ± 0.1	3.1 ± 0.2

Superscripts mean significance ($p < 0.05$), a: A vs. B, b: A vs. C, c: A vs. D. T; total.

the difference in the *cna* locus between clusters A and C might be one reason for clinical difference. Thus, we next investigated bacterial characteristics, focusing on the *cna* locus. WGS revealed 17 specific SNPs at the *cna* locus, all of which were present in cluster C strains but not in cluster A strains (Figure 5a). Ten out of 17 SNPs had changes in the predictable cna amino acid sequence. Moreover, 7 amino acid replacements were located in the functional domain, and some were accumulated in close site (Figure 5b).

CV. catheter and cna mutations were predictors for BSI

Next, we explored whether these mutations at *cna* loci were a predictor for the incidence of BSI. Univariate logistic analysis

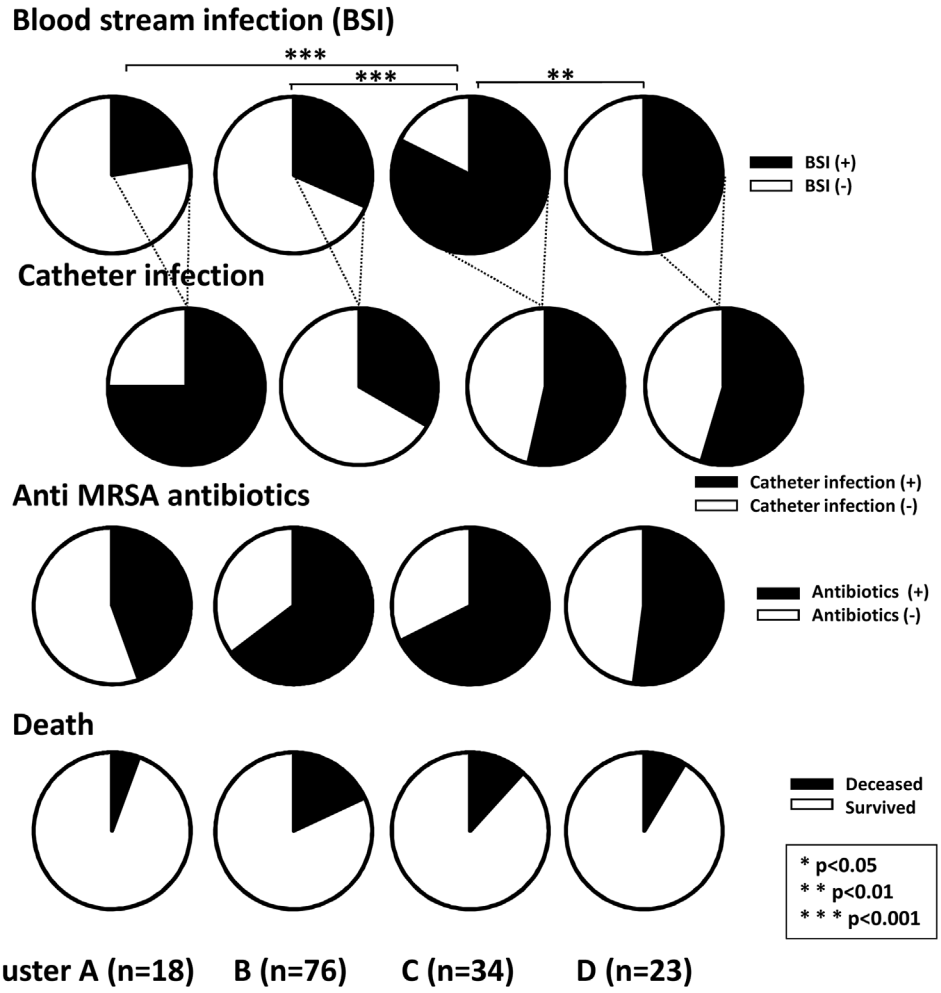

Figure 3. Clinical features of the patients in each cluster.
Clinical manifestations of the patients were investigated in each cluster. Notably, the rate of BSI was higher in cluster C patients than those in cluster A, B and D. Especially, catheter infection is the leading cause of BSI in patients of cluster C. The percentage of the patients needing treatment with anti MRSA agents and the rate of all-cause of death were similar among the clusters.

cluster C, before infection 21.7 ± 1.7; after infection, 16.0 ± 1.7). The incidence of AKI and the change of total bilirubin were similar in the 4 groups (Figure 4).

Bacterial characteristics in each cluster

While both cluster A and C strains had the *cna* locus, they exhibited different clinical manifestations. We hypothesized that

showed that CV catheter insertion (hazard ratio [HR], 9.35; 95% confidence interval [95% CI], 3.948–22.116; $p < 0.0001$), platelet counts (HR, 0.96; 95% CI, 0.924–0.995; $p = 0.021$) and mutation at *cna* loci (HR, 7.7; 95% CI, 2.730–21.693; $p < 0.0001$) were predictors of BSI. Multivariate analysis revealed that CV catheter insertion (HR, 5.01; 95% CI, 1.710–15.080; $p = 0.003$) and mutation at *cna* loci (HR, 5.1; 95% CI, 1.442–19.620; $p = 0.009$) were high risk for BSI (Table 3).

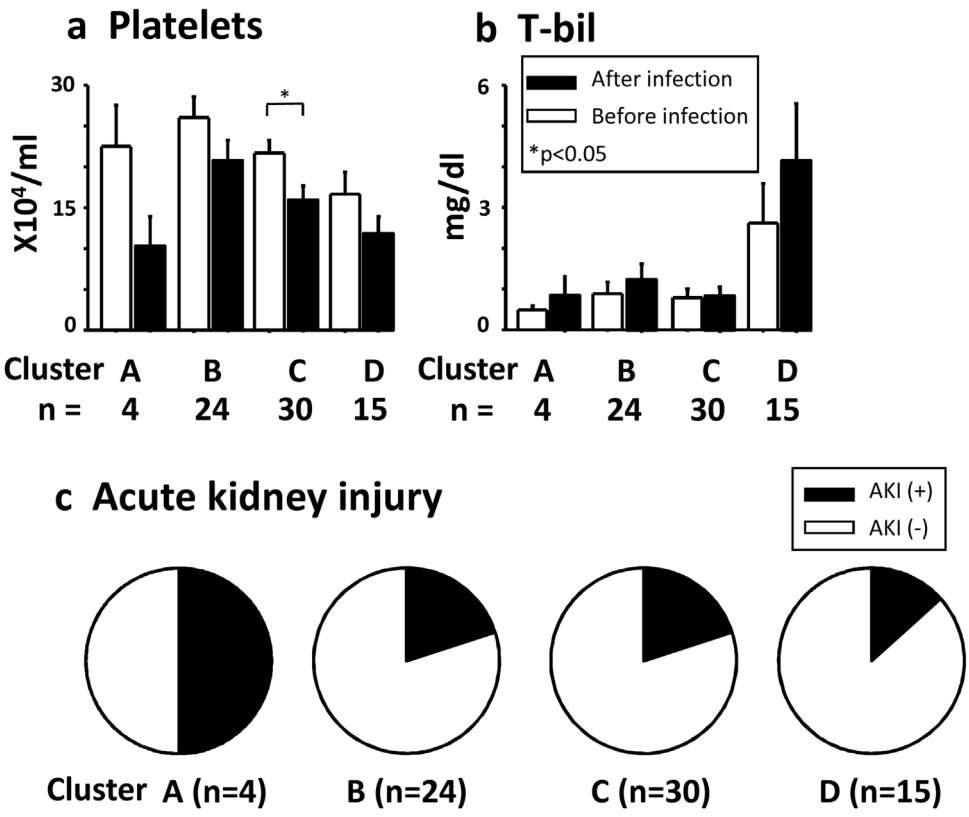

Figure 4. Laboratory data during infection in each cluster.
To compare the severity of infection, we analyzed the laboratory data acquired during therapy with anti MRSA agents from all surviving patients. The platelet counts were lower in cluster C patients during the therapy (platelet count $\times 10^4/\mu L$: cluster C, before infection 21.7 ± 1.7; after infection, 16.0 ± 1.7). The incidence of AKI and the change of total bilirubin were similar in the 4 groups (Figure 3). Cr; creatinine, T-bil; total bilirubin.

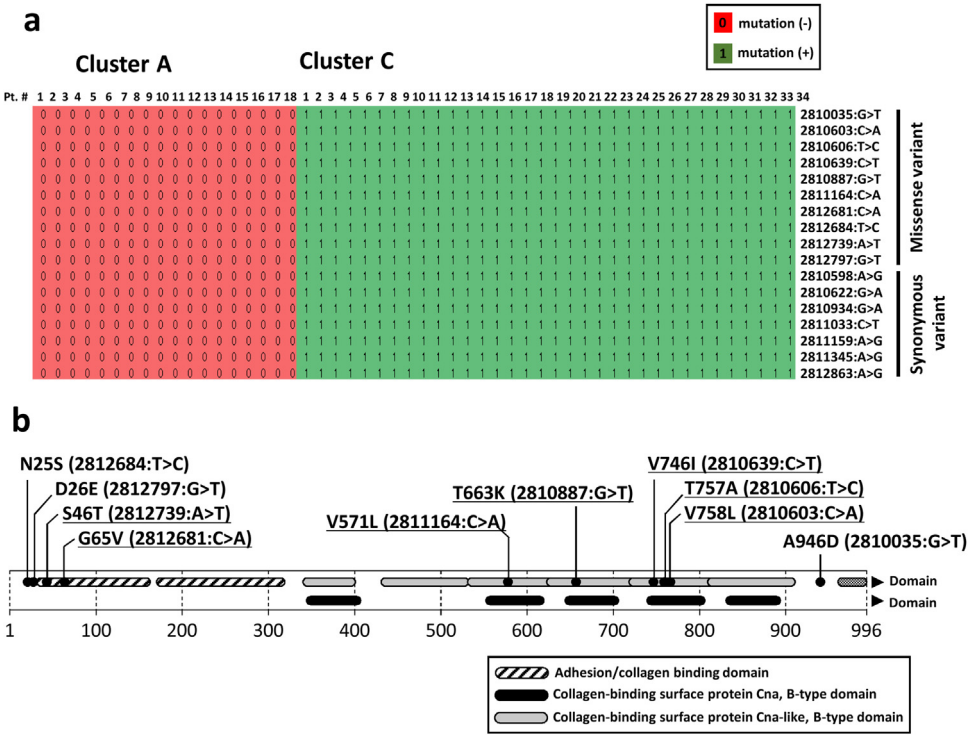

Figure 5. The characteristics of *cna* locus in cluster A and C.
WGS revealed 17 specific SNPs on *cna* locus that exist in all of the cluster C strains (a). Ten out of 17 SNPs changed the predictable amino acid sequence of the *cna* protein. Moreover, 7 amino acid replacements were located on the functional domain and some of them were observed at close site in cna (b). The domains of amino acid replacements were detected with InterPro (http://www.ebi.ac.uk/interpro/).

Table 3
Univariate and multivariate logistic analysis of predictors associated with BSI.

	Univariate analysis			Multivariate analysis		
	Hazard ratio	95% C.I.	*P*-value	Hazard ratio	95% C.I.	*P*-value
Age (yrs)	1.01	0.997–1.035	0.084			
Peformance status	1.19	0.970–1.457	0.094			
Diabetes	1.66	0.652–4.249	0.283			
CV catheter	9.35	3.948–22.116	<0.0001	5.01	1.710–15.080	0.003
Platelets ($\times 10^4/\mu L$)	0.96	0.924–0.995	0.021	0.95	0.900–1.014	0.1
Creatinine (mg/dL)	0.95	0.731–1.247	0.731			
T. bilirubin (mg/dL)	1.1	0.901–1.347	0.322			
Albumin (g/dL)	0.72	0.380–1.347	0.296			
Cna mutation	7.7	2.730–21.693	<0.0001	5.1	1.442–19.260	0.009

Abbreviations are CV; central venous, T; total, C.I.; confidence interval.

Mutant CNA protein showed lower attachment ability

To evaluate the functional impact of mutant CNA, we performed an attachment assay of CNA protein to collagen. Surprisingly, mutant CNA showed lower attachment ability to the collagen from calf skin, pig tendon and rat tail as compared to CNA without mutation (Figure 6a). Even at the different concentrations, the attached CNA to collagen was less as compared to CNA without mutation (Figure 6b). Recent data showed the biofilm development consists of some stages such as bacterial multiplication and exodus (Moormeier et al., 2014). In particular, the exodus of bacteria allows for further dissemination of bacterial cells, resulting in probable severe infection. Thus, we assessed the exodus stages during biofilm formation on the strains with/without *cna* mutation. The strains with *cna* mutant exhibited a higher ratio of detached/attached bacteria in the early stage of biofilm formation, indicating more bacterial dissemination as compared to the strains without *cna* mutation (Figure 6c). Nevertheless, the biofilm formation showed similar levels between the strains at 24 h. The ratio of detached/attached bacteria in the early stage of biofilm formation was also higher in the strains with *cna* mutant on collagen-I coated plate (Figure 6d). To elucidate whether CNA contributes to the early stage of biofilm development, the gene expressions of the adhesion/biofilm formation related molecules were evaluated. The expression level of *cna* was higher at 4 to 8 h of biofilm formation. Interestingly, the expression levels of *cna* is lower in *cna* mutant strain as compared to non-mutant strain (Figure 6e). Some other molecules, such as *icaA, icaB, icaC, icaR, eno,* and *fnbA* also showed increased expression at the early stage of biofilm development. While the expression levels of some genes were decreased at 12 and 24 h in the non-mutant strain, the expression levels of genes such as *icaA, icaB, icaC, icaD, icaR, sarS, sdrD, dltD* and *fnbA,* remained at high levels in the mutant strain (Figure 6e). Next, the gene mutation in the functional domain of these molecules was analyzed to see if gene mutation could contribute to changing the function of the molecules. The frequency of mutation in the functional domain is relatively high in the functional domain in the *cna* locus (Figure 6f).

Discussion

We observed that the mutations in the *cna* locus have an impact on the clinical manifestations of MRSA infection. Gene mutations in loci associated with adhesion and biofilm formation correlated with disease onset/bacterial colonization in MRSA infection. The profiles of SNPs and indels divided the MRSA strains into several clusters that matched the pattern of the presence/absence of the *cna* locus. The frequency of BSI, especially catheter infection, was higher in the cluster in which the *cna* locus was present. Interestingly, several specific gene mutations and amino acid substitutions were detected in the *cna* gene in cluster C strains. Moreover, mutation in the *cna* locus as well as CV catheter insertion were the predictors for the incidence of BSI. Interestingly, mutant CNA protein showed lower attachment ability to collagen, suggesting that mutant CNA might contribute to the exodus of bacteria and the dissemination of infection.

WGS provides sufficient discriminatory genomic information to identify bacterial lineages. Several studies have reported the use of WGS to trace the path of MRSA hospital outbreaks (Harris et al., 2013; Harris et al., 2010; Köser et al., 2012b). In addition, WGS has been used to elucidate the genetic pathogenesis of resistance to antibiotics (Alm et al., 2014; Leopold et al., 2014; Iwata et al., 2017). Another group has reported that WGS revealed the insertion of gene loci from a close relative MRSA strain, resulting in the de-repression of cytotoxin expression and increased virulence (Benson et al., 2014). In our study, whole genome analysis revealed that specific strains are more likely to cause BSI. Since there was no difference in host factors except the younger age with high PS in cluster A and lower albumin levels in cluster B, these differences in clinical manifestation could be attributable to MRSA genotype differences.

Several groups have explored the relationship between MRSA genotype characteristics and BSI. Xiong et al. (2009a) reported that persistent MRSA bacteremia isolates had a higher frequency of several drug resistance genes. van der Mee-Marquet et al. (2004) demonstrated clonal diffusion of the 3 major French MRSA clones among patients with BSI using pulsed-field gel electrophoresis. Other studies determined the genotypes of bacteremia isolates from patients with liver disease (Török et al., 2014) and kidney disease (McNicholas et al., 2011). While these studies showed the genotypic diversity of MRSA in BSI, limited data are available regarding whether certain genotypic lineages are more likely to cause BSI.

Recent studies employed logistic prediction analysis, such as Random Forest, to model the genetic risk to some diseases (Zhen et al., 2017; Chuang and Kuo, 2017), including MRSA infection (Laabei et al., 2014). Twelve mutations of top 1000 mutations, which showed high correlation with MRSA infection, were raised as important mutations in this study. Of these mutations, 2810892: T → C at *cna* locus, showed different expression patterns in each cluster. Interestingly, the presence/absence patterns of the *cna* locus matched the WGS cluster pattern.

The cluster C strain that was associated with a higher frequency of BSI had the *cna* locus. Cna is a cell-wall-anchored protein that contributes to the adhesion to collagen-rich tissue (Foster et al., 2014). A recent study revealed that cna inhibits the activation of the complement system by binding to C1q (Kang et al., 2013). An association between the *cna* locus and BSI has been reported in several studies. Persistent bacteremia isolates exhibited a higher frequency of bearing cna than that in resolving bacteremia isolates (Xiong et al., 2009b). Clusters that were positive for the *cna* locus were observed in 50% of BSI cases (Yu et al., 2012) and 71% of

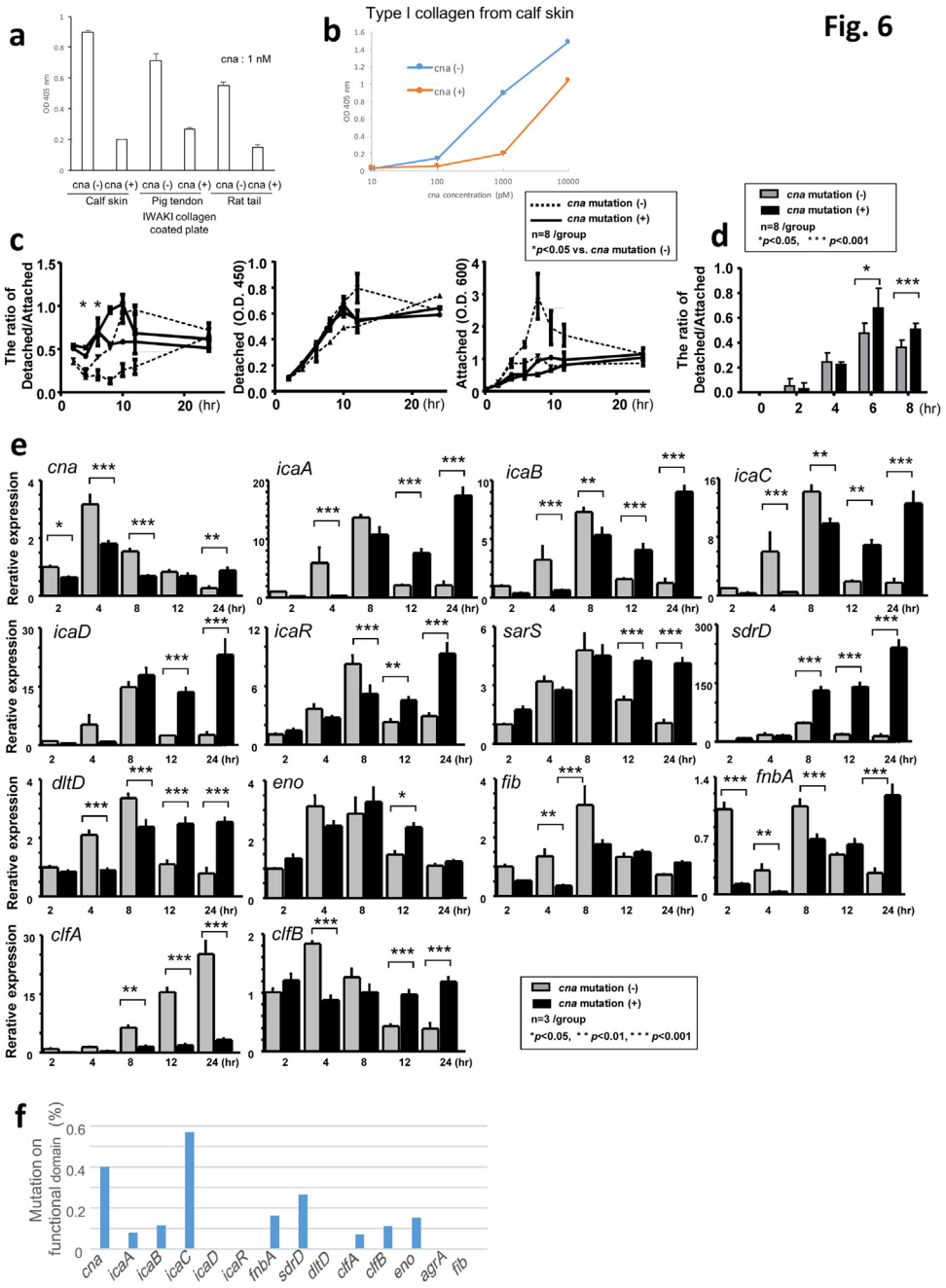

Figure 6. The functional assay of CNA protein.
The attachment ability of CNA proteins with/without mutation were evaluated using various types of collagen. Mutant CNA showed lower attachment ability to collagen as compared to without mutation (a). Even at different concentrations, the attached CNA to collagen was less as compared to CNA without mutation (b). The attachment/ biofilm assay showed that the exodus on early stage of biofilm development is increased in *cna* mutant strains as compared to the strains without *cna* mutation on non-coated plate (c) as well as collagen coated plate (d). The gene expression of *cna* was up-regulated at an early time point during the biofilm formation. Moreover, the expression levels of *cna* were lower in *cna* mutant strain as compared to non-mutant strain (d). The frequency of mutations on the functional domain in each locus (e).

catheter-associated BSI cases (Fowler et al., 2005) caused by *S. aureus*. Blomfeldt A. et al. reported S. aureus clonal complex 30 genotype was associated with mortality in patients with BSI (Blomfeldt et al., 2016). Moreover, S. aureus clonal complex 30 has been reported to possess *cna* locus (Nienaber et al., 2011). These reports showed the association of *cna* locus with high incidence of BSI with/without high mortality.

Although the cluster A strain has the *cna* locus, the frequency of BSI was lower in cluster A than that in cluster C patients. This observation could result from differences of the *cna* locus between bacteria in clusters A and C. To explore the possibility, we next investigated bacterial characteristics related to the *cna* locus. The cluster C strain had several specific gene mutations and showed amino acid replacements that would affect the predicted structure of the cna protein. These results indicate that the specific gene mutations may change the cna protein structure, thereby modifying its function in the cluster C strain.

Univariate and multivariate analysis also revealed that *cna* mutation was an independent predictor for the incidence of BSI. Surprisingly, hazard ratio of *cna* mutations is almost same as that of CV catheter insertion. This result may suggest that the MRSA strain carrying *cna* mutations has a high impact for BSI. However, more

analysis, including prospective cohort study, is needed to clarify the risk for BSI of the strain.

Interestingly, the attachment ability to collagen is lower in mutant CNA as compared to that without mutation. Recent study revealed that detachment in the early phase of biofilm formation causes the dissemination of bacteria and allows for possible expansion of infection (Moormeier et al., 2014). In this regard, the strains with mutant *cna* showed a higher ratio of detached and attached bacterial numbers as compared to normal strain, especially in the early phase of biofilm formation. Moreover, the gene expression of *cna* was increased at the early time point of biofilm formation, which suggests that cna might contribute to the exodus of bacteria. In addition, the expression level of *cna* was lower in *cna* mutant strain as compared to non-mutant strain, which would indicate that not only the function of CNA but also the low expression levels of CNA might decrease the attachment ability in the *cna* mutant strain. Some other molecules, such as *icaA, icaR, fnbA, sdrD, AgrA and SarS* also showed increased expression at the early stage of biofilm development. However, the frequency of gene mutation on functional domain is higher in cna as compared to the other molecules, suggesting that the mutation of CNA might be associated with the exodus of bacteria in the early phase of biofilm formation. However, more precise studies are needed to clarify the contribution of gene mutation to biofilm formation, especially exodus of bacteria.

In summary, WGS revealed a cluster of MRSA patients with a high frequency of BSI and infection by strains with a mutation in the *cna* locus. These findings suggest that the bacterial genotype affects the clinical characteristics of MRSA infection. In particular, patients with strains carrying *cna* mutations could have an increased risk of BSI.

Funding source

This work was supported by the Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research on Innovative Areas program (Inflammation Cellular Sociology, 17H06394, YI, NS, KF and TW) and JSPS KAKENHI (18K08426, YI). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest

The authors have declared that no conflict of interest exists.

Ethical approval

Informed consent was obtained from all subjects. All data analyzed were anonymized. All experiments were performed in accordance with approved guidelines of Kanazawa University. This study was conducted with the approval of the ethics committee of Kanazawa University (IRB approval number:1522).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ijid.2019.11.003.

References

Alm RA, McLaughlin RE, Kos VN, et al. Analysis of *Staphylococcus aureus* clinical isolates with reduced susceptibility to ceftaroline: an epidemiological and structural perspective. J Antimicrob Chemother 2014;69:2065–75.

Benson MA, Ohneck EA, Ryan C, et al. Evolution of hypervirulence by a MRSA clone through acquisition of a transposable element. Mol Microbiol 2014;93:664–81.

Blomfeldt A, Eskesen AN, Aamot HV, et al. Population-based epidemiology of *Staphylococcus aureus* bloodstream infection: clonal complex 30 genotype is associated with mortality. Eur J Clin Microbiol Infect Dis 2016;35:803–13.

Cassat JE, Smeltzer MS, Lee CY. Investigation of biofilm formation in clinical isolates of *Staphylococcus aureus*. Methods Mol Biol 2014;1085:195–211.

Chuang LC, Kuo PH. Building a genetic risk model for bipolar disorder from genome-wide association data with random forest algorithm. Sci Rep 2017;7:39943.

Foster TJ, Geoghegan JA, Ganesh VK, et al. Adhesion, invasion and evasion: the many functions of the surface proteins of *Staphylococcus aureus*. Nat Rev Microbiol 2014;12:49–62.

Fowler Jr VG, Justice A, Moore C, et al. Risk factors for hematogenous complications of intravascular catheter-associated *Staphylococcus aureus* bacteremia. Clin Infect Dis 2005;40:695–703.

Fujita S, Senda Y, Iwagami T, et al. Rapid identification of staphylococcal strains from positive-testing blood culture bottles by internal transcribed spacer PCR followed by microchip gel electrophoresis. J Clin Microbiol 2005;43:1149–57.

Harris SR, Feil EJ, Holden MT, et al. Evolution of MRSA during hospital transmission and intercontinental spread. Science 2010;327:469–74.

Harris SR, Cartwright EJ, Török ME, et al. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. Lancet Infect Dis 2013;13:130–6.

Iwata Y, Satou K, Tsuzuku H, Furuichi K, Senda Y, Sakai-Takemori Y, et al. Down-regulation of the two-component system and cell-wall biosynthesis-related genes was associated with the reversion to daptomycin susceptibility in daptomycin non-susceptible methicillin-resistant *Staphylococcus aureus*. Eur J Clin Microbiol Infect Dis 2017;36(October (10)):1839–45.

JANIS. Japan Nosocomial Infections Surveillance (JANIS). 2019. https://janis.mhlw.go.jp/report/open_report/2017/3/1/ken_Open_Report_201700.pdf.

Kang M, Ko YP, Liang X, et al. Collagen-binding microbial surface components recognizing adhesive matrix molecule (MSCRAMM) of Gram-positive bacteria inhibit complement activation via the classical pathway. J Biol Chem 2013;288:20520–31.

Köser CU, Holden MT, Ellington MJ, et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N Engl J Med 2012a;366:2267–75.

Köser CU, Holden MT, Ellington MJ, et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N Engl J Med 2012b;366:2267–75.

Laabei M, Recker M, Rudkin JK. Predicting the virulence of MRSA from its genome sequence. Genome Res 2014;24:839–49.

Leopold SR, Goering RV, Witten A, et al. Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. J Clin Microbiol 2014;52:2365–70.

Matsumura T, Sugawara Y, Yutani M, Amatsu S, Yagita H, Kohda T, et al. Botulinum toxin A complex exploits intestinal M cells to enter the host and exert neurotoxicity. Nat Commun 2015;6(February):6255.

McNicholas S, Shore AC, Coleman DC, et al. DNA microarray genotyping and virulence and antimicrobial resistance gene profiling of methicillin-resistant *Staphylococcus aureus* bloodstream isolates from renal patients. J Clin Microbiol 2011;49:4349–51.

Moormeier DE, Bose JL, Horswill AR, Bayles KW. Temporal and stochastic control of *Staphylococcus aureus* biofilm development. MBio 2014;5(October (5))e01341-14.

Nienaber JJ, Sharma Kuinkel BK, Clarke-Pearson M, et al. Methicillin-susceptible *Staphylococcus aureus* endocarditis isolates are associated with clonal complex 30 genotype and a distinct repertoire of enterotoxins and adhesins. J Infect Dis 2011;204:704–13.

Török ME, Harris SR, Cartwright EJ, et al. Zero tolerance for healthcare-associated MRSA bacteraemia: is it realistic?. J Antimicrob Chemother 2014;69:2238–45.

van der Mee-Marquet N, Domelier AS, Girard N, Bloodstream Infection Study Group of the Relais d'Hygiène du Centre, et al. Epidemiology and typing of *Staphylococcus aureus* strains isolated from bloodstream infections. J Clin Microbiol 2004;42:5650–7.

Xiong YQ, Fowler VG, Yeaman MR, et al. Phenotypic and genotypic characteristics of persistent methicillin-resistant *Staphylococcus aureus* bacteremia in vitro and in an experimental endocarditis model. J Infect Dis 2009a;199:201–8.

Xiong YQ, Fowler VG, Yeaman MR, et al. Phenotypic and genotypic characteristics of persistent methicillin-resistant *Staphylococcus aureus* bacteremia in vitro and in an experimental endocarditis model. J Infect Dis 2009b;199:201–8.

Yu F, Li T, Huang X, et al. Virulence gene profiling and molecular characterization of hospital-acquired *Staphylococcus aureus* isolates associated with bloodstream infection. Diagn Microbiol Infect Dis 2012;74:363–8.

Zhen Y, Xinghui Z, Chao W, et al. Several microRNAs could predict survival in patients with hepatitis B-related liver cancer. Sci Rep 2017;7:45195.

Zong Y, Xu Y, Liang X, Keene DR, Höök A, Gurusiddappa S, et al. A 'Collagen Hug' model for *Staphylococcus aureus* CNA binding to collagen. EMBO J 2005;24 (December (24)):4224–36.