Mechanisms of disease

Genome and virulence determinants of high virulence community-acquired MRSA

Tadashi Baba, Fumihiko Takeuchi, Makoto Kuroda, Harumi Yuzawa, Ken-ichi Aoki, Akio Oguchi, Yoshimi Nagai, Natsuko Iwama, Kazuyuki Asano, Timothy Naimi, Hiroko Kuroda, Longzhu Cui, Kenji Yamamoto, Keiichi Hiramatsu

Summary

Background A new type of meticillin-resistant *Staphylococcus aureus* (MRSA), designated community-acquired MRSA, is becoming increasingly noticeable in the community, some strains of which cause fatal infections in otherwise healthy individuals. By contrast with hospital-acquired MRSA, community-acquired MRSA is more susceptible to non β-lactam antibiotics. We investigated the high virulence potential of certain strains of this bacterium.

Methods We ascertained the whole genome sequence of MW2, a strain of community-acquired MRSA, by shotgun cloning and sequencing. MW2 caused fatal septicaemia and septic arthritis in a 16-month-old girl in North Dakota, USA, in 1998. The genome of this strain was compared with those of hospital-acquired MRSA strains, including N315 and Mu50.

Findings Meticillin resistance gene (*mecA*) in MW2 was carried by a novel allelic form (type IVa) of staphylococcal cassette chromosome *mec* (SCC*mec*), by contrast with type II in N315 and Mu50. Type IVa SCC*mec* did not carry any of the multiple antibiotic resistance genes reported in type II SCC*mec*. By contrast, 19 additional virulence genes were recorded in the MW2 genome. All but two of these virulence genes were noted in four of the seven genomic islands of MW2.

Interpretation MW2 carried a range of virulence and resistance genes that was distinct from those displayed on the chromosomes of extant *S aureus* strains. Most genes were carried by specific allelic forms of genomic islands in the MW2 chromosome. The combination of allelic forms of genomic islands is the genetic basis that determines the pathogenicity of medically important phenotypes of *S aureus*, including those of community-acquired MRSA strains.

Lancet 2002; **359:** 1819–27
See Commentary page 1791

Department of Bacteriology, Juntendo University, 2-1-1 Hongo, Bunkyo-ku, Tokyo 113-8421, Japan (T Baba PhD, M Kuroda PhD, H Yuzawa PhD, H Kuroda MS, L Cui MD, Prof K Hiramatsu MD)**; National Institute of Technology and Evaluation, Tokyo** (A Oguchi ASc, Y Nagai ASc, K Aoki BSc, K Asano MSc)**; New Energy and Industrial Technology Development Organisation, Tokyo** (N Iwama MSc)**; Research Institute, International Medical Centre of Japan, Tokyo** (F Takeuchi PhD, K Yamamoto MD)**; Organisation for Pharmaceutical Safety and Research, Tokyo** (F Takeuchi)**; and Minnesota Department of Health, Minneapolis, MN, USA** (T Naimi MD)

Correspondence to: Prof Keiichi Hiramatsu
(e-mail: hiram@med.juntendo.ac.jp)

Introduction

Meticillin-resistant *Staphylococcus aureus* (MRSA), which arose in the 1960s, typically causes infections in inpatients, who have risk factors associated with health care. In the past 5 years however, MRSA infections have been described in the general population. These infections arise in the community, and thus affected people lack traditional risk factors such as recent admittance, surgery, or long-term residence in care facilities. Community-acquired MRSA infections can cause serious and even fatal infections in otherwise healthy hosts.[1,2]

Community-acquired MRSA typically affects children and young adults, and it causes a range of infections similar to those caused by community-acquired meticillin-susceptible *S aureus* (MSSA).[3] Furthermore, these infections are more likely to be susceptible to many antimicrobial classes and to have different pulsed-field gel electrophoresis subtypes than isolates that are associated with health care. There is an ominous sign that MRSA strains circulating beyond nosocomial settings are replacing community-acquired MSSA as the flora of healthy human beings in some communities.

Recently, we established the whole genome sequence of two MRSA strains, N315 and Mu50, both of which are strains associated with health care.[4] Two further hospital-acquired MRSA strains have been sequenced by others (COL, E-MRSA-16 [strain 252]; see Methods). However, the genome of any community-acquired strain with high virulence has not been analysed. In this study, we sequenced the whole genome of the community-acquired MRSA strain MW2, and aimed to investigate the genetic basis for this strain's high virulence by comparison of its genome with those of MRSA strains that are associated with health care.

Methods

MW2 is a typical community-acquired strain of MRSA, which was isolated in 1998 in North Dakota, USA.[1] MW2 is so-called because strains of similar pulsed-field gel electrophoresis genotypes are reported widely in mid-western USA. MW2 caused fatal septicaemia and septic arthritis in a 16-month-old American-Indian girl who had no risk factors associated with health care. She died within 2 h of arrival at hospital. The MW2 strain is susceptible to many antibiotic classes, apart from β-lactam antibiotics, to which the strain shows a typical heterogeneous-type resistance.

We established the whole genome sequence as described previously,[4] with the exception that we used genomic DNA fragments of 1·5–3·0 kb instead of 1·0–2·2 kb to make the plasmid library. We based the final genome sequence of MW2 on about 64 000 sequences. We have entered the whole genome sequence of MW2 primarily in the DNA Database of Japan, with accession numbers AP004822–AP004831 (chromosome), AP004832 (plasmid), and BA000033 (11 entries and contig).

GLOSSARY

BLA OPERON
Usually present on a plasmid in *S aureus*; its function is β-lactam-induced production of penicillinase. The operon is composed of three genes: *blaZ*, encoding a penicillin-hydrolysing enzyme (penicillinase), and its transcription regulator genes, *blaR1* and *blaI*. When penicillin is in the environment, membrane-bound signal transducer protein BlaR1 recognises it and transmits the signal to the cytoplasm. Then, the repressor protein BlaI, which binds near to the promoter of *blaZ* preventing its transcription, is cleaved off, leaving *blaZ* to be transcribed efficiently to produce penicillinase.

GENOMIC ISLAND ALLOTYPING
Clinically important properties, drug resistance, and the pathogenic potential of *S aureus* strains are mostly established by the set of genomic islands that strain possesses. Each island is found in several alternative forms with different sets of virulence or resistance genes among *S aureus* strains—ie, it exists in several allelic forms. Since different allelic forms of islands encode different properties, identifying the allelic set of islands that the strain carries gives information on the overall properties of the strain, including its pathogenic potential and the pattern of antibiotic susceptibilities.

MEC OPERON
Composed of *mecA*, *mecR1,* and *mecI*. The latter two are gene homologues of *blaR1* and *blaI,* respectively. MecR1 is the membrane-bound signal transducer recognising β-lactam antibiotics, and BlaI is a repressor protein. Therefore, the same sequence of events arises as with *bla* operon. The only difference is that *mecA* encodes cell-wall synthesis enzyme PBP2′ instead of penicillinase encoded by *blaZ*: β-lactams induce meticillin resistance in this case.

PANTON-VALENTINE LEUKOCIDIN
Member of the leukocidin group of bacterial toxins that kill leucocytes by creation of pores in the cell membrane. Apart from *S aureus* Panton-Valentine leukocidin, the group includes *Pasteurella haemolytica* leukotoxin, *Actinobacillus actimycetemcomitans* leukotoxin, *Listeria monocytogenes* listeriolysin, *Escherichia coli* haemolysin, and *Fusobacterium necrophorum* leukotoxin.

RESTRICTION-MODIFICATION SYSTEM
Bacteria protect themselves from invasion of foreign DNA (eg, bacteriophage DNA) by cleaving DNA (restriction) at specific recognition sites. The restriction enzyme HsdR (HSD is the acronym of host-specificity determinant) cleaves the specific DNA site that HsdS (S stands for specificity) recognises. To protect the bacteria's own DNA from restriction, DNA methylase, HsdM, and HsdS together modify the specific site by methylation and methylated recognition sites are not cleaved by HsdR. When the *hsdM* gene encoding HsdM is present in the genomic island, the island is considered to be protected from spontaneous loss from the chromosome, because such cells may not survive the self-digestion of the chromosome that occurs when *hsdM* gene is lost.

STAPHYLOCOCCAL CASSETTE CHROMOSOME (SCC)
A unique family of mobile genetic elements found on the chromosomes of *Staphylococcal* spp. In *S aureus* chromosome, it is always integrated at a fixed site near the origin of replication. If SCC carries meticillin resistance gene (*mecA*), it is called SCC*mec*, the integration of which in the *S aureus* chromosome generates meticillin-resistant *S aureus* (MRSA).

SUPERANTIGENS
A variety of molecules that share the ability to activate large populations of T lymphocytes through co-ligation between major histocompatibility complex (MHC) class II molecules on antigen-presenting cells and the variable portion of the T-cell antigen receptor β chain; the pattern of Vβ activation is specific for each of these superantigens. They include staphylococcal enterotoxins, exotoxins, and toxic-shock-syndrome toxin.

SYNONYMOUS CODON USAGE
Each aminoacid of protein is coded by a set of three nucleotides (a codon) in the gene. Some codons, for example GCT, GCC, GCA, GCG, code the same aminoacid, alanine. Such codons are called synonymous codons. The proportion of usage of synonymous codons to encode each aminoacid is known to vary for different proteins and for different species.

We initially identified open reading frames (*orfs*)—sections of the genome likely to encode proteins—with GLIMMER[5] and rbsfinder[6] software; we ignored *orfs* comprising 30 or fewer codons. We reviewed predicted *orfs* individually with the GAMBLER program (Xanagen, Kawasaki, Japan), and mutually overlapping *orfs* were identified and removed in some cases. We searched a non-redundant protein database with ORFs—peptides encoded by *orfs*—with BLAST software.[7] We identified transfer RNA and tmRNA[8] genes by tRNAscan-SE[9] and with a procedure available online.[10]

The best-hit entries identified by BLASTP[7] searches with e-values of $1{\cdot}0\times10^{-5}$ or less were classified into taxonomic groups in accordance with the NCBI (National Center for Biotechnology Information) taxonomy database.[11] We excluded hits against entries for Staphylococcus genus from this analysis.

We calculated bias in SYNONYMOUS CODON USAGE against averages of all *orfs* longer than 150 bp (B_{all}) and of ribosomal proteins (B_{rp}) and GC content at the third codon position (GC3), as described previously.[4] With these values, we defined putative highly expressed genes and alien genes as *orfs* with $B_{rp}<0{\cdot}35$ and $B_{all}>0{\cdot}30$, and $B_{rp}>0{\cdot}52$ and $B_{all}>0{\cdot}345$, respectively. We also regarded *orfs* as possible alien genes when their GC3 values differed more than 1·5 SD from the average of all *orfs* longer than 150 bp, which corresponds to GC3<0·16 or GC3>0·29.

To compare the MW2 genome with those of other *S aureus* strains, we first processed homologous regions of the two chromosomes with bl2seq (blast for two sequences),[12] and we visualised them for gross comparison. To do this, we divided a 2·8-Mb genome into three 1-Mb fragments, and blast homology was calculated for the nine pairs of fragments. Then, we processed blast results by a script written in Perl (http://www.perl.com) and visualised with Mathematica (Wolfram Research, Champaign, IL, USA). Similarity between the two virtual chromosomes—which are obtained by removal of all genomic islands and the mobile genetic elements IS*1181*, Tn*554*, and Tn*5801* from the chromosomes—was assessed by AVID whole genome alignment as the proportion of matched nucleotides in the entire aligned nucleotides including gaps.[13] To verify some of the ideas derived from the comparison, we compared the genome sequence of MW2 with those of NCTC8325,[14] E-MRSA-16 (strain 252),[15] and MRSA strain COL.[16]

To assess spontaneous excision of genomic islands, LightCycler PCR was done with the LightCycler-FastStart DNA Master SYBR Green I kit (Roche Diagnostics, Mannheim, Germany) with specific primers designed in and around the genomic islands. To calculate frequency of the chromosome from which the island in question was precisely cut out, and to ascertain copy number of closed circular DNA formed by the excised island, we did quantitative PCR amplification, and then established the relative copy numbers of each PCR product with LightCycler software.

Role of the funding source
The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Strains used for comparison
N315 is a MRSA strain named after Nagasaki University Hospital, Japan, isolate number 315. Mu50 is a vancomycin-resistant MRSA (VRSA) isolate from the pus of a postoperative wound. NCTC8325 is a registered strain of the National Collection of Type Cultures and has been widely used as laboratory *S aureus* strains. E-MRSA-16 (strain 252) is an epidemic MRSA from an outbreak in 1991–92 in the UK. The classification of number 16 was

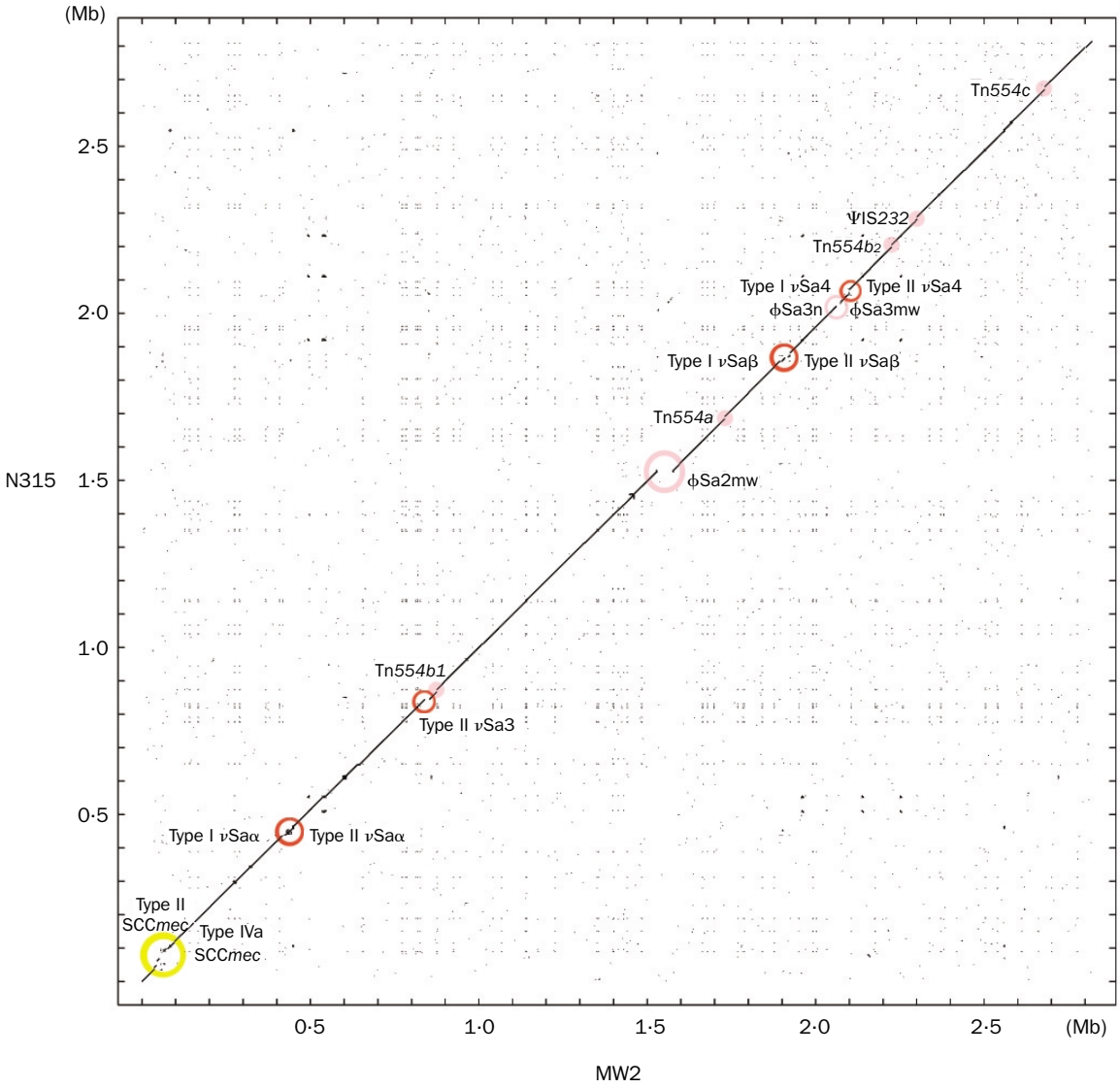

Figure 1: **Seven genomic islands as major differences between N315 and MW2 chromosomes**
Genomic islands in MW2 chromosome=open circles. Yellow=SCC*mec*; pink=prophages; red=ν elements (genomic islands apart from SCC and prophages). Insertion of mobile elements in N315 genome=closed circle. Name of the allelic form of each genomic island in N315 and MW2 is given above and below the diagonal line, respectively. No allelic genomic island is recorded in N315 for νSa3 and ϕSa2.

established by phage-typing. MRSA COL was isolated in the UK in 1965.

MW2 chromosome compared with N315

Figure 1 shows overall nucleotide sequence similarity of MW2 (abscissa) and N315 (ordinate) chromosomes. The discontinuous regions of the central diagonal line of homology correspond with seven staphylococcal genomic islands that discriminate the main functional properties of the two genomes (see below; figures 1 and 2). Other minor discontinuities correspond with insertion of mobile elements in the N315 genome. The term genomic island is defined as a chromosomal region whose exogenous derivation is inferable. STAPHYLOCOCCAL CASSETTE CHROMOSOME *mec* (SCC*mec*; *mec* stands for meticillin resistance) and prophage are included, together with regions previously called pathogenicity islands. These islands carried genes associated with pathogenicity of *S aureus*. The reason for adoption of the term genomic island instead of pathogenicity island is that we are now finding new islands that are very similar to the previous ones, but no genes associated with pathogenicity are found on them (eg, type I νSa3 and type II νSa4). Well-established transposons and insertion sequences are excluded from genomic island terminology.

Figure 1 also shows many dots forming lattice-like patterns that cover the entire chromosomes of both MW2 and N315. This pattern indicates presence of multiple repeats that are well conserved between the two chromosomes. The core sequence 5′-TGTTGGG-GCCC-3′, designated staphylococcus uniform repeats, is present in most lattice cross-sections. There are 60–85 copies of these repeats in the six *S aureus* chromosomes and 18 copies in a *S epidermidis* chromosome (based on the draft sequence obtained from The Institute for Genomic Research), but are not reported in virus, archaea, or eubacteria other than *Staphylococcus* spp.

Table 1 lists general information about the MW2 genome compared with those of N315 and Mu50. MW2 has six ribosomal RNA operons compared with five in N315 and Mu50 (figure 2). Transposons and insertion sequences are scarcely recorded in the MW2 genome (see also figure 2). A truncated copy of insertion sequence IS*1272* is reported within SCC*mec* at a site downstream of

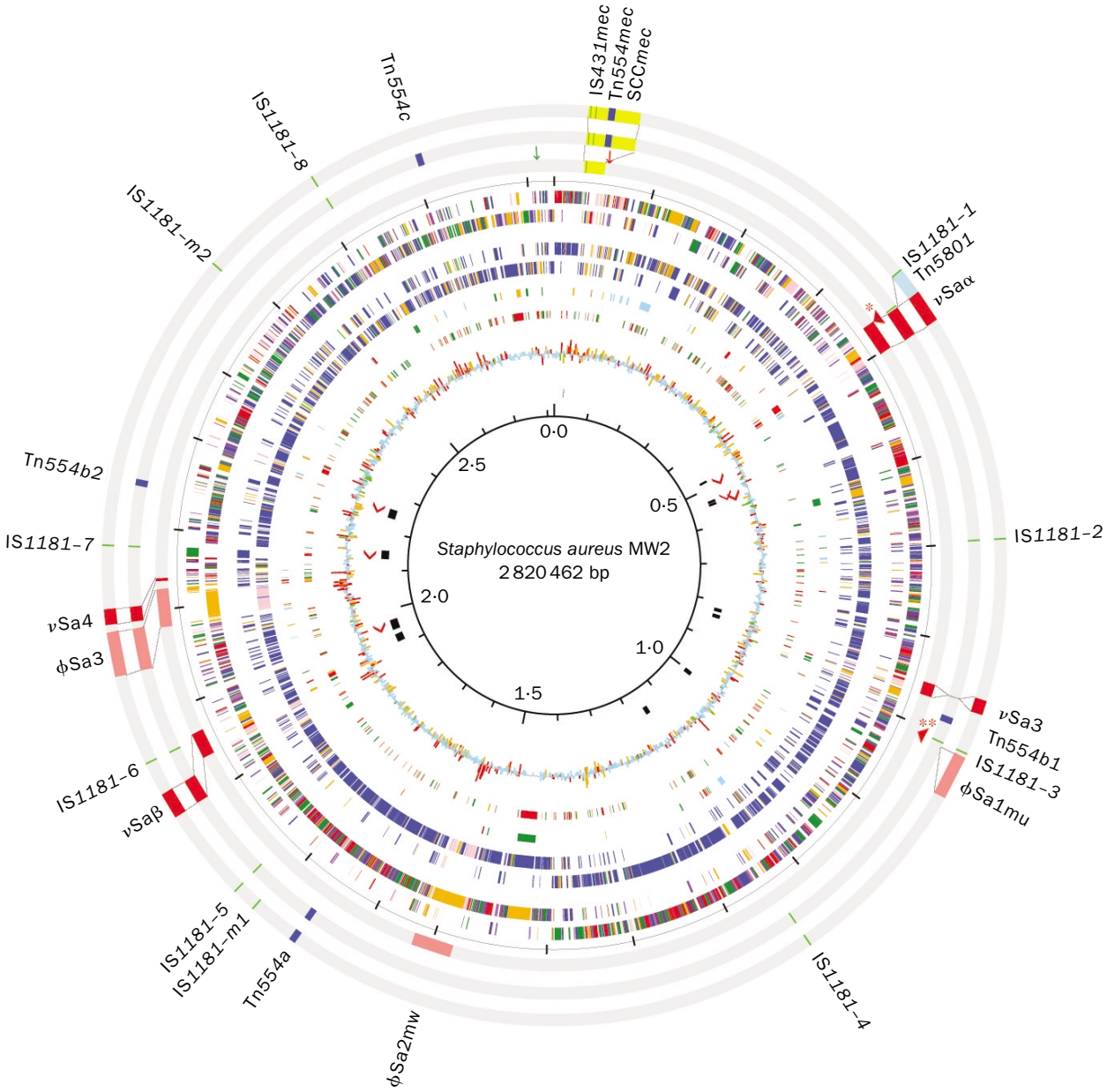

Figure 2: **Functional genomic organisation of MW2 chromosome compared with N315 and Mu50**
From the outside inward: first three circles show distribution of genomic islands (GIs) (yellow=SCC*mec*; red=islands; pink=prophages), transposons (light blue=Tn*5801*; dark blue=Tn*554*), and insertion sequences (green) in the chromosomes of Mu50 (first circle), N315 (second circle), and MW2 (third circle). Red arrowheads indicate the attachment sites for GI families νSa2 (*) and νSa1 (**). Third circle shows location of virulence genes *cna* (green arrow) and *seh* (red arrow). Fourth circle shows every 100th open reading frame (*orf*). Fifth and sixth circles show *orf*s on plus and minus strand, respectively. Blue=cell envelope and cellular processes; green=intermediary metabolism; orange=information pathways; yellow=other functions; magenta=similar to unknown proteins; pink=no similarity. Seventh and eighth circles show taxonomic distribution of BLAST best-hit entries on the plus and minus strand, respectively. Blue=Bacillus/Clostridium group; green=firmicutes (gram-positive eubacteria); pink=viruses/insertion sequences/transposons; orange=archaea/eubacteria/eukaryota; white=no hit or ribosomal and transfer RNAs. Ninth circle shows virulence-associated *orf*s. Red=toxins; green=adhesins; orange=exoproteins; blue=others. Tenth circle shows *orf*s whose BLAST first hit entry is a human pathogen (red=streptococci; orange=*Bacillus anthracis* and *Bacillus cereus*; green=others). Eleventh circle shows GC content at the third codon (GC3) and synonymous codon-usage bias of each *orf*. Green=highly expressed *orf*s; red=putative alien *orf*s; orange=possible alien *orf*s based on GC3 skew; blue=other *orf*s. Size of the coloured bar=deviation of GC3 value of each *orf* from the average. Red arrowhead=rRNA and its orientation. Black bars=locations of tRNAs. Twelfth circle shows nucleotide position in Mb.

a truncated copy of the *mec* regulator gene, *mecR1* (part of the MEC OPERON), and the remains of a Tn*552*-related transposon carrying BLA OPERON (encoding penicillinase and its regulator function) were noted on plasmid pMW2. IS*431mec* is the only apparently functional insertion sequence in the MW2 genome that is always present linked to *mecA* in SCC*mec*.[17]

Evolutionary association of the three strains was assessed by alignment of their virtual chromosomes. Percentage of matched nucleotides between N315 and Mu50 was 99·7%, whereas that between MW2 and N315 and Mu50 was 94·8% and 94·7%, respectively. Data suggest that N315 and Mu50 are closely related to each other, whereas MW2 is relatively distant from N315 and Mu50.

Genetic basis for non-multiresistance of MW2

Absence of multiresistance to non β-lactam antibiotics is a characteristic of community-acquired MRSA strains, a feature also shared by MW2. Consistent with this

	MW2	N315*	Mu50*
Chromosome			
Length of sequence (bp)	2 820 462	2 814 816	2 878 040
G+C content			
Total genome	32·8%	32·8%	32·9%
Protein coding region	33·5%	33·5%	33·6%
RNAs	49·4%	49·5%	48·7%
Non-coding region	27·8%	28·0%	29·3%
Open reading frames			
Percentage coding	83·5%	83·4%	83·8%
Protein coding regions	2632	2593	2714
Ribosomal RNAs			
16S	**6**	5	5
23S	**6**	5	5
5S	**7**	6	6
Transfer RNAs	61	62	60
tmRNAs	1	1	1
Insertion sequences			
IS*1181*	**0**	8	10
IS*431* (on SCC*mec*)	1	2	2
ψ IS*1272* (on SCC*mec*)	1	0	0
Others (remnants)	4	10	11
Transposons			
Tn*554*	**0**	5	2
Tn*5801* (conjugative)	0	0	1
Genomic islands			
Prophages	2	1	2
SCC*mec*	1 (type IVa)	1 (type II)	1 (type II)
νSa islands†	4	3	4
Plasmid			
Length of sequence (bp)	20 654	24 653	25 107
G+C content	28·3%	28·7%	28·9%
Open reading frames			
Percentage coding	75·6%	78·0%	80·9%
Protein coding regions	27	29	34

Prominent features in bold. *Based on the updated nucleotide sequences. †Previously designated pathogenicity islands.

Table 1: **General features of NM2 genome compared with those of N315 and Mu50**

phenotype, the MW2 genome contains no antibiotic resistance genes apart from *blaZ* (encoding penicillinase) on plasmid pMW2 and *mecA* encoding penicillin-binding protein 2′ (PBP2′) found within SCC*mec*.

The structure of SCC*mec* in MW2 is quite distinct from that of hospital-acquired MRSA strains.[18] Figure 3 shows the structure of SCC*mec* in MW2 (designated type IVa) compared with type I SCC*mec* (reported in hospital-acquired MRSA strains discovered in the 1960s, such as NCTC10442 and COL),[19] and type II SCC*mec* (carried by hospital-acquired MRSA strains such as N315 and Mu50).[6] The type IVa allelic form of SCC*mec* of MW2 is much smaller in size than the previously identified types in hospital-acquired MRSA strains. The boundary of SCC*mec* is defined by two functional direct repeats, at which excision of the element takes place. Type IVa SCC*mec* of MW2 comprises two allelic elements—class-B *mec*-gene complex (*mecA* and its regulatory genes) and type-2 *ccrA and B* genes.[19] All SCC*mec* elements are integrated at exactly the same site, and yet each has distinct genetic organisation, presumably as a result of complex series of mutual crossovers. No other resistance genes, apart from *mecA*, were recorded on MW2 SCC*mec*.

It is noteworthy that *seh*, encoding staphylococcal enterotoxin H, and a truncated copy of *seo*, encoding enterotoxin O, are present next to SCC*mec* (figure 3). The *seh* gene exists beyond the right boundary of SCC*mec*, linked to the second copy of the left extremity region of SCC (containing the unique gene *mcrB* encoding a putative restriction enzyme). These genes are only present in MW2: they were not recorded in any of the other five complete genomes. Work in animals and clinical observation suggest that Staphylococcal enterotoxin H is involved in an acute toxic-shock-like syndrome.[20] This enterotoxin is also reported to be produced in disproportionately high amounts compared with other SUPERANTIGENS.[21] The *seh* gene could be associated with a strong promoter, or its location near the origin of replication, at which more gene copy numbers per cell are expected because of the uncoupled nature of cell division and chromosome replication, could be important.

Virulence genes in islands of MW2

MW2 chromosome has four genomic islands besides two prophages and a SCC. One of the islands, νSa3 (ν stands for island), is unique to MW2. The island carries two new allelic forms of enterotoxin genes, *sel2* and *sec4*. It shares exactly the same attachment site (5′-TCCCGCCGTC-TCCAT-3′) with νSa3mu of Mu50 (previously designated SaGIm). The island νSa4 also exists in allelic forms. Three islands, νSa4mw (reported in MW2), νSa4n (N315), and νSa4mu (Mu50), share the same integrase and integration site (5′-GTTTTACATCATTCCCGGCAT-3′) (figure 2, table 2). However, the superantigen genes *sel, sec3,* and *tst* that are carried by νSa4n and νSa4mu (type I) are not recorded in the type II island in MW2. This finding

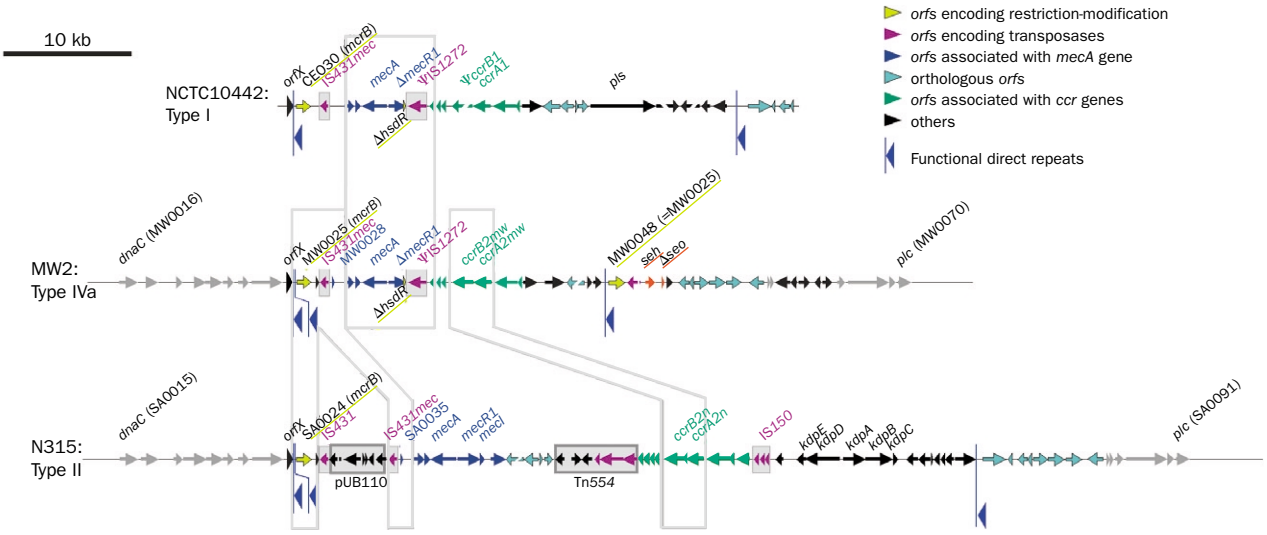

Figure 3: **Comparison of three allelic forms of SCC*mec*, in MW2, N315, and NCTC10442 strains**
Arrows with gaps indicate frameshifted *orfs*.

Genomic island family*	Location of integration†	Allelic islands found in N315, Mu50 and MW2 [strain] (previous name)	Characteristic genes carried by the island	Recombination genes found in the island	Other allelic elements belonging to the family
SCC	34 150	Type II SCC*mec* [N315, Mu50] **Type IVa SCC*mec* [MW2]**	*mecA*, *ermA*, *bleO*, *addD* *mecA*	*ccrA2*, *ccrB2* *ccrA2*, *ccrB2*	Type I, III, IVb SCC*mec*
ϕSa1	887 316	ϕSa1mu [Mu50](ϕMu50B)	(72 orfs)	*int ϕ Sa1*	ϕ 11, ϕ ETA
ϕSa2	1 529 123	**ϕSa2mw [MW2]**	***lukS-PV*, *lukF-PV***	*int ϕ Sa2*	ϕ SLT
ϕSa3	2 046 206	ϕSa3n [N315] (ϕN315)	*sak*, *sep*	*int ϕ Sa3*	ϕ 42
		ϕSa3mu [Mu50] (ϕMu50A)	*sak*, *sea*	*int ϕ Sa3*	
		ϕSa3mw [MW2]	*sak*, *sea*, ***seg2***, ***sek2***	*int ϕ Sa3*	
νSa1	879 794	None		*int ν Sa1*	SaPI1, SaPI3
νSa2	415 781	None		*int ν Sa2*	SaPIbov
νSa3	839 358	Type I νSa3 [Mu50] (SaGlm)	*fhuD*	*int ν Sa3*	None
		Type II νSa3 [MW2]	***ear*, *sel2*, *sec4***	*int ν Sa3*	
νSa4	2 097 809	Type I νSa4 [N315] (SaPln1)	*sel*, *sec3*, *tst*	*int ν Sa4*	None
		Type I νSa4 [Mu50] (SaPlm1)	*sel*, *sec3*, *tst*	*int ν Sa4*	
		Type II νSa4	(4 unknown *orf*s)	*int ν Sa4*‡	
νSaα		Type I νSaα [N315] (SaPln2)	*set6–15*, *lukDE*, *lpl1–9*	*tnp ν Saα1*‡	None
	MW0368 through MW0401	Type I νSaα [Mu50] (SaPlm2)	*set6–15*, *lukDE*, *lpl1–9*	*tnp ν Saα1*‡	
		Type II νSaα [MW2]	***set16–26***, *lukDE*, ***lpl10–14***	None	
νSaβ		Type I νSaβ [N315] (SaPln3)	*spl*, *lukDE*, *seg*, *sen*, *sei*, *sem*, *seo*	*tnp ν Saβ1*‡	None
	MW1745 through MW1770	Type I νSaβ [Mu50] (SaPlm3)	*spl*, *lukDE*, *seg*, *sen*, *sei*, *sem*, *seo*	*tnp ν Saβ1*‡	
		Type II νSaβ [MW2]	*spl*, *lukDE*, ***bsa*** (bacteriocin)	*tnp ν Saβ2*‡	

Bold=features unique to MW2. *int*=integrase gene; *tnp*=transposase gene. *Defined by the location of integration. †Number corresponds with nucleotide position of MW2. ‡Frameshifted or truncated.

Table 2: **Classification of *S aureus* genomic islands**

formally negates involvement of TSST-1 (toxic shock syndrome toxin-1) in high virulence of MW2, which is reported to be the case with an exceptionally virulent hospital-acquired MRSA strain.[22]

Islands νSa3 and νSa4 are spontaneously excised from chromosomes of all the strains tested and form extrachromosomal closed circular DNA. νSa3mw is spontaneously excised from MW2 chromosome at a frequency of 0·0044, but it exists at a 17 times greater copy number than the chromosome from which it was excised, indicating the replication capability of the excised island. Frequency of spontaneous excision of νSa4mw from MW2 chromosome is much lower, which could be because the integrase gene of the νSa4mw is frameshifted by a point mutation (table 2).

The other two islands, νSaα and νSaβ, are reported in all six *S aureus* strains and all clinical strains so far tested (K Hiramatsu, unpublished). νSaα is distinctive because it carries many putative staphylococcal exotoxin (*set*) genes (figure 4).[6,23] However, none of 11 *set* genes of MW2 encodes protein of identical aminoacid sequence to that encoded by 10 *set* genes of N315 (figure 4, A). Maximum aminoacid identity noted among these 21 putative exotoxins was 98%. Pairwise comparison of orthologous *set* gene products indicated that the difference comprised many non-synonymous mutations accumulated in the T-cell receptor-binding domain of the exotoxins.[24]

Structure of νSaβ also differs greatly from strain to strain. The superantigen gene cluster (composed of five enterotoxin genes) carried by islands of N315 and Mu50 (type I) are missing from νSaβ of MW2 (type II) (table 2). Instead, νSaβmw has a novel gene cluster, designated *bsa* (bacteriocin of *S aureus*), which encodes a putative bacteriocin (toxin or antibiotic to other bacteria) and its modification factors, which are similar to those reported in *S epidermidis*[25] (figure 4, B). The type II allelic form of the island is probably important for MW2 to compete with other natural flora for successful colonisation in man.

Genomic islands of S aureus

The two island families, νSaα and νSaβ, both have a broken transposase gene and a RESTRICTION-MODIFICATION SYSTEM composed of *hsdS* and *hsdM* (*hsd* stands for host specificity determinant.) The broken transposase gene feature is consistent with the fact that these two islands are not spontaneously cut out from the chromosome compared with νSa3 and νSa4. The common restriction-modification system seems to contribute to stabilisation of islands in the *S aureus* genome.[4] In support of this view, HsdM methylases encoded by these islands are very well conserved across the six *S aureus* genomes, the aminoacid identity of which range from 99% to 100%. However, the HsdS protein that establishes target specificity of methylation, which is encoded by *hsdS* located next to *hsdM* on each island, exists in three allelic forms whose mutual aminoacid identities are smaller than 66% (figure 4, A and B; table 2). Their aminoacid differences are mainly localised in the putative nucleotide-sequence-recognition regions.[26] For each island family, it is noted that mutual association of the three *hsdS* alleles is well correlated with the mutual structural similarity of the three islands on which they are located. Therefore, νSaα and νSaβ themselves exist in three allelic forms with their *hsdS* genes as landmarks.

νSaα and νSaβ arise in different allelic combinations in each of the six *S aureus* chromosomes. N315 and Mu50 carry type I νSaα and type I νSaβ, NCTC8325 and COL carry type I νSaα and type II νSaβ, whereas MW2 carries type II νSaα and type II νSaβ (table 2). E-MRSA-16 (strain 252) has different allelic forms for both islands (type III). This distribution of allelic forms of island is consistent with the view that the two islands were once mobile genetic elements, and were independently acquired by intraspecies genetic transfer between *S aureus* strains. In support of this hypothesis is the fact that every island of this family, apart from type II νSaα of MW2, has a remnant of transposase gene (figure 4, A and B), and notably, the allotype of this gene also correlates with the allotype of the island on which it is located.

Virulence genes in prophages

All the sequenced *S aureus* strains, except for COL, have prophage ϕSa3. Two new enterotoxin gene alleles, *seg2* and *sek2*, are reported in ϕSα3mw, which encodes putative enterotoxin G and K homologues. These toxin genes are

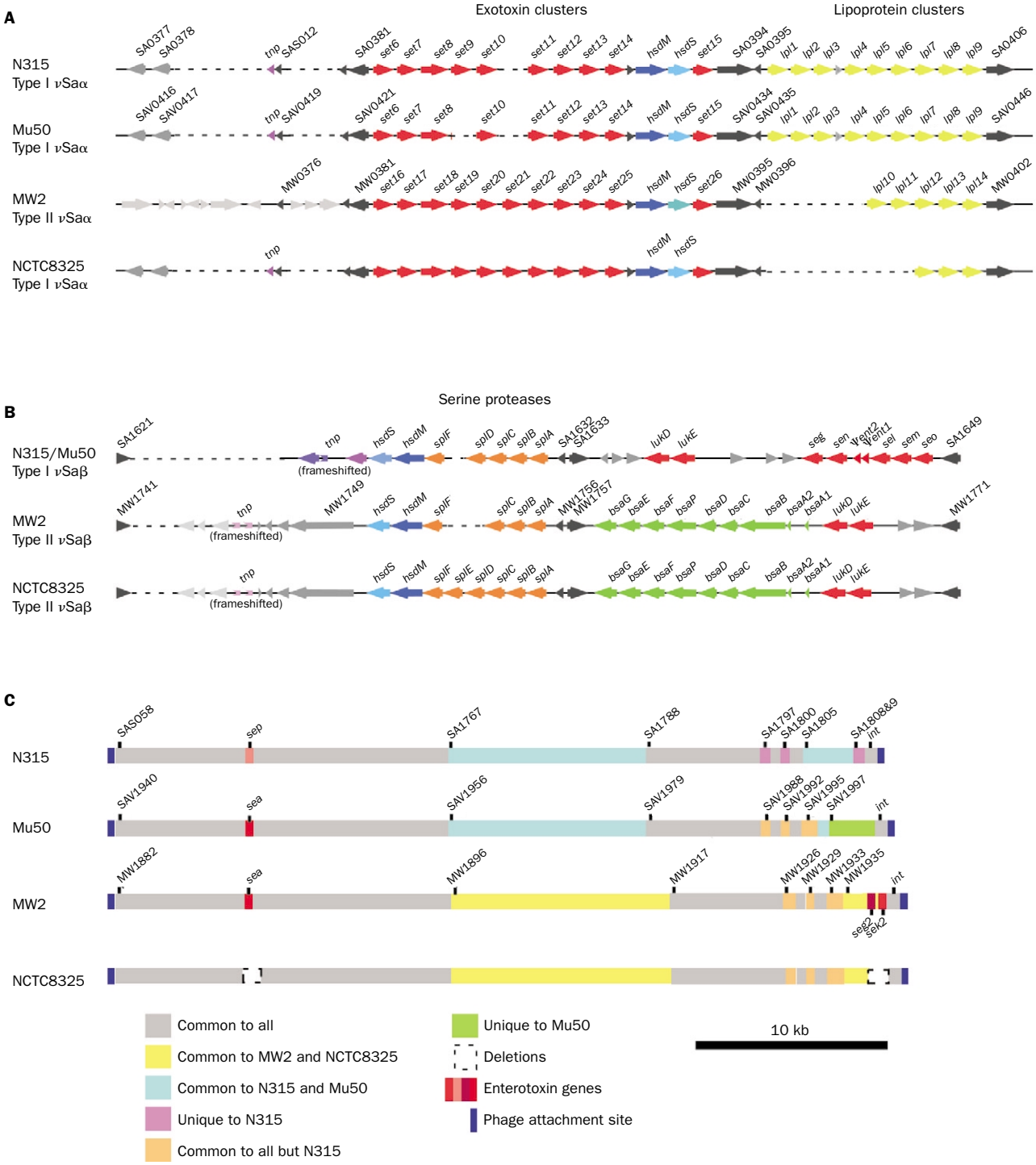

Figure 4: **Staphylococcal genomic islands in strains N315, Mu50, NCTC8325, and MW2**
tnp=transposase gene. Arrows and arrowheads=*orfs* and their direction of transcription. Broken lines=missing *orfs*. No *orf* number has been assigned to NCTC8325 coding regions. Superantigen genes are filled in red. Regions with greater than 90% of their nucleotide sequences identical are filled with the same colour. (A) Genomic island νSaα. (B) Genomic island νSaβ. (C) Prophage ϕSa3. For (A) and (B), red=toxins; yellow=lipoproteins; orange=serine proteases; light green=bacteriocin-related genes. Other unique colours are given by the gene names. Colour for *hsdS* differs by νSa type. Grey arrows=functionally unknown *orfs*. Dark grey=identical *orfs* to all strains shown; light grey=*orfs* not identical to all strains. (C) see key.

not carried by any ϕSa3 prophages found in comparator genomes. Alignment of some of these prophages highlights their mosaic structures that have been generated by multiple crossovers, which is reminiscent of the case of SCC*mec* (figure 4, C). This finding shows clearly that prophages also exist in allelic forms.

The other prophage on MW2 chromosome, ϕSa2mw, carries genes *lukF-PV* and *lukS-PV*, which encode PANTON-VALENTINE LEUKOCIDIN components, which have a potent toxic effect on human white-blood cells. The toxin is strongly associated with severe forms of pneumonia (necrotic pneumonia) caused by community-acquired *S aureus* strains.[27] The ϕSa2 attachment site is present in all six chromosomes, and three strains MW2, NCTC8325, and E-MRSA-16 (strain 252) do have integrated prophages. However, again they exist in

different allelic forms, and only ϕSa2mw has Panton-Valentine leukocidin genes.

Virulence genes of MW2 found outside genomic islands

Collagen-adhesin protein (CNA)[28] is uniquely encoded in the MW2 genome near the origin of replication (*orf* MW2612, indicated by green arrow in figure 2). This protein is reported to be implicated in pathogenesis of osteomyelitis and septic arthritis.[29,30] Small direct repeat sequences are found around *cna*, but there is no evidence that the gene is a part of a mobile genetic element.

Discussion

Transposons and insertion sequences can integrate themselves into any chromosome loci by illegitimate recombination.[31] Thus, they tend to shuffle genome structure and are thought to contribute much to adaptability of *S aureus* to the adverse environment. From this viewpoint, it is curious to note that MW2 has scarcely any transposons or insertion sequences, whereas N315 and Mu50 have many (table 1).

It would be reasonable to assume that hospitals are a severe environment for microorganisms to survive in, because they are constantly exposed to various antiseptics and new antibiotics. Multiple insertions of transposons and insertion sequences in hospital-acquired MRSA genomes might be testament to the evolutionary ordeal they have gone through. In support of this view, we have previously shown that the copy number of insertion sequences of an MRSA strain increases in time during its propagation within a hospital.[32] By contrast, MW2 is thought to represent a successful clone of *S aureus* in the community. The bacteriocin operon in MW2 might be indispensable for community-acquired *S aureus* strains; in this environment, competition with many other bacterial species for colonisation serves as a big selective pressure for survival.

Growth rate is also important. In fact, MW2 grows much faster than hospital-acquired MRSA strains, as doubling time shows, which—in drug-free Mueller-Hinton broth—is 23·5, 34·8, and 46·8 min for MW2, N315, and Mu50, respectively. As an already successful part of the natural flora of healthy human beings, MW2 seems to establish an additional advantage by acquisition of a simple and less encumbering allelic version of SCC*mec* (type IVa) to survive sporadic exposure to limited varieties of antibiotics prescribed in the outpatients' clinic.

MW2 has a total of 18 toxins, which are not recorded in any of the five comparative *S aureus* genomes (N315, Mu50, E-MRSA-16 [strain 252], NCTC8325, and COL). The Panton-Valentine leukocidin components recorded on ϕSa2mw should at least partly contribute to virulence of MW2, since results of an epidemiological study showed a close association of Panton-Valentine leukocidin with necrotic pneumonia,[27] a symptom that is seen in patients infected with MW2 and related strains.[1]

The superantigen staphylococcal enterotoxin H, which is reported to have the highest binding affinity to major histocompatibility complex class II molecules among staphylococcal enterotoxins, is also uniquely reported in MW2 genome.[33] Besides this superantigen, MW2 has 15 unique superantigen genes: 11 exotoxin (*set*) genes on νSaα island and four enterotoxin genes on νSa3 and ϕSa3 genomic islands. These putative superantigens have at least 2% aminoacid differences compared with their extant homologues, and together with staphylococcal enterotoxin H, constitute an unique allelic superantigen range in MW2. These superantigens are expected to trigger human T cells in a quite distinct fashion from those of other *S aureus* strains. The culture supernatant of MW2 contained ten times stronger human T cell proliferation activity than N315, and its activity is directed towards distinct T-cell subsets compared with those activated by N315 (T Uchiyama, unpublished observation). This strong stimulation of distinct T-cell subsets might be correlated with high virulence of MW2.

Some of the virulence genes on MW2 probably contribute to its pathogenicity in healthy human hosts, but it is not clear which of these virulence factors is most important. For example, although work in animals shows that MW2 has an LD_{50} that is five times less than N315, results of gene knockout studies show that no single gene, such as *lukS-PV, lukF-PV*, *cna*, or *seh*, accounts for this difference (T Baba, unpublished). Therefore, it is likely that the total pathogenic potential of MW2 is a result of synthetic contribution of many virulence genes listed in this study (table 2, figure 2). Moreover, although some genes (eg, *cna*) arise singly in the chromosome, most virulence genes tend to exist as a set of allelic genes carried by genomic islands. This fact was first noticed from scrutiny of νSaα, in which the *set* gene cluster encoded a totally different range of superantigens between N315 and MW2. This finding strongly suggests that the pathogenic potential of *S aureus* is established first, by the genomic island families it is part of, and second, by whatever allelic form of each island family it takes. Since most of the genomic islands are potentially mobile, pathogenic potential of each strain of *S aureus* is not an intrinsic trait and is likely to change as a result of intraspecies exchange of virulence genes via lateral gene transfer.

On the basis of our findings, we propose that GENOMIC ISLAND ALLOTYPING is a novel approach to *S aureus* genome typing. This process allows us to predict the pathogenic capability of an *S aureus* clinical strain, and would even help us predict the symptom, severity, and prognosis of the illness that it causes. Genomic island allotyping, done either by PCR, microarray hybridisation, or nucleotide sequence determination of certain regions of genomic islands, would add medically relevant information to extant genotypings, such as pulsed-field gel electrophoresis or multilocus sequence typing,[34] which indicate only the physical genomic information of *S aureus* chromosome.

Contributors

All investigators contributed to design of the study. K Hiramatsu was responsible for overall experimental design and interpretation of data. T Baba did comparative genome analysis and interpretation of structures of the pathogenicity islands and prophages of MW2. F Takeuchi and K Yamamoto did computer-directed analysis on MW2 genome and its comparison with other strains. M Kuroda constructed the genomic library and annotated the MW2 genome. H Yuzawa analysed structural differences of SCC*mec* between MW2 and other strains. A Oguchi, Y Nagai, and N Iwama established the MW2 genomic sequence. K Aoki and K Asano assembled contigs and did the BLAST search. H Kuroda-Murakami studied mobile capability of genomic islands by PCR with a Light Cycler. T Naimi did clinical and epidemiological characterisation of MW2 and associated mid-western community-acquired MRSA strains. L Cui analysed allelic genes of MW2 and comparator strains.

Conflict of interest statement

None declared.

Acknowledgments

We thank Hisashi Kikuchi, chairman of National Institute of Technology and Evaluation, for his support; Chuntima Tiensasitorn, Rika Nishiko and Satomi Suzuki for technical assistance; Susan Johnson, Kathy LeDell, Ruth Lynfield, Richard Danila, and Michael Osterholm (Minnesota Department of Health), Larry Shireley (North Dakota Department of Health), and Fred C Tenover (Centers for Disease Control and Prevention, Atlanta, USA) for clinical and epidemiological characterisation of MW2 strain; Hiroyuki Takase and Ken-ichi Sato (Daiichi Pharmaceutical Co, Japan) for virulence evaluation of MW2 in animal infection model; and Yuh Morimoto for help in preparation of computer graphics.

This work was supported by Grant-in-Aid for Scientific Research on Priority Areas (13226114) from the Ministry of Education, Science, Sports, Culture, and Technology of Japan; and the Core University System Exchange Programme under the Japan Society for the Promotion of Science, coordinated by the University of Tokyo Graduate School of Medicine and Mahidol University. The study was also partly supported by a grant for International Health Cooperation Research (11C-4) from the Ministry of Health and Welfare, Tokyo, Japan, and by the Research for the Future Program of the Japan Society for the Promotion of Science.

References

1 CDC. Four pediatric deaths from community-acquired meticillin-resistant *Staphylococcus aureus*: Minnesota and North Dakota, 1997–1999. *MMWR Morb Mortal Wkly Rep* 1999; **48:** 707–10.

2 Naimi TS, LeDell KH, Boxrud DJ, et al. Epidemiology and clonality of community-acquired meticillin-resistant *Staphylococcus aureus* in Minnesota, 1996–1998. *Clin Infect Dis* 2001; **33:** 990–96.

3 Groom AV, Wolsey DH, Naimi TS, et al. Community-acquired meticillin-resistant *Staphylococcus aureus* in a rural American Indian community. *JAMA* 2001; **286:** 1201–05.

4 Kuroda M, Ohta T, Uchiyama I, et al. Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. *Lancet* 2001; **357:** 1225–40.

5 TIGR. Glimmer. http://www.tigr.org/softlab/glimmer/glimmer.html (accessed May 7, 2002).

6 Suzek BE, Ermolaeva MD, Schreiber M, Salzberg SL. A probabilistic method for identifying start codons in bacterial genomes. *Bioinformatics* 2001; **17:** 1123–30.

7 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215:** 403–10.

8 Ray BK, Apirion D. Characterization of 10S RNA: a new stable RNA molecule from *Escherichia coli*. *Mol Gen Genet* 1979; **174:** 25–32.

9 Washington University in St Louis. tRNAscan-SE search server. http://www.genetics.wustl.edu/eddy/tRNAscan-SE/ (accessed May 7, 2002).

10 The tmRNA website. http://www.indiana.edu/~tmrna/ (accessed May 3, 2002).

11 The NCBI taxonomy homepage. http://www.ncbi.nlm.nih.gov/Taxonomy (accessed May 3, 2002).

12 Tatusova TA, Madden TL. BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* 1999; **174:** 247–50.

13 Bray N, Fabrikant A, Lord J, Schwartz J, Dubchak I, Pachter L. AVID: a global alignment program for large genomic sequences. Available at http://www-gsd.lbl.gov/vista (accessed May 1, 2002).

14 Oklahoma University. *Staphylococcus aureus* NCTC 8325 Genome Sequencing. http://www.genome.ou.edu/staph.html (accessed May 3, 2002).

15 Sanger Centre. *Staphylococcus aureus*. http://www.sanger.ac.uk/Projects/S_aureus/ (accessed May 3, 2002).

16 The Institute for Genomic Research. TIGR databases. http://www.tigr.org/tdb/ (accessed May 3, 2002).

17 Katayama Y, Ito T, Hiramatsu K. Genetic organization of the chromosome region surrounding *mecA* in clinical staphylococcal strains: role of IS*431*-mediated *mecI* deletion in expression of resistance in *mecA*-carrying, low-level methicillin-resistant *Staphylococcus haemolyticus*. *Antimicrob Agents Chemother* 2001; **45:** 1955–63.

18 Ma XX, Ito T, Tiensasitorn C, et al. Novel type of staphylococcal cassette chromosome *mec* identified in community-acquired methicillin-resistant *Staphylococcus aureus* strains. *Antimicrob Agents Chemother* 2002; **46:** 1147–52.

19 Hiramatsu K, Cui L, Kuroda M, Ito T. The emergence and evolution of methicillin-resistant *Staphylococcus aureus*. *Trends Microbiol* 2001; **9:** 486–93.

20 Ren K, Bannan JD, Pancholi V, et al. Characterization and biological properties of a new staphylococcal exotoxin. *J Exp Med* 1994; **180:** 1675–83.

21 Omoe K, Ishikawa M, Shimoda Y, Hu DL, Ueda S, Shinagawa K. Detection of *seg*, *seh*, and *sei* genes in *Staphylococcus aureus* isolates and determination of the enterotoxin productivities of *S aureus* isolates harboring *seg*, *seh*, or *sei* genes. *J Clin Microbiol* 2002; **40:** 857–62.

22 Cui L, Kasegawa H, Murakami Y, Hanaki H, Hiramatsu K. Postoperative toxic shock syndrome caused by a highly virulent methicillin-resistant *Staphylococcus aureus* strain. *Scand J Infect Dis* 1999; **31:** 208–09.

23 Williams RJ, Ward JM, Henderson B, et al. Identification of a novel gene cluster encoding staphylococcal exotoxin-like proteins: characterization of the prototypic gene and its protein product, SET1. *Infect Immun* 2000; **68:** 4407–15.

24 Papageorglou AC, Acharya KR. Microbial superantigens: from structure to function. *Trends Microbiol* 2000; **8:** 369–75.

25 Schnell N, Entian KD, Schneider U, et al. Prepeptide sequence of epidermin, a ribosomally synthesized antibiotic with four sulphide-rings. *Nature* 1988; **333:** 276–78.

26 Fuller-Pace FV, Murray NE. Two DNA recognition domains of the specificity polypeptides of a family of type I restriction enzymes. *Proc Natl Acad Sci USA*. 1986; **83:** 9368–72.

27 Gillet Y, Issartel B, Vanhems P, et al. Association between *Staphylococcus aureus* strains carrying gene for Panton-Valentine leukocidin and highly lethal necrotising pneumonia in young immunocompetent patients. *Lancet* 2002; **359:** 753–59.

28 Patti JM, Jonsson H, Guss B, et al. Molecular characterization and expression of a gene encoding a *Staphylococcus aureus* collagen adhesin. *J Biol Chem* 1992; **267:** 4766–72.

29 Eliasri MO, Thomas JR, Skinner RA, et al. *Staphylococcus aureus* collagen adgesin contributes to the pathogenesis of osteomyelitis. *Bone* 2002; **30:** 275–80.

30 Patti JM, Boles JO, Hook M. The *Staphylococcus aureus* collagen adhesin is a virulence determinant in experimental septic arthritis. *Infect Immun* 1994; **62:** 152–61.

31 Murphy E. Transposable elements in Gram-positive bacteria. In: Berg DE, Howe MM, eds. Mobile DNA. Washington, DC: ASM, 1989: 269–88.

32 Yoshida T, Kondo N, Abu Hanifah Y, Hiramatsu K. Combined use of ribotyping, PFGE typing and IS*431* typing in the discrimination of nosocomial strains of methicillin-resistant *Staphylococcus aureus*. *Microbiol Immunol* 1997; **41:** 687–95.

33 Nilsson H, Bjork P, Dohlsten M, Antonsson P. Enterotoxin H displays unique MHC class II-binding properties. *J Immunol* 1999; **163:** 6686–93.

34 Enright MC, Spratt BG. Multilocus sequence typing. *Trends Microbiol* 1999; **7:** 482–87.