Complete genome sequencing of *Staphylococcus aureus*: insights into virulence and antimicrobial resistance

Benjamin Howden

Department of Microbiology,
Monash University, VIC

Infectious Diseases Department
Austin Health
Studley Rd, Heidelberg, VIC 3084
Fax: (03) 9496 6677
Tel: (03) 9496 6676
Email: Benjamin.Howden@austin.org.au

John K Davies

Department of Microbiology,
Monash University, VIC

Tim Stinear

Department of Microbiology,
Monash University, VIC

***Staphylococcus aureus* remains one of the most important opportunistic bacterial pathogens of humans. It is characterised by the presence of many virulence factors, as well as an ability to rapidly develop or acquire antimicrobial resistance.**

Traditionally, two major patterns of *S. aureus* epidemiology existed, with hospitalised patients acquiring somewhat less virulent, but more antibiotic-resistant strains of *S. aureus* (especially methicillin-resistant *S. aureus* – MRSA), while patients in the community often acquired more virulent, less antibiotic-resistant strains. The recent emergence of highly pathogenic, community-onset MRSA (CO-MRSA) has seen a convergence of these characteristics, and a blurring of the margins between typical hospital and community-acquired *S. aureus* infections. In addition, the recent acquisition of vancomycin resistance in *S. aureus* represents a significant step in the evolution of antimicrobial resistance in *S. aureus*. Genome sequencing has provided deep insights into the evolution of *S. aureus* as a significant human pathogen, with recent projects focusing on better understanding the emergence of CO-MRSA and low-level vancomycin resistance. The advent of low-cost DNA sequencing means that *S. aureus* genome projects are becoming very common so, rather than global assessments, genomics are now being used to address specific questions regarding mechanisms of pathogenesis and resistance.

S. aureus has a clonal population structure, which has been best elucidated by the use of multi-locus sequence typing (MLST) [1]. This process uses differences in the sequences of seven *S. aureus* house-keeping genes to define an MLST sequence type (ST). Isolates with six or seven matching genes belong to the same clonal complex (CC). It has been shown that most MRSA strains can be grouped into five CCs –CC5, CC8, CC22, CC30, CC45 – and in fact almost 90% of all *S. aureus* strains group into one of the 11 most common CCs [1, 2]. Previous genome sequencing projects have predominately included strains from MLST CC5 and CC8. Further sequencing of unrelated strains could provide new insights into the structure of the *S. aureus* genome.

Genome sequencing technology: then and now

The sequencing of DNA molecules first began in the 1970s, initially using the Maxam-Gilbert method; however, the major sequencing technology has been the Sanger method which was initially developed by Frederick Sanger in 1975 [3]. This method involves the use of chain-terminator nucleotides labelled with fluorescent dyes which are combined with fragmented DNA, primers and DNA polymerase. Each nucleotide becomes labelled with a fluorescent dye, and the sequence is resolved in an acrylamide polymer. This was the technology used to complete the first human genome project and the majority of bacterial genome projects to date.

Recently, high throughput, low-cost DNA sequencing technology has been developed, and it has opened the door to the use of whole bacterial genome sequencing as a routine tool in microbiological research. The first of these new technologies, developed by 454 Life Sciences, is based on a microfluidics platform and an alternative sequencing chemistry called pyrosequencing, which was originally developed by Pal Nyren in the 1990s [4].

In this technique, a single-stranded DNA molecule is sequenced from a template, and a series of enzymatic reactions occurs at the sequential addition of each of the four bases, resulting in a detectable light signal. Although the sequence read lengths are somewhat shorter than traditional Sanger sequencing and the quality is lower, the throughput is much higher and the final coverage can be much deeper. Currently, Sanger sequencing can sequence about 2.8 million bases in a 24 hour period, while 454 Sequencing™ can complete 300 million bases within the same timeframe. At its current rate, 454 Sequencing™ could complete a human genome sequence in 10 days [3].

Other companies have developed even higher throughput, with low cost technology. The Illumina Genome Analyzer, based on massively parallel short-read sequencing (<35 nucleotides) can currently generate 2 Giga-bases of sequence in a single 4 day run for a fraction of the cost of Sanger sequencing [5].

The advent of new technologies has changed the framework for understanding bacterial genetics. Indeed, these approaches are not just making genomics more accessible, they are also revolutionising gene expression and DNA-protein interaction studies. The coming years will see an explosion in the number of bacterial genomes sequenced, with greater insights into the molecular mechanisms of pathogens and antimicrobial resistance. Where a successful PhD project once included sequencing a handful of genes, students now have the possibility to sequence one or more bacterial genomes as part of their project.

History of *S. aureus* genome sequencing

The laboratory *S. aureus* strain NCTC8325 was used to first generate the genome map of *S. aureus* [6]. In 2000 the complete sequencing and annotation of the strain was commenced. The complete genome sequence of NCTC8325 is now available and, in addition, the genome sequence of an additional 14 *S. aureus* strains is also available, the majority of which are clinical isolates (Table 1) [6-16]. There are, in fact, more complete genome sequences of *S. aureus* than any other bacterial species.

Initially projects aimed to describe the overall structure and components of the *S. aureus* genome [7, 15] while more recent projects have aimed to determine the genetics behind particular strain characteristics. For example, the sequencing of the USA300 genome aimed to determine features associated with the virulence and highly efficient colonising potential of this strain that is causing significant morbidity and mortality in the United States [10]. Another recent sequencing project mapped the mutations associated with the acquisition of antimicrobial resistance in a pair of *S. aureus* strains [16].

The majority of *S. aureus* genome sequencing projects have utilised traditional Sanger sequencing technology. Our research group has recently undertaken a *S. aureus* genome sequencing project utilising 454 Sequencing™ technology.

Our own genome sequencing project aimed to compare the sequences of two clinical isolates from the same patient which were associated with persistent infection and the emergence of vancomycin-intermediate *S. aureus* during therapy. The initial wound culture isolate was a vancomycin susceptible MRSA, while the later blood culture isolate was a vancomycin-intermediate *S. aureus* [17]. The isolates were identical by pulsed-field gel electrophoresis (PFGE), indicating that the isolates were related to each other, and that significant genetic rearrangement had not occurred during the evolution of low-level vancomycin resistance.

We have previously attempted to define the genetic changes leading to resistance using microarray transcriptional comparisons [18]. Despite significant transcriptional changes in the vancomycin-intermediate *S. aureus* (VISA) strain JKD6008, compared to the vancomycin-susceptible *S. aureus* (VSSA) JKD6009, the genetic determinants of resistance were not easily defined. In particular, we were unable to locate mutations that led to resistance. A complete genome comparison was therefore undertaken. The use of 454 Sequencing™ technology allowed raw sequence data to be generated rapidly (within 3 weeks) with approximately 2.8Mb of unique sequence for each isolate. Mutation detection was then carried out to assess for genetic changes in the VISA isolate; this identified a limited number of mutations which could be responsible for resistance.

Characteristics of the *S. aureus* genome

The availability of multiple *S. aureus* genome sequences makes it possible to determine the identification of the 'core' *S. aureus* genome (i.e. present in all strains) and of the 'accessory' genome (i.e. present in some of the strains) [19].

Figure 1 gives an overview of the 'typical' *S. aureus* genome which has a size of approximately 2.74-2.91Mb, and contains approximately 2600 protein coding sequences. About 75% of the genome comprises the core component, which is generally conserved and composed of genes present in all strains. Figure 2 provides a comparison of the genome sequences for a number of strains in relation to the early MRSA strain COL [20]. This figure demonstrates that the different strains have similar genomic architecture throughout most of the genome. The core genome contains genes required for growth and survival, and also contains some genes associated with pathogenesis (Figure 1). Although the core genome is conserved between strains, many genes encode hypothetical proteins of unknown function. In addition, some parts of the 'core' genome are not as stable as

the term suggests, and the core genome has therefore been subdivided into 'stable core' and 'core variable genome' [21]. The core variable genome tends to encode virulence factors, and has a higher rate of nucleotide variability, and a higher number of variable number tandem repeats (VNTRs). These changes can lead to alterations in protein function or specificity.

The accessory genome is mainly composed of mobile genetic elements such as plasmids, transposons, phages and pathogenicity islands. These elements are able to move horizontally between *S. aureus* strains, and can provide the basis for differences in antibiotic resistance or pathogenicity. However, not all of these elements retain the ability to be mobile.

The genetic diversity of the *S. aureus* genome, and hence the ability of the strain to behave differently in a clinical context, can arise from the accessory genome with different elements present in different strains. In addition, variability can arise from the core genome, where subtle genetic diversity due to point mutations in core genes can affect gene or protein function [19]. Examples of this include the evolution of methicillin resistance in *S. aureus*, where the *mecA* gene is acquired as part of a mobile genetic element and is part of the 'accessory genome'. In contrast, it appears that the evolution of low-level vancomycin resistance in *S. aureus* is at least in part related to point mutations in important regulatory genes of the core genome [8, 16].

Key features of previous *S. aureus* genome sequencing projects

In 2001 the complete genome sequences of the Japanese MRSA strain N315 and the VISA strain Mu50 were published [7]. This project characterised the basic structure of the *S. aureus* genome and discovered a number of novel mobile genetic elements. It also gave insight into the mechanisms by which *S. aureus* generates genetic diversity. A project that compared the MRSA COL genome sequence to that of a *Staphylococcus epidermidis* strain (RP62a) also suggested that the variability and virulence of *S. aureus* has arisen by lateral gene transfer with other low-GC-content gram-positive bacteria [15]. Additional phenotypic

Table 1. Summary of *S. aureus* genome sequences.

Strain	MLST (CC)	Features
NCTC8325 [6]	ST8 (CC8)	Laboratory strain, UK, before 1949
N315 [7]	ST5 (CC5)	Hospital-acquired MRSA, Japan, 1982
Mu50 [7]	ST5 (CC5)	Hospital-acquired VISA, Japan, 1997
Mu3 [8]	ST5 (CC5)	Hospital-acquired hVISA, Japan, 1997
MW2 [9]	ST1 (CC1)	PVL positive CO-MRSA, fatal childhood infection, USA, 1998
USA300 [10]	ST8 (CC8)	Skin and fort tissue infection isolate, PVL positive, San Francisco, USA, 2000
TCH1516 [11]	ST8 (CC8)	USA300-HOU-methicillin-resistant, adolescent with severe sepsis, USA
TCH959 [11]	ST8 (CC8)	USA300-HOU-methicillin-susceptible, 12 yo with buttock abscess, USA
MRSA252 [12]	ST36 (CC30)	EMRSA-16, hospital-acquired epidemic strain, UK, 1997
MSSA476 [12]	ST1 (CC1)	Community-acquired MSSA, UK, 1998
RF122 [13]	ST151	Associated with mastitis in cattle, Ireland, 1993
Newman [14]	ST8 (CC8)	Human infection, 1952. No antibiotic resistance. Pathogenesis extensively investigated
COL [15]	ST250 (CC8)	Early MRSA strain, UK, 1961
JH1 [16]	ST5 (CC5)	Hospital-acquired VSSA, vancomycin MIC 1mg/L, USA, 2000
JH9 [16]	ST5 (CC5)	Hospital-acquired VISA, vancomycin MIC 8mg/L, USA, 2000. Clinical derivative of JH1

MLST multi-locus sequence type
CC clonal complex
MRSA methicillin-resistant *S. aureus*
VISA vancomycin-intermediate *S. aureus*
hVISA hetero-VISA
PVL Panton-Valentine Leukocidin
CO-MRSA community-onset MRSA
MSSA methicillin-susceptible *S. aureus*
VSSA vancomycin-susceptible *S. aureus*

differences were thought to be attributable to single nucleotide polymorphisms, especially in cell envelope genes.

Holden *et al.* sequenced the UK EMRSA-16 strain (MRSA252), a successful hospital clone of MRSA, and compared it to the community methicillin-susceptible *S. aureus* (MSSA) isolate, MSSA476 [12]. This comparison demonstrated the importance of the accessory genome in determining clinical characteristics of the isolates. MRSA252 is one of the most genetically diverse strains sequenced to date, with ~6% of the genome containing novel elements. Significant differences in the presence of antibiotic resistance determinants and virulence genes were found in the accessory genomes of the strains. Interestingly, the methicillin-susceptible strain MSSA476 carried SCC*mec*-like element, which did not contain the *mecA* gene, but did carry a gene conferring fusidic acid resistance.

The CO-MRSA strain MW2 caused the deaths of a number of children in North Dakota in 1998. The genome sequence of this strain was published in 2002, and revealed the novel type IV SSC*mec* which is characteristic of CO-MRSA strains and divergent from typical SSC*mec* found in hospital MRSA strains [9]. In particular, the novel type IV SSC*mec* does not carry additional antimicrobial resistance genes, and partly explains the non-multi-resistant nature of CO-MRSA strains. In addition, novel virulence factors were discovered, and the presence of *lukF-PV* and *lukS-PV* encoding Panton-Valentine Leukocidin (PVL) was documented. This important cytotoxin was carried on the prophage ϕSa2, often also found in other *S. aureus* genomes.

Diep *et al.,* sequenced the CO-MRSA strain USA300, which is a major cause of community-acquired infections in the USA and Europe [10]. Interestingly the authors found a novel mobile genetic element which was part of the accessory genome in USA300, and which encoded an arginine deaminase pathway and an oligopeptide permease system which could contribute to growth and survival of USA300. This mobile element (arginine catabolic mobile element – ACME) was not found in other *S. aureus* strains, but is highly prevalent in *S. epidermidis,* generating the hypothesis that the element may confer increased ability of the strain to colonise skin. The sequencing of a second USA300 CO-MRSA and a USA300 MSSA strain revealed some differences in plasmid content which could have an effect on clinical virulence [11].

Mwangi *et al.* (2007) sequenced a pair of clinical isolates from a patient with MRSA bacteraemia in whom VISA developed after failure of vancomycin therapy. In addition to the emergence of vancomycin-resistance, the MRSA strain developed resistance to rifampicin and low-level resistance to daptomycin [16]. Genome sequencing of the first and last clinical isolate detected 35

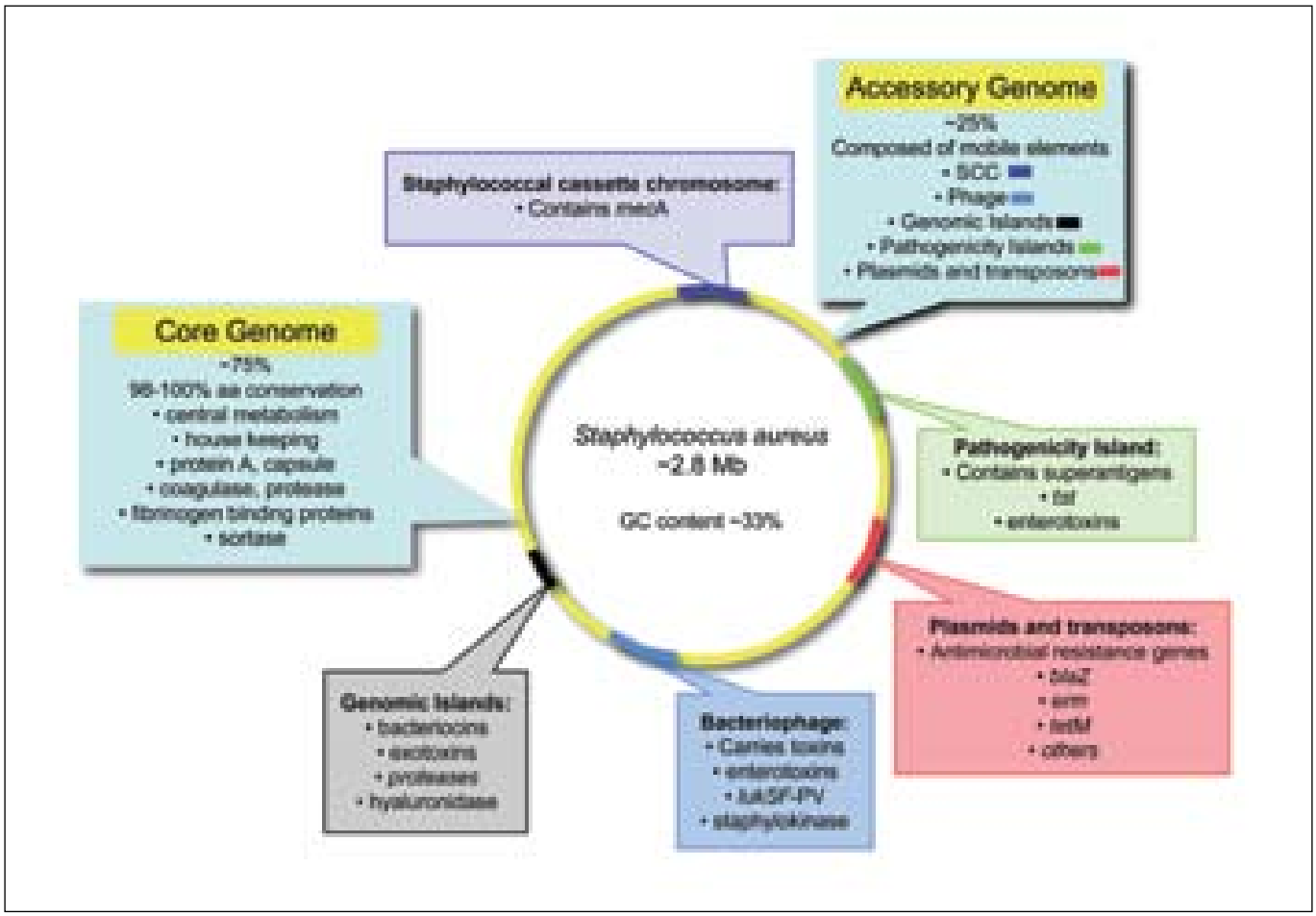

Figure 1. Overview of a 'typical' *S. aureus* genome.

point mutations in 31 loci. Important mutations which were potentially linked to changes in vancomycin MIC were found in a global regulatory operon called *vraSR*. Neoh *et al.* compared the genome sequences of the hetero-VISA (hVISA) strain Mu3 and the VISA strain Mu50, and discovered an important point mutation in the response regulator *graR* that was associated with generation of VISA from hVISA [8]. The results of our genome sequencing comparison of VSSA compared to VISA have demonstrated mutations in different global regulatory genes and will provide additional information on the genetic pathways to resistance [unpublished data].

Where to from here?

Although we now have a good understanding of the *S. aureus* genome structure, many questions remain to be answered. In particular, understanding the determinants of colonisation and invasion could lead to improvements in strategies to control *S. aureus*. Further detailed analysis of VISA strains will help determine the exact mechanisms of resistance. Additionally, understanding the mode of transfer of mobile genetic elements will also be important. Further insights into the structure and function of the *S. aureus* genome will aid in the development of strategies for improved prevention and treatment of *S. aureus* disease.

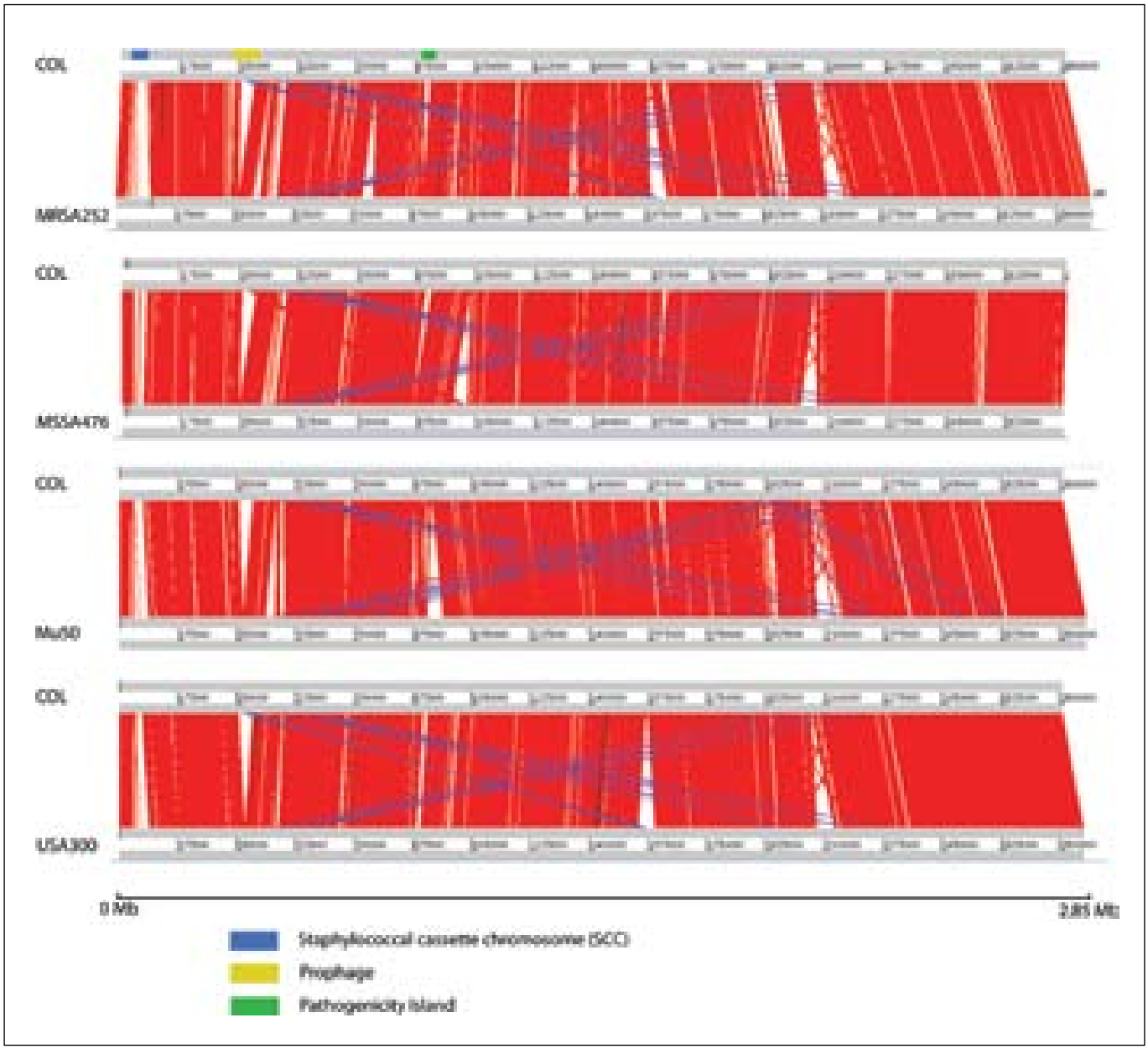

Figure 2. Comparison of the chromosomes of four sequenced *S. aureus* strains compared to the MRSA COL genome sequence. The alignments were generated using a genome comparison tool, artemis [20]. The lines connecting the genomes represent orthologous matches identified by BLAST analysis. If the connecting lines are red, the orthologues are in the same orientation; if they are blue, the orthologues are in the reverse orientation. Some of the accessory genome elements are highlighted and demonstrate that these regions tend to confer the variability between genomes.

References

1. Feil, E.J. and Enright, M.C. (2004) Analyses of clonality and the evolution of bacterial pathogens. *Curr. Opin. Microbiol.* 7, 308-313.
2. Enright, M.C. *et al.* (2002) The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). *Proc. Natl. Acad. Sci. USA* 99, 7687-7692.
3. Available at 454 Life Sciences Corporation website: http://www.454.com
4. Nyrén, P. (2007) The history of pyrosequencing. *Methods Mol. Biol.* 373,1-14.
5. Available at: http://www.illumina.com/
6. Gillaspy, A.F. *et al.* (2006) The *Staphylococcus aureus* NCTC 8325 Genome. In: Gram Positive Pathogens, pp.381-411.
7. Kuroda, M. *et al.* (2001) Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. *Lancet* 357, 1225-1240.
8. Neoh, H.M. *et al.* (2007) Mutated response-regulator *graR* is responsible for hetero-VISA to VISA phenotypic conversion of vancomycin resistance in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 52, 45-53.
9. Baba, T. *et al.* (2002) Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet* 359, 1819-1827.
10. Diep, B.A. *et al.* (2006) Complete genome sequence of USA300, an epidemic clone of community-acquired methicillin-resistant *Staphylococcus aureus*. *Lancet* 367, 731-739.
11. Highlander, S.K. *et al.* (2007) Subtle genetic changes enhance virulence of methicillin resistant and sensitive *Staphylococcus aureus*. *BMC Microbiol.* 7, 99.
12. Holden, M.T. *et al.* (2004) Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. *Proc. Natl. Acad. Sci. USA* 99, 101, 9786-9791
13. Herron-Olson, L. *et al.* (2007) Molecular correlates of host specialization in *Staphylococcus aureus*. *PLoS ONE* 2, e1120.
14. Baba, T. *et al.* (2008) Genome sequence of *Staphylococcus aureus* strain Newman and comparative analysis of Staphylococcal genomes: polymorphism and evolution of two major pathogenicity islands. *J. Bacteriol.* 190, 300-310.
15. Gill, S.R. *et al.* (2005) Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. *J. Bacteriol.* 187, 2426-2438.
16. Mwangi, M.M. *et al.* (2007) Tracking the *in vivo* evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing. *Proc. Natl. Acad. Sci. USA* 99, 9451-9456.
17. Howden, B.P. *et al.* (2006) Isolates with low-level vancomycin resistance associated with persistent methicillin-resistant *Staphylococcus aureus* bacteremia. *Antimicrob. Agents Chemother.* 50, 3039-3047.
18. Howden, B.P. *et al.* (2008) Different bacterial gene expression patterns and attenuated host immune responses are associated with the evolution of low-level vancomycin resistance during persistent methicillin-resistant *Staphylococcus aureus* bacteraemia. *BMC Microbiol.* 8, 39.
19. Lindsay, J.A. and Holden, M.T. (2006) Understanding the rise of the superbug: investigation of the evolution and genomic variation of *Staphylococcus aureus*. *Funct. Integr. Genomic.* 6, 186-201.
20. Available at the Sanger Institute website: http://www.sanger.ac.uk
21. Feng, Y. *et al.* (2007) Evolution and pathogenesis of *Staphylococcus aureus*: lessons learned from genotyping and comparative genomics. *FEMS Microbiol. Rev.* 32, 23-37.

Ben Howden is an Infectious Diseases Physician and Medical Microbiologist at Austin Health. He is completing a PhD in the Department of Microbiology at Monash University, investigating the mechanisms of low-level vancomycin resistance in clinical isolates of *Staphylococcus aureus*, with a focus on comparative and functional genomic comparisons of vancomycin susceptible and resistant strains.

Tim Stinear received his PhD in Microbiology from Monash University in 2001 followed by a 3-year postdoc with Professor Stewart Cole at the Institut Pasteur to sequence the genome of *Mycobacterium ulcerans*. Since 2005 he has led his own research group in the Department of Microbiology at Monash University, studying bacterial pathogenesis.

John Davies is the Head of the Department of Microbiology at Monash University. His major research focus is the pathogenesis of *Neisseria gonorrheae* and *Neisseria meningitidis*, with a focus on gene regulation.

Evolution of multiple drug resistance in staphylococci

Slade O Jensen, Stephen M Kwong
Bruce R Lyon & Neville Firth

School of Biological Sciences, University of Sydney, NSW
Tel: (02) 9351 3369
Fax: (02) 9351 4771
E-mail: slade.jensen@bio.usyd.edu.au
E-mail: skwong@bio.usyd.edu.au
E-mail: b.lyon@usyd.edu.au
E-mail: nfirth@bio.usyd.edu.au

Since the dawn of the era of antibiotic therapy, beginning with the introduction of the first penicillins (ß-lactams) in the 1940s, strains of *Staphylococcus aureus* exhibiting resistance to antibiotics have become more and more prevalent in both clinical and community settings. Foremost among these strains are the so-called methicillin-resistant *S. aureus* (MRSA), which have gained worldwide notoriety as hospital 'superbugs'. The name MRSA belies the true nature of these organisms, as they are not only resistant to penicillin and the ß-lactamase-resistant penicillins, such as methicillin, oxacillin and flucloxacillin, but are usually also resistant to a significant array of other antibiotics (Table 1), representing most of the available drug classes.

In the USA every year it is estimated that *S. aureus* infections are associated with approximately 292,000 hospitalisations, with around 126,000 caused by MRSA – around 19,000 die [1]. In Australia, it is thought that more than 1700 deaths are associated with *S. aureus* bacteraemia per year [2]. Treatment of infections caused by MRSA is problematic, as few antibiotics are available to control the organism. In this regard, reports of *S. aureus* strains that are resistant to vancomycin [3], once considered the antibiotic of last resort but increasingly used as a front-line treatment, are particularly concerning.

How did we get from a situation of apparent universal susceptibility of *S. aureus* to antibiotics to almost total resistance in a little over 50 years? Research has highlighted the accumulation of a