Materials and Methods

Bacterial Isolates & Storage. This study utilised a temporal and geographically dispersed collection of 123 ST239 *S. aureus* isolates. Of these, 87 were recovered from the Australian states of Victoria (n=52), New South Wales (n=10), and Queensland (n=10), as well as New Zealand (n=2), between 1980 and 2012. The Australasian isolates were selected from available collections of *S. aureus* to achieve the widest possible temporal and geographic diversity. In addition to this, 13 isolates were selected as whole genome sequence data was already available. Isolates were identified as ST239 using pulsed-field gel electrophoresis, or an ST239 specific PCR as previously described (1). Twenty-seven isolates are from previously published studies (2-4), and the remaining 60 are novel. The majority of these represent clinical isolates causing either septicaemia (n=50), or tissue infection (n=7) in non-duplicated patients. Twelve represent paired, or triplet, isogenic isolates recovered from patients who developed reduced glycopeptide susceptible variants during persistent infection. The remaining 17 isolates, all recovered in the early 1980s, had no clinical data available. This collection was supplemented with the genome sequences of an additional 35 international ST239 isolates (5)

Bacterial isolates were stored in glycerol broth at -80°C and subcultured twice on horse blood agar (Oxoid) prior to all testing. If liquid culture was required the second subculture was performed in heart infusion (HI) broth.

Genome Sequencing & Analysis

A total of 71 isolates were sequenced in this study. Genomic DNA was extracted using the GenElute Genomic DNA Kit (Sigma), multiplex short read DNA libraries were created with

the Nextera XT DNA Preparation Kit (Illumina), and whole genome sequencing was performed on the MiSeq (Illumina) using 2 x 150 bp chemistry; all performed as per the manufacturers recommendations.

The genome sequence reads of previously published isolates are available under the following accession numbers: TW20 (GenBank, FN433596), the 35 international ST239 isolates (Sequence Read Archive, ERA000102), and JKD6008 (GenBank, CP002120). All other sequences were submitted to the National Centre for Biotechnology Information GenBank and are available under BioProject SAMN03083454.

A read mapping approach was used to align short read sequences of all 123 isolates to reference ST239 strain TW20 using SHRiMP v2.0 (6). Identification of mutation in read alignments were performed as previously described (7). Briefly, singles nucleotide polymorphisms (SNPs) were identified using a python utility that creates a record of all putative differences, using a stringency threshold of 80%, for each nucleotide position in the alignment. The core genome is then defined as those nucleotide positions represented in each and every isolate.

Detection of recombination. Core genome SNPs potentially introduced by recombination were identified using a modified method of that proposed by Croucher *et. al*. (8). A phylogeny was established using RAxML (9), then SNPs were mapped onto the branches using PAML v4 (10); each branch representing a unique group of temporally related SNPs. Each group was examined for spatial clustering using a chi-squared analysis with a Bonferroni correction applied to determine significance. Clustered SNPs were removed and the analysis repeated with no further clustering identified.

Phylogenomic analyses. An initial global population model was estimated using PhyML v3 (11), under a general time reversible (GTR) model of nucleotide substitution with 1000 bootstrap replicates and rooted with the non-ST239 *S. aureus* strain FPR3757 (GenBank, NC_007793.1). This phylogeny was used to assess the correlation between branch length and year of isolation with Path-o-Gen v1.3 (12).

Phylogenetic analysis of the core genome, non-recombinant SNPs was completed using BEAST v1.7.5 (refer to http://beast.bio.ed.ac.uk for all software in this paragraph) (12), also using the GTR model. Three molecular clock models were initially trialled (strict clock, uncorrelated relaxed log-normal (UCLN) and random local clock) with the best likelihood achieved using the UCLN model (unpublished data), a result consistent with Gray *et.al.* (13). A constant size tree prior was used for modelling phylogeny and a Bayesian Skyline model for estimating the effective population size; run with the default parameters of a piecewise-constant model and 10 groups. For the global phylogeny (Fig. 1A) four 100 million Monte Carlo Markov Chain (MCMC) analyses were run and combined, and for individual clade analyses four 10 million MCMC chains were run and combined, sampling every 1000^{th} steps to achieve effective sampling. The burnin was set at the earliest point at which a stable likelihood was achieved in each run (interpreted visually in Tracer v1.5). No runs were discarded as the marginal likelihoods of all iterations was ≤ 2 (14). Runs were combined using LogCombiner v1.7.5 and re-sampled at a frequency that would achieve approximately 10,000 samples. TreeAnnotator v1.7.5 was used to select the topology with maximum clade credibility and annotate the tree with mean heights. Final alterations to tree topology, including branch colour and annotation, were undertaken in FigTree v1.4.2. EPS models (Fig. 2A) were generated in Tracer v 1.5 using the Bayesian Skyline reconstruction tool, run using

default parameters. Calculation of the molecular clock as SNPs per site per year was completed by multiplying the rate estimated in BEAST with the number of variant sites tested, then dividing the total by the approximate size of a *S. aureus* genome (2.8Mb) (13).

Gene Presence/Absence. Sequence reads were de novo assembled as previously described (7), and the resulting contigs were used to construct an orthologous protein matrix with proteinortho v5 (14), using blastp+ with a default e-value of $1x10^{-5}$. This matrix was used to generate a multidimensional scaling plot, in FriPan (http://drpowell.github.io/FriPan/, unpublished), representing the variation in gene content between isolates (Figure 2B). Identification of the CDSs that differentiate each clade was completed with an in-house pearl script that reconstructs the orthologous protein matrix under defined criteria. ie. The inclusion of each clade, run independently, with and without the exclusion of the eight Clade 1 isolates.

Phenotypic Analysis.

Antimicrobial susceptibility testing. The minimum inhibitory concentration (MIC) for each isolate were determined for vancomycin, teicoplanin and daptomycin, using Etests (bioMérieux), following the manufacturers recommendations. Results were interpreted following the Clinical Laboratory Standards Institute (CLSI) guidelines (15). To determine reduced glycopeptide susceptibility Macro-Etests, using a 2.0 McF solution, were performed as previously described by Walsh et. al. (16). *S. aureus* ATCC25923 and Mu3 were included as controls.

Genome wide association study (GWAS). Allelic association was performed using the GWAS software Plink (17). Phenotypic groups were defined by a vancomycin MIC of ≤ 2 μg/mL or ≥ 3 μg/mL. Core genome SNPs that resulted in non-synonymous amino acid changes were compared testing each SNP individually, as well as grouping SNPs based on location and assigning an affected or unaffected status for each CDS with or without mutation, respectively. A Bonferroni correction was used to account for multiple testing; the critical p-value determined as ≤ 0.05/ number of SNPs or CDSs tested.

Growth curves. The average doubling time for each isolate was determined using growth curves. These were performed as previously described (18), with the one exception that Mueller Hinton broth (BD) was used. Isolates were tested in both biological and technical triplicate. To convert optical density into doubling time, technical replicates were averaged and graphed on a log scale. The last time point in which all isolates entered exponential growth (40 minutes) and first time point at which the first isolate entered stationary phase (130 minutes) was used as time points 0 and 1, respectively. The following two equations were then applied to each biological replicate and the final doubling times averaged for each isolate.

$$Generation\ Time\ (Gt) = \frac{log10(\ Time\ 1\ ) - \ log10(\ Time\ 0\ )}{\log10(2)}$$

$$Doubling\ Time\ (Dt) = (Gt\ \times\ \Delta Time)\ \times 60$$

Detection of delta-haemolysin. The delta-haemolysin assay was performed as previously described (19). Briefly, isolates were streaked perpendicular to *S. aureus* strain RN4220, a producer of only beta-haemolysin, on sheep blood agar (Oxoid) and incubated for 24 hours at 37°C. Augmentation to the zone of haemolysis for the query strain when it overlaps with the zone of beta-haemolysis is considered a positive result for the presence of delta-haemolysin. Each isolate was tested in biological triplicate and any inconsistencies were resolved by accepting the results seen in two of the three tests.

Reverse Transcription - quantitative Polymerase Chain Reaction (RT-qPCR). A total of 17 isolates were randomly selected, using a random number generator in Excel, from within three groups: Clade 1 isolates recovered during the 1980s (n = 6), Clade 1 isolates recovered after 1990 (n = 6) and Clade 2 isolates (n = 5). Total RNA was extracted from cells grown to an OD_{600} of 1.0-1.1 in HI broth using the RNeasy Mini Kit (Qiagen), as per the manufacturer's recommendations with the inclusion of two on-column DNase I digestion steps. Synthesis of cDNA and qPCR were performed as previously described (3), using RNAIII primers from (20). Control strain JKD6159 was used to establish a standard curve, and the average C_t, from three technical replicates, was used to calculate a fold difference for each isolate compared to JKD6159. The average of two biological replicates is displayed in Figure 4A.

Murine sepsis model. Wild-type 8-week-old female BALB/c mice were injected via the tail vein with 100µl of a phosphate buffered saline (PBS) solution containing approximate 1×10^6 colony forming units (CFU) of either BPH2032, BPH2057, BPH2070, BPH2073, BPH2080, BPH2086, BPH2095 or BPH2098. These isolates represent the first eight isolates chosen in

the random selection of isolates for RT-qPCR (see RT-qPCR section above). Mice were monitored every eight hours for signs of illness. Any mice that developed severe illness were euthanized by CO_2 inhalation. All surviving mice were euthanized at day seven post-injection. Bacterial burden within the liver, kidney and spleen were determined by tissue homogenisation in 5ml PBS and serial dilution. The minimum detectable CFU count was 5 x 10^4 CFU/mL. All animal work was performed in accordance with the Animal Research Ethics Committee at the University of Melbourne.

Statistical analyses. All statistical analyses were performed in either R v2.15.3 (21), or Prism GraphPad v6.04 (22). Comparison of growth rates and MIC values were conducted in R using a Welsh two sample t-test. Significance was determined as a p values ≤ 0.05. In both analyses, the results from the four isolates belonging to the small clade (BPH2006, BPH2008, BPH2010 and BPH2011) were included with Group 2. A log transformation was applied to MIC values during comparison. Analysis of the *in vivo* sepsis model was completed using Prism Graphpad. Kaplan-Meier curves were established and log-rank tests used to compare survival. Differences in percentage weight loss from pre-injection weight was completed using either a unpaired t-test for comparison of temporal groups or a one way ANOVA for RNA III expression groups. Significance was determined as a p value ≤ 0.05.

All images were compiled and finalised using Inkscape v0.48.5 (23).

References

1. **Feil, E. J., E. K. Nickerson, N. Chantratita, V. Wuthiekanun, P. Srisomang, R. Cousins, W. Pan, G. Zhang, B. Xu, N. P. J. Day, and S. J. Peacock.** 2008. Rapid detection of the pandemic methicillin-resistant *Sta*phylococcus *aureus* clone ST239, a dominant strain in Asian hospitals. J. Clin. Micro **46:**1520-1522.
2. **Kelley, P. G., W. Gao, P. B. Ward, and B. P. Howden.** 2011. Daptomycin non-susceptibility in vancomycin-intermediate *Staphylococcus aureus* (VISA) and heterogeneous-VISA (hVISA): implication for therapy after vancomycin treatment failure. J. Antimicrob. Chemother **66:**1057-1060.
3. **Howden, B. P., P. D. R. Johnson, P. B. Ward, T. P. Stinear, and J. K. Davies.** 2006. Isolates with low-level vancomycin resistance associated with persistent methicillin-resistant *Staphylococcus aureus* bacteremia. Antimicrob. Agents Chemother. **50:**3039-3047.
4. **Holmes, N. E., J. D. Turnidge, W. J. Munckhof, J. O. Robinson, T. M. Korman, M. V. N. O'Sullivan, T. L. Anderson, S. A. Roberts, W. Gao, K. J. Christiansen, G. W. Coombs, P. D. R. Johnson, and B. P. Howden.** 2011. Antibiotic choice may not explain poorer outcomes in patients with *Staphylococcus aureus* bacteremia and high vancomycin minimum inhibitory concentration. J. Infect. Dis. **204:**340-347.
5. **Harris, S. R., E. J. Feil, M. T. G. Holden, M. A. Quail, E. K. Nickerson, N. Chantratita, S. Gardete, A. Tavares, N. Day, J. A. Lindsay, J. D. Edgeworth, H. de Lencastre, J. Parkhill, S. J. Peacock, and S. D. Bentley.** 2010. Evolution of MRSA during hospital transmission and intercontinental spread. Science **327:**469-474.
6. **David, M., M. Dzamba, L. D, L. Ilie, and M. Brudno.** 2009. SHRiMP2: sensitive yet practical short read mapping. Bioinform. **25:**1966-7.

7. **Stinear, T. P., K. E. Holt, K. Y. L. Chua, J. Stepnell, K. L. Tuck, G. W. Coombs, P. F. Harrison, T. Seemann, and B. P. Howden.** 2014. Adaptive change inferred from genomic population analysis of the ST93 epidemic clone of community-associated methicillin-resistant *Staphylococcus aureus*. Genome Biol. Evol. **6:**366-378.

8. **Croucher, N. J., S. R. Harris, C. Fraser, M. A. Quail, J. Burton, M. van der Linder, L. McGee, A. von Gottberg, J.-H. Song, K. S. Ko, B. Pichon, S. Baker, C. M. Parry, L. M. Lambertsen, D. Shahinas, D. R. Pillai, T. J. Mitchell, G. Dougan, A. Tomasz, K. P. Klugman, J. Parkhill, W. P. Hanage, and S. D. Bentley.** 2011. Rapid phenumococcal evolution in response to clinical intervention. Science **331:**430-434.

9. **Stamatakis, A.** 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinform. **22:**2688-2690.

10. **Yang, Z.** 2007. PAML4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24:**1586-1591.

11. **Guindon, S., F. Delsuc, J. F. Dufayard, and O. Gascuel.** 2009. Estimating maximum liklihood phylogenies with PhyML. Methods Mol. Biol. **537:**113-137.

12. **Rambaut, A.** 2010, posting date. Path-O-Gen v1.3: temporal signal investigation tool. [http://tree.bio.ed.ac.uk/software/pathogen/]

13. **Gray, R. R., A. J. Tatem, J. A. Johnson, A. V. Alekseyenko, O. G. Pybus, M. A. Suchard, and M. Salemi.** 2011. Testing spatiotemporal hypothesis of bacterial evolution using methicillin-resistant *Staphylococcus aureus* ST239 genome-wide data within a bayesian framework. Mol. Biol. Evol. **28:**1593-1603.

14. **Lechner, M., S. Findei, L. Steiner, M. Marz, P. F. Stadler, and S. J. Prohaska.** 2011. Proteinortho: Detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics **12**.

15. **CLSI.** 2012. Performance standards for antimicrobial susceptibility testing: twenty-second informational supplement, vol. CLSI document M100-S22. Clinical Laboratory Standards Institute, Wayne, PA.

16. **Walsh, T. R., A. Bolmstrom, A. Qwarnstrom, P. Ho, M. Wooton, R. A. Howe, A. P. MacGowan, and D. Diekema.** 2001. Evaluation of current methods for detection of staphylococci with reduced susceptibility to glycopeptides. J. Clin. Micro **39:**2439-2444.

17. **Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham.** 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. **81:**559-575.

18. **Lam, M. M. C., T. Seemann, N. J. Tobias, H. Chen, V. Haring, R. J. Moore, S. Ballard, L. M. Grayson, P. D. R. Johnson, B. P. Howden, and T. P. Stinear.** 2013. Comparative analysis of the complete genome of an epidemic hospital sequence type 203 clone of vancomycin-resistant *Enterococcus faecium*. BMC Genomics **14**.

19. **Herbert, S., A.-K. Ziebandt, K. Ohlsen, T. Schafer, M. Hecker, D. Albrecht, R. P. Novick, and F. Gotz.** 2010. Repair of global regulators in *Staphylococcus aureus* 8325 and comparative analysis with other clinical isolates. Infect. Immun. **78:**2877-2889.

20. **Howden, B. P., C. R. E. McEvoy, D. L. Allen, K. Chua, W. Gao, P. F. Harrison, J. Bell, G. Coombs, V. Bennett-Wood, J. L. Porter, R. Robins-Browne, J. K. Davies, T. Seemann, and T. P. Stinear.** 2011. Evolution of multidrug resistance

during *Staphylococcus aureus* infection involves mutation of the essential two component regulator WalKR. PLoS Pathog. **7:**e1002359.

21. **R Core Team.** 2013. R v2.15.3: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [http://www.r-project.org/]

22. **GraphPad Software.** 2014. GraphPad Prism v6.04. GraphPad Software, Inc., San Diego, California USA. [http://www.graphpad.com/scientific-software/prism/]

23. **The Inkscape Team** 2014, posting date. Inkscape v0.48.5. [https://inkscape.org/en/]