

The continuing evolution of community-associated MRSA ST93-IV in Australia

Stanley Pang (✉ s.pang@murdoch.edu.au)

Murdoch University <https://orcid.org/0000-0001-9165-5704>

Denise A Daley

Fiona Stanley Hospital

Shafi Sahibzada

Murdoch University

Shakeel Mowlaboccus

Murdoch University

Marc Stegger

Statens Serum Institut

Geoffrey W Coombs

Murdoch University

Research article

Keywords: Staphylococcus aureus, GWAS, bacteraemia, phylogenomics, Australia

Posted Date: December 11th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-65512/v2>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background The global emergence of community-associated methicillin-resistant *Staphylococcus aureus* (CA-MRSA) has seen the dominance of specific clones in different regions around the world with the PVL-positive ST93-IV as the predominant CA-MRSA clone in Australia. In this study we applied a genome-wide association study (GWAS) approach on a collection of Australian ST93-IV MRSA genomes to identify genetic traits that may have assisted the ongoing transmission of ST93-IV in Australia. We also compared the genomes of ST93-IV bacteraemia and non-bacteraemia isolates to identify potential virulence factors associated with bacteraemia.

Results Based on single nucleotide polymorphism phylogenetics we identified two distinct ST93-IV clades circulating concurrently in Australia. One of the clades contained isolates primarily isolated in the northern regions of Australia whilst isolates in the second clade were distributed across the country. Analyses of the ST93-IV genome plasticity over a 15-year period (2002-2017) revealed an observed gain in accessory genes amongst the clone's population. The GWAS analysis on the bacteraemia isolates identified two genes that have also previously been associated to this kind of infection.

Conclusions The emergence of a ST93-IV clade containing additional virulence genes may explain the high prevalence of ST93-IV infections amongst the indigenous population living in the northern regions of Australia. In summary, this study has shown ST93-IV is evolving with multiple additional genes possibly contributing to its dominance in the Australian community.

Background

Over the last three decades, community-associated methicillin-resistant *Staphylococcus aureus* (CA-MRSA) has emerged globally. Although polyclonal, a small number of CA-MRSA clones are dominant in different regions of the world such as multilocus sequence type (ST) 8-IV (USA300) in North America, ST80-IV in Europe and Northern Africa, ST59-IV/V in Asia, ST772-V and ST22-IV in the Indian subcontinent, and ST30-IV in the West Pacific region (1). Transmission of the dominant clones in other regions has occurred, and characteristically they harbour the *lukS/F-PV* genes that encode the Pantone-Valentine leukocidin (PVL) toxin (2).

In Australia, the dominant CA-MRSA clone is PVL-positive ST93-IV(3). Colloquially known as the "Queensland CA-MRSA clone", ST93-IV was first described in the early 2000s. Although known to cause severe infections including necrotizing pneumonia, ST93-IV is typically associated with skin and soft tissue infections (4). Reported across Australia, the clone is frequently isolated in the indigenous Australian population where its dominance is believed to be linked to overcrowding (5), poor hygiene and healthcare (6). Using whole genome sequencing (WGS) and temporal and geographical analysis, ST93 has been shown to be an early diverging and recombinant lineage genetically related to ST59/ST121 and to an unknown *S. aureus* lineage that emerged in the 1970s in the North Western region of Australia (5). Although earlier studies into the genetic diversity of ST93 showed multiple rearrangements of the *spa*

sequence, the core regions of the genome were very stable (2). However in 2014, Stinear *et al.* suggested ST93 clone was under pressure for adaptive change due to a reduction in both exotoxin expression and oxacillin minimum inhibitory concentration (7).

To determine the association between gene content and disease, genome-wide association studies (GWAS) can be performed by analysing single nucleotide polymorphisms (SNPs), and the accessory genes provided by WGS data. For example, GWAS performed on isolates from children with acute *S. aureus* osteomyelitis identified an association in the number of virulence genes present and the severity of disease (8). In contrast, when applied to *S. aureus* bacteraemia isolates, no obvious associations in the number of virulence genes present in isolates from patients with and without *S. aureus* infective endocarditis were identified (9). GWAS can also be used to examine the evolution of a bacterial clone. For example recent GWAS performed on livestock-associated CC398 MRSA, showed the clone frequently lost antimicrobial resistance genes and acquired human specific virulence genes in relation to the origin of the host (10).

In this study, we performed GWAS on a collection of Australian ST93 MRSA bacteraemia isolates collected over a three-year period (2015-2017) and a collection of previously published ST93 MRSA genomes (2002-2012). Phylogenetic analysis of the genomes was performed by examining SNPs in the core genome and investigating the absence/presence of accessory genes. To determine what genetic traits may have assisted the ongoing transmission of ST93-IV in Australia we correlated the absence and presence of accessory genes in the ST93-IV genomes to time, location and whether they originated from a bloodstream infection.

Results

The 423 ST93-IV were isolated across Australia from the following states and mainland territories: Northern Territory (n=141), Queensland (n=98), New South Wales (n=64), Western Australia (n=54), Victoria (n=43), South Australia (n=19), Australia Capital Territory (n=3) and Tasmania (n=1). Overall, there were 302 bacteraemia and 121 non-bacteraemia isolates. The non-bacteraemia isolates were limited to four geographical regions: New South Wales, Victoria, Western Australia and Northern Territory.

Based on core genome SNPs, the rooted phylogeny based on 1383 SNPs depicted the ST93 population to cluster primarily in two main clades (Figure 1). Clade 1 contained 111 bacteraemia isolates predominantly from northern Australia whilst clade 2 contained 185 bacteraemia and 119 non-bacteraemia isolates collected across Australia.

Comparison between Principal Component Analysis (PCA) and Phylogenetic Clustering

By examining the presence and absence of accessory genes, PCA identified two distinct clusters (Figure 2). Isolates in the two PCA clusters correlated with isolates in the two SNP derived phylogenetic clades.

GWAS Comparison between Bacteraemia and Non-bacteraemia ST93 Isolates

GWAS revealed nine accessory genes correlated with the bacteraemia isolates ($p < 0.001$ and odds ratio > 1) (Table 1). However, as seven of the genes were clade 1 specific they were not considered bacteraemia factors.

Table 1: GWAS showing genes significantly correlating to bacteraemia using the presence (+) and absence (-) of each gene in 423 isolates (Bonferroni p value < 0.001 and a odds ratio > 1)

Gene	Function	Bacteraemia	Non
		Isolates	Bacteraemia
		N (%)	Isolates
			N (%)
<i>clfA</i>	Clumping factor A	240 (79.4)	53 (43.8)
<i>hsdM_1</i>	Type I restriction enzyme EcoKI M protein	269 (89)	55 (45)
<i>ohrR</i>	Organic hydroperoxide resistance transcriptional regulator*	103 (34.1)	0
<i>acul</i>	Putative acrylyl-CoA reductase Acul*	102 (33.7)	0
<i>ypuA</i>	Hypothetical protein*	105 (34.7)	1 (0.7)
<i>hutl_2</i>	Hypothetical protein*	101 (33.4)	1 (0.7)
<i>entE</i>	Enterotoxin type E*	101 (33.4)	0
<i>soj</i>	Chromosome partitioning ATPase*	101 (33.4)	1 (0.7)
<i>entA_2</i>	Enterotoxin type A*	99 (32.7)	0

The two non-clade specific genes that correlated with bacteraemia were *hsdM* (type I restriction enzyme EcoKI M protein) and *clfA* (clumping factor A) (Figure 1). Overall, of the 302 bacteraemia isolates, 76% (n=230) carried both genes; 16% (n=49) carried one of the genes, and the remaining 7% (n=23) carried neither gene. Only 43% and 45% of the non-bacteraemia isolates carried the *clfA* and *hsdM* genes respectively. The difference in the prevalence of *clfA* and *hsdM* in the bacteraemia and non-bacteraemia isolates was statistically significant (p value < 0.001).

The seven clade 1 specific accessory genes were *ohrR* (organic hydroperoxide resistance transcriptional regulator), *acul* (putative acrylyl-Coa reductase), *ypuA* (hypothetical protein), *hutl_2* (hypothetical protein), *entE* (enterotoxin E), *soj* (chromosome-partitioning ATPase) and *entA_2* (enterotoxin A) (Figure 1). Approximately 88% (n=98/111) of the clade 1 genomes harboured all seven genes, with seven isolates containing none of the seven genes. The seven genes were located on five different contigs, with *entE* and *acul* co-located with *soj* and *ohR* respectively. The higher read coverage of the seven genes relative to the chromosome suggests the seven genes have a mobile genetic element origin.

Genomic diversity of ST93 over Time and Location

No significant differences in the presence or absence of accessory genes over time or location were identified.

Recombination/rearrangement of the ST93 genome

When we analysed conserved gene neighbourhoods, we observed some genes that had re-arrangements correlating to bacteraemia. For example, some bacteraemia isolates contained different rearrangements of the *sftA* (DNA translocase), *sdrF* (serine-aspartate repeat-containing protein F), *pIs* (surface protein), and *setC* (sugar efflux transporter C) genes. Analysis of the contigs carrying these genes show *sftA* and *setC* are co-located, while *sdrF* and *pIs* are located separately.

Discussion

In the current study we have identified two distinct ST93-IV clades circulating concurrently in Australia. The identification of the two clades by SNP analysis of the core region was supported by the PCA based on the absence and presence of genes matrix. The clade 1 isolates were primarily isolated in the northern regions of Australia spread over three states/territories (Western Australia, Northern Territory and Queensland) whilst the clade 2 isolates were distributed across the country. Based on the genomic data of the van Hal *et al.* (5) historic ST93-IV isolates that were located at the root of the phylogenetic tree, we believe the two clades recently diverged from a common ancestor.

Clade 1 isolates differed from the clade 2 isolates by having acquired up to seven additional accessory genes. The known biological significance of these accessory genes varies. The *entA* and *entE* genes, encode the superantigen enterotoxins A and E respectively and play an important role in serious staphylococcal infections by triggering an overexpression of inflammatory mediators. The *ohrR* gene, which has previously been identified in *Pseudomonas aeruginosa* (11) and *Bacillus subtilis* (12), is known to increase an organism's resistance to organic hydroperoxides. The ability to resist peroxide provides the organism a growth advantage and increases its survival in host cells (13). The *soj* gene, a *parA* homologue involved in chromosome segregation during DNA replication, is not normally found in *S. aureus* (14). Typically, chromosome segregation in *S. aureus* is performed by the *parB* homologue *spo0J*, which was identified in all ST93-IV genomes. In *Bacillus subtilis*, *soj* and *spo0J* are present and work together to prevent premature midcell Z ring assembly (15). By having acquired *soj*, clade 1 isolates may have an advantage over non-clade 1 isolates as they have a more efficient DNA replication system. The roles of the three remaining accessory genes, *acul* (a putative protein), and *ypuA* and *hutI* (both hypothetical proteins) are not known. The addition of the seven accessory genes, which are likely to have originated on mobile genetic elements, may explain the high rates of ST93-IV skin infections amongst indigenous children living in the northern regions of Australia (16). Further studies are required to determine if clade 1 has become the predominant ST93-IV strain in the region's indigenous communities.

Based on the variability of the ST93-IV accessory genes over time and location we attempted to identify clade 1 or 2 specific subclades. Although minor accessory gene variations occurred in a small number of isolates, (for example, four isolates contained *qacA* [antiseptic resistance protein], *qacR* [HTH-type transcriptional regulator] and *tnsB* [transposon] which were all located on the same contig), no significant differences in the absence or presence of accessory genes related to specific subclades were observed.

GWAS for Bacteraemia vs Non-bacteraemia MRSA

In 2017 a GWAS performed by Lilje and colleagues was not able to identify genetic differences between *S. aureus* bacteraemia and non-bacteraemia genomes (9). Their results however may have been influenced by studying a variety of *S. aureus* lineages and clonal complexes. To identify if specific genetic factors are harboured by *S. aureus* bacteraemia genomes our study was limited to a single *S. aureus* lineage. GWAS identified two genes associated with the ST93-IV bacteraemic isolates. The *hsdM* gene has recently been shown to be a hotspot for chromosome rearrangements in staphylococcus causing phenotype switching associated with persistent infections (17). The *clfA* gene, which mediates staphylococcal binding to fibrin coated surfaces, has previously been linked to bacteraemia (18).

Chromosome rearrangements of genes may lead to altered gene expression (19). The 23 bacteraemic genomes that did not harbor *hsdM* and *clfA* all carried rearrangements of the *pls*, *sdrF*, *setC* and *sftA* genes. The *pls* and *sdrF* genes encode surface proteins. *pls*, which mediates bacterial aggregation and binding to glycolipids and human epithelial cells (20, 21), has been shown in mice models to be an important factor in causing sepsis (22). SdrF, which is a microbial surface components recognising adhesive matrix molecule (MSCRAMM), allows staphylococcus to attach to and colonise host cells (23). The *sdr* gene in the ST93 genome JKD6159 has previously been reported to be the most diverse amongst different *S. aureus* clones, suggesting acquisition by horizontal transfer has occurred. In the Huping *et al.* study the *sdr* in the ST93 genome was classified as *sdrC* (24). However, an updated annotation database has identified the gene as *sdrF* which had previously only been reported in *S. epidermidis*. SdrF adheres to human keratinocytes and epithelial cells facilitating *S. epidermidis* colonisation of the skin (25). The *sftA* gene encodes for a DNA translocase. DNA translocases play an important role in allowing an organism to convert to the "L-form" cell wall deficient state which is resistant to cell wall-targeting antibiotics (26). DNA translocases have also been shown to be necessary in the formation of biofilms (27). Although the role of SftA in *S. aureus* is believed to be performed by SpoIIIE, in *Bacillus subtilis* SftA and SpoIIIE have been shown to have different functions during cell division. SftA aids in moving DNA from the closing septum while SpoIIIE translocates septum-entrapped DNA (28). We therefore hypothesise the rearrangement of *sftA* in some ST93-IV causing bacteraemia may provide the organism with a more efficient cell division pathway. *setC*, which encodes a sugar efflux transporter, is not a known bacteraemia factor. However, the gene is co-located with *sftA*.

Conclusion

GWAS is a powerful tool to identify associations using large datasets however there are limitations such as the patient's age and prior medical condition which are factors associated with bacteraemia. In the current study we have identified accessory genes and gene rearrangements that have strong associations with ST93-IV bacteraemia. However, to confirm this association there is a need to expand the dataset to include more ST93 non bacteraemia isolates. Furthermore, to determine if these genes are bacteraemia determinants other *S. aureus* lineages should be examined. Phylogenetic analysis has shown ST93-IV has recently gained accessory virulence genes which may be contributing to the clone's persistence in the Australian indigenous communities.

Methods

Bacterial Strains and Genome Assembly

A total of 300 ST93 MRSA bacteraemia isolates were identified in the 2015 (29), 2016 (30) and 2017 (31) Australian Group on Antimicrobial Resistance (AGAR) Australian *Staphylococcus aureus* Sepsis Outcome Programs (ASSOPs). All isolates collected were from patients with systemic infections. As part of ASSOP, MRSA isolates were referred to a central reference laboratory where genomic libraries were prepared using the Illumina Nextera[®] XT DNA Library Prep Kit (Illumina, United States) according to the manufacturer's protocol. WGS was performed on the Miseq or Nextseq platforms using the Miseq Reagent Kit V3 (600 cycle) and the Nextseq 500/550 Mid Output Kit V2.5 (300 cycles), respectively. The raw sequence reads were assembled *de novo* using SPAdes V3.12 (32). Sequencing quality control was determined based on average sequencing depth. Thirty-one had genomes less with than 40x coverage and therefore were excluded. The MLST profiles of the remaining 269 genomes were determined using the mlst tool described by Seeman *et al.* (33).

In addition to the 269 ASSOP ST93 MRSA, whole genome sequences for 154 ST93 MRSA collected between 2002 to 2012 from Van Hal *et al.* study (5) were included (Supplementary Table 1).

All sequence data obtained from this study were deposited to the NCBI Sequence Read Archive under BioProject ID PRJNA644215.

Phylogenetic Analysis

Using the chromosome of *S. aureus* CC398 reference strain S0395 (GenBank accession ID AM990992) as the reference genome, the bacterial variant calling tool snippy V4.1.0 (34) was used to extract SNPs from the core genome. The 423 ST93 genomes were used to generate a rooted maximum parsimony phylogenetic tree using MEGA V10.1.7 (35). Phylogenetic clades were defined as a cluster of isolates sharing multiple common SNP mutations. The iTOL V3 web service was used to visualise the phylogenetic tree and the corresponding metadata (36).

Genome-Wide Association Study (GWAS)

Genes from the 423 assembled *S. aureus* genome sequences were annotated with Prokka V1.13 (37) and the pan-genome was extracted by Roary V3.12.0 (38) using the -s option of no paralog splitting. The pan-genome matrix of gene presence or absence in each genome was used as input for Scoary V1.6.16 (39) with the following traits; SNP phylogeny clades, location (states and territories), year of isolation and whether the isolate was from a bloodstream infection. Genes that scored a Bonferroni corrected p value $\leq 10^{-5}$ and odds ratio > 1 were further investigated similar to the method described by Arnoud H. M. van Vliet (40). In addition to Scoary analysis, principal component analysis (PCA) of binomial variables on the pan-genome matrix was performed for determination of association with the statistical package R version 3.5.1 (41) and ggplot2 V3.2.1 to confirm relationships between genes identified in Scoary and traits. A chi-square test of independence was performed to examine the significance of any relationship. An alpha level of 0.05 was used as a significance criterion for all statistical tests.

The default Roary parameters using paralogous gene splitting was used to detect rearranged genes. Sequences upstream and downstream of the rearranged genes were analysed to confirm recombination.

List Of Abbreviations

AGAR: Australian Group on Antimicrobial Resistance

ASSOPs: Australian *Staphylococcus aureus* Sepsis Outcome Programs

CA-MRSA: Community-Associated Methicillin-Resistant *Staphylococcus aureus*

CC: Clonal Complex

DNA: deoxyribonucleic acid

GWAS: Genome Wide Association Study

MSCRAMM: microbial surface components recognising adhesive matrix molecule

MLST: Multi Locus Sequence typing

PCA: Principal Component Analysis

PVL: Pantone-Valentine Leukocidin

SNP: Single Nucleotide Polymorphisms

ST: Sequence Type

WGS: Whole Genome Sequencing

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The data that support the findings of this study are openly available on the SRA database under Bioproject: PRJNA644215.

Competing Interest

The authors declare that they have no competing interests

Funding

This work received no specific grant from any funding agency.

Authors' contributions

SP conceived of and designed the study and performed the literature search, generated the figures and tables, and wrote the manuscript. DD, SS and MS collected and analyzed the data, and critically reviewed the manuscript. GC supervised the study and with SM reviewed the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

References

1. Lakhundi S, Zhang K. Methicillin-Resistant *Staphylococcus aureus*: Molecular Characterization, Evolution, and Epidemiology. Clin Microbiol Rev. 2018;31(4).
2. Coombs GW, Goering RV, Chua KY, Monecke S, Howden BP, Stinear TP, et al. The molecular epidemiology of the highly virulent ST93 Australian community *Staphylococcus aureus* strain. PLoS One. 2012;7(8):e43037.
3. Coombs GW, Daley DA, Mowlaboccus S, Lee YT, Pang S, Australian Group on Antimicrobial R. Australian Group on Antimicrobial Resistance (AGAR) Australian *Staphylococcus aureus* Sepsis Outcome Programme (ASSOP) Annual Report 2018. Commun Dis Intell (2018). 2020;44.
4. Peleg AY, Munckhof WJ. Fatal necrotising pneumonia due to community-acquired methicillin-resistant *Staphylococcus aureus* (MRSA). Med J Aust. 2004;181(4):228-9.

5. van Hal SJ, Steinig EJ, Andersson P, Holden MTG, Harris SR, Nimmo GR, et al. Global Scale Dissemination of ST93: A Divergent *Staphylococcus aureus* Epidemic Lineage That Has Recently Emerged From Remote Northern Australia. *Front Microbiol.* 2018;9:1453.
6. Uhlemann AC, Otto M, Lowy FD, DeLeo FR. Evolution of community- and healthcare-associated methicillin-resistant *Staphylococcus aureus*. *Infect Genet Evol.* 2014;21:563-74.
7. Stinear TP, Holt KE, Chua K, Stepnell J, Tuck KL, Coombs G, et al. Adaptive change inferred from genomic population analysis of the ST93 epidemic clone of community-associated methicillin-resistant *Staphylococcus aureus*. *Genome Biol Evol.* 2014;6(2):366-78.
8. Collins A, Wakeland EK, Raj P, Kim MS, Kim J, Tareen NG, et al. The impact of *Staphylococcus aureus* genomic variation on clinical phenotype of children with acute hematogenous osteomyelitis. *Heliyon.* 2018;4(6):e00674.
9. Lilje B, Rasmussen RV, Dahl A, Stegger M, Skov RL, Fowler VG, et al. Whole-genome sequencing of bloodstream *Staphylococcus aureus* isolates does not distinguish bacteraemia from endocarditis. *Microb Genom.* 2017;3(11).
10. Sieber RN, Larsen AR, Urth TR, Iversen S, Moller CH, Skov RL, et al. Genome investigations show host adaptation and transmission of LA-MRSA CC398 from pigs into Danish healthcare institutions. *Sci Rep.* 2019;9(1):18655.
11. Atichartpongkul S, Fuangthong M, Vattanaviboon P, Mongkolsuk S. Analyses of the regulatory mechanism and physiological roles of *Pseudomonas aeruginosa* OhrR, a transcription regulator and a sensor of organic hydroperoxides. *J Bacteriol.* 2010;192(8):2093-101.
12. Fuangthong M, Atichartpongkul S, Mongkolsuk S, Helmann JD. OhrR is a repressor of ohrA, a key organic hydroperoxide resistance determinant in *Bacillus subtilis*. *J Bacteriol.* 2001;183(14):4134-41.
13. Li X, Tao J, Han J, Hu X, Chen Y, Deng H, et al. The gain of hydrogen peroxide resistance benefits growth fitness in mycobacteria under stress. *Protein Cell.* 2014;5(3):182-5.
14. Murray H, Errington J. Dynamic control of the DNA replication initiation protein DnaA by Soj/ParA. *Cell.* 2008;135(1):74-84.
15. Hajduk IV, Mann R, Rodrigues CDA, Harry EJ. The ParB homologs, Spo0J and Noc, together prevent premature midcell Z ring assembly when the early stages of replication are blocked in *Bacillus subtilis*. *Mol Microbiol.* 2019;112(3):766-84.
16. Bowen AC, Mahe A, Hay RJ, Andrews RM, Steer AC, Tong SY, et al. The Global Epidemiology of Impetigo: A Systematic Review of the Population Prevalence of Impetigo and Pyoderma. *PLoS One.* 2015;10(8):e0136789.
17. Guerillot R, Kostoulias X, Donovan L, Li L, Carter GP, Hachani A, et al. Unstable chromosome rearrangements in *Staphylococcus aureus* cause phenotype switching associated with persistent infections. *Proc Natl Acad Sci U S A.* 2019;116(40):20135-40.
18. Perez-Montarelo D, Viedma E, Larrosa N, Gomez-Gonzalez C, Ruiz de Gopegui E, Munoz-Gallego I, et al. Molecular Epidemiology of *Staphylococcus aureus* Bacteremia: Association of Molecular Factors With the Source of Infection. *Front Microbiol.* 2018;9:2210.

19. Raeside C, Gaffe J, Deatherage DE, Tenailon O, Briska AM, Ptashkin RN, et al. Large chromosomal rearrangements during a long-term evolution experiment with *Escherichia coli*. mBio. 2014;5(5):e01377-14.
20. Huesca M, Peralta R, Sauder DN, Simor AE, McGavin MJ. Adhesion and virulence properties of epidemic Canadian methicillin-resistant *Staphylococcus aureus* strain 1: identification of novel adhesion functions associated with plasmin-sensitive surface protein. J Infect Dis. 2002;185(9):1285-96.
21. Roche FM, Meehan M, Foster TJ. The *Staphylococcus aureus* surface protein SasG and its homologues promote bacterial adherence to human desquamated nasal epithelial cells. Microbiology. 2003;149(Pt 10):2759-67.
22. Josefsson E, Juuti K, Bokarewa M, Kuusela P. The surface protein Pls of methicillin-resistant *Staphylococcus aureus* is a virulence factor in septic arthritis. Infect Immun. 2005;73(5):2812-7.
23. Foster TJ, Geoghegan JA, Ganesh VK, Hook M. Adhesion, invasion and evasion: the many functions of the surface proteins of *Staphylococcus aureus*. Nat Rev Microbiol. 2014;12(1):49-62.
24. Xue H, Lu H, Zhao X. Sequence diversities of serine-aspartate repeat genes among *Staphylococcus aureus* isolates from different hosts presumably by horizontal gene transfer. PLoS One. 2011;6(5):e20332.
25. Trivedi S, Uhlemann AC, Herman-Bausier P, Sullivan SB, Sowash MG, Flores EY, et al. The Surface Protein SdrF Mediates *Staphylococcus epidermidis* Adherence to Keratin. J Infect Dis. 2017;215(12):1846-54.
26. Glover WA, Yang Y, Zhang Y. Insights into the molecular basis of L-form formation and survival in *Escherichia coli*. PLoS One. 2009;4(10):e7316.
27. Chang Y, Gu W, Zhang F, McLandsborough L. Disruption of Imo1386, a putative DNA translocase gene, affects biofilm formation of *Listeria monocytogenes* on abiotic surfaces. Int J Food Microbiol. 2013;161(3):158-63.
28. Kaimer C, Gonzalez-Pastor JE, Graumann PL. SpoIIIE and a novel type of DNA translocase, SftA, couple chromosome segregation with cell division in *Bacillus subtilis*. Mol Microbiol. 2009;74(4):810-25.
29. Coombs GW, Daley DA, Lee YT, Pang S, Bell JM, Turnidge JD, et al. Australian Group on Antimicrobial Resistance (AGAR) Australian *Staphylococcus aureus* Sepsis Outcome Programme (ASSOP) Annual Report 2015. Commun Dis Intell (2018). 2018;42.
30. Coombs GW, Daley DA, Lee YT, Pang S, Australian Group on Antimicrobial R. Australian Group on Antimicrobial Resistance (AGAR) Australian *Staphylococcus aureus* Sepsis Outcome Programme (ASSOP) Annual Report 2016. Commun Dis Intell (2018). 2018;42.
31. Coombs GW, Daley DA, Lee YT, Pang S. Australian Group on Antimicrobial Resistance (AGAR) Australian *Staphylococcus aureus* Sepsis Outcome Programme (ASSOP) Annual Report 2017. Commun Dis Intell (2018). 2019;43.

32. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455-77.
33. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics.* 2010;11:595.
34. Seemann T. Snippy: fast bacterial variant calling from NGS reads. 2015.
35. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol.* 2018;35(6):1547-9.
36. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 2016;44(W1):W242-5.
37. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30(14):2068-9.
38. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics.* 2015;31(22):3691-3.
39. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* 2016;17(1):238.
40. van Vliet AH. Use of pan-genome analysis for the identification of lineage-specific genes of *Helicobacter pylori*. *FEMS Microbiol Lett.* 2017;364(2).
41. R_core_Team. R: A Language and Environment for Statistical Computing. 2019.

Figures

Tree scale: 100

Bacteremic

- Yes
- No

Genes

- Present
- Absent

Location

- ACT
- NSW
- NT
- QLD
- SA
- WA
- VIC
- TAS

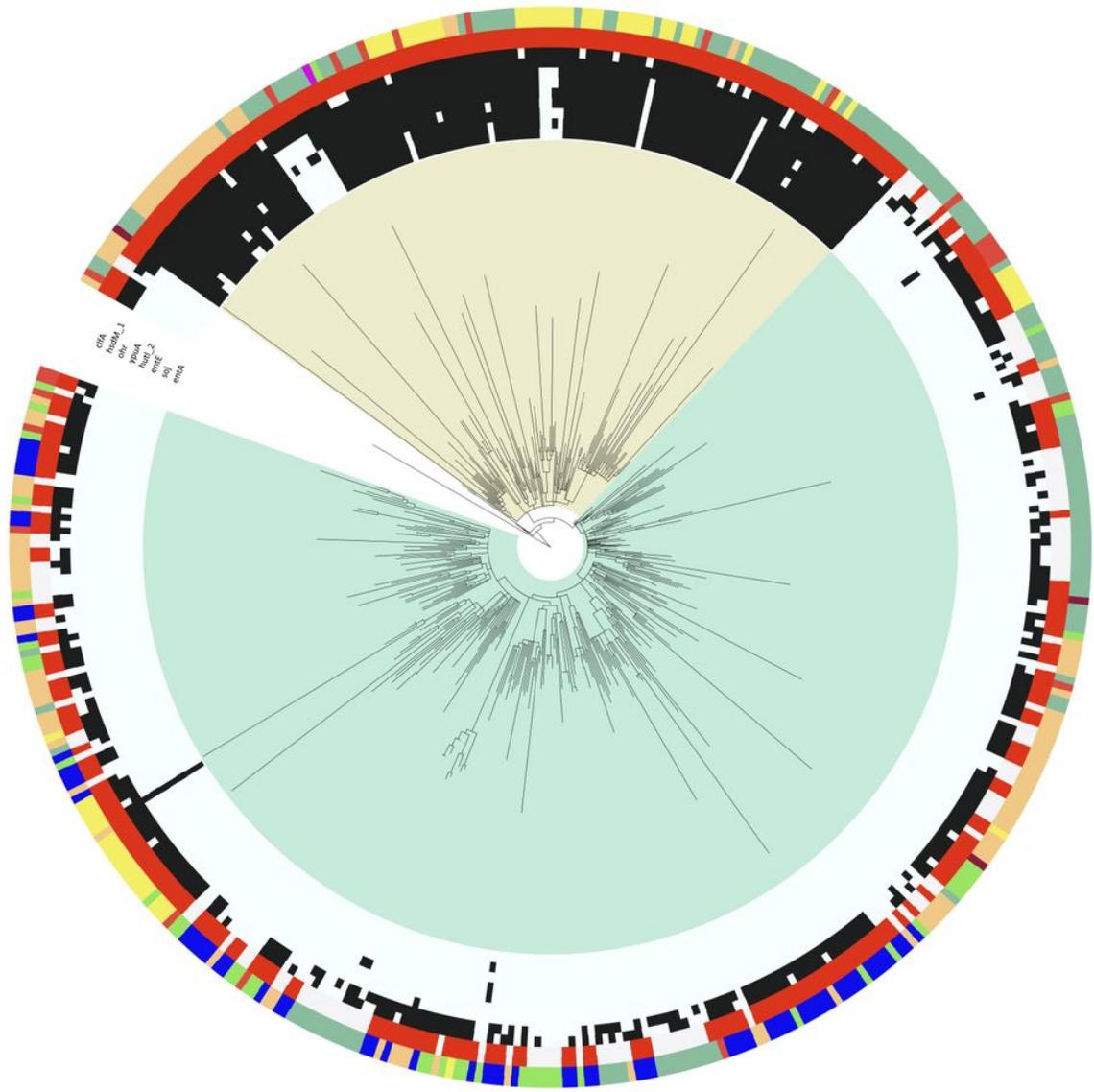


Figure 1

Rooted Phylogenetic tree of 423 ST93 *S. aureus* bacteraemia and non-bacteraemia genomes represented as red and white respectively (outer ring). Location is represented by the abbreviation of Australian states and territories: Australian Capital Territory (ACT), New South Wales (NSW), Northern Territory (NT), Queensland (Qld), South Australia (SA), Western Australia (WA), Victoria (Vic) and Tasmania (Tas). Genes present (black) and absence (grey) that correlate with bacteraemia are listed in the order (outer to inner); *clfA*, *hsdM_1*, *ohrR*, *acul*, *ypuA*, *hutl_2*, *entE*, *soj* and *entA_2*.

Tree scale: 100

Bacteremic

- Yes
- No

Genes

- Present
- Absent

Location

- ACT
- NSW
- NT
- QLD
- SA
- WA
- VIC
- TAS

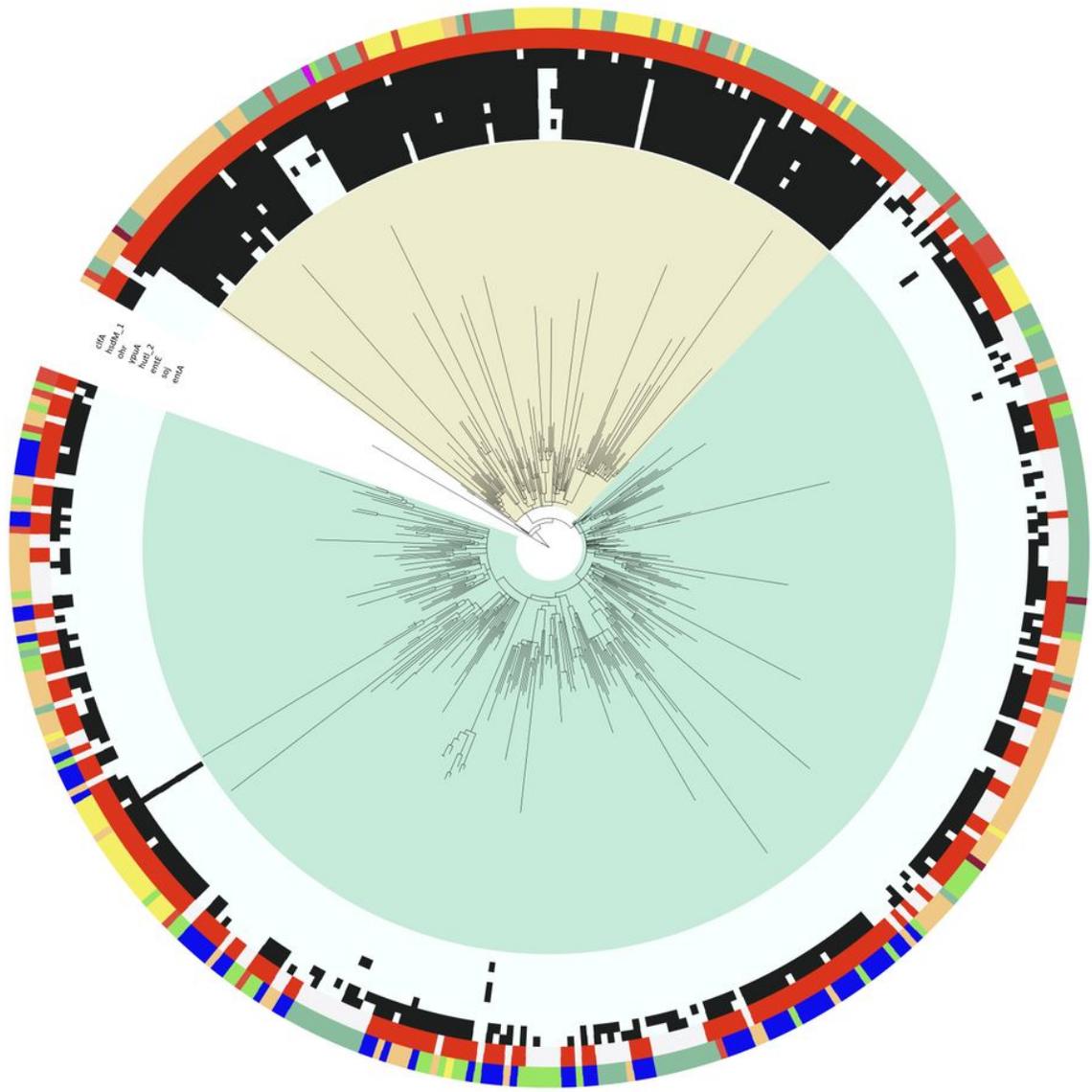


Figure 1

Rooted Phylogenetic tree of 423 ST93 *S. aureus* bacteraemia and non-bacteraemia genomes represented as red and white respectively (outer ring). Location is represented by the abbreviation of Australian states and territories: Australian Capital Territory (ACT), New South Wales (NSW), Northern Territory (NT), Queensland (Qld), South Australia (SA), Western Australia (WA), Victoria (Vic) and Tasmania (Tas). Genes present (black) and absence (grey) that correlate with bacteraemia are listed in the order (outer to inner); *clfA*, *hsdM_1*, *ohrR*, *acul*, *ypuA*, *hutl_2*, *entE*, *soj* and *entA_2*.

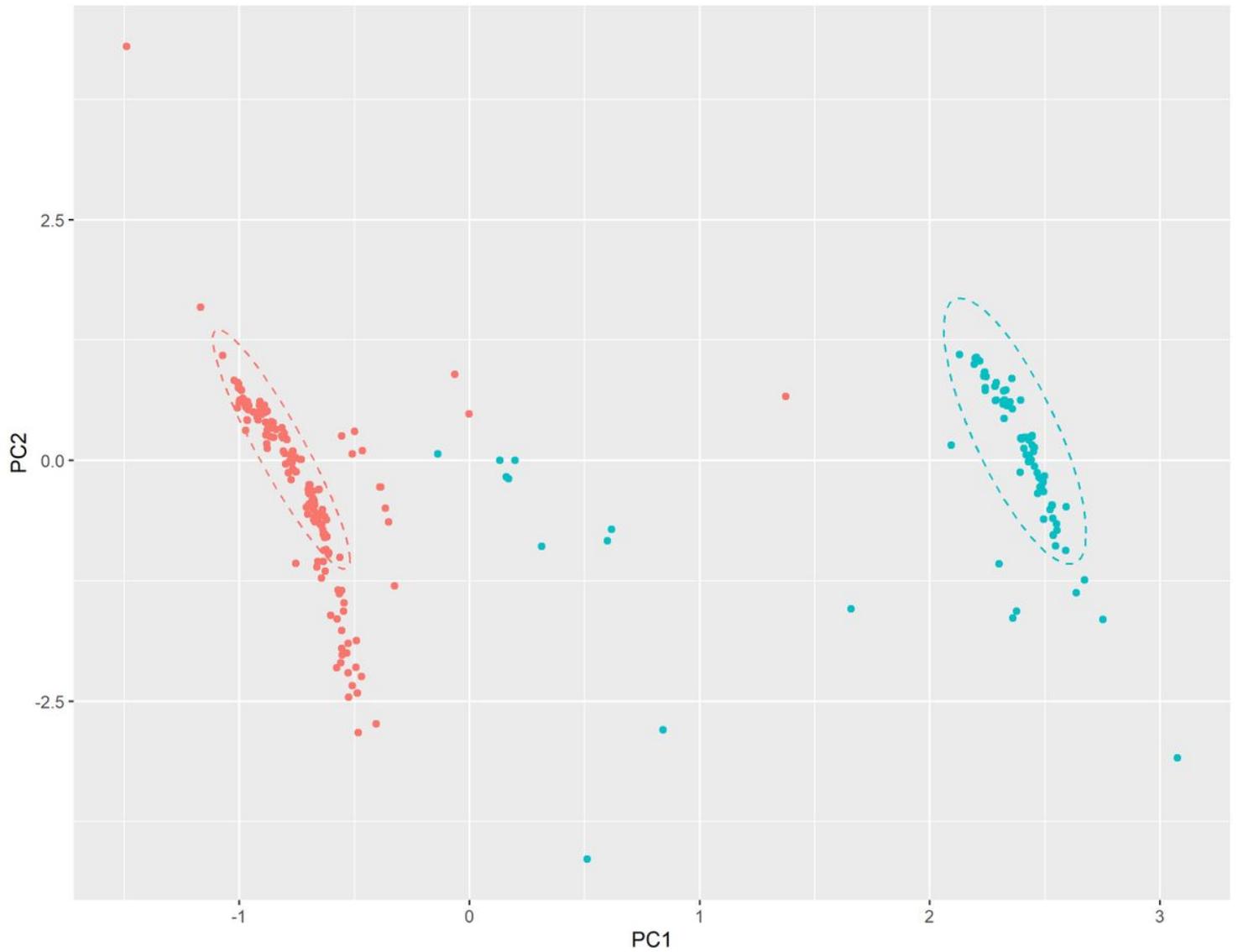


Figure 2

Principal Component Analysis of pan-genome gene matrix of ST93-IV isolates. The green coloured dots represent isolates in clade 1, while the red coloured dots represent isolates in clade 2. Non-clade 1 and 2 isolates are grey coloured dots. The ellipse is generated using the multivariate t- distribution with CI= 95%.

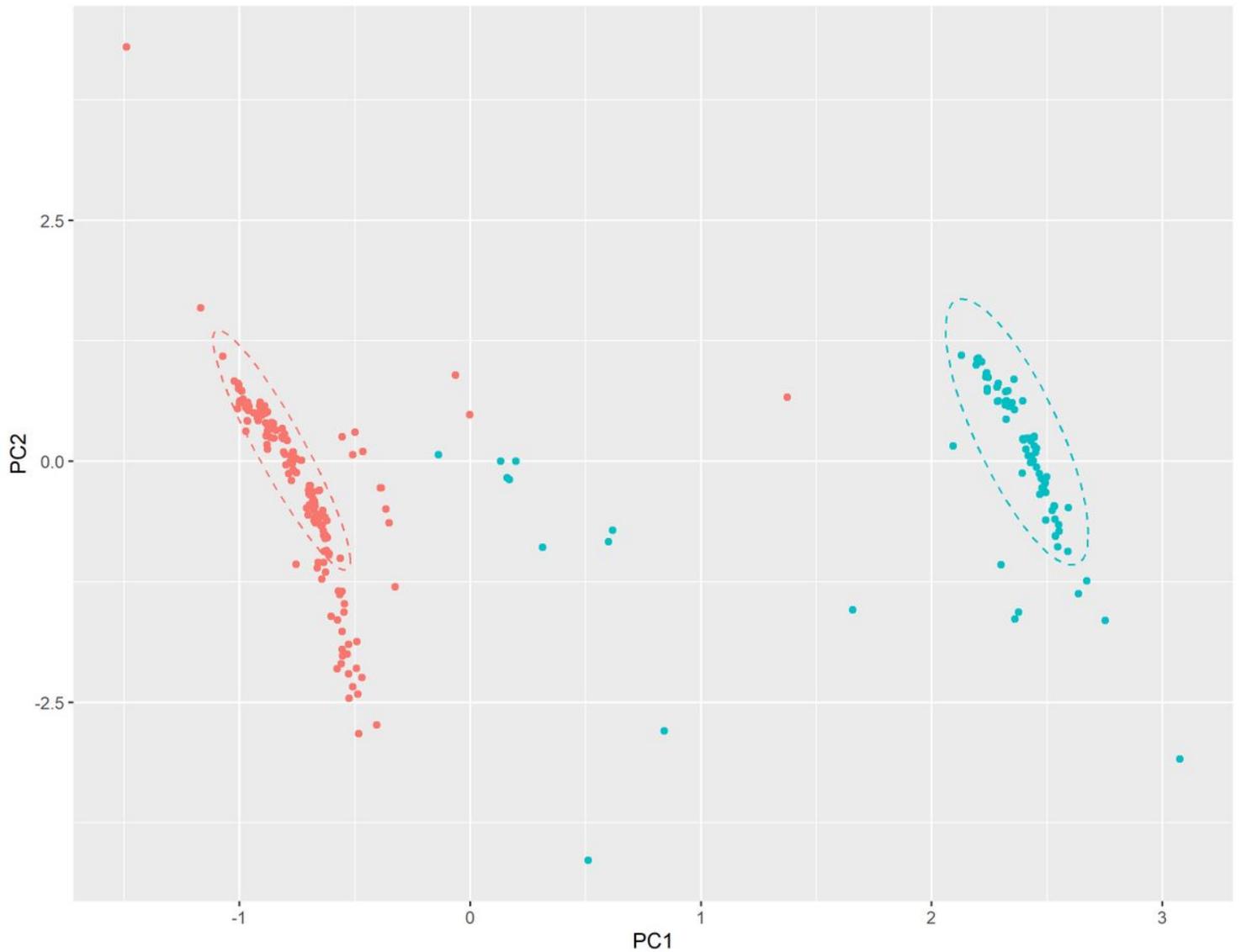


Figure 2

Principal Component Analysis of pan-genome gene matrix of ST93-IV isolates. The green coloured dots represent isolates in clade 1, while the red coloured dots represent isolates in clade 2. Non-clade 1 and 2 isolates are grey coloured dots. The ellipse is generated using the multivariate t- distribution with CI= 95%.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1strainlist.xlsx](#)
- [SupplementaryTable1strainlist.xlsx](#)