A phylogenetically distinct *Staphylococcus aureus* lineage prevalent among Indigenous communities in Northern Australia.

Short title: A divergent *Staphylococcus aureus* lineage

JacklynWei Sze Ng[1], Deborah C. Holt[1], Rachael A. Lilliebridge[1], Alex J. Stephens[2], Flavia Huygens[2], Steven Y.C. Tong[1], Bart J. Currie[1], Philip M. Giffard[1*]

[1]Menzies School of Health Research, Institute of Advanced Studies, Charles Darwin University, Darwin, Northern Territory, Australia

[2]Institute of Health and Biomedical Innovation, Queensland University of Technology, Kelvin Grove, Brisbane, Queensland, Australia

*Corresponding author:

Philip Giffard

Menzies School of Health Research

PO Box 41096

Casuarina

0811 Northern Territory

Australia

E-mail: phil.giffard@menzies.edu.au

Tel: +61(8) 8922 8196

Fax: +61(8) 8927 5187

Abstract

The aim was to determine the evolutionary position of the *Staphylococcus aureus* CC75 that is prevalent in tropical north Australia. Sequencing of *gap*, *rpoB*, *sodA, tuf*, and *hsp60* and the multilocus sequence typing loci revealed a clear separation between conventional *S. aureus* and CC75, and significant diversity within CC75.

To date, more than fifty species and subspecies of the genus *Staphylococcus* have been described (5, 13) The most prominent species, *Staphylococcus aureus,* is a major human pathogen that can cause a wide variety of hospital and community acquired infections (9). Strains carrying the mobile genetic element SCC*mec* are resistant to a broad range of β-lactam antibiotics and are termed methicillin-resistant *S. aureus* (MRSA).

Evolutionary relationships within the genus *Staphylococcus* have previously been determined using homologous housekeeping genes (3, 6, 7, 14, 15, 20), and *S. aureus* has been shown to be well separated from the other species. Multilocus sequence typing (MLST) of large numbers of *S. aureus* isolates has revealed that the MLST loci are in general no more than 3% divergent (4). This is considerably less than other bacterial species (2, 8). Therefore, the picture that has emerged is that extant *S. aureus* are descended from a relatively recent common ancestor, and there is a wide evolutionary gulf between *S. aureus* and any other species.

Recent evidence indicates that this model is an oversimplification. Surveillance of staphylococcal carriage and disease in the tropical north of the Northern Territory of Australia yielded methicillin-susceptible and methicillin-resistant *Staphylococcus* isolates that are highly divergent from *S. aureus* at the MLST loci, and are therefore difficult to classify. These isolates were shown to be closely related to each other, and were designated clonal complex 75 (CC75) (16). They accounted for 23% of all isolates and 70% of MRSA isolates recovered during a community study of impetigo (16), and clinical isolates were associated with a variety of disease states (22). Similarly to other

community acquired MRSA, the majority of SCC*mec* cassettes found in CC75 were type IV (16).

The aims of this study were to determine the evolutionary position of CC75 within the genus *Staphylococcus*, and to test the hypothesis that CC75 encompasses significant diversity, even within a single geographical region with a small human population. The reporting of this study was complicated by the uncertain taxonomic status of CC75. In this manuscript we have regarded CC75 as belonging to *S. aureus*, but made a comment regarding taxonomic status when discussing the results. In the interests of clarity, we have used the term "CC75"to refer to this lineage, and used the term "conventional *S. aureus"* to denote "*S. aureus* not including CC75".

Five CC75 isolates were included in this study. HS2, HS42 and M34 were derived from surveillance of skin lesions in remote Indigenous communities in the north of the Northern Territory, Australia, with M34 being included in a previously reported study (16). SCC1302 and SCC1119 are clinical isolates obtained from the Pathology Department of Royal Darwin Hospital (18). They were identified as belonging to CC75 using the SNP-based method described by McDonald et al (16), and all are methicillin resistant . The isolates were routinely propagated on horse blood agar.

The evolutionary position of CC75 was deduced from the sequences of six gene fragments previously used to study the relationships between *Staphylococcus* species (3, 6, 7, 14, 15, 20). Sequences were obtained from HS2, HS42, M34 and SCC1302. To extract genomic DNA, colonies were picked into 5mL of Todd-Hewitt yeast broth and

grown at 37^{o}C overnight with agitation. DNA was extracted from 1mL of pelleted cells using a QIAamp DNA mini kit (QIAGEN) according to the manufacturer's instructions, using 100ug/mL lysostaphin (Sigma). Fragments of the 16S rRNA, *gap*, *rpoB*, *sodA*, *tuf* and *hsp60* genes were amplified and sequenced. PCR was carried out using oligonucleotides listed in Table 1 using the reaction conditions specified in the references to the primer sequences. PCR products were purified and sequenced by either Bioscience North Australia, Charles Darwin University, Darwin, Australia, or Macrogen Inc. Seoul, Korea.

The CC75 sequences were compared to the corresponding sequences from five conventional *S. aureus* strains and a representative of seven other *Staphylococcus* species (Table 2). In the case of 16S rRNA, no differences between CC75 and conventional *S. aureus* were found over the region sequenced. For the remaining genes fragments, sequences were aligned and neighbour-joining trees deduced using the Mega software package (21) (Figure 1). It can be seen that there is a consistent pattern that CC75 is significantly diverged from conventional *S. aureus,* however CC75 and conventional *S. aureus* are each other's closest relatives. There were an average of eight times as many polymorphisms separating CC75 and conventional *S. aureus* as there were separating the most divergent conventional *S. aureus* from each other. It was concluded that the most recent common ancestor of CC75 and conventional *S. aureus* existed considerably before the most recent common ancestor of conventional *S aureus,* but considerably after the most recent common ancestor of conventional *S. aureus* and any other recognized staphylococcal species. Also of interest was the presence of similar levels of diversity within the CC75 and conventional *S. aureus* sequences.

In order to further examine the relationship between CC75 and conventional *S. aureus*, the sequences of their MLST loci were compared. Neighbour-joining trees were deduced from the allele sequences in the *S. aureus* MLST database. The tree for the *yqiL* locus is shown in Figure 2 and the trees from the other loci are essentially equivalent (with the exception of the *aroE* locus- see below). The *yqiL* alleles associated with CC75 isolates (alleles 49, 105 and 131) are clearly removed from all the other alleles. It can be seen that this confirms that CC75 is distinct from conventional *S. aureus*.

There are currently five sequence types (STs) in the *S. aureus* MLST database (http://saureus.mlst.net) that clearly belong to CC75. These are ST75, ST850, ST883 ST1223 and ST1304. There are apparent anomalies in that all of these STs apart from ST1223 possess six "CC75" alleles but a conventional *S. aureus aroE* allele (the precise allele varies from ST to ST.) We were completely unable to amplify a product for the *aroE* locus for any CC75 isolate using the previously reported *S. aureus* MLST primers (4) and also had difficultly amplifying some of the other loci on several occasions. As a result we designed an alternative primer set for MLST of CC75 isolates (Table 1), utilizing sequence flanking the alleles, or, in the case of the *aroE* locus, the sequence of the ST1223 *aroE* allele, to design the primers. We then performed MLST analysis on four CC75 isolates (HS2, M34, SCC1119 and SCC1302). The sequence of each locus was compared to that in the *S. aureus* MLST database. None of the isolates corresponded to known sequence types, with between one and three loci of each isolate having an allele sequence not present on the MLST database (Table 3). The two CC75 isolates from community surveillance (HS2 and M34) possessed different alleles to each other at all of

the seven MLST loci. The two clinical CC75 isolates possessed the same alleles as each other, but differed from one of the community isolates at five loci and the other community isolate at all seven loci. Thus there is significant diversity in alleles even within this small number of CC75 isolates. The *aroE* sequences were closely related to the ST1223 *aroE* allele, with between 0 and 5 polymorphisms. We consider it possible that the *aroE* alleles in CC75 sequence types ST75, ST850, ST883 and ST1304 in the MLST database are not correct, and are due to the conjunction of primer template mismatches, and PCR contamination. We were unable to submit our new alleles to the MLST web site because of the truncated *aroE* sequences. However, they have been submitted to Genbank (accession numbers xxxx, xxxx etc.). There is a clear case for a separate MLST scheme for CC75.

The two ST75, and single ST850 and ST883 isolates listed in the MLST database all originated in northern Australia (18, 19), and it is stated on the MLST web-site that the ST1304 isolate is from the state of Western Australia, has a very large area in the tropics. Thus until recently, to the best of our knowledge CC75 was confined to tropical Northern Australia. However the single ST1223 isolate originates from Cambodia. One unexpected finding was the close relationship between ST1223 and one of the two Northern Australia community isolates from this study. These differ at just one of the seven MLST loci. This is in marked contrast to the considerable difference between the two community isolates (Table 3). This suggests that international dissemination of clones or clonal complexes within CC75 takes place. While we have grouped these isolates as ‘clonal complex 75’, there is in fact much greater variation within this group

than as classically defined for a clonal complex comprising of a founder and its single and double locus variants.

We have been unable to find any phenotype that is diagnostic for CC75. The CC75 isolates were tested using the API Staph biochemical test strips (biomérieux – Australia Pty. Ltd, Balkam Hills, NSW), and yielded results typical for conventional *S. aureus* . It is the opinion of the authors that it would be logical to regard CC75 as a new sub-species of *S. aureus*. However, the lack of a known diagnostic phenotype may make acceptance of a valid subspecies name difficult under current taxonomic guidelines. However, the explosion of genome sequence data will likely impact on taxonomic guidelines for bacteria in the near future, and we anticipate that definition of a valid sub-species name for CC75 on the basis of multilocus sequence data only, or in combination with geographical range, will at some point become possible.

The significance of this study lies not just in the characterization of a new phylogenetic group of *Staphylococcus*, but also in that CC75 appears to be a highly abundant human pathogen in some regions of the world. The level of diversity in CC75 excludes the notion that CC75 in northern Australia represents the rapid radiation of a clone. Rather, it appears that CC75 in Northern Australia is long established and endemic. The geographical range of CC75 remains unknown, but its detection in Cambodia suggests it may be widely distributed but largely unnoticed. This is possibly because it is indistinguishable from conventional *S. aureus* on the basis of conventional microbiological diagnostic methods and biochemical tests, and possibly because its primary habitat is less developed regions of the tropical Asia Pacific region. Technical

difficulties in performing MLST on CC75 isolates using conventional MLST primers may also contribute to under-reporting.

This group of bacteria is significant as it is genotypically distinct from conventional *S. aureus*, but is still able to cause *S. aureus*-like disease. More detailed comparative genomic analysis of CC75, conventional *S. aureus* and other less pathogenic Staphylococcal species may better reveal the core genomic features which contribute to the pathogenicity of conventional *S. aureus* and CC75 compared to other staphylococcal species.

Acknowledgments

This work was supported by the Menzies School of Health Research, and the Cooperative Research Centres program of the Australian Federal Government. The authors thank the staff of the Pathology Department at Royal Darwin Hospital for the diagnosis, propagation and donation of clinical isolates.

References

1. **Baba, T., T. Bae, O. Schneewind, F. Takeuchi, and K. Hiramatsu.** 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:**300-10.

2. **Boerlin, P., P. Kuhnert, D. Hussy, and M. Schaellibaum.** 2003. Methods for Identification of *Staphylococcus aureus* Isolates in Cases of Bovine Mastitis. J. Clin. Microbiol. **41:**767-771.

3. **Drancourt, M., and D. Raoult.** 2002. *rpoB* Gene Sequence-Based Identification of *Staphylococcus* Species. J. Clin. Microbiol. **40:**1333-1338.

4. **Enright, M. C., N. P. J. Day, C. E. Davies, S. J. Peacock, and B. G. Spratt.** 2000. Multilocus Sequence Typing for Characterization of Methicillin-Resistant and Methicillin-Susceptible Clones of *Staphylococcus aureus*. J. Clin. Microbiol. **38:**1008-1015.

5. **Euzeby, J. P.** 1997. List of Bacterial Names with Standing in Nomenclature: a Folder Available on the Internet. Int J Syst Bacteriol **47:**590-592.

6. **Ghebremedhin, B., F. Layer, W. Konig, and B. Konig.** 2008. Genetic Classification and Distinguishing of Staphylococcus Species Based on Different Partial *gap*, 16S rRNA, *hsp60*, *rpoB*, *sodA*, and *tuf* Gene Sequences. J. Clin. Microbiol. **46:**1019-1025.

7. **Goh, S., S. Potter, J. Wood, S. Hemmingsen, R. Reynolds, and A. Chow.** 1996. HSP60 gene sequences as universal targets for microbial species identification: studies with coagulase-negative staphylococci. J. Clin. Microbiol. **34:**818-823.

8. **Grady, R., D. Blanc, P. Hauser, and J. Stanley.** 2001. Genotyping of European isolates of methicillin-resistant *Staphylococcus aureus* by fluorescent amplified-fragment length polymorphism analysis (FAFLP) and pulsed-field gel electrophoresis (PFGE). J. Med. Microbiol. **50:**588-593.

9. **Grundmann, H., M. Aires-de-Sousa, J. Boyce, and E. Tiemersma.** 2006. Emergence and resurgence of meticillin-resistant *Staphylococcus aureus* as a public-health threat. Lancet **368:**874-885.

10. **Herron-Olson, L., J. R. Fitzgerald, J. M. Musser, and V. Kapur.** 2007. Molecular correlates of host specialization in Staphylococcus aureus. PLoS ONE **2:**e1120.

11. **Highlander, S. K., K. G. Hulten, X. Qin, H. Jiang, S. Yerrapragada, E. O. Mason, Jr., Y. Shang, T. M. Williams, R. M. Fortunov, Y. Liu, O. Igboeli, J. Petrosino, M. Tirumalai, A. Uzman, G. E. Fox, A. M. Cardenas, D. M. Muzny, L. Hemphill, Y. Ding, S. Dugan, P. R. Blyth, C. J. Buhay, H. H. Dinh, A. C. Hawes, M. Holder, C. L. Kovar, S. L. Lee, W. Liu, L. V. Nazareth, Q. Wang, J. Zhou, S. L. Kaplan, and G. M. Weinstock.** 2007. Subtle genetic changes enhance virulence of methicillin resistant and sensitive Staphylococcus aureus. BMC Microbiol **7:**99.

12. **Kuroda, M., T. Ohta, I. Uchiyama, T. Baba, H. Yuzawa, I. Kobayashi, L. Cui, A. Oguchi, K. Aoki, Y. Nagai, J. Lian, T. Ito, M. Kanamori, H. Matsumaru, A. Maruyama, H. Murakami, A. Hosoyama, Y. Mizutani-Ui, N. K. Takahashi, T. Sawano, R. Inoue, C. Kaito, K. Sekimizu, H. Hirakawa, S. Kuhara, S. Goto, J. Yabuzaki, M. Kanehisa, A. Yamashita, K. Oshima, K. Furuya, C. Yoshino, T. Shiba, M. Hattori, N. Ogasawara, H. Hayashi, and K. Hiramatsu.** 2001. Whole genome sequencing of meticillin-resistant Staphylococcus aureus. Lancet **357:**1225-40.

13. **Kwok, A. Y. C., and A. W. Chow.** 2003. Phylogenetic study of *Staphylococcus* and *Macrococcus* species based on partial hsp60 gene sequences. Int. J. Syst. Evol. Microbiol. **53:**87-92.

14. **Lane, D. J.** 1991. 16S/23S rRNA sequencing, p. 115-175. *In* E. Stackebrandt and M. Goodfellow (ed.), Nucleic acid techniques in bacterial systematics. John Wiley & Sons, Inc., Chichester, United Kingdom.

15. **Martineau, F., F. J. Picard, D. Ke, S. Paradis, P. H. Roy, M. Ouellette, and M. G. Bergeron.** 2001. Development of a PCR Assay for Identification of Staphylococci at Genus and Species Levels. J. Clin. Microbiol. **39:**2541-2547.

16. **McDonald, M., A. Dougall, D. Holt, F. Huygens, F. Oppedisano, P. M. Giffard, J. Inman-Bamber, A. J. Stephens, R. Towers, J. R. Carapetis, and B. J. Currie.** 2006. Use of a single-nucleotide polymorphism genotyping system to demonstrate the unique epidemiology of methicillin-resistant Staphylococcus aureus in remote aboriginal communities. J. Clin. Microbiol. **44:**3720-7.

17. **Neoh, H. M., L. Cui, H. Yuzawa, F. Takeuchi, M. Matsuo, and K. Hiramatsu.** 2008. Mutated response regulator graR is responsible for phenotypic conversion of Staphylococcus aureus from heterogeneous vancomycin-intermediate resistance to vancomycin-intermediate resistance. Antimicrob Agents Chemother **52:**45-53.

18. **O'Brien, F. G., T. T. Lim, F. N. Chong, G. W. Coombs, M. C. Enright, D. A. Robinson, A. Monk, B. Said-Salim, B. N. Kreiswirth, and W. B. Grubb.** 2004. Diversity among Community Isolates of Methicillin-Resistant *Staphylococcus aureus* in Australia. J. Clin. Microbiol. **42:**3185-3190.

19. **Okuma, K., K. Iwakawa, J. D. Turnidge, W. B. Grubb, J. M. Bell, F. G. O'Brien, G. W. Coombs, J. W. Pearman, F. C. Tenover, M. Kapi, C. Tiensasitorn, T. Ito, and K. Hiramatsu.** 2002. Dissemination of New Methicillin-Resistant *Staphylococcus aureus* Clones in the Community. J. Clin. Microbiol. **40:**4289-4294.

20. **Poyart, C., G. Quesne, C. Boumaila, and P. Trieu-Cuot.** 2001. Rapid and Accurate Species-Level Identification of Coagulase-Negative Staphylococci by Using the sodA Gene as a Target. J. Clin. Microbiol. **39:**4296-4301.

21. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24:**1596-1599.

22. **Tong, S., E. Bishop, R. Lilliebridge, A. Cheng, Z. Spasove-Penkova, D. Holt, P. Giffard, M. McDonald, B. Currie, and C. Boutlis.** Community-associated methicillin-resistant *Staphylococcus aureus* and methicillin-susceptible *S. aureus* in Indigenous northern Australia: epidemiology and outcomes. Journal of Infectious Diseases **In Press**.

Figure Legends

Figure 1. Neighbour joining trees of *gap, rpoB, sodA, tuf* and *hsp60* DNA sequences The following fragments of the coding sequences were used: *gap*: 58-941, *rpoB*: 1-492, *sodA*: 11-380, *hsp60*:1-538, *tuf*: 360-730 (conventional *S. aureus* numbering). .

Figure 2. Neighbour joining tree of representative *yqiL* alleles from the *S. aureus* MLST database. To aid presentation, only a subset of alleles were included. These were selected to encompass as much diversity as possible.

Table 1

Oligonucleotides used for PCR and sequencing.

Locus	Primer name	Primer sequence (5'- 3')	Amplicon size (bp)	Reference
Phylogenetic analysis				
16S	16S 27F	AGAGTTTGATCMTGGCTCAG	950	(14)
	16S 907R	CCGTCAATTCMTTTRAGTTT		
gap	gap1-F	ATGGTTTTGGTAGAATTGGTCGTTTA	931	(6)
	gap2-R	GACATTTCGTTATCATACCAAGCTG		
rpoB	rpoB-2491F	AACCAATTCCGTATIGGTTT	751	(3)
	rpoB-3241R	GCIACITGITCCATACCTGT		
sodA	sodA-F	CCITAYICITAYGAYGCIYTIGARCC	429	(20)
	sodA-R	ARRTARTAIGCRTGYTCCCAIACRTC		
tuf	tuf-Tseq271F	AAYATGATIACIGGIGCIGCICARATGGA	885	(15)
	tuf-Tseq1138R	CCIACIGTICKICCRCCYTCRCG		
hsp60	hsp60-F	GAATTCGAIIIIGCIGGIGAYGGIACIACIAC	600	(7)
	hsp60-R	CGCGGGATCCYKIYKITCICCRAAICCIGGIGCYTT		
MLST primers				
arcC	*arcC*-Up	TTG ATT CAC CAG CGC GTA TTG TC	456	(4)
	arcC-Dn	AGG TAT CTG CTT CAA TCA GCG		
aroE	*aroE*-F75 internal	AAC TTT AAG TCT TTA GGG TTA GA	412	This Study
	aroE-R2 75 internal	CAT TTC AGC ATC GTT TAA CGT		
glpF	*glpF*-Up	CTA GGA ACT GCA ATC TTA ATC C	465	This Study
	glpF-R75	AGG TAA AAT AGC ATG TGC AAT TC		
gmk	*gmk*-F75	ATC GTT TTA TCA GGG CCA TC	429	This Study
	gmk-R75	TCA TTA ACT ACT ACG TAA TCA TA		
pta	*pta*-Up	GTT AAA ATC GTA TTA CCT GAA GG	474	(4)
	pta-Dn	GAC CCT TTT GTT GAA AAG CTT AA		
tpi	*tpi*-F75	TCG GAC ATT CTG AAC GTC GTG AA	402	This Study
	tpi-Dn	TTT GCA CCT TCT AAC AAT TGT AC		
yqiL	*yqiL*-F75	CAG CAT ATA GAA CAC CAA TTG GC	516	This Study
	yqiL-R75	CGT TGT GGA ATT GAT ACC GGA AC		

Table 2

GenBank accession numbers of sequences used for the phylogenetic analyses.

Staphylococcus species (strain)	Genome sequence	*gap*	16S rRNA	*rpoB*	*sodA*	*hsp60*	*tuf*
S. aureus Mu3 (17)	AP009324						
S. aureus N315(12)	BA000018						
S. aureus Newman (1)	AP009351						
S. aureus RF122 (10)	AJ938182						
S. aureus USA300 (11)	CP000730						
CC75 HS2		To be submitted	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
CC75 HS42		To be submitted	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
CC75 M34		To be submitted	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
CC75 SCC1302		To be submitted	To be submitted	To be submitted	To be submitted	To be submitted	To be submitted
S. capitis		DQ321676	AY688039	AF325885	AJ343940	AF036322	AF298798
S. epidermidis		DQ321683	AY688053	AF325872	AJ343906	AF029245	AF298800
S. haemolyticus		DQ321687	AY688062	AF325888	AJ343910	U92809	AF298801
S. hominis		DQ321688	AY688064	AF325875	AJ343911	AF053572	AF298802
S. hyicus		DQ321689	AY688066	AF325876	AJ343913	AF019778	EU571080
S. intermedius		DQ321690	AY688070	AF325869	AJ343914	AY123723	EU571083
S. lugdunensis		DQ321693	AY688076	AF325870	AJ343917	AF053570	AF298803
S. sciuri		DQ321697	AB212276	AY820256	AY820257	AY820255	AY763434
S. simulans		DQ321698	AY688101	AF325877	AJ343956	AF053584	AF298805
S. warneri		DQ321699	AY688106	AF325887	AJ343958	AF053569	AF298806

Table 3

MLST alleles of CC75 isolates

Isolate	***arcC***	***aroE***	***glpF***	***gmk***	***pta***	***tpi***	***yqiL***
ST75	36	3[a]	43	34	39	52	49
ST850	36	3[a]	43	79	39	52	49
ST883	36	4[a]	43	34	107	116	105
ST1223	151	187	20	101	145	150	131
ST1304	36	45a	43	34	39	52	49
CC75-HS2	151	(187)[b]	20	new 1	145	150	131
CC75-M34	new 1	new 1	43	new 2	107	116	105
CC75-SCC1119	36	new 2	43	new 2	39	52	49
CC75-SCC1302	36	new 2	43	new 2	39	52	49

[a] possible incorrect alleles, so ST75 and ST1304 may be identical

b identical to allele 187, but slightly less than the full *aroE* sequence was determined (see text).