
Accepted Manuscript

Genomic epidemiology of methicillin-susceptible \textit{Staphylococcus aureus} across colonisation and skin and soft tissue infection

Alex Grinberg, Patrick J. Biggs, Ji Zhang, Stephen Ritchie, Zachary Oneroa, Charlotte O'Neil, Ali Karkaba, Niluka S. Velathanthiri, Geoffrey W. Coombs

PII: S0163-4453(17)30256-6
DOI: 10.1016/j.jinf.2017.07.010
Reference: YJINF 3960

To appear in: \textit{Journal of Infection}

Received Date: 27 June 2017
Revised Date: 0163-4453 0163-4453
Accepted Date: 29 July 2017


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Genomic epidemiology of methicillin-susceptible *Staphylococcus aureus* across colonisation and skin and soft tissue infection

**Running title:** Genomic epidemiology of MSSA


a Massey University, Institute of Veterinary, Animal and Biomedical Sciences, Private Bag 11,222, Palmerston North (4442), New Zealand.

b University of Auckland, Faculty of Medical and Health Sciences, Molecular Medicine and Pathology, 85 Park Rd, Grafton, Auckland, 1023, New Zealand.

c School of Veterinary and Life Sciences, Murdoch University, 90 South Street, Murdoch, WA, Australia

* Corresponding author: Alex Grinberg, Infectious Diseases Group, Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Private Bag 11,222, Palmerston North (4442), New Zealand;

Tel: (+) 64 6 3569099; Email: a.grinberg@massey.ac.nz
Summary

Objectives: *Staphylococcus aureus* skin and soft tissue infection (Sa-SSTI) places a significant burden on healthcare systems. New Zealand has a high incidence of Sa-SSTI, and here most morbidity is caused by a polyclonal methicillin-susceptible (MSSA) population. However, MSSA also colonise asymptomatically the cornified epithelia of approximately 20% of the population, and their divide between commensalism and pathogenicity is poorly understood. We aimed to see whether MSSA is genetically differentiated across asymptomatic colonisation and SSTI; and given the close interactions between people and pets, whether strains isolated from pets differ from human strains.

Methods: We compared the genomes of contemporaneous asymptomatic colonisation and clinical MSSA isolates obtained in New Zealand from humans and pets.

Results: Core and accessory genome comparisons revealed a homogeneous bacterial population across colonisation, disease, humans, and pets. The rate of MSSA colonisation in dogs was comparatively low (5.4%).

Conclusions: In New Zealand, most Sa-SSTI morbidity is caused by a random sample of the colonising MSSA population, consistent with the opportunistic infection paradigm rather than a model distinguishing pathogenic from innocuous strains. Thus, studies of host factors determining colonisation and immune-escape may be more beneficial than comparative virulence studies. Contact with household pets may pose low zoonotic risk.

Keywords: *Staphylococcus aureus*, skin and soft tissue infection, SSTI, whole genome sequencing, zoonoses
Introduction

Skin and soft tissue infection (SSTI) places a significant burden on healthcare systems due to its propensity to recur and complicate [1]. *Staphylococcus aureus* is the predominant pathogen isolated from SSTI, globally [2]. The incidence of *S. aureus* SSTI (Sa-SSTI) has increased significantly in many developed countries [3, 4]. Variations in the rate of isolation of methicillin-resistant versus methicillin-susceptible *S. aureus* strains (MRSA; MSSA) are observed between continents, with a higher rate of MRSA registered in North America than in Europe [2]. New Zealand has a high incidence of Sa-SSTI, but a recent national survey indicated that here most morbidity is caused by a polyclonal MSSA population [5].

*S. aureus* also colonises asymptptomatically the cornified epithelia and is found in the nares of 20-30% of the population at any point in time [6]. Some observations indicate colonisation and disease-causing isolates are approximately equally distributed among the predominant lineages [7, 8]. Nevertheless, although asymptomatic MSSA colonisation is common, only a small proportion of colonised people develop illness, and this dual lifestyle as a commensal and a pathogen has puzzled researchers for decades [9, 10]. On one side of the spectrum, the host-centred model of pathogenesis considers MSSA as opportunistic agents and attributes a central pathogenetic role to the variations or breaches in host defences [11, 12]. This model predicts a homogeneous MSSA population across asymptomatic carriage and disease. Alternatively, the microorganism-centred model distinguishes between innocuous and pathogenic strains [13]. The conflict between these models remains unresolved mainly because population-based studies rarely compare the relative abundances of contemporaneous strains across asymptomatic colonisation and disease.

Colonised people are at increased risk of developing an endogenous MSSA infection [14], but the exogenous sources of Sa-SSTI are elusive [15]. Modern society fosters close physical interactions
between people and pets, and MSSA are isolated sporadically from a range of infection sites in dogs and cats [16], and also from the nares and skin of asymptomatic dogs [16, 17]. Recent studies showed that pets can be colonised by the same MRSA lineages causing infection in people [18], but the relatedness of MSSA colonizing pets and causing Sa-SSTI in humans is not known.

The comparative genomics approach enables an understanding of the differentiation of bacterial populations across different ecological niches. Most recent studies comparing colonising and clinical S. aureus isolates concerned MRSA. Extrapolation of these data to MSSA is problematic as MRSA belong to a relatively small spectrum of lineages, whereas MSSA are highly diverse [19]. Hence, we performed a comparative genomics study of contemporaneous human and pet colonisation and clinical MSSA isolates obtained in New Zealand. Our study was motivated by two main questions. Firstly, we wanted to ascertain whether MSSA is differentiated across asymptomatic colonisation and SSTI, which is critical to understand the pathogenesis of Sa-SSTI. Secondly, considering the potential role of pets as reservoir of zoonotic Sa-SSTI, we aimed to assess MSSA host-specificity.

**Materials and Methods**

**MSSA isolates**

We used 85 nearly contemporaneous isolates obtained from four epidemiological niches: human asymptomatic nasal colonisation (n=27); human clinical (n=17); pet asymptomatic colonisation (n=15); and pet clinical (n=26) isolates. The human nasal isolates were obtained in 2014 from 77 consenting adults (>15 years old) admitted to Auckland City Hospital for acute or elective orthopaedic care, within 48 hours of admission. Exclusion criteria included hospital admission within the last 6 months, or more than two admissions in the past 12 months; living in a long term care facility; having more than two long term conditions; taking more than two regular medications; receiving antibiotics; and suspected of
having an orthopaedic infection. Sampling was performed by rotating a swab in each anterior nares ten times. The Northern B New Zealand Health and Disability Ethics Committee approved the sampling. Swabs were placed in transport media, plated onto Mannitol-Salt agar the same day and incubated for 48 hours at 37°C. Colonies resembling *S. aureus* were sub-cultured and identified by phenotypic tests and later confirmed by genotyping.

The human clinical MSSA isolates originated from a large collection of isolates obtained in 2014 from clinical microbiology laboratories throughout the country. The collection had been genotyped previously for a national survey [5]. For this study, 17 clinical isolates were selected at random to represent the predominant seven-loci multi-locus sequence types (MLST)/spa-type combinations identified in the survey. Fourteen of these isolates were recorded as originating from SSTI, one from an eye swab, one from an ear swab, and one from a urine specimen. Nine isolates were from patients with community-onset infections, seven from healthcare-associated infections, and one was unknown (definitions of healthcare-associated and community-onset infection are provided in [4]).

The 26 clinical isolates from pets (dogs: 19; cats: 7) were isolated by two New Zealand animal diagnostic laboratory networks from a range of infection sites between June 2012 - June 2013 [16] and February - July 2016. The 15 colonisation isolates from pets (dogs: 13; cats: 2) were isolated from swabs taken from a cross sectional survey of 391 animals (colonisation prevalence: 3.8%; prevalence in dogs: 5.4%; in cats: 1.2%). Of these, 257 (148 dogs; 109 cats) were sampled in Auckland between June 2012 and June 2013 as previously described [15]. Additional animals (89 dogs and 45 cats) brought for veterinary care to the Massey University Veterinary Teaching Hospital, Palmerston North, were sampled between February and July 2016. Animals were sampled with owners’ consent by inserting and gently rotating a swab (Transystem™; Copan, Brescia, Italy) into the nostrils of dogs, or gently rubbing it on the perianal skin of cats (due to their small nostrils). The Massey University Animal Ethics Committee approved the sampling. Swabs were sent to the University microbiology laboratory and processed within 24 hours.
from arrival. Methods for isolation and identification of S. aureus included enrichment, as previously described [15]. All MSSA isolates were frozen at -80°C for future testing.

**DNA microarray genotyping**

Initially, the 85 isolates were typed using a commercial DNA microarray providing data on >300 genetic markers (CLONDIAG, Alere Technologies, Jena, Germany). Microarray enabled the prediction of the clonal complex (CC) of each isolate, and the exclusion of any MRSA [20].

**Whole genome sequencing (WGS) and genome annotation**

Eighty isolates were submitted for WGS on an Illumina MiSeq (Illumina Inc, San Diego, CA). DNA libraries were prepared using the Nextera XT kit and sequenced using version 2 chemistry to generate 300-bp pair-ended reads. We analysed paired sequence reads with an in-house software that performed read quality analysis and adapter removal. We assembled adapter-free filtered reads de novo using SPAdes [21] in the “careful” mode, with other parameters set as default. The assemblies were annotated with Prokka version 1.11 using Staphylococcus specific BLAST databases [22]. Outputs were stored in a MySQL database (Oracle Corporation, Redwood Shores, CA). Read curation methods and genomic metrics for the five most common CC (CC1, CC5, CC188, CC30 and CC15) are reported in Supplementary Text and Supplementary Table 1.

**Multi-locus sequence typing and spa-typing**

The classic seven-loci MLST profile of the isolates were determined by comparing the assembled sequences to the reference allele sequence definitions and then to the reference MLST profiles (https://pubmlst.org/saureus/, accessed 27.01.2017) with BLASTN [23]. Spa typing was performed by mapping the reads to a spa sequence database created using data compiled from the Ridom SpaServer (http://spa.ridom.de/spatypes.shtml). When spa-types could not be defined in silico, we performed PCR
and Sanger sequencing of the variable region of the protein A gene [24] and searched the obtained sequences in the Ridom server (searches were performed between June 2016 and February 2017).

Core genome comparisons

The S. aureus population exists as several globally distributed clonal complexes [25]. Hence, in order to maximize the number of genes available for the analysis, core genome comparisons between the niches were performed for each CC separately. The relationships between the isolates within each CC were resolved using Neighbour Joining (NJ) trees, assuming equality of substitution rates. Firstly, the trees were constructed using the pairwise Hamming distance matrix found through ad hoc whole-genome MLST (wgMLST) analysis with Genome Profiler version 2.0 [26]. A second core genome analysis within the CCs used NJ trees generated using the pairwise amino-acid Hamming distance matrix. Briefly, for each isolate we extracted the amino acid predictions and clustered genes using OrthoMCL [27]. Within each CC, a set of orthologous amino acid predictions observed only once in all the isolates was investigated further. If the lengths of the prediction in an orthologous cluster were the same, that cluster formed part of what we dubbed ‘shared genome’ (SG) (a proxy of the core genome). The amino-acid predictions were concatenated per isolate to produce the SG and the relationships between the concatemers were resolved using the NJ trees. All NJ trees were visualized using Geneious R10 [28].

Accessory genome comparisons

Approximately 20% of the S. aureus genome is variably present and represents the accessory genome. It encodes a spectrum of drug and metal resistance, substrate utilization, virulence factors, and more [29]. Accessory genome analysis is essential for the understanding of bacterial niche adaptation [30, 31]. In this study we dubbed the set of genes that were not observed in all the isolates ‘variably shared genome’ (VSG), a proxy of the accessory genome. We analysed the accessory genomes of all the isolates used for core genome comparisons, using a number of complementary techniques.
Firstly, we tabulated the annotations identified in the sample and created a binary matrix reporting the presence or absence of each gene, in each isolate. To visualise the relationships between the isolates, we created a pairwise Hamming distance matrix between the isolates and used it for a Principal Coordinate Analysis (PCoA) on PAST software [32]. The matrix was further explored using heatmaps constructed in R (R Foundation for Statistical Computing, Vienna, Austria; http://www.R-project.org/).

Secondly, we performed a Clusters of Orthologous Groups (COG) analysis to assess differences in the distribution of gene abundances of 23 protein functional groups between the niches [33]. Briefly, an analysis of the PROKKA gene predictions was performed using the COGsoft set of tools (available at ftp://ftp.ncbi.nih.gov/pub/wolf/COGs/COGsoft). Each genome was analysed with the PSI-BLAST algorithm against itself, and then against the COG database with and without filtering. The results were parsed using the COGmakehash, COGreadblast and COGnitor programs with default parameters, to generate a set of COG assignments for each isolate. A table containing the number of proteins per COG functional group for each isolate was used for a Principal Component Analysis (PCO), performed using PAST software.

**Gene-by-gene comparisons**

It has been hypothesized that the presence or absence, or truncations or mutations of single genes might profoundly influence the *S. aureus* phenotype [34, 35]. Thus, we complemented the genome-wide statistical analysis with gene-by-gene comparisons. Briefly, we consulted the literature and extracted the sequences of the following potentially influential genes from the assembled genomes: the coding sequences of the *agrC* (GenBank accession number JQ066322.1); *fnBPA* (GenBank Acc. ET052801); *rsbU* (GenBank Acc. CAA68929); *isdH* (GenBank Acc. ADC37900); *fusC* (GenBank Acc. 2YB5_A); *cfB* (GenBank Acc. WP_0724333232); *coa* (GenBank Acc. WP_000744061); *isdA* (GenBank Acc. Q7A152); *sarA* (GenBank Acc. AA805396); and *vwBP* (GenBank Acc. BAB94631) (gene extraction
methods are described in Supplementary Text). Extracted sequences were compared using multiple alignments. In order to exclude genes containing artifactual truncations, as a consequence of gaps between the contigs, we only used isolates yielding gene sequences contained in a single contig amidst unrelated sequence. Finally, we explored the microarray data to determine the presence in each isolate, of the mobile elements LukF-PV and and LukS-PV, *scn*, *sak*, 17 enterotoxin genes, TST (‘human allele’), the ACME element and the biofilm-associated *bap* gene.

**Results**

**Clonal complexes, multi-locus sequence types and spa-types**

Microarray correctly predicted the CC of the 17 previously genotyped human clinical isolates. Five out of six predominant CCs identified in human clinical isolates in the national survey (CC1, CC188, CC5, CC30 and CC15) were also the most frequently identified CCs among the human colonisation sample and among the pet colonisation and clinical samples. Moreover, the predominant MLST and spa types reported in the national survey were also the most common combinations found in these samples (Table 1). An unweighted pair group method with arithmetic mean (UPGMA) dendrogram constructed in PAST software using the microarray presence/absence data showed each CC as a monophyletic group, and within each CC many colonisation and clinical isolates from human and animals had identical hybridisation patterns (Figure 1).

**Core genome comparisons**

The NJ trees resolving the wgMLST relationships within CC1, CC188, CC5 and CC30 are shown in Figure 2 (CC15 contained a small number of isolates and is not shown). As expected, in all the trees, each isolate is resolved in a separate branch. Isolates from the same niche did not form monophyletic branches.
Shallow clusters of isolates from the same niche were observed in CC1, but these were not observed in the concatenated shared genome trees, where human and animal isolates from colonisation and disease co-segregated (concatenated trees are not shown). These results did not support a differentiation of the core genomes between the niches.

**Accessory genome comparisons**

The PCoA results showed a marked segregation of the VSG according to their CC, but within each CC many isolates from different niches intertwined in the bi-dimensional space delimited by the two main co-ordinates (Figure 3). Interestingly, CC30 and CC5 clustered close to each other, suggesting the composition of their accessory genomes was similar, but this was not further explored. In agreement with the PCoA, heatmaps showed a high degree of similarity within each CC, but not within the niches (Figure 4). The PCA plot resolving the COG data showed a fairly homogeneous scattering of the four niches (Supplementary Figure 1). These results did not support a differentiation of the accessory genomes between the niches.

**Gene-by-gene comparisons**

Multiple alignments of the agrC sequences showed a segregation of the sequences according to CC, but within each CC the isolates from the different niches displayed identical sequences. Also the RsbU, IsdH, VwBP, ClfB, Coa, and FnBPA alignments showed segregation by CC but identical sequences within each CC, whereas the SarA had a small number of nucleotide variations in one isolate only. CC1, CC5 and CC188 showed identical IsdA sequences. However, the IsdA sequences of CC15 and CC30 were distinct. Consistent with the observation of a high prevalence of fusidic acid resistance in CC1 in New Zealand [36], ten CC1 isolates harbouried the fusC gene. By microarray, the scn, sak, TST (‘human allele’), fnbA, fnbB, LukF-PV and LukS-PV and most enterotoxin genes were identified in the four niches.
The *Bap* gene and the ACME element were not found in this study (microarray results are provided in Supplementary File 1).

**Discussion**

There are many open questions regarding the pathogenesis of Sa-SSTI. Several mobile genetic elements have been associated with increased *S. aureus* virulence, but many elements are found across a range of clones, precluding a clear delineation of strains according to their pathogenicity [37-39]. The results of our genomic comparison indicate a homogeneous MSSA population across colonisation, SSTI, humans, and pets.

Our study has strengths and limitations. Firstly, data suggest that subtle genetic differences might alter the *S. aureus* in vitro phenotype. Hence, traditional genotyping methods might be considered inadequate to discern commensals from pathogens. In this study we used the comparative genomics approach. Other comparative strengths were the use of nearly contemporaneous samples of isolates from New Zealand, which precluded major temporal and geographical biases, and the availability of national molecular data, which allowed us to include all the predominant CCs causing SSTI in New Zealand. Limitations included the cross-sectional nature of the sampling and the lack of detailed clinical data, which did not allow us to draw conclusions about the duration of colonisation or the severity of the infections.

The genomic comparison proceeded in steps. The initial comparison of microarray and spa-typing results with the data from the national survey revealed a homogeneous clonal landscape dominated by the same CC/MLST/spa type combinations across colonisation and disease, in humans and pets. Therefore, notwithstanding the sampling variation we succeeded in capturing the predominant MLST/CC/spa-types causing Sa-SSTI in New Zealand using sample sizes of only 27 human and 15 animal
colonisation isolates. Moreover, several uncommon CC in this study also displayed uncommon spa-types reported in clinical isolates in the national survey (Table 1). Interestingly, the same clones were also common among the infection site isolates obtain from pets.

Core genome comparisons within CC1, CC188, CC5 and CC30 failed to detect any niche-specific differentiation, and accessory genomes displayed segregation of isolates according to their CC, but not by epidemiological niche. We also performed gene-by-gene comparisons of several elements which might determine niche-specificity, and again, these elements differentiated according to the CC, but not by niche. Hence, MSSA causing SSTI in New Zealand appear to be a random sample of the nasal colonisation population, consistent with the opportunistic model of infection. Studies of host factors determining colonisation and immune-escape phenomena may therefore be more beneficial than comparative bacterial virulence studies.

It is important to note that these results refer to the predominant strains found in New Zealand and do not completely preclude the existence of variability in infectivity or virulence. In fact, outbreaks of Sa-SSTI caused by single strains have been documented [40], and such strains may be more infectious than other strains. Moreover, our results may not be relevant in other regions, where highly infectious minority colonisation variants might contribute disproportionately to morbidity. This could be the case in the USA after the emergence of the USA300, but this remains to be established as comparisons of the relative abundance of contemporaneous USA300 isolates across colonisation and disease, are lacking.

Informed interventions aimed at controlling infectious diseases require knowledge of the ecological niches where pathogens thrive or survive. The importance of endogenous S. aureus infections is widely acknowledged, but the exogenous sources of infection are not well characterised. S. aureus are also important animal pathogens. In particular, MSSA is a common cause of bovine contagious mastitis, worldwide [41]. Yet, cross-transmission of MSSA between cattle and humans has only been documented
sporadically [42], and studies have shown a population differentiation among cattle and humans [43, 44]. The situation in pets is less clear. Recent studies showed that pets can be colonised by the same MRSA lineages causing infection in people [18], and genomic studies have inferred cross-transmission of MRSA between dogs and humans [45]. Not surprisingly, therefore, MRSA (but not MSSA) are increasingly listed as potential zoonotic agents transmissible from pets [46]. We found that pets are colonised and infected by the same MSSA population colonising and causing infection in people. However, the 5.4% colonisation prevalence in dogs is significantly lower than the 18% nasal colonisation found in humans in New Zealand using cross-sectional sampling [47]. We may have underestimated the colonisation prevalence in pets, as 30% of the animals sampled in Auckland received antimicrobial treatment on the day of sampling [16]. Nevertheless, comparably low nasal MSSA colonisation rates have been previously reported in dogs [17, 48], suggesting *S. aureus* colonisation is somehow hindered in this species. We postulate that pets may pose low zoonotic risk, but we would consider house-hold pets as potential exogenous sources of MSSA in case of recurrent SSTI infections.

In conclusion, we found a homogeneous MSSA population across colonisation and SSTI in humans and pets, consistent with the opportunistic model of infection and a lack of host-specificity. A low prevalence of colonisation in pets signals low zoonotic risk. A myriad of statistical methods not used in this study are available for genomic comparisons. Hence, we deposited the nucleotide reads with the associated data in the Sequence Read Archive (SRA) of the US National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/sra/; submission number SUB2765041), and welcome others to use them.

**Conflict of Interest:** The authors declare that they have no conflict of interest.
Acknowledgments

The human clinical MSSA isolates were provided by Helen Heffernan, Institute of Environmental Science & Research Ltd, Porirua, New Zealand. We thank Barry Hedgespeth, BVSc, for facilitating the sampling of dogs and cats in Palmerston North. This work was supported by the Mc George Research Fund, New Zealand.

References


5 - Heffernan H, Bakker S, Woodhouse R, Dyet K, Williamson D. Institute of Environmental Science & Research Ltd, Porirua. Demographics, antimicrobial susceptibility and molecular epidemiology of


Figure 1. Unweighted Pair Group Method with Arithmetic Mean (UPGMA) dendrogram obtained using the microarray presence/absence data matrix. Each isolate is represented by the initials of its epidemiological niche followed by the clonal complex (CC), its seven-loci multi-locus sequence type (ST) and spa type (t). HN: human nasal (black); HK: human clinical (red); CN: canine nasal (purple); CK: canine clinical (green); FN: feline perianal (brown); FK: feline clinical (blue). Microarray markers containing ambiguous hybridization results were excluded from the analysis.

Figure 2: Unrooted neighbour joining trees resolving the relationships between the core genomes of the isolates within clonal complex (CC) 1 (upper left), CC5 (upper right), CC30 (lower left), and CC188 (lower right). Trees were built using the Jukes-Cantor model considering the Hamming distance matrix obtained from pairwise comparisons between the whole genome multi-locus sequence types (wgMLST) of the isolates. Within each CC, isolates are named by a serial number followed by the niche denomination. The ST and spa type of each isolate are reported in Supplementary Table 1. HN, human nasal; HK, human clinical; CN, canine nasal; FN, feline nasal; CK canine clinical; FK feline clinical isolates.

Figure 3. Principal co-ordinate analysis plots obtained using the MSSA binary gene presence/absence matrix. The isolates are represented by data points positioned in the bi-dimensional space given by the two main coordinates. Blue, clonal complex (CC) 188; black, CC1; green, CC30; red, CC5. In A, colonization isolates are represented as triangles and clinical isolates as dots. In B, human isolates are represented as triangles and animal isolates as dots. Note the segregation of the isolates according to their CC, except CC30 and CC5, which do not appear to markedly segregate. Note also an admixture of clinical and carriage isolates, and animal and human isolates within each clonal complex (a sub-cluster of four segregated animal isolates appear in CC1, but this was not further analysed). Two isolates (one from CC188 and one from CC5) were excluded as the COG analysis returned thousands of ‘not-determined’ coding sequences for these genomes.

Figure 4. Heatmaps constructed using the binary gene presence/absence matrix of the sequenced isolates belonging to CC1, CC5, CC188 and CC30. The gradient white -> yellow -> red indicates more similar-> less similar isolates (a scale reporting the number of differences is presented between the heatmaps). Left heatmap: isolates sorted by clonal complex (red bar: CC1; blue: CC5; yellow: CC30; purple: CC188). Right heatmap: isolates sorted by epidemiological niche (red bar: canine and feline colonization; blue: canine and feline clinical; green: human colonization; yellow: human clinical). Note the pattern of similarity within each CC (in A), and the lack of a distinctive pattern of the isolates sorted by niche (in B).
Table 1: Common methicillin-susceptible *Staphylococcus aureus* (MSSA) clonal complexes (CC), multi-locus sequence types (ST) and spa-types (t) observed in New Zealand, and in this study.

<table>
<thead>
<tr>
<th>ST+spa-type (in this study)*§</th>
<th>Predominant CC/spa-type in NZ† (% of MSSA morbidity)</th>
<th>Number of human clinical isolates (in this study)</th>
<th>Number of human colonisation isolates (in this study)</th>
<th>Number of animal clinical isolates (in this study)</th>
<th>Number of animal colonisation isolates (in this study)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC1</td>
<td>t127 (15.3%)</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>ST1 + t127</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST1 + tnew</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST1 + t207</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST1 + t601</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST2851 + t127</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC188</td>
<td>t189 (10.2%)</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>ST188 + t189</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST188 + t8914</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST188 + t5558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC5</td>
<td>t002 + t1265 + t179 (9.9% in total)</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>ST5 + t002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST5 + t1265</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST5 + t179</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST5 + t5213</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST5 + t6690</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST5 + t1550</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST5 + t15693</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST5 + t15692</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST835 + t002</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST1259 + t5402</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC121</td>
<td>t645 + t659 + t159 (7.1% in total)</td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ST121 + t645</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST2276 + t659</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC15</td>
<td>t084 (2.8%)</td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ST15 + t084</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST15 + t085</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST15 + t13775</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST582 + t1509</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC30</td>
<td>t019 and t012 (2.9% in total)</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ST30 + t019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST30 + t012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST30 + t1128</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST30 + t11414</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST30 + t1623</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST30 + t1227</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST34 + t089</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other CC/ST (this study)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST72 + t4727</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST45 + t065</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST1281 + t2094</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST672 + t3841</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST672 + t1781</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST2479</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST97 + t267</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC8 + CC9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1 + 1</td>
</tr>
<tr>
<td>CC59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC7 + CC45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* the spa types observed in this study and accounting for ≥1% or the morbidity in New Zealand are underlined; § data extracted from Heffernan et al 2015 [3].
Figure 1
Figure 2
Figure 3.
Figure 4
Highlights

- We compared the genomes of MSSA across colonisation and disease, humans, and pets
- Skin and soft tissue infection is caused by a random sample of the colonising MSSA
- This result supports the opportunistic model of infection
- Pets pose low zoonotic risk