Extensive Horizontal Gene Transfer within and between Species of Coagulase-Negative *Staphylococcus*

Joshua T. Smith[1,*] and Cheryl P. Andam[2,*]

[1]Department of Molecular, Cellular and Biomedical Sciences, University of New Hampshire, Durham, New Hampshire, USA

[2]Department of Biological Sciences, University at Albany, State University of New York, New York, USA

*Corresponding authors: E-mails: candam@albany.edu; jts1044@wildcats.unh.edu.

Accepted: 1 September 2021

Abstract

Members of the gram-positive bacterial genus *Staphylococcus* have historically been classified into coagulase-positive *Staphylococcus* (CoPS) and coagulase-negative *Staphylococcus* (CoNS) based on the diagnostic presentation of the coagulase protein. Previous studies have noted the importance of horizontal gene transfer (HGT) and recombination in the more well-known CoPS species *Staphylococcus aureus*, yet little is known of the contributions of these processes in CoNS evolution. In this study, we aimed to elucidate the phylogenetic relationships, genomic characteristics, and frequencies of HGT in CoNS, which are now being recognized as major opportunistic pathogens of humans. We compiled a data set of 1,876 publicly available named CoNS genomes. These can be delineated into 55 species based on allele differences in 462 core genes and variation in accessory gene content. CoNS species are a reservoir of transferrable genes associated with resistance to diverse classes of antimicrobials. We also identified nine types of the mobile genetic element SCC*mec*, which carries the methicillin resistance determinant *mecA*. Other frequently transferred genes included those associated with resistance to heavy metals, surface-associated proteins related to virulence and biofilm formation, type VII secretion system, iron capture, recombination, and metabolic enzymes. The highest frequencies of receipt and donation of recombined DNA fragments were observed in *Staphylococcus capitis*, *Staphylococcus caprae*, *Staphylococcus hominis*, *Staphylococcus haemolyticus*, and members of the *Saprophyticus* species group. The variable rates of recombination and biases in transfer partners imply that certain CoNS species function as hubs of gene flow and major reservoir of genetic diversity for the entire genus.

Key words: *Staphylococcus*, coagulase-negative, genome, horizontal gene transfer, recombination.

Significance

Little is known of the evolutionary history and genome structure of coagulase-negative *Staphylococcus* (CoNS), which are now being recognized as major opportunistic pathogens of humans. Here, we report that CoNS are more phylogenetically diverse than previously recognized, with frequent but biased recombination contributing to its evolution. Defining the evolutionary processes that contribute to CoNS diversification is the first step toward effective surveillance, diagnostics, and control of CoNS infections.

Introduction

Members of the gram-positive bacterial genus *Staphylococcus* have historically been classified into coagulase-positive *Staphylococcus* (CoPS) and coagulase-negative *Staphylococcus* (CoNS). This binomial classification system is based on whether a species exhibits clotting of plasma by the coagulase enzyme, which converts fibrinogen to fibrin (Crosby et al. 2016). Coagulase production as a diagnostic method was developed in the 1940s to discriminate the highly pathogenic *Staphylococcus aureus* from other species that were considered to be less clinically relevant (Fairbrother 1940). At that time, CoPS was almost exclusively represented by *S. aureus*. Since

then, numerous CoPS and CoNS species have been identified. There are also coagulase-variable species, which are composed of strains that are either coagulase-positive or -negative. For example, *Staphylococcus schleiferi* consists of a coagulase-negative subspecies (*S. schleiferi* subsp. *schleiferi*) and a coagulase-positive subspecies (*S. schleiferi* subsp. *coagulans*), but the phenotypic distinction between the two is often blurry (Vandenesch et al. 1994). Two other species are coagulase-variable (*Staphylococcus hyicus* and *Staphylococcus agnetis*; Hassler et al. 2008; Taponen et al. 2012). In 2014, the genus consisted of 47 species and 23 subspecies, with 38 belonging to CoNS (Becker et al. 2014). There are currently 75 species and 30 subspecies identified in the List of Prokaryotic names with Standing in Nomenclature (LPSN; Parte 2014; https://lpsn.dsmz.de/search?word=staphylococcus as of August 1, 2021). In 2012, Lamers et al. developed a multilocus sequence classification scheme based on four loci (16S rRNA, *dnaJ*, *rpoB*, and *tuf* gene fragments) to classify *Staphylococcus* genus into six species groups (*Auricularis*, *Hyicus-Intermedius*, *Epidermidis-Aureus*, *Saprophyticus*, *Simulans*, and *Sciuri*; Lamers et al. 2012; Becker et al. 2014).

There is mounting evidence from recent diagnostic studies that CoNS species pose a significant health burden as opportunistic pathogens. Because CoNS represent a major component of the microbiota of skin and mucous membranes in humans and animals, colonization appears to be a key source of endogenous infections (Akiyama et al. 1998; Fux et al. 2005; Zingg et al. 2009). Breaching the skin barrier through use of transiently or permanently inserted medical devices (e.g., intravenous catheters, vascular grafts, prosthetic joints) thus offers CoNS opportunities to gain access to host tissues and form biofilms on foreign body surfaces (von Eiff et al. 2006; Zheng et al. 2018). While almost all CoNS species generate infections given opportune environmental conditions (Becker et al. 2014), a few CoNS species are particularly notable. *Staphylococcus epidermidis* and *Staphylococcus haemolyticus* are most commonly associated with infections of indwelling or implanted foreign bodies (Ahmed et al. 2019; Both et al. 2021). *Staphylococcus saprophyticus* is the common cause of urinary tract infections (Widerström et al. 2007; Lawal et al. 2021), whereas *Staphylococcus lugdunensis* can cause infectious endocarditis (Non and Santos 2017). *Staphylococcus capitis* has been implicated in sepsis in preterm infants in neonatal intensive care units (Stenmark et al. 2019; Wirth et al. 2020). CoNS are also reported to be responsible for a variety of laryngological diseases (i.e., diseases of the upper respiratory tract and other organs present in the neck and head; Michalik et al. 2020). Worldwide, increasing antimicrobial resistance in CoNS species in healthcare and community settings further exacerbates the health burden they cause and limits available therapeutic options (May et al. 2014; Pedroso et al. 2018; Asante et al. 2021).

Previous studies have noted the importance of horizontal gene transfer (HGT) and recombination in the more well-known *S. aureus* (Castillo-Ramírez et al. 2012; Driebe et al. 2015; Spoor et al. 2015; Murray et al. 2017), yet little is known of the contributions of these processes in CoNS evolution. In this study, we aimed to elucidate the phylogenetic relationships and frequencies of genome-wide HGT in CoNS species. Using 1,876 publicly available high-quality genomes representing 55 species, we showed that CoNS species are remarkably diverse in terms of their core allelic and gene content variation, including antimicrobial resistance and virulence genes. Extensive HGT within and between species plays a major role in CoNS evolution, but frequencies of receipt and donation of recombined DNA fragments varied significantly between transfer partners. These results provide important insights into the evolutionary history of this obscure group of *Staphylococcus* species. Our findings are relevant for developing an improved classification system, surveillance, and management options of CoNS-related infections.

Results

CoNS Are Phylogenetically, Geographically, and Genomically Diverse

We first sought to determine the diversity and phylogenetic relationships of the 1,876 named CoNS genomes retrieved from NCBI. Genome sizes ranged from 2.10 to 3.12 Mb (mean = 2.53 Mb), whereas the number of homologous genes per genome ranged from 1,992 to 3,080 (mean = 2,410; supplementary table S1, Supplementary Material online). The entire CoNS pan-genome consisted of 59,238 homologous gene clusters (supplementary table S2, Supplementary Material online), of which 462 genes were present in ≥95% of the genomes (i.e., 239 core + 223 soft-core genes). The accessory genome consisted of 2,267 shell genes and 56,509 cloud genes. Of the latter, a total of 10,667 genes were singletons, that is, genes detected in only a single genome, and represented 18% of the pan-genome. Using the 95% average nucleotide identity (ANI) threshold to define a species (Jain et al. 2018), we identified a total of 55 species in the data set. To ensure that all genomes belong to the same genus, we calculated the percentage of conserved proteins (POCP; Qin et al. 2014). Due to high computational requirement of comparing every possible pair of the 1,876 genomes in our data set, we selected a representative genome from each of the 55 species. Our analysis showed pairwise POCP values ranging from 59.0% and 100% (supplementary fig. S1 and table S3, Supplementary Material online). POCP values of at least 50% had been proposed to delineate the boundaries of a prokaryotic genus (Qin et al. 2014).

CoNS genomes came from all seven continents, albeit in highly variable numbers due to differences in sampling efforts (supplementary fig. S2, Supplementary Material online). Majority came from North America and Europe, with 825

and 552 genomes, respectively. The data set also included 83 genomes that were derived from the International Space Station. The strains were derived from a variety of sources (human, animal, environment, and food items; supplementary table S1, Supplementary Material online). A total of 994 genomes came from human samples.

We built a maximum-likelihood phylogenetic tree using sequence variation in 462 core genes (fig. 1). The core genome phylogeny showed clearly defined species boundaries among the 55 ANI-defined species and among the six species groups. The largest species group *Epidermidis* consisted of 1,357 genomes, which can be delineated into 17 species. *S. epidermidis* represented the species with the highest number of genomes ($n = 732$). This was not surprising as *S. epidermidis* is an opportunistic pathogen associated with major human infections and thus has been widely sampled (Méric et al. 2018). The midpoint-rooted tree also showed the *Sciuri* species group, consisting of five species, located at the base of the phylogeny. This result is consistent with previous studies that establish it is the ancestral group of *Staphylococcus*

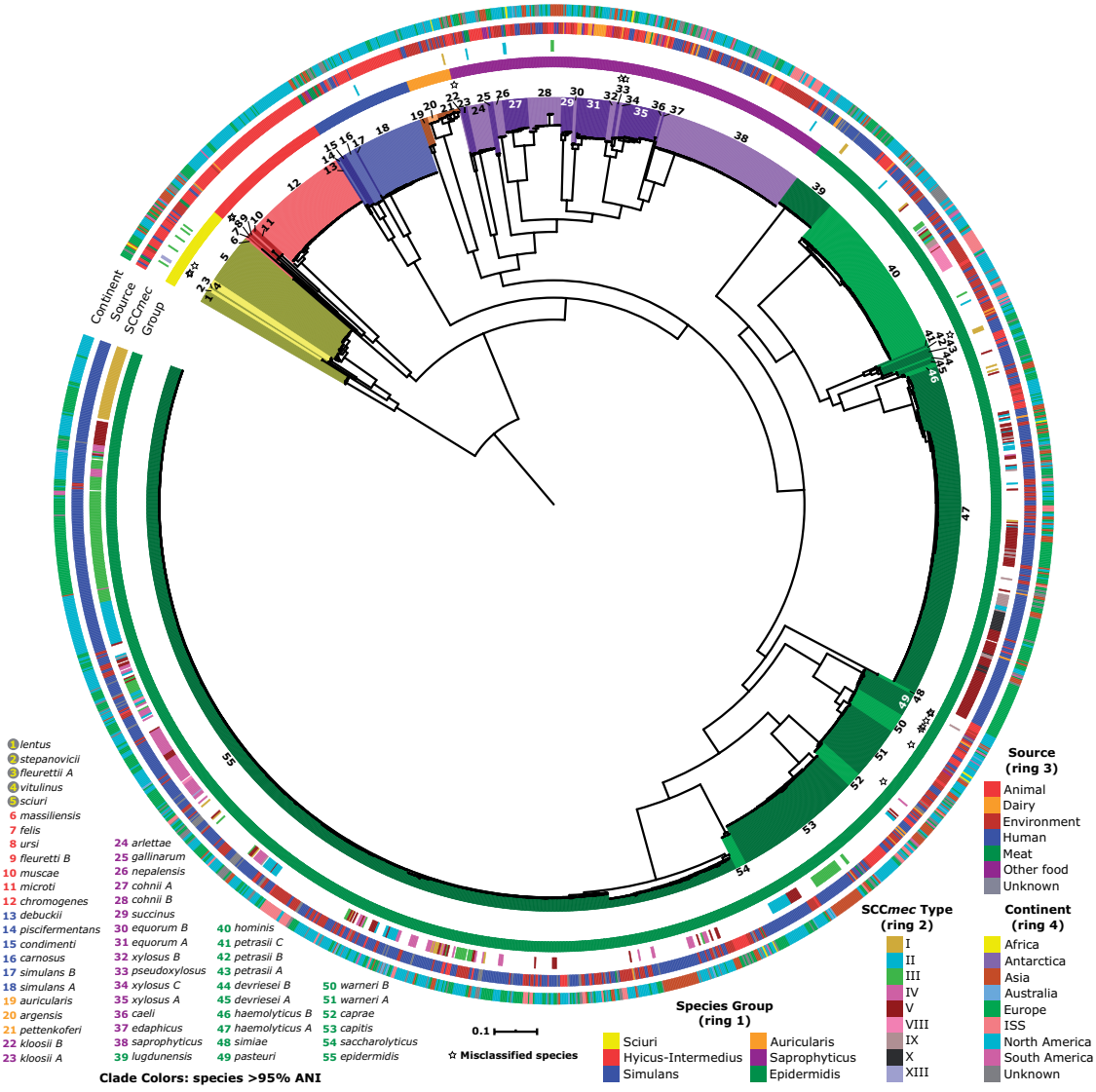

Fig. 1.—Maximum-likelihood phylogenetic tree of 1,876 CoNS genomes representing 55 species. The midpoint-rooted tree was built using the concatenation of 462 core genes. Species delineation is based on the ≥95% ANI threshold. Species are indicated by colored branches and numbered 1–55. Tree scale represents the number of nucleotide substitutions per site. Rings outside the tree (inner to outer) show the species group, SCC*mec* type, source, and geographical location. Species group is based on the classification by Lamers et al. (2012) and member species of each species group are indicated by alternating light and dark shades of the same color. Numbers in the color legend that represent the 55 species are color-coded according to their membership in species groups. Some species were designated A, B, or C if they were named the same species in NCBI but likely represent different species based on phylogenetic position and pairwise ANI values of <95%. ISS, International Space Station. Stars indicate those 16 genomes with taxonomic ambiguities.

(Lamers et al. 2012; Rolo et al. 2017). The four other species groups *Auricularis*, *Hyicus-Intermedius*, *Saprophyticus*, and *Simulans* were also clearly defined on the phylogeny and consisted of 31 (three species), 91 (seven species), 269 (17 species), and 70 (six species) genomes, respectively. Within each species group, estimation of genome-wide ANI values also revealed distinct boundaries between species (supplementary fig. S3 and table S4, Supplementary Material online).

We detected 16 genomes that showed taxonomic discrepancies based on their phylogenetic placement and ANI values (marked with a star in fig. 1; supplementary table S5, Supplementary Material online). Some genomes that were initially identified as belonging to a single species can be delineated into more than one species (labeled A, B, and C in fig. 1). *Staphylococcus fleurettii* A and B species were found in species groups *Sciuri* (consistent with the results of reference; Lamers et al. 2012) and *Hyicus-Intermedius*, respectively. Two genomes named as *Staphylococcus pasteuri* in NCBI (GCF_000494875.1 and GCF_002276895.1) were more closely related to genomes in *Staphylococcus warneri* A. The two *Staphylococcus massiliensis* genomes were more closely related to the *Hyicus-Intermedius* species group, rather than the *Saprophyticus* group as previously reported (Lamers et al. 2012). *Staphylococcus pettenkoferi* clusters within the *Auricularis* species group, rather than the *Saprophyticus* species group as previously reported (Lamers et al. 2012). The use of genomic analysis to resolve complex or conflicting taxonomic assignments in *Staphylococcus* has been previously demonstrated (Madhaiyan et al. 2020; Lavecchia et al. 2021), and our results will further inform future classification efforts in this genus.

Overall, we found that CoNS is more diverse than previously recognized, highlighting the need for a more extensive sampling of this obscure bacterial group. Our compiled dataset shows that numerous species were poorly represented in current genomic sequencing efforts. Moreover, we also found 16 out of 1,876 genomes that have ambiguous taxonomic assignments.

CoNS Species Are a Reservoir of Diverse and Transferrable Resistance Genes

The horizontally acquired antimicrobial resistance genes we detected in silico represented a total of 16 antimicrobial classes (table 1, supplementary fig. S4 and table S6, Supplementary Material online). The most commonly detected resistance genes were those that encode resistance to beta-lactams (present in 927 genomes), fluoroquinolones (742 genomes), and diaminopyrimidines (727 genomes). Several antimicrobial resistance genes were found to be unique to certain species. In addition to *norA* which was identified in almost every *S. epidermidis* strain, we also detected streptogramin A resistance genes *vatB* and *vgaB*, respectively ($n = 9$ genomes) which are often transferred together on a single plasmid (Haroche et al. 2003). Unique to *S. haemolyticus* A was the clindamycin resistance gene *lsaB* ($n = 3$ genomes). Two genes which confer resistance to macrolide, lincosamide, and streptogramin B were detected only in *Staphylococcus lentus* (*erm(43)*; $n = 5$ genomes) and *Staphylococcus chromogenes* (*ermT*; $n = 5$ genomes). Unique to every *S. saprophyticus* genome ($n = 112$ genomes) was *fusD*. The *fusF* gene (fusidic acid resistance; $n = 23$) was detected only in the two *S. cohnii* species A and B. Finally, the *salA* gene was found in majority of *Staphylococcus sciuri* genomes (resistance to lincosamides and streptogramins; $n = 28$ genomes).

The protein product of *mecA* is the penicillin-binding protein PBP2a, which has a low affinity to methicillin and other broad-spectrum beta-lactams (Fishovitz et al. 2014). We detected the *mecA* gene in 18 species (fig. 1 and supplementary fig. S4, Supplementary Material online). The mobile SCC*mec*, which carries the *mecA* gene, can be differentiated into 14 structurally diverse types (International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements [IWG-SCC] 2009). Across our entire CoNS data set, we detected nine SCC*mec* types (I, II, III, IV, V, VIII, IX, X, and XIII). The most common SCC*mec* types were type V ($n = 168$ genomes), type III ($n = 158$ genomes), and type IV ($n = 107$ genomes). *Staphylococcus epidermidis*, which has highest number of genomes, had eight SCC*mec* types (all except type V).

CoNS Species Harbor Many Virulence Genes

We also examined the presence of virulence genes using in silico methods (supplementary fig. S5 and table S7, Supplementary Material online). We identified 32 virulence genes, of which the most common was *clpP*. It was detected in 1,144 non-*S. epidermidis* genomes and 50 *S. epidermidis* genomes. The proteolytic activity of ClpP plays a major role in virulence, stress response, and physiology of *S. aureus* (Michel et al. 2006) and other pathogenic species (Knudsen et al. 2013; Zhao et al. 2016; Zheng et al. 2020). Other virulence genes that were frequently detected were particularly notable. The genes *hld* and *icaA* were found almost exclusively in *S. epidermidis* and closely related species. A total of 520 *S. epidermidis* genomes, three *Staphylococcus saccharolyticus* genomes and two *Staphylococcus simiae* genomes contained the *hld* gene. The *hld* gene codes for a delta-hemolysin that is cytolytic to neutrophils and erythrocytes and can induce chronic inflammatory skin disease by triggering mast cell degranulation in *S. aureus* (Nakamura et al. 2013). The *icA* gene was detected in 330 *S. epidermidis* genomes, 94 *S. capitis* genomes, 11 *Staphylococcus caprae* genomes, 3 *S. saccharolyticus* genomes, and 2 *S. simiae* genomes. The *ica* operon, of which *icaA* is part of, functions in the production of polysaccharide intercellular adhesin which mediates cell-to-cell adhesion and biofilm formation

Table 1.

Number of genomes per species carrying horizontally acquired genes that encode resistance for each antimicrobial class

Species	Aminogl-ycoside	Diamino-pyrimidine	Fluoro-quinolone	Fosfo-mycin	Fusidic acid	Lincos-amide	Macr-olide	Mupi-rocin	N-glycoside	Oxazolid-inone	Phenicol	Pleuro-mutilin	QAC	Strepto-gramin	Tetra-cycline	Beta-lactam
argensis	0	0	0	0	0	7	1	0	0	0	0	0	0	7	0	0
arlettae	1	3	3	0	0	4	6	0	0	1	6	0	3	6	4	16
auricularis	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2
caeli	2	2	0	0	0	2	2	0	0	0	2	0	0	2	2	0
capitis	44	17	2	3	8	28	19	2	0	12	13	0	32	30	8	67
caprae	5	0	1	0	0	3	4	0	0	0	1	0	0	4	0	8
carnosus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
chromogenes	2	2	0	0	1	7	7	0	0	0	3	0	0	7	9	12
cohnii_A	2	2	0	5	19	6	13	0	0	0	3	0	0	13	1	2
cohnii_B	3	2	3	2	5	11	24	1	0	2	3	0	6	22	4	2
condimenti	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
debuckii	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
devriesei_A	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
devriesei_B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
edaphicus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
epidermidis	346	675	731	5	148	288	406	167	26	19	47	0	379	419	110	635
equorum_A	2	2	0	2	0	1	5	0	0	0	2	0	0	5	1	3
equorum_B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
felis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
fleurettii_A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
fleurettii_B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
gallinarum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
haemolyticus_A	1	0	0	0	3	3	12	0	0	0	0	0	11	13	0	5
haemolyticus_B	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0
hominis	45	2	2	1	16	48	67	13	6	0	4	0	76	73	21	95
kloosii_A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
kloosii_B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
lentus	1	1	0	1	0	6	6	0	0	1	1	0	0	6	0	1
lugdunensis	1	0	0	0	1	2	2	0	0	0	0	0	0	2	0	13
massiliensis	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
microti	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
muscae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
nepalensis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pasteuri	1	1	0	0	2	1	4	0	0	0	1	0	1	4	3	7
petrasii_A	3	0	0	0	2	1	2	1	2	0	0	0	0	2	0	7
petrasii_B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
petrasii_C	0	0	0	0	0	0	1	0	0	0	0	0	1	1	1	0
pettenkoferi	1	0	0	1	0	9	3	1	0	0	0	0	0	9	1	5
piscifermentans	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(continued)

Table 1. Continued

Species	Aminogl-ycoside	Diamino-pyrimidine	Fluoro-quinolone	Fosfo-mycin	Fusidic acid	Lincos-amide	Macr-olide	Mupi-rocin	N-glycoside	Oxazolid-inone	Phenicol	Pleuro-mutilin	QAC	Strepto-gramin	Tetra-cycline	Beta-lactam
pseudoxylosus	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
saccharolyticus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
saprophyticus	0	8	0	0	112	9	18	0	0	0	0	0	4	18	25	4
sciuri	5	1	0	0	0	30	3	1	0	3	4	28	0	3	4	4
simiae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
simulans_A	0	0	0	0	0	8	2	0	0	0	2	0	0	5	4	7
simulans_B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
stepanovicii	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
succinus	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
ursi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
vitulinus	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
warneri_A	5	3	0	0	2	5	11	0	0	0	3	0	3	11	1	19
warneri_B	0	3	0	0	0	0	1	0	0	0	0	0	2	1	3	4
xylosus_A	0	0	0	0	0	1	1	0	0	0	0	0	0	1	3	1
xylosus_B	0	0	0	1	0	1	1	0	0	0	0	0	0	0	1	0
xylosus_C	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Total	470	727	742	23	319	485	622	186	34	38	95	28	519	665	211	927

NOTES.—QAC, quaternary ammonium compounds. Details of the distribution of specific resistance genes per genome are presented in supplementary figure S5 and table S7, Supplementary Material online.

(Cue et al. 2012). This adhesin has been implicated in catheter-associated infections (Arciola et al. 2001). Lastly, many non-*epidermidis* species carried the *cap8* genes, which are involved in the synthesis of the serotype 8 capsular polysaccharide and is the most prevalent capsule type in clinical isolates of *S. aureus* (Luong and Lee 2002; O'Riordan and Lee 2004). In *S. aureus*, the capsule enhances virulence by impeding phagocytosis and stimulating colonization and persistence on mucosal surfaces (Luong and Lee 2002; O'Riordan and Lee 2004). Capsule genes had been previously detected in *S. lugdunensis* (Lebeurre et al. 2019), and our results greatly expand the taxonomic distribution of *cap8* genes. We detected four types of *cap8* genes that were frequently found in *S. sciuri, S. chromogenes, S. lugdunensis*, and *S. haemolyticus* A genomes.

Ten CoNS Species Exhibit Frequent Intraspecies Recombination

Recombination is known to contribute to the evolution and ecology of *S. aureus* (Castillo-Ramírez et al. 2012; Driebe et al. 2015; Spoor et al. 2015; Murray et al. 2017), yet little is known of the frequency of recombination in CoNS. We can only carry out estimates of intraspecies recombination for those species with sufficient number of representative genomes; hence, we selected 10 species with more than 30 representative genomes. We used the mcorr program to estimate six different evolutionary and recombination parameters for each of the ten most common species (fig. 2 and supplementary table S8, Supplementary Material online). The species with the lowest sequence diversity were *S. epidermidis* (0.002356) and *S. warneri* A (0.002437), whereas *S. sciuri* (0.006427) and *Staphylococcus simulans* (0.00641) have the highest values. The recombinational divergence (ϕ) and mutational divergence (θ) also varied among the species. The recombinational and mutational divergence values refer to the average number of recombination events and mutation events, respectively, per locus since coalescence of a pair of individual strains (Lin and Kussell 2019). The relative rate of recombination to mutation (ϕ/θ or γ/μ) ranged from 2.45 in *S. haemolyticus* A to 8.30 in *Staphylococcus hominis*. The mean fragment size ($\bar{f}$) of recombination events ranged from 433.44 bp in *S. lugdunensis* to 955.20 bp in *S. capitis*. Finally, the recombination coverage (c), which represents the proportion of sites in the core genome whose diversity was derived from recombination, was lowest in *S. warneri* A (10.91%) and highest in *S. lugdunensis* (30.72%), *S. simulans* A (30.16%), and *S. sciuri* (27.89%). For comparison, *S. aureus* ($n = 308$ genomes from reference Lin and Kussell 2019) showed lower values for diversity (0.015), recombination divergence (0.042), and γ/μ (1.0) and higher values for mutational divergence (0.042) and recombination coverage (36%). The mean fragment length of recombined DNA in CoNS is

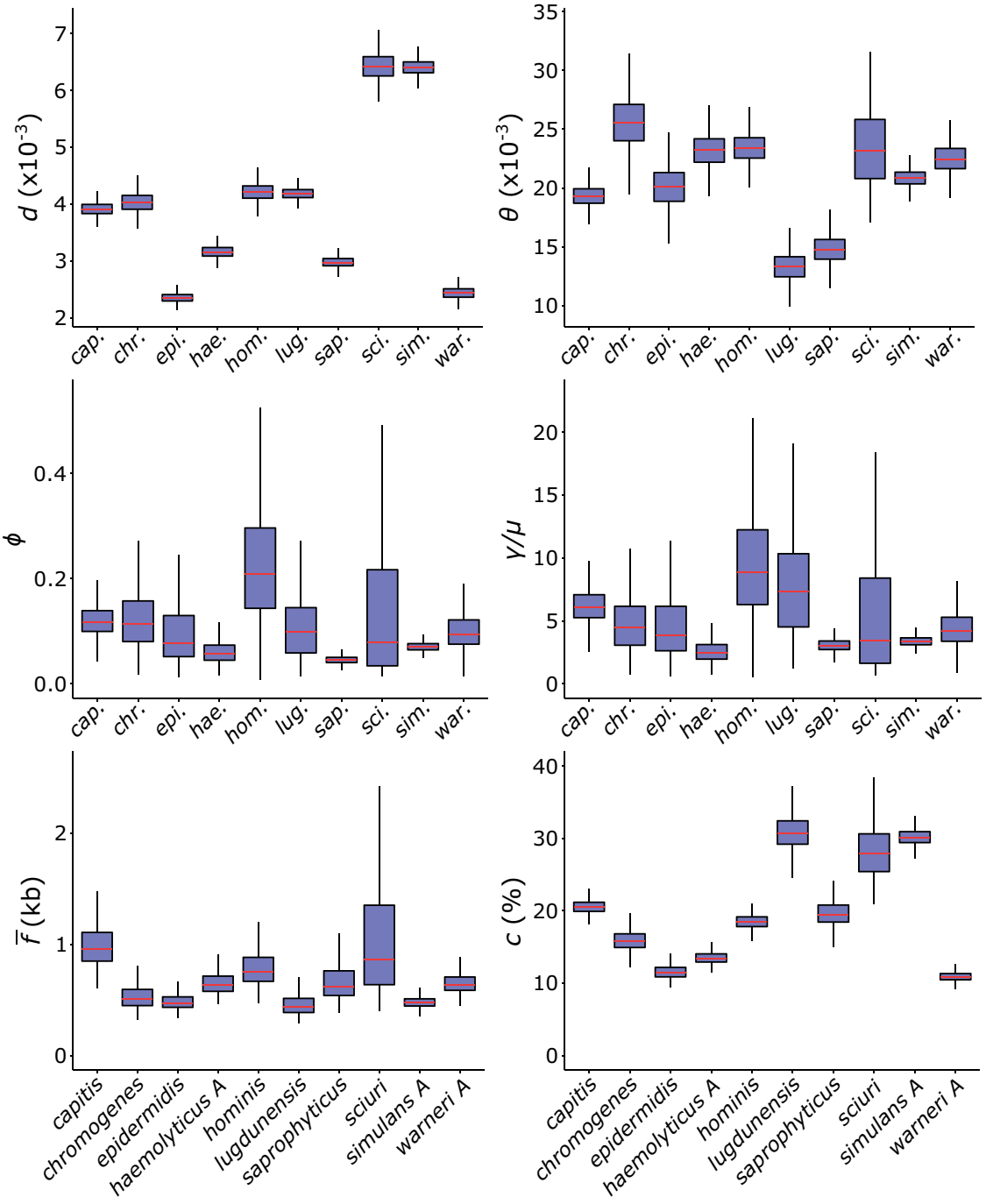

Fig. 2.—Intraspecies recombination in ten most common CoNS species. Recombination parameters were calculated by mcorr (Lin and Kussell 2019). Core genome alignment of each species was used as input in mcorr with 1,000 bootstrapped replicates. d, diversity brought into the population by recombination and clonal diversity; θ, mutational divergence; ϕ, recombinational divergence; γ/μ relative rate of recombination to mutation (equivalent to ϕ/θ); $\bar{f}$, mean fragment size of a recombination event; c, recombination coverage.

comparable to that found in *S. aureus* (550 bp; Lin and Kussell 2019).

To further examine within-species recombination, we also identified the specific genes in the core and shared accessory genomes that were frequently recombined in the ten most common species (supplementary fig. S6 and table S9, Supplementary Material online). Five species were characterized as having a quarter of their pan-genome consisting of genes that have had a history of recent and ancestral recombination. These included *S. epidermidis* (2,077 genes, representing 25.03% of its pan-genome), *S. chromogenes* (1,025 genes; 27.63%), *S. hominis* (1,449; 24.55%), *S. haemolyticus*

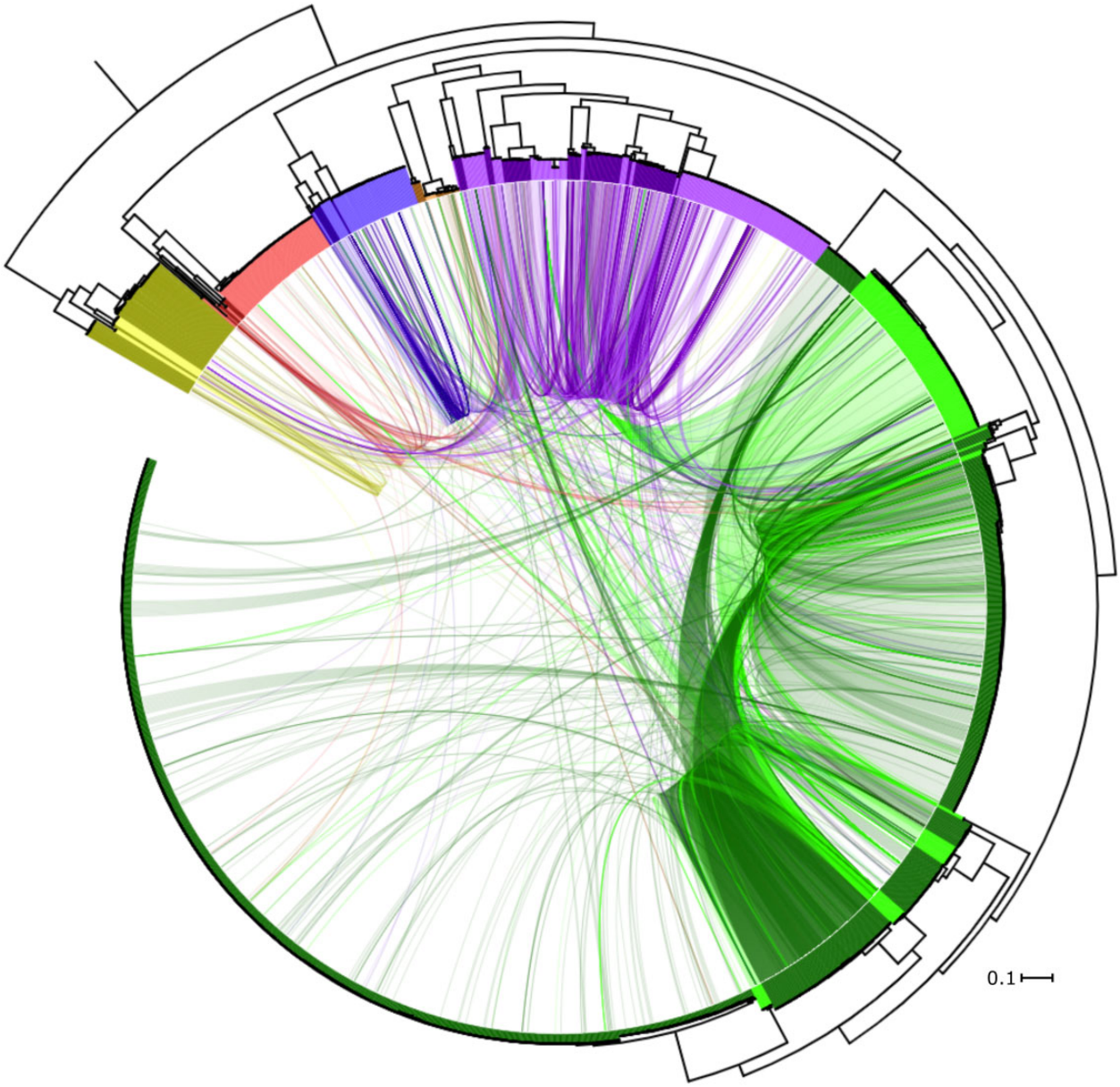

Fig. 3.—Interspecies HGT. Results of the fastGEAR recombination detection program were mapped to the core genome phylogenetic tree. For each recombination event, the donor and recipient genomes were linked with a colored line. The color of the connecting line is based on the color of the recipient genome. The tree is identical to that in figure 1. For visual clarity, recombination between genomes within a species was not mapped onto the tree. Only recent recombinations (and not ancestral recombinations) were included.

A (1,570; 23.73%), and *S. sciuri* (1,152; 21.42%). The other five species have fewer genes that had experienced recombination: *S. lugdunensis* (185 genes, representing 5.74% of its pan-genome), *S. warneri* A (345 genes; 8.41%), *S. saprophyticus* (642; 10.99%), *S. capitis* (711; 16.57%), and *S. simulans* A (773; 18.44%). Some of the most common functions of highly recombined genes across the ten species included those associated with response and/or resistance to antimicrobials (*bceB*, *bmrA*, *swrC*, *gyrA*, and *lnrL*) and heavy metals (*arsB*), surface-associated and adherence-related proteins related to virulence and biofilm formation (*ebh*, *fbe*, *sdrF*, *sraP*, *sdrG*, *sasG*, *uafA*, *bca*, and *sdrC*), type VII secretion system (*essC*), iron capture (*isdB*), recombination process (*bin3*, *hin*, *sbcC*, and *xerC*), and metabolic processes that involve ion binding (*bioB*, *gltA*, *ipdC*, *lip*, *lipA*, *pgcA*, *phoB*, and *ppdK*).

Highways of Interspecies HGT in CoNS

Mapping the recent recombination events across the CoNS phylogeny allowed us to ask whether there was heterogeneity

in the frequency of recent transfers between species. Using fastGEAR results for the entire data set (supplementary table S10, Supplementary Material online), we found that although interspecies HGT was widespread, species tended to transfer with certain partners that were members of their own species group (fig. 3). This was particularly notable for species that comprise the species groups *Epidermidis*, *Saprophyticus*, and *Simulans*. For example, the most common species pairs with the highest frequencies of interspecies transfer were *S. capitis* and *S. caprae*, *S. haemolyticus* A and *S. haemolyticus* B, and *S. saprophyticus* and *Staphylococcus edaphicus* (fig. 3). For each pair of species, we detected a total of 889, 616, and 381 transfer events, respectively. The species pairs with frequent transfer events traversing boundaries between species groups were *S. hominis* and *S. cohnii* B (128 transfer events), *S. capitis* and *Staphylococcus arlettae* (31 transfer events), and *S. haemolyticus* A and *Staphylococcus equorum* A (28 transfer events).

We also characterized frequent donors and recipients of transfer events across the entire CoNS phylogeny. The most frequent donors to other species were *S. caprae* (971 transfers) and two of the ten species with the highest number of representative genomes (*S. haemolyticus* A [693 transfers] and *S. saprophyticus* [644 transfers]). When adjusted for the number of genomes per species, those with the highest number of donated DNA per genome were *Staphylococcus devriesei* B (275 transfers per genome), *S. haemolyticus* B (126 transfers per genome), and *S. caprae* (88 donations per genome). The most frequent recipients of transferred DNA were *S. capitis* (1,627 receipts), *S. haemolyticus* A (667 receipts), and *S. edaphicus* (506 receipts). When adjusted for the number of genomes per species, the most common recipients were *S. edaphicus* (506 receipts per genome), *Staphylococcus caeli* (217 receipts per genome) and *S. simulans* B (187 receipts per genome). Overall, these results show that certain CoNS species were frequent participants (either as donors or recipients) in interspecies HGT.

To explore the contributions of ecology to the observed biases in interspecies HGT, we compared the frequencies of recombination between genomes that came from humans, live animals (domestic and wild), environmental (including plants), and animal-derived foods (consisting of milk, cheese, poultry, beef, pork, and fermented seafood and meats; supplementary fig. S7, Supplementary Material online). When the total number of recombination events was considered, frequent interspecies recombination involved human-associated strains, whether between strains from humans and animals ($n = 1{,}793$ transfers), between strains from humans only ($n = 1{,}154$ transfers) and between strains from humans and environment ($n = 973$ transfers). However, this result is likely due to human-associated strains making up more than half the data set, thus providing more potential recombination partners. When normalizing the ecological group sizes for more even comparisons (i.e., the number of observed recombination events relative to the total number of possible recombining pairs), we found that strains from animal-derived foods recombined with each other at a frequency of 0.0182, indicating that 1.82% of all possible transfers were detected. Frequent transfers also occurred between strains from animals ($n = 0.0108$) and between strains from animal and animal-derived foods ($n = 0.0080$).

Discussion

CoNS species were historically defined by delimitation from the more pathogenic coagulase-producing *S. aureus* (Fairbrother 1940). Although coagulase production was established initially for diagnostic procedure, it gradually became a clinical classification scheme to exclude the purportedly nonpathogenic species of *Staphylococcus*. This consequently led to the lack of recognition for the clinical importance and poor taxonomic sampling of CoNS. As such, the diversity CoNS species have not been confidently established.

Our study revealed the importance of recombination and HGT in CoNS evolution. The variable rates of recombination and biases in transfer partners imply that certain CoNS species function as hubs of gene flow and major reservoir of genetic diversity for the entire genus. We detected biases in HGT between certain species within and between species groups. The highest frequencies of interspecies receipt and donation of recombined DNA were detected in *S. capitis*, *S. caprae*, *S. hominis*, *S. haemolyticus*, and members of the *Saprophyticus* species group. The distribution of virulence and antimicrobial resistance genes, including the intermingled distribution of different SCC*mec* types across the phylogeny, suggests multiple independent acquisition events in different lineages. Such extensive HGT and shuffling of clinically relevant genes between strains and species can facilitate the emergence of high-risk clones with unpredictable phenotypes. Bacterial pathogens from animal-derived food items and live animals may act to facilitate transfers with strains from other sources, including humans (Silva et al. 2014; Thanner et al. 2016; Hernández-González and Castillo-Ramírez 2020). This has important implications to designing agricultural practices managing foodborne disease outbreaks and transmission. These data support the concept that surveillance of *Staphylococcus* in both human and nonhuman sources could play a critical role in the early identification of bacterial clones that have acquired novel genetic material and/or phenotypes.

A pair of lineages or species that exchange DNA more frequently with each other than they do with others is said to be linked by a highway of gene sharing (Beiko et al. 2005). These highways have been previously reported in higher taxonomic groups, as in the case of phyla (e.g., HGT between Aquificae, Proteobacteria and Archaea; Eveleigh et al. 2013) and genera (e.g., HGT between *Synechococcus* and *Prochlorococcus*; Zhaxybayeva et al. 2009). Frequent

recombination between distantly related taxa may originate from co-occupancy of the same ecological niche over a long period of time or to syntrophic or other close symbiotic relationships (Russell and Cavanaugh 2017; Zeng et al. 2017; Arroyo et al. 2019). At lower taxonomic levels, highways of recombination have also been previously reported. In *Streptococcus pneumoniae*, frequent DNA donors and recipients were those strains that lack genes for capsule biosynthesis (Chewapreecha et al. 2014). Noncapsulated *S. pneumoniae* are considered to be less associated with disease than the capsulated strains and thus are excluded from the targets of currently available polysaccharide vaccines (Daniels et al. 2016; Bradshaw and McDaniel 2019). Heterogeneity in recombination frequency between capsulated and noncapsulated *S. pneumoniae* may be a reflection of differential rates of response and adaptation to selection pressures from the environment or host (Chewapreecha et al. 2014). Highways of recombination have also been observed between different subspecies of *Salmonella enterica*, including those subspecies that are not often implicated in disease (Park and Andam 2020). In these two pathogenic species, it appears then that the limited association with disease or clinical interventions may have facilitated greater opportunities for genetic recombination. For CoNS, interspecies highways of HGT are likely shaped by ecology, with animal hosts and animal-derived foods acting as major hubs of genetic exchange. Our findings suggest that certain CoNS species can act as a major reservoir for many transferrable antimicrobial resistance determinants and virulence-related genes. These results also suggest that CoNS species may be an important reservoir of genetic diversity for the wider *Staphylococcus* genus. For example, frequent interspecies transfer has been reported between the CoPS species *S. aureus* and CoNS species *S. epidermidis* due to overlapping niche space in the human skin and nasal pharynx (Méric et al. 2015).

We recognize the limitations of this study. First, our reliance on publicly available genomes means that certain CoNS species and geographical regions were disproportionately represented in our analysis. Rare species such as those consisting of less than ten genomes remain poorly characterized in terms of estimating species-level genomic variation, recombination rates, and population structure. Limited ecological data on many non-*aureus* species can obfuscate inferences of the contributions of shared or overlapping ecological niches to their evolution. Second, existing databases for querying antimicrobial resistance genes, virulence-associated genes and SCC*mec* types are limited to those built for the more intensively studied *S. aureus*. Hence, CoNS-specific genetic variants are likely to exist but remain invisible from current in silico detection methods and databases. This makes it problematic to track their prevalence and spread over the long term. Notwithstanding these limitations, we obtained sufficient representation of CoNS that can be used as basis for developing genus-wide databases for characterizing CoNS species in the future. Third, recombination between closely related strains with very similar DNA sequences is difficult to detect using current methods. Hence, our results are likely an underestimation of the true extent of recombination between CoNS found in nature. Lastly, inferring highways of transfer between ancestral lineages remains challenging. This is because of the uncertainty due to missing lineages that can obscure true donors and those DNA segments that have been horizontally acquired and subsequently lost.

The results presented here spawn outstanding questions and open multiple avenues for future investigations. First, the ecology of specific CoNS species remains poorly studied. This includes the genetic basis of adaptation to different ecological niches within the human body and specific clinical presentations. Many CoNS species are also known to colonize a diverse array of animals (Bhargava and Zhang 2012; Kern and Perreten 2013; Mama et al. 2019), yet the range of host preferences and instances of switching between hosts (including animal to human) remain poorly studied. A systematic surveillance and sampling from both human and animals will greatly advance our understanding of CoNS ecology. Moreover, overlapping niches between CoNS and CoPS may reveal previously unrecognized ecological interactions, including recombination and HGT (Méric et al. 2015), that contribute to each other's success as commensals and pathogens. Second, our results on species-specific genomic variation will form the foundation of in vitro and in vivo investigations of clinically relevant phenotypes, such as adhesion, biofilm formation, and internalization by and persistence in host cells. This knowledge will be particularly useful for developing effective therapeutic options for treatment of CoNS infections. Third, species-specific genetic variants can aid in precise laboratory detection in carriage and disease. Difficulties in species identification may lead to a lack of noted infections caused by *Staphylococcus* other than *S. aureus*. Hence, the real impact of CoNS species as etiological factors of human infections may remain underreported or overlooked. The implementation of reliable genetic methods in clinical practice will therefore improve the identification process and result in more precise diagnosis of staphylococcal infections. Defining the evolutionary processes such as recombination and HGT that contribute to CoNS diversity and speciation is the first step toward effective surveillance, diagnostics, and control of CoNS infections.

Materials and Methods

Data Set

We retrieved 2,258 genome assemblies of all named CoNS species from the National Center for Biotechnology Information (NCBI) Reference Sequence database in November 2020. To reduce ambiguity, we excluded those species with coagulase-variable phenotypes (*Staphylococcus agentis, S. hyicus*, and *S. schleiferi*). To ensure that only high-quality genomes were included, we used QUAST v.5.0.2

(Gurevich et al. 2013) to identify and exclude genomes with >200 contigs and <40,000 bp N50 scores. We also used CheckM v.1.1.2 to exclude genomes with <90% completeness and >5% contamination (Parks et al. 2015). A total of 1,876 genomes passed these criteria and were subsequently used in all downstream analyses (supplementary table S1, Supplementary Material online). The number of contigs ranged from 1 to 194 and N50 values ranged from 40,254 to 2,941,886 bp. To ensure that all genomes in the data set belong to the *Staphylococcus* genus, we calculated the POCP using CompareM v.0.1.2 (https://github.com/dparks1134/CompareM; last accessed August 1, 2021) and Basic Local Alignment Search Tool (BLAST) P option to compare amino acid identity (Konstantinidis and Tiedje 2005). As defined by Qin et al. (2014), the POCP should be greater than 50% between genomes to be considered members of the same genus. Because carrying out a full pairwise comparison of all 1,876 genomes was not computationally feasible, we chose one representative genome from each of the 55 species. Where available, the reference genome as identified by NCBI was used. In cases when there was no reference genome for the species available in NCBI, the highest-quality genome was chosen. In cases whereby the highest-quality genome (based on N50, number of contigs, completeness, and contamination) and the selection of representative genome was difficult to assess, we randomly chose a representative genome of a species. Representative genomes used for POCP analysis are indicated in supplementary table S1, Supplementary Material online. To confirm species identity, we used fastANI v.1.1 to calculate the ANI for all pairs of genomes (Jain et al. 2018). We used the ≥95% ANI threshold to define a species (Jain et al. 2018). Genomes identified as a single species in NCBI but exhibited <95% ANI were designated with capital letters (e.g., *Staphylococcus xylosus* A, B, or C) to distinguish them as separate species.

Pan-Genome Analysis and Phylogenetic Reconstruction

We used Prokka v.1.14.6 to annotate the genome sequences (Seemann 2014). To estimate the pan-genome of the entire data set, we used Panaroo v.1.2.3 that employs a graph-based algorithm (Tonkin-Hill et al. 2020) to cluster gene families using CD-HIT (Fu et al. 2012). Panaroo enables stringent quality control checks to correct annotation errors, fragmented assemblies, mistranslation, and contamination of the genomes (Tonkin-Hill et al. 2020). We used Panaroo's strict mode for contamination and erroneous annotation removal to ensure that only high-quality genomes were included. The presence or absence of a gene in each genome was used to define the core and accessory genes. The genes in the pan-genome were classified into core genes (present in ≥99% of genomes), soft-core genes (present in 95% to <99% of genomes), shell genes (present in 15% to <95% of genomes), and cloud genes (present in <15% of genomes). Sequences of each gene family were aligned using MAFFT (Katoh et al. 2009). The aligned sequences of the core genes were then concatenated and were used as input to build a maximum-likelihood phylogenetic tree using IQTree v.2.0.3 (Minh et al. 2020). We used a general time reversible nucleotide substitution model (Tavaré 1986) and a FreeRate rate heterogeneity model with four categories. We used the IQTree built-in UFBoot2 (Hoang et al. 2018) to carry out 1,000 ultrafast bootstrap replicates. We also reran the Panaroo pipeline (Tonkin-Hill et al. 2020) for each species that has at least 30 genomes each to determine the gene content of each species. Phylogenetic trees were visualized using the online platform Interactive Tree of Life (IToL; Letunic and Bork 2019).

In Silico Detection of Resistance Genes, Virulence Genes, Plasmids, and SCC*mec*

We used ABRicate v.0.8.13 (https://github.com/tseemann/abricate; last accessed August 1, 2021) to detect horizontally acquired genes conferring resistance to different antimicrobial classes and genes associated with virulence. ABRicate employs BLAST (Altschul et al. 1990) to screen the Resfinder (Bortolaia et al. 2020) and Virulence Factor databases (Liu et al. 2019; last updated in January 2021). To increase the likelihood of accurate gene detection, we used 95% sequence identity and 95% coverage thresholds for resistance and virulence gene detection. Plasmids were identified using the PlasmidFinder database (Carattoli and Hasman 2020). We used SCCmecFinder v.1.2 (Kaya et al. 2018) to identify the presence and type of *mecA*-carrying SCC*mec* using default minimum thresholds of >60% for sequence coverage and >90% sequence identity.

Recombination Detection

We used two approaches to detect genome-wide recombination. First, we used the coalescent-based mcorr program, which measures the degree to which any two loci in the core genome and are separated by defined number of nucleotides have correlated synonymous substitutions (Lin and Kussell 2019). The correlation profiles were then used to calculate six parameters related to recombination: diversity, mutational divergence, recombinational divergence, proportion of sites in the genome whose diversity was derived from recombination (or recombination coverage), mean recombination fragment size, and relative rate of recombination to mutation (Lin and Kussell 2019).

We also used fastGEAR to identify specific core and shared accessory genes that have experienced recent and ancestral recombination (Mostowy et al. 2017). A recent recombination involves a few strains, whereas an ancestral recombination affects entire lineages (Mostowy et al. 2017). FastGEAR uses a hidden Markov model (HMM) to compare polymorphic sites in the strain's sequence and compares them to those found in members of its own lineage and in other lineages.

The comparison is made over multiple HMM iterations, with parameters updated for each iteration based on results from the prior run. To test the significance of the inferred recombinations and identify false-positive results, we used a diversity test implemented in fastGEAR to compare sequence variation of the recombined fragment to its background (Mostowy et al. 2017). By default, fastGEAR determines the recipient genome of a recent recombination sequence and the lineage from which the sequence was donated. To predict the origin of the recently recombined regions, we identified the specific donor genome by developing a custom script to extract the recombined sequence from the recipient genome and then compare it to members of the donor lineage using BLAST (Altschul et al. 1990). The closest matching donor with a BLAST *e*-value of 1e−5 or less was then paired with the recipient. Recent recombinations were visualized in IToL (Letunic and Bork 2019) and the postprocessing scripts provided by fastGEAR.

We used the default parameters for each program unless indicated otherwise.

Statistical Tests

To account for variation in the number of genomes per species, the frequencies of DNA donations were adjusted by dividing the number of observed donations per species by the number of genomes of a species. A similar calculation was done for DNA receipts. To account for the differences in the number of genomes from each ecological source, we used the combinations formula as a normalizing factor to compare frequencies of recombination between strains from the same source:

$$\frac{2r}{p(p-1)}$$

and different sources:

$$\frac{r}{pq},$$

where r is the number of recombination events, p is the number of genomes from the first source, and q is the number of genomes from the second source.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was supported by the National Institutes of Health (NIH) (Grant no. 1R35GM142924) and startup funds from the College of Arts and Sciences of the University at Albany, State University of New York to C.P.A. This work was also supported by the University of New Hampshire Graduate School Dissertation Year Fellowship to J.T.S. The funders had no role in study design, data collection and analysis, decision to publish, and preparation of the manuscript and the findings do not necessarily reflect views and policies of the authors' institutions and funders.

The authors thank the UNH Resource Computing Center where all bioinformatics analyses were performed. They thank Cooper J. Park, Isaiah P. A. Lee, and Kiarash Rahmani Eliato for useful discussions and thoughtful suggestions in our analyses. They acknowledge the use of publicly available world map from simplemaps.com. The authors declare no competing interests.

Author Contributions

J.T.S. carried out all bioinformatics analyses. J.T.S. and C.P.A. wrote the manuscript. All authors read and approved the final manuscript.

Data Availability

All genome sequences and associated metadata used in this study are publicly available from NCBI. Accession numbers for each genome are listed in supplementary table S1, Supplementary Material online.

Literature Cited

Ahmed A, et al. 2019. Catheter related recurrent blood stream infection caused by linezolid-resistant, methicillin resistant *Staphylococcus haemolyticus*; an emerging super bug. J Pak Med Assoc. 69(2):261–263.

Akiyama H, Kanzaki H, Tada J, Arata J. 1998. Coagulase-negative staphylococci isolated from various skin lesions. J Dermatol. 25(9):563–568.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410.

Arciola CR, Baldassarri L, Montanaro L. 2001. Presence of icaA and icaD genes and slime production in a collection of staphylococcal strains from catheter-associated infections. J Clin Microbiol. 39(6):2151–2156.

Arroyo FA, Pawlowska TE, Choat JH, Clements KD, Angert ER. 2019. Recombination contributes to population diversification in the polyploid intestinal symbiont Epulopiscium sp. type B. ISME J. 13(4):1084–1097.

Asante J, et al. 2021. Multidrug-resistant coagulase-negative staphylococci isolated from bloodstream in the uMgungundlovu district of KwaZulu-Natal Province in South Africa: emerging pathogens. Antibiotics (Basel) 10(2):198.

Becker K, Heilmann C, Peters G. 2014. Coagulase-negative staphylococci. Clin Microbiol Rev. 27(4):870–926.

Beiko RG, Harlow TJ, Ragan MA. 2005. Highways of gene sharing in prokaryotes. Proc Natl Acad Sci U S A. 102(40):14332–14337.

Bhargava K, Zhang Y. 2012. Multidrug-resistant coagulase-negative Staphylococci in food animals. J Appl Microbiol. 113(5):1027–1036.

Bortolaia V, et al. 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother. 75(12):3491–3500.

Both A, et al. 2021. Distinct clonal lineages and within-host diversification shape invasive Staphylococcus epidermidis populations. PLoS Pathog. 17(2):e1009304.

Bradshaw JL, McDaniel LS. 2019. Selective pressure: rise of the nonencapsulated pneumococcus. PLoS Pathog. 15(8):e1007911.

Carattoli A, Hasman H. 2020. PlasmidFinder and in silico pMLST: identification and typing of plasmid replicons in whole-genome sequencing (WGS). Methods Mol Biol. 2075:285–294.
Castillo-Ramírez S, et al. 2012. Phylogeographic variation in recombination rates within a global clone of methicillin-resistant *Staphylococcus aureus*. Genome Biol. 13(12):R126.
Chewapreecha C, et al. 2014. Dense genomic sampling identifies highways of pneumococcal recombination. Nat Genet. 46(3):305–309.
Crosby HA, Kwiecinski J, Horswill AR. 2016. *Staphylococcus aureus* aggregation and coagulation mechanisms, and their function in host-pathogen interactions. Adv Appl Microbiol. 96:1–41.
Cue D, Lei MG, Lee CY. 2012. Genetic regulation of the intercellular adhesion locus in staphylococci. Front Cell Infect Microbiol. 2:38.
Daniels CC, Rogers PD, Shelton CM. 2016. A review of pneumococcal vaccines: current polysaccharide vaccine recommendations and future protein antigens. J Pediatr Pharmacol Ther. 21(1):27–35.
Driebe EM, et al. 2015. Using whole genome analysis to examine recombination across diverse sequence types of *Staphylococcus aureus*. PLoS One 10(7):e0130955.
Eveleigh RJM, Meehan CJ, Archibald JM, Beiko RG. 2013. Being *Aquifex aeolicus*: untangling a hyperthermophile's checkered past. Genome Biol Evol. 5(12):2478–2497.
Fairbrother R. 1940. Coagulase production as a criterion for the classification of the staphylococci. J Pathol. 50(1):83–88.
Fishovitz J, Hermoso JA, Chang M, Mobashery S. 2014. Penicillin-binding protein 2a of methicillin-resistant *Staphylococcus aureus*. IUBMB Life. 66(8):572–577.
Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28(23):3150–3152.
Fux CA, et al. 2005. Dynamics of hemodialysis catheter colonization by coagulase-negative staphylococci. Infect Control Hosp Epidemiol. 26(6):567–574.
Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29(8):1072–1075.
Haroche J, et al. 2003. Clonal diversity among streptogramin A-resistant *Staphylococcus aureus* isolates collected in French hospitals. J Clin Microbiol. 41(2):586–591.
Hassler C, Nitzsche S, Iversen C, Zweifel C, Stephan R. 2008. Characteristics of *Staphylococcus hyicus* strains isolated from pig carcasses in two different slaughterhouses. Meat Sci. 80(2):505–510.
Hernández-González IL, Castillo-Ramírez S. 2020. Antibiotic-resistant *Acinetobacter baumannii* is a One Health problem. Lancet Microbe. 1(7):e279.
Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 35(2):518–522.
International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC). 2009. Classification of staphylococcal cassette chromosome mec (SCCmec): guidelines for reporting novel SCCmec elements. Antimicrob Agents Chemother. 53(12):4961–4967.
Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 9(1):5114.
Katoh K, Asimenos G, Toh H. 2009. Multiple alignment of DNA sequences with MAFFT. Methods Mol Biol. 537:39–64.
Kaya H, et al. 2018. SCCmecFinder, a web-based tool for typing of Staphylococcal Cassette Chromosome mec in *Staphylococcus aureus* using whole-genome sequence data. mSphere 3(1):e00612-00617. doi:10.1128/mSphere.00612-17.
Kern A, Perreten V. 2013. Clinical and molecular features of methicillin-resistant, coagulase-negative staphylococci of pets and horses. J Antimicrob Chemother. 68(6):1256–1266.
Knudsen GM, et al. 2013. ClpP deletion causes attenuation of *Salmonella typhimurium* virulence through mis-regulation of RpoS and indirect control of CsrA and the SPI genes. Microbiology (Reading) 159(Pt 7):1497–1509.
Konstantinidis KT, Tiedje JM. 2005. Towards a genome-based taxonomy for prokaryotes. J Bacteriol. 187(18):6258–6264.
Lamers RP, et al. 2012. Phylogenetic relationships among Staphylococcus species and refinement of cluster groups based on multilocus data. BMC Evol Biol. 12:171.
Lavecchia A, et al. 2021. Comparative genomics suggests a taxonomic revision of the Staphylococcus cohnii species complex. Genome Biol Evol. 13(4):evab020.
Lawal OU, et al. 2021. Foodborne origin and local and global spread of *Staphylococcus saprophyticus* causing human urinary tract infections. Emerg Infect Dis. 27(3):880–893.
Lebeurre J, et al. 2019. Comparative genome analysis of Staphylococcus lugdunensis shows clonal complex-dependent diversity of the putative virulence factor, ess/type VII locus. Front Microbiol. 10:2479.
Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 47(W1):W256–W259.
Lin M, Kussell E. 2019. Inferring bacterial recombination rates from large-scale sequencing datasets. Nat Methods. 16(2):199–204.
Liu B, Zheng D, Jin Q, Chen L, Yang J. 2019. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res. 47(D1):D687–D692.
Luong TT, Lee CY. 2002. Overproduction of type 8 capsular polysaccharide augments *Staphylococcus aureus* virulence. Infect Immun. 70(7):3389–3395.
Madhaiyan M, Wirth JS, Saravanan VS. 2020. Phylogenomic analyses of the Staphylococcaceae family suggest the reclassification of five species within the genus Staphylococcus as heterotypic synonyms, the promotion of five subspecies to novel species, the taxonomic reassignment of five Staphylococcus species to Mammaliicoccus gen. nov., and the formal assignment of Nosocomiicoccus to the family Staphylococcaceae. Int J Syst Evol Microbiol. 70(11):5926–5936.
Mama OM, et al. 2019. High diversity of coagulase negative staphylococci species in wild boars, with low antimicrobial resistance rates but detection of relevant resistance genes. Comp Immunol Microbiol Infect Dis. 64:125–129.
May L, Klein EY, Rothman RE, Laxminarayan R. 2014. Trends in antibiotic resistance in coagulase-negative staphylococci in the United States, 1999 to 2012. Antimicrob Agents Chemother. 58(3):1404–1409.
Méric G, et al. 2015. Ecological overlap and horizontal gene transfer in *Staphylococcus aureus* and *Staphylococcus epidermidis*. Genome Biol Evol. 7(5):1313–1328.
Méric G, et al. 2018. Disease-associated genotypes of the commensal skin bacterium *Staphylococcus epidermidis*. Nat Commun. 9(1):5034.
Michalik M, et al. 2020. Coagulase-negative staphylococci (CoNS) as a significant etiological factor of laryngological infections: a review. Ann Clin Microbiol Antimicrob. 19(1):26.
Michel A, et al. 2006. Global regulatory impact of ClpP protease of *Staphylococcus aureus* on regulons involved in virulence, oxidative stress response, autolysis, and DNA repair. J Bacteriol. 188(16):5783–5796.
Minh BQ, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 37(5):1530–1534.
Mostowy R, et al. 2017. Efficient Inference of recent and ancestral recombination within bacterial populations. Mol Biol Evol. 34(5):1167–1182.
Murray S, et al. 2017. Recombination-mediated host adaptation by avian *Staphylococcus aureus*. Genome Biol Evol. 9(4):830–842.
Nakamura Y, et al. 2013. Staphylococcus δ-toxin induces allergic skin disease by activating mast cells. Nature 503(7476):397–401.

Non LR, Santos CAQ. 2017. The occurrence of infective endocarditis with *Staphylococcus lugdunensis* bacteremia: a retrospective cohort study and systematic review. J Infect. 74(2):179–186.
O'Riordan K, Lee JC. 2004. Staphylococcus aureus capsular polysaccharides. Clin Microbiol Rev. 17(1):218–234.
Park CJ, Andam CP. 2020. Distinct but intertwined evolutionary histories of multiple *Salmonella enterica* subspecies. mSystems 5(1):e00515-00519. doi:10.1128/mSystems.00515-19.
Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25(7):1043–1055.
Parte AC. 2014. LPSN–list of prokaryotic names with standing in nomenclature. Nucleic Acids Res. 42(Database issue):D613–D616.
Pedroso SHSP, et al. 2018. Coagulase-negative staphylococci isolated from human bloodstream infections showed multidrug resistance profile. Microb Drug Resist. 24(5):635–647.
Qin Q-L, et al. 2014. A proposed genus boundary for the prokaryotes based on genomic insights. J Bacteriol. 196(12):2210–2215.
Rolo J, et al. 2017. Evolutionary origin of the staphylococcal cassette chromosome mec (SCCmec). Antimicrob Agents Chemother. 61(6):e02302-02316. doi:10.1128/AAC.02302-16.
Russell SL, Cavanaugh CM. 2017. Intrahost genetic diversity of bacterial symbionts exhibits evidence of mixed infections and recombinant haplotypes. Mol Biol Evol. 34(11):2747–2761.
Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30(14):2068–2069.
Silva C, Calva E, Maloy S. 2014. One Health and food-borne disease: salmonella transmission between humans, animals, and plants. Microbiol Spectr. 2(1):OH-0020-2013.
Spoor LE, et al. 2015. Recombination-mediated remodelling of host-pathogen interactions during *Staphylococcus aureus* niche adaptation. Microb Genom. 1(4):e000036.
Stenmark B, Hellmark B, Söderquist B. 2019. Genomic analysis of Staphylococcus capitis isolated from blood cultures in neonates at a neonatal intensive care unit in Sweden. Eur J Clin Microbiol Infect Dis. 38(11):2069–2075.
Taponen S, et al. 2012. *Staphylococcus agnetis* sp. nov., a coagulase-variable species from bovine subclinical and mild clinical mastitis. Int J Syst Evol Microbiol. 62(Pt 1):61–65.
Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lectures Math Life Sci (Amer Math Soc). 17(2):57–86.
Thanner S, Drissner D, Walsh F. 2016. Antimicrobial resistance in agriculture. mBio 7(2):e02227-15.
Tonkin-Hill G, et al. 2020. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol. 21(1):180.
Vandenesch F, et al. 1994. Clotting activity in *Staphylococcus schleiferi* subspecies from human patients. J Clin Microbiol. 32(2):388–392.
von Eiff C, Arciola CR, Montanaro L, Becker K, Campoccia D. 2006. Emerging Staphylococcus species as new pathogens in implant infections. Int J Artif Organs. 29(4):360–367.
Widerström M, Wiström J, Ferry S, Karlsson C, Monsen T. 2007. Molecular epidemiology of *Staphylococcus saprophyticus* isolated from women with uncomplicated community-acquired urinary tract infection. J Clin Microbiol. 45(5):1561–1564.
Wirth T, et al.; ESGS Study Group of ESCMID. 2020. Niche specialization and spread of Staphylococcus capitis involved in neonatal sepsis. Nat Microbiol. 5(5):735–745.
Zeng Q, et al. 2017. Recombination of virulence genes in divergent *Acidovorax avenae* strains that infect a common host. Mol Plant Microbe Interact. 30(10):813–828.
Zhao B-B, Li X-H, Zeng Y-L, Lu Y-J. 2016. ClpP-deletion impairs the virulence of *Legionella pneumophila* and the optimal translocation of effector proteins. BMC Microbiol. 16(1):174.
Zhaxybayeva O, Doolittle WF, Papke RT, Gogarten JP. 2009. Intertwined evolutionary histories of marine Synechococcus and *Prochlorococcus marinus*. Genome Biol Evol. 1:325–339.
Zheng J, et al. 2020. ClpP participates in stress tolerance, biofilm formation, antimicrobial tolerance, and virulence of *Enterococcus faecalis*. BMC Microbiol. 20(1):30.
Zheng Y, He L, Asiamah TK, Otto M. 2018. Colonization of medical devices by staphylococci. Environ Microbiol. 20(9):3141–3153.
Zingg W, Demartines N, Imhof A, Senn G, Ruef C. 2009. Rapid colonization with methicillin-resistant coagulase-negative staphylococci after surgery. World J Surg. 33(10):2058–2062.

Associate editor: Tal Dagan