Check for updates

Comprehensive Phylogenetic Analysis of Bovine Non-*aureus* Staphylococci Species Based on Whole-Genome Sequencing

Sohail Naushad[1,2], Herman W. Barkema[1,2], Christopher Luby[2,3], Larissa A. Z. Condas[1,2], Diego B. Nobrega[1,2], Domonique A. Carson[1,2] and Jeroen De Buck[1,2]*

[1] Department of Production Animal Health, Faculty of Veterinary Medicine, University of Calgary, Calgary, AB, Canada,
[2] Canadian Bovine Mastitis and Milk Quality Research Network, St-Hyacinthe, QC, Canada, [3] Department of Large Animal Clinical Sciences, Western College of Veterinary Medicine, University of Saskatchewan, Saskatoon, SK, Canada

Non-*aureus* staphylococci (NAS), a heterogeneous group of a large number of species and subspecies, are the most frequently isolated pathogens from intramammary infections in dairy cattle. Phylogenetic relationships among bovine NAS species are controversial and have mostly been determined based on single-gene trees. Herein, we analyzed phylogeny of bovine NAS species using whole-genome sequencing (WGS) of 441 distinct isolates. In addition, evolutionary relationships among bovine NAS were estimated from multilocus data of 16S rRNA, *hsp60*, *rpoB*, *sodA*, and *tuf* genes and sequences from these and numerous other single genes/proteins. All phylogenies were created with FastTree, Maximum-Likelihood, Maximum-Parsimony, and Neighbor-Joining methods. Regardless of methodology, WGS-trees clearly separated bovine NAS species into five monophyletic coherent clades. Furthermore, there were consistent interspecies relationships within clades in all WGS phylogenetic reconstructions. Except for the Maximum-Parsimony tree, multilocus data analysis similarly produced five clades. There were large variations in determining clades and interspecies relationships in single gene/protein trees, under different methods of tree constructions, highlighting limitations of using single genes for determining bovine NAS phylogeny. However, based on WGS data, we established a robust phylogeny of bovine NAS species, unaffected by method or model of evolutionary reconstructions. Therefore, it is now possible to determine associations between phylogeny and many biological traits, such as virulence, antimicrobial resistance, environmental niche, geographical distribution, and host specificity.

Keywords: Non-*aureus* staphylococci, coagulase-negative staphylococci, bovine intramammary infection, phylogenetic trees, whole-genome sequencing

INTRODUCTION

The genus *Staphylococcus* currently consists of 52 species and 28 subspecies. Phylogeny and classification of this genus is under active investigation and has been subject to substantial revisions (Kloos et al., 1998a,b; Svec et al., 2004; Taponen et al., 2012). The non-*aureus* staphylococci (NAS), previously referred to as coagulase-negative staphylococci [although some species have

a variable response to the coagulase test (Dos Santos et al., 2016)], are the most frequently isolated microorganisms from the mammary gland of dairy cows and increasingly recognized as etiologic agents of intramammary infection (IMI) in cattle worldwide (Pyörälä and Taponen, 2009; Sampimon et al., 2009; Thorberg et al., 2009; De Vliegher et al., 2012). Although the NAS are primarily associated with subclinical or mild clinical mastitis (Honkanen-Buzalski et al., 1994; Supré et al., 2011), IMI with NAS decrease quality and quantity of milk produced and can moderately increase somatic cell count (Schukken et al., 2009; Taponen and Pyörälä, 2009; De Vliegher et al., 2012; Condas et al., in review). Some species of NAS are capable of persisting in the udder for months or even throughout the lactation (Aarestrup et al., 1995; Laevens et al., 1997; Taponen et al., 2007; Thorberg et al., 2009). Additionally, NAS contain important virulence factors (Zhang et al., 2003; Otto, 2013; Vanderhaeghen et al., 2014), have a high level of antimicrobial resistance (Rajala-Schultz et al., 2009; Frey et al., 2013; Taponen et al., 2015), and can cause chronic IMI (Gillespie et al., 2009; De Vliegher et al., 2012). In contrast, NAS have also been reported to have a protective role against major IMI-related pathogens (Matthews et al., 1990; De Vliegher et al., 2004). Contradictory findings among studies regarding pathogenicity of NAS is potentially related to classifying disparate species and strains of staphylococci as one group (Woodward et al., 1987, 1988; Matthews et al., 1990; Piepers et al., 2009; Vanderhaeghen et al., 2014). However, NAS are a heterogeneous group of numerous species and it is, therefore, expected that individual NAS species will have different effects on udder health and production (Piepers et al., 2009; Vanderhaeghen et al., 2014). To determine effects of individual or closely related species on udder health and to investigate whether these variable effects align with their phylogenetic relationship, a complete understanding of species relatedness in the NAS group is essential.

Early approaches to understanding staphylococcal relationships and phylogenies were primarily based on sequence analysis of the 16S rRNA gene and other genes, such as *rpoB* (β-subunit of RNA polymerase), *tuf* (elongation factor Tu), cpn60 (heat shock protein 60), and *dnaJ* (heat shock protein 40) (Takahashi et al., 1999; Kwok and Chow, 2003; Shah et al., 2007; Ghebremedhin et al., 2008; Lamers et al., 2012). However, phylogenies based on these single genes displayed contradictory topologies when compared to each other (Ghebremedhin et al., 2008; Lamers et al., 2012). In contrast, utilization of phylogenies based on larger datasets of sequences from multiple genes provides significantly greater resolution of phylogenetic relationships between organisms (Rokas et al., 2003; Wu et al., 2009; Naushad et al., 2015a,b). For example, a recent multi-gene phylogeny of *Staphylococcus* spp., based on the 16S rRNA, *dnaJ*, *rpoB*, and *tuf* genes (Lamers et al., 2012), provided clear and robust support for many interspecies relationships among recently diverged NAS species. Notwithstanding, for more ancient lineages, there were several conflicting phylogenies among various gene trees.

In this study, we completed whole genome sequencing (WGS) of 441 bovine NAS isolates and utilized WGS data to construct robust and well-resolved phylogenetic trees based on all genes in the core genome of the NAS, thereby providing a highly reliable foundation for understanding interspecies relatedness among the NAS. We also produced several housekeeping (ribosomal and information transfer) gene-based phylogenetic trees to investigate the reliability of specific single gene trees and to determine the gene or group of genes which most closely approximate the phylogeny produced by the core genome. Lastly, we investigated effects of various phylogenetic reconstruction methods on NAS phylogenies in order to determine effects of various parameters and models of evolution on the final topology of NAS phylogenies.

MATERIALS AND METHODS

Selection of Isolates

A total of 441 NAS isolates (**Table 1**) were selected for WGS from 5507 isolates obtained from bovine staphylococcal IMI stored at the Mastitis Pathogen Collection of the Canadian Bovine Mastitis and Milk Quality Research Network (CBMQRN). Selection was randomized over the 25 NAS species previously identified (Mellmann et al., 2006; Condas et al., in review) which originated from 87 herds distributed over four geographic regions encompassing Atlantic Canada (Nova Scotia, Prince Edward Island, New Brunswick), Central Canada (Québec and Ontario), and Western Canada (Alberta) (Reyher et al., 2011). Inclusion criteria of isolates were: (1) inclusion of 68 NAS isolated from clinical mastitis samples; (2) inclusion of 26 NAS isolates with multi-drug resistance (MDR); (3) for uncommon species (defined as <20 unique isolations at cow level), inclusion of one arbitrarily selected isolate of that species per cow; (4) one isolate per cow for all other species, until achieving the final number. Our criteria allowed two isolates of the same NAS species isolated from the same cow to be included (i.e., *S. chromogenes* isolated from clinical mastitis and subclinical mastitis with MDR in the same cow), although the majority of isolates originated from different cows (**Table 1**).

DNA Extraction and Whole Genome Sequencing

All isolates obtained were grown on 5% defibrinated sheep blood agar plates (BD Diagnostics, Mississauga, ON, Canada) with subsequent incubation at 35°C for 24 h to yield single colonies. For each of these isolates, colonies were picked from blood agar plates and suspended in Bacto Brain Heart Infusion broth (BD Diagnostics), and incubated at 37°C overnight. Genomic DNA for all isolates was extracted with DNeasy Blood and Tissue Kit (Qiagen, Toronto, ON, Canada), according to the corresponding protocol for Gram-positive bacteria. Quality and quantity of DNA was checked using a NanoVue plus spectrophotometer (GE Healthcare Life Sciences, Mississauga, ON, Canada) and the Qubit 2.0 fluorometer (Invitrogen, Burlington, ON, Canada). Each DNA sample was diluted to a final concentration of 0.2 ng/μl. Sequencing of these samples was performed using the Illumina MiSeq platform (Illumina, San Diego, CA, USA); DNA libraries for sequencing were prepared using a Nextera XT DNA library preparation kit (Illumina, San Diego, CA, US). All sequencing steps, including cluster generation, paired-end

TABLE 1 | Distribution of NAS species that were selected for WGS, obtained from Canadian Bovine Mastitis and Milk Quality Research Network.

Species name	Total isolates sequenced	Alberta			Ontario			Québec			Atlantic Canada*		
		N	Herds	Cows	N	Herds	Cows	N	Herds	Cows	N	Herds	Cows
S. agnetis	13	3	1	2	5	4	4	3	3	3	2	2	2
S. arlettae	15	3	2	3	9	9	9	3	3	3	0	0	0
S. auricularis	2	1	1	1	1	1	1	0	0	0	0	0	0
S. capitis	22	0	0	0	10	8	10	6	5	6	6	6	6
S. caprae	1	0	0	0	1	1	1	0	0	0	0	0	0
S. chromogenes	83	30	12	27	29	17	26	6	5	6	18	11	17
S. cohnii	24	2	2	2	12	9	12	8	6	7	2	2	2
S. devriesei	8	3	2	3	3	3	3	0	0	0	2	2	2
S. epidermidis	26	5	2	5	8	8	8	7	6	7	6	5	6
S. equorum	17	8	6	8	5	5	5	3	2	3	1	1	1
S. fleurettii	2	2	1	2	0	0	0	0	0	0	0	0	0
S. gallinarum	21	3	3	3	8	7	7	7	5	5	3	3	3
S. haemolyticus	29	9	5	9	8	8	8	5	5	5	7	6	7
S. hominis	11	0	0	0	2	2	2	7	5	7	2	2	2
S. hyicus	3	0	0	0	0	0	0	2	2	2	1	1	1
S. kloosii	1	0	0	0	1	1	1	0	0	0	0	0	0
S. nepalensis	2	0	0	0	2	2	2	0	0	0	0	0	0
S. pasteuri	6	0	0	0	0	0	0	6	4	6	0	0	0
S. saprophyticus	16	0	0	0	5	5	5	8	6	8	3	3	3
S. sciuri	29	8	6	8	9	9	9	8	7	7	4	3	4
S. simulans	42	6	5	6	19	13	18	13	10	13	4	4	4
S. succinus	15	3	3	3	5	4	5	1	1	1	6	5	6
S. vitulinus	6	2	2	2	1	1	1	0	0	0	3	3	3
S. warneri	19	2	2	2	8	6	7	5	2	5	4	4	4
S. xylosus	28	8	5	8	8	5	8	11	9	11	1	1	1
Total	441	98	60	94	159	128	152	109	86	105	75	64	74

Nova Scotia, Prince Edward Island, and New Brunswick.

sequencing (2 × 250 bp), and primary data analysis for quality control, were performed on the instrument.

Genome Assembly and Annotation

Sequence reads obtained from the MiSeq platform were checked for poorly sequenced regions and Illumina adapters sequences were trimmed using cutadapt (Martin, 2011), implemented in Trim Galore! 0.4.0 (with default parameters). After trimming sequences, filtered reads were assembled into contigs using the *de novo* assembly program SPAdes version 3.6.0 (Nurk et al., 2013), employing built-in error correction and default parameters. To determine average coverage of sequencing, reads were mapped back to the assembled genome using BWA 0.7.12-r1039 (Li and Durbin, 2010) and depth of sequencing for each contig was plotted using BEDTools (Quinlan and Hall, 2010). Genome annotations and gene predictions for contigs larger than 200 bp were performed with Prokka 1.11 (Seemann, 2014), using the provided *Staphylococcus* database. Quality of the assembled genomes and assembly metrics was determined using Quast (Gurevich et al., 2013). The entire genome assembly process was automated using the Snakemake workflow engine (Köster and Rahmann, 2012). The data was submitted to NCBI under BioProject ID PRJNA342349.

Single Protein Based Phylogenetic Analysis

Phylogenetic analyses were performed on 24 highly conserved housekeeping proteins (Supplementary Table 1). Sequences for each of these 24 proteins for all NAS species were retrieved from the genomes using BLAST+2.2.31 (Camacho et al., 2009). The sequence for *Macrococcus caseolyticus*, an outgroup species, was downloaded from NCBIs GenBank database. Multiple sequence alignments for these proteins were created using MUSCLE v3.8.31 (Edgar, 2004). The resulting alignments were used for phylogenetic analysis. Phylogenetic trees, based on 100 bootstrap replicates, were constructed by employing Maximum-Likelihood (ML), Maximum-Parsimony (MP), and Neighbor-Joining (NJ) methods using MEGA 6.0 (Tamura et al., 2013). Evolutionary distances for ML and NJ methods were computed using a JTT matrix-based model (Jones et al., 1992). Maximum-Parsimony trees were obtained using the Subtree-Pruning-Regrafting (SPR) algorithm (Nei and Kumar, 2000).

Single Gene/s and Multilocus Sequence Analysis

Phylogenetic trees were constructed based on full-length sequences of 16S rRNA, *hsp60*, *rpoB*, *sodA*, and *tuf* genes. Full-length sequences of these genes for NAS species were obtained

using BLAST+2.2.31 (Camacho et al., 2009). Multiple sequence alignments for each of gene were created using MUSCLE v3.8.31 (Edgar, 2004). Maximum-Likelihood, MP and NJ trees based on these sequence alignments were created using 100 bootstrap replicates in MEGA 6.0 (Tamura et al., 2013).

Multilocus sequence analysis was performed on nucleotide sequences of the 16S rRNA, *hsp60*, *rpoB*, *sodA*, and *tuf* genes. Individual gene alignments were manually concatenated to create a combined dataset of these five genes. Poorly aligned regions from this concatenated alignment were removed using Gblocks 0.92 (Castresana, 2000) with default settings except the allowed gap position parameter which was changed to 0.5 (50%). Maximum-Likelihood, MP and NJ trees based on 100 bootstrap replicates of this dataset were constructed using MEGA 6.0 (Tamura et al., 2013). For all trees, evolutionary distances for ML, MP, and NJ methods were computed using the General Time Reversible model (Nei and Kumar, 2000), Subtree-Pruning-Regrafting (Nei and Kumar, 2000), and Kimura 2-parameter model (Kimura, 1980), respectively. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from final alignments. All trees were rooted using *Macrococcus caseolyticus*, and gene sequences for this species were downloaded from NCBI GenBank database.

NAS Phylogenetic Tree—PhyloPhlAn

A genome-based phylogenetic tree of 441 NAS isolates was constructed using the published pipeline PhyloPhlAn (Segata et al., 2013). Briefly, the PhyloPhlAn approach is based on the use of 400 ubiquitous and phylogenetically informative proteins. Orthologs of these proteins in NAS genomes were detected using USEARCH v5.2.32 (Edgar, 2010). Multiple sequence alignments of these proteins were generated using MUSCLE v3.8.31 (Edgar, 2004). A final concatenated dataset containing 4231 aligned amino acid positions was generated. Final tree construction was performed using FastTree version 2.1 (Price et al., 2010). To determine effects of various evolutionary models on the resulting tree, Maximum-Likelihood, MP, and NJ trees based on 100 bootstrap replicates of the final alignment were also constructed using MEGA 6.0 (Tamura et al., 2013). For ML and NJ trees, evolutionary distances were computed using a JTT matrix-based model (Jones et al., 1992), whereas the Subtree-Pruning-Regrafting algorithm (Nei and Kumar, 2000) was used for MP tree construction.

NAS Core Genome Phylogenetic Tree

A phylogenetic tree of all NAS isolates, rooted using *Macrococcus caseolyticus*, was created based on the core genome of the bovine NAS group. The core set of NAS proteins were identified using the UCLUST algorithm (Edgar, 2010). Protein families with at least 30% sequence identity and 50% sequence length in 441 NAS isolates were considered core. However, protein families present in <80% of the input genomes were excluded from further analysis. Proteins families which contained potential paralogous sequences (duplicated sequence in same genome) were also excluded from further analysis. Each protein family was individually aligned using MAFFT 7 (Katoh and Standley, 2013). Aligned amino acid positions which contained gaps in more than

50% of genomes, were excluded from further analysis. Remaining amino acid positions were concatenated to create a combined dataset consisting of 128,080 aligned amino acids. A maximum-likelihood tree based on this alignment was constructed using FastTree 2.1 (Price et al., 2010), using the Whelan and Goldman substitution model (Whelan and Goldman, 2001) and JTT matrix-based model (Jones et al., 1992).

Comparison of Phylogenetic Trees

Genome-based phylogenetic trees (PhyloPhlAn, ML, MP, NJ) and Core-Genome-Tree (CGT) were visually inspected and compared. Single gene and protein trees were compared with each other and with CGT. Topological differences among them were computed using Robinson-Foulds (RF) distance matrix (Robinson and Foulds, 1981; Makarenkov and Leclerc, 2000), implemented as webserver; T-REX (Tree and reticulogram REConstruction, www.trex.uqam.ca; Boc et al., 2012). To facilitate comparisons of our results, RF distance scores were normalized using the maximum possible distance between two trees, calculated according to 2(N-3) where N represents the total number of taxa on a tree (Kupczok et al., 2010; Bernard et al., 2016). We designated these as normalized-Robinson-Foulds (nRF) scores, with values ranging from 0 (0%) to 1 (100%), and nRF = 0 indicating that topologies of two trees under investigation are congruent. Consequently, higher nRF score indicate a low level of congruence (or high level of incongruence) between two tree topologies.

RESULTS

Whole-Genome Based Phylogenetic Analysis of NAS

Five phylogenetic trees were constructed based on WGS data of NAS isolates. The first of these trees was constructed using a core genome dataset of NAS species (**Figure 1**), whereas the second tree (Supplementary Figure 1) was constructed using PhyloPhlAn (Segata et al., 2013). The same dataset from aligned sequences, obtained from PhyloPhlAn was also used to construct the ML (Supplementary Figure 2), MP, (Supplementary Figure 3), and NJ trees (Supplementary Figure 4). Within these trees, species of the NAS group branched into five distinct monophyletic and statistically well-supported clades (100% nodal support) in all phylogenetic reconstructions. We referred to them as Clade A, Clade B, Clade C, Clade D, and Clade E, which contained 3, 3, 1, 8, and 10 NAS species, respectively (**Figure 1**). The phylogeny of NAS species was also constructed by excluding 26 multidrug resistant isolates, as they could have introduced selection bias. However, no differences in phylogeny of the NAS species were seen (data not shown).

In all phylogenetic trees, Clade A branched as first or as the most ancient divergence of NAS group and was composed of three NAS species: *S. sciuri*, *S. fleurettii*, and *S. vitulinus*. Within this clade, *S. sciuri* formed a distant cluster basal to *S. fleurettii* and *S. vitulinus,* which appeared as sister taxa to each other (**Figure 1** and Supplementary Figures 1–4). The second or Clade B of NAS group contained *S. chromogenes*, *S. hyicus*, and *S. agnetis*, with *S. hyicus* appearing as sister taxon to *S.*
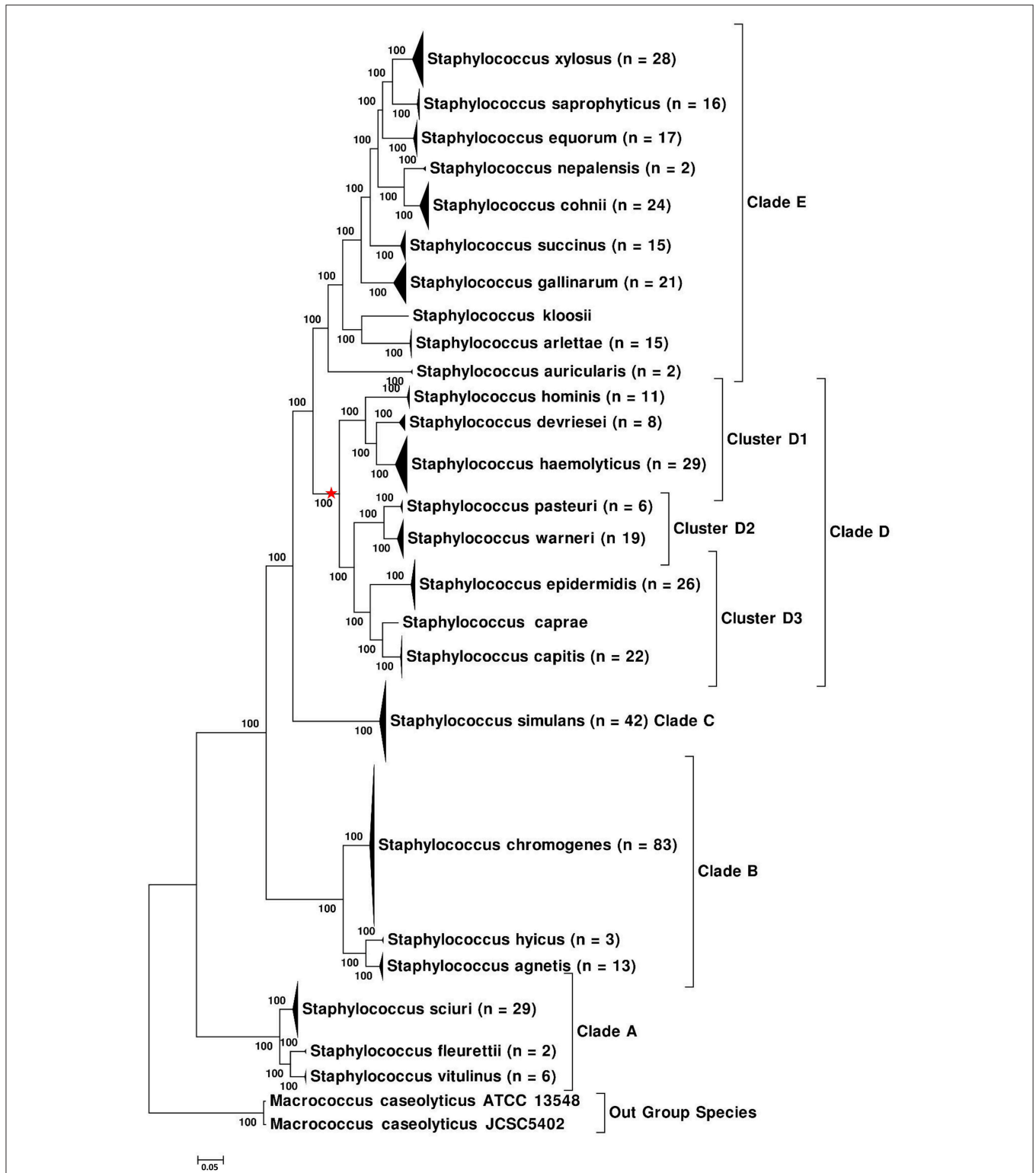
FIGURE 1 | Core-Genome-Tree of NAS, showing the branching of NAS species into five distinct clades. Multiple isolates of the same species were collapsed and total number of isolates for each species are shown after the species name. The evolutionary history of NAS was inferred using the Maximum Likelihood method based on the Whelan and Goldman substitution model (Whelan and Goldman, 2001) and JTT model (Jones et al., 1992), resulting in identical topology for both models. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in number of substitutions per site. Red star on the tree indicates phylogenetic placement of *S. aureus*.

agnetis and *S. chromogenes* forming the basal group (**Figure 1** and Supplementary Figures 1–3). However, there was a discrepancy in branching order among these species in the phylogenetic tree constructed using NJ method (Supplementary Figure 4), which placed *S. chromogenes* as sister taxon to *S. hyicus* and displayed *S. agnetis* branching separately. The next lineage of NAS to diverge was constituted by all isolates of a single species *S. simulans* (Clade C). All 42 sequenced isolates of this species diverged independently from other NAS members by a long branch, forming a strongly supported clade (**Figure 1** and Supplementary Figures 1–4).

Clade D contained eight species comprised of *S. hominis, S. devriesei, S. haemolyticus, S. pasteuri, S. warneri, S. epidermidis, S. caprae,* and *S. capitis* (**Figure 1** and Supplementary Figures 1–4). *Staphylococcus aureus,* a major pathogen of IMI, also shared common ancestor with species of Clade D. Phylogenetic placement of *S. aureus* is indicated on the tree (**Figure 1**). The eight NAS species of this clade were split into three clusters. The first cluster (D1) was composed of *S. hominis, S. devriesei,* and *S. haemolyticus.* Within this cluster, *S. hominis* appeared as a deeper branching species, whereas *S. devriesei* and *S. haemolyticus* branched together, representing divergence from a common ancestor (**Figure 1** and Supplementary Figures 1–4). The second cluster (D2) was formed by two NAS species, *S. pasteuri* and *S. warneri,* grouping together with 100% support in all genomic phylogenies (**Figure 1** and Supplementary Figures 1–4). Cluster D3 displayed a closer relationship of *S. caprae* to *S. capitis,* whereas *S. epidermidis* appeared as the basal lineage of this cluster. Monophyletic branching of these clusters within Clade D was supported by 100% bootstrap scores in all trees. The branching order and phylogenetic placement of these cluster groups was the same in all phylogenies (**Figure 1** and Supplementary Figures 1–4).

The fifth clade or Clade E was the most recently diverged clade of NAS and contained 10 species, comprised of *S. auricularis, S. kloosii, S. arlettae, S. gallinarum, S. succinus, S. nepalensis, S. cohnii, S. equorum, S. saprophyticus,* and *S. xylosus* (**Figure 1** and Supplementary Figures 1–4). Within this clade, *S. auricularis* was the deepest branching species, followed by a cluster group shared by *S. kloosii* and *S. arlettae.* Next, *S. gallinarum* and *S. succinus* diverged as two well-supported separate lineages (**Figure 1** and Supplementary Figures 1–4). Within the remaining species, there was a closer relationship of *S. nepalensis* to *S. cohnii* and *S. saprophyticus* to *S. xylosus* (**Figure 1** and Supplementary Figures 1–4). All isolates of *S. equorum* branched together as sister taxa to the *S. xylosus—S. saprophyticus* group.

Single Gene/s and Multilocus Sequence Analysis

Phylogenetic relationships among NAS species were also inferred using 16S rRNA, *hsp60, rpoB, sodA,* and *tuf* gene sequences separately and as a multilocus data set. The ML, MP, and NJ trees produced from these genes were compared with each other and with the CGT. The nRF distance-based comparisons among ML, NJ single trees and CGT (**Tables 2A,B**) and single genes MP trees and CGT (**Table 2C**) are presented in **Table 2**. Consistent

TABLE 2 | Comparisons of normalized-Robinson-Foulds (nRF) scores among CGT and single gene ML (A), single gene NJ (B), and single gene MP (C) trees.

	CGT	Multilocus	16S rRNA	hsp60	rpoB	sodA	tuf
(A)							
CGT	0*						
Multilocus	0.1	0					
16S rRNA	0.3	0.3	0				
hsp60	0.3	0.3	0.3	0			
rpoB	0.1	0.1	0.3	0.3	0		
sodA	0.3	0.3	0.4	0.5	0.3	0	
tuf	0.3	0.3	0.5	0.4	0.3	0.5	0
(B)							
CGT	0*						
Multilocus	0.1	0					
16S rRNA	0.4	0.4	0				
hsp60	0.3	0.3	0.4	0			
rpoB	0.1	0.1	0.4	0.3	0		
sodA	0.3	0.3	0.3	0.5	0.3	0	
tuf	0.3	0.3	0.5	0.4	0.3	0.5	0
(C)							
CGT	0*						
Multilocus	0.2	0					
16S rRNA	0.5	0.5	0				
hsp60	0.6	0.5	0.7	0			
rpoB	0.3	0.3	0.5	0.5	0		
sodA	0.3	0.3	0.6	0.6	0.4	0	
tuf	0.3	0.3	0.7	0.5	0.5	0.4	0

The colors of the cells represent heat map generated from the nRF scores. nRF value zero (dark green) indicates that the two trees are topologically congruent. Higher nRF scores (dark red) indicate a lower level of congruence between two tree topologies.

with the CGT, multilocus ML (**Figure 2**) and NJ (Supplementary Figure 5) trees displayed branching of NAS species into five distinct clades (Clade A, Clade B, Clade C, Clade D, and Clade E). The MP-multilocus tree (Supplementary Figure 6), however, had a different pattern of clade divergence, and displayed branching of *S. simulans* with *S. auricularis* within Clade E. Hence, Clade C was absent from MP-based phylogeny. Except for the multilocus MP tree, species composition within each clade was identical to the CGT. However, branching order within Clade D and Clade E differed among each other and with CGT. Within Clade D, in the multilocus MP tree, the cluster group containing *S. pasteuri* and *S. warneri* suggested a deeper branching cluster compared to other species of this clade (Supplementary Figure 6). The notable difference within clade E was branching of *S. succinus* and *S. equorum* as sister taxa in all multilocus trees (**Figure 2** and Supplementary Figures 5, 6).

Phylogenetic trees (ML, MP, and NJ) based upon 16S rRNA gene (**Figure 3** and Supplementary Figures 7, 8), *hsp60* (Supplementary Figures 9–11), *rpoB* (Supplementary Figures 12–14), *sodA* (Supplementary Figures 15–17), and *tuf* (Supplementary Figures 18–20) gene sequences were constructed. The nRF distance-based analyses for ML trees (**Table 2A**), NJ trees (**Table 2B**), and MP trees (**Table 2C**)
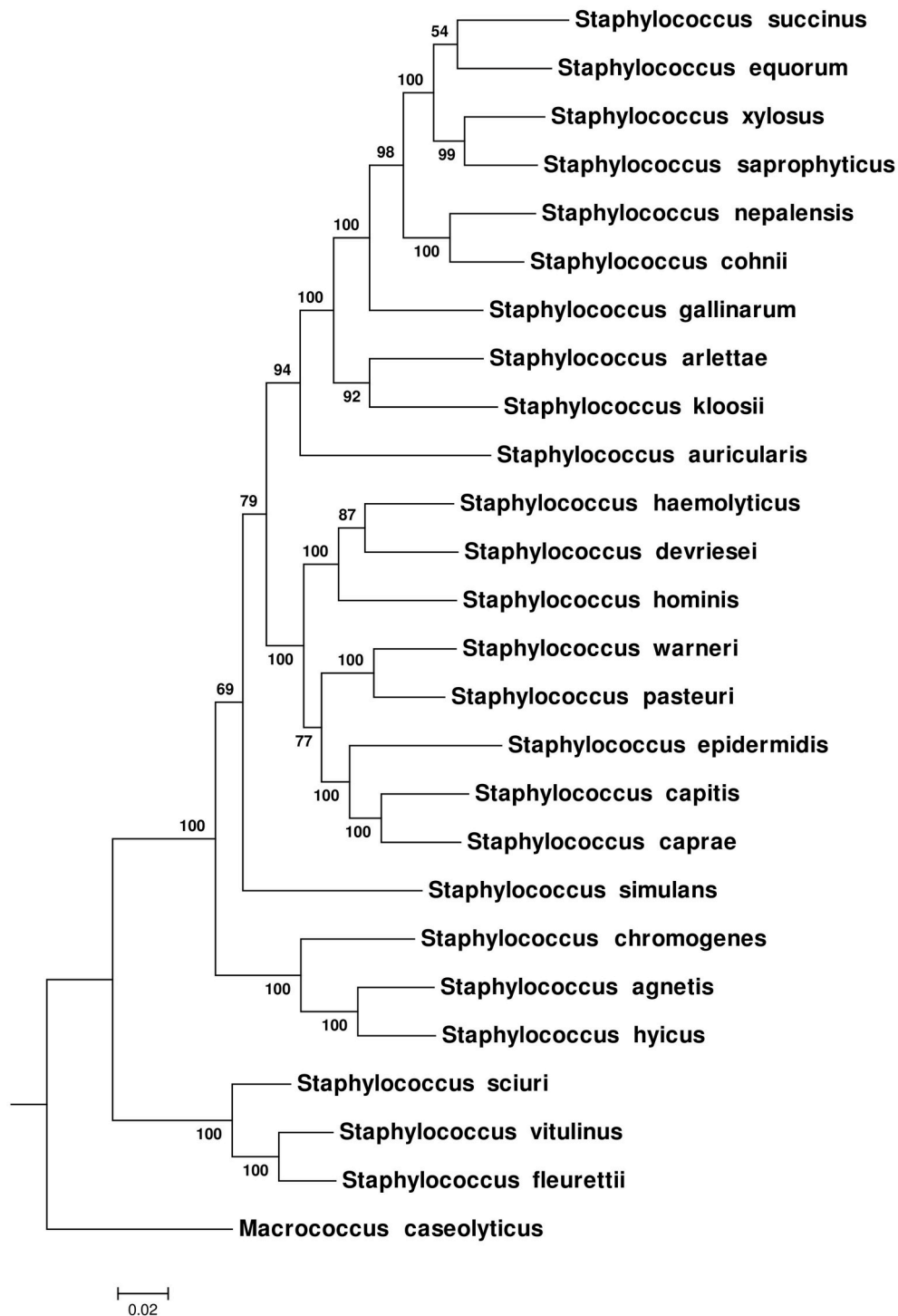
FIGURE 2 | Maximum-Likelihood phylogenetic tree obtained from multilocus data set of 16S rRNA, *hsp60, rpoB, sodA*, and *tuf* genes. The tree was constructed using a General Time Reversible model (Nei and Kumar, 2000). Bootstrap scores for each node are presented next to the branches. The tree was drawn to scale, with branch lengths measured in number of substitutions per site.

based upon these five genes, indicated that these trees differed from each other and from the CGT (**Table 2**). Within the 16S rRNA gene based phylogenies, the only reliably identified

relationship was the composition and branching order of *S. sciuri, S. fleurettii*, and *S. vitulinus* within Clade A (**Figure 3** and Supplementary Figures 7, 8). All other species had varying

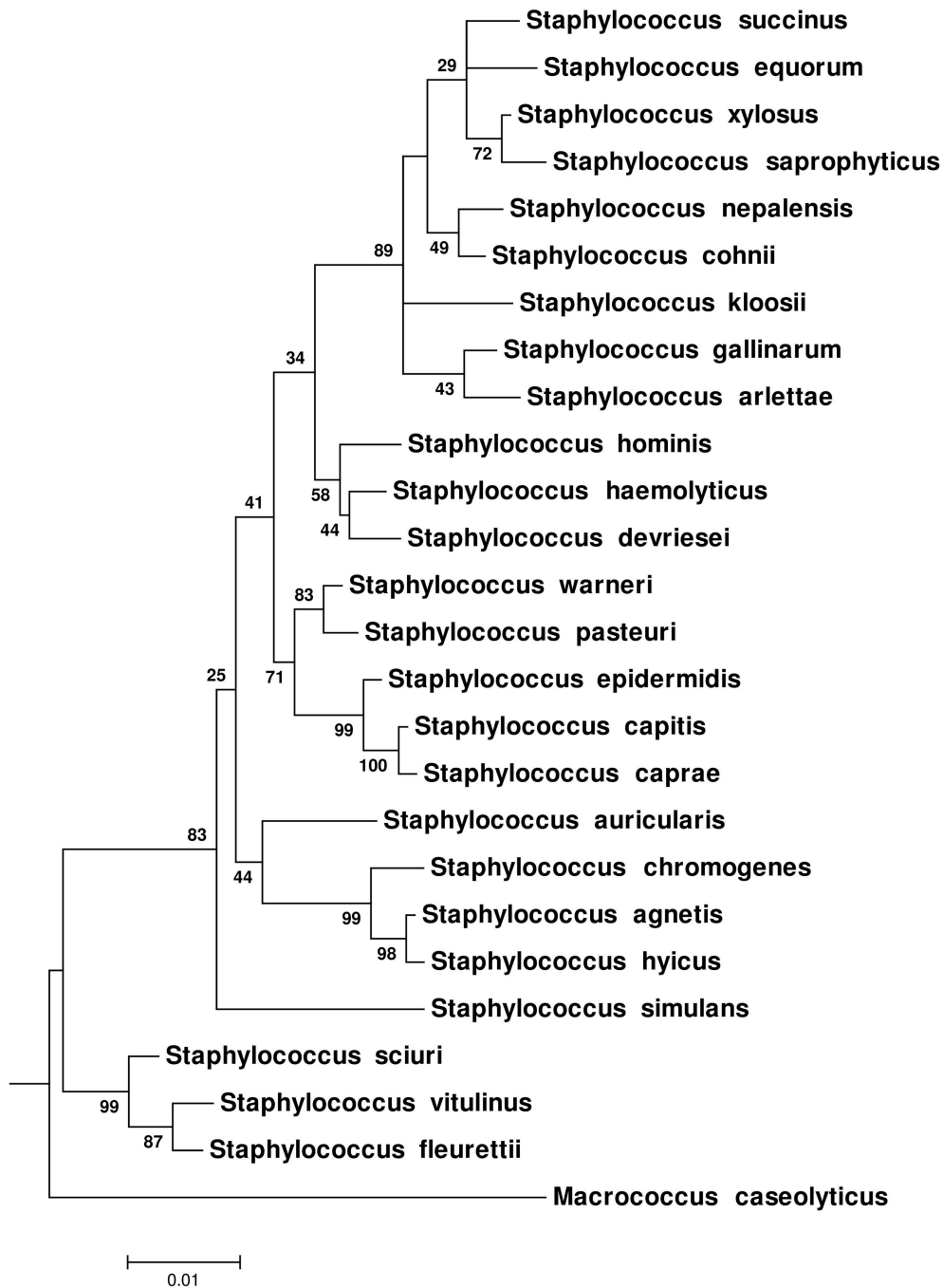**FIGURE 3 | Maximum-Likelihood phylogenetic tree based on 16S rRNA gene sequences of NAS species.** The tree was constructed using General Time Reversible model (Nei and Kumar, 2000). The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. The values on the nodes represent bootstrap support for each relationship.

patterns of evolutionary placement. Although *S. simulans* branched distinctly in ML and NJ trees, the order of its divergence was different compared to the CGT. In these cases, *S. simulans* branched after divergence of Clade A from the rest of the NAS species (**Figure 3** and Supplementary Figure 8), as opposed to branching after separation of Clades A and B in the CGT

(**Figure 1** and Supplementary Figures 1–4). The branching of S. *simulans* after Clade A had 83 and 80% statistical support in ML (**Figure 3**) and NJ (Supplementary Figure 8) trees, respectively. In the 16S-rRNA-MP tree, *S. simulans* clustered with other NAS species (Supplementary Figure 7). Within Clade B, *S. auricularis* branched with *S. chromogenes*, *S. agnetis*, and *S. hyicus* (**Figure 3**).

However, Clade D was not recovered as a monophyletic lineage. The cluster group containing *S. hominis*, *S. haemolyticus*, and *S. devriesei* appeared as separate taxa distinct from the rest of Clade D species (**Figure 3**). Except for *S. auricularis*, which clustered with clade B, species composition was invariable within Clade E. However, there was discordance among branching of these species. Additionally, phylogenetic positioning of *S. kloosii* and *S. equorum* was not resolved properly within Clade E (**Figure 3**).

Phylogenies constructed using *hsp60* gene sequence differed significantly from each other, both in species composition and their hierarchy of branching pattern. For three *hsp60* trees (ML, MP, and NJ), Clade A branches as monophyletic group with 100% bootstrap scores (Supplementary Figures 9–11). However, species of other clades were intermixed. The ML and NJ phylogenetic trees constructed using *rpoB* gene sequence revealed the same major clades. However, in both trees, *S. simulans* branched with *S. chromogenes*, *S. agnetis*, and *S. hyicus*, species of Clade B. Within Clade D, in both the ML and NJ trees, *S. pasteuri* and *S. warneri* formed a basal branching group compared to other Clade D species. Within Clade E, there were similar branching patterns among ML and NJ trees; the only difference was branching of *S. succinus* after the *S. cohnii—S. nepalensis* pair, as opposed to its divergence after *S. gallinarum* in CGT. Except for Clade A, the MP tree (Supplementary Figure 13) based on the *rpoB* gene, had a largely different pattern of evolutionary relationships among NAS species. Phylogenetic trees based on the *sodA* gene only revealed formation of Clades A and B, similar to CGT (Supplementary Figures 15–17), whereas remaining species of other clades (Clade C, D, and E) had a polyphyletic grouping pattern. For ML and NJ method-based phylogenies obtained from *tuf* gene, NAS species branched into the same five distinct clades as the CGT (Supplementary Figures 18, 20). However, in the MP (*tuf*) tree, *S. simulans* grouped with species of Clade D (Supplementary Figure 19). Hierarchy and clustering patterns among members of Clade D and Clade E differed among various phylogenies, based on the *tuf* gene.

Single Protein/s (Translated Genes) Based Phylogeny of NAS

Evolutionary relationships of NAS species were also investigated, based on 24 housekeeping protein sequences (Supplementary Table 1). Phylogenies obtained from each protein sequence were examined by various methods of evolutionary estimations (ML, MP, and NJ). Three trees (ML, MP, and NJ) produced from each single protein, were compared with each other and with CGT. The nRF distances between each of ML, NJ, and MP single protein trees and CGT are presented (**Tables 3–5**, respectively). Differences and similarities in species composition and branching order among species of various clades among trees are presented (**Table 6**).

Species composition and branching order for Clade A were retrieved in most phylogenetic reconstructions (**Table 3**) except MP trees [Elongation factor Tu (EF-Tu) and ribosomal protein S11 (RibS11)] and NJ tree [ribosomal protein S11 (RibS11)] which could not recover Clade A. Among protein trees that depicted branching of Clade A as monophyletic group, there

were branching order differences among species of Clade A in MP trees, based upon cytidine diphosphate (CDP) diglyceride synthetase (CdsA), ribosomal protein S2 (RibS2) and RNA polymerase subunits α and β protein sequences (RpoA and RpoB). Discordance in branching of Clade A species was also apparent in NJ trees constructed from RecA, ribosomal protein S2 (RibS2) and ML tree reconstructed from CDP diglyceride synthetase (CdsA), elongation factor Tu (EF-Tu), cell division protein FtsY, RecA, ribosomal protein S2 (RibS2), and RNA polymerase α subunit (RpoA) protein sequences.

Similar to Clade A, most phylogenetic estimation methods produced Clade B as an independent and distinct monophyletic group, except in MP trees built from cell division protein FtsY, gyrase B (GyrB), ribosomal protein L2 (RibL2), RibS11, and RNA polymerase β′ (RpoC) subunit protein sequences. Within Clade B, dissimilarities in branching among its species were seen in MP [serine hydroxymethyl transferase (GlyA) and gyrase A (GyrA)], NJ [ribosomal protein L15 (RibL15)] and ML trees [cell division protein FtsY, serine hydroxymethyl transferase (GlyA), gyrase B (GyrB), and HSP60]. Clade C was considered accurate only when it branched after Clades A and B in phylogenetic trees. Several protein sequences, namely, CDP diglyceride synthetase (CdsA), gyrase A (GryA), HSP70 and ribosomal protein S8 (RibS8), produced precise branching of Clade C in ML, MP, and NJ trees. Accurate branching of Clade C was also recovered in MP trees constructed from ribosomal protein L1 (RibL1) and RibS2, preprotein translocase SecA, SecY, and NJ trees, based on cell division protein FtsY, ribosomal protein L2, L15, and S2 (RibL2, RibL15, and RibS2) and PcrA helicase. ML trees constructed using sequences from cell division protein FtsY and preprotein translocase SecY also produced correct branching of Clade C (identical to CGT).

Species composition for Clade D, matching to CGT, was obtained in ML, MP, and NJ phylogenies constructed using sequences from elongation factor G (EF-G), cell division protein FtsY, Gyrase A (GyrA), RecA, RNA polymerase subunit β (RpoB), preprotein translocase SecY and PcrA helicase. However, identical branching patterns among species of Clade D were only observed in MP (GyrA) and NJ trees constructed from serine hydroxymethyl transferase (GlyA), preprotein translocase SecA and PcrA helicase protein sequences. Accurate species composition within Clade E was recovered in all phylogenetic estimations (ML, MP, and NJ) obtained from CDP diglyceride synthetase (CdsA), cell division protein FtsY, serine hydroxymethyl transferase (GlyA), gyrase A (GyrA), HSP60, RNA polymerase subunit β (RpoB), and preprotein translocase SecY. However, the matching branching order among the species of Clade E to CGT was only present in one MP tree [Serine hydroxymethyl transferase (GlyA)] and three NJ trees [HSP60, RNA polymerase subunit β (RpoB), and preprotein translocase SecA].

DISCUSSION

The role of NAS species in bovine udder health is controversial. Several studies concluded NAS are the principal cause of IMI (Piepers et al., 2007; Sampimon et al., 2009). In contrast,

TABLE 3 | Normalized-Robinson-Foulds distance score comparisons of CGT and single proteins ML trees.

	CGT	CdsA	EF-G	EF-P	EF-Tu	FtsY	GlyA	GyrA	GyrB	HSP60	HSP70	KsgA	RecA	RibL1	RibL2	RibL15	RibS2	RibS8	RibS11	RpoA	RpoB	RpoC	SecA	SecY	PcrA
CGT	0*																								
CdsA	0.5	0																							
EF-G	0.4	0.6	0																						
EF-P	0.3	0.6	0.4	0																					
EF-Tu	0.6	0.7	0.7	0.6	0																				
FtsY	0.3	0.5	0.6	0.5	0.7	0																			
GlyA	0.2	0.4	0.4	0.5	0.6	0.5	0																		
GyrA	0.3	0.4	0.4	0.4	0.7	0.4	0.3	0																	
GyrB	0.4	0.7	0.5	0.5	0.6	0.6	0.5	0.5	0																
HSP60	0.4	0.5	0.6	0.4	0.7	0.3	0.3	0.4	0.5	0															
HSP70	0.4	0.6	0.6	0.6	0.7	0.5	0.5	0.5	0.5	0.5	0														
KsgA	0.4	0.6	0.5	0.5	0.7	0.5	0.4	0.5	0.6	0.6	0.5	0													
RecA	0.4	0.7	0.4	0.4	0.7	0.4	0.5	0.5	0.5	0.6	0.7	0.5	0												
RibL1	0.3	0.6	0.5	0.5	0.5	0.5	0.4	0.5	0.4	0.5	0.5	0.5	0.5	0											
RibL2	0.3	0.5	0.5	0.5	0.6	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.6	0.5	0										
RibL15	0.5	0.6	0.6	0.5	0.7	0.5	0.5	0.4	0.4	0.5	0.5	0.5	0.6	0.6	0.6	0									
RibS2	0.4	0.6	0.5	0.5	0.7	0.5	0.5	0.5	0.6	0.7	0.6	0.4	0.4	0.6	0.6	0.6	0								
RibS8	0.2	0.5	0.5	0.5	0.5	0.4	0.3	0.3	0.4	0.5	0.3	0.3	0.5	0.4	0.3	0.4	0.4	0							
RibS11	0.7	0.7	0.7	0.7	0.7	0.8	0.7	0.7	0.7	0.8	0.7	0.7	0.8	0.7	0.6	0.6	0.7	0.5	0						
RpoA	0.3	0.5	0.4	0.5	0.5	0.5	0.4	0.5	0.4	0.6	0.4	0.4	0.4	0.5	0.5	0.4	0.3	0.2	0.6	0					
RpoB	0.3	0.6	0.6	0.5	0.7	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.3	0.4	0.5	0.6	0.4	0.7	0.5	0				
RpoC	0.3	0.6	0.5	0.6	0.6	0.6	0.3	0.5	0.5	0.6	0.5	0.5	0.6	0.4	0.4	0.6	0.5	0.4	0.7	0.4	0.4	0			
SecA	0.1	0.5	0.4	0.3	0.7	0.3	0.3	0.3	0.5	0.3	0.5	0.4	0.3	0.3	0.4	0.4	0.5	0.3	0.7	0.4	0.3	0.4	0		
SecY	0.2	0.4	0.4	0.4	0.7	0.3	0.3	0.2	0.5	0.4	0.5	0.4	0.4	0.4	0.3	0.4	0.5	0.3	0.7	0.4	0.3	0.4	0.1	0	
PcrA	0.3	0.6	0.3	0.4	0.7	0.5	0.4	0.4	0.5	0.5	0.6	0.4	0.3	0.5	0.5	0.5	0.4	0.4	0.7	0.3	0.5	0.5	0.3	0.4	0

*The colors of the cells represent heat map generated from the nRF scores. nRF value zero (dark green) indicates that the two trees are topologically congruent. Higher nRF scores (dark red) indicate a lower level of congruence between two tree topologies.

TABLE 4 | Normalized-Robinson-Foulds distance score comparisons of CGT with single protein NJ trees.

	CGT	CdsA	EF-G	EF-P	EF-Tu	FtsY	GlyA	GyrA	GyrB	HSP60	HSP70	KsgA	RecA	RibL1	RibL2	RibL15	RibS2	RibS8	RibS11	RpoA	RpoB	RpoC	SecA	SecY	PcrA
CGT	0*																								
CdsA	0.2	0																							
EF-G	0.4	0.5	0																						
EF-P	0.3	0.4	0.5	0																					
EF-Tu	0.6	0.7	0.6	0.7	0																				
FtsY	0.2	0.3	0.5	0.3	0.7	0																			
GlyA	0.2	0.2	0.5	0.4	0.5	0.3	0																		
GyrA	0.2	0.2	0.5	0.4	0.6	0.3	0.2	0																	
GyrB	0.4	0.5	0.5	0.5	0.7	0.4	0.4	0.4	0																
HSP60	0.4	0.4	0.6	0.4	0.7	0.3	0.5	0.4	0.5	0															
HSP70	0.4	0.5	0.6	0.6	0.6	0.4	0.5	0.5	0.5	0.5	0														
KsgA	0.3	0.4	0.4	0.5	0.7	0.5	0.4	0.4	0.5	0.5	0.5	0													
RecA	0.4	0.4	0.5	0.4	0.6	0.3	0.4	0.4	0.5	0.6	0.7	0.5	0												
RibL1	0.3	0.4	0.6	0.5	0.7	0.4	0.4	0.4	0.5	0.5	0.5	0.6	0.7	0											
RibL2	0.3	0.3	0.5	0.4	0.6	0.3	0.3	0.3	0.5	0.5	0.5	0.4	0.5	0.5	0										
RibL15	0.4	0.4	0.6	0.6	0.6	0.5	0.4	0.4	0.6	0.5	0.5	0.5	0.6	0.6	0.5	0									
RibS2	0.3	0.4	0.5	0.5	0.6	0.5	0.4	0.4	0.5	0.5	0.5	0.3	0.4	0.4	0.4	0.4	0								
RibS8	0.2	0.3	0.5	0.5	0.5	0.4	0.3	0.3	0.4	0.4	0.4	0.3	0.5	0.4	0.4	0.4	0.4	0							
RibS11	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0						
RpoA	0.3	0.4	0.4	0.5	0.6	0.5	0.3	0.4	0.5	0.6	0.4	0.4	0.5	0.6	0.5	0.5	0.3	0.3	0.7	0					
RpoB	0.3	0.3	0.6	0.5	0.7	0.3	0.4	0.4	0.4	0.4	0.5	0.6	0.5	0.4	0.4	0.5	0.5	0.4	0.7	0.6	0				
RpoC	0.4	0.4	0.6	0.6	0.7	0.4	0.4	0.4	0.5	0.6	0.6	0.6	0.6	0.6	0.4	0.6	0.6	0.5	0.7	0.6	0.4	0			
SecA	0.0	0.2	0.4	0.3	0.6	0.3	0.1	0.2	0.4	0.4	0.5	0.3	0.4	0.4	0.3	0.5	0.4	0.3	0.7	0.3	0.3	0.4	0		
SecY	0.2	0.2	0.5	0.4	0.6	0.3	0.2	0.2	0.4	0.5	0.5	0.4	0.3	0.4	0.3	0.5	0.5	0.3	0.7	0.4	0.3	0.3	0.2	0	
PcrA	0.2	0.3	0.4	0.4	0.7	0.4	0.3	0.3	0.4	0.5	0.5	0.3	0.4	0.5	0.3	0.5	0.3	0.3	0.7	0.4	0.5	0.5	0.3	0.3	0

*The colors of the cells represent heat map generated from the nRF scores. nRF value zero (dark green) indicates that the two trees are topologically congruent. Higher nRF scores (dark red) indicate a lower level of congruence between two tree topologies.

TABLE 5 | Normalized-Robinson-Foulds distance score comparisons of CGT with single protein MP trees.

	CGT	CdsA	EF-G	EF-P	EF-Tu	FtsY	GlyA	GyrA	GyrB	HSP60	HSP70	KsgA	RecA	RibL1	RibL2	RibL15	RibS2	RibS8	RibS11	RpoA	RpoB	RpoC	SecA	SecY	PcrA
CGT	0*																								
CdsA	0.3	0																							
EF-G	0.5	0.7	0																						
EF-P	0.4	0.6	0.5	0																					
EF-Tu	0.5	0.7	0.6	0.7	0																				
FtsY	0.2	0.4	0.6	0.4	0.6	0																			
GlyA	0.3	0.4	0.6	0.5	0.6	0.4	0																		
GyrA	0.3	0.4	0.6	0.4	0.7	0.4	0.4	0																	
GyrB	0.6	0.7	0.7	0.7	0.7	0.7	0.5	0.7	0																
HSP60	0.4	0.6	0.7	0.5	0.6	0.4	0.5	0.5	0.7	0															
HSP70	0.3	0.5	0.6	0.6	0.6	0.3	0.4	0.5	0.7	0.5	0														
KsgA	0.4	0.6	0.6	0.6	0.5	0.5	0.5	0.6	0.7	0.6	0.6	0													
RecA	0.4	0.6	0.6	0.5	0.7	0.5	0.5	0.5	0.7	0.7	0.6	0.5	0												
RibL1	0.5	0.6	0.7	0.7	0.6	0.5	0.6	0.5	0.7	0.6	0.5	0.7	0.7	0											
RibL2	0.6	0.6	0.7	0.6	0.8	0.5	0.6	0.5	0.8	0.6	0.6	0.8	0.7	0.7	0										
RibL15	0.5	0.5	0.6	0.6	0.8	0.5	0.5	0.5	0.7	0.7	0.6	0.7	0.6	0.7	0.7	0									
RibS2	0.6	0.6	0.7	0.7	0.7	0.6	0.7	0.6	0.7	0.7	0.6	0.5	0.6	0.7	0.7	0.7	0								
RibS8	0.4	0.5	0.6	0.7	0.7	0.4	0.5	0.6	0.7	0.6	0.5	0.5	0.6	0.7	0.7	0.6	0.6	0							
RibS11	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0						
RpoA	0.4	0.5	0.7	0.7	0.7	0.6	0.5	0.6	0.7	0.6	0.6	0.5	0.6	0.7	0.8	0.7	0.7	0.5	0.9	0					
RpoB	0.5	0.4	0.7	0.7	0.6	0.4	0.4	0.6	0.5	0.5	0.6	0.6	0.6	0.6	0.7	0.7	0.7	0.6	0.9	0.6	0				
RpoC	0.5	0.6	0.6	0.6	0.7	0.5	0.5	0.6	0.7	0.7	0.6	0.6	0.6	0.6	0.7	0.7	0.7	0.6	0.9	0.6	0.5	0			
SecA	0.2	0.4	0.5	0.4	0.6	0.4	0.3	0.3	0.6	0.5	0.5	0.5	0.4	0.6	0.6	0.5	0.6	0.5	0.9	0.5	0.5	0.6	0		
SecY	0.2	0.3	0.5	0.4	0.6	0.3	0.3	0.2	0.6	0.5	0.5	0.5	0.4	0.5	0.6	0.4	0.6	0.4	0.8	0.5	0.4	0.5	0.2	0	
PcrA	0.4	0.6	0.5	0.6	0.6	0.5	0.5	0.6	0.7	0.5	0.5	0.3	0.6	0.7	0.7	0.6	0.7	0.4	0.9	0.5	0.6	0.6	0.5	0.5	0

*The colors of the cells represent heat map generated from the nRF scores. nRF value zero (dark green) indicates that the two trees are topologically congruent. Higher nRF scores (dark red) indicate a lower level of congruence between two tree topologies.

TABLE 6 | Description of similarities and divergences among phylogenetic trees of 24 proteins most commonly applied in CNS species identification and phylogenetic estimations.

Protein/Method	Clade A Species			Clade A Branching			Clade B Species			Clade B Branching			Clade C Species			Clade D Species			Clade D Branching			Clade E Species			Clade E Branching		
	MP*	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML
CdsA	MP§	NJ	ML	MP§	NJ	.	MP	NJ	ML	MP	NJ	.	MP	NJ	ML	.	.	.	.	.	.	MP	NJ	ML	.	.	.
EF-G	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	.	.	.	MP	NJ	ML	.	.	.	MP	NJ	ML	.	.	.
EF-P	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EF-Tu	.	NJ	ML	.	NJ	.	MP	NJ	ML	.	.	ML	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
FtsY	MP	NJ	ML	MP	NJ	.	.	NJ	ML	.	NJ	.	MP	NJ	ML	MP	NJ	ML	.	NJ	.	MP	NJ	ML	MP	.	.
GlyA	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	.	NJ	.	.	.	.	MP	NJ	.	.	.	.	MP	NJ	ML	.	.	.
GyrA	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	.	.	MP	NJ	ML	.	.	.
GyrB	MP	NJ	ML	MP	NJ	ML	MP	.	.	.	NJ	.	.	.	.	MP	NJ	.	.	.	.	.	.	.	.	.	.
HSP60	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	.	.	.	.	MP	NJ	ML	.	.	.
HSP70	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	.	.	.	.	.	.	MP	.	ML	.	NJ	.
KsgA	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	.	.	.	MP	NJ	ML	.	.	.	.	.	.	.	.	.
RecA	MP	NJ	ML	MP	.	.	MP	NJ	ML	MP	NJ	ML	MP	.	.	.	NJ	ML	.	.	.	.	.	.	.	.	.
RibL1	MP	NJ	ML	MP	.	ML	MP	.	.	.	.	.	.	.	.	.	NJ	.	.	.	.	.	NJ	ML	.	.	.
RibL15	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	.	NJ	.	.	NJ	.	.	.	.	.	.	.	.	.	.	.	.	.
RibL2	MP	NJ	ML	MP	NJ	ML	.	.	.	.	NJ	.	.	NJ	ML	MP	NJ	ML	.	.	.	MP	NJ	ML	.	.	.
RibS11	.	.	ML	.	.	ML	.	.	.	.	.	.	.	.	ML	.	.	.	.	.	.	.	.	.	.	.	.
RibS2	MP	NJ	ML	MP	NJ	.	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	.	.	.	.	.	.	.	.	.	.	.	.
RibS8	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	.	ML	.	.	.	.	NJ	ML	.	.	.
RpoA	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	.	MP	NJ	.	.	.	.	.	.	.	.	.	.
RpoB	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	.	MP	NJ	ML	.	.	.	MP	NJ	ML	.	NJ	.
RpoC	MP	.	ML	.	.	ML	.	.	ML	.	.	ML	.	.	ML	.	.	ML	.	.	.	.	.	.	.	.	.
SecA	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	.	.	MP	NJ	ML	MP	NJ	.	MP	NJ	.	.	NJ	.
SecY	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	.	NJ	ML	MP	NJ	ML	.	.	.	MP	NJ	ML	.	.	.
PcrA	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	MP	NJ	ML	.	.	.	MP	NJ	ML	.	NJ	.	.	.	.	.	.	.

Methods of tree construction MP, Maximum parsimony; NJ, Neighbor-joining; ML, Maximum parsimony.

§*Color shaded areas in orange, green, blue, yellow, and pink indicates the methods in each the presence of species, and branching at respective clades were in agreement to clades observed in CGT.*

$*Gray areas indicate disagreement between the species, and branching at respective clade compared to CGT.*

other researchers reported a protective effect of naturally occurring IMI with NAS (Matthews et al., 1990; De Vliegher et al., 2004). Species-specific studies aiming to determine the role of individual NAS species have been conducted, but these studies have used various and often suboptimal genotyping techniques (Vanderhaeghen et al., 2015). Current understanding regarding evolutionary relationships of NAS and other staphylococcal species is primarily based on information obtained from phylogenetic trees of 16S rRNA gene and various other individual gene/protein sequences (Takahashi et al., 1999; Drancourt and Raoult, 2002; Kwok and Chow, 2003; Shah et al., 2007; Ghebremedhin et al., 2008), which often produces conflicting phylogenies (Ghebremedhin et al., 2008; Lamers et al., 2012). Secondly, most previous studies attempted to resolve the *Staphylococcus* phylogeny from a variety of hosts, including humans and animals. Unfortunately, a reliable hierarchical classification of NAS species of bovine origin, which reveals true evolutionary histories and speciation, has not yet been reported.

In this study, WGS-based phylogenetic trees, constructed using ML, MP, and NJ methods, employing different evolutionary models, divided bovine NAS species into five distinct monophyletic clades. Earlier studies based on single genes and concatenated sequences suggested from 4 to 11 species groups in staphylococci (Kloos et al., 1998a; Takahashi et al., 1999; Drancourt and Raoult, 2002; Kwok and Chow, 2003; Shah et al., 2007; Ghebremedhin et al., 2008; Lamers et al., 2012). Variability in these groupings was attributed to input data sets, methods, and evolutionary models of tree construction.

WGS-based phylogenies of bovine NAS, constructed in this study, produced consensus grouping, and interspecies relationships when tested against various methods (ML, MP, and NJ) and evolutionary models [Jones-Taylor-Thornton (JTT), Whelan and Goldman (WAG), and Subtree-Pruning-Regrafting (SPR) models]. There were similar major clades in our multilocus trees (ML and NJ) when tested against General Time Reversible, Subtree-Pruning-Regrafting, and Kimura 2-parameter models. However, there were variable relationships of the species within larger clades in these trees. Division of bovine NAS into five major clades represent ancient events in evolutionary history of NAS, whereas the branching order within clades (toward the tip of the trees) represent recent evolutionary changes. A recent study on staphylococcal phylogeny, using a multilocus data set of five genes, inferred consistent placement of species near the tips of the trees, but also indicated limitations of inferring phylogenies of higher taxa (Lamers et al., 2012). However, in our WGS-based phylogenies, there was a consistent grouping of larger clades (ancient evolution) and species relationships (recent evolutionary histories) for NAS species obtained in all five WGS trees.

For this study, we also constructed 72 single protein trees (24 ML, 24 MP, and 24 NJ), 15 single gene trees (5 ML, 5 MP, and 5 NJ), and three multilocus trees (ML, MP, and NJ). The purpose of these phylogenetic reconstructions was to observe effects of various data, methods, and evolutionary models on inferring phylogenetic relationships among bovine NAS species. Interestingly, none of the single gene/protein trees, constructed using any of the phylogenetic estimation methods, produced identical consensus phylogeny to WGS phylogenetic reconstruction. There were several unresolved, inconsistent, and conflicting relationships among NAS in single genes/proteins trees. Therefore, we inferred that single genes/proteins may not be suitable for identification and phylogeny of NAS species. Conflicting results of single genes/proteins are understandable, as phylogenetic reconstructions based upon single genes/proteins are potentially affected by many parameters. Incomplete lineage sorting, gene gain/loss, gene duplication, presence of paralogs, convergent evolution, and horizontal gene transfers interfere with tree construction (Maddison and Knowles, 2006; Burbrink and Pyron, 2011; Chung and Ané, 2011; Mendes and Hahn, 2016; Stadler et al., 2016). Additionally, each gene may be under different selection pressure and therefore may have been evolving differently, resulting in differing phylogenies (Degnan and Rosenberg, 2006; Burbrink and Pyron, 2011).

In our single proteins phylogenetic estimations of bovine NAS species, Clades A and B were recovered in most reconstructions (**Table 6**). However, there were more conflicting results for species composition and species relationships in Clades D and E. Therefore, we inferred that that evolutionary interval after speciation may also have affected phylogenetic relationships. For example, in case of Clades A and B of bovine NAS (which represent earliest lineages), sufficient evolutionary changes have accumulated to differentiate respective species of these clades. This is in contrast with Clade D and Clade E, which were more recent lineages that have not resolved properly in most of our single proteins phylogenies, as previously concluded (Takahashi et al., 1999; Kwok and Chow, 2003; Shah et al., 2007; Ghebremedhin et al., 2008). Additionally, it is well-established that single-gene trees reflect evolution of particular genes and do not necessarily relate to speciation. Hence, single gene phylogenies may not be optimal for studying species relationships (Beiko et al., 2005; Puigbò et al., 2010).

All species descending from a common ancestor are expected to share some biochemical, phenotypical or physiological characteristics. In our WGS trees, placement of *S. sciuri*, *S. fleurettii*, and *S. vitulinus* into Clade A, which represented oxidase-positive and novobiocin-resistant members, often defined as "Sciuri group," was consistent with previous studies (Drancourt and Raoult, 2002; Kwok and Chow, 2003; Svec et al., 2004; Shah et al., 2007; Lamers et al., 2012). All remaining bovine NAS species were oxidase-negative, indicating that the gene responsible for oxidase activity was lost after divergence of Clade A. Species included in Clade B were oxidase-negative, novobiocin-susceptible, and coagulase variable, including *S. chromogenes* which is coagulase-negative and *S. hyicus* and *S. agnetis* which are coagulase variable. Within this clade, *S. chromogenes* is, with ∼50% of NAS isolations from bovine milk samples, the most prevalent species in IMI in Canada (Condas et al., in review) and worldwide (Sampimon et al., 2009; Piessens et al., 2011). Placement of Clade B as second monophyletic lineage of NAS species, supported by 100% nodal support, was unique to our WGS and multilocus trees. Species of this clade was given variable placement and species relatedness in previous reconstructions from multilocus data set (Lamers et al., 2012) and various single genes trees, including,16S rRNA, *hsp60*,

rpoB, *sodA*, *gap*, *tuf*, and *dnaJ* sequences (Takahashi et al., 1999; Drancourt and Raoult, 2002; Kwok and Chow, 2003; Shah et al., 2007; Ghebremedhin et al., 2008). All our WGS trees and multilocus trees had monophyletic branching of Clade B as the second lineage of NAS. However, variable phylogenetic placement of this clade was observed in single genes/proteins trees.

Clade C appeared as third lineage in our WGS trees and is composed of multiple isolates of *S. simulans*, the second most prevalent NAS species found in bovine IMI in Canada (Condas et al., in review). These isolates are novobiocin-susceptible, coagulase and oxidase-negative in phenotype. Our placement of *S. simulans* as third lineage was consistent with earlier studies (Lamers et al., 2012). However, discrepancy in branching of this clade was noted in many of our single-protein phylogenies (**Tables 3–6**) and in previous single-gene phylogenetic reconstructions (Takahashi et al., 1999; Kwok and Chow, 2003; Ghebremedhin et al., 2008). Species included in Clade D of our WGS trees, were novobiocin-susceptible, oxidase-negative, and coagulase-negative. The species of this clade were further divided into three cluster groups, consistent with a previous classification (Lamers et al., 2012). Species included in Clade E were novobiocin-resistant, oxidase-negative and coagulase-negative, except *S. auricularis* which was novobiocin susceptible, oxidase-negative and coagulase-negative (Lamers et al., 2012). Our placement of *S. auricularis* sharing a common ancestor with the other species of Clade E, was generally in agreement with previous 16S rRNA gene based phylogenies (Takahashi et al., 1999; Ghebremedhin et al., 2008) but contradictory to another phylogeny (Lamers et al., 2012) that proposed *S. auricularis* as a second divergence group within Staphylococci. However, there was a different pattern of branching for *S. auricularis* in their 16S rRNA and *dnaJ* base phylogenetic reconstructions (Lamers et al., 2012). Our placement of *S. auricularis* with other species of Clade E was consistent in all WGS and multilocus trees. Perhaps differences between current and previous (Lamers et al., 2012) multilocus reconstructions were due to our analyses restricted to NAS of bovine origin, and excluded NAS species from other hosts. Notwithstanding, WGS data enabled us to construct robust phylogeny of bovine NSA isolates obtained from CBMQRN. Based on these results, an association can now be investigated between phylogeny and epidemiology of these isolates. This study represents the first multi-genome-scale phylogenetic analysis of bovine NAS isolates.

CONCLUSIONS

Based on WGS, we established a robust phylogeny of bovine NAS species, supported by various methods and models of evolutionary estimations. Evolutionary and interspecies relationships revealed by this phylogeny will allow determination, among related species, of common biological traits, such as virulence, environmental niche, and geographical distribution, and host specificity. Knowing the true phylogeny of NAS species will also provide the basis for studies to elucidate the role and significance of individual and related NAS species in udder health.

We also compared WGS phylogenetic trees with single genes/proteins and multilocus trees. Based on our results, we conclude that whole genome sequences, which encompass the entire genetic information of organisms, offer great accuracy in reconstructing evolutionary relationships and should be "the method of choice" for identification and elucidation of evolutionary histories of bacterial organisms. Inferences derived from this work will be helpful in selection of parameters, such as input data, evolutionary models and methods of phylogenetic reconstruction for many other groups of bacteria.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: SN, HB, CL, and JD. Performed lab experiments: SN, LC, DN, and DC. Computational analyses and analyzed the data: SN and JD. Wrote the paper: SN, HB, and JD.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2016.01990/full#supplementary-material

REFERENCES

Aarestrup, F. M., Wegener, H. C., Rosdahl, V. T., and Jensen, N. E. (1995). Staphylococcal and other bacterial species associated with intramammary infections in Danish dairy herds. *Acta Vet. Scand.* 36, 475–487.

Beiko, R. G., Harlow, T. J., and Ragan, M. A. (2005). Highways of gene sharing in prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14332–14337. doi: 10.1073/pnas.0504068102

Bernard, G., Chan, C. X., and Ragan, M. A. (2016). Alignment-free microbial phylogenomics under scenarios of sequence divergence, genome rearrangement and lateral genetic transfer. *Sci. Rep.* 6:28970. doi: 10.1038/srep28970

Boc, A., Diallo, A. B., and Makarenkov, V. (2012). T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res.* 40, W573–W579. doi: 10.1093/nar/gks485

Burbrink, F. T., and Pyron, R. A. (2011). The impact of gene-tree/species-tree discordance on diversification-rate estimation. *Evolution* 65, 1851–1861. doi: 10.1111/j.1558-5646.2011.01260.x

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinform.* 10:421. doi: 10.1186/1471-2105-10-421

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

Chung, Y., and Ané, C. (2011). Comparing two Bayesian methods for gene tree/species tree reconstruction: simulations with incomplete lineage sorting and horizontal gene transfer. *Syst. Biol.* 60, 261–275. doi: 10.1093/sysbio/syr003

Degnan, J. H., and Rosenberg, N. A. (2006). Discordance of species trees with their most likely gene trees. *PLoS Genet.* 2:e68. doi: 10.1371/journal.pgen.0020068

De Vliegher, S., Fox, L. K., Piepers, S., McDougall, S., and Barkema, H. W. (2012). Invited review: mastitis in dairy heifers: nature of the disease, potential impact, prevention, and control. *J. Dairy Sci.* 95, 1025–1040. doi: 10.3168/jds.2010-4074

De Vliegher, S., Opsomer, G., Vanrolleghem, A., Devriese, L. A., Sampimon, O. C., Sol, J., et al. (2004). *In vitro* growth inhibition of major mastitis pathogens by *Staphylococcus chromogenes* originating from teat apices of dairy heifers. *Vet. Microbiol.* 101, 215–221. doi: 10.1016/j.vetmic.2004.03.020

Dos Santos, D. C., Lange, C. C., Avellar-Costa, P., Dos Santos, K. R., Brito, M. A., and Giambiagi-deMarval, M. (2016). Staphylococcus chromogenes, a coagulase-negative staphylococcus species that can clot plasma. *J. Clin. Microbiol.* 54, 1372–1375. doi: 10.1128/JCM.03139-15

Drancourt, M., and Raoult, D. (2002). *rpoB* gene sequence-based Identification of *Staphylococcus* species. *J. Clin. Microbiol.* 40, 1333–1338. doi: 10.1128/JCM.40.4.1333-1338.2002

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Frey, Y., Rodriguez, J. P., Thomann, A., Schwendener, S., and Perreten, V. (2013). Genetic characterization of antimicrobial resistance in coagulase-negative staphylococci from bovine mastitis milk. *J. Dairy Sci.* 96, 2247–2257. doi: 10.3168/jds.2012-6091

Ghebremedhin, B., Layer, F., König, W., and König, B. (2008). Genetic classification and distinguishing of *Staphylococcus* species based on different partial gap, 16S rRNA, *hsp60*, *rpoB*, *sodA*, and *tuf* gene sequences. *J. Clin. Microbiol.* 46, 1019–1025. doi: 10.1128/JCM.02058-07

Gillespie, B. E., Headrick, S. I., Boonyayatra, S., and Oliver, S. P. (2009). Prevalence and persistence of coagulase-negative *Staphylococcus* species in three dairy research herds. *Vet. Microbiol.* 134, 65–72. doi: 10.1016/j.vetmic.2008.09.007

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Honkanen-Buzalski, T., Myllys, V., and Pyörölö, S. (1994). Bovine clinical mastitis due to coagulase-negative staphylococci and their susceptibility to antimicrobials. *Zentralbl. Veterinarmed. B* 41, 344–350. doi: 10.1111/j.1439-0450.1994.tb00237.x

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Compu. Appl. Biosci.* 8, 275–282. doi: 10.1093/bioinformatics/8.3.275

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120. doi: 10.1007/BF01731581

Kloos, W. E., Ballard, D. N., George, C. G., Webster, J. A., Hubner, R. J., Ludwig, W., et al. (1998a). Delimiting the genus Staphylococcus through description of *Macrococcus caseolyticus* gen. nov., comb. nov. and *Macrococcus equipercicus* sp. nov., and *Macrococcus bovicus* sp. no. and *Macrococcus carouselicus* sp. nov. *Int. J. Syst. Bacteriol.* 48(Pt 3), 859–877. doi: 10.1099/00207713-48-3-859

Kloos, W. E., George, C. G., Olgiate, J. S., Van Pelt, L., McKinnon, M. L., Zimmer, B. L., et al. (1998b). *Staphylococcus hominis* subsp. *novobiosepticus* subsp. nov., a novel trehalose- and N-acetyl-D-glucosamine-negative, novobiocin- and multiple-antibiotic-resistant subspecies isolated from human blood cultures. *Int. J. Syst. Bacteriol.* 48(Pt 3), 799–812. doi: 10.1099/00207713-48-3-799

Köster, J., and Rahmann, S. (2012). Snakemake–a scalable bioinformatics workflow engine. *Bioinformatics* 28, 2520–2522. doi: 10.1093/bioinformatics/bts480

Kupczok, A., Schmidt, H. A., and von Haeseler, A. (2010). Accuracy of phylogeny reconstruction methods combining overlapping gene data sets. *Algorithms Mol. Biol.* 5:37. doi: 10.1186/1748-7188-5-37

Kwok, A. Y., and Chow A., W. (2003). Phylogenetic study of *Staphylococcus* and *Macrococcus* species based on partial *hsp60* gene sequences. *Int. J. Syst. Evol. Microbiol.* 53, 87–92. doi: 10.1099/ijs.0.02210-0

Laevens, H., Deluyker, H., Schukken, Y. H., De Meulemeester, L., Vandermeersch, R., De Muêlenaere, E., et al. (1997). Influence of parity and stage of lactation on the somatic cell count in bacteriologically negative dairy cows. *J. Dairy Sci.* 80, 3219–3226. doi: 10.3168/jds.S0022-0302(97)76295-7

Lamers, R. P., Muthukrishnan, G., Castoe, T. A., Tafur, S., Cole, A. M., Parkinson, C. L. (2012). Phylogenetic relationships among Staphylococcus species and refinement of cluster groups based on multilocus data. *BMC Evol. Biol.* 12:171. doi: 10.1186/1471-2148-12-171

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698

Maddison, W. P., and Knowles, L. L. (2006). Inferring phylogeny despite incomplete lineage sorting. *Syst. Biol.* 55, 21–30. doi: 10.1080/10635150500354928

Makarenkov, V., and Leclerc, B. (2000). Comparison of additive trees using circular orders. *J. Comput. Biol.* 7, 731–744. doi: 10.1089/106652701446170

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10–12. doi: 10.14806/ej.17.1.200

Matthews, K. R., Harmon, R. J., and Smith, B. A. (1990). Protective effect of *Staphylococcus chromogenes* infection against *Staphylococcus aureus* infection in the lactating bovine mammary gland. *J. Dairy Sci.* 73, 3457–3462. doi: 10.3168/jds.S0022-0302(90)79044-3

Mellmann, A., Becker, K., von Eiff, C., Keckevoet, U., Schumann, P., and Harmsen, D. (2006). Sequencing and staphylococci identification. *Emer. Infect. Dis.* 12, 333–336. doi: 10.3201/eid1202.050962

Mendes, F. K., and Hahn, M. W. (2016). Gene tree discordance causes apparent substitution rate variation. *Syst. Biol.* 65, 711–721. doi: 10.1093/sysbio/syw018

Naushad, S., Adeolu, M., Goel, N., Khadka, B., Al-Dahwi, A., and Gupta, R. S. (2015a). Phylogenomic and molecular demarcation of the core members of the polyphyletic *Pasteurellaceae* genera *Actinobacillus*, *Haemophilus*, and *Pasteurella*. *Int. J. Genomics* 2015:198560. doi: 10.1155/2015/198560

Naushad, S., Adeolu, M., Wong, S., Sohail, M., Schellhorn, H. E., and Gupta, R. S. (2015b). A phylogenomic and molecular marker based taxonomic framework for the order *Xanthomonadales*: proposal to transfer the families *Algiphilaceae* and *Solimonadaceae* to the order *Nevskiales* ord. nov. and to create a new family within the order *Xanthomonadales*, the family *Rhodanobacteraceae* fam. nov., containing the genus *Rhodanobacter* and its closest relatives. *Antonie Van Leeuwenhoek* 107, 467–485. doi: 10.1007/s10482-014-0344-8

Nei, M., and Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. Oxford: Oxford University Press.

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A. A., Korobeynikov, A., Lapidus, A., et al. (2013). Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J. Comput. Biol.* 20, 714–737. doi: 10.1089/cmb.2013.0084

Otto, M. (2013). Coagulase-negative staphylococci as reservoirs of genes facilitating MRSA infection: staphylococcal commensal species such as *Staphylococcus epidermidis* are being recognized as important sources of genes promoting MRSA colonization and virulence. *Bioessays* 35, 4–11. doi: 10.1002/bies.201200112

Piepers, S., De Meulemeester, L., de Kruif, A., Opsomer, G., Barkema, H. W., and De Vliegher, S. (2007). Prevalence and distribution of mastitis pathogens in subclinically infected dairy cows in Flanders, Belgium. *J. Dairy Res.* 74, 478–483. doi: 10.1017/S0022029907002841

Piepers, S., De Vliegher, S., de Kruif, A., Opsomer, G., and Barkema, H. W. (2009). Impact of intramammary infections in dairy heifers on future udder health, milk production, and culling. *Vet. Microbiol.* 134, 113–120. doi: 10.1016/j.vetmic.2008.09.017

Piessens, V., Van Coillie, E., Verbist, B., Supré, K., Braem, G., Van Nuffel, A., et al. (2011). Distribution of coagulase-negative *Staphylococcus* species from milk and environment of dairy cows differs between herds. *J. Dairy Sci.* 94, 2933–2944. doi: 10.3168/jds.2010-3956

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5:e9490. doi: 10.1371/journal.pone.0009490

Puigbò, P., Wolf, Y. I., and Koonin, E. V. (2010). The tree and net components of prokaryote evolution. *Genome Biol. Evol.* 2, 745–756. doi: 10.1093/gbe/evq062

Pyörälä, S., and Taponen, S. (2009). Coagulase-negative staphylococci—Emerging mastitis pathogens. *Vet. Microbiol.* 134, 3–8. doi: 10.1016/j.vetmic.2008.09.015

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Rajala-Schultz, P. J., Torres, A. H., Degraves, F. J., Gebreyes, W. A., and Patchanee, P. (2009). Antimicrobial resistance and genotypic characterization of coagulase-negative staphylococci over the dry period. *Vet. Microbiol.* 134, 55–64. doi: 10.1016/j.vetmic.2008.09.008

Reyher, K. K., Dufour, S., Barkema, H. W., Des Coteaux, L., Devries, T. J., Dohoo, I. R., et al. (2011). The National Cohort of Dairy Farms–a data collection platform for mastitis research in Canada. *J. Dairy Sci.* 94, 1616–1626. doi: 10.3168/jds.2010-3180

Robinson, D. F., and Foulds, L. R. (1981). Comparison of phylogenetic trees. *Math. Biosci.* 53, 131–147. doi: 10.1016/0025-5564(81)90043-2

Rokas, A., Williams, B. L., King, N., and Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425, 798–804. doi: 10.1038/nature02053

Sampimon, O. C., Barkema, H. W., Berends, I. M. G. A., Sol, J., and Lam, T. J. G. M. (2009). Prevalence and herd-level risk factors for intramammary infection with coagulase-negative staphylococci in Dutch dairy herds. *Vet. Microbiol.* 134, 37–44. doi: 10.1016/j.vetmic.2008.09.010

Schukken, Y. H., González, R. N., Tikofsky, L. L., Schulte, H. F., Santisteban, C. G., Welcome, F. L., et al. (2009). CNS mastitis: nothing to worry about? *Vet. Microbiol.* 134, 9–14. doi: 10.1016/j.vetmic.2008.09.014

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Segata, N., Börnigen, D., Morgan, X. C., and Huttenhower, C. (2013). PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* 4, 2304. doi: 10.1038/ncomms3304

Shah, M. M., Iihara, H., Noda, M., Song, S. X., Nhung, P. H., Ohkusu, K., et al. (2007). *dnaJ* gene sequence-based assay for species identification and phylogenetic grouping in the genus *Staphylococcus*. *Int. J. Syst. Evol. Microbiol.* 57, 25–30. doi: 10.1099/ijs.0.64205-0

Stadler, T., Degnan, J. H., and Rosenberg, N. A. (2016). Does gene tree discordance explain the mismatch between macroevolutionary models and empirical patterns of tree shape and branching times? *Syst. Biol.* 65, 628–639. doi: 10.1093/sysbio/syw070

Supré, K., Haesebrouck, F., Zadoks, R. N., Vaneechoutte, M., Piepers, S., and De Vliegher, S. (2011). Some coagulase-negative *Staphylococcus* species affect udder health more than others. *J. Dairy Sci.* 94, 2329–2340. doi: 10.3168/jds.2010-3741

Svec, P., Vancanneyt, M., Sedlácek, I., Engelbeen, K., Stetina, V., Swings, J., et al. (2004). Reclassification of *Staphylococcus pulvereri* Zakrzewska-Czerwinska et al. 1995 as a later synonym of *Staphylococcus vitulinus* Webster et al. 1994. *Int. J. Syst. Evol. Microbiol.* 54, 2213–2215. doi: 10.1099/ijs.0.63080-0

Takahashi, T., Satoh, I., and Kikuchi, N. (1999). Phylogenetic relationships of 38 taxa of the genus *Staphylococcus* based on 16S rRNA gene sequence analysis. *Int. J. Syst. Bacteriol.* 49(Pt 2), 725–728. doi: 10.1099/00207713-49-2-725

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Taponen, S., Koort, J., Björkroth, J., Saloniemi, H., and Pyörälä, S. (2007). Bovine intramammary infections caused by coagulase-negative staphylococci may persist throughout lactation according to amplified fragment length polymorphism-based analysis. *J. Dairy Sci.* 90, 3301–3307. doi: 10.3168/jds.2006-860

Taponen, S., Nykäsenoja, S., Pohjanvirta, T., Pitkälä, A., and Pyörälä, S. (2015). Species distribution and *in vitro* antimicrobial susceptibility of coagulase-negative staphylococci isolated from bovine mastitic milk. *Acta Vet. Scand.* 58, 12. doi: 10.1186/s13028-016-0193-8

Taponen, S., and Pyörälä, S. (2009). Coagulase-negative staphylococci as cause of bovine mastitis—Not so different from *Staphylococcus aureus*? *Vet. Microbiol.* 134, 29–36. doi: 10.1016/j.vetmic.2008.09.011

Taponen, S., Supré, K., Piessens, V., Van Coillie, E., De Vliegher, S., and Koort, J. M. (2012). *Staphylococcus agnetis* sp. nov., a coagulase-variable species from bovine subclinical and mild clinical mastitis. *Int. J. Syst. Evol. Microbiol.* 62, 61–65. doi: 10.1099/ijs.0.028365-0

Thorberg, B. M., Danielsson-Tham, M. L., Emanuelson, U., and Persson Waller, K. (2009). Bovine subclinical mastitis caused by different types of coagulase-negative staphylococci. *J. Dairy Sci.* 92, 4962–4970. doi: 10.3168/jds.2009-2184

Vanderhaeghen, W., Piepers, S., Leroy, F., Van Coillie, E., Haesebrouck, F., and De Vliegher, S. (2014). Invited review: effect, persistence, and virulence of coagulase-negative *Staphylococcus* species associated with ruminant udder health. *J. Dairy Sci.* 97, 5275–5293. doi: 10.3168/jds.2013-7775

Vanderhaeghen, W., Piepers, S., Leroy, F., Van Coillie, E., Haesebrouck, F., and De Vliegher, S. (2015). Identification, typing, ecology and epidemiology of coagulase negative staphylococci associated with ruminants. *Vet. J.* 203, 44–51. doi: 10.1016/j.tvjl.2014.11.001

Whelan, S., and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 18, 691–699. doi: 10.1093/oxfordjournals.molbev.a003851

Woodward, W. D., Besser, T. E., Ward, A. C., and Corbeil, L. B. (1987). *In vitro* growth inhibition of mastitis pathogens by bovine teat skin normal flora. *Can. J. Vet. Res.* 51, 27–31.

Woodward, W. D., Ward, A. C. S., Fox, L. K., and Corbeil, L. B. (1988). Teat skin normal flora and colonization with mastitis pathogen inhibitors. *Vet. Microbiol.* 17, 357–365. doi: 10.1016/0378-1135(88)90049-1

Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N. N., et al. (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* 462, 1056–1060. doi: 10.1038/nature08656

Zhang, Y.-Q., Ren, S.-X., Li, H.-L., Wang, Y.-X., Fu, G., Yang, J., et al. (2003). Genome-based analysis of virulence genes in a non-biofilm-forming *Staphylococcus epidermidis* strain (ATCC 12228). *Mol. Microbiol.* 49, 1577–1593. doi: 10.1046/j.1365-2958.2003.03671.x