



OPEN

Genomic comparisons and phylogenetic analysis of mastitis-related staphylococci with a focus on adhesion, biofilm, and related regulatory genes

Lucas José Luduverio Pizauro^{1✉}, Camila Chioda de Almeida², Saura Rodrigues Silva¹, Janet I. MacInnes³, Andrew M. Kropinski³, Luiz Francisco Zafalon⁴, Fernando Antônio de Avila² & Alessandro de Mello Varani¹

Mastitis is a common and costly disease on dairy farms, commonly caused by *Staphylococcus* spp. though the various species are associated with different clinical outcomes. In the current study, we performed genomic analyses to determine the prevalence of adhesion, biofilm, and related regulatory genes in 478 staphylococcal species isolated from clinical and subclinical mastitis cases deposited in public databases. The most prevalent adhesion genes (*ebpS*, *atl*, *pls*, *sasH* and *sasF*) were found in both clinical and subclinical isolates. However, the *ebpS* gene was absent in subclinical isolates of *Staphylococcus arlettae*, *S. succinus*, *S. sciuri*, *S. equorum*, *S. galinarum*, and *S. saprophyticus*. In contrast, the *coa*, *eap*, *emp*, *efb*, and *vWbp* genes were present more frequently in clinical (vs. subclinical) mastitis isolates and were highly correlated with the presence of the biofilm operon (*icaABCD*) and its transcriptional regulator, *icaR*. Co-phylogenetic analyses suggested that many of these adhesins, biofilm, and associated regulatory genes could have been horizontally disseminated between clinical and subclinical isolates. Our results further suggest that several adhesins, biofilm, and related regulatory genes, which have been overlooked in previous studies, may be of use for virulence profiling of mastitis-related *Staphylococcus* strains or as potential targets for vaccine development.

Bovine mastitis is one of the costliest diseases seen on dairy farms, with an estimated global loss of 19.7 to 32 billion US\$ due to reduced milk production and withdrawal periods related to antibiotic usage¹. Mastitis may also cause death directly or lead to the slaughter of chronically infected animals¹. *S. aureus* is generally considered the most important cause of both clinical and subclinical mastitis, while coagulase-negative staphylococci or non-aureus *Staphylococcus* spp. are thought to be of lesser importance or as opportunistic pathogens². A higher prevalence of subclinical versus clinical infections has been reported³. Moreover, the prevalence of subclinical mastitis is likely to be underestimated due to the lack of obvious signs except for changes in milk quantity and quality (which can only be detected by specific tests such as the California Mastitis Test and by somatic cell counting)¹. It is generally believed that the *Staphylococcus* spp. isolates associated with chronic infections are different from those that cause acute infections and are more likely to be transmitted and persist in the herd due to better host adaptation and the absence of overt clinical signs^{1,2}.

Attachment and colonization are crucial in the development of mastitis. In *Staphylococcus* spp., factors such as the fibronectin-binding proteins (*fnbA* and *fnbB*), elastin binding proteins (*ebpS*), clumping factors (*clfA* and *clfB*), and collagen-binding protein (*cna*) play important roles in binding to host cells, colonization, and invasion². The *pls* gene, which encodes the plasmin-sensitive protein, also has an important role in bacterial adherence⁴. The surface proteins encoded by *sasH* and *sasF*, play a significant role in virulence because they bind to host extracellular matrix and plasma components. Recently, they have been reported to be prevalent

¹Department of Technology, Sao Paulo State University, Jaboticabal, Sao Paulo, Brazil. ²Department of Microbiology, Sao Paulo State University, Jaboticabal, Sao Paulo, Brazil. ³Department of Pathobiology, Ontario Veterinary College, University of Guelph, Guelph, ON, Canada. ⁴Brazilian Agricultural Research Corporation (EMBRAPA), Embrapa Southeast Livestock, Sao Carlos, Sao Paulo, Brazil. ✉email: lucaspizauro@gmail.com

Species	Mastitis	
	Clinical	Subclinical
<i>S. agnetis</i>	2.50% (2)	3.52% (14)
<i>S. arlettae</i>	1.25% (1)	3.27% (13)
<i>S. aureus</i>	18.7% (15)	3.27% (13)
<i>S. auricularis</i>	0.00% (0)	0.50% (2)
<i>S. capitis</i>	1.25% (1)	5.28% (21)
<i>S. caprae</i>	0.00% (0)	0.50% (2)
<i>S. chromogenes</i>	28.7% (23)	15.5% (62)
<i>S. cohnii</i>	1.25% (1)	5.78% (23)
<i>S. devriesei</i>	1.25% (1)	1.76% (7)
<i>S. epidermidis</i>	5.00% (4)	5.53% (22)
<i>S. equorum</i>	0.00% (0)	4.52% (18)
<i>S. fleurettii</i>	0.00% (0)	0.50% (2)
<i>S. gallinarum</i>	1.25% (1)	5.03% (20)
<i>S. haemolyticus</i>	5.00% (4)	6.28% (25)
<i>S. hominis</i>	1.25% (1)	3.27% (13)
<i>S. hyicus</i>	0.00% (0)	0.50% (2)
<i>S. kloosii</i>	0.00% (0)	0.25% (1)
<i>S. nepalensis</i>	0.00% (0)	0.50% (2)
<i>S. pasteurii</i>	0.00% (0)	1.51% (6)
<i>S. saprophyticus</i>	0.00% (0)	4.02% (16)
<i>S. sciuri</i>	10.0% (8)	5.28% (21)
<i>S. simulans</i>	20.0% (16)	6.78% (27)
<i>S. succinus</i>	0.00% (0)	3.77% (15)
<i>S. vitulinus</i>	0.00% (0)	1.51% (6)
<i>S. warneri</i>	0.00% (0)	4.77% (19)
<i>S. xylosus</i>	2.50% (2)	6.53% (26)
Total	100% (80)	100% (398)

Table 1. Relative and absolute frequency of staphylococcal species obtained from complete genome of staphylococci associated with bovine/bubaline mastitis.

adhesins in genomes of *Staphylococcus* spp. isolates from cattle⁵. In addition, the *ica* genes, which are associated with the synthesis of polysaccharide intercellular adhesin (PIA), are thought to play a crucial role in biofilm development in these bacteria.

Molecular epidemiology-based methods such as specific PCR assays, MLST, and PFGE have been used to analyze the genetic diversity and virulence factors and to track the dissemination of *Staphylococcus* spp. infections, but they have limitations¹. Accordingly, the current work aimed to ascertain the prevalence of adhesion and biofilm genes by investigating whole genome sequences of *Staphylococcus* spp. from clinical and subclinical mastitis cases and evaluate the phylogenetic relationship of these isolates and determine if any adhesin or biofilm genes associated with acute bovine/bubaline mastitis.

Results

Assessment of clinical and subclinical mastitis *Staphylococcus* spp. isolates. *Staphylococcus chromogenes* (28.7%), *S. simulans* (20.0%), *S. aureus* (18.7%) and the *S. sciuri* (10.0%) were the most prevalent clinical mastitis species deposited in the NCBI GenBank database. The most frequent staphylococcal species associated with subclinical mastitis were *S. chromogenes* (15.6%), *S. simulans* (6.8%), *S. xylosus* (6.5%), *S. haemolyticus* (6.3%), *S. cohnii* (5.8%), *S. epidermidis* (5.5%), *S. capitis* (5.3%), *S. sciuri* (5.3%) (Table 1, Fig. 1).

Distribution of adhesin, biofilm and regulatory genes. In the mastitis related staphylococcal genomes analyzed (n=478) the most prevalent genes associated with adhesion and biofilm formation were: *ebpS* (71.3%), *atl* (70.9%), *sasF* (70.7%), *sasH* (53.3%), *araC* (52.1%), *tcaR* (52.1%), *sarA* (52.1%), *sigB* (52.1%) *pls* (44.6%), *sasA* (37.2%) and *sasC* (30.8%) (Table 2, Fig. 2).

In strains associated with clinical mastitis, the *ebpS* (83.8%), *atl* (83.8%), *sasF* (83.8%), *sasH* (77.5%), *rbf* (56.3%), *tcaR* (56.3%), *sarA* (56.3%), *sigB* (56.3%), *pls* (55.0%), *sasA* (47.5%), *pls* (37.5%) and *sasC* (30.0%) genes were most frequently detected. The carriage of adhesin/biofilm related genes in isolates associated with subclinical mastitis were less frequent (e.g., *ebpS* (68.8%), *atl* (68.3%), *sasF* (68.1%), *atl* (51.3%), *rbf* (51.3%), *tcaR* (51.3%), *sarA* (51.3%), *sigB* (51.3%), *sasH* (48.5%), *pls* (42.5%), *sasA* (35.2%) and *sasC* (30.9%) (Table 2).

Most of the subclinical isolates of *S. aureus*, *S. capitis*, *S. chromogenes*, *S. cohnii*, *S. epidermidis*, *S. haemolyticus*, *S. warneri* and *S. simulans* had the *ebpS*, *atl*, *pls*, *sasH*, *sasC* and *sasF* adhesion-associated genes with the *clfB* and

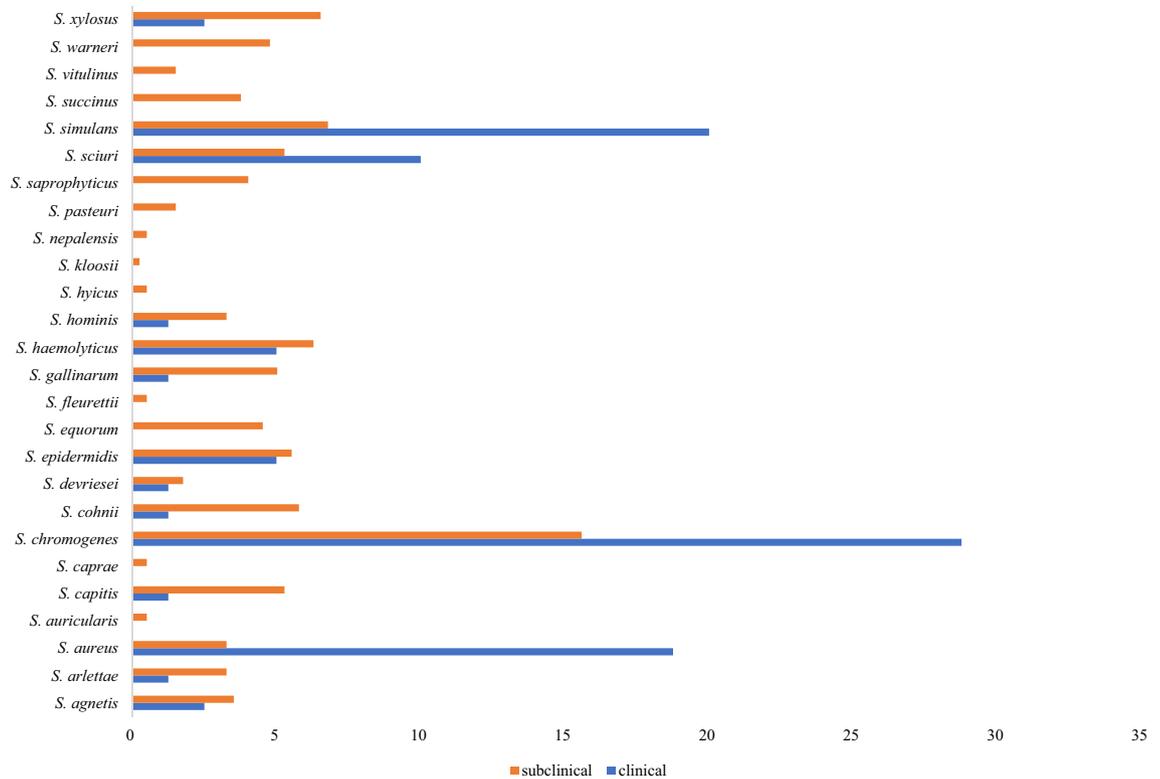


Figure 1. Frequency of staphylococcal species associated with clinical and subclinical mastitis.

emp genes found only in *S. aureus* strains. The PIA production operon *icaADBC* and its regulator *icaR* was only present in *S. aureus*, *S. chromogenes*, *S. capitis*, and *S. epidermidis*; most *S. cohnii* and *S. saprophyticus* carried the *icaC* gene. In the clinical isolates, the *icaADBC* operon and *icaR* gene were present only in *S. aureus* isolates. Other biofilm regulatory genes i.e., *rbf*, *tcaR*, *sarA* and *sigB* were found in subclinical isolates of *S. aureus*, *S. chromogenes*, *S. epidermidis* and *S. haemolyticus*, but not in *S. simulans* isolates (Supplementary Table 1).

Phylogenetic analyses reveal no clear relationship between clinical and subclinical isolates. Analysis of the 16S RNA genes from the genome sequences of the *Staphylococcus* spp. from bovine and buffalo mastitis cases revealed that the clinical and subclinical isolates (n = 478) are present in a wide variety of clades and do not show any clear relationship (Supplementary Fig. 1). The 16S RNA gene phylogeny also indicated that the mastitis related *S. aureus*, *S. epidermidis*, and *S. capitis* have a close phylogenetic relationship. These species also possess many adhesion genes (avg. no. = 26, 11, and 17 respectively), followed by *S. chromogenes* and *S. warneri* (avg. no. = 9 and 12, respectively). *S. capitis* has a close phylogenetic relationship to the species that are mainly associated with clinical mastitis (*S. aureus* and *S. epidermidis*). *S. chromogenes*, which was implicated in cases of clinical (n = 23/80) and subclinical mastitis (n = 61/398), is most closely related to *S. agnetis* and *S. hyicus* species that were only associated with subclinical mastitis. “Subclinical species” *S. saprophyticus*, *S. xylosus*, *S. gallinarum* and *S. arlettae* formed a distinct node with few strains involved in clinical mastitis and with most of these strains not carrying known adhesion/biofilm related genes. The “subclinical species” *S. warneri* and *S. pasteurii* were also phylogenetically related and carried biofilm/adhesion associated genes (n = 35; avg. no. of genes = 12 and 10, respectively). No specific pattern was observed between clinical and subclinical strains based on *ebpS*, *rbf*, *sarA*, *sasH*, *sigB*, and *tcaR* gene phylogeny (Supplementary Figs. 2–7, respectively). Generally, the clinical and subclinical strains of most species were in the same clade.

The co-phylogenetic analysis suggested that there may be movement of virulence genes by horizontal gene transfer (HGT) in staphylococcal spp. For example, the *ebpS* gene may have moved among clinical and subclinical isolates of *S. simulans*, *S. chromogenes* and *S. aureus* (Supplementary Fig. S8); the *pls* gene among clinical and subclinical isolates of *S. haemolyticus*, *S. chromogenes* and *S. simulans* (Supplementary Fig. S9); the *rbp* gene among clinical and subclinical isolates of *S. chromogenes*, *S. aureus* and *S. haemolyticus* (Supplementary Fig. S10), and the *sarA* gene, among clinical and subclinical isolates of *S. chromogenes* and *S. aureus*, respectively (Supplementary Fig. S11). Possible HGT was also observed for the *sasH* gene among *S. aureus*, *S. simulans* and *S. chromogenes* (Supplementary Fig. S12) and the *tcaR* gene among the *S. aureus*, and *S. chromogenes* clinical and subclinical isolates (Supplementary Fig. S13). Although the co-phylogenetic analysis suggested that HGT might be taking place, further investigation of the association of these virulence genes with mobile genetic elements, such as transposons and *S. aureus* pathogenicity islands (SaPIs), must be carried out to test the HGT hypothesis.

Data analysis indicates adhesion and biofilm genes exclusively related to clinical isolates. Hierarchical clustering analysis based on the presence/absence of the adhesin, biofilm, and regulatory

Role	Gene	Mastitis	
		Clinical	Subclinical
Clump factor A	<i>clfA</i>	18.7% (15)	9.05% (36)
Clump factor B	<i>clfB</i>	16.2% (13)	3.27% (13)
Collagen adhesion	<i>cna</i>	6.25% (5)	4.77% (19)
Fibronectin binding protein A	<i>fnbA</i>	21.2% (17)	7.29% (29)
Fibronectin binding protein B	<i>fnbB</i>	11.2% (9)	6.03% (24)
Elastin binding protein	<i>ebpS</i>	83.8% (67)	68.8% (274)
Staphylococcal protein A	<i>spa</i>	20.0% (16)	7.04% (28)
Ser-Asp rich fibrinogen-binding protein C	<i>sdrC</i>	17.5% (14)	9.80% (39)
Ser-Asp rich fibrinogen-binding protein D	<i>sdrD</i>	7.50% (6)	2.26% (9)
Ser-Asp rich fibrinogen-binding protein E	<i>sdrE</i>	22.5% (18)	11.5% (46)
Staphylocoagulase	<i>coa</i>	17.5% (14)	3.52% (14)
Extracellular adherence protein Eap/Map	<i>eap</i>	17.5% (14)	3.52% (14)
Extracellular matrix protein-binding protein	<i>emp</i>	17.5% (14)	3.27% (13)
Fibrinogen binding protein	<i>efb</i>	17.5% (14)	3.77% (15)
Secreted von Willebrand factor-binding protein	<i>vWbp</i>	18.7% (15)	6.78% (27)
Bifunctional autolysin	<i>atl</i>	83.8% (67)	68.3% (272)
Accumulation associated protein	<i>aap</i>	2.50% (2)	5.78% (23)
Surface protein	<i>pls</i>	55.0% (44)	42.4% (169)
Surface protein G	<i>sasG</i>	10.0% (8)	2.76% (11)
Surface protein H	<i>sasH</i>	77.5% (62)	48.4% (193)
Surface protein A	<i>sasA</i>	47.5% (38)	35.1% (140)
Surface protein C	<i>sasC</i>	30.0% (24)	30.9% (123)
Surface protein D	<i>sasD</i>	15.0% (12)	13.5% (54)
Surface protein F	<i>sasF</i>	83.8% (67)	68.0% (271)
Surface protein I	<i>sasI</i>	0.00% (0)	0.00% (0)
Surface protein K	<i>sasK</i>	8.75% (7)	2.76% (11)
Biofilm operon protein A	<i>icaA</i>	16.2% (13)	10.5% (42)
Biofilm operon protein B	<i>icaB</i>	16.2% (13)	10.5% (42)
Biofilm operon protein C	<i>icaC</i>	17.5% (14)	17.3% (54)
Biofilm operon protein D	<i>icaD</i>	16.2% (13)	10.5% (42)
Biofilm negative transcriptional regulator	<i>icaR</i>	16.2% (13)	13.5% (54)
regulation of biofilm formation	<i>rbf</i>	56.3% (45)	51.2% (204)
Transcriptional regulator involved in teicoplanin susceptibility	<i>tcaR</i>	56.3% (45)	51.2% (204)
Transcriptional regulator involved in biofilm formation process	<i>sarA</i>	56.3% (45)	51.2% (204)
RNA polymerase sigma factor B	<i>sigB</i>	56.3% (45)	51.2% (204)

Table 2. Relative and absolute frequency of adhesin, biofilm genes, and related regulatory genes of staphylococcal species associated with clinical and subclinical mastitis.

genes revealed 20 different clusters (Supplementary Table 1). One hundred and twenty-seven (26.5%) strains (13 clinical and 114 sub-clinical) of the 478 genomes evaluated lacked the 35 adhesion and biofilm-associated genes identified by the RAST annotation tool. The staphylococcal species lacking these genes included: *S. arlettae*, *S. equorum*, *S. gallinarum*, *S. sciuri*, *S. succinus*, *S. vitulinus* and *S. xylosus*. In species heatmaps (Fig. 3), the pattern of adhesion/biofilm genes in clinical isolates differed from that of sub-clinical isolates. The presence of the *clfA*, *clfB*, *fnbA*, *spa*, *sdrC*, *coa*, *eap*, *emp*, *vWbp*, *sasD*, *icaABCD* and *icaR* genes was highly correlated in clinical isolates, while in subclinical isolates, no specific gene correlations were observed (Spearman coefficient > 0.8). Based on hierarchical matrix clustering, clusters 9 and 10; 19 and 18 and 4 and 5 (Supplementary Table 1) contained most of the strains that harbored a typical pattern of nine genes (*rbf*, *pls*, *sasF*, *sarA*, *atl*, *sasH*, *sigB*, *tcaR* and *ebpS*) in both clinical and subclinical isolates. This pattern is also demonstrated in the heatmap of the gene frequency (Fig. 4).

Discussion

Staphylococcus spp. are the most common etiologic agents of mastitis, with *S. aureus* thought the most important, while coagulase-negative staphylococci and non-aureus staphylococci considered less significant⁶. Based on 16S RNA identification of the 478 available genome sequences, *S. chromogenes* (28.7%) and *S. simulans* (20.0%) were the staphylococcal species most frequently associated with clinical mastitis. *S. aureus* was the next most prevalent species (18.7%) associated with clinical mastitis and it was rarely (3.3%) associated with subclinical

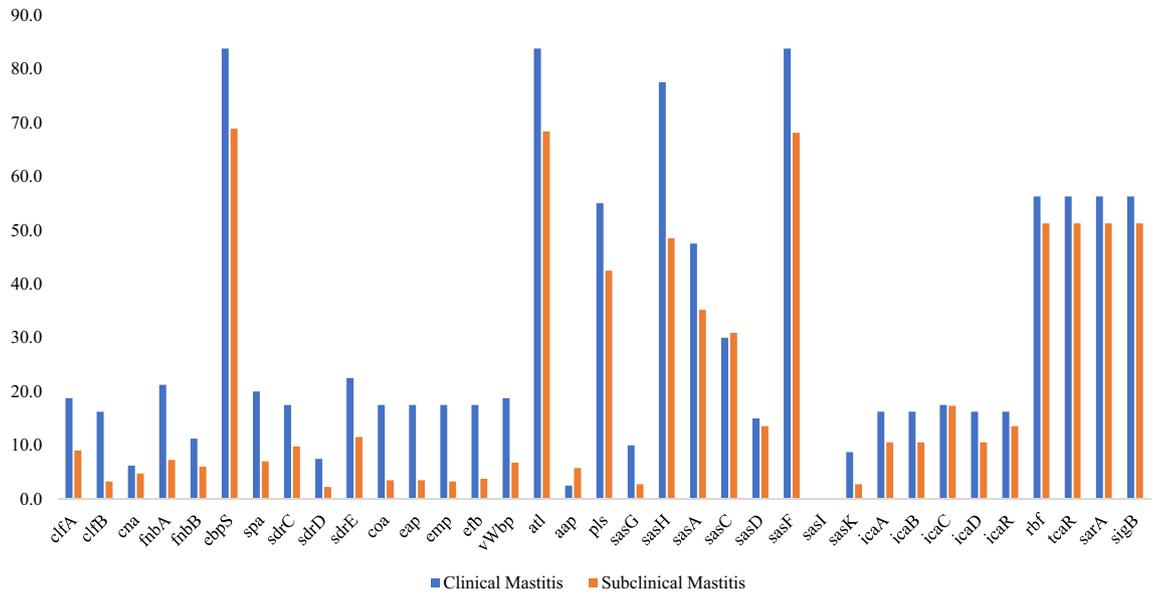


Figure 2. Frequency of adhesin and biofilm-related genes in mastitis-associated staphylococcal isolates.

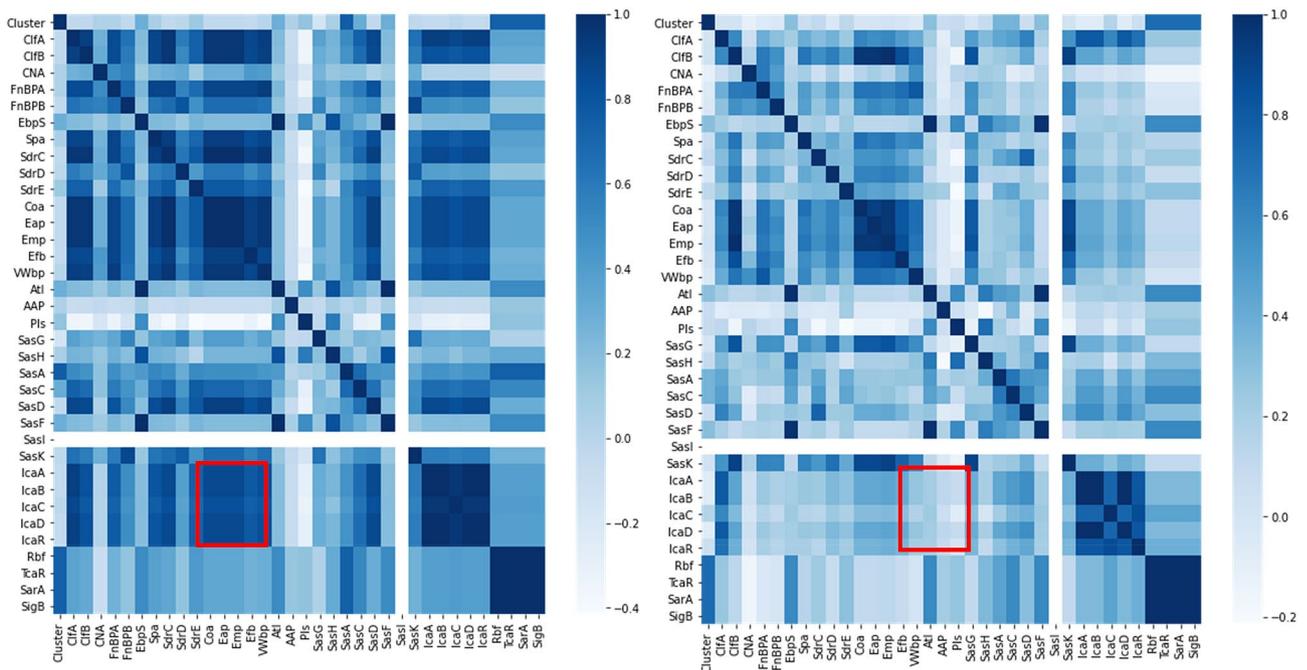


Figure 3. Heat map of adhesion and biofilm genes in clinical and subclinical staphylococcal isolates (Left—Clinical mastitis; Right—Subclinical mastitis). A higher correlation between the presence of the *icaABCD* and the *coa, eap, emp, efb* and *vWbp* was found in clinical (vs. subclinical) isolates (red box).

mastitis deposits in GenBank. Most subclinical cases were associated with *S. chromogenes* (15.6%) followed by *S. simulans* (6.8%), *S. xyloso* (6.5%), *S. haemolyticus* (6.3%), *S. cohnii* (5.8%), *S. epidermidis* (5.5%), *S. capitis* (5.3%), *S. sciuri* (5.3%), *S. gallinarum* (5.0%), *S. warneri* (4.8%), *S. equorum* (4.5%), *S. saprophyticus* (4.0%), *S. succinus* (3.8%), *S. arlettae* (3.5%), and *S. agnetis* (3.3%). These findings are consistent with a growing number of studies which report that coagulase-negative staphylococci are emerging pathogens associated with mastitis and persistence of intramammary infection in bovine worldwide⁷. As observed in this study, in a recent Canadian study, the *S. chromogenes* and *S. simulans* were among the most common species found in clinical mastitis cases⁸.

Adherence is considered the first step of staphylococcal infection and the presence of biofilm aids in the process. Accordingly, adhesion-related genes are thought to be key virulence factors⁸. In the current study, the most frequently observed adherence and biofilm-forming genes were *ebpS, atl, pls, sasH, sasF, rbf, tcaR, sarA* and *sigB* in both clinical and subclinical isolates. The genomes of some subclinical species i.e., *S. arlettae, S. succinus*,

expression is regulated by a conserved octanucleotide sequence (COS) and since they are involved in modulating the immune response to *S. aureus* infections or antibiotic, it is possible to assume that the *emp* gene would work with the *vWbp* gene in *S. aureus* immune response evasion^{27,28}. The *eap* gene product has recently been shown to suppress the formation of “neutrophil extracellular traps” (NETs), which are thought to function as a neutrophil-mediated extracellular trapping mechanism²⁹.

The *icaABCD* operon is the most studied *Staphylococcus* biofilm forming genes and it is most frequently reported in mastitis isolates highlighting their potential to form biofilm³⁰. In the current study, the *icaC* gene was the most prevalent in clinical and subclinical isolates, in contrast to a previous report of which *icaA* and *icaD* are the most prevalent³¹. Also, the finding of this study observed that coagulase negative *Staphylococcus* (CoNS) and *S. aureus* possessed the *icaA* and/or *icaD* gene in contrast with a previous findings that the *icaA* was only observed in CoNS strain while the *icaD* was found both in CoNS and *S. aureus*. The most prevalent biofilm regulatory genes detected were: *rbf*, *tcaR*, *sarA* and *sigB* which is in agreement with previous finding of high presence of *sarA*, *tcaR* in *S. aureus* from bovine subclinical mastitis isolates³². The *rbf* gene is an important biofilm regulatory gene and its inactivation results in a biofilm negative phenotype³³. It has recently been shown that *rbf* mutants exhibit significantly increased pathogenicity compared to the wild type *S. aureus* strains³⁴ suggesting an important role in host adaptation. The *rbf* gene product negatively regulates hemolytic activity by repressing the expression of the *hla* and *psmA* genes. It also upregulates *sarX*, which, in turn, activates the *icaADBC* locus leading to biofilm production³⁵.

The *tcaR* gene increases the production of PIA by regulating the expression of the *icaADBC* operon and the *spa*, *sasF* and *sarS* genes³⁵. Given the high frequency of the *sasF* gene observed in this study, detection of its transcriptional regulators was not unexpected. The *sarA* family of transcription regulators proteins are responsible for controlling many target genes involved in virulence. Most notably, SarA is responsible for regulating the *agr* loci, which is a pivotal regulator of virulence in *S. aureus*³⁶. The presence of the *sarA* gene in mastitis was observed in a recent study in all of the 84 *S. aureus* isolates from mastitis cases in Xinjiang, China³⁶. The rRNA polymerase sigma factor (SigB) has a central role in stress homeostasis. This protein contributes to the synthesis of several virulence determinants defining staphylococcal pathogenesis, including the transcriptional activation of many surface proteins (such as *clfA* and *fnbA*) while downregulating the production of secreted toxins and proteases (such as *Aur*, *SspA*, *SspB*)³⁷.

Phylogenetic analysis of *Staphylococcus* spp. related to mastitis has been done before, studies to date have focused on comparatively few isolates and mainly on *S. aureus* and observed that strains that had different origin were clustered together³⁸. In this study, phylogenetic analysis of the 16S RNA genes indicated that *S. aureus*, *S. epidermidis*, *S. caprae* and *S. capitis* have a close relationship as observed in human clinical isolates³⁹. Also, these species all possessed and shared a large number of adhesion genes. In a previous study, some authors have suggested that dairy cows can be subclinically infected with *S. aureus* subtypes that can cause clinical mastitis if the right conditions are present³⁸. In the current study, some clinical and subclinical strains clustered together based on their 16S RNA sequences, but they had different biofilm and coagulase gene contents. Also, this study shows that *S. chromogenes* isolates from cases of clinical and subclinical mastitis were closely related to *S. agnetes* and *S. hyicus* suggesting that these species could also demonstrated the same potential to become an emerging mastitis agent as *S. chromogenes*⁴⁰.

S. aureus is reported to acquire and disseminate SaPIs through HGT events mediated by phages⁴¹. Moreover, *S. aureus* colonization of different host species is known to be facilitated by the HGT of virulence factors across different staphylococcal species⁴². It is further known that biofilm growth can increase the rate of HGT of virulence determinants such as antibiotic resistance genes⁴³. In this study, the co-phylogenetic analysis suggested that HGT amongst clinical and subclinical isolates of *S. chromogenes*, *S. aureus*, and *S. simulans* (mainly *ebpS*, *rbp*, *sarA*, *tcaR*, *pls*) and *sigB* gene in *S. aureus* may occur. Moreover, the phylogenetic relationship of the adhesion and biofilm genes: *ebpS*, *sasH*, *atl*, *sarA*, *rbf* and *tcaR* are different from 16S phylogenetic distribution. This finding is consistent with the notion that HGT occurs among clinical and subclinical isolates⁴⁴. It is therefore tempting to speculate that virulence factors may arise in staphylococcal species not generally associated with clinical mastitis by known *Staphylococcus* HGT mechanisms, but further study is needed to demonstrate this.

Although the arbitrary source of isolates used (478 staphylococcal spp. isolated from clinical and subclinical mastitis from Brazil, Canada, India, Netherlands, and United States which had sequenced) might have introduced some biases, a number of the adhesins, biofilm, and related regulatory genes identified in this study might be useful for virulence profiling or as targets for vaccine development for mastitis-related staphylococcus species.

Methods

Genomic data. The genomes of *Staphylococcus* spp. from clinical (n = 80) and subclinical (n = 398) mastitis cases worldwide were downloaded from the National Center for Biotechnology Information (NCBI). An initial advanced search of the NCBI Biosample database I with “mastitis” and “staphylococcus” as keywords resulted in 925 entries. After this initial step, only complete genomes that were identified as mastitis isolates from *Bos taurus* or *Bubalus bubalis* and which were the sole agent associated with the diseases were evaluated. Also, for more detailed information, the publications or their BioSample descriptions were evaluated (Supplementary Table 1). It was assumed that mastitis states were classified according to the clinical presentation and standard triage test described by Radostits et al.³. Genomes in this study were from bacteria isolated in Brazil⁴⁵, Canada⁴⁶, India, Netherlands, and United States (information regarding isolates BioSample available on Supplementary Table 1).

Genome annotation and adhesion-related gene identification. Genomes were annotated using Rapid Annotation using Subsystem Technology (RAST)^{47,48}. The sequences of the genes classified as adhesion/adhesins or implicated in biofilm formation and their respective regulatory genes were downloaded and ana-

lyzed manually. The genes were considered to encode adhesins or be play a role in biofilm formation based on their classification in the VFDB reference database for bacterial virulence factors⁴⁹ and/or in the RAST annotation engine⁴⁷. 16S rRNA gene sequences were obtained from the complete genomes using the Basic Rapid Ribosomal RNA Predictor (Barrnap) v 0.9 (<https://github.com/tseemann/barrnap>).

Data analysis. The presence or absence of selected genes was used in hierarchical clustering analysis with PAST software v4.03⁵⁰. Clusters of the isolates were created based on the most and least frequent genes. The Spearman test was used to analyze the correlation between the presence/absence of adhesin and biofilm genes in both clinical and subclinical mastitis isolates (A coefficient close to 1.0 indicates a high correlation). Gene profiling by frequency heatmaps was calculated using Numpy v1.20.3⁵¹. Graphs were made using Matplotlib v3.4.2⁵² and, when needed, with R-software v4.1.0⁵³. The statistical significance of gene presence and mastitis state was obtained by logistic regression with R software.

Phylogenetic analysis and tree construction. The phylogenetic correlation of the 16S RNA, *araC*, *pls*, *sasF*, *sarA*, *atl*, *sasH*, *sigB* *tcaR*, and *ebpS* genes was determined and phylogenetic trees were generated with maximum likelihood approach using IQ-TREE2 from MAFFT v7⁵⁴ gene alignment using default parameter. For clade support, we performed ultrafast bootstrap analyses with 1000 pseudoreplicates implemented in IQ-tree⁵⁵. Tree visualization was done with iTOL v5⁵⁶. The co-phylogenetic tree construction was done using phytools v0.7-80 in R-software v4.1.0⁵³. In order to create a rooted tree, the closely related *Escherichia coli* strains (accession numbers): 2014C-3057 (NZ_CP027387.1); 2015C-4944 (NZ_CP027390.1); 2013C-4538 (NZ_CP027582.1); E2865 (NZ_AP018808.1); 97-3250 (NZ_JHEW00000000.1); FORC_028 (NZ_CP012693.1); 2013C-4225 (NZ_CP027577.1); 2014C-3050 (NZ_CP027472.1); 2012C-4606 (NZ_CP027352.1) and CFSAN027343 (NZ_CP037943.1) were used as an outgroup.

Received: 17 April 2021; Accepted: 10 August 2021

Published online: 30 August 2021

References

- Rocha, L. S. *et al.* Comparative genomics of *Staphylococcus aureus* associated with subclinical and clinical bovine mastitis. *PLoS ONE* **14**, 1–19 (2019).
- Schukken, Y. H. *et al.* Host-response patterns of intramammary infections in dairy cows. *Vet. Immunol. Immunopathol.* **144**, 270–289 (2011).
- Radostits, O. M., Gay, C. C., Hinchcliff, K. W. & Constable, P. D. Diseases of the mammary gland. In *Veterinary Medicine—A Textbook of the Diseases of Cattle, Horses, Sheep, Pigs, and Goats*. 673–763 (2007).
- Efthimiou, G., Tsiamis, G., Typas, M. A. & Pappas, K. M. Transcriptomic adjustments of *Staphylococcus aureus* COL (MRSA) forming biofilms under acidic and alkaline conditions. *Front. Microbiol.* **10**, 2393 (2010). <https://doi.org/10.3389/fmicb.2019.02393>
- Åvall-Jääskeläinen, S. *et al.* Comparative genome analysis of 24 bovine-associated *Staphylococcus* isolates with special focus on the putative virulence genes. *PeerJ* **6**, e4560 (2018).
- Wald, R., Hess, C., Urbantke, V., Wittek, T. & Baumgartner, M. Characterization of staphylococcus species isolated from bovine quarter milk samples. *Animals* **9**, 1–16 (2019).
- Corbeil, A. *et al.* Short communication: Search for superantigen genes in coagulase-negative staphylococci isolated from bovine milk in Canada. *J. Dairy Sci.* **102**, 2008–2010 (2019).
- Zuniga, E. *et al.* Occurrence of genes coding for MSCRAMM and biofilm-associated protein Bap in *Staphylococcus* spp. isolated from bovine subclinical mastitis and relationship with somatic cell counts. *Microb. Pathog.* **89**, 1–6 (2015).
- Oikonomou, G. *et al.* Milk Microbiota: What are we exactly talking about?. *Front. Microbiol.* **11**, 1–15 (2020).
- Taponen, S. *et al.* Bovine milk microbiome: A more complex issue than expected. *Vet. Res.* **50**, 1–15 (2019).
- Nováková, D. *et al.* *Staphylococcus equorum* and *Staphylococcus succinus* isolated from human clinical specimens. *J. Med. Microbiol.* **55**, 523–528 (2006).
- Kengkoom, K. & Ampawong, S. *Staphylococcus sciuri* associated to subcutaneous abscess and dermatitis in ICR mouse. *Arq. Bras. Med. Vet. e Zootec.* **69**, 117–122 (2017).
- Latham, R. H., Running, K. & Stamm, W. E. Urinary tract infections in young adult women caused by *Staphylococcus saprophyticus*. *JAMA J. Am. Med. Assoc.* **250**, 3063–3066 (1983).
- Yang, F. *et al.* Short communication: Detection and molecular characterization of methicillin-resistant *Staphylococcus aureus* isolated from subclinical bovine mastitis cases in China. *J. Dairy Sci.* **103**, 840–845 (2019).
- Ahangari, Z., Ghorbanpoor, M., Shapouri, M. R. S., Gharibi, D. & Ghazvini, K. Methicillin resistance and selective genetic determinants of *Staphylococcus aureus* isolates with bovine mastitis milk origin. *Iran. J. Microbiol.* **9**, 152–159 (2017).
- Kot, B., Szweđa, P., Frankowska-maciejewska, A. & Wolska, K. Virulence gene profiles in *Staphylococcus aureus* isolated from cows with subclinical mastitis in eastern Poland. *J Dairy Res.* **83**(2), 228–235. <https://doi.org/10.1017/S002202991600008X> (2016) (**Epub 2016 Apr 1**).
- Singh, V. K. High level expression and purification of *atl*, the major autolytic protein of *Staphylococcus aureus*. *Int J Microbiol.* **2014**, 615965. <https://doi.org/10.1155/2014/615965> (2014) (**Epub 2014 Jan 19**).
- Josse, J., Laurent, F. & Diot, A. Staphylococcal adhesion and host cell invasion: Fibronectin-binding and other mechanisms. *Front. Microbiol.* **8**, 1–8 (2017).
- Porayath, C. *et al.* Autolysin mediated adherence of *Staphylococcus aureus* with fibronectin, gelatin and heparin. *Int. J. Biol. Macromol.* **110**, 179–184 (2018).
- Thammavongsa, V., Missiakas, D. M. & Schneewind, O. *Staphylococcus aureus* degrades neutrophil extracellular traps to promote immune cell death. *Bone* **23**, 1–7 (2013).
- Fursova, K. *et al.* Virulence factors and phylogeny of *Staphylococcus aureus* associated with bovine mastitis in Russia based on genome sequences. *Front. Vet. Sci.* **7**, 1–10 (2020).
- Hoekstra, J. *et al.* Genomic analysis of European bovine *Staphylococcus aureus* from clinical versus subclinical mastitis. *Sci. Rep.* **10**, 1–11 (2020).

23. Jenkins, A. *et al.* Differential expression and roles of *Staphylococcus aureus* virulence determinants during colonization and disease. *MBio* **6**, 1–10 (2015).
24. Akers, R. M. & Nickerson, S. C. Mastitis and its impact on structure and function in the ruminant mammary gland. *J. Mammary Gland Biol. Neoplasia* **16**, 275–289 (2011).
25. Javid, F. *et al.* Molecular typing of *Staphylococcus aureus* based on coagulase gene. *Vet. World* **11**, 423–430 (2018).
26. Trivedi, U. *et al.* *Staphylococcus aureus* coagulases are exploitable yet stable public goods in clinically relevant conditions. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E11771–E11779 (2018).
27. McCarthy, A. J. & Lindsay, J. A. Genetic variation in *Staphylococcus aureus* surface and immune evasion genes is lineage associated: Implications for vaccine design and host-pathogen interactions. *BMC Microbiol.* **15**(10), 173. <https://doi.org/10.1186/1471-2180-10-173> (2010).
28. Harraghy, N., Homerova, D., Herrmann, M. & Kormanec, J. Mapping the transcription start points of the *Staphylococcus aureus* *eap*, *emp*, and *vWb* promoters reveals a conserved octanucleotide sequence that is essential for expression of these genes. *J. Bacteriol.* **190**, 447–451 (2008).
29. Eisenbeis, J. *et al.* The *Staphylococcus aureus* extracellular adherence protein *Eap* is a DNA binding protein capable of blocking neutrophil extracellular trap formation. *Frontiers in Cellular and Infection Microbiology* **8**:235 (2018). <https://doi.org/10.3389/fcimb.2018.00235>
30. Azmi, K., Qrei, W. & Abdeen, Z. Screening of genes encoding adhesion factors and biofilm production in methicillin resistant strains of *Staphylococcus aureus* isolated from Palestinian patients. *BMC Genomics* **20**, 1–12 (2019).
31. Salina, A., Guimarães, F. F., Richini Pereira, V. B., Menozzi, B. D. Rall, V. L. M. & Langoni, H. Detection of *icaA*, *icaD*, and *bab* genes and biofilm production in *Staphylococcus aureus* and non-*aureus* staphylococci isolated from subclinical and clinical bovine mastitis. *Arq. Bras. Med. Veterinária e Zootec.* **72**, 1034–1038 (2020).
32. Ren, Q., Liao, G., Wu, Z., Lv, J. & Chen, W. Prevalence and characterization of *Staphylococcus aureus* isolates from subclinical bovine mastitis in southern Xinjiang, China. *J. Dairy Sci.* **103**, 3368–3380 (2020).
33. Rowe, S. E. *et al.* AraC-type regulator Rbf controls the *Staphylococcus epidermidis* biofilm phenotype by negatively regulating the *icaADBC* repressor SarR. *J. Bacteriol.* **198**, 2914–2924 (2016).
34. Fang, B., Liu, B. & Sun, B. Transcriptional regulation of virulence factors *Hla* and phenol-soluble modulins α by AraC-type regulator *rbf* in *Staphylococcus aureus*. *Int. J. Med. Microbiol.* **310**, 151436 (2020).
35. Hoang, T. M. *et al.* Transcriptional regulation of *icaADBC* by both *icaR* and *tcaR* in *Staphylococcus epidermidis*. *J. Bacteriol.* **201**, 1–17 (2019).
36. Ballal, A. & Manna, A. C. Expression of the *sarA* family of genes in different strains of *Staphylococcus aureus*. *Microbiology* **155**, 2342–2352 (2009).
37. Marbach, H. *et al.* Within-host evolution of bovine *Staphylococcus aureus* selects for a SigB-deficient pathotype characterized by reduced virulence but enhanced proteolytic activity and biofilm formation. *Sci. Rep.* **9**, 1–12 (2019).
38. Ronco, T. *et al.* Genomic investigation of *Staphylococcus aureus* isolates from bulk tank milk and dairy cows with clinical mastitis. *Vet. Microbiol.* **215**, 35–42 (2018).
39. Watanabe, S. *et al.* Complete genome sequencing of three human clinical isolates of *Staphylococcus caprae* reveals virulence factors similar to those of *S. epidermidis* and *S. capitis*. *BMC Genomics*. **19**(1), 810 (2018). <https://doi.org/10.1186/s12864-018-5185-9>.
40. dos Santos, D. C. *et al.* *Staphylococcus chromogenes*, a coagulase-negative *Staphylococcus* species that can clot plasma. *J. Clin. Microbiol.* **54**, 1372–1375 (2016).
41. Tormo, M. Á. *et al.* *Staphylococcus aureus* pathogenicity island DNA is packaged in particles composed of phage proteins. *J. Bacteriol.* **190**, 2434–2440 (2008).
42. McCarthy, A. J. *et al.* Extensive horizontal gene transfer during *Staphylococcus aureus* co-colonization *in vivo*. *Genome Biol. Evol.* **6**, 2697–2708 (2014).
43. Savage, V. J., Chopra, I. & O'Neill, A. J. *Staphylococcus aureus* biofilms promote horizontal transfer of antibiotic resistance. *Antimicrob. Agents Chemother.* **57**, 1968–1970 (2013).
44. Xue, H., Lu, H. & Zhao, X. Sequence diversities of serine-aspartate repeat genes among *Staphylococcus aureus* isolates from different hosts presumably by horizontal gene transfer. *PLoS One* **2011**;6(5):e20332. <https://doi.org/10.1371/journal.pone.0020332>.
45. Pizauro, L. J. L. *et al.* Complete Genome Sequences of 11 *Staphylococcus* spp. strains isolated from buffalo milk and milkers' hands. *Microbiol. Resour. Announc.* **8**(47), e01264-19. <https://doi.org/10.1128/MRA.01264-19> (2019).
46. Naushad, S. *et al.* Comprehensive phylogenetic analysis of bovine non-*aureus* staphylococci species based on whole-genome sequencing. *Front. Microbiol.* **20**(7), 1990 (2016). <https://doi.org/10.3389/fmicb.2016.01990>.
47. Overbeek, R. *et al.* The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **42**, 206–214 (2014).
48. Aziz, R. K. *et al.* The RAST server: Rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008).
49. Chen, L. *et al.* VFDB: A reference database for bacterial virulence factors. *Nucleic Acids Res.* **33**, 325–328 (2005).
50. Hammer, Ø., David A. T. & Harper, P. D. R. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electron.* **4**, 1–9 (2001).
51. Van Der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy array: A structure for efficient numerical computation. *Comput. Sci. Eng.* **13**, 22–30 (2011).
52. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 99–104 (2007).
53. R Core, T. R. *A Language and Environment for Statistical Computing*, vol. 2. <http://www.R-project.org/> (2004).
54. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
55. Minh, B. Q. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
56. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).

Author contributions

L.J.L.P and A.M.V. conceived the experiment. L.J.L.P., C.C.A. and S.R.S. conducted the experiment, J.I.M., A.M.K., L.F.Z. and F.A.A. helped elaborating the draft. All authors contributed in the analysis of the results and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-96842-2>.

Correspondence and requests for materials should be addressed to L.J.L.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021